CE-Net: Context Encoder Network for 2D Medical Image Segmentation

Zaiwang Gu, Jun Cheng, Huazhu Fu, Kang Zhou, Huaying Hao, Yitian Zhao, Tianyang Zhang, Shenghua Gao, and Jiang Liu

***Abstract*— Medical image segmentation is an important step in medical image analysis. With the rapid development of a convolutional neural network in image processing, deep learning has been used for medical image segmentation, such as optic disc segmentation, blood vessel detection, lung segmentation, cell segmentation, and so on. Previously, U-net based approaches have been proposed. However, the consecutive pooling and strided convolutional operations led to the loss of some spatial information. In this paper, we propose a context encoder network (CE-Net) to capture more high-level information and preserve spatial information for 2D medical image segmentation. CE-Net mainly contains three major components: a feature encoder module, a context extractor, and a feature decoder module. We use the pretrained ResNet block as the fixed feature extractor. The context extractor module is formed by a newly proposed dense atrous convolution block and a residual multi-kernel pooling block. We applied the proposed CE-Net to different 2D medical image segmentation tasks. Comprehensive results show that the proposed method outperforms the original U-Net method and other state-of-the-art methods for optic disc segmentation, vessel detection, lung segmentation, cell contour segmentation, and retinal optical coherence tomography layer segmentation.**

***Index Terms*— Medical image segmentation, deep learning, context encoder network.**

I. Introduction

Medical image segmentation is often an important step in medical image analysis, such as optic disc segmentation [1]–[3] and blood vessel detection [4]–[8] in retinal images, cell segmentation [9]–[11] in electron microscopic (EM) recordings, lung segmentation [12]–[16] and brain segmentation [17]–[22] in computed tomography (CT) and magnetic resonance imaging (MRI). Previous approaches to medical image segmentation are often based on edge detection and template matching [15]. For example, circular or elliptical Hough transform are used in optic disc segmentation [3], [23]. Template matching is also used for spleen segmentation in MRI sequence images [24] and ventricular segmentation in brain CT images [22].

Deformable models are also proposed for medical image segmentation. The shape-based method using level sets [25] has been proposed for two-dimensional segmentation of cardiac MRI images and three-dimensional segmentation of prostate MRI images. In addition, a level set-based deformable model is adopted for kidney segmentation from abdominal CT images [26]. The deformable model has also been integrated with the Gibbs prior models for segmenting the boundaries of organs [27], with an evolutionary algorithm and a statistical shape model to segment the liver [16] from CT volumes. In optic disc segmentation, different deformable models have also been proposed and adopted, such as mathematical morphology, global elliptical model, local deformable model [28], and modified active shape model [29].

Learning based approaches are proposed to segment medical images as well. Aganj *et al.* [30] proposed the local center of mass based method for unsupervised learning based image segmentation in X-ray and MRI images. Kanimozhi and Bindu [31] applied the stationary wavelet transform to obtain the feature vectors, and self-organizing map is adopted to handle these feature vectors for unsupervised MRI image segmentation. Tong *et al.* [32] combined dictionary learning and sparse coding to segment multi-organ in abdominal CT images. Pixel classification based approaches [1], [33] are

Manuscript received December 24, 2018; revised February 24, 2019; accepted March 4, 2019. Date of publication March 7, 2019; date of current version October 1, 2019. This work was supported in part by the Cixi Institute of Biomedical Engineering, Chinese Academy of Sciences, under Grant Y60001RA01 and Grant Y80002RA01, in part by the Ningbo 3315 Innovation Team Grant, Zhejiang Province, under Grant Y61102DL03, in part by the National Natural Science Foundation of China under Grant 61601029, and in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LZ19F010001. *(Corresponding author: Jun Cheng.)*

Z. Gu is with the School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200072, China, and also with the Cixi Institute of Biomedical Engineering, Chinese Academy of Sciences, Zhejiang 315201, China (e-mail: guzaiwang@nimte.ac.cn).

J. Cheng, H. Hao, Y. Zhao, and T. Zhang are with the Cixi Institute of Biomedical Engineering, Chinese Academy of Sciences, Zhejiang 315201, China (e-mail: chengjun@nimte.ac.cn; haohuaying@nimte.ac.cn; yitian.zhao@nimte.ac.cn; zhangtianyang@nimte.ac.cn).

H. Fu is with the Inception Institute of Artificial Intelligence, Abu Dhabi, United Arab Emirates, and also with the Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore 138632 (e-mail: huazhufu@gmail.com).

K. Zhou and S. Gao are with the School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China (e-mail: zhoukang@shanghaitech.edu.cn; gaoshh@shanghaitech.edu.cn).

J. Liu is with the Department of Computer Science and Engineering, Southern University of Science and Technology, Guangdong 518055, China, and also with the Cixi Institute of Biomedical Engineering, Chinese Academy of Sciences, Zhejiang 315201, China (e-mail: liuj@sustech.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2019.2903562

also learning based approaches which train classifiers based on pixels using pre-annotated data. However, it is not easy to select the pixels and extract features to train the classifier from the larger number of pixels. Cheng *et al.* [1] used the superpixel strategy to reduce the number of pixels and performed the optic disc and cup segmentation using superpixel classification. Tian *et al.* [34] adopted a superpixel-based graph cut method to segment 3D prostate MRI images. In [35], superpixel learning based method is integrated with restricted regions of shape constrains to segment lung from CT images.

The drawbacks of these methods lie in the utilization of hand-crafted features to obtain the segmentation results. On the one hand, it is difficult to design the representative features for different applications. On the other hand, the designed features working well for one type of images often fail on another type. Therefore, there is a lack of general approach to extract the feature.

With the development of convolutional neural network (CNN) in image and video processing [36] and medical image analysis [37], [38], automatic feature learning algorithms using deep learning have emerged as feasible approaches for medical image segmentation. Deep learning based segmentation methods are pixel-classification based learning approaches. Different from traditional pixel or superpixel classification approaches which often use hand-crafted features, deep learning approaches learn the features and overcome the limitation of hand-crafted features.

Earlier deep learning approaches for medical image segmentation are mostly based on image patches. Ciresan *et al.* [39] proposed to segment neuronal membranes in microscopy images based on patches and sliding window strategy. Then, Kamnitsas *et al.* [40] employed a multi-scale 3D CNN architecture with fully connected conditional random field (CRF) for boosting patch based brain lesion segmentation. Obviously, this solution introduces two main drawbacks: redundant computation caused from sliding window and the inability to learn global features.

With the emerging of the end-to-end fully convolutional network (FCN) [41], Ronneberger *et al.* [10] proposed U-shape Net (U-Net) framework for biomedical image segmentation. U-Net has shown promising results on the neuronal structures segmentation in electron microscopic recordings and cell segmentation in light microscopic images. It has becomes a popular neural network architecture for biomedical image segmentation tasks [42]–[45]. Sevastopolsky [43] applied U-Net to directly segment the optic disc and optic cup in retinal fundus images for glaucoma diagnosis. Roy *et al.* [44] used a similar network for retinal layer segmentation in optical coherence tomography (OCT) images. Norman *et al.* [42] used U-Net to segment cartilage and meniscus from knee MRI data. The U-Net is also applied to directly segment lung from CT images [45].

Many variations have been made on U-Net for different medical image segmentation tasks. Fu *et al.* [4] adopted the CRF to gather the multi-stage feature maps for boosting the vessel detection performance. Later, a modified U-Net framework (called M-Net) [2] is proposed for joint optic disc and cup segmentation by adding multi-scale inputs and deep supervision into the U-net architecture. Deep supervision mainly introduces the extra loss function associated with the middle-stage features. Based on the deep supervision, Chen *et al.* [46] proposed a Voxresnet to segment volumetric brain, and Dou *et al.* [47] proposed 3D deeply supervised network (3D DSN) to automatically segment lung in CT volumes.

To enhance the feature learning ability of U-Net, some new modules have been proposed to replace the original blocks. Apostolopoulos *et al.* [48] proposed a branch residual U-network (BRU-net) to segment pathological OCT retinal layer for age-related macular degeneration diagnosis. BRU-net relies on residual connection and dilated convolutions to enhance the final OCT retinal layer segmentation. Gibson *et al.* [49] introduced dense connection in each encoder block to automatically segment multiple organs on abdominal CT. Kumar *et al.* [21] proposed an InfiNet for infant brain MRI segmentation. Besides the above achievements for U-Net based medical image segmentation, some researchers have also made progress to modify U-Net for general image segmentation. Peng *et al.* [50] proposed a novel global convolutional network to improve semantic segmentation. Lin *et al.* [51] proposed a multi-path refinement network, which contains residual convolution unit, multi-resolution fusion and chained residual pooling. Zhao *et al.* [52] adopted spatial pyramid pooling to gather the extracted feature maps to improve the semantic segmentation performance.

A common limitation of the U-Net and its variations is that the consecutive pooling operations or convolution striding reduce the feature resolution to learn increasingly abstract feature representations. Although this invariance is beneficial for classification or object detection tasks, it often impedes dense prediction tasks which require detailed spatial information. Intuitively, maintaining high-resolution feature maps at the middle stages can boost segmentation performance. However, it increases the size of feature maps, which is not optimal to accelerate the training and ease the difficulty of optimization. Therefore, there is a trade-off between accelerating the training and maintaining the high resolution. Generally, the U-Net structures can be considered as Encoder-Decoder architecture. The Encoder aims to reduce the spatial dimension of feature maps gradually and capture more high-level semantic features. The Decoder aims to recover the object details and spatial dimension. Therefore, it is spontaneous to capture more high-level features in the encoder and preserve more spatial information in the decoder to improve the performance of image segmentation.

Motivated by the above discussions and also the Inception-ResNet structures [53], [54] which make the neural network wider and deeper, we propose a novel dense atrous convolution (DAC) block to employ atrous convolution. The original U-Net architecture captures multi-scale features in the limited scaling range by adopting the consecutive 3×3 convolution and pooling operations in the encoding path. Our proposed DAC block could capture wider and deeper semantic features by infusing four cascade branches with multi-scale atrous convolutions. In this module, the residual connection is utilized to prevent the gradient vanishing. In addition, we also propose a residual multi-kernel pooling (RMP) motivated

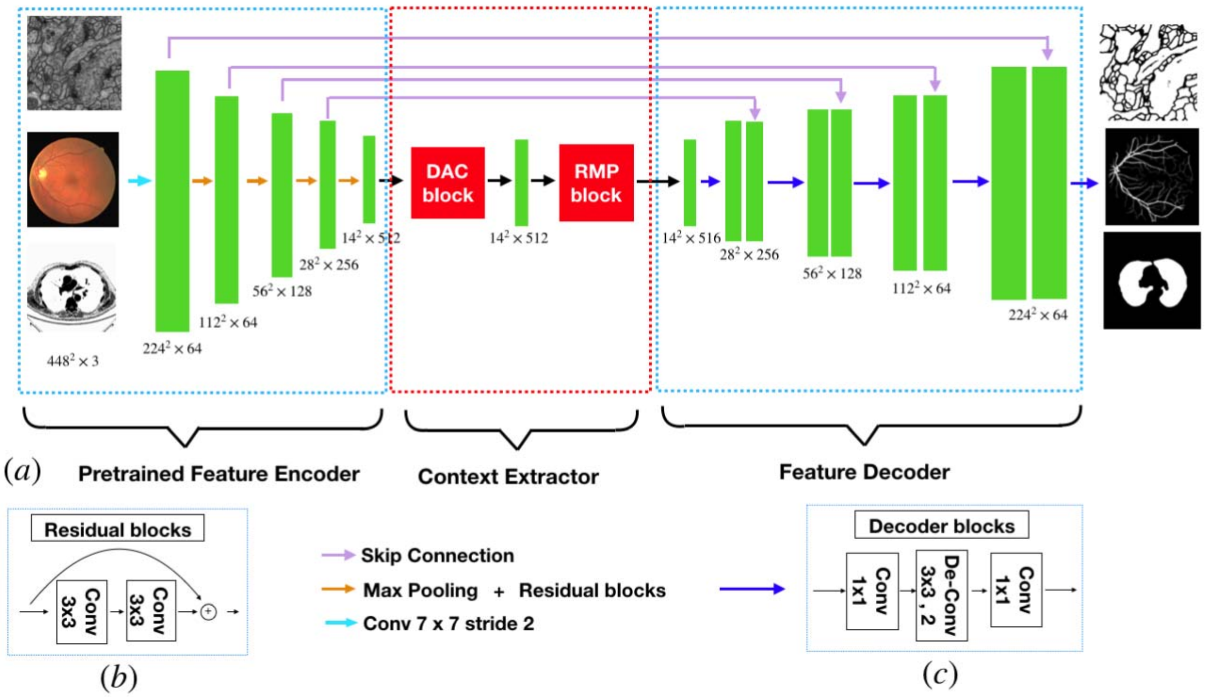

Fig. 1. Illustration of the proposed CE-Net. Firstly, the images are fed into a feature encoder module, where the ResNet-34 block pretrained from ImageNet is used to replace the original U-Net encoder block. The context extractor is proposed to generate more high-level semantic feature maps. It contains a dense atrous convolution (DAC) block and a residual multi-kernel pooling (RMP) block. Finally, the extracted features are fed into the feature decoder module. In this paper, we adopt a decoder block to enlarge the feature size, replacing the original up-sampling operation. The decoder block contains 1×1 convolution and 3×3 deconvolution operations. Based on skip connection and the decoder block, we obtain the mask as the segmentation prediction map.

from spatial pyramid pooling [55]. The RMP block further encodes the multi-scale context features of the object extracted from the DAC module by employing various size pooling operations, without the extra learning weights. In summary, the DAC block is proposed to extract enriched feature representations with multi-scale atrous convolutions, followed by the RMP block for further context information with multi-scale pooling operations. Integrating the newly proposed DAC block and the RMP block with the backbone encoder-decoder structure, we propose a novel context encoder network named as CE-Net. It relies on the DAC block and the RMP block to get more abstract features and preserve more spatial information to boost the performance of medical image segmentation.

The main contributions of this work are summarized as follows:

1) We propose a DAC block and RMP block to capture more high-level features and preserve more spatial information.
2) We integrate the proposed DAC block and RMP block with encoder-decoder structure for medical image segmentation.
3) We apply the proposed method in different tasks including optic disc segmentation, retinal vessel detection, lung segmentation, cell contour segmentation and retinal OCT layer segmentation. Results show that the proposed method outperforms the state-of-the-art methods in these different tasks.

The remainder of this paper is organized as follows. Section II introduces the proposed method in details. Section III presents the experimental results and discussions. In Section IV, we draw some conclusions.

II. Method

The proposed CE-Net consists of three major parts: the feature encoder module, the context extractor module, and the feature decoder module, as shown in Fig. 1.

A. Feature Encoder Module

In U-Net architecture, each block of encoder contains two convolution layers and one max pooling layer. In the proposed method, we replace it with the pretrained ResNet-34 [53] in the feature encoder module, which retains the first four feature extracting blocks without the average pooling layer and the fully connected layers. Compared with the original block, ResNet adds shortcut mechanism to avoid the gradient vanishing and accelerate the network convergence, as shown in Fig. 1(b). For convenience, we use the modified U-net with pretrained ResNet as backbone approach.

B. Context Extractor Module

The context extractor module is a newly proposed module, consisting of the DAC block and the RMP block. This module extracts context semantic information and generates more high-level feature maps.

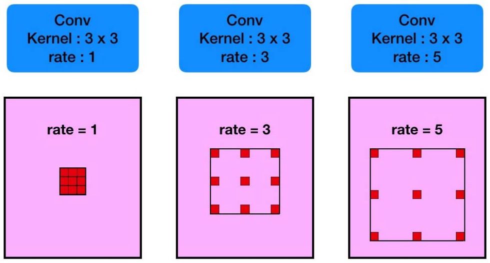

Fig. 2. The illustrations of atrous convolution.

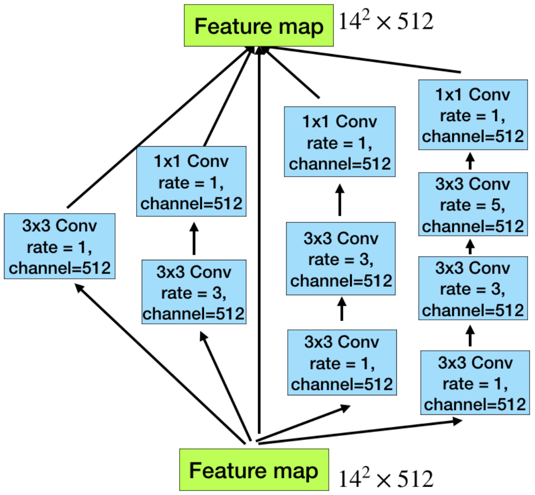
Feature map
$14^2 \times 512$
3x3 Conv rate = 1, channel=512
1x1 Conv rate = 1, channel=512
3x3 Conv rate = 3, channel=512
1x1 Conv rate = 1, channel=512
3x3 Conv rate = 3, channel=512
3x3 Conv rate = 1, channel=512
1x1 Conv rate = 1, channel=512
3x3 Conv rate = 5, channel=512
3x3 Conv rate = 3, channel=512
3x3 Conv rate = 1, channel=512
Feature map
$14^2 \times 512$

Fig. 3. The illustrations of dense atrous convolution block. It contains four cascade branches with the gradual increment of the number of atrous convolution, from 1 to 1, 3, and 5, then the receptive field of each branch will be 3, 7, 9, 19. Therefore, the network can extract features from different scales.

1) Atrous Convolution: In semantic segmentation tasks and object detection tasks, deep convolutional layers have shown to be effective in extracting feature representations for images. However, the pooling layers lead to the loss of semantic information in images. In order to overcome this limitation, atrous convolution is adopted for dense segmentation [56]:

The atrous convolution is originally proposed for the efficient computation of the wavelet transform. Mathematically, the atrous convolution under two-dimensional signals is computed as follows:

$$\boldsymbol{y}[\boldsymbol{i}] = \sum_{\boldsymbol{k}} \boldsymbol{x}[\boldsymbol{i} + r\boldsymbol{k}]\boldsymbol{w}[\boldsymbol{k}], \tag{1}$$

where the convolution of the input feature map $\boldsymbol{x}$ and a filter $\boldsymbol{w}$ yields the output $\boldsymbol{y}$, and the atrous rate r corresponds to the stride with which we sample the input signal. It is equivalent to convolute the input $\boldsymbol{x}$ with upsampled filters produced by inserting $r - 1$ zeros between two consecutive filter values along each spatial dimension (hence the name atrous convolution in which the French word atrous means holes in English). Standard convolution is a special case for rate $r = 1$, and atrous convolution allows us to adaptively modify filter's field-of-view by changing the rate value. See Fig. 2 for illustration.

2) Dense Atrous Convolution Module: Inception [54] and ResNet [53] are two classical and representative architectures in the deep learning. Inception-series structures adopt different receptive fields to widen the architecture. On the contrary, ResNet employs shortcut connection mechanism to avoid the exploding and vanishing gradients. It makes the neural network break through up to thousands of layers for the first time. Inception-ResNet [54] block, which combines the Inception and ResNet, inherits the advantages of both approaches. Then it becomes a baseline approach in the field of deep CNNs.

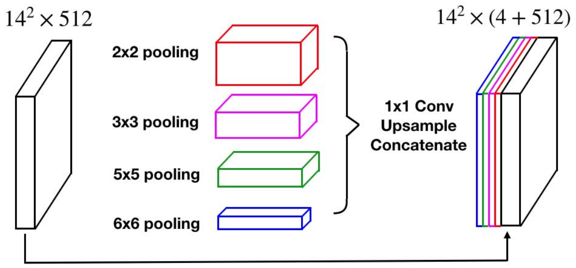

Fig. 4. The illustrations of residual multi-kernel pooling (RMP) strategy. The proposed RMP gather context information with four different-size pooling kernels. Then features are fed into 1×1 convolution to reduce the dimension of feature maps. Finally, the upsampled features are concatenated with original features.

Motivated by the Inception-ResNet-V2 block and atrous convolution, we propose dense atrous convolution (DAC) block to encode the high-level semantic feature maps. As shown in Fig. 3, the atrous convolution is stacked in cascade mode. In this case, DAC has four cascade branches with the gradual increment of the number of atrous convolution, from 1 to 1, 3, and 5, then the receptive field of each branch will be 3, 7, 9, 19. It employs different receptive fields, similar to Inception structures. In each atrous branch, we apply one 1×1 convolution for rectified linear activation. Finally, we directly add the original features with other features, like shortcut mechanism in ResNet. Since the proposed block looks like a densely connected block, we name it dense atrous convolution block. Very often, the convolution of large reception field could extract and generate more abstract features for large objects, while the convolution of small reception field is better for small object. By combining the atrous convolution of different atrous rates, the DAC block is able to extract features for objects with various sizes.

3) Residual Multi-Kernel Pooling: A challenge in segmentation is the large variation of object size in medical image. For example, a tumor in middle or late stage can be much larger than that in early stage. In this paper, we propose a residual multi-kernel pooling to address the problem, which mainly relies on multiple effective field-of-views to detect objects at different sizes.

The size of receptive field roughly determines how much context information we can use. The general max pooling operation just employs a single pooling kernel, such as 2×2. As illustrated in Fig. 4, the proposed RMP encodes global context information with four different-size receptive fields: 2×2, 3×3, 5×5 and 6×6. The four-level outputs contain the feature maps with various sizes. To reduce the dimension of weights and computational cost, we use a 1×1 convolution after each level of pooling. It reduces the dimension of the feature maps to the $\frac{1}{N}$ of original dimension, where N represents number of channels in original feature maps. Then we upsample the low-dimension feature map to get the same size features as the original feature map via bilinear interpolation. Finally, we concatenate the original features with upsampled feature maps.

C. Feature Decoder Module

The feature decoder module is adopted to restore the high-level semantic features extracted from the feature encoder module and context extractor module. The skip connection takes some detailed information from the encoder to the decoder to remedy the information loss due to consecutive pooling and striding convolutional operations. Similar to [48], we adopted an efficient block to enhance the decoding performance. The simple upscaling and deconvolution are two common operations of the decoder in the U-shape Networks. The upscaling operation increases the image size with linear interpolation, while deconvolution (also called transposed convolution) employs convolution operation to enlarge the image. Intuitively, the transposed convolution could learn a self-adaptive mapping to restore feature with more detailed information. Therefore, we choose to use the transposed convolution to restore the higher resolution feature in the decoder. As illustrated in Fig. 1(c), it mainly includes a 1×1 convolution, a 3×3 transposed convolution and a 1×1 convolution consecutively. Based on skip connection and the decoder block, the feature decoder module outputs a mask, the same size as the original input.

D. Loss Function

Our framework is an end-to-end deep learning system. As illustrated in Fig. 1, we need to train the proposed method to predict each pixel to be foreground or background, which is a pixel-wise classification problem. The most common loss function is cross entropy loss function.

However, the objects in medical images such as optic disc and retinal vessels often occupy a small region in the image. The cross entropy loss is not optimal for such tasks. In this paper, we use the Dice coefficient loss function [57], [58] to replace the common cross entropy loss. The comparison experiments and discussions are also conducted in the following section. The Dice coefficient is a measure of overlap widely used to assess segmentation performance when ground truth is available, as in Equation (2):

$$L_{dice} = 1 - \sum_{k}^{K} \frac{2\omega_k \sum_i^N p_{(k,i)} g_{(k,i)}}{\sum_i^N p_{(k,i)}^2 + \sum_i^N g_{(k,i)}^2} \tag{2}$$

where N is the pixel number, $p_{(k,i)} \in [0, 1]$ and $g_{(k,i)} \in \{0, 1\}$ denote predicted probability and ground truth label for class k, respectively. K is the class number, and $\sum_k \omega_k = 1$ are the class weights. In our paper, we set $\omega_k = \frac{1}{K}$ empirically.

The final loss function is defined as:

$$L_{loss} = L_{dice} + L_{reg} \tag{3}$$

where L_{reg} represents the regularization loss (also called to weight decay) [59] used to avoid overfitting.

To evaluate the performance of CE-Net, we apply the proposed method to five different medical image segmentation tasks: optic disc segmentation, retinal vessel detection, lung segmentation, cell contour segmentation and retinal OCT layer segmentation.

III. Experiment

A. Experimental Setup

In this section, we first introduce the image preprocessing and data augmentation strategies used in training and testing phases.

1) Training Phase: Because of the limited number of training images, the datasets are augmented to reduce the risk of overfitting [36]. Firstly, we do data augmentation in an ambitious way, including horizontal flip, vertical flip and diagonal flip. In this way, each image in the original dataset is augmented to 2×2×2 = 8 images. Next, the solutions of image preprocessing mainly include scaling from 90% to 110%, color jittering in HSV color space and image shifting randomly. The random image preprocessing method can enhance the data augmentation capability.

2) Testing Phase: To improve the robustness of medical image segmentation method, we also adopt test augmentation strategy, as that in [60] and [61], including image horizontal flip, vertical flip and diagonal flip (equal to predicting each image 8 times). Then we average the 8 predictions to get the final prediction map. All baseline approaches utilize the same strategy during testing phase.

3) Experiment Settings: Our proposed network is based on the ResNet pretrained on ImageNet. The implementation is based on the public PyTorch platform. The training and testing bed is Ubuntu 16.04 system with the NVidia GeForce Titan graphics cards, which has 12 Gigabyte memory.

During the training, we adopt mini-batch stochastic gradient descent (SGD) with batch size 8, momentum 0.9 and weight decay 0.0001, other than Adam optimization. We use SGD optimization since recent studies [62], [63] show that SGD often achieves a better performance, though the Adam optimization convergences faster. In addition, we use the "poly" learning rate policy where the learning rate is multiplied by $(1 - \frac{iter}{max\_iter})^{power}$ with power 0.9 and initial learning rate $4e^{-3}$ [52]. The maximum epoch is 100. We have released our codes on Github. [1]

B. Optic Disc Segmentation

We first test the proposed CE-Net on optic disc segmentation. Three datasets, ORIGA [66], Messidor [67] and RIM-ONE-R1 [68], are used in our experiments. ORIGA dataset contains 650 images with dimension 3072 × 2048. It has been divided into 2 sets: *Set A* for training and *Set B* for testing [69]. In this paper, we follow the same partition of the data set to train and test our models. Messidor dataset is a public dataset provided by the Messidor program partners. It consists of 1200 images with three different sizes: 1440 × 960, 2240 × 1488, 2340 × 1536. The Messidor dataset is originally collected for Diabetic Retinopathy (DR) grading. Later, disc

[1] https://github.com/Guzaiwang/CE-Net

TABLE I

Comparison With Different Methods for OD Segmentation on the ORIGA, Messidor and RIM-ONE-R1 Datasets (Mean ± Standard Deviation)

Method	ORIGA	Messidor	RIM-ONE-R1					
			Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Overall
Superpixel[1]	0.102±0.104	0.125±0.113	0.178	0.229	0.243	0.183	0.181	0.203±0.104
U-Net [10]	0.115±0.068	0.069±0.121	0.137	0.149	0.156	0.171	0.149	0.152±0.107
M-Net[2]	0.071±0.047	0.113±0.089	0.128	0.135	0.153	0.142	0.117	0.135±0.098
Faster RCNN [64]	0.069±0.056	0.079±0.058	0.101	0.152	0.161	0.149	0.104	0.133±0.107
DeepDisc [65]	0.069±0.040	0.064±0.039	0.077	**0.107**	**0.119**	0.101	0.079	0.097±0.045
CE-Net	**0.058±0.032**	**0.051±0.033**	**0.058**	0.112	0.125	**0.080**	**0.059**	**0.087±0.039**

boundary for each image has also been provided from the official website.[2] RIM-ONE dataset consists of three releases. The numbers of image are 169, 455 and 159 respectively. In this paper, we use first released dataset (RIM-ONE-R1), and there are five different expert annotations in RIM-ONE-R1 dataset. We follow the partition in [70] to get the training and testing images in the Messidor and RIM-ONE-R1 datasets. It should be noted that the ORIGA and Messidor datasets provide full image while the RIM-ONE-R1 provides cropped image.

In order to segment the optic disc in the retinal fundus images based on their original resolution, we crop an 800 × 800 area around the brightest point as motivated in [71], except for RIM-ONE-R1 dataset where the region with optic disc has already been cropped and provided.

To evaluate the performance, we adopt the overlapping error, which has been commonly used to evaluate the accuracy of optic disc segmentation:

$$E = 1 - \frac{Area(S \cap G)}{Area(S \cup G)}, \tag{4}$$

where S and G denote the segmented and the manual ground truth optic disc respectively. Beside the average values, we also calculate the corresponding standard deviations.

We compare our methods with state-of-the-art algorithms. Five different algorithms are compared, including superpixel classification method [1], U-Net [10], M-Net method [2], faster RCNN method [72] and DeepDisc method [65]. All of baseline models are adopted from their original implementations.

Table I shows the mean and standard deviation of the overlapping errors of these methods. As we can see, the proposed CE-Net outperforms the state-of-the-art optic disc segmentation methods. In particular, it achieves an overlapping error of 0.058 in the ORIGA dataset, a relative reduction of 15.9% from 0.069 by the latest Faster RCNN or DeepDisc methods. In Messidor dataset, CE-Net achieves an overlapping error of 0.051, which is a relative reduction of 20.3% from 0.064 by DeepDisc. The RIM-ONE-R1 dataset has five independent annotations. In our experiments, we follow the same setting in [70] to use cross validation to get the results. Although it performs slightly worse than DeepDisc in comparison with the annotation by Expert 2 and Expert 3, the overall results still show that CE-Net outperforms DeepDisc and other methods.

We also show four sample results in Fig. 5 to visually compare our method with some competitive methods, including superpixel based method, M-Net and DeepDisc. The images show that our method obtain more accurate segmentation results.

C. Retinal Vessel Detection

The second application is the retinal vessel detection. We use the public DRIVE [73] dataset which contains 40 images. In DRIVE, two expert manual annotations are provided, the first of which is chosen as the ground truth for performance evaluation in the literature [4]. The 40 images are divided into 20 images for training and 20 images for testing. To compare performance of the vessel detection, we compute two evaluation metrics, the sensitivity (Sen) and the accuracy (Acc), which are also calculated in [4] and [6].

$$Sen = \frac{TP}{TP + FN} \tag{5}$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

where TP, TN, FP and FN represent the number of true positives, true negatives, false positives and false negatives, respectively. In addition, we also introduce the area under receiver operation characteristic curve (AUC) to measure segmentation performance.

We compare the proposed CE-Net with the state-of-the-art algorithms [5], [7], [8]. In addition, some classical deep learning based methods [4], [10], [74] are also included into the comparison. Table II shows the comparison among these methods. From the comparison, the CE-Net achieves 0.8309, 0.9545 and 0.9779 in Sen, Acc and AUC respectively, better than other methods. Comparing with the backbone, the Sen increases from 0.7781 to 0.8309 by 6.8%, the Acc increases from 0.9477 to 0.9545 and the AUC increases from 0.9705 to 0.9779, which shows that the proposed DAC and RMP blocks are beneficial for retina vessel detection as well. We show some examples for visual comparison in Fig. 6.

D. Lung Segmentation

The next application is lung segmentation task, which is to segment lung structure in 2D CT images from the Lung Nodule Analysis (LUNA) competition. The LUNA competition is originally conducted for the following challenge tracks: nodule

[2] http://www.uhu.es/retinopathy/

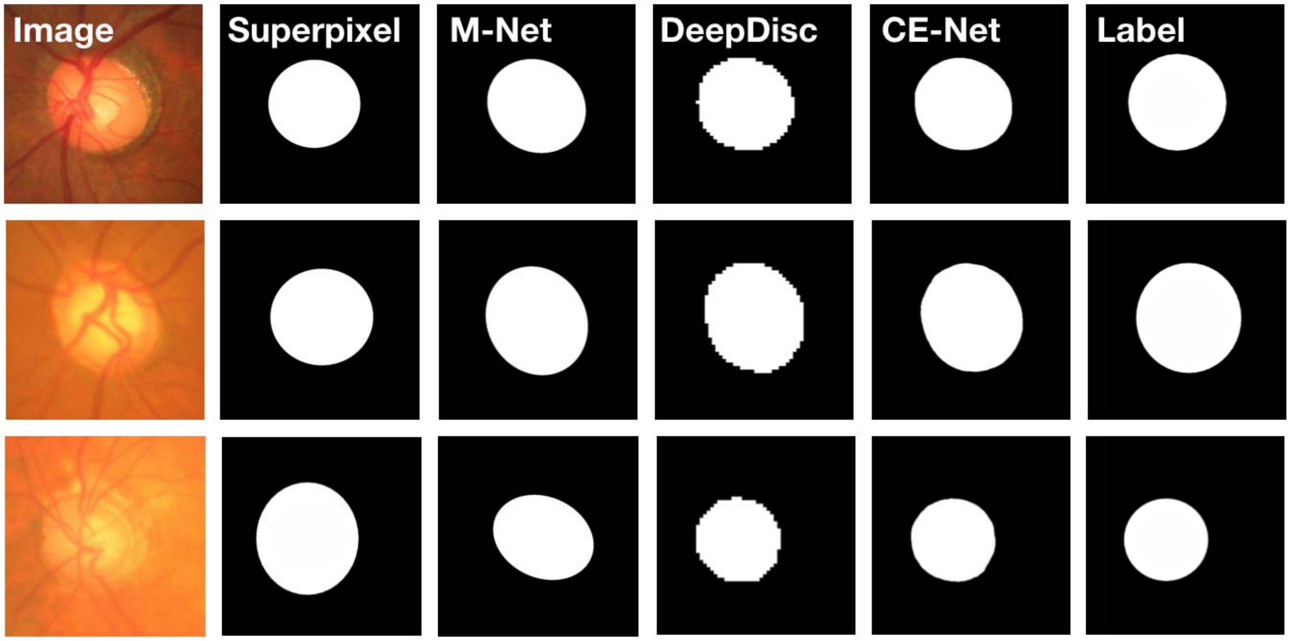

Fig. 5. Sample results. From left to right: original fundus images, state-of-the-art results obtained by superpixel based method [1], M-Net [2], DeepDisc [65], CE-Net and ground-truth masks.

TABLE II
PERFORMANCE COMPARISON OF VESSEL DETECTION

Method	Sen	Acc	AUC
Azzopardi [8]	0.7655	0.9442	0.9614
Roychowdhury [7]	0.7250	0.9520	0.9672
zhao [5]	0.7420	0.9540	0.8620
HED [74]	0.7364	0.9434	0.9723
U-Net [10]	0.7537	0.9531	0.9601
DeepVessel [4]	0.7603	0.9523	0.9752
Backbone	0.7781	0.9477	0.9705
CE-Net	**0.8309**	**0.9545**	**0.9779**

detection and false positive reduction. Because the segmented lungs are fundamental for further lung nodule candidates, we adopt the challenge dataset to evaluate our proposed CE-Net. The dataset contains 534 2D samples (512×512 pixels) with respective label images and can be freely downloaded from the official website. [3] We use 80% of the images for training and the rest for testing, and cross validation is also conducted. The evaluation metrics include the overlapping error, accuracy and sensitivity, similar to those in optic disc segmentation and vessel detection. Beside the average values, we also calculate the corresponding standard deviations in Table III.

From the comparison shown in Table III, the CE-Net achieves 0.038 in overlapping error, 0.8309 in Sensitivity score and 0.9545 in Accuracy score, better than the U-Net. We also compare CE-Net with the backbone, and the overlapping error decreases from 0.044 to 0.038 by 13.6%, the sensitivity score increases from 0.967 to 0.980 while the accuracy increases from 0.988 to 0.990, which further supports that our proposed DAC and RMP blocks are beneficial for lung segmentation. We also give a few examples for visual comparison of lung segmentation in Fig. 6.

TABLE III
PERFORMANCE COMPARISON OF LUNG SEGMENTATION (MEAN ± STANDARD DEVIATION)

Method	E	Acc	Sen
U-Net [10]	0.087±0.090	0.975±0.032	0.938
Backbone	0.044±0.063	0.988±0.024	0.967
CE-Net	**0.038±0.061**	**0.990±0.023**	**0.980**

E. Cell Contour Segmentation

The fourth application is cell contour segmentation. The cell segmentation task is to segment neuronal structures in electron microscopic recordings. The dataset is provided by the EM challenge, which started at ISBI 2012 and is still open for new contributions [75]. The training set contains 30 images (512×512 pixels), and could be downloaded from the official website.[4] The testing set consists of 30 images, and is publicly available as well. However, the corresponding ground truths are kept unknown. The results on the testing set are obtained by sending the prediction maps to the organizers, who will then compute and release the results. From the statement on the official website, the following metrics are the best for the quantitative evaluation of segmentation results: foreground-restricted rand scoring after border thinning (V^{Rand}) and foreground-restricted information theoretic scoring after border thinning (V^{Info}). The V^{Rand} mainly computes the weighted harmonic mean by jointing the Rand split score and Rand merge score, which are used to measure

[3] https://www.kaggle.com/kmader/finding-lungs-in-ct-data/data/

[4] http://brainiac2.mit.edu/

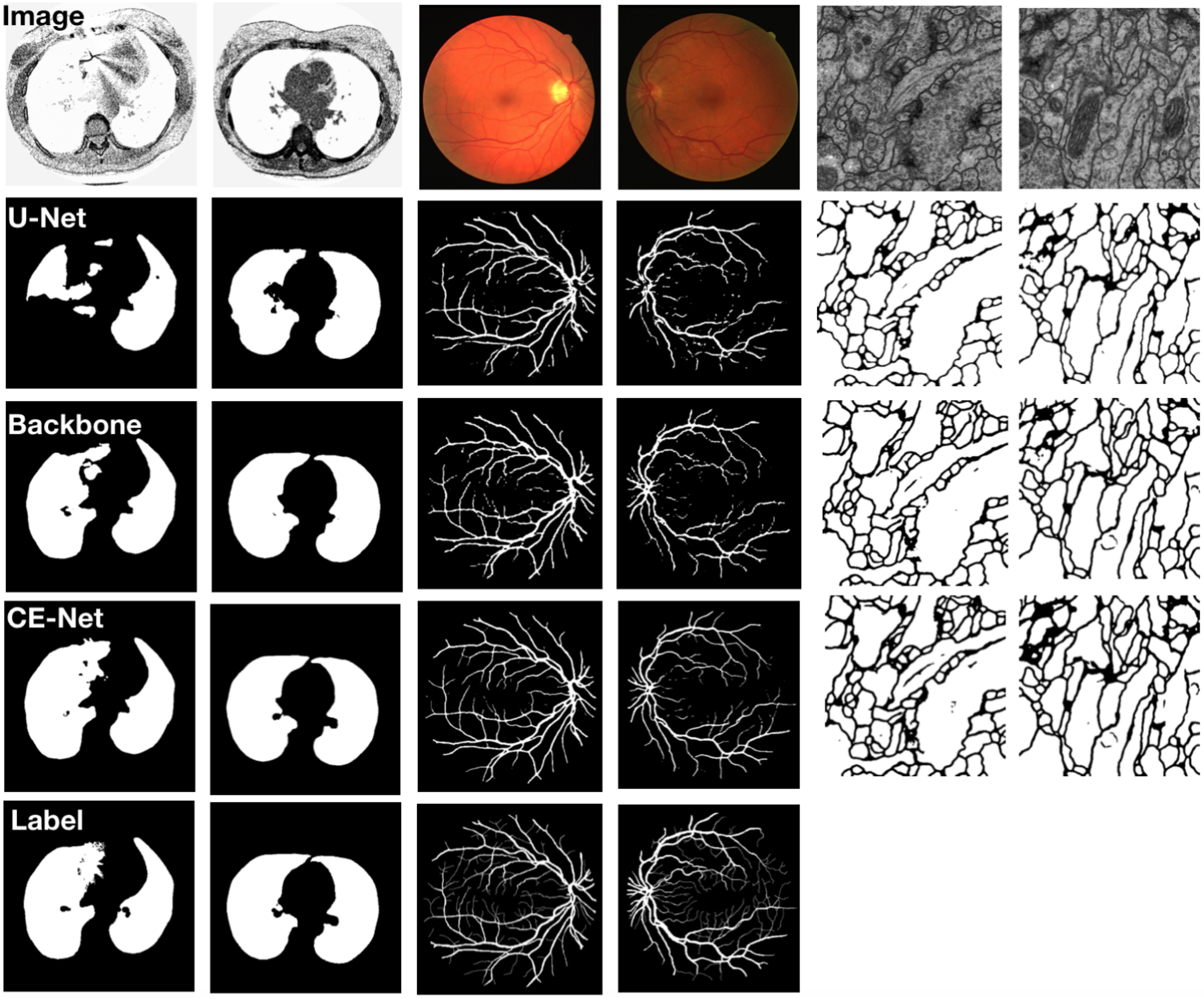

Fig. 6. Sample results of lung segmentation, vessel detection and cell contour segmentation. From top to bottom: original images, U-Net, Backbone, CE-Net and ground truth (the ground truth for cell images is not given).

the segmentation performance. Similarly, the V^{Info} mainly computes weighted harmonic mean of information theoretic score. The higher scores represent the better segmentation performance. The specific computation process and more details of these two algorithms could be found in [76].

We compare our CE-Net with the original U-Net and backbone, and the final results are shown in Table IV. Our CE-Net outperforms the U-Net and Backbone. It indicates that our proposed CE-Net is effective for cell contour segmentation task. We also give a few examples for visual comparison in Fig. 6, though the ground truth is not available.

TABLE IV
PERFORMANCE COMPARISON OF CELL CONTOUR SEGMENTATION

Method	V^{Rand}	V^{Info}
U-Net [10]	0.9432	0.9562
Backbone	0.9569	0.9716
CE-Net	**0.9743**	**0.9878**

F. Retinal OCT Layer Segmentation

The above four application are conducted on two-class segmentation problems where we only need to segment foreground objects from background. In this paper, we also show that our method is applicable for multi-class segmentation tasks. We use the retinal OCT layer segmentation as an example to apply CE-Net to segment 11 retinal layers [77]. This dataset contains 20 3D volumes and each volume has 256 2D scans. Ten boundaries have been manually demarcated to divide each 2D image into 11 parts: boundary 1 corresponding to internal limiting membrane(ILM); boundary 2 between nerve fiber layer and the ganglion cells layer (NFL/GCL); boundary 3 between inner plexiform layer and the inner nuclear layer (IPL/INL); boundary 4 between the inner nuclear layer and the outer plexiform layer (INL/OPL); boundary 5 between the outer plexiform layer and the outer nuclear layer (OPL/ONL); boundary 6 corresponding to the external limiting membrane (ELM); boundary 7 corresponding to the upper boundary of inner segment (up IS); boundary

TABLE V
THE COMPARISON RESULTS ON TOPCON DATASET

Method	ILM	NFL/GCL	IPL/INL	INL/OPL	OPL/ONL	ELM	Up IS/OS	Low IS/OS	OS/RPE	BM/Choroid	Overall
Topcon [77]	1.61	2.09	2.10	2.27	-	2.17	1.82	-	1.65	1.80	-
SRR [77]	1.61	**2.02**	**2.02**	1.91	-	1.86	1.63	-	1.62	1.80	-
FCN [41]	2.10	4.41	3.77	4.54	4.78	4.52	3.84	4.36	5.06	7.88	4.53
U-Net [10]	1.38	3.05	2.70	2.77	3.30	2.34	1.86	2.00	2.42	2.65	2.45
Backbone	2.13	2.70	2.52	2.20	2.79	1.91	1.26	1.60	2.02	2.70	2.18
CE-Net w/ CE	1.45	2.48	2.20	2.08	2.55	1.66	1.19	**1.04**	1.52	1.82	1.80
CE-Net w/ Dice	**1.37**	**2.02**	2.08	**1.80**	**2.47**	**1.48**	**1.10**	1.26	**1.48**	**1.74**	**1.68**

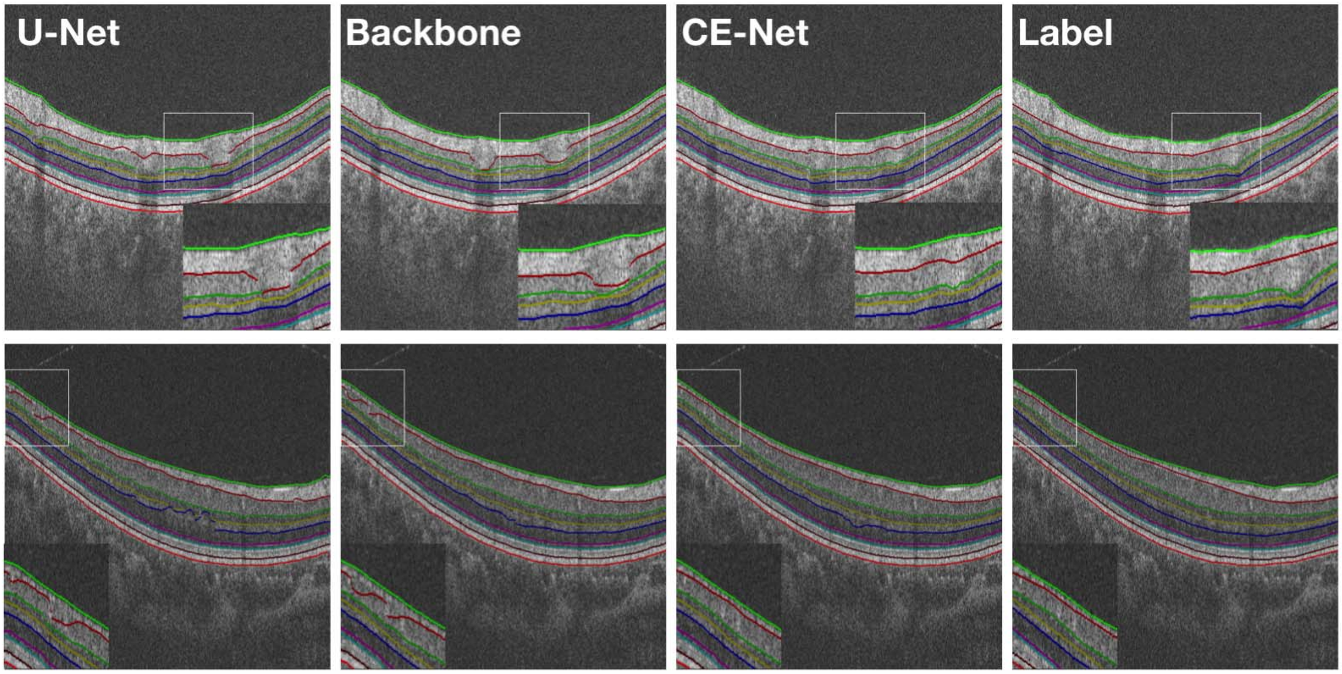

Fig. 7. Sample results. From left to right: U-Net, Backbone, CE-Net and ground-truth masks. The edges between different layers have been marked with colored lines

8 corresponding to the lower boundary of inner segment (low IS); boundary 9 between the outer segments and the retinal pigment epithelium (OS/RPE); boundary 10 between Bruchs membrane and the choroid (BM/Choroid). To evaluate the performance, we adopt the mean absolute error [77], which has been commonly used to evaluate the accuracy of retinal OCT layer segmentation.

We compare our proposed method with some state-of-the-art OCT layer segmentation approaches: Topcon built-in method in [77], Speckle Reduction by Reconstruction (SRR) method [77], FCN [41] and U-Net [10].

The performance comparisons are summarized in Table V. Compared with U-Net and the backbone approach, our CE-Net achieves an overall mean absolute error of 1.68, which is a relative reduction of 31.4% from 2.45 and 22.9% from the 2.18, respectively. Compared to the Topcon built-in method and SRR, our CE-Net also achieves better results in most scenarios. This indicates that our proposed CE-Net could also be applied to multi-class segmentation tasks. Further, we also conduct the comparison experiments between cross entropy loss and dice loss. Table V shows that the CE-Net with dice loss is superior to that with cross entropy loss.

We also present some sample results in Fig. 7 to visually compare our method with U-Net and the Backbone approach. The images clearly show more accurate segmentation results by our CE-Net.

G. Ablation Study

In order to justify the effectiveness of the pretrained ResNet, DAC block and RMP block in the proposed CE-Net, we conduct the following ablation studies using the ORIGA and DRIVE datasets as examples:

1) Ablation Study for Adopting Pretrained ResNet Model: Our proposed method is based on U-Net, therefore U-Net is the most fundamental baseline model. We employed the residual block to replace the original encoder block of U-Net, aiming at enhancing the learning capability. We call the modified U-shape network with pretrained residual block and feature decoder as 'Backbone'. Recent work [78] points out that ImageNet pre-training largely helps to circumvent optimization problems and fine-tuning from pretrained weights converges faster than that from scratch. We have also conducted experiments to compare the results with pre-training to those without. Fig. 8 shows how the losses change in the two

TABLE VI
Ablation Study for Each Component on ORIGA and DRIVE Datasets

Method	ORIGA	DRIVE	
	E	Acc	AUC
U-Net	0.115±0.068	0.939±0.006	0.960±0.006
Backbone	0.075±0.068	0.943±0.004	0.971±0.005
Backbone + Inception-block	0.068±0.059	0.950±0.004	0.972±0.005
Backbone + DAC w/o atrous	0.073±0.050	0.952±0.004	0.970±0.005
Backbone + DAC with atrous	0.061±0.043	0.953±0.004	0.977±0.006
Backbone + RMP	0.061±0.044	0.952±0.004	0.974±0.005
Backbone + Inception-ResNet-block	0.065±0.042	0.951±0.004	0.974±0.005
CE-Net	**0.058±0.032**	**0.955±0.003**	**0.978±0.006**

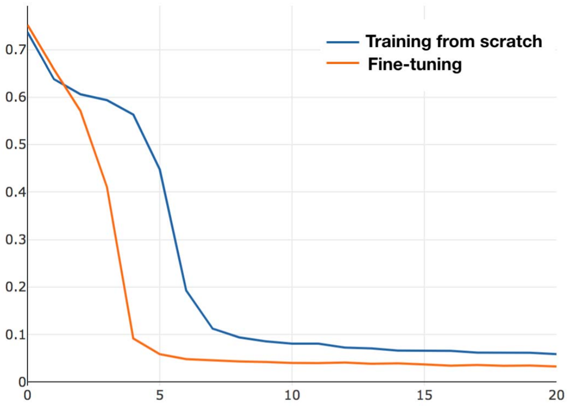

Fig. 8. The orange line represents the training loss of fine-tuning from pretrained weights, while the blue represents loss of end-to-end training from scratch.

scenarios. As we can see, the loss decreases faster in the case with pretraining than that without. Table VI shows the segmentation results. By adopting the pretrained ResNet blocks, the Backbone approach achieved a better performance. For OD segmentation, the overlapping error is decreased by 34.8% from 0.115 to 0.075. For retinal vessel detection, the *Acc* and *AUC* are increased from 0.939 and 0.960 to 0.943 and 0.971, respectively. The results indicate that pretrained ResNet blocks are beneficial.

2) Ablation Study for Dense Atrous Convolution Block: The proposed DAC block employs the atrous convolution with different rates, assembled in the Inception-like block. Therefore, we first conduct experiments to validate the usefulness of the atrous convolution. We use regular convolution to replace the atrous convolution in DAC block (referred to Backbone + DAC w/o atrous). As shown in Table VI, our proposed DAC module (referred to Backbone + DAC with atrous) reduces the overlapping error by 16.4% from 0.073 to 0.061 in OD segmentation and improves the *Acc* and *AUC* in retinal vessel detection. This indicates that atrous convolution helps to extract high-level semantic features, compared to the regular convolution. We also compare our proposed DAC block with the regular Inception-V2 block (referred to Backbone + Inception-block). The comparison results show that the DAC block outperforms the regular inception block, with a relative reduction of 10.3% from 0.068 to 0.061 in overlapping error for OD segmentation. Finally, the overlapping error is reduced by 18.7% from 0.075 of Backbone to 0.061 (Backbone + DAC). This shows that the proposed DAC block is able to further extract global information to get high-level semantic feature maps with high resolution, which is useful for our segmentation task.

3) Ablation Study for Residual Multi-Kernel Pooling Module: Table VI also shows the effect of RMP, which boosts the performance of OD segmentation. The Backbone with RMP module is referred to as 'Backbone + RMP'. Compared to the Backbone, the overlapping error decreased by 18.7% from 0.075 to 0.061 in OD segmentation, while the *Acc* and *AUC* scores increased from 0.943 and 0.971 to 0.952 and 0.974 for retinal vessel detection. The RMP module could encode the global information and change the combination way of feature maps.

4) Ablation Study for Network With Similar Complexity: Researchers have shown that the complexity is an embodiment of the network capability [79] and an increased complexity often leads to better performance. Therefore, there is a concern that the improvements might come from the increased complexity of the network. To ease such a concern, we compare our network with a network with similar complexity. In this paper, we compare it with the aforementioned backbone backed up by regular Inception-ResNet-V2 blocks (Backbone + Inception-ResNet-block). Table VI shows that our CE-Net is better, with an overlapping error reduction from 0.065 to 0.058 in OD segmentation and the *Acc* and *AUC* scores increase from 0.951 and 0.974 to 0.955 and 0.978.

IV. Conclusions

Medical image segmentation is important in the medical image analysis. In this paper, we propose an end-to-end deep learning framework named CE-Net for medical image segmentation. Compared with U-Net, the proposed CE-Net adopts pretrained ResNet block in the feature encoder. A newly proposed dense atrous convolution block and residual multi-kernel pooling are integrated to the ResNet modified U-Net structure to capture more high-level features and preserve more spatial information. Our method can be applied to a new application by fine-tuning our model using the new training data and the manual ground truth. Our experimental results show that the proposed method is able to improve the medical image segmentation in different tasks, including optic disc segmentation, retinal vessel detection, lung segmentation, cell contour segmentation and retinal OCT layer segmentation. It is believed that the approach is a general one and can be applied to other 2D medical image segmentation tasks. In this paper, our method is validated on 2D images now and the extension to 3D data would be a possible future work.

References

[1] J. Cheng *et al.*, "Superpixel classification based optic disc and optic cup segmentation for glaucoma screening," *IEEE Trans. Med. Imag.*, vol. 32, no. 6, pp. 1019–1032, Jun. 2013.

[2] 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 Trans. Med. Imag.*, vol. 37, no. 7, pp. 1597–1605, Jul. 2018.

[3] A. Aquino, M. E. Gegundez-Arias, and D. Marin, "Detecting the optic disc boundary in digital fundus images using morphological, edge detection, and feature extraction techniques," *IEEE Trans. Med. Imag.*, vol. 29, no. 11, pp. 1860–1869, Nov. 2010.

[4] H. Fu, Y. Xu, S. Lin, D. W. K. Wong, and J. Liu, "Deepvessel: Retinal vessel segmentation via deep learning and conditional random field," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Salmon Tower Building, NY, USA: Springer, pp. 132–139, Oct. 2016.

[5] Y. Zhao, L. Rada, K. Chen, S. P. Harding, and Y. Zheng, "Automated vessel segmentation using infinite perimeter active contour model with hybrid region information with application to retinal images," *IEEE Trans. Med. Imag.*, vol. 34, no. 9, pp. 1797–1807, Sep. 2015.

[6] P. Liskowski and K. Krawiec, "Segmenting retinal blood vessels with deep neural networks," *IEEE Trans. Med. Imag.*, vol. 35, no. 11, pp. 2369–2380, Nov. 2016.

[7] S. Roychowdhury, D. D. Koozekanani, and K. K. Parhi, "Iterative vessel segmentation of fundus images," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 7, pp. 1738–1749, Jul. 2015.

[8] G. Azzopardi, N. Strisciuglio, M. Vento, and N. Petkov, "Trainable COSFIRE filters for vessel delineation with application to retinal images," *Med. Image Anal.*, vol. 19, no. 1, pp. 46–57, Jan. 2015.

[9] Y. Al-Kofahi, W. Lassoued, W. Lee, and B. Roysam, "Improved automatic detection and segmentation of cell nuclei in histopathology images," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 4, pp. 841–852, Apr. 2010.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Salmon Tower Building, NY, USA: Springer, Nov. 2015, pp. 234–241.

[11] T.-H. Song, V. Sanchez, H. ElDaly, and N. M. Rajpoot, "Dual-channel active contour model for megakaryocytic cell segmentation in bone marrow trephine histology images," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 12, pp. 2913–2923, Dec. 2017.

[12] S. Wang *et al.*, "Central focused convolutional neural networks: Developing a data-driven model for lung nodule segmentation," *Med. Image Anal.*, vol. 40, pp. 172–183, Aug. 2017.

[13] W. Shen *et al.*, "Learning from experts: Developing transferable deep features for patient-level lung cancer prediction," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Salmon Tower Building, NY, USA: Springer, Oct. 2016, pp. 124–131.

[14] J. Song *et al.*, "Lung lesion extraction using a toboggan based growing automatic segmentation approach," *IEEE Trans. Med. Imag.*, vol. 35, no. 1, pp. 337–353, Jan. 2016.

[15] 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 Trans. Med. Imag.*, vol. 20, no. 7, pp. 595–604, Jul. 2001.

[16] T. Heimann, H.-P. Meinzer, and I. Wolf, "A statistical deformable model for the segmentation of liver CT volumes using extended training data," in *Proc. MICCAI Work*, Oct. 2007, pp. 161–166.

[17] A. Makropoulos *et al.*, "Automatic whole brain MRI segmentation of the developing neonatal brain," *IEEE Trans. Med. Imag.*, vol. 33, no. 9, pp. 1818–1831, Sep. 2014.

[18] B. H. Menze *et al.*, "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Trans. Med. Imag.*, vol. 34, no. 10, pp. 1993–2024, Oct. 2015.

[19] V. Cherukuri, P. Ssenyonga, B. C. Warf, A. V. Kulkarni, V. Monga, and S. J. Schiff, "Learning based segmentation of CT brain images: Application to postoperative hydrocephalic scans," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 8, pp. 1871–1884, Aug. 2018.

[20] M. Huang, W. Yang, Y. Wu, J. Jiang, W. Chen, and Q. Feng, "Brain tumor segmentation based on local independent projection-based classification," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 10, pp. 2633–2645, Oct. 2014.

[21] S. Kumar, S. Conjeti, A. G. Roy, C. Wachinger, and N. Navab, "Infinet: Fully convolutional networks for infant brain MRI segmentation," in *Proc. 15th Int. Symp. Biomed. Imaging (ISBI)*, Apr. 2018, pp. 145–148.

[22] W. Chen, R. Smith, S.-Y. Ji, K. R. Ward, and K. Najarian, "Automated ventricular systems segmentation in brain CT images by combining low-level segmentation and high-level template matching," *BMC Med. Inf. Decis. Making*, vol. 9, no. 1, p. S4, Dec. 2009.

[23] X. Zhu and R. M. Rangayyan, "Detection of the optic disc in images of the retina using the Hough transform," in *Proc. Eng. Med. Biol. Soc.*, Aug. 2008, pp. 3546–3549.

[24] A. Mihaylova and V. Georgieva, "Spleen segmentation in MRI sequence images using template matching and active contours," *Proc. Comput. Sci.*, vol. 131, pp. 15–22, Dec. 2018.

[25] A. Tsai *et al.*, "A shape-based approach to the segmentation of medical imagery using level sets," *IEEE Trans. Med. Imag.*, vol. 22, no. 2, pp. 137–154, Feb. 2003.

[26] F. Khalifa *et al.*, "A new deformable model-based segmentation approach for accurate extraction of the kidney from abdominal CT images," in *Proc. 18th IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2011, pp. 3393–3396.

[27] T. Chen and D. Metaxas, "Image segmentation based on the integration of Markov random fields and deformable models," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Salmon Tower Building, NY, USA: Springer, Oct. 2000, pp. 256–265.

[28] J. Lowell *et al.*, "Optic nerve head segmentation," *IEEE Trans. Med. Imag.*, vol. 23, no. 2, pp. 256–264, Feb. 2004.

[29] J. Xu, O. Chutatape, E. Sung, C. Zheng, and P. C. T. Kuan, "Optic disk feature extraction via modified deformable model technique for glaucoma analysis," *Pattern Recognit.*, vol. 40, no. 7, pp. 2063–2076, Jul. 2007.

[30] I. Aganj, M. G. Harisinghani, R. Weissleder, and B. Fischl, "Unsupervised medical image segmentation based on the local center of mass," *Sci. Rep.*, vol. 8, no. 1, Aug. 2018, Art. no. 13012.

[31] M. Kanimozhi and C. H. Bindu, "Brain MR image segmentation using self organizing map," *Brain*, vol. 2, no. 10, pp. 261–274, Oct. 2013.

[32] T. Tong *et al.*, "Discriminative dictionary learning for abdominal multi-organ segmentation," *Med. Image Anal.*, vol. 23, no. 1, pp. 92–104, 2015.

[33] M. D. Abràmoff *et al.*, "Automated segmentation of the optic disc from stereo color photographs using physiologically plausible features," *Invest. Ophthalmol. Vis. Sci.*, vol. 48, pp. 1665–1673, Apr. 2007.

[34] Z. Tian, L. Liu, Z. Zhang, and B. Fei, "Superpixel-based segmentation for 3D prostate MR images," *IEEE Trans. Med. Imag.*, vol. 35, no. 3, pp. 791–801, Mar. 2016.

[35] T. Kitrungrotsakul, X.-H. Han, and Y.-W. Chen, "Liver segmentation using superpixel-based graph cuts and restricted regions of shape constrains," in *Proc. Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 3368–3371.

[36] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[37] K. Zhou *et al.*, "Multi-cell multi-task convolutional neural networks for diabetic retinopathy grading," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 2724–2727.

[38] Z. Wang, Y. Yin, J. Shi, W. Fang, H. Li, and X. Wang, "Zoom-in-Net: Deep mining lesions for diabetic retinopathy detection," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Salmon Tower Building, NY, USA: Springer, Sep. 2017, pp. 267–275.

[39] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 2843–2851.

[40] K. Kamnitsas *et al.*, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Med. Image Anal.*, vol. 36, pp. 61–78, Feb. 2017.

[41] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.

[42] B. Norman, V. Pedoia, and S. Majumdar, "Use of 2D U-Net convolutional neural networks for automated cartilage and meniscus segmentation of knee mr imaging data to determine relaxometry and morphometry," *Radiology*, vol. 288, no. 1, Mar. 2018, Art. no. 172322.

[43] A. Sevastopolsky, "Optic disc and cup segmentation methods for glaucoma detection with modification of U-net convolutional neural network," *Pattern Recognit. Image Anal.*, vol. 27, no. 3, pp. 618–624, Jul. 2017.

[44] A. G. Roy *et al.*, "ReLayNet: Retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks," *Biomed. Opt. Express*, vol. 8, no. 8, pp. 3627–3642, 2017.

[45] B. A. Skourt, A. El Hassani, and A. Majda, "Lung CT image segmentation using deep neural networks," *Proc. Comput. Sci.*, vol. 127, pp. 109–113, Dec. 2018.

[46] H. Chen, Q. Dou, L. Yu, and P.-A. Heng. (Aug. 2016). "VoxResNet: Deep voxelwise residual networks for volumetric brain segmentation." [Online]. Available: https://arxiv.org/abs/1608.05895

[47] Q. Dou, H. Chen, Y. Jin, L. Yu, J. Qin, and P.-A. Heng, "3D deeply supervised network for automatic liver segmentation from CT volumes," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Springer, 2016, pp. 149–157.

[48] S. Apostolopoulos, S. De Zanet, C. Ciller, S. Wolf, and R. Sznitman, "Pathological oct retinal layer segmentation using branch residual U-shape networks," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Salmon Tower Building, NY, USA: Springer, Oct. 2017, pp. 294–301.

[49] E. Gibson *et al.*, "Automatic multi-organ segmentation on abdominal CT with dense V-networks," *IEEE Trans. Med. Imag.*, vol. 37, no. 8, pp. 1822–1834, Aug. 2018.

[50] C. Peng, X. Zhang, G. Yu, G. Luo, and J. Sun, "Large kernel matters—Improve semantic segmentation by global convolutional network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* 2017, pp. 4353–4361.

[51] G. Lin, A. Milan, C. Shen, and I. Reid, "Refinenet: Multi-path refinement networks for high-resolution semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1925–1934.

[52] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2881–2890.

[53] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[54] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proc. AAAI*, vol. 4, 2017, p. 12.

[55] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, 2015.

[56] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A.-L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2017.

[57] W. R. Crum, O. Camara, and D. L. G. Hill, "Generalized overlap measures for evaluation and validation in medical image analysis," *IEEE Trans. Med. Imag.*, vol. 25, no. 11, pp. 1451–1461, Nov. 2006.

[58] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis.*, Oct. 2016, pp. 565–571.

[59] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970.

[60] J. Dai, Y. Li, K. He, and J. Sun, "R-fcn: Object detection via region-based fully convolutional networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 379–387.

[61] Y. Zhang, Z. Qiu, T. Yao, D. Liu, and T. Mei, "Fully convolutional adaptation networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 6810–6818.

[62] A. C. Wilson, R. Roelofs, M. Stern, N. Srebro, and B. Recht, "The marginal value of adaptive gradient methods in machine learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 4148–4158.

[63] N. S. Keskar and R. Socher. (Dec. 2017)."Improving generalization performance by switching from Adam to SGD," [Online]. Available: https://arxiv.org/abs/1712.07628

[64] Y. Jiang *et al.*, "Optic disc and cup segmentation with blood vessel removal from fundus images for glaucoma detection," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Jul. 2018, pp. 862–865.

[65] Z. Gu *et al.*, "Deepdisc: Optic disc segmentation based on atrous convolution and spatial pyramid pooling," in *Proc. Int. Workshop Comput. Pathol.* Salmon Tower Building, NY, USA: Springer, Sep. 2018, pp. 253–260.

[66] Z. Zhang *et al.*, "ORIGA$^{-\text{light}}$: An Online retinal fundus image database for glaucoma analysis and research," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol.*, Aug./Sep. 2010, pp. 3065–3068.

[67] E. Decencière *et al.*, "Feedback on a publicly distributed image database: The messidor database," *Image Anal. Stereol.*, vol. 33, no. 3, pp. 231–234, 2014.

[68] F. Fumero, S. Alayon, J. L. Sanchez, J. Sigut, and M. Gonzalez-Hernandez, "Rim-one: An open retinal image database for optic nerve evaluation," in *Proc. 24th Int. Symp. Comput.-Based Med. Syst.*, Jun. 2011, pp. 1–6.

[69] J. Cheng, "Sparse range-constrained learning and its application for medical image grading," *IEEE Trans. Med. Imag.*, vol. 37, no. 12, pp. 2729–2738, Dec. 2018.

[70] A. Li *et al.*, "Learning supervised descent directions for optic disc segmentation," *Neurocomputing*, vol. 275, pp. 350–357, Jan. 2018.

[71] Z. Zhang *et al.*, "Optic disc region of interest localization in fundus image for glaucoma detection in ARGALI," in *Proc. 5th IEEE Conf. Ind. Electron. Appl.*, Jun. 2010, pp. 1686–1689.

[72] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[73] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Trans. Med. Imag.*, vol. 23, no. 4, pp. 501–509, Apr. 2004.

[74] S. Xie and Z. Tu, "Holistically-nested edge detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1395–1403.

[75] A. Cardona *et al.*, "An integrated micro- and macroarchitectural analysis of the drosophila brain by computer-assisted serial section electron microscopy," *PLoS Biol.*, vol. 8, no. 10, Oct. 2010, Art. no. e1000502.

[76] I. Arganda-Carreras *et al.*, "Crowdsourcing the creation of image segmentation algorithms for connectomics," *Frontiers Neuroanatomy*, vol. 9, p. 142, Nov. 2015.

[77] J. Cheng *et al.*, "Speckle reduction in 3D optical coherence tomography of retina by a-scan reconstruction," *IEEE Trans. Med. Imag.*, vol. 35, no. 10, pp. 2270–2279, Oct. 2016.

[78] K. He, R. Girshick, and P. Dollár. (Nov. 2018). "Rethinking imagenet pre-training." [Online]. Available: https://arxiv.org/abs/1811.08883

[79] G. Huang, Z. Liu, K. Q. Weinberger, and L. van der Maaten, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4700–4708.