基于一致性感知特征融合的高动态范围成像方法

印佳丽 韩 津 陈 斌 刘西蒙

(福州大学计算机与大数据学院 福州 350108)

(福州大学网络系统信息安全福建省高校重点实验室 福州 350108)

摘 要 高动态范围成像(High Dynamic Range Imaging,HDRI)技术是指通过融合多张低动态范围(Low Dynamic Range, LDR)图像拓展图像动态范围、完整图像内容的方法,其为解决由于相机传感器动态范围有限而导致所拍摄图像内容丢失的问题提供了实际的解决方案.通过数十年的研究,众多有效的 HDRI 方法已被提出,并在无物体运动、内容曝光良好的静态场景中取得接近最优的性能.然而,现实场景中物体移动和相机偏移无法避免,直接使用传统 HDRI 方法会在融合后的 HDR 图像中产生严重的重影和伪影.这使得仅包含简单融合过程的 HDRI 方法并不适用于实际应用,现实场景中的 HDRI 任务仍然具有一定挑战.因此,针对动态场景下的 HDRI 研究迅速发展.近期的方法集中在借助深度卷积神经网络(Convolutional Neural Network,CNN)的力量以期实现更好的性能.在这些基于 CNN 的方法中,特征融合对于恢复图像完整内容、消除图像伪影方面起着至关重要的作用.传统的特征融合方法通过借助跳跃连接或注意力模块,首先将 LDR 图像的特征进行拼接,并通过堆叠的卷积操作逐渐关注不同的局部特征.然而,此类方案通常忽略了 LDR 图像序列之间丰富的上下文依赖关系,且未充分利用特征之间的纹理一致性.为解决这一问题,本文提出了一种全新的一致性感知特征融合(Coherence-Aware Feature Aggregation,CAFA)方案,该方案在卷积过程中对输入特征中位于不同空间位置但具有相同上下文信息的特征信息进行采样,从而显式地将上下文一致性纳入特征融合中.基于 CAFA,本文进一步提出了一种结合 CAFA 的动态场景下一致性感知高动态范围成像网络 CAHDRNet.为更好地嵌合 CAFA 方案,本文通过设计三个额外的可学习模块来构建 CAHDRNet.首先,使用基于在 ImageNet 上预训练的 VGG-19 构建可学习特征提取器,并在模型训练期间不断更新该特征提取器的参数.这种设计可实现 LDR 图像的联合特征学习,为 CAFA 中的上下文一致性评估奠定了坚实基础.接着,应用所提出的 CAFA 模块,通过在图像特征中采样具有相同上下文的信息进行特征融合.最后,本文提出使用一种多尺度残差补全模块来处理融合后的特征,利用不同扩张率进行特征学习,以实现更强大的特征表示并在图像缺失区域中进行可信细节填充.同时,设计一个软注意力模块来学习不同图像区域的重要性,以便在跳跃连接期间获得与参考图像互补的所需特征.多种实验验证了 CAHDRNet 的有效性并证实其优于现有最先进的方法.具体而言,本文所提出的 CAHDRNet 在 Kalantari 数据集上 *HDR-VDP*-2 和 *PSNR*-L 等指标相较于次好方法 AHDRNet 分别提升了 1.61 和 0.68.

关键词 高动态范围成像;图像融合;特征融合;上下文一致性;卷积采样

中图法分类号 TP391 **DOI 号** 10.11897/SP.J.1016.2024.02352

High Dynamic Range Imaging Based on Coherence-Aware Feature Aggregation

YIN Jia-Li HAN Jin CHEN Bin LIU Xi-Meng

(*College of Computer Science and Big Data, Fuzhou University, Fuzhou* 350108)

(*Fujian Provincial Key Laboratory of Information Security of Network Systems, Fuzhou University, Fuzhou* 350108)

Abstract High Dynamic Range Imaging (HDRI) is a technology of fusing multiple Low Dynamic Range (LDR) images to extend image dynamic range, restore image contents and generate high dynamic range (HDR) images. It provides a practical solution to the problem of content loss in

收稿日期:2023-06-29;在线发布日期:2024-07-08.本课题得到国家自然科学基金(62202104,62102422,62072109,U1804263)、福建省自然科学基金(2021J05129,2021J06013)资助.**印佳丽**,博士,研究员,主要研究领域为计算摄影学、神经网络鲁棒性.E-mail:jlyin@fzu.edu.cn.**韩 津**,硕士,主要研究方向为底层图像处理.**陈 斌**,博士研究生,主要研究方向为神经网络对抗攻防、计算机视觉.**刘西蒙**(通信作者),博士,教授,主要研究领域为大数据存储及处理、数据安全.E-mail:snbnix@gmail.com.

captured images due to the limited dynamic range of the camera sensors. With decades of studies, numerous promising approaches have been proposed and near-optimal performance has been achieved for the HDRI static scenes with no object motions and well-exposed contents. However, object motions or camera shifts are inevitable in practical scenarios. Directly using traditional HDRI methods would induce severe ghosting artifacts into the merged HDR image. This makes HDRI with simple merging process inapplicable in real-world applications, which poses a challenge to the HDRI task. Thus, the study on HDRI of dynamic scenes has grown rapidly. Recent advances focus on exploring the power of deep convolutional neural networks (CNNs) to achieve a better performance. Among these CNN-based methods, the feature aggregation plays a crucial role in completing image contents and eliminating ghosting artifacts. Equipped with skip connections or attention modules, the features derived from multiple LDR images are first concatenated and then gradually focus on different local aspects via stacked convolutions. However, such aggregation schemes generally neglect to utilize the rich contextual dependencies across LDR image sequence, the textural coherence among features have not been fully exploited. To address this issue, this paper proposes a novel Coherence-Aware Feature Aggregation (CAFA) scheme that samples grids with the same contextual information instead of the same position across input features during convolutional operations, so that contextual coherence can be explicitly incorporated into feature aggregation. Based on CAFA, this paper further proposes Coherence-Aware HDR Network (CAHDRNet) for HDRI of dynamic scenes. To facilitate the incorporation of CAFA, the proposed CAHDRNet is constructed by designing three additional learnable modules. Firstly, a learnable feature extractor, which is built upon a VGG-19 pre-trained on ImageNet, is used to extract features from each LDR image. The parameters will be updated during end-to-end training. Such a design enables a joint feature learning of LDR images which creates a solid foundation for applying the coherence evaluation in CAFA. Then, the proposed CAFA module is applied to aggregate the features by sampling grids with the same contextual information in each image features. Next, a Multi-Scale Residual Hallucinating (MSRH) module is proposed to process the aggregated features, in which the features are learnt across different scales of dilated rates to achieve a more powerful feature representation and hallucinate plausible details in the missing regions. Also, a soft attention module is equipped to learn the importance of different image regions for obtaining the features that are complementary to the reference image during skip connection. Various experiments are conducted to validate the effectiveness of our proposed CAHDRNet, where it demonstrates superior performance over state-of-the-art (SOTA) methods. Specifically, the proposed CAHDRNet improves the *HDR-VDP*-2 and *PSNR*-L values on Kalantari's dataset over the second-best AHDRNet by 1.61 and 0.68, respectively.

Keywords　high dynamic range imaging; image fusion; feature aggregation; contextual coherence; convolutional sampling

1　引　言

高动态范围成像(High Dynamic Range Imaging, HDRI)是一种通过融合多张具有不同曝光水平的低动态范围(Low Dynamic Range, LDR)图像来拓展图像动态范围、恢复图像内容、形成高动态范围(High Dynamic Range, HDR)图像的技术[1]. 该技术在现实中具有广泛的应用场景,如摄影[2-3]、电影[4-5]和电子游戏[6]等领域. 经过数十年的研究,专家学者们已提出众多有效的 HDRI 方法[7-12],并在场景静态、内容固定以及曝光良好 LDR 图像的 HDRI 中取得了接近最优的性能. 然而,在现实场景中,大规模前景运动和手持拍摄等原因常会导致 LDR 图像之间

的内容错位,光线等自然条件也会导致图像内容存在不同面积的过曝或欠曝.使用传统的 HDRI 方法所生成的 HDR 图像中往往掺杂伪影、内容不全且缺乏细节.因此,复杂场景下的高动态范围成像仍然是一个具有挑战性的问题.

近年来,针对复杂场景下的 LDR 图像序列内容错位的问题,专家学者提出了使用"对齐-融合"框架来缓解融合过程中的伪影生成.具体来说,这类方法首先将 LDR 图像通过光流法进行图像对齐,然后应用卷积神经网络(Convolutional Neural Networks,CNNs)模型来融合对齐后的图像[13-15].然而,这种基于光流的对齐方法仅依赖于简单像素级特征,对于包含复杂的运动物体的图像很难有良好的效果,特别是对于具有不同曝光水平的图像序列来说,对齐不良反而会在融合结果中加重伪影的生成.因此,一些研究选择跳过这种像素级图像对齐过程,专注于使用带有残差块的卷积网络进行图像融合[16-17],同时尝试使用具有跳连接[1,18]或注意力模块[19]的卷积网络来进一步改善融合质量.

尽管现有的大多数方法已经取得了显著的效果,但它们大多都忽视了 LDR 图像之间丰富的上下文一致性,仅通过堆叠的卷积操作来融合特征.具体而言,几乎所有的现有方法都只使用简单的特征串联来组合特征,然后使用具有局部和固定感受野的卷积模块从中提取更复杂的特征.由于被融合的特征是从每个图像的相同位置区域中学习得到,因此忽视了输入图像之间丰富的上下文一致性,将直接导致融合图像中伪影的产生.如图 1 所示,由于物体在移动,每个图像中的同一位置区域可能代表不同的上下文信息,若进行简单的传统特征串联式融合方案,伪影的产生将不可避免.因此,打破传统特征融合中的基于空间位置关系的融合对于 HDRI 任务至关重要.

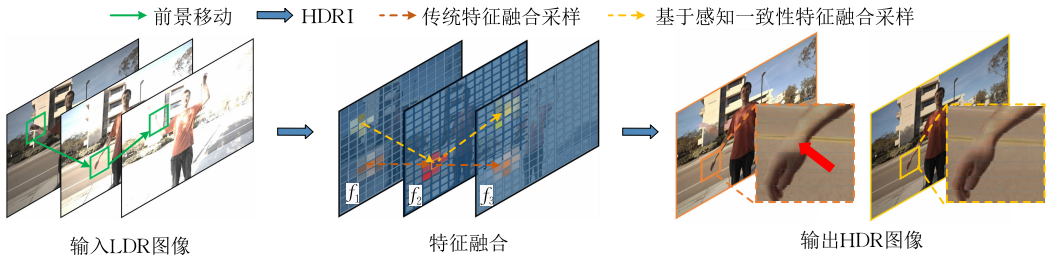


图 1 动机图(传统特征融合方案应用卷积操作从每个图像中的相同位置区域进行学习,忽视了 LDR 图像序列之间丰富的上下文对应关系,导致伪影的生成.本文所提出的一致性感知的特征融合方法(CAFA),使模型能够从具有相同上下文信息的区域进行学习)

目前,已经有一些研究尝试改进上述提到的问题.其中,Yan 等人[20]提出了一种非局部网络,通过应用非局部操作来利用 LDR 图像之间的全局上下文关系.Liu 等人[21]则提出了一个先粗糙后细节的卷积神经网络,通过使用基于可变形非局部卷积的方法来进行特征融合.然而,这些方法并没有明确地对 LDR 图像之间的全局相关性进行建模,导致无法在实际上有效促进细节特征的融合.为克服这一问题,本文研究直接把图像序列中上下文一致性纳入特征融合过程中,并鼓励模型从具有相同上下文信息的区域中学习特征.因此,本文提出了一种新的一致性感知特征融合方案(Coherence-Aware Feature Aggregation,CAFA).在融合过程中,CAFA 首先评估图像特征之间的上下文相关性,其后在卷积过程中对具有相同或相似上下文信息的输入图像进行采样,进而生成新的特征.与以往的方法相比,该方案可以从具有相同纹理信息而不同空间位置的区域进行融合特征,进而更好地形成细粒度特征.如图 1 所示,本文所提出的基于感知一致性特征融合方案不再在每个图像中的同一空间位置区域进行采样,而是在具有一致纹理信息的区域采样.

为了将所提出的 CAFA 方案用于处理动态场景的高动态范围成像,本文进一步提出了一个基于 CAFA 的一致性感知高动态范围成像网络(Coherence-Aware HDR Network,CAHDRNet).具体来说,本文通过设计三个额外的可学习模块来构建 CAHDRNet,以促进 CAFA 发挥最大效果.首先,本文基于在 ImageNet 上预训练的 VGG-19 模型构建了一个可学习的特征提取器,用于从每个图像中提取特征.提取器的参数将在端到端的训练过程中不断更新,从而促进网络更好地学习不同 LDR 图像的特征,为 CAFA 中一致性感知评估和应用打下坚实基础.接着,应用所提出的 CAFA 模块,即通过在各个图像中采样具有相同上下文信息的特征进行融合.在特

征融合之后，本文构建并使用一个多尺度残差补全模块(Multi-Scale Residual Hallucinating，MSRH)模块来处理融合特征. 该模块将在不同扩张率的不同尺度上进行学习，从而生成更强大精准的特征，并在缺失区域中生成可信的细节. 此外，本文还设计了一个软注意力模块，该模块通过学习不同图像区域对于获取所需特征的重要性，在网络解码跳跃连接过程中对融合结果进行细节内容补充. 本文进行了充分的实验来验证所提出的 CAHDRNet 的有效性，实验结果表明本文所提出的方法在 HDRI 任务性能上广泛优于现有最先进的方法. 综上所述，本文的贡献可概括为以下三点：

(1)本文提出了一种针对 HDRI 的一致性感知特征融合(CAFA)方案，该方案明确地将 LDR 图像间的上下文相关性纳入特征融合过程中，以减少图像融合过程中的伪影生成并有效增强生成图像内容.

(2)基于所提出的一致性感知特征融合方案 CAFA，本文进一步提出了一致性感知高动态范围成像网络(CAHDRNet)，通过构建可学习特征提取器、多尺度残差补全、软注意力机制等多个模块，CAHDRNet 可在 HDRI 任务上进行端到端的训练.

(3)本文通过在多种数据集上进行实验，并在所提出的网络的不同结构进行全面比较，充分探索并验证了所提出的 CAHDRNet 网络架构的有效性和高效性.

2　相关工作

高动态范围成像作为底层计算机视觉领域中的一个基础和关键任务，在过去几十年中得到了广泛的研究. 早期的 HDRI 方法[7,11]在融合对齐良好的 LDR 图像上表现优秀，但在处理包含运动物体的复杂场景时会出现严重的伪影和失真. 为了克服这个问题，许多专家学者开始研究复杂场景下的高动态范围成像方法.

2.1　基于偏移像素去除的 HDRI 方法

为了解决 LDR 图像之间的前后景内容偏移而造成的融合伪影问题，最直接的方法就是去除 LDR 图像之间偏移的像素，也就是在融合过程中将偏移像素作为离群值进行排除. 利用基于色差[22-24]、图像梯度[25]、局部熵[26]以及秩最小化[27-28]等不同的偏移检测方法，目前已经涌现出许多 HDRI 方案. 具体来说，Raman 等人[23]提出了一种基于超像素分组的自底向上分割算法来检测偏移，在移除偏移像素后使用像素级图像融合方法进行融合. Jacobs 等人[29]提出了一种基于局部熵差的指标，用于衡量不同 LDR 图像之间的运动偏移. Zhang 等人[25]提出了两种基于梯度的质量度量标准，即可见度和一致性，并通过使用这两种度量标准进行 LDR 图像融合. Lee 等人[27]开发了一种低秩矩阵框架，通过对移动物体进行稀疏性、连通性约束以及对欠曝光和过曝光区域的先验约束，实现了无伪影的高动态范围成像. 然而，这些方法仅使用像素级别的图像特征来检测运动区域，LDR 图像序列中由于曝光度不同，因此简单的基于像素级别的图像特征无法保证其融合过程中像素值移除的可靠性. 此外，进一步地，即使能够实现高精度的运动像素检测，去除这些运动像素也会显著降低用于重建 HDR 图像的内容信息，进而降低图像融合质量.

2.2　基于输入图像对齐的 HDRI 方法

与上述直接去除偏移区域像素的方法不同，基于输入图像对齐的方法尝试先将输入的 LDR 图像进行校正对齐，然后再将它们融合成 HDR 图像. 目前用于 HDRI 的 LDR 图像对齐的方法可分为两类，即基于图像块的方法[30-31]和基于光流[14-15,32]的方法. 在基于图像块的方法中，Sen 等人[30]提出了一种基于图像块的能量最小化公式来合成 HDR 图像. Hu 等人[31]则通过变换域上的亮度和梯度一致性特质来对齐输入图像. 在基于光流的方法中，Bogoni[14]通过光流进行局部运动矢量，进而对齐输入图像. Kang 等人[15]则将 LDR 图像转换到亮度域，然后在亮度域中获取光流. Zimmer 等人[32]通过最小化一个由图像梯度和平滑度组成的能量函数来计算光流. 这些方法在融合过程之前对输入图像进行对齐，有效地提高了 HDRI 的性能. 然而这些方法的局限性在于，他们不仅缺乏对实际样例的学习，同时也忽视了图像融合过程中的优化，因此仍无法有效处理存在大规模前景运动的复杂情况.

2.3　基于深度神经网络的 HDRI 方法

近年来，随着深度神经网络的广泛应用，很多学者也开始研究基于深度学习的 HDRI 方法. 其中一些方法[2,33-38]利用神经网络直接学习 LDR 到 HDR 的映射，以从单个 LDR 图像生成 HDR 图像. 例如，Eilertsen 等人[33]提出了一种编码器-解码器网络，可以直接根据单个 LDR 图像生成 HDR 图像. Endo 等人[34]则先使用输入的单张 LDR 图像生成具有不同曝光的多个 LDR 图像，然后再通过深度网络融合这些图像. 然而，由于单张图像中可用内容的不足，

这些方法通常无法恢复输入图像中严重过曝或欠曝的区域. Prabhakar 等人[10]和 Xu 等人[11]分别提出了可以有效融合静态 LDR 图像的网络架构,但应用于具有前景运动的 LDR 图像时都会产生严重的重影伪影. 对于具有不对齐输入的 HDRI,Kalantari 等人[13]提出首先使用光流对齐 LDR 图像,然后再使用 CNN 融合对齐的 LDR 图像. Wu 等人[16]提出了一种带有深度残差的编码器-解码器框架用于将多个 LDR 图像转换为 HDR 图像. Yan 等人[19]提出了 AHDR 方法,通过应用注意力模块,从 LDR 图像中有选择地提取偏移区域以外的特征,并使用基于膨胀卷积的网络来修补缺失的细节. 后续的工作还采用了双重注意力机制[39]、非局部操作[20]、额外的残差连接[1,18,40]、可变形卷积[21]等方法,以捕获更强大的局部特征.

与上述基于卷积模型的 HDRI 方法不同,一些学者[41-42]尝试使用近年来在自然语言处理和计算机视觉领域大放异彩的 Transformer 模型来解决高动态范围成像问题. 其中,Zhou 等人[41]结合 Transformer 和卷积网络,通过空间注意力模块来抑制伪影的产生. Yan 等人[42]通过设计基于块的内容对齐模块进行运动物体的对齐,接着结合 Transformer 和可变形卷积设计了一个基于窗口的可变形 Transformer 层来融合图像.

尽管这些方法都在一定程度上减少了生成图像中的伪影,但其大多都是基于空间位置关系进行特征融合,最后再使用自注意力机制抑制伪影的产生. 即使有些方法会先对图像进行对齐,但因为在特征融合中对相同位置区域而非相同上下文信息区域进行采样,其过程中形成的伪影很难在后期去除. 由于没有充分考虑到输入图像之间的丰富上下文对应关系,没有打破基于空间位置关系的融合思想,上述方法最终的重建效果并不理想.

2.4 特征融合

特征融合是指将不同尺度或清晰度的特征图进行优化组合的过程,是深度学习工作如图像分割、目标分类中提升性能的重要手段. 传统的特征融合方法通常忽略输入图像之间丰富的上下文对应关系,只使用简单的拼接将特征连接在一起,然后使用卷积进行融合. 其过程可以表示为

$$\boldsymbol{Z}=Conv(\boldsymbol{T}_1\oplus\boldsymbol{T}_2\oplus\boldsymbol{T}_3) \tag{1}$$

其中,$\boldsymbol{Z}$ 表示融合后的特征. 值得注意的是,在每个卷积单元中,它会在每个图像特征中采样相同的特征区域. 具体来说,对于特征 $\boldsymbol{Z}$ 中的第 i 个位置,有

$$\boldsymbol{Z}^i=\sum_{\Delta\in\mathcal{R}}\boldsymbol{w}^i\cdot[\boldsymbol{T}_1^{i+\Delta}\oplus\boldsymbol{T}_2^{i+\Delta}\oplus\boldsymbol{T}_3^{i+\Delta}] \tag{2}$$

其中,Δ 表示网格区域 $\mathcal{R}$ 内的偏移量,$\boldsymbol{w}$ 表示卷积中的学习参数. 如图 2(a)所示,传统特征融合方法针对每个特征的相同区域进行采样.

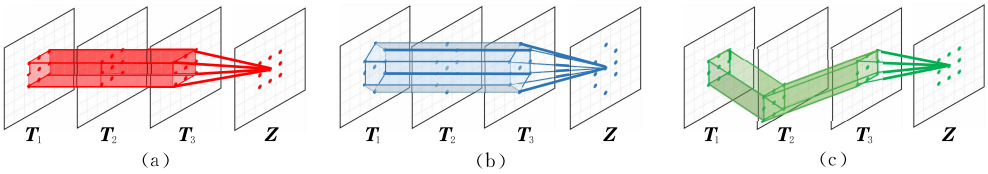


图 2 传统特征融合方法与本文所提出的一致性感知特征融合方法对比示意图((a)基于传统卷积的融合:在输入中采样具有相同位置的网格;(b)基于可变形卷积的融合:在不同输入中采样具有不同感受野但位置仍相同的网格;(c)本文所提出的 CAFA:在融合过程中采样具有相同上下文信息的网格)

与传统卷积操作固定方形的感受野不同,可变形卷积[43]学习偏移量移动采样点,进而灵活调整感受野的形状. 通过拥有不同形状的感受野,可变形卷积可以适应不同形状的特征,更好地处理复杂场景. 最近的工作[21]应用可变形卷积来融合特征,有更好的对齐效果. 可变形卷积可以描述为

$$\boldsymbol{Z}^i=\sum_{\Delta\in\mathcal{R}+\Delta_r}\boldsymbol{w}^i\cdot[\boldsymbol{T}_1^{i+\Delta}\oplus\boldsymbol{T}_2^{i+\Delta}\oplus\boldsymbol{T}_3^{i+\Delta}] \tag{3}$$

其中,Δ_r 表示在卷积过程中学习的偏移量. 可变形卷积的采样过程如图 2(b)所示. 可变形卷积的优势来自于采样过程中不同形状的感受野. 虽然可变形卷积可以处理偏移变化,但它仍然忽略了输入之间的相关性,因此在高动态范围成像任务中仍存在局限性.

3 方 法

给定一系列曝光不同、内容未对齐的 LDR 图像,记为 $\{\boldsymbol{I}_1,\boldsymbol{I}_2,\boldsymbol{I}_3,\cdots,\boldsymbol{I}_n\}$,$\boldsymbol{I}_n\in R^{X\times Y\times 3}$. HDRI 任务首先从 LDR 图像序列中选定一张参考图像 $\boldsymbol{I}_r$,作为融合后图像的内容基础,然后通过 HDRI 方法重建得到 HDR 图像 $\boldsymbol{I}_h$,其内容与参考图像 $\boldsymbol{I}_r$ 一致且曝光良好. 本文参考之前的相关工作,以三张 LDR 图像

融合为例进行方法描述和实验，即 n=3. 同时选择中等曝光的图像 $\boldsymbol{I}_2$ 作为参考图像，即 $\boldsymbol{I}_r=\boldsymbol{I}_2$. 本文重点研究使用卷积网络进行 HDR 图像重建，首先从图像 $\boldsymbol{I}_n$ 中提取特征图 $\boldsymbol{T}_n$，然后融合这些特征生成 HDR 图像 $\boldsymbol{I}_h$. 在下文中，本文首先描述所提出的 CAFA 方案，然后介绍基于 CAFA 方案所构建的端到端 HDR 图像重建网络 CAHDRNet. 为方便表示与理解，相关数学符号定义如表 1 所示.

表 1　本文所使用符号及其含义说明

数学符号	含义
$\boldsymbol{Z}$	融合特征，$\boldsymbol{Z}^i$ 为特征 $\boldsymbol{Z}$ 第 i 个位置的值
$\boldsymbol{T}$	提取的图像特征，$\boldsymbol{T}_j^i$ 表示第 i 张图所提取出的特征在第 j 个位置的值，t 表示由 $\boldsymbol{T}$ 所拆分获得的特征区块
$\boldsymbol{I}$	输入图像，$\boldsymbol{I}^i$ 为第 i 张输入图像，$\boldsymbol{I}^r$ 为参考图像，$\boldsymbol{I}^h$ 为 HDR 图像
$\boldsymbol{r}$	余弦相似度，$\boldsymbol{r}^{i,j}$ 为参考图像中第 i 个特征块与非参考图像第 j 个特征块的相似度
$\boldsymbol{h}$	映射索引，$\boldsymbol{h}^i$ 表示非参考图像中与图像中第 i 个特征块最相关的特征块的位置索引
$\boldsymbol{s}$	置信度，$\boldsymbol{s}^i$ 表示非参考图像中与图像中第 i 个特征块最相关的特征块的置信度

3.1　一致性感知特征融合

为解决传统特征融合方法在 HDRI 任务中的局限性，本文提出了一种明确考虑输入图像之间上下文一致性的方法，称为 CAFA(Coherence-Aware Feature Aggregation). 该方法旨在融合具有上下文相关信息的特征，进而显式利用输入图像之间的纹理相关性，通过采样图像特征中具有相同纹理信息的网格来融合特征. 具体来说，给定特征块 $\boldsymbol{T}_1$ 和 $\boldsymbol{T}_2$ 之间的相关性矩阵 $\boldsymbol{h}_1$，以及特征块 $\boldsymbol{T}_3$ 和 $\boldsymbol{T}_2$ 之间的相关性矩阵 $\boldsymbol{h}_3$，CAFA 通过以下方式融合特征：

$$\boldsymbol{Z}^i=\sum_{\Delta\in\mathcal{R}}\boldsymbol{w}^i\cdot\left[\boldsymbol{T}_1^{h_1^i+\Delta}\oplus\boldsymbol{T}_2^{i+\Delta}\oplus\boldsymbol{T}_3^{h_3^i+\Delta}\right]\tag{4}$$

其中，h_1^i 和 h_3^i 分别表示 $\boldsymbol{h}_1$ 和 $\boldsymbol{h}_3$ 中的第 i 个元素，它记录了在非参考特征中与参考特征中的第 i 个特征块对应的索引. 相关性嵌入的具体方案将在 3.2 节中介绍. 如图 2(c)所示，因为特征之间的采样位置不同，CAFA 可以通过学习不同输入图像之间上下文一致的区域来生成特征. 总的来说，本文所提出的 CAFA 与可变形卷积的区别在于：第一，CAFA 建立了跨输入图像之间的相关性，而可变形卷积只利用了输入图像内部的相关性；第二，CAFA 明确地建模了相关性，而可变形卷积通过学习偏移量来生成偏移.

3.2　一致性感知的高动态范围成像网络

在所提出的 CAFA 特征融合方案基础之上，本文进一步设计了一个支持动态场景的基于一致性感知的高动态范围成像网络(Coherence-Aware HDR Network，CAHDRNet). 具体来说，CAHDRNet 由四个模块组成：(1)用于提取特征的可学习特征提取器模块；(2)用于特征融合的 CAFA 模块；(3)用于识别互补特征的软注意力模块；以及(4)用于填补缺失细节的多尺度残差补全模块. 其网络架构概述如图 3 所示.

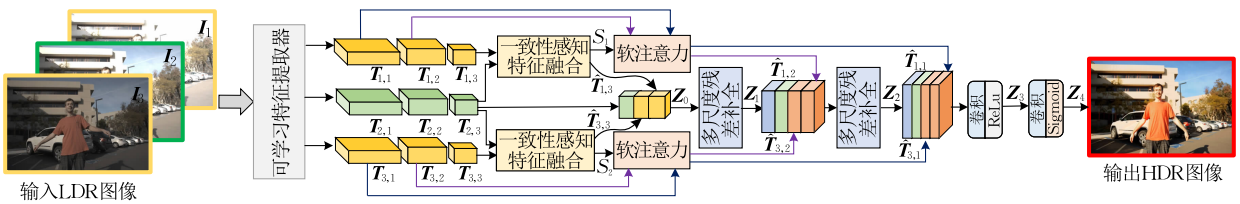


图 3　CAHDRNet 框架示意图(首先，可学习特征提取器将从输入的 LDR 图像序列生成相应特征. 其次，通过本文所提出的基于上下文一致性的特征融合方案(CAFA)对特征进行融合. 接着，使用多尺度残差补全(MSRH)模块将融合后的特征进一步解码得到最终结果图像. 同时，本文使用软注意力模块对 LDR 图像的浅层特征进行识别，获取与融合后特征形成互补的特征，并将这些特征通过跳跃连接机制加入到 MSRH 模块的解码过程中，提高生成图像的质量)

特征提取器. 在 HDRI 任务中，精准的 LDR 图像语义特征将有助于生成无伪影、高质量的高动态范围图像，因此准确地提取 LDR 图像的特征至关重要. 本文不采用直观的卷积操作来提取特征，而是使用预训练于 ImageNet 数据集上的 VGG-19 模型作为骨干网络. 这个网络的参数将在端到端的训练过程中不断学习更新. 这样的设计鼓励了网络可以更好地学习输入的多张 LDR 图像的特征，从而为后续的特征融合提供了坚实的基础. 特征提取的过程可以描述为

$$\boldsymbol{T}_n=\mathcal{E}(\boldsymbol{I}_n)\tag{5}$$

其中，$\mathcal{E}(\cdot)$ 表示可学习的特征提取器，$\boldsymbol{T}_n$ 表示从输入的 LDR 图像 $\boldsymbol{I}_n$ 中提取的特征图. 具体来说，本文从 VGG-19 网络中能够更好为特征精准融合提供基础的中间特征层，即第 5 层、第 9 层和第 13 层中提取特征，分别表示为 $\boldsymbol{T}_{n,k}$，$k\in\{1,2,3\}$.

基于一致性感知的特征融合模块. 在获得图像特征后，本文使用所提出的 CAFA 方案来融合深度

特征 $\boldsymbol{T}_{n,3}$. 如前所述,CAFA 的目的是在卷积过程中采样输入图像之间具有相同或相似上下文信息的特征. 因此,如何计算特征之间的一致性就成为了 CAFA 的核心问题. 本文使用直接相关嵌入(Direct Relevance Embedding,DRE)[44]来评估特征之间的一致性. 其流程如图 4 所示. 首先对非参考图像特征 $\boldsymbol{T}_n$,$n\in 1,3$ 和参考图像特征 $\boldsymbol{T}_r$,$\boldsymbol{T}_r=\boldsymbol{T}_2$ 进行相关嵌入. 具体而言,将 $\boldsymbol{T}_n$ 和 $\boldsymbol{T}_r$ 都展开成图块,分别表示为 $\boldsymbol{t}_n^i$ 和 $\boldsymbol{t}_r^i$,其中 $i\in[1,X\times Y]$. 对于参考图像特征 $\boldsymbol{T}_r$ 中的每个图块 $\boldsymbol{t}_r^i$,计算其与LDR图像特征 $\boldsymbol{T}_n$ 中每个图块 $\boldsymbol{T}_n$ 的余弦相似度:

$$\boldsymbol{r}^{i,j}=\left\langle \frac{\boldsymbol{t}_r^i}{\|\boldsymbol{t}_r^i\|},\ \frac{\boldsymbol{t}_n^j}{\|\boldsymbol{t}_n^j\|}\right\rangle \tag{6}$$

其中,$\boldsymbol{r}^{i,j}$ 表示 $\boldsymbol{T}_r$ 中 $\boldsymbol{t}_r^i$ 和 $\boldsymbol{T}_n$ 中 $\boldsymbol{t}_n^i$ 之间的纹理相关性. 接着,对于参考图像特征 $\boldsymbol{T}_r$ 中的第 i 个图像块,可获取 $\boldsymbol{T}_n$ 中最相关的图块位置索引. 本文使用一个硬注意力映射 H 来记录这个索引,可以计算为

$$h^i=\arg\max_j \boldsymbol{r}^{i,j} \tag{7}$$

其中,h_i 是映射 H 中的第 i 个元素,记录了对于参考图像特征 $\boldsymbol{T}_r$ 中的第 i 个图块最相关的 $\boldsymbol{T}_n$ 中的图块位置索引. 此外,本文还使用一个软注意力映射 S 来记录相关性的置信度,其中元素 s^i,$i\in[1,X\times Y]$ 计算如下:

$$s^i=\max \boldsymbol{r}^{i,j} \tag{8}$$

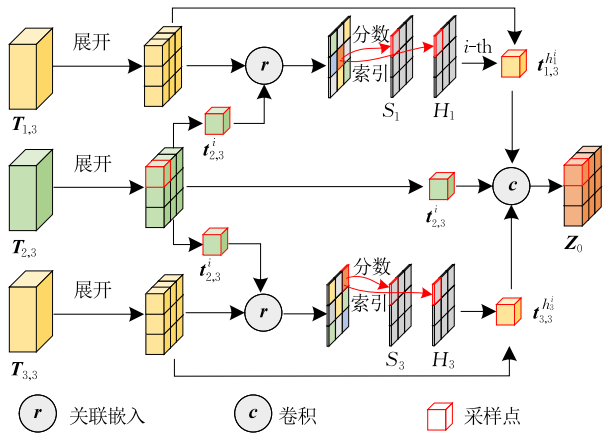
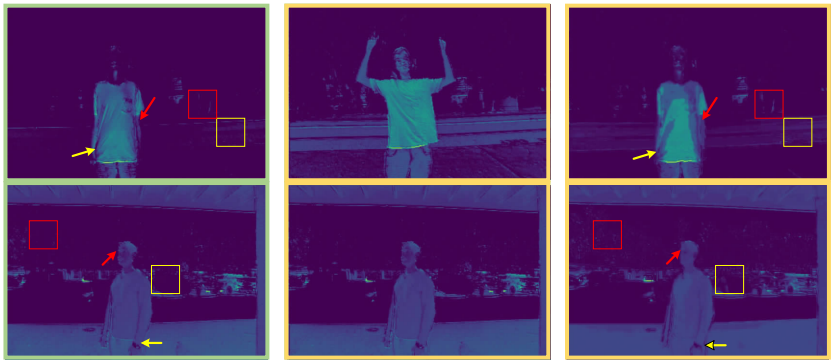


图 4 一致性感知特征融合方案示意图

简单的说,对于 $\boldsymbol{T}_r$ 中的图块 $\boldsymbol{t}_r^i$,根据 H 中的索引 h^i,在 $\boldsymbol{T}_n$ 中找到最相关的图块 $\boldsymbol{T}_n^{h^i}$. 最后,使用式(4)来融合特征.

为使 CAFA 过程更清晰、直观,本文进一步对 CAFA 进行质化分析. 对于输入特征 $\boldsymbol{T}_n$ 和 $\boldsymbol{T}_r$,本文根据 H 中记录的最相关图块将 $\boldsymbol{T}_n$ 中的图像块重新组织为 $\hat{\boldsymbol{T}}_n$. 因此可以直观地看到参考图块 $\boldsymbol{T}_r$ 在融合过程中采样的 $\boldsymbol{T}_n$ 图块. 示例如图 5 所示. 如预期,本文所提出的 CAFA 在融合过程中网络采样了具有相同纹理信息的图块. 因此,通过使用 CAFA 方案,可以有效地抑制内容错位的影响,进而显著提高 HDRI 去伪影性能. 本文将在第 4 节中进行更多的讨论和分析.



(a) 参考图像特征 $\boldsymbol{T}_r$ (b) 非参考图像特征 $\boldsymbol{T}_n$ (c) 非参考图像重构特征 $\hat{\boldsymbol{T}}_n$

图 5 CAFA 的质化分析(为了直观地展示 CAFA,本文根据 $\boldsymbol{T}_r$(a)和 $\boldsymbol{T}_n$(b)之间相关性的采样索引 H,将 $\boldsymbol{T}_n$(b)重新组织为 $\hat{\boldsymbol{T}}_n$(c). 如图所示,在 $\hat{\boldsymbol{T}}_n$(c)中的采样区域在每个位置上都与 $\boldsymbol{T}_r$(a)具有上下文连贯性)

软注意力模块. 由于本文所提出的 CAHDRNet 是一个编码器-解码器网络,在解码过程存在跳跃连接. 为了避免有害特征对融合结果的影响,同时增强特征中与参考图像互补的所需特征,本文提出了一个软注意力模块. 该模块可以识别出浅层特征表示 $\boldsymbol{T}_{n,1}$ 和 $\boldsymbol{T}_{n,2}$ 中的有利区域. 不同于 Yan 等人[19]直接使用 LDR 图像生成注意力图,本文提出的网络使用的是特征融合过程中获得的相关性嵌入来生成软注意力图 S. 这样获取的软注意力图 S 利用了非参考图像和参考图像之间嵌入的纹理相关性. 为了适应不同的特征尺度,本文使用反卷积层来扩大和校准 S. 其架构如图 6 所示. 同时,在进行元素乘法之前,本文使用 Sigmoid 函数将软注意力图归一化到 $[0,1]$ 区间.

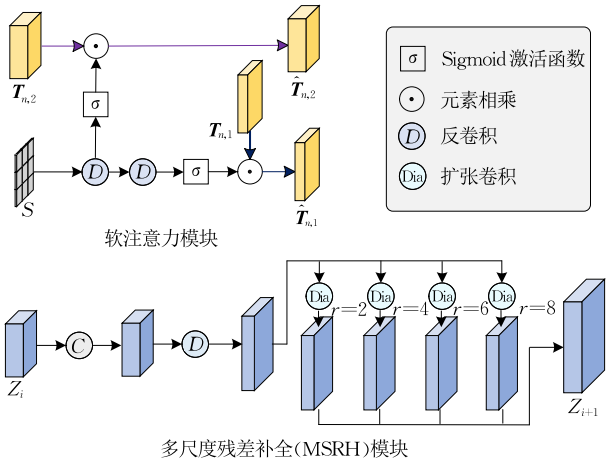


图 6　软注意力模块和多尺度残差补全(MSRH)模块的架构示意图

多尺度残差补全模块. 为了更好地学习 LDR 图像特征并补全内容缺失区域的细节，受文献[19]中密集膨胀残差模块的结构启发，本文在解码过程中引入了膨胀卷积，利用扩大的卷积感受野有效地补全图像细节. 具体而言，本文提出了一个多尺度残差补全(Multi-Scale Residual Hallucinating，MSRH)模块，通过扩大感受野并集成不同膨胀比例来聚合特征. 如图 6 所示，在每个 MSRH 模块中，先使用反卷积层对特征图进行上采样，然后将跳跃连接的结果加入到特征中，接着再使用一个卷积层，此过程可以描述为

$$\boldsymbol{Z}_i'=Conv(DeConv(\boldsymbol{Z}_i')\oplus\hat{\boldsymbol{T}}_{1,k-i}\oplus\boldsymbol{T}_{2,k-i}\oplus\hat{\boldsymbol{T}}_{3,k-i}) \tag{9}$$

接着应用 4 个并行的膨胀卷积层，每个膨胀卷积层具有不同的膨胀率，能够获得更大的感受野，进而更好地补全图像细节. 具体可表示为

$$\begin{aligned}\boldsymbol{Z}_{i+1}=&DiaConv_{r=2}(\boldsymbol{Z}_i')\oplus DiaConv_{r=4}(\boldsymbol{Z}_i')\oplus\\&DiaConv_{r=6}(\boldsymbol{Z}_i')\oplus DiaConv_{r=8}(\boldsymbol{Z}_i')\end{aligned} \tag{10}$$

其中 $DiaConv$ 表示膨胀卷积操作，r 是膨胀率. 在此设计下，网络可以在多个尺度上学习整合跳跃连接的对齐特征，从而获得更强大的特征，保证了 CAHDRNet 能够有效地恢复内容，提高 HDR 重建质量.

3.3　实现细节

图像预处理. 若输入的 LDR 图像不是 RAW 格式，本文首先使用相机响应函数(Camera Response Function，CRF)将图像转换到线性空间，并使其归一化在[0，1]的范围内. 在将 LDR 图像输入到网络之前，首先应用伽马校正编码[16]将输入 LDR 图像 $\boldsymbol{I}_n$ 映射到 HDR 域，生成相应的一组 $\boldsymbol{I}_n'$：

$$\boldsymbol{I}_n'=\frac{\boldsymbol{I}_n^{\gamma}}{e_n} \tag{11}$$

其中 γ 是伽马校正因子，本文中设置为 2.2，e_n 表示 LDR 图像 $\boldsymbol{I}_n$ 的曝光时间. 接着本文遵循文献[16]中的建议，沿着通道维度连接图像 $\boldsymbol{I}_n$ 和 $\boldsymbol{I}_n'$，最后得到一个 6 通道的张量作为网络的输入. 也就是说，本文方法将同时输入 LDR 域和 HDR 域图像. 其中，LDR 域的图像有助于检测过饱和区域，而 HDR 域图像则鼓励检测偏移区域，因此以 6 通道作为网络输入可以最好地提升网络融合质量. 在此过程中，所有的图像预处理操作都是以浮点数的形式进行，同时将参考曝光的范围定义为单位辐射.

损失函数. 与前人工作[19]一致，本文使用经过色调映射的 HDR 图像来进行损失函数的计算. 使用经过色调映射的 HDR 图像来计算损失函数可提高计算效率和稳定性. 通过用经过色调映射的 LDR 图像来计算损失函数，使得像素值范围适合于标准图像处理算法和显示设备，可以将计算限制在 LDR 范围内，从而减少计算资源的需求. 此外，由于 LDR 图像像素值较小，其在数值上更稳定，可以减少由于像素值过大而引起的数值不稳定性和计算误差. 本文采用 μ-law 来压缩 HDR 图像的范围，该过程可表示为

$$\mathcal{T}(\boldsymbol{I}_h)=\frac{\log(1+\mu\boldsymbol{I}_h)}{\log(1+\mu)} \tag{12}$$

其中 μ 代表压缩参数. 按照参考文献[13]，本文设置其参数值为 5000. 本文的损失函数定义如下：

$$\mathcal{L}=\left\|\mathcal{T}(\hat{\boldsymbol{I}}_h)-\mathcal{T}(\boldsymbol{I}_h)\right\|_1 \tag{13}$$

其中 $\hat{\boldsymbol{I}}_h$ 代表网络生成的 HDR 图像，$\boldsymbol{I}_h$ 则代表对应的正解 HDR 图像.

4　实　验

本节介绍本文所提出的 CAHDRNet 的相关实验结果. 首先在 4.1 节中描述实验设置，包括实验中使用的数据集、模型训练细节和量化指标. 接着，在 4.2 节中，本文基于视觉质量和量化指标，将 CAHDRNet 与现有的最先进的 HDRI 方法进行比较. 4.3 节中展示了一系列消融实验结果，以验证所提出的 CAHDRNet 的核心组件的有效性. 为进一步验证本文所提出的 CAFA 方案的特征融合效果，4.4 节中将 CAFA 方案与基于可变形卷积、非局部操作的方案进行比较，进一步证实 CAFA 在 HDRI 任务上的优越性. 4.5 节对不同方法的时间复杂度进行比较. 为了进一步探索所提出方法的可扩展性，4.6 节中展示了基于 CAHDRNet 的扩展应用，并给出了相应的实验结果.

4.1　实验设置

数据集. 本文在 Kalantari 数据集[13]上进行模

型的训练和测试. 其中,训练集包含 74 对数据,每对数据包括三个输入 LDR 图像和一个正解 HDR 图像. 测试集包含 10 对数据,本文对该测试集进行质化和量化实验,并与其他方法进行比较. 为了评估所提出方法的泛化性能,本文还在 Tursun 数据集[45]上进行了额外测试.

模型训练. 本文的实验设备为一台装有 Intel Core i7、64 GB 内存和 NVIDIA GeForce 2080 Ti GPU 的计算机,实验架构基于 PyTorch 深度学习框架完成. 对于给定的训练图像,本文首先进行数据增强操作以避免训练过程中的过拟合现象. 具体来说,将图像裁剪为 512×512 的区块,并通过翻转、旋转及平移等数据增强手段,生成丰富且多样化的训练小块. 在网络训练阶段,本文首先采用 Xavier 方法对权重进行初始化,随后通过学习率为 1×10^{-5} 的 Adam 优化器在 100 轮迭代中优化权重.

量化指标. 实验使用峰值信噪比 *PSNR*(Peak Signal to Noise Ratio)[46]、结构相似性指数 *SSIM*(Structural SIMilarity Index)和 *HDR-VDP-2*[47]指标来评估 HDR 重建结果. 具体来说,本文使用 μ-*law* 作为参数对图像进行色调映射,分别对 Matlab 的 tonemap 函数后的图像以及线性域上的图像进行计算,得到多组 *PSNR* 和 *SSIM* 值,分别记为 *PSNR*-μ 和 *SSIM*-μ、*PSNR*-M和*SSIM*-M以及 *PSNR*-L 和 *SSIM*-L. 对于 Tursun 的数据集[45],由于缺少参考 HDR 图像,实验使用自然度图像质量评估器 *NIQE*(Naturalness Image Quality Evaluator)[48]和盲色调映射质量指数 *BTMQI*(Blind Tonemapping Quality Index)[49]指标进行图像质量评估.

4.2 与现有方法对比

实验将本文所提出的 CAHDRNet 与文献中的五种代表性方法进行了比较. 这些方法分别是:Sen 等人[30]提出的基于图像块的方法、Kalantari 等人[13]基于光流的方法、Wu 等人[16]基于 CNN 的 HDRNet 方法、Yan 等人[19](2019)提出的 AHDRNet 方法、Yan 等人[20](2020)的方法以及 Prabhakar 等人[18]的方法. 为了公平比较,本文使用各自作者发布的模型来获取各自的结果. 其中,对于 Yan 等人[20](2020)的方法,由于目前作者还未发布训练好的模型,因此本文使用其代码默认设置重新训练该模型.

4.2.1 质化评估

现有 HDRI 方法都能在偏移或过曝区域较小的 LDR 图像上获得接近最优的结果,因此在视觉上很难看出各方法之间的差异. 为了进行更好的比较,本文实验在一些具有大量偏移区域或过曝区域的挑战性图像上进行. 而大多数现有算法无法处理这种复杂图像中的伪影问题,因此可以更好地观察到不同方法效果的差异.

在 Kalantari 数据集中的挑战性图像上各方法的质化比较如图 7 和图 8. 图中第一和第二列分别为待融合的 LDR 图像以及输入 CAHDRNet 后获得的色调映射结果. 第三列为从输入 LDR 图像中裁剪的图像块. 第四至第九列分别为不同 HDRI 方法产生的结果. 最后的第十列为正解 HDR 图像. 如图 7 和图 8 的第四列所示,Sen 等人的结果虽然与正解图像对齐良好,但他们的结果存在严重的伪影(如图 7第一行图像块),并且存在过度平滑. Kalantari 等人方法的结果也存在类似的问题,即并不能很好消除伪影(如图 8 中手臂区域中的道路伪影). 这是因为基于像素级别的对齐是不可靠的,尤其是对于具有不同曝光水平的 LDR 图像序列,很容易在最终 HDR 结果中产生伪影. 相比之下,Wu 等人和 Yan 等人(2019)的方法的结果在视觉上更好. 不仅与参考图像对齐良好,并且图像中的细节清晰可见. 但它们在处理具有大块偏移区域(如图 7 中的树木)的图像时仍然会出现伪影. Yan 等人(2020)的方法和 Prabhakar 等人的方法虽然可以有效地纠正运动并抑制伪影,但结果存在模糊、细节丢失和颜色失真的问题. 总的来说,基于 CNN 的方法比基于图像块或光流的方法能产生更好的结果,但是在处理具有大量偏移区域或过曝区域的挑战性图像时,这些方法使用的简单的跳跃连接、注意力或非局部操作不能

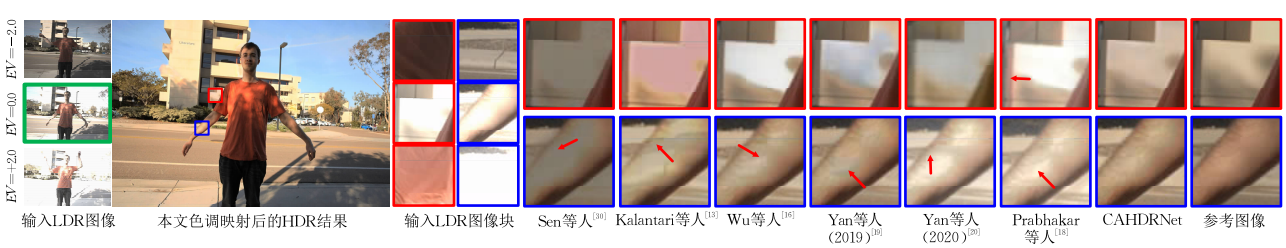


图 7 使用来自 Kalantari 数据集[13]"007"测试图像序列的实验结果比较(左侧部分依次为 LDR 图像、本文方法所得到的色调映射后结果、输入 LDR 图像块. 右侧部分展示了不同方法结果的放大区域比较)

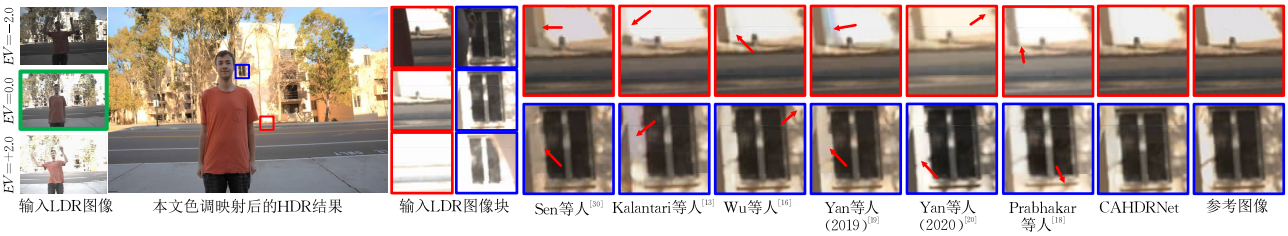


图 8　使用来自 Kalantari 数据集[13]“008”测试图像序列的实验结果比较(左侧部分依次为 LDR 图像、本文方法所得到的色调映射后结果、输入 LDR 图像块. 右侧部分展示了不同方法结果的放大区域比较)

完全消除伪影.

与以上对比方法相比,本文所提出的方法就能够在处理这些挑战性图像时产生高质量的无伪影的结果. 无论是存在摄像机移位(如图 7 第一行图像块)、大曝光区域(如图 7 第二行图像块),还是大型物体运动(如图 8 第一行图像块)、包含过曝内容的物体运动(如图 8 第二行图像块),本文所提出的方法都能够产生无可见伪影、细节丰富的结果. 值得注意的是,相比其他方法,本文方法生成的结果在语义上也更加合理.

在图 9 和图 10 中,本文在没有参考图像的 Tursun 数据集上测试了包括本文提出方法在内的各种 HDRI 方法. 可以看出,Sen 等人的方法会产生具有严重重影伪影的过饱和图像块. 通过观察图 9 和图 10 的第五列的结果,可以发现 Kalantari 等人的结果虽然能够减少伪影,但产生的图像存在区域. 在基于 CNN 的方法中,Wu 等人的方法会产生黑色区域,而 Yan 等人(2020)的方法的结果则会导致图像过度饱和与过度平滑. Yan 等人(2019)AHDRNet 方法和 Prabhakar 等人的方法虽然能够生成更美观的结果,但其结果中仍然存在细节损失和颜色失真的问题. 作为对比,如图 9 和图 10 的最后一列所展示的本文方法的结果避免了过度饱和及颜色失真,保持了图像原始的颜色和细节.

图 9　使用来自 Tursun 等人[45]的数据集中“Fire”测试图像序列的质化比较(左侧部分依次为 LDR 图像、本文的色调映射后结果和 LDR 图像块. 右侧部分展示了不同方法结果的放大区域比较)

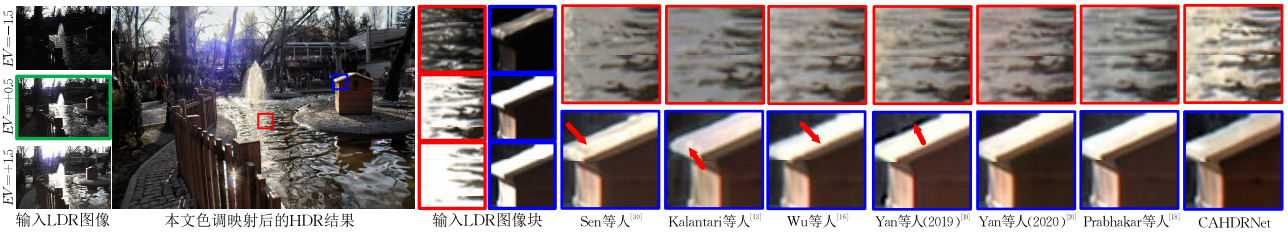


图 10　使用来自 Tursun 等人[45]的数据集中“Lake”测试图像序列的质化比较(左侧部分依次为 LDR 图像、本文的色调映射后结果和 LDR 图像块. 右侧部分展示了不同方法结果的放大区域比较)

4.2.2　量化评估

为了更好评估各种方法的结果质量,本文使用 *PSNR*、*SSIM* 和 *HDR-VDP*-2 等指标进行了量化比较,比较结果如表 2 所示. 从结果可以看出,Sen 等人和 Kalantari 等人的方法效果较差,其 *PSNR*、*SSIM* 和 *HDR-VDP*-2 值较低. 这主要是因为它们所使用的像素级的对准会引起伪影. Wu 等人的结果在 *HDR-VDP*-2 值上优于 Sen 等人和 Kalantari 等人,但在 *PSNR* 和 *SSIM* 值上表现较差. Yan 等人(2019)的方法通过使用注意力模块避免了伪影,提高了 *PSNR*、*SSIM* 和 *HDR-VDP*-2 值. Yan 等人(2020)的方法由于图像模糊和过度平滑的问题(如图 7 和图 8 所示),在 *PSNR*、*SSIM* 和 *HDR-VDP*-2 值上表现较低. 作为对比,本文所提出的方法克服了上述方法的缺陷,在 *PSNR*、*SSIM* 和 *HDR-VDP*-2 等指标上均优于现有技术水平.

实验进一步使用 *BTMQI* 和 *NIQE* 评估各个方法在无参考的 Tursun 数据集上的泛化性能. 其中,*BTMQI* 和 *NIQE* 值越低表示图像质量越好. 因为 Yan 等人(2020)的方法存在过曝问题,生成的图像具有较差的自然性和结构性,反映在指标上就是该方法结果具有最高的 *BTMQI* 和 *NIQE* 值(从表 2 的最后一列可以看出),这些问题也能在质化结果图 9 和图 10 中看出. Sen 等人和 Kalantari 等人的方法由于去除伪影的效果有限,也有较高的 *BTMQI* 值. 但有趣的是,因为像素级处理可以更好地保留纹理信息,这些方法在 *NIQE* 值方面表现更好. Wu 等人的方法改善了结果中的伪影问题,具有较低的 *BTMQI* 值. Yan 等人的方法通过采用注意力机制提高了去除伪影的效果,具有较低的 *BTMQI* 值,但结果仍然存在细节损失和色彩失真的问题,因此具有较高的 *NIQE* 值. 作为对比,本文方法的结果在 *BTMQI* 和 *NIQE* 值方面均取得了最低值,优于其他 5 种方法.

表 2 本文方法与现有方法的量化比较结果(加粗数字表示最好的结果)

方法	Kalantari 数据集[13]							Tursun 数据集[45]	
	PSNR-μ ↑	*PSNR-M* ↑	*PSNR-L* ↑	*SSIM-μ* ↑	*SSIM-M* ↑	*SSIM-L* ↑	*HDR-VDP-2* ↑	*BTMQI* ↓	*NIQE* ↓
Sen 等人[30]	35.6275	30.1273	34.8588	0.9798	0.9631	0.9634	50.8771	5.9428	3.6432
Kalantari 等人[13]	39.4773	36.9541	40.1358	0.9813	0.9800	0.9846	62.0245	5.8091	4.1064
Wu 等人(HDRNet)[16]	41.6565	37.5783	40.8854	0.9878	0.9812	**0.9865**	60.4955	5.4886	4.1907
Yan 等人(2019, AHDRNet)[19]	42.2276	38.9626	40.8572	**0.9899**	0.9831	0.9856	62.3119	5.4401	5.0251
Yan 等人(2020)[20]	34.9230	30.1264	38.3637	0.9765	0.9705	0.9793	61.1673	6.1211	6.2512
Prabhakar 等人[10]	40.5210	36.5947	41.0346	0.9771	0.9702	0.9865	62.0005	5.2761	4.0658
本文方法	**42.9120**	**39.3412**	**41.2798**	0.9863	**0.9815**	0.9775	**63.9274**	**5.0252**	**3.3040**

4.3 消融实验

为了充分验证本文所提出的 CAHDRNet 中各个模块(包括可学习特征提取器(LFE)模块、CAFA 模块、跳跃连接(SC)模块、软注意力(SA)模块和多尺度残差补全(MSRH)模块的有效性,在本章节中,本文设置以下几种 CAHDRNet 变体来进行实验比较:

U-Net. U-Net 是图像处理中的一种基准网络. 实验构建了一个 U-Net 结构的编码器-解码器框架作为基准网络,其中编码器是不可训练的特征提取器,而解码器由四个反卷积层组成.

MSRH-Net. 为了评估 MSRH 模块的有效性,实验用 MSRH 模块替换了 U-Net 结构中的解码器部分,命名为 MSRH-Net.

LFE-MSRH-Net. 为了评估可学习特征提取器(LFE)模块的有效性,实验将 MSRH-Net 中的特征提取器改为可学习的特征提取器模块,命名为 LFE-MSRH-Net 模块.

CA-MSRH-Net. CAFA 方案是本文的最重要贡献之一,为了评估该方法的有效性,实验在 LFE-MSRH-Net 的基础上引入了 CAFA 模块,命名为 CA-MSRH-Net 模块. 注意,该模块可以理解为是一个没有软注意力跳跃连接的 CAHDRNet 网络.

SC-MSRH-Net. 为了评估软注意力模块的有效性,实验设计了 SC-MSRH-Net. 该网络没有使用软注意力模块,而是直接在 CA-MSRH-Net 上添加了简单的跳跃连接得到. 通过和使用软注意力模块的 CAHDRNet 进行对比,可评估软注意力模块的有效性.

(1) 多尺度残差补全模块. 为了验证本文所提出的多尺度残差补全模块的有效性,本文将 U-Net 和 MSRH-Net 进行实验比较. 如图 11 中的第一列和第二列所示,添加了 MSRH 模块后的 MSRH-Net

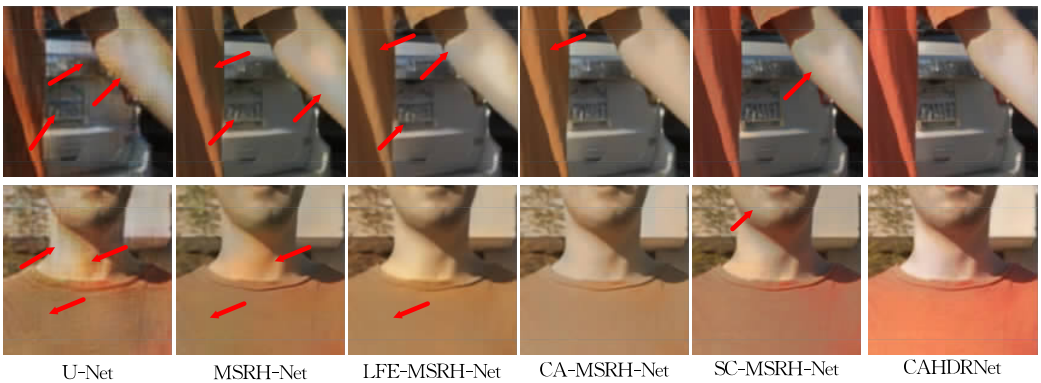


图 11 不同 CAHDRNet 变体所得到结果的视觉质量比较示意图

有效地减轻了伪影并恢复了细节. 表 3 中的量化比较也显示，MSRH-Net 的性能优于 U-Net.

表 3　不同 CAHDRNet 变体的量化比较
（粗体表示最好的结果）

网络结构	*PSNR*-L ↑	*SSIM*-L ↑	*HDR-VDP*-2 ↑
U-Net	29.2127	0.9306	48.2036
MSRH-Net	32.2042	0.9428	53.4659
LFE-MSRH-Net	32.2417	0.9478	55.2365
CA-MSRH-Net	36.3721	0.9726	58.6548
SC-MSRH-Net	38.9949	0.9860	60.7726
CAHDRNet	**41.2798**	**0.9881**	**63.9274**

(2) 可学习的特征提取模块. 如图 11 的第二列和第三列所示，相比于特征提取器不可训练的 MSRH-Net，LFE-MSRH-Net 在去除伪影方面的效果有显著提升. 同时表 3 中的量化指标表明，LFE-MSRH-Net 的性能优于 MSRH-Net.

(3) 一致性感知特征融合模块. CAFA 模块是本文 CAHDRNet 中的关键机制. 如图 11 所示，相对于 MSRH-Net 和 LFE-MSRH-Net，通过在特征融合过程中建模 LDR 图像间的上下文一致性，CA-MSRH-Net 可以更有效地消除伪影. 同时从图 11 的第四列中可以看出，本文所提出的 CAFA 方法不会因内容偏移产生伪影. 有关 CAFA 对于特征采样的直观可视化结果，请参见 3.2 节中的图 5. 此外，根据表 3 的量化比较结果，可以观察到 CA-MSRH-Net 取得了比 LFE-MSRH-Net 更高的分数，这说明 CAFA 可以有效提升 HDRI 中特征融合效果.

(4) 软注意力模块. 虽然 CA-MSRH-Net 也可以消除伪影，但是它容易导致褪色和纹理模糊（例如图 11 中第四列中 T 恤和字符的颜色）. 本文在 CA-MSRH-Net 的基础上添加跳跃连接得到 SC-MSRH-Net，从对比结果可以看出，跳跃连接可以有效地校正颜色并修复纹理. 但是，简单的跳跃连接也会导致结果中出现伪影，如图 11 的第五列所示. 因此，本文提出使用一个软注意力模块来加强跳跃连接，也就是本文的最终网络 CAHDRNet. 通过软注意力模块，本文在保留跳跃连接优点的同时减少伪影的产生. 如表 3 中的量化结果所示，本文所提出的 CAHDRNet 具有比其他模型更好的性能.

4.4　CAFA 与可变形卷积、非局部操作的比较

为验证本文所提出的 CAFA 方案的有效性，本小节将 CAFA 与目前两种前沿的融合方法可变形卷积（Deformable Convolution）和非局部操作（Non-local operation）进行对比. 具体来说，本章节实验构建了两个变体网络，分别用可变形卷积和非局部操作替换了 CAFA，并在包含大面积偏移内容的图像上进行了测试. 图 12 中展示了质化比较结果，表 4 中展示了量化比较结果. 从图 12 所示的结果中可以观察到，在存在较大偏移区域和过曝区域的挑战性场景中，本文所提出的 CAFA 可以更好地消除伪影并维持纹理细节. 表 4 中量化指标的显著提升也表明，相较于可变形卷积和非局部操作，CAFA 方案在处理 HDRI 任务方面具有更好的性能.

一致性感知融合　　非局部操作融合　　可变形卷积融合

图 12　使用不同特征融合方案的质化比较示意图（图像块来自 Kalantari 数据集[13] 中最具挑战性的场景）

表 4　不同特征融合方案的量化比较

特征融合方法	*PSNR*-L ↑	*SSIM*-L ↑	*HDR-VDP*-2 ↑
可变形卷积融合	33.4483	0.9687	59.5440
非局部操作融合	37.0070	0.9775	60.7682
一致性感知融合	**37.4696**	**0.9792**	**62.1308**

注：粗体表示最好的结果.

4.5　时间复杂度

表 5 以参数数量和每张图像的平均处理时间为度量标准，对比不同方法的实现效率. 为公平比较，实验将 Kalantari 数据集[13] 中的测试图像切割为 512×512 图像块，并在 CPU 环境下对各个方法进行对比. 如表 5 所示，Sen 等人的方法处理一张图像的时间约为 75 s，其中最多的时间都用于基于补丁的重建. Kalantari 的方法参数量最少，运行速度更快. Yan 等人（2019）的方法只采用简单的网络和简单的注意力模块，因此参数数量较小，效率较高. Wu 等人和 Yan 等人（2020）的方法更为复杂，参数数量

更大,运行时间更长.相比之下,本文所提出的方法不仅取得了优越的结果,同时在时间消耗和参数数量方面也与其他方法保持在一个量级.本文方法的主要时间消耗几种在于输入对齐的部分,也就是执行语义对齐操作以纠正图像特征过程.虽然本文方法处理时间并不是最优,但可以通过在多个处理单元之间进行分布式推理来提高运行速度.

表 5 本文方法与现有方法的量化比较结果

方法	时间/s	参数量/M
Sen 等人[30]	75.44	—
Kalantari 等人[13]	0.46	0.38
Wu 等人[16]	5.14	33.21
Yan 等人(2019)[19]	1.84	1.50
Yan 等人(2020)[20]	2.40	31.53
本文方法	**3.90**	**21.13**

注:粗体表示最好的.

4.6 针对更多输入的扩展应用

为了进一步评估所提出网络的可扩展性,本节将探究网络在融合超过 3 张输入 LDR 图像时的效果.在实验过程中,对于多个 LDR 输入,首先指定其中一个输入图像作为参考图像,然后进行语义对齐以对齐非参考图像的特征,最后将这些图像的所有特征进行融合并解码得到融合图像.实验选择 Sen 数据集[30]的训练数据集作为输入,因为它包含多个 LDR 图像和 HDR 正解图像.本文在 Sen 和 Tursun 数据集上测试了扩展后的网络.质化结果如图 13 所示,本文构造的扩展网络生成的 HDR 图像与参考图像对齐,并具有清晰的细节.此外,值得注意的是,由于更多的输入图像包含了更多的可用信息,图 13 右图中生成的 HDR 图像比图 10 中的 HDR 图像具有更好的视觉质量,结果也更加清晰.

(a) (b)

图 13 使用来自 Tursun 等人[45]的数据集中"Lake"测试图像序列的质化比较(左侧部分显示了 LDR 图像、本文的色调映射后结果和 LDR 图像块.右侧部分展示了不同方法结果的放大区域比较)

5 总 结

本文提出了一致性感知高动态范围成像网络(CAHDRNet)来解决动态场景的高动态范围成像问题.具体而言,本文设计了一种新颖的一致性感知特征融合方案,通过在 LDR 图像间采样具有相同上下文信息的网格来提取特征进行融合,进而显式地建模输入 LDR 图像之间的对应关系.此外,CAHDRNet 还设计了三个附加模块(可学习特征提取器、软注意力模块以及多尺度残差补全模块)以进一步提高图像融合质量.本文通过实验在多种数据集上证明了本文所提出的 CAHDRNet 广泛优于最先进的 HDRI 方法,显著提升了去除伪影的效果.

本文局限性及展望. 本文的核心思想在于改变特征融合时的采样位置,通过采样不同输入中具有相同上下文的区域进行特征融合.由于拆分的特征块都是固定大小,因此本文方法的感受野同传统卷积一样,为固定矩形状局部区域.因此,在采样点个数及具体局部位置可进行进一步研究,将相关采样点进一步细粒化.此外,结合模型可以看出,本文计算上下文一致性的主要对象是不同的特征之间,这虽然更好地利用了不同曝光程度下的信息,但参考图像内部本身存在的短距离依赖关系的权重将被减弱.因此,本文下一步的研究方向将侧重于:(1)考虑使用 SwinTransformer[50]等窗口可变自注意力机制让模型能够更好地处理参考图像内部的信息从而衍生并完整图像内容;(2)结合可变形卷积中感受野调节机制,进一步对特征融合过程中的感受野进行可变调节;(3)精简网络架构,结合轻量级网络如 SBCFormer[51]实现对特征提取及融合的高效处理.

参 考 文 献

[1] Niu Yuzhen, Wu Jianbin, Liu Wenxi, et al. HDR-GAN: HDR image reconstruction from multi-exposed LDR images with large motions. IEEE Transactions on Image Processing, 2021, 30: 3885-3896

[2] Kim J, Lee S, Kang S J. End-to-end differentiable learning

to HDR image synthesis for multi-exposure images//Proceedings of the AAAI Conference on Artificial Intelligence. Online, 2021: 1780-1788

[3] Santos M S, Ren T I, Kalantari N K. Single image HDR reconstruction using a CNN with masked features and perceptual loss. ACM Transactions on Graphics, 2020, 39(4): 80:1-80:10

[4] Kim S Y, Oh J, Kim M. JSI-GAN: GAN-based joint super-resolution and inverse tone-mapping with pixel-wise task-specific filters for UHD HDR video//Proceedings of the AAAI Conference on Artificial Intelligence. New York, USA, 2020: 11287-11295

[5] Chen Xiangyu, Zhang Zhengwen, Ren Jimmy S, et al. A new journey from SDRTV to HDRTV//Proceedings of the IEEE/CVF International Conference on Computer Vision. Online, 2021: 4500-4509

[6] Anderson E F, McLoughlin L. Critters in the classroom: A 3D computer-game-like tool for teaching programming to computer animation students//Proceedings of the ACM SIGGRAPH 2007 Educators Program. San Diego, USA, 2007: 7-es

[7] Mertens T, Kautz J, van Reeth F. Exposure fusion//Proceedings of the 15th Pacific Conference on Computer Graphics and Application. Hawaii, USA, 2007: 382-390

[8] Kou F, Li Z, Wen C, et al. Multi-scale exposure fusion via gradient domain guided image filtering//Proceedings of the 2017 IEEE International Conference on Multimedia and Expo. Hong Kong, China, 2017: 1105-1110

[9] Li Zhengguo, Wei Zhe, Wen Changyun, et al. Detail-enhanced multi-scale exposure fusion. IEEE Transactions on Image Processing, 2017, 26(3): 1243-1252

[10] Prabhakar K R, Srikar V S, Babu R V. DeepFuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy, 2017: 4714-4722

[11] Xu Han, Ma Jiayi, Zhang Xiao-Ping. MEF-GAN: Multi-exposure image fusion via generative adversarial networks. IEEE Transactions on Image Processing, 2020, 29: 7203-7216

[12] Zhu Minfeng, Pan Pingbo, Chen Wei, et al. EEMEFN: Low-light image enhancement via edge-enhanced multi-exposure fusion network//Proceedings of the AAAI Conference on Artificial Intelligence. New York, USA, 2020, 34(7): 13106-13113

[13] Kalantari N K, Ramamoorthi R. Deep high dynamic range imaging of dynamic scenes. ACM Transactions on Graphics, 2017, 36(4): 144:1-144:12

[14] Bogoni L. Extending dynamic range of monochrome and color images through fusion//Proceedings of the 15th International Conference on Pattern Recognition. Barcelona, Spain, 2000, 3: 7-12

[15] Kang S B, Uyttendaele M, Winder S, et al. High dynamic range video. ACM Transactions on Graphics, 2003, 22(3): 319-325

[16] Wu Shangzhe, Xu Jiarui, Tai Yu-Wing, et al. Deep high dynamic range imaging with large foreground motions//Proceedings of the European Conference on Computer Vision. Munich, Germany, 2018: 117-132

[17] Yan Q, Gong D, Zhang P, et al. Multi-scale dense networks for deep high dynamic range imaging//Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision. Hawaii, USA, 2019: 41-50

[18] Prabhakar K R, Senthil G, Agrawal S, et al. Labeled from unlabeled: Exploiting unlabeled data for few-shot deep HDR deghosting//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Online, 2021: 4875-4885

[19] Yan Qingsen, Gong Dong, Shi Qinfeng, et al. Attention-guided network for ghost-free high dynamic range imaging//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019: 1751-1760

[20] Yan Qingsen, Zhang Lei, Liu Yu, et al. Deep HDR imaging via a non-local network. IEEE Transactions on Image Processing, 2020, 29: 4308-4322

[21] Liu Zhen, Lin Wenjie, Li Xinpeng, et al. ADNet: Attention-guided deformable convolutional network for high dynamic range imaging//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Online, 2021: 463-470

[22] Gallo O, Gelfandz N, Chen Wei-Chao, et al. Artifact-free high dynamic range imaging//Proceedings of the 2009 IEEE International Conference on Computational Photography. Kyoto, Japan, 2009: 1-7

[23] Raman S, Chaudhuri S. Reconstruction of high contrast images for dynamic scenes. The Visual Computer, 2011, 27: 1099-1114

[24] Grosch T. Fast and robust high dynamic range image generation with camera and object movement//Proceedings of the Vision, Modeling and Visualization. Aachen, Germany, 2006: 277-284

[25] Zhang Wei, Cham Wai-Kuen. Gradient-directed multiexposure composition. IEEE Transactions on Image Processing, 2011, 21(4): 2318-2323

[26] Heo Y S, Lee K M, Lee S U, et al. Ghost-free high dynamic range imaging//Proceedings of the Asian Conference on Computer Vision. Queenstown, New Zealand, 2010: 486-500

[27] Lee C, Li Yuelong, Monga V. Ghost-free high dynamic range imaging via rank minimization. IEEE Signal Processing Letters, 2014, 21(9): 1045-1049

[28] Oh T H, Lee J Y, Tai Y W, et al. Robust high dynamic range imaging by rank minimization. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 37(6): 1219-1232

[29] Jacobs K, Loscos C, Ward G. Automatic high-dynamic range image generation for dynamic scenes. IEEE Computer Graphics and Applications, 2008, 28(2): 84-93

[30] Sen P, Kalantari N K, Yaesoubi M, et al. Robust patch-based HDR reconstruction of dynamic scenes. ACM Transactions on Graphics, 2012, 31(6): 203:1-203:11

[31] Hu Jun, Gallo O, Pulli K, et al. HDR deghosting: How to deal with saturation?//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Portland, USA, 2013: 1163-1170

[32] Zimmer H, Bruhn A, Weickert J. Freehand HDR imaging of moving scenes with simultaneous resolution enhancement. Computer Graphics, 2011, 30(2): 405-414

[33] Eilertsen G, Kronander J, Denes G, et al. HDR image reconstruction from a single exposure using deep CNNs. ACM Transactions on Graphics, 2017, 36(6): 1-15

[34] Endo Y, Kanamori Y, Mitani J. Deep reverse tone mapping. ACM Transactions on Graphics, 2017, 36(6): 177:1-177:10

[35] Zhang Jinsong, Lalonde J F. Learning high dynamic range from outdoor panoramas//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy, 2017: 4519-4528

[36] Yang Xin, Xu Ke, Song Yibing, et al. Image correction via deep reciprocating HDR transformation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 1798-1807

[37] Chen Xiangyu, Liu Yihao, Zhang Zhengwen, et al. HDRUNet: Single image HDR reconstruction with denoising and dequantization//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Online, 2021: 354-363

[38] Robidoux N, Capel L E G, Seo D, et al. End-to-end high dynamic range camera pipeline optimization//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Online, 2021: 6297-6307

[39] Ye Qian, Xiao Jun, Lam K, et al. Progressive and selective fusion network for high dynamic range imaging//Proceedings of the 29th ACM International Conference on Multimedia. Chengdu, China, 2021: 5290-5297

[40] Xiong Pengfei, Chen Yu. Hierarchical fusion for practical ghost-free high dynamic range imaging//Proceedings of the 29th ACM International Conference on Multimedia. Chengdu, China, 2021: 4025-4033

[41] Zhou Fangfang, Fu Zhengming, Zhang Dan. High dynamic range imaging with context-aware transformer//Proceedings of the 2023 International Joint Conference on Neural Networks. Queensland, Australia, 2023: 1-8

[42] Yan Qingsen, Chen Weiye, Zhang Song, et al. A unified HDR imaging method with pixel and patch level//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada, 2023: 22211-22220

[43] Dai Jifeng, Qi Haozhi, Xiong Yuwen, et al. Deformable convolutional networks//Proceedings of the IEEE International Conference on Computer Vision. Honolulu, USA, 2017: 764-773

[44] Yang Fuzhi, Yang Huan, Fu Jianlong, et al. Learning texture transformer network for image super-resolution//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2020: 5791-5800

[45] Tursun O T, Akyüz A O, Erdem A, et al. An objective deghosting quality metric for HDR images. Computer Graphics Forum, 2016, 35(2): 139-152

[46] Yin Jia-Li, Chen Bo-Hao, Li Ying. Highly accurate image reconstruction for multimodal noise suppression using semisupervised learning on big data. IEEE Transactions on Multimedia, 2018, 20(11): 3045-3056

[47] 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. ACM Transactions on Graphics, 2011, 30(4): 1-14

[48] Mittal A, Soundararajan R, Bovik A C. Making a 'completely blind' image quality analyzer. IEEE Signal Processing Letters, 2012, 20(3): 209-212

[49] Gu K, Wang S, Zhai G, et al. Blind quality assessment of tone-mapped images via analysis of information, naturalness, and structure. IEEE Transactions on Multimedia, 2016, 18(3): 432-443

[50] Liu Ze, Lin Yutong, Cao Yue, et al. Swin transformer: Hierarchical vision transformer using shifted windows//Proceedings of the IEEE/CVF International Conference on Computer Vision. Online, 2021: 10012-10022

[51] Lu X, Suganuma M, Okatani T. SBCFormer: Lightweight network capable of full-size ImageNet classification at 1 FPS on single board computers//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Hawaii, USA, 2024: 1123-1133

YIN Jia-Li, Ph. D., professor. Her research interests include computational photography, and robustness of deep neural networks.

HAN Jin, M. S. His research interest is low-level image processing.

CHEN Bin, Ph. D. candidate. His main research interests are adversarial attack and defense, and computer vision.

LIU Xi-Meng, Ph. D., professor. His research interests include big data storage and processing, and data security.

Background

High Dynamic Range Imaging (HDRI) aims to recover a High Dynamic Range (HDR) image by merging stacked Low Dynamic Range (LDR) images with different exposure levels, which can benefit various multimedia applications. With decades of studies, numerous promising approaches have been proposed and achieved remarkable progress, especially the deep learning based approaches, in which the LDR images are first mapped into the feature space and the HDR image is obtained by fusing the stacked high-level features.

Although impressive improvements have been achieved, most of these networks strive to merge the features by stacked convolutional operations without paying much attention to the rich textural coherence across the LDR images. Primarily, feature aggregation is widely used in CNN-based HDRI models to fuse feature representations from multiple LDR images, while almost all the existing HDRI networks only use simple feature concatenation to put the features together and then apply stacked convolution modules with local and fixed respective fields to form more complex features. On one hand, the simple concatenation makes rich contextual dependencies among the LDR images ignored. On the other hand, the aggregated features are generated by learning from position-corresponded regions in each image feature representation, since the objects are moving, the position-corresponded areas in each image always represent different contextual information, thus bringing in ghosting artifacts in the final result.

Our motivation is to directly incorporate contextual coherence into the feature aggregation and encourage it to learn features from regions sharing the same contextual information across input image features. We introduce a new Coherence-Aware Feature Aggregation (CAFA) scheme to enable such coherence-aware feature learning. During the aggregation process, we first evaluate the contextual correlations between image features and then generate new features by sampling grids with the same or similar contextual information across input images during convolutions. Our CAFA draws closer to the respective fields with the same textural information across LDR images so that the aggregated features are expected to be explicitly fine-grained formed.

Furthermore, we propose a Coherence-Aware HDR Network (CAHDRNet) incorporated with CAFA for HDRI of dynamic scenes. To facilitate the incorporation of CAFA, we construct our CAHDRNet by designing three additional learnable modules, including a learnable feature extractor for creating the solid foundation for applying the coherence evaluation, a Multi-Scale Residual Hallucinating (MSRH) module for processing the aggregated features, and a soft attention module for enhancing the desired features which are complementary to the reference image during skip connection. Extensive experiments are conducted to validate the effectiveness of our proposed CAHDRNet, where our method demonstrates superior performance over state-of-the-art (SOTA) methods.

This paper was supported in part by the National Natural Science Foundation of China under Grant Nos. 62202104, 62102422, 62072109, and U1804263; and in part by the Natural Science Foundation of Fujian Province under Grant Nos. 2021J06013 and 2021J05129.