

SAM-Retina:基于SAM的双模态视网膜图像动静脉分割

许恒宇, 陈坤, 徐琳, 孙明斋, 陆洲

引用本文

许恒宇, 陈坤, 徐琳, 孙明斋, 陆洲. [SAM-Retina:基于SAM的双模态视网膜图像动静脉分割](#)[J]. 计算机科学, 2025, 52(10): 123-133.

XU Hengyu, CHEN Kun, XU Lin, SUN Mingzhai, LU Zhou. [SAM-Retina:Arteriovenous Segmentation in Dual-modal Retinal Image Based on SAM](#)[J]. Computer Science, 2025, 52(10): 123-133.

相似文献推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[ZKFERP:计算成本恒定的通用高效范围证明方案](#)

ZKFERP:Universal and Efficient Range Proof Scheme with Constant Computational Cost
计算机科学, 2022, 49(10): 335-343. <https://doi.org/10.11896/jsjcx.210900044>

[基于双目叠加仿生的微换衣行人再识别](#)

Moderate Clothes-Changing Person Re-identification Based on Bionics of Binocular Summation
计算机科学, 2022, 49(8): 165-171. <https://doi.org/10.11896/jsjcx.210600140>

[俄语多模态情感语料库的构建及应用](#)

Construction and Application of Russian Multimodal Emotion Corpus
计算机科学, 2021, 48(11): 312-318. <https://doi.org/10.11896/jsjcx.200900088>

[基于CPNet的相对图像质量评估](#)

Relative Image Quality Assessment Based on CPNet
计算机科学, 2020, 47(11): 159-167. <https://doi.org/10.11896/jsjcx.190900052>

[认知视角下的文本情感计算](#)

Text Affective Computing from Cognitive Perspective
计算机科学, 2010, 37(12): 182-185.

SAM-Retina: 基于 SAM 的双模态视网膜图像动静脉分割

许恒宇¹ 陈坤² 徐琳^{1,3} 孙明斋^{4,5} 陆洲¹

1 安徽师范大学物理与电子信息学院光电信息材料功能调控与应用安徽省重点实验室 安徽 芜湖 241002

2 中国科学技术大学精密机械与精密仪器系 合肥 230000

3 安徽师范大学计算机与信息学院工业智能数据安全安徽省重点实验室 安徽 芜湖 241002

4 中国科学技术大学生物医学工程学院 合肥 230026

5 中国科学技术大学苏州高等研究院 江苏 苏州 215000

(2121021186@ahnu.edu.cn)

摘要 动脉与静脉在 RGB 视网膜成像中形态高度相似,且其本身结构兼具细微性和复杂性,导致现阶段多数视网膜图像处理所使用的动静脉分割模型难以取得理想效果。为提高动静脉分割的准确性,同时降低训练成本,提出了一种基于 SAM(Segment Anything Model)的视网膜分割模型——SAM-Retina。SAM-Retina 采用特征融合器-适配型图像编码器-掩码解码器架构,使用同时包含 RGB 图像以及 570 nm 和 610 nm 单波长图像的结构-功能双模态视网膜图像代替原有的单模态(RGB)图像作为输入,利用特征融合器融合这 3 种图像的特征;通过在视觉转换器中插入 Adapter 模块并对其加以更新,保留图像编码器在大规模自然图像数据集上的预训练参数;使用静态提示嵌入代替提示编码器,去除原有 SAM 分割流程中的提示输入过程和提示编码过程。实验阶段将模型在 DualModal2019 和 HRF 数据集上进行训练和评估,并与 U-Net,CRU-Net 和 TW-GAN 进行对比。结果表明,相较于对比模型,SAM-Retina 在各项评估指标上效果更好,尤其是双模态图像的引入,使得在无需扩大模型规模的前提下,有效提升了分割性能。

关键词: 双模态视网膜图像;动静脉分割;图像编码器;视觉转换器;静态提示嵌入

中图分类号 TP391

SAM-Retina: Arteriovenous Segmentation in Dual-modal Retinal Image Based on SAM

XU Hengyu¹, CHEN Kun², XU Lin^{1,3}, SUN Mingzhai^{4,5} and LU Zhou¹

1 Anhui Key Laboratory for Control and Applications of Optoelectronic Information Materials, School of Physics and Electronic Information, Anhui Normal University, Wuhu, Anhui 241002, China

2 Department of Precision Machinery and Precision Instrumentation, University of Science and Technology of China, Hefei 230000, China

3 Anhui Provincial Key Laboratory of Industrial Intelligence Data Security, School of Computer and Information, Anhui Normal University, Wuhu, Anhui 241002, China

4 School of Biomedical Engineering, University of Science and Technology of China, Hefei 230026, China

5 Suzhou Institute for Advanced Research, University of Science and Technology of China, Suzhou, Jiangsu 215000, China

Abstract The shapes of arteries and veins are highly similar in RGB retinal imaging, and their inherent structures are both subtle and complex, making it difficult for most retinal image processing models to achieve ideal results. To improve the accuracy of arteriovenous segmentation and reduce training costs, a retinal segmentation model based on segment anything model is proposed—SAM-Retina. SAM-Retina adopts a feature fusion, adaptive image encoder, and mask decoder architecture, using structure-and-function dual-modal retinal images simultaneously containing RGB as well as 570 nm and 610 nm single wavelength images instead of a single mode(RGB) image as input. The features of these three images are fused through a feature fusion. While retaining the pre-trained parameters of the image encoder on a large-scale natural image dataset, the model's feature extraction capability on retinal medical images is enhanced by inserting an adapter module and updating it within the vision transformer(ViT) block. The

到稿日期:2024-08-02 返修日期:2024-11-25

基金项目:国家自然科学基金面上项目(22073001);网络与信息安全安徽省重点实验室课题(AHNIS2021001);芜湖市重点研发与成果转化项目(2023yf110);安徽师范大学校级科研项目培育项目(2022xjxm052)

This work was supported by the National Natural Science Foundation of China(22073001), Anhui Provincial Key Laboratory of Network and Information Security Project(AHNIS2021001), Key R&D and Achievement Transformation Projects in Wuhu(2023yf110) and Anhui Normal University School-level Research Project Incubation Project(2022xjxm052).

通信作者:徐琳(linxu@ahnu.edu.cn)

static prompt embedding instead of prompt encoder is adopted to remove the input and encoding process of prompts in the original SAM segmentation process. During the experimental phase, the model is trained and evaluated on the DualModal2019 and HRF datasets, and compared with U-Net, CRU-Net, and TW-GAN. The experimental results show that SAM-Retina is more advanced than other models in various evaluation indicators and the employment of dual-modal image also improves segmentation performance without increasing the model size.

Keywords Dual-modal retinal image, Arteriovenous segmentation, Image encoder, Vision transformer, Static prompt embedding

视网膜是人体的重要器官,也是唯一可以通过非侵入式成像技术直接观察到血管的组织。视网膜动静脉的结构参数与多种眼科及全身性疾病密切相关。通过分析视网膜图像中的动静脉的结构参数,能够对糖尿病视网膜病变^[1]、青光眼^[2]等眼部疾病以及阿尔茨海默症^[3]、脑卒中^[4]等中枢系统疾病进行诊断。不同疾病对视网膜动脉和静脉的影响各异,例如青光眼患者的动脉直径和阿尔兹海默症患者的静脉直径会显著减小;高血压和高胆固醇水平会同时影响动静脉直径^[5],从而改变动静脉直径比值(Arteriolar-to-Venular diameter Ratio, AVR)。因此,AVR^[6]已成为诊断多种眼科相关疾病的重要指标。此外,视网膜动静脉的归一化曲率(Tortuosity)^[7-8]、分形维数^[9](Fractal Dimension, FD)和血管密度^[10]已被证明与多种疾病有关联性。综上可知,精确分割视网膜动静脉对于利用动静脉特征进行病理分析和疾病诊断具有重大意义。

传统的动静脉分割研究主要分为两类:基于图形和基于特征。前者通过构建血管的树状拓扑结构来实现分割^[11-12],通常需要手动进行初始化,单个分类错误可能导致整个血管树分类不当。后者针对以目标像素为中心的视网膜像素区块,对其特征展开分析,从而实现目标像素的分类^[13-14]。具体地,先从血管树的分割结果中提取血管中心线,再从中心线像素中提取特征,并据此将血管段分类为动脉或静脉。

近年来,深度学习方法被成功应用到许多医学图像分析中。与传统方法不同的是,深度学习方法通过多个卷积层和池化层自动学习和提取图像中的复杂特征,无需手动设计或选择特征^[15]。在医学图像分割领域,U-Net^[16]网络以其出色的性能成为主流方法。近年来,多数视网膜动静脉分割方法都以U-Net为基础进行改进和优化。Hemelings等^[17]首次使用改进的U-Net进行视网膜动静脉分割,并使用将血管分割和动静脉分类结合的策略,该策略在动静脉的中心线像素上达到了96.98%的分类准确性。Karlsson等^[18]提出了一种基于串联U-Nets的卷积神经网络,将前一个U-Net网络的输出和原始图像一起作为当前U-Net网络的输入,其在DRIVE数据集上的表现优于人类专家。除了U-Net网络之外,生成对抗网络(Generative Adversarial Networks, GAN)也被应用于视网膜动静脉分割领域。Chen等^[19]提出了一种新的拓扑和宽度感知生成对抗网络TW-GAN,首次将拓扑连通性和血管宽度信息整合到深度学习框架中用于动静脉分类。Zhang等^[20]则使用生成对抗网络对U-Net网络进行改进,在RITE数据集上的评估结果显示,该网络的动静脉分类准确性达到96.87%。

动静脉在RGB视网膜图像中形态相似度高,结构细微并具有高度复杂性,使得从RGB图像中提取出充分的特征信息

非常困难。为了更有效地处理这些细微的特征,通常需要大规模运算资源和复杂的模型。如何在尽可能提高动静脉分割准确性的同时降低训练成本,是目前研究的主要问题。结构-功能双模态视网膜成像技术的出现和大型基础分割模型的发展,为这一问题带来了新的解决方案。利用双模态视网膜成像技术,可以同时采集视网膜的RGB图像、570nm图像和610nm图像,不同图像之间的差异性可以使模型在不提升规模的情况下获得更丰富的动静脉特征信息。同时已有研究表明^[21-24],基础分割模型在自然图像数据集上能够通过预训练学习得到大量通用的、可迁移的特征,在将其应用于特定的医学图像分割任务时,表现出更快的训练收敛速度和更高的分割精度。

针对上述解决方案,本文提出了基于大型基础分割模型(SAM)^[25]的视网膜动静脉分割模型SAM-Retina。主要创新点如下:

- 1) 针对RGB图像在动静脉末端的细小分支上特征信息不足的问题,使用双模态图像作为输入,并添加特征融合器融合双模态图像的特征,增强模型可提取的动静脉特征信息;
- 2) 使用Adapter模块调整图像编码器,在保留原模型从自然图像上学习获得通用视觉表示的同时,提升模型对视网膜医学图像的特征提取能力;
- 3) 使用静态提示嵌入代替提示编码器,在精简模型的同时进一步提高模型性能,避免了人为动态选择提示对分割结果的影响。

1 工作基础

1.1 结构-功能双模态视网膜影像

现阶段,视网膜动静脉分割方法大多采用从RGB图像中提取特征的方式。图1(a)展示了动脉和静脉在血管壁厚度和血液含氧量上的差异,在RGB图像中表现为静脉比动脉更粗且颜色更暗。然而,明暗特征容易受到照明条件和成像设备差异等多种因素影响,许多疾病造成的病理改变同样会对动静脉结构造成影响。尤其在细小动静脉分支上,上述差异特征被进一步弱化,这导致多数使用RGB图像进行特征提取的模型对细小动静脉的分割能力较差。为解决这一问题,文中提出新的SAM-Retina模型,使用双模态眼底成像技术所获取的图像组作为输入。

与传统成像技术不同,双模态成像技术能够同时针对视网膜的功能和结构进行成像;在功能成像方面,可以从单波长图像中根据朗伯-比尔定律计算视网膜血氧饱和度。本文方法的图像采集设备为合肥奥布斯生物科技有限公司研发的双模态眼底相机采集系统(OT-110M, Hefei Orbis Biotech

LTD,China),该系统以氙灯作为光源,光路中包含两个滤光片,使中心波长为570 nm和610 nm的光波透过,因此能够同时拍摄3张视网膜图像,即RGB图像、570 nm图像和610 nm图像(下文统称为双模态图像)。在570 nm波长下,血管中的氧合血红蛋白(HbO₂)和脱氧血红蛋白(Hb)有几乎相同的消光系数^[26],使得动脉和静脉在此波长下都能清晰地被可视化,其成像如图1(b)所示。相比之下,氧合血红蛋白对610 nm波长的光吸收率极低,导致动脉的可见度在610 nm图像中远低于静脉,其结果如图1(c)所示。这种在不同波长下动静脉可见度的差异为模型提供了更多的特征信息,该优势在动静脉末端细小分支的分割上尤其突出。

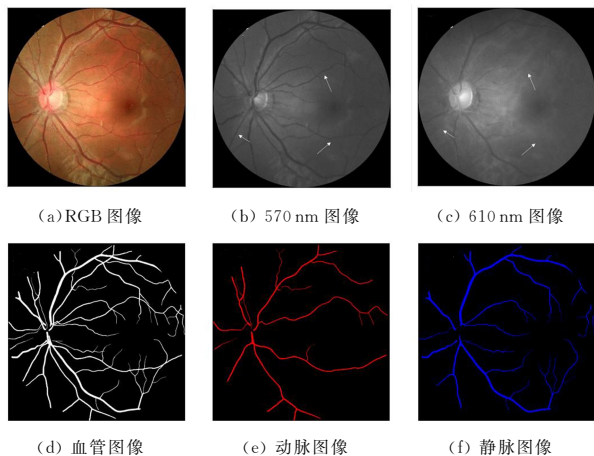


图1 双模态视网膜图像

Fig. 1 Dual-modal retinal images

1.2 基础分割模型

近年来,人工智能领域经历了基础模型发展和应用的飞速增长。这些模型通过在大规模数据集上进行训练,具备了在诸多任务和领域中的泛化能力。在自然语言处理领域,如ChatGPT和Gemini^[27]这样的大语言模型已得到广泛应用,并泛化到不同的细分场景;在图像生成领域,Stable Diffusion^[28]和DALL-E^[29]等扩散模型能够根据文本提示创造出高清图像。随着SAM的出现,这种由数据驱动的模式也进入图像分割领域。

SAM由图像编码器(Image Encoder)、提示编码器(Prompt encoder)和掩码解码器(Mask Decoder)3部分组成。其中,图像编码器通过多个重复堆叠的ViT块对图像进行16倍下采样得到图像特征嵌入,提示编码器通过对提示点进行傅里叶位置编码得到提示嵌入。掩码解码器由两个Transformer层组成,通过对提示嵌入和图像特征嵌入反复进行交叉注意力得到最终的预测掩码。

Wald等^[21]和Deng等^[24]通过在不同医学图像分割数据集上的评估,展示了SAM的Zero-shot能力。尽管其性能尚未达到目前最先进水平,但是SAM是复杂医学图像分割任务专门化研究的有效基石。Ma等^[22]冻结SAM的图像编码器和提示编码器,仅对掩码解码器进行再训练。在21个3D医学图像分割任务上的评估显示,迁移训练的SAM在不同任务中的Dice系数可达到59.46%~93.31%。此外,Wu等^[30]通过冻结提示编码器,调整图像编码器和掩码解码器微

调SAM,并且通过引入深度和广度分支的注意力操作,增加了模型对3D样本的分割能力。实验结果显示,调整后的模型在各数据集上获得了平均88.3%的Dice系数,超过了一些经典的医学分割网络。

上述研究表明,基础分割模型在庞大的自然图像数据集上学习到的通用特征提取能力,有助于医学图像分割任务的下游适配。基于此,本文选择SAM作为基础模型,并对其加以调整以适配视网膜动静脉分割领域的特异性。

2 SAM-Retina

SAM-Retina的网络架构由特征融合器、适配型图像编码器和掩码解码器3部分组成,其架构如图2所示。其中,特征融合器为一个新增的模块,适配型图像编码器则是经过优化的模块。在模型反向传播过程中,只对特征融合器和图像编码器中的Adapter模块的参数进行梯度更新,同时保持掩码解码器参数处于冻结状态。在分割过程中,采用端到端的方法,即同时进行血管分割任务和动静脉分类任务,避免在多重任务中分类任务对分割结果的依赖。具体地,特征融合器将同一视网膜的双模态图像融合成一张三通道的特征融合图像;该特征融合图像随后被图像编码器处理转化为图像嵌入(image embedding);图像嵌入与静态提示嵌入(static prompt embedding)被共同输入到掩码解码器中,经过解码器处理后输出动脉和静脉的预测掩码。

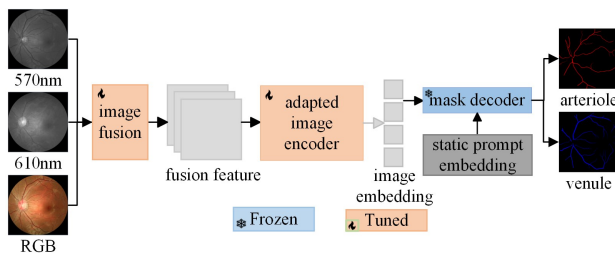


图2 SAM-Retina网络架构

Fig. 2 Architecture of SAM-Retina network

2.1 特征融合器

特征融合器用于融合双模态图像的特征,是模型中新加入的模块。如图3所示,特征融合器在通道维度上将双模态图像进行堆叠,形成一个五通道输入。随后,该五通道输入被送入两个顺序排列的卷积层,每个卷积层后都跟有一个批量归一化层,以稳定学习过程并加速收敛。第一个卷积层通过64个过滤器提取图像每个通道的初级特征。第二个卷积层使用3个过滤器,以输出与RGB图像通道相匹配的特征图。在这两个卷积层之后,采用残差连接的策略进一步融合特征并保留原始图像的信息。特征融合器的计算过程如式(1)所示:

$$\mathbf{F}_{\text{res}} = \text{Conv}(\text{Stack}(\mathbf{I}_{\text{RGB}}, \mathbf{I}_{570\text{nm}}, \mathbf{I}_{610\text{nm}})) + \mathbf{I}_{\text{RGB}} \quad (1)$$

其中,Stack表示堆叠操作,Conv表示连续两次卷积操作。为了将融合图像的特征矩阵值控制在[0,1]之间,在获得最终输出之前通过一个Sigmoid激活函数将 \mathbf{F}_{res} 归一化,运算过程如式(2)所示:

$$\mathbf{S} = \text{sigmoid}(\mathbf{F}_{\text{res}}) = \frac{1}{1 + e^{-\mathbf{F}_{\text{res}}}} \quad (2)$$

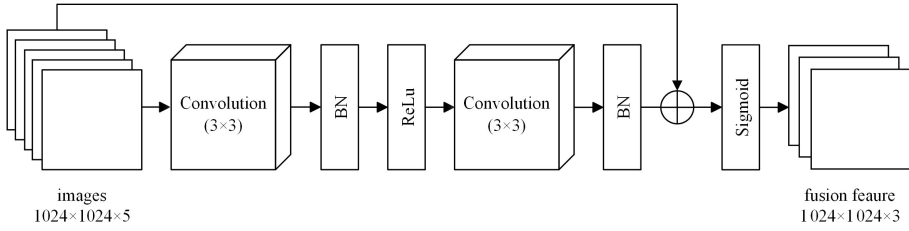


图3 新加入的特征融合器的结构

Fig. 3 Structure of newly added feature fusion module

2.2 适配型图像编码器

SAM 使用了基于视觉转换器 Vision Transformer^[31] (ViT) 的图像编码器来提取图像特征。在输入编码器之前, 图像首先被分割成 16×16 像素的小块, 并通过一个线性变换层(也称为嵌入层)转换为固定维度的图像块令牌(token)。由于 Transformer 架构本身不具备处理输入顺序的能力, 因此需要为每个图像块令牌添加位置编码, 用于向模型提供图像块在原始图像中的位置信息。编码器由多个重复堆叠的 ViT 块组成, 具体结构如图 4(a) 所示。每个 ViT 块包含一个多头自注意力 (Multi-Head Self-Attention) 层和一个多层感知器 (Multiple Perceptron, MLP) 层。第 l 个 ViT 块的运算过程如式 (3)、式 (4) 所示:

$$\hat{\mathbf{X}} = \text{Multihead}(\text{Norm}(\mathbf{X}^{l-1})) + \mathbf{X}^{l-1} \quad (3)$$

$$\mathbf{X}^l = \text{MLP}(\text{Norm}(\hat{\mathbf{X}})) + \hat{\mathbf{X}} \quad (4)$$

其中, Norm 表示层归一化操作, Multihead 表示多头自注意力机制, MLP 表示多层感知器。多头注意力的具体计算式为:

$$\mathbf{Q}_i = \mathbf{X}\mathbf{W}_i^Q, \mathbf{K}_i = \mathbf{X}\mathbf{W}_i^K, \mathbf{V}_i = \mathbf{X}\mathbf{W}_i^V \quad (5)$$

$$\text{head}_i = \text{Attention}(\mathbf{Q}_i, \mathbf{K}_i, \mathbf{V}_i) = \text{softmax}\left(\frac{\mathbf{Q}_i\mathbf{K}_i^T}{\sqrt{d_k}}\right)\mathbf{V}_i \quad (6)$$

$$\text{Multihead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)\mathbf{W}^O \quad (7)$$

其中, $\mathbf{Q}_i, \mathbf{K}_i$ 和 \mathbf{V}_i 分别表示查询(query)、键(key)和值(value)的权重矩阵; \mathbf{X} 表示输入特征; $\mathbf{W}_i^Q, \mathbf{W}_i^K$ 和 \mathbf{W}_i^V 表示对应的转换矩阵; d_k 表示矩阵的维度; head_i 表示单个注意力头的输入; \mathbf{W}^O 表示多头注意力的权重。经过若干层 ViT 块的处理后, 图像块令牌被转换为图像嵌入。

SAM 图像编码器中 ViT 的预训练在自然图像上进行。为了使其泛化到视网膜动静脉分割场景, 使用 Adapter 模块对图像编码器进行调整^[32], 在训练时仅更新添加的两个 Adapter 模块中的参数, 其余图像编码器参数全部冻结。Adapter 模块使用了低秩自适应 (Low-Rank Adaptation, LoRA) 的思想, 通过引入低秩矩阵分解来减少参数更新的数量。具体来说, 对于预训练模型中的每个权重矩阵 \mathbf{W} , LoRA 用两个低秩矩阵 \mathbf{A} 和 \mathbf{B} 的乘积来近似, 即 $\mathbf{W} \approx \mathbf{A}\mathbf{B}$ 。这样在微调过程中, 只需要更新这两个较小的低秩矩阵, 不用更新整个权重矩阵。

如图 4(b) 所示, 每个 ViT 块都被部署了两个 Adapter 模块, 第一个 Adapter 模块置于多头注意力之后和残差连接之前, 第二个 Adapter 模块放置在多头注意力之后 MLP 层的残差路径中。第二个 Adapter 模块之后使用了一个缩放系数 (scale factors) 对嵌入进行缩放, 允许模型调整不同层输出的比重。

Adapter 模块均为 Bottleneck 结构, 首先通过一个下投影

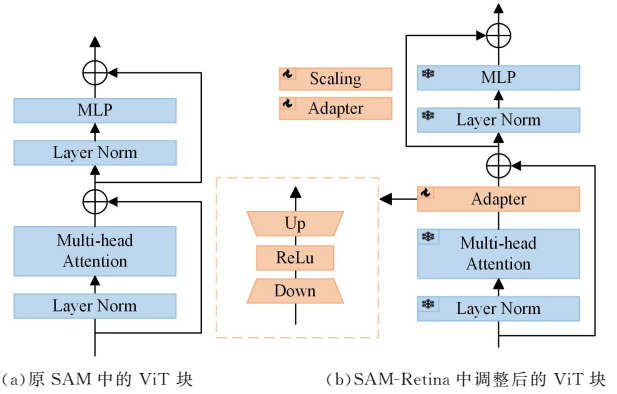
D_{down} 将输入特征嵌入 \mathbf{x}_{in} 从原始维度压缩到隐藏维度, 再使用 ReLU 激活函数进行非线性变换, 最后通过一个上投影 D_{up} 将激活后的特征扩充到原始维度。具体过程如下所示:

$$\mathbf{x}_{\text{hid}} = D_{\text{down}}(\mathbf{x}_{\text{in}}) = \mathbf{W}_{\text{down}} \times \mathbf{x}_{\text{in}} + \mathbf{b}_{\text{down}} \quad (8)$$

$$\mathbf{x}_{\text{act}} = \text{ReLU}(\mathbf{x}_{\text{hid}}) = \max(0, \mathbf{x}_{\text{hid}}) \quad (9)$$

$$\mathbf{x}_{\text{out}} = D_{\text{up}}(\mathbf{x}_{\text{act}}) = \mathbf{W}_{\text{up}} \times \mathbf{x}_{\text{act}} + \mathbf{b}_{\text{up}} \quad (10)$$

其中, \mathbf{W}_{down} 和 \mathbf{W}_{up} 分别表示上投影和下投影的权重矩阵, \mathbf{b}_{down} 和 \mathbf{b}_{up} 表示对应的偏置矩阵。



(a) 原 SAM 中的 ViT 块

(b) SAM-Retina 中调整后的 ViT 块

图4 Vision Transformer 块网络结构调整

Fig. 4 Adapted Vision Transformer block network structure

2.3 静态提示嵌入

SAM 中使用提示编码器将所输入的提示(prompt)转换为稀疏提示嵌入和密集提示嵌入。提示嵌入被用于引导模型的注意力和理解, 这类类似于自然语言处理中的参数微调(prompt-tuning)。在视觉任务中, 这些提示被视为一种先验知识, 可以帮助模型更有效地识别和处理图像中的关键特征。SAM 支持 4 种类型的提示: 点(point)、边界框(box)、掩码(mask)和文本(text)。

鉴于视网膜血管的细微特性和复杂性, 人工选取有效的提示不仅效率低下, 而且容易出错。为了解决这一问题, SAM-Retina 移除了 SAM 的提示编码器, 并采用了一种新的策略, 具体描述如下。

在训练集上计算出图像中被标记为血管概率最大的像素, 如式 (11) 所示:

$$(\hat{i}, \hat{j}) = \arg \max_{i,j} \left(\frac{1}{N} \sum_{k=1}^N I_k(i, j) \right) \quad (11)$$

其中, $I_k(i, j)$ 为第 k 张真值图像中位置 (i, j) 的像素值, 如果为真值则为 1, 否则为 0; N 为训练集中的图片总数; (\hat{i}, \hat{j}) 为所求最大概率为血管的像素坐标。

得到坐标后, 先将其数值归一化, 再进行位置编码运算, 最后加上预训练中正标签的权重矩阵, 得到该点的提示嵌入。

具体计算过程如下:

$$p_{\text{norm}} = \left(\frac{\hat{i}+0.5}{W}, \frac{\hat{j}+0.5}{H} \right) \quad (12)$$

$$pe = \sin(\mathbf{W}_{\text{pos}} \cdot p_{\text{norm}}) + \cos(\mathbf{W}_{\text{pos}} \cdot p_{\text{norm}}) \quad (13)$$

$$e = pe + \mathbf{W}_{\text{embed}} \quad (14)$$

其中, W 和 H 分别表示图像的宽和高, \mathbf{W}_{pos} 表示位置编码的权重矩阵, $\mathbf{W}_{\text{embed}}$ 表示正标签的权重矩阵。

新策略的核心是将计算出的提示嵌入 e 作为稀疏提示嵌入, 稀疏提示嵌入与 SAM 中默认的密集提示嵌入一起组成静态提示嵌入, 之后与图像嵌入一起输入到掩码解码器中, 减少了人为提供提示对分割结果的影响。

2.4 损失函数

在动静脉分割任务中, 动静脉交汇处的分割尤其具有挑战性。为了解决这一问题, SAM-Retina 的损失函数增加了对动静脉交汇处的关注。该损失函数综合考虑了动脉、静脉和整体血管上的分割损失, 同时对动静脉交汇处的分割结果损失给予了额外的权重, 以对其加以强调, 减少此关键区域的分割错误。计算方法如式(15)所示:

$$L_{\text{total}} = \lambda_1 (L(y_a, \hat{y}_a) + L(y_v, \hat{y}_v)) + \lambda_2 (L(\max(y_a, y_v), \max(\hat{y}_a, \hat{y}_v))) + \lambda_3 (L(\min(y_a, y_v), \min(\hat{y}_a, \hat{y}_v))) + \lambda_4 (L(\min(y_a, y_v), \min(\hat{y}_a, \hat{y}_v))) \quad (15)$$

其中, y_a 和 y_v 分别为动静脉分割真值, \hat{y}_a 和 \hat{y}_v 为预测的概率值, $\min(y_a, y_v)$ 为动脉静真值之间的交集, $\min(\hat{y}_a, \hat{y}_v)$ 为动脉预测值和静脉真值之间的交集, $\min(y_a, \hat{y}_v)$ 为静脉预测值和动脉真值之间的交集。 L 为交叉熵损失, 如式(16)所示:

$$L = - \sum_{i=1}^N [y^{(i)} \log(\hat{y}^{(i)}) + (1 - y^{(i)}) \log(1 - \hat{y}^{(i)})] \quad (16)$$

其中, N 表示像素总数, $\hat{y}^{(i)}$ 表示单个像素预测为真值的概率。

3 实验及结果分析

3.1 实验环境

本实验所采用的所有模型均基于 Python 3.8 版本, 利用 PyTorch 框架进行实现。实验都在一台有 4 个 GeForce GTX 3090 GPU 的 Linux 服务器上完成。训练时, batch-size 设置为 2, 使用 Adam 优化器, 学习率设置为 0.0001。为加快后期拟合过程, 每训练 10 个 epoch, 学习率下降为原来的 3/10, 共训练 80 个 epoch。

3.2 数据集预处理

3.2.1 数据集

本实验使用 DualModal2019 和 HRF 两个公开数据集进行训练和评估。

DualModal2019 数据集包含 30 组高分辨率(1024×1024 像素)的图像, 每组由 3 张图像构成: 1 张 RGB 彩色图像以及 2 张对应的 570 nm 和 610 nm 单色图像。在 DualModal2019 数据集中随机抽取 5 组图像作为测试集, 用于评估模型在未见过的数据上的性能表现, 其余 25 组图像进行五折交叉验证。文中每个量化表格里, 所有性能指标下面都标注了五折交叉验证结果的平方差。

HRF 数据集(High-Resolution Fundus image database)

包含 45 张高分辨率 RGB 视网膜图像, 这些图像分为 3 类, 分别为健康眼底图像、糖尿病视网膜病变眼底图像和青光眼眼底图像。为了方便和其他模型进行对比, 本实验采用文献[19]中的划分方法, 即 3 类图像中每一类的前 5 张作为测试集, 用于评估模型性能。

3.2.2 预处理

为了进一步增强模型的泛化能力并避免过拟合, 对数据集中每张图像都进行了分割处理: 以 128 像素为步长, 将每张图像分割成多个 256×256 像素的小图像块。此外, 在训练过程中, 引入了多种数据增强技术以提高模型对于图像变换的鲁棒性, 包括对图像进行随机角度旋转、随机水平翻转和随机垂直翻转等。

3.3 评价指标

以往的动静脉分割结果评估通常采用分阶段的评估策略, 即首先评估模型从眼底图像中分割出视网膜血管的性能, 然后评估从这些分割结果中进一步分类出动脉和静脉的准确性。这种方法虽然在流程上较为简明, 但其评估的准确性受限于血管分割结果, 导致最终动静脉分类的准确率评估可能不够全面。本实验模型采用了一种端到端的分割策略。为了更准确和直观地反映模型在动静脉分割任务上的性能, 本实验对分割结果进行了像素级别的评估, 直接计算正确标记和错误标记的动静脉像素数量, 并在此基础上使用准确度(Accuracy)、灵敏度(Sensitivity)、特异度(Specificity)、精确度(Precision)、F1 分数(F1-Score)以及马修斯相关系数(MCC)来综合评估模型分割结果的质量。评估指标的计算式如表 1 所列。

表 1 评估指标计算方式

评估指标	计算方式
Accuracy	$\frac{TP+TN}{TP+TN+FN+FP}$
Sensitivity	$\frac{TP}{FN+TP}$
Specificity	$\frac{TN}{FP+TN}$
Precision	$\frac{TP}{FP+TP}$
F1-Score	$2 \times \frac{TP}{2 \times TP + FP + FN}$
MCC	$\frac{TP \times TN - FP \times FN}{\sqrt{(FP+TP)(FN+TP)(FP+TN)(FN+TN)}}$

表 1 中, TP 和 TN 分别表示被正确预测为动静脉和背景(非动静脉)的像素数量, FP 和 FN 分别表示被错误分类为动静脉和背景(非动静脉)的像素数量。

3.4 对比实验及评估

本文的对比实验主要分为两部分: 第一部分使用 DualModal2019 数据集, 将 SAM-Retina 和现有的主流医学分割模型 U-Net 和 CRU-Net^[33] 进行对比, 具体分为两个场景, 在第一个场景中仅使用 RGB 图像, 在第二个场景中使用双模态图像, 即同时使用 RGB 图像、570 nm 图像和 610 nm 图像; 第二部分使用 HRF 数据集, 将 SAM-Retina 与 U-Net, CRU-Net 和 TW-GAN 进行对比。

3.4.1 DualModal2019 数据集

1) 仅使用 RGB 图像

在仅使用 RGB 图像的场景下, 3 个模型的评估指标结

果如表 2 所列。F1 分数是评估网络性能更全面的指标，为了更直观地展示对比差异，使用 F1 分数绘制了如图 5(a)所示的柱状图。可以明显看出，在仅使用 RGB 图像的场下，SAM-Retina 的动静脉分割结果 F1 分数均值相比 U-Net 分别提升了 5.28 个百分点和 5.50 个百分点，相比于 CRU-Net 分别提升了 3.44 个百分点和 4.56 个百分点。MCC 指标对比的结果与 F1 分数类似。图 5(b)和图 5(c)分别展示了动静脉分割的精确度-召回率 (PR) 曲线，用于评估

在不同阈值设置下精确度和召回率 (Recall) 之间的关系。可以看出，SAM-Retina 的动静脉 PR 曲线将 CRU-Net 和 U-Net 的动静脉 PR 曲线包围。AUC 为 PR 曲线和坐标轴所围面积，其值越高，表示模型的性能越好。SAM-Retina 动脉分割结果的 AUC 均值相比 U-Net 提升了 14.63 个百分点，相较于 CRU-Net 提升了 4.77 个百分点；静脉分割结果的 AUC 均值相比 U-Net 提升了 13.61 个百分点，相较于 CRU-Net 提升了 4.88 个百分点。

表 2 单 RGB 图像输入下 U-Net, CRU-Net, SAM-Retina 的性能对比

Table 2 Performance comparison of U-Net, CRU-Net, and SAM-Retina solely using RGB images as input (%)

任务	模型	F1	MCC	Sensitivity	Precision	Accuracy	Specificity
动脉	U-Net	77.40±0.61	76.54±0.63	77.30±0.95	77.60±0.71	98.31±0.04	99.13±0.04
	CRU-Net	79.24±0.67	78.54±0.68	80.57±1.67	78.35±1.39	98.43±0.05	99.12±0.08
	SAM-Retina	82.68±0.13	82.09±0.17	80.62±0.72	84.96±1.04	98.74±0.02	99.44±0.05
静脉	U-Net	79.19±0.44	78.49±0.42	84.53±0.68	74.60±1.14	98.27±0.05	98.83±0.08
	CRU-Net	80.04±0.45	79.36±0.43	83.30±1.76	77.38±2.26	98.39±0.07	99.00±0.14
	SAM-Retina	84.69±0.14	84.08±0.16	85.40±0.47	84.04±0.41	98.80±0.01	99.34±0.02

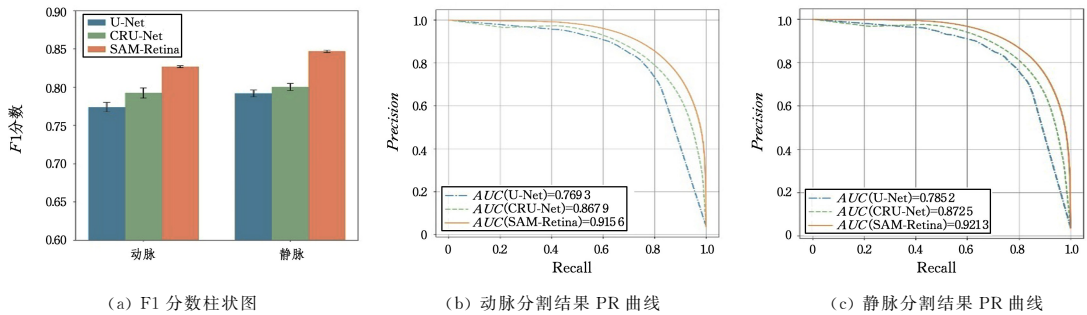


图 5 单 RGB 图像输入下各模型评估指标的可视化对比

Fig. 5 Visual comparison of metrics for various models under single RGB image input

图 6(a)展示了 RGB 图像输入条件下，测试集图片中的动静脉真值与不同模型的动静脉分割结果，图 6(c)是图 6(a)

的局部放大图。红色的血管代表动脉，蓝色的血管代表静脉，紫色为动静脉交汇区域，即同时标注为动脉与静脉的血管。

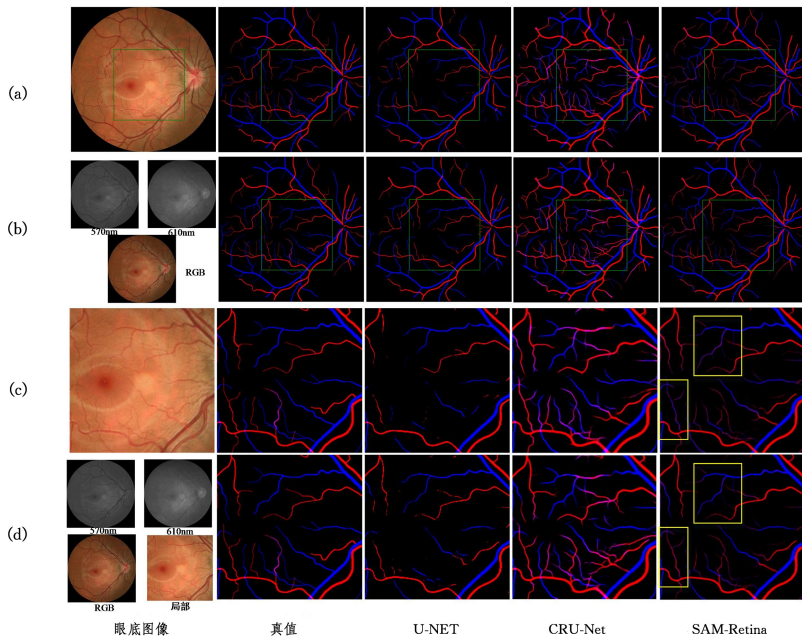


图 6 RGB 图像与多模态输入下各模型分割可视化结果(电子版为彩图)

Fig. 6 Visualization results of segmentation of various models under RGB image and dual-modal image

将3个模型的分割结果与真值进行对比,可以观察到以下特点:U-Net网络在血管末端和细小血管分叉处的分割效果较差;CRU-Net虽然在一定程度上处理了细小血管,但在这些区域出现了大量的误分割情况;而SAM-Retina虽然仍存在将某些细小血管同时分类为动脉和静脉的情况,但是相比U-Net和CRU-Net误分割情况显著减少。

2)使用双模态图像

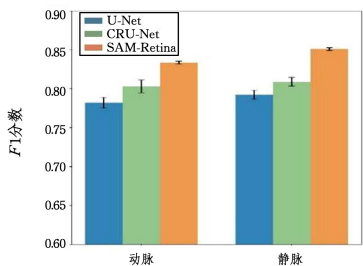
在使用双模态图像的场景下,3个模型均使用Dual-Modal2019数据集进行训练和评估。图6(b)为融合双模态3组图像后的分割结果,图6(d)为图6(b)的局部放大图。图6(c)和图6(d)为眼底同一位置放大区域,但采用不同模态图像的分割对比结果。可以看出,双模态图像提供了更为丰富的动静脉信息,使3个模型在动静脉区分结果上均有明显改善。

3个模型的定量评估指标结果数据如表3和图7所示,其中图7(a)为F1分数柱状图。从表3和图7(a)中可以明显看出;SAM-Retina的动静脉分割结果F1分数均值,相比于U-Net分别提升5.17个百分点和5.87个百分点,相比于CRU-Net分别提升3.06个百分点和4.22个百分点;同时,SAM-Retina的MCC指标也优于U-Net和CRU-Net模型。动静脉分割结果的PR曲线分别如图7(b)和图7(c)所示,可以看出双模态输入场景下SAM-Retina的动静脉PR曲线同样将CRU-Net和U-Net的动静脉PR曲线包围,SAM-Retina动脉分割结果的AUC均值相比U-Net提升了11.61个百分点,相比CRU-Net提升了5.35个百分点,静脉分割结果的AUC均值相较于U-Net和CRU-Net分别提升了11.47个百分点和5.22个百分点。

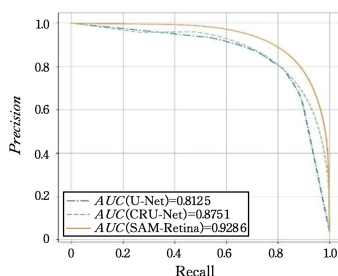
表3 双模态图像输入下U-Net,CRU-Net和SAM-Retina的性能对比

Table 3 Performance comparison of U-Net,CRU-Net,and SAM-Retina using dual-modal image as input (%)

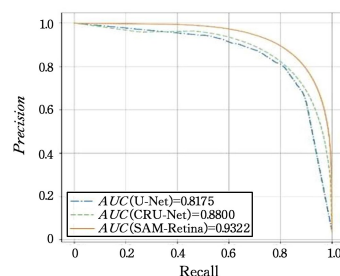
任务	模型	F1	MCC	Sensitivity	Precision	Accuracy	Specificity
动脉	U-Net	78.18±0.66	77.35±0.69	78.45±1.60	78.00±1.88	98.36±0.07	99.13±0.11
	CRU-Net	80.29±0.81	79.63±0.83	81.01±1.40	79.98±2.49	98.52±0.09	99.20±0.13
	SAM-Retina	83.35±0.19	82.79±0.20	80.85±0.49	86.08±0.36	98.79±0.01	99.49±0.02
静脉	U-Net	79.22±0.56	78.51±0.60	84.57±0.90	74.58±0.70	98.27±0.05	98.83±0.04
	CRU-Net	80.87±0.56	80.19±0.56	83.14±1.16	79.02±1.17	98.47±0.06	99.09±0.10
	SAM-Retina	85.09±0.19	84.50±0.20	85.48±0.39	84.76±0.21	98.84±0.01	99.38±0.01



(a) F1分数柱状图



(b) 动脉分割结果PR曲线



(c) 静脉分割结果PR曲线

图7 双模态图像输入下各模型评估指标的可视化对比

Fig. 7 Visual comparison of metrics for various models under dual-modal images input

图6(c)和图6(d)中黄色矩形框为SAM-Retina融合多模态信息后与单RGB图像输入场景的对应区域细节比较结果。由于血管末端缺乏动静脉的特征信息,这两处细小血管在单RGB图像中同时被分类为动脉和静脉。在利用了多模态动静脉可见性差异后,SAM-Retina给出了正确的分类。

进一步综合两种场景(使用RGB图像和双模态图像)下的定量评估指标(结合表2和表3中的SAM-Retina数据),得出不同模态下的性能评估结果,如表4所列。可以看出,SAM-Retina动静脉分割结果的F1分数均值分别提高了0.67个百分点和0.40个百分点,AUC分别提高了1.3个百分点和1.09个百分点,且在其他评估指标上也表现出了一定程度的提升。这充分表明,相比于传统的RGB图像,双模态图像中动静脉的可视性差异能够为模型提供更丰富的信息,从而在处理细节丰富且结构复杂的视网膜图像时,获得更准确的分割结果。

表4 SAM-Retina在RGB和双模态图像下的性能对比

Table 4 Performance comparison of SAM-Retina in RGB and dual-modal images (%)

任务	图像	F1	Precision	Accuracy
动脉	RGB	82.68±0.13	84.96±1.04	98.74±0.02
	双模态	83.35±0.19	86.08±0.36	98.79±0.01
静脉	RGB	84.69±0.14	84.04±0.41	98.80±0.01
	双模态	85.09±0.19	84.76±0.21	98.84±0.01

3.4.2 HRF数据集

为了进一步评估SAM-Retina模型的分割性能,同样在HRF数据集上进行对比实验,并引用了TW-GAN模型的公开分割结果进行对比。由于HRF数据集中只有RGB图像,因此本实验所有的训练和评估均在单RGB图像上进行。图8(a)为健康人眼底图像,其图像包括大量纹理噪声;图8(b)为糖尿病性视网膜病变患者眼底图像,其图像包括大量的出血斑。分别与真值作对比,SAM-Retina动静脉的分割准确度

都是最高的,尤其是在动静脉交汇区域(紫色部分)和包含病变区域(黄色矩形为代表区域之一)。U-Net 在动静脉末端分支细节的检测方面面临较大的挑战,CRU-Net 易受到噪声与病变的干扰,而 TW-GAN 在多种病理状况下对动静脉的区

分能力存在一定的局限性。这表明,虽然眼部疾病的存在增加了图像分割的难度,但 SAM-Retina 凭借其卓越的特征提取和处理能力,依然保持较好的分割效果,展现了强大的鲁棒性和泛化能力。

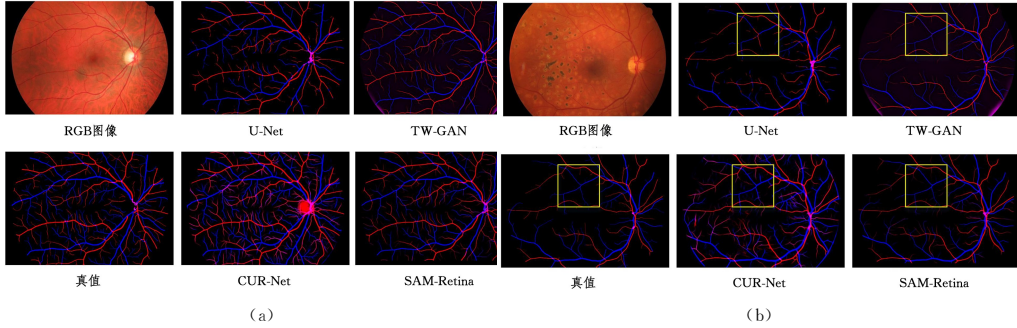


图 8 HRF 数据集各模型分割的可视化结果(电子版为彩图)

Fig. 8 Visualization results of segmentation for each model on HRF dataset

结合表 5 的定量结果和图 9(a)可以看出, SAM-Retina 动脉分割结果的 F1 分数相较于 U-Net, CRU-Net 和 TW-GAN 分别领先了 4.38 个百分点、1.89 个百分点和 3.30 个百分点, 静脉分割结果分别领先 3.39 个百分点、4.81 个百分点和 1.89 个百分点。动静脉分割结果的 PR 曲线分别如图 9(b) 和图 9(c) 所示, 可以看出 SAM-Retina 的 PR 曲线将 U-Net 和 TW-GAN 完全包围, 将 CRU-Net 部分包围, SAM-Retina

动脉分割结果的 AUC 均值相较于 U-Net, CRU-Net 和 TW-GAN 分别提升了 12.42 个百分点、4.1 个百分点和 4.59 个百分点, 静脉分割结果的 AUC 均值则分别提升了 10.51 个百分点、8.12 个百分点和 2.53 个百分点。综合上述数据可以发现, 在 HRF 数据集上 SAM-Retina 同样优于其余模型, 这进一步证明了 SAM-Retina 在处理复杂和具有挑战性的动静脉分割任务时的强大性能。

表 5 HRF 数据集上 U-Net, CRU-Net, TW-GAN 和 SAM-Retina 的性能对比

Table 5 Performance comparison of U-Net, CRU-Net, TW-GAN, and SAM-Retina on HRF dataset

任务	模型	F1	MCC	Sensitivity	Precision	Accuracy	Specificity
动脉	U-Net	71.40	70.76	72.77	71.14	98.18	99.07
	CRU-Net	73.89	73.37	78.46	70.74	98.28	98.98
	TW-GAN	72.48	72.04	72.07	74.54	98.30	99.23
	SAM-Retina	75.78	75.37	75.89	77.02	98.50	99.29
静脉	U-Net	75.75	75.06	75.82	76.76	98.11	99.08
	CRU-Net	74.33	73.72	77.36	73.00	97.94	98.84
	TW-GAN	77.25	76.68	76.99	78.78	98.24	99.17
	SAM-Retina	79.14	78.59	77.90	81.38	98.41	99.30

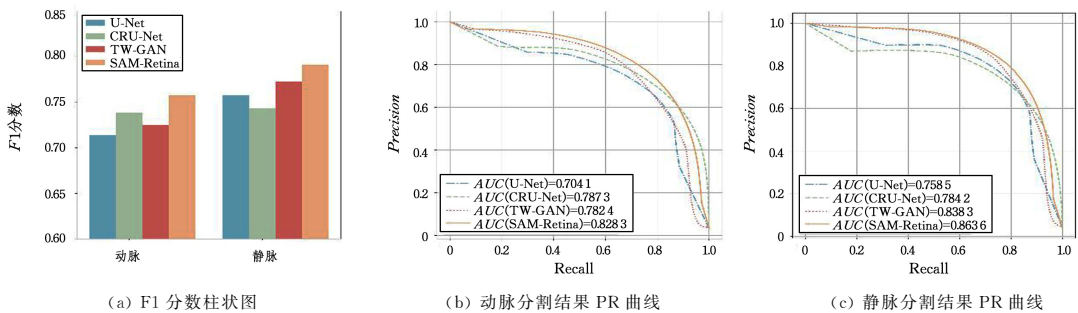


图 9 HRF 数据集中模型性能的可视化对比

Fig. 9 Visualized comparison of model performance on HRF dataset

3.5 消融对比实验

为了验证对 SAM 调整的有效性, 将 SAM-Retina 与 SAM 在测试集上进行对比。具体地, 针对模型(见图 2)的特征融合器、适配型图像编码器、静态提示嵌入以及损失函数分别做了对比实验及分析。

3.5.1 对特征融合器的消融对比实验

由于特征融合器的作用是融合双模态图像, 当 SAM-

Retina 模型只使用 RGB 作为输入时, 对比使用双模态数据作为输入, 结果如图 6 和表 4 所示。图 6(c) 与图 6(d) 中 SAM-Retina 的结果对比(黄色矩形区域)表明, 特征融合器有效地利用了动静脉在不同波长下所展现的差异, 显著改善了动静脉误分类的情况。此外, 表 4 中所提方法各项指标的显著提升进一步证明了特征融合器的融合过程对于提升分割性能具有积极作用。

3.5.2 对适配型图像编码器的消融对比实验

对适配型图像编码器的消融实验,仍然需要去除特征融合器的影响,即基于同样的单模态数据,将 SAM-Retina 和 SAM 模型进行对比。图 10 展示了这两个模型的可视化分割结果。由于不能融合多模态特征,只能选择 3 类图片中的一种进行训练。文中选择 RGB 图像作为输入,SAM 分别使用自动分割(无提示)、1 个采样点和 3 个采样点提示,SAM-Retina 使用静态提示嵌入。从图 10 中可以明显看到,与 SAM-Retina 相比,SAM 只是粗略地提取出了眼底医学图像的大致特征,缺乏提取动静脉特征信息的能力,以至于完全无法从眼底图像中分割出视网膜动静脉。实验的定量评估指标数据对比如表 6 所列,其中 MCC 分数是一个衡量二元分类问题中分类器性能的重要指标,它的取值范围为 $-1 \sim 1$ 。具体地,1 表示完全准确的预测,0 表示预测的结果相当于随机预测, -1 表示预测的结果和真值完全相反。由表 6 中数据可以得出,无论是在取 1 个点作为提示还是取 3 个点作为

提示的场景中,SAM 的动静脉分割预测结果都类似于随机预测,验证了上文所说的 SAM 作为一个在自然图像训练集上训练的模型,缺乏对特定领域医学影像的特征提取能力。

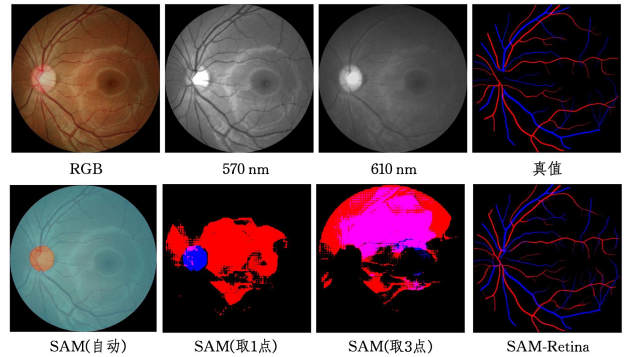


图 10 SAM 和 SAM-Retina 的可视化分割结果对比

Fig. 10 Visualized segmentation results comparison of SAM and SAM-Retina

表 6 SAM 与 SAM-Retina 在 RGB 图像中的性能对比

Table 6 Performance comparison of SAM and SAM-Retina on RGB image

(%)							
任务	模型	F1	MCC	Sensitivity	Precision	Accuracy	Specifcity
动脉	SAM(1 point)	7.16	-0.03	93.64	3.72	9.07	5.78
	SAM(3 point)	7.48	0	93.50	3.90	9.84	6.44
	SAM-Retina	82.53	81.9	81.20	83.99	98.72	99.40
静脉	SAM(1 point)	7.20	0	94.04	3.74	9.23	5.92
	SAM(3 point)	7.52	0.03	94.64	3.92	9.16	5.69
	SAM-Retina	84.58	83.96	84.60	84.58	98.80	99.37

3.5.3 对静态提示嵌入模块的消融对比实验

为了探究动态提示的选择以及静态提示嵌入对 SAM-Retina 分割性能的影响,使用 SAM-Retina 在 DualModal2019 双模态数据集上进行了对比实验。实验设计了 SAM-Retina 的 3 种不同场景:1)在动静脉处随机选择某一点作为提示;2)不使用提示;3)使用静态提示嵌入。

表 7 列出了 SAM-Retina 在不使用/使用静态提示嵌入模块下的评估指标的对比结果,全部数据均为五折交叉验证的综合结果。其中,将使用静态提示嵌入和使用一个正采样点作为提示进行对比,动静脉分割结果的 F1 分数均值分别

提高了 3.11 个百分点和 3.09 个百分点,标准差分别降低了 1.47 个百分点和 1.67 个百分点。当将使用静态提示嵌入和不使用提示进行对比时,动静脉分割结果的 F1 分数均值分别提高了 3.04 个百分点和 2.49 个百分点,同时标准差分别降低了 0.87 个百分点和 0.40 个百分点。结果表明,使用静态提示嵌入不仅提升了分割性能,还减少了结果的变异性。此外,将不使用提示与使用单采样点作为提示相比,动静脉 F1 分数的标准差下降了 0.56 个百分点和 1.27 个百分点。这进一步证实了由于视网膜血管结构的微小和复杂特征,提示的动态选择可能会带来不确定性,对分割结果产生负面影响。

表 7 SAM-Retina 不同提示输入下模型性能的对比

Table 7 Performance comparison of SAM-Retina under different prompt inputs

(%)							
任务	提示	F1	MCC	Sensitivity	Precision	Accuracy	Specifcity
动脉	1 point	80.24±1.66	80.19±1.29	72.16±3.70	90.69±1.96	98.68±0.07	99.71±0.08
	None	80.31±1.08	80.17±0.77	72.99±1.84	89.69±2.98	98.67±0.04	99.66±0.13
	static prompt embedding	83.35±0.19	82.79±0.20	80.85±0.49	86.08±0.36	98.79±0.01	99.49±0.02
静脉	1 point	82.00±1.86	81.67±1.56	76.71±5.66	88.57±3.35	98.70±0.07	99.58±0.16
	None	82.60±0.59	82.05±0.49	80.31±4.07	85.36±3.85	98.68±0.05	99.42±0.20
	static prompt embedding	85.09±0.19	84.50±0.20	85.48±0.39	84.76±0.21	98.84±0.01	99.38±0.01

3.5.4 对损失函数的消融对比实验

本研究的损失函数(见式(15))由动静脉损失、动静脉交汇处损失和血管整体损失 3 部分组成。为了进一步验证该损失函数的有效性,实验对损失函数的后两项进行了消元处理,

即每次消除其中一部分损失。实验在 DualModal2019 双模态数据集上进行,量化结果如表 8 所列。实验结果表明: SAM-Retina 损失函数消去动静脉交汇处损失或血管整体损失均会使动静脉分割结果的评估指标不同程度地下降。当消去动静脉

脉交汇处损失时,动脉分割结果的 F1 分数下降了 0.6 个百分点,静脉的 F1 分数下降了 0.18 个百分点。因此,如果不额外

增加动静脉交汇处损失的权重,模型更倾向于将交汇处分割为静脉。

表 8 SAM-Retina 对损失函数的消融对比实验

Table 8 Ablation comparison experiment on the loss function for SAM-Retina

(%)

任务	模型	F1	MCC	Sensitivity	Precision	Accuracy	Specificity
动脉	消去动静脉交汇处损失	82.75±0.34	81.24±0.36	79.54±0.71	84.62±0.93	98.62±0.05	99.41±0.03
	消去血管整体损失	83.11±0.21	81.95±0.24	80.82±0.48	85.63±0.73	98.63±0.02	99.46±0.02
	SAM-Retina 损失函数	83.35±0.19	82.79±0.20	80.85±0.49	86.08±0.36	98.79±0.01	99.49±0.02
静脉	消去动静脉交汇处损失	84.91±0.20	84.32±0.23	84.37±0.38	84.69±0.32	98.79±0.01	99.37±0.01
	消去血管整体损失	84.78±0.21	84.24±0.25	84.27±0.75	84.12±0.54	98.63±0.02	99.32±0.02
	SAM-Retina 损失函数	85.09±0.19	84.50±0.20	85.48±0.39	84.76±0.21	98.84±0.01	99.38±0.01

结束语 本文提出了一种基于 SAM 的视网膜动静脉分割模型 SAM-Retina。不同于以往大多数分割方法, SAM-Retina 使用双模态图像代替 RGB 图像作为输入,通过新增特征融合器融合了双模态图像的特征,以便在不扩大模型图像编码器规模的情况下增加模型能够获取到的特征信息;在图像编码器中新插入 Adapter 模块,提高了模型对视网膜医学图像特征的提取能力;同时去除原有的动态提示编码器,使用静态提示嵌入,减少了人工操作导致的失误。多组实验结果表明, SAM-Retina 在视网膜动静脉分割领域不仅在精度上领先于传统医学图像分割网络,而且表现出强大的鲁棒性和泛化能力,能够有效抑制噪声干扰,精确处理健康和病变患者的眼底图像。相比于 SAM 基础模型,所提模型架构更精简,迁移训练的参数量约为原来 1/10,但模型自动化程度更好,综合性能更优。当前研究结果对于推动复杂医学影像分析的自动化和智能化进程具有重要的应用价值和意义。

参考文献

[1] CHEN Y X, WEN L, PEI C W, et al. Changes of microvascular diameter in non-proliferative diabetic retinopathy[J]. International Eye Science, 2021, 9(21): 1632-1636.

[2] MURSCH-EDLMAYR A S, BOLZ M, STROHMAIER C. Vascular aspects in glaucoma: from pathogenesis to therapeutic approaches[J]. International Journal of Molecular Sciences, 2021, 22(9): 4662.

[3] CHEUNG C Y, MOK V, FOSTER P J, et al. Retinal imaging in Alzheimer's disease[J]. Journal of Neurology, Neurosurgery & Psychiatry, 2021, 92(9): 983-994.

[4] RIM T H, TEO A W J, YANG H H S, et al. Retinal vascular signs and cerebrovascular diseases[J]. Journal of Neuro-Ophthalmology, 2020, 40(1): 44-59.

[5] IKRAM M K, DE JONG F J, VINGERLING J R, et al. Are retinal arteriolar or venular diameters associated with markers for cardiovascular disorders? The Rotterdam Study[J]. Investigative Ophthalmology & Visual Science, 2004, 45(7): 2129-2134.

[6] DASHTBOZORG B, MENDONÇA A M, CAMPILHO A. An automatic method for the estimation of arteriolar-to-venular ratio in retinal images[C]// Proceedings of the 26th IEEE International Symposium on Computer-based Medical Systems. IEEE, 2013: 512-513.

[7] CIURICĂ S, LOPEZ-SUBLET M, LOEYS B L, et al. Arterial tortuosity: novel implications for an old phenotype[J]. Hyper-

tension, 2019, 73(5): 951-960.

[8] TELISCHAK N A, YEDAVALI V, MASSOUD T F. Tortuosity of superior cerebral veins: comparative magnetic resonance imaging morphometrics in normal subjects and arteriovenous malformation patients[J]. Clinical Anatomy, 2021, 34(3): 326-332.

[9] YU S, LAKSHMINARAYANAN V. Fractal dimension and retinal pathology: a meta-analysis[J]. Applied Sciences, 2021, 11(5): 2376.

[10] MILANI P, MONTESANO G, ROSSETTI L, et al. Vessel density, retinal thickness, and choriocapillaris vascular flow in myopic eyes on OCT angiography[J]. Graefes' Archive for Clinical and Experimental Ophthalmology, 2018, 256: 1419-1427.

[11] DASHTBOZORG B, MENDONÇA A M, CAMPILHO A. An Automatic Graph-Based Approach for Artery/Vein Classification in Retinal Images[J]. IEEE Transactions on Image Processing, 2014, 23(3): 1073-1083.

[12] ZHAO Y, XIE J, ZHANG H, et al. Retinal vascular network topology reconstruction and artery/vein classification via dominant set clustering[J]. IEEE Transactions on Medical Imaging, 2019, 39(2): 341-356.

[13] WELIKALA R A, FOSTER P J, WHINCUP P H, et al. Automated arteriole and venule classification using deep learning for retinal images from the UK Biobank cohort[J]. Computers in Biology and Medicine, 2017, 90: 23-32.

[14] REMESEIRO B, MENDONÇA A M, CAMPILHO A. Automatic classification of retinal blood vessels based on multilevel thresholding and graph propagation[J]. The Visual Computer, 2021, 37: 1247-1261.

[16] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]// Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer, 2015: 234-241.

[17] HEMELINGS R, ELEN B, STALMANS I, et al. Artery-vein segmentation in fundus images using a fully convolutional network[J]. Computerized Medical Imaging and Graphics, 2019, 76: 101636.

[18] KARLSSON R A, HARDARSON S H. Artery vein classification in fundus images using serially connected U-Nets[J]. Computer Methods and Programs in Biomedicine, 2022, 216: 106650.

- [19] CHEN W, YU S, MA K, et al. TW-GAN: Topology and width aware GAN for retinal artery/vein classification[J]. *Medical Image Analysis*, 2022, 77: 102340.
- [20] ZHANG J, YANG K, SHEN Z, et al. End-to-End Automatic Classification of Retinal Vessel Based on Generative Adversarial Networks with Improved U-Net[J]. *Diagnostics*, 2023, 13(6): 1148.
- [21] WALD T, ROY S, KOEHLER G, et al. SAM, MD: Zero-shot medical image segmentation capabilities of the Segment Anything Model[C]// *Medical Imaging with Deep Learning*, 2023.
- [22] MA J, HE Y, LI F, et al. Segment Anything in Medical Images[J]. arXiv:2304.12306, 2023.
- [23] MAZUROWSKI M A, DONG H, GU H, et al. Segment anything model for medical image analysis: an experimental study[J]. *Medical Image Analysis*, 2023, 89: 102918.
- [24] DENG R, CUI C, LIU Q, et al. Segment anything model(sam) for digital pathology: Assess zero-shot segmentation on whole slide imaging[J]. arXiv:2304.04155, 2023.
- [25] KIRILLOV A, MINTUN E, RAVI N, et al. Segment Anything[C]// *2023 IEEE/CVF International Conference on Computer Vision*, 2023.
- [26] CHEN H Y. Calibration of Retinal Oximetry Devices by Fundus-simulating Phantoms[D]. Hefei: University of Science and Technology of China, 2019.
- [27] TEAM G, ANIL R, BORGEAUD S, et al. Gemini: a family of highly capable multimodal models[J]. arXiv:2312.11805, 2023.
- [28] ROMBACH R, BLATTMANN A, LORENZ D, et al. High-resolution image synthesis with latent diffusion models[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022: 10684-10695.
- [29] RAMESH A, PAVLOV M, GOH G, et al. Zero-shot text-to-image generation[C]// *International Conference on Machine Learning*, PMLR, 2021: 8821-8831.
- [30] WU J, ZHANG Y, FU R, et al. Medical SAM Adapter: Adapting Segment Anything Model for Medical Image Segmentation[J]. arXiv:2306.12620, 2023.
- [31] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv:2010.11929, 2020.
- [32] HU E J, SHEN Y, WALLIS P, et al. LORA: Low-rank adaptation of large language models[J]. arXiv:2106.09685, 2021.
- [33] ZHANG S, ZHENG R, LUO Y, et al. Simultaneous Arteriole and Venule Segmentation of Dual-Modal Fundus Images Using a Multi-Task Cascade Network[J]. *IEEE Access*, 2019, 7: 57561-57573.



XU Hengyu, born in 1999, postgraduate, is a member of CCF(No. X8654G). His main research interest is deep learning.



XU Lin, born in 1981, master, lecturer, master supervisor. Her main research interests include smart health and computer vision.

(责任编辑:何杨)