Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Level set guided region prototype rectification network for retinal vessel segmentation

Yifei Liu [a], Qingtian Wu [b], Xueyu Liu [a], Junyu Lu [a], Zhenhuan Xu [a], Yongfei Wu [a,*], Shu Feng [c,*]

[a] *College of Data Science, Taiyuan University of Technology, Taiyuan, Shanxi, 030024, China*
[b] *Faculty of Science and Technology, University of Macau, Taipa, Macao Special Administrative Region of China*
[c] *Department of Foundation, Shanxi Agricultural University, Taigu, Shanxi, 030801, China*

ARTICLE INFO

Keywords:
Retinal vessel segmentation
Level-set
Region prototype rectification
Self-supervised and supervised loss

ABSTRACT

Retinal vessel segmentation refers to extracting the vessel region with continuous and smooth boundaries from retinal images, which is of great significance in clinical practices. However, due to the weak and blurry edges of targets as well as interference (such as optic cup and disc) in the background, current deep neural network-based methods struggle in extracting features with discriminative semantics while preserving continuous edges. To enforce continuous predictions of weak edges, we propose a level set guided region prototype rectification (LSRPR) framework and a novel level set loss (LS-loss) with learnable and self-guided mechanisms. Specifically, the LSRPR firstly takes features of the last layer from the decoders of a U-Net version as input and rectified the region prototype by an auxiliary self-supervised level set loss, then the pre-trained model is fine-tuned by using supervised level set loss. The LS-loss facilitates the model to generate reliable guidance and enhances the continuous of edges among the decoders of neural network model. The proposed method is simple, yet effective, which can easily be extended to other frameworks. The quantitative and qualitative experimental results on public retinal vessel datasets indicate the effectiveness of the region prototype rectification compared to other SOTA models. Our code is available at Github:https://github.com/tweedlemoon/LSRPR.

1. Introduction

Accessing and analyzing the structural information of the retinal vessel are indispensable for healthcare professionals to understand and diagnose the related diseases. For instance, research [1] revealed that severe cardiovascular disorders of children may be linked to the retinal arteriolar tortuosity, and microcirculatory abnormalities in the retinal vasculature are often found in hypertension and diabetes mellitus according to [2]. Therefore, automatic and accurate segmentation of retinal vessel images have the potential to gather structural information and quickly assist professional in making decision.

In the literature, various methods and algorithms have been proposed to solve the segmentation task of retinal vessel, ranging from the classical level set method, graph cut and convolution neural networks (CNNs). CNNs have shown superior performance in many image segmentation tasks benefitting from its powerful capability of learning informative multi-level hierarchical features from data itself. Fully convolution networks (FCNs) [3] are first proposed and used to segment objects in early time. Based on the FCN network, Ronneberger et al. proposed the U-Net [4] dedicated to medical image segmentation by designing encoder–decoder structures with a skip connection between them to fuse features of encoder and decoder. Afterwards, a large number of improvements for fine-tuning U-Net and combining them with other methods related to machine learning have been proposed to apply to various image segmentation task [5–7]. For the retinal vessel segmentation, some techniques are employed to avoid overfitting and improve the segmentation performance of U-Net in handling such small dataset, e.g., dropblock [8] in S-UNet [9] proposed by Hu et al. and SA-UNet [7] proposed by Guo et al. Although many customized CNN models [10,11] have been proposed to extract the edges of retinal vessel from the eye fundus image, it is rather difficult to precisely locate the edges of vessel because of complex tubular structure of retinal vessels and low imaging contrast. The CNN traning only considers each pixel information independently and is hard to learn the advanced abstract concepts (such as edges continuation and object topology) at the object level. Fig. 1 shows that the CNN based methods struggle in extracting features with discriminative semantics while preserving continuous edges of vessels. Active contour model (ACM) or level set (LS) methods are the well-known approaches which can embody the

[*] Corresponding authors.
E-mail addresses: eflier.liu@gmail.com (Y. Liu), wuyongfei@tyut.edu.cn (Y. Wu), fengshu@sxau.edu.cn (S. Feng).

https://doi.org/10.1016/j.bspc.2023.105428
Received 21 April 2023; Received in revised form 10 July 2023; Accepted 12 September 2023
Available online 20 September 2023

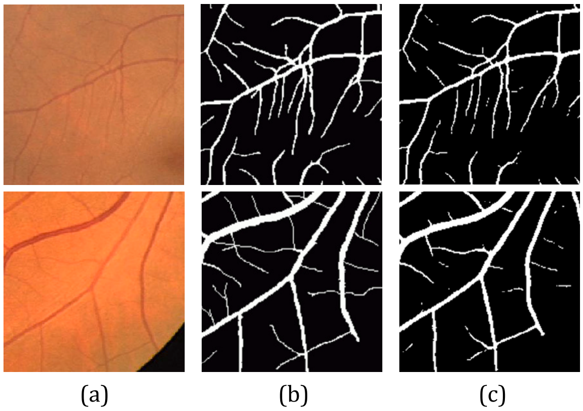

Fig. 1. CNN based methods struggling in extracting features with discriminative semantics while preserving continuous edges of vessels. (a) Vessel image; (b) Ground truth; (c) Attention U-Net's [5] segmentation result.

high level concepts of object shapes and guarantee the smoothness of contours and a consistent topology [12].

Recently, many methods have been developed to improve the segmentation performance via combining ACM technology into CNN framework. Among these methods, a novel loss function derived from the energy functional of ACM, such as Mumford-Shah loss [13], AC loss [14,15] and level set loss [16], is skillfully designed to learn the compact and uniform edge maps. Meanwhile, some researchers also attempted to incorporate ACM or LS into deep learning frameworks in an end-to-end fashion. For examples, Hu et al. [17] proposed a deep level set network to produce salience maps for detecting salient objects more precisely. DSAC [18], DCAC [19] and DARNet [20] trained a CNN to learn ACM parameter initialization automatically. In the field of medical image segmentation, Le et al. [21] used level set combined with multiplexed FCNs to obtain better results on the brain tumor dataset for medical treatment. Chen et al. [22] proposed the AC Loss function using DenseNet as a deep learning network. Ma et al. [23], using VNet as the backbone, combined the level set loss in conjunction with the geodesic loss and completed the segmentation of datasets such as LiTs. Le et al. [24] improved on the traditional level set by proposing a residual level set, which also achieved a better performance on the saliency detection task. Hatamizadeh et al. [25] developed a novel architecture called as DALS for segmenting lesions in medical imaging by combining ACM and CNN. Wang et al. [26] used level set information as the prior information to do liver segmentation. Under general circumstances, level set methods aforementioned can provide prior information and add interpretability to results. However, these methods have corresponding limitations more or less, such as need to fine tuning of parameters in the level set formulation [17], sensitive to the manual initialization of target contours [18,20], and trainable of ACM module [25].

To alleviate the limitations of CNN, we designed a level set guided regional prototype rectification (LSRPR) framework to correct the regional information and produce continuous edges for retinal vessel segmentation in this study. Specifically, we present a two-stage structure which uses the classic Attention U-Net as the CNN backbone, and a novel level set loss as auxiliary function to enhances the performance of the networks. In particular, we use level set loss in self-supervised way to speed up the convergence of networks in the first stage, and fine-tune the networks by using the level set loss as supervised one in the second stage. The proposed LSRPR method is trainable and free of fine tuning parameters. Experimental results on the three public retinal vessels datasets (i.e. DRIVE, CHASE_DB1 and RITE) verify that the proposed LSRPR can improve the segmentation performance by specifying the continuity of vessels when compared to state-of-the-art deep learning segmentation networks. In summary, the main contributions of this work are listed as follows:

- A level set guided regional prototype rectification framework (LSRPR) is presented to generates reliable guidance and enhances the continuous of edges among the decoders of neural network model.
- A two-stage learning pipeline is designed, in which a self-supervised learning of level loss is utilized in the first stage to speed up the training of network and a supervised learning of level loss is employed in the second stage to fine-tune the segmentation accuracy of network.
- We formulate two expressions for the self-supervised level set loss and supervised one, respectively.
- The proposed LSRPR obtains state-of-the-art performance on three public retinal vessels segmentation datasets. More generally, it can easily be extended to other frameworks to improve boundary accuracy.

The rest of this paper is structured as follows. In Section 2, we give brief review for some works which are related with our work. Our proposed segmentation framework for retinal vessels is presented in Section 3. Section 4 provides experimental results to verify the performance of the proposed method including ablation study and comparison with state-of-the-art method. In the end, we draw conclusion about this paper in Section 5.

2. Related work

In this section, we briefly review level set method, CNN for retinal vessel segmentation and combination of CNN with level set that related with our proposed retinal vessel segmentation method.

2.1. Level set method

The level set (LS) method [27] was widely applied in image segmentation with active contour [28], due to its capability of automatically handling various topological changes (splitting and merging). The basic idea is to define an implicit function in a higher dimension to represent contours as the zero level set. The function is referred as level set function and evolved according to a partial differential equation (PDE) derived from a Lagrangian formulation of variational active contour model. Fig. 2 shows an example of how level set works.

Traditionally, the energy functional with respect to level set ϕ is usually defined based on the difference of image features, such as intensity, color and texture, between foreground and background. In [28], Chan and Vese proposed a well-known variational level set model for image segmentation as

$$\begin{aligned} L_{CV}(\phi) = \lambda_1 & \int_{\Omega} (I(x) - c_1)^2 H(\phi(x)) dx \\ & + \lambda_2 \int_{\Omega} (I(x) - c_2)^2 (1 - H(\phi(x))) dx \\ & + \mu \int_{\Omega} |\nabla H(\phi(x))|, \end{aligned} \tag{1}$$

where μ, λ_1 and λ_2 are positive parameters, ϕ is a level set function, and H is the Heaviside function. The two piecewise constants c_1 and c_2 are defined as

$$c_1 = \frac{\int_{\Omega} I(x) H(\phi(x)) dx}{\int_{\Omega} H(\phi(x)) dx}, \quad c_2 = \frac{\int_{\Omega} I(x)(1 - H(\phi(x))) dx}{\int_{\Omega} (1 - H(\phi(x))) dx}. \tag{2}$$

which represent the mean intensity values of I in $\Omega_1 = \{x \in \Omega | \phi(x) > 0\}$ and $\Omega_2 = \{x \in \Omega | \phi(x) < 0\}$, respectively.

Then, gradient descent method is applied to minimize an energy function and update the value of level set function. Under the circumstances, information such as shape and regions can be integrated into the energy functional to enhance the segmentation performance of level set. However, it is difficult to exploit more high level features to deal with complex images, which limited the application of level set segmentation in practice.

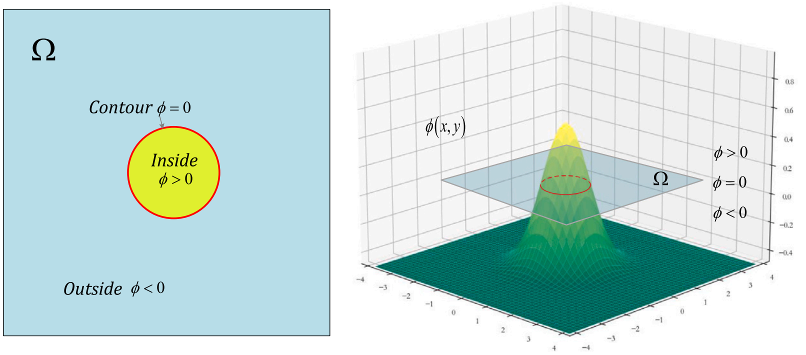

Fig. 2. How level set function works. The mountain-like shape denotes the function value of ϕ, and Ω denotes the whole image. $\phi(x, y) > 0$ denotes the pixel (x, y) belongs to the foreground. The red line, where $\phi(x, y) = 0$ denotes the contour of the object.

2.2. CNN-based segmentation

Convolutional neural networks for image processing can be traced back to Alexnet [29], which is a classical network for image classification that dramatically improves classification performance. Afterwards, Long et al. [3] proposed full convolution neural network (FCN) by superseding all the fully connected layers of CNN with convolutional layers, which realizes pixel-level image segmentation and promotes the improvement of segmentation accuracy. On the basis of FCN, U-Net introduced by Ronneberger et al. [4] has been preferably studied and improved. It employed the encoder–decoder structure consisting of a contracting branch and an expanding branch in the semantic segmentation domain, and added a skip connection to the encoder and the decoder to fuse low level semantic features.

For the retinal vessel segmentation, increasingly complex and sophisticated CNN based on U-Net version have been pushing performance of segmentation retinal vessel. Jiang et al. [30] used a migration learning approach to adapt the task to regional vessel element identification and result merging, improving segmentation accuracy. Guo et al. [31] proposed a multiscale deeply supervised short connection network (BTS-DSN) for vessel segmentation. Alom et al. fused RCNN into U-Net to create R2UNet [32]. Guo et al. created the SD-UNet [33] by incorporating the Dropblock proposed by Ghiasi et al. [8] in the UNet structure. Hu et al. first reduced the UNet to a structure called MiUNet and then cascaded them to obtain the S-UNet [9]. Jin et al. created DUNet [6] which was inspired by Dai et al.'s variable convolutional neural network [34]. Alfonso Francia et al. [35] linked the two U-Nets to obtain a cascaded segmentation model. Guo et al. [36] added the spatial attention block to S-UNet to create SA-UNet. Valanarasu et al. [11] proposed KiUNet inspired by KiteNet. Dong et al. [37] proposed CRAUNet under the idea of Zhao's saliency segmentation [38]. Wang et al. [39] used CRF-RNN to extend U-Net and proposed an EE-UNet to segment optic disc and cup in fundus images. However, the features extracted by convolution in retinal vessel segmentation by the U-Net versions contain only local contextual information, which is limited for capturing long-range dependencies and can easily lead to incorrect semantic understanding in the continuity of vessel segmentation tasks.

2.3. Combination of CNN with level set

To combine the advantages of LS and CNN network, some researchers have develop many techniques to integrate the LS into CNN, achieving good segmentation performance on several datasets.

Hu et al. [17] combined the level set and super-pixel approach into CNN to process the saliency detection task, achieving relatively good performance on several datasets. To incorporate the deep convolutional network with the level set method, a energy functional about level set ϕ is defined as

$$L = \alpha \left[\int_{\Omega} \left| H(\phi(x)) - c_1 \right|^2 H(\phi(x))\, dx \right] + \beta \left[\int_{\Omega} \left| H(\phi(x)) - c_2 \right|^2 (1 - H(\phi(x)))\, dx \right], \tag{3}$$

where Ω denotes pixel space of the input image, the saliency values output by the CNN is linearly shifted into $[-0.5, 0.5]$ and treat it as the level set ϕ.

Constants c_1 and c_2 refer as average saliency values of inside(ϕ) and outside(ϕ), respectively. Keeping ϕ fixed and minimizing the energy function with respect to c_1 and c_2, these two constants can be expressed as

$$c_1 = \frac{\int_{\Omega} H(\phi(x))\, H(\phi(x))\, dx}{\int_{\Omega} H(\phi(x))\, dx}, \quad c_2 = \frac{\int_{\Omega} H(\phi(x))\, (1 - H(\phi(x)))\, dx}{\int_{\Omega} (1 - H(\phi(x)))\, dx}. \tag{4}$$

The proposed level set based loss function can be incorporated into the convolutional network in an end-to-end fashion to detect salient objects, which can handles object boundaries more accurately.

Inspired by the linkage of Heaviside function and Softmax function, Kim et al. [13] replaced the Heaviside function with softmax output of the last CNN layer and intergraded the Mumford-Shah energy functional into CNN framework, improving the segmentation accuracy and giving the model some unsupervised learning capability. Specifically, for two-phase segmentation, the softmax output of two feature channels at the last CNN layers is computed by

$$y_1(x) = \frac{e^{z_1(x)}}{e^{z_1(x)} + e^{z_2(x)}}, \quad y_2(x) = \frac{e^{z_2(x)}}{e^{z_1(x)} + e^{z_2(x)}}, \tag{5}$$

where $x \in \Omega$, and $z_1(x)$ and $z_2(x)$ denote the network output at x from the preceding layer before the softmax. The output values of Eq. (5) of $z_1(x)$ and $z_2(x)$ are close to 1 when the pixel value at x belongs to the respective class.

Based on the Eq. (5), the Mumford-Shah functional is modified as the following loss function of CNN:

$$L_{MScnn}(\Theta; I) = \int_{\Omega} \left| I(x) - c_1 \right|^2 y_1(x; \Theta)\, dx + \int_{\Omega} \left| I(x) - c_2 \right|^2 y_2(x; \Theta), \tag{6}$$

where $I(x)$ is the original image, $y_1(x; \Theta)$ and $y_2(x; \Theta)$ is the output of softmax layer in Eq. (5), Θ refers as the learnable network parameters.

The two constants c_1 and c_2 are computed by

$$c_1 = \frac{\int_{\Omega} I(x)\, y_1(x; \Theta)\, dx}{\int_{\Omega} y_1(x; \Theta)\, dx}, \quad c_2 = \frac{\int_{\Omega} I(x)\, y_2(x; \Theta)\, dx}{\int_{\Omega} y_2(x; \Theta)\, dx}. \tag{7}$$

which represent the average value of pixels that belong to the foreground and background, respectively.

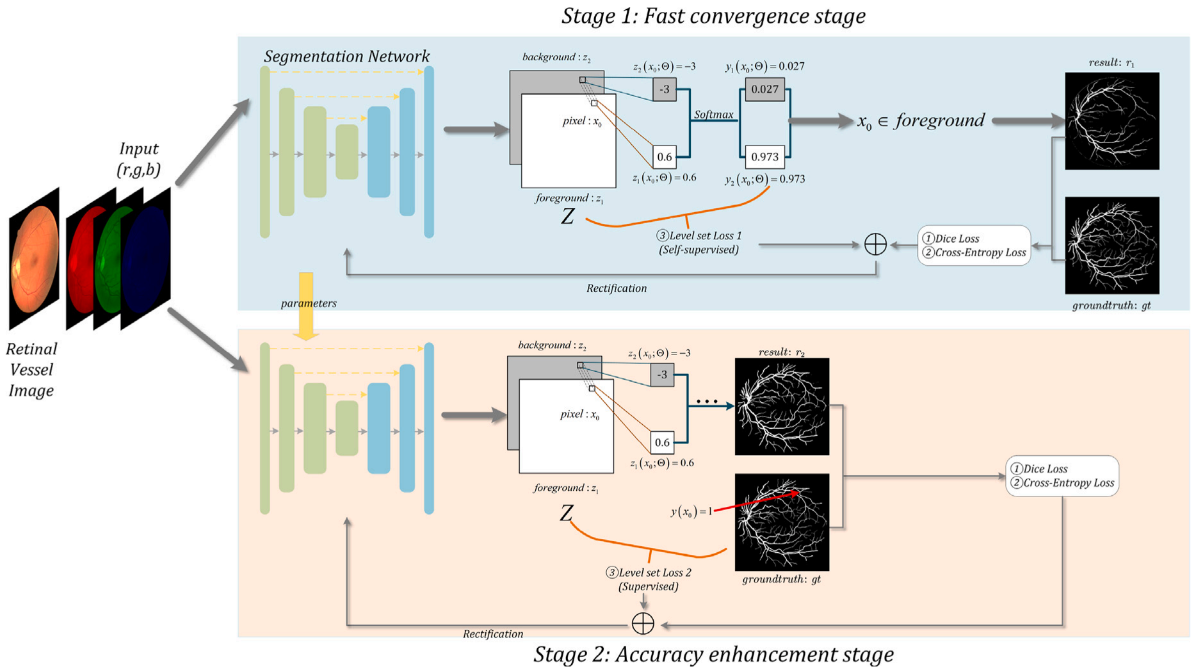

Fig. 3. The overall architecture of our LSRPR method.

As previously mentioned, even if the level set method has self-supervised learning capability, the simple level set method struggles to accurately model complex images and can produce wildly divergent results due to different initializations. By using the softmax layer in CNN as a differentiable approximation of the characteristic function, the loss function (6) can be minimized by back-propagation in the training process. Therefore, the combination of level set method and CNN can avoid the randomness of the results of level set method and give CNN prior knowledge and the capability of self-supervised learning. For a better fusing level set method with CNN, we will develop a similar strategy in our network and describe it in detail below.

3. Methodology

In this section, we first introduce the overall architecture of our proposed LSRPR model, and then describe the stages of fast convergence and accuracy enhancement in the model. Finally, a self-supervised and supervised level set loss combined with Dice and cross-entropy (CE) loss functions are designed to train the model.

3.1. Overview of model architecture

Fig. 3 shows the overall architecture of our proposed LSRPR model. The upper part in the blue background is the fast convergence stage, which use a self-supervised loss derived from level set to guide the network quickly convergent. Similarly, the lower part in the orange background is the accuracy enhancement stage by using supervised level set loss to fine-tune the segmentation network. In these two stages, the self-supervised and supervised level set loss functions are constructed based on the output feature maps of backbone network. In this work, we employ the Attention U-Net [5] proposed by Oktay et al. as the backbone network via experimental verification (see Section 4.3). Due to introducing the attention gate (AG) mechanism in skip connection between the encoder and decoder, Attention U-Net suppresses irrelevant regions compared to previous work, highlights the features of the task, and improves recognition accuracy. Next, we will systematically explain the implementation detail of the two stages .

3.2. Fast convergence stage

To make the network quickly convergent, a initial model that can basically segment the retinal vessel image is obtained via a auxiliary level set loss function in self-supervised way to guide the rectification of region prototype in the this stage. Specifically, an original retinal vessel image are fed into the backbone network and output the last feature mappings with quantified channel numbers related with segmentation task. For the binary segmentation of retinal vessel, the channel of the last feature mappings of network is set to 2, in which one feature mapping is responsible for predicting the foreground and the other is in charge of predicting the background. Without loss of generality, we name one of feature mappings as $z_1(x)$ and the other as $z_2(x)$. For the features mapping $z_1(x)$ and $z_2(x)$, both of them do not embody region information due to the limitation of CNN. To make the network learning more high abstract conception (such region prototype), we use level set method to obtain the prototypes of foreground and background.

Accordingly, we design the following level set loss function based on the network output before using softmax as

$$L_{LSself}(\Theta) = \alpha \int_{\Omega} \left|z_1(x;\Theta) - c_1\right|^2 y_1(x;\Theta)\,dx + \beta \int_{\Omega} \left|z_2(x;\Theta) - c_2\right|^2 y_2(x;\Theta)\,dx, \tag{8}$$

where α and β are two positive parameters, Ω is the full domain of the feature mapping, $y_1(x;\Theta)$ and $y_2(x;\Theta)$ is the softmax output of feature mapping for the last layers of the neural network.

The prototypes c_1 and c_2 of foreground and background can be computed as

$$c_1 = \frac{\int_{\Omega} z_1(x;\Theta)\, y_1(x;\Theta)\,dx}{\int_{\Omega} y_1(x;\Theta)\,dx}, \quad c_2 = \frac{\int_{\Omega} z_2(x;\Theta)\, y_2(x;\Theta)\,dx}{\int_{\Omega} y_2(x;\Theta)\,dx}. \tag{9}$$

It is worth noting that the loss computation of level set in this stage is free of ground truth, and thus it has the self-supervised learning capability. With the help of Eqs. (8) and (9), The prototypes of two feature maps in the last network layer can be rectified and iteratively make the positive/negative pixels judged by the network closer to the average of all positive/negative pixels' values predicted by the network. In other words, the proposed unsupervised level set loss can help the network to fast cluster. From this perspective, the proposed level set loss can guide the network to converge in the right direction, thus boosting the fast convergence of the network. We will give ablation studies and explanations in Section 4.3.3.

3.3. Accuracy enhancement stage

The trained network in the fast convergence stage just identify the pixels that can be easily classified as the foreground or background due to the utilization of self-supervised learning strategy. In this circumstance, the pixels with uncertainty may be misclassified. To improve the classification accuracy of the network, we propose accuracy enhancement stage, in which we define a supervised level set loss via combining the ground truth. Specifically, in the accuracy enhancement stage, we keep training the network parameters that are obtained in the fast convergence stage in a supervised fashion.

Accordingly, the level set loss is formulated as

$$L_{LS}(\Theta) = \alpha \int_{\Omega} \left|z_1(x;\Theta) - c_1\right|^2 y(x)\, dx + \beta \int_{\Omega} \left|z_2(x;\Theta) - c_2\right|^2 (1 - y(x))\, dx, \tag{10}$$

Similarly, the two prototypes c_1 and c_2 can be computed as

$$c_1 = \frac{\int_{\Omega} z_1(x;\Theta)\, y(x)\, dxdy}{\int_{\Omega} y(x)\, dx}, \quad c_2 = \frac{\int_{\Omega} z_2(x;\Theta)(1 - y(x))\, dx}{\int_{\Omega} (1 - y(x))\, dx}. \tag{11}$$

where $y(x)$ represents the ground truth.

Different from the fast convergence stage, the accuracy enhancement stage employs the ground truth to calculate the level set loss. By means of the supervised information, the network can focus on the pixels with uncertainty and achieve accuracy improvement. Specifically, based on Eqs. (10) and (11), the prototypes of two feature maps in the last network layer can be fine-tuned and are closer to the average values of two classes computed from the ground truth. Here, the proposed level set loss just plays an auxiliary correction role, which can identify the information missed by the network and improve the output of the network, e.g. eliminating idiosyncratic pixel points. For more information and visualization, we give corresponding ablation study and explanation in Section 4.3.2.

3.4. Total loss function

In the proposed method, the defined self-supervised and supervised level set loss function are employed to rectify the region prototypes. In order to optimize the model in the direction of the correct outcome and learn more generalized representations, it is necessary to impose additional supervision on the network training by using the standard Dice loss and cross-entropy loss.

The Dice loss and cross-entropy loss has the following formulation:

$$L_{Dice} = 1 - \frac{2\left|P \cap Y\right|}{|P| + |Y|} = 1 - \frac{2\sum(P \cdot Y)}{\sum P + \sum Y}, \tag{12}$$

$$L_{CE} = -\left[Y \log P + (1 - Y)\log(1 - P)\right], \tag{13}$$

where P and Y denote the predicted results and ground truth, respectively.

Therefore, the overall definition of the loss function for two stages of the model are as follows:

$$L_{stage1} = \chi_1 L_{Dice} + \chi_2 L_{CE} + \chi_3 L_{LSself}, \quad L_{stage2} = \lambda_1 L_{Dice} + \lambda_2 L_{CE} + \lambda_3 L_{LS}. \tag{14}$$

where the hyper-parameters $\chi = \left(\chi_1, \chi_2, \chi_3\right)$ and $\lambda = \left(\lambda_1, \lambda_2, \lambda_3\right)$ are utilized to ensure the training stability and effectiveness of two stages. The specific values of these hyper-parameters are given in Section 4.1.3.

4. Experiment and result

In this section, we first valid and evaluate our proposed method on three public dataset of retinal vessel images including DRIVE [40], CHASE_DB1 [41] and RITE [42], and then conduct a comprehensive ablation study to verify the impact of each part of our proposed method. Finally, we conduct the comparison experiment with state-of-the-art methods to show the performance of our proposed method.

4.1. Dataset and experimental setting

4.1.1. Dataset

We evaluate the performance of the proposed method on three public retinal vessel datasets, namely DRIVE, CHASE_DB1 and RITE. The DRIVE dataset contains 40 images with a size of 565 × 584 and is divided into training and testing sets originally, of which 20 are the training set and 20 are the test set. Each image in the test set is labeled by two experts and provided with a mask to focus on the segmentation of blood vessels in the region. In the DRIVE dataset, we use the training set for training and the annotation results of expert 2 to calculate the evaluation data. The CHASE_DB1 dataset contains 28 images with the image size of 999 × 960. The dataset is not divided into training and testing sets. To prove the generalization capacity of our proposed model, we select 75% (21 images) randomly for training and the rest 25% (7 images) for testing for 10 times, and we averaged the annotation results of the 7 images to calculate the evaluation data. The RITE dataset is divided into training set, test set, and validation set originally, which contains 80 images that serve as the training set and 10 images that serve as the test set and 10 images that serve as the validation set, all of which have the size of 512 × 512.

4.1.2. Evaluation criteria

For the semantic segmentation task based on CNN, pixels usually are classified into target or background by predicting the classification results of all pixels in the image. In the retinal vessel segmentation, the vessel pixels in the retinal fundus image are deemed as the target class, and other nonvessel pixels are set as the background class. We could calculate four basic indexes which include false negative (FN), false positive (FP), true negative (TN) and true positive (TP) by comparing the prediction results of the semantic segmentation model with the ground truth segmentation. We use four metrics to evaluate the performance of our proposed method: sensitivity (Sen), accuracy (Acc), F1-score (F1) and mean intersection over union (mIoU). The four criteria are calculated as follows:

$$Sen = \frac{TP}{TP + FN}, \tag{15}$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}, \tag{16}$$

$$F1 = \frac{2TP}{2TP + FN + FP}, \tag{17}$$

$$mIoU = \frac{1}{k+1} \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{i=0}^{k} p_{ij} + \sum_{i=0}^{k} p_{ji} - p_{ii}}. \tag{18}$$

Among these criteria, Sen indicates the ratio of correctly predicted vessel pixels in all true vessel pixels, representing the predictive power of positive cases. Acc indicates the ratio of correctly predicted vessels and background pixels in all pixels, representing the overall accuracy rate. F1 is an overall indicator that is defined as the harmonic average of precision and recall, representing a complete and accurate search for blood vessels. mIoU takes into account a combination of positive and negative examples and represents the predictive power of the model for both aspects. In Eq. (18), k denotes the class number excluding the background, which is 1 in this task. p_{ij} denotes the number of pixels predicted by the network which should have been class i but are incorrectly predicted into class j. Similarly, p_{ii} denotes the number of pixels which be well predicted into class i.

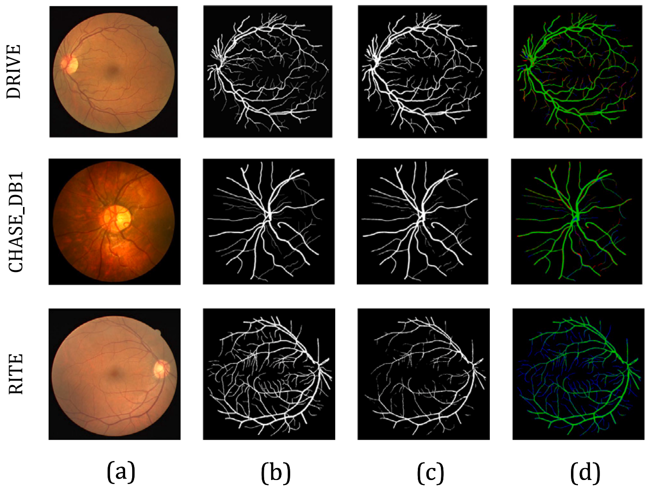

Fig. 4. Segmentation results of examples taken from three datasets by our proposed method. (a) original images; (b) ground truths; (c) corresponding segmentation results; (d) mixed results of ground truth and prediction result.

Table 1
Segmentation performance of our proposed LSRPR method on three public datasets.

Dataset	Sen	F1	Acc	mIoU
DRIVE	0.8342	0.8430	0.9735	0.8502
CHASE_DB1	0.8347	0.8102	0.9724	0.8190
RITE	0.6329	0.7566	0.9655	0.7861

4.1.3. Implementation details

Our proposed method is implemented in Pytorch 1.8.2+ with python 3.7+, and all of experiments are conducted on Linux sever with an NVIDIA Tesla V100, 32 GB GPU. To avoid overfitting, we used random cropping and padding operations to augment the dataset and rescale the image to certain size by making the input size 480 × 480 for DRIVE, 960 × 960 for CHASE_DB1, and 480 × 480 for RITE. Additionally, there is a 50% chance of flipping the images up and down as well as a 50% chance of flipping them left and right in training process of model. The augmented images are fed into the model for network training. We train independently between the three datasets without any pre-trained model. In our experiments, we found that when the batch size was set to 2, our video memory footprint was only 3–5 GB. For the parameters setting, we set the batch size as 4. The hyper-parameters of Eq. (14) are empirically set as $\chi_1 = \lambda_1 = 1.0$, $\chi_2 = \lambda_2 = 1.0$, and $\chi_3 = \lambda_3 = 10^{-6}$. We use SGD optimizer with a momentum of 0.9 to solve the oscillation problem. The learning rate is set to 0.01 and decreases to the 10^{-6} level with the epoch increasing dynamically. The training epoches is set as 100 in the fast convergence stage and 200 in the accuracy enhancement stage, respectively. We select the Attention U-Net [5] as our backbone network by experimentally comparing with other backbone networks (see Section 4.3.1).

4.2. Segmentation performance of the proposed method

The segmentation performances of the proposed method on the three datasets have been listed in Table 1. We can observe that the values of Sen, F1, Acc, and mIoU for our model achieve 0.8342/0.8347/0.6329, 0.8430/0.8120/0.7566, 0.9735/0.9724/0.9655, 0.8502/0.8190/0.7861 on the DRIVE, CHASE_DB1 and RITE datasets, respectively.

Table 2
Performance of different backbones with/without our two-stage LS loss on DRIVE dataset.

Method	Year	F1	Acc	mIoU
U-Net [4]	(2015)	0.8142	0.9681	0.8291
U-Net+LS		0.8422(↑)	0.9688(↑)	0.8323(↑)
R2-UNet [32]	(2018)	0.8266	0.9711	0.8384
R2-UNet+LS		0.8295(↑)	0.9708(↓)	0.8389(↑)
Attention U-Net [5]	(2018)	0.8420	0.9687	0.8310
Attention U-Net+LS		**0.8430**(↑)	**0.9735**(↑)	**0.8502**(↑)
SA-UNet [7]	(2021)	0.8241	0.9697	0.8343
SA-UNet+LS		0.8339(↑)	0.9714(↑)	0.8423(↑)

Fig. 4 illustrates the visual segmentation results of one case taken from each dataset, in which the original images, ground truths, corresponding segmentation results and overlapped results of ground truths and segmentation results are displayed in the first column, the second column, the third column and the fourth column, respectively. For the overlapped results, green pixels mean true positive predictions, red pixels indicate false positive predictions, and blue pixels represent false negative predictions. As can be obviously seen, the vessels can be well segmented, showing that our proposed model can accurately segment the vessels without subjecting to the influence of the optic disc and cup. Additionally, our proposed model can segment vessels with weak boundaries and produce more continuous segmentation boundaries due to the utilization of level set. However, there are still some misclassified pixels, which is connected to the functionality of the backbone network. We discovered that in the DRIVE dataset, the prediction error points were typically located at the ends of the vessels, and the network's predictions of the vessels were frequently thicker than the actual vessels. This finding may be related to the difficulty in extracting the tiny edge information. The fact that the blood vessels and background in the original image of the RITE dataset have a low contrast may be the reason why there are more positive examples of prediction errors in the RITE dataset.

4.3. Ablation study

To investigate the influence of each branch (backbone, self-supervised level loss and supervised level set loss) of our proposed LSRPR on the segmentation results, we perform ablation experiments on DRIVE dataset.

4.3.1. Backbone selection

In this section, we conduct several comparative experiments with different backbones to find the best backbone. In the comparison experiments, we select U-Net, R2-UNet, Attention U-Net, and SA-UNet as backbones to complete the segmentation task, which are well-known backbone networks for medical image segmentation. All of them are individually trained against the DRIVE dataset before embedding them into LSRPR for training. Table 2 shows the segmentation performance of different backbones with or without our two-stage LS loss on DRIVE dataset. As can be clearly seen, the Attention U-Net obtains the best segmentation performance compared to other networks. Moreover, the performance of Attention U-Net can be improved when combining the level set loss into the network training. Specifically, the Attention U-Net combined level set increases the F1 value by 0.1% to 3%, the accuracy by 0.01% to 1.5%, and the mIoU by 0.6% to 1.9%. In addition, the results demonstrate that not all of backbone networks can benefit from level set loss, e.g. R2-UNet shows that the accuracy value decrease when adding LS loss. However, level set loss could produces better outcomes in most of cases. Based on the observation, we choose Attention U-Net as the backbone of our LSRPR due to its distinct statistical advantage over all other compared backbones.

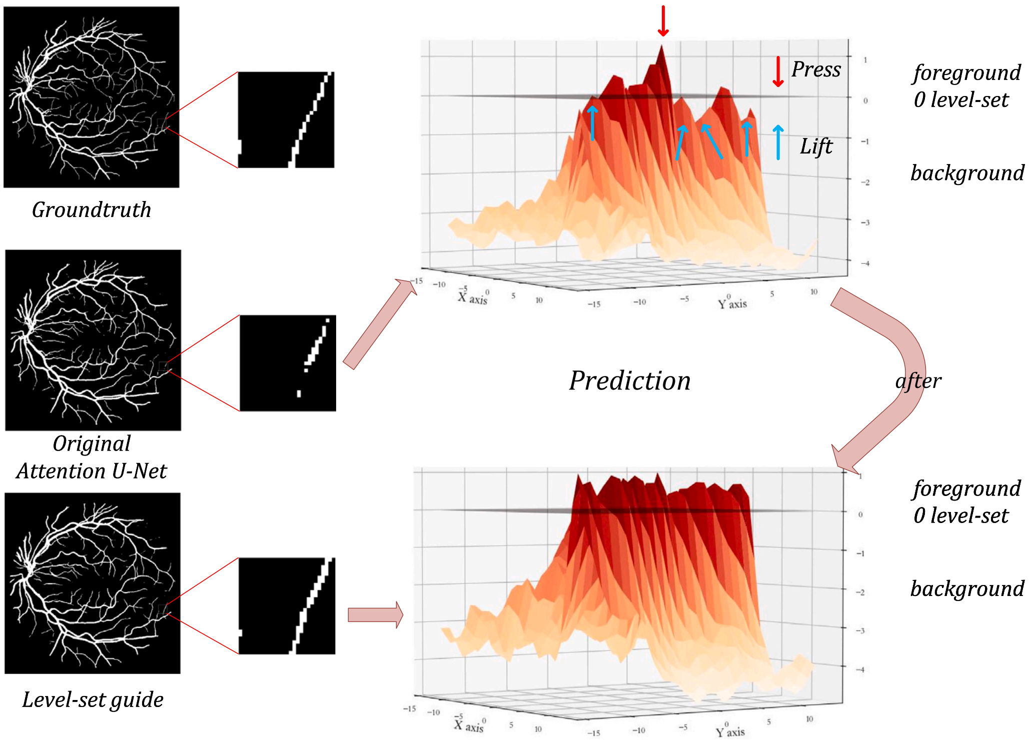

Fig. 5. Effect of level set loss. The level set loss provides the ability to smooth the shape of the segmentation, and it presses the highest hills and lifts the sunk valley.

4.3.2. *Effect of level set loss*

To show the effect of two-stage level set loss used in our proposed method, we provide experiment to observe the prototype rectification that level set brings to the network training. Fig. 5 illustrates the visualization result of the first image in the DRIVE dataset by using Attention U-Net with and without intervention of level set loss. It is worth noting that we take the feature mappings of CNN's output, specifically Z in Fig. 3, and reverse the value when the pixel belongs to the background to visualize. From a general view as shown in the left part of Fig. 5, the predicted results with level set loss guidance have less white-point shaped noise in comparison to the original Attention U-Net's result. In the local region as shown in the right part of Fig. 5, our method pulls down the pixels with over-predicted values and pulls up the pixels with incorrect predictions so that they are predicted correctly. From a theoretical perspective, level set method has the ability to force the foreground pixels or background pixels to be similar. In other words, it has the ability to cluster. Additionally, the neural network's adding the level set loss as a loss function is a plug-and-play design that has no impact on network optimization. This is the property of the level set function that can smooth the predicted values and this is also the meaning of Region Prototype Rectification.

The situation shown in the example above is also prevalent in the three datasets of our experiments, which is not an isolated case. We find some evidence in several images to support our theory that the proposed LS-loss maintains the vessels' connectivity. We trained the aforementioned methods on the DRIVE dataset, and Fig. 6 selects the 2nd image from the test set. From the figure, we can see that SA-UNet and Attention U-Net more effectively extract the blood vessel details than U-Net and R2U-Net do. In comparison, our LSRPR, based on Attention U-Net, smooths out the predictions and reduces the noise, resulting in more continuous vessels and closer to the ground truth. We choose the 3rd left-eye image in the CHASE_DB1 dataset as an example in Fig. 7 and we also choose the 8th image of RITE dataset as an example in Fig. 8. In CHASE_DB1 dataset, the situation is different. R2U-Net and SA-UNet are negatively impacted by the optic cup and optic disc, which are incorrectly extracted as features, whereas U-Net and Attention U-Net are unaffected. However, U-Net, SA-UNet, and Attention U-Net perform better in the RITE dataset than R2U-Net. In these two figures, we can clearly see that the regions are completed with prototype rectifications and the vessels are connected together by our LSRPR.

4.3.3. *Interpretability of designing two stages*

In this section, we interpret the significance of our two-stage design by organizing ablation experiments, and describe the purpose of each stage by using the training results of the experiments.

In the first stage, the self-supervised level set loss can help the network parameters iterate more quickly. In the second stage, the supervised level set loss can fine tune the network parameters for precisely predict the vessels. We subsequently add Stages 1 and 2 to the training process of networks to see how they affect the performance of networks. In this experiment, we use DRIVE as the dataset, and we only consider accuracy metric of the model to keep things simple. The training accuracy of the network by adding Stages 1 and 2 are listed in Table 3. We from Table 3 can see that Stage 1 makes the model less accurate than direct training of networks, while the Stage 2 improves the final outcome accuracy.

We show the convergence processes of networks training with or without self-supervised level set (Stage 1) in Fig. 9. The left portion of this figure shows how Dice loss varies with epoch increasing, and the right part shows the network's output image for given epoch. We see that in the first stage, our method converges faster than that without the level set function's guidance. One of the explanations is that, the unsupervised level set loss function quickly raises the values of the points around the pixel points with positive output values, which promotes faster convergence. According to the theoretical characteristics

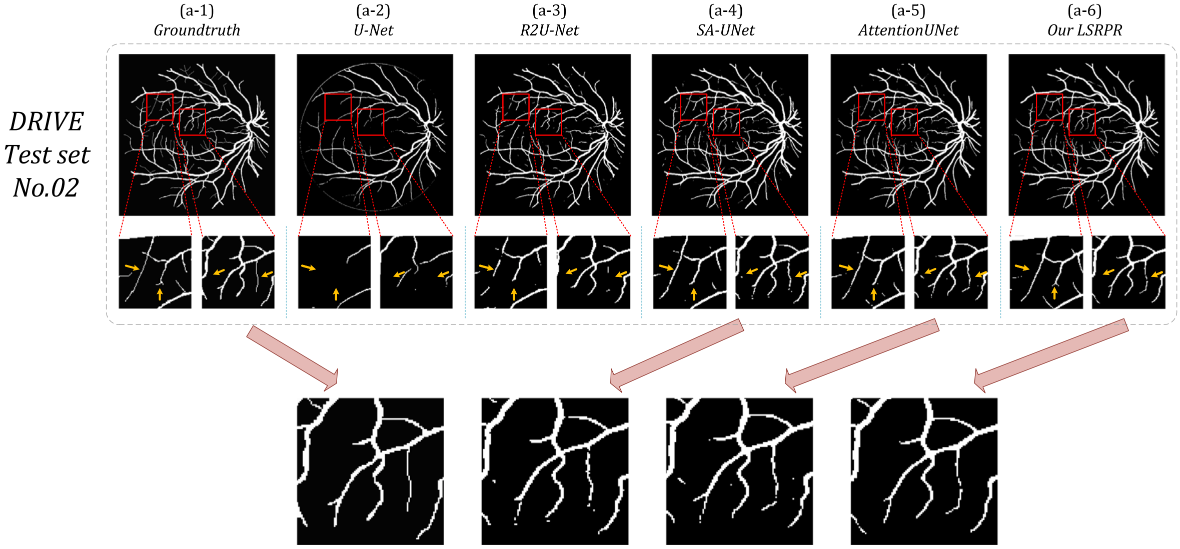

Fig. 6. Backbone selection: qualitative comparison with different models on the DRIVE dataset.

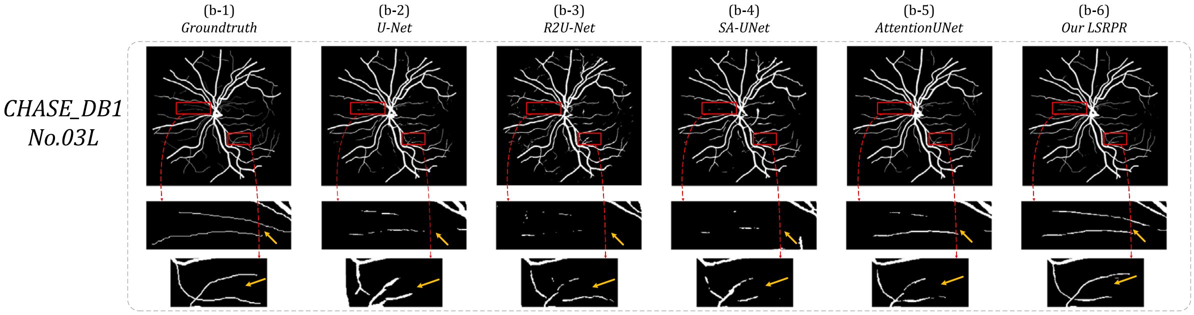

Fig. 7. Backbone selection: qualitative comparison with different models on the CHASE_DB1 dataset.

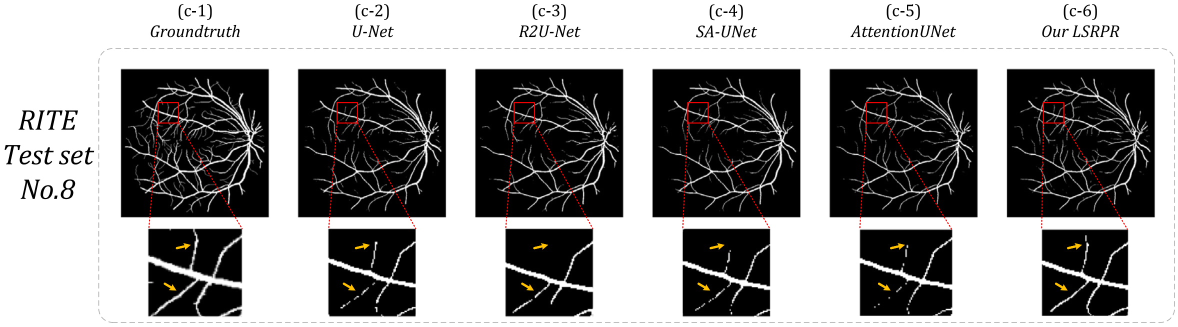

Fig. 8. Backbone selection: qualitative comparison with different models on the RITE dataset.

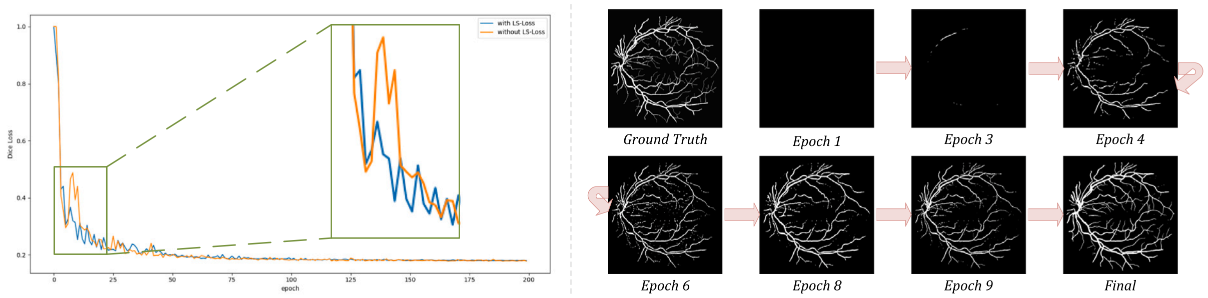

Fig. 9. Converges speed and object extraction process with or without the level set guidance.

Table 3
Accuracy comparison of Attention U-Net with or without level set loss.

Model	Acc
Attention U-Net	0.9687
Attention U-Net+Stage 1 (Self-supervised)	0.9622 (↓)
Attention U-Net+Stage 1+Stage 2 (Supervised)	**0.9735** (↑)

of the level set method, the level set loss has the prior knowledge to direct the CNN to iterate correctly. In the left portion of Fig. 9, the fast converge stage guided by the level set shows a smooth downward trend compared to no level set. However it is worth mentioning that Stage 1 does not make the results more accurate, because this is not the goal of Stage 1. On the contrary, the addition of level set function in Stage 1 may make the results less accurate, but it is in exchange for a significant increase in the speed of convergence. The right part of Fig. 9 shows that the model first predicts all pixels as background at initial epoch, and then some pixels are predicted as foreground through the training of the loss function to get the appropriate results gradually. Suppose that the supervised level set loss function of Stage 2 is directly added in the very beginning of training. Since the pixels are all predicted to be background, then the average of all pixel values that should be foreground will also be negative, at which point the supervised level set loss in Stage 2 prevents the pixel predictions from lifting, creating a confrontation with Dice and the cross-entropy loss function, resulting in an overall failure of networks. Due to the risk of the network failing to converge, we are unable to use the supervised level set loss for directly training the network. On the contrary, if the level set function of Stage 1 is used, this level set function will unconditionally obey the decision result trained by Dice with cross-entropy loss, which will uplift all positive values and suppress all negative values. That is also the fundamental reason why this level set function can accelerate convergence but not exact results. In this way, the two stages complement each other and work together to get a reasonable split result.

4.4. Performance comparisons with state-of-the-art methods

In the following, we compare our LSRPR with some of the most sophisticated models created specifically for segmenting retinal vessels. In other words, these models used the same dataset as ours in their respective papers. In order to respect the results of the original authors, the experimental results we selected for the comparison are taken from the respective papers, which means the criteria values are not calculated by us. As different papers select various metrics for evaluation, we use a horizontal line to indicate the metrics that are not mentioned in their respective papers. It is worth mentioning that few articles have used mIoU for evaluation in retinal vessel segmentation, while it is an important index for evaluating segmentation, so we show our mIoU here just for reference. We list the results of other methods on these three datasets, with the highest bolded.

Table 4
Quantitative comparison of LSRPR with other methods on DRIVE dataset.

Method	Year	Sen	F1	Acc	mIoU
U-Net [4]	(2015)	–	0.8142	–	–
R2U-Net [32]	(2018)	0.7792	0.8171	0.9556	–
LadderNet [43]	(2018)	0.7856	0.8202	0.9561	–
ET-Net [44]	(2019)	–	–	0.9560	0.7744
DUNet [6]	(2019)	–	0.8237	0.9697	–
IterNet [10]	(2020)	0.7791	0.8218	0.9574	–
NFN+ [45]	(2020)	0.7996	–	0.9582	–
IBA-U-Net [46]	(2021)	0.7858	–	0.9550	–
SA-UNet [36]	(2021)	0.8212	0.8263	0.9698	–
BSEResU-Net [47]	(2021)	0.8324	–	0.9574	–
MD-Net [48]	(2021)	0.8065	–	0.9676	–
SCS-Net [49]	(2021)	0.8289	–	0.9697	–
CSGNet [50]	(2022)	0.7984	0.8312	0.9709	–
FANet [51]	(2022)	–	0.8183	–	0.6927
CRAUNet [37]	(2022)	0.7954	0.8302	0.9586	–
LSRPR (Ours)		**0.8342**	**0.8430**	**0.9735**	**0.8502**

Table 5
Quantitative comparison of LSRPR with other methods on CHASE_DB1 dataset.

Method	Year	Sen	F1	Acc	mIoU
U-Net [4]	(2015)	–	0.7783	–	–
R2U-Net [32]	(2018)	0.7756	0.7928	0.9634	–
LadderNet [43]	(2018)	0.7978	0.8031	0.9656	–
DUNet [6]	(2019)	–	0.7883	0.9724	–
IterNet [10]	(2020)	0.7970	0.8072	0.9760	–
NFN+ [45]	(2020)	0.7963	–	0.9672	–
SA-UNet [36]	(2021)	**0.8573**	0.8153	0.9755	–
MD-Net [48]	(2021)	0.7504	–	0.9731	–
SCS-Net [49]	(2021)	0.8365	–	0.9744	–
CSGNet [50]	(2022)	0.7945	**0.8246**	**0.9779**	–
FANet [51]	(2022)	–	0.8108	–	0.6820
CRAUNet [37]	(2022)	0.8259	0.8156	0.9659	–
LSRPR (Ours)		0.8347	0.8102	0.9724	**0.8190**

Table 4 shows the quantitative comparison of LSRPR with other methods on DRIVE dataset. As can be obviously seen, our proposed method performed better than all others in the DRIVE dataset experiment for sensitivity, F1, accuracy, and mIoU. The sensitivity and accuracy represent the capability of the segmentation model, and we are slightly ahead of the best method in terms of sensitivity (by about 0.2 percentage points), F1 value (by about 1.3 percentage points), and accuracy (by about 0.3 percentage points), and the largest increase in F1 value represents that the prediction becomes accurate, which is the significance of the rectification in our proposed method.

Table 5 shows the quantitative comparison of LSRPR with other methods on CHASE_DB1 dataset. We observe that our proposed method ranks in the top for each of these values, despite the fact that it does not achieve the highest values of all models in terms of individual values. However, compared to other models, our method can significantly improve the F1 values without lagging too much in accuracy, and this is attributed to doing region prototype rectification to connect the vessels.

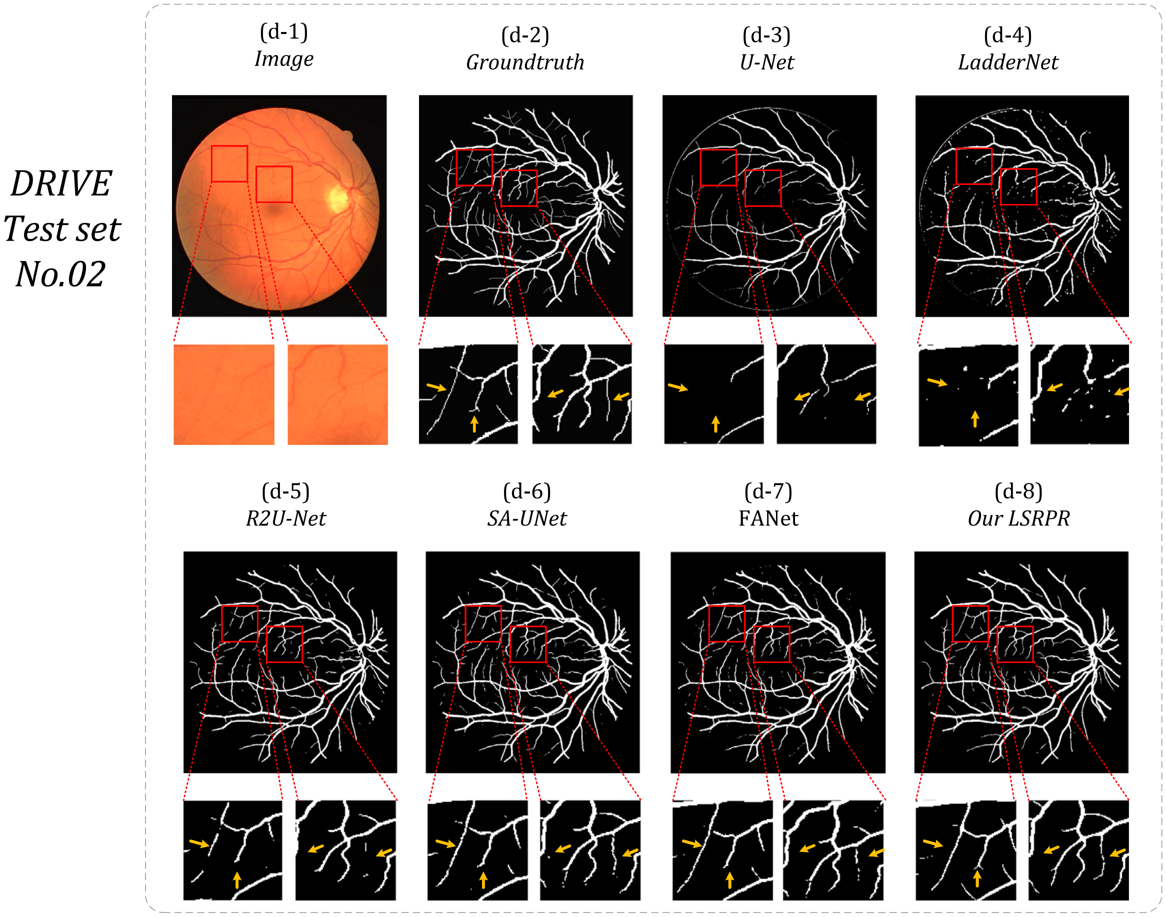

Fig. 10. Performance comparisons: visualization of some state-of-art models listed above on the DRIVE dataset.

Table 6
Quantitative comparison of LSRPR with other methods on RITE dataset.

Method	Year	F1
U-Net [4]	(2015)	0.5524
SegNet [52]	(2017)	0.5223
KiU-Net [11]	(2021)	0.7517
LSRPR (Ours)		**0.7566**

Table 6 illustrates the quantitative comparison of LSRPR with other methods on RITE dataset. Given a few number of papers using the RITE dataset, we only came up with recent model for comparison in terms of F1 values. We observe that our method outperforms the competing methods.

In particular, we have trained and visualized the output of some open-source models. Fig. 10 selects 5 open-source state-of-art methods listed in Table 4. As we can see, the original image of these datasets are already difficult to distinguish for the human eyes. Some models are unable to probe the details of blood vessels, such as LadderNet and R2UNet. Other models probe for vascular details but cannot guarantee vascular continuity, such as SA-UNet and FANet. The same situation occurs in CHASE_DB1 dataset in Fig. 11, with 5 open-source state-of-art methods listed in Table 5. Due to the high resolution of the images in this dataset and U-Net's ability to train on large-scale images with greater detail than other models, it performs better. Likewise, Fig. 12 selects 2 open-source state-of-art methods listed in Table 6. The effects of the selected methods do not differ significantly on this dataset, but our method clearly corrects the vessel disconnections. Due to the possibility of using different training methods, such as optimizers, learning rates, etc., our trained results may deviate a little from the results of their respective paper, which is within the error margin. From the visualization results, thanks to AttetionUNet's own ability to identify details and our LSRPR's rectification ability, our final model achieves excellent results in both probing vascular details and maintaining vascular continuity, which we believe plays a key role in improving criteria.

5. Conclusion

In this paper, we present a level set guided prototype rectification framework (LSRPR) for the retinal vessel segmentation task, which hires the level set loss as an auxiliary guidelines to modify the region prototype for the retinal vessel segmentation aspect and consists of two stages: the fast convergence stage and the accuracy enhancement stage. Each of these two stages is controlled by a corresponding level set loss, which has a variety of effects, including an unsupervised level set loss for faster convergence and a supervised level set loss for fine-tuning to get greater accuracy. We also conduct ablation experiments to select the best backbone and investigate how level set loss enhances the backbone to complete the segmentation task better. As the result shows, our proposed method can improve the segmentation accuracy and enhance the continuity of vessels compared to the recent segmentation methods. However, the two-stage structure in our LSRPR is not an end-to-end structure, which provides additional tedious operations for the training process. Moreover, level set method, as a mature image segmentation theory, is supposed to have unsupervised learning capability. In other

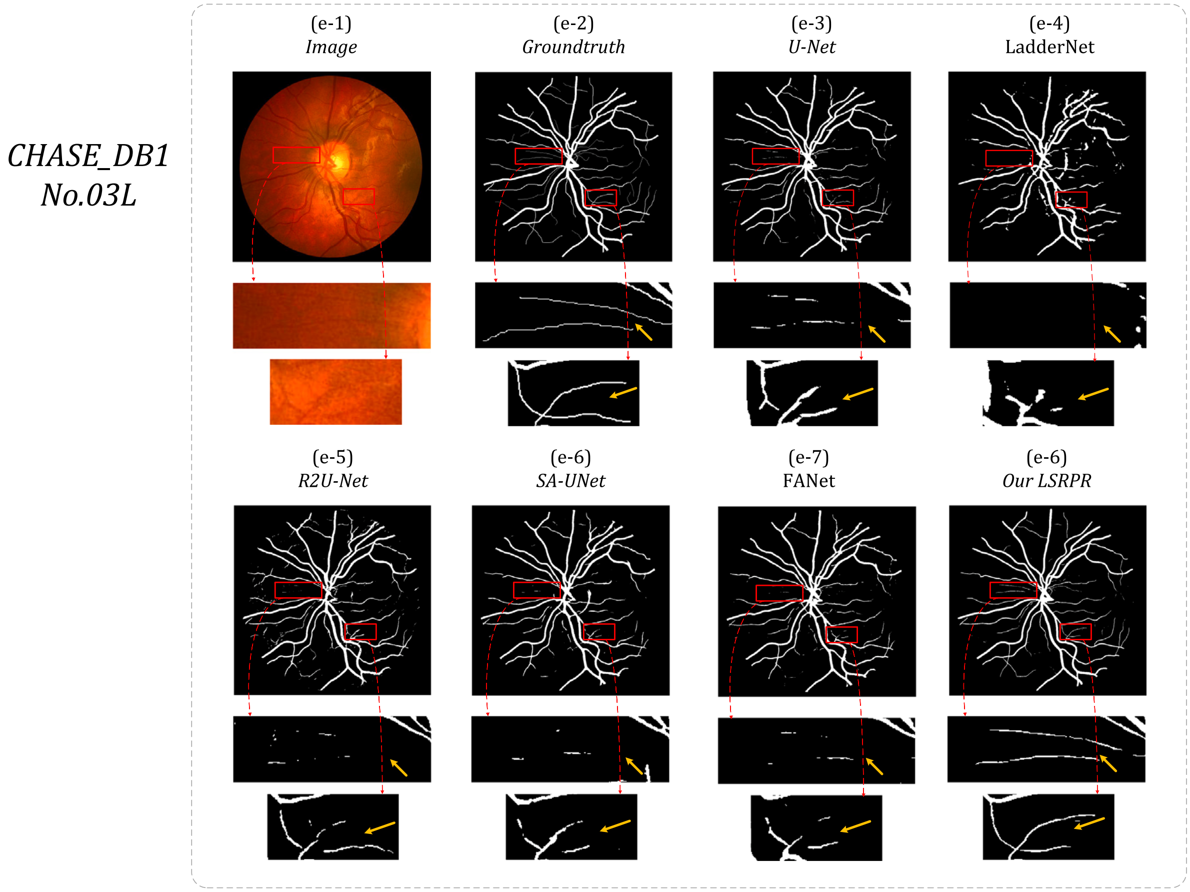

Fig. 11. Performance comparisons: visualization of some state-of-art models listed above on the CHASE_DB1 dataset.

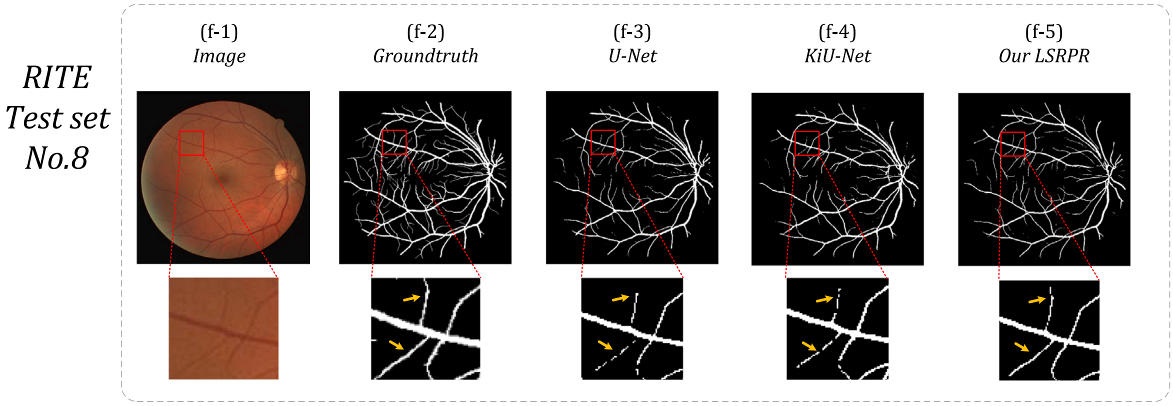

Fig. 12. Performance comparisons: visualization of some state-of-art models listed above on the RITE dataset.

words, we have not fully explored the potential of level set method. In addition, we only applied the level set method to the end of the network, but in reality, the level set can provide a priori information to each layer of the network, which is also the direction of our future research.

CRediT authorship contribution statement

Yifei Liu: Methodology, Software, Validation, Formal analysis, Writing – original draft, Visualization. **Qingtian Wu:** Formal analysis, Writing – review & editing, Supervision. **Xueyu Liu:** Investigation, Writing – review & editing. **Junyu Lu:** Investigation, Writing – review & editing. **Zhenhuan Xu:** Investigation, Writing – review & editing. **Yongfei Wu:** Conceptualization, Methodology, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Shu Feng:** Conceptualization, Writing – review & editing, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors sincerely thank the editor and reviewers for their valuable advice to improve our work. This work was supported by the National Natural Science Foundation of China (Grant Nos. 61901292, 62201331); the Natural Science Foundation of Shanxi Province of China (Grant Nos. 202103021223050, 20210302124543).

References

[1] C.G. Owen, A.R. Rudnicka, C.M. Nightingale, R. Mullen, S.A. Barman, N. Sattar, D.G. Cook, P.H. Whincup, Retinal arteriolar tortuosity and cardiovascular risk factors in a multi-ethnic population study of 10-year-old children; The child heart and health study in England (CHASE), Arterioscler. Thromb. Vasc. Biol. 31 (8) (2011) 1933–1938.

[2] N. Witt, T.Y. Wong, A.D. Hughes, N. Chaturvedi, B.E. Klein, R. Evans, M. McNamara, S.A.M. Thom, R. Klein, Abnormalities of retinal microvascular structure and risk of mortality from ischemic heart disease and stroke, Hypertension 47 (5) (2006) 975–981.

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

[4] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.

[5] O. Oktay, J. Schlemper, L.L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N.Y. Hammerla, B. Kainz, et al., Attention U-Net: Learning where to look for the pancreas, 2018, arXiv preprint arXiv:1804.03999.

[6] Q. Jin, Z. Meng, T.D. Pham, Q. Chen, L. Wei, R. Su, DUNet: A deformable network for retinal vessel segmentation, Knowl.-Based Syst. 178 (2019) 149–162.

[7] C. Guo, M. Szemenyei, Y. Yi, W. Wang, B. Chen, C. Fan, Sa-Unet: Spatial attention U-Net for retinal vessel segmentation, in: 2020 25th International Conference on Pattern Recognition, ICPR, IEEE, 2021, pp. 1236–1242.

[8] G. Ghiasi, T.-Y. Lin, Q.V. Le, Dropblock: A regularization method for convolutional networks, Adv. Neural Inf. Process. Syst. 31 (2018).

[9] J. Hu, H. Wang, S. Gao, M. Bao, T. Liu, Y. Wang, J. Zhang, S-unet: A bridge-style U-Net framework with a saliency mechanism for retinal vessel segmentation, IEEE Access 7 (2019) 174167–174177.

[10] L. Li, M. Verma, Y. Nakashima, H. Nagahara, R. Kawasaki, Iternet: Retinal image segmentation utilizing structural redundancy in vessel networks, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2020, pp. 3656–3665.

[11] J.M.J. Valanarasu, V.A. Sindagi, I. Hacihaliloglu, V.M. Patel, Kiu-Net: Overcomplete convolutional architectures for biomedical image and volumetric segmentation, IEEE Trans. Med. Imaging 41 (4) (2021) 965–976.

[12] B. Al-Diri, A. Hunter, D. Steel, An active contour model for segmenting and measuring retinal vessels, IEEE Trans. Med. Imaging 28 (9) (2009) 1488–1497.

[13] B. Kim, J.C. Ye, Mumford–Shah loss functional for image segmentation with deep learning, IEEE Trans. Image Process. 29 (2019) 1856–1866.

[14] X. Chen, B.M. Williams, S.R. Vallabhaneni, G. Czanner, R. Williams, Y. Zheng, Learning active contour models for medical image segmentation, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2019, pp. 11624–11632, http://dx.doi.org/10.1109/CVPR.2019.01190.

[15] S. Gur, L. Wolf, L. Golgher, P. Blinder, Unsupervised microvascular image segmentation using an active contours mimicking neural network, in: 2019 IEEE/CVF International Conference on Computer Vision, ICCV, 2019, pp. 10721–10730, http://dx.doi.org/10.1109/ICCV.2019.01082.

[16] Y. Kim, S. Kim, T. Kim, C. Kim, CNN-based semantic segmentation using level set loss, in: 2019 IEEE Winter Conference on Applications of Computer Vision, WACV, 2019, pp. 1752–1760, http://dx.doi.org/10.1109/WACV.2019.00191.

[17] P. Hu, B. Shuai, J. Liu, G. Wang, Deep level sets for salient object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2300–2309.

[18] D. Marcos, D. Tuia, B. Kellenberger, L. Zhang, M. Bai, R. Liao, R. Urtasun, Learning deep structured active contours end-to-end, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 8877–8885.

[19] A. Hatamizadeh, D. Sengupta, D. Terzopoulos, End-to-end deep convolutional active contours for image segmentation, 2019, arXiv preprint arXiv:1909.13359.

[20] D. Cheng, R. Liao, S. Fidler, R. Urtasun, Darnet: Deep active ray network for building segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 7431–7439.

[21] T. Le, R. Gummadi, M. Savvides, Deep recurrent level set for segmenting brain tumors, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2018, pp. 646–653.

[22] X. Chen, B.M. Williams, S.R. Vallabhaneni, G. Czanner, R. Williams, Y. Zheng, Learning active contour models for medical image segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 11632–11640.

[23] J. Ma, J. He, X. Yang, Learning geodesic active contours for embedding object global information in segmentation CNNs, IEEE Trans. Med. Imaging 40 (1) (2020) 93–104.

[24] T.H.N. Le, K.G. Quach, K. Luu, C.N. Duong, M. Savvides, Reformulating level sets as deep recurrent neural network approach to semantic segmentation, IEEE Trans. Image Process. 27 (5) (2018) 2393–2407.

[25] A. Hatamizadeh, A. Hoogi, D. Sengupta, W. Lu, B. Wilcox, D. Rubin, D. Terzopoulos, Deep active lesion segmentation, in: Machine Learning in Medical Imaging: 10th International Workshop, MLMI 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Proceedings, Vol. 10, Springer, 2019, pp. 98–105.

[26] X. Wang, X. Jiang, Retinal vessel segmentation by a divide-and-conquer funnel-structured classification framework, Signal Process. 165 (2019) 104–114.

[27] S. Osher, J.A. Sethian, Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations, J. Comput. Phys. 79 (1) (1988) 12–49.

[28] T.F. Chan, L.A. Vese, Active contours without edges, IEEE Trans. Image Process. 10 (2) (2001) 266–277.

[29] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Commun. ACM 60 (6) (2017) 84–90.

[30] Z. Jiang, H. Zhang, Y. Wang, S.-B. Ko, Retinal blood vessel segmentation using fully convolutional network with transfer learning, Comput. Med. Imaging Graph. 68 (2018) 1–15.

[31] 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. Inf. 126 (2019) 105–113.

[32] M.Z. Alom, M. Hasan, C. Yakopcic, T.M. Taha, V.K. Asari, Recurrent residual convolutional neural network based on U-Net (r2u-net) for medical image segmentation, 2018, arXiv preprint arXiv:1802.06955.

[33] C. Guo, M. Szemenyei, Y. Pei, Y. Yi, W. Zhou, SD-UNet: A structured dropout U-Net for retinal vessel segmentation, in: 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering, BIBE, IEEE, 2019, pp. 439–444.

[34] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, Y. Wei, Deformable convolutional networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 764–773.

[35] G.A. Francia, C. Pedraza, M. Aceves, S. Tovar-Arriaga, Chaining a U-Net with a residual U-net for retinal blood vessels segmentation, IEEE Access 8 (2020) 38493–38500.

[36] C. Guo, M. Szemenyei, Y. Yi, W. Wang, B. Chen, C. Fan, Sa-Unet: Spatial attention U-Net for retinal vessel segmentation, in: 2020 25th International Conference on Pattern Recognition, ICPR, IEEE, 2021, pp. 1236–1242.

[37] F. Dong, D. Wu, C. Guo, S. Zhang, B. Yang, X. Gong, CRAUNet: A cascaded residual attention U-Net for retinal vessel segmentation, Comput. Biol. Med. (2022) 105651.

[38] T. Zhao, X. Wu, Pyramid feature attention network for saliency detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 3085–3094.

[39] J. Wang, X. Li, Y. Cheng, Towards an extended EfficientNet-based U-Net framework for joint optic disc and cup segmentation in the fundus image, Biomed. Signal Process. Control 85 (2023) 104906.

[40] J. Staal, M. Abramoff, M. Niemeijer, M. Viergever, B. van Ginneken, Ridge-based vessel segmentation in color images of the retina, IEEE Trans. Med. Imaging 23 (4) (2004) 501–509, http://dx.doi.org/10.1109/TMI.2004.825627.

[41] M.M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A.R. Rudnicka, C.G. Owen, S.A. Barman, An ensemble classification-based approach applied to retinal blood vessel segmentation, IEEE Trans. Biomed. Eng. 59 (9) (2012) 2538–2548, http://dx.doi.org/10.1109/TBME.2012.2205687.

[42] Q. Hu, M.D. Abràmoff, M.K. Garvin, Automated separation of binary overlapping trees in low-contrast color retinal images, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013: 16th International Conference, Nagoya, Japan, September 22-26, 2013, Proceedings, Part II 16, Springer, 2013, pp. 436–443.

[43] J. Zhuang, LadderNet: Multi-path networks based on U-net for medical image segmentation, 2018, arXiv preprint arXiv:1810.07810.

[44] Z. Zhang, H. Fu, H. Dai, J. Shen, Y. Pang, L. Shao, Et-Net: A generic edge-attention guidance network for medical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2019, pp. 442–450.

[45] Y. Wu, Y. Xia, Y. Song, Y. Zhang, W. Cai, NFN+: A novel network followed network for retinal vessel segmentation, Neural Netw. 126 (2020) 153–162.

[46] S. Chen, Y. Zou, P.X. Liu, Iba-U-Net: attentive bconvlstm U-Net with redesigned inception for medical image segmentation, Comput. Biol. Med. 135 (2021) 104551.

[47] D. Li, S. Rahardja, BSEResU-Net: An attention-based before-activation residual U-Net for retinal vessel segmentation, Comput. Methods Programs Biomed. 205 (2021) 106070.

[48] Z. Shi, T. Wang, Z. Huang, F. Xie, Z. Liu, B. Wang, J. Xu, MD-Net: A multi-scale dense network for retinal vessel segmentation, Biomed. Signal Process. Control 70 (2021) 102977.

[49] H. Wu, W. Wang, J. Zhong, B. Lei, Z. Wen, J. Qin, Scs-Net: A scale and context sensitive network for retinal vessel segmentation, Med. Image Anal. 70 (2021) 102025.

[50] S. Guo, CSGNet: Cascade semantic guided net for retinal vessel segmentation, Biomed. Signal Process. Control 78 (2022) 103930.

[51] N.K. Tomar, D. Jha, M.A. Riegler, H.D. Johansen, D. Johansen, J. Rittscher, P. Halvorsen, S. Ali, Fanet: A feedback attention network for improved biomedical image segmentation, IEEE Trans. Neural Netw. Learn. Syst. (2022).

[52] V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: A deep convolutional encoder-decoder architecture for image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (12) (2017) 2481–2495.