

基于自适应补偿网络的视网膜血管分割

张林¹, 吴闯¹, 范心宇¹, 官朝举^{1,2}, 李甦雁^{1,2}, 刘辉^{1*}¹中国矿业大学信息与控制工程学院, 江苏 徐州 221116;²徐州市第一人民医院眼科, 江苏 徐州 221116

摘要 基于眼底图像的视网膜血管精确分割对眼科疾病诊断意义重大。但视网膜血管结构高度复杂,多尺度及前、背景比例失衡,自动分割困难。因此,本文提出自适应补偿网络(SACom)实现端到端的视网膜血管精确分割。SACom以U型网络为基本框架,首先在编码器端引入可变形卷积提高复杂血管结构信息学习能力;然后在U型网络底部设计自适应多尺度对齐上下文模块提取并聚合多尺度上下文信息,对齐上下文特征;最后在解码器端设计协同补偿分支,融合多级输出提升模型的映射能力,实现精细分割。实验结果表明,SACom可有效提高视网膜血管的分割精度,在DRIVE、CHASE_DB1和STARE三个公共数据集上的准确率分别达到0.9695、0.9763和0.9753,灵敏度分别达到0.8403、0.8748和0.8506,曲线下面积(AUC)分别达到0.9880、0.9917和0.9919。

关键词 图像处理; 视网膜血管; 可变形卷积; 上下文对齐; 特征自适应融合

中图分类号 TP391 **文献标志码** A

DOI: 10.3788/AOS230599

1 引言

眼睛是人类重要的视觉器官,随着老龄化程度加深及全民近视发病率增长,眼科疾病患者逐年增加,早期筛查有助于降低眼科疾病的发病率^[1]。视网膜血管的形态信息被发现与糖尿病性视网膜病变、青光眼等多种眼科疾病密切相关,已广泛应用于临床诊断^[2]。眼底图像以非入侵的方式呈现视网膜血管等眼底组织结构,广泛应用于眼科临床^[3]。但眼底图像本身存在噪声、对比度低和血管背景像素比例不平衡等问题,其中视网膜血管错综复杂,具有高曲率、多尺度等特点,形态信息获取较难。医生手工标注存在费时费力、主观性强等问题,无法满足临床大量患者快速诊断的需求^[4]。因此,基于眼底图像的血管自动精确分割一直都是业内的研究热点。

近年来,国内外众多学者已设计大量的视网膜血管自动分割方法,大体上可分为传统的和基于深度学习的方法。传统方法主要基于形态学。例如,Azzopardi等^[5]设计了选择性响应血管的滤波器结合阈值处理实现血管分割,但易受噪声影响。Neto等^[6]先通过自适应局部阈值法对血管进行粗分割,再通过曲率分析和形态学重建细化分割,但步骤繁琐、精度有限。Staal等^[7]和You等^[8]将视网膜血管的分割转化为像素级分类问题,前者结合图像脊线和K近邻分类来

实现,后者则结合径向投影和半监督自训练框架来实现。这些基于传统机器学习的方法依赖手工设计的特征,且易受图像质量和病变区域的影响,分割性能及泛化能力均有待提高。

随着深度学习的发展,卷积神经网络(CNN)主导的深度学习架构获得广泛应用。Fu等^[9]将视网膜血管分割视为边界检测任务,结合CNN和条件随机场实现血管分割,但存在严重的欠分割及断裂问题。Li等^[10]通过多次迭代U-Net构建IterNet,从分割的血管图像本身找到血管细节,但不能捕捉多尺度特征,且模型参数量太大。Guo等^[11]引入空间注意力模块强化重要特征,并使用结构化随机失活卷积块抑制过拟合,但血管边界和微小血管的识别不理想。Zhang等^[12]将循环神经网络集成到CNN以传递上下文,并提出基于块的损失权重映射纠正血管背景失衡,但忽略了全局上下文信息,导致灵敏度和准确率较低。综上,现有的视网膜血管分割方法在多尺度特征提取、上下文信息的利用、参数量和分割精度等方面仍有待改善。因此,本文针对视网膜血管形态复杂、存在多尺度特性、微小血管捕捉困难等问题,构建自适应补偿网络(SACom),该工作有望为医生提供更精准的眼底血管分割,辅助疾病诊断。

2 本文方法

SACom模型结构如图1所示。针对视网膜血管复

收稿日期: 2023-02-27; 修回日期: 2023-03-18; 录用日期: 2023-04-06; 网络首发日期: 2023-05-08

基金项目: 国家自然科学基金(61971422)、徐州市科技创新计划-社会发展重点专项(KC22112, KC21153)

通信作者: *hui.liu@cumt.edu.cn

杂多变的形态结构,使用可变形卷积提升模型学习结构信息的能力;针对图像上下文信息的提取和处理,设计自适应多尺度对齐上下文模块 (AMAC),提取、聚合多尺度的对齐的上下文信息,引导血管更好地重建;为了

充分利用网络的中间特征,提出协同补偿分支(CCB),其多级输出分别有助于定位血管主要结构和微小血管细节,通过特征层平均化自适应融合与解码器末端输出特征层融合补充,以生成更加精确的分割结果。

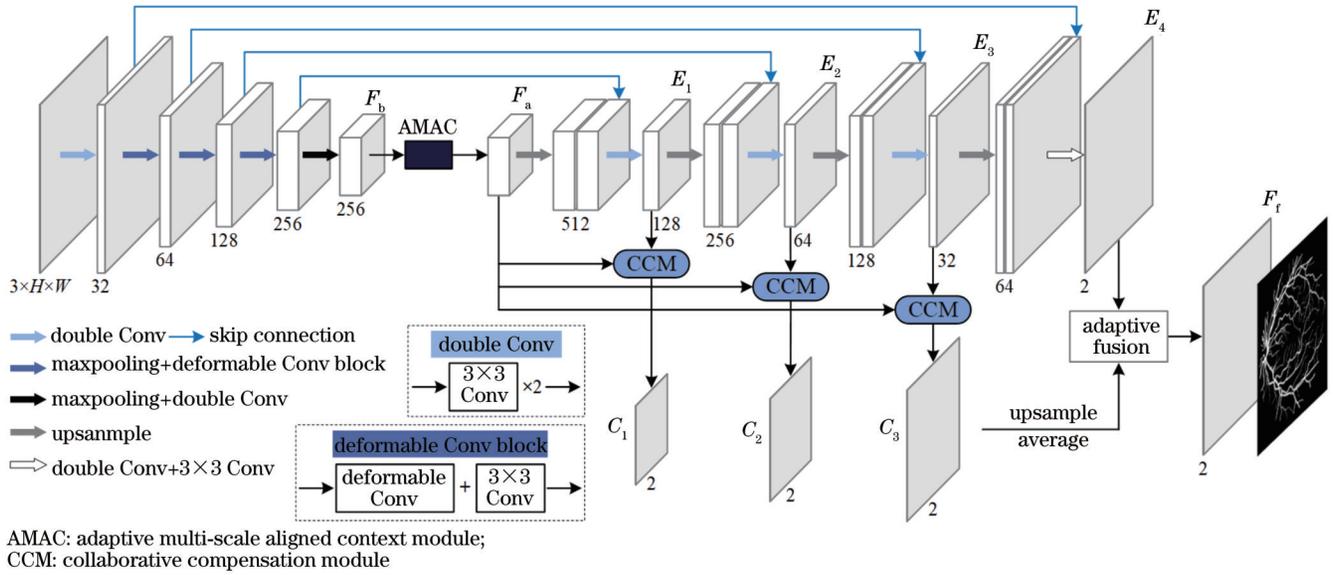


图1 SACom模型整体结构
Fig. 1 Overall structure of SACom model

2.1 可变形卷积

常规卷积固有的卷积核结构限制了模型的几何变换,不利于形态复杂的视网膜血管的建模。可变形卷积通过引入额外的偏移量改变常规卷积固定的空间采样位置,形成可根据目标尺度和形状自适应调整的任意形状卷积核^[13],能够更好地学习视网膜血管的结构信息、捕捉微小血管细节。给定 $K \in \{(-1, -1), (-1, 0), \dots, (0, 1), (1, 1)\}$ 为一个膨胀率为1的 3×3 卷积核,网格 K 定义了规则的感受野。对于可变形卷积输出特征图 y 上的每个位置 p_0 :

$$y(p_0) = \sum_{p_i \in K} W(p_i) \cdot X(p_0 + p_i + \Delta p_i), \quad (1)$$

式中: X 为输入特征图; $W(p_i)$ 为各个采样位置 p_i 上对应的权值, p_i 枚举了网格 K 中的所有位置。由于卷积核中每个参数加入了偏移量 Δp_i , 常规卷积采样网格变成了不规则和偏移的位置 $p_i + \Delta p_i$, 从而具有自适应确定尺度或感受野大小的能力。偏移量 Δp_i 是通过另外一个平行的常规卷积单元计算得到, 如图2所示。

可变形卷积的优越性来源于其可以适应物体几何形变的能力, 空间支持比常规卷积更符合目标结构。但采样范围可能会超出感兴趣的区域, 导致特征受到无关内容的影响, 因此加入一个调节机制, 自由地调整空间支持区域^[14]。带调节机制的可变形卷积为

$$y(p_0) = \sum_{p_i \in K} W(p_i) \cdot X(p_0 + p_i + \Delta p_i) \cdot \Delta m_i, \quad (2)$$

式中, $\Delta m_i (0 \leq \Delta m_i \leq 1)$ 为需要学习的调节标量, 可以

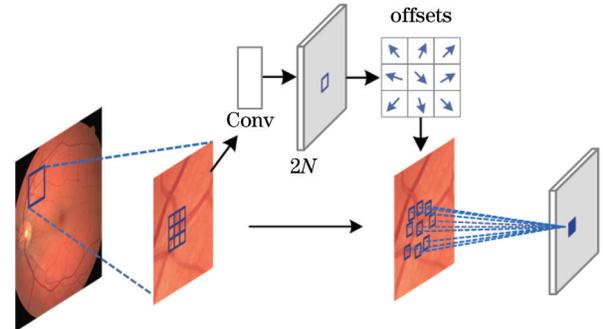


图2 可变形卷积
Fig. 2 Deformable convolution

对学习特征幅度进行调节。

2.2 自适应多尺度对齐上下文模块

上下文信息可以为图像分割提供丰富的语义指导, 每一个像素点都与其相邻像素存在关联, 在识别不同尺度大小的目标时需捕捉像素之间不同距离的依赖关系。金字塔池化模块^[15]、空间金字塔池化模块^[16]均以非自适应的方式提取多尺度上下文信息, 当输入图像大小改变时则难以匹配目标尺寸。考虑到插值算法上采样、残差连接引入的特征不对齐, 以及特征聚合过程中不加区分的上下文信息融合造成的错位, 最终会导致错误分类问题^[17-18], Huang等^[18]提出的对齐上下文模块将池化所得上下文特征进行了对齐, 但并非多尺度的。

针对眼底图像中视网膜血管的多尺度特性, 同时

为解决上述问题,本文提出一种 AMAC 模块。通过设置多个自适应平均池化分支来提取多尺度的上下文特征,根据输入特征图大小自适应调整各分支输出尺寸以解决输入图片大小改变时感受野和目标尺寸不匹配问题,如图 3 所示。4 个平均池化分支输出特征图的边

长分别设置为 $b_i = \frac{\sqrt{h \cdot w}}{n_i}$ ($i=1, 2, 3, 4$), 其中 h, w

为输入特征图 F_b 的高和宽, n_i 为从 2 开始依次增大且能被 $\sqrt{h \cdot w}$ 整除的 4 个整数。例如,当 AMAC 模块输入特征图的大小为 30×30 时, n_i 分别为 2、3、5、6, 4 个池化分支的输出尺寸分别为 15×15 、 10×10 、 6×6 、 5×5 。

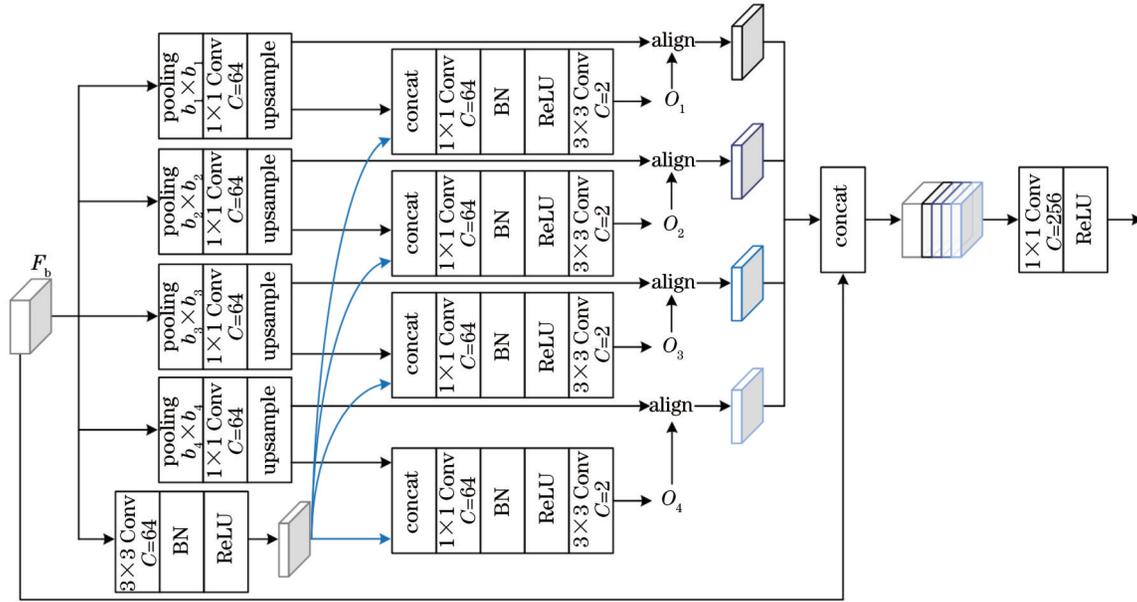


图 3 自适应多尺度对齐上下文模块

Fig. 3 Adaptive multi-scale aligned context module

输入特征图 F_b 依次经过卷积、批量归一化(BN)和修正线性单元(ReLU),得到局部特征 $F_i \in \mathbb{R}^{64 \times h \times w}$ 。双线性插值上采样后的各分支上下文信息特征图分别与 F_i 通道方向拼接,再通过一系列卷积层学习位置偏移量 $O_i \in \mathbb{R}^{2 \times h \times w}$,由于每个位置有两个方向的偏移,所以 O_i 通道数为 2。位置偏移量用于上下文特征的对齐,最后将所有分支输出和输入特征图叠加实现多尺度上下文信息的融合,如下式所示:

$$F_{\text{fusion}} = \text{concat} [F_b, A(F_i, O_i)], \quad (3)$$

式中: $A(\cdot)$ 为上下文特征对齐操作; $\text{concat}(\cdot)$ 为通道方向级联; F_{fusion} 为再经 1×1 卷积调整通道数和 ReLU 得到的输出 F_a 。AMAC 模块不同于以往对池化生成的上下文特征上采样后直接拼接融合的做法,而是通过学习二维偏移量来引导上下文特征根据局部特征调整空间位置,从而为每个像素选择由池化获得的上下文,再进行聚合。多尺度地对齐上下文信息,有助于血管结构的重建,同时解决血管的连通性问题。

2.3 CCB

编码器-解码器结构的卷积神经网络,通常把解码器最后一个卷积层的输出特征图用于最终预测^[19-20]。但解码器中存在多个卷积层输出,它们是编码器中浅层特征和解码器中高级特征融合的结果,具有丰富的

语义信息,被利用程度较低。为了充分利用解码器各层输出的特征,同时复用网络底部高级语义特征,本文提出了 CCB。如图 4 所示,解码器中不同阶段特征图 E_i ($i=1, 2, 3$) 分别与 F_a 作为各分支协同补偿模块 (CCM) 的输入,进一步提取特征得到输出 $\{C_1, C_2, C_3\}$,然后单独与调整到相应尺寸大小的真实标签计算损失,进行反向传播,从而提升各分支输出的准确性。 C_i ($i=1, 2, 3$) 与解码器末端输出特征图 E_4 的分辨率依次增大,对应逐步细化的血管结构,它们相互融合补充能够实现视网膜血管更加全面的分割。

CCM 对输入特征图 F_a 进行两次卷积,以消除与解码器中特征图 E_i ($i=1, 2, 3$) 之间的语义隔阂,每次卷积后依次执行 BN 和 ReLU,其中第 2 个卷积为 1×1 卷积,输出通道数分别与输入 E_i ($i=1, 2, 3$) 相同;上采样后与对应输入 E_i ($i=1, 2, 3$) 进行通道方向拼接,再经过一次卷积输出通道数为 2、分辨率与 E_i ($i=1, 2, 3$) 相同的特征图 C_i ($i=1, 2, 3$)。

2.4 特征层平均化自适应融合

不同于相加、取平均、通道方向拼接等常规的特征图聚合方式,本文受文献^[21]的启发,让模型自适应地学习不同特征层对于最终输出的贡献度。不同之处在于:1) 融合对象并非只针对解码器中特征层,而是 CCB 的各级输出以及解码器末端特征层;2) 参与融合

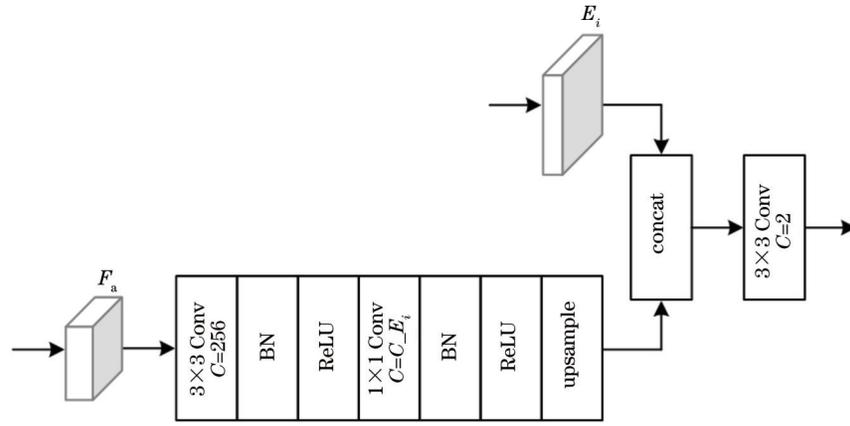


图 4 协同补偿模块

Fig. 4 Collaborative compensation module

的所有特征层并非直接进行自适应融合,而是CCB的各级输出特征层先经上采样和平均化,再与解码器末端输出 \$E_4\$ 进行自适应融合,生成最终用于预测像素级分割结果的特征图 \$F_f\$,如图 5 所示。该过程可表示为

$$F_f = \alpha \cdot \frac{1}{3} \sum_{i=1}^3 U(C_i) + \beta \cdot E_4 \quad (4)$$

CCB 不同层级的输出特征图 \$C_i (i=1, 2, 3)\$,经双线性上采样 \$U\$ 调整分辨率到输入图片大小后求平均。结合自适应权重参数 \$\alpha (0 \leq \alpha \leq 1)\$、\$\beta (0 \leq \beta \leq 1)\$,最终得到特征层 \$F_f\$。为了确保在训练阶段的数值稳定,本文设定 \$\alpha, \beta\$ 的初值均为 0.5。

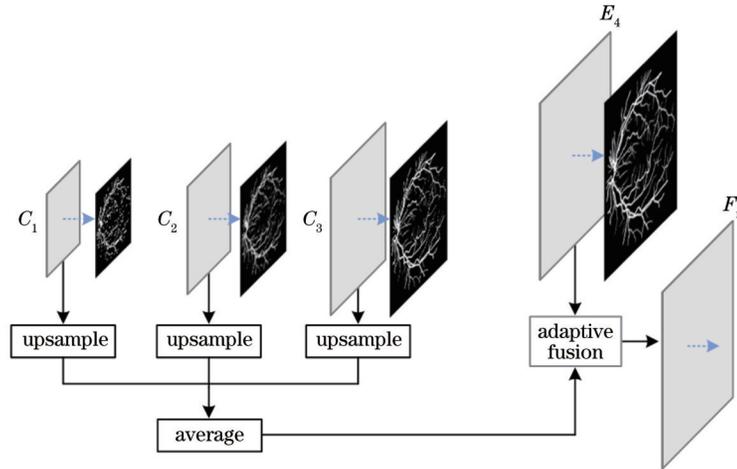


图 5 特征层平均化自适应融合

Fig. 5 Feature layer averaging adaptive fusion

2.5 损失函数

眼底图像中血管像素和背景像素严重不平衡,导致血管特征学习困难,带权重的交叉熵损失可以缓解该问题,定义为

$$L_{wce} = -\frac{1}{N} \sum_{n=1}^N [W_f \cdot y_n \log_2 p_n + (1 - y_n) \log_2 (1 - p_n)] \quad (5)$$

式中: \$y_n \in \{0, 1\}\$ 为真实标签; \$p_n \in [0, 1]\$ 为像素为血管类别的预测概率; \$N\$ 为给定图片像素点的总个数; \$W_f\$ 为血管像素的权重,本文设置为 2.0。

Dice 损失函数^[22]本质上是通过衡量两个样本的相似度,实现对样本的监督,同样适用于血管像素和背景像素比例不平衡的情况,定义为

$$L_{dice} = 1 - \frac{2 \sum_{n=1}^N |p_n y_n|}{\sum_{n=1}^N |p_n| + \sum_{n=1}^N |y_n|} \quad (6)$$

式中, \$y_n \in \{0, 1\}\$ 既可以是前景的真实标签,也可以是背景的真实标签。实际计算 Dice 损失时,1 减去的是血管和背景两个通道上 Dice 系数的平均值。Dice 损失函数侧重于挖掘前景区域,但对小目标不友好,且容易不稳定,极端情况下还会出现梯度饱和现象。因此,本文同时采用带权重的交叉熵损失函数和 Dice 损失函数对网络训练过程进行监督,以达到更好的效果。最终损失函数 \$L\$ 可表示为

$$L = L_{wce} + L_{dice} \quad (7)$$

3 实验结果及分析

3.1 数据集

本文在 DRIVE、CHASE_DB1、STARE 三个公共数据集上开展实验以验证 SACom 的分割性能及鲁棒性。其中, DRIVE 数据集包含 40 张 $565 \text{ pixel} \times 584 \text{ pixel}$ 的眼底图像, 划分成训练集和测试集各 20 张, 训练集中每张图像只有 1 号专家的标注, 测试集每张图像则有 1 号和 2 号两位专家标注; CHASE_DB1 数据集包含从 14 名儿童双眼采集的 28 张眼底图像, 图像分辨率为 $999 \text{ pixel} \times 960 \text{ pixel}$, 实验中将前 20 张划分为训练集, 剩余的 8 张作为测试集; STARE 数据集包含 20 张分辨率为 $700 \text{ pixel} \times 605 \text{ pixel}$ 的眼底图像, 采用 2 折交叉验证进行实验。CHASE_DB1 和 STARE 中图像均有 2 位专家的标注, 本文采用 1 号专家标注作为真实标签完成算法的性能评估。

此外, 考虑到数据集中样本量较小, 本文通过在线数据增强的方式增加训练样本的多样性, 提高模型泛化能力、抑制过拟合。具体地, 在训练数据输入网络前对图像进行一系列随机变换, 包括随机的尺寸缩放、颜色抖动、水平和垂直方向翻转, 以及在 $[-\pi, \pi]$ 角度范围内随机旋转, 最后再随机剪裁到相应大小。其中, DRIVE 数据集图像的分辨率随机剪裁至 $480 \text{ pixel} \times 480 \text{ pixel}$, CHASE_DB1 数据集图像随机剪裁至 $832 \text{ pixel} \times 832 \text{ pixel}$, STARE 数据集图像随机剪裁至 $576 \text{ pixel} \times 576 \text{ pixel}$, 以满足输入图像的边长为下采样率 16 的整数倍。

3.2 实验细节

模型实现基于 PyTorch 框架, 计算机处理器为 Intel(R) Xeon(R) W-2175 CPU @ 2.50 GHz, 64 GB 内存, 搭配两张显卡 NVIDIA GeForce RTX 2080Ti GPU。模型训练初始学习率设置为 10^{-2} , 采用 SGD 优化器, 同时使用 0.9 的动量和 10^{-4} 的权重衰减。参考“ploy”学习率策略, 初始学习率乘以 $1 - \left(\frac{\text{iter}}{\text{max_iter}}\right)^{\text{power}}$, 其中 power 为 0.9, iter 为当前训练迭代步数, max_iter 为整个训练过程中所需迭代的步数^[23]。同时, 训练阶段 DRIVE 和 STARE 数据集的批量大小设置为 2, 训练 600 轮。CHASE_DB1 数据集的批量大小设置为 2, 共训练 500 轮。

3.3 评估指标

本文通过准确率 (A_{cc})、灵敏度 (S_e)、特异度 (S_p)、 F_1 分数和曲线下面积 (AUC) 等重要指标来评估模型的分割性能, 具体计算方法如下:

$$A_{cc} = (S_{TP} + S_{TN}) / (S_{TP} + S_{TN} + S_{FP} + S_{FN}), \quad (8)$$

$$S_e = S_{TP} / (S_{TP} + S_{FN}), \quad (9)$$

$$S_p = S_{TN} / (S_{TN} + S_{FP}), \quad (10)$$

$$F_1 = 2S_{TP} / (2S_{TP} + S_{FP} + S_{FN}), \quad (11)$$

式中: S_{TP} 是真阳性的变量, 对应正确分割的血管像素点个数; S_{TN} 是真阴性的变量, 对应正确分割背景像素点个数; 相应地, S_{FP} 和 S_{FN} 为假阳性和假阴性的变量, 分别表示错误分割为血管的背景像素个数和错误分割为背景的血管像素个数; F_1 分数被定义为模型精确率和召回率的调和平均值, AUC 被定义为受试者工作特征 (ROC) 曲线下与坐标轴围成的面积, 二者的值越接近 1, 模型的分割性能越好。

3.4 模块及策略的有效性验证

以 DRIVE 数据集为例进行对比实验, 探究 L_{wcc} 中血管像素不同权重对模型分割性能的影响, 证明针对眼底图像视网膜血管分割提出模型中各模块和策略均为最优。

3.4.1 血管像素权重对性能的影响

为确定加权交叉熵损失函数 L_{wcc} 中的权重, 本文在给定 $W_i = \{1.0, 1.5, 2.0, 2.5, 3.0\}$ 不同权重情况下, 对 DRIVE 数据集进行分割, 结果如表 1 所示。可以看出, 当权重 W_i 较大时, 模型灵敏度提高, 但特异度较低; 反之, 当权重 W_i 较小时, 模型表现出更高的特异度, 但灵敏度降低。这是因为较大的 W_i 会加强模型对血管像素的监督, 其错误分割会导致更大的惩罚, 从而提高血管像素的分类能力, 但同时带来更高的假阳性问题, 导致特异度降低; 而由于背景像素远多于血管像素, 导致权重 W_i 较小时准确率较高的现象。当 W_i 为 2.0 时, 可获得 F_1 和 AUC 均最高的综合性能, 因此, 本文最终设置 W_i 为 2.0。

表 1 不同权重下的性能比较

Table 1 Performance comparison under different weights

Weight	A_{cc}	S_e	S_p	F_1	AUC
1.0	0.9703	0.8002	0.9866	0.8252	0.9878
1.5	0.9700	0.8156	0.9848	0.8263	0.9875
2.0	0.9695	0.8403	0.9819	0.8285	0.9880
2.5	0.9681	0.8549	0.9790	0.8243	0.9874
3.0	0.9669	0.8723	0.9759	0.8218	0.9878

3.4.2 不同监督和融合策略的比较

CCB 各级输出 $C_i (i=1, 2, 3)$ 先上采样到输入图像大小, 再与解码器最后一个卷积层输出 E_4 进行特征融合。针对 4 种不同的特征融合策略: 1) $C_i (i=1, 2, 3)$ 上采样特征图和 E_4 直接进行自适应融合得到最终输出, 4 个特征图的权重初始值均为 0.25; 2) $C_i (i=1, 2, 3)$ 上采样特征图和 E_4 一起平均化输出; 3) $C_i (i=1, 2, 3)$ 上采样后特征图先取平均, 再和 E_4 进行自适应融合, 初始权重为 0.5, 且 $C_i (i=1, 2, 3)$ 上采样特征图、 E_4 以及融合后所得特征图 F_i 均与真实标签计算损失函数 L 进行联合监督; 4) 不同于策略 3 中 $C_i (i=1, 2, 3)$ 上采样后再计算损失函数, 而是 $C_i (i=1, 2, 3)$ 分别与调整到自身相同尺寸大小的真实标签计算损失函数 L , 其余操作均与策略 3 相同。定量对比结果如表 2

所示。

表 2 不同监督和融合策略的比较

Table 2 Comparison of different supervision and fusion strategies

Strategy	A_{cc}	S_e	S_p	F_1	AUC
1	0.9688	0.8374	0.9814	0.8246	0.9871
2	0.9684	0.8369	0.9810	0.8227	0.9868
3	0.9692	0.8397	0.9816	0.8269	0.9875
4	0.9695	0.8403	0.9819	0.8285	0.9880

由表 2 可得,本文模型采用策略 4 进行训练监督和特征融合时,各评估指标均为最佳,可以实现眼底图像视网膜血管分割的优异性能。

3.4.3 消融实验

对 7 个网络进行消融实验,采用上述的评估指标以及模型的参数量对比不同模型,结果如表 3 所示(其中, $U-Net_n$ 中的 n 表示 U-Net 的第一个卷积层输出通道数; DC_{ijk} 表示在编码器的第 i, j, k 个卷积块中引入可变形卷积; Baeline+ DC_{234} +AMAC+CCB 即为本文所提出的自适应补偿网络 SACom)。 $U-Net_{32}$ 在数据量较少的眼底图像数据集上表现出的性能优于 $U-Net_{64}$, 且参数量不足 $U-Net_{64}$ 的 1/4, 因此本文将 $U-$

Net_{32} 作为基线模型^[24](一些编码器-解码器结构的卷积神经网络第一个卷积层输出通道数为 64^[25-26], 经过若干次通道数迭代, 整个网络的参数量较大, 推理时间增加, 数据量较少时还容易导致过拟合现象的发生^[27])。可变形卷积有助于捕捉视网膜血管的结构信息, 但是部署到模型中不同位置时取得的效果也有差别。从表 3 结果可以看出, 在基线模型的基础上, 同样是 3 个可变形卷积层, 当部署在编码器的第 2、3、4 个卷积块中时, 模型的分割性能提升最多, 且参数量较少。以 AUC 为例, 相较基线的 0.9761 提升至 0.9802, 提升了 0.42%。在此基础上, AMAC 模块自适应地聚合多尺度对齐上下文信息, 有助于视网膜血管的重建, 模型的分割性能全面提升, 各评估指标均有提高。相比于仅引入可变形卷积, AUC 又提升了 0.48%。另一方面, CCB 充分利用解码器中特征的同时复用网络底部深层语义信息, 各层级输出对应逐步细化的血管结构, 与解码器末端输出特征图融合补充, 以生成更精确的视网膜血管分割图。引入 CCB 和上述训练监督及特征融合策略使得分割性能进一步提升, 且 AUC 达到 0.9880, 再次提升 0.31%。总之, SACom 各个模块均有利于提高分割性能, 且相较于基线模型仅增加少量参数。

表 3 消融实验结果对比

Table 3 Comparison of ablation experimental results

Model	A_{cc}	S_e	S_p	F_1	AUC	Params /M
$U-Net_{64}$	0.9551	0.7841	0.9800	0.8167	0.9758	17.26
$U-Net_{32}$ (Baseline)	0.9554	0.7849	0.9802	0.8175	0.9761	4.32
Baseline+ DC_{123}	0.9642	0.7883	0.9810	0.7939	0.9789	4.20
Baeline+ DC_{234}	0.9647	0.7914	0.9814	0.7972	0.9802	3.79
Baeline+ DC_{345}	0.9612	0.7869	0.9779	0.7802	0.9765	3.87
Baeline+ DC_{234} +AMAC	0.9667	0.8277	0.9801	0.8134	0.9849	4.17
Baeline+ DC_{234} +AMAC+CCB	0.9695	0.8403	0.9819	0.8285	0.9880	6.01

3.5 实验结果对比

3.5.1 分割效果对比

SACom 在三个数据集上的分割结果如图 6 所示, 图中虚线圆圈部分为代表性的局部效果对比。可见, 本文方法对粗血管和微小血管都能够较全面地识别, 血管的连通性也更为理想; 且对于不同的数据集, SACom 都能够较好地分割血管, 说明了其良好的泛化能力。此外, 人工标注有一定的误差, 专家 1 和专家 2 的标注存在微小差别, 因此, 采用不同的专家标注进行评估所得结果不同, 如表 4 所示。

进一步深入研究发现, 分割结果中还存在专家 1 的标注中未出现, 以及两位专家均未给予标注但在眼底图像中却真实存在的微小血管, 如图 7 中虚线圆圈区域所示。可以看出, SACom 具有较强的血管分割能力, 能够更好地识别血管像素, 避免了人工标注主观性较强的问题。

3.5.2 与其他先进方法对比

为了进一步证明 SACom 的有效性, 本文在 DRIVE、CHASE_DB1 和 STARE 三个公共数据集上与当前具有代表性的工作以及近年在视网膜血管分割任务上性能表现优秀的网络 $U-Net$ ^[24]、 $DUNet$ ^[28]、 $AA-UNet$ ^[29]、 $CSU-Net$ ^[30]、 $CMNet$ ^[21] 开展对比实验, 如表 5 所示, 其中加粗字体为最优。

由表 5 可见, 所提出模型在 DRIVE 数据集上各评估指标均为最优, 且相对其他方法在灵敏度和 AUC 等指标上有较大提升。其灵敏度为 0.8403, 比其他方法高出 4.11%~7.06%, 且 AUC 达到 0.9880。说明本文方法能够准确地分割血管, 具有较好的综合性能; SACom 凭借可变形卷积和 AMAC 模块较强的自适应能力以及整体方法的优越性, 在分辨率较高的 CHASE_DB1 数据集上分割性能仍然优秀。全局准确率较高的同时灵敏度高达 0.8748, 远超其他方法,

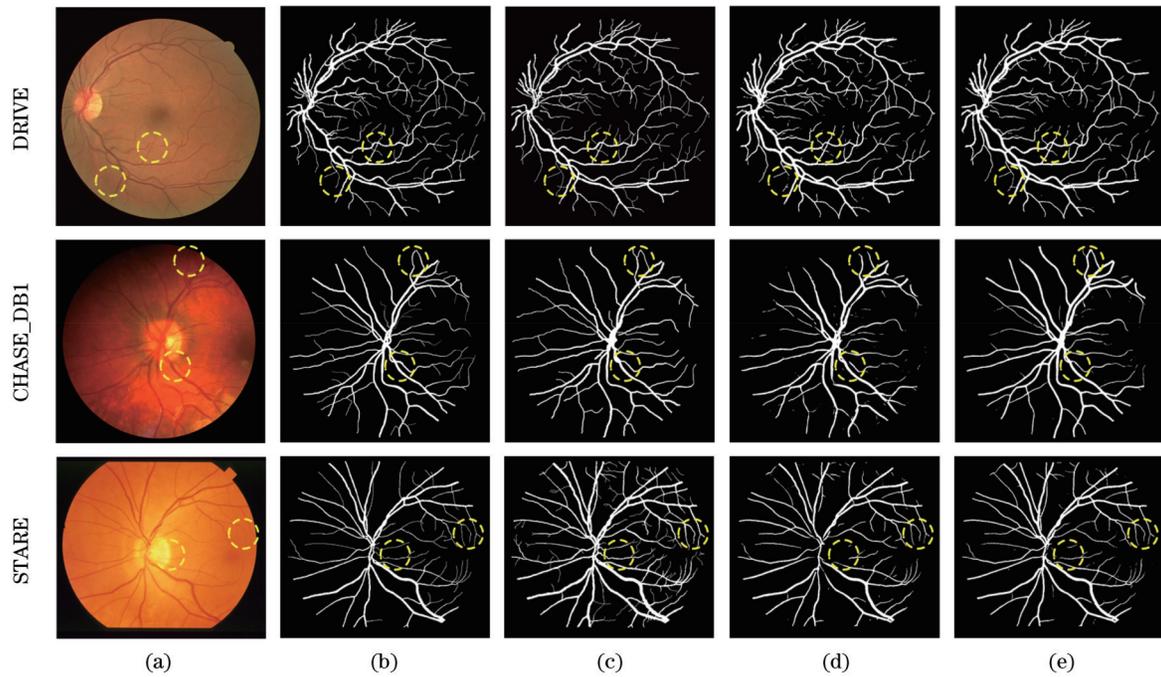


图 6 不同数据集上的可视化结果。(a)原图;(b)专家1的标注;(c)专家2的标注;(d)U-Net分割结果;(e)SCom分割结果
Fig. 6 Visualization results on different datasets. (a) Original images; (b) ground truth by expert 1; (c) ground truth by expert 2; (d) segmentation results of U-Net; (e) segmentation results of SCom

表 4 采用DRIVE测试集不同专家标注评估SCom性能对比
Table 4 Performance comparison of SCom with different expert annotations on DRIVE test set

Label	A_{cc}	S_c	S_p	F_1	AUC
1	0.9695	0.8403	0.9819	0.8285	0.9800
2	0.9737	0.8713	0.9832	0.8484	0.9910

且 F_1 、AUC 分别达到了 0.8229、0.9917, 同样具有显著优势。特异度指标略低于 CMNet 和 CSU-Net, 这

是由于 SCom 对于血管像素区分能力较强, 分割出了较多 1 号专家未标注到的微小血管, 导致出现较多相对于真实标签的假阳性, 但仍可以达到 0.9831; 在 STARE 数据集上, 本文方法的特异度指标虽略低于 U-Net 和 DUNet, 但是在其他评估指标上均远高于它们, 其中灵敏度分别比它们高出 11.36% 和 11.99%, F_1 分别高出 3.18% 和 3.08%。分析可知, 这是由于 U-Net 使用普通卷积和最大池化下采样的组合, 编解

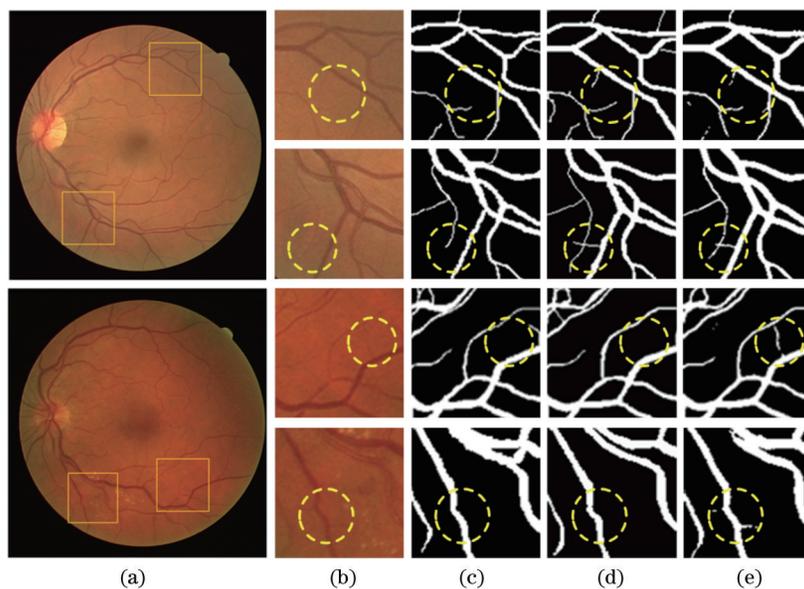


图 7 DRIVE 数据集分割结果细节对比。(a)原图;(b)局部图;(c)标签 1 局部图;(d)标签 2 局部图;(e)SCom 分割结果局部图
Fig. 7 Detail comparison of DRIVE dataset segmentation results. (a) Original images; (b) local images; (c) local images of ground truth 1; (d) local images of ground truth 2; (e) local images of SCom segmentation results

码过程中会丢失大量血管结构信息, DUNet 的数据处理方式是图像裁剪成块, 这会导致部分上下文信息丢失, 影响血管结构的重建。CSU-Net 的 F_1 在

所有方法里最高, 但其他指标都显著低于本文方法, 且本文方法的 AUC 达到了 0.9919, 比其高出 1.33%。

表 5 不同方法分割性能对比

Table 5 Segmentation performance comparison of different methods

Dataset	Method	A_{cc}	S_e	S_p	F_1	AUC
DRIVE	U-Net	0.9554	0.7849	0.9802	0.8175	0.9761
	DUNet	0.9566	0.7963	0.9800	0.8237	0.9802
	AA-UNet	0.9558	0.7941	0.9798	0.8216	0.9847
	CSU-Net	0.9565	0.8071	0.9782	0.8251	0.9801
	CMNet	0.9664	0.8069	0.9818	0.8064	0.9840
	SACom	0.9695	0.8403	0.9819	0.8285	0.9880
CHASE_DB1	U-Net	0.9576	0.7304	0.9717	0.7793	0.9782
	DUNet	0.9610	0.8155	0.9752	0.7883	0.9804
	AA-UNet	0.9608	0.8176	0.9704	0.7892	0.9865
	CSU-Net	0.9706	0.8427	0.9836	0.8105	0.9824
	CMNet	0.9742	0.8461	0.9836	0.8097	0.9879
	SACom	0.9763	0.8748	0.9831	0.8229	0.9917
STARE	U-Net	0.9636	0.7638	0.9867	0.8135	0.9789
	DUNet	0.9641	0.7595	0.9878	0.8143	0.9832
	AA-UNet	0.9640	0.7598	0.9878	0.8142	0.9824
	CSU-Net	0.9702	0.8432	0.9845	0.8516	0.9825
	SACom	0.9753	0.8506	0.9856	0.8394	0.9919

总之, 本文提出模型 SACom 的整体性能优于其他先进方法, 并具有较高的灵敏度, 而眼底图像中血管像素和背景像素比例失衡, 高灵敏度至关重要。较强

的泛化能力使其在多个数据集上都可以达到较好的分割效果。图 8 展示了不同方法在各数据集上的 AUC 值, 可以更加直观地比较出本文方法性能的优越性。

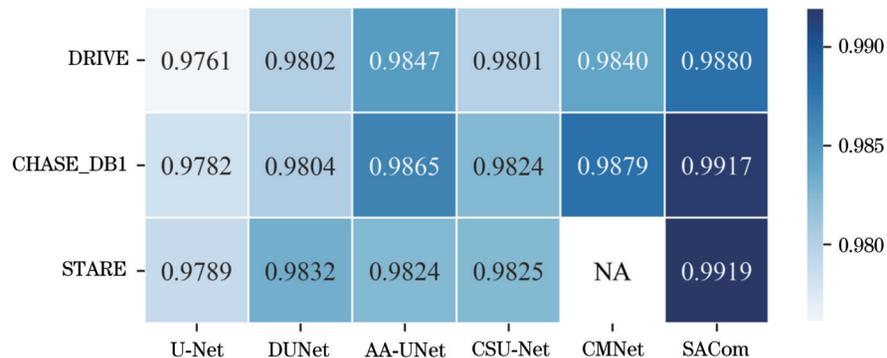


图 8 SACom 和其他先进方法 AUC 指标可视化

Fig. 8 Visualization of AUC for SACom and other advanced methods

4 结 论

针对眼底图像中视网膜血管精确分割的难题, 本文提出了一种高效的自动分割算法 SACom。SACom 借鉴 U-Net 的网络架构, 在编码器中引入可变形卷积以增强对血管结构信息的学习能力。U 型网络底部设计自适应多尺度对齐上下文模块, 提取和聚合多尺度的对齐上下文信息, 以适应视网膜血管的多尺度问题, 使得粗血管和微小血管都可以更好地得

到重建, 还有助于解决血管的连通性问题。随后, 为了充分利用解码器中特征层以及复用网络底部高级语义特征, 提出了 CCB, 将自适应多尺度对齐上下文模块输出特征层和解码器中特征层作为 CCM 的输入, 进一步提取特征得到不同分辨率的多级输出, 并采用分别与调整到相应尺寸的真实标签计算损失的训练监督策略。最后再将 CCB 各级输出取平均, 与解码器末端输出特征图自适应融合, 通过多级特征图相互补充以实现精细分割。在三个公共数据集上对

SACom 进行了评估,结果表明,该方法对不同像素分类的泛化能力较好,尤其是对血管像素的区分能力更强,在所有数据集上的综合分割性能均优于其他先进方法。此外,本文算法的参数数量较少,有利于部署和临床实际应用。但受图像亮度、对比度和病变区域的影响,分割结果中仍然存在少量血管分割不足、断裂等现象。未来将进一步优化算法实现更好的分割结果,以及探讨眼底图像视网膜血管的域适应分割算法。

参 考 文 献

- [1] Soomro T A, Khan T M, Khan M A U, et al. Impact of ICA-based image enhancement technique on retinal blood vessels segmentation[J]. *IEEE Access*, 2018, 6: 3524-3538.
- [2] Fraz M M, Remagnino P, Hoppe A, et al. Blood vessel segmentation methodologies in retinal images-a survey[J]. *Computer Methods and Programs in Biomedicine*, 2012, 108(1): 407-433.
- [3] Li T, Bo W, Hu C Y, et al. Applications of deep learning in fundus images: a review[J]. *Medical Image Analysis*, 2021, 69: 101971.
- [4] 薛文渲, 刘建霞, 刘然, 等. 改进 U 型网络的眼底视网膜血管分割方法[J]. *光学学报*, 2020, 40(12): 1210001.
Xue W X, Liu J X, Liu R, et al. An improved method for retinal vascular segmentation in U-net[J]. *Acta Optica Sinica*, 2020, 40(12): 1210001.
- [5] Azzopardi G, Strisciuglio N, Vento M, et al. Trainable COSFIRE filters for vessel delineation with application to retinal images[J]. *Medical Image Analysis*, 2015, 19(1): 46-57.
- [6] Neto L C, Ramalho G L B, Neto J F S R, et al. An unsupervised coarse-to-fine algorithm for blood vessel segmentation in fundus images[J]. *Expert Systems With Applications*, 2017, 78: 182-192.
- [7] Staal J, Abrámoff M D, Niemeijer M, et al. Ridge-based vessel segmentation in color images of the retina[J]. *IEEE Transactions on Medical Imaging*, 2004, 23(4): 501-509.
- [8] You X G, Peng Q M, Yuan Y, et al. Segmentation of retinal blood vessels using the radial projection and semi-supervised approach[J]. *Pattern Recognition*, 2011, 44(10/11): 2314-2324.
- [9] Fu H Z, Xu Y W, Lin S, et al. DeepVessel: retinal vessel segmentation via deep learning and conditional random field [M]//Ourselin S, Joskowicz L, Sabuncu M, et al. *Medical image computing and computer-assisted intervention-MICCAI 2016. Lecture notes in computer science*. Cham: Springer, 2016, 9901: 132-139.
- [10] Li L Z, Verma M, Nakashima Y, et al. IterNet: retinal image segmentation utilizing structural redundancy in vessel networks [C]//2020 IEEE Winter Conference on Applications of Computer Vision (WACV), March 1-5, 2020, Snowmass, CO, USA. New York: IEEE Press, 2020: 3645-3654.
- [11] Guo C L, Szemenyei M, Yi Y G, et al. SA-UNet: spatial attention U-net for retinal vessel segmentation[C]//2020 25th International Conference on Pattern Recognition (ICPR), January 10-15, 2021, Milan, Italy. New York: IEEE Press, 2021: 1236-1242.
- [12] Zhang Y, He M, Chen Z N, et al. Bridge-Net: context-involved U-net with patch-based loss weight mapping for retinal blood vessel segmentation[J]. *Expert Systems With Applications*, 2022, 195: 116526.
- [13] Dai J F, Qi H Z, Xiong Y W, et al. Deformable convolutional networks[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 764-773.
- [14] Zhu X Z, Hu H, Lin S, et al. Deformable ConvNets V2: more deformable, better results[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2020: 9300-9308.
- [15] Zhao H S, Shi J P, Qi X J, et al. Pyramid scene parsing network [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6230-6239.
- [16] 罗文劫, 韩国庆, 田学东. 多尺度注意力解析网络的视网膜血管分割方法[J]. *激光与光电子学进展*, 2021, 58(20): 2017001.
Luo W J, Han G Q, Tian X D. Retinal vessel segmentation method based on multi-scale attention analytic network[J]. *Laser & Optoelectronics Progress*, 2021, 58(20): 2017001.
- [17] Li X T, You A S, Zhu Z, et al. Semantic flow for fast and accurate scene parsing[M]//Vedaldi A, Bischof H, Brox T, et al. *Computer vision-ECCV 2020. Lecture notes in computer science*. Cham: Springer, 2020, 12346: 775-793.
- [18] Huang Z L, Wei Y C, Wang X G, et al. AlignSeg: feature-aligned segmentation networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(1): 550-557.
- [19] 蔡乾宏, 刘宇红, 张荣芬. 基于改进 U-Net 的两阶段视网膜血管分割方法[J]. *激光与光电子学进展*, 2021, 58(16): 1617002.
Cai Q H, Liu Y H, Zhang R F. Two-stage retinal vessel segmentation based on improved U-net[J]. *Laser & Optoelectronics Progress*, 2021, 58(16): 1617002.
- [20] 赵凤, 钟蓓蓓, 刘汉强. 基于注意力机制的多尺度残差 U-Net 眼底血管分割[J]. *激光与光电子学进展*, 2022, 59(18): 1810002.
- [21] Zhao F, Zhong B B, Liu H Q. Multi-scale residual U-Net fundus blood vessel segmentation based on attention mechanism [J]. *Laser & Optoelectronics Progress*, 2022, 59(18): 1810002.
- [22] 罗凌, 薛定宇, 冯兴隆. 基于紧凑混合网络的视网膜血管自动分割[J]. *控制与决策*, 2022, 37(2): 353-360.
Luo L, Xue D Y, Feng X L. Automatic segmentation of retinal vessel via compact mixed network[J]. *Control and Decision*, 2022, 37(2): 353-360.
- [23] Milletari F, Navab N, Ahmadi S A. V-net: fully convolutional neural networks for volumetric medical image segmentation[C]//2016 Fourth International Conference on 3D Vision (3DV), October 25-28, 2016, Stanford, CA, USA. New York: IEEE Press, 2016: 565-571.
- [24] Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(4): 834-848.
- [25] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation[M]//Navab N, Hornegger J, Wells W, et al. *Medical image computing and computer-assisted intervention-MICCAI 2015. Lecture notes in computer science*. Cham: Springer, 2015, 9351: 234-241.
- [26] Gu Z W, Cheng J, Fu H Z, et al. CE-net: context encoder network for 2D medical image segmentation[J]. *IEEE Transactions on Medical Imaging*, 2019, 38(10): 2281-2292.
- [27] Zhang Y, Fang J, Chen Y, et al. Edge-aware U-net with gated convolution for retinal vessel segmentation[J]. *Biomedical Signal Processing and Control*, 2022, 73: 103472.
- [28] 郑婷月, 唐晨, 雷振坤. 基于全卷积神经网络的多尺度视网膜血管分割[J]. *光学学报*, 2019, 39(2): 0211002.
Zheng T Y, Tang C, Lei Z K. Multi-scale retinal vessel segmentation based on fully convolutional neural network[J]. *Acta Optica Sinica*, 2019, 39(2): 0211002.
- [29] Jin Q G, Meng Z P, Pham T D, et al. DUNet: a deformable network for retinal vessel segmentation[J]. *Knowledge-Based Systems*, 2019, 178: 149-162.

- [29] Lü Y, Ma H, Li J N, et al. Attention guided U-net with atrous convolution for accurate retinal vessels segmentation[J]. IEEE Access, 2020, 8: 32826-32839.
- [30] Wang B, Wang S P, Qiu S, et al. CSU-net: a context spatial U-net for accurate blood vessel segmentation in fundus images[J]. IEEE Journal of Biomedical and Health Informatics, 2021, 25(4): 1128-1138.

Retinal Vessel Segmentation via Self-Adaptive Compensation Network

Zhang Lin¹, Wu Chuang¹, Fan Xinyu¹, Gong Chaoju^{1,2}, Li Suyan^{1,2}, Liu Hui^{1*}

¹*School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, Jiangsu, China;*

²*Department of Ophthalmology, The First People's Hospital of Xuzhou, Xuzhou 221116, Jiangsu, China*

Abstract

Objective Human eye is a crucial component of vision, but the number of patients suffering from ocular illnesses is growing every year. It has been discovered that the morphological characteristics of retinal blood vessels are strongly associated with several ocular conditions including diabetic retinopathy and glaucoma, and they are frequently employed in clinical diagnosis. Therefore, precise segmentation of retinal blood vessels based on color fundus images is crucial for the diagnosis of ocular illnesses. However, the fundus image itself displays noise, poor contrast, and an unbalanced distribution of blood vessels and background pixels. Additionally, morphological information gathering is challenging due to the delicate, highly curved, and multi-scale properties of retinal blood vessels. The time-consuming, difficult, and subjective nature of doctors' manual segmentation makes it ineffective for providing a large number of patients with a speedy diagnosis. To achieve precise automatic segmentation of retinal blood vessels from end to end, we propose the self-adaptive compensation network (SACom).

Methods SACom employs the U-shaped network as its fundamental structure. First, deformable convolution is incorporated into the encoder to enhance the model's capacity to learn information about morphological structures of retinal blood vessels. An adaptive multi-scale aligned context (AMAC) module is then developed at the bottom of the U-shaped network to extract and aggregate multi-scale context information and align the context features produced by pooling. It can adaptively extract context features according to the input image size and utilize the image context information correctly. Finally, a collaborative compensation branch (CCB) is proposed to fully leverage the feature layer in the decoder and high-level semantic features at the bottom of the network. Its multi-level outputs are helpful for positioning the overall structure of the blood vessel to fine details. Then they are fused with the output feature layer of the decoder end through the feature layer averaging adaptive fusion to improve the mapping capability of the model.

Results and Discussions The segmentation accuracy of retinal vessels can be effectively improved by the proposed SACom model. Each module is beneficial to improve segmentation performance according to the ablation experiment. Compared with the baseline model, SACom just adds a small number of extra parameters (Table 3). The proposed approach can thoroughly detect both thick blood vessels and thin blood vessels, and the connectedness of blood vessels is also more ideal, according to the visualization results of the segmentation (Fig. 6). Subsequent investigation reveals that there are microscopic blood vessels in the SACom segmentation results that are not labeled by experts but exist in fundus images (Fig. 7). It is clear that SACom has a good ability to segment blood vessels and identify blood vessel pixels more accurately, thereby addressing strong subjectivity in manual labeling. SACom performs better than other state-of-the-art methods generally (Table 5), with high sensitivity. The accuracy reaches 0.9695, 0.9763, and 0.9753, the sensitivities are 0.8403, 0.8748, and 0.8506, and the respective AUC values are 0.9880, 0.9917, and 0.9919 for DRIVE, CHASE_DB1, and STARE datasets, respectively.

Conclusions An effective automatic segmentation algorithm called SACom is put forth to achieve precise segmentation of retinal vessels in fundus images. SACom integrates deformable convolution into the encoder based on the network architecture of U-Net to improve the learning capacity of vascular structural information. The bottom of the U-Net is constructed with an AMAC module that can collect and aggregate multi-scale aligned context information to adapt to the multi-scale issue of retinal blood vessels. Finally, a CCB is proposed. Its multi-level outputs calculate loss respectively

and conduct backpropagation to improve the accuracy of each branch's output result. The outputs of the CCB are averaged and then adaptively fused with the output feature map of the decoder for accurate segmentation. The experimental results on three datasets reveal that the method has excellent generalization capability for different pixel classifications, especially for blood vessel pixels, and its comprehensive segmentation performance is better than other state-of-the-art algorithms. In addition, the proposed algorithm does not need too much computation load, which makes it easy to deploy in clinical applications.

Key words image processing; retinal vessels; deformable convolution; context alignment; feature adaptive fusion