

# 基于多尺度残差卷积神经网络的视杯视盘联合分割



袁鑫<sup>1</sup>, 郑秀娟<sup>1</sup>, 吉彬<sup>1,2</sup>, 李淼<sup>1</sup>, 李彬<sup>1</sup>

1. 四川大学 电气工程学院 自动化系 (成都 610065)
2. 中国移动 (成都) 产业研究院 (成都 610041)

**【摘要】** 青光眼是不可逆性失明的首要原因, 早期症状不明显, 容易被忽视, 因此青光眼早期筛查尤为重要。杯盘比是临床上用于青光眼筛查的重要指标, 所以精准分割视杯视盘是计算杯盘比的关键。本文提出了一种基于全卷积分多尺度残差神经网络的视杯视盘分割方法。首先, 对眼底图像进行对比度增强, 并引入极坐标变换。随后, 构造 W-Net 作为主体网络, 用带残差多尺度全卷积模块来替代标准卷积单元, 输入端口加入图像金字塔来构造多尺度输入, 侧输出层作为早期的分类器生成局部预测输出。最后, 提出了一种新的多标签损失函数来指导网络分割。实验采用 REFUGE 数据集验证, 最终视杯、视盘分割的平均交并比分别为 0.904 0、0.955 3, 重叠误差分别为 0.178 0、0.066 5。结果表明, 该方法不仅实现了视杯视盘联合分割, 而且有效提高了其分割精度。该方法将有助于大规模青光眼早期筛查的推广。

**【关键词】** 深度学习; 全卷积神经网络; 视盘分割; 视杯分割; 青光眼筛查

## Joint optic disc and cup segmentation based on residual multi-scale fully convolutional neural network

YUAN Xin<sup>1</sup>, ZHENG Xiujuan<sup>1</sup>, JI Bin<sup>1,2</sup>, LI Miao<sup>1</sup>, LI Bin<sup>1</sup>

1. Department of Automation, College of Electrical Engineering, Sichuan University, Chengdu 610065, P.R.China
2. China Mobile (Chengdu) Industrial Research Institute, Chengdu 610041, P.R.China

Corresponding author: ZHENG Xiujuan, Email: xiujuanzheng@scu.edu.cn

**【Abstract】** Glaucoma is the leading cause of irreversible blindness, but its early symptoms are not obvious and are easily overlooked, so early screening for glaucoma is particularly important. The cup to disc ratio is an important indicator for clinical glaucoma screening, and accurate segmentation of the optic cup and disc is the key to calculating the cup to disc ratio. In this paper, a full convolutional neural network with residual multi-scale convolution module was proposed for the optic cup and disc segmentation. First, the fundus image was contrast enhanced and polar transformation was introduced. Subsequently, W-Net was used as the backbone network, which replaced the standard convolution unit with the residual multi-scale full convolution module, the input port was added to the image pyramid to construct the multi-scale input, and the side output layer was used as the early classifier to generate the local prediction output. Finally, a new multi-tag loss function was proposed to guide network segmentation. The mean intersection over union of the optic cup and disc segmentation in the REFUGE dataset was 0.904 0 and 0.955 3 respectively, and the overlapping error was 0.178 0 and 0.066 5 respectively. The results show that this method not only realizes the joint segmentation of cup and disc, but also improves the segmentation accuracy effectively, which could be helpful for the promotion of large-scale early glaucoma screening.

**【Key words】** deep learning; fully convolutional neural network; optic disc segmentation; optic cup segmentation; glaucoma screening

DOI: 10.7507/1001-5515.201909006

基金项目: 国家自然科学基金项目 (81201146); 四川科技计划项目 (2019YFS0140); 成都市技术创新研发项目 (2020-YF05-01386-SN)

通信作者: 郑秀娟, Email: xiujuanzheng@scu.edu.cn

## 引言

青光眼是全世界第二大致盲眼病(仅次于白内障),也是造成不可逆性失明的首要原因<sup>[1]</sup>。青光眼造成的视力损伤是不可逆的,且早期症状不易发现,因此对青光眼的早期筛查与诊断至关重要。目前,眼底图像和三维光学相干层析成像(optical coherence tomography, OCT)常被用于辅助诊断青光眼。其中OCT图像成本相对昂贵且普及率低,不适用于大规模青光眼筛查,故大多数医生常选择使用成本较低和容易获得的眼底图像进行青光眼筛查与诊断。当下,评估眼底图像中视神经乳头的技术是主流的青光眼筛查技术<sup>[2]</sup>,它采用一个二分类来判断是否属于青光眼疾病。然而眼底图像常由经验丰富的眼科医生手动标注,费时费力且带有很大的主观性,因此手工标注眼底图像不适用于青光眼的大规模筛查。

在眼底彩色图像中,视盘呈现亮黄色且形状近似椭圆,可分为两个明显的区域:中间明亮区(视杯)和外围区(视神经网膜边缘)。青光眼大规模筛查技术中,多种自动评估视神经乳头的的方法被相继提出,例如垂直杯盘比(vertical cup to disc ratio, CDR)<sup>[3]</sup>、视杯(optic cup, OC)视盘(optic disc, OD)面积比以及视盘直径<sup>[4]</sup>等。而在临床上,医生主要采用杯盘比评估视神经头。杯盘比是指垂直杯径与垂直盘径的比。通常情况下,杯盘比值越大,则患青光眼的概率越大。因此,准确地分割视杯视盘是筛查和诊断青光眼的关键。

在医学图像分割算法中,主要分为基于手工提取特征的传统方法<sup>[5-8]</sup>和利用卷积神经网络来自动提取特征的深度学习框架<sup>[9-11]</sup>。早期研究中的视杯视盘分割方法是利用手工提取视觉特征进行分割,其主要包括颜色、纹理、对比度阈值、边缘检测、分割模型和区域分割方法<sup>[12-19]</sup>,但这些方法容易受到眼底图像拍摄环境和图像本身质量的影响,从而影响目标的分割效果。此外,从众多像素中提取出特征训练分类器不易实现,故有学者<sup>[20]</sup>提出采用超像素策略来减少像素数,并采用超像素分类进行视杯视盘分割。但该方法需要手工构造特征来获得分割结果,其实现过程繁琐且可重复性差。深度学习在计算机视觉任务中克服了人工设计特征的局限性,并可自动学习高度的可区分性特征进行表示。在医学图像分割领域,早期的深度学习方法大多是基于图像块<sup>[21-22]</sup>,其局限性是滑动窗会导致冗余计算和无法学习到图像的全局特征。接着,端

到端的全卷积神经网络(fully convolutional neural network, FCN)被提出并在图像分类和分割中得到广泛应用<sup>[23-24]</sup>。在全卷积神经网络基础上,Ronneberger等<sup>[25]</sup>提出在医学图像分割领域具有卓越性能的U-Net结构,并且该结构已经成为该领域的重要结构。通常,U-Net<sup>[25]</sup>结构可以被认为是编、解码器的结构。编码器的目的是逐步减少特征映射的空间维数,来获取更高层次的语义特征。而解码器则是为了恢复目标中的空间维数和结构细节。因此,在编码器中应尽可能捕获更多的高级特征,而在解码器中应尽可能地保留空间信息。之后,在U-Net基础上又发展出许多改进网络结构,包括M-Net<sup>[26-27]</sup>、U-Net++<sup>[28-29]</sup>等,均在医学图像分割任务中取得了不错的效果。

现有眼底图像分析方法大多是对视杯视盘分别进行分割<sup>[12-15,30]</sup>。文献<sup>[12-13]</sup>将视杯视盘分割分为两个独立阶段,即分别利用各自的特征进行分割。在分割视杯时,先定位到视盘,把视盘分割出来后再提取感兴趣区域来分割视杯,造成了视盘分割信息的浪费。文献<sup>[14]</sup>则提出将视杯视盘分割集成在同一分割框架内。但 these 方法忽略了视杯和视盘之间存在的先验信息,例如形状约束和结构约束,即视杯视盘均接近椭圆形,视杯包含于视盘之中。同时,上述几种方法都将视杯视盘看作非独立标签,每个像素只对应一个标签(即视杯、视盘或背景),割裂了两者的联系。

针对上述问题,本文以U-Net<sup>[25]</sup>与M-Net<sup>[26]</sup>为基础,提出了一个新的W-Net网络结构用于视杯视盘联合分割。相比于U-Net<sup>[25]</sup>或M-Net<sup>[26]</sup>,W-Net主网络结构是以W型卷积神经网络为主体框架,加深了网络深度,有利于网络进行深层信息学习。此外,W-Net引入了多尺度输入和不同网络深度的监督,其中多尺度输入可构造图像金字塔,使输入图像信息更加丰富;另外在网络中加入不同的深度监督,可使侧输出层计算梯度能更容易地反向传播到前面的卷积层中,可有效抑制梯度消失问题和有利于网络的早期训练。同时,本文还提出了一种残差多尺度全卷积学习模块来代替U-Net和M-Net中的标准卷积学习单元,加深了网络宽度,从而能在有限的尺度范围内捕获更多尺度特征。最后,提出了一种新的多标签损失函数来指导网络进行视杯视盘分割。

## 1 方法

### 1.1 数据集

本文选用 REFUGE 数据集 (<https://refuge.grand-challenge.org/Home/>) 进行方法验证。该数据库是目前青光眼眼底照片精标数据库中最全面的标注数据库, 主要包括青光眼与非青光眼两种类型数据, 其中青光眼和非青光眼图像的比例分别为 10% 和 90%。每张眼底图像分别包含诊断、图像分割及定位三方面信息, 由七位专家人工标记并融合, 克服了之前许多青光眼公开数据集存在的只有诊断标签信息, 无视杯、视盘等关键结构的标注信息, 且参与标注的专家较少等缺点。所有图像均以后极 (posterior pole) 为中心, 同时含有黄斑和视盘。在这个数据集中, 由训练集、验证集和测试集三个组成。训练集中有 400 张像素为  $2142 \times 2056$  的眼底图像, 是使用 Zeiss Visucam 500 眼底相机拍摄的, 而验证集和测试集各由 400 张像素均为  $1634 \times 1634$  的眼底图像组成, 是使用 Canon CR-2 眼底相机拍摄的。由于各个集合被不同的眼底照相机拍摄, 故而图像在颜色、纹理等光学性质上会呈现不同, 如图 1 所示。

## 1.2 图像预处理

**1.2.1 极坐标变换** 对于图像的预处理操作, 如图 2 所示。在该方法中, 最初得到的是一张包含了感兴趣区域的眼底图像全景图, 首先使用现有的视盘自动检测方法对视盘的中心进行定位<sup>[31]</sup>, 按照指定的像素大小对视盘区域进行分割, 以得到感兴趣区域。然后, 将得到的感兴趣区域进行极坐标变换, 像素级坐标变换将原始的眼底图像映射到极坐标系统, 其转换关系如式 (1) 所示:

$$\begin{cases} r = \sqrt{u^2 + v^2} \\ \theta = \tan^{-1}v/u \end{cases} \quad (1)$$

其中  $u, v$  表示某个像素在笛卡尔坐标系下的纵横坐标;  $\theta, r$  表示某个像素在极坐标下的方向角和半径。

极坐标变换的作用主要有: ① 空间约束。原始的眼底图像中, 存在一个有效的几何约束是视杯包含于视盘之中, 且两者形状是近似椭圆形, 这些信息的表达和约束在网络中的实现比较困难, 且各个图像的差异性 (大小和位置等) 较大。但在极坐标系统下, 可以容易地将这种几何约束转换成空间关系, 即视杯、视盘和背景呈现出有序的层结构。正因为有这种层结构, 本文便将其作为先验空间约束。具体来说, 在视杯分割结果中, 预测视杯像素不可能存在于预测标注图像的下方 1/2 处, 可直接将其设置为背景部分; 同理, 在视盘分割的预测标注图像的下方 1/3 处可以直接将其设置为背景部分。② 平衡杯盘比。在原始的眼底图像中, 视杯

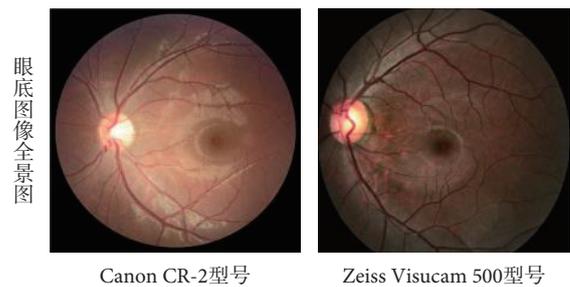


图 1 不同眼底照相机拍摄的眼底图像  
Fig.1 Fundus images taken by different fundus cameras

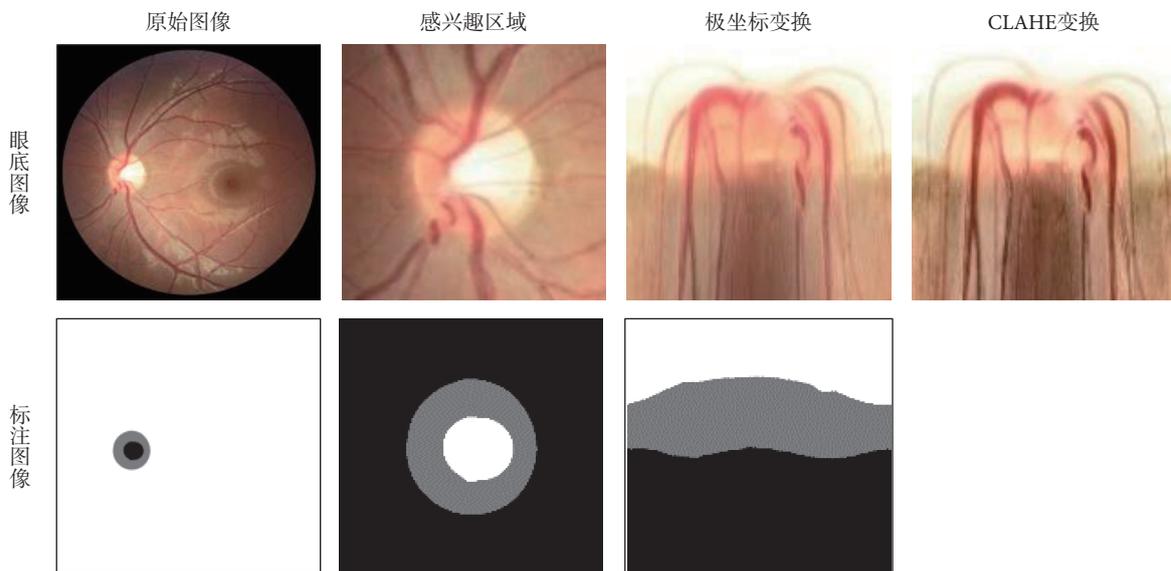


图 2 眼底图像和标注图像的预处理过程  
Fig.2 The preprocessing process of fundus image and annotated image

视盘与背景的像素分布有很大的偏差。即使截取到的感兴趣区域中，视杯区域仅占 4% 左右，这种极不平衡的结构区域很容易造成神经网络在训练过程中的偏差与过拟合，大大影响视盘、视杯分割的精度。极坐标变换是基于视盘中心的图像平坦化，可以通过插值来扩大视杯在图像中的比例。经过极坐标变换的视杯区域比率比感兴趣区域部分高 23.4%<sup>[26]</sup>，能有效提高杯盘比，平衡数据集，防止过拟合和提高分割精度。

**1.2.2 数据增强** 由于数据集中眼底图像数量有限，在得到感兴趣区域并进行极坐标变换之后，使用数据集增强来防止模型过拟合。随机图像处理方法可以提高数据增强的能力，对图像进行随机变换，包括随机水平翻转、图形随机移动和 HSV (Hue, Saturation, Value) 颜色空间中的颜色抖动。

**1.2.3 归一化处理** 通常眼底图像存在光照不均匀、对比度低和血管干扰等问题，这些因素都会降低青光眼诊断的有效性。在视杯视盘分割任务中，先将视杯视盘和背景的对对比度增大，再对增强后的图像进行限制对比度自适应直方图均衡化处理 (contrast limited adaptive histogram equalization, CLAHE)<sup>[32]</sup>和颜色归一化以纠正不均匀的光照和改善低对比度，最后将处理后的图像与处理前的图像组合起来，构成一个新的六通道图像数据，以此来丰富图像信息，让模型学习到具有更强可分性的特征。

### 1.3 全卷积多尺度神经网络模型的建立

本文所提出的 W-Net 是一个端到端的多标签深度神经网络，由四个主要部分组成，如图 3 所示。第一部分是多尺度层，用于构造图像金字塔的输入，实现不同尺度和光学性质的原始语义信息输入，极大地丰富语义信息。第二部分是采用 W 型的卷积神经网络作为主体结构，用于学习图像中所包含的不同层次的深浅特征。第三部分是在不同网络深度层次的多个侧输出层来实现深层监控，让梯度信息更好反向回传到前层，有效抑制神经网络的梯度消失和有利于网络的训练学习。此外，还提出了带残差多尺度全卷积模块来代替普通卷积模块，实现在不同尺度感受野下提取更多可分性强的特征，增加网络的特征提取能力。最后，提出了一个多标签损失函数来指导视盘、视杯的联合分割实现。

**1.3.1 主网络结构** 借鉴 U-Net<sup>[25]</sup>、M-Net<sup>[26]</sup>，本文将两个 U-Net 进行级联作为网络的主体，如图 3 所示。U-Net 是一种高效的生物医学图像分割的全卷积神经网络，由编码路径和解码路径两部分组成，但是对于较复杂的多标签任务，往往 U-Net 的分割精度不高，本文主网络是将两个 U-Net 进行级联，前 U 型网络为后 U 型网络提供了较为浅层的语义表达，后 U 型网络对前 U 型网络提供的浅层信息进行进一步语义抽象与增强。这种结构加深了神经的深度，从而使得网络能在训练过程中学习到更

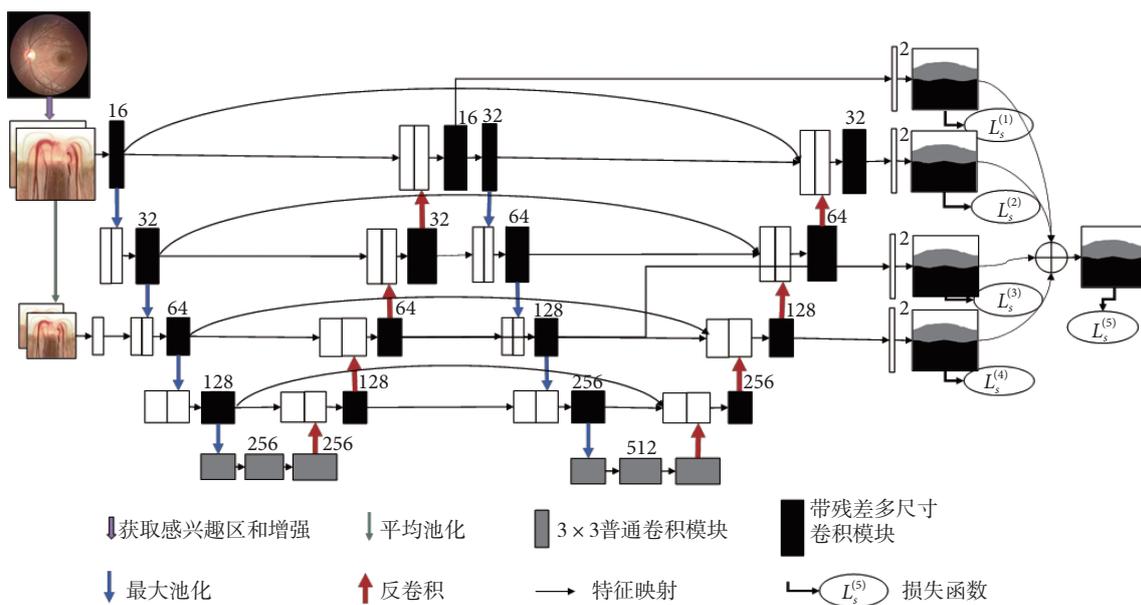


图 3 W-Net 主网络结构图 包括一个多尺度输入层、一个 W 形卷积网络、四个侧输出和多标签损失函数

Fig.3 Main network structure including a multi-scale output layer, a W-shaped convolutional network, four side outputs, and a multi-label loss function

多深层结构信息和更加抽象的语义信息；再加之网络中的跳跃连接，可以实现网络中同一尺度深、浅不同层次之间的语义信息融合和减少不同语义之间的语义鸿沟，来指导网络提取深层的可分性语义表示。此外，浅层特征可最大程度保留图像原始的结构信息，而深层特征则包含更多抽象结构信息，两者进行融合有利于在最后的卷积通道中恢复目标中的空间维数和结构细节，进而提高分割精度。

每个编码路径与卷积核一起执行卷积层以产生一组编码器特征映射，并且利用修正线性单元 (rectified linear unit, ReLU) 激活函数，编码器路径还利用卷积层来输出解码器特征图。最后，将最终的解码器层输出的高维特征表示输入到可训练的多标签分类卷积层中，最终分类器采用 Sigmoid 激活函数来生成预测概率图，实现像素级分类。本文将视杯视盘多标签分割任务输出为三个通道 (分别为视杯、视盘和背景) 的概率预测图，概率映射的预测结果即为在每个像素处具有最大概率的类别。

**1.3.2 多尺度输入层** 文献[26]对多尺度输入或图像金字塔已经进行验证，可有效提高网络分割性能。经过图像的预处理之后，将 RGB 图像拓展到了六个通道。本文使用平均池化对图像进行降采样，并在编码路径中构造多尺度输入。这样可以将多尺度输入集成到编码层之中，有效避免参数大量增长，以及等效增加编码路径的网络宽度。与其他工作<sup>[26]</sup>不同，我们发现当图像金字塔的深度与网络深度一致时，网络学习性能并非能达到最佳。因为在持续的降采样过程中，虽然会丢失一些图像信息，同时得到一些新的结构信息，但这之中存在大量冗余信息，会导致网络存在过拟合和泛化性能差等问题。因此在本文中进行了一次降采样跨网络深度的处理，既增加了网络深度，又能较好地抑制网络中的输入信息冗余，有效实现两者的平衡。

**1.3.3 侧输出层** 侧输出层充当一个早期分类器，为早期层生成一个局部伴随输出映射。M-Net<sup>[26]</sup>中有四个侧输出层，并将四个侧输出层保持权重一致、简单平均叠加，得到最后的输出。本文发现，网络的几个侧输出层对最后的输出的贡献并非相等。侧输出层选择不同深度网络的一、三通道，与前面的多尺度输出层结构保持一致，这样既保证网络输出之间冗余信息的有效减少，又可对网络进行有效深度监督。最后，将不同的侧输出层进行加权融合，给出下式的目标函数：

$$\sigma_s = \sum_{n=0}^N \alpha_n L_s^{(n)} \quad (2)$$

其中， $\alpha_n$  表示每个侧输出层的损失函数的融合权重， $L_s^{(n)}$  表示第  $n$  个侧输出层的多标签损失， $N$  表示侧输出层的个数。在本文中，设置最后一层的损失函数权重为 0.6，其余权重均为 0.1。

侧输出层的主要优点是可以将边输出的损失函数反向传播到对应的浅层卷积层中，可有效抑制梯度消失问题，有利于实现早期层的训练。此外，侧输出层的多尺度融合实现了特征高性能的融合，侧输出层对各尺度的输出预测结果实现深度监督，从而获得更好的输出结果。

**1.3.4 多标签损失函数** 视杯、视盘联合分割任务是一个多标签问题。现有的分割方法通常属于多类设置，它将每个实例按照某种规则对其进行唯一标记。而在多标签方法中，每一类都是一个独立的二进制分类，某一实例按照某种规则对每一类进行判断，这意味着某一实例可分别属于多个类别。这一点在视杯、视盘分割任务中十分重要，因为视杯、视盘有严格的几何约束，视盘区域包含视杯区域，表明属于视杯的像素也一定属于视盘。在病变的青光眼之中，视盘区域中排除掉视杯区域的部分仅仅只是一个很薄的圆环，这就会造成视盘与背景区域像素面积不平衡，而采用多标签的方法可有效提高视盘与背景的比例，可以解决类别极不平衡的问题。在该方法中，我们提出一个基于 DICE<sup>[33]</sup> 和 Focal-Loss<sup>[34]</sup> 的多标签损失函数，见公式 (3) ~ (5)。DICE 系数是一个重叠的度量，广泛用于评估分割性能。DICE 损失函数指示前景掩码重叠比，并且可以处理前景像素与背景的不平衡问题。Focal-Loss 则主要为了解决数据集不平衡问题，因为它更关注于小样本中难以分割的区域。

$$L_D = - \sum_k^K \frac{2 \sum_i^N p_i g_i + \epsilon}{\sum_i^N p_i^2 + \sum_i^N g_i^2 + \epsilon} \quad (3)$$

$$L_F = \sum_k^K [-\alpha g_i (1 - p_i)^\gamma \log(p_i) - \alpha (1 - g_i) p_i^\gamma \log(1 - p_i)] \quad (4)$$

$$L_s = L_D + L_F \quad (5)$$

其中， $g_i$  和  $p_i$  分别表示第  $i$  个像素的实际类别和真实类别； $\epsilon$ 、 $\alpha$ 、 $\gamma$  为常值系数，设置  $\epsilon = 0.9$ ， $\alpha = 0.5$ ， $\gamma = 2$ 。

**1.3.5 残差多尺度卷积模块** 在深度学习网络中，Inception<sup>[35]</sup> 和 ResNet<sup>[36]</sup> 是两个经典的框架。

Inception 结构通过拓宽网络宽度来接收不同尺度的感受野。而 ResNet 则引入了跳跃连接机制来克服网络退化和梯度消失的问题,这使得神经网络深度首次突破 1 000 层以上的有效学习。Inception-ResNet<sup>[37]</sup>模块结合了 Inception 和 ResNet,继承了这两种方法的优点,使之成为深度卷积神经网络中一种基本的方法。在 Inception-ResNet-V4 模块的启发之下,我们提出了一种残差多尺度卷积模块来提取高层次的语义特征映射,如图 4 所示。在通常的情况下,具有大感受野的卷积核可以提取大目标和产生丰富的抽象特征,而小的感受野卷积核则更适合提取小目标和细节的特征。在残差多尺度卷积模块中,使用两个 3×3 的卷积单元代替一个 5×5 的卷积模块,使用 1×7 和 7×1 两个模块来代替 7×7 的模块,且每个通道采用原来 1/4 的卷积核。这些改进不仅减少了网络参数,同时能够保持较大的感受野。除了通过四条卷积通道进行特征映射,还引入了类似 ResNet 中的跳跃连接。然后再经过一个标准 3×3 卷积单元和 1×1 的卷积单元来进行降维和非线性激励,以提升网络的表达能力。最后对所得到的卷积单元再与输入特征进行跳跃连接,实现不同深、浅层次特征融合。

1.4 评价指标

为了评价所提出 W-Net 网络的分割性能,引入了一系列评价指标来分析所得视杯、视盘和背景区域的预测分割结果,这些指标一般都用来评价视杯视盘分割的准确性<sup>[38]</sup>,包括了精确度 (pixels accuracy, PA)、平均精确度 (mean accuracy, MA)、平均交并比 (mean intersection over union, MIoU)、加权交并比 (frequency weighted intersection over union, FWIoU)、以及重叠误差 (overlapping error, OE)。具体表达式见公式 (6) ~ (10)。

$$PA = \frac{\sum_i n_{ii}}{t} \tag{6}$$

$$MA = \frac{1}{n_{cl}} \sum_i \frac{n_{ii}}{t_i} \tag{7}$$

$$MIoU = \frac{1}{n_{cl}} \sum_i \frac{n_{ii}}{t_i + \sum_j n_{ji} - n_{ii}} \tag{8}$$

$$FWIoU = \frac{1}{t} \sum_i \frac{t_i n_{ii}}{t_i + \sum_j n_{ji} - n_{ii}} \tag{9}$$

$$OE = 1 - \sum_i \frac{n_{ii}}{t_i + \sum_j n_{ji} - n_{ii}} \tag{10}$$

其中,  $n_{cl}$  表示类别数;  $t = \sum_i t_i$  表示所有像素的数目;  $t_i = \sum_j n_{ij}$  表示属于第  $i$  类的像素总数。 $n_{ji}$  表示实际类别为第  $i$  类却错分为第  $j$  类的像素数目。

2 结果

2.1 实验环境

在实验中,主要步骤:① 图像预处理;② 利用网络训练来学习系统最优参数,并在每个训练周期结束后的验证阶段保存最佳网络模型;③ 测试最佳网络模型并预测结果;④ 预测结果评估。本文实验是基于 Keras 的 Python 和 Tensorflow 后端实现的,采用的显存为 8 GB 的 Nvidia GeForce GTX 1 080 GPU 和主频 3.50 GHz 的 Inter(R)i7-5 930k CPU。在训练过程采用 Adam 优化器对深度模型进行优化。网络基于图像数据集,采用数据集作为参考进行有监督训练,其中 75% 即 600 张眼底图像用于训练,12.5% 即 100 张图像用于验证,12.5% 即 100 张图像用于测试。训练中每一次选取的样本数设置为 6,训练集每训练一个周期结束,对验证集进行一次整体评测,训练过程持续 200 个周期。实验测试结果与专家标注结果作比较,采用测试图片定量评价的平均值作为定量分析的结果。

2.2 实验结果

本研究在 REFUGE 眼底图像数据库上进行实

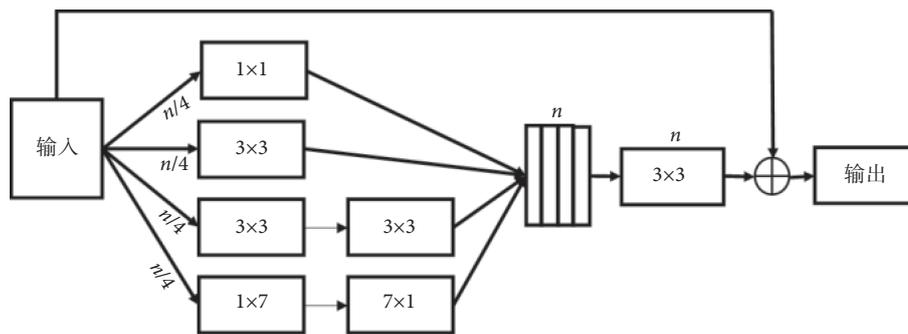


图 4 残差多尺度卷积模块结构图

Fig.4 Structure of residual multi-scale convolution module

验并与当前最先进的方法进行了对比,包括 U-Net<sup>[25]</sup>、M-Net<sup>[26]</sup>以及基于它们的一些改进模型。为了验证所提出的残差多尺度卷积单元相较于标准卷积模块的优越性能,使用 U-Net<sup>[25]</sup>、M-Net<sup>[26]</sup>作为主体框架,将其中的普通卷积单元替换为残差多尺度卷积单元,分别称之为 U-Net-Mcon 和 M-Net-Mcon,再与本文所提出的 W-Net 为主体框架的模型分别以普通卷积单元(W-Net)和残差多尺度卷积单元(W-Net-Mcon)进行比较。通过定量分析 ACC、MA、MIoU、FWIoU 以及 OE 五个评价指标,分别评估提出的 W-Net-Mcon 在视杯、视盘的分割性能。

表1 显示了在 REFUGE 数据集<sup>[31]</sup>中使用包括 U-Net<sup>[25]</sup>、M-Net<sup>[26]</sup>在内的六种网络方法进行视杯、视盘联合分割所得到的性能指标。不难看出,本文提出的 W-Net-Mcon 网络的视杯、视盘分割结果在上述五个评价指标均优于其他网络。在实验结果中,视盘分割结果的 ACC、MA、MIoU、FWIoU、OE 分别为 0.983 3、0.978 6、0.955 3、0.967 4 和 0.066 5,视杯分割结果分别为 0.987 0、0.949 6、0.904 0、0.975 3 和 0.178 0。

图5 显示了网络优化过程训练集和验证集的精度和损失函数变化情况。可以看出,W-Net-Mcon 在精度上明显高于其他网络,同时损失函数能够快速收敛且取得最优。

图 6 显示了在 REFUGE 数据集得到的视杯、视盘分割实验结果实例。为了证明所提方法的泛化性能,选取了数据集中两种不同光学性质的眼底图像,分别为图 6 中的 1、2、3 组和 4、5、6 组。不难看出,该方法在处理对比度较高或较低的两类不同眼底图像时,均可以实现视杯视盘的高精度联合分割。

### 3 讨论

本研究利用深度学习中的全卷积神经网络对

眼底图像中的视杯视盘进行联合分割。为了解决传统图像处理算法无法实现低对比度图像和不同光学图像之间低泛化性的视杯视盘联合分割问题,本研究利用全卷积神经网络强大的特征提取和特征表达能力,提出一种新型网络结构和带残差多尺度卷积学习模块的网络(W-Net-Mcon),实现了高维与低维特征的融合,提高了特征的使用率。

从图 5 可以看出,与其他网络相比,本文构建的网络在不同方面均表现出了一定的优势。在精度上,本文构建的网络高于其他网络,能够使损失函数快速收敛且取得最优。这表明本文方法相较于其他网络,在训练集上有较强的拟合能力和特征提取能力。而在验证集上的高精度则表明网络没有出现因对训练集的过度学习而引发的过拟合,该方法具有较强的泛化性和鲁棒性,能够学习图像语义信息中的强可分性特征表示,从而使得网络具有更好的分割性能。这些优势也体现在表 1 的结果中。

表 1 罗列的是评价指标的定量分析。首先,在各个网络框架的分割效果中的优势:实验中选用了 U-Net、M-Net 和 W-Net 进行比较,W-Net 各项指标均高于另外两组数据,尤其是在视盘、视杯的 OE 指标表现优异,这表明本文提出的 W-Net 网络结构在某些方面优于其他网络。其优势主要由于:① W-Net 的多尺度输入层相较于 U-Net 加深了网络宽度以提供丰富的原始语义信息,同时相较于 M-Net 又减少了原始语义信息中的重复、冗余信息,有效实现两者平衡;② W-Net 中不同网络深度的侧输出层实现了网络各层的深度监督,从而指导网络学习特征信息;③ 引入跳跃连接,将同一尺度不同深、浅层次的特征信息融合,有利于减少网络参数和网络梯度反向传播,可避免梯度消失问题。

其次,在所提出的残差多尺度卷积模块的优势:实验将 U-Net、U-Net-Mcon, M-Net、M-Net-

表 1 六种网络对于视杯、视盘联合分割的性能评价指标比较

Tab.1 Comparison of performance evaluation indicators of the six types of networks for joint segmentation of cup and disc

网络	视盘					视杯				
	OE	ACC	MA	MIoU	FWIoU	OE	ACC	MA	MIoU	FWIoU
U-Net <sup>[25]</sup>	0.087 7	0.977 2	0.971 6	0.940 9	0.956 9	0.236 4	0.982 8	0.909 8	0.872 6	0.966 9
M-Net <sup>[26]</sup>	0.081 9	0.981 1	0.974 5	0.942 8	0.960 2	0.195 9	0.985 4	0.938 4	0.897 0	0.973 2
U-Net-Mcon	0.078 1	0.979 3	0.971 3	0.947 1	0.960 7	0.217 5	0.984 0	0.926 8	0.882 7	0.969 7
M-Net-Mcon	0.077 2	0.979 8	0.975 2	0.948 1	0.961 0	0.192 2	0.985 8	0.938 7	0.898 1	0.973 5
W-Net	0.080 5	0.981 9	0.974 8	0.945 8	0.960 8	0.190 7	0.986 1	0.940 2	0.899 3	0.973 8
W-Net-Mcon	<b>0.066 5</b>	<b>0.983 3</b>	<b>0.978 6</b>	<b>0.955 3</b>	<b>0.967 4</b>	<b>0.178 0</b>	<b>0.987 0</b>	<b>0.949 6</b>	<b>0.904 0</b>	<b>0.975 3</b>

注:最佳结果用加粗字体标出

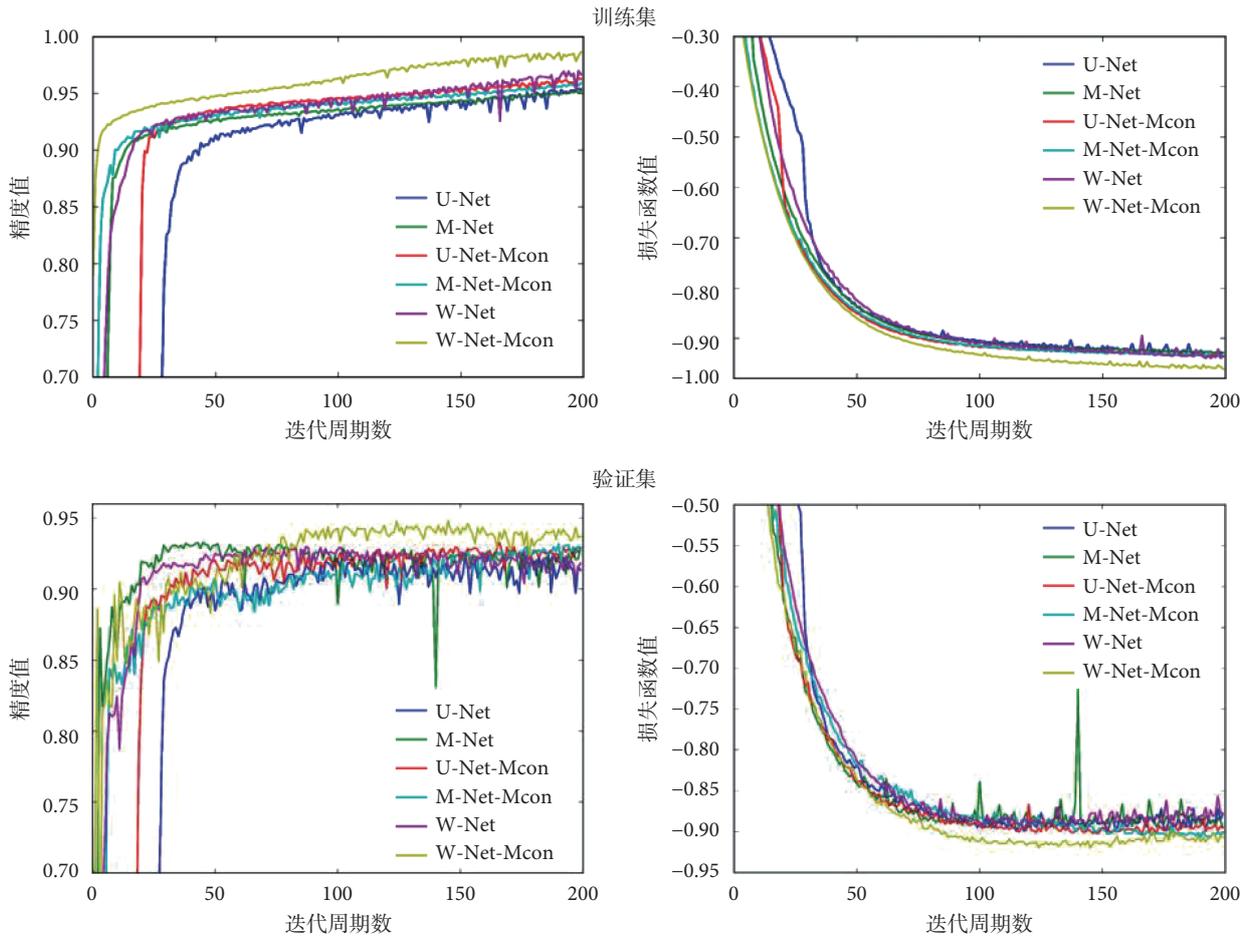


图5 训练集和验证集的精度与损失函数值在网络训练过程中随周期迭代的变化情况

Fig.5 The accuracy and loss function values of the training set and the validation set varying with the cycle iteration during the network training process

Mcon 和 W-Net、W-Net-Mcon 两两分别对比, 在 U-Net 框架中, 除了视盘中 MA 指标略低, U-Net-Mcon 的各项指标均高于原始 U-Net; 在 M-Net 中, 除视盘中 ACC 指标略低, M-Net-Mcon 的其他指标均比 M-Net 好; 对于 W-Net 结构, 使用残差多尺度卷积模块的 W-Net-Mcon 所有指标均比 W-Net 的结果好。通过实验对比, 在相同的网络框架下, 使用带残差多尺度卷积模块的网络性能明显优于普通卷积单元, 这是因为残差多尺度卷积具有不同尺度的感受野来提取网络中不同层次的特征和丰富语义信息。同时, 将不同尺度的特征映射连接起来进行特征融合, 另有外面的跳跃连接构成残差学习进而使得网络更加容易训练, 可有效抑制梯度消失问题和网络退化问题。因此, W-Net 虽加深了网络, 但其性能并未退化, 并使得分割效果有所提高。

最后, 本文将带残差多尺度卷积模块和 W-Net 深度神经网络框架进行结合, 即 W-Net-Mcon。W-Net-Mcon 与其他网络比较, 在所有评价指标中均

最好。这是因为 W-Net 通过易于实现的级联方式, 拓展了网络的深度使得网络提取到更深层次的语义信息。同时为了解决网络加深可能导致的网络退化和梯度消失问题, 引入了跳跃连接和深度监督。一方面, 可以使网络在梯度反向传播时更容易到达浅层的神经层; 另一方面, 可鼓励特征复用, 加强特征传播, 实现不同深浅层次的特征融合, 进而能抑制梯度消失并减少训练参数的数量。通过上述分析, 我们可以认为所提的网络 W-Net-Mcon 明显优于原始的 U-Net<sup>[25]</sup>和 M-Net<sup>[26]</sup>, 视杯、视盘分割结果在主要评价指标 OE 中分别降至 0.178 0 和 0.066 5, 达到最精准的分割效果。

图 6 中直观地展示了不同网络的分割效果。从总体上来看, W-Net-Mcon 网络输出的分割结果形状更接近于专家标注的图像形状, 且噪声点更少; 从细节来看, W-Net-Mcon 网络输出的分割结果显示边缘不规则形状更少, 且更加平滑, 这与表 1 的结果是一致的。因此, W-Net-Mcon 会产生更加精确的视杯、视盘分割结果, 能够有效帮助医生快

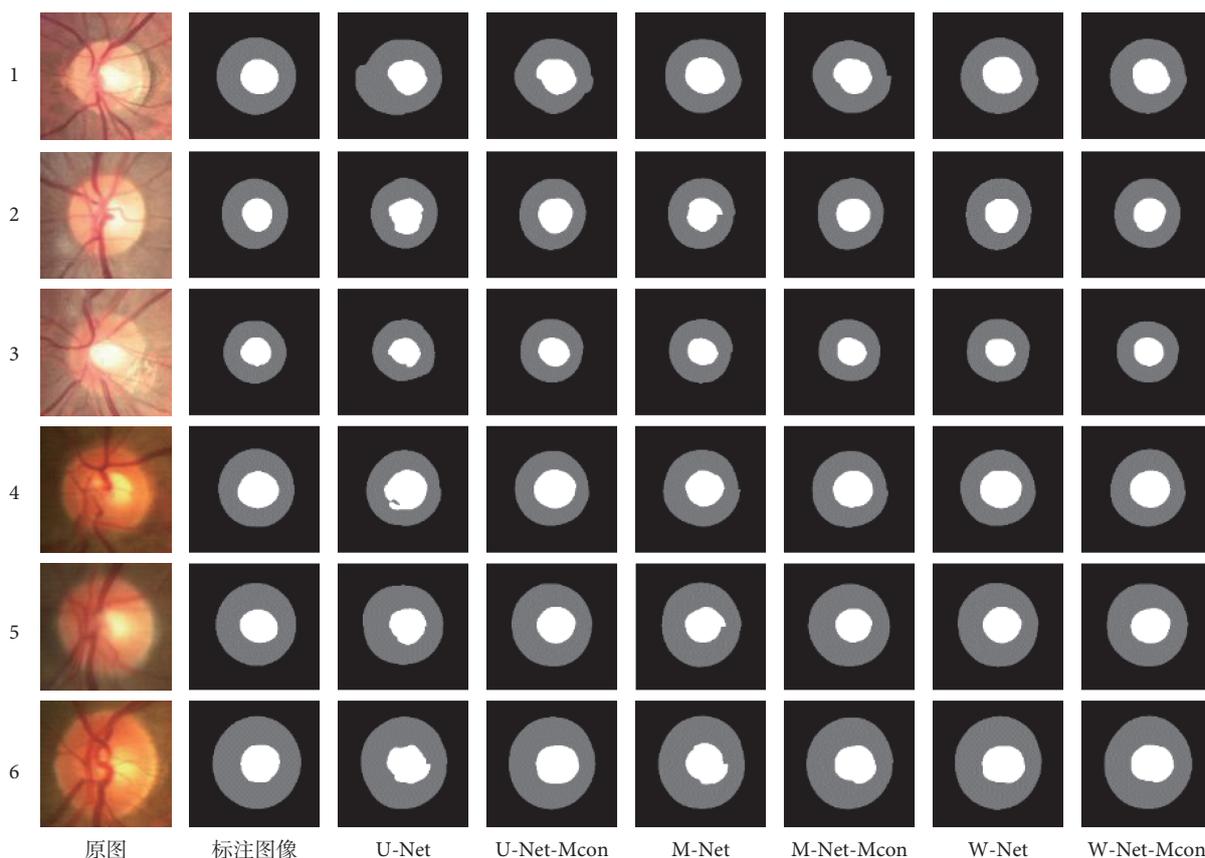


图6 各网络视杯视盘联合分割的效果比较

Fig.6 Comparison of the joint segmentation of the disc and cup of each network

速分割出视杯视盘，进而评估杯盘比和完成青光眼的辅助筛查与诊断，从而提高医生的效率和降低因医生疲劳或主观性可能造成的误诊风险，也可将其推广到青光眼大规模筛查任务中。

此外，本研究对未来工作提出了两个重要方向。一是泛化数据集，需要在更多的数据集上进行验证，以确保方法具有更强的泛化性能和鲁棒性。因此我们计划完成并发布一个包含眼底图像视杯视盘分割的多目标数据集，利用数据驱动深度学习技术发展。二是关于超参数的选择，本文中一些超参数的选择主要还是根据经验和实验效果来人工选择的，缺乏系统的调参方式，这也是目前深度学习面临的一个重要问题，因此我们将不断对其进行探究与完善。

#### 4 结论

本文提出了一种称为 W-Net-Mcon 的残差多尺度全卷积深度神经网络结构，并将视盘、视杯联合分割任务看作单阶段多标签标记问题。在图像预处理阶段，我们引入了极坐标变换来实现空间约束和平衡杯盘比。我们提出的 W-Net-Mcon 是以 W

型的全卷积网络作为主体框架，在网络输入口加入图像金字塔来构造多尺度输入层，同时在与之对应相同尺度的网络中构造多个侧输出层作为早期的分类器来生成局部预测输出。此外，本文还提出了带残差多尺度全卷积模块，利用不同大小的感受野来提取不同层次的语义信息，再加上同一尺度的不同深度层次的跳跃连接，起到了鼓励特征复用、加强特征传播、不同深浅层次的特征融合作用。最后，提出了一种多标签损失函数来指导网络分割。我们已经证明该系统在 REFUGE 数据集<sup>[31]</sup>获得了较好的分割结果。实验结果表明，该网络实现了视盘、视杯联合分割并有效提高了分割精度，这将有助于推进大规模的青光眼早期筛查，因此 W-Net-Mcon 的残差多尺度全卷积深度神经网络结构具有良好的应用前景。

利益冲突声明: 本文全体作者均声明不存在利益冲突。

#### 参考文献

- 1 Tham Y C, Li Xiang, Wong T Y, *et al.* Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology*, 2014, 121(11): 2081-2090.

- 2 Garway-Heath D F, Hitchings R A. Quantitative evaluation of the optic nerve head in early glaucoma. *Br J Ophthalmol*, 1998, 82(4): 352-361.
- 3 Jonas J B, Bergua A, Schmitz-Valckenberg P, *et al.* Ranking of optic disc variables for detection of glaucomatous optic nerve damage. *Invest Ophthalmol Vis Sci*, 2000, 41(7): 1764-1773.
- 4 Michael D H O D. Optic disc size, an important consideration in the glaucoma evaluation. *Clin Eye Vis Care*, 1999, 11(2): 59-62.
- 5 Wang Jinke, Cheng Yuanzhi, Guo Changyong, *et al.* Shape-intensity prior level set combining probabilistic atlas and probability map constrains for automatic liver segmentation from abdominal CT images. *Int J Comput Assist Radiol Surg*, 2016, 11(5): 817-826.
- 6 Shi Changfa, Cheng Yuanzhi, Wang Jinke, *et al.* Low-rank and sparse decomposition based shape model and probabilistic atlas for automatic pathological organ segmentation. *Med Image Anal*, 2017, 38: 30-49.
- 7 Zhang Jianpeng, Xia Yong, Xie Yutong, *et al.* Classification of medical images in the biomedical literature by jointly using deep and handcrafted visual features. *IEEE J Biomed Health Inform*, 2018, 22(5): 1521-1530.
- 8 Shi Changfa, Cheng Yuanzhi, Liu Fei, *et al.* A hierarchical local region-based sparse shape composition for liver segmentation in CT scans. *Pattern Recognit*, 2016, 50: 88-106.
- 9 de Oliveira L A Jr, Medeiros H R, Macêdo D, *et al.* SegNetRes-CRF: A deep convolutional encoder-decoder architecture for semantic image segmentation//2018 International Joint Conference on Neural Networks (IJCNN). Rio de Janeiro: IEEE, 2018, 39(12): 2481-2495.
- 10 Guo Z, Li X, Huang H, *et al.* Deep learning-based image segmentation on multi-modal medical imaging. *IEEE Trans Radiat Plasma Med Sci*, 2019, 3(2): 162-169.
- 11 Liu Shouqiang, Li Miao, Li Min, *et al.* Research of animals image semantic segmentation based on deep learning. *Concurrency and Computation: Practice and Experience*, 2020, 32(1): e4892.
- 12 Joshi G D, Sivaswamy J, Krishnadas S R. Optic disk and cup segmentation from monocular color retinal images for glaucoma assessment. *IEEE Trans Med Imaging*, 2011, 30(6): 1192-1205.
- 13 Dehghani A, Moghaddam H A, Mohammad-Shahram M. Optic disc localization in retinal images using histogram matching. *EURASIP Journal on Image and Video Processing*, 2012, 2012(1): 19.
- 14 肖志涛, 邵一婷, 张芳, 等. 基于眼底结构特征的彩色眼底图像视盘定位方法. *中国生物医学工程学报*, 2016, 35(3): 257-263.
- 15 Almazroa A, Burman R, Raahemifar K, *et al.* Optic disc and optic cup segmentation methodologies for glaucoma image detection: a survey. *J Ophthalmol*, 2015, 2015: 180972.
- 16 郑绍华, 陈健, 潘林, 等. 基于定向局部对比度的眼底图像视盘检测方法. *中国生物医学工程学报*, 2014, 33(3): 289-296.
- 17 Zheng Yuanjie, Stambolian D, O'Brien J, *et al.* Optic disc and cup segmentation from color fundus photograph using graph cut with priors. *Med Image Comput Comput Assist Interv*, 2013, 16(Pt 2): 75-82.
- 18 Aquino A, Gegúndez-Arias M E, Marín D. Detecting the optic disc boundary in digital fundus images using morphological, edge detection, and feature extraction techniques. *IEEE Trans Med Imaging*, 2010, 29(11): 1860-1869.
- 19 刘振宇, 汪森. 改进区域生长算法在视杯图像分割中的应用. *辽宁大学学报: 自然科学版*, 2017, 44(2): 105-113.
- 20 Cheng Jun, Liu Jiang, Xu Yanwu, *et al.* Superpixel classification based optic disc and optic cup segmentation for glaucoma screening. *IEEE Trans Med Imaging*, 2013, 32(6): 1019-1032.
- 21 Sironi A, Turetken E, Lepetit V, *et al.* Multiscale centerline detection. *IEEE Trans Pattern Anal Mach Intell*, 2016, 38(7): 1327-1341.
- 22 Kamnitsas K, Ledig C, Newcombe V F J, *et al.* Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal*, 2017, 36: 61-78.
- 23 Shen Wei, Zhou Mu, Yang Feng, *et al.* Learning from experts: Developing transferable deep features for patient-level lung cancer prediction//Ourselin S, Joskowicz L, Sabuncu M, *et al.* Medical image computing and computer-assisted intervention: Lecture notes in computer science. Cham: Springer, 2016, 9901: 124-131.
- 24 Song Jiangdian, Yang Caiyun, Fan Li, *et al.* Lung lesion extraction using a toboogan based growing automatic segmentation approach. *IEEE Trans Med Imaging*, 2016, 35(1): 337-353.
- 25 Ronneberger O, Fischer P, Brox T, *et al.* U-net: Convolutional networks for biomedical image segmentation//Navab N, Hornegger J, Wells W, *et al.* Medical image computing and computer-assisted intervention: Lecture notes in computer science. Cham: Springer, 2015: 234-241.
- 26 Fu Huazhu, Cheng Jun, Xu Yanwu, *et al.* Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. *IEEE Trans Med Imaging*, 2018, 37(7): 1597-1605.
- 27 Mehta R, Sivaswamy J. M-net: A Convolutional Neural Network for deep brain structure segmentation//2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). Melbourne: IEEE, 2017: 437-440.
- 28 Zhou Z, Rahman Siddiquee M M, Tajbakhsh N, *et al.* UNet++: A nested U-Net architecture for medical image segmentation//Stoyanov D, Taylor Z, Carneiro G, *et al.* Deep learning in medical image analysis and multimodal learning for clinical decision support. DLMIA 2018, ML-CDS 2018. Lecture notes in computer science. Cham: Springer, 2018, 11045: 3-11.
- 29 Yu F, Wang Dequan, Shelhamer E, *et al.* Deep layer aggregation//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 2403-2412.
- 30 Sevastopolsky A. Optic disc and cup segmentation methods for glaucoma detection with modification of U-Net convolutional neural network. *Pattern Recogn Image Anal*, 2017, 27(3): 618-624.
- 31 Aganj I, Harisinghani M G, Weissleder R, *et al.* Unsupervised medical image segmentation based on the local center of mass. *Sci Rep*, 2018, 8(1): 13012.
- 32 Zuiderveld K. Contrast limited adaptive histogram equalization. San Diego: Academic Press Professional, Inc, 1994.
- 33 Crum W R, Camara O, Hill D L. Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE Trans Med Imaging*, 2006, 25(11): 1451-1461.
- 34 Lin T Y, Goyal P, Girshick R, *et al.* Focal loss for dense object detection//Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 2980-2988..
- 35 Huang G, Liu Z, Van Der Maaten L, *et al.* Densely connected convolutional networks//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017: 4700-4708.
- 36 He Kaiming, Zhang Xiangyu, Ren Shaoqing, *et al.* Deep residual learning for image recognition//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016: 770-778.
- 37 Szegedy C, Ioffe S, Vanhoucke V, *et al.* Inception-v4, inception-resnet and the impact of residual connections on learning//Thirty-First AAAI Conference on Artificial Intelligence. San Francisco: AAAI, 2017: 1-12.
- 38 Gu Zaiwang, Cheng Jun, Fu Huazhu, *et al.* CE-Net: context encoder network for 2D medical image segmentation. *IEEE Trans Med Imaging*, 2019, 38(10): 2281-2292.

收稿日期: 2019-09-03 修回日期: 2020-04-23

本文编辑: 李蓓兰

