计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
2025,34(3):27−39 [doi: 10.15888/j.cnki.csa.009775] [CSTR: 32024.14.csa.009775]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于 Retinex 理论的双重注意力 Transformer 的低光照图像增强①

李　佳, 王　婷, 杨文杰, 王弘扬

(成都信息工程大学 计算机学院, 成都 610225)
通信作者: 王　婷, E-mail: wangting@cuit.edu.cn

摘　要: 在低光照图像增强的研究中, 虽然现有技术在提升图像亮度方面取得了进展, 但细节恢复不足和颜色失真等问题仍然存在. 为了解决这些问题, 本文提出一种基于 Retinex 理论具有双重注意力的 Transformer 增强网络——DARFormer. 该网络由光照估计网络和损坏修复网络两部分组成, 旨在提升低光照图像的亮度, 同时保留更多的细节并防止颜色失真. 光照估计网络是基于图像先验来估计亮度映射项, 用于低光照图像亮度增强; 损坏修复网络则优化亮度增强后的图像质量, 采用具有空间注意力和通道注意力的 Transformer 架构. 在 LOL_v1、LOL_v2 和 SID 公开数据集上进行实验表明: 与主流的增强方法相比, DARFormer 在定量和定性指标上取得了更好的增强结果.

关键词: 图像增强; 低照度图像; 空间注意力; 通道注意力; Transformer

引用格式: 李佳,王婷,杨文杰,王弘扬.基于 Retinex 理论的双重注意力 Transformer 的低光照图像增强.计算机系统应用,2025,34(3):27–39. http://www.c-s-a.org.cn/1003-3254/9775.html

Dual-attention Retinex Theory-based Transformer for Low-light Image Enhancement

LI Jia, WANG Ting, YANG Wen-Jie, WANG Hong-Yang

(School of Computer Science, Chengdu University of Information Technology, Chengdu 610225, China)

Abstract: In the research on low-light image enhancement, although existing technologies make progress in improving image brightness, the issues of insufficient detail restoration and color distortion still persist. To tackle these problems, this study introduces a dual-attention Retinex-based Transformer network—DARFormer. The network consists of an illumination estimation network and corruption restoration network, which aims to enhance the brightness of low light images while preserving more details and preventing color distortion. Illumination estimation network uses an image prior to estimate the brightness mapping, which is used to enhance the brightness of low-light images. The corruption restoration network optimizes the quality of the brightness-enhanced image, employing a Transformer architecture with spatial attention and channel attention. Experiments carried out on public datasets LOL_v1, LOL_v2, and SID show that compared with the prevalent enhancement methods, DARFormer achieves better enhancement results in quantitative and qualitative indicators.

Key words: image enhancement; low-light image; spatial attention; channel attention; Transformer

在图像采集过程中, 由于环境光线不足、拍摄设备的局限性以及设备参数设置不当等原因, 常会导致采集到的图像出现亮度低、对比度差、细节丢失和噪声增多等问题. 低光图像增强方法 (low-light image enhancement, LLIE) 正是针对这些问题, 通过技术手段对图像进行改善, 旨在恢复图像中的光照并减少由于

① 基金项目: 四川省科技厅重点研发项目 (2023YFG0099, 2023YFG0261)
收稿时间: 2024-08-02; 修改时间: 2024-08-27; 采用时间: 2024-09-10; csa 在线出版时间: 2025-01-16
CNKI 网络首发时间: 2025-01-17

光照恢复过程中引入的噪声、伪影和克服颜色失真等问题. 作为一种基础的视觉处理任务, 低光图像增强对于提升高级视觉任务的性能至关重要, 比如夜间对象识别[1]、目标检测[2]和动驾驶等场景. 因此, 深入研究和开发有效的 LLIE 方法, 对于推动相关领域的技术进步具有显著的应用价值和实际意义.

近年来, 人们开发了多种 LLIE 方法, 主要分为两大类: 传统方法和基于深度学习的方法. 传统方法包括直方图均衡化[3]、伽马校正[4]和 Retinex 理论[5]方法. 直方图均衡化和伽马校正都是基于分布映射的方法, 它们通过映射输入图像的低光照分布, 放大暗区域的像素值, 以提升图像的整体亮度. 由于此类方法忽略了图像中的语义信息, 生成的结果往往会存在曝光不当、细节丢失和颜色失真等缺陷. Retinex 理论是一种基于模型优化的方法, 它认为图像可以被分解为反射分量和光照分量. 基于 Retinex 模型的方法往往依赖于人工设定的模型参数, 这限制了其在不同场景下的性能一致性.

与传统的方法相比, 基于卷积神经网络 (convolutional neural network, CNN) 的方法能够理解和利用图像中的语义信息, 并且能够从大量数据中自动学习低光照图像与正常光照图像之间的复杂映射, 无需手动设计参数. 研究者们使用 CNN 来实现 LLIE, 主要的代表方法有: EnlightenGAN[6]、URetinex-Net[7]、R2RNet[8]、EEMEFN[9]、Bread[10]和 MSRNet[11]. 然而, 受感受野 (即网络中单个神经元能够“看到”的图像区域) 的限制, 基于 CNN 的方法在捕获全局信息方面存在局限性. 这导致图像增强结果存在信息丢失的问题. Transformer[12]模型在 2017 年被首次提出, 它弥补了基于 CNN 的方法的缺陷. Transformer 模型的核心优势在于其自注意力机制, 它通过单个自注意力层就可以获得远程依赖关系, 这一方法提供了强大的全局上下文建模能力. 然而, 标准的 Transformer 模型的计算复杂度与输入序列的长度呈平方关系, 这阻碍了其在实际应用中的发展. 因此, 混合 CNN 与 Transformer 网络架构被提出[13–15], 成为该方向的先驱. 尽管混合 CNN 与 Transformer 有效地用于 LLIE, 但仍然依赖于繁重的 CNN 主干, 并且为了减少 Transformer 计算量, 一般采取的措施是控制输入的序列长度在可接受的范围内, 通常采用 CNN 进行下采样来达到目的. 这种方式不可避免地导致信息丢失.

为了克服现有方法存在的细节丢失和颜色失真的问题, 本文提出基于 Retinex 理论具有双重注意力的 Transformer 增强网络 (dual-attention Retinex-based Transformer, DARFormer). DARFormer 由光照估计网络和损坏修复网络组成. 光照估计网络负责接收低光照图像, 并结合图像先验信息来实现亮度的增强. 损坏修复网络对亮度增强的图像进行进一步的处理, 以修复图像存在的损坏, 在这整个修复过程受到亮度特征图的指导, 以确保修复效果与光照条件相适应. 为了在增强图像的同时保留更多的细节, 损坏修复网络采用了一种结合空间注意力和通道注意力的 Transformer 架构. 这种架构以 U-Net 的形式设计, 使得网络能够在空间和通道两个维度上进行更精确的特征提取和增强, 从而在提升图像质量的同时, 最大程度地保留图像的细节和色彩. 本文的主要创新点如下.

(1) 提出了基于 Retinex 理论具有双重注意力的 Transformer 方法 DARFormer, 用于低光照图像增强.

(2) 设计一种新颖的光照指导的双重注意力 Transformer 模块, 该模块集成了空间注意力和通道注意力机制, 旨在对亮度增强后的图像进行损坏修复, 同时减少细节丢失和颜色失真.

(3) 在 LOL_v1、LOL_v2_real、LOL_v2_synthetic 和 SID 数据集上进行大量的定量和定性实验, 证明了本文提出的方法在多个评价指标上普遍优于现有的主流方法.

1 相关工作

1.1 基于 Retinex 模型的低光照图像增强

Retinex 理论为 LLIE 提供了直接的物理描述, 根据这一理论, 低光照图像和正常光照图像之间存在一种点除关系, 即通过低光照图像点除光照分量可以获得接近正常光照的图像. 一些研究者利用传统的 Retinex 模型用于 LLIE[16–18]. 例如, Wang 等人[17]基于 Retinex 模型设计一个光照滤波器, 旨在增强非均匀光照图像, 同时保留图像的自然度, 但该方法的性能并不稳定, 时常出现细节缺失和亮度不足的现象. 考虑到黑暗场景下隐藏的噪声和伪影, Li 等人[16]结合 Retinex 模型和去噪模型, 设计基于增广 Lagrange 乘子的 ADM 算法来估计噪声图. 然而, 该方法仍然会出现亮度不足和对比度不足的结果. 此外, 一些研究者侧重于光照估计. 例如 Guo 等人[18]将图像的 R、G、B 这 3 个颜色通道中

像素最大值作为初始光照, 后利用保留边缘的平滑方法[19]优化初始光照, 最后利用 Retinex 理论合成增强图像. 该方法取得了良好的性能, 但是在某些情况会出现曝光过度的现象. 总之, 基于传统的 Retinex 模型的核心工作是人工设计的先验, 通常需要针对实际情况进行手动调整参数, 这导致此类方法具有较差的泛化能力. 此外, 不准确的先验会导致增强的结果产生伪影, 过度曝光或曝光不足等问题.

随着深度学习的崛起, 基于深度学习的 LLIE 方法[6–11,13,14,20–24]应运而生, 克服了基于 Retinex 模型方法的局限性. 受 Retinex 理论的启发, Wei 等人[24]提出了基于 Retinex 的低光照图像增强网络 RetinexNet, 由分解网络和增强网络组成, 其中增强网络包括光照估计和反射分量估计两个模块组成. 此方法是多阶段训练, 对分解后的光照分量和反射分量分别进行调整, 这导致生成的结果会出现未知的伪影, 细节丢失的现象. 进一步地, Jiang 等人[8]在 Wei 等人的工作基础上, 提出从真实低光照图像到真实正常光照图像的增强方法 R2R-Net, 该网络与 RetinexNet 类似, 在反射分量重建部分增加了细节重建模块, 该模块使用快速傅里叶变换提取图像的频域信息, 从而实现对图像细节的保留. 然而, 由于该方法仍然是在分解后对反射分量进行恢复, 细节丢失的问题依然存在, 并且基于深度学习的方法缺乏可解释性.

为了有效防止细节丢失并提高可解释性和灵活性, Wu 等人[7]结合基于模型和基于学习方法的优势, 提出了一种基于 Retinex 模型的深度展开网络 URetinex-Net, 该网络继承了基于模型方法的灵活性和可解释性. 此外, 该方法利用 CNN 的强大建模能力对反射分量、光照分量同时进行优化增强. 然而, URetinex-Net 忽略了噪声的影响, 并且以不明确的方式处理照明和反射先验, 这导致产生不切实际的结果, 并伴有颜色失真. 与以往方法图像到图像的预测不同, Wang 等人[23]提出图像到光照映射的增强网络, 但该方法没有考虑到图像的干扰信息, 这导致增强后的图像存在很大的噪声且颜色失真. 总之, 这些方法普遍存在以下限制: 大多依赖准确的基于 Retinex 理论的分解, 这在复杂的光照环境下具有挑战性; 其次, 基于 Retinex 理论的深度学习的方法往往采用分阶段的训练策略, 这使得训练过程变得较为复杂和耗时; 最后, 卷积核的感受野限制了网络捕捉图像全局信息的能力.

1.2 基于 Transformer 的低光照图像增强

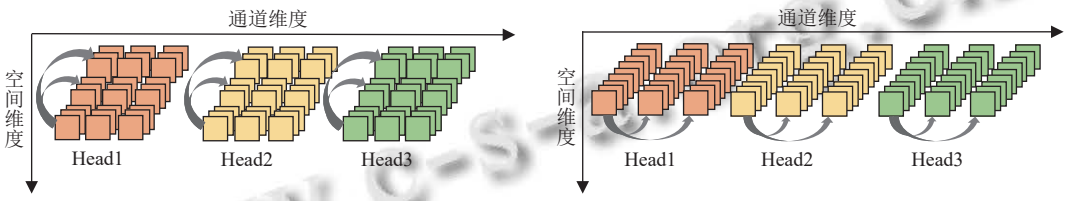
ViT[25]是首个将 Transformer 引入到视觉领域的模型, Transformer 模型及其变体已经成功应用于图像识别[25]、对象检测[26]、超分辨率[27]、分割[28]以及图像恢复[29]等多个视觉任务中. 近年来, 一些基于 CNN 与 Transformer 的混合网络被提出[13–15,30]. 其中大部分方法采用双分支结构[13–15], 即局部分支和全局分支. 局部分支使用 CNN 来捕获局部信息, 全局分支采用 Transformer 来构建全局上下文. 例如, Xu 等人[13]提出了一个基于 Transformer 的信噪比感知网络框架 SNR-Net. 该框架在处理高信噪比区域时, 局部分支通过残差网络结构专注于利用局部细节信息进行图像增强, 以突出清晰度和纹理. 对于信噪比较低的区域, 全局分支发挥作用, 利用 Transformer 的自注意力机制, 捕捉图像中的上下文信息, 从而有效地增强这些区域, 改善整体的视觉质量和亮度. 然而, 这类方法为了减少 Transformer 的计算量, 往往通过压缩输入序列长度方式来达到目的, 这可能导致关键信息丢失, 影响细节保留.

为了更好地获取细节信息, Cai 等人[22]提出一种基于 Retinex 理论的 Transformer 模型 Retinexformer, 由光照估计和损坏恢复两部分组成. 该方法与以往基于 Retinex 理论的深度学习方法的多阶段训练过程不同, 它通过单一训练阶段来实现图像增强. 在 Retinexformer 模型中, 采用 Transformer 结构来处理经过亮度增强后的图像, 以实现损坏修复. 然而, 该模型为了减少的计算量, 对自注意力机制进行修改, 使复杂度与输入序列长度呈线性关系, 这导致全局上下文信息在该方法中只被部分捕获.

1.3 注意力机制

SE (squeeze-and-excitation) 注意力[31]是由 Hu 等人提出的一种通道注意力, 在图像分类任务上取得了优异表现. 这种机制允许网络自适应地调整不同通道的特征响应, 突出重要信息的同时抑制不重要的特征. 然而, SE 注意力忽略了特征映射上的空间重要性. 因此, Woo 等人[32]在 SE 注意力基础上, 进一步提出了 CBAM (convolutional block attention module) 注意力, CBAM 引入空间注意力, 结合空间和通道两个维度, 能够更全面地捕获特征信息. 众所周知, Transformer 的自注意力机制[25]是聚合全局信息的有效方法, 因其计算复杂度较高, 这限制了在高分辨率图像处理中的发展. 因此, El-Nouby 等人[33]提出了协方差注意力 (cross-covariance

attention, XCA), 可以高效处理高分辨率图像, 该注意力是转置版的自注意力, 通过转置查询向量和键向量, 将计算复杂度降低到与输入序列长度呈线性关系. 随后, Ding 等人[34]提出了 DaViT, 一种包含空间注意力和通道注意力的 Transformer 架构, 在图像分类、目标检测、图像分割任务上进行实验, 均取得了最好的性能. 类似地, Azad 等人[28]结合空间注意力[35]和通道注意力提出一种方法用于医学图像分割, 称为 DAE-former, 达到了领域内最佳水平. 这些研究结果证明了空间注意力和通道注意力在提升图像处理性能方面的互补性和有效性. 尽管这些注意力机制在多个领域都显示出了其价值, 但在低光照图像处理领域, 空间和通道注意力的重要性尚未得到充分研究. 因此, 本文提出了一种新型的光照指导的双重注意力 Transformer 模块, 该模块集成了空间注意力和通道注意力, 旨在更全面地捕获特征间的全局关系, 从而有效提升图像的增强效果. 图 1(a) 和图 1(b) 分别显示了图像的输入序列的空间注意和通道注意力.

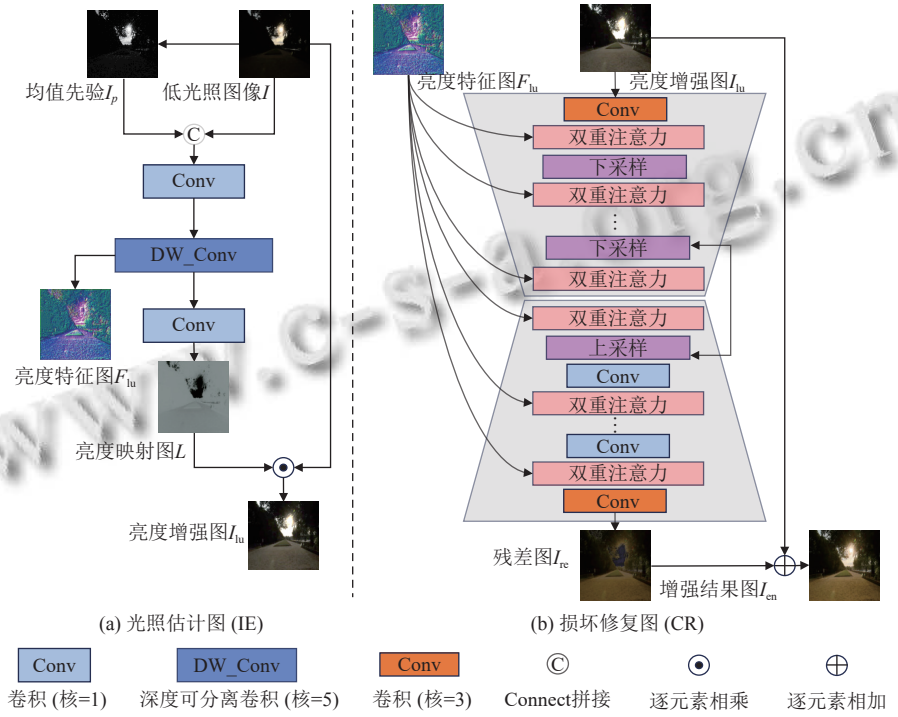


(a) 空间注意力 (b) 通道注意力

图 1 输入序列的空间注意力和通道注意力的可视化

2 方法

本文提出的基于 Retinex 理论具有双重注意力的 Transformer 网络 DARFormer, 如图 2 所示. DARFormer 由图 2(a) 光照估计 (illumination estimation, IE) 网络和图 2(b) 损坏修复 (corruption restoration, CR) 网络两部分组成.



(a) 光照估计图 (IE) (b) 损坏修复图 (CR)

图 2 DARFormer 的总体结构

IE 采用简单的 CNN 网络结构, 负责对低光照图像进行亮度增强. CR 由编码器和解码器组成, 使用卷积层进行下采样和上采样, 并通过跳跃连接将编码器与解码器的特征进行融合, 以进一步处理亮度增强后的

图像, 修复其中存在的损坏. 以光照指导的具有双重注意力的 Transformer 模块作为核心模块, 它通过集成空间注意力和通道注意力来全面捕捉特征间的全局信息, 从而实现效果提升. 首先, 将低光照图像与平均先验图像 I_p 进行通道连接, 输入 IE 中. IE 负责生成亮度增强图像 (light-up image, I_{lu}) 和亮度特征图 (light-up feature, F_{lu}). 接下来, 将 I_{lu} 和 F_{lu} 作为输入, 送入 CR 中. CR 的任务是去除亮度增强过程中引入的干扰信息, 从而恢复图像的细节和色彩, 进一步提升图像质量.

2.1 基于 Retinex 理论的框架

Retinex 理论认为图像 $I\in\mathbb{R}^{H\times W\times 3}$ 可以被分解为两个部分: 反射分量 $R\in\mathbb{R}^{H\times W\times 3}$ 和光照分量 $L\in\mathbb{R}^{H\times W\times 1}$. 在数学中的表达如式 (1) 所示:

$$I = R\cdot L \tag{1}$$

其中, $\cdot$ 表示的是逐元素相乘. 反射分量 R 代表了图像的固有属性, 如图像的颜色、纹理等, 是一个相对稳定的量. 而光照分量 L 则代表了场景的光照条件. 传统的基于 Retinex 理论的 LLIE 方法通常假设低光照图像的反射分量 R 可以近似作为正常光照图像, 从而可以通过将低光照图像的光照分量从低光照图像中分离出来恢复图像. 然而此类方法忽略了低光照条件下反射分量受噪声和伪影的影响. 为了解决这一问题, 引入干扰项 $\tilde{L}\in\mathbb{R}^{H\times W\times 1}$ 和 $\tilde{R}\in\mathbb{R}^{H\times W\times 3}$, 分别作为光照分量和反射分量的干扰项, 然后, 通过消除图像中的干扰项来达到增强效果. 因此, 将低光照图像 I 重新表示如式 (2) 所示:

$$I = \left(R+\tilde{R}\right)\cdot\left(L+\tilde{L}\right) = R\cdot L + R\cdot\tilde{L} + \tilde{R}\cdot L + \tilde{R}\cdot\tilde{L} \tag{2}$$

根据传统的基于 Retinex 理论的假设, 将 R 视为正常曝光的图像. 因此, 引入亮度映射图 $\bar{L}\in\mathbb{R}^{H\times W\times 1}$ 来进行亮度增强, 在等式两边同乘 $\bar{L}$, 使得 $L\cdot\bar{L}=1$, 因此得到亮度增强图 I_{lu}, I_{lu} 表示如式 (3) 所示:

$$I_{\text{lu}} = I\cdot\bar{L} = R + R\cdot\tilde{L}\cdot\bar{L} + \tilde{R} + \tilde{R}\cdot\tilde{L}\cdot\bar{L} \tag{3}$$

然后, 用 C 表示亮度增强后图像包含干扰的所有项, 简化式 (3) 得到式 (4).

$$I_{\text{lu}} = R + C \tag{4}$$

通过消除所有干扰项就能得到增强结果. 因此本文 DARFormer 方法可以表示为式 (5)、式 (6):

$$(I_{\text{lu}}, F_{\text{lu}}) = IE\left(I, I_p\right) \tag{5}$$

$$I_{\text{en}} = CR(I_{\text{lu}}, F_{\text{lu}}) \tag{6}$$

其中, IE 为光照估计网络, CR 为损坏恢复网络, $I_p\in\mathbb{R}^{H\times W\times 1}$ 代表图像平均先验, I_p 如式 (7) 所示, $I_{\text{lu}}\in\mathbb{R}^{H\times W\times 3}$ 代表亮度增强后的图像, $F_{\text{lu}}\in\mathbb{R}^{H\times W\times 40}$ 代表亮度特征图, $I_{\text{en}}\in\mathbb{R}^{H\times W\times 3}$ 代表增强结果图像.

$$I_p = \frac{1}{C}\sum_{c=1}^{C} I(H,W,C) \tag{7}$$

2.2 光照估计网络

如图 2(a) 结构所示, 首先, 将低光照图像 I 和平均图像先验 I_p 在通道的维度上进行连接. 接着, 通过一个 3 层卷积神经网络进行特征提取. 首先, 使用一个 1×1 卷积来初步融合 I 和 I_p 的特征. 随后, 采用一个 5×5 深度可分离卷积来捕捉不同区域的光照变化, 生成亮度特征图 F_{lu}. 最后, 使用一个 1×1 卷积来生成一个亮度映射图 $\bar{L}$. 最终, 将这个亮度映射图 $\bar{L}$ 和输入图像 I 进行逐元素相乘得到亮度增强图像 I_{lu}.

2.3 损坏修复网络

如图 2(b) 结构所示, 损坏修复网络采用以光照指导的双重注意力 Transformer 模块为基础块的编码器和解码器架构. 编码器负责下采样, 而解码器则执行上采样, 两者在结构上呈对称性. 首先, 将亮度增强图像 I_{lu} 通过一个 3×3 卷积, 进行通道维度的映射, 以匹配亮度特征图 F_{lu} 的通道数. 随后, 将 I_{lu} 和 F_{lu} 输入编码器中. 编码器由 3 层堆叠的编码器块组成, 每个编码器块包含若干个双重注意力 Transformer 层和一个 4×4 卷积, 在双重注意力 Transformer 层中, 串联的光照指导的多头注意力和通道注意力模块共同用于特征提取, 这意味着每个通道序列包含了整个图像的空间信息. 图 3 显示了双重注意力 Transformer 的结构细节. 采用 4×4 卷积执行下采样, 其功能是降低输入特征的空间分辨率至原尺寸的一半, 同时将特征的通道数翻倍. 这一过程允许网络在不同层次上捕捉并整合多尺度的特征信息. 解码器与编码器结构类似, 使用 4×4 的卷积进行上采样, 将特征图的空间尺寸放大至原尺寸的两倍, 同时将通道数减半. 通过跳跃连接与编码器层的特征沿通道维度进行拼接, 实现了特征的整合. 紧接着, 利用一个 1×1 卷积对拼接后的特征进行融合, 以增强特征的表达能力. 融合后的特征随后输入到双重注意力的 Transformer 模块中, 该模块进一步提炼和优化特征. 最终, 解码器的输出通过一个 3×3 卷积, 将其通道数映射为 3 通道的 RGB 格式, 得到残差图 I_{re}, 然后将残差图 I_{re} 和亮度增强图像 I_{lu} 进行残差相加, 得到最终的增强结果图 I_{en}.

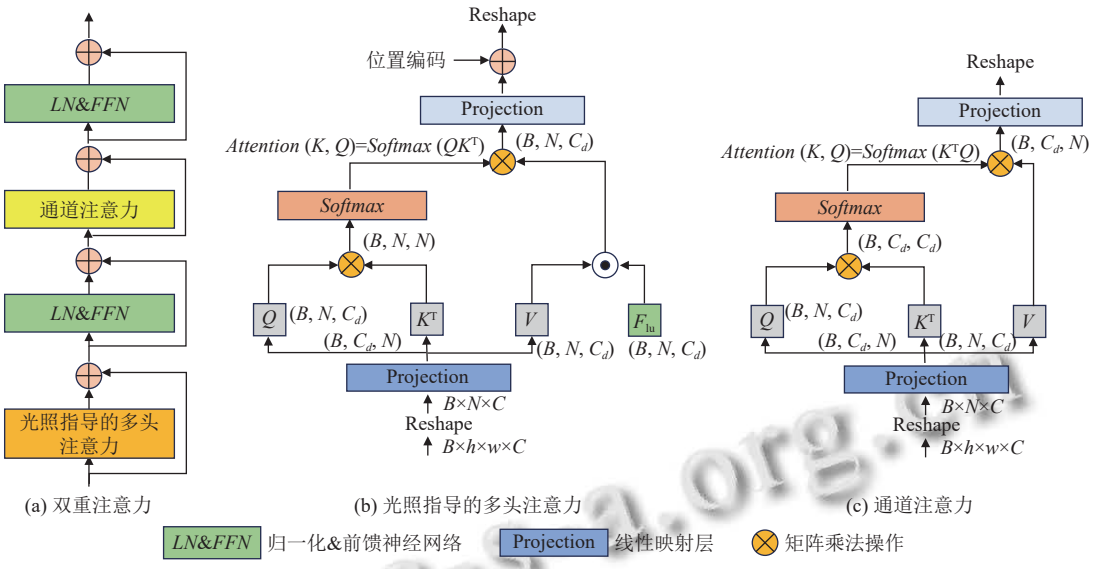


图 3 双重注意力的结构图

2.3.1 光照指导的多头注意力

标准的 Transformer 自注意机制如式 (8) 所示:

$$\begin{cases} A(Q,K,V)=Concat\left(head_1,\cdots,head_{N_h}\right) \\ head_i=Attention\left(Q_i,K_i,V_i\right) \\ Attention\left(Q_i,K_i,V_i\right)=Softmax\left[\dfrac{Q_i(K_i)^{\mathrm{T}}}{\sqrt{C_h}}\right]V_i \end{cases} \tag{8}$$

其中, Q, K, V 分别代表查询、键和值向量; $Q_i=X_iW_i^Q$, $K_i=X_iW_i^K$, $V_i=X_iW_i^V$, Q_i, K_i和V_i代表第 i 个头的视觉特征, 维度都是$\mathbb{R}^{HW\times C_d}$; W_i^Q, W_i^K 和W_i^V 分别表示第 i 个头的 Q, K, V 的映射权值, 维度为$\mathbb{R}^{C_d\times C_d}$. $X=Concat\left(X_1,\cdots,X_{N_h}\right)$, X 代表总的输入特征的序列大小, 其中 X_i 的维度是$\mathbb{R}^{HW\times C_d}$, N_h 表示多头注意力的头数, $N_h\times C_d=C$, C表示的是输入特征的总的通道数量.

本文提出的光照指导的多头注意力, 在一个标准的 Transformer 的自注意力机制的基础上, 给值向量添加一个光照特征权重 F_{lu}, 把 F_{lu} 分成N_h个头, 如式 (9) 所示, 使F_{lu_i}符合V_i的维度, 通过$V_i\times F_{\mathrm{lu}_i}$操作进行光照指导. N_h是由输入特征的通道数和 $feat$ 值动态确定的, 如式 (10) 所示, $feat$ 表示第 1 次下采样前的特征维度. 光照指导的多头注意力如式 (11) 所示:

$$F_{\mathrm{lu}}=Concat\left(F_{\mathrm{lu}_1},\cdots,F_{\mathrm{lu}_{N_h}}\right) \tag{9}$$

$$N_h=\frac{C}{feat} \tag{10}$$

$$Attention\left(Q_i,K_i,V_i\right)=Softmax\left[\frac{Q_i(K_i)^{\mathrm{T}}}{a_i}\right]\left(V_i\times F_{\mathrm{lu}_i}\right) \tag{11}$$

其中, a_i是一个可学习的参数. 光照指导的多头注意力采用标准的 Q 和 K 的点积操作, 以捕捉输入特征映射的空间信息. 更多的细节如图 3(b) 所示.

2.3.2 通道注意力

通道注意力的细节如图 3(c) 所示. 通过转置查询向量和键向量的操作, 使其沿着通道维度运行, 而不是像自注意力一样沿着空间维度运行. 通道注意力如式 (12)、式 (13) 所示:

$$T(Q,K,V)=VA_{XC}(K,Q) \tag{12}$$

$$A_{XC}(K,Q)=Softmax\left(K^{\mathrm{T}}Q/\tau\right) \tag{13}$$

其中, A_{XC}表示的是通道注意力. 并且引入可学习的温度参数 τ, 这个参数在 $Softmax$ 函数之前对内积结果进行缩放, 允许模型在生成注意力权重时有更大的灵活性. 通过调整温度参数, 可以控制注意力分布的集中程度或平滑程度. 将通道分为 8 个组, 并在每个组中进行自关注, 设置 N_d=8, C_d=C/8. 通过这种方式, 可以实现跨通道进行交互.

2.3.3 双重注意力 Transformer 网络

双重注意力结构如图 3(a) 所示. 由光照指导的多头注意力和通道注意力组成. 在这两个注意力的上方, 包括一个 LN (layer normalization) 层和一个 FFN (feed-forward network) 层. 其中 Add 表示残差连接, 用于防止网络退化. LN 用于对每一层的激活值进行归一化. FFN 层代表前馈网络层, 是由一系列卷积层和 $GELU$ 激活函数来实现的. 双重注意 Transformer 用数学公式

表示如式 (14)–式 (17) 所示:

$$S_{\text{block}}(X,Q_s,K_s,V_s)=A_s(Q_s,K_s,V_s\times F_{\text{lu}})+X \tag{14}$$

$$T=FFN1(S_{\text{block}})=FFN(LN(S_{\text{block}}))+S_{\text{block}} \tag{15}$$

$$C_{\text{block}}(T,Q_c,K_c,V_c)=A_c(Q_c,K_c,V_c)+T \tag{16}$$

$$Y=FFN2(C_{\text{block}})=FFN(LN(C_{\text{block}}))+C_{\text{block}} \tag{17}$$

其中, A_s 和 A_c 分别表示光照指导的多头注意力和通道注意力. S_{block} 和 C_{block} 分别表示光照指导的多头注意力块和通道注意力块. Q_s、K_s、V_s 分别表示输入特征 X 映射出的查询、键和值向量. Q_c、K_c、V_c 分别表示输入特征 T 映射出的查询、键和值向量. $FFN1$ 和 $FFN2$ 分别表示光照指导的多头注意力和通道注意力的前馈网络, FFN 用数学公式表示如式 (18) 所示:

$$FFN(X)=Conv(GELU(DW\_Conv(GELU(Conv(X))))) \tag{18}$$

其中, $Conv$ 表示 1×1 的卷积, $GELU$ 指的是 $GELU$ 激活函数[36], DW_Conv 指的是一个深度可分离卷积.

2.4 损失函数

本文采用平均绝对误差 (mean absolute error, MAE) 作为损失函数. MAE 定义为预测值与真实值之间差的绝对值的平均, 数学表达式如式 (19) 所示:

$$MAE=\frac{1}{n}\sum_{i=1}^{n}|y_i-\hat{y}_i| \tag{19}$$

其中, y_i 表示真实图像的第 i 个像素值, 而 $\hat{y}_i$ 表示增强结果的第 i 个像素值, n 为总像素数. 该函数对于异常值具有鲁棒性. 低光图像中有大量的噪声和不准确的亮度, 使用 MAE 能够使增强结果的质量更加稳定.

3 实验

3.1 数据集

本文方法在公开数据集 LOL_v1[24]、LOL_v2[37] 和 SID[38] 上进行性能评估. 这些数据集都是成对的数据. LOL 数据集是用于低光图像增强的 RGB 基准数据集, 有 LOL_v1 和 LOL_v2 两个版本. 其中 LOL_v1 数据集, 包含室内、室外等多个场景, 458 对图像用于训练, 15 对图像用于测试. LOL_v2 是对 LOL_v1 的补充, LOL_v2 分为 LOL_v2_real 和 LOL_v2_synthetic 两个子集. LOL_v2_real 子集是通过改变 ISO 和曝光时间拍摄得到的, 589 对图像用于训练, 100 对用于测试. LOL_v2_synthetic 子集是基于 RAW 图像合成得到的. 900 对图像用于训练, 100 对用于测试. SID 数据集是基于 RAW 格式的图像, 包含富士和索尼两个子集. 本文采用索尼相机捕获的子集, 该数据集包含室内和室外两个场景, 其中室内图像是在极暗的条件下拍摄的, 室外图像通常是在路灯或者月光等照明条件下拍摄的. 本文使用 SID 提供的脚本将低光照图像从 RAW 转换为 RGB 格式. 2 099 对图像用于训练, 598 对用于测试.

3.2 细节说明

本文所提 DARFormer 方法使用 PyTorch (CUDA 11.3, Python 3.8, PyTorch 1.11) 以端到端的方式实现, 所有的实验都是在 Linux 系统下进行训练和测试的, 使用 1 张 RTX 4090 的显卡. 实验中本文将 patch_size 的大小设置为 128×128, batch_size 设置为 8, 使用 Adam[39] 作为优化器, 设置动量项参数为 0.9, RMSprop 参数为 0.999. 使用余弦退火方法来动态设置学习率. 初始的学习率为 2×10^{-4}, 随后稳步降低到 1×10^{-4}. 实验的目的是最小化增强结果和真实图像之间的平均绝对误差.

3.3 评价

3.3.1 定量评价

本文用到的图像质量评价指标有峰值信噪比 (peak signal to noise ratio, PSNR)[40]、结构相似度 (structural similarity index, SSIM)[41]、学习感知图像块相似度 (learned perceptual image patch similarity, LPIPS)[42] 和均方根误差 (root mean square error, RMSE). SSIM 和 PSNR 是图像增强中最常用的两个指标, PSNR 是一种衡量信噪比的方法, 计算增强图像和真实图像之间的峰值信噪比, PSNR 值越高表示图像清晰度越高, 细节保留能力更强, 图像质量越好. SSIM 分别从亮度、对比度、结构 3 个方面度量图像的相似性. SSIM 值越大, 表示图像越相似, 图像更符合人眼视觉特性. LPIPS 是基于深度学习方法的, 可以更好地捕捉人类视觉系统对图像质量的感知, LPIPS 值越低表示图像越相似. RMSE 通过计算两张图像之间像素的差值的平方, 然后求均值再开方, 是一种非常直观的误差度量方法, RMSE 值越小, 表示图像越相似.

表 1 展示了本文方法 DARFormer 与其他 7 种方法在 4 个数据集上的定量结果. 其中“↑”表示值越高越好, “↓”表示值越小越好. 加粗的字体为最优值. 表 1 中所有数据都是使用原论文提供的公开代码进行训练和测试获得.

表 1 各方法在 4 个数据集上的 PSNR、SSIM、RMSE、LPIPS 结果

方法	LOL_v1				LOL_v2_real				LOL_v2_synthetic				SID			
	PSNR (dB)↑	SSIM↑	RMSE↓	LPIPS↓	PSNR (dB)↑	SSIM↑	RMSE↓	LPIPS↓	PSNR (dB)↑	SSIM↑	RMSE↓	LPIPS↓	PSNR (dB)↑	SSIM↑	RMSE↓	LPIPS↓
LIME[18]	14.02	0.57	10.04	0.37	17.14	0.57	10.21	0.34	17.63	0.78	9.95	0.2	12.4	0.17	10.08	0.93
Zero-Dce[20]	14.86	0.57	10.22	0.34	18.06	0.59	10.02	0.31	17.76	0.7	10.03	0.17	13.07	0.18	10.07	0.9
EnlightenGAN[6]	17.46	0.64	10.08	0.32	16.05	0.61	10.08	0.42	16.57	0.73	10.08	0.36	16.33	0.32	10.17	0.56
R2RNet[8]	18.14	0.61	10.01	0.29	17.9	0.65	10.19	0.29	16.05	0.55	10.21	0.35	14.71	0.39	10.51	0.54
SNR-Net[13]	**24.61**	**0.82**	**8.54**	0.16	21.48	0.81	9.69	0.16	24.14	0.93	8.74	**0.06**	22.56	0.61	9.24	0.44
LLformer[21]	23.65	0.79	8.78	0.17	21.43	0.81	**9.58**	**0.14**	17.16	0.69	9.96	0.24	16.48	0.32	9.98	0.71
Retinexformer[22]	23.21	0.81	9.22	0.15	21.25	0.82	9.73	0.16	25.09	0.92	8.63	**0.06**	24.14	0.72	8.45	0.34
DARFormer	24.13	**0.82**	8.78	**0.14**	**21.81**	**0.84**	9.60	**0.14**	**25.78**	**0.93**	**8.46**	**0.06**	**24.56**	**0.74**	**8.30**	**0.33**

从表 1 可以看到, DARFormer 在 SSIM 指标上全面超越其他方法, 这一结果凸显了其在图像亮度、对比度和结构信息方面的卓越表现. 然后, 在 PSNR 指标上, DARFormer 同样展现出色的表现, 除了在 LOL_v1 数据集上略低于 SNR-Net 方法 0.48 dB 外, 在其他数据集上的 PSNR 值均优于 SNR-Net, 分别提升 0.33 dB、1.64 dB 和 2.00 dB. PSNR 值的提升能够直接反映图像清晰度的增加和细节保留能力的增强. 此外, DARFormer 在 LPIPS 指标上也取得了最高分, 这说明 DARFormer 能够生成更符合人类视觉系统感知特性的图像. RMSE 作为衡量像素级差异的指标, DARFormer 在 LOL_v2_synthetic 和 SID 数据集上取得了最优值, 而在 LOL_v1 和 LOL_v2_real 数据集上位列第 2, 这验证了该方法在图像保真度方面的优势.

本文还分析了 DARFormer 与其他 3 种 Transformer 模型的复杂度, 包括参数量、浮点运算次数 (FLOPs) 和推理时间. 使用 NVIDIA A4000 显卡进行实验, 控制图片分辨率和批处理大小相同. 结果如表 2 所示. 明显地, DARFormer 方法在计算资源的消耗上面有所增加, 但结合图像质量的评价指标来看, 它可以保持较低参数量和计算成本, 提供高质量的图像增强效果.

表 2 各 Transformer 方法的复杂度分析对比

方法	参数量 (M)	FLOPs (G)	推理时间 (s)	批处理大小	分辨率大小
SNR-Net[13]	4.01	57.63	0.16	8	128×128
LLformer[21]	24.52	44.07	0.51	8	128×128
Retinexformer[22]	1.61	34.04	0.19	8	128×128
DARFormer	1.93	55.11	0.23	8	128×128

3.3.2 定性评价

图 4–图 7 分别展示了各方法在 LOL_v1、LOL_v2_real、LOL_v2_synthetic 和 SID 数据集上的视觉对比. 红色框标记的区域被特别关注, 并在下方提供了这些区域的放大视图, 以便更细致地观察细节.

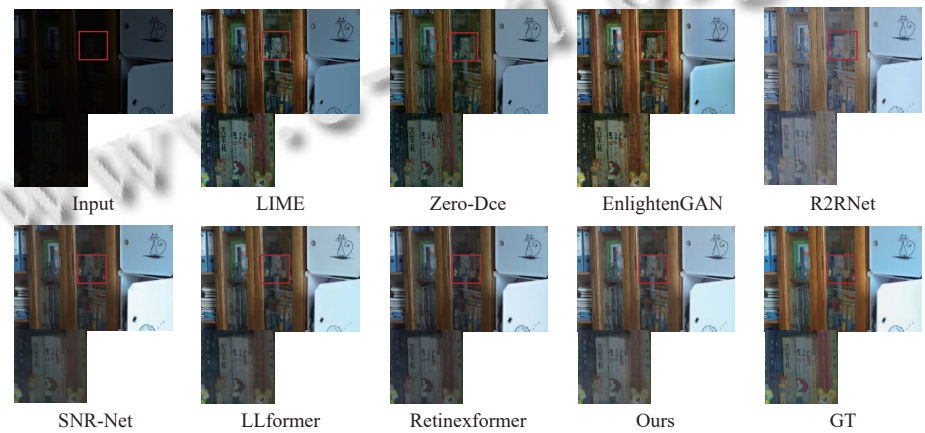


图 4 各方法在 LOL_v1 数据集上的视觉对比图

通过观察这些图像, 可以发现 LIME、Zero-DCE 和 EnlightenGAN 这 3 种方法在噪声抑制和伪影控制方面存在明显不足, 它们的增强结果整体偏暗, 在视觉上也缺乏应有的明亮度和清晰度. R2RNet 虽然在清晰度方面有所表现, 但在颜色准确性和对比度上却不尽如人意.

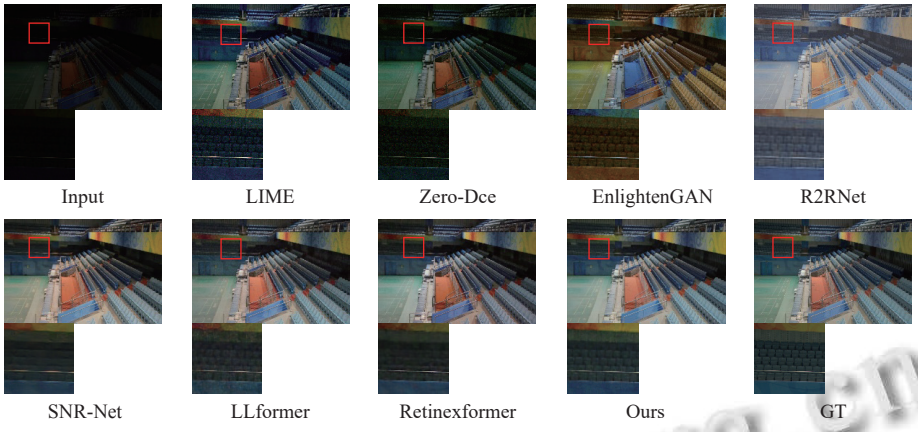


图 5 各方法在 LOL_v2_real 数据集上的视觉对比图

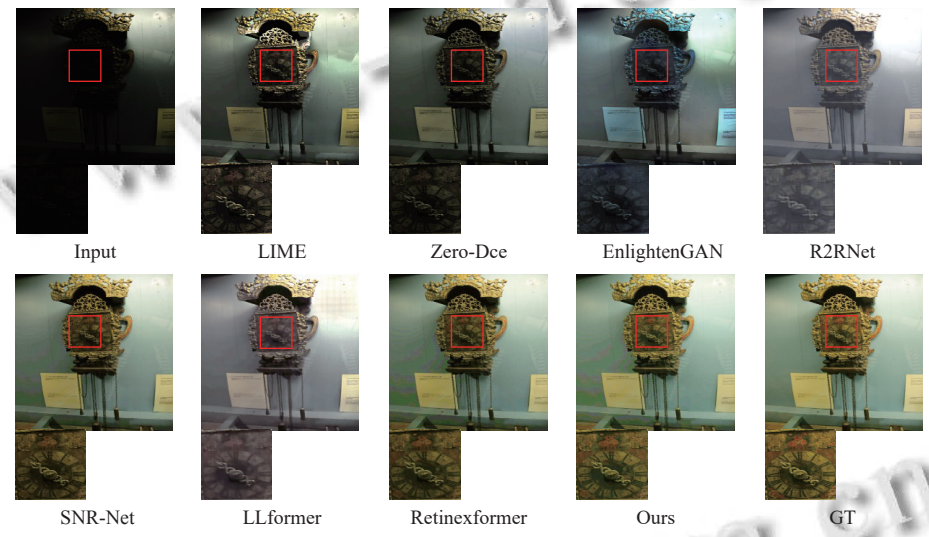


图 6 各方法在 LOL_v2_syhthetic 数据集上的视觉对比图

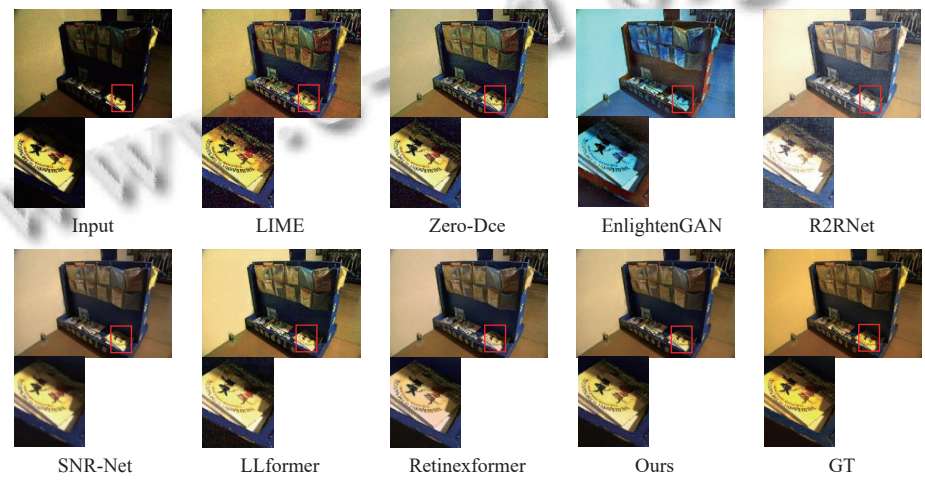


图 7 各方法在 SID 数据集上的视觉对比图

尽管 SNR-Net、LLFormer 和 Retinexformer 在光照提升方面取得了一定成效, 基本能产生令人满意的图像, 但在处理复杂纹理和极暗区域时仍面临挑战. 具体来看, 图 4 展示了各方法在 LOL_v1 数据集上的视觉

对比. 虽然 SNR-Net 实现了最高的 PSNR, 但其去噪过程的过度平滑化却导致了细节的模糊和丢失, 如书籍名字“苦荼花”文本细节处较为明显. LLFormer 在增强图中书柜架子的极暗区域时, 增强结果中引入了明显的黑色伪影. Retinexformer 虽然在细节保留方面表现良好, 但在亮度增强方面却不如本文的 DARFormer 方法. DARFormer 不仅能够有效恢复光照, 还能够有效地保留图像的细节与色彩.

图 5 展示了各方法在 LOL_v2_real 数据集上的视觉对比. SNR-Net 和 Retinexformer 在重建体育馆座椅的清晰度上存在明显不足. 而 LLFormer 则因噪声放大问题, 导致整体视觉效果受损. 相比之下, DARFormer 不仅成功抑制了噪声, 还保持了图像细节的完整性.

图 6 展示了各方法在 LOL_v2_synthetic 数据集上的视觉对比. LLFormer 增强结果整体出现明显颜色失真、过曝情况, 整体颜色偏向单一的黑白色. Retinexformer 和 SNR-Net 则在亮度恢复和颜色恢复上有所欠缺. 与之相比, DARFormer 增强结果具有更均匀的光照和更接近真实场景的颜色.

图 7 展示了各方法在 SID 数据集上的视觉对比. Retinexformer 在颜色一致性上存在问题, 将黄色背景光照增强为偏淡粉色背景光照. SNR-Net 细节丢失问题依然明显, 在图书的本文细节中尤为明显. LLFormer 增强结果则出现整体噪声扩大, 且墙角区域引入绿色的杂色. 相比之下, DARFormer 不仅有效抑制了噪声, 还恢复了图像的细节和颜色信息. 通过以上详细对比, DARFormer 在细节保留、颜色恢复和整体视觉等多个方面具有优越性.

3.4 消融实验

表 3 显示了在 4 个数据集上的消融实验的定量结果. 其中“↑”表示值越高越好, “↓”表示值越小越好. 加粗的字体为最优值.

表 3 在 4 个数据集上消融实验的定量结果

方法	LOL_v1				LOL_v2_real				LOL_v2_synthetic				SID			
	PSNR (dB)↑	SSIM↑	RMSE↓	LPIPS↓	PSNR (dB)↑	SSIM↑	RMSE↓	LPIPS↓	PSNR (dB)↑	SSIM↑	RMSE↓	LPIPS↓	PSNR (dB)↑	SSIM↑	RMSE↓	LPIPS↓
方法1	21.69	0.8	9.89	0.17	**22.01**	0.83	9.55	0.17	24.64	0.92	8.62	0.07	23.88	0.70	8.43	0.40
方法2	19.98	0.64	9.81	0.47	19.98	0.63	9.81	0.62	18.51	0.71	9.92	0.34	21.51	0.60	9.25	0.51
方法3	22.66	0.81	9.51	0.16	21.4	0.83	9.63	0.18	24.80	0.91	8.68	0.08	22.88	0.67	8.80	0.40
DARFormer	**24.13**	**0.82**	**8.78**	**0.14**	21.81	**0.84**	**9.60**	**0.14**	**25.78**	**0.93**	**8.46**	**0.06**	**24.56**	**0.74**	**8.30**	**0.33**

方法 1 采用通道注意力和一个大核注意力 LKA[43] 作为双重注意力块. 将光照特征图引入到通道注意力中, 进行有光照指导的通道注意力. 然后再输入到 LKA 注意力进行空间注意力的特征交互. 方法 2 采用一个有效注意力[35]和一个通道注意力作为双重注意力, 在有效注意力中, 引入光照特征图指导空间信息的捕获. 其中设置有效注意力头数 head=1. 然后再输入 head=8 的通道注意力中. 方法 3 采用光照指导的多头注意力和一个通道注意力, 连接方式采用并行的方式, 然后使用一个卷积进行特征融合.

在进行消融实验的过程中, 本研究采用了不同空间注意力和通道注意力机制的组合, 以期对亮度增强后的图像进行精细的损坏修复. 从表 3 可以得到: DARFormer 方法在多个评价指标上均优于其他消融实验方法.

进一步地, 图 8 显示了在 4 个数据集上消融实验的视觉对比. 通过观察可以发现: 方法 1 存在颜色失真和伪影问题, 如 SID 中窗户的颜色偏绿; LOL_v1 中立柜部分存在黑色伪影. 方法 2 在整体上表现出纹理丢失, 如 LOL_v2_real 中座椅细节模糊不清. 方法 3 则与方法 1 有着类似的问题, 不仅在颜色恢复上表现不佳, 其抑制噪声的能力也相对较弱, 如 LOL_v2_syhthetic 中花朵的颜色不够鲜艳, LOL_v1 中毛衣有大量噪声. 相比之下, DARFormer 则展现出了其在噪声抑制、颜色和细节恢复方面的卓越能力, 提供了更为优秀的视觉效果.

4 总结

本文针对现有的 LLIE 方法存在细节丢失和颜色失真等问题, 提出了基于 Retinex 理论具有双重注意力的 Transformer 增强网络 DARFormer. 基于 Retinex 理论, 考虑到亮度增强后图像存在部分信息损坏, 将网络设计为两个部分: 光照估计网络和损坏修复网络. 光照估计网络中估计亮度映射来点亮低光照图像, 然后在损坏恢复网络中, 设计具有空间注意力和通道注意力的双重注意力机制来捕获全局上下文, 来进行损坏修复. 在空间特征捕获上, 通过引入了亮度特征图进行光

照指导, 增强网络对图像中不同区域的光照变化的理解, 从而在修复过程中保持图像的自然性和一致性. 然后, 利用通道注意力进一步丰富空间信息. 这种设计不仅有效增强图像的亮度, 还有助于保留更多的细节和色彩. 大量的实验结果验证了 DARFormer 与其他主流算法相比的竞争优势. 从算法的复杂性来看, DARFormer 算法以其较低的参数量和延迟时间, 展现出在自动驾驶领域应用的潜力.

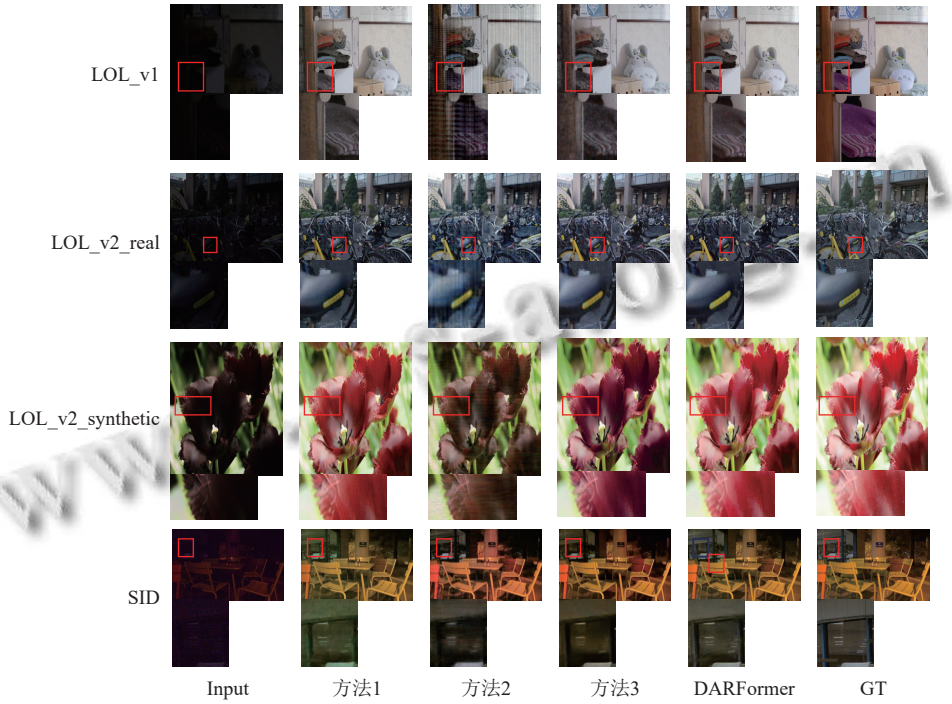


图 8 在 4 个数据集上的消融实验的视觉对比

参考文献

1 Huang YK, Zha ZJ, Fu XY, *et al*. Real-world person re-identification via degradation invariance learning. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020. 14072–14082. [doi: 10.1109/cvpr42600.2020.01409]

2 Zhao ZQ, Zheng P, Xu ST, *et al*. Object detection with deep learning: A review. IEEE Transactions on Neural Networks and Learning Systems, 2019, 30(11): 3212–3232. [doi: 10.1109/TNNLS.2018.2876865]

3 Pizer SM, Amburn EP, Austin JD, *et al*. Adaptive histogram equalization and its variations. Computer Vision, Graphics, and Image Processing, 1987, 39(3): 355–368. [doi: 10.1016/S0734-189X(87)80186-X]

4 Wang ZG, Liang ZH, Liu CL. A real-time image processor with combining dynamic contrast ratio enhancement and inverse gamma correction for PDP. Displays, 2009, 30(3): 133–139. [doi: 10.1016/j.displa.2009.03.006]

5 Land EH. The Retinex theory of color vision. Scientific American, 1977, 237(6): 108–128. [doi: 10.1038/scientificamerican1277-108]

6 Jiang YF, Gong XY, Liu D, *et al*. EnlightenGAN: Deep light enhancement without paired supervision. IEEE Transactions on Image Processing, 2021, 30: 2340–2349. [doi: 10.1109/TIP.2021.3051462]

7 Wu WH, Weng J, Zhang PP, *et al*. URetinex-Net: Retinex-based deep unfolding network for low-light image enhancement. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. 5891–5900. [doi: 10.1109/CVPR52688.2022.00581]

8 Hai J, Xuan Z, Yang R, *et al*. R2RNet: Low-light image enhancement via real-low to real-normal network. Journal of Visual Communication and Image Representation, 2023, 90: 103712. [doi: 10.1016/j.jvcir.2022.103712]

9 Zhu MF, Pan PB, Chen W, *et al*. EEMEFN: Low-light image enhancement via edge-enhanced multi-exposure fusion network. Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York: AAAI, 2020. 13106–13113. [doi: 10.1609/AAAI.V34I07.7013]

10 Guo XJ, Hu QM. Low-light image enhancement via breaking down the darkness. International Journal of Computer Vision, 2023, 131(1): 48–66. [doi: 10.1007/s11263-022-01667-9]
11 Priyadarshini R, Bharani A, Rahimankhan E, *et al*. Low-light image enhancement using deep convolutional network. In: Raj JS, Iliyasu AM, Bestak R, *et al*. eds. Innovative Data Communication Technologies and Application. Singapore: Springer, 2021. 695–705.
12 Vaswani A, Shazeer N, Parmar N, *et al*. Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6000–6010.
13 Xu XG, Wang RX, Fu CW, *et al*. SNR-aware low-light image enhancement. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. 17693–17703. [doi: 10.1109/CVPR52688.2022.01719]
14 Cui ZT, Li KC, Gu L, *et al*. You only need 90k parameters to adapt light: A light weight Transformer for image enhancement and exposure correction. Proceedings of the 33rd British Machine Vision Conference (BMVC). London: BMVA Press, 2022. 238.
15 Dang JC, Zhong Y, Qin XL. PPformer: Using pixel-wise and patch-wise cross-attention for low-light image enhancement. Computer Vision and Image Understanding, 2024, 241: 103930. [doi: 10.1016/j.cviu.2024.103930]
16 Li MD, Liu JY, Yang WH, *et al*. Structure-revealing low-light image enhancement via robust Retinex model. IEEE Transactions on Image Processing, 2018, 27(6): 2828–2841. [doi: 10.1109/TIP.2018.2810539]
17 Wang SH, Zheng J, Hu HM, *et al*. Naturalness preserved enhancement algorithm for non-uniform illumination images. IEEE Transactions on Image Processing, 2013, 22(9): 3538–3548. [doi: 10.1109/TIP.2013.2261309]
18 Guo XJ, Li Y, Ling HB. LIME: Low-light image enhancement via illumination map estimation. IEEE Transactions on Image Processing, 2017, 26(2): 982–993. [doi: 10.1109/TIP.2016.2639450]
19 Xu L, Yan Q, Xia Y, *et al*. Structure extraction from texture via relative total variation. ACM Transactions on Graphics, 2012, 31(6): 139. [doi: 10.1145/2366145.2366158]
20 Guo CL, Li CY, Guo JC, *et al*. Zero-reference deep curve estimation for low-light image enhancement. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 1777–1786. [doi: 10.1109/CVPR42600.2020.00185]
21 Wang T, Zhang KH, Shen TR, *et al*. Ultra-high-definition low-light image enhancement: A benchmark and Transformer-based method. Proceedings of the 37th AAAI Conference on Artificial Intelligence. Washington: AAAI, 2023. 2654–2662. [doi: 10.1609/aaai.v37i3.25364]
22 Cai YH, Bian H, Lin J, *et al*. Retinexformer: One-stage retinex-based Transformer for low-light image enhancement. Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision. Paris: IEEE, 2023. 12470–12479. [doi: 10.1109/ICCV51070.2023.01149]
23 Wang RX, Zhang Q, Fu CW, *et al*. Underexposed photo enhancement using deep illumination estimation. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 6842–6850. [doi: 10.1109/CVPR.2019.00701]
24 Wei C, Wang WJ, Yang WH, *et al*. Deep Retinex decomposition for low-light enhancement. Proceedings of the 2018 British Machine Vision Conference (BMVC). Newcastle: BMVA Press, 2018. 155.
25 Dosovitskiy A, Beyer L, Kolesnikov A, *et al*. An image is worth 16x16 words: Transformers for image recognition at scale. Proceedings of the 9th International Conference on Learning Representations (ICLR). OpenReview.net, 2021.
26 Carion N, Massa F, Synnaeve G, *et al*. End-to-end object detection with Transformers. Proceedings of the 16th European Conference on Computer Vision. Glasgow: Springer, 2020. 213–229. [doi: 10.1007/978-3-030-58452-8_13]
27 Zhou YP, Li Z, Guo CL, *et al*. SRFormer: Permuted self-attention for single image super-resolution. Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision. Paris: IEEE, 2023. 12734–12745. [doi: 10.1109/ICCV51070.2023.01174]
28 Azad R, Arimond R, Aghdam EK, *et al*. DAE-former: Dual attention-guided efficient Transformer for medical image segmentation. Proceedings of the 6th International Workshop on Predictive Intelligence in Medicine. Vancouver: Springer, 2023. 83–95. [doi: 10.1007/978-3-031-46005-0_8]
29 Zamir SW, Arora A, Khan S, *et al*. Restormer: Efficient Transformer for high-resolution image restoration. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. 5718–5729. [doi: 10.1109/CVPR52688.2022.00564]
30 Wang KQ, Cui ZT, Jia JR, *et al*. Linear array network for low-light image enhancement. arXiv:2201.08996. 2022.

31 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 7132–7141. [doi: 10.1109/CVPR.2018.00745]

32 Woo S, Park J, Lee JY, *et al*. CBAM: Convolutional block attention module. Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich: Springer, 2018. 3–19. [doi: 10.1007/978-3-030-01234-2_1]

33 El-Nouby A, Touvron H, Caron M, *et al*. XCiT: Cross-covariance image Transformers. Proceedings of the 35th International Conference on Neural Information Processing Systems. Curran Associates Inc., 2021. 1531.

34 Ding MY, Xiao B, Codella N, *et al*. DaViT: Dual attention vision Transformers. Proceedings of the 17th European Conference on Computer Vision. Tel Aviv: Springer, 2022. 74–92.

35 Shen ZR, Zhang MY, Zhao HY, *et al*. Efficient attention: Attention with linear complexities. Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision. Waikoloa: IEEE, 2021. 3530–3538. [doi: 10.1109/WACV48630.2021.00357]

36 Hendrycks D, Gimpel K. Gaussian error linear units (GELUs). arXiv:1606.08415, 2016.

37 Yang WH, Wang WJ, Huang HF, *et al*. Sparse gradient regularized deep Retinex network for robust low-light image enhancement. IEEE Transactions on Image Processing, 2021, 30: 2072–2086. [doi: 10.1109/TIP.2021.3050850]

38 Chen C, Chen QF, Xu J, *et al*. Learning to see in the dark. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 3291–3300. [doi: 10.1109/CVPR.2018.00347]

39 Kingma DP, Ba J. Adam: A method for stochastic optimization. Proceedings of the 3rd International Conference on Learning Representations. San Diego, 2015.

40 Kellman P, McVeigh ER. Image reconstruction in SNR units: A general method for SNR measurement. Magnetic Resonance in Medicine, 2005, 54(6): 1439–1447. [doi: 10.1002/mrm.20713]

41 Wang Z, Bovik AC, Sheikh HR, *et al*. Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing, 2004, 13(4): 600–612. [doi: 10.1109/TIP.2003.819861]

42 Zhang R, Isola P, Efros AA, *et al*. The unreasonable effectiveness of deep features as a perceptual metric. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 586–595. [doi: 10.1109/CVPR.2018.00068]

43 Lau KW, Po LM, Rehman YAU. Large separable kernel attention: Rethinking the large kernel attention design in CNN. Expert Systems with Applications, 2024, 236: 121352. [doi: 10.1016/j.eswa.2023.121352]

(校对责编: 张重毅)