MORE-UNET: MULTI-OUTPUTS RECURRENT UNET FOR MEDICAL IMAGE SEGMENTATION

[1]SONG-TOAN TRAN, [2]CHING-HWA CHENG, [3]THANH-TUAN NGUYEN, [4]MINH-HAI LE, [5]DON-GEY LIU

[1]Ph.D. Program in Electrical and Communication Engineering, Feng Chia University, 40724, Taichung, Taiwan, R.O.C.
[2]Department of Electronic Engineering, Feng Chia University, 40724 Taichung, Taiwan, R.O.C.
E-mail: [1]tstoan1512@tvu.edu.vn, [2]chengch@fcu.edu.tw, [3]nttuan@kgc.edu.vn, [4]lmhai@tvu.edu.vn, [5]dgliu@fcu.edu.tw

Abstract - Medical image segmentation has been an interesting topic in recent years. Deep learning is showing a superiority to traditional methods. The disadvantage of deep learning networks used standard convolution neural network is that the size of the model is often large, resulting in a lot of memory consumption and large computation time. In this paper, we propose a new model named MORe-Unet, which combines the use of Recurrent convolutional neural network (RCNN) based on Unet architecture to reduce the network size with exploitation of all output feature maps in convolution units of the network nodes. We evaluated the effectiveness of the models on liver segmentation application on LiTS 2017 dataset and Spleen segmentation on Medical Segmentation Decathlon Challenge 2018 datasets. Experimental results show that the proposed model has smaller size and improved performance compared to current models.

Keywords - Recurrent Convolutional Neural Network, Medical Image Segmentation, Liver Segmentation, Spleen Segmentation, Unet Architecture.

I. INTRODUCTION

Medical image segmentation is one of the challenges that has been addressed in recent years. There are many proposed methods for medical image segmentation challenges including: traditional image processing, basic machine learning, and deep learning [1]. Deep learning is the most commonly used method, the number of published studies has increased by more than 100 times between 2014 and 2018 [2]. There are many medical image segmentation applications that used convolution neural network (CNN) such as liver and liver tumor [3]-[5], brain and brain tumor [6]-[8], lung and lung nodule [9], [10], etc.

Unet is the most commonly used deep learning architecture [11]. However, Unet [12] and the architectures based on Unet are mostly large in size and time consuming. Changing the number of filters as well as the number of nodes will increase the size of the model. Some models have improved this weakness by using dense architecture [13], [14], or changing the structure of the convolution function [15], or using the recurrent convolutional neural network (RCNN) [16], [17]. Dense architectures were decreased the number of parameters but the network structure would be more complex, computation time was also increased. It is proposed to change the structure of the convolution to improve the performance of the model, but the network size is not reduced. RCNN reduced the network size by reusing the weights of convolution. In this study, we use recurrent convolutional neural network (RCNN) instead of traditional CNN structure. The advantage of RCNN structure is to reduce the number of parameters while ensuring the quantity of convolution function. Therefore, the network size will be reduced. In this article, we also combined the extraction of all feature maps from convolution functions in Unet's nodes. Fig. 1 depicts the difference between traditional CNN and RCNN proposed in this study. In our network, the output feature maps are used for next layers and the decoder side of Unet structure.

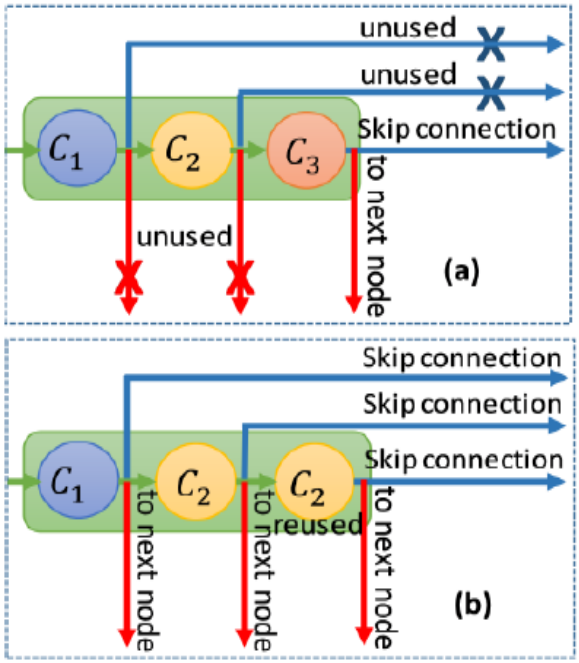


Figure 1. The difference between the node of (a) traditional Unet and (b) the proposed model

In summary, in this research, we have contributed the following technical problems: (1) proposed a new network model, MORe-Unet, combining the RCNN structure with the exploitation of feature maps in the network node; (2) evaluated the proposed model on liver segmentation and spleen segmentation; (3) experimental results showed the improvements of our

proposed model: the network size was significantly reduced and the segmentation performance was better than the current models.

The content of the paper is divided into 5 main sections: section I is introduction, section II describes the MORe-Unet network architecture, the experimental are presented in section III, section IV is the results and discussion, and the final section is conclusion.

II. PROPOSED NETWORK- MORE-UNET

A. Network architecture

Fig. 2 presents the overview of MORe-Unet models. The skip connection path, max-pooling, and up-convolution path consist of more than one path. The reason is that all of output feature maps from the encoder node were used. The encoder nodes and decoder nodes were the RCL blocks.

The output states of the RCL block depend on the number of convolutional units reused. In this study, we investigate two cases that are t = 1 and t = 2, which is the number of times the convolution unit is reused in RCL. In case t = 1, MORe-Unet will consist of 3 sub-Unet models (Fig. 3 (a)). When t = 2, MORe-Unet will consist of 4 sub-Unet models (Fig. 3 (b)). Each convolution unit in the RCL block consists of a convolutional function with a kernel size of 3x3 followed by a ReLU activation function. The details of the connection between the encoder node and the decoder node are shown in Fig. 4.

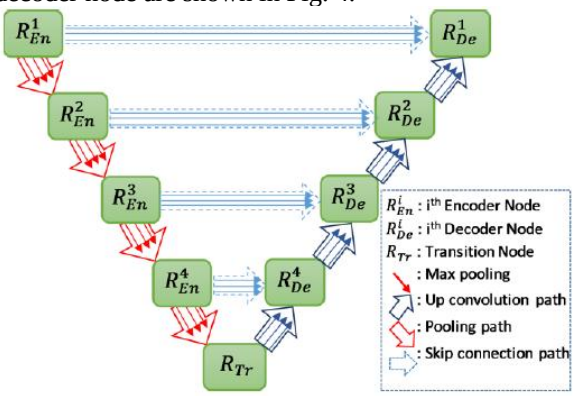


Figure 2. The overview structure of MORe-Unet

B. Connectivity details

In this section, connections and output feature map will be analyzed in detail. Let ${}_iE^j$ be the output feature map of the jth convolution unit of the ith encoder node. ${}_iE^j$ is given by:

$$E_i^j = \begin{cases} C\left(M\left(E_{i-1}^j\right)\right) & j = 1 \\ C\left(\left[M\left(E_{i-1}^j\right), E_i^{j-1}\right]\right) & j > 1 \end{cases} \quad (1)$$

where C (•) is the convolution function, [•] is the concatenate function, M (•) denotes the max-pooling function, and *j* is the number of convolution unit.

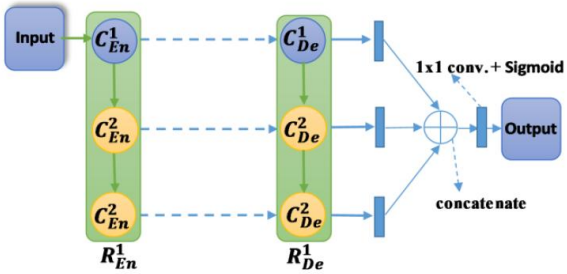


Figure 4. The connection between encoder node and decoder node in the first level.

In the decoder side, the inputs of the convolution unit consist of the feature maps after applied up-convolution from the previous nodes, the feature maps from the encoder node at the same level, and the feature maps from the previous convolution units in the same node. The calculation equation is described as

$$D_i^j = \begin{cases} C\left(\left[T\left(D_{i+1}^j\right), E_i^j\right]\right) & j = 1 \\ C\left(\left[T\left(D_{i+1}^j\right), E_i^j, D_i^{j-1}\right]\right) & j > 1 \end{cases} \quad (2)$$

where D_j is the output feature of the jth convolution in the ith decoder node, T (•)denotes the transposed convolution.

The final output of the MORe-Unet is obtained by concatenating all of output feature maps from the first level decoder node then applied the 1x1 convolution and sigmoid function. Let Y is the final output, which is given by:

$$Y = \zeta\left(\Psi\left(\left[\zeta(\Psi(D^j))\right]_{k=1}^{j}\right)\right) \quad (3)$$

where $\zeta(\bullet)$ is the sigmoid function, $(\bullet)$ depicts the 1x1 convolution. The filters number in the convolution units in the nodes are {16; 32; 64;128; 256;128; 64; 32;16} corresponding with $\left\{R^{En}; R^{En}; R^{En}; R^{En}; R^{Tr}; R^{De}; R^{De}; R^{De}; R^{De}\right\}$.

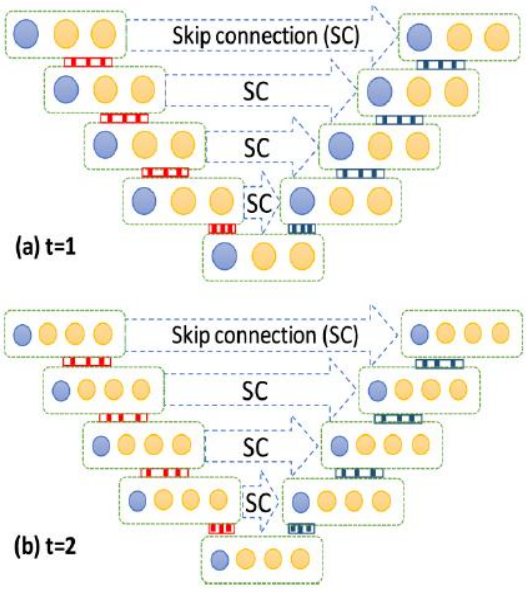


Figure 3. The details of MORe-Unet when (a) t = 1 and (b) t =2)

III. EXPERIMENTS

A. Dataset and preprocessing

To evaluate the model's performance, we implemented two fractions applications, liver segmentation and spleen fractionation. Details of datasets and preprocessing are presented in this section.

For the liver segmentation application, the dataset used for training is provided by the 2017 LiTS challenge. It consists of 131 CT volumes for training and validation and 70 volumes for testing. However, the annotation is included 131 volumes, thereby, in our experiments, we only use the volume which included the ground truth. There are two main goals in the LiTS challenge that are liver segmentation and liver tumor segmentation. In this study, we only consider to implement the liver segmentation task. Tumor labels are referred to as liver labels. The total slices of 131 volumes is 58638, the size of the slices is 512x512. For training and testing, the dataset was divided into three subsets: 90 volumes for training, 11 volumes for validation, and 30 volumes for testing. The preprocessing of the dataset was implemented as follow: firstly, the slice was cropped to the size of 448x448 then scaled to the size of 224x224. This process helps to reduce the computation time. Secondly, we applied the Hounsfield unit threshold in range of [-200,250]. This helps remove the redundant parts around the liver and the noise. Finally, we combine three adjacent slices into a RGB image. We also removed some slices without the liver inside. In summary, there are total 22109 images used in our 4494 images for validation, and 7059 images for testing, validation, and testing.

For the spleen segmentation application, the dataset is provided by the Medical Segmentation Decathlon Challenge 2018. There are 63 CT volumes in the dataset. However, there are only 41 volumes were included the ground truth.

The total number of slices in 41 volumes is 3650 with the size of 512x512. The dataset was randomly divided into three subsets: training set includes 2920 slices, validation set consists of 584 slices, and testing set include 730 slices. For preprocessing, we also applied the Hounsfield Unit threshold in the range of [-200,250]. The image was cropped to a size of 224x224 before used for training.

B. Loss function

The loss function used in our study was a hybrid loss function. Because of the imbalance data, we combine the weighted cross-entropy and the dice loss. The expressions of the loss function are following as

$$L_{total} = L_W + L_D, \tag{4}$$

$$L_W = -\frac{1}{M}\sum_{i=1}^{M}\left((1-\omega)\, p_i \log \hat{p}_i + \omega(1-p_i)\log(1-\hat{p}_i)\right) \tag{5}$$

$$L_D = 1 - \frac{2\sum_{i=1}^{M}(p_i \hat{p}_i) + \varepsilon}{\sum_{i=1}^{M}(p_i + \hat{p}_i) + \varepsilon} \tag{6}$$

Where L_{total} is the total loss, L_W denotes weighted cross-entropy loss, and L_D presents the dice loss. p_i is the background pixel recognition probability, $\hat{p}_i$ is the foreground pixel recognition probability. M is the total pixels number The ω is the weight of foreground. The ε is tiny value for preventing the infinitive problem.

C. Experimental setting

Our model was implemented by Keras package. The number of the epochs was 100 for liver segmentation and 200 for spleen segmentation. The batch size was 8. We also applied the learning rate scheduler with the initial value was 3e-4. The learning rare was update by *initial* $*(0.9^{epochs/10})$. The early-stopping mechanism was also used to save the training time. A work station computer with Intel Xeon Silver 4114 CPU, GRID Virtual GPU V100D-8Q, and 32GB of RAM memory was used to train our models. The augmentation technique was used for the spleen dataset.

IV. RESULTS AND DISCUSSION

A. Evaluation metrics

The metrices used to evaluate the models performance are: Dice coefficient (DC), F1-score (F1), and mean Intersection over Union (IoU). They are described as follow,

$$DC(X, \hat{X}) = \frac{2 \times |X \cap \hat{X}|}{|X| + |\hat{X}|} \tag{7}$$

$$IoU(X, \hat{X}) = \frac{|X \cap \hat{X}|}{|X \cup \hat{X}|} \tag{8}$$

$$F1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \tag{9}$$

where X is the number of positive class pixels of ground truth, $\hat{X}$ is the number of pixels predicted as the positive class. The TP, FP, and FN are cases number of true positive, false positive, and false negative, respectively.

B. Segmentation results

In this study, we trained total of four network model: Unet, Unet++ [14], MORe-Unet (t=1), and MORe-Unet (t=2). Table 1 compares the segmentation results on liver segmentation and parameters number of proposed model with Unet, Unet++.

Models	No. paras (million)	Metrics (%) DC	F1	IoU
Unet	7.77 M	91.26	91.31	84.07
Unet++	9.51 M	94.75	95.19	91.08
MORe-Unet (t=1)	**3.48 M**	***95.86***	***96.04***	***92.06***
MORe-Unet (t=2)	***4.64 M***	**96.25**	**96.43**	**92.78**

Table 1. The segmentation results on liver segmentation

As seen in Table 1, our proposed models achieved the better results than Unet and Unet++. For MORe-Unet (t = 1), the metrics are increased by 4.60%, 4.73%, and 7.99% compared with Unet in terms of DC, F1, and IoU, respectively. For MORe-Unet (t=2), the increaments are 4.99%, 5.12%, and 8.71%, respectively. Furthermore, our proposed models consume less memory than Unet and Unet++ (3.48 million, 4.64 million against 7.77 million, 9.51 million).

Table 2 shows the segmentation results on the spleen dataset. We can obsever that the MORe-Unet (t =2) is obtained the best results on all of the metrics. Compared with Unet model, our proposed models show the sigfinicant improvements.

Models	No. paras (million)	Metrics (%) DC	F1	IoU
Unet	7.77 M	90.28	90.30	82.93
Unet++	9.51 M	95.00	95.42	90.49
MORe-Unet (t = 1)	**3.48 M**	***96.30***	***96.56***	***92.88***
MORe-Unet (t = 2)	***4.64 M***	**96.44**	**96.62**	**93.15**

Table 2. The segmentation results on spleen dataset

Fig. 5 presents some examples of segmentation results on liver and spleen segmentations. We can see that, our proposed models show the better segmentation results on liver and spleen segmentation than that of Unet and Unet++. That proves the effectiveness of our proposed network models.

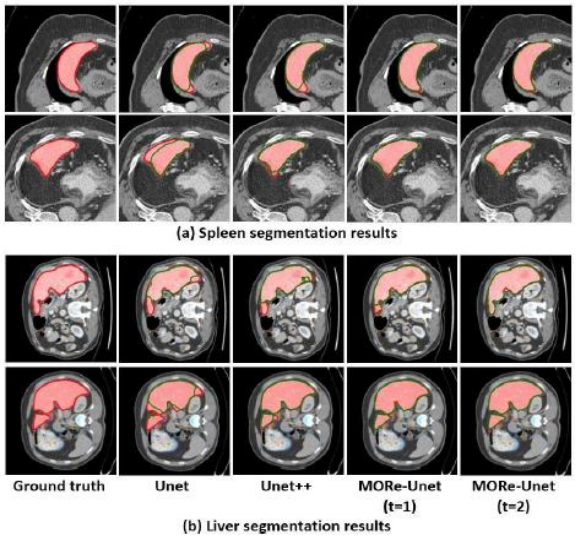


Figure 5. Examples of segmentation results on (a) liver slices and (b) spleen slices. Red curve denotes the ground true, green one depicts the boundaries of the predicted organs

C. Discussion

Unet architecture is demonstrating outstanding success in the medical image segmentation challenges. However, the main disadvantage of network models is that the large network size leads to a lot of memory consumption and training time. Changing the convolution architecture, improving the skip connection, or using dense structures are responsible for the increase in the network size. RNN has been shown the effectiveness in reducing network size. The adoption of RCLs in place of the Unet's network node has significantly reduced the network size. Combining with exploiting all the output features of the convolutional units has improved model performance. Fig. 6 illustrates the evolution of the validation loss and validation dice values on spleen segmentation. In the figure, we see that our proposed models achieve the lowest validation loss and the highest validation dice.

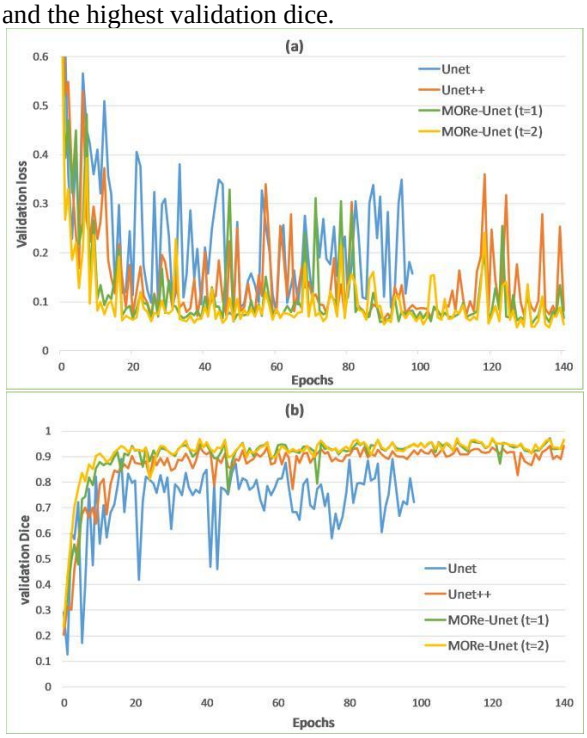


Figure 6. The learning curves of training process on spleen segmentation: (a) validation loss and (b) validation dice.

Table 3 compares the results of the proposed model with popular models. Compared with current models, the proposed model has outperformed results except for Unet 3+ on liver segmentation. However, compared to the network size, Unet 3+ has a much larger network size than the proposed models (26.97 million against 3.48 million and 4.64 million).

Models	DC (%) Liver	Spleen*
PSPNet [18]	92.42	92.40
Deeplab V3+ [19]	91.86	92.90
Attention Unet [20]	93.41	93.24
Unet 3+ [13]	**96.75**	96.20
MORe-Unet (t = 1)	96.30	***96.56***
MORe-Unet (t = 2)	***96.44***	**96.62**

* comparing with different dataset

Table 3. Comparing with popular network models

V. CONCLUSION

In this study, we proposed a new network model for medical image segmentation. The new model, MORe-Unet, is the combination of RCNN and the exploitation the output feature maps in the node of the network. The experiments have shown the effectiveness of proposed network. The MORe-Unet not only reduced the network size but also outperformed the popular networks on the segmentation results. This would be a new approach for medical image segmentation in the future.

REFERENCES

[1] T. Lei, R. Wang, Y. Wan, X. Du, H. Meng, and A. K. Nandi, "Medical Image Segmentation Using Deep Learning: A Survey," ArXiv200913120 Cs Eess, Sep. 2020. [Online].

[2] T. Zhou, S. Ruan, and S. Canu, "A review: Deep learning for medical image segmentation using multi-modality fusion," Array, vol. 3–4, p. 100004, Sep. 2019.

[3] 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 Trans. Med. Imaging, vol. 37, no. 12, pp. 2663–2674, Dec. 2018.

[4] H. Seo, C. Huang, M. Bassenne, R. Xiao, and L. Xing, "Modified U-Net (mU-Net) With Incorporation of Object-Dependent High Level Features for Improved Liver and Liver-Tumor Segmentation in CT Images," IEEE Trans. Med. Imaging, vol. 39, no. 5, pp. 1316–1325, May 2020.

[5] X.-F. Xi, L. Wang, V. S. Sheng, Z. Cui, B. Fu, and F. Hu, "Cascade U-ResNets for Simultaneous Liver and Lesion Segmentation," IEEE Access, vol. 8, pp. 68944–68952, 2020.

[6] B. H. Menze et al., "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," IEEE Trans. Med. Imaging, vol. 34, no. 10, pp. 1993–2024, Oct. 2015.

[7] L. Wang et al., "Nested Dilation Networks for Brain Tumor Segmentation Based on Magnetic Resonance Imaging," Front. Neurosci., vol. 13, p. 285, Apr. 2019.

[8] J. Zhang, X. Lv, H. Zhang, and B. Liu, "AResU-Net: Attention Residual U-Net for Brain Tumor Segmentation," Symmetry, vol. 12, no. 5, p. 721, May 2020.

[9] G. Aresta et al., "iW-Net: an automatic and minimalistic interactive lung nodule segmentation deep network," Sci. Rep., vol. 9, no. 1, p. 11591, Dec. 2019.

[10] N. V. Keetha, S. A. B. P, and C. S. R. Annavarapu, "U-Det: A Modified U-Net architecture with bidirectional feature network for lung nodule segmentation," ArXiv200309293 Cs Eess Stat, Mar. 2020.

[11] S. Panayides et al., "AI in Medical Imaging Informatics: Current Challenges and Future Directions," IEEE J. Biomed. Health Inform., vol. 24, no. 7, pp. 1837–1857, Jul. 2020.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, vol. 9351, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.

[13] H. Huang et al., "UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation," in ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, May 2020, pp. 1055–1059.

[14] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation," IEEE Trans. Med. Imaging, vol. 39, no. 6, pp. 1856–1867, Jun. 2020.

[15] Y. Chen et al., "Channel-Unet: A Spatial Channel-Wise Convolutional Neural Network for Liver and Tumors Segmentation," Front. Genet., vol. 10, p. 1110, Nov. 2019.

[16] Ming Liang and Xiaolin Hu, "Recurrent convolutional neural network for object recognition," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015, pp. 3367-3375.

[17] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation," ArXiv180206955 Cs, May 2018.

[18] H. S. Zhao, J. P. Shi, X. J. Qi, X. G. Wang and J. Y. Jia, "Pyramid scene parsing network", The IEEE Conference on Computer Vision and Pattern Recognition, pp. 2881-2890, 2017.

[19] L-C. Chen, Y. K. Zhu, G. Papandreou and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation", Proceedings of the European Conference on Computer Vision, 2018.

[20] O. O et al., "Attention u-net: Learning where to look for the pancreas", Medical Imaging with Deep Learning, 2018.

★★★