

HYBRID CASCADED NEURAL NETWORK FOR LIVER LESION SEGMENTATION

Raunak Dey, Yi Hong

Department of Computer Science, University of Georgia

ABSTRACT

Automatic liver lesion segmentation is a challenging task while having a significant impact on assisting medical professionals in the designing of effective treatment and planning proper care. In this paper, we propose a cascaded system that combines both 2D and 3D convolutional neural networks to segment hepatic lesions effectively. Our 2D network operates on a slice-by-slice basis in the axial orientation to segment liver and large liver lesions; while we use a 3D network to detect small lesions that are often missed in a 2D segmentation design. We employ this algorithm on the LiTS challenge obtaining a Dice score per subject of 68.1%, which performs the best among all non-pre-trained models and the second-best among published methods. We also perform two-fold cross-validation to reveal the over- and under-segmentation issues in the annotations of the LiTS dataset.

Index Terms— Liver lesion segmentation, Hybrid neural network, Small lesion segmentation.

1. INTRODUCTION

Liver lesions are groups of abnormal cells in the liver, and some of them lead to cancer. Liver cancer is one of the leading causes of cancer deaths worldwide, and more than 700,000 deaths are reported each year, according to the American Cancer Society. For the liver cancer screening, the Computer Tomography (CT) is the most commonly used imaging tool, and the technique of the automatic liver lesion segmentation from a CT scan has great impacts on cancer diagnosis, surgery planning, and treatment evaluations.

Due to the heterogeneous and diffusive appearance of hepatic lesions, liver lesion segmentation is a challenging task. Researchers have proposed many segmentation algorithms based on the classical segmentation techniques, e.g., thresholding [1], region growing [2], active contour [3], and ensemble [4]. Recently, deep neural networks have been widely used in the liver lesion segmentation and have shown improved performance in detecting and segmenting liver lesions. These algorithms either use 2D convolutional neural networks [5, 6, 7], 3D networks [8], or a combination of both [9, 10]. One of the issues that most existing methods

are facing is the segmentation of the small lesions, which are often missed in their final predictions.

In this paper, we introduce a hybrid cascaded segmentation network to segment the liver lesions, especially the small ones. The network utilizes both 2D and 3D convolutional networks to effectively find both large and small lesions in the liver region. We use a 2D convolutional neural network (CNN) to obtain the liver mask from the input CT volume of the abdomen region, which locates the liver region for further processing. Another 2D CNN is then employed to extract large lesions from 2D slices, while a dedicated 3D CNN is proposed to segment out small lesions from 3D volumes.

We evaluate our method on the Liver Tumor Segmentation (LiTS) dataset [11] and obtain a Dice score per subject of 68.1% without any pre-training or post-processing. This result ranks first among the currently published non-pre-trained networks. Moreover, with the proposed special treatment of the small lesion segmentation using a 3D CNN, we improve the network performance of segmenting the large lesions only by 7.1% in terms of the Dice score per subject. In addition, using a two-fold cross-validation on the LiTS training set, we observe that our network can segment unannotated regions which share the similar intensity as that of the lesion region annotated on the same slice in the LiTS dataset.

2. METHOD

2.1. Overview

In this section, we describe our hybrid cascaded network to handle the liver lesion segmentation problem. We first segment the liver from the original CT scan slice-by-slice using a 2D neural network. The generated liver mask allows us to focus on the lesion segmentation located inside of the liver. At the stage of lesion segmentation, we treat large and small lesions separately with a 2D network to segment large lesions and a 3D network to segment small lesions. This hybrid design was motivated by the goal of balancing computational efficiency without sacrificing accuracy. Compared to the 3D segmentation network, a 2D network is more efficient in terms of both computation time and memory cost; however, on a 2D slice, small lesions are likely to be confused with some normal tissues in the liver region. Therefore, we propose to use an adaptive 3D network to handle the small lesion segmen-

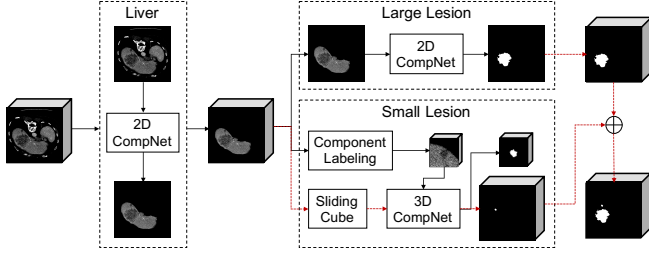


Fig. 1. Overview of our proposed hybrid cascaded network for the liver lesion segmentation. We adopt 2D CompNets (see Fig. 2 for the detailed network architecture) to segment the liver region and large lesions and apply a 3D CompNet (Fig. 4) to segment the small lesions. The red dashed lines indicate the test phase of the liver lesion segmentation.

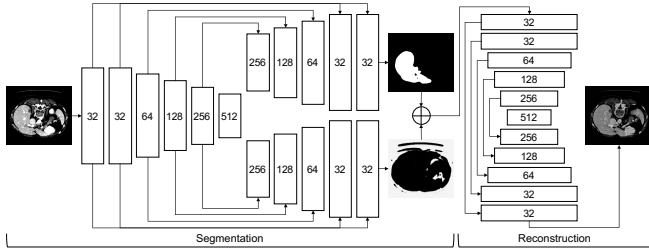


Fig. 2. Network architecture of the 2D CompNet for the liver/large lesion segmentation. This U-Net-like network has the object segmentation (top left), its complementary segmentation (bottom left), and the image reconstruction (right). Each block has two convolutional layers with the numbers of feature maps shown in the block box.

tation, which greatly reduces such false positives. Figure 1 depicts the overview of the proposed hybrid cascaded network. Its backbone network is the CompNet [12], which was demonstrated to be more robust in addressing the segmentation problem on pathological images than the U-Net [13].

2.2. Liver Segmentation

A liver CT scan typically has hundreds of slices, and each slice has an image resolution of 512×512 pixels. Directly working on such 3D images would cause time and memory issues using a segmentation neural network under an encoder-and-decoder design. Fortunately, the segmentation of the liver region inside the abdominal cavity is relatively easier than the segmentation of its lesions. As demonstrated by recent works [5, 6, 14, 15] and our experiments, liver segmentation relied only on 2D slices can achieve more than 95% overall dice accuracy. Therefore, we use a 2D network to segment out the liver from a 3D CT scan.

We follow the design of the basic CompNet proposed in [12] and adapt it to our liver segmentation task, resulting in the 2D network architecture shown in Fig. 2. This net-



Fig. 3. False positives of small liver lesions on a 2D slice.

work takes 2D slices as input and outputs 2D liver masks that are stacked into a 3D mask. The encoder branches of the liver segmentation, its complementary segmentation, and the reconstruction sections consist of multiple blocks of two convolutional layers followed by the batch normalization layers. All convolutional layers use a filter size of 3×3 and the filter numbers in the blocks are as follows 32, 32, 64, 128, 256, and the transition block of 512. Following each block, we have a pooling layer. The decoder branches of the network mirror the encoder ones with the pooling layers being replaced by 2D transposed convolutional layers. We multiply the input 3D scan with the obtained liver mask to retain only the liver region for the following lesion segmentation.

2.3. Two-Step Lesion Segmentation

After having the liver segmented, we might directly apply a 3D segmentation network to extract the liver lesions. However, a 3D Network operating only in the liver region is still computationally expensive and may suffer from a lack of 3D training scans. To mitigate these issues, we propose a hybrid strategy to handle the segmentation of large and small lesions separately. In particular, we use a 2D network to scan slice-by-slice and predict large lesions if present. However, this method is not sufficient for predicting small lesions, as the false positive examples shown in Fig. 3. This happens due to the appearance of the small lesions matching that of other tissues or vessels on 2D liver slices. A 3D network learning from cropped volumes with small lesions can reduce such false positives because it leverages the observation that the location of the tissue or vessel appears to travel considerably in-between slices, whereas the liver lesion appears to be roughly stationary across slices. Regarding the size threshold used to separate the large and small lesions, it may vary for different applications and datasets. Here, we set the threshold as a pixel dimension of $32 \times 32 \times 32$ by experimentally testing on the LiTS dataset, which balances the prediction accuracy and the computation time.

Large Lesion Segmentation. We use another 2D CompNet like the one used for the liver segmentation (Fig. 2) to ex-

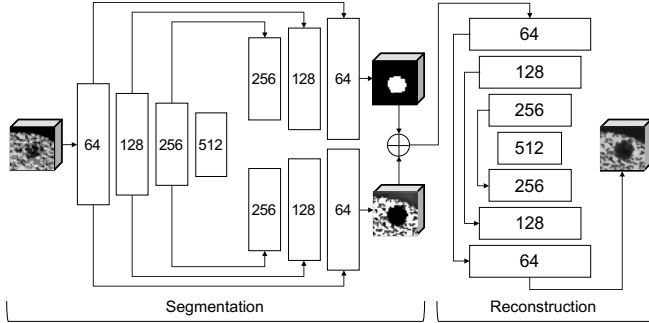


Fig. 4. Network architecture of the 3D CompNet used for the small lesion segmentation. This U-Net-like network has the small lesion segmentation (top left), its complementary segmentation (bottom left), and the image reconstruction (right). Each block has two convolutional layers with the numbers of feature maps shown in the block box.

tract lesions larger than 32×32 pixels from the input slices with the liver only. Since this network takes care of the prediction of large lesions, we clean the training masks by removing all small lesion masks, which is implemented using the connected component labeling algorithm provided by the OpenCV library [16]. The horizontal and vertical dimensions of the removed small lesions are both less than or equal to 32 on a 2D slice in the axial orientation. In the test phase, we also remove the detected lesions less than 32×32 using component labeling.

Small Lesion Segmentation. We use a 3D CompNet, as shown in Fig. 4, to segment the small lesions. This 3D network takes a volume of $32 \times 32 \times 32$ as input. The encoder branches of the lesion segmentation, its complementary segmentation, and the reconstruction consist of three 3D convolutional blocks. Each block is composed of two 64, 128, and 256 feature maps, respectively, each followed by a 3D pooling layer with a pool size (2, 2, 2). The transition block has 512 feature maps. The decoder branches mirror the encoder branches with the pooling layers replaced by the 3D transposed convolutional layers.

To generate the training samples for this 3D network, we use the component labeling to locate the center of a lesion and estimate its dimensions on a 2D slice. We sample a 3D volume for a lesion whose horizontal and vertical dimensions on the 2D slice are smaller than or equal to 32. Since the small lesions often occupy a few slices, we choose a cube with a size of $32 \times 32 \times 32$ to cover the lesion. In particular, around the lesion center on a slice, we select 15 slices above and 16 slices below to create a 3D volume for training. During the testing phase, we use a 3D sliding cube over the liver volume to predict the small lesions using the above trained 3D network. To choose an appropriate stride size for the sliding cube, we test on 32, 16, 8, and 4 voxels, respectively. It turns out that the stride 32 and 16 are too large to capture tiny lesions. The

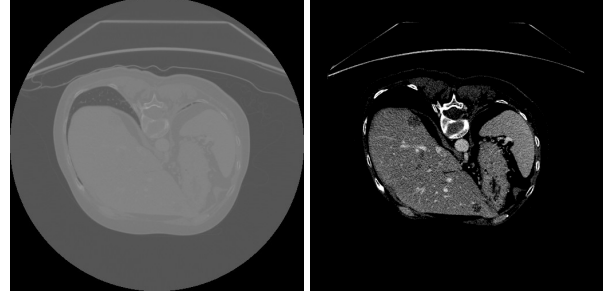


Fig. 5. An example of a liver CT scan before (left) and after (right) preprocessing.

stride 8 is an appropriate one since a further reduction of the stride size does not reveal any new changes but increasing the prediction time. To fuse the overlap predictions at the same voxel, we take their average and set the value greater than 0.5 to 1 as the final prediction.

3. EXPERIMENTS

Dataset and Preprocessing. To test our proposed network, we use the public Liver Tumor Segmentation (LiTS) [11] dataset, which consists of 130 abdomen CT scans for training and 70 for testing. To train the 2D liver segmentation network, we fully use all training scans with a total of 58,638 2D slices. In the network training of the large liver lesion segmentation, we focus on the slices with the liver present, resulting in 19,163 2D slices in total. The 3D network is trained on 11,503 3D small lesion samples with size of $32 \times 32 \times 32$. We preprocess the liver CT scans using a histogram-based thresholding method. We select the rightmost peak of the intensity histogram distribution of a CT scan for normalization and use the histogram equalization algorithm to generate the enhanced images, as shown in Fig. 5. The implementation of the preprocessing and the liver lesion segmentation network is available online ¹.

Experimental Settings. We use Keras with the TensorFlow backend to implement our proposed network. The 2D CompNet for the liver segmentation is trained for 40 epochs using the Adam optimizer with a learning rate of $5e-5$. Next, we train the 2D and 3D CompNets for the large and small lesion segmentation in the same manner, i.e., we first train the networks using the Adam optimizer with a learning rate of $5e-5$, the same as that used in [5], and having an early stopping scheme with the tolerance being set to 5. We then train the networks with a learning rate of $1e-6$ using an early stopping with a tolerance of 10 trials. Both steps have 150 maximum number of epochs for training. Also, we use an L2 regularization with a parameter of $2e-4$ and a dropout with a rate of 0.3 after all pooling and upsampling layers to mitigate overfitting.

¹https://github.com/raun1/LITS_Hybrid_Comp_Net

Method	Dice Per Subject	Dice Overall	W/o Pre-training	W/o Post-processing
H-Dense U-Net [14]	0.722	0.824		✓
Multiple U-Nets [5]	0.680	0.796	✓	
2.5 D U-Net [6]	0.670	—	✓	
CDNN [15]	0.657	0.820	✓	✓
FED-Net [17]	0.650	0.766	✓	
AH-Net [10]	0.634	0.834	✓	✓
RA U-Net [18]	0.595	0.795	✓	✓
BS-U-Net [7]	0.552	0.729	✓	✓
Ours (w/o small lesions)	0.681 (0.610)	0.813 (0.776)	✓	✓

Table 1. Comparison among published approaches and ours on the LiTS challenge.

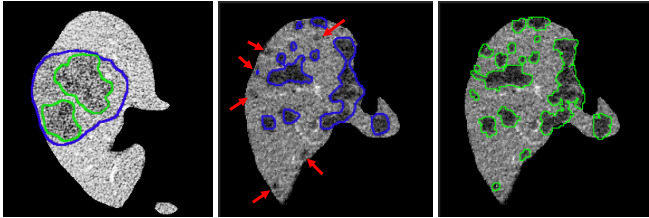


Fig. 6. Comparison between our predictions (green) and the LiTS annotations (blue) using the two-fold cross-validation on the LiTS training set. The red arrows indicate the artifacts that share similar intensity distribution to lesions annotated on the same slice but are missing in the annotations while being predicted by our method.

Experimental Results. Since the LiTS dataset does not include the segmentation ground truth for the test set, we first perform the two-fold cross-validation on the training set to quantitatively and qualitatively evaluate the performance of our proposed method. We obtain 67.3% Dice per subject and some visual samples are presented in Fig. 6. By evaluating the test set of the LiTS challenge, we list in Table 1 our result, which is generated by our model submitted to the challenge, with comparison to those of the currently published approaches. According to the Dice score per subject, the most important metric for measuring an algorithm’s performance on the LiTS challenge, our approach is at the second rank, following after the H-Dense U-Net [14], which however uses pre-training. In addition, our small lesion segmentation could be an add-on component to their method for further improvement in segmenting small lesions. Our experiment shows that the dice per subject can be improved from 0.61 to 0.681 with the consideration of the small lesion segmentation. In particular, among the methods without pre-training and post-processing, our method has the best performance in terms of the dice-per-subject score. Figure 7 reports some visual results of our predictions on the LiTS test set.

Observations on LiTS Annotations. As shown in Fig. 6, the LiTS annotations have both over-segmentation and under-segmentation issues. For the large lesions in both cases

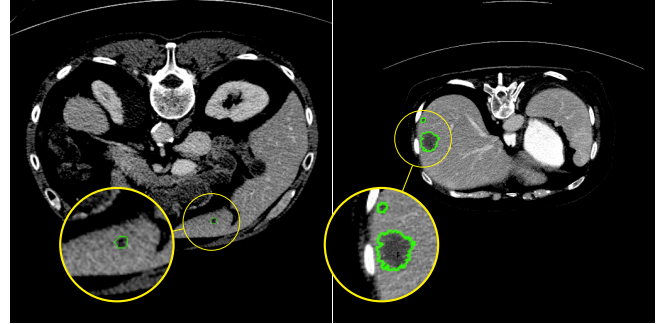


Fig. 7. Examples of our predictions (indicated by the green lines) on the LiTS test set.

shown in Fig. 6, our predictions better fit lesions compared to the ground truth, while for the small lesions, our predictions locate more lesions potentially missing in the ground truth. Similar observations have been reported in [5]. Due to the imperfect ground truth provided by the LiTS challenge, we argue that the metrics computed against the ground truth probably could not be the only way to compare the segmentation results. Visual results could be considered as well, and ours visually present reasonable liver lesion segmentation predicted by our method.

4. DISCUSSION AND CONCLUSION

In this paper, we proposed a hybrid 2D and 3D neural network for segmenting the liver lesions. Especially, we designed a dedicated 3D segmentation network for the small lesions in the liver. This 3D segmentation network could be an add on to a network that suffers from segmenting small objects. Moreover, we observed the imperfect annotations provided by the LiTS data set, which hinders the further improvement of a network’s learning performance and makes the evaluation results questionable for comparison. How to handle and fully leverage such imperfect labels will be our future work. We also plan to extend our framework to other medical applications with the lesion/tumor or small object segmentation.

5. REFERENCES

- [1] Jan Hendrik Moltz, Lars Bornemann, Volker Dicken, and H Peitgen, "Segmentation of liver metastases in ct scans by adaptive thresholding and morphological processing," in MICCAI workshop, 2008, vol. 41, p. 195.
- [2] Jan Hendrik Moltz, Lars Bornemann, Jan-Martin Kuhnigk, Volker Dicken, et al., "Advanced segmentation techniques for lung nodules, liver metastases, and enlarged lymph nodes in ct scans," IEEE Journal of selected topics in signal processing, vol. 3, no. 1, pp. 122–134, 2009.
- [3] Laurent Massotier and Sergio Casciaro, "A new fully automatic and robust algorithm for fast segmentation of liver tissue and tumors from ct scans," European radiology, vol. 18, no. 8, pp. 1658, 2008.
- [4] Weimin Huang, Yongzhong Yang, Zhiping Lin, Guang-Bin Huang, Jiayin Zhou, Yuping Duan, and Wei Xiong, "Random feature subspace ensemble based extreme learning machine for liver tumor detection and segmentation," in EMBC. IEEE, 2014, pp. 4675–4678.
- [5] Grzegorz Chlebus, Andrea Schenk, Jan Hendrik Moltz, Bram van Ginneken, Horst Karl Hahn, and Hans Meine, "Automatic liver tumor segmentation in ct with fully convolutional neural networks and object-based post-processing," Scientific reports, vol. 8, no. 1, pp. 15497, 2018.
- [6] Xiao Han, "Automatic liver lesion segmentation using a deep convolutional neural network method," arXiv:1704.07239, 2017.
- [7] Song Li and Geoffrey Kwok Fai Tso, "Bottleneck supervised u-net for pixel-wise liver and tumor segmentation," arXiv:1810.10331, 2018.
- [8] Huiyan Jiang, Tianyu Shi, Zhiqi Bai, and Liangliang Huang, "Ahcnet: An application of attention mechanism and hybrid connection for liver tumor segmentation in ct volumes," IEEE Access, 2019.
- [9] Patrick Ferdinand Christ, Florian Ettlinger, Felix Grün, Mohamed Ezzeldin A Elshaera, et al., "Automatic liver and tumor segmentation of ct and mri volumes using cascaded fully convolutional neural networks," arXiv:1702.05970, 2017.
- [10] Siqi Liu, Daguang Xu, S Kevin Zhou, Olivier Pauly, Sasa Grbic, Thomas Mertelmeier, Julia Wicklein, Anna Jerebko, Weidong Cai, and Dorin Comaniciu, "3d anisotropic hybrid network: Transferring convolutional features from 2d images to 3d anisotropic volumes," in MICCAI. Springer, 2018, pp. 851–858.
- [11] Patrick Bilic, Patrick Ferdinand Christ, Eugene Vorontsov, Grzegorz Chlebus, et al., "The liver tumor segmentation benchmark (lits)," arXiv:1901.04056, 2019.
- [12] Raunak Dey and Yi Hong, "Compnet: Complementary segmentation network for brain mri extraction," in MICCAI. Springer, 2018, pp. 628–636.
- [13] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in MICCAI. Springer, 2015, pp. 234–241.
- [14] Xiaomeng Li, Hao Chen, Xiaojuan Qi, Qi Dou, Chi-Wing Fu, and Pheng-Ann 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.
- [15] Yading Yuan, "Hierarchical convolutional-deconvolutional neural networks for automatic liver and tumor segmentation," arXiv:1710.04540, 2017.
- [16] G. Bradski, "The OpenCV Library," Dr. Dobb's Journal of Software Tools, 2000.
- [17] Xueying Chen, Rong Zhang, and Pingkun Yan, "Feature fusion encoder decoder network for automatic liver lesion segmentation," arXiv:1903.11834, 2019.
- [18] Qiangguo Jin, Zhaopeng Meng, Changming Sun, Leyi Wei, and Ran Su, "Ra-unet: A hybrid deep attention-aware network to extract liver and tumor in ct scans," arXiv:1811.01328, 2018.