

基于双重注意力网络的高动态范围图像重建及色调映射

陈子豪, 胡 楷, 阮明航

(中国海洋大学 信息科学与工程学部计算机科学与技术学院, 山东 青岛 266404)

摘要 在高动态范围(High Dynamic Range, HDR)图像重建任务中, 当输入图像过曝光或者欠曝光时, 常见的基于深度学习的 HDR 图像重建方法容易出现细节信息丢失和色彩饱和度差的问题。为了解决这一问题, 本文提出了一种基于双重注意力网络的 HDR 图像重建方法。首先, 该方法利用双重注意力模块(Dual Attention Module, DAM)分别从像素和通道两个维度的注意力机制对过曝光和欠曝光的两张源图像进行特征提取并融合, 得到一张初步融合图像。接着, 在此基础上构建特征增强模块(Feature Enhancement Module, FEM)分别对初步融合图像进行细节增强和颜色校正。然后, 将图像传入色调映射模块(Visual-Saliency-Based Tone Mapping, VsbTM), 利用显著性感知加权和所提出的滤波器的 HDR 图像局部色调映射算法, 改善注意显著区域的视觉质量。最后, 引用对比学习使生成图像更加接近参考图像的同时远离源图像。经过多次训练, 最终生成 HDR 图像。实验结果表明本文方法在 PSNR、SSIM 和 LPIPS 指标上取得最优评价结果, 生成的 HDR 图像色彩饱和度和好且细节信息精准完整, 且 HDR 图像色调映射 LDR 图像无色晕伪影。

关键词 高动态范围成像, 双重注意力机制, 图像重建, 色调映射, 显著性感知加权

1 引言

1.1 方案选择

我们选择基于《基于双重注意力网络的高动态范围图像重建》^[1]及《Visual-Saliency-Based Tone Mapping for High Dynamic Range Images》^[2]两篇文章进行论文复现, 并对前文已有方法进行了改进。

1.2 背景介绍

1.2.1 高动态范围 HDR

照度是决定图像动态范围的关键因素。在自然光环境下, 照度的变化幅度相当大, 白天阳光明媚时, 照度可能在 30000 至 130000 勒克斯之间, 而到了夜晚, 照度则降至 0.001 至 0.02 勒克斯^[3]。然而, 当前市场上的摄影设备所捕捉到的图像动态范围往往远不及自然环境。这种低动态范围(LDR)的图像常常会出现过度曝光或曝光不足的区域, 导致图像细节的丢失和色彩饱和度的下降^[4]。与此相对, 高动态范围(HDR)技术能够提供更宽广的亮度范围、更丰富的色彩表现以及更完整的细节展示, 因此受到广泛关注。近年来, HDR 成像技术已经在摄影^[5]、视频游戏、电影电视制作^{[6][7]}、医学成像以及一些特定工业领域^[8]中得到广泛应用。

1.2.2 视觉显著性

目前大多数传统显示设备只有有限的动态范围,无法显示 HDR 图像。HDR 图像与显示设备的范围存在巨大差异,因此有必要对 HDR 图像进行压缩,以便在这些普通的 LDR 显示设备上同时再现两个极端的光影区域的外观。另一方面,研究表明,我们的人类视觉系统(HVS)被大量的视觉信息所淹没,它无法完全处理^[9]。视觉显著性反映了图像区域或对象从其周围突出的程度,并提供了优先考虑视觉处理的机制^[10]。视觉显著性被广泛应用于传统 LDR 图像的处理,如图像/视频压缩、视觉搜索、目标识别等^[11]。由于视觉显著性的目的是预测观察者观看场景时的注意力注视,因此对 HDR 图像,特别是 HDR 图像的显示提出了很高的要求。

2 相关工作

2.1 图像动态范围

在传统的 HDR 图像重建方法中, Pual 等人^[12]和 Mertens 等人^[13]提出了一种典型的方法,即利用多曝光图像去重建 HDR 图像。Chen 等人^[14]提出了一种新的数据结构双边网格,将二维的图像数据升维,同时将图像的像素值作为新的维度。Banterle 等人^[15]利用逆色调映射算法进行全局增强,然后通过估计的边缘制导图放大饱和区域。然而,这些方法在重建 HDR 图像时会出现伪影、鬼影等问题。尤其是当图像引入运动场景时,这一问题尤为严重。针对这一问题,Mitsunaga 等人^[16]主要集中在减轻这些伪影上。在深度学习的 HDR 图像重建技术中,研究者们主要构建端到端网络去直接生成 HDR 图像。Kalantari 等人^[17]首次提出了基于学习的动态场景 HDR 图像重建方法,该方法使用卷积神经网络对图像进行对齐和合并。Kim 等人^[18]以及 Sharma 等人^[19]利用深度学习在 HDR 重建任务中针对伪影的问题进行研究。Huang 等人^[20]建立了基于曝光归一化和补偿(ENC)模块的多重曝光校正框架,ENC 模块包括一个将不同的曝光特征映射到曝光不变特征空间的曝光归一化部分,以及整合曝光归一化部分中未处理初始特征的补偿部分,以确保信息的完整性。Hu 等人^[21]提出了一种光适应 HDR 恢复框架,对自适应低曝光、正常曝光和高曝光细节放大和多细节融合,通过这两个部分对 HDR 图像进行预测,在不同光照范围内对细节进行放大的多图像生成和最终融合。Masia 等人^[22]提出了一种基于机器学习的全局伽马模型,在保持理想局部纹理的同时自适应扩展亮度范围。Endo 等人^[23]使用两个神经网络从正常曝光的单个 LDR 图像中推断出过曝光和欠曝光的图像,然后对三张曝光图像进行融合去得到一张 HDR 图像。随后 Lee 等人^{[24][25][26]}定义了两个神经网络,表示具有相对曝光值的图像之间的关系。所提出的结构可以很好地随生成图像的数量而扩展,而不需要重新训练或添加更多的子网络。作者后来改进了

之前的工作, 使用 GAN 结构递归生成多曝光堆栈。Xu 等人^[27]提出了统一的无监督融合网络, 适用于多模态、多曝光、多聚焦多种图像融合。Le 等人^[28]提出了一种弱监督学习方法, 通过学习从单个图像生成多次曝光来反转 HDR 重建的物理图像形成过程。Hu 等人^[29]提出了一种光适应 HDR 恢复框架, 对自适应低光、正常和高光进行细节放大和多细节融合, 通过这两个部分对 HDR 图像进行预测。

这些基于深度学习的 HDR 图像重建方法虽然取得了优异成果, 但仍有局限性。对于基于像素的方法虽然可以描绘整体像素强度, 但在曝光过度或者曝光不足的区域会出现较差的色彩饱和度, 而基于块的方法在细节还原上具有优势, 但在平衡整体对比度方面有限。为了解决这些问题, 本文设计了一个基于双重注意力网络(Dual Attention Network, DANet)的高动态范围图像重建算法。该算法有以下 3 个优点: 1) 提出双重注意力模块(DAM)分别针对过曝光和欠曝光的源图像去提取潜在特征信息并进行融合, 有利于将提取到的特征信息更好地共享与互补。2) 构建特征增强模块(FEM)对融合图像进行颜色校正以及细节增强, 有利于图像显著恢复鲜艳色彩的同时抑制重要信息的丢失。3) 结合对抗损失、感知损失以及全局/局部的对比学习损失构成的损失函数, 进一步提升 HDR 图像的质量, 并增加网络的鲁棒性。

2.2 色调映射

为了在 LDR 显示设备上显示 HDR 图像, 开发了许多全局和局部色调映射算法^[30]。全局色调映射算法简单, 并保留了 HDR 图像的自然性, 因为全局色调映射算法使用单调曲线压缩 HDR 图像^[30]。然而, 全局色调映射算法在保留 HDR 图像的精细细节方面面临一个根本性的困难^[31]。为了解决这个问题, 提出了许多局部色调映射算法, 它们的重点是避免在色调映射图像中出现光环伪影。Reinhard 等^[32]提出了一种空间变化算子^[33]。

邻域的局部性由一个圆形滤波器获得, 该滤波器的大小通过计算局部精细细节的度量来适应每个像素。Farbman 等^[34]采用基于二次优化的保边图像平滑技术, 将 HDR 图像亮度分解为基础层和细节层。基础层由尖锐边缘等大尺度变化组成, 细节层由精细细节等小尺度变化组成。

使用对数域的比例因子缩小基础层的比例, 以减小动态范围。尽管基于全局优化的方法通常会产生出色的质量, 但它们具有很高的计算成本。Durand and Dorsey^[35]采用双边滤波器^[36]将 HDR 图像的亮度分解为基层和细节层, 提出了一种快速局部映射方法。He 等人^[37]引入了一种用于 HDR 图像分解的引导滤波器。然而, 基于局部过滤的方法不能像基于全局优化的方法那样保留尖锐的

边缘^[38]。因此，局部滤波器在平滑某些边缘时，光晕是不可避免的。因此，需要设计一种新的局部滤波器，以更好地保留锐利的边缘，从而避免在最终图像中出现光晕伪影。此外，视觉显著性提供了一种优先处理视觉处理的机制，它被证明对传统 LDR 图像的处理很有用^[39]。

可以预期，视觉显著性与^[40]中的感知局部对比度和颜色饱和度等其他感知质量测量一样，对 HDR 图像有用。但它在现有的局部色调映射算法中没有得到很好的利用。因此，本文希望通过考虑视觉显著性来建立局部色调映射算法。

3 研究方法

3.1 网络整体结构

在整体的网络结构图中，如图 1 所示，网络主要由双重注意力模块(DAM)和特征增强模块(FEM)以及色调映射模块(VsbTM)组成。双重注意力模块用于提取过曝和欠曝图像的特征，之后做初步的融合。如图 1 示例图中，在同一场景中，欠曝图像的阳光部分可以清楚呈现，而沙滩部分细节缺失，相比之下，过曝图像阳光部分细节缺失而沙滩部分细节清楚。因此图像融合可以很好地对过曝光和欠曝光图像进行互补，使图像可以呈现更加完整的细节。它们的融合过程可以表示为：

$$I_f = (I_o \odot A_o(I_o)) \oplus (I_u \odot A_u(I_u)) \quad (1)$$

其中， I_o 、 I_u 分别为输入的过曝光图像和输入的欠曝光图像， A_o 对应过曝光图像的 DAM 模块、 A_u 对应欠曝光图像的 DAM 模块， I_f 表示初始融合图像， \odot 和 \oplus 分别表示元素的点乘和加法。

初步融合图像会出现颜色饱和度低、色彩不均匀等问题。为了改善融合图像颜色的低饱和度，并保持与源图像纹理的一致性，本文提出了一个特征增强模块，用于对图像的细节特征做进一步增强以及色彩方面进行校正。损失函数采用对抗损失、感知损失和对比损失函数对生成图像和参考图进行约束。网络的所有组成部分具体介绍如下：

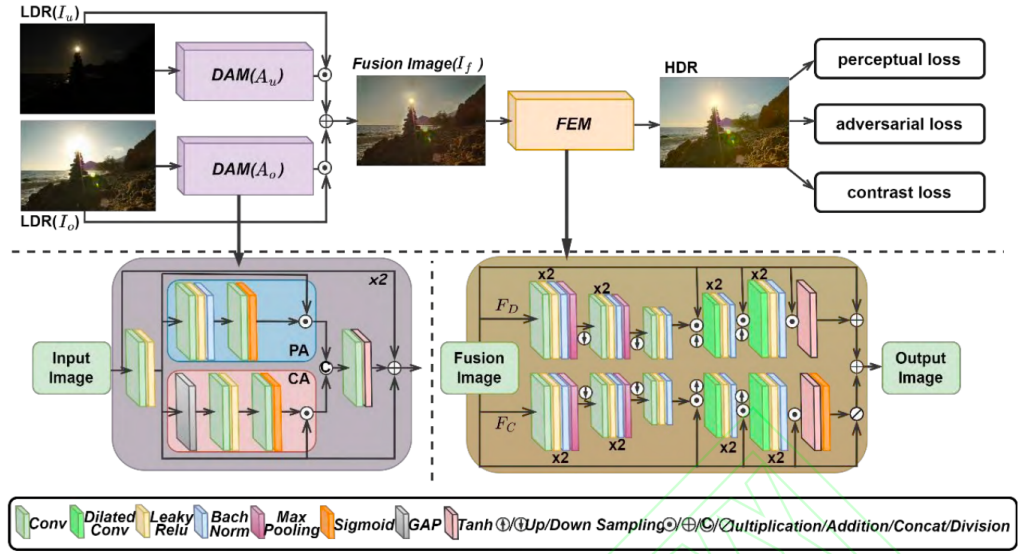


图 1 网络整体结构图

3.2 双重注意力模块(DAM)

注意力机制在图像处理方面得到了广泛应用，其目的是自适应提取图像的显著特征，利用网络的相关部分生成相关权重去指导图像融合过程。例如，多曝光图像融合任务中，通过网络提取源图像的共同曝光区域特征然后生成学习权重进而指导图像融合过程。现有的很多 HDR 成像处理方法中忽略了照度特征和颜色信息的相关性。因此，本文设计了一个双重注意力模块，在该模块中提出了两个注意力机制：像素注意力(Pixel Attention, PA)机制和通道注意力(Channel Attention, CA)机制。其中，像素注意力机制关注颜色信息，通道注意力机制关注照度信息。

1) 像素注意力(PA): 像素注意力机制主要关注像素值，在一定范围内，像素值越高代表输入图像有更高的照度和更亮的色彩。如图 1 下方蓝色框图内容所示，该模块首先使用两个卷积层对源图像的像素值进行提取，提取到的特征形状由 $C \times H \times W$ 变为 $1 \times H \times W$ ，其中， C 为通道数， H 、 W 分别代表图像的高度和宽度。最后将像素特征与 PA 的输入特征点乘。

2) 通道注意力(CA): 通道注意力主要为每个通道分配一个权重值去平衡颜色信息并且调节不同通道之间光照不均匀的问题。如图 1 下方粉色框图内容所示，该模块首先使用全局平均池化(Global Average Pooling, GAP)将输入图像特征进行通道转换，特征形状由 $C \times H \times W$ 变为 $C \times 1 \times 1$ ，然后采用两层卷积层来学习权重值。最后将学习到的权重值与的输入特征相乘。

如图 1 紫色框所示，DAM 模块首先使用卷积层提取过曝光或者欠曝光图像的特征信息，然后利用像素注意力机制和通道注意力机制分别学习相应的特征，将学习到的特征进行拼接使通道数增加，之后经过卷积层对图像再次对特征信息进行提取，从而获得更加精细的图像特征信息。学习到的特征以像素为单位。

通过双重注意力模块可以从两幅源图像中生成一个具有互补信息的图像。然而，尽管注意力机制提取并且整合了不同曝光水平图像的丰富信息，但颜色和曝光等特征信息还需要一定的校正。因此，还需要对初始融合图像做后续的处理。

3.3 特征增强模块(FEM)

在 HDR 图像处理任务中，重建细节丰富性和校正色彩饱和度是一个非常复杂的问题。除此之外，还需要对初始融合图像进行处理使其能够更符合人类视觉感受、更接近参考图像，因此设计一个双流特征增强模块来解决这一问题。该模块包括两个基于 U 型的网络流：细节补充流 FD 和由 Retinex 理论^[42]启发的颜色校正流 FC，整个网络过程可以表示为：

$$I_{\text{hdr}} = (I_f \oplus F_D(I_f)) \oplus (I_f \oslash F_C(I_f)) \quad (2)$$

其中， I_{hdr} 为生成的 HDR 图像， \oslash 表示对图像像素进行除法操作。

Retinex 理论将图像分解得到照度分量和反射分量。我们期望 FC 生成接近真实颜色的照度图，并通过逐元素划分来纠正照度分量的颜色偏差。但是这个过程可能会影响反射分量，干扰图像的纹理细节。因此，需要引入 FD 来弥补细节的损失。如图 1 橙色框所示，该双流网络由不同尺度的卷积层组成，并设置多个残差连接以防止信息丢失，最后将它们各自的输出相加，得到最终的 HDR 图像。

该模块在上采样过程中采用空间金字塔式的膨胀卷积^[43]对图像进行特征提取。膨胀卷积是一种卷积神经网络中的卷积操作，旨在扩展卷积核的感受野，以捕获更广泛的上下文信息。与传统的卷积相比，膨胀卷积有以下两点优势。

1) 扩大感受野。膨胀卷积是通过在卷积核元素之间添加一些间隔，这些间隔称为膨胀率。膨胀率决定了卷积核的感受野大小。例如，膨胀率为 1 时，相当于传统的卷积；膨胀率为 2 时，卷积核每隔一个像素计算一次；膨胀率为 3 时，卷积核每隔两个像素计算一次，依此类推。通过增加膨胀率，膨胀卷积可以有效地扩大卷积核的有效感受野，从而能够捕获更多远处像素的信息，提供更广阔的上下文信息。

2) 减少计算量。传统的卷积操作使用固定大小的卷积核在输入图像上滑动，每次计算一个输出值。然而，对于较大的感受野，需要使用较大的卷积核，这会导致参数数量急剧增加。与传统的增大卷积核不同，膨胀卷积能够实现更大的感受野，同时不增加卷积核的参数数量，因此在一定程度上控制了参数计算和计算机运行内存的开销。

如图 2 所示，膨胀卷积首先采用 1×1 卷积核的标准卷积将高维的特征映射到低维空间，然后采用 3×3 卷积核以及不同膨胀率的卷积对低维的特征映射进行重新采样，最后将使用不同膨胀率采样得到的特征信息进行融合得到最终的特征映射。

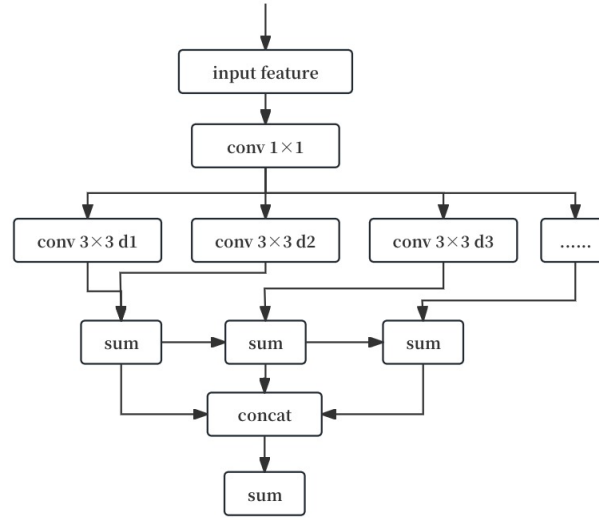


图 2 金字塔式膨胀卷积流程图

膨胀卷积在不改变特征图尺寸的同时能够扩大卷积核的感受野，从而提供更丰富的位置信息和上下文特征，有助于使网络所学习到的特征更加准确和丰富。

3.4 损失函数

在 HDR 图像重建任务中，需要在参考图的引导下去推进任务，以纠正可能出现的颜色偏差和图像失真。为此本文引用了对抗损失、感知损失和全局/局部对比损失。对抗损失基于生成对抗网络(GAN),由卷积和批处理归一化操作组成的判别器实现。感知损失和对比损失基于预训练的 VGG-16 网络实现。

在 HDR 图像处理中，首先使用 MSE 损失函数作为初级像素级约束，表示为：

$$\mathcal{L}_{\text{MSE}} = \| I_{\text{hdr}} - I_{\text{g}} \|^2 \quad (3)$$

其中, I_{hdr} 表示生成的 HDR 图像, I_g 表示标准参考图像。

由于传统的 MSE 损失不足以表达人类视觉系统对图像的直观感受, 因此, 本文引入了模拟人类视觉感受的多曝光图像融合的结构相似度损失^[44], 多曝光图像融合的结构相似度 (MEF_SSIM) 在一般图像结构相似度的基础上提出从局部计算图像的亮度、对比度和结构, 然后赋予不同的权重去计算两张图像结构相似度的值。

$$\mathcal{L}_{\text{SSIM}} = 1 - S_{\text{MEF_SSIM}}(I_{\text{hdr}}, I_g) \quad (4)$$

其中, $S_{\text{MEF_SSIM}}$ 表示生成 HDR 图像和参考图的结构相似度的值。

为了平衡像素强度分布和抑制不良伪影, 引入了对抗损失来约束生成的融合图像与参考图像的一致性。使用 LSGAN^[45] 损失对生成器和鉴别器进行如下约束:

$$\mathcal{L}_D = \mathbb{E}_{X \sim P_{\text{real}}} [(D(I_g) - 1)^2] + \mathbb{E}_{X \sim P_{\text{fack}}} [(D(I_{\text{hdr}}) - 0)^2] \quad (5)$$

$$\mathcal{L}_G = \mathbb{E}_{X \sim P_{\text{fack}}} [(D(I_{\text{hdr}}) - 1)^2] \quad (6)$$

总的对抗损失 \mathcal{L}_{adv} 是以上两部分之和, 表示为:

$$\mathcal{L}_{\text{adv}} = \mathcal{L}_D + \mathcal{L}_G \quad (7)$$

$$\mathcal{L}_{\text{per}} = \sum_{l=1}^L \frac{1}{C_l H_l W_l} \|\varphi_l(I_{\text{hdr}}) - \varphi_l(I_g)\|_1^2 \quad (8)$$

其中, φ_l 表示 VGG-16 的第 l 层, $L = \{1, 3, 5, 9, 13\}$ 。C、H、W 分别表示相对应层特征图的通道数, 高度和宽度。

本文 HDR 图像重建主要是在输入图像是在过曝图像和欠曝图像的基础上进行研究, 对过曝区域和欠曝区域进行细节恢复并抑制非均匀光照, 因此在损失函数中引入了全局和局部约束来解决这一问题。首先, 对抗学习通过约束像素来确定图像是真是假, 与对抗学习不同的是对比学习是从特征层面学习样本数据的特征表示。其次, 与简单的感知损失不同, 对比学习从正、负两方面并行推进, 在逼近正标签分布的同时, 利用负标签的反应来提高模型的泛化程度, 加快网络的收敛速度。对比学习充分利用输入的过曝图像和欠曝图像使生成图像的特征可以有效避免源图像的错误表示。

全局对比约束旨在指导生成图像的整体恢复，我们希望生成图像更接近参考图的特征分布，而远离过曝光和欠曝光的特征分布，定义如下：

$$\Gamma_G = \sum_{l=1}^L \left(\frac{\|\varphi_l(I_{\text{hdr}}) - \varphi_l(I_g)\|_1^2}{\|\varphi_l(I_{\text{hdr}}) - \varphi_l(I_o)\|_1^2} + \frac{\|\varphi_l(I_{\text{hdr}}) - \varphi_l(I_g)\|_1^2}{\|\varphi_l(I_{\text{hdr}}) - \varphi_l(I_u)\|_1^2} \right) \quad (9)$$

其中， I_o 和 I_u 分别表示过曝和欠曝图像， I_{hdr} 和 I_g 分别表示生成的 HDR 图像和参考图像。

然而，在实践中，同一张图片中的曝光水平也会由于空间中的遮挡而改变。例如，在一张户外拍摄的照片中，如果天空区域和地面区域的物体遮挡情况不同，那么它们所接收到的阳光强度也会有所不同，进而导致天空区域比地面区域明亮或者昏暗。这种情况下，图片中的曝光水平就会因为物体遮挡而产生不同的变化。小面积的过度曝光或曝光不足需要更好地平衡，以克服可能出现的色彩偏差。因此，本文分别从过曝图像、欠曝图像、生成图像和参考图像中随机裁剪图像块 P ，从局部解决这一问题，使生成的 HDR 图像块更趋近于参考图像块，并且与源图像的异常曝光区域块存在显著差异，局部对比约束的定义如下：

$$\Gamma_L = \sum_{l=1}^L \sum_{p=1}^P \left(\frac{\|\varphi_l(I_{\text{hdr}_p}) - \varphi_l(I_{g_p})\|_1^2}{\|\varphi_l(I_{\text{hdr}_p}) - \varphi_l(I_{o_p})\|_1^2} + \frac{\|\varphi_l(I_{\text{hdr}_p}) - \varphi_l(I_{g_p})\|_1^2}{\|\varphi_l(I_{\text{hdr}_p}) - \varphi_l(I_{u_p})\|_1^2} \right) \quad (10)$$

其中， I_{hdr_p} 、 I_{g_p} 、 I_{o_p} 、 I_{u_p} 分别表示生成 HDR 图像、参考图像、过曝图像和欠曝图像所对应的图像块。 P 设置为 4。

综上所述全局及局部对比损失定义如下：

$$\Gamma_{\text{cl}} = \Gamma_G + \Gamma_L \quad (11)$$

最后，总的损失定义为：

$$L_{\text{total}} = L_{\text{MSE}} + \alpha L_{\text{SSIM}} + \beta L_{\text{adv}} + \gamma L_{\text{per}} + \delta \Gamma_{\text{cl}} \quad (12)$$

其中，超参数 α 、 β 、 γ 、 δ 都设置为 0.1。

3.5 色调映射

本文的色调映射模块(VsbTM)利用了一种基于显著性感知的 HDR 图像局部色调映射算法。在现有的^[38]的显著性模型中，选择^[46]和^[47]的显著性模型，由于其简单性和鲁棒性，从 LDR 域扩展到 HDR 域。采用扩展显著性模型建立 HDR

图像处理的显著性感知加权。将提出的显著性感知加权和新的边缘感知加权融合在一起，形成内容感知加权，并将其纳入^[37]中的引导图像滤波器中，形成感知引导图像滤波器。然后应用新的滤波器和显著性感知加权来设计 HDR 图像的局部映射算法。本文提出的局部色调映射算法的三个主要组成部分是 HDR 亮度分量分解为基础层和细节层、基础层压缩和细节层放大。将该滤波器应用于 HDR 图像的亮度分量分解。由于所提出的滤波器比^[37]中的引导滤波器更好地保留了基础层的锐利边缘，因此在色调映射图像中显著减少了晕影。采用提出的显著性感知权重，设计了一种显著性感知的全局色调映射算法，用于基层压缩。^[49]中的色调映射结果可以再现真实场景的整体外观，这使得 Reinhard 的方法在图像自然度方面接近全局映射算法。论文^[48]表明 Reinhard 的摄影色调再现^[49]是最好的局部色调映射算法之一。总体而言，色调映射模块主要实现了：1)HDR 图像的简单显著性感知加权；2)内容感知引导图像滤波器；3) ^[49]中扩展显著性感知加权和全局色调映射算法的无缝集成。实验结果表明，所提出的局部色调映射算法可以提高色调映射图像的视觉质量，特别是注意显著区域的视觉质量。

首先扩展^[46]和^[47]中的显著性模型，为 HDR 图像建立显著性感知加权 $[\Gamma_b(p')]$ 。然后，利用 HDR 图像的局部方差和局部均值，提出了 HDR 图像的边缘感知加权 $\Gamma_e(p')$ 。然后计算建议的内容感知权重为

$$W(p') = \frac{\Gamma_e(p')}{\Gamma_b(p')} \quad (13)$$

设 $I_h(p') = [R_h(p'), G_h(p'), B_h(p')]$ 为 HDR 图像中的一个像素。与^[46]和^[47]中的显著性模型类似，我们的 HDR 图像显著性模型基于 HDR 图像的图像共现直方图(ICH)。由于 HDR 图像的每个颜色分量通常由浮点数表示，因此首先对其进行量化，使其由整数表示。为简单起见，将总箱数记为 K ，并以颜色分量 R_h 为例说明过程。量化的颜色分量记为 $Q(R_h(p'))$ 。颜色分量 R_h 在强度域 $H_{i,R}(m,n)$ 的 ICH，表示像素强度的发生/共现，定义为

$$H_{i,R} = [H_{i,R}(m,n)]; \quad 1 \leq m, n \leq K \quad (14)$$

$H_{i,R}$ 是一个大小为 $K \times K$ 的对称方阵。

设 $\Omega_{\rho_1}(p')$ 为以半径为 ρ_1 的像素 p' 为中心的方形窗口。本文选取 ρ_1 的值为 4。

ICH 元素 $H_{i,R}(m,n)$ 是窗口 $\Omega_{\rho_1}(p')$ 内图像值 m 和 n 的共出现次数。然后通过对 ICH

矩阵 $H_{i,R}$ 进行归一化得到概率质量函数(PMF) $P_{i,R}$ 。由于显著性通常与发生/共发生负相关，因此倒置 PMF $\bar{p}_{i,R}(m,n)$ 计算为

$$\begin{cases} 0; & \text{if } P_{i,R}(m,n) = 0 \\ 0; & \text{if } P_{i,R}(m,n) > \frac{1}{U_{i,R}} \\ \frac{1}{U_{i,R}} - P_{i,R}(m,n); & \text{if } P_{i,R}(m,n) \leq \frac{1}{U_{i,R}} \end{cases} \quad (15)$$

其中 $P_{i,R}(m,n)$ 是 $P_{i,R}$ 中的一个元素， $U_{i,R}$ 是 $H_{i,R}$ 中非零项的总数。视觉显著性由倒置 PMF 计算。对于像素 p' ，视觉显著性 $S_{i,R}(p')$ 定义如下

$$S_{i,R}(p') = \sum_{p \in \Omega_{\rho_1}(p')} \bar{p}_{i,R}(Q(R_h(p')), Q(R_h(p))) \quad (16)$$

同样，对于颜色分量 G_h 和颜色分量 B_h ，计算 $S_{i,G}(p')$ 和 $S_{i,B}(p')$ 。将它们归一化为[0,1]，并相加形成强度域中的显著性模型 $S_i(p')$ 。除了在强度域中计算显著性外，还将梯度取向纳入显著性计算。对于每个像素，计算梯度方向并量化为180个区间。从量子化的梯度方向可以构造方向共发生直方图 $H_{o,R}$ 。对应的倒置 PMF $\bar{P}_{o,R}$ 和显著性模型 $S_{o,R}(p')$ 可以由 $H_{o,R}$ 生成。同样地，计算出 $S_{o,G}(p')$ 和 $S_{o,B}(p')$ 。将它们归一化为[0,1]，并相加形成梯度域的显著性模型 $S_o(p')$ 。

$S_i(p')$ 和 $S_o(p')$ 的和记为 $S(p')$ 。最终的视觉显著性定义为 $S(p')$ 的归一化值。

$$S_F(p') = \frac{1}{N} \sum_{p=1}^N \frac{S(p')}{S(p)} \quad (17)$$

其中 N 为图像的总像素数。

提出的显著性感知权重 $\Gamma_b(p')$ 定义为

$$\Gamma_b(p') = \begin{cases} 1; & \text{if } S_F(p') \leq 1 \\ S_F^\kappa(p'); & \text{otherwise} \end{cases} \quad (18)$$

其中 κ (通常在 0.5 和 1 之间)是一个常数，本文选择其值为 0.75。

根据 $\Gamma_b(p')$ 的值分配像素 p' 的处理优先级。具有较大 $\Gamma_b(p')$ 的像素被分配更高的优先级。当像素 p' 位于注意显著区域时， $\Gamma_b(p')$ 的值通常大于 1。因此，通常给予注意突出区域的像素更高的优先级。这与我们的 HVS 特征相匹配，即我们的 HVS 更关注注意突出区域内的信息，而不是注意突出区域外的信息^[38]。

计算红、绿、蓝分量的线性组合得到 HDR 图像的亮度值，方法如下

$$Y_h(p) = 0.299R_h(p) + 0.587G_h(p) + 0.114B_h(p) \quad (19)$$

对数域中的亮度分量记为 $L_h(p)$ 。设 $\sigma_{L_h, \rho_2}^2(p)$ 和 $\mu_{L_h, \rho_2}(p)$ 分别为窗口 $\Omega_{\rho_2}(p)$ 中 L_h 分量的方差和均值，本文选取 ρ_2 的值为 15。考虑两个像素 p' 和 p 。当 $L_h(p)$ 在边缘， $L_h(p')$ 在平坦区域时， $\sigma_{L_h, \rho_2}^2(p)/\mu_{L_h, \rho_2}^2(p)$ 的值通常大于 $\sigma_{L_h, \rho_2}^2(p')/\mu_{L_h, \rho_2}^2(p')$ 的值。基于这一观察，通过使用所有像素的归一化局部方差计算边缘感知权重 $\Gamma_e(p')$

$$\Gamma_e(p') = \frac{1}{N} \sum_{p=1}^N \frac{\left(\frac{\sigma_{L_h, \rho_2}^2(p') + v_1}{\mu_{L_h, \rho_2}^2(p') + v_2} \right)^\zeta}{\left(\frac{\sigma_{L_h, \rho_2}^2(p) + v_1}{\mu_{L_h, \rho_2}^2(p) + v_2} \right)^\zeta} \quad (20)$$

其中 ζ (通常在 0.5 和 1 之间)是一个常数，其值被选择为 0.75。当 $\sigma_{L_h, \rho_2}^2(p)$ 接近于零时，为了避免不稳定， v_1 被包括在内，其值计算为 $(0.001 * L)^2$ ，其中 L 为输入图像的动态范围^[50]。 v_2 是一个很小的常数(通常是 10^{-9})。

如果 p' 位于边缘，则 $\Gamma_e(p')$ 的值通常大于 1，如果 p' 位于光滑区域，则值通常小于 1。显然，通过使用(20)中提出的权重 $\Gamma_e(p')$ ，边缘上的像素比平坦区域中的像素赋予更大的权重。

心理物理学研究表明：1) HVS 在进一步处理之前似乎选择了注意突出区域，从而降低了场景分析的复杂性；2)我们的 HVS 对注意突出区域之外的信息并非视而不见^[38]。因此，在 HDR 图像的动态范围按比例缩小后，允许根据视觉显著性模型发生扭曲是有意义的。给定一张 HDR 图像，将(19)中 HDR 图像的亮度分量分解为

$$Y_h(p) = Y_b(p)Y_d(p) \quad (21)$$

其中 Y_b 和 Y_d 分别为基础层和细节层。细节层 Y_d 的动态范围很小，而基础层 Y_b 的变化可能很大，通常比细节层 Y_d 变化大得多。通过使用全局色调映射算法

将基础层 Y_b 缩小为 \hat{Y}_b 来减小整体动态范围。细节层 Y_d 被保留甚至放大为 \hat{Y}_d ，以增强局部对比度。压缩后的亮度值 Y_l 是 \hat{Y}_b 与 \hat{Y}_d 的乘积，最后采用 HDR 图像 $\{R, G, B\}_h$ 、 Y_h 、 Y_l 生成输出的 LDR 图像 $\{R, G, B\}_l$ 。

从“刚显著差异”实验^[51]中可以看出，对数函数近似于 HVS 视网膜的转换。受此启发，分解在对数域中执行为

$$L_h(p) = L_b(p) + L_d(p) \quad (22)$$

其中 $L_b(p)$ 和 $L_d(p)$ 分别是 $\log(Y_b(p))$ 和 $\log(Y_d(p))$ 。

本文提出的局部色调映射模块的三个主要组成部分是：1)将 L_h 分解为 L_b 和 L_d ；2)通过全局色调映射压缩 L_b ；3) L_d 的放大。

受^[37]中的引导滤波器的启发，假设 L_b 是窗口 $\Omega_{\rho_2}(p')$ 中 L_h 的线性变换。

$$L_b(p) = a_{p'}L_h(p) + b_{p'}, \forall p \in \Omega_{\rho_2}(p') \quad (23)$$

其中 $a_{p'}$ 和 $b_{p'}$ 在窗口 $\Omega_{\rho_2}(p')$ 中应该是恒定的。

在保持线性模型(23)的情况下，通过最小化 L_h 和 L_b 之间的差得到线性系数 $(a_{p'}, b_{p'})$ ，即最小化以下成本函数：

$$\sum_{p \in \Omega_{\rho_2}(p')} \left[W(p') (a_{p'}L_h(p) + b_{p'} - L_h(p))^2 + \lambda a_{p'}^2 \right] \quad (24)$$

将(24)中的代价函数 $E(a_{p'}, b_{p'})$ 与^[37]中的代价函数进行比较，可以发现(24)中的代价函数 $E(a_{p'}, b_{p'})$ 包含一个内容感知权重 $W(p')$ ，该权重根据显著性感知权重负责减少光环伪影并保留精细节。

由线性模型(23)可知： $\nabla L_b(p) = a_{p'} \nabla L_h(p)$ 。显然， $\Omega_{\rho_2}(p')$ 中 L_b 的平滑度取决于 $a_{p'}$ 的值。这意味着(24)中的数据项和正则化项与^[34]中的相似，数据项测量 L_b 相对于输入 L_h 的保真度，正则化项测量 L_b 的平滑度。主要区别在于所提出的滤波器和^[37]中的滤波器是基于局部优化的，而^[34]中的优化问题是全局的。 $a_{p'}$ 和 $b_{p'}$ 的最优值计算为

$$a_{p'} = \frac{W(p')\sigma_{L_h, \rho_2}^2(p')}{W(p')\sigma_{L_h, \rho_2}^2(p') + \lambda} \quad (25)$$

$$b_{p'} = (1 - a_{p'})\mu_{L_h, \rho_2}(p') \quad (26)$$

值得注意的是，在所有重叠的窗口 $\Omega_{\rho_2}(p')$ 中都包含一个像素 p 。 $L_b(p)$ 有许多不同的值。一种简单的方法是将 $L_b(p)$ 的所有可能值取平均值，如下所示：

$$L_b(p) = \bar{a}_p L_h(p) + \bar{b}_p \quad (27)$$

其中 \bar{a}_p 和 \bar{b}_p 被定义为

$$\bar{a}_p = \frac{1}{|\Omega_{\rho_2}(p)|} \sum_{p' \in \Omega_{\rho_2}(p)} a_{p'} \quad (28)$$

$$\bar{b}_p = \frac{1}{|\Omega_{\rho_2}(p)|} \sum_{p' \in \Omega_{\rho_2}(p)} b_{p'} \quad (29)$$

采用(18)中提出的显著性感知加权，设计一种显著性感知的全局色调映射算法，用于压缩基层 L_b 。基层 L_b 的压缩对生成的 LDR 图像的视觉质量非常重要。由于边缘保留在基础层 L_b 中，因此在压缩基础层 L_b 时，只需要减小边缘的大小，同时保持边缘的方向不变。否则，色调映射的图像会出现晕晕伪影。如果使用全局映射算法压缩基础层 L_b ，则可以实现此目标。通过将(18)中的显著性感知权重纳入^[49]中的全局色调映射算法，提出了一种简单的显著性感知全局色调映射算法模块(VsbTM)。

L_h 的平均值(记为 \bar{L}_h)在^[32]和^[49]的色调映射算法中起着重要作用。值得注意的是，当在^[32]和^[49]中计算 \bar{L}_h 值时，所有像素都被赋予了相同的权重。现在采用(18)中提出的显著性感知权重来计算 \bar{L}_h

$$\bar{L}_h = \frac{\sum_{p=1}^N \Gamma_b(p) L_h(p)}{\sum_{p=1}^N \Gamma_b(p)} \quad (30)$$

设 a 为整个 HDR 图像的整体键值，它决定了 HDR 图像的整体亮度^[32]。与^[49]相同， a 定义为

$$a = 0.18 * 4 \frac{2\bar{L}_h - L_h^{min} - L_h^{max}}{L_h^{max} - L_h^{min}} \quad (31)$$

其中 L_h^{max} 和 L_h^{min} 分别是 HDR 图像中所有 $L_h(p)(\neq 0)$ 的最大值和最小值。为了提高稳定性，我们排除了一些最亮和最暗的像素。

然后将基础层 L_b 压缩为

$$\hat{L}_b(p) = \log(a) + L_b(p) - \bar{L}_h - \log(1 + a \exp(L_b(p) - \bar{L}_h)) \quad (32)$$

显然，色调映射图像的整体亮度在很大程度上取决于注意突出区域。在最终图像中，可以更好地保留注意显著区域的动态范围。因此，与不考虑显著性感知权重的色调映射算法相比，区域的视觉质量也得到了改善。

放大细节层 L_d 也很重要，这样最终的 LDR 图像看起来更清晰。色调映射模块提出了一种简单的方法来实现这一点：

$$\hat{L}_d(p) = \theta L_d(p) \quad (33)$$

其中 θ (通常在 1 和 2 之间)是一个常数，本文选择其值为 1.5。

为了显示色调映射后的图像，将压缩后的 \hat{L}_b 和处理后的 \hat{L}_d 分别转换回强度域，表示为 $\hat{Y}_b(p)$ 和 $\hat{Y}_d(p)$ 。色调映射的亮度分量 Y_l 是压缩的基层 \hat{Y}_b 和放大的细节层 \hat{Y}_d 的乘积。然后 HDR 图像的颜色通道按如下方式缩放

$$\{R, G, B\}_l(p) = \left(\frac{Y_l(p)}{Y_h(p)} (R, G, B)_h(p) \right)^s \quad (34)$$

其中 $\{R, G, B\}_l$ 表示输出 LDR 图像的颜色通道。指数 s (一般在 0.375 ~ 0.625 之间)是校正显示器件衰减效应的饱和度参数，其值为 0.5。

4 实验分析与结果

4.1 实验设置

本实验是基于 Ubuntu20.04, CUDA11.3, Python3.8 搭建的深度学习环境，采用 Pytorch 深度学习框架，在搭载了 NVIDIA GeForce RTX2070 GPU 的计算机上进行了训练。实验在训练过程中使用 Adam 优化器对网络和鉴别器参数进行优化，学习率为 0.0002，每次实验进行 200 轮训练。

4.2 实验数据

本次实验采用公开的 SICE 数据集。在数据集中，其中一部分的图像序列中仅包含 3 张曝光图像或者只有一系列的欠曝图像，不符合实验要求。因此，本

实验首先手动去除了个别无效数据，并收集了 589 个图像序列，每个序列包括一系列曝光图像，即过度曝光图像、正常曝光图像和曝光不足图像，其中每个序列都对应有一张高质量的 HDR 参考图像。然后分别对每个序列抽取第 1 张、第 2 张图像作为欠曝图像，抽取第 6 张、第 7 张图像作为过曝图像，以此来扩充训练数据集。本次实验共收集图像 3573 张，其中，从 SICE 数据集随机抽取 318 张图像以及实拍图像 39 张作为测试数据集，其余 3216 张图像作为训练数据集。批量大小设置为 4，用于训练的图像块和用于对比学习的图像块分别被随机裁剪为 320×320 和 32×32 。

4.3 评估指标

本次实验共采用 3 个指标对生成图像进行评价，包括基于信息理论的峰值信噪比(Peak Signal to Noise Ratio, PSNR)，基于图像结构相似度的度量值(Structural Similarity Index Measure, SSIM)以及学习感知图像块相似度(Learned Perceptual Image Patch Similarity, LPIPS)。

(1) 峰值信噪比(PSNR) ^[52]PSNR 是评价图像质量最常见、应用最广泛的客观指标，其定义为：

$$\text{PSNR} = 10\log_{10} \frac{r^2}{\text{MSE}(I_g, I_{\text{hdr}})} \quad (35)$$

其中， r 为 I_{hdr} 的像素峰值，该值为 255。MSE 的定义为：

$$\text{MSE}(I_g, I_{\text{hdr}}) = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W (I_g(i, j) - I_{\text{hdr}}(i, j))^2 \quad (36)$$

PSNR 值越大表明生成的 HDR 图像更接近参考图像。然而，该指标没有考虑到人眼的视觉特性，导致有时与人的主观判断不一致。因此，PSNR 指标的使用需要其他指标的辅助。

(2) 结构相似度(SSIM) ^[53]SSIM 模拟了人类视觉系统对敏感图像的破坏和失真。分为亮度、对比度和结构三部分进行估计，定义分别如下：

$$\begin{aligned} l(I_{\text{hdr}}, I_g) &= \frac{2\mu_{\text{hdr}}\mu_g + C_1}{\mu_{\text{hdr}}^2 + \mu_g^2 + C_1} \\ c(I_{\text{hdr}}, I_g) &= \frac{2\sigma_{\text{hdr}}\sigma_g + C_2}{\sigma_{\text{hdr}}^2 + \sigma_g^2 + C_2} \\ s(I_{\text{hdr}}, I_g) &= \frac{\sigma_{\text{hdr}g} + C_3}{\sigma_{\text{hdr}}\sigma_g + C_3} \end{aligned} \quad (37)$$

其中, l, c, s 分别表示亮度、对比度和结构, μ 、 σ 分别表示均值和方差, σ_{hdr} 表示 I_{hdr} 和 I_g 的协方差, C_1 、 C_2 、 C_3 指常数。

(3) 学习感知图像块相似度(LPIPS)^[54]

LPIPS 学习生成图像到参考图像的反向映射强制生成器学习从生成图像中重构参考图像的反向映射, 并优先处理它们之间的感知相似度, 更符合人类的感知情况:

$$\text{LPIPS}(I_{\text{hdr}}, I_g) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} \| \mathcal{U}_l \odot (I_{\text{hdr}hw}^l - I_{ghw}^l) \|_2^2 \quad (38)$$

其中, l 是指图像通道, \mathcal{U}_l 指缩放因子, H_l, W_l 分别表示图像的高和宽。

4.4 实验结果

为了证明本文所提方法的有效性, 该算法进行了一系列对比实验, 将本文的方法与 Lee 等人(DRHDR)^[39], Marnierides 等人(ExpandNet)^[55], Xu 等人(U2Fusion)^[46], Le 等人^[47]进行了比较。为了公平的比较, 所有比较方法都使用本文方法采用的数据集重新进行了训练, 并且在重新训练时保持了各自原有的损失和训练方法。

(1) 定性比较

本次实验采用 SICE 数据集将本文方法与之前的代表性方法进行比较, 分别针对室内室外的不同场景进行测试, 如图 3 所示。其中, image1 到 image4 是来自 SICE 数据集, image5 和 image6 是来自于实拍图。在 DRHDR 方法中提高了图像的整体亮度, 但是色彩与参考图存在差异, 整体图像发灰, 且存在光晕与噪声, 如图 image3 所示; ExpandNet 方法在本次实验中效果比较差, 图像色差比较严重, 如图 image2 中的窗口部分, image5 的天空细节缺失, 且存在光晕, 噪声等问题。在 U2Fusion 方法中, 图像亮度整体有很大提升, 色彩饱和度提高, 如图 image1、image3、image5 所示, 但是同样存在噪声, 而且在细节上处理比较模糊; Le 等人的方法的效果比较好, 图像细节与纹理方面处理的也较为出色, 但在色彩处理上还是存在色偏现象, 图像整体偏暗, 如图 image6 的墙体部分。与以上几种方法对比, 本文方法获得了更接近参考图的效果, 在色彩还原以及细节处理上本文方法都达到了最优效果, 不存在颜色失真, 光晕等现象。

(2) 定量比较

除了定性比较之外，在对比实验中采用了 3 个指标来进行定量分析，一是峰值信噪比(PSNR)，该指标常用来评估数字图像质量，PSNR 的值越高，表明生成的 HDR 图像与参考图像在像素值上越逼近，图像的质量越好。二是结构相似性指数(SSIM)，该指标用来测量两张图像之间的结构相似性，SSIM 的值越大表明这两张图像的结构相似性越大。三是学习感知图像块相似度(LPIPS)，该指标用来衡量图像视觉感知相似度，LPIPS 越小则表明图像的感知相似度越大。这 3 个指标从不同角度来评估生成的 HDR 图像，以反映生成图像的质量。表 1 展示了实验定量评价的结果。

表 1 实验定量评价结果

Tab. 1 Quantitative evaluation results of the experiment

Method	PSNR(↑)	SSIM(↑)	LPIPS(↓)
DRHDR	17.052	0.785	0.164
ExpandNet	16.239	0.594	0.200
U2Fusion	17.296	0.743	0.159
Le et al	18.377	0.853	0.143
Ours	18.496	0.882	0.149

Method	PSNR(↑)	SSIM(↑)	LPIPS(↓)
DRHDR	17.052	0.785	0.164
ExpandNet	16.239	0.594	0.200
U2Fusion	17.296	0.743	0.159
Le et al	18.377	0.853	0.143
Papers	18.496	0.882	0.149
Ours	19.121	0.879	0.150

其中，红色表示最优结果，蓝色表示次优结果。结果显示，与其他方法相比，本文方法在 PSNR、SSIM、LPIPS 这三个评价指标中都得到了最优的结果。定量结果表明，本文的 HDR 图像重建方法不仅纹理精确度高，而且色彩饱和度佳。

4.5 消融实验

为了验证本文所提出网络模型的有效性，对该网络模型进行了消融实验。

本次实验首先去掉了双重注意力模块(DAM)，所得实验结果如表 2 所示，从表中可以看出去掉双重注意力模块后，PSNR、SSIM 的值分别下降了 0.985、0.114；LPIPS 的值上升了 0.108。从主观图来看，如图 4 所示，去掉双重注意力模块图像出现细节丢失，噪声等现象。

其次，去掉特征增强模块(FEM)，从表 2 中可以看出 PSNR、SSIM 的数值分别下降了 1.459、0.292，LPIPS 的数值上升了 0.056，从图 4 中可以看出去掉特征增强模块后，图像出现色偏、伪影等问题。最后，去掉膨胀卷积后，定量评价结果如表 2 所示，PSNR、SSIM 的值分别下降了 1.2、0.142；LPIPS 的值上升了 0.052，实验效果对比图如图 4 所示，图像整体亮度偏暗且存在噪声。

表 2 消融实验定量评价结果

Tab. 2 Quantitative evaluation results of ablation experiments

Method	PSNR(↑)	SSIM(↑)	LPIPS(↓)
Without DAM	17.511	0.768	0.257
Without FEM	17.037	0.759	0.205
Without Dilated Conv	17.296	0.740	0.201
Ours	18.496	0.882	0.149



图 4 消融实验效果对比图

Fig. 4 Comparison chart of ablation experiment effect

通过消融实验证明了本文所提出的基于双重注意力网络模型生成的图像质量更高，色彩饱和度也更好，更接近参考图。

4.6 色调映射

色调映射部分所有结果都是从视觉质量的角度进行比较，主要是进行定性比较。

首先通过设置 $\Gamma_b(p)$ 的值为 1 并测试三组 HDR 图像来测试(4)中提出的边缘感知加权的效率。对于^[37]中的滤波器和所提出的滤波器，分别取 ρ 和 λ 的值为 15 和 1。由^[34]可知， $\gamma, \epsilon, \lambda$ 分别取^[34]中全局优化方法的 1.2、0.001 和 2。使用^[32]和

[49]中算法的默认设置来生成色调映射图像。如图 1 所示，所提出的滤波器比[37]中的引导滤波器更好地保留了基础层中的边缘。

因此，通过使用[37]中的引导滤波器，所提出的色调映射算法可以避免在色调映射图像中出现光晕伪影，而在色调映射的图像中存在光晕伪影。所提滤波器处理的色调映射图像与[34]的全局优化方法处理的色调映射图像相当，而所提滤波器的复杂度远低于[34]的全局优化方法。尽管[32]和[49]中的局部色调映射算法可以避免在色调映射图像中出现光晕伪影，但它不能像其他三种局部色调映射算法那样保留细节。

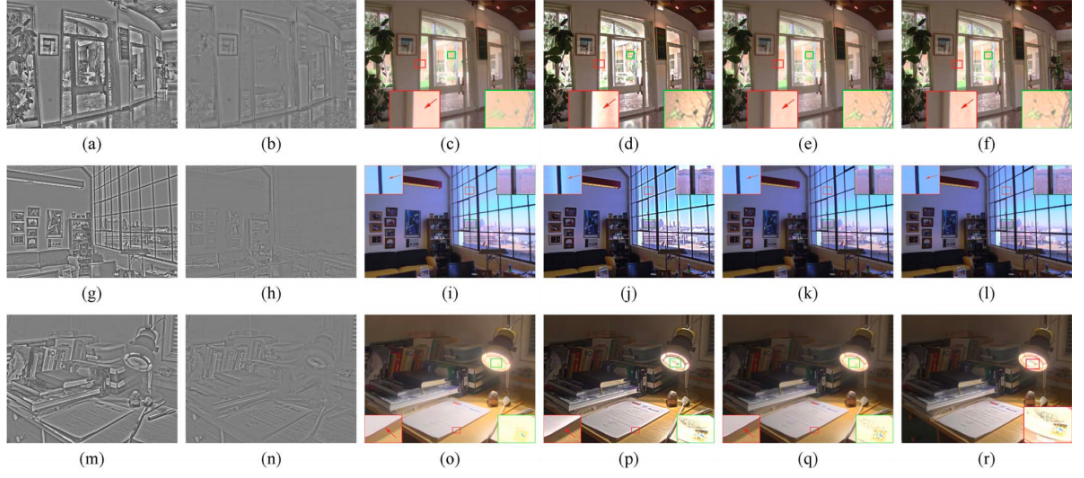
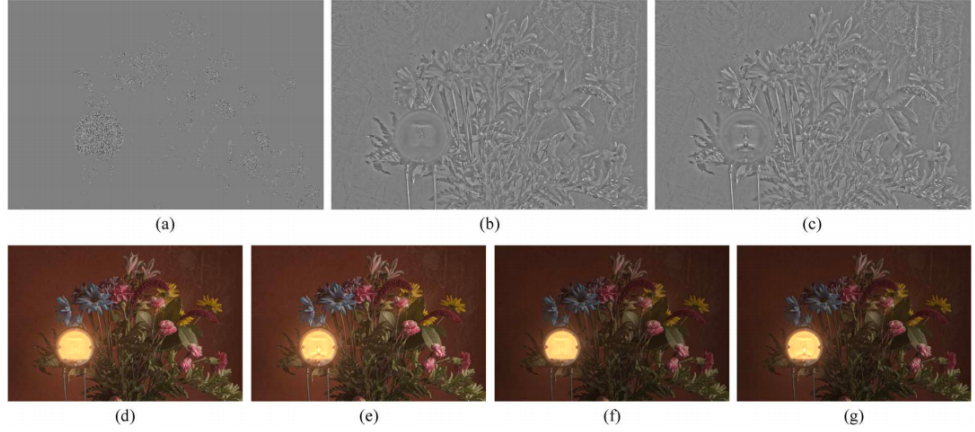


Fig. 1. Comparison on tone-mapped images. (a), (g), and (m) Extracted fine details by the filter in [18]. (b), (h), and (n) Extracted fine details by the proposed filter. (c), (i), and (o) Tone-mapped images by the algorithm in [13] and [25]. (d), (j), and (p) Tone-mapped images by the filter in [18]. (e), (k), and (q) Tone-mapped images by using the L_2 -norm-based optimization method in [15]. (f), (l), and (r) Tone-mapped images by the proposed filter.



5 总结讨论

5.1 成果应用分析

本文提出了一种基于双重注意力网络模型的 HDR 图像重建方法。利用双重注意力模块进行图像初步融合，之后经过特征增强模块对初步融合图像进行颜

色校正和细节增强，最终生成更接近人眼感知的 HDR 图像。与此同时，本文在公开的数据集上以及实拍图像与其他方法进行了比较，结果表明，本文的方法产生的 HDR 图像不仅具有更好的视觉效果，而且具有最接近参考图的像素分布。

在色调映射的相关工作中，本文提出了一种新的 HDR 图像显著性感知加权和边缘感知加权方法。应用它们设计了一种新的局部色调映射算法，用于在有限动态范围的显示设备上显示 HDR 图像。实验结果表明，该方法可以避免色晕伪影出现在色调映射图像中，提高了色调映射图像的视觉质量，特别是注意显著区域。

5.2 不足与思考

6 个人贡献

陈子豪：

胡楷：

阮明航：

参考文献

- [1] 王仙峰, et al. "基于双重注意力网络的高动态范围图像重建." 激光与光电子学进展 1-15.
- [2] Li, Zhengguo, and Jinghong Zheng. "Visual-saliency-based tone mapping for high dynamic range images." *IEEE Transactions on Industrial Electronics* 61.12 (2014): 7076-7082.
- [3] Huang, Li, et al. "Multi-exposure image fusion based on feature evaluation with adaptive factor." *IET Image Processing* 15.13 (2021): 3211-3220.
- [4] Liu, Jinyuan, et al. "HoLoCo: Holistic and local contrastive learning network for multi-exposure image fusion." *Information Fusion* 95 (2023): 237-249.
- [5] 林森, and 刘世本. "基于多尺度生成对抗网络的水下图像增强." 激光与光电子学进展 58.16(2021):306-315.
- [6] 辛元雪, et al. "基于改进增强型超分辨率生成对抗网络的图像超分辨率重建算法." 激光与光电子学进展 59.04(2022):381-391.

- [7] 汤璐, et al."基于 YCbCr 空间融合的高动态范围成像方法."激光与光电子学进展 59.14(2022):308-319.
- [8] 秦艳, et al."基于高动态范围成像的极端亮度测量方法研究."应用光学 44.01(2023):168-174.
- [9] Tsotsos, John K. "Analyzing vision at the complexity level." *Behavioral and brain sciences* 13.3 (1990): 423-445.
- [10] Borji, Ali, Dicky N. Sihite, and Laurent Itti. "Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study." *IEEE Transactions on Image Processing* 22.1 (2012): 55-69.
- [11] Grgic, Sonja, Mislav Grgic, and Branka Zovko-Cihlar. "Performance analysis of image compression using wavelets." *IEEE Transactions on industrial electronics* 48.3 (2001): 682-695.
- [12] Debevec, Paul E., and Jitendra Malik. "Recovering high dynamic range radiance maps from photographs." *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*. 2023. 643-652.
- [13] Mertens, Tom, Jan Kautz, and Frank Van Reeth. "Exposure fusion: A simple and practical alternative to high dynamic range photography." *Computer graphics forum*. Vol. 28. No. 1. Oxford, UK: Blackwell Publishing Ltd, 2009.
- [14] Chen, Jiawen, Sylvain Paris, and Frédo Durand. "Real-time edge-aware image processing with the bilateral grid." *ACM Transactions on Graphics (TOG)* 26.3 (2007): 103-es.
- [15] Banterle, Francesco, et al. "A framework for inverse tone mapping." *The Visual Computer* 23 (2007): 467-478.
- [16] Mitsunaga, Tomoo, and Shree K. Nayar. "Radiometric self calibration." *Proceedings. 1999 IEEE computer society conference on computer vision and pattern recognition (Cat. No PR00149)*. Vol. 1. IEEE, 1999.
- [17] Kalantari, Nima Khademi, and Ravi Ramamoorthi. "Deep high dynamic range imaging of dynamic scenes." *ACM Trans. Graph.* 36.4 (2017): 144-1.
- [18] Kim, Dae-Eun, and Munchurl Kim. "Learning-based low-complexity reverse tone mapping with linear mapping." *IEEE Transactions on Circuits and Systems for Video Technology* 30.2 (2019): 400-414.
- [19] Sharma, Aashish, Robby T. Tan, and Loong-Fah Cheong. "Estimation of camera response function using prediction consistency and gradual refinement with an extension to deep learning." *arXiv preprint arXiv:2010.04009* (2020).
- [20] Huang, Jie, et al. "Exposure normalization and compensation for multiple-exposure correction." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.
- [21] Hu, Xiangyu, et al. "LA-HDR: Light Adaptive HDR Reconstruction Framework for Single LDR Image Considering Varied Light Conditions." *IEEE Transactions on Multimedia* (2022).
- [22] Masia, Belen, Ana Serrano, and Diego Gutierrez. "Dynamic range expansion based on image statistics." *Multimedia Tools and Applications* 76 (2017): 631-648.
- [23] Endo, Yuki, Yoshihiro Kanamori, and Jun Mitani. "Deep reverse tone mapping." *ACM Trans. Graph.* 36.6 (2017): 177-1.
- [24] Lee, Siyeong, Gwon Hwan An, and Suk-Ju Kang. "Deep chain hdri: Reconstructing a high dynamic range image from a single low dynamic range image." *IEEE Access* 6 (2018): 49913-49924.

- [25] Lee, Siyeong, Gwon Hwan An, and Suk-Ju Kang. "Deep recursive hdri: Inverse tone mapping using generative adversarial networks." *proceedings of the European Conference on Computer Vision (ECCV)*. 2018.
- [26] Lee, Siyeong, et al. "Learning to generate multi-exposure stacks with cycle consistency for high dynamic range imaging." *IEEE Transactions on Multimedia* 23 (2020): 2561-2574.
- [27] Xu, Han, et al. "U2Fusion: A unified unsupervised image fusion network." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.1 (2020): 502-518.
- [28] Le, Phuoc-Hieu, et al. "Single-image hdr reconstruction by multi-exposure generation." *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. 2023.
- [29] Hu, Xiangyu, et al. "LA-HDR: Light Adaptive HDR Reconstruction Framework for Single LDR Image Considering Varied Light Conditions." *IEEE Transactions on Multimedia* (2022).
- [30] Debevec, Paul, et al. "High dynamic range imaging." *ACM SIGGRAPH 2004 Course Notes*. 2004. 14-es.
- [31] Fattal, Raanan, Dani Lischinski, and Michael Werman. "Gradient domain high dynamic range compression." *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*. 2023. 671-678.
- [32] Reinhard, Erik, et al. "Photographic tone reproduction for digital images." *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*. 2023. 661-670.
- [33] A. Adams, *The Print*. ser. The Ansel Adams Photography series. New York, NY, USA: Little, Brown and Company, 1983.
- [34] Farbman, Zeev, et al. "Edge-preserving decompositions for multi-scale tone and detail manipulation." *ACM transactions on graphics (TOG)* 27.3 (2008): 1-10.
- [35] Durand, Frédo, and Julie Dorsey. "Fast bilateral filtering for the display of high-dynamic-range images." *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*. 2002.
- [36] Tomasi, Carlo, and Roberto Manduchi. "Bilateral filtering for gray and color images." *Sixth international conference on computer vision (IEEE Cat. No. 98CH36271)*. IEEE, 1998.
- [37] He, Kaiming, Jian Sun, and Xiaoou Tang. "Guided image filtering." *IEEE transactions on pattern analysis and machine intelligence* 35.6 (2012): 1397-1409.
- [38] Itti, Laurent, Christof Koch, and Ernst Niebur. "A model of saliency-based visual attention for rapid scene analysis." *IEEE Transactions on pattern analysis and machine intelligence* 20.11 (1998): 1254-1259.
- [39] Judd, Tilke, et al. "Learning to predict where humans look." *2009 IEEE 12th international conference on computer vision*. IEEE, 2009.
- [40] Shen, Rui, Irene Cheng, and Anup Basu. "QoE-based multi-exposure fusion in hierarchical multivariate Gaussian CRF." *IEEE Transactions on Image Processing* 22.6 (2012): 2469-2478.
- [41] Liu, Shibin, et al. "Adaptive learning attention network for underwater image enhancement." *IEEE Robotics and Automation Letters* 7.2 (2022): 5326-5333.
- [42] Land, Edwin H. "The retinex theory of color vision." *Scientific american* 237.6 (1977): 108-129.
- [43] Mehta, Sachin, et al. "Espnet: Efficient spatial pyramid of dilated convolutions for semantic segmentation." *Proceedings of the european conference on computer vision (ECCV)*. 2018.

- [44] Ma, Kede, Kai Zeng, and Zhou Wang. "Perceptual quality assessment for multi-exposure image fusion." *IEEE Transactions on Image Processing* 24.11 (2015): 3345-3356.
- [45] Mao, Xudong, et al. "Least squares generative adversarial networks." *Proceedings of the IEEE international conference on computer vision*. 2017.
- [46] Lu, Shijian, and Joo-Hwee Lim. "Saliency modeling from image histograms." *Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part VII* 12. Springer Berlin Heidelberg, 2012.
- [47] Lu, Shijian, Cheston Tan, and Joo-Hwee Lim. "Robust and efficient saliency modeling from image co-occurrence histograms." *IEEE transactions on pattern analysis and machine intelligence* 36.1 (2013): 195-201.
- [48] Kuang, Jiangtao, et al. "Evaluating HDR rendering algorithms." *ACM Transactions on Applied Perception (TAP)* 4.2 (2007): 9-es.
- [49] Reinhard, Erik. "Parameter estimation for photographic tone reproduction." *Journal of graphics tools* 7.1 (2002): 45-51.
- [50] Wang, Zhou, et al. "Image quality assessment: from error visibility to structural similarity." *IEEE transactions on image processing* 13.4 (2004): 600-612.
- [51] Hurvich, Leo M., and Dorothea Jameson. *The perception of brightness and darkness*. 1966.
- [52] Huynh-Thu, Quan, and Mohammed Ghanbari. "Scope of validity of PSNR in image/video quality assessment." *Electronics letters* 44.13 (2008): 800-801.
- [53] Wang, Zhou, Eero P. Simoncelli, and Alan C. Bovik. "Multiscale structural similarity for image quality assessment." *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*. Vol. 2. Ieee, 2003.
- [54] Zhang, Richard, et al. "The unreasonable effectiveness of deep features as a perceptual metric." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.
- [55] Marnierides, Demetris, et al. "Expandnet: A deep convolutional neural network for high dynamic range expansion from low dynamic range content." *Computer Graphics Forum*. Vol. 37. No. 2. 2018.