

# 基于多任务学习的眼底图像红色病变点分割\*

郭松<sup>1</sup>, 李涛<sup>2,3</sup>, 李宁<sup>1</sup>, 康宏<sup>1,4</sup>, 张玉军<sup>5</sup>, 王恺<sup>1</sup>



<sup>1</sup>(南开大学 计算机学院, 天津 300350)

<sup>2</sup>(天津市网络与数据安全重点实验室(南开大学), 天津 300350)

<sup>3</sup>(计算机体系结构国家重点实验室(中国科学院), 北京 100190)

<sup>4</sup>(北京上工医信科技有限公司, 北京 100176)

<sup>5</sup>(中国科学院 计算技术研究所, 北京 100190)

通讯作者: 王恺, E-mail: wangk@nankai.edu.cn

**摘要:** 糖尿病性视网膜病变(糖网病)是导致成年人视觉损失的主要因素之一.早期的眼底筛查可以显著降低这种视觉损失的可能性.彩色眼底图像由于具有采集便利、对人体无伤害等特点,常被用于大规模的眼底筛查工作.对眼底图像中的红色病变点而言,微动脉瘤是轻度非增殖性糖网病的主要标志,出血点与中度及重度非增殖性糖网病的诊断有关,因此,眼底图像中出血点和微动脉瘤的准确分割对糖网病分级诊断具有重要参考价值.提出一种基于多任务学习的分割模型 Red-Seg 来对出血点和微动脉瘤进行分割.该网络包含两个分支,每个分支处理一种病变点.设计了一种两阶段训练算法,并且两个阶段使用不同的损失函数:第 1 阶段使用改进的 Top-k 带权交叉熵损失函数,将模型训练集中在难样本上;第 2 阶段将最小化假阳性和假阴性作为 Red-Seg 模型训练的优化目标,进一步减少病变点误分.最后,在 IDRiD 数据集上进行模型验证,并与其他病变点分割方法进行对比.实验结果表明,在应用 Red-Seg 模型进行微动脉瘤和出血点红色病变点分割时,两阶段训练算法可以显著减少病变点误分情况,尤其是出血点分割的准确率和召回率都提高 2.8%.同时,与 HED、FCRN、DeepLabv3+和 L-Seg 等图像级分割模型相比,Red-Seg 模型在微动脉瘤分割上获得了更好的 AUC<sub>PR</sub>.

**关键词:** 眼底图像;糖网病;微动脉瘤分割;出血点分割;多任务学习

**中图法分类号:** TP391

中文引用格式: 郭松,李涛,李宁,康宏,张玉军,王恺.基于多任务学习的眼底图像红色病变点分割.软件学报,2021,32(11): 3646–3658. <http://www.jos.org.cn/1000-9825/6038.htm>

英文引用格式: Guo S, Li T, Li N, Kang H, Zhang YJ, Wang K. Red lesion segmentation of fundus image with multi-task learning. Ruan Jian Xue Bao/Journal of Software, 2021,32(11):3646–3658 (in Chinese). <http://www.jos.org.cn/1000-9825/6038.htm>

## Red Lesion Segmentation of Fundus Image with Multi-task Learning

GUO Song<sup>1</sup>, LI Tao<sup>2,3</sup>, LI Ning<sup>1</sup>, KANG Hong<sup>1,4</sup>, ZHANG Yu-Jun<sup>5</sup>, WANG Kai<sup>1</sup>

<sup>1</sup>(College of Computer Science, Nankai University, Tianjin 300350, China)

<sup>2</sup>(Tianjin Key Laboratory of Network and Data Science Technology (Nankai University), Tianjin 300350, China)

<sup>3</sup>(State Key Laboratory of Computer Architecture (Chinese Academy of Sciences), Beijing 100190, China)

\* 基金项目: 国家自然科学基金(61872200); 国家重点研发计划(2016YFC0400709, 2018YFB2100300); 天津市自然科学基金(18YFYZCG00060, 19JCZDJC31600); 天津市教学成果奖重点培育项目(PYGJ-018)

Foundation item: National Natural Science Foundation of China (61872200); National Key Research and Development Program of China (2016YFC0400709, 2018YFB2100300); Natural Science Foundation of Tianjin Municipality (18YFYZCG00060, 19JCZDJC31600); Key Cultivation Project of Education Achievement Award of Tianjin Municipality (PYGJ-018)

收稿时间: 2019-07-18; 修改时间: 2019-11-04; 采用时间: 2020-03-27

<sup>4</sup>(Beijing Shanggong Medical Technology Co. Ltd., Beijing 100176, China)

<sup>5</sup>(Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China)

**Abstract:** Diabetic retinopathy (DR) is the leading cause of vision loss for adult individuals, and early fundus screening can significantly reduce this visual loss. Color fundus image is often used in large-scale fundus screening due to the acquisition convenience and its human-harmless. As a kind of red lesions in fundus images, the appearance of microaneurysms is the main marker of mild non-proliferative DR, and hemorrhage, as another kind of red lesions, is related to moderate and severe non-proliferative DR. So that red lesions in fundus images are important indicators for the screening of DR. This study proposes a multi-task network, named Red-Seg, for red lesion segmentation. The network contains two individual branches, each is used for one kind of lesion segmentation. Meantime, a two-stage training algorithm is presented where different loss functions are used in different stages. In the first stage, modified Top- $k$  balanced cross-entropy loss is used to push the network focuses on hard-to-classify samples. And, in the second stage, false positive and false negative are integrated as loss function into training to reduce misclassification further. At last, extensive experiments are employed on the IDRiD dataset, and the lesion segmentation results are compared with other methods. Experimental results show that proposed two-stage training algorithm can lead to much higher precision and recall, which means this method can reduce misclassification in some certain. Specifically for hemorrhage segmentation, both recall and precision increased by at least 2.8%. Meanwhile, compared with other image-level lesion segmentation models, such as HED, FCRN, DeepLabv3+, and L-Seg, Red-Seg achieves much higher AUC\_PR on microaneurysm segmentation.

**Key words:** fundus image; diabetic retinopathy; microaneurysms segmentation; hemorrhage segmentation; multi-task learning

糖尿病性视网膜病变(简称糖尿病,diabetic retinopathy,简称 DR)是糖尿病引起的眼部疾病,是致盲的主要原因之一。据报道,在美国、欧洲和亚洲,大约 1/3 的糖尿病患者患有不同程度的糖尿病<sup>[1]</sup>。研究表明,通过早期筛查,可以显著降低糖尿病的致盲率,而糖尿病筛查的主要方法是分析彩色眼底图像。按照国际眼科中心定义的糖尿病分级标准,与糖尿病相关的眼底病变点主要包括软渗、硬渗、出血点、微动脉瘤、静脉串珠等<sup>[1]</sup>,其中,出血点和微动脉瘤的颜色偏红(如图 1 所示),统称为红色病变点。轻度非增殖性糖尿病在眼底图像上的表现是只存在微动脉瘤(microaneurysm,简称 MA),中度及重度非增殖性糖尿病的表现之一是眼底存在出血点(hemorrhage,简称 HE)。因此,眼底红色病变点的识别和检测对糖尿病的分级诊断具有重要的指示意义。但由于病变点形态大小不一,导致人工进行病变点分析耗时耗力,并且与大规模待检人员相比,眼科医生严重匮乏,这使得大规模的糖尿病筛查难以有效开展。因此,开发一种自动化的眼底图像病变点分析工具非常必要。

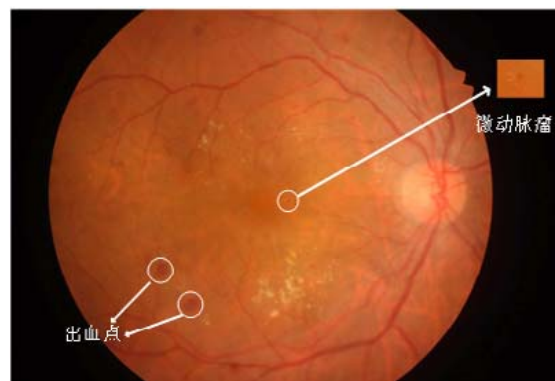


Fig.1 A fundus image from the IDRiD dataset

图 1 IDRiD 数据集中的眼底图像

目前,大量工作表明,深度学习方法在眼底图像分析领域获得了比传统方法更优的性能,比如在糖尿病筛查、眼底血管分割、眼底视杯视盘分割、病变点检测等领域<sup>[2-10]</sup>。Google 在 2016 年提出了一种基于深度学习模型的糖尿病筛查算法,首先通过多位职业眼科医生对将近 13 万张眼底图像进行病变等级标注,然后采用 inception-v3 网络对其进行训练和测试,在由 1 万多张图像组成的测试集上获得了 90.3%的敏感性和 98.1%的特

异性<sup>[2]</sup>.Guo 等人提出的 BTS-DSN 血管分割网络,在 DRIVE 数据集上,其 AUC 结果达到了 0.980 6,血管分割的准确率超过了第 2 组标注专家<sup>[4]</sup>.Fu 等人提出的 M-Net 对眼底视盘和视杯进行分割,在 ORIGA 数据集上达到了 92.9%的视盘分割重叠率和 77.0%的视杯分割重叠率,在青光眼自动分类上也获得了较好结果<sup>[7]</sup>.但目前已有的眼底病变点分割模型大多只能分割一种病变点,比如 HED<sup>[11]</sup>、U-Net<sup>[12]</sup>、FCRN<sup>[13]</sup>等,这就意味着完成多种病变点的分割需要更大的计算量;同时,由于病变点尺寸不一,特征不明显,现有模型存在较多的误识别情况,尤其在微动脉瘤(如图 1 所示)分割上.

为了解决上述问题,本文提出了一种基于多任务学习的多分枝分割模型 Red-Seg 来对眼底出血点和微动脉瘤进行分割.Red-Seg 共包括两个任务:出血点分割和微动脉瘤分割.为了实现这种多任务学习,模型首先通过若干个共享卷积层对眼底图像进行特征提取;随后,通过两个独立的分支得到两种病变点的分割结果.同时,针对目前眼底病变点分割模型存在的误识别问题,本文提出一种两阶段训练算法来减少病变点误分的情况.其中,第 1 阶段使用改进的 Top-k 带权交叉熵损失<sup>[9]</sup>,使保留的负样本个数随着网络训练逐渐减少,从而迫使网络去学习能区分这些难分样本的特征;在第 2 阶段,将最小化假阳性和假阴性作为模型训练优化目标,进一步减少误分情况.最后,本文在公开数据集 IDRiD<sup>[14]</sup>上验证了 Red-Seg 的分割效果,并且与 IDRiD 竞赛队伍和其他方法进行了对比.在 IDRiD 排行榜上,Red-Seg 在微动脉瘤和出血点分割上都排名第三.但是 IDRiD 排行榜前三名队伍的模型属于图像块级(patch-level)分割模型,需要多次前向传播才可得到一张图像的分割结果;而 Red-Seg 属于图像级(image-level)分割模型,只需一次前向传播便可得到整幅图像的分割结果.因此,Red-Seg 在分割效率上具有一定优势.与 HED,FCRN 和 Deeplabv3+相比,Red-Seg 在微动脉瘤和出血点分割上获得了更高的 AUC\_PR(PR 曲线下面积).与 L-Seg<sup>[8]</sup>相比,Red-Seg 在微动脉瘤分割上 AUC\_PR 指标高出 2%;与 Clément 等人的方法<sup>[15]</sup>相比,出血点的 AUC\_PR 指标高出 3.9%.本文的创新点如下:

- 1) 本文基于 VGGNet 主干网络提出了一种多任务的网络模型 Red-Seg,它利用不同分支分别处理微动脉瘤和出血点分割任务,实现单个网络同时分割眼底图像中的两种红色病变点.
- 2) 本文提出一种 Red-Seg 模型的两阶段训练算法:第 1 阶段使用改进的 Top-k 带权交叉熵损失;第 2 阶段将最小化假阳性和假阴性作为模型训练的优化目标,有利于减少病变点误分.
- 3) 本文在 IDRiD 数据集上对 Red-Seg 模型和训练算法进行验证,结果表明了两阶段训练算法的有效性,进而使得 Red-Seg 与其他图像级分割方法相比获得了更好的 AUC\_PR.

## 1 相关工作

对于眼底图像病变点分割,基于机器学习的方法主要利用阈值分割法或者模板匹配法确定候选区域,随后使用分类器对候选区域进行分类.传统的眼底图像分析方法需要人工提取特征,存在鲁棒性不强、泛化性差等缺点.基于深度学习的方法主要包括分类网络和分割网络,

- 分类网络对每个像素单独进行预测,网络的输入是待预测像素周围的一图像块,输出为对该图像块的类别预测.这种方法需要对眼底图像中所有像素都单独进行预测,进而得到整幅图像的分割结果.比如,Tan 等人设计了一种分类网络,实现了对微动脉瘤、出血点和硬渗这 3 种病变点的自动化检测<sup>[16]</sup>.这种方法的问题在于运行效率低、检测时间长,并且检测出的病变点区域明显偏大.
- 区别于分类网络,基于分割网络的病变点分割方法根据输入图像方式的不同可进一步划分为图像块级分割网络和图像级分割网络.
  - 图像块级分割网络需要将一张眼底图像分成若干个图像块,对每个图像块独立进行病变点分割,最后将所有的分割结果组合起来形成一张眼底图像的分割结果.比如,Chudzik 等人使用图像块级的全卷积神经网络对每个图像块中的微动脉瘤进行检测,并在 3 个微动脉瘤数据集上对模型进行验证.实验结果表明,该方法获得了比传统方法更高的灵敏度<sup>[17]</sup>.图像块级分割网络的一个缺点是需要多次前向传播才可得到整幅图像的病变点分割结果,与图像级分割网络相比,其运行效率较低.

- 图像级分割网络的输入是一张眼底图像,输出为对应的分割结果,只需一次前向传播.比如,Mo 等人设计了一种图像级的分割网络模型 FCRN 来对眼底硬渗点进行分割,取得了不错的分割效果<sup>[13]</sup>.Guo 等人设计了一种端到端的多病变点分割网络 L-Seg,对于一张眼底图像,可以同时得到 4 种病变点的分割结果,并且该方法引入了一种新的损失函数来减少误分情况<sup>[8]</sup>.Yang 等人提出了一种两阶段的深度卷积神经网络模型,可检测出血点、渗出点等眼底病变,并对糖尿病进行分级预测<sup>[18]</sup>.本文设计的分割网络即属于图像级分割网络.

在训练病变点分割网络方面,常见的基于交叉熵的损失函数包括带权交叉熵<sup>[11]</sup>、Focal Loss<sup>[19]</sup>、Top- $k$  带权交叉熵<sup>[9]</sup>等.除此之外,还存在一些直接对评价指标进行优化的损失函数,如 Dice Loss<sup>[12]</sup>、IoU Loss<sup>[20]</sup>等.带权交叉熵损失由于未考虑难分样本和易分样本,在微动脉瘤和出血点分割这种易将背景误分成病变的情况下,会存在严重的误分问题.Dice Loss 和 IoU Loss 存在的问题是它们的梯度不稳定,有时会出现难以收敛的情况.现有的分割网络训练算法大致可以分为两类:一类是训练过程中只使用单一的损失函数;另一类是使用混合的损失函数,如 Dice Loss+Focal Loss<sup>[21]</sup>等.单一的损失函数不能结合其他损失函数的优点,而混合的损失函数需要考虑二者之间的平衡.本文提出了一种两阶段的训练算法,每个阶段使用不同的损失函数,第 2 阶段在第 1 阶段训练的基础之上再进行训练.这种方法既发挥了两种损失函数的优点,也无需像混合的损失函数那样考虑二者之间的平衡.

## 2 分割模型及训练算法

### 2.1 红色病变点分割模型

分割模型通常包括编码器和解码器<sup>[22]</sup>两部分:编码器进行特征学习,解码器恢复特征图的分辨率并得到分割结果.本文提出了一种多分支的网络模型 Red-Seg,该模型也遵循这种编码器+解码器的结构,其编码器结构使用 VGGNet<sup>[23]</sup>,共包括 5 组卷积,每组卷积都由若干个卷积操作和非线性激活操作组成.VGGNet 中,从 conv2 到 conv5 特征图的分辨率依次降低为输入图像的 1/2、1/4、1/8 和 1/16.解码器结构中包含两个分支:一个分支用来进行微动脉瘤的分割,另一个分支用来进行出血点的分割.如图 2 所示为本文提出的红色病变点分割网络结构图.对于微动脉瘤分割,因为微动脉瘤偏小,所以网络融合了 conv2 到 conv4 的特征.而出血点大小不一,存在大块的出血点,也存在小如微动脉瘤的出血点,因此出血点分割分支融合了 conv3 到 conv5 的特征.因为 conv1 的感受野过小,包括太多噪音(如后文图 7 所示),所以在两个分支中都未融合 conv1 的特征.另外,为了进一步提高分割效果,网络中使用了与 HED<sup>[11]</sup>和 GoogLeNet<sup>[24]</sup>类似的深度监督方法,除了两个分支各自的损失函数之外,每一个分支还存在 3 个监督损失,分别对应不同尺度下的分割结果.

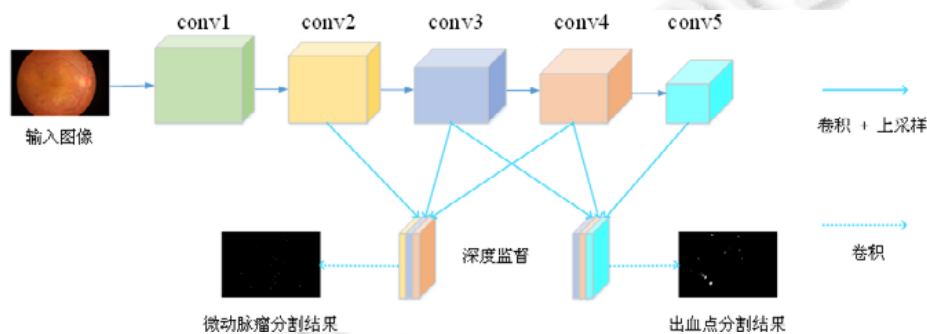


Fig.2 Architecture of Red-Seg

图 2 Red-Seg 模型结构图

解码器中微动脉瘤分割分支如图 3 所示.3 组卷积 conv2~conv4 各包含若干个卷积操作,在每组卷积的最后一个卷积操作之后接一个  $1 \times 1$  的卷积操作,分别得到 3 种不同尺度的特征图,3 个特征图的尺寸分别为  $H/2 \times W/2$ ,

$H/4 \times W/4$  和  $H/8 \times W/8$ . 然后使用双线性插值对 3 个特征图进行上采样, 分别上采样 2 倍、4 倍和 8 倍. 上采样之后的特征图与输入眼底图像拥有相同的分辨率. 训练时, 每个特征图接一个监督损失, 因此这 3 个特征图可以理解为不同尺度下的微动脉瘤分割结果. 随后对 3 个特征图进行拼接操作, 得到一个厚度为 3 的特征图, 再次使用  $1 \times 1$  的卷积操作, 得到一个融合之后的厚度为 1 的特征图. 训练时, 此特征图后接一个损失函数. 在测试时, 对融合后的特征图进行 sigmoid 运算, 即得到微动脉瘤的分割概率图. 因此, 分割分支的输出是多尺度融合的结果. 从图 3 中可以看出, 微动脉瘤解码分支一共包括 4 个损失, 其中 3 个为监督损失.

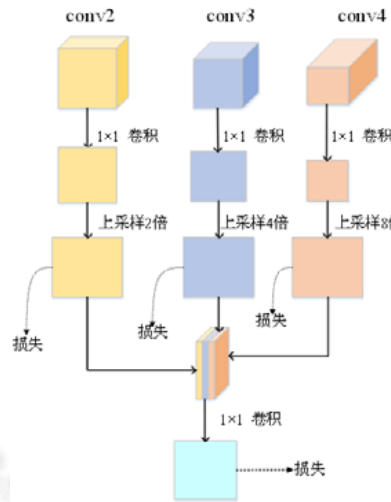


Fig.3 Microaneurysms segmentation branch

图 3 微动脉瘤分割分支

解码器中出血点分割分支如图 4 所示. 与微动脉瘤分割分支不同的是, 出血点分割分支从 conv3~conv5 得到 3 个特征图, 随后进行上采样操作. 另外, 从 conv5 卷积得到的特征图需要上采样 16 倍恢复到原图的分辨率. 随后对 3 个特征图进行拼接操作, 得到一个厚度为 3 的特征图, 再次使用  $1 \times 1$  的卷积操作, 得到一个融合之后的厚度为 1 的特征图. 随后的操作与微动脉瘤分割分支中的一致. 此分支中同样存在 4 个损失函数, 其中 3 个为监督损失 (如图 4 所示).

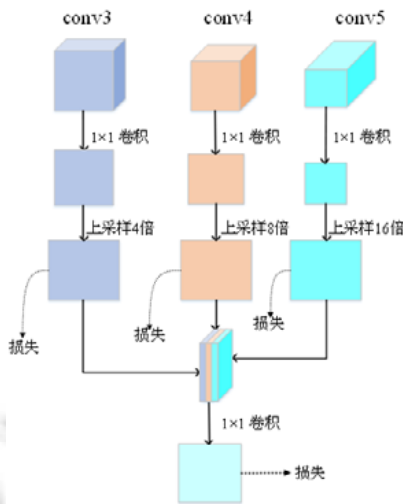


Fig.4 Hemorrhage segmentation branch

图 4 出血点分割分支

最后,网络参数见表 1.为了方便,表中只列出了卷积层的参数,包括卷积核的大小和卷积核个数,省略了池化、ReLU 等信息.

**Table 1** Detailed configurations of Red-Seg (only list settings of convolutional layers)

**表 1** Red-Seg 模型详细配置表(只列出了卷积层设置参数)

|                 | conv1               | conv2                | conv3                | conv4                | conv5                |
|-----------------|---------------------|----------------------|----------------------|----------------------|----------------------|
| 主网络<br>(VGGNet) | conv1_1<br>(3×3-64) | conv2_1<br>(3×3-128) | conv3_1<br>(3×3-256) | conv4_1<br>(3×3-512) | conv5_1<br>(3×3-512) |
|                 | conv1_2<br>(3×3-64) | conv2_2<br>(3×3-128) | conv3_2<br>(3×3-256) | conv4_2<br>(3×3-512) | conv5_2<br>(3×3-512) |
|                 | -                   | -                    | conv3_3<br>(3×3-256) | conv4_3<br>(3×3-512) | conv5_3<br>(3×3-512) |
| 微动脉瘤分割分支        | -                   | conv<br>(1×1-1)      | conv<br>(1×1-1)      | conv<br>(1×1-1)      | -                    |
| 出血点分割分支         | -                   | -                    | conv<br>(1×1-1)      | conv<br>(1×1-1)      | conv<br>(1×1-1)      |

可以看出,主网络使用 VGGNet 的前 5 组卷积,共包括 13 个卷积层,每个卷积层都使用 3×3 的卷积核;而且随着网络深度的增加,卷积核的个数从 conv1 中的 64 增加到 conv5 中的 512.在微动脉瘤和出血点分割分支中,每个卷积层中卷积核的大小为 1×1,且只包括一个卷积核.

## 2.2 损失函数及训练算法

交叉熵损失函数通常被用来训练分割网络,而普通的带权交叉熵损失并未考虑难分与易分样本的问题,因此,在这种病变点和背景点极易混淆的场景下,眼底病变点分割极易出现严重的误识别问题.本文使用改进的 Top- $k$  带权交叉熵损失函数<sup>[9]</sup>,通过不断丢弃易分样本,使网络更多地集中在难分样本的学习上,从而提高网络的判别能力.同时,为了进一步减少误分,本文直接将最小化误分(假阴性+假阳性)作为优化目标去训练 Red-Seg 模型,提出了一种两阶段训练算法.

### 2.2.1 第 1 阶段的损失函数

Top- $k$  带权交叉熵损失函数<sup>[9]</sup>定义如下:

$$L_1(p, y) = -\beta \sum_{y_i=1} \log p_i - (1-\beta) \sum_{i \in y_{top}} \log(1-p_i),$$

其中,  $p$  为激活值,由 *sigmoid* 函数计算得来;  $y$  为标注;  $\beta$  为权重因子. Top- $k$  损失需要将负样本按照其损失降序排序,然后选择损失最大的一部分负样本进行优化.形式化如下:

$$y_{top} = \{t_1, t_2, \dots, t_\lambda\},$$

其中,  $t_i$  为损失第  $i$  大的负样本编号,  $\lambda$  为选中的负样本个数. 负样本确定之后,可用如下公式定义权重因子:

$$\beta = \frac{\lambda}{\lambda + N_+},$$

其中,  $N_+$  为正样本个数. 可见,权重因子为负样本占有所有记入损失样本的比例.

原始的 Top- $k$  损失使用的是固定的  $\lambda$  值,每次网络迭代都选择固定数量的负样本,这就使得随着网络的训练,选中的负样本中易分样本的比例逐渐增加.为了解决这个问题,本文提出了一种随着训练迭代  $\lambda$  线性衰减的调度策略.定义如下:

$$\lambda(r) = \lambda_0 - \frac{\lambda_0}{M/S} \left\lfloor \frac{r}{S} \right\rfloor,$$

其中,  $\lambda_0$  为初始值,  $M$  为网络最大迭代次数,  $r$  为当前迭代次数,  $S$  为衰减周期. 当  $r \in [1, S-1]$  时,  $\lambda(r) = \lambda_0$ ; 当  $r \in [S, 2S-1]$

时,  $\lambda(r) = \lambda_0 \left(1 - \frac{1}{M/S}\right)$ .

### 2.2.2 第 2 阶段的损失函数

为了将误分情况集成到损失函数中,首先定义假阴性(false negative,简称 FN)和假阳性(false positive,简称

FP)如下:

$$FN = \sum_{i=1}^N y_i(1-p_i)^2, FP = \sum_{i=1}^N p_i^2(1-y_i),$$

其中,  $N$  为总样本个数,也就是一张眼底图像中总像素个数;  $y_i$  为第  $i$  个样本(像素)的标注,取值为 0 或者 1;  $p_i$  为第  $i$  个样本的激活概率值,由 *sigmoid* 函数计算而来.

第 2 阶段的损失函数定义如下:

$$L_2(p, y) = FN + FP = \sum_{i=1}^N y_i(1-p_i)^2 + \sum_{i=1}^N p_i^2(1-y_i) = \sum_{y_i=1} (1-p_i)^2 + \sum_{y_i=0} p_i^2 = \sum_{i=1}^N (y_i - p_i)^2.$$

从公式中可以看出,第 2 阶段的损失函数  $L_2$  经过化简之后就是均方差损失函数,也就是将分类问题转化为回归问题,使得对于每一个样本  $i$ ,其激活概率值要接近真实的标注.

### 2.2.3 训练算法

首先使用 VGGNet 在 ImageNet<sup>[25]</sup>上预训练的权重来初始化模型  $M$  的参数  $\theta$ .第 1 阶段,从训练数据  $D$  中随机选择一个样本,然后网络前向传播得到对该样本的分割概率图  $p$ ,随后使用  $L_1$  损失函数计算当前样本的损失值,并计算梯度  $g$ ,最后使用随机梯度下降算法更新参数  $\theta$ .第 2 阶段与第 1 阶段类似,区别是使用  $L_2$  损失函数训练网络,第 2 阶段训练结束之后,得到分割模型  $M$ .如算法 1 所示为训练算法伪代码,为了方便表示,伪代码中并未考虑权重衰减和动量.

**算法 1.** 两阶段训练算法.

(a) 第 1 阶段

输入:训练数据  $D=\{(x^1, y^1), \dots, (x^N, y^N)\}$ ,超参数(学习率  $\eta_1, \eta_2$ ,最大迭代次数  $Max1, Max2$  等).

输出:分割模型  $M$ .

initialize  $\theta$  from VGGNet;

set  $r=1$ ;

**repeat**

  sample  $(x^i, y^i)$  from  $D$ ;

$p=f(x^i|\theta)$ ;

$Loss1=L_1(p, y^i|\theta)$ ;

$g = \frac{\partial L_1(p, y^i | \theta)}{\partial \theta}$ ;

$\theta=\theta-\eta_1g$ ;

$r=r+1$ ;

**until** ( $r>Max1$ )

(b) 第 2 阶段

set  $r=1$ ;

**repeat**

  sample  $(x^i, y^i)$  from  $D$ ;

$p=f(x^i|\theta)$ ;

$Loss2=L_2(p, y^i|\theta)$ ;

$g = \frac{\partial L_2(p, y^i | \theta)}{\partial \theta}$ ;

$\theta=\theta-\eta_2g$ ;

$r=r+1$ ;

**until** ( $r>Max2$ )

**return**  $\theta$ ;



### 3 实验及分析

本文在 Caffe<sup>[26]</sup> 框架下实现网络模型,使用 NVIDIA GTX 1080Ti 显卡加速网络训练.另外,实验中使用的操作系统为 Ubuntu 16.04,CUDA 和 CUDNN 的版本分别为 8.0 和 5.1.下面对数据集、网络中具体的参数设置、评价指标和实验结果进行具体介绍.

#### 3.1 数据集

IDRiD<sup>[14]</sup>是在国际会议 ISBI 2018 举办的视网膜眼底图像分级和分割挑战赛中提出的,提供了眼底 4 种病变点的像素级标注,包括微动脉瘤、出血点、硬渗和软渗.该数据集共包括 81 张眼底图像,其中,训练集有 54 张,测试集有 27 张,图像的分辨率为 4288×2848(宽度×高度).另外,由于原始眼底图像的分辨率太高无法直接放入显存中,文中先将图像和对应病变点标注下采样 4 倍到 1072×712.随后使用旋转和镜像操作对训练集进行扩充,扩充之后的训练集共包括 324 张眼底图像.

#### 3.2 参数设置

训练使用 ImageNet 上训练好的 VGGNet 初始化分割网络参数,使用随机梯度下降算法优化损失函数.动量项设置为 0.9,权重衰减项设置为 0.000 5.在训练的第 1 阶段,使用固定的学习率调度策略,学习率初始化为 1e-8,训练一共迭代 97 200 次(300 轮).Top-k 带权交叉熵损失中, $\lambda_0$  初始化为 0.14,S 设置为 10 000.在训练的第 2 阶段,使用第 1 阶段训练的网络初始化权重,由于第 1 阶段已经学习到了稳定的病变点特征,因此第 2 阶段的学习率设置为第 1 阶段的 1/10(1e-9),而且总迭代次数设置为 6 480 次(20 轮),避免对第 1 阶段学习到的特征进行过度调整.

#### 3.3 评价指标

本文使用准确率 P、召回率 R 和 AUC\_PR(P-R 曲线下面积)这 3 个指标来评价病变点的分割效果.

$$P = \frac{TP}{TP + FP}, R = \frac{TP}{TP + FN},$$

其中,TP 为正确分类的病变点像素个数,FP 为背景像素误分成病变点的个数,FN 为病变点误分成背景点的个数.召回率(查全率)反映的是在所有的病变点中正确找到的病变点的比例,准确率(查准率)反映的是在模型找到的所有病变点中正确找到的病变点的比例.从公式中可以看出,准确率和召回率只能评估单一阈值下的分割效果.因此,本文使用 AUC\_PR 来综合考虑多个阈值下的准确率和召回率.

#### 3.4 实验结果

模型在微动脉瘤和出血点分割上的指标见表 2,可以看出,仅仅使用第 1 阶段损失或者第 2 阶段损失均无法取得较好的分割结果,尤其是仅仅使用第 2 阶段损失,会由于样本不均衡造成模型无法训练;但是,当使用两阶段训练算法时,病变点分割结果的准确率和召回率都得到了提升.这说明了第 2 阶段损失需要在第 1 阶段训练的权重基础上进行学习,第 1 阶段和第 2 阶段是缺一不可的.同时,从表 2 还可以看出,第 2 阶段训练进一步减少了误分情况,尤其是在出血点分割上,召回率提高了将近 3%,准确率提高了 2.8%.从图 1 中可以看出,分割微动脉瘤的难度要大于出血点.尽管如此,两阶段的训练算法也使得微动脉瘤的 AUC\_PR 提高了 1.7%,准确率提高了 1.3%.

**Table 2** Comparison of segmentation performance of microaneurysms and hemorrhages under different training methods

**表 2** 不同训练方法下微动脉瘤和出血点分割性能对比

| 方法          | 微动脉瘤           |                |                | 出血点            |                |                |
|-------------|----------------|----------------|----------------|----------------|----------------|----------------|
|             | 准确率            | 召回率            | AUC_PR         | 准确率            | 召回率            | AUC_PR         |
| 一阶段训练       | 0.487 5        | 0.492 6        | 0.466 9        | 0.598 7        | 0.595 5        | <b>0.591 6</b> |
| 只使用第 2 阶段损失 | -              | -              | -              | -              | -              | -              |
| 两阶段训练       | <b>0.500 9</b> | <b>0.499 3</b> | <b>0.483 9</b> | <b>0.626 9</b> | <b>0.625 1</b> | 0.569 5        |



从图5可以看到,因为第1阶段的损失函数会丢弃易分样本,所以图5(c)和图5(e)中存在低概率值的背景点.但是经过第2阶段的训练之后,这些背景点的概率值被抑制,这在图5(d)中尤为明显.图5中两阶段分割效果图的变化与表2中准确率和召回率的提高是相对应的.最后,从图6可以看出网络已经收敛.

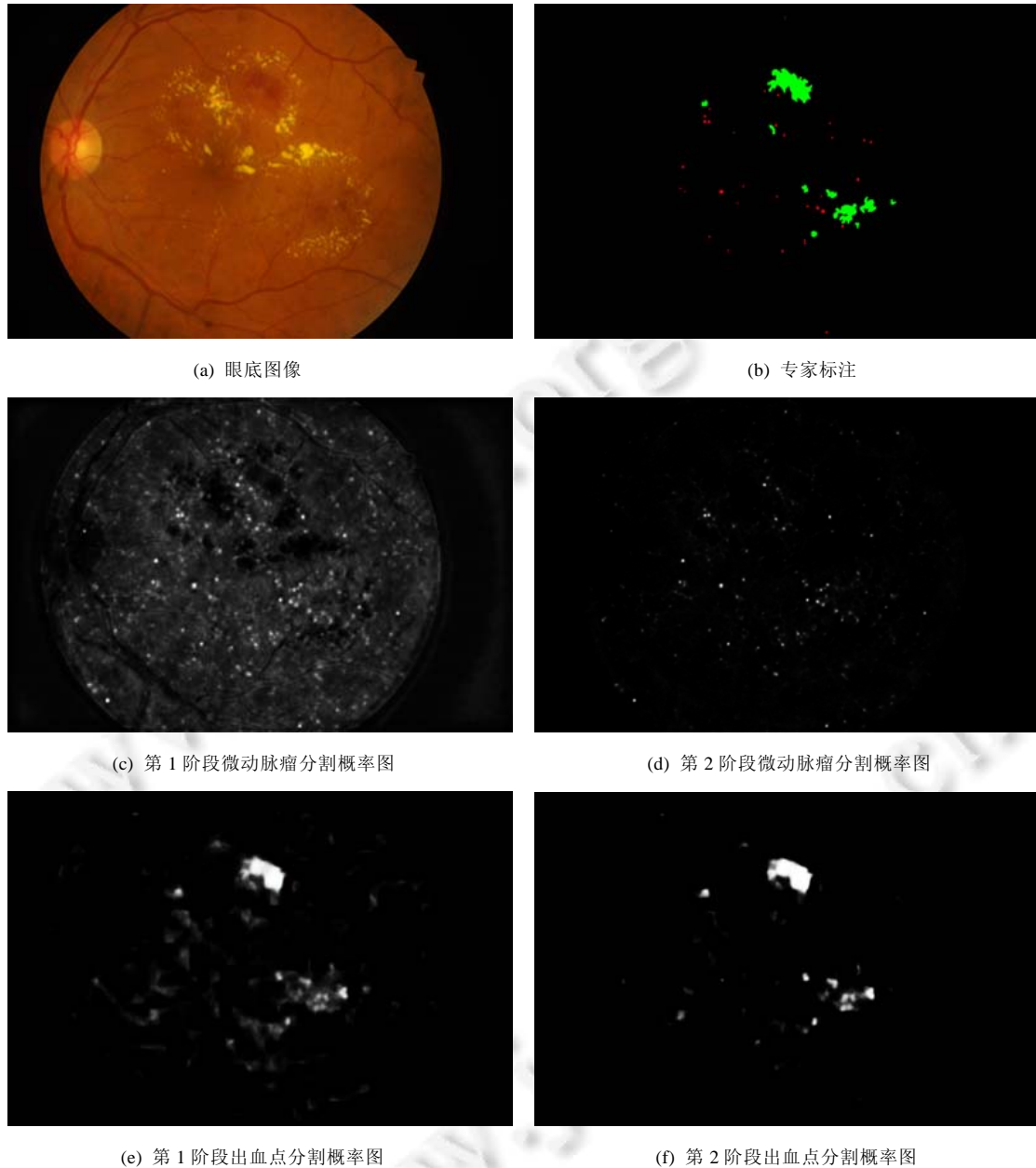


Fig.5 Segmentation probability maps for microaneurysms and hemorrhages

图5 微动脉瘤和出血点分割概率图

表3是本文方法与其他方法的对比.与 HED,FCRN,Deeplabv3+和 Clément 等人的这些方法相比,Red-Seg 在两种病变点分割任务上取得了更好的 AUC\_PR.主要原因在于:这些模型使用简单的交叉熵损失函数训练,并没有考虑难分样本的问题,所以存在比较严重的误分情况,以致于这些模型的准确率和召回率都比 Red-Seg 低.

与 L-Seg 相比,同样都是图像级的多病变点分割网络,L-Seg 和 Red-Seg 都考虑了难分样本的问题,但是 Red-Seg 微动脉瘤的分割效果要好,主要存在两个原因.

- 一是 L-Seg 中使用多通道的分桶损失函数(multi-channel bin loss),该损失每次迭代保留的负样本个数不确定,而且个数不是随着训练迭代逐渐较少;由于 Red-Seg 使用改进的 Top- $k$  带权交叉熵损失函数,所以随着训练迭代保留的负样本个数逐渐较少,从而网络更集中在难分样本上.
- 二是 L-Seg 仅使用一个损失函数训练,而 Red-Seg 使用两阶段的训练算法训练,在第 2 阶段直接将误分样本作为优化目标进行训练,因此能够获得更高的分割精度.

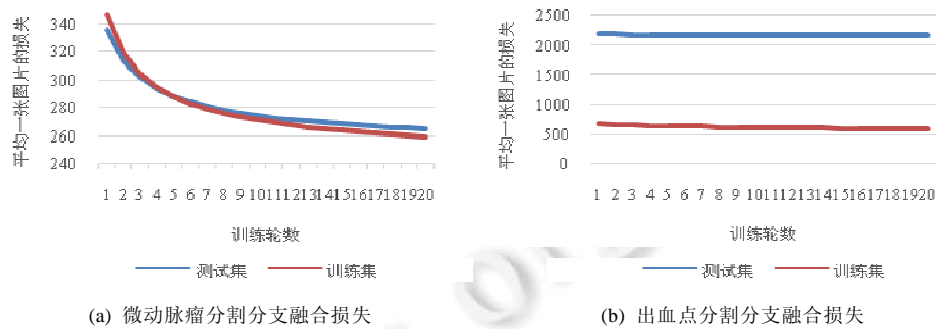


Fig.6 Loss curves of Red-Seg on the training set and test set

图 6 Red-Seg 模型在训练集和测试集上的损失曲线

Table 3 Comparison with other lesion segmentation methods

表 3 与其他病变点分割模型的对比

| 方法                                   | AUC_PR(微动脉瘤)   | AUC_PR(出血点)    |
|--------------------------------------|----------------|----------------|
| HED <sup>[8]</sup>                   | 0.427 1        | 0.508 4        |
| FCRN <sup>[8]</sup>                  | 0.338 3        | 0.420 0        |
| DeepLabv3+ <sup>[8]</sup>            | 0.151 4        | 0.477 2        |
| L-Seg <sup>[8]</sup>                 | 0.462 7        | <b>0.637 4</b> |
| Clément <i>et al</i> <sup>[15]</sup> | -              | 0.530 0        |
| LWENet <sup>[27]</sup>               | 0.344 3        | 0.376 9        |
| Red-Seg                              | <b>0.483 9</b> | 0.569 5        |

表 4 是 Red-Seg 与 IDRiD 竞赛前 4 名的对比.

Table 4 Comparison with IDRiD challenge teams

表 4 与 IDRiD 竞赛队伍的对比

| 方法                | AUC_PR(微动脉瘤)   | AUC_PR(出血点)    |
|-------------------|----------------|----------------|
| VRT (1st)         | 0.495 1        | <b>0.680 4</b> |
| PAtech (2nd)      | 0.474 0        | 0.649 0        |
| iFLYTEK-MIG (3rd) | <b>0.501 7</b> | 0.558 8        |
| SOONER (4th)      | 0.400 3        | 0.539 5        |
| Red-Seg           | 0.483 9        | 0.569 5        |

从表 4 中可以看出,Red-Seg 在微动脉瘤和出血点分割上排名第三位.其中,Red-Seg 微动脉瘤的 AUC\_PR 比第一名仅低 1.8%.另外,Red-Seg 属于图像级的分割网络,而竞赛前三名的队伍使用的模型属于图像块级分割网络.图像块级分割网络与图像级分割网络相比存在两个缺点.

- 第一,图像块级分割网络的训练数据需要人为地根据某些规则从整幅图像中截取,图像块的选取对分割网络的性能至关重要;而图像级分割网络的训练数据为整幅图像,并不需要人为干涉.
- 第二,图像级分割网络只需一次前向传播便可以得到眼底图像的分割图,而图像块级分割网络需要多次的前向传播才能得到一幅图像的分割图.

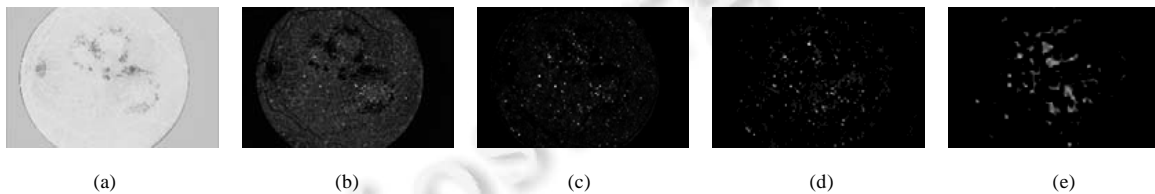
因此,Red-Seg 模型在训练过程和运行效率上比 IDRiD 竞赛队伍的模型更有优势.

表 5 是在融合不同尺度特征情况下,微动脉瘤的分割效果对比.与 Red-Seg 中只融合了 conv2~conv4 的特征相比,加入 conv5 的特征并没有带来分割效果的提升.主要原因是 conv5 的特征图是原图下采样 16 倍得到的,丢失了大量细节信息,所以经过上采样之后,分割结果较粗糙(如图 7(e)所示).另外,在 Red-Seg 中融合 conv1 的特征也没有带来性能提升.主要原因是 conv1 的特征图感受野小,虽然保留了细节信息,但是没有学到太多语义信息(如图 7(a)所示).

**Table 5** Comparison of segmentation performance of microaneurysms under different feature fusion methods (using one-stage training)

**表 5** 不同特征融合方式下微动脉瘤的分割性能对比(使用一阶段训练)

| 方法                      | 微动脉瘤           |                |                |
|-------------------------|----------------|----------------|----------------|
|                         | 准确率            | 召回率            | AUC PR         |
| conv2+conv3+conv4       | <b>0.487 5</b> | <b>0.492 6</b> | <b>0.466 9</b> |
| conv2+conv3+conv4+conv5 | 0.485 2        | 0.487 9        | 0.447 7        |
| conv1+conv2+conv3+conv4 | 0.486 4        | 0.489 7        | 0.464 7        |



**Fig.7** From left to right, segmentation probability maps for microaneurysms of conv1~conv5 correspond to the supervision loss

**图 7** 从左到右依次为 conv1~conv5 对应监督损失的微动脉瘤分割概率图

#### 4 总结与展望

分割眼底病变点是一件具有挑战性的任务.本文提出了一种基于多任务学习的、可以同时分割出血点和微动脉瘤的网络模型 Red-Seg.网络中使用两个分支分别对不同的病变点进行分割,并且网络中使用多尺度特征融合和深度监督的方法得到更准确的分割结果.同时,本文提出了一种两阶段训练算法:第 1 阶段使用改进的 Top-k 带权交叉熵损失函数,第 2 阶段直接最小化假阳性和假阴性来减少病变点分割中的误分情况.本文在 IDRiD 数据集上进行实验,结果表明,两阶段训练算法可以显著减少误分,而且 Red-Seg 与其他方法和 IDRiD 竞赛队伍相比体现出了较好的优势.最后,模型将在 <https://github.com/guomugong/Red-Seg> 开源.

在眼底病变点分割方向仍存在若干问题有待解决.

- 第一,从数据的角度来看,目前仍缺少大规模的公开眼底病变点标注数据集.IDRiD 数据集的规模仅仅为 81 张,与深度学习动辄上亿的参数量对比,这种数据量是远远不够的.数据不足的问题已经成为影响深度学习在眼底病变点检测领域应用的关键因素.
- 第二,从技术的角度来看,目前深度学习的基本假设是训练集和测试集服从同一个分布,而在不同的眼底相机拍摄的眼底图像中病变点的特征存在差异,这就使得在一个眼底相机拍摄的图像上训练的模型在其他眼底相机拍摄的图像上预测时会造成诊断效果的下降.

因此,如何提高分割网络的泛化性也是临床应用上的一个重要问题.

#### References:

- [1] Wong TY, Sun J, Kawasaki R, *et al.* Guidelines on diabetic eye care: The Int'l council of ophthalmology recommendations for screening, follow-up, referral, and treatment based on resource setting. *Ophthalmology*, 2018,125(10):1608–1622. [doi: 10.1016/j.ophtha.2018.04.007]

- [2] Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, Venugopalan S, Widner K, Madams T, Cuadros J. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Journal of the American Medical Association*, 2016,316(22):2402–2410. [doi: 10.1001/jama.2016.17216]
- [3] Pang H, Wang C. Deep learning model for diabetic retinopathy detection. *Ruan Jian Xue Bao/Journal of Software*, 2017,28(11): 3018–3029 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5332.htm> [doi: 10.13328/j.cnki.jos.005332]
- [4] Guo S, Wang K, Kang H, Zhang YJ, Gao YQ, Li T. BTS-DSN: Deeply supervised neural network with short connections for retinal vessel segmentation. *Int'l Journal of Medical Informatics*, 2019,126:105–113. [doi: 10.1016/j.ijmedinf.2019.03.015]
- [5] Liang LM, Liu BW, Yang HL, Shi F, Chen XJ. Supervised blood vessel extraction in retinal images based on multiple feature fusion. *Chinese Journal of Computers*, 2018,41(11):2566–2580 (in Chinese with English abstract). [doi: 10.11897/SP.J.1016.2018.02566]
- [6] Pellegrini E, Robertson G, Macgillivray T, Hemert JV, Trucco E. A graph cut approach to artery/vein classification in ultra-widefield scanning laser ophthalmoscopy. *IEEE Trans. on Medical Imaging*, 2018,37(2):516–526. [doi: 10.1109/TMI.2017.2762963]
- [7] Fu HZ, Cheng J, Xu YW, Wong DWK, Liu J, Cao XC. Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. *IEEE Trans. on Medical Imaging*, 2018,37(7):1597–1605. [doi: 10.1109/TMI.2018.2791488]
- [8] Guo S, Li T, Kang H, Li N, Zhang YJ, Wang K. L-Seg: An end-to-end unified framework for multi-lesion segmentation of fundus images. *Neurocomputing*, 2019,349:52–63. [doi: 10.1016/j.neucom.2019.04.019]
- [9] Guo S, Wang K, Kang H, Liu T, Gao YQ, Li T. Bin loss for hard exudates segmentation in fundus images. *Neurocomputing*, 2020,392:314–324. [doi: 10.1016/j.neucom.2018.10.103]
- [10] Van GM, Van GB, Hoyng C, Theelen T, Sanchez C. Fast convolutional neural network training using selective data sampling: Application to hemorrhage detection in color fundus images. *IEEE Trans. on Medical Imaging*, 2016,35(5):1273–1284. [doi: 10.1109/TMI.2016.2526689]
- [11] Xie S, Tuo ZW. Holistically-nested edge detection. In: *Proc. of the IEEE Int'l Conf. on Computer Vision*. 2015. 1395–1403. [doi: 10.1109/ICCV.2015.164]
- [12] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: *Proc. of the Int'l Conf. on Medical Image Computing and Computer Assisted Intervention*. 2015. 234–241. [doi: 10.1007/978-3-319-24574-4\_28]
- [13] Mo J, Zhang L, Feng Y. Exudate-based diabetic macular edema recognition in retinal images using cascaded deep residual networks. *Neurocomputing*, 2018,290:161–171. [doi: 10.1016/j.neucom.2018.02.035]
- [14] Porwal P, Pachade S, Kokare M, Deshmukh G, Son J, *et al.* IDRiD: Diabetic retinopathy—Segmentation and grading challenge. *Medical Image Analysis*, 2020,59:101561. [doi: 10.1016/j.media.2019.101561]
- [15] Clément P, Renaud D, Farida C. A novel weakly supervised multitask architecture for retinal lesions segmentation on fundus images. *IEEE Trans. on Medical Imaging*, 2019,38(10):2432–2444. [doi: 10.1109/TMI.2019.2906319]
- [16] Tan JH, Fujita H, Sivaprasad S, Bhandary SV, Rao AK, Chua KC, Acharya UR. Automated segmentation of exudates, haemorrhages, microaneurysms using single convolutional neural network. *Information Sciences*, 2017,420:66–76. [doi: 10.1016/j.ins.2017.08.050]
- [17] Chudzik P, Majumdar S, Calivá F, AlDiri B, Hunter A. Microaneurysm detection using fully convolutional neural networks. *Computer Methods and Programs in Biomedicine*, 2018,158:185–192. [doi: 10.1016/j.cmpb.2018.02.016]
- [18] Yang YH, Li T, Li W, Wu HS, Fan W, Zhang WS. Lesion detection and grading of diabetic retinopathy via two-stages deep convolutional neural networks. In: *Proc. of the Int'l Conf. on Medical Image Computing and Computer Assisted Intervention*. 2017. 533–540. [doi: 10.1007/978-3-319-66179-7\_61]
- [19] Lin TY, Goyal P, Girshick R, He KM, Dollár P. Focal loss for dense object detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2020,42(2):318–327. [doi: 10.1109/TPAMI.2018.2858826]
- [20] Rahman MA, Wang Y. Optimizing intersection-over-union in deep neural networks for image segmentation. In: *Proc. of the Int'l Symp. on Visual Computing*. 2016. 234–244. [doi: 10.1007/978-3-319-50835-1\_22]

- [21] Zhu WT, Huang YF, Zeng L, Chen XM, Liu Y, Qian Z, Du N, Fan W, Xie XH. AnatomyNet: Deep learning for fast and fully automated whole—Volume segmentation of head and neck anatomy. *Medical Physics*, 2019,46(2):576–589. [doi: 10.1002/mp.13300]
- [22] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2017,39(4):640–651. [doi: 10.1109/TPAMI.2016.2572683]
- [23] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [24] Szegedy C, Liu W, Jia YQ, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2015. 1–9. [doi: 10.1109/CVPR.2015.7298594]
- [25] Jia D, Wei D, Richard S, Lijia L, Kai L, Li FF. ImageNet: A large-scale hierarchical image database. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2009. 248–255. [doi: 10.1109/CVPR.2009.5206848]
- [26] Jia YQ, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S, Darrell T. Caffe: Convolutional architecture for fast feature embedding. In: *Proc. of the 22Nd ACM Int'l Conf. on Multimedia*. 2014. 675–678. [doi: 10.1145/2647868.2654889]
- [27] Guo S, Li T, Wang K, Zhang C, Kang H. A lightweight neural network for hard exudate segmentation of fundus image. In: *Proc. of the Int'l Conf. on Artificial Neural Networks*. 2019. 189–199. [doi: 10.1007/978-3-030-30508-6\_16]

#### 附中文参考文献:

- [3] 庞浩,王枫.用于糖尿病视网膜病变检测的深度学习模型.软件学报,2017,28(11):3018–3029. <http://www.jos.org.cn/1000-9825/5332.htm> [doi: 10.13328/j.cnki.jos.005332]
- [5] 梁礼明,刘博文,杨海龙,石霏,陈新建.基于多特征融合的有监督视网膜血管提取.计算机学报,2018,41(11):2566–2580. [doi: 10.11897/SP.J.1016.2018.02566]



郭松(1991—),男,博士,主要研究领域为医疗影像分析,深度学习.



康宏(1973—),男,博士,讲师,CCF 专业会员,主要研究领域为数据库技术,机器学习.



李涛(1977—),男,博士,教授,博士生导师,CCF 杰出会员,主要研究领域为异构计算,机器学习,智能物联网.



张玉军(1976—),男,博士,研究员,博士生导师,CCF 高级会员,主要研究领域为计算机网络.



李宁(1994—),男,硕士,主要研究领域为医疗影像分析,深度学习.



王恺(1979—),男,博士,副教授,CCF 专业会员,主要研究领域为人工智能,计算机视觉,机器学习.