

文章编号 1004-924X(2023)23-3482-08

融合注意力机制的 TransGLnet 脉络膜自动分割

黄文博*, 屈超凡, 燕 杨

(长春师范大学 计算机科学与技术学院, 吉林 长春 130032)

摘要:针对脉络膜与巩膜间对比度低, 脉络膜分割存在脉络膜下边界模糊, 难以界定等问题, 提出了融合注意力机制的 TransGLnet 脉络膜自动分割网络, 在卷积层引入全局注意力模块(Global Attention Module, GAM), 在特征之间应用矩阵乘法, 在整体空间位置的多个特征之间建立非线性交互, 在不使用大量参数的情况下提取全局特征; 在卷积层和 Transformer 编码器之间引入局部注意力模块(Local Attention Module, LAM), 以 1/4 特征图为基本单元探索局部特征, 特征图元素位置移动规则为保持行位置的元素不变, 将列位置的元素由大到小重新排列。两模块融合可令网络有效兼顾全局与局部特征。实验结果表明, TransGLnet 网络的 Dice 值为 0.91, 准确率为 0.98, 平均交并比为 0.89, F1 值为 0.90, 豪斯多夫距离为 6.56。与现有脉络膜自动分割方法相比, 本文方法的各项性能指标均有提高。TransGLnet 脉络膜自动分割网络具有较好的稳定性, 可供临床借鉴。

关键词:医学图像处理; 脉络膜自动分割; TransUnet; 全局注意力; 局部注意力

中图分类号: TP391.7 **文献标识码:** A **doi:** 10.37188/OPE.20233123.3482

Automatic segmentation of choroid by TransGLnet integrating attention mechanism

HUANG Wenbo*, QU Chaofan, YAN Yang

(School of Computer Science and Technology, Changchun Normal University,
Changchun 130032, China)

* Corresponding author, E-mail: huangwenbo@sina.com

Abstract: Addressing the challenge posed by the low contrast between the choroid and sclera in choroid segmentation, this research introduces the TransGLnet choroid automatic segmentation network, employing an attention mechanism. The incorporation of a Global Attention Module (GAM) within the convolutional layer involves matrix multiplication between features, establishing nonlinear interactions across multiple features in the global spatial context. This enables the extraction of global features without an excessive number of parameters. To explore local features, a Local Attention Module (LAM) is introduced between the convolution layer and Transformer encoder, focusing on a 1/4 feature graph. The movement rule for feature graph elements maintains row position consistency while rearranging elements in the column position from largest to smallest. The integration of these two modules ensures that the network effectively considers both global and local features. Experimental results showcase the efficacy of the proposed TransGLnet network with a Dice value of 0.91, accuracy at 0.98, equal crossover ratio of 0.89, F1 value

收稿日期: 2023-06-15; 修订日期: 2023-07-24.

基金项目: 吉林省科技厅科技发展计划与自然科学基金联合基金项目(No. YDZJ202101ZYTS147)

reaching 0.90, and a Hausdorff distance of 6.56. Comparative analysis against existing automatic choroidal segmentation methods reveals notable improvements in performance metrics. The network presented in this study demonstrates robustness and stability, rendering it suitable for clinical reference.

Key words: medical image processing; automatic choroidal segmentation; TransUnet; global attention; local attention

1 引言

脉络膜层是位于眼底视网膜下方的血管层,其主要功能是为视网膜提供氧气和营养^[1]。临床研究表明,观察脉络膜厚度的变化对于预测相关眼部疾病至关重要。如成人年龄及近视度数的增加会导致脉络膜厚度变薄,糖尿病视网膜病变会导致脉络膜厚度变厚^[2]。因此,脉络膜分割及厚度分析在眼部疾病的早期预测及诊断中具有重要意义。

目前,眼科医生想要获得准确的脉络膜边界,需在眼底 OCT 图像中手动测量其厚度,耗时费力。眼底 OCT 图像脉络膜区域中,脉络膜与巩膜之间对比度较低,脉络膜自动分割存在脉络膜下边界模糊,难以界定等困难^[3]。针对这些问题,国内外学者对脉络膜分割展开了较为深入的研究。Yan 等^[4]提出了上下文高效自适应网络(Context Efficient Adaptive Network,CEA-Net),用于 OCT 图像中脉络膜的自动分割,以 CE-Net 为主干模型,整合高效通道注意力(Efficient Channel Attention,ECA)模块以实现精确的分割性能。许祥丛等^[5]提出了压缩-激励-拼接(Squeeze-Excitation-Connection,SEC)模块,并与 U-Net 相结合称为 SEC-Unet。SEC 模块作为注意力机制通过自适应学习通道之间相互关系提高网络的表达能力和性能,使网络更适应脉络膜分割任务,但注意力的计算仅在局部范围内进行,对全局上下文没有充分关注。Zhu 等^[6]提出脉络膜 U 型网络,捕捉脉络膜层和脉络膜血管之间的相关特征和特定特征,生成像素级分类,实现脉络膜层的自动分割。Khaing 等^[7]提出以 U-Net 作为主干模型,集成具有不同膨胀率的膨胀卷积,实现脉络膜分割。膨胀卷积的空洞间隔不合适,可能导致感受野过大或过小,因此没能捕获更多的上下文信息来进行预测。Xu 等^[8]提出结合图像增强和基于注意力的密集(Attention-

based Dense,AD)U-Net 模型,在 U-Net 基础上加入注意力门机制,使网络能够加强对图像感兴趣的区域关注,从而对脉络膜进行分割。WU 等^[9]提出边界增强编码器-解码器架构,边界增强模块构成每个编码器-解码器层的主干,特征提取分别使用不同扩张率的扩张卷积来获取不同感受野下的相关图像特征。这意味着不同位置上计算的特征主要关注局部区域,而不是图像的全局信息。通道增强分支的重点仍是局部区域特征增强,缺乏全局信息获取。

综上所述,现有方法实现脉络膜分割多数只关注局部信息,虽能对脉络膜细节进行学习,但欠缺对图像上下文和全局结构信息的学习,从而没能有效兼顾脉络膜 OCT 图像的全局信息及局部细节,导致脉络膜在 OCT 图像分割任务中存在定位和分割困难及分割结果边界模糊等问题,脉络膜自动精准分割仍面临挑战。

本文提出一种基于 TransGLnet 网络的脉络膜自动分割方法,用于在 OCT 图像中自动分割脉络膜层。为了解决网络在训练过程中产生全局信息丢失及局部细节不准确等问题,引入专注于全局信息和局部细节的全局注意力模块^[10](Global Attention Module,GAM)和局部注意力模块(Local Attention Module,LAM^[11]),帮助网络捕捉图像的整体上下文和语义信息,同时网络聚焦于图像的细节和局部特征,跨越语义鸿沟,使得网络能够综合考虑到图像的全局结构和局部细节,从而更全面地理解图像内容和语义信息,显著提升模型性能。

2 网络设计

2.1 网络搭建

本文搭建的 TransGLnet 网络由编码器和解码器两部分组成,如图 1 所示。其主干网络为 TransUnet^[12]。编码器由卷积层、特征层及

Transformer 编码器组成。在特征提取过程中,卷积层有获取局部信息的能力,但对获取全局上下文的能力很弱,Transformer 对来自卷积神经网络特征映射的标记化图像块进行编码,作为提取全局上下文的输入序列,主要衡量序列中所有元素之间的相关性,但忽略了局部细节信息。为了增强网络对全局与局部信息获取的能力,探索特征图中元素之间的空间位置关系,本文引入

GAM 和 LAM 来探索不同范围内元素之间的相关性以及特征图的局部细节。为了提高图像分割的精度和感兴趣区域的重建效果,这里采用与编码器对称的解码器来恢复图像中的感兴趣区域。由于下采样在特征提取的过程中不可避免地会丢失特征图中的一些前景信息,为了解决这个问题,通过跳跃连接将编码器的多尺度特征信息和上采样提取的上下文信息进行特征融合。

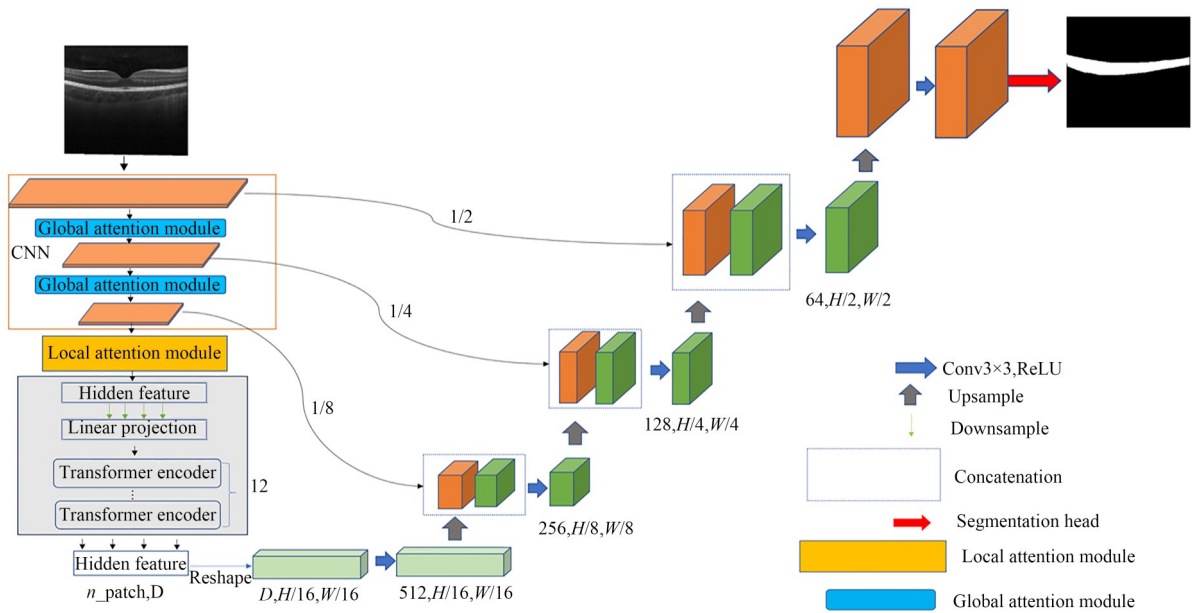


图 1 TransGLnet 整体框架

Fig. 1 Overall framework of TransGLNet

2.2 全局注意力模块

全局注意力模块 GAM 嵌入到卷积层之中,如图 2 所示。卷积层与 ResNet50^[13] 网络架构基本相同。本文输入的原始特征图尺寸为 $3 \times 224 \times 224$ 。经过一系列卷积运算后,最终输出的特征图尺寸为 $1\ 024 \times 14 \times 14$ 。在计算空间有限

的情况下,使用 Transformer 计算特征图中元素之间的相关性往往需要大量的计算资源。由于较大的输入批量往往会因计算资源不足而导致实验失败,因此,采用组归一化来克服批归一化在小批量下实验效果不佳的缺点,使用自注意力机制来实现 GAM。为了提取全局特征,GAM 从

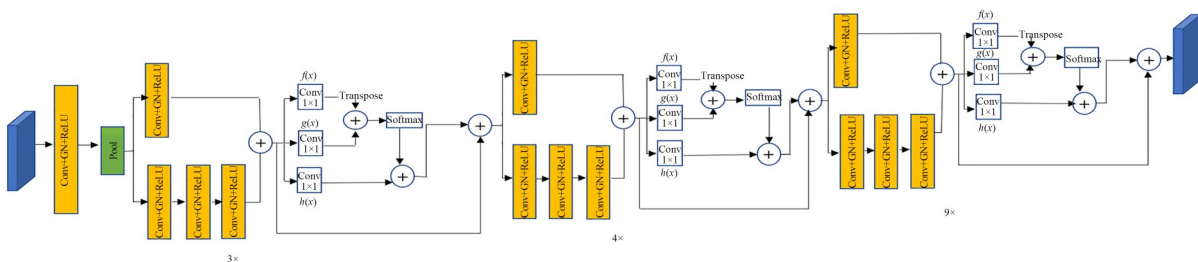


图 2 全局注意力模块结构

Fig. 2 Structure of global attention module

输入特征的线性卷积开始,每个卷积都具有 1×1 内核。这种线性变换的结果假设成 $f(x)$, $g(x)$ 和 $h(x)$ 。GAM 的核心思想是特征之间的矩阵乘法,如下:

$$x + h(x)S[f(x)^T g(x)]^T \gamma, \quad (1)$$

其中 S 代表 Sigmoid 函数; γ 是一个超参数, 本文设置为 1。

将注意力输出结果乘以标量 γ , 以调节 GAM 对输出的贡献。如果 γ 为零, 则整个全局信息模块的贡献为零, 输入 x 直接传递到下一层。使用矩阵乘法能够创建多个不同特征之间的非线性交互。例如, $f(x)$ 中的每个特征都与 $g(x)$ 中的所有特征相乘, 从而帮助捕获全局特征, 且无需额外的参数。在将 Softmax 函数应用于矩阵乘法输出之后, 每个特征图都成为一个注意力掩码, 可以使网络集中注意力于 $h(x)$ 中的特定区域, 忽略其他区域。将 Softmax 函数输出与 $h(x)$ 相乘可以过滤掉注意力为零的区域, 从而确保只有关注的区域传递到下一层。

2.3 局部注意力模块

直接使用 Transformer 探索全局元素的关系, 网络很难关注特征图的细节信息。为了进一步探索特征图中元素之间的空间位置关系, 这里使用 LAM 来探索局部范围内元素之间的位置关系。LAM 流程如图 3 所示。输入到网络的特征图是 $R^{C \times H \times W}$, 其中 C 为通道数, H 和 W 分别为特征图的高度和宽度。为了方便介绍此模块, 这里 $C=1, H=W=4$ 。

特征融合过程如图 4 所示, 输入到网络的原始特征图为 $F_1 \in R^{1 \times 4 \times 4}$, 并且以 $1/4$ 的特征图为基础

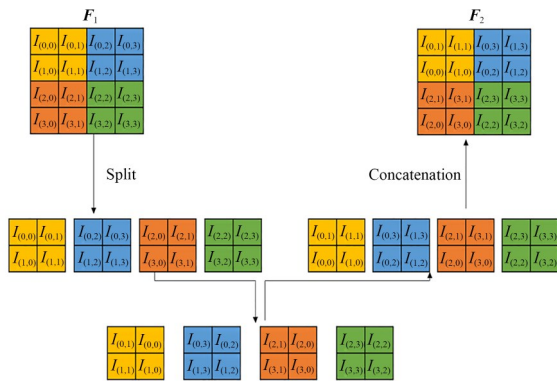


图 3 局部注意力模块流程

Fig. 3 Process of local attention module

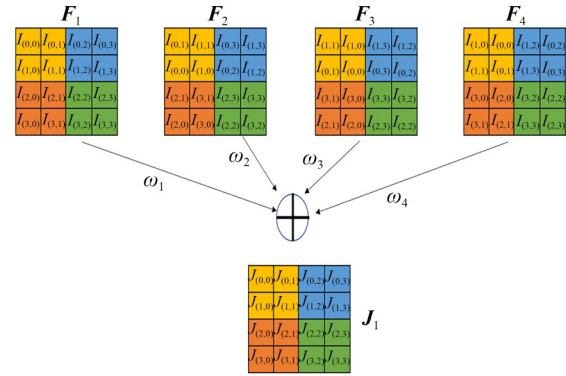


图 4 特征融合过程示意图

Fig. 4 Schematic of feature fusion process

本单位探索元素间的空间位置关系。对输入的原始特征图 F_1 进行切割, 得到 4 个特征块, 每个特征块的尺寸为 $H/4 \times W/4$ 。接着, 针对每个特征块进行所有元素的每行位置不变、每列位置从大到小重新排列, 得到新的特征块。然后, 对新特征块中元素进行转置操作, 得到改变后的新特征块。将完成元素转置的特征块拼接, 得到特征图 F_2 。类似地, 特征图 F_3 和 F_4 是在特征图 F_2 和 F_3 的基础上, 按照相同的特征图切割、元素空间位置偏移、转置以及特征块组合得到的。最后, 通过 4 个尺寸为 1×1 的卷积分别作用于特征图 F_1, F_2, F_3 和 F_4 , 将它们相加得到特征图 J_1 。 J_1 中每个元素都是通过特征图 F_1 中的局部区域元素交互得到的。LAM 以 $1/4$ 特征图为基础单元, 感受野较小, 因此更有利于网络提取图像的局部细节信息。

2.4 损失函数

医学图像中待分割的前景信息通常占据整幅图像的较小比例, 并且不同类别的前景区域的尺寸和形状也不同, 这使得在学习和训练过程中, 网络容易出现过拟合现象。为了应对这种情况, 这里采用 Dice loss^[14] 作为损失函数的一部分, 从而改善网络的泛化能力和抗过拟合能力。Dice loss 基于 Dice 系数, 用来度量分割结果与真实标签之间的相似度。使用 Dice Loss 可以使网络更好地学习图像的边界和细节信息, 从而提高分割精度, 其公式如下:

$$L_{\text{Dice}} = 1 - 2 \times \frac{\sum_i P_i G_i}{\sum_i P_i + \sum_i G_i}, \quad (2)$$

其中: i 属于元素的具体索引值, P_i 为预测图中第

i 个元素值属于某一类别前景的概率, G_i 是标签图中第 i 个元素的真实值。除此之外, 本文使用交叉熵损失 (Cross Entropy Loss, CEL) 作为损失函数的一部分, 即:

$$L_{\text{CEL}} = -\frac{1}{H \times W} \log(P_{i(\text{correct})}), \quad (3)$$

其中: $P_{i(\text{correct})}$ 表示某点的元素值被正确预测的概率, $H \times W$ 表示最终预测图的尺寸。

最后, 损失函数由 CEL 损失函数和 Dice 损失函数组成, 具体公式如下:

$$L_{\text{Loss}} = \lambda_1 \times L_{\text{Dice}} + \lambda_2 \times L_{\text{CEL}}. \quad (4)$$

所有实验均默认在 λ_1 和 λ_2 为 0.5 的情况下进行。

3 实验及结果分析

实验在 Pytorch 深度学习框架下搭建, 实验环境为 python 3.9, pytorch 2.0.0, cuda 12.0。将数据集中 80% 的图像划分为训练集, 20% 的图像划分为测试集。采用指数衰减调整学习率 ExponentialLR, 其中 Gamma 为 0.9, 优化器为 Adam, 训练 epoch 为 50, batch_size 为 2。

3.1 数据集

本文采用的数据集是广州妇女儿童医学中心的视网膜 OCT 图像^[15]。从 OCT 数据集中随机选取 300 幅图像。脉络膜层分割黄金标准由两位眼科医师使用 Image J 软件在 OCT 图像上进行手工标注。在标记脉络膜下边界时, 眼科医师最初通过使用分段线工具选择稀疏点来标记, 然后通过样条线拟合操作进行平滑处理。最后进行微调, 以确保绘制的曲线边界清晰, 具体过程如图 5 所示。

当出现边界标注不一致时, 两位医师会讨论标注边界的异议, 共同解决标注不确定性, 达成一致意见。若无法达成一致, 邀请第三位眼科医

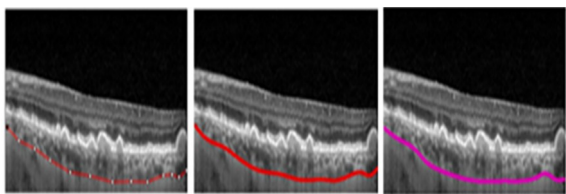


图 5 脉络膜标记过程

Fig. 5 Choroidal labeling process

师评估标注结果, 作为仲裁者解决异议, 最终达成一致的脉络膜边界标注结果。

3.2 评价指标

本文使用 Dice 系数、准确率 (Accuracy)、均交并比 (Miou)、F1 值以及豪斯多夫距离 (Hausdorff Distance, HD) 5 个指标来评估性能。其计算公式如下:

$$\text{Dice} = \frac{2 \times TP}{2 \times TP + FP + FN}, \quad (5)$$

$$\text{Acc} = \frac{TP + TN}{TP + TN + FP + FN}, \quad (6)$$

$$\text{MIoU} = \frac{1}{k+1} \sum_{i=0}^k \frac{TP}{TP + FP + FN}, \quad (7)$$

其中: TP 为预测正确的正样本像素数量, FP 为预测错误的正样本像素数量, TN 为预测正确的负样本像素数量, FN 为预测错误的负样本像素数量, $k+1$ 为包括背景在内的类的数量。

$$F_1 = 2 \times \frac{PR}{P + R}, \quad (8)$$

$$\text{其中: } P = \frac{TP}{TP + FP}, R = \frac{TP}{TP + FN}.$$

$$\text{HD} = \max \left\{ \max_{x \in X} \min_{y \in Y} d(x, y), \max_{y \in Y} \min_{x \in X} d(x, y) \right\}, \quad (9)$$

其中: X 和 Y 分别表示预测图和标签图, x 和 y 分别表示预测图和标签图中的像素; $\{\min_{y \in Y} d(x, y)\}$ 表示预测图中像素到标签图中像素的最短距离集合, $\{\max_{y \in Y} \min_{x \in X} d(x, y)\}$ 表示该最短距离集合中的最大距离; $\{\min_{y \in Y} d(x, y)\}$ 表示标签图中像素到预测图中像素的最短距离集合, $\{\max_{y \in Y} \min_{x \in X} d(x, y)\}$ 表示该最短距离集合中的最大距离。

3.3 实验结果

由表 1 可知, 本文方法的 Dice 系数为 0.91, 准确率达到 0.98, MIoU 达到 0.89, F1 值达到 0.9, 豪斯多夫距离为 6.56。实验结果表明, 使用 TransGLnet 网络训练取得了较好的分割结果, 相对于 TransUnet, 本文网络的 Dice 提高了 10%。相对于其他脉络膜自动分割方法, 本文方法的各项性能指标均有提高, 证明本文训练策略具有很高的稳定性。

图 6 展现了脉络膜 OCT 图像训练和测试结果的可视化。其中, 第一张图片是原始图片, 第

表 1 不同网络脉络膜分割结果比较

Tab.1 Comparison of choroid segmentation results among different networks

Net	Dice	Acc	MIOU	F1_score	HD
U-Net	0.87	0.97	0.87	0.87	8.34
AttU-Net	0.86	0.97	0.86	0.86	8.43
R2U-Net	0.88	0.97	0.88	0.88	7.25
R2AU-Net	0.87	0.97	0.88	0.87	7.79
TransUnet	0.81	0.96	0.83	0.90	10.32
Swin-Unet	0.86	0.97	0.86	0.86	8.41
Ours	0.91	0.98	0.89	0.91	6.56

二张图片是标签图片,接下来的几张图片是不同分割方法的可视化结果。将本文方法与现有的脉络膜自动分割方法进行比较,包括 U-Net^[16], AttU-Net^[17], R2U-Net^[18], R2AU-Net^[19], TransU-

net^[20], Swin-Unet^[21]。

U-Net 是一个由跳跃连接将编码器-解码器连接的 U 形结构,该网络广泛应用于医学图像分割任务。AttU-Net 以 U-Net 为主干,在跳跃连接处加入注意力门机制,抑制输入图像中的不相关区域,突出特定局部区域的显著特征。R2U-Net 以 U-Net 为主干,结合残差连接和循环连接,增强 U-Net 在处理长距离依赖和特征传递方面的能力。R2AU-Net 结合递归残差卷积层,更好地表示分割任务的特征和注意力区域,以突出显示通过跳跃连接传递的显著特征。TransUnet 将 Transformer 和 U-Net 结合利用 Transformer 的自注意力机制来捕捉图像中的全局上下文信息。Swin-Unet 基于纯 Transformer 的 U 形编码器-解码器网络用于医学图像分割,将 SwinTransformer 块作为特征表示和远程语义信息交互学习的基本单元。

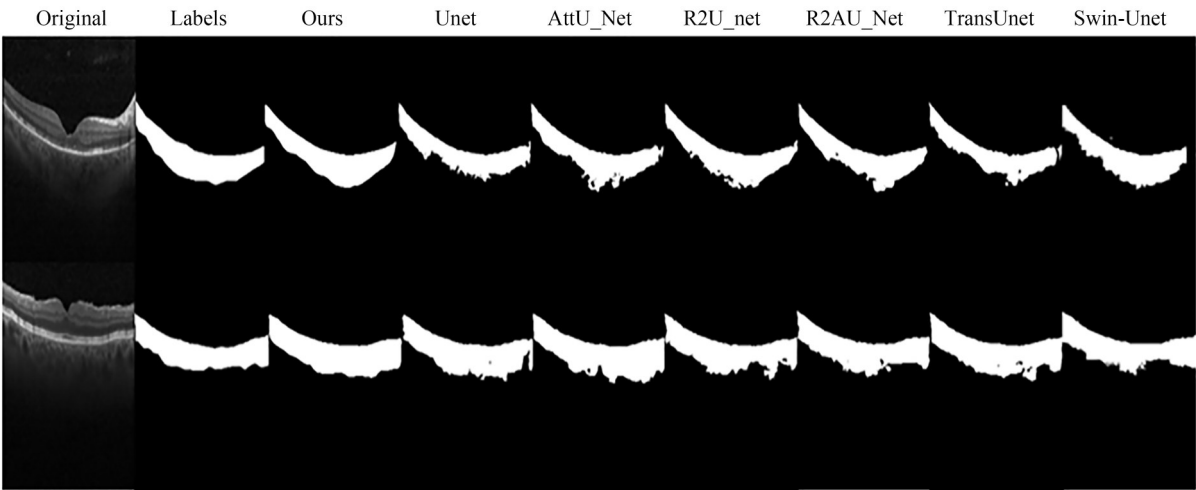


图 6 脉络膜分割实例结果的可视化

Fig. 6 Visualization of choroid segmentation example results

分割结果显示,由于脉络膜下边界模糊,在其他网络处理图像中明显可见脉络膜下边界参差不齐且不连续,相比之下,本文方法分割脉络膜得到了相对平滑和完整的层边界。在特征提取中,卷积层获取局部信息的能力很强,但对全局的建模能力较弱。Transformer 主要衡量序列中所有元素间的相关性,却忽略了局部细节信息的重要性。本文提出的方法得益于其对全局的建模能力的增强和增加对局部细节信息的重视,

因此相比于基准网络和其他网络,本文网络的学习能力得到增强。

为验证本文模型的有效性,还针对美国国立卫生研究院国家医学图书馆提供的 500 张肺部 X 射线图片^[22]进行分割,分割结果如图 7 所示。实验结果表明,本文提出的 TransGLnet 模型在该数据集上也体现出优秀的性能, Dice 系数达到 0.94, HD 为 8.92。通过观察分割结果发现, TransGLnet 模型对于全局建模能力以及局部细

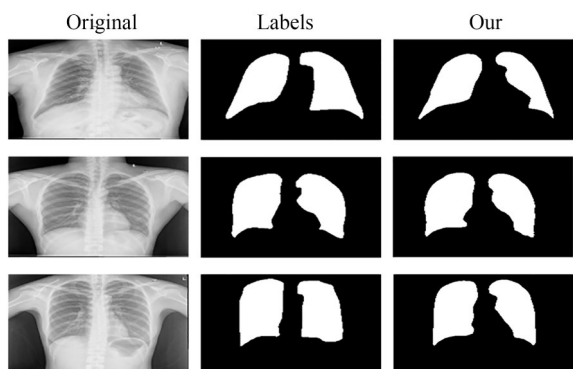


图 7 肺部分割实例结果的可视化

Fig. 7 Visualization of results of lung segmentation example

节信息的重视,使其在肺部分割任务上表现出较好的效果,进一步证明了 TransGLnet 模型具有较强的泛化能力,能够在不同数据集和任务中取得良好的性能表现。

4 消融实验

为验证添加模块的有效性,本文进行了消融实验,消融模型 1(Ablation 1)代表只包含 GAM,消融模型 2(Ablation 2)代表只包含 LAM,使用相同的训练和测试集对消融模型进行训练和验证。

图 8 为消融实验结果。实验结果表明,分别添加两个模块到主干网络效果都有提升,Dice 分别提高了 4% 和 6%,同时添加两模块,Dice 提高

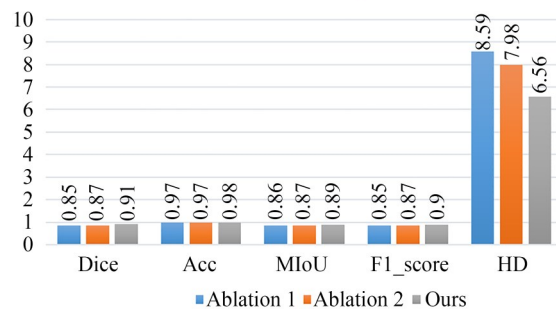


图 8 消融实验结果

Fig. 8 Ablation experiment results

了 10%,模型的整体性能显著提升。

5 结 论

脉络膜分析对于确定眼病的发病机制和治疗策略至关重要,使用计算机快速准确地分割脉络膜可以辅助医生预测相关眼部疾病。本文以 TransUnet 为主干网络,融入 GAM 和 LAM 模块。通过对下采样过程中的特征进行处理,将它与上采样路径中的特征进行拼接,以增加图像细节并更充分地利用每一层的多尺度信息。实验结果表明,本文方法的 Dice 系数为 0.91,准确率达到 0.98,MIoU 达到 0.89,F1 值达到 0.9,HD 为 6.56,相比现有方法各项指标均有所提高,可为临床提供可准确的脉络膜分割结果。

参考文献:

- [1] NICKLA D L, WALLMAN J. The multifunctional choroid [J]. *Progress in Retinal and Eye Research*, 2010, 29(2): 144-168.
- [2] ALONSO-CANEIRO D, READ S A, COLLINS M J. Automatic segmentation of choroidal thickness in optical coherence tomography [J]. *Biomedical Optics Express*, 2013, 4(12): 2795-2812.
- [3] TIAN J, MARZILIANO P, BASKARAN M, et al. Automatic measurements of choroidal thickness in EDI-OCT images[C]. 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. August 28-September 1, 2012. San Diego, CA. IEEE, 2012: 5360-5363.
- [4] YAN Q F, GU Y Y, ZHAO J Y, et al. Automatic choroid layer segmentation in OCT images via context efficient adaptive network[J]. *Applied Intelligence*, 2023, 53(5): 5554-5566.
- [5] 许祥丛, 陈俊彦, 王雪花, 等. 通过 SEC-UNet 精准分割糖尿病视网膜病变眼底 OCT 图像脉络膜层[J]. *生物化学与生物物理进展*, 2022, 49(12): 2450-2457.
- [6] XU X C, CHEN J Y, WANG X H, et al. Accurate segmentation of choroidal layer in OCT images of diabetic retinopathy by SEC-UNet[J]. *Progress in Biochemistry and Biophysics*, 2022, 49(12): 2450-2457. (in Chinese)
- [7] ZHU L, LI J M, ZHU R L, et al. Synergistically segmenting choroidal layer and vessel using deep learning for choroid structure analysis[J]. *Physics in Medicine and Biology*, 2022, 67(8). DOI: 10.1088/1361-6560/ac5ed7.

- [7] KHAING T T, OKAMOTO T, YE C, *et al.* ChoroidNET: a dense dilated U-net model for choroid layer and vessel segmentation in optical coherence tomography images [J]. *IEEE Access*, 2021, 9: 150951-150965.
- [8] XU X C, WANG X H, LIN J Y, *et al.* Automatic segmentation and measurement of choroid layer in high myopia for OCT imaging using deep learning [J]. *Journal of Digital Imaging*, 2022, 35 (5): 1153-1163.
- [9] WU W J, GONG Y, HAO H Y, *et al.* Choroidal layer segmentation in OCT images by a boundary enhancement network[J]. *Frontiers in Cell and Developmental Biology*, 2022, 10: 1060241.
- [10] WANG X L, GIRSHICK R, GUPTA A, *et al.* Non-local neural networks [C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. June 18-23, 2018. Salt Lake City, UT, USA. IEEE, 2018: 7794-7803.
- [11] XU L, WANG L, LI Y, *et al.* Big Model and Small Model: remote modeling and local information extraction module for medical image segmentation [J]. *Applied Soft Computing*, 2023, 136: 110128.
- [12] CHEN J, LU Y, YU Q, *et al.* TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation [EB/OL]. 2021: *arXiv*: 2102.04306. <https://arxiv.org/abs/2102.04306.pdf>.
- [13] HE K M, ZHANG X Y, REN S Q, *et al.* Deep residual learning for image recognition [C]. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 27-30, 2016. Las Vegas, NV, USA. IEEE, 2016: 770-778.
- [14] MILLETARI F, NAVAB N, AHMADI S A. V-net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation [EB/OL]. 2016: *arXiv*: 1606.04797. <https://arxiv.org/abs/1606.04797.pdf>.
- [15] KERMANY D, ZHANG K, GOLDBAUM M. Labeled optical coherence tomography (oct) and chest x-ray images for classification [J]. *Mendeley data*, 2018, 2(2): 651.
- [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 International Publishing, 2015: 234-241.
- [17] WANG S, LI L, ZHUANG X. AttU-Net: attention U-Net for brain tumor segmentation [C]. *International MICCAI Brainlesion Workshop*. Cham: Springer International Publishing, 2021: 302-311.
- [18] ALOM M Z, HASAN M, YAKOPCIC C, *et al.* Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation [J]. *arXiv preprint arXiv:1802.06955*, 2018.
- [19] ZUO Q, CHEN S, WANG Z. R2AU-Net: attention recurrent residual convolutional neural network for multimodal medical image segmentation [J]. *Security and Communication Networks*, 2021, 2021: 1-10.
- [20] CHEN J, LU Y, YU Q, *et al.* Transunet: Transformers make strong encoders for medical image segmentation [J]. *arXiv preprint arXiv: 2102.04306*, 2021.
- [21] CAO H, WANG Y, CHEN J, *et al.* Swin-unet: Unet-likeic pure transformer for medical image segmentation [C]. *European Conference on Computer Vision*. Cham: Springer Nature Switzerland, 2022: 205-218.
- [22] JAEGER S, CANDEMIR S, ANTANI S, *et al.* Two public chest X-ray datasets for computer-aided screening of pulmonary diseases [J]. *Quantitative Imaging in Medicine and Surgery*, 2014, 4 (6): 475.

作者简介:



黄文博(1980—),男,吉林长春人,博士,教授,硕士生导师,2018年于吉林大学获得博士学位,主要从事模式识别、图像处理及自然语言处理方面的研究。E-mail: huangwenbo@sina.com



屈超凡(1999—),男,江苏南通人,硕士研究生,主要从事医学图像自动分割算法的研究。E-mail: fcqkzy@163.com