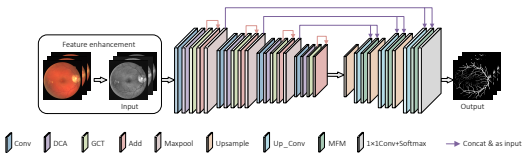
OEE | 光电工程 Opto-Electronic Engineering

Article

2026 年 3 月, 第 53 卷, 第 3 期

分组通道 Transformer 与调制融合的视网膜血管分割算法



梁礼明, 王灵亮*, 蔡锦辉, 王成斌

江西理工大学电气工程与自动化学院, 江西 赣州 341000

摘要：针对视网膜血管分割存在微细血管表征能力有限、复杂背景干扰和血管结构断裂等问题，提出一种分组通道 Transformer 与调制融合的视网膜血管分割算法。首先设计双维协同注意力模块于编码部分，增强微细血管的特征表达，有效提升血管分割能力。其次在编码端构建分组通道 Transformer 模块，旨在捕获血管之间的长程依赖关系，精确捕捉血管纹理信息。最后在解码端引入调制融合模块，有效融合编解码特征信息，抑制病灶区域噪声干扰。在数据集 DRIVE、CHASE_DB1 和 STARE 上进行实验，所提算法准确率分别为 0.9711、0.9762 和 0.9758，灵敏度分别为 0.8026、0.8219 和 0.8042，特异性分别为 0.9873、0.9866 和 0.9899。该方法在整体分割效果方面优于其他算法，展现出更好的分割性能。

关键词：视网膜血管分割；分组通道 Transformer；双维协同注意力；调制融合

The English title and abstract appear at the end of the paper.

DOI: 10.12086/oee.2026.250264 | CSTR: 32245.14.oee.2026.250264 | 中图分类号：TP391 | 文献标识码：A

梁礼明, 王灵亮, 蔡锦辉等. 分组通道 Transformer 与调制融合的视网膜血管分割算法 [J]. 光电工程, 2026, **53**(3): 250264

Liang L M, Wang L L, Cai J H et al. Grouped channel Transformer and modulation fusion retinal vessel segmentation algorithm[J]. *Opto-Electron Eng*, 2026, **53**(3): 250264

1 引　言

视网膜血管形态变化是多种疾病诊断的重要依据，如糖尿病视网膜病变、高血压及白内障等。在医疗诊断中，医生通常分析眼底图像中血管的直径和弯曲程度来辅助判断病情。然而，视网膜血管结构复杂多变，存在血管细小分支密集、图像对比度低及病变区域干扰等问题，给血管的精确分割带来了较大挑战。因此，研发一种高效、准确的视网膜血管分割技术至关重要[1-2]。

在视网膜血管分割领域，现有方法主要可分为无监督方法和有监督方法两大类。无监督方法不依赖人工标注数据，主要通过分析血管的形态特征实现分割，常见的方法包括基于匹配滤波[3]、基于边缘检测[4] 和基于模型[5] 等。然而，这类方法往往依赖人工构建的特征提取流程，难以全面捕捉血管复杂的拓扑结构，从而导致分割性能不高。相比之下，有监督方法则需专家提供的标注数据作为训练样本，在血管分割任务上实现更高的精度，此类方法可进一步分为传统机器学习方法与深度学习方法。传统机器学习方法包括支持向量机 (SVM)[6] 和集成特征提取[7] 等，然而其特征提取过程仍需人工设计，

收稿日期：2025-09-04；修回日期：2025-11-12；录用日期：2025-11-14；网络出版日期：2026-03-19

基金项目：国家自然科学基金资助项目 (51365017，61463018)；江西省自然科学基金资助项目 (20192BAB205084)；江西省教育厅科学技术研究青年项目 (GJJ2200848)

*通信作者：王灵亮，2325286035@qq.com

版权所有©2026 中国科学院光电技术研究所

无法充分表征血管的多尺度结构特征。

随着深度学习技术的快速发展，视网膜血管分割领域迎来了较大突破。Ronneberger 等[8] 提出的 U-Net 架构通过其独特的编码器-解码器结构和跳跃连接机制，在血管分割中表现出色，但其存在特征层级间的语义差异，难以精确提取血管的细节信息。随后，研究者们对 U-Net 进行了多种改进。Getahun 等[9] 构建一种基于编解码架构的全尺度网络，通过结合残差块与自适应阈值算法提升血管分割的精度。Radha 等[10] 设计一种融合 Gabor 滤波与深度可分离卷积的 U 型网络，提升对血管特征的捕获能力。尽管上述算法在分割性能上有所提升，但仍存在特征提取不充分和全局上下文信息捕获能力有限等问题。Transformer[11] 模型通过自注意力机制获取序列间的长程依赖性，在自然语言处理任务中取得较大成功。Dosovitskiy 等[12] 提出视觉 Transformer (ViT) 模型，通过将图像划分为块序列并应用自注意力机制，增强模型对全局依赖关系的建模能力，为视网膜血管分割任务提供新的解决方案。Wu 等[13] 构建一种跨尺度级联分层 Transformer 网络，通过在编码器中构建局部-全局跨尺度 Transformer 结构，并在解码器中引入位置感知 Transformer 机制，有效保留血管末端位置信息。Cui 等[14] 设计一种基于 Transformer 的联合注意力视网膜血管分割算法，通过多尺度特征融合与注意力机制捕捉上下文信息，增强血管特征权重。上述算法虽在血管分割任务中表现较好，但仍存在微小血管分割不足、血管边缘模糊和易受病变区域干扰等问题。

针对上述问题，本文提出一种分组通道 Transformer 与调制融合的视网膜血管分割算法 (grouped channel Transformer and modulation fusion network, GTMF-Net)，主要工作：1) 设计双维协同注意力模块于编码部分，增强血管特征的可辨识度，提升模型对微血管的分割精度；2) 应用分组通道 Transformer 模块构建双编码结构，建立全局信息依赖，有效捕获血管细节信息；3) 在解码端引入调制融合模块，增强编码部分和解码部分的特征融合，提升模型抗噪声干扰能力。

2 本文算法

2.1 算法设计

本文在 U 型网络结构的基础上，提出一种分组通道 Transformer 与调制融合的视网膜血管分割算法，其结构如图 1 所示。GTMF-Net 由特征强化单元、双维协同注意力模块、分组通道 Transformer 模块和调制融合模块四部分组成。首先特征强化单元采用滑窗裁剪对眼底图像进行扩充，对其应用灰度变换、直方图均衡化等多种预处理方法以增强图像对比度。然后在编码端设计双维协同注意力模块，自适应调整模型对不同血管区域的关注程度，增强微细血管结构信息的权重。再后借助分组通道 Transformer 模块设计双编码架构，有效捕获上下文信息，实现血管纹理特征的精确提取。最后将调制融合模块置于解码端以充分融合多尺度特征，同时抑制图像中的噪声信息。

2.2 双维协同注意力模块

在视网膜血管分割任务中，由于眼底图像中血管形态复杂多变，且血管与背景的对比度低，导致微细血管分割较困难。Yu 等[15] 提出基于离散余弦变换 (DCT) 的空间注意力模块，其对通道内各个空间位置分配不同权重，增强血管区域响应，实现模型对微血管信息的充分捕获。本文受文献 [15] 启发，设计一种双维协同注意

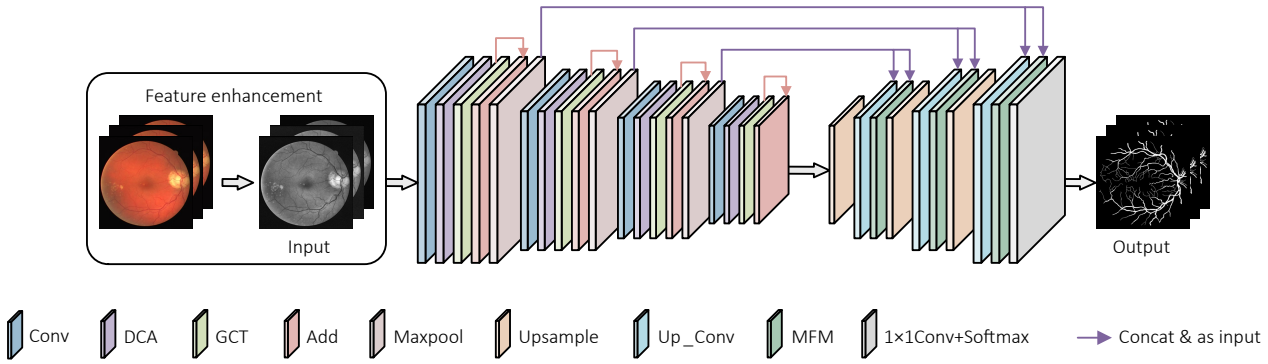


图 1 分组通道 Transformer 与调制融合的视网膜血管分割算法

Fig. 1 Grouped channel Transformer and modulation fusion retinal vessel segmentation algorithm

力模块 (dual-dimensional collaborative attention module, DCA)，旨在捕捉通道间和空间内的相互依赖关系，有效维持血管形态结构的完整性。其结构如图 2 所示。

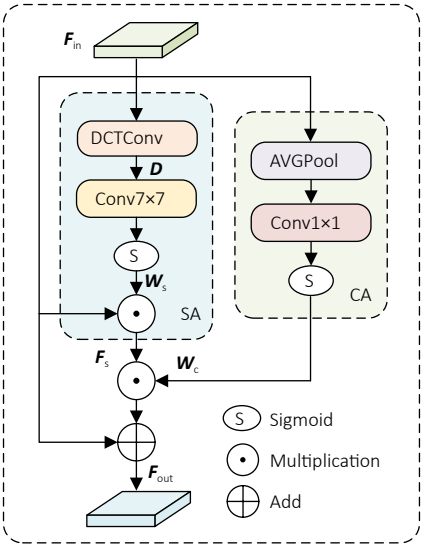


图 2 双维协同注意力模块

Fig. 2 Dual-dimensional collaborative attention module

DCA 包括空间注意力 (spatial attention, SA) 分支和通道注意力 (channel attention, CA) 分支，其中 SA 分支通过对各个空间位置进行加权，抑制无关背景区域。CA 分支通过调整各个通道的权重，使网络聚焦于血管特征。在 SA 分支中，首先对输入特征$\boldsymbol{F}_{\text{in}}$应用离散余弦变换卷积操作，以提取完整的频率特征 $\boldsymbol{D}$。然后通过 7×7 卷积层来建模位置间的相互依赖，并使用 Sigmoid 激活函数得到空间注意力权重$\boldsymbol{W}_{\text{s}}$。最后将权重$\boldsymbol{W}_{\text{s}}$与输入特征图$\boldsymbol{F}_{\text{in}}$进行相乘得到空间增强特征$\boldsymbol{F}_{\text{s}}$。其数学表达式分别为

$$\boldsymbol{D} = Conv_{\text{DCT}}(\boldsymbol{F}_{\text{in}}), \tag{1}$$

$$\boldsymbol{W}_{\text{s}} = \sigma(Conv_{7\times7}(\boldsymbol{D})), \tag{2}$$

$$\boldsymbol{F}_{\text{s}} = \boldsymbol{F}_{\text{in}} \odot \boldsymbol{W}_{\text{s}}, \tag{3}$$

式中：$Conv_{\text{DCT}}$表示离散余弦变换卷积操作；$Conv_{7\times7}$表示 7×7 卷积层；σ表示 Sigmoid 函数；⊙表示相乘。在 CA 分支中，首先对输入特征$\boldsymbol{F}_{\text{in}}$应用平均池化操作和 1×1 卷积操作，并通过 Sigmoid 函数得到通道注意力权重。然后将通道注意力权重$\boldsymbol{W}_{\text{c}}$与特征$\boldsymbol{F}_{\text{s}}$进行相乘，以突出血管区域。最后将上述结果与输入特征$\boldsymbol{F}_{\text{in}}$相加，得到输出特征$\boldsymbol{F}_{\text{out}}$。其数学表达式分别为

$$\boldsymbol{W}_{\text{c}} = \sigma(Conv_{1\times1}(AVGPool(\boldsymbol{F}_{\text{in}}))), \tag{4}$$

$$\boldsymbol{F}_{\text{out}} = (\boldsymbol{F}_{\text{s}} \odot \boldsymbol{W}_{\text{c}}) + \boldsymbol{F}_{\text{in}}, \tag{5}$$

式中：$AVGPool$表示平均池化操作；$Conv_{1\times1}$表示 1×1 卷积操作。

2.3 分组通道 Transformer 模块

由于传统卷积网络侧重于局部区域特征的提取，难以捕捉长程依赖关系。Li 等[16] 提出分组通道自注意力模块，其将输入特征图按通道维度划分为若干组，在各组内进行通道注意力计算，有效捕获全局上下文信息，增强模型对血管语义信息的表征能力。本文受文献 [16] 启发，并结合 ViT，提出分组通道 Transformer 模块 (grouped channel Transformer module, GCT)，其结构如图 3 所示。

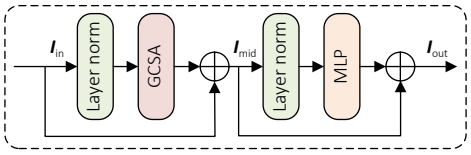


图 3 分组通道 Transformer 模块

Fig. 3 Grouped channel Transformer module

GCT 模块由分组通道自注意力模块 (grouped channel-wise self-attention module, GCSA)、层归一化 (layer norm, LN) 和多层感知机 (multilayer perceptron, MLP) 组成。其过程分别表示如下：

$$\boldsymbol{I}_{\text{mid}} = GCSA(LN(\boldsymbol{I}_{\text{in}})) + \boldsymbol{I}_{\text{in}}, \tag{6}$$

$$\boldsymbol{I}_{\text{out}} = MLP(LN(\boldsymbol{I}_{\text{mid}})) + \boldsymbol{I}_{\text{mid}}, \tag{7}$$

式中：$\boldsymbol{I}_{\text{in}}$表示输入；$\boldsymbol{I}_{\text{out}}$表示输出。

2.3.1 分组通道自注意力模块

GCSA 能实现对全局信息的有效建模，增强模型对微血管的识别能力，其结构如图 4 所示。首先给定输入特征$\boldsymbol{X} \in R^{C\times H\times W}$，应用线性投影操作和深度卷积操作，得到查询矩阵 $\boldsymbol{Q}$、键矩阵 $\boldsymbol{K}$ 和值矩阵 $\boldsymbol{V}$。然后将上述特征沿通道维度划分为 G 个子组并重塑为$\boldsymbol{Q}_{\text{g}}$、$\boldsymbol{K}_{\text{g}}$和$\boldsymbol{V}_{\text{g}}$。其次将$\boldsymbol{Q}_{\text{g}}$和$\boldsymbol{K}_{\text{g}}$进行点乘和缩放，并经过 Softmax 函数归一化后得到注意力矩阵$\boldsymbol{A}_{\text{g}}$。最后利用矩阵$\boldsymbol{A}_{\text{g}}$对$\boldsymbol{V}_{\text{g}}$进行加权求和，将加权结果进行拼接及线性投影操作得到输出特征 $\boldsymbol{Y}$。其在数学上分别表示如下：

$$(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = DConv_{d=2}(\Omega(\boldsymbol{X})), \tag{8}$$

$$(\boldsymbol{Q}_{\text{g}},\boldsymbol{K}_{\text{g}},\boldsymbol{V}_{\text{g}}) = GroupReshape_{\text{g}}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}), \tag{9}$$

$$\boldsymbol{A}_{\mathrm{g}}=Softmax\left(\frac{\boldsymbol{Q}_{\mathrm{g}}\boldsymbol{K}_{\mathrm{g}}}{t}\right), \tag{10}$$

$$\boldsymbol{Y}=\Omega(Concat(\boldsymbol{A}_1\boldsymbol{V}_1,\boldsymbol{A}_2\boldsymbol{V}_2,\cdots,\boldsymbol{A}_{\mathrm{g}}\boldsymbol{V}_{\mathrm{g}})), \tag{11}$$

式中：$\boldsymbol{Q}_{\mathrm{g}}\in\mathbf{R}^{\frac{C}{G}\times HW}$、$\boldsymbol{K}_{\mathrm{g}}\in\mathbf{R}^{HW\times\frac{C}{G}}$、$\boldsymbol{V}_{\mathrm{g}}\in\mathbf{R}^{HW\times\frac{C}{G}}$；$\Omega$表示线性投影操作；$DConv_{d=2}$表示膨胀率为 2 的深度卷积；$GroupReshape_{\mathrm{g}}$表示分组与形状变换；$t$ 表示可学习的参数。

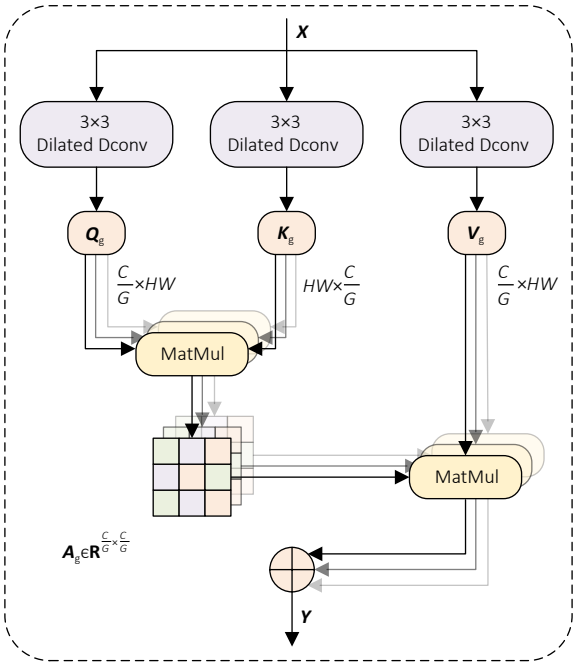


图 4 分组通道自注意力模块

Fig. 4 Grouped channel-wise self-attention module

2.4 调制融合模块

在原始 U-Net 架构中，编码路径与解码路径通过跳跃连接直接耦合，这种方式虽能传递空间细节信息，但编码端提取的低级特征与解码端提取的高级特征在语义上存在差异。为了将编解码特征充分融合，提高模型对血管信息的捕获能力并抑制背景噪声干扰，本文引入调制融合模块 (modulation fusion module, MFM)[17]，如图 5 所示。

首先将编码器特征$\boldsymbol{F}_{\mathrm{enc}}$与解码器特征$\boldsymbol{F}_{\mathrm{dec}}$进行相加，得到融合特征$\boldsymbol{F}_{\mathrm{sum}}$。然后将特征$\boldsymbol{F}_{\mathrm{sum}}$通过全局平均池化操作、多层感知机 (MLP) 和 Softmax 归一化运算，以生成通道注意力权重$\boldsymbol{A}_{\mathrm{attn}}$。再后将权重$\boldsymbol{A}_{\mathrm{attn}}$按通道维度分别与特征$\boldsymbol{F}_{\mathrm{enc}}$和特征$\boldsymbol{F}_{\mathrm{dec}}$相乘，并将上述结果相加得到特征$\boldsymbol{F}_{\mathrm{fus}}$，以增强血管特征的语义表达。最后将特征$\boldsymbol{F}_{\mathrm{fus}}$与特征$\boldsymbol{F}_{\mathrm{enc}}$进行拼接，对拼接结果应用卷积操作得到输出特征$\boldsymbol{F}_{\mathrm{out}}$。其在数学上分别表示为

$$\boldsymbol{F}_{\mathrm{sum}}=\boldsymbol{F}_{\mathrm{enc}}\oplus\boldsymbol{F}_{\mathrm{dec}}, \tag{12}$$

$$\boldsymbol{A}_{\mathrm{attn}}=Softmax(MLP(GAP(\boldsymbol{F}_{\mathrm{sum}}))), \tag{13}$$

$$\boldsymbol{F}_{\mathrm{fus}}=(\boldsymbol{A}_{\mathrm{attn}}\otimes\boldsymbol{F}_{\mathrm{enc}})\oplus(\boldsymbol{A}_{\mathrm{attn}}\otimes\boldsymbol{F}_{\mathrm{dec}}), \tag{14}$$

$$\boldsymbol{F}_{\mathrm{out}}=Conv(Concat(\boldsymbol{F}_{\mathrm{fus}},\boldsymbol{F}_{\mathrm{enc}})), \tag{15}$$

式中：*GAP* 表示全局平均池化操作；*Conv* 表示卷积操作；⊕表示相加；⊗表示相乘。

3 实验结果与分析

3.1 实验设计

本文实验在 64 位 Windows 11 操作系统环境下进行，硬件配置为 CPU Intel (R) Core (TM) i7-14650HX，运行内存 16 GB，GPU NVIDIA GeForce RTX 4060，显存 8 GB。算法基于 PyTorch 框架搭建，采用 Python 编程语言实现。在训练阶段，模型通过 Adam 优化器更新参数，使用交叉熵损失函数对训练过程进行监督。学习率设为 0.0005，总训练轮次为 20，批次大小固定为 6。为平衡模型收敛效率与过拟合风险，引入余弦退火策略动

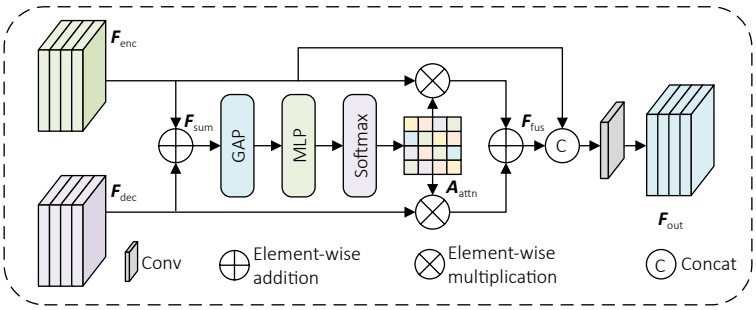


图 5 调制融合模块

Fig. 5 Modulation fusion module

态调整学习率，并设置早停机制，若连续 10 个训练周期未产生更优权重则自动终止训练。

3.2 数据集及预处理

本文采用 DRIVE[18]、CHASE_DB1[19] 和 STARE[20] 三个公开视网膜图像数据集进行算法验证。DRIVE 数据集包含 40 幅彩色眼底图像，每幅图像的分辨率为 565 pixel×584 pixel，其中 20 幅用于训练，另外 20 幅用于测试。CHASE_DB1 数据集采集自 14 名学龄儿童的双眼，共 28 幅分辨率为 999 pixel×960 pixel 的彩色视网膜图像，实验中选取前 20 幅作为训练集，其余 8 幅作为测试集。STARE 数据集包含 20 幅 605 pixel×700 pixel 的彩色眼底图像，因官方未提供训练集和测试集划分，故采用五折交叉验证法进行实验评估。上述 3 个数据集中，每幅视网膜图像均附有两位专家独立标注的金标准，本文选取第一位专家的标注结果作为基准参考。

由于原始眼底图像普遍存在亮度不均、对比度低及噪声干扰等问题，上述因素会影响模型的分割性能，故对数据集进行预处理。如图 6 所示，首先对原始图像进行灰度转换，降低计算冗余并突出血管形态结构。然后通过标准化操作调整像素分布范围，缓解亮度分布不均对图像质量的影响。再后应用限制对比度直方图均衡化，增强血管纹理与背景之间的局部对比度。最后采用伽马校正对整体亮度分布进行非线性调整，优化微细血管在较暗区域的显现效果。考虑到公开数据集样本量有限，为缓解训练过程中的过拟合风险，对数据集进行扩充处理。具体来说，采用 64×64 大小的滑动窗口从预处理后的原始图像及对应的金标准图像中提取 10000 个图像块。

3.3 评价指标

为了客观评估模型的性能，本文选取准确率 (Acc)、灵敏度 (Sen)、特异性 (Spe)、F1 分数 (F1) 和 ROC 曲线下面积 (AUC) 作为评价指标。其中，准确率表示正确分割的血管与背景像素之和占总像素的比例；灵敏度表示正确分割的血管像素占实际血管像数的比例；特异性表示正确分割的非血管像素占实际背景像素的比例；F1 分数综合评估模型对血管结构的捕获能力。其计算公式分别为

$$Acc = \frac{TP+TN}{TP+TN+FP+FN}, \tag{16}$$

$$Sen = \frac{TP}{TP+FN}, \tag{17}$$

$$Spe = \frac{TN}{TN+FP}, \tag{18}$$

$$F_1 = \frac{2TP}{2TP+FP+FN}, \tag{19}$$

式中：TP 为真阳性；TN 为真阴性；FP 为假阳性；FN 为假阴性。

3.4 不同分割算法结果对比

3.4.1 整体分割结果对比

为了验证本文算法的性能优势，在相同实验环境下，将 GTMF-Net 分割结果和其余四种算法分割结果进行比较，包括 U-Net[8]、Attention U-Net[21]、FR U-Net[22] 和 GT-DLA-dsHFF[23]。U-Net 是基准模型，Attention U-Net 是在 U-Net 基础上引入注意力机制的经典改进模型，FR U-Net 是近年来分割性能较好的模型，GT-DLA-dsHFF 是近年来将 Transformer 结构与 U-Net 框架相结合的先进模型。如图 7 所示，其中第一、二行为 DRIVE 数据集中眼底图像，第三、四行为 CHASE_DB1 数据集中眼底图像，第五、六行为 STARE 数据集中眼底图像。

观察 DRIVE 数据集第一行分割结果可知，U-Net 无法有效提取交叉处微细血管的形态特征，导致其丢失较多的血管信息，Attention U-Net 分割细小血管时出现血管分割不足问题，FR U-Net 和 GT-DLA-dsHFF 在分支血管末端处存在血管断裂现象，而 GTMF-Net 在微细血管的分割上取得良好的效果。观察 DRIVE 数据集第二行分割结果可知，由于视网膜边界区域的背景与前景

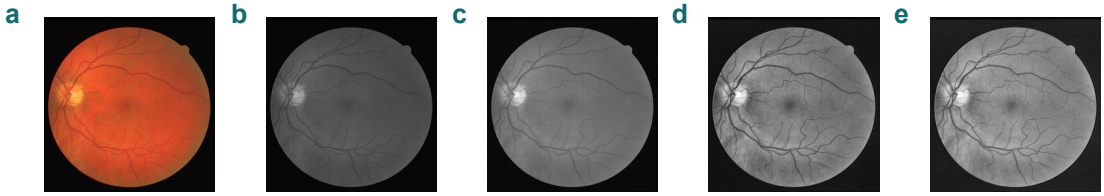


图 6 预处理图像。(a) 原图；(b) 灰度图；(c) 标准化图；(d) 直方图均衡化图；(e) 伽马校正图

Fig. 6 Preprocessed image. (a) Original image; (b) Grayscale image; (c) Normalized image; (d) CLAHE image; (e) Gamma corrected image

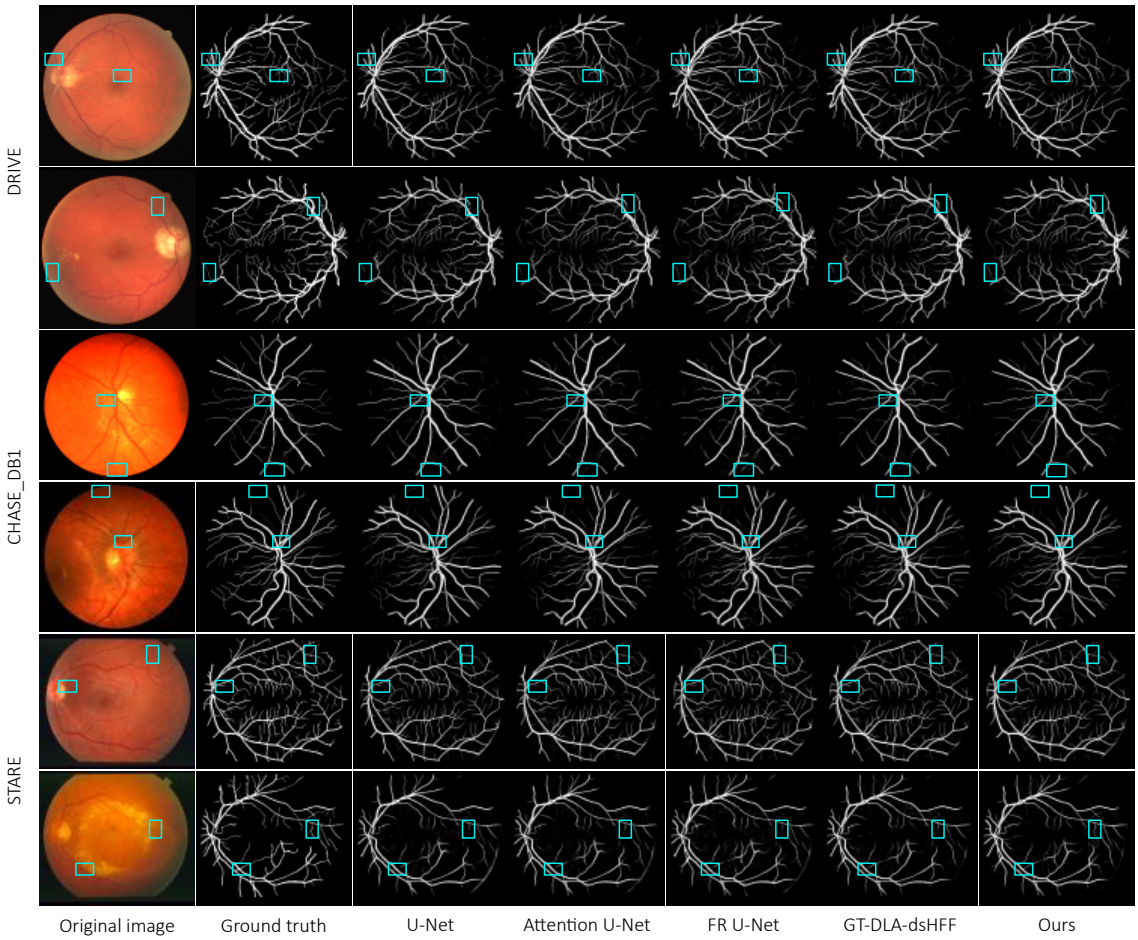


图 7 不同算法分割结果对比

Fig. 7 Comparison of segmentation results with different algorithms

对比度低，U-Net 和 Attention U-Net 出现血管识别不充分问题，FR U-Net、GT-DLA-dsHFF 和 GTMF-Net 相对较好地保留血管信息，但 FR U-Net和 GT-DLA-dsHFF 存在主血管轮廓不清晰现象。观察 CHASE_DB1 数据集第三和第四行分割结果可知，U-Net、FR U-Net 和 GT-DLA-dsHFF 将病灶区域错误分类为血管，Attention U-Net 出现主血管像素缺失问题，而 GTMF-Net 能够准确区分背景与血管，较完整地分割出主血管。观察 STARE 数据集第五行分割结果可知，由于视盘区域周围光照不均匀，U-Net、Attention U-Net 和 FR U-Net 出现微血管不连续问题，GT-DLA-dsHFF 和 GTMF-Net 存在问题相对较小。观察 STARE 数据集第六行分割结果可知，受中央视网膜动脉阻塞病变影响，U-Net、Attention U-Net 和 FR U-Net 存在血管分割模糊现象，GT-DLA-dsHFF 出现血管末梢断连问题，GTMF-Net 分割出的血管结构更加清晰，表明 GTMF-Net 在视网膜图像分割中具有更优的鲁棒性。

3.4.2 细节分割结果对比

为了更深入地体现所提算法的优势，图 8 展示了不同算法在 DRIVE、CHASE_DB1 和 STARE 数据集上的局部放大细节，其中图 8 第一、第二和第三行细节分别源自图 7 第一、第四和第六行的标注区域。

由第一行的分割细节图可知，U-Net 出现分支处细小血管信息缺失问题较为严重，Attention U-Net、FR U-Net 和 GT-DLA-dsHFF 存在血管结构断连现象，GTMF-Net 则较好地保持血管结构的完整性。由第二行的分割细节图可知，U-Net、FR U-Net 和 GT-DLA-dsHFF 受病灶区域影响，将病理信息分割为血管，Attention U-Net 出现主血管分割不完全问题，而 GTMF-Net 能较清晰地分割出主血管结构。由第三行的分割细节图可知，U-Net、Attention U-Net 和 FR U-Net 在血管交叉区域存在分割断裂及微细血管像素缺失问题，GT-DLA-dsHFF 和

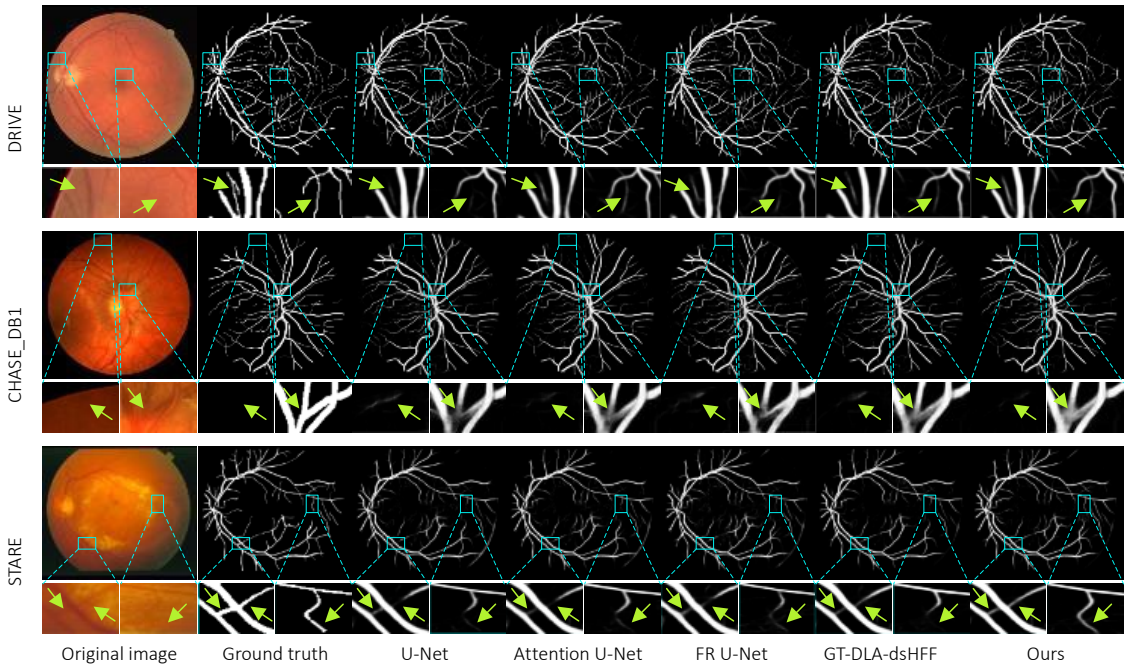


图 8 不同算法分割结果细节对比

Fig. 8 Detailed comparison of segmentation results with different algorithms

GTMF-Net 相对较好地保留交叉处的血管信息，但 GT-DLA-dsHFF 存在微细血管分割不足现象，相比之下，GTMF-Net 能够更准确地分割出血管轮廓，分割结果与金标准基本吻合。实验表明，GTMF-Net 在 DRIVE、CHASE_DB1 和 STARE 数据集上实现对血管的准确识别，相较其余对比算法具有更优的分割性能。

3.4.3 性能指标对比

为了定量评估所提算法的性能，在相同实验环境下，将 GTMF-Net 与文献 [8] 及文献 [21-23] 在 DRIVE、CHASE_DB1 和 STARE 这 3 个数据集上进行比较，性能指标对比结果如表 1 所示。

观察表 1 可知，在 DRIVE 数据集上，Attention U-

表 1 不同算法性能指标对比

Table 1 Comparison of performance metrics of different algorithms

Dataset	Method	Acc	Sen	Spe	AUC	F1
DRIVE	U-Net[8]	0.9704	0.7954	0.9872	0.9871	0.8250
	Attention U-Net[21]	0.9710	0.7963	**0.9878**	0.9882	0.8282
	FR U-Net[22]	0.9705	0.8013	0.9867	0.9876	0.8264
	GT-DLA-dsHFF[23]	0.9710	0.8006	0.9874	0.9881	0.8289
	Ours	**0.9711**	**0.8026**	0.9873	**0.9884**	**0.8298**
CHASE_DB1	U-Net[8]	0.9747	0.8085	0.9859	0.9885	0.8015
	Attention U-Net[21]	0.9760	0.8037	**0.9876**	0.9899	0.8091
	FR U-Net[22]	0.9742	0.8043	0.9856	0.9878	0.7975
	GT-DLA-dsHFF[23]	0.9758	0.8075	0.9871	0.9897	0.8081
	Ours	**0.9762**	**0.8219**	0.9866	**0.9901**	**0.8136**
STARE	U-Net[8]	0.9753	0.7924	0.9903	0.9903	0.8305
	Attention U-Net[21]	0.9756	0.7880	**0.9910**	0.9905	0.8312
	FR U-Net[22]	0.9748	0.8039	0.9889	0.9901	0.8299
	GT-DLA-dsHFF[23]	0.9756	0.7997	0.9901	0.9908	0.8335
	Ours	**0.9758**	**0.8042**	0.9899	**0.9911**	**0.8350**

Net 利用注意力门机制动态调整特征响应，其特异性达到最优值，但 GTMF-Net 的准确率、灵敏度、AUC 值和 F1 分数均比其高，表明 Attention U-Net 虽然能有效识别非血管区域，但其对血管和背景的整体判别效果有待提高。在 CHASE_DB1 数据集上，GT-DLA-dsHFF 利用全局 Transformer 建模像素间长距离依赖关系，其在特异性上高于 GTMF-Net，但 GTMF-Net 其余指标均为最优，其中灵敏度较 GT-DLA-dsHFF 提升 1.44%，表明 GTMF-Net 捕获微细血管的能力较强。在 STARE 数据集上，除 FR U-Net 算法外，其余算法的特异性较 GTMF-Net 均有提升，原因是 GTMF-Net 利用分组通道自注意力机制有效提升微血管分割能力，但会导致模型将一些背景区域识别为血管，进而影响特异性的表现；但 GTMF-Net 在准确率、灵敏度、AUC 值和 F1 分数等指标上均取得最优值，表明 GTMF-Net 在血管分割的综合表现上更为出色。

为了直观体现所提算法的性能，图 9 和图 10 展示了上述 5 个模型在 DRIVE 和 CHASE_DB1 数据集上的 ROC 曲线和 PR 曲线对比图。由图 9 和图 10 可知，GTMF-Net 的 ROC 曲线更接近坐标系的左上区域，PR 曲线更接近坐标系的右上区域，AUC (ROC) 值和 AUC (PR) 值均为最高，表明 GTMF-Net 能够更全面地识别血管结构，较其余对比算法分割性能更优。

3.4.4 与近年先进算法比较

为了进一步验证所提算法的先进性，将 GTMF-Net 与近年其他先进算法[24-31]在 Acc、Sen、Spe、AUC 和 F1 这 5 个指标上进行对比，如表 2~表 4 所示，其数据均引自原文献。

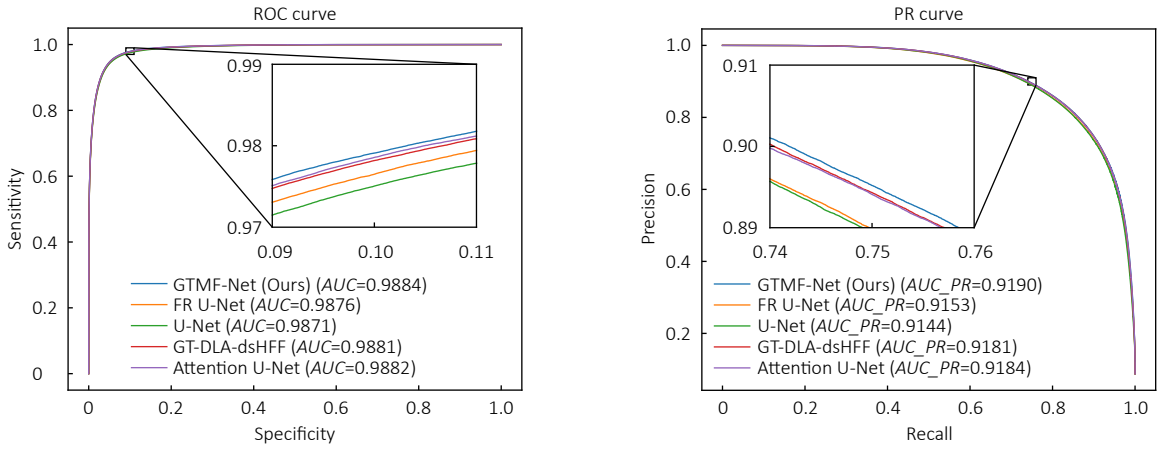


图 9 不同算法在 DRIVE 数据集上 ROC 曲线与 PR 曲线对比

Fig. 9 Comparison of ROC and PR curves of different algorithms on the DRIVE dataset

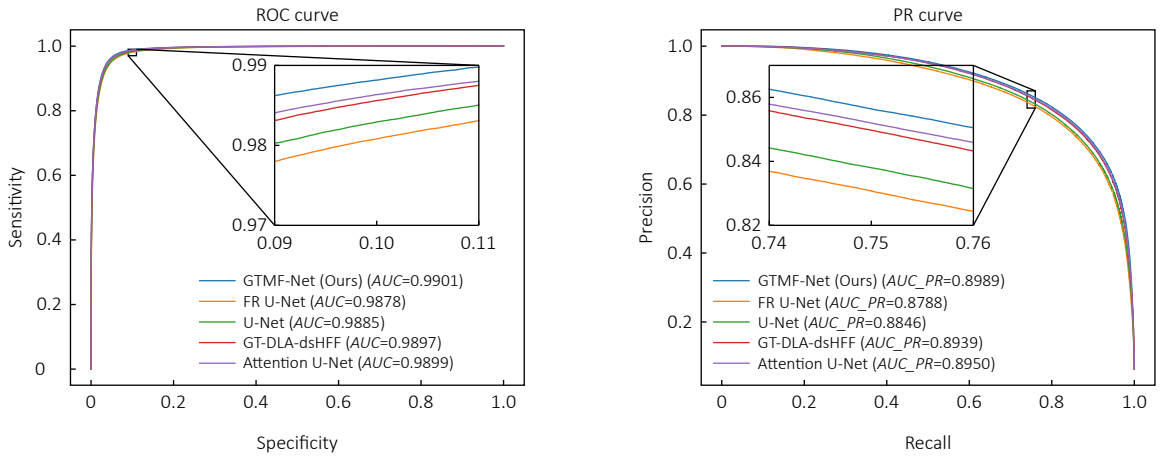


图 10 不同算法在 CHASE_DB1 数据集上 ROC 曲线与 PR 曲线对比

Fig. 10 Comparison of ROC and PR curves of different algorithms on the CHASE_DB1 dataset

由表 2 可知，在 DRIVE 数据集上，DS2TUNet 通过 Transformer 整合局部和全局特征，并结合深度监督机制实现多层级监督，其灵敏度高于 GTMF-Net，但其余指标均低于 GTMF-Net。AttUKAN 利用注意力门机制选择性增强血管区域特征，并结合标签引导的像素级对比损失函数强化前景血管像素的判别能力，其灵敏度达到最优值，但 GTMF-Net 其余指标均比其高。此外，GTMF-Net 的准确率和 AUC 值均优于其余对比算法，表明 GTMF-Net 在辨别血管像素和背景像素方面具有更好的性能。由表 3 可知，在 CHASE_DB1 数据集上，ISAM-UNet 通过深度融合多尺度上下文信息和优化特征提取过程，有效减少血管特征损失，其特异性和 F1 分数高于 GTMF-Net，但 GTMF-Net 的准确率、灵敏度和 AUC 值分别比 ISAM-UNet 提升 0.95%、4.4% 和 0.13%，表明 GTMF-Net 能够增强复杂背景区域中血管的识别能力，精确提取血管细节特征。由表 4 可知，在 STARE 数据集上，RCAR-UNet 提出一种粗糙通道注意力机制，自适应校准特征通道，有效抑制无关区域的特征响应，

表 2　DRIVE 数据集性能指标对比结果

Table 2　Performance comparison results on the DRIVE dataset

Dataset	Method	Acc	Sen	Spe	AUC	F1
DRIVE	ResDO-UNet[24]	0.9561	0.7985	0.9791	—	0.8229
	DS2TUNet[25]	0.9664	0.8071	0.9823	—	0.8195
	ISAM-UNet[26]	0.9565	0.8243	0.9758	0.9807	0.8284
	PGU-Net[27]	0.9572	0.7915	0.9814	0.9812	0.8250
	GCC-UNet[28]	0.9574	0.8032	0.9821	0.9813	0.8278
	RCAR-UNet[29]	0.9537	0.7487	0.9836	—	0.8047
	RSF-UNet[30]	0.9574	0.7995	0.9805	0.9816	0.8270
	AttUKAN[31]	0.9549	**0.8393**	0.9721	0.9795	0.8250
	Ours	**0.9711**	0.8026	**0.9873**	**0.9884**	**0.8298**

表 3　CHASE_DB1 数据集性能指标对比结果

Table 3　Performance comparison results on the CHASE_DB1 dataset

Dataset	Method	Acc	Sen	Spe	AUC	F1
CHASE_DB1	ResDO-UNet[24]	0.9672	0.8020	0.9794	—	0.8236
	DS2TUNet[25]	0.9741	0.8101	0.9869	—	0.8362
	ISAM-UNet[26]	0.9667	0.7779	**0.9900**	0.9888	**0.8372**
	PGU-Net[27]	0.9658	0.8051	0.9817	0.9857	0.8099
	GCC-UNet[28]	0.9659	0.8123	0.9815	0.9850	0.8086
	RCAR-UNet[29]	0.9566	0.7475	0.9798	—	0.7470
	RSF-UNet[30]	0.9672	0.8126	0.9826	0.9874	0.8179
	AttUKAN[31]	0.9631	0.8061	0.9811	0.9821	0.8134
	Ours	**0.9762**	**0.8219**	0.9866	**0.9901**	0.8136

表 4　STARE 数据集性能指标对比结果

Table 4　Performance comparison results on the STARE dataset

Dataset	Method	Acc	Sen	Spe	AUC	F1
STARE	ResDO-UNet[24]	0.9567	0.7963	0.9792	—	0.8172
	ISAM-UNet[26]	0.9660	0.7811	0.9880	0.9864	0.8298
	PGU-Net[27]	0.9654	0.7368	0.9901	0.9862	0.8062
	GCC-UNet[28]	0.9658	0.7806	0.9877	0.9856	0.8282
	RCAR-UNet[29]	0.9594	0.6979	**0.9905**	—	0.7850
	RSF-UNet[30]	0.9719	0.8002	0.9882	0.9899	0.8318
	AttUKAN[31]	0.9643	0.7688	0.9868	0.9796	0.8114
	Ours	**0.9758**	**0.8042**	0.9899	**0.9911**	**0.8350**

其特异性最高，较 GTMF-Net 提升 0.06%，但 GTMF-Net 除特异性外，其余指标均为各项的最优值，表明 GTMF-Net 在血管分割任务上实现了更高的精度，综合性能优于其余先进算法。综合表 2~表 4 可知，GTMF-Net 泛化能力强，分割性能优越。

3.5 模型复杂度

表 5 展示了 GTMF-Net 与其他几种模型在参数量和浮点计算量方面的对比。结果表明，与 U-Net、Attention U-Net 和 GT-DLA-dsHFF 相比，GTMF-Net 在保持相近浮点计算量的同时，其参数量仅为 11.360 M，显著低于上述模型。尽管 FR U-Net 的参数量略低于 GTMF-Net，但其性能指标和分割效果均不及 GTMF-Net。综上所述，GTMF-Net 在模型复杂度与分割性能之间取得较好平衡，是一种具备较低计算开销与较高精度的分割模型。

3.6 消融实验

为证明本文算法各模块的有效性，在 DRIVE、CHASE_DB1 和 STARE 数据集上进行消融实验，其结果如表 6 所示。其中，每个数据集的首行对应原始 U-Net 模型的实验结果。

观察表 6 可知，当 DCA 单独加入或与其他模块组合加入时，三个数据集中准确率和 F1 分数均有一定的

表 5 模型参数量与浮点计算量对比

Table 5 Comparison of model parameters and floating-point computations

Method	Params/M	FLOGs/G
U-Net[8]	34.526	4.091
Attention U-Net[21]	34.877	4.160
FR U-Net[22]	**7.375**	**3.682**
GT-DLA-dsHFF[23]	29.819	7.407
Ours	11.360	3.901

表 6 消融实验结果

Table 6 Ablation experiment results

Dataset	DCA	GCT	MFM	Acc	Sen	Spe	AUC	F1
DRIVE				0.9704	0.7954	0.9872	0.9871	0.8250
	√			0.9709	0.8004	0.9873	0.9882	0.8284
		√		0.9708	0.7976	0.9874	0.9880	0.8272
			√	0.9709	0.7943	0.9878	0.9881	0.8272
	√	√		0.9710	0.8018	0.9872	0.9881	0.8289
	√		√	**0.9711**	0.7963	**0.9879**	0.9883	0.8285
		√	√	0.9710	0.7986	0.9876	0.9883	0.8287
	√	√	√	**0.9711**	**0.8026**	0.9873	**0.9884**	**0.8298**
CHASE_DB1				0.9747	0.8085	0.9859	0.9885	0.8015
	√			0.9754	0.8046	0.9869	0.9875	0.8050
		√		0.9759	**0.8230**	0.9862	0.9900	0.8118
			√	0.9758	0.8073	0.9871	0.9896	0.8078
	√	√		0.9758	0.8186	0.9864	0.9898	0.8104
	√		√	0.9760	0.8094	**0.9872**	0.9900	0.8096
		√	√	0.9761	0.8151	0.9869	0.9899	0.8116
	√	√	√	**0.9762**	0.8219	0.9866	**0.9901**	**0.8136**
STARE				0.9753	0.7924	0.9903	0.9903	0.8305
	√			0.9755	0.7906	0.9907	0.9907	0.8310
		√		0.9757	0.8022	0.9899	0.9909	0.8344
			√	0.9756	0.7939	0.9905	0.9906	0.8323
	√	√		0.9755	0.7965	0.9902	0.9906	0.8324
	√		√	**0.9759**	0.7945	**0.9908**	0.9910	0.8339
		√	√	0.9756	0.8005	0.9900	0.9907	0.8335
	√	√	√	0.9758	**0.8042**	0.9899	**0.9911**	**0.8350**

提升，表明 DCA 能够引导网络聚焦于血管相关区域，有效提升分割精度。当 GCT 单独加入时，CHASE_DB1 数据集中灵敏度达到最优；当 GCT 与 DCA 组合加入时，DRIVE 和 STARE 数据集中灵敏度较仅加入 DCA 时进一步提升，表明 GCT 能够有效增强模型对微细血管的捕获能力。当 MFM 单独加入时，三个数据集中准确率、特异性和 AUC 值均有所提升；当 MFM 与 DCA 组合加入时，三个数据集中特异性均达到最优，且在 DRIVE 和 STARE 数据集中准确率达到最高，表明 MFM 能够抑制背景噪声干扰，提高模型对血管信息的捕获能力。完整的 GTMF-Net 模型在三个数据集中特异性均未达到最高，其原因在于，GCT 利用了分组通道自注意力机制，该机制通过在通道维度上对权重进行重分配，强化局部血管特征响应。虽其有助于提升模型对微细血管特征的提取能力，但当特征分布中存在噪声或背景纹理与血管特征较为相似时，模型可能会错误地放大这些无关区域的响应，将一些背景区域识别为血管，进而影响特异性的表现。但完整的 GTMF-Net 模型除特异性外其余指标均表现较好，整体性能最优。综上所述，GTMF-Net 有效且合理。

4 结　论

针对视网膜血管分割存在复杂背景干扰、微细血管分割不足和血管结构断裂等问题，本文提出分组通道 Transformer 与调制融合的视网膜血管分割算法。首先，在算法的编码端设计双维协同注意力模块，建立通道间和空间位置上的依赖关系，有效保留血管形态结构信息；其次，应用分组通道 Transformer 模块构建双编码结构，建立血管像素间的全局语义关联，促进微细血管局部信息的充分捕获；最后，在算法的解码端引入调制融合模块，保证血管结构的连续性，降低无关背景区域噪声的影响。实验结果表明，GTMF-Net 在 DRIVE、CHASE_DB1 和 STARE 数据集上展现出良好的血管分割能力，但算法在低对比度区域的特征提取能力有限，导致其仍存在误分割现象，后续研究将优化算法设计，提升病变区域的血管识别精度。

作者贡献

梁礼明：指导论文选题、论文修改；王灵亮：设计并完成实验，论文写作；蔡锦辉：研究资料的收集；王成斌：协助完成论文修改。

利益冲突

所有作者声明无利益冲突

参考文献

1. Liang L M, Zhong Y, Chen K Q et al. Fusion dual-attention retinal disease grading algorithm with PVTv2 and DenseNet121[J]. *Opto-Electron Eng*, 2025, **52**(4): 240273.
梁礼明, 钟奕, 陈康泉, 等. 融合 PVTv2 和 DenseNet121 的双注意力视网膜病变分级算法[J]. 光电工程, 2025, **52**(4): 240273.
2. Xiao Z J, Tian H, Zhang J H et al. Fusion of dynamic features enhances remote sensing building segmentation[J]. *Opto-Electron Eng*, 2025, **52**(3): 240231.
肖振久, 田昊, 张杰浩, 等. 融合动态特征增强的遥感建筑物分割[J]. 光电工程, 2025, **52**(3): 240231.
3. Li Q, You J, Zhang D. Vessel segmentation and width estimation in retinal images using multiscale production of matched filter responses[J]. *Expert Syst Appl*, 2012, **39**(9): 7600−7610.
4. Ooi A Z H, Embong Z, Abd Hamid A I et al. Interactive blood vessel segmentation from retinal fundus image based on canny edge detector[J]. *Sensors (Basel)*, 2021, **21**(19): 6380.
5. Cheng Y Z, Hu X, Wang J et al. Accurate vessel segmentation with constrained B-snake[J]. *IEEE Trans Image Process*, 2015, **24**(8): 2440−2455.
6. Ghosh S K, Ghosh A. A novel retinal image segmentation using rSVM boosted convolutional neural network for exudates detection[J]. *Biomed Signal Process Control*, 2021, **68**: 102785.
7. Wang S L, Yin Y L, Cao G B et al. Hierarchical retinal blood vessel segmentation based on feature and ensemble learning[J]. *Neurocomputing*, 2015, **149**: 708−717.
8. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation[C]//*Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention*, Munich, 2015: 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.
9. Getahun M N, Rogov O Y, Dylov D V et al. FS-Net: full scale network and adaptive threshold for improving extraction of micro-retinal vessel structures[J]. *Pattern Recognit Lett*, 2025, **189**: 188−194.
10. Radha K, Karuna Y. Gabor-modulated depth separable convolution for retinal vessel segmentation in fundus images[J]. *Comput Biol Med*, 2025, **188**: 109789.
11. Vaswani A, Shazeer N, Parmar N et al. Attention is all you need[C]//*Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, 2017: 6000–6010.
12. Dosovitskiy A, Beyer L, Kolesnikov A et al. An image is worth 16x16 words: transformers for image recognition at scale[C]//*Proceedings of the 9th International Conference on Learning Representations*, Astria, 2021.
13. Wu C W, Guo M, Ma M et al. TLTNet: a novel transscale cascade layered transformer network for enhanced retinal blood vessel segmentation[J]. *Comput Biol Med*, 2024, **178**: 108773.
14. Cui Y, Su J J, Tian Y et al. TA-Net: transformer based joint attention retinal vessel segmentation network[J]. *Signal Image Video Process*, 2025, **19**(6): 475.
15. Yu K R, Zhang T Q, Wang H W et al. FSTA-SNN: frequency-based spatial-temporal attention module for spiking neural networks[C]//*Proceedings of the 30th AAAI Conference on Artificial Intelligence*,

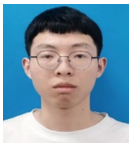
Philadelphia, 2025: 22227–22235. https://doi.org/10.1609/aaai.v39i21.3437.
16. Li J Y, Zhang Z L, Zuo W M. Rethinking transformer-based blind-spot network for self-supervised image denoising[C]//*Proceedings of the 39th AAAI Conference on Artificial Intelligence*, Philadelphia, 2025: 4788–4796. https://doi.org/10.1609/aaai.v39i5.32506.
17. Zhang Y F, Zhou S, Li H F. Depth information assisted collaborative mutual promotion network for single image dehazing[C]// *Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 2024: 2846–2855. https://doi.org/10.1109/CVPR52733.2024.00275.
18. Staal J, Abràmoff M D, Niemeijer M et al. Ridge-based vessel segmentation in color images of the retina[J]. *IEEE Trans Med Imaging*, 2004, **23**(4): 501−509.
19. Owen C G, Rudnicka A R, Mullen R et al. Measuring retinal vessel tortuosity in 10-year-old children: validation of the computer-assisted image analysis of the retina (CAIAR) program[J]. *Invest Ophthalmol Vis Sci*, 2009, **50**(5): 2004−2010.
20. Hoover A D, Kouznetsova V, Goldbaum M. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response[J]. *IEEE Trans Med Imaging*, 2000, **19**(3): 203−210.
21. Oktay O, Schlemper J, Le Folgoc L et al. Attention U-Net: Learning where to look for the pancreas[Z]. arXiv: 1804.03999, 2018. https://doi.org/10.48550/arXiv.1804.03999.
22. Liu W T, Yang H H, Tian T et al. Full-resolution network and dual-threshold iteration for retinal vessel and coronary angiograph segmentation[J]. *IEEE J Biomed Health Inf*, 2022, **26**(9): 4623−4634.
23. Li Y, Zhang Y, Liu J Y et al. Global transformer and dual local attention network via deep-shallow hierarchical feature fusion for retinal vessel segmentation[J]. *IEEE Trans Cybern*, 2023, **53**(9): 5826−5839.
24. Liu Y H, Shen J, Yang L et al. ResDO-UNet: a deep residual network for accurate retinal vessel segmentation from fundus images[J]. *Biomed Signal Process Control*, 2023, **79**: 104087.
25. Liu Q Y, Zhou F, Shen J X et al. A fundus vessel segmentation method based on double skip connections combined with deep supervision[J]. *Front Cell Dev Biol*, 2024, **12**: 1477819.
26. Ma Z D, Li X B. An improved supervised and attention mechanism-based U-Net algorithm for retinal vessel segmentation[J]. *Comput Biol Med*, 2024, **168**: 107770.
27. Wang Z Y, Jia L V, Liang H C. Partial class activation mapping guided graph convolution cascaded U-Net for retinal vessel segmentation[J]. *Comput Biol Med*, 2024, **178**: 108736.
28. Wei X X, Lin X, Liu H Y et al. Retinal vessel segmentation with deep graph and capsule reasoning[Z]. arXiv: 2409.11508, 2024. https://doi.org/10.48550/arXiv.2409.11508.
29. Ding W P, Sun Y, Huang J S et al. RCAR-UNet: retinal vessel segmentation network algorithm via novel rough attention mechanism[J]. *Inform Sci*, 2024, **657**: 120007.
30. Sun Z H, Wang H, Xie Q et al. RSF-Conv: rotation-and-scale Equivariant Fourier parameterized convolution for retinal vessel segmentation[J]. *IEEE Trans Neural Netw Learn Syst*, 2025, **36**(9): 16549−16563.
31. Zeng S, Lee C H, Nnamdi M C et al. Novel extraction of discriminative fine-grained feature to improve retinal vessel segmentation[Z]. arXiv: 2505.03896, 2025. https://doi.org/10.48550/arXiv.2505.03896.

作者简介



梁礼明 (1967-)，男，硕士，教授，硕士生导师，主要研究方向为机器学习、模式识别与图像处理等。

E-mail：lianglm67@163.com



【通信作者】王灵亮 (2001-)，男，硕士研究生，主要研究方向为医学图像处理。

E-mail：2325286035@qq.com

Grouped channel Transformer and modulation fusion retinal vessel segmentation algorithm

Liang Liming, Wang Lingliang*, Cai Jinhui, Wang Chengbin

School of Electrical Engineering and Automation, Jiangxi University of Science and Technology, Ganzhou, Jiangxi 341000, China

Abstract

Objective Changes in retinal vascular morphology serve as crucial indicators for diagnosing various diseases, such as diabetic retinopathy, hypertension, and cataracts. Traditional diagnosis of retinal diseases primarily relies on manual assessment by clinicians, typically involving the analysis of retinal vascular curvature, width, and branching patterns based on imaging results from fundus examinations, optical coherence tomography, and fluorescein angiography. However, these imaging outcomes are susceptible to interference from factors such as illumination and noise, which increase the difficulty for clinicians in distinguishing retinal vessels from the background. Furthermore, inherent ocular structures like the optic cup and optic disc regions, as well as pathological alterations such as hard exudates, soft exudates, and the macular area, can also impede clinicians' judgment. Diagnostic results often vary subjectively among different physicians, potentially leading to misdiagnosis due to insufficient experience or fatigue. Automated retinal vessel segmentation technology can not only significantly reduce the time and effort required for manual annotation by clinicians and improve diagnostic efficiency but also minimize the impact of human factors to the greatest extent, providing more objective results to assist physicians in decision-making. Therefore, developing an accurate and automated retinal vessel segmentation technique holds substantial practical significance.

Methods In the field of retinal vessel segmentation, existing methods can be broadly categorized into unsupervised and supervised approaches. Unsupervised methods do not rely on manually annotated data and primarily achieve segmentation by analyzing the morphological features of vessels. Common techniques include matched filtering-based, edge detection-based, and model-based methods. Matched filtering methods mainly utilize Gaussian kernel functions or their variants, which resemble the cross-sectional shape of blood vessels, to construct two-dimensional linear filter banks. These methods enhance tubular structures in retinal images through multi-scale and multi-directional filter responses, combined with thresholding for preliminary detection of vessel pixels. Edge detection-based methods capture boundary features between vessels and the background by calculating gradient changes in pixel intensity or using differential operators such as Canny and Sobel, often integrated with local or global thresholding strategies to extract vascular contours with continuous topological structures. Model-based methods rely on geometric or statistical models, such as active contour models, level set methods, or vascular tubular models, to mathematically describe vessel structure and distribution. They iteratively optimize energy functionals or probabilistic fitting processes to gradually evolve fine segmentation boundaries that adaptively conform to vessel morphology. However, such methods often depend on manually designed feature extraction pipelines, making it difficult to fully capture the complex topological structures of vessels, thereby resulting in limited segmentation performance. In contrast, supervised methods require expert-annotated data as training samples to achieve higher accuracy in vessel segmentation tasks. These methods can be further divided into traditional machine learning methods and deep learning methods. Traditional machine learning methods typically involve two steps: manual feature extraction and classification. Handcrafted features mainly include edge responses, wavelet responses, principal component information, and color intensity information. Based on the choice of classifier, they can be roughly categorized into k-nearest neighbors, random forests, support vector machines, and Gaussian mixture models. Traditional machine learning methods extract segmentation targets based on vascular features and fundus image characteristics. Compared to manual screening, they offer advantages such as shorter processing time and faster segmentation. However, their feature extraction process still requires manual design and cannot adequately represent the multi-scale structural features of vessels. With the continuous advancement and development of deep learning, it has gradually been adopted across various tasks in computer vision. In the field of medical image analysis, segmentation algorithms have been continuously updated and iterated. The introduction of various deep learning models and modules has further improved the accuracy and efficiency of fundus vessel segmentation algorithms. Compared to traditional machine learning methods, deep learning methods can automatically learn optimal features and achieve predictive modeling without human intervention, offering higher segmentation efficiency. Nevertheless, challenges such as limited representation of fine vessels, interference from complex backgrounds, and discontinuities in vessel structures remain. To address the aforementioned issues, this paper proposes a retinal vessel segmentation algorithm based on Grouped Channel Transformer and Modulated Fusion. First, a dual-dimensional cooperative attention module is designed in the encoder to enhance the distinguishability of

Foundation item: National Natural Science Foundation of China (51365017, 61463018), the Natural Science Foundation of Jiangxi Province (20192BAB205084), and the Youth Project of Science and Technology Research of Jiangxi Provincial Department of Education (GJJ2200848)

*Correspondence: Wang L L, E-mail: 2325286035@qq.com

vessel features and improve the model's accuracy in segmenting microvessels. Second, a Grouped Channel Transformer module is applied to construct a dual-encoder architecture, establishing global dependencies to effectively capture detailed vessel information. Finally, a Modulated Fusion module is introduced at the decoder to enhance feature fusion between the encoder and decoder, thereby improving the model's robustness against noise interference.

Results and Discussions Experiments conducted on the DRIVE, CHASE_DB1, and STARE datasets demonstrate that the proposed algorithm achieves accuracy rates of 0.9711, 0.9762, and 0.9758, sensitivities of 0.8026, 0.8219, and 0.8042, and specificities of 0.9873, 0.9866, and 0.9899, respectively. The proposed method outperforms other algorithms in overall segmentation performance, demonstrating superior segmentation capability.

Conclusions GTMF-Net demonstrates robust vessel segmentation capabilities on the DRIVE, CHASE_DB1, and STARE datasets. It achieves a favorable balance between model complexity and segmentation performance, offering both relatively low computational cost and high segmentation accuracy. However, the algorithm currently exhibits limitations in feature extraction within low-contrast regions, which can lead to mis-segmentation. Future work will focus on optimizing the algorithm's design to improve the accuracy of vessel identification in pathological areas.

Keywords: retinal vessel segmentation; grouped channel Transformer; dual-dimension collaborative attention; modulation fusion

Liang L M, Wang L L, Cai J H et al. Grouped channel Transformer and modulation fusion retinal vessel segmentation algorithm[J]. *Opto-Electron Eng*, 2026, **53**(3): 250264; DOI: 10.12086/oee.2026.250264

Opto-Electronic Journals Group
www.oejournal.org

扫描二维码，获取 PDF 全文