Extraction of Liver Vessel Systems From CT-Image (Trích xuất hệ thống mạch máu gan từ ảnh chụp CT)

Nguyen Le Nhat Duong[1,2], Pham Khanh Trinh[1,2], Le Thanh Sach[2,3*]
[1]Faculty of Computer Science and Engineering, Ho Chi Minh City University of Technology (HCMUT), 268 Ly Thuong Kiet Street, District 10, Ho Chi Minh City, Vietnam
[2]Office for International Study Programs, Ho Chi Minh City University of Technology (HCMUT), 268 Ly Thuong Kiet Street, District 10, Ho Chi Minh City, Vietnam
[3]Vietnam National University Ho Chi Minh City, Linh Trung Ward, Thu Duc District, Ho Chi Minh City, Vietnam
*Corresponding author: ltsach@hcmut.edu.vn

Abstract

Extraction of the blood vessel system of a liver is a challenging task in the field of medical image processing. Normally, doctors must examine each slice manually to achieve accurate vessel segmentation. However, manual vessel segmentation is tedious and relies on the experience and skill of the practitioners. Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) scans contain a lot of information about human anatomy. As a result, automatic liver vessel segmentation has drawn research attention in recent years. For these reasons, in this work, we propose our customized MultiResUNet, called MultiResDenseUNet, as an approach to solve this challenging task. To evaluate the results of our proposed method, we conducted experiments with 20 data in different contrast levels from the public IRCAD dataset. The results demonstrate impressive performance in comparison with the state-of-the-art methods.
Trích xuất hệ thống mạch máu của gan là một nhiệm vụ đầy thách thức trong lĩnh vực xử lý hình ảnh y tế. Thông thường, các bác sĩ phải kiểm tra từng lát cắt bằng tay để phân đoạn mạch chính xác. Tuy nhiên, việc phân loại bằng thủ công cần sự tỉ mỉ rất cao và còn phụ thuộc vào kinh nghiệm và kỹ năng của người thực hiện. Ảnh từ chụp cộng hưởng từ (MRI) và chụp cắt lớp vi tính (CT) chứa rất nhiều thông tin về giải phẫu người. Chính vì thế, việc áp dụng trích xuất mạch máu của gan một cách tự động đã thu hút sự rất nhiều công trình nghiên cứu trong những năm gần đây. Vì những lý do trên, chúng tôi đề xuất sử dụng một biến thể của mô hình MultiResUNet, được chúng tôi đặt tên là MultiResDenseUNet, như một cách tiếp cận để giải quyết bài toán đầy thách thức này. Để đánh giá kết quả của phương pháp trên, chúng tôi đã tiến hành thử nghiệm với 20 dữ liệu có mức độ tương phản khác nhau từ tập dữ liệu công khai IRCAD. Kết quả cho thấy phương pháp này có hiệu suất ấn tượng so với các phương pháp hiện đại.
Keywords: *Blood vessel, vessel segmentation, Magnetic Resonance Imaging, Computed Tomography, MultiResUNet, MultiResDenseUNet, IRCAD.*

1. Introduction

Liver is one of the most vital organs in our body; however, due to the complexity of the liver vessel system, hepatic diseases are often quite hard to diagnose. In order to help experts with diagnosis and treatment planning related to these diseases, accurate segmentation methods of the vessels are in high demand in clinical practice. In recent days, liver vessel segmentation methods could be classified into: traditional methods and deep learning-based methods.

Traditional methods

Traditional edge-based methods can be further classified into image filtering, enhancement algorithm and tracking-based algorithm [1, 2]. Image filtering and enhancement method is based on the tree structure of the liver vessels as a whole. It utilizes the relationship between Hessian matrix eigenvalues to extract the tubular structure from the image [3, 4]. Tracking-based algorithms use pre-defined vessel models and track the minimum cost path. Friman et al. [5] suggested tracking hypothetical vessel trajectories at the same time to improve the results in low contrast conditions. Cetin and Unal [6], Cetin et al. [7] proposed the tubular structure segmentation method, a second-order tensor from directional intensity measurement and higher-order tensor based on cylindrical flux-based to construct the vascular structure.

Deep learning-based methods

The evolution of computer science in research on image segmentation has revealed interest in the reconstruction and interpretation of complex organic structures. The accuracy of the segmentation has been enhanced recently because of deep learning. U-Net [8], which originated from the FCN [9], is considered to be the most widely used deep learning model for image segmentation. The

majority of deep learning models rely on the robustness of CNN-based architectures, specifically, U-Net [10], V-Net [11], FCN and its variants. In chronological order, early segmentation on the retina is based on 2D methods. However, 3D methods recently became the mainstream. Huang et al. [13] proposed a variant of 3D-UNet which fits the problem well, but this model has to be trained on a large dataset with high-quality labels and the noise labels of current datasets affect the model's performance. Yang et al. [14] proposed a method of liver vessel segmentation based on an improved V-Net network which utilizes the dilated convolution, 3D deep supervision mechanism and inter-scale dense connections in the decoder of the network. This method prevents the loss of high-level semantic information during the decoding process and effectively integrates multi-scale feature information. However, its performance score is not high because of complex structures and the contrast of surrounding tissues from CT images. Wu et al. [15] proposed an end-to-end segmentation network called Inductive BIased Multi-Head Attention Vessel Net (IBIMHAV-Net) by utilizing swin transformer, proposed by Liu et al. [16], and the combination of convolution and self-attention. This method depends heavily on the number of transformer blocks and the corresponding upsampling and downsampling kernel blocks; otherwise, it cannot lead to convergence. Consequently, this method may not work well for some datasets. Considering the related works, we propose a convolution-based architecture for liver vessel segmentation which better captures global features of biomedical image segmentation and reduces the loss of spatial information during pooling operations.

2. Our proposed method

In U-Net architecture [10], after each pooling layer and transposed convolutional layers, sequences of two 3 × 3 convolutional layers are used, as shown in Figure 1. As explained in Szegedy et al. [17], two 3 × 3 convolutional operations in a row resembles a 5 × 5 convolutional operation.

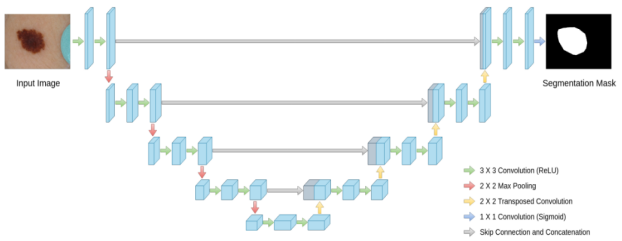


Figure 1: The U-Net architecture comprises an encoder and decoder pathway with concatenation between the corresponding layers from [18].

Based on the approach of Inception network, to augment U-Net with multi-resolutional analysis is to incorporate 3×3, and 7×7 convolution operations in parallel to the 5×5 convolution operation, as shown in Figure 2a. However, the implementation of additional layers in parallel increases the size of the memory significantly. To solve this problem, Ibhetaz et al. [18] borrow an idea from [17] which is replacing the bigger 5×5 and 7×7 convolutional layers by using a sequence of smaller 3×3 convolutional blocks, as shown in Figure 2b. [18] also add a residual connection, as shown in Figure 2c, because of their efficacy in biomedical image segmentation, and for additional spatial information from the 1×1 convolutional layers. This is called a *MultiRes block* [18].

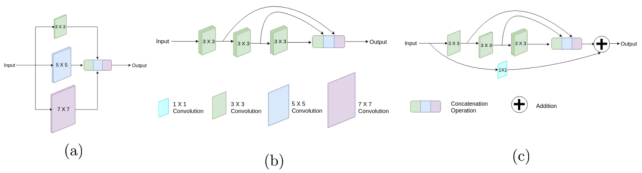


Figure 2: The MultiRes block from [18].

With respect to the MultiRes block [18], we have added some skip connections, as red lines in Figure 3, between mutual convolutional blocks and input, in order to generalize the data much better. The input to the next layer is the concatenation of the previous layers. This idea is quite similar to the DenseNet [23] architecture.

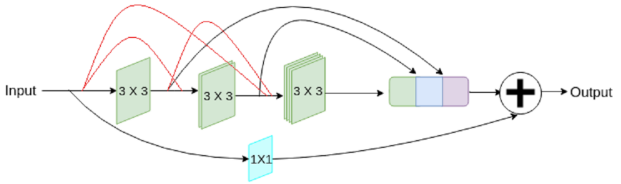


Figure 3: Our proposed MultiRes++ block.

[18] first pass the feature maps through a sequence of convolutional layers with residual connections, and then concatenate them with the decoder features. As in [18], it is called *Res path*, illustrated in Figure 3. As a usual, 3×3 and 1×1 filters are used in the convolutional layers to accompany the residual connections.

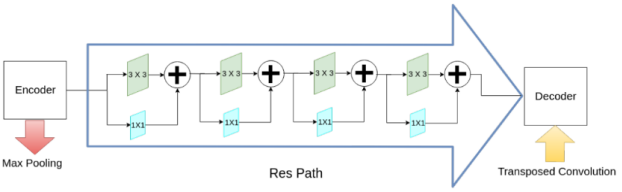


Figure 4: The Res path from [18].

In general, all of these components are integrated into our proposed MultiResDenseUNet, as shown in Figure 5.

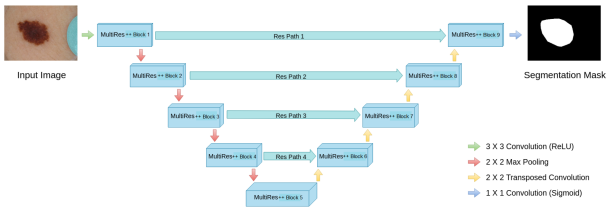


Figure 5: Our proposed MultiResDenseUNet.

Based on the MultiResDenseUNet architecture, we added some state-of-the-art techniques to improve the performance of the model. All the convolutional layers except for the output layer, used in this network are activated by the ReLU activation function and batch-normalized. This is a segmentation procedure; therefore, we chose not to use log loss cross-entropy. In cross entropy loss, the loss is calculated as the average of per-pixel loss, and the per-pixel loss is calculated independently without knowing whether its adjacent pixels are boundaries or not. Therefore, cross entropy only considers loss in a micro sense rather than considering it globally, and this is not enough for segmentation level. Dice Coefficient is a widely used metric to evaluate the segmentation output as it is a region-based coefficient. It has also been used as a loss function as it fulfills the mathematical representation of segmentation objectives. However, due to its non-convex property, it may fail in convergence, thus not leading to the optimal results. Inspired by Milletari et al. [11] and Moshagen et al. [12], we decided to use log-cosh dice loss as the loss function of the model. It is represented as a combination of log-cosh and dice loss:

$$Loss = log(cosh(DiceLoss)) \quad (1)$$

where $DiceLoss$ is defined as followed:

$$DiceLoss(y, \hat{p}) = 1 - \frac{2y\hat{p} + \epsilon}{y + \hat{p} + \epsilon} \quad (2)$$

with ϵ as a *smooth* value, usually 1, added in numerator and denominator to ensure the function is not undefined in edge case scenarios such as when $y = \hat{p} = 0$.

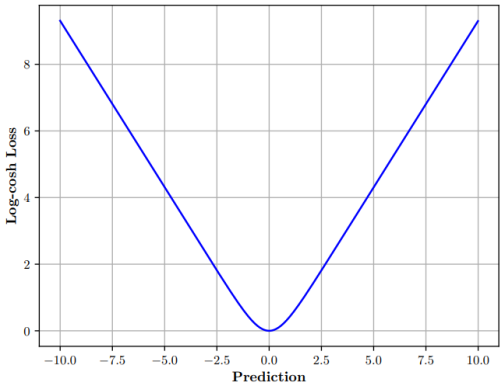


Figure 6: Plot of log-cosh loss with the range of predicted values: [-10,10] from [12].

$log(cosh(x))$ is nearly equal to $\frac{x^2}{2}$ for small x and $|x| - log(2)$ for large x. The log-cosh loss function acts as Mean Absolute Error for large values and Mean Squared Error for small values as proven by Moshagen et al.'s work. It can learn a big cluster and give a stable solution. Another reason for us to choose log-cosh dice loss is because hyperbolic functions are used by deep learning algorithms for their non-linearities. These functions are easily differentiated. The range of $cosh(x)$ can go up to infinity, so log space is used to capture its range. As a result, we have:

$$L(x) = log(cosh(x)) \quad (3)$$

and by using the chain rule, we have:

$$L'(x) = \frac{sinh(x)}{cosh(x)} = tanh(x) \quad (4)$$

which is continuous and finite as $tanh(x)$ range is $[-1, 1]$. Based on the proof that log-cosh remains continuous and finite after first order differentiation, we choose and propose the log-cosh dice loss function as the loss function of the model.

3. Experiments

Evaluation metrics

The evaluation metrics are Dice Coefficient and Accuracy:

$$Dice = \frac{2TP}{2TP + FP + FN} \quad (5)$$

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \quad (6)$$

where TP and TN are the numbers of correct pixels categorized by liver vessels and background, FP and FN are the numbers of incorrect pixels categorized by the liver vessels and background.

IRCAD dataset

In comparison with other hepatic dataset, IRCADB dataset offers a high variety and complexity of livers and its lessons are publicly available. The IRCAD dataset, which is provided by the Research Institute against Digestive Cancer, includes 20 venous phase enhanced CT volumes from various European hospitals with different CT scanners. 20 three-dimensional images of enhanced portal venous phase with the number of slices ranging from 74 to 260, and a single-layer resolution of 512×512. Manually, we chose 12 cases for training and 8 cases for testing. The size of the input image after preprocessing is 512×512×1 because we extracted each slice from the 3D volume for training and inference purposes.

Experimental setup

The hardware configuration for the experiment is an NVIDIA Tesla P100 – PCIE GPU (16 GB memory) with development tools are Python 3.7 and Keras. The whole procedure is finally conducted in 100 epochs. Adam optimizer was selected for the training process. The batch size was 4. In addition, we used the Cyclical learning rate method, proposed by Leslie N. Smith [19]. Our Cyclical options are: $1e-4$ for minimum bound (base learning rate) and $1e-3$ for maximum bound (maximum learning rate). One of the important points of the setup is the He initialization, which is invented by He et al. [20]. He initialization is a method for weight initialization that works well for the non-linear activation functions, especially ReLU. A proper initialization method should avoid reducing or magnifying the magnitudes of input signals exponentially.

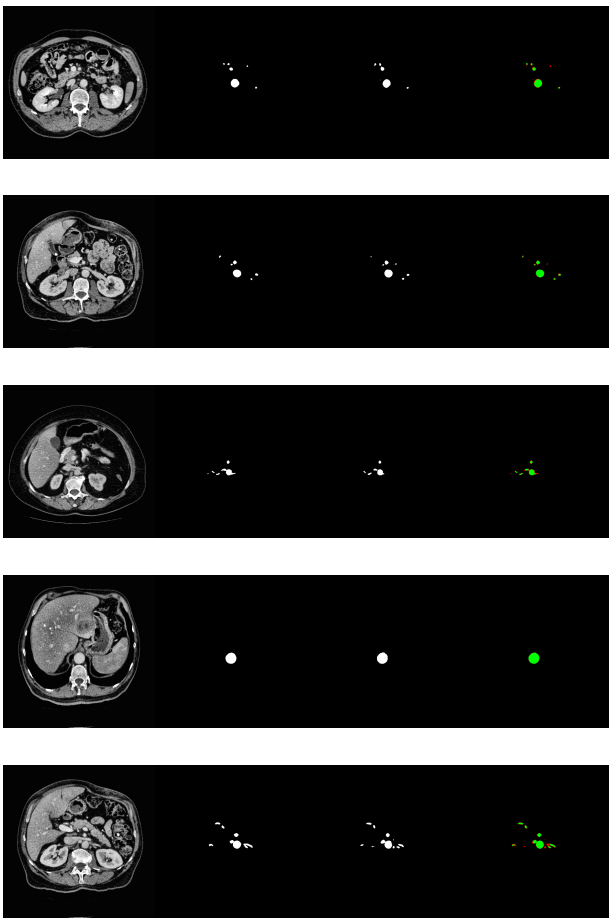
4. Results

Our method was tested on 4 people from IRCAD datasets as mentioned. As shown in Table 1, after introducing the MultiResDenseUNet with an appropriate loss function, weight initialization, and hyperparameters, the average dice and accuracy are 79.4 and 0.998 respectively.

Table 1: The Dice Coefficient of our model and other related models. The Dice Coefficient here is rounded to percentage.

Method	Dice (%)
Frangi et al. [21]	66.4
Huang et al. [13]	67.5
V-Net [11]	69.1
ResUNet [22]	70.6
Improved inter-scale V-Net [14]	71.6
U-Net [10]	72.3
IBIMHAV-Net [15]	74.8
MultiResUNet [18]	**77.4**
Ours - MultiResDenseUNet	**79.4**

Our mechanism can alleviate the vanishing gradient or explosion of the network during the training process because of the residual connection. Concatenation between encoder and decoder can help the network better learn different features without losing information.



a b c d

Figure 7: Some visualizations and comparison of the groundtruth labels and the predictions of our proposed method. Row indicates different data samples. Three columns: (a) The input image; (b) the groundtruth label; (c) the prediction; and (d) the intersection: Red is the Difference and Green is the Similarity between (b) and (c).

With reference to Figure 4, our proposed deep neural network works well with a wide range of CT input images. However, in some sophisticated samples with continued vessels, the last sample in Figure 4 for example, the network could not predict the whole vessel but the correct shape with discontinued vessel instead.

5. Conclusions

We have suggested an architecture to automatically segment vessels in a CT image. The proposed method's features are learned from 2D slices, which are first preprocessed by normalization, clipping and CLAHE. These preprocessing methods can deal with any intensity variation in the CT data. The results achieved with the IRCAD dataset confirm that our proposed model better generalizes and improves the accuracy with respect to

current state-of-the-art methods. In the future, we would try to increase the segmentation accuracy by introducing more powerful and precise datasets, and updating the model architecture in parallel, in order to solve sophisticated and challenging cases.

6. References

1. Jonas Lamy, Odyssée Merveille, Bertrand Kerautret, Nicolas Passat, and Antoine Vacavant. *Vesselness filters: A survey with benchmarks applied to liver imaging*. In 2020 25th International Conference on Pattern Recognition (ICPR), pages 3528–3535. IEEE, 2021.

2. Ha Manh Luu, Camiel Klink, Adriaan Moelker, Wiro Niessen, and Theo Van Walsum. *Quantitative evaluation of noise reduction and vesselness filters for liver vessel segmentation on abdominal CT images*. Physics in Medicine & Biology, 60(10):3905, 2015.

3. A. Foruzan, R. Zoroofi, Y. Sato, M. Hori, A Hessian-based filter for vascular segmentation of noisy hepatic CT scans, Int. J. Comput. Assisted Radiol. Surg., 7 (2012), 199–205.

4. J. Li, M. Zhang, Y. Gao, Vessel segmentation of liver CT images by hessian-based enhancement, International Conference on Image and Graphics, (2019), 442–445.

5. Ola Friman, Milo Hindennach, Caroline Kühnel, and Heinz-Otto Peitgen. *Multiple hypothesis template tracking of small 3d vessel structures*. Medical image analysis, 14(2):160–171, 2010.

6. Suheyla Cetin and Gozde Unal. *A higher-order tensor vessel tractography for segmentation of vascular structures*. IEEE transactions on medical imaging, 34(10):2172–2185, 2015.

7. Suheyla Cetin, Ali Demir, Anthony Yezzi, Muzaffer Değertekin, and Gozde Unal. *Vessel tractography using an intensity based tensor model with branch detection.* IEEE transactions on medical imaging, 32(2):348–363, 2012.

8. Ronneberger, O.; Fischer, P.; Brox, T. *U-Net: Convolutional Networks for biomedical image segmentation*. In MICCAI; Springer:Cham, Switzerland, 2015.

9. Long, J.; Shelhamer, E.; Darrell, T. *Fully convolutional networks for semantic segmentation*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.

10. Suheyla Cetin and Gozde Unal. *A higher-order tensor vessel tractography for segmentation of vascular structures*. IEEE transactions on medical imaging, 34(10):2172–2185, 2015.

11. Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. *V-net: Fully convolutional neural networks for volumetric medical image segmentation*. In 2016 fourth international conference on 3D vision (3DV), pages 565–571. IEEE, 2016.

12. Thilo Moshagen, Nihal Acharya Adde, Ajay Navilarekal Rajgopal, *Finding hidden-feature depending laws inside a data set and classifying it using Neural Network*, 2021.

13. Qing Huang, Jinfeng Sun, Hui Ding, Xiaodong Wang, and Guangzhi Wang. *Robust liver vessel extraction using 3d u-net with variant dice loss function*. Computers in biology and medicine, 101:153–162, 2018.

14. Jhinzu Yang, Meihan Fu, and Ying Hu. *Liver vessel segmentation based on inter-scale V-Net*. Mathematical Biosciences and Engineering, 18(4) : 4327-4340, 2021.

15. Mian Wu, Yinling Qian, Xiangyun Liao, Qiong Wang, and Pheng-Ann Heng. *Hepatic vessel segmentation based on 3D swin-transformer with inductive biased multi-head self-attention*. 2021.

16. Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. *Swin Transformer: Hierarchical Vision Transformer using Shifted Windows*. Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.

17. Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. *Rethinking the inception architecture for computer vision*. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2818–2826, 2016.

18. Nabil Ibtehaz, and M. Sohel Rahman. *MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation*. Neural Networks 121 (2020): 74-87.

19. Leslie N. Smith. *Cyclical Learning Rates for Training Neural Networks*. IEE winter conference on applications of computer vision 2017.

20. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. *Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification*. Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

21. Alejandro F. Frangi, Wiro J. Niessen, Koen L. Vincken, and Max A. Viergever. *Multiscale vessel enhancement filtering*. International conference on medical image computing and computer-assisted intervention. Springer, Berlin, Heidelberg. 1998.

22. Wei Yu, Bin Fang, Yongqing Liu, Mingqi Gao, Shenhai Zheng, and Yi Wang. *Liver vessels segmentation based on 3D residual U-Net*. In 2019 IEEE International Conference on Image Processing (ICIP), pages 250–254. IEEE, 2019.

23. Gao Huang, Zhuang Liu, Laurens van der Maaten, Kilian Q. Weinberger, *Densely Connected Convolutional Networks*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018.