

基于特征交互融合与注意力的光场图像超分辨率

徐欣宜^{1,2*}, 邓慧萍^{1,2}, 向森^{1,2}, 吴谨^{1,2}¹武汉科技大学信息科学与工程学院, 湖北 武汉 430081;²武汉科技大学冶金自动化与检测技术教育部工程研究中心, 湖北 武汉 430081

摘要 光场图像包含丰富的空间信息和角度信息,在三维重建、虚拟现实领域有广泛应用。但由于光场相机的内部限制,光场图像的低空间分辨率阻碍了其应用发展,具体表现为图像边缘区域的模糊。考虑到光场子孔径图像中空间信息包含着丰富的纹理和细节,而角度信息则对应不同视图之间的相关性,提出一种基于特征交互融合与注意力的光场图像超分辨率网络。通过特征提取和特征交互融合模块充分融合光场的空间角度信息;通过特征通道注意力模块自适应地学习有效信息,抑制冗余信息并细化图像的高频细节;通过光场结构一致性模块保持光场图像间的视差结构。在 5 个光场数据集上的实验结果表明,所提网络得到的超分辨率结果性能普遍优于所比较的超分辨率网络。

关键词 图像处理; 超分辨率; 深度学习; 光场图像; 注意力机制

中图分类号 TP391

文献标志码 A

DOI: 10.3788/LOP221911

Light Field Image Super-Resolution Based on Feature Interaction Fusion and Attention Mechanism

Xu Xinyi^{1,2*}, Deng Huiping^{1,2}, Xiang Sen^{1,2}, Wu Jin^{1,2}¹*School of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan 430081, Hubei, China;*²*Engineering Research Center for Metallurgical Automation and Measurement Technology of Ministry of Education, Wuhan University of Science and Technology, Wuhan 430081, Hubei, China*

Abstract Light field images contain rich spatial and angle information and are, therefore, widely used in three-dimensional reconstruction and virtual reality; however, the limited spatial resolution of light field pictures, notably the blurring of the image edge area, prevents their application and development due to the inherent constraints of light field cameras. A light field image super-resolution network based on feature interactive fusion and attention is proposed here because the spatial information in a light field subaperture image contains rich texture and high-frequency details and the angle information corresponds to the correlation between different views. Here, the feature extraction and feature interactive fusion modules completely fuse the spatial and angle information of the light field; the feature channel attention module refines high-frequency aspects of the images by adaptively learning effective information and suppressing redundant information; and the optical field structure consistency module preserves the parallax structure between optical field pictures. The performance of the proposed network is typically superior to that of the compared super-resolution network, according to the experimental results from five light field datasets.

Key words image processing; super-resolution; deep learning; light field image; attention mechanism

1 引言

图像超分辨率(SR)是指由一幅低分辨率图像或图像序列恢复出高分辨率图像。高分辨率图像中含有丰富的信息,这些丰富的信息在图像分割^[1]、图像深度估计^[2]和图像显著性检测^[3]等方面有着广泛应用。

光场(LF)相机通过在主透镜和传感器之间加入微透镜阵列,不仅记录光线的强度,还记录光线的方向^[4]。但由于光场相机内部传感器技术的限制,目前光场相机获取的子孔径图像分辨率普遍较低,低空间分辨率的问题阻碍了光场成像的发展。在实际应用中需要用到高分辨率的图像,因此有必要提升低分辨率光场图

收稿日期: 2022-06-24; 修回日期: 2022-08-05; 录用日期: 2022-09-26; 网络首发日期: 2022-10-06

通信作者: *731403114@qq.com

像的分辨率,生成高分辨率光场图像。

在光场图像超分辨率领域,传统的方法和基于深度学习的方法都得到了广泛的应用。传统方法一般通过设计滤波器和估计深度来提高分辨率。Wanner 等^[5]提出一种基于深度信息估计的光场图像超分辨率变分方法。Mitra 等^[6]通过高斯混合模型编码光场结构,以实现深度估计、视图合成和光场图像超分辨率提升。Alain 等^[7]将 BM3D 滤波扩展到光场中,提出针对光场图像对和光场图像去噪的 LFBM5DSR。Rossi 等^[8]开发了一种基于图的方法,通过图优化来实现光场图像的超分辨率提升。虽然这些模型很好地编码了光场结构,但由于这些图像对于先验信息的表示能力较差,空间信息无法得到充分利用。

近年来,基于深度学习的图像超分辨率方法在空间信息开发中的表现优于传统的方法。近几年深度学习在单视图超分辨率领域取得了巨大成功,受到这些工作的启发,最近的光场图像超分辨率方法采用深度卷积网络来提高其性能。基于深度学习的光场图像超分辨率方法利用不同视图之间的互补信息来学习低分辨率到高分辨率视图的映射关系。Yoon 等^[9]首先使用卷积神经网络(CNN)来处理光场图像,提出的 LFCNN 先通过 SRCNN 提升相邻子孔径图像对的空间分辨率,再通过一个结构相似的 SRCNN 来合成相邻子孔径图像对之间的新视图。Yuan 等^[10]利用极平面图(EPI)的特点提出 LF-DCNN,该网络同样分为两个部分,先通过 EDSR 网络对每个子孔径图像进行空间超分辨,然后提取出 EPI 输入 EPI 增强网络,该 EPI 增强网络由几个密集残差块组成,最后将重建的 EPI 转换成高分辨率的子孔径图像。LFCNN 和 LF-DCNN 的第 1 部分都只是单纯地对空间信息进行了提升,没有用到角度信息。Wang 等^[11]将双向循环卷积网络(BRCNN)引入光场图像超分辨率,提出 LFNet,他们建立了两个相同结构的双向循环卷积子网络,分别对一列和一子孔径图像的相邻视图迭代地学习空间相关性,最后通过堆栈泛化技术将两个子网络的输出融合生成高分辨率光场。Zhang 等^[12]利用 EPINet 模型提出 resLF,他们将 4 个角度方向的子孔径图像堆叠起来形成输入,构成 4 个不同的分支,对 4 个分支的视图都用多个残差块进行特征提取,最后将它们融合起来提高中心子孔径图像超分辨率。然而,这些网络只利用了特定方向视图的信息,没有用到所有视图的有效空间信息,信息利用不充分。为了解决这些问题,Yeung 等^[13]提出 LFSSR,利用空间和角度的交替卷积来提高所有视图的超分辨率。Jin 等^[14]提出 LF-ATO,将网络分为多个分支,每个分支利用其他所有视图的信息来辅助超分辨其中一个视图,以实现所有视图的高超分辨率。然而这些方法虽然充分利用了空间信息,却忽略了光场的角度域信息,同样造成了信息的不充分利用、边缘区域细节的丢失,使重建视图较模糊。

Zhang 等^[15]提出 MEG-Net,将不同方向上的子孔径图像堆叠起来分别作为输入,提取图像间的亚像素信息。Wang 等^[16]提出 LF-InterNet,分别提取视图的角度信息和空间信息,并用角度信息辅助空间信息。这些方法虽然加入了角度信息,但是没有考虑到信息的冗余问题,在一些细节上的结果仍然不是很理想。

与普通的 2D 图像不同,光场图像除了具有空间信息,还具有丰富的角度信息,现有的深度学习方法没有利用丰富的角度域信息,造成图像的纹理特别是边缘区域模糊。考虑到这些问题,本文提出一种基于特征交互融合与注意力的光场图像超分辨率网络。该网络充分利用角度信息来辅助空间信息进行空间超分辨,同时通过注意力机制学习有效信息,提高超分辨的性能,解决边缘模糊问题,保留清晰的纹理细节。总的来说,所提网络具有以下特点:

1)设计空间角度特征提取模块(SAFE_module),利用交替的残差块和空洞空间金字塔池化(ASPP)来提取深层空间特征、增大感受野、提取不同尺度上丰富的上下文信息,有助于恢复纹理细节;将角度信息融入空间信息实现角度信息对空间超分辨的辅助,利用互补的角度信息恢复缺失的细节,使图像的边缘区域更清晰。

2)由于加入丰富的角度信息,光场子孔径图像之间的特有相关性会造成大量的信息冗余,因此设计特征通道注意力模块(FCA_module)来细化空间细节,通过堆叠的特征通道注意力使主干网络只学习有效的空间和角度信息、自适应地给不同通道赋予不同的权重以抑制冗余信息,获得更清晰的重建图像。

3)融合角度与空间信息进行上采样后的高分辨率图像难以保持视差结构,而大多数方法通过几何映射保持视差,会丢失高频信息。引入结构一致性模块(SC_module),利用空间域和角度域的交替卷积来对高分辨率图像进行进一步优化,隐式地学习空间角度相关性,以保持重建视图之间的视差结构,实现光场结构一致性的正则化,进一步提高超分辨性能。

2 光场图像超分辨率算法

2.1 网络结构

4D 光场图像可以表示为 $L(x, y, u, v)$, 其中, (x, y) 表示光场的空间坐标, (u, v) 表示角度坐标。低分辨率光场图像可以描述为 $L \in \mathbf{R}^{U \times V \times H \times W}$, 其中, U 和 V 表示角度域, H 和 W 表示空间域, 若光场被排列成子孔径图像阵列, 则可以看作含有 $U \times V$ 个子孔径图像, 每个子孔径图像的空间分辨率为 $H \times W$ 。为了得到高空间分辨率的光场图像 $L \in \mathbf{R}^{U \times V \times aH \times aW}$ (a 为图像放大倍数), 需要学习到一个函数 f , 其可将低空间分辨率图像通过映射得到高空间分辨率图像:

$$L(\hat{x}, \hat{y}, u, v) = f[L(x, y, u, v)], \quad (1)$$

式中： $L(\hat{x}, \hat{y}, u, v)$ 表示高空间分辨率光场； $L(x, y, u, v)$ 表示低空间分辨率光场； f 表示需要学习的空间超分辨函数。

相比于 2D 图像，光场图像具有更丰富的四维信息，光场图像超分辨率的核心问题是充分利用角度信息和空间信息并恢复纹理细节来提升图像质量。所提基于特征交互融合与注意力的光场图像超分辨率网络既可以充分利用空间角度信息又可以保持良好的光场结构一致性、恢复高频细节，提升重建质量。

所提网络的总体结构如图 1 所示，网络分为两条

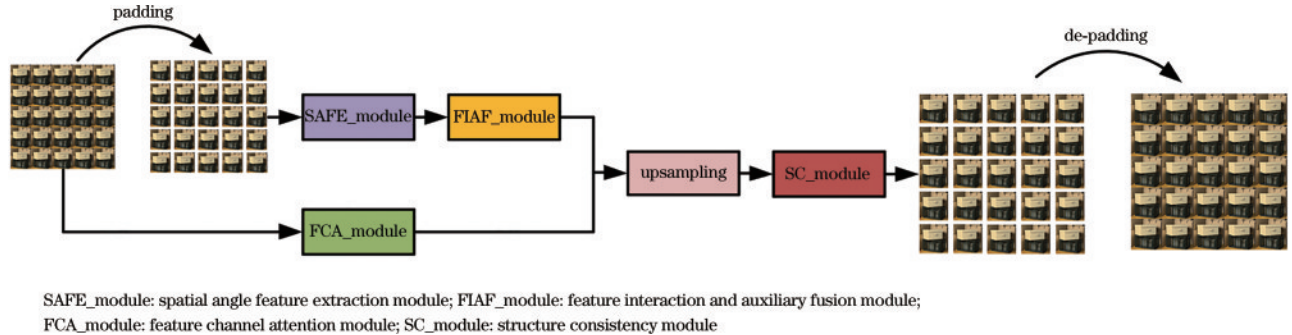


图 1 所提网络总体结构

Fig. 1 Overall architecture of proposed network

2.2 特征提取模块

4D 光场具有丰富的四维特征，这些空间和角度特征是紧密联系的，并且对于超分辨率网络的贡献是不同的，因此很难有效地利用空间信息或是角度信息。考虑到这点，所提 SAFE_module 通过两种特征提取卷积即空间特征提取卷积 (SFECConv) 和角度特征提取卷积 (AFECConv) 来分别提取空间信息和角度信息，从而消除不同特征间的干扰，提高特征的有效性。另外，很多超分辨率网络都依赖于不同的角度分辨率，每当改变角度分辨率就要训练不同的模型导致计算量增大。因此，所提网络用密集排列的光场子孔径图像作为输入，使空间特征提取卷积和角度特征提取卷积独立于光场角度分辨率，从而可以使用单一网络来训练不同角度分辨率的光场图像，降低训练复杂度。具有丰富的空间上下文信息的特征有利于后面的特征交互和特征辅助融合，因此需要大感受野来提取深层特征。为此，使用 ASPP 提取深层空间特征，增大感受野并获得不同尺度的上下文信息。

对于光场子孔径图像而言，每个子孔径图像包含着空间信息，不同的子孔径图像之间则对应着角度信息。以 $H \times W \times A \times A$ 的子孔径图像阵列作为输入，首先在每个相邻的子孔径图像间进行一个像素的 0 填充，形成 $(AH + A - 1) \times (AH + A - 1)$ 大小的子孔径图像，然后在填充后的图像阵列进行空间特征提取卷积和角度特征提取卷积形成两路特征，其中：SFECConv 为核大小 3×3 、步长为 1 的卷积，用来在每

支路：以子孔径图像为输入，在第 1 条支路中先通过空间角度特征提取模块提取图像的空间特征和角度特征并对两种特征进行处理；随后经过特征交互和辅助融合模块让空间特征和角度特征进行充分交互，并用角度特征辅助空间特征得到融合后的空间特征；另外，对输入图像加入特征通道注意力抑制冗余特征，形成第 2 条支路，突出有用的边缘特征，恢复图像高频细节；将 2 条支路融合并经过上采样后，通过交替的空间角度卷积来保持光场的结构一致性，最后进行视图间的 de-padding 得到密集的高分辨率光场子孔径图像。

个子孔径图像中提取空间信息，和角度信息无关；AFECConv 为核大小 3×3 、步长为 1、dilation 为 $[H + 1, W + 1]$ 、padding 为 $[H + 1, W + 1]$ 的卷积，用来提取不同的子孔径图像间相同空间位置的角度信息，与空间信息无关。接着用 4 个交替的残差块和残差 ASPP 对空间特征支路提取深层空间特征，ASPP 采用 3 个 3×3 的空洞卷积，空洞率分别为 1、2、4，提取出不同尺度的层次特征，将 3 个尺度的特征连接起来后再用 1×1 的卷积融合。图 2 为特征提取模块结构图。

2.3 特征交互和辅助融合模块

在提取光场的空间信息和角度信息后，为了充分利用空间信息和角度信息，需要对空间信息和角度信息进行交互和融合，因此加入特征交互和融合模块 (FIAF_module)。一方面，角度特征和输入的空间特征连接，经过一个核大小为 3×3 、步长为 1 的卷积和一个 ReLU 层让两个特征融合，实现由角度信息引导的空间特征融合；另一方面，输入的空间特征通过核大小为 3×3 、步长为 1、dilation 为 $[H + 1, W + 1]$ 、padding 为 $[H + 1, W + 1]$ 的角度特征提取卷积形成新的角度特征，与输入的角度特征连接，然后通过一个 1×1 的卷积和 ReLU 层进行特征融合，完成角度特征的更新。同时，空间特征与角度特征都采用局部残差连接输入特征和交互后的特征。每一次空间角度特征的交互形成一个信息交互块，生成新的空间信息和角度信息，将多个信息交互块堆叠起来以生成多个空间信息和角度信息，图 3 为 FIAF_module 结构图。

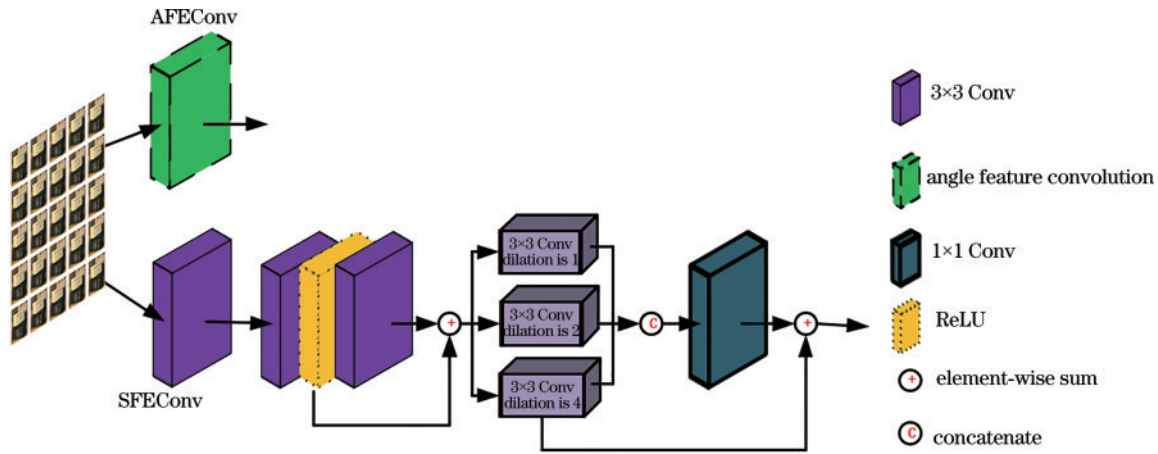


图 2 特征提取模块

Fig. 2 Feature extraction module

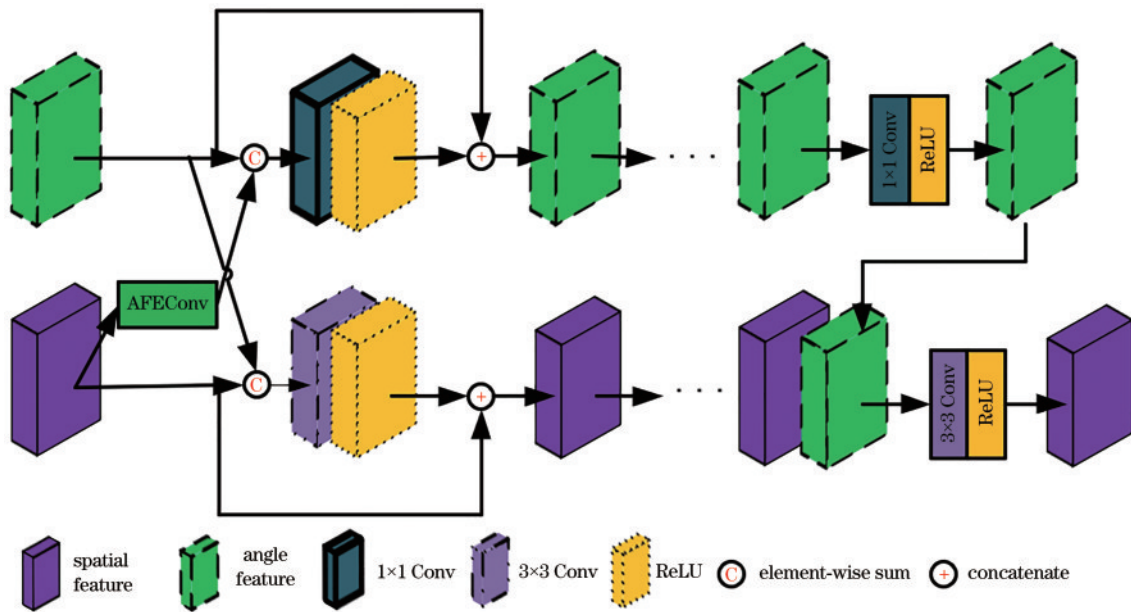


图 3 FIAF_module 结构图

Fig. 3 Structure diagram of FIAF_module

信息交互后需要进行信息的融合。将每个信息交互块生成的空间信息和角度信息分别连接起来,然后进行特征融合。具体地,先用一个 1×1 的卷积将连接后的角度特征压缩到 C 个通道,然后再经过上采样和交互模块输出的空间特征相连接,连接后的空间特征再次通过空间特征提取卷积压缩通道,将所有特征进行融合得到最后的空间特征。

2.4 特征通道注意力模块

图像空间超分辨率任务的本质就是恢复高频信息,高频即图像中的边缘、细节部分。现有的大部分网络直接从低分辨率图像中提取特征,并对每个通道都做一样的处理,在低频信息上耗费了太多算力,阻碍了网络的性能提升。其次,由于光场子孔径图像之间具有很强的相关性,在获取充分的光场图像特征信息的同时会造成大量的信息冗余,很多冗余信息都用不上。

注意力机制是解决信息冗余问题的好方法,近年来,越来越多的图像处理方法中引入了注意力机制,如文献 [17-19]。本实验设计了一个由注意力模型组成的 FCA_module,用来将图像的细节部分细化,将重点放在有效的高频信息上并抑制无效的冗余信息。该模块通过学习每个通道特征的重要性,为每个通道赋予不同的权重,给更有价值的通道更高的权重。

与以往的通道注意力机制(CA)^[17]不同,所提网络通过一个 3×3 的卷积提取浅层特征,然后将其输入多个级联的特征通道注意力块中,每个特征通道注意力块的结构如图 4 所示。先用两个卷积层提取深层信息,再用特征通道注意力赋予权重。具体地,用全局最大池化来细化边缘高频信息,用全局平均池化来聚合全局空间信息,随后用两个 1×1 卷积分别对两种池化后的结果进行降维和升维以捕获跨通道间的依赖性,并将两路支

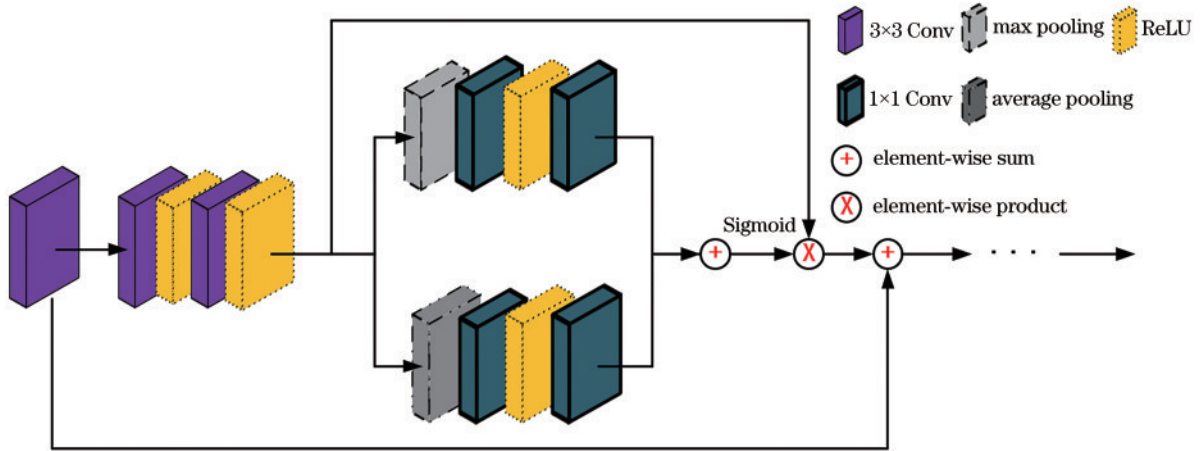


图 4 特征通道注意力模块

Fig. 4 Feature channel attention module

路融合提高获得有效高频信息的能力,最后通过 Sigmoid 激活函数激活每个通道,并与输入通道特征相乘来自适应地调整不同通道的权重。另外,将输入浅层特征添加到注意力后形成残差从而进行局部残差学习。将信息融合后得到的特征图和细化模块得到的特征图连接并融合,最后将空间特征通过卷积进行通道扩张,将通道数变成 a^2C ,然后通过 PixelShuffle 层上采样将通道数降为 C ,生成高分辨率的光场子孔径图像。

2.5 结构一致性模块

光场图像具有丰富的空间角度四维信息,这就表示其在相邻的子孔径图像间都具有特定的视差结构。视差结构作为光场图像的重要特性,在超分辨后应得到很好的保存。现有的大多数模型大致采用两种方法来保持光场结构一致性:1)在传统方法中,通过使对应的像素共享相似的强度值来提高结构一致性,这种方法通常需要通过几何映射得到深度图,这导致结果依赖于深度图的质量,不能很好地应对遮挡区域,还会产生高频损失;2)在深度学习中,不同子孔径图像间的相

关性仅在低分辨率时被应用,高分辨率图像间的相关性没有很好的保持。因此,受到 Yeung 等^[13]的启发,所提网络使用交替的空间角度卷积组成 SC_module 来保持高分辨率光场视图的结构一致性。同时还能代替 4D 卷积处理光场数据的高维性,大大降低了参数量,与前面的特征交互融合模块一样可以充分利用光场图像间的互补信息。

具体地,先用核大小为 3×3 、步长为 1 的卷积层对前面生成的每个高分辨率子孔径图像提取特征,得到尺寸为 $aH \times aW \times C \times MN$ 的特征图。再用 2D 空间卷积对特征图进行空间域的卷积,随后将空间域转换为角度域,即 $M \times N \times C \times a^2HW$,然后用 2D 角度卷积对特征图进行角度域的卷积,这样就完成了一个空间角度的卷积。将多个空间角度卷积叠加形成空间角度域的交替卷积。最后,将生成的空间角度深度特征再经过一次空间卷积得到重建的残差图,添加到输入的子孔径图像形成残差学习。图 5 为 SC_module 结构图。

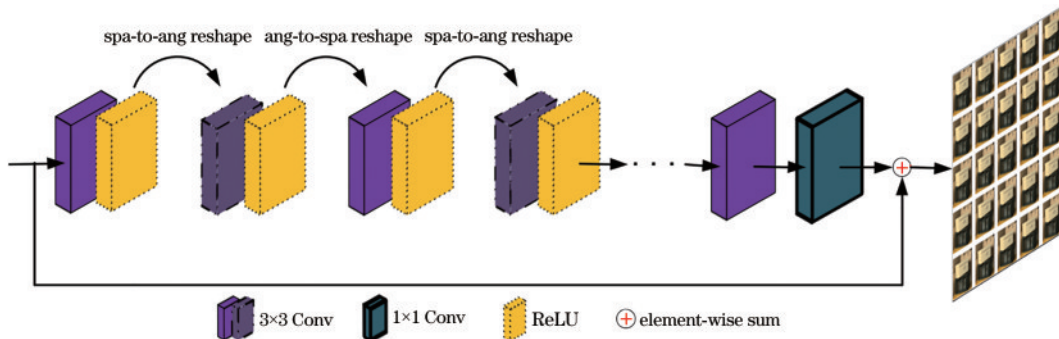


图 5 SC_module 结构图

Fig. 5 Structure diagram of SC_module

3 实验结果及分析

3.1 训练细节

使用 5 个公共 4D 光场数据集来验证算法的有效

性,这 5 个数据集分别为 EPFL^[20]、HCInew^[21]、HCIOld^[22]、INRIA^[23]、STFgantry^[24],训练集和测试集中所有光场的角度分辨率都为 9×9 ,表 1 为各个数据集中训练集和测试集的场景数。在训练前,将每个子

孔径图像裁剪成大小为 64×64 的 patch, 然后用双线性下采样(下采样因子为 2 和 4)生成低分辨率图像作为网络的输入。为了解决数据不足的问题, 使用随机翻转和 90° 旋转将训练数据增加 8 倍, 在每次数据增强过程中, 所有的子孔径图像都需要沿着空间和角度方向进行翻转和旋转, 以保持完整的光场结构。损失函数为 L1 损失函数, 其对图像超分辨率有良好的效果, 并对异常点有更好的鲁棒性:

$$L_{L1} = \| \mathbf{y}_i - \mathbf{y}_j \|_1, \quad (2)$$

式中: \mathbf{y}_i 表示第 i 个像素的真实值; \mathbf{y}_j 表示第 j 个像素的估计值; $\| \cdot \|_1$ 表示 1-范数, 为真实值和预测值之间距离的均值。

表 1 5 个数据集上训练集和测试集的场景数

Table 1 Number of scenarios for training and test sets on 5 datasets

Dataset	EPFL	HCInew	HCInold	INRIA	STFgantry	Total
Training	70	20	10	35	9	144
Test	10	4	2	5	2	23

在训练过程中: 特征交互融合模块中特征交互块数设置为 16, 每 4 个形成 1 个特征交互组, 每个特征交互组生成 1 个空间特征和 1 个角度特征, 最后生成 4 个

空间特征和 4 个角度特征作为特征融合模块的输入; 特征通道注意力模块中的注意力块数设置为 20; 结构一致性模块中的交替卷积数设置为 4; 默认训练的角度分辨率为 5×5 。

实验在 NVIDIA GTX3080Ti GPU 上运行, 使用 PyTorch 1.6 作为框架, 优化器使用 Adam, 采用随机梯度下降, batch size 设为 8, 初始学习率设为 0.0005, 每 10 个 epoch 学习率下降 0.5。在第 16 个 epoch 后收敛, 总共训练 40 个 epoch, 训练大约花了 5 天。

3.2 算法性能比较及分析

为了验证所提算法的有效性, 在 4D 光场数据集上与一些经典算法进行定性和定量比较, 包括: 最新传统光场超分辨率算法 GB^[8]、LFBM5D^[7]; 基于学习的 2D 图像超分辨率算法 VDSR^[25]、EDSR^[26]、RCAN^[27]; 基于学习的光场超分辨率算法 LFNet^[11]、resLF^[12]、LFSSR^[13]、LF-ATO^[14]、LF-InterNet^[16]、MEG-Net^[15]。采用双线性插值作为基准。定量评价的指标采用 PSNR 和 SSIM, PSNR 表示峰值信噪比, SSIM 表示结构相似度指数。

表 2、表 3 为各算法在 $\times 2$ 、 $\times 4$ 倍超分辨率下在 5 个数据集上的 PSNR/SSIM 定量结果对比, 其中, 每列中黑色加粗为最优值, 加下划线为次优值。

表 2 $\times 2$ 超分辨率 PSNR/SSIM 指标对比

Table 2 Comparison of $\times 2$ super-resolution PSNR/SSIM indicators

Algorithm	Scale	EPFL	HCInew	HCInold	INRIA	STFgantry
Bicubic	$\times 2$	29.740/0.9376	31.887/0.9356	37.686/0.9785	31.331/0.9577	31.063/0.9498
VDSR	$\times 2$	32.498/0.9598	34.371/0.9561	40.606/0.9867	34.439/0.9741	35.541/0.9789
EDSR	$\times 2$	33.089/0.9629	34.828/0.9592	41.014/0.9874	34.985/0.9764	36.296/0.9818
RCAN	$\times 2$	33.159/0.9634	35.022/0.9603	41.125/0.9875	35.046/0.9769	36.670/0.9831
GB	$\times 2$	31.485/0.9564	33.736/0.9523	39.885/0.9853	33.297/0.9713	33.983/0.9719
LFBM5D	$\times 2$	31.243/0.9563	33.735/0.9527	40.029/0.9856	33.118/0.9713	33.932/0.9716
LFNet	$\times 2$	33.106/0.9684	36.051/0.9707	42.621/0.9920	34.889/0.9789	37.105/0.9873
resLF	$\times 2$	33.617/0.9706	36.685/0.9739	43.422/0.9932	35.395/0.9804	38.354/0.9904
LFSSR	$\times 2$	33.671/0.9744	36.802/0.9749	43.811/0.9938	35.279/0.9832	37.944/0.9898
LF-ATO	$\times 2$	34.166/0.9742	37.080/0.9742	43.640/0.9936	35.985/0.9833	39.582/0.9920
LF-InterNet	$\times 2$	34.314/0.9752	36.997/0.9757	44.363/0.9942	35.986/0.9839	38.371/0.9908
MEG-Net	$\times 2$	<u>34.332/0.9761</u>	<u>37.186/0.9766</u>	43.965/0.9940	<u>36.049/0.9839</u>	38.678/0.9912
Proposed algorithm	$\times 2$	34.493/0.9767	37.468/0.9790	<u>44.252/0.9943</u>	36.366/0.9843	<u>38.894/0.9916</u>

从表 2、表 3 中可以看出, 所提算法在 5 个数据集上都有很好的结果。其中: $\times 2$ 超分辨率结果在 STFgantry 数据集中 PSNR/SSIM 指标为次优, 在 HCInold 数据集中 PSNR 指标为次优, 在其他数据集中 PSNR/SSIM 指标均为最优; $\times 4$ 超分辨率结果在 STFgantry 数据集中 PSNR/SSIM 指标同样为次优, 在 EPFL 数据集中 PSNR 指标为次优, 在其他数据集中 PSNR/SSIM 指标均为最优。

图 6、图 7 分别为不同超分辨率算法的视觉效果对比。图 6 为超分辨率 $\times 2$ 时测试数据集中选取的两个

场景; 图 7 为超分辨率 $\times 4$ 时测试数据集中选取的两个场景。

从图 6、图 7 可以直观看出, 基于学习的图像超分辨率算法普遍优于传统光场图像超分辨率算法, 而基于学习的光场图像超分辨率算法优于基于学习的 2D 单个图像超分辨率算法。图 6 中, 所提算法明显更接近真实值, 特别是在具有密集条纹的位置, 如 HCInew-origami 中的横条纹和 INRIA-sculpture 的阶梯部分, 而其他算法特别是传统算法和 2D 单个图像超分辨率算法的结果非常模糊, 边缘细节无法被很好恢复。在

表 3 $\times 4$ 超分辨率 PSNR/SSIM 指标对比Table 3 Comparison of $\times 4$ super-resolution PSNR/SSIM indicators

Algorithm	Scale	EPFL	HCInew	HCInold	INRIA	STFgantry
Bicubic	$\times 4$	25.264/0.8324	27.715/0.8517	32.576/0.9344	26.952/0.8867	26.087/0.8452
VDSR	$\times 4$	27.246/0.8777	29.308/0.8823	34.810/0.9515	29.186/0.9204	28.506/0.9009
EDSR	$\times 4$	27.833/0.8854	29.591/0.8869	35.176/0.9536	29.656/0.9257	28.703/0.9072
RCAN	$\times 4$	27.907/0.8863	29.694/0.8886	35.359/0.9548	29.805/0.9276	29.021/0.9131
GB	$\times 4$	26.900/0.8737	29.128/0.8817	34.473/0.9523	28.774/0.9183	28.162/0.8953
LFBM5D	$\times 4$	26.810/0.8774	29.351/0.8885	34.797/0.9573	28.805/0.9246	28.184/0.9035
LFNet	$\times 4$	27.899/0.8991	30.382/0.9067	36.159/0.9657	29.967/0.9381	29.718/0.9322
resLF	$\times 4$	28.260/0.9035	30.723/0.9107	36.705/0.9682	30.338/0.9412	30.191/0.9372
LFSSR	$\times 4$	28.596/0.9118	30.928/0.9145	36.907/0.9696	30.585/0.9467	30.570/0.9426
LF-ATO	$\times 4$	28.514/0.9115	30.880/0.9135	36.999/0.9699	<u>30.711/0.9484</u>	30.607/0.9430
LF-InterNet	$\times 4$	28.708/0.9139	<u>30.977/0.9163</u>	<u>37.078/0.9714</u>	<u>30.670/0.9485</u>	30.491/0.9423
MEG-Net	$\times 4$	28.656/0.9137	<u>31.002/0.9162</u>	<u>37.109/0.9709</u>	30.590/0.9478	30.619/0.9435
Proposed algorithm	$\times 4$	<u>28.701/0.9148</u>	31.131/0.9208	37.210/0.9724	30.734/0.9503	<u>30.610/0.9432</u>

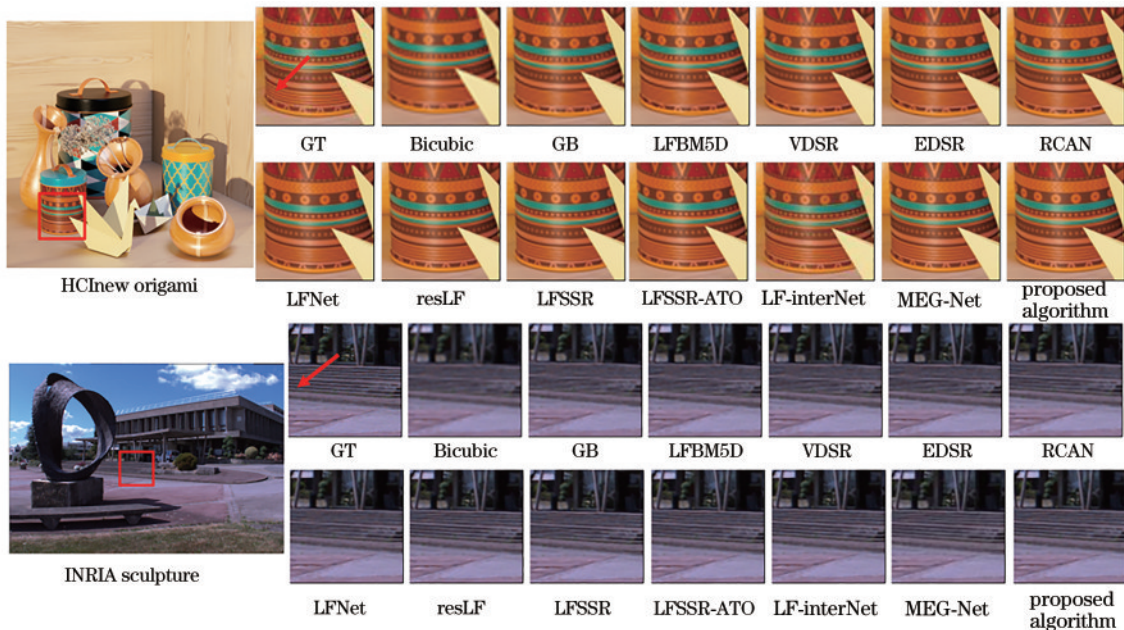
图 6 $\times 2$ 超分辨率不同算法视觉效果对比Fig. 6 Comparison of visual effects of different algorithms with $\times 2$ super-resolution

图 7 中,恢复 $\times 4$ 超分辨率是非常困难的,基于学习的单图像超分辨率算法的结果非常模糊,这是因为 $\times 4$ 超分辨率的输入缺失大量细节,而单个图像中的空间信息无法重建高质量的高分辨率图像。相比之下,所提算法可以使用不同视图之间的角度互补信息来恢复缺失的细节,从而获得更好的超分辨率结果,得到清晰的纹理和边缘细节。

表 4 为 $\times 2$ 超分辨率时所提算法与各种基于深度学习的算法的参数量 (Params) 及在 5 个数据集上运行时间的对比。

从表 4 可以看出,所提算法在参数量较小的情况下,相比于其他大多数基于深度学习的超分辨率算法具有明显的速度优势,且具有较高的超分辨率重建

质量。

为了表明所提算法的有效性,将其直接应用于几个真实世界场景的光场图像中,由于这种情况下无法使用地面真实值 (GT),与几个光场超分辨率深度学习算法的视觉性能进行了比较,图 8 为其中两个场景的视觉效果。

从图 8 可以看出,所提算法在细节上恢复得更清晰,如第 1 个图中的窗户边缘细节和第 2 个图中的“DELL”字样,证明其可应用于低频相机生成高质量的高分辨率光场图像。

3.3 消融实验

在 EFPL 数据集上对 $\times 2$ 超分辨率进行了消融实验,分别验证空间信息和角度信息与 FIAF_module、

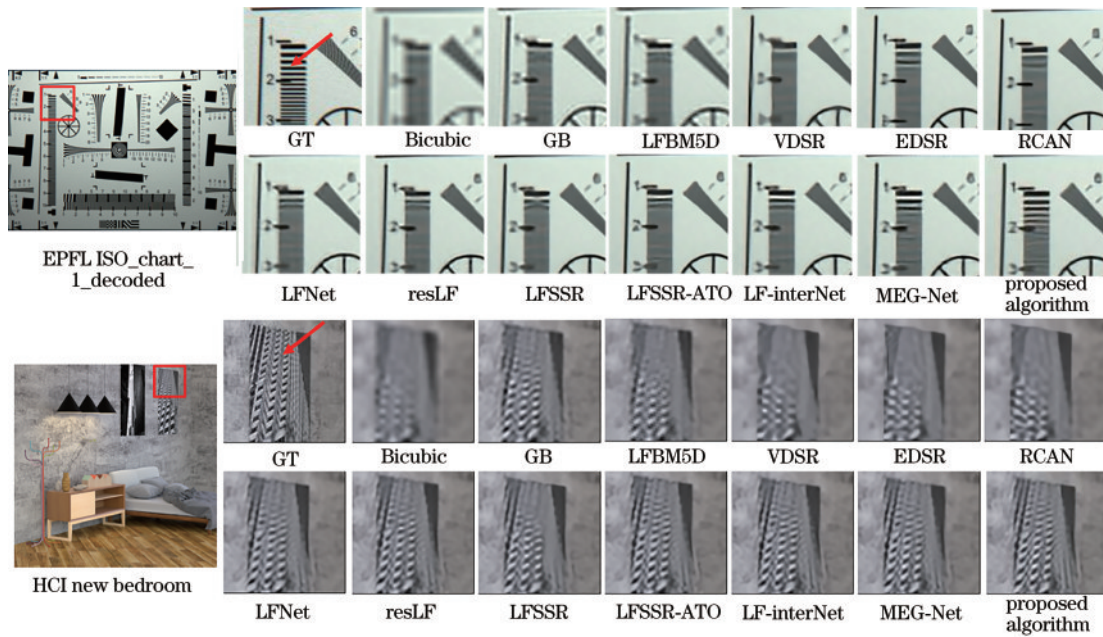


图7 ×4超分辨率不同算法视觉效果对比

Fig. 7 Comparison of visual effects of different algorithms with ×4 super-resolution

表4 ×2超分辨率模型参数量与运行时间(s)对比

Table 4 Comparison of ×2 super-resolution model parameters and running time (s)

Algorithm	Params / 10 ⁶	EPFL	HCInew	HCIdold	INRIA	STFgantry
VDSR	0.665	104.55	40.00	45.47	49.78	20.63
EDSR	38.890	727.34	267.79	299.01	360.05	133.24
RCAN	15.360	178.69	60.14	69.80	82.69	32.31
LFNet	10.385	203.57	84.35	95.92	107.48	48.02
resLF	8.648	749.12	272.48	310.60	368.29	136.84
LFSSR	1.774	79.51	30.30	33.73	37.46	15.00
LF-ATO	1.364	561.25	205.72	233.35	280.79	104.66
LF-InterNet	5.483	193.83	72.06	81.27	96.98	36.79
MEG-Net	1.775	285.45	104.00	119.35	141.54	53.08
Proposed algorithm	4.273	162.43	43.56	48.51	57.70	18.74

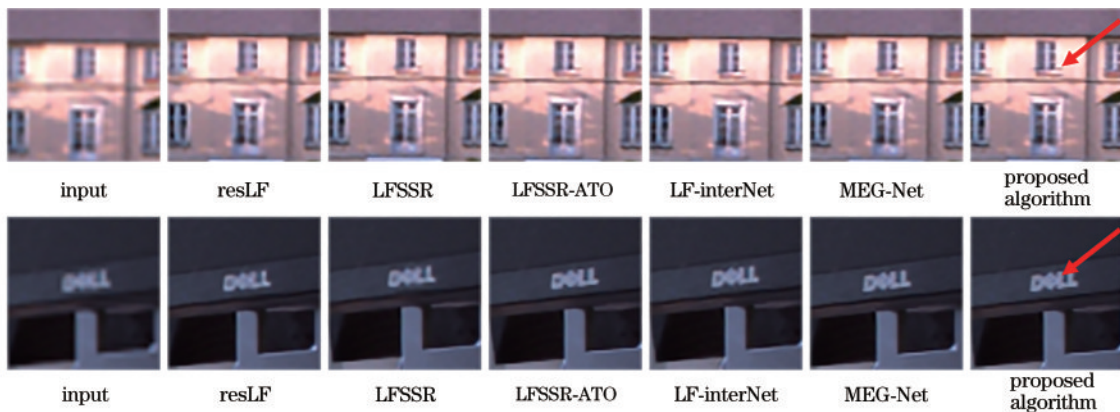


图8 真实世界场景下不同算法视觉效果对比

Fig. 8 Comparison of the visual effects of different algorithms in real-world scenes

FCA_module、SC_module对网络做出的贡献和对超分辨率结果产生的影响,以PSNR和SSIM作为定量评价指标,数据结果如表5所示,公平起见,其中,未加

入的模块用和模块中深度相同的卷积块代替,卷积核大小为3×3。

从表5的结果可以看出:只使用空间信息或只使用

表 5 消融实验的结果比较

Table 5 Comparison of results of ablation experiments

Model	PSNR	SSIM
Bicubic	29.740	0.9376
Spatial only	32.853	0.9464
Angular only	30.674	0.9398
Non-FIAF_module	33.887	0.9587
Non-FCA_module	34.185	0.9638
Non-SC_module	34.401	0.9695
All module	34.493	0.9767

角度信息时最后的重建结果较差,同时应用角度和空间信息可以显著提高重建质量;当去掉空间角度信息交互融合模块时,两个指标下降较多,可见将空间信息和角度信息充分交互并用角度信息辅助空间信息可以更充分利用光场信息,提高重建质量;当去掉空间细节细化模块时,造成了信息的冗余,无法突出有用信息;当去掉结构一致性模块时,降低了相邻视图间的一致性;当所有模块都添加进网络时,PSNR 和 SSIM 分别达到了最高值,重建效果最好,显著提高了网络的性能。

4 结 论

提出一种基于特征交互融合与注意力的光场图像超分辨率网络,充分利用光场图像视图间的互补信息来提升重建质量。通过空间特征提取卷积和角度特征提取卷积分别提取光场图像的空间信息和角度信息,同时用残差块和空间空洞金字塔提取深层空间信息,从而得到多尺度空间信息来增大感受野;为了充分利用所有视图的空间角度信息,将空间信息和角度信息充分交互并将空间角度信息融合达到利用角度信息辅助空间信息的目的;为了解决信息冗余的问题,加入通道注意力模型来抑制冗余信息,突出有用信息,恢复高频信息;最后,通过一致性模块来校正光场的结构一致性,保持相邻视图间的视差结构。在多个光场数据集上的实验结果表明,所提算法可以有效恢复图像的高频细节,使边缘更加清晰,比现有的多数的 2D 图像和光场图像超分辨率算法得到更好的重建效果。但是对于大视差场景的重建质量不是很好,未来的工作中考虑用金字塔模型来处理大视差问题。

参 考 文 献

- [1] 孟俊熙,张莉,曹洋,等.基于 Deeplab v3+ 的图像语义分割算法优化研究[J].激光与光电子学进展,2022,59(16):1610009.
Meng J X, Zhang L, Cao Y, et al. Optimization of image semantic segmentation algorithms based on Deeplab v3+ [J]. Laser & Optoelectronics Progress, 2022, 59(16): 1610009.
- [2] 袁梦凯,朱新军,侯林鹏.基于 R2U-Net 的单帧投影条纹图深度估计[J].激光与光电子学进展,2022,59(16):

1610001.

Yuan M K, Zhu X J, Hou L P. Depth estimation from single-frame fringe projection patterns based on R2U-Net [J]. Laser & Optoelectronics Progress, 2022, 59(16): 1610001.

- [3] 邱森森,柴雄力,邵枫.一种立体全景图像显著性检测模型[J].激光与光电子学进展,2021,58(8):0810002.
Qiu M M, Chai X L, Shao F. Saliency detection model for stereoscopic panoramic images[J]. Laser & Optoelectronics Progress, 2021, 58(8): 0810002.
- [4] Guo M T, Hou J H, Jin J, et al. Deep spatial-angular regularization for light field imaging, denoising, and super-resolution[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(10): 6094-6110.
- [5] Wanner S, Goldluecke B. Variational light field analysis for disparity estimation and super-resolution[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(3): 606-619.
- [6] Mitra K, Veeraraghavan A. Light field denoising, light field superresolution and stereo camera based refocussing using a GMM light field patch prior[C]//2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, June 16-21, 2012, Providence, RI, USA. New York: IEEE Press, 2012: 22-28.
- [7] Alain M, Smolic A. Light field super-resolution via LFBM5D sparse coding[C]//2018 25th IEEE International Conference on Image Processing, October 7-10, 2018, Athens, Greece. New York: IEEE Press, 2018: 2501-2505.
- [8] Rossi M, Frossard P. Geometry-consistent light field super-resolution via graph-based regularization[J]. IEEE Transactions on Image Processing, 2018, 27(9): 4207-4218.
- [9] Yoon Y, Jeon H G, Yoo D, et al. Learning a deep convolutional network for light-field image super-resolution[C]//2015 IEEE International Conference on Computer Vision Workshop, December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 57-65.
- [10] Yuan Y, Cao Z Q, Su L J. Light-field image superresolution using a combined deep CNN based on EPI[J]. IEEE Signal Processing Letters, 2018, 25(9): 1359-1363.
- [11] Wang Y L, Liu F, Zhang K B, et al. LFNet: a novel bidirectional recurrent convolutional neural network for light-field image super-resolution[J]. IEEE Transactions on Image Processing, 2018, 27(9): 4274-4286.
- [12] Zhang S, Lin Y F, Sheng H. Residual networks for light field image super-resolution[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 16-21, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 11038-11047.
- [13] Yeung H W F, Hou J H, Chen X M, et al. Light field spatial super-resolution using deep efficient spatial-angular separable convolution[J]. IEEE Transactions on Image Processing, 2019, 28(5): 2319-2330.
- [14] Jin J, Hou J H, Chen J, et al. Light field spatial super-resolution via deep combinatorial geometry embedding and structural consistency regularization[C]//2020 IEEE/

- CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 2257-2266.
- [15] Zhang S, Chang S, Lin Y F. End-to-end light field spatial super-resolution network using multiple epipolar geometry[J]. IEEE Transactions on Image Processing, 2021, 30: 5956-5968.
- [16] Wang Y Q, Wang L G, Yang J G, et al. Spatial-angular interaction for light field image super-resolution[M]// Vedaldi A, Bischof H, Brox T, et al. Computer vision-ECCV 2020. Lecture notes in computer science. Cham: Springer, 2020, 12368: 290-308.
- [17] Hu J, Shen L, Albanie S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.
- [18] Mei Y Q, Fan Y C, Zhang Y L. Pyramid attention networks for image restoration[EB/OL]. (2020-04-28) [2022-03-05]. <https://arxiv.org/abs/2004.13824v4>.
- [19] Wang Q L, Wu B G, Zhu P F, et al. ECA-net: efficient channel attention for deep convolutional neural networks [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), August 23-28, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 11531-11539.
- [20] Rerabek M, Ebrahimi T. New light field image dataset [EB/OL]. [2022-06-03]. <https://infoscience.epfl.ch/record/218363>.
- [21] Honauer K, Johannsen O, Kondermann D, et al. A dataset and evaluation methodology for depth estimation on 4D light fields[M]//Lai S H, Lepetit V, Nishino K, et al. Computer vision-ACCV 2016. Lecture notes in computer science. Cham: Springer, 2017, 10113: 19-34.
- [22] Wanner S, Meister S, Goldlücke B. Datasets and benchmarks for densely sampled 4D light fields[EB/OL]. [2022-03-06]. http://publications.lightfield-analysis.net/WMG13_vmv.pdf.
- [23] Pendu M L, Jiang X R, Guillemot C. Light field inpainting propagation via low rank matrix completion[J]. IEEE Transactions on Image Processing, 2018, 27(4): 1981-1993.
- [24] Vaish V, Adams A. The (new) stanford light field archive [EB/OL]. [2022-06-05]. <http://lightfield.stanford.edu/>.
- [25] Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 26-July 1, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 1646-1654.
- [26] Lim B, Son S, Kim H, et al. Enhanced deep residual networks for single image super-resolution[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 1132-1140.
- [27] Zhang Y L, Li K P, Li K, et al. Image super-resolution using very deep residual channel attention networks[EB/OL]. (2018-07-08) [2022-03-06]. <https://arxiv.org/abs/1807.02758>.