

# 光电工程

## Opto-Electronic Engineering

中文核心期刊 中国科技核心期刊  
Scopus CSCD

### LF-UMTI: 基于多尺度空角交互的无监督多曝光光场图像融合

李玉龙, 陈晔曜, 崔跃利, 郁梅

#### 引用本文:

李玉龙, 陈晔曜, 崔跃利, 等. LF-UMTI: 基于多尺度空角交互的无监督多曝光光场图像融合[J]. *光电工程*, 2024, 51(6): 240093.

Li Y L, Chen Y Y, Cui Y L, et al. LF-UMTI: unsupervised multi-exposure light field image fusion based on multi-scale spatial-angular interaction[J]. *Opto-Electron Eng*, 2024, 51(6): 240093.

<https://doi.org/10.12086/oe.2024.240093>

收稿日期: 2024-04-23; 修改日期: 2024-06-01; 录用日期: 2024-06-03

## 相关论文

### 角度差异强化的光场图像超分网络

吕天琪, 武迎春, 赵贤凌

*光电工程* 2023, 50(2): 220185 doi: [10.12086/oe.2023.220185](https://doi.org/10.12086/oe.2023.220185)

### 基于多尺度特征融合的遥感图像小目标检测

马梁, 苟于涛, 雷涛, 靳雷, 宋怡萱

*光电工程* 2022, 49(4): 210363 doi: [10.12086/oe.2022.210363](https://doi.org/10.12086/oe.2022.210363)

### 伪标签细化引导的相机感知无监督行人重识别方法

程思雨, 陈莹

*光电工程* 2023, 50(12): 230239 doi: [10.12086/oe.2023.230239](https://doi.org/10.12086/oe.2023.230239)

更多相关论文见光电期刊集群网站 



<http://cn.ojournal.org/oe>



 OE\_Journal



Website

DOI: 10.12086/oe.2024.240093

# LF-UMTI: 基于多尺度空角交互的无监督多曝光光场图像融合

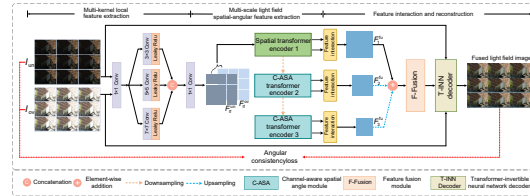
李玉龙<sup>1</sup>, 陈晔曜<sup>1</sup>, 崔跃利<sup>2</sup>, 郁梅<sup>1\*</sup><sup>1</sup>宁波大学信息科学与工程学院, 浙江 宁波 315211;<sup>2</sup>台州学院电子信息与工程学院, 浙江 台州 318000

**摘要:** 光场成像可同时捕获真实场景中光线的强度和方向信息。但受限于成像传感器的势阱容量, 现光场相机单曝光捕获的光场图像难以完整记录真实场景中所有的细节信息。为了解决上述问题, 本文提出了一种基于多尺度空角交互的无监督多曝光光场成像方法。该方法采用多尺度空角交互策略, 以有效提取光场空角特征, 同时利用通道维上建模策略以降低计算量来适应光场高维结构。其次, 构建了由可逆神经网络导向的光场重建模块, 以避免融合伪影并恢复更多细节信息。最后, 设计了一种角度一致性损失, 其考虑了边界子孔径图像和中心子孔径图像之间的视差变化, 以保证融合结果的视差结构。为评估所提方法的性能, 建立了一个面向真实场景的多曝光光场基准数据集。实验结果表明, 所提方法可在保证角度一致性的前提下重建出具备高对比度和丰富细节的光场图像。与现有方法相比, 所提方法在客观质量和主观视觉两方面均取得更好的结果。

**关键词:** 光场成像; 多曝光融合; 多尺度空角交互; 无监督学习; 角度一致性

中图分类号: TP394.1

文献标志码: A



李玉龙, 陈晔曜, 崔跃利, 等. LF-UMTI: 基于多尺度空角交互的无监督多曝光光场图像融合 [J]. 光电工程, 2024, 51(6): 240093

Li Y L, Chen Y Y, Cui Y L, et al. LF-UMTI: unsupervised multi-exposure light field image fusion based on multi-scale spatial-angular interaction[J]. *Opto-Electron Eng*, 2024, 51(6): 240093

## LF-UMTI: unsupervised multi-exposure light field image fusion based on multi-scale spatial-angular interaction

Li Yulong<sup>1</sup>, Chen Yeyao<sup>1</sup>, Cui Yueli<sup>2</sup>, Yu Mei<sup>1\*</sup><sup>1</sup>Faculty of Information Science and Engineering, Ningbo University, Ningbo, Zhejiang 315211, China;<sup>2</sup>School of Electronic and Information Engineering, Taizhou University, Taizhou, Zhejiang 318000, China

**Abstract:** Light field imaging can simultaneously capture the intensity and direction information of light in a real-world scene. However, due to the limited capacity of imaging sensors, light field images captured with a single exposure struggle to fully record all the details in the real scene. To address the aforementioned issue, an unsupervised multi-exposure light field imaging method based on multi-scale spatial-angular interactions is

收稿日期: 2024-04-23; 修回日期: 2024-06-01; 录用日期: 2024-06-03

\*通信作者: 郁梅, yumei@nbu.edu.cn

版权所有©2024 中国科学院光电技术研究所

proposed in this paper. A multi-scale spatial-angular interaction strategy is adopted to effectively extract spatial-angular features of the light field. Additionally, a channel-wise modeling strategy is employed to reduce computational complexity and adapt to the high-dimensional structure of the light field. Furthermore, a light field reconstruction module guided by reversible neural networks is constructed to avoid fusion artifacts and recover more detailed information. Lastly, an angle consistency loss is designed, considering the disparity variations between boundary sub-aperture images and the central sub-aperture image, to ensure the disparity structure of the fusion result. To evaluate the performance of the proposed method, a benchmark dataset for multi-exposure light field imaging is created, targeting real-world scenes. Experimental results demonstrate that the proposed method can reconstruct light field images with high contrast and rich details while ensuring angular consistency. Compared with the existing methods, the proposed method achieves superior results in both objective quality and subjective visual perception.

**Keywords:** light field imaging; multi-exposure fusion; multi-scale spatial-angular interaction; unsupervised learning; angular consistency

## 1 引言

随着计算摄像的发展,新的成像技术层出不穷,光场(light field, LF)成像技术就是其中的典型代表。光场相机通过将微透镜阵列放置在主透镜后面的设计,可同时捕获入射光线的强度和方向信息以完备记录光线分布。受益于其捕获的丰富信息,光场在深度估计<sup>[1-3]</sup>、语义分割<sup>[4]</sup>、新视图合成<sup>[5]</sup>等众多领域有着广泛应用。最近,消费级光场相机的兴起使得用户可以轻松地获取到不同角度分辨率的光场图像。但为获得丰富的角度信息,微透镜阵列中子透镜只拥有非常小的孔径尺寸,导致采集的光场图像的动态范围非常有限,容易出现欠曝光或过曝光现象,阻碍了光场图像在现实应用中的优势。

作为一种经济高效的高动态范围成像解决方案,多曝光融合(multi-exposure fusion, MEF)旨在通过多次曝光保留图像中感知最佳的部分,同时调整整体曝光不足和过度曝光的区域。根据现有理论,现有的传统MEF方法大致可分为三类:变换域方法、空间域方法和基于深度学习的方法<sup>[6]</sup>。变换域方法旨在融合变换域空间中的不同系数,主要分为三个阶段组成:图像变换、系数融合和逆变换。空间域方法可以充分考虑不同像素之间的关系,因此可以获得比变换域方法更均匀的光照分布结果。Li等人提出了一种易于实现的空间域MEF方法,使用均值滤波器来分离图像基础层和细节层,并采用引导滤波器来引导不同子层的对应融合<sup>[7]</sup>。Liu等人提出了一种基于密集尺度不变特征变换的MEF方法,其方法既可以处理静态场景,也可以在动态场景中得到应用<sup>[8]</sup>。Lee等人定义

了基于整体亮度和全局梯度构造的两个权重函数,在此基础上构造了一种自适应权重多曝光融合算法<sup>[9]</sup>。Ulucan等人设计了一种利用主成分分析、自适应曝光区域和显著图生成无鬼影的权重图框架,可以获得较好的视觉融合效果<sup>[10]</sup>。此外,为了提高光场图像的动态范围,直接使用普通多曝光图像的方法来获取4D光场图像中的单幅子孔径图像的传统算法相继出现<sup>[11-12]</sup>。这类方法可以看作是对不同角度视点分别进行处理,因而可能会破坏光场角度一致性。Le Pendu等人提出了一种基于加权低秩近似的高动态范围光场算法,利用密集视点采样高动态光场。但该方法只考虑恢复饱和区域的信息,无法恢复场景低暗区域的细节<sup>[13]</sup>。上述传统方法虽然可以在一定程度上提高光场图像的动态范围,但对提取的特征进行建模以及设计合理有效的融合规则都极其耗时和费力,限制了MEF的性能。近年来,基于深度学习的多曝光图像融合研究越来越受到重视,通过构建深度学习的端到端模型,可以自适应地融合和重建出视觉良好的图像。

## 2 相关工作

基于深度学习的多曝光融合方法可以分为有监督和无监督二类。有监督的多曝光融合方法中参考图像的获取通常是从多种算法的融合结果中挑选主观打分较高的图像作为真实标签。按照这种方法,Yin等人提出了一种基于编码器-解码器的方法来融合具有不同曝光的两幅图像<sup>[14]</sup>。Xu等人提出了一种基于生成对抗网络的方法来实现动态场景的多曝光融合<sup>[15]</sup>。

Liu 等人首次将自监督中的常用对比学习方法引入多曝光融合领域中, 设计了一种基于整体和局部的对比学习网络, 借助新构造的正负样本的损失函数实现了的良好融合性能<sup>[16]</sup>。Liu 等人提出了一种注意力引导的全局和局部对抗学习网络, 其方法利用 Retinex 理论设计了一个光照网络来估计不同场景下的光照以增强融合后图像的颜色<sup>[17]</sup>。基于学习的光场高动态范围重建任务也逐渐得到关注。Chen 等人提出了一种同时增强光场视角和动态范围的成像算法, 通过配准、曝光恢复和多曝光图像融合三个子任务最终实现高质量的光场图像重建<sup>[18]</sup>。

现有的基于深度学习的有监督多曝光融合方法已经取得了一些进展, 但仍存在一些问题有待解决。首先, 基于有监督学习的方法性能在很大程度上取决于高质量的参考图像, 但多曝光融合方法中的参考图像通常需要耗费大量时间才能获取。因此, 开发无监督的多曝光融合算法更具有实用性。DeepFuse 是首次利用无监督深度学习方法解决 MEF 问题的尝试<sup>[19]</sup>。Ma 等人提出了一种快速灵活的 MEF 方法, 通过将下采样的源图像输入网络来生成权重图<sup>[20]</sup>。Qu 等人用 Transformer 模块与 CNN 模块提取的特征相加的设计来兼顾关注局部和全局信息, 而且经过三个根据多曝光融合特点设计的自监督图像重建任务来让网络充分学习到多曝光融合所需的特征<sup>[21]</sup>。Zheng 等人利用自适应引导滤波器构建新的融合范式的网络, 并提出了一种在梯度域上计算的图像保真度损失函数<sup>[22]</sup>。近来, 对比学习的思想也被引入到无监督的多曝光算法当中。Xu 等人提出了一种基于对比学习的无监督多曝光图像融合方法, 不再受到需要真实标签构造正样本的限制, 且减少了源图像不当曝光和颜色对最终融合图像的影响<sup>[23]</sup>。Zhang 等人深入探究了成像场景中纹理结构、光照分布和可见颜色分布跟最终融合图像的反射率、阴影和颜色分量之间的关系, 设计了一种基于图像分解的无监督多曝光融合网络<sup>[24]</sup>。

以上融合方法是专门针对多曝光融合任务所设计, 还有几种基于通用深度学习的方法来服务于各种图像融合任务。Xu 等人提出了名为 FusionDN 的融合网络, 将不同的图像融合任务用一个统一的密集连接网络实现, 并通过应用弹性权重共享来避免在连续训练多个任务时忘记从先前任务中学到的先验知识<sup>[25]</sup>。U2Fusion 网络对 FusionDN 信息保留的分配策略进行了改进, 其信息保留程度是通过提取的特征进行信

息测量来分配的, 而不再是基于源图像中的信息测量<sup>[26]</sup>。另一种类似于 FusionDN 的方法是 PMGI 网络, 它也以统一的方式实现了各种图像融合任务。区别在于 PMGI 将图像融合问题统一为源图像的梯度和强度比例保持问题<sup>[27]</sup>。

综上所述, 普通图像的多曝光融合算法在有监督与无监督方法上都已经取得了较大的进步, 但目前利用深度学习网络解决光场相机动态范围较小的问题并同时兼顾光场的高维数据特性的方法还较少。而且, 现有光场多曝光融合方法在处理极端曝光的场景时容易出现细节纹理弱、对比度低和角度一致性问题。此外, 尽管现有的基于有监督学习的多曝光融合算法性能得到较大程度的提高, 但其训练过程极其依赖真实值, 然而高质量真实光场多曝光融合图像的获取仍然是一个挑战。

为了解决上述问题, 本文提出了一种基于多尺度空角交互的无监督多曝光光场图像融合方法, 在不依赖真实高动态范围 (high dynamic range, HDR) 光场图像的前提下, 实现了高对比度和丰富细节的多曝光光场图像融合。本文主要贡献包括: 1) 提出了通道建模的多尺度空间角度的光场特征提取网络, 且利用不同尺度宏像素区域覆盖不同视差范围的特点来捕获不同视差范围的角度特征; 2) 设计了一种基于可逆神经网络与 Transformer 的光场重建网络, 利用可逆神经网络无损信息传输的特性恢复出更多的光场图像细节; 3) 为了确保融合过程中角度信息不发生丢失, 从中心子孔径与周围子孔径的关系出发设计了角度一致性损失函数。

### 3 所提出方法

目前主流光场数据可视化方法包括子孔径图像 (sub-aperture images, SAI) 和微透镜图像 (macro-pixel image, MacPI)。通过固定光场角度维度, 光场图像可以以子孔径图像阵列的形式来可视化, 每幅子孔径图像对应于一个视点, 因而它们之间具有视差关系。类似地, 通过固定光场空间维度, 光场图像可以可视化为包含角度信息的微透镜图像, 其中的每个宏像素可视作由 SAI 中坐标位置相同的像素点所构成的集合。本文根据两种不同光场可视化数据的特点, 从子孔径图像来提取空间信息, 而微透镜图像中的宏像素则被用来提取光场的角度信息。

图 1 给出了所提出的无监督多曝光光场融合的总

体框架, 其主要包括多核局部特征提取模块, 多尺度光场空角特征提取模块以及光场图像特征交互与重建模块三部分。首先, 通过不同感受野的卷积核提取不同曝光值下的子孔径阵列的局部信息。然后, 引入空角全局特征提取模块来获取光场空间和角度维度的多层次的全局信息。不同尺度下的光场欠曝光与过曝光特征会被送入光场特征交互模块来获得不同层次的融合特征。最后, 融合后的特征经过利用可逆神经网络构建的光场解码器重建出细节丰富且对比度高的光场图像。

### 3.1 多核局部特征提取模块

为了解决 Transformer 模型缺少局部建模的问题, 本文采用卷积核大小不同的卷积层获取子孔径图的局部特征, 同时映射到特征空间的光场特征图, 以直接用作后续空角 Transformer 编码器的输入。如图 1 所示, 输入的子孔径图经过  $1 \times 1$  卷积后会经过三条并联且卷积核大小依次递增的多核卷积层, 获取到的不同感受野的局部特征经过通道间的相加最终得到初始光场特征, 表示为  $F_{if}^i \in \mathbb{R}^{(U \times V) \times H \times W \times C}$ , 其中  $U, V$  表示光场的角度维度,  $H, W$  表示光场的空间维度,  $C$  表示通道维度,  $i=ov$  或  $un$ , 分别表示输入的过曝光和欠曝光光场图像。

### 3.2 多尺度光场空角特征提取模块

由于实际计算资源的限制, 初始的 Transformer 无法直接提取光场图像的空间和角度特征。为了解决这个限制, 本文采用通道建模的 Transformer 来避免高分辨率光场需要消耗庞大计算资源的问题, 并利用多尺度的模块结构获取不同视差范围内的角度特征,

同时, 特征下采样的操作可以进一步减少模块的复杂度。

空间 Transformer 编码器利用在通道维度上建模的自注意力机制对输入的光场数据进行空间维度的处理。子孔径特征需要产生相应的嵌入特征  $Q_i^n \in \mathbb{R}^{(UV) \times (HW) \times C}, K_i^n \in \mathbb{R}^{(UV) \times C \times HW}, V_i^n \in \mathbb{R}^{(UV) \times (HW) \times C}, n=1, 2, 3$ , 其过程可以表示为

$$Q_i^n = W_d^Q F_{if}^{i,n}, K_i^n = W_d^K F_{if}^{i,n}, V_i^n = W_d^V F_{if}^{i,n}, \quad (1)$$

其中:  $W_d^{(\cdot)} \in \mathbb{R}^{C \times C}$  表示  $1 \times 1$  卷积与  $3 \times 3$  深度可分离卷积级联的模块。

然后, 在通道维度计算  $Q_i^n$  与  $K_i^n$  之间大小为  $\mathbb{R}^{C \times C}$  的自注意力图。相比于常规空间建模需要计算大小为  $\mathbb{R}^{HW \times HW}$  的自注意力图, 该方法可减少大量计算资源的消耗。具体计算过程可以表示为

$$Attention(Q_i^n, K_i^n, V_i^n) = softmax\left(\frac{Q_i^n (K_i^n)^T}{\alpha}\right) V_i^n, \quad (2)$$

$$\tilde{F}_{if}^{i,n} = W_p Attention + F_{if}^{i,n}, \quad (3)$$

其中:  $F_{if}^{i,n}$  表示输入的光场特征图,  $\tilde{F}_{if}^{i,n}$  表示输出的光场特征图;  $\alpha$  是一个可学习的缩放参数, 用来控制自注意力图的大小,  $W_p^{(\cdot)}$  表示  $1 \times 1$  卷积。

在前馈网络中引入深度可分离卷积, 以对空间相邻像素位置的信息进行编码, 这有助于弥补在 Transformer 中局部信息缺少的问题。具体计算过程可以表示为

$$\tilde{F}_{if}^{i,n} = W_p^{i,n} (GELU(W_d^1(LN(\tilde{F}_{if}^{i,n}))) \odot W_d^2(LN(\tilde{F}_{if}^{i,n}))) + \tilde{F}_{if}^{i,n}, \quad (4)$$

其中:  $GELU$  表示非线性激活函数,  $\odot$  表示元素相乘,  $\tilde{F}_{if}^{i,n}$  表示提取的光场空间特征图。

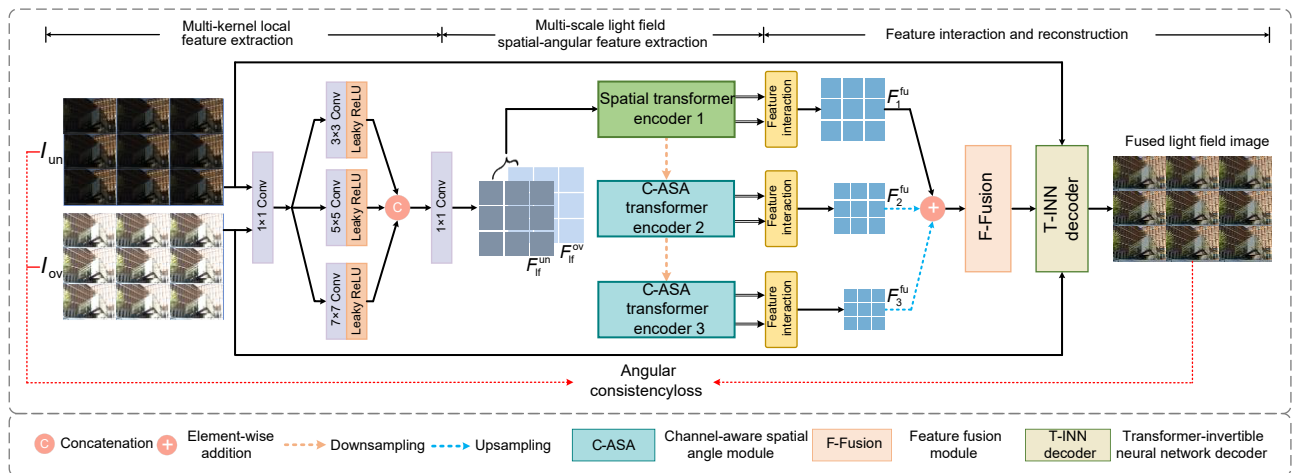


图 1 本文方法整体网络框图  
Fig. 1 Overall network diagram of the proposed method

本文为了提高对光场视差变化的鲁棒性, 采用了在不同尺度的宏像素区域内进行全局角度交互, 而不是单个宏像素, 并以此构建了 C-ASA Transformer 编码器来处理光场数据的角度信息, 其具体结构如图 2 所示。光场空间特征在送入 C-ASA Transformer 编码器前需要从子孔径图像转化为微透镜图像形式。将微透镜形式的光场图像下采样 $\lambda$ 倍后, 由 $M \times M$ 个宏像素组成的正方形区域内的有效视差范围为 $[-\lambda M, \lambda M]$ 。随着 $M$ 的增加, 可以获得更大的视差范围, 但大宏像素区域包含的像素数量会给网络的训练过程带来沉重的计算负担。同时, 考虑到匹配所拍摄的多曝光光场数据集的视差范围, 本文在每个尺度分支上所选取的 $M$ 均为 1。具体来说, 在送入 C-ASA Transformer 编码器进行特征提取之前, 微透镜形式的光场特征上 $M \times M$ 个宏像素会被聚合成大小为 $MU \times MV$ 的子区域。然后, 将这些宏像素子区域展平为一维序列 $F_M^{i,n} \in \mathbb{R}^{(M^2 UV) \times C \times (HW/M^2)}$ 输入到后续的角度特征提取模块。角度 Transformer 和空间 Transformer 的处理基本类似, 但由于光场角度维度的自注意力图的计算量通常较小, 因此只采用了基本的 Transformer 网络结构来实现。

具体计算过程可以表示为

$$Q_{i,N}^M = W_s^{Q,N} F_M^{i,N}, K_{i,N}^M = W_s^{K,N} F_M^{i,N}, V_{i,N}^M = W_s^{V,N} F_M^{i,N}, \quad (5)$$

$$Attention(Q_{i,N}^M, K_{i,N}^M, V_{i,N}^M) = softmax(\frac{Q_{i,N}^M K_{i,N}^M}{\alpha}) V_{i,N}^M, \quad (6)$$

$$\tilde{F}_M^{i,N} = W_p Attention + F_M^{i,N}, \quad (7)$$

$$\bar{F}_M^{i,N} = W_p(GELU(DWConv(W_p(LN(\tilde{F}_M^{i,N})))))) + \tilde{F}_M^{i,N}, \quad (8)$$

其中:  $W_s^{Q,N}$ 、 $W_s^{K,N}$ 、 $W_s^{V,N}$ 、 $W_p \in \mathbb{R}^{C \times C}$  均表示线性层,  $DWConv$  表示深度可分离卷积,  $\bar{F}_M^{i,N} \in \mathbb{R}^{(M^2 UV) \times C \times (HW/M^2)}$  表示第 $N$ 个尺度分支下角度 Transformer 的输出,  $N=2, 3$ 。最后, 对输出后的序列进行转化得到子孔径形式的特征输出 $\hat{F}_i^n \in \mathbb{R}^{(UV) \times H \times W \times C}$ 。通过这种方式, 多尺度空角特征提取模块可以覆盖不同光场图像的视差变化。

### 3.3 光场图像特征交互与重建网络

为了获得高质量的多曝光融合光场图像, 需要自适应地聚合来自欠曝光和过曝光光场图像中的互补信息, 以获得有效的光场融合特征。为此, 本文引入交叉注意力融合方法来实现光场特征的交互。如图 3 所

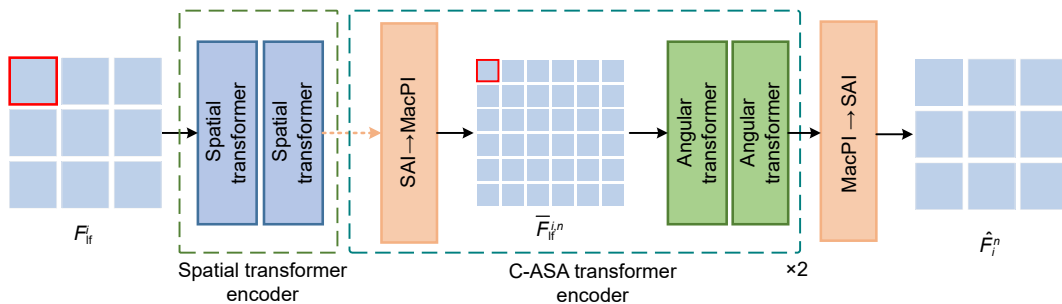


图 2 光场空间角度特征提取模块

Fig. 2 Light field spatial angular feature extraction module

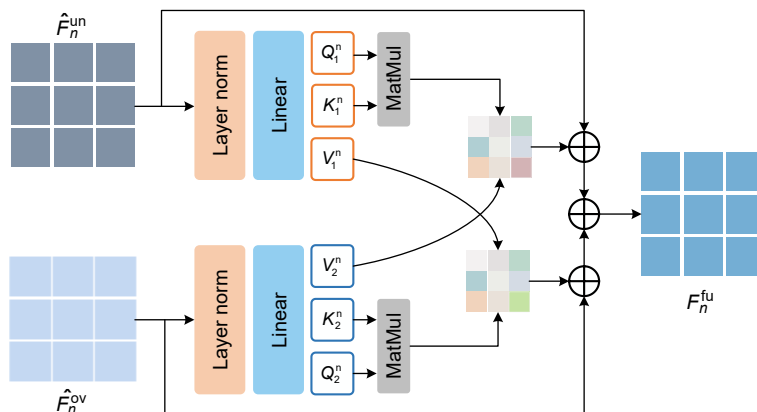


图 3 光场信息交互模块

Fig. 3 Light field information interaction module

示, 经过空角特征提取网络得到的不同尺度下的特征 ( $\hat{F}_n^{ov}, \hat{F}_n^{un}$ ) 首先被送入归一化层, 然后将它们输入  $3 \times 3$  深度可分离卷积产生形状为  $\mathbb{R}^{(U \times V) \times (H \times W) \times C}$  的张量, 即从  $\hat{F}_n^{ov}$  中产生  $(Q_{ov}^{C,n}, K_{ov}^{C,n}, V_{ov}^{C,n})$ , 从  $\hat{F}_n^{un}$  中产生  $(Q_{un}^{C,n}, K_{un}^{C,n}, V_{un}^{C,n})$ , 然后通过交换不同曝光输入之间的  $Q_i^{C,n}$  值实现光场信息的交互。利用交换后的  $Q_{ov}^n \in \mathbb{R}^{(UV) \times (HW) \times C}$  与转置后的  $K_{un}^n \in \mathbb{R}^{(UV) \times C \times (HW)}$  计算其大小为  $\mathbb{R}^{C \times C}$  注意力图, 它代表的是两个不同特征之间的相似度, 然后与  $V_{un}^n$  相乘后用来选择其中的互补特征。最后, 将所选取的特征形状转化为子孔径形式并和输入相加得到互补后的特征。具体计算过程可以表示如下:

$$F_n^{fu} = \eta(G_{ov}^n + G_{un}^n) + \hat{F}_n^{un} + \hat{F}_n^{ov}$$

$$G_{ov}^n = \text{softmax}\left(\frac{Q_{ov}^{C,n} K_{un}^{C,n}}{\alpha}\right) V_{un}^{C,n}, \quad (9)$$

$$G_{un}^n = \text{softmax}\left(\frac{Q_{un}^{C,n} K_{ov}^{C,n}}{\alpha}\right) V_{ov}^{C,n}$$

其中:  $G_{ov}^n$ 、 $G_{un}^n$  均表示经过信息交互后的光场特征,

$F_n^{fu}$  表示最终信息交互后输出的光场特征,  $\eta = 0.5$ 。

三个不同尺度的光场信息交互分支会产生三种大小不同的光场交互特征, 这些光场特征经过上采样转化成相同大小。然后, 如图 4(a) 所示, 特征融合模块根据每个分支特征的重要程度来有选择性地保留相应的特征。经过选择性保留的光场特征  $F^S \in \mathbb{R}^{(UV) \times H \times W \times C}$  被送入光场重建模块。考虑到重建特征中的边缘和纹理信息对于图像融合任务非常重要, 而可逆神经网络可以通过网络输入和输出特征相互生成, 使得输入信息能够得到更好的保留, 因此, 本文提出了一种结合可逆神经网络与 Transformer 的光场重建模块, 即图 1 中的 T-INN Decoder(transformer-invertible neural network decoder), 具体结构如图 5 所示。首先通过通道建模的 Transformer 和卷积模块以及通道注意力获得重建特征  $R_k^s$ 。此外, 在重建特征送入后续网络之前, 由密集卷积模块组成的残差连接会与之相加, 以避免梯度爆炸并加快网络收敛速度。重建特征被送入可逆神经网络解码器中, 在每个可逆层中特征传递的方式

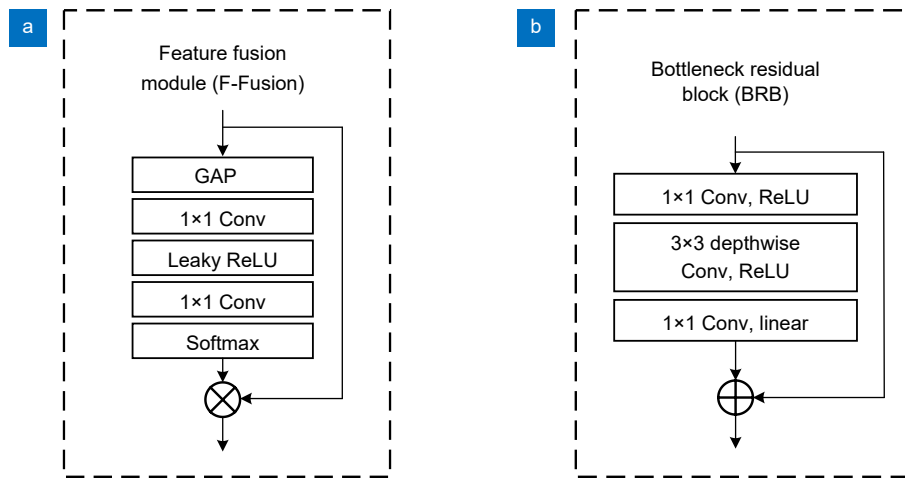


图 4 (a) 特征融合模块; (b) 瓶颈残差模块

Fig. 4 (a) Feature fusion module; (b) Bottleneck residual module

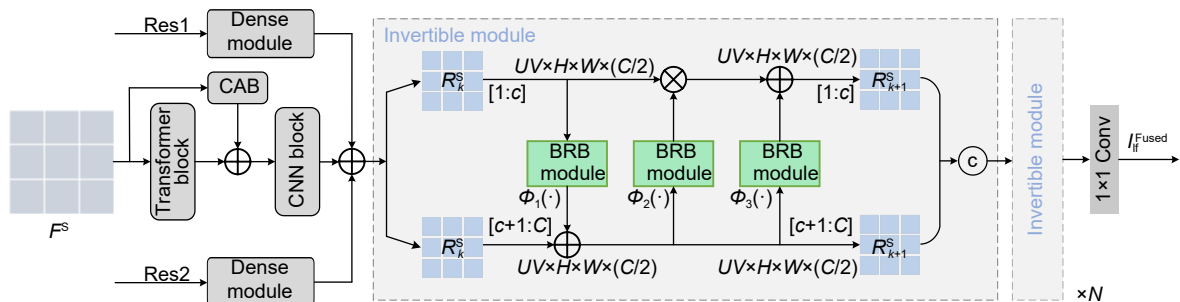


图 5 光场重建模块

Fig. 5 Light field reconstruction module

可以表示为<sup>[28]</sup>

$$\begin{aligned} R_{k+1}^S[c+1:C] &= R_k^S[c+1:C] + \Phi_1(R_k^S[1:c]), \\ R_{k+1}^S[1:c] &= R_k^S[1:c] \odot \exp(\Phi_2(R_{k+1}^S[c+1:C])) \\ &\quad + \Phi_3(R_{k+1}^S[c+1:C]), \\ R_{k+1}^S &= \text{Concat}\{R_{k+1}^S[1:c], R_{k+1}^S[c+1:C]\}, \end{aligned} \quad (10)$$

其中:  $R_k^S[1:c]$  是第  $k$  个可逆层输入前  $c$  个通道的特征值,  $\text{Concat}$  表示通道级联运算,  $\Phi_1$ 、 $\Phi_2$ 、 $\Phi_3$  均采用瓶颈残差模块 (bottleneck residual module, BRB) 作为特征映射层, 具体结构在图 4(b) 中给出。最后, 使用一层  $1 \times 1$  卷积将特征通道数变为 1 实现光场图像的重建。

### 3.4 损失函数

本方法的损失函数包含两个主要部分, 即融合损失  $L_{\text{fuse}}$  和角度一致性损失  $L_{\text{con}}$ 。  $L_{\text{fuse}}$  包括四个普通图像融合中常用的损失函数, 即多曝光结构相似性损失 (MEFSSIM)<sup>[29]</sup>、像素损失、结构相似性损失 (SSIM)<sup>[30]</sup> 和梯度损失, 可以表示为

$$L_{\text{fuse}} = L_{\text{mefssim}} + L_{\text{grad}} + \alpha L_{\text{pixel}} + \beta L_{\text{ssim}}, \quad (11)$$

其中:  $\alpha$ 、 $\beta$  为权重参数。

融合损失当中,  $L_{\text{pixel}}$ 、 $L_{\text{ssim}}$  损失主要保证融合结果不偏离输入光场图像的基本内容。此外, 考虑到像素损失容易使融合结果中高频纹理变得模糊, 因此增加梯度损失来保证相应的细节信息恢复。它们均在光场图像的每个子孔径图像下定义, 表达如下:

$$\begin{aligned} L_{\text{mefssim}} &= 1 - \frac{1}{2UV} \\ &\quad \cdot \sum_{u=1}^U \sum_{v=1}^V (\text{MEFSSIM}(I_{\text{F}}^{(u,v)}, I_{\text{ov}}^{(u,v)}) + \text{MEFSSIM}(I_{\text{F}}^{(u,v)}, I_{\text{un}}^{(u,v)})), \end{aligned} \quad (12)$$

$$L_{\text{grad}} = \frac{1}{UV} \sum_{u=1}^U \sum_{v=1}^V (\|\nabla I_{\text{F}}^{(u,v)}\| - \max(\|\nabla I_{\text{ov}}^{(u,v)}\|, \|\nabla I_{\text{un}}^{(u,v)}\|)), \quad (13)$$

$$L_{\text{pixel}} = \frac{1}{n} \sum_{l=1}^n (\|I_{\text{F}} - I_{\text{ov}}\|_1 + \|I_{\text{F}} - I_{\text{un}}\|_1), \quad (14)$$

$$\begin{aligned} L_{\text{ssim}} &= 1 - \frac{1}{2UV} \\ &\quad \cdot \sum_{u=1}^U \sum_{v=1}^V (\text{SSIM}(I_{\text{F}}^{(u,v)}, I_{\text{ov}}^{(u,v)}) + \text{SSIM}(I_{\text{F}}^{(u,v)}, I_{\text{un}}^{(u,v)})), \end{aligned} \quad (15)$$

其中:  $\nabla$  表示梯度算子,  $I_{\text{F}}^{(u,v)}$  表示经过融合后的光场图像,  $I_{\text{ov}}^{(u,v)}$ 、 $I_{\text{un}}^{(u,v)}$  分别表示输入的过曝光和欠曝光光场图像,  $(u, v)$  表示子孔径图像的角度坐标,  $\|\cdot\|_1$  表示

L1 范数,  $n$  表示像素的数量。

相比于普通相机, 光场相机在主镜头和相机传感器之间插入一个微透镜阵列, 可以同时记录光线的强度和角度信息。而光场图像的角度信息可以利用相邻子孔径图像之间的视差变化来体现, 其主要分为水平视差和垂直视差。可是单一的水平或垂直视差难以完整概括整幅光场图像的视差变化。因此, 为了充分挖掘光场图像的视差关系, 同时利用视差关系来约束重建后的光场图像, 本文提出了一种角度一致性损失函数。该损失函数首先取出过曝光和欠曝光的中心子孔径图像  $I_{\text{ov}}^{\text{center}}$ 、 $I_{\text{un}}^{\text{center}}$ , 然后将其与对应光场图像各个位置上的子孔径图像相减, 以获取整幅光场的视差变化图。同理, 按上述方法获取重建图像的视差变化图。考虑到过曝光与欠曝光的互补关系, 将获取的过曝光和欠曝光视差变化图进行取最大值操作。用重建图像的视差变化图与取最大值后的视差变化图计算 L1 损失。具体过程表达如下:

$$\begin{aligned} L_{\text{con}} &= \frac{1}{UV} \sum_{u=1}^U \sum_{v=1}^V \left( \| (I_{\text{F}}^{(u,v)} - I_{\text{F}}^{\text{center}}) \right. \\ &\quad \left. - \max((I_{\text{ov}}^{(u,v)} - I_{\text{ov}}^{\text{center}}), (I_{\text{un}}^{(u,v)} - I_{\text{un}}^{\text{center}})) \|_1 \right), \end{aligned} \quad (16)$$

其中:  $I_{\text{F}}^{\text{center}}$ 、 $I_{\text{ov}}^{\text{center}}$ 、 $I_{\text{un}}^{\text{center}}$  分别代表融合图像、过曝光、欠曝光光场子孔径阵列中的中心子孔径图像。

最后, 总的损失函数可以表示为

$$L_{\text{total}} = L_{\text{fuse}} + L_{\text{con}}. \quad (17)$$

## 4 实验结果与分析

### 4.1 数据集的建立

目前光场成像领域缺乏真实场景的多曝光光场数据集, 然而深度网络的学习过程中高质量数据集的作用至关重要。因此, 为了能训练出一个性能良好的无监督多曝光光场融合网络, 本文构建了用于静态场景的多曝光光场成像任务的数据集, 其中包括两种曝光值的光场图像, 具体构建过程如下:

首先, 将 Lytro Illum 光场相机安装在三脚架上来减少拍摄过程中人为造成的抖动。然后, 利用光场相机的连续曝光模式来捕捉静态场景下的不同曝光的多曝光光场图像。考虑到采用手动快门进行静态场景数据集拍摄易受到手动操作影响而产生细微的伪影, 因此设置光场相机的自动快门来获取严格静态的多曝光光场图像。为了模拟极端曝光情况, 欠曝光和过曝光的曝光间隔设置为 4, 例如图 6 中的  $EV=-2.0$  和



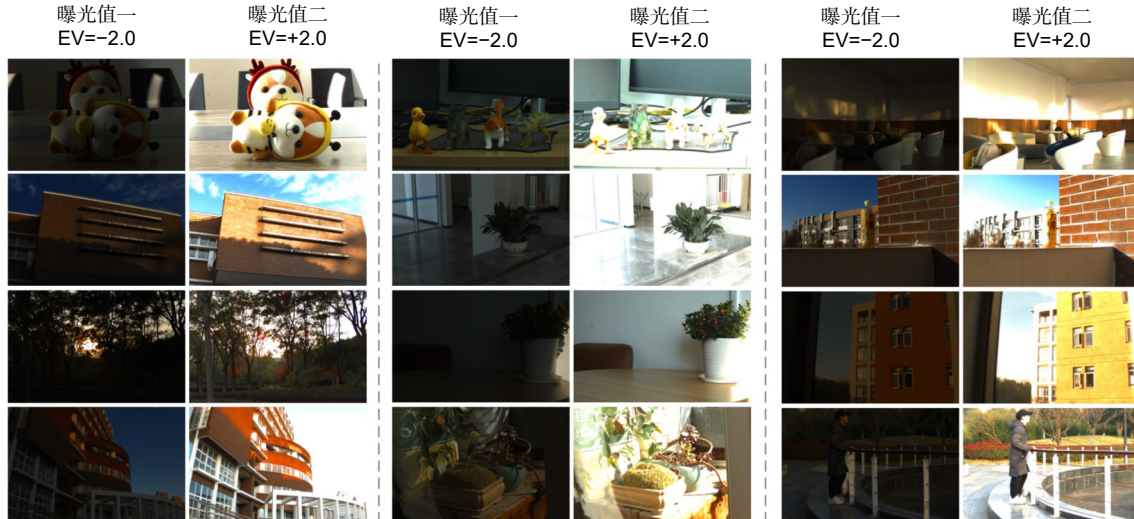


图 6 本文建立的基准数据集中的部分场景示例

Fig. 6 Some scene examples in the benchmark dataset established in this work

EV=+2.0。

光场相机拍摄的原始光场数据需要经过光场工具箱解码以及频域滤波、颜色校正等预处理操作, 以获得视觉上自然的子孔径形式的光场数据。此外, 在解码过程中本文设置光场图像的角度分辨率为 7, 并将原始子孔径图像大小裁剪至 400×600, 以便于后续光场图像的处理。基于上述处理, 本文最终可以获得静态场景下的多曝光光场图像。

在实际构建数据集过程中, 初始共拍摄了 344 组真实场景下的静态多曝光光场图像。然而, 真实场景下的一些静态场景存在难以避免的运动区域, 例如晃动的树叶或背景区域存在移动物体。这些不符合严格静态的场景可能会对网络的训练产生错误的引导作用。因此, 经过严格筛选, 淘汰了超过 40% 初始拍摄的场景, 最终选取了共 200 组高质量的场景用于训练和测试, 其中训练集包含 100 组场景, 测试集与验证集各包含 50 组场景, 且训练和测试的场景互不重叠。图 6 显示了经过挑选后的数据集中场景的一些示例。可以观察到, 所拍摄数据集涵盖了日常生活中常见的室内、室外场景, 且欠曝光和过曝光图像的曝光差异较大, 符合真实场景中极端曝光的实际情况。

#### 4.2 训练细节

本文中的融合处理是对源图像的亮度通道进行的。为了防止融合结果出现颜色失真, 本文采用普通彩色图像融合算法中常用的策略来融合颜色通道 ( $C_b$  和  $C_r$ ), 如下所示:

$$C_f = \frac{C_1(|C_1 - \tau|) + C_2(|C_2 - \tau|)}{|C_1 - \tau| + |C_2 - \tau|}, \quad (18)$$

其中:  $C_1$  和  $C_2$  表示从过曝光和欠曝光的图像中提取的  $C_b$ (或  $C_r$ ) 通道。 $C_f$  代表相应的融合颜色通道, 参数  $\tau$  设置为 128。最后, 将融合的亮度和颜色通道组合并转换到 RGB 颜色空间, 即可获得颜色与曝光良好的结果。

所提出的方法在 PyTorch 框架上实现, 实验环境设置为 12th Gen Intel(R) Core(TM) i7-12700 CPU @2.10GHz、16GBRAM 和 NVIDIA RTX 3090 显卡。网络参数使用 Adamw 优化器进行优化。初始学习率为  $1e-3$ , 并在每轮训练后使用指数衰减方式调整优化器的学习率, 衰减率设置为 0.9。在具体训练过程中, 光场多曝光图像被裁剪为角度分辨率为  $7 \times 7$ 、空间分辨率为  $64 \times 64$  的光场图像块作为网络的输入进行训练。

#### 4.3 与现有方法的对比

在本节中, 将所提出的方法与普通图像多曝光融合的几种先进方法进行比较, 包括 DSIFT-EF<sup>[8]</sup>、GFF<sup>[7]</sup>、MEFAW<sup>[9]</sup>、PAS-MEF<sup>[10]</sup>、DeepFuse<sup>[20]</sup>、PMGI<sup>[27]</sup>、U2Fusion<sup>[26]</sup>、TransMEF<sup>[21]</sup>、FFMEF<sup>[22]</sup>。在选取的九种普通图像多曝光融合方法中, 前四种属于传统方法, 不需要训练。对于后五种基于学习的方法, 在所构建的数据集上对除 TransMEF<sup>[21]</sup> 以外的算法重新训练, 以确保后续比较的公平性和可靠性。TransMEF<sup>[21]</sup> 未在本文数据集上重新训练的原因主要是由于其在经过大量自然图像的预训练后并未进行融合网络的训练, 因此采用与其论文中相同的融合方法直接进行融合。

具体来说, 本文使用光场图像的子孔径数据形式作为普通图像的样本来训练所有比较方法。

一种客观指标的高低并不能证明采用某种方法得到的多曝光融合结果所具有的视觉效果的好坏。因此, 为了全面评估不同方法的融合性能, 采用了八种客观指标进行对比分析, 其中分别包括: 基于图像信息量的指标 NMI<sup>[31]</sup>、 $Q_{\text{nice}}$ <sup>[32]</sup>; 基于图像特征的指标 AG<sup>[33]</sup>、 $Q_{\text{abf}}$ <sup>[34]</sup>、SD<sup>[35]</sup>、SF<sup>[36]</sup>; 基于图像结构相似度的指标 MEFSSIM<sup>[29]</sup> 和基于人眼视觉感知的指标  $Q_{\text{cv}}$ <sup>[37]</sup>。共挑选了 50 组不与训练集重叠的多曝光光场图像对进行客观指标的对比, 并在表 1 中给出了客观指标的对比实验结果, 其中最高的性能指标用粗体标记, 次优的性能指标用下划线标记。

如表 1 所示, 在从不同评价角度选取的 8 个评估指标中, 本文方法在其中 3 个指标上取得了最好的结果, 在另外 2 个指标上取得了次优的结果; 综合上述不同方法在不同指标上的排名, 本文方法最终取得了最为靠前的排名, 证明本文方法具有良好的多曝光融合性能。尤其是在 AG<sup>[33]</sup>、SD<sup>[35]</sup> 指标上有着显著的提高。AG<sup>[33]</sup> 上最优的结果意味着本文方法可以利用可逆神经网络构建的光场图像解码器准确地重建出光场融合图像中更多的细节信息。SD<sup>[35]</sup> 通常用于衡量图像亮度分布的均匀性和对比度。较高的 SD<sup>[35]</sup> 体现了使用本文方法融合后的图像具有更大的亮度差异和更丰富的细节, 可以以无监督的方式获取到动态范围大的光场图像。此外,  $Q_{\text{cv}}$ <sup>[37]</sup> 的最佳结果也体现了本文方法在保留更多细节的同时取得了良好的视觉感知效果。需要指出的是传统方法在 MEFSSIM<sup>[29]</sup>、 $Q_{\text{abf}}$ <sup>[34]</sup>、SF<sup>[36]</sup> 等评价指标上超越了本文方法, 但传统方法缺少非线性拟合能力而容易出现明显的亮暗分界区域,

影响了多曝光融合图像的观感, 这种缺陷也使得传统方法在与眼感知有关的评价指标  $Q_{\text{cv}}$ <sup>[37]</sup> 上获得较差的排名。

图 7 展示了本文方法与现有方法在本文测试集上的主观多曝光融合结果, 从中可以看到传统的多曝光融合方法在处理曝光差异大的区域时失去了全局结构的一致性, 并出现了明显的亮暗过渡区域, 如 DSIFT-EF<sup>[8]</sup> 和 GFF<sup>[7]</sup> 在图 7(a) 中白色墙面的融合结果。基于学习的方法可以获得比传统方法更加均匀的光照信息, 但难以调整极度曝光区域中的光照分布和恢复出足够的细节。FFMEF<sup>[22]</sup> 和 U2Fusion<sup>[26]</sup> 在图 7(a) 的融合结果中出现细节模糊的问题。DeepFuse<sup>[20]</sup>、PMGI<sup>[27]</sup> 和 TransMEF<sup>[21]</sup> 方法都存在共同的问题, 即无法恢复暗部的光场图像细节, 如图 7(b) 所示。整体而言, 本文方法能够恢复极端曝光条件下的场景细节信息, 并能增强局部暗区域, 以达到更佳视觉效果。如图 7(a) 的红框放大图中的结果表明本文方法可以增强暗部的结构并恢复相应结构的纹理信息。虽然传统方法也能够保留该区域的细节, 但全局不均匀的光照在很大程度上降低了其性能。

#### 4.4 光场角度一致性

与普通成像不同, 光场成像可以利用捕获的多视图信息提升深度估计的性能, 即光场深度图估计的精确度与光场图像的角度质量密切相关。借助光场这一特性, 本文采用估计光场深度图的方法来体现不同融合方法在保持光场角度一致性方面的性能。本文共选取五种在客观指标上表现良好的多曝光融合方法进行主观对比。图 8 展示了不同方法深度图估计的主观结果, 从中可以看出由于普通图像的多曝光融合方法单

表 1 50 个测试场景上不同方法的客观指标对比

Table 1 Comparison of objective indicators among different methods on 50 testing scenarios

Method	SD $\uparrow$	MEFSSIM $\uparrow$	$Q_{\text{cv}}\downarrow$	SF $\uparrow$	$Q_{\text{abf}}\uparrow$	$Q_{\text{nice}}\uparrow$	NMI $\uparrow$	AG $\uparrow$	Rank $\downarrow$
DISFT_EF <sup>[8]</sup>	58.2897(5)	0.9610(3)	679.7548(8)	23.5287(2)	0.7489(3)	0.8118(10)	0.6411(8)	5.4570(7)	46
GFF <sup>[7]</sup>	60.2416(3)	<u>0.9624(2)</u>	592.7575(10)	<b>25.8524(1)</b>	<b>0.7648(1)</b>	0.8154(7)	0.6074(9)	5.7301(3)	36
MEFAW <sup>[9]</sup>	54.2736(6)	<b>0.9639(1)</b>	625.0993(9)	22.2241(3)	0.7507(2)	0.8163(6)	0.5215(10)	5.5581(4)	41
PAS_MEF <sup>[10]</sup>	48.7852(10)	0.8887(9)	517.8809(7)	16.2884(7)	0.5452(6)	0.8146(8)	0.6610(7)	5.5482(5)	59
DeepFuse <sup>[20]</sup>	53.3335(7)	0.9040(6)	321.1736(4)	16.3365(6)	0.5408(7)	0.8202(3)	0.8482(3)	5.4888(6)	42
PMGI <sup>[27]</sup>	51.0411(9)	0.8933(8)	361.4652(6)	16.2283(8)	0.5209(8)	0.8194(4)	0.8285(4)	5.4126(8)	55
U2Fusion <sup>[26]</sup>	52.7949(8)	0.8973(7)	316.6257(3)	15.5335(10)	0.5061(9)	0.8183(5)	0.7999(5)	5.3375(10)	57
TransMEF <sup>[21]</sup>	<u>60.6602(2)</u>	0.9328(5)	<u>281.9047(2)</u>	18.1766(5)	0.6264(5)	<b>0.8210(1)</b>	<b>0.8758(1)</b>	6.0909(2)	<u>23</u>
FFMEF <sup>[22]</sup>	58.9095(4)	0.8595(10)	334.6085(5)	16.0121(9)	0.4360(10)	0.8135(9)	0.7276(6)	5.4097(9)	62
Proposed	<b>67.5961(1)</b>	0.9491(4)	<b>238.4925(1)</b>	20.5533(4)	0.6870(4)	<u>0.8205(2)</u>	<u>0.8510(2)</u>	<b>6.7186(1)</b>	<b>19</b>

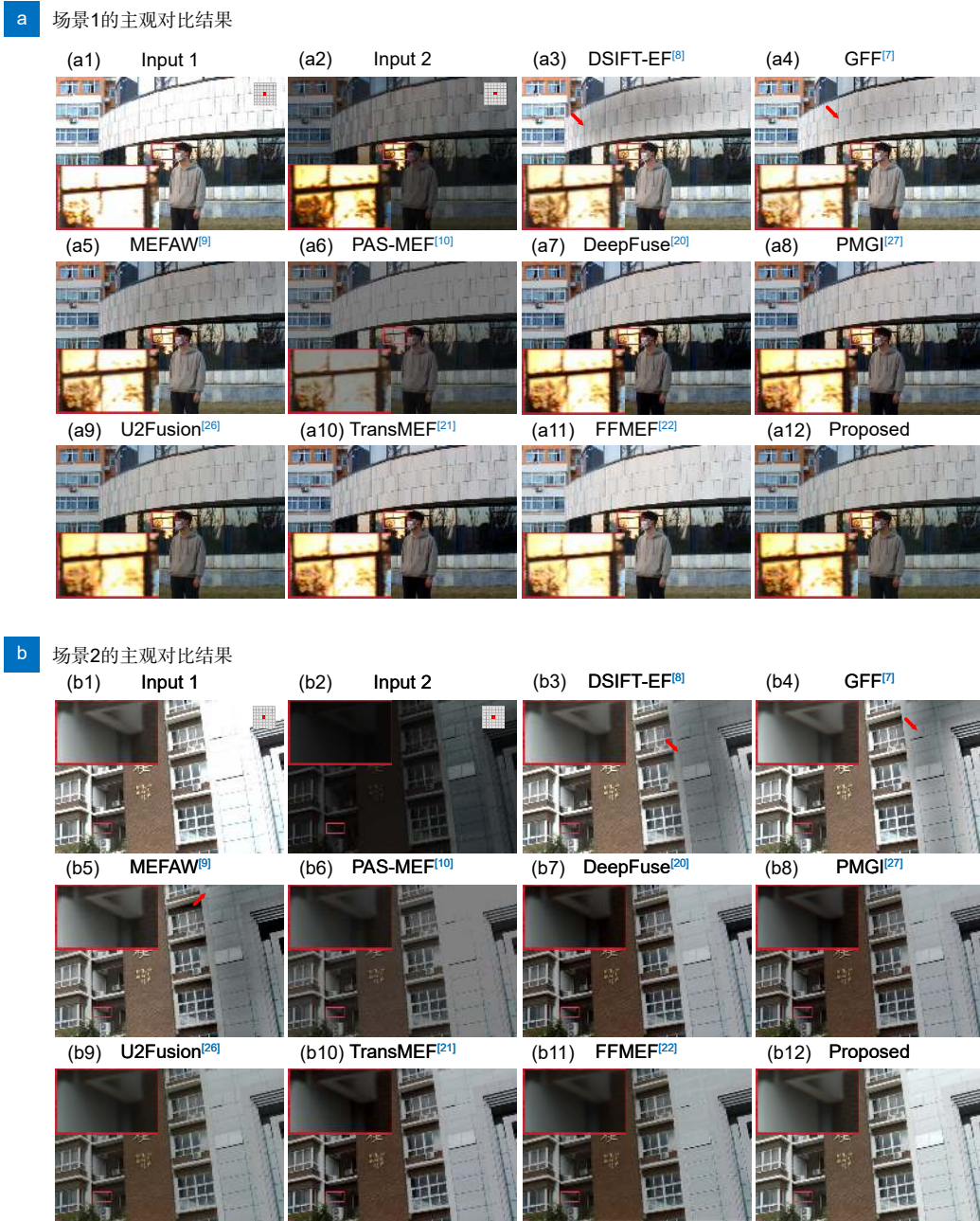


图 7 不同方法在所建立的基准数据集上的主观对比结果

Fig. 7 Subjective comparison results of different methods on the established benchmark dataset

独对每个子孔径图像进行融合的策略忽略了光场图像的角度维度的联合处理, 在一些具有挑战性的区域中产生了深度估计误差, 例如图 8 中人物裤子和衣领区域。相比之下, 利用本文方法融合的结果所估计的深度图表现出更少的误差区域和更清晰的边缘, 这意味着所提出的方法具有更好的性能。

#### 4.5 消融实验

在本节中, 以具体的融合场景为示例对本文提出的方法进行消融实验, 包括多尺度空角交互特征提取

模块、联合可逆神经网络与 Transformer 设计的光场重建模块、多核局部特征提取模块以及为保证光场角度一致性而设计的损失函数。消融实验均在本文所构建的光场数据集上进行。

1) 多尺度空角交互特征提取模块 (C-ASA) 的有效性。为了验证本文提出的空角上下文交互模块可以有效地提取到光场高维特征信息, 消融实验中使用普通卷积层替换了这一部分。表 2 给出了定量对比结果, 从中可以看出本文所提出的模块由于正确处理光场的

高维数据结构, 因此融合图像的质量得到很大的提升。从图 9 中定性结果也可以看出, 普通的卷积层无法有效地提取到不同曝光程度下的光场特征, 导致最终的融合图像整体曝光偏向欠曝光, 恢复出来的动态范围极为有限, 而本文提出的方法获得了动态范围更大的光场图像。

2) 联合可逆神经网络与 Transformer 搭建的光场重建模块 (T-INN) 的有效性。为了在光场图像重建过程中保留更多的细节信息, 在重建阶段引入可以进行无损信息传输的可逆神经网络。为了验证本文设计的光场重建模块的有效性, 利用普通的  $3 \times 3$  卷积和相应的激活函数来替代本文的光场重建网络。表 2 给出了这一消融实验的定量评价结果。当用普通卷积替代光

场重建模块后, 融合图像的质量在八种指标上出现了不同程度的下降。从图 9 所示的主观对比结果中也可以发现, 使用卷积替换后的融合结果中出现了拼接伪影, 而且融合结果中的天空出现了不自然的颜色与亮暗不均匀的斑点区域。

3) 多核局部特征提取模块的有效性 (MK)。本文通过设置三条并行但卷积核大小不同的局部特征提取分支, 弥补了后续上下文交互的光场空角特征提取模块局部特征缺少的问题, 且扩大了局部特征的感受野大小。为了验证这一部分设计的有效性, 使用  $3 \times 3$  卷积来代替多核的局部特征提取模块进行消融实验。如图 9 所示, 相同感受野的局部特征提取模块在处理极端曝光区域时难以恢复出令人满意的细节信息。相比

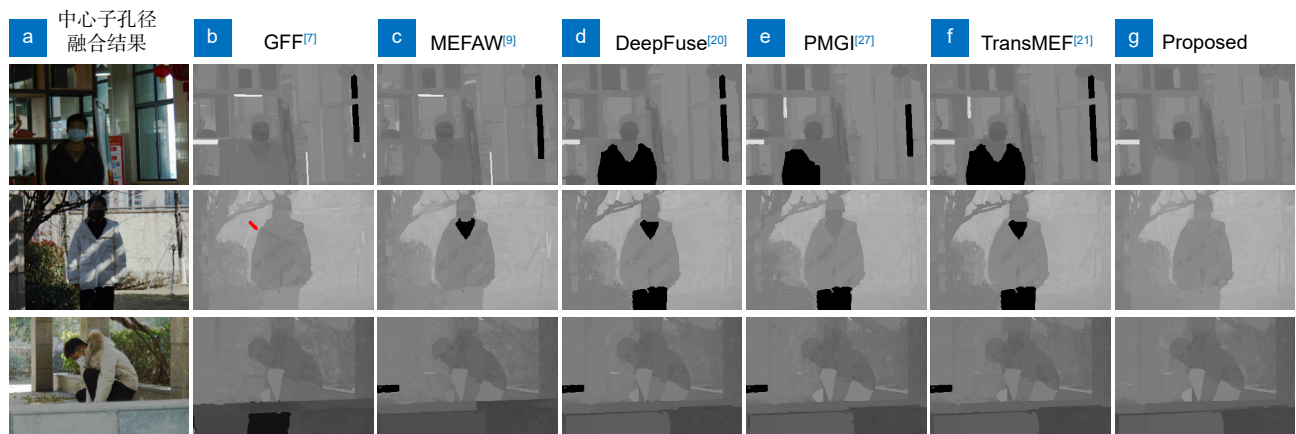


图 8 从不同方法的多曝光光场融合结果中估计出深度图的主观对比结果

Fig. 8 Subjective comparison of depth maps estimated from the fused light field images obtained with different fusion methods

表 2 本文主要网络结构消融实验的客观指标对比结果

Table 2 Comparative results of objective indicators of the main network structure ablation experiments in this article

Method	SD $\uparrow$	MEFSSIM $\uparrow$	Q <sub>cv</sub> $\downarrow$	SF $\uparrow$	Q <sub>abr</sub> $\uparrow$	Q <sub>nice</sub> $\uparrow$	NMI $\uparrow$	AG $\uparrow$
w/o C-ASA	<b>74.6707</b>	0.8747	313.7698	16.8144	0.4682	0.8143	0.6906	5.7951
w/o T-INN	40.1623	0.7105	948.2726	16.4418	0.3252	0.8095	0.4769	6.1205
w/o MK	66.819	0.8999	252.7716	19.3284	0.5027	0.8142	0.6666	6.3629
Proposed	67.5961	<b>0.9491</b>	<b>238.4925</b>	<b>20.5533</b>	<b>0.687</b>	<b>0.8205</b>	<b>0.851</b>	<b>6.7186</b>

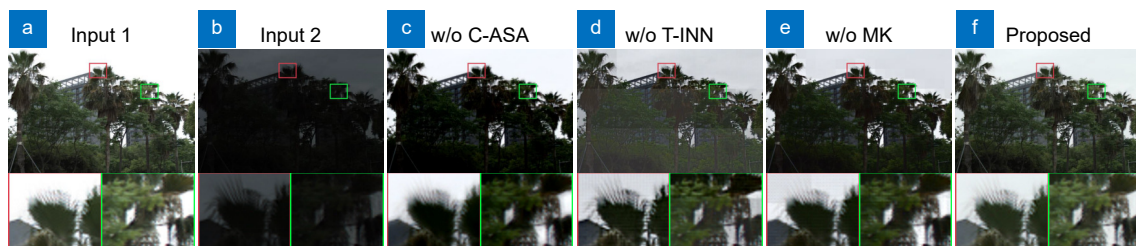


图 9 本文主要网络结构消融实验的主观对比结果

Fig. 9 Subjective comparison results of the main network structure ablation experiments

之下, 本文所设计的多核结构实现了细节丰富的融合结果。

4) 角度一致性损失函数 (AC loss) 的有效性。在本文工作中, 针对融合任务设计了一种新的损失函数, 以监督融合网络在输出高质量的融合结果的同时不破坏光场的角度一致性。为了验证此损失函数的有效性, 本文在训练过程中移除了角度一致性损失后重新训练网络。从图 10 估计的深度图中可以看出, 由于缺少角度一致性损失的约束, 使得融合过程中产生了角度维度的信息丢失, 在最终的深度图中出现了结构模糊、误差区域增大等问题, 阻碍了光场后续的应用。

#### 4.6 运行时间及模型复杂度分析

在实际应用中, 运行时间综合反映了模型的处理速度。因此, 本文共选取六种在客观指标上表现良好的多曝光融合方法, 并在相同环境下对比较方法的运行时间和参数量进行了测试。其中, GFF<sup>[7]</sup>、MEFAW<sup>[9]</sup>和 DISFT\_EF<sup>[8]</sup>属于传统多曝光融合方法, 在 CPU 上执行, 不包含可学习的参数, 而其他方法依赖于深度学习, 在 GPU 上执行。为了保证对比实验的公平性,

在同一实验环境中测试不同方法的运行时间, 环境设置为 12th Gen Intel(R) Core(TM) i7-12700 CPU @2.10 GHz、16 GB RAM 和 NVIDIA RTX 3090 显卡。表 3 给出了测试结果, 其中运行时间表示利用输入的两幅极端曝光的光场图像重建出角度空间分辨率为  $7 \times 7 \times 400 \times 600$  的光场图像所需要的时间。由于普通图像的多曝光融合方法无法联合处理光场图像, 故将光场数据转化为子孔径形式并以单张子孔径图像输入来计算最终的运行时间。从表 3 的对比结果可以看出, 普通图像的多曝光融合方法不需要针对光场图像的高维特性进行网络设计, 因此相较于本文方法具有更少的参数量。但也正因为缺少对光场图像的联合处理, 造成了不同程度的客观质量下降并无法保证光场数据的角度一致性, 相关对比实验可以在表 1 和图 8 中找到。

## 5 结论

在本文的工作中, 针对多曝光光场融合任务提出了一种名为 LF-UMTI 的融合方法。面对现有方法对

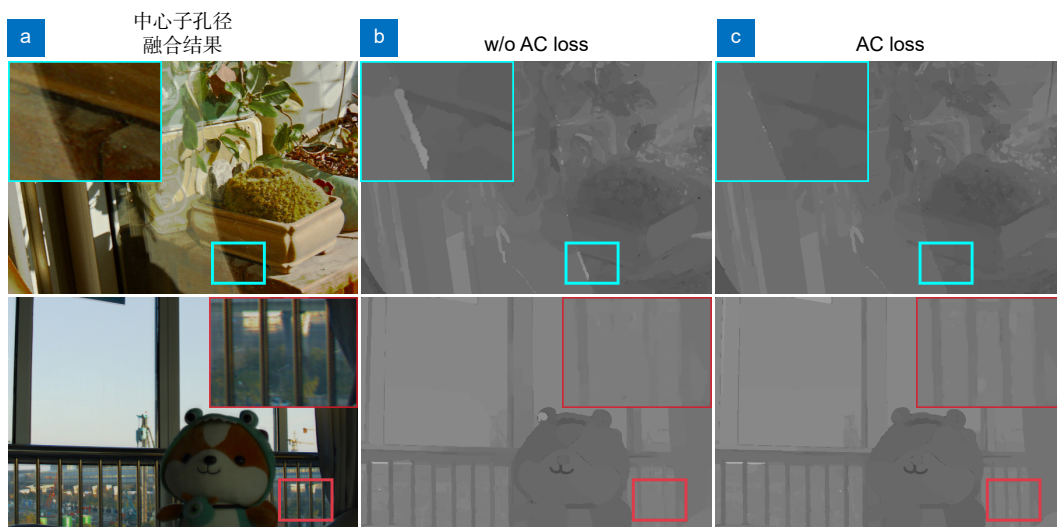


图 10 角度一致性损失消融实验的主观对比结果

Fig. 10 Subjective comparison results of angle consistency loss ablation experiments

表 3 不同方法的运行时间及参数量比较结果

Table 3 Comparison results of running time and parameter quantities of different methods

Method	GFF <sup>[7]</sup>	MEFAW <sup>[9]</sup>	DISFT_EF <sup>[8]</sup>	DeepFuse <sup>[20]</sup>	PMGI <sup>[27]</sup>	TransMEF <sup>[21]</sup>	Proposed
Runtime/s	13.6808	30.5074	23.0349	14.8176	6.9727	11.2749	16.8701
Params/M	-	-	-	0.283	0.0401	18.1696	0.9734
Platform	MATLAB	MATLAB	MATLAB	Pytorch	Pytorch	Pytorch	Pytorch
Device	CPU	CPU	CPU	GPU	GPU	GPU	GPU

光场角度维度信息考虑不足的现象, 本文设计了一种空角交互网络, 以无监督的方式解决融合任务。多尺度空角交互特征模块利用通道维度上建模的自注意力机制与多感受野的卷积层, 从源光场图像中学习不同尺度上的丰富特征信息。光场信息交互与重建模块中交叉注意力的引入使得光场不同曝光度的特征能够以互补的方式相互融合。而以可逆神经网络原理构建的重建模块则最大程度上保留了光场的细节信息。此外, 角度一致性损失函数的提出有效地减少了融合过程中角度信息的损失。为了训练所提出的无监督网络, 本文还构建了一个真实场景的多曝光光场数据集。经过大量主客观对比实验和消融实验验证, 本文所提方法在纹理细节、视觉亮度与光场角度一致性方面均表现良好。

## 参考文献

- [1] Cui Z L, Sheng H, Yang D, et al. Light field depth estimation for non-lambertian objects via adaptive cross operator[J]. *IEEE Trans Circuits Syst Video Technol*, 2024, **34**(2): 1199–1211.
- [2] Ma S, Wang N, Zhu L C, et al. Light field depth estimation using weighted side window angular coherence[J]. *Opto-Electron Eng*, 2021, **48**(12): 210405.  
马帅, 王宁, 朱里程, 等. 基于边框加权角相关的光场深度估计算法[J]. *光电工程*, 2021, **48**(12): 210405.
- [3] Wu D, Zhang X D, Fan Z G, et al. Depth acquisition of noisy scene based on inline occlusion handling of light field[J]. *Opto-Electron Eng*, 2021, **48**(7): 200422.  
吴迪, 张旭东, 范之国, 等. 基于光场内联遮挡处理的噪声场景深度获取[J]. *光电工程*, 2021, **48**(7): 200422.
- [4] Cong R X, Yang D, Chen R S, et al. Combining implicit-explicit view correlation for light field semantic segmentation[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023: 9172–9181. <https://doi.org/10.1109/CVPR52729.2023.00885>.
- [5] Han L, Zhong D W, Li L, et al. Learning residual color for novel view synthesis[J]. *IEEE Trans Image Process*, 2022, **31**: 2257–2267.
- [6] Xu F, Liu J H, Song Y M, et al. Multi-exposure image fusion techniques: a comprehensive review[J]. *Remote Sens*, 2022, **14**(3): 771.
- [7] Li S T, Kang X D, Hu J W. Image fusion with guided filtering[J]. *IEEE Trans Image Process*, 2013, **22**(7): 2864–2875.
- [8] Liu Y, Wang Z F. Dense SIFT for ghost-free multi-exposure fusion[J]. *J Visual Commun Image Represent*, 2015, **31**: 208–224.
- [9] Lee S, 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]//*Proceedings of the 25th IEEE International Conference on Image Processing (ICIP)*, 2018: 1737–1741. <https://doi.org/10.1109/ICIP.2018.8451153>.
- [10] Ulucan O, Ulucan D, Turkan M. Ghosting-free multi-exposure image fusion for static and dynamic scenes[J]. *Signal Process*, 2023, **202**: 108774.
- [11] Gul M S K, Wolf T, Bätz M, et al. A high-resolution high dynamic range light-field dataset with an application to view synthesis and tone-mapping[C]//*2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, 2020: 1–6.  
<https://doi.org/10.1109/ICMEW46912.2020.9105964>.
- [12] Li C, Zhang X. High dynamic range and all-focus image from light field[C]//*Proceedings of the 7th IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM)*, 2015: 7–12.  
<https://doi.org/10.1109/ICCIS.2015.7274539>.
- [13] Le Pendu M, Guillemot C, Smolic A. High dynamic range light fields via weighted low rank approximation[C]//*Proceedings of the 25th IEEE International Conference on Image Processing (ICIP)*, 2018: 1728–1732.  
<https://doi.org/10.1109/ICIP.2018.8451584>.
- [14] Yin J L, Chen B H, Peng Y T. Two exposure fusion using prior-aware generative adversarial network[J]. *IEEE Trans Multimedia*, 2021, **24**: 2841–2851.
- [15] Xu H, Ma J Y, Zhang X P. MEF-GAN: multi-exposure image fusion via generative adversarial networks[J]. *IEEE Trans Image Process*, 2020, **29**: 7203–7216.
- [16] Liu J Y, Wu G Y, Luan J S, et al. HoLoCo: holistic and local contrastive learning network for multi-exposure image fusion[J]. *Inf Fusion*, 2023, **95**: 237–249.
- [17] Liu J Y, Shang J J, Liu R S, et al. Attention-guided global-local adversarial learning for detail-preserving multi-exposure image fusion[J]. *IEEE Trans Circuits Syst Video Technol*, 2022, **32**(8): 5026–5040.
- [18] Chen Y Y, Jiang G Y, Yu M, et al. Learning to simultaneously enhance field of view and dynamic range for light field imaging[J]. *Inf Fusion*, 2023, **91**: 215–229.
- [19] Ram Prabhakar K, Sai Srikar V, Venkatesh Babu R. DeepFuse: a deep unsupervised approach for exposure fusion with extreme exposure image pairs[C]//*Proceedings of 2017 IEEE International Conference on Computer Vision*, 2017: 4724–4732.  
<https://doi.org/10.1109/ICCV.2017.505>.
- [20] Ma K D, Duanmu Z F, Zhu H W, et al. Deep guided learning for fast multi-exposure image fusion[J]. *IEEE Trans Image Process*, 2020, **29**: 2808–2819.
- [21] Qu L H, Liu S L, Wang M N, et al. TransMEF: a transformer-based multi-exposure image fusion framework using self-supervised multi-task learning[C]//*Proceedings of the 36th AAAI Conference on Artificial Intelligence*, 2022: 2126–2134.  
<https://doi.org/10.1609/AAAI.v36i2.20109>.
- [22] Zheng K W, Huang J, Yu H, et al. Efficient multi-exposure image fusion via filter-dominated fusion and gradient-driven unsupervised learning[C]//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023: 2804–2813.  
<https://doi.org/10.1109/CVPRW59228.2023.00281>.
- [23] Xu H, Haochen L, Ma J Y. Unsupervised multi-exposure image fusion breaking exposure limits via contrastive learning[C]//*Proceedings of 37th AAAI Conference on Artificial Intelligence*, 2023: 3010–3017.  
<https://doi.org/10.1609/AAAI.v37i3.25404>.
- [24] Zhang H, Ma J Y. IID-MEF: a multi-exposure fusion network based on intrinsic image decomposition[J]. *Inf Fusion*, 2023, **95**: 326–340.
- [25] Xu H, Ma J Y, Le Z L, et al. FusionDN: a unified densely

- connected network for image fusion[C]//*Proceedings of the 34th AAAI Conference on Artificial Intelligence*, 2020: 12484–12491. <https://doi.org/10.1609/AAAI.v34i07.6936>.
- [26] Xu H, Ma J Y, Jiang J J, et al. U2Fusion: a unified unsupervised image fusion network[J]. *IEEE Trans Pattern Anal Mach Intell*, 2022, 44(1): 502–518.
- [27] Zhang H, Xu H, Xiao Y, et al. Rethinking the image fusion: a fast unified image fusion network based on proportional maintenance of gradient and intensity[C]//*Proceedings of the 34th AAAI Conference on Artificial Intelligence*, 2020: 12797–12804. <https://doi.org/10.1609/AAAI.v34i07.6975>.
- [28] Zhou M, Huang J, Fang Y C, et al. Pan-sharpening with customized transformer and invertible neural network[C]//*Proceedings of the 36th AAAI Conference on Artificial Intelligence*, 2022: 3553–3561. <https://doi.org/10.1609/aaai.v36i3.20267>.
- [29] Ma K D, Zeng K, Wang Z. Perceptual quality assessment for multi-exposure image fusion[J]. *IEEE Trans Image Process*, 2015, 24(11): 3345–3356.
- [30] Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity[J]. *IEEE Trans Image Process*, 2004, 13(4): 600–612.
- [31] Hossny M, Nahavandi S, Creighton D. Comments on 'Information measure for performance of image fusion'[J]. *Electron Lett*, 2008, 44(18): 1066–1067.
- [32] Wang Q, Shen Y, Jin J. Performance evaluation of image fusion techniques[M]//Stathaki T. *Image Fusion: Algorithms and Applications*. Amsterdam: Academic Press, 2008: 469–492. <https://doi.org/10.1016/B978-0-12-372529-5.00017-2>.
- [33] Cui G M, Feng H J, Xu Z H, et al. Detail preserved fusion of visible and infrared images using regional saliency extraction and multi-scale image decomposition[J]. *Opt Commun*, 2015, 341: 199–209.
- [34] Xydeas C S, Petrovic V. Objective image fusion performance measure[J]. *Electronics Letters*, 2000, 36(4): 308–309.
- [35] Rao Y J. In-fibre Bragg grating sensors[J]. *Meas Sci Technol*, 1997, 8(4): 355–375.
- [36] Eskicioglu A M, Fisher P S. Image quality measures and their performance[J]. *IEEE Trans Commun*, 1995, 43(12): 2959–2965. <https://doi.org/10.1109/26.477498>.
- [37] Chen H, Varshney P K. A human perception inspired quality metric for image fusion based on regional information[J]. *Inf Fusion*, 2007, 8(2): 193–207.

## 作者简介



李玉龙 (2000-), 男, 安徽合肥人, 硕士研究生, 主要从事光场多曝光图像融合等方面的研究。

E-mail: [1946734397@qq.com](mailto:1946734397@qq.com)



【通信作者】郁梅 (1968-), 女, 江苏无锡人, 博士, 教授, 博士生导师, 2000 年于韩国亚洲大学 (Ajou University) 获得博士学位, 主要从事多媒体信号处理与通信、计算成像、视觉感知与编码、图像与视频质量评价等方面的研究。

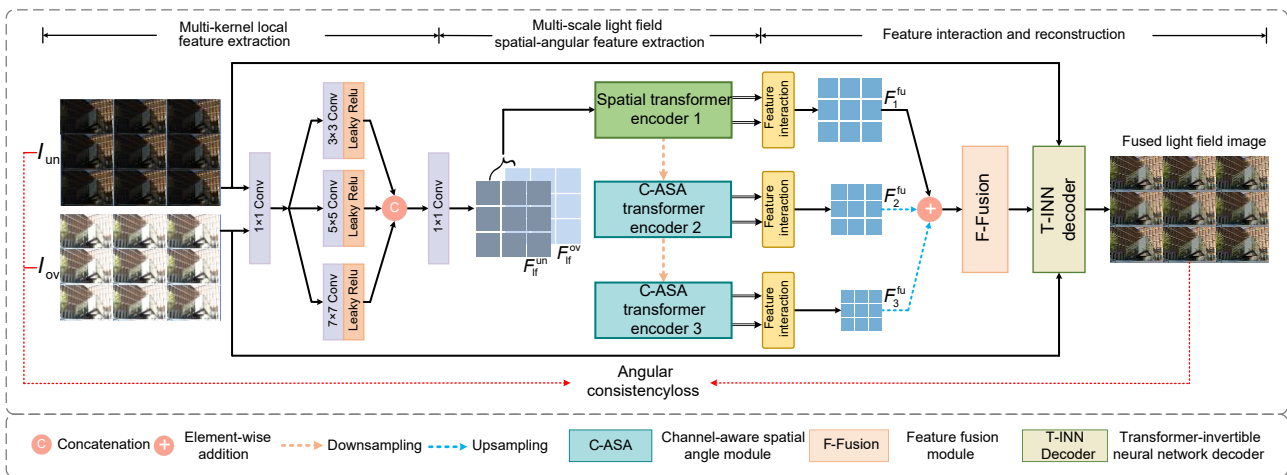
E-mail: [yumei@nbu.edu.cn](mailto:yumei@nbu.edu.cn)



扫描二维码, 获取PDF全文

# LF-UMTI: unsupervised multi-exposure light field image fusion based on multi-scale spatial-angular interaction

Li Yulong<sup>1</sup>, Chen Yeyao<sup>1</sup>, Cui Yueli<sup>2</sup>, Yu Mei<sup>1\*</sup>



**Overview:** Light field imaging has unique advantages in many applications such as refocusing and depth estimation, since it can simultaneously capture spatial and angular information of light rays. However, due to the limited dynamic range of the camera, the light field images may suffer from over-exposure and under-exposure issues, bringing challenges to capturing all the details of the real scene and posing difficulties for subsequent light field applications. In recent years, deep learning has shown powerful nonlinear fitting capabilities and has achieved good results in multi-exposure fusion for conventional images. However, the high-dimensional characteristics of light field images make it necessary to consider not only the issues of traditional images suffered from, but also the angular consistency of the fused light field images during multi-exposure fusion. In this paper, an unsupervised multi-exposure light field imaging method (LF-UMTI) based on multi-scale spatial-angular interaction is proposed. Firstly, a multi-scale spatial-angular interaction strategy is employed to extract spatial-angular features and explore complementary information of source light field images at different scales. A channel-dimensional modeling strategy is also employed to reduce computational complexity and adapt to the high-dimensional structure of light fields. Secondly, a light field reconstruction module guided by reversible neural networks is constructed to avoid fusion artifacts and recover more detailed information. Lastly, an angular consistency loss is designed, which takes into account the disparity variations between boundary sub-aperture images and central sub-aperture images to ensure the disparity structure of the fusion result. To evaluate the performance of the proposed method, a benchmark dataset of multi-exposure light field images of the real scenes is established. Through subjective and objective quality evaluations of the fused light field images as well as ablation experiments conducted on the proposed dataset, the effectiveness of the proposed method is demonstrated in reconstructing high-contrast and detail-rich light field images while preserving angular consistency. Considering future research tasks and analyzing the limitations of the network, simplifying the model and improving the operational speed will be key directions for future research tasks.

Li Y L, Chen Y Y, Cui Y L, et al. LF-UMTI: unsupervised multi-exposure light field image fusion based on multi-scale spatial-angular interaction[J]. *Opto-Electron Eng*, 2024, 51(6): 240093; DOI: 10.12086/oe.2024.240093

<sup>1</sup>Faculty of Information Science and Engineering, Ningbo University, Ningbo, Zhejiang 315211, China; <sup>2</sup>School of Electronic and Information Engineering, Taizhou University, Taizhou, Zhejiang 318000, China

\* E-mail: yumei@nbu.edu.cn