

# 基于多重注意力和感知加权学习的 单图像高动态范围重建

柯德涨<sup>1</sup>, 陈晔曜<sup>1</sup>, 徐海勇<sup>2</sup>, 金充充<sup>3</sup>, 蒋刚毅<sup>1\*</sup>

(1. 宁波大学信息科学与工程学院, 浙江宁波 315211; 2. 宁波大学数学与统计学院, 浙江宁波 315211;  
3. 宁波大学科学技术学院, 浙江宁波 315300)

**摘要:** 单图像高动态范围(High Dynamic Range, HDR)重建能够避免多曝光HDR成像可能造成的鬼影伪像, 正受到广泛研究. 然而, 现有方法由于缺乏对重要信息的关注, 尚不能很好地恢复曝光不良区域的细节信息. 为解决该问题, 本文提出了一种基于多重注意力和感知加权学习的单图像HDR重建方法, 旨在从单幅低动态范围图像中推断出高保真的HDR图像. 具体而言, 考虑到恢复曝光不良区域需参考其他区域的补偿信息, 本文设计了具有全局-局部感受野的多重注意力视觉Transformer (Multi-Attention Vision Transformer, MA-ViT), 将其深度可分离卷积和注意力机制相结合, 从而实现更有效的全局和局部特征提取与交互. 此外, 还提出了一种损失感知加权图以引导网络聚焦曝光不良区域, 进一步提升HDR重建质量. 本文在多个基准数据集上构建了全面的对比实验, 结果表明所提出方法相较于目前最先进的方法在平均峰值信噪比(Peak Signal to Noise Ratio, PSNR)上提高了0.23 dB, 同时生成了具有更高视觉质量的HDR重建结果.

**关键词:** 单图像高动态范围重建; 深度学习; 逆色调映射; 注意力机制; 感知加权

**基金项目:** 国家自然科学基金(No.62401301, No.62271276, No.62171243); 浙江省自然科学基金(No.LQ24F010002)

**中图分类号:** TP391 **文献标识码:** A **文章编号:** 0372-2112(2025)06-2063-16

**电子学报 URL:** <http://www.ejournal.org.cn>

**DOI:** 10.12263/DZXB.20241055

## Single-Image High Dynamic Range Reconstruction Based on Multi-Attention and Perceptual Weighted Learning

KE De-zhang<sup>1</sup>, CHEN Ye-yao<sup>1</sup>, XU Hai-yong<sup>2</sup>, JIN Chong-chong<sup>3</sup>, JIANG Gang-yi<sup>1\*</sup>

(1. Faculty of Information Science and Engineering, Ningbo University, Ningbo, Zhejiang 315211, China;

2. School of Mathematics and Statistics, Ningbo University, Ningbo, Zhejiang 315211, China;

3. College of Science and Technology, Ningbo University, Ningbo, Zhejiang 315300, China)

**Abstract:** Single-image high dynamic range (HDR) reconstruction can avoid ghosting artifacts that may be caused by multi-exposure HDR imaging, and is receiving widespread research. However, existing methods still struggle to effectively restore detail information in poorly exposed regions due to a lack of focus on critical information. To address this issue, this paper proposes a single-image HDR reconstruction method based on multi-attention and perceptual weighted learning, which aims to infer a high-fidelity HDR image from a single low dynamic range (LDR) image. Specifically, considering that the restoration of poorly exposed regions requires reference to compensation information from other regions, a multi-attention vision transformer (MA-ViT) with global-local receptive fields is designed. It combines depthwise separable convolution and attention mechanisms to achieve more effective global and local feature extraction and interaction. In addition, a loss aware weighted map is proposed to guide the network to focus on poorly exposed regions, further enhancing the quality of HDR reconstruction. Comprehensive comparative experiments are conducted on multiple benchmark datasets, and the results show that the proposed method improves the average peak signal to noise ratio (PSNR) by 0.23 dB compared to the state-of-the-art method, while generating HDR reconstruction results with higher visual quality.

**Key words:** single-image high dynamic range reconstruction; deep learning; inverse tone mapping; attention mechanism; perceptual weighting

Foundation Item(s): National Natural Science Foundation of China (No.62401301, No.62271276, No.62171243);  
Natural Science Foundation of Zhejiang Province (No.LQ24F010002)

## 1 引言

在计算摄影和电子信息领域,高动态范围(High Dynamic Range, HDR)重建一直是一个重要课题.近年来, HDR 显示技术逐渐得到普及,但 HDR 影像内容还相对匮乏.特别地,由于传感器的限制,目前大多数消费级数码相机仍无法完备捕获真实世界中的亮度变化<sup>[1,2]</sup>,进而导致采集的图像遭受低动态范围(Low Dynamic Range, LDR),即存在欠曝光和过曝光现象.基于多曝光的 HDR 成像<sup>[3-7]</sup>是解决该问题的常见途径,其通过合并多幅以不同曝光值采集的 LDR 图像,以重现更广泛的亮度范围和更细致的内容信息.尽管该方法在静态条件下能够生成高质量的 HDR 图像,但对动态场景容易造成鬼影伪像.

与多曝光方式不同,单图像 HDR 重建旨在从单幅 LDR 图像中重建高质量的 HDR 图像.由于规避了多曝光数据采集的要求,该方式更具灵活性和便利性.此外,目前互联网上存在的数字图像大多仅包含单曝光信息,因而单图像 HDR 重建可有效兼容现有图像数据,即特别适用于先前拍摄得到或网络上获取的图像.传统方法通常是基于各类数学模型并结合人类视觉系统知识以设计 HDR 重建方法.例如, Landis 等人<sup>[8]</sup>提出了基于幂函数模型的动态范围扩展方法. Huo 等人<sup>[9]</sup>提出了基于人类视觉系统视网膜响应的方法,其在良好曝光条件下可产生较好的重建结果.然而,对于严重曝光不良的图像,它们难以恢复高保真的细节信息.此外,由于不完善的数学模型和先验知识,上述方法难以处理各类型的图像.

近年来,随着以卷积神经网络(Convolutional Neural Network, CNN)为主的各种深度学习方法在图像信号处理和计算机视觉领域取得显著成功,许多基于学习的单图像 HDR 重建方法也被相继提出. Eilertsen 等人<sup>[10]</sup>针对亮度饱和问题,提出了一个基于自编码网络的单图像 HDR 重建方法,其通过饱和掩膜以引导网络聚焦高亮区域的细节恢复. Marnerides 等人<sup>[11]</sup>设计了一种基于多尺度架构的 HDR 重建网络,其整合局部、扩张和全局分支来恢复 HDR 信息. Khan 等人<sup>[12]</sup>利用反馈机制,通过多次更新的隐状态来学习 LDR 到 HDR 的非线性映射关系. Wang 等人<sup>[13]</sup>构建了一种基于相机成像原理的 Kunet 网络. Zhang 等人<sup>[14]</sup>设计了一种逐级重建 HDR 信息的多阶段渐进式网络.值得注意的是,上述方法主要基于 CNN 架构,缺乏对长范围信息的建模以及对曝光不良区域的重点关注.因此,它们难以细致地恢复极端曝光区域的细节信息.

针对以上问题,本文提出了一种基于多重注意力和感知加权学习的单图像 HDR 重建方法.具体而言,设计了具有全局-局部感受野的多重注意力视觉 Transformer (Multi-Attention Vision Transformer, MA-ViT),通过将深度可分离卷积与注意力机制相结合,以实现更有效的全局和局部特征提取与交互.此外,基于对曝光区域的分析,提出了一种损失感知加权图,其能引导网络关注曝光不良区域,进而提升 HDR 重建质量.广泛的实验结果表明,所提出方法在定量和定性比较中均优于现有代表性方法.

## 2 相关工作

在本节中,将回顾现有的单图像 HDR 重建方法,其主要分为两类:传统方法和基于学习的方法.

### 2.1 传统方法

传统的单图像 HDR 重建方法主要结合各类数学模型以及人类视觉系统的相关知识. Landis 等人<sup>[8]</sup>提出了基于幂函数模型的动态范围扩展方法. Akyüz 等人<sup>[15]</sup>提出了一种简单线性扩展方法,通过对最大-最小值归一化后的图像进行亮度乘法运算来拓展动态范围. Masia 等人<sup>[16]</sup>提出了基于 Gamma 曲线的扩展算子.上述方法均属于全局模型,即对所有像素使用同一扩展函数.尽管它们拥有逻辑简单且计算量小的优点,但该处理方式可能会损失图像的局部对比度,并引入色彩失真.不同于全局处理, Meylan 等人<sup>[17]</sup>提出了一种色调缩放函数,分别对图像的漫反射和高光区域的亮度进行不同程度的线性缩放,进而分段拓展图像的动态范围. Kuo 等人<sup>[18]</sup>提出了一种内容自适应的逆色调映射算子,可对不同场景特征产生不同响应. Kovaleski 等人<sup>[19]</sup>提出了一种基于交叉双边滤波的逆色调映射算子,可为不同曝光输入生成 HDR 图像和视频. Wang 等人<sup>[20]</sup>提出了一种基于用户交互的方法,可人为框选出欠曝光或过曝光区域,而后通过纹理合成技术来重建该区域的细节信息.尽管以上方法对曝光良好或异常曝光不显著区域可生成较好的 HDR 重建结果,但难以应对严重曝光不良的情况,并且容易引入颜色失真以及细节丢失等问题.

### 2.2 基于学习的方法

近年来,以 CNN 为主的深度学习方法在计算机视觉中取得巨大进步.受此启发,许多基于深度学习的单图像 HDR 重建方法被相继提出来提升单图像 HDR 重建性能. Eilertsen 等人<sup>[10]</sup>提出了一个基于自编码网络的方法,旨在恢复过曝光/饱和区域的细节信息.

Marnerides 等人<sup>[11]</sup>提出了一种基于多尺度架构的 ExpNet 网络,通过局部分支、扩张分支和全局分支来捕获不同尺度下的特征细节. Khan 等人<sup>[12]</sup>根据反馈思想,通过多次迭代更新隐状态以学习 LDR 到 HDR 的信息重建. Santos 等人<sup>[21]</sup>设计了一种特征掩蔽机制,使生成图像避免依赖饱和和区域中的无效信息. Chen 等人<sup>[22]</sup>考虑到 HDR 重建中存在的噪声和量化误差,提出了包含基础网络,加权网络和条件网络的 HDRUNet 网络. Wang 等人<sup>[13]</sup>将 HDR 到 LDR 之间的成像知识集成到 UNet 架构中而设计了一种 HDR 重建网络(称为 Kunet). Bei 等人<sup>[23]</sup>将生成对抗网络(Generative Adversarial Networks, GAN)应用于单图像 HDR 重建. 然而, GAN 生成的图像可能包含内容偏差和不切实际的伪影,这会严重影响良好曝光区域的重建质量. Moriwaki 等人<sup>[24]</sup>使用针对单图像 HDR 重建优化的感知损失来提高生成图像质量. Liu 等人<sup>[25]</sup>将原型相机处理管道建模为动态范围截断、非线性映射,以及量化 3 个阶段. 基于此,他们构建了 3 个子网络来学习这些阶段的逆转过程以重建最终的 HDR 图像. 在此基础上, She 等人<sup>[26]</sup>针对现有单帧 HDR 图像重建方法存在色域不自然、映射函数求解空间大,以及数据集不足等问题,提出了一种半监督式的双边学习网络. 值得注意的是,上述方法均是直接学习 LDR 域到 HDR 域的非线性映射.

不同地,考虑到 LDR 域与 HDR 域之间的巨大差距,直接从 8 位 LDR 图像中推断 32 位 HDR 图像是相对困难的,有研究者提出间接学习方法. Endo 等人<sup>[27]</sup>构建了一个基于二维卷积和三维反卷积的逆色调映射网络,其根据单曝光图像来学习生成具有更高和更低曝光的多幅 LDR 图像. 之后,他们利用现有多曝光融合方法来从生成的多曝光堆栈中重建 HDR 图像. Lee 等人<sup>[28]</sup>将 GAN 应用于多曝光堆栈生成. 进一步,LEE 等人<sup>[29]</sup>还

提出了基于循环一致性的多曝光图像生成网络. Jo 等人<sup>[30]</sup>提出了一种新的曝光控制网络,并构建了一个亮度特征生成器,用于生成恢复高频分量所需区域的信息. Zhang 等人<sup>[31]</sup>提出了一种只需估计两幅曝光图像的方法,并使用跨模型注意力来利用它们的信息. Chen 等<sup>[32]</sup>提出了连续曝光值表示方法,利用循环训练策略来监督网络学习,使其在缺乏标签数据的情况下仍可生成具有任意曝光值的图像.

相比于传统方法,基于学习的单图像 HDR 重建方法展现了更好的性能优势. 然而,现有方法大多基于 CNN 架构,其有限的感受野阻碍了长范围信息的获取. 此外,现有方法对曝光不良区域的强调基本上局限于依照 LDR 图像亮度而设计的掩膜,缺乏对 HDR 图像在退化过程中造成的视觉信息损失分布的考虑.

### 3 本文提出的方法

基于上述分析,本文提出了一种基于多重注意力和感知加权学习的单图像 HDR 重建方法.

#### 3.1 总述

图 1 展示了所提出方法的整体框架. 鉴于 Transformer 架构<sup>[33,34]</sup>具有强大的长范围建模能力,所提出方法构建多重注意力视觉 Transformer(MA-ViT)作为主要部件,可捕获不同感受野下的特征依赖. 为促进全局和局部特征的提取与融合,所提出的 MA-ViT 结合了深度可分离卷积和注意力机制,从不同维度上聚焦更具价值的信息. 同时,为保证重建 HDR 图像的保真度,设计了两个增强分支以分别学习结构和颜色特征表示. 此外,所提出方法在损失函数中引入损失感知加权图以引导网络关注感知质量降级区域,并考虑 HDR 图像在退化过程中的视觉信息损失分布,从而提升 HDR 重建质量.

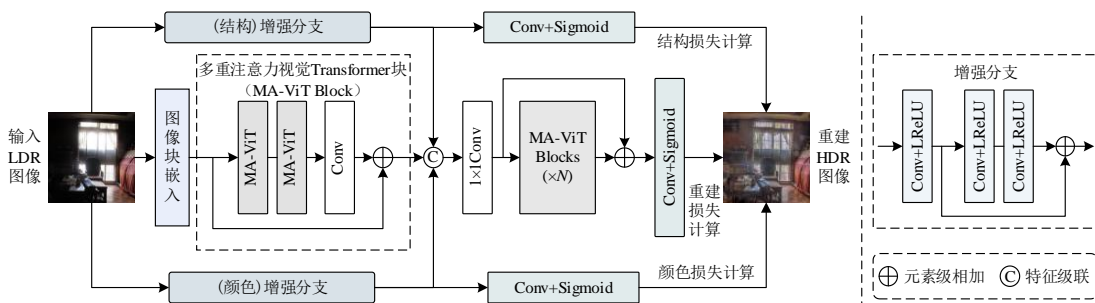


图 1 所提出方法的整体框架

#### 3.2 多重注意力视觉 Transformer

由于曝光不良区域丢失了原始场景中的细节信息,所以需要更大的感受野和特征关注机制,以迫使网络可从邻域或非局部区域中推断出曝光不良区域的细节信息. 现有方法大多基于 CNN 架构,其有限的感受

野阻碍了长范围信息的获取. 针对此,本文设计了具有全局-局部感受野的 MA-ViT,以有效捕获不同感受野的信息并实现融合,进而提高对有效特征的关注.

如图 2 所示,所提出的 MA-ViT 采用双分支结构,在全局分支上使用基于窗口的 Transformer<sup>[35]</sup>来捕获全局

信息. 该Transformer由两个层归一化(Layer Normalization, LN), 以及分别紧随其后的多头自注意力(Multi-Head Attention, MSA)和前馈网络(Feed-Forward Network, FFN)组成. 通过窗口移位机制, 可建立特征之间的长范围依赖. 给定输入特征  $F_{in}$ , 全局分支的处理过程如下所示:

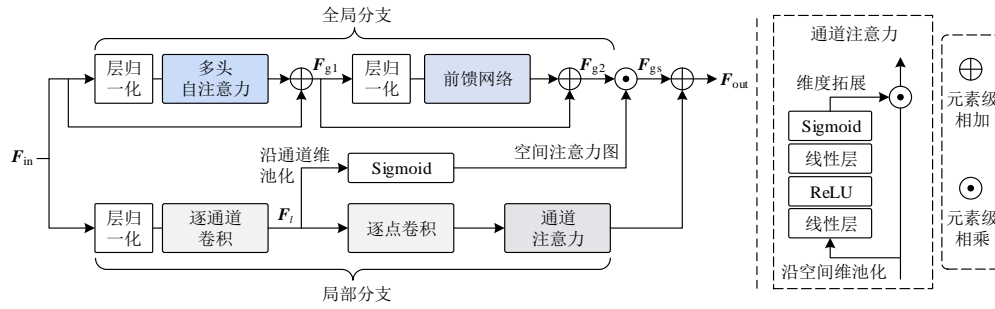


图2 所提出的多重注意力视觉Transformer的具体结构

为保持局部细节, 在局部分支上将深度可分离卷积与空间、通道注意力相结合, 使得网络在高效捕获局部信息的同时, 可更多地关注有价值特征. 具体地, 首先对层归一化后的特征使用  $5 \times 5$  逐通道卷积以捕获局部特征(记为  $F_l$ ). 之后, 沿通道维对局部特征进行均值池化, 以聚合多通道信息. 基于此, 利用 Sigmoid 函数来获得空间注意力图, 从而对全局分支的输出(即  $F_{g2}$ )进行空间调制, 如下所示:

$$F_{gs} = F_{g2} \odot \text{Sigmoid}(\text{Pool}_c(F_l)) \quad (3)$$

其中,  $F_{gs}$  表示经空间调制后的全局特征,  $\text{Pool}_c(\cdot)$  表示通道池化,  $\odot$  表示逐元素相乘.

此外, 对于局部特征  $F_l$ , 应用逐点卷积来引入跨通道的信息交互, 从而克服逐通道卷积在不同特征通道间信息隔离的局限. 随后对逐点卷积的结果使用通道注意力<sup>[36]</sup>以强调重要特征并抑制冗余信息. 最后, 将该特征与经空间调制的全局特征  $F_{gs}$  进行元素级相加, 以得到全局与局部融合后的特征  $F_{out}$ , 如下所示:

$$F_{out} = F_{gs} + \text{CA}(\text{Conv}_p(F_l)) \quad (4)$$

其中,  $\text{Conv}_p(\cdot)$  表示逐点卷积,  $\text{CA}(\cdot)$  表示通道注意力.

### 3.3 增强分支

为缓解重建图像出现视觉不适的结构和颜色失真, 所提出方法构建了增强分支, 如图1右侧所示. 该分支由包含残差连接的3个卷积层组成. 在增强分支后, 利用单层卷积层以及 Sigmoid 函数来将输出特征投影为像素级图像, 以与标签图像进行特定损失计算, 进而有效控制该分支的学习倾向. 本文构建了两个增强分支来分别聚焦学习结构和颜色信息复原. 之后, 两个增强分支的输出和第一个 MA-ViT 的输出进行级联, 并通过  $1 \times 1$  卷积以调整通道维度, 其结果被馈送到后续

$$F_{g1} = \text{MSA}(\text{LN}(F_{in})) + F_{in} \quad (1)$$

$$F_{g2} = \text{FFN}(\text{LN}(F_{g1})) + F_{g1} \quad (2)$$

其中,  $F_{g2}$  表示捕获长范围信息后的特征,  $F_{g1}$  表示经 MSA 后的中间特征,  $\text{LN}(\cdot)$  表示层归一化,  $\text{MSA}(\cdot)$  表示多头自注意力,  $\text{FFN}(\cdot)$  表示由两个线性层组成的前馈网络.

的 MA-ViT 中进行更深层的特征学习. 为了调查增强分支的作用, 图3上部分前两张图可视化了两个增强分支所输出的特征图, 第三张图是对应的真值图像; 另外, 图3下部分前两张图可视化了两个增强分支所重建的图像, 第三张图是所提出方法最终重建的图像. 需要指出, 考虑到特征的高维性, 首先在通道维度上对结构增强分支和颜色增强分支输出的特征图执行主成分分析, 而后取出包含大部分能量的第一个分量, 进而得到结构和颜色特征图的可视化结果. 从图3上半部分可知, 结构特征图呈现出更加清晰的结构和纹理, 而颜色特征图在不同颜色区域产生了不同的响应. 进一步地, 从图3下半部分给出的重建结果可以看到, 结构和颜色增强分支所重建的图像分别在细节和色调上更加接近于最终的重建图像, 这表明两个增强分支可有效学习各自的信息复原.

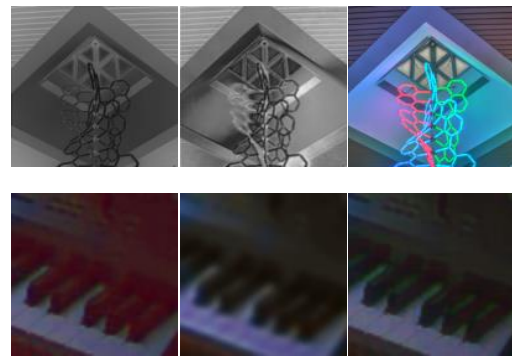


图3 增强分支效果验证的相关图示

### 3.4 感知加权建模

考虑到过/欠曝光区域的信息复原难于正常曝光区域, 本文设计了一种损失感知加权图, 以隐式引导网络

聚焦由曝光不良而导致视觉感知受损的区域. 该加权图类似于掩膜, 可看作图像信息重要性的一种度量. Santos 等人<sup>[21]</sup>直接将掩膜与特征图进行相乘以消除饱和区域的无用信息. 与之不同, 本文将掩膜的值逆转以作为加权图, 并将其应用于损失计算来强调曝光不良区域的信息复原. 现有掩膜生成往往基于手工设计的函数和人工指定的参数. 例如, 在大于某个阈值时, 越接近饱和的区域往往获得越低的掩膜值. 然而, 该方式仅考虑亮度信息, 而没有考虑不良曝光造成的感知受损程度.

以过曝光图像为例, 如图 4 所示, LDR 图像中大部分树木区域趋于饱和, 并且路面纹路也由于过曝光而丢失细节. 显然, 所提出方法需重点关注这些区域, 以提升 HDR 重建质量. 然而, 采用固定阈值(基于亮度)而得到的掩膜加权图, 即图 4(c)会在强调树木区域的同时过度关注天空区域(即权值为 1). 另外, 如图 4(d)所示, 利用 Santos 等人<sup>[21]</sup>提出的掩膜方式而生成的加权图依然会对天空区域赋予更大的权值, 且会忽略路面区域的一些纹理(即权值接近 0). 进一步地, 通过观察色调映射图像可发现, 天空区域的亮度高于其他区域, 并且平坦无纹理. 换言之, 相较于树木区域和路面纹路, 天空区域的感知质量损失并不大, 这可能导致网络无意义的关注. 为此, 本文设计了一种损失感知加权图, 可引导网络关注更需复原的区域.

具体地, 首先设定一个具有固定阈值的掩膜来初步检测过曝光区域, 如下所示:

$$M_o(x, y) = \begin{cases} 0, & \text{if } I_{\text{LDR}}(x, y) < \alpha \\ 1, & \text{otherwise} \end{cases} \quad (5)$$

其中,  $M_o$  表示过曝光掩膜;  $I_{\text{LDR}}$  表示输入的 LDR 图像;  $\alpha$

表示过曝光阈值, 其值为 0.95.

考虑到 HDR 图像(记为  $I_{\text{HDR}}$ )处于线性域, 对其进行  $\mu$  律压缩以实现域转换, 如下所示:

$$I_\mu = \frac{\ln(1 + \mu I_{\text{HDR}})}{\ln(1 + \mu)} \quad (6)$$

其中,  $I_\mu$  表示域转换后的 HDR 图像;  $\mu$  为控制压缩水平的自定义参数, 其值设为 5 000.

基于上述, 利用  $I_\mu$  和  $I_{\text{LDR}}$  进行差异归一化计算, 并将结果与  $M_o$  相乘, 以得到初步加权图. 之后, 对其中数值排列前 2% 的像素点截断到它们之中的最低值. 最后, 对截断结果进行归一化处理, 如下所示:

$$W_o = \text{Normal} \left( \text{Clip} \left( \left( \frac{I_{\text{LDR}} - I_\mu}{I_\mu + \varepsilon} \times M_o \right) \right) \right) \quad (7)$$

其中,  $W_o$  表示针对过曝光区域的损失感知加权图;  $\text{Clip}(\cdot)$  表示对 2% 像素进行截断;  $\text{Normal}(\cdot)$  表示最大-最小值归一化;  $\varepsilon$  为防止除数为 0 的极小值, 其值设为  $1 \times 10^{-4}$ .

图 4(e)可视化了所提出的损失感知加权图. 可以看到, 天空区域的权值减小, 并且树木区域的权值相较于 Santa 等人<sup>[21]</sup>的掩膜更高; 同时路面纹路也获得了较高的权值. 因此, 所提出的损失感知加权图可有效引导网络关注更需复原的区域.

### 3.5 损失函数

所提出方法涉及的损失函数包含 3 个部分, 对应于两个增强分支的结果以及最终重建结果, 如下所示:

$$L_{\text{total}} = L_S + L_C + L_F \quad (8)$$

其中,  $L_S$ ,  $L_C$  和  $L_F$  分别为对应于结构增强分支、颜色增强分支以及最终重建结果的损失.

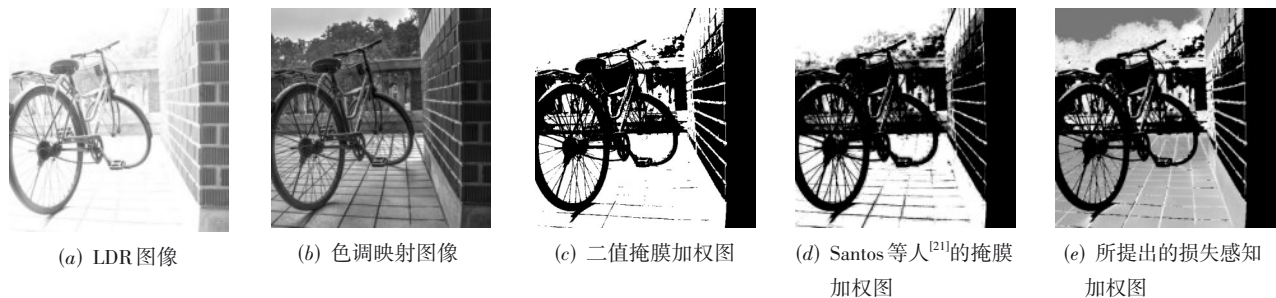


图 4 所提出的感知加权建模的相关图示

对于  $L_S$ , 采用基于 VGG-19 的感知损失<sup>[37]</sup>以从多尺度上对图像结构及纹理复原增加约束, 如下所示:

$$L_S = \lambda_S L_{\text{VGG}}(I_S, I_{\text{GT}}) \quad (9)$$

其中,  $I_S$  和  $I_{\text{GT}}$  分别表示结构增强分支的输出和真值图像;  $L_{\text{VGG}}$  表示 VGG 损失, 它是采用预训练的 VGG-19 网

络提取多尺度特征, 进而比较生成图像与真值图像在特征空间中的差异, 以此指导模型优化, 使其更加注重语义结构和纹理一致性. 需要指出, 本文采用  $L_1$  损失来量化特征差异.  $\lambda_S$  为权重参数, 其值为 10.

对于  $L_C$ , 除使用像素  $L_1$  损失外, 利用余弦相似度来

约束每个 RGB 向量的颜色准确性,如下所示:

$$L_{CS}(I_C, I_{GT}) = 1 - \frac{1}{K} \sum_{i=1}^K \frac{I_{C,i} \cdot I_{GT,i}}{\|I_{C,i}\|_2 \|I_{GT,i}\|_2} \quad (10)$$

$$L_C = L_1(I_C, I_{GT}) + \lambda_C L_{CS}(I_C, I_{GT}) \quad (11)$$

其中,  $L_{CS}$  表示基于余弦相似度的损失;  $I_C$  表示颜色增强分支的输出;  $I_{C,i}$  和  $I_{GT,i}$  表示第  $i$  个像素的 RGB 向量;  $K$  为像素总数;  $\lambda_C$  为权重参数, 其值为 5.

为了引导网络关注感知受损严重的区域, 在  $L_F$  中加入了所提出的损失感知加权图, 其包含像素加权损失 ( $L_{1,W}$ ) 和感知加权损失 ( $L_{VGG,W}$ ) 两部分.  $L_{1,W}$  定义如下:

$$L_{1,W}(I_F, I_{GT}) = \frac{\frac{1}{N} \sum_{i=1}^N |I_{F,i} - I_{GT,i}| (1 + \lambda_W W_{o,i})}{\frac{1}{N} (1 + \lambda_W W_{o,i})} \quad (12)$$

其中,  $I_{F,i}$ 、 $I_{GT,i}$  和  $W_{o,i}$  分别表示最终重建结果、真值图像和感知加权图中第  $i$  个像素位置的值;  $\lambda_W$  为权重参数, 其设置为 0.1.

$L_{VGG,W}$  为计算感知加权后不同尺度下的特征差异, 如下所示:

$$L_{VGG,W} = \sum_j^M \lambda_{VGG,j} L_{1,W}(f_{F,j}, f_{GT,j}) \quad (13)$$

其中,  $M$  表示特征尺度, 其设置为 5;  $\lambda_{VGG,j}$  表示第  $j$  层特征对应的权重参数, 依次为 1/32, 1/16, 1/8, 1/4, 1;  $f_{F,j}$  和  $f_{GT,j}$  分别表示重建结果和真值图像对应的第  $j$  层特征图.

## 4 实验结果与分析

在本节中, 首先描述实验细节, 包括数据集、参数设置以及评估指标. 然后, 将给出所提出方法与现有 10 种先进的单图像 HDR 重建方法的定量和定性比较结果, 以全面验证所提出方法的性能. 最后, 将给出消融结果以深入调查所提出方法的结构.

### 4.1 实验设置

#### 4.1.1 数据集

如表 1 所示, 采用 HDR-Real<sup>[25]</sup>, HDR-Synth<sup>[25]</sup>, LDR and HDR pair<sup>[38]</sup>, Lee 等人<sup>[39]</sup> 和 Kalantari 等人<sup>[40]</sup> 等 5 个数据集中提供的训练数据来训练所提出方法. 这其中, HDR-Synth<sup>[25]</sup> 数据集缺乏可用的 LDR 图像, 本文使用 ExpNet<sup>[11]</sup> 提供的方法来对 HDR 图像进行随机色调映射以生成对应 LDR 图像. Lee 等人<sup>[39]</sup> 的数据集包含多个曝光的 LDR 图像, 本文选取曝光值为 -2, 0, +2 的 LDR 图像来构造训练集. Kalantari 等人<sup>[40]</sup> 的数据集原本是用于动态多曝光 HDR 成像, 因而本文选择与参考 HDR 图像具有匹配前景的中间曝光 LDR 图像用作训练集. 最终共收集到 11 028 组包含各种场景分布 (室内和室外、白天和黑夜、人、动植物) 和不同分辨率的训练图像对 (即 LDR-HDR 对). 在训练中, 采用包含翻转和旋

转等数据增强方式.

对于测试集, 本文采用 HDR-Real<sup>[25]</sup>, Lee 等人<sup>[39]</sup>, HDR-Eye<sup>[41]</sup> 和 Hanji 等人<sup>[42]</sup> 等 4 个数据集, 其分别包含 1 838、144、42 和 181 组测试图像对. 因此, 最终共得到 2 205 组测试图像对.

表 1 本文实验所使用数据集的详细介绍

数据集	训练集	测试集
HDR-Real <sup>[25]</sup>	9 786	1 838
HDR-Synth <sup>[25]</sup>	496	/
LDR and HDR pair <sup>[38]</sup>	528	/
Lee 等人 <sup>[39]</sup>	144	144
Kalantari 等人 <sup>[40]</sup>	74	/
HDR-Eye <sup>[41]</sup>	/	42
Hanji <sup>[42]</sup>	/	181
合计	11 028	2 205

#### 4.1.2 实验环境

实验环境配置为 Intel Core i5, 32 G 内存, NVIDIA RTX 4060Ti (16 G). 所提出方法采用 PyTorch 框架实现, 使用  $\beta_1=0.9$  和  $\beta_2=0.999$  的 Adam 优化器来训练网络. 初始学习率设置为  $1 \times 10^{-4}$ , 并使用 StepLR 学习率衰减调度器, 步长为 5, 缩放因子为 0.8. 在训练期间, 每幅图像被裁剪为尺寸为  $224 \times 224$  的图像块, 并且批量尺寸设置为 4. 总体训练时间约为 2 天.

#### 4.1.3 评价指标

考虑到对 HDR 图像直接进行评估难以得到符合人眼感知的结果, 本文采用感知统一技术 (Perceptually Uniform21, PU21)<sup>[43]</sup> 来对重建图像和标签图像进行编码; 其中, HDR 显示器峰值亮度设置为  $4\ 000\ \text{cd/m}^2$ . 同时, 如文献 [42, 44] 中所述, 本文采取相机响应函数 (Camera Response Function, CRF) 矫正方法对重建图像进行相应于标签图像的 CRF 矫正, 以重点关注它们的结构差异. 在此基础上, 本文采用 PSNR, SSIM 以及 MS-SSIM 作为客观评价指标. 此外, 还采用特定于 HDR 数据的评价指标, 即 HDR-VDP-3<sup>[45]</sup>. 需要指出, HDR-VDP-3 所涉及的相关超参数, 即显示器尺寸、分辨率和观看距离分别设置为 24 寸、 $1\ 920 \times 1\ 080$  和 0.5 m. 上述 4 个指标均是越高表示 HDR 图像质量越好.

### 4.2 与最先进方法的比较

为实现全面评估, 将所提出方法与现有 10 种具有代表性的单图像 HDR 重建方法进行比较, 包括 3 种传统方法, 即 LEO<sup>[8]</sup>、KEO<sup>[18]</sup> 和 KOEO<sup>[19]</sup>, 以及 8 种基于学习的方法, 即 ExpNet<sup>[11]</sup>、FHDR<sup>[12]</sup>、MaskCNN<sup>[21]</sup>、SHDR<sup>[25]</sup>、HDRUNe<sup>[22]</sup>、Kunet<sup>[13]</sup> 和 MSP<sup>[14]</sup>.

#### 4.2.1 定量评估

表 2 中给出了定量结果, 其中粗体表示最好的结果, 下划线表示第二好的结果, 后文表格同理. 可以观

察到,所提出方法在多个数据集上对于大多数指标取得了较好的结果.例如,对于HDR-Eye数据集上,所提出方法在HDR-VDP-3指标上比第二好方法高了0.106;对于Hanji数据集上,所提出方法在PSNR指标上超越了第二好方法0.27 dB.更具体地,LEO<sup>[8]</sup>的全局处理机制忽视了场景内容特性,因而获得了较低的质量分数.KEO<sup>[18]</sup>和KOE<sup>[19]</sup>通过局部处理以考虑不同内容,获得了改进的性能.相比于上述传统方法,基于学习的方

法实现了相对更高的客观质量.ExpNet<sup>[11]</sup>和FHDR<sup>[12]</sup>由于简单的模型架构而缺少足够的特征表征能力,因而改进有限.HDRUnet<sup>[22]</sup>和Kunet<sup>[14]</sup>主要关注过曝光区域的重建和去噪,因此在场景亮度整体较高的Hanji数据集上取得了不错的性能,但在包含许多欠曝光场景的HDR-Real数据集上表现欠佳.SHDR<sup>[25]</sup>通过专门设计的恢复流程而实现了相对较好的HDR重建结果,但仍弱于所提出方法获得的结果.

表 2 不同单图像 HDR 重建方法的定量比较

数据集	方法	传统方法			基于学习的方法							
		评价指标	LEO <sup>[8]</sup>	KEO <sup>[18]</sup>	KOE <sup>[19]</sup>	ExpNet <sup>[11]</sup>	FHDR <sup>[12]</sup>	HDRUnet <sup>[22]</sup>	MaskCNN <sup>[21]</sup>	SHDR <sup>[25]</sup>	Kunet <sup>[13]</sup>	MSP <sup>[14]</sup>
HDR-Eye <sup>[41]</sup>	PU-PSNR	18.30	23.49	23.27	23.98	24.18	21.95	22.97	<u>24.73</u>	22.02	24.13	<b>24.83</b>
	PU-SSIM	0.690 9	0.838 5	0.832 8	0.842 6	0.846 3	0.813 3	0.831 8	<b>0.855 7</b>	0.802 9	0.847 4	<u>0.853 0</u>
	PU-MS-SSIM	0.788 2	0.948 0	0.947 0	0.949 0	0.954 7	0.930 5	0.945 9	<u>0.956 2</u>	0.925 0	0.952 2	<b>0.957 3</b>
	HDR-VDP-3	4.462	7.663	7.590	7.605	<u>7.770</u>	6.877	7.474	7.761	6.717	7.765	<b>7.876</b>
Hanji <sup>[42]</sup>	PU-PSNR	23.98	27.86	26.89	28.55	28.57	<u>28.61</u>	28.37	28.31	28.04	28.08	<b>28.88</b>
	PU-SSIM	0.877 2	0.920 0	0.912 0	0.933 8	0.926 7	<b>0.958 5</b>	0.920 0	0.931 5	<u>0.953 8</u>	0.932 5	0.932 9
	PU-MS-SSIM	0.901 1	0.969 5	0.961 6	0.963 4	0.971 0	<b>0.973 2</b>	<u>0.972 3</u>	0.971 1	0.968 5	0.968 0	0.971 6
	HDR-VDP-3	5.776	7.502	7.135	7.428	7.591	<b>7.943</b>	7.271	7.686	<u>7.694</u>	7.644	7.689
HDR-Real <sup>[25]</sup>	PU-PSNR	19.91	23.65	23.51	22.98	23.83	22.40	23.50	<u>24.56</u>	22.34	23.23	<b>24.59</b>
	PU-SSIM	0.727 5	0.834 1	0.826 3	0.845 7	0.836 5	0.801 5	0.825 7	0.854 6	0.782 6	0.828 7	<b>0.856 6</b>
	PU-MS-SSIM	0.791 0	0.904 7	0.904 1	0.903 3	0.908 4	0.878 8	0.903 8	<u>0.917 1</u>	0.867 5	0.889 6	<b>0.917 4</b>
	HDR-VDP-3	4.032	6.157	6.243	6.078	6.358	5.845	6.434	<b>6.676</b>	5.886	6.460	<u>6.618</u>

#### 4.2.2 定性评估

图 5 显示了不同方法的定性比较结果,使用 Reinhard 算子<sup>[46]</sup>对 HDR 图像进行压缩以可视化.其中,图 5(a)的 LDR 图像属于严重欠曝光.可以发现,LEO<sup>[8]</sup>出现了大范围的颜色偏差,而 KEO<sup>[18]</sup>,KOE<sup>[19]</sup>和 MaskCNN<sup>[21]</sup>的重建结果出现了明显的紫色色块.ExpNet<sup>[11]</sup>产生了偏黄的色调偏移,并对窗户倒影的细节恢复不佳.FHDR<sup>[12]</sup>的结果整体偏紫色,这可能是由于其反馈机制中的错误信息逐渐累积导致.HDRUnet<sup>[22]</sup>和 Kunet<sup>[13]</sup>在设计上侧重过曝光区域的恢复以及噪声去除,这导致对欠曝光图像的鲁棒性较差,即重建结果出现大面积令人不适的黑色模糊区域.SHDR<sup>[25]</sup>产生了不自然的色调,这可能是其线性化网络进行了错误的 CRF 逆转.MSP<sup>[14]</sup>不能恢复严重欠曝光的内容,同时产生了错误的颜色外观.相比之下,所提出方法更好地恢复了暗区域的信息,并且在色调和细节上更接近真值 Ground truth.

图 5(b)展现了对于中等曝光图像的重建结果,其中,过曝光的伞面和阴影中的门框是恢复的难点.KEO<sup>[18]</sup>对伞面恢复较好,但在门框部分出现了许多红色伪影.KOE<sup>[19]</sup>在伞面区域生成了不自然的纹理.ExpNet<sup>[11]</sup>对伞面和门口细节的恢复欠佳.FHDR<sup>[12]</sup>产生的结果整体色调偏暗紫.HDRUnet<sup>[22]</sup>在伞面上出现了令人不适的伪影,MaskCNN<sup>[21]</sup>和 Kunet<sup>[13]</sup>不能很好地恢

复伞面细节,同时 Kunet<sup>[13]</sup>在门框部分过于模糊.SHDR<sup>[25]</sup>在褐色伞面出现红色伪影,并且和 MSP<sup>[14]</sup>一样在门框部分的内容恢复不佳,整体偏暗.与之相比,所提出方法在保证整体色调不出现偏差的情况下,较好地恢复了曝光不良区域的细节.

在图 5(c)中,可重点关注过曝光的雕像区域.KEO<sup>[18]</sup>,KOE<sup>[19]</sup>,MaskCNN<sup>[21]</sup>和 Kunet<sup>[13]</sup>无法恢复雕像头部的纹理,同时 KEO<sup>[18]</sup>和 KOE<sup>[19]</sup>在花盆处引入了明显的噪声.ExpNet<sup>[11]</sup>和 FHDR<sup>[12]</sup>能恢复一定的细节,但其结果颜色偏黄并且过渡不柔和.HDRUnet<sup>[22]</sup>产生了过于光滑的效果,MSP<sup>[14]</sup>产生结果的整体对比度较低.相比之下,所提出方法能够较好地恢复纹理,同时避免了令人不适的伪影,具有较好的自然度.上述视觉结果证明了所提出方法具有更优越的 HDR 重建性能.

#### 4.2.3 效率比较

本节比较了不同基于学习的单图像 HDR 重建方法的计算效率,包含参数量、计算量(Floating Point Operations, FLOPs)和运行时间,如表 3 所示.其中,参数量单位为 M, FLOPs 单位为 G, 运行时间单位为 s.所有方法均在统一环境下测试,环境配置为 Intel Core i5, 32 G 内存, NVIDIA RTX 4060Ti (16G), 测试图像尺寸为 256 × 256.为避免随机性,运行时间是对 20 幅图像进行统计并取均值来展示.联合表 2 和表 3 的数值可以发现,相较于

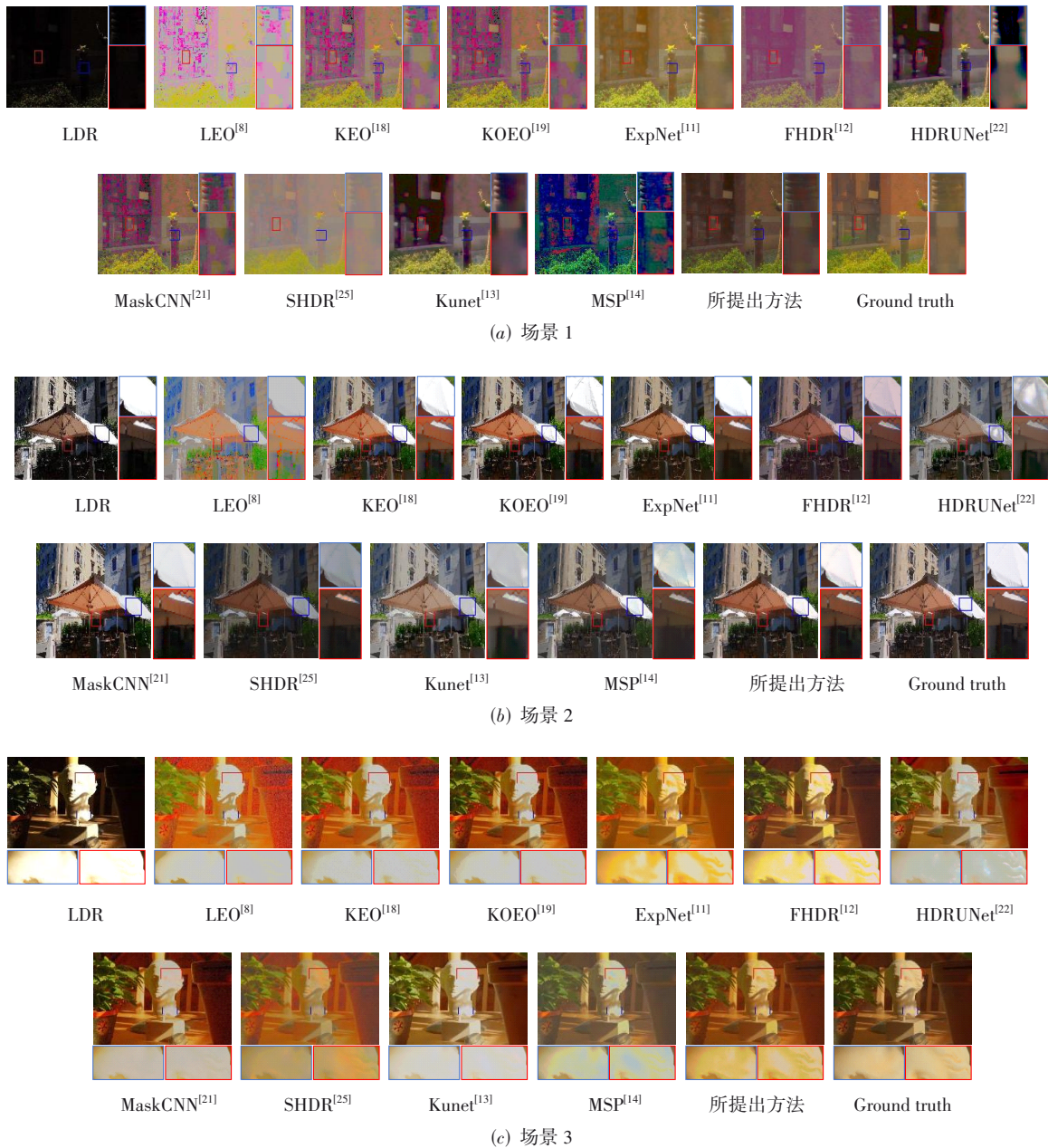


图 5 不同的单图像HDR重建方法的视觉结果比较

基于CNN的方法,如ExpNet<sup>[11]</sup>,MaskCNN<sup>[21]</sup>等,所提出方法的运行时间相对更长,这归因于自注意力机制带来的计算开销.进一步地,所提出方法在包含较少的参数量和可接受的计算量下取得了有竞争力的HDR重建结果,这表明所提出方法更适用于高精度需求的场合.

#### 4.3 不同色调映射算子下的比较

色调映射算子的不同可能会造成可视化效果的差异,因而只使用特定的算子可能会导致比较不公平.为此,本文进一步对不同色调映射算子下的可视化内容进

表 3 不同基于学习的单图像HDR重建方法的效率比较

方法	参数量/M	计算量/G	运行时间/s
ExpNet <sup>[11]</sup>	0.456 9	13.621 9	0.001
FHDR <sup>[12]</sup>	0.571 3	72.280 4	0.012
HDRUNet <sup>[22]</sup>	1.651 5	23.274 2	0.006
MaskCNN <sup>[21]</sup>	51.541 9	18.943 7	0.003
Kunet <sup>[13]</sup>	2.160 9	58.829 3	0.008
MSP <sup>[14]</sup>	12.570 6	316.153 7	0.059
SHDR <sup>[25]</sup>	28.316 5	112.787 5	0.514
所提出方法	0.770 6	50.221 0	0.071

行了全面比较.

#### 4.3.1 定量评估

本文使用了经典指标对不同色调映射算子下各方法重建的HDR结果进行客观评价,如表4所示.表4中每一种指标对应的3行数据从上到下分别对应 Reinhard<sup>[46]</sup>, Mantiuk<sup>[47]</sup>和 Drago<sup>[48]</sup>色调映射算子处理下

图像的客观质量.这些数值会随着所使用色调映射算子的变化而变化,因此在评测HDR重建上并不具备绝对意义,但仍可在一定程度上验证不同方法的性能.可以看到,所提出方法在大多数指标上取得了较好的结果,这表明所提出方法对于不同色调映射算子的泛用性,也具备更优异的HDR重建性能.

表4 不同色调映射算子下各方法的定量比较

色调映射算子	方法	数据集								
		HDR-Eye <sup>[41]</sup>			Hanji <sup>[42]</sup>			HDR-Real <sup>[25]</sup>		
		PSNR	SSIM	MS-SSIM	PSNR	SSIM	MS-SSIM	PSNR	SSIM	MS-SSIM
Reinhard <sup>[46]</sup>	LEO <sup>[8]</sup>	15.04	0.552 7	0.733 6	15.29	0.730 5	0.753 8	11.00	0.467 4	0.550 1
	KEO <sup>[18]</sup>	18.54	0.668 3	0.868 6	15.29	0.750 4	0.767 3	14.39	0.568 3	0.632 4
	KOEO <sup>[19]</sup>	18.52	0.673 7	0.867 3	15.03	0.738 9	0.761 4	14.56	0.574 3	0.641 9
	ExpNet <sup>[11]</sup>	19.68	0.717 7	0.905 5	16.23	0.787 5	0.839 4	16.51	<u>0.692 0</u>	0.744 8
	FHDR <sup>[12]</sup>	<u>20.61</u>	0.707 2	0.915 3	17.99	0.818 8	0.859 9	16.95	0.590 4	0.758 4
	HDRUNet <sup>[22]</sup>	18.30	0.668 4	0.865 0	16.10	0.794 6	0.875 6	12.92	0.481 9	0.689 3
	MaskCNN <sup>[21]</sup>	18.41	0.670 6	0.858 2	17.47	0.790 6	0.843 3	14.99	0.585 4	0.680 9
	SHDR <sup>[25]</sup>	19.63	0.728 5	0.888 6	<b>18.88</b>	<b>0.840 3</b>	0.864 9	<b>17.93</b>	0.679 2	<b>0.778 8</b>
	Kunet <sup>[13]</sup>	18.08	0.683 3	0.861 4	15.55	0.799 3	0.877 5	13.60	0.572 1	0.670 3
	MSP <sup>[14]</sup>	20.34	<b>0.750 4</b>	<b>0.924 9</b>	<u>18.22</u>	<u>0.833 1</u>	<u>0.866 1</u>	15.59	0.602 0	0.726 2
所提出方法	<b>21.09</b>	<u>0.731 7</u>	<u>0.923 4</u>	17.35	0.803 3	<b>0.867 2</b>	<u>17.89</u>	<b>0.703 5</b>	<u>0.766 5</u>	
Mantiuk <sup>[47]</sup>	LEO <sup>[8]</sup>	14.73	0.374 4	0.701 3	15.43	0.371 4	0.798 4	17.74	0.460 5	0.714 8
	KEO <sup>[18]</sup>	17.56	0.593 6	0.862 9	17.55	0.539 5	0.808 6	19.96	0.546 7	0.781 7
	KOEO <sup>[19]</sup>	18.48	0.604 5	0.860 8	19.13	0.552 9	0.792 3	19.16	0.502 7	0.766 6
	ExpNet <sup>[11]</sup>	19.80	0.655 8	0.890 7	20.21	0.591 8	0.854 9	19.74	0.563 0	0.799 5
	FHDR <sup>[12]</sup>	20.29	0.652 1	0.907 9	22.64	0.630 6	0.876 3	21.92	0.552 9	0.841 4
	HDRUNet <sup>[22]</sup>	17.15	0.531 6	0.824 3	15.60	0.405 3	0.824 5	15.33	0.350 0	0.741 6
	MaskCNN <sup>[21]</sup>	19.06	0.621 1	0.854 6	22.80	<u>0.639 3</u>	0.903 4	20.20	0.542 7	0.803 6
	SHDR <sup>[25]</sup>	19.76	0.592 7	0.856 2	<b>24.91</b>	<b>0.708 3</b>	<b>0.941 9</b>	<b>25.05</b>	<b>0.675 8</b>	<b>0.880 2</b>
	Kunet <sup>[13]</sup>	16.70	0.528 5	0.820 8	15.77	0.417 5	0.835 1	14.99	0.377 1	0.738 9
	MSP <sup>[14]</sup>	<u>21.38</u>	<b>0.699 2</b>	<u>0.917 7</u>	19.25	0.471 0	0.862 2	18.54	0.473 6	0.780 9
所提出方法	<b>21.68</b>	<u>0.688 5</u>	<b>0.919 7</b>	<u>23.66</u>	<u>0.632 2</u>	<u>0.923 6</u>	<u>22.54</u>	<u>0.618 8</u>	<u>0.843 5</u>	
Drago <sup>[48]</sup>	LEO <sup>[8]</sup>	14.86	0.205 5	0.687 2	14.54	0.588 1	0.784 4	16.06	0.281 2	0.676 5
	KEO <sup>[18]</sup>	18.70	0.575 3	0.856 0	15.00	0.467 0	0.840 6	18.45	0.514 5	0.771 2
	KOEO <sup>[19]</sup>	19.66	0.605 0	0.867 2	15.21	0.389 8	0.859 3	18.72	0.514 2	0.779 3
	ExpNet <sup>[11]</sup>	21.51	0.696 4	0.905 2	17.81	0.469 0	0.893 5	19.60	0.560 9	0.811 4
	FHDR <sup>[12]</sup>	21.49	0.691 0	0.916 5	19.69	0.474 6	0.915 0	21.35	0.540 4	0.841 1
	HDRUNet <sup>[22]</sup>	17.15	0.552 7	0.833 4	13.88	0.273 4	0.863 2	16.01	0.430 1	0.768 0
	MaskCNN <sup>[21]</sup>	20.38	0.637 7	0.874 8	23.18	0.596 9	0.915 2	19.79	0.536 9	0.813 2
	SHDR <sup>[25]</sup>	21.87	0.683 0	0.896 4	<b>27.09</b>	<b>0.742 3</b>	<b>0.935 4</b>	<b>23.68</b>	<b>0.629 4</b>	<b>0.870 0</b>
	Kunet <sup>[13]</sup>	16.68	0.551 8	0.830 5	13.98	0.278 7	0.873 8	15.69	0.423 9	0.763 5
	MSP <sup>[14]</sup>	<u>22.00</u>	<b>0.719 5</b>	<u>0.919 9</u>	22.27	0.633 0	0.904 5	18.91	0.500 6	0.790 2
所提出方法	<b>22.85</b>	<u>0.717 2</u>	<b>0.927 7</b>	<u>24.21</u>	<u>0.688 7</u>	<u>0.933 2</u>	<u>22.29</u>	<u>0.602 0</u>	<u>0.848 8</u>	

#### 4.3.2 定性评估

图6中展示了同一组HDR图像在不同色调映射算子下的可视化结果,所使用的色调映射算子依次为 Reinhard<sup>[46]</sup>, Mantiuk<sup>[47]</sup>和 Drago<sup>[48]</sup>,图分题中ENet, HU-

Net, MCNN分别为ExpNet, HDRUNet和MaskCNN的缩写.可以观察到,对于传统方法,LEO<sup>[8]</sup>的结果在不同色调映射下差别偏大,并且难以恢复正确的颜色和结构信息;KEO<sup>[18]</sup>和KOEO<sup>[19]</sup>会出现色彩断层.对于基于

学习的方法, ExpNet<sup>[11]</sup>出现了一定程度上的色调偏差, FHDR<sup>[12]</sup>在曝光不良的情况下会产生整体偏紫的结果. HDRUnet<sup>[22]</sup>在过曝光区域产生了视觉伪影以及网格效应,在欠曝光处产生了模糊结果. SHDR<sup>[25]</sup>在某些色调映射函数下和 Ground truth(即标签图像)相距甚远. Ku-

net<sup>[13]</sup>产生了带有伪影和较低饱和度的结果. MSP<sup>[14]</sup>产生了明显伪像. 与之相比,所提出方法生成了最接近 Ground truth 的结果. 综上所述,所提出方法在具有优异性能的同时,还对不同的色调映射算法具有较好的适应性和鲁棒性.

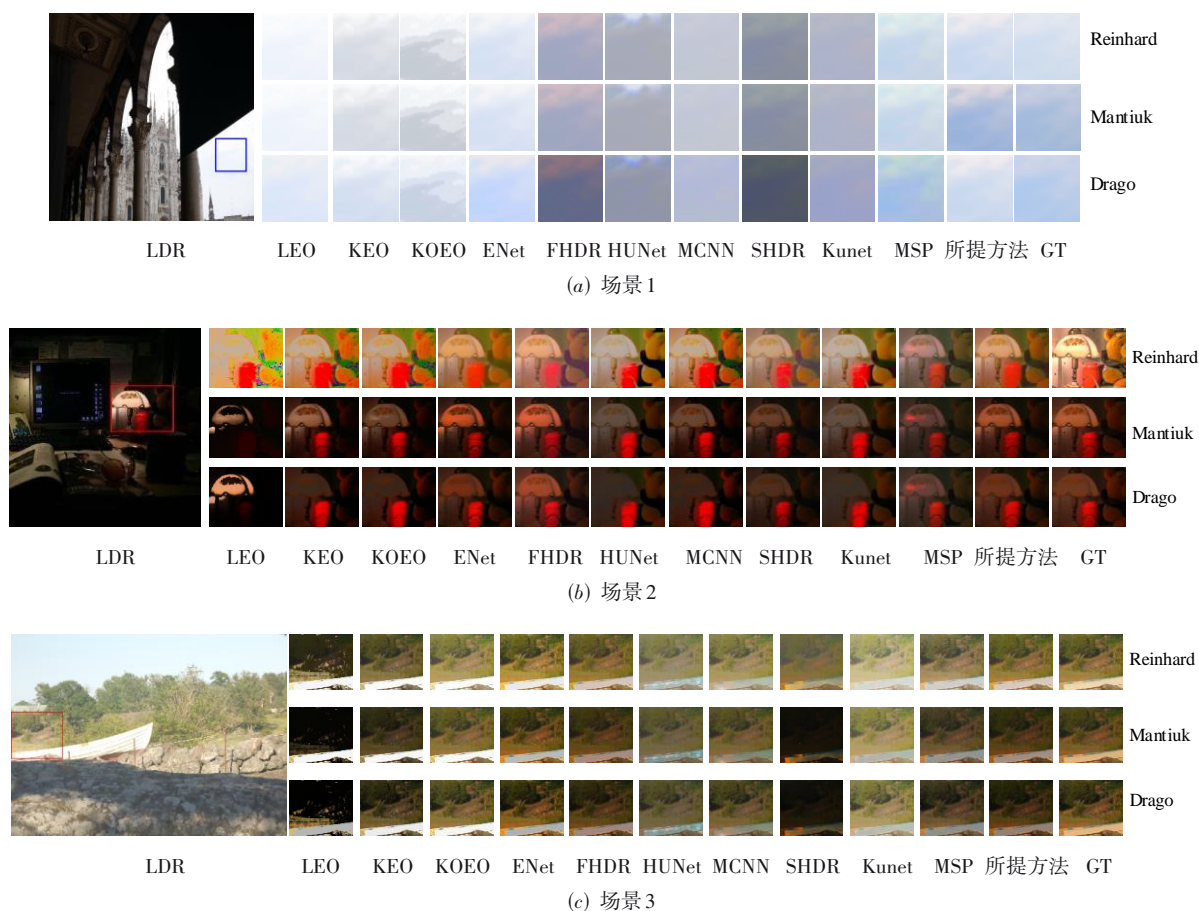


图6 不同色调映射算子下的视觉结果

#### 4.4 不同EV下HDR重建性能比较

##### 4.4.1 定量评估

为了更详细地观察不同方法在不同曝光值(Exposure Value, EV)下的表现,本文在Lee等人<sup>[39]</sup>提供的数据集上实验了不同方法在EV为+2,0,-2上的HDR重建结果.

表5给出了定量结果,其中,单元格内的3个指标分别表示在EV=+2, EV=0, EV=-2下的性能. HDRUnet<sup>[22]</sup>, Kunet<sup>[13]</sup>和MSP<sup>[14]</sup>由于在恢复欠曝光区域的糟糕表现,在EV=-2情况下得分较低. MaskCNN<sup>[21]</sup>对于过曝光区域的恢复欠佳. 所提出方法凭借全局和局部的联合建模,以及对于曝光不良区域的更多关注,在各个EV值下都取得了较好的结果.

##### 4.4.2 定性评估

图7展示了在不同EV下各方法重建结果的视觉比

较,其由Reinhard<sup>[46]</sup>色调映射算子进行可视化. 在图7上半部分中,对于EV=+2的情况,对比方法在高亮墙壁区域出现较差的重建效果,LEO<sup>[8]</sup>,KEO<sup>[18]</sup>和KOKO<sup>[19]</sup>的结果亮度过高同时泛白;HDRUnet<sup>[22]</sup>和SHDR<sup>[25]</sup>的结果出现光斑伪影;MaskCNN<sup>[21]</sup>和Kunet<sup>[13]</sup>的结果过于模糊. 本文方法在无明显伪影的情况下较好地保持了墙壁的结构和纹理. 对于EV=-2的情况,LEO<sup>[8]</sup>产生了较大色调偏差, FHDR<sup>[12]</sup>整体色调偏紫;另外,KEO<sup>[8]</sup>,KOEO<sup>[19]</sup>,ExpNet<sup>[11]</sup>,HDRUnet<sup>[22]</sup>,Kunet<sup>[13]</sup>和MSP<sup>[14]</sup>均无法较好地恢复门上的装饰;HDRUnet<sup>[22]</sup>和Kunet<sup>[13]</sup>还在该处产生了绿色伪像. 在图7下半部分中,对于EV=+2的情况,LEO<sup>[8]</sup>无法恢复树顶过曝光区域的细节,HDRUnet<sup>[22]</sup>,MaskCNN<sup>[21]</sup>,SHDR<sup>[25]</sup>,Kunet<sup>[13]</sup>和MSP<sup>[14]</sup>均产生了不同程度的形态或颜色伪像. 此外,对于EV=0时,Kunet<sup>[13]</sup>出现了光斑伪影. 对于EV=-2的情况,LEO<sup>[8]</sup>,FHDR<sup>[12]</sup>和

表 5 文献[39]数据集的定量评估

方法	PU-PSNR	PU-SSIM	PU-MS-SSIM	HDR-VDP-3
/	EV=+2, EV=0, EV=-2			
LEO <sup>[8]</sup>	20.55/21.99/19.17	0.868 4/0.858 2/0.795 5	0.890 4/0.884 6/0.787 8	5.278/5.762/3.522
KEO <sup>[18]</sup>	22.16/28.89/29.14	0.933 9/0.967 2/0.962 7	0.928 8/0.980 7/0.979 7	5.932/8.278/8.506
KOEO <sup>[19]</sup>	20.76/26.44/29.44	0.885 8/0.946 5/0.963 1	0.888 5/0.960 7/0.983 9	5.383/7.658/8.721
ExpNet <sup>[11]</sup>	22.46/28.63/28.54	0.924 4/0.959 6/0.948 9	0.915 5/0.974 4/0.977 6	5.821/8.018/8.358
FHDR <sup>[12]</sup>	22.59/28.61/27.45	0.930 1/0.967 1/0.965 0	0.929 4/0.981 9/0.984 9	5.842/8.161/8.653
HDRUNet <sup>[22]</sup>	20.78/25.03/24.12	0.884 0/0.926 9/0.886 4	0.884 6/0.948 8/0.944 0	5.923/7.360/6.935
MaskCNN <sup>[21]</sup>	20.66/29.69/27.33	0.908 3/0.963 5/0.960 1	0.892 6/0.985 7/0.973 6	6.388/8.674/8.378
SHDR <sup>[25]</sup>	<u>23.19/29.71/29.45</u>	0.932 0/0.968 0/0.966 1	0.928 9/0.982 1/0.985 2	<u>7.062/8.535/8.321</u>
Kunet <sup>[13]</sup>	20.03/24.40/25.05	0.888 3/0.917 7/0.861 3	0.868 2/0.931 8/0.924 2	6.101/7.161/6.904
MSP <sup>[14]</sup>	<u>23.94/28.54/24.22</u>	<u>0.942 2/0.958 4/0.899 8</u>	<u>0.930 4/0.969 4/0.917 6</u>	<u>6.821/8.213/7.201</u>
本文提出方法	22.86/29.42/29.32	<u>0.935 9/0.968 4/0.965 6</u>	<u>0.933 3/0.983 0/0.985 6</u>	6.133/8.365/8.815

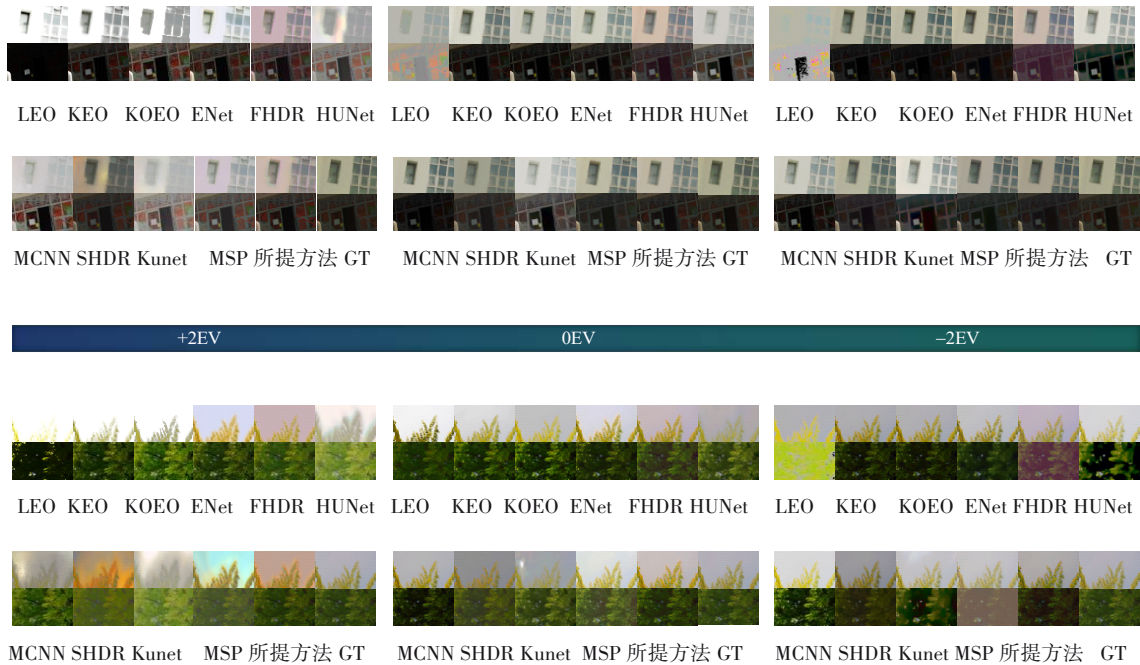


图 7 不同EV下各方法的视觉结果

SHDR<sup>[25]</sup>的结果产生了色调误差, HDRUNet<sup>[22]</sup>和 Kunet<sup>[13]</sup>引入了模糊伪像. 相比之下, 本文方法没有产生令人不适的伪像, 并且生成了和真值更为接近的结果.

#### 4.5 真实场景上的验证

为了进一步验证所提出方法的泛化能力, 利用佳能 EOS 90D 数码相机拍摄了多组真实 LDR 图像, 并选取表现较好的深度学习方法<sup>[13, 14, 22, 25]</sup>构造了相应的对比实验. 图 8 展示了视觉比较结果. 可以观察到, 所提出方法重建的 HDR 图像具有更高的视觉质量. 具体而言, 在场景 1 中, 大部分对比方法难以恢复模型猫在腿部过曝光区域的信息, 所提出方法较好地恢复了黄色斑纹细节. 在场景 2 中, 所提出方法重建的 HDR 图

像具有视觉恰当的亮度和对比度, 并表现出了最高的自然度和保真度. 在场景 3 中, HDRUNet<sup>[22]</sup>和 Kunet<sup>[13]</sup>重建的结果在地面区域出现颜色失真, SHDR<sup>[25]</sup>重建的结果亮度过低; 相比之下, 所提出方法重建了具有更高颜色饱和度和更高亮度的结果. 上述实验结果验证了所提出方法在未见过的真实场景中的有效性和泛化能力.

#### 4.6 消融实验

在本节中, 首先构建了 4 种消融方案来深入调查所提出方法中核心部分的有效性.

方案 1. 只使用由窗口 Transformer 块构成网络, 即去除局部交互.

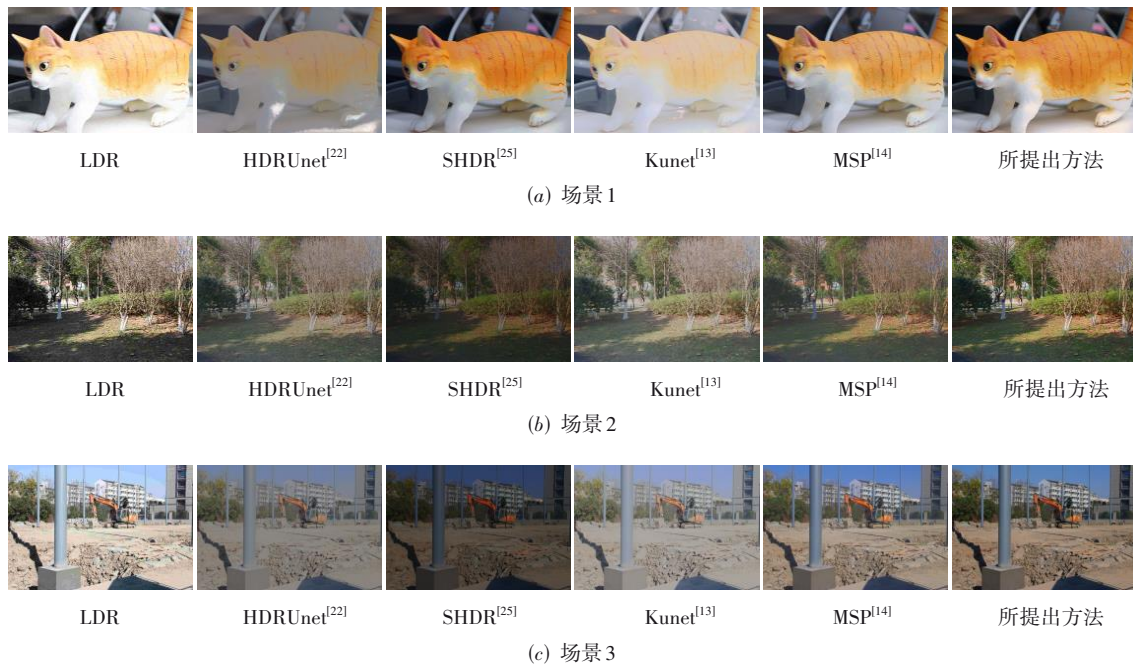


图8 在真实场景下不同HDR重建方法的视觉结果

方案2. 使用MA-ViT代替方案1中的Transformer块.

方案3. 在方案2的基础上,使用所提出的损失感知加权图.

方案4. 在方案2的基础上,引入针对结构和颜色信息复原的增强分支.

表6给出了消融实验的定量结果. 相较于只使用窗口Transformer的方案1,所提出的MA-ViT带来了可观的性能增益,在各项指标上均有不同程度的提升. 对比方案2和方案3可知,在引入感知加权图后,方案3在多数指标上取得了更好的表现,如在PSNR上比方案2平均高出0.07 dB. 方案4的指标说明增强分支在某些数据集下能够起到重要作用,如在HDR-Real数据集上,方案4的各项指标均高于方案2. 总之,所提出方法结合了以上部件,在大多数指标上取得了最好的结果.

其次,针对HDR重建损失设计,本文对比了不同的

损失函数组合. 为了追求更好的感知质量,本文引入了LPIPS (Learned Perceptual Image Patch Similarity)<sup>[49]</sup>损失  $L_{LPIPS}$ . 同时,构建了如图9所示的判别器以构成生成对抗学习架构,判别损失  $L_D$  采用二元交叉熵损失. 实验结果如表7所示,其中,  $L_1+L_{VGG}$  为实际采用的损失组合. 可以看到,在所比较的4种损失组合中,  $L_1+L_{LPIPS}$  在各项指标上表现平平;在此基础上加入感知损失,即  $L_{VGG}$  后,各项指标都得到了可观的增益. 另外,  $L_1+L_{VGG}+L_D$  的结果不尽如人意,这可能是由于不同损失的特性导致网络学习困难. 值得注意的是,本文使用的  $L_1+L_{VGG}$  在绝大多数指标上取得了最高的定量结果,因此将其作为所提出方法的最终损失组合.

此外,考虑到在所提出的MA-ViT中,空间注意力加权结果与通道注意力加权结果之间的不同融合方式可能带来性能差异,本文对乘法和加法两种融合方式

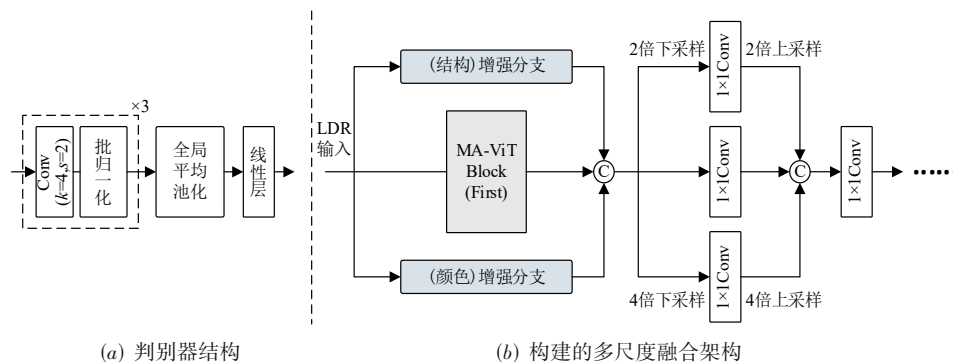


图9 消融实验中构造的相关网络示意图

表 6 所提出方法中核心部件的消融结果

指标	消融方案	HDR-Eye <sup>[41]</sup>	Hanji <sup>[42]</sup>	HDR-Real <sup>[25]</sup>
PU-PSNR	方案 1	19.19	21.75	21.01
	方案 2	24.68	28.62	24.52
	方案 3	24.74	<u>28.68</u>	<b>24.60</b>
	方案 4	<u>24.79</u>	28.57	24.58
	本文提出方法	<b>24.83</b>	<b>28.88</b>	<u>24.59</u>
PU-SSIM	方案 1	0.822 6	0.858 9	0.821 2
	方案 2	0.851 9	<u>0.930 9</u>	0.854 0
	方案 3	<b>0.853 0</b>	0.925 0	0.855 1
	方案 4	0.850 8	0.928 2	<u>0.855 6</u>
	本文提出方法	<b>0.853 0</b>	<b>0.932 9</b>	<b>0.856 6</b>
PU-MS-SSIM	方案 1	0.955 5	0.904 0	0.869 7
	方案 2	0.956 0	<u>0.971 0</u>	0.915 5
	方案 3	<u>0.956 5</u>	0.969 3	<u>0.916 4</u>
	方案 4	0.955 7	0.968 5	0.915 9
	本文提出方法	<b>0.957 3</b>	<b>0.971 6</b>	<b>0.917 4</b>
HDR-VDP-3	方案 1	6.800	6.295	5.775
	方案 2	7.863	<b>7.712</b>	6.618
	方案 3	<u>7.873</u>	7.652	<b>0.658</b>
	方案 4	7.862	7.576	<u>6.649</u>
	本文提出方法	<b>7.876</b>	<u>7.689</u>	6.618

表 7 不同损失函数组合的消融结果

指标	损失组合	HDR-Eye <sup>[41]</sup>	Hanji <sup>[42]</sup>	HDR-Real <sup>[25]</sup>
PU-PSNR	$L_1+L_{LPIPS}$	24.62	28.21	24.22
	$L_1+L_{LPIPS}+L_{VGG}$	<u>24.80</u>	<u>28.51</u>	<u>24.56</u>
	$L_1+L_{VGG}+L_D$	19.59	25.06	21.04
	$L_1+L_{VGG}$	<b>24.83</b>	<b>28.88</b>	<b>24.59</b>
PU-SSIM	$L_1+L_{LPIPS}$	0.844 2	0.932 3	0.838 9
	$L_1+L_{LPIPS}+L_{VGG}$	<u>0.852 1</u>	<b>0.936 4</b>	<u>0.854 5</u>
	$L_1+L_{VGG}+L_D$	0.810 1	0.913 6	0.815 2
	$L_1+L_{VGG}$	<b>0.853 0</b>	<u>0.932 9</u>	<b>0.856 6</b>
PU-MS-SSIM	$L_1+L_{LPIPS}$	0.952 5	0.964 4	0.907 6
	$L_1+L_{LPIPS}+L_{VGG}$	<u>0.956 8</u>	<u>0.970 2</u>	<u>0.916 7</u>
	$L_1+L_{VGG}+L_D$	0.913 2	0.952 3	0.877 2
	$L_1+L_{VGG}$	<b>0.957 3</b>	<b>0.971 6</b>	<b>0.917 4</b>
HDR-VDP-3	$L_1+L_{LPIPS}$	7.720	7.523	6.476
	$L_1+L_{LPIPS}+L_{VGG}$	<u>7.851</u>	<u>7.593</u>	<u>6.609</u>
	$L_1+L_{VGG}+L_D$	6.656	6.749	5.700
	$L_1+L_{VGG}$	<b>7.876</b>	<b>7.689</b>	<b>6.618</b>

进行研究. 表 8 给出了乘法融合和所提出方法(即加法融合)的定量比较, 其中, 加法融合为实际采用的融合方式. 可以看到, 相较于乘法融合, 加法融合在各个指标上都取得了更高的质量值. 原因可能在于乘法操作难以互补不同特征间的信息, 而加法操作可有效整合

不同特征信息, 进而提升重建质量.

最后, 为了研究多尺度架构对单图像 HDR 重建性能的影响, 本文在增强分支和第一个 MA-ViT 的融合节点处构建了多尺度融合架构进行实验. 如图 9(b) 所示, 采用双线性插值以获得不同尺度下的特征图, 进而通过卷积来实现多尺度特征融合.

表 8 不同融合方式的消融结果

指标	融合架构	HDR-Eye <sup>[41]</sup>	Hanji <sup>[42]</sup>	HDR-Real <sup>[25]</sup>
PU-PSNR	乘法融合	24.75	28.51	24.52
	加法融合	<b>24.83</b>	<b>28.88</b>	<b>24.59</b>
PU-SSIM	乘法融合	0.852 8	0.928 2	0.853 5
	加法融合	<b>0.853 0</b>	<b>0.932 9</b>	<b>0.856 6</b>
PU-MS-SSIM	乘法融合	0.956 7	0.968 0	0.915 5
	加法融合	<b>0.957 3</b>	<b>0.971 6</b>	<b>0.917 4</b>
HDR-VDP-3	乘法融合	7.868	7.512	6.592
	加法融合	<b>7.876</b>	<b>7.689</b>	<b>6.618</b>

表 9 给出了多尺度架构与所提出方法的定量比较. 可以看到, 在某些指标上, 多尺度架构能带来一定的性能增益. 然而, 单尺度架构在大多数指标上取得了更好的表现, 这主要归因于曝光不良图像在不同尺度之间的特征差异并不明显; 在融合过程中, 语义特征和细节特征可能相互掩盖. 另外, 所提出方法使用全局自注意力以及局部空间与通道注意力而考虑了不同感受野下的特征学习. 因此, 使用单尺度融合架构在满足任务要求的同时, 减轻了计算资源的开销.

表 9 不同融合架构的消融结果

指标	融合架构	HDR-Eye <sup>[41]</sup>	Hanji <sup>[42]</sup>	HDR-Real <sup>[25]</sup>
PU-PSNR	多尺度架构	24.81	28.97	24.59
	单尺度架构	24.83	28.88	24.59
PU-SSIM	多尺度架构	0.852 7	0.932 1	0.856 2
	单尺度架构	0.853 0	0.932 9	0.856 6
PU-MS-SSIM	多尺度架构	0.957 0	0.971 4	0.917 5
	单尺度架构	0.957 3	0.971 6	0.917 4
HDR-VDP-3	多尺度架构	7.872	7.638	6.620
	单尺度架构	7.876	7.689	6.618

## 5 结论

考虑到现有方法对曝光不良区域的关注不足, 在这项工作中, 提出了一个新的基于多重注意力和感知加权学习的单图像高动态范围(HDR)重建方法. 所提出的多重注意力视觉 Transformer 可联合建模全局与局部特征依赖, 并在不同维度上强调重要特征, 从而为高质量 HDR 重建实现有效的特征学习. 此外, 本文设计了一种感知损失加权图以引导网络聚焦由曝光不良而

导致信息丢失严重的区域,这可为增强网络重建能力提供有效指导. 在多个基准数据集上的广泛实验表明,所提出方法能在提高动态范围的同时,较好地恢复曝光不良区域的细节信息,并且能在多种色调映射函数处理下生成高质量的可视化内容;此外,本文方法对不同曝光的输入也具有鲁棒性. 在未来工作中,我们将探索更有效的注意力机制使用方式,并探索新的生成模型,如扩散模型等,以进一步提高单图像 HDR 重建的性能.

#### 参考文献

- [1] LI H F, YANG Z M, ZHANG Y F, et al. Single-image HDR reconstruction assisted ghost suppression and detail preservation network for multi-exposure HDR imaging[J]. *IEEE Transactions on Computational Imaging*, 2024, 10: 429-445.
- [2] 张雨童, 邓欣, 徐迈. 动态场景下深度自监督多曝光图像融合方法[J]. *电子学报*, 2024, 52(1): 264-273.  
ZHANG Y T, DENG X, XU M. Deep self-supervised multi-exposure image fusion for dynamic scenes[J]. *Acta Electronica Sinica*, 2024, 52(1): 264-273. (in Chinese)
- [3] LIU S Z, ZHANG X D, SUN L C, et al. Joint HDR denoising and fusion: A real-world mobile HDR image dataset[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2023: 13966-13975.
- [4] 印佳丽, 韩津, 陈斌, 等. 基于一致性感知特征融合的高动态范围成像方法[J]. *计算机学报*, 2024, 47(10): 2352-2367.  
YIN J L, HAN J, CHEN B, et al. High dynamic range imaging based on coherence-aware feature aggregation[J]. *Chinese Journal of Computers*, 2024, 47(10): 2352-2367. (in Chinese)
- [5] HU T, YAN Q S, QI Y K, et al. Generating content for HDR deghosting from frequency view[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2024: 25732-25741.
- [6] 王仙峰, 刘世本, 田建东, 等. 基于双重注意力网络的高动态范围图像重建[J]. *激光与光电子学进展*, 2024, 61(12): 1237005.  
WANG X F, LIU S B, TIAN J D, et al. High dynamic range image reconstruction based on dual-attention network[J]. *Laser & Optoelectronics Progress*, 2024, 61(12): 1237005. (in Chinese)
- [7] 郭露露, 易红伟. 基于局域加权叠加的高动态范围图像融合算法[J]. *光子学报*, 2022, 51(11): 344-353.  
GUO L L, YI H W. High dynamic range image fusion algorithm based on local weighted superposition[J]. *Acta Photonica Sinica*, 2022, 51(11): 344-353. (in Chinese)
- [8] LANDIS H. Production-ready global illumination[J]. *Signature Course Notes*, 2002, 10: 87-101
- [9] HUO Y Q, YANG F, DONG L, et al. Physiological inverse tone mapping based on retina response[J]. *The Visual Computer*, 2014, 30(5): 507-517.
- [10] EILERTSEN G, KRONANDER J, DENES G, et al. HDR image reconstruction from a single exposure using deep CNNs[J]. *ACM Transactions on Graphics*, 2017, 36(6): 1-15.
- [11] MARNERIDES D, BASHFORD-ROGERS T, HATCHETT J, et al. ExpandNet: A deep convolutional neural network for high dynamic range expansion from low dynamic range content[J]. *Computer Graphics Forum*, 2018, 37(2): 37-49.
- [12] KHAN Z, KHANNA M, RAMAN S. FHDR: HDR image reconstruction from a single LDR image using feedback network[C]//2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP). Piscataway: IEEE, 2019: 1-5.
- [13] WANG H, YE M, ZHU X T, et al. KUNet: Imaging knowledge-inspired single HDR image reconstruction[C]//Proceedings of the 31st International Joint Conference on Artificial Intelligence. Virtual Event: IJCAI Organization, 2022: 1408-1414.
- [14] ZHANG W, JIANG G Y, CHEN Y Y, et al. Multi-stage coarse-to-fine progressive enhancement network for single-image HDR reconstruction[J]. *Displays*, 2024, 84: 102791.
- [15] AKYÜZ A, FLEMING R, RIECKE B, et al. Do HDR displays support LDR content?: A psychophysical evaluation[J]. *ACM Transactions on Graphics*, 2007, 26(3): 1-8.
- [16] MASIA B, AGUSTIN S, FLEMING R W, et al. Evaluation of reverse tone mapping through varying exposure conditions[J]. *ACM Transactions on Graphics*, 2009, 28(5): 1-8.
- [17] MEYLAN L, DALY S, SÜSSTRUNK S. The reproduction of specular highlights on high dynamic range displays[J]. *Color and Imaging Conference*, 2006, 14(1): 333-338.
- [18] KUO P H, TANG C S, CHIEN S Y. Content-adaptive inverse tone mapping[C]//2012 Visual Communications and Image Processing. Piscataway: IEEE, 2012: 1-6.
- [19] KOVALESKI R P, OLIVEIRA M M. High-quality re-

- verse tone mapping for a wide range of exposures[C]//2014 27th SIBGRAPI Conference on Graphics, Patterns and Images. Piscataway: IEEE, 2014: 49-56.
- [20] WANG T H, CHIU C W, WU W C, et al. Pseudo-multiple-exposure-based tone fusion with local region adjustment[J]. *IEEE Transactions on Multimedia*, 2015, 17(4): 470-484.
- [21] SANTOS M S, REN T I, KALANTARI N K. Single image HDR reconstruction using a CNN with masked features and perceptual loss[J]. *ACM Transactions on Graphics*, 2020, 39(4): 1-10.
- [22] CHEN X Y, LIU Y H, ZHANG Z W, et al. HDRUNet: Single image HDR reconstruction with denoising and dequantization[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway: IEEE, 2021: 354-363.
- [23] 贝悦, 王琦, 程志鹏, 等. 基于条件生成对抗网络的HDR图像生成方法[J]. *北京航空航天大学学报*, 2022, 48(1): 45-52.
- BEI Y, WANG Q, CHENG Z P, et al. HDR image generation method based on conditional generative adversarial network[J]. *Journal of Beijing University of Aeronautics and Astronautics*, 2022, 48(1): 45-52. (in Chinese)
- [24] MORIWAKI K, YOSHIHASHI R, KAWAKAMI R, et al. Hybrid loss for learning single-image-based HDR reconstruction[EB/OL]. (2018-12-18) [2024-11-23]. <https://arxiv.org/abs/1812.07134v1>.
- [25] LIU Y L, LAI W S, CHEN Y S, et al. Single-image HDR reconstruction by learning to reverse the camera pipeline[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 1651-1660.
- [26] SHE L, YE M, LI S, et al. Single-image HDR reconstruction by dual learning the camera imaging process[J]. *Engineering Applications of Artificial Intelligence*, 2023, 120: 105947.
- [27] ENDO Y, KANAMORI Y, MITANI J. Deep reverse tone mapping[J]. *ACM Transactions on Graphics*, 2017, 36(6): 1-10.
- [28] LEE S, AN G H, KANG S J. Deep recursive HDRI: Inverse tone mapping using generative adversarial networks[C]//Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 613-628.
- [29] LEE S, JO S Y, AN G H, et al. Learning to generate multi-exposure stacks with cycle consistency for high dynamic range imaging[J]. *IEEE Transactions on Multimedia*, 2020, 23: 2561-2574.
- [30] JO S Y, LEE S, AHN N, et al. Deep arbitrary HDRI: Inverse tone mapping with controllable exposure changes[J]. *IEEE Transactions on Multimedia*, 2021, 24: 2713-2726.
- [31] ZHANG N, YE Y Y, ZHAO Y, et al. Revisiting the stack-based inverse tone mapping[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2023: 9162-9171.
- [32] CHEN S K, YEN H L, LIU Y L, et al. Learning continuous exposure value representations for single-image HDR reconstruction[C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2023: 12944-12954.
- [33] 陈羽中, 陈友昆, 林闽沪, 等. 基于边缘辅助和多尺度Transformer的无参考屏幕内容图像质量评估[J]. *电子学报*, 2024, 52(7): 2242-2256.
- CHEN Y Z, CHEN Y K, LIN M H, et al. No-reference screen content image quality assessment based on edge assistance and multi-scale transformer[J]. *Acta Electronica Sinica*, 2024, 52(7): 2242-2256. (in Chinese)
- [34] 雷涛, 翟钰杰, 许叶彤, 等. 基于边缘引导和动态可变形Transformer的遥感图像变化检测[J]. *电子学报*, 2024, 52(1): 107-117.
- LEI T, ZHAI Y J, XU Y T, et al. Edge guided and dynamically deformable transformer network for remote sensing images change detection[J]. *Acta Electronica Sinica*, 2024, 52(1): 107-117. (in Chinese)
- [35] LIANG J Y, CAO J Z, SUN G L, et al. SwinIR: Image restoration using swin transformer[C]//2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). Piscataway: IEEE, 2021: 1833-1844.
- [36] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(8): 2011-2023.
- [37] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]//2015 International Conference on Learning Representations. San Diego: ICLR, 2015: 1-14.
- [38] JANG H, BANG K, JANG J, et al. Dynamic range expansion using cumulative histogram learning for high dynamic range image generation[J]. *IEEE Access*, 2020, 8: 38554-38567.
- [39] LEE S, AN G H, KANG S J. Deep chain HDRI: Reconstructing a high dynamic range image from a single low dynamic range image[J]. *IEEE Access*, 2018, 6: 49913-49924.

- [40] KALANTARI N K, RAMAMOORTHI R. Deep high dynamic range imaging of dynamic scenes[J]. ACM Transactions on Graphics, 2017, 36(4): 1-12.
- [41] NEMOTO H, KORSHUNOV P, HANHART P, et al. Visual attention in LDR and HDR images[C]//2015 9th International Workshop on Video Processing and Quality Metrics for Consumer Electronics. Scottsdale: VPQM, 2015: 1-6.
- [42] HANJI P, MANTIUK R, EILERTSEN G, et al. Comparison of single image HDR reconstruction methods: The caveats of quality assessment[C]//Special Interest Group on Computer Graphics and Interactive Techniques Conference Proceedings. New York: ACM, 2022: 1-8.
- [43] MANTIUK R K, AZIMI M. PU21: A novel perceptually uniform encoding for adapting existing quality metrics for HDR[C]//2021 Picture Coding Symposium (PCS). Piscataway: IEEE, 2021: 1-5.
- [44] EILERTSEN G, HAJISHARIF S, HANJI P, et al. How to cheat with metrics in single-image HDR reconstruction[C]//2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). Piscataway: IEEE, 2021: 3981-3990.
- [45] MANTIUK R K, HAMMOU D, HANJI P. HDR-VDP-3: A multi-metric for predicting image differences, quality and contrast distortions in high dynamic range and regular content[EB/OL]. (2023-08-26) [2024-11-23]. <https://arxiv.org/abs/2304.13625v1>.
- [46] REINHARD E, STARK M, SHIRLEY P, et al. Photographic tone reproduction for digital images[J]. ACM Transactions on Graphics, 2002, 21(3): 267-276.
- [47] MANTIUK R, DALY S, KEROFISKY L. Display adaptive tone mapping[J]. ACM Transactions on Graphics, 2008, 27(3): 1-10.
- [48] DRAGO F, MYSZKOWSKI K, ANNEN T, et al. Adaptive logarithmic mapping for displaying high contrast scenes[J]. Computer Graphics Forum, 2003, 22(3): 419-426.
- [49] ZHANG R, ISOLA P, EFROS A A, et al. The unreasonable effectiveness of deep features as a perceptual metric[EB/OL]. (2018-08-10) [2024-11-23]. <https://arxiv.org/abs/1801.03924v2>.

#### 作者简介



**柯德涨** 男,2000年10月生,浙江瑞安人。宁波大学信息科学与工程学院硕士研究生。主要研究方向为高动态范围成像、光场图像处理等。  
E-mail: 15869640695@163.com



**陈晔曜** 男,1994年10月生,浙江嘉兴人。宁波大学信息科学与工程学院讲师。主要研究方向为高动态范围成像、光场成像、视频处理等。中国电子学会会员编号:E190158389M。  
E-mail: cheneyao@nbu.edu.cn



**徐海勇** 男,1981年12月生,浙江金华人。宁波大学数学与统计学院教授、博士生导师。主要研究方向为水下图像增强、视觉感知等。中国电子学会会员编号:E190013185M。  
E-mail: xuhaiyong@nbu.edu.cn



**金充充** 女,1994年10月生,浙江绍兴人。宁波大学科学技术学院讲师。主要研究方向为视频图像质量评价、视觉感知等。  
E-mail: jinchongchong@nbu.edu.cn



**蒋刚毅** 男,1964年8月生,浙江绍兴人。宁波大学信息科学与工程学院教授、博士生导师。主要研究方向为多媒体信号处理、图像处理与视频压缩、计算成像与视觉感知等。  
E-mail: jianggangyi@nbu.edu.cn