

DOI:10.13232/j.cnki.jnju.2022.02.013

## 融合全局和局部特征的光场图像空间超分辨率算法

井花花, 晏涛\*, 刘渊

(江南大学人工智能与计算机学院, 无锡, 214122)

**摘要:** 光场相机传感器有限的空间分辨率阻碍了光场图像处理相关研究的进展. 提出一种融合全局和局部特征的光场图像空间超分辨率算法, 提高了对光场子视点全局关系建模的能力. 由于光场相机捕捉的图像亮度较低, 严重影响了超分辨率图像的质量, 提出一个改进的 4D 零参考深度曲线估计网络 (4D Zero-DCE-Net), 充分利用光场全部子视点信息来提高光场图像的亮度. 为了解决光场图像空间分辨率低的问题, 提出一个基于生成对抗网络的光场图像空间超分辨率网络模型. 生成器包含三个部分: 第一部分是 Transformer 和 4D 卷积以并行方式结合的网络结构, 能以较浅的网络层捕捉图像的全局和局部细节信息; 第二部分是一个交互融合注意力模块 IFAM (Interactive Fusion Attention Module), 能有效地融合上述两个分支得到的全局自注意力和局部细节信息; 第三部分是一个重建模块 PS-PA (Pixel Shuffle-Pixel Attention), 能提高整个光场的空间分辨率. 最后, 利用相对判别器来指导生成器的训练. 实验结果表明, 提出的算法和其他算法相比, 峰值信号比 (PSNR) 至少提升了 1 dB.

**关键词:** 光场图像, 超分辨率, Transformer, 4D 卷积

**中图分类号:** TP391

**文献标志码:** A

## A spatial super-resolution method for light field images by fusing global and local features

Jing Huahua, Yan Tao\*, Liu Yuan

(School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, 214122, China)

**Abstract:** The limited spatial resolution of sensor of light field camera hinders the progress of light field image processing related research. This paper proposes a spatial super-resolution algorithm for light field images by integrating global and local features, which improves the ability of modeling the global relationship between light field sub-views. Since brightness of captured light field images is low and seriously affects the quality of the super-resolution image, this paper proposes an improved 4D Zero-DCE-Net to make full use of all sub-views of a light field to enlighten light field images. In order to solve the problem of low spatial resolution of light field images, we propose a spatial super-resolution network model of light field images based on generative adversarial network. The generator consists of three parts. The first part is a network structure that combines Transformer and 4D convolution in a parallel manner. It captures global and local details of the images with a shallower network layer. The second part proposes an interactive fusion attention module (IFAM) to effectively fuse the global self-attention and local detail information from the above two branches. The third part is a reconstruction module (PS-PA) to improve the spatial resolution of the entire light field. Finally, the relative discriminator is used to guide the training of the generator. Extensive experimental results show that our proposed method improves the PSNR (Peak Signal to Noise Ratio) performance index by at least 1 dB than other methods.

**Key words:** light field image, super resolution, Transformer, 4D convolution

基金项目: 国家自然科学基金 (61902151)

收稿日期: 2021-10-29

\* 通讯联系人, E-mail: yantao@jiangnan.edu.cn

随着光场相机的快速发展,光场图像处理算法日益增多.光场图像与传统的2D图像不同,一次曝光即可记录整个四维光场的结构信息,包括光线的强度和方向信息,因此可以更好地支撑多个计算机视觉领域的研究,例如重聚焦<sup>[1]</sup>、三维重建<sup>[2-5]</sup>、虚拟/增强现实<sup>[6-7]</sup>、图像去除运动模糊<sup>[8]</sup>和去雨<sup>[9]</sup>等.

通常使用两平面参数化来表示4D光场,即 $L(x, y, s, t)$ , $(x, y)$ 表示空间平面, $(s, t)$ 表示角度平面.近年来,研究者们已经提出多个传统的光场图像超分辨率算法,对视点之间的关系进行物理建模,将超分辨率视作优化问题.Wanner and Goldluecke<sup>[10-11]</sup>在EPI (Epipolar Plane Image)上使用结构张量来估计视差图,该视差图可用在变体框架中实现空间和角度超分辨率.Mitra and Veeraraghavan<sup>[12]</sup>提出一个通用框架,使用以视差值为条件的高斯混合模型对光场进行建模.Cho et al<sup>[13]</sup>提出一种基于稀疏编码的框架来训练低分辨率和高分辨率光场图像对的字典来提高算法的性能.

由于光场数据的高维特征,上述传统超分辨率算法的效果非常有限.随着深度神经网络研究的快速发展和成功,涌现了一些基于深度学习的算法来解决光场空间超分辨率的问题,它们主要利用多视点冗余性以及视点之间的补充信息来学习从低分辨率到高分辨率的映射关系.Rossi and Frossard<sup>[14]</sup>提出一个基于图的正则化器来增强光场的几何结构,从而提高光场全部视点的分辨率.Yoon et al<sup>[15-16]</sup>首次提出使用CNN(Convolutional Neural Networks)处理光场数据,同时实现角度和空间超分辨率.Wang et al<sup>[17]</sup>提出一个双向循环卷积神经网络,分别对水平和垂直两个方向上相邻子视点的空间相关性进行建模,然后利用堆栈泛化技术将两者集成起来获得高分辨率光场图像.Zhang et al<sup>[18]</sup>提出一个残差网络来实现光场图像的超分辨率,将四个不同方向的视点堆叠起来进入网络的四个分支提取特征,并将四个分支的残差信息融合在一起得到重建结果.Yeung et al<sup>[19]</sup>提出一种空间角度可分离的卷积模块,可以更高效地提取空间和角度的特征,进而更

好地恢复光场图像的细节信息.Meng et al<sup>[20]</sup>提出一种高维密集残差卷积神经网络来重建光场,将光场的所有子视点作为输入,并使用4D卷积获取视点之间的关系.Wang et al<sup>[21]</sup>提出一个空间角度交互网络实现光场图像的超分辨率,首先从宏像素图像中提取角度和空间特征,然后通过交互模块实现空间-角度的信息交互,最后利用光场重塑得到高分辨率图像.Jin et al<sup>[22]</sup>提出一种All-to-One模块,通过当前视点与参考视点的组合相关性充分利用光场视点之间的补充信息,并使用结构一致性正则化模块来恢复光场的几何结构.

本文在光场超分辨率之前首先对图像进行亮度增强.由于光场相机捕捉的图像亮度较低,造成光场分解得到的子视点图像因存在较多噪声/坏点等导致图像失真的问题,影响了超分辨率图像的质量.真实场景的光场图像没有真值,无法用有监督的方法训练网络.Guo et al<sup>[23]</sup>提出零参考深度曲线估计网络(Zero-DCE-Net),增强了单张图像的亮度,效果比较好,但由于光场子视点从边缘到中心的亮度范围不同,边缘子视点会偏暗.因此,本文在Zero-DCE-Net的基础上将2D卷积改成4D卷积,提出改进的4D Zero-DCE-Net,利用各子视点之间亮度信息(光场的几何结构信息)来提高整个光场图像的亮度.

为了保证光场多个子视点结构的完整性和一致性,需要获取光场图像的全局特征,然后再提取局部特征来补充图像的细节信息.4D卷积神经网络在光场图像超分辨率上已经取得不错的效果,但是其全局建模能力较弱,不能有效地捕捉每个子视点内长距离的依赖关系,所以需要不断增加卷积层获得更大感受野来对图像的全局关系建模,这不但会增加大量的参数,而且低水平特征会逐渐消失,进而影响生成图像的质量.因此,本文分开提取全局和局部特征.Transformer的感受野比较大,它对序列到序列建模和并行计算的能力可以很好地对光场数据的全局关系建模.4D残差卷积充分地利用光场几何结构信息获取全部子视点的局部特征,使其能够生成局部细节丰富和全局信息一致的高分辨率光场图像.

本文提出一个基于深度学习的光场图像空间超分辨率算法,如图1所示.首先对光场图像进行

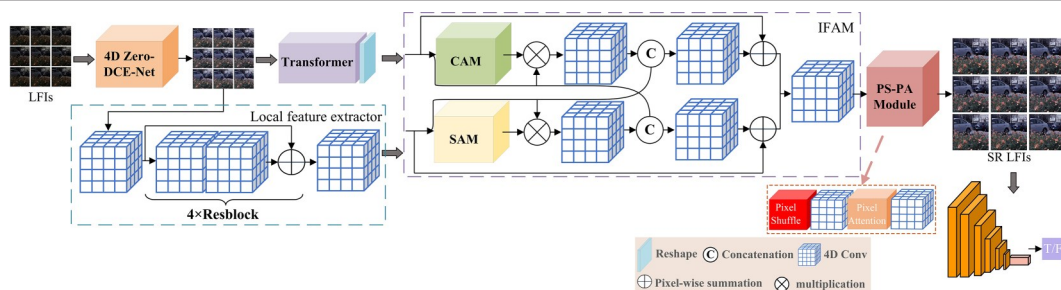


图 1 本文的网络结构图

Fig. 1 Network structure proposed in this paper

亮度增强的预处理,然后提出一个生成对抗网络(Generative Adversarial Networks, GAN)来实现光场图像超分辨率.生成器由三部分组成:第一部分是由 Transformer 和 4D 残差卷积以并行方式结合的网络结构来提取全局和局部信息;第二部分利用交互融合注意力模块(IFAM)融合两个分支的信息;第三部分通过重建模块(PS-PA)来提高整个光场图像的空间分辨率.最后,通过相对判别器<sup>[24]</sup>指导神经网络的训练.

## 1 算法原理

本文算法由光场图像亮度增强和光场图像超分辨率两个模块组成,具体步骤如下:(1)提出改进的 4D 零参考深度曲线估计网络(4D Zero-DCE-Net),对光场图像做亮度增强;(2)利用 Transformer 建模全局关系的能力和 4D 卷积对全部视点空间相关性建模的归纳偏置,构建一个由 Transformer 和 4D 卷积以并行方式结合的网络模型;(3)提出一个融合模块(IFAM)融合两个分支的信息,促进对全局和局部信息的提取并进一步增强图像的细节信息;(4)提出一个重建模块(PS-PA),使网络关注有用的信息来提高整个光场图像的空间分辨率.下面详细介绍算法的每个部分.

**1.1 光场图像亮度增强(预处理)** 低亮度的光场图像会影响视觉效果和图像传递信息的质量,因此,本文在零参考深度曲线估计网络(Zero-DCE-Net)<sup>[23]</sup>的基础上设计一个 4D 零参考深度曲线估计网络(4D Zero-DCE-Net),充分利用光场的几何结构信息来实现整个光场图像的亮度增强,将图像亮度增强任务建模为一个基于轻量级

深度神经网络的图像特定曲线估计问题.为了估计像素级的高阶亮度增强曲线以实现输入图像亮度的动态范围调整,算法主要依据像素的取值范围、单调性、可微性,设计了一组无参考的损失函数来隐式测量亮度增强的效果并驱动神经网络的学习.4D Zero-DCE-Net 的核心是亮度增强曲线、4D 深度曲线估计网络(4D DCE-Net)和零参考的损失函数.

首先,设计一种能够将低亮度图像自动映射到增强图像的曲线,曲线参数是自适应的且仅取决于输入图像.曲线需要满足增强图像的像素值归一化为 $[0, 1]$ 、曲线是单调的以及在梯度反向传播过程中是可导的三个要求,如式(1)所示:

$$LE(I(x); \alpha) = I(x) + \alpha I(x)(1 - I(x)) \quad (1)$$

其中,  $x$  为像素坐标,  $LE(I(x); \alpha)$  为输入图像  $I(x)$  的增强结果,  $\alpha \in [-1, 1]$  为可训练的曲线参数.

为了使模型能适应弱光条件的挑战以及可以在更宽的动态范围内调整图像,通过迭代和重新定义  $\alpha$  为一个逐像素参数的方式来对式(1)做进一步改进,如式(2)所示:

$$LE_n(x) = LE_{n-1}(x) + M_n(x)LE_{n-1}(x)(1 - LE_{n-1}(x)) \quad (2)$$

其中,  $LE_n(x)$  为输入图像亮度增强的结果,  $n$  为迭代次数,  $M$  为与输入图像大小一样的参数图.

然后,利用 4D DCE-Net 来学习输入图像和曲线参数图之间的映射关系.4D DCE-Net 由具有对称连接的七个 4D 卷积层构成,输入为低亮度图像,输出为一组用于高阶曲线的曲线参数图.最后,设置四个无参考损失函数:空间一致损失、曝光控制损失、颜色恒等损失以及平滑度损失来



约束网络,关于损失函数的更多细节可参考文献[23].

**1.2 光场图像超分辨率** 超分辨率网络结构是一个生成对抗网络,生成器主要由 Transformer 和 4D 卷积以并行方式结合的网络模型、融合模块和重建模块三部分组成,判别器采用相对判别器<sup>[24]</sup>. 本文的主要任务是将低分辨率的光场图像  $I^L \in \mathbb{R}^{S \times T \times X \times Y}$  重建为高分辨率的光场图像  $I^H \in \mathbb{R}^{S \times T \times rX \times rY}$ , 其中,  $r$  表示上采样因子, 本文设置为 4,  $(S, T)$  和  $(X, Y)$  分别表示角度坐标和空间坐标.

**1.2.1 全局特征提取器** 本文将光场的全部视点输入 Transformer<sup>[25]</sup> 中提取全局特征, 这与传统的 Transformer 不同. 传统的 Transformer 只对单张图像进行注意力操作, 而 Bertasius et al<sup>[25]</sup> 的 Transformer 可以对多张图像进行注意力操作, 它通过对不同视点相同位置的图像块进行注意力操作, 再对同一视点不同位置的图像块进行空间注意力操作来提取光场全部子视点的信息. 如图 2 所示, 同时取多个视点同一位置的图像块(如图中左上角红色框位置)进行注意力操作, 无红色框的图像块表示没有进行注意力操作.



图2 不同视点相同位置的图像块的注意力计算示意图  
Fig. 2 Schematic diagram of attention computation for image patches at the same position for all sub-views

首先将所有视点表示为  $F \in \mathbb{R}^{H \times W \times C \times V}$ , 每个视点分成  $N$  个大小为  $P$  的图像块, 并把这些图像块展开为向量  $f_{(p,t)} \in \mathbb{R}^{P^2 \times C}$ , 其中,  $N = HW/P^2$ ,  $p = 1, \dots, N$  表示空间位置,  $t = 1, \dots, V$  表示视点下标. 为了保存每个图像块的位置, 加入一个可学习的位置编码  $E_{(p,t)} \in \mathbb{R}^D$ ,  $D$  表示多头注意力的总维度. 通过矩阵  $M \in \mathbb{R}^{D \times P^2 \times C}$  将每个  $f_{(p,t)}$  映射到嵌入向量  $Z_{(p,t)}^{(0)}$ , 表示为  $Z_{(p,t)}^{(0)} = Mf_{(p,t)} + E_{(p,t)}$ , 令其作为 Transformer 的输入. Transformer 包含  $l$  个

编码块, 对于每个编码块,  $q_{(p,t)}^{(l,h)}, k_{(p,t)}^{(l,h)}, v_{(p,t)}^{(l,h)}$  可以根据前一个编码块得到的编码  $Z_{(p,t)}^{(l-1)}$  计算得到:

$$\begin{aligned} q_{(p,t)}^{(l,h)} &= W_q^{(l,h)} LN(Z_{(p,t)}^{(l-1)}) \\ k_{(p,t)}^{(l,h)} &= W_k^{(l,h)} LN(Z_{(p,t)}^{(l-1)}) \\ v_{(p,t)}^{(l,h)} &= W_v^{(l,h)} LN(Z_{(p,t)}^{(l-1)}) \end{aligned} \quad (3)$$

其中,  $h$  表示多头注意力的索引,  $LN$  是层归一化操作,  $W_q^{(l,h)}, W_k^{(l,h)}$  以及  $W_v^{(l,h)}$  表示权重.

自注意力权重是通过点乘计算得到的, 首先计算不同视点相同位置的图像块的自注意力权重, 如式(4)所示:

$$\mu_{(p,t)}^{(l,h) \text{ views}} = SM \left( \left( \frac{q_{(p,t)}^{(l,h)}}{\sqrt{D_h}} \right)^T \cdot \left[ k_{(0,0)}^{(l,h)} \{ k_{(p',t)}^{(l,h)} \}_{p'=1, \dots, V} \right] \right) \quad (4)$$

其中,  $D_h$  表示每个头注意力的维度,  $SM$  表示 softmax 激活函数,  $( )^T$  表示转置操作,  $\mu_{(p,t)}^{(l,h) \text{ views}}$  表示不同视点相同位置的图像块的自注意力权重.

同一视点图像块的空间注意力的权重计算如式(5)所示:

$$\mu_{(p,t)}^{(l,h) \text{ space}} = SM \left( \left( \frac{q_{(p,t)}^{(l,h)}}{\sqrt{D_h}} \right)^T \cdot \left[ k_{(0,0)}^{(l,h)} \{ k_{(p',t)}^{(l,h)} \}_{p'=1, \dots, N} \right] \right) \quad (5)$$

其中,  $\mu_{(p,t)}^{(l,h) \text{ space}}$  表示同一视点图像块的空间注意力的权重.

第  $l$  块的编码  $Z_{(p,t)}^{(l)}$  是通过每个头注意力的自注意力系数来计算值向量的加权和, 然后将来自所有头注意力的这些向量级联在一起得到的.  $Z_{(p,t)}^{(l)}$  根据式(3)获得新的  $q_{(p,t)}^{(l,h)}, k_{(p,t)}^{(l,h)}, v_{(p,t)}^{(l,h)}$ , 利用式(5)来计算空间自注意力, 然后计算值向量的加权和, 将来自所有头注意力的这些向量级联在一起得到向量  $Z_{(p,t)}^{(l) \text{ space}}$ , 再输入  $MLP$  中得到在  $l$  块的编码  $Z_{(p,t)}^{(l)}$ , 最后将结果重建为  $H \times W \times C \times V$  的特征图  $F_1$ . 如式(6)所示:

$$\begin{aligned} Z_{(p,t)}^{(l)} &= MLP \left( LN \left( Z_{(p,t)}^{(l) \text{ space}} \right) \right) + Z_{(p,t)}^{(l) \text{ space}} \\ F_1 &= Reshape \left( Z_{(p,t)}^{(l)} \right) \end{aligned} \quad (6)$$

其中,  $MLP$  表示多层感知机,  $LN$  是层归一化操作.

**1.2.2 局部特征提取器** 本文利用浅层(仅四层)4D 残差网络从光场的全部视点中学习局部几

何特征,充分利用了视点之间的补充信息以及光场的几何结构信息.通过浅层的4D残差网络,不仅缓解了梯度消失的问题,保留了低水平光场几何特征,而且能更好地恢复光场的结构和细节信息.4D残差模块中包含两层步长为1,卷积核尺寸为 $3 \times 3 \times 3 \times 3$ 的4D卷积.如式(7)所示:

$$F_2 = H_{4DRBs}(I^r) \quad (7)$$

其中, $I^r$ 表示光场全部子视点, $H_{4DRBs}$ 表示4D残差卷积操作.

**1.2.3 融合模块(IFAM)** 为了有效地融合从Transformer和4D卷积获取的特征,本文提出一个新的融合模块IFAM,它包含空间注意力(SAM)和通道注意力(CAM),并以交互的方式实现两个分支信息的融合.SAM首先对特征图进行两个池化操作,然后进入 $7 \times 7$ 卷积和Sigmoid激活函数得到空间上的权重.CAM使特征图经过自适应的平均和最大池化操作,池化后的特征图再分别通过 $3 \times 3$ 卷积,然后进行像素相加,最后通过Sigmoid激活函数得到通道上的权重.利用通道注意力模块促进来自Transformer分支全局信息的提取,利用空间注意力模块增强来自4D残差网络的局部细节信息.经过注意力模块得到的权重与原始图像相乘后的特征图再与另外一个分支的信息经过 $3 \times 3$ 卷积融合在一起后和原始的信息相加,最后再通过 $3 \times 3$ 卷积把两个分支的信息融合在一起.如式(8)所示:

$$\begin{aligned} F'_1 &= \text{Conv}[\text{Conv}(CA(F_1) \times F_1), F_2] + F_1 \\ F'_2 &= \text{Conv}[\text{Conv}(SA(F_2) \times F_2), F_1] + F_2 \\ F &= \text{Conv}[F'_1, F'_2] \end{aligned} \quad (8)$$

其中, $F_1$ 和 $F_2$ 分别表示全局和局部特征, $CA$ 和 $SA$ 分别表示通道和空间注意力, $\text{Conv}$ 表示卷积, $F'_1$ 和 $F'_2$ 分别表示经过注意力模块后与另外一个分支信息融合后的特征, $F$ 表示最终融合后的特征.

**1.2.4 重建模块(PS-PA)** PS-PA主要由PixelShuffle层、卷积层和像素注意力层组成,融合后的信息进入PS-PA中得到高分辨率的光场图像.首先,融合后的信息进入PixelShuffle层,经过像素重排来放大图像,具体是通过卷积获得通道数为 $r$ 的平方的特征图,然后利用PS(Periodic

Shuffling)获得高分辨率图像,则 $r$ 是上采样因子;然后进入 $3 \times 3$ 卷积提取特征,获取的特征进入像素注意力模块中通过 $1 \times 1$ 卷积和Sigmoid激活函数为图像的每个像素都分配权重,从而使网络关注有用的信息;最后,通过 $3 \times 3$ 卷积得到高分辨率图像.本文在重建模块引入像素注意力,既可以提高卷积的表达能力,又能更好地恢复光场图像的细节信息.如式(9)所示:

$$I^{sr} = R_{pspa}(F) \quad (9)$$

其中, $I^{sr}$ 表示最终超分辨率后的结果, $R_{pspa}$ 表示重建操作, $F$ 表示融合后的特征.

网络训练的总损失函数由三部分组成:重建损失、感知损失以及GAN损失,如下所示:

$$L_{\text{total}} = (1 - \lambda_1 - \lambda_2)L_{\text{rec}} + \lambda_1 L_{\text{per}} + \lambda_2 L_{\text{Gs}} \quad (10)$$

其中, $L_{\text{rec}}$ 是重建损失, $L_{\text{per}}$ 是感知损失, $L_{\text{Gs}}$ 是GAN损失, $\lambda_1$ 和 $\lambda_2$ 是 $L_{\text{per}}$ 和 $L_{\text{Gs}}$ 在总损失函数中的权重.

重建损失被定义为生成的高分辨率图像和真值的 $L_1$ 损失,如下所示:

$$L_{\text{rec}} = \frac{1}{ST} \sum_{s=1}^S \sum_{t=1}^T \|I_{s,t}^{HR} - I_{s,t}^{SR}\|_1 \quad (11)$$

其中, $I_{s,t}^{HR}$ 是真值, $I_{s,t}^{SR}$ 是生成的结果, $S$ 和 $T$ 是光场的角度坐标.

为了提高超分辨率结果的视觉质量,引入感知损失:

$$L_{\text{per}} = \frac{1}{ST} \sum_{s=1}^S \sum_{t=1}^T \|\phi(I_{s,t}^{HR}) - \phi(I_{s,t}^{SR})\|_2^2 \quad (12)$$

其中, $\phi$ 表示从VGG中提取的特征图.

判别器和生成器损失函数如下所示:

$$\begin{aligned} L_{\text{Ds}} &= \frac{1}{ST} \sum_{s=1}^S \sum_{t=1}^T (D_{\text{Ra}}(I_{s,t}^{HR}, I_{s,t}^{SR}) - 1)^2 + D_{\text{Ra}}(I_{s,t}^{SR}, I_{s,t}^{HR})^2 \\ L_{\text{Gs}} &= \frac{1}{ST} \sum_{s=1}^S \sum_{t=1}^T (D_{\text{Ra}}(I_{s,t}^{SR}, I_{s,t}^{HR}) - 1)^2 + D_{\text{Ra}}(I_{s,t}^{HR}, I_{s,t}^{SR})^2 \end{aligned} \quad (13)$$

其中, $L_{\text{Ds}}$ 和 $L_{\text{Gs}}$ 分别表示判别器和生成器损失函数, $D_{\text{Ra}}$ 表示相对判别器.

## 2 实验结果

**2.1 网络训练细节** 利用斯坦福真实场景光场数据集<sup>[26]</sup>和LytroIllum光场相机采集的光场图像同时进入网络训练,训练过程中输入 $7 \times 7$ 的光场图像子视点阵列.共使用180组光场数据,其中,

120组光场图像用来训练,20组光场图像用来评估模型,40组光场图像用来测试。

使用PyTorch框架在Titan X GPU上进行网络训练,采用Adam优化器更新参数,学习率设置为 $2\text{E}-4$ ,训练和测试的批量大小分别取2和1,训练次数epoch设为500, $\lambda_1$ 和 $\lambda_2$ 分别设为0.04和0.01.不同 $\lambda_1$ 和 $\lambda_2$ 实验结果的峰值信号比(Peak Signal to Noise Ratio, *PSNR*)与结构相似性(Structural Similarity, *SSIM*)的平均值如表1所示,表中黑体字是最优结果.由表可见,本文拟定的深度神经网络参数可以使实验结果达到一个理想的指标。

表1 不同超参数的*PSNR*与*SSIM*平均值

Table 1 Average value of *PSNR* and *SSIM* under different hyperparameters

Settings	S1	S2	S3	S4	S5	S6
$\lambda_1$	0.02	0.02	0.04	0.04	0.06	0.06
$\lambda_2$	0.01	0.02	0.01	0.02	0.01	0.02
<i>PSNR</i>	31.37	30.89	<b>34.46</b>	33.78	32.86	32.02
<i>SSIM</i>	0.921	0.916	<b>0.963</b>	0.942	0.931	0.928

实验阶段,首先将低亮度的光场图像输入4D Zero-DCE-Net中实现亮度增强,将经过降采样得到的低分辨率光场图像输入由Transformer和4D残差卷积以并行方式结合的网络中提取全局和局部特征;接着,进入IFAM中融合两者的信息,再进入PS-PA中得到高分辨率的光场图像.4D Zero-DCE-Net在训练时,输入低亮度的光场图像作为训练集进行无监督的训练,得到一个最优模型,再将测试集输入模型进行测试.超分辨率网络在训练时,将低分辨率的光场图像作为训练集和验证集输入网络来训练模型,得到一个最优的模型.将测试集输入模型中进行测试,然后利用真值对测试结果进行评估来判断模型的性能.这里注意,超分辨率网络模型输入的是已经提高亮度后的低分辨率图像。

**2.2 光场图像亮度增强结果定量分析** 对光场图像亮度增强的结果进行评估,由于没有真值和参考图,所以利用自然图像质量评估器(NIQE)进行定量比较.NIQE是一种无参考图像质量评估工具,NIQE越低表示图像的视觉质量越好.如

表2所示,亮度增强后的图像的NIQE明显低于输入图像,图像的质量也更好。

表2 光场图像亮度增强的定量比较结果

Table 2 Quantitative comparison results of light field image brightness enhancement

Input	NIQE ↓
Original image	4.235
Result	3.363

**2.3 光场图像亮度增强结果定性分析** 光场图像亮度增强的结果如图3所示,图中第一行和第三行是原始的光场图像,第二行和第四行是亮度增强后的结果.由图可见,提亮后的光场图像视觉效果更好,一些细节更清晰。

**2.4 图像超分辨率结果定量分析** 将本文算法与基于优化的超分辨率算法<sup>[13]</sup>以及基于深度学习的算法<sup>[21-22,27-28]</sup>进行定量分析对比,其中Dai et al<sup>[27]</sup>,Jo et al<sup>[28]</sup>分别是单张图像和视频图像的超分辨率算法.为了实验的公平性,所有算法都在我们的训练数据集上训练,在 $7\times 7$ 光场的中心子视点上采用*PSNR*,*SSIM*,*PI*(Perceptual Index)<sup>[29]</sup>以及*BRISQUE*(Blind/Referenceless Image Spatial Quality Evaluator)<sup>[30]</sup>四个度量标准来对本文的算法和其他算法进行定量分析.其中,*PI*用来评估图像的感知质量,*BRISQUE*用来评估在空间域图像的质量,*PI*和*BRISQUE*越低,图像的质量越高.定量分析结果如表3所示,表中黑体字表示最优的结果.由表可见,本文算法的*PSNR*和*SSIM*均高于其他算法,*PI*和*BRISQUE*也都低于其他算法。

本文也在自己采集的光场图像上进行测试,由于采集的光场图像没有经过降采样操作直接输入网络进行测试,所以没有其对应的高分辨率图像,只能利用*PI*来评估结果.如图4至图9所示,每张图下都标注了各算法结果对应的*PI*,可以看出,本文算法的*PI*是最低的,因而图像的感知质量是最高的。

**2.5 图像超分辨率结果定性分析** 如图4、图5和图6所示,第一行从左到右依次是插值、Dai et al<sup>[27]</sup>、Jo et al<sup>[28]</sup>和Rossi and Frossard<sup>[14]</sup>的实验结果,第二行从左到右依次是Jin et al<sup>[22]</sup>、Wang et





图 3 光场图像亮度增强结果

Fig. 3 Brightness enhancement results of light field images

表 3 光场超分辨率的定量比较结果(40组测试数据)

Table 3 Quantitative comparison results of light field super-resolution (40 arrays of test data)

Method	PSNR $\uparrow$	SSIM $\uparrow$	PI $\downarrow$	BRISQUE $\downarrow$
Bicubic	25.67	0.802	8.47	66.87
SAN <sup>[27]</sup>	30.88	0.912	6.22	50.36
DUF <sup>[28]</sup>	30.45	0.902	6.26	51.07
GB <sup>[13]</sup>	28.17	0.865	6.48	57.62
LFSR-ATO <sup>[22]</sup>	31.15	0.928	6.16	49.29
LFIT <sup>[21]</sup>	32.22	0.934	6.01	48.53
Ours	<b>34.46</b>	<b>0.963</b>	<b>5.23</b>	<b>42.29</b>

al<sup>[21]</sup>、本文算法的结果和真值(GT)。由图可见, Wang et al<sup>[21]</sup>和 Jin et al<sup>[22]</sup>的算法表现良好, Dai et al<sup>[27]</sup>和 Jo et al<sup>[28]</sup>的算法在某些场景的结果也不错,但在纹理区域上还有一定程度的模糊,而本文算法的结果更加锐利,在一些复杂的场景的优势更明显。由图4可见, Wang et al<sup>[21]</sup>的算法表现优于前面几个算法,但也没本文算法的结果锐利。

本文还在自己采集的数据集上进行了定性分析,结果是本文算法的结果明显优于其他算法。如图7和图8所示,本文算法在文字的细节恢复方

面明显好于其他算法。图9的结果也远胜其他算法。同时可以看出, Rossi and Frossard<sup>[14]</sup>和 Wang et al<sup>[21]</sup>的算法在重建过程中损失了很多高频细节信息。更多的实验结果请参考 <https://github.com/jhh1996/LFSR/tree/main/experiment>。

**2.6 消融实验** 为了证明提出的模块能有效地提高算法性能,进行了四组消融实验:用残差卷积来代替 IFAM 融合两个分支的信息、PS-PA 去掉像素注意力模块只用 PixelShuffle、利用另外一个 4D 残差卷积网络代替 Transformer 分支提取特征以及利用另外一个 Transformer 代替 4D 残差卷积网络分支来提取特征。实验结果如表4所示,表中黑体字表示最优结果。由表可见,本文提出的融合模块和改进的重建模块明显好于基础模块。IFAM 和残差卷积能有效地融合两个分支的信息并增强细节信息,重建模块引入像素注意力模块后使网络更加关注有用的信息,从而有利于图像细节信息的恢复。从表4还可以看出,第一种情况利用另外一个 4D 残差卷积网络代替 Transformer 分支提取特征,缺少了全局信息;第二种情况只利用



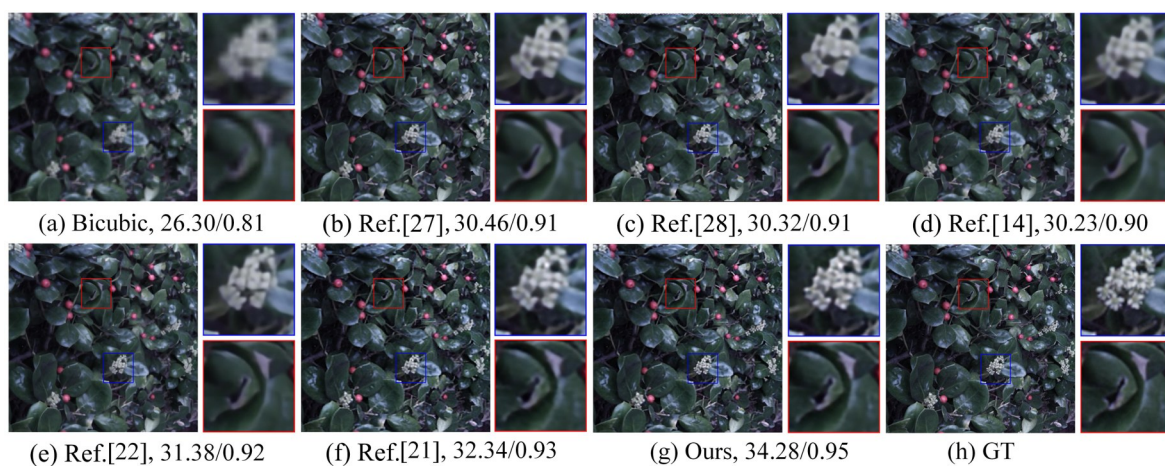


图4 光场图像超分辨率结果比较(场景1)

Fig. 4 Super-resolution results of light field images (Scene 1)

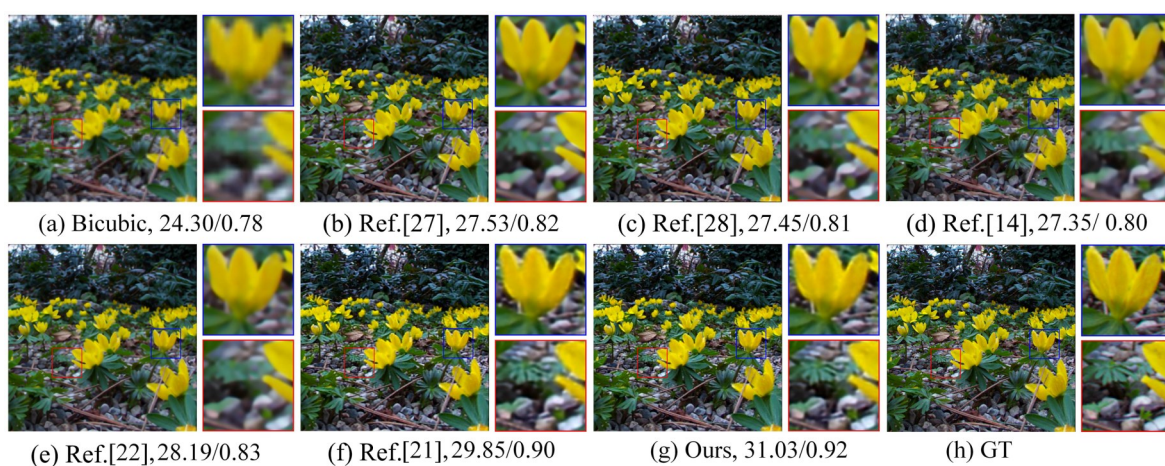


图5 光场图像超分辨率结果比较(场景2)

Fig. 5 Super-resolution results of light field images (Scene 2)

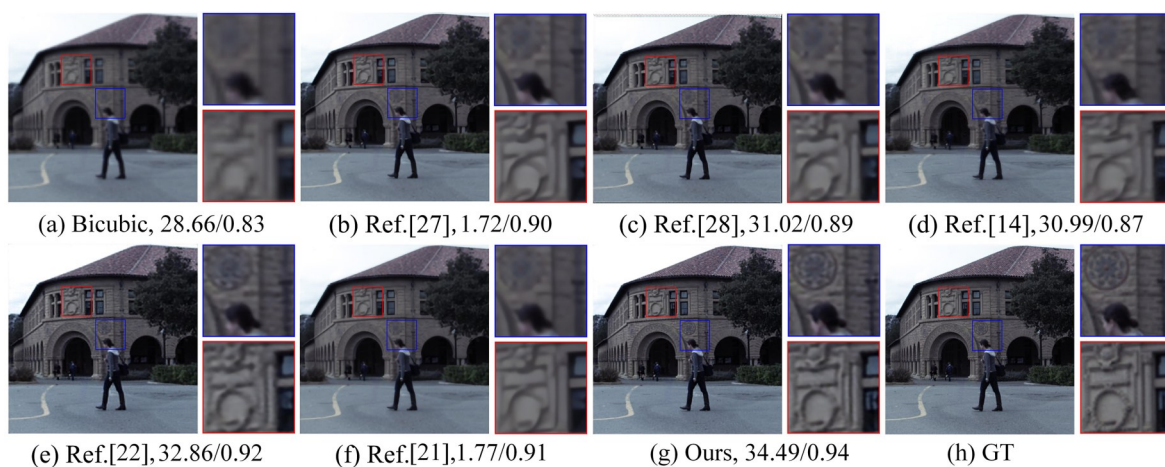


图6 光场图像超分辨率结果比较(场景3)

Fig. 6 Super-resolution results of light field images (Scene 3)



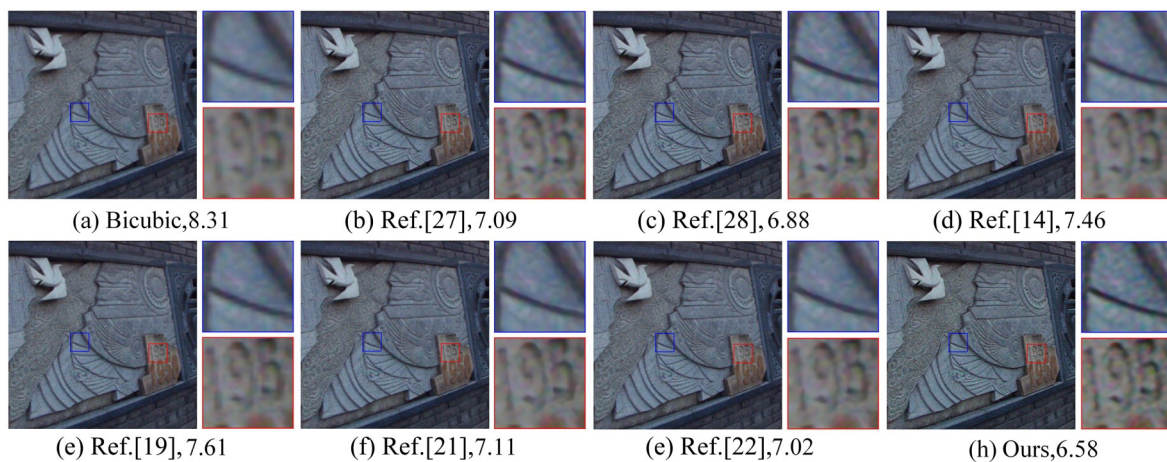


图 7 光场图像超分辨率结果比较(场景 4)

Fig. 7 Super-resolution results of light field images (Scene 4)

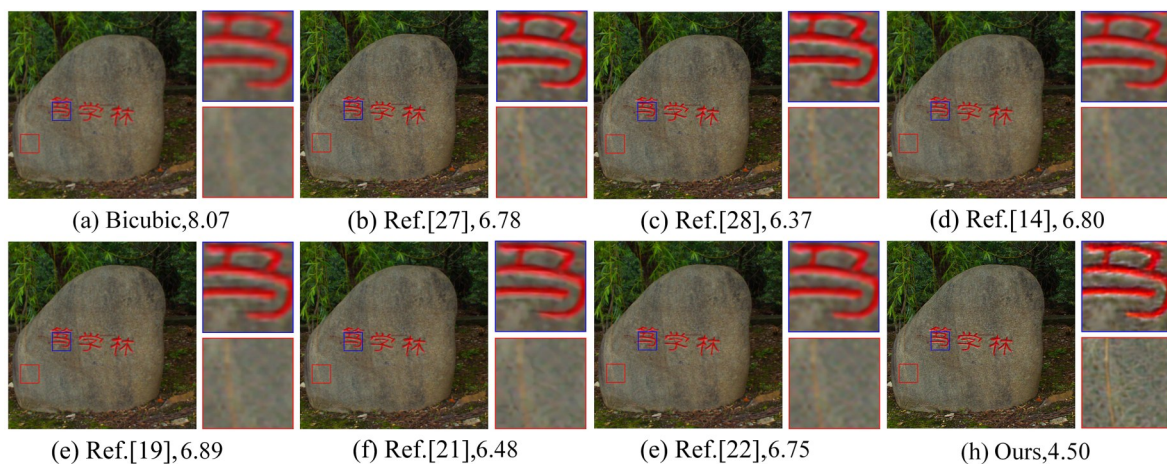


图 8 光场图像超分辨率结果比较(场景 5)

Fig. 8 Super-resolution results of light field images (Scene 5)



图 9 光场图像超分辨率结果比较(场景 6)

Fig. 9 Super-resolution results of light field images (Scene 6)

表4 不同模块的超分辨率结果

Table 4 Super resolution results of different modules

Model	PSNR $\uparrow$	SSIM $\uparrow$	PI $\downarrow$	BRISQUE $\downarrow$
Concat+Res	32.43	0.932	5.95	47.84
PixelShuffle	33.58	0.954	5.82	46.56
Dual 4D Resnet	31.68	0.926	6.08	48.45
Dual Transformer	31.45	0.923	6.11	48.89
Ours	<b>34.46</b>	<b>0.963</b>	<b>5.23</b>	<b>42.29</b>

Transformer 提取特征,缺少了局部细节信息,这两种情况下模型的性能都明显下降。

### 3 结论

本文利用无监督的4D深度学习网络模型对光场图像进行亮度增强,提高初始图像传递信息的质量;提出一种基于深度神经网络的光场图像空间超分辨率算法,通过Transformer和4D卷积两个网络分支充分利用光场全部子视点之间的信息,以一个较浅的网络层捕捉到图像的全局和局部细节信息;通过IFAM将上述两个分支获得的信息有效地融合在一起,最后进入PS-PA重建模块得到高分辨率的光场图像。在真实光场数据集上的实验结果表明,本文算法能有效地提高整个光场的空间超分辨率,与其他算法相比,细节恢复得更好。但是,真实场景的光场图像内容往往比较复杂,所以本文的实验结果在一些细节方面的修复还不够好,下一步需要研究改进网络模型和训练策略,可以借助高分辨率的单反图来进一步提升光场图像超分辨率的质量。

#### 参考文献

- [1] Fiss J, Curless B, Szeliski R. Refocusing plenoptic images using depth-adaptive splatting//2014 IEEE International Conference on Computational Photography. Santa Clara, CA, USA: IEEE, 2014: 1–9.
- [2] Zhu H, Wang Q, Yu J Y. Occlusion-model guided antiocclusion depth estimation in light field. IEEE Journal of Selected Topics in Signal Processing, 2017, 11(7):965–978.
- [3] Kim C, Zimmer H, Pritch Y, et al. Scene reconstruction from high spatio-angular resolution light fields. ACM Transactions on Graphics, 2013, 32(4):73.
- [4] Zhu H, Zhang Q, Wang Q. 4D light field superpixel and segmentation//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017:6709–6717.
- [5] Si L P, Wang Q. Dense depth-map estimation and geometry inference from light fields via global optimization//Proceedings of the 13<sup>th</sup> Asian Conference on Computer Vision. Springer Berlin Heidelberg, 2016:83–98.
- [6] Huang F C, Luebke D, Wetzstein G. The light field stereoscope//ACM SIGGRAPH 2015 Emerging Technologies. Los Angeles, CA, USA: ACM, 2015: Article No.24.
- [7] Yu J Y. A light-field journey to virtual reality. IEEE MultiMedia, 2017, 24(2):104–112.
- [8] Srinivasan P P, Ng R, Ramamoorthi R. Light field blind motion deblurring//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017:2354–2362.
- [9] Ding Y Y, Li M Y, Yan T, et al. Rain streak removal from light field images. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(2):467–482.
- [10] Wanner S, Goldluecke B. Spatial and angular variational super-resolution of 4D light fields//European Proceedings of the 12<sup>th</sup> Conference on Computer Vision. Springer Berlin Heidelberg, 2012: 608–621.
- [11] Wanner S, Goldluecke B. Variational light field analysis for disparity estimation and super-resolution. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(3):606–619.
- [12] Mitra K, Veeraraghavan A. Light field denoising, light field superresolution and stereo camera based refocussing using a GMM light field patch prior//2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. Providence, RI, USA: IEEE, 2012: 22–28.
- [13] Cho D, Lee M, Kim S, et al. Modeling the calibration pipeline of the Lytro camera for high quality light-field image reconstruction//2013 IEEE International



- Conference on Computer Vision. Sydney, Australia: IEEE, 2013: 3280—3287.
- [14] Rossi M, Frossard P. Geometry-consistent light field super-resolution via graph-based regularization. *IEEE Transactions on Image Processing*, 2018, 27(9): 4207—4218.
- [15] Yoon Y, Jeon H G, Yoo D, et al. Learning a deep convolutional network for light-field image super-resolution//2015 IEEE International Conference on Computer Vision Workshop. Santiago, Chile: IEEE, 2015: 57—65.
- [16] Yoon Y, Jeon H G, Yoo D, et al. Light-field image super-resolution using convolutional neural network. *IEEE Signal Processing Letters*, 2017, 24(6): 848—852.
- [17] Wang Y L, Liu F, Zhang K B, et al. LFNet: A novel bidirectional recurrent convolutional neural network for light-field image super-resolution. *IEEE Transactions on Image Processing*, 2018, 27(9): 4274—4286.
- [18] Zhang S, Lin Y F, Sheng H. Residual networks for light field image super-resolution//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE, 2019: 11038—11047.
- [19] Yeung H W F, Hou J H, Chen X M, et al. Light field spatial super-resolution using deep efficient spatial-angular separable convolution. *IEEE Transactions on Image Processing*, 2019, 28(5): 2319—2330.
- [20] Meng N, So H K H, Sun X, et al. High-dimensional dense residual convolutional neural network for light field reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(3): 873—886.
- [21] Wang Y Q, Wang L G, Yang J G, et al. Spatial-angular interaction for light field image super-resolution//Proceedings of the 16<sup>th</sup> European Conference on Computer Vision. Springer Berlin Heidelberg, 2020: 290—308.
- [22] Jin J, Hou J H, Chen J, et al. Light field spatial super-resolution via deep combinatorial geometry embedding and structural consistency regularization//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE, 2020: 2257—2266.
- [23] Guo C L, Li C Y, Guo J C, et al. Zero-reference deep curve estimation for low-light image enhancement//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE, 2020: 1777—1786.
- [24] Jiang Y F, Gong X Y, Liu D, et al. EnlightenGAN: Deep light enhancement without paired supervision. *IEEE Transactions on Image Processing*, 2021(30): 2340—2349.
- [25] Bertasius G, Wang H, Torresani L. Is space-time attention all you need for video understanding?. 2021, arXiv:2102.05095.
- [26] <http://lightfield.stanford.edu/lfs.html>.
- [27] Dai T, Cai J R, Zhang Y B, et al. Second-order attention network for single image super-resolution//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE, 2019: 11057—11066.
- [28] Jo Y, Oh S W, Kang J, et al. Deep video super-resolution network using dynamic upsampling filters without explicit motion compensation//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 3224—3232.
- [29] Vasu S, Madam N T, Rajagopalan A N. Analyzing perception-distortion tradeoff using enhanced perceptual super-resolution network//European Conference on Computer Vision. Springer Berlin Heidelberg, 2018: 114—131.
- [30] Mittal A, Moorthy A K, Bovik A C. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 2012, 21(12): 4695—4708.

(责任编辑 杨可盛)