

An Evaluation of CNN-based Liver Segmentation Methods using Multi-types of CT Abdominal Images from Multiple Medical Centers

Hong Son Hoang¹, Cam Phuong Pham², Daniel Franklin^{3§}, Theo van Walsum^{4§}, and Manh Ha Luu^{1,4§*}

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

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

³SEDE/FEIT, University of Technology, Sydney, Australia

⁴BIGR, Department of Radiology and Nuclear Medicine, Erasmus MC, Rotterdam, the Netherlands

§ IEEE member

* Corresponding author email: halm@vnu.edu.vn

Abstract—Automatic segmentation of CT images has recently been applied in several clinical liver applications. Convolutional Neural Networks (CNNs) have shown their effectiveness in medical image segmentation in general and also in liver segmentation. However, liver image quality may vary between medical centers due to differences in the use of CT scanners, protocols, radiation dose, and contrast enhancement. In this paper, we investigate three well known CNNs for liver segmentation using data from several medical centers: FCN-CRF, DRIU, and V-net. We perform qualitative evaluation of the CNNs based on Dice score, Hausdorff distance, mean surface distance and false positive rate. The results show that all three CNNs achieved a mean Dice score of over 90% in liver segmentation with typical contrast enhanced CT images of the liver. The results also demonstrate that those CNNs have reduced performance in liver segmentation in the case of low-dose and non-contrast enhanced CT images. In conclusion, these promising results enable further investigation of alternative deep learning based approaches to liver segmentation using CT images.

Index Terms—liver segmentation, CT images, U-net, V-net, low-dose

I. INTRODUCTION

Liver cancer is the sixth most common cancer worldwide [1], with a high incidence in developing countries in Eastern Asia, South-Eastern Asia, Northern Africa and Southern Africa [2]. Liver cancer is one of the most common causes of death from cancer in Vietnam [3]. Less than 15% of patients with liver cancer can survive without treatment for more than 5 years [4]. The size and shape of the liver varies considerably from patient to patient. Clinical assessment of liver cancer and treatment planning therefore requires accurate knowledge of the liver of each individual patient. For instance, in liver surgery, surgeons require precise liver segmentations before making the decision to excise the liver segment(s) containing the tumors [5]. Liver segmentation is also used in image registration techniques in RFA liver intervention [6], [7], and for delineating regions of interest for liver tumor segmentation [8]. Conventionally, a liver segmentation can be created by annotating the liver and liver lesion by on a slice-by-slice,

which is time-consuming and complicated [5]. Therefore, there is a need for the use of computer-based liver segmentation methods in clinical practice [9]. However, liver segmentation from CT volumes is a challenging task due to the low intensity contrast between the liver and other neighboring organs [5]. Also, the quality of CT images may differ between medical centers because of variations in the use of CT scanners, as well as the amount of injected contrast agent and radiation dose in each particular application (see Figure 1). Therefore, a robust automatic liver segmentation method, although greatly needed, also is challenging to implement in practice, and this problem has recently become an active area of research.

Several liver segmentation methods have been proposed in the literature, including region growing, intensity thresholding, graph cut, and deformable models [9], [10]. Nevertheless, these methods are based on hand-crafted features, and thus have limited feature representation capability. Recently, Convolution Neural Networks (CNNs), a typical type of deep learning neural network, have achieved great success in a wide range of medical imaging problems such as classification, segmentation and object detection, achieving state-of-the-art performance comparable to human oncologists/radiologists [11], [12]. One of the reasons for this success is that CNNs have the ability to learn a hierarchical representation of images, without the need for handcrafted features [13]. In the liver segmentation task, CNN-based segmentation methods have been shown to outperform classical statistical and image-processing approaches [8], [11], [14]. Ronneberger et al. (2015) introduced the well-known U-net architecture [11], and Christ et al. (2016) applied this CNN to segment the liver [8] (see Figure 2). Christ et al. (2017) proposed a fully convolutional neural network (FCN) combined with conditional random field (CRF) to segment the liver in both CT and MRI images, and a mean Dice score of 94% was reported [14]. Li et al. (2018) created the H-dense U-net by combining a 2D dense U-net and a 3D counterpart and reported a Dice score of 96.5% for liver segmentation [15]. Bellver et al. (2017)

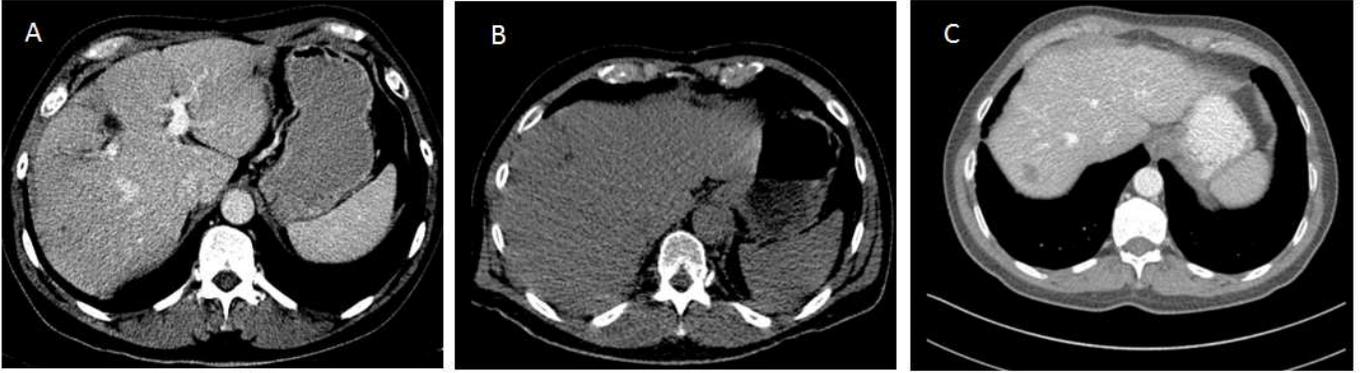


Fig. 1. Examples of CT image of livers: a low-dose contrast enhanced CT (A), a low-dose non-contrast enhanced CT (B) and a contrast enhanced CT (C). Those images were acquired from two medical centers, yielding the large differences in the quality of the liver CT image.

modified the original OSVOS neural network to segment the liver [16] and achieved a Dice score of 94%. In general, those CNN based liver segmentation methods can be classified into two categories: 2D fully convolutional neural networks (2D FCNs) and 3D fully convolutional neural networks (3D FCNs) [15]. 2D FCN methods [8], [14], [16] perform on a single slice or three continuous slices extracted from a 3D volume as the input images. The final 3D segmentation volume is created by stacking the 2D segmentation outputs in the corresponding order. 3D FCNs [13], [15], [17] utilize 3D convolutional filter kernels instead of 2D convolutional filter kernels, and the input is a complete 3D volume. In contrast to 2D FCNs, 3D FCNs can use 3D spatial information for segmentation; however, this comes at the cost of higher computational complexity and GPU memory. Theoretically, the high memory consumption enables a reduction in the depth of the network and the filter’s field-of-view, which are supposed to be the main factors for performance improvements [18]. However, the performance of 3D FCN versus 2D FCN in the task of liver segmentation is still under debate [19].

One of the well-known characteristics of CNNs is that a huge amount of data is required in the training stage to achieve high segmentation performance [20]. However, large datasets of suitable medical images are generally not readily available (due to privacy concerns) and CT images of the liver are often large - potentially in excess of one gigabyte - leading to limited availability of training data, often only originating from one or a few medical centers and thus potentially limiting the performance and generality of the developed methods. Based on our study of the literature, most related works train their model and evaluate their methods on just one or two datasets; most of them are from the MICCAI grand challenges [21]. In practice, besides the contrast enhanced CT images typically used, the dose and contrast agent awareness in use produces multiple types of CT image of the liver [7]. Therefore, we intend to investigate how well those methods perform on a larger variety of liver CT datasets. In this paper, our main contributions are:

- Firstly, we implement three well-known state-of-the-art

- CNN architectures, Cascaded-FCN [8], [11], [14], V-net [13] and DRIU [16], and train DRIU network for the task of liver segmentation using a multi-centers dataset of CT images of the liver.

- Secondly, we evaluate those methods on four CT datasets from four medical centers/sources including contrast enhanced CT, non-contrast enhanced CT and low-dose contrast enhanced CT images which are used in several clinical applications [7]. The details of each method, datasets, and evaluation metrics are described in detail in the next section.

II. METHODS

A. Neural network architectures:

1) *Cascaded Fully CNNs (CFCN) with conditional random fields (CRF)*: The CFCN introduced by Christ et al. (2017) contains two U-net networks to segment the liver and liver tumors [8]. In this study, we only implement the first U-net for liver segmentation. The key idea of the U-net architecture is that it has the ability to learn a hierarchical representation of the training image in 2D [14]. It contains 19 layers divided into two sections: the encoder and the decoder. The encoder acts a classifier for the contextual information of the pixels in the image, while the decoder, comprising connections from the layers in the encoder, provides spatial information regarding the pixels. Given a 2D input slice, the output of the U-net is a 2D probability map as a soft prediction for each corresponding pixel in the original slice. For the optimization process, cross entropy CE is used as the objective loss function:

$$CE = - \sum_i^C t_i \log(s_i) \quad (1)$$

where C are the two classes of liver and non-liver regions, t_i is ground truth and s_i is soft prediction score at the location i . Next, a 3D-dense conditional random field (CRF) is applied to combine the 2D probability maps, enabling consideration of both spatial coherence and appearance information [8].

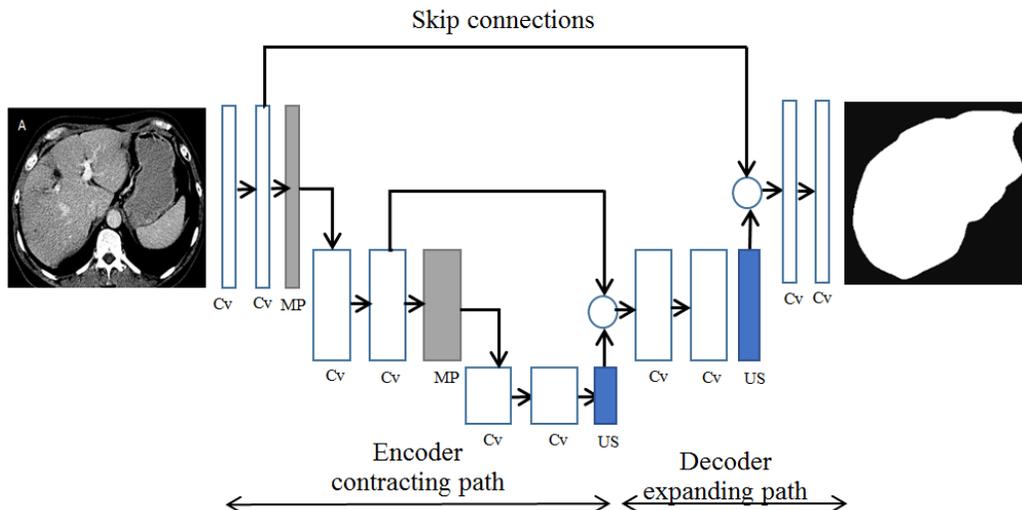


Fig. 2. A well-known CNNs architecture, U-net, is designed to automatically segmentat the liver from CT images. The 3-level neural network architecture contains two parts: the encoder and the decoder. The contracting path acts to classify pixels of the 2D image while the expanding path is performs matching between max pooling layers (MP) and upsampling layers (US) to provide locations of the classified pixels in the original image. The figure is adapted from [12].

2) *V-Net: Fully CNNs for Volumetric Medical Image Segmentation*: The key idea of the V-net is that while most CNNs are only able to process 2D images, the V-net is able to segment 3D volumes using volumetric convolutions and fully convolutional neural networks [13], [15], [17]. Similar to the U-net architecture, V-net also contains two paths: the down sampling (encoding) path of the network consists of a compression path, which is followed by the up sampling (decoding) path that decompresses the feature map until it reaches the original size of the input volume. The direct connections from the encoding to the decoding path provide location information and hence improve the accuracy of the final segmentation prediction. In this study, Dice loss is used as the objective function for the optimizer [13]:

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

where p_i and g_i are voxel values of the predicted segmentation and the ground truth, respectively, and N is the number of voxels in the volumes. Note that the segmentation and the ground truth have the same size.

3) *DRIU: Deep retinal image understanding*: DRIU was first used by Bellver et al. (2017) for liver segmentation using CT images [16]. The network architecture is based on VGG-16 [16] without the fully-connected layers, but still containing fully convolutional layers, ReLU, and max-pooling layers. Similar to U-net, the DRIU network consists of a set of paired convolutional layers, each having the same size of feature map, followed by max-pooling layers. The deeper layers of the network may capture more abstract information at a coarser scale. In contrast, in the more shallow layers, the network is able to capture feature maps that work at a

higher resolution which contain local spatial information of the object. In the end, DRIU combines the all feature maps by resizing and linearly adding them into a single output image. In this way, the final output segmentation contains information of the object at multiscale resolution. In this work, weighted Binary Cross Entropy CEw is used as objective loss function as in [16]:

$$CEw = - \sum_i^C w_i t_i \log(s_i) \quad (3)$$

Where w_i , with $\sum_i^C w_i = 1$, are the weights which balance the relative importance of the pixel classes.

B. Data

In this study, we collected four datasets of CT images from multiple sources/medical centers, containing contrast enhanced, non-contrast enhanced, and low-dose CT images of the liver. All of the datasets were anonymized by their own cite before involving in this study. The first dataset is from the Liver Tumour Segmentation (LiTS) challenge in the MICCAI grand challenge in NIFTI format [21]. The images were acquired on a variety of CT scanners and protocols from several medical centers, with in-plane spatial resolution varying from 0.55 mm to 1.0 mm, slice spacing varying from 0.45 mm to 6.0 mm, and the number of slices between 74 and 986. We use 115 labelled images, divided into two subsets: 105 for training as similar in [16] and 15 for testing. The second dataset is randomly selected from the Mayo Clinic (Mayo) with 10 images acquired on a Siemens CT scanner under full radiation dose protocol. The images have in-plane resolutions between 0.64 mm and 0.84 mm and slice spacing of 3 mm. The original images were cropped in the z dimension

in order to reduce the number of slices such that the liver is preserved, resulting in the number of slices being between 46 and 112. The images were acquired at 100 kVP, with CTDIvol of 18-21 mGy. The third and the fourth dataset are randomly selected from Erasmus MC with 15 patients scanned by Siemens scanners with low-dose protocol [22]. 15 data of these are contrast enhanced (EMC_LD) and 15 data are non-contrast enhanced CT images (EMC_NC_LD). The in-plane resolution of those images is between 0.56 and 0.89 mm, and slice thickness is between 2 mm and 5 mm, with the number of slices between 27 and 68 for the contrast dataset and from 21 to 89 for the non-contrast dataset. The images were acquired during radio frequency ablation intervention at 80-120 kVP, with CTDIvol of 4-9 mGy, resulting in noisy images due to the low radiation dose (see Figure 1). The dataset from Erasmus MC and Mayo were annotated by two experts for the ground truth, which is used in the evaluation sections (Section 3).

C. Implementation

We have implemented DRIU and V-net using Python 3 and the FCN-CRF network using Python 2. We used the Tensorflow 1.18 platform, and CUDA version 9.1.

The DRIU network was trained on a Linux PC with Intel Core i9 9900K CPU, 8 cores, clock of 3.6-5 GHz; 16 MB cache memory, NVIDIA Titan V GPU (11 GB RAM version), 64 GB RAM, and PSU Seasonic 1000W. The parameter setup is inherited from the original work of Bellver et al [16] with the batch size of 1; 15000 to 50000 iterations for each channel; the initial learning rate of 10⁻⁸ and Momentum optimizer. Training time on the 105 training dataset was approximately 2 days.

For the FCN-CRF network, we modified the source code from [8] to obtain a complete pipeline for 3D liver segmentation and reutilized its trained model. Meanwhile, we implemented V-net and reused the trained model on the same LiTS dataset, based on the source code and introduction from Chen's website <https://github.com/junqiangchen/LiTS—Liver-Tumor-Segmentation-Challenge>

D. Evaluation Criteria

1) *Dice score*: We use Dice score (DSC) to evaluate the liver segmentation performance. Given a segmentation X and ground truth Y , DSC is defined as:

$$DSC = \frac{2|X \cap Y|}{|X \cup Y|} \quad (4)$$

where $|\cdot|$ is an operator to count number of segmentation/ground truth voxels in the intersection domain or the union domain. DSC reaches a maximum value of 1 when the predicted segmentation X perfectly matches the ground truth Y . In contrast, the DSC is minimized when X and Y do not overlap at all.

2) *Hausdorff distance and mean surface distance*: Let U and V be two boundaries of liver segmentation and

ground truth, respectively. We define their Hausdorff distance $d_H(U, V)$ by:

$$d_H(U, V) = \max \left\{ \sup_{u \in U} \inf_{v \in V} d(u, v), \sup_{v \in V} \inf_{u \in U} d(u, v) \right\} \quad (5)$$

where \sup represents the supremum and \inf denotes the infimum. Mean surface distance $d_M(U, V)$ is defined as following:

$$d_M(U, V) = \frac{1}{|V|} \sum_{v \in V} \inf_{u \in U} d(u, v) \quad (6)$$

3) *False positive rate*: The False positive rate (FPR) can be used to evaluate false positive segmentation outside the ground truth. It can be formulated as following:

$$FPR(X, Y) = 100 \times \frac{|X \setminus Y|}{|Y|} \quad (7)$$

where $X \setminus Y$ denotes the part of X does not overlap with Y . Results of evaluation using these criteria are reported in the next section.

E. Results and Discussion

The evaluation scores of the three CNN architectures are summarized in Table I. FCN-CRF and DRIU perform very well on the LiTS dataset, with both achieving a mean Dice score of over 90% (see the first row cluster of Table I), the threshold for success used in other applications [7]. Those results are similar to the result reported in the original works [8], [16]. In contrast, V-net shows poor performance on this dataset, achieving a mean Dice score of 73.65%. By visually checking the data, we see that the predicted segmentations by V-net contain a large number of non-connected components in the area outside the liver, include the the areas of the spleen, the stomach, etc. These false positive segmentations result in the high FPR score of 19.2%. We hypothesize that a post-processing step may help to eliminate these false positive segmentations and thus further improve the segmentation result. Due to the large volume size, the three networks require a period of 30 seconds to almost 3 minutes for segmentation. FCN-CRF consists of the conditional random field step which significantly increases processing time, while the 3D CNN approach, i.e. V-net, consumes more time than the other methods.

In the second row cluster of the Table I, the evaluation with the Mayo dataset shows that the three network architectures perform well with similar scores. Note that in this dataset, the 3D images were manually cropped to fit the liver volumes, thus part of the false positive segmentation in the V-net does not appear in the segmentation evaluation setup. FCN-CRF achieves the best score of 52.47 mm for the Hausdorff distance metric, while the mean surface distance of DRIU is the smallest at 2.84 mm. Since the volume size of the dataset is much smaller than for the LiTS dataset, the processing time of the three networks is just a few seconds per image. This means the actual time to generate a liver segmentation is very small. A pre-processing step to crop the region of interest will

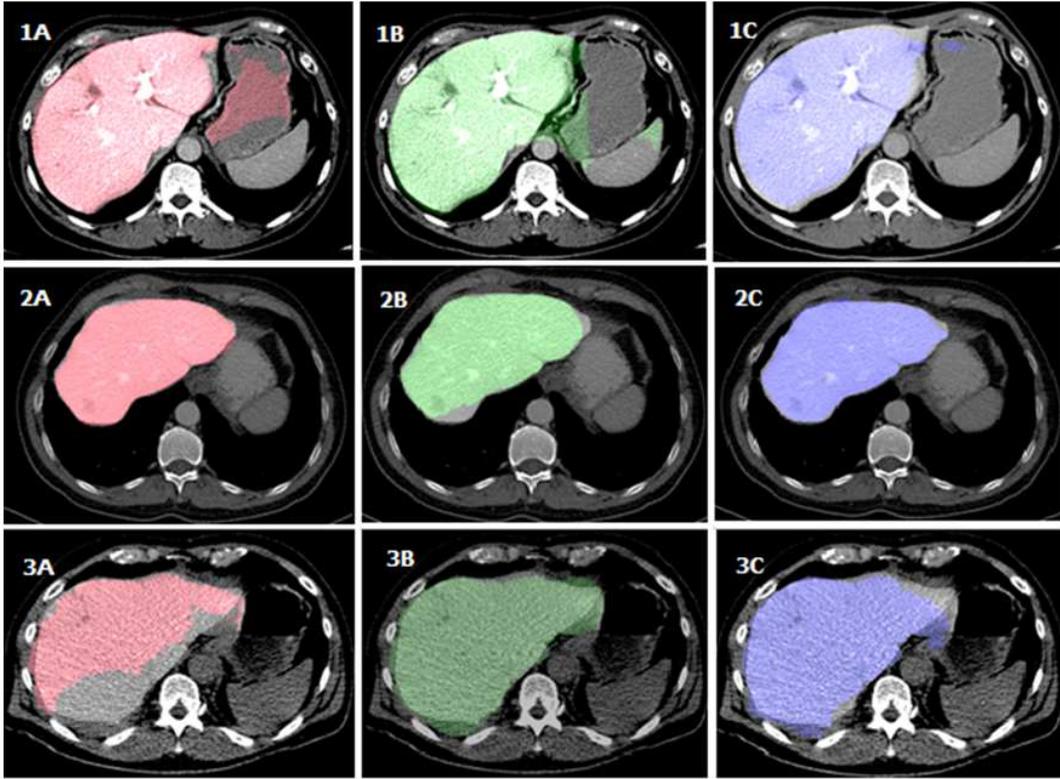


Fig. 3. Examples of liver segmentation by FCN-CRF (in red), DRIU (in green) and V-net (in blue). Each row shows a CT scan acquired from an individual patient. The first row is liver segmentation on the EMC low-dose contrast dataset (EMC_LD), the second row is an image from Mayo dataset with the segmentation overlaid on top. The last row illustrates the liver segmentation of a low-dose, non-contrast enhanced CT image from the EMC_NC_LD dataset.

Dataset	CNNs	Dice (%)	Hausdorff (mm)	Mean surface distance (mm)	FPR (%)	Processing time (s)
LiTS	FCN-CRF	92.4 ± 6.1	207.7 ± 69.5	6.3 ± 23.6	8 ± 10.2	50-77
	DRIU	93.8 ± 1.2	428 ± 1.2	9.6 ± 41.9	4.6 ± 1.5	33-39
	V-net	73.7 ± 15.9	381.7 ± 15.9	59.7 ± 100.6	19.2 ± 12.4	56-83
Mayo	FCN-CRF	91.8 ± 3.4	52.5 ± 62.1	6.9 ± 19.8	5 ± 3	7-7.3
	DRIU	90 ± 2.4	193.8 ± 39.1	2.8 ± 19.8	8.3 ± 2	5.6-5.9
	V-net	91.6 ± 4.0	126.6 ± 71.2	5.5 ± 19.8	9 ± 2.4	6.1-9.2
EMC_LD	FCN-CRF	80.4 ± 14.1	141.8 ± 18.5	11.7 ± 27.7	21.9 ± 16.7	2.8-5.7
	DRIU	83.8 ± 6.2	147.1 ± 26.7	10.4 ± 122.3	14.7 ± 6.7	2-4.5
	V-net	85.2 ± 9.9	118 ± 44	7.4 ± 23	15.9 ± 7.3	3.4
EMC_NC_LD	FCN-CRF	67.4 ± 28.3	97 ± 37.5	11.8 ± 20.8	32.2 ± 29.6	2.6-7
	DRIU	74.4 ± 25.2	131.7 ± 44.3	13.1 ± 19.5	26.1 ± 27.0	1.5 -6
	V-net	81 ± 14.7	105.8 ± 37.0	8.2 ± 17.9	15.1 ± 15.6	2.2-7.7

TABLE I
PERFORMANCE PARAMETERS OF THE THREE CNNs ACROSS ALL OF THE DATASETS

have a large impact on liver segmentation work in clinical applications such as liver interventions which require limited time.

The third and the last row cluster of Table I present the scores of the three networks on the EMC_LD and EMC_NC_LD datasets, respectively. From the results we can conclude that the performance of the three CNNs reduces dramatically due to the impact of low dose noise. That can be explained by the fact that apparently the low-dose images is not in the training set, hence the networks did not work

well for the image type that are not represented in the training set. Furthermore, in general, the mean Dice scores evaluated using the EMC_NC_LD dataset are lower than those in the one using the contrast enhanced dataset. This can be explained by the fact that contrast agent not only enhances the liver vessels but also enhances the liver parenchyma, resulting in a clearer boundary between the liver and other organs (see Figure 3). These results suggest that liver segmentation in low-dose CT image and non-contrast enhanced CT image of the liver is still a challenging task and need further improvements before it can

be applied in clinical use. The first attempt to investigate may be to add these types of data in the training set and retrain the CNN models.

Although this study was carried out on multiple-site datasets and using state of the art methods, there are some limitations in this study. First, the dataset for evaluation only contains 10-15 cases. However, since the datasets were randomly selected, we suppose these are representative. In addition, because the images are three-dimensional and contain several dozen to hundreds of slices per image, we assume this is sufficient for liver segmentation evaluation. Second, there have been some variants of the three CNNs for liver segmentation published recently which have demonstrated even higher Dice scores [15]. Nevertheless, our study aims to investigate how flexible CNNs are with respect to multiple CT image types of the liver, and we suppose that other CNN-based approaches will show similar performance to the three well-known CNNs evaluated in this study. Third, the three CNNs model were either reused from public sources or trained with a setup inherited from the related works [16], leading to limited in the ability in handling the image segmentation. Still, that could be addressed in a larger study with more data and fine tuning hyper parameters, data argumentation involved in the training process [20].

III. CONCLUSIONS

We have successfully implemented and evaluated three CNN architectures for liver segmentation on CT images. The datasets are from several hospitals/medical centers and include contrast enhanced, non-contrast enhanced, and low-dose CT images. The qualitative evaluation result showed that the CNN based segmentation approach for the liver using typical contrast enhanced CT images all achieve good performance while liver segmentation for low-dose and non-contrast enhanced CT images is still a challenging problem. However, with the current development of CNN based methods, we believe that better results for these problems may be realized in the near future, making liver segmentation available for use in clinical practice.

ACKNOWLEDGMENT

This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 102.01-2018.316. 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.

REFERENCES

- [1] K. A. McGlynn, J. L. Petrick, and W. T. London, "Global epidemiology of hepatocellular carcinoma: an emphasis on demographic and regional variability," *Clinics in liver disease*, vol. 19, no. 2, pp. 223–238, 2015.
- [2] M. Mohammadian, N. MahdaviFar, A. Mohammadian-Hafshejani, and H. Salehiniya, "Liver cancer in the world: epidemiology, incidence, mortality and risk factors," *World Cancer Res J*, vol. 5, no. 2, p. e1082, 2018.
- [3] T. T. Hong, N. P. Hoa, S. M. Walker, P. S. Hill, and C. Rao, "Completeness and reliability of mortality data in viet nam: Implications for the national routine health management information system," *PloS one*, vol. 13, no. 1, p. e0190755, 2018.
- [4] <https://www.cancer.org/cancer/liver-cancer/detection-diagnosis-staging/survival-rates.html>. [Online; accessed 28-June-2019].
- [5] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.
- [6] G. Foltz, "Image-Guided Percutaneous Ablation of Hepatic Malignancies," *Seminars in Interventional Radiology*, vol. 31, pp. 180–186, June 2014.
- [7] H. M. Luu, A. Moelker, S. Klein, W. Niessen, and T. van Walsum, "Quantification of nonrigid liver deformation in radiofrequency ablation interventions using image registration," *Physics in Medicine & Biology*, vol. 63, no. 17, p. 175005, 2018.
- [8] P. F. Christ, M. E. A. Elshaer, F. Ettliger, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, M. Armbruster, F. Hofmann, M. DăŃAnastasi, et al., "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, 2016.
- [9] A. Gotra, L. Sivakumaran, G. Chartrand, K.-N. Vu, F. Vandenbroucke-Menu, C. Kauffmann, S. Kadoury, B. Gallix, J. A. de Guise, and A. Tang, "Liver segmentation: indications, techniques and future directions," *Insights into imaging*, vol. 8, no. 4, pp. 377–392, 2017.
- [10] T. Heimann, B. Van Ginneken, M. A. Styner, Y. Arzhaeva, V. Aurich, C. Bauer, A. Beck, C. Becker, R. Beichel, G. Bekes, et al., "Comparison and evaluation of methods for liver segmentation from ct datasets," *IEEE transactions on medical imaging*, vol. 28, no. 8, pp. 1251–1265, 2009.
- [11] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.
- [12] G. Chartrand, P. M. Cheng, E. Vorontsov, M. Drozdal, S. Turcotte, C. J. Pal, S. Kadoury, and A. Tang, "Deep learning: a primer for radiologists," *Radiographics*, vol. 37, no. 7, pp. 2113–2131, 2017.
- [13] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*, pp. 565–571, IEEE, 2016.
- [14] P. F. Christ, F. Ettliger, F. Grün, M. E. A. Elshaera, J. Lipkova, S. Schlecht, F. Ahmaddy, S. Tatavarty, M. Bickel, P. Bilic, et al., "Automatic liver and tumor segmentation of ct and mri volumes using cascaded fully convolutional neural networks," *arXiv preprint arXiv:1702.05970*, 2017.
- [15] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, "H-denseunet: hybrid densely connected unet for liver and tumor segmentation from ct volumes," *IEEE transactions on medical imaging*, vol. 37, no. 12, pp. 2663–2674, 2018.
- [16] M. Bellver, K.-K. Maninis, J. Pont-Tuset, X. Giró-i Nieto, J. Torres, and L. Van Gool, "Detection-aided liver lesion segmentation using deep learning," *arXiv preprint arXiv:1711.11069*, 2017.
- [17] A. A. Novikov, D. Major, M. Wimmer, D. Lenis, and K. Bühler, "Deep sequential segmentation of organs in volumetric medical scans," *IEEE transactions on medical imaging*, 2018.
- [18] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [19] H. Meine, G. Chlebus, M. Ghafoorian, I. Endo, and A. Schenk, "Comparison of u-net-based convolutional neural networks for liver segmentation in ct," *arXiv preprint arXiv:1810.04017*, 2018.
- [20] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.
- [21] MICCAI grand challenge. <https://grand-challenge.org/challenges/>. [Online; accessed 28-June-2019].
- [22] H. M. Luu, C. Klink, W. Niessen, A. Moelker, and T. van Walsum, "Non-rigid registration of liver ct images for ct-guided ablation of liver tumors," *PloS one*, vol. 11, no. 9, p. e0161600, 2016.