Medical Image Analysis 55 (2019) 216–227

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Automated segmentation of macular edema in OCT using deep neural networks

Junjie Hu, Yuanyuan Chen, Zhang Yi*

Machine Intelligence Laboratory, College of Computer Science, Sichuan University, Chengdu 610065, PR China

ARTICLE INFO

Article history:
Received 25 December 2018
Revised 23 April 2019
Accepted 9 May 2019
Available online 10 May 2019

Keywords:
Macular edema segmentation
Deep neural networks
Atrous convolution

ABSTRACT

Macular edema is an eye disease that can affect visual acuity. Typical disease symptoms include subretinal fluid (SRF) and pigment epithelium detachment (PED). Optical coherence tomography (OCT) has been widely used for diagnosing macular edema because of its non-invasive and high resolution properties. Segmentation for macular edema lesions from OCT images plays an important role in clinical diagnosis. Many computer-aided systems have been proposed for the segmentation. Most traditional segmentation methods used in these systems are based on low-level hand-crafted features, which require significant domain knowledge and are sensitive to the variations of lesions. To overcome these shortcomings, this paper proposes to use deep neural networks (DNNs) together with atrous spatial pyramid pooling (ASPP) to automatically segment the SRF and PED lesions. Lesions-related features are first extracted by DNNs, then processed by ASPP which is composed of multiple atrous convolutions with different fields of view to accommodate the various scales of the lesions. Based on ASPP, a novel module called stochastic ASPP (sASPP) is proposed to combat the co-adaptation of multiple atrous convolutions. A large OCT dataset provided by a competition platform called “AI Challenger” are used to train and evaluate the proposed model. Experimental results demonstrate that the DNNs together with ASPP achieve higher segmentation accuracy compared with the state-of-the-art method. The stochastic operation added in sASPP is empirically verified as an effective regularization method that can alleviate the overfitting problem and significantly reduce the validation error.

1. Introduction

The macula is located in the central part of the retina and is the main sensory region for detecting light signals. Impairment of the macula impacts visual acuity to varying degrees. Macular edema, characterized by an accumulation of fluid near the fovea in the macula, is an common reason for vision loss (Goatman, 2006). Macular edema does not signify a particular ocular disease, but a series of macular responses to retinal environment alternation (Coscas et al., 2010), including pigment epithelium detachment (PED) and subretinal fluid (SRF). Common causes of macular edema include age-related macular degeneration, diabetic retinopathy, and intraocular surgery. Early diagnosis and the timely treatment of macular edema can help to reduce vision loss.

Optical coherence tomography (OCT) is a non-invasive imaging technology introduced in 1991 that has been widely used in the clinical evaluation of macular edema (Huang et al., 1991; Coscas et al., 2010; Tranos et al., 2004). Compared with other imaging modalities such as fluorescein angiography, the advantages of OCT include its non-invasive property, rapid imaging, high reproducibility, and safety profile (Trichonas and Kaiser, 2014). OCT is based on the interferometric principle that generates cross-sectional images in high resolution. It can delineate multiple retinal layers and visualize the structural changes of the retina in large areas (Wolf and Wolf-Schnurrbusch, 2010). The A scan (one dimension) in depth is reconstructed using the reflective signal, and a number of A scans are used to construct the B scan slice (two dimensions) (Drexler et al., 2003). The complete OCT volume is composed of multiple B scan slices. In the OCT slices, normal and abnormal tissues have a different appearance because of their distinct reflectivity patterns. Fig. 1 shows four slices, which represent the appearance of normal and abnormal tissues.

The accurate segmentation of macular edema related lesions is required for quantification in clinical practice. However, the manual annotation of lesions is subjective, labor-intensive, and prone to errors. This is mainly caused by the limited use of prior knowledge about the distorted morphology and the blurred boundaries near the lesions (Xu et al., 2017). Many computer-aided systems have been proposed to assist ophthalmologists in the clinical diagnosis of macular edema. Conventional segmentation methods

* Corresponding author.
E-mail address: zhangyi@scu.edu.cn (Z. Yi).

https://doi.org/10.1016/j.media.2019.05.002
1361-8415/© 2019 Elsevier B.V. All rights reserved.

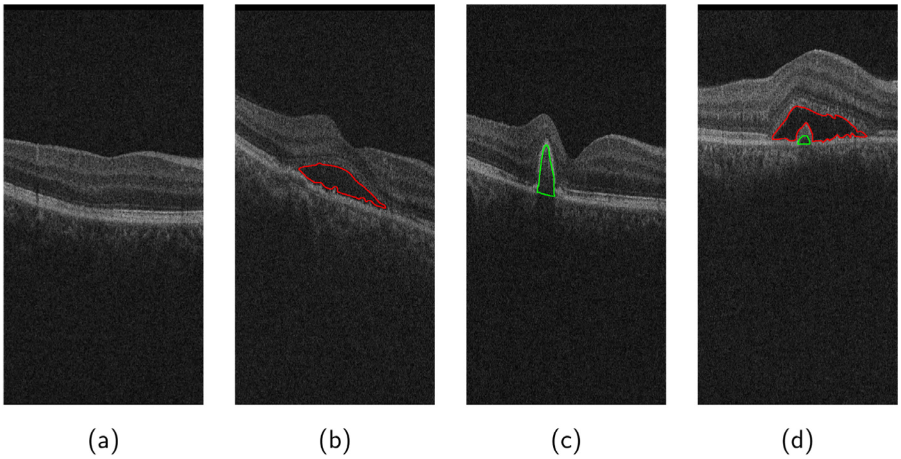

Fig. 1. Appearance of normal and macular edema tissues in OCT. The lesions of SRF and PED are marked by red and green, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

in these systems include threshold-based and graph-based approaches that use hand-crafted features (Chen et al., 2013; Penha et al., 2012; Ahlers et al., 2008). However, these low-level features are sensitive to image quality and require significant domain knowledge. In this paper, we aim to automatically segment SRF and PED lesions in OCT images using deep neural networks (DNNs). Different from conventional methods that use elaborate hand-crafted features, DNNs learn to extract different levels of features from low to high as the depth of the network increases. Many novel networks have been proposed to perform segmentation tasks, including FCN (Long et al., 2015), UNet (Ronneberger et al., 2015), SegNet (Badrinarayanan et al., 2017), DeepLab (Chen et al., 2018a), etc. Based on these networks, some studies have been conducted to segment macular edema related lesions in an end-to-end manner. Significantly, ReLayNet (Roy et al., 2017) and 3D U-Net (De Fauw et al., 2018), which are DNNs-based methods, have achieved better performance in fluid segmentation task compared with traditional methods. These results demonstrate that DNNs are promising approaches in macular edema segmentation tasks.

Although DNNs-based models have achieved remarkable performance in macular edema segmentation tasks, the depth of neural networks in current studies is relatively shallow. For example, the encoder and decoder parts in ReLayNet are both networks with six layers. The question is whether the depth of the model could help to improve segmentation accuracy. Moreover, macular edema related lesions typically have multiple scales from small to large. The conventional method to manage this is to feed the model with the inputs rescaled multiple times, and then aggregate the features. However, this approach has a high computational cost. Based on the above considerations, in this paper, DNNs with a considerably greater depth (e.g., ResNet50 (He et al., 2016b)) is used as an encoder to extract highly abstract features from the OCT image. Then atrous spatial pyramid pooling (ASPP) (Chen et al., 2018a), which has been successfully applied to the semantic image segmentation task, is used to process these features. ASPP is based on atrous convolution (Chen et al., 2018a), which is a special convolution operation with “holes” inserted between the elements of the kernel. Atrous convolution can effectively increase the field of view of the kernel, but does not incur more learnable parameters and computational cost. ASPP works by using multiple atrous convolutions with different fields of view to capture features at multiple scales. In the experimental section, we empirically demonstrate that ASPP together with features from ResNet50 has better segmentation accuracy than ReLeyNet and U-Net.

Based on ASPP, a novel module that combines atrous convolution and randomness is proposed in this paper. The proposed module is called stochastic ASPP (sASPP) because features from atrous convolutions in ASPP may be randomly dropped (or retained) during the training phase. An example of sASPP in the encoder-decoder segmentation model is presented in Fig. 2. The input of sASPP is the highly abstract feature from the DNNs-based encoder network, and is independently processed by the atrous convolutions. The main difference between ASPP and sASPP is the random drop operation applied to the feature maps produced by the atrous convolutions. In this example, there are two feature maps dropped (indicated by the dashed line). sASPP is proposed to prevent the co-adaptation of the atrous convolutions in ASPP. Moreover, the proposed sASPP can be regarded as 2^n possible models, suppose there are n atrous convolutions. The proposed sASPP should have better performance compared with ASPP, because the latter uses the multiple atrous convolutions statically, whereas sASPP attempts to use a number of possible combinations of them. The experimental results demonstrate that the proposed sASPP is an effective regularization method that achieves lower errors on the validation dataset. The main contributions of this study are summarized as follows:

(i) ASPP together with an encoder-decoder model is used to segment the SRF and PED lesions in OCT images. The method has superior accuracy compared with the state-of-the-art method.
(ii) A module called sASPP is further proposed, which stochastically drops the feature maps from the atrous convolutions in ASPP. Randomness in sASPP is assumed to prevent the co-adaptation of the atrous convolutions.
(iii) The experimental results demonstrate that sASPP is an effective regularization method to alleviate the overfitting problem. Compared with ASPP, the proposed sASPP can achieve higher segmentation accuracy while maintaining lower errors on the validation dataset.

2. Related works

In this section, an overview of previous studies based on the traditional graph theory method for macular edema segmentation is presented, followed by a brief introduction of DNNs and their applications.

2.1. Traditional methods for segmenting macular edema

Graph theory has been widely used to segment anatomical and pathological structures in OCT images (Haeker et al., 2007; Garvin et al., 2009; Chiu et al., 2010; 2015). Each OCT slice image can be

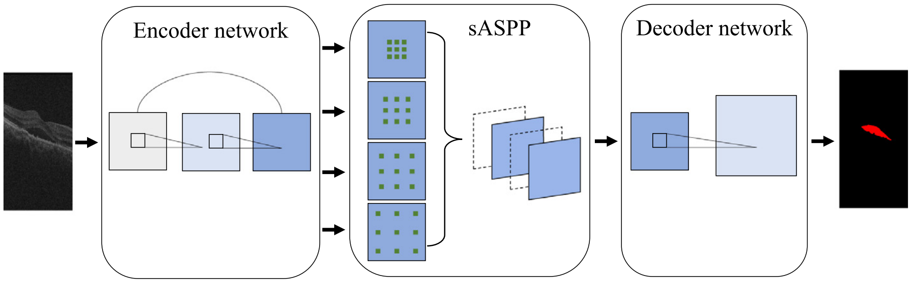

Fig. 2. Proposed sASPP in the encoder-decoder based segmentation model. The dashed line denotes that the feature maps are dropped during training.

considered as a graph, where each pixel corresponds to a node in the graph. Nodes are connected by edges with assigned weights. The task is to travel across the graph from a start node to an end node with the minimum sum of weights. Then the preferred pathway is the boundary between the pathological and normal structures in the segmentation task. The appropriate weight assignment to the edge and the solution of the minimum sum of weights between two nodes are the two keys in graph-based method. Particularly, the intensity gradients are used to estimate the weights, followed by Dijkstra's algorithm to determine the lowest weighted path of the graph in Chiu et al. (2010) to segment the retinal layers. This method was further improved by a subsequent study that applied constrains to add prior information (Dufour et al., 2013). Moreover, Chiu et al. (2015) proposed segmenting the retinal layers together with the fluid regions using a two-stage segmentation method. In the first stage, the kernel regression method is used to generate coarse segmentation results. Graph theory is then used to refine the boundaries in the second stage.

In addition to the aforementioned graph based method, there also exist other machine learning based methods, such as k-nearest neighbor (k-NN) (Chen et al., 2012), random forest (Lang et al., 2015), etc. These traditional methods have achieved progress in the retinal layer or fluid segmentation task; however, some limitations should be noted. First, these methods depend heavily on hand-crafted, low-level features, which have limited representation ability and are sensitive to image quality. Second, the traditional methods typically divide the segmentation procedure into several stages, and the errors in the previous stage may be amplified in the latter stages. Moreover, significant domain knowledge is required in the designation of hand-crafted features and segmentation stages. These potential limitations have constrained the wide application of conventional segmentation methods in the clinical analysis of macular edema.

2.2. Deep neural networks for segmenting macular edema

DNNs are composed of multiple layers that transform the input nonlinearly layer by layer in a data-driven manner. Two types of neural networks, that is, feed-forward neural networks (FNNs) (LeCun et al., 1998; He et al., 2016a) and recurrent neural networks (RNNs) (Williams and Zipser, 1989; Yi and Tan, 2004; Yi, 2010), have been heavily studied during the recent decades. Since 2012 when AlexNet (Krizhevsky et al., 2012) won the ILSVRC-2012 competition, there have been revolutionary advances in computer vision tasks using DNNs. These achievements can be attributed to several factors, including novel network architecture, graphics processing units (GPUs) with powerful computation ability, large-scale annotated dataset, etc. In addition to computer vision tasks, DNNs have been successfully applied to medical image analysis applications, such as retinopathy of prematurity (Hu et al., 2018), lymph node (Anthimopoulos et al., 2016), breast cancer classification tasks (Carneiro et al., 2017), etc.

A recent study that used a U-Net based network called ReLayNet (Roy et al., 2017), was the first work to use fully convolutional neural networks in retinal layer and fluid segmentation tasks. ReLayNet is composed of two parts, an encoder and decoder, both of which are based on neural networks with six layers. The encoder network is constructed by convolutional and pooling layers in turn, which are designed to extract features from the input OCT image. The decoder network attempts to predict the category of each raw pixel by processing features from the encoder network using alternating convolutional and unpooling layers. The unpooling layer upsamples the features using the indices from the corresponding pooling layer, which helps to accurately segment the fluid with small regions. The upsampled features are then concatenated with those from the encoder network that have an identical spatial size. ReLayNet is trained end-to-end and has demonstrated superior performance compared with conventional segmentation methods. Besides ReLayNet, a 3D U-Net (De Fauw et al., 2018) network is also used to segment the lesions in OCT images. Based on the segmentation results produced by the 3D U-Net, classification networks are further used to give diagnoses and referrals. The two-stage framework can reduce the error caused by data from different devices.

The depth of neural networks is a critical factor, where the network with deep depth probes extracts features with a high abstraction level. However, one major limitation of ReLayNet and 3D U-Net is the comparatively shallow depth in the encoder, which may impede segmentation performance. In this paper, a novel module called sASPP is proposed to accurately segment macular edema related lesions. sASPP is based on ASPP, which is a module composed of multiple atrous convolutions with different fields of view, whereas sASPP stochastically drops features from each atrous convolution. sASPP has three advantages over ReLayNet. First, the input of sASPP is from a very deep neural network (e.g., ResNet50), which can discriminatively represent the data. Second, the multiple atrous convolutions with different fields of view in sASPP can adapt to objects of various scales. Third, the randomness in sASPP helps to alleviate the overfitting problem in the segmentation task.

3. Methodology

In this section, the ASPP is introduced, including the definition of atrous convolution and features in multiple scales. Then the proposed ASPP with randomness is illustrated in detail.

3.1. Atrous spatial pyramid pooling

3.1.1. Atrous convolution

In the convolution operation, the parameters are kernels that connect with the input locally. An example of a convolution operation is shown in Fig. 3(a). The convolution operation can be

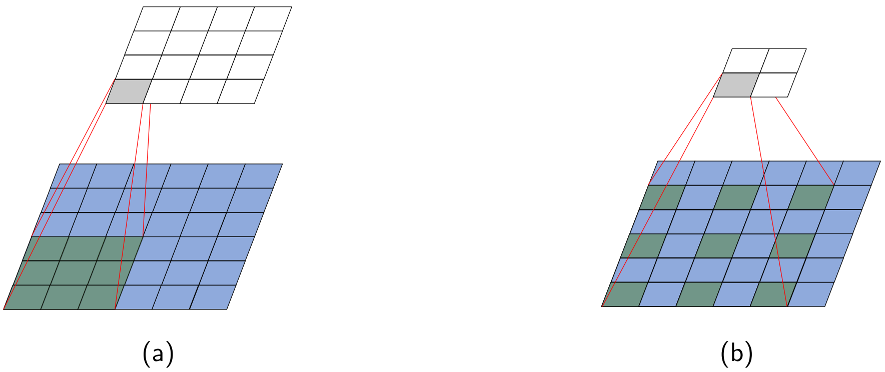

Fig. 3. Comparison between convolution and atrous convolution operations. (a) Convolution operation for which the kernel size is 3 and stride is 1. (b) Atrous convolution operation for which the kernel size is 3, stride is 1, and rate is 2. Both kernels in the two types of operations are shared among each location in the inputs.

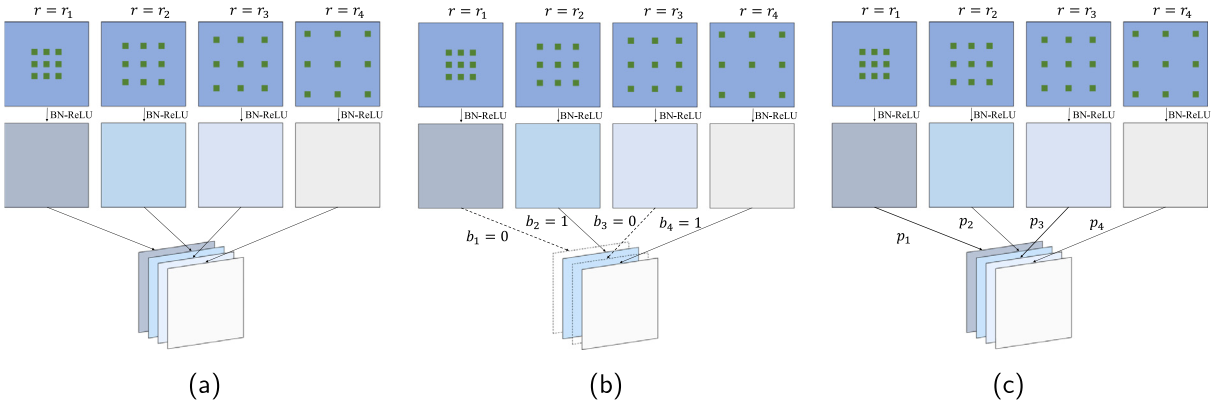

Fig. 4. Comparison between ASPP and the proposed sASPP. (a) ASPP is composed of four atrous convolutions with rates ranging from r_1 to r_4. The kernel sizes of all the operations are equal to 3. (b) sASPP in a particular state during the training phase. sASPP has the same rates and kernel size as ASPP shown in (a). The value of b denotes the existence of the feature map, where $b = 0$ denotes the dropped state, and vice versa. (c) sASPP during the test phase. All the feature maps from the atrous convolutions are retained during the test phase, with their magnitude scaled by the retention probability in the training phase.

formulated as

$$z^{l+1}_{v,u} = \sum_{c=0}^{C^l-1} \sum_{p=0}^{P^l-1} \sum_{q=0}^{Q^l-1} a^l_{v+p,u+q,c} \cdot w^l_{p,q,c}, \tag{1}$$

where P, Q, and C denote the dimensions of the convolution kernel, and their corresponding lower-case letters denote the cursor in kernel w^l. $z^{l+1}_{v,u}$ denotes the input of location (v, u) in the $l+1$ layer, which is computed by the convolution operation of kernel w^l and the previous layer's output a^l. $a^{l+1}_{v,u}$ can be computed by applying non-linear transformations (e.g., Batch normalization Ioffe and Szegedy, 2015, ReLU Glorot et al., 2011 etc.) to $z^{l+1}_{v,u}$. Typically, convolution operations in modern networks often have small kernels, such as 3×3. Increasing the field of view of the kernel to capture a large context has the disadvantage of increasing the number of computations and parameters to learn. Different from the conventional operation, the atrous convolution operation (Papandreou et al., 2015) can effectively enlarge the field of view of the kernel without suffering the above problems. The newly introduced parameter in atrous convolution is rate r, which controls the spacing in the kernel. It is equivalent to insert $r - 1$ zeros between consecutive elements in the kernel. An example of an atrous convolution operation with $r = 2$ is shown in Fig. 3(b). The atrous convolution operation can be formulated as

$$z^{l+1}_{v,u} = \sum_{c=0}^{C^l-1} \sum_{p=0}^{P^l-1} \sum_{q=0}^{Q^l-1} a^l_{v+r\cdot p,u+r\cdot q,c} \cdot w^l_{p,q,c}. \tag{2}$$

It can be seen from Eq. (2) that atrous convolution is equivalent to standard convolution when $r = 1$. Using atrous convolution in neural network models, it is convenient for us to arbitrarily control the field of view of kernel by setting an appropriate value of r.

3.1.2. Multiple scale features

Typically, there often exist objects with multiple scales in segmentation tasks. To capture them simultaneously, the traditional method is to resample the input to multiple scales to train the model (Farabet et al., 2013; Lin et al., 2015). However, this approach increases the computational burden significantly. To solve this problem while maintaining the computational cost, ASPP (Chen et al., 2018a) is used to extract features with multiple scales from the input directly. ASPP is inspired by spatial pyramid pooling (SPP) (He et al., 2014), and the main difference between them is the specific operations applied to the input. Specifically, SPP uses multiple pooling operations with different scales to process the input, whereas ASPP uses multiple atrous convolutions. To obtain an intuitive understanding of ASPP, an example of four atrous convolutions with different rates is shown in Fig. 4(a). Intuitively, ASPP can be regarded as allowing the model itself to learn the optimal representation of features using multiple atrous convolutions with different rates. The four parallel atrous convolutions are independently applied to the same input to capture features with different scales. After the non-linear transformation is applied, the output is stacked prior to the next operation. Note that the number of atrous convolutions and their corresponding rates can be customized according to the specific task when applying ASPP in neural network models. The formulation of ASPP is as follows

$$\boldsymbol{a}^l = \left[\boldsymbol{a}^l_1, \ldots, \boldsymbol{a}^l_n\right]. \tag{3}$$

The multiple atrous convolutions in ASPP help to capture features with different scales; however, the optimal configuration of ASPP can only be determined by trial and error. The question is whether it is possible that the combination of the first m (assuming $m < n$) atrous convolutions perform better than applying all the n operations. Instead of traversing n atrous convolutions to obtain an optimal combination, a novel architecture is proposed that applies randomness to feature maps from atrous convolutions. The motivation and detailed analysis are described in the next section.

3.2. Stochastic atrous spatial pyramid pooling

3.2.1. Multiple scale features with randomness

Motivated by stochastic depth networks (Huang et al., 2016), which is a neural network architecture that combines the dropout method (Srivastava et al., 2014) with the residual block in ResNets, a novel method is proposed to regularize ASPP with randomly dropped feature maps from atrous convolutions. Qualitatively, stochastic depth networks can be regarded as a combination of many networks with varying depth because each residual block may be dropped independently, where the proposed architecture can be considered as ASPP with varying width. The proposed architecture is called sASPP because randomness is combined with ASPP. An example of the proposed sASPP during the training phase is shown in Fig. 4(b). $b_i \in \{0, 1\}$ is a binary variable that controls the existence of feature maps from each atrous convolution. Clearly, the first and third feature maps are dropped, and $\boldsymbol{a}^l$ is equal to $\left[\boldsymbol{0}, \boldsymbol{a}_2^l, \boldsymbol{0}, \boldsymbol{a}_4^l\right]$. Consistent with the dropout technique, scalar variable b_i obeys the Bernoulli distribution parameterized by p_i. This indicates that feature maps from each atrous convolution in ASPP have a probability of p_i of being retained and $1 - p_i$ of being dropped when b_i is one or zero, respectively. Moreover, Eq. (4) can be reformulated as

$$\boldsymbol{a}^l = \left[b_1\boldsymbol{a}_1^l, \ldots, b_n\boldsymbol{a}_n^l\right]. \tag{4}$$

Note that the dropout used in sASPP is slightly different from the original dropout method. The dropout here applies independent random variable b_i to the entire feature, rather than to each element of the feature used in the original dropout. This is to consider the complete representation of feature maps from each atrous convolution. p_i is the newly introduced parameter in sASPP, and denotes the retained probability of feature maps. The larger p_i, the greater contributions by the feature maps from the i-th atrous convolution. The reverse also applies. Intuitively, there are two options for setting p_i. One is to set p_i identically for all the feature maps (e.g. 0.8), which means that they make the same contribution to the segmentation results. The other option is set p_i as a function of i, which indicates that feature maps from different atrous convolutions contribute unequally according to the size of the field of view. Typically, p_i can increase (or decrease) linearly from p_1 to p_n. The case of linearly increasing p_i is as follows

$$p_i = p_1 + \frac{p_n - p_1}{n - 1} \cdot (i - 1). \tag{5}$$

The detailed validation of the impact of p_i on the segmentation results is presented in Section 4.

3.2.2. Model ensemble

By applying the stochastic drop operation in ASPP, feature maps from atrous convolutions are likely to be involved in the segmentation. The proposed sASPP can be considered as a combination of 2^n possible networks for the existence status of $\boldsymbol{a}_i$ during the training phase. For the test phase, one possible strategy is to traverse every combination of $\boldsymbol{a}_i$ and average the segmentation output as the final prediction. However, this would significantly increase the computational cost because of the exponential number of models. Following the strategy used in dropout, all the feature maps from atrous convolutions are retained, with their values scaled by p_i during the test phase. An example of sASPP during the test phase is shown in Fig. 4(c). This strategy can be illustrated by combining all possible networks with different widths into a single test architecture, which is an ensemble mechanism. The training and test procedures of sASPP are shown in Algorithm 1. In the training phase, n binary variables are independently sampled from Bernoulli distributions for each mini-batch. $\boldsymbol{a}^{l+1}$ is calculated by concatenating the feature maps multiplied by the binary variables. Note that the stochastic gradient descent algorithm remains the same as that in conventional DNNs. In the test phase, sASPP becomes a deterministic model that uses all the feature maps, with their magnitudes scaled by the retention probability.

Algorithm 1 Training and test phases of sASPP.

```
Input: sASPP with trainable parameters $\boldsymbol{W}$; target $\boldsymbol{y}$; learning rate $\alpha$
for training mini-batch $j$ do
    for atrous convolution $i \in [1, n]$ do
        $\boldsymbol{a}_i^{l+1} \leftarrow F^l(\boldsymbol{W}_i^l * \boldsymbol{a}^l)$
        $b_i \leftarrow Bernoulli(p_i)$
    end for
    $\boldsymbol{a}^{l+1} \leftarrow \left[b_1\boldsymbol{a}_1^{l+1}, \ldots, b_n\boldsymbol{a}_n^{l+1}\right]$
    $\boldsymbol{W} \leftarrow \boldsymbol{W} - \alpha\frac{\partial C(\boldsymbol{a}^L, \boldsymbol{y})}{\partial \boldsymbol{W}}$
end for
for test sample $j$ do
    for atrous convolution $i \in [1, n]$ do
        $\boldsymbol{a}_i^{l+1} \leftarrow F^l(\boldsymbol{W}_i^l * \boldsymbol{a}^l)$
    end for
    $\boldsymbol{a}^{l+1} \leftarrow \left[p_1\boldsymbol{a}_1^{l+1}, \ldots, p_n\boldsymbol{a}_n^{l+1}\right]$
end for
```

4. Experimental setup and results

In this section, the experimental setup is presented, including the dataset and specific configuration of the proposed method. Then the results of control experiments are given and analyzed in detail.

4.1. Experimental setup

4.1.1. Dataset

The dataset is sourced from a competition platform called AI Challenger[1] which is hosted by several well-known Internet enterprises in China. The dataset is well annotated at the pixel level. SRF and PED lesions are used to validate our proposed method. The dataset is composed of training, validation, and test parts that contain 70, 15, and 15 cases, respectively. Each case contains 128 slices, with a resolution of 512 × 1024. Note that only the training and validation datasets' annotations have been released. The team that participated in the challenge is required to train and assess its model using the training and validation dataset, and upload the segmentation prediction on the test dataset. The final rank on the leaderboard is based on the comparison between the uploaded prediction with the ground truth of the test dataset. There are two rounds in the challenge, namely round A and B. We won the first and second places in the two rounds, respectively. The test dataset in both rounds are the same, and the difference relies on the updates of the leaderboard. In round A, the participants can upload their prediction results three times per week and the score is updated immediately. In round B, there are total two chances to

[1] https://challenger.ai/competition/fl2018.

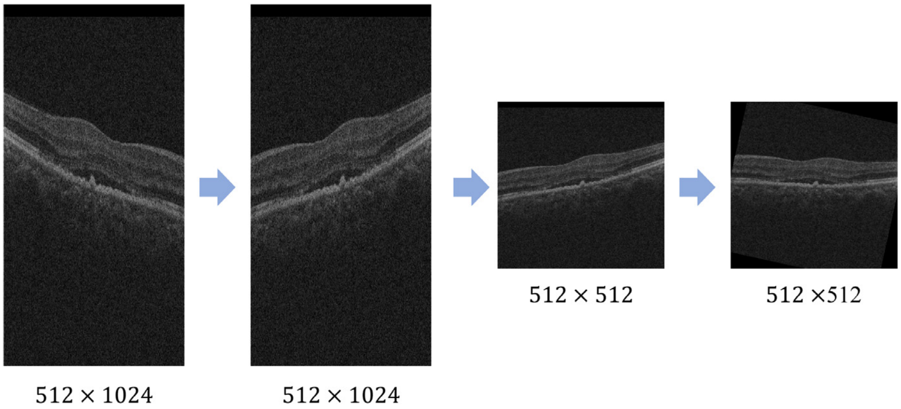

Fig. 5. Data augmentation pipeline in the training phase. The number below each image denotes the width and height. From left to right are the original, horizontal flipped, resized, and rotated images. All the operations are randomly performed.

Table 1
The leaderboard of macular edema segmentation competition in AI Challenger. The score is calculated by the mean DSC value of each lesion type.

Team	Score of Round A	Team	Score of Round B
Our	**0.7521**	965728310	0.7612
965728310	0.7466	**Our**	**0.7521**
Looking	0.7458	Looking	0.7460
DeepSeg	0.7441	Viking	0.7442
Menelvagor	0.7398	DeepSeg	0.7441

upload the prediction results, while the score is invisible till the end of the challenge. The final score in each round is based on the highest one. The leaderboard is shown in Table 1. Because the annotations of the test dataset in the competition are not provided, the models in the following experiments are evaluated on the validation dataset.

4.1.2. Configurations

ASPP and the proposed sASPP are types of modules that aim to extract features with multiple fields of view. To evaluate their performance on the macular edema segmentation task, the DeepLabv3+ (Chen et al., 2018b) model combined with ASPP or sASPP is used. DeepLab is a series of segmentation models that have achieved state-of-the-art performance on the PASCAL VOC 2012 semantic segmentation task. DeepLabv3+ is the most recent model that has delivered better performance than the previous series models. The distinction between DeepLabv3+ and the other versions is the encoder and decoder paradigm. Specifically, a backbone network (e.g., ResNet or Xception) together with atrous convolutions is used as the encoder. The decoder part uses the low-level features from the middle part of the encoder and outputs of ASPP to obtain the final segmentation results. In this study, a 1×1 convolution together with three atrous convolutions with the kernel size of 3×3 (rates are 6, 12, and 18) are used in ASPP and sASPP. ResNet50 is used as the backbone network. sASPP with three configurations are used in the control experiments, including linearly increased, decreased, and identical values of p_i, which are denoted as sASPP ($p\uparrow$), sASPP ($p\downarrow$), and sASPP ($p = 0.8$), respectively. p_i in the three configurations are {0.5, 0.6, 0.7, 0.8}, {0.8, 0.7, 0.6, 0.5}, and {0.8, 0.8, 0.8, 0.8}. For ReLayNet, the implementation is based on the open source code provided by the original authors.[2] For 3D U-Net, we implement the network structure that is described in De Fauw et al. (2018).

The size of the raw input is 512×1024, which is too large for the model. To alleviate the computational burden, the size of the input is set to 512×512 during the training phase. The dataset in the training phase is extensively augmented to mitigate the problem of overfitting. The augmentation pipeline is shown in Fig. 5. Note that the resize augmentation represented by the third image is composed of random scale and crop/padding operations. If the output of the scale operation is larger than 512, then the image is cropped to 512; otherwise it is padded with zeros. In the experiments, the scale ratio is in [0.75, 1.5] and the maximum rotation angle is 15. At the end of the augmentation pipeline, the data is divided by 255 to ensure the pixel values are located in [0,1].

DeepLabv3+ with ASPP or sASPP is trained with the back-propagation algorithm by minimizing the cross-entropy cost function with respect to the parameters

$$\mathcal{J} = -\frac{1}{M}\sum_{m=1}^{M} \boldsymbol{y}_m^{\top} ln(\boldsymbol{a}_m^L), \tag{6}$$

where M denotes the number of images in a batch and $\boldsymbol{a}^L$ denotes the output of the model after applying the softmax function. To optimize the above cost function, stochastic gradient descend with momentum algorithm is used. The learning rate and momentum are set to 1×10^{-7} and 0.9, respectively. L2 regularization with the weight decay 5×10^{-4} is also used. The number of images in each batch is 10, and the training process is completed when 60 epochs is reached.

The true positive volume fraction (TPVF), positive predictive value (PPV), and dice similarity coefficient (DSC) are used to evaluate the model's performance. These metrics have been widely used to assess the performance of segmentation tasks (Udupa et al., 2006; Papandreou et al., 2015; Roy et al., 2017; Xu et al., 2017). Their definitions are as follows

$$TPVF = \frac{|V_S \bigcap V_G|}{|V_G|}, \tag{7}$$

$$PPV = \frac{|V_S \bigcap V_G|}{|V_S|}, \tag{8}$$

$$DSC = 2 \times \frac{|V_S \bigcap V_G|}{|V_S \bigcup V_G|}, \tag{9}$$

where V_S and V_G denote the volume of the model's segmentation results and ground truth, respectively.

4.1.3. Implementation

The proposed model is implemented using PyTorch (Paszke et al., 2017). All experiments are carried out on a server with Linux OS and hardware of CPU Intel Xeon E5-2620 @2.4 GHz, four NVIDIA Tesla K40m GPUs, and 64GB of RAM.

[2] https://github.com/abhi4ssj/relaynet_pytorch.

Table 2
Performance comparison between ReLayNet, 3D U-Net, and ASPP in the segmentation of SRF and PED lesions.

Model	SRF			PED		
	DSC	TPVF	PPV	DSC	TPVF	PPV
ReLayNet	0.5472 ± 0.10	0.4550 ± 0.10	0.6862 ± 0.09	0.5820 ± 0.13	0.6224 ± 0.16	**0.5464** ± 0.14
3D U-Net	0.8060 ± 0.08	0.7328 ± 0.12	**0.8954** ± 0.07	0.6131 ± 0.11	0.7785 ± 0.17	0.5057 ± 0.07
ASPP	**0.8447** ± 0.07	**0.8214** ± 0.11	0.8693 ± 0.09	**0.6370** ± 0.09	**0.8341** ± 0.10	0.5153 ± 0.11

4.2. Results

4.2.1. Comparison with the state-of-the-art

The quantitative comparison between ASPP, ReLayNet, and 3D U-Net is described as follows. Table 2 shows the values of the three metrics for the segmentation of SRF and PED lesions. For the segmentation of SRF, the ASPP demonstrates significantly superior performance compared with ReLayNet and 3D U-Net. The DSC of ASPP is 0.8447, which is over 0.5472 and 0.8060 in ReLayNet and 3D U-Net by a large margin. The advantages of ASPP can be also observed in the metrics of TPVF, where ASPP achieves higher scores than ReLayNet and 3D U-Net. The 3D U-Net achieves the highest PPV score which is 0.8954 among the three models. For the segmentation of PED, the DSC score of ASPP is 0.6370, which is still higher than 0.5820 and 0.6131 in ReLayNet and 3D U-Net. For the TPVF metric, ASPP substantially exceeds ReLayNet and 3D U-Net where the values for ASPP, ReLayNet, and 3D U-Net are 0.8341, 0.6224, and 0.7785, respectively. The higher score demonstrates that ASPP can recognize more PED lesions compared with the other two models. However, the PPV in ASPP is 0.5153, which is inferior to 0.5464 in ReLayNet. The lower PPV score in ASPP indicates that the positive samples predicted by ASPP are more prone to be incorrect than those in ReLayNet. In terms of ASPP, the low PPV score also causes the DSC score in the segmentation of PED to be much lower than that in SRF, because the DSC is a balanced metric that is determined by TPVF and PPV simultaneously. These quantitative experimental results indicate that ASPP performs much better than ReLayNet and 3D U-Net on the SRF and PED segmentation tasks. To further validate the effectiveness of ASPP, we also carry out experiment on the fluid segmentation task of Duke dataset (Chiu et al., 2015) for which ReLayNet reported their results. The split of training and test dataset is kept the same with the one in (Roy et al., 2017). The DSC, TPVF, and PPV scores of ASPP are 0.8025 ± 0.04, 0.8619 ± 0.08, and 0.7490 ± 0.02, respectively. The DSC score of ASPP is higher than the reported 0.77 of ReLayNet.

In the following, we qualitatively compare the segmentation results of ASPP, 3D U-Net, and ReLayNet. The groundtruth together with the prediction of ASPP, 3D U-Net, and ReLayNet of five OCT slices are shown in Fig. 6. It can be seen that both the three method accurately segment the SRF lesions in the first slice. For the second slice, ASPP identifies the majority of the SRF lesions, except the precise segmentation of the anomalous boundaries, which is mainly caused by the built-in transformation invariance ability in deep convolutional neural networks. Both the 3D U-Net and ReLayNet only identify a small part of the lesion. The same result can be observed in the third slice. In the fourth slice, the segmentation results of ASPP and 3D U-Net are closer to the groundtruth than those of ReLayNet. Moreover, ReLayNet misclassifies a small portion of PED to SRF. Both the three methods recognize the majority of the SRF lesions in the fifth slice. However, ReLayNet fail to recognize the PED lesions.

The superiority of ASPP can be attributed to the following two reasons. First, a key factor in the segmentation task is the effective features that can fully represent lesions. Note that the input of ASPP is the highly abstract features produced from ResNet50. Compared with ResNet50, the encoder network in both ReLayNet and 3D U-Net have much shallow depth. Second, the multiple atrous convolutions in ASPP can capture the features in multiple scales, which is favorable in macular edema segmentation tasks. However, the relatively shallow depth of the encoder network and the fixed size of convolution kernel in ReLayNet and 3D U-Net may impede the model's ability to extract features from lesions of various shapes and sizes.

Table 3
Performance comparison between ASPP and sASPP with different configurations.

Model	SRF			PED		
	DSC	TPVF	PPV	DSC	TPVF	PPV
ASPP	0.8447	0.8214	0.8693	0.637	**0.8341**	0.5153
sASPP ($p\uparrow$)	0.8716	0.8648	0.8785	0.6837	0.6705	**0.6975**
sASPP ($p\downarrow$)	0.8166	0.7557	**0.8881**	0.7189	0.8255	0.6368
sASPP ($p = 0.8$)	**0.8759**	**0.8799**	0.8719	**0.7371**	0.8147	0.6729

4.2.2. Stochastic atrous spatial pyramid pooling

The performance comparison between ASPP and sASPP is described as follows. Table 3 shows the quantitative segmentation results of ASPP and sASPP. As can be seen in the table, sASPP ($p\uparrow$) achieves higher scores than ASPP in the segmentation of SRF. sASPP ($p\uparrow$) also obtains a higher DSC score in the segmentation of PED. However, this superiority mainly benefits from the higher PPV score, where the values in sASPP ($p\uparrow$) and ASPP are 0.6975 and 0.5153, respectively. In terms of the TPVF, the value in sASPP ($p\uparrow$) is 0.6705, which is much lower than 0.8341 in ASPP. The performance of sASPP ($p\downarrow$) is also evaluated, which is shown in the third line of Table 3. In contrary to sASPP (↑), sASPP (↓) achieves a lower DSC score compared with ASPP in the segmentation of SRF. For the segmentation of PED, sASPP ($p\downarrow$) achieves a higher DSC score compared with both ASPP and sASPP ($p\uparrow$).

Note that sASPP randomly drop the feature maps produced from the atrous convolutions during training phase, where parameter p denotes the retained probability. A larger value of p indicates that the corresponding atrous convolution plays a more important role. For sASPP ($p\uparrow$) and sASPP ($p\downarrow$), the atrous convolution with large and small fields of view dominate the segmentation results. Based on the DSC scores of ASPP and sASPP with linear variation, it can be observed that the large and small fields of view matter in the segmentation of SRF and PED lesions, respectively. To further verify the effectiveness of parameter p, all the retained probabilities of the feature maps are set to 0.8 and the results are shown in the last row of Table 3. We find that sASPP ($p = 0.8$) achieves the highest DSC score among all the models in both segmentations of SRF and PED.

A qualitative comparison of the convergence speed between ASPP and sASPP is also performed. The training and test loss in each epoch are shown in Fig. 7. It can be seen that the training loss of ASPP decreases the fastest among the three models, whereas sASPP ($p\uparrow$) decreases slightly faster than sASPP ($p = 0.8$). Note that the training loss in sASPP oscillates much more than that in ASPP, which is partially caused by sASPP's adaption to the various possible combinations of feature maps in the training phase. Moreover, contrary convergence results in terms of test loss can be observed. The order of methods with respect to the test errors, from

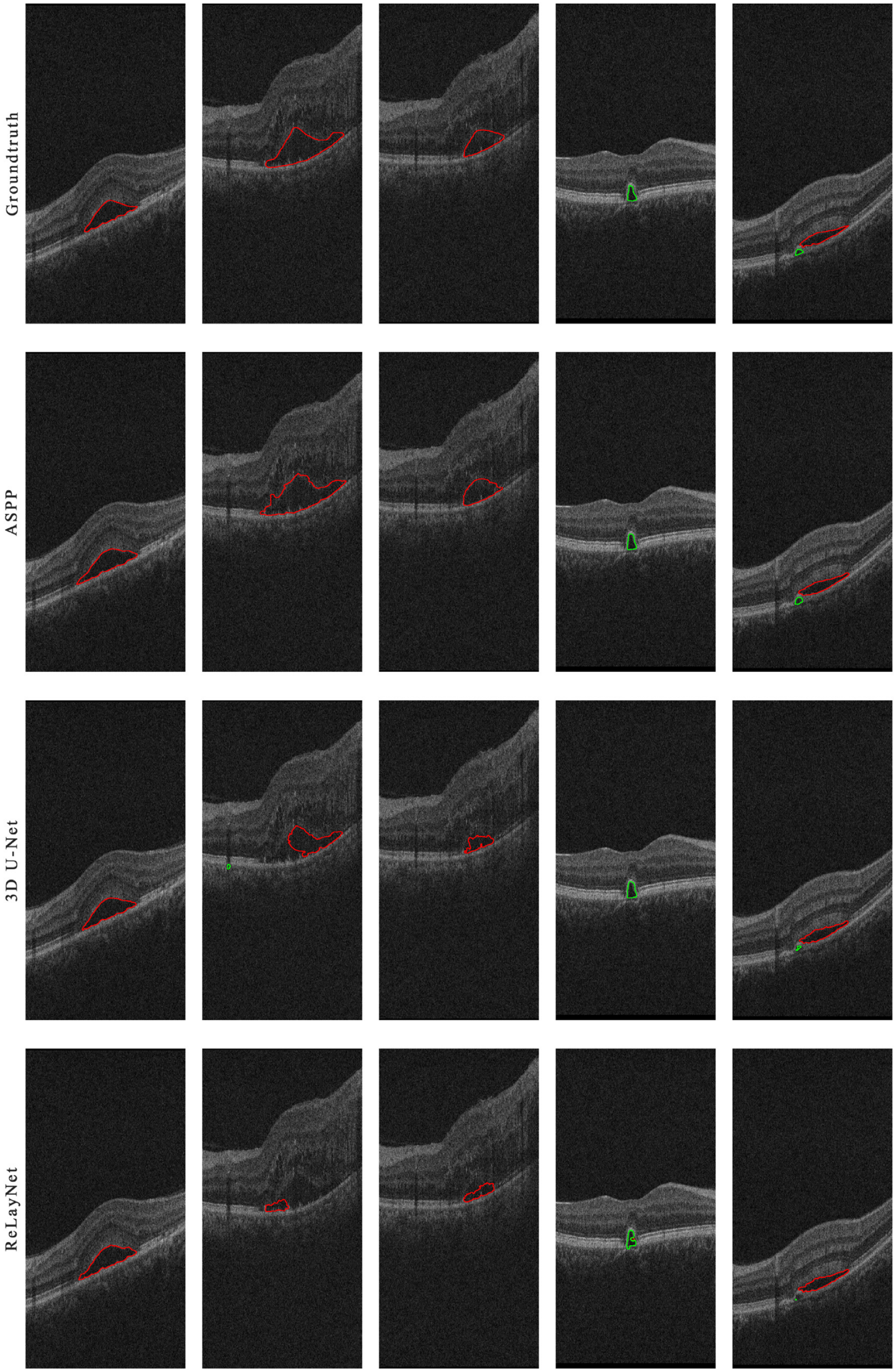

Fig. 6. Qualitative comparison between ASPP, 3D U-Net, and ReLayNet. The four rows denote the groundtruth, prediction results of ASPP, 3D U-Net, and ReLayNet. Each column represents a specific slice in OCT.

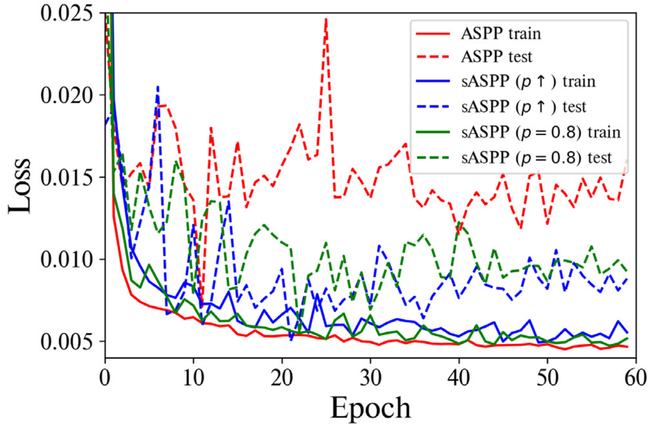

Fig. 7. Comparison of training and test loss between ASPP and sASPP.

Table 4
DSC score of ASPP and sASPP with different combination of atrous convolutions. A tick denotes the involvement of atrous convolution.

Model	r = 1	r = 6	r = 12	r = 18	DSC	
					SRF	PED
ASPP	√				0.5812	0.5745
ASPP	√	√			0.7862	0.6674
ASPP	√	√	√		0.8312	0.6525
ASPP	√	√	√	√	0.8447	0.6370
sASPP (p↑)	√				0.0051	0.0134
sASPP (p↑)	√	√			0.6192	0.0364
sASPP (p↑)	√	√	√		0.8351	0.4287
sASPP (p↑)	√	√	√	√	0.8716	0.6837
sASPP (p↓)	√				0.6886	0.3482
sASPP (p↓)	√	√			0.7867	0.7032
sASPP (p↓)	√	√	√		0.7978	0.7303
sASPP (p↓)	√	√	√	√	0.8166	0.7189
sASPP ($p = 0.8$)	√				0.4164	0.0115
sASPP ($p = 0.8$)	√	√			0.8131	0.5120
sASPP ($p = 0.8$)	√	√	√		0.8639	0.7189
sASPP ($p = 0.8$)	√	√	√	√	0.8759	0.7371

low to high, is sASPP (p↑), sASPP ($p = 0.8$), and ASPP. Compared with ASPP, sASPP achieves lower test errors whereas the training errors also decreases slower. In terms of the gap between training and test losses, it is clear to observe that the one in sASPP is much smaller than that of ASPP. These convergence results demonstrate that the stochastic drop operation is an effective regularization method that can alleviate the overfitting problem in the segmentation task.

The Bland–Altman plots which measure the limits of agreement of SRF and PED are shown in Fig. 8. The voxel size of the OCT data is $12 \times 47 \times 1.95$ μm^3. It is noted that the volume of SRF is much larger than that of PED in the dataset. It can be seen that the 95% limits of agreement between sASPP and groundtruth is narrower than that between ASPP and groundtruth in both SRF and PED. Moreover, the average discrepancy between sASPP and groundtruth in the lesion of SRF is −0.0027, which is smaller than 0.0307 between ASPP and groundtruth in terms of the absolute value. For the lesion of PED, the absolute average discrepancy between sASPP and groundtruth is larger than that between ASPP

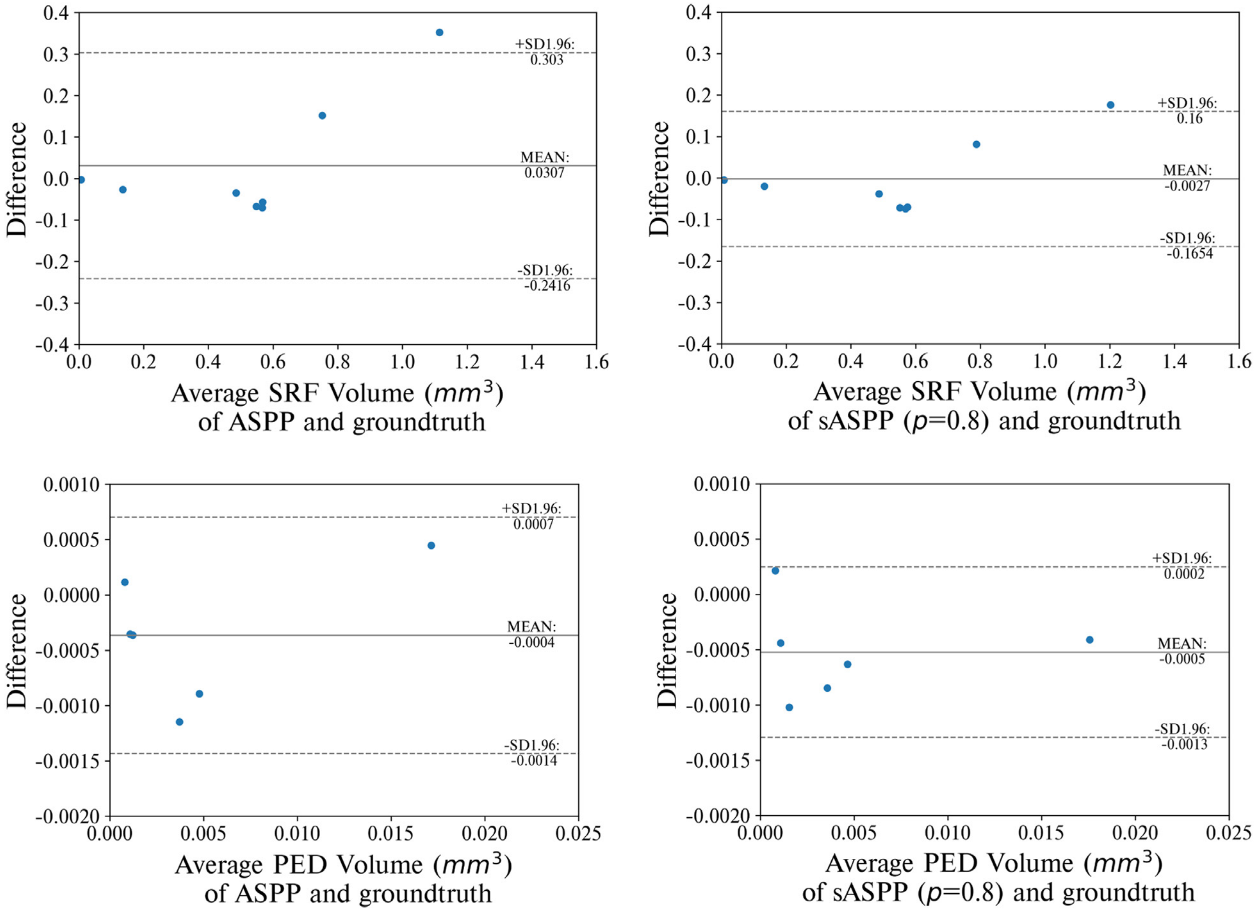

Fig. 8. The Bland-Altman plots which measure the limits of agreement of SRF and PED.

Table 5
Performance comparison between ASPP and sASPP on 3DIRCADb.

Model	Liver			Tumor		
	DSC	TPVF	PPV	DSC	TPVF	PPV
ASPP	0.9432 ± 0.01	0.9591 ± 0.02	**0.9278** ± 0.01	0.6258 ± 0.01	0.4911 ± 0.13	**0.8624** ± 0.11
sASPP ($p = .8$)	**0.9486** ± 0.01	**0.9706** ± 0.02	0.9275 ± 0.02	**0.6774** ± 0.008	**0.5714** ± 0.09	0.8314 ± 0.11

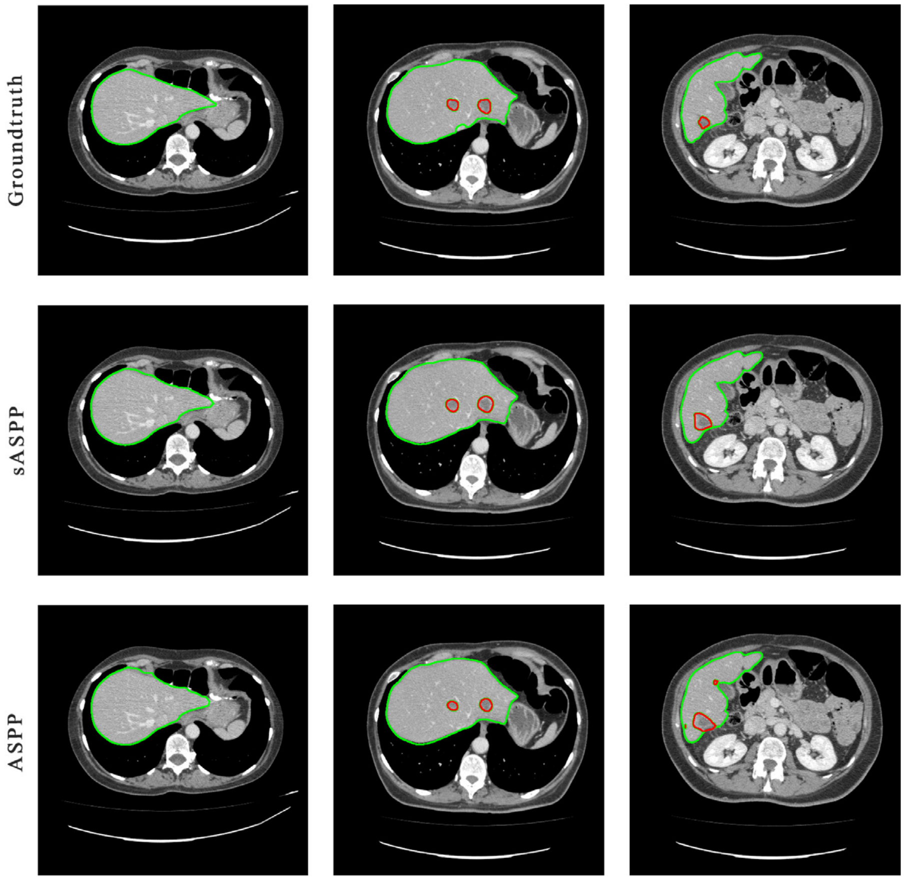

Fig. 9. Qualitative comparison between sASPP and ASPP. The three rows denote the groundtruth, prediction results of sASPP, and ASPP. Each column represents a specific scan in a CT volume. The red and green lines represent the edges of livers and tumors.

and groundtruth. It might be caused by the lower TPVF value of sASPP when compared with ASPP.

4.2.3. Analytical experiment

To empirically validate the importance of atrous convolutions in ASPP and sASPP, control experiments are carried out using gradually enabled atrous convolutions. The DSC scores are shown in Table 4. For ASPP, the DSC score of the SRF monotonically increasing when more atrous convolutions are involved. This uptrend for SRF can be also observed in the model of sASPP with different settings of p, which indicates all the four components in ASPP contribute to the segmentation of SRF. However, these contributions are unequally distributed. It can be seen that the increment of the DSC score caused by the introduction of $r = 18$ is less than that for $r = 6$. For sASPP ($p\uparrow$) and sASPP ($p = 0.8$), the DSC score is very low when only using atrous convolutions with $r = 1$ or $r = 6$, which is mainly caused by the intensity scale mechanism during the test phase.

For ASPP in the segmentation of PED, the model achieves the highest DSC score when atrous convolutions with $r = 1$ and $r = 6$ are involved. Introducing atrous convolution with a large rate (e.g., $r = 12$ or $r = 16$) in ASPP decreases the DSC score. This negative impact is alleviated in sASPP, where all DSC scores in the segmentation of PED monotonically increasing, except for $r = 18$ in sASPP ($p\downarrow$). The above results demonstrate that sASPP can use the atrous convolutions with multiple rates better than ASPP.

4.2.4. Comparison on 3DIRCADb dataset

To further validate the generalization of the proposed method, 3DIRCADb (Soler et al., 2010) dataset which contains 3D CT-scans of livers and lesions, is further used to conduct experiments. The dataset includes 20 volumes where 15 of them contain hepatic tumors. Among those 15 volumes, 5 and 10 volumes are used as test and training dataset, respectively. Two tasks are performed, including the segmentation of livers and tumors. Table 5 shows the segmentation performance of ASPP and sASPP ($p = 0.8$). For the segmentation of liver, it can be seen that the sASPP achieves higher DSC and TPVF scores compared with ASPP. The superiority of sASPP can be also observed in the segmentation of tumors in terms of the DSC and TPVF scores. However, the PPV score of ASPP in the segmentation of tumor is higher than that of sASPP. Furthermore, Fig. 9 shows the qualitative comparison of segmentation results between sASPP and ASPP. For the first two scans, it can be seen that both the ASPP and sASPP have accurately segment the livers and tumors. For the third scan, the ASPP wrongly segment a small part of normal tissues into tumors, where the sASPP segment correctly. These results further demonstrate the effectiveness and robustness of the proposed method.

5. Conclusion

In this paper, an encoder-decoder model together with ASPP is used to segment SRF and PED lesions in OCT images. The encoder is based on ResNet50, which has a large capacity to learn to extract high-level features of lesions from the raw OCT image. Then the features are fed into ASPP, which is composed of multiple atrous convolutions with different rates. Benefiting from the advantages of atrous convolution, ASPP effectively process the features by applying various fields of view without increasing the computational cost and learnable parameters. The experimental results demonstrate that ASPP is favorable in the macular edema segmentation task and substantially outperformed the state-of-the-art model.

A novel module called sASPP, which combines randomness with ASPP is further proposed. sASPP stochastically drop the feature maps produced by the atrous convolutions in ASPP during the training phase. The retained probability of each feature map is equivalent to its contribution, where the larger probability represents the feature map is more important in the segmentation. During the test phase, all the feature maps are retained, with their intensity scaled by the retention probability during the training phase. This stochastic operation is an effective ensemble mechanism to prevent the co-adaptation of multiple atrous convolutions in ASPP. The results of the control experiments indicate that the various atrous convolutions are nonequivalent in the segmentation of SRF and PED. The atrous convolutions with large fields of view are more preferable in the segmentation of SRF than that of PED. Compared with ASPP, the proposed sASPP further improves the DSC score while maintaining a significantly lower error on the test dataset. Experiments on 3DIRCADb dataset also validates the superiority of the proposed sASPP. In future studies, we intent to apply the proposed method to more types of macular edema related lesions. Additionally, learning mechanisms will be integrated into sASPP to better adjust the importance of the multiple atrous convolutions.

Conflict of interest

No other relationships/conditions/circumstances that present a potential conflict of interest.

Acknowledgements

This work was supported by the National Natural Science Foundation of China [grant 61432012 and U1435213].

References

Ahlers, C., Simader, C., Geitzenauer, W., Stock, G., Stetson, P., Dastmalchi, S., Schmidt-Erfurth, U., 2008. Automatic segmentation in three-dimensional analysis of fibrovascular pigmentepithelial detachment using high-definition optical coherence tomography. Br. J. Ophthalmol. 92 (2), 197–203.

Anthimopoulos, M., Christodoulidis, S., Ebner, L., Christe, A., Mougiakakou, S., 2016. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. IEEE Trans. Med. Imaging 35 (5), 1207–1216.

Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. Segnet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach.Intell. (12) 2481–2495.

Carneiro, G., Nascimento, J., Bradley, A.P., 2017. Automated analysis of unregistered multi-view mammograms with deep learning. IEEE Trans. Med. Imaging 36 (11), 2355–2365.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2018a. Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Trans. Pattern Anal. Mach.Intell. 40 (4), 834–848.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018b. Encoder-decoder with atrous separable convolution for semantic image segmentation. arXiv preprint arXiv:1802.02611.

Chen, Q., Leng, T., Zheng, L., Kutzscher, L., Ma, J., de Sisternes, L., Rubin, D.L., 2013. Automated drusen segmentation and quantification in SD-OCT images. Med. Image Anal. 17 (8), 1058–1072.

Chen, X., Niemeijer, M., Zhang, L., Lee, K., Abràmoff, M.D., Sonka, M., 2012. Three-dimensional segmentation of fluid-associated abnormalities in retinal oct: probability constrained graph-search-graph-cut. IEEE Trans. Med. Imaging 31 (8), 1521–1531.

Chiu, S.J., Allingham, M.J., Mettu, P.S., Cousins, S.W., Izatt, J.A., Farsiu, S., 2015. Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema. Biomed. Optics Express 6 (4), 1172–1194.

Chiu, S.J., Li, X.T., Nicholas, P., Toth, C.A., Izatt, J.A., Farsiu, S., 2010. Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation. Optics Express 18 (18), 19413–19428.

Coscas, G., Cunhavaz, J., Soubrane, G., 2010. Macular edema: definition and basic concepts. Dev. Ophthalmol. 47, 1–9.

De Fauw, J., Ledsam, J.R., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., Askham, H., Glorot, X., O'Donoghue, B., Visentin, D., et al., 2018. Clinically applicable deep learning for diagnosis and referral in retinal disease. Nature Med. 24 (9), 1342.

Drexler, W., Sattmann, H., Hermann, B., Ko, T.H., Stur, M., Unterhuber, A., Scholda, C., Findl, O., Wirtitsch, M., Fujimoto, J.G., et al., 2003. Enhanced visualization of macular pathology with the use of ultrahigh-resolution optical coherence tomography. Arch. Ophthalmol. 121 (5), 695–706.

Dufour, P.A., Ceklic, L., Abdillahi, H., Schroder, S., De Dzanet, S., Wolf-Schnurrbusch, U., Kowal, J., 2013. Graph-based multi-surface segmentation of oct data using trained hard and soft constraints. IEEE Trans. Med. Imaging 32 (3), 531–543.

Farabet, C., Couprie, C., Najman, L., Lecun, Y., 2013. Learning hierarchical features for scene labeling. IEEE Trans. Pattern Anal. Mach.Intelligence 35 (8), 1915–1929.

Garvin, M.K., Abramoff, M.D., Wu, X., Russell, S.R., Burns, T.L., Sonka, M., 2009. Automated 3-d intraretinal layer segmentation of macular spectral-domain optical coherence tomography images. IEEE Trans. Med. Imaging 28 (9), 1436–1447.

Glorot, X., Bordes, A., Bengio, Y., 2011. Deep sparse rectifier neural networks. In: Proceedings of the fourteenth International Conference on Artificial Intelligence and Statistics, pp. 315–323.

Goatman, K.A., 2006. A reference standard for the measurement of macular oedema. Br. J. Ophthalmol. 90 (9), 1197–1202.

Haeker, M., Sonka, M., Kardon, R., Shah, V.A., Wu, X., Abràmoff, M.D., 2007. Automated segmentation of intraretinal layers from macular optical coherence tomography images. In: Medical Imaging 2007: Image Processing, 6512. International Society for Optics and Photonics, p. 651214.

He, K., Zhang, X., Ren, S., Sun, J., 2014. Spatial pyramid pooling in deep convolutional networks for visual recognition. In: European Conference on Computer Vision, pp. 346–361.

He, K., Zhang, X., Ren, S., Sun, J., 2016a. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.

He, K., Zhang, X., Ren, S., Sun, J., 2016b. Identity mappings in deep residual networks. In: European Conference on Computer Vision. Springer, pp. 630–645.

Hu, J., Chen, Y., Zhong, J., Ju, R., Yi, Z., 2018. Automated analysis for retinopathy of prematurity by deep neural networks. IEEE Trans. Med. Imaging doi:10.1109/TMI.2018.2863562.

Huang, D., Swanson, E.A., Lin, C.P., Schuman, J.S., Stinson, W.G., Chang, W., Hee, M.R., Flotte, T., Gregory, K., Puliafito, C.A., et al., 1991. Optical coherence tomography. Science 254 (5035), 1178–1181.

Huang, G., Sun, Y., Liu, Z., Sedra, D., Weinberger, K.Q., 2016. Deep networks with stochastic depth. In: European Conference on Computer Vision, pp. 646–661.

Ioffe, S., Szegedy, C., 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning, pp. 448–456.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105.

Lang, A., Carass, A., Swingle, E.K., Al-Louzi, O., Bhargava, P., Saidha, S., Ying, H.S., Calabresi, P.A., Prince, J.L., 2015. Automatic segmentation of microcystic macular edema in oct. Biomed. Optics Express 6 (1), 155–169.

LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86 (11), 2278–2324.

Lin, G., Shen, C., Hengel, A. V. D., Reid, I., 2015. Efficient piecewise training of deep structured models for semantic segmentation. arXiv preprint arXiv:1504.01013.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440.

Papandreou, G., Kokkinos, I., Savalle, P.-A., 2015. Modeling local and global deformations in deep learning: epitomic convolution, multiple instance learning, and sliding window detection. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 390–399.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A., 2017. Automatic differentiation in pytorch. Advances in Neural Information Processing Systems.

Penha, F.M., Rosenfeld, P.J., Gregori, G., Falcão, M., Yehoshua, Z., Wang, F., Feuer, W.J., 2012. Quantitative imaging of retinal pigment epithelial detachments using spectral-domain optical coherence tomography. Am. J. Ophthalmol. 153 (3), 515–523.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 234–241.

Roy, A.G., Conjeti, S., Karri, S.P.K., Sheet, D., Katouzian, A., Wachinger, C., Nassir, N., 2017. Relaynet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks. Biomed. Optics Express 8 (8), 3627–3642.

Soler, L., Hostettler, A., Agnus, V., Charnoz, A., Fasquel, J., Moreau, J., Osswald, A., Bouhadjar, M., Marescaux, J., 2010. 3d image reconstruction for comparison of algorithm database: a patient specific anatomical and medical image database.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15 (1), 1929–1958.

Tranos, P.G., Wickremasinghe, S.S., Stangos, N.T., Topouzis, F., Tsinopoulos, I., Pavesio, C.E., 2004. Macular edema. Surv. Ophthalmol. 49 (5), 470–490.

Trichonas, G., Kaiser, P.K., 2014. Optical coherence tomography imaging of macular oedema. Br. J. Ophthalmol. 98 (Suppl 2), ii24–ii29.

Udupa, J.K., Leblanc, V.R., Ying, Z., Imielinska, C., Schmidt, H., Currie, L.M., Hirsch, B.E., Woodburn, J., 2006. A framework for evaluating image segmentation algorithms. Comput. Med. Imaging Graphics 30 (2), 75–87.

Williams, R.J., Zipser, D., 1989. A learning algorithm for continually running fully recurrent neural networks. Neural Comput. 1 (2), 270–280.

Wolf, S., Wolf-Schnurrbusch, U., 2010. Spectral-domain optical coherence tomography use in macular diseases: a review. Ophthalmologica 224 (6), 333–340.

Xu, Y., Yan, K., Kim, J., Wang, X., Li, C., Su, L., Yu, S., Xu, X., Feng, D.D., 2017. Dual-stage deep learning framework for pigment epithelium detachment segmentation in polypoidal choroidal vasculopathy. Biomed. Optics Express 8 (9), 4061–4076.

Yi, Z., 2010. Foundations of implementing the competitive layer model by Lotka—Volterra recurrent neural networks. IEEE Trans. Neural Netw. 21 (3), 494–507.

Yi, Z., Tan, K.K., 2004. Convergence Analysis of Recurrent Neural Networks. Kluwer Academic Publishers.