U-Net with Dense Encoder, Residual Decoder and Depth-wise Skip Connections

1st Weiqin Ying
School of Software Engineering
South China University of Technology
Guangzhou 510006, PR China

2nd Junhui Li
School of Software Engineering
South China University of Technology
Guangzhou 510006, PR China

3rd Yu Wu*
School of Computer Science and Cyber Eng.
Guangzhou University
Guangzhou 510006, PR China
wuyu@gzhu.edu.cn (Corresponding author)

4th Kaijie Zheng
School of Software Engineering
South China University of Technology
Guangzhou 510006, PR China

4th Yali Deng
School of Software Engineering
South China University of Technology
Guangzhou 510006, PR China

5th Jiachen Li
School of Software Engineering
South China University of Technology
Guangzhou 510006, PR China

***Abstract*—For tasks like medical image segmentation and understanding, U-Net is one of the most prominent convolutional neural networks (CNNs) in recent years. Most of the models for image segmentation today are the variants of the classical U-Net. By applying some improvements on the convolution blocks and skip connections in U-Net, this paper proposes a dense-residual depth-wise U-Net (DR-DW U-Net). The DR-DW U-Net aims at extracting more useful features and alleviating the pressure during gradient descent. It adopts dense blocks in the left encoding path while using residual blocks in the right decoding path. In addition, the skip connections of DR-DW U-Net are injected with additional convolution blocks in the form of depth-wise convolution. The experimental results on two medical segmentation datasets indicate that the DR-DW U-Net achieves competitive performances and notable improvements over the classical U-Net as well as some well-designed variants of U-Net.**

***Index Terms*—convolutional neural network, medical image segmentation, U-Net, convolution block, skip connection**

I. Introduction

The correct and accurate segmentation or detection of lesion is critical in the procedure of the diagnosis of many diseases. The introduction of convolutional neural networks has brought a significant breakthrough in computer vision compared with the traditional algorithm. In segmentation tasks, the state-of-the-art models for medical image segmentation are variants of the encoder-decoder architecture like U-Net [1], while other more common tasks are variants of fully convolutional networks [2] and Mask R-CNN [3]. Many researchers have been working together for many years to find better models to correctly extract lesions or organs in different tasks, such as optic disc segmentation [4], and organs segmentation on abdominal CT [5].

Before deep learning, approaches to medical image segmentation were often based on edge detection and template matching [6]. Unfortunately, these traditional methods cannot always promise a satisfactory performance. The development of deep learning and CNNs offered a different theory in image processing and understanding. From the very beginning, CNNs focused on classification problems with the help of fully connected layers. Then fully convolution networks [1], [2] were developed to solve pixel-level classification problems, where the resolution of output of networks was expected to be the same as the input, rather than collapsing into short vectors produced by fully connected layers.

However, the correct detection and segmentation in medical images and videos are still great challenges. The reasons can be concluded as follows. In most cases, the amount of training and testing samples of medical images is much smaller than widely-used datasets like the COCO dataset [7], resulting in the higher risk of overfitting. The annotations of some medical images could be slightly uncertain. Some ground truth labels of medical images might be incorrectly annotated.

In this paper, a dense-residual depth-wise U-Net (DR-DW U-Net) is presented to ameliorate the ability of information propagation and feature extraction of U-Net, as well as to reduce the probability of overfitting and degradation. All the convolution blocks in encoder stages appear in the form of dense blocks, while the decoder stages adopt residual blocks. And additional depth-wise convolution blocks are injected in the skip connections in DR-DW U-Net.

II. Related work

The classical U-Net consists of an encoder path for capturing context and a decoder path that enables precise localization with skip connections from the previous encoder path. An example of the classical U-Net model is shown in Fig. 1.

This work was supported in part by the Natural Science Foundation of Guangdong Province under Grant 2020A1515011491, in part by the National Natural Science Foundation of China under Grant 61203310 and Grant 61503087, in part by the Pearl River S&T Nova Program of Guangzhou under Grant 2014J2200052, in part by the Research and Development Program in Key Areas of Guangdong Province under Grant 2019B010154004, in part by the Fundamental Research Funds for the Central Universities, SCUT, under Grant 2017MS043, and in part by the Platform Development Program for Innovation and Entrepreneurship at Colleges in Guangzhou under Grant 2019PT103.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

The encoder and decoder of the classical U-Net are connected horizontally by the skip connections using concatenation.

Many variations of U-Net have been developed for different medical image segmentation tasks. A modified U-Net framework, called M-Net [4], introduces multi-scale inputs and deep supervision for joint optic disc and cup segmentation. By adding residual connections and dilated convolutions, Stefanos et al. proposed a branch U-network (BRU-net) [8] to segment pathological OCT retinal layer for age-related macular degeneration diagnosis. Gibson et al. introduced dense connection in each encoder block to automatically segment multiple organs on abdominal CT [5]. U-Net++ [9] aims at reducing the semantic gap between the feature maps of the encoder and decoder. U-Net++ can be globally viewed as a triangle nested network where the skip connections here look like dense net [10]. And cascaded CNNs can be selected as potential solutions if the target organs in medical images show large inter-patient variation in terms of shape and size.

The main reasons that U-Net and its related architectures work quite well can be concluded as follows. First, down-sampling makes the model more robust to the variations of the input images, reduce the computation cost, decrease the probability of over-fitting and increase the acceptive field to capture high semantic information, while up-sampling can gradually recover the resolution of the feature maps. Then, the skip connections help recover the full spatial resolution at the network output and heighten the suitability of fully convolutional methods for semantic segmentation.

However, it is not always feasible to copy the architecture from one task to others. Some of the drawbacks of the classical U-Net include the following. First, it can be a little hard to train the classical U-Net, since all the convolution operations in the classical U-Net are plain stack of convolution layers. Second, the downsampling method and upsampling method in the classical U-Net are max-pooling and upsampling with interpolation, which may result in information loss. Finally, the ReLU activation function in the classical U-Net might also lead to information loss when transforming high-dimensional features to low-dimensional features.

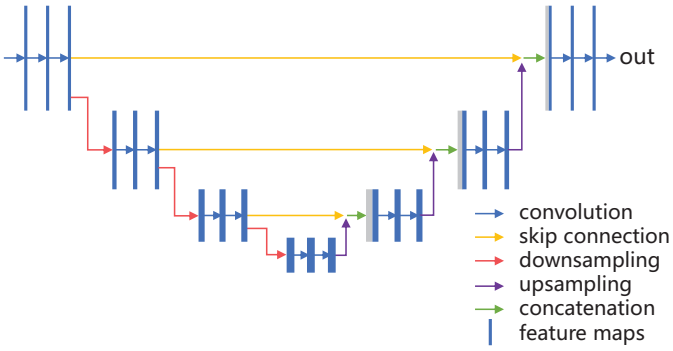


Fig. 1. An example of U-Net architecture.

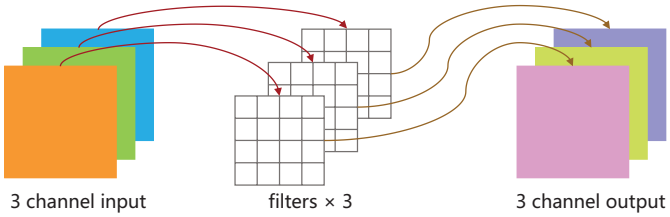


Fig. 2. An example of depth-wise convolution. The input here is 3-channel feature maps. Each filter corresponds to one channel of input, yielding one channel of output feature map.

III. Methodology

To address the issues mentioned above and improve the performance of U-Net, a dense-residual depth-wise U-Net (DR-DW U-Net) is proposed in this paper. The detailed design of the structure of DR-DW U-Net is introduced in this section.

A. Max pooling and strided depth-wise convolution

Both the average pooling and max pooling might result in information loss [11]. On the contrary, a convolution layer with stride 2 helps to retain more spatial information, as well as to shrink the size of the feature maps. One drawback of the strided convolution is that it introduces more trainable parameters.

In this paper, the down-sampling operation combines the max pooling and strided depth-wise convolution together by pointwise summation. In the depth-wise convolution, each channel of the input corresponds to only one of the convolution filters, as shown in Fig. 2. It helps to learn features specifically for each channel, as well as to reduce the number of parameters [12]. The depth-wise convolution here can be viewed as complement for pooling operation. Symmetrically, the upsampling operation combines the traditional upsampling with deconvolution by pointwise summation.

B. Residual block and dense block

Adding more layers to a suitably deep model may lead to higher training error where the accuracy gets saturated and then degrades rapidly. A residual network with shortcut connections was proposed [13] to address this problem. A residual block aims at learning the mapping of $H(x) = F(x) + x$ by feedforward neural networks with shortcut connections to improve information flow and ease training, as shown in Fig. 3.

Furthermore, a dense block [10] connects all layers directly with each other. It tries to improve models performance and information flow by introducing direct connections from any layer to all subsequent layers. The residual block and dense block can even be injected in skip connections in U-Net++ [9]. Inspired by these methods, the basic convolution blocks and skip connections are modified in DR-DW U-Net. The details of implementation are discussed in the following subsections.

C. Activation function

The rectified linear unit activation function (ReLU) is the common choice of activation function nowadays. Assuming

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

x is a one-dimension input, the formula of ReLU could be described as follow:

$$f(x) = max(x, 0). \tag{1}$$

It is obvious that the negative part of x does not have any gradient during back propagation, which can lead to some dead cells in networks. In this regard, a parameterized ReLU function [14], PReLU, is adopted as the activation function in DR-DW U-Net, except the final layer. The PReLU goes as:

$$f(x) = \begin{cases} x \ , \ x \geq 0, \\ \gamma x \ , \ x < 0. \end{cases} \tag{2}$$

The trainable parameter γ determines the way how the negative part of x is activated. And γ can be learned using gradient descent. In other words, the activation function actually depends on the different datasets.

D. DR-DW U-Net

In the general U-Net architecture, the convolution blocks used in the contracting path and expansive path are usually identical. In this paper, DR-DW U-Net has different convolution blocks in the contracting path and expansive path, as shown in Fig. 3. The DR-DW U-Net can be globally viewed as a U-Net with a dense net as the encoder backbone and a residual net as the decoder. Notice that a weighted average of the outputs from the last dense block and the final residual block is calculated at the top level of DR-DW U-Net, as shown in the green arrow in Fig. 3.

The implementation of dense blocks and residual blocks in DR-DW U-Net are shown in Fig. 3. In each dense block of DR-DW U-Net, there exists four middle convolution layers which learn additional feature maps based on the "collective knowledge" gained by the previous layers. Dense blocks help to strengthen feature propagation, encourage feature reuse and alleviate the vanishing-gradient problem. Thus, dense blocks are adopted to extract more useful information in the encoder. However, in the destination of skip connection, the feature maps from the encoder and decoder are combined through concatenation. The dense block in the decoder can output much larger feature maps, leading to more computation cost. Thus, the residual blocks are adopted in the decoder of DR-DW U-Net. Another reason is that the residual function is easier to learn using gradient descent.

To extract more useful features while retaining high resolution representation, the top-level skip pathway contains an additional dense block, depth-wise convolution and residual block. The resolution remains unchanged at the top-level of DR-DW U-Net to alleviate information loss due to the downsampling and upsampling. To further improve the features propagation and heighten the robustness of the predictions, the outputs from the last dense block A and the final residual block B are summed up to a weighted mean Y by $Y = \alpha A + \beta B$. A and B are of the same shape. Both α and β have the same number of channels as A and B. α and β are learnable parameters in DR-DW U-Net. It helps to ensure that all the convolution blocks do make their contributions. Another hypothesis is that the feature maps near the input of the network can retain more contour details and texture features, while the feature maps deep down in network have more semantic information. Fusing these two kinds of features helps to obtain a better result.

In order to reduce the computation cost and the number of parameters, the skip connections at the other levels of DR-DW U-Net include only one depth-wise block. One of the reasons is that the features from the skip connections are not always meaningful and useful. To some extent, the additional depth-wise convolution acts as a simple attention gate to decide which features should be retained.

IV. Experiment

The experiment in this paper tried to validate whether the DR-DW U-Net is superior to baseline U-Net, and whether the DR-DW U-Net is competitive against some recent work on U-Net.

A. Networks

Several networks were implemented in our experiment to act as control experiment. For the clarity of notation, the classic U-Net is referred as U-Net, the U-Net with residual blocks as RB U-Net, the U-Net with dense blocks as DB U-Net, the U-Net with residual blocks and residual skip connection as RB-RC U-Net. In RB-RC U-Net, the number of residual blocks in skip connections decreases as the depth go deeper, just like U-Net++.

To summarize, U-Net, RB U-Net, DB U-Net, RB-RC U-Net and DR-DW U-Net are trained and compared on different datasets. It is important to note that the networks mentioned above were carefully scaled so that the number of trainable parameters of all networks in our experiment can reach to one million, and the amount difference between each network is no more than 0.05 million.

B. Datasets

The datasets used in our experiment are the DRIVE dataset [15] (Digital Retinal Images for Vessel Extraction) for retina vessel segmentation and Kaggle 2017 dataset of Finding and Measuring Lungs in CT Data for lung segmentation. The DRIVE contains 40 colored fundus photographs, and the plane resolution of DRIVE is 565×584. The task here is to correctly segment the vessels in retinal images. The Kaggle 2017 lung dataset used in our experiment is a collection of 2D images with manually segmented lungs.

C. Experimental configurations

Our experiments are based on the public Keras 2.2.4 platform. The python version is 3.6 and the CUDA version is 10.0. The training and testing were carried on Ubuntu 16.04 system with one NVIDIA GeForce GTX 1080Ti GPU.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

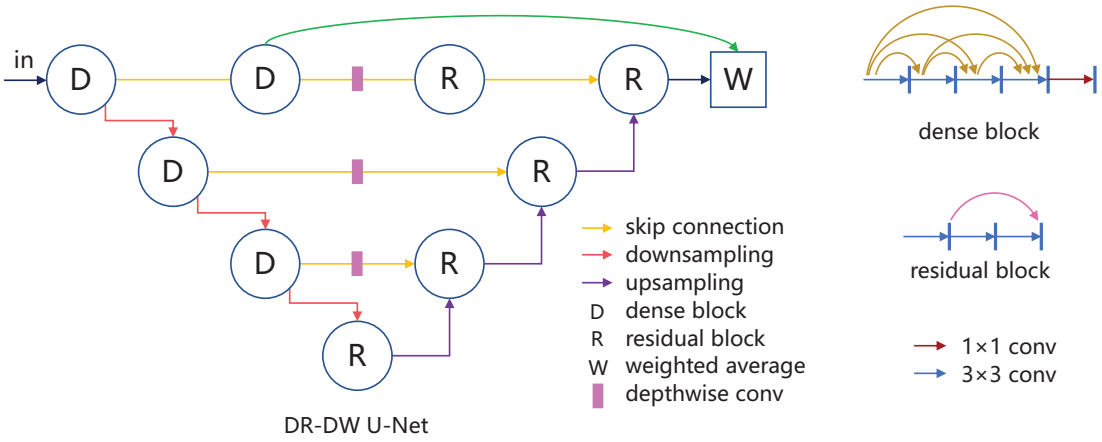


Fig. 3. The architecture of DR-DW U-Net. The capital letter D represents a dense block and R represents a residual block.

1) Image preprocessing: For the DRIVE datasets, before training, both the training and testing images are pre-processed with the following transformations: RGB to gray-scale conversion, standardization and contrast-limited adaptive histogram equalization (CLAHE). The training data are augmented by random horizontal flip and rotation, as well as random gamma adjustment and random Gaussian blur (excluding the ground truth masks). For the Kaggle 2017 lung datasets, random width shift, height shift, random rotation and horizontal flip are applied to the training images and masks.

2) Training: For the DRIVE dataset, the training of the neural network is performed on sub-images (patches) of the pre-processed full images. Each patch, of dimension 64 × 64, is obtained by randomly selecting a small part of the full image to feed into networks. During the training phase, each training image yields 1000 small patches to make up the training set. Although the patches may overlap with each other, it does not harm the performance of networks.

For the Kaggle 2017 lung dataset, all the lung images and masks were rescaled to 256 × 256 to reduce the computation cost required for training and inference. The whole image and corresponding ground truth mask were simply passed to the neural network instead of extracting small patches.

In order to alleviate the problem of class-imbalance, the loss function in our experiments is the summation of the cross-entropy loss and the dice coefficient loss. The cross-entropy loss for a batch containing n images is defined as follows:

$$CE(Y_b, P_b) = -\frac{1}{n}\sum_{b=1}^{n} Y_b \log P_b + (1 - Y_b)\log(1 - P_b). \quad (3)$$

The dice coefficient is a measure of overlap widely used to assess the performance of segmentation when the ground truth is available, especially when the number of pixels of different categories are highly imbalanced:

$$Dice(Y_b, P_b) = \frac{1}{n}\sum_{b=1}^{n} \frac{2 \cdot Y_b \cdot P_b}{Y_b + P_b}. \quad (4)$$

Thus, the overall loss function is described as:

$$Loss(Y_b, P_b) = 1 - Dice(Y_b, P_b) + CE(Y_b, P_b). \quad (5)$$

In the equations mentioned above, Y_b and P_b denote the predicted probabilities and the ground truth labels of the pixels in the given image, respectively. The adam optimization [16] is employed for training instead of stochastic gradient descent (SGD).

3) Inference: The test set of DRIVE dataset contains 20 images and corresponding masks to allow researchers to test their network. In order to improve the performance, the prediction of each pixel is obtained by averaging multiple predictions. The results on the test set are obtained by averaging the predicted probability over multiple overlapping patches covering the same pixel. The Kaggle 2017 lung dataset contains 267 2D samples with respective label ground truth images. Notice that since the official website didn't divide the dataset into a training set and testing set, 90% of the images are used for training and the rest for testing in our experiments.

4) Evaluation metrics: For the DRIVE dataset, the evaluation metrics here include the global accuracy, sensitivity, the area under receiver operation characteristic curve (AUC of the sensitivity-specificity curve) and specificity. For the Kaggle 2017 lung dataset, only the global accuracy and dice coefficient are considered.

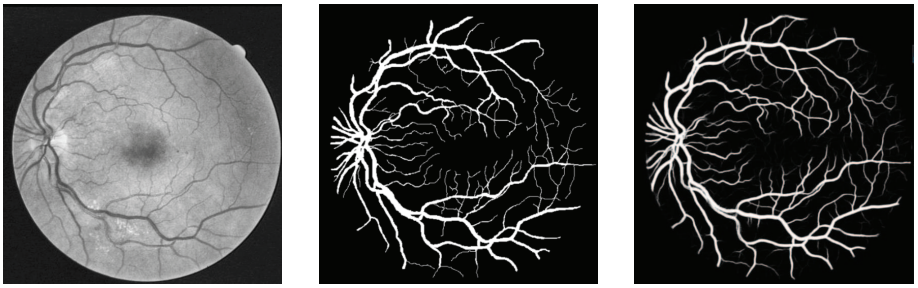
V. Results and Analysis

In this section, the results of U-Net, RB U-Net, DB U-Net, RB-RC U-Net and DR-DW U-Net are reported. It is noticed that all of the evaluation scores reported in our experiments are the average scores of multiple repeated experiments.

A. Retinal vessel segmentation

Testing is performed on the DRIVE testing dataset, using the gold standard binary mask as the ground truth. And the evaluation metrics are calculated on the test set. With a stride of 5 pixels in both height and width, multiple consecutive overlapping patches are extracted in each testing image. Then, for each pixel, the vessel probability is obtained by averaging

978-1-7281-6926-2/20/$31.00 ©2020 IEEE



(a) (b) (c)

Fig. 4. A test sample from the DRIVE dataset in our experiments. (a) The input image after pre-processing. (b) The corresponding ground truth. (c) The prediction obtained by DR-DW U-Net in our experiments.

(a) (b) (c)

Fig. 5. A test sample from the Kaggle 2017 lung dataset in our experiments. (a) The input image. (b) The corresponding ground truth. (c) The prediction obtained by DR-DW U-Net in our experiments.

TABLE I
COMPARISONS OF EVALUATION SCORES ACHIEVED BY DIFFERENT NETWORKS ON THE DRIVE DATASET.

architecture	accuracy	sensitivity	AUC	specificity
U-Net	0.9559	0.7712	0.9790	**0.9835**
RB U-Net	0.9570	0.8190	**0.9813**	0.9780
DB U-Net	0.9568	0.7948	0.9798	0.9800
RB-RC U-Net	0.9567	0.8244	0.9807	0.9775
DR-DW U-Net	**0.9573**	0.8185	0.9811	0.9788
CE-Net [17]	0.9545	**0.8309**	0.9779	-
MS-NFN [18]	0.9567	0.7844	0.9807	0.9819

probabilities over all the predicted patches covering that pixel. A test sample from the DRIVE dataset is shown in Fig. 4.

The evaluation metrics and corresponding scores on the DRIVE dataset are shown in Table I. The results of CE-Net [17] and MS-NFN [18] are also included here for comparison. Each modification of DR-DW U-Net on the classical U-Net helps improve the performance. It is observed that the accuracy of each network is almost the same. And surprisingly, DR-DW U-Net achieves the better performance than CE-Net and MS-NFN. Considering all of the evaluation metrics, it is fair to say that the RB U-Net and DR-DW U-Net achieve the relatively best performances in retinal vessel segmentation. But it is notice that the superiority of RB U-Net and DR-DW U-Net over the other architectures are not very significant. It seems that the residual and dense blocks indeed help to capture more information.

B. Lung segmentation

For the Kaggle 2017 lung dataset, 10% of the images are used for testing. The evaluation metrics and scores on the Kaggle 2017 lung dataset in our experiments are shown in Table II. A test sample from the Kaggle 2017 lung dataset is shown in Fig. 5. The results reported from CE-Net are also included to validate the effectiveness of DR-DW U-Net. The letter H denotes that the corresponding architecture is very hard to train on the dataset.

Although the total number of trainable parameters of the RB-RC U-Net is just a little larger than that of RB U-Net, it is very hard to converge to an approximate optimal solution. Several typical sets of hyperparameters were selected. But it seemed that the RB-RC U-Net fell into some saddle points. The reason is probably that the complexity level of the Kaggle 2017 lung dataset is quite low, where more complex networks may suffer from vanishing gradient, overfitting, model degration or saddle points. On the other hand, the results are

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

TABLE II
COMPARISONS OF EVALUATION SCORES ACHIEVED BY DIFFERENT NETWORKS ON THE KAGGLE 2017 LUNG DATASET.

architecture	accuracy	dice score
U-Net	0.9750	0.9528
RB U-Net	0.9917	0.9804
DB U-Net	0.9879	0.9697
RB-RC U-Net	H	H
DR-DW U-Net	**0.9926**	**0.9813**
CE-Net [17]	0.9900	-

reasonable, for the reason that the residual architecture eases the training of the network and enable the network to go deeper, while the dense architecture fuses the information from the anterior feature maps to the subsequent feature maps, if the dataset is large enough.

VI. CONCLUSION

The classic U-Net architecture have been proved to achieve good performances for different biomedical segmentation applications. With the help of different strategies of data augmentation, it only needs relatively fewer annotated images. However, since different datasets can vary in image resolution, dataset size and organ structure, the classic U-Net architecture is not perfect. In this paper, the downsampling, upsampling and activation function of U-Net are modified, and a DR-DW U-Net architecture is proposed. DR-DW U-Net adopts dense blocks in the encoder and uses residual blocks in the decoder. The skip connections in DR-DW U-Net are supplemented with depth-wise convolutions. Experimental results indicate that the DR-DW U-Net is competitive for Retinal vessel segmentation and lung segmentation.

Our future work will focus on the improvements of DR-DW U-Net by cascade networks, multiple inputs and outputs, deep supervision, super resolution, irregular shaped convolution kernels, and hyper parameters search.

REFERENCES

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.

[2] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.

[3] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2961–2969.

[4] H. Fu, J. Cheng, Y. Xu, D. W. K. Wong, J. Liu, and X. Cao, "Joint optic disc and cup segmentation based on multi-label deep network and polar transformation," *IEEE Transactions on Medical Imaging*, vol. 37, no. 7, pp. 1597–1605, 2018.

[5] E. Gibson, F. Giganti, Y. Hu, E. Bonmati, S. Bandula, K. Gurusamy, B. Davidson, S. P. Pereira, M. J. Clarkson, and D. C. Barratt, "Automatic multi-organ segmentation on abdominal CT with dense v-networks," *IEEE Transactions on Medical Imaging*, vol. 37, no. 8, pp. 1822–1834, 2018.

[6] Y. Lee, T. Hara, H. Fujita, S. Itoh, and T. Ishigaki, "Automated detection of pulmonary nodules in helical CT images based on an improved template-matching technique," *IEEE Transactions on Medical Imaging*, vol. 20, no. 7, pp. 595–604, 2001.

[7] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Proceedings of European Conference on Computer Vision*. Springer, 2014, pp. 740–755.

[8] S. Apostolopoulos, S. De Zanet, C. Ciller, S. Wolf, and R. Sznitman, "Pathological OCT retinal layer segmentation using branch residual U-shape networks," in *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 294–301.

[9] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. 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*. Springer, 2018, pp. 3–11.

[10] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2017, pp. 4700–4708.

[11] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: The all convolutional net," *arXiv preprint arXiv:1412.6806*, 2014.

[12] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2017, pp. 1251–1258.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2016, pp. 770–778.

[14] ——, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proceedings of the IEEE International Conference on Computer Vision*. IEEE, 2015, pp. 1026–1034.

[15] J. Staal, M. D. Abràmoff, M. Niemeijer, M. A. Viergever, and B. Van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Transactions on Medical Imaging*, vol. 23, no. 4, pp. 501–509, 2004.

[16] N. S. Keskar and R. Socher, "Improving generalization performance by switching from Adam to SGD," *arXiv preprint arXiv:1712.07628*, 2017.

[17] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "CE-Net: context encoder network for 2D medical image segmentation," *IEEE Transactions on Medical Imaging*, vol. 38, no. 10, pp. 2281–2292, 2019.

[18] Y. Wu, Y. Xia, Y. Song, Y. Zhang, and W. Cai, "Multiscale network followed network model for retinal vessel segmentation," in *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 119–126.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE