动态场景下深度自监督多曝光图像融合方法

张雨童[1],邓 欣[2*],徐 迈[1]
(1. 北京航空航天大学电子信息工程学院,北京 100191;2. 北京航空航天大学网络空间安全学院,北京 100191)

摘 要: 近年来,面向动态场景的多曝光图像融合技术取得重大进展. 其中,基于深度学习的方法在视觉效果和运算效率上都远超传统算法,成为高动态范围成像技术的主流. 然而,现有基于深度学习的融合方法都以有监督学习的方式实现,过度依赖真值图像,难以被广泛应用于实际场景中. 本文提出了一个基于深度自监督学习的动态多曝光图像融合网络,主要贡献包括:设计自监督的动态多曝光融合网络框架,探索高动态范围图像与低动态范围图像序列的内在关联;提出基于注意力机制的全局去伪影模块,使用全局文本模块减少动态融合产生的运动伪影,增强图像细节;提出融合重建模块,通过残差和稠密连接实现多层次特征之间的信息流动;设计运动掩膜引导的自监督损失函数,用于网络的高效训练. 实验表明,与现有方法相比,本文提出的方法在高动态范围图像重建的主观和客观质量上均表现较好,运算效率显著提升.

关键词: 高动态范围成像;多曝光图像融合;深度学习;自监督学习
基金项目: 国家自然科学基金(No.62001016)
中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112(2024)01-0264-10
电子学报 URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20220893

Deep Self-Supervised Multi-Exposure Image Fusion for Dynamic Scenes

ZHANG Yu-tong[1], DENG Xin[2*], XU Mai[1]
(1. *School of Electronic and Information Engineering*, *Beihang University*, *Beijing* 100191, *China*;
2. *School of Cyber Science and Technology*, *Beihang University*, *Beijing* 100191, *China*)

Abstract: In recent years, significant progress has been made in multi-exposure image fusion in dynamic scenes. In particular, the deep learning based methods have shown great visual performance in dynamic multi-exposure image fusion, which have become the mainstream methods in high dynamic range (HDR) imaging. However, the current deep learning based methods are mostly implemented in a supervised manner, which heavily rely on the ground-truth images. That makes it difficult for them to work in real scenes. In this paper, we propose a self-supervised multi-exposure image fusion network for dynamic scenes. The main contributions of this paper are as follows: we design a self-supervised fusion network to explore the latent relationship between HDR and low dynamic range (LDR) images; we propose an attention mechanism based global deghosting module, to reduce the ghosting artifacts caused by moving objects; we propose a merging reconstruction module with residual and dense connections, to improve the reconstruction details; we design a motion mask guided self-supervised loss function to train the proposed network efficiently. Experimental results demonstrate the effectiveness of the proposed method. Compared with the state-of-the-art methods, our method achieves higher objective and subjective quality on reconstructed HDR images, with faster running speed.

Key words: high dynamic range imaging; multi-exposure image fusion; deep learning; self-supervised learning
Foundation Item(s): National Natural Science Foundation of China (No.62001016)

收稿日期:2022-07-28;修回日期:2022-11-08;责任编辑:朱梅玉
*通讯作者:邓欣

1 引言

现实自然场景下的图像亮度、色彩和对比度通常具有高动态范围(High Dynamic Range,HDR),然而由于硬件限制,普通的相机传感器只能拍摄低动态范围(Low Dynamic Range,LDR)的图像,难以描述和反映自然场景中真实的色彩亮度和纹理细节. 多曝光图像融合(Multi-Exposure image Fusion,MEF)技术提供了一种高效的HDR图像获取方法,通过融合一系列不同曝光度下的LDR图像生成清晰的HDR图像. 许多研究工作[1~4]通过融合静态场景下的LDR图像序列实现HDR图像重建. 然而,当LDR图像序列中存在物体运动或相机抖动时,静态的多曝光图像融合方法会产生严重的运动伪影或污点伪影.

为克服多曝光图像融合产生的运动伪影现象,学者们提出基于图像配准的融合方法[5~10]和基于运动检测的融合方法[11~15]. 其中,基于图像配准的方法使用光流或单应性变换,在全局或局部范围内对动态LDR图像序列进行配准,然后将配准后的LDR图像融合成一张HDR图像. 然而,这类方法的效果极度依赖配准方法的准确度. 另一类基于运动检测的方法先检测图像中的运动区域,然后通过移除这些区域防止图像中出现运动伪影. 然而,这类方法只能处理小幅度运动,当场景中出现大幅度运动时,由于去除运动区域过多,生成的HDR图像信息不全. 同时,上述方法虽可在一定程度上减少运动伪影,但计算复杂度较高.

为了解决传统方法的弊端,研究人员提出了基于深度学习的方法[16~23],以较低的运行复杂度实现动态场景下的多曝光图像融合. 例如,Wu等人[17]提出了编码器-解码器结构的神经网络来同时配准和融合动态LDR图像序列,以生成高质量的HDR图像. 然而,现有基于深度学习的方法都是采用有监督的学习模式,需要真值HDR图像作为指导;而动态场景下真值HDR图像的获取是十分困难的,或者说并不存在真正的真值图像. 这限制了现有深度学习方法在实际动态HDR图像重建任务中的应用.

综上所述,传统方法可以在一定程度上减少运动产生的融合伪影,但其计算复杂度高,且无法处理场景中存在的大幅度运动,难以在实际场景中应用. 而现有的基于深度学习的方法通常采用有监督学习的方式,过度依赖真值HDR图像的存在性和精确性,忽视了高动态范围图像与低动态范围图像序列之间的紧密联系,难以适应复杂多变的现实场景. 同时,现有深度学习数据集中的真值图像是使用简单的静态加权融合方法得到的,缺乏准确性和可靠性. 因此,通过自监督学习探索动态场景下高动态范围与低动态范围图像之间的内在关联,是将动态场景下的高动态范围成像技术广泛应用于现实生活中的关键.

面向动态场景下的多曝光图像融合,本文提出了一种高效的基于深度自监督学习的融合方法,在不依赖真值HDR图像的前提下,实现动态场景下的高效HDR图像重建. 本文主要贡献包括:(1)提出深度自监督学习框架实现动态场景下多曝光图像融合;(2)提出了全局去伪影模块(Global Deghosting Module,GDM),使用注意力机制减少图像中的运动伪影,增强图像细节;(3)设计了融合重建模块(Merging Reconstruction Module,MRM),通过残差和稠密连接增强多层次特征间的信息流动,丰富重建后HDR图像的细节信息;(4)设计了运动掩膜引导的损失函数,用于网络的高效训练.

2 相关工作

2.1 传统动态场景下多曝光图像融合方法

传统动态场景下的HDR图像重建方法根据其核心思想可以大致分为两类:基于图像配准的HDR图像重建方法和基于运动检测的HDR图像重建方法.

2.1.1 基于图像配准的多曝光图像融合方法

基于图像配准的HDR图像重建方法首先在局部或全局范围内对LDR图像序列进行配准,然后将配准后的LDR图像序列融合成一张HDR图像. 例如,Bogoni等人[5]和Kang等人[6]提出从LDR图像序列中选择一张图像作为参考图像,然后使用光流将其余非参考图像转换至参考图像的运动状态,从而实现对LDR图像的配准. 然而这些传统的光流方法无法准确估计运动变化,难以减轻运动造成的伪影现象. 为解决上述问题,Sen等人[9]和Hu等人[10]摒弃了光流方法,将动态HDR重建过程建模为联合优化问题. 具体来说,Sen等人[9]通过优化基于图像块的能量最小化方程来同时配准和重建图像;Hu等人[10]提出同时优化能量函数以及色彩、梯度的连续性,从而将动态图像序列转化为静态图像序列,用于HDR融合. 与基于光流的方法相比,基于优化的方法能够取得更好的重建效果,但其难以处理存在大幅度运动的场景,并且通常计算复杂度很高.

2.1.2 基于运动检测的多曝光图像融合方法

基于运动检测的HDR图像重建方法首先从LDR图像序列中选择参考图像,然后利用LDR图像中的不变特征检测非参考图像中的运动像素. 在此基础上,将检测到的运动像素丢弃,并由静止像素进行补偿,从而达到去除运动伪影的效果. 例如,Gallo等人[11]使用基于图像块的对数强度值来检测非参考图像与参考图像

之间连续的区域,并对这些区域的强度估计求平均值来获得无伪影的HDR图像.然而,这种基于图像块的运动检测方法忽略了图像块之间的信息关联,损失了图像块之间的内容联系.为了保留图像中信息的全局关联性,Oh等人[12]提出了一种秩最小化算法来检测存在运动的异常像素,实现HDR图像的高鲁棒性重建.Ma等人[13]提出了基于结构化块分解的方法SPD-MEF来实现动态场景下的HDR图像重建.该方法首先使用强度映射函数生成潜在图像,然后利用结构向量检测图像中的运动区域.然而SPD-MEF计算复杂度较高,且生成图像中易存在色彩失真.Li等人[14]在SPD-MEF的基础上去除了正则化步骤,并引入了多尺度的方法,在加快运算速度的同时提高动态融合效果.

在场景中物体较少、运动幅度较小的情况下,基于运动检测的方法能够取得较好的融合效果.然而,当场景中物体运动幅度较大时,由于去除的运动区域过多,重建的HDR图像中存在大量信息缺失.

2.2 基于深度学习的动态场景下多曝光图像融合方法

近年来,基于深度学习的方法在动态场景的HDR图像重建任务中取得了显著的效果.例如,Kalantari等人[16]首先利用传统光流方法将输入LDR图像序列进行配准,然后使用卷积神经网络对其进行融合.由于传统光流方法难以精确配准物体动作,Prabhakar等人[23]使用神经网络PWC-Net[24]对光流进行预测,实现对输入LDR图像序列的准确对齐.随着注意力机制在探索图像的全局关联性和保留图像细节上展示出强大的能力,Yan等人[19]将注意力机制引入到动态图像融合任务中,设计端到端的深度卷积神经网络减少了HDR图像重建过程中产生的运动伪影.进一步,Yan等人[20]提出了基于非局部(non-local)注意力机制的神经网络来探索图像序列中的局部和全局内容关联,通过对局部和全局特征进行整合实现高质量HDR图像的重建.

上述基于深度学习的动态HDR图像重建方法能够有效减少重建图像中的运动伪影,增强图像中的纹理细节,融合效果和运行效率都远超传统算法.然而,上述基于深度学习的方法均需要在有真值图像监督的条件下进行网络训练,图像融合精度受限于真值图像的准确性,同时忽视了LDR图像序列与HDR图像之间的信息关联.为克服上述缺点,本文提出采用自监督学习的方式来解决动态场景下的HDR图像重建问题,在无需真值图像的条件下实现高质量的HDR图像重建.

3 本文方法

动态场景下多曝光图像融合旨在将N张曝光度不同、运动状态各异的LDR图像融合为一张曝光度适中、色彩纹理细节丰富的HDR图像.本文使用3张分别在曝光不足、曝光适中和曝光过度的条件下拍摄的LDR图像进行动态HDR重建.为减少运动产生的伪影现象,本文提出了全局去伪影模块,并设计了运动掩膜引导的自监督损失函数以去除运动伪影.下面详细介绍本文网络结构、运动掩膜生成过程及自监督损失函数设计.

3.1 网络结构

本文提出的网络结构如图1所示,整体网络由特征提取模块(Feature Extraction Blocks,FEB)、全局去伪影模块(Global Deghosting Module,GDM)和融合重建模块(Merging Reconstruction Module,MRM)构成.

本文网络的原始输入为3张具有不同曝光度等级的LDR图像,包括曝光不足图像$\boldsymbol{I}_u$、曝光中等图像$\boldsymbol{I}_r$和曝光过度图像$\boldsymbol{I}_o$.其中,曝光中等图像$\boldsymbol{I}_r$为参考图像.在将LDR图像序列输入网络之前,使用直方图匹配策略[25]将曝光不足和曝光过度的LDR图像$\boldsymbol{I}_u$和$\boldsymbol{I}_o$向参考图像$\boldsymbol{I}_r$进行对齐,生成潜在图像$\boldsymbol{I}_{\hat{u}}$和$\boldsymbol{I}_{\hat{o}}$:

$$\boldsymbol{I}_{\hat{u}}=f_{\mathrm{HM}}\left(\boldsymbol{I}_u,\boldsymbol{I}_r\right) \tag{1}$$

$$\boldsymbol{I}_{\hat{o}}=f_{\mathrm{HM}}\left(\boldsymbol{I}_o,\boldsymbol{I}_r\right) \tag{2}$$

其中,$f_{\mathrm{HM}}(\cdot)$表示直方图匹配的操作,生成的潜在图像$\boldsymbol{I}_{\hat{u}}$和$\boldsymbol{I}_{\hat{o}}$分别被用于补偿$\boldsymbol{I}_u$和$\boldsymbol{I}_o$中的运动区域,这些潜在图像将和原始的3张LDR输入图像一起输入到整体网络中.

3.1.1 特征提取模块

特征提取模块(FEB)的结构如图1(a)所示,每一个特征提取模块包括4个带PReLU激活函数的卷积层和一个残差连接,其中前三个卷积层的卷积核大小为3×3,最后一个卷积层的卷积核为1×1,且不同特征提取模块之间不共享参数.具体来说,特征提取模块FEB从5张输入图像中分别提取初始特征,以参考图像$\boldsymbol{I}_r$为例,FEB可从中提取特征$\boldsymbol{G}_r$:

$$\boldsymbol{G}_r=f_{\mathrm{FEB}}(\boldsymbol{I}_r) \tag{3}$$

其中,$f_{\mathrm{FEB}}(\cdot)$表示特征提取模块函数.类似地,可以从其余输入图像中获取初始特征$\boldsymbol{G}_u$,$\boldsymbol{G}_{\hat{u}}$,$\boldsymbol{G}_o$和$\boldsymbol{G}_{\hat{o}}$.

在特征提取之后,将提取出的特征分为两组各自联结:(1)曝光不足的图像特征$\boldsymbol{G}_u$和潜在图像特征$\boldsymbol{G}_{\hat{u}}$与参考图像特征$\boldsymbol{G}_r$进行联结,以保持参考图像中的运动状态与曝光不足图像中的亮度信息;(2)曝光过度的图像特征$\boldsymbol{G}_o$和潜在图像特征$\boldsymbol{G}_{\hat{o}}$与参考图像特征$\boldsymbol{G}_r$进行联结,以保持参考图像中的运动状态与曝光过度图像中的亮度信息,从而得到两组联结特征$\boldsymbol{F}_u$和$\boldsymbol{F}_o$:

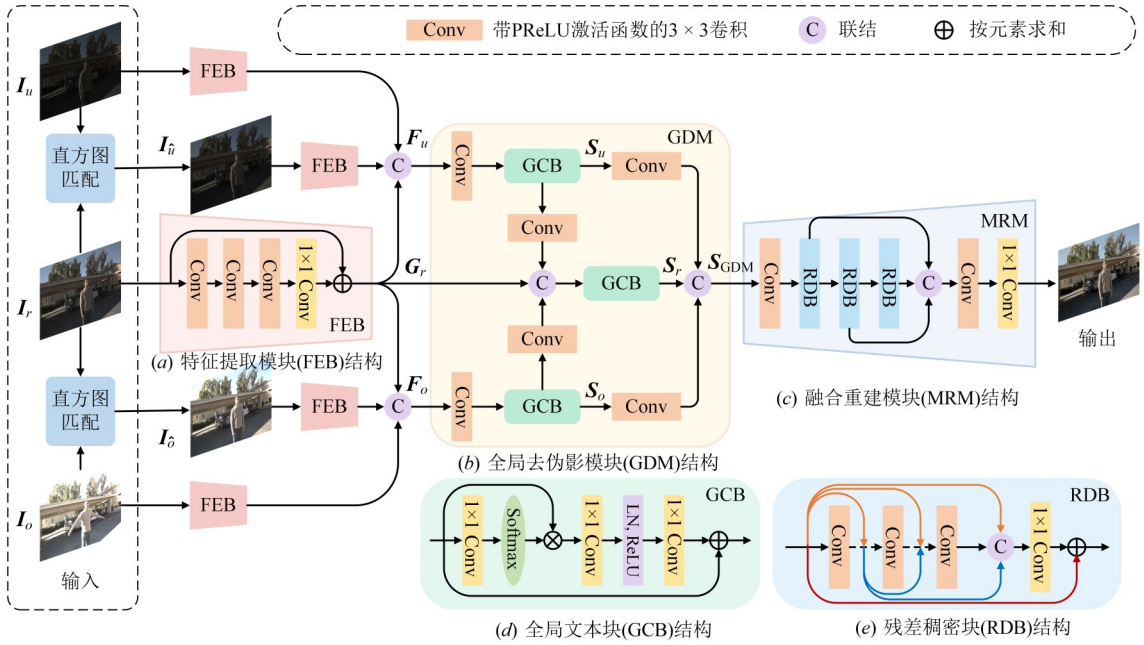


图1　动态自监督多曝光图像融合神经网络

$$\boldsymbol{F}_u=\mathrm{Concat}\left(\boldsymbol{G}_u,\boldsymbol{G}_{\hat{u}},\boldsymbol{G}_r\right) \tag{4}$$

$$\boldsymbol{F}_o=\mathrm{Concat}\left(\boldsymbol{G}_o,\boldsymbol{G}_{\hat{o}},\boldsymbol{G}_r\right) \tag{5}$$

其中,Concat(·)表示在通道维度上进行联结. 此后,以上生成的特征 $\boldsymbol{F}_u$, $\boldsymbol{G}_r$ 和 $\boldsymbol{F}_o$ 被传入全局去伪影模块(GDM)中进行伪影去除和细节增强.

3.1.2　全局去伪影模块

全局去伪影模块(GDM)结构如图1(*b*)所示,它由3个全局文本块(Global Context Block,GCB)[26]构成,目的是高效地整合LDR图像序列中的全局文本信息,并减少融合图像过程中产生的运动伪影. GDM模块的输入为欠曝光图像特征 $\boldsymbol{F}_u$、参考图像特征 $\boldsymbol{G}_r$ 和过曝光图像特征 $\boldsymbol{F}_o$. 首先将欠曝光图像特征 $\boldsymbol{F}_u$ 和过曝光图像特征 $\boldsymbol{F}_o$ 进行特征通道数压缩,然后分别使用GCB模块对全局文本信息进行探索:

$$\boldsymbol{S}_u=f_{\mathrm{GCB}}\left(C_{3\times3}\left(\boldsymbol{F}_u\right)\right) \tag{6}$$

$$\boldsymbol{S}_o=f_{\mathrm{GCB}}\left(C_{3\times3}\left(\boldsymbol{F}_o\right)\right) \tag{7}$$

在此基础上,将上述GCB的输出 $\boldsymbol{S}_u$, $\boldsymbol{S}_o$ 与参考图像特征 $\boldsymbol{G}_r$ 进行结合,再次经过GCB模块实现对参考图像运动状态的强调和对图像细节的增强,得到特征 $\boldsymbol{S}_r$. 最后将3个GCB模块的输出 $\boldsymbol{S}_u$, $\boldsymbol{S}_o$ 和 $\boldsymbol{S}_r$ 结合,作为融合重建模块(MRM)的输入. 这一过程可以表示为

$$\boldsymbol{S}_r=f_{\mathrm{GCB}}\left(\mathrm{Concat}\left(C_{3\times3}\left(\boldsymbol{S}_u\right),\boldsymbol{G}_r,C_{3\times3}\left(\boldsymbol{S}_o\right)\right)\right) \tag{8}$$

$$\boldsymbol{S}_{\mathrm{GDM}}=\mathrm{Concat}\left(C_{3\times3}\left(\boldsymbol{S}_u\right),\boldsymbol{S}_r,C_{3\times3}\left(\boldsymbol{S}_o\right)\right) \tag{9}$$

其中,$f_{\mathrm{GCB}}(\cdot)$表示全局文本块函数,$C_{3\times3}$ 表示带PReLU激活函数的3×3卷积层.

作为GDM的核心模块,全局文本块GCB结合了非局部块(Non-Local Block,NLB)[27]和压缩激活块(Squeeze-and-Excitation Block,SEB)[28]的优点,可实现对图像中的全局信息的高效表征,其结构如图1(*d*)所示. 首先使用1×1卷积和Softmax函数获取输入特征的加权图谱,与输入特征相乘后经过卷积和激活函数处理,再与输入特征通过残差连接相加,得到输出特征. 上述GCB对输入特征的处理流程可表示为

$$\boldsymbol{S}_{\mathrm{out}}=\boldsymbol{S}_{\mathrm{in}}+C_{1\times1}\left(\mathrm{LN}\left(C_{1\times1}\left(\boldsymbol{S}_{\mathrm{in}}\cdot\mathrm{Softmax}\left(C_{1\times1}\left(\boldsymbol{S}_{\mathrm{in}}\right)\right)\right)\right)\right) \tag{10}$$

其中,$\boldsymbol{S}_{\mathrm{in}}$ 和 $\boldsymbol{S}_{\mathrm{out}}$ 表示GCB的输入和输出特征;$C_{1\times1}$ 表示卷积核为1×1的卷积层;LN(·)表示带ReLU激活函数的层归一化函数;Softmax(·)表示Softmax函数.

3.1.3　融合重建模块

融合重建模块(MRM)的结构如图1(*c*)所示,它由多个卷积层和3个残差稠密块(Residual Dense Block,RDB)[29]串联构成. 融合重建模块以全局去伪影模块的输出 $\boldsymbol{S}_{\mathrm{GDM}}$ 作为输入,通过卷积操作和RDB残差稠密块对HDR图像进行重建. 具体来讲,全局去伪影模块的输出特征 $\boldsymbol{S}_{\mathrm{GDM}}$ 首先经过卷积操作压缩特征通道数,之后输入到连续的多个RDB块中进行特征融合与重建. 当前的RDB块的输出作为下一个RDB块的输入,对于第 i 个RDB块,其输入 $\boldsymbol{S}_{i-1}$ 与输出 $\boldsymbol{S}_i$ 的关系可表示为

$$\boldsymbol{S}_i=f_{\mathrm{RDB}}\left(\boldsymbol{S}_{i-1}\right) \tag{11}$$

其中,$f_{\mathrm{RDB}}(\cdot)$表示残差稠密块操作. 所有RDB模块的

输出彼此联结以实现不同层次的特征融合，最后通过3×3和1×1的卷积重建HDR图像，整个过程可以表示为

$$\boldsymbol{H}=C_{1\times1}(C_{3\times3}(\mathrm{Concat}(\boldsymbol{S}_1,\boldsymbol{S}_2,\boldsymbol{S}_3))) \tag{12}$$

作为MRM的核心模块，RDB的结构如图1(*e*)所示. 多个串联的卷积层间建立稠密连接，使网络能够从不同的卷积层中整合多层次的特征，有助于恢复融合图像的细节. RDB的输入与输出之间形成残差连接，保证图像特征中信息完备性，增强网络中的信息流动，使网络易于训练.

3.2 运动掩膜生成

运动掩膜在本文损失函数设计中具有重要的作用，可有效降低运动伪影失真. 根据文献[7]，曝光不足的LDR图像$\boldsymbol{I}_u$可被分解为信号强度c_u、信号结构$\boldsymbol{s}_u$和平均强度l_u，即

$$\begin{aligned}\boldsymbol{I}_u&=\left\|\boldsymbol{I}_u-\boldsymbol{\mu}_{\boldsymbol{I}_u}\right\|_2\cdot\frac{\boldsymbol{I}_u-\boldsymbol{\mu}_{\boldsymbol{I}_u}}{\left\|\boldsymbol{I}_u-\boldsymbol{\mu}_{\boldsymbol{I}_u}\right\|_2}+\boldsymbol{\mu}_{\boldsymbol{I}_u}\\&=\left\|\widetilde{\boldsymbol{I}}_u\right\|_2\cdot\frac{\widetilde{\boldsymbol{I}}_u}{\left\|\widetilde{\boldsymbol{I}}_u\right\|_2}+\boldsymbol{\mu}_{\boldsymbol{I}_u}\\&=c_u\cdot\boldsymbol{s}_u+\boldsymbol{l}_u\end{aligned} \tag{13}$$

本文使用信号结构分量$\boldsymbol{s}_u$来检测图像中的运动区域，曝光不足的LDR图像$\boldsymbol{I}_u$和参考图像$\boldsymbol{I}_r$之间的信号结构分量内积$\boldsymbol{\Gamma}_{ur}$可以被计算为

$$\boldsymbol{\Gamma}_{ur}=\boldsymbol{s}_r^{\mathrm{T}}\boldsymbol{s}_u=\frac{(\boldsymbol{I}_r-\boldsymbol{l}_r)^{\mathrm{T}}(\boldsymbol{I}_u-\boldsymbol{l}_u)}{\left\|\boldsymbol{I}_r-\boldsymbol{l}_r\right\|_2\left\|\boldsymbol{I}_u-\boldsymbol{l}_u\right\|_2} \tag{14}$$

其中，$\boldsymbol{\Gamma}_{ur}$中的每个元素取值范围是[−1,1]，元素值越大表示参考图像$\boldsymbol{I}_r$和欠曝光图像$\boldsymbol{I}_u$之间的像素值连续性越高. 为了获得运动区域图谱，本文设置了阈值θ将$\boldsymbol{\Gamma}_{ur}$映射为二值掩膜$\boldsymbol{B}_u$，即

$$\boldsymbol{B}_u(i,j)=\begin{cases}1, & \text{if}\quad \boldsymbol{\Gamma}_{ur}(i,j)\geqslant\theta\\0, & \text{otherwise}\end{cases} \tag{15}$$

其中，(i,j)表示图像中元素的位置. 图2展示了过曝光LDR图像和欠曝光LDR图像的二值掩膜，其中白色区域表示相对静止区域，黑色区域则表示存在运动的区域. 在计算运动掩膜引导的损失函数之前，需要对二值掩膜进行软化操作，以生成平滑的运动掩膜，即

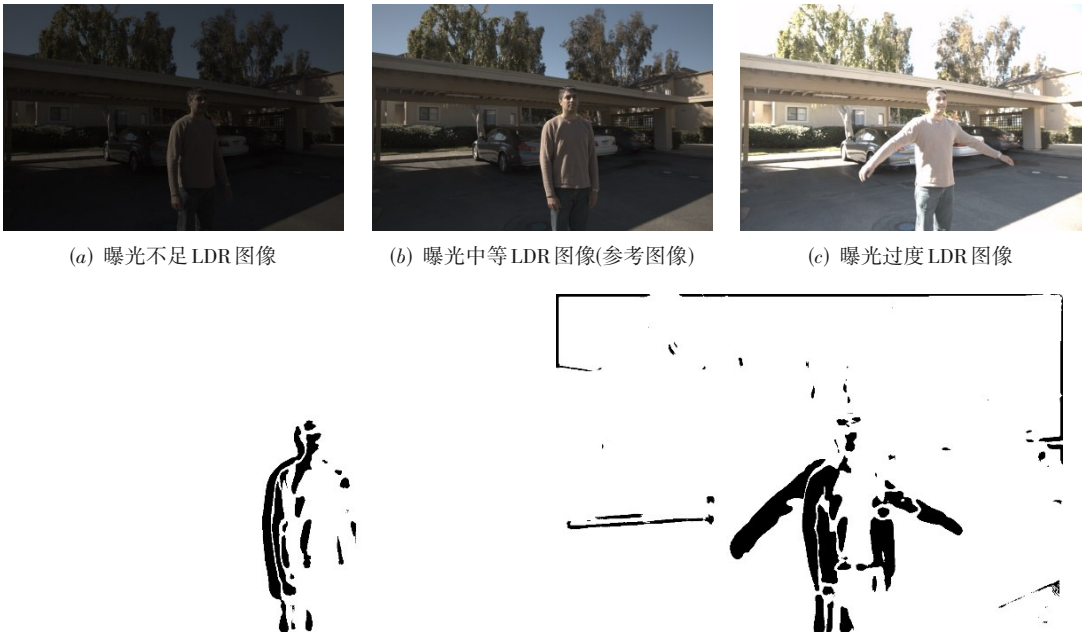
$$\boldsymbol{M}_u(i,j)=\beta\boldsymbol{B}_u(i,j)+(1-\beta) \tag{16}$$

类似地，我们可以获得曝光过度图像$\boldsymbol{I}_o$的运动掩膜$\boldsymbol{M}_o$，对于潜在图像$\boldsymbol{I}_{\hat{u}}$和$\boldsymbol{I}_{\hat{o}}$，其运动掩膜分别与$\boldsymbol{M}_u$和$\boldsymbol{M}_o$呈现互补的态势，即

$$\boldsymbol{M}_{\hat{u}}=\boldsymbol{E}-\boldsymbol{M}_u \tag{17}$$

$$\boldsymbol{M}_{\hat{o}}=\boldsymbol{E}-\boldsymbol{M}_o \tag{18}$$

其中，$\boldsymbol{E}$表示元素全为1的矩阵，参考图像$\boldsymbol{I}_r$的运动掩膜$\boldsymbol{M}_r$设置为$\boldsymbol{E}$.



(*a*) 曝光不足LDR图像　(*b*) 曝光中等LDR图像(参考图像)　(*c*) 曝光过度LDR图像

(*d*) 曝光不足LDR图像对应的运动掩膜　(*e*) 曝光过度LDR图像对应的运动掩膜

图2 多曝光LDR图像中的二值化运动掩膜

3.3 损失函数

为了训练本文提出的自监督神经网络，本文设计了运动掩膜引导的自监督损失函数，来减少动态多曝光图像融合过程中产生的伪影，损失函数的具体组成如下：

$$\mathcal{L}=\sum_{i=1}^{5}\boldsymbol{M}_i\cdot\left\{l_1(\boldsymbol{I}_i,\boldsymbol{H})+\lambda\left[1-\mathrm{SSIM}(\boldsymbol{I}_i,\boldsymbol{H})\right]\right\} \tag{19}$$

其中,$\boldsymbol{I}_i$为输入的LDR图像,$\boldsymbol{M}_i$为图像$\boldsymbol{I}_i$对应的运动掩膜,具体指代如表1所示.$\boldsymbol{H}$表示网络重建的HDR图像,$l_1(\cdot)$表示平均绝对误差(Mean Absolute Error, MAE)损失,$(1-\mathrm{SSIM}(\cdot))$为结构相似性指数(Structural Similarity Index Metric,SSIM)损失,λ为权重参数.损失函数中的平均绝对误差损失主要目的是保留图像细节,而SSIM损失被用于保持图像的结构和锐利的边缘.

表1　损失函数中符号指代

i	1	2	3	4	5
$\boldsymbol{M}_i$	$\boldsymbol{M}_u$	$\boldsymbol{M}_{\hat{u}}$	$\boldsymbol{M}_r$	$\boldsymbol{M}_{\hat{o}}$	$\boldsymbol{M}_o$
$\boldsymbol{I}_i$	$\boldsymbol{I}_u$	$\boldsymbol{I}_{\hat{u}}$	$\boldsymbol{I}_r$	$\boldsymbol{I}_{\hat{o}}$	$\boldsymbol{I}_o$

4　实验结果与分析

4.1　实验设置

本文使用Kalantari等人[16]提出的训练集进行网络训练,共包括74个动态多曝光图像序列,并通过数据增强策略对训练集进行扩充,最终用于训练的图像块总数约为35 000,图像块尺寸为256×256. 在测试集方面,本文使用Kalantari数据集中的15组测试图像序列,以及Sen等人[9]提出的8组动态多曝光图像序列,共计23组图像序列对网络性能进行测试. 在具体的参数设置中,针对式(16)和式(17)中的参数θ和β,我们通过经验和多次实验确定θ为0.9,β为0.8;式(20)中权重参数λ是结构相似性指数SSIM损失函数的权重. 在本文任务中,由于需要减少融合图像中的运动伪影,因此需要对输出图像的结构进行更强的约束,增大SSIM损失函数的权重能够使输出图像的结构更加完整,边缘更加锐利,因此本文将λ设置为10. 在网络训练中,使用默认参数的Adam优化器进行网络训练,初始学习率为1×10^{-4}.

4.2　对比算法

本文的对比算法包括5种目前最先进的动态场景HDR图像融合方法:Sen等人[9]提出的基于图像配准的方法,基于运动检测的方法SPD-MEF[13],FMMEF[14]和MSPD-MEF[15],以及基于有监督深度学习的方法DeepHDR[17]. 所有对比算法的结果都由官方开源的代码库得到,同时,为保证对比公平性,基于深度学习的对比算法DeepHDR[17]在本文的训练数据集上进行了重新训练.

4.3　算法性能客观评价

由于本文提出的HDR重建方法无需真值图像,因此只能采用无参考质量评价标准对融合后的HDR图像进行质量评价. 本文采用3种无参考的客观质量评价指标,包括信息熵(information entropy)[30]、盲色调映射质量指标(Blind Tone-Mapped Quality Index,BTMQI)[31]和动态场景多曝光图像融合指标(Multi-Exposure Fusion Metric for dynamic Scenes,MEF-SSIMd)[32]. 其中,信息熵和MEF-SSIMd的指标值越大、BTMQI的指标值越小,表示HDR图像重建质量越好.

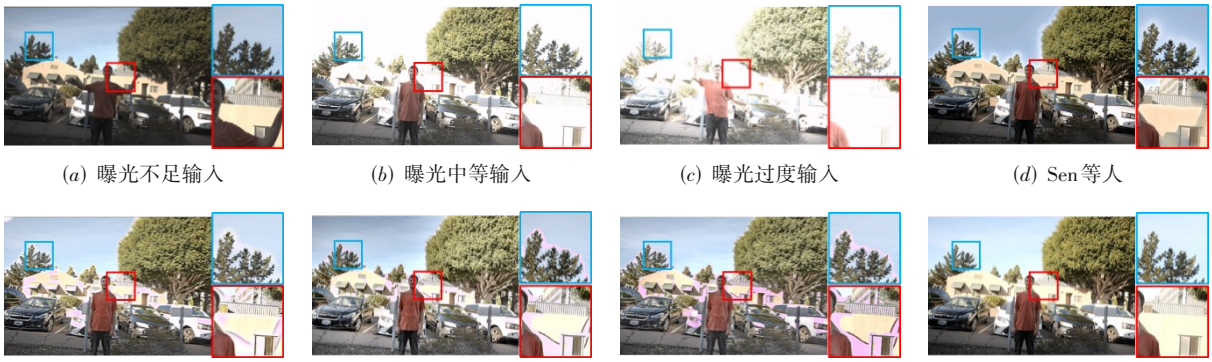
表2展示了本文方法和对比算法在Kalantari数据集[16]和Sen数据集[9]上的客观质量评价数值结果,加粗数据表示最优结果,加横线的数据为次优结果. 可以看到,在上述两个数据集上,本文方法在评价指标BTMQI上达到了最优,超过了所有对比算法的结果;在Entropy指标上达到了次优,仅次于Sen等人提出的方法. 在MEF-SSIMd指标上超过了传统算法的结果,在Sen数据集上超过了深度监督学习方法的结果,在Kalantari数据集上仅次于DeepHDR[17]且指标十分接近. Sen等人提出的方法在多尺度下通过图像配准和迭代优化的方式得到HDR图像,尽可能利用到输入图像中所有信息进行融合,因此Entropy指标上达到了最优,然而其他指标结果较差. 上述实验结果表明,本文提出的方法能够在不依赖真值图像的条件下,以自监督学习的方式在动态场景下实现高质量的HDR图像重建,甚至在多个指标上超越了有监督学习算法DeepHDR. 主要原因在于本文提出的自监督学习方法不依赖真值图像,通过充分建模输入和输出图像的内在关联对信息进行高效融合,在增强融合图像细节的同时减少了运动伪影. 与自监督学习相比,有监督学习方法的融合性能十分依赖真值图像,在真值图像不准确的情况下融合性能十分受限.

表2　不同算法的HDR图像重建结果在Kalantari数据集和Sen数据集上的客观质量结果对比

数据集	Kalantari数据集[16]			Sen数据集[9]		
评价指标	Entropy ↑	BTMQI ↓	MEF-SSIMd ↑	Entropy ↑	BTMQI ↓	MEF-SSIMd ↑
Sen等人[9]	**7.604 1**	3.103 2	0.648 7	**7.650 3**	3.732 6	0.781 8
SPD-MEF[13]	7.449 3	3.061 6	0.623 0	7.260 0	3.726 9	0.763 6
FMMEF[14]	7.477 8	3.562 0	0.614 5	7.386 3	3.687 6	0.739 1
MSPD-MEF[15]	7.277 9	<u>2.887 2</u>	0.617 2	7.368 8	<u>2.970 2</u>	0.743 2
DeepHDR[17]	7.463 8	3.065 2	**0.652 8**	7.409 2	3.061 8	<u>0.789 4</u>
本文方法	<u>7.481 8</u>	**2.869 5**	<u>0.652 3</u>	<u>7.458 2</u>	**2.967 7**	**0.795 0**

4.4 算法性能主观评价

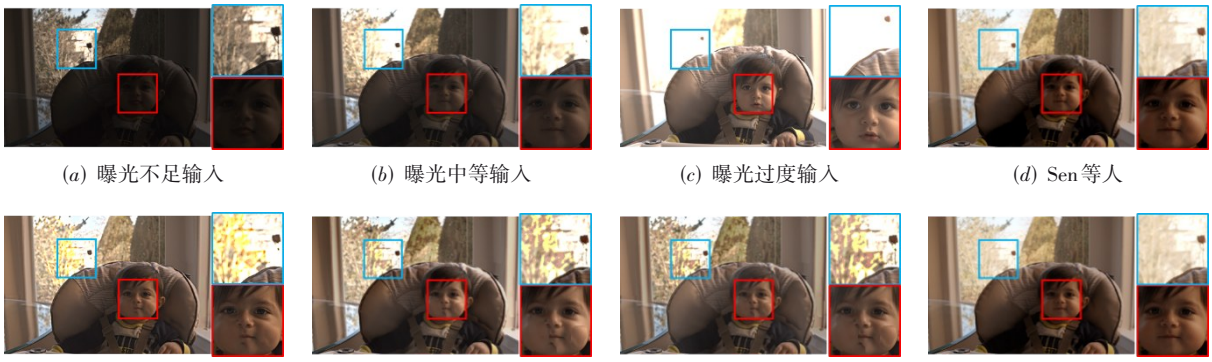
图3展示了本文方法与对比算法在Kalantari测试集上的HDR融合结果. 在图3中可以看到,对比算法在树冠位置存在严重的光晕效应和色彩失真,而本文方法则有效保留了图像的色彩和纹理细节. 此外,对比算法生成的图像中没有去除由于手臂运动产生的运动伪影,从而严重影响了图像的视觉质量,而本文方法则有效地去除了运动伪影并恢复了运动区域的色彩和图像信息. 图4展示了本文方法与对比算法在Sen测试集上的HDR融合结果. 可以看到,本文方法更好地恢复了背景中树木的色彩和轮廓,并减少了人脸上的运动伪影.



(a) 曝光不足输入 (b) 曝光中等输入 (c) 曝光过度输入 (d) Sen等人

(e) SPD-MEF (f) FMMEF (g) MSPD-MEF (h) 本文方法

图3 Kalantari测试集中图像010的多曝光图像融合视觉效果对比



(a) 曝光不足输入 (b) 曝光中等输入 (c) 曝光过度输入 (d) Sen等人

(e) SPD-MEF (f) FMMEF (g) MSPD-MEF (h) 本文方法

图4 Sen测试集中图像HighChair的多曝光图像融合视觉效果对比

4.5 运动掩膜的作用

为了验证运动掩膜在本文损失函数设计中的作用,我们分别使用带有和不带有运动掩膜的损失函数对网络进行训练,并在Kalantari数据集[16]上进行测试. 表3展示了是否使用运动掩膜对HDR图像融合效果的影响,加粗数据表示最优结果. 可以看出,运动掩膜可显著改善BTMQI和MEF-SSIMd指标.

表3 是否使用运动掩膜对融合效果的影响

训练策略	BTMQI ↓	MEF-SSIMd ↑
不使用运动掩膜	3.571 7	0.378 3
使用运动掩膜	**2.869 5**	**0.652 3**

此外,我们将是否使用运动掩膜训练网络的HDR融合结果进行可视化比较,如图5所示. 可以看到,在使用运动掩膜的条件下,有效去除了重建HDR图像中的运动伪影,有效提高了图像的视觉质量.

4.6 算法运行时间比较

算法运行的时间复杂度决定了其能否被高效应用于实际场景中,为了证明本文方法的高效性,我们在Kalantari数据集[16]和Sen数据集[9]上进行实验,分别计算本文方法和对比算法生成一张HDR图像所需要的平均时间. 根据作者提供的官方代码,Sen等人的方法[9]及SPD-MEF[13]、FMMEF[14]和MSPD-MEF[15]方法使用Linux环境下的MATLAB,在Intel Core i9-9960X CPU上进行运行和测试,DeepHDR[17]及本文方法在Linux环境下使用PyTorch框架,在NVIDIA GeForce 3090 GPU上进行训练和测试. 对比结果如表4所示,可以看到,本文方法能够以最短的运算时间生成HDR图像,平均每张图像仅需要0.48 s,远低于对比算法的运行时间.

(*a*) 输入动态LDR图像序列

(*b*) 不带有运动掩膜加权的损失函数训练后的融合效果

(*c*) 带有运动掩膜加权的损失函数训练后的融合效果

图5　运动掩膜对HDR图像融合视觉效果对比

表4　本文方法与对比算法运行时间比较

方法	Sen等人[9]	SPD-MEF[13]	FMMEF[14]	MSPD-MEF[15]	DeepHDR[17]	本文方法
运行时间/s	57.08	11.76	3.93	3.67	0.64	**0.48**

算法运行时间产生差异的主要原因源自算法的实现思路. Sen等人的方法[9]对输入图像序列中每一张图像在多个尺度下进行迭代配准和目标函数优化,需要大量的运算时间和资源;而SPD-MEF[13],FMMEF[14]和MSPD-MEF[15]方法将图像拆分为图像块,以图像块为单位进行分解、融合和重建,同样增加了运算时间;本文方法采用基于端到端神经网络的深度学习方法,可将整张图像直接输入网络,无需划分图像块,因此极大缩短了算法的运行时间.

5　总结

本文首次提出了一种自监督的神经网络模型,在不依赖真值HDR图像的前提下,实现动态场景下的高质量HDR图像重建. 为了减少HDR图像融合过程中由于物体运动产生的伪影,本文提出全局去伪影模块,使用注意力机制减少图像中的运动伪影,增强图像细节. 此外,本文设计了运动掩膜引导的自监督损失函数,用于网络的高效训练. 在多个图像数据集上的实验结果表明,相较于现有算法,本文方法在客观和主观质量评价指标上均取得更加优秀的HDR图像融合效果,并且时间复杂度更低.

参考文献

[1] MA K D, ZENG K, WANG Z. Perceptual quality assessment for multi-exposure image fusion[J]. IEEE Transactions on Image Processing, 2015, 24(11): 3345-3356.

[2] PRABHAKAR K R, SRIKAR V S, BABU R V. DeepFuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs[C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2017: 4724-4732.

[3] DENG X, DRAGOTTI P L. Deep convolutional neural network for multi-modal image restoration and fusion[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(10): 3333-3348.

[4] DENG X, ZHANG Y T, XU M, et al. Deep coupled feedback network for joint exposure fusion and image super-resolution[J]. IEEE Transactions on Image Processing, 2021, 30: 3098-3112.

[5] BOGONI L. Extending dynamic range of monochrome and color images through fusion[C]//Proceedings 15th International Conference on Pattern Recognition. Piscataway: IEEE, 2002: 7-12.

[6] KANG S B, UYTTENDAELE M, WINDER S, et al. High dynamic range video[J]. ACM Transactions on Graphics, 22(3): 319-325.

[7] JINNO T, OKUDA M. Motion blur free HDR image acquisition using multiple exposures[C]//2008 15th IEEE International Conference on Image Processing. Piscataway:

IEEE, 2008: 1304-1307.

[8] 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.

[9] SEN P, KALANTARI N K, YAESOUBI M, et al. Robust patch-based HDR reconstruction of dynamic scenes[J]. ACM Transactions on Graphics, 31(6): 203.

[10] HU J, GALLO O, PULLI K, et al. HDR deghosting: How to deal with saturation? [C]//2013 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2013: 1163-1170.

[11] GALLO O, GELFANDZ N, CHEN W C, et al. Artifact-free high dynamic range imaging[C]//2009 IEEE International Conference on Computational Photography (ICCP). Piscataway: IEEE, 2009: 1-7.

[12] OH T H, LEE J Y, TAI Y W, et al. Robust high dynamic range imaging by rank minimization[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(6): 1219-1232.

[13] MA K D, LI H, YONG H W, et al. Robust multi-exposure image fusion: A structural patch decomposition approach[J]. IEEE Transactions on Image Processing, 2017, 26(5): 2519-2532.

[14] LI H, MA K D, YONG H W, et al. Fast multi-scale structural patch decomposition for multi-exposure image fusion[J]. IEEE Transactions on Image Processing, 2020: 32310768.

[15] LI H, CHAN T N, QI X B, et al. Detail-preserving multi-exposure fusion with edge-preserving structural patch decomposition[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(11): 4293-4304.

[16] KALANTARI N K, RAMAMOORTHI R. Deep high dynamic range imaging of dynamic scenes[J]. ACM Transactions on Graphics, 36(4): 144.

[17] WU S Z, XU J R, TAI Y W, et al. Deep high dynamic range imaging with large foreground motions[C]//Computer Vision - ECCV 2018: 15th European Conference. New York: ACM, 2018: 120-135.

[18] PRABHAKAR K R, ARORA R, SWAMINATHAN A, et al. A fast, scalable, and reliable deghosting method for extreme exposure fusion[C]//2019 IEEE International Conference on Computational Photography (ICCP). Piscataway: IEEE, 2019: 1-8.

[19] YAN Q S, GONG D, SHI Q F, et al. Attention-guided network for ghost-free high dynamic range imaging[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2019: 1751-1760.

[20] YAN Q S, ZHANG L, LIU Y, et al. Deep HDR imaging via a non-local network[J]. IEEE Transactions on Image Processing, 2020, 29: 4308-4322.

[21] LIU Z, LIN W J, LI X P, et al. ADNet: Attention-guided deformable convolutional network for high dynamic range imaging[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway: IEEE, 2021: 463-470.

[22] YAN Q S, GONG D, SHI J Q, et al. Dual-attention-guided network for ghost-free high dynamic range imaging[J]. International Journal of Computer Vision, 2022, 130(1): 76-94.

[23] PRABHAKAR K R, SENTHIL G, AGRAWAL S, et al. Labeled from unlabeled: Exploiting unlabeled data for few-shot deep HDR deghosting[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 4873-4883.

[24] SUN D Q, YANG X D, LIU M Y, et al. PWC-net: CNNs for optical flow using pyramid, warping, and cost volume [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 8934-8943.

[25] GROSSBERG M D, NAYAR S K. Determining the camera response from images: What is knowable? [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(11): 1455-1467.

[26] CAO Y, XU J R, LIN S, et al. GCNet: Non-local networks meet squeeze-excitation networks and beyond[C]//2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). Piscataway: IEEE, 2019: 1971-1980.

[27] ZHANG Y L, TIAN Y P, KONG Y, et al. Residual dense network for image super-resolution[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 2472-2481.

[28] WANG X L, GIRSHICK R, GUPTA A, et al. Non-local neural networks[C]//2018 IEEE/CVF Conference on

Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 7794-7803.

[29] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 7132-7141.

[30] TSAI D Y, LEE Y, MATSUYAMA E. Information entropy measure for evaluation of image quality[J]. Journal of Digital Imaging, 2008, 21(3): 338-347.

[31] 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.

[32] 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, 2019: 31535996.

作者简介

张雨童　男，1999 年 2 月生，四川成都人. 2020年在北京航空航天大学获得学士学位. 现为北京航空航天大学硕士研究生. 主要研究方向为深度学习和图像融合.

E-mail: yutongzhang@buaa.edu.cn

邓　欣　女，1991 年 1 月生，山东威海人. 博士毕业于英国伦敦帝国理工学院获博士学位. 现为北京航空航天大学网络空间安全学院副研究员. 主要研究方向为多模态图像处理和可解释神经网络.

E-mail: cindydeng@buaa.edu.cn

徐　迈　男，1981 年 2 月生，江苏无锡人. 博士毕业于英国伦敦帝国理工学院获博士学位. 现为北京航空航天大学电子信息工程学院教授. 主要研究方向为图像处理和人工智能. 中国电子学会会员编号：E190014800S.

E-mail: maixu@buaa.edu.cn