

中图法分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(2020)10-2259-12

论文引用格式: Gao Y Q, Guo S, Li N, Wang K, Kang H and Li T. 2020. Arteriovenous classification method in fundus images based on semantic fusion. Journal of Image and Graphics 25(10): 2259-2270(高颖琪, 郭松, 李宁, 王恺, 康宏, 李涛. 2020. 语义融合眼底图像动静脉分类方法. 中国图象图形学报 25(10): 2259-2270 [DOI: 10.11834/jig.200187])

语义融合眼底图像动静脉分类方法

高颖琪¹, 郭松¹, 李宁¹, 王恺^{1,2}, 康宏^{1,3}, 李涛¹

1. 南开大学计算机学院, 天津 300350; 2. 天津市医药数据分析与统计研究重点实验室, 天津 300071;

3. 北京上工医信科技有限公司, 北京 100176

摘要: 目的 眼底图像中的动静脉分类是许多系统性疾病风险评估的基础步骤。基于传统机器学习的方法操作复杂,且往往依赖于血管提取的结果,不能实现端到端的动静脉分类,而深度语义分割技术的发展使得端到端的动静脉分类成为可能。本文结合深度学习强大的特征提取能力,以提升动静脉分类精度为目的,提出了一种基于语义融合的动静脉分割模型 SFU-Net(semantic fusion based U-Net)。方法 针对动静脉分类任务的特殊性,本文采用多标签学习的策略来处理该问题,以降低优化难度。针对动静脉特征的高度相似性,本文以 DenseNet-121 作为 SFU-Net 的特征提取器,并提出了语义融合模块以增强特征的判别能力。语义融合模块包含特征融合和通道注意力机制两个操作:1) 融合不同尺度的语义特征从而得到更具有判别能力的特征;2) 自动筛选出对目标任务更加重要的特征,从而提升性能。针对眼底图像中血管与背景像素之间分布不均衡的问题,本文以 focal loss 作为目标函数,在解决类别不均衡问题的同时重点优化困难样本。结果 实验结果表明,本文方法的动静脉分类的性能优于现有绝大多数方法。本文方法在 DRIVE(digital retinal images for vessel extraction)数据集上的灵敏度(sensitivity)与目前最优方法相比仅有 0.61% 的差距,特异性(specificity)、准确率(accuracy)和平衡准确率(balanced-accuracy)与目前最优方法相比分别提高了 4.25%、2.68% 和 1.82%;在 WIDE 数据集上的准确率与目前最优方法相比提升了 6.18%。结论 语义融合模块能够有效利用多尺度特征并自动做出特征选择,从而提升性能。本文提出的 SFU-Net 在动静脉分类任务中表现优异,性能超越了现有绝大多数方法。

关键词: 眼底图像;动静脉分类;深度学习;语义分割;特征融合

Arteriovenous classification method in fundus images based on semantic fusion

Gao Yingqi¹, Guo Song¹, Li Ning¹, Wang Kai^{1,2}, Kang Hong^{1,3}, Li Tao¹

1. College of Computer Science, Nankai University, Tianjin 300350, China;

2. Key Laboratory for Medical Data Analysis and Statistical Research of Tianjin, Tianjin 300071, China;

3. Beijing Shangong Medical Technology and Development Co. Ltd., Beijing 100176, China

Abstract: **Objective** Arteriovenous (A/V) classification in fundus images is a fundamental step for the risk assessment of many systemic diseases. A/V classification methods based on traditional machine learning require complicated feature engineering, consistently rely on the results of blood vessel extraction, and cannot achieve end-to-end A/V classification. The development of deep semantic segmentation technology makes end-to-end A/V classification possible, and has been com-

收稿日期: 2020-05-19; 修回日期: 2020-07-06; 预印本日期: 2020-07-13

基金项目: 国家自然科学基金项目(61872200); 天津市自然科学基金项目(19JCZDJC31600)

Supported by: National Natural Science Foundation of China(61872200); Tianjin Municipal Natural Science Foundation (19JCZDJC31600)

monly used in fundus image analysis. In this paper, a segmentation model semantic fusion based U-Net (SFU-Net) is proposed combined with the powerful feature extraction capabilities of deep learning to improve the accuracy of A/V classification. **Method** First, the arteries and veins in the fundus image belong to blood vessels and are highly similar in structure. Existing deep learning-based A/V classification methods frequently treat this problem as a multiclassification problem. This paper proposes a multilabel learning strategy to address this problem for reducing the difficulty of optimization and deal with the situation where the arteries and veins in the fundus image cross. The lower layers of the network are mainly responsible for extracting the common features of the two structures. The upper layers of the network learn two binary classifiers and extract the arteries and veins independently. Second, considering the high similarity of the description features of arteries and veins in color and structure, this paper improves the U-Net architecture in two aspects. 1) The original simple feature extractor of U-Net is replaced by DenseNet-421. The original U-Net encoder is composed of 10 convolutional layers and four maximum pooling layers, and the feature extraction capability is extremely limited. By contrast, DenseNet-421 has many convolutional layers, and the introduction of dense connections makes the feature utilization rate high, the transmission efficiency of features and gradients in the network is high, and the feature extraction ability is strong. This paper reduces four downsampling operations of U-Net to three, and the input image is downsampled by eight times to avoid the loss of detailed information. 2) A semantic fusion module is proposed. The semantic fusion module includes two operations, namely, feature fusion and channelwise attention mechanism. Low-level features have high resolution and contain many location and detail information, but few semantic features and many noises. High-level features have strong semantic information, but their resolution is extremely low and the detail information is few. The features from different layers are first fused to enhance their distinguishing ability. For the fused features, the channelwise attention mechanism is used to select the features. The convolution filter can only capture local information. The global average pooling operation is performed on the input features in the channel dimension to capture global context information. Each element of the resulting vector is a concentrated representation of its corresponding channel. Two nonlinear transformations are then performed on the vector to model the correlation between channels and reduce the amount of parameters and calculations. The vector is restored to its original dimension and normalized to 0-1 through the sigmoid gate. Each element in the obtained vector is regarded as the importance of each channel in the input feature, and each feature channel of the input feature is weighted through a multiplication operation. Through the channel attention mechanism, the network can automatically focus on the feature channels that are important to the task while suppressing the features that are unimportant, thereby improving the performance of the model during the training process. Third, considering the problem of uneven distribution between blood vessels and background pixels in the fundus image, this paper takes focal loss as the loss function to solve the problem of class imbalance and focus on difficult samples at the same time. Focal loss introduces parameters α and γ in the cross-entropy loss function. Parameter α is used to balance the difference between positive and negative samples. Parameter γ adjusts the degree where the loss of simple samples is reduced, thereby amplifying the difference between the loss values of difficult and simple samples. The values of the two parameters are determined through cross-validation. The overall optimization goal is the sum of the focal loss of arteries and veins, thereby optimizing the arteries and veins during training. **Result** The proposed method is verified on two public datasets, namely, digital retinal images for vessel extraction (DRIVE) and WIDE, and its performance is evaluated from two perspectives, namely, segmentation and classification. Experimental results demonstrate that the proposed method shows better performance than most existing methods. The proposed method achieves an area under the curve of 0.968 6 and 0.973 6 for segmenting arteries and veins on the DRIVE dataset, and the sensitivity, specificity, accuracy and balanced-accuracy of A/V classification are 88.39%, 94.25%, 91.68%, and 91.32%, respectively. Compared with state-of-the-art method, the sensitivity of the proposed method only decreases by 0.61%, and specificity, accuracy, and balanced-accuracy have absolute improvements of 4.25%, 2.68%, and 1.82%, respectively. The proposed method achieves an accuracy of 92.38%, which is 6.18% higher than the state-of-the-art method. **Conclusion** The fusion module can effectively use multi-scale features and automatically select many important features, thereby improving performance. The proposed method performs well in A/V classification, exceeding most existing methods.

Key words: fundus images; arteriovenous classification; deep learning; semantic segmentation; feature fusion

0 引言

视网膜血管是唯一可以采用成像技术直接观察到的体内血管,具有非侵入性的优点。视网膜血管的形态学变化可以作为许多系统性疾病生物标志,如直径、曲折度和分形维数(Abramoff等,2010)等。其中,动静脉直径比值是一种基本的特征,其变化可用于多种系统性疾病的发生和死亡风险的预测。具体而言,动静脉直径的比值越小,表示发生高血压(Wong等,2001)、脑卒中(Cheung等,2017)、冠心病(Liew和Wang,2011)和动脉粥样硬化(Ikram等,2004)等疾病的风险越高。因此,作为测量动静脉直径比值的基础步骤,区分视网膜血管中的动脉和静脉,对许多系统性疾病早期预测和观察病情进展具有重要意义。

彩色眼底图像是目前最成熟,也是临床上应用最广泛的视网膜成像方式,具有采集简单、存储方便以及价格适宜的优势,适合于大规模人群的疾病筛查(Abramoff等,2010)。眼底图像动静脉分类的难点主要在于:这两种结构描述特征的高度相似性,以及眼底图像内部和图像之间亮度和对比度的差异。

在临床上,医生主要通过视觉和几何特征来区分这两种结构(Rothaus等,2009; Dashtbozorg等,2014):1) 动脉较亮而静脉较暗,因为动脉中的氧含量高于静脉中的氧含量,并且眼底图像中动脉的颜色都是相似的;2) 动脉直径小于相邻静脉直径;3) 动脉壁较厚,中央反射带较宽;4) 动脉和静脉可以相互交叉,而动脉和动脉、静脉和静脉永远不会交叉,但动脉和静脉都可以分叉形成更细的血管。

手动区分动静脉是一项非常费时费力的工作,自动的动静脉分类能够弥补医生数量不足的问题,具有高效、低成本的特点。现有的动静脉分类方法总体上可以分为两类:基于传统机器学习的方法和基于深度学习的方法。基于传统机器学习的动静脉分类方法一方面需要复杂的手工提取特征,这就需要研究者对于专业背景知识有更加深刻的理解;另一方面其依赖于血管分割的结果,不能实现端到端的动静脉分类,操作流程较为复杂。而深度学习(LeCun等,2015)凭借其强大的特征提取能力,在医学影像分析包括眼底图像分析中表现出了超越传统方法的性能,但在眼底图像动静脉分类方面的研究

相对较少,由于血管结构的复杂性,分割结果往往存在准确率低以及细血管丢失等问题。

为了解决动静脉特征高度相似、区分难度大的问题,本文提出一种基于语义融合的动静脉分类网络SFU-Net(semantic fusion based U-Net),SFU-Net以改进的U-Net作为基础分割网络,融合不同层次的特征并筛选出对分类任务更加有用的特征,最后用两个二分类器分别提取动脉和静脉,并在DRIVE(digital retinal images for vessel extraction)和WIDE两个公开数据集上进行了验证。

1 眼底图像动静脉分类方法概述

眼底图像中的动静脉分类方法可以分为两类:基于传统机器学习的方法和基于深度学习的方法。

基于传统机器学习的动静脉分类方法通常包含两个步骤:血管提取和血管分类。第一个全自动的动静脉分类方法是Grisan和Ruggeri(2003)提出的,首先将视盘周围的同心圆区域划分成4个象限,对每个象限内最粗的5根血管进行分类,然后采用血管跟踪技术将动静脉标签传播到区域外部,该方法在手动标注的443根血管的验证集上整体分类错误率为12%。与Grisan和Ruggeri(2003)的方法类似,Vázquez等人(2013)先将视盘周围的同心圆区域的血管分段,然后用颜色信息对每个半径上的血管片段分别进行分类,最后采用血管跟踪的方法,将属于同一条血管不同半径上的血管片段连接起来,所有相连的血管片段投票决定最终的分类结果。在100张手动标注的测试图片上,得到了87.68%的准确率。为了计算动静脉比值,Niemeijer等人(2011)和Muramatsu等人(2011)主要关注感兴趣区域(距离视盘中心2~3倍视盘半径的圆环区域)内的动静脉分类问题,算法流程主要包括:提取血管、提取视盘以确定感兴趣区域、提取特征对血管分类、测量血管直径并选择主血管对以及计算动静脉直径比值等。Mirsharif等人(2013)首先确定了主血管的类型,然后利用血管的结构特点,在血管的交点和分支点处检测小血管,小血管的类型由与之相连的主血管确定。Xu等人(2017)通过图像间的正则化和来自同一个受试者的图像的归一化,来减少特征空间的差异,并提取一阶和二阶纹理特征来提升分类性能。从特征提取和特征选择的角度出发,Huang等

人(2018)首先从血管中心线像素提取了大量特征,然后应用基于遗传搜索的特征选择技术,来获得动静脉分类的最优特征子集。马志扬(2015)提出了一种四分段高斯血管模型以定量描述血管片段的灰度横截信息,并应用支持向量机(support vector machine, SVM)算法判定血管片段的动静脉类型。薛岚燕等人(2017)采用K均值聚类实现感兴趣测量区域内动静脉的自动分类。

上述方法都利用局部信息来区分动脉和静脉,为了结合全局结构信息,研究者们引入了图论的方法对血管的整体结构进行建模,以提升分类性能。2009年,Rothaus等人(2009)第一次用图来表示血管,用于半自动的动静脉分类。视网膜上血管的解剖特征被建模为双重约束图,然后采用启发式AC-3算法来解决双层约束搜索问题,以克服NPH(non-deterministic polynomial hard)问题的计算复杂性。Dashtbozorg等人(2014)将整个血管网络建模成一个无向图,然后根据节点的度、边的夹角和血管直径等结构特点将图划分为多个子图,每个子图上的所有血管像素属于同一个类别,最后对每个子图的所有中心线像素点提取30维特征,采用线性判别分析、二次判别分析和K最近邻分类器进行分类,每个子图所属的类别由所有中心线像素点投票决定。该算法在CT-DRIVE数据集上的血管中心线像素分类准确率达到87.4%。Estrada等人(2015a)提出了一种半自动的动静脉分类方法,该方法能够更好地对中小血管进行分类。他们建立了一个强大的可能性模型,结合了交点处的颜色、重叠和局部生长方面的先验知识,然后迭代搜索可能的解空间。Hu等人(2015)通过建立一个强连接的网络来修复血管之间的连接性,并进一步将血管网络划分为动静脉树。使用手动标注的血管作为输入,该算法的分类准确率为88.15%,使用算法自动提取的血管作为输入,分类准确率为86.11%。Joshi等人(2014)将血管分割图转换成血管片段图,建立图结构,用Dijkstra算法通过最小化累积边成本来搜索血管子树,最后用K均值聚类算法来区分动脉子树和静脉子树。Joshi等人(2014)在来自50个受试者的50幅眼底图像上进行了测试,血管像素正确分类的准确率为91.44%,主血管片段的分类准确率为96.42%。针对用扫描激光检眼镜获取的超广角视网膜眼底图像中的动静脉分类问题,Pellegrini等人(2018)提出了

一种基于图割的方法,从生成的图表示中计算全局最佳的动脉和静脉网络之间的划分。这是第一个全自动处理此类图像中动静脉分类问题的方法,在WIDE数据集上像素级准确率为86.2%,片段级准确率达到86.4%。Zhao等人(2020)采用优势集聚类的方法,将视网膜血管拓扑估计和动静脉分类建模为成对聚类问题,在5个公开数据集上的实验表明,该方法能够准确重建血管拓扑结构并对血管进行分类。此外,他们还手动标注了INSPIRE, IOSTAR, VICAVR和DRIVE数据集的血管拓扑结构并公开发布,以方便研究人员使用。

深度语义分割技术的发展使得端到端的动静脉分类成为可能。Xu等人(2018)采用一个改进的全卷积网络来同时分割动脉和静脉,并公开了一个新的动静脉分类数据集REVEAL。考虑到类型不确定的血管像素,Galdran等人(2019)将动静脉分类建模成一个四分类问题,即动脉、静脉、背景以及血管类型不确定的像素。Girard等人(2019)将深度学习和图传播的方法相结合,分割血管的同时将血管分为动脉和静脉。该方法包含两个步骤:1)用U-Net进行血管分割和分类;2)U-Net输出的标签通过血管的图表示进行传播,该图的节点为血管分支,边的权重定义为连接两个分支对的成本。为了高效地传播标签,将图简化为其最小生成树。该方法在DRIVE数据集上的血管分割准确率达到94.8%,在CT-DRIVE数据集上动静脉分类的灵敏度为93.7%,特异性为92.9%。虽然同样采用深度学习的方法来处理动静脉分类问题,Welikala等人(2017)采用分类的方式而不是语义分割的方式。他们提出了一个具有3个卷积层和3个全连接层的分类网络,对于血管片段上的每一个血管中心线像素,以该目标像素为中心,提取一个 31×31 像素的图像块,然后将其随机裁剪为 25×25 像素的图像块输入分类网络对该像素进行分类,血管中心线像素的最终标签取决于提取的10个不同的 25×25 像素的图像块的结果,血管中心线像素投票的均值作为整个血管片段的标签。

与传统机器学习方法相比,深度学习具有强大的特征提取能力,简化了动静脉分类任务的流程。然而,基于深度学习的动静脉分类方法往往将该问题作为一个三分类问题来处理,即将眼底图像中的每个像素分为动脉、静脉和背景(或四分类,动脉、

静脉、背景以及不确定的像素)。但是动静脉分类的特殊性在于,动脉和静脉都属于血管像素,并且具有非常相似的结构,因此本文将动静脉分类作为一个多标签而不是多分类问题来处理,首先提取动静脉的共同特征,然后用两个独立的二分类器分别提取动脉和静脉。

2 动静脉分类网络 SFU-Net

2.1 模型概述

本文提出的基于语义融合的动静脉分割网络 SFU-Net 主要由 3 个模块构成,模型结构如图 1 所

示。首先,一个 U 形的卷积网络作为主体结构从输入图像中提取丰富的语义特征;然后,语义融合模块融合不同层次的语义特征并自动筛选对分类更加重要的特征;最后,多标签分类层实现同时分割动脉和静脉。

2.2 U 形基础网络

U-Net(Ronneberger 等,2015) 是生物医学图像分割领域非常有效的一种语义分割网络,SFU-Net 以改进的 U-Net 作为主体结构。与最初的 U-Net 相似,SFU-Net 中的 U 形基础网络由编码器和解码器构成,编码器负责特征提取,解码器负责恢复空间分辨率。

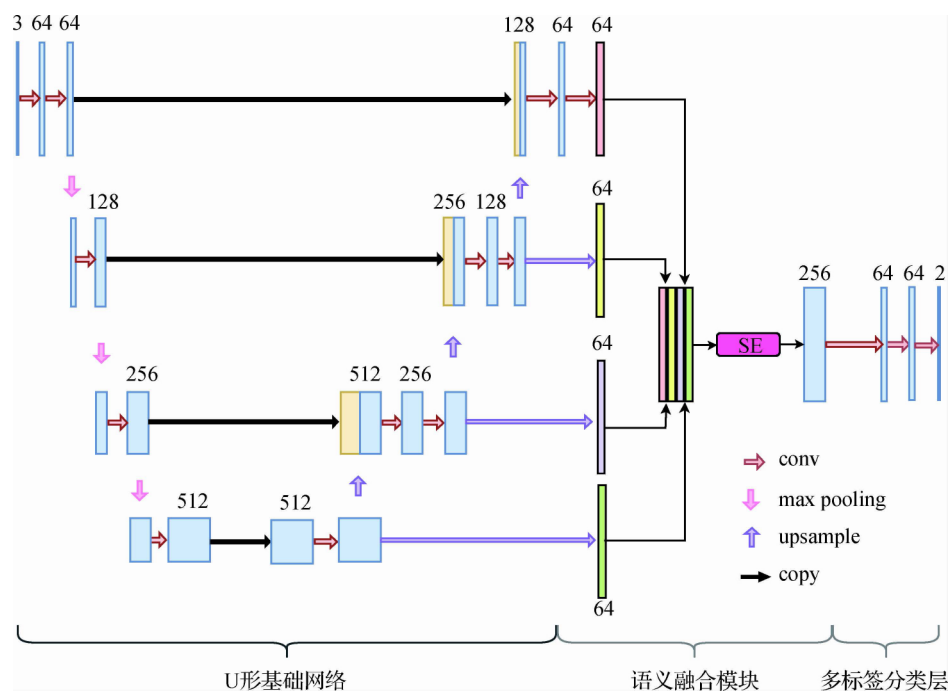


图1 SFU-Net 模型结构

Fig.1 Overview of proposed SFU-Net

改进后的 U 形网络以 DenseNet-121(Huang 等,2017) 作为编码器,DenseNet-121 由 4 个密集块构成,每两个相邻块之间有一个过渡层,通过卷积和池化操作来改变特征的大小。

解码器与 U-Net 的解码器类似,将来自编码器的特征上采样 2 倍,然后利用跳跃连接将上采样后的特征与来自编码器的同分辨率的特征进行拼接,从而缩小编码器和解码器特征之间的差异,然后将通道数减半,逐步恢复到原始尺寸。

2.3 语义融合模块

语义融合模块包含两个操作:特征融合和基于

通道的注意力机制。

1) 特征融合。U 形基础网络从输入图像中提取特征并逐渐恢复空间细节信息,虽然短连接和逐层上采样的方式能够在一定程度上避免细节信息的丢失,然而网络低层仍然保留了较多的细节信息而语义信息较少,网络高层具有丰富的语义信息,但细节信息有所丢失。因此,融合不同层次的特征能够有效提升分类性能,并获得更加精细的分割结果。

首先对来自解码器不同层的特征进行上采样操作,恢复至原图尺寸,得到 4 个特征图,记为 X_1 , X_2 , X_3 和 X_4 ,每个特征图包含 64 个特征通道,然后将

不同层上采样后的特征进行级联,得到含有 256 个特征通道的特征图,即

$$X = [X_1, X_2, X_3, X_4] \quad (1)$$

2) 基于通道的注意力机制。即 SE(squeeze and excitation) 操作,从输入的 256 个通道中筛选对分类任务最重要的特征(Hu 等,2018),其操作流程如图 2 所示。

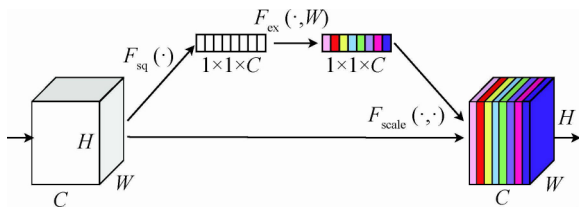


图 2 SE 操作示意图

Fig. 2 SE operation diagram

首先,考虑到卷积滤波器只能捕捉到该区域的局部信息,为了能够利用全局上下文信息,对维度为 $H \times W \times C$ (SFU-Net 中 $C = 256$) 的输入特征 X 的每个通道进行全局平均池化操作 $F_{sq}(\cdot)$,得到向量 $z \in \mathbf{R}^C$, z 中第 c 个元素作为 X 的第 c 个通道的全局信息的浓缩表示,即

$$z_c = F_{sq}(X_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W x_c(i, j) \quad (2)$$

然后,采用两个全连接层对通道间的相关性进行建模,第 1 个全连接层将输入特征维度降低到原来的 $1/r$,经过非线性激活后通过第 2 个全连接层将特征升回原来的维度,进一步通过 sigmoid 门将权重归一化到 0-1 之间,这一系列操作记为 $F_{ex}(\cdot, W)$,即

$$s = F_{ex}(z, W) = \sigma(W_2 \delta(W_1 z)) \quad (3)$$

式中, δ 表示非线性激活函数,实验中采用 ReLU 函数, $W_1 \in \mathbf{R}^{\frac{C}{r} \times C}$, $W_2 \in \mathbf{R}^{C \times \frac{C}{r}}$ 。采用两个全连接层意味着引入了额外的非线性激活,从而更好地拟合通道间的相关性,并减少计算量和参数量。

最后,将 s 中的每个元素看做是对输入特征 X 进行特征选择后每个通道的重要性,通过乘法操作对 X 中的每个特征通道进行加权,该操作表示为 $F_{scale}(\cdot, \cdot)$,即

$$\tilde{X}_c = F_{scale}(X_c, s_c) = s_c \cdot x_c \quad (4)$$

最终得到的特征为 $\tilde{X} = [\tilde{X}_1, \tilde{X}_2, \tilde{X}_3, \tilde{X}_4]$ 与输入特征 X 具有相同的维度。

通过 SE 操作,网络能够在增加极少的计算量和参数数量的条件下,在训练过程中自动关注对任务更重要的特征通道,抑制那些不重要的特征,从而提升模型的性能。

2.4 多标签分类层

多标签分类层首先将语义融合模块得到的 256 维通道的特征图降为 64 维,然后利用两层连续的具有 64 个过滤器的卷积层进一步提取区分动脉和静脉的特征,最后用两个二分类器独立地区分动脉和静脉。眼底图像中背景像素和血管像素的分布严重不平衡,据统计,DRIVE 数据集中血管和背景像素的比例约为 1:10,而动脉和静脉像素数目相差不大。为了解决类别不平衡的问题,两个二分类器均以 focal loss 作为损失函数。

focal loss 在交叉熵损失函数中引入了参数 α 和 γ ,定义为

$$FL(p, y) = - \sum_{i=1}^{H \times W} (\alpha (1 - p_i)^\gamma y_i \log p_i + p_i^\gamma (1 - \alpha) (1 - y_i) \log (1 - p_i)) \quad (5)$$

式中, H 和 W 为分割图的高度和宽度, $y \in \{0, 1\}$ 为真实标签, $p \in [0, 1]$ 为预测概率值。参数 $\alpha \in [0, 1]$ 称为平衡因子,用来平衡正负样本本身的差异,对于正样本引入权重 α ,对于负样本引入权重 $1 - \alpha$ 。参数 $\gamma \geq 0$ 调节简单样本损失被降低的程度,从而将困难样本和简单样本损失的差异放大。超参数 α 和 γ 的值通过消融实验来确定,本文所使用的 α 和 γ 的值分别为 0.7 和 2。

通过这两个参数的引入,focal loss 不仅解决了类别不平衡的问题,而且区分了简单和困难样本,使得在网络训练过程中更加关注困难的、错分的样本。

多标签损失函数定义为动脉和静脉的 focal loss 之和,即

$$L(p, y) = \sum_{k=1}^2 FL(p_k, y_k) \quad (6)$$

从而在网络训练过程中同时优化动脉和静脉。

3 实验结果与分析

3.1 实验设置

3.1.1 数据集

本文实验在两个公开数据集 DRIVE(Staal 等,2004) 和 WIDE(Estrada 等,2015b) 上进行。

DRIVE 是眼底图像血管分割和动静脉分类中最常用的数据集,包含 40 幅彩色眼底图像。这些图像是从 400 幅糖尿病受试者的眼底图像中随机选择的,所有图像均以黄斑为中心,由佳能 CR5 非散瞳 3CCD 相机以 45° 视角拍摄,分辨率为 565 × 584 像素。第三方机构对 DRIVE 数据集进行了动静脉的标注,公开的有 3 个版本,AV-DRIVE、CT-DRIVE 和 RITE。本文采用 RITE 标注(Hu 等, 2015)作为动静脉的真实标签,每幅图像均标注了 4 种类型的血管:红色表示动脉,蓝色表示静脉,绿色表示动脉和静脉重合的部分,白色表示类型不确定的血管。

WIDE 数据集采用 Optos 200Tx 超广角设备(Optos plc, Dunfermline, Scotland, UK)拍摄,共包含 30 幅高分辨率眼底图像。与 RITE 相同,WIDE 也标注了 4 种类型的血管。

3.1.2 数据预处理

为了避免过拟合并提升模型的性能,本实验采用旋转、镜像和加入随机噪声等方式将训练集扩充为 60 倍,类型不确定的血管被视为背景。对于 DRIVE 数据集,图像以原尺寸输入网络,对于 WIDE 数据集,图像被缩小为 700 × 500 像素输入网络。

3.1.3 实验环境

实验采用一台配置有双 CPU 6 核 Intel E5 处理器,128 GB 内存,Ubuntu 16.04 操作系统和 4 块 Nvidia GTX 1080Ti 显卡的工作站。模型基于 PyTorch 深度学习框架实现,PyTorch 版本为 1.0.0,Python 版本为 2.7。

3.1.4 模型训练

本文采用 RMSprop 优化算法来训练 SFU-Net,利用在 ImageNet 上训练好的 DenseNet-121 进行微调,采用 L2 正则化来避免过拟合,正则化项权重为 0.000 5,初始学习率设置为 10^{-4} ,训练周期为 300。DRIVE 已经划分好了训练集和测试集,对于 WIDE 数据集,采用两折交叉验证,随机划分训练集和测试集。

3.2 评价指标

本文从分割和分类两个方面来评价模型的性能,动脉和静脉均以 0.5 为阈值将概率图转化为二值图。

对于动静脉分割,本文采用 ROC 曲线(receiver operating characteristic curve)下面积(area under curve, AUC)作为评价指标。ROC 曲线以真正例率

(灵敏度)为纵轴,以假正例率(1 - 特异性)为横轴,通过改变阈值进行绘制,AUC 值越接近 1,性能越好。DRIVE 提供了 FOV(field of view)掩膜的标注,计算 AUC 时只计算 FOV 内的像素;对于 WIDE 数据集,在整幅图像上计算。

对于动静脉分类,本文采用灵敏性(sensitivity)、特异性(specificity)、准确率(accuracy)和平衡准确率(balanced-accuracy, b_{acc}) 4 个指标来评价,并将动脉视为正例,静脉视为负例,指标定义为

$$se = \frac{TP}{TP + FN} \quad (7)$$

$$sp = \frac{TN}{TN + FP} \quad (8)$$

$$acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

$$b_{acc} = \frac{se + sp}{2} \quad (10)$$

式中, TP 表示正确识别的动脉数目, TN 表示正确识别的静脉数目, FP 表示被误分为动脉的静脉像素数目, FN 表示被误分为静脉的动脉像素数目。 se 描述模型检测动脉的性能, sp 描述模型检测静脉的性能, acc 描述模型区分动脉和静脉的性能, b_{acc} 定义为 se 和 sp 的均值,反映了 se 和 sp 之间的平衡。

3.3 消融实验

为了验证语义融合模块的有效性,本文在 DRIVE 数据集上进行了消融实验,加入语义融合模块后 DRIVE 数据集上的性能对比如表 1 所示。

表 1 加入语义融合模块前后 DRIVE 数据集上的结果对比
Table 1 Comparison of results on the DRIVE dataset with or without semantic fusion module

特征融合	SE	动脉	静脉	动静脉分类 / %			
		AUC	AUC	se	sp	acc	b_{acc}
×	×	0.959 7	0.964 4	89.55	90.49	89.94	90.02
×	✓	0.970 7	0.975 4	87.58	93.62	90.91	90.60
✓	✓	0.968 6	0.973 6	88.39	94.25	91.68	91.32

注:加粗字体为每列最优值。

从表 1 中可以看出,只加入特征融合时,模型分割动脉和静脉的性能同时提升了 1% 以上,动静脉分类的 sp , acc 和 b_{acc} 均有所提升。引入 SE 操作后,与不加入 SE 时相比,分割性能下降了约 0.2%,但是分类的 acc 和 b_{acc} 分别提升了 0.77% 和

0.72%。总的来说,与U形基础网络相比,特征融合和SE操作的引入同时带来了分割和分类性能的提升,验证了语义融合模块的有效性。

3.4 与其他方法对比

1) DRIVE数据集。本文将SFU-Net与现有的动静脉分类方法以及经典的自然图像分割模型DeepLabv3+ (Chen等2018)进行对比,DRIVE数据集上的性能对比如表2所示。从表中可以看出,以RITE作为动静脉标注,在相同的实验配置和评价体系下,SFU-Net像素级分类的 se 略低于UA-AV的89%,其余3个指标 sp , acc 和 b_{acc} 均超越了现有的其他方法。需要注意的是,Xu等人(2018)采用十折交叉验证,意味着他们使用了更多的训练数据,而

SFU-Net只使用20幅图像作为训练集,整体的 acc 和 b_{acc} 仍然高于Xu等人(2018)的方法。Welikala等人(2017)将所有图像划分为25幅训练图像、5幅验证图像和10幅测试图像,在私有标注的测试图像上得到了91.97%的 acc 。Srinidhi等人(2019)采用图搜索元启发式方法在CT-DRIVE和AV-DRIVE数据集上得到了非常高的指标,然而,其方法的局限在于他们使用血管标注图进行分类。一方面,在实际应用中,由于工作量大且耗时,血管标注图很难获得;另一方面,对于血管标注图进行建模太过理想化,而算法自动提取的血管总是存在缺陷,比如血管不连续,这会对后面的分析带来不利的影响。

表2 DRIVE数据集上动静脉分类方法对比
Tabel 2 Comparison of arteriovenous classification methods on DRIVE dataset

方法	动静脉标注	$se/\%$	$sp/\%$	$acc/\%$	$b_{acc}/\%$	描述
Huang等人(2018)		70.9	73.8	72.0	72.25	
Hu等人(2015)		-	-	86.11	-	
UA-AV	RITE	89	90	89	89.5	所有血管像素
DeepLabv3+		84.05	85.64	84.94	84.84	
SFU-Net(本文)		88.39	94.25	91.68	91.32	
Xu等人(2018)	RITE	89.7	90.4	-	90.05	十折交叉验证,对所有血管像素进行评价
Hemelings等人(2019)	RITE	-	-	94.25	-	血管中心线,仅限于宽度超过2像素的血管
Srinidhi等人(2019)	AV-DRIVE	96.6	92.9	94.7	94.75	基于血管标注图
Srinidhi等人(2019)	CT-DRIVE	95.0	91.5	93.2	93.25	宽度超过3像素的血管
Dashtbozorg等人(2014)	CT-DRIVE	90	84	87.4	87	血管中心线,仅限于宽度超过3个像素的血管
Girard等人(2019)	CT-DRIVE	93.7	92.9	-	93.3	血管中心线,仅限于宽度超过3个像素的血管
Mirsharif等人(2013)	未知	82.65	85.74	84.05	84.19	血管中心线,仅限于宽度超过3个像素的血管
Welikala等人(2017)	未知	-	-	91.97	-	像素级评估,数据集划分为25幅图像训练,5幅验证,10幅测试
Zhao等人(2020)	未知	94.2	92.7	-	93.45	血管中心线
Xu等人(2017)	私有	91.5	92.9	92.3	92.2	73 003个中心线像素

注:加粗字体为对所有血管像素进行评价时RITE上的最优值,“-”代表无数据。

DRIVE数据集上动静脉分割的ROC曲线如图3所示,曲线上红色圆点为该曲线的最佳截断点。SFU-Net分割动脉和静脉的ROC曲线均完全覆盖DeepLabv3+,呈现出远优于DeepLabv3+的分割性能。

图4展示了SFU-Net和DeepLabv3+在DRIVE

数据集上的分类结果图,分割图中红色表示动脉,蓝色表示静脉,绿色表示动静脉交叉像素,白色表示类型不确定的血管,黄色框内为同一块区域上SFU-Net和DeepLabv3+的结果对比。DeepLabv3+虽然在自然图像分割中表现优异,但眼底图像中血管较细,且动脉和静脉之间特征高度相似,因此Deep-

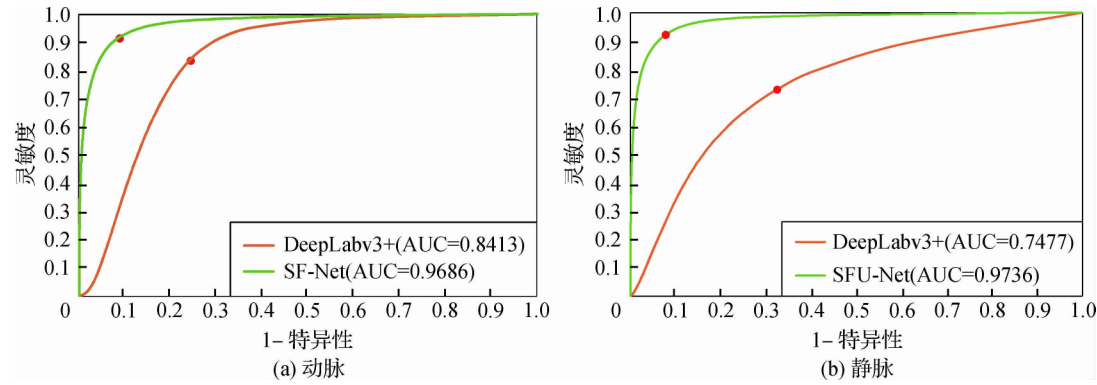


图 3 DRIVE 数据集上动静脉分割的 ROC 曲线

Fig. 3 ROC curve of arteriovenous segmentation on DRIVE dataset ((a) artery; (b) vein)

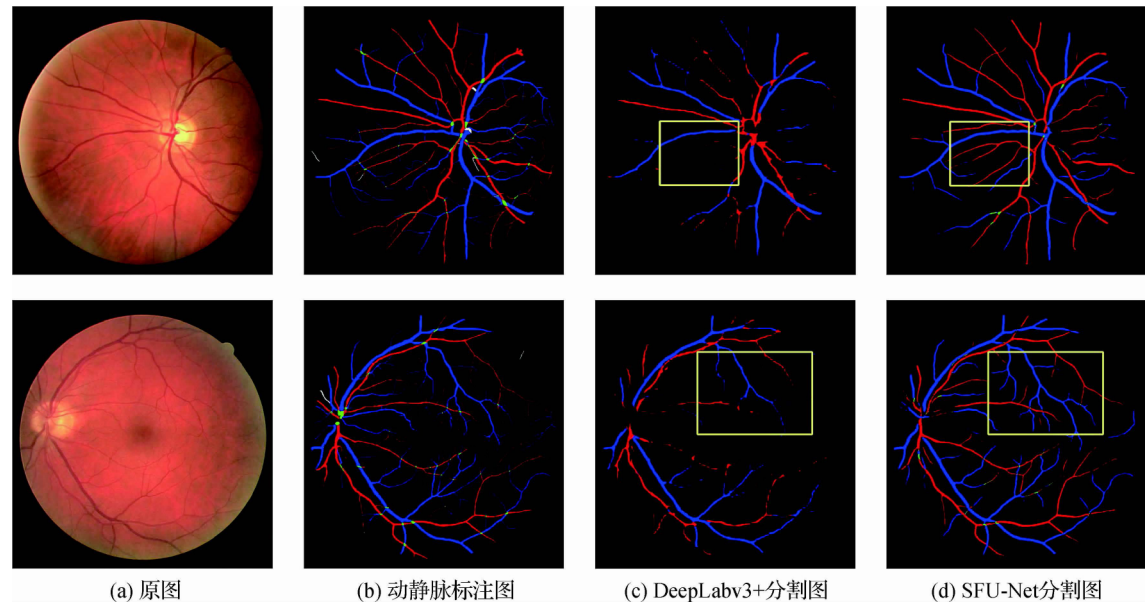


图 4 DRIVE 数据集上动静脉分类结果图

Fig. 4 Arteriovenous classification results on DRIVE dataset

((a) original images; (b) ground truth; (c) segmentation results of DeepLabv3 + ; (d) segmentation results of SFU-Net)

Labv3 + 并不适合于动静脉分割,所呈现的分割结果中大量血管消失,分类准确率也较低。而本文提出的 SFU-Net 识别血管的完整性和准确率均优于 DeepLabv3 + 。

2) WIDE 数据集。不同方法在 WIDE 数据集上的性能对比如表 3 所示。WIDE 数据集是超宽视野眼底图像,血管相素比普通眼底图像更为细小,对 DeepLabv3 + 这类自然图像分割方法来说难度更大,DeepLabv3 + 的分类 acc 只有 76.56%。而本文 SFU-Net 像素级分类的 acc 达到 92.38%,比 Pellegrini 等人(2018)的图割方法高出 6.18%,比 DeepLabv3 + 高出 15.82%。

表 3 WIDE 数据集上动静脉分类方法对比
Tabel 3 Comparison of arteriovenous classification methods on WIDE dataset

方法	$se/\%$	$sp/\%$	$acc/\%$	$b_{acc}/\%$
Pellegrini 等人(2018)	-	-	86.2	-
DeepLabv3 +	71.27	79.13	76.56	75.20
SFU-Net(本文)	91.19	93.39	92.38	92.29

注:加粗字体为每列最优值,“-”代表无数据。

WIDE 数据集上动静脉分割的 ROC 曲线如图 5 所示, SFU-Net 分割动脉和静脉的 AUC 分别比 DeepLabv3 + 高出 0.076 3 和 0.085 6。图 6 展示了 WIDE 数据集上的分割结果图, SFU-Net 的分割效果

虽然优于 DeepLabv3 + ,但是捕捉细血管的能力有

所欠缺,血管断裂的情况比较明显。

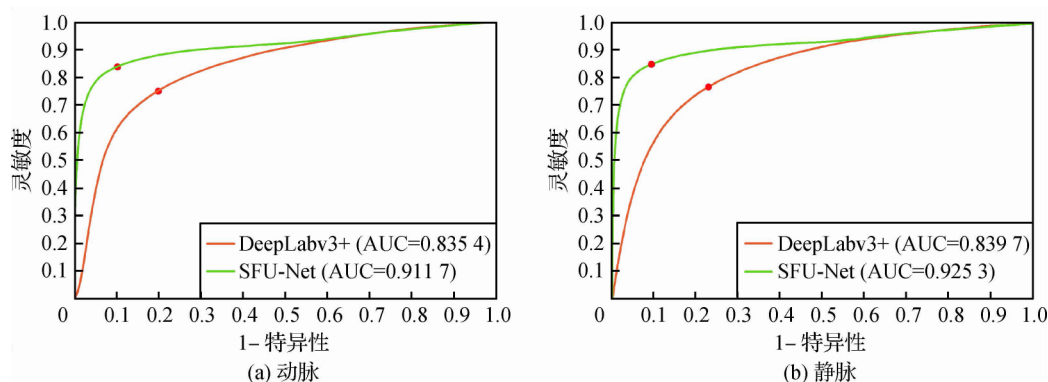


图5 WIDE数据集上动静脉分割的ROC曲线

Fig. 5 ROC curve of arteriovenous segmentation on WIDE dataset ((a) artery; (b) vein)

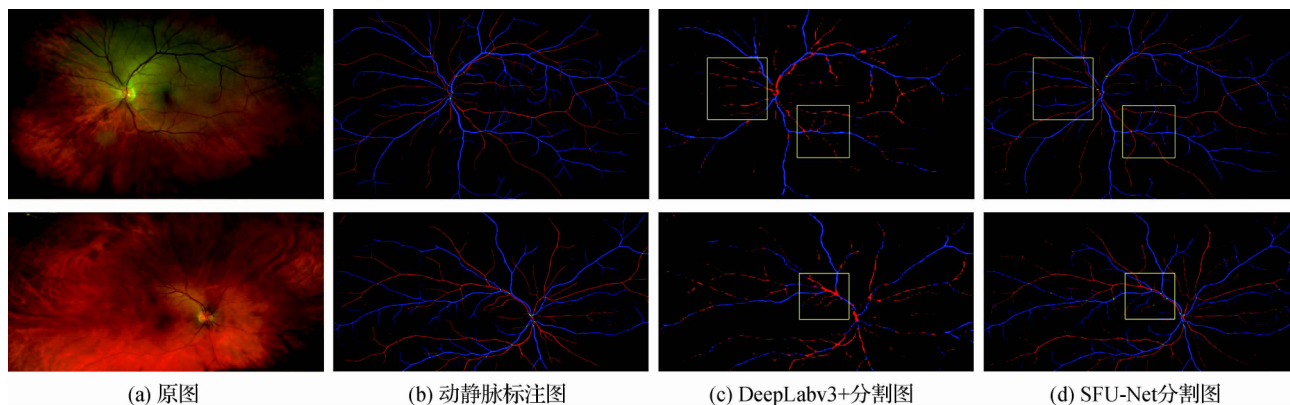


图6 WIDE数据集上动静脉分类结果图

Fig. 6 Arteriovenous classification results on WIDE dataset

((a) original images; (b) ground truth; (c) segmentation results of DeepLabv3 +; (d) segmentation results of SFU-Net)

4 结论

本文基于医学图像分割网络 U-Net ,分析眼底图像中动静脉分类任务的特点,提出了一种基于多标签学习和语义融合的动静脉分类方法 SFU-Net ,以提升分类的精度。最后,在两个公开数据集上与现有方法进行对比,通过对比实验发现,本文方法提取动脉和静脉的完整性、区分动静脉的准确率远远优于 DeepLabv3 + ,并且与现有动静脉分类方法相比获得了最高的 acc 和 b_{acc} ,验证了本文方法在眼底图像动静脉分类任务中的优越性。

但本文算法也存在局限性,主要体现在: 1) 本文使用的 focal loss 仍然是像素级的损失函数,而没有考虑到血管的结构特征,导致分割图中存在血管断裂的情况; 2) 本文实验所使用的训练集和测试集

都来自于同一个数据集,当训练集和测试集数据分布差异较大时,模型性能会受到影响。

未来的工作将主要从以下方面展开: 1) 设计基于血管结构的损失函数。考虑血管结构的先验信息,有助于减少血管断裂并提升分类的准确率。2) 域自适应学习。在实际应用中,由于采集设备、采集条件,甚至地域、种族等的差异,眼底图像本身的差异很大,因此,采用领域自适应的方法来提升模型的泛化能力是未来研究的重点。

参考文献(References)

- Abramoff M D , Garvin M K and Sonka M. 2010. Retinal imaging and image analysis. IEEE Reviews in Biomedical Engineering , 3: 169–208 [DOI: 10.1109/RBME.2010.2084567]
- Chen L C , Zhu Y K , Papandreou G , Schroff F and Adam H. 2018.

- Encoder-decoder with atrous separable convolution for semantic image segmentation//Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich: Springer: 801-818 [DOI: 10.1007/978-3-030-01234-2_49]
- Cheung C Y L, Ikram M K, Chen C and Wong T Y. 2017. Imaging retina to study dementia and stroke. *Progress in Retinal and Eye Research*, 57: 89-107 [DOI: 10.1016/j.preteyeres.2017.01.001]
- Dashtbozorg B, Mendonça A M and Campilho A. 2014. An automatic graph-based approach for artery/vein classification in retinal images. *IEEE Transactions on Image Processing*, 23(3): 1073-1083 [DOI: 10.1109/TIP.2013.2263809]
- Estrada R, Allingham M J, Mettu P S, Cousins S W, Tomasi C and Farsiu S. 2015a. Retinal artery-vein classification via topology estimation. *IEEE Transactions on Medical Imaging*, 34(12): 2518-2534 [DOI: 10.1109/TMI.2015.2443117]
- Estrada R, Tomasi C, Schmidler S C and Farsiu S. 2015b. Tree topology estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(8): 1688-1701 [DOI: 10.1109/TPAMI.2014.2382116]
- Galdran A, Meyer M, Costa P, Mendonça and Campilho A. 2019. Uncertainty-aware artery/vein classification on retinal images//2019 IEEE 16th International Symposium on Biomedical Imaging. Venice, Italy: IEEE: 556-560 [DOI: 10.1109/ISBI.2019.8759380]
- Girard F, Kavalec C and Cheriet F. 2019. Joint segmentation and classification of retinal arteries/veins from fundus images. *Artificial Intelligence in Medicine*, 94: 96-109 [DOI: 10.1016/j.artmed.2019.02.004]
- Grisan E and Ruggeri A. 2003. A divide et impera strategy for automatic classification of retinal vessels into arteries and veins//Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Cancun, Mexico: IEEE: 890-893 [DOI: 10.1109/IEMBS.2003.1279908]
- Hemelings R, Elen B, Stalmans I, Van Keer K, De Boever P and Blaschko M B. 2019. Artery-vein segmentation in fundus images using a fully convolutional network. *Computerized Medical Imaging and Graphics*, 76: #101636 [DOI: 10.1016/j.compmedimag.2019.05.004]
- Hu J, Shen L and Sun G. 2018. Squeeze-and-excitation networks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE: 7132-7141 [DOI: 10.1109/CVPR.2018.00745]
- Hu Q, Abramoff M D and Garvin M K. 2015. Automated construction of arterial and venous trees in retinal images. *Journal of Medical Imaging*, 2(4): #044001 [DOI: 10.1117/1.JMI.2.4.044001]
- Huang F, Dashtbozorg B, Tan T and ter Haar Romeny B M. 2018. Retinal artery/vein classification using genetic-search feature selection. *Computer Methods and Programs in Biomedicine*, 161: 197-207 [DOI: 10.1016/j.cmpb.2018.04.016]
- Huang G, Liu Z, Van Der Maaten L and Weinberger K Q. 2017. Dense-ly connected convolutional networks//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE: 4700-4708 [DOI: 10.1109/CVPR.2017.243]
- Ikram M K, de Jong F J, Vingerling J R, Witteman J C, Hofman A, Breteler M M B and de Jong P T V M. 2004. Are retinal arteriolar or venular diameters associated with markers for cardiovascular disorders? The rotterdam study. *Investigative Ophthalmology and Visual Science*, 45(7): 2129-2134 [DOI: 10.1167/iovs.03-4390]
- Joshi V S, Reinhardt J M, Garvin M K and Abramoff M D. 2014. Automated method for identification and artery-venous classification of vessel trees in retinal vessel networks. *PLoS One*, 9(2): e88061 [DOI: 10.1371/journal.pone.0088061]
- LeCun Y, Bengio Y and Hinton G. 2015. Deep learning. *Nature*, 521(7553): 436-444 [DOI: 10.1038/nature14539]
- Liew G and Wang J J. 2011. Retinal vascular signs: a window to the heart. *Revista Española de Cardiología (English Edition)*, 64(6): 515-521 [DOI: 10.1016/j.rec.2011.02.017]
- Ma Z Y. 2015. An Automatic Method for the Artery/Vein Classification in Retinal Images. Beijing: Beijing Institute of Technology (马志扬. 2015. 基于彩色眼底图像的视网膜血管动静脉分类研究. 北京: 北京理工大学)
- Mirsharif Q, Tajeripour F and Pourreza H. 2013. Automated characterization of blood vessels as arteries and veins in retinal images. *Computerized Medical Imaging and Graphics*, 37(7/8): 607-617 [DOI: 10.1016/j.compmedimag.2013.06.003]
- Muramatsu C, Hatanaka Y, Iwase T, Hara T and Fujita H. 2011. Automated selection of major arteries and veins for measurement of arteriolar-to-venular diameter ratio on retinal fundus images. *Computerized Medical Imaging and Graphics*, 35(6): 472-480 [DOI: 10.1016/j.compmedimag.2011.03.002]
- Niemeijer M, Xu X Y, Dumitrescu A V, Gupta P, Van Ginneken B, Folk J C and Abramoff M D. 2011. Automated measurement of the arteriolar-to-venular width ratio in digital color fundus photographs. *IEEE Transactions on Medical Imaging*, 30(11): 1941-1950 [DOI: 10.1109/TMI.2011.2159619]
- Pellegrini E, Robertson G, MacGillivray T, van Hemert J, Houston G and Trucco E. 2018. A graph cut approach to artery/vein classification in ultra-widefield scanning laser ophthalmoscopy. *IEEE Transactions on Medical Imaging*, 37(2): 516-526 [DOI: 10.1109/TMI.2017.2762963]
- Ronneberger O, Fischer P and Brox T. 2015. U-net: convolutional networks for biomedical image segmentation//Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich, Germany: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]
- Rothaus K, Jiang X Y and Rhiem P. 2009. Separation of the retinal vascular graph in arteries and veins based upon structural knowledge. *Image and Vision Computing*, 27(7): 864-875 [DOI: 10.1016/j.

imavis. 2008. 02. 013]

Srinidhi C L , Aparna P and Rajan J. 2019. Automated method for retinal artery/vein separation via graph search metaheuristic approach. IEEE Transactions on Image Processing , 28(6) : 2705-2718 [DOI: 10. 1109/TIP. 2018. 2889534]

Staal J , Abràmoff M D , Niemeijer M , Viergever M A and Van Ginneken B. 2004. Ridge-based vessel segmentation in color images of the retina. IEEE Transactions on Medical Imaging , 23(4) : 501-509 [DOI: 10. 1109/tmi. 2004. 825627]

Vázquez S G , Cancela B , Barreira N , Penedo M G , Rodríguez-Blanco M , Seijo M P , de Tuero G C , Barceló M A and Saez M. 2013. Improving retinal artery and vein classification by means of a minimal path approach. Machine Vision and Applications , 24(5) : 919-930 [DOI: 10. 1007/s00138-012-0442-4]

Welikala R A , Foster P J , Whincup P H , Rudnicka A R , Owen C G , Strachan D P , Barman S A and the UK Biobank Eye and Vision Consortium. 2017. Automated arteriole and venule classification using deep learning for retinal images from the UK Biobank cohort. Computers in Biology and Medicine , 90: 23-32 [DOI: 10. 1016/j. compbiomed. 2017. 09. 005]

Wong T Y , Klein R , Klein B E K , Tielsch J M , Hubbard L and Nieto F J. 2001. Retinal microvascular abnormalities and their relationship with hypertension , cardiovascular disease , and mortality. Survey of Ophthalmology , 46(1) : 59-80 [DOI: 10. 1016/S0039-6257(01) 00234-X]

Xu X Y , Ding W X , Abràmoff M D and Cao R F. 2017. An improved arteriovenous classification method for the early diagnostics of various diseases in retinal image. Computer Methods and Programs in Biomedicine , 141: 3-9. [DOI: 10. 1016/j. cmpb. 2017. 01. 007]

Xu X Y , Wang R D , Lv P L , Gao B , Li C , Tian Z Q , Tan T and Xu F. 2018. Simultaneous arteriole and venule segmentation with domain-specific loss function on a new public database. Biomedical Optics Express , 9(7) : 3153-3166 [DOI: 10. 1364/BOE. 9. 003153]

Xue L Y , Cao X R , Lin J W , Zheng S H and Yu L. 2017. Artery/vein automatic classification in retinal images and vessel diameter Measurement. Chinese Journal of Scientific Instrument , 38(9) : 2307-2316 (薛岚燕 , 曹新容 , 林嘉雯 , 郑绍华 , 余轮. 2017. 动静脉血管自动分类方法及其管径测量. 仪器仪表学报 , 38(9) : 2307-2316) [DOI: 10. 3969/j. issn. 0254-3087. 2017. 09. 027]

Zhao Y T , Xie J Y , Zhang H Z , Zheng Y L , Zhao Y F , Qi H , Zhao Y C , Su P , Liu J and Liu Y H. 2020. Retinal vascular network topology reconstruction and artery/vein classification via dominant set clustering. IEEE Transactions on Medical Imaging , 39 (2) : 341-356 [DOI: 10. 1109/TMI. 2019. 2926492]

作者简介



高颖琪 ,1995 年生 ,女 ,硕士研究生 ,主要研究方向为深度学习和医疗影像分析。

E-mail: gaoyingqi@mail. nankai. edu. cn



康宏 通信作者 ,男 ,讲师 ,主要研究方向为大数据处理、机器学习。

E-mail: kanghong@nankai. edu. cn

郭松 男 ,博士研究生 主要研究方向为深度学习和医疗影像分析。E-mail: guomugong@hotmail. com

李宁 男 ,硕士研究生 主要研究方向为深度学习和医疗影像分析。E-mail: kuangge_1994@163. com

王恺 男 ,副教授 主要研究方向为图像处理、深度学习。E-mail: wangk@nankai. edu. cn

李涛 男 ,教授 主要研究方向为异构计算、机器学习、智能物联网。E-mail: litao@nankai. edu. cn