



基于注意力编解码器及多残差网络的逆半色调方法

邬凡^{1,2}, 杨俊², 桂江生¹

(1. 浙江理工大学计算机科学与技术学院, 杭州 310018; 2. 嘉兴学院信息科学与工程学院, 浙江嘉兴 314001)

摘要: 针对当前逆半色调方法恢复的图像存在细节不清晰甚至丢失的问题, 提出了一种基于注意力编解码器及多残差网络(Encoder-decoder with attention and multi-residual network, EDAMRNet)的逆半色调方法。首先, 设计融合注意力机制的编解码器结构, 在其跳跃连接处添加非对称特征融合模块, 以有效提取图像上下文信息; 然后, 构造多残差网络, 捕获并保留图像空间细节信息; 最后, 应用监督注意力模块对图像上下文信息进行加强, 再传递到多残差网络, 以恢复出高质量的连续色调图像。实验结果表明: 该方法与现有最优方法相比, 在 Urban100 和 Manga109 数据集下的峰值信噪比平均值均提高了 0.1 dB, 结构相似性平均值分别提高了 0.0010 和 0.0005。该方法能够在提取图像上下文信息的同时保留图像空间细节信息, 可更好地恢复图像纹理信息, 提高图像清晰度, 为图像逆半色调方法研究提供了一种新的方案。

关键词: 逆半色调; 图像恢复; 注意力机制; 编解码器; 多残差网络; 清晰度

中图分类号: TP391

文献标志码: A

文章编号: 1673-3851(2024)05-0369-09

引文格式: 邬凡, 杨俊, 桂江生. 基于注意力编解码器及多残差网络的逆半色调方法[J]. 浙江理工大学学报(自然科学), 2024, 51(3): 369-377.

Reference Format: WU Fan, YANG Jun, GUI Jiangsheng. An inverse halftoning method based on encoder-decoder with attention and multi-residual network[J]. Journal of Zhejiang Sci-Tech University, 2024, 51(3): 369-377.

An inverse halftoning method based on encoder-decoder with attention and multi-residual network

WU Fan^{1,2}, YANG Jun², GUI Jiangsheng¹

(1. School of Computer Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China; 2. College of Information Science and Engineering, Jiaying University, Jiaying 314001, China)

Abstract: An inverse halftoning method based on encoder-decoder with attention and multi-residual network (EDAMRNet) was proposed to address the issue of unclear or even lost details in image restoration using current reverse halftone methods. Firstly, the encoder-decoder structure with attention was designed, and the asymmetric feature fusion modules were added at its skip connections to effectively capture image context information. Then, the multi-residual network was constructed to capture and retain spatial details of the image. Finally, the supervised attention module was applied to enhance the image context information, which was then transmitted to multi-residual network to restore high-quality continuous-tone images. The experimental results showed that compared to the existing optimal method, the proposed method improved the average peak signal-to-noise ratio by 0.1 dB and the average structural similarity by 0.0010 and 0.0005, respectively on the Urban100 and Manga109 datasets. The method can extract image context information while preserving spatial details, better restoring image texture information and improving image clarity. It provides a new scheme for the study of image inverse

收稿日期: 2023-07-14 网络出版日期: 2023-12-12

基金项目: 浙江省基础公益计划项目(LGG22F020021); 嘉兴市科技计划项目(2021AY10071)

作者简介: 邬凡(1999—), 男, 江西南昌人, 硕士研究生, 主要从事数字图像处理、深度学习等方面的研究。

通信作者: 杨俊, E-mail: juneryoung@zjxu.edu.cn

half-toning.

Key words: inverse half-toning; image restoration; attention mechanism; encoder-decoder; multi-residual network; clarity

0 引言

半色调方法是将连续色调图像转换为二值图像,使得转换后的图像在一定距离的视觉效果与连续色调图像相似;有序抖动和误差扩散是两种有效的半色调方法^[1]。由于半色调图像所需存储空间小,传输速度较快,被广泛应用于印刷和打印等领域。经半色调方法生成的图像丢失了色调和细节等信息,导致图像质量降低,若直接对半色调图像进行缩放、旋转和识别等处理,会产生图像失真、识别精度低等问题。逆半色调方法是半色调方法的逆过程,即将二值图像转换为连续色调图像。然而,通过部分已有信息恢复全部未知信息,是一个不定问题。因此,在逆半色调处理过程中,如何恢复出高质量的连续色调图像是一个具有挑战性的任务。

传统的逆半色调方法包括滤波法^[2]和查找表法^[3]两类。Xiong等^[4]提出了一种基于小波的逆半色调方法,该方法利用小波变换的多分辨率特性,分离出半色调图像的高频信息与低频信息,然后在图像的高频信号中实现去噪功能,低频信号中完成保留边缘信息处理,最后通过小波逆变换使图像从频域转换到空间域,恢复出连续色调图像;该方法虽然比简单的滤波方法能有效提高恢复图像质量,但其计算速度慢,所需计算资源高。Mese等^[5]提出了一种基于查找表的逆半色调方法,首先建立半色调图像值与连续色调图像值之间的查找表,然后通过查表的方法恢复出连续色调图像;这种方法能快速恢复连续色调图像,然而恢复的图像在平缓区域存在噪声。

近年来,随着人工智能技术在不同领域广泛应用,深度学习方法在超分辨率、图像恢复等领域已经取得了非凡的成绩,为克服传统逆半色调方法存在的缺陷提供了新的思路。Hou等^[6]提出了一种基于U-Net的逆半色调方法,该方法的网络结构简单,训练时间短,但恢复的连续色调图像细节不清晰。Xiao等^[7]提出了一种梯度引导网络,通过输入的半色调图像来预测梯度图,将获得的梯度图与半色调图像融合,恢复出连续色调图像。在Xiao等^[7]的基础上,Yuan等^[8]提出了一种梯度残差引导网络,通过引入残差更好地恢复出连续色调图像细节。Xia等^[9]提出了一种渐进残差学习网络,首先使用

内容聚合模块抑制半色调噪声,恢复初步的连续色调图像,然后使用细节增强模块提取图像的细节信息,最后将细节信息与连续色调图像融合,恢复出连续色调图像。Son等^[10]提出了一种结构感知深度神经网络,抑制半色调图像在平缓区域的噪声,恢复出连续色调图像的纹理细节。Shao等^[11]提出了一种融合注意力机制的深度神经网络,使用注意力机制引导网络抑制半色调噪声,并通过多阶段损失函数加速网络优化,恢复出连续色调图像。李梅等^[12]提出了一种融合注意力机制的多尺度残差网络,从不同尺度抑制半色调噪声,提高了连续色调图像在平滑区域和纹理细节上的质量。Li等^[13]提出了一种渐进式多尺度残差网络,融合不同阶段提取的特征信息,提升了连续色调图像的质量。上述逆半色调方法^[6-13]虽取得一定效果,但存在恢复出的连续色调图像不够清晰、空间细节丢失等问题。

为提升恢复的连续色调图像清晰度,保留更多的空间细节信息,本文提出了一种基于注意力编解码器及多残差网络(Encoder-decoder with attention and multi-residual network, EDAMRNet)的逆半色调方法。考虑到注意力机制能够有选择地处理图像特征信息,编解码器的层次化结构能够提取图像上下文信息,本文设计了融合注意力机制的编解码器,对连续色调图像的上下文信息进行提取;由于残差网络能够捕获图像的空间细节信息,本文构造了多残差网络,提取连续色调图像的空间细节信息。在Urban100、Manga109、Set5、Set14和Place365公开数据集上对本文提出的方法与其他方法进行对比,以验证该方法的有效性。本文构建了EDAMRNet模型,为图像逆半色调的网络模型研究提供了新的角度和思路,增加该领域研究的多样性,为学者进一步深入研究逆半色调网络模型提供参考和基础。

1 方法设计

本文提出了一种端到端的基于EDAMRNet的逆半色调方法,在编码阶段使用残差块对下采样图像进行特征提取,提取图像的深层特征;在解码阶段通过特征注意力模块(Feature attention module, FAM)^[14]调整图像特征信息的权重,提升模型对有用特征信息的关注度;在编解码器的跳跃连接处添

加非对称特征融合模块(Asymmetric feature fusion module, AFFM)^[14],融合不同层级输出的特征,使模型提取到丰富的图像上下文信息。然后,使用多残差网络提取并保留图像空间细节信息。最后,采用监督注意力模块(Supervised attention module, SAM)^[15]对图像上下文信息进行加强,再传递到多残差网络,以端到端的方式恢复出连续色调图像。

1.1 网络结构

本文提出的 EDAMRNet 结构示意图如图 1 所示。图像的高、宽和通道数分别为 H 、 W 和 C 。输入的半色调图像 $I \in \mathbf{R}^{H \times W \times 1}$ 经过卷积层和残差块,得到图像浅层特征 $F_s \in \mathbf{R}^{H \times W \times C}$ 。浅层特征计算过程可由式(1)表示:

$$F_s = M_{RB}(J_{3 \times 3}(I)) \quad (1)$$

其中: $J_{3 \times 3}$ 表示卷积核为 3×3 的卷积操作, M_{RB} 表示残差操作。浅层特征输入到编解码器,得到图像

上下文信息 $T \in \mathbf{R}^{H \times W \times C}$ 。图像上下文信息计算过程可由式(2)表示:

$$T = E_{ED}(F_s) \quad (2)$$

其中: E_{ED} 表示编解码操作。图像上下文信息和半色调图像同时输入到 SAM,得到注意力特征 $F_{out} \in \mathbf{R}^{H \times W \times C}$ 。注意力特征计算过程可由式(3)表示:

$$F_{out} = S_{SAM}(T, I) \quad (3)$$

其中: S_{SAM} 表示应用监督注意力机制。注意力特征和浅层特征拼接后,经过残差块和卷积层,得到残差图像 $X \in \mathbf{R}^{H \times W \times 1}$ 。残差图像与半色调图像相加,得到恢复的连续色调图像 \tilde{I} 。上述计算过程可由式(4)~(5)表示:

$$X = J_{3 \times 3}(M_{RB}(P_{concat}(F_{out}, F_s))) \quad (4)$$

$$\tilde{I} = X \oplus I \quad (5)$$

其中: P_{concat} 表示特征拼接操作; \oplus 表示按元素相加。

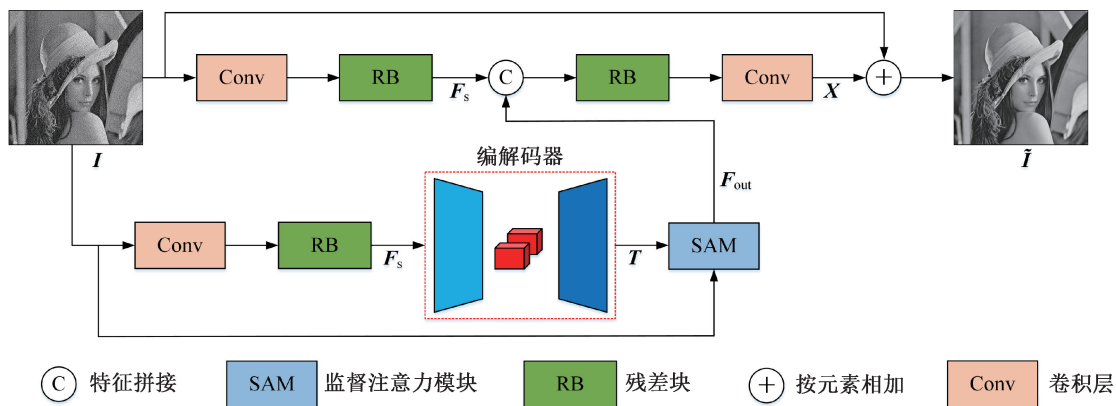


图 1 EDAMRNet 结构示意图

1.2 融合注意力机制的编解码器结构

图像上下文信息是指图像中单个像素与其周围像素之间相互关联的信息。在逆半色调处理过程中,模型通过学习图像中像素之间的关系,从而获得更准确的语义表达。编解码器的层次化结构有利于提取图像上下文信息;通过注意力机制为图像上下文信息中的元素赋予不同权重,可使图像上下文信息被充分利用。因此,本文设计了融合注意力机制的编解码器。首先,在编码阶段使用像素重组的方式进行下采样,以降低图像分辨率,扩大感受野,接着使用残差块对下采样图像进行特征提取,获取图像的深层特征信息;然后,在解码阶段还原图像的浅层特征时,使用反卷积增大图像分辨率,容易产生棋盘伪影,因此采用像素重组的方式进行上采样,上采样图像再经过 FAM,使模型关注有价值的特征信息;最后,不同于直接连接上下采样层,本文使用跳

跃连接,将编解码器中相同维度的层级连接起来,融合图像的深层特征和浅层特征。为进一步提升融合效果,在跳跃连接处添加 AFFM,将编码器中不同层级提取的特征进行融合。融合编码器融合的特征与解码器还原的特征,使特征信息在整个编解码器中流动,有助于提取图像上下文信息。

1.2.1 AFFM

在传统的编解码器结构中,特征信息都是自上而下地流动,信息流动性受限。为加强特征信息在各个层级之间的流动性,在编解码器的跳跃连接处添加 AFFM。AFFM 将不同层级输出的特征信息进行融合,使模型提取丰富的图像上下文信息。同时,通过 AFFM 融合不同层级输出的特征,可有效减少图像信息丢失。AFFM 结构示意图如图 3 所示,其中: B_n^{out} 表示编码器第 n ($n=1, 2, 3$) 层残差块输出的特征, $Resize$ 表示对特征进行放大或缩小。

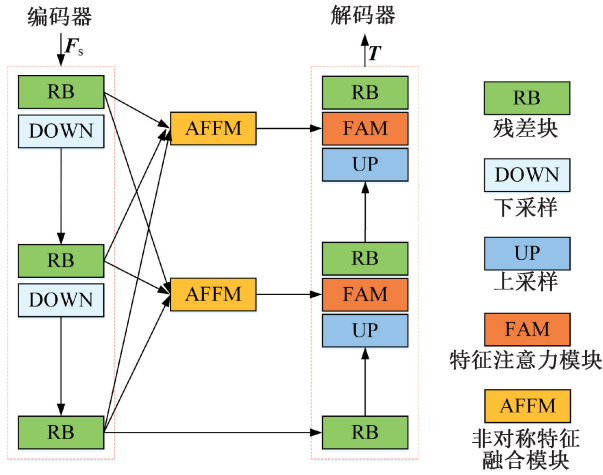


图2 融合注意力机制的编解码器结构示意图

为确保相同大小的特征进行拼接,将编码器中不同层级残差块输出的特征分别进行放大或缩小。然后进行拼接,再经过2个卷积层,得到融合后的特征 A_n^{out} 。上述计算过程可由式(6)—(7)表示:

$$A_1^{\text{out}} = J_{3 \times 3} (J_{1 \times 1} (P_{\text{concat}} (B_1^{\text{out}}, B_{2,u}^{\text{out}}, B_{3,u}^{\text{out}}))) \quad (6)$$

$$A_2^{\text{out}} = J_{3 \times 3} (J_{1 \times 1} (P_{\text{concat}} (B_1^{\text{out}}, B_{2,d}^{\text{out}}, B_{3,u}^{\text{out}}))) \quad (7)$$

其中: $B_{n,u}^{\text{out}}$ 表示编码器中第 n ($n=2,3$) 层残差块输出的放大后的特征; $B_{n,d}^{\text{out}}$ 表示编码器中第 n ($n=1$) 层残差块输出的缩小后的特征; $J_{1 \times 1}$ 表示卷积核为 1×1 的卷积操作; A_n^{out} 表示第 n ($n=1,2$) 个 AFFM 输出的特征。

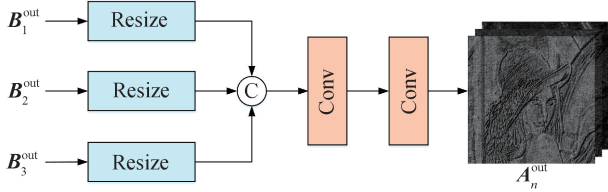


图3 AFFM结构示意图

1.2.2 FAM

通过注意力机制可以增加模型对图像有用特征的关注度。为有效利用图像恢复所需的关键特征信息,在上采样之后添加 FAM。FAM 为图像上下文

信息中的元素分配不同权重,使图像上下文信息被充分利用。FAM 结构示意图如图 4 所示。上采样输出的特征 F_u 与 AFFM 输出的特征 A_n^{out} 进行相乘,然后输入到 3×3 卷积层,再与 A_n^{out} 进行相加,得到有效的特征 F_m 。上述计算过程可由式(8)表示:

$$F_m = J_{3 \times 3} (A_n^{\text{out}} \otimes F_u) \oplus A_n^{\text{out}} \quad (8)$$

其中: \otimes 表示按元素相乘。

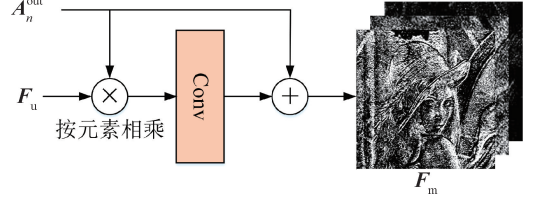


图4 FAM结构示意图

1.3 多残差网络

传统网络直接学习输入图像和目标图像之间的映射关系,但在网络较深时存在梯度消失和信息丢失问题。残差网络通过引入跳跃连接,缓解了梯度消失,使模型更容易地捕捉图像空间细节信息。使用跳跃连接还可以直接传递梯度,优化梯度在网络中的传播,加速网络收敛。

本文通过残差块组成的多残差网络处理半色调图像,提取并保留图像空间细节信息。多残差网络结构示意图如图 5 所示。半色调图像 $I \in \mathbf{R}^{H \times W \times 1}$ 经过卷积层和残差块,得到浅层特征 $F_s \in \mathbf{R}^{H \times W \times C}$ 。浅层特征再经过残差块和卷积层,得到残差图像 $X \in \mathbf{R}^{H \times W \times 1}$ 。为保留丰富的高分辨率图像特征,上述过程不对图像进行上下采样。最后将残差图像和半色调图像相加,得到恢复的连续色调图像 \tilde{I} 。上述计算过程可由式(9)—(11)表示:

$$F_s = M_{\text{RB}} (J_{3 \times 3} (I)) \quad (9)$$

$$X = J_{3 \times 3} (M_{\text{RB}} (F_s)) \quad (10)$$

$$\tilde{I} = X \oplus I \quad (11)$$

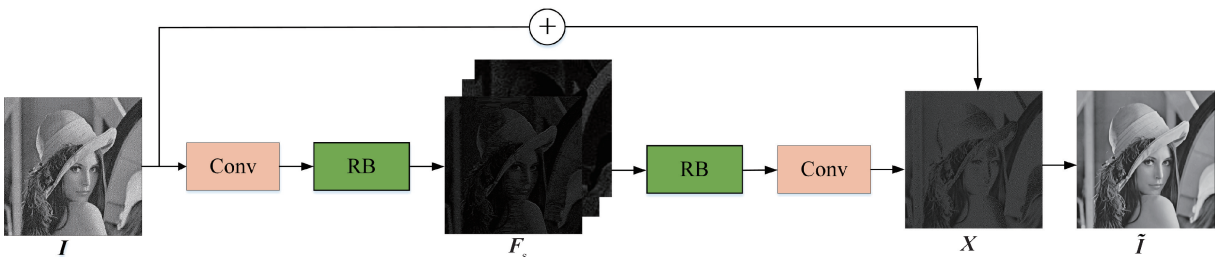


图5 多残差网络结构示意图

1.4 SAM

本文在编解码器和多残差网络之间添加 SAM,

SAM 引入真实图像作为监督信号,抑制信息量较少的特征,选择对图像恢复有用的特征传递到多残差

网络,以更准确地恢复出连续色调图像。SAM 结构示意图如图 6 所示。首先,特征 $F_{in} \in \mathbf{R}^{H \times W \times C}$ 经过 1×1 卷积层生成的图像,与半色调图像 I 相加,得到初步的连续色调图像 $N \in \mathbf{R}^{H \times W \times 1}$ 。然后,图像 N 经过 1×1 卷积层,再通过 Sigmoid 激活函数自适应分配权重,得到特征权重图 $Q \in \mathbf{R}^{H \times W \times C}$ 。最后,将特征权重图 Q 和经过 1×1 卷积层的特征 F_{in} 相乘,

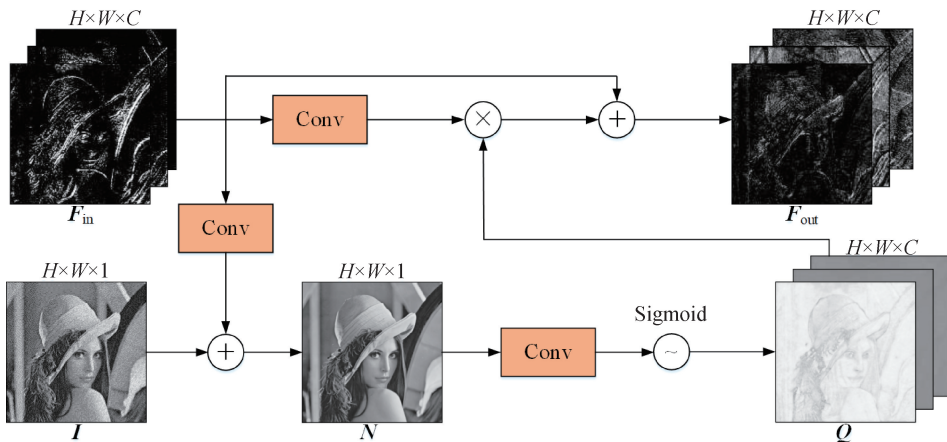


图 6 SAM 结构示意图

1.5 损失函数

损失函数是衡量预测值与真实值之间的差异,数值越小说明网络模型的性能更优。由于 L_1 损失可以防止潜在的梯度爆炸,因此与均方差损失相比较, L_1 损失对异常值更加鲁棒。 L_1 损失提供了真实连续色调图像与恢复的连续色调图像之间的差异度量,计算可由式(15)表示:

$$L_1 = \frac{1}{K} \sum_{k=1}^K \|\tilde{\mathbf{I}} - \mathbf{Y}\|_1 \quad (15)$$

其中: \mathbf{Y} 表示真实图像; k 表示训练集中的第几个样本; K 表示训练集的样本总数。

图像频域包含图像的不同频率分量,通过快速傅里叶变换(Fast Fourier transform, FFT)^[17]可将图像从空间域转换为频域表示,其高频分量对应图像的纹理和空间细节信息。 L_{FFT} 损失提供了真实连续色调图像与恢复的连续色调图像在频域上的差异度量,有助于恢复图像的高频分量。 L_{FFT} 损失计算可由式(16)表示:

$$L_{\text{FFT}} = \frac{1}{K} \sum_{k=1}^K \|V_{\text{FFT}}(\tilde{\mathbf{I}}) - V_{\text{FFT}}(\mathbf{Y})\|_1 \quad (16)$$

其中: V_{FFT} 表示将图像从空间域转换到频域表示。

为综合考虑图像空间域和频域的信息,本文采用的损失函数 L_{total} ,计算如式(17)所示:

$$L_{\text{total}} = L_1 + \lambda L_{\text{FFT}} \quad (17)$$

其中: λ 为超参数,取值为 0.1。

再与特征 F_{in} 相加,得到注意力特征 $F_{out} \in \mathbf{R}^{H \times W \times C}$ 。上述计算过程可由式(12)–(14)表示:

$$\mathbf{N} = J_{1 \times 1}(\mathbf{F}_{in}) \oplus \mathbf{I} \quad (12)$$

$$\mathbf{Q} = \sigma(J_{1 \times 1}(\mathbf{N})) \quad (13)$$

$$\mathbf{F}_{out} = (\mathbf{Q} \otimes (J_{1 \times 1}(\mathbf{F}_{in}))) \oplus \mathbf{F}_{in} \quad (14)$$

其中: σ 表示 Sigmoid 激活函数。

2 结果与讨论

2.1 数据集

本文使用在图像恢复任务中被广泛使用的公开数据集 VOC2012^[18]。该数据集包括 20 个不同类别的图像,随机挑选 13841 张图像作为训练集,3000 张图像作为验证集。训练前使用误差扩散法生成对应的半色调图像,形成训练集和验证集。训练集和验证集被裁剪成大小为 128×128 像素,通道数为 1,每张图像通过随机水平或垂直旋转进行数据增强。本文选择公开数据集 Urban100、Manga109、Set5、Set14 和 Place365 作为测试集。

2.2 实验过程

实验设备和配置如下,硬件使用 Intel(R) Core (TM) i9-10900X CPU @ 3.70 GHz 处理器和 GeForce RTX 3090Ti 显卡,实验环境使用 64 位 Ubuntu 系统、Python 3.8.0、PyTorch 1.11.0 和 CUDA 11.3。在实验中,EDAMRNet 中编解码器和多残差网络中残差块的个数分别设为 4 和 8,提取浅层特征时残差块的个数设为 8。训练过程中采用 AdamW 优化器,批量大小设置为 16,训练轮数约 460 轮,初始学习率为 2×10^{-4} ,使用余弦退火策略优化学习率。

2.3 实验结果和对比分析

为验证本文方法的有效性,选择逆半色调方法

PRL^[9]、MSPRL^[13]和其他基于卷积神经网络的图像恢复方法^[14-15,19-20]进行比较。在这些图像恢复方法中, Lim 等^[19]提出了增强深度超分辨率网络(EDSR),通过移除批量归一化提升了网络性能; Cho 等^[14]提出了单图像去模糊网络(MIMOUNet),采用由粗到细的策略,具有多输入多输出的 U-Net 结构,可快速和高效地实现图像去模糊; Zamir 等^[15]提出了多阶段渐进式图像恢复网络(MPRNet),通过逐步恢复每阶段退化的图像,提升了图像恢复效果; Chen 等^[20]提出了非线性无激活网络(NAFNet),网络结构简单且计算量低,实现了图像去噪和图像去模糊。这些方法可扩展用于逆半色调任务。为准确评价实验结果,以上方法与本文方法使用相同数据集进行训练,本文从客观评价、主观评价和训练性能对恢复的连续色调图像质量进行评估。

2.3.1 客观评价

本文采用峰值信噪比(Peak signal-to-noise

ratio, PSNR)和结构相似性(Structural similarity, SSIM)为量化指标,对不同的逆半色调方法进行客观评估。PSNR 可以量化恢复的连续色调图像与真实连续色调图像之间的差异程度,其值越高,表明图像失真越少,质量越高。SSIM 可以评估恢复的连续色调图像与真实连续色调图像之间的结构相似程度,考虑了图像对比度和结构等因素,其值越高,表明两幅图像越相似。本文方法与其他方法在 5 个基准数据集上的 PSNR 和 SSIM 平均值见表 1,从表中可以看出:本文方法相比目前最优的逆半色调方法 MSPRL^[13], PSNR 平均值提高 0.02~0.10 dB, SSIM 平均值提高 0.0001~0.0010;本文方法与最优的图像恢复方法 MPRNet^[15]相比, PSNR 平均值提高 0.03~0.48 dB, SSIM 平均值提高 0.0006~0.0008。与其他方法相比,本文方法恢复的连续色调图像与真实连续色调图像之间差异最小,并且结构上更加相似。

表 1 不同方法恢复的连续色调图像的 PSNR 平均值和 SSIM 平均值

方法	PSNR/dB					SSIM				
	Urban100	Manga109	Set5	Set14	Place365	Urban100	Manga109	Set5	Set14	Place365
EDSR ^[19]	30.32	34.03	33.51	30.98	30.52	0.9119	0.9475	0.9231	0.8866	0.8965
PRL ^[9]	29.39	32.55	32.06	29.97	29.23	0.9017	0.9365	0.9103	0.8746	0.8840
NAFNet ^[20]	30.50	34.29	33.54	31.13	30.57	0.9159	0.9484	0.9230	0.8840	0.8981
MIMOUNet ^[14]	30.58	34.50	33.63	31.16	30.59	0.9156	0.9495	0.9245	0.8890	0.8983
MPRNet ^[15]	—	34.47	—	—	30.76	—	0.9515	—	—	0.9017
MSPRL ^[13]	31.10	34.85	33.81	31.40	30.77	0.9226	0.9518	0.9264	0.8925	0.9020
本文方法	31.20	34.95	33.84	31.46	30.79	0.9236	0.9523	0.9266	0.8926	0.9023

2.3.2 主观评价

通过本文方法与其他方法恢复的连续色调图像相比较,对图像恢复效果进行主观视觉评价。本文方法与其他方法恢复的连续色调图像 Urban026 (Urban100 数据集)和 ARMS(Manga109 数据集)见图 7 和图 8。图像中的小矩形框表示图像的局部区域,大矩形框表示放大后的局部区域。从图 7 前两行可以看出:在 Urban026 图像中,EDSR^[19]、PRL^[9]、NAFNet^[20]和 MIMOUNet^[14]方法恢复的图像整体上不清晰,纹理和细节较模糊;MSPRL^[13]方法能够恢复图像的部分纹理信息,本文方法能够恢复图像的整体纹理信息,去除半色调噪声更彻底。从图 8 前两行可以看出:在 ARMS 图像中,EDSR^[19]和 PRL^[9]方法恢复效果较差,部分线条之间存在黏连现象;NAFNet^[20]、MIMOUNet^[14]和 MSPRL^[13]方法恢复效果较好,但从大矩形框中可以看出,线条附近均出现伪影且边缘较模糊。通过

观察对比图可以发现,本文方法恢复的连续色调图像具有最优的视觉效果,整体上更加清晰,且不存在伪影,保留更多的空间细节信息。

2.3.3 训练性能

为验证本文网络模型的训练性能,将本文方法与基于 MIMOUNet^[14]、MPRNet^[15]和 NAFNet^[20]网络模型的方法进行对比分析。在训练过程中, PSNR 平均值随训练轮数的变化见图 9,从图 9 中可以看出:NAFNet^[20]模型收敛速度较慢, PSNR 平均值较低;MIMOUNet^[14]、MPRNet^[15]和 EDAMRNet 模型均在 350 轮之后基本保持稳定。上述结果说明本文网络模型在训练时具有较好的稳定性, PSNR 平均值最高,整体性能最好。

2.4 消融实验

本文网络模型主要由编解码器、多残差网络和 SAM 组成。为验证该网络模型设计的合理性及有效性,在训练中,使用 Urban100 作为测试数据集,

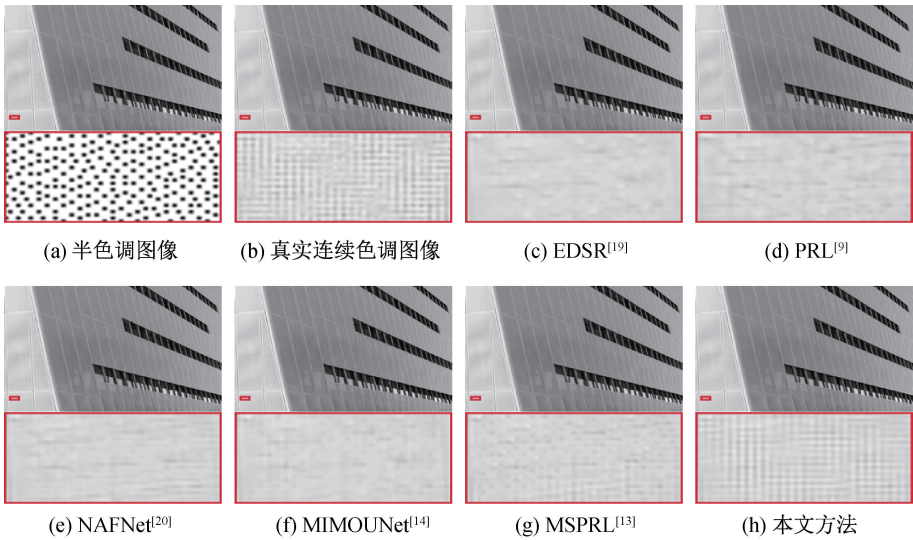


图 7 不同方法恢复的 Urban026 图像

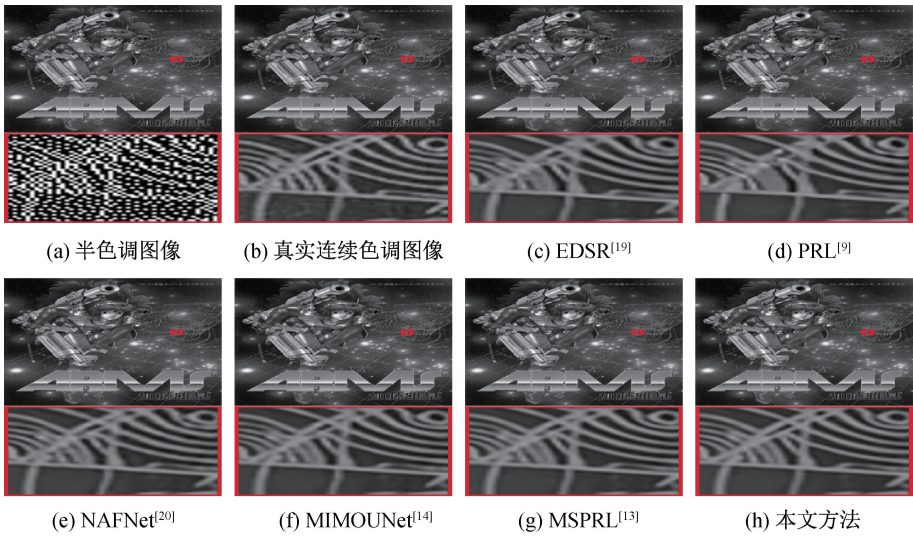


图 8 不同方法恢复的 ARMS 图像

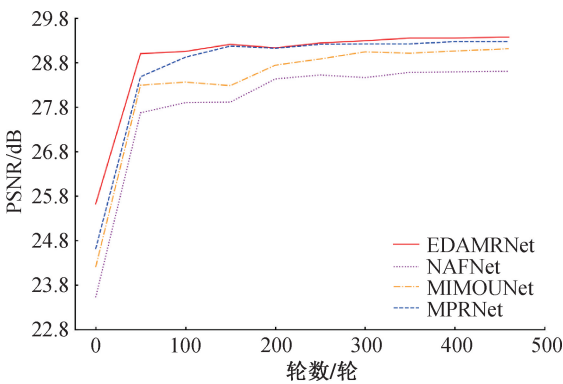


图 9 EDAMRNet、NAFNet、MIMOUNet 和 MPRNet 的 PSNR 平均值变化曲线

迭代次数设置为 10k 次。对比结果见表 2,从表中可以看出:相较于单一的编解码器,采用编解码器和多残差网络使 PSNR 平均值和 SSIM 平均值分别提高 0.44 dB 和 0.0058,图像恢复效果有明显提升;

通过在编解码器和多残差网络之间引入 SAM,使 PSNR 平均值和 SSIM 平均值分别提高 0.06 dB 和 0.0007,恢复出更高质量的连续色调图像。

表 2 不同网络结构和模块的对比结果				
编解码器	多残差网络	SAM	PSNR/dB	SSIM
✓			30.50	0.9148
✓	✓		30.94	0.9206
✓	✓	✓	31.00	0.9213

注:“✓”表示使用该网络结构或模块。

在训练中,分别采用 L_1 损失函数和 L_{total} 损失函数。对比结果见表 3,从表中可以看出:相较于 L_1 损失函数,采用 L_{total} 损失函数使 PSNR 平均值和 SSIM 平均值分别提高 0.15 dB 和 0.0003。上述结果说明 L_{total} 损失函数使模型能够同时提取图像空间域和频域的信息,提升了连续色调图像的质量。

表 3 不同损失函数的对比结果

损失函数	PSNR/dB	SSIM
L_1	30.85	0.9210
L_{total}	31.00	0.9213

通道数对模型的特征提取能力至关重要。在训练中,通道数分别设为 48 和 64。对比结果见表 4,从表中可以看出:当通道数由 48 增加到 64 时,PSNR 平均值和 SSIM 平均值分别提高 0.11 dB 和 0.0012。上述结果说明增加通道数能够使模型更好地提取图像特征信息。

表 4 不同通道数的对比结果

通道数/个	PSNR/dB	SSIM
48	30.89	0.9201
64	31.00	0.9213

3 结 论

本文提出了一种基于 EDAMRNet 的逆半色调方法,以端到端的方式,从半色调图像直接恢复出连续色调图像。通过在编解码器和多残差网络之间引入 SAM,对图像上下文信息进行加强,然后传递到多残差网络;采用组合损失函数优化网络,提高了恢复图像的质量。本文方法与现有最优的逆半色调方法相比,在 Urban100 和 Manga109 数据集下的 PSNR 平均值均提高了 0.1 dB,SSIM 平均值分别提高了 0.0010 和 0.0005,恢复出的连续色调图像保留了更多空间细节,整体更加清晰。

参考文献:

[1] Guo J M, Sankarasrinivasan S. Digital halftone database (DHD): A comprehensive analysis on halftone types[C] //2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). Honolulu, HI, USA. IEEE, 2019: 1091-1099.

[2] 张燕, 张二虎. 图像逆半色调技术研究[J]. 西安理工大学学报, 2017, 33(3): 282-289.

[3] 孔月萍, 曾平, 何波, 等. LUT 与 Elman 网络相结合的图像逆半调算法[J]. 中国图象图形学报, 2007, 12(11): 1988-1991.

[4] Xiong Z X, Orchard M T, Ramchandran K. Inverse halftoning using wavelets[J]. IEEE Transactions on Image Processing, 1999, 8(10): 1479-1483.

[5] Mese M, Vaidyanathan P P. Look-up table (LUT) method for inverse halftoning[J]. IEEE Transactions on

Image Processing, 2001, 10(10): 1566-1578.

[6] Hou X X, Qiu G P. Image companding and inverse halftoning using deep convolutional neural networks [EB/OL]. (2017-07-01)[2023-07-15]. <https://arxiv.org/abs/1707.00116>.

[7] Xiao Y, Pan C, Zheng Y, et al. Gradient-guided DCNN for inverse halftoning and image expanding[C]//Asian Conference on Computer Vision. Cham: Springer International Publishing, 2019: 207-222.

[8] Yuan J, Pan C, Zheng Y, et al. Gradient-guided residual learning for inverse halftoning and image expanding[J]. IEEE Access, 2019, 8: 50995-51007.

[9] Xia M, Wong T T. Deep inverse halftoning via progressively residual learning[C]//Computer Vision-ACCV 2018: 14th Asian Conference on Computer Vision. Perth, Australia. Cham: Springer International Publishing, 2019: 523-539.

[10] Son C H. Inverse halftoning through structure-aware deep convolutional neural networks [J]. Signal Processing, 2020, 173: 107591.

[11] Shao L H, Zhang E H, Li M. An efficient convolutional neural network model combined with attention mechanism for inverse halftoning[J]. Electronics, 2021, 10(13): 1574.

[12] 李梅, 张二虎. 融合注意力机制的多尺度深度网络的逆半调方法[J]. 包装工程, 2022, 43(11): 283-291.

[13] Li F Y, Yang J. Rethinking PRL: A multiscale progressively residual learning network for inverse halftoning [EB/OL]. (2023-05-27) [2023-07-15]. <https://arxiv.org/abs/2305.17355>.

[14] Cho S J, Ji S W, Hong J P, et al. Rethinking coarse-to-fine approach in single image deblurring[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, QC, Canada. IEEE, 2022: 4621-4630.

[15] Zamir S W, Arora A, Khan S, et al. Multi-stage progressive image restoration [C] // 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA. IEEE, 2021: 14816-14826.

[16] Shi W Z, Caballero J, Huszár F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA. IEEE, 2016: 1874-1883.

[17] Cochran W T, Cooley J W, Favin D L, et al. What is the fast Fourier transform? [J]. Proceedings of the

IEEE, 1967, 55(10): 1664-1674.

[18] Everingham M, Ali Eslami S M, Van Gool L, et al. The pascal visual object classes challenge: A retrospective[J]. International Journal of Computer Vision, 2015, 111(1): 98-136.

[19] Lim B, Son S, Kim H, et al. Enhanced deep residual networks for single image super-resolution[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Honolulu, HI, USA. IEEE, 2017: 1132-1140.

[20] Chen L Y, Chu X J, Zhang X Y, et al. Simple baselines for image restoration[C]//17th European Conference on Computer Vision (ECCV). Tel Aviv, Israel. Cham: Springer Nature Switzerland, 2022: 17-33.

(责任编辑:康 锋)