Opto-Electronic Engineering
光 电 工 程

Article

2020 年，第 47 卷，第 12 期

DOI: 10.12086/oee.2020.200007

融合多尺度特征的光场图像超分辨率方法

赵圆圆，施圣贤*

上海交通大学机械与动力工程学院，上海 200240

摘要：光场相机作为新一代的成像设备，能够同时捕获光线的空间位置和入射角度，然而其记录的光场存在空间分辨率和角度分辨率之间的制约关系，尤其子孔径图像有限的空间分辨率在一定程度上限制了光场相机的应用场景。因此本文提出了一种融合多尺度特征的光场图像超分辨网络，以获取更高空间分辨率的光场子孔径图像。该基于深度学习的网络框架分为三大模块：多尺度特征提取模块、全局特征融合模块和上采样模块。网络首先通过多尺度特征提取模块学习 4D 光场中固有的结构特征，然后采用融合模块对多尺度特征进行融合与增强，最后使用上采样模块实现对光场的超分辨率。在合成光场数据集和真实光场数据集上的实验结果表明，该方法在视觉评估和评价指标上均优于现有算法。另外本文将超分辨后的光场图像用于深度估计，实验结果展示出光场图像空间超分辨率能够增强深度估计结果的准确性。

关键词：超分辨；光场；深度学习；多尺度特征提取；特征融合

中图分类号：TP391.4　　**文献标志码：**A

引用格式：赵圆圆，施圣贤. 融合多尺度特征的光场图像超分辨率方法[J]. 光电工程，2020，**47**(12): 200007

Light-field image super-resolution based on multi-scale feature fusion

Zhao Yuanyuan, Shi Shengxian*

School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

Abstract: As a new generation of the imaging device, light-field camera can simultaneously capture the spatial position and incident angle of light rays. However, the recorded light-field has a trade-off between spatial resolution and angular resolution. Especially the application range of light-field cameras is restricted by the limited spatial resolution of sub-aperture images. Therefore, a light-field super-resolution neural network that fuses multi-scale features to obtain super-resolved light-field is proposed in this paper. The deep-learning-based network framework contains three major modules: multi-scale feature extraction, global feature fusion, and up-sampling. Firstly, inherent structural features in the 4D light-field are learned through the multi-scale feature extraction module, and then the fusion module is exploited for feature fusion and enhancement. Finally, the up-sampling module is used to achieve light-field super-resolution. The experimental results on the synthetic light-field dataset and real-world light-field dataset showed that this method outperforms other state-of-the-art methods in both visual and numerical evaluations. In addition, the super-resolved light-field images were applied to depth estimation in this paper, the results illustrated that the disparity map was enhanced through the light-field spatial super-resolution.

收稿日期：2020-01-03； 收到修改稿日期：2020-04-15

基金项目：国家自然科学基金资助项目(11772197)

作者简介：赵圆圆(1995-)，女，硕士研究生，主要从事计算机视觉、光场成像技术的研究。E-mail：ZhaoYuanyuan_236@163.com

通信作者：施圣贤(1980-)，男，博士，副教授，主要从事机器视觉、光场三维测试技术的研究。E-mail：kirinshi@sjtu.edu.cn

Keywords: super-resolution; light-field; deep learning; multi-scale feature extraction; feature fusion
Citation: Zhao Y Y, Shi S X. Light-field image super-resolution based on multi-scale feature fusion[J]. *Opto-Electronic Engineering*, 2020, **47**(12): 200007

1 引 言

光场成像的概念最早由 Lippmann[1]于 1908 年提出，经过较长一段时间的发展，Adelson 和 Wang[2]于 1992 年搭建了全光相机模型，随后 Ng 等人[3]于 2005 年设计出了手持式光场相机。作为新一代的成像设备，近年来光场相机已经被广泛应用到三维测试领域，如：三维流场测试[4-8]、三维火焰温度场重建[9]以及三维物体形貌重建[10-11]等。与传统相机不同，光场相机在主透镜与成像平面(CCD/CMOS)之间安装了一个微透镜阵列，可通过单次拍摄同时捕获空间中光线的空间位置和入射角度，因此能够从单张原始光场图像中还原出所拍摄场景的三维信息。然而由于光场相机的固有结构设计，其空间分辨率与角度分辨率之间存在一定的制约关系[3]。以商用光场相机 Lytro Illum 为例，其捕获的光场为 7728 pixels×5368 pixels，而经过光场渲染得到的 15×15 的子图像阵列中每张子孔径图像的分辨率仅为 625 pixels×434 pixels。过低的子图像空间分辨率导致光场深度估计算法得到的深度图分辨率过低，同时对深度估计结果的准确性造成一定的影响。因此，越来越多的学者投入到光场超分辨率研究中，以拓展光场相机的应用场景。

目前，主流的光场超分辨率主要分为空间超分辨率、角度超分辨率和时间超分辨率以及三者的任意组合。具体地，利用 4D 光场中的冗余信息并提出其所遵循的模型框架来实现超分辨率，这些光场超分辨率方法大致分为三大类：基于几何投影的方法、基于先验假设的优化方法和基于深度学习的方法[12]。基于几何投影的方法主要是根据光场相机的成像原理，通过获取不同视角子孔径图像之间的亚像素偏移来对目标视图进行超分辨。Lim 等人[13]通过分析得出，光场 2D 角度维度上的数据中暗含着不同视角图像在空间维度上的亚像素偏移信息，继而提出了利用数学模型将其投影至凸集上进行迭代优化来获取高分辨率图像的方法。Georgiev 等人[14]建立了专门针对聚焦型光场相机的超分辨框架，通过子图像中的对应点找出相邻视图之间的亚像素偏移，然后将相邻视图中的像素传播至目标视图中得到超分辨率结果。基于先验假设的方法是研究人员为了重建出更真实的高分辨率视图所提出的。这类方法在利用 4D 光场结构的同时加入了对实际拍摄场景的先验假设，由此提出相应的物理模型对光场超分辨率问题进行优化求解。Bishop 等人[15]在光场成像模型中加入了朗伯反射率和纹理保留的先验假设，并在变分贝叶斯框架中对光场图像进行超分辨，实验表明该算法在真实图像上有较好的表现。Rossi 等人[16]提出了一种利用不同光场视图信息并结合图正则化器来增强光场结构并最终得到高分辨率视图的方法。考虑实际光场图像中的噪声问题，Alain 和 Smolic[17]提出了一种结合 SR-BM3D[18]单图像超分辨滤波器和 LFBM5D[19]光场降噪滤波器的方法，通过在 LFBM5D[19]滤波步骤和反投影步骤之间反复交替以实现光场超分辨。基于深度学习的光场超分辨率近年来也在逐渐兴起。Yoon 等人[20]首次采用深度卷积神经网络对光场图像进行空间和角度超分辨。在他们的工作中，首先通过空间超分辨网络对每个子孔径图像进行上采样并结合 4D 光场结构对其增强细节，然后通过角度超分辨网络生成相邻视图之间新的视角图像。Wang 等人[21]将光场子图像阵列看作是 2D 图像序列，用双向递归卷积神经网络对光场中相邻视角图像之间的空间关系进行建模，并设计了一种隐式多尺度融合方案来进行超分辨重建。Zhang 等人[22]提出了一种使用残差卷积神经网络的光场图像超分辨方法(ResLF)，通过学习子图像阵列中水平、竖直和对角方向上的残差信息，并将其用于补充目标视图的高频信息，实验结果表明该方法在视觉和数值评估上均表现出优良的性能。

为了充分利用 4D 光场的冗余信息，需要结合光场中 2D 空间维度和 2D 角度维度上的数据来学习 4D 光场中的固有结构特征和丰富的纹理细节，以最终实现光场超分辨率。受基于深度学习的立体图像超分辨率网络框架 PASSRnet[23]的启发，本文提出了一种融合多尺度特征的光场超分辨率网络结构。该方法的核心思想是：在无遮挡情况下，光场 $N\times N$ 的子图像阵列中，中心视角图像中的像素点与其他周围视角图像中与之对应的像素点之间存在特定的变换关系。通过利用这一几何约束，某一视图的纹理细节特征可被来自其他视图的补充信息所增强。本文所提出的超分辨率网络框架中，首先通过原子空间金字塔池化(atrous spatial

pyramid pooling，ASPP)模块[24]来扩大感受野以学习到光场中 2D 空间维度上的多尺度特征，然后经由融合模块对所提取的特征进行融合并结合光场中 2D 角度维度上的几何约束进行全局特征增强，最后由上采样模块对光场图像进行空间超分辨。该网络通过对多尺度特征的融合与增强，能够累积到光场中丰富的纹理细节信息，在×2 超分辨率任务中，该方法在遮挡和边缘区域也能表现出良好的重建效果，平均信噪比(peak signal to noise ratio，PSNR)比现有方法提高了 0.48 dB。本文将超分辨后的光场图像用于深度估计，以探索光场空间超分辨率对深度估计结果的增强作用。

2 本文方法

2.1 算法模型与框架

在本文中，使用 G^{HR} 表示原始的高分辨率光场，G^{LR} 表示对应的经过下采样得到的低分辨率光场。由于高分辨率光场与对应的低分辨率光场之间保持一致性，因此 G^{LR} 可看作是 G^{HR} 经过光学模糊和下采样而得到的。考虑到上述过程中引入的噪声问题，可对 G^{LR} 和 G^{HR} 之间的一致性关系进行如下数学建模[25]：

$$G^{\mathrm{LR}} = \boldsymbol{S}\cdot\boldsymbol{B}\cdot G^{\mathrm{HR}} + N \quad , \tag{1}$$

式中：$\boldsymbol{B}$ 为模糊矩阵，$\boldsymbol{S}$ 表示下采样矩阵，N 代表过程中可能会引入的误差项。

光场超分辨率重建任务可看作是式(1)描述过程的逆过程，即对 G^{LR} 进行上采样并进一步去除模糊，从而得到超分辨率后的光场 G^{SR}。具体过程可被数学描述为

$$G^{\mathrm{SR}} = \boldsymbol{B}^{-1}\boldsymbol{S}^{-1}G^{\mathrm{LR}} = G^{\mathrm{HR}} + \boldsymbol{B}^{-1}\boldsymbol{S}^{-1}N \quad , \tag{2}$$

式中：$\boldsymbol{B}^{-1}$ 表示去模糊矩阵，$\boldsymbol{S}^{-1}$ 为上采样矩阵，$\boldsymbol{B}^{-1}\boldsymbol{S}^{-1}N$ 表示超分辨率后的光场 G^{SR} 与原始的高分辨率光场 G^{HR} 之间的误差。由上式可以看出，在超分辨率重建任务中，利用更多的纹理细节信息可以较大程度上重建出更接近原始数据的光场。由于真实图像中往往存在噪声，故在超分辨率重建过程中加入抗噪声模块，将会进一步提升超分辨率算法的性能。

本文所提出的算法框架将 G^{LR} 和 G^{HR} 作为网络的输入数据和真实数据，以训练得到上采样映射，从而输出光场超分辨率重建结果 G^{SR}。如图 1(a)所示，该网络结构分为三大模块：多尺度特征提取模块、全局特征融合模块和上采样模块。首先，多尺度特征提取模块分别对低分辨率光场 G^{LR} 中的每个视图进行特征提取，以得到 $N\times N$ 的特征图阵列；然后生成的特征图阵列经过堆叠后被发送至融合模块进行特征融合，同时该模块利用光场中角度维度上的约束对所提取的特征进行全局增强；而后获得的光场结构特征被发送至上采样模块以最终输出超分辨后的光场子图像阵列。每个模块的网络结构设计与作用将在下一节介绍。

2.1.1 多尺度特征提取

纹理信息对于大多数图像处理任务具有十分重要的意义。在超分辨率任务中，对高频信息的有效提取和利用决定了能否详实地重建出高分辨率图像中的细节部分。因此，本文采用 ASPP 块来扩大接收域并分别从每张光场子孔径图像中提取多尺度特征。如图 1(b)的示例，该 ASPP 块由膨胀率不同的原子空洞卷积组成。不同膨胀率的原子空洞卷积的感受野不同，因此 ASPP 块可以累积来自图像中不同区域的纹理细节信

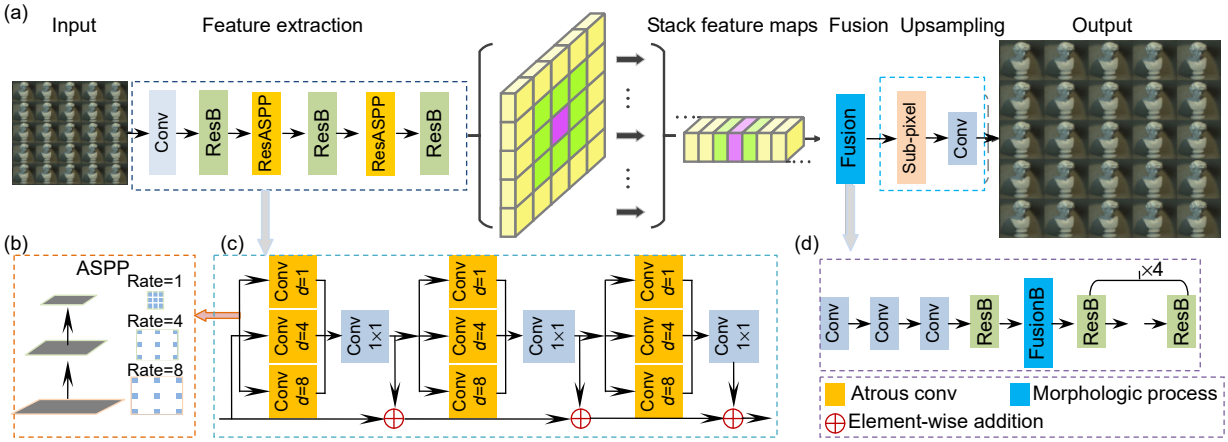


图 1 融合多尺度特征的光场超分辨率网络示意图。

(a) 网络结构；(b) ASPP 块；(c) ResASPP 块结构；(d) 融合模块结构

Fig. 1 Schematic diagram of light-field super-resolution based on multi-scale features. (a) Structure of the network; (b) ASPP block; (c) ResASPP block; (d) Structure of the fusion module

息。本文算法在 ASPP 块结构基础上加入了残差式的设计，组成了 ResASPP(residual atrous spatial pyramid pooling，ResASPP)块的网络子结构。如图 1(c)所示，将 3 个结构参数相同的 ASPP 块级联并以残差的形式加到上游输入中即为 1 个 ResASPP 块。在每个 ASPP 块中，首先 3 个原子空洞卷积分别以 d=1,4,8 的膨胀率对上游输入进行特征提取，然后再由 1 个 1×1 的卷积核对所得到的多尺度特征进行融合。整体的多尺度特征提取模块的操作流程为：低分辨率光场 G^{LR} 中的子孔径图像经过 1 个常规卷积和 1 个残差块(residual block，ResB)的处理提取出低级特征；接着，由交替出现两次的 ResASPP 块和残差块对低级特征进行多尺度特征提取及特征融合，从而得到每张子孔径图像的中级特征。如图 1(a)所示，多尺度特征提取模块分别对 $N\times N$ 的低分辨率子图像阵列中的每个视图进行操作，最终提取出与之相对应的 $N\times N$ 特征图阵列。在多尺度特征提取环节，网络主要对 4D 光场中 2D 空间维度上的信息加以利用并从中获取图像空间中的纹理细节特征。

2.1.2 全局特征融合

4D 光场具有冗余性，因此在提取到光场中空间维度上的特征后，需进一步利用角度线索来搜寻 4D 光场结构中的几何约束。如图 2(a)左侧的光场子图像阵列，在该 4D 光场结构中，不同视角图像中的对应匹配点之间遵循立体几何约束。在此，使用 $G_{s,t}(u,v)$ 表示 (s,t) 视角图像中位于 (u,v) 位置的像素点，该像素点的视差值记为 $d_{u/v}$。不考虑遮挡情况，空间中的某个物点，被记录在视图 (s,t) 中的 (u,v) 位置，同时也会被记录在视图 (s',t') 中的 (u',v') 像素位置。

$$\begin{aligned} G_{s,t}(u,v) &= G_{s',t'}(u+(s'-s)\cdot d_{u/v}, v+(t'-t)\cdot d_{u/v}) \\ &= G_{s',t'}(u',v') \quad , \end{aligned} \tag{3}$$

其中：视图 (s,t) 中的 (u,v) 像素点与视图 (s',t') 中的 (u',v') 像素点为立体视觉中的对应匹配点。

光场中每张低分辨率视图是从略微不同的角度来捕获场景，因此某一视图中未获取的纹理细节可能会被另一个视图捕获到。即一个视图的纹理细节特征可被来自其他视图的补充信息所增强。考虑到光场中每个视角之间的基线很小，中心视角图像可通过一定的“翘曲变换”(warping transformation)生成其他周围视角图像，反之亦然。中心视图生成周围视图的过程可被数学描述为

$$G_{s',t'} = \boldsymbol{W}_{st\to s't'}\cdot G_{s,t} + N_{st\to s't'} \quad , \tag{4}$$

式中：$G_{s,t}$ 表示中心视角图像，$G_{s',t'}$ 表示其他周围视角图像，$\boldsymbol{W}_{st\to s't'}$ 为“翘曲矩阵”(warping matrix)，而 $N_{st\to s't'}$ 为经“翘曲变换”后生成的视图与该视角原本的子孔径图像 $G_{s',t'}$ 之间的误差项。该误差项部分来源于遮挡问题，如 $G_{s',t'}$ 中某像素点由于在中心视角中位于被遮挡区域，故该像素点不能由 $G_{s,t}$ 生成。考虑遮挡情况下，式(4)描述的模型可改为

$$G_{s',t'} = \boldsymbol{M}_{st\to s't'}\cdot \boldsymbol{W}_{st\to s't'}\cdot G_{s,t} + N_{st\to s't'} \quad , \tag{5}$$

其中：$\boldsymbol{M}_{st\to s't'}$ 是一个“mask”矩阵，用于去除上述遮挡问题的影响。

基于上述的思想，本文设计了图 2(a)所示的融合块(fusion block，FusionB)以对所提取到的多尺度特征

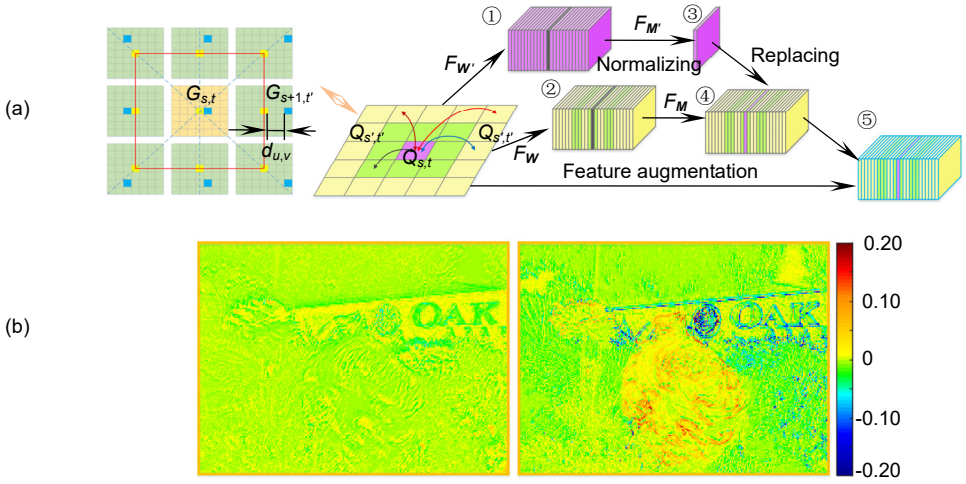


图 2 融合块 FusionB 原理示意图。(a) FusionB 结构；(b) 多尺度特征与经过 FusionB 融合后的特征对比

Fig. 2 Schematic diagram of fusion block. (a) Structure of FusionB; (b) Comparison of multi-scale features and the features fused by FusionB

进行特征融合与增强。如图 2(a)所示，$N\times N$ 特征图阵列中的周围视角特征 $Q_{s',t'}$ 经过“翘曲变换” $\boldsymbol{W}_{s't'\to st}$ 可分别生成中心视角特征 $Q'_{s,t}$，如标号为①的特征块所示。同样地，中心视角特征 $Q_{s,t}$ 经过“翘曲变换” $\boldsymbol{W}_{st\to s't'}$ 也可相应地生成周围视角特征 $Q_{s',t'}$，如图 2(a)中标号为②的特征块所示。前述过程可表示为

$$\begin{cases} Q'_{s,t}=F_{W'}(Q_{s't'})=\boldsymbol{W}_{s't'\to st}\otimes Q_{s't'} \\ Q'_{s't'}=F_{W}(Q_{s,t})=\boldsymbol{W}_{st\to s't'}\otimes Q_{s,t} \end{cases}, \tag{6}$$

式中：$\otimes$ 为分批矩阵乘法(batch-wise matrix multiplication)。继而，该模块对特征块①和②分别进行“mask”处理以去除遮挡问题的影响。获取“mask”矩阵的方法为：得出生成视图与原本视图之间的误差项的绝对值，绝对值越大，说明该区域为遮挡区域，具体地：

$$\boldsymbol{M}_{s't'\to s,t}(i,j)=\begin{cases} 0, & \text{if}(||Q'_{s,t}(i,j)-Q_{s,t}(i,j)||_1)>T \\ 1, & \text{otherwise} \end{cases}, \tag{7}$$

其中：$T=0.9\times\max(||Q'_{s,t}-Q_{s,t}||_1)$ 为算法中设置的经验阈值，“mask”矩阵 $\boldsymbol{M}_{st\to s't'}$ 的求法与 $\boldsymbol{M}_{s't'\to s,t}$ 类似。而后，对特征块①和②中的遮挡区域进行过滤：

$$\begin{cases} Q^M_{s,t}=F_{M'}(Q'_{s,t})=\boldsymbol{M}_{s't'\to s,t}\otimes Q'_{s,t} \\ Q^M_{s't'}=F_{M}(Q'_{s't'})=\boldsymbol{M}_{st\to s't'}\otimes Q'_{s't'} \end{cases}, \tag{8}$$

式中：$Q^M_{s,t}$ 和 $Q^M_{s't'}$ 分别为经过“mask”处理后得到的特征块。由于在上述过程中，共生成了 $n=N\times N-1$ 个中心视角特征图，故对 $Q^M_{s,t}$ 进行归一化处理，得到图 2(a)中所示的标号为③的特征图 $Q^{M_c}_{s,t}$：

$$Q^{M_c}_{s,t}=\frac{1}{n}\sum_{k\in[1,\,n]}Q^M_{s,t}(:,:,k)\ , \tag{9}$$

式中：k 为 $N\times N$ 特征图阵列中除中心视图外，其他视图按照从左上到右下的顺序排列时的索引值；$Q^M_{s,t}(:,:,k)$ 则表示由中心视图生成的经过“mask”处理后所得到的第 k 个其他周围视角特征图。进一步将 $Q^M_{s't'}$ 中心位置的特征图替换为标号③的特征图，即可得到经全局融合后的特征块④。该特征块④将累加至原先输入的多尺度特征上以实现特征增强，最终得到经过特征融合与增强的特征块⑤。图 2(b)中分别展示了任意选取的某一视角图像所对应的多尺度特征及其经过 FusionB 融合后所得到的特征。结果表明，融合块 FusionB 结合了 4D 光场中角度维度上的信息而使得输出的特征图中包含了更多的纹理细节信息，如图 2(b)右侧特征图中明显的边缘部分，这也展示出该融合块对输入特征起到了增强作用。

全局特征融合块主要利用 4D 光场结构来对所提取的多尺度特征进行融合与增强。如图 1(a)所示，多尺度特征图阵列 $Q_0\in R^{NH\times NW\times C}$ 中的每个视图按照从左上到右下的顺序在通道 C 上进行堆叠，从而得到特征图 $Q\in R^{H\times W\times(N\times N\times C)}$。堆叠后的特征图 $Q\in R^{H\times W\times(N\times N\times C)}$ 将作为输入被发送至全局特征融合模块。全局特征融合模块的结构如图 1(d)所示，堆叠后的多尺度特征首先经由 3 个常规卷积进行特征再提取，接着经由 1 个残差块进行特征融合，然后进入融合块以实现特征增强。该融合块 FusionB 通过提取 4D 光场中的角度特征，可以在原有特征上累加更多的纹理细节信息。继而，经过增强的特征将被送至 4 个级联的残差块进行特征充分融合，并最终生成可用于光场图像超分辨率重建的 4D 光场结构特征。

2.1.3 上采样模块及损失函数

在特征提取和融合模块完成对 4D 光场结构特征的学习后，上采样模块将对获取的特征图进行超分辨率重建。该模块采用了超分辨率网络常用的上采样方法—子像素卷积(sub-pixel convolution)，或被称为像素洗牌操作(pixel shuffle)[26]。子像素卷积模块首先从输入的通道数为 C 的特征图中产生 r^2 个通道数为 C 的特征图，然后对得到的通道数为 $r^2\times C$ 的特征图进行抽样操作，并由此生成分辨率为 r 倍的高分辨率特征图[26]。该高分辨率特征图被发送至 1 个常规的卷积层中进行特征融合并最终生成超分辨后的光场子图像阵列。在训练过程中，超分辨后的光场子孔径图像分别与实际的高分辨率光场子孔径图像进行一一对比，本文的损失函数使用 L1 范数(如式(10)所示)，因为在神经网络训练过程中表现出了更好的性能。另外，网络采用泄露因子为 0.1 的带泄露修正线性单元(leaky ReLU)作为激活函数以避免训练过程中神经元不再进行信息传播的情况。

$$L_{\text{loss}}=\sum_{s,t}\sum_{u,v}(||G^{\text{SR}}(u,v,s,t)-G^{\text{HR}}(u,v,s,t)||_1)\ 。 \tag{10}$$

2.2 算法性能评价指标

本文选用图像超分辨率重建领域常用的 PSNR 和结构相似性(structural similarity，SSIM)评价指标对算法性能进行评价。对于超分辨率重建后得到的光场 G^{SR} 和真实光场数据 G^{HR} 可计算出光场中每张子孔径图像对应的 PSNR 值(用 R_{PSNR} 表示，单位 dB)：

$$R_{\text{PSNR}}(s,t)=10\cdot\log_{10}\left(\frac{(2^n-1)^2}{e_{\text{MSE}}(s,t)}\right)\ , \tag{11}$$

式中：e_{MSE} 为超分辨率重建得到的图像与原始的高分辨率图像之间的均方误差，求解过程参见徐亮等人[27]

在相关工作中的介绍，n 为图像的位深度。本文中，训练数据为 RGB 三通道图像。因此在做结构相似性评价时，首先将光场子孔径图像转换至 YCbCr 颜色空间，然后提取出 Y 通道图像。通过对比真实数据的 Y 通道图像 $G_{\mathrm{Y}}^{\mathrm{HR}}$ 与超分辨后的 Y 通道图像 $G_{\mathrm{Y}}^{\mathrm{SR}}$ 来计算 SSIM(用 M_{SSIM} 表示)值：

$$M_{\mathrm{SSIM}}(s,t)=F_{\mathrm{SSIM}}(G_{\mathrm{Y}}^{\mathrm{SR}}(s,t),\ G_{\mathrm{Y}}^{\mathrm{HR}}(s,t))\ ,\qquad (12)$$

式中：F_{SSIM} 表示计算图像之间结构相似性的函数，具体见徐亮等人[27]在其工作中介绍的 SSIM 求解过程。

3 实验结果与分析

实验使用了来自 HCI1[28]和 HCI2[29]的 4D 合成光场图像以及由 Lytro Illum 光场相机拍摄的分别来自 Stanford[30]和 EPFL[31]的真实图像。从 Stanford 和 EPFL 数据集中分别随机取出约 5/6 的光场数据与 HCI2 数据集组合作为训练集，并把 Stanford 和 EPFL 数据集中剩下的光场数据与 HCI1 中的光场数据组合作为测试集。本实验中训练集和测试集分别有 419 个和 91 个光场图像。所有的实验 LF 图像均按照 5×5 的角度分辨率进行预处理，然后使用双三次插值对高分辨率光场 G^{HR} 进行空间×2 降采样以获得低分辨率光场 G^{LR}，再使用本文方法对 G^{LR} 进行超分辨处理以得到超分辨率重建结果 G^{SR}。超分辨率重建结果的质量由 PSNR 和 SSIM 来进行定量评估。在实验中，将本文方法与×2 单张图像超分辨方法 FALSR[32]和传统光场超分辨方法 GBSR 以及基于深度学习的光场图像超分辨方法 ResLF 进行对比，用于展示所提方法的性能与潜力。

在训练过程中，G^{LR} 中的低分辨率(low resolution，LR)子孔径图像被以 32 pixels 的步长裁剪成了空间大小为 64 pixels×64 pixels 的小块，G^{HR} 中的高分辨率(high resolution，HR)子孔径图像也对应地被裁剪成大小为 128 pixels×128 pixels 的小块，由此构成网络的输入数据和真实数据。在实验中，通过水平和垂直地随机翻转图像来进行数据增强。本文搭建的神经网络在 Nvidia GTX 1070 GPU 的 PC 上基于 Pytorch 框架进行训练，模型使用 Adam 优化方法[33]并且使用 Xaviers 方法[34]初始化每一层卷积层的权重。模型的初始学习率设置为 2×10^{-4}，每 20 个周期衰减 0.5 倍，经过 80 个周期后停止训练，整个训练过程大约需要 2 天左右的时间。在测试过程中，分别将模型应用到合成数据集和真实数据集上以评估本文所采用的超分辨网络的性能，进一步地将超分辨率光场 G^{SR} 中的超分辨率(super resolution，SR)子孔径图像阵列应用到深度估计上以观察光场空间超分辨率对视差计算的影响。

3.1 合成光场图像

在训练过程中使用了 HCI2 合成数据集中的光场图像，因此使用 HCI1 中的合成光场图像对各超分辨率方法进行性能测试。实验中基于深度学习的超分辨率算法 FALSR 和 ResLF 采用的是作者发布的预先训练好的模型。另外，双三次插值图像超分辨率方法在实验中被当作基准算法。图 3 展示了各算法对场景 Buddha、Mona 和 Papillon 的超分辨率重建结果。实验结果表明，双三次插值重建出的图像整体上比较模糊，这是由于该方法主要利用了图像中的低频信息而忽略了对高频信息的有效利用。而基于光场几何约束的传统方法 GBSR 能够较为真实地重建出超分辨率图

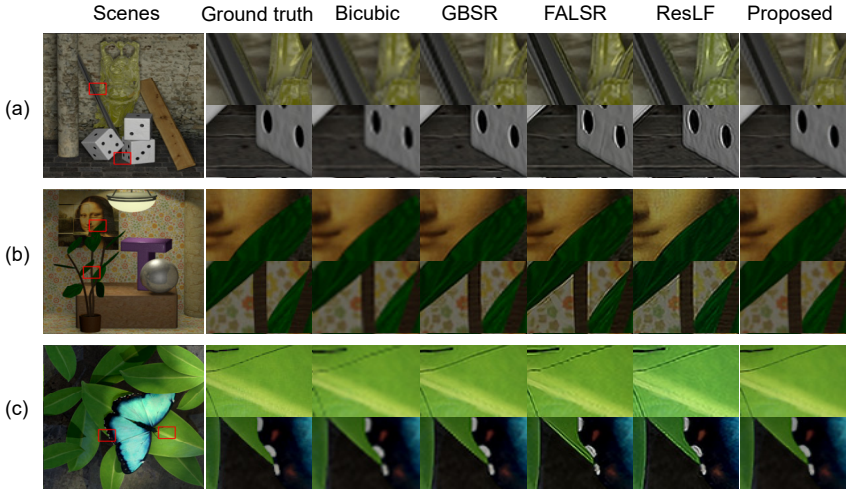


图 3 合成数据光场超分辨结果。(a) Buddha 场景；(b) Mona 场景；(c) Papillon 场景

Fig. 3 Light-field super resolution results on synthetic data. (a) Buddha; (b) Mona; (c) Papillon

像，整体上表现出了不错的性能，然而该方法重建场景中的边缘部分会出现模糊或过度锐化问题。另外，GBSR 算法完成 1 个场景的光场超分辨率重建大概需要 2 h~3 h，十分耗时。基于深度学习的×2 单张图像超分辨方法 FALSR 对于场景中同一物体内部区域的重建效果较好，但由于仅利用单张视图而没有考虑 4D 光场结构中的隐含线索，因此无法重建复杂的纹理，同时该方法存在较大程度的锐化过度问题。而基于深度学习的光场超分辨方法 ResLF 通过结合 EPI 空间中的极线约束可较为真实地重建出图像中的纹理细节，但由于没有用到光场中的全部视角图像从而导致对遮挡边缘部分的重建结果有些失真。本文提出的超分辨率网络通过利用光场中的所有视角图像，能够更为真实地重建出高分辨率图像中的纹理信息，同时全局特征融合模块一定程度上改善了边缘模糊失真和锐化过度的情况，在主观视觉上表现出了更好的重建性能。定量的性能评估结果如表 1 所示，蓝色字体标注了除本文方法外的评估指标最高的算法，红色字体则标注了本文方法优于蓝色字体所标注方法的场景。由表 1 看出 ResLF 重建出的图像保持着较高的结构相似度，GBSR 在合成图像重建上整体获得了次佳的分数，而本文方法在 PSNR 和 SSIM 上均优于其他方法。

3.2 真实光场图像

超分辨率算法常被用于真实图像任务，本文进一步地使用 Lytro Illum 相机拍摄的真实光场数据来测试各超分辨率算法的性能。真实图像往往存在许多实际问题，特别是 Lytro 拍摄的光场图像存在较多噪点，这对光场超分辨率重建以及视差计算造成了一定的困难。通常 HR 光场子图像中 1 个像素位置的噪点经过下采样-上采样过程之后在 SR 图像中将会呈现出 2×2 像素区域大小的噪点。噪点在图像中随机离散地分布，较多的噪点导致真实图像超分辨率结果的 PSNR 值与合成图像相比整体偏低。

图 4 展示了真实光场数据的超分辨率重建结果。其中，FALSR 由于没有利用来自其他视角图像中的冗余信息而导致重建效果不佳，甚至重建图像中物体的边缘可能会存在较大程度的扭曲变形，如图 4(a)中重建的栅栏的边缘部分。ResLF 通过利用多个方向的 EPI 信息能够较为详实地还原图像中复杂的纹理细节，特别是对于图像空间中方向为水平、竖直和对角的纹理。但该方法超分辨率重建图像中的边缘部分仍会存在过度平滑和模糊的现象，如图 4(b)中重建出的车牌号码中的字母“A”。本文所提方法能够较好地重建出各个方向的纹理信息，包括圆滑的边缘信息，整体上表现出了较高的光场超分辨率重建性能，如图 4(c)中的校徽以及花瓣的边缘。

定量评估结果如表 2 所示，蓝色字体标注了除本文方法外的评价指标最高的算法，红色字体标注了本文方法优于蓝色字体所标注方法的场景。由表 2，本文方法重建出的 Fence 场景的 PSNR 低于 ResLF 和 FALSR。这是由于该场景的原始图像中存在较多噪点，而本文网络在设计中没有特别考虑降噪问题，同时融合模块过多地累加了噪点的多尺度特征。另外，本文方法在 Fence 场景下的 SSIM 略低于 ResLF，这是因为 ResLF 对水平、竖直和对角的纹理有较强的超分辨率重建能力，而 Fence 场景中存在较多的对角纹理。在 Cars 和 Flowers 场景中，本文方法在 PSNR 和 SSIM 上的表现均优于其他方法。将本文网络用于测试集中的 Stanford 真实光场图像的超分辨率重建上，得到的平均 PSNR/SSIM 为 38.30 dB/0.9778，比 ResLF 文献中在 Stanford 数据集上得到的 PSNR/SSIM(35.48 dB/0.9727)值略高，且比 FALSR 文献中在公开数据集 Set5[35]上×2 超分辨所得的 PSNR(37.82 dB)值也略高。综合地看，本文所提出的超分辨网络在主观视觉和评价指标上处于相对领先的水平。

表 1 不同超分辨算法在合成数据上的性能比较

Table 1 Performance comparison of different image super resolution algorithms on synthetic data

Method	Buddha		Mona		Papillon	
	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM
Bicubic	33.0865	0.9208	32.6579	0.9301	33.4031	0.9365
GBSR	35.7463	0.9568	38.1479	0.9769	38.7855	0.9802
FALSR	34.9493	0.9373	34.8104	0.9412	34.7569	0.9504
ResLF	35.4988	0.9689	34.3314	0.9614	35.1983	0.9754
Proposed	39.8095	0.9807	41.5483	0.9865	41.0616	0.9852

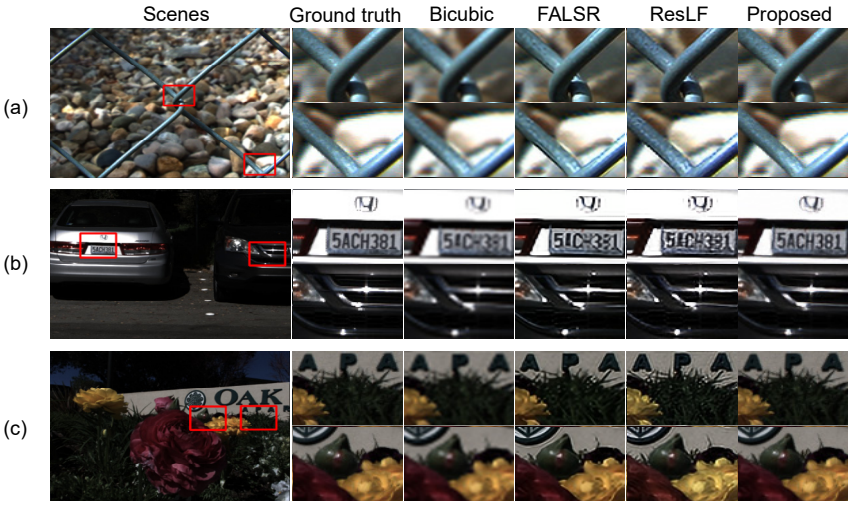


图 4 真实数据光场超分辨结果。(a) Fence 场景；(b) Cars 场景；(c) Flowers 场景

Fig. 4 Light-field super resolution results on real-world data. (a) Fence; (b) Cars; (c) Flowers

表 2 不同超分辨算法在真实数据上的性能对比

Table 2 Performance comparison of different image super resolution algorithms on real-world data

Method	Fence		Cars		Flowers	
	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM
Bicubic	30.8720	0.9541	31.3657	0.9401	30.2619	0.9194
FALSR	35.1476	0.9639	31.5821	0.9422	31.5795	0.9192
ResLF	34.9172	0.9844	31.6191	0.9722	31.3748	0.9538
Proposed	31.5522	0.9816	35.2929	0.9800	40.6967	0.9874

3.3 应用于深度估计

为了观察光场空间超分辨率对视差计算结果的影响，本节分别对经下采样得到的低分辨率光场 G^{LR} 和经本文方法得到的超分辨率光场 G^{SR} 进行了深度估计。深度估计算法统一采用 POBR[36]，图 5 中分别展示了场景真实的视差图(Ground truth)、由 G^{LR} 计算得到的低分辨率视差图(LR depth)和由 G^{SR} 计算得到的高分辨率视差图(SR depth)。值得一提的是，SR depth 的分辨率为 LR depth 的 2 倍，因此对光场进行空间超分辨率可进一步获得高分辨率的深度图。另一方面，深度估计结果表明，高分辨率的光场子图像阵列中包含更为丰富的纹理信息，尤其能为遮挡或边缘区域提供更多的线索，因此可以更加准确地还原出所拍摄场景的深度信息，如图 5 中黑色方框标记的部分。

为了更直观地展示深度估计结果的优劣，本文将 LR depth 和 SR depth 分别与 Ground truth 对比，得到图 6 所示的误差图。在实验中，SR depth 与 Ground truth 直接相减求绝对值以得到误差图。而由于 LR depth 和 Ground truth 的分辨率不同无法直接做差，因此先采用双三次插值对 Ground truth 进行下采样，然后再将经下采样得到的视差图与 LR depth 对比来得到 LR depth 的误差图。双三次插值的本质是对图像进行平滑滤波，这会使得 Ground truth 下采样后的视差数据值较小幅度地偏离原始数据，并且使得视差数据的极大值变小。因此 LR depth 的误差图与 SR depth 的误差图相比，其中绝大部分像素位置的数值会偏大，而在纹理边缘所对应的像素位置的数值会偏小。故图 6 所展示的误差图对比是一种略失公允的对比，但对比结果依然能够说明一定的问题。如图 6 中红色方框标记区域的视差计算误差，超分辨后的光场深度估计结果优于低分辨率光场的深度估计结果，这与在图 5 上的直观视觉对比结果相一致。另外，本文分别对两个场景的平均视差误差 LR depth error 和 SR depth error 进行了计算。其中，Mona 场景中低分辨率光场视差的平均计算误差为 0.1699 pixels，高分辨率光场视差的平均计算误差为 0.0075 pixels。而 Flower 场景中低分辨率光场视差的平均计算误差为 0.2013 pixels，高分辨率光场视差的平均计算误差为 0.0434 pixels。

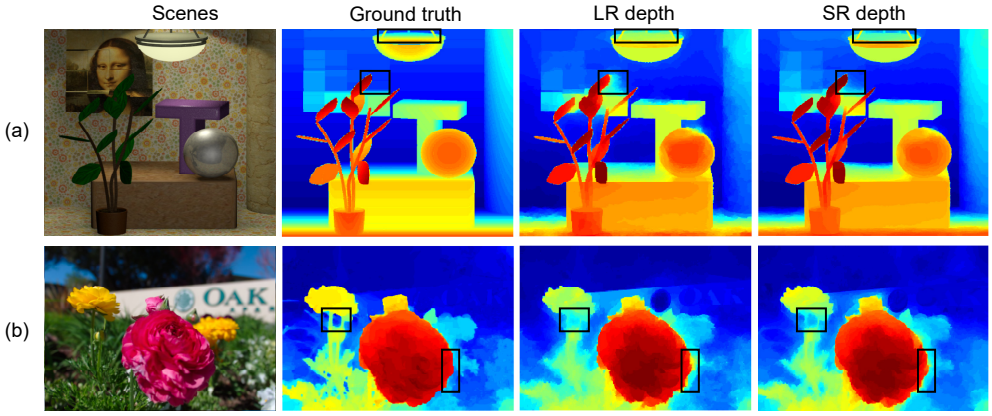


图 5 视差估计结果。(a) Mona 场景视差图；(b) Flowers 场景视差图

Fig. 5 Depth estimation results. (a) Disparity map of Mona; (b) Disparity map of Flowers

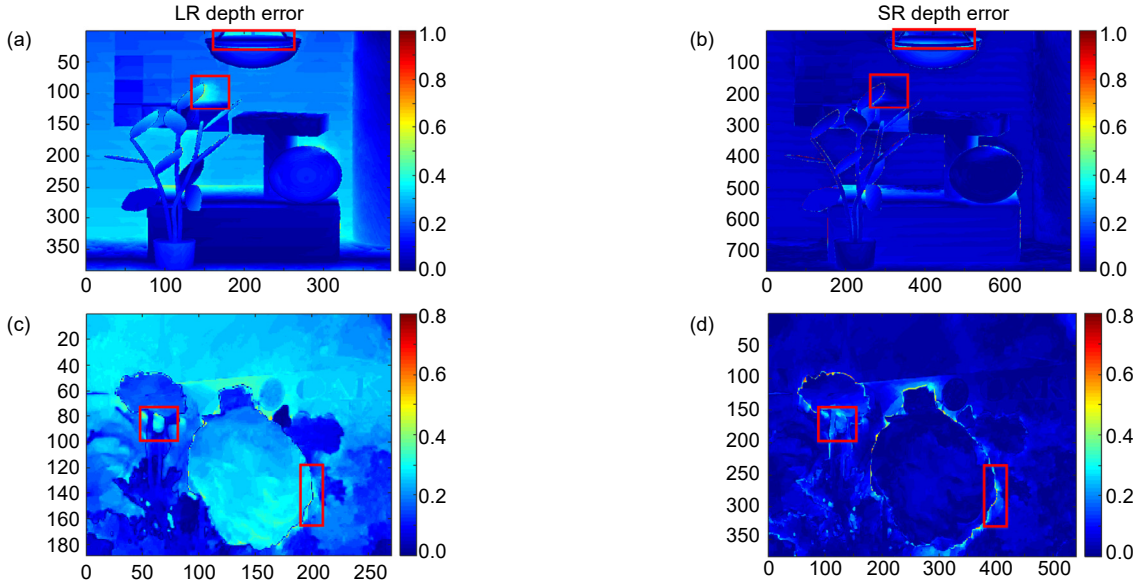


图 6 深度估计结果与真实深度之间的误差图(单位：像素)。(a)，(b) Mona 场景误差图；(c)，(d) Flowers 场景误差图

Fig. 6 Error maps between depth estimation results and the ground truth (unit: pixel). (a), (b) Mona's error map; (c), (d) Flowers's error map

4 结 论

本文提出了一种融合多尺度特征的光场超分辨网络以提高光场子孔径图像的空间分辨率。在所提的网络框架中，通过多尺度特征提取模块探索 4D 光场中的固有结构信息，然后采用融合模块对提取到的纹理信息进行融合和增强，最后使用上采样模块实现光场子图像阵列的超分辨率。实验结果表明，该方法在合成光场数据集和真实光场数据集上均表现出了较好的性能，×2 超分辨率重建情况下，平均 PSNR 比单图超分辨方法 FALSR 高 0.48 dB，平均 SSIM 比光场超分辨率方法 ResLF 评价指标高 0.51%。另外，该方法在主观视觉上也表现出了良好的超分辨率重建性能。该方法不仅能够重建图像空间中水平、竖直和对角方向的纹理，同时还可用于其他各个方向的复杂纹理重建。进一步地，本文将超分辨结果用于光场深度估计，发现其能够为遮挡或边缘区域提供更多的线索，实验结果展示出光场图像空间超分辨率在一定程度上增强了视差计算结果的准确性。

参考文献

[1] Lippmann G. Épreuves réversibles donnant la sensation du relief[J]. *Journal de Physique Théorique et Appliquée*, 1908, **7**(1): 821–825.

[2] Adelson E H, Wang J Y A. Single lens stereo with a plenoptic camera[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1992, **14**(2): 99–106.

[3] Ng R, Levoy M, Brédif M, *et al*. Light field photography with a

hand-held plenoptic camera[R]. Stanford Tech Report CTSR 2005-02, 2005.

[4] Tan Z P, Johnson K, Clifford C, *et al*. Development of a modular, high-speed plenoptic-camera for 3D flow-measurement[J]. *Optics Express*, 2019, **27**(9): 13400–13415.

[5] Fahringer T W, Lynch K P, Thurow B S. Volumetric particle image velocimetry with a single plenoptic camera[J]. *Measurement Science and Technology*, 2015, **26**(11): 115201.

[6] Shi S X, Ding J F, New T H, *et al*. Volumetric calibration enhancements for single-camera light-field PIV[J]. *Experiments in Fluids*, 2019, **60**(1): 21.

[7] Shi S X, Ding J F, New T H, *et al*. Light-field camera-based 3D volumetric particle image velocimetry with dense ray tracing reconstruction technique[J]. *Experiments in Fluids*, 2017, **58**(7): 78.

[8] Shi S X, Wang J H, Ding J F, *et al*. Parametric study on light field volumetric particle image velocimetry[J]. *Flow Measurement and Instrumentation*, 2016, **49**: 70–88.

[9] Sun J, Xu C L, Zhang B, *et al*. Three-dimensional temperature field measurement of flame using a single light field camera[J]. *Optics Express*, 2016, **24**(2): 1118–1132.

[10] Shi S X, Xu S M, Zhao Z, *et al*. 3D surface pressure measurement with single light-field camera and pressure-sensitive paint[J]. *Experiments in Fluids*, 2018, **59**(5): 79.

[11] Ding J F, Li H T, Ma H X, *et al*. A novel light field imaging based 3D geometry measurement technique for turbomachinery blades[J]. *Measurement Science and Technology*, 2019, **30**(11): 115901.

[12] Cheng Z, Xiong Z W, Chen C, *et al*. Light field super-resolution: a benchmark[C]//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, Long Beach, CA, 2019.

[13] Lim J, Ok H, Park B, *et al*. Improving the spatail resolution based on 4D light field data[C]//*Proceedings of the 16th IEEE International Conference on Image Processing*, Cairo, Egypt, 2009, **2**: 1173–1176.

[14] Georgiev T, Chunev G, Lumsdaine A. Superresolution with the focused plenoptic camera[J]. *Proceedings of SPIE*, 2011, **7873**: 78730X.

[15] Bishop T E, Favaro P. The light field camera: extended depth of field, aliasing, and superresolution[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, **34**(5): 972–986.

[16] Rossi M, Frossard P. Graph-based light field super-resolution[C]//*Proceedings of the IEEE 19th International Workshop on Multimedia Signal Processing*, Luton, UK, 2017: 1–6.

[17] Alain M, Smolic A. Light field super-resolution via LFBM5D sparse coding[C]//*Proceedings of the 25th IEEE International Conference on Image Processing*, Athens, Greece, 2018: 1–5.

[18] Egiazarian K, Katkovnik V. Single image super-resolution via BM3D sparse coding[C]//*Proceedings of the 23rd European Signal Processing Conference*, Nice, France, 2015: 2849–2853.

[19] Alain M, Smolic A. Light field denoising by sparse 5D transform domain collaborative filtering[C]//*Proceedings of the IEEE 19th International Workshop on Multimedia Signal Processing*, Luton, UK, 2017: 1–6.

[20] Yoon Y, Jeon H G, Yoo D, *et al*. Learning a deep convolutional network for light-field image super-resolution[C]//*Proceedings of 2015 IEEE International Conference on Computer Vision Workshop*, Santiago, Chile, 2015: 57–65.

[21] 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.

[22] Zhang S, Lin Y F, Sheng H. Residual networks for light field image super-resolution[C]//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, USA, 2019: 11046–11055.

[23] Wang L G, Wang Y Q, Liang Z F, *et al*. Learning parallax attention for stereo image super-resolution[C]//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, USA, 2019: 12250–12259.

[24] Chen L C, Zhu Y K, Papandreou G, *et al*. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//*Proceedings of the European Conference on Computer Vision*, Glasgow, United Kingdom, 2018: 801–818.

[25] Wang R G, Liu L L, Yang J, *et al*. Image super-resolution based on clustering and collaborative representation[J]. *Opto-Electronic Engineering*, 2018, **45**(4): 170537.
汪荣贵, 刘雷雷, 杨娟, 等. 基于聚类和协同表示的超分辨率重建[J]. 光电工程, 2018, **45**(4): 170537.

[26] Shi W Z, Caballero J, Huszár F, *et al*. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, USA, 2016: 1874–1883.

[27] Xu L, Fu R D, Jin W, *et al*. Image super-resolution reconstruction based on multi-scale feature loss function[J]. *Opto-Electronic Engineering*, 2019, **46**(11): 180419.
徐亮, 符冉迪, 金炜, 等. 基于多尺度特征损失函数的图像超分辨率重建[J]. 光电工程, 2019, **46**(11): 180419.

[28] Wanner S, Meister S, Goldluecke B. Datasets and benchmarks for densely sampled 4D light fields[M]//Bronstein M, Favre J, Hormann K. *Vision*, *Modeling* & *Visualization*, Lugano, Switzerland: The Eurographics Association, 2013: 225–226.

[29] Honauer K, Johannsen O, Kondermann D, *et al*. A dataset and evaluation methodology for depth estimation on 4D light fields[C]//*Proceedings of the Asian Conference on Computer Vision*, Taipei, Taiwan, China, 2016: 19–34.

[30] Raj S A, Lowney M, Shah R, *et al*. Stanford lytro light field archive[EB/OL]. http://lightfields.stanford.edu/LF2016.html. 2016.

[31] Rerabek M, Ebrahimi T. New light field image dataset[C]//*Proceedings of the 8th International Conference on Quality of Multimedia Experience*, Lisbon, Portugal, 2016.

[32] Chu X X, Zhang B, Ma H L, *et al*. Fast, accurate and lightweight super-resolution with neural architecture search[Z]. arXiv: 1901.07261, 2019.

[33] Kingma D P, Ba L J. Adam: a method for stochastic optimization[C]//*Proceedings of the International Conference on Learning Representations*, San Diego, America, 2015.

[34] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks[C]//*Proceedings of the 13th International Conference on Artificial Intelligence and Statistics*, Sardinia, Italy, 2010: 249–256.

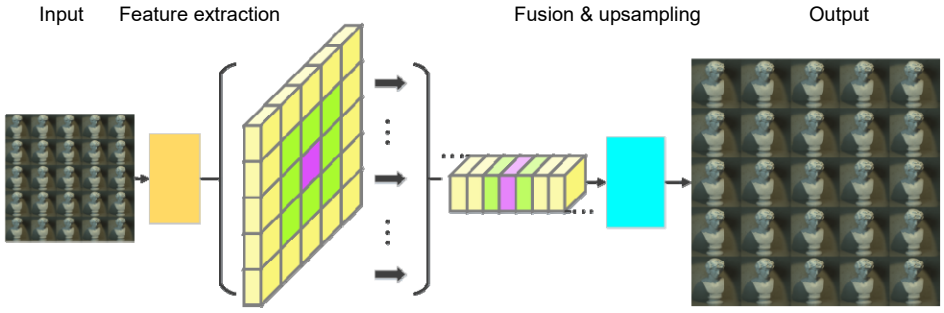
[35] Bevilacqua M, Roumy A, Guillemot C, *et al*. Low-complexity single-image super-resolution based on nonnegative neighbor embedding[C]//*British Machine Vision Conference*, Guildford, UK, 2012.

[36] Chen J, Hou J H, Ni Y, *et al*. Accurate light field depth estimation with superpixel regularization over partially occluded regions[J]. *IEEE Transactions on Image Processing*, 2018, **27**(10): 4889–4900.

Light-field image super-resolution based on multi-scale feature fusion

Zhao Yuanyuan, Shi Shengxian*

School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China



Structure of light-field image super resolution network

Overview: As a new generation of imaging equipment, a light-field camera can simultaneously capture the spatial position and incident angle of light rays. However, the recorded light-field has a trade-off between spatial resolution and angular resolution. Especially the limited spatial resolution of sub-aperture images limits the application scenarios of light-field cameras. Therefore, a light-field super-resolution network that fuses multi-scale features to obtain super-resolved light-field is proposed in this paper. The deep-learning-based network framework contains three major modules: multi-scale feature extraction module, global feature fusion module, and up-sampling module. The design ideas of different modules are as follows.

a) Multi-scale feature extraction module: To explore the complex texture information in the 4D light-field space, the feature extraction module uses ResASPP blocks to expand the perception field and to extract multi-scale features. The low-resolution light-field sub-aperture images are first sent to a Conv block and a Res block for low level feature extraction, and then a ResASPP block and a Res block are alternated twice to learn multi-scale features that accumulate high-frequency information in the 4D light-field.

b) Global feature fusion module: The light-field images contain not only spatial information but also angular information, which implies inherent structures of 4D light-field. The global feature fusion module is proposed to geometrically reconstruct the super-resolved light-field by exploiting the angular clues. It should be noted that the feature maps of all the sub-images from the upstream are first stacked in the channel dimension of the network and then are sent to this module for high-level features extraction.

c) Up-sampling module: After learning the global features in the 4D light-field structure, the high-level feature maps could be sent to the up-sampling module for light-field super resolution. This module uses sub-pixel convolution or pixel shuffle operation to obtain 2 spatial super-resolution, after feature maps are sent to a conventional convolution layer to perform feature fusion and finally output a super-resolved light-field sub-images array.

The network proposed in this paper was applied to the synthetic light-field dataset and the real-world light-field dataset for light-field images super-resolution. The experimental results on the synthetic light-field dataset and real-world light-field dataset showed that this method outperforms other state-of-the-art methods in both visual and numerical evaluations. In addition, the super-resolved light-field images were applied to depth estimation, and the results illustrated the parallax calculation enhancement of light-field spatial super-resolution, especially in occlusion and edge regions.

Citation: Zhao Y Y, Shi S X, *et al*. Light-field image super-resolution based on multi-scale feature fusion[J]. *Opto-Electronic Engineering*, 2020, **47**(12): 200007

Supported by National Natural Science Foundation of China (11772197)

* E-mail: kirinshi@sjtu.edu.cn