应用光学
Journal of Applied Optics
SCOPUS,等来源期刊
中文核心期刊、中国科技核心期刊

基于亚像素和梯度引导的光场图像超分辨率

韦玮 陈芬 张华波 罗英国 张鹏 彭宗举

Light field images super-resolution based on sub-pixel and gradient guide

WEI Wei, CHEN Fen, ZHANG Huabo, LUO Yingguo, ZHANG Peng, PENG Zongju

引用本文:
韦玮, 陈芬, 张华波, 等. 基于亚像素和梯度引导的光场图像超分辨率[J]. 应用光学, 2024, 45(5): 956-965. DOI: 10.5768/JAO202445.0502003
WEI Wei, CHEN Fen, ZHANG Huabo, et al. Light field images super-resolution based on sub-pixel and gradient guide[J]. Journal of Applied Optics, 2024, 45(5): 956-965. DOI: 10.5768/JAO202445.0502003

在线阅读 View online: https://doi.org/10.5768/JAO202445.0502003

文章编号：1002-2082 (2024) 05-0956-10

基于亚像素和梯度引导的光场图像超分辨率

韦 玮，陈 芬，张华波，罗英国，张 鹏，彭宗举
（重庆理工大学 电气与电子工程学院，重庆 400054）

摘 要：针对光场相机捕获到的光场图像空间分辨率较低等问题，提出一种基于亚像素和梯度引导的光场图像超分辨率方法。设计了多重亚像素信息提取模块，该模块将子孔径图像分为水平、垂直、对角和反对角 4 个图像堆，并分别提取各图像堆的亚像素信息。同时，考虑到梯度先验可为预测高频细节提供有效线索，在重建过程中融合了子孔径图像的梯度多重亚像素信息。在 5 个公开数据库上的实验结果表明，本文方法不仅在客观指标上普遍优于现有方法，在主观视觉效果上也有更好的表现，边缘纹理细节更加清晰。

关键词：超分辨率；光场图像；亚像素信息；深度学习
中图分类号：TN201;TP391.4 **文献标志码：**A **DOI:** 10.5768/JAO202445.0502003

Light field images super-resolution based on sub-pixel and gradient guide

WEI Wei, CHEN Fen, ZHANG Huabo, LUO Yingguo, ZHANG Peng, PENG Zongju
（School of Electronics and Electronic Information Engineering, Chongqing University of Technology, Chongqing 400054, China）

Abstract: For the problem of low spatial resolution of light field images captured by light field cameras, a super-resolution method for light field images based on sub-pixel and gradient guide was proposed. A multiple sub-pixel information extraction module was designed, which divided the sub-aperture images into four image stacks: horizontal, vertical, diagonal and anti-diagonal, and extracted the sub-pixel information of each image stack separately. Meanwhile, considering that the gradient prior could provide effective clues for predicting high-frequency details, the gradient multiple sub-pixel information of the sub-aperture images was fused in the reconstruction process. Experimental results on five publicly available databases show that the proposed method not only generally outperforms the existing methods in terms of objective indexes, but also has better performance in the subjective visual effect, which the edge texture details are clearer.

Key words: super-resolution; light field images; sub-pixel information; deep learning

引言

光场同时包含了场景的空间信息和角度信息，在计算机视觉和计算机图形学等领域具有广泛的应用前景，可用于目标检测[1]、深度估计[2]和三维（three dimension, 3D）重建[3]等。光场作为一种高维数据，在 3D 世界中很难被表示出来，需要特殊的光场成像设备记录。与普通相机相比，光场相机在主透镜和图片传感器之间嵌入了一个微透镜阵列，从多个角度同时记录空间光线的方向和强度，能为用户提供更为丰富的场景和运动信息。由于受到光场相机中微透镜阵列的限制，光场图像的分辨率普遍较低，阻碍了光场的应用，因此出现了大量关于光场图像超分辨率的研究。

图像超分辨率是指采用图像处理和机器学习

收稿日期：2023-06-25； **修回日期：**2023-10-28
基金项目：重庆市自然科学基金 (cstc2021jcyj-msxmX0411, CSTB2022NSCQ-MSX0873)；重庆理工大学研究生教育高质量发展行动计划资助成果 (gzlcx20233138, gzlcx20233081)
作者简介：韦玮（1999—），男，硕士研究生，主要从事光场超分辨率研究。E-mail: 545645775@qq.com
通信作者：陈芬（1973—），女，博士、教授，主要从事图像和视频信号处理研究。E-mail: chenfen@cqut.edu.cn

技术, 从已有的低分辨率(low-resolution, LR)图像重建高分辨率(high-resolution, HR)图像的技术。传统单图像超分辨率(single image super resolution, SISR)只需要考虑单幅图像的空间信息, 而光场图像超分辨率还需要考虑角度信息。提高光场图像的空间分辨率最直接的方法是用 SISR 算法单独超分所有视图, 但这种方法忽略了光场图像的角度信息和不同视图间的相关性, 重建图像质量较低。

光场图像各视图之间不是独立的, 而是具有高度相关性并存在大量信息冗余。根据光场图像的一致性, 同一像素可以在不同视图中被找到, 因此在光场图像超分辨率时, 通过同时利用多个视图的空间信息可以提高重建图像的质量。视图中物体的形状和位置会随着视图角度的不同而变化, 因此光场图像中各视图间存在着像素差异性, 这种差异主要体现在各像素的光线方向和强度上, 在超分辨率过程中忽略这些差异将导致重建图像出现伪影、失真。此外, 在图像超分辨率过程中会出现纹理细节丢失问题, 由于图像边缘是最重要的图像特征之一, 一些方法[4-5]利用梯度引导单图像超分辨率, 增强了重建图像的边缘细节。

为了解决光场图像各视图之间的像素差异性问题, 同时更好地保护重建图像的边缘纹理细节, 本文提出一种基于亚像素和梯度引导的光场图像超分辨率方法。通过提取光场图像中的亚像素信息来弥补视图间的像素差异性, 同时融合梯度先验来增强重建图像的边缘细节。具体而言, 首先构建了多重亚像素信息提取(multiple sub-pixel information extraction, MSPIE)模块提取光场图像中的多重亚像素信息, 该模块将输入的子孔径图像(sub-aperture image, SAI)分为水平、垂直、对角和反对角图像堆, 通过特征提取块提取各个图像堆的亚像素信息; 其次通过特征表示增强模块对各视图间的全局上下文信息进行特征增强; 然后融合了各子孔径图像的梯度多重亚像素增强特征; 最后通过重建模块重建出高分辨率的光场图像。

1　相关工作

由于传感器分辨率的限制, 光场相机很难同时获取具有高空间和高角度分辨率的光场图像。为了在现有硬件基础上解决这个问题, 许多研究人员把研究重点放在算法角度上。YOON Y 等人[6]首次将卷积神经网络引入到光场图像超分辨率中, 将所有视图输入到空间超分辨率网络中, 提高其空间分辨率, 再将超分辨率后的图像分堆后输入到不同角度的超分辨率网络, 从而完成空间和角度的超分辨率。与传统方法相比, 该方法性能有了较大提升, 因此, 此后基于深度学习的超分辨率方法成为了光场图像超分辨率中的主要研究方向。

光场图像可以被表示为宏像元图像(macro-pixel image, MacPI)、SAI 和极平面图像(epipolar plane image, EPI), 并且 3 种图像可以相互转换。MacPI 中混合了角度信息和空间信息, 一些方法利用 MacPI 作为输入, 有效提取和合并了空间信息和角度信息。WANG Y Q 等人[7]设计了不同的特征提取器分别提取空间和角度特征, 通过空间-角度相互作用模块合并空间和角度信息, 该特征提取器在重建过程中有效提取了光场图像的空间信息和角度信息, 提高了重建质量。YEUNG H W F 等人[8]设计了空间-角度可分离卷积模块来分别提取空间和角度特征, 该模块首先对空间特征进行空间卷积, 然后将空间特征重塑为角度特征, 再用角度卷积提取角度特征, 该方法提出的可分离卷积能充分提取空间和角度信息。WU G C 等人[9]首先对输入的 EPI 进行模糊操作, 然后用细节恢复网络恢复 EPI 的角度细节, 最后再用去模糊操作恢复空间细节, 该方法有效处理了光场图像的空间和角度维度的信息不对称问题。安平等人[10]将输入图像分为水平和垂直图像堆, 然后提取对应水平和垂直 EPI 特征, 该方法有效利用了光场图像的纹理信息, 提高了重建结果的几何一致性。

SAI 可以视为在不同角度下记录的图像, 每个 SAI 包含了空间信息, 角度信息被隐式地包含在所有 SAI 中。将 SAI 作为输入便于提取光场图像的空间信息。ZHANG S 等人[11]提出了基于残差网络的方法, 将 SAI 分为水平、垂直和对角图像堆, 提取不同空间方向的特征。ZHANG S 等人[12]在特征提取部分用 3D 卷积替代传统二维卷积, 重建质量有了较大提升。JIN J 等人[13]通过组合所有视图的信息来超分辨率每个视图。此外, 在网络中加入的空间角度可分离卷积可以保护光场的结构一致性。WANG Y Q 等人[14]设计了角度可变形对齐模块, 完成空间和角度特征对齐, 该模块的可变形卷积可以根据不同情况调整卷积核的形状, 能更好

地提取特征。WANG S Z 等人[15]首次将 Transformer 引入到光场图像超分辨率，并将光场图像超分辨率分为图像内容超分辨率和图像梯度超分辨率，提出了 DPT，引入的 Transformer 可以增强不同视图之间的相关性，对梯度的超分辨率可以在重建过程中保护图像细节，取得了较好的超分结果，受到了广泛关注。

尽管这些方法取得了不错的成效，但现有方法主要通过级联多个卷积层获得大的感受野来覆盖差异范围，不能很好地解决光场图像超分辨率时，各视图之间的像素差异性问题。此外，大部分方法在重建过程中没有注重对光场图像边缘细节的保护，导致重建结果的边缘区域细节比较模糊。在光场图像超分辨率过程中，若充分利用周围视图为中心视图提供的不同空间方向的亚像素信息，通过在不同视图中找到相应的亚像素偏移，可有效提高中心视图的空间分辨率。同时，考虑梯度先验引导的图像超分辨率，可有效增强重建图像的边缘纹理细节。因此，本文提出一种基于亚像素和梯度引导的光场图像超分辨率方法。

2 本文方法

光场可以被表示为 $L(x,y,s,t)$，其中 (x,y) 表示空间维度，(s,t) 表示角度维度，$L\in\mathbf{R}^{(H\times W\times S\times T)}$。光场图像可以看作 $S\times T$ 个 SAI 的阵列，每个 SAI 的分辨率为 $H\times W$ 像素。光场图像超分辨率的目标是从 LR 光场 $L_{\rm LR}(x,y,s,t)$ 重建出超分辨率光场 $L_{\rm SR}(x,y,s,t)$。$L_{\rm SR}\in\mathbf{R}^{(aH\times aW\times S\times T)}$，其中 a 为空间放大倍数。

$$L_{\rm SR}=f(L_{\rm LR}) \tag{1}$$

式中 f 表示从 LR 光场映射到 HR 光场的函数。

本文方法的网络结构如图 1 所示。网络主要分为两个分支，一个为主干分支，另一个为梯度分支，主干分支的输入为 SAI，梯度分支提取所有 SAI 梯度图作为输入。主干分支通过 MSPIE 模块将 SAI 分为水平、垂直、对角和反对角图像堆，提取各个图像堆的亚像素信息，然后经过增强 Transformer 进行特征表示增强。梯度分支和主干分支进行相同操作提取到 SAI 的梯度多重亚像素信息，通过融合 Transformer 将主干分支和梯度分支的输出特征进行融合，再通过重建模块完成上采样，最后与经过双三次插值的原始输入图像相加完成全局残差学习，得到最终重建出的高分辨率光场图像。

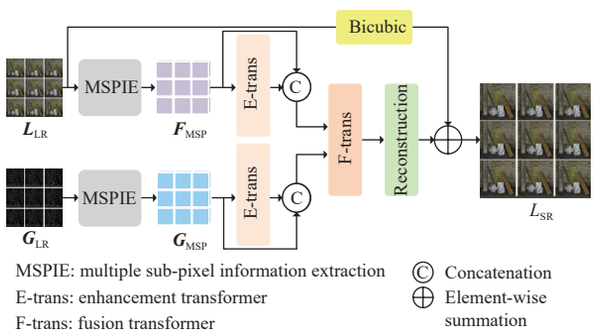

图 1 网络总体结构

Fig. 1 Overall structure diagram of proposed network

2.1 多重亚像素信息提取

与传统的多视图图像不同，光场图像中的视图有不同的角度方向，在一个角度方向的视图包含了中心视图在对应空间方向的亚像素偏移。图 2 为周围视图的亚像素信息，红星表示中心视图的像素信息，黄星、绿星、紫星和蓝星分别表示水平、垂直、对角和反对角方向的亚像素信息。图 2 表明了不同空间方向的亚像素偏移可以在对应角度方向的视图中找到。根据光场的图像一致性，如果相邻视图之间的像素差异为 1/3，通过从周围 3 个视图中提取对应像素，就可以将该区域上采样放大 3 倍。因此，根据差异信息找到不同视图之间的亚像素偏移，从而提高每个视图的空间分辨率。

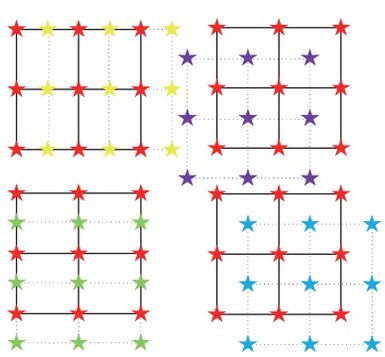

图 2 周围视图的亚像素信息

Fig. 2 Sub-pixel information from surrounding view images

将 SAI 按不同的角度方向叠加成图像堆，水平、垂直和对角方向的图像堆可以表示为

$$S^i_{0^\circ}=L(:,:,m,:),i\in\{1,2,\cdots,A\} \tag{2}$$

$$S^i_{90^\circ}=L(:,:,m,:),i\in\{1,2,\cdots,A\} \tag{3}$$

$$S^i_{45^\circ}=L(:,:,m,i+1-m),i\in\{1,2,\cdots,2A-1\} \tag{4}$$

$$S^i_{135^\circ}=L(:,:,m,A-i+m),i\in\{1,2,\cdots,2A-1\} \tag{5}$$

式中$m\in\{1,\cdots,A\}$。当$1\leqslant i\leqslant A$时, $1\leqslant m\leqslant A$; 当$i>A$时, $i-A\leqslant m\leqslant 2A-i$。水平和垂直方向各有$A$个图像堆, 每个图像堆有$A$张图像, 对角方向有$2A$-1个图像堆, 每个图像堆的图像数量不同。

通过固定光场图像L的一个角度维度(s,t)和空间维度(x,y), 可以得到对应的EPI。EPI同时包含了在空间维度和角度维度的信息, 并且可以有效反映光场的场景几何一致性。此外, EPI中线条的斜率反应了对应像素点的差异, 因此利用EPI可以解决光场超分辨率中差异问题。将SAI沿着固定方向叠成图像堆, 图像堆沿着某一方向的二维切片就是对应方向的EPI。为了进一步解释空间维度和角度维度之间的联系, 图3展示了水平方向的EPI中的亚像素信息。水平图像堆$S^i_{0^\circ}$可以通过将SAI沿着水平方向堆叠, 然后沿着水平方向切片, 得到的二维切片$E_{\rm H}(x,u)$就是水平EPI。

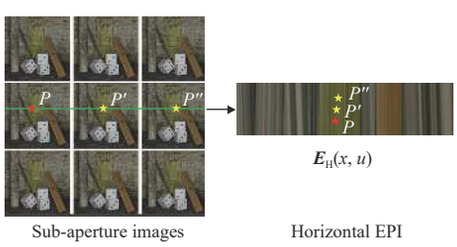

图3　水平极平面图像中的亚像素信息

Fig. 3　Sub-pixel information in horizontal EPI

在光场图像不同视图中的像素可以表示为

$$L(x,y,s,t)=L(x+\Delta x,y+\Delta y,s+\Delta s,t+\Delta t)\quad(6)$$

式中: $\Delta x=\Delta s\times d$, $\Delta y=\Delta t\times d$; d为参考点和其他点的像素差异大小。

根据d, 一个视图中的像素点可以与其他视图中的另一像素点相匹配。由于在角度维度的密集采样, 在一个视图中的亚像素信息可以在其他视图中找到。如图3所示, 参考点P的水平亚像素信息可以在周围2个视图中找到。参考点$L_P(x_P,y_P,s_P,t_P)$和$L_P{}'(x_P+\Delta x,y_P,s_P+\Delta x/d,t_P)$对应, 表示相应的水平亚像素信息可以在相邻的视图中提取到。在图3的水平EPI中, 参考点P对应P'和P''。此外, 垂直和对角亚像素信息可以分别在对应的垂直和对角EPI中找到。一旦找到了在中心EPI中4个方向的周围亚像素信息, 对应的区域就能完成超分辨率。

MSPIE模块用于提取不同方向的亚像素信息。如图4所示, MSPIE模块分为4个分支, 每个分支由特征提取块提取亚像素信息。特征提取块由1个3D卷积层及4个残差块组成, 残差块由1个PReLu激活层和3D卷积层组成。为了在角度维度对齐卷积特征, 加入了角度对齐模块(angular alignment module, AAM), AAM在文献[14]中的角度可变对齐模块的基础上进行了改进, 由于只需要对齐卷积特征, 因此移除了其中的可变卷积。由于EPI可以视为沿图像堆方向的二维切片, 因此, 特征提取块中的3D卷积可以提取到对应方向EPI中的亚像素信息。为了保证有足够大的感受野来提取亚像素信息, 卷积核大小设置为3×3×3, 并且每个特征提取块包含4个残差块。

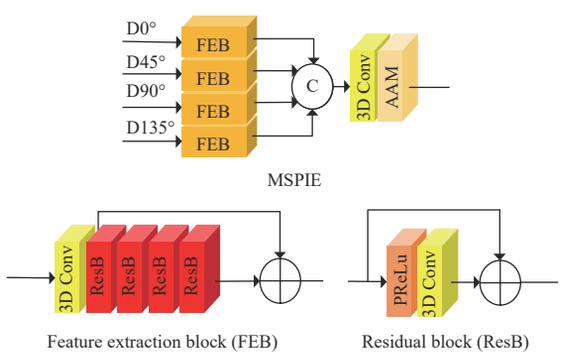

图4　多重亚像素信息提取模块

Fig. 4　Multiple sub-pixel information extraction blocks

经过特征提取模块的输出特征为F^i_θ:

$$F^i_\theta=H_{\rm FEB}(S^i_\theta)\quad(7)$$

式中: 当$\theta\in\{0^\circ,90^\circ\}$时, $i\in\{1,2,\cdots,A\}$; 当$\theta\in\{45^\circ,135^\circ\}$时, $i\in\{1,2,\cdots,2A-1\}$。

将F^i_θ重塑为5维数据$F_\theta\in\mathbf{R}^{(H\times W\times S\times T\times C)}$, 其中$C$为特征通道。将4个方向的特征连接, 通过1个卷积层进行降维, 最后通过AAM进行特征对齐, 得到多重亚像素特征$F_{\rm MSP}\in\mathbf{R}^{(H\times W\times S\times T\times C)}$。

2.2　特征表示增强

考虑到卷积特征只能独立捕捉每个SAI内的局部上下文信息, 因此缺乏不同SAI之间的全局上下文信息。与传统卷积不同, 视觉Transformer把一张图像视为一个符号的序列, 并且通过自注意力机制建立所有符号之间的关系, 因此加入Transformer可进一步丰富特征表示。由于光场图像由多个视图组成, 并且每个视图的相关性强, 因此可以将光场图像视为一个连续序列。受到文献[16]启发, 本文将视频超分辨率中的Transformer引入到光场图像超分辨率中用于特征表示增强, 以更好地捕捉各视图间的全局上下文信息, 该Transformer包含4个空间角度增强自注意力层。

将所有SAI视为一个序列, 特征序列$F_{\rm s}\in\mathbf{R}^{(L\times H\times W\times N)}$

作为 Transformer 的输入，其中 L 为序列长度，$N=4C$。图 5 为 Transformer 中的空间角度增强自注意力层，其中：U 为展开操作，用于从特征图中提取局部块；F 为折叠操作，用于组合局部块为原始特征图大小；T 为转置操作；$f_{\boldsymbol{Q}}$、$f_{\boldsymbol{K}}$、$f_{\boldsymbol{V}}$ 为卷积操作。当输入为图像时，Vit[17] 使用线性投影来提取几个特征块。与 Vit 不同，空间角度增强自注意力层将特征序列输入到 Transformer 以后，首先通过 3 个卷积分别提取空间特征，用展开操作来提取局部块，然后生成矩阵 $\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$，并对 $\boldsymbol{K}$ 进行转置操作生成矩阵 $\boldsymbol{K}^{\mathrm{T}}$。将 $\boldsymbol{K}^{\mathrm{T}}$ 和 $\boldsymbol{Q}$ 进行矩阵相乘得到 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 的相似性矩阵，再将相似性矩阵和 $\boldsymbol{Q}$ 进行矩阵相乘得到最终局部特征块，用折叠操作组合所有局部特征块，最后将 1×1 卷积层 f_{p} 的输出与 F_{s} 相加，得到增强特征 $F_{\mathrm{T}}\in\mathbf{R}^{(L\times H\times W\times N)}$。

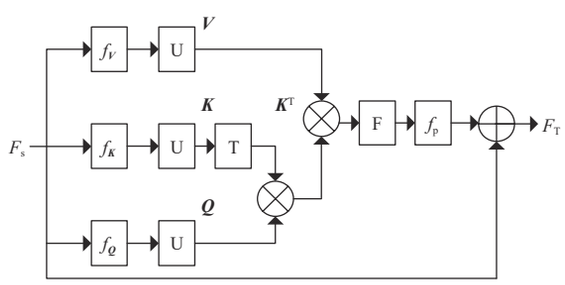

图 5 空间角度增强自注意力层

Fig. 5 Spatial-angular enhanced self-attention layer

2.3 梯度特征融合

为了在超分辨率过程中更好地保护重建图像的边缘细节，本文利用梯度来引导光场图像超分辨率。由于梯度图中包含有丰富的图像结构细节，为了充分利用其细节信息，设计了梯度分支以进一步提取梯度信息。与 DPT 利用多个卷积层提取梯度信息的方式不同，本文通过 MSPIE 模块提取多重亚像素梯度信息，从消融实验对比结果可知，所提取的多重亚像素梯度信息可进一步提高重建质量。具体过程如下：提取所有 SAI 的梯度图得到梯度阵列 $G_{\mathrm{LR}}\in\mathbf{R}^{(H\times W\times S\times T)}$，然后用 MSIPE 模块提取梯度多重亚像素特征 $G_{\mathrm{MSP}}\in\mathbf{R}^{(H\times W\times S\times T\times C)}$，再进行特征表示增强得到增强梯度特征 $G_{\mathrm{T}}\in\mathbf{R}^{(L\times H\times W\times N)}$，最后将光场图像增强特征和梯度图像增强特征融合完成结构细节增强。与空间角度增强自注意力层注重相同序列之间的关系不同，梯度特征融合需要注重梯度特征序列和原始特征序列之间的关系，因此对空间角度增强自注意力层进行了改进，将上面两个分支的输入改为梯度特征序列，下面分支的输入改为原始特征序列。图 6 为梯度融合跨注意力层，输出为融合特征 $F_{\mathrm{f}}\in\mathbf{R}^{(L\times H\times W\times N)}$。

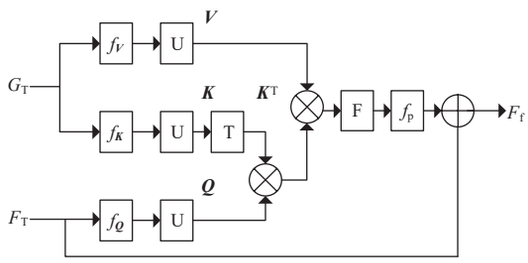

图 6 梯度融合跨注意力层

Fig. 6 Gradient fusion cross-attention layer

2.4 重建和上采样

融合梯度增强特征之后，再经过重建模块，重建模块由信息多重蒸馏模块[18] 和上采样模块组成，其中上采样模块包含 2 个卷积层和 1 个像素洗牌层。第 1 个卷积层用于将特征升维到 a^2N，然后用一个像素洗牌层上采样重建特征，得到目标分辨率 $aH\times aW$，最后用一个 1×1 卷积将通道降为 1。完成上采样后，再与经过双三次插值的原始输入图像相加得到重建结果 $L_{\mathrm{SR}}\in\mathbf{R}^{(aH\times aW\times S\times T)}$。

3 实验结果与分析

为了验证算法的有效性，训练和测试所用的 5 个公共数据库分别为 EPFL[19]、HCI_new[20]、HCI_old[21]、INRIA_Lytro[22] 和 Stanford_Gantry[23]。选取其中 144 个场景用于训练，23 个场景用于测试。所有光场图像的原始角度分辨率为 9×9 像素，训练和测试时将原始光场图像裁剪到合适分辨率，角度分辨率为 5×5 像素，空间分辨率 64×64 像素。使用双三次插值对原始图像进行下采样生成 LR 图像。

3.1 训练细节

本文网络在 NVIDIA RTX 4090 GPU 的 PC 上基于 Pytorch 框架进行训练。首先将光场图像从 RGB 三通道图像转换为 YCbCr 颜色空间，提取 Y 通道图像进行超分。生成视觉对比结果时，对 Cb 和 Cr 通道用双三次插值进行上采样。训练时将原始光场图像裁剪为 64×64 像素的图像块，步幅为 32，然后用双三次插值将图像块下采样到 32×32 像素作为网络的 LR 输入图像。对裁剪后的训练数据进行随机水平翻转、垂直翻转和 90°旋转来增加训练数据。由于光场图像中两个视图间的像素差异 d 较小，d 的范围在 0～1 之间，因此在训练时需

要重点考虑参考视图和前后两个视图的像素信息。3D 卷积核大小设置为 3×3×3, 其中空间尺寸为 3×3, 卷积深度为 3, 即利用到 3 个视图的信息。初始学习率设置为 $2e^{-4}$, 学习率优化方式为 ADAM, β_1=0.9, β_2=0.999。学习率每隔 15 个 epoch 减半, 总共训练 59 个 epoch, batchsize 为 8, 损失函数为平均绝对误差 $L1$ 损失函数。

$$l_{\mathrm{Loss}}(x,y)=\frac{1}{n}\sum_{i=1}^{n}|y_i-f(x_i)| \tag{8}$$

式中: y_i为原始高分辨率图像; $f(x_i)$为重建图像。

3.2　算法性能比较

为了评估本文方法的性能, 在 5 个公开数据集上与一些方法进行了对比, 包括了 VDSR[24]、EDSR[25]、resLF[11]、LFSSR[6]、LFSSR-ATO[13]、MEG-Net[12]、LF-Inter[6]、DPT[15], 其中 VDSR 和 EDSR 直接用对应 SISR 方法来超分光场图像中的每个视图。评价超分辨率性能的指标为峰值信噪比(peak signal to noise ratio, PSNR)和结构相似度(structural similarity, SSIM)。

表 1 和表 2 分别为 2 倍超分辨率结果对比和 4 倍超分辨率结果对比, 其中加粗的为最优结果, 加下划线的为次优结果。从表中结果可以看出, 本文方法的指标都为最优或次优, 证明了本文方法具有较好的超分辨率性能。

表 1　不同方法 $2^{\times}$超分辨率 PSNR(dB)/SSIM 对比

Table 1　PSNR(dB)/SSIM quantitative comparison of different methods $2^{\times}$ super-resolution

Method	Scale	EPFL	HCI_new	HCI_old	INRIA_Lytro	Stanford_Gantry
Bicubic	×2	29.88/0.902 4	31.65/0.937 3	37.46/0.969 5	31.77/0.908 6	30.75/0.944 1
VDSR	×2	32.50/0.959 8	34.37/0.956 1	40.61/0.986 7	34.44/0.974 1	35.54/0.978 9
EDSR	×2	33.09/0.962 9	34.83/0.959 2	41.01/0.987 4	34.98/0.976 4	36.30/0.981 8
ResLF	×2	33.62/0.970 6	36.69/0.973 9	43.42/0.993 2	35.39/0.980 4	38.35/0.990 4
LFSSR	×2	33.67/0.974 4	36.80/0.974 9	43.81/0.993 8	35.28/0.983 2	37.94/0.989 8
LFSSR-ATO	×2	34.27/0.975 7	37.24/0.976 7	44.21/0.994 2	36.17/0.984 2	39.64/0.992 9
LF-Inter	×2	34.11/0.976 0	37.17/0.976 3	44.57/0.994 6	35.83/0.984 3	38.43/0.990 9
MEG-Net	×2	34.31/**0.977 3**	37.42/0.977 7	44.10/0.994 2	36.10/**0.984 9**	38.77/0.991 4
DPT	×2	34.49/0.975 8	37.35/0.977 1	44.30/0.994 3	36.41/0.984 3	39.43/0.992 6
Proposed	×2	**34.53**/0.976 4	**37.64/0.978 5**	**44.65/0.994 6**	**36.43**/0.984 8	**39.89/0.993 3**

表 2　不同 $4^{\times}$超分辨率方法 PSNR(dB)/SSIM 对比

Table 2　PSNR(dB)/SSIM quantitative comparison of different methods $4^{\times}$ super-resolution

Method	Scale	EPFL	HCI_new	HCI_old	INRIA_Lytro	Stanford_Gantry
Bicubic	×4	25.46/0.771 1	27.58/0.863 6	32.42/0.914 3	27.30/0.805 8	25.91/0.832 2
VDSR	×4	27.25/0.877 7	29.31/0.882 3	34.81/0.951 5	29.76/0.920 4	28.51/0.900 1
EDSR	×4	27.83/0.885 4	29.59/0.886 9	35.18/0.953 6	29.65/0.925 6	28.70/0.907 2
ResLF	×4	28.26/0.903 5	30.72/0.910 7	36.71/0.968 2	30.34/0.941 2	30.19/0.937 2
LFSSR	×4	28.60/0.911 2	30.93/0.914 5	36.91/0.969 6	30.59/0.946 8	30.57/0.942 6
LFSSR-ATO	×4	28.51/0.911 5	30.88/0.913 5	37.00/0.969 9	30.71/0.948 4	30.61/0.943 0
LF-Inter	×4	28.81/0.916 2	30.96/0.916 1	37.15/0.971 6	30.78/0.949 1	30.36/0.940 9
MEG-Net	×4	28.75/0.916 0	31.10/0.917 7	37.29/0.971 6	30.67/0.949 0	30.77/0.945 3
DPT	×4	**28.94**/0.917 0	31.20/0.918 8	37.41/0.972 1	30.96/0.950 3	31.15/0.948 8
Proposed	×4	28.88/**0.918 0**	**31.34/0.921 2**	**37.55/0.972 9**	**31.01/0.951 3**	**31.25/0.950 0**

从表 1 可以看出, 在 2 倍超分辨率结果中, 本文提出的方法在 5 个数据库上均取得了不错的表现。其中 PSNR 指标在 5 个数据库上均为最优, SSIM 指标除了在 EPFL 和 INRIA_Lytro 数据库为次优, 在其余 3 个数据库上均为最优。与其他光场图像超分辨率方法相比, VDSR 和 EDSR 直接用 SISR 方法超分光场中的每个视图, 忽略了不同视图之间的相关性, 导致超分结果与其他

方法差距较大。由于 MEG-Net 和本文方法在超分辨率过程中均利用到了 EPI，而 EPI 可以有效反映光场的场景几何一致性，因此 SSIM 指标高于其他方法。在 4 倍超分辨率结果中，本文方法的 PSNR 指标在 EPFL 数据库上为次优，在其余 4 个数据库上均为最优，SSIM 指标在 5 个数据库上均为最优。由此可证明，本文方法在不同倍数超分辨率任务下，均能保持较高的超分辨率质量。

图 7 和图 8 为不同方法 2 倍超分辨率的视觉效果对比。图中左侧展示了对应场景的中心视图，右侧展示了不同方法超分辨率中心视图后的放大区域，并在下面列出了对应方法在该场景下的 PSNR 和 SSIM。图 7 为选取的 INRIA_Lytro 数据库中的 Building_Decoded 场景，图 8 为 HCI_new 数据库中的 Origami 场景。

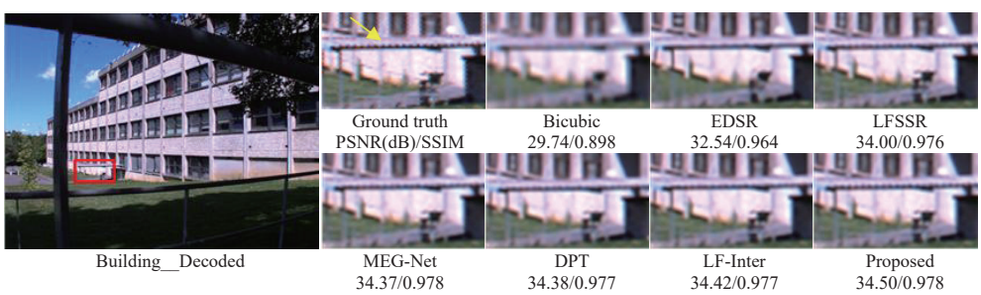

图 7　不同方法视觉效果对比 (场景 Building_Decoded)

Fig. 7　Comparison of visual effects of different methods (scene Building_Decoded)

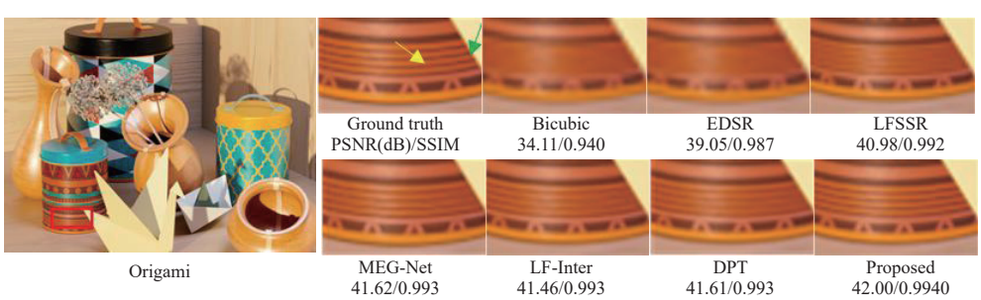

图 8　不同方法视觉效果对比 (场景 Origami)

Fig. 8　Comparison of visual effects of different methods (scene Origami)

可以看出，将单图像超分辨率算法直接运用到光场图像中的视觉效果明显不如光场图像超分辨率算法。从图 7 中可以看出，本文所提方法更接近原始光场图像，在黄色箭头所指的黑色纹理处，与其他方法相比，本文方法的超分辨率结果更加清晰且能复原更多纹理细节；EDSR 方法没有还原出纹理细节，并且图像整体较模糊；其余 4 种光场图像超分辨率算法虽然复原了一部分纹理细节，但还原的纹理细节部分还不够清晰。从图 8 中可以看出，在黄色箭头所指的黑色条纹处，虽然 LF-Inter、DPT 和 MEG 方法都重建出了黑色条纹，但是本文方法重建出的黑色条纹明显更加清晰；此外，在绿色箭头所指的边缘细节处，其余方法的重建效果较差，由于本文方法在重建过程中融合了梯度多重亚像素信息，因此重建结果还原出了光场图像的边缘纹理细节。

3.3　计算效率分析

为了进一步验证本文方法的计算效率，本文从网络模型的参数量（Params）和计算复杂度（Flops）角度来评估不同方法的计算效率。表 3 为不同超分辨率模型的参数量、Flops 和在 5 个数据库上的平均 PSNR 和 SSIM。计算 Flops 时，设置输入光场图像的大小为 5×5×32×32 像素。从表 3 可以看出，本文方法的模型在参数量和 Flops 较小的情况下，平均 PSNR 和 SSIM 的值最高，分别达到了 38.63 dB 和 0.985 5，具有较高的重建质量。与 EDSR 和 LFSSR-ATO 相比，本文方法的 Flops 明显更低。为了在重建中保护图像的边缘细节，本文方法设计了梯度分支，因此参数量和 Flops 略高于其他方法。

表 3　$2^{\times}$超分辨率模型参数量、Flops 和 PSNR/SSIM 对比

Table 3　Comparison of $2^{\times}$ super-resolution model parameters, Flops and PSNR/SSIM

Method	Params/M	Flops/G	PSNR (dB)/SSIM
EDSR	38.62	988.47	36.04/0.973 5
ResLF	8.65	37.06	37.49/0.981 7
LF-Inter	5.04	42.33	38.02/0.984 4
MEG-Net	1.69	48.39	38.14/0.985 1
LFSSR-ATO	1.22	453.03	38.31/0.984 7
DPT	3.66	57.32	38.40/0.984 8
Proposed	4.51	79.04	**38.63/0.985 5**

3.4　消融实验

为了探索光场图像超分辨率过程中亚像素信息和梯度的有效性, 本文进行了 4 组消融实验。实验在 EPFL 数据库上进行验证, 其中 70 个场景用于训练, 10 个场景用于测试。评价指标为 PSNR 和 SSIM。实验结果如表 4 所示。用于对比的特征提取部分与文献 [14] 和 DPT 相同, 由两个残差空洞空间金字塔池化模块(residual atrous spatial pyramid pooling blocks, ResASPP)和残差块(residual blocks, RB)构成, 即 ResASPP+RB 结构, MSPIE 为本文提出的多重亚像素信息提取模块。

表 4　消融实验对比

Table 4　Comparison of ablation experiment

Gradient branch		Main branch		Params/M	Flops/G	PSNR (dB)/SSIM
ResASPP+RB	MSPIE	ResASPP+RB	MSPIE			
—		√		2.57	47.48	33.54/0.974 0
—			√	2.99	58.35	33.95/0.975 8
√		√		3.66	57.32	33.80/0.974 8
√			√	4.09	68.18	34.02/0.975 5
	√		√	4.51	79.04	**34.17/0.976 0**

表 4 中的第 1 组和第 2 组实验为仅有主干分支的对比, 第 3 组、第 4 组和第 5 组实验为加上梯度分支之后的对比。第 1 组和第 2 组实验结果表明, 与 ResASPP+RB 结构相比, 本文方法的 MSPIE 模块在参数量和 Flops 增大较小的情况下, PSNR 和 SSIM 指标都有提升, 证明了特征提取部分中 MSPIE 模块的有效性, 通过利用光场图像的亚像素信息可以有效提升重建质量。其次验证梯度细节的作用, 第 3 组、第 4 组和第 5 组实验结果表明, 加入梯度细节能明显提高重建质量, 证明了在重建过程中融合梯度能有效保护光场图像的边缘纹理细节, 从而提升重建质量。此外, 第 4 组和第 5 组实验的梯度信息提取方式不同, 梯度分支的特征提取部分分别为 ResASPP+RB 结构和 MSPIE 模块, 结果表明利用 MSPIE 提取的多重亚像素梯度信息更有利于提高光场图像的重建质量。

4　结论

本文提出了一种基于亚像素信息和梯度引导的光场图像超分辨率方法, 通过提取 SAI 的亚像素信息对光场图像进行超分辨率, 有效解决了光场图像超分辨率中的像素差异性问题, 同时还利用梯度引导超分辨率过程增强了重建图像的边缘纹理细节。为了提取 SAI 的亚像素信息, 构建了 MSPIE 模块, 该模块首先将 SAI 分成水平、垂直、对角和反对角图像堆, 然后通过 3D 特征提取块提取图像堆的亚像素信息。为了丰富不同 SAI 之间的全局上下文信息, 引入 Transformer 进行特征表示增强。为了解决光场图像在重建过程中存在边缘细节丢失的问题, 利用梯度引导超分辨率过程, 消融实验表明, 梯度引导的光场超分辨率能有效保护图像的边缘细节。在公开数据库上的实验表明, 本文方法可以有效提高光场图像的空间分辨率, 并且复原光场图像的边缘纹理细节。

参考文献:

[1] JIA C, SHI F, ZHAO M. Object detection based on light field imaging[C]// Proceedings of 2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design, May 4-6, 2022, Hangzhou, China. New York: IEEE, 2022: 239-244.

[2] HAN L, SHI Z, ZHENG S N, et al. Light field depth estimation using RNN and CRF[C]// Proceedings of 2022 7th International Conference on Image, Vision and Com-

puting, July 26-28, 2022, Xi'an, China. New York: IEEE, 2022: 725-729.

[3] 宋力争, 林冬云, 彭侠夫, 等. 基于深度学习的 patch-match 双目三维重建[J]. 应用光学, 2022, 43(3): 436-443.
SONG Lizheng, LIN Dongyun, PENG Xiafu, et al. Patch-match binocular 3D reconstruction based on deep learning[J]. Journal of Applied Optics, 2022, 43(3): 436-443.

[4] XU Y F, ZHAO J K. Deep multi-levels edge-guided network for super-resolution[C]// Proceedings of 2022 19th International Computer Conference on Wavelet Active Media Technology and Information Processing, December 16-18, 2022, Chengdu, China. New York: IEEE, 2022: 1-5.

[5] LIU Y H, LI S M, LIU A Q. Two-way guided super-resolution reconstruction network based on gradient prior[C]// Proceedings of 2021 IEEE International Conference on Image Processing, September 19-22, 2021, Anchorage, AK, USA. New York: IEEE, 2021: 1819-1823.

[6] YOON Y, JEON H G, YOO D, et al. Light field image super-resolution using convolutional neural network[J]. IEEE Signal Processing Letters, 2017, 24(6): 848-852.

[7] WANG Y Q, WANG L G, YANG J G, et al. Spatial-angular interaction for light field image super-resolution[C]// Proceedings of the 16th European Conference on Computer Vision, August 23-28, 2020, Glasgow, United Kingdom. Berlin: Springer, 2020: 290-308.

[8] 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.

[9] WU G C, ZHAO M D, WANG L Y, et al. Light field reconstruction using deep convolutional network on EPI[C]// Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 1638-1646.

[10] 安平, 陈欣, 陈亦雷, 等. 基于视点图像与 EPI 特征融合的光场超分辨率[J]. 信号处理, 2022, 38(9): 1818-1830.
AN Ping, CHEN Xin, CHEN Yilei, et al. Light field super-resolution based on viewpoint image and EPI feature fusion[J]. Signal Processing, 2022, 38(9): 1818-1830.

[11] ZHANG S, LIN Y, SHENG H. Residual networks for light field image super-resolution[C]// Proceeding of 2019 IEEE Conference on Computer Vision and Pattern Recognition, June 16-21, 2019, Long Beach, CA, USA. New York: IEEE, 2019: 11038-11047.

[12] ZHANG S, CHANG S, LIN Y. End-to-end light field spatial super-resolution network using multiple epipolar geometry[J]. IEEE Transactions on Image Processing, 2021, 30: 5956-5968.

[13] JIN J, HOU J H, CHEN J, et al. Light field spatial super-resolution via deep combinatorial geometry embedding and structural consistency regularization[C]// Proceedings of 2020 IEEE Conference on Computer Vision and Pattern Recognition, June 13-19, 2020, Seattle, WA, USA. New York: IEEE, 2020: 2260-2269.

[14] WANG Y Q, YANG J G, WANG L G, et al. Light field image super-resolution using deformable convolution[J]. IEEE Transactions on Image Processing, 2021, 30: 1057-1071.

[15] WANG S Z, ZHOU T F, LU Y, et al. Detail-preserving transformer for light field image super-resolution[J]. AAAI Conference on Artificial Intelligence, 2022, 36(3): 2522-2530.

[16] CAO J Z, LI Y W, ZHANG K, et al. Video super-resolution transformer[EB/OL]. [2023-06-15]. https://arxiv.org/abs/2106.06847.

[17] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: transformers for image recognition at scale[EB/OL]. [2023-06-15]. https://www.cnblogs.com/lucifer1997/p/18210509.

[18] HUI Z, GAO X B, YANG Y C, et al. Lightweight image super-resolution with information multi-distillation network[C]// Proceedings of the 27th ACM International Conference on Multimedia, October, 2019, New York, NY, USA. New York: ACM, 2019: 2024-2032.

[19] RERABEK M, EBRAHIMI T. New light field image dataset[C]// Proceedings of 8th International Conference on Quality of Multimedia Experience. Lisbon: Portugal, 2016: 2409457.

[20] HONAUER K, JOHANNSEN O, KONDERMANN D, et al. A dataset and evaluation methodology for depth estimation on 4D light fields[C]// 13th Asian Conference on Computer Vision. Taiwan: Springer , 2016, 19-34.

[21] WANNER S, MEISTER S, GOLDLUECKE B. Datasets and benchmarks for densely sampled 4D light fields[J]. Vision, Modelling and Visualization, 2013, 13: 225-226.

[22] PENDU M L, JIANG X R, GUILLEMOT C. Light field in painting propagation via low rank matrix completion[J]. IEEE Transactions on Image Processing, 2018,

27(4): 1981-1993.

[23] VAISH V, ADAMS A. The (new) standford light field archive[EB/OL]. [2023-06-15]. http://lightfield.stanford.edu/.

[24] KIM J, LEE J K, LEE K M. Accurate image super-resolution using very deep convolutional networks[C] //IEEE Conference on Computer Vision and Pattern Recognition, June 26-July 1, 2016, Las Vegas, USA. New York: IEEE, 2016: 1646-1654.

[25] 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, July 21-26, 2017, Honolulu, USA. New York: IEEE, 2017: 1132-1140.