

基于动态 Token 聚合注意力机制的双模态 OCTA 分割网络

李鸿飞, 王文念, 梅嫣嫣, 彭磊

山东第一医科大学 (山东省医学科学院) 医学信息与人工智能学院, 山东 泰安 271016

DOI: 10.61369/TACS.2026040036

摘要 : 针对光学相干断层扫描血管成像 (OCTA) 中细小血管信号微弱、受散斑噪声干扰导致单模态分割算法易产生末梢断裂与连通性下降等问题, 本文提出一种双模态视网膜血管分割网络 DDTA-UNet。网络以 U-Net 为基础框架, 融合 OCTA 血流信息与 OCT 结构先验作为双通道输入, 在编码端引入多头动态 Token 聚合注意力模块 (MHDTAA), 以高效建模血管全局拓扑依赖, 在跳跃连接处串联 CBAM 与注意力门控机制对多尺度特征进行联合筛选, 抑制背景噪声干扰。实验结果表明, DDTA-UNet 在 OCTA-500 数据集的 FAZ 分割与大血管分割任务上均展现出较好性能, 验证了双模态融合与动态注意力机制在 OCTA 视网膜血管分割场景中的有效性。

关键词 : 光学相干断层扫描血管造影 (OCTA); 视网膜血管分割; 动态 Transformer; 注意力机制; 双模态融合

Dual-Modal OCTA Segmentation Network Based on Dynamic Token Aggregation Attention Mechanism

Li Hongfei, Wang Wennian, Mei Yanyan, Peng Lei

College of Medical Information and Artificial Intelligence Shandong First Medical University(Shandong Academy of Medical Sciences), Tai'an, Shandong 271016

Abstract : To address the problems of weak tiny vessel signal and speckle noise interference in Optical Coherence Tomography Angiography (OCTA), which lead to tip breakage and poor connectivity of single-modal segmentation algorithms, this paper proposes a dual-modal retinal vessel segmentation network named DDTA-UNet. Based on the U-Net framework, the network takes OCTA blood flow information and OCT structural prior as dual-channel input, introduces a Multi-Head Dynamic Token Aggregation Attention (MHDTAA) module at the encoder to efficiently model the global topological dependence of blood vessels, and serially connects CBAM and attention gating mechanism at the skip connections to jointly filter multi-scale features and suppress background noise interference. Experimental results show that DDTA-UNet achieves excellent performance on both FAZ segmentation and large vessel segmentation tasks on the OCTA-500 dataset, which verifies the effectiveness of dual-modal fusion and dynamic attention mechanism in OCTA retinal vessel segmentation scenarios.

Keywords : Optical Coherence Tomography Angiography (OCTA); retinal vessel segmentation; dynamic Transformer; attention mechanism; dual-modal fusion

引言

目前, 眼科疾病是较为普遍的疾病之一。在眼底疾病的诊断过程中, 视网膜眼底图像提供了宝贵的病变特征信息^[1,2]。细致分析这些图像, 可以获取关于眼底病变的病理学数据, 视网膜血管的密度和形态特征是诊断眼底疾病时不可或缺的重要指标。医学成像技术的飞速发展显著提升了疾病诊断的精确度。光学相干断层扫描血管造影 (OCTA) 技术^[3,4]保留了光学相干断层扫描 (OCT) 技术的无创性和三维微米级分辨率特性。OCTA 不仅能详尽地揭示视网膜结构, 还增添了血流动态的可视化展示, 极大地丰富了临床诊断信息。

传统的 OCTA 技术虽然能获取三维视网膜数据, 但是临床量化指标如中央凹无血管区 (FAZ) 面积通常基于二维投影图进行, 涉及视网膜层的精确分割。近年来, 有研究将 U-Net 和 3D U-Net 等网络模型应用于图像分割领域^[5,6]。然而, 这些模型在处理 OCTA 数据以提取二维视网膜指标时, 仍面临依赖大量标注训练数据的难题。

本文提出一种融合 OCTA 血流信息与 OCT 结构先验的双模态视网膜血管分割网络 DDTA-UNet。该网络以 U-Net 为基础框架, 通过动态 Token 聚合注意力机制在控制计算开销的前提下建模血管全局拓扑依赖, 并在跳跃连接处引入多尺度串联注意力筛选以抑制背景噪声干扰, 主要面向视网膜大血管分割与 FAZ 分割两项任务。

一、技术背景

（一）视网膜血管分割

OCT 图像在血流成像方面的局限性导致了相关研究的不充分。OCTA 技术通过提供高分辨率的三维血管成像，可量化视网膜血管分析。

现有的分割方法，虽然可以抑制噪声，但是难以区分大血管与毛细血管，且依赖投影图。本文所提 DDTA-UNet 在编码器端引入 MHDTAA 模块，在有效控制计算复杂度的前提下建模血管全局结构，弥补卷积方法在全局建模方面的不足，能够提升分割精度并降低层分割误差对定量分析的影响。

（二）无血管区分割

在 OCTA 技术普及之前，FAZ 的边界描述较为模糊，相关研究相对匮乏。本文通过引入 OCT 图像提供的结构先验，与 OCTA 血流信息进行双模态融合，从数据层面强化 FAZ 区域的结构约束，以改善边界分割准确性，避免了对视网膜层分割的依赖和错误层分割的影响。

（三）注意力机制与多模态融合

在医学图像分割领域，U-Net 等网络模型展现了优异的性能。然而，这些网络利用三维数据生成二维投影图时存在局限，无法有效提取特征。

Transformer 结构的引入显著提升了全局特征建模能力，TransUNet 和 OCT²Former 均在 OCTA 血管分割任务中取得了较好性能。然而，标准 Transformer 的自注意力计算复杂度随图像分辨率呈平方级增长，在高分辨率 OCTA 图像处理中计算开销较大，且直接跳跃连接引入的浅层背景噪声会干扰解码器的精细重建。CBAM 与注意力门控机制分别通过通道-空间联合加权和语义引导门控对跳跃连接特征实施自适应筛选，本文将二者串联部署，系统性地抑制噪声特征对解码器路径的干扰。

二、方法

本节依照 DDTA-UNet 的前向信息流阐述各核心模块的设计原理，网络整体架构如图 1 所示。

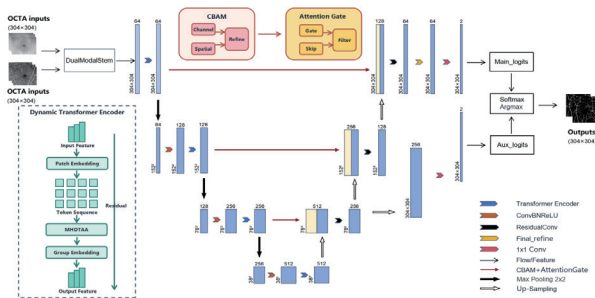


图 1 DDTA-UNet 整体网络架构图

（一）总体架构

DDTA-UNet 以 U-Net 的 U 型编码器-解码器框架为基

础，以配对的 OCTA-3M 与 OCT-3M 图像作为双通道输入。网络前端 DualModalStem 模块完成跨模态浅层特征融合，编码器由四个编码块（enc1~enc4）组成，以动态 Transformer 编码模块为核心，逐层提取多尺度全局结构特征，每层后接 2×2 最大池化进行空间下采样。编码器浅层特征（e1~e3）在传入解码器之前经 CBAM 与注意力门控机制串联筛选，解码器通过双线性上采样与 ResidualConv 逐步重建空间细节，输出头前附加 final_refine 模块执行特征精炼，训练阶段同时激活辅助输出头以加速中间层收敛。

（二）动态 Transformer 编码模块

标准 Transformer 自注意力在高分辨率特征图上的计算复杂度为 $O(N^2)$ ($N=H \times W$)，在显存受限条件下难以直接部署；同时纯 Transformer 架构缺乏局部归纳偏置，在小规模医学数据集上收敛偏慢。动态 Transformer 编码模块由 PatchEmbed、MHDTAA、GroupEmbedding 与残差连接四个子模块构成，在 enc1 ~ enc4 中逐层部署。PatchEmbed 将二维特征图展平为 Token 序列，其数学表达如式 (1) 所示：

$$X = \text{PatchEmbed}(F) \quad (1)$$

MHDTAA 引入 M 个可学习查询向量 ($M \ll N$)，通过交叉注意力将 N 个 Token 动态聚合为 M 个语义 Token 并迭代，更新规则如式 (2) 所示：

$$Q^{(t+1)} = \text{softmax}\left(\frac{Q_{\text{learn}} \cdot K^T}{\sqrt{d}}\right) \cdot V \quad (2)$$

该操作将计算复杂度降至 $O(MN)$ 。GroupEmbedding 通过分组卷积补偿 Token 序列化引发的空间信息损失，并叠加可学习位置嵌入，最终输出与输入残差相加。该模块在 enc1 ~ enc4 中逐层复用，有效降低高分辨率特征图的计算开销，同时保持视网膜血管拓扑的全局连通性。

（三）跳跃连接模块

为抑制标准跳跃连接引入的背景噪声干扰，本文在 e1~e3 传入解码器之前串联 CBAM 与注意力门控两个子模块。CBAM 依次施加通道注意力与空间注意力对特征执行自适应重标定，其中 σ 为 Sigmoid 函数，GAP 与 GMP 分别为全局平均池化与全局最大池化， $f_{7 \times 7}$ 为 7×7 卷积，通道注意力与空间注意力分别如式 (3)、式 (4) 所示：

$$M_c(F) = \sigma\left(MLP(GAP(F)) + MLP(GMP(F))\right) \quad (3)$$

$$M_s(F) = \sigma\left(f_{7 \times 7}\left([GAP_c(F); GMP_c(F)]\right)\right) \quad (4)$$

注意力门控以解码器深层特征 d_j 为门控信号，对 CBAM 筛选后的浅层特征 e_i' 实施语义引导。权重分配，其中 α 为逐元素门控系数， $\diamond$ 为逐元素张量积， W_s 与 W_g 分别为跳跃路径与门控路径的线性投影，门控系数计算及加权输出如式 (5) 所示：

$$\alpha = \sigma\left(BN(W_s(e_i') + W_g(d_j))\right) \quad (5)$$

$$e_i = \alpha \diamond e_i'$$

两阶段串联筛选使跳跃连接特征在通道-空间维度完成自适应重标定后，进一步受解码器深层语义的引导，系统性地压制背景噪声对解码器路径的干扰。

（四）其他模块

DualModalStem 为双模态融合入口，通过两个独立子网络分别提取 OCTA 与 OCT 浅层特征并计算差异图，三路特征拼接后经降维与联合筛选完成跨模态整合。ConvBNReLU 为全网通用的卷积 - 批归一化 - 激活三元组基础单元，承担逐层特征变换。ResidualConv 在双层卷积路径上引入残差捷径连接，缓解深层网络梯度消失问题。final_refine 部署于主输出头前，对解码器最终特征图执行精炼卷积，以提升血管边界预测的精度。

（五）训练策略

针对血管像素占比低的类别不平衡问题，采用交叉熵损失与 Dice 损失联合构成分割损失（式 (6)），二者互补以同时约束像素级分类准确性与区域重叠程度：

$$L_{seg} = L_{CE} + L_{Dice} \tag{6}$$

为加速中间层收敛，网络引入辅助监督分支（AUX 头），训练总损失由主输出损失与辅助损失加权叠加构成（式 (7)）；辅助分支仅在训练阶段激活，推理阶段自动关闭：

$$L_{total} = L_{main} + 0.2 \times L_{aux} \tag{7}$$

（六）可视化结果分析

1. 大血管分割结果

图2展示了 DDTA-UNet 在大血管分割任务中3例代表性样本的可视化结果，可见 DDTA-UNet 的优势主要集中于极细末梢毛细血管区域。

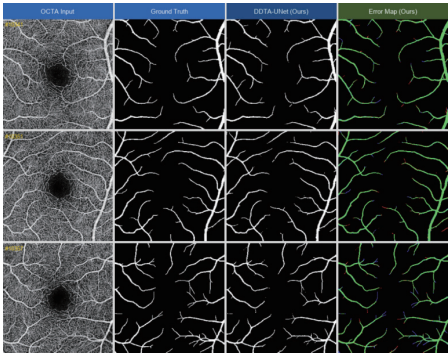


图 2 DDTA-UNet 大血管分割可视化结果

2. 无血管区（FAZ）分割结果

图3展示了 FAZ 分割任务中 DDTA-UNet 与 OCT²Former 的对比结果。可见 OCT²Former 在部分样本 FAZ 边界区域出现分割碎裂，DDTA-UNet 输出的 FAZ 边界整体更为平滑连续，错误像素极少。

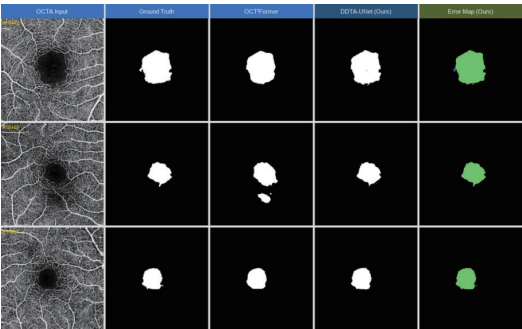


图 3 FAZ 分割可视化对比结果

3. 定量对比分析

由表1和表2可知，本文方法在 OCTA 图像的无血管区分割和血管区分割上的 Dice 系数分别达 97.30% 和 91.85%，JAC、精确率均高于对比方法，充分证明了该网络具有优异的分割性能。

表 1 大血管区分割定量对比结果

方法	Dice	Jaccard	精确率	召回率
U-Net	90.60	82.88	--	--
CE-Net	90.87	83.33	98.80	89.34
TransUNet	89.89	81.70	98.66	87.88
OCT ² Former	91.93	85.13	98.94	91.45
本文方法	91.85	85.20	98.68	91.54

表 2 无血管区分割定量对比结果

方法	Dice	Jaccard	灵敏度	特异度
OCT ² Former	90.53	83.19	91.56	99.62
本文方法	97.30	94.73	96.44	99.79
提升幅度	+6.77	+11.54	+4.88	+0.17

三、结语

本文针对 OCTA 图像低对比度与单模态信息不足的问题，提出双模态视网膜血管分割网络 DDTA-UNet。通过 MHDTAA 动态 Token 聚合机制实现高效全局拓扑建模，结合 CBAM 与注意力门控机制对跳跃连接特征实施联合筛选，并以双模态融合模块引入 OCT 结构先验信息。

在 OCTA-500 数据集上，DDTA-UNet 在 FAZ 分割 Dice 系数达 97.30%，大血管分割 Dice 系数达 91.85%，在 Jaccard 系数与召回率方面表现突出。后续可从扩充训练数据与模型轻量化方向继续优化，以更好地服务于眼科智能筛查临床应用。

参考文献

[1] 朱锦亭, 冯洁. 光学相干断层扫描血管成像术评估成年近视患者黄斑区视网膜血流密度和视网膜厚度 [J]. 中国临床研究, 2024, 37(12): 1895-1899.

[2] 杨雄一, 易果果, 陈艳霞, 等. 应用 OCTA 观察糖尿病视网膜病变的微血管变化 [J]. 国际眼科杂志, 2025, 25(2): 179-190.

[3] 陈子华, 张宏. 光学相干断层扫描血管成像在眼前节疾病中的研究现状 [J]. 中国医药科学, 2024, 14(8): 62-65.

[4] 刘自强, 接传红, 王建伟, 等. OCTA 在神经眼科疾病中应用的研究进展 [J]. 中国中医眼科杂志, 2024, 34(5): 477-480.

[5] 汪有崧, 裴峻鹏, 李增辉, 等. 深度学习的视网膜血管分割研究综述 [J]. 计算机科学与探索, 2024, 18(8): 1960-1978.

[6] 刘新娟, 韩旭, 方二喜. 基于并行 U-Net 模型的眼底微血管图像分割方法 [J]. 中国激光, 2024, 51(21): 103-116.