A Deep Bottleneck U-Net Combined with Saliency Map for Classifying Diabetic Retinopathy in Fundus Images

Vo Thi Hong Tuyet[1,2,3], Nguyen Thanh Binh[1,2(✉)], Dang Thanh Tin[2,4]

[1] Faculty of Computer Science and Engineering, Ho Chi Minh City University of Technology (HCMUT), Ho Chi Minh City, Vietnam
[2] Vietnam National University Ho Chi Minh City, Ho Chi Minh City, Vietnam
[3] Faculty of Information Technology, Ho Chi Minh City Open University, Ho Chi Minh City, Vietnam
[4] Faculty of Electrical and Electronics Engineering, Ho Chi Minh City University of Technology (HCMUT), Ho Chi Minh City, Vietnam
`ntbinh@hcmut.edu.vn`

Abstract—Early detection of retinopathy plays an important role in the care of people with diabetes. Classification of diabetic retinopathy in fundus images is very challenging because the blood vessels in the retinal images are too small. Morphology of objects with multi-level saliency is the recent choice because of the activation of feature extraction. However, the challenges of the input models are very complex with the blood. The color, lighting or context can become the reasons that create the decline of the primary key for training. This paper proposes a method for classification of diabetic retinopathy using saliency and shape detection of objects based on a deep Bottleneck U-Net (DbU-Net) and support vector machines in retinal blood vessels. The proposed method includes four stages: preprocessing, feature extraction using DbU-Net, saliency prediction and classification based on the support vector machine. To evaluate this method, its results are compared to the results of the other methods by using the same datasets of STARE and DRIVE for testing with evaluation criteria such as sensitivity, specificity, and accuracy. The accuracy of the proposed method is about 97.1% in these datasets. To assess the levels of diabetes, the diagnostician must initially identify the retinal image with diabetes or not. The result of this paper may help the diagnostician to easily do this.

Keywords—Bottleneck U-Net, saliency, classification, diabetic retinopathy, retinal blood vessel

1   Introduction

The diagnosis of abnormality in medical images depends on many factors such as doctor skills, equipments, etc. Every object has a different shape, size and position in the medical image. The morphology of objects includes shape and direction of gradients. The retinal blood vessel image contains a lot of useful information for the diagnosis of disease. The analysis of signs and characteristics of the retinal blood vessels

can help doctors diagnose the disease making the patient suffering. Consequently, image segmentation and classification are very useful for the diagnosis of doctors. The gray level and salient regions are the primary key for detection.

Diabetes damages the blood vessels of all organs in the human body. It is most evident in the microvasculature. In eye, due to the damage to the retinal capillaries, it increases vascular permeability, plasma leakage into the retina causing edema. When capillaries are destroyed causing blockage, it causes retinal ischemia. The body then responds by secreting factors that stimulate the growth of new blood vessels to nourish these areas of the retina. However, these blood vessels are fragile, causing complications such as vitreous hemorrhage, fibrosis, and retinal detachment.

The retina is the light-sensitive organ, in which the macula is the center of the retina, where our vision is most delicate. The effects of diabetes on the retina make the retina severely damaged, severely affecting our visual function. Early detection of retinopathy plays an important role in the care of people with diabetes. Today, the application of machine learning to assist doctors in diagnosing diseases, especially eye-related diseases, is of interest to researchers. Currently, there are many studies applying machine learning to the diagnosis of diabetic retinopathy. This paper proposed a method for diabetic classification in retinal vessels based on the deep Bottleneck U-Net with saliency. The proposed method includes four stages: preprocessing, feature extraction using deep Bottleneck U-Net (DbU-Net), saliency prediction and classification based on support vector machines. To compare the results of the proposed method with those of the other methods, the STARE and DRIVE datasets are used for testing with evaluation criteria such as sensitivity, specificity, and accuracy.

Main contributions of this paper are as follows:

1. Proposing the deep Bottleneck U-Net (DbU-Net) to match diabetic classification in retinal vessels, the structure of U-Net adapts with multi levels of saliency and retinal blood vessels images.
2. Increasing the accuracy of the classification by DbU-Net combined with saliency prediction.

The rest of the paper is organized as the following: literature review is presented in Section 2. The proposed method for diabetic classification in retinal vessels is presented in Section 3. The experimentation and evaluation results are presented in Section 4. Finally, Section 5 is conclusions and future works.

2 Literature review

In recent times, many researchers have proposed segmentation methods for retinal images [1, 2, 3, 4, 5, 6]. Marwan [1] had applied the histogram equalization for region detection. This research used the automatic threshold for selection based on multi levels. Morphology of retinal blood vessels was proposed in [4, 5, 6] with shape and the kernel chosen for presentation. The other approaches are developed based on the surface and structures of retinal vessels for segmentation on the diabetic patients [7, 9, 10,12]. Molina [8] based on intra- and inter-structure of the relationship between

neighbor knowledge to extract features. Sundaram [11] extracted the retinal blood vessels through Hybrid segmentation. Memari [12] had taken the advantage of Fuzzy C-means clustering in a levels set for retinal blood segmentation. The matching of these algorithms was the morphology which was collected by the relationship of knowledge between the histogram/kernels and multi levels. The disadvantages of these methods are the relationship of parameters around the pixels which are unimportant. To overcome the disadvantages of morphology in segmentation, Rahebi [13] proposed the neural network depending on gray level which was applied in matrix-based features to segment retinal blood vessels. Ronneberger [14] proposed U-net to modify and extend fully convolutional network architecture for biomedical image segmentation. In these networks, the authors applied the encoder and decoder for the conversion of the structure of the input images.

Besides, the researchers had improved the neural networks with the transfer learning for segmentation and classification [15, 18, 21]. This transfer was continued with multi scales for deep learning in [16, 17, 22]. Moreover, the disease classification algorithms had proposed such as: multi-kernel support vector machine [19], CT image of coronavirus of Li [20], EfficientNet for the leaf classification [22], ultrasonography with the automated classification of liver [23], hypertension risk by photoplethysmography with deep learning [24] or convolutional neural network [26], the saliency with convolutional neural network for brain tumour classification [25], edge detection based on salient map [28], convolutional neural network based on transfer learning for diabetic [34], machine learning algorithm for diabetes classification [35], diabetic retinopathy fundus images classification based on hyperparameter tuning deep learning [36], etc. The feature extraction by the learning models could be the input or output of the segmentation process. The prediction models had given the parameters which included context, color channels, noise/blur values, etc. The author of these researches changed the number of layers in the encoder of the input model. However, the complex problem is the pre-processing which has a wide range of unclear pixels. Domain/transform is one of the pre-processing methods. The dividing of decomposition is also the condition for the number of the chosen parameters. The reason to use the domain for segmentation or classification is to smooth results from thresholding or filtering in their steps. Laplacian and dynamid are the filters in contourlet transform. Nageswara [27] had proposed this domain with an active contour model for activation of the object boundary of Brain MR image segmentation. The variety of pre-processing of deep learning is also the combination of projection context or saliency together [29, 30, 31]. Convolutional neural network with the saliency-level input for segmentation and the index prediction of the content-based medical image retrieval is shown in [32]. On the other hand, the tree structure with a graph in global context for Covid-19 lung segmentation is also presented in [33]. Abou [42] applied convolutional neural network to detect and classify normal white blood cells. Yanni [44] had used hybrid algorithm structure convolutional neural network for segmentation. As a result, the segmentation contributes to the index prediction and object detection.

Another application of deep learning is classification. The index of the kind of images is selected by feature extraction and prediction layer. The standard convolutional neural network architecture [25, 26, 43] gives a downsampling factor to the network.

Fully convolutional network [15] has 13 layers and feature maps by the encoding network. The final layer of full convolution is the FC layer with the prediction value from 0 to 1. As a result, the results of these deep learning models give the learning values to some prediction systems. On the other hand, they can be the final results of the system. Their characters are the pre-processing with the number of blocks with other size strides or windows of each layer.

3 The disease classification based on deep Bottleneck U-Net for saliency prediction

The segmentation or classification based on the feature extraction. Deep learning is a collection of layers with the stride and size kernels. The parameters will be changed in each block. The concept about block is the combination between convolution and maxpooling layers. The neural network with the tensors is height × width × the number of channels from the chosen input. The architecture of any deep learning is the model with tensors which have been arranged. These outputs are concerned as: convolution (transposed, dilated), pooling (maxpooling, average pooling, max unpooling, etc). Besides that the upsampling for the decoder can be the primary condition for feature extraction. The equation (1) is the calculating method of pooling layer where m is the number of kernels and n is the number of stride.

$$p_k^{l+1}(m,n) = max_{1 \le g \le u} x(m,(n-1)*u+g) \tag{1}$$

The connection of many convolution and pooling layers creates the multi-level features maps. In the real truth, the ReLU is a final layer of block. The equation (2) is the ReLU layer:

$$c_k^{l+1}(m,n) = ReLU\left(\sum_{g=1}^u x\left(m,n-g+\frac{u+1}{2}\right)w_k^l(g)+b_k^l\right) \tag{2}$$

where the i-th node of the l-th layer to the j-th node of the l+1-th layer are weight values, w is the random weight, and b being bias values. The decrease of the size is done by fully connected. This step will be an operator for saliency features maps. The window size of 3 × 3, the convolution layer of 1 × 1 will be done. The bilinear upsampling overcomes the prediction for segmentation.

The shape of objects in medical images is complex and various. Their boundary structures are dependent on the features and the patient's body parts. Most methods for the abnormal detection are contour detection, segmentation, or prediction. This section presents a method for diabetic classification in retinal blood vessels. The proposed method is presented clearly as the Figure 1.
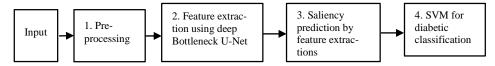
Input → 1. Pre-processing → 2. Feature extraction using deep Bottleneck U-Net → 3. Saliency prediction by feature extractions → 4. SVM for diabetic classification

Fig. 1. The proposed method for diabetic classification.

In the Figure 1, the proposed method includes four stages: preprocessing, feature extraction using deep Bottleneck U-Net (DbU-Net), saliency prediction and classification based on support vector machines (SVM). The object's features in the retinal blood vessels are influenced by a wide range of the surrounding elements, such as background, lighting, the blur/ noise value of the arround neighbor-pixels. The output of the original U-Net has two informations being the mask and contour. In this paper, the proposed method uses one output channel which is the salient mask. Saliency levels are detected and given to the next step which is classification by SVM. Now, the Figure 1 will be explained step by step as below.

3.1 The pre-processing for the input medical images stage

Threshold or filter is a vital position. In some researches, all transformations such as normalization, augmentation, etc will keep information important to the input of the below step. On the other hand, the background color is not the same as the blood vessel color in retinal blood vessels. In the diabetic retinal blood vessel, the disease bases on the color and the continuity of pixels together. Therefore, the concept of superpixels are presented. However, the superpixels are rarely used at the input of learning model because two reasons are:

1. The original convolution layer, which is the vital component of models, is defined over the grid and limited in the irregular superpixels.
2. The superpixels are calculated from the first block to the last block.

The position and size of retinal blood vessel images are stable. The input images are only divided into multi regions based on the neighbor-pixels in range of about 10 superpixels. This result will be used as the condition to divide each class one by one for the next step. This task is only a preparation for the first window in DbU-Net.

3.2 Feature extraction using DbU-Net stage

The traditional U-Net architecture also includes the collection between convolution layer and maxpooling layer in each block. On the other hand, the concatenation is done in the interlaced input of maxpooling layer. The number of blocks is 4 or 5 blocks in encoder and decoder processes. The base U-Net includes:

— Transition block with the matrix window and the number of filters: the sum of 3 convolution layers (2 layers of 3×3 and 1 layer of 5×5) for the input of the final convolution layer in this block. Then, activating the batch norm.
— Encoder (downsampling) has 5 blocks, each block has 2 convolution layers and 1 maxpooling layer.
— Dense block interlacing between 3 convolution layers (3×3) and 3 batch norm layers.
— Decoder (upsampling) has 3 convolution layers (3×3) and 1 batch norm.
— The final step is the convolution layer.

The characteristic of this network is the opposition of each block in encoder and decoder by U character. The final step of the original U-Net had given the mask and the contour for segmentation. Therefore, the feature extraction of the object in the image is still limited, leading to low accuracy when classifying objects. To overcome this limitation, this paper proposed an improvement of the architecture of the traditional U-net as follows:

— The input in the size of first layer in transition block,
— The change of bottleneck by the number of batch normalization in training for the output of encoder,
— The final layer for output and keeping the salient mask for segmentation.
— The union of sigmoid activation functions for saliency map and saliency prediction by pixel-superpixel.

The DbU-Net architecture for saliency prediction is presented in Figure 2. The DbU-Net architecture has the encoder and decoder for saliency prediction of the below steps, each block is the combination of convolution, activation, batch norm and ReLU, the maxpooling for the next section of each two blocks. The DbU-Net architecture also includes encoder and decoder parts. However, the input of the model will be changed. Here, each convolution layer has 1 convolution layer, 1 activation layer, 1 batch normalization layer, 1 ReLU.
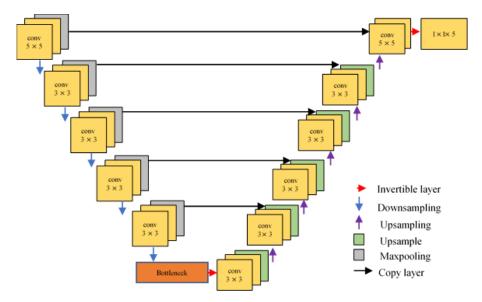

Fig. 2. The DbU-net architecture for saliency prediction

In the pre-processing step, the image is resized to 576×576 pixels. The plan of changing for the first layer was 2 convolution blocks with the 64 filters of the size 5×5. Its input is the retinal blood vessel image. The size of the input image is $M \times M$ with three color channels ($M \times M \times 3$). Therefore, the feature map has the size of $5 \times$

5×64 filters. The 2 convolution blocks and 1 maxpooling are a combo. In Figure 2, there are 5 combos in the encoder. The second combo with $3 \times 3 \times 64$ filters is for the first layer. The pooling layer is with M/2 and filters $\times 2$ (288×288 and 3×3 $\times 128$). The output of the maxpooling layer is used for the next blocks with 3×3 and double filters then the combo of 2 convolution blocks and 1 maxpooling.

The next period of this model is the Bottleneck concept. That includes:

— 01 batch normalization for the size keeping of the feature map in deep learning models.
— 02 batch normalization for training with the number of filters being 64×16, the same in padding.
— Increasing the double size of the window shape for concatenating in the decoder.

The development of the bottleneck in this method is that the 1 window 3×3 is applied in the first batch norm for checking the key feature extraction in U-Net. In decoder periods, each combo (2 convolution layers and 1 upsample layer) has the filters which are half off then upsampling. This decreases the nearest neighbor 2×2. Besides, the output of the final layer in each encoder block becomes the concatenation of each decoder block. The sigmoid activation function is applied at the final layer to generate saliency for segmentation. These results unify with the saliency by pixel-superpixel in the equation (3) for the final segmentation.

3.3 The saliency prediction based on the feature extraction stage

The pixel-superpixel association and superpixel centers with v iterations were calculated. The saliency level (R(x)) at pixel x in feature map of image is calculated in the equation (3):

$$
R(x) = \begin{cases}
0 & if \ 0 \leq S(x) < \frac{255}{K} \times 1 \\
\frac{255}{K-1} & if \ \frac{255}{K} \times 1 \leq S(x) < \frac{255}{K} \times 2 \\
& \quad . \\
& \quad . \\
& \quad . \\
255 & if \ frac255K \times (K-1) \leq S(x) < \frac{255}{K} \times K
\end{cases} \tag{3}
$$

where, S(x) is the saliency value at pixel x and the number of levels is K. The superpixel map was done as the below pseudo code.

The pre-processing period gives the 10 superpixels for clustering. In these steps, the distance from each pixel to 9 surrounding superpixels is calculated by the approximate nearest neighbor. The Figure 3 presented the segmentation results of saliency periods.
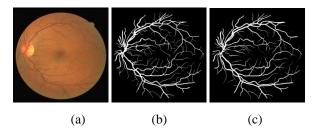
(a)   (b)   (c)

Fig. 3. The segmentation results by the Db U-net for saliency prediction. (a) The retinal vessels images in the DRIVE dataset. (b) The ground truth in the DRIVE dataset. (c) The segmentation result

```
Pseudo code:
Input: the feature extraction by Db U-Net
Output: the superpixel map for saliency prediction
  Function cal_SuperpixelMap(feature extraction)
      Features 𝓕ₙ×ₖ = 𝓕(i)
      The superpixel centers S = the average features i
      in grid
      while iteration t from 1 to v do:
         The association between pixel p and superpixel
         i is: 𝐐ᵗₚᵢ = e^(−‖𝓕ₚ−Sᵢᵗ⁻¹‖²)
         superpixel centers = ∑ association
      end while
         return argmax (association)
  End function
```

The output of this stage is the feature map and is the input for the SVM stage. However, the normalization for the saliency U-net has a wide range of complexes. Consequently, the below tasks will be to limit this range, they are:

1. Converting the input value from [0, 255] to [0, 1] and removing the lighting values.
2. Increasing the training data by data augmentation method by with operations of horizontal flip, vertical flip, crop, zoom, etc.

The fusion and non-saliency map are calculated for saliency map. Each level of saliency map is the basis for the distance between the average feature vectors of pixels. The contrast saliency value ($c_{i,j}$) for pixel at position (i, j) is calculated as in the equation (4):

$$c_{i,j} = D\left[\left(\frac{1}{N_1}\sum_{p=1}^{N_1}v_p\right),\left(\frac{1}{N_2}\sum_{p=1}^{N_2}v_q\right)\right] \tag{4}$$

3.4 The optimizing SVM parameters for classification stage

The optimization from soft margin is kernel margin for detection of support vectors. The soft margin (S) in SVM classifier is calculated as in the equation (5):

$$S = \left[\frac{1}{n}\sum_{i=1}^{n} \max(0, 1 - y_i(w^T x_i - b))\right] + \lambda \|w\|^2 \tag{5}$$

where, λ is the yield of the linearly classifiable input, x_i is the margin's boundary, w is the linear combination of support vectors.

The final results of this process is to classify the diabetic retinal blood vessels. The classification gives the features of saliency segmentation. The next section will present experimental results.

4 Experimentation and Evaluation Results

4.1 Materials and Datasets

Experiments are developed in Python, with the computer of Intel core i7, 2.9 GHz CPU, 16 GB DDR2 memory. The STARE [37] and DRIVE [38] dataset are used for experimentation.
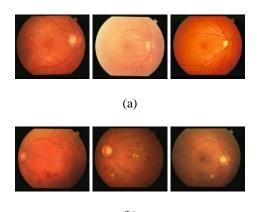
The STARE database contains 81 images, including 30 images of healthy retinas and 51 images of diseased retinas. The size of images is subsequently digitized at 605 × 700 pixels in resolution and 24 bits per pixel (standard RGB). These images were acquired using a TopCon TRV-50 fundus camera at 350 FOV [37]. Each image has been defined under ppm format.

The DRIVE dataset includes 400 retinal vessel images of diabetic patients aged between 25 and 90 years. The images were acquired using a Canon CR5 non-mydriatic 3CCD camera with a 45 degrees field of view (FOV) [38]. The size of images is subsequently digitized at 768 × 584 pixels in resolution and 8 bits per pixel (standard RGB). Each image has been defined under jpeg format.

With each of the above dataset, the experimentation used 80% dataset for training and 20% for testing to evaluate the results. The images are cropped to the size of 512*512. And these images augmented the dataset by rotation, scaling, clipping, etc. as in the Table 1. Consequently, the size of the training dataset is expanded to 2000 images. Some retinal blood vessel images in the STARE and DRIVE dataset are presented in the Figure 4 and Figure 5.

Table 1. Data augmentation parameters.

Transformation Type	Description
Rotation	Randomly rotate image between (−15˚, 15˚).
Clipping	Randomly clip images with angle between −10˚ and 10˚
Flipping	Horizontal and vertical flip images
Translation	Randomly shift between −5% and 5% of pixels

(a)

(b)

Fig. 4. Some retinal blood vessel images in the STARE dataset of diabetes. (a) Normal retinal blood vessel images. (b) Diabetic retinopathy image
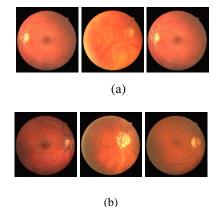

(a)

(b)

Fig. 5. Some retinal blood vessel images in the DRIVE dataset of non-diabetes. (a) Normal retinal blood vessel images. (b) Diabetic retinopathy image

4.2 Evaluation metrics and experimental results

With the evaluation of the segmentation based on a wide range of deep learning models, Jaccard Index (JI) value was used to evaluate the results of the proposed method. Suppose that, A and B are the image segmentation and the image which has ground truth, respectively. The JI(A, B) was calculated by the equation (6). JI value ranges from 0 to 100. The higher the value of JI is, the better it is.

$$JI(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{6}$$

Sensitivity (Se) defines the ability to detect diabetic retinopathy images. It is the ratio between the numbers of diabetic retinopathy images detected by the proposed

method and the total number of diabetic retinopathy images in the test dataset, and ranges from 0 to 1.

Specificity (Sp) defines the ability to distinguish images that have diabetic retinopathy or not. It is the ratio between the number of normal images that the proposed method can distinguish and the total number of normal images in the test dataset, and ranges from 0 to 1.

Accuracy (Acc) represents the result accuracy of the proposed method in the test dataset, and ranges from 0 to 1 (equivalent to the range from 0% to 100%).

The sensitivity, specificity and accuracy are used to evaluate metrics and are calculated as in the equation (7), equation (8) and equation (9), respectively.

$$Sensitivity = \frac{TP}{TP+FN} \tag{7}$$

$$Specificity = \frac{TN}{TN+FP} \tag{8}$$

$$ACC = \frac{TP+TN}{TP+FN+TN+FP} \tag{9}$$

where,

- True Positive (TP) is the number of the result pixels which are the pixels of blood vessels in ground truth.
- True Negative (TN) is the number of the result pixels which are not the pixels of blood vessels in ground truth.
- False Positive (FP) is the number of the result pixels which are used to detect a blood vessel, but that is wrong with the ground truth.
- False Negative (FN) is the number of the result pixels which are used to detect not a blood vessel, but that is wrong with the ground truth.

As presented in Section 3, the proposed method includes four stages. To evaluate the effectiveness of using salient map versus morphometry in the proposed method, we compared the JI values between using salient map versus morphometry at stage 3 based on some deep learning models such as: Convolutional neural network (CNN), Fully convolutional network (FCN) and U-Net, respectively. Table 2 and Table 3 presented the average Jaccard Index (JI) values of the morphology and saliency segmentation stages based on other models in the STARE and DRIVE datasets. The values in Table 2 and Table 3 are visualized in Figure 6.

Table 2. The average Jaccard Index (JI) values of the morphology and saliency segmentation stages based on other models in STARE dataset

Using	CNN	FCN	U-Net	DbU-Net
Morphology method	87.44	88.02	87.85	92.27
Saliency method	88.61	89.89	90.08	94.89

Table 3. The average Jaccard Index (JI) values of the morphology and saliency segmentation based on other models in DRIVE dataset

Using	CNN	FCN	U-Net	DbU-Net
The morphology method	86.71	88.63	88.95	93.71
Saliency method	89.04	90.26	91.07	95.69

(a)   (b)

Fig. 6. Comparison of the average JI value of the morphology and saliency segmentation based on other models in STARE dataset (a) and DRIVE dataset (b)

The morphology method for segmentation is not better than the saliency in U-Net, CNN, FCN and DbU-Net. However, the average Jaccard Index values of saliency prediction in Table 2 and Table 3 help strengthen DbU-Net in the diabetic retinal blood vessel segmentation. Figure 7 is a case of the result of saliency segmentation in non-diabetes images based on DbU-Net, U-Net, CNN and FCN in the DRIVE dataset. The Jaccard Index is calculated by the ground truth in the dataset with the segmentation results of each method.

From Table 2, Table 3, Figure 6 and Figure 7, the saliency prediction by DbU-net adapts to segment better. The reasons why saliency adapts to deep learning better are the feature extraction. If algorithms depend on the morphology of objects, the information will have shape and weak connections. Its adaptation is only the input of neural networks or deep learning. Therefore, the low pixels are the main contribution in morphology. Saliency predictions based on deep learning are given from the feature extraction. The number of encoder and decoder blocks of any model which detects segmentation helps choose the high pixels.

The evaluation of sensitivity, specificity, and accuracy criteria are proposed to objectively evaluate the effectiveness of the proposed method. The results of the proposed method are compared with Zhitao method [39], Memari method [40], Charu method [41] in the STARE dataset. These results comparisons are shown clearly in Table 4.
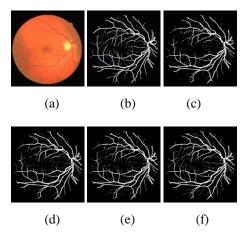
(a)   (b)   (c)

(d)   (e)   (f)

Fig. 7. The saliency segmentation of methods in non-diabetes image in DRIVE database.(a) The original image. (b) The ground truth of image (a) in the dataset. (c) The saliency segmentation based on the CNN method (JI = 89.57). (d) The saliency segmentation based on the FCN method (JI = 91.3). (e) The saliency segmentation based on the U-Net method (JI = 92.08). (f) The saliency segmentation based on the DbU-Net method (JI = 95.48).

Table 4. The average percent for the classification of methods in STARE dataset

Method	STARE dataset		
	Sensitivity	*Specificity*	*Accuracy*
Zhitao method [39]	0.862	0.934	0.921
Memari method [40]	0.882	0.981	0.942
Charu method [41]	0.913	0.984	0.958
Proposed method	0.947	0.982	0.968

With the STARE dataset, the accuracy of the proposed method is 96.8% whereas the accuracy of Zhitao method, Memari method and Charu method are 92.1%, 94.2% and 95.2%, respectively.

The results of the proposed method are also compared with Zhitao method [39], Memari method [40] in the DRIVE dataset. These comparisons are shown clearly in Table 5. With the DRIVE dataset, the accuracy of the proposed method is 97.1% whereas the accuracy of Zhitao method and Memari method are 92.9% and 95.1%, respectively. From the above evaluation, the DbU-Net for saliency prediction fits perfectly with the diabetic classification based on the retinal blood vessels.

Table 5. The average percent for classification of methods in DRIVE dataset

Method	DRIVE dataset		
	Sensitivity	*Specificity*	*Accuracy*
Zhitao method [39]	0.871	0.942	0.929
Memari method [40]	0.891	0.988	0.959
Proposed method	0.956	0.987	0.971

As presented in Section 3.2 and Section 3.4, the multi levels of the encoder in two blocks contribute to increasing the vital characteristics of deep learning. Then, the strong area prediction of objects is a wide range of kernels for SVM to classify. This is the reason why the results of the proposed method are better than the results of the other methods in the same data sets.

5 Conclusions

Diabetic retinopathy damages the small blood vessels that nourish the retina. Classification of diabetic retinopathy fundus image is very challenging because the blood vessels in the retinal images are too small. Consequently, using the retinal image processing method must be selective, avoiding loss of information in the image. This paper proposed the method for classification of diabetic retinopathy using saliency and shape detection of objects based on deep Bottleneck U-Net in the retinal blood vessels. Then, the diabetic classification depends on the support vector machine. The proposed method gives the results better than those of the other methods in STARE dataset and DRIVE dataset. To assess the levels of diabetes, the diagnostician must initially identify the retinal image with diabetes or not. The result of this paper may help the diagnostician to easily do this. However, the proposed method is not able to classify different types of diabetic retinopathy. To overcome this shortcoming, in the future of work, the proposed method will be improved to detect other features of the diabetic. Since then, it can classify different types of diabetic and improve the efficiency of classification.

6 Acknowledgement

7 References

[1] Marwan D. Saleh, C. Eswaran, and Ahmed Mueen (2011). An Automated Blood Vessel Segmentation Algorithm Using Histogram Equalization and Automatic Threshold Selection. J. Digit. Imaging, 24(4):564-572. https://doi.org/10.1007/s10278-010-9302-9

[2] Odstrcilik, Jan and Kolar, Radim and Budai, Attila and Hornegger, Joachim and Jan, Jiří and Gazarek, Jiri and Kubena, Tomas and Cernosek, Pavel and Svoboda, Ondrej and Angelopoulou, Elli (2013). Retinal vessel segmentation by improved matched filtering: Evaluation on a new high-resolution fundus image database. Image Processing, IET, 7: 373-383. https://doi.org/10.1049/iet-ipr.2012.0455

[3] Binh N.T., Tuyet V.T.H., Vinh P.C. (2014). Ultrasound Images Denoising Based Context Awareness in Bandelet Domain. Context-Aware Systems and Applications, ICCASA 2013, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Tele-

communications Engineering, Springer, 128. https://doi.org/10.1007/978-3-319-14227-2_12

[4] Gehad Hassan, Nashwa El-Bendary, Aboul Ella Hassanien, Ali Fahmy, Shoeb Abullah M. and Vaclav Snasel (2015). Retinal Blood Vessel Segmentation Approach Based on Mathematical Morphology. Procedia Computer Science, 65: 612 - 622. https://doi.org/10.1016/j.procs.2015.09.005

[5] Raja, D. and Sivamurugesan, Vasuki (2015). Automatic Detection of Blood Vessels in Retinal Images for Diabetic Retinopathy Diagnosis. Computational and Mathematical Methods in Medicine, 2015: 219-279. https://doi.org/10.1155/2015/419279

[6] Dash, Jyotiprava and Bhoi, Nilamani (2017). A Thresholding Based Technique to Extract Retinal Blood Vessels from Fundus Images. Future Computing and Informatics Journal, 2: 103 - 109. https://doi.org/10.1016/j.fcij.2017.10.001

[7] Holm, Sven and Russell, Greg and Nourrit, Vincent (2017). DR HAGIS - a fundus image database for the automatic extraction of retinal surface vessels from diabetic patients. Journal of Medical Imaging, 4: 414-503. https://doi.org/10.1117/1.jmi.4.1.014503

[8] Molina-Casado, José and Carmona, Enrique and García-Feijoó, Julián (2017). Fast detection of the main anatomical structures in digital retinal images based on intra- and inter-structure relational knowledge. Computer Methods and Programs in Biomedicine, 149: 55-68. https://doi.org/10.1016/j.cmpb.2017.06.022

[9] T. Jemima Jebaseeli, C. Anand Deva Durai and J. Dinesh Peter (2018). Retinal Blood Vessel Segmentation from Depigmented Diabetic Retinopathy Images. IETE Journal of Research, 0: 1-18. https://doi.org/10.1080/03772063.2018.1535918

[10] Vo Thi Hong Tuyet, Nguyen Mong Hien Pham Bao Quoc, Nguyen Thanh Son, Nguyen Thanh Binh (2019). Adaptive Content-Based Medical Image Retrieval Based On Local Features Extraction In Shearlet Domain. EAI Endorsed Transactions on Context-aware Systems and Applications, 6(17). https://doi.org/10.4108/eai.18-3-2019.159351

[11] Sundaram, Ramakrishnan and KS, Ravichandran and Jayaraman, Premaladha and B, Venkatraman (2019). Extraction of Blood Vessels in Fundus Images of Retina through Hybrid Segmentation Approach. Mathematics, 7: 172-169. https://doi.org/10.3390/math7020169

[12] Memari, Nogol and Ramli, Abd Rahman and Saripan, M. Iqbal Bin and Mashohor, Syamsiah and Moghbel, Mehrdad (2019). Retinal Blood Vessel Segmentation by Using Matched Filtering and Fuzzy C-means Clustering with Integrated Level Set Method for Diabetic Retinopathy Assessment. Journal of Medical and Biological Engineering, 39: 713-731. https://doi.org/10.1007/s40846-018-0454-2

[13] Rahebi, J. and Hardalaç, Fırat (2014). Retinal Blood Vessel Segmentation with Neural Network by Using Gray-Level Co-Occurrence Matrix-Based Features. Journal of medical systems, 38: 79-85. https://doi.org/10.1007/s10916-014-0085-2

[14] O. Ronneberger, P. Fischer and T. Brox (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv: Computer Vision and Pattern Recognition.

[15] Zhexin Jiang, Hao Zhang, Yi Wang, Seok-Bum Ko (2018). Retinal blood vessel segmentation using fully convolutional network with transfer learning. Computerized medical imaging and graphics, 68: 1 – 15. https://doi.org/10.1016/j.compmedimag.2018.04.005

[16] M. Li, Q. Yin and M. Lu (2018). Retinal Blood Vessel Segmentation Based on Multi-Scale Deep Learning. 2018 Federated Conference on Computer Science and Information Systems (FedCSIS), 1-7. https://doi.org/10.15439/2018f127

[17] Jayme Garcia Arnal Barbedo (2018). Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plat disease classification. Computer and Electronics in Agriculture, 153: 46 – 53. https://doi.org/10.1016/j.compag.2018.08.013

[18] Patricia Melin, Ivette Miramontes, German Prado-Arechiga (2018). A hydrid model based on modular neural networks and fuzzy systems for classification of blood pressure and hypertension risk diagnosis. Expert systems with applications, 106(1):146 – 164. https://doi.org/10.1016/j.eswa.2018.04.023

[19] Shankar, K., Lakshmanaprabu, S.K., Gupta, D. et al (2020). Optimal feature-based multi-kernel SVM approach for thyroid disease classification. J Supercomput, 76: 1128–1143. https://doi.org/10.1007/s11227-018-2469-4

[20] Li, K., Fang, Y., Li, W. et al (2020). CT image visual quantitative evaluation and clinical classification of coronavirus disease (COVID-19). Eur Radiol, 30: 4407–4416. https://doi.org/10.1007/s00330-020-06817-6

[21] Ibrahim Atli, Osman Serdar Gedik (2021). Sine-Net: A fully convolutional deep learning architecture for retinal blood vessel segmentation. Engineering Science and Technology, an Internatinal Journal, 24(2): 271 – 283. https://doi.org/10.1016/j.jestch.2020.07.008

[22] Umit Atila, Murat Ucar, Kemal Akyol, Emine Ucar (2021). Plant leaf disease classification using EfficientNet deep learning model. 61(101182). https://doi.org/10.1016/j.ecoinf.2020.101182

[23] Lee J.H., Joo I., Kang, T.W. et al. (2020). Deep learning with ultrasonography: automated classification of liver fibrosis using a deep convolutional neural network. Eur Radiol 30: 1264–1273. https://doi.org/10.1007/s00330-019-06407-1

[24] Liang, Y.; Chen, Z.; Ward, R.; Elgendi, M. (2018). Photoplethysmography and Deep Learning: Enhancing Hypertension Risk Stratification. Biosensors, 8(101). https://doi.org/10.3390/bios8040101

[25] Uthra Devi, K., Gomathi, R (2020). Brain tumour classification using saliency driven non-linear diffusion and deep learning with convolutional neural networks (CNN). J Ambient Intell Human Computing. https://doi.org/10.1007/s12652-020-02200-x

[26] Desmond Chuang Kiat Soh, E.Y.K. Ng, V.Jahmunah, Shu Lih Oh, Ru San Tan, U.Rajendra Acharya (2020). Automated diagnostic tool for hypertension using convolutional neural network. Computers in Biology and medicine, 126(103999). https://doi.org/10.1016/j.compbiomed.2020.103999

[27] P. Nageswara Reddy, C.P.V.N.J. Mohan Rao Ch. Satyanarayana (2017). Brain MR image segmentation by modified Active contours and Contourlet transform. ICTACT Journal on image and video processing, 8(2): 1645-1650. https://doi.org/10.21917/ijivp.2017.0231

[28] Sen He, Nicolas Pugeaulty (2018). Salient region segmentation. Computer Vision and Pattern Recognition, 1-6.

[29] Jingfei Hu, Hua Wang, Shengbo Gao, Mingkun Bao, Tao Liu, Yaxing Wang, Jicong Zhang (2019). S-UNet: A Bridge-Style U-Net Framework With a Saliency Mechanism for Retinal Vessel Segmentation. IEEE Access, 7: 174167-174177. https://doi.org/10.1109/access.2019.2940476

[30] Xue, Ly., Lin, Jw., Cao, Xr. et al (2019). A saliency and Gaussian net model for retinal vessel segmentation. Frontiers Inf Technol Electronic Eng, 20: 1075–1086. https://doi.org/10.1631/fitee.1700404

[31] Y. Sun, M. Prabhushankar and G. AlRegib (2020). Implicit Saliency In Deep Neural Networks. 2020 IEEE International Conference on Image Processing (ICIP), 2915-2919. https://doi.org/10.1109/icip40778.2020.9191186

[32] Tuyet V.T.H., Binh N.T., Quoc N.K., Ashish Khare (2021). Content Based Medical Image Retrieval Based on Salient Regions Combined with Deep Learning", Mobile Netw Application. https://doi.org/10.1007/s11036-021-01762-0

[33] H. Huang et al. (2021). Graph-Based Pyramid Global Context Reasoning With a Saliency-Aware Projection for Covid-19 Lung Infections Segmentation. ICASSP 2021 - 2021 IEEE

International Conference on Acoustics, Speech and Signal Processing (ICASSP), 1050-1054. https://doi.org/10.1109/icassp39728.2021.9413957

[34] X. Li, T. Pang, B. Xiong, W. Liu, P. Liang and T. Wang (2017). Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification. 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 1-11. https://doi.org/10.1109/cisp-bmei.2017.8301998

[35] Deepti Sisodia, Dilip Singh Sisodia (2018). Prediction of Diabetes using Classification Algorithms. Procedia Computer Science, 132: 1578-1585. https://doi.org/10.1016/j.procs.2018.05.122

[36] K. Shankar, Y. Zhang, Y. Liu, L. Wu and C. Chen (2020). Hyperparameter Tuning Deep Learning for Diabetic Retinopathy Fundus Image Classification. IEEE Access, 8: 118164-118173. https://doi.org/10.1109/access.2020.3005152

[37] https://cecas.clemson.edu/~ahoover/stare/ (last accessed 03rd Jan 2021).

[38] https://drive.grand-challenge.org/ (last accessed 03rd Jan 2021).

[39] Zhitao Xiao, Xinpeng Zhang, Lei Geng, Fang Zhang, Jun Wu, Jun Tong, Philip O. Ogunbona. and Chunyan Shan (2017). Automatic non-proliferative diabetic retinopathy screening system based on color fundus image. BioMedical Engineering OnLine,1-19. https://doi.org/10.1186/s12938-017-0414-z

[40] Memari. N, Abd Rahman Ramli, M. Iqbal Bin Saripan, Syamsiah Mashohor, Mehrdad Moghbel (2019). Retinal blood vessel segmentation by using matched filtering and fuzzy c-means clustering with integrated level set method for diabetic retinopathy assessment. Journal of Medical and Biological Engineering, 39: 713–731. https://doi.org/10.1007/s40846-018-0454-2

[41] Charu Bhardwaj, Shruti Jain, Meenakshi Sood (2020). Diabetic retinopathy lesion discriminative diagnostic system for retinal fundus images. Advanced Biomedical Engineering, 9: 71-82. https://doi.org/10.14326/abe.9.71

[42] Abou El-Seoud, S., Siala, M. H., & McKee, G. (2020). Detection and Classification of White Blood Cells through Deep Learning Techniques. International Journal of Online and Biomedical Engineering, 16(15), pp. 94–105. https://doi.org/10.3991/ijoe.v16i15.15481

[43] Shweikeh, E., Lu, J., & Al-Rajab, M. (2021). Detection of Cancer in Medical Images using Deep Learning. International Journal of Online and Biomedical Engineering, 17(14), pp. 164–171. https://doi.org/10.3991/ijoe.v17i14.27349

[44] Yanni, R. M. T., El-Ghitany, N. E.-K., Amer, K., Riad, A., & El-Bakry, H. (2021). A New Model for Image Segmentation Based on Deep Learning. International Journal of Online and Biomedical Engineering, 17(07), pp. 28–47. https://doi.org/10.3991/ijoe.v17i07.21241

8 Authors

Vo Thi Hong Tuyet received the Bachelor of Science degree in computer science from Ho Chi Minh City University of Pedagogical - Vietnam and the Master of Technology degree in computer science from Ho Chi Minh city University of Technology – Vietnam National University in Ho Chi Minh city (VNU-HCM), in 2011 and 2015, respectively. Now, she is a lecturer at Faculty of Information Technology, Ho Chi Minh City Open University, Vietnam. She is working as a PhD student at Faculty of Computer science and Engineering, the Ho Chi Minh City University of Technology, VNU-HCM. Her research interests include recognition, image processing (Email: vthtuyet.sdh19@hcmut.edu.vn, tuyet.vth@ou.edu.vn).

Nguyen Thanh Binh received the Bachelor of Engineering degree from Ho Chi Minh City University of Technology -Vietnam National University Ho Chi Minh City (VNU-HCM), Viet Nam, in 2000, the Master's degree and Ph.D degree in computer science both from University of Allahabad, India, in 2005 and 2011, respectively. Now, he is currently an Associate Professor in the Department of Information Systems, Faculty of Computer science and Engineering, the Ho Chi Minh City University of Technology, VNU-HCM. He has published one book, one book chapter and more than 65 research papers. His research interests include recognition, image processing, multimedia information systems, decision support system, and time series data.

Dang Thanh Tin received his M.S. degree in Electronic Engineering from the Ho Chi Minh city University of Technology, Viet Nam, and his Ph.D degree in Information system engineering in the cooperation of the Ho Chi Minh city University of Technology and Osaka Sangyo University, Japan, in 1998 and in 2006, respectively. From 2007 to 2013, he was at the Ho Chi Minh city University of Technology as Assistant Professor. He is currently an Associate Professor in the Faculty of Electrical and Electronics Engineering, the Ho Chi Minh city University of Technology, Viet Nam. In 2006, 2009, and 2016 he was a research fellow, and visiting professor in Information Systems Engineering Laboratory, Faculty of Electrical and Computer Engineering at University of Illinois at Urbana Champaign (USA), and in Lab of LCS (Laboratoire de Conception et d'Intégration des Systèmes) at Grenoble INP-Esisar (FRANCE), where he was responsible for research programs in the area of Computer systems engineering and Information systems engineering. Dang Thanh Tin is the author of over 58 technical publications. His research interests include image processing topics in biomedical engineering, medical expert systems topics (Email: dttin@hcmut.edu.vn).