

RESEARCH

Open Access



Retinal vessel segmentation with constrained-based nonnegative matrix factorization and 3D modified attention U-Net

Yang Yu and Hongqing Zhu*

*Correspondence:

hqzhu@ecust.edu.cn

School of Information Science & Engineering, East China University of Science and Technology, No.130 Meilong Road, 200237 Shanghai, China

Abstract

Due to the complex morphology and characteristic of retinal vessels, it remains challenging for most of the existing algorithms to accurately detect them. This paper proposes a supervised retinal vessels extraction scheme using constrained-based nonnegative matrix factorization (NMF) and three dimensional (3D) modified attention U-Net architecture. The proposed method detects the retinal vessels by three major steps. First, we perform Gaussian filter and gamma correction on the green channel of retinal images to suppress background noise and adjust the contrast of images. Then, the study develops a new within-class and between-class constrained NMF algorithm to extract neighborhood feature information of every pixel and reduce feature data dimension. By using these constraints, the method can effectively gather similar features within-class and discriminate features between-class to improve feature description ability for each pixel. Next, this study formulates segmentation task as a classification problem and solves it with a more contributing 3D modified attention U-Net as a two-label classifier for reducing computational cost. This proposed network contains an upsampling to raise image resolution before encoding and revert image to its original size with a downsampling after three max-pooling layers. Besides, the attention gate (AG) set in these layers contributes to more accurate segmentation by maintaining details while suppressing noises. Finally, the experimental results on three publicly available datasets DRIVE, STARE, and HRF demonstrate better performance than most existing methods.

Keywords: Nonnegative matrix factorization, Retinal vessel segmentation, Within-class and between-class constrained, 3 Dimension, Attention U-Net

1 Introduction

The retina is the only part of the human body that allows direct non-invasive visualization of its anatomical components. There is a close relationship between retinal vascular system and many diseases such as diabetic retinopathy, stroke, and cardiovascular disease. Manual labeling of blood vessels in fundus images is accepted by the medical community, but it is a long and time-consuming task, which requires medical specialists to be

competent. Therefore, the automatic detection of blood vessels instead of using only manual depiction is the most critical step for computer-aided diagnosis systems.

In this paper, a hierarchical classification framework for retinal vessel extraction is developed by using within-class and between-class constraint nonnegative matrix factorization (NMF) and three-dimensional (3D) modified attention U-Net architecture. The proposed framework performs on full green channel (G-channel) images directly and contains three major steps: (i) before feature extraction and classifier training, the input retinal images are pre-processed by Gaussian filtering, gamma correction, and region processing, respectively. This step is to reduce noise and outlier and adjust image contrast for better detection process of retinal vessel; (ii) this study considers the spatial relationship for each pixel by generating a vector ranking its neighboring 9×9 pixels. By placing each element of this vector as the row and the numbers of pixels as the column, thus, a nonnegative data matrix is formed. Then, we incorporate within-class and between-class constraints into the standard NMF objective function to obtain the nonnegative low-dimensional representation of the neighborhood information of each pixel. The within-class and between-class constraints are applied respectively into objective function of classical NMF to better discriminate features between different classes. This could be achieved by assimilating same classes eigenvectors with within-class constraint and differentiating different classes' eigenvectors by between-class constraint. By applying NMF, the coefficient matrix at lower dimension with 20 channels containing meaningful neighboring information is ready for final retinal vessel segmentation using network; (iii) we present a modified attention U-Net structure aim more precisely on extracting vessels by limited computation. This proposed network model is a symmetric U-shaped structure with attention mechanism so that the contraction path and expansion path can highlight salient feature useful for segmentation task. Unlike the conventional U-Net [1] that a large number of parameters would be trained in several input feature channels, we design a modified U-shape structure aim more precisely on extracting vessels. More specifically, we add an upsampling layer before the U-Net structure and encode information with only three max-pooling layers based on the feature maps with higher resolution obtained from upsampling layer. Similarly, three upsampling layers in the decoder path following by one max-pooling layer is built symmetrically to achieve end-to-end classifier. At the same time, an attention gate (AG) at each layer is set to record and convey detail information to decode path so that a more accurate identification of all vessels would be achieved in general. Being quantitatively and qualitatively verified on three public datasets DRIVE, STARE, and HRF, the proposed approach achieves better performance over other related algorithms.

The rest of the paper is organized as follows: Section 2 introduces some related works. Section 3 presents details of the proposed framework. All experiments and corresponding analyses are displayed in Section 4. At last, Section 5 outlines the concluding remarks and future research directions.

2 Related works

Existing segmentation approaches of retinal vessels could be divided into two categories: supervised and unsupervised by the use of manual labeled ground truth or not.

Unsupervised algorithms are designed according to inherent features of the retinal vessels without relying on artificial labeled images. Recent proposed unsupervised

approaches can be roughly divided into matching filter methods [2], vascular tracing methods [3], level set methods [4], model-based methods [5], hierarchical image matting models [6], etc. Generally, although unsupervised algorithms improve segmentation performance, thin vessels which affect the whole performance considerably is difficult to be detected [7].

Supervised algorithms require samples of vessels and non-vessels pixels from training databases with help of ophthalmologists to classify pixels for vessel detection. The algorithm usually uses extracted feature vector to train the classifier and to identify whether pixels belong to vascular or non-vascular. For example, Zhu et al. [8] designed a multi-dimensional discriminative feature vector extracting local features for vessel detection. Other supervised segmentation approaches include using Gaussian mixture model (GMM) [5], support vector machine (SVM) [9], random forest [10], various clustering strategies [11], etc. Supervised methods rely on hand-design feature extraction schemes predefined with prior knowledge. Features must be carefully defined in advance before entering the classifier, while features needed to be redesigned as dataset changes. Some other algorithms combined several types of features into one feature vector, while dimensionality problem may emerge [12]. Soares et al. [13] proposed a feature-based Bayesian extractor that build a 7-D feature vector for every pixel by Gabor wavelet transform. Lupascu et al. [14] adopted another 41-D feature vector for classification. Therefore, NMF [15], a linear dimensionality reduction technique commonly used for extracting basic and latent features from high-dimensional data matrices, is widely adopted. However, latent semantic structure within data set may not be discovered well by the basis vectors in classical NMF while high-dimensional data are represented by low-dimensional vectors [16]. In addition, since features are extracted from similar images, some inherent relations should have existed in these features whereas sometimes failed to. To overcome these problems, some local-based feature representation NMF algorithms by integrating sparseness constraints and graph constraints were presented [17, 18].

Recently, deep learning-based schemes have shown enormous success on pixel-wise classification problems due to its good performance in feature learning [19]. Carefully designed convolution neural network (CNN) could well serve instead of manual selection of features on vessel detection task. For example, Szkulmowski et al. [20] trained a CNN for vessel detection using augmented retinal vessel data. Soomro et al. [7] proposed a strided-CNN model that is very effective for thin vessel detection. This model is an encoder and decoder architecture where the pooling layers are replaced with strided convolutional layers. Another deeply supervised neural network with short connections to transfer semantic information between side-output layers [21]. In [22], Guo et al. formulated retinal vessel extraction task as a classification problem and solve it using CNN as a two labels classifier. Although CNN-based architectures can automatically learn features by convolution layers and pooling operations without prior knowledge, one of the main drawbacks of these methods is large number of training data required. Recently, a symmetric encoder and decoder structure U-Net was introduced and was approved to segmentation tasks with a small amount of data [1]. Wang et al. [23] introduced a modified U-Net architecture to capture more semantic information of fundus images by designing two encoders: spatial path and context path. Bhatkalkar et al. [24] integrated attention module in skip-connections between encoders and decoders of U-Net to highlight salient features. Recently, attention block is widely applied to emphasize targets and reduce the

effect of noise. In [25], Zhang et al. introduced an attention guided network (AG-Net) to achieve the retinal blood map. Li et al. [26] designed a mini-UNets architecture performed based on the output of classical U-Net that further achieved the obscured detail of vessel.

3 Methodology

The entire process of the proposed approach for extracting retinal vessels from fundus image consists of three main phases: (i) pre-processing of fundus images, (ii) reduce dimension using constrained NMF, and (iii) segmentation via 3D modified attention U-Net. Figure 1 shows the block diagram of this proposed approach.

3.1 Image preprocessing

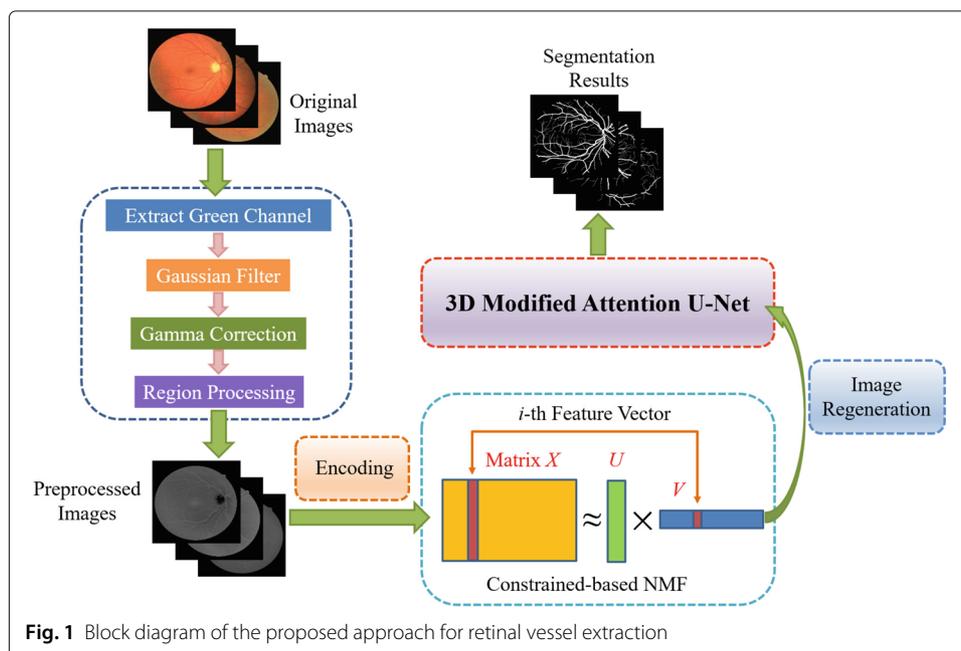
The main purpose of image pre-processing is to suppress background noise in images through Gaussian filter, and to equalize illumination of the optic disc and the fovea via Gamma correction. At last, the vascular features are emphasized by using region processing operator. In this paper, the G-channel of the retinal image is applied since it reflects the highest contrast, as shown in Fig. 2.

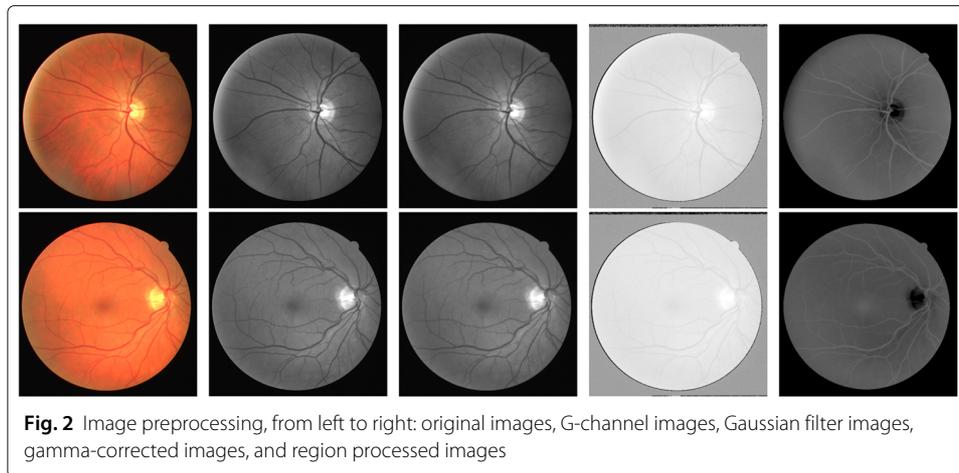
The first stage is to reduce background noise with Gaussian filter, which is a highly effective measure dealing with random noise. The filter function of this filter is written as follows

$$G_{\sigma}(x, y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{1}$$

where σ denotes the standard deviation. The value for σ is about 0.8 in this paper.

In the second stage, the gamma correction is used to adjust the contrast of images and to enhance local details. Besides, it can reduce the impact of local shadow and light variance of images. According to [27], the formula based on gamma correction is defined by





$$f(I) = I^\gamma. \tag{2}$$

In the experiment section, we will discuss how to set the best value of γ , From Fig. 2, we can observe that γ correction method provides high contrast image.

Region processing is the last stage of pre-processing, the gamma corrected images are converted, and the background around the retina is replaced with a density level of the same region in its mask image. The label image is binary image, where vessel area is 1 and non-vessel area is 0. But the corrected G-channel image showed vessel areas closing to 0, non-vessel area approaching to 1, and the mask area equals 0. Therefore, we reversed the image first, so that vessel area approaches to 1, non-vessel area approaches to 0, similar to the representation of label image. Mask area showing 0 is returned to zero according to the mask image.

3.2 Within-class and between-class constraint NMF

1) *Image coding*: This algorithm would first extract features for every pixel by its surrounding information. This algorithm constructs a 9×9 window with the observed pixels as center, the nearest 80 pixels having its G-channel density value collected. Figure 3 displays the block diagram of the proposed encoding method. This algorithm uses a column vector of 81 (including itself) as the original feature vector of every pixel. After all the original images (including training images and test images) are encoded into one matrix X , the original feature vectors are extracted from every pixel of all images and then all pixels are spliced to a matrix X with size of $m \times n$ by columns. In matrix X , one column denotes neighborhood information of a pixel.

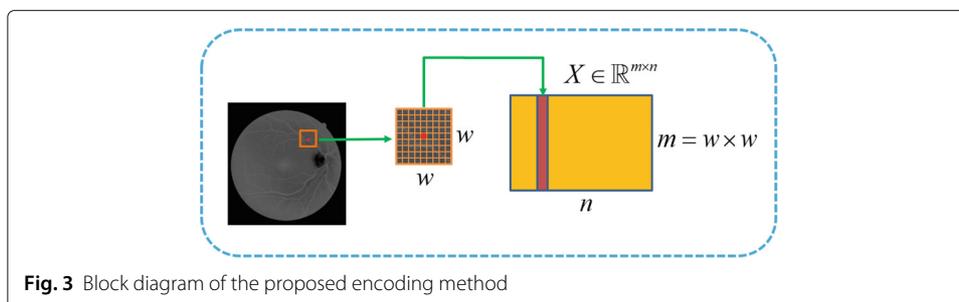


Fig. 3 Block diagram of the proposed encoding method

2) *Proposed constraint-based NMF*: Given a data matrix $X = [x_{ij}] = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{m \times n}$, the standard NMF aims to find two nonnegative matrices $U = [u_{ik}] = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_t] \in \mathbb{R}^{m \times t}$ and $V = [v_{kj}] = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n] \in \mathbb{R}^{t \times n}$ to approximate the given matrix X using

$$X \approx UV. \tag{3}$$

The objective function of classical NMF can be formulated by

$$\min_{U, V} J(U, V) = \|X - UV\|_F^2, \quad \text{s.t.} \quad U, V \geq 0, \tag{4}$$

where $\|\cdot\|_F$ denotes Frobenius norm (F-norm). Lee et al. [28] proposed that local minimum can be found by using the following multiplicative updates rules.

$$u_{ik} \leftarrow u_{ik} \frac{(XV^T)_{ik}}{(UVV^T)_{ik}}, \quad v_{kj} \leftarrow v_{kj} \frac{(U^T X)_{kj}}{(U^T UV)_{kj}}. \tag{5}$$

In this paper, we proposed a new objective function to properly obtain the optimal solution. Let r and c represent row and column of an image I , the total number of images in its dataset and training set is N and \bar{T} , respectively. Let the feature matrix of image I be $X \in \mathbb{R}^{m \times n}$, where $m = w \times w$ and $n = r \times c \times N$, \mathbf{x}_i ($i = 1, \dots, n$) represents the feature of the i th pixel. The feature of each pixel is an m -dimensional vector, which is achieved by using a square window at size of $w \times w$ ($w = 9$ in this paper) so that m is the total number of neighboring pixels around the current pixel including itself. Thus, the spatial relationship between neighboring pixels is taken into account.

The coefficient matrix is given as $U \in \mathbb{R}^{m \times t}$, and the inner dimension t is set to 20 in this paper. The new feature matrix after reducing dimension is defined as $V \in \mathbb{R}^{t \times n}$. In this case, the proposed objective function consists of the three terms could be given as follows

$$\min_{U, V, G} J(U, V, G) = \|X - UV\|_F^2 + \alpha \|L - UG\|_F^2 + \beta \|G - GA\|_F^2 (\tau \|GB\|_F^2)^{-1}, \tag{6}$$

where α and β adjust the contributions of two constrained terms. In this model, we randomly select \bar{T} images from dataset as training set. According to ground truth of these images, we split the coefficient matrix $L \in \mathbb{R}^{m \times \tau}$ into $L = [L_a L_b]$, where $L_a \in \mathbb{R}^{m \times a}$, $L_b \in \mathbb{R}^{m \times b}$, $\tau = r \times c \times \bar{T}$, $\tau = a + b$. The first L_a column of the matrix L represents that all of these training images have a pixels belonging to blood vessels, and b pixels belong to the background. After NMF decomposition, $G \in \mathbb{R}^{t \times \tau}$ is a $t \times \tau$ matrix with the first a column denoting the features of pixels which belong to blood vessels. The aim of the second regularized term on the right-hand side of (6) is to find the two nonnegative matrixes G and U in order to maintain the data space structure of original selecting training images on low-dimensional data space obtained from the matrix factorization. The first few columns of this matrix G represent the pixels that belong to vessels. The goal of the third regularized term of (6) is to take the within-class and between-class distances as additional constraint on the proposed objective function. Let matrix A being

$$A = \begin{bmatrix} \frac{1}{a} & \frac{1}{a} & \dots & \frac{1}{a} & & & & & \\ \frac{1}{a} & \frac{1}{a} & \dots & \frac{1}{a} & & & & & \\ \vdots & \vdots & \ddots & \vdots & & & & & \\ \frac{1}{a} & \frac{1}{a} & \dots & \frac{1}{a} & & & & & \\ & & & & \frac{1}{b} & \frac{1}{b} & \dots & \frac{1}{b} & \\ \mathbf{0} & & & & \frac{1}{b} & \frac{1}{b} & \dots & \frac{1}{b} & \\ & & & & \vdots & \vdots & \ddots & \vdots & \\ & & & & \frac{1}{b} & \frac{1}{b} & \dots & \frac{1}{b} & \end{bmatrix}. \tag{7}$$

The constrained term $\|G - GA\|_F^2$ is designed in such a way that within-class distance between the feature vector of each pixel and the mean value is expected to approach zero. The study also considers the F -norm as a constraint on matrix GB to take between-class scatter distance into consideration in our model. To do is, we define the vector $B \in \mathbb{R}^{\tau \times 1}$

$$B = \left[\frac{1}{a} \dots \frac{1}{a} -\frac{1}{b} \dots -\frac{1}{b} \right]^T. \tag{8}$$

In our method, two constrained terms $\|G - GA\|_F^2$ and $\|GB\|_F^2$ are negatively correlated. Specially, the ideal situation is that the within-class distance $\|G - GA\|_F^2$ decreases while between-class distance $\|GB\|_F^2$ increases. The constraint term $\|GB\|_F^2$ is multiplied by the coefficient τ to balance its weight values and $\|G - GA\|_F^2$.

For convenience of calculations, α in (6) is set to one, and β is defined τ ($\alpha = 1, \beta = \tau$). Thus, the right-hand side of the third term of (6) becomes $\|G - GA\|_F^2 \|GB\|_F^{-2}$. In order to avoid the initial value of this matrix at the beginning of the iteration being zero while considering that $\|GB\|_F^2 \gg 1$, this paper rewrites the objective function (6) as follows

$$\min_{U, V, G} J(U, V, G) = \|X - UV\|_F^2 + \|L - UG\|_F^2 + \|G - GA\|_F^2 (\|GB\|_F^2 + 1)^{-1}. \tag{9}$$

The following expression is obtained by taking partial derivatives of J with respect to coefficient matrix U and feature matrix V .

$$\frac{\partial J}{\partial U} = -XV^T + UVV^T - LG^T + UGG^T, \tag{10}$$

$$\frac{\partial J}{\partial V} = -U^T X + U^T UV. \tag{11}$$

Taking a derivative of J with respect to G , we have

$$\begin{aligned} \frac{\partial J}{\partial G} &= -U^T L + U^T UG \\ &+ \frac{(G - GA - GA^T + GAA^T) (\|GB\|_F^2 + 1) - GBB^T \|G - GA\|_F^2}{(\|GB\|_F^2 + 1)^2}. \end{aligned} \tag{12}$$

According to Karush-Kuhn-Tucker (KKT) conditions [29] $\psi_{ik} u_{ik} = 0$ and $\phi_{kj} v_{kj} = 0$, we have the following formulas for u_{ik} and v_{kj} .

$$\left(2UVV^T + 2UGG^T \right)_{ik} u_{ik} - \left(2XV^T - 2LG^T \right)_{ik} u_{ik} = 0, \tag{13}$$

$$\left(2U^T UV \right)_{kj} v_{kj} - \left(2U^T X \right)_{kj} v_{kj} = 0. \tag{14}$$

Similarly, $\varphi_{kl}g_{kl} = 0$ leads to the following expression.

$$\begin{aligned} & \left(2 \left(G - GA - GA^T + GAA^T \right) \left(\|GB\|_F^2 + 1 \right) - \frac{2GBB^T \|G - GA\|_F^2}{\left(\|GB\|_F^2 + 1 \right)^2} \right)_{kl} g_{kl} \\ & + \left(-2U^T L \right)_{kl} g_{kl} + \left(2U^T UG \right)_{kl} g_{kl} = 0. \end{aligned} \tag{15}$$

Then, the updating rules of U and V can be deduced from the above equations as follows.

$$u_{ik} \leftarrow u_{ik} \frac{(XV^T + LG^T)_{ik}}{(UVV^T + UGG^T)_{ik}}, \quad v_{kj} \leftarrow v_{kj} \frac{(U^T X)_{kj}}{(U^T UV)_{kj}}. \tag{16}$$

Using similar strategy, the multiplicative update of matrix G would be

$$g_{kl} \leftarrow g_{kl} \frac{\left(U^T L \left(\|GB\|_F^2 + 1 \right)^2 + \left(GA + GA^T \right) \left(\|GB\|_F^2 + 1 \right) + GBB^T \|G - GA\|_F^2 \right)_{kl}}{\left(U^T UG \left(\|GB\|_F^2 + 1 \right)^2 + \left(G + GAA^T \right) \left(\|GB\|_F^2 + 1 \right) \right)_{kl}}. \tag{17}$$

Because matrix A is a symmetric positive definite matrix, that is $GA^T = GA$, $GAA^T = GA$, thus, (17) can be rewritten as follows

$$g_{kl} \leftarrow g_{kl} \frac{\left(U^T L \left(\|GB\|_F^2 + 1 \right)^2 + 2GA \left(\|GB\|_F^2 + 1 \right) + GBB^T \|G - GA\|_F^2 \right)_{kl}}{\left(U^T UG \left(\|GB\|_F^2 + 1 \right)^2 + \left(G + GA \right) \left(\|GB\|_F^2 + 1 \right) \right)_{kl}}. \tag{18}$$

The optimizing scheme of this proposed constrained-based NMF is summarized in **Algorithm 1**.

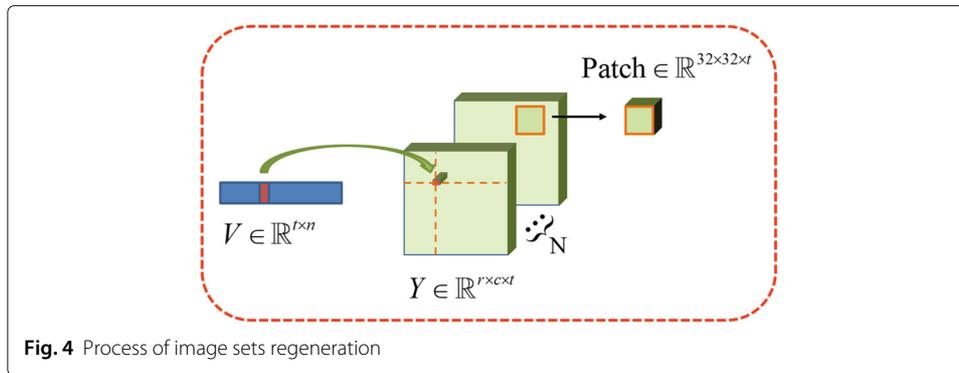
Algorithm 1 The optimizing scheme of constrained-based NMF

Input: X, L ;

Output: V ;

- 1: **Initial:** U by using the random initialization, V and G with one. Construct matrices A and B ;
 - 2: **Repeat:**
 - Update U according to (16);
 - Update V using (16);
 - Update G by using (18);
 - 3: **Until** the objective function (9) convergence;
 - 4: **Return** V ;
-

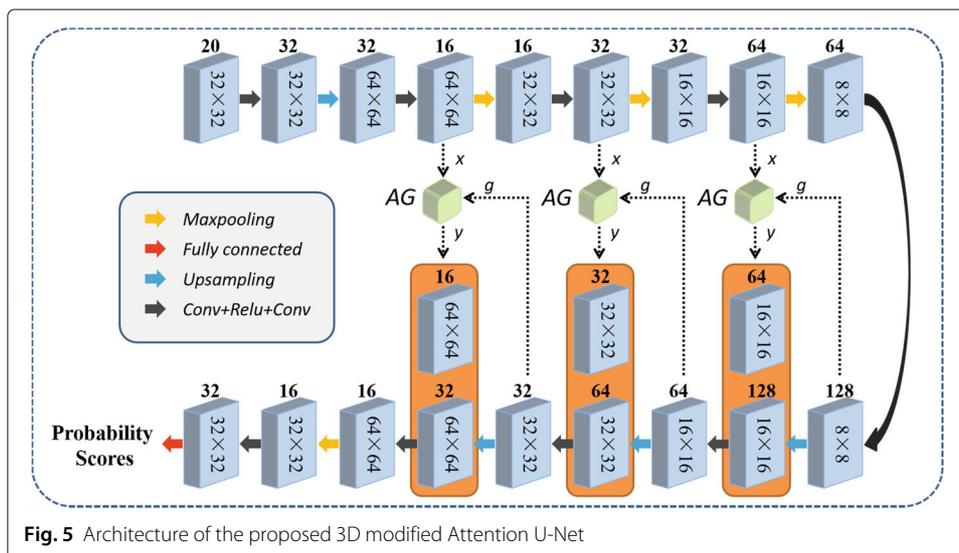
3) *Image regeneration:* After final generation of a more contributing feature matrix V where columns represent neighborhood information of n pixels, only 20 ($t = 20$ in this paper) rows remained for feature description. For further utilization of this contributing low dimension feature description method, we convert matrix V back to images with every pixel of the same encoding sequence. Thus, every pixel of the processed image with neighborhood feature information of 20 dimensions would be conveyed to proposed 3D segmentation networks for further vessel classification. Figure 4 shows the block diagram of image set regeneration.

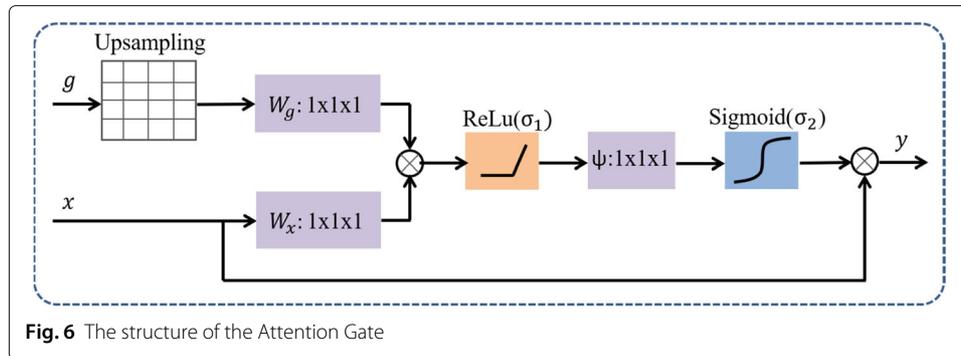


3.3 Proposed 3D modified attention U-Net

Computer vision-based blood vessel detection requires algorithms with high accuracy and relatively convenient computation. By comprehensively describing neighborhood information using our constrained-based NMF, this study designs an end-to-end 3D modified attention U-Net architecture as a trainable classifier for vessel extraction. The architecture of this proposed network is shown in Fig. 5. Considering that current attention U-Net [24] plays only basic function in classification, the proposed 3D modified attention U-Net aims at reducing computational complexity with limited resource devoting to region of intense classification need.

Specifically, input data obtaining from N images where one image would be saved as $r \times c$ with 20 channels representing neighborhood features. These data would be firstly divided into patches at size $32 \times 32 \times 20$ and conveyed into the whole network. This network set an upsampling layer raising patch size by doubling the original input size before conventional maxpooling layers. Besides, only three maxpooling layers remain in our network compared to original four in U-Net compressing patch size. Symmetrically, three upsampling layers followed by a maxpooling recovering patch size is set in concert with encode path. This new design achieves comprehensive result among networks in terms of computational cost and segmentation accuracy.





Considering that retinal vessel segmentation is a highly specific task on few image regions with intense meaningful information, three AGs conveying contributing information are inserted as shown in Fig. 6. Here, x signal conveying attention maps maintaining fine-grained details adding with g signal from former layers would generate a y output multiplying with former feature maps. In this way, more focus on salient features would be distributed on more detailed feature maps, and vessel areas would gain more learning resources, where large area of non-vessel retina would be suppressed.

4 Results and discussions

In this section, we first describe the datasets and metrics used in the experiment, and then detail the experimental results and performance analysis of the proposed method on some publicly used benchmark datasets.

4.1 Benchmark datasets

The framework will be evaluated based on high-resolution images from three publicly available datasets: DRIVE¹ [30], START² [31], and HRF³.

The DRIVE contains 40 fundus images which the size of each image is 565×584 pixels with 8 bits per color channel. All of the images have been segmented manually as a ground truth and their field of view (FOV) binary masks are also provided. The DRIVE dataset used in this paper contains 40 fundus images, in which training set is consisted of 35 color retinal images, and the other 5 images are adopted for testing.

The START contains 20 fundus images with resolution 700×605 pixels and 8 bits per color channel, in which 10 with pathologies and 10 without any pathologies. All images in this dataset are manually segmented by two observers. The results of the first observer are regarded as the ground truth. In the experiment, we randomly selects 16 images with hand labeled results for training and left 4 images for testing.

The HRF database [32] consists of 45 high-resolution eye fundus images at size 3504×2336 segmented by a group of experts working in the field of retinal image analysis and clinicians from the cooperated ophthalmology clinics. In specific, the dataset could be divided into 15 images of healthy patients, 15 images of patients with diabetic retinopathy and 15 images of glaucomatous patients. One ground truth image and a mask determinate field of view (FOV) is attached for each image. In this paper, we randomly select 41 images for training and the remaining 4 images for test. Testing images contain

¹<http://www.isi.uu.nl/Research/Databases/DRIVE/>

²<http://cecas.clemson.edu/~ahoover/stare/>

³<https://www5.cs.fau.de/research/data/fundus-images/>

one healthy patient image (No.02_h) and three glaucomatous patient images (No.05_g, No.09_g, No.10_g).

4.2 Experimental environment and evaluation metrics

This subsection is to evaluate the segmentation performance on DRIVE and START datasets and compares the proposed method with state-of-the-art algorithms. All experiments are run on a small server with Intel (R) Core (TM) i7-9700KF CPU (4.8 GHz) with NVIDIA GeForce RTX 2080 Ti GPU. Our architecture was built based on a publicly available Python 3.7 platform and was implemented on Tensorflow backend Keras deep learning library.

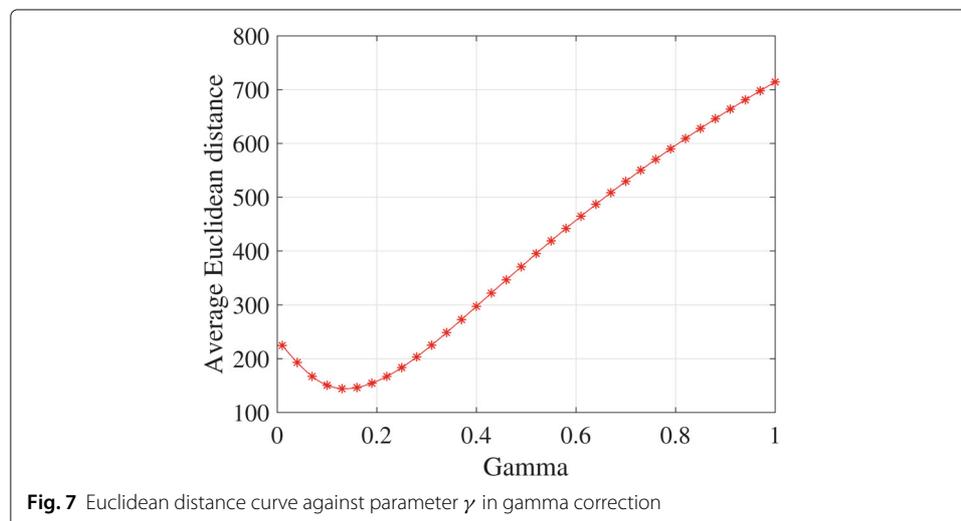
The performance of the vessel segmentation is measured using sensitivity (*SE*), specificity (*SP*), accuracy (*ACC*), *precision*, and *recall*. They are defined as follows:

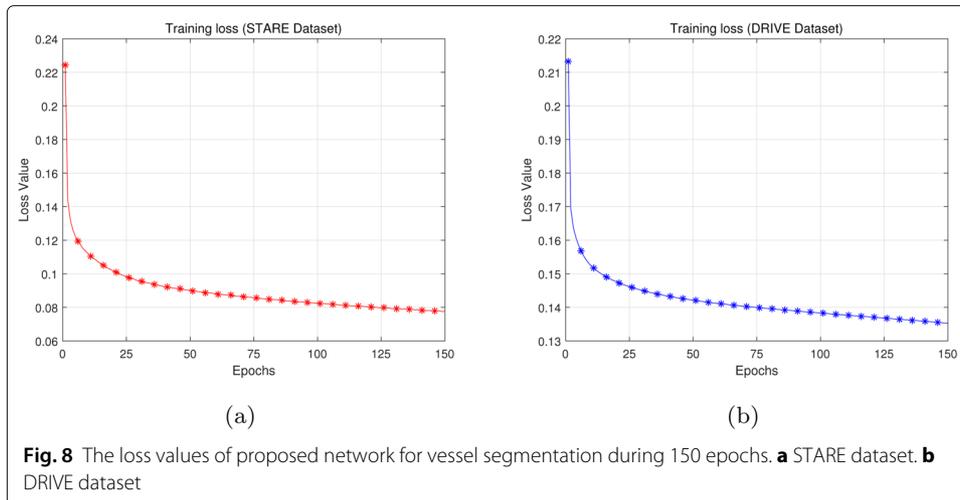
$$SE = \frac{TP}{TP + FN}, \quad SP = \frac{TN}{FP + TN}, \tag{19}$$

$$ACC = \frac{TP + TN}{TP + FP + FN + TN}, \tag{20}$$

$$Precision = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN}, \tag{21}$$

where *TP* and *TN* denote the number of pixels correctly classified as vessel pixels and non-vessel, respectively. *FN* represents the number of vessel pixels incorrectly labeled as non-vessel. *FP* is the number of non-vessel pixels incorrectly labeled as vessels. Precision and Recall measure the exactness and completeness of model performance. In addition, the performances have been examined in terms of standard indexes, such as AUC (area under the curve) and ROC (receiver operating characteristic curve) [33]. The AUC value is calculated using the trapezoidal rule. The closer the AUC value is to 1, the better the performance of the corresponding blood vessel segmentation algorithm. The ROC curve is a plot of *SE* versus $1-SP$ by varying the threshold on probability map.





4.3 Selection of parameter

This subsection discusses which value of γ from the gamma correction phase is most appropriate for the proposed approach. Thus, the pre-processing performances on ten images randomly selected from DRIVE dataset and compares the enhanced images with the ground truth in terms of Euclidean distance. Figure 7 shows the effect of varying value of γ in gamma correction on Euclidean distance. It is experimentally found that there is an obvious bottom point of the Euclidean distance while parameter γ increases, and the

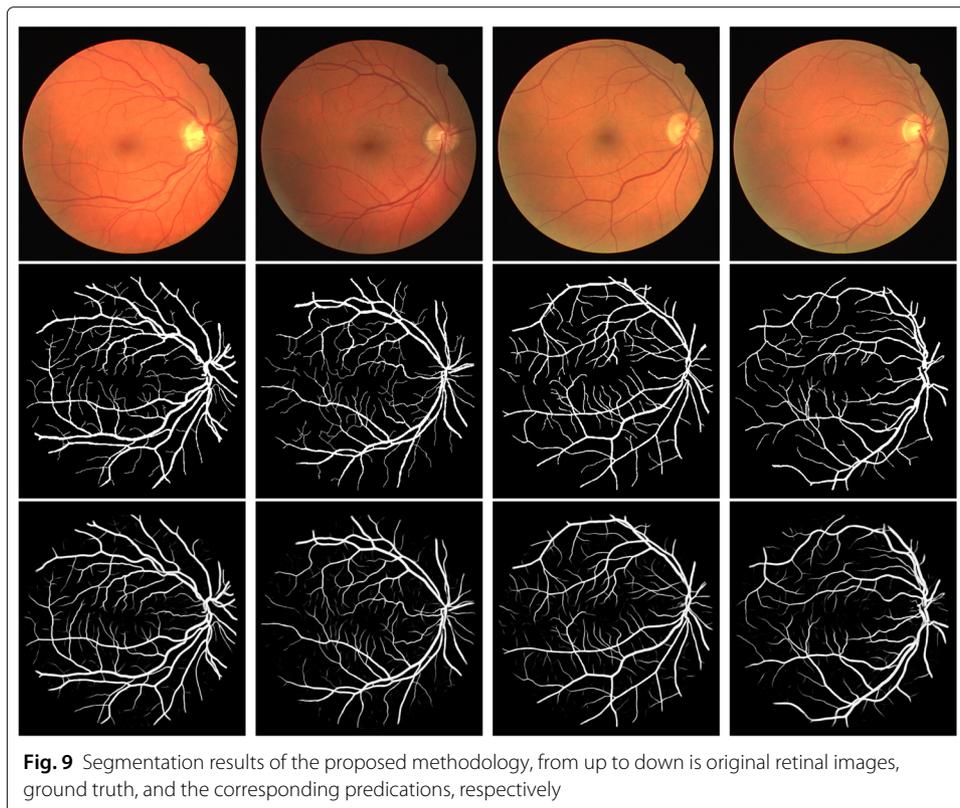


Table 1 Performance of proposed method on different datasets

STARE dataset		Predicted condition	
True condition	Condition positive	TP=87800	FN=11064
	Condition negative	FP=16627	TN=817366
DRIVE dataset		Predicted condition	
True condition	Condition positive	TP=85926	FN=15948
	Condition negative	FP=17268	TN=789017

minimum is achieved at $\gamma = 0.14$. Based on this parameter study, we adopt $\gamma = 0.14$ in the following experiments as our method can produce the best results in this case.

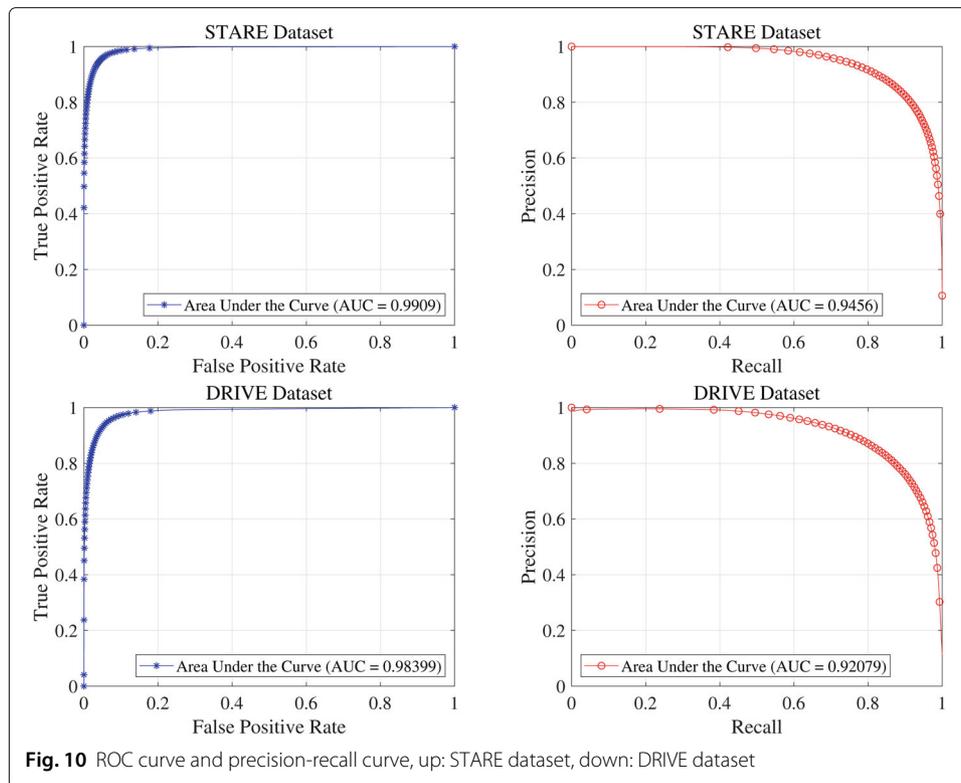
4.4 Experiment on retinal vessel extraction

The training processing is summarized as follows: After the network model generated, some processes are implemented before training. Reading the feature matrix, mask, and ground truth images of the original training set respectively, pixels outside the mask regarded as the region of interesting (ROI) could be gathered into a label set. Through this measure, useless information could be abandoned before training. We train and test the proposed network model on both STARE and DRIVE datasets and fine-tuned our network with a learning rate of $5e - 5$, a weight decay of $1e - 6$. A dropout rate of 0.2 was used between two convolutional layers. Batch size was set to 32, and 150 epochs were used to ensure convergence. In the training phase, we used the Adam optimizer [34]. The loss values verse epochs obtained in the process of model training are given in Fig. 8.

During the image prediction phase, a similar data process as the training model, including pre-processing and constrained-based NMF, is carried out as well. Reading the information of each pixel and locating them in mask images of the original testing set, pixels which do not fall in the mask areas are conveyed into the proposed network model for testing. After a new label set (segmentation result) is outputted by the predicting process implemented by our model, pixels in the label set are filled back into the image with the order they are picked. Figure 9 shows some examples generated by the proposed methodology on DRIVE dataset, from which we can observed that our method is available to extract abundant vascular branches at different thickness. To prove the validity of the proposed, we evaluate quantitatively the retinal vessel segmentation results on test sets of both DRIVE and STARE by comparing the average values of the predictions with ground truth. Four different evaluation metrics ACC, SE, SP, and precision are applied, where all of them are computed from TP, FN, FP, and TN. Tables 1 and 2 list the evaluation results obtained by using the proposed framework on different datasets. The ROC curves of two databases are measured to quantify the proposed predication results and are provided in Fig. 10. As can be observed, our method performs better in detecting vessels on STARE than DRIVE. Our model also generates high AUC on two test tasks, at 0.9909 for

Table 2 Performance of proposed method on different datasets

Datasets	Accuracy	Sensitivity	Specificity	Precision
STARE	0.9703	0.8881	0.9801	0.8408
DRIVE	0.9634	0.8434	0.9786	0.8626



STARE, and 0.9839 for DRIVE. These values demonstrate the validation of our proposed framework on predicting retinal vessels.

4.5 Comparison with other network models

To test the effectiveness of the proposed framework, we compared the output of our approach with several advanced algorithms U-Net⁴ [1], AG-UNet⁵ [35], IterNet⁶ [26], DenseNet⁷ [36], and V-GAN⁸ [37] on STARE and DRIVE. Their segmentation results are obtained by running publicly available codes.

All these deep convolution network-based algorithms are able to extract most of these vessels, while the proposed method performs well on most images, even when image contrast is low. Four images demonstrating retinal vessel segmentation algorithms results and ground truth from DRIVE dataset are shown in Fig. 11. Figure 12 shows the enlarged images of six models by bilinear interpolation to the size of 200×200 pixels on three images from STARE dataset. It could be observed that crossing vessel branches and thick vessels are two most significant factors for misclassification, which a different level is presented under different algorithms. U-Net and AG-UNet achieve similar segmentation while the inserted AG presents slightly better preservation of vessel details. However, thin vessels still remain broken or blur. In contrast, DenseNet and V-GAN almost capture all suspicious vein compared to the other four methods. But all predicted vascular networks seem to be exaggerated so that distinction of vessel thickness are not significant,

⁴ <https://github.com/ternaus/robot-surgery-segmentation>

⁵ <https://github.com/ztzhao6/Liver-Segmentation-with-AttentionUNet>

⁶ <https://github.com/conscienceli/IterNet>

⁷ <https://github.com/DeepTrial/Retina-VesselNet>

⁸ <https://github.com/jaeminSon/V-GAN>

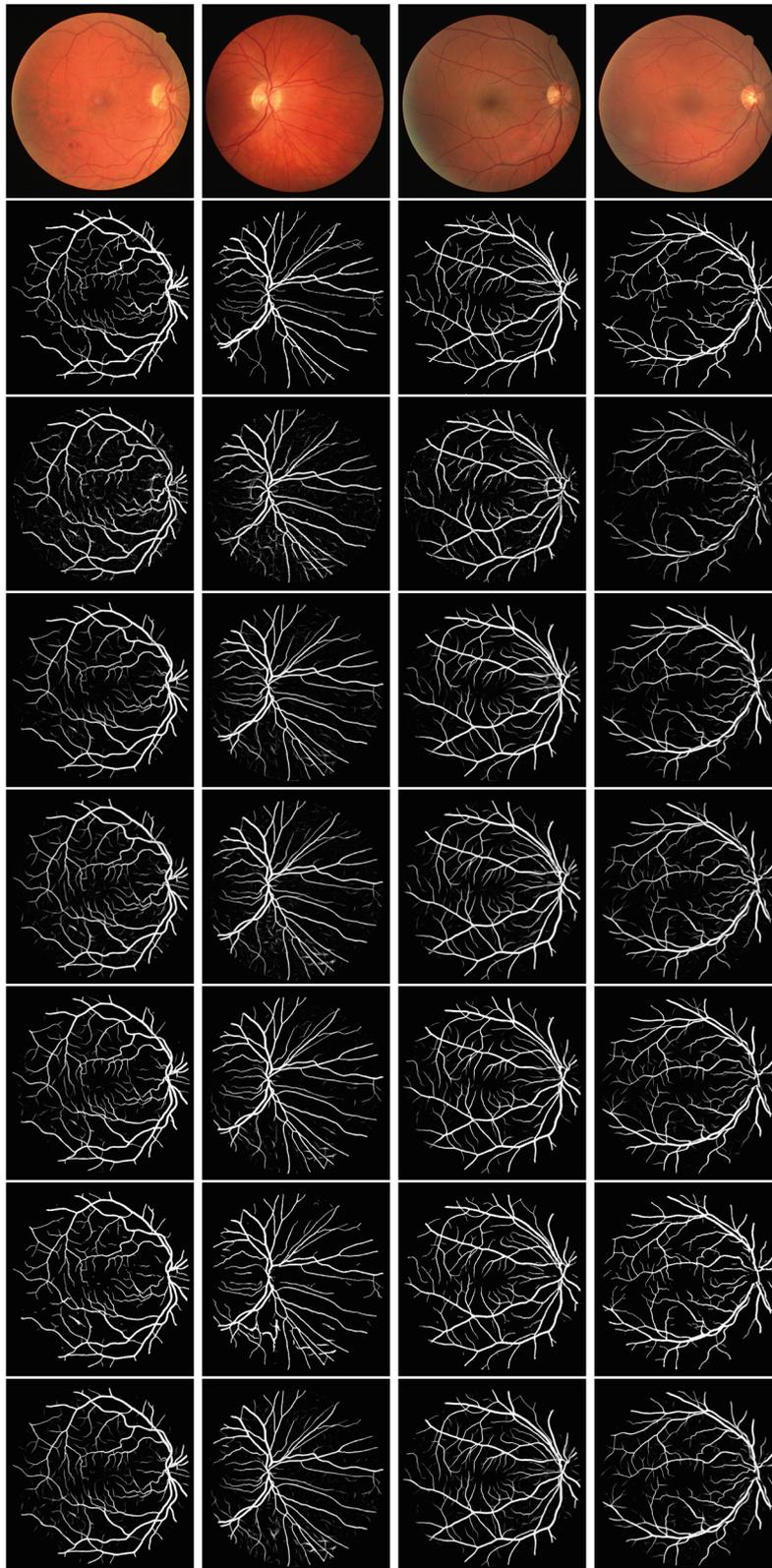


Fig. 11 Comparison of vessel segmentation results of existing algorithms with the proposed method on DRIVE dataset, from up to down is input images, ground truth, U-Net, AG-UNet, IterNet, DenseNet, V-GAN, and ours, respectively

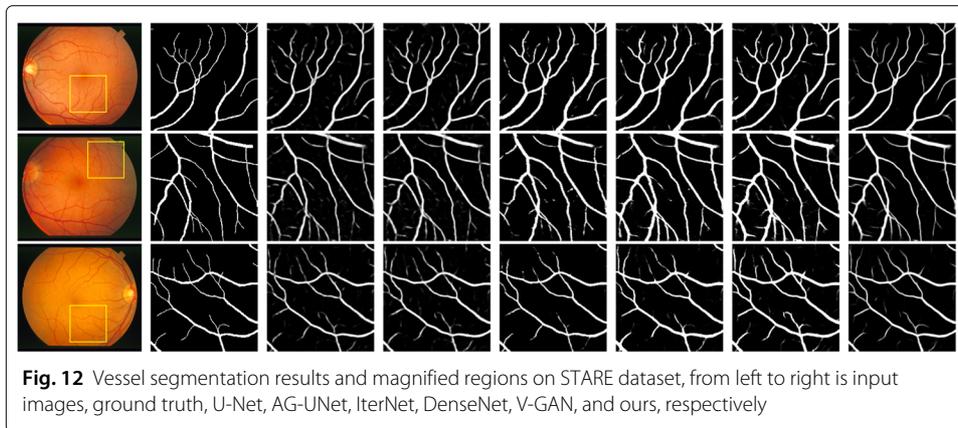


Fig. 12 Vessel segmentation results and magnified regions on STARE dataset, from left to right is input images, ground truth, U-Net, AG-UNet, IterNet, DenseNet, V-GAN, and ours, respectively

and some vessels are excessive detected. IterNet, however, presents both problems of the above methods where detected areas are all of similar thickness, and details are detected less manifest. Comparing with all these methods, our model results could avoid detecting either too coarse or blurred. Interrupting strips are also less likely to be falsely captured, whereas inapparent blood vessels could be identified more approximate the ground truth. For more visually convenient comparison, we list six methods in a reasonable sequence as shown in Fig. 12.

For more validation of the proposed method, we calculate the evaluation metrics of vessels of resulted images using their corresponding ground truth. After that, we calculate the SE, SP, ACC, and precision of vessels of DRIVE and STARE datasets as shown in Tables 3 and 4. Higher sensitivity assures all potential vessel areas being detected, and higher specificity assure correctness among detected area. Consistent with our visual analysis, the SP value of U-Net and AG-UNet is higher on both datasets at about 0.97. This indicates that the identification is relatively conservative and basic so that uncertain areas such as thin and blur vessels maybe missed. But the SE of V-GAN are higher than other network at about 0.95 on STARE and 0.85 on DRIVE. This means the classification are relatively coarse and suspicious areas are of high likelihood to be identified as vessels. The results of IterNet also consists with former evaluations, where the SEs are smaller than DenseNet and V-GAN but larger than U-Net and AG-UNet. Also, the SPs are either similar with methods from both trends or staying at a moderate level. However, it could be observed that these methods have their own characteristics, while neither being too specific nor too sensitive meet the requirement of efficient real-world application. Our method shows a good tradeoff between both metrics, while qualified results are reached compared to all methods. If we look at ACC and precision, which higher figure means

Table 3 Performance analysis of all algorithms on STARE databases with respect to the measuring metrics

Methods	Accuracy	Specificity	Sensitivity	Precision
U-Net [1]	0.9561	0.9765	0.7758	0.7885
AG-UNet [35]	0.9636	0.9778	0.8437	0.8188
IterNet [26]	0.9517	0.9575	0.9027	0.7157
DenseNet [36]	0.9495	0.9503	0.9424	0.6925
V-GAN [37]	0.9482	0.9475	0.9540	0.6832
Ours	0.9703	0.9801	0.8881	0.8408

Table 4 Performance analysis of all algorithms on DRIVE databases with respect to the measuring metrics

Methods	Accuracy	Specificity	Sensitivity	Precision
U-Net [1]	0.9491	0.9796	0.7071	0.8148
AG-UNet [35]	0.9599	0.9775	0.8201	0.8221
IterNet [26]	0.9574	0.9831	0.7791	0.8691
DenseNet [36]	0.9604	0.9750	0.8449	0.8106
V-GAN [37]	0.9560	0.9689	0.8541	0.7763
Ours	0.9634	0.9786	0.8434	0.8626

better precise of a network, U-Net and AG-UNet present generally higher values than the other three methods. Nonetheless, our method achieves the highest ACC at 0.9703 and 0.9634, and the highest precision at 0.8726 and 0.8408.

4.6 Comparison of segmentation results on high-resolution dataset

Since high resolution images are becoming common in clinical use, we evaluate our method on HRF dataset (image size 3504×2336 pixels) in this experiment. We compare the proposed approach with several state-of-the-art methods and retinal vessel segmentation results are displayed in Fig. 13. Among these methods, Soares et al. [13] is a standard segmentation algorithm while the others are all based on CNN. Some competitive methods proposed recently [37, 38] are also engaged in comparative experiment. It could be seen from Fig. 13 that traditional segmentation method [13] shows a relatively blur vessel outlines than CNN-based approaches. U-Net and AG-UNet still seem to miss thin vessels, and DenseNet and V-GAN depict relatively thick blood vessels for any suspicious area. Different from segmentation results on STARE and DRIVE, IterNet fails to capture vessels around the optic discs on images from HRF dataset. M-GAN and our approach both achieve competitive results on these high-resolution images, where our approach is slightly more sensitive to thin vessels. It could be observed from Table 5 that traditional segmentation method [13] show less satisfied results in terms of precision and sensitivity, although test time is noticeably less than other CNN-based methods. All methods achieve similar level of specificity where M-GAN is 0.003 slightly higher than ours. Different from previous results, AG-UNet is slightly more sensitive than other approaches under HRF database while sensitivity of V-GAN dropped to 0.8196, second only to AG-UNet. Our approach and M-GAN both achieve competitive results where M-GAN achieves the highest accuracy at 0.97 and our approach arrives at the highest precision at 0.8947. Nevertheless, our approach trains for 89 seconds per epoch and tests for 17 s per image which is relatively quicker than M-GAN. In general, all experimental results keep the same level of figures as on previous datasets. Our approach shows advantage on high-resolution database in terms of specificity and precision.

The advantage of our method is mainly due to the proposed 3D modified Attention U-Net architecture and the use of constrained-based NMF, which not only offers highly discriminative features that help us to classify small segments from non-vessel pixels, but also improves global spatial consistency of the results. It can be seen from the experiences that our method outperforms these competitive methods in terms of reasonable and accurate vessels detecting for application purpose.

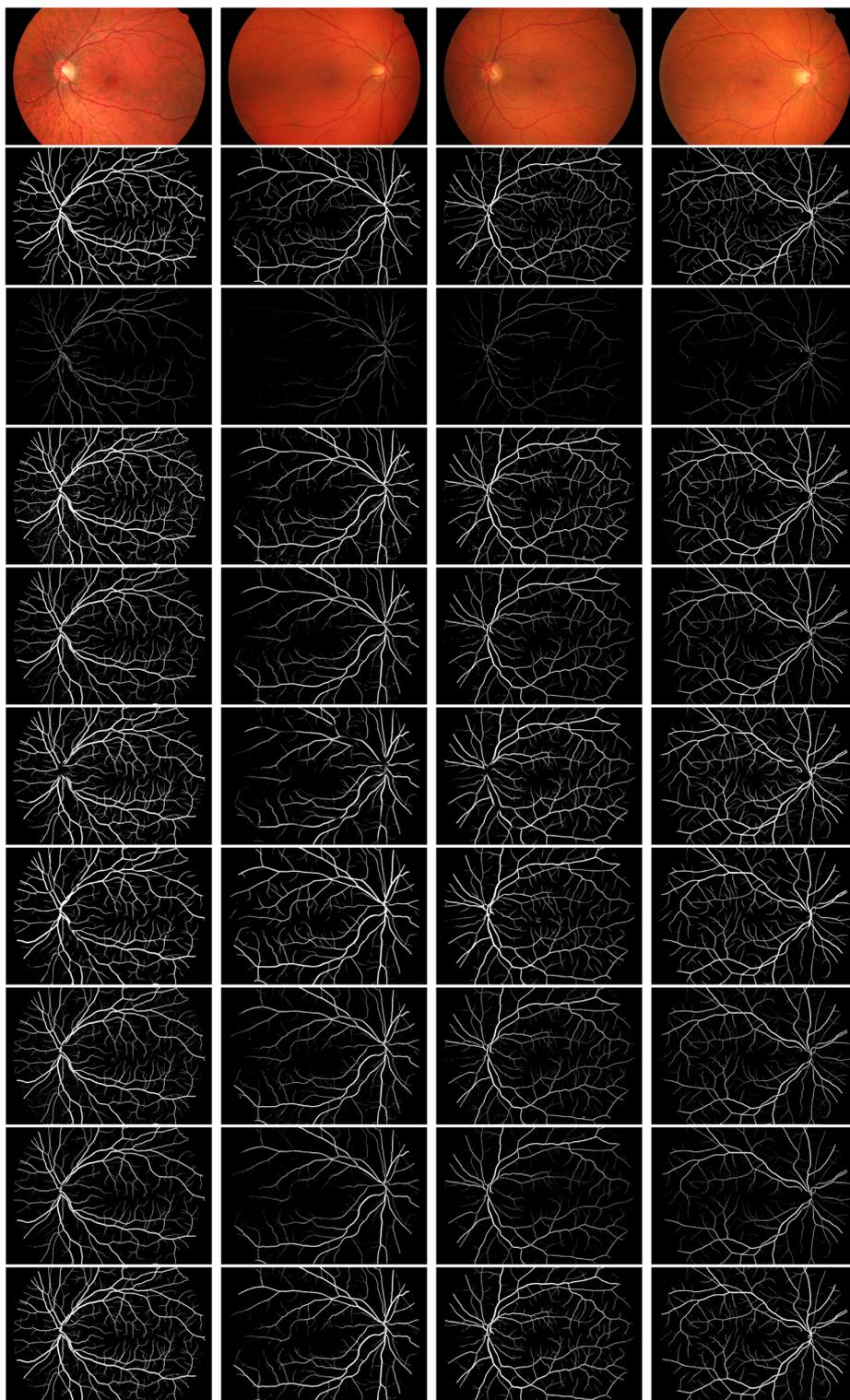


Fig. 13 Vessel segmentation results on HRF dataset, from up to down is input images, ground truth, Soares et al. [13], U-Net, AG-UNet, IterNet, DenseNet, V-GAN, M-GAN, and ours, respectively

Table 5 Performance analysis of all algorithms on HRF databases with respect to the measuring metrics, as well as training time per epoch and prediction time on one image (in second)

Methods	Accuracy	Specificity	Sensitivity	Precision	Training	Prediction
Soares et al. [13]	0.9373	0.9671	0.6253	0.6444	–	6
U-Net [1]	0.9577	0.9701	0.6577	0.7308	12	10
AG-U-Net [35]	0.9600	0.9875	0.8297	0.8268	29	13
IterNet [26]	0.9623	0.9743	0.7524	0.7284	428	53
DenseNet [36]	0.9639	0.9812	0.7610	0.7747	67	22
V-GAN [37]	0.9617	0.9858	0.8196	0.8213	107	32
M-GAN [38]	0.9700	0.9931	0.6948	0.8800	112	36
Ours	0.9688	0.9903	0.7451	0.8947	89	17

5 Conclusions

In this paper, we proposed a 3D modified attention U-Net architecture along with constrained-based NMF to extract retinal vessels accurately especially for thin vessels. The pre-processing steps include gamma correction and region processing to achieve well contrast images for subsequent calculation. Next, we proposed a novel NMF algorithm with within-class and between-class constraints to encode and extract neighborhood feature information of each pixel, while image dimension reducing. Our constrained-based NMF approach also provide a new choice for computer vision research while compressing dimension is necessary. Next, a 3D modified attention U-Net with an upsampling beforehand and a downsampling after the three max-pooling layer is proposed. At the same time, the AGs used in the skip connection highlight useful feature information and suppress irrelevant content. Finally, to measure the effectiveness of the proposed framework, we tested the proposed model on three datasets DRIVE, STARE, and HRF. The obtained results and related comparisons shown that the performances of this proposed scheme were better than most of the exist. The proposed retinal vessel extraction scheme can be extended to other similar vessel segmentation focused tasks such as cardiovascular extraction.

Abbreviations

NMF: Nonnegative matrix factorization; 3D: Three-dimensional; AG: Attention gate; G-channel: Green channel; GMM: Gaussian mixture model; SVM: Support vector machine; CNN: Convolution neural network; AG-Net: Attention guided network; F-norm: Frobenius norm; KKT: Karush-Kuhn-Tucker; FOV: Field of view; SE: Sensitivity; SP: Specificity; ACC: Accuracy; AUC: Area under the curve; ROC: Receiver operating characteristic curve; ROI: Region of interesting

Acknowledgements

Not applicable.

Authors' contributions

Yang Yu implemented the proposed methodology, original draft preparation, software, and validation. Hongqing Zhu took part in writing and general supervision of the final version of this paper. The authors read and approved the final manuscript.

Authors' information

Yang Yu is currently working towards his Ph.D. in East China University of Science and Technology, Shanghai, China, and has received B.S. degree in Electronic Information Science from Jiangsu University of Science and Technology in 2018. His research domains include medical image processing, deep learning, and computer vision. Hongqing Zhu received the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2000. From 2003 to 2005, she was a Post-Doctoral Fellow with the Department of Biology and Medical Engineering, Southeast University, Nanjing, China. She is currently a professor with the East China University of Science and Technology, Shanghai. Her current research interests include medical image processing, deep learning, computer vision, and pattern recognition. She is a member of IEEE and IEICE.

Funding

This work was supported by the National Nature Science Foundation of China under Grant 61872143.

Availability of data and materials

The image datasets used to support the findings of this study can be downloaded from the public websites whose hyperlinks are provided in the article.

Competing interests

The authors declare that they have no competing interests.

Received: 25 August 2020 Accepted: 10 January 2021

Published online: 28 January 2021

References

1. O. Ronneberger, P. Fischer, T. Brox, in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, U-Net: convolutional networks for biomedical image segmentation, (2015), pp. 234–241
2. N. P. Singh, R. Srivastava, Retinal blood vessels segmentation by using Gumbel probability distribution function based matched filter. *Comput. Methods Prog. Biomed.* **129**, 40–50 (2016)
3. J. De, H. Li, L. Cheng, Tracing retinal vessel trees by transductive inference. *BMC Bioinformatics.* **15**(1), 20 (2014)
4. N. Memari, M. I. B. Saripan, S. Mashohor, M. Moghbel, Retinal blood vessel segmentation by using matched filtering and fuzzy c-means clustering with integrated level set method for diabetic retinopathy assessment. *J. Med. Biol. Eng.* **39**(5), 713–731 (2019)
5. D. Kaba, A. G. Salazar-Gonzalez, Y. Li, X. Liu, A. Serag, in *Proceedings of the Health Information Science*, Segmentation of retinal blood vessels using Gaussian mixture models and expectation maximisation, (2013), pp. 105–112
6. Z. Fan, J. Lu, C. Wei, H. Huang, X. Cai, X. Chen, A hierarchical image matting model for blood vessel segmentation in fundus images. *IEEE Trans. Image Process.* **28**(5), 2367–2377 (2019)
7. T. A. Soomro, A. J. Afifi, J. Gao, O. Hellwich, L. Zheng, M. Paul, Strided fully convolutional neural network for boosting the sensitivity of retinal blood vessels segmentation. *Expert Syst. Appl.* **134**, 36–52 (2019)
8. C. Zhu, B. Zou, R. Zhao, J. Cui, X. Duan, Z. Chen, Y. Liang, Retinal vessel segmentation in colour fundus images using extreme learning machine. *Comput. Med. Imaging Graph.* **55**, 68–77 (2017)
9. D. Relan, T. Macgillivray, L. Ballerini, E. Trucco, in *Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Automatic retinal vessel classification using a least square-support vector machine in vampire, (2014), pp. 142–145
10. S. Wang, Y. Yin, G. Cao, B. Wei, Y. Zheng, G. Yang, Hierarchical retinal blood vessel segmentation based on feature and ensemble learning. *Neurocomputing.* **149**, 708–717 (2015)
11. W. Wiharto, E. Suryani, The comparison of clustering algorithms k-means and fuzzy c-means for segmentation retinal blood vessels. *Acta Informatica Med.* **28**, 42 (2020)
12. C.-Y. Lin, C.-Y. Kang, T.-Y. Huang, M.-K. Chang, A novel non-negative matrix factorization technique for decomposition of Chinese characters with application to secret sharing. *EURASIP J. Adv. Signal Process.* **35**, 1–8 (2019)
13. J. Soares, J. J. G. Leandro, R. M. Cesar, H. F. Jelinek, M. J. Cree, Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification. *IEEE Trans. Med. Imaging.* **25**(9), 1214–1222 (2006)
14. C. A. Lupascu, D. Tegolo, E. Trucco, FABC: retinal vessel segmentation using adaboost. *IEEE Trans. Inf. Technol. Biomed.* **14**(5), 1267–1274 (2010)
15. P. O. Hoyer, Non-negative matrix factorization with sparseness constraints. *J. Mach. Learn. Res.* **5**(1457–1469) (2004)
16. M. T. Belachew, N. Del Buono, Robust embedded projective nonnegative matrix factorization for image analysis and feature extraction. *Pattern. Anal. Applic.* **20**(4), 1045–1060 (2017)
17. X. Cai, F. Sun, Supervised and constrained nonnegative matrix factorization with sparseness for image representation. *Wirel. Pers. Commun.* **102**, 3055–3066 (2018)
18. J. Zhang, Y. Rao, J. Zhang, Y. Zhao, Trigraph regularized collective matrix tri-factorization framework on multiview features for multilabel image annotation. *IEEE Access.* **7**, 161805–161821 (2019)
19. S. Baghersalimi, B. Bozorgtabar, P. Schmidaugeon, H. K. Ekenel, J. Thiran, DermoNet: densely linked convolutional neural network for efficient skin lesion segmentation. *EURASIP J. Image Video Process.* **2019**(1), 71 (2019)
20. M. Szkulmowski, P. Liskowski, B. Wieloch, K. Krawiec, B. L. Sikorski, Convolutional neural networks for artifact free OCT retinal angiography. *Investig. Ophthalmol. Vis. Sci.* **58**, 649–649 (2017)
21. S. Guo, K. Wang, H. Kang, Y. Zhang, Y. Gao, T. Li, BTS-DSN: deeply supervised neural network with short connections for retinal vessel segmentation. *Int. J. Med. Inform.* **126**, 105–113 (2019)
22. Y. Guo, U. Budak, L. Vespa, E. S. Khorasani, A. Sengur, A retinal vessel detection approach using convolution neural network with reinforcement sample learning strategy. *Measurement.* **125**, 586–591 (2018)
23. B. Wang, S. Qiu, H. He, in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Dual encoding U-net for retinal vessel segmentation, (2019), pp. 84–92
24. B. J. Bhatkalkar, D. R. Reddy, S. Prabhu, S. V. Bhandary, Improving the performance of convolutional neural network for the segmentation of optic disc in fundus images using attention gates and conditional random fields. *IEEE Access.* **8**, 29299–29310 (2020)
25. S. Zhang, H. Fu, Y. Yan, Y. Zhang, Q. Wu, M. Yang, M. Tan, Y. Xu, in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Attention guided network for retinal image segmentation, (2019), pp. 797–805
26. L. Li, M. Verma, Y. Nakashima, H. Nagahara, R. Kawasaki, in *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, IterNet: Retinal Image Segmentation Utilizing Structural Redundancy in Vessel Networks, (Snowmass Village, 2020), pp. 3645–3654
27. S. Kansal, R. K. Tripathi, Adaptive gamma correction for contrast enhancement of remote sensing images. *Multimed. Tools Appl.* **78**(18), 25241–25258 (2019)
28. D. D. Lee, H. S. Seung, Learning the parts of objects by nonnegative matrix factorization. *Nature.* **401**, 788–791 (1999)
29. H. W. Kuhn, Nonlinear programming: a historical view. *Traces Emergence Nonlinear Program.* **31**, 393–414 (2013)

30. J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, B. Van Ginneken, Ridge-based vessel segmentation in color images of the retina. *IEEE Trans. Med. Imaging.* **23**(4), 501–509 (2004)
31. A. D. Hoover, V. L. Kouznetsova, M. H. Goldbaum, Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Trans. Med. Imaging.* **19**(3), 203–210 (2000)
32. A. Budai, R. Bock, A. Maier, J. Hornegger, G. Michelson, Robust vessel segmentation in fundus images. *Int. J. Biomed. Imaging.* **2013**, 1–12 (2013)
33. E. R. Delong, D. R. Delong, D. L. Clarkepearson, Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* **44**(3), 837–845 (1988)
34. D. P. Kingma, J. L. Ba, in *Proceedings of the International Conference on Learning Representations (ICLR)*, vol. 12, Adam: a method for stochastic optimization, (San Diego, 2015), pp. 1–15
35. J. Schlemper, O. Oktay, M. Schaap, M. P. Heinrich, B. Kainz, B. Glocker, D. Rueckert, Attention gated networks: learning to leverage salient regions in medical images. *Med. Image Anal.* **53**, 197–207 (2019)
36. G. Huang, Z. Liu, L. V. Der Maaten, K. Q. Weinberger, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Densely connected convolutional networks, (2017), pp. 2261–2269
37. J. Son, S. J. Park, K. H. Jung, Retinal vessel segmentation in fundoscopic images with generative adversarial networks. arXiv preprint arXiv:1706.09318 (2017)
38. K.-B. Park, S. H. Choi, J. Y. Lee, M-gan: Retinal blood vessel segmentation by balancing losses through stacked deep fully convolutional networks. *IEEE Access.* **8**, 146308–146322 (2020)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
