

中图分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2022)03-0961-12

论文引用格式: Tan X, Diao Y C, Chen X J, Shi F, Fan Y, Xie J M and Zhu W F. 2022. Deep-supervision and feature-aggregation network for linear lesion segmentation of high myopia. Journal of Image and Graphics 27(03):0961-0972(谭晓,刁逸超,陈新建,石霏,樊莹,谢嘉旻,朱伟芳. 2022. 面向高度近视条纹损伤的深监督特征聚合网络. 中国图象图形学报 27(03):0961-0972 [DOI: 10.11834/jig.210642]

面向高度近视条纹损伤的深监督特征聚合网络

谭晓¹, 刁逸超¹, 陈新建^{1,2}, 石霏¹, 樊莹³, 谢嘉旻³, 朱伟芳^{1*}

1. 苏州大学电子信息学院, 苏州 215006; 2. 苏州大学放射医学与辐射防护国家重点实验室, 苏州 215123;

3. 上海交通大学附属第一人民医院, 上海 200080

摘要: 目的 条纹状损伤是高度近视向病理性近视发展过程中的一种重要眼底改变。临床研究表明,在无创的眼底光学相干断层扫描(optical coherence tomography, OCT)图像中,条纹状损伤主要表现为视网膜色素上皮层—Bruch's膜—脉络膜毛细血管复合体(retinal pigment epithelium-Bruch's membrane-choriocapillario complex, RBCC)损伤,具体包括RBCC异常和近视牵引纹。由于OCT图像中条纹损伤存在目标小、边界模糊等问题,其自动分割极具挑战性。本文提出了一种基于特征聚合下采样和密集语义流监督的特征融合分割网络(deep-supervision and feature-aggregation based network, DSFA-Net),用于实现OCT图像中条纹损伤的自动分割。方法 为减少网络参数量,DSFA-Net以通道减半的U-Net为基准网络,在编码器路径中设计并嵌入特征聚合下采样模块(feature aggregation pooling module, FAPM),使得网络在下采样过程中保留更多的上下文和局部信息,在解码器路径中设计并嵌入一种新的深监督模块——密集语义流监督模块(dense semantic flow supervision module, DSFSM),使得网络在解码过程中聚合不同分辨率图像之间的细节和语义信息。结果 方法在上海市第一人民医院提供的751幅2维眼底OCT B扫描图像上进行验证与评估。相比于U-Net,本文方法在参数量降低53.19%的情况下,平均Dice相似系数(Dice similarity coefficient, DSC)、Jaccard和敏感度指标分别提高了4.30%、4.60%和2.35%。与多种较流行的语义分割网络相比,本文网络在保持最小网络参数量的同时,分割性能取得了较明显提升。为了更加客观地评估FAPM与DSFSM模块的性能,本文同时进行了一系列消融实验。在基准网络添加FAPM模块后,平均DSC、Jaccard和敏感度指标分别提高了1.05%、1.35%和3.35%。在基准网络添加DSFSM模块后,平均DSC、Jaccard和敏感度指标分别提高了4.90%、5.35%和5.90%。在基准网络添加FAPM与DSFSM模块后,平均DSC、Jaccard和敏感度指标分别提高了6.00%、6.45%和5.50%。消融实验的结果表明,本文提出的DSFSM和FAPM模块可以有效提升网络的分割性能。结论 本文提出的DSFA-Net提升了眼底OCT图像中条纹损伤的分割精度,具备潜在的临床应用价值。

关键词: 高度近视; 条纹损伤; 光学相干断层扫描(OCT); 深监督; 特征聚合; 卷积神经网络(CNN); 医学图像分割

Deep-supervision and feature-aggregation network for linear lesion segmentation of high myopia

Tan Xiao¹, Diao Yichao¹, Chen Xinjian^{1,2}, Shi Fei¹, Fan Ying³, Xie Jiamin³, Zhu Weifang^{1*}

1. School of Electronic and Information Engineering, Soochow University, Suzhou 215006, China;

2. National Key Laboratory of Radiology and Radiation Protection, Soochow University, Suzhou 215123, China;

3. The First People's Hospital Affiliated to Shanghai Jiao Tong University, Shanghai 200080, China

收稿日期: 2021-08-16; 修回日期: 2021-12-26; 预印本日期: 2022-01-04

* 通信作者: 朱伟芳 wfzhu@suda.edu.cn

基金项目: 国家重点研发计划资助(2018YFA0701700); 国家自然科学基金项目(U20A20170)

Supported by: National Key R&D Program of China (2018YFA0701700); National Natural Science Foundation of China (U20A20170)

Abstract: **Objective** Linear lesion is an important symptom in the progressive development of high myopia to pathological myopia. Clinical studies have shown that linear lesion appears as retinal pigment epithelium-Bruch's membrane-choriocapillaris complex (RBCC) disruption in non-invasive retinal optical coherence tomography (OCT) images; it includes RBCC disorder and myopic stretch line. Recently, convolutional neural networks (CNNs) have demonstrated excellent performance on computer vision tasks, and many convolutional neural network based methods have been applied for medical segmentation tasks. However, the automatic segmentation of linear lesion is extremely challenging due to the small target attribution and blurred boundary problem. To tackle this issue, a novel deep-supervision and feature-aggregation based network (DSFA-Net) is proposed for the segmentation of linear lesion in OCT image with high myopia. **Method** To reduce the network parameters, the proposed DSFA-Net considers the U-Net with half channels the baseline. A novel feature aggregation pooling module (FAPM) is proposed and embedded in the encoder path to preserve more details for small targets. It can aggregate the contextual information and local spatial information during the downsampling operation. FAPM is performed using two steps. First, the input feature map is parallel fed into three pathways. The first two pathways contain a horizontal and vertical strip pooling layer followed by a 1D convolutional layer with kernel size of 1×3 and 3×1 and a reshape layer to capture contextual information. The third pathway contains a 2D convolutional layer with kernel size 7×7 followed by a sigmoid function to enable each pixel to obtain a normalized weight between 0 and 1. These weights are multiplied with the original input feature and fed into a reshape layer to capture the local spatial information. Second, the output features of these pathways are combined by the element-wise addition to obtain the aggregated output feature. A novel dense semantic flow supervision module (DSFSM) is proposed and embedded in the decoder path to aggregate the details and semantic information between features with different resolutions during the feature decoding. This approach combines the advantages of deep supervision and dense semantic flow supervision strategy and increases the effective feature maps in the hidden layers of the network. The proposed DSFA-Net is implemented and trained based on Python3.8 and Pytorch with NVIDIA TITAN X GPU, Intel i7-9700KF CPU. The initial learning rate is set to 0.001, and the batch size is set to 2. Stochastic gradient descent (SGD) with a momentum of 0.9 and weight decay of 0.0001 is adopted as the optimizer. Binary cross entropy (BCE) loss and Dice loss are combined as the total loss function for the proposed DSFA-Net because linear lesion has variant sizes, which cause data unbalance problem. **Result** The proposed DSFA-Net was evaluated on 751 2D retinal OCT B-scan images provided by the First People's Hospital Affiliated to Shanghai Jiao Tong University. The size of each OCT B-scan image is 256×512 pixels. The ground truth is manually annotated under the supervision of the experienced ophthalmologists. Compared with the original U-Net, the proposed DSFA-Net decreases the network parameter number by 53.19%, and the average Dice similarity coefficient (DSC), Jaccard, and sensitivity indicators increase by 4.30%, 4.60%, and 2.35%, respectively. Compared with the seven other existing image semantic segmentation networks, such as CE-Net, SegNet, and Attention-UNet, the proposed DSFA-Net has achieved state-of-the-art segmentation performance while maintaining the minimum amount of network parameters. Several ablation experiments have been designed and conducted to evaluate the performance of the proposed FAPM and DSFSM modules. With the embedding of FAPM into the encoder path of the baseline (baseline + FAPM), the average DSC, Jaccard, and sensitivity indicators increase by 1.05%, 1.35%, and 3.35%, respectively. With the embedding of DSFSM into the decoder path of the baseline (baseline + DSFSM), the average DSC, Jaccard, and sensitivity indicators increase by 4.90%, 5.35%, and 5.90%, respectively. With the embedding of FAPM and DSFSM into the baseline (the proposed DSFA-Net), the average DSC, Jaccard, and sensitivity indicators increase by 6.00%, 6.45%, and 5.50%, respectively. The results of the ablation experiment show that the proposed FAPM and DSFSM modules can effectively improve the segmentation performance of the network. **Conclusion** We propose a novel deep-supervision and feature-aggregation based network for the segmentation of linear lesion in OCT image with high myopia. The proposed FAPM and DSFSM modules can be inserted into convolutional neural networks conveniently. The experimental results prove that the proposed DSFA-Net improves the accuracy of linear lesion segmentation in retinal OCT images, indicating the potential clinical application value.

Key words: high myopia; linear lesion; optical coherence tomography (OCT); deep supervision; feature aggregation; convolutional neural network (CNN); medical image segmentation

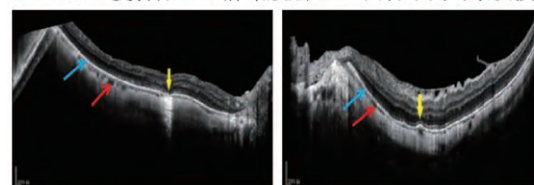
0 引言

人类从外界获取的信息中 70% 以上来自视觉 (Marr 和 Vaina, 1982)。随着手机、电脑等电子产品的普及,世界范围内由于过度用眼导致近视的发病率正在逐年上升。一般当近视度数超过 600° 或者眼轴大于 26 mm 时,临床诊断为高度近视 (Tang 等, 2015)。世界卫生组织的一项研究表明预计到 2050 年,全球高度近视人数将占总人口的 10% (Mariotti 等 2015)。部分高度近视患者存在严重的视觉健康隐患,主要表现为近视度数不断加深,视力下降甚至最终失明 (Tokoro, 1988),这类发生眼底病变并可能致盲的近视称为病理性近视 (Huang 等, 2009)。目前,我国是世界近视人口最多的国家,高度近视占近视人口 20% 以上,其中由病理性近视导致的视力损伤占 17.6% ~ 40%,病理性近视已成为 45 ~ 59 岁人群不可逆致盲的首要病因 (Wang 等, 2013)。条纹状损伤 (linear lesion) 是高度近视渐进发展为病理性近视过程中的一种重要眼底特征,与脉络膜新生血管及黄斑萎缩等严重眼底疾病有着密切相关性 (Ohno-Matsui 等 2003),

65% 以上的条纹损伤若不加以控制将会继续恶化 (Fang 等, 2018)。临床研究表明,在无创的光学相干断层扫描 (optical coherence tomography, OCT) 图像中,条纹状损伤主要表现为色素上皮—Bruch's 膜—脉络膜毛细血管复合体 (retinal pigment epithelium—Bruch's membrane—choriocapillaris complex, RBCC) 损伤,具体包括 RBCC 异常和近视牵引纹 (Shinohara 等 2014)。

条纹损伤的自动分割以及定量分析对于病理性近视的预防与早诊早治具有非常重要的临床意义。但如图 1 所示,首先,由于在 OCT 图像中 RBCC 异常和近视牵引纹均属于小目标,前景与背景像素数量相差悬殊,存在严重的数据不平衡问题。其次, RBCC 异常和近视牵引纹在 OCT 图像与其他正常组织结构在像素灰度、形态结构等方面存在极大相似性,这对条纹损伤尤其是 RBCC 异常的分割会造成严重干扰。因此,基于 OCT 图像的条纹损伤自动分割极具挑战性。目前,有关条纹损伤的分析与诊断

→ BBCC 复合体 → 脉络膜层 → 两种不同的条纹损伤



(a) RBCC异常

(b) 近视牵引纹

图1 OCT 图像中的条纹状损伤

Fig. 1 Linear lesion in OCT image

((a) RBCC disorder; (b) myopic stretch line)

主要依赖于医生的主观经验,相关自动分割技术鲜有报道,目前仅有 Jiang 等人 (2019)、Feng 等人 (2020) 基于吲哚青绿血管造影 (indocyanine green angiography, ICGA) 图像对条纹损伤的自动分割进行了研究,分别提出了基于部分密集连接与改进 c-GAN (conditional generative adversarial network) 的分割网络和基于上下文金字塔聚合的分割网络,较好实现了 ICGA 图像中条纹状损伤的自动分割。虽然 ICGA 是目前临床诊断条纹损失的金标准,但需要注射吲哚青绿造影剂,属于有创成像方式,部分病人会对造影剂产生过敏甚至休克等不良反应。因此,基于眼底 OCT 图像的条纹损伤自动分割对于眼科临床条纹损伤的大规模筛查和无创检测具有重要意义。为实现 OCT 图像中条纹损伤的自动分割,本文主要工作包括:

1) 针对传统池化下采样方式在小目标分割任务中易造成信息丢失,最终导致预测图中小目标的漏检和误检问题,本文提出一种新的特征聚合下采样模块 (feature aggregation pooling module, FAPM),可以在下采样过程中保留更多的上下文和局部信息。

2) 提出了一种新的深监督模块——密集语义流监督模块 (dense semantic flow supervision module, DSFSM),以增加网络隐藏层中的有效特征图,并聚合不同分辨率图像之间的细节和语义信息。

3) 提出一种基于特征聚合下采样模块和密集语义流监督模块的特征聚合分割网络 (deep-supervision and feature-aggregation based network, DSFA-Net),首次较好地实现了眼底 OCT 图像中条纹损伤的自动分割。

1 本文方法

1.1 整体网络结构

U-Net(Ronneberger 等 2015) 是一种为医学图像分割任务设计的卷积神经网络,其通过跳跃连接,结合网络浅层信息,可以缓解网络上采样过程中信息不足的缺陷。但是由于小目标分割任务的特殊性,仅通过传统跳跃连接来实现深层信息与浅层信息的融合往往会造成细粒度信息缺失等问题。为此本文提出了一种基于特征聚合下采样和密集语义流监督的特征融合分割网络 DSFA-Net 来解决这个问题。如图 2 所示。为减少网络参数量,本文提出的 DSFA-Net 以通道减半的 U-Net 为基准网络,采用 5 层 U 型编码器—解码器结构,每层编码器由 2 个基本编码单元组成,每个基本编码单元由 1 个 3×3 的卷积和 1 个修正线性单元激活函数(rectified linear unit, ReLU) 组成。第 1 层编码器的通道数为 32,每经过一层编码器,特征通道数增加一倍。下采样阶段使用本文提出的特征聚合下采样模块 FAPM。与原始 U-Net 相比,结构上的改进主要包括: 1) 放弃编码器中传统的最大池化或者平均池化下采样模式,提出并嵌入了一种新的下采样模块——特征聚合下采样模块 FAPM,减少下采样过程中小目标信息的损失; 2) 在解码器路径中,提出并嵌入了一种新的深监督模块——密集语义流监督模块 DSFSM,在训练过

程中将语义信息更有效地从网络深层传递到浅层并实现特征聚合,同时能够加快网络的收敛速度。

1.2 特征聚合下采样模块

卷积神经网络中的池化下采样能够在降低特征维度的同时保留特征有效信息,在保持特征旋转、平移和尺度不变性的情况下减少网络过拟合(Nagi 等 2011)。但传统的平均池化或者最大池化下采样方式存在感受野小、未考虑全局信息等问题(Boureau 等 2010)。在小目标分割任务中会造成严重的信息丢失,最终导致预测图中小目标的漏分和误分。目前,已有一些改进的下采样方法尝试解决上述问题(张新良 等 2020),比如采用条纹状池化(Hou 等 2020) 来获取上下文信息,采用局部重要性池化(Gao 等 2019) 获取局部信息等。本文融合了条纹状池化和局部重要性池化的优点,提出了一个新的特征聚合下采样模块 FAPM,其结构如图 3 所示。FAPM 模块主要分为两部分: 1) 图 3 中左边所示的上下文信息获取模块,包含两个支路,分别获取垂直方向和水平方向的上下文信息。第 1 个支路的输入特征图经过垂直条纹池化和一个 1×3 卷积后,尺寸从 $H \times W$ 变为 $1 \times W$,再通过双线性插值将特征图尺寸调整为 $H/2 \times W/2$,用于获取垂直方向的上下文信息;第 2 个支路的输入特征图通过横向条纹池化和一个 3×1 卷积,尺寸从 $H \times W$ 变为 $H \times 1$,然后通过双线性插值将特征图尺寸调整为 $H/2 \times W/2$,用于获取水平方向的上下文信息。最终,两条支

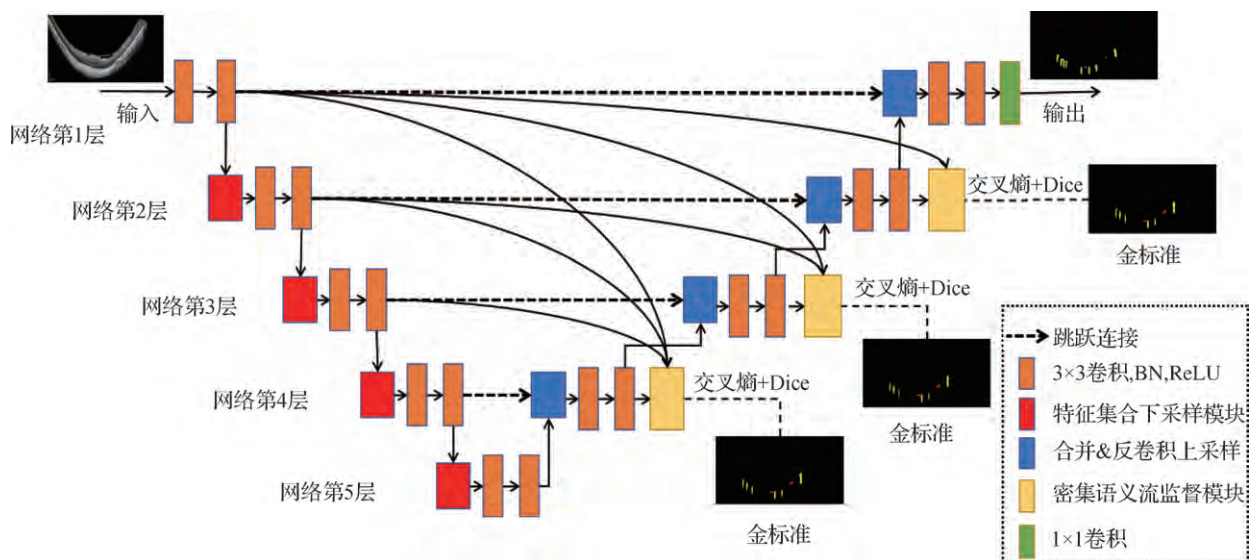


图 2 DSFA-Net 整体网络结构

Fig. 2 The overall network of DSFA-Net

路的输出相加聚合,使网络实现对上下文信息的获取。2)图3右边所示的局部信息获取模块,首先输入特征图经过一个 7×7 卷积,然后通过 Sigmoid 函数使得每一个像素获得(0,1)之间的归一化权重,再将这些权重与输入特征图相乘,经过双线性插值将特征图尺寸调整为 $H/2 \times W/2$,使得输出特征图可以有效聚合每一个像素点的权重,减少信息丢失。将上述两部分的输出特征图相加聚合作为 FAPM 模块最终的输出特征图,构成一种新的下采样方式,实现上下文信息和局部信息特征聚合。

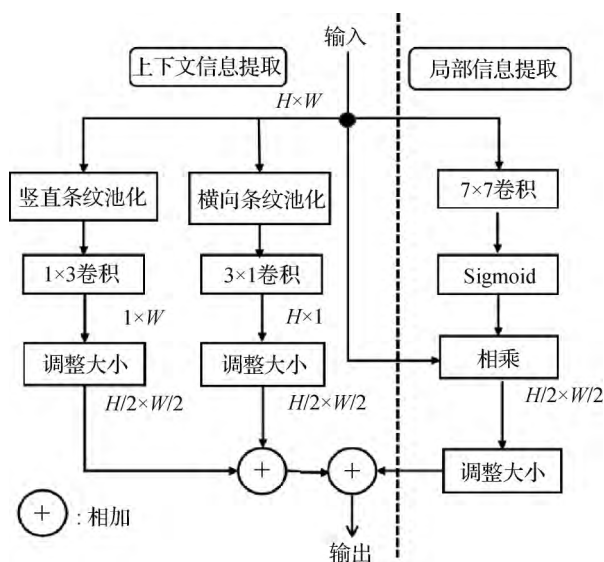


图3 特征聚合下采样模块
Fig.3 The structure of FAPM

1.3 密集语义流监督模块

由于条纹损伤目标较小,随着U型卷积神经网络层数的加深,深层特征往往因为更关注于高级语义信息而造成小目标信息丢失问题,而仅通过简单的跳跃连接进行深层信息与浅层信息的融合,使得条纹损伤边缘区域的决策边界难以判定,从而造成模糊边缘区域难以分割的问题。为此,本文将深监督机制(Wang等2015)融入分割网络,在为输出层提供语义监督的基础上,为隐藏层提供直接的金标准监督,并将金标准信息通过反向传播改变隐藏层参数,解决梯度消失和收敛速度过慢等问题。影响语义分割性能的一个主要因素是网络深隐藏层中具有强语义信息的特征图的分辨率过低,缺乏足够的空间细节信息,而浅层特征图存在分辨率高但缺乏语义信息的问题,因此如何让网络获得具有强语义信息的高分辨率特征图,是提高网络分割性能

的关键。Li等人(2020)为此提出了一种流对齐模块(flow alignment module, FAM),用于学习相邻层特征图之间的语义流,并有效地将高层语义信息与高分辨率细节信息进行对齐。由于条纹损伤分布较广且目标较小,如果能通过语义流将分割目标的位置信息对齐,这将有助于提升条纹损伤的分割性能。

因此,为了增加网络隐藏层中的有效特征图,并聚合不同分辨率图像之间的细节和语义信息,本文提出了DSFSM。以图2中解码器第4层的DSFSM模块为例说明设计原理,其网络结构如图4所示。DSFSM模块共包含3个分支,分别将解码器第4层特征图先通过 1×1 卷积和双线性上采样方式使其通道数和尺寸分别与编码器第1、2、3层特征图的通道数和尺寸保持一致,然后与编码器各层特征图分别级联合并,经过 3×3 卷积产生语义流场,进行如图5所示的语义流对齐,即先将语义流场每一个像素点生成的偏置对应到低分辨率特征图中的一个小数坐标像素点,然后根据这个小数坐标,找到4个最近的整数相邻像素进行可微分双线性上采样(Jaderberg等2015),实现低分辨率特征与高分辨率特征的语义对齐。语义对齐后的特征图再与各自编码器特征图相加聚合。最后3个分支的输出上采样达到与金标准尺寸一致后进行级联合并,经过 1×1 卷积将通道数降为1后与金标准实现深监督,在网络训练过程中通过反向传播改变解码器和编码器各层的参数。至此,编码器第1、2、3层中都融入了来自解码器第4层的语义信息。

相比于原始语义流对齐模块,本文DSFSM模块最大的优点是通过密集连接将解码器中隐藏层的语义信息传递给了全部低层编码器的隐藏层,而原始语义流对齐模块只是将其传递给上一层编码器的隐藏层。此外,DSFSM模块还有来自金标准的深监督机制来辅助优化网络的训练,对网络的各个隐藏层进行直接的辅助分类。因此,DSFSM模块可以使得网络将语义信息和细节信息聚合更充分,进而提升分割性能。

2 实验设置

采用基于PyTorch的深度学习框架实现提出的DSFA-Net,编程语言采用Python, GPU型号为NVIDIA TITAN X,显存12 GB。网络训练过程中的

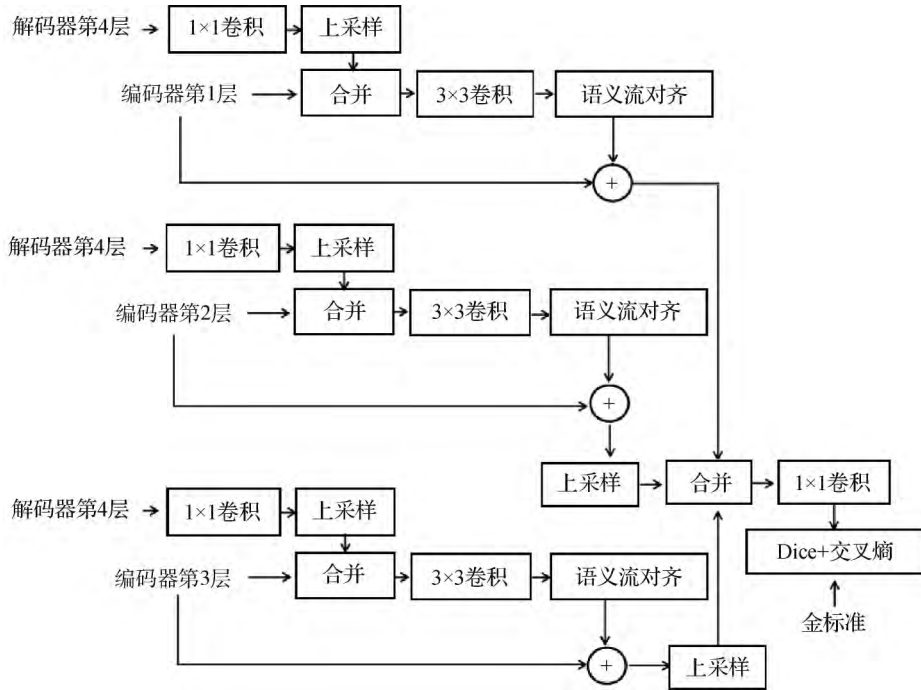


图4 密集语义流监督模块
Fig. 4 The structure of DSFSM

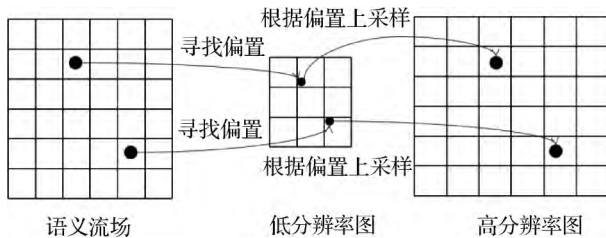


图5 语义流对齐过程
Fig. 5 The alignment of semantic flow

初始学习率设置为 0.001,采用动量为 0.9、权值衰减系数为 0.000 1 的随机梯度下降(Ruder 2016)算法对网络进行优化。网络训练过程中批尺寸设置为 2 采用 3 折交叉验证策略,每一折训练进行 120 次迭代,保存在验证集上 Dice 相似系数(Dice similarity coefficient, DSC)最高的模型。受限于数据集数量,为了提高网络的泛化能力、减少过拟合风险,采用了在线随机数据扩增(Salamon 和 Bello 2017)方式,包括随机旋转、上下翻转、左右翻转和增加高斯噪声。由于 OCT 图像中 RBCC 异常以及近视牵引纹目标区域较小,存在严重的数据不平衡问题,因此本文采用交叉熵损失函数 $Loss_{CE}$ (Boer 等 2005) 与 Dice 损失函数 $Loss_{Dice}$ (Milletari 等 2016) 之和作为网络整体损失函数 $Loss_{total}$,降低数据不平衡的影响。

$$Loss_{CE} = -\frac{1}{n} \sum_{i=1}^n \sum_{c=1}^C y_{i,c} \log(p_{i,c}) \quad (1)$$

$$Loss_{Dice} = 1 - \frac{1}{n} \sum_{i=1}^n \frac{2x_i y_i}{x_i^2 + y_i^2} \quad (2)$$

$$Loss_{total} = Loss_{CE} + Loss_{Dice} \quad (3)$$

式中, n 为图像中的像素总数, i 表示图像中第 i 个像素的值, $y_{i,c}$ 表示金标准中 i 属于类别 c 的概率, $p_{i,c}$ 表示预测结果中 i 预测为类别 c 的概率, C 表示类别数, y 为金标准中的像素值, x 为网络预测的像素值。

3 实验结果与分析

实验采用的眼底 OCT 图像来自上海市第一人民医院,由德国海德堡公司的 Heidelberg SPECTRALIS OCT 仪器采集,包含来自 22 只高度近视眼共计 751 幅 256×512 像素的 2 维 OCT B 扫描图像。按背景、RBCC 异常及近视牵引纹 3 类进行金标准标注,所有金标准标注均在两位专业眼科医生的指导下完成。

3.1 评价指标

为了全面、客观地评估提出的 DSFA-Net 的分割性能,采用了 DSC、Jaccard(Shelhamer 等 2017)指

数和敏感度(sensitivity)作为评价指标。式(4)所定义的DSC指标采用预测结果正确区域像素的两倍与预测结果和金标准区域像素和的比值,是医学图像分割中常见的指标;式(5)定义的Jaccard系数可用于衡量预测值与金标准的交集与并集的比值;式(6)定义的敏感度表示在所有前景中被正确标记为前景的像素的比例。分别计算为

$$f_{\text{DSC}} = \frac{2TP}{2TP + FP + FN} \quad (4)$$

$$f_{\text{Jac}} = \frac{TP}{TP + FP + FN} \quad (5)$$

$$f_{\text{Sen}} = \frac{TP}{TP + FN} \quad (6)$$

式中,TP为真阳性,表示预测值与金标准都为前景的点;FP为假阳性,表示预测值为前景但金标准为背景的点;TN为真阴性,表示预测值与金标准都为背景的点;FN为假阴性,表示预测为背景但金标准为前景的点。上述3个评价指标取值都在[0,1],结果越靠近1表示网络分割性能越好,越靠近0表示网络分割性能越差。

3.2 消融实验

为了验证本文提出的特征聚合下采样模块FAPM以及密集语义流监督模块DSFSM对网络分割性能提升的有效性,以通道数减半的U-Net网络为基准网络,进行消融实验:1)基准网络;2)基准网络+FAPM;3)基准网络+DSFSM;4)基

准网络+DSFSM+FAPM,即本文提出的DSFA-Net。

表1是上述消融实验结果的定量性能比较。由表1可见,对于近视牵引纹的分割,基准U-Net网络的DSC系数为0.687、Jaccard指标为0.569、敏感度指标为0.783;基准网络加入FAPM模块后,DSC指标提升了2.8%、Jaccard指标提升了3.2%、敏感度指标提升了5.3%;基准网络加入DSFSM模块后,DSC指标提升了5.6%、Jaccard指标提升了6.2%、敏感度指标提升了2.9%;基准网络加入DSFSM和FAPM模块后,3项指标均得到进一步提升,最终DSC、Jaccard和敏感度指标较基准网络分别提升了6.4%、7.0%和3.8%。对于RBCC异常的分割,基准U-Net的DSC系数为0.590、Jaccard指标为0.451、敏感度指标为0.607。基准网络加入FAPM模块后,DSC和Jaccard指标略有下降,原因是RBCC异常的边界较为模糊,视网膜其他组织结构病变干扰误导了FAPM模块的判断;基准网络加入DSFSM模块后,DSC指标提升了4.2%、Jaccard指标提升了4.5%、敏感度指标提升了5.6%;基准网络加入DSFSM和FAPM模块后,3项指标均得到进一步提升,DSC、Jaccard和敏感度指标较基准网络分别提升了5.6%、4.9%和6.6%,说明DSFSM和FAPM模块的结合可以有效提取RBCC异常的特征信息,减少误检和漏检。

表1 消融实验结果

Table 1 The results of ablation experiments

网络	近视牵引纹			RBCC异常		
	DSC	Jaccard	敏感度	DSC	Jaccard	敏感度
基准网络	0.687	0.569	0.783	0.590	0.451	0.607
基准网络+FAPM	0.715	0.601	0.836	0.582	0.445	0.621
基准网络+DSFSM	0.743	0.631	0.812	0.632	0.496	0.663
基准网络+DSFSM+FAPM	0.751	0.639	0.827	0.646	0.510	0.673

注:加粗字体表示每列最优结果。

图6展示了4例消融实验的可视化分割结果。如图6(c)基准网络的预测图中,RBCC异常区域误检以及近视牵引纹漏检现象比较严重,说明基准网络采用传统下采样方式以及通过简单跳跃连接方式实现深层信息与浅层信息融合的方式无法解决条纹损伤分割任务中小目标缺失以及边界模糊等问题;

如图6(d)所示,在基准网络中加入FAPM模块后,预测图中近视牵引纹漏检和RBCC异常误检得到有效缓解,表明本文提出的FAPM模块可有效减少下采样过程中细节信息缺失的问题,从而较有效地解决小目标分割问题;如图6(e)所示,在基准网络中加入DSFSM模块后,预测图中的RBCC异常与近视

牵引纹的边缘区域分割结果更接近于金标准,表明本文提出的 DSFSM 模块通过密集语义流监督机制可以有效融合深层信息与浅层信息,从而有效缓解模糊边界区域难以分割的问题;图 6(f) 所示,在基准网络中同时加入 FAPM 和 DSFSM 模块后,

RBCC 异常和近视牵引纹的分割结果得到进一步改善,表明在 DSFA-Net 中特征聚合下采样方式和密集语义流监督机制得到进一步有效融合,更好地解决了条纹损伤分割任务中小目标分割和边界模糊问题。

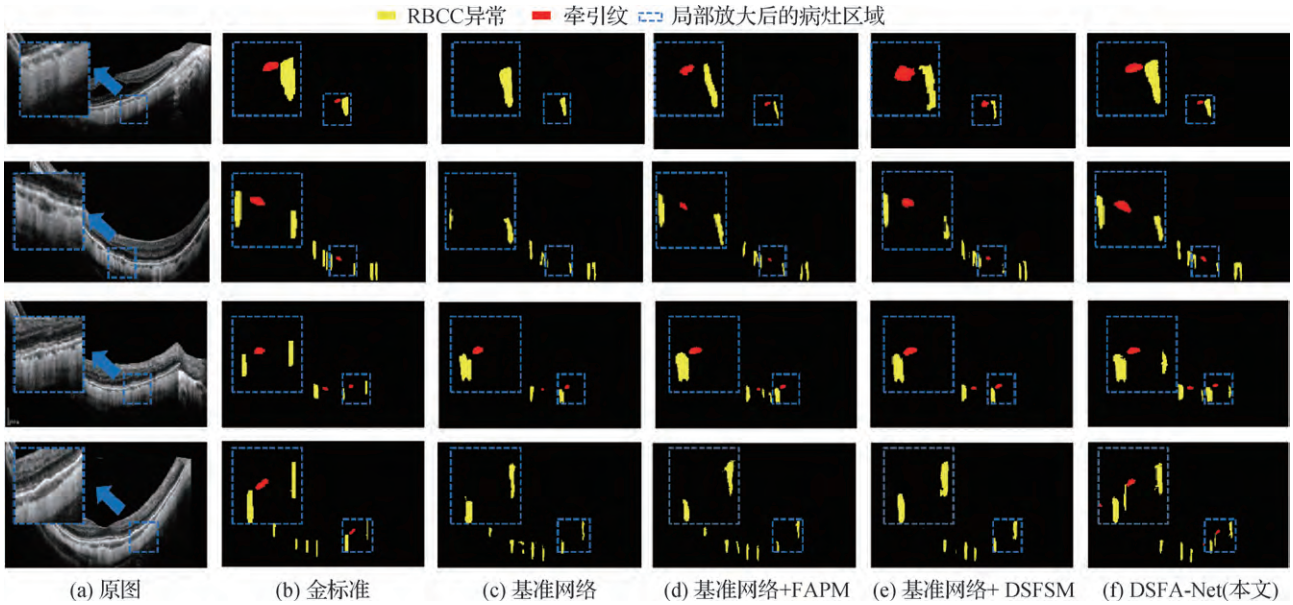


图 6 消融实验分割结果

Fig. 6 Segmentation results of ablation experiments

((a) original images; (b) ground truth; (c) baseline; (d) baseline + FAPM; (e) baseline + DSFSM; (f) DSFA-Net(ours))

3.3 对比实验

为了更加客观地评估本文方法的性能,进行了对比实验,主要包括以下 3 组: 1) 下采样方式对比实验。基于基准网络,对比本文提出的特征聚合下采样模块 DSFSM 与最大池化、平均池化、步长为 2 的 3×3 卷积、条纹池化和局部重要池化等 5 种较为流行的下采样方式。2) 深监督策略对比实验。基于基准网络,对比本文提出的密集语义流监督模块 FAPM 与传统深

监督、语义流深监督等深监督策略。3) 分割性能对比实验。将本文 DSFA-Net 与 SegNet (Badrinarayanan 等, 2017)、PSPNet (pyramid scene parsing network) (Zhao 等, 2017)、CE-Net (context encoder network) (Gu 等, 2019)、DeeplabV3 (Chen 等, 2017)、Attention U-Net (Oktay 等, 2018)、TransUNet (Chen 等, 2021) 和 U-Net 等优秀语义分割网络进行了对比实验。3 组对比实验的结果分别如表 2—表 4 所示。

表 2 下采样方式对比实验结果

Table 2 The comparisons of different down-sampling methods

下采样方式	近视牵引纹			RBCC 异常		
	DSC	Jaccard	敏感度	DSC	Jaccard	敏感度
基准网络 + 最大池化	0. 687	0. 569	0. 783	0. 590	0. 451	0. 607
基准网络 + 平均池化	0. 678	0. 560	0. 792	0. 593	0. 455	0. 625
基准网络 + 步长为 2 的 3×3 卷积	0. 672	0. 556	0. 800	0. 581	0. 442	0. 598
基准网络 + 条纹状池化	0. 693	0. 576	0. 778	0. 578	0. 439	0. 614
基准网络 + 局部重要池化	0. 704	0. 587	0. 819	0. 597	0. 459	0. 618
基准网络 + FAPM	0. 715	0. 601	0. 836	0. 582	0. 445	0. 621

注: 加粗字体表示每列最优结果。

表3 深监督策略对比实验结果

Table 3 The comparisons of different deep supervision methods

深监督策略	近视牵引纹			RBCC 异常		
	DSC	Jaccard	敏感度	DSC	Jaccard	敏感度
基准网络 + 传统深监督	0.724	0.612	0.802	0.632	0.495	0.650
基准网络 + 语义流深监督	0.728	0.614	0.799	0.631	0.496	0.655
基准网络 + DSFSM	0.743	0.631	0.812	0.632	0.496	0.663

注: 加粗字体表示每列最优结果。

表4 不同网络对比实验结果

Table 4 The comparisons of different networks

网络	近视牵引纹			RBCC 异常			参数量/M
	DSC	Jaccard	敏感度	DSC	Jaccard	敏感度	
基准网络	0.687	0.569	0.783	0.590	0.451	0.607	7.76
DeepLabV3	0.480	0.376	0.630	0.442	0.318	0.489	58.16
PSPNet	0.653	0.532	0.757	0.578	0.440	0.626	27.76
SegNet	0.692	0.579	0.803	0.589	0.461	0.653	29.44
CE-Net	0.669	0.549	0.753	0.550	0.412	0.585	29.00
U-Net	0.699	0.583	0.800	0.612	0.474	0.653	31.04
Attention U-Net	0.711	0.612	0.807	0.626	0.502	0.698	44.08
TransUNet	0.724	0.621	0.819	0.588	0.451	0.689	105.5
本文 DSFA-Net	0.751	0.639	0.827	0.646	0.510	0.673	14.53

注: 加粗字体表示每列最优结果。

由表2可知,在下采样方式对比实验中,本文提出的FAPM模块对于近视牵引纹的分割取得了最好的结果,说明FAPM模块能更好地提取和聚合近视牵引纹的上下文信息和局部信息。在RBCC异常的分割中,局部重要池化下采样方式取得了最高的DSC和Jaccard指标,可能的原因是局部重要池化下采样对RBCC异常表现出的高亮区域更敏感,更准确地提取了RBCC异常的形状;平均池化下采样方式获得了最高的敏感度指标,平均池化下采样虽然对RBCC异常分割的真阳性较高,但是假阳性也高,因此出现了敏感度较高但DSC和Jaccard指标较低的情况。

在表3所示的深监督策略对比实验中,对于近视牵引纹和RBCC异常的分割,提出的DSFSM模块均具有优势。在DSC、Jaccard和敏感度3项指标上,DSFSM模块不仅均优于传统深监督机制,也均优于DSFSM模块设计灵感来源的语义流结合深监督机制,其主要原因是相比于语义流模块,DSFSM

模块通过密集连接方式可以将更多解码器隐藏层的语义信息融合到编码器的隐藏层中,使得整体网络中语义和细节的特征信息聚合更加充分。

表4对比了本文DSFA-Net与DeepLabV3、PSPNet、CE-Net、U-Net、SegNet、Attention U-Net、TransUNet等语义分割网络的性能,同时还对各个网络的参数量进行了比较。由表4可见,本文DSFA-Net将通道减半的U-Net作为基准网络,结合FAPM和DSFSM模块后,在尽可能保留细节信息的同时,将深层信息与浅层信息合理结合,在参数量仅为14.53M的情况下,分割性能整体优于其他语义分割网络,表明网络结构的合理设计对特定任务分割性能的提升至关重要。

图7展示了4例采用不同分割网络的分割结果。由图7(d)(e)以及表4可以看出,对于编码器—解码器网络结构,相较于通道减半的U-Net,原始U-Net在增加参数的情况下,RBCC异常误分割现象得到减少,但模糊边界区域的分割结果未得到

有效改善。从图 7(c) 可以看出,在网络参数量为 DSFA-Net 4 倍的 DeeplabV3 的预测图中 RBCC 异常与近视牵引纹的误检以及漏检现象非常严重,说明对于小目标分割任务,高效地提取细节信息至关重要。如果无法对细节进行合理的提取,即使参数量增大也难以优化分割结果。本文提出的 DSFA-Net 以较小的参数量与合理的网络结构,提高了对 RBCC 异常与近视牵引纹细节特征的提取,更好地实现了

RBCC 异常和近视牵引纹的分割,有效缓解了条纹损伤分割任务中小目标以及模糊边界区域难以分割的问题。

综上所述,本文提出的 DSFA-Net 以合理的网络结构和模块设计较好地实现了 RBCC 异常和近视牵引纹分割任务,为 RBCC 异常和近视牵引纹的自动分割任务提供了一个可靠的技术手段,具有很大的临床意义。

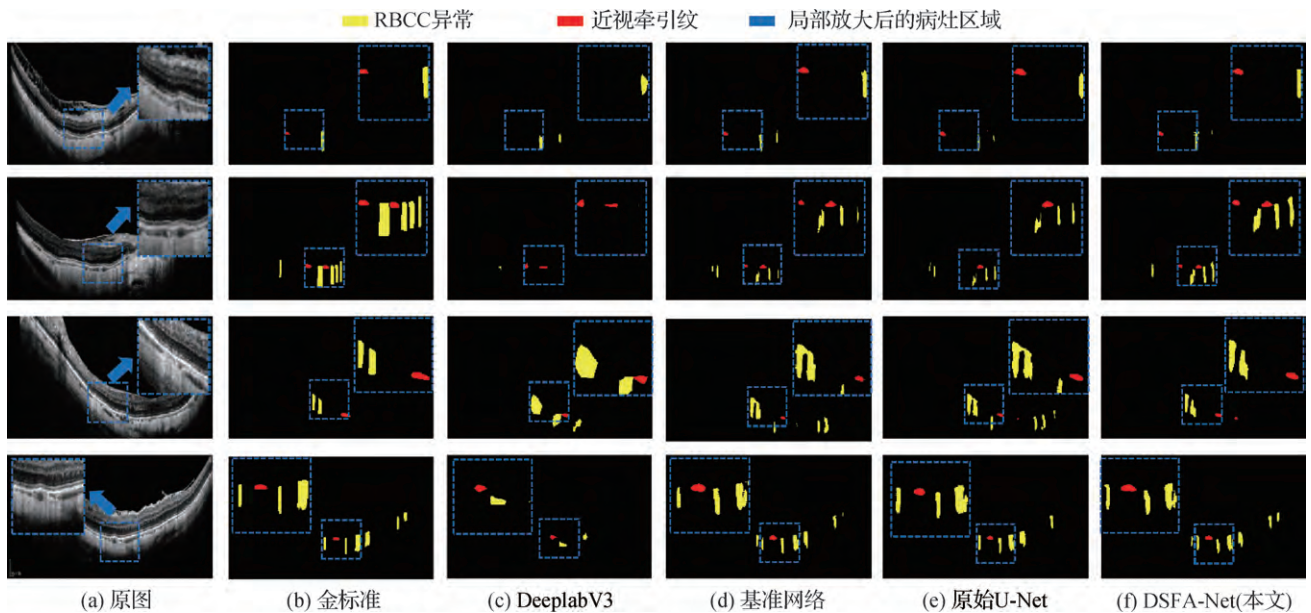


图 7 不同网络的分割结果

Fig. 7 Segmentation results of different networks

((a) original image; (b) ground truth; (c) DeeplabV3; (d) baseline; (e) U-Net; (f) DSFA-Net(ours))

4 结 论

眼底 OCT 图像中条纹损伤的自动分割对于条纹状损伤的无创诊断以及病理性近视的预防与早诊早治具有非常重要的临床意义。本文提出了一种基于特征聚合下采样模块和密集语义流监督模块的 U 型卷积神经网络 DSFA-Net,减少了下采样过程中的信息丢失,聚合了隐藏层不同分辨率图像之间的细节和语义信息。消融实验的结果表明,本文提出的两个模块能有效缓解数据不平衡问题,较好地克服了小目标分割和边界模糊难题;对比实验的结果表明,本文提出的 DSFA-Net 网络相比其他网络更具优势。

虽然提出的 DSFA-Net 网络在高度近视条纹损伤分割任务中取得了较好的结果,但是分割性能仍

有较大的提升空间。首先,由于采用的眼底 OCT 图像数据有限,分割网络的泛化性能有待进一步提升;其次,由于条纹损伤区域较小,本文方法虽然较好地缓解了数据分布严重不平衡问题,但仍需研究如何进一步解决该问题。

因此,针对以上问题,在后续的研究工作中将通过采集更多 OCT 数据或者采用传统、深度学习方法生成更多 OCT 数据来对数据集进行扩充,使得分割网络得到充分的学习和训练;同时尝试将目标检测与语义分割任务相结合,以进一步解决数据不平衡问题,进而提高网络的分割性能和泛化能力。

参考文献(References)

Badrinarayanan V, Kendall A and Cipolla R. 2017. SegNet: a deep convolutional encoder-decoder architecture for image segmentation.

- IEEE Transactions on Pattern Analysis and Machine Intelligence , 39(12) : 2481-2495 [DOI: 10.1109/TPAMI.2016.2644615]
- Boureau Y L , Bach F , LeCun Y and Ponce J. 2010. Learning mid-level features for recognition//Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco , USA: IEEE: 2559-2566 [DOI: 10.1109/CVPR.2010.5539963]
- Boer P D , Kroese D P , Mannor S and Rubinstein R Y. 2005. A tutorial on the cross-entropy method. *Annals of operations research* , 134(1) : 19-67 [DOI: 10.1007/s10479-005-5724-z]
- Chen J N , Lu Y Y , Yu Q H , Luo X D , Adeli E , Wang Y , Lu L , Yuille A L and Zhou Y Y. 2021. TransUNet: transformers make strong encoders for medical image segmentation [EB/OL]. [2021-07-19]. <https://arxiv.org/pdf/2102.04306.pdf>
- Chen L C , Papandreou G , Schroff F and Adam H. 2017. Rethinking atrous convolution for semantic image segmentation [EB/OL]. [2021-07-19]. <https://arxiv.org/pdf/1706.05587.pdf>
- Fang Y X , Yokoi T , Nagaoka N , Shinohara K , Onishi Y , Ishida T , Yoshida T , Xu X , Jonas J B and Ohno-Matsui K. 2018. Progression of myopic maculopathy during 18-year follow-up. *Ophthalmology* , 125(6) : 863-877 [DOI: 10.1016/j.ophtha.2017.12.005]
- Feng S L , Zhao H M , Shi F , Cheng X N , Wang M , Ma Y H , Xiang D H , Zhu W F and Chen X J. 2020. CPFNet: context pyramid fusion network for medical image segmentation. *IEEE Transactions on Medical Imaging* , 39(10) : 3008-3018 [DOI: 10.1109/TMI.2020.2983721]
- Gao Z T , Wang L M and Wu G S. 2019. LIP: local importance-based pooling//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul , Korea (South) : IEEE: 3354-3363 [DOI: 10.1109/ICCV.2019.00345]
- Gu Z W , Cheng J , Fu H Z , Zhou K , Hao H Y , Zhao Y T , Zhang T Y , Gao S H and Liu J. 2019. CE-Net: context encoder network for 2D medical image segmentation. *IEEE Transactions on Medical Imaging* , 38(10) : 2281-2292 [DOI: 10.1109/TMI.2019.2903562]
- Hou Q B , Zhang L , Cheng M M and Feng J S. 2020. Strip pooling: rethinking spatial pooling for scene parsing//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle , USA: IEEE: 4002-4011 [DOI: 10.1109/CVPR42600.2020.00406]
- Huang S S , Zheng Y F , Foster P J , Huang W Y and He M G. 2009. Prevalence and causes of visual impairment in Chinese adults in urban southern China: the Liwan Eye Study. *Archives of Ophthalmology* , 127(10) : 1362-1367 [DOI: 10.1001/archophthalmol.2009.138]
- Jaderberg M , Simonyan K , Zisserman A and Kavukcuoglu K. 2015. Spatial transformer networks//Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal , Canada: MIT Press: 2017-2025
- Jiang H J , Chen X J , Shi F , Ma Y H , Xiang D H , Ye L , Su J Z , Li Z Y , Chen Q Y , Hua Y H , Xu X , Zhu W F and Fan Y. 2019. Improved cGAN based linear lesion segmentation in high myopia IC-GA images. *Biomedical Optics Express* , 10(5) : 2355-2366 [DOI: 10.1364/BOE.10.002355]
- Li X T , You A S , Zhu Z , Zhao H L , Yang M K , Yang K Y , Tan S H and Tong Y H. 2020. Semantic flow for fast and accurate scene parsing//Proceedings of the 16th European Conference on Computer Vision. Glasgow , UK: Springer: 775-793 [DOI: 10.1007/978-3-030-58452-8_45]
- Mariotti S P , Kocur I , Resnikoff S , Jong M , Naidoo K S , He M G , Holden B A , Salomão S R , Sankaridurg P , Jonas J B , Saw S M , Smith E L III , Kedir J , Trier K , Wong T Y , Minto H , Yekta A A , Vitale S , Morgan I G , Ohno-Matsui K , Pärssinen O , Rao G and Zhao J L. 2015. The impact of myopia and high myopia: report of the Joint World Health Organization-Brien Holden Vision Institute Global Scientific Meeting on Myopia [EB/OL]. [2021-07-19]. <https://www.researchgate.net/publication/318216691>
- Marr D and Vaina L. 1982. Representation and recognition of the movements of shapes. *Proceedings of the Royal Society of London. Series B. Biological Sciences* , 214(1197) : 501-524 [DOI: 10.1098/rspb.1982.0024]
- Milletari F , Navab N and Ahmadi S A. 2016. V-Net: fully convolutional neural networks for volumetric medical image segmentation//Proceedings of the 4th International Conference on 3D Vision. Stanford , USA: IEEE: 565-571 [DOI: 10.1109/3DV.2016.79]
- Nagi J , Ducatelle F , Di Caro G A , Cireşan D , Meier U , Giusti A , Nagi F , Schmidhuber J and Gambardella L M. 2011. Max-pooling convolutional neural networks for vision-based hand gesture recognition//Proceedings of 2011 IEEE International Conference on Signal and Image Processing Applications. Kuala Lumpur , Malaysia: IEEE: 342-347 [DOI: 10.1109/ICSIPA.2011.6144164]
- Ohno-Matsui K , Yoshida T , Futagami S , Yasuzumi K , Shimada N , Kojima A , Tokoro T and Mochizuki M. 2003. Patchy atrophy and lacquer cracks predispose to the development of choroidal neovascularisation in pathological myopia. *British Journal of Ophthalmology* , 87(5) : 570-573 [DOI: 10.1136/bjo.87.5.570]
- Oktay O , Schlemper J , Le Folgoc L , Lee M , Heinrich M , Misawa K , Mori K , McDonagh S , Hammerla N Y , Kainz B , Glocker B and Rueckert D. 2018. Attention U-Net: learning where to look for the pancreas [EB/OL]. [2021-07-19]. <https://arxiv.org/pdf/1804.03999.pdf>
- 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]
- Ruder S. 2016. An overview of gradient descent optimization algorithms [EB/OL]. [2021-07-19]. <https://arxiv.org/pdf/1609.04747.pdf>

- Salamon J and Bello J P. 2017. Deep convolutional neural networks and data augmentation for environmental sound classification. *IEEE Signal Processing Letters*, 24(3): 279-283 [DOI: 10.1109/LSP.2017.2657381]
- Shelhamer E, Long J and Darrell T. 2017. Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4): 640-651 [DOI: 10.1109/tpami.2016.2572683]
- Shinohara K, Moriyama M, Shimada N, Tanaka Y and Ohno-Matsui K. 2014. Myopic stretch lines: linear lesions in fundus of eyes with pathologic myopia that differ from lacquer cracks. *Retina*, 34(3): 461-469 [DOI: 10.1097/IAE.0b013e3182a6b494]
- Tang Y T, Wang X F, Wang J C, Huang W, Gao Y P, Luo Y and Lu Y. 2015. Prevalence and causes of visual impairment in a Chinese adult population: the Taizhou Eye Study. *Ophthalmology*, 122(7): 1480-1488 [DOI: 10.1016/j.ophtha.2015.03.022]
- Tokoro T. 1988. On the definition of pathologic myopia in group studies. *Acta Ophthalmologica*, 66(S185): 107-108 [DOI: 10.1111/j.1755-3768.1988.tb02681.x]
- Wang L H, Huang W Y, He M, Zheng Y F, Huang S S, Liu B, Jin L, Congdon N G and He M G. 2013. Causes and five-year incidence of blindness and visual impairment in urban southern China: the Liwan Eye Study. *Investigative Ophthalmology and Visual Science*, 54(6): 4117-4121 [DOI: 10.1167/iovs.13-11911]
- Wang L W, Lee C Y, Tu Z W and Lazebnik S. 2015. Training deeper convolutional networks with deep supervision [EB/OL]. [2021-07-19]. <https://arxiv.org/pdf/1505.02496.pdf>
- Zhang X L, Fu P F, Zhao Y J, Xie H and Wang W R. 2020. Point cloud data classification and segmentation model using graph CNN and different pooling functions. *Journal of Image and Graphics*, 25(6): 1201-1208 (张新良, 付鹏飞, 赵运基, 谢恒, 王琬如. 2020. 融合图卷积和差异性池化函数的点云数据分类分割模型. *中国图象图形学报*, 25(6): 1201-1208) [DOI: 10.11834/jig.190367]

- Zhao H S, Shi J P, Qi X J, Wang X G and Jia J Y. 2017. Pyramid scene parsing network//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 6230-6239 [DOI: 10.1109/CVPR.2017.660]

作者简介



谭晓, 1998年生, 男, 硕士研究生, 主要研究方向为医学图像处理与分析、图像分割。
E-mail: txiao_corey@163.com



朱伟芳, 通信作者, 女, 副教授, 主要研究方向为医学图像处理与分析、模式识别、机器学习。
E-mail: wfzhu@suda.edu.cn

刁逸超, 男, 硕士研究生, 主要研究方向为医学影像处理与分析。E-mail: 457805700@qq.com

陈新建, 男, 教授, 主要研究方向为医学图像处理与分析及其在临床诊断中的应用。E-mail: xjchen@suda.edu.cn

石霏, 女, 副教授, 主要研究方向为医学图像处理与分析、模式识别、时频分析。E-mail: shifei@suda.edu.cn

樊莹, 女, 主任医师, 主要研究方向为眼底病。
E-mail: mdfanying@sjtu.edu.cn

谢嘉旻, 女, 博士研究生, 主要研究方向为眼底病。
E-mail: galanthus_nivalis@126.com