2018 17th International Symposium on Distributed Computing and Applications for Business Engineering and Science

Mask Convolution for Filtering on Irregular-Shaped Image

Yiping Jiao
School of Automation
Southeast University, Nanjing, China
ping@seu.edu.cn

Chengqi Qian
Institute of Life Science
Southeast University, Nanjing, China
sw20117801@163.com

Shumin Fei
School of Automation
Southeast University, Nanjing, China
smfei@seu.edu.cn

***Abstract*— In this paper, we proposed the concept of mask convolution, which operates on irregular shaped foreground, and could gain improvement on the margin of the object. In particular, we prove that this operation could be done in an efficient way. The extra calculation than normal convolution only lies on mask image convolution, and one time elements-wise multiply and division. To verify the necessity of mask convolution, we use mean filtering on retina images to eliminate spots. Compared to normal convolution, the proposed method could obtain clearer image on the margin of foreground, which is beneficial for subsequent CNN-based vessel segmentation model. We also proved that this method can be extended to other convolution kernels such as Gaussian and sharpen filters.**

Keywords- ***mask convolution; convolutional neural networks; retina image; blood vessel segmentation; image processing***

I. INTRODUCTION

Recent developments in deep learning have led to many breakthroughs in image processing, speech recognition and computer aided diagnosis. Lots of efforts have been paid on structural design [1], normalization [2], as well as details such as initialization, activation function, and optimization algorithm. However, less research has been focused on convolution, which is the core of CNN. It is known that many different choices can be made when dealing with the margin of an image, for example, zero padding fill zeros outside the image to make the operation valid. If only the overlapping area is considered, it is called valid padding, which may lead to size shrinking after convolution. Another alternative strategy is to copy the nearest valid pixel while the convolution exceeds the boundary [3], which seems more natural and is commonly used. What's more, convolution is also an important technique in traditional image processing, which is widely used in edge detection and smoothing.

Traditional convolution works well when dealing with large-size rectangular shaped images, however, for those objects which have complicated boundary, it may be less effective, or loss some important information. These cases frequently happened in modern computer aided diagnosis system. For example, in retina images, we often have a circle-shaped foreground mask, known as FOV (Field of View). The rest of these image is background, which can be either black or white. If an indiscriminative operation is executed on such image, the margin of the foreground will be badly affected. Things will be worse when it comes to pathology images, due to the arbitrary foreground boundary. Compared to the tricky problem, the foreground could be recognized much easier, for example, by a simple thresholding, or an advanced CNN based tissue segmentation systems [4]. Furthermore, some models can generate class-specific segmentation masks which may be useful for pathologists[5,6]. To make use of the mask, and operate without background pixels, this induces a mask operation problem.

In this paper, we will consider mask convolution, which is the most general operation employed in many scenes, such as Gaussian smooth and convolutional neural network. We will give an example of eliminating locally uneven illumination in DRIVE dataset[7]. An example is showed as Fig.1, where obvious uneven color could be observed in the left bottom corner of the right image. It can be very hard for an automated recognition system to deal with the tedious situation directly. We will use a strategy which eliminate the local mean, to make the color be even in the whole image. However, the areas near the margin should be tackled carefully, avoid of taking the black background into consideration, that is the reason why a mask convolution can be useful. Except for the observation of the processed images, we will also show that a significant improve could be obtained when the preprocess implemented with a subsequent CNN segmentation model. The applicability to other filters will also be discussed.

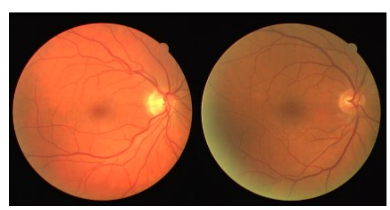


Figure 1. Uneven illumination of retina images

II. METHODOLOGY

A. Basic Conception

Considering the simplest example showed as Fig.2, in which we want to convolute the image patch with a 2x2 size kernel. The mask corresponding to the image patch is given as $m_i \in \{0,1\}$, where 1 indicates a foreground pixel and 0 for background pixel. Traditional convolution will calculate regardless of the mask, as:

978-1-5386-7445-1/18/$31.00 ©2018 IEEE
DOI 10.1109/DCABES.2018.00039

$$R_{\text{origin}} = \sum k_i p_i = k_1 p_1 + k_2 p_2 + k_3 p_3 + k_4 p_4 \tag{1}$$

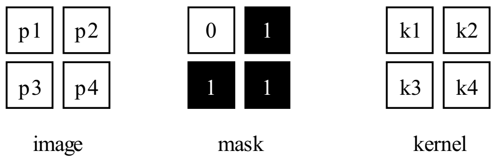


Figure 2. An example for mask convolution

If we extent the region outside the mask by zero padding, it will seem like:

$$R_{\text{zero-padding}} = \sum k_i p_i m_i = k_2 p_2 + k_3 p_3 + k_4 p_4 \tag{2}$$

However, as the mask abandons pixel p_1 , and as the proportion of k_1 is ignored, the response R will trends to be a smaller value than a patch which is fully composed of foreground pixels (under cases that k_i is nonnegative). To eliminate the difference, we claim that a multiplier coefficient should be considered, as:

$$\begin{aligned} R_{fine} &= \sum k_i p_i m_i \times \frac{\sum k_i}{\sum k_i m_i} \\ &= \left(k_2 p_2 + k_3 p_3 + k_4 p_4\right) \times \frac{k_1 + k_2 + k_3 + k_4}{k_2 + k_3 + k_4} \end{aligned} \tag{3}$$

B. *Mask Convolution*

The idea is quite intuitive and can be easily implemented by element-wise calculation, especially in mean filtering. However, the efficiency remains a problem. It is known that convolution makes use of FFT to speed up the calculation. Whether it is possible to turn the mask convolution to traditional convolution is worthy to be investigated.

By zero-padding-like convolution, we will get $R_{\text{zero-padding}}=\sum k_i p_i m_i$. If we apply the same convolution to mask map, we can obtain:

$$R_{\text{mask}} = \sum k_i m_i \tag{4}$$

Thus we can reform what we desired with convolution as:

$$R_{\text{fine}} = \frac{\left(P \times M\right) \otimes K}{M \otimes K} \times \sum k_i \tag{5}$$

And under $\sum k_i = 1$, which is a normal case in filtering, more simplified form could be obtained:

$$R_{\text{fine}} = \frac{\left(P \times M\right) \otimes K}{M \otimes K} \tag{6}$$

where $\otimes$ denotes the convolution between two matrices, multiply and division employed are element-wise operations. It is obvious that this is equivelant to normal convolution except for margin cases. But considering the amplitude and potential zero-division, the results may need to be truncated within valid range in an addition step.

We noticed that NVIDIA used a partial convolution to reconstruct images with holes[7], which is formulated as:

$$X' = W^T \left(X \otimes M\right) \frac{1}{sum\left(M\right)} + b \tag{7}$$

It is slightly different from our fomulation. Partial convolution is written in a CNN-compatible way, in which trainable weight W could include kernel K and linear coefficient $\sum k_i$. The difference lies on the denominator. Partial convolution uses the counting of the involved pixels, whereas our formulation also considered the information within the masked region.

C. *Local Illumination Normalization*

To deal with DRIVE dataset [8], we defined the illumination normalization (IN) as moving the mean pixel intensity to a predefined value within a certain window, namely:

$$\begin{aligned} I_{\text{norm}}\left(i,j\right) = \; & I_0 + I\left(i,j\right) \\ & -k \times \text{mean}\left\{I\left(t,s\right) \middle| \left(t,s\right) \in W\left(i,j\right) \cap M\right\} \end{aligned} \tag{8}$$

where I_0 is the predefined constant, $W\left(i,j\right)$ is the set of pixels within the window centered at $\left(i,j\right)$, it should be noticed that pixels outside the mask are excluded by the intersection with M .

As we proved above, it is unnecessary to implement the complex logical rule in an element-wise way. By define a filter as:

$$K = \frac{1}{k^2} \mathbf{1}_{k \times k} \tag{9}$$

where k is the window size, then the local mean can be obtained by a *M*-masked convolution with kernel K .

III. MATERIAL AND EXPERIMENT DESCRIPTION

A. *Dataset*

As a case study, we conducted following experiments based on DRIVE [8], where the aim is to distinguish blood vessels from the whole retinal images. DRIVE database were gathered from a diabetic retinopathy screening program in Netherlands. There are in total 40 images, each of which is 768×584 RGB 3-channel image. The challenge is to get segmentation of the blood vessels within the image. The images were separated into training set and test set, each of which is composed of 20 samples. More details about the dataset is available in [8] and [9].

B. *Preprocessing*

As a preprocessing approach, we considered 3 cases. In the following, "none" means without any preprocessing, while "IN without mask" used IN without considering the given mask, and "IN with mask" refers to the strategy claimed as (8). While IN is considered, the parameters involved are set as $k = 1.5$, $I_0 = 128$, and it was done in Red, Green and Blue channel independently.

C. *Segmentation Model*

To build a segmentation model, we chose a simple CNN, which is composed of 3 conv layers, together with intermediate pooling layers, and followed by fully connected layers and 2-way softmax layer. The network structure is showed as Fig.3.

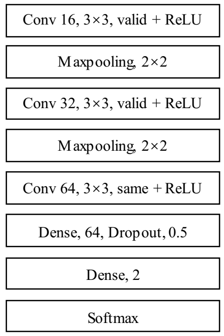


Figure 3. Structure of vessel segmentation CNN

To train the model, patches of size 25×25 is chopped from the image with/without illumination normalization, the ground-truth label of which is determined by whether the center pixel of the patch belongs to blood vessels. We generate 2000 positive and negative patches within a training image respectively, which produce 80000 training patches. 30% of the patches were chosen randomly as validation set, in case of overfitting.

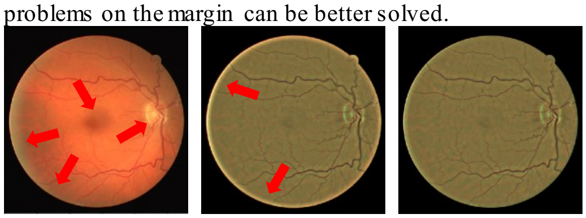
We implemented the network using Keras, whose backend is Tensorflow. We used categorical cross-entropy as loss function, and default Adam optimizer. Norm-2 regularizer was introduced in all convolutional layers and the first dense layer, whose coefficient was set as 1e-4.

In fact, finer results could be get by more complex auxiliary preprocessing or post-processing approaches, such as ZCA whitening [10], data augmentation [10], morphology operation [11], and conditional random field [12]. However, as we focused on mask convolution as a preprocessing approach, none of them were assembled, to make the results more immediate and sensitive.

IV. RESULTS

A. Illumination Normalization

Here we show the validity of IN as Fig.4. The left one is the original image without any preprocessing, it can be seen that the margin of the foreground suffers from imperfect exposure, and light or dark spots can also be found within the image, which is believed to be optic disc and cup[13] or pathological conditions. All these may affect the recognition and are marked by red arrows. In the middle we show IN without mask, where significant improvement is achieved. But the margin of the object is still abnormal, which is also marked out. In the right one, masked IN is implemented and it can be seen that the problems on the margin can be better solved.



(a) without IN (b) IN without Mask (c) IN with Mask

Figure 4. An example of Illumination Normalization

B. Vessel Segmentation

TABLE I. TRAINING ACCURACY

Preprocessing	Accuracy	
	Training Set	***Validation Set***
without IN	91.75%	90.19%
IN without mask	92.16%	89.67%
IN with mask	92.81%	89.01%

CNNs under different preprocessing were trained using the configurations described above. The performance on training set is given in Tab.1, from which only subtle different could be seen. By varying the segmentation threshold θ , the entire ROC curve on test could be obtained as Fig.5. It can be seen that, by considering the IN preprocess, the performance indeed improved, while the mask-IN achieved best results.

We take the 6th retinal image in the test set as example. It can be found that in Fig.6(a) and (b), the model suffers from obvious false-positive on the edge of the target, which are marked by arrows. While the model is implemented with mask convolution based IN, the result will be better, which is showed in in Fig.6(c).

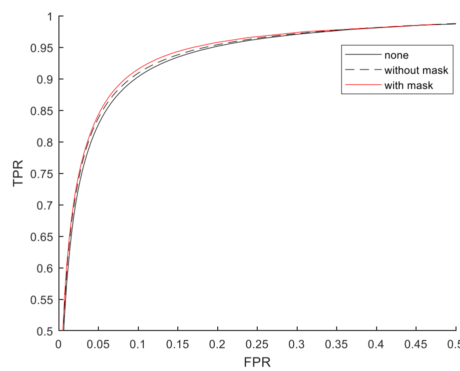
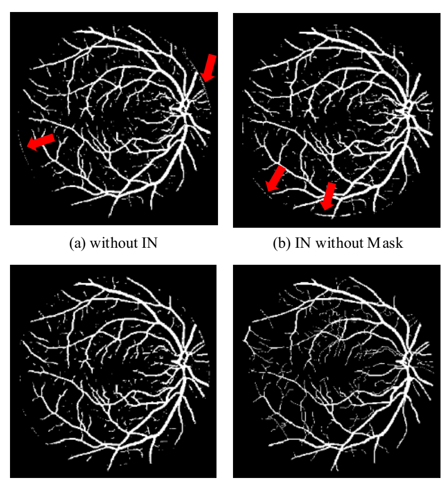


Figure 5. ROC Curves on Test Set



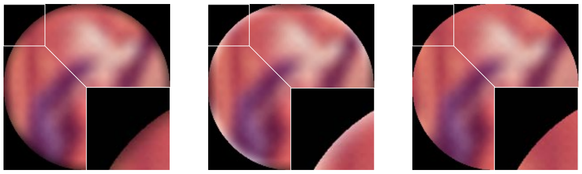
(a) without IN (b) IN without Mask

(c) IN with Mask (d) ground truth

Figure 6. Segmentation of 6th Test Image

V. DISCUSSION

Above we use mask convolution to implement mean filtering, here we will discuss the applicability for other filter kernels.



(a) On black BG (b) On white BG (c) Mask convolution

Figure 7. Convolution Using Large-size Gaussian Filter

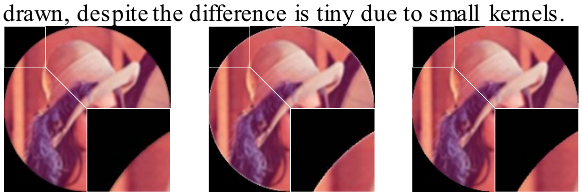
We firstly take a large-size Gaussian filter, whose parameter is $s = 50$, $\sigma = 20$, which is very similar to the above mean filter. We fill the background with black (which is equivalent to zero-padding) or white, and impose filtering towards the two images. Meanwhile, the mask convolution is also applied to the image. Results showed as Figure.7. Note that we impose masking after convolution, considering the content outside mask is meaningless.

It can be seen that the normal convolution was applied to black or white background images, the margin of the result will be affected. By making use of the mask information, the mask convolution could obtain more reasonable results.

Similarly, we also conducted the experiment on smaller Gaussian filter ($s = 10$, $\sigma = 3$) and 5x5 sharpen filter:

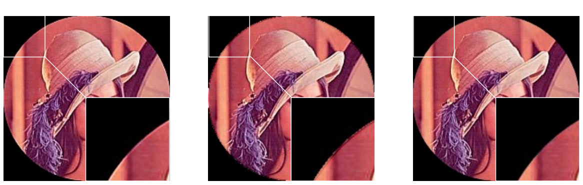
$$K = \frac{1}{8}\begin{bmatrix} -1 & -1 & -1 & -1 & -1 \\ -1 & 2 & 2 & 2 & -1 \\ -1 & 2 & 8 & 2 & -1 \\ -1 & 2 & 2 & 2 & -1 \\ -1 & -1 & -1 & -1 & -1 \end{bmatrix}$$

Results are showed as Figure.8 and Figure.9 respectively, from which the coherent conclusion could be drawn, despite the difference is tiny due to small kernels.



(a) On black BG (b) On white BG (c) Mask convolution

Figure 8. Convolution Using Small-size Gaussian Filter



(a) On black BG (b) On white BG (c) Mask convolution

Figure 9. Convolution Using Sharpen Filter

VI. CONCLUSION

From all the experiments above, we could conclude that given the mask of an object, the proposed mask convolution will perform better on the margin of the object than the traditional one. Such processing can gain profit in subsequent image analysis procedure, for example, a CNN for segmentation task in this paper. The mask convolution is computational cheap, and can be extended to other situations where convolution is employed. Whether this technique can be implemented within popular convolutional neural networks is worth of further investigation.

REFERENCES

[1] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[2] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[J]. arXiv preprint arXiv:1502.03167, 2015.

[3] Zhang H, Liu D, Xiong Z. CNN-based text image super-resolution tailored for OCR[C]//Visual Communications and Image Processing (VCIP), 2017 IEEE. IEEE, 2017: 1-4.

[4] Bándi P, van de Loo R, Intezar M, et al. Comparison of different methods for tissue segmentation in histopathological whole-slide images[C]//Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on. IEEE, 2017: 591-595.

[5] Bulten W, Hulsbergen-van de Kaa C A, van der Laak J, et al. Automated segmentation of epithelial tissue in prostatectomy slides using deep learning[C]//Medical Imaging 2018: Digital Pathology. International Society for Optics and Photonics, 2018, 10581: 105810S.

[6] de Bel T, Hermsen M, Smeets B, et al. Automatic segmentation of histopathological slides of renal tissue using deep learning[C]//Medical Imaging 2018: Digital Pathology. International Society for Optics and Photonics, 2018, 10581: 1058112.

[7] Liu G, Reda F A, Shih K J, et al. Image Inpainting for Irregular Holes Using Partial Convolutions[J]. arXiv preprint arXiv:1804.07723, 2018.

[8] Staal J, Abràmoff M D, Niemeijer M, et al. Ridge-based vessel segmentation in color images of the retina[J]. IEEE transactions on medical imaging, 2004, 23(4): 501-509.

[9] Niemeijer M, Staal J, van Ginneken B, et al. Comparative study of retinal vessel segmentation methods on a new publicly available database[C]//Medical Imaging 2004: Image Processing. International Society for Optics and Photonics, 2004, 5370: 648-657.

[10] Liskowski P, Krawiec K. Segmenting retinal blood vessels with deep neural networks[J]. IEEE transactions on medical imaging, 2016, 35(11): 2369-2380.

[11] Joshi S, Karule P T. Retinal blood vessel segmentation[J]. International Journal of Engineering and Innovative Technology (IJEIT), 2012, 1(3): 175-178.

[12] Fu H, Xu Y, Lin S, et al. Deepvessel: Retinal vessel segmentation via deep learning and conditional random field[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2016: 132-139.

[13] Sevastopolsky A, Drapak S, Kiselev K, et al. Stack-U-Net: Refinement Network for Image Segmentation on the Example of Optic Disc and Cup[J]. arXiv preprint arXiv:1804.11294, 2018.