

DNF: 黑暗中的解耦和反馈网络

孟宇昂¹ 翻译*

Xin Jin^{1†} Ling-Hao Han^{1†} Zhen Li¹ Chun-Le Guo^{1‡} Zhi Chai² Chongyi Li³

¹VCIP, CS, Nankai University ²Hisilicon Technologies Co. Ltd.

³S-Lab, Nanyang Technological University

{xjin, lhhan}@mail.nankai.edu.cn, zhenli1031@gmail.com, guochunle@nankai.edu.cn,

chaizhi2@huawei.com, chongyi.li@ntu.edu.sg

<https://github.com/Srameo/DNF>

Abstract

RAW 数据的独特属性在低光照图像增强方面展现出巨大潜力。然而,单阶段和多阶段方法的性能受到现有体系结构固有的局限性的制约。在单阶段方法中,在两个不同领域之间的混合映射,即噪声到清晰和 *RAW* 到 *sRGB*,由于领域模糊性,误导了单阶段方法。多阶段方法仅通过每个阶段产生的图像传播信息,忽略了在有损图像级数据流中丰富的特征。本文探索了这些瓶颈的通用解决方案,并提出了一个名为解耦和反馈 (*Decouple and Feedback, DNF*) 的框架。为了缓解领域模糊性,我们解耦了领域特定的子任务,并充分利用了 *RAW* 和 *sRGB* 领域的独特属性。通过反馈机制跨阶段进行特征传播,避免了由于图像级数据流造成信息丢失的问题。我们方法的两个关键见解令人满意地解决了基于 *RAW* 数据的低光图像增强的固有局限性,使我们的方法仅用 19% 的参数就大大超越了以前最先进的方法,在 *SID* 的 *Sony* 和 *Fuji* 子集上实现了 0.97dB 和 1.30dB 的 *PSNR* 改进。

1. Introduction

低光场景下的成像越来越受到关注,尤其是随着智能手机和监控系统夜视模式的普及。然而,由于信噪比极低,低光图像增强 (LLIE) 是一项具有挑战性的任务。

*mya@mail.nankai.edu.cn

†Equal contribution.

‡C. L. Guo is the corresponding author.

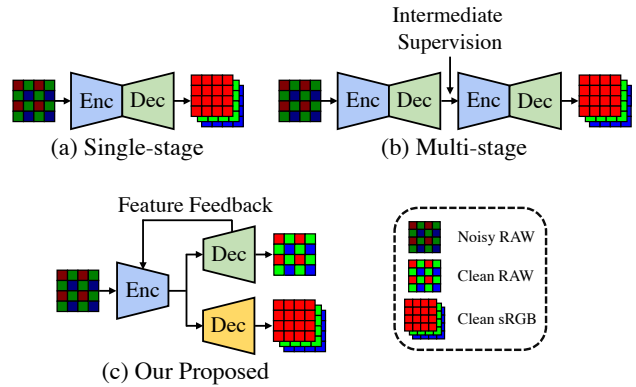


图 1. 不同基于 RAW 的低光照图像增强方法的缩略图。(a) 单阶段方法。(b) 具有中间监督的多阶段方法。(c) 所提出的 DNF。

最近,深度学习解决方案已被广泛研究以解决不同数据领域的这一任务,从基于 *sRGB* 的方法 [14, 15, 21, 40] 到基于 *RAW* 的方法 [2, 7, 35, 47]。与 *sRGB* 数据相比,未处理过的 *RAW* 数据在三个方面是独一无二的,这对 LLIE 有益: 1) 信号与 *RAW* 域中的光子计数呈线性相关, 2) 在图像信号处理 (ISP) 管道 [33] 之前, *RAW* 图像上的噪声分布是可处理的, 3) *RAW* 格式的更高位深记录了更易区分的低强度信号。

开创性的工作 *SID* [2] 为基于 *RAW* 的 LLIE 提出了一个大规模配对数据集,重新激发了人们对数据驱动方法的兴趣。如 Fig. 1 所示,一条工作线 [2, 5, 12, 13, 22, 42] 专注于设计单级网络架构,另一条工作线 [4, 7, 35, 47] 利用多级网络进行渐进增强。尽管性能得到了很大的提升,但这两种架构仍然受到固有限制

的瓶颈。首先，当前的单阶段方法迫使神经网络学习从嘈杂的 RAW 域到干净的 sRGB 域的直接映射。两个不同域（嘈杂到干净和 RAW 到 sRGB）之间的混合映射会误导整体增强过程，从而导致域模糊问题。例如，在颜色空间转换期间，RAW 图像中可处理的噪声将映射到不可预测的分布。因此，最终结果不可避免地会出现颜色偏移和未处理的噪声。其次，现有的多阶段方法由级联子网络组成 pipeline，每个子网络负责基于上一阶段的输出图像进行逐步增强。在图像级数据流的设计下，只有图像在多个阶段之间向前传播，而后一阶段仅从前一阶段的结果中获取信息。同时，由于下采样操作或单独的目标函数 [41]，每个阶段中的每个子网络都可能造成信息丢失。因此，次优性能与有损图像级数据流紧密相关。误差随着阶段的推移而传播、累积和放大，最终无法重建纹理细节。

为了充分发挥 RAW 图像在 LLIE 中的潜力，需要一种通用的 pipeline，以超越上述两个限制。具体来说，神经网络应该利用不同域中上述优点 [7]，而不是被域模糊性所困扰。根据 RAW 和 sRGB 域的独特属性，将增强分解为特定域的子任务至关重要。在探索了 RAW 域中的线性和可处理噪声之后，可以有意识地执行从 RAW 域到 sRGB 域的色彩空间转换，而不会受到噪声干扰。此外，pipeline 不能妨碍跨阶段的通信，而是仅允许一小部分有损信息通过的图像级数据流。由于子任务的多样性，每个级别的中间特征往往是互补的 [20, 46]。同时，多尺度特征保留了纹理和上下文信息，为后续阶段提供了额外的指导 [41]。因此，不同阶段的特征需要跨数据流传播，聚合丰富的特征并维持完整的信息。领域特定的解耦与特征级数据流一起，促进了可学习性，从而实现了更好的增强性能，并保留了该方法的解释性。

基于这些原则，我们提出了一个解耦和反馈框架，其设计针对基于 RAW 的 LLIE 进行了以下定制。增强过程被解耦为两个特定领域的子任务：RAW 域中的去噪 [30, 33, 45, 48] 和 sRGB 域中的色彩恢复 [8, 28, 39]，如 Fig. 1(c) 所示。在先前研究 [27] 中常用的编码器-解码器架构下，子网络中的每个模块都来自每个域的独特属性：用于 RAW 去噪的通道独立去噪 (CID) 块和用于色彩渲染的矩阵色彩校正 (MCC) 块。此外，我们不使用不准确的去噪 RAW 图像，而是利用 RAW 解码器的多尺度特征作为去噪先验。然后，特征通过提出

的门控融合模块 (GFM) 流入共享 RAW 编码器，自适应地区分纹理细节和剩余噪声。经过去噪先验反馈后，信号与噪声进一步区分，在 RAW 域中产生完整且丰富的特征。得益于特征级数据流，MCC 块解码器可以高效处理剩余增强和到 sRGB 域的颜色转换。

主要贡献总结如下：

- 特定领域任务解耦扩展了 RAW 和 sRGB 域中独特属性的利用，避免了域歧义。
- 由去噪先验反馈支持的特征级数据流减少了错误累积并跨阶段聚合了互补特征。
- 与之前最先进的方法相比，所提出的方法仅以 19% 的参数和 63% 的 FLOPs 获得了显著的改进，例如在 Sony SID 数据集上 PSNR 改进了 0.97dB，在 Fuji SID 数据集上 PSNR 改进了 1.30dB。

2. Related Work

2.1. 基于 RAW 的低光图像增强

正如我们在 Sec. 1 中提到的，由于 RAW 图像具有独特的属性，它已被广泛用于极低光照条件下的图像增强。如 Fig. 1 所示，基于 RAW 的方法通常根据是否有中间监督分为两类：单阶段和多阶段。单阶段方法 [2, 5, 12, 13, 22, 42] 旨在强制深度学习从嘈杂的 RAW 域到干净的 sRGB 域的直接映射。为了获得更好的性能，人们采用了多种尝试，包括相似性和感知损失 [42]、残差学习 [22]、多尺度特征 [5] 和轻量级 [12, 13]。然而，由于域模糊性，上述单阶段方法往往无法恢复纹理细节。为了解决单阶段方法的局限性，提出了多阶段方法 [4, 7, 35, 47]。通过对 sRGB 域进行中间监督，EEMEFN [47] 和 LDC [35] 在第二阶段重建了细节。对不同域的中间监督用于不同的目的，例如 RAW 域用于解耦 [7]，单色域用于低光信息补充 [4]。然而，所有现有的多阶段方法都共享相同的架构，即级联编码器-解码器 [27]。它们的图像级数据流会导致跨阶段的错误积累。我们提出的方法在两个方面不同于现有方法。1) 采用特定于域的解耦架构来充分利用 RAW 和 sRGB 格式的属性。2) 采用特征级反馈架构来处理图像级数据流的错误积累。

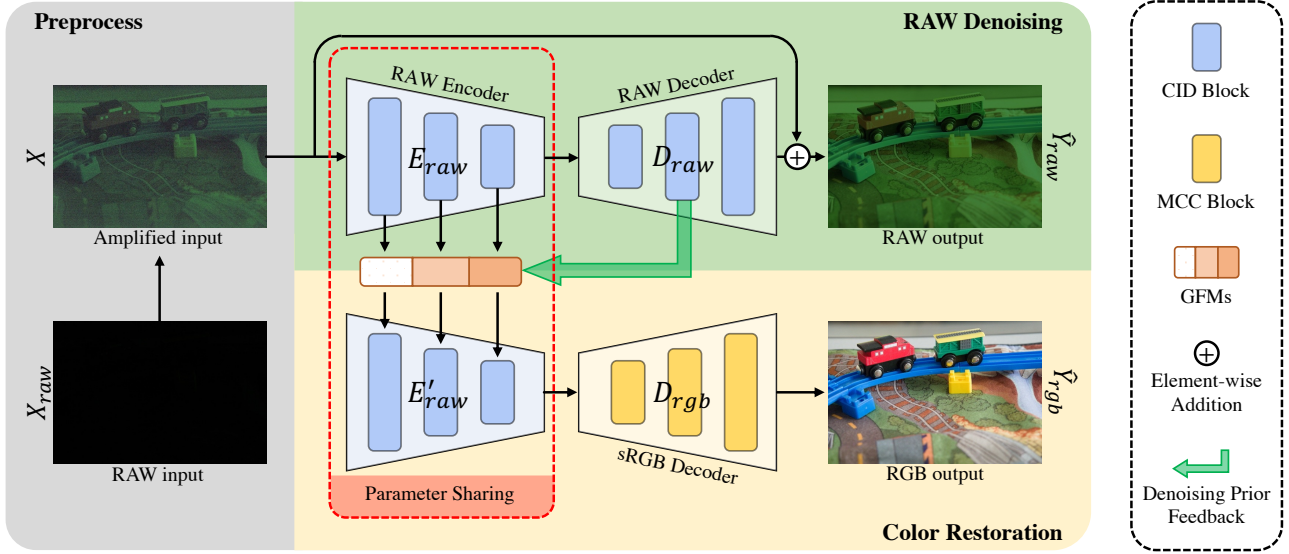


图 2. 所提出的 DNF 的概述。我们的 DNF 包含三个主要组件：1) 一个 RAW 编码器和一个 RAW 解码器，在 RAW 域中进行去噪，并对 RAW 图像输出进行辅助监督，2) 处理反馈特征的门控融合模块（GFMs），3) 一个 sRGB 解码器，执行颜色空间转换并输出最终结果。

2.2. 解耦机制

解耦机制旨在将原始任务分解为几个更简单的子任务，然后明确地攻克它们。通过适当的解耦，神经网络可以更容易收敛，从而获得更好的性能。李等人 [18] 将外推任务解耦为边界框布局生成、分割布局生成和图像生成。合理的解耦通过边界框和分割布局两个立足点缩小了文本和图像之间的域差距。最近关于修复的研究 [16, 25, 26] 旨在将修复任务解耦为结构和纹理重建，以获得更好的结果。在高级任务 [11] 中，将域自适应任务解耦为特征分布对齐和分割可以提高性能。我们通过领域特定任务解耦来实现这一机制，将 LLIE 任务解耦为 RAW 去噪和色彩恢复。

2.3. 反馈机制

反馈机制使网络能够从先前的状态访问掌握。这个想法已经应用于许多任务，例如分类 [37]、超分辨率 [17, 19] 和点云完成 [36]。在涉及反馈机制的情况下，Li 等人 [36] 采用了课程学习策略进行逐步恢复。Yan 等人 [36] 打算使用反馈机制用高分辨率特征丰富低分辨率特征。所有现有的方法都应用反馈机制来逐步完成单一任务，这与我们的方法不同。我们的反馈机制使我们的网络能够在两个不同的子任务之间进行通信，也可以在不同的领域中进行通信。

3. Methodology

如 Fig. 2 所示，我们提出的 DNF 框架由两个阶段组成：RAW 去噪和颜色渲染，以逐步增强低光照 RAW 图像。给定输入图像 X_{raw} ，在乘以预定义的放大比例 [2] 后，放大后的图像 X 首先通过 RAW 域中的编码器 E_{raw} 和解码器 D_{raw} 进行去噪。然后，我们没有使用不准确的 \hat{Y}_{raw} 进行颜色渲染，而是将去噪特征 \mathbf{F}_{dn} 从 D_{raw} 反馈回 E'_{raw} ，进一步区分信号和噪声，并在 RAW 域中组合丰富的特征。最后，sRGB 解码器 D_{rgb} 采用 RAW 域中的多尺度特征来渲染在 sRGB 域的最终输出 \hat{Y}_{rgb} 。

具体来说，一个共享的编码器 E_{raw} 和两个解码器 (D_{raw} 和 D_{rgb}) 是专门为通过领域特定任务解耦与任务特定模块而解耦的子任务设计。引入了通道独立去噪 (CID) 模块，以学习 RAW 域中不同颜色通道中的可处理且独立的噪声分布。根据颜色空间的定义，矩阵颜色校正 (MCC) 模块使用全局矩阵变换来完成剩余的增强以进入 sRGB 域。此外，我们引入了一个去噪先验反馈机制，以避免跨阶段的误差累积。通过从 RAW 解码器提取的去噪特征 \mathbf{F}_{dn} ，RAW 编码器用高频信息丰富了浅层特征。此外，我们提出了一个带门控机制的门控融合模块 (GFM) [17]，用于自适应地探索噪声中的细节。

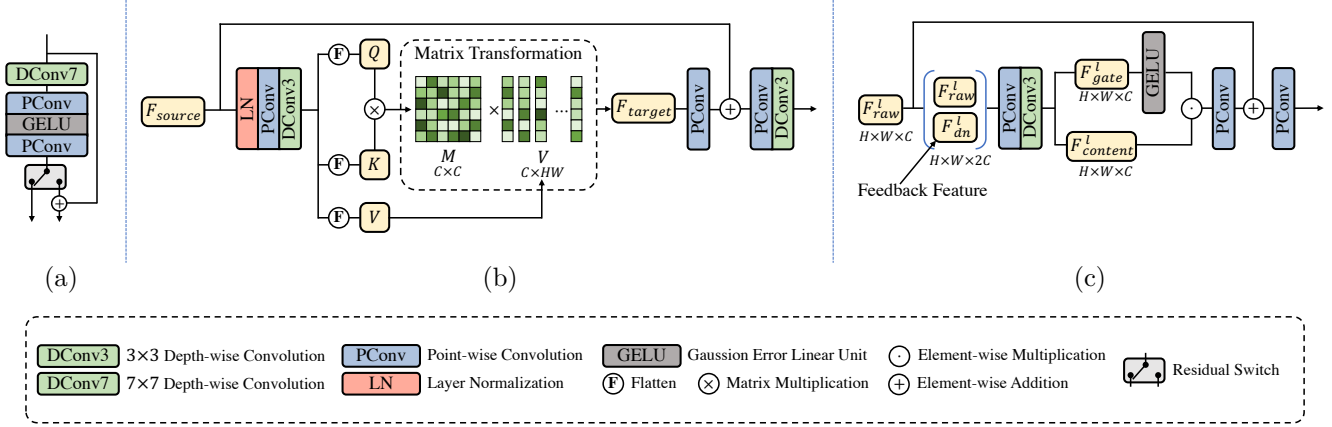


图 3. 所提出的任务特定块和融合模块的详细架构：(a) 通道独立去噪（CID）块，(b) 矩阵颜色校正（MCC）块，以及 (c) 门控融合模块（GFM）。

3.1. 领域特定任务解耦

我们提出领域特定任务解耦来处理由噪声到清晰和 RAW 到 sRGB 的混合映射引起的领域模糊。噪声的 RAW 和清晰的 sRGB 领域之间的鸿沟使网络难以学习横跨两个领域的直接映射。因此，作者提出在噪声的 RAW 领域中引入中间监督，以缓解直接学习混合映射的难度。通过在干净的 RAW 领域的中间监督，可以：1) 将增强分解为 RAW 去噪和颜色恢复，2) 充分利用这一特性，即在 RAW 图像上的噪声分布是可处理的，用于去噪，3) 减少在颜色恢复期间的噪声干扰，从而减少颜色偏移。

在 RAW 域内去噪。如图 2 所示，作者通过堆叠多个通道独立去噪（CID）块来实现 RAW 编码器 E_{raw} 和 RAW 解码器 D_{raw} 。CID 块的设计基于以下两个先验知识：1) RAW 格式的低光照图像受到与信号无关的噪声影响，该噪声遵循零均值分布 [9, 33]，2) 不同通道的信号本质上相关性较小，因此在 RAW 域中噪声分布在各个通道上趋于独立 [24, 34]。因此，需要对几乎相同的信号（邻近像素）进行 burst observations，以消除零均值噪声的干扰。同时，在去噪过程中防止通道间信息交换对于处理通道独立的噪声分布至关重要。根据上述讨论，在 CID 块中引入了大核深度卷积进行去噪。CID 块的详细结构如 Fig. 3 (a) 所示。具体来说，对于输入特征 F_{in} ，经过通道独立去噪块后的输出特征 F_{out} 可以表示为：

$$F_{out} = \text{MLP}(\text{DConv7}(F_{in})) + F_{in}, \quad (1)$$

其中 DConv7 是一个带有 7×7 核的深度卷积。MLP 是

通过两个逐点卷积层和一个 GELU [6] 非线性函数实现的。此外，设置了一个残差开关，以使用共享权重的 CID 块执行两种不同的功能，详细内容见 Sec. 3.2。

RAW 到 sRGB 的颜色校正。矩阵变换通常在规范的 ISP 流程 [23] 中使用。由于全局共享设置，如环境照明和颜色空间规范，图像的颜色主要通过通道矩阵变换来增强或转换到另一种颜色空间。遵循这一原则，作者引入了一个矩阵颜色校正块来执行全局颜色增强和局部细化，如图 Fig. 3 (b) 所示。对于 sRGB 解码器 D_{rgb} ，作者堆叠了多个 MCC 块进行颜色校正。这个块的设计受益于最近的转置自注意力 [38] 的进展。全局感受野和通道操作非常适合规范 ISP 中的颜色校正。给定输入源特征 $F_{source} \in \mathbb{R}^{C \times H \times W}$ ，查询 $Q \in \mathbb{R}^{C \times HW}$ 、键值 $K \in \mathbb{R}^{C \times HW}$ 、值 $V \in \mathbb{R}^{C \times HW}$ 依次通过 1×1 卷积、 3×3 深度卷积和展平操作生成。然后，通过矩阵乘法获得变换矩阵 $M \in \mathbb{R}^{C \times C}$ 。这个过程可以表示为：

$$Q, K, V = \text{Flatten}(\text{DConv3}(\text{PConv}(F_{source}))), \quad (2)$$

$$M = \text{Softmax}(Q \cdot K^T / \lambda), \quad (3)$$

其中，缩放系数 λ 用于数值稳定性。然后，颜色向量 V 通过矩阵 M 进行变换，在特征层面执行颜色空间转换。颜色变换后的目标特征可以通过 $F_{target} = M \cdot V$ 获得。作为对全局矩阵变换的补充，我们使用深度卷积和逐点卷积进一步细化局部细节。

3.2. 去噪先验反馈机制

在以前的基于 RAW 的增强方法中 [4, 7, 35, 47]，一部分高频内容被错误地识别为噪声，严重恶化了

最终结果，导致细节丢失和数据流的损失。为了避免现有多阶段方法的有损图像级数据流，我们提出了具有特征级信息传播的去噪先验反馈机制。记 $\mathbf{F}_{dn} = \{F_{dn}^1, F_{dn}^2, \dots, F_{dn}^L\}$ 为从 RAW 解码器 D_{raw} 提取的一组去噪特征，其中 L 表示阶段数。 \mathbf{F}_{dn} 的每个元素主要包含 RAW 域中不同尺度的最终噪声估计信息。具体来说，这些特征使噪声更容易区分，并作为进一步去噪的指导。通过将去噪特征集合 \mathbf{F}_{dn} 链接到具有多个反馈连接的 RAW 编码器的相应阶段 [1, 19, 29]，编码器逐渐生成更好的去噪特征，以进行进一步增强。因此，sRGB 解码器 D_{rgb} 可以更专注于颜色校正。反馈的 pipeline 如 Fig. 2 所示，可以表示为：

$$\mathbf{F}_{dn} = D_{raw}(E_{raw}(X)), \quad F_{rdn} = E'_{raw}(X, \mathbf{F}_{dn}), \quad (4)$$

其中 F_{rdn} 表示将被转发到 sRGB 解码器的精细去噪特征。 E'_{raw} 表示 RAW 编码器，不仅包含 E_{raw} 的权重，还配备了 L 个门控融合模块 (GFMs)。每个 GFM 负责处理来自 \mathbf{F}_{dn} 的一个反馈特征。

门控机制。 GFM 旨在通过门控机制自适应地将反馈噪声估计与初始去噪特征融合 [17]。在特征门控期间，我们希望有用信息被自适应地选择并沿着空间和通道维度合并。为了提高效率，使用逐点卷积和深度卷积 [3] 来聚合通道和局部内容信息。然后，我们将混合特征沿通道维度分成两个块，即 F_{gate}^l 和 F_{con}^l 。在通过 GELU 非线性函数激活后， F_{gate}^l 通过逐点乘法对 F_{con}^l 进行门控。通过这种门控机制实现了空间和通道的适应性。GFM 的详细结构如图 3 (c) 所示。第 l 阶段 l -th ($l \in \{1, 2, \dots, L\}$) 的操作可以表示为：

$$F_{gate}^l, F_{con}^l = \text{DConv3}(\text{PConv}([F_{raw}^l, F_{dn}^l])), \quad (5)$$

$$F_{fuse}^l = \text{PConv}(F_{con}^l \odot \text{GELU}(F_{gate}^l)) + F_{raw}^l, \quad (6)$$

其中 DConv3 和 PConv 分别表示具有 3×3 核的深度卷积和逐点卷积。 \odot 表示哈达玛积。 F_{raw}^l 是在原始 RAW 编码器中经过第 l 个上采样层后获得的特征。 F_f^l 是对应的融合特征。逐点卷积对该融合特征进行通道混合。混合特征被输入到 RAW 编码器中的下一个 CID 块进行进一步细化。

残差开关机制。 我们只在 RAW 域的去噪阶段 [22, 43] 保留 global shortcut，以实现更好的去噪效果，而在颜色恢复阶段则去除它，以避免在噪声的 RAW 域和干净的 sRGB 域之间产生模糊的连接，如 Fig. 2 所示。因

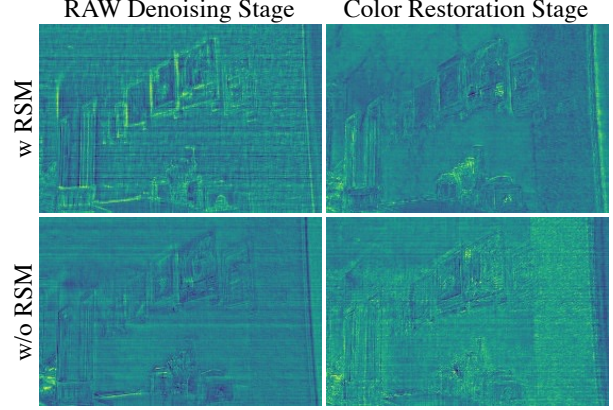


图 4. 共享 RAW 编码器的特征可视化，带或不带 RSM (请放大以获得最佳视图)。

此，编码器在去噪时需要执行噪声估计，而在颜色恢复期间则需要重建信号。为了实现单一编码器中这两种相互矛盾的功能，我们提出了一个简单但有效的残差开关机制，如 Fig. 3 (a) 所示，它赋予共享 RAW 编码器中的 CID 块产生两种相互矛盾的特征：噪声和信号。在带有全局残差连接的去噪阶段，局部残差 shortcuts 被关闭以估计噪声。相反，在渲染阶段，局部残差被激活，通过在 shortcut 上的原始特征抵消噪声，最终重建信号。如 Fig. 4 所示，带有 RSM 的共享 RAW 编码器的 CID 块能够在不同阶段产生两种不同的特征。然而，如果没有 RSM，权重共享的 CID 块在颜色恢复阶段无法区分噪声和信号，导致特征模糊。剩余的噪声限制了颜色校正过程，并再次引入了领域模糊。

3.3. 训练目标

为了通过特定于领域的任务解耦顺序地完成 RAW 去噪和颜色恢复的子任务，我们在不同领域引入了两种不同的监督，即清晰的 RAW 和清晰的 sRGB。真实值是清晰的 RAW 图像 Y_{raw} 。我们将去噪解码器的输出 RAW 图像表示为 \hat{Y}_{raw} 。我们网络的损失函数为：

$$L = \left\| Y_{raw} - \hat{Y}_{raw} \right\|_1 + \left\| Y_{rgb} - \hat{Y}_{rgb} \right\|_1, \quad (7)$$

其中 Y_{rgb} 是真实的 sRGB 图像。值得注意的是，在我们的方法中，仅使用 L_1 损失用于 RAW 监督和 sRGB 监督，而不是像之前的方法那样混合复杂的损失函数 [7, 30, 32, 42, 47]。训练细节和详细的网络架构可以在补充材料中找到。

表 1. 基于 RAW 的 LLIE 方法在 SID [2] 的 Sony 和 Fuji 子集上的定量结果。最好的结果用**粗体**表示，而第二好的结果用下划线表示。带有 \uparrow 和 \downarrow 的指标分别表示越高越好和越低越好。带 * 的方法表示模型是在下采样分辨率下训练和推理的，我们在测试时手动将结果上采样到原始分辨率。带 # 的方法表示模型仅在 SID 数据集上的小数字增益 ($\times 100$) 图像上训练和推理。“-”表示结果不可用。

Category	Method	Params.	FLOPs	Sony			Fuji		
				PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
Single-Stage	SID [2]	7.7 M	48.5 G	28.96	0.787	0.356	26.66	0.709	0.432
	DID [22]	2.5 M	669.2 G	29.16	0.785	0.368	-	-	-
	SGN [5]	19.2 M	75.5 G	29.28	0.790	0.370	<u>27.41</u>	0.720	0.430
	LLPackNet [12]	1.2 M	7.2 G	27.83	0.755	0.541	-	-	-
	RRT [13]	0.8 M	5.2 G	28.66	0.790	0.397	26.94	0.712	0.446
Multi-Stage	EEMEFN [47]	40.7 M	715.6 G	29.60	0.795	0.350	27.38	<u>0.723</u>	<u>0.414</u>
	LDC* [35]	8.6 M	124.1 G	29.56	0.799	0.359	27.18	0.703	0.446
	MCR# [4]	15.0 M	90.5G	<u>29.65</u>	<u>0.797</u>	<u>0.348</u>	-	-	-
	RRENet [7]	15.5 M	96.8 G	29.17	0.792	0.360	27.29	0.720	0.421
	Ours	2.8 M	57.0 G	30.62	<u>0.797</u>	0.343	28.71	0.726	0.391

4. Experiments and Analysis

4.1. 数据集和评估指标

我们在两个不同的基于 RAW 的低光照图像增强数据集上对提出的 DNF 进行了基准测试，即 See-In-the-Dark (SID) [2] 数据集和 Mono-Colored Raw Paired (MCR) [4] 数据集。SID [2] 数据集包含 5094 张极低光照 RAW 图像，以及由两台相机拍摄的正常光照参考图像：Sony A7S2 带有 Bayer 传感器，分辨率为 4240×2832 ，以及带有 X-Trans 传感器的 Fuji X-T2，分辨率为 6000×4000 。低光照图像的曝光时间从 0.1 秒到 0.033 秒不等，参考图像的曝光时间是低光照图像的 100 到 300 倍。请注意，Sony 子集的测试集中长曝光和短曝光的图像对存在不对齐问题，因此我们在测试阶段遵循以前的方法，丢弃了这些图像 [22, 47]。为了公平比较，所有比较的方法都在相同的设置下进行评估。MCR [4] 数据集包含 4980 张图像，分辨率为 1280×1024 ，用于训练和测试，包括 3984 张低光照 RAW 图像，498 张单色图像和 498 张 sRGB 图像。它包括两种不同场景，室内和室外，设置了不同的曝光时间，室内场景为 $1/256$ 秒到 $3/8$ 秒，室外场景为 $1/4096$ 秒到 $1/32$ 秒。然而，没有提供 RAW 格式的地面真实图像，这对于训练提出的方法至关重要。因此，我们选

表 2. 基于 RAW 的 LLIE 方法在 MCR 数据集 [4] 上的定量结果。最好的结果用**粗体**表示，而第二好的结果用下划线表示。带有 \uparrow 和 \downarrow 的指标分别表示越高越好和越低越好。

Category	Method	PSNR \uparrow	SSIM \downarrow
Single-Stage	RRT [13]	25.74	0.851
	SGN [5]	26.29	0.882
	DID [22]	26.16	0.888
	SID [2]	29.00	0.906
Multi-Stage	LDC [35]	29.36	0.904
	MCR [4]	<u>31.69</u>	<u>0.908</u>
	Ours	32.00	0.915

择每个场景中曝光时间最长的图像作为 RAW 地面真值。我们将 PSNR、SSIM [31] 和 LPIPS [44] 分别作为像素评估、结构评估和感知评估的定量评估指标。

4.2. 与 SOTA 方法对比

我们在 SID [2] 和 MCR [4] 数据集的两个子集，Sony 和 Fuji 上评估我们的 DNF，并与最先进的基于 RAW 的 LLIE 方法进行比较，包括单阶段方法 SID [2]、DID [22]、SGN [5]、LLPackNet [12] 和 RRT [13]，以及多阶段方法 EEMEFN [47]、LDC [35]、RRENet [7]

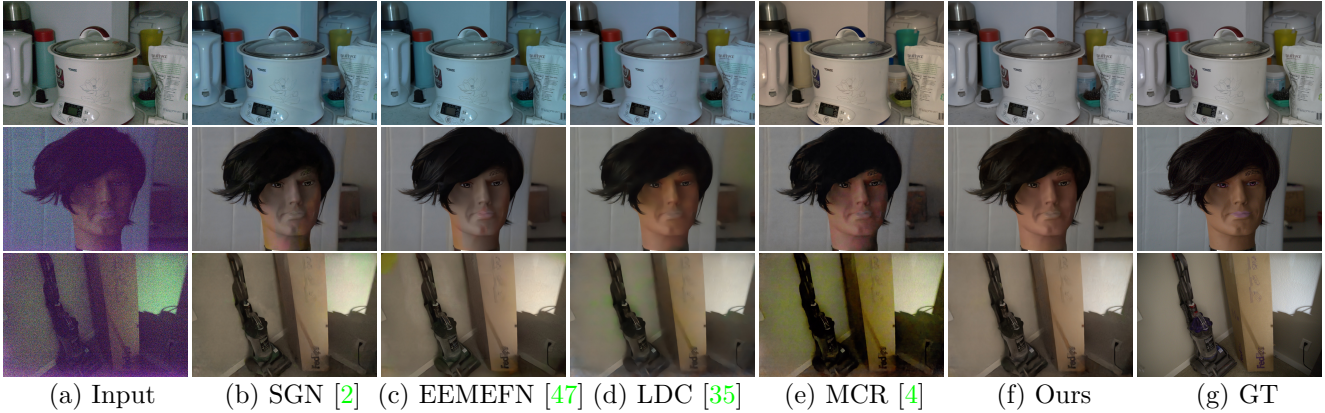


图 5. 我们的方法与最先进的方法之间的视觉比较（请放大以获得最佳视图）。我们对输入图像进行了放大和后处理，并使用了图像信号处理（ISP）进行可视化 [2]。

和 MCR [4]。

定量评估。如 Tab. 1 和 Tab. 2 所示，我们的方法在性能上远远超过了以前的最先进方法。在 SID 数据集上，DNF 在 PSNR 和 LPIPS 得分上取得了最佳成绩，与第二佳方法相比，在 Sony 和 Fuji 子集上分别实现了 0.97 dB 和 1.30 dB 的 PSNR 提升，以及 0.005 和 0.023 的 LPIPS 提升。关于复杂性，DNF 在参数和 FLOPs 方面明显少于当前最佳方法（即 MCR 和 EEMEFN）。我们的网络使用的参数是 MCR 和 EEMEFN 的 1/5 和 1/15，FLOPs 分别是它们的 3/5 和 1/13。在 MCR [4] 数据集上，DNF 在 Tab. 2 中展示了最佳 PSNR 和 SSIM 得分，与以前的最先进方法相比，分别提高了 0.31dB 和 0.07，同时参数和 FLOPs 更少。

定性评估。 Fig. 5 和 Fig. 6 显示了在 SID [2] 数据集上的定性结果。可以看出，与我们提出的方法相比，比较方法增强的结果存在严重的内容失真和伪影，这是由于它们有限的去噪能力造成的。此外，得益于 DNF 架构，颜色更准确地转换和增强，没有噪声干扰，因此展现出更好的颜色一致性以及更真实和生动的颜色渲染。我们的方法成功地抑制了强烈的噪声，同时保留了丰富的纹理细节。

4.3. 消融实验

我们对提出的 DNF 进行了广泛的消融研究。所有实验均在 SID [2] 的 Sony 子集上进行。

领域特定任务解耦。为了更好地评估 DNF 的领域特定任务解耦的影响，我们对比了对去噪解码器的不同种类的中间监督方法，如 Tab. 3 所示。1) 没有监督 (w/o

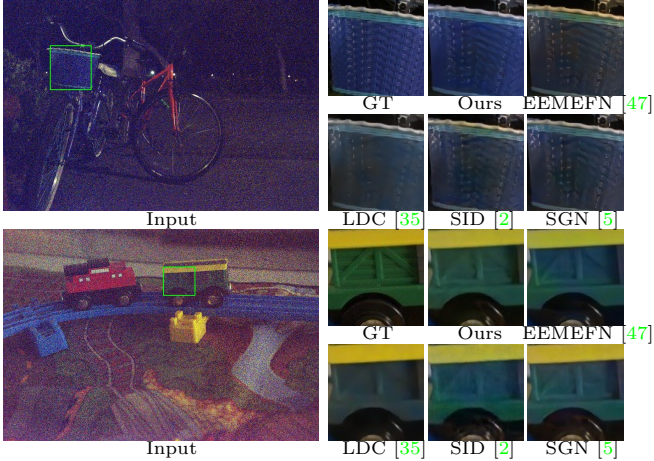


图 6. 我们的方法与最先进的方法之间的视觉比较。与最先进的方法相比，我们的方法在纹理保留和颜色恢复方面表现更好。

Sup.) 无法将去噪先验反馈回共享 RAW 编码器，导致 PSNR 下降了 0.14dB。2) sRGB 监督 (sRGB Sup.) 将主任务分解为第一阶段增强和细节重建，如 [35, 47]。第一阶段增强由于直接从嘈杂的 RAW 域学习到干净的 sRGB 域，导致领域歧义，PSNR 下降了 0.42dB。sRGB Sup. 和 w/o Sup. 之间的比较 (0.28dB↓) 表明，领域歧义严重限制了网络的性能。

去噪先验反馈机制。为了验证 DNF 的基于反馈机制的框架的有效性，我们首先检查了框架的单阶段和多阶段的变体 [4, 7, 35, 47])，如 Tab. 3 所示。1) 单阶段变体通过直接串联 RAW 编码器和 sRGB 解码器，导致 PSNR 下降了 0.46dB。2) 多阶段变体简单地串联了两个 UNets [27]，分别配备了 CID 块和 MCC 块。图像级数据流的损失严重恶化了性能，PSNR 下降了 0.30dB。

表 3. 对解耦和反馈框架的消融研究。Sup. 表示去噪解码器的监督。

Module	Replacement	PSNR	SSIM
RAW Sup.	w/o Sup.	30.48	0.795
	sRGB Sup.	30.20	0.796
Feedback	Single-Stage	30.16	0.792
	Mult-Stage	30.32	0.795
GFM	Conv	30.40	0.795
	w/o Gate	30.35	0.794
	SKFF [40]	30.37	0.795
Original		30.62	0.797

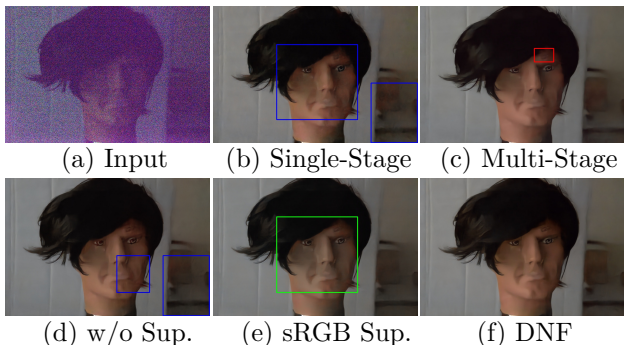


图 7. 我们的方法与消融模型之间的视觉比较 (请放大以获得最佳视图)。蓝色、红色、绿色方框分别表示剩余噪声、细节损失和颜色偏移。

门控机制。采用了三种其他的融合模块：传统的卷积层 (0.22dB↓)，没有门控机制的深度卷积 (0.27dB↓)，以及专门为特征融合设计的 SKFF [10, 40] 模块 (0.25dB↓)。由于门控机制提供的逐像素选择，所提出的方法享有最佳性能。

残差开关机制。如 Tab. 4所示，与完全没有残差快捷连接相比，利用全局残差快捷连接可以提高性能 (0.06dB↑)。然而，在颜色恢复阶段的全局快捷连接会通过引入领域歧义来限制性能 (0.03dB↓)。有无所有局部快捷连接的实验都引入了功能上的矛盾，从而导致不同程度的性能下降 (与 DNF 相比, 0.33dB↓ 和 0.30dB↓)。与另一种实现残差开关机制的方法相比：在去噪时打开开关或在颜色恢复期间关闭开关，CID 块在颜色恢复期间的局部快捷连接提供了更多关于图像内容的信息，从而带来了更高的性能。

与特征级数据流多阶段框架相比。如 Tab. 5所示，与

表 4. 对 CID 块的残差机制的消融研究。Global 和 Local 分别表示编码器和解码器之间的全局残差快捷方式以及 CID 块中的局部快捷方式。RAW 和 sRGB 分别表示 RAW 去噪阶段和颜色恢复阶段。最后一行表示我们提出的 DNF 的实现和性能。

Global		Local		PSNR	SSIM
RAW	sRGB	RAW	sRGB		
✓				30.26	0.794
✓				30.32	0.795
✓	✓			30.29	0.794
✓		✓		30.48	0.794
✓			✓	30.29	0.794
✓			✓	30.62	0.797

表 5. 与其他特征级数据流的比较。Multi-Stage* 表示一个特征级的多阶段框架。

Method	w/o RSM	DNF	Multi-Stage*
PSNR	30.32	30.62	30.46
SSIM	0.794	0.797	0.796

特征级数据流多阶段框架相比，作者的模型提供了最佳性能，这验证了残差开关机制 (RSM) 的有效性。特征级多阶段框架保留了门控融合模块，但涉及两个不同的 RAW 编码器。结果表明，使用作者提出的 RSM，权重共享的编码器可以执行两种不同的功能。此外，噪声估计和信号重建这两个互补的功能相互补充，共同实现了更好的性能。

5. Conclusion

针对 RAW 格式的独特属性，我们提出了一种用于基于 RAW 的低光照图像增强的 DNF 框架。作为一个通用的 pipeline，所提出的 DNF 克服了以前方法的固有限制。领域特定任务解耦消除了单阶段方法引入的领域模糊，而去噪先验反馈则取代了存在有损图像级数据流的多阶段方法。

限制。所提出的框架，也与大多数现有方法一样存在一个限制，即在极低光照条件下，根据曝光时间预定义输入图像的放大比例。在现实世界场景中，估计正常照明是至关重要且困难的。

致谢。本研究得到中国国家重点研发计划 (NO. 2018AAA0100400)、国家自然科学基金 (NO. 62225604) 和中国博士后科学基金 (NO. 2021M701780) 的资助。还得到了 CAAI-华为 MindSpore 开放基金的赞助。

参考文献

- [1] Chunshui Cao, Xianming Liu, Yi Yang, Yinan Yu, Jiang Wang, Zilei Wang, Yongzhen Huang, Liang Wang, Chang Huang, Wei Xu, et al. Look and think twice: Capturing top-down visual attention with feedback convolutional neural networks. In *ICCV*, 2015. 5
- [2] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *CVPR*, 2018. 1, 2, 3, 6, 7
- [3] François Chollet. Xception: Deep learning with depth-wise separable convolutions. In *CVPR*, 2017. 5
- [4] Xingbo Dong, Wanyan Xu, Zihui Miao, Lan Ma, Chao Zhang, Jiewen Yang, Zhe Jin, Andrew Beng Jin Teoh, and Jiajun Shen. Abandoning the bayer-filter to see in the dark. In *CVPR*, 2022. 1, 2, 4, 6, 7
- [5] Shuhang Gu, Yawei Li, Luc Van Gool, and Radu Timofte. Self-guided network for fast image denoising. In *ICCV*, 2019. 1, 2, 6, 7
- [6] Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (GELUs). *arXiv:1606.08415*, 2016. 4
- [7] Haofeng Huang, Wenhan Yang, Yueyu Hu, Jiaying Liu, and Ling-Yu Duan. Towards low light enhancement with raw images. *IEEE TIP*, 2022. 1, 2, 4, 5, 6, 7
- [8] Andrey Ignatov, Luc Van Gool, and Radu Timofte. Replacing mobile camera isp with a single deep learning model. In *CVPR Workshops*, 2020. 2
- [9] Geonwoon Jang, Wooseok Lee, Sanghyun Son, and Kyoung Mu Lee. C2n: Practical generative noise modeling for real-world denoising. In *ICCV*, 2021. 4
- [10] Aupendu Kar, Sobhan Kanti Dhara, Debashis Sen, and Prabir Kumar Biswas. Zero-shot single image restoration through controlled perturbation of koschmieder’s model. In *CVPR*, 2021. 8
- [11] Xin Lai, Zhuotao Tian, Xiaogang Xu, Yingcong Chen, Shu Liu, Hengshuang Zhao, Liwei Wang, and Jiaya Jia. Decouplenet: Decoupled network for domain adaptive semantic segmentation. *arXiv:2207.09988*, 2022. 3
- [12] Mohit Lamba, Atul Balaji, and Kaushik Mitra. Towards fast and light-weight restoration of dark images. *arXiv:2011.14133*, 2020. 1, 2, 6
- [13] Mohit Lamba and Kaushik Mitra. Restoring extremely dark images in real time. In *CVPR*, 2021. 1, 2, 6
- [14] Chongyi Li, Chunle Guo, and Change Loy Chen. Learning to enhance low-light image via zero-reference deep curve estimation. *TPAMI*, 2021. 1
- [15] Chongyi Li, Chunle Guo, Linghao Han, Jun Jiang, Ming-ming Cheng, Jinwei Gu, and Chen Change Loy. Low-Light Image and Video Enhancement Using Deep Learning: A Survey. *TPAMI*, 2021. 1
- [16] Jingyuan Li, Fengxiang He, Lefei Zhang, Bo Du, and Dacheng Tao. Progressive reconstruction of visual structure for image inpainting. In *ICCV*, 2019. 3
- [17] Qilei Li, Zhen Li, Lu Lu, Gwanggil Jeon, Kai Liu, and Xiaomin Yang. Gated multiple feedback network for image super-resolution. *BMVC*, 2019. 3, 5
- [18] Yijun Li, Lu Jiang, and Ming-Hsuan Yang. Controllable and progressive image extrapolation. In *WACV*, 2021. 3
- [19] Zhen Li, Jinglei Yang, Zheng Liu, Xiaomin Yang, Gwanggil Jeon, and Wei Wu. Feedback network for image super-resolution. In *CVPR*, 2019. 3, 5
- [20] Yang Liu, Zhaowen Wang, Hailin Jin, and Ian Wassell. Multi-task adversarial network for disentangled feature learning. In *CVPR*, 2018. 2
- [21] Feifan Lv, Feng Lu, Jianhua Wu, and Chongsoon Lim. Mblen: Low-light image/video enhancement using cnns. In *BMVC*, 2018. 1
- [22] Paras Maharjan, Li Li, Zhu Li, Ning Xu, Chongyang Ma, and Yue Li. Improving extreme low-light image denoising via residual learning. In *ICME*, 2019. 1, 2, 5, 6
- [23] Junichi Nakamura. *Image sensors and signal processing for digital still cameras*. CRC press, 2017. 4
- [24] Seonghyeon Nam, Youngbae Hwang, Yasuyuki Matsushita, and Seon Joo Kim. A holistic approach to cross-channel image noise modeling and its application to image denoising. In *CVPR*, 2016. 4
- [25] Kamyar Nazari, Eric Ng, Tony Joseph, Faisal Qureshi, and Mehran Ebrahimi. Edgeconnect: Structure guided image inpainting using edge prediction. In *ICCV Workshops*, 2019. 3
- [26] Yurui Ren, Xiaoming Yu, Ruonan Zhang, Thomas H Li, Shan Liu, and Ge Li. Structureflow: Image inpainting via structure-aware appearance flow. In *ICCV*, 2019. 3

- [27] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *ICML*, 2015. 2, 7
- [28] Eli Schwartz, Raja Giryes, and Alex M Bronstein. Deepisp: Toward learning an end-to-end image processing pipeline. *IEEE TIP*, 2018. 2
- [29] Abhinav Shrivastava, Rahul Sukthankar, Jitendra Malik, and Abhinav Gupta. Beyond skip connections: Top-down modulation for object detection. *arXiv:1612.06851*, 2016. 5
- [30] Yuzhi Wang, Haibin Huang, Qin Xu, Jiaming Liu, Yiqun Liu, and Jue Wang. Practical deep raw image denoising on mobile devices. In *ECCV*, 2020. 2, 5
- [31] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE TIP*, 2004. 6
- [32] Kaixuan Wei, Ying Fu, Jiaolong Yang, and Hua Huang. A physics-based noise formation model for extreme low-light raw denoising. In *CVPR*, 2020. 5
- [33] Kaixuan Wei, Ying Fu, Yinqiang Zheng, and Jiaolong Yang. Physics-based noise modeling for extreme low-light photography. *TPAMI*, 2021. 1, 2, 4
- [34] Jun Xu, Lei Zhang, David Zhang, and Xiangchu Feng. Multi-channel weighted nuclear norm minimization for real color image denoising. In *ICCV*, 2017. 4
- [35] Ke Xu, Xin Yang, Baocai Yin, and Rynson WH Lau. Learning to restore low-light images via decomposition-and-enhancement. In *CVPR*, 2020. 1, 2, 4, 6, 7
- [36] Xuejun Yan, Hongyu Yan, Jingjing Wang, Hang Du, Zhihong Wu, Di Xie, Shiliang Pu, and Li Lu. Fb-net: Feedback network for point cloud completion. In *ECCV*, 2022. 3
- [37] Amir R Zamir, Te-Lin Wu, Lin Sun, William B Shen, Bertram E Shi, Jitendra Malik, and Silvio Savarese. Feedback networks. In *CVPR*, 2017. 3
- [38] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. *arXiv:2111.09881*, 2021. 4
- [39] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Cycleisp: Real image restoration via improved data synthesis. In *CVPR*, 2020. 2
- [40] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for real image restoration and enhancement. In *ECCV*, 2020. 1, 8
- [41] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *CVPR*, 2021. 2
- [42] Syed Waqas Zamir, Aditya Arora, Salman Khan, Fahad Shahbaz Khan, and Ling Shao. Learning digital camera pipeline for extreme low-light imaging. *Neurocomputing*, 2021. 1, 2, 5
- [43] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE TIP*, 2017. 5
- [44] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018. 6
- [45] Yi Zhang, Hongwei Qin, Xiaogang Wang, and Hongsheng Li. Rethinking noise synthesis and modeling in raw denoising. In *ICCV*, 2021. 2
- [46] Muming Zhao, Jian Zhang, Chongyang Zhang, and Wenjun Zhang. Leveraging heterogeneous auxiliary tasks to assist crowd counting. In *CVPR*, 2019. 2
- [47] Minfeng Zhu, Pingbo Pan, Wei Chen, and Yi Yang. EEMEFN: Low-light image enhancement via edge-enhanced multi-exposure fusion network. In *AAAI*, 2020. 1, 2, 4, 5, 6, 7
- [48] Yunhao Zou and Ying Fu. Estimating fine-grained noise model via contrastive learning. In *CVPR*, 2022. 2