

基于改进 Real-ESRGAN 的超广角眼底图像 超分辨率重建

张艳珠¹, 崔新赞¹, 李婷雪¹, 史铭宇²

(1. 沈阳理工大学 自动化与电气工程学院, 沈阳 110159; 2. 中国医科大学附属第四医院, 沈阳 110005)

摘要: 超广角眼底成像 (ultra-widefield fundus imaging, UWF) 技术在视网膜病变的早期筛查与诊断中具有重要应用价值, 但受制于成像设备性能及光学条件, UWF 图像存在分辨率不足、伪影明显等问题, 影响诊断的准确性。为此, 本文在真实场景增强超分辨率生成对抗网络 (Real-ESRGAN) 框架基础上融入注意力机制和多尺度特征提取模块, 优化生成器网络结构, 并改进判别器损失函数, 提出一种改进 Real-ESRGAN 的超广角眼底图像超分辨率重建方法, 强化对关键细节与纹理信息的重建能力。基于 UWF 数据集进行眼底图像的 2 倍与 4 倍超分辨率重建实验, 并将本文方法与 ESRGAN、CATANet、MSRGAN、SwinIR、RCAN-it 及原始 Real-ESRGAN 等主流方法进行对比, 结果表明, 本文改进方法在峰值信噪比 (PSNR) 和结构相似性指数 (SSIM) 等客观指标上均优于其他方法, 主观评价结果进一步验证了其在图像病灶区域清晰度提升与伪影抑制方面的优势, 本文方法可为临床诊断提供更为可靠的图像支持。

关键词: 超广角眼底成像; Real-ESRGAN; 多尺度特征提取; 注意力机制

中图分类号: TP391.4 文献标志码: A DOI: 10.3969/j.issn.1003-1251.2026.04.002

Super-resolution Reconstruction of Ultra-widefield Fundus Images Using an Enhanced Real-ESRGAN Framework

ZHANG Yanzhu¹, CUI Xinzan¹, LI Tingxue¹, SHI Mingyu²

(1. Shenyang Ligong University, Shenyang 110159, China;

2. The Fourth Affiliated Hospital of China Medical University, Shenyang 110005, China)

Abstract: Ultra-widefield fundus imaging (UWF) is an advanced ophthalmic imaging technology that holds significant value in the early screening and diagnosis of diseases such as diabetic retinopathy and retinal detachment. However, due to limitations in imaging device performance and optical conditions, UWF images often suffer from issues like low resolution and prominent artifacts, which severely hinder diagnostic accuracy. To address these challenges, in this study, attention mechanisms and multi-scale feature extraction module are integrated into the Real-ESRGAN framework, to optimize the generator network structure, and improve the discriminator and loss function, which helps thereby enhancing the ability to reconstruct critical details and texture information. Super-resolution reconstruction experiments were conducted on the UWF dataset at magnification of 2 and 4 times, and a comprehensive comparison was made among prominent methods, including ESRGAN, CATANet, MSRGAN, SwinIR, and RCAN-it. Experimental results demonstrate that the proposed meth-

od outperforms the others in terms of objective metrics such as peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM). Subjective assessments further validate its superiority in improving lesion region clarity and suppressing artifacts, being able to provide more reliable imaging support for clinical diagnosis.

Key words: ultra-widefield fundus imaging; Real-ESRGAN; multi-scale feature extraction; attention mechanism

目前,视网膜成像方式可分为传统眼底摄影(TFP)和超广角眼底摄影(UWFP)两种^[1]。TFP的视野范围为 $30^\circ \sim 60^\circ$,主要用于观察黄斑区及视神经盘,在视网膜周边区域的成像存在一定局限性。相比而言,UWFP可提供的眼底视野范围至少为 200° ,在视网膜疾病的诊断与评估中具有更重要的临床应用价值^[2]。然而,由于成像设备性能及光学环境等因素的影响,超广角眼底成像(ultra-widefield fundus imaging, UWF)存在分辨率较低、伪影明显等问题,限制了其临床应用的准确性^[3]。为解决上述问题,研究者将人工智能技术与眼底成像相结合,提出图像超分辨率重建方法,即通过对低分辨率图像进行重建,生成具有更高分辨率与更丰富细节的图像,从而增强其临床诊断价值^[4-5]。

图像超分辨率重建方法主要包括三类:基于插值^[6]、基于重建^[7]和基于学习^[8]的方法。基于插值和基于重建的方法在恢复高频细节方面存在局限,易产生边缘模糊现象;基于深度学习的方法能够有效重建图像高频细节,显著提升感知质量。近年来,随着深度学习技术的发展,超分辨率重建方法的研究取得了一定进展^[9]。Sun等^[10]将全卷积神经网络(FCN)与传统插值相结合,提高了图像重建速度与质量;Ahmad等^[11]提出了一种渐进式上采样的生成式对抗网络(GAN)模型,改善了图像真实感;Gu等^[12]基于超分辨率生成对抗网络(SRGAN),构建了残差全图注意网络(RW-WAN),增强了对关键区域的关注能力;Mukadam等^[13]将增强型超分辨率生成对抗网络(ESRGAN)与自适应非抽样轮廓变换(ANSCT)相结合,显著提升了图像边缘清晰度;Lu等^[14]在SwinIR网络中引入通道注意力机制,有效融合了局部与全局特征;Wang等^[15]改进了真实场景增强超分辨率生成对抗网络(Real-ESRGAN),设计了一种结合动态通道与空间注意力机制的残差密集模块(RRDB-DCSA),有效减少了图像中的伪影。

上述研究中提出的图像超分辨率重建方法均

在一定程度上提高了低分辨率超广角眼底图像重建质量,但仍存在不同程度的纹理丢失、伪影残留与细节还原度不足等问题。为此,本文在Real-ESRGAN框架基础上,优化其网络结构,提出一种改进的Real-ESRGAN方法。在生成器中引入多尺度残差结构与注意力机制,以带有光谱归一化机制的VGG19作为判别器,采用Charbonnier损失函数,以有效提升UWF图像的细节表现力与诊断可用性。

1 Real-ESRGAN 模型

Real-ESRGAN模型是在ESRGAN基础上提出的改进方法,其生成器结构与ESRGAN一致^[16]。在网络架构方面进行了两项关键性优化:其一,引入高阶降质建模(high-order degradation modeling, HDM),能够更加真实地模拟复杂的实际图像退化过程;其二,判别器采用结合光谱归一化(spectral normalization, SN)机制的U-Net编码-解码结构,可增强网络的判别能力和训练稳定性^[17]。Real-ESRGAN的整体网络结构如图1所示。

降质过程包括四个阶段:模糊处理、下采样、噪声添加与JPEG压缩。传统降质模型通常将高分辨率图像与模糊核进行卷积处理后,进行缩放因子为 r 的下采样,并添加噪声,再经过JPEG图像压缩处理生成低分辨率图像;高阶降质模型则在传统降质过程基础上进行多次迭代,从而更真实地还原复杂退化情形。一阶降质和高阶降质过程的数学表达式分别为

$$\mathbf{x}_c = D(\mathbf{y}) = [(\mathbf{y} * \mathbf{k}) \downarrow_r + \mathbf{n}]_{\text{JPEG}} \quad (1)$$

$$\mathbf{x} = D_i(\mathbf{y}) = (D_i \times \cdots \times D_2 \times D_1)(\mathbf{y}) \quad (2)$$

式中: \mathbf{x}_c 为经一阶降质模型退化后的低分辨率图像; \mathbf{y} 为原始高分辨率图像; D 表示一阶降质过程; \mathbf{k} 为模糊核;*表示卷积操作; \downarrow_r 表示缩放因子为 r 的下采样操作; \mathbf{n} 为噪声项;下标JPEG表示图像压缩处理; \mathbf{x} 为经高阶降质模型退化后的低分辨率图像; D_i 表示 i 阶降质过程。

高分辨率图像经二阶降质处理后,其低分辨

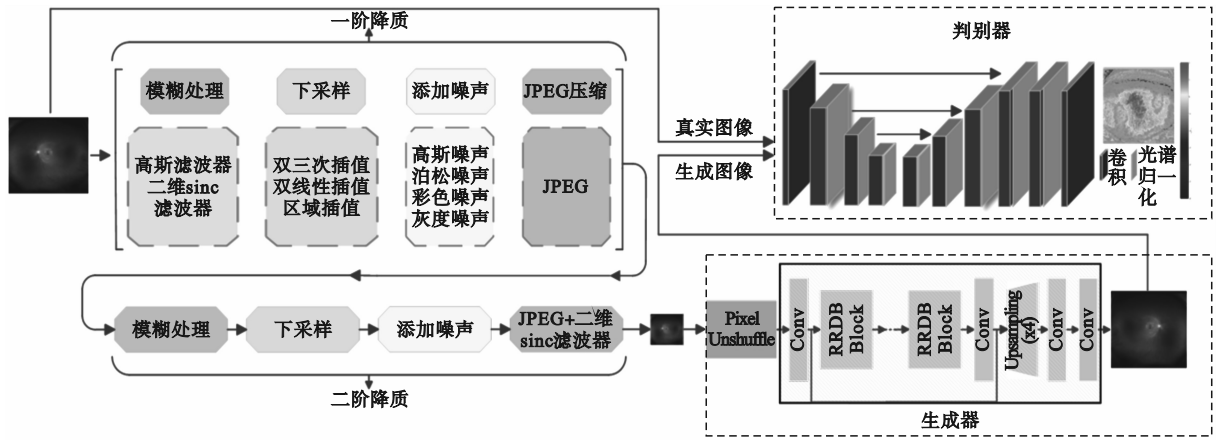


图1 Real-ESRGAN网络结构

Fig. 1 The architecture of Real-ESRGAN network

率图像被输入生成器进行超分辨率重建。Real-ESRGAN的生成器基于多级残差密集块(residual-in-residual dense block, RRDB)构建而成,通过残差连接与密集连接的协同作用,有效提升特征提取能力与信息传递效率。该生成器结构包括Pixel Unshuffle模块、初始卷积层(Conv)、多个RRDB模块(RRDB Block)、特征融合卷积层、上采样层(Upsampling)及输出卷积层,能够在逐步提升图像分辨率的同时,充分保留图像的细节与纹理信息。

判别器接收真实图像与生成图像作为输入,进行真伪判别,采用编码-解码结构的U-Net架构,并引入光谱归一化机制。U-Net通过跳跃连接融合多尺度特征,使判别器兼顾图像的局部细节与全局结构,从而增强对复杂降质图像的判别能力。光谱归一化机制通过约束权重矩阵的最大奇异值,有效抑制梯度爆炸,并缓解梯度消失问题,进一步提升训练稳定性与模型泛化能力。上述改进可为生成器提供更稳定且更具区分性的对抗反馈,从而促进其合成更真实、纹理更丰富的超分辨率图像。

2 Real-ESRGAN模型的改进

2.1 生成器网络结构

为增强UWF图像中病变区域的清晰度与细节纹理,本文对Real-ESRGAN生成器结构进行优化。首先,将原始RRDB模块替换为并行多尺度卷积结构,通过引入 3×3 与 5×5 卷积核分支,兼顾局部细节保持与大感受野信息提取,提升模型对不同尺度特征的建模能力;其次,在RRDB模块后引入SE(squeeze and excitation)通道注意力机制,增强关键通道的特征表达能力。将并行多尺

度卷积、SE注意力机制与RRDB结构相结合,构建MSRRDB(multi-scale squeeze-and-excitation RRDB)模块,以有效提升模型对病灶区域结构与纹理的重建效果。

并行多尺度卷积通过在单一网络层中并行引入具有不同感受野的卷积核(如 3×3 和 5×5)来增强特征提取能力,该结构能够在同一层次上同时捕捉局部与全局信息,实现图像多尺度特征的联合建模。 3×3 卷积核具有较小的感受野,主要用于增强图像的边缘、纹理等细节的恢复能力,从而提升图像的清晰度与锐度; 5×5 卷积核具有较大的感受野,有助于获取更广域的结构信息,有效缓解由局部感知不足引起的边缘模糊与结构扭曲等问题。通过在通道维度上融合不同尺度卷积核提取的特征,不仅能够保留图像的高频细节信息,还能兼顾整体结构的完整性,从而提升模型识别和提取复杂图像信息的能力。通过引入残差缩放系数(β),在特征融合过程中自适应地调节残差结构的权重,从而有效抑制梯度爆炸与过拟合,提升模型训练的稳定性。这种多尺度融合策略可显著提高网络的特征表达丰富性和鲁棒性,对于抑制细节信息丢失、保持图像结构一致性具有重要意义。将并行多尺度卷积与RRDB结构结合组成MRRDB(multi-scale RRDB)模块,其结构如图2所示,图中Dense Block表示密集块,Conv表示卷积,LReLU为激活函数。

为进一步增强模型对关键特征的关注程度,在MRRDB模块基础上引入SE通道注意力机制,构建MSRRDB模块。SE模块通过“压缩-激发”操作自适应地调整各通道特征的重要性,提升模型对病变区域细节和纹理的感知能力,从而增强图像超分辨率重建质量。

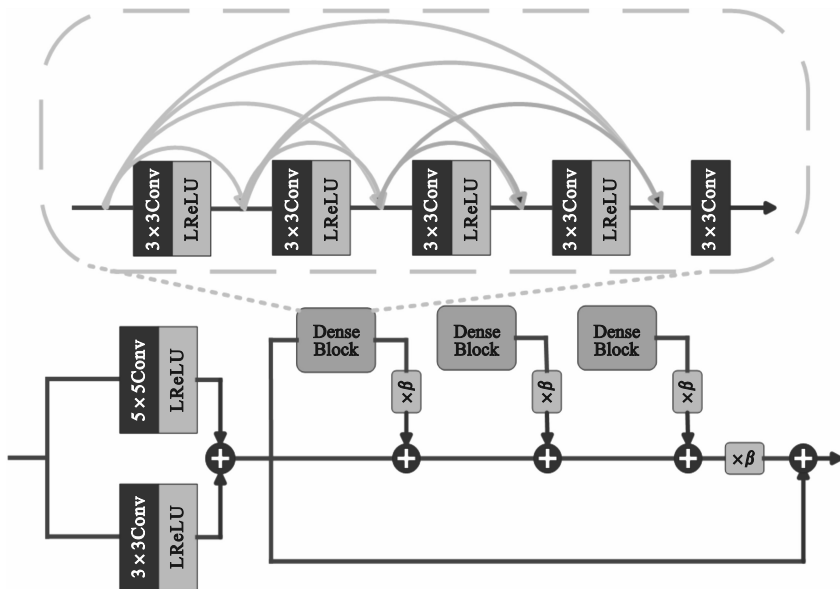


图 2 MRRDB 模块结构

Fig. 2 Architecture of MRRDB module

SE 模块结构如图 3 所示。SE 机制实现过程包括四个步骤:首先,对输入特征图 X ($X \in \mathbf{R}^{C' \times H' \times W'}$, 其中 C' 、 H' 、 W' 分别为输入特征图的通道数、高度和宽度)进行卷积转型(该过程用 F_u 表示)得到新的特征图 U ($U \in \mathbf{R}^{C \times H \times W}$, 其中 C 、 H 、 W 分别为转型后输出特征图的通道数、高度和宽度);然后,将特征图 U 进行全局平均池化或最大池化,压缩为通道描述向量 $[C, 1, 1]$,提取每个通道的全局统计信息(该过程用 $F_{sq}(\cdot)$ 表示);再后,该通道描述向量经过包含两个全连接层的多层感知机(MLP),先降维以减小计算开销,再升维以恢复通道数量,并结合非线性激活函数生成通道权重(该过程用 $F_{ex}(\cdot, w)$ 表示,其中 w 为可学习权重);最后,将通道权重与原始特征图进行逐通道相乘(该过程用 $F_{scale}(\cdot, \cdot)$ 表示),实现特征的加权调控,增强关键信息,抑制无关内容。调整后的特征图 \tilde{X} 用于后续图像重建。

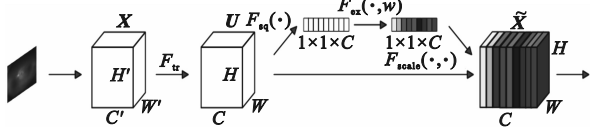


图 3 SE 模块结构

Fig. 3 Structure of SE module

MSRRDB 结构如图 4 所示。首先通过三个 MRRDB 块的密集连接实现多尺度特征融合,然后通过卷积层进行特征整合与非线性变换,最后引入 SE 通道注意力机制以自适应校准特征通道

的重要性权重。

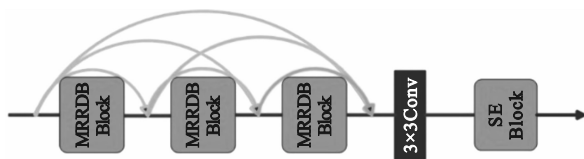


图 4 MSRRDB 模块结构

Fig. 4 Structure of MSRRDB module

假设 MRRDB 输入特征图为 f_0 , 每一层 MRRDB 块的输出特征均作为后续所有层的输入, 则第 i 个 MRRDB 块的输出特征图 f_i ($i=1, 2, 3$) 可表示为

$$f_i = f_{i-1} + F[f_0, f_1, \dots, f_{i-1}] \quad (3)$$

式中 $F(\cdot)$ 表示多尺度特征提取函数。

采用 3×3 和 5×5 卷积核提取的特征图 x_1 和 x_2 分别表示为

$$x_1 = \phi(W_1 * [f_0, f_1, \dots, f_{i-1}]) \quad (4)$$

$$x_2 = \phi(W_2 * [f_0, f_1, \dots, f_{i-1}]) \quad (5)$$

式中: W_1 和 W_2 分别表示 3×3 和 5×5 卷积核的权重参数; $\phi(\cdot)$ 表示 LeakyReLU 激活函数。

将 x_1 和 x_2 在通道维度上进行拼接, 融合来自不同感受野的特征信息。为进一步整合多尺度特征并压缩通道数, 通过一个 3×3 卷积进行特征融合与降维, 得到多尺度融合后的特征图 f , 计算过程表示为

$$f = F([f_0, f_1, \dots, f_{i-1}]) = W_{out} * \phi([x_1, x_2]) \quad (6)$$

式中: W_{out} 为在输出阶段进行通道融合时所使用的卷积核权重参数; $[x_1, x_2]$ 表示通道维度拼接后

的多尺度特征图。

经过卷积层特征整合与非线性变换后,加入 SE 模块,经过 SE 模块后输出的特征图为 f_{SE} ,其计算过程表示为

$$f_{SE} = f \cdot \sigma(W_4 \cdot \delta(W_3 \cdot \text{GlobalPooling}(f))) \quad (7)$$

式中: $\text{GlobalPooling}(\cdot)$ 表示全局平均池化; W_3 为第一个全连接层的权重, $W_3 \in \mathbf{R}^{c \times c}$; $\delta(\cdot)$ 表示 ReLU 激活函数; W_4 为第二个全连接层的权重, $W_4 \in \mathbf{R}^{c \times \frac{c}{r}}$; $\sigma(\cdot)$ 表示 Sigmoid 激活函数。

MSRRDB 通过多尺度特征融合结构与注意力机制的结合,既可保留局部细节,又可提升全局结构建模能力,有效增强模型对医学图像关键区域的感知力。

2.2 判别器网络结构

Real-ESRGAN 中原始的 U-Net 判别器采用对称的编码-解码结构,虽能有效捕获全局信息,但在超分辨率重建任务中存在一定局限性。一方面,跳跃连接虽有助于融合多尺度信息,却削弱了对高频细节的约束,导致纹理丢失;另一方面,反卷积上采样易引入伪影,影响边缘锐利度。同时,U-Net 更侧重全局一致性,对局部精细结构建模能

力不足,易使生成图像过于平滑,降低感知质量。

为解决上述问题,本文将判别器由 U-Net 替换为 VGG19,以增强对高频细节的建模能力。VGG19 由多层 3×3 卷积和 ReLU 激活函数构成,具备较强的特征提取与表达能力,能够逐层捕捉丰富的多尺度信息,其较深的网络深度和逐层缩小的特征图尺寸有助于多尺度特征建模,输出层通过 softmax 激活函数进行判别。VGG19 在纹理、边缘等细节层面的判别效果相比 U-Net 更优,有助于提升对抗训练中对真实图像的区分能力。此外,VGG19 的层级特征可用于计算感知损失,引导生成器在高层语义空间中优化纹理表达,提升重建图像的真实性与视觉质量。由于经过大规模图像数据预训练,VGG19 还具有良好的泛化能力,可为生成器提供更有效的优化方向,使重建结果更加清晰自然。同时,在判别器中引入光谱归一化机制(SN),通过约束权重矩阵的最大奇异值,有效抑制梯度爆炸或消失,提升训练过程的稳定性与模型的泛化能力。引入光谱归一化机制的 VGG19 模块结构如图 5 所示,图中 FC1 和 FC2 分别为第一个和第二个全连接层。

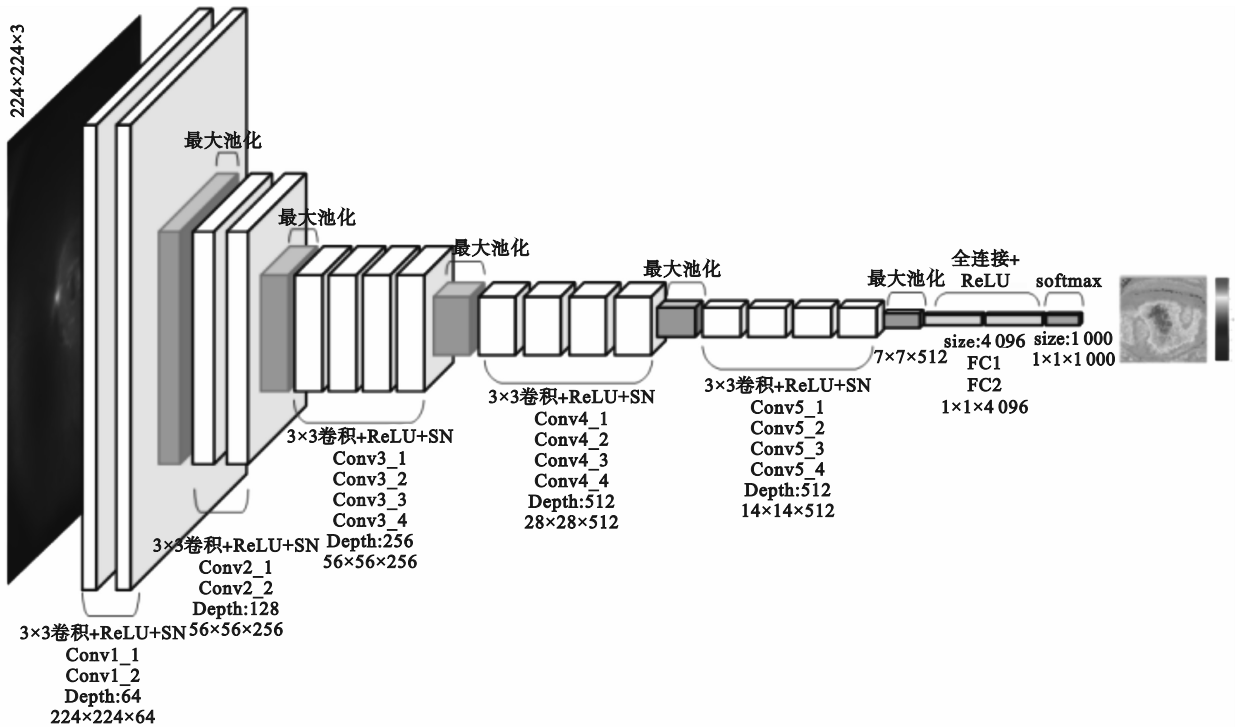


图 5 引入光谱归一化机制的 VGG19 模块结构

Fig. 5 Structure of VGG19 module with spectral normalization

2.3 损失函数

Real-ESRGAN 中原始的 L_1 损失函数广泛用

于图像重建任务,其核心思想是通过计算生成图像与真实图像之间的像素级差异,使生成图像尽

可能接近真实图像。 L_1 损失函数的数学表达式为

$$L_1 = \frac{1}{N} \sum_{i=1}^N |p_i - g_i| \quad (8)$$

式中: p_i 和 g_i 分别为生成图像和真实图像的第 i 个像素值; N 是图像的总像素数。

由于 L_1 损失函数在零点处不可微,当 $p_i = g_i$ 时会出现梯度不连续现象,在梯度下降过程中可能引发梯度更新不稳定,从而影响模型的收敛性;此外,尽管 L_1 损失函数对异常值具有较高的鲁棒性,但在处理包含噪声的图像时,其对误差的敏感性不足,无法有效抑制噪声对训练过程的干扰,从而限制模型在高噪声环境下的表现,影响图像重建质量。因此,在处理超分辨率图像时,采用 L_1 损失函数可能导致生成图像缺乏细节、对比度较低,无法有效保留图像的自然感和真实感。

为克服 L_1 损失函数的局限性,本文引入 Charbonnier 损失。其核心思想是:在计算像素间差异时引入平方根运算和一个极小常数,从而在误差较小时实现梯度的平滑变化。Charbonnier 损失(用 L_c 表示)的数学表达式为

$$L_c = \frac{1}{N} \sum_{i=1}^N \sqrt{(p_i - g_i)^2 + \epsilon^2} \quad (9)$$

式中 ϵ 为一个极小的正数,本文取 1×10^{-3} 。

采用 Charbonnier 损失不仅能够控制全局像素误差,更有效地恢复局部细节,且可克服传统 L_1 损失在高频信息保留方面的不足,生成的高分辨率图像在视觉效果上更为清晰、自然,更接近真实图像的感知效果。

3 实验结果与分析

3.1 实验环境和参数设置

实验环境如表 1 所示。在训练期间使用 Adam 优化器,学习率设置为 0.000 1,一阶动量项设置为 0.9,二阶动量项设置为 0.999。整个训练过

表 1 实验环境

Table 1 Experimental environments

实验环境	相关配置
操作系统	Windows 11
CPU	Inter(R) Core(TM) i7-13700H
GPU	NVIDIA GeForce RTX 4060
深度学习框架	Pytorch2.5.1
开发环境	Python3.10.4
CUDA	12.3

程共进行 40 万次迭代,批次大小为 8,每迭代 5 000 次自动保存一次模型权重参数。

3.2 实验数据集及模型评估指标

本文使用的图像数据由某医院提供,采用激光扫描检眼镜(Daytona 型,Optos 公司)采集获得。共收集七种超广角眼底图像,包括正常眼(NE)及视网膜脱落(RD)、视网膜静脉阻塞(RVO)、视神经病变(ON)、视网膜色素变性(RP)、糖尿病视网膜病变(DR)和病理性近视(PM)六种疾病类型,其中分辨率为 $2\,600 \times 2\,048$ 像素的图像共 814 张,分辨率为 $3\,900 \times 3\,072$ 像素的图像共 306 张,所有图像均由专业眼科医生进行分类标注。

为构建训练样本,对原始的超广角眼底图像进行裁剪,生成尺寸为 650×512 像素的子图像,共获得 5 092 张图像样本。将数据集按照 8:1:1 的比例划分为训练集、验证集和测试集,以确保模型训练的充分性和评估结果的可靠性。

选用双三次插值法对原始图像分别进行 2 倍与 4 倍下采样,生成对应尺度的低分辨率(LR)图像。LR 图像与原始 UWF 图像构成配对数据,用于训练与评估图像超分辨率重建模型。数据集部分样本如图 6 所示,随着图像的放大倍数增加,可明显观察到许多细节丢失。

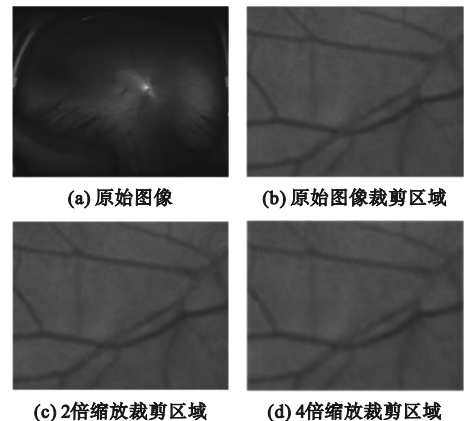


图 6 原始图像与不同放大倍数的低分辨率图像对比

Fig. 6 Comparison between original image and low resolution images of different multiples

此外,为进一步验证本文提出方法在真实应用场景中的泛化能力,采用 Optos 官网发布的超广角眼底图像构建外部测试集。该外部测试集图像由超广角荧光素眼底血管造影机(California 型,Optos 公司)采集获得,共 27 张,分辨率为 $4\,000 \times 4\,000$ 像素,用于模型的独立测试与性能对比分析。该部分数据未参与模型训练过程,有助于更客观、

更全面地评估模型的鲁棒性与实际应用效果。

本文从客观指标与主观评价两方面对模型性能进行评估。客观指标包括结构相似性指数 (SSIM) 和峰值信噪比 (PSNR), 用于量化图像重建质量; 主观评价通过人工观察图像的清晰度与细节保留程度, 综合判断模型的实际表现。

3.3 消融实验

为验证本文提出改进策略对超广角眼底图像超分辨率重建的效果, 设计消融实验, 系统评估生成器、判别器及损失函数三个模块对模型整体性能的影响, 其中缩放因子为 2 的图像超分辨率重建消融实验结果如表 2 所示。

表 2 消融实验结果

Table 2 Ablation experiment results

生成器	判别器	损失函数	退化方式	PSNR/dB	SSIM	训练时间/h
RRDB	U-Net	L_1	双三次插值法	39.67	0.949 2	15.5
RRDB	VGG19	L_1	双三次插值法	39.52	0.947 8	17.0
RRDB	U-Net	L_c	双三次插值法	39.71	0.951 0	16.3
RRDB	VGG19	L_c	双三次插值法	39.58	0.948 6	17.8
MSRRDB	U-Net	L_1	双三次插值法	39.84	0.952 7	18.2
MSRRDB	VGG19	L_1	双三次插值法	40.03	0.953 2	19.6
MSRRDB	U-Net	L_c	双三次插值法	39.75	0.950 4	18.9
MSRRDB	VGG19	L_c	双三次插值法	40.26	0.957 2	19.8
MSRRDB	VGG19	L_c	HDM	40.43	0.958 1	23.5
MRRDB	VGG19	L_c	双三次插值法	40.18	0.954 6	19.4
MRRDB	VGG19	L_c	HDM	40.52	0.959 0	22.8

由表 2 可见: 在保持 RRDB 生成器不变的情况下, 仅将判别器由 U-Net 替换为 VGG19 导致重建性能下降, 说明判别能力过强可能破坏生成器与判别器之间的训练平衡, 从而抑制生成器的有效学习; 保持 RRDB 生成器不变的情况下, 仅将损失函数由传统的 L_1 损失替换为 Charbonnier 损失, 在不同模型配置下均带来细微但稳定的性能提升, 表现为 PSNR 和 SSIM 更高; 与采用 RRDB 生成器相比, 采用改进的多尺度生成器 MSRRDB 后, 模型整体性能显著提升, 说明生成器结构的优化在图像超分辨率重建任务中起着关键作用。综合比较各模型配置, 采用 MSRRDB 生成器、VGG19 判别器和 Charbonnier 损失的组合取得了最优性能, 表明生成器、判别器与损失函数三者之间需协同设计, 同时发挥结构改进与损失函数优化的联合优势, 避免单一模块性能过强造成训练失衡。

在上述研究基础上, 进一步分析 SE 模块与退化方式之间的交互作用。SE 模块的引入增强了模型对通道特征的建模能力, HDM 退化方式能够更真实地模拟复杂的低质量图像退化过程, 两者结合可显著提升重建图像的质量, 但也伴随着训练时间的大幅增加。相比之下, 采用双三次插值退化方式并结合 SE 模块的模型, 尽管在

PSNR 和 SSIM 指标上略低于采用 HDM 退化方式的模型, 但训练效率显著提升, 重建质量仍保持在较高水平, 表现出较好的性能与效率权衡。上述结果表明: 在不同退化方式下, SE 模块的引入对模型性能均有正向作用; 双三次插值方案在训练时间方面更具优势, 适合训练资源或周期存在一定限制的任务场景。

3.4 可视化分析实验

为增强模型在图像超分辨率重建过程中的可解释性与可视化效果, 本文在原始 Real-ESRGAN 模型及本文改进模型的基础上引入梯度加权类激活映射 (gradient-weighted class activation mapping, Grad-CAM) 机制^[18], 用以分析生成器在重建过程中所关注的图像区域。具体而言: 对于原始 Real-ESRGAN 模型, Grad-CAM 作用于生成器网络的最后一个卷积层, 通过提取其输出特征图的梯度信息, 实现对模型关注区域的可视化; 对于本文改进 Real-ESRGAN 模型, Grad-CAM 作用于最后一个 MSRRDB 模块内部的注意力机制层, 以更精细地刻画模型在高频信息恢复阶段所聚焦的区域。可视化结果直观呈现出模型在不同结构下对图像细节关注程度的差异, 图 7 所示为六类病变图像原图与两种模型下的热力图对比结果。

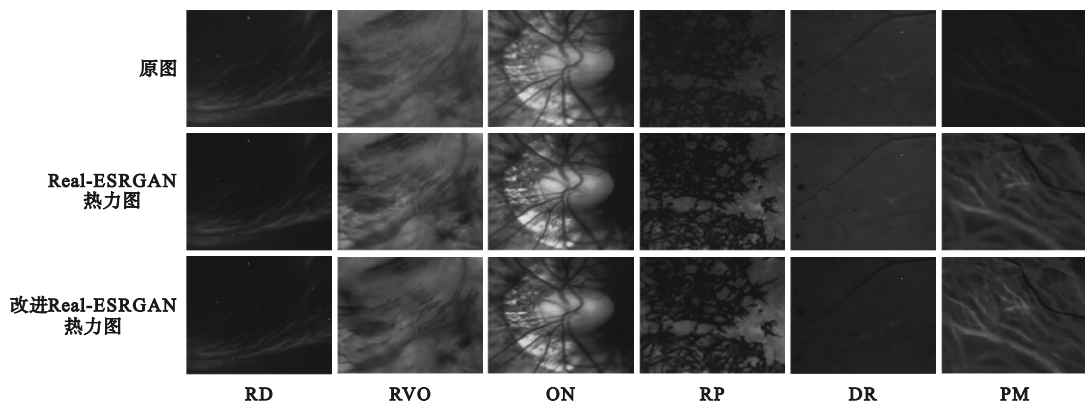


图 7 六类病变图像原图与两种模型下的热力图对比结果

Fig. 7 Comparison of original images of six lesion categories and corresponding heatmaps generated by two different models

为进一步评估改进模型中引入的 SE 模块在聚焦关键解剖结构方面的有效性,本文基于 Grad-CAM 生成的热力图强度分布,提出一种定量分析方法:通过设定强度阈值将热力图划分为高响应区域与低响应区域,并分别计算其面积占比 (high activation area ratio, HAAR) 与平均响应强度 (high activation mean strength, HAMS),用于衡量模型关注区域的集中程度及响应强度。不同模型在多类眼底病图像上的 HAAR 与 HAMS 比较如表 3 所示,改进后的 Real-ESRGAN 模型在多类眼底病图像中的 HAAR 和 HAMS 指标呈现出更具结构差异性的分布特征。RD、ON 与 PM 图像的病灶区

域主要集中于热力图的高亮区域,表现为较高的 HAAR 和 HAMS 值,说明模型能更精准地聚焦于这些类别图像中的关键病灶结构;在 RVO、RP 和 DR 图像中,病变区域多位于响应较弱的暗色区域,相较于原始模型,改进模型在这些区域的响应有所提升,展现出良好的病灶覆盖效果。

综上,本文改进模型在不同病灶类型上展现出更符合结构特征的响应模式,提升了对关键区域的聚焦能力。热力图分析结果验证了 SE 模块在细粒度结构建模和病灶识别中的有效性,并进一步佐证了其在增强超分辨率图像病理相关性方面的重要作用。

表 3 不同模型在多类眼底病图像上的 HAAR 与 HAMS 比较

Table 3 Comparison of HAAR and HAMS between different models on various types of fundus disease images

指标	模型	病变类型					
		RD	RVO	ON	RP	DR	PM
HAAR	Real-ESRGAN	10.15	47.49	30.49	28.23	2.64	17.24
	改进 Real-ESRGAN	14.63	34.97	35.16	27.60	2.40	22.20
HAMS	Real-ESRGAN	0.228 5	0.438 5	0.529 6	0.371 6	0.271 1	0.353 9
	改进 Real-ESRGAN	0.239 2	0.427 4	0.545 8	0.361 9	0.217 9	0.354 4

3.5 对比实验

为验证本文提出改进 Real-ESRGAN 模型的有效性,选取 ESRGAN^[19]、CATANet^[20]、MedSRGAN^[12]、SwinIR^[21]、RCAN-it^[22] 及原始 Real-ESRGAN 作为对比模型,在不同放大倍数下进行超广角眼底图像超分辨率重建实验,计算各模型的 PSNR 与 SSIM,结果如表 4 和表 5 所示。可见,随着放大倍数的提升,各模型的 PSNR 与 SSIM 均呈下降趋势,说明高倍数图像重建任务更具挑战性;本文改进 Real-ESRGAN 模型在各倍数下均取得最高的 PSNR 与 SSIM,显示出其优越的重建性能。

为直观展示本文改进 Real-ESRGAN 模型在重建图像中的视觉效果,在超广角眼底图像测试集中分别对多种超分辨率重建方法进行 2 倍和 4 倍放大实验,结果如图 8 和图 9 所示。为便于对比不同模型在关键病灶细节还原上的性能差异,图 8(a)与图 9(a)中采用黑色标注框表示针对病变部位所执行的超分辨率重建区域。由图 8 和图 9 可以观察到,各超分辨率重建方法(按照图(c)~(i)的顺序)在病变区细节恢复能力上呈现出逐步提升的趋势。ESRGAN 对视网膜整体结构能够实现较为准确的重建,但在细节处仍易出现模糊和

表 4 不同模型超分辨率重建图像 PSNR 比较

Table 4 PSNR comparison of super-resolution reconstructed images by different models

dB

数据集	放大倍数	不同模型的 PSNR						
		ESRGAN	CATANet	MedSRGAN	SwinIR	RCAN-it	Real-ESRGAN	改进 Real-ESRGAN
内部测试集	2	38.73	38.87	38.92	39.86	39.87	39.67	40.26
	4	38.19	38.50	38.43	39.25	39.32	39.14	39.75
外部测试集	2	38.26	38.70	38.58	39.44	39.67	39.45	39.89
	4	38.93	38.32	38.19	38.97	39.22	39.17	39.64

表 5 不同模型超分辨率重建图像 SSIM 比较

Table 5 SSIM comparison of super-resolution reconstructed images by different models

数据集	放大倍数	不同模型的 SSIM						
		ESRGAN	CATANet	MedSRGAN	SwinIR	RCAN-it	Real-ESRGAN	改进 Real-ESRGAN
内部测试集	2	0.933 6	0.937 1	0.955 6	0.956 4	0.954 3	0.949 2	0.957 2
	4	0.932 8	0.933 7	0.953 1	0.954 4	0.950 6	0.946 3	0.955 2
外部测试集	2	0.932 2	0.936 7	0.945 8	0.954 9	0.954 7	0.947 2	0.956 1
	4	0.932 1	0.935 4	0.941 7	0.951 0	0.948 3	0.945 3	0.955 4

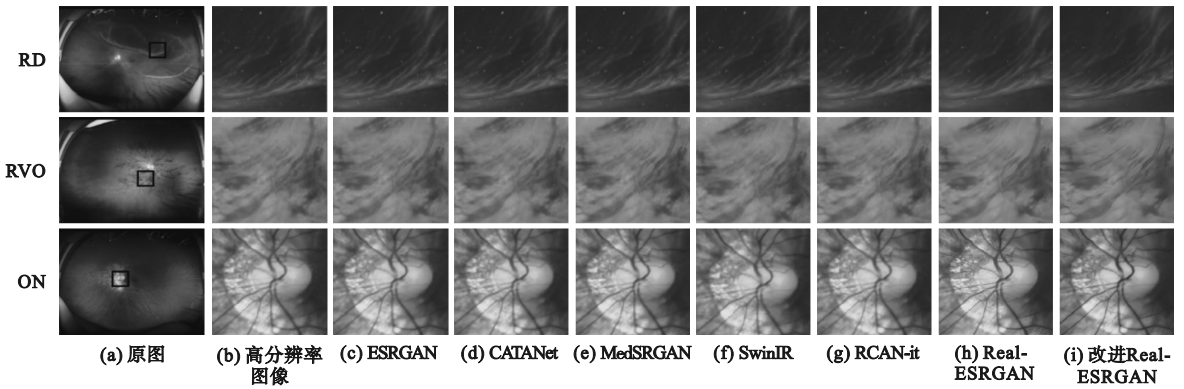


图 8 各模型 2 倍超分辨率重建结果

Fig. 8 Super-resolution reconstructed results of different models at 2 times magnification

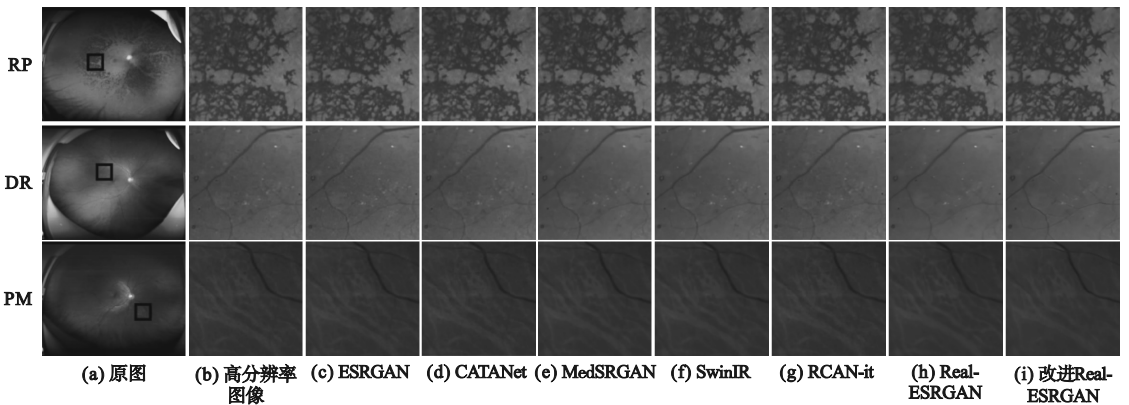


图 9 各模型 4 倍超分辨率重建结果

Fig. 9 Super-resolution reconstructed results of different models at 4 times magnification

棋盘式伪影;轻量化模型 CATANet 引入通道注意力与任务感知机制,能够增强病灶特征并减少噪

声,但在纹理细节上略显模糊;MedSRGAN 通过感知损失强化了重建图像的纹理细节,但噪声和

伪影依旧影响了病灶边界的清晰度;SwinIR 引入自注意力机制,较好还原了高频细节,但在4倍放大时会出现轻微模糊;RCAN-it 通过残差通道注意力机制提升了结构建模能力,在纹理恢复和边缘重构方面优于 SwinIR;Real-ESRGAN 更注重对真实图像退化过程的建模,但仍存在对微小病变特征重建过于平滑、局部纹理丢失的问题;相较而言,本文提出的改进型 Real-ESRGAN 能够更精准地恢复病变区的纹理细节,呈现出最佳的视觉重建效果。

4 结论

在 Real-ESRGAN 基础上提出了改进型超分辨率重建模型:在生成器中引入多尺度残差结构并融合 SE 通道注意力机制,增强了特征表达能力;在判别器中采用基于 VGG19 的架构,并结合光谱归一化机制,提升了对高频纹理细节的判别性能;将原有的 L_1 损失函数替换为更具鲁棒性的 Charbonnier 损失函数,实现了更平滑的梯度传播与更稳定的训练过程。在超广角眼底图像测试集上的对比实验结果表明,本文改进模型在 PSNR、SSIM 等客观指标和主观视觉评价中均取得最优结果,显著提升了重建图像质量与纹理细节还原性能,验证了其在超广角眼底图像超分辨率重建任务中的有效性。

参考文献 (References):

- [1] Yoo T K, Ryu I H, Kim J K, et al. Deep learning can generate traditional retinal fundus photographs using ultra-widefield images via generative adversarial networks [J]. *Computer Methods and Programs in Biomedicine*, 2020, 197: 105761.
- [2] 邓轩,张静琳.超广角眼底成像技术的发展及应用[J].*眼科学报*, 2022, 37(7): 585-594.
Deng X, Zhang J L. Development and application of ultra-wide-field fundus imaging [J]. *Eye Science*, 2022, 37(7): 585-594. (in Chinese)
- [3] Lee K G, Song S J, Lee S, et al. FQ-UWF: unpaired generative image enhancement for fundus quality ultra-widefield retinal images [J]. *Bioengineering*, 2024, 11(6): 568.
- [4] Wei Q L, Gu Z Y, Tan W M, et al. Development and validation of an automatic ultrawide-field fundus imaging enhancement system for facilitating clinical diagnosis: a cross-sectional multicenter study [J]. *Engineering*, 2024, 41: 179-188.
- [5] Nandal P, Pahal S, Khanna A, et al. Super-resolution of medical images using real ESRGAN [J]. *IEEE Access*, 2024, 12: 176155-176170.
- [6] Joshi S, Forjaz A, Han K S, et al. InterpolAI: deep learning-based optical flow interpolation and restoration of biomedical images for improved 3D tissue mapping [J]. *Nature Methods*, 2025, 22(7): 1556-1567.
- [7] Wu J H, Wyman O, Tang Y D, et al. Multi-view 3D reconstruction based on deep learning: a survey and comparison of methods [J]. *Neurocomputing*, 2024, 582: 127553.
- [8] Muhammad Y, Feng J C, Kaleem A, et al. Deep learning-based image reconstruction for different medical imaging modalities [J]. *Computational and Mathematical Methods in Medicine*, 2022, 2022: 8750648.
- [9] El-Shafai W, Mahmoud Mohamed E, Zeghid M, et al. Hybrid single image super-resolution algorithm for medical images [J]. *Computers, Materials & Continua*, 2022, 72(3): 4879-4896.
- [10] Sun N, Li H N. Super resolution reconstruction of images based on interpolation and full convolutional neural network and application in medical fields [J]. *IEEE Access*, 2019, 7: 186470-186479.
- [11] Ahmad W, Ali H, Shah Z, et al. A new generative adversarial network for medical images super resolution [J]. *Scientific Reports*, 2022, 12(1): 9533.
- [12] Gu Y C, Zeng Z T, Chen H B, et al. MedSRGAN: medical images super-resolution using generative adversarial networks [J]. *Multimedia Tools and Applications*, 2020, 79(29): 21815-21840.
- [13] Mukadam S B, Patil H Y. Fusion of ESRGAN, adaptive NSCT, and multi-attention CNN with wavelet transform for histopathological image classification [J]. *IEEE Access*, 2024, 12: 129977-129993.
- [14] Lu W J, Jiang J H, Tian H, et al. Asymmetric convolution Swin Transformer for medical image super-resolution [J]. *Alexandria Engineering Journal*, 2023, 85: 177-184.
- [15] Wang S P, Liu P Y, Chen S J, et al. A method for image super-resolution reconstruction with attention mechanism in generative adversarial networks [C]//*Proceedings of the 2024 13th International Conference on Computing and Pattern Recognition*. Tianjin, China: ACM, 2024: 112-119.
- [16] Wang X T, Xie L B, Dong C, et al. Real-ESRGAN: training real-world blind super-resolution with pure synthetic data [C]//*2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*. Montreal, BC, Canada: IEEE, 2021: 1905-1914.
- [17] Zhu Z W, Lei Y S, Qin Y L, et al. IRE: improved image super-resolution based on real-ESRGAN [J]. *IEEE Access*, 2023, 11: 45334-45348.
- [18] Selvaraju R R, Cogswell M, Das A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization [J]. *International Journal of Computer Vision*, 2020, 128(2): 336-359.
- [19] Wang X T, Yu K, Wu S X, et al. ESRGAN: enhanced super-resolution generative adversarial networks [C]//*Computer Vision European Conference on Computer Vision (ECCV) 2018 Workshops*. Munich, Germany: Springer, 2018: 63-79.
- [20] Liu X, Liu J, Tang J, et al. CATANet: efficient content-aware token aggregation for lightweight image super-resolution [C]//*2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Nashville, TN, USA: IEEE, 2025: 17902-17912.
- [21] Liang J Y, Cao J Z, Sun G L, et al. SwinIR: image restoration using Swin Transformer [C]//*2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*. Montreal, QC, Canada: IEEE, 2021: 1833-1844.
- [22] Lin Z, Garg P, Banerjee A, et al. Revisiting RCAN: improved training for image super-resolution [PP/OL]. *arXiv* (2022-01-27) [2025-05-17]. <https://doi.org/10.48550/arXiv.2201.11279>.