

光学学报

基于神经辐射场的光场角度域超分辨

苗源, 刘畅, 邱钧*

北京信息科技大学应用数学研究所, 北京 100101

摘要 由于数据通量限制, 光场数据采集存在空间-角度分辨率折中, 无法达到诸多应用场景对精细三维重建的要求。本文使用双平面参数化光场数据训练神经辐射场, 由神经辐射场生成光场视点平面上新的视点图像, 进而实现光场角度域超分辨。由于神经辐射场能够对光场场景进行隐式表达, 为高分辨率的四维光场拟合出准确的隐式函数, 并将输入变量通过位置编码映射到该变量的傅里叶特征。因此该方法可以准确表达具有复杂不利条件的光场场景, 并且能够有效地解决场景高频纹理信息较难拟合的问题。本文通过超分辨光场的角度域信息突破空间-角度分辨率折中、带宽积的限制, 在实验中将角度分辨率从 5×5 提升到 9×9 , 并且峰值信噪比 (PSNR) 平均提升了 13.8%, 结构相似度 (SSIM) 平均提升了 9.19%。该研究结果可为后续的光场计算成像工作提供参考。

关键词 成像系统; 光场; 神经辐射场; 角度域超分辨; 体渲染

中图分类号 TP391 文献标志码 A

DOI: 10.3788/AOS230549

1 引言

光场成像^[1]是一种新型的成像技术, 它相对于传统的二维成像技术而言, 不仅可以记录位置信息, 还可以捕捉光线的方向信息^[2], 为计算机视觉和计算成像提供了新的研究方法和信息支持。光场成像已成为成像领域的一个研究热点, 并被广泛用于计算机视觉的深度估计^[3]和目标检测识别^[4]等领域。

由于光场数据采集系统的带宽积受到限制, 光场数据的空间分辨率和角度分辨率之间存在相互限制^[5]。例如在基于微透镜的光场相机中, 通过在主镜头和图像传感器之间插入一个微透镜阵列来采集四维光场。然而传感器的分辨率有限, 光场相机以牺牲角度分辨率为代价记录光的方向信息^[6]。其中角度域超分辨是突破光场采集系统带宽积限制的重要手段。光场角度域超分辨的目的是从稀疏视图集中重建密集采样的光场, 即沿角度维的超分辨光场。光场角度域超分辨技术能够避免进行直接密集采样所需的大量成本和工作量。同时, 通过从稀疏视图集中重建密集采样的光场, 还可以避免图像采集时可能带来的误差和噪声, 从而提高了重建的精度和质量。在三维重建和虚拟现实应用中, 也可以通过光场角度域超分辨技术实时生成高质量的虚拟场景^[7]。而目前光场角度域超分辨受到许多不利条件的干扰, 如遮挡、深度变化和背景干扰

等, 并且对于纹理丰富细节处也较难表达, 因此很难对上述情况的光场信息进行高质量的角度域超分辨。本文提出用神经辐射场刻画表示双平面光场, 实现光场角度域超分辨, 并解决目前光场角度域超分辨面临的问题, 为后续的光场计算成像工作提供支撑并带来新的思路和方法。

2 相关工作

2.1 光场角度域超分辨相关工作

光场角度域超分辨旨在从稀疏的视图集中重建一个密集采样的光场, 即沿角度维的超分辨光场。光场角度域超分辨技术能够在光场场景同一平面的新视点生成高质量的图像, 实现高质量视点渲染。由于密集采样的光场可以提供平滑的视差转移和自然重新聚焦的细节, 并可以应用到三维重建和虚拟现实等相关应用中^[8], 因此光场角度域超分辨近年来已被广泛研究。

光场角度域超分辨技术根据对几何信息使用程度可分为两类: 基于场景几何的新视点生成和基于隐式场景几何的新视点生成。基于场景几何的新视点生成包括以下方法: Wanner 等^[9]提出利用光场核极平面图像 (EPI) 的结构来计算连续视差图, 再使用视差图和变分法来计算新视点图像; Georgiev 等^[10]提出新的相机设计, 得到具有更高空间分辨率的光场数据, 通过光流法对视点进行插值以得到新视点图像补偿光场数据

收稿日期: 2023-02-14; 修回日期: 2023-03-14; 录用日期: 2023-03-24; 网络首发日期: 2023-05-08

基金项目: 国家自然科学基金 (62171044, 61871042, 61931003)、北京市自然科学基金 (432004)、北京信息科技大学“勤信人才”培育计划 (QXTCPB202105)

通信作者: *qiu jun@bistu.edu.cn

中较低的角度分辨率;Penner等^[11]在新视点生成过程中,建模了一个可以生成连续新视点图像的场景几何模型,并增强了深度重建的鲁棒性;Zhou等^[12]提出了一个可以快速合成场景的新视点图像的学习框架,该框架利用了多平面图像(MPIs)的分层表示方式;Mildenhall等^[13]提出局部光场融合(LLFF),通过离散化将稀疏光场的每个视点分解成多平面图,从中合成局部光场以实现宽视场,该方法的优点在于可以通过少量的输入视图即可进行高保真度的自然场景视图合成。

隐式场景几何的新视点生成不需要显式的几何模型,相反它们需要图像之间的特征对应,比如Seitz等^[14]基于点对应生成沿着两个原始相机中心线之间的新视点图像;光场渲染^[15]通过对预先获取的一组样本进行适当的滤波和插值,生成场景的新视点图像。Yoon等^[16]提出的光场卷积神经网络(LFCNN)同时适用于空间和角度的超分辨。他们首先对每个输入的子孔径图像(SAIs)进行了单独的超解,然后根据它们的相邻视图依次合成新的视图。Wang等^[17]提出了一个伪4D卷积神经网络(CNN)来恢复角度域超分辨的EPI容积的高频细节。Yeung等^[18]提出了一个通过空间-角度交替卷积的从粗到细的光场重建网络。Wu等^[19]提出了基于CNN的EPI光场重建(LFEPICNN),是一种基于CNN的角度域超分辨方法,用于从稀疏视图集重建光场。该方法能够更好地处理高频噪声和失真,提高图像的清晰度和细节;具有较强的鲁棒性和通用性,可以处理各种不同类型的图像。Xie等^[20]提出了采用生成对抗网络重建的超分辨方法。然而,这些技术均存在一些限制,如光场场景中可能会出现遮挡、深度变化和背景干扰等问题,而且在处理纹理丰富、细节复杂的区域时也存在困难。因此,要对出现上述情况的光场进行高质量的角度域超分辨处理较为困难。

2.2 神经辐射场相关工作

因神经渲染领域的快速发展,神经辐射场的概念被提出。这类方法可以在各种应用中使用神经网络作为隐函数来表示三维模型,包括图像生成、视点生成和重新照明^[21]。以前的研究需要利用已知的三维形状当作监督信息,但是在实际情况中,三维形状的真值并不总是能轻易获取。因此,接下来的研究中人们尝试仅使用图像作为直接输入,不需要已知的几何形状。为此,可微的绘制方法被研究人员提出,即在神经网络中增加一个绘制步骤,直接从图像绘制误差中学习神经网络。然而,这些方法的绘制结果过于平滑,只能处理几何结构不太复杂的简单形状。

在此背景下,Mildenhall等^[22]提出了新的表示方法,即神经辐射场(NeRF),它同样只使用输入图像作为可监督的信息,并能将精确的隐函数与高分辨率的几何形状相适应,从而使复杂的场景能够得到高质量高分辨率的视点合成结果。NeRF使用多层感知器来

表示一个描述三维模型的几何和颜色信息的五维向量函数。输入是三维空间位置和二维观察方向,输出是该位置的体积密度和该位置相应方向的颜色矢量。在训练期间,通过使用梯度下降法使预测图像和真实图像之间的误差最小化来拟合上述函数。因为体绘制过程是可微的,因此可将其加入到上述的神经网络训练中,以实现只以图像的颜色值作为监督的学习过程。此外,NeRF对输入变量进行预编码以避免丢失合成新视点图像的高频信息^[23]。实验数据表明,高频信息的复杂匹配问题可以通过这种映射来解决。

除了NeRF方法本身的分析和优化外,还有一些研究将其扩展到更多样化和更复杂的应用场景中。在文献[24]中,NeRF算法被扩展到使用网上人们拍摄的同场景的多视点照片来合成自由视点,并提出了非限制性拍摄条件下的神经辐射场(NeRF-W)算法。它是基于以下原则:由于一天中的时间、天气和光线等条件变化很大,而且不同的照片中经常出现人群等前景物体,原始的NeRF无法准确训练出这种场景的神经辐射场。这种情况可能会导致即使照片来自相同的位置和角度,也会产生不同的图像。考虑到这一特点,NeRF-W通过对不同输入图像的照明条件进行解码和编码的方式准确地学习场景的形状^[25]。为了应对临时性物体的存在,如在不同图像之间变化的前景,NeRF-W用所有输入图像共有的元素和依赖于每个输入图像的元素对场景进行建模。这使得该模型能够有效地区分场景中静态的内容和临时性的对象。然而,当场景中包含动态物体,依赖于多视图的几何一致性会被破坏,因此NeRF不能表示包含动态物体的场景。因此,扩展NeRF以表示动态场景也是一个重要的研究领域。在文献[26]中,有研究人员提出了可变形神经辐射场(D-NeRF)来表示含有动态物体的场景。该方法对所有输入图像的可见外观进行编码,以补偿图像之间的可见度差异,如曝光和白平衡。通过描述场景的变形,该方法可以准确地实现包含动态物体场景的视点合成。

基于NeRF来对光场场景进行隐式表达,并且利用位置编码将输入变量映射到其傅里叶特征,便可实现由高分辨率的四维光场拟合出准确的隐函数。不仅能处理如遮挡、深度变化和背景干扰等问题,并且对于纹理丰富细节处也能进行高质量的还原。

3 基于NeRF的光场角度域超分辨方法

3.1 利用光场数据训练NeRF

辐射场表示在空间中每个点发出的辐射,而光场表示沿着一条射线移动的总综合辐射。由此利用光场数据训练出的NeRF包含了全部的场景信息,甚至包含了未知视点的信息,由此提升场景表示性能。假设本文在正面的场景上工作,并且辐射沿着光线保持不

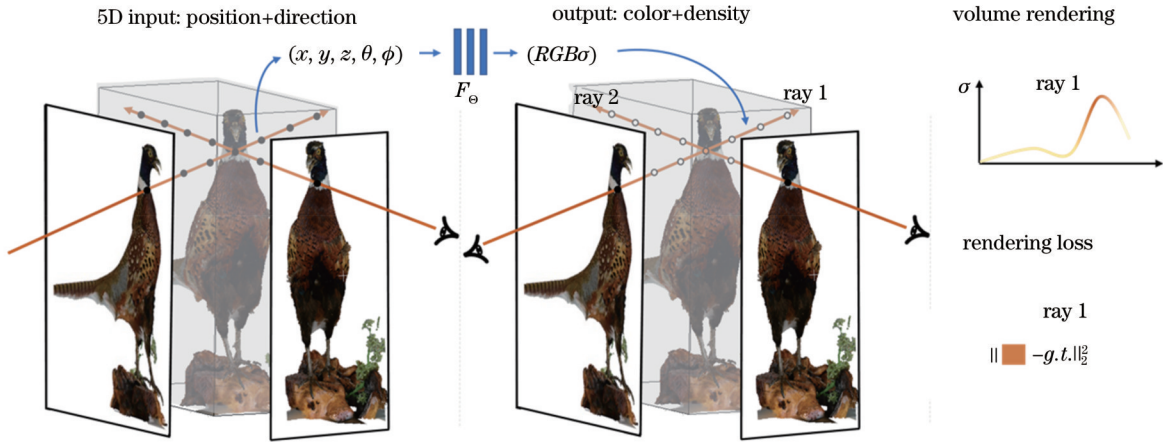


图 1 利用光场数据训练 NeRF 的概述。沿着相机射线对光场 5D 坐标(位置和观察方向)进行采样,将这些位置送入 MLP 以产生颜色和体积密度,并使用体积渲染技术将这些值合成对应的图像。利用梯度下降法拟合渲染损失来优化场景表示
 Fig. 1 Overview of training NeRF using light field data. By sampling light field 5D coordinates (position and viewing directions) along camera rays, feeding these positions into MLP to produce color and volume densities, and using volume rendering techniques to synthesize these values into corresponding image. Rendering losses are fitted using gradient descent to optimize scene representation

变,可以通过构建一个双平面光场来表示场景,取这些光线分别在平面上的局部交点坐标 (u, v) 、 (x, y) 将四维光场参数化:

$$L: r = (u, v, x, y) \rightarrow (c, \sigma). \quad (1)$$

如图 1 所示,将光场参数化后便可利用光场数据训练 NeRF。NeRF 使用多层感知器(MLP)来表达五维的向量函数^[22],该函数描述了三维模型的几何形状和颜色信息。输入为三维的空间位置向量 $x = (x, y, z)$ 和二维的观察方向向量 $d = (\theta, \varphi)$,输出则为该视点位置的体密度 σ 以及从该视点沿 d 方向发射出的颜色向量 $c = (r, g, b)$ 。利用 MLP 来表达这个连续的五维函数:

$$F_{\theta} = (x, d) \rightarrow (c, \sigma). \quad (2)$$

通过体绘制的方式由 σ 和 c 对图像的颜色进行预测。对于图像中任意点的像素,沿着相机光心 o 穿过该像素发出光线,该光线计为 $r = o + td$,近边界和原边界分别为 t_n 和 t_t ,则该像素的颜色值可由如下积分获得:

$$C(r) = \int_{t_n}^{t_t} T(t) \sigma(o + td) c(o + td, d) dt, \quad (3)$$

其中 $T(t)$ 为

$$T(t) = \exp \left[- \int_{t_n}^t \sigma(o + sd) ds \right]. \quad (4)$$

函数 $T(t)$ 表示沿射线从 t_n 到 t 的累积透过率,即射线从 t_n 传播到 t 并且没有击中任何其他粒子的概率。可以通过计算相机画面中每个像素对应的光线积分颜色,继而合成整张视图。

为了有效拟合光场场景中的高频信息,采用位置编码对输入变量映射到其傅里叶特征,大大改善了渲

染视图中的精细细节重建。在最初的实施中,以下的位置编码 γ 被应用于场景坐标 x (归一化为 $[-1, 1]$) 和观察方向单位矢量 d 的每个分量中。

$$\gamma(v) = \left[\sin(2^0 \pi v), \cos(2^0 \pi v), \dots, \sin(2^{N-1} \pi v), \cos(2^{N-1} \pi v) \right], \quad (5)$$

式中, N 是用户确定的编码维度参数,在 NeRF 原始报道中 x 的维度参数为 $N=10$, d 的维度参数为 $N=4$ 。

3.2 用 NeRF 实现光场角度域超分辨

光场角度域超分辨,也称为光场重建或视图合成,目的是从稀疏视图集重建密集采样的光场,即沿角域的超分辨率光场^[27]。本文提出一种基于 NeRF 的光场角度域超分辨方法,通过 NeRF 来表示光场,编辑光场中相机姿态获得子孔径图像间的新视角,完成角度维度的密集采样。该方法主要优势是基于 NeRF 来对光场场景进行隐式表达,为高分辨率的四维光场拟合出准确的隐式函数,准确表达具有复杂不利条件的光场场景,能够解决无法超分辨出光场场景有遮挡及深度具有变化的问题;并且将输入变量通过位置编码映射到该变量的傅里叶特征,能够有效地解决光场场景高频纹理信息较难拟合的问题。

本文使用 Colmap^[28] 来生成一组 $n*n$ 的光场子孔径图像的相机姿态,相机姿态可以被表示为一个相机到世界的变换矩阵:

$$T = [R|t], \quad (6)$$

式中, R 和 t 分别代表相机的旋转和平移,将姿态矩阵和原始图像输入利用 NeRF 表示的光场中,通过使用梯度下降法,最小化真实图像和预测图像间的误差对函数进行拟合,即

$$\min_{\sigma, c} \frac{1}{n} \sum_{i=1}^n \|I_i - I_i^p(c, \sigma)\|, \quad (7)$$

式中： I_i^p 为训练后得到的第*i*个预测图像； I_i 为第*i*个真实图像。

如图 2 所示,根据光场数据训练完网络之后,可以通过编辑变换矩阵 T 来得到想要的渲染姿态,进而生成 $(n+m)*(n+m)$ 的子孔径图像,以完成光场角度域的超分辨。

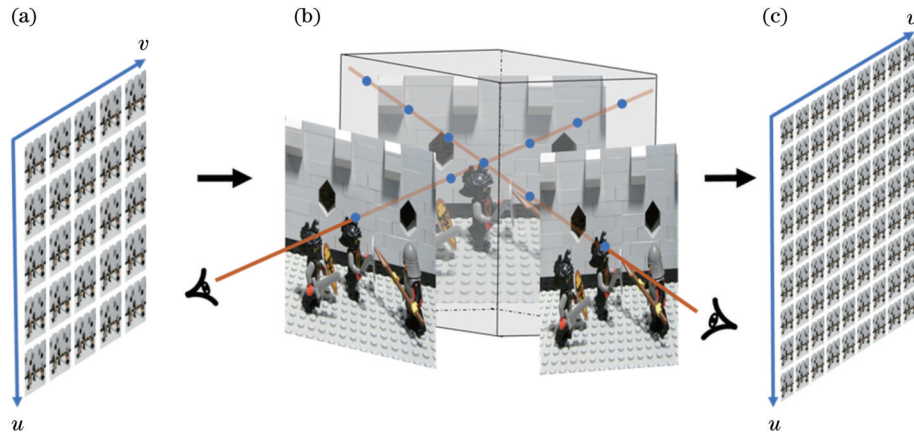


图 2 基于 NeRF 对光场进行角度域超分辨的过程。(a)输入 5*5 子孔径图像;(b)利用 NeRF 来表达光场;(c)输出 9*9 子孔径图像
Fig. 2 Process of NeRF based light field super-resolution in angular domain. (a) Input 5*5 sub-aperture image; (b) expression of optical field using NeRF; (c) output 9*9 sub-aperture image

本文的训练网络结构如图 3 所示,辐射场函数由 MLP 表达。MLP 将 x (三维坐标点和视图方向) 作为

输入,计算在任意位置和观察方向下 σ 和 红绿蓝 (RGB) 值。图中 ReLU 和 Sigmoid 为激活函数。

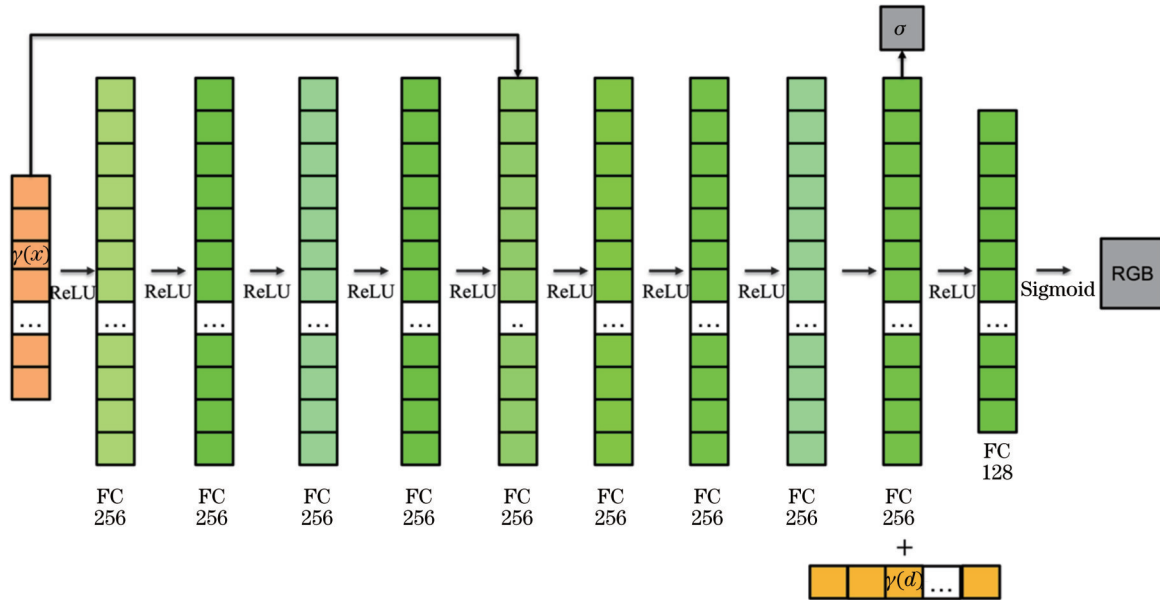


图 3 NeRF 训练网络结构图。MLP 将 x (三维坐标点和视图方向) 作为输入通过前 8 个全连接层 (FC), 之后特征向量将与输入观察方向的位置编码 $\gamma(d)$ 相连接, 输出 σ 和 RGB 值
Fig. 3 Diagram of NeRF training network structure. MLP takes x (3D coordinate points and view directions) as input through first eight fully connected layers (FC), after which feature vectors will be connected to position encoding $\gamma(d)$ of input observation directions, with output σ and RGB values

在场景训练完成后,通过合成未参与训练的测试集图片来评估算法合成的视图质量。本文选择使用峰值信噪比 (PSNR)^[29]、结构相似度 (SSIM)^[30] 和学习感知图像块相似度 (LPIPS)^[31] 三项得分来评估该算法合成的目标视图和真实原图的差异。这些指标试图在有

(全参考) 或无 (无参考) 地面真实图像的情况下评估单个图像的质量。其中, PSNR 是衡量图像质量的常用指标之一, 它主要用于衡量图像的噪声水平和失真程度。PSNR 的值越高, 表示失真越小, 图像质量越好。SSIM 是一种结构相似性指标, 用于比较两张图像之

间的相似性。SSIM 的值越接近 1, 表示两张图像越相似, 图像质量越好。LPIPS 是一种基于深度学习的图像相似性度量方法, 它通过学习感知性损失函数来衡量图像之间的相似性。LPIPS 的值越小, 表示图像之间的相似性越高, 图像质量越好。

4 数据实验

4.1 模拟数据实验

本文首先选择了模拟光场数据进行光场角度域超分辨实验。数据来源于 HCI 光场数据集^[32], 其中包含超过 1000 个高分辨率和高质量的光场图像, 包括静态和动态场景, 如室内和室外环境、自然景观、人体和物体等, 并且还包含了一些特殊场景, 例如微缩环境、照明变化和材质变化等。通过该数据集分别获得正向场景 antinous 和 boardgames 的 5*5 张子孔径图像, 图像分辨率为 512 pixel×512 pixel。通过本文的方法分别在

行和列角度域超分辨后生成 9*9 张图像。为验证本文所提方法的准确性, 将本文方法与其他角度域超分辨方法 LLFF^[13]和 LFEPICNN^[19]进行了对比。

可以发现在图 4 放大图中, LFEPICNN 对于子孔径图像雕像颈部的丰富纹理处还原效果较好, 但是不同深度的背景墙面处纹理效果较差; LLFF 在生成的新视点图像中前景后景交界处有明显的模糊; 而本文方法不论是对于具有深度的墙面背景处纹理还是雕像表面复杂纹理都能够较好地还原。

如图 5 所示, 在 boardgames 场景的放大图中可以发现, LLFF 在生成的新视点子孔径图像中深度发生变化处有大量的噪点, 这表明场景的深度并没有得到有效的还原; 而 LFEPICNN 虽然对于子孔径图像整体的深度处还原效果较好, 但是在棋盘表面格子处纹理效果较差; 本文方法同样能够较好地还原这两处的纹理。

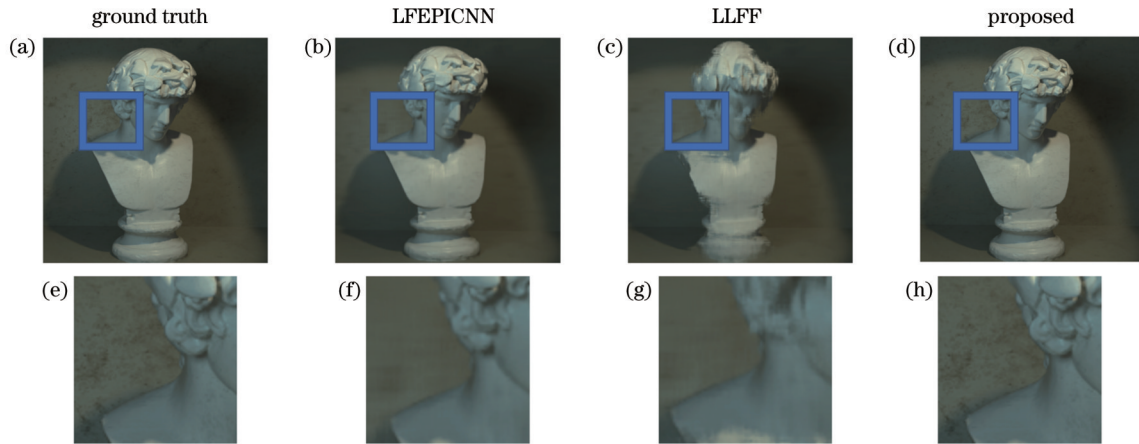


图 4 不同方法对 HCI antinous 场景角度域超分辨生成的新视角。(a) 真实场景; (b) LFEPICNN 方法; (c) LLFF 方法; (d) 提出方法; (e)、(f)、(g)、(h) 分别为 antinous 场景使用以上方法后的局部放大图

Fig. 4 New perspective for angular super-resolution generation of HCI antinous scenes using different methods. (a) Ground truth; (b) LFEPICNN method; (c) LLFF method; (d) proposed method; (e), (f), (g), (h) local zooms of antinous scene after using above methods, respectively

为了验证该方法的效果, 本文还对角度域超分辨前后的 EPI 进行了对比。当固定一个角度域坐标和一个空间域坐标时, 也可以产生 4D 光场的一个 2D 切片, 称之为 EPI。EPI 中同时包含了空间信息和角度信息, 每个物点在 EPI 中对应一条直线, 其斜率与深度相关, EPI 中不同斜率的直线代表不同深度的物点, 可以通过对比角度域超分辨前后相同位置的 EPI 比较判断超分辨效果。

如图 6 所示, 由于该场景在高度为 200 pixel 时深度变化较为明显, 因此固定 $u=3, y=200$ 绘制 EPI。图 6(d) 中角度域超分辨后的 EPI 呈不同深度斜率的直线, 再次证明场景的深度得到还原。图 6(b) 和 6(c) 角度域超分辨后对应两张完整 EPI 的 PSNR 值为 44.3, 这说明原位置的 EPI 和 ground truth 的 EPI 高度重合。而在图 6(d) 中可以发现, 场景中同一物点对应斜率在

角度域超分辨后的 EPI 中明显具有更高的清晰度和分辨率, 证明角度域超分辨后重建出的新视点位置没有噪声和失真情况, 表明场景的高频信息与纹理细节得到还原。表 1 中的相关评估指标也验证了该结论, 其中 PSNR 平均提升了 16.2%, SSIIM 平均提升了 13.7%, LPIPS 平均降低了 67.5%。

4.2 实际数据实验

为了验证本文提出方法在实际场景中的表现, 本文随后选择了斯坦福实际光场数据进行实验。斯坦福光场数据集^[33] (Stanford Lytro Light Field Archive) 包含来自商用的 Lytro 相机和由该大学开发的实验性 Illum 相机的 364 个光场图像样本。这些图像代表了各种场景和主题, 如自然风景、建筑物、室内家居等。本文从斯坦福光场数据集分别获得正向场景 knights 和 bunny 的 5*5 张子孔径图像, 图像分辨率为

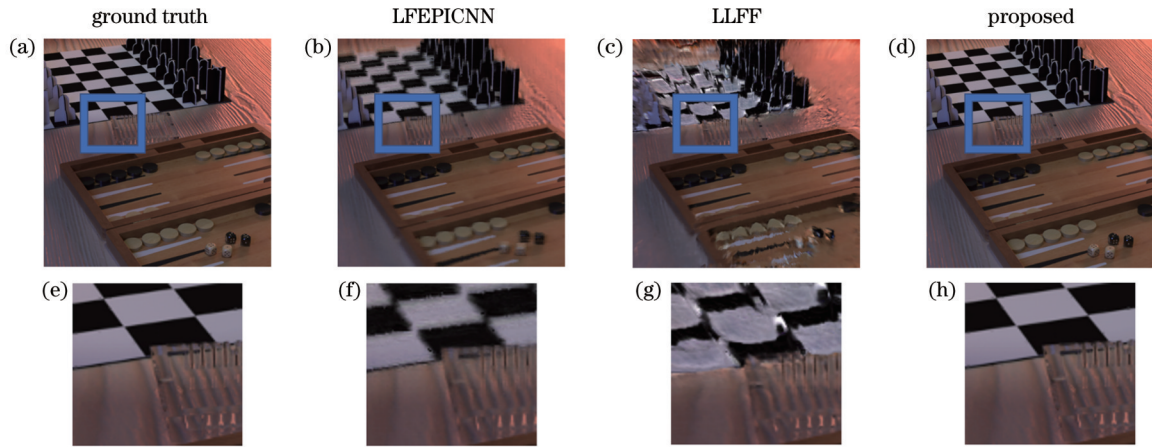


图 5 不同方法对 HCI boardgames 场景角度域超分辨生成的新视角。(a)真实场景;(b)LFEPICNN方法;(c)LLFF方法;(d)提出方法;(e)、(f)、(g)、(h)分别为 boardgames 场景使用以上方法后的局部放大图

Fig. 5 New perspective for angular super-resolution generation of HCI boardgames scenes using different methods. (a) Ground truth; (b) LFEPICNN method; (c) LLFF method; (d) proposed method; (e), (f), (g), (h) local zooms of boardgames scene after using above methods, respectively

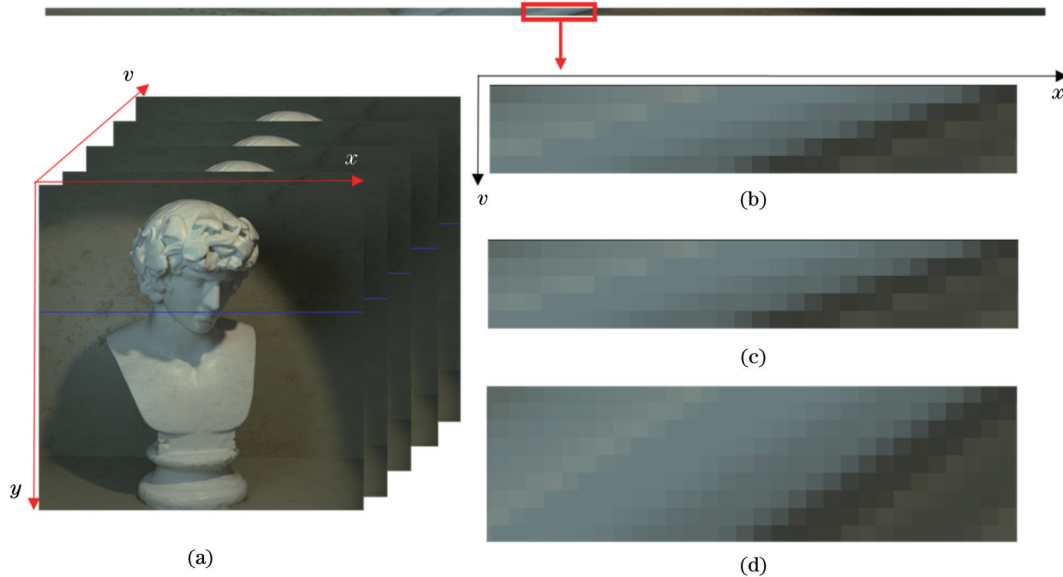


图 6 HCI boardgames 场景角度域超分辨前后的 EPI 局部图对比。(a)固定 $u=3, y=200$ 绘制的 EPI;(b)原始数据 EPI 局部图;(c)角度域超分辨后与原始数据对应位置的 EPI 局部图;(d)角度域超分辨后的 EPI 局部图

Fig. 6 Comparison of partial EPI before and after angular super-resolution of HCI boardgames scene. (a) Fix $u=3, y=200$ to draw EPI; (b) partial EPI of original data; (c) partial EPI of position corresponding to original data after angular super-resolution; (d) partial EPI after angular super-resolution

表 1 HCI 模拟光场数据集的指标评估

Table 1 Metrics evaluation on HCI simulated light field dataset

	Antinous			Boardgames		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
LFEPICNN	36.25	0.8542	0.094	38.75	0.9023	0.068
LLFF	33.86	0.7830	0.028	35.33	0.8372	0.046
Proposed	43.23	0.9936	0.011	42.15	0.9829	0.013

512 pixel×512 pixel,通过本文方法分别在行和列角度域超分辨后生成 9×9 张图像。同样将本文方法与其他角度域超分辨方法 LLFF 和 LFEPICNN 进行了对比。

在实际场景中,本文选择了 knights(图 7)和 bunny(图 8)场景,在 knights 场景的士兵头盔附近放大,可以发现 LLFF 在背后墙面空洞处附近有较多的噪声,该方法同样未能成功还原出场景细节中的深度;在 LFEPICNN 中虽然能够较好地还原深度,但在具有丰富纹理的头盔表面效果仍旧较差,这说明 LFEPICNN 并不能对场景中的高频信息进行处理;在 bunny 场景中,LFEPICNN 和 LLFF 在兔子的腿部处纹理细节表现都较为模糊;而本文方法不仅能够较好地还原深度,而且能对高频信息进行较好的处理。

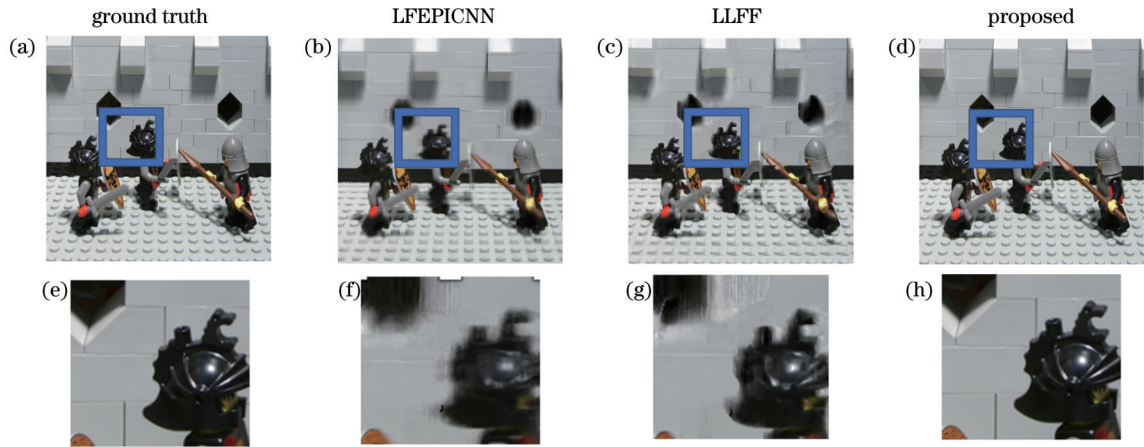


图 7 不同方法对斯坦福 knights 场景角度域超分辨生成的新视角。(a)真实场景;(b)LFEPICNN 方法;(c)LLFF 方法;(d)提出方法;(e)、(f)、(g)、(h)分别为 knights 场景使用以上方法后的局部放大图

Fig. 7 New perspective for angular super-resolution generation of Stanford knights scenes. (a) Ground truth; (b) LFEPICNN method; (c) LLFF method; (d) proposed method; (e), (f), (g), (h) local zooms of knights scene after using above methods, respectively

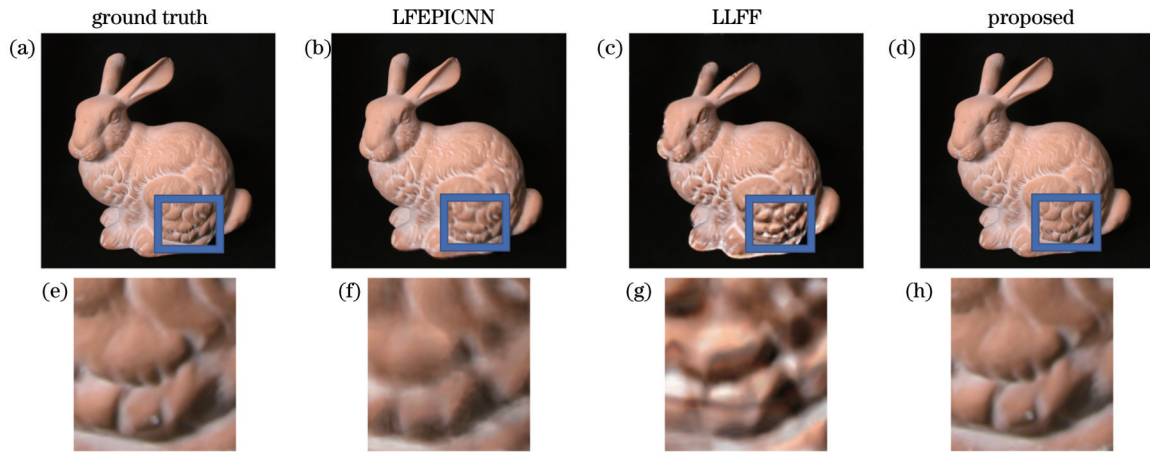


图 8 不同方法对斯坦福 bunny 场景角度域超分辨生成的新视角。(a)真实场景;(b)LFEPICNN 方法;(c)LLFF 方法;(d)提出方法;(e)、(f)、(g)、(h)分别为 bunny 场景使用以上方法后的局部放大图

Fig. 8 New perspective for angular super-resolution generation of Stanford bunny scenes. (a) Ground truth; (b) LFEPICNN method; (c) LLFF method; (d) proposed method; (e), (f), (g), (h) local zooms of bunny scene after using above methods, respectively

在斯坦福实际光场中也进行了角度域超分辨之后的 EPI 绘制,固定 $u=3, y=210$ 绘制 EPI。如图 9 所示,将实际数据角度域超分辨后的 EPI 和超分辨之前进行对比,图 9 (b)和 9(c)角度域超分辨后对应两张完整 EPI 的 PSNR 值为 43.92,并且在图 9 (d)中同一物点对应斜率在角度域超分辨后的 EPI 中明显具有更高的清晰度和分辨率,同样证明场景的深度、高频信息得到还原,并且达到了纹理细节还原度高的新视点重建效果。表 2 中的相关评估指标也验证了该结论。其中 PSNR 平均提升了 13.8%,SSIM 平均提升了 9.19%,LPIPS 平均降低了 63.2%。

5 讨 论

受光场采集系统的限制,光场数据在空间分辨率与角度分辨率之间存在相互制约,为了在现有光场相

机的基础上得到具有更多信息的光场数据,通常要提高光场的角度分辨率或空间分辨率。以前的大多数研究工作都是基于显式场景几何来进行新视点生成的。然而,在实际的光场场景信息中,不利条件如遮挡、深度变化和背景干扰等问题出现较多,对于纹理丰富细节处也较难表达,因为很难对上述情况的光场信息进行高质量的超分辨。

本文提出一种基于 NeRF 的光场角度域超分辨方法,通过 NeRF 来表示光场,编辑光场中相机姿态获得子孔径图像间的新视角,完成角度维度的密集采样。本文方法相对于所选方法 LLFF 和 LFEPICNN 的主要优势是基于 NeRF 来对光场场景进行隐式表达,为高分辨率的四维光场拟合出准确的隐式函数,准确表达具有复杂不利条件的光场场景,能够解决 LLFF 无法重建出场景有遮挡及深度具有变化的问题;并将输

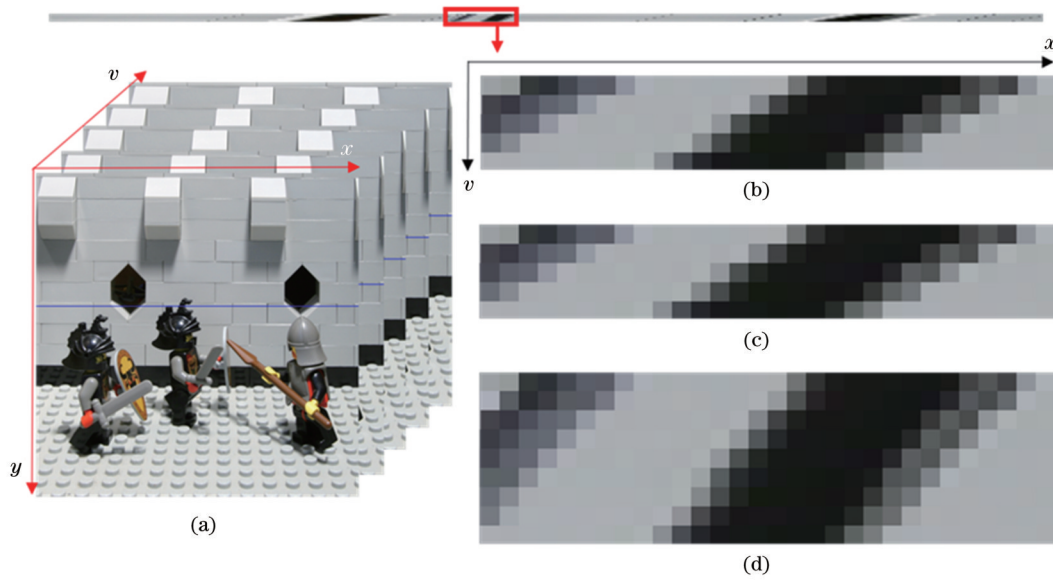


图9 斯坦福 knights 场景角度域超分辨前后的 EPI 局部图对比。(a) 固定 $u=3, y=210$ 绘制 EPI; (b) 原始数据 EPI 局部图; (c) 角度域超分辨后与原始数据对应位置的 EPI 局部图; (d) 角度域超分辨后的 EPI 局部图

Fig. 9 Comparison of partial EPI before and after angular super-resolution of Stanford knights scene. (a) Fix $u=3, y=210$ to draw EPI; (b) partial EPI of original data; (c) partial EPI of position corresponding to original data after angular super-resolution; (d) partial EPI after angular super-resolution

表 2 斯坦福实际光场数据集的指标评估

Table 2 Metrics evaluation on Stanford actual light field dataset

	Bunny			Knights		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
LFEPICNN	36.53	0.8971	0.096	35.79	0.8542	0.128
LLFF	35.34	0.8332	0.053	37.44	0.9023	0.045
Proposed	40.72	0.9436	0.014	41.35	0.9521	0.012

入变量通过位置编码映射到该变量的傅里叶特征,能够有效地解决 LFEPICNN 对场景高频纹理信息较难拟合的问题。

在 HCI 光场数据集和斯坦福光场数据集的实验结果表明,本文方法在光场角度域超分辨具有优异的性能,并且几个评估指标达到了最佳性能,这证明了所提方法的有效性。然而,在计算过程中还发现 NeRF 中的位置编码模块增加了模型的计算量,这导致计算量较大,无法满足对光场场景新视点的实时渲染。因此,优化模型计算将是后续工作的重要方向。

6 结 论

本文提出了一种新的基于 NeRF 的光场角度域超分辨算法。通过 NeRF 来表示光场,编辑光场中相机姿态获得子孔径图像间的新视角,完成角度维度的密集采样。在场景的隐式表达中,本文利用位置编码将输入变量映射到其傅里叶特征,来解决对于高频信息难以拟合的问题。在 HCI 模拟光场数据集的实验表明,本文方法在几个超分辨指标方面取得了最佳结果,并且明显优于其他方法。不仅是在模拟数据,在斯坦

福实际光场数据集的实验结果也论证了本文方法的有效性。总的来说,本文的超分辨方法不仅能处理如遮挡、深度变化和背景干扰出现的情况,并且也能对纹理丰富细节处进行高质量的还原。未来,将用该方法进行大型场景的实时渲染和场景重构。作为一种场景表示的新范式,NeRF 为光场计算成像提供了新的思路和新的方法,后续将进一步结合场景的几何信息和物理信息,提升计算成像性能和场景表示性能。

参 考 文 献

- [1] Gershun A. The light field[J]. Journal of Mathematics and Physics, 1939, 18(1/2/3/4): 51-151.
- [2] 殷永凯, 于锴, 于春展, 等. 几何光场三维成像综述[J]. 中国激光, 2021, 48(12): 1209001.
Yin Y K, Yu K, Yu C Z, et al. 3D imaging using geometric light field: a review[J]. Chinese Journal of Lasers, 2021, 48(12): 1209001.
- [3] Li J Q, Lu M L, Li Z N. Continuous depth map reconstruction from light fields[J]. IEEE Transactions on Image Processing, 2015, 24(11): 3257-3265.
- [4] Amit Y, Felzenszwalb P, Girshick R. Object detection[M]//Ikeuchi K. Computer vision. Cham: Springer, 2020: 1-9.
- [5] Boominathan V, Mitra K, Veeraraghavan A. Improving resolution and depth-of-field of light field cameras using a hybrid imaging system[C]//2014 IEEE International Conference on Computational Photography (ICCP), May 2-4, 2014, Santa Clara, CA, USA. New York: IEEE Press, 2014.
- [6] Adelson E H, Bergen J R. The plenoptic function and the elements of early vision[J]. Computational Models of Visual Processing, 1991, 1(2): 3-20.
- [7] 唐于珺, 王林波, 文刚, 等. 结构光照明超分辨成像图像重建算法研究进展[J]. 激光与光电子学进展, 2022, 59(6): 0617009.
Tang Y J, Wang L B, Wen G, et al. Recent advances in structured illumination microscope super-resolution image

- reconstruction[J]. *Laser & Optoelectronics Progress*, 2022, 59(6): 0617009.
- [8] Wang Z J, Zhao T Y, Hao H W, et al. High-speed image reconstruction for optically sectioned, super-resolution structured illumination microscopy[J]. *Advanced Photonics*, 2022, 4(2): 026003.
- [9] 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.
- [10] Georgiev T G, Zheng K C, Curless B, et al. Spatio-angular resolution tradeoffs in integral photography[J]. *Rendering Techniques*, 2006, 2006: 21.
- [11] Penner E, Zhang L. Soft 3D reconstruction for view synthesis[J]. *ACM Transactions on Graphics*, 2017, 36(6): 1-11.
- [12] Zhou T H, Tucker R, Flynn J, et al. Stereo magnification: learning view synthesis using multiplane images[EB/OL]. (2018-04-24)[2022-11-06]. <https://arxiv.org/abs/1805.09817>.
- [13] Mildenhall B, Srinivasan P P, Ortiz-Cayon R, et al. Local light field fusion: practical view synthesis with prescriptive sampling guidelines[J]. *ACM Transactions on Graphics*, 2019, 38(4): 1-14.
- [14] Seitz S M, Dyer C R. View morphing[C]//*Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, August 4-9, 1996, New Orleans, LA, USA. New York: ACM Press, 1996: 21-30.
- [15] Levoy M, Hanrahan P. Light field rendering[C]//*Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, August 4-9, 1996, New Orleans, LA, USA. New York: ACM Press, 1996: 31-42.
- [16] 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 (ICCVW)*, December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2016: 57-65.
- [17] Wang Y L, Liu F, Wang Z L, et al. End-to-end view synthesis for light field imaging with pseudo 4DCNN[M]//*Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science*. Cham: Springer, 2018, 11206: 340-355.
- [18] Yeung H W F, Hou J H, Chen J, et al. Fast light field reconstruction with deep coarse-to-fine modeling of spatial-angular clues[M]//*Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science*. Cham: Springer, 2018, 11210: 138-154.
- [19] Wu G C, Liu Y B, Fang L, et al. Light field reconstruction using convolutional network on EPI and extended applications[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, 41(7): 1681-1694.
- [20] 谢一博, 徐乃涛, 周顺, 等. 分布式红外阵列相机的超分辨率图像重建研究[J]. *激光与光电子学进展*, 2022, 59(16): 1611004. Xie Y B, Xu N T, Zhou S, et al. Super-resolution image reconstruction of distributed infrared array camera[J]. *Laser & Optoelectronics Progress*, 2022, 59(16): 1611004.
- [21] 常远, 盖孟. 基于NeRF的视点合成算法综述[J]. *图学学报*, 2021, 42(3): 376-384. Chang Y, Gai M. A review on neural radiance fields based view synthesis[J]. *Journal of Graphics*, 2021, 42(3): 376-384.
- [22] Mildenhall B, Srinivasan P P, Tancik M, et al. NeRF: representing scenes as neural radiance fields for view synthesis[J]. *Communications of the ACM*, 2022, 65(1): 99-106.
- [23] Vaswani A, Shazeer N, Parmar N, et al. Attention is all You need[C]//*Proceedings of the 31st International Conference on Neural Information Processing Systems*, December 4-9, 2017, Long Beach, California, USA. New York: ACM Press, 2017: 6000-6010.
- [24] Martin-Brualla R, Radwan N, Sajjadi M S M, et al. NeRF in the wild: neural radiance fields for unconstrained photo collections[C]//*2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 7206-7215.
- [25] Bojanowski P, Joulin A, Lopez-Paz D, et al. Optimizing the latent space of generative networks[EB/OL]. (2017-07-18)[2022-11-09]. <https://arxiv.org/abs/1707.05776>.
- [26] Park K, Sinha U, Barron J T, et al. Nerfies: deformable neural radiance fields[C]//*2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, October 10-17, 2021, Montreal, QC, Canada. New York: IEEE Press, 2022: 5845-5854.
- [27] Meng N, So H K H, Sun X, et al. High-dimensional dense residual convolutional neural network for light field reconstruction[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(3): 873-886.
- [28] Schönberger J L, Frahm J M. Structure-from-motion revisited[C]//*2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 4104-4113.
- [29] Huynh-Thu Q, Ghanbari M. Scope of validity of PSNR in image/video quality assessment[J]. *Electronics Letters*, 2008, 44(13): 800-801.
- [30] Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity[J]. *IEEE Transactions on Image Processing*, 2004, 13(4): 600-612.
- [31] Zhang R, Isola P, Efros A A, et al. The unreasonable effectiveness of deep features as a perceptual metric[C]//*2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 586-595.
- [32] 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.
- [33] Raj A S, Lowney M, Shah R, et al. Stanford lytro light field archive[EB/OL]. [2022-11-09]. <http://lightfields.stanford.edu/LF2016.html>.

Neural Radiance Field-Based Light Field Super-Resolution in Angular Domain

Miao Yuan, Liu Chang, Qiu Jun*

Institute of Applied Mathematics, Beijing Information Science and Technology University, Beijing 100101, China

Abstract

Objective In response to the trade-off of spatial-angular resolution in light field data acquisition due to data flux limitations, we propose a neural radiance field-based method to achieve high-quality light field super-resolution in the angular domain. The occlusion, depth variations, and background interference make super-resolution in the angular domain a challenging task, and it is difficult to express the rich details of the texture. In order to address these issues, many solutions are proposed in terms of novel view synthesis based on explicit and implicit scene geometry. However, both explicit and implicit scene geometry methods generate new viewpoint images of the scene from the geometric features of the scene, which are prone to problems such as noise interference and difficult reconstruction of textural details. Therefore, we propose neural radiance field-based light field super-resolution in the angular domain to reconstruct densely sampled light fields from sparse viewpoint sets, which can avoid errors and noises that may be introduced during image acquisition and improve the accuracy and quality of subsequent three-dimensional (3D) reconstruction.

Methods By training the neural network with the light field data, the neural radiance field captures the complete scene information, even for novel viewpoints, and thus enhances the scene representation performance. In order to achieve this, a multilayer perceptron is utilized to express a five-dimensional vector function that describes the geometry and color information of the 3D model. The image color is then predicted using volume rendering. The light field is subsequently represented by a neural radiance field, and dense sampling of the angular dimension is achieved by adjusting the camera pose in the light field to obtain new perspectives between the sub-aperture images. This approach overcomes the limitations of prior techniques, including occlusion, depth variation, and background interference in light field scenes. Additionally, the input variable is mapped to the Fourier features of that variable by positional encoding, effectively addressing the challenge of fitting to the high-frequency textural information of the scene.

Results and Discussions We propose the neural radiance field-based light field super-resolution in the angular domain by representing the light field by the neural radiance field. The main advantage of the proposed method over the selected experimental methods, such as local light field fusion (LLFF) and light field reconstruction using convolutional network on EPI (LFEPICNN) is that the proposed method is based on the neural radiance field to implicitly represent the light field scene, which can fit an accurate implicit function for the high-resolution four-dimensional light field and accurately represent the light field scene with complex conditions. The experimental results show that the super-resolution method based on the neural radiance field proposed can improve the angular resolution from 5×5 to 9×9 . The peak signal to noise ratio (PSNR) is improved by 13.8% on average, and the structural similarity (SSIM) is improved by 9.19% on average (Table 1 and Table 2).

Conclusions We propose a novel method of neural radiance field-based light field super-resolution in the angular domain. By representing the light field with the neural radiance field, the new perspective images between sub-aperture images are generated to achieve the dense sampling of the angular dimension. In the implicit representation of the scene, position encoding is utilized to map the input variables to their Fourier features to address the problem of difficult fitting for high-frequency information. Experiments on the HCI simulated light field dataset show that the proposed method achieves the best results in several super-resolution metrics and significantly outperforms other methods. Experimental results on the Stanford light field real dataset demonstrate the effectiveness of the method. Overall, the super-resolution method is not only able to deal with occlusions, depth variations, and background interference but also to obtain high-quality reproduction of rich textural details. In the future, the proposed method will be used for real-time rendering and scene reconstruction of large scenes. As a new paradigm for scene representation, neural radiance fields provide new ideas and methods for computational imaging of light fields, and we will further combine scenes' geometric and physical information to improve computational imaging performance and scene representation performance.

Key words imaging system; light field; neural radiance field; angular super-resolution; volume rendering