



激光与光电子学进展  
*Laser & Optoelectronics Progress*  
ISSN 1006-4125, CN 31-1690/TN

## 《激光与光电子学进展》网络首发论文

题目: 基于双重注意力网络的高动态范围图像重建  
作者: 王仙峰, 刘世本, 田建东, 赵娟平, 刘雅静, 郝春晖  
收稿日期: 2023-07-21  
网络首发日期: 2023-09-22  
引用格式: 王仙峰, 刘世本, 田建东, 赵娟平, 刘雅静, 郝春晖. 基于双重注意力网络的高动态范围图像重建[J/OL]. 激光与光电子学进展.  
<https://link.cnki.net/urlid/31.1690.tn.20230920.1827.080>



**网络首发:** 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

**出版确认:** 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

# 基于双重注意力网络的高动态范围图像重建

王仙峰<sup>1,2,3</sup>, 刘世本<sup>2,3</sup>, 田建东<sup>2,3\*</sup>, 赵娟平<sup>1</sup>, 刘雅静<sup>2,3</sup>, 郝春晖<sup>1,2,3</sup>

<sup>1</sup>沈阳化工大学 信息工程学院, 辽宁 沈阳 110142;

<sup>2</sup>中国科学院沈阳自动化研究所 机器人学国家重点实验室, 辽宁 沈阳 110016

<sup>3</sup>中国科学院机器人与智能制造创新研究院, 辽宁 沈阳 110169

**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。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, 且生成的 HDR 图像色彩饱和度好且细节信息精准完整。

**关键词** 图像重建; 高动态范围成像; 图像融合; 双重注意力机制

**中图分类号** TP391 **文献标志码** A

## High Dynamic Range Image Reconstruction Based on Dual Attention Network

Wang Xianfeng<sup>1,2,3</sup>, Liu Shibei<sup>2,3</sup>, Tian Jiandong<sup>2,3\*</sup>, Zhao Juanping<sup>1</sup>, Liu Yajing<sup>2,3</sup>,  
Hao Chunhui<sup>1,2,3</sup>

<sup>1</sup> College of Information Engineering, Shenyang University of Chemical Technology,  
Shenyang 110142, Liaoning, China;

<sup>2</sup> State Key Laboratory of Robotics, Shenyang Institute of Automation Chinese Academy  
of Sciences, Shenyang 110016, Liaoning, China;

<sup>3</sup> Institute of Robotics and Intelligent Manufacturing Innovation, Chinese Academy of  
Sciences, Shenyang 110169, Liaoning, China

**Abstract** In the high dynamic range (HDR) image reconstruction task, the existing HDR image reconstruction methods based on deep learning are prone to the loss of detail information and poor color saturation as the input image is overexposed or underexposed. To solve this problem, We proposes a dual attention network based HDR image reconstruction method. Firstly, this method utilizes the Dual Attention Module (DAM) to apply attention mechanism from pixel and channel dimensions respectively to extract and fuse the features of the source image, and then obtain a preliminary fusion image. On this basis, a feature enhancement module is constructed to perform detail enhancement and color correction respectively. The final reference to contrastive learning is

to generate images that are closer to the reference image and away from the source image. After multiple trainings, the HDR image is finally generated. The experimental results show that the method in this paper achieves the best evaluation results on PSNR, SSIM and LPIPS, and the generated HDR image has good color saturation and accurate details.

**Key words** image reconstruction; high dynamic range image; image fusion; dual attention mechanism

**OCIS codes** 100.3020;100.3010;100.2000

## 1 引言

照度是影响图像动态范围的重要因素。通常,自然场景中的照度变化很大,在阳光充足的白天照度范围可达 30000~130000lux,而夜晚的照度范围仅有 0.001~0.02lux<sup>[1]</sup>。然而,目前市场上的大部分摄影设备获取到图像的动态范围是远低于自然场景的。而这种低动态范围(Low Dynamic Range, LDR)图像会不可避免地会出现过曝或者欠曝的区域,造成图像细节信息丢失,色彩饱和度差等问题<sup>[2]</sup>。相对地,高动态范围(HDR)技术可以获得更广的亮度范围、更丰富的色彩显示以及更完整的细节呈现等优势而备受关注。因此,近年来, HDR 成像方法被广泛应用于相机拍摄<sup>[3]</sup>, 视频游戏, 影视行业<sup>[4][5]</sup>, 医学成像, 以及一些特殊的工业场景<sup>[6]</sup>等领域。

在传统的 HDR 图像重建方法中, Pual 等人<sup>[7]</sup>和 Mertens 等人<sup>[8]</sup>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<sup>[9]</sup>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<sup>[10]</sup>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<sup>[11]</sup>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<sup>[12]</sup>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<sup>[13]</sup>以及 Sharma 等人<sup>[14]</sup>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<sup>[15]</sup>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<sup>[16]</sup>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<sup>[17]</sup>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<sup>[18]</sup>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<sup>[19][20][21]</sup>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<sup>[22]</sup>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<sup>[23]</sup>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<sup>[24]</sup>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. 提出方法

### 2.1 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)两部分组成。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和，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：



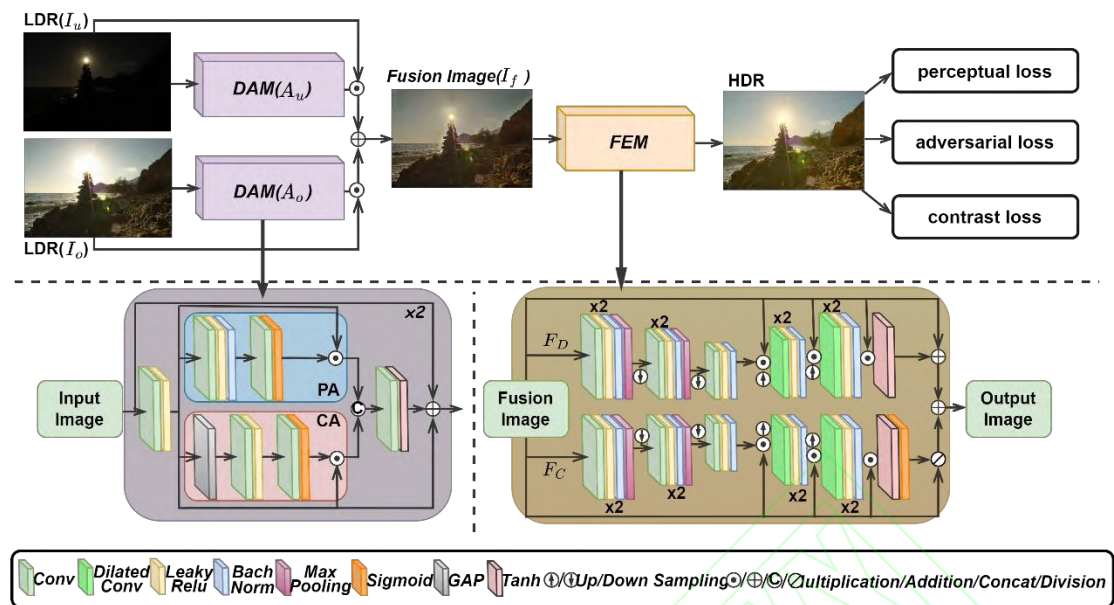


图 1 网络整体结构图

Fig. 1 Overall Network Structure Diagram

## 2.2 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<sup>[25]</sup>。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

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。

### 2.3 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  $F_D$  和由 Retinex 理论<sup>[26]</sup>启发的颜色校正流  $F_C$ ,整个网络过程可以表示为:

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

其中,  $I_{\text{hdr}}$  为生成的 HDR 图像,  $\odot$  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  $F_C$  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  $F_D$  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<sup>[27]</sup>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  $1 \times 1$  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  $3 \times 3$  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。

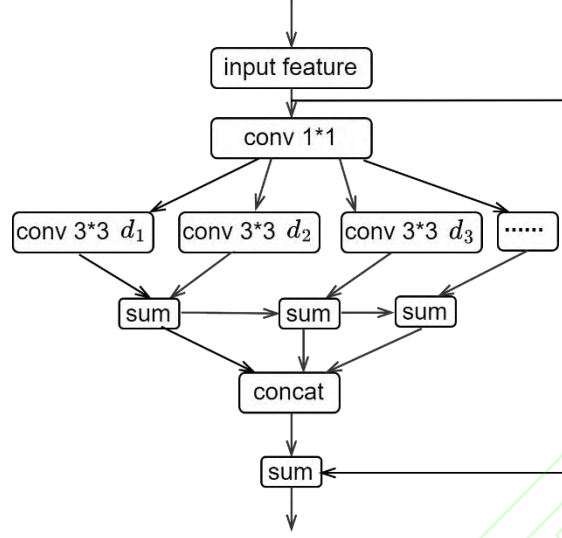


图 2 金字塔式膨胀卷积流程图

Fig 2 Pyramidal dilation convolution flowchart

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。

## 2.4 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,  $I_{\text{hdr}}$  表示生成的 HDR 图像,  $I_{\text{g}}$  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<sup>[28]</sup>,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。一般图像结构相似度的具体定义在下文 3.2 节。

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

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<sup>[29]</sup>损失对生成器和鉴别器进行如下约束

$$\mathcal{L}_D = E_{X \sim P_{\text{real}}} \left[ \left( D(I_g) - 1 \right)^2 \right] + E_{X \sim P_{\text{fick}}} \left[ \left( D(I_{\text{hdr}}) - 0 \right)^2 \right] \quad (5)$$

$$\mathcal{L}_G = E_{X \sim P_{\text{fick}}} \left[ \left( D(I_{\text{hdr}}) - 1 \right)^2 \right] \quad (6)$$

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

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

从整体风格角度考虑，本文引入了感知损失，它有助于测量特征域的差异，通过预训练 VGG-16 的多层特征距离计算得到

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

其中， $\varphi_l$  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<sup>[2]</sup>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：

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

其中， $I_o$  和  $I_u$  分别表示过曝和欠曝图像， $I_{\text{hdr}}$  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， $I_{\text{hdr}_p}$ 、 $I_{g_p}$ 、 $I_{o_p}$ 、 $I_{u_p}$  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：

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

其中，超参数  $\alpha$ 、 $\beta$ 、 $\gamma$ 、 $\delta$  都设置为 0.1。

## 3 实验结果与分析

### 3.1 实验设置

本次实验采用公开的 SICE<sup>[30]</sup>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  $320 \times 320$  和  $32 \times 32$ 。

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。

### 3.2 评价指标

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)<sup>[31]</sup>

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：

$$\text{PSNR} = 10 \log_{10} \frac{r^2}{\text{MSE}(I_g, I_{\text{hdr}})} \quad (13)$$

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

$$\text{MSE}(I_g, I_{\text{hdr}}) = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W (I_g(i, j) - I_{\text{hdr}}(i, j))^2 \quad (14)$$

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)<sup>[32]</sup>

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：

$$l(I_{\text{hdr}}, I_g) = \frac{2\mu_{\text{hdr}}\mu_g + C_1}{\mu_{\text{hdr}}^2 + \mu_g^2 + C_1} \quad (15)$$

$$c(I_{\text{hdr}}, I_g) = \frac{2\sigma_{\text{hdr}}\sigma_g + C_2}{\sigma_{\text{hdr}}^2 + \sigma_g^2 + C_2} \quad (16)$$

$$s(I_{\text{hdr}}, I_g) = \frac{\sigma_{\text{hdr}g} + C_3}{\sigma_{\text{hdr}}\sigma_g + C_3} \quad (17)$$

其中， $l$ 、 $c$ 、 $s$  分别表示亮度、对比度和结构， $\mu$ 、 $\sigma$  分别表示均值和方差， $\sigma_{\text{hdr}g}$  表示  $I_{\text{hdr}}$  和  $I_g$  的协方差， $C_1$ 、 $C_2$ 、 $C_3$  指常数。

$$\text{SSIM}(I_{\text{hdr}}, I_g) = \frac{1}{N} \sum_{k=1}^N S(I_{\text{hdr}_k}, I_{g_k}) \quad (18)$$

其中， $S = l(I_{\text{hdr}_k}, I_{g_k}) \cdot c(I_{\text{hdr}_k}, I_{g_k}) \cdot s(I_{\text{hdr}_k}, I_{g_k})$ 。N 表示滑窗截取图像块的个数。

### (3) 学习感知图像块相似度(LPIPS)<sup>[33]</sup>

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。

$$\text{LPIPS}(I_{\text{hdr}}, I_g) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} \|u_l \odot (I_{\text{hdr}_{hw}}^l - I_{g_{hw}}^l)\|_2^2 \quad (19)$$

其中， $l$  是指图像通道， $u_l$  指缩放因子， $H_l$ 、 $W_l$  分别表示图像的高和宽。

## 3.3 实验分析

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)<sup>[20]</sup>，Marnerides 等人(ExpandNet)<sup>[34]</sup>，Xu 等人(U2Fusion)<sup>[22]</sup>，Le 等人<sup>[23]</sup>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。



图 3 实验效果对比图

Fig .3 Comparison chart of experimental results

### (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

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 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。

## 5 结 论

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。

## 参考文献

- [1] Li Huang, Zhengping Li, Chao Xu, et al, Multi-exposure image fusion based on feature evaluation with adaptive factor[J]. IEEE Transactions on Image Processing. 2021, 15(13) 3211 - 3220.
- [2] Liu J, Wu G, Luan J, et al. HoLoCo: Holistic and local contrastive learning network for multi-exposure image fusion[J]. Information Fusion, 2023, 95: 237-249.
- [3] 林森, 刘世本. 基于多尺度生成对抗网络的水下图像增强[J]. 激光与光电子学进展, 2021, 58(16): 1610017.  
Lin S, Liu S B. Underwater Image Enhancement Based on Multiscale Generative Adversarial Network[J]. Laser & Optoelectronics Progress, 2021, 58(16): 1610017.
- [4] 辛元雪, 朱凤婷, 史朋飞, 等. 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[J]. 激光与光电子学进展, 2022, 59(4): 0420002.  
Xin Y X, Zhu F T, Shi P F, et al. Super-Resolution Reconstruction Algorithm of Images Based on Improved Enhanced Super-Resolution Generative Adversarial Network [J]. Laser & Optoelectronics Progress, 2022, 59(4): 0420002.



- [5] 汤璐, 卢荣胜, 史艳琼, 等. 基于 YCbCr 空间融合的高动态范围成像方法[J]. 激光与光电子学进展, 2022, 59(14): 1415029.
- Tang L, Lu R S, Shi Y Q, et al. High Dynamic Range Imaging Method Based on YCbCr Color Space Fusion[J]. Laser & Optoelectronics Progress, 2022, 59(14): 1415029.
- [6] 秦艳, 刘瑞星, 康臻, 等. 基于高动态范围成像的极端亮度测量方法研究[J]. 应用光学, 2023, 44(1): 168.
- Qin Y, Liu R X, Kang Z, et al. Extreme luminance measurement method based on high dynamic range imaging[J]. Journal of Applied Optics, 2023, 44(1): 168.
- [7] Paul E. Debevec and Jitendra Malik. 1997. Recovering high dynamic range radiance maps from photographs.[C] In Proceedings of the 24th annual conference on Computer graphics and interactive techniques (SIGGRAPH '97). ACM Press/Addison-Wesley Publishing Co.USA, 369–378.
- [8] Mertens T, Kautz J, Van Reeth F. Exposure fusion: A simple and practical alternative to high dynamic range photography[C]//Computer graphics forum. Oxford, UK: Blackwell Publishing Ltd, 2009, 28(1): 161-171.
- [9] Chen J, Paris S, Durand F. Real-time edge-aware image processing with the bilateral grid[J]. ACM Transactions on Graphics (TOG), 2007, 26(3): 103-es.
- [10] Banterle F, Ledda P, Debattista K, et al. A framework for inverse tone mapping[J]. The Visual Computer, 2007, 23: 467-478.
- [11] Mitsunaga T, Nayar S K. Radiometric self calibration[C]//Proceedings. 1999 IEEE computer society conference on computer vision and pattern recognition (Cat. No PR00149). IEEE, 1999, 1: 374-380.
- [12] Kalantari N K, Ramamoorthi R. Deep high dynamic range imaging of dynamic scenes[J]. ACM Transactions on Graphics., 2017, 36(4): 144:1-144:12.
- [13] Kim D E, Kim M. Learning-based low-complexity reverse tone mapping with linear mapping[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 30(2): 400-414.
- [14] Sharma A, Tan R T, Cheong L F. Estimation of Camera Response Function using Prediction Consistency and Gradual Refinement with an Extension to Deep Learning[J]. arXiv preprint arXiv:2010.04009, 2020.
- [15] Huang J, Liu Y, Fu X, et al. Exposure normalization and compensation for multiple-exposure correction[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 6043-6052.
- [16] Hu X, Shen L, Jiang M, et al. LA-HDR: Light Adaptive HDR Reconstruction Framework for Single LDR Image Considering Varied Light Conditions
- [17] Masia B, Serrano A, Gutierrez D. Dynamic range expansion based on image statistics[J]. Multimedia Tools and Applications, 2017, 76: 631-648.
- [18] Endo Y, Kanamori Y, Mitani J. Deep reverse tone mapping[J]. ACM Transactions on Graphics, 2017, 36(6): 177:1-177:10.

- [19] 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.
- [20] Lee S, An G H, Kang S J. Deep recursive hdri: Inverse tone mapping using generative adversarial networks[C]//proceedings of the European Conference on Computer Vision (ECCV). 2018: 596-611.
- [21] 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.
- [22] Xu H, Ma J, Jiang J, et al. U2Fusion: A unified unsupervised image fusion network[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 44(1): 502-518.
- [23] Le P H, Le Q, Nguyen R, et al. Single-image hdr reconstruction by multi-exposure generation[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2023: 4063-4072.
- [24] Hu X, Shen L, Jiang M, et al. LA-HDR: Light Adaptive HDR Reconstruction Framework for Single LDR Image Considering Varied Light Conditions[J]. IEEE Transactions on Multimedia, 2022. .3183404.
- [25] Liu S, Fan H, Lin S, et al. Adaptive learning attention network for underwater image enhancement[J]. IEEE Robotics and Automation Letters, 2022, 7(2): 5326-5333.
- [26] Land E H. The retinex theory of color vision[J]. Scientific american, 1977, 237(6): 108-129.
- [27] Mehta S, Rastegari M, Caspi A, et al. Espnet: Efficient spatial pyramid of dilated convolutions for semantic segmentation[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 552-568.
- [28] Ma K, Zeng K, Wang Z. Perceptual quality assessment for multi-exposure image fusion[J]. IEEE Transactions on Image Processing, 2015, 24(11): 3345-3356.
- [29] Mao X, Li Q, Xie H, et al. Least squares generative adversarial networks[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2794-2802.
- [30] Cai J, Gu S, Zhang L. Learning a deep single image contrast enhancer from multi-exposure images[J]. IEEE Transactions on Image Processing, 2018, 27(4): 2049-2062.
- [31] Huynh-Thu Q, Ghanbari M. Scope of validity of PSNR in image/video quality assessment[J]. Electronics letters, 2008, 44(13): 800-801.
- [32] Wang Z, Simoncelli E P, Bovik A C. Multiscale structural similarity for image quality assessment[C]//The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003. IEEE, 2003, 2: 1398-1402.
- [33] Zhang R, Isola P, Efros A A, et al. The unreasonable effectiveness of deep features as a perceptual metric[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 586-595.
- [34] 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[C]//Computer Graphics Forum. 2018, 37(2): 37-49.

网络首发:

标题: 基于双重注意力网络的高动态范围图像重建

作者: 王仙峰, 刘世本, 田建东, 赵娟平, 刘雅静, 郝春晖

收稿日期: 2023-07-21

录用日期: 2023-09-06

DOI: 10.3788/LOP231770

引用格式:

王仙峰, 刘世本, 田建东, 赵娟平, 刘雅静, 郝春晖. 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): 1210006.

网络首发文章内容与正式出版的有细微差别, 请以正式出版文件为准!

您感兴趣的其他相关论文:

一种基于螺旋波带片的无透镜边缘增强成像技术

田占全 李丽娜 孙达 马建设 苏萍

清华大学深圳国际研究生院, 广东 深圳 518055

激光与光电子学进展, 2023, 60(8): 0811022

基于 $L^{1/2}$ 范数的扇束X射线荧光CT重建

杨双 蒋上海 胡新宇 罗彬彬 赵明富 汤斌 龙邹荣 石胜辉 邹雪 周密

重庆理工大学光纤传感与光电检测重庆市重点实验室, 重庆 400054

激光与光电子学进展, 2023, 60(6): 0610010

基于迭代的纯相位全息图生成算法比较研究

王晓诗 桂进斌 李俊昌 宋庆和

昆明理工大学理学院物理系, 云南 昆明 650500

激光与光电子学进展, 2023, 60(6): 0609001

基于全广义变分约束加权最小二乘的低剂量计算机断层重建方法

牛善洲 张梦真 邱洋 李硕 梁礼境 刘宏 刘国良

赣南师范大学数学与计算机科学学院赣州市计算成像重点实验室, 江西 赣州 341000

激光与光电子学进展, 2023, 60(4): 0411001

基于图像增强和滚动引导滤波的红外与可见光图像融合

梁佳明 杨莘 田立凡

武汉科技大学信息科学与工程学院, 湖北 武汉 430081

激光与光电子学进展, 2023, 60(2): 0210006