

位置注意力引导下的青光眼自动筛查方法^①



赵天泽¹, 刘巧红¹, 李新宇², 林敏³

¹(上海健康医学院 医疗器械学院, 上海 201318)

²(上海理工大学 健康科学与工程学院, 上海 200093)

³(上海市同济医院, 上海 200065)

通信作者: 林敏, E-mail: linm_doc@163.com

摘要: 为了提高青光眼疾病的预测和诊断的准确性, 避免人工筛查造成的误差累积, 本文提出了一种位置注意力引导下的青光眼自动筛查方法. 所提出的模型包含了眼底图像注意力预测和青光眼疾病分类两个部分. 首先, 提出了一个基于结合深度理解卷积核和通道激励连接空间金字塔的 U 型网络进行眼底图像注意力预测, 并将解码过程中的特征图作为空间信息引导青光眼分类. 其次, 提出了在青光眼分类模型中使用的位置注意力机制, 该注意力机制结合不同来源的通道信息与空间信息对来自外部编码器的特征图进行动态调整. 青光眼分类模型的主分支堆叠了多个位置注意力模块和残差模块用于实现分类任务, 同时设计了一个用于分割任务的辅助分支协助模型训练和优化, 提高分类精度. 所提方法基于青光眼 LAG 数据集测试的精确度、召回率和 AUC 指标分别达到 97.84%、97.75% 和 98.57%, 表现优于所有对比模型. 通过可视化注意力激活热图得到的模型决策关注区域更加准确, 辅助临床诊断中对病灶的定位, 并为临床诊断的结果提供有效的参考.

关键词: 青光眼筛查; 卷积神经网络; 深度理解卷积核; 通道激励连接空间金字塔; 位置注意力

引用格式: 赵天泽,刘巧红,李新宇,林敏.位置注意力引导下的青光眼自动筛查方法.计算机系统应用. <http://www.c-s-a.org.cn/1003-3254/9840.html>

Glaucoma Automatic Screening Method Based on Position Attention Guidance

ZHAO Tian-Ze¹, LIU Qiao-Hong¹, LI Xin-Yu², LIN Min³

¹(College of Medical Instruments, Shanghai University of Medicine and Health Sciences, Shanghai 201318, China)

²(School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

³(Shanghai Tongji Hospital, Shanghai 200065, China)

Abstract: To improve the accuracy of predicting and diagnosing glaucoma and avoid the accumulation of errors caused by manual screening, this study proposes an automatic glaucoma screening method guided by position attention. The proposed method includes two parts: attention prediction of fundus images and glaucoma disease classification. First, a U-shaped network based on the combination of deep understanding convolution kernels and channel excitation connection spatial pyramids is proposed to predict the attention of fundus images. Feature maps in the decoding process are used as spatial information to guide glaucoma classification. Second, a position attention mechanism used in the glaucoma classification model is proposed, which combines channel information and spatial information from different sources to dynamically adjust the feature maps from external encoders. The main branch of the glaucoma classification model stacks multiple position attention modules and residual modules to fulfill the classification task. At the same time, an auxiliary branch for segmentation tasks is designed to assist in model training and optimization to improve classification accuracy. The precision, recall, and AUC of the proposed method based on the glaucoma LAG dataset test reach 97.84%, 97.75%, and 98.57% respectively, which outperform all the comparative models. The model decision attention area obtained by

^① 收稿时间: 2024-10-16; 修改时间: 2024-11-07, 2024-11-29; 采用时间: 2024-12-04; csa 在线出版时间: 2025-02-28

visualizing the attention activation heat map is more accurate, assisting in locating the lesions in clinical diagnosis and providing an effective reference for the results of clinical diagnosis.

Key words: glaucoma screening; convolutional neural network (CNN); deep understanding convolutional kernel; channel excitation link spatial pyramid; position attention

青光眼是一种慢性进行性视神经病变,表现为视乳头萎缩及凹陷、视野缺损、视力下降等,据布莱恩霍尔顿视觉研究所的相关报道,青光眼已经成为全球第二大致盲原因,其患病率预计将从2020年的7600万人增加到2040年的1.118亿人^[1]。青光眼通常在早期无明显症状,多数患者在未感到明显不适的情况下病情恶化至晚期,症状一旦明显意味着患者已遭受到不可逆的视野和视力损失。因此,早筛查、早诊断并进行及时干预是预防和控制青光眼的关键。然而,青光眼的检查主要依赖于临床医生对于眼底图像进行经验性的分析和判断,同时检查过程耗时复杂也不利于大规模的筛查。近年来,人工智能和计算机辅助诊断技术的发展为青光眼的快速准确筛查和诊断提供了新的途径。

目前在利用深度学习技术针对青光眼的智能诊断方面,主要集中在眼底图像的分割和分类任务上。基于深度学习算法实现对眼底图像中视杯和视盘的有效分割,从而计算出杯盘比可用于青光眼的初步筛查。后者利用深度学习算法学习眼底图像特征并将图像分类为正常眼底图像和青光眼眼底图像,判断是否患有青光眼。

自2015年图像分割的基准网络U-Net^[2]被提出并成功应用到医学图像分割任务上,不少针对U-Net的改进模型被设计用来实现眼底图像的视杯和视盘分割。Fu等人^[3]在U-Net的基础上增加了侧输出层,提出了一种基于多标签深度网络M-Net的视杯和视盘联合分割方法,并将眼底图像转换到极坐标系中,利用空间约束、等价增加、平衡杯状比例来提高分割精度。Tulsani等人^[4]为了评估青光眼改进了UNet++网络模型对视杯和视盘进行自动分割,并提出一个混合交叉熵和Dice系数的损失函数以解决类别不平衡的问题。Mallick等人^[5]提出了一种改进后的U-Net模型RFAUCNxt,由预训练的ConvNeXt基础作为编码器和修改后的ConvNeXt模块作为轻量级的解码器组成,进一步设计了一种新的注意力门模块融合上采样过程中的低级和高级语义信息,从而提高了视盘和视杯的分割精度。梅

华威等人^[6]将U-Net的编码部分融入ECA注意力模块进行特征提取,并设计了残差空洞卷积模块和路径增强模型进一步提取多尺度的上下文语义信息,增强整个特征层次,结合多标签损失函数,提高了视杯盘与背景区域的像素比例,指导网络分割。

随着视盘分割技术的不断发展,近来,将分割出的视盘作为指导进行青光眼的筛查成为一种提高判别准确率的有效方式。Fu等人^[7]提出了一种用于自动青光眼筛查的新型视盘感知集成网络DENet,它整合了全局眼底图像和局部视盘区域的深度分层上下文,将全局眼底图像、分割引导网络、局部视盘区域和极坐标下的视盘区域4个不同输出概率融合作为最终的筛查结果,实验表明融合模型的准确率有着显著地提升。Kashyap等人^[8]基于改进的U-Net架构和DenseNet-201模型的迁移学习策略,训练青光眼检测与分类的模型,对自建的眼底图像数据集进行青光眼评估,有效地从视网膜眼底图像中分割出视杯区域,并提取相关特征,进而提高青光眼诊断的准确性。吴星等人^[9]基于人工智能眼底分析技术建立了青光眼病灶诊断系统,该系统包含了分割-分类一体化预测模型和杯盘比计算模块,在诊断青光眼疾病的同时探索青光眼疾病过程中杯盘比的发展情况。

由于视盘分割图只有前景背景两种分类信息,其仅能提供简单的位置信息,而不包含更深层次的空间权重信息,这成为分类任务效果进一步提升的瓶颈。而模拟医生临床诊断时的眼球活动所得到的注意力图,其包含了更深层的权重信息,所以使用注意力图取代分割图成为突破这一瓶颈的有效方式。Liu等人^[10]建立了包含由模拟眼球追踪形成的注意力图的公开数据集LAG database,并提出了包含注意力预测、病理区域定位与青光眼分类3部分子模块的新网络AG-CNN。郭璠等人^[11]提出一种语义特征图引导的青光眼筛查方法,利用分割模型进行视盘区域的分割定位,并且根据定位结果提取用于青光眼筛查的重点感兴趣区域,再通过设计注意力模块生成注意力图,强化特征分类相关

区域而抑制非相关区域. 将重点关注区域进行分割后结合分类任务, 提升了单一图像分类模型在复杂场景下的准确性和实际应用效果.

不过目前已有的利用视盘注意力图对青光眼进行分类的方法, 利用注意力图的方式较为单一, 缺少不同任务, 即视盘注意力预测任务与青光眼分类任务之间的特征信息的转换. 为了解决这一问题, 提高青光眼疾病的预测和诊断的准确性, 本文提出了一种视盘注意力引导下的青光眼自动筛查方法. 具体而言, 首先提出了一种结合深度理解卷积核 (deep understanding convolutional kernel, DUCK)^[12]和通道激励连接空间金字塔 (channel excitation link spatial pyramid, CELSP) 的视盘注意力预测网络, 对眼底图像进行注意力预测. 进一步提出了一种基于位置注意力 (position attention, PA) 的青光眼分类模型, 将视盘注意力预测网络中解

码过程的特征图作为位置信息与分类模型中的特征图进行融合, 使得模型更加关注于病变相关区域从而抑制不相关区域, 提升模型预测的准确性. 最后, 利用梯度加权类别激活映射性 (gradient-weighted class activation mapping, Grad-CAM)^[13]算法实现注意力激活热图可视化, 增强网络可解释性.

1 方法概述与总体流程

图 1 是整个视盘位置注意力引导下的青光眼自动筛查方法的工作流程. 首先将原始的眼底图像输入结合深度理解卷积核和通道激励连接空间金字塔的注意力图预测网络, 得到最终预测的眼底注意力图. 再将注意力图与原始眼底图像拼接后作为基于位置注意力的青光眼分类模型的输入, 结合位置注意力机制提高对相关区域的识别精度, 最终输出分类识别的结果.

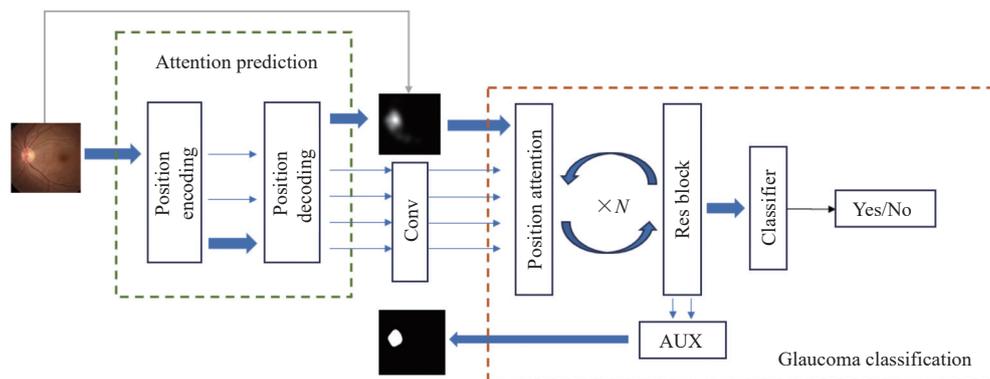


图 1 视盘位置注意力引导下的青光眼自动筛查方法的工作流程

注意力预测网络基于 U 型网络构架设计, 包括了位置编码器 (position encoding module, PEM) 和位置解码器 (position decoding module, PDM). 其中 PEM 由残差卷积模块组成, 在提高特征提取能力的同时避免梯度消失问题. PDM 利用了深度理解卷积核模块和通道激励连接空间金字塔模块, 深度理解卷积核模块在不同尺度上增大感受野进一步融合特征的同时不增加模型参数, 通道激励连接空间金字塔模块加强了通道间的相互作用和空间关联性, 同时平衡了高级特征和空间细节.

青光眼分类模型由主分支和辅助分支组成, 其中主分支包含了残差卷积模块和位置注意力模块, 由来自 PDM 的外部特征图作为位置注意力模块的额外输入, 使得模型更加关注于感兴趣的区域而抑制非相关区域. 辅助分支则以分割为任务, 将分割标签作为先验

信息, 增强模型对定位能力的学习.

2 位置注意力引导的青光眼分类方法

该方法的网络结构由注意力预测模型与分类模型两个子模型组成, 其结构如图 2 所示. 首先由注意力预测模型得到模拟医生临床诊断时眼球活动的注意力图, 该注意力图将与原始图像在通道方向拼接后, 再作为分类模型的输入. 不仅注意力预测模型的最终输出会参与分类任务, 注意力预测过程中的特征图, 即 PDM 中的特征图, 也将在通过一个 3×3 的卷积层后, 与分类模型中的特征图一同输入至 PA 模块参与分类任务.

2.1 视盘注意力预测模型

2.1.1 注意力预测模型网络结构

注意力预测网络模型为 U 型结构, 由 PEM 与 PDM 组成, 其结构如图 2 左侧 Attention prediction 部

分所示. 其中 PEM 使用了基于 ResNet50^[14]的网络结构, 由 4 个残差模块 Res Block 组成, 但其中的部分残差模块将普通卷积改为了空洞卷积以维持特征图大小. 如图 2 的 PEM 部分, 一个 Res Block 中由 N 个子模块组成, 第 1 个子模块通过步距为 S 的卷积调整输入特征的大小, 并将输入特征的通道数扩张到 4 倍, “短链接”中使用一个 1×1 卷积调整通道数. 剩余的

$N-1$ 个子模块先对通道数进行压缩, 最后再还原至原本通道数, 从而减少参数量, 避免参数浪费. 在该网络中, 第 2 个 Res Block 进行了 2 倍的下采样以减小特征图尺寸. 第 3 和第 4 个 Res Block 不进行下采样, 使用空洞卷积通过增加扩张系数 D 来扩大感受野, 有助于网络捕捉更大范围的上下文信息, 提高分割的准确性.

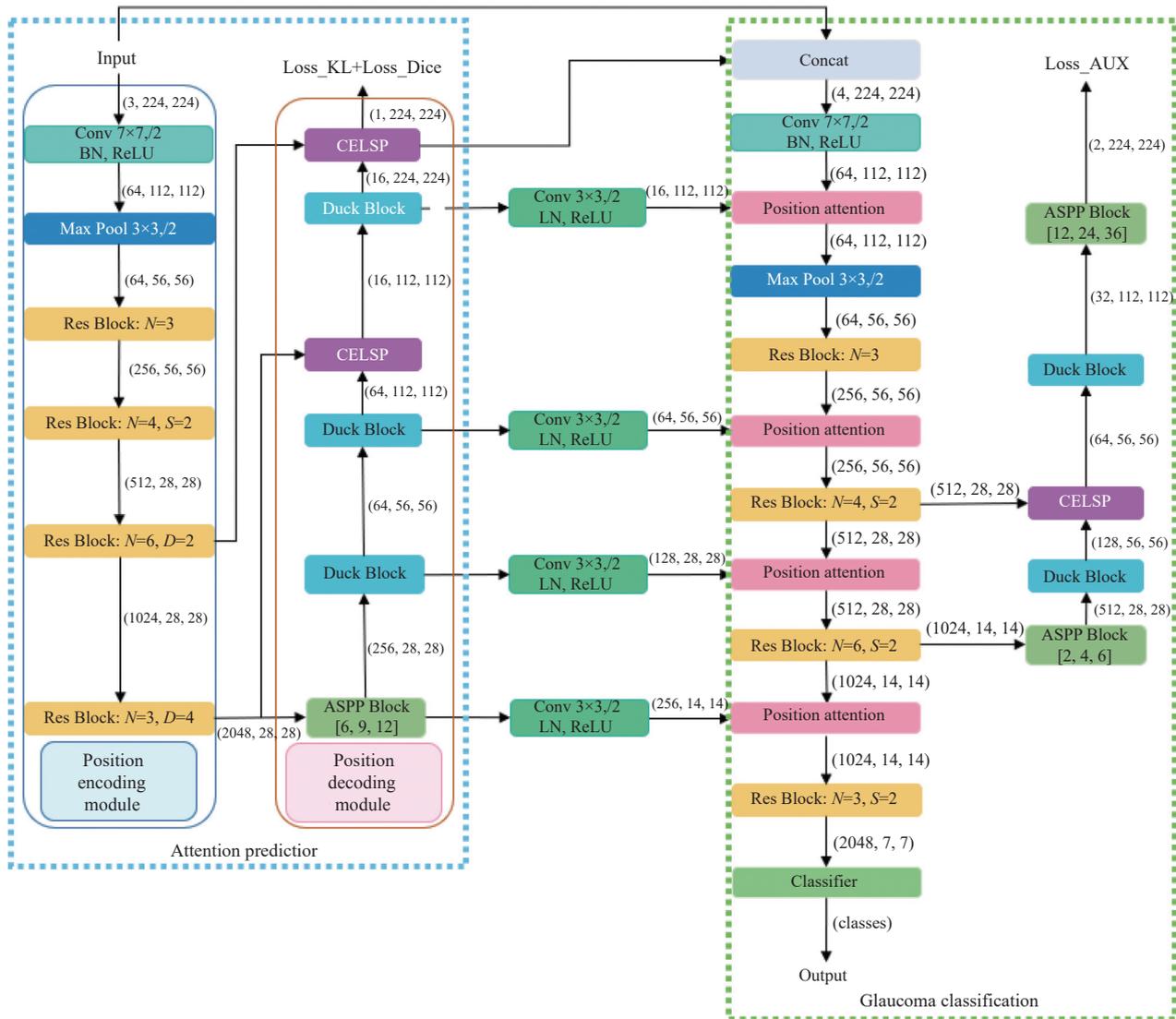


图 2 由两个子模型组成的主模型结构图

PDM 混合了 1 个 ASPP^[15]模块、3 个 DUCK 模块和 2 个 CELSP 模块, 在双线性插值进行上采样的同时, 使用深度理解卷积核模块 DUCK 推理编码器中丢失的信息, 并提出新的通道激励连接空间金字塔模块 CELSP, 引入通道激励后的编码信息融合. PDM 使用

ASPP 模块作为第 1 个模块, 其能在多种尺度上提取解码器输出的信息, 进而生成更适合作为输入的深层语义特征. 为增强 ASPP 模块的泛化能力, 在其基础结构中增设了失活层、 3×3 卷积及 1×1 卷积, 后者调整通道数以优化信息处理. 此外, 为避免编码器与解码器间

的语义鸿沟导致的语义特征丢失, 设计了双链接结构, 将编码器中的后两个 Res Block 的输出跳转连接到相应的 CELSP 模块, 减少了解码器中特征的重复使用, 以提升网络识别新特征的能力.

2.1.2 深度理解卷积核模块

深度理解卷积核模块 DUCK^[12]是一种新型的卷积核, 其设计了 6 种并行的卷积核, 使得网络在训练时采用最合适的卷积核. 本文所采用的 DUCK 块如图 3 所示, 其中第 1 个分支和第 2 个分支分别模拟 15×15 和 7×7 的大卷积核, 使用扩展卷积来减少模拟大卷积核所

需要的参数, 同时使得网络更好地理解高级别的特征, 有助于保持分割的整体结构完整性. 最后一个分支利用通过组合空间可分离卷积, 组合 1×N 卷积核和 N×1 卷积核来模拟 N×N 大小的卷积核, 这样可以使用更少的参数量来实现相同的感受野, 但由于受到卷积核方向的限制, 会难以捕捉感受野内对角线方向的信息. 中间 3 个分支分别组合了不同数量的残差块 Residual Block 来获得不同的感受野, 模拟 5×5、9×9 与 13×13 的卷积核大小, 这对于捕获图像中不同尺寸区域的细微特征至关重要.

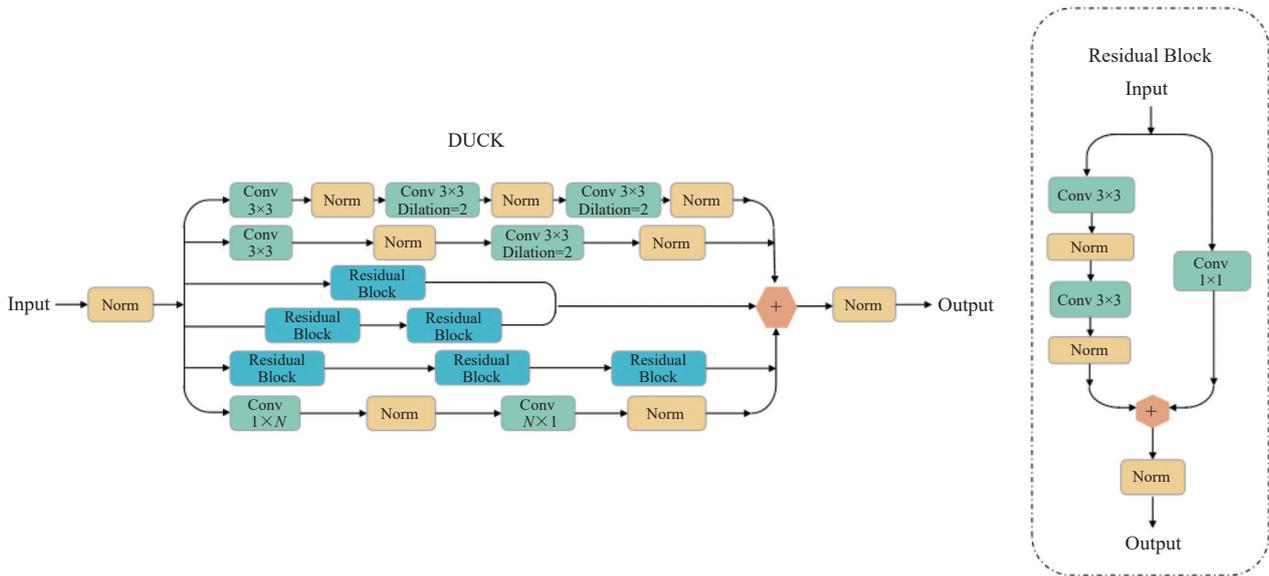


图 3 深度理解卷积核模块

DUCK 模块通过并行操作 6 种不同的卷积核, 使得网络能够在学习过程中选择最适合的特征提取策略. 这个模块的扩展感受野有助于在特征图尺寸逐步扩大的插值过程中, 有效地推理和调整放大区域的细节. 另外本文在 DUCK 模块之前采用双线性插值上采样, 以优化整体特征恢复效果.

2.1.3 通道激励连接空间金字塔模块

在深度网络捕提高级语义特征的过程中, 空间特征的重要性逐渐减少, DUCK 模块为了更好地获取广泛的深层信息, 可能会牺牲一些低级的细节信息, 为了弥补这一缺陷, 本文提出了一个新的通道激励连接空间金字塔模块 CELSP. 该模块通过强化通道间的相互作用和空间关联性, 旨在改善高级特征和空间细节的平衡, 其结构如图 4 所示.

新设计的 CELSP 模块, 在使用空间金字塔 (spatial

pyramid, SP) 整合 DUCK 捕获的信息的同时, 通过通道提取链接 (channel excitation link, CEL) 模块引入编码器中包含更多原始图像信息的低级特征, 从而弥补 DUCK 模块可能丢失的低级细节. 这两个模块的输出相加后, 通过残差结构处理以产生最终输出.

CEL 模块的输入特征图来自解码器, 其本质为一个通道激励模块^[16]. 对输入特征进行全局平均池化和 1×1 卷积后, 再使用 Sigmoid 函数激活, 得到一个可学习的通道权重, 将通道权重与通过 1×1 卷积调整通道后的特征图相乘, 得到通道激励后的特征. 进行双线性插值上采样调整大小与 1×1 卷积调整通道后的输出, 可以视为包含了浅层信息的低级特征.

SP 模块在使用 1×1 卷积之外, 分别使用了 3×3 卷积, 3×3 的空间可分离卷积, 以及空洞率为[1,3]和 3×3 空洞卷积与 3×3 空洞空间可分离卷积. SP 模块在

不同尺寸的较小感受野中进行特征捕获,与 DUCK 模块主要在较大区域捕捉特征的特点形成互补.考虑到在眼底图像中,眼视盘的形状通常是椭圆形,长轴可能在不同的方向,因此在 SP 模块中使用了空间可分离卷

积,有助于分别学习图像的水平 and 垂直特征,从而更好地适应眼视盘的方向性和形态变化.实验表明,在 SP 模块中使用空间可分离卷积确实提高了模型的精度.

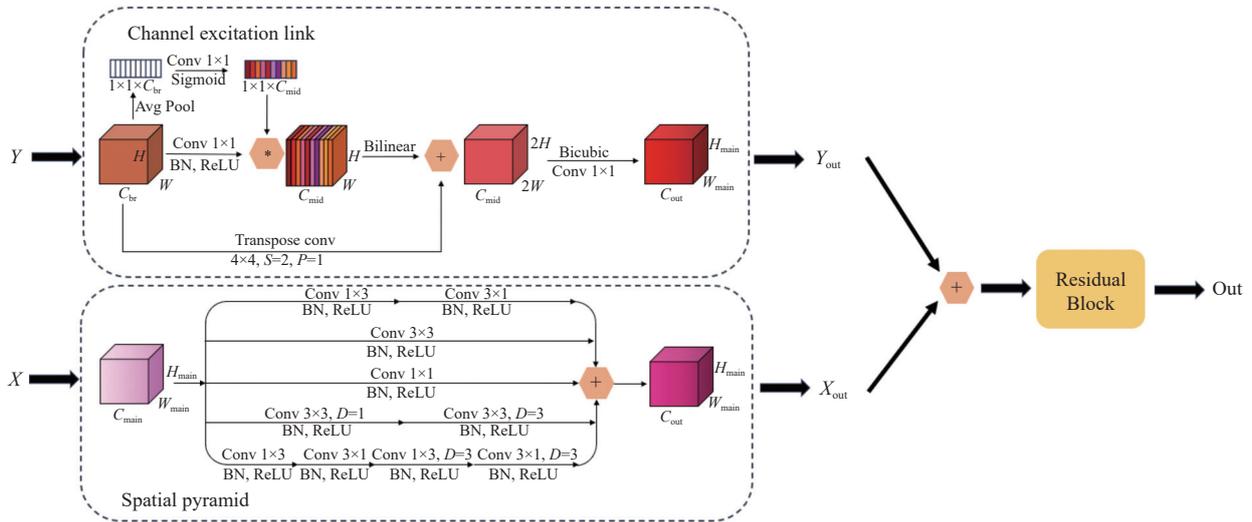


图 4 通道激励连接空间金字塔模块

2.1.4 注意力预测任务损失函数

实验中,模型训练的损失函数为 KL 散度 (Kullback-Leibler divergence) 与 Dice_Loss 函数,仅使用 KL 散度作为损失函数,得到的注意力图可能不够集中,所以额外使用 Dice_Loss 函数对注意力区域进行约束.

KL 散度的数学表示为:

$$L_{KL} = \frac{1}{N} \sum_{i=1}^N y_i \log \left(\frac{y_i}{p_i} \right) \quad (1)$$

Dice_Loss 函数的数学表示为:

$$L_{Dice} = 1 - \frac{2 \cdot \sum_{i=1}^N p^i y^i}{\sum_{i=1}^N p^i + \sum_{i=1}^N y^i + \epsilon} \quad (2)$$

其中, N 是样本的总数, p^i 表示样本的像素 i 的预测值; y^i 表示样本注意力图标签的像素 i 归一化的值. Dice_Loss 中的 ϵ 表示一个平滑项,即接近于 0 的小常数,用于确保分母不为 0.

模型总损失则为:

$$L = \alpha \times L_{KL} + \beta \times L_{Dice} \quad (3)$$

其中, α 与 β 为平衡损失值的超参数,建议 $\alpha \geq \beta$.

2.2 青光眼分类模型

2.2.1 分类模型网络结构

在眼底图像中,青光眼相关的病变区域多为细微的局部区域,这些局部病变区域在深度卷积模型的下采样操作中可能存在不可逆的信息丢失,导致有效的判别性局部区域被忽略,模型趋向过拟合.因此,针对这一问题,本文提出了一个基于位置注意力的青光眼分类模型,利用预测的注意力图及其过程中产生的特征图对分类过程进行引导.同时,增加了一个以分割为任务的辅助分支,帮助 PA 模块从先验信息中学习空间信息.其结构如图 2 右侧 Glaucoma classification 部分所示.

分类任务的主干网络同样基于 ResNet50,由多个残差模块组成,但本文在残差模块之间插入了原创的 PA 模块.

辅助分支 (auxiliary branch, AUX) 辅助分支首先使用扩张系数为 [2,4,6] 的 ASPP 模块扩大感受野,接下来通过 DUCK 模块进行上采样,同时捕获更大范围的深层特征.由于浅层网络的梯度调整较难,辅助分支中的 CELSP 模块引入第 2 个残差模块的输出,以更有效地调整主分支的前两个 PA 模块.最终,通过第 2 个 DUCK 模块上采样后,通过扩张系数为 [12, 24, 36] 的

ASPP 模块得到最终的输出. 辅助分支基于主干网络实现了分割任务, 虽然分割不是该模型的主要任务, 但在进行分割时, 模型需要识别眼底图像中的细节特征并理解其中不同部分的结构关系, 这些细粒度特征与结构信息对于分类任务非常有帮助. 此外, 通过辅助分支输出分割结果, 可以帮助网络更准确地定位视盘位置, 使分类网络更好地定位关键区域.

2.2.2 位置注意力模块

为了实现在注意力预测任务与分类任务之间的特征转换, 使注意力预测任务中的特征图帮助分类网络更加关注于病变相关区域从而抑制不相关区域, 本文提出位置注意力模块 PA, 帮助网络融合来自 PDM 包

含了位置信息的深层特征图, 其结构如图 5 所示. 该模块对来自 PDM 的特征图与分类模型中的特征图进行分别处理.

分类模型内部的特征图, 将通过两个不同的 1×1 卷积核, 得到 query 矩阵与 key 矩阵, 在对两个矩阵进行展平处理后, 再将 query 矩阵进行转置. 将转置后的 query 矩阵与 key 矩阵进行矩阵乘法, 得到 relation 矩阵, 该矩阵包含了二维空间中各像素位置之间的相关性. 对 relation 矩阵进行以行为单位的线性运算, 并将结果的尺寸调整为宽高与输入特征图相同, 最终得到 value 矩阵, 其代表了每个像素位置与其他像素位置的相关性加权.

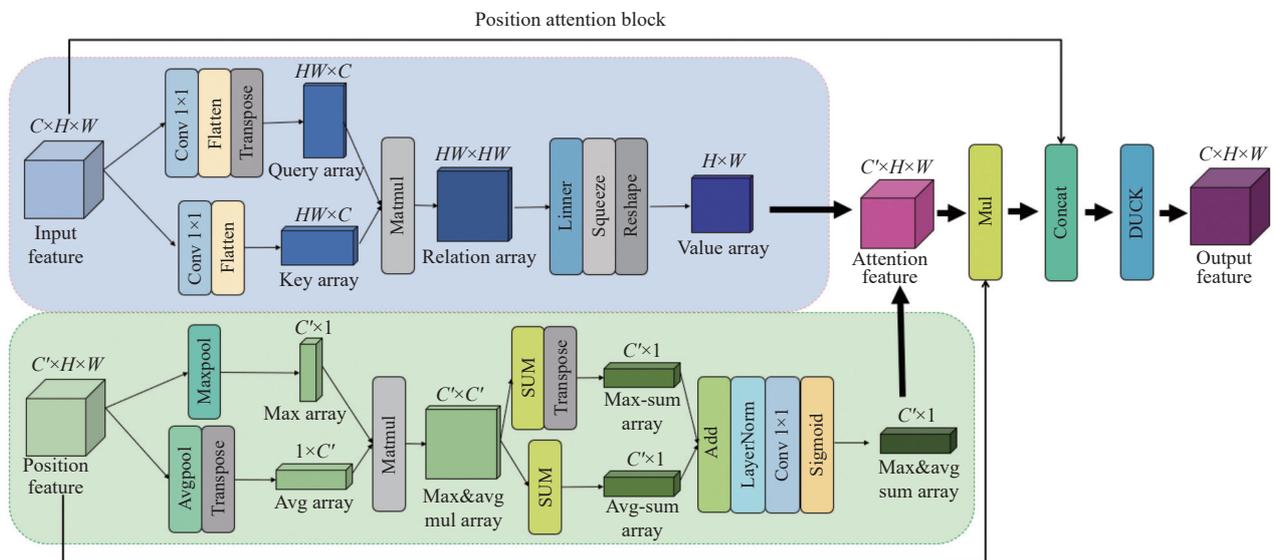


图 5 位置注意力模块

来自 PDM 的外部特征图, 将分别进行最大值池化与平均值池化, 得到最大值矩阵与平均值矩阵. 对其中之一进行转置, 并进行矩阵乘法, 得到的矩阵中, 每一个元素均代表了两个通道或同一通道的最大值与平均值的乘积. 通过对该矩阵的行列方向分别求和, 得到每个通道的平均值与其他通道的最大值乘积之和, 与每个通道的最大值与其他通道的平均值乘积之和. 将不同方向求和所得的矩阵相加, 并经过层归一化、 1×1 卷积层以及 Sigmoid 激活函数后, 得到最终的通道注意力权重.

由分类模型内部特征图得到的 value 矩阵, 与来自 PDM 的外部特征图的通道注意力权重, 通过二者可以对外部特征图进行空间与通道上的调整, 得到深层

的位置信息. 将最终的位置信息与内部特征图进行通道方向的拼接, 并经过一个 DUCK 模块还原至原本的通道数, 即 PA 模块的最终输出. 这种策略避免了直接将外部特征图强硬地与内部特征图融合在一起, 而是结合内部特征图的空间信息和外部特征图自身的通道权重, 对外部特征图进行通道和空间上的调整, 实现注意力预测任务与分类任务之间的特征转换, 有效地提升了模型对细微病变区域的识别能力.

2.2.3 分类器

在主分支末尾处使用的分类器为基于 vision Transformer 的多层感知器 (MLP) 分类器^[17], 其结构如图 6 所示. 输入特征图首先经过 1×1 卷积层, 以提取局部特征并调整维度. 在 Patch embedding 层, 特征图被分割

成多个块,并转换为平坦化的二维向量. MLP 包含多个全连接层, GELU 激活函数和 Dropout 层,旨在提取高级特征并建立图像各部分之间的长距离联系. 特征在归一化后需要重新排列,随后通过全局平均池化和展

平操作处理,最终通过全连接层输出模型的分​​类预测. 对比于由全连接层或最大池化函数组成的传统分类器,该结构能够处理更复杂的图像数据,并通过这些连续的变换提取对于图像分类任务更有意义的特征.

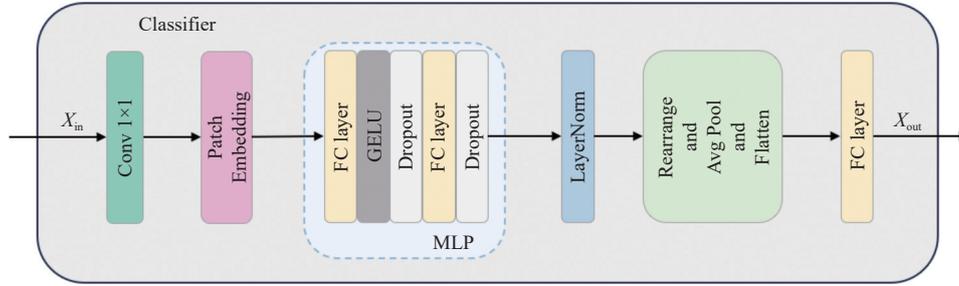


图 6 分类器

2.2.4 分类任务损失函数

本网络中,模型损失函数由主分支和辅助分支两部分组成. 主分支的损失值由交叉熵损失函数计算,辅助分支的损失值则由交叉熵损失函数与 Dice_Loss 函数计算.

交叉熵损失函数的数学表示为:

$$L_{CE} = -\frac{1}{N} \sum_{i=1}^N [y^i \cdot \log(p^i) + (1 - y^i) \cdot \log(1 - p^i)] \quad (4)$$

模型总损失的计算公式如下:

$$L = a \times L_{CE}^{main} + (b \times L_{CE}^{aux} + c \times L_{Dice}^{aux}) \quad (5)$$

其中, L_{CE}^{main} 为主分支的交叉熵损失, L_{CE}^{aux} 为辅助分支的交叉熵损失, L_{Dice}^{aux} 为辅助分支的 Dice_Loss. a 、 b 与 c 为平衡损失值的超参数,建议 $a \gg b = c$.

3 分析与讨论

3.1 实验数据集

为了验证本文提出方法的有效性,实验采用了包含了青光眼专家标签的公开眼底数据集 LAG database^[10],该数据集由 4854 幅彩色眼底图像组成,其中非青光眼眼底图像 3143 幅,疑似青光眼患者眼底图像 1711 幅. 该数据集中所有眼底数据均由合格的青光眼专家综合考虑眼内压、视野以及视盘病变而得出诊断结果. 每幅眼底图像都配有对应的注意力图,其通过捕获专家诊断过程中专家的注意力区域生成. 由于该数据集中并没有提供分割标签,本文对注意力图进行二值化,取阈值为 100,将得到的结果作为分类网络辅助分支的训

练标签. 是数据集中眼底图像、对应注意力区域以及经过处理后用于分割模型训练的二值化标签.

对于注意力预测任务,本文在人工去除小部分质量较差的样本后,将 LAG database 数据集依据 8:2 比例随机划分,作为分割网络的训练集和测试集. 考虑到非青光眼眼底图像与疑似青光眼患者眼底图像的样本数差距较大,为避免分类网络的预测倾向于样本数更多的一类,在去除小部分质量较差的样本后,随机取出 60% 的非青光眼眼底图像,其与全部的疑似青光眼患者眼底图像作为分类网络的数据集,训练集和测试集的划分比例为 8:2. 数据样本与标签如图 7 所示.

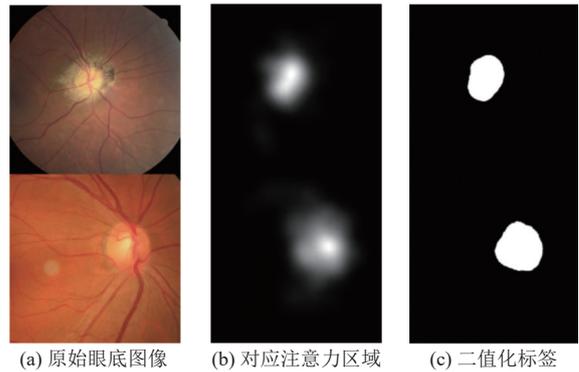


图 7 原始数据集及处理结果

3.2 参数设置及模型训练

本文实验基于 Python 语言以及 PyTorch 深度学习框架实现,采用 RTX 4060Ti 16 GB GPU 加速训练. 在模型训练阶段,使用随机裁剪、随机旋转和水平翻转等数据增强方法用于增强训练样本多样性. 采用的优化器为 AdamW,训练批量大小为 32,初始学习率设置

为 $1E-5$. 训练过程使用 ReduceLRonPlateau 作为学习率调度器, 在模型训练达到瓶颈时能够按照一定比例减少学习率, 对模型进行更微小的调整. 输入图像的像素大小为 (3, 224, 224), 尺寸调整使用双线性插值.

为了加速模型收敛并缓解数据规模偏小的问题, 本文在分类网络中使用迁移学习, 基于 ImageNet 数据集训练好的 ResNet50 初始化分类网络中残差模块的参数. 网络中除使用预训练权重的残差模块以外, 其他卷积层参数均采用 Kaiming 初始化.

模型训练可以使用一体训练, 将注意力预测模型与分类模型的损失值全部计算完毕后相加, 同时训练两个模型; 也可以进行分体训练, 在注意力预测模型训练完毕后, 读取其权重并将其冻结, 再对分类模型进行训练.

本文为了节约计算资源, 使用分体训练, 先对注意力预测模型进行训练, 注意力预测模型的损失收敛曲线如图 8 所示.

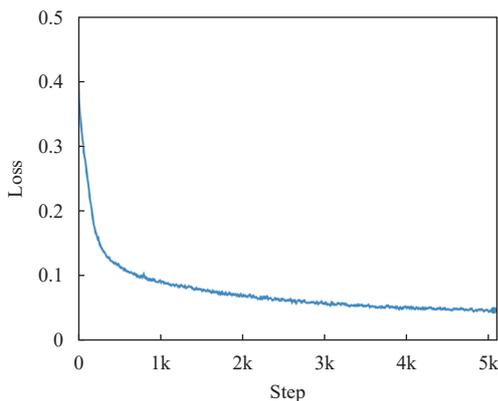


图 8 注意力预测模型损失收敛曲线

该模型的损失值最终稳定在 0.05 以下, 并且波动较小, 表示该模型在性能与鲁棒性上均有不错的表现.

在注意力预测模型训练完毕后, 读取该模型的权重并冻结, 对分类模型进行训练. 分类模型的损失收敛曲线如图 9 所示.

分类模型最终损失值浮动在 0.004 左右, 表示该分类模型的学习能力十分优秀. 并且损失值在训练之初迅速下降, 这说明了 PA 模块一定程度上加速了分类模型的拟合速度.

3.3 网络性能评估

为了验证本文提出的掩膜引导下的分类模型的有效性, 将本文方法与经典分类模型 ResNet^[14]、MobileNetV3^[18]、Swin-Transformer^[19]以及同为针对青光眼

分类任务所提出的 Chen 等^[20]方法和 AG-CNN 模型^[10]进行了实验对比, 并采用精准度 (Precision) 和召回率 (Recall)、AUC 指标 3 个评价指标对比客观结果, 反映每种模型在图像分类任务中的表现.

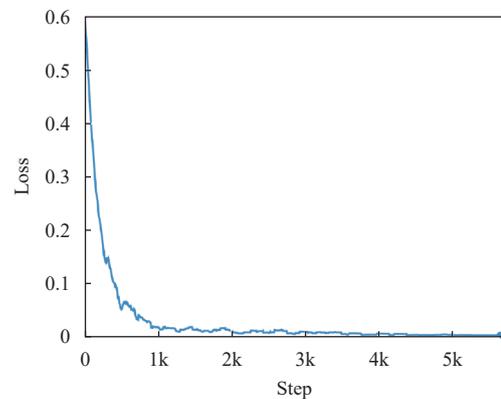


图 9 分类模型损失收敛曲线

从表 1 中可以看出, 所提分类模型在各项指标上均达到了最优, 表明模型有很强的学习能力, 能够有效地提升青光眼分类鉴别的精度. 对比于次优的 AG-CNN 模型, 精准度、召回率和 AUC 上分别提升了 0.65%、0.67% 和 0.82%.

表 1 不同方法在 LAG 数据集分类对比结果

Method	Precision	Recall	AUC
ResNet50	0.9571	0.9593	0.9638
MobileNetV3-Large	0.9518	0.9535	0.9586
Swin-Transformer	0.9657	0.9638	0.9682
文献[20]	0.9692	0.9686	0.9754
AG-CNN	0.9719	0.9708	0.9775
Ours	0.9784	0.9775	0.9857

AG-CNN 作为仅次于本文的方法, 其通过全卷积网络预测出注意力区域, 然后通过该区域对分类网络特征图加权, 这在一定程度保留了特征图上分类相关特征区域并抑制非相关区域, 从而提升了整体分类性能.

本文方法不同于 AG-CNN 那样单一地利用注意力区域, 而是在将预测的注意力与图像拼接的同时, 通过 PA 模块调整并融合预测注意力过程中所产生的深层特征图. 数值结果证明本文方法的优越性主要得益于位置注意力模块引导了分类模型对于病灶区域的关注度, 进而提升了分类性能.

3.4 消融实验

3.4.1 位置注意力消融实验

为了验证本文分类模型中提出的 PA 模块的有效性, 基于 LAG database 数据集, 通过 Grad-CAM^[13]方法

对模型决策区域进行可视化. 图 10 给出了移除 PA 模块和未移除 PA 模型的第 2 个残差模块和第 4 个残差模块决策区域的可视化图. 实验结果显示, 使用 PA 模块时, 通过 Grad-CAM 方法分析的决策区域与样本原

图中的视盘区域一致性更高. 第 2 个残差模块的输出显示 PA 模块可以在初级阶段有效地帮助模型快速定位关键视盘区域, 第 4 个残差模块的输出则进一步验证了该模块在精确定位视盘区域方面的有效性.

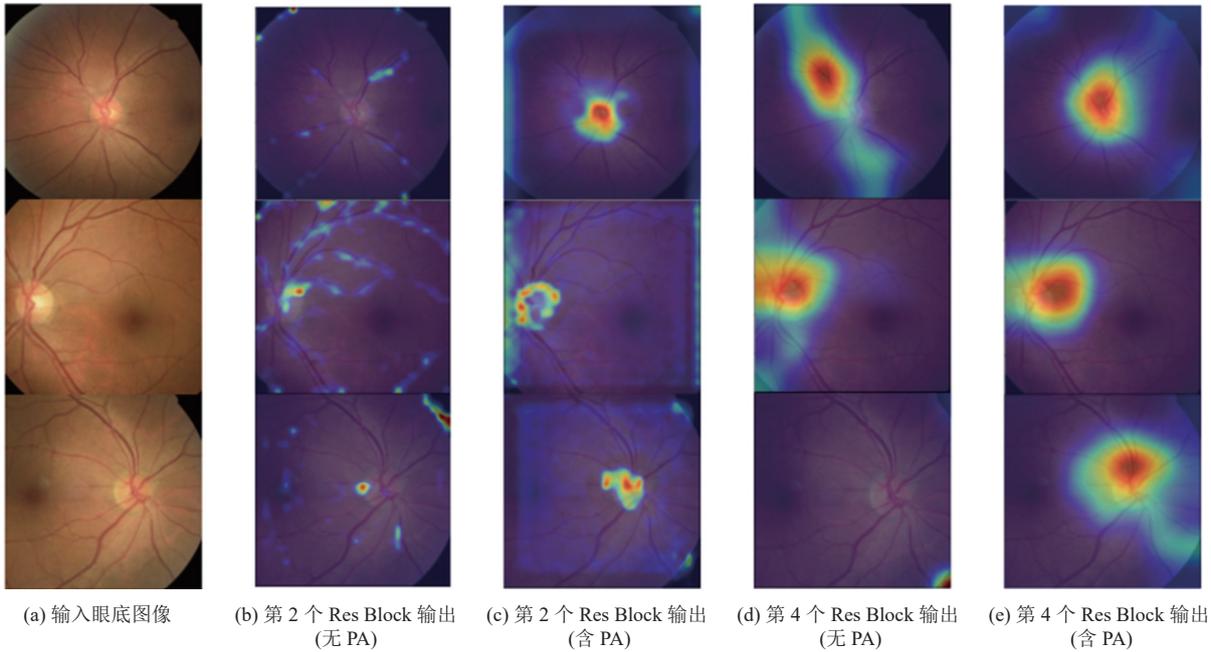


图 10 位置注意力消融实验对比结果

3.4.2 网络结构消融实验

为了验证分类模型中各种结构对于模型性能的影响, 进行了消融实验. 结构包括: 去除所有的 PA 模块和辅助分支、去除所有的 PA 模块和辅助分支的同时, 根据注意力区域裁剪数据集、去除所有的 PA 模块以及去除辅助分支 4 种设置. 并给出了不同结构的评价指标, 如表 2 所示.

Method	Precision	Recall	AUC
去除PA模块与辅助分支	0.9573	0.9589	0.9640
去除PA模块与辅助分支并对注意力区域进行裁剪	0.9554	0.9558	0.9538
仅去除PA模块	0.9648	0.9631	0.9642
仅去除辅助分支	0.9747	0.9742	0.9745
Ours	0.9784	0.9775	0.9857

在消融实验中, 同时去除 PA 模块与辅助分支, 性能会产生大幅度下降. 若此时根据注意力图对原始眼底图像进行裁剪, 模型的性能会进一步大幅度下降. 合理推测出, 完全删除背景区域, 只保留注意力集中的区域, 可能会导致细节信息的不可逆丢失. 并且裁剪后的数据集尺寸很难保持一致, 这在模型的训练过程中也

产生了不可忽略的负面影响.

仅去除 PA 模块与仅去除辅助分支的两种情况, 模型性能均产生了一定程度的下降, 这验证了两者在模型训练中发挥了有效作用. 对比仅去除 PA 模块与仅去除辅助分支的两种情况, 可以认为 PA 模块相比于辅助分支发挥了更重要的作用. 考虑到辅助分支带来的额外计算开销, 在计算资源有限的情况, 禁用辅助分支可能是一个更好的选择.

4 结论与展望

为了更有效地利用视盘注意力图, 提高青光眼早期筛查的准确性, 本文提出了一种位置注意力引导下的青光眼自动筛查方法. 首先提出了一种用于预测视盘区域注意力的网络, 该模型结合深度理解卷积核对于和通道激励连接空间金字塔, 提高模型的预测精度. 进一步将预测的注意力图与原始图像拼接, 并将解码器中的特征图作为位置注意力的额外输入, 设计了一种基于位置注意力的青光眼分类模型, 由来自解码器的深层位置信息引导模型更加关注于病变相关区域从而抑制不相关区域, 并结合辅助分支对于模型训练的

协助, 从而提升模型对于分类结果的准确性. 实验结果显示, 本文所提方法基于 LAG 数据集测试的精准度、召回率和 AUC 分别达到 97.84%、97.75% 和 98.57%, 在对比实验中, 相比于次优模型, 精准度、召回率和 AUC 指标上分别提升了 0.65%、0.67% 和 0.82%. 总的来说, 本研究不仅提高了青光眼筛查的自动化水平和准确性, 也为深度学习在医学图像分析领域的应用提供了新的方法.

参考文献

- 1 Tham YC, Li X, Wong TY, *et al.* Global prevalence of glaucoma and projections of glaucoma burden through 2040. *Ophthalmology*, 2014, 121(11): 2081–2090. [doi: [10.1016/j.ophtha.2014.05.013](https://doi.org/10.1016/j.ophtha.2014.05.013)]
- 2 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. *Proceedings of the 18th International Conference on Medical Image Computing and Computer-assisted Intervention*. Munich: Springer, 2015. 234–241.
- 3 Fu HZ, Cheng J, Xu YW, *et al.* Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. *IEEE Transactions on Medical Imaging*, 2018, 37(7): 1597–1605. [doi: [10.1109/TMI.2018.2791488](https://doi.org/10.1109/TMI.2018.2791488)]
- 4 Tulsani A, Kumar P, Pathan S. Automated segmentation of optic disc and optic cup for glaucoma assessment using improved UNet++ architecture. *Biocybernetics and Biomedical Engineering*, 2021, 41(2): 819–832. [doi: [10.1016/j.bbe.2021.05.011](https://doi.org/10.1016/j.bbe.2021.05.011)]
- 5 Mallick S, Paul J, Sil J. Response fusion attention U-ConvNext for accurate segmentation of optic disc and optic cup. *Neurocomputing*, 2023, 559: 126798. [doi: [10.1016/j.neucom.2023.126798](https://doi.org/10.1016/j.neucom.2023.126798)]
- 6 梅华威, 尚虹霖, 苏攀, 等. 融合残差上下文编码和路径增强的视杯视盘分割. *中国图象图形学报*, 2024, 29(3): 637–654.
- 7 Fu HZ, Cheng J, Xu YW, *et al.* Disc-aware ensemble network for glaucoma screening from fundus image. *IEEE Transactions on Medical Imaging*, 2018, 37(11): 2493–2501. [doi: [10.1109/TMI.2018.2837012](https://doi.org/10.1109/TMI.2018.2837012)]
- 8 Kashyap R, Nair R, Gangadharan SMP, *et al.* Glaucoma detection and classification using improved U-Net deep learning model. *Healthcare*, 2022, 10(12): 2497. [doi: [10.3390/healthcare10122497](https://doi.org/10.3390/healthcare10122497)]
- 9 吴星, 黄焯霖, 叶子, 等. 人工智能眼底分析技术对青光眼病灶的诊断价值研究. *解放军医学院学报*, 2022, 43(10): 1014–1018. [doi: [10.3969/j.issn.2095-5227.2022.10.003](https://doi.org/10.3969/j.issn.2095-5227.2022.10.003)]
- 10 Liu L, Xu M, Wang XF, *et al.* Attention based glaucoma detection: A large-scale database and CNN model. *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019. 10563–10572.
- 11 郭璠, 李伟清, 赵鑫, 等. 语义特征图引导的青光眼筛查方法. *计算机辅助设计与图形学学报*, 2021, 33(3): 363–375.
- 12 Dumitru RG, Peteleaza D, Craciun C. Using DUCK-Net for polyp image segmentation. *Scientific Reports*, 2023, 13(1): 9803. [doi: [10.1038/s41598-023-36940-5](https://doi.org/10.1038/s41598-023-36940-5)]
- 13 Selvaraju RR, Cogswell M, Das A, *et al.* Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Proceedings of the 2017 IEEE International Conference on Computer Vision*. Venice: IEEE, 2017. 618–626.
- 14 He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016. 770–778.
- 15 Chen LC, Papandreou G, Kokkinos I, *et al.* DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(4): 834–848. [doi: [10.1109/TPAMI.2017.2699184](https://doi.org/10.1109/TPAMI.2017.2699184)]
- 16 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018. 7132–7141.
- 17 Yi SL, Zhang G, Qian CX, *et al.* A multimodal classification architecture for the severity diagnosis of glaucoma based on deep learning. *Frontiers in Neuroscience*, 2022, 16: 939472. [doi: [10.3389/fnins.2022.939472](https://doi.org/10.3389/fnins.2022.939472)]
- 18 Howard A, Sandler M, Chen B, *et al.* Searching for MobileNetV3. *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision*. Seoul: IEEE, 2019. 1314–1324.
- 19 Liu Z, Lin YT, Cao Y, *et al.* Swin Transformer: Hierarchical vision Transformer using shifted windows. *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision*. Montreal: IEEE, 2021. 9992–10002.
- 20 Chen XY, Xu YW, Yan SC, *et al.* Automatic feature learning for glaucoma detection based on deep learning. *Proceedings of the 18th International Conference on Medical Image Computing and Computer-assisted Intervention*. Munich: Springer, 2015. 669–677.

(校对责编: 张重毅)