

基于稠密局部-全局特征融合的超高清 多曝光图像融合方法

贾修一¹, 林乔万尼¹, 郑卓然¹, 石争浩²

(1. 南京理工大学计算机科学与工程学院, 江苏南京 210094; 2. 西安理工大学计算机科学与工程学院, 陕西西安 710048)

摘要: 随着超高清(Ultra-High-Definition, UHD)成像技术的应用,生成高质量的UHD图像通常需要融合多幅曝光水平不同的UHD图像。然而,目前基于深度学习的多曝光图像融合方法直接融合从不同曝光水平的图像中提取的特征图,未能充分利用不同曝光级别图像中的特征信息,而这些特征信息对于获得良好的多曝光融合结果至关重要。为解决这一问题,我们提出了一种新颖的UHD多曝光图像融合方法,该方法结合了图像的局部和长距离依赖特征,旨在挖掘不同曝光级别图像之间的依赖关系,提取出更高级的语义和特征。进而,利用不同级别的短连接来聚合不同粒度的特征。最后,为了过滤带噪声的特征,我们还提出了带有门控机制的多层感知器来生成高质量的超高清图像。为了更好地展示实验结果,我们还针对多曝光融合任务建立了一个UHD图像数据集。实验结果表明,在单个显存24G的GPU上执行UHD多曝光图像融合任务时,我们的方法明显优于现有方法。

关键词: 超高清图像;多曝光图像融合;稠密特征融合;双分支;实时处理

基金项目: 国家自然科学基金(No.62176123)

中图分类号: TP397.4

文献标识码: A

文章编号: 0372-2112(2025)01-0238-10

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20230840

UHD Multi-Exposure Image Fusion via Dense Local-Global Feature Aggregation

JIA Xiu-yi¹, LIN Qiao-wan-ni¹, ZHENG Zhuo-ran¹, SHI Zheng-hao²

(1. School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, Jiangsu 210094, China;

2. School of Computer Science and Engineering, Xi'an University of Technology, Xi'an, Shaanxi 710048, China)

Abstract: With the deployment of ultra-high-definition (UHD) imaging technology, generating high-quality UHD images typically involves fusing multiple UHD images with varying exposure levels. However, current multi-exposure image fusion (MEF) methods based on deep learning perform direct fusion of feature maps extracted from images with different exposure levels. These methods fail to fully exploit the feature information in images with varying exposure levels, which is essential for achieving successful MEF outcomes. To address this problem, we develop a UHD multi-exposure image fusion approach that incorporates both local and long-range characteristics of images, and it aims to mine the dependencies of images with different exposure levels. By enforcing translation invariance and self-attention on images with varying exposure levels, we can extract higher-level semantics and features. Furthermore, we aggregate the resulting features of different granularity by utilizing shortcut connections at various levels. Finally, we propose the Gate-MLP with a gating mechanism for filtering features with noise to generate a high-quality UHD image. To better demonstrate the work for UHD MEF task, we also establish a UHD image dataset for MEF task. Extensive experimental results demonstrate that our approach significantly outperforms existing approaches for UHD multi-exposure image fusion task on a single 24G RAM GPU.

Key words: UHD images; multi-exposure image fusion; dense feature aggregation; dual-branch; real-time processing
Foundation Item(s): National Natural Science Foundation of China (No.62176123)

1 引言

随着手持摄像设备的不断发展和对高质量超高清(Ultra-High-Definition, UHD)图像需求的增加,获取视觉上令人愉悦的UHD图像已成为亟待解决的问题.多曝光图像融合技术提供了一种经济高效的解决方案,可从多个低动态范围图像中获取清晰的高动态范围图像.迄今为止,人们已经提出了一系列传统的多曝光融合方法^[1].然而,传统的多曝光融合方法通常依赖于手工创建的特征来融合多重曝光图像,这在性能和鲁棒性方面会有不足,使得在测试数据中可能会出现伪影.相反,基于深度学习的方法得益于它们能够自动学习和提取相关特征,在多曝光融合任务中取得了令人满意的效果.

目前,这些方法大多利用卷积神经网络(Convolutional Neural Networks, CNN)进行特征提取,并依靠具有不同曝光水平的多个低动态范围图像来生成高动态范围图像.然而,由于依赖大量低动态范围图像,这种方法往往导致处理时间变慢,存储消耗增加.为解决这一问题,一些方法只使用2幅或3幅曝光水平不同的低动态范围图像作为图像融合的输入^[2,3].此外,除了基于CNN的技术外,Qu等人^[4]提出了一种基于Transformer的多曝光融合任务框架.然而,上述方法存在以下问题:(1)如图1所示,这些方法都无法在单个显存24G的GPU上直接实时处理一对超高清图像.这种限制源于这些方法没有采用下采样或低阶处理技术,无法直接处理大分辨率($3\ 840 \times 2\ 160$)图像.(2)在特征提取方面,由于CNN的感受野有限,因此在捕捉长距离特征方面存在局限性.虽然Transformer^[5]具有执行远距离建模能力,但由于缺乏感应式建模,它通常需要大量的训练数据才能达到最佳性能.(3)这些方法忽略了将局部学习算法与全局学习算法相结合的好处.相反,它们只专注于采用单一建模策略,而没有充分利用2种方法同时使用所带来的优势.为了应对这些挑战,引入了一种新的稠密局部-全局特征融合方法,主要利用成对的CNN模块和Transformer模块进行特征提取.这些模块在低分辨率图像上并行运行,以捕捉局部和全局特征.为防止梯度消失,使用多个短连接对输出特征图进行融合.这种方法能利用CNN和Transformer的优势,实现更有效的特征提取.为了在特征融合阶段最大限度地利用提取的特征信息,引入稠密特征融合方法,而不是直接融合全局颜色特征和局部纹理特征.此外,还提出了门控多层感知器(Gate-MuLtilayer Perceptron, Gate-MLP)技术,以动态过滤提取的特征,减少特征图中的噪声.为了便于全面了解超高清图像的多曝光融合,还创建了一个超高清文档图像数据集,该数据集专门设计用于多曝光融合任务.该数据集包含发票、合同等多种文档类型.

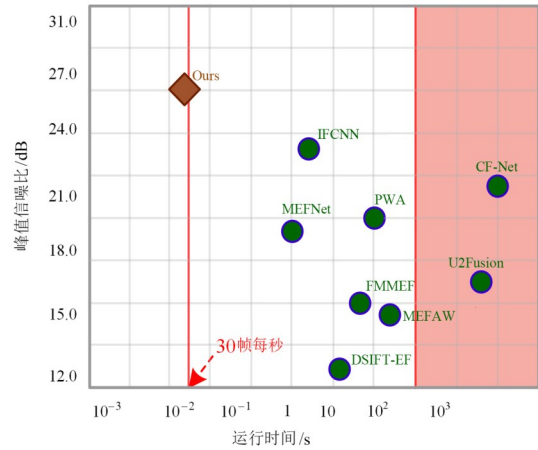


图1 多曝光融合方法在本文数据集上的峰值信噪比与运行时间对比

本文的贡献可概括如下:

(1)提出了一种稠密的局部-全局特征融合方法,该方法表现出很高的性能,并且能够在显存24G的单个RTX 3090 GPU上以30帧每秒的帧速率处理一对UHD($3\ 000 \times 4\ 000$)多重曝光图像.

(2)定制了稠密融合策略以融合从多个分支提取的特征.这一策略解决了单次曝光图像中特征信息不足的问题.此外,为了减少在特征交互过程中引入过多噪声,提出了Gate-MLP. Gate-MLP采用了一种特征动态选择机制来过滤相互作用的特征,确保只有相关的特征才能被识别.

(3)构建了一个重要而全面的超高清图像数据集,专门用于多曝光融合任务.将所提方法与目前最先进的传统方法和基于深度学习的方法进行比较,实验结果表明了本文方法在融合性能和图像质量方面的有效性和优越性.

2 相关工作

2.1 多曝光图像融合

现有的多曝光图像融合方法大致可分为传统方法和深度学习方法两大类,传统方法处理流程中通常包括图像转换、信息丰富度量度和融合策略设计3个主要处理部分,根据其进行分析处理的图像域不同可以分为基于空间域和基于转换域的技术方法.传统的图像融合方法主要包括基于加权平均的方法、基于小波变换的方法、基于主成分分析的方法、基于稀疏表示的方法等.这些传统方法一般依赖于手工设计的特征和策略进行融合.近年来,深度学习技术得到了快速的发展,因其强大的表征能力,研究人员提出了许多基于深度学习的多曝光图像融合方法.在这些工作中,很多研究人员广泛探索了卷积神经网络、生成对抗网络、自动编码器等深度网络模型应用于图像融合任务的可能性.

传统的多曝光融合方法大致可分为两大类:基于空间域的方法^[6]和基于变换域的方法^[7]. 空间域方法直接应用模型对 RGB 图像进行色彩和细节重建. 基于空间域的多曝光融合方法利用特定的空间特征,按照特定规则直接在空间域融合输入源图像. 这些方法为每个输入图像生成一个权重图(权重张量),并以所有输入图像的加权平均值获得融合图像. 基于变换域的多曝光融合方法通常包括 3 个阶段:图像变换、系数融合和反演. 首先,通过图像分解或表示等技术将输入图像转换到另一个域,如频域;然后,根据预先设计的融合规则对转换后的系数进行融合;最后,对融合后的系数进行相应的反相处理,得到最终的融合图像.

深度学习在多曝光融合领域备受关注,取得了一些进展. 例如,Liu 等人^[8]提出了一种具有细节校正功能的分层注意力引导学习方法. 在另一项研究中,Deng 等人^[9]设计了一种新颖的深度卷积神经网络来处理多模态图像修复和融合问题. 其方法能自动分离不同模态之间共享的共同信息和属于每种模态的独特信息. 此外,Yang 等人^[10]提出了一种基于生成式对抗网络的新型多曝光融合方法,通过增加判别器的数量,以区分融合图像和极端曝光图像对,从而在最终的融合输出中保留极端曝光图像的内容. Jung 等人^[11]设计了一种深度图像融合网络,该网络采用无监督深度学习框架. 该方法在整个过程中采用卷积神经网络,包括特征提取、特征融合和图像重建. 这些基于深度学习的方法取得了令人满意的成果,通过利用深度神经网络的强大功能来提高融合性能,从而推动了多曝光融合技术的发展.

2.2 实时处理超高清图像

随着手持摄像设备的快速发展,对实时处理超高清图像的需求也在不断增长. 研究者已关注到这一需求,并提出了专门用于实现超高清图像实时处理的方法. Hu 等人^[12]设计了一种基于金字塔结构的轻量级高效网络. 该网络利用基于梯度的搜索策略来搜索分层补丁和缩放方案,从而可以在没有严格限制的情况下进行更灵活的处理. Liang 等人^[13]设计了一种专门用于实时处理超高清图像的拉普拉斯金字塔变换网络,该网络作为一种轻量级网络,能在降低分辨率的同时有效地转换低频成分. 此外,该网络采用渐进式屏蔽策略,有效地细化了高频成分. Lin 等人^[14]提出了一种专为实时处理高分辨率图像而定制的背景替换技术. 他们的方法能以每秒 30 帧的速度处理 4K 分辨率图像,并在现代 GPU 上以每秒 60 帧的速度处理高清图像. Lin 等人^[15]为 4K 分辨率弱光图像增强提出了一种可解释的双边学习框架. 它能以每秒 100 帧的速度处理 4K (3 840 × 2 160) 分辨率图像.

3 方法

要处理一对低曝光和过曝光的图像,首先使用双线性插值法对它们进行下采样,以获得低分辨率 (256 × 256) 图像. 为了获得不同曝光度图像的局部纹理特征信息和全局色彩特征信息,使用卷积神经网络和 Transformer 在双路径上对图像进行处理. 在得到图像的局部纹理特征信息和全局色彩特征信息之后,对这些特征图进行上采样使其分辨率达到原输入图像的分辨率 (3 840 × 2 160),再将高分辨率特征图与输入图像相乘,以获得高质量的特征图. 为了降低特征融合中噪声对融合后的特征的代表能力的影响,设计了 Gate-MLP. 该模块使用门控机制对特征图中包含的特征信息进行动态筛选,以此增强特征图中有效信息的表征能力. 将获得的高质量特征图用 Gate-MLP 处理后,再将不同曝光度图像的局部纹理特征和全局色彩特征进行稠密融合,即两两交互融合,从而获得高质量的超高清图像. 图 2(a) 展示了所提方法的架构,左半部分是基于双路径的特征提取,右半部分是交互式的稠密特征融合.

3.1 特征提取

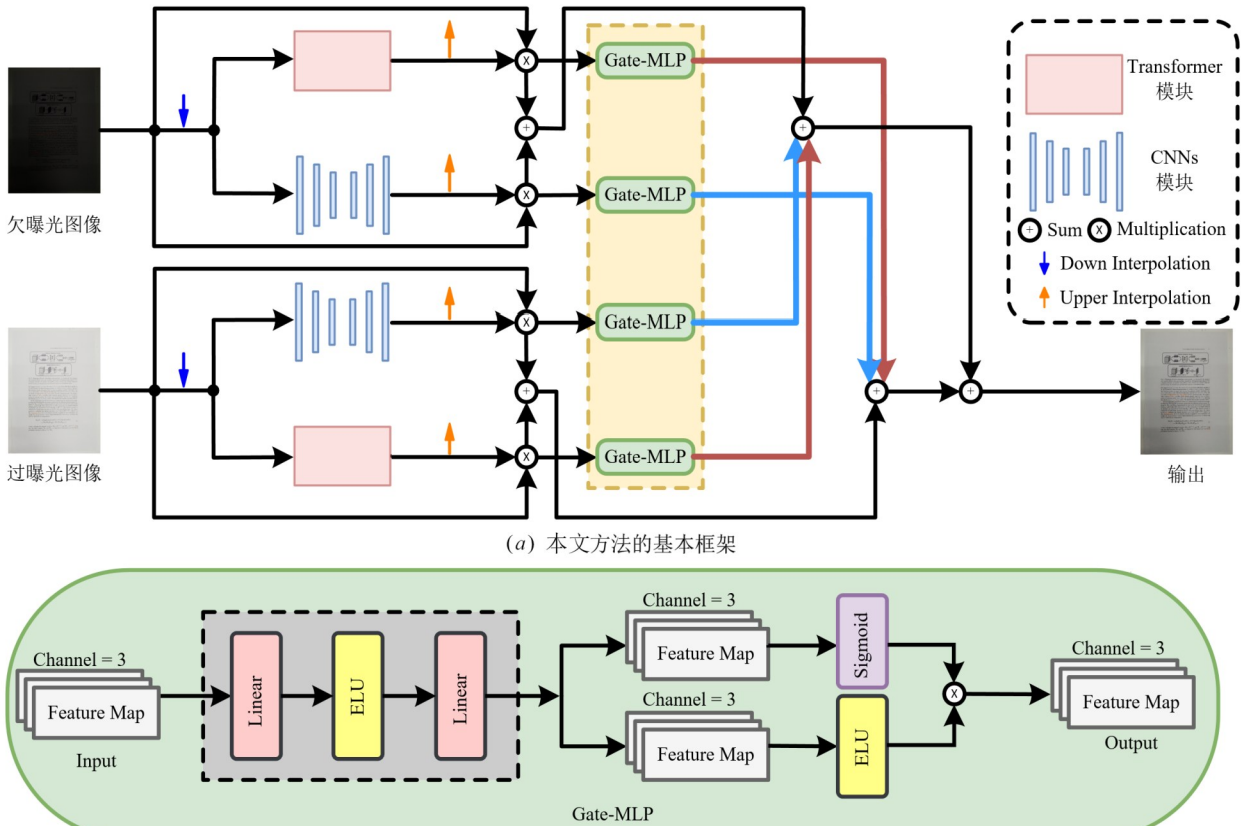
提出的特征提取方法通过一个多分支网络作用于多个曝光图像的输入. 每个曝光图像由一组网络 N_s 学习,其中一个为局部信息提取网络 N_L 和全局信息建模网络 N_G . 具体来说,对于给定的一对欠曝光图像 I_U 和过曝光图像 I_O ,首先对这两幅图像进行下采样操作,以获得固定的低分辨率 (256 × 256) 欠曝光图像 I'_U 和过曝光图像 I'_O . 然后,并行(并行网络可在多个 GPU 上加速)对获得的低分辨率图像 I'_U 和 I'_O 进行特征提取. 下面将详细讨论 N_s 中包含的局部特征提取网络 N_L 和全局特征建模网络 N_G .

(1) 局部特征提取. 受 U-Net^[16] 的启发,U 型编解码结构对图像的局部特征信息更为敏感,因此使用标准的 U-Net 作为 CNN 模块来进行特征提取. 首先对原始图像进行 4 倍下采样,在每一次下采样后将特征通道的数量增加 1 倍. 然后对特征图进行上采样 (4 倍或更高),每次上采样后特征通道数减半. 局部纹理特征的提取可以写成以下形式:

$$F_L = \text{UTB} \uparrow (\text{Conv}_{1 \times 1}(F'_L)) \odot I \quad (1)$$

其中,UTB \uparrow 表示使用双线性插值进行上采样; I 表示原始高分辨率图像; F'_L 表示 CNN 模块处理后的特征图;Conv_{1×1} 用于在通道维度上挤压 F'_L . 将 F'_L 的通道数挤压到 3 之后,高分辨率 (3 000 × 4 000) 特征图向上采样并乘以 I ,生成局部纹理特征 f_L . UTB 作用于 I 所生成的特征图可以被视为生成了一个具有局部特征的仿射变换系数.

(2) 全局色彩特征. 受 Zamir 等人^[17] 的启发,使用



(a) 本文方法的基本框架
(b) Gate-MLP基本架构
图2 本文方法的架构图

Transformer 模块来提取全局颜色特征. 首先,用卷积层对图像 I 进行预处理(卷积核大小为 1×1),以获得浅层特征,然后通过 N_G 将浅层特征转化为深层特征. 模块 N_G 与 ViT^[18] 的不同之处在于,它通过计算通道上的注意力来隐含编码全局上下文信息,而不是空间注意力.

$$F_G = UTB \uparrow (\text{Conv}_{1 \times 1}(F'_G) + I) \odot I \quad (2)$$

其中, I' 表示低分辨率图像; F'_G 表示经过 Transformer 模块处理后的特征图; $\text{Conv}_{1 \times 1}$ 用于在通道维度中压缩 F'_G 使特征图中的通道数为 3. 将 F'_G 挤压到通道数为 3, 然后上采样回高分辨率 ($3\,000 \times 4\,000$) 图像, 得到全局特征 f_g .

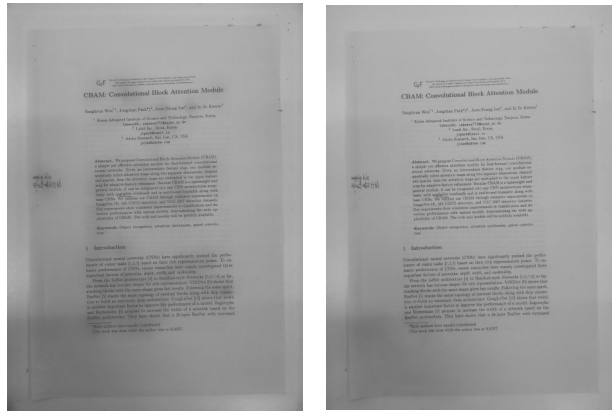
(3) 稠密特征融合. 提取图像 I_U 和 I_O 的特征 f_U 和 f_g 后, 需要对这些特征进行融合, 以获得清晰的 UHD 图像. 超高清图像经过大范围下采样后, 会丢失大量的纹理和色彩信息, 即使采用了注意力下采样也是如此. 因此, 设计了一种稠密特征融合的方法, 以保留原始图像的有效纹理和色彩信息. 具体来说, 首先, 以端到端的方式获取输入图像的整体特征图. 然后, 通过加法运算将特征 f_U 、 f_g 和 I 进行融合. 最后, 得到一个包含原始图像信息的特征图 f_m . 随着融合程度的增加, 特征图或原始图像中不相关的细节也会大量叠加, 从而导致整个

模型的学习空间不稳定. 为解决该问题, 引入一个门控多层感知器.

3.2 Gate-MLP

为了充分发挥多特征图的优势并避免融合过程中不相关噪声的引入, 设计了 Gate-MLP 来进行动态特征选择, 它就像一个过滤器, 作用于每个通道和每个空间位置的特征图, 对这些特征图中的信息进行筛选处理, 最后得到包含噪声更少的特征图. 为了更直观地反映 Gate-MLP 的作用, 将有无 Gate-MLP 的网络融合结果可视化, 结果如图 3 所示. 很明显, 在不带 Gate-MLP 的网络融合结果中存在大量阴影, 而在带 Gate-MLP 的网络融合结果中却没有这种阴影. 此外, 还对经过和未经 Gate-MLP 处理的各个通道的特征图进行了可视化, 结果如图 4 所示. 图 4 中第 1 行图片是未经 Gate-MLP 处理的可视化特征图, 第 2 行图片是经过 Gate-MLP 处理的可视化特征图. Gate-MLP 作用于 RGB 三通道的特征图, 在 R 通道的对比图中, Gate-MLP 对于纹理细节特征进行了过滤, 在第 2 行的特征图中可以很明显的看出文字的大致轮廓. 在 G 通道的对比图中, Gate-MLP 对于色彩特征进行了过滤, 经过 Gate-MLP 处理的特征图具有较为真实的色彩特征. 在 B 通道的对比图中, Gate-MLP

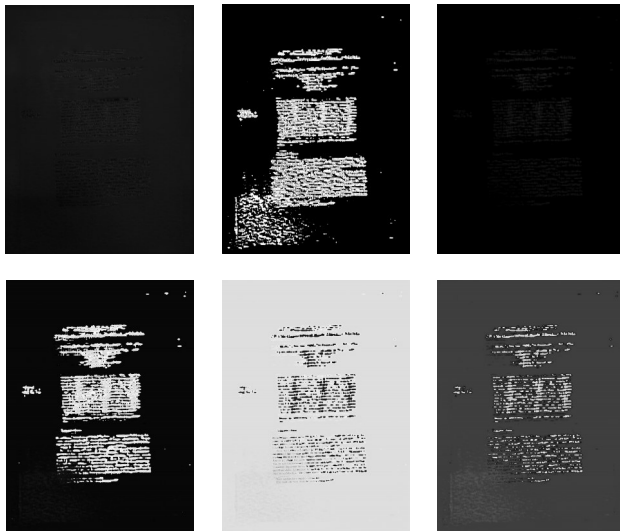
对于整体的纹理细节和色彩特征做了一定的处理, B通道的特征图经过处理之后, 在图片色彩以及文字轮廓方面都有一定程度的恢复。



(a) 不含 Gate-MLP 模块

(b) 包含 Gate-MLP 模块

图3 网络中有无 Gate-MLP 的可视化结果



(a) R 通道特征图

(b) G 通道特征图

(c) B 通道特征图

图4 经过或未经过 Gate-MLP 处理的特征图的可视化结果

Gate-MLP 的具体实现细节如图 2(b) 所示, 通过 2 个线性层 (Linear) 和 2 个 ELU 层将特征图 f_m 进行通道数扩充再拆分得到 2 个通道数一致的特征图. 其中一个特征图经 Sigmoid 函数处理后将归入 $[0, 1]$, 以表明每个局部区域的有效性并抑制噪声. 另一个特征图经 ELU 激活函数处理后, 与 Sigmoid 函数处理后的特征图逐个像素相乘, 得到滤波后的特征图. 特征过滤可以表示为如下形式:

$$GM(F) = \text{Sigmoid}(M_1(F)) \times \text{ELU}(M_2(F)) \quad (3)$$

GM 表示提出的 Gate-MLP 过滤操作. 其中, F 表示待处理的特征图 $F = \text{MLP}(f_m)$. $M_1(F)$ 和 $M_2(F)$ 表示经

过 MLP 处理后将通道数扩充到 6, 然后通过分割操作将输出在通道数上一分为二, 得到 2 个通道数为 3 的特征图.

最后, 通过稠密特征融合得到的 2 张不同曝光度图像的特征图可以表示为如下形式:

$$F_U = F_U^L + GM(F_U^L) + F_U^G + GM(F_U^G) \quad (4)$$

$$F_O = F_O^L + GM(F_O^L) + F_O^G + GM(F_O^G) \quad (5)$$

$$\text{Output} = F_U + F_O \quad (6)$$

其中, F_U 和 F_O 分别表示欠曝光图像和过曝光图像的最终特征图; F_U^L 和 F_U^G 分别表示低曝光图像的局部特征和全局特征; F_O^L 和 F_O^G 分别表示过曝光图像的局部特征和全局特征.

4 实验结果

在合成数据集和真实世界数据集上进行实验, 对所提出的方法进行评估. 对比方法是 8 种多曝光融合方法, 包括 DSIFT-EF^[19]、FMMEF^[20]、PWA^[21]、MEFAW^[22]、CF-Net (Correlation Filter Net)^[2]、IFCNN (Information Fusion and Convolutional Neural Network)^[23]、MEFNet (Managed Extensibility Framework Net)^[24] 和 U2Fusion^[25]. 在这些对比方法中, DSIFT-EF、FMMEF、PWA 和 MEFAW 被认为是传统的多曝光融合方法, 主要通过人工选取的特征和预定义的融合规则. 相比之下, CF-Net、IFCNN、MEFNet 和 U2Fusion 是基于深度学习的多曝光融合方法, 主要利用卷积神经网络或其他深度学习架构来学习特征表征并自动确定融合规则.

4.1 训练数据

首先在 SICE 数据集^[26]上进行对比, 该数据集包含 5 000 多张不同曝光级别的 UHD 图像. 数据集中的最高分辨率达到 $5\,472 \times 3\,648$ 像素. 在该数据集中, 具有不同曝光水平的多张图像对应一张真实图像. 鉴于方法是对 2 种图像进行融合, 从 SICE 数据集中选择曝光最少的图像和曝光最多的图像作为网络的欠曝光和过曝光输入. 此外, 也创建了一个名为 Doc-4K 的 UHD 文档图像数据集进行对比. Doc-4K 数据集由使用 UHD 摄像机拍摄的 500 张高质量 UHD 文档图像组成. 为了生成用于训练的欠曝光和过曝光图像, 受 Lv 等人^[27]的研究启发, 采用线性变换和伽马变换的组合, 变换数学公式如下:

$$I_{\text{out}} = \beta \cdot (\alpha \cdot I)^\gamma \quad (7)$$

其中, β 、 γ 和 α 从均匀分布中随机取样; I 表示未处理的原始图像; I_{out} 表示处理后的图像. 当符号 γ 大于 1.0 时, 图像暗部的动态范围将被压缩, α 和 β 将控制亮部的最大值, 这种方法会使正常亮度的图像变暗. 因此, 可以通过设置 $\gamma > 1$ 来生成欠曝光的图像, 通过 $\gamma < 1$ 来生成过曝光的图像. 参数的设置参考了 Lv 等人^[27]的研究:

$\alpha \sim U(0.9, 1)$, $\beta \sim U(0.5, 1)$. 对于过曝图像, 设置 $\gamma \sim U(0, 1)$, 对于欠曝图像, 设置 $\gamma \sim U(1, 5)$. 最后, 选择其中的 450 对图像作为训练数据集. 由于传统的多曝光融合算法处理 4K 图像耗时过长, 只从剩余图像中选择 50 对进行测试. SICE 数据集中的超高清图像包含丰富色彩信息; Doc-4K 数据集中的超高清图像中主要是包含丰富的纹理细节, 例如文字、线条等, 但是部分图片中还是有彩色区域出现. 在这 2 个侧重点不同的数据集上的测试结果, 可以表明所提方法不仅在色彩恢复方面表现良好, 在细节纹理恢复方面表现也很优秀.

4.2 实验设置

模型基于 PyTorch 实现, 使用 Adam 优化器进行训练, 训练时使用的批次大小为 8. 初始学习率设置为 0.001. 网络共训练了 600 轮次. 训练过程中使用 L1 损失作为损失函数. 本文所提的方法不仅能处理超高清 (4K 分辨率) 图像, 还能处理低分辨率图像. 输入图像在拍摄时的 ISO 尽量控制在 100~400 之间, 因为过低的 ISO 会导致图像暗区域过大, 过高的 ISO 会导致图像中出现大量的噪点. 由于 U2Fusion 和 CF-Net 无法直接处理 UHD 图像, 选择对原始图像进行缩减采样, 使其大小接近这些模型所能处理的最大尺寸. 就 U2Fusion 而言, 在模型处理之前, 图像会被下采样到 $1\ 024 \times 1\ 024$ 的分辨率. 然而, 对于 CF-Net, 降低的分辨率甚至更小, 仅为 512×512 . 然后, U2Fusion 和 CF-Net 生成的输出结果会通过上采样过程恢复到原始分辨率. 其他对比

方法可以直接处理 4K 图像, 而无需下采样或上采样.

除此之外, 对所有基于深度学习的模型进行了微调, 主要采用随机剪切和大小调整技术, 并将其调整到一致的大小, 作为模型的输入. 然后对这些模型得出的结果进行插值, 以获得最终输出结果, 并用于与其他方法进行比较.

4.3 评估和结果

(1) 合成数据集评估. 在 SICE 和 Doc-4K 这 2 个数据集上进行评估. 图 5、图 6 是 SICE 数据集的代表性样本结果, 图 7、图 8 是 Doc-4K 数据集的代表性样本结果. 由图 4~图 8 可知, 当只有 2 幅输入图像时, DSIFT-EF、FMMEF、MEFAW 和 PWA 等传统方法在恢复图像结构方面表现良好. 然而, 它们在颜色和边缘恢复方面却不尽人意. 此外, 这些方法无法实时处理 UHD 图像. CF-Net、IFCNN、MEFNet 和 U2Fusion 等基于深度学习的模型可以恢复一些图像细节, 但效果仍不尽如人意, 而且这些方法都无法满足 UHD 图像的实时处理需求. 而本文方法生成的融合结果 (如图 5(k)、图 6(k)、图 7(k)、图 8(k) 所示) 与真实图像 (如图 5(l)、图 6(l)、图 7(l)、图 8(l) 所示) 非常接近. 这表明提出的方法能产生高质量的融合结果. 表 1 列出了对比方法在这 2 个数据集上峰值信噪比 (Peak Signal-to-Noise Ratio, PSNR)、结构相似性 (Structural Similarity, SSIM)、多曝光结构相似性 (Multi-Exposure Image Fusion and Structural Similarity, MEF-SSIM) 和运行时间 4 个指标的定量结果, 进一步展示了所提方法的有效性.

表 1 在 Doc-4K ($3\ 000 \times 4\ 000$) 和 SICE 数据集 (最高图像分辨率为 $5\ 472 \times 3\ 648$) 上进行的 PSNR、SSIM、MEF-SSIM 和运行时间定量评估

		CF-Net	DSIFT-EF	FMMEF	IFCNN	MEFAW	MEFNet	PWA	U2Fusion	Ours
SICE	PSNR	18.620 0	13.340 0	16.520 0	19.290 0	17.070 0	18.180 0	16.150 0	19.510 0	20.030 0
	SSIM	0.686 4	0.615 1	0.624 3	0.703 2	0.659 2	0.677 9	0.632 1	0.721 2	0.768 2
	MEF-SSIM	0.792 1	0.729 4	0.734 5	0.798 5	0.756 7	0.780 1	0.751 2	0.871 8	0.843 1
	运行时间/s	—	21.330 0	55.710 0	3.180 0	79.820 0	1.230 0	60.150 0	—	0.030 0
Doc-4K	PSNR	21.330 0	12.410 0	15.990 0	22.860 0	15.820 0	18.910 0	19.890 0	16.260 0	26.880 0
	SSIM	0.844 9	0.676 9	0.762 8	0.871 3	0.827 5	0.861 5	0.907 9	0.770 3	0.932 8
	MEF-SSIM	0.899 8	0.772 8	0.825 6	0.910 5	0.881 4	0.908 9	0.948 7	0.851 5	0.973 5
	运行时间/s	—	20.310 0	50.830 0	3.070 0	71.560 0	1.120 0	54.750 0	—	0.030 0

注: 图像最好的结果加粗表示, 运行时间行中的“—”表示该方法无法处理 UHD.

(2) 真实世界数据集评估. 接下来, 将评估所提方法在真实世界欠曝光和过曝光图像上的性能. 使用实际拍摄的分辨率为 4K ($3\ 000 \times 4\ 000$) 的多重曝光图像, 并将结果与各种多曝光融合方法进行比较. 图 9 展示了一幅具有挑战性的真实世界图像的定性比较结果.

在评估过程中, 发现 MEFAW 会导致融合结果中黑色区域周围出现光晕. FMME 无法有效恢复纹理细节,

而 PWA 则在融合图像中引入了一些色彩失真. MEFNet 的鲁棒性有限, 生成的融合图像要么欠曝光, 要么过曝光. 此外, U2Fusion 恢复的图像整体颜色偏暗. 相比之下, 所提方法在恢复清晰的纹理细节和生成逼真的色彩方面表现出色, 融合效果优于上述其他方法.

4.4 消融研究

为了评估网络中每个组件的有效性, 做了一系列消融实验, 实验结果如表 2 所示. 此外, 还研究了改

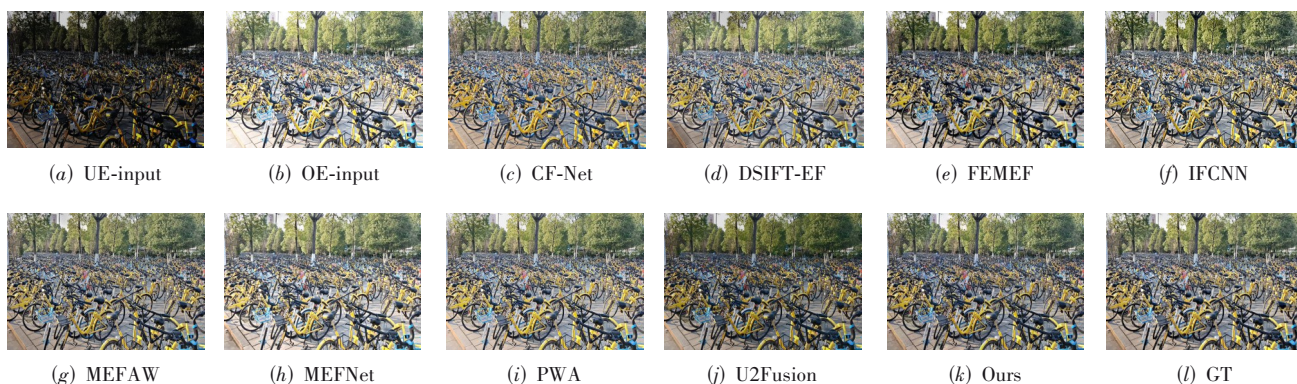


图5 SICE 数据集的融合结果(一)

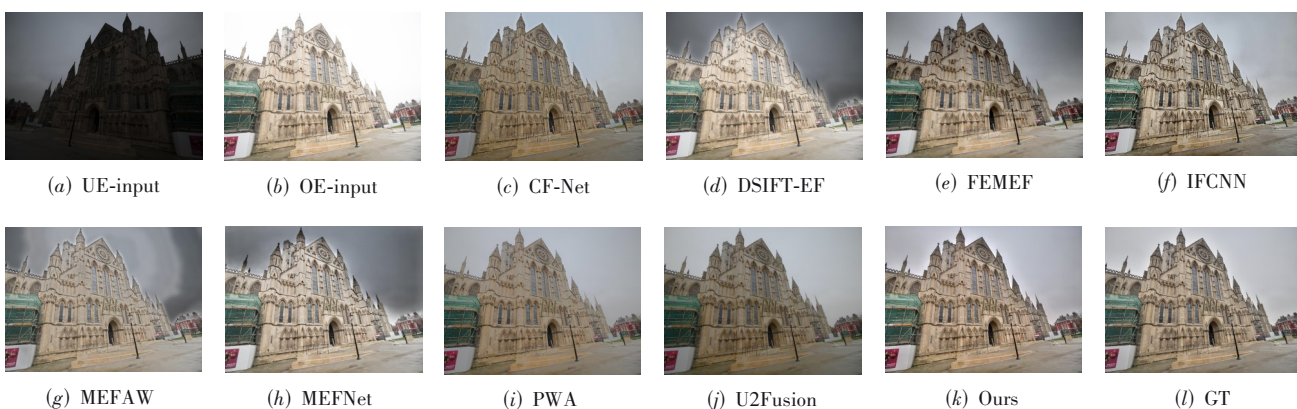


图6 SICE 数据集的融合结果(二)

变下采样图像的大小对模型性能的影响,实验结果如表3所示.

(1) Gate-MLP. 在评估表2、表3过程中,比较了有无 Gate-MLP 组件的网络. Gate-MLP 充当了输入特征图的过滤角色. 从表2所示的结果可以看出,在3个评估指标 PSNR、SSIM 和 MEF-SSIM 上,网络都优于不带 Gate-MLP 的网络.

(2) 稠密特征融合. 对有稠密特征融合的网络和无稠密特征融合的网络进行了评估. 稠密特征融合的实施使本文网络能够充分利用不同曝光度图像的局部纹理特征和全局色彩特征. 结果显示本文网络在所有3个评估指标上都优于未进行稠密特征融合的网络.

(3) 并行特征提取. 对基于双分支的并行特征提取网络和串行特征提取网络进行了比较. 并行特征提取方法可同时捕捉图像的局部纹理特征和全局色彩特征. 结果清楚地表明网络在所有3个评估指标上都优于采用串行特征提取的网络. 此外,还尝试用 Transformer 模块或 CNN 模块替换双分支中的所有特征提取模块. 从结果可以看出,与所提的网络相比,使用单一

类型特征提取模块的网络性能有所下降.

上述网络组件都不会影响模型实时处理 UHD 图像的能力. 此外,还进行了测试,以评估下采样图像的大小对模型性能的影响. 表3所示的结果表明,随着图像尺寸的增大,PSNR、SSIM 和 MEF-SSIM 指标都有所提高. 然而,这也会导致处理时间的增加,从而无法满足实时处理的需求. 目标是选择更大的向下采样尺寸,以获得更好的视觉质量,同时要确保模型仍能实时处理 UHD 图像,最终在本文方法中采用尺寸为 256×256 .

4.5 运行时间

在同一台机器上对所有深度模型进行了评估,这台机器配备了 Intel(R) Xeon(R) CPU 和 NVIDIA RTX 3090 GPU. 测量的运行时间只包括 GPU 的处理时间,不考虑任何 I/O 操作. 所有模型在 Doc-4K 和 SICE 上的平均运行时间如表1所示. 使用传统的多曝光融合方法处理一对 UHD 图像可能会非常慢. 虽然一些基于深度学习的方法可以在几秒钟内处理一对 UHD 图像,但它们无法在单 24G RAM 的 GPU 上运行. 而本文方法能够以 30 帧每秒的速度实时处理一对 UHD 图像.

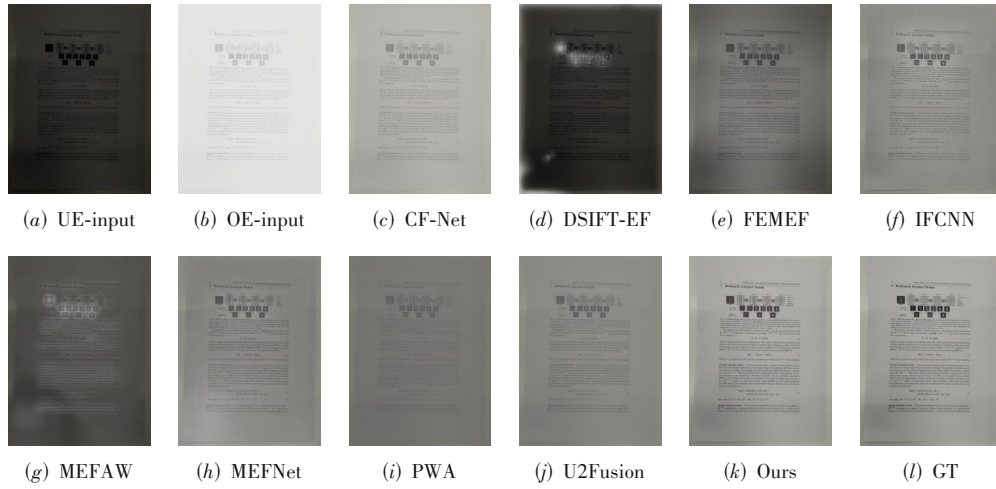


图7 Doc-4K 数据集的融合结果(一)

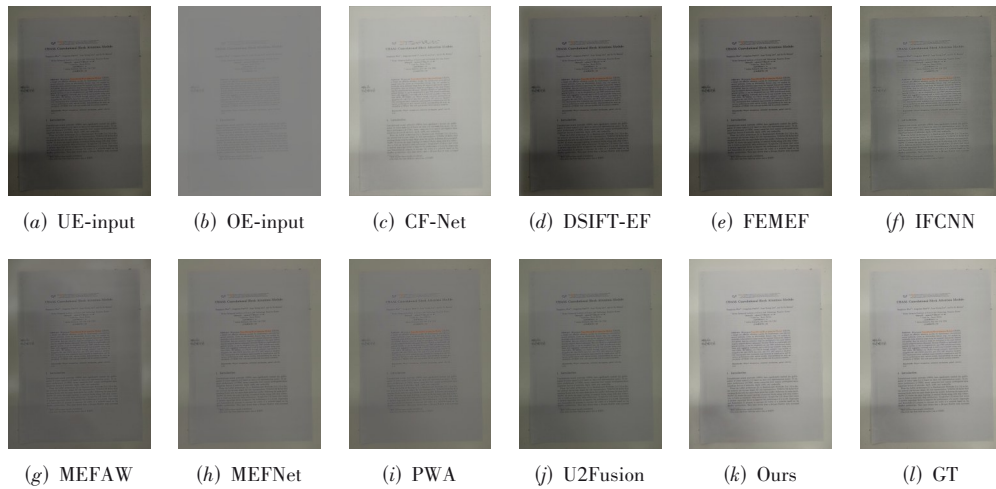


图8 Doc-4K 数据集的融合结果(二)

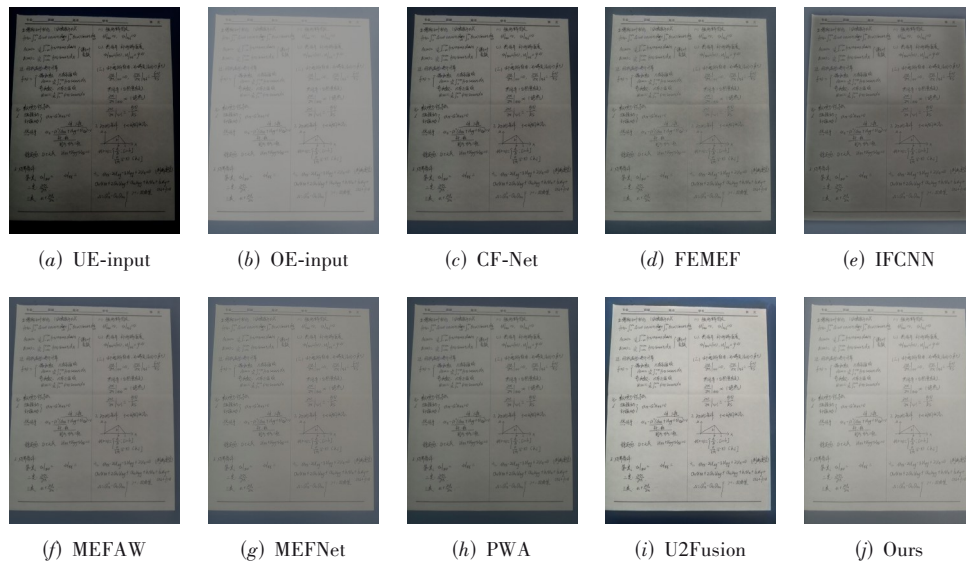


图9 真实拍摄的UHD图像的融合结果

表 2 网络中引入的每个部分的有效性实验

Size	PSNR	SSIM	MEF-SSIM	运行时间/s
64 × 64	20.92	0.927 6	0.968 1	0.01
128 × 128	23.32	0.928 7	0.968 9	0.02
256 × 256	26.88	0.932 8	0.973 5	0.03
512 × 512	26.95	0.932 2	0.975 8	0.08

表 3 下采样图像的大小对模型性能的影响

项目	PSNR	SSIM	MEF-SSIM	运行时间/s
w/o Gate-MLP	20.59	0.886 5	0.938 9	0.02
w/o Dense Fusion	19.11	0.875 4	0.927 5	0.02
w/o Dual-Branch	23.88	0.918 9	0.960 1	0.03
w/o Transformer	24.44	0.908 8	0.956 1	0.03
w/o CNN	25.17	0.911 5	0.958 9	0.03
Ours	26.88	0.932 8	0.973 5	0.03

5 结论

本文针对 UHD 图像提出了一种新颖的多曝光融合方法,该方法使用稠密的局部-全局特征融合,通过基于双分支的并行方法从输入图像中提取特征,并从不同曝光级别的局部纹理和全局色彩中聚合高密度特征,以补偿单次曝光图像中缺失的特征.进一步地,为了解决融合过程中与噪声相关的问题,设计了一种 Gate-MLP 模块来过滤提取的特征信息.此外,还专门为多曝光融合任务设计了一个新的 UHD 图像数据集.定量和定性分析表明所提的方法在效率和效果方面优于当前最先进的多曝光融合方法.虽然方法成功地恢复了局部区域的细节,但仍有改进的余地,尤其是在色彩鲜艳的区域.在未来的工作中,目标是从局部区域提取更丰富的色彩信息,避免融合图像过程中色彩失真.

参考文献

- [1] BHANDARI A K, SUBRAMANI B, VELUCHAMY M. Multi-exposure optimized contrast and brightness balance color image enhancement[J]. *Digital Signal Processing*, 2022, 123: 103406.
- [2] 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.
- [3] 张雨童, 邓欣, 徐迈. 动态场景下深度自监督多曝光图像融合方法[J]. *电子学报*, 2024, 52(1): 264-273.
ZHANG Y T, DENG X, XU M. Deep self-supervised multi-exposure image fusion for dynamic scenes[J]. *Acta Electronica Sinica*, 2024, 52(1): 264-273. (in Chinese)
- [4] QUL H, LIU S L, WANG M N, et al. TransMEF: A transformer-based multi-exposure image fusion framework using self-supervised multi-task learning[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022, 36(2): 2126-2134.
- [5] VASWANI A, SHAZEER N, PAEMAR N, et al. Attention is all you need[C]//*Proceedings of Advances in Neural Information Processing Systems*. New York: ACM, 2017: 6000-6010.
- [6] QI G Q, CHANG L, LUO Y Q, et al. A precise multi-exposure image fusion method based on low-level features[J]. *Sensors*, 2020, 20(6): 1597.
- [7] WANG Q T, CHEN W H, WU X M, et al. Detail-enhanced multi-scale exposure fusion in YUV color space[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2020, 30(8): 2418-2429.
- [8] LIU J Y, SHANG J J, LIU R S, et al. Halder: Hierarchical attention-guided learning with detail-refinement for multi-exposure image fusion[C]//2021 IEEE International Conference on Multimedia and Expo (ICME). Piscataway: IEEE, 2021: 1-6.
- [9] 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.
- [10] YANG Z G, CHEN Y P, LE Z L, et al. GANFuse: A novel multi-exposure image fusion method based on generative adversarial networks[J]. *Neural Computing and Applications*, 2021, 33(11): 6133-6145.
- [11] JUNG H, KIM Y, JANG H, et al. Unsupervised deep image fusion with structure tensor representations[J]. *IEEE Transactions on Image Processing*, 2020, 29: 3845-3858.
- [12] HU X B, REN W Q, YU K C, et al. Pyramid architecture search for real-time image deblurring[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2021: 4278-4287.
- [13] LIANG J, ZENG H, ZHANG L. High-resolution photorealistic image translation in real-time: A Laplacian pyramid translation network[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 9392-9400.
- [14] LIN S C, RYABTSEV A, SENGUPTA S, et al. Real-time high-resolution background matting[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 8758-8767.
- [15] LIN Q W N, ZHENG Z R, JIA X Y. UHD low-light image enhancement via interpretable bilateral learning[J]. *Information Sciences*, 2022, 608: 1401-1415.

- [16] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015. Cham: Springer International Publishing, 2015: 234-241.
- [17] ZAMIR S W, ARORA A, KHAN S, et al. Restormer: Efficient transformer for high-resolution image restoration[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2022: 5718-5729.
- [18] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[C]//Proceedings of the International Conference on Learning Representations, Dublin: ICLR, 2021: 1-21.
- [19] LIU Y, WANG Z F. Dense SIFT for ghost-free multi-exposure fusion[J]. Journal of Visual Communication and Image Representation, 2015, 31: 208-224.
- [20] 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, 29: 5805-5816.
- [21] MA K D, WANG Z. Multi-exposure image fusion: A patch-wise approach[C]//2015 IEEE International Conference on Image Processing (ICIP). Piscataway: IEEE, 2015: 1717-1721.
- [22] LEE S H, PARK J S, CHO N I. A multi-exposure image fusion based on the adaptive weights reflecting the relative pixel intensity and global gradient[C]//2018 25th IEEE International Conference on Image Processing (ICIP). Piscataway: IEEE, 2018: 1737-1741.
- [23] ZHANG Y, LIU Y, SUN P, et al. IFCNN: A general image fusion framework based on convolutional neural network[J]. Information Fusion, 2020, 54: 99-118.
- [24] MA K D, DUANMU Z F, ZHU H W, et al. Deep guided learning for fast multi-exposure image fusion[J]. IEEE Transactions on Image Processing, 2019, 29: 2808-2819.
- [25] XU H, MA J Y, JIANG J J, et al. U2Fusion: A unified unsupervised image fusion network[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(1): 502-518.
- [26] CAI J R, GU S H, ZHANG L, et al. Learning a deep single image contrast enhancer from multi-exposure images[J]. IEEE Transactions on Image Processing, 2018, 27(4): 2049-2062.
- [27] LV F F, LI Y, LU F. Attention guided low-light image enhancement with a large scale low-light simulation dataset[J]. International Journal of Computer Vision, 2021, 129(7): 2175-2193.

作者简介



贾修一 男,1983年出生于山东省日照市.教授,博士生导师.主要研究方向为机器学习、数据挖掘、计算机视觉.
E-mail: jiaxy@njust.edu.cn



郑卓然 男,1992年出生于山东省枣庄市.毕业于南京理工大学并获博士学位.主要研究方向为图像增强、计算机视觉.
E-mail: zhengzr@njust.edu.cn



林乔万尼 男,1999年出生于浙江省温州市.毕业于南京理工大学并获硕士学位.主要研究方向为图像处理、计算机视觉.
E-mail: qwnlin@njust.edu.cn



石争浩 男,1968年出生于陕西省富平县.教授,博士生导师.主要研究方向为机器视觉、医学图像处理及机器学习.
E-mail: ylshi@xaut.edu.cn