

DOI:10.11784/tdxbz202407012

基于通道协同与色彩校正的 RAW 数据到 RGB 图像重建方法

刘昱¹, 陈永祺¹, 王瀚¹, 刘人赫¹, 汪少初²

(1. 天津大学微电子学院, 天津 300072; 2. 天津市测绘院有限公司, 天津 300381)

摘要: 图像传感器直接输出的原始 RAW 数据需经过重建操作后才能转换为 RGB 图像, 用于重建 RGB 的传统多步骤图像信号处理 (ISP) 算法往往具有流程复杂、灵活性低、误差累积等问题, 而现有的一些基于深度学习的图像信号处理算法存在未能充分利用 RAW 图像模式信息与未能充分利用 RGB 通道相关性等局限。基于此, 本文设计了基于深度学习的 ISP-Net 模型, 针对上述不足进行改进。该模型包含模式信息复原、跨通道重建和色调调整 3 个主要模块。首先, 模式信息复原模块通过掩码提取 RAW 图像的颜色通道信息, 并通过通道间的融合操作最大化地学习 RAW 图像的模式信息; 其次, 跨通道重建模块通过通道间相互指导的方式, 充分利用通道相关性提升图像重建质量; 最后, 色调调整模块通过构建色调调整矩阵, 提升颜色复原的准确性。所设计的模型在合成数据集与真实数据集上均进行了测试。在主观评价上, 本文方法在色彩准确度、纹理细节重建方面更接近真实图像; 在客观指标上, 特别地, 在噪声强度为 15 时, 本文方法的峰值信噪比在 Kodak、McMaster、WED-CDM 和 Urban 100 数据集上均为最优指标, 分别为 32.33 dB、32.44 dB、32.22 dB 及 31.67 dB, 最大可优于同数据集次优指标 0.24 dB。在真实数据集上的峰值信噪比为 28.98 dB, 相较于传统图像信号处理方法 FlexISP、SEM 和 ADMM, 峰值信噪比均高出至少 17 dB, 相比于基于深度学习的 RAW 图像重建方法 Deepjoint 和 MGCC-B 峰值信噪比分别提升了 4.03 dB 和 1.55 dB。

关键词: 图像信号处理; RAW 图像重建; 模式信息复原; 跨通道重建; 色调调整

中图分类号: TN911.73

文献标志码: A

文章编号: 0493-2137(2025)04-0395-11

A Method for RAW Data to RGB Image Reconstruction Based on Channel Synergism and Color Correction

Liu Yu¹, Chen Yongqi¹, Wang Han¹, Liu Renhe¹, Wang Shaochu²

(1. School of Microelectronics, Tianjin University, Tianjin 300072, China;

2. Tianjin Institute of Surveying and Mapping Co., Ltd., Tianjin 300381, China)

Abstract: The original RAW data output by image sensors needs to be reconstructed into RGB images. Traditional multistep image signal processing (ISP) algorithms often face challenges such as complex processes, low flexibility, and error accumulation. Furthermore, some deep-learning-based ISP algorithms struggle to fully utilize the mode information of RAW images and the correlation between RGB channels. To address these deficiencies, ISP-Net—a deep learning model designed for improved RGB image reconstruction—is proposed herein. The model comprises a mode information restoration module, a cross-channel reconstruction module, and a tonal adjustment module. The mode information restoration module extracts color channel data from RAW images using masks and learns mode information through channel fusion. The cross-channel reconstruction module utilizes channel correlations to enhance image reconstruction quality through mutual guidance between channels. Finally, the tonal adjustment mod-

收稿日期: 2024-07-21; 修回日期: 2024-10-30.

作者简介: 刘昱 (1976—), 男, 博士, 教授, liuyu@tju.edu.cn.

通信作者: 汪少初, wangshaochu20@hotmail.com.

基金项目: 天津市自然科学基金资助项目 (22JCZDJC00270); 天津市研究生科研创新资助项目 (2021YJSO2B12).

Supported by the Natural Science Foundation of Tianjin, China (No. 22JCZDJC00270), the Research and Innovation Project for Postgraduates in Tianjin, China (No. 2021YJSO2B12).

ule employs a tonal adjustment matrix for precise color restoration. Experiments conducted on synthetic and real datasets reveal that in terms of subjective evaluation, the ISP-Net model achieves good color accuracy and texture detail, aligning closely with real images. In terms of objective metrics, particularly with noise intensities of 15, the peak signal-to-noise ratio (PSNR) of ISP-Net exhibits optimal performance across the Kodak, McMaster, WED-CDM, and Urban 100 datasets, with values of 32.33 dB, 32.44 dB, 32.22 dB, and 31.67 dB, respectively. These values outperform the second-best metrics within the same datasets by up to 0.24 dB. On real datasets, the PSNR reaches 28.98 dB, which surpasses the traditional ISP algorithms, namely FlexISP, SEM, and ADMM, by at least 17 dB. Compared to those of the deep-learning-based RAW image reconstruction algorithms such as Deepjoint and MGCC-B, the PSNR is improved by 4.03 dB and 1.55 dB, respectively.

Keywords: image signal processing; RAW image reconstruction; mode information restoration; cross-channel reconstruction; tonal adjustment

随着互联网技术的飞速发展,媒体文件,尤其是图片和视频,已成为人们日常生活中不可或缺的信息载体.移动物联网的高速发展进一步推动了边缘端设备,如智能手机、虚拟现实(virtual reality, VR)设备、自动驾驶车辆等,在图像采集与处理方面的广泛应用.这些设备所采集的图像质量,对于用户体验和后续图像处理至关重要.

然而,从图像传感器直接输出的原始 RAW 数据到高质量 RGB 图像的呈现,需要经过一系列复杂的图像处理流程,即图像信号处理(image signal processing, ISP)算法.传统图像信号处理算法是一个由多种图像处理技术组成的综合系统.其通过分解复杂任务为简单步骤的方法,有效促进了算法的开发和维护,缩短了开发周期.该算法将 CMOS 传感器捕获的原始光信号转换为彩色图像,涉及到暗电流校正^[1]、去马赛克^[2]、降噪^[3]、白平衡^[4]、颜色校正^[5]和伽马校正^[6]等步骤,旨在提升图像质量.然而,这种流水线设计存在流程复杂、灵活性低、误差累积等问题.此外,传统的图像信号处理算法在黑夜等极端环境下难以生成高质量图像,且其开发需要完备的专业色彩知识.

近年来,卷积神经网络(convolutional neural network, CNN)在图像处理领域取得了显著进展,为图像信号处理流程的优化提供了新的思路.CNN 已被成功应用于去马赛克^[7]、去噪^[8]、色调变换^[9]、亮度调节^[10]和去模糊^[11]等领域,并且性能优异.除此之外,也有将深度学习模型用于实现完整图像信号处理功能的方案,即 RAW 图像重建.例如,PyNET 方法^[12]引入了类金字塔网络来融合图像的局部和全局特征.TSCNN 方法^[13]则采用了两阶段处理流程,充分融合了滤光片信息和高低频分解策略.Guo 等^[14]提出的两阶段网络,分别专注于处理去马赛克和去噪任务.而 MicroISP 方法^[15]设计了轻量级带注意力机

制的三分支卷积网络.考虑到 RAW 图像重建任务中绿色通道采样率是红色和蓝色的 2 倍,GCP-Net 方法^[16]针对绿色通道特性进行优化.MGCC 方法^[17]注重不同颜色通道间的交互和关联性.InC 方法^[18]则使用 3 个独立的网络来解耦绿色、红绿色差和蓝绿色差特征,从而实现高效的 RAW 图像重建.

尽管这些工作已取得了一定的成果,但仍存在诸多亟待解决的问题.例如,一些工作^[12,15-17]未能将 RAW 图像模式信息,如各颜色像素分布及其相对位置关系等,充分用于重建过程;一些工作^[12-15]在处理 RGB 通道时忽视了通道间的相关性.为解决这些问题,本文提出了一种基于通道协同与色彩校正的 RAW 数据到 RGB 图像重建方法,即设计了一种可替代传统图像信号处理流程的深度学习模型 ISP-Net.首先,本文为模型设计了模式信息复原(mode information restoration, MIR)模块,用于充分将 RAW 图像模式信息充分用于重建过程.该模块通过掩码提取 RAW 图像的颜色通道信息,并利用通道间的融合操作最大化地学习 RAW 图像的模式信息.其次,为了充分利用 RGB 通道相关性,本文为模型引入了跨通道重建(cross-channel reconstruction, CCR)模块.该模块通过通道间相互指导的方式,充分利用颜色通道信息及其相关性进行 RAW 图像重建.再次,本文为模型设计了色调调整(tonal adjustment, TA)模块,以进一步优化图像色彩复原的质量.最后,本文构建了一个真实的复杂光照场景 RAW-RGB (complex lighting RAW-RGB, CLRR)图像数据集,为不同场景 RAW 图像重建模型的训练和测试提供了支持.

1 本文方法

1.1 模型总体结构

本文设计了 ISP-Net 模型,其主要由 3 部分构

成：模式信息复原模块、跨通道重建模块和色调调整模块。图 1 展示了其模型结构，其中， H 和 W 分别代表特征图的长度和宽度， R 、 G 、 B 分别表示红、绿、蓝 3 种颜色的特征分量。在该模型中，RAW 图像首先被输入到模式信息复原模块，用于从原始图像中提取并恢复关键的模式信息，以支持后续模块的 RAW 图像重建工作。随后，处理后的模式信息被传递至跨通道

重建模块。此模块通过残差通道注意力组 (residual group, RG)^[19]与跨通道指导部分多次重复堆叠(红框所示部分)，可充分利用不同颜色通道间的交互信息，进而实现更为精确的 RAW 图像重建。接着，图像特征数据被输入至色调调整模块。该模块负责调整图像色彩，确保输出的图像色彩准确自然。最终，模型可以重建出高质量的 RGB 图像。

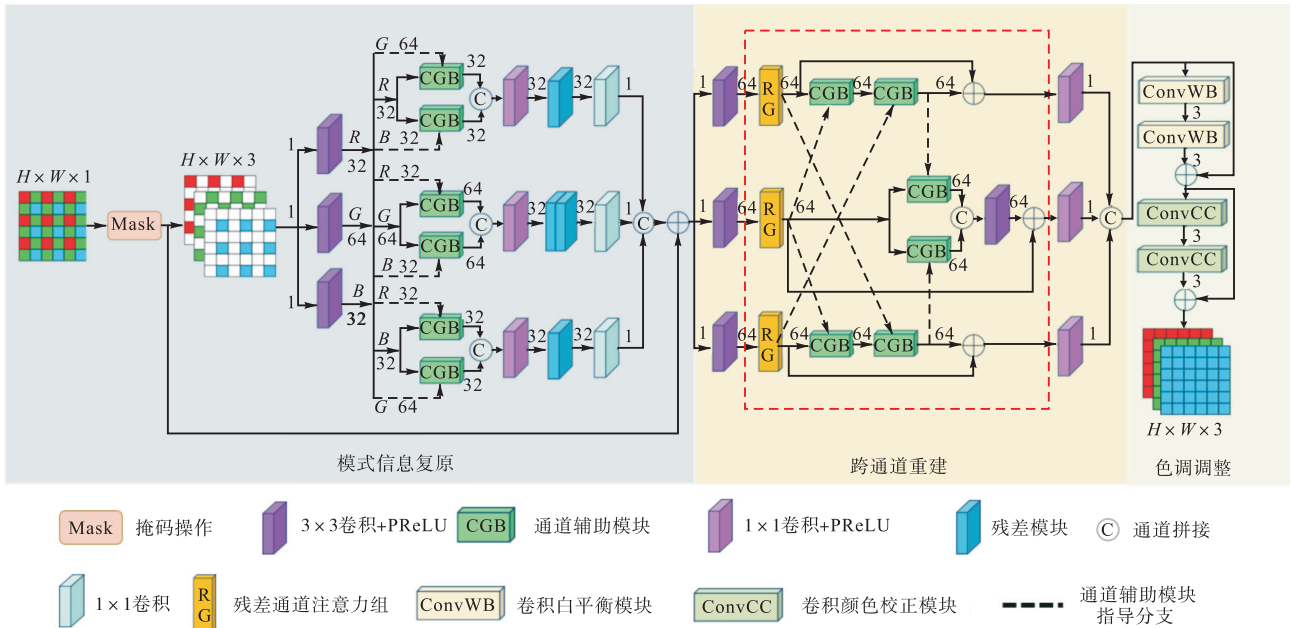


图 1 ISP-Net 模型结构
Fig.1 ISP-Net model structure

1.2 模式信息复原模块

在输入阶段需要考虑如何最大化利用各颜色通道模式信息，即各通道的像素信息及其相对位置关系，以提升特征提取效果。常规的神经网络模型在处理 RAW 图像时，通常使用拜耳采样函数或其他类似的滤光片阵列采样函数来分别提取 R_r (绿色偏红色)、 G_b (绿色偏蓝色)、 B 4 个特征分量。然而，这种处理方法在 RAW 图像重建阶段存在一些问题，例如像素位置映射关系不明确，以及直接提取的颜色特征分量尺寸为原图 1/4，模型需要进行上采样，从而可能导致性能下降等。为了有效地解决这些问题，本文提出了一种创新的模式信息复原模块，它能更有效地利用图像中的模式信息，进而提升 RAW 图像重建的质量。

该模块的工作过程可以简要概括为输入图像的掩码处理以及通道信息的特征提取和融合。首先，本文通过设计好的掩码操作 (Mask) 对 RAW 图像进行处理，生成 RGB 三通道掩码特征图。具体来说，使用 3 个不同的掩码来分别保留 R 、 G 、 B 3 个颜色通道的

信息，忽略其余颜色通道信息，如图 1 左侧所示。这一做法的优势在于有效提取颜色信息的同时避免上采样，确保特征图的分辨率不发生改变，从而避免该过程中图像信息丢失的问题。

接下来，对掩码处理后的特征图使用卷积进行初步的特征提取。在此过程中，由于人眼对绿色信息更敏感，以及绿色通道信息在 RAW 图像中采样率是其他颜色通道的 2 倍，因此在设置卷积核数量时，本文将绿色通道的卷积核数量设为其他颜色通道的 2 倍，即红色和蓝色通道的卷积核数量为 $N = 32$ ，绿色通道的卷积核数量为 $2N = 64$ 。

之后，考虑到不同颜色通道信息在 RAW 图像重建过程中存在互补性，本文引入了通道辅助模块 (channel guiding block, CGB)^[17]。该模块能够整合其他颜色通道的信息，为特定颜色通道的重建提供指导，具体结构如图 2 所示，其中， N_1 和 N_2 分别代表特征通道数， r 代表扩张率。其工作原理是选择一个辅助颜色通道作为指导分支，同时选定一个主颜色通道作为重建分支。在指导分支中，首先使用一个 3×3

卷积将辅助颜色通道的特征通道数转换为与主颜色通道相同的特征通道数. 接着使用卷积注意力模块(convolutional block attention module, CBAM)^[20]对辅助颜色通道信息进行初步特征提取. 随后, 通过 3×3 卷积将特征通道数翻倍, 并分别采用 sigmoid 和 tanh 函数各处理一半通道, 从而获取像素级的缩放参数和偏置参数.

在重建分支中, 主颜色通道应用特征缩放和偏置参数进行像素级缩放和偏置调整. 接着, 利用 3 个扩张率分别为 1、2、5 的 3×3 卷积对调整后的主颜色通道特征进行处理, 捕捉多尺度细节. 所得特征经过通道拼接(concatenation)操作后, 再利用卷积进一步

提取特征. 最后, 将原始主颜色通道信息与所提取特征逐点相加, 得到通道辅助模块的最终输出. 此过程可表示为

$$O_i = \sum_{i \in \Omega} [w(\alpha_i I_i + \beta_i) + b] \quad (1)$$

式中: O_i 为以位置 i 为中心的卷积窗口输出结果; w 为应用于整个输入特征的共享卷积权重; α_i 和 β_i 分别为以位置 i 为中心的卷积窗口内像素级的缩放参数和偏置参数; I_i 为以 i 为中心的卷积窗口内的输入特征; Ω 为该卷积窗口; b 为所有图像特征位置上共享的卷积偏置.

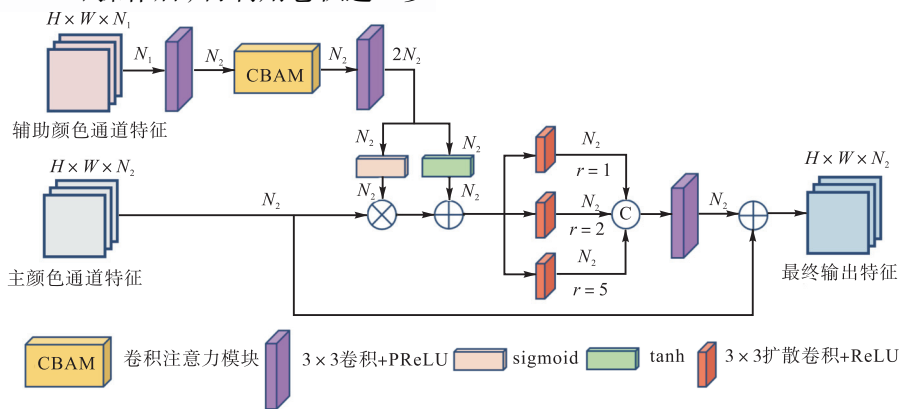


图 2 CGB 结构

Fig.2 Channel guiding block structure

通过对式(1)进行等价变换, 可证明通道辅助模块的作用可等效于为每个像素位置分配独特的卷积权重, 即

$$O_i = \sum_{i \in \Omega} (w\alpha_i I_i + w\beta_i + b) \quad (2)$$

$$O_i = \sum_{i \in \Omega} (w_i I_i + b_i) \quad (3)$$

式中: b_i 为以位置 i 为中心的卷积窗口对应的特定卷积偏置; w_i 为该窗口对应的特定卷积权重.

因此, 在 ISP-Net 模型中, 通过通道辅助模块可实现红、绿、蓝 3 种颜色通道的并行图像重建, 如图 1 左侧所示, 即利用蓝色和绿色通道指导红色通道的图像重建, 并将得到的特征图进行通道拼接操作, 相同操作也用于绿色和蓝色通道, 从而完成初步的特征提取和融合. 随后, 对红、绿、蓝 3 种颜色通道分别应用 1×1 卷积进行通道数和特征的调整. 为进行更深层次的处理, 本文对各颜色通道特征采用残差模块(ResBlock)^[21]进行特征提取. 特别地, 在信息更为丰富的绿色通道上部署了 2 个残差模块, 其余颜色通道则部署 1 个残差模块. 之后, 通过 1×1 卷积将所有颜色特征图的通道数降至 1, 并进行通道拼接. 最

终, 将融合后的特征与经过掩码处理的中间图像逐点相加后输出.

1.3 跨通道重建模块

在第 1 阶段完成对通道模式信息的有效提取后, 如何高效地利用这些颜色通道信息以进行高质量的 RAW 图像重建是设计跨通道重建模块的关键. 由于 MGCC^[17]网络中跨通道交叉指导策略能自适应地学习通道间相关性, 极大地提升 RAW 图像重建的质量. 因此, 在构建 ISP-Net 模型时, 本文将其融入跨通道重建模块的设计中, 并根据网络的具体需求进行细致的调整和优化, 以确保其性能和效率的最大化.

跨通道重建模块作为 ISP-Net 的核心组件, 通过颜色通道间的相互指导实现图像的高效重建. 该模块利用不同颜色通道间的互补信息, 以自适应的方式进行特征学习和融合, 同时抑制冗余信息, 从而提升重建图像的质量. 在模块的初始阶段, 本文采用 3×3 卷积分别对 R 、 G 、 B 3 个颜色通道进行了浅层特征提取. 随后, 这些特征被送入残差通道注意力组中进行深层次关系的自适应学习. 该残差组由 20 个含通道注意力机制的单元堆叠而成, 每个单元都能够自适应

学习颜色通道不同层次的特征。

在残差组处理之后,再次引入了通道辅助模块来实现跨颜色通道间的自适应指导。以红色通道为例,其残差组输出的特征会依次被绿色与蓝色通道通过通道辅助模块进行特征指导,随后输出结果与红色通道残差组的原始输出进行逐点相加。上述操作旨在最大化地利用这两个通道的有用信息来增强红色通道的特征表示,蓝色通道的处理过程类似。同样考虑到绿色通道信息的重要性,本文采用了特殊的处理方式以期进一步提升重建效果。绿色通道的残差组输出会同时接受红、蓝两个通道的通道辅助模块指导,并将所得结果进行通道拼接后利用 3×3 卷积进行特征提取和融合,再将得到的特征与残差组的原始输出进行逐点相加。

最终,对经过通道辅助模块跨通道指导后的各颜色通道特征分别使用 1×1 卷积进行通道数的调整,将各颜色通道数调整为 1,进一步实现特征调整和融合。值得注意的是,模块可以通过增加 3 种颜色的残差通道注意力组和通道辅助模块跨通道指导部分(图 1 红框部分)的重复堆叠次数,从而实现更好的 RAW 图像重建效果。经过实验验证,重复堆叠次数为 4 时 ISP-Net 模型达到最佳性能。

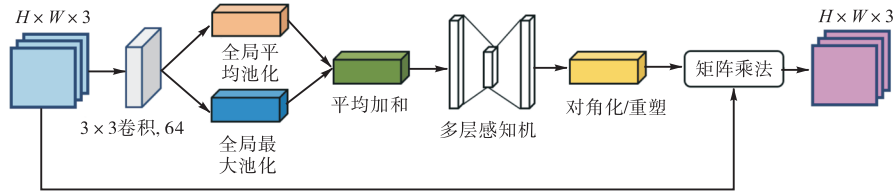


图 3 卷积白平衡/颜色校正模块

Fig.3 ConvWB/ConvCC module

在色调调整模块的结构设计中,如图 1 右侧所示,首先采用两个串联的卷积白平衡模块对输入特征数据进行处理,之后将其输出结果与原始输入的特征数据进行逐点相加,从而在网络中更精准地实现了白平衡功能,即

$$\begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = \begin{bmatrix} w_{11} & 0 & 0 \\ 0 & w_{22} & 0 \\ 0 & 0 & w_{33} \end{bmatrix} \begin{bmatrix} w'_{11} & 0 & 0 \\ 0 & w'_{22} & 0 \\ 0 & 0 & w'_{33} \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} w_{11}w'_{11}+1 & 0 & 0 \\ 0 & w_{22}w'_{22}+1 & 0 \\ 0 & 0 & w_{33}w'_{33}+1 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (4)$$

式中: R' 、 G' 、 B' 为经过模块处理后的红、绿、蓝像

1.4 色调调整模块

考虑到传统的图像信号处理流水线中,最终处理阶段通常包括色域变换、伽马校正等色调调整。在 ISP-Net 模型中,本文设计了色调调整模块来显式的构建色调调整矩阵以实现这一功能。该模块主要由 2 个卷积白平衡 (convolutional white balance, ConvWB) 模块和 2 个卷积颜色校正 (convolutional color correction, ConvCC) 模块串联而成,这两种模块的结构均可用图 3 表示。具体区别为多层感知机 (multilayer perceptron, MLP) 的输出参数个数与特征矩阵的处理方法不同。例如,在卷积白平衡模块工作过程中,特征数据首先通过 3×3 卷积进行特征提取,再并行的通过全局平均池化 (global average pooling, AvgPool) 和全局最大池化 (global max pooling, MaxPool) 操作整合特征,然后将所得特征平均加和,得到进一步整合特征的结果。接着,通过在多层感知机中对上一步骤的输出结果进行处理并输出 3 个关键参数,接着对所得参数进行对角化处理形成对角矩阵,并与图像特征进行矩阵乘法。在卷积颜色校正模块工作过程中,多层感知机则输出 9 个参数,重塑形成 3×3 矩阵后,再与图像特征进行矩阵乘法。

素数值; w_{ij} 、 w'_{ij} 均为卷积白平衡参数。

接着,将前述流程的输出结果输入至类似流程设计的卷积颜色校正模块中,完成颜色的精细校正,则有

$$\begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix} \begin{bmatrix} c'_{11} & c'_{12} & c'_{13} \\ c'_{21} & c'_{22} & c'_{23} \\ c'_{31} & c'_{32} & c'_{33} \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^3 (c_{1i}c'_{i1})+1 & \sum_{i=1}^3 (c_{1i}c'_{i2}) & \sum_{i=1}^3 (c_{1i}c'_{i3}) \\ \sum_{i=1}^3 (c_{2i}c'_{i1}) & \sum_{i=1}^3 (c_{2i}c'_{i2})+1 & \sum_{i=1}^3 (c_{2i}c'_{i3}) \\ \sum_{i=1}^3 (c_{3i}c'_{i1}) & \sum_{i=1}^3 (c_{3i}c'_{i2}) & \sum_{i=1}^3 (c_{3i}c'_{i3})+1 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (5)$$

式中 c_{ij} 、 c'_{ij} 均为卷积颜色校正参数。

2 实验与结果分析

2.1 数据集及实验设置

本文通过在合成数据集及真实的复杂光照场景数据集上实验对所提方法进行了全面的性能评估。其中合成数据集由真实 RGB 图像添加噪声后通过拜耳函数采样而成,真实数据集则是通过个人移动设备拍摄制作而成。在合成数据方面,本文遵循了与 Zhang 等^[17]、Liu 等^[22]、Chen 等^[23]相同的训练集设置。采用 800 张 DIV2K^[24]和 2 650 张 Flickr2K^[21]的高分辨率图像作为训练集。测试数据涵盖了包括 Kodak、McMaster^[25]、WED-CDM^[26]、MIT moiré^[27]和 Urban 100 5 个基准数据集。Kodak 数据集包含 24 张分辨率为 768 像素 × 512 像素的彩色图像,展现了多样化的视觉场景。相比之下,McMaster 数据集的 18 张 500 像素 × 500 像素图像呈现出更为饱和的色彩与锐利的结构特征。WED-CDM 数据集则包括 100 张高质量全彩图像,涵盖更广泛的场景与色阶变化。MIT moiré 数据集的测试数据有 1 000 张 128 像素 × 128 像素图像,拍摄对象均为具有密集纹理结构且易产生摩尔纹的物品,这对 RAW 图像重建工作造成一定的困难。而 Urban 100 数据集则拥有 100 张城市场景的高分辨率图像。这些合成数据集共同确保了本文方法在不同设备和场景下得到全面的评估。

考虑到合成数据集是先对 RGB 图像添加不同标准差的高斯噪声,再通过拜耳阵列采样而形成的。这种方式并不能完全体现图像信号处理方法在实际应用场景中的效果。为更贴合实际并特别检验图像信号处理方法在复杂光照环境下的性能,本文提出了 CLRR 图像数据集。该数据集是在不同的地点及天气条件下拍摄并制作而成,拍摄所采用的设备为 IQOO Neo5s。通过 python 的 rawpy 函数分析得知其 RAW 图像采用 10 位比特存储,滤光片阵列为 RGGB 拜耳模式。RAW 图像文件和彩色 RGB 图像分别由设备直接提供的 RAW 图像输出以及该设备自动采用的图像信号处理算法输出所得。通过后后期裁剪并筛选出了较为清晰及纹理信息丰富的 500 个 RAW-RGB 图像对,分辨率为 128 像素 × 128 像素,其中 400 个图像对用于构建训练集,而剩余的 100 个图像对作为测试集。

在模型训练阶段,采用随机裁剪输入 64 像素 × 64 像素大小的数据块,并设定批量大小为 8。为增强模型的泛化能力,输入数据进行了包括旋转(0°、90°、

180°、270°)和镜像(上下、左右)在内组合的多种数据增强操作。需要注意的是,在进行这些操作时,需要保持滤光片阵列排布方式的一致性。实验使用 Adam 优化器和 L1 Loss 损失函数进行优化。L1 Loss 的数学表达式为

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |f(x_i) - y_i| \quad (6)$$

式中: n 为样本数量; x_i 为输入的第 i 个样本; f 为采用的 RGB 重建算法; y_i 为模型第 i 个样本对应的真实标签。

此外,本文方法可通过修改掩码函数来灵活地适应不同拜耳模式(如 GRBG、BGGR、GBRG)或其他颜色滤波阵列模式(如富士 X-Trans)。为客观评估模型性能,本文采用了峰值信噪比(peak signal-to-noise ratio, PSNR)和结构相似度(structural similarity, SSIM)指数作为评价指标,分别从像素级误差和结构相似度两个方面对重建图像的质量进行了量化评估。

2.2 合成数据集实验结果

合成数据集包含合成 RAW 图像及其对应的原始图像。其构建过程如下:对原始输入图像添加不同标准差强度的高斯噪声,随后采用拜耳阵列对其进行采样,并保存所生成的合成 RAW 图像与输入原图。为了全面评估所提模型在 RAW 图像重建上的性能,本文挑选了 9 种现有的 RAW 图像重建方法作为对比,包括 FlexISP^[28]、SEM^[29]和 ADMM^[30] 3 种传统方法,以及 Deepjoint^[27]、SGNet^[22]、Wild-JDD^[23]、MGCC-B^[17]、Inc^[18]和 Guo 等^[14]6 种基于深度学习的方法。上述方法的性能数据为直接援引相关论文中所报告的数据,主观对比图由笔者提供的开源代码运行生成。而针对部分未进行开源的论文方法,则直接引用相关论文数据,且实验的主观对比图不进行完整对照。同时,考虑到一些方法使用噪声水平作为辅助输入,这在实际场景中是不现实的,因此不进行比较。

本文选择添加 4 种标准差强度的高斯噪声进行性能测试,即噪声强度 σ 分别为 0、5、10 和 15。值得注意的是,当噪声强度不为 0 时,图像信号处理的任务为联合去马赛克去噪;而当噪声强度为 0 时,图像信号处理的任务则转变为单纯地去马赛克任务。表 1 详细呈现了各方法在 5 个不同测试数据集上,对应于 4 种噪声强度的性能测试结果。

由表 1 结果可知,当噪声强度为 0、5、10 时,本文方法在 Kodak、McMaster、WED-CDM、MIT moiré 和 Urban 100 这 5 个数据集上得到的峰值信噪比数值和结构相似度数值优于其他方法,且本文方法重建

结果指标为最优的次数增多. 特别是, 在噪声强度为 15 时, 本文方法的峰值信噪比在 Kodak、McMaster、WED-CDM 和 Urban 100 数据集上均为最优指标, 分

别为 32.33 dB、32.44 dB、32.22 dB 和 31.67 dB, 最大可优于次优指标 0.24 dB. 而在 MIT moiré 数据集上为次优指标, 仅低于最优指标 0.21 dB.

表 1 不同噪声强度下各方法在 5 个合成数据集上 RAW 图像重建结果比较

Tab.1 Comparison of RAW image reconstruction results among various methods on five synthetic datasets under different noise intensities

方法	σ	Kodak 数据集		McMaster 数据集		WED-CDM 数据集		MIT moiré 数据集		Urban 100 数据集	
		峰值信噪比/dB	结构相似度	峰值信噪比/dB	结构相似度	峰值信噪比/dB	结构相似度	峰值信噪比/dB	结构相似度	峰值信噪比/dB	结构相似度
FlexISP ^[28]	0	38.83	0.990 3	38.29	0.984 2	37.94	0.983 2	34.61	0.955 6	37.36	0.986 1
SEM ^[29]		37.41	0.986 0	34.12	0.962 7	34.74	0.974 7	32.15	0.944 0	34.61	0.978 9
ADMM ^[30]		31.48	0.941 9	32.37	0.951 1	31.56	0.949 9	28.43	0.898 5	28.73	0.935 2
Deepjoint ^[27]		40.10	0.992 4	37.88	0.982 9	37.93	0.985 8	35.01	0.968 6**	36.51	0.986 1
MGCC-B ^[17]		42.18	0.994 9*	39.14	0.986 4*	39.78	0.990 0*	35.59	0.958 8	38.61	0.989 9*
InC ^[18]		42.81**	—	39.62**	—	—	—	—	—	39.16	—
Guo 等 ^[14]		42.76*	0.989 3	39.61*	0.972 5	40.22*	0.983 1	36.53**	0.953 3	39.52**	0.986 4
ISP-Net(本文方法)		42.81**	0.995 4**	39.54	0.987 6**	40.40**	0.990 6**	36.05*	0.959 6*	39.39*	0.990 6**
FlexISP ^[28]	5	33.93	0.953 0	33.35	0.945 4	33.19	0.948 7	31.65	0.934 1	33.00	0.956 1
SEM ^[29]		34.51	0.967 1	32.36	0.943 0	32.74	0.955 8	30.64	0.928 2	32.49	0.964 3
ADMM ^[30]		31.22	0.937 2	32.05	0.941 6	31.15	0.937 4	28.32	0.890 6	28.73	0.927 6
Deepjoint ^[27]		36.06	0.975 2	35.46	0.969 8	35.17	0.971 5	32.75	0.953 0**	34.02	0.973 2
SGNet ^[22]		—	—	—	—	—	—	32.15	0.904 3	34.54	0.953 3
Wild-JDD ^[23]		36.88	0.952 0	35.85	0.942 5	35.92	0.954 3	32.29	0.898 7	34.70	0.953 4
MGCC-B ^[17]		37.06	0.979 0*	36.26	0.974 3*	36.21	0.978 0*	33.11	0.946 6	35.30	0.979 1*
Guo 等 ^[14]		37.12*	0.951 9	36.41*	0.943 3	36.40*	0.956 3	33.81**	0.928 9	35.77**	0.969 4
ISP-Net(本文方法)	37.17**	0.979 7**	36.48**	0.975 6**	36.45**	0.978 8**	33.31*	0.947 2*	35.64*	0.979 9**	
FlexISP ^[28]	10	29.36	0.866 3	29.15	0.860 1	28.99	0.865 4	28.05	0.868 8	28.97	0.886 5
SEM ^[29]		29.74	0.886 4	28.68	0.858 6	28.86	0.875 5	27.62	0.867 1	28.71	0.898 5
ADMM ^[30]		29.96	0.916 5	30.33	0.905 8	29.63	0.907 8	27.43	0.867 5	27.96	0.907 0
Deepjoint ^[27]		33.14	0.952 5	33.08	0.947 6	32.71	0.948 5	30.50	0.926 5	31.70	0.952 7
SGNet ^[22]		—	—	—	—	—	—	30.09	0.861 9	32.14	0.922 9
Wild-JDD ^[23]		33.81	0.912 7	33.53	0.912 3	33.44	0.924 4	30.30	0.864 5	32.42	0.928 8
MGCC-B ^[17]		34.08*	0.959 8*	33.94	0.959 1*	33.76	0.963 0*	30.95	0.928 1*	33.00	0.966 1*
Guo 等 ^[14]		34.08*	0.914 3	34.03*	0.915 2	33.88*	0.929 4	31.40**	0.893 9	33.04*	0.948 2
ISP-Net(本文方法)	34.13**	0.960 2**	34.09**	0.960 5**	33.90**	0.964 0**	31.08*	0.928 9**	33.27**	0.967 4**	
FlexISP ^[28]	15	26.32	0.761 1	26.28	0.757 3	26.09	0.766 4	25.42	0.787 6	26.13	0.802 9
SEM ^[29]		25.80	0.746 2	25.43	0.732 1	25.46	0.748 1	24.72	0.768 4	25.33	0.787 3
ADMM ^[30]		28.12	0.874 4	28.20	0.854 1	27.72	0.865 4	26.04	0.830 8	26.57	0.891 9
Deepjoint ^[27]		31.24	0.929 3	31.27	0.921 1	30.92	0.923 8	28.89	0.898 3	29.94	0.930 4
SGNet ^[22]		—	—	—	—	—	—	28.60	0.818 8	30.37	0.892 3
Wild-JDD ^[23]		31.92	0.876 5	31.90	0.884 6	31.75	0.896 5	28.89	0.831 0	30.79	0.905 5
MGCC-B ^[17]		32.27*	0.941 4*	32.32	0.943 8*	32.11	0.948 4*	29.50	0.909 7*	31.43*	0.953 7*
Guo 等 ^[14]		32.24	0.880 7	32.40*	0.890 2	32.20*	0.904 8	29.83**	0.859 7	31.21	0.926 8
ISP-Net(本文方法)	32.33**	0.942 0**	32.44**	0.945 4**	32.22**	0.949 5**	29.62*	0.9107**	31.67**	0.9551**	

注: **标识为最优指标; *标识为次优指标; 由于 InC^[18]、Guo 等^[14]、SGNet^[22]和 Wild-JDD^[23]未开源代码, 表中只比较了其原论文公开数据, 空缺部分以“—”表示.

因此, 本文推测噪声会影响图像结构的相关性, 从而削弱不同颜色通道之间的引导作用, 但是每个颜色通道仍然能够为其他颜色通道的重建提供有用的信息. 特别地, 本文方法在不同数据集上表现出的优势也不相同. 本文方法用于 Kodak、McMaster、WED-CDM 数据集上大多为最优指标, 当本文的方法用于 MIT moiré 数据集和 Urban 100 数据集时, 随着噪声

强度的增加本文方法的优势也在不断地增加. 为了进一步展示各方法的性能差异, 本文对各方法的测试结果进行了可视化展示. 由于 MIT moiré 数据集中包含了更多密集纹理且易产生摩尔纹的图片, 这对 RAW 图像重建工作提出了更高的挑战. 因此本文选择在 MIT moiré 数据集上进行去马赛克任务比较, 即当噪声强度为 0 时, 展示各方法进行 RAW 图像重建

复杂光照场景进行 RAW 重建优化. 此外, 实际应用中还需根据特定传感器的特性来调整动态范围, 并需适配不同设备间的色彩域差异. 这些因素的综合作用共同导致传统图像信号处理算法在跨设备使用时效果不尽如人意. 相较于传统图像信号处理方法, 深度学习的方法在图像质量评估指标峰值信噪比和结构相似度上具备更优性能. 如表 2 所示, 本文所提出的方法在与其他两种基于深度学习的 RAW 图像重建方法 Deepjoint^[27]和 MGCC-B^[17]比较时, 峰值信噪比分别提升了 4.03 dB 和 1.55 dB, 同时结构相似度也分别提高了 0.060 0 和 0.012 6.

表 2 真实复杂光照场景数据集 CLRR 上各方法 RAW 图像重建结果

Tab.2 RAW image reconstruction results of different methods on the real complex lighting RAW-RGB dataset CLRR

方法	峰值信噪比/dB	结构相似度
FlexISP ^[28]	11.55	0.296 8
SEM ^[29]	11.56	0.434 1
ADMM ^[30]	11.64	0.314 8
Deepjoint ^[27]	24.95	0.861 4
MGCC-B ^[17]	27.43	0.908 8
ISP-Net(本文方法)	28.98	0.921 4

实验的可视化结果如图 5 所示, 其中原始 RGB 图像为 IQOO Neo5s 手机自带的图像信号处理算法处理所得结果. 从图中可看出, 本文方法在复杂光照环境下的色彩复原能力优于其他方法. 具体地, ADMM^[30]方法与 FlexISP^[28]方法进行 RAW 重建后

的图像存在明显的色调偏绿和亮度偏高的问题; SEM^[29]方法则整体呈现出亮度偏低的特征; 其中上述传统方法重建结果均存在整体的色调和亮度重建失真问题, 其原因可能在于传统方法无法针对复杂的光照场景进行动态地适配, 从而影响了图像重建的准确性. Deepjoint^[27]方法在宣传栏文字图片的重建中出现了色调偏红的现象, 且在不规则墙体图片的重建中则产生了不应有的亮斑; MGCC-B^[17]方法在宣传栏文字图片的重建中出现了彩虹纹, 同时在处理含英文文字的户外场景图片时, 文字背景部分颜色发生了明显的色偏. 相较于其他方法, 本文方法通过精确提取模式信息, 最大化利用通道间相关性, 并通过色调调整, 实现了高质量图像重建. 因此在复杂光照场景的图像处理中, ISP-Net 方法展现出更高的复原准确性与真实度, 有效避免了多种常见问题.

2.4 消融实验

为深入探究模型结构对 RAW 图像重建性能的影响, 本文设计了一系列消融实验. 首先, 验证跨通道重建模块的有效性及其引入的必要性. 接着, 采用控制变量法, 依次在模型中添加了模式信息复原模块、跨通道重建模块以及色调调整模块, 并在每个集成阶段逐步进行性能验证. 所有实验均在合成数据集上进行训练和验证, 且所有实验设置保持一致, 仅通过调整模型结构来比较性能差异. 在噪声强度为 0 的测试数据集上, 获得了如表 3 所示的定量对比结果.



(a) ADMM^[30] (b) FlexISP^[28] (c) SEM^[29] (d) Deepjoint^[27] (e) MGCC-B^[17] (f) 本文方法 (g) 原始 RGB 图像

图 5 在真实复杂光照场景数据集 CLRR 上各方法可视化结果

Fig.5 Visualization results of various methods on the real complex lighting RAW-RGB dataset CLRR

通过对表 3 数据深入分析, 发现单独采用跨通道重建模块在 RAW 图像重建实验中展现出了良好的性能. 它深入提取并高效综合 RAW 图像特征与各颜色通道信息, 显著提升了重建图像的准确性和质量,

为模块优化奠定了基础. 而单独采用模式信息复原模块进行 RAW 图像重建时效果不佳, 因其仅对 RAW 图像模式信息进行了浅层特征的提取, 缺乏深层处理. 随后, 在模式信息复原模块基础上集成了跨

通道重建模块. 实验结果显示, 图像评价指标峰值信噪比和结构相似度均有大幅提高, 且优于单独采用跨通道重建模块时的性能. 这表明这两个模块的组合能充分提取 RAW 图像模式信息, 并深层次挖掘通道间信息, 以优化 RAW 图像重建质量. 最后, 通过引

入色调调整模块, 进一步优化了 RAW 图像重建效果. 该模块精准校正了由设备和场景因素引起的颜色失真, 使图像色彩更加真实自然, 从而显著提高了重建图像的整体视觉质量.

表 3 在噪声强度为 0 的合成数据集上进行消融实验结果

Tab.3 Results of ablation experiments on synthetic datasets with a noise intensity of 0

模块	Kodak 数据集		McMaster 数据集		WED-CDM 数据集		MIT moiré 数据集		Urban 100 数据集	
	峰值信噪比/dB	结构相似度	峰值信噪比/dB	结构相似度	峰值信噪比/dB	结构相似度	峰值信噪比/dB	结构相似度	峰值信噪比/dB	结构相似度
跨通道重建	41.79	0.994 5	38.74	0.985 8	39.50	0.989 3	34.39	0.953 9	37.60	0.987 7
模式信息复原	15.86	0.699 1	17.93	0.749 0	14.70	0.720 7	14.28	0.655 0	13.72	0.704 1
跨通道重建 + 模式信息复原	42.65	0.995 3	39.48	0.987 5	40.29	0.990 7	35.94	0.958 9	39.17	0.990 3
跨通道重建 + 模式信息复原 + 色调调整 (ISP-Net)	42.81	0.995 4	39.54	0.987 6	40.40	0.990 6	36.05	0.959 6	39.39	0.990 6

3 结 语

本文提出了一种基于通道协同与色彩校正的 RAW 数据到 RGB 图像重建方法, 即设计了一个基于深度学习的 ISP-Net 模型, 旨在替代传统图像信号处理算法对 RAW 图像进行高质量重建. 该模型由模式信息复原模块、跨通道重建模块以及色调调整模块构成, 分别负责从 RAW 图像中学习模式信息、利用颜色通道相关性重建图像以及优化图像色彩复原质量. 为了验证 ISP-Net 模型性能, 本文采用 5 个公开的合成数据集 Kodak、McMaster、WED-CDM、MIT moiré 和 Urban 100 进行测试. 测试结果表明, 本文方法在噪声强度分别为 0、5 和 10 时, 5 个数据集上得到的峰值信噪比数值和结构相似度数值优于其他方法, 且本文方法重建结果指标为最优的次数增多. 特别是, 在噪声强度为 15 时, 本文方法的峰值信噪比在 Kodak、McMaster、WED-CDM 和 Urban100 数据集上均为最优指标, 分别为 32.33 dB、32.44 dB、32.22 dB 及 31.67 dB, 最大可优于次优指标 0.24 dB. 而在 MIT moiré 数据集上为次优指标, 仅低于最优指标 0.21 dB. 除此之外, 为了验证图像信号处理方法对真实复杂光照环境的 RAW 图像重建能力, 本文构建了一套真实复杂光照场景的 CLRR 数据集, 该数据集包含 500 对分辨率为 128 像素×128 像素的 RAW-RGB 图像, 涵盖了不同光照条件场景. 在该数据集上训练并与现有方法进行对比. 测试结果表明本文方法相较于传统图像信号处理方法 FlexISP^[28]、SEM^[29]和 ADMM^[30], 峰值信噪比均高出至少 17 dB, 相比于基于深度学习的 RAW 图像重建方法 Deepjoint^[27]和 MGCC-B^[17]峰值信噪比分别提升了

4.03 dB 和 1.55 dB, 并进一步证明了其在 RAW 图像重建任务上的优越性. 同时, 可视化实验结果证明, 本文设计的 ISP-Net 模型还可较好地处理 RAW 图像重建中存在的复杂光照场景及密集纹理内容. 此外, 本文设计的模型也存在模型结构较为复杂的问题, 未来的工作有望优化出更简洁高效的模型结构框架.

参考文献:

- [1] 郭智慧, 魏 巍. 基于 CIS 暗电平校正算法研究与实现[J]. 电子技术与软件工程, 2023 (5): 89-93. Guo Zhihui, Wei Wei. Research and implementation based on CIS black level correction algorithm[J]. Electronic Technology & Software Engineering, 2023 (5): 89-93 (in Chinese).
- [2] Li X, Gunturk B, Zhang L. Image demosaicing: A systematic survey[C]//Visual Communications and Image Processing 2008. San Jose, USA, 2008: 489-503.
- [3] Buades A, Coll B, Morel J M. A non-local algorithm for image denoising[C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). San Diego, USA, 2005: 60-65.
- [4] Gijzenij A, Gevers T, van de Weijer J. Improving color constancy by photometric edge weighting[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 34 (5): 918-929.
- [5] Kwok N M, Shi H Y, Ha Q P, et al. Simultaneous image color correction and enhancement using particle swarm optimization[J]. Engineering Applications of Artificial Intelligence, 2013, 26 (10): 2356-2371.
- [6] 张美的, 余顺园. 基于局部自适应伽马校正低照度图像增强[J]. 计算机与现代化, 2023 (8): 74-78, 118. Zhang Meidi, Yu Shunyu. Low-light image enhancement based on adaptive local gamma correction[J]. Computer and Modernization, 2023 (8): 74-78, 118 (in Chinese).
- [7] Tan D S, Chen W Y, Hua K L. DeepDemosaieking:

- Adaptive image demosaicking via multiple deep fully convolutional networks[J]. *IEEE Transactions on Image Processing*, 2018, 27(5): 2408-2419.
- [8] Liu Y, Qin Z Y, Anwar S, et al. Invertible denoising network: A light solution for real noise removal[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville, USA, 2021: 13365-13374.
- [9] He J W, Liu Y H, Qiao Y, et al. Conditional sequential modulation for efficient global image retouching [C]// *European Conference on Computer Vision-ECCV 2020*. Glasgow, UK, 2020: 679-695.
- [10] Fu X Y, Zeng D L, Huang Y, et al. A fusion-based enhancing method for weakly illuminated images[J]. *Signal Processing*, 2016, 129: 82-96.
- [11] Guo C L, Yan Q X, Anwar S, et al. Image dehazing transformer with transmission-aware 3D position embedding[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans, USA, 2022: 5812-5820.
- [12] Ignatov A, van Gool L, Timofte R. Replacing mobile camera ISP with a single deep learning model[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. Seattle, USA, 2020: 536-537.
- [13] Chang K, Li H X, Tan Y, et al. A two-stage convolutional neural network for joint demosaicking and super-resolution[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021, 32(7): 4238-4254.
- [14] Guo Y, Jin Q Y, Morel J M, et al. Joint demosaicking and denoising benefits from a two-stage training strategy[J]. *Journal of Computational and Applied Mathematics*, 2023, 434: 115330.
- [15] Ignatov A, Sycheva A, Timofte R, et al. MicroISP: Processing 32mp photos on mobile devices with deep learning[C]// *European Conference on Computer Vision-ECCV 2022 Workshops*. Tel Aviv, Israel, 2022: 729-746.
- [16] Guo S, Liang Z T, Zhang L. Joint denoising and demosaicking with green channel prior for real-world burst images[J]. *IEEE Transactions on Image Processing*, 2021, 30: 6930-6942.
- [17] Zhang Y, Sun W J, Chen Z Z. Joint image demosaicking and denoising with mutual guidance of color channels[J]. *Signal Processing*, 2022, 200: 108674.
- [18] Niu Y, Zhang L, Li C. Independent and collaborative demosaicking neural networks[C]// *Proceedings of the 5th ACM International Conference on Multimedia in Asia*. Tainan, China, 2023: 1-7.
- [19] Zhang Y L, Li K P, Li K, et al. Image super-resolution using very deep residual channel attention networks[C]// *Proceedings of the European Conference on Computer Vision(ECCV)*. Munich, Germany, 2018: 286-301.
- [20] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]// *Proceedings of the European Conference on Computer Vision(ECCV)*. Munich, Germany, 2018: 3-19.
- [21] Lim B, Son S, Kim H, et al. Enhanced deep residual networks for single image super-resolution[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. Honolulu, USA, 2017: 136-144.
- [22] Liu L, Jia X, Liu J Z, et al. Joint demosaicking and denoising with self guidance[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA, 2020: 2240-2249.
- [23] Chen J R, Wen S, Chan S H G. Joint demosaicking and denoising in the wild: The case of training under ground truth uncertainty[C]// *Proceedings of the AAAI Conference on Artificial Intelligence*. Vancouver, Canada, 2021: 35(2): 1018-1026.
- [24] Agustsson E, Timofte R N. Challenge on single image super-resolution: Dataset and study[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. Honolulu, USA, 2016: 21-26.
- [25] Zhang L, Wu X L, Buades A, et al. Color demosaicking by local directional interpolation and nonlocal adaptive thresholding[J]. *Journal of Electronic Imaging*, 2011, 20(2): 023016.
- [26] Tan R X, Zhang K, Zuo W M, et al. Color image demosaicking via deep residual learning[C]// *IEEE International Conference on Multimedia and Expo(ICME)*. Hong Kong, China, 2017: 1-6.
- [27] Gharbi M, Chaurasia G, Paris S, et al. Deep joint demosaicking and denoising[J]. *ACM Transactions on Graphics(ToG)*, 2016, 35(6): 1-12.
- [28] Heide F, Steinberger M, Tsai Y T, et al. Flexisp: A flexible camera image processing framework[J]. *ACM Transactions on Graphics(ToG)*, 2014, 33(6): 1-13.
- [29] Klatzer T, Hammernik K, Knobelreiter P, et al. Learning joint demosaicking and denoising based on sequential energy minimization[C]// *2016 IEEE International Conference on Computational Photography (ICCP)*. Evanston, USA, 2016: 1-11.
- [30] Tan H L, Zeng X R, Lai S M, et al. Joint demosaicking and denoising of noisy bayer images with ADMM [C]// *2017 IEEE International Conference on Image Processing(ICIP)*. Beijing, China, 2017: 2951-2955.

(责任编辑:孙立华)