

一种利用特征预对齐的高动态范围图像重建网络

谭台哲^{1,2}, 龚智远¹, 杨卓^{1,3}

(1. 广东工业大学计算机学院, 广东 广州 510006;

2. 河源市湾区数字经济技术发展有限公司, 广东 河源 517400;

3. 广东省人类运动表现科学重点实验室, 广东 广州 510500)

摘要: 基于多张不同曝光度的低动态范围(LDR)图像来重建高动态范围(HDR)图像,是一项具有挑战性的任务,尤其在相机和物体运动的情况下,运动区域常常会产生伪影,从而影响最终的重建图像质量。导致这一问题的根本原因在于多张 LDR 图像在内容上不对齐,图像之间的几何差异会显著影响重建效果。为了解决该问题,提出一种基于特征预对齐的 HDR 图像重建网络,旨在通过特征的预对齐来提高 HDR 重建的效果。该网络由特征预对齐和 HDR 重建 2 个主要部分组成:在特征预对齐部分,设计一个特征预对齐网络(FPAN),该网络能够引导输入图像的特征与参考图像对齐,从而减少由运动引起的伪影;在 HDR 重建部分,通过选择性状态空间模型来对预对齐后的特征进行全局上下文的建模,并通过简化的 HDR 恢复网络生成最终的 HDR 图像。为了评估所提网络的性能,在 2 个数据集上进行广泛的实验,结果显示,该网络在多个客观评估指标上优于对比方法,在主观视觉上也呈现出良好的效果,且具有一定的泛化能力。

关键词: 高动态范围图像;图像重建;特征预对齐;选择性状态空间模型;运动伪影

源代码链接: <https://github.com/ONGGZY/PAFHDR>

中图分类号: TP391.41

文献标志码: A

DOI: 10.19678/j.issn.1000-3428.0070697

A High Dynamic Range Image Reconstruction Network Using Feature Pre-Alignment

TAN Taizhe^{1,2}, GONG Zhiyuan¹, YANG Zhuo^{1,3}

(1. School of Computer, Guangdong University of Technology, Guangzhou 510006, Guangdong, China;

2. Heyuan Bay Area Digital Economy Technology Development Co., Ltd., Heyuan 517400, Guangdong, China;

3. Guangdong Key Laboratory of Human Sports Performance Science, Guangzhou 510500, Guangdong, China)

【Abstract】 Reconstructing High Dynamic Range (HDR) images from multiple Low Dynamic Range (LDR) images with varying exposures is challenging, especially in scenarios involving camera and object motion, where motion regions often introduce artifacts, thereby affecting the quality of the final reconstructed image. This issue primarily arises from the misalignment of content among multiple LDR images, where geometric differences between images significantly impact the reconstruction outcome. To address this issue, we propose an HDR image reconstruction network based on feature pre-alignment for enhancing HDR reconstruction quality. This network has two stages: feature pre-alignment and HDR reconstruction. In the feature pre-alignment stage, a Feature Pre-Alignment Network (FPAN) guides the alignment of features from input images with those from the reference image, thereby reducing artifacts caused by motion. In HDR reconstruction stage, a selective state space model is employed for modeling the global context of the pre-aligned features, and a simplified HDR restoration network generates the final HDR image. Extensive experiments are conducted on two datasets to evaluate the performance of the proposed network. The results show that the proposed network outperforms comparative methods across multiple objective evaluation metrics, exhibits satisfactory subjective visual effects, and demonstrates certain generalization capabilities.

【Key words】 High Dynamic Range (HDR) images; image reconstruction; feature pre-alignment; selective state space model; motion artifacts

0 引言

在使用普通相机进行摄影时,相机的传感器

往往难以捕捉到与人类视觉相匹配的亮度范围,这导致得到的低动态范围(LDR)图像在过曝区域或暗部会有细节缺失问题。为此,高动态范围

基金项目: 国家自然科学基金(62237001,61907009)。

作者简介: 谭台哲,男,副教授,博士,主研方向为计算机视觉、机器学习;龚智远(通信作者),硕士研究生;杨卓,讲师、博士。

收稿日期: 2024-12-10

修回日期: 2025-01-13

E-mail: 1139580423@qq.com

(HDR)图像技术应运而生,它通过结合多个不同曝光程度的图像来扩展图像的动态范围。HDR 图像不仅在视觉效果上更为逼真,而且在计算机视觉、虚拟现实、游戏图形等领域有着广泛的应用前景。

在 HDR 成像的早期,针对静态场景,主要通过双重曝光合成来捕捉亮部和暗部细节的技术来得到 HDR 图像。随着技术的发展,HDR 成像开始应用于动态场景,但是动态场景的剧烈运动会导致重建的 HDR 图像出现伪影。为了解决这一问题,一些方法通过对输入的多张 LDR 图像进行对齐^[1-3],或在图像融合前拒绝未对齐的像素^[4-6]来去除伪影。然而,准确的对齐很难实现,并且当通过不准确的像素拒绝而丢弃掉有用信息时,整体 HDR 图像效果会降低。

目前,基于深度学习的方法在学术界引起了显著关注。与传统方法相比,基于深度学习的方法提高了重建结果的视觉质量,这些方法主要分为显示对齐^[7-9]和隐式对齐^[10-12]两种方式。在显示对齐技术中,通常依赖于光流来调整图像,使其在视觉上的一致,但这种方法的有效性往往受限于光流估计的精确度,当遇到图像遮挡或过曝区域时,光流估计的准确性会大幅降低。为了克服这一局限,隐式对齐方法提供了一种替代方案,它不直接依赖于光流估计,从而能够规避相关挑战。在隐式对齐方法中,基于卷积神经网络(CNN)的方法提高了重建结果的质量。然而,上述方法主要抑制多重曝光图像融合过程中由于不对齐而产生的伪影,往往忽略了在曝光过度或曝光不足的区域的信息恢复,且由于卷积的固有局部性,会在结果中产生意外的失真伪影。

视觉 Transformer(ViT)^[13]因其出色的长距离建模能力而被引入到多种视觉任务中^[14-15]。因此,研究者们提出了融合 CNN 和 Transformer 的 HDR 图像重建方法。文献[16]提出了一种新的上下文感知视觉 Transformer,用于无伪影的 HDR 成像,其采用双分支架构,能够同时捕获全局和局部依赖性。文献[17]提出了一种 HDR 去伪影网络 HyHDRNet,该网络通过学习参考图像和非参考图像之间的复杂关系,来生成视觉上令人愉悦的 HDR 图像。文献[18]使用可变形卷积对 CNN 提取的 LDR 图像特征进行对齐,并在融合过程中引入 Transformer 来捕获特征之间的长距离关系。

上述方法利用了 Transformer 能有效捕获全局特性信息的能力,然而,这些方法未能有效解决在欠

曝区域有效信息的保存和丢失信息的恢复问题,这一局限性妨碍了重建结果视觉质量的提升。针对这些问题,本文设计一个特征预对齐的 HDR 重建网络,该网络主要包括 2 个部分,即多张 LDR 图像特征预对齐和 HDR 图像重建。在特征预对齐阶段,设计一个特征预对齐模块,该模块包含一个对齐损失函数来引导特征预对齐,以促进后续的 HDR 重建;在重建网络部分,结合一种新型的选择性状态空间模型 Mamba^[19],利用该模型进行全局上下文特征关系建模,以及设计一个简化的恢复网络,来得到最终的重建结果,该部分主要简化了常见网络的结构,如通道注意力和门控单元。

1 利用特征预对齐的 HDR 图像重建网络

1.1 网络输入预处理

假设有 3 张多曝光 LDR 源图像,分别记为 $\{L_1, L_2, L_3\}$ 。在通常情况下,为了获得满意的重建结果,建议对源图像执行伽马校正,即将 LDR 图像映射到 HDR 域得到 $\{\bar{L}_1, \bar{L}_2, \bar{L}_3\}$:

$$\bar{L}_i = \frac{L_i^\gamma}{t_i}, i=1,2,3 \quad (1)$$

式中: t_i 表示 L_i 的曝光时间; γ 是伽马校正参数,根据以往经验知识,通常将其设置为 2.2。校正后,将 L_i 和 \bar{L}_i 在通道维度上进行拼接,得到 $X_i = [\bar{L}_i, L_i]$,然后将其作为网络的输入。这种策略在文献[20]中被使用,因为 LDR 图像有助于检测噪声或饱和区域,而伽马校正的图像有助于检测错位。在多曝光 HDR 重建中,通常会选择中等曝光度的图像作为参考图像,而其他曝光程度的图像会作为补充信息被整合到参考图像中。

1.2 网络整体结构

如图 1 所示(彩色效果见《计算机工程》官网 HTML 版,下同),本文模型的整体结构主要由 2 个部分构成,即特征预对齐网络(FPAN)和 HDR 重建网络。

1) 特征预对齐网络。在多张不同曝光度的 LDR 图像中,存在由于相机运动和物体运动引起的图像配准误差。为了解决这一问题,特征预对齐网络的主要作用是对这些输入图像的特征进行预对齐,确保不同曝光图像之间的关键特征能够精确对齐到参考图像上。通过这一过程,能够有效减少由于运动引起的伪影,尤其在运动区域,确保生成的 HDR 图像更加真实和稳定。

2) HDR 重建网络。经过特征预对齐后,输入图像的特征已被精确对齐到参考图像中。HDR 重建

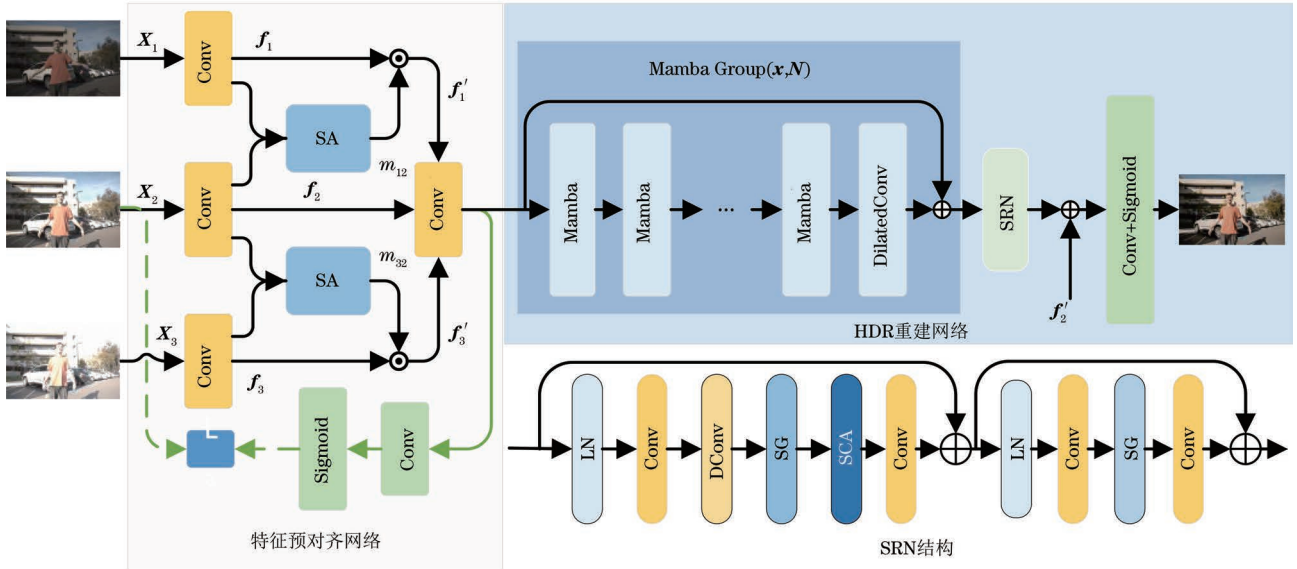


图 1 本文网络整体架构

Fig.1 The overall architecture of the network in this paper

网络的主要任务是将对齐后的特征进行融合,从而恢复出高质量的 HDR 图像。为了处理图像中的长距离依赖关系,本文采用了选择性状态空间模型,使网络能够在全局范围内捕捉图像的细节。最后,使用简化的 HDR 恢复模块,生成高质量的 HDR 图像。

总体来说,给定 3 张 LDR 图像,在经过式(1)的处理之后,作为输入,首先经过特征预对齐网络,得到对齐后的特征 f ,随后输入到重建网络部分,得到重建结果 $\hat{H} \in \mathbb{R}^{H \times W \times 3}$ 。

1.3 特征预对齐网络

FPAN 的结构如图 1 所示,对于输入的图片数据 $X_i \in \mathbb{R}^{H \times W \times 6}$, $i = \{1, 2, 3\}$,首先通过 3 个独立的卷积层提取粗糙特征 $f_i \in \mathbb{R}^{H \times W \times C}$,其中, C 是通道数。然后将每个非参考特征 f_1 和 f_3 分别与参考特征 f_2 拼接,并通过空间注意力(SA)模块计算得到注意力图 m_{12} 和 m_{32} :

$$\begin{cases} m_{12} = \text{SA}(f_1, f_2) \\ m_{32} = \text{SA}(f_3, f_2) \end{cases} \quad (2)$$

随后,将结果分别与对应的非参考特征 f_1 和 f_3 进行逐点相乘运算,得到 f'_1 和 f'_3 :

$$\begin{cases} f'_1 = f_1 \odot m_{12} \\ f'_3 = f_3 \odot m_{32} \end{cases} \quad (3)$$

式中: \odot 表示逐点相乘运算。

最后,将 f'_1 、 f_2 和 f'_3 进行拼接得到特征 f :

$$f = \text{cat}[f'_1, f_2, f'_3] \quad (4)$$

式中: cat 表示拼接操作。

对于上述流程,SA 模块的作用是根据图像中的空间位置来赋予不同区域不同的权重,关注更重

要的区域,抑制不重要的区域^[21]。由于 HDR 的重建是在参考图像的基础上进行的,因此设计对齐损失函数 L_a ,来引导特征 f 隐式地对齐到参考图像中。具体而言,将特征 f 经过一个卷积层和 Sigmoid 函数,得到一个与参考图像 X_2 大小相同的结果 Y ,然后利用 Sobel 算子得到它们各自的梯度图,最后计算它们的 L_a 损失:

$$Y = \text{Sigmoid}(\text{Conv}(f))$$

$$L_a = \|\phi(Y) - \phi(X_2)\|_1 \quad (5)$$

式中: $\text{Conv}(\cdot)$ 表示卷积; $\phi(\cdot)$ 表示利用 Sobel 算子来生成梯度图。通过这种方式,将特征 f 隐式地对齐到参考图像中,从而降低后续重建的难度,达到更好的重建效果。

1.4 HDR 图像重建网络

HDR 图像重建网络结构如图 1 所示,该部分由多个 Mamba 组(MG)和一个简化的恢复网络(SRN)组成。虽然 Mamba 模型相比 Transformer 在建模长距离依赖关系时具有更小的计算复杂度,但是 Mamba 模型最初是为自然语言处理而设计的,这是因为 Mamba 以递归方式处理展平的一维图像序列,这可能导致空间上接近的像素在展平序列中相距甚远,从而导致局部像素遗忘的问题。随后,经过改进的状态空间模型逐渐被运用到计算机视觉领域,如图像去模糊、超分辨率重建、图像分类等^[22-24],并证实了其有效性。为了将 Mamba 应用于 HDR 图像重建任务,本文参考文献[23],将原始 Mamba 的选择性扫描模块改进为二维选择性扫描模块,如图 2 所示,将原本对单个方向序列的处理变为对 4 个方向序列的处理,从而缓解局部像素遗忘的问题。

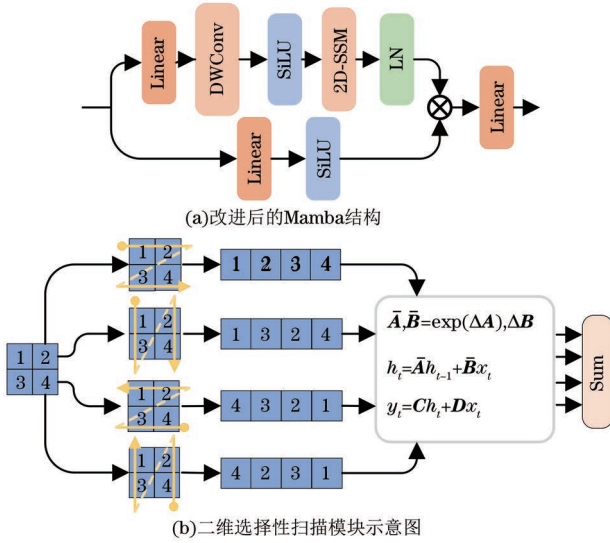


图 2 改进的 Mamba 结构和二维选择性扫描模块
Fig.2 Improved Mamba structure and 2D selective scanning module

在将 Mamba 的核心思想进行调整和优化后,使其能够有效处理图像中全局特征的关系,从而提高图像重建的效果。而简化的恢复网络则通过精简的结构设计,聚焦于从提取的特征中恢复最终的高质量 HDR 图像,进一步提升网络的效率与性能。具体而言,对于该部分网络的 MG 部分,每个 MG 包含 6 个 Mamba 和 1 个空洞卷积层,一共有 3 个 MG 模块,对于第 n 个 MG 输入 $F_{n,0}$,第 m 个 Mamba 的输出可以表示为:

$$F_{n,m} = \text{Mamba}(F_{n,m-1}) \quad (6)$$

式中: $n = \{1, 2, 3\}; m = \{1, 2, \dots, 6\}$ 。

然后,将第 6 个 Mamba 的输出输入到一个空洞卷积层中,这个空洞卷积层可以用来增加上下文的视野,为了更好地收敛,还采用了残差连接,因此,第 n 个 MG 的输出可以表示为:

$$F_n = \text{DConv}(F_{n,6}) + F_{n,0} \quad (7)$$

式中: $\text{DConv}(\cdot)$ 表示空洞卷积层。

恢复网络 SRN 的具体结构可以见图 1。受文献[25]的启发,该部分通过简化网络结构,从而降低了计算复杂性,使得网络在保持性能的同时更加高效。假设从输入特征 X 得到输出特征 Y ,则 SRN 可以表达为:

$$\begin{cases} X_1 = X \oplus W_p^2(\text{SCA}(\text{SG}(W_d^1 W_p^1(\text{LN}(X)))))) \\ Y = X_1 \oplus W_p^4(\text{SG}(W_p^3(\text{LN}(X_1)))) \\ \text{SG}(\cdot) = X_{f1} \odot X_{f2} \\ \text{SCA}(\cdot) = X_{f3} \cdot W_p^5 \cdot \text{GAP}(X_{f3}) \end{cases} \quad (8)$$

式中: $W_p^{(\cdot)}$ 是 1×1 的卷积; $W_d^{(\cdot)}$ 是 3×3 的深度可

分离卷积; $\text{GAP}(\cdot)$ 是全局池化层; $\text{SG}(\cdot)$ 和 $\text{SCA}(\cdot)$ 分别是简化的门控单元和通道注意力; $X_{f1}, X_{f2} \in \mathbb{R}^{H \times W \times C/2}$ 是通过划分 X_{f0} 通道而获得的。

更多关于 $\text{SG}(\cdot)$ 和 $\text{SCA}(\cdot)$ 的计算细节如图 3 所示。总体来说,最后一个 MG 模块的输出 F_3 经过 SRN 后再输入到最后的卷积层和 Sigmoid 函数,得到恢复的 HDR 图像 $\hat{H} \in \mathbb{R}^{H \times W \times 3}$:

$$\hat{H} = \text{Sigmoid}(\text{Conv}(\text{SRN}(F_3) + f_2)) \quad (9)$$

式中: $\text{Conv}(\cdot)$ 表示一个卷积层。

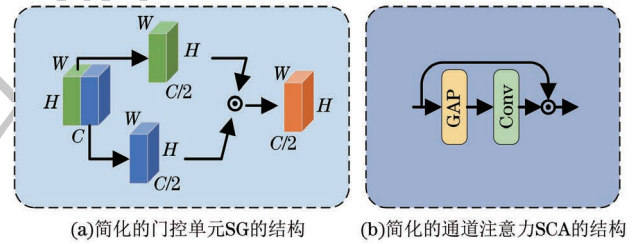


图 3 SG 和 SCA 的结构
Fig.3 SG and SCA structures

1.5 损失函数

在实际应用中, HDR 图像一般会通过色调映射处理后展示。因此,本文在训练网络时,除了在线性域的结果上设计损失函数,还在进行色调映射处理之后的结果上设计损失函数。假设 HDR 域中的图像为 H ,使用 μ -law 函数对 H 进行色调映射:

$$T(H) = \frac{\log_a(1 + \mu H)}{\log_a(1 + \mu)} \quad (10)$$

式中: μ 是压缩因子; $T(H)$ 表示色调映射后的图像。根据文献[7]的做法,设置 $\mu = 5\ 000$ 。损失函数设计 $L_{r,w}$ 和 $L_{ssim,w}$, $w = \{u, l\}$:

$$\begin{cases} L_{r,u} = \|T(H) - T(\hat{H})\|_1 \\ L_{ssim,u} = 1 - \text{SSIM}(T(H), T(\hat{H})) \\ L_{r,l} = \|H - \hat{H}\|_1 \\ L_{ssim,l} = 1 - \text{SSIM}(H, \hat{H}) \end{cases} \quad (11)$$

式中: u 和 l 分别表示色调映射后的域和原始的线性域; $\text{SSIM}(\cdot)$ 是结构相似性指数。加上前面提到的对齐损失函数 L_a ,因此,总的损失函数表示为:

$$L = L_a + L_{r,w} + L_{ssim,w} \quad (12)$$

2 实验与结果分析

2.1 实验数据集和评估指标

1) 数据集。训练网络所使用的数据集参考文献[7],在这个数据集中,74 个样本用于训练,15 个样本用于测试,每个样本包含 3 个曝光值分别为

$\{-2, 0, 2\}$ 或 $\{-3, 0, 3\}$ 的 LDR 图像, 每组图像都附有真实图像。文献[26]的数据集提供 8 个场景的多曝光 LDR 图像, 这个数据集不包括真实数据的 HDR 图像。本文使用文献[7]数据集的训练集训练所提网络, 随后在文献[7]数据集的测试集和文献[26]数据集上进行测试。

2) 评估指标。对于包含真实 HDR 图像的文献[7]数据集, 使用 PSNR-L、SSIM-L、PSNR-u、SSIM-u 以及 HDR-VDP-2^[27-28] 作为评估标准。对于没有真实 HDR 图像的文献[26]数据集, 参考以往的工作^[29], 使用 BTMQI^[30] 和 MEF-SSIMd^[31] 作为评估指标。PSNR-L 和 SSIM-L 通过计算线性域中的 PSNR 和 SSIM, 来衡量重建 HDR 图像与真实 HDR 图像的相似度; PSNR-u 和 SSIM-u 对比经过色调映射处理后的重建图像和对应的真实图像; HDR-VDP-2 专门用于评估 HDR 图像的视觉质量; BTMQI 通过评估色调映射处理后的 HDR 图像在信息熵、自然度和结构保真度方面的表现, 来衡量图像的整体质量, 且 BTMQI 值越小, 表明重建效果越好; MEF-SSIMd 分别计算 HDR 重建图像在动态区域和静态区域与真实图像的 SSIM 值, 并将两者的

评估结果进行平均, 以综合评估重建质量, MEF-SSIMd 值越大, 表示重建效果越好。

2.2 实施细节

本文网络模型是使用 PyTorch 实现的, 使用 Adam 优化器, 采用余弦退火策略。对于优化器超参数的设置, 按照以往工作经验, 设置 $\beta_1 = 0.9$ 、 $\beta_2 = 0.999$ 、 $\epsilon = 1 \times 10^{-8}$ 。初始学习率为 2×10^{-4} , 逐渐降低到 1×10^{-6} 。本文网络的输入为从训练集中随机裁取的 128×128 像素的图像块, 并将批量大小设置为 2。对于数据增强方面, 采取水平和垂直翻转策略。网络共训练 5 500 轮, 并在 1 张 NVIDIA 3090 GPU 上完成训练, 消融实验也采用同样的训练策略。

2.3 对比实验

2.3.1 文献[7]数据集上的对比实验

为了验证所提方法的有效性, 选取文献[7]方法、DeepHDR^[9]、HDRI^[10]、AHDRNet^[32]、SGARN^[33]、HDR-GAN^[34] 以及文献[29]方法这 7 个具有代表性的方法, 在文献[7]数据集上进行比较。

过曝和大运动区域场景下的实验结果如图 4 所示。从图 4 结果来看: 文献[7]方法依赖于光流估计



图 4 文献[7]测试数据集中的“停车”场景下不同方法的 HDR 重建结果比较

Fig.4 Comparison of HDR reconstruction results using different methods in the "parking" scenario from the test dataset in reference [7]

的准确性,若光流估计不准确,会导致重建结果产生伪影;DeepHDR 没有考虑在饱和区域中丢失细节的恢复,导致生成的 HDR 图像中细节丢失严重;AHDRNet 利用注意力机制抑制伪影效应,但在重建 HDR 图像的视觉质量上仍有显著的改进空间;HDR-GAN 没有考虑由移动对象引起的空间不一致性,从而导致重建结果中有明显伪影;HDRI 和 SGARN 没有充分解决过曝区域中丢失细节恢复的问题,导致重建的 HDR 图像细节有所缺失;本文方法在对图像进行重建之前,先进行一个特征预对齐,使其他 LDR 图像特征配准到参考图像,降低由于不对齐而增加的重建难度,并且利用选择性状态空间模型,对预对齐后的特征进行全局上下文信息的建

模,充分挖掘参考和非参考图像之间的信息,为后续恢复网络在参考图像上进行 HDR 重建提供更多的互补信息,以提高最终的重建质量。此外,本文网络在过曝和剧烈运动的情况下也能取得良好的视觉效果。

为了客观评估不同方法的重建质量,列出不同方法在文献[7]数据集的 15 个测试场景中的平均客观评估结果,如表 1 所示,最优结果加粗标注,次优结果加下划线标注。这些结果表明,所提方法在多个指标上均优于对比方法,在 SSIM-L、SSIM-u 和 PSNR-u 指标上也表现出了次优的性能。其中, HDR-VDP-2 这一指标与显示器参数有关,不同的参数配置会产生不同的结果,因此仅作参考。

表 1 文献[7]数据集上的定量比较结果

Table 1 Quantitative comparison results on the dataset from reference [7]

方法	PSNR-u/dB	SSIM-u	PSNR-L/dB	SSIM-L	HDR-VDP-2
文献[7]方法	42.74	0.987 7	40.72	0.982 4	62.87
DeepHDR	41.62	0.986 5	40.88	0.985 8	57.37
HDRI	43.65	0.989 4	41.76	0.986 7	64.46
AHDRNet	43.74	0.991 3	41.69	0.986 9	63.51
SGARN	43.96	0.990 7	41.51	0.987 4	<u>65.11</u>
HDR-GAN	43.92	0.990 5	41.57	0.986 5	64.70
文献[29]方法	44.11	0.990 9	<u>42.00</u>	0.988 4	64.93
本文方法	<u>44.10</u>	0.991 1	42.22	<u>0.988 0</u>	65.63

2.3.2 文献[26]数据集上的对比实验

为了进一步验证本文网络模型的泛化能力,在文献[7]数据集上训练所提网络,随后将该训练好的网络在文献[26]数据集上进行测试,并与文献[7]方法、DeepHDR^[9]、AHDRNet^[32]、HDR-GAN^[34]、HDRI^[10]、文献[29]方法这 6 个具有代表性的方法进行比较。从图 5 中的放大区域可以看出,相比其他方法,本文方法呈现出了一个更好的主观视觉效果,特别是对纹理细节恢复有着更清晰的呈现。

此外,文献[26]数据集上各方法的定量评估结果如表 2 所示。本文方法在 BTMQI 和 MEF-SSIMd 指标上分别取得了最优和次优的性能,这进一步展示了该方法具有一定的泛化能力。

2.4 消融实验

本文方法的主要改进在于特征预对齐网络和恢复网络。因此,本文将完整网络中的对齐损失函数和 SRN 去除并作为基线模型,进行消融实验。定义 3 种方法:1)w/o L_a ,表示不使用对齐损失函数;2)w/o SRN,表示不使用 SRN;3)baseline,表示

基线模型。这 3 种方法采用的数据集都是文献[7]数据集。消融实验结果如表 3 所示,在使用预对齐损失函数或 SRN 时,能表现出更好的效果,两者同时存在时,效果能达到最优。其中,预对齐损失函数能将特征预对齐到参考图像,使得重建结果中减少了由于不对齐而产生的伪影,而在 SRN 恢复网络中,使用了简化的通道注意力和门控单元,通道注意力机制通过自适应地给每个通道分配一个权重,自动学习对最终 HDR 恢复任务最重要的通道,而门控单元通过引入门控结构来决定最终的 HDR 恢复任务中哪些信息需要传递,哪些信息需要保留或丢弃。

消融实验的定性结果如图 6 所示,可以明显看出,结合所有组件的网络模型显著提升了图像的细节恢复、运动区域的清晰度,这些结果验证了本文网络架构中各个模块的协同作用,并证明了每个模块在提高图像质量方面的有效性。

2.5 计算成本分析

为了分析本文方法的计算成本,以输入大小为 128×128 像素为标准,对不同方法的模型参数数量和

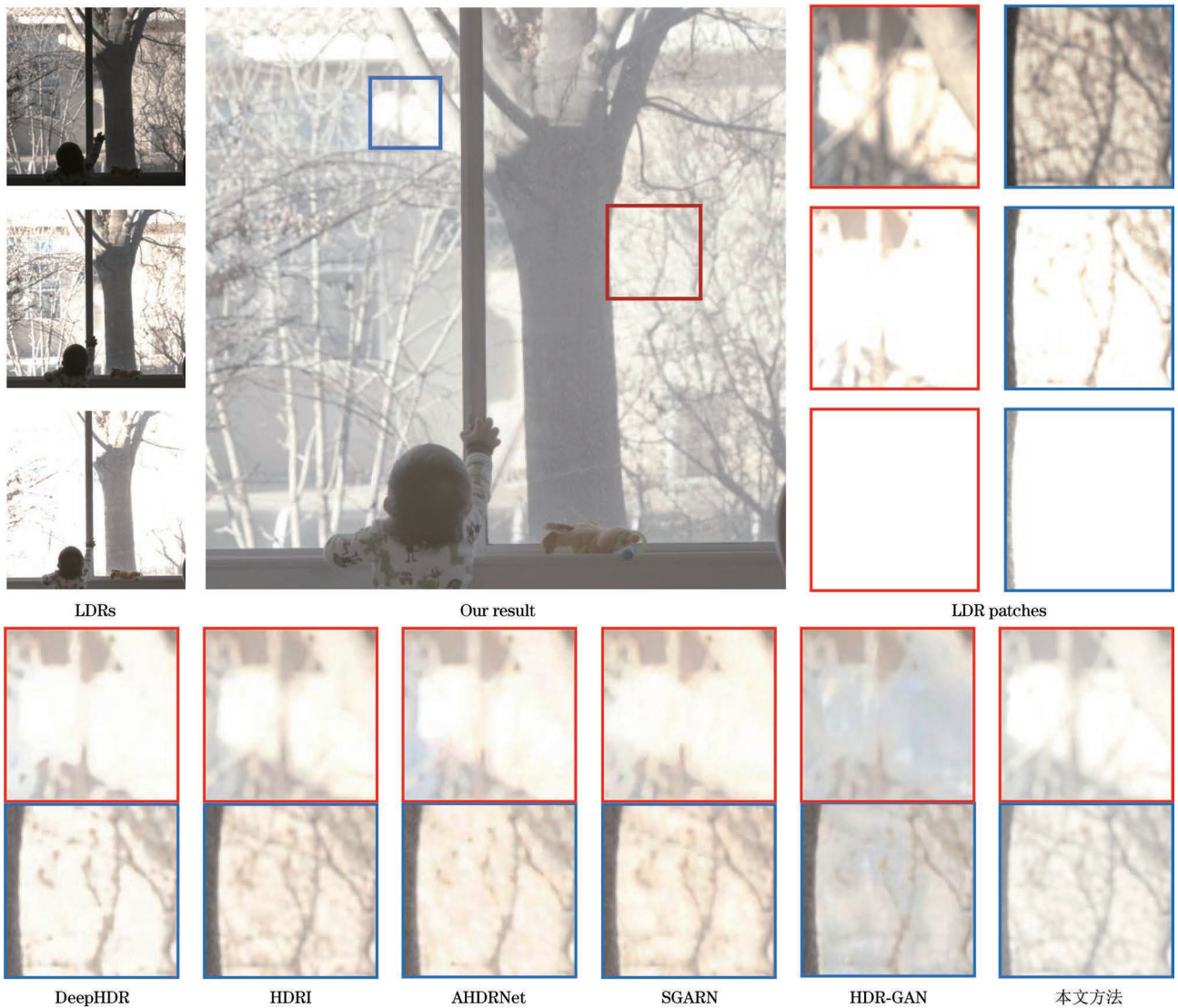


图 5 文献[26]数据集上的定性结果

Fig.5 Qualitative results on the dataset from reference [26]

表 2 文献[26]数据集上的定量评估结果

Table 2 Quantitative evaluation results on the dataset from reference [26]

方法	BTMQI	MEF-SSIMd
文献[7]方法	3.767 9	0.763 6
DeepHDR	3.774 4	0.799 0
AHDRNet	3.687 7	0.795 7
HDR-GAN	7.134 0	0.799 2
HDRI	5.503 6	0.802 4
文献[29]方法	<u>3.523 9</u>	0.805 0
本文方法	3.512 1	<u>0.803 4</u>

FLOPs 进行比较。

如表 4 所示,本文方法的网络参数数量为 1.30×10^6 , FLOPs 为 21.22×10^9 , FLOPs 的大小尽管没

表 3 消融实验定量结果

Table 3 Quantitative results of ablation experiment

方法	PSNR-L/dB	PSNR-u/dB	SSIM-L	SSIM-u
baseline	40.52	42.84	0.983 2	0.985 9
w/o L_a	41.97	43.41	0.986 5	0.990 5
w/o SRN	42.00	43.85	0.988 2	0.990 7
full	42.22	44.10	0.988 0	0.991 1

有取得一个令人满意的值,但是本文网络的参数量较小,总体来说其计算成本在可接受范围内。对于 FLOPs 值偏大的原因,本文分析是网络中采取的 Mamba 模型,其从 4 个方向上进行全局上下文特征建模,这种做法尽管能增强建模能力,但是同时也会增加计算负担。针对该点缺陷,未来考虑设计一个计算开销更低且重建质量良好的 HDR 重建网络。

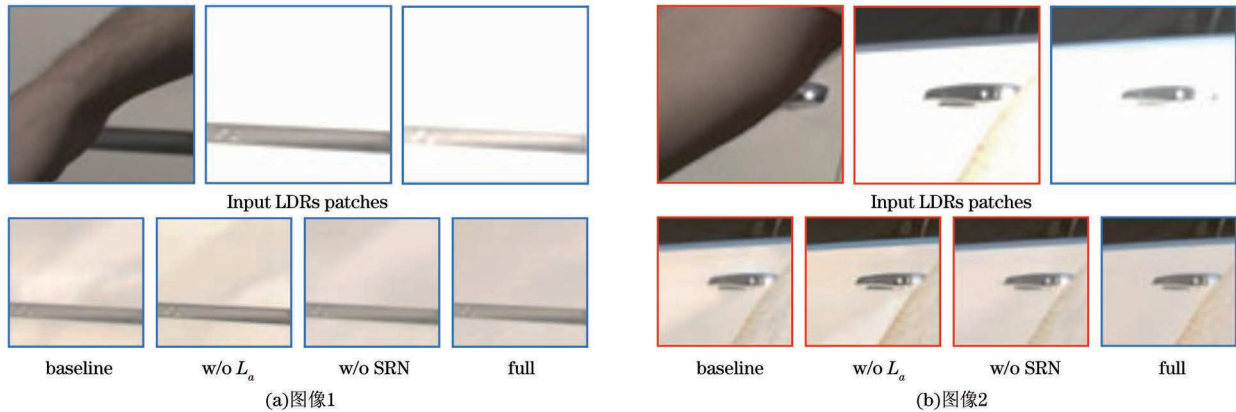


图 6 消融实验定性结果

Fig. 6 Qualitative results of ablation experiment

表 4 不同方法的 FLOPs 和模型参数量对比结果

Table 4 Comparison results of FLOPs and model parameter counts for different methods

方法	FLOPs/10 ⁹	Parameters/10 ⁶
文献[7]方法	—	0.30
DeepHDR	15.93	16.61
HDR-GAN	—	7.76
AHDRNet	23.70	1.44
HDR1	—	6.70
SGARN	13.76	0.89
文献[29]方法	114.66	6.33
本文方法	21.22	1.30

3 结束语

基于多张不同曝光度的 LDR 图像来重建 HDR 图像,是一项具有挑战性的任务,尤其在相机和物体运动的情况下,常常会导致重建的 HDR 图像在运动区域产生伪影。为了应对这一挑战,本文提出一种基于特征预对齐的 HDR 图像重建方法。通过特征预对齐网络对输入图像的特征进行隐式对齐,有效减少运动区域的伪影,并在后续的重建阶段保持图像的高质量。实验结果表明,所提方法在有真实标签的文献[7]数据集上达到了较优的客观指标和主观视觉效果,有效解决了运动伪影问题。另外,为了验证本文所提方法的泛化能力,针对不存在真实标签的文献[26]数据集,本文将在文献[7]数据集上训练出的网络应用到文献[26]数据集上进行测试,结果表明,所提网络具有一定的泛化能力和鲁棒性。

尽管本文模型在多个测试集上取得了理想的效果,但仍存在一些改进的空间。未来的工作将致力于进一步提升网络的计算效率,使其在实时处理和

大规模数据集上更加高效。另外,考虑设计一个更合理的特征预对齐模块和对应的对齐损失函数,来增强特征的预对齐,以此提升重建效果。

参考文献

- [1] HU J, GALLO O, PULLI K, et al. HDR deghosting: how to deal with saturation? [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2013: 1163-1170.
- [2] KANG S B, UYTENDAELE M, WINDER S, et al. High dynamic range video [J]. ACM Transactions on Graphics, 2003, 22(3): 319-325.
- [3] ZIMMER H, BRUHN A, WEICKERT J. Freehand HDR imaging of moving scenes with simultaneous resolution enhancement [J]. Computer Graphics Forum, 2011, 30(2): 405-414.
- [4] GALLO O, GELFANDZ N, CHEN W C, et al. Artifact-free high dynamic range imaging [C] // Proceedings of the IEEE International Conference on Computational Photography (ICCP). Washington D. C., USA: IEEE Press, 2010: 1-7.
- [5] JACOBS K, LOSCOS C, WARD G. Automatic high-dynamic range image generation for dynamic scenes [J]. IEEE Computer Graphics and Applications, 2008, 28(2): 84-93.
- [6] PECE F, KAUTZ J. Bitmap movement detection: HDR for dynamic scenes [C] // Proceedings of the Conference on Visual Media Production. Washington D. C., USA: IEEE Press, 2011: 1-8.
- [7] KALANTARI N K, RAMAMOORTHI R. Deep high dynamic range imaging of dynamic scenes [J]. ACM Transactions on Graphics, 2017, 36(4): 1-12.
- [8] PRABHAKAR K R, AGRAWAL S, SINGH D K, et al. Towards practical and efficient high-resolution HDR deghosting with CNN [EB/OL]. [2024-08-05]. https://link.springer.com/chapter/10.1007/978-3-030-58589-1_30.
- [9] WU S Z, XU J R, TAI Y W, et al. Deep high dynamic range imaging with large foreground motions [EB/OL]. [2024-08-05]. <https://arxiv.org/abs/1711.08937>.
- [10] CHUNG H, CHO N I. High dynamic range imaging of dynamic scenes with saturation compensation but without explicit motion compensation [C] // Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Washington D. C., USA: IEEE Press, 2022: 61-71.
- [11] LIU Z, LIN W J, LI X P, et al. ADNet: attention-guided deformable convolutional network for high dynamic range

- imaging[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Washington D. C., USA:IEEE Press, 2021: 463-470.
- [12] TAN X, CHEN H A, XU K, et al. High dynamic range imaging for dynamic scenes with large-scale motions and severe saturation[J]. *IEEE Transactions on Instrumentation and Measurement*, 2022, 71: 5003415.
- [13] ALEXEY D. An image is worth 16×16 words: Transformers for image recognition at scale[EB/OL]. [2024-08-05]. <https://arxiv.org/abs/2010.11929>.
- [14] 田永林, 王雨桐, 王建功, 等. 视觉 Transformer 研究的关键问题: 现状及展望[J]. *自动化学报*, 2022, 48(4): 957-979. TIAN Y L, WANG Y T, WANG J G, et al. Key problems and progress of vision Transformer: the state of the art and prospects[J]. *Acta Automatica Sinica*, 2022, 48(4): 957-979. (in Chinese)
- [15] 彭斌, 白静, 李文静, 等. 面向图像分类的视觉 Transformer 研究进展[J]. *计算机科学与探索*, 2024, 18(2): 320-344. PENG B, BAI J, LI W J, et al. Survey on visual Transformer for image classification[J]. *Journal of Frontiers of Computer Science and Technology*, 2024, 18(2): 320-344. (in Chinese)
- [16] LIU Z, WANG Y L, ZENG B, et al. Ghost-free high dynamic range imaging with context-aware Transformer[EB/OL]. [2024-08-05]. <https://arxiv.org/abs/2208.05114>.
- [17] YAN Q S, CHEN W Y, ZHANG S, et al. A unified HDR imaging method with pixel and patch level[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Washington D. C., USA:IEEE Press, 2023: 22211-22220.
- [18] CHEN R F, ZHENG B L, ZHANG H, et al. Improving dynamic HDR imaging with fusion Transformer [J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023, 37(1): 340-349.
- [19] GU A, DAO T. Mamba: linear-time sequence modeling with selective state spaces[EB/OL]. [2024-08-05]. <https://arxiv.org/abs/2312.00752>.
- [20] 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.
- [21] 贺姗, 蔺素珍, 王彦博, 等. 基于特征融合的多波段图像描述生成方法[J]. *计算机工程*, 2024, 50(6): 236-244. HE S, LIN S Z, WANG Y B, et al. Multi-band image caption generation method based on feature fusion [J]. *Computer Engineering*, 2024, 50(6): 236-244. (in Chinese)
- [22] GUO H, LI J M, DAI T, et al. MambaIR: a simple baseline for image restoration with state-space model [EB/OL]. [2024-08-05]. <https://arxiv.org/abs/2402.15648>.
- [23] ZHU L H, LIAO B C, ZHANG Q, et al. Vision Mamba: efficient visual representation learning with bidirectional state space model [EB/OL]. [2024-08-05]. <https://arxiv.org/html/2401.09417v3>.
- [24] CHEN S, ATAPOUR A A, ZHANG H Z, et al. MxT: Mamba x Transformer for image inpainting[EB/OL]. [2024-08-05]. <https://arxiv.org/abs/2407.16126>.
- [25] CHEN L Y, CHU X J, ZHANG X Y, et al. Simple baselines for image restoration [EB/OL]. [2024-08-05]. <https://arxiv.org/abs/2204.04676>.
- [26] SEN P, KALANTARI N K, YAESOUBI M, et al. Robust patch-based HDR reconstruction of dynamic scenes[J]. *ACM Transactions on Graphics*, 2012, 31(6): 1-11.
- [27] MANTIUK R, KIM K J, REMPEL A G, et al. HDR-VDP-2: a calibrated visual metric for visibility and quality predictions in all luminance conditions [J]. *ACM Transactions on Graphics*, 2011, 30(4): 1-14.
- [28] EILERTSEN G, HAJISHARIF S, HANJJP, et al. How to cheat with metrics in single-image HDR reconstruction[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). Washington D. C., USA:IEEE Press, 2021: 3981-3990.
- [29] 杨珍妹, 李华锋, 张亚飞. 面向高动态范围成像的内容恢复和鬼影抑制网络[J]. *模式识别与人工智能*, 2024, 37(4): 313-327. YANG Z M, LI H F, ZHANG Y F. Missing content restoration and ghosting suppression network for high dynamic range imaging[J]. *Pattern Recognition and Artificial Intelligence*, 2024, 37(4): 313-327. (in Chinese)
- [30] GU K, WANG S Q, ZHAI G T, et al. Blind quality assessment of tone-mapped images via analysis of information, naturalness, and structure [J]. *IEEE Transactions on Multimedia*, 2016, 18(3): 432-443.
- [31] FANG Y M, ZHU H W, MA K D, et al. Perceptual evaluation for multi-exposure image fusion of dynamic scenes [J]. *IEEE Transactions on Image Processing*, 2020, 29: 1127-1138.
- [32] YAN Q S, GONG D, SHI Q F, et al. Attention-guided network for ghost-free high dynamic range imaging [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Washington D. C., USA:IEEE Press, 2020: 1751-1760.
- [33] TANG L F, HUANG H, ZHANG Y F, et al. Structure-embedded ghosting artifact suppression network for high dynamic range image reconstruction [J]. *Knowledge-Based Systems*, 2023, 263: 110278.
- [34] NIU Y Z, WU J B, LIU W X, et al. HDR-GAN: HDR image reconstruction from multi-exposed LDR images with large motions[J]. *IEEE Transactions on Image Processing*, 2021, 30: 3885-3896.

文字编辑 吴云芳
栏目编辑 宋圆