Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

RECEIVED
12 December 2022

REVISED
3 April 2023

ACCEPTED FOR PUBLICATION
13 April 2023

PUBLISHED
3 May 2023

PAPER

Multi-scale and local feature guidance network for corneal nerve fiber segmentation

Wei Tang[1], Xinjian Chen[1,2], Jin Yuan[3], Qingquan Meng[1], Fei Shi[1], Dehui Xiang[1], Zhongyue Chen[1] and Weifang Zhu[1,*]

[1] MIPAV Lab, School of Electronic and Information Engineering, Soochow University, People's Republic of China
[2] MIPAV Lab, School of Electronic and Information Engineering and State Key Laboratory of Radiation Medicine and Protection, Soochow University, People's Republic of China
[3] Zhongshan Ophthalmic Center, Sun Yat-Sen University, People's Republic of China
[*] Author to whom any correspondence should be addressed.

E-mail: wfzhu@suda.edu.cn

Keywords: corneal confocal microscopy, convolutional neural network, nerve fiber segmentation

Abstract

Objective. Corneal confocal microscopy (CCM) is a rapid and non-invasive ophthalmic imaging technique that can reveal corneal nerve fiber. The automatic segmentation of corneal nerve fiber in CCM images is vital for the subsequent abnormality analysis, which is the main basis for the early diagnosis of degenerative neurological systemic diseases such as diabetic peripheral neuropathy. *Approach*. In this paper, a U-shape encoder–decoder structure based multi-scale and local feature guidance neural network (MLFGNet) is proposed for the automatic corneal nerve fiber segmentation in CCM images. Three novel modules including multi-scale progressive guidance (MFPG) module, local feature guided attention (LFGA) module, and multi-scale deep supervision (MDS) module are proposed and applied in skip connection, bottom of the encoder and decoder path respectively, which are designed from both multi-scale information fusion and local information extraction perspectives to enhance the network's ability to discriminate the global and local structure of nerve fibers. The proposed MFPG module solves the imbalance between semantic information and spatial information, the LFGA module enables the network to capture attention relationships on local feature maps and the MDS module fully utilizes the relationship between high-level and low-level features for feature reconstruction in the decoder path. *Main results*. The proposed MLFGNet is evaluated on three CCM image Datasets, the Dice coefficients reach 89.33%, 89.41%, and 88.29% respectively. *Significance*. The proposed method has excellent segmentation performance for corneal nerve fibers and outperforms other state-of-the-art methods.

1. Introduction

The geometric and topological features such as length, density, and tortuosity of corneal nerve fibers are important indicators for the early diagnosis of degenerative neurological systemic diseases such as diabetic peripheral neuropathy (Daousi *et al* 2004, Mehra *et al* 2007, Tavakoli *et al* 2010a, Kang and Kim 2015, Li *et al* 2019), human immunodeficiency virus (Kemp *et al* 2017), Parkinson's disease (Misra *et al* 2017), multiple sclerosis (Petropoulos *et al* 2017) and various dementias (Ponirakis *et al* 2019, Testa *et al* 2020), in which the thin nerve fibers are affected first predominantly (Petropoulos *et al* 2020). Corneal confocal microscopy (CCM) (Tavakoli *et al* 2010b) is a rapid and non-invasive ophthalmic imaging technique that can reveal corneal nerve fiber well. As CCM imaging is widely and fRequently used in disease screening and clinical trials, automatic and accurate methods for corneal nerve fiber segmentation in CCM images are urgently needed, which is the basis for the quantitative analysis of geometric and topological features. Figure 1 shows some CCM images with different intensity distributions, contrasts, and tortuosity levels in which corneal nerve fibers are curvilinear structures with various orientations, lengths, and thicknesses. As can be seen from figure 1, corneal nerve fibers

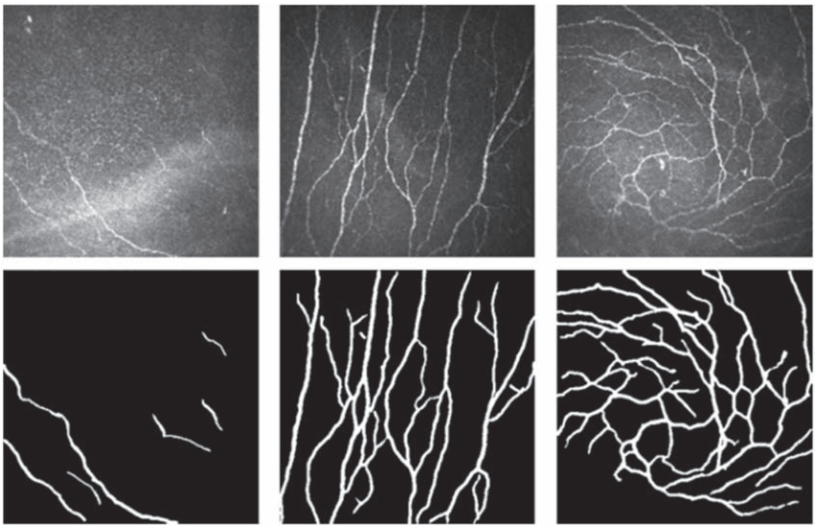

Figure 1. Corneal confocal microscopy(CCM) images with different intensity distributions, contrasts and tortuosity levels.

may appear very faint due to differences of imaging depth, and CCM images also contain small bright and non-nerve fiber structures (usually cells), which increases the challenge of identifying nerve fibers.

1.1. Traditional machine learning based methods

There are many traditional machine learning based methods for corneal nerve fiber segmentation including filtering based methods (Dabbah *et al* 2010, 2011, Ferreira *et al* 2012, Poletti and Ruggeri 2013, Annunziata *et al* 2016), clustering based methods (Scarpa *et al* 2008, Ronneberger *et al* 2015, Chen *et al* 2016, Lagali *et al* 2018) and classification based methods (Wang *et al* 2020, Zhong *et al* 2022). (1) Filtering based methods: Dabbah *et al* proposed a dual-model corneal nerve fiber detection algorithm based on Gabor and Gaussian filters (Dabbah *et al* 2010). They proposed a multi-scale adaptive dual-model based detection algorithm later, which utilized the curvilinear structure property of the nerve fibers (Dabbah *et al* 2011). Ferreira *et al* proposed a wavelet transform filtering based phase symmetry analysis to identify the nerve structures and used morphological operations for nerve reconstruction, in which the reconstruction performance dependents on the selection of correct seed points and needs sophisticated post-processing (Ferreira *et al* 2012). (2) Clustering based methods: Poletti *et al* identified a set of seed points and connecting seeds by means of minimum cost paths to trace nerve fibers (Poletti and Ruggeri 2013). Aunnunziata *et al* proposed a hybrid nerve fiber segmentation method, which consisted of a scale and curvature-invariant ridge detector based appearance model and a K-means clustering based context filters. It was specifically designed for tortuous and fragmented structures and the segmentation results were used for the further tortuosity estimation (Annunziata *et al* 2016). Scarpa *et al* proposed a nerve tracing method based on Gabor filter and fuzzy C-means clustering, in which several post-processing procedures were adopted to remove false recognitions and to link sparse segments into continuous structures (Scarpa *et al* 2008). (3) Classification based methods: Chen *et al* used dual-model filter and dual-tree complex wavelet transform based feature descriptors to training the neural network/random forest for nerve fiber detection. The detection results were used for the evaluation of nerve fiber quantification (Chen *et al* 2016). Lagali *et al* proposed a support vector machine based method for nerve fiber recognition (Lagali *et al* 2018). As traditional machine learning based methods are not end-to-end, they are generally less efficient and accurate.

1.2. Deep learning based methods

With deep learning showing its strong superiority, more and more studies have focused on convolutional neural network based networks to segment structures with curvilinear characteristics, such as nerve fiber and blood vessel. There are few deep learning based methods focusing on corneal nerve fiber segmentation. Considering blood vessels and nerve fibers are both curvilinear structures with similar characteristics, it is worth learning from the methods for blood vessel segmentation task. There are many methods for retinal vessel segmentation based on U-shape encoder–decoder structure (Ronneberger *et al* 2015). Zhong *et al* proposed MMDC-Net to extract multi-layer and multi-scale information to sharpen the vessel details for retinal vessel segmentation (Zhong *et al* 2022). Wang *et al* proposed RVSeg-Net with dilated convolution for retinal vessel segmentation (Chen *et al* 2017, Wang *et al* 2020). However, the use of dilated convolution may lead to the loss of detailed vessel information while increasing the receptive field. Sun *et al* proposed UCR-Net to fuse context attention information with context attention exploration module and global and spatial attention module to capture context features for vessel segmentation (Sun *et al* 2022). Gu *et al* proposed CE-Net which used the pretrained ResNet as backbone and was embedded with context extractor module for vessel segmentation (He *et al* 2016, Gu *et al* 2019). But five downsampling operations in ResNet may destroy the structure of thin vessels. Feng *et al* proposed CPFNet which combined two pyramidal modules to fuse global and multi-scale context information and shew good performance on four challenging medical image segmentation tasks including retinal linear lesion segmentation in indocyanine green angiography (ICGA) images (Feng *et al* 2020). There are also some networks based on the U-shape encoder–decoder structure for the corneal nerve fiber segmentation. U-Net was directly used for corneal nerve fiber segmentation in (Colonna *et al* 2018, Williams *et al* 2020) for the following analysis. Zhang *et al* incorporated attention gatemodules to improve the network's ability to distinguish the nerve fiber from background. However, this network did not take advantage of multi-scale information, which is important for the simultaneous segmentation of thick and thin nerve fibers (Zhang *et al* 2020). Mou *et al* proposed the CS-Net (Mou *et al* 2019) embedded with channel and spatial attention module to capture the attention relationship in channel direction and spatial direction for nerve fiber centerline tracing, which was later extended to CS^2 -Net (Mou *et al* 2021) for dealing with 3D curvilinear structure segmentation. Chen *et al* developed a modified UNet++ (Zhou *et al* 2018) model (by adding skip connections in the upsampling path) for nerve fiber segmentation and developed a centerline extraction algorithm based on neighborhood statistics (Chen *et al* 2021). However, the added dense skip connections in U-Net++ did not result in a significant improvement in segmentation performance. Yang *et al* proposed a multi-discriminator adversarial convolutional network (MDACN), where both the generator and the two discriminators emphasize multi-scale feature representations, combined with an improved loss function which enables the network to pay more attention to thin fibers (Yang *et al* 2021).

1.3. Overview and contributions

The difficulty of nerve fiber segmentation in CCM images mainly lies in the thin and faint nerve fibers and the low contrast of the images. To address these difficulties, we propose a novel multi-scale and local feature guidance neural network (MLFGNet) embedded with three novel modules, including multi-scale progressive guidance (MFPG) module, local feature guided attention (LFGA) module, and multi-scale deep supervision (MDS) module, which are designed from both multi-scale information fusion and local information extraction perspectives to enhance the network's ability to discriminate the global and local structure of nerve fibers. In order to recover more thin and faint fiber related information in the decoding stage and suppress background noise, the MFPG module is designed, which uses high-level features to guide low-level features and aggregates information level by level to shrink the information gap between different levels progressively. In order to improve the network's ability to discriminate nerve fibers with low contrast, the LFGA module is proposed. LFGA module first splits the feature map into m patches. Then pixel-wise correlation and linear dependency are captured in parallel on each patch, which enables the network to pay more attention to local features. To achieve the efficient optimization of the network during training, the MDS module is designed, which allows the gradient information to flow between different stages in the decoder sufficiently in the backpropagation. The main contributions of this paper are as follows:

(1) We propose a novel MLFGNet equipped with MFPG, LFGA, and MDS modules for corneal nerve fiber segmentation.

(2) From the perspective of taking full use of features from different layers and scales, the proposed MFPG module can solve the imbalance between the semantic information and spatial information which enhances the network's ability to recover the structure of nerve fibers, and the MDS module can fully utilize the relationship between high-level and low-level features for feature reconstruction in the decoder path which further optimizes the performance of the network and improves its optimization efficiency.

(3) From the perspective of highlighting local features, the proposed LFGA module enables the network to capture pixel-wise correlation and linear dependency on local feature maps, which can improve the network's ability to discriminate faint and thin nerve fibers.

(4) The proposed MLFGNet is evaluated on two public corneal nerve fiber datasets and one in-house dataset and achieves state-of-the-art segmentation performance.

The rest of this paper is organized as follows: in section 2, we introduce the detailed structure of the proposed MLFGNet. Section 3 presents experimental settings, and the results, including ablation experiments and comparison experiments with state-of-the-art methods. Conclusion is presented in section 4.

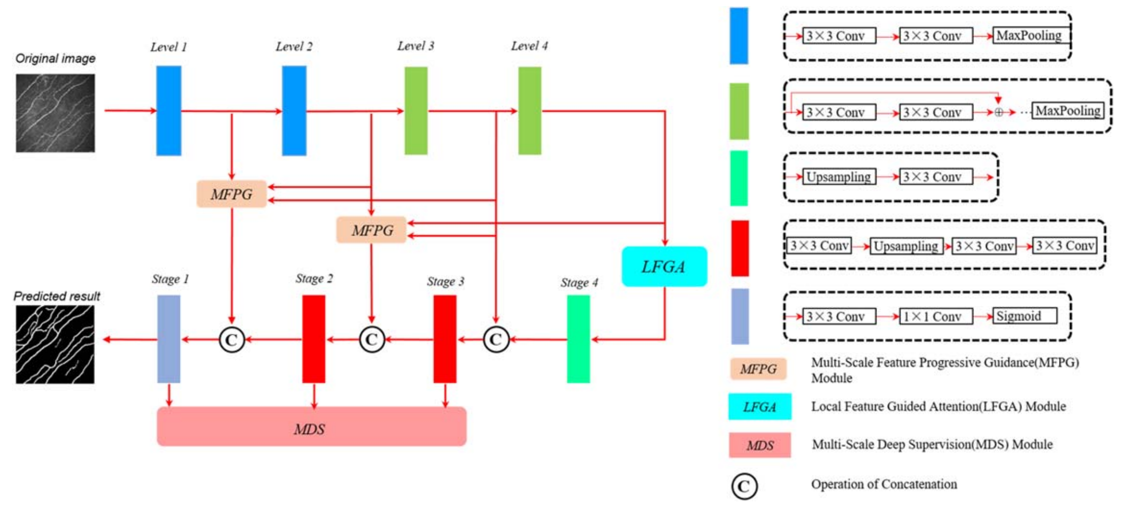

Figure 2. The overall architecture of the proposed MLFGNet. The original image is fed into the encoder to obtain the high-level features, and the local attention information is captured the by the proposed local feature guided attention (LFGA) module. Next, the features are recovered by the decoder, the multi-scale information flows are introduced by the proposed multi-scale progressive guidance (MFPG) module. Meanwhile, the multi-scale deep supervision (MDS) module is introduced as deep supervision mechanism. The final predicted result is obtained via a threshold segmentation operation with threshold of 0.5.

2. Methods and materials

2.1. Overview of the architecture of MLFGNet

Figure 2 shows the whole framework of the proposed multi-scale feature guidance neural network (MLFGNet), which is based on a U-shape encoder–decoder structure and consists of five parts: the encoder path, the multi-scale progressive guidance (MFPG) module, the LFGA module, the MDS module, and the decoder path.

2.2. Feature encoder and decoder

The encoder path includes four levels. As can be seen from figure 1 that nerve fiber is a kind of thin and curvilinear structure, keeping its spatial information as much as possible while extracting its deep semantic information is essentially necessary. For this purpose, in the first two levels (Level 1 and 2) of the encoder, we use conv-block as U-Net (Ronneberger *et al* 2015) to extract spatial features, which consists of two consecutive 3 × 3 convolutions, batch normalization, and ReLu activation. In the last two levels (Level 3 and 4) of the encoder, pre-trained ResNet layers are employed to extract deep semantic features. In the decoder stage, the decoder fuses the feature maps from the corresponding encoder module through skip connection and then upsamples the fused feature map via bilinear upsampling.

2.3. Multi-scale feature progressive guidance module

In the decoder path, the semantic and spatial information will gradually lose in the process of feature upsampling. To address this problem, U-Net (Ronneberger *et al* 2015) introduces skip connection to recover this information. However, some previous researchers have found that the simple fusion of low-level and high-level features could be less effective due to the information gap between different levels (Zhang *et al* 2018, Guo *et al* 2019). These studies tried to deal with this problem (Feng *et al* 2020, Xu *et al* 2020), but they simply upsampled the deep features and concatenated with the shallow features, in which the information gap between different levels was not fully considered. In this paper, we propose a novel multi-scale feature progressive guidance (MFPG) module, in which information is progressively aggregated from high-level features to low-level ones to shrink the information gap between different levels. MFPG module can effectively suppress background noise and retain detailed spatial and semantic information.

Figure 3 shows the MFPG module between Level 1 of the encoder path and Stage 1 of the decoder path. Given a feature map $X_3 \in \mathbb{R}^{C_3 \times H_3 \times W_3}$ (C_3 represents the channel number of X_3, and H_3 and W_3 represent the height and width of X_3) from Level 3 of the encoder and an upsampling ratio k (k is set to 2 here). First, a dynamic upsampling was implemented instead of interpolation upsampling. To be specific, let X_3 through a content encoder layer (composed of consecutive 3×3 convolution, batch normalization, and Relu activation) and generate a new feature map $X_3^{\mathrm{En}} \in \mathbb{R}^{C_2 * k * k \times H_3 \times W_3}$. Then, sub-pixel convolution (Shi *et al* 2016) with stride $\frac{1}{k}$ $(k = 2)$ is used to transform X_3^{En} to $X_3^{\mathrm{Up}} \in \mathbb{R}^{C_2 \times H_3 * k \times W_3 * k}$, which is learnable and more flexible than bilinear upsampling. Concatenate X_3^{Up} with Level 2 feature map X_2 to generate $X_2^{\mathrm{Guide}} \in \mathbb{R}^{2 * C_2 \times H_2 \times W_2}$, which integrates

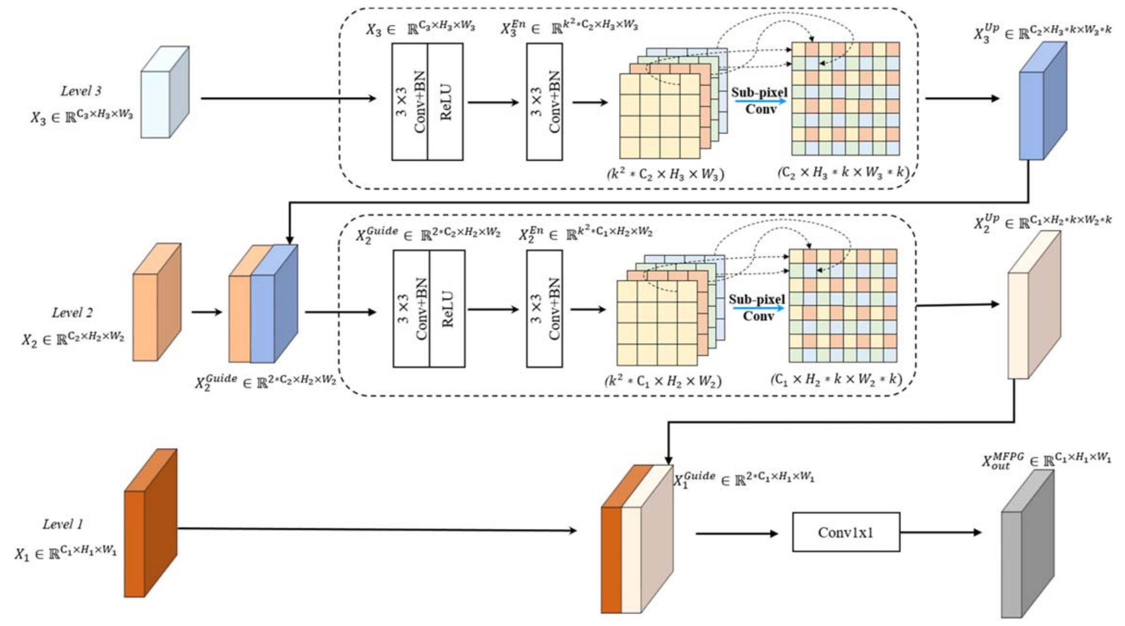
Level 3
$X_3 \in \mathbb{R}^{C_3\times H_3\times W_3}$
$X_3 \in \mathbb{R}^{C_3\times H_3\times W_3}$
$X_3^{En} \in \mathbb{R}^{k^2*C_2\times H_3\times W_3}$
3×3 Conv+BN
ReLU
3×3 Conv+BN
Sub-pixel Conv
$(k^2 * C_2 \times H_3 \times W_3)$
$(C_2 \times H_3 * k \times W_3 * k)$
$X_3^{Up} \in \mathbb{R}^{C_2\times H_3*k\times W_3*k}$
Level 2
$X_2 \in \mathbb{R}^{C_2\times H_2\times W_2}$
$X_2^{Guide} \in \mathbb{R}^{2*C_2\times H_2\times W_2}$
$X_2^{Guide} \in \mathbb{R}^{2*C_2\times H_2\times W_2}$
$X_2^{En} \in \mathbb{R}^{k^2*C_1\times H_2\times W_2}$
3×3 Conv+BN
ReLU
3×3 Conv+BN
Sub-pixel Conv
$(k^2 * C_1 \times H_2 \times W_2)$
$(C_1 \times H_2 * k \times W_2 * k)$
$X_2^{Up} \in \mathbb{R}^{C_1\times H_2*k\times W_2*k}$
Level 1
$X_1 \in \mathbb{R}^{C_1\times H_1\times W_1}$
$X_1^{Guide} \in \mathbb{R}^{2*C_1\times H_1\times W_1}$
Conv1x1
$X_{out}^{MFPG} \in \mathbb{R}^{C_1\times H_1\times W_1}$

Figure 3. The multi-scale feature progressive guidance (MFPG) module. Taking Stage 1 as an example, MFPG uses high-level features to guide low-level features and aggregates information level by level to shrink the information gap between different levels progressively.

the feature information from the local level (Level 2) and the higher level (Level 3). We also let $X_2^{\rm Guide}$ through a content encoder layer to generate a new feature map $X_2^{\rm En} \in \mathbb{R}^{C_1*k*k\times H_2\times W_2}$ and use sub-pixel convolution to transform it to $X_2^{\rm Up} \in \mathbb{R}^{C_1\times H_2*k\times W_2*k}$. Then concatenate $X_2^{\rm Up}$ to Level 1 feature map X_1 to generate $X_1^{\rm Guide} \in \mathbb{R}^{2*C_1\times H_1\times W_1}$, which fuses the features from X_1, X_2, and X_3 progressively and are both rich in spatial details and semantic information, solving the nuisance of feature mismatch caused by information gap between different levels. Finally, we use a 1×1 convolution to adjust the channel of $X_1^{\rm Guide}$ to C_1 and get the output feature $X_{\rm out} \in \mathbb{R}^{C_1\times H_1\times W_1}$. The process can be formulated as follows,

$$X_2^{\rm Guide} = {\rm concat}\left[\!\!\left[X_2\ ,\ {\rm Sub}_{\rm conv}\left[\oint(x_3)\right]\right]\!\!\right] \in \mathbb{R}^{2*C_2\times H_2\times W_2} \tag{1}$$

$$X_1^{\rm Guide} = {\rm concat}\left[\!\!\left[X_1,\ {\rm Sub}_{\rm conv}\left[\oint(X_2^{\rm Guide})\right]\right]\!\!\right] \in \mathbb{R}^{2*C_1\times H_1\times W_1} \tag{2}$$

$$X_{\rm out} = f\ [\![X_1^{\rm Guide}]\!] \in \mathbb{R}^{C_1\times H_1\times W_1} \tag{3}$$

Where Sub_conv[·] means sub-pixel convolution, $\oint(\cdot)$ means content encoder operation. concat $[\![\ \cdot\ ]\!]$ means concatenation operation. $f\ [\![\ \cdot\ ]\!]$ means 1×1 convolution.

2.4. Local feature guided attention module

Due to the difference in imaging depth, the intensity of corneal nerve fibers in CCM images often varies greatly, which means that the contrast between nerve fibers and the background varies greatly. In order to enable the network to pay attention to local feature information and improve the network's ability to discriminate nerve fibers with low contrast, inspired by (Shi *et al* 2016, Wang *et al* 2018, Yuan *et al* 2021, Hou *et al* 2020), a novel LFGA module is proposed, which splits the feature map into m patches and pixel-wise correlation and linear dependency are captured in parallel on each patch. Figure 4. shows the architecture of the LFGA module, which mainly consists of linear dependency capture path and pixel-wise correlation capture path.

Let $X_{\rm in} \in \mathbb{R}^{C\times H_0\times W_0}$ represent the original input feature and $P^{i_{\rm th}} \in \mathbb{R}^{C\times H\times W}$, $i_{\rm th} \in [1,\ m]$ represent the $i_{\rm th}$ patched feature map , in which C denotes the number of channels, H_0 and W_0 are the height and width of $X_{\rm in}$ and H and $W$are the height and width of $P^{i_{\rm th}}$ respectively ($H = H_0/\sqrt{m}$, $W = W_0/\sqrt{m}$, m is set to 16 in this paper). In the linear dependency capture path, horizontal and vertical pooling layers are applied to the input feature $P^{i_{\rm th}} \in \mathbb{R}^{C\times H\times W}$ first to generate four new feature maps $P_{x1}^{i_{\rm th}}$, $P_{x2}^{i_{\rm th}} \in \mathbb{R}^{C\times 1\times W}$($x_1$ and x_2 represent max-pooling and average-pooling in horizontal direction respectively) and $P_{y1}^{i_{\rm th}}$, $P_{y2}^{i_{\rm th}} \in \mathbb{R}^{C\times H\times 1}$($y_1$ and y_2 represent max-pooling and average-pooling in vertical direction respectively), which represent the linear dependency features captured in the horizontal and vertical directions respectively. Second, $P_{x1}^{i_{\rm th}}$, $P_{x2}^{i_{\rm th}}$, $P_{y1}^{i_{\rm th}}$ and $P_{y2}^{i_{\rm th}}$ are expanded to original image size, concatenated and followed by a 1×1 convolution and sigmoid activation to obtain the attention map $P'_{xy} \in \mathbb{R}^{H\times W}$, Third, element-wise multiplication is performed between P and the

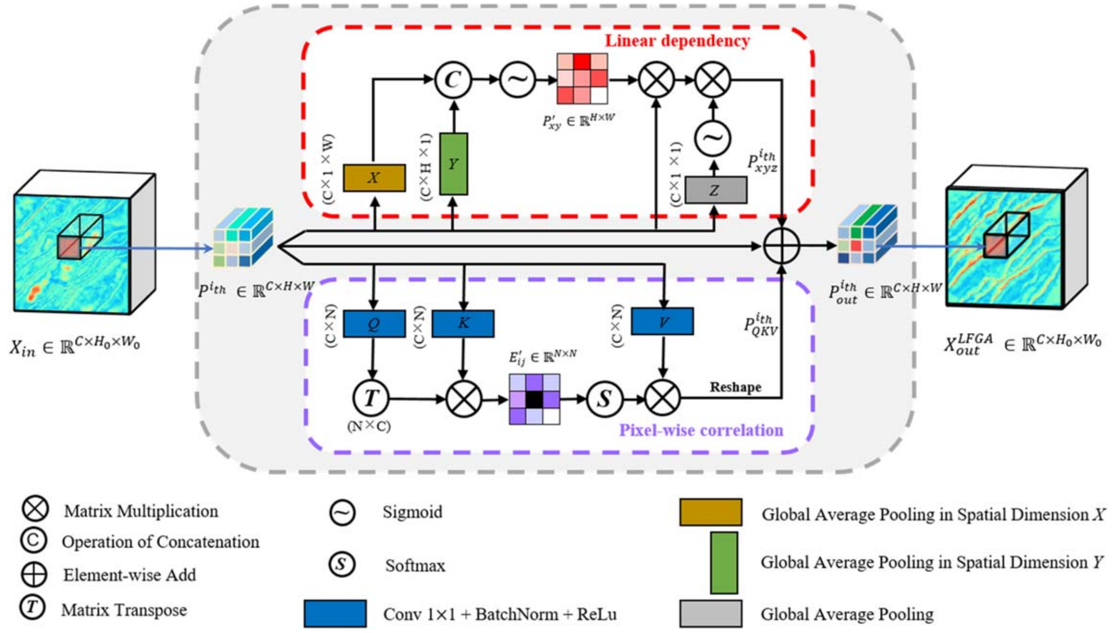
Linear dependency
C
$P'_{xy} \in \mathbb{R}^{H\times W}$
(C×1 ×W)
X
(C×H ×1)
Y
(C×1 ×1)
Z
$P^{i_{th}}_{xyz}$
$X_{in} \in \mathbb{R}^{C\times H_0\times W_0}$
$P^{i_{th}} \in \mathbb{R}^{C\times H\times W}$
$P^{i_{th}}_{out} \in \mathbb{R}^{C\times H\times W}$
$X^{LFGA}_{out} \in \mathbb{R}^{C\times H_0\times W_0}$
(C×N)
Q
(C×N)
K
(C×N)
V
$P^{i_{th}}_{QKV}$
$E'_{ij} \in \mathbb{R}^{N\times N}$
T
(N×C)
S
Reshape
Pixel-wise correlation
Matrix Multiplication
Operation of Concatenation
Element-wise Add
Matrix Transpose
Sigmoid
Softmax
Conv 1×1 + BatchNorm + ReLu
Global Average Pooling in Spatial Dimension X
Global Average Pooling in Spatial Dimension Y
Global Average Pooling

Figure 4. Local feature guided attention (LFGA) module. Pixel-wise correlation and Linear dependency are captured on local feature map to improve the network's ability to discriminate nerve fibers.

attention map P'_{xy} to generate a new feature map $P^{i_{\text{th}}}_{xy}$, which fuses linear dependencies both in horizontal and vertical directions. Fourth, a new feature $P^{i_{\text{th}}}_{z} \in \mathbb{R}^{C\times 1\times 1}$ is obtained from $P^{i_{\text{th}}}$ via a global average pooling layer followed by sigmoid activation to capture the global information. Finally, $P^{i_{\text{th}}}_{z}$ is multiplied with $P^{i_{\text{th}}}_{xy}$ to generate the output feature map P_{xyz}, which cannot only capture linear dependencies from different directions, but also obtain the global information. The construction of the linear dependency capture path can be formulated as follows,

$$P'_{xy} = \delta \, [\![f\,(\text{concat}[P^{i_{\text{th}}}_{x1}, P^{i_{\text{th}}}_{x2}, P^{i_{\text{th}}}_{y1}, P^{i_{\text{th}}}_{y2}])]\!] \in \mathbb{R}^{H\times W} \tag{4}$$

$$P^{i_{\text{th}}}_{xy} = \text{Mul}\,[\![P^{i_{\text{th}}}, P'_{xy}]\!] \in \mathbb{R}^{C\times H\times W} \tag{5}$$

$$P^{i_{\text{th}}}_{xyz} = \text{Mul}\,[\![\delta(P^{i_{\text{th}}}_{z}), P^{i_{\text{th}}}_{xy}]\!] \in \mathbb{R}^{C\times H\times W}, \tag{6}$$

where $\delta\,(\cdot)$ means sigmoid activation, $f\,(\cdot)$ means 1×1 convolution, Mul $[\![\,\cdot\,]\!]$ means element-wise multiplication, concat$[\cdot]$ means concatenation operation.

In pixel-wise correlation path, the input feature map $P^{i_{\text{th}}} \in \mathbb{R}^{C\times H\times W}$ is reshaped to $Q \in \mathbb{R}^{C\times N}$, $K \in \mathbb{R}^{C\times N}$ and $V \in \mathbb{R}^{C\times N}$ first, where $N = H\times W$. E'_{ij} is obtained by applying a softmax layer on the matrix multiplication of the transpose of Q and K, which denotes the j_{th} position's impact on the i_{th} position. Second, matrix multiplication is performed between E'_{ij} and V to obtain the attention enhanced feature and reshape it to $\mathbb{R}^{C\times H\times W}$ to construct the pixel-wise correlation based feature $P^{i_{\text{th}}}_{QKV}$, which can help the pixels belonging to the nerve fibers promote each other and enhance the contrast between nerve fibers and background. The acquisition of the pixel-wise correlation can be formulated as follows,

$$E'_{ij} = \frac{\exp(Q^T_y \cdot K_x)}{\sum_{x=1}^{N} \exp(Q^T_y \cdot K_x)} \in \mathbb{R}^{N\times N} \tag{7}$$

$$P^{i_{\text{th}}}_{QKV} = \sum\nolimits_{i=1}^{N} V_i E'_{ij} \in \mathbb{R}^{C\times H\times W} \tag{8}$$

Finally, the linear dependency feature $P^{i_{\text{th}}}_{xyz}$, the pixel-wise correlation feature $P^{i_{\text{th}}}_{QKV}$ and the original input feature $P^{i_{\text{th}}}$ are added to get feature $P^{i_{\text{th}}}_{\text{out}}$. $P^{i_{\text{th}}}_{\text{out}}$ features are reorganized into the final global reconstructed feature $X^{\text{LFGA}}_{\text{out}}$,

$$P^{i_{\text{th}}}_{\text{out}} = P^{i_{\text{th}}}_{xyz} + P^{i_{\text{th}}}_{QKV} + P^{i_{\text{th}}} \in \mathbb{R}^{C\times H\times W} \tag{9}$$

$$X^{\text{LFGA}}_{\text{out}} = \mathfrak{R}\left[\!\!\left[\sum_{i_{\text{th}}=1}^{m} P^{i_{\text{th}}}_{\text{out}} \right]\!\!\right] \in \mathbb{R}^{C\times H_0\times W_0} \tag{10}$$

where $\mathfrak{R}(\cdot)$ means the reorganization operation.

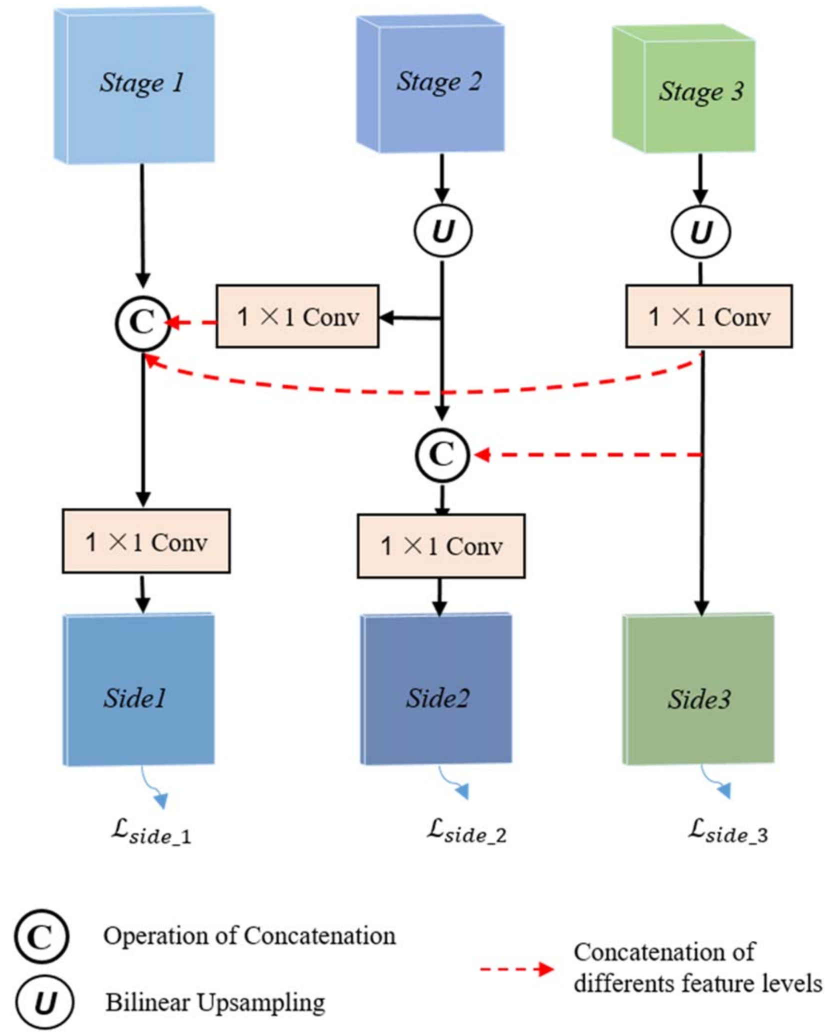

Figure 5. Multi-scale deep supervision (MDS) module. Semantic information from the deep layers are transfered to the shallow ones in the decoder path, which fully utilizes the relationship between high-level and low-level features for feature reconstruction.

2.5. Multi-scale deep supervision module

There are some empirical evidences demonstrating that the deeper the neural network, the more difficult it is to be optimized. Many previous studies utilized deep supervision mechanism to alleviate this problem (Fu *et al* 2018, Qin *et al* 2020, Wu *et al* 2021). M-Net (Fu *et al* 2018) generated side-output prediction maps as deep supervision, in which the weights for all the side-output layers are set equally. This may cause the ignorance of the different characteristics of the different side-output layers. SCS-Net (Wu *et al* 2021) set weights adaptively for the different side-output layers and got better performance. However, none of them have considered the relationship between high-level features and low-level features in the decoder path, which helps reduce the semantic information gap between different stages.

To fully utilize the relationship between high-stage and low-stage features for feature reconstruction in the decoder path, a novel MDS module is proposed, which is shown in figure 5. The MDS module enables feature map in each stage to interact with the ones in other stages to fuse information between low and high stages first. Then, the fused feature map is used to generate the side-outputs. Take the side-output of Stage 2 (Side_ 2) as an example. Feature map of Stage 3 is first up-sampled. Then, through a 1 × 1 convolution, it is concatenated with the feature map of Stage 2 to generate the fused feature map. Finally, the side-output of Stage 2 (Side_ 2) is generated by applying a 1 × 1 convolution to the fused feature map. The side-outputs transfer semantic information from the deep stages to the shallow ones, which can propagate effective information through backward propagation and let the auxiliary loss better optimize the network. The side-outputs from different stages can be formulated as follows,

$$\mathrm{Side\_1} = \psi(\mathrm{Up}[\mathrm{Stage3}]) \tag{11}$$

$$Side\_2 = \varphi(\mathrm{concat}[\mathrm{Up}[\mathrm{Stage2}],\, \psi(\mathrm{Up}[\mathrm{Stage3}]]) \tag{12}$$

Table 1. Details of the datasets used to evaluate the proposed method.

Datasets	Number (384 × 384)	Tortuosity level (numbers)	Public/Private	Source
Dataset 1	90	Normal (50) pathological (40)	Private	Zhongshan Ophthalmic Center
Dataset 2	114	Level 1 (30)	Public	Cixi Institute of Biomedical Engineering
		Level 2 (30)		
		Level 3 (30)		
		Level 4 (24)		
Dataset 3	30	Level 1 (10)	Public	University of Padova
		Level 2 (10)		
		Level 3 (10)		

$$\text{Side\_3} = \Theta(\text{concat}[\text{Stage1}, \phi(\text{Up}[\text{Stage2}]), \psi(\text{Up}[\text{Stage3}])]) \tag{13}$$

where Side_1, Side_2 and Side_3 represent the side-outputs of Stage 1, Stage 2 and Stage 3 respectively, $\psi(\cdot)$, $\varphi(\cdot)$, $\phi(\cdot)$ and $\Theta(\cdot)$ mean 1 × 1 convolution with parameters $\mathscr{W}^{\psi}$, $\mathscr{W}^{\varphi}$, $\mathscr{W}^{\phi}$ and $\mathscr{W}^{\Theta}$ respectively, Up$(\cdot)$means the bilinear interpolation, and concat $[\![\cdot]\!]$ means the concatenation operation.

2.6. Training and inference detail

To overcome the data imbalance problem, the Dice loss is used as the segmentation cost function $\mathscr{L}_{\text{seg}}$ to make the network pay more attention to the foreground. The binary cross entropy (BCE) loss is used as the deep supervision loss $\mathscr{L}_{\text{side\_}k}$ to fully leverage the multiple side-outputs.

$$\mathscr{L}_{\text{seg}} = 1 - \frac{2\sum_{i=1}^{N} g_i p_i + \varepsilon}{\sum_{i=1}^{N}(g_i + p_i) + \varepsilon} \tag{14}$$

$$\mathscr{L}_{\text{side\_}k} = -\frac{1}{N}\sum_{i=1}^{N} g_i \cdot \log(p_i) + (1 - g_i) \cdot \log(1 - p_i)(k = 1, 2, 3), \tag{15}$$

where N represents the total number of pixels in each feature map, $g_i \in \{0,1\}$ and $p_i \in [0,1]$ represent the value of the i_{th} pixel in the ground truth and the predicted probability map respectively. In order to speed up the convergence in the network training, the Laplace smoothing factor ε is added to $\mathscr{L}_{\text{seg}}$, is which is set to 1 in our experiments.

The total loss function $\mathscr{L}_{\text{total}}$ adopted for the proposed MLFGNet is the combination of $\mathscr{L}_{\text{seg}}$ and $\mathscr{L}_{\text{side\_}k}$, which is defined as follows,

$$\mathscr{L}_{\text{total}} = \mathscr{L}_{\text{seg}} + \sum_{k=1}^{3} \lambda_k \mathscr{L}_{\text{side}_k}(16)$$

where λ_k is a trade-off between segmentation loss $\mathscr{L}_{\text{seg}}$ and deep supervision loss $\mathscr{L}_{\text{side\_}k}$. In our experiments, $\lambda_1 = 1$, $\lambda_2 = 0.8$ and $\lambda_3 = 0.4$.

2.7. Datasets

As shown in table 1, three CCM image Datasets were used to evaluate the performance of the proposed MLFGNet. Images in all three Datasets were obtained by Heidelberg Retina Tomograph with a Rostock Cornea Module microscope. The size of all images is 384 × 384. The field of view (FOV) is 400 μm × 400 μm. The ground truth of three datasets are manually labeled under the supervision of an ophthalmologist with extensive clinical experience.

Dataset 1 is a private dataset acquired from Zhongshan Ophthalmic Center, China, containing 90 two-dimensional CCM images. Among them, 50 images were taken from four normal eyes and 40 from four eyes with diabetic keratopathy. The collection and analysis of image data were approved by the Institutional Review Board of Zhongshan Ophthalmic Center and adhered to the tenets of the Declaration of Helsinki.

Dataset 2 is a public CCM image dataset with 404 images, provided by the Cixi Institute of Biomedical Engineering, Chinese Academy of Sciences (Mou *et al* 2021). According to the tortuosity of never fiber, the images were graded into levels 1–4. As there are only 24 images in level 4, 30 CCM images are randomly selected from each other three levels to keep the number of images of different levels balanced. So 114 images are included in Dataset 2.

Dataset 3 is also a public CCM image dataset provided by the Laboratory of Biomedical Imaging, University of Padova, Italy (Scarpa *et al* 2011), which contains 30 CCM images from 30 normal or pathological subjects (diabetes, pseudoexfoliation syndrome, and keratoconus).

Table 2. Ablation experiments about the proposed MFPG, LFGA, and MDS modules on Dataset 1 (Mean ± Standard deviation).

MFPG	LFGA	MDS	Dice (%)	IoU (%)	Sen (%)	AUC (%)	p-value
◊	◊	◊	87.34 ±5.02	77.85 ±7.55	86.75 ±5.34	92.78 ±2.69	—
◊	◊	√	88.25 ±4.82	79.29 ±7.39	88.72 ±4.93	93.75 ±2.50	<0.001
◊	√	◊	87.69 ±4.62	78.35 ±7.05	86.90 ±4.94	92.88 ±2.47	0.004
√	◊	◊	87.97 ±4.69	78.82 ±7.19	88.59 ±4.65	93.65 ±2.31	<0.001
◊	√	√	88.45 ±4.51	79.57 ±6.98	89.28 ±3.96	94.02 ±2.01	<0.001
√	◊	√	88.78 ±4.71	80.13 ±7.31	88.24 ±5.22	93.60 ±2.64	<0.001
√	√	◊	88.27 ±4.90	79.34 ±7.54	87.69 ±4.91	93.30 ±2.50	<0.001
√	√	√	**89.33** ±4.22	**80.97** ±6.62	**88.73** ±5.02	**93.86** ±2.50	<0.001

A four-fold cross-validation strategy is used to objectively evaluate the proposed MLFGNet, which is also adopted in all the comparison experiments. For Dataset 1, the images were randomly divided into four groups according to subjects, and each group contained both normal and pathological subjects. For Dataset 2 and Dataset 3, the images were randomly divided into four folds according to the levels of tortuosity.

2.8. Evaluation metrics

Five evaluation metrics including Dice, intersection over union (IoU), sensitivity (Sen), specificity (Spe) and area under the ROC curve (AUC) are employed in our experiments. To evaluate the statistical significance of the improvement, the Wilcoxon signed-rank test between the proposed MLFGNet and other methods is conducted on Dice coefficient in both comparison and ablation experiments.

2.9. Implementation details

All the experiments are implemented on the public Pytorch platform and NVIDIA RTX 2080Ti with 11GB memory. In the training process, stochastic gradient descent (SGD) algorithm with poly learning rate policy is used to optimize the weights of the network. The learning rate lr is as follows,

$$lr = lr_b \times \left(1 - \frac{Iter}{Iter_t}\right)^p \tag{17}$$

where $Iter$ and $Iter_t$ represent the current number of iterations and the total number of iterations respectively. The basic learning rate lr_b is set to 0.01 and the declining index p is set to 0.9. The batch size is set to 2 and the epochs is set to 80.

3. Results and discussion

3.1. Ablation experiments

3.1.1. Ablation experiments about the proposed modules

Eight ablation experiments about the proposed LFGA module, MFPG module, and MDS module are conducted on Dataset 1, Dataset 2, and Dataset 3, including Baseline (only including the encoder path and the decoder path shown in figure 2), Baseline with MDS module (Baseline + MDS), Baseline with LFGA module (Baseline + LFGA), Baseline with MFPG module (Baseline + MFPG), Baseline with MDS and LFGA module (Baseline + MDS + LFGA), Baseline with MFPG and MDS module (Baseline + MFPG + MDS), Baseline with MFPG and PA module (Baseline + MFPG + LFGA) and Baseline with MFPG, LFGA and MDS module (Baseline + MFPG + LFGA + MDS, the proposed MLFGNet). The results of ablation experiments on Dataset 1, Dataset 2, and Dataset 3 are shown in tables 2, 3 and 4 respectively. For Dataset 1, compared with the Baseline, the improvements on Dice index are 0.91%, 0.35%, 0.63%. 1.11%, 1.44%, 0.93% and 1.99% and reach 88.25%, 87.69%, 87.97%, 88.45%, 88.78%, 88.27% and 89.33% respectively. For Dataset 2, as shown in table 3, the improvements on Dice index are 0.8%, 0.26%, 0.56%, 0.85%, 1.27%, 0.87% and 1.43% and reach 88.78%, 88.24%, 88.54%, 88.83%, 89.25%, 88.85%, and 89.41% respectively. For Dataset 3, as shown in table 4, the improvements on Dice index are 1.34%, 1.12%, 1.58%, 3.17%, 4.22%, 3.3%, and 5.48%, and reach 84.15%, 83.93%, 84.39%, 85.98%, 87.03% ,86.11%, and 88.29% respectively. All the p-values (Dice index, Wilcoxon signed-rank test) are less than 0.05, indicating that all the proposed modules have achieved significant improvements compared with the Baseline.

Figure 6 shows the segmentation results of ablation experiments, which show that the proposed MFPG module, LFGA module, and MDS module can improve the segmentation performance of the proposed network

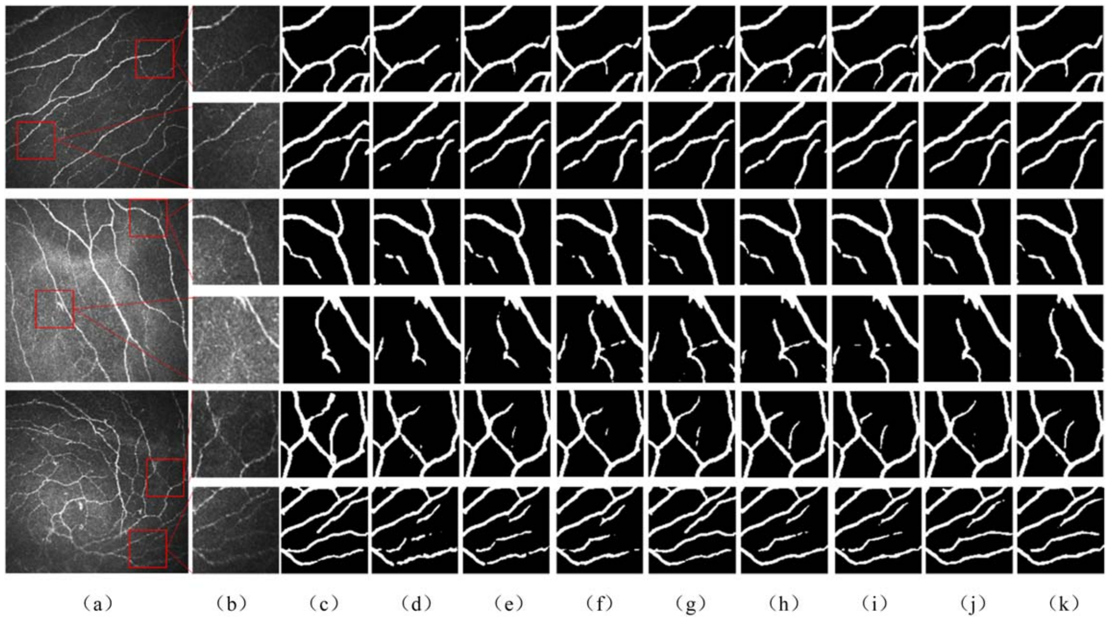

Figure 6. The segmentation results of ablation experiments in our study. (a) Original images; (b) detailed view; (c) ground truth; (d) Baseline; (e) Baseline + MFPG; (f) Baseline + LFGA; (g) Baseline + MDS; (h) Baseline + MFPG + LFGA; (i) Baseline + MFPG + MDS; (j) Baseline + LFGA + MDS; (k) proposed.

Table 3. Ablation experiments about the proposed MFPG, LFGA, and MDS modules on Dataset 2 (Mean ± Standard deviation).

MFPG	LFGA	MDS	Dice (%)	IoU (%)	Sen (%)	AUC (%)	*p*-value
◊	◊	◊	87.98 ±4.45	78.80 ±6.50	87.60 ±6.51	93.20 ±3.18	—
◊	◊	√	88.78 ±3.89	80.04 ±6.03	88.42 ±5.37	93.66 ±2.64	<0.001
◊	√	◊	88.24 ±4.17	79.19 ±6.31	87.54 ±6.44	93.19 ±3.14	0.004
√	◊	◊	88.54 ±4.06	79.67 ± 6.26	88.00 ± 5.75	93.45 ± 2.83	<0.001
◊	√	√	88.83 ± 4.32	80.15 ± 6.58	88.25 ± 6.62	93.57 ± 3.22	<0.001
√	◊	√	89.25 ± 3.84	80.79 ± 5.99	89.02 ± 5.51	93.96 ± 2.71	<0.001
√	√	◊	88.85 ± 3.82	80.14 ± 5.98	87.85 ± 5.81	93.40 ± 2.84	<0.001
√	√	√	**89.41** ± 3.74	**81.05** ± 5.88	**88.38** ± 5.87	**93.69** ± 2.87	<0.001

Table 4. Ablation experiments about the proposed MFPG, LFGA, and MDS modules on Dataset 3 (Mean ± Standard deviation).

MFPG	LFGA	MDS	Dice (%)	IoU (%)	Sen (%)	AUC (%)	*p*-value
◊	◊	◊	82.81 ± 5.57	71.04 ± 8.06	82.95 ± 7.59	90.84 ± 3.72	—
◊	◊	√	84.15 ± 5.62	73.03 ± 8.36	82.82 ± 8.05	90.91 ± 3.97	<0.001
◊	√	◊	83.93 ± 5.00	72.61 ± 7.35	83.38 ± 6.99	91.13 ± 3.42	0.004
√	◊	◊	84.39 ± 5.48	73.37 ± 8.14	84.53 ± 7.02	91.68 ± 3.46	<0.001
◊	√	√	85.98 ± 5.27	75.77 ± 8.06	86.85 ± 5.41	92.85 ± 2.72	<0.001
√	◊	√	87.03 ± 5.08	77.38 ± 7.82	86.53 ± 6.37	92.81 ± 3.19	<0.001
√	√	◊	86.11 ± 4.63	75.90 ± 7.03	86.07 ± 7.03	92.52 ± 3.49	<0.001
√	√	√	**88.29** ± 4.11	**79.27** ± 6.51	**87.43** ± 6.09	**93.31** ± 3.04	<0.001

effectively. It can also be seen from figure 6 that with the addition of the proposed modules, the breaks in the topological structure of the nerve fibers are gradually filled and the connectivity is higher.

3.1.2. Ablation experiments about backbones

Table 5 shows the ablation studies of different backbones on all three datasets. The results of MLFGNet/Res (only use ResNet34 as backbone) and MLFGNet/Conv-block (only use conv-block as backbone) on three

Table 5. Ablation experiments about different backbones (Mean ± Standard deviation).

Methods	Dice (%)	IoU (%)	Sen (%)	AUC (%)
		Dataset 1		
MLFGNet/Res	88.37 ± 3.67	79.46 ± 5.68	87.49 ± 4.80	93.24 ± 2.30
MLFGNet/Conv-block	88.32 ± 4.54	79.27 ± 7.01	88.10 ± 4.80	93.46 ± 2.42
MLFGNet	**89.33** ± 4.22	**80.97** ± 6.62	**88.73** ± 5.02	**93.86** ± 2.50
		Dataset 2		
MLFGNet/Res	87.57 ± 3.96	78.10 ± 5.96	86.90 ± 6.24	92.83 ± 3.02
MLFGNet/Conv-block	88.65 ± 4.10	79.84 ± 6.30	87.93 ± 5.93	93.43 ± 2.91
MLFGNet	**89.41** ± 3.74	**81.05** ± 5.88	**88.38** ± 5.87	**93.69** ± 2.87
		Dataset 3		
MLFGNet/Res	86.85 ± 4.96	77.08 ± 7.62	85.70 ± 7.14	92.42 ± 3.54
MLFGNet/Conv-block	87.31 ± 5.05	77.81 ± 7.78	86.45 ± 7.37	92.79 ± 3.67
MLFGNet	**88.29** ± 4.11	**79.27** ± 6.51	**87.43** ± 6.09	**93.31** ± 3.04

Table 6. Ablation experiments about different patch numbers in LFGA module on Dataset 1 (Mean ± Standard deviation).

m-value	Dice (%)	IoU (%)	Sen (%)	AUC (%)
1	88.52 ± 4.57	79.70 ± 7.05	88.48 ± 4.47	93.68 ± 2.25
4	88.18 ± 4.93	79.19 ± 7.51	87.12 ± 5.49	93.04 ± 2.76
16(**selected**)	**89.33** ± 4.22	**80.97** ± 6.62	**88.73** ± 5.02	**93.86** ± 2.50
36	88.55 ± 4.61	79.74 ± 7.10	87.58 ± 4.80	93.29 ± 2.41
64	88.20 ± 4.91	79.22 ± 7.52	87.18 ± 5.27	93.06 ± 2.68
144	88.32 ± 4.87	79.40 ± 7.43	87.60 ± 5.07	93.27 ± 2.56

datasets are not good, which indicate that only using conv-block as backbone is too shallow to extract deep semantic information and lead to the limited receptive field, resulting in the degradation of the nerve fiber segmentation performance. Only using ResNet34 as backbone can deepen the depth of the network and increase the receptive field. Meanwhile, it will destroy the thin and curvilinear structure of nerve fibers and thus lead to the decline in performance. The best results can be obtained in the proposed MLFGNet, in which 2 conv-blocks and 2 ResNet34 layers are jointly applied. Specifically, 2 conv-blocks are used in the first and second layers of the encoder (Level 1 and Level 2) to retain more spatial information of nerve fibers. 2 pre-trained ResNet34 layers are used in the third and fourth layers (Level 3 and Level 4), which can deepen the network, expand the receptive field, and obtain rich semantic information. This combination backbone strategy takes into account the spatial information and semantic information of nerve fibers in the feature extraction and achieves the best segmentation performance.

3.1.3. Ablation experiments about patch numbers in LFGA module

Table 6 shows the ablation studies about different patch numbers (m) of LFGA module on Dataset 1. As can be seen from table 6, too many or too few patches will affect the segmentation performance. When $m = 16$, the best result is achieved, which means that the original feature map is split into 16 patches. We select $m = 16$ in all experiments.

Figure 7 shows the predicted probabilities of pixels as curvilinear structures before and after applying the proposed LFGA module respectively. As can be seen from figure 7, without LFGA module, the curvilinear structures have not been highlighted consistently, which means that the network does not extract significant curvilinear features. With LFGA module, the network can focus on the curvilinear structures well and suppress the influence of bright background noise, which means that the LFGA module can extract and aggregate the curvilinear structure features and enhance the network’s ability to distinguish nerve fibers.

3.2. Comparison experiments

In order to evaluate the performance of our proposed MLFGNet, the proposed segmentation framework is compared with some state-of-the-art deep learning based segmentation networks, including CS-Net (Mou *et al* 2019), CPFNet (Zhang *et al* 2020), U-Net++ (Zhou *et al* 2018), U^2-Net (Qin *et al* 2020), CE-Net (Colonna *et al* 2018), U-Net (Chen *et al* 2017), Attention U-Net (Scarpa *et al* 2011), MDACN (Guo *et al* 2019), and MMDC-Net (Chen *et al* 2017). All the experimental implementation details are kept consistent for fair including four-fold cross validation strategy, learning rate strategy, batch size, etc. The comparison experiment results on the three datasets are shown in tables 7, 8, and 9 respectively. As can be seen from the tables, the proposed MLFGNet

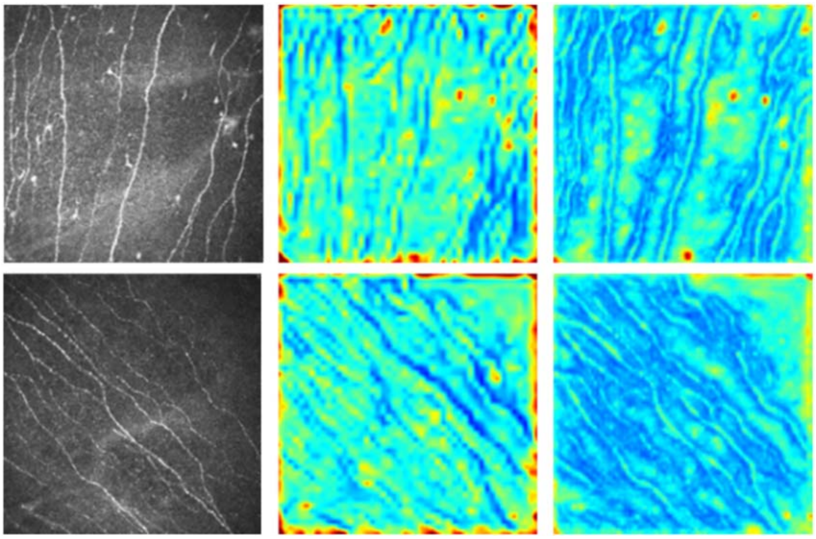

Figure 7. The output of the proposed LFGA module on randomly select images. From left to right: the original image, the predicted probabilities of pixels as curvilinear structure before and after applying the proposed LFGA module respectively.

Table 7. Comparison results on Dataset 1 (Mean ± Standard deviation).

Methods	Dice (%)	IoU (%)	Sen (%)	AUC (%)	p-value
CS-Net (Mou *et al* 2019)	84.95 ± 6.44	74.33 ± 8.97	84.15 ± 5.74	91.42 ± 2.93	<0.001
CPFNet (Feng *et al* 2020)	85.44 ± 4.55	74.85 ± 6.69	84.78 ± 5.37	91.70 ± 2.61	<0.001
U-Net++ (Zhou *et al* 2018)	86.90 ± 5.23	77.20 ± 7.80	85.57 ± 5.61	92.23 ± 2.84	<0.001
U^2-Net (Qin *et al* 2020)	87.15 ± 5.32	77.6 ± 7.88	85.51 ± 5.92	92.23 ± 2.96	<0.001
CE-Net (Gu *et al* 2019)	87.25 ± 4.28	77.63 ± 6.49	86.44 ± 5.13	92.62 ± 2.50	<0.001
U-Net (Ronneberger *et al* 2015)	87.32 ± 5.66	77.91 ± 8.38	86.12 ± 5.44	92.52 ± 2.78	<0.001
MMDC-Net (Zhong *et al* 2022)	87.51 ± 4.45	78.06 ± 6.80	87.12 ± 4.75	92.99 ± 2.33	<0.001
Attention U-Net (Oktay *et al* 2018)	87.42 ± 4.97	77.98+7.48	87.16 ± 5.18	92.98 ± 2.59	<0.001
MDACN (Yang *et al* 2021)	89.21 ± 4.13	80.77 ± 6.48	88.40 ± 4.90	93.72 ± 2.45	0.45
MLFGNet	**89.33** ± 4.22	**80.97** ± 6.62	**88.73** ± 5.02	**93.86** ± 2.50	—

Table 8. Comparison results on Dataset 2 (Mean ± Standard deviation).

Methods	Dice (%)	IoU (%)	Sen (%)	AUC (%)	p-value
CS-Net (Mou *et al* 2019)	85.32 ± 5.33	74.74 ± 7.45	83.48 ± 6.59	91.14 ± 3.24	<0.001
CPFNet (Feng *et al* 2020)	85.13 ± 4.31	74.34 ± 6.26	83.53 ± 6.56	91.08 ± 3.19	<0.001
U-Net++ (Zhou *et al* 2018)	87.17 ± 4.37	77.52 ± 6.51	86.41 ± 5.63	92.61 ± 2.78	<0.001
U^2-Net (Qin *et al* 2020)	87.46 ± 4.19	77.95 ± 6.36	86.21 ± 6.28	92.55 ± 3.07	<0.001
CE-Net (Gu *et al* 2019)	86.81 ± 4.51	76.95 ± 6.64	85.53 ± 7.02	92.16 ± 3.41	<0.001
U-Net (Ronneberger *et al* 2015)	87.79 ± 4.71	78.51 ± 6.82	86.62 ± 6.51	92.77 ± 3.20	<0.001
MMDC-Net (Zhong *et al* 2022)	87.61 ± 4.42	78.21 ± 6.69	86.24 ± 5.74	92.57 ± 2.82	<0.001
Attention U-Net (Oktay *et al* 2018)	88.17 ± 3.86	79.04 ± 5.97	87.31 ± 5.50	93.09 ± 2.70	<0.001
MDACN (Yang *et al* 2021)	89.08 ± 4.13	80.53 ± 6.16	88.12 ± 6.00	93.55 ± 2.83	<0.001
MLFGNet	**89.41** ± 3.74	**81.05** ± 5.88	**88.38** ± 5.80	**93.69** ± 2.87	—

outperforms all the state-of-the-art networks on Dataset 1, Dataset 2, and Dataset 3. To evaluate the statistical significance of the improvement, the Wilcoxon signed-rank test is conducted on Dice coefficient. As can be seen from tables 8 and 9, all the p-values are less than 0.05, indicating that the proposed MLFGNet has achieved a significant improvement compared with other networks on Dataset 2 and Dataset 3. As shown in table 7, although the proposed MLFGNet reaches the best performance, the p-value between the proposed MLFGNet and MDACN is 0.45, indicating that there is no significant difference between them. The possible reason is that more than half of the CCM images in Dataset 1 are normal (50/90), that is, most of the CCM images contain nerve fibers with low grades of tortuosity, which are relatively easy to be segmented. The proposed MLFGNet

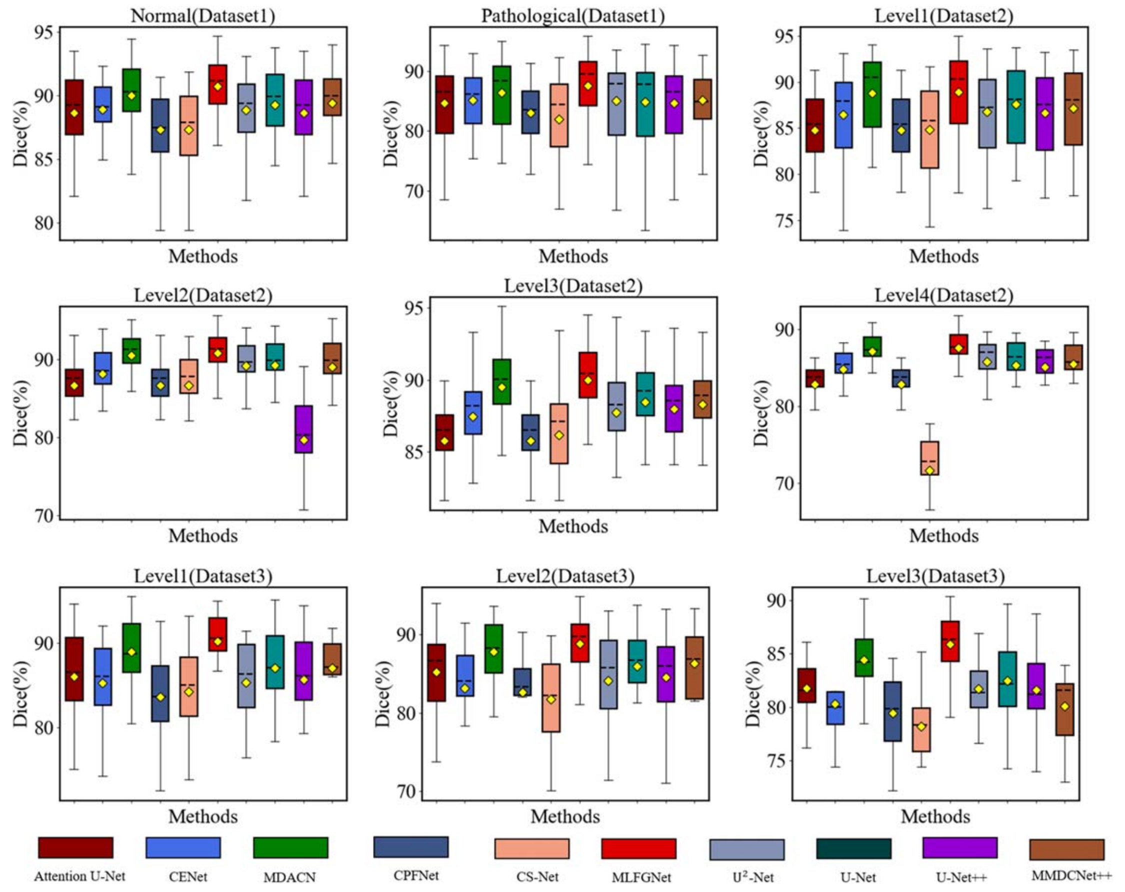

Figure 8. Performances of different methods on all three datasets. Box plots show the median as the center lines, upper and lower quartiles as box limits, whiskers as maximum and minimum values, yellow diamond as average values. The *y*-axis represents the dice coefficient, the *x*-axis represents the different methods.

Table 9. Comparison results on Dataset 3 (Mean ± Standard deviation).

Methods	Dice (%)	IoU (%)	Sen (%)	AUC (%)	*p*-value
CS-Net (Mou *et al* 2019)	81.37 ± 6.30	69.05 ± 8.90	80.90 ± 7.14	89.80 ± 3.57	<0.001
CPFNet (Feng *et al* 2020)	81.90 ± 5.15	69.66 ± 7.36	81.20 ± 7.61	89.93 ± 3.78	<0.001
U-Net++ (Zhou *et al* 2018)	83.95 ± 5.84	72.76 ± 8.57	82.49 ± 8.28	90.74 ± 4.09	<0.001
U^2-Net (Qin *et al* 2020)	83.72 ± 5.78	72.40 ± 8.47	82.54 ± 7.84	90.74 ± 3.86	<0.001
CE-Net (Gu *et al* 2019)	82.92 ± 5.81	71.22 ± 8.28	82.25 ± 8.34	90.53 ± 4.11	<0.001
U-Net (Ronneberger *et al* 2015)	85.14 ± 5.41	74.51 ± 8.15	84.20 ± 7.09	91.61 ± 3.53	<0.001
MMDC-Net (Zhong *et al* 2022)	84.46 ± 4.99	73.42 ± 7.48	84.30 ± 5.45	91.63 ± 2.65	<0.001
Attention U-Net (Oktay *et al* 2018)	84.35 ± 5.67	73.34 ± 8.47	83.07 ± 7.82	91.03 ± 3.87	<0.001
MDACN (Yang *et al* 2021)	87.62 ± 4.27	78.22 ± 6.78	86.79 ± 5.39	92.97 ± 2.65	<0.001
MLFGNet	**88.29** ± 4.11	**79.27** ± 6.51	**87.43** ± 6.09	**93.31** ± 3.04	—

achieves stable segmentation performance on all three datasets, indicating that our network is suitable for nerve fiber segmentation with different grades of tortuosity.

To evaluate the effect of different grades of tortuosity on the nerve fiber segmentation performance, the Dice indexes are re-calculated according to the tortuosity levels of the image on all three datasets respectively. The performances of different methods on all three datasets are shown as box plots in figure 8. The proposed MLFGNet shows the best performance on all three datasets, which also indicates our network is designed to deal with nerve fibers from both normal subjects and pathological ones.

Figure 9 shows some nerve fiber segmentation results of different networks, including both normal (Row 1) and pathological (Row 2 and 3) subjects. As can be seen from figure 9, all methods can achieve good segmentation results for thick nerve fibers, but most of them may fail to detect the faint and thin ones. The proposed MLFGNet can capture these faint and thin nerve fibers well due to the purposely designed LFGA

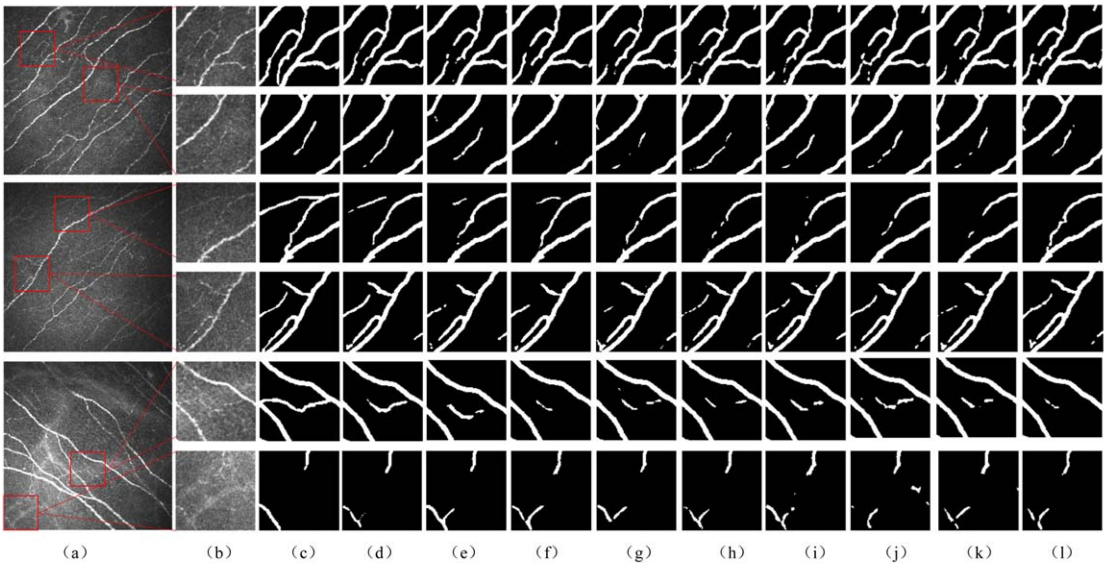

Figure 9. Segmentation results of different methods. (a) Original images; (b) detailed view; (c) ground truth; (d) the proposed MLFGNet; (e) MDACN; (f) attention U-Net; (g) U-Net; (h) U^2-Net; (i) U-Net++; (j) CE-Net; (k) CS-Net; (l) MMDC-Net.

Table 10. Cross-dataset evaluation (Dataset 2: training set, Dataset 3: validation set, and Dataset 1: test set).

Methods	Dice (%)	IoU (%)	Sen (%)	AUC (%)	p-value
CS-Net (Mou *et al* 2019)	84.20	73.12	85.24	91.82	<0.01
CPFNet (Feng *et al* 2020)	85.59	75.02	86.46	92.45	<0.01
U-Net++ (Zhou *et al* 2018)	85.96	75.63	84.95	91.87	<0.01
U^2-Net (Qin *et al* 2020)	86.37	76.29	85.19	92.01	<0.01
CE-Net (Gu *et al* 2019)	86.29	76.11	87.71	93.32	<0.01
U-Net (Ronneberger *et al* 2015)	85.90	75.56	86.45	92.52	<0.01
MMDC-Net (Zhong *et al* 2022)	87.56	78.10	87.61	93.16	<0.01
Attention U-Net (Oktay *et al* 2018)	86.33	76.20	84.37	91.65	<0.01
MDACN (Yang *et al* 2021)	87.97	78.72	**88.10**	**93.42**	0.23
MLFGNet	**88.12**	**78.98**	87.94	93.37	—

module, which can make the network pay more attention to local feature information and improve its ability to discriminate nerve fibers with low contrast. The segmentation results of the proposed method are of higher continuity, with more faint and thin fiber segments detected.

To verify the robustness of the proposed MLFGNet, a cross-dataset evaluation is performed, in which Dataset 2 (114 images) is used as the training set, Dataset 3 (30 images) is adopted as the validation set, and Dataset 1 (90 images) is used as the test set. The same cross-dataset evaluation strategy is also adopted for other methods. As shown in table 10, the proposed MLFGNet achieves the best performance with an average Dice coefficient of 88.12%. The p-value of the Wilcoxon signed-rank test indicates that there is no significant difference between the proposed MLFGNet and MDACN. The possible reason is that the proportion of thick nerve fibers is generally higher than that of thin ones in the CCM images, which means although our method segments thin nerve fibers better than MDACN, the Dice index is not obviously improved. However, figure 9 shows that the segmentation results of the proposed method have fewer breaks and thus better topological connectivity.

To further verify the effectiveness and superiority of the proposed method, the intermediate feature maps from different stages of the decoder path in our proposed MLFGNet and three state-of-the-art networks including MDACN, U-Net and CE- Net are visualized. Figure 10(a) are the original images and the corresponding ground truth. As shown in figures 10(b) and (c), by analyzing and comparing the nerve fibers in the attention maps from Stage 1 to Stage 3, we note that the proposed model can focus on curvilinear structures. The curvilinear structures gradually become brighter and clearer from top to bottom. In high-stage feature maps, the proposed MLFGNet can clearly focus on the curvilinear structures, while they are almost invisible in the comparison methods. In low-stage feature maps, the highlighted areas are mainly distributed around the curvilinear structures with purer background in the proposed MLFGNet. While the comparison methods either

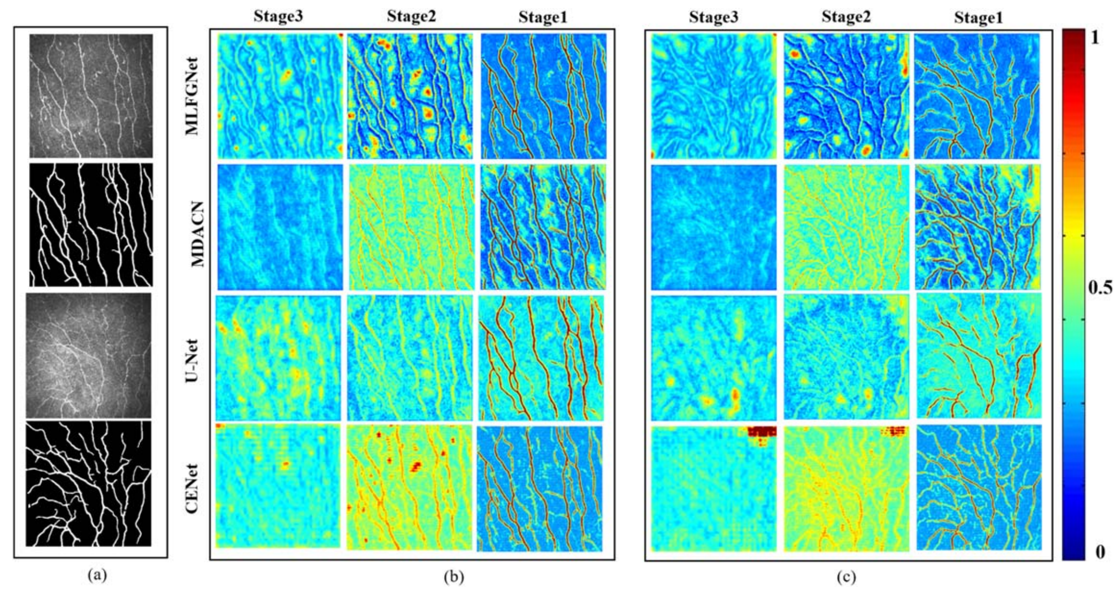

Figure 10. The output feature maps of intermediate stages of the decoder path. (a) Two original images and the corresponding ground truth; (b) and (c) feature maps for these two original images. From top to bottom: the feature maps of the proposed MLFGNet, MDACN, U-Net and CE-Net. From left to right: Stage 3, Stage 2 and Stage 1 display the feature maps representing the incremental refinement for nerve fiber segmentation.

loses a lot of curvilinear structures, or the background is too cluttered. The final segmentation results also show that our method can detect more nerve fibers. Overall, it can be observed from the comparison of each column that the proposed MLFGNet has a stronger response than the comparison methods which proves the proposed network has a higher ability to capture the curvilinear structure of nerve fibers and is more powerful in suppressing the background interference.

4. Conclusion

In this paper, we propose an end-to-end deep learning based framework named MLFGNet for nerve fiber segmentation in confocal corneal microscopy images. Based on the U-shape encoder–decoder structure, the proposed multi-scale feature progressive guidance (MFPG) modules are embedded as skip connections, in which information is progressively aggregated from high-level features to low-level ones to shrink the information gap between different levels. A novel LFGA module is proposed and embedded into the top of the encoder, which splits the feature map into m patches and pixel-wise correlation and linear dependency are captured in parallel on each patch. LFGA module enables the network to pay attention to local feature information and improves the network's ability to discriminate nerve fibers with low contrast. The multi-scale deep supervision (MDS) module is proposed to fully utilize the relationship between high-stage and low-stage features for feature reconstruction in the decoder path. The proposed MLFGNet is evaluated on three CMM image datasets and achieves state-of-the-art performance.

Although the proposed MLFGNet performs well for nerve fiber segmentation on three CCM image datasets, there are still some limitations. The hyper-parameters such as patch number in the LFGA module and the trade-off λ_k in the loss function used in this paper may be invalid for new datasets due to the possible domain shift problem and need to be re-tuned. The domain adaptive ability of the transfer learning (Sahu *et al* 2021) may be adopted to relieve this problem. There is room for segmentation performance improvement. Graph convolutions are good at capturing topology structure of the target (Shin *et al* 2019), which may help us further improve the segmentation performance. In future work, we will still try to improve the nerve fiber segmentation performance in both normal and pathological images, and we will further investigate the tortuosity classification based on the nerve fiber segmentation results.

Acknowledgments

This study was supported in part by the National Key Research and Development Program of China under Grant 2018YFA0701700, part by the National Nature Science Foundation of China under Grant U20A20170 and

Grant 62271337, and part by Natural Science Research of Jiangsu Higher Education Institutions of China under Grant 21KJB510021.

Data availability statement

The data cannot be made publicly available upon publication because they contain sensitive personal information. The data that support the findings of this study are available upon reasonable request from the authors.

Conflict of interest statement

The authors declare that they have no conflict of interests regarding this paper.

ORCID iDs

Wei Tang https://orcid.org/0000-0001-7580-4132
Dehui Xiang https://orcid.org/0000-0001-7873-9778
Weifang Zhu https://orcid.org/0000-0001-9540-4101

References

Annunziata R, Kheirkhah A, Aggarwal S, Hamrah P and Trucco E 2016 A fully automated tortuosity quantification system with application to corneal nerve fibres in confocal microscopy images *Med. Image Anal.* **32** 216–32

Chen L C, Papandreou G, Kokkinos I, Murphy K and Yuille A L 2017 Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs *IEEE Trans. Pattern Anal. Mach. Intell.* **40** 834–48

Chen X, Graham J, Dabbah M A, Petropoulos I N, Tavakoli M and Malik R A 2016 An automatic tool for quantification of nerve fibers in corneal confocal microscopy images *IEEE Trans. Biomed. Eng.* **64** 786–94

Chen Z, Mo Y, Chen J and Mo J 2021 Corneal nerve fiber segmentation and centerline extraction *Optics in Health Care and Biomedical Optics XI* vol 11900 (Nantong, Jiangsu, China: SPIE) pp 196–201

Colonna A, Scarpa F and Ruggeri A 2018 Segmentation of corneal nerves using a U-Net-based convolutional neural network *Computational Pathology and Ophthalmic Medical Image Analysis* (Cham: Springer) pp 185–92

Dabbah M A, Graham J, Petropoulos I, Tavakoli M and Malik R A 2010 Dual-model automatic detection of nerve-fibres in corneal confocal microscopy images *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention* (Berlin, Heidelberg: Springer) pp 300–7

Dabbah M A, Graham J, Petropoulos I, Tavakoli M and Malik R A 2011 Automatic analysis of diabetic peripheral neuropathy using multi-scale quantitative morphology of nerve fibres in corneal confocal microscopy imaging *Med. Image Anal.* **15** 738–47

Daousi C, MacFarlane I A, Woodward A, Nurmikko T J, Bundred P E and Benbow S J 2004 Chronic painful peripheral neuropathy in an urban community: a controlled comparison of people with and without diabetes *Diabetic Med.* **21** 976–82

Feng S, Zhao H, Shi F, Cheng X, Wang M, Ma Y, Xiang D, Zhu W and Chen X 2020 CPFNet: context pyramid fusion network for medical image segmentation *IEEE Trans. Med. Imaging* **39** 3008–18

Ferreira A, Morgado A M and Silva J S 2012 A method for corneal nerves automatic segmentation and morphometric analysis *Comput. Methods Programs Biomed.* **107** 53–60

Fu H, Cheng J, Xu Y, Wong D W K, Liu J and Cao X 2018 Joint optic disc and cup segmentation based on multi-label deep network and polar transformation *IEEE Trans. Med. Imaging* **37** 1597–605

Gu Z, Cheng J, Fu H, Zhou K, Hao H, Zhao Y, Zhao T, Gao S and Liu J 2019 CE-Net: context encoder network for 2d medical image segmentation *IEEE Trans. Med. Imaging* **38** 2281–92

Guo S, Wang K, Kang H, Zhang Y, Gao Y and Li T 2019 BTS-DSN: deeply supervised neural network with short connections for retinal vessel segmentation *Int. J. Med. Informatics* **126** 105–13

He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition* pp 770–8

Hou Q, Zhang L, Cheng M M and Feng J 2020 Strip pooling: rethinking spatial pooling for scene parsing *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition* pp 4003–12

Kang M S and Kim C H 2015 Management of Diabetic Peripheral Neuropathy *Clinical Diabetes* **23** 9–15

Kemp H I *et al* 2017 Use of corneal confocal microscopy to evaluate small nerve fibers in patients with human immunodeficiency virus *JAMA Ophthalmol.* **135** 795–800

Lagali N S *et al* 2018 Wide-field corneal subbasal nerve plexus mosaics in age-controlled healthy and type 2 diabetes populations. *Sci. Data* **5** 1–12

Li Q *et al* 2019 Quantitative analysis of corneal nerve fibers in type 2 diabetics with and without diabetic peripheral neuropathy: comparison of manual and automated assessments *Diabetes Res. Clin. Pract.* **151** 33–8

Mehra S, Tavakoli M, Kallinikos P A, Efron N, Boulton A J M, Augustine T and Malik R A 2007 Corneal confocal microscopy detects early nerve regenerationafter pancreas transplantation in patients with type 1 diabetes *Diabetes Care* **30** 2608–12

Misra S L, Kersten H M, Roxburgh R H, Danesh-Meyer H V and McGhee C N 2017 Corneal nerve microstructure in Parkinson's disease *J. Clin. Neurosci.* **39** 53–8

Mou L *et al* 2021 CS2-Net: deep learning segmentation of curvilinear structures in medical imaging *Med. Image Anal.* **67** 101874

Mou L, Zhao Y, Chen L, Cheng J, Gu Z, Hao H, Qi H, Zheng Y, Frandi A and Liu J 2019 CS-Net: channel and spatial attention network for curvilinear structure segmentation *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention* (Cham: Springer) pp 721–30

Oktay O *et al* 2018 Attention u-net: learning where to look for the pancreas arXiv.1804.03999

Petropoulos I N, Kamran S, Li Y, Khan A, Ponirakis G, Akhtar N, Deleu D, Shuaib A and Malik R A 2017 Corneal confocal microscopy: an imaging endpoint for axonal degeneration in multiple sclerosis *Investigative Ophthalmol. Vis. Sci.* **58** 3677–81

Petropoulos I N, Ponirakis G, Khan A, Gad H, Almuhannadi H, Brines M and Malik R A 2020 Corneal confocal microscopy: ready for prime time *Clin. Exp. Optom.* **103** 265–77

Poletti E and Ruggeri A 2013 Automatic nerve tracking in confocal images of corneal subbasal epithelium *Proc. of the 26th IEEE Int. Sym. on Computer-Based Medical Systems* (IEEE) 119–24

Ponirakis G *et al* 2019 Association of corneal nerve fiber measures with cognitive function in dementia *Ann. Clin. Trans. Neurol.* **6** 689–97

Qin X, Zhang Z, Huang C, Dehghan M, Zaiane O R and Jagersand M 2020 U2-Net: going deeper with nested U-structure for salient object detection *Pattern Recognit.* **106** 107404

Ronneberger O, Fischer P and Brox T 2015 U-net: convolutional networks for biomedical image segmentation *MICCAI* (Cham : Springer) pp 234–41 Int. Conf. on Medical Image Computing and Computer-assisted Intervention

Sahu M, Mukhopadhyay A and Zachow S 2021 Simulation-to-real domain adaptation with teacher–student learning for endoscopic instrument segmentation *Int. J. Comput. Assis. Radiol. Surg.* **16** 849–59

Scarpa F, Grisan E and Ruggeri A 2008 Automatic recognition of corneal nerve structures in images from confocal microscopy *Investigative Opthalmol. Vis. Sci.* **49** 4801–7

Scarpa F, Zheng X, Ohashi Y and Ruggeri A 2011 Automatic evaluation of corneal nerve tortuosity in images from *in vivo* confocal microscopy *Investigative Ophthalmol. Vis. Sci.* **52** 6404–8

Shi W, Caballero J, Huszár F, Totz J, Aitken A P, Bishop R, Rueckert D and Wang Z 2016 Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition* pp 1874–83

Shin S Y, Lee S, Yun I D and Lee K M 2019 Deep vessel segmentation by learning graphical connectivity *Med. Image Anal.* **58** 101556

Sun Q, Dai M, Lan Z, Cai F, Wei L, Yang C and Chen R 2022 UCR-Net: U-shaped context residual network for medical image segmentation *Comput. Biol. Med.* **151** 106203

Tavakoli M, Quattrini C, Abbott C, Kallinikos P, Marshall A, Finnigan J, Morgan P, Efron N, Boulton A J M and Malik R A 2010a Corneal confocal microscopy: a novel noninvasive test to diagnose and stratify the severity of human diabetic neuropathy *Diabetes Care* **33** 1792–7

Tavakoli M, Quattrini C, Abbott C, Kallinikos P, Marshall A, Finnigan J, Morgan P, Efron N, Boulton A J M and Malik R A M 2010b Corneal confocal microscopy *Diabetes Care* **33** 1792–7

Testa V *et al* 2020 Neuroaxonal degeneration in patients with multiple sclerosis: an optical coherence tomography and *in vivo* corneal confocal microscopy study *Cornea* **39** 1221–6

Wang W, Zhong J, Wu H, Wen Z and Qin J 2020 Rvseg-net: an efficient feature pyramid cascade network for retinal vessel segmentation *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention* (Cham: Springer)

Wang X, Girshick R, Gupta A and He K 2018 Non-local neural networks *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition* pp 7794–803

Williams B M *et al* 2020 An artificial intelligence-based deep learning algorithm for the diagnosis of diabetic neuropathy using corneal confocal microscopy: a development and validation study *Diabetologia* **63** 419–30

Wu H, Wang W, Zhong J, Lei B, Wen Z and Qin J 2021 Scs-net: a scale and context sensitive network for retinal vessel segmentation *Med. Image Anal.* **70** 102025

Xu R, Liu T, Ye X, Liu F, Lin L, Li L, Tanaka S and Chen Y W 2020 Joint extraction of retinal vessels and centerlines based on deep semantics and multi-scaled cross-task aggregation *IEEE J. Biomed. Health Inform.* **25** 2722–32

Yang C, Zhou X, Zhu W, Xiang D, Chen Z, Yuan J, Chen X and Shi F 2021 Multi-discriminator adversarial convolutional network for nerve fiber segmentation in confocal corneal microscopy images *IEEE J. Biomed. Health Inform.* **26** 648–59

Yuan Y, Huang L, Guo J, Zhang C, Chen X and Wang J 2021 OCNet: object context for semantic segmentation *Int. J. Comput. Vision* **129** 2375–98

Zhang D, Huang F, Khansari M, Berendschot T T, Xu X, Dashtbozorg B, Sun Y, Zhang J and Tan T 2020 Automatic corneal nerve fiber segmentation and geometric biomarker quantification *Eur. Phys. J. Plus* **135** 266

Zhang Z, Zhang X, Peng C, Xue X and Sun J 2018 Exfuse: enhancing feature fusion for semantic segmentation *Conference on Computer Vision (ECCV)* pp 269–84 Proc. of the European Conf. on Computer Vision (ECCV)

Zhong X, Zhang H, Li G and Ji D 2022 Do you need sharpened details? Asking MMDC-Net: multi-layer multi-scale dilated convolution network for retinal vessel segmentation *Comput. Biol. Med.* **150** 106198

Zhou Z, Siddiquee M M R, Tajbakhsh N and Liang J 2018 Unet++: a nested u-net architecture for medical image segmentation *DLMIA ML-CDS 2018* (*Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*) vol 11045 (Cham: Springer) pp 3–11