NeRF：将场景表示为视图合成的神经辐射场

本·米尔登霍尔[1] *普拉图普。斯里尼瓦桑[1] *马修Tancik[1] *乔纳森T。巴伦[2]拉瓦莫奥蒂[3]吴仁[1]

[1]加州大学伯克利分校[2]谷歌研究[3]加州大学圣地亚哥分校

摘要我们提出了一种方法，通过使用稀疏的输入视图集来优化底层的连续体积场景函数，来实现合成复杂场景的新视图的最新结果。我们的算法使用一个全连接（非卷积）深度网络表示一个场景，该网络的输入是一个连续的5D坐标（空间位置（x，y，z）和观看方向（e，））），其输出是该空间位置的体积密度和视图相关的发射辐射。ϕ我们通过查询沿着相机光线的5D坐标来合成视图，并使用经典的体积渲染技术将输出的颜色和密度投影到图像中。因为体积渲染自然是可分割的，所以优化我们的表示所需的唯一输入是一组具有已知相机姿态的图像。我们描述了如何简单地优化神经辐射，以渲染具有复杂几何和外观的场景的逼真新视图，并展示了在神经渲染和视图合成方面的结果。查看综合结果最好是作为视频来查看，所以我们敦促读者查看我们的补充视频，以进行令人信服的比较。

关键词：场景表示、视图合成、基于图像的渲染、体渲染、3D深度学习

介绍

1

在这项工作中，我们以一种新的方式解决了长期存在的视图合成问题，即直接优化连续五维场景表示的参数，以最小化呈现一组捕获图像的误差。

$\theta\phi$我们将一个静态场景表示为一个连续的5D函数，它输出空间中每个点（x，y，z）在每个方向（，）发射的辐射，以及每个点的密度，它就像一个环形不透明度，控制光线通过（x，y，z）累积的辐射。我们的方法优化了一个没有任何卷积层（通常称为MLP）的多层感知器或MLP)的深度全连接神经网络，通过从单个5D坐标（x，y，z，，）回归到单个体积密度和视图相关的RGB颜色来表示这个函数。$\theta\phi$以保持此神经辐射（NeRF）

* 作者对这项工作的贡献相等。

图1：我们提出了一种方法，从一组输入图像中优化一个场景的连续5维神经辐射场表示（任何连续位置的体积密度和视图依赖的颜色）。我们使用卷渲染的技术沿光线积累场景表示的样本，从而从任何视点渲染场景。在这里，我们可视化了在周围半球随机捕获的合成鼓场景的100个输入视图集，并展示了从我们优化的NeRF表示中呈现的两个新视图。

从一个特定的角度我们：1)3月相机射线通过场景生成一组采样的3d点，2)使用这些点及其相应的2d查看方向输入神经网络产生一个输出的颜色和密度，和3)使用经典体积渲染技术积累这些颜色和密度到一个2d图像。因为这个过程自然是可分散的，我们可以使用梯度下降来优化这个模型，通过最小化每个观察图像和相应的视图从我们的表示呈现之间的误差。在多个视图中最小化这个错误，可以鼓励网络通过为包含真实的底层场景内容的位置分配高体积密度和准确的颜色来预测场景的连贯模型。图2可视化了这个整体管道。

我们认为，为一个复杂的场景优化神经辐射场表示的基本实现并不收敛于一个合理的高分辨率表示，并且在每个摄像机射线所需的样本数量上是微不足道的。ffiffi我们通过将输入的5D坐标与位置编码进行转换来解决这些问题，使MLP能够表示更高频率的函数，我们提出了一种分层采样程序，以减少充分采样这种高频场景表示所需的查询数量。 我们的方法继承了体积表示的优点：两者都可以表示复杂的真实世界的几何形状和外观，并且非常适合使用投影图像进行基于梯度的优化。至关重要的是，我们的方法克服了在高分辨率建模复杂场景时离散体素网格的高昂存储成本。总之，我们的技术贡献包括：

- 一种用复杂的几何图形和材料表示连续场景为5D神经辐射图像的方法，参数化为基本的MLP网络。
- 一个基于经典的体积渲染技术的可分割的渲染过程，我们使用它来优化来自标准RGB图像的这些表示。这包括一个分层采样策略，将MLP的容量分配到具有可见场景内容的空间。

-位置编码将每个输入的5D坐标映射到更高维空间，使我们能够成功地优化神经辐射图像来表示高频场景内容。

我们证明了我们所得到的神经辐射等方法在定量和定性上优于最先进的视图合成方法，包括它对场景的神经三维表示的工作，以及训练深度卷积网络来预测采样体积表示的工作。据我们所知，本文提出了第一个连续的神经场景表示方法，它能够从自然环境中捕获的RGB图像中呈现真实物体和场景的高分辨率逼真的新视图。

2相关工作

计算机视觉的一个很有前途的最近方向是在MLP的权重中编码对象和场景，它直接从三维空间位置映射到形状的隐式表示，例如该位置的有符号距离[6]。然而，到目前为止，这些方法还无法像使用三角形网格或体素网格等离散表示场景一样的保真度来再现具有复杂几何形状的真实场景。在本节中，我们将回顾这两项工作，并将它们与我们的方法进行对比，这增强了神经场景表示的能力，从而产生了渲染复杂的现实场景的最先进的结果。

使用MLPs从低维坐标映射到颜色的类似方法也被用于表示其他图形功能，如图像[44]、纹理材料[12,31,36,37]和间接照明值[38]。

神经三维形状表示最近的工作通过优化深度网络，将xyX坐标映射到有符号的距离函数[15,32]或占用率[11,27]，研究了连续三维形状的隐式表示。然而，这些模型受到其访问地面真实三维几何的需求的限制，通常来自合成的三维形状数据集，如ShapeNet [3]。随后的工作通过制定可分割的渲染函数，放宽了地面真实三维形状的要求，允许神经隐式形状表示征仅使用二维图像进行优化。Niemeyer等。[29]将曲面表示为三维占用值，并使用数值方法来识别每条光线的曲面交点，然后使用隐式扩散法计算精确的导数。每个射线交叉位置作为神经三维纹理的输入，该纹理预测该点的中断颜色。西茨曼等人。[42]使用一种不那么直接的神经3D表示，它只是在每个连续的3D坐标上输出一个特征向量和RGB颜色，并提出一个可分割的呈现函数，由递归神经网络沿着每条射线行进，以决定表面的位置。

虽然这些技术可能代表复杂和高分辨率的几何，但到目前为止，它们仅限于低几何复杂度的简单形状，导致过度平滑渲染。我们展示了一种优化网络来编码5D辐射系数(三维体积

与二维视图依赖的外观)可以代表更高分辨率的几何图形和外观，以呈现复杂场景的逼真的新视图。

给定了一个视图的密集采样，可以通过简单的光场样本插值技术[21,5,7]来重建逼真的新视图。对于具有稀疏视图采样的新视图合成，计算机视觉和图形社区通过从观察到的图像中预测传统的几何图形和外观表示，取得了显著的进展。一种流行的方法是使用基于网格的场景表示，它们具有中断的[48]或依赖于视图的[2,8,49]外观。可分散光栅化器[4,10,23,25]或路径追踪器[22,30]可以直接优化网格表示，以使用梯度下降复制一组输入图像。然而，基于图像重投影的基于梯度的网格优化往往是针状的，可能是由于局部最小值或损失景观的条件较差。此外，该策略要求在优化[22]之前提供一个具有混合拓扑的模板网格作为初始化，这对于无约束的真实场景通常是不可用的。

另一类方法使用体积表示来解决从一组输入的RGB图像中进行高质量逼真的视点合成的任务。体积方法能够真实地表示复杂的形状和材料，非常适合基于梯度的优化，并且往往比基于网格的方法产生更少的视觉上分散注意力的伪影。早期的体积测量方法使用观察到的图像来直接彩色体素网格[19,40,45]。最近，[9,13,17,28,33,43,46,52]的几种方法使用多个场景的大型数据集来训练深度网络，这些网络从一组输入图像中预测采样的体积表示，然后使用阿尔法合成[34]或学习合成在测试时呈现新的视图。其他工作已经优化了每个特定场景的卷积网络（CNNs）和采样体素网格的组合，这样CNN就可以补偿低分辨率体素网格[41]的离散伪影，或者允许预测体素网格根据输入时间或动画控制[24]而变化。虽然这些体积技术在新视图合成方面取得了令人印象深刻的结果，但由于其离散采样，它们的高分辨率图像的时空复杂性较差——渲染高分辨率图像需要三维空间的iner采样。我们通过编码一个深度全连接神经网络的参数内的连续体积来规避这个问题，这不仅产生比之前的体积方法更高质量的渲染，而且只需要这些采样体积表示的存储成本的一小部分。

3神经辐射场场景表示

我们将一个连续的场景表示为一个5D向量值函数，它的输入是一个三维位置x =（x，y，z）和2D观看方向（，），其输出是一个发射的颜色c =（r，g，b）和体积密度a。$\theta\phi$在实践中，我们表达

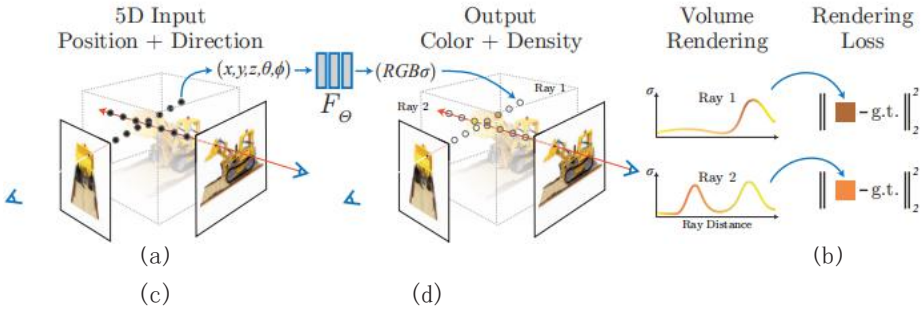


图2：对我们的神经辐射等地呈现场景表示和可分割的渲染程序的概述。我们通过沿着相机射线(a)采样5D坐标（位置和观看方向）来合成图像，将这些位置输入MLP以产生颜色和体积密度(b)，并使用体积渲染技术将这些值合成成图像(c).这个渲染函数是可分割的，所以我们可以通过最小化合成和地面真实观测图像(d).之间的残差来优化我们的场景表示

方向作为三维笛卡尔单位向量d。我们用MLP网络F：来近似（x，d）场景表示！ $_{\Theta}$ （c，a），并优化其权重，从每个输入的5D坐标映射到其相应的体积密度和方向发射的颜色。

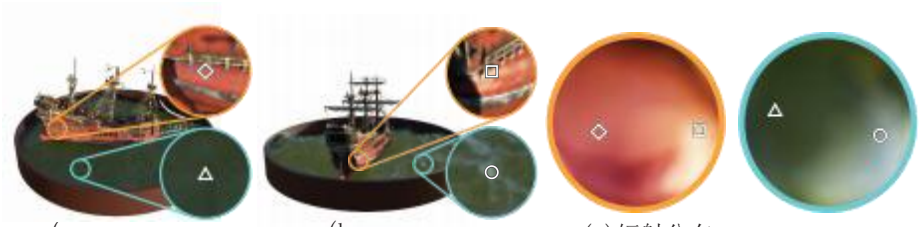
我们通过限制网络预测体积密度a作为位置x的函数，同时允许RGB颜色c被预测为多视图一致。$_{\Theta}$为了实现这一点，MLP F首先处理具有8个全连接层的输入3D坐标x（使用ReLU激活和每层256个通道），并输出a和一个256维的特征向量。然后，这个特征向量与相机射线的查看方向连接起来，并传递到一个额外的全连接层（使用ReLU激活和128个通道），输出与视图相关的RGB颜色。

见图。3为我们的方法如何使用输入查看方向来表示非兰伯特缺陷的一个例子。如图所示。ffi4、一个没有视图依赖（只有x作为输入）训练的模型具有表示推测的缺陷。

4具有发光字段的卷体渲染

我们的5D神经辐射表示一个场景作为体积密度和定向发射辐射在空间的任何点。我们使用经典体积渲染[16]的原理来渲染任何通过场景的光线的颜色。体积密度a (x)可以解释为一条射线在x位置处终止于一个微小粒子的周向概率。具有近边界和远边界t的相机光线=(t)的期望颜色C（+）n和tf是

$$C(\mathbf{r}) = \int_{t_n}^{t_f} T(t)\sigma(\mathbf{r}(t))\mathbf{c}(\mathbf{r}(t),\mathbf{d})dt\,, \text{ where } T(t) = \exp\left(-\int_{t_n}^{t}\sigma(\mathbf{r}(s))ds\right). \quad (1)$$



(a 视图1 (b 视图2 (c)辐射分布

图3：一种与视图相关的发射辐射的可视化方法。我们的神经辐射场表示输出RGB颜色作为空间位置x和观察方向d的5D函数。在这里，我们在船舶场景的神经表示中可视化两个空间位置的方向颜色分布。在(a)和(b)中，我们展示了两个来自两个不同相机位置的混合3D点的外观：一个在船的一侧（橙色插图），另一个在水面上（蓝色插图）。我们的方法预测了这两个三维点的不断变化的镜面外观，并且在(c)中，我们展示了这种行为如何在观看方向的整个半球连续地推广。

函数T (t)表示沿t的累积透射率nt，我。e.，光线从t开始传播的概率n到t而不击中任何其他粒子。从我们的连续神经辐射场渲染一个视图需要估计通过所需的虚拟相机的每个像素跟踪的相机射线的这个积分C (r)。

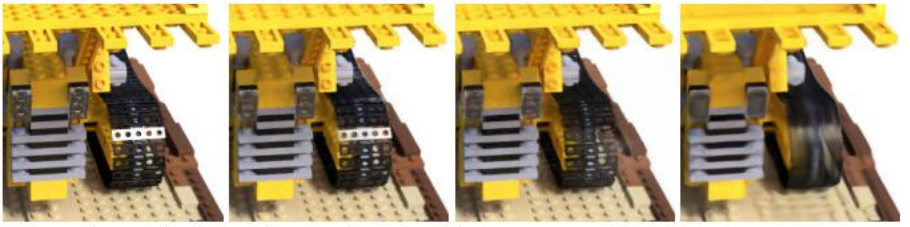
我们用求积法对这个连续积分进行了数值估计。确定性求交通常用于渲染离散体素网格，它会极大地限制我们表示的分辨率，因为MLP只会在一个离散位置集上查询。相反，我们使用分层抽样方法，其中我们划分[tn， tf]分成N个均匀间隔的箱子，然后从每个箱子内均匀地随机抽取一个样本：

$$t_i \sim \mathcal{U}\left[t_n + \frac{i-1}{N}(t_f - t_n),\ t_n + \frac{i}{N}(t_f - t_n)\right]. \tag{2}$$

虽然我们使用一组离散的样本来估计积分，但分层抽样使我们能够表示一个连续的场景表示，因为它导致MLP在优化过程中在连续的位置进行评估。我们使用这些样本来估计C (r)与求积规则在Max [26]讨论的体积渲染审查中：

$$\hat{C}(\mathbf{r}) = \sum_{i=1}^{N} T_i(1 - \exp(-\sigma_i\delta_i))\mathbf{c}_i\,,\ \text{where } T_i = \exp\left(-\sum_{j=1}^{i-1}\sigma_j\delta_j\right), \tag{3}$$

其中6i= ti+1— ti是相邻样本之间的距离。这个函数用于从(c)的集合中计算(r)$\hat{C}$i， ai)的值是简单的可区分的，并简化为传统的alpha组合与alpha值α i=1-exp(-ai6i).



地面真相，完整的模型，没有视图依赖性，没有位置编码

图4：在这里，我们可视化了我们的完整模型是如何通过表示与视图相关的发射辐射和通过高频位置编码传递我们的输入坐标的。去除视图依赖关系可防止模型在推土机胎面上重新创建镜面反射。去除位置编码大大降低了模型表示高频几何图形和纹理的能力，导致了过度平滑的外观。

5优化神经辐射场

在前一节中，我们已经描述了将场景建模为神经辐射并从此表示呈现新视图所需的核心组件。然而，我们观察到，这些组件并不适合实现最先进的质量，如第6.4节所示)。我们引入了两个改进来表示高分辨率的复杂场景。第一个是输入坐标的位置编码，帮助MLP表示高频函数，第二个是分层采样程序，允许我们对这种高频表示进行精确的采样。

.15位置编码

尽管神经网络是通用的函数逼近器[14]，但我们发现，让网络F直接作用于xyz输入坐标，会导致渲染图在表示颜色和几何形状的高频变化方面表现不佳。这与拉哈曼等人最近的研究结果是一致的。[35]，这表明深度网络倾向于学习低频函数。他们还表明，在将输入的高频函数传递给网络之前，使用高频函数将输入映射到高维空间，可以更好地拟合包含高频变化的数据。

我们在神经场景表征的背景下利用这些信息，并表明将F重新定义为两个函数F = F O V的组成，一个是学习的，另一个不是，显著提高了性能(见图。4和表2）。这里V是从R到更高维空间R的映射，而F仍然只是一个普通的MLP。在形式上，我们使用的编码函数是：

$$V(p) = (\sin(2^0 p)\pi, \cos(2^0 p)\pi, \cdots, \sin(2^{L-1} p)\pi, \cos(2^{L-1} p)\pi). \quad (4)$$

该函数V（•）分别应用于x中的三个坐标值（它们被归一化为位于[-1,1])和的三个分量

笛卡尔观测方向单位向量d（构造为在-1]中）。在我们的实验中，我们将V (x)的L = 10设置为V (d).，将L = 4设置为V (d).

在流行的变压器架构[47]中也使用了类似的映射，其中它被称为位置编码。然而，变形金刚使用它来实现一个不同的目标，即提供序列中标记的离散位置，作为不包含任何顺序概念的体系结构的输入。相反，我们使用这些函数将连续输入坐标映射到更高维空间，使我们的MLP更容易接近更高的频率函数。同时研究从投影建模三维蛋白质结构的相关问题的工作，[51]也利用了类似的输入坐标映射。

.25分层体积采样

我们的渲染策略是在每个摄像机射线的N个查询点上密集评估神经辐射的影响网络是有害的：对渲染图像没有贡献的自由空间和遮挡区域仍然被重复采样。ffi我们从体积渲染[20]的早期工作中获得灵感，并提出了一种层次表示，通过与inal渲染上的预期效果成比例分配样本来增加渲染的非显著性。ffi

我们不是仅仅使用一个网络来表示场景，而是同时优化两个网络：一个“粗”和一个“ine”。我们首先抽样得到一个N的集合c使用分层抽样的位置，并评估这些位置的“粗糙”网络。2和3。给定这个“粗糙”网络的输出，然后我们会沿着每条射线产生一个更知情的点采样，其中样本偏向于体积的相关部分。为此，我们首先从粗网络中重写alpha合成的颜色$\hat{C}_c(r)$在等式。3作为所有采样颜色的加权和c_i沿着光线：

$$\hat{C}_c(\mathbf{r}) = \sum_{i=1}^{N_c} w_i c_i\,, \qquad w_i = T_i(1-\exp(-\sigma_i\delta_i))\,. \tag{5}$$

将这些重量标准化为$\hat{w}_i=w_i/\sum_{j=1}^{N_c} w_j$沿着射线产生一个分段常数的PDF。我们对第二组N进行采样f使用逆变换抽样从这个分布中得到的位置，在第一集和第二集样本的并集处评估我们的“ine”网络，并计算射线的最终渲染颜色 $\hat{C}_f(r)$使用方程。3，但使用了所有的N_c+N_f样品这个过程将更多的样本分配到我们希望包含可见内容的区域。这解决了一个与重要性抽样类似的目标，但我们使用采样值作为整个积分域的非均匀离散化，而不是将每个样本作为整个积分的独立概率估计。

5.3实施细节

我们为每个场景优化了一个单独的神经连续体积表示网络。这只需要一个被捕获的场景的RGB图像的数据集，

相应的相机姿态和内在参数，以及场景边界（我们使用地面真实相机姿态、内在边界和合成数据，并使用COLMAP运动结构软件包[39]来估计真实数据的这些参数）。在每次优化迭代中，我们从数据集中所有像素的集合中随机抽取一批相机光线，然后按照Sec中描述的分层采样进行采样。5.2查询N_c样本来自粗网络和$N_c+ N_f$来自ine网络的样本。然后，我们使用Sec中描述的卷渲染过程。4来渲染来自两组样本中的每条射线的颜色。我们的损失只是粗渲染和正弦渲染的渲染颜色和真像素颜色之间的总平方误差：

$$\mathcal{L} = \sum_{\mathbf{r}\in\mathcal{R}} \left[\left\| \hat{C}_c(\mathbf{r}) - C(\mathbf{r}) \right\|_2^2 + \left\| \hat{C}_f(\mathbf{r}) - C(\mathbf{r}) \right\|_2^2 \right] \tag{6}$$

其中R是每批中的光线集合，C (r)，$\hat{C}_c$(r)，以及$\hat{C}_f$(r)分别为射线r的地面真实值、粗体积预测值和ine体积预测的RGB颜色。请注意，即使inal渲染来自于$\hat{C}_f$(r)，我们也最小化了损失的损失$\hat{C}_c$(r)，使来自粗网络的权值分布可以用来在ine网络中分配样本。

在我们的实验中，我们使用了4096条射线的批次大小，每条采样在N_c= 64在粗体积和N中的坐标$_f$= 128中的附加坐标。我们使用Adam优化器[18]，其学习速率从5 10开始$\times^{-4}$然后呈指数级衰减到5 10$\times^{-5}$在优化过程中(其他Adam超参数的默认值为β_1= 0.9， β_2=为0.999，和e=为10^{-7}).对单个场景的优化通常需要大约100-300k的迭代才能收敛到单个NVIDIA V100 GPU上（大约1-2天）。

6结果

我们定量的是表1)和定性的是图。8和6)表明，我们的方法优于之前的工作，并提供了广泛的消融研究来验证我们的设计选择（表2）。我们敦促读者观看我们的补充视频，以更好地欣赏我们的方法在呈现新视图的平滑路径时比基线方法的重大改进。

.16个数据集

我们首先展示了在两个物体的合成渲染数据集上的实验结果(表1，“混淆合成360”^{O}和“现实合成360”$^{O''}$).深度体素[41]数据集包含四个具有简单几何形状的兰伯特对象。×每个对象从上半球采样的视点中呈现512 512像素（479作为输入，1000用于测试）。此外，我们还生成了我们自己的数据集，其中包含8个物体的路径跟踪图像，显示出复杂的几何形状和真实的非兰伯特材料。6个是从上半球采样的视点渲染，2个是从全球上采样的视点渲染。我们渲染每个场景的100个视图作为输入，200个用于测试，所有的像素都是800 800像素。×

方法PSNR“SSIM”LPIPSt SRN [42] 33: 20 0：963 0：073 NV [24] 29：62 0：929 0：099 LLFF [28] 34：38 0：985 0：048我们的40：15 0：991 0：023	二熔断器合成360^0[41]			现实合成360^0			真实的前向[28]		
	PSNR″	ssim”	LPIPSt	PSNR″	ssim“	LPIPSt	PSNR″	ssim“	LPIPSt
	22:26	0:846	0:170	22:84	0:	0:378			
	26:05	0:893	0:160	–	668	–			
	24:88	0:911	0:114	24:13	_	0:212			
	31:01	**0:947**	0:081	26:50	0:798	0:250			

表1：我们的方法在定量上优于之前的合成和真实图像数据集上的工作。我们报告了PSNR/SSIM（越高越好）和LPIPS[50]（越低越好）。DeepVoxels [41]数据集由4个具有简单的几何形状的分解对象组成。我们的真实合成数据集由8个几何上复杂的物体与复杂的非兰伯材料的路径跟踪渲染图组成。真实的数据集由8个真实世界场景的手持式正向捕获组成（NV不能在这个数据上进行评估，因为它只在一个有限制的体积内重建对象）。虽然LLFF实现了稍好的LPIPS，但我们敦促读者查看我们的补充视频，我们的方法获得了更好的多视图一致性，并产生比所有基线更少的工影。

复杂场景的真实图像我们展示了用大致向前的图像捕获的复杂真实世界场景的结果（表1，“真实的面向”）。这个数据集包括用手持手机拍摄的8个场景（5个取自LLFF纸，3个由我们拍摄），拍摄了20到62张图像，并保持不变1/8个用于测试集。所有图像均为1008 756像素。×

6.2比较

为了评估我们的模型，我们将其与当前性能最好的视图合成技术进行了比较，详细介绍如下。所有方法使用相同的输入视图训练一个单独的网络为每个场景除了本地光场融合[28]，训练一个三维卷积网络在一个大数据集，然后使用相同的训练网络处理输入图像的新场景在测试时间。

神经体积（NV）[24]合成了一些物体的新观点，这些物体完全位于一个独特的背景前的有限体积内（必须在没有感兴趣的物体的情况下单独捕获）。它优化了一个深度三维卷积网络来预测一个离散的RGBα体素网格^{3}样品以及一个3D翘曲网格与32个^{3}样品该算法通过使摄像机光线通过扭曲的体素网格来呈现新的视图。

场景表示网络（SRN）[42]将一个连续的场景表示为一个不透明的表面，由一个MLP隐式定义，它将每个（x，y，z）坐标映射到一个特征向量。他们利用任何三维坐标上的特征向量来训练循环神经网络沿着光线的下一步大小，通过场景表示。从inal步骤开始的特征向量被解码为表面上该点的单一颜色。请注意，SRN是同一作者对Deep体素[41]的一个更好的后续表现，这就是为什么我们不包括与Deep体素的比较。

图5：对使用基于物理的渲染器生成的新合成数据集的场景的测试集视图进行比较。我们的方法能够恢复几何和外观的细节，如船的索具，乐高的齿轮和踏板，麦克风闪亮的站架和网格格栅，和材料的非兰伯特阻力。LLFF在麦克风支架上展示带状文物，材料的物品边缘，以及在船的桅杆和乐高物体内部的重影物品。SRN在每种情况下都会产生模糊和扭曲的渲染图。神经体积不能捕捉麦克风格栅或乐高齿轮的细节，它完全不能恢复船舶索具的几何形状。

图6：对真实世界场景的测试集视图进行比较。LLFF是专门为这个用例（对真实场景的正向捕获）而设计的。与LLFF相比，我们的方法能够在渲染视图中更一致地表示ine几何图形，如Fern的叶子和T-rex中的骨架肋骨和栏杆所示。我们的方法还正确地重建了LLFF难以清晰渲染的部分封闭区域，如底部蕨类作物叶子后面的黄色架子和底部兰花作物背景下的绿叶。在多个渲染之间混合也会导致LLFF的重复边缘，如在顶级兰花作物中看到的。SRN捕捉每个场景中的低频几何形状和颜色变化，但不能重现任何弦细节。

局部光场融合（LLFF）[28] LLFF是设计为良好采样的正面场景产生逼真的新视图。它使用训练好的三维卷积网络直接预测每个输入视图的离散的反采样RGBα网格（多平面图像或MPI [52]），然后通过alpha合成和将附近的MPI混合到新的视点中来呈现新的视图。

.36讨论

我们完全优于两个基线，它们也优化了每个场景中单独的网络（NV和SRN）。此外，我们只使用它们的输入图像作为我们的整个训练集，在定性和定量上产生更好的渲染优于LLFF（除了一个指标以外的所有指标）。

SRN方法产生高度平滑的几何和纹理，其视图合成的表示能力受到限制，因为每个相机射线只选择一个深度和颜色。NV基线能够捕获合理详细的体积几何形状和外观，但它使用了一个底层的显式128^3体素网格阻止它缩放以表示高分辨率的i个细节。LLFF特别提供了一个不超过输入视图之间差异64像素的“采样指南”，因此它经常无法估计包含多达400-500像素视图之间差异的合成数据集中的正确几何图形。此外，LLFF混合了不同的场景表示，以呈现不同的视图，导致感知分散的不一致，这在我们的补充视频中很明显。

这些方法之间最大的实际权衡是时间与空间。所有比较的单个场景方法对每个场景至少需要12个小时的训练。相比之下，LLFF可以在10分钟内处理一个小的输入数据集。然而，LLFF为每个输入图像产生一个大的3D体素网格，导致了巨大的存储需求（一个“现实合成”场景超过15GB）。我们的方法只需要5 MB的网络权值（相对于LLFF，相对压缩为3000），这甚至比我们任何数据集的单个场景的单独输入图像的内存更少。×

.46消融研究

我们在表2中通过广泛的消融研究验证了我们的算法的设计选择和参数。我们展示了我们的“现实合成360”的结果^{0}景色第9行显示了我们的完整模型作为一个参考点。第1行显示了我们的模型的极简版本，没有位置编码（PE）、视图依赖（VD）或分层采样(H)。在第2-4行中，我们一次从完整模型中删除这三个组件，观察到位置编码（第2行）和视图依赖（第3行）提供了最大的定量支持，其次是分层采样（第4行）。第5-6行显示了我们的性能是如何随着输入图像数量的减少而下降的。请注意，当我们的方法提供100张图像时，它们在所有指标上的性能仍然超过了NV、SRN和LLFF（见补充材料）。在第7-8行中，我们验证了我们对最大频率的选择

	输入	#Im.	L	(Nc,	PSNR″	ssim”	LPIPSt
1)无PE、VD、H	xyz	100	–	Nf)	26 28	0。	0.136
2)没有职位。编码	xyzeϕ	100	–	(256, –	27	906	0.108
3)无视图依赖性	xyz	100	10	) (64,	30..6	0.924	0.117
4)无层次结构	xyzeϕ	100	10	128)	7.77.6	0.925	0.109
5)图像少得多	xyzeϕ	25	10	(64,	606	0.938	0.107
6)图像较少	xyzeϕ	50	10	128)	27.78	0.925	0.096
7)频率较低	xyzeϕ	100	5	(256, –	29.79	0.940	0.088
8)更多频率	xyzeϕ	100	15	) (64,	30.59	0.944	0.096

表2：对我们的模型进行的消融研究。指标是来自我们的真实合成数据集的8个场景的平均值。看到秒。6.4详细说明。

L在我们的位置编码中使用的x（d使用的最大频率是按比例缩放的）。只使用5个频率会降低性能，但将频率的数量从10个增加到15个并不能提高性能。我们认为增加L的好处是有限的，一次2^L超过了采样输入图像中出现的最大频率（在我们的数据中大约为1024个）。

7结论

我们的工作直接解决了之前使用mlp将对象和场景表示为连续函数的工作的不足。我们证明，将场景表示为5d神经辐射轴（一个MLP输出体积密度和视图依赖的发射辐射作为3d位置和2d观看方向的函数）产生更好的渲染比以前训练深度卷积网络输出离散体素表示的主要方法。ffi虽然我们已经提出了一种分层采样策略，使渲染样本更有效（用于训练和测试），但在研究精确优化和渲染神经辐射图像的技术方面仍有很大的进展。ffi未来工作的另一个方向是可解释性：体素网格和网格等采样表示允许对渲染视图和失败模式的预期质量进行推理，但当我们在深度神经网络的权值中对场景进行编码时，还不清楚如何分析这些问题。我们相信，这项工作在基于真实世界图像的图形管道方面取得了进展，其中复杂的场景可以由从实际物体和场景的图像中优化出来的神经辐射图像组成。

感谢曹凯文、杨国伟和拉加万的评论和讨论。RR承认资金来自ONR拨款N000141712687和N000142012529和Ronald L。格雷厄姆主席。BM由赫兹基金会奖学金资助，MT由美国国家科学基金会研究生奖学金资助。谷歌通过BAIR公共项目提供了大量的云计算积分捐赠。我们感谢以下内容

混合交换用户中使用的模型：格雷格扎尔（船），1DInc（椅子）、布里亚纳霍斯（鼓）、赫伯霍尔德（法庭）、无埃里克（热狗）、海因泽尼斯（乐高）、埃尔布鲁德拉特里布（材料）和向上3d。de（麦克风）。

参考文献

1. 阿巴迪，阿加瓦尔， A.， 巴勒姆，布莱夫多，E.，陈，Z.，城市， C.， 科拉多，G。S.，戴维斯， A.， 迪恩，德文，先生，格赫马瓦特，古德费罗，我，竖琴， A.， 欧文，G.，伊萨德，贾，Y.，J.，凯泽，库德鲁尔，M.，莱文伯格，J。˝长鬃毛 D.， 蒙加，摩尔，默里， D.， 欧拉 C.， 舒斯特尔，施伦斯，斯坦纳， B.， 萨茨克弗，塔尔瓦，塔克，万霍克，瓦苏德万，维加斯，奥，沃登，瓦滕伯格，威克，余，郑，张流：异构系统的大型机器学习（2015）
2. 比勒，C.，博斯，M.，麦克米兰，戈特勒，科恩，M.：非结构化发光渲染。in：签名（2001）
3. 张，A。X.，芬克豪瑟，T.，吉巴斯，L.，汉拉汉，P.，黄，Q，李，Z，萨瓦雷斯，S，萨瓦，M，宋，S，苏，H，等：沙佩内：信息丰富的3d模型库。arXiv:1512.03012 (2015)
4. 陈，W.，高，J.，凌，H.，史密斯，E。J.，莱赫蒂宁，雅各布森，菲德勒：学习用一个基于插值的可分割渲染器来预测三维对象。在： NeurIPS（2019）
5. 科恩，戈特勒，S。J.，泽利斯基，R.，格雷泽利斯基。在：签名（1996）
6. 参考文献：：一种从范围图像中建立复杂模型的体积方法。在：签名（1996）
7. 戴维斯，A.，利维伊，M.，杜兰德，F.：非结构化的照明设备。在：欧洲图形（2012）
8. Debevec，P.，泰勒，C。J.，Malik，J.：从照片中建模和渲染架构：一种基于几何和图像的混合方法。在：签名（1996）
9. 弗林，布罗克斯顿，德贝维克，杜瓦尔，费夫，奥弗贝克，R.，斯纳弗利，N。深度视图：查看与学习到的梯度下降的合成。在： CVPR (2019)
10. 热诺亚，科尔，马斯奇诺特， A.，萨纳 A.， 瓦， D.，， 弗里曼，W。T.：三维可变形模型回归的无监督训练。输入：CVPR（2018）
11. 热诺亚，科尔，弗，苏德， A.，萨纳 A.，T.：三维形状的局部深度隐式函数。输入：CVPR（2020）
12. 亨兹勒，P.，米特拉，N。J.，里歇尔，T.：从2d范例中学习神经三维纹理空间。输入：CVPR（2020）
13. 单图像断层扫描：2d颅x射线的3d体积。在：欧洲图形（2018）
14. 例如：多层前馈网络是一种通用的逼近器。神经网络（1989）
15. 江市， C.， Sud, A.， Makadia, A.， 黄建华先生：三维场景的局部隐式网格表示。输入：CVPR（2020）
16. KajiyaJ。T.，HerzenB。P.V.：光线追踪的体积密度。计算机图形学（签名图）（1984年）
17. Kar, A.，H¨ane, C.， 学习多视角立体声机。在： NeurIPS (2017)
18. 金玛，D。P.，Ba，J.： Adam：一种随机优化的方法。在： ICLR (2015)

16 B.米尔登霍尔，P。P.斯里尼瓦桑。Tancik等人。

19.Kutulakos，K.N.，Seitz。M.：一种通过空间雕刻而形成的形状理论。国际计算机视觉杂志（2000年）

20.ffiM.：对体积数据的有效射线追踪。图形上的ACM事务
(1990)

21.利沃伊，M.，汉拉汉，P.：光ield渲染。在：签名（1996）

22.李，T。M.，阿塔拉，M.，杜兰德，F.，莱赫蒂宁，J.：通过边缘采样进行可区分的蒙特卡洛射线追踪。ACM图形交易（亚洲）
(2018)

23.刘，S.，李，t.，陈，W.，李，H.：软光栅化器：一个基于图像的三维推理的可区分渲染器。输入：ICCV（2019）

24.隆巴迪，西蒙，萨拉吉，施瓦茨，莱尔曼，A.，谢赫，Y。：神经体积：从图像中学习动态可渲染的体积。ACM图形交易（签名图）（2019年）

25.罗珀，M。M.，黑色，M。J.： OpenDR：一个近似的可区分的渲染器。输入：ECCV（2014）

26.马克斯，名词：直接体积渲染的光学模型。IEEE可视化与计算机图形学学报（1995）

27.美国，美国，A.：占用网络：在功能空间中学习三维重建。输入：CVPR（2019）

28.米尔登霍尔，B.，斯里尼瓦桑，P。P.，Ortiz-Cayon，R，卡兰塔里，N。K.，Ramamoorthi，R.局部光场融合：具有规范抽样指南的实用观点综合。ACM图形交易（签名图）（2019年）

29.可发散的体积渲染：在没有三维监督的情况下学习隐式的三维表示。在： CVPR
(2019)

30.三菱2：一个可延迟的向前和反向渲染器。ACM图形交易（亚洲）
(2019)

31.陈，陈，陈，陈，陈：在函数空间中学习纹理表示。输入：ICCV（2019）

32.公园，J。J.，弗罗伦斯，S.，S.，学习连续符号距离函数。输入：CVPR（2019）

33.张力伟：视图合成的软三维重建。ACM图形交易（亚洲）（2017）

34.波特，T.，杜夫，T.：合成数字图像。计算机图形学（签名图）（1984年）

35.Rahaman，N.，Baratin，A.，Arpit，D.，阿克斯勒博士，¨.，林，汉普雷希特，F。A.，本吉奥，Y.，考维尔， A.C.：关于神经网络的频谱偏差。输入：ICML（2018）

36.王，陈，王，王，陈，王：btf的统一神经编码。计算机图形学论坛（欧洲图形学）（2020年）

37.雷纳，G.，雅各布，W.，高希，A.，韦里奇，T.：神经BTF压缩和插值。计算机图形学论坛（欧洲图形学）（2019年）

38.任、P、王、J、龚、M，林、S、童、X、郭、B.：具有辐射回归函数的全局照明。ACM图形事务处理（2013年）

39.Sch¨onberger，J.L.，弗拉姆，J。M.：结构。输入：CVPR（2016）

40.Seitz。M.，戴尔，C。R.：通过体素着色重建逼真的场景。国际计算机视觉杂志（1999）

41.西茨曼，¨.，蒂斯，¨.，海德，¨.，尼.，韦茨斯坦，¨，¨：深度体素：学习持久3D特征嵌入。输入：CVPR（2019）

42.场景表示网络：连续的3d结构感知神经场景表示。在： NeurIPS（2019）

43. 斯里尼瓦桑，P。P.，塔克，R.，巴伦，J。T.，T.，T.，T.，T.。：用多平面图像推动视图外推的边界。在： CVPR (2019)
44. 斯坦利，K。O.：构成模式产生网络：一种新的发展抽象。基因编程和可进化的机器（2007）
45. 具有透明度和垫子的立体声匹配。在：ICCV (1998)
46. 图尔西亚尼，S.，周，T.，埃弗罗斯，A。A.，J.：通过可分散射线一致性的单视图重建的多视图监督。输入：CVPR（2017）
47. Vaswani，A.，谢泽尔，帕马尔，谢泽尔，尤斯科雷特，琼斯，戈麦斯， A.N.，Kaiser，L-。我：你所需要的就是注意力。输入：NeurIPS（2017）
48. 韦西特，新，戈西尔：让有颜色吧！三维重建的大规模纹理化。输入：ECCV（2014）
49. 木材，D。N.，AzumaD。I.，阿尔丁格，K.，无曲线， B.，杜尚，t， D.H.，斯图茨尔，W：3D摄影。in：签名（2000）
50. 张，R.，伊索拉，P.，Efros，A。A.，深度特征作为知觉度量的不合理有效性。输入：CVPR（2018）
51. 钟，E。D.，贝普勒，T.，戴维斯，J。H.，Berger，B.：从低温电子显微镜图像中重建三维蛋白质结构的连续分布。在： ICLR（2020）
52. 周，伟，伟，伟，伟伟：立体放大：使用多平面图像学习视图合成。ACM图形交易（签名图）（2018年）

一个附加的实施细节

网络架构图。7详细介绍了我们简单的全连接架构。

我们的方法通过查询沿着摄像机射线的连续5D坐标下的神经辐射场表示来呈现视图。对于合成图像的实验，我们放大场景，使其位于以原点为中心的边长为2的立方体内，并且只查询这个边界体内的表示。我们的真实图像数据集包含的内容可以存在于最近点和近点之间的任何位置，所以我们使用标准化的设备坐标将这些点的深度范围映射到[-1；1]。这将所有的射线原点转移到场景的近平面，将相机的透视射线映射到转换体积中的平行射线，并使用视差（逆深度）而不是度量深度，因此现在所有坐标都是有界的。

对于真实场景数据，我们通过在优化过程中向输出a值（通过ReLU之前）添加零均值和单位方差的随机高斯噪声来规范我们的网络，这就略微提高了呈现新视图的视觉性能。我们在张量低的[1]中实现了我们的模型。

为了在测试时渲染新的视图，我们通过粗网络每条射线采样64点，通过ine网络每条射线采样64 + 128 = 192点，每条射线总共有256个网络查询。我们现实的合成

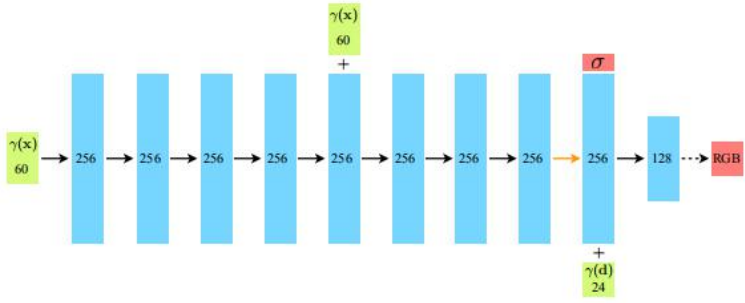


图7：这是对我们的全连接网络架构的可视化。输入向量用绿色表示，中间隐藏层用蓝色表示，输出向量用红色表示，每个块内的数字表示向量的维数。所有的层都是标准的全连接层，黑色箭头表示ReLU激活的层，橙色箭头表示没有激活的层，虚线黑色箭头表示s型激活的层，“+”表示向量连接。输入位置（V（x））的位置编码通过8个全连接的ReLU层，每个层有256个通道。我们遵循DeepSDF [32]架构，并包括一个跳过连接，它将此输入连接到第5层的激活中。另一个层输出体积密度a（使用ReLU进行整流，以确保输出的体积密度是非负的）和一个256维的特征向量。这个特征向量与输入查看方向的位置编码（V（d））连接起来，并由一个额外的具有128个通道的全连接ReLU层进行处理。一层（s型激活）输出位置x输出发射的RGB辐射，由方向为d的射线观察。

数据集需要每幅图像64万条光线，而我们的真实场景每幅图像需要762万光线，导致每个渲染图像有1.5亿到2亿个网络查询。在NVIDIA V100上，每帧大约需要30秒。

B附加的基线方法细节

神经卷（NV）[24]我们使用由作者在https://github上开源的NV代码。并遵循他们的程序，在单一场景中进行训练。

场景表示网络（SRN）[42]我们使用由作者在https://github上开源的SRN代码。并按照他们的程序对单个场景进行训练。

局部光场融合（LLFF）[28]我们使用了由作者在https://github上开源的预先训练好的LLFF模型。com/Fyusion/LLFF.

作者发表的SRN实现需要大量的GPU内存，并且即使在4 NVIDIA V100GPU上并行化，图像分辨率也被限制在512 512像素。×我们为合成数据集计算512 512像素的SRN定量指标，为真实数据集计算504 376像素的SRN，而可以在更高分辨率下运行的其他方法分别为800 800和1008 752。××××

C NDC射线空间的推导

我们在归一化的设备坐标（NDC）空间中使用“正面”捕获来重建真实的场景，这通常作为三角形栅格化管道的一部分。这个空间很方便，因为它在将z轴（相机轴）转换为线性差异时保留了平行线。

在这里，我们推导了应用于射线从相机空间映射到NDC空间的变换。齐次坐标的标准三维透视投影矩阵为：

$$M = \begin{pmatrix} \frac{n}{r} & 0 & 0 & 0 \\ 0 & \frac{n}{t} & 0 & 0 \\ 0 & 0 & \frac{-(f+n)}{f-n} & \frac{-2fn}{f-n} \\ 0 & 0 & -1 & 0 \end{pmatrix} \tag{7}$$

其中，n，f是近、远的剪切平面，r和t是近剪切平面上场景的右界和上界。（请注意，这是在惯例中，照相机是在看-z的方向。）投影一个齐次点（x、y、z、1）]，我们向左乘以M，再除以第四个坐标：

$$\begin{pmatrix} \frac{n}{r} & 0 & 0 & 0 \\ 0 & \frac{n}{t} & 0 & 0 \\ 0 & 0 & \frac{-(f+n)}{f-n} & \frac{-2fn}{f-n} \\ 0 & 0 & -1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{n}{r}x \\ \frac{n}{t}y \\ \frac{-(f+n)}{f-n}z - \frac{-2fn}{f-n} \\ -z \end{pmatrix} \tag{8}$$

$$\text{project} \rightarrow \begin{pmatrix} \frac{n}{r}\frac{x}{-z} \\ \frac{n}{t}\frac{y}{-z} \\ \frac{(f+n)}{f-n} - \frac{2fn}{f-n}\frac{1}{-z} \end{pmatrix} \tag{9}$$

投影点现在处于归一化的设备坐标（NDC）空间中，在那里，原始的查看结果已经被映射到立方体$[-1,1]^3$.

我们的目标是取射线o++并计算射线原点\方向d\在NDC空间中，对于每个t，都存在一个新的t\其中（o + td）= oπ\+ t**d**\π（其中是使用上述矩阵进行的投影）。换句话说，原始射线和NDC空间射线的投影可以追踪出相同的点（但不一定以相同的速率）。

让我们从方程式中重写投影点。9作为(a$_X$北/z， a$_Y$y/z， a$_Z$+ b$_Z$/z)$^\intercal$.新的起源的组成部分o\方向d\必须满足：

$$\begin{pmatrix} a_x \frac{o_x+td_x}{o_z+td_z} \\ a_y \frac{o_y+td_y}{o_z+td_z} \\ a_z + \frac{b_z}{o_z+td_z} \end{pmatrix} = \begin{pmatrix} o'_x + t'd'_x \\ o'_y + t'd'_y \\ o'_z + t'd'_z \end{pmatrix}. \tag{10}$$

为了消除一定程度的自由度，我们决定用t\=0和t = 0应该映射到同一点。替换t = 0和t\= 0 Eqn.10直接给出了我们的NDC空间原点o\:

$$\mathbf{o}' = \begin{pmatrix} o'_x \\ o'_y \\ o'_z \end{pmatrix} = \begin{pmatrix} a_x \frac{o_x}{o_z} \\ a_y \frac{o_y}{o_z} \\ a_z + \frac{b_z}{o_z} \end{pmatrix} = \pi(\mathbf{o}). \tag{11}$$

这正是原始射线原点的投影T (o)。通过把它代回Eqn。10对于任意的t，我们可以确定t的值\和d\:

$$\begin{pmatrix} t'd'_x \\ t'd'_y \\ t'd'_z \end{pmatrix} = \begin{pmatrix} a_x \frac{o_x+td_x}{o_z+td_z} - a_x \frac{o_x}{o_z} \\ a_y \frac{o_y+td_y}{o_z+td_z} - a_y \frac{o_y}{o_z} \\ a_z + \frac{b_z}{o_z+td_z} - a_z - \frac{b_z}{o_z} \end{pmatrix} \tag{12}$$

$$= \begin{pmatrix} a_x \frac{o_z(o_x+td_x)-o_x(o_z+td_z)}{(o_z+td_z)o_z} \\ a_y \frac{o_z(o_y+td_y)-o_y(o_z+td_z)}{(o_z+td_z)o_z} \\ b_z \frac{o_z-(o_z+td_z)}{(o_z+td_z)o_z} \end{pmatrix} \tag{13}$$

$$= \begin{pmatrix} a_x \frac{td_z}{o_z+td_z}\left(\frac{d_x}{d_z} - \frac{o_x}{o_z}\right) \\ a_y \frac{td_z}{o_z+td_z}\left(\frac{d_y}{d_z} - \frac{o_y}{o_z}\right) \\ -b_z \frac{td_z}{o_z+td_z}\frac{1}{o_z} \end{pmatrix} \tag{14}$$

分解一个只依赖于t的共同表达式给我们：

$$t' = \frac{td_z}{o_z+td_z} = 1 - \frac{o_z}{o_z+td_z} \tag{15}$$

$$\mathbf{d}' = \begin{pmatrix} a_x\left(\frac{d_x}{d_z} - \frac{o_x}{o_z}\right) \\ a_y\left(\frac{d_y}{d_z} - \frac{o_y}{o_z}\right) \\ -b_z \frac{1}{o_z} \end{pmatrix}. \tag{16}$$

请注意，根据需要，t\=0当t = 0时。此外，我们还看到了t\!1作为t！1．回到原始的投影矩阵，我们的常数是：

$$a_x = -\frac{n}{r} \tag{17}$$

$$a_y = -\frac{n}{t} \tag{18}$$

$$a_z = \frac{f+n}{f-n} \tag{19}$$

$$b_z = \frac{2fn}{f-n} \tag{20}$$

使用标准的针孔照相机模型，我们可以重新参数化为：

$$a_x = -\frac{f_{cam}}{W/2} \tag{21}$$

$$a_y = -\frac{f_{cam}}{H/2} \tag{22}$$

其中W和H是图像的像素和f_{cam}是照相机的焦距。

在我们真实的正面捕捉中，我们假设远场景的边界是初始的(这花费我们很少，因为NDC使用z维来表示逆深度，i。e.，不同在这个限制下，z常数简化为：

$$a_z = 1 \tag{23}$$

$$b_z = 2n\,. \tag{24}$$

将所有内容结合在一起：

$$\mathbf{o}' = \begin{pmatrix} -\frac{f_{cam}}{W/2}\frac{o_x}{o_z} \\ -\frac{f_{cam}}{H/2}\frac{o_y}{o_z} \\ 1 + \frac{2n}{o_z} \end{pmatrix} \tag{25}$$

$$\mathbf{d}' = \begin{pmatrix} -\frac{f_{cam}}{W/2}\left(\frac{d_x}{d_z} - \frac{o_x}{o_z}\right) \\ -\frac{f_{cam}}{H/2}\left(\frac{d_y}{d_z} - \frac{o_y}{o_z}\right) \\ -2n\frac{1}{o_z} \end{pmatrix}. \tag{26}$$

在我们的实现中有一个重要的细节：我们通过o将o转移到z = -n与近平面的交点（在这个NDC转换之前）n= o + tnd表示tn= -(n+ oz)/dz.一旦我们转换为NDC射线，这就允许我们简单地采样t\从0到1线性，以便在原始空间中得到n到1的线性采样。

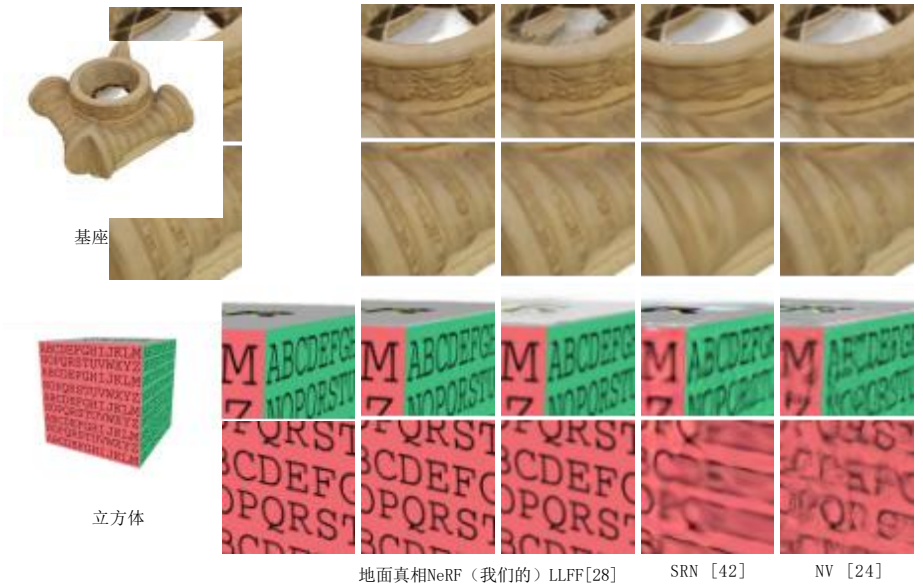


图8：对来自DeepVoxels [41]合成数据集的场景的测试集视图的比较。这个数据集中的对象具有简单的几何形状和完美的折折率。由于大量的输入图像（479个视图）和渲染对象的简单性，我们的方法和LLFF [28]在这个数据上执行得几乎很完美。当在其3D卷之间进行插值时，LLFF仍然偶尔会出现工件，如在每个对象的顶部插入中。SRN [42]和NV [24]没有呈现这些细节的代表性能力。

D其他结果

每个场景的分解表3、4、5和6包括将主要论文中提出的定量结果分解为每个场景的指标。每个场景的细分与本文中提出的总体定量指标是一致的，其中我们的方法在定量上优于所有基线。尽管LLFF实现了稍微更好的LPIPS指标，但我们敦促读者查看我们的补充视频，其中我们的方法获得了更好的多视图一致性，并比所有基线产生更少的工件。

	PSNR“基座			SSIM“基座			LPIPSt基座		
	主席	立方体	瓦斯	主席	立方体	瓦斯	主席	立方体	瓦斯
DeepVoxels [41]	33:45	32:35 28:42	27:99	0:99	0:97 0:97	0:96	–	——	–
SRN [42]	36:67	35:91 28:74	31:46	0:982	0:957 0:944	0:969	0:093	0:081 0:074	0:044
NV [24]	35:15	36:47 26:48	20:39	0:980	0:963 0:916	0:857	0:096	0:069 0:113	0:117
LLFF [28]	36:11	35:87 32:58	32:97	**0:992**	0:983 0:983	0:983	0:051	0:039 0:064	0:039
我们的	**42:65**	**41:44** **39:19**	**37:32**	0:991	**0:986** **0:996**	**0:992**	**0:047**	**0:024** **0:006**	**0:017**

表3：来自DeepVoxels [41]数据集的每个场景的定量结果。这个数据集中的“场景”都是具有简单几何图形的废弃对象，由由3D扫描器捕获的纹理映射网格渲染。深度体素方法的指标直接取自他们的论文，该论文没有报告LPIPS，只报告了SSIM的两个重要指标。

	信号-噪音功率比							
	主席	鼓	无花果属植物	热狗	垒高拼装玩具	材料	麦克风	船
SRN [42]	26 96 :	17:18	20:73	26:81	20:85	18:09	26:85	20:60
NV [24]	28 33 :	22:58	24:79	30:71	26:08	24:22	27:78	23:93
LLFF [28]	28 72 :	21:13	21:79	31:41	24:54	20:72	27:48	23:22
我们的	**33 00** **:**	**25:01**	**30:13**	**36:18**	**32:54**	**29:62**	**32:91**	**28:65**

	ssim”							
	主席	鼓	无花果属植物	热狗	垒高拼装玩具	材料	麦克风	船
SRN [42]	0:910	0:766	0:849	0:923	0:809	0:808	0:947	0:757
NV [24]	0:916	0:873	0:910	0:944	0:880	0:888	0:946	0:784
LLFF [28]	0:948	0:890	0:896	0:965	0:911	0:890	0:964	0:823
我们的	**0:967**	**0:925**	**0:964**	**0:974**	**0:961**	**0:949**	**0:980**	**0:856**

	LPIPSt							
	主席	鼓	无花果属植物	热狗	垒高拼装玩具	材料	麦克风	船

SRN [42]	0:106	0:267	0:149	0:100	0:200	0:174	0:063	0:299
NV [24]	0:109	0:214	0:162	0:109	0:175	0:130	0:107	0:276
LLFF [28]	0:064	0:126	0:130	**0:061**	0:110	0:117	0:084	0:218
我们的	**0:046**	**0:091**	**0:044**	0:121	**0:050**	**0:063**	**0:028**	**0:206**

表4：来自我们的现实合成数据集的每个场景的定量结果。这个数据集中的“场景”都是具有更复杂的几何测量和非兰伯材料的物体，使用搅拌机的循环追踪器渲染。

信号-噪音功率比

	房间	羊齿植物	树叶	堡垒	兰花	花	霸王龙	角
SRN [42]	27:29	21:37	18:24	26:63	17:37	24:63	22:87	24:33
LLFF [28]	28:42	22:85	19:52	29:40	18:52	25:46	24:15	24:70
我们的	**32:70**	**25:17**	**20:92**	**31:16**	**20:36**	**27:40**	**26:80**	**27:45**

ssim”

	房间	羊齿植物	树叶	堡垒	兰花	花	霸王龙	角
SRN [42]	0:883	0:611	0:520	0:641	0:449	0:738	0:761	0:742
LLFF [28]	0:932	0:753	**0:697**	0:872	0:588	**0:844**	0:857	**0:840**
我们的	**0:948**	**0:792**	0:690	**0:881**	**0:641**	0:827	**0:880**	0:828

LPIPSt

	房间	羊齿植物	树叶	堡垒	兰花	花	霸王龙	角
SRN [42]	0:240	0:459	0:440	0:453	0:467	0:288	0:298	0:376
LLFF [28]	**0:155**	**0:247**	**0:216**	0:173	**0:313**	**0:174**	**0:222**	**0:193**
我们的	0:178	0:280	0:316	**0:171**	0:321	0:219	0:249	0:268

表5：来自我们的真实图像数据集的每个场景的定量结果。这个数据集中的场景都是用一个前置的手持手机捕获的。

NeRF：将场景表示为神经辐射场，以进行视图合成25

	PSNR“热						
	主席	鼓	无花果属植物	狗乐高	材料	麦克风	船
1)无PE、VD、H	28:44	23:11	25:17	32:24 26:38	24:69	28:16	25:12
2)没有职位。编码	30:33	24:54	29:32	33:16 27:75	27:79	30:76	26:55
3)无视图依赖性	30:06	23:41	25:91	32:65 29:93	24:96	28:62	25:72
4)无层次结构	31:32	24:55	29:25	35:24 31:42	29:22	31:74	27:73
5)图像少得多 6)图像较少	30:92 :19 32 :	22:62 23:70	24:3 9 27:4 5	32:77 27:97 34:91 31:53	26:55 28:54	30:4 7 32:3 3	26:5 7 27:6 7
7)频率较低 8)更多频率	32:19 :87 32 :	25:29 24:65	30:73 29:92	36:0 6 35:7 8	30:7 7 32:5 0	**29:77** 29:54	31:6 6 32:8 6	28:2 6 28:3 4
9)完整模型	**33:00 :**	25:01	30:13	**36:18 32:54**		29:62	**32:91**	**28:65**

	SSIM“热						
	主席	鼓	无花果属植物	狗乐高	材料	麦克风	船
1)无PE、VD、H	0:919	0:896	0:926	0:955 0:882	0:905	0:955	0:810
2)没有职位。编码	0:938	0:918	0:953	0:956 0:903	0:933	0:968	0:824
3)无视图依赖性	0:948	0:906	0:938	0:961 0:947	0:912	0:962	0:828
4)无层次结构	0:951	0:914	0:956	0:969 0:951	0:944	0:973	0:844
5)图像少得多 6)图像较少	0:95 6 0:96 3	0:895 0:911	0:92 2 0:94 8	0:966 0:930 0:971 0:957	0:925 0:941	0:97 2 0:97 9	0:83 2 0:84 7
7)频率较低 8)更多频率	0:95 9 0:96 7	0:928 0:921	0:965 0:962	0:972 0:947 0:973 0:961	**0:952** 0:948	0:973 0:980	0:85 3 0:85 3
9)完整模型	**0:967**	0:925	0:964	**0:974 0:961**	0:949	0:980	**0:856**

	LPIPSt热						
	主席	鼓	无花果属植物	狗乐高	材料	麦克风	船
1)无PE、VD、H	0:095	0:168	0:084	0:104 0:178	0:111	0:084	0:261
2)没有职位。编码	0:076	0:104	0:050	0:124	0:079	0:041	0:261

				0:128			
3)无视图依赖性	0:075	0:148	0:113	0:112	0:102	0:073	0:220
				0:088			
4)无层次结构	0:065	0:177	0:056	0:130	0:080	0:039	0:249
				0:072			
5)图像少得多	0:05	0:173	0:08	0:123	0:079	0:03	0:22
6)图像较少	8	0:166	2	0:081	0:068	5	9
	0:05		0:05	0:121		0:02	0:22
	1		7	0:055		9	3
7)频率较低	0:05	0:143	0:038	0:087	**0:060**	0:029	0:21
8)更多频率	5	0:158	0:045	0:071	0:064	0:027	9
	0:04			0:116			0:26
	7			0:050			1
9)完整模型	**0:046**	**0:091**	0:044	**0:121**	0:063	0:028	**0:206**
				0:050			

表6：来自我们的消融研究的每个场景的定量结果。这里使用的场景与表4中使用的场景相同。