Dynamic Gaussian Convolution for Retinal Optical Coherence Tomography Segmentation

Ziyun Yang
Department of Biomedical Engineering
Duke University
Durham, NC 27707
ziyun.yang@duke.edu

Xiaorui Peng
Department of Biomedical Engineering
Duke University
Durham, NC 27707
xiaorui.peng@duke.edu

Zhanghao Yang
Department of Biomedical Engineering
Duke University
Durham, NC 27707
zhanghao.yang@duke.edu

Abstract

Increasing the imaging equality of spectral-domain optical coherence tomography (SD-OCT) is a key factor to expand OCT-relevant application, such as OCT segmentation and further diagnosis. In this paper, the method of denoising the noisy detected image by adding a trainable Gaussian kernel as a denoising physical layer(named dynamic Gaussian convolution) is investigated. Combined with traditional segmentation network, we showed that using our dynamic Gaussian convolution as physical layer is better than traditional convolution layer in both segmentation and denoising task. Multiple factors like noise levels, role of physical layer, training strategy are also discussed.

Keywords: Frequency-domain optical coherence tomography; Retinal layer segmentation; Physical layer

1 Introduction

Spectral-domain optical coherence tomography (SD-OCT) is a widely used non-invasive imaging modality for the diagnosis of soft tissues. Especially, OCT is the preferred modality for retinal imaging, for its high resolution on cross-sectional tissue imaging. [5] The segmentation of retinal OCT is extremely important for retinal diagnosis. Our general aim in this project is to improve the layers segmentation accuracy of retinal OCT.

Meanwhile, if we consider the real-world OCT equipment, the task would become harder if the detector of the OCT has noise such as background noise or thermal noise. Thus, in order to optimize the physical performance of our imaging system, we will first simulate a Gaussian noise and apply it to our OCT dataset. Then we will add a physical layer aiming to reduce the noise and therefore optimize the performance of the OCT segmentation network. This physical layer(the denoising filter) would be simulated using a convolution layer/filter at first(which may be changed as we go deeper) right after the input and connected by a segmentation network. In a conclusion, in our neural network, we will optimize both hardware performance (denoising the detector noise) of the OCT system and the performance of the neural network which is focused on the expected segmentation task.

2 Related Work

2.1 Deep Biomedical Image Segmentation

Image segmentation aims to classify every pixel in a image. Correctly segment the biomedical images is very important for biomedical detection and diagnosis. Recently, deep learning based image segmentation becomes the main trend of image segmentation.the segmentation model of deep learning usually contains two parts: the down-sampling feature extraction path and the up-sampling size recovering path. This structure was first proposed by Long et al. in [3]. Further, Ronneberger et al.[4] proposed a novel 'U' shape network called U-Net pipeline for biomedical data segmentation. The U-Net then became a powerful baseline of medical image segmentation and have been widely applied in different retinal OCT segmentation tasks. In 2018, Zhou et al.[7] proposed a nested U-Net structure which has nested, denser skip connections for medical image segmentation. Alom et al.[1] proposed a recurrent residual U-Net called R2U-Net which combines residual and recurrent modules with vanilla U-Net.

2.2 Retinal OCT Segmentation

For retinal OCT segmentation, there are also many deep learning-based methods proposed by researchers. Based on U-Net, Roy et al. proposed Relaynet [5] for both retinal layer OCT segmentation and fluid region segmentation by changing the kernel sizes and up-sampling operation in U-Net. This work is the first end-to-end CNN model for retinal layer and fluid region segmentation that does not require extra prior knowledge and pre-processing operation. Sui et al. proposed a deep learning model to learn the graph-edge weight to further segment the choroid regions in retinal OCT [6].

3 Methods

In order to learn a better denoising physical lens from the deep learning network, we introduced a denoising physical layer called Dynamic Gaussian Convolution. We will add this layer before Relaynet [5].

3.1 Dynamic Gaussian Convolution

Gaussian filter is a basic noise filter for Gaussian noise removal. In 2-D discrete field, the Gaussian filter is defined as:

$$g(x,y) = \frac{1}{2\pi\sigma_f^2} * e^{-\frac{x^2+y^2}{2\sigma_f^2}} \tag{1}$$

where x, y are the coordinates of filter entries and σ_f is the standard deviation, describing the level of smoothing. However, traditional Gaussian filter has a fixed σ_f and is not learnable. Thus, in order to built a trainable Gaussian filter, we first constrained the convolution kernel with a Gaussian distribution. Then we set the value of σ_f as a parameter in PyTorch. Thus, the σ_f can be learned with the following network. We name this structure dynamic Gaussian convolution.

3.2 Relaynet with dynamic Gaussian filter

After the physical layer, our next segmentation network will use the same structure showed in Relaynet. Generally, Relaynet has a U shape structure and uses un-pooling instead of transpose convolution for up sampling, comparing to U-Net. Together with Relaynet, we built our model for this project as figure 1.

4 Experiment Details

4.1 Data source(s)

We used the human retinal OCT dataset on the Duke SD-OCT publicly available dataset for Diabetic Macular Edema (DME) patients.[2] The dataset consists of 110 annotated SD-OCT B-scan images

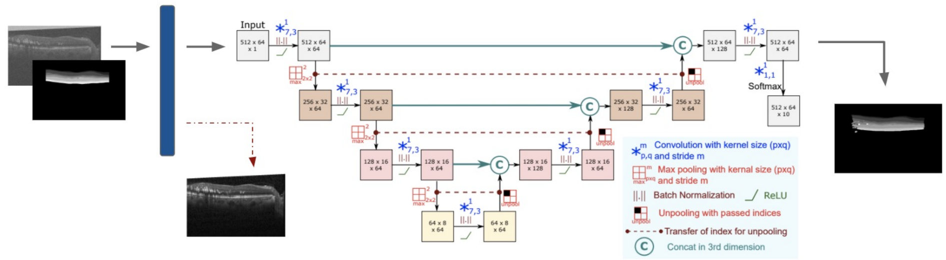

Figure 1: Network structure. Combined the dynamic Gaussian convolution with the Relaynet[5].

of size 512×740 obtained from 10 patients suffering from DME (11 B-scans per patient). The 11 B-scans per patient were annotated centered at the fovea and 5 frames on either side of the fovea (foveal slice and scans laterally acquired at $\pm 2, \pm 5, \pm 10, \pm 15$ and ± 20 from the foveal slice). There are 11 segmentation classes for each image. Two expert clinicians annotated these 110 B-scans for the retinal layers and fluid regions.

4.2 Experiment setting

4.2.1 Hyperparameters

All the experiments are coded on PyTorch. We used Adam optimizer and a gradually decreased learning strategy for learning rate, with the initial value 0.005. We used a weighted cross entropy loss sums a dice loss as our loss function. For evaluation, we use dice coefficient for segmentation result evaluation and we used the peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM) for denoising evaluation.

4.2.2 Data preprocessing

We randomly sampled 9 patients for training and 1 patient for testing each time and we repeated this operation for four times. Thus, we generated four different subsets. Every experiment is conducted on all of these four subset and the average results are concerned. For every input image, we randomly add Gaussian noise on them every time. Thus we created the noisy input with simulated detector noise (Gaussian noise).

4.2.3 Pretraining strategy

In order to get the optimized physical layer and segmentation result. We used a pretraining strategy, which is, for every single experiment, we did several different sub experiments. We first pretrained both the Relaynet and physical layer. After this, based on the optimized model of last step, we froze the Relaynet and trained the network again. This means in the second step, only the physical layer was updated. Next step we froze the physical layer and unfroze the Relaynet and trained again based on model from last step. We then repeated this for one more time. We did these steps in order to find the local optima for both the Relaynet and physical layer.

5 Results and Analysis

5.1 Ablation Studies: Noise level

In order to evaluate the robustness of our model, we also did some ablation studies.

First, we studied the effect of the noise levels. Figure 2 shows the result of the study. From left to right, we gradually increased the noise level (σ, i.e., the standard derivation of noise) from 0 to 40. We used a Relaynet as the baseline, the dotted line here. From the figure, as the images getting noisier, the gap between Relaynet and the Gaussian physical layer becomes larger. which means, the noisier the signal is, we will have better performance of our physical layer.

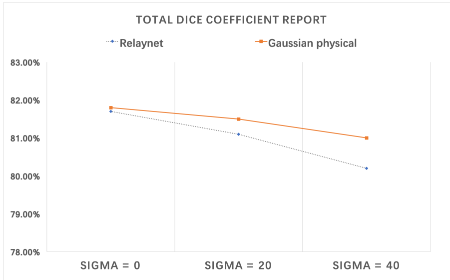

Figure 2: Noise Levels

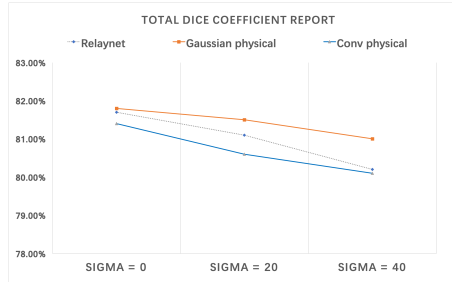

Figure 3: Dynamic Gaussian Conv vs. Conv2D.

5.2 Ablation Studies: selection of the physical layer

Then we studied the selection of the physical layer. We used Gaussian convolution and traditional convolution and did this comparison. Gaussian convolution, which is the orange line showed in figure 3, outperformed the traditional one at any noise level.

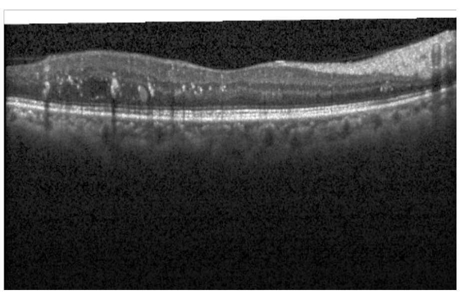

Figure 4: Clean image

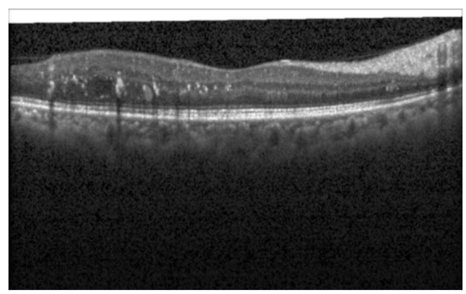

Figure 5: Dynamic Gaussian Convolution

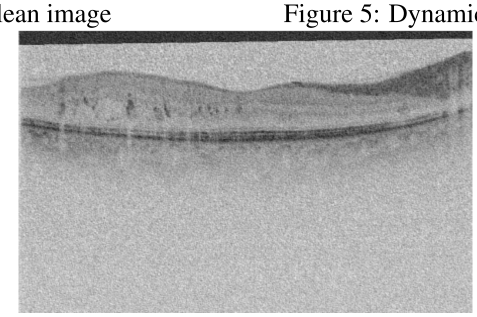

Figure 6: Traditional Convolution

We also plot the outputs of our physical layer. As we can see, the output of our dynamic Gaussian convolution in figure 5 is nearly the same with the clean image in figure 4, which means our dynamic algorithm recovered the image at a satisfying level. Our model denoised the image while the traditional convolution did something undesired, as showed in the figure 6.

5.3 Ablation Studies: Pretraining Mechanism

Then, we studied how the pretraining process performs. As mentioned before, we pretrained the network at first and then fix different parts to get their optimized state. The pretrained one has a way better result as showed in figure 7.

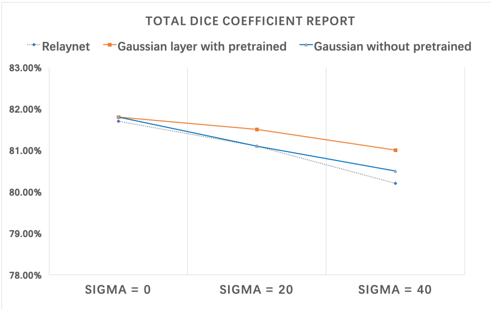

Figure 7: Ablation Study: Importance of Pretraining.

5.4 Performance of Denoising

Last, we would like to show the performance of the physical layer. After the training, we got the learned value of σ_f and the corresponding Gaussian filter. Here, we will illustrate two of the learned σ_f for Gaussian filter at noise level σ = 20 and 40, respectively. As showed in Table 1, as the image get noisier, the learned σ_f is larger.

In order to show the visualization results of denoising, here we use the case when noise level σ=40 as an example. To evaluate the denoising results, we used peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM), which is a metric to describe the structural similarity, as out metrics.

Showed in figure 8-11 and table 2, the PSNR value increased from 16.09 to 29.53 and the SSIM index of Gaussian convolution with pretrained has increased from 0.223 to 0.874 and it is very close to the original clean image.

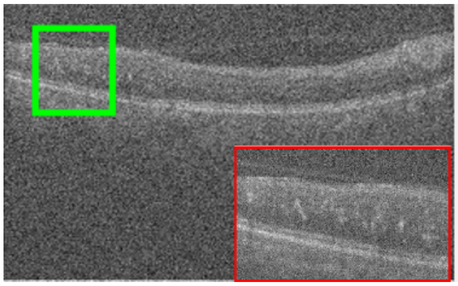

Figure 8: Noisy Input

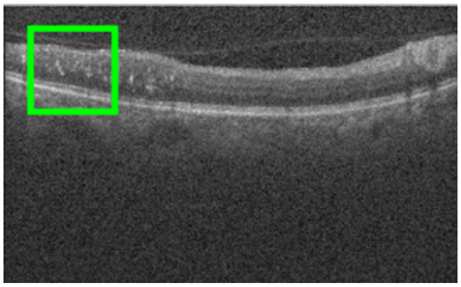

Figure 9: Gaussian without pretrained

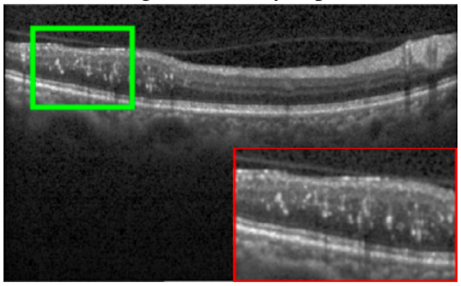

Figure 10: Gaussian with Pretrained

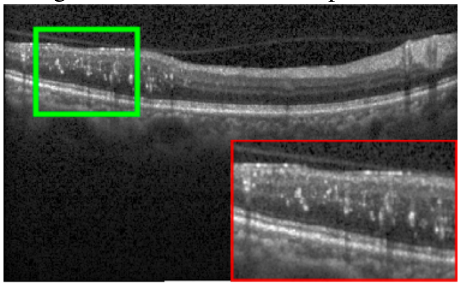

Figure 11: Original clean image

6 Future Work

First, we can try different Gaussian kernel sizes, see if the kernel sizes impact the final output. The dynamic Gaussian kernel can be combined with more network structures, such as U-Net.

Table 1: Learned σ_f value at different noise level

Noise level σ	Learned σ_f
σ =20	0.6479
σ =40	1.3539

Table 2: PSNR and SSIM result when noise level σ = 40

Name	PSNR	SSIM
Noisy Input	16.09	0.223
Gaussian without Pretrained	23.88	0.626
Gaussian with Pretrained	29.53	0.874

The structure should be tested on more noise model, for example, mixture Gaussian noise, Poisson noise, speckle noise, and see how well it can deal with the different noises.

References

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

[2] Stephanie J Chiu, Michael J Allingham, Priyatham S Mettu, Scott W Cousins, Joseph A Izatt, and Sina Farsiu. Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema. *Biomedical optics express*, 6(4):1172–1194, 2015.

[3] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. 2014.

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

[5] Abhijit Guha Roy, Sailesh Conjeti, Sri Phani Krishna Karri, Debdoot Sheet, Amin Katouzian, Christian Wachinger, and Nassir Navab. Relaynet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks. *Biomedical optics express*, 8(8):3627–3642, 2017.

[6] Xiaodan Sui, Yuanjie Zheng, Benzheng Wei, Hongsheng Bi, Jianfeng Wu, Xuemei Pan, Yilong Yin, and Shaoting Zhang. Choroid segmentation from optical coherence tomography with graph-edge weights learned from deep convolutional neural networks. *Neurocomputing*, 237:332 – 341, 2017.

[7] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. 2018.