1. Attention Unet++: A Nested Attention-Aware U-Net for Liver CT Image Segmentation

Accession number: 20210109724396
Authors: Li, Chen (1); Tan, Yusong (1); Chen, Wei (1); Luo, Xin (1); Gao, Yuanming (1); Jia, Xiaogang (1); Wang, Zhiying (1)
Author affiliation: (1) National University of Defense Technology, College of Computer, China
Source title: Proceedings - International Conference on Image Processing, ICIP
Abbreviated source title: Proc. Int. Conf. Image Process. ICIP
Volume: 2020-October
Part number: 1 of 1
Issue title: 2020 IEEE International Conference on Image Processing, ICIP 2020 - Proceedings
Issue date: October 2020
Publication year: 2020
Pages: 345-349
Article number: 9190761
Language: English
ISSN: 15224880
ISBN-13: 9781728163956
Document type: Conference article (CA)
Conference name: 2020 IEEE International Conference on Image Processing, ICIP 2020
Conference date: September 25, 2020 - September 28, 2020
Conference location: Virtual, Abu Dhabi, United arab emirates
Conference code: 165772
Sponsor: The Institute of Electrical and Electronics Engineers Signal Processing Society
Publisher: IEEE Computer Society
Abstract: Liver cancer is one of the cancers with the highest mortality. In order to help doctors diagnose and treat liver lesion, an automatic liver segmentation model is urgently needed due to manually segmentation is time-consuming and error-prone. In this paper, we propose a nested attention-aware segmentation network, named Attention UNet++. Our proposed method has a deep supervised encoder-decoder architecture and a redesigned dense skip connection. Attention UNet++ introduces attention mechanism between nested convolutional blocks so that the features extracted at different levels can be merged with a task-related selection. Besides, due to the introduction of deep supervision, the prediction speed of the pruned network is accelerated at the cost of modest performance degradation. We evaluated proposed model on MICCAI 2017 Liver Tumor Segmentation (LiTS) Challenge Dataset. Attention UNet++ achieved very competitive performance for liver segmentation. © 2020 IEEE.
Number of references: 11
Main heading: Image segmentation
Controlled terms: Computerized tomography - Diseases
Uncontrolled terms: Attention mechanisms - Competitive performance - Encoder-decoder architecture - Liver lesions - Liver segmentation - Liver tumor segmentations - Performance degradation - Related selections
Classification code: 723.5 Computer Applications
DOI: 10.1109/ICIP40778.2020.9190761
Compendex references: YES
Database: Compendex

ATTENTION UNET++: A NESTED ATTENTION-AWARE U-NET FOR LIVER CT IMAGE SEGMENTATION

Chen Li, Yusong Tan, Wei Chen, Xin Luo, Yuanming Gao, Xiaogang Jia, Zhiying Wang*

College of Computer, National University of Defense Technology, China

ABSTRACT

Liver cancer is one of the cancers with the highest mortality. In order to help doctors diagnose and treat liver lesion, an automatic liver segmentation model is urgently needed due to manually segmentation is time-consuming and error-prone. In this paper, we propose a nested attention-aware segmentation network, named Attention UNet++. Our proposed method has a deep supervised encoder-decoder architecture and a redesigned dense skip connection. Attention UNet++ introduces attention mechanism between nested convolutional blocks so that the features extracted at different levels can be merged with a task-related selection. Besides, due to the introduction of deep supervision, the prediction speed of the pruned network is accelerated at the cost of modest performance degradation. We evaluated proposed model on MICCAI 2017 Liver Tumor Segmentation (LiTS) Challenge Dataset. Attention UNet++ achieved very competitive performance for liver segmentation.

Index Terms— Attention, UNet++, Liver Segmentation

1. INTRODUCTION

Liver cancer is one of the most common internal malignancies and the leading cause of cancer death, which pose a huge threat to human health in the world. Early detection and treatment are the keys to increasing liver cancer survival. Identifying the position of liver is a preparation step in diagnosing and plays an indispensable role in disease treatment. Generally, researches on liver segmentation can be categorized in to 2 classes: (1)manual and semi-automatic segmentation, (2)automatic segmentation. Manual segmentation largely relies on experts with advanced technical skills to perform such tasks. Besides, the quality of the segmentation relies heavily on the judgment of experts, which is time-consuming and error-prone. These factors make manual segmentation impractical. Meanwhile, semi-automatic segmentation still requires manual intervention, which leads to biases and errors. As a consequence, automated medical image segmentation has become a preferred choice in this field and has been extensively studied. The models commonly used in medical image segmentation are variants of U-Net and FCN. For instance, Li et al.[1] present a H-Dense UNet for liver and tumor segmentation through hybrid feature fusion layer. Christian et al.[2] proposed a pipeline of two fully convolutional networks for automatic multi-label heart segmentation from CT and MRI volume. Liao et al.[3] proposed a 3D deep neural convolutional network to find all malignant nodules from CT images for assessing lung cancer. However, these methods usually divide the segmentation task into two steps: localization and segmentation. The extra localization step will increase the amount of model parameters and bring extra time consumption. The accuracy of model segmentation also depends heavily on the first step positioning accuracy. Even minor errors in medical image segmentation may lead doctors to diagnose patients mistakes, so the precise segmentation of medical images is still challenging. To address the need for more accurate segmentation result in medical images and inspired by the attention mechanism, we propose Attention UNet++, a reliable segmentation model based on nested U-Net architecture and attention mechanism.

The contributions of our work can be listed as follows:

1) Attention UNet++ is introduced for liver CT image segmentation.
2) Experiments conducted on dataset LiTS show that Attention Gate can focus on target area of the whole image.
3) Attention U-Net++ has the ability to increase the weight of the target region while inhibiting the background region that is unrelated to the segmentation task.
4) Attention UNet++ has superior performance on liver segmentation task compared to other UNet-based models.
5) Pruned Attention UNet++ is accelerated at the cost of little performance due to the introduction of deep supervision.

Our paper is organized as follows. Section 2 will briefly review related works. Section 3 will detail our segmentation methodology. Then we illustrate the experimental setups in Section 4. After that, the experimental results will be displayed and analyzed. Conclusions and future works are given in Section 5.

This work is supported by National Key Research and Development Program of China (No. 2018YFB0204301).

*Corresponding author: ystan@nudt.edu.cn.

2. RELATED WORKS

Nested UNet architecture : Since U-Net[4] was proposed, it has been widely applied in medical image segmentation. Most researches used it as the backbone and made some changes for different segmentation tasks in the past. UNet++[5] is one of the most representative UNet-based segmentation architectures. What distinguishes UNet++ from UNet is that the former redesigned dense skip connections between encoder and decoder at different levels and used nested convolutional blocks. Each nested convolutional block in UNet++ extracts semantic information by several convolution layers. And every convolution layer in the block is connected through dense skip connections so that concatenation layer can fuse different levels semantic information.

Attention Gate(AG): Attention mechanism firstly emerged in the natural language processing(NLP) and quickly gained dominance. It was Non-local[6] proposed by He Kaiming's team that first introduced the attention mechanism to computer vision. Since then, [7] combined the shared network to the field of semantic segmentation. [8] combined with residual and attention mechanism to obtain a deep network. In order to focus on locations that are relevant for target organ, we refer to the method proposed by PASSRnet[9] to add a simple but effective Attention Gate to the network architecture.The architecture of Attention Gate is shown in Fig.1. The inputs of the Attention Gate are the upsampling

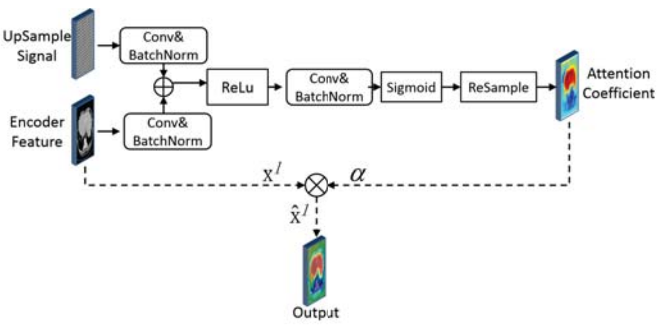


Fig. 1. Diagram of Attention Gate.

feature in the expansion path and the corresponding feature from encoder. The former one is used as gating signal to enhance the learning of the target area related to the segmentation task while suppressing the area irrelevant in the task. Therefore, Attention Gate can improve the efficiency of propagating semantic information through skip connections. Next, the S-shaped activation function sigmoid is selected to train the convergence of the parameters in the Gate and to get the attention coefficient α. Finally, the output can be obtained by multiplying the encoder feature by coefficient α pixel by pixel.

3. METHODOLOGY

Attention Nested UNet(Attention UNet++): We design an integrated network called Attention UNet++ for medical image segmentation. The high-level overview of the Attention UNet++ is shown in Fig.2.

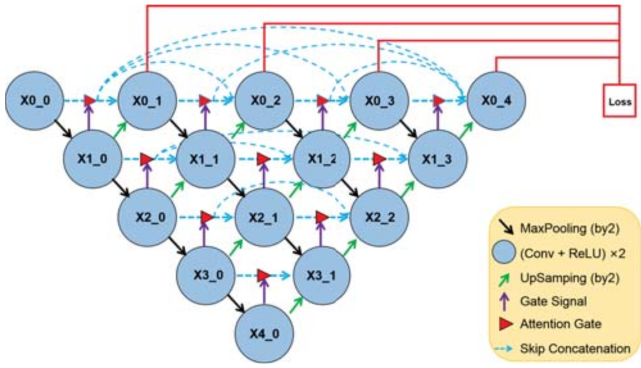


Fig. 2. Architecture of Attention UNet++.

As we can see, Attention UNet++ uses nested U-Net as the basic network framework. Encoder and decoder are symmetrically arranged on both sides of the network. The context information extracted by the encoder is propagated to the decoder of the corresponding layers through the dense skip connections, so that more efficient hierarchical features can be extracted. After received the features from each layer, decoder restores the features in a bottom-up manner.

We define the feature map as follows, let $X_{i,j}$ represents the output of block where i denotes the feature depth in the encoder and j denotes the depth of the convolution layer in the nested block along the skip connection, so extracted feature map of convolution layer can be defined as:

$$X_{i,j} = \begin{cases} \Phi\left[X_{i-1,j}\right] & ,j=0 \\ \Phi\left[\int_{k=0}^{j-1} Ag(X_{i,k}), Up\left(X_{i+1,j-1}\right)\right] & ,j>0 \end{cases}$$

where $\Phi[\ ]$ denotes the convolution block followed by concatenation merger, $Up()$ and $Ag()$ mean upsampling and the attention gate selection respectly. Detailed analysis of the first skip pathway in Attention UNet++ is shown in Fig.3.

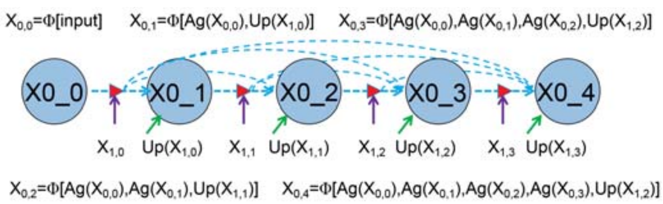


Fig. 3. Detailed analysis of dense skip connections of the first level nested blocks in Attention UNet++.

One of the main innovations in Attention UNet++ is: The network extracts feature from encoder and transfers it to decoder through dense skip connections for integration of hierarchical representations. Besides, Attention Gate mechanism is added between nested convolutional blocks, so that the features extracted at different levels can be merged with a focused selection in the expansion path. As a consequence, the accuracy of Attention UNet++ is enhanced.

Deep supervision: In order to introduce deep supervision, Attention UNet++ adds 1x1 convolutional layer and sigmoid activation function after every output nodes(X_{0_1},X_{0_2}, X_{0_3},X_{0_4}). Fig.4 shows the diagram of deep supervision. Because of dense skip connections in the nested blocks, Attention UNet++ obtains full resolution feature maps at different semantic levels from nodes. In order to fully integrate these semantic information, we redesign a hybrid loss function combining with soft dice coefficient loss and cross entropy loss, which is defined as follows:

$$Loss = \sum_{i=1}^{4}\left(1 - Mean\left[\frac{Y_i \times log\overline{Y}}{2} + \frac{2 \times Y_i \times \overline{Y}}{Y_i^2 + \overline{Y}^2}\right]\right)$$

where $\overline{Y}$ is the real result and Y_i is the segmentation output from node X_{0_i}, $Mean()$ is the average function. $\frac{Y_i \times log\overline{Y}}{2}$ is the dice coefficient loss while $\frac{2 \times Y_i \times \overline{Y}}{Y_i^2+\overline{Y}^2}$ is the cross entropy loss.

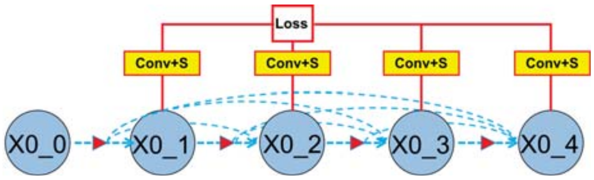
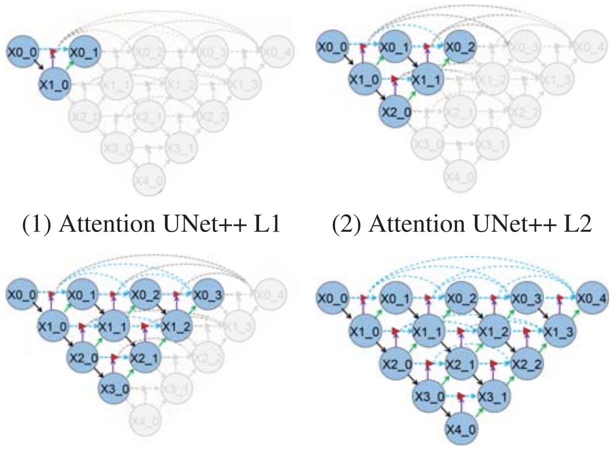


Fig. 4. Diagram of deep supervision in Attention UNet++.

Model pruning: When the trained network predicts the segmentation result, the decoder path at depth d is independent of the deeper decoder path. As a consequence, we completely remove the irrelevant decoder paths and use smaller trained Attention UNet++ to segment at depth d. In Fig.5, We use Attention UNet++ L1,L2,L3,L4 to denote network pruned at four different depth. Pruned Attention UNet++ LN means taking final result from X_{0_N}. Selecting the extent of model pruning is weighed by evaluating the performance and inference time of the four models during validation. For example, if final result comes from X_{0_1}, Attention UNet++ L1 is a maximally pruned architecture. Similarly, there is no pruning in Attention UNet++ L4 when final result comes from X_{0_4}.

4. EXPERIMENTS AND RESULTS

We evaluate proposed method on the public dataset from the MICCAI 2017 Liver Tumor Segmentation challenge(LiTS).



(1) Attention UNet++ L1 (2) Attention UNet++ L2

(3) Attention UNet++ L3 (4) Attention UNet++ L4

Fig. 5. Trained Attention UNet++ with deep supervision makes segmentation results available at different level. Region in gray means these nodes and attention gates are removed during predicting.

4.1. Dataset setup and Preprocessing

LiTS dataset has 131 training CT scans and 70 test CT scans. The training images were annotated. We split the annotated cases at 5:1 ratio into training dataset and test dataset respectively. In LiTS, the ground truth segmentation provides three different labels: liver, tumor(lesion) and background. For image preprocessing, we only consider liver as positive class and others as negative class. And then, we truncate the Hounsfield units(HU) value range of all images in the database to [-200,200] to remove irrelevant useless details.

4.2. Evaluation Metrics

We use dice similarity coefficient, intersection over union (IoU), precision and recall as performance indicators to evaluate the performance of liver CT image segmentation. The larger the values of these four indicators, the larger the overlapping area between the segmentation result and the ground true, the higher the similarity, and the greater the accuracy of the segmentation.

4.3. Experimental results

Segmentation results: Table 1 compares UNet, R2U-Net, UNet++, Attention U-Net, Attention R2U-Net and Attention UNet++ for the liver segmentation experiment. It can be known that under the evaluation of four indicators, the proposed Attention Nested UNet(Attention UNet++) outperforms for liver CT image segmentation in six methods, achieving the IoU ratio increased 7.99%, the Dice coefficient increased by 3.7%, and the precision increased 5%, recall rate increased 4% over U-Net. The comparison between the seg-

Table 1. Segmentation results of methods on the test dataset

Method	IoU(%)	Dice(%)	Precision	Recall
U-Net[4]	89.49	94.45	0.93	0.95
R2U-Net[10]	90.69	05.11	0.94	0.96
UNet++[5]	94.46	97.15	0.98	0.96
Attention UNet[11]	93.39	96.58	0.97	0.96
Attention R2U-Net	92.38	96.03	0.97	0.95
Attention U-Net++	**97.48**	**98.15**	**0.98**	**0.99**

mentation results and the manually annotated result is shown in Fig.6.

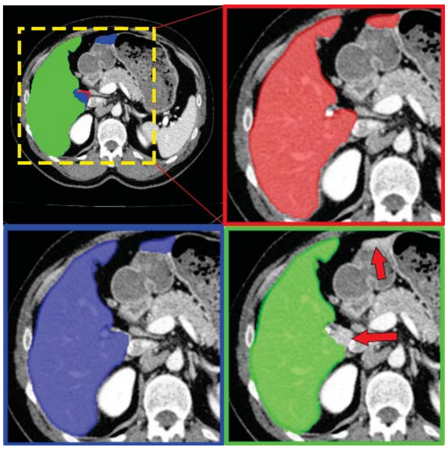


Fig. 6. Comparison between liver segmentation results and manually annotated result. The ground true is shown in a red region, the prediction results using the Attention UNet++ and U-Net are displayed in a blue region and green region. U-Net's prediction error area is highlighted with a red arrow.

Attention learning results: As shown in Fig.7, Attention coefficient α gradually changes during training, weights in the target organ (red) where related to the segmentation task is enhancing while confusing tissue (blue) is suppressing. So introducing attention mechanism can enhance the learning of target regions and increases the effectiveness of skip connections. Besides, there is no need to trim the ROI and locate the target object in our model.

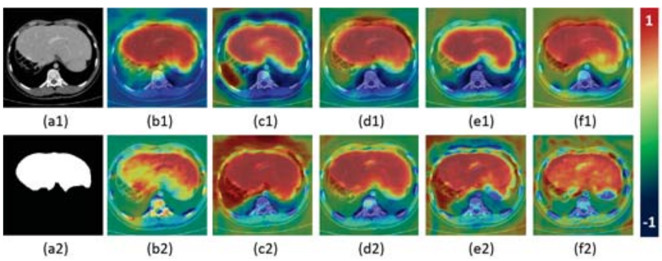


Fig. 7. (a1) and (a2) are the CT images and groud true in LiTS. (b1) to (f1) are the attention coefficients in the first layer of attention gates in different training periods. (b2) to (f2) are the attention coefficients in the second layer of attention gates in different training periods.

Model pruning results: As we can see in Fig.8, Attention UNet++ L3 achieves on average 7.734% reduction in prediction time and 75.512% reduction in parameters while decreasing IoU by only 0.615% and decreasing Dice by only 2.558%. Shallowest Attention UNet++ L1 achieves on average 17.128% reduction in prediction time and 98.84% reduction in parameters while decreasing IoU by 13.354% and decreasing Dice by 27.18%. So we conclude that model prun-

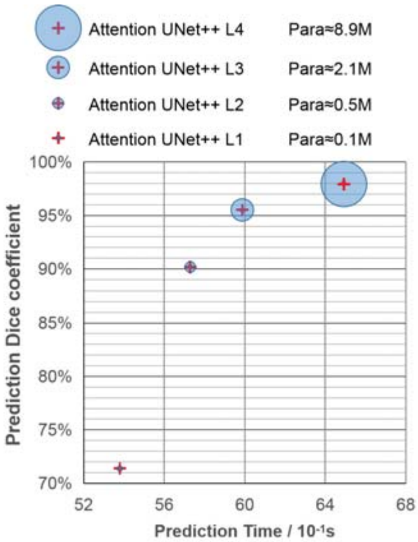


Fig. 8. Parameters, inference time, and Dice coefficient of Attention UNet++ under different extent of pruning.

ing can significantly reduce the model parameters and prediction time, but segmentation performance decline at same time. Therefore, we need to make a reasonable judgment based on the actual situation before pruning. In addition, since most deep CNN segmentation models take a long time to calculate and require a large amount of memory, it is more practical to apply the pruned segmentation model to computer-aided diagnosis in small computers, especially in mobile devices.

5. CONCLUSIONS

In this paper, a nested attention-aware segmentation network(Attention UNet++) is proposed and applied to liver segmentation. Our experiment on the LiTS dataset demonstrated that the competitive performance of proposed Attention UNet++ in liver CT image segmentation. The improvement is attributed to the the combination of dense skip connection and attention mechanism. Experiment also proved that Attention UNet++ has the ability to increase the weight of the target region while inhibiting the background region that is unrelated to the segmentation task. Besides, due to the introduction of deep supervision, the prediction speed of the pruned Attention UNet++ is accelerated at the cost of little performance degradation.

6. REFERENCES

[1] 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.

[2] Christian Payer, Darko Štern, Horst Bischof, and Martin Urschler, "Multi-label whole heart segmentation using cnns and anatomical label configurations," in *International Workshop on Statistical Atlases and Computational Models of the Heart*. Springer, 2017, pp. 190–198.

[3] Fangzhou Liao, Ming Liang, Zhe Li, Xiaolin Hu, and Sen Song, "Evaluate the malignancy of pulmonary nodules using the 3-d deep leaky noisy-or network," *IEEE transactions on neural networks and learning systems*, 2019.

[4] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[5] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 3–11. Springer, 2018.

[6] Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He, "Non-local neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7794–7803.

[7] Liang-Chieh Chen, Yi Yang, Jiang Wang, Wei Xu, and Alan L. Yuille, "Attention to scale: Scale-aware semantic image segmentation," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[8] 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 preprint arXiv:1811.01328*, 2018.

[9] Longguang Wang, Yingqian Wang, Zhengfa Liang, Zaiping Lin, Jungang Yang, Wei An, and Yulan Guo, "Learning parallax attention for stereo image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 12250–12259.

[10] Md Zahangir Alom, Mahmudul Hasan, Chris Yakopcic, Tarek M Taha, and Vijayan K Asari, "Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation," *arXiv preprint arXiv:1802.06955*, 2018.

[11] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al., "Attention u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.