李晓辉, 肖晓霞, 何昕,等. 基于全局和局部特征融合的多损失优化视网膜图像质量评估模型[J]. 智能计算机与应用,2026,16(4):9-19. DOI:10.20169/j.issn.2095-2163.25111801

基于全局和局部特征融合的多损失优化视网膜图像质量评估模型

李晓辉[1], 肖晓霞[1], 何 昕[1], 刘灵儿[1], 彭清华[2], 晏峻峰[1], 李 阳[1]
(1 湖南中医药大学 信息科学与工程学院, 长沙 410208; 2 湖南中医药大学 中医学院, 长沙 410208)

摘 要: 视网膜图像质量评估(Retinal Image Quality Assessment, RIQA)是确保眼科诊断可靠性的关键技术环节。针对现有基于深度学习的方法过度依赖局部特征而忽视全局质量信息的问题,本研究提出一种多损失优化的全局-局部特征融合网络MLS-Net。该网络采用双分支架构,分别处理原始 RGB 图像和经 G 通道对比度受限自适应直方图均衡化(CLAHE)增强的图像,以协同利用全局质量信息和局部细节特征。全局分支引入频域引导的注意力机制,通过低频能量分布建模图像模糊与清晰度,自适应增强关键通道特征;局部分支采用结构引导的注意力机制,通过突出视盘、血管等关键解剖区域的显著性来捕捉细粒度质量信息。通过设计的双向交叉注意力模块实现特征融合,并结合多损失优化策略提升整体效能。实验结果表明,MLS-Net 在公共数据集 EyeQ 上取得精确率为 0.892 7、召回率为 0.901 0、*F*1 分数为 0.896 7 和 Kappa 值为 0.922 4。在私有数据集 RIQA-RFMiD 上实现精确率为 0.732 8、召回率为 0.717 8、*F*1 分数为 0.713 2 和 Kappa 值为 0.748 9。该研究验证了全局与局部特征融合策略在提升 RIQA 性能方面的有效性。

关键词: 视网膜图像质量评估; 频域;结构; 多损失优化; 交叉注意力
中图分类号: TP391 **文献标志码**: A **文章编号**: 2095-2163(2026)04-0009-11

Multi-loss optimization model for retinal image quality assessment based on global and local feature fusion

LI Xiaohui[1], XIAO Xiaoxia[1], HE Xin[1], LIU Ling′er[1], PENG Qinghua[2], YAN Junfeng[1], LI Yang[1]
(1 School of Information Science and Engineering, Hunan University of Chinese Medicine, Changsha 410208, China;
2 School of Traditional Chinese Medicine, Hunan University of Chinese Medicine, Changsha 410208, China)

Abstract: Retinal image quality assessment (RIQA) is crucial for ensuring reliable ophthalmic diagnosis. Current deep learning-based RIQA methods predominantly rely on local features while overlooking global quality information, which limits performance improvement. To address this limitation, this study proposes MLS-Net, a multi-loss optimized global-local feature fusion network. The architecture employs a dual-branch framework that processes both original RGB images and G-channel contrast-limited adaptive histogram equalization (CLAHE) enhanced images to synergistically utilize global quality information and local detailed features. The global branch incorporates a frequency-guided attention mechanism that models image blur and sharpness through low-frequency energy distribution to adaptively enhance critical channel features. The local branch utilizes a structure-guided attention mechanism that captures fine-grained quality information by emphasizing salient anatomical regions such as the optic disc and blood vessels. Feature integration is achieved through a designed bidirectional cross-attention module, complemented by a multi-loss optimization strategy to enhance overall performance. Experimental results demonstrate that MLS-Net achieves precision of 0.892 7, recall of 0.901 0, *F*1-score of 0.896 7, and Kappa value of 0.922 4 on the public EyeQ dataset. On the private RIQA-RFMiD dataset, it attains precision of 0.732 8, recall of 0.717 8, *F*1-score of 0.713 2, and Kappa value of 0.748 9. This research validates the effectiveness of global-local feature fusion strategy in enhancing RIQA performance.

Key words: retinal image quality assessment; frequency domain; structure; multi-loss optimization; bidirectional cross-attention

基金项目: 国家自然科学基金青年科学基金(62402180);湖南省高校联合基金(2025JJ90031);湖南省教育厅科学研究项目(23A0273);湖南省中医药管理局重点项目(A2023048);湖南省自然科学基金青年基金(2024JJ6338);长沙市自然科学基金(kq2402173)。

作者简介: 李晓辉(1997—),男,硕士研究生,主要研究方向:深度学习,眼底图像; 何 昕(2001—),女,硕士研究生,主要研究方向:深度学习,眼底图像; 刘灵儿(2002—),女,硕士研究生,主要研究方向:深度学习,面部识别; 彭清华(1964—),男,博士,教授,主任医师,博士生导师,主要研究方向:中医眼科学; 晏峻峰(1965—),女,博士,教授,博士生导师,主要研究方向:中医人工智能; 李 阳(1993—),女,博士,主要研究方向:医学图像处理,多模态医疗数据分析。
通信作者: 肖晓霞(1982—),女,博士,教授,主要研究方向:中医信息化智能化基础研究。Email: amily_x@hnucm.edu.cn。

收稿日期: 2025-11-18

0 引 言

近年来,全球人口老龄化进程加速,视网膜疾病(如年龄相关性黄斑变性、糖尿病性视网膜病变和青光眼)的发病率持续上升。通过分析视网膜图像,可实现疾病的早期检测与干预,从而降低医疗成本并改善患者生活质量[1]。然而,由于成像设备差异及操作者技术水平不一,获取的图像质量常存在显著差异。低质量图像可能导致视盘、血管及病灶等关键结构模糊,影响临床诊断。因此,自动化的视网膜图像质量评估(Retinal Image Quality Assessment, RIQA)对于筛除低质量样本、保障诊断可靠性具有重要意义[2]。

传统视网膜图像质量评估(RIQA)方法主要分为两类:基于通用质量参数[3]和基于结构质量参数[4]。前者通过分析清晰度、对比度和光照均匀性等指标来评估整体图像质量。例如,Bartling[5]等将视网膜图像划分为若干非重叠区域,分别计算亮度与对比度指标,再合成为整体质量得分。然而,该方法忽略了图像结构间的关联性,易导致整体评估偏差。后者则基于关键结构(如黄斑、视盘和血管)的可见性。例如,Usher 等利用方向匹配滤波与区域分割进行血管检测,并以血管清晰度与面积为质量指标[6]。该类方法依赖血管分割精度,分割误差会显著影响最终评估结果。

近年来,卷积神经网络(CNNs)在图像分类任务中取得了巨大成功[7-8],CNNs 也被广泛引入眼底图像质量评估领域。例如,2016 年,Mahapatra 等[9]将无监督的显著图信息与有监督的 CNN 特征相结合来评估视网膜图像质量。2017 年,Yu 等[10]使用显著区域检测和 AlexNet 提取特征,然后通过支持向量机(SVM)分类器进行分类。然而,此类方法在低质量或复杂场景下易出现显著区域检测失败,导致性能下降。2019 年,Fu 等[11]提出多颜色空间融合网络(MCF-Net),通过 RGB、HSV 与 LAB 三分支提取特征,虽提升了性能,但计算开销显著增加。2020 年,Xu 等[12]利用显著结构检测器获取解剖结构的显著图,然后将其拼接起来输入到 CNN 中,与基于结构质量参数的方法类似,这种方法严重依赖于血管分割[13]的准确性。2022 年,Xu 等[14]提出了一种由暗通道先验和亮通道先验引导的深度网络用于视网膜图像质量评估。该方法将暗-亮通道先验引入深度网络,而无需增加额外参数。暗-亮通道先验假设在所有视网膜图像中都有效,但在光照不均、严重眩光或高噪声的条件下可能会失效。Guo 等[15]提出了一种基于标签正则化的视网膜图像质量评估模型,但标签正则化的引入可能导致模型过度依赖训练集中的标签信息,从而在新测试数据上表现不佳。

上述讨论的基于深度学习的方法存在一些共性问题。首先,许多方法严重依赖特定的先验[16]或假设,这些假设可能在复杂场景或低质量图像中失效。其次,许多方法倾向于使用 DenseNet[17]这种密集连接的网络结构作为主干网络。虽然 DenseNet 通过特征重用和密集连接在局部特征提取方面表现出色,但这种设计也会导致模型在处理视网膜图像时过度依赖局部特征,而忽视了全局特征的重要性。为了解决这些问题,结合全局和局部特征的方法已成为一种潜在的解决方案。全局特征能够捕捉图像的整体结构和上下文信息,而局部特征则专注于细节和关键区域。两者的结合可以弥补单一特征提取方法的局限性,增强模型在复杂场景下的鲁棒性和准确性。

为此,本文提出一种基于全局和局部特征融合的多损失优化网络(MLS-Net),实现全局与局部特征的协同建模,以获得更可靠的视网膜图像质量评估结果。本文主要贡献如下:

1) 双分支双输入特征提取模块(EGL Module):模型采用双分支、双输入结构。原始 RGB 图像输入全局特征提取分支,G 通道经 CLAHE 增强的 RGB 图像输入局部特征提取分支,以同时获取全局与细粒度质量特征。

2) 频域全局注意力(FGA):在全局分支中,引入基于频域低频能量分布的全局注意力机制,以刻画图像锐度与模糊程度,自适应调节通道权重,突出全局质量相关特征。

3) 结构局部注意力(SLA):在局部分支中,融合血管梯度信息与黄斑、视盘区域先验,通过可学习偏移实现结构自适应,引导网络关注关键解剖区域,增强细粒度特征表达。

4) 双向交叉注意力融合模块(CAF Module):设计双向交叉注意力机制,实现全局与局部特征的交互式融合,充分发挥两者在宏观感知与细节捕获方面的互补优势,提升特征表达能力。

5) 多任务损失加权机制:提出多任务自适应加权策略,在优化主分类任务的同时,平衡全局与局部特征学习。

1　方法

1.1　MLS-Net 总体架构

为解决当前视网膜图像质量评估中过度依赖局部特征的问题，本文提出了一种融合全局和局部特征的多损失优化网络模型，命名为 MLS-Net。该模型包含一个全局和局部特征提取模块（EGL-Module）和一个双向交叉注意力特征融合模块（CAF-Module），其整体架构如图 1 所示。

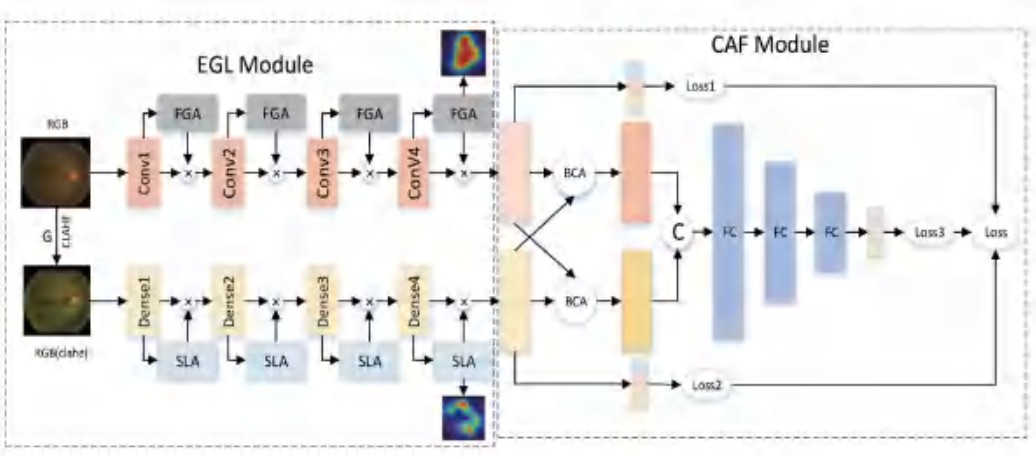


图 1　MLS-Net 总体架构图

Fig. 1　Overall architecture of MLS-Net

在特征提取模块中，原始 RGB 图像用作全局特征提取分支的输入。对 RGB 图像的 G 通道进行 CLAHE 增强后生成一个新的 RGB 图像，该图像用作局部特征提取分支的输入。设 $I=(\mathrm{R},\mathrm{G},\mathrm{B})$ 表示原始 RGB 图像，$I'=(\mathrm{R},\mathrm{G}_{\mathrm{CLAHE}},\mathrm{B})$ 表示增强后的 RGB 图像。全局特征提取分支中引入了频域全局注意力机制（FGA），而局部特征提取分支中则引入了结构局部注意力机制（SLA）。

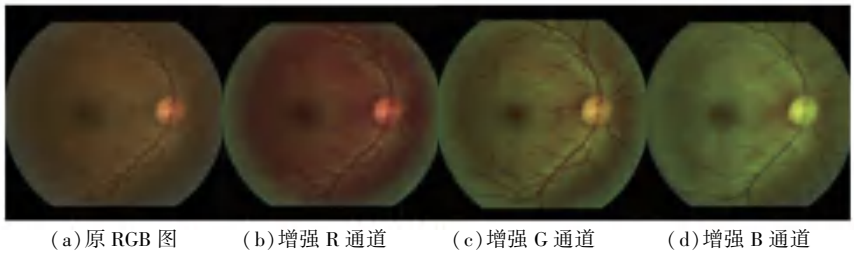
在特征融合阶段，本文使用双向交叉注意力（BCA）机制来融合特征提取模块提取的全局和局部特征。将每个分支融合后的特征进行拼接，然后通过若干全连接层进行降维处理，最后进行最终的分类。MLS-Net 采用了多路径加权损失策略，以有效平衡单分支学习和特征融合的重要性。

1.2　G 通道增强的 RGB 图像

为了清晰突出视网膜图像中的关键结构特征并增强局部特征的提取能力，本文对原始 RGB 图像进行了增强处理。与其他通道相比，G 通道在血管结构、视盘和黄斑区表现出了更高的对比度。对不同通道进行 CLAHE 增强得到增强后的 RGB 图像，如图 2 所示。

G 通道的 CLAHE 增强流程如图 3 所示。首先，从尺寸为 $H\times W$ 的 RGB 图像中提取 G 通道，并将其划分为 $m\times n$（超参数，通常取 8×8 或 16×16）的网格。每个网格独立处理，确保增强效果适应局部区域。网格数量计算如下式：

$$N_{\mathrm{tiles}}=\frac{H}{m}\times\frac{W}{n} \tag{1}$$



（a）原 RGB 图　（b）增强 R 通道　（c）增强 G 通道　（d）增强 B 通道

图 2　不同通道 CLAHE 增强后的 RGB 图像

Fig. 2　RGB image after CLAHE enhancement on different channels

对于每个网格，计算其灰度直方图 $H(g)$，其中灰度级 g 定义为 $g\in[0,L-1]$（L 通常设为 256）。$H(g)$ 的计算如下式：

$$H(g)=\sum_{(i,j)\in\mathrm{tiles}}\delta(I(i,j)=g) \tag{2}$$

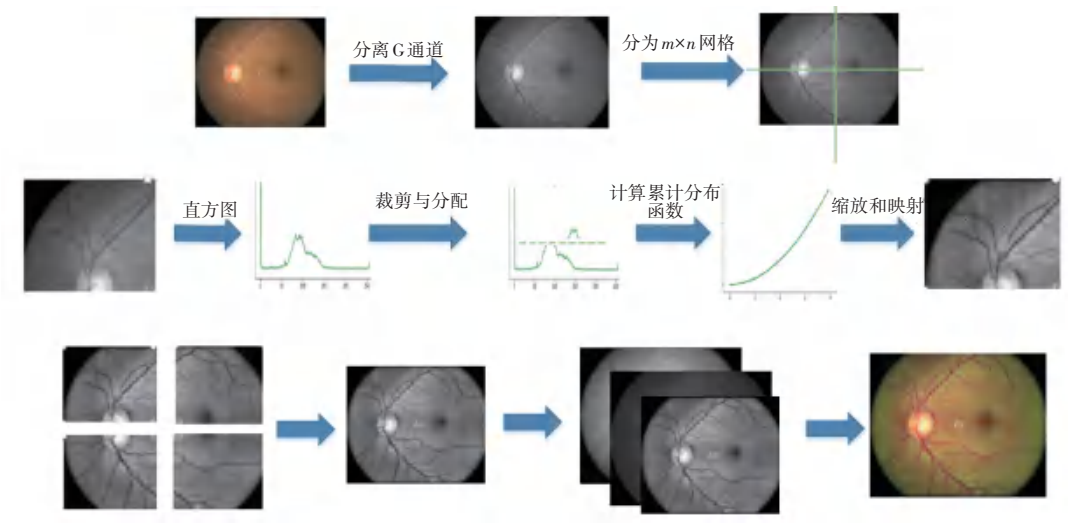


图 3 G 通道的增强过程

Fig. 3 CLAHE enhancement applied to the G channel

此处，$l(i,j)$ 表示图像像素值，$\delta(*)$ 为指示函数(当条件满足时取值为 1，否则为 0)。该方法可获取各网格的灰度分布，为后续图像增强提供必要数据基础。为抑制对比度过度增强，进行直方图裁剪。裁剪阈值记为 T，计算如下式：

$$T = \text{clipLimit} = C \times \frac{m \times n}{L} \tag{3}$$

其中，C 为对比度限制系数(超参数，一般取值范围为 2 ~ 4)，$m \times n$ 表示每个图像块的总像素数。若某灰度级的像素频数 $H(g)$ 超过阈值 T，则执行裁剪操作。裁剪运算规则如下式：

$$H'(g) = \begin{cases} H(g), & \text{if } H(g) \leqslant T \\ T, & \text{if } H(g) > T \end{cases} \tag{4}$$

为维持总像素数不变，将裁剪的像素频数均匀分配至所有灰度级。修正后的直方图频数 $H''(g)$ 计算公式如下式：

$$H''(g) = H'(g) + \frac{\sum_{g=0}^{L-1} \max(0, H(g) - T)}{L} \tag{5}$$

基于裁剪后的直方图 $H''(g)$，通过下式计算累积分布函数 $C(g)$ 以实现灰度值映射：

$$C(g) = \sum_{k=0}^{g} H''(K) \tag{6}$$

计算所得 $C(g)$ 需进行归一化处理。通过下式将每个网格内像素的灰度值 g 映射至目标灰度范围(通常为[0,255])，最终得到增强像素值 g'：

$$g' = \text{round}\left(\frac{C(g) - C_{\min}}{C_{\max} - C_{\min}}\right) \times (L - 1) \tag{7}$$

其中，$C_{\min}$ 表示最小非零累积频数，$C_{\max}$ 为最大累积频数。

为确保增强图像的过渡平滑性，在相邻网格边界处应用双线性插值。将增强后的 G 通道与 R、B 通道合成后，最终获得增强的 RGB 图像。

1.3 全局特征提取分支

如图 4 所示，为精准捕捉视网膜图像的全局质量特征，本文设计了专用全局特征提取分支，采用 ConvNeXt[18] 作为骨干网络，ConvNeXt 融合了 Transformer[19] 的设计理念，具备更大感受野和层归一化，在增强全局信息捕捉能力的同时保持卷积计算的高效性，尤其适用于中小规模数据集上的视网膜图像质量评估任务。

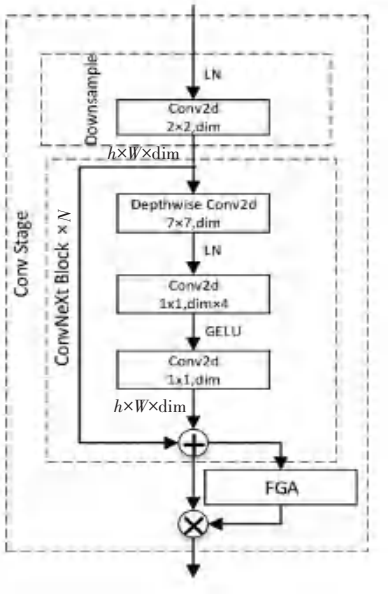


图 4 全局特征提取分支主体结构

Fig. 4 Main structure of the global feature extraction branch

该分支由 4 组 Conv Stage 构成，每个 Stage 包含

一个下采样层和若干 ConvNeXtBlock 模块。为进一步强化全局特征提取，本文提出频域引导的全局注意力机制（FGA），通过对每个通道应用二维傅里叶变换，提取频谱幅值并计算中心低频区域能量，以此刻画通道重要性[20]。由于视网膜图像的整体质量主要受光照、模糊程度和对比度等全局因素影响，这些特征在频域中集中于低频分量。该机制能够从全局尺度揭示图像的对比度与纹理能量分布，基于低频能量生成的通道权重可自适应增强关键特征并抑制高频噪声，从而精准感知图像质量。该设计在提升特征表达能力的同时保障了计算效率，为后续质量评估提供了可靠的全局信息基础。具体结构如图 5 所示。

设输入图像为 I，全局特征提取流程如下：

F_i 表示第 i 阶段的输出特征，F_0 初始化为 I，F_i 的计算如下式：

$$F_i = \mathrm{FGA}(\mathrm{ConvNeXtBlock}_i(\mathrm{Down}_i(F_{i-1}))),\quad i = 1,2,3,\cdots,N \tag{8}$$

计算所得 F_N 需经全局平均池化（GAP）与层归一化（LN）处理，获得最终全局特征 F_{global}：

$$F_{\mathrm{global}} = \mathrm{LN}(\mathrm{GAP}(F_N)) \tag{9}$$

频域引导全局注意力的具体结构如图 5 所示。首先，对输入特征图 $x \in R^{B\times C\times H\times W}$ 中的每个通道独立执行二维傅里叶变换（fft），得到复数频谱 $F(u,v)$，计算公式如下：

$$F(u,v) = \sum_{h=0}^{H-1}\sum_{w=0}^{W-1} x(h,w)\cdot e^{-j2\pi\left(\frac{uh}{H}+\frac{vw}{W}\right)} \tag{10}$$

其中，u 和 v 分别表示水平与垂直频率索引。通过计算幅值谱，得到频率分布：

$$M(u,v) = |F(u,v)| \tag{11}$$

接着，对幅值谱进行中心化处理（fftshift），并在频谱中心区域截取 $k\times k$ 的低频窗口。窗口大小由超参数 if_ratio（超参数，通常设置为 0.125）控制，其计算公式如下式：

$$k = \max(1,\lfloor \min(H,W)\cdot lf\_ration \rfloor) \tag{12}$$

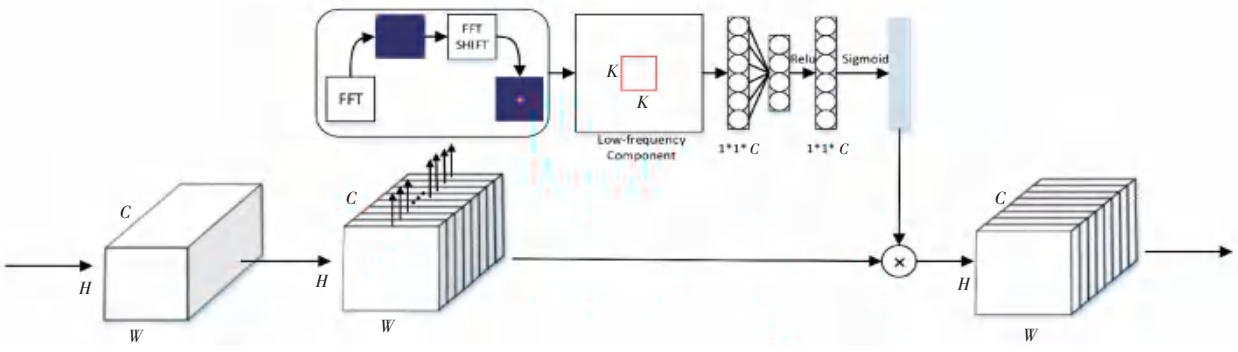


图 5　频域引导全局注意力

Fig. 5　Frequency-domain Guided Global Attention

在该低频窗口内，计算每个通道的低频能量 E_C，公式如下式：

$$E_C = \sum_{i=0}^{k-1}\sum_{j=0}^{k-1} M_c(i,j) \tag{13}$$

为减弱不同样本间能量尺度差异，进行归一化处理。最后将归一化能量向量通过两个全连接层生成通道权重，对输入特征进行加权：

$$y = x\cdot\sigma(FC_2(FC_1(\tilde{\boldsymbol{E}}_c))) \tag{14}$$

其中，$\tilde{\boldsymbol{E}}_c$ 表示归一化处理后的能量向量，σ 为 Sigmoid 函数。

1.4　局部特征提取分支

在视网膜图像质量评估中，黄斑中心区、血管结构和视盘等局部特征至关重要。为此，本文设计了基于 DenseNet 的局部特征提取分支。相较于 ConvNeXt，DenseNet 采用较小的卷积核和感受野，更擅长捕捉细粒度纹理特征。其密集连接机制可将浅层特征传递至所有后续层，从而保留边缘和纹理等局部细节信息。

该分支包含 4 个 Dense Stage，其中前 3 个由 DenseBlock 与 Transition 层组成，最后一层仅包含 DenseBlock。为进一步强化局部特征提取能力，本文引入结构引导的局部注意力机制（SLA）。SLA 通过可学习梯度算子提取血管与边缘信息，并利用高斯建模引入视盘和黄斑区域先验，同时设置可学习偏移以适应个体差异。融合后的结构特征生成空间注意力图，对局部特征进行自适应加权。不同于依赖精确分割的硬约束，SLA 以软引导方式通过端到端学习选择性利用结构提示信息，从而显著增强局部结构表征能力。Dense Stage 结构如图 6 所示。

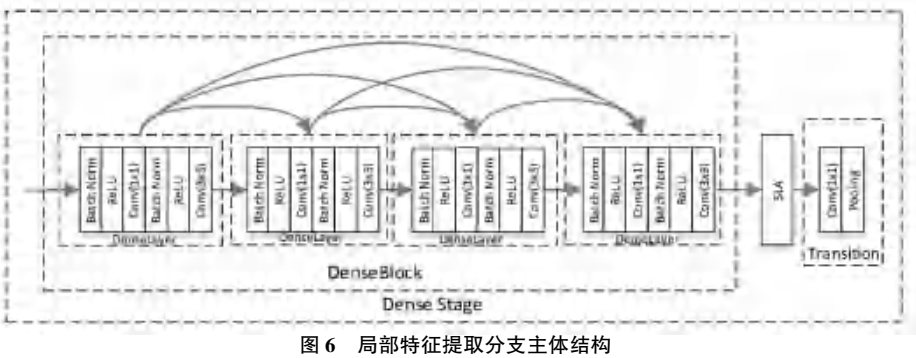


图 6 局部特征提取分支主体结构

Fig. 6 Main structure of the local feature extraction branch

该分支的输入为增强后的 G 通道 RGB 图像 I'，局部特征 F_i 的计算流程如下：

$$F_0 = \mathrm{Stem}(I') \tag{15}$$

$$F_i = \mathrm{SLA}(\mathrm{DenseBlock}_i(F_{i-1})), i = 1,2,\cdots,N \tag{16}$$

$$F_{i+1} = \mathrm{Transition}_i(F_i), i = 1,2,\cdots,N-1 \tag{17}$$

其中 Stem(·)表示局部特征提取分支的初始卷积与池化层组。对计算所得 F_N 进行批归一化(BN)与全局平均池化(GAP)后，获得最终局部特征 F_{local}：

$$F_{\mathrm{local}} = \mathrm{GAP}(\mathrm{BN}(F_N)) \tag{18}$$

结构引导局部注意力的具体结构如图 7 所示。

假设输入特征图 $x \in R^{B\times C\times H\times W}$。为了提取血管等局部结构特征，首先对输入特征图 x 分别应用 Sobel X 和 Sobel Y 核：

$$K_X = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}, K_Y = \begin{pmatrix} -1 & 2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix} \tag{19}$$

通过深度卷积得到水平和垂直梯度：

$$F_x = \mathrm{DWConv}(X;K_x), F_y = \mathrm{DWConv}(X;K_y) \tag{20}$$

梯度幅值计算公式如下：

$$F_{\mathrm{mag}} = \sqrt{F_x^{\,2} + F_y^{\,2} + \in}, \in = 10^{-6} \tag{21}$$

然后通过卷积降维，得到血管结构特征：

$$F_{\mathrm{struct}} = \mathrm{Conv}(F_{\mathrm{mag}}) \tag{22}$$

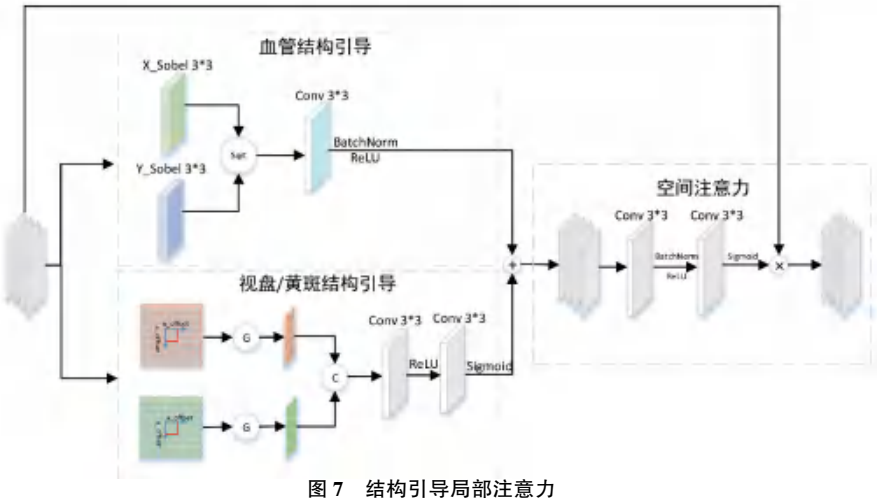


图 7 结构引导局部注意力

Fig. 7 Structure-guided local attention

为了突出关键区域(黄斑和视盘)，通过可学习高斯热力图引导特征关注。初始中心坐标定义为：

$$C_{\mathrm{mac}} = \left(\frac{H}{2}, \frac{W}{2}\right) + \Delta_{\mathrm{mac}}, \; C_{\mathrm{disc}} = \left(\frac{H}{4}, \frac{W}{4}\right) + \Delta_{\mathrm{disc}} \tag{23}$$

其中，Δ_{mac}、Δ_{disc} 为学习偏移参数。之后对每个像素 (i,j) 生成高斯热力图：

$$M_{\mathrm{macula}}(i,j) = \exp\left(-\frac{(i-y_{\mathrm{mac}})^2 + (j-x_{\mathrm{mac}})^2}{2\sigma_{\mathrm{mac}}^2}\right) \tag{24}$$

$$M_{\mathrm{disc}}(i,j) = \exp\left(-\frac{(i-y_{\mathrm{disc}})^2 + (j-x_{\mathrm{disc}})^2}{2\sigma_{\mathrm{disc}}^2}\right) \tag{25}$$

将两个热力图拼接后，通过卷积生成区域先验特征。

$$F_{\mathrm{region}} = \mathrm{Conv}(\mathrm{Conv}([M_{\mathrm{macula}}, M_{\mathrm{disc}}])) \tag{26}$$

之后将血管结构特征和区域先验特征进行通道融合，再通过卷积生成空间注意力图，最后将空间注意力加权输出特征。

$$y = x \cdot \mathrm{Conv}(\mathrm{Conv}(F_{\mathrm{region}})) \tag{27}$$

1.5　双向交叉注意力特征融合模块

为了充分挖掘全局特征与局部特征的互补性，本文设计了双向交叉注意力特征融合模块（CAF-Module）。该模块通过双向信息交互机制，使全局特征能动态调整以捕获更细粒度的细节，同时让局部特征融合全局上下文，从而强化对整体结构的理解。双向交叉注意力结构如图 8 所示。

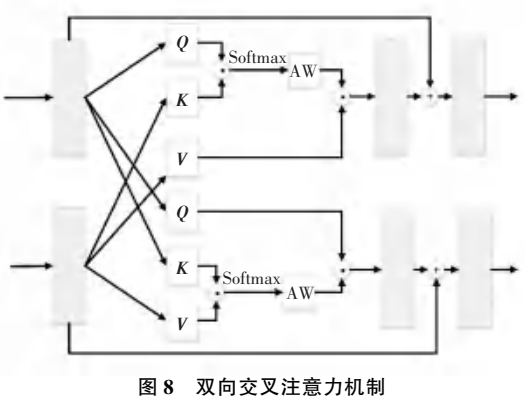


图 8　双向交叉注意力机制

Fig. 8　Bidirectional Cross-Attention Mechanism

该注意力机制的输入由用于分类的最终一维特征向量构成，即分别通过全局分支和局部分支提取的全局特征 F_{global} 与局部特征 F_{local}。在全局特征提取分支中，将 F_{global} 作为查询向量 $\boldsymbol{Q}$，F_{local} 则作为键向量 $\boldsymbol{K}$ 和值向量 $\boldsymbol{V}$，从而更好地捕获局部分支为全局分支提供的互补信息。为保证注意力计算的有效性，本文先将一维特征向量沿通道维度划分为多个 token，从而形成具有序列长度的矩阵特征。随后，通过线性变换生成查询、键和值向量：

1）特征投影。对划分后的序列特征 F_{glocal} 与 F_{local} 应用线性变换，生成查询向量 $\boldsymbol{Q}$、键向量 $\boldsymbol{K}$ 和值向量 $\boldsymbol{V}$。其计算过程如下式所示：

$$\boldsymbol{Q} = \boldsymbol{W}_{\boldsymbol{Q}} F_{\text{global}},\ \boldsymbol{K} = \boldsymbol{W}_{\boldsymbol{K}} F_{\text{local}},\ \boldsymbol{V} = \boldsymbol{W}_{\boldsymbol{V}} F_{\text{local}} \tag{28}$$

其中，$\boldsymbol{W}_{\boldsymbol{Q}}, \boldsymbol{W}_{\boldsymbol{K}}, \boldsymbol{W}_{\boldsymbol{V}}$ 为可学习的权重矩阵。

2）相似度计算。通过式（29）计算查询向量 $\boldsymbol{Q}$ 与键向量 $\boldsymbol{K}$ 的相似度，以捕捉全局特征与局部特征间的相关性。

$$AW = \text{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\text{T}}}{\sqrt{d_k}}\right) \tag{29}$$

其中 d_k 表示键向量 $\boldsymbol{K}$ 的维度，缩放因子用于防止数值溢出。

3）加权求和与拼接。获得注意力权重 AW 后，按下式对值向量 $\boldsymbol{V}$ 进行加权求和，得到特征向量 $\boldsymbol{F}_{\text{ca}}$：

$$\boldsymbol{F}_{\text{ca}} = AW \cdot \boldsymbol{V} \tag{30}$$

4）为稳定训练并保留原始信息，将双向交叉注意力输出的 $\boldsymbol{F}_{\text{ca}}$ 与原始全局特征 F_{global} 通过残差连接融合：

$$F'_{\text{global}} = \boldsymbol{F}_{\text{ca}} + F_{\text{global}} \tag{31}$$

对于局部分支，采用相同处理流程，F_{local} 作为查询向量 $\boldsymbol{Q}$，F_{global} 同时作为键向量 $\boldsymbol{K}$ 和值向量 $\boldsymbol{V}$，最终输出更新后的局部特征 F'_{local}。

该双向交叉注意力机制充分挖掘了全局与局部特征的互补性，使两类特征在不同层级相互增强。将输出的 F'_{glocal} 和 F'_{local} 拼接形成融合特征 $F_{\text{fusion.}}$，随后输入若干全连接层进行高阶信息提取与非线性特征映射，最终用于分类任务。

1.6　多损失优化策略

为充分发挥各分支特性并增强特征提取能力，本文采用三路径加权损失策略，分别为全局、局部和融合分支提供独立监督。该设计使全局分支专注整体信息，局部分支聚焦细节特征，融合分支则深度整合二者的交互信息，从而在避免相互干扰的同时实现更优的特征学习。各分支输出的特征分别输入分类器，计算得到全局损失值与局部损失值。同时，两分支特征经双向交叉注意力模块交互并拼接，再通过全连接层降维后输入分类器，生成融合损失值。

2　实验

2.1　数据集

本文在两个代表性视网膜图像质量数据集上验证了所提 MLS-Net 的有效性：Eye-Q[11] 与 RIQA-RFMiD[21]。

Eye-Q 数据集包含从 EyePACS 数据集新标注的约 3 万张图像。其中 12 543 张用于训练，按质量分为 3 级："好"（8 347 张）、"可接受"（1 876 张）、"差"（2 320 张）；测试集含 16 249 张图像，三类分布为："好"（8 470 张）、"可接受"（4 559 张）、"差"（3 220 张）。

RIQA-RFMiD 数据集由 Fu 等联合湖南省常德市第一人民医院专家基于 RFMiD 数据集构建，用于评估视网膜图像质量。该数据集包含 1 920 张训练图像与 640 张验证图像。其中 2 239 张标注为"好"，194 张为"可接受"，其余 127 张为"差"。

2.2　实验环境

所有实验均在一台配备 NVIDIA RTX 3090 GPU（24 GB 显存）、Intel Xeon Gold CPU 和 512 GB 内存的工作站上完成，运行环境为 Ubuntu 20.04 操作系统与 Python 3.8。模型实现基于 PyTorch 深度学习框架。在训练过程中，本文采用渐进式优化策略：以 AdamW[22] 作为优化器，并使用交叉熵损失函数作为监督信号。初始学习率设为 1e-4，并结合基于验证

集损失动态调整学习率的 ReduceLROnPlateau 机制[23]。实验配置的批大小为 16,完整训练共进行 100 个周期,同时使用早停策略来防止过拟合。

3 结果

3.1 统计分析

表 1 展示了本文提出的 MLS-Net 在 EyeQ 数据集上的视网膜图像质量评估(RIQA)结果,并与 7 种已有方法进行对比。表 1 中的对比结果显示,不同方法之间存在显著的性能差异。传统方法(如 AlexNet+SVM[10])因特征表达能力有限,其在精确率(Precision)、召回率(Recall)和 $F1$ 分数($F1$ - score)上表现较差。相比之下,基于 InceptionV3[24]、DenseNet 和 ResNet[25] 的后续方法取得了更好的性能。例如,引入更深网络结构的 InceptionV3 与 AlexNet+SVM 相比,其精确率和 Kappa 系数分别提升了 6.15%和 8.85%,展现了更强的特征提取能力。MCFNet[11]、Dual-SalStructIQA[12] 和 GuidedNet[14] 均以 DenseNet 为主干网络,但策略各有不同:MCFNet 采用多颜色空间特征融合策略(RGB、HSV、Lab),充分利用不同颜色空间在图像质量评估中的互补信息,其 $F1$ 分数达到 0.860 6,Kappa 系数为 0.895 5;Dual-SalStructIQA 通过显著性结构引导的双分支设计增强模型对关键区域质量的评估能力,Kappa 系数提升至 0.907 6;GuidedNet[14] 引入暗-亮通道先验,无需额外参数即可优化特征提取,最终将精确率和 Kappa 系数分别提升至 0.886 5 和 0.912 7。DualRegNet[15] 以 ResNet 为主干网络并结合标签正则化策略,同样在多指标上实现了性能提升。

本文提出的 MLS-Net 以 DenseNet121 和 ConvNeXt-tiny 作为主干网络,旨在全面提取视网膜图像的全局与局部质量特征,并通过双向交叉注意力模块充分挖掘全局与局部特征间的关联性,增强模型判别能力。如表 1 所示,MLS-Net 在 EyeQ 数据集上取得了最佳性能,其精确率为 0.892 7,召回率为 0.901 0,$F1$ 分数为 0.896 7,Kappa 系数为 0.922 4,显著优于其他方法。

表 2 展示了在私有数据集 RIQA-RFMiD 上的对比结果。可以看到,本文的方法在精确率、召回率、$F1$ 分数和 Kappa 系数四项关键指标上均达到最优,进一步验证了全局与局部特征融合策略的有效性。更为重要的是,在跨域测试中(即由不同采集设备与成像条件构成的 RIQA-RFMiD 数据集),MLS-Net 的 Kappa 系数由 GuidedNet 的 0.686 6 提升至 0.748 9,实现了约 8.3% 的相对提升。这一结果充分说明,适度增加的模型复杂度显著提升了模型的跨域泛化能力。换言之,MLS-Net 通过在结构上引入全局-局部交互建模机制,实现了对成像差异、光照变化及噪声干扰的更强鲁棒性,从而在不同数据分布下均保持稳定且可靠的性能表现。

表 1 在 EyeQ 数据集上与现有方法的对比

Table 1 Comparison with existing methods on the EyeQ dataset

Methods	Backbone	Precision	Recall	$F1$ - score	Kappa	Parameters/M
AlexNet+SVM[10]	AlexNet	0.795 3	0.820 2	0.805 2	0.792 0	56.90
Inception v3[24]	InceptionV3	0.856 8	0.824 0	0.835 0	0.880 5	23.20
MCFNet[11]	DenseNet	0.865 1	0.857 4	0.860 6	0.895 5	28.26
Dual-SalStructIQA[12]	DenseNet	0.874 8	0.872 1	0.872 3	0.907 6	13.92
GuidedNet[14]	DenseNet	0.886 5	0.879 1	0.882 8	0.912 7	6.96
DualRegNet[15]	DenseNet	0.886 8	0.878 6	0.882 0	0.913 8	23.90
本文	DenseNet+ConvNeXt	**0.892 7**	**0.901 0**	**0.896 7**	**0.922 4**	**37.50**

表 2 在私有 RIQA-RFMiD 数据集上与现有方法的对比

Table 2 Comparison with existing methods on the private RIQA-RFMiD dataset

Methods	Backbone	Precision	Recall	$F1$ - score	Kappa	Parameters/M
DenseNet[17]	DenseNet	0.701 1	0.611 5	0.633 7	0.654 7	7.98
ConvNeXt[18]	ConvNeXt	0.612 2	0.686 8	0.614 1	0.697 3	28.60
MCFNet[11]	DenseNet	0.685 0	0.647 7	0.658 6	0.660 4	28.26
GuideNet[14]	DenseNet	0.697 6	0.657 4	0.676 9	0.686 6	6.96
本文	DenseNet+CpnvNeXt	**0.732 8**	**0.717 8**	**0.713 2**	**0.748 9**	**37.50**

综上所述，MLS-Net 在保持较高计算效率的同时，通过合理的结构性复杂度提升，实现了精度与泛化性的双重优化，在实际临床场景下具备较高的应用潜力。

3.2　消融分析

在本研究中，本文提出了一系列创新方法以提升视网膜图像质量评估性能，包括改进的网络架构、G 通道 CLAHE 增强技术、频域全局注意力、结构局部注意力以及多损失机制。为验证这些方法的有效性，本文在 EyeQ 数据集上进行了消融实验，系统评估各组件对模型性能的影响。

如表 3 所示，实验设计以仅包含局部特征提取的基线模型为起点，逐步引入全局特征提取分支、特征融合模块、对 G 通道应用 CLAHE 增强的预处理策略、频域全局注意力、结构局部注意力，最终构建完整的 MLS-Net 模型。结果表明，每个新增组件均以不同程度促进了模型性能的提升。

表 3　在 EyeQ 数据集上的消融实验结果

Table 3　Ablation experiment results on the EyeQ dataset

Model	Precision	Recall	*F*1 - score	Kappa
Local Path	0.843 3	0.883 0	0.860 0	0.893 5
+Global Path	0.885 3	0.877 8	0.880 9	0.907 0
+Fusion Module	0.890 3	0.894 5	0.892 1	0.916 8
+G(Clahe)	0.886 7	0.898 9	0.892 0	0.917 2
FGA	0.887 5	0.900 9	0.893 8	0.919 5
SLA	**0.892 7**	**0.901 0**	**0.896 7**	**0.922 4**

在各项改进中，全局特征提取分支的加入使 Kappa 系数提升 1.35%，精确率(Precision)提高 4.2%，凸显了全局信息在视网膜图像质量评估中的重要性。进一步引入特征融合模块后，Kappa 系数与召回率(Recall)分别增加 0.98%和 1.67%，说明全局与局部特征的互补能显著增强模型的判别能力。随后对局部特征提取分支的输入应用 CLAHE 预处理后，各项指标也得到小幅提升，最后结合频域全局注意力与结构局部注意力后，模型性能达到最优。

为验证全局与局部分支中骨干网络选型的合理性，本文进一步设计了对比实验。具体设置如下：

(1)采用同构骨干网络配置，即全局与局部分支 DenseNet 或均使用 ConvNeXt；

(2)交换骨干网络配置，即在全局分支中采用 DenseNet，局部分支中采用 ConvNeXt。实验结果见表 4。

表 4　在 EyeQ 数据集上不同骨干网络组合的对比结果

Table 4　Comparison of different backbone combinations on the EyeQ dataset

Model	Precision	Recall	*F*1 - score	Kappa
DenseNet+DenseNet	0.876 3	0.866 6	0.870 3	0.897 5
ConvNeXt+ConvNeXt	0.868 1	0.891 6	0.859 9	0.908 6
DenseNet+ConvNeXt	0.869 7	0.877 2	0.872 9	0.902 0
ConvNeXt+DenseNet	**0.892 7**	**0.901 0**	**0.896 7**	**0.922 4**

实验结果表明，相较于本文所提的骨干组合(全局分支使用 ConvNeXt，局部分支使用 DenseNet)，上述两类对比设置的模型性能均出现一定下降。其中，本文所采用的组合在各项评价指标上均表现最优，表明该异构骨干设计能够有效发挥两种网络结构在不同特征层面的优势，具备良好的合理性与互补性。

本文进一步开展了一组对比实验，以验证所提出注意力机制的有效性。在局部分支中，引入了结构引导局部注意力(SLA)，该机制通过利用血管走向及视盘-黄斑等关键结构先验，使网络能够更有针对性地聚焦于与病理相关的区域特征。在全局分支中，提出了频域引导全局注意力(FDA)，其通过频域分析强化图像整体纹理与对比度建模，从而提升全局表征能力。为评估所提出模块的优势，本文将其与典型注意力机制组合(如 SE[26]+SA)进行了对比，实验结果见表 5。结果显示，FDA 与 SLA 的联合应用在 Precision、Recall、*F*1 - score 和 Kappa 四项指标上均取得最优表现，表明该双重引导的注意力机制能够更有效地实现全局与局部特征的互补建模。

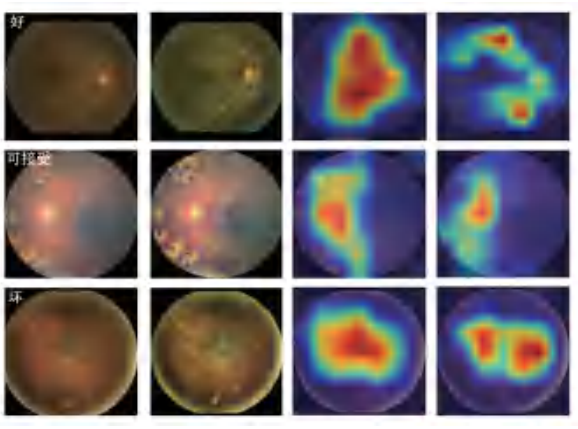
表 5　在 EyeQ 数据集上不同注意力机制的对比结果

Table 5　Comparison of results for different attention mechanisms on the EyeQ dataset

Attention Block	Precision	Recall	*F*1 - Score	Kappa
SE[26]+SA	0.889 8	0.898 8	0.893 8	0.919 1
ECA[27]+SA	0.891 4	0.898 6	0.894 7	0.920 2
CBAM[28]+CBAM	0.889 0	0.897 0	0.892 3	0.916 8
FDA+SLA(本文)	**0.892 7**	**0.901 0**	**0.896 7**	**0.922 4**

3.3　可视化实验

本文采用梯度加权类激活映射(Grad-CAM)[29]对双分支网络的注意力区域进行可视化分析，结果如图 9 所示。该图自上而下展示了“好”、“可接受”和“坏”3 个质量类别的代表性眼底图像样本。每行从左至右依次为：原始 RGB 图像、增强后的 RGB 图像、全局分支的类激活图以及局部分支的类激活图。



(a) RGB 图像 (b) 增强后图像 (c) 全局分支(d) 局部分支

图 9 不同分支的 Grad-CAM 可视化示例

Fig. 9 Examples of Grad - CAM visualizations from different branches

全局分支能广泛关注视盘、黄斑等整体结构,宏观把握图像质量;局部分支则敏锐捕捉视盘边缘、血管分叉等细节特征,负责微观质量评估。二者互补融合,使模型能从宏观与微观双重视角全面评估眼底图像质量,显著提升鲁棒性与判别能力。

4 结束语

本文提出的 MLS-Net 通过全局-局部特征融合与多损失联合优化,有效提升了视网膜图像质量评估的准确性与鲁棒性。该方法不仅在结构设计上兼顾了全局感知与局部细节,还在优化策略上实现了分支间的动态平衡,为自动化眼底图像质量控制提供了新的思路。未来工作将致力于进一步提升模型的轻量化与泛化能力,探索其在临床筛查系统及移动端设备中的应用潜力。

参考文献

[1] ZANG P, HORMEL T T, HWANG T S, et al. Deep-learning-aided diagnosis of diabetic retinopathy, age - related macular degeneration, and glaucoma based on structural and angiographic OCT[J]. Ophthalmology Science, 2023, 3(1): 100245.

[2] SHI C, LEE J, WANG G, et al. Assessment of image quality on color fundus retinal images using the automatic retinal image analysis[J]. Scientific Reports, 2022, 12(1): 10455.

[3] GONÇALVES M B, NAKAYAMA L F, FERRAZ D, et al. Image quality assessment of retinal fundus photographs for diabetic retinopathy in the machine learning era: A review[J]. Eye, 2024, 38(3): 426-433.

[4] ABRAMOVICH O, PIZEM H, VAN EIJGEN J, et al. FundusQ-Net: A regression quality assessment deep learning algorithm for fundus images quality grading [J]. Computer Methods and Programs in Biomedicine, 2023, 239: 107522.

[5] BARTLING H, WANGER P, MARTIN L. Automated quality evaluation of digital fundus photographs[J]. Acta Ophthalmologica, 2009, 87(6): 643-647.

[6] USHER D, DUMSKYJ M, HIMAGA M, et al. Automated detectionof diabetic retinopathy in digital retinal images: A tool for diabetic retinopathy screening[J]. Diabetic Medicine, 2004, 21(1): 84-90.

[7] ZHAO X, WANG L, ZHANG Y, et al. A review of convolutional neural networks in computer vision [J]. Artificial Intelligence Review, 2024, 57(4): 99.

[8] LI L. Convolutional Neural Networks (CNNs)-basedmathod for medical image analysis[C]//Proceedings of the 1st International Conference on Engineering Management, Information Technology and Intelligence, Scitepress-Science and Technology Publications. 2024: 546-552.

[9] MAHAPATRA D, ROY P K, SEDAI S, et al. Retinal image quality classification using saliency maps and CNNs [C]// Proceedings of International Workshop on Machine Learning in Medical Imaging. Cham: Springer , 2016: 172-179.

[10] YU F L, SUN J, LI A, et al. Image quality classification for DR screening using deep learning[C]//Proceedings of the 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Piscataway, NJ: IEEE, 2017: 664-667.

[11] FU H, WANG B, SHEN J, et al. Evaluation of retinal image quality assessment networks in different color - spaces [C]// Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2019: 48-56.

[12] XU Z, ZOU B, LIU Q. A deep retinal image quality assessment network with salient structure priors[J]. Multimedia Tools and Applications, 2023, 82(22): 34005-34028.

[13] 万昊玟. 基于双教师协同类别不平衡眼底图像分割方法[J/OL]. 智能计算机与应用(2025-05-29)[2025-12-01]. DOI: 10. 20169/j. issn. 2095-2163. 25052702.

[14] XU Z, ZOU B, LIU Q. A dark and bright channel prior guided deep network for retinal image quality assessment [J]. Biocybernetics and Biomedical Engineering, 2022, 42(3): 772-783.

[15] GUO T, LIANG Z, GU Y, et al. Learning for retinal image quality assessment with label regularization [J]. Computer Methods and Programs in Biomedicine, 2023, 228: 107238.

[16] 杨旭恺, 王凯祎, 陈璐, 等. 全身 PET/CT 图像下结合健康人先验知识的肿瘤分割[J/OL]. 智能计算机与应用(2025-04-19)[2025-12-01]. DOI: 10. 20169/j. issn. 2095-2163. 25031303.

[17] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 4700-4708.

[18] LIU Z, MAO H, WU C Y, et al. A convnet for the 2020s[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2022: 11976-11986.

[19] 李翔, 张涛, 张哲, 等. Transformer 在计算机视觉领域的研究综述[J]. 计算机工程与应用, 2023, 59(1): 1-15.

[20] ZOU F, LIU Y, CHEN Z, et al. Fourier channel attention powered lightweight network for image segmentation[J]. IEEE Journal of Translational Engineering in Health and Medicine, 2023, 11(4): 252-260.

[21] PANCHAL S, NAIK A, KOKARE M, et al. Retinal Fundus

Multi－Disease Image Dataset（RFMiD）2. 0：A dataset of frequently and rarely identified diseases［J］. Data，2023，8(2)：29.

［22］ZHOU P，XIE X，LIN Z，et al. Towards understanding convergence and generalization of AdamW［J］. IEEE Transactions on Pattern Analysis and Machine Intelligence，2024，46(9)：6486－6493.

［23］AL－KABABJI A，BENSAALI F，DAKUA S P. Scheduling techniques for liver segmentation：Reducelronplateau vs onecyclelr［C］//Proceedings of International Conference on Intelligent Systems and Pattern Recognition. Cham：Springer，2022：204－212.

［24］SAHA S K，XIAO D，KANAGASINGAM Y. A novel method for correcting non－uniform/poor illumination of color fundus photographs［J］. Journal of Digital Imaging，2018，31(4)：553－561.

［25］XU W，FU Y L，ZHU D. ResNet and its application to medical image processing：Research progress and challenges［J］. Computer Methods and Programs in Biomedicine，2023，240：107660.

［26］HU J，SHEN L，SUN G. Squeeze－and－excitation networks［C］//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway，NJ：IEEE，2018：7132－7141.

［27］WANG Q，WU B，ZHU P，et al. ECA－Net：Efficient channel attention for deep convolutional neural networks［C］//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway，NJ：IEEE，2020：11534－11542.

［28］LIU B，ZHAO X，HU H，et al. Detection of esophageal cancer lesions based on CBAM faster R－CNN［J］. Journal of Theory and Practice of Engineering Science，2023，3(12)：36－42.

［29］ZHANG Y，ZHU Y，LIU J，et al. An interpretability optimization method for deep learning networks based on grad－CAM［J］. IEEE Internet of Things Journal，2025，12：3961－3970.