arXiv:2303.04322v2 [cs.RO] 6 Aug 2023

DroNeRF：用于计算神经辐射场的实时多智能体无人机位姿优化

Dipam Patel、Phu Pham 和 Aniket Bera 美国普渡大学计
算机科学系 补充版本：ideas.cs.purdue.edu/research/dronerf

摘要——**我们提出了一种名为*DroNeRF* 的新型优化算法，仅使用少量图像即可实现围绕物体的单目相机无人机自主定位，从而进行实时三维重建。神经辐射场（NeRF）是一种新颖的视图合成技术，可根据一组输入图像生成物体或场景的新视角。将无人机与 NeRF 结合，可以为生成场景的新视角提供一种独特且动态的方式，尤其适用于移动受限、场景能力有限的情况。我们的方法专注于计算各架无人机的优化位姿，仅依赖物体几何结构，而不使用任何外部定位系统。数据采集阶段独特的相机定位方式会显著影响三维模型的质量。为评估所生成新视角的质量，我们计算了多种感知指标，例如峰值信噪比（PSNR）和结构相似性指数（SSIM）。我们的工作展示了：在无人机机动能力有限的情况下，对多架无人机进行最佳布置，有助于生成感知质量更佳的结果。**

一、引言

相机的位置对于捕捉任何场景都至关重要，因为它会直接影响视觉系统解读视觉数据的性能。尤其是在三维重建这类任务中，所研究对象的相机位姿和曝光情况会显著影响最终重建模型的质量。当系统必须应对随机的相机运动——这种运动不同寻常，偏离日常场景——便可能导致重建效果不佳甚至失败。此时，场景中不可预测的变化会增加重建难度 [1]。对于涉及复杂场景理解的应用，使用无人机进行新视角合成，可以为生成场景或物体的新视角提供一种独特且动态的方式，从而为虚拟现实、游戏和娱乐等领域开辟新的应用可能。NeRF 模型的质量在很大程度上取决于图像的相机位置（尤其是图像数量较少，即少于十张时）。本文聚焦于单目相机无人机群的特定摄像机布置问题。我们的目标是根据物体的几何形状，为每架无人机确定独立的最优位姿，以尽可能少的图像完成拍摄。

我们的算法优先覆盖目标的关键区域，因为这会让重建结果极具实用价值。例如，在拍摄一个人进食的图像时，人的脸部、双手的位置以及桌上的食物，比桌子的其他部分或人体的其他部位更重要。尤其对于凹面物体，为了尽可能充分地捕捉其几何形状，必须将更多关注点放在靠近凹陷处的位置。本文展示了目标的朝向、形状和尺寸如何导致无人机围绕目标调整位姿 [2]。

另一方面，神经辐射场（NeRF） [3]最近实现了基于二维

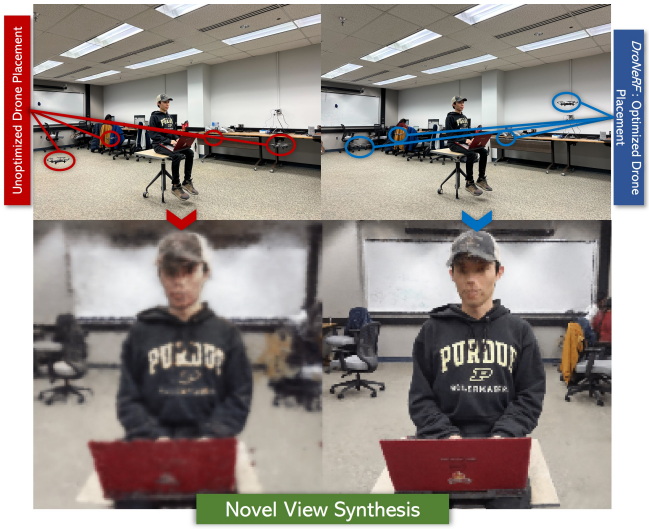


图 1：不同飞行配置下三维重建结果的比较——左半部分展示了无人机按照传统圆形轨迹飞行，且不优先考虑物体几何形状的情况。右半部分展示了无人机经过优化后的位置，重点关注人物的上半身。因此，正对人物的无人机比其余无人机飞得更高。

图像的新型三维重建，可生成高度逼真且准确的结果。它适用于渲染真实物体的高分辨率照片级真实感新视角。此类三维重建通常使用昂贵的设备，如 RGB-D相机或激光雷达传感器。然而，前者在直射阳光下可能会失效，而后者则更笨重且成本更高。一种简单、廉价且轻便的替代方案是使用单目摄像机，这将是更加便捷的解决方案。

M¨uller 等人 [4] 仅凭带位姿图像即可实时拟合辐射场。Chng 等人 [5]和 Lin 等人 [6]指出，只要初始估计足够准确，就不一定需要真实位姿。他们利用连续图像流构建准确的辐射场，无需将真实位姿或深度图作为输入。

在决定于现实世界中拍摄物体照片时，在计算机视觉领域的大多数最先进应用中， [1], 大多数相机都对称地布置在物体周围的虚拟半球上。这种相机分布通常效果良好，但前提是物体形状相对凸起且分布均匀。对于形状奇特、不对称或带有凹陷结构的物体，这种方法则会失效。

我们的方法是一种优化技术，能够迭代理解物体几何结构，并重新调整无人机/相机的位置，以捕捉更多网格细节（即让相机能够看到更多网格三角形）。

由于本研究并不侧重于在环境中自主寻找感兴趣的主体，因此这种初始放置有助于在寻找最佳相机位姿时，相较于常规位置优先考虑该主体。我们的方法受到 Kong 等人的区域检测策略的启发。 [7]此外，也受到 Chabra 等人对表面朝向的估计方法的启发。 [2]。图1直接比较了不同飞行配置及其对应的三维重建结果。

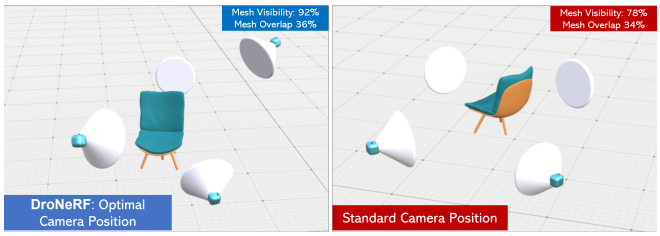


图 2： 用于计算网格可见性和网格重叠的相机模拟定位。与标准圆形模式位姿相比，*DroNeRF* 优化后的网格可见性高出近 *15%*，因为其优化依据是物体几何结构。

我们工作的主要贡献可概括如下：

- 我们设计了一个端到端流程，***DroNeRF***，利用多智能体无人机优化单目相机的位置，从而快速完成三维场景重建。
- 我们的工作能够根据无人机围绕物体采集并传输的图像构建可靠的辐射场，无需外部定位。
- ***DroNeRF*** 的表现优于采用未经优化的无人机位置的配置。我们比较了两种配置下不同物体的三维重建结果。

本文其余部分安排如下：第2节总结相关工作。第3节详细介绍系统概述和方法。第4节讨论评估指标、结果和分析。最后，第5节探讨未来方向与局限性，并总结我们的方法。

II. 相关工作

我们回顾了两条不同研究方向的相关文献：场景的最优摄像机布置，以及基于单目图像的神经辐射场，以构建融合这两者的算法。

A. 相机位姿优化

在场景周围布置多台相机以捕捉更大范围的视野，其效果很容易优于单相机系统。Tarabanis 等人 [8]，Mavrinac 等人 [9]，Zhao 等人 [10],和 Fu 等人 [11] 从视频监控的角度阐述了在三维空间中布置不同相机以进行二维图像分析的问题。Fu 等人 [11]还研究了在不同限制条件下实现最优摄像机布置的挑战。正如 Watras 等人 [12] 和 Zhao 等人 [10]所述，相机布置可以以使用尽可能少的相机覆盖场景为目标，也可以在相机数量固定的情况下，最大限度地扩大所捕捉的场景范围。本文尝试将这两种问题表述整合为一种。

Fleishman 等人 [13]在相机布置和覆盖范围问题上做出了重要贡献。为了使任何视觉系统达到最先进的性能，他们的工作强调相机视场、分辨率以及目标场景的体积。假设已预先知晓场景的几何形状，他们提出了一种相机定位架构，用于根据二维图像生成三维模型。他们还纳入了所有预期的相机位姿，以涵盖场景中的每一项特征。

正如 State 等人所述， [14],一种创新方法是在仿真中使用渲染后的相机，以了解每台相机的鲁棒性。他们还介绍了一款交互式模拟器，可在三维环境中设计、操控和测试不同的多相机配置。该模拟器在三维表面上实现了一种独特的纹理映射形式，以展示每台相机的覆盖范围 [2], [15]。此外， [16] 的工作通过可见性多边形优化了场景理解。

B. 神经辐射场

近年来，人们对各种场景表示进行了广泛研究，包括网格 [17], [18],、点云 [19],和隐式函数 [20], [21]。此外，许多研究提出了神经表示，旨在实现高质量渲染或自然信号表示。其中一个值得关注的模型是 NeRF [3],，它利用辐射场解决新视角合成问题，并实现了照片级真实感。尽管 NeRF 的渲染质量令人印象深刻，但由于渲染每个像素都需要进行大量计算开销高昂的多层感知机（MLP）求值，其渲染速度较慢。近期论文提出了混合表示，将快速的显式场景表示与可学习的神经网络组件相结合，相比纯隐式方法显著提升了速度。研究者探索了各种显式表示，包括稀疏体素 [22], [23], 、三平面 [24], [25], [26],和张量分解 [27]。

上述基于 NeRF 的方法的一个缺点是，它们依赖相机位姿，而相机位姿通常无法获取，因此无法用于训练模型。神经辐射场领域越来越多的研究旨在消除对不准确已知相机位姿的依赖。这种独立性对于创建 NeRF 尤具吸引力，因为它免去了获取图像相机位姿所需的劳动密集、耗时的预处理步骤 [28]。

在多智能体规划（用于群体导航）方面，已有大量相关研究 [29], [30], [31], [32],[33], [34], , 但考虑到计算开销，我们实现了 3D-RVO 的基线版本 [35]。我们的方法不假设特定的先验条件，可以从随机初始化的相机位姿开始优化。在传统的运动恢复结构或同步定位与建图中，通过优化目标函数，同时重建场景的结构和外观。受经典运动恢复结构中光束法平差联合优化方法的启发，我们提出了一种无需已知位姿相机的结构感知 NeRF 方法 [36]。

III. 概述与方法论

找到最优相机位置以高效进行三维重建，对计算机视觉和计算机图形学都很重要。目标是从不同视角拍摄图像，以重建场景的三维模型。我们的方法是构建一个优化问题，找出使特定代价函数最小化的相机位置。代价函数可基于多种因素，例如最终三维模型的质量、所需相机的数量以及相机之间的距离。基本思路

是构建一个优化问题，找出使特定代价函数最小化的相机位置。

正如 Holt 等人 [37]所述，此处采用的一种简单技术是：将无人机放置在以物体为中心、半径固定且任意的圆周上，作为起始位置。由于可根据所用无人机的数量将圆周等分，因此无人机能够均匀分布。不过，一个重要的考虑因素是尽量减少无人机数量。由于无人机彼此相邻放置，其各自的视场会相互重叠，因此应尽可能减少这种重叠。这样一组相机便能从各个方向覆盖物体特征，并实现最大覆盖范围。



图 3：每台相机单独拍摄物体（如左图所示），并在整个几何形状上计算出不同区域。随后，分别为这些区域拟合边界框（如中图所示）。最后，将来自各无人机视角的所有边界框合并，形成一个三维边界框（如右图所示）。

A. 概述

为了计算环绕物体的相机的最佳位置，系统必须即时理解场景。我们的目标是用尽可能少的图像实时构建物体的三维模型。只有当无人机优先捕捉物体的三维几何形状及其主要特征，而非帧中的噪声数据时，这种重建才切实可行。该流程有助于在最短时间内实现高效、可靠的重建，因为收集到的训练数据将包含最多的物体特征。图 2 清楚地展示了我们提出的配置以及传统的均匀布置方式。

B. 无人机配置计算

受 Chabra 等人工作的启发， [2], 我们采用离散优化技术，为给定场景寻找最优的无人机配置。我们将可能的无人机配置和位姿空间划分为一组 P 由 D个位姿组成，其中 D 是潜在无人机位姿的总数。我们的目标是确定一组经过优化的无人机配置O，其中 O 是 P的子集，且 K 表示使用的无人机数量。我们首先使用贪心算法初始化 O ，然后使用模拟退火（SA）对其进行优化。从贪心算法生成的初始配置集开始，模拟退火算法在每次迭代中通过调整 O 中相机的位姿，或从集合 P中选择新的相机位姿，来修改相机配置。如果新配置的适应度值F_{fit} 优于迄今发现的最佳配置，算法便接受该配置。我们根据 [2]计算适应度值。如果适应度值变差，模拟退火算法会以一定概率接受新配置，该概率取决于当前温度和玻尔兹曼选择[1]。为避免陷入局部极大值，SA 在运行初期以较高概率接受质量指标低于已知最佳配置的配置。随着算法推进，接受适应度值低于已知最佳配置的配置的概率会逐渐降低，直到算法只接受优于当前最佳配置的解。算法 1受 [2], 启发，给出了一个基本伪代码，用于通过适应度函数为一组无人机选择最优位姿。计算适应度函数时，需要输入相机 C、表面 S 、帧 F 以及 D 。 F_{fit}

算法1 多相机位姿优化的适应度函数

需要： C, S, D, F
确保： F_{fit}
  $F_{fit} = 0$
  对于 每个相机位姿 c（位于集合 C 中）**执行**
    对于 每一帧 f （位于集合 F 中）**执行**
      s_{near} = NULL
      $s_{nearDist}$ = ∞
      对于每个表面 s 属于集合 S **执行**
        d = 表面距离(c, s, D, f)
        如果 $d < s_{nearDist}$ **则**
          $s_{nearDist} = d$
          $s_{near} = s$
        结束 if
      结束 for
      如果 $s_{near} \neq$ NULL **则**
        $(x, y)_{expect}$ = 预期图像位置(c, s_{near})
        $(x, y)_{real}$ = 真实图像位置(c, s_{near}, D, f)
        F' = 差值$((x, y)_{expect}, (x, y)_{real})$
        $F_{fit} += F'$
      结束如果
    结束循环
  结束循环
  返回 F_{fit}

这里的问题在于，我们需要获取目标物体的几何形状，才能计算无人机的最佳位置。为此，我们将该物体粗略且满足最低需求的几何形状表示为一组三维包围盒。在以下小节中，我们将说明如何高效地计算这些几何形状。

C. 种子几何计算

为了计算场景的最低可行几何形状，我们从环绕物体的一组 4 架无人机获取多个快照。这些无人机彼此呈90度角排列。这样便可在两个正交轴上计算出包围物体的“区域”／边缘。得到这些边缘集合后，我们会创建多个三维包围盒，从两个轴的方向包围这些边缘。

我们将“区域”定义为一个区域，它可以概括为一组具有共同属性（如颜色或强度）的相邻像素，这些像素构成一个易于与周围环境区分的整体。与边缘或角点相比，区域提供了互补信息，这也构成了区域的关键特征。

高斯拉普拉斯核可表示为

$$\nabla^2 LoG = \frac{1}{\pi\sigma^4}\left(1 - \frac{x^2+y^2}{2\sigma^2}\right) e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (1)$$

其中 σ 为标准差，(x, y) 是图像 $I(x, y)$ 的像素坐标。

由 (1) 可知，LoG 取决于图像中的区域结构大小（x，y）以及用于平滑的高斯核 σ 的大小。随着核尺寸不断增大，表面会变得更加平滑，这意味着图像导数的幅值会变小。

根据 Lindeberg [38], , 导数会随 σ的增大而指数级衰减。为补偿这一点，LoG 需要用 σ^2进行归一化。此外，为了捕捉图像域中不同大小的区域，需要采用多尺度方法，自动根据 σ的不同取值调整选择。因此，尺度归一化拉普拉斯算子定义为

$$\nabla^2_{norm} LoG = \frac{1}{\pi\sigma^2}\left(1 - \frac{x^2+y^2}{2\sigma^2}\right) e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (2)$$

最后，根据 (2)，与图像 $I(x, y)$ 卷积后的归一化高斯拉普拉斯算子为

$$(\hat{x}, \hat{y}; \hat{\sigma}) = argmax_{(x,y;\sigma)}(\nabla^2_{norm} LoG * I(x, y)) \quad (3)$$

其中，$(\hat{}x, \hat{}y)$ 表示区域的位置，$\hat{}\sigma$ 表示区域的大小。

根据 Lindeberg [38], , 区域位于极大值处；这些点同时是 $\nabla^2_{norm} LoG$相对于空间和尺度的局部极大值。

D. 边界框估计

如 LoG所述，该方法可以识别不同大小的区域。然而，仅考虑最大的区域，因为假设从像素坐标来看，目标物体相较于背景具有更大的尺寸。获得适当估计的最简单方法，是使用 OpenCV 的 boundingRect 函数计算包围最大区域的边界框。该方法从区域中估计出轴对齐的边界框，而不是旋转边界框，因为无人机无法保持该位置，否则会在数据集中引入滚转和偏航误差。因此，所得边界框将占据图像中最大的区域之一。根据（3），从所有（x、y）位置中找出区域大小最大的区域σ

$$B_{box} = boundingRect(max(\hat{\sigma}_{(\hat{x},\hat{y})}))$$

其中 B_{box} 表示（x，y，w，h）为边界框的坐标，其提供了 C_{box}，即边界框的质心。图3 概述了如何从各个区域推导边界框。

质心 C_{box} 用于通过*PID* 控制器控制无人机的位置。由于无人机没有额外的传感器，也没有用于反馈的外部定位系统，因此这种控制策略属于开环系统。不过，根据物体几何形状进行误差估计，使其能够作为闭环系统运行。

关于*PID* 相对于 C_{box} 通过将无人机平移至目标位置来对其进行对齐。然而，无人机的朝向仍然不正确。无人机的偏航控制至关重要，否则采集的数据将完全错位，导致模型构建效果不佳。为了解决这一问题，受 Chabra 等人工作的启发 [2], 我们将边界框质心的朝向 C_{box} 相对于单目相机无人机C 设为

$$O(C, C_{box}) = \frac{\overrightarrow{C} - \overrightarrow{C_{box}}}{||\overrightarrow{C} - \overrightarrow{C_{box}}||}.\overrightarrow{C_{box_n}} \quad (4)$$

其中，$\overrightarrow{C_{box_n}}$ 垂直于边界框的质心，用于控制偏航误差估计。

使用 (4) 和 C_{box}，*PID* 输出会调整无人机的 R_i、P_i、T_i 和 Y_i （滚转、俯仰、油门和偏航），以减小质心 C_{box} 与图像帧中心（x_{cent}，y_{cent}）之间的误差。其中，R_i 和 T_i 会减小*X*和*Z* 方向上的误差，而 P_i 会减小*Y* 方向上的误差，从而最大化物体面积。同样，Y_i 会减小 θ 平面上的误差。系统会持续独立监测每架无人机的位置，以计算误差并实时更新 PID 输出，从而准确维持所需的设定点。

完成初始迭代后，无人机会按照预定义的圆形轨迹移动固定距离，从而产生位姿变化。此后，所提出的算法仅按照上述方式微调无人机位姿。因此，不再需要外部定位系统。整个过程重复 5 次，每个物体总共仅采集 20 张图像。

E. 场景重建

我们的目标是近乎实时地重建三维场景。为此，我们采用 Instant-NGP [4] 中的多分辨率哈希编码方法。

与 NeRF 一样，我们将连续场景表示为一个五维向量值函数，其输入变量由一个三维坐标（x，y，z）和一个二维观察方向（θ，ϕ）组成。输出变量包括发射颜色（r，g，b）和体密度 σ。在他们的原始研究中，Mildenhall 等人 [3]提出，使用位置编码对五维输入向量进行编码，有助于模型捕获包含高频变化的数据。

与 NeRF 不同，Instant-NGP 使用基于哈希的编码方案，从而实现更快的收敛。其思路是，给定一个神经网络 m，其中不仅包含可训练权重**w**，还包含用于对输入进行编码的可训练编码参数 θ **y** = encode(**x** θ)。这些参数被组织成 L 个层级，每个层级都关联着 T 个特征向量。

哈希编码过程可描述如下：

- 对于给定的输入坐标 **x**，在每个 L 分辨率层级上确定相邻体素，并使用哈希函数为其角点分配索引。
- 从哈希表 θ_l 中检索所有所得角点索引对应的 F 维特征向量。
- 根据 **x** 在相应l^{th} 体素内的相对位置，对特征向量进行线性插值。
- 将每个层级的结果与任何辅助输入 $\xi \in R^E$ （观察方向、材料参数等）拼接起来，生成编码后的输入向量 $\mathbf{y} \in R^{LF+E}$。
- 随后将该向量 **y** 输入多层感知机（MLP）以训练模型。与原始 NeRF 模型一样，输出将是发射颜色 **c** 和体密度 σ 。

网格分辨率采用等比数列，在最粗和最精细的分辨率之间进行选择。有关哈希编码技术的完整概述，可参阅 M¨uller 等人 [4] 的原始论文，了解更多详细说明。我们的处理流程采用了这种多分辨率哈希编码方法。

F. 实验

本研究使用的无人机是 Ryze Robotics Tello EDU 版本，这是一款小巧、轻便且价格低廉的可编程无人机。我们使用四架这样的无人机组成无人机群。Tello 配备一台 720p、500 万像素的相机，分辨率为 960 x 720，帧率为 30 FPS，飞行时间约为 13 分钟。无人机底部配有视觉定位系统（VPS），由一台相机和一个 3D 红外模块组成。该系统有助于保持无人机的当前位置，实现精确悬停。

Tello 拥有自己的软件开发工具包（SDK），其中包含用于访问飞行和配置设置的命令列表。这些无人机内置热点，手机或 PC 可以连接至该热点，并通过其应用程序或任何开源 Python 模块，借助 UDP 套接字通信协议发送飞行命令。

Tello EDU 有两种不同模式：

- **接入点模式：**无人机充当即时接入点，可与任何设备建立直接的一对一连接，以接收命令并将相机视图流传输至该设备。在此模式下，所有无人机均使用相同的默认 IP 地址——192.168.10.1。
- **站点模式：**站点模式是一种网状连接，多个无人机可同时连接到同一个接入点。因此，连接到同一接入点的中央 PC 可以同时向所有无人机发送命令，从而实现无人机群。

受硬件限制，站点模式不支持相机流传输，因为所有无人机默认都在相同端口地址上传输相机帧。由于它们都连接到同一网络，不同流之间会产生干扰。为解决此问题，我们在接入点模式下将所有无人机分别连接至不同的 USB WiFi 适配器。此外，所有 WiFi 适配器与之前一样连接到同一个 WiFi 网络，形成一个中央 PC 所连接的网状网络。该新设置既有利于组建无人机群，也支持各无人机单独进行相机流传输。如前所述，在起飞前，这些无人机从各个方向以对称配置布置在目标物体周围，以覆盖物体的所有表面。

我们所有实验均在 Ubuntu 22.04.2 系统上进行，并使用配备 12 GB GDDR6X 显存的 NVIDIA GeForce RTX 4090 作为中央 PC，用于接收各路独立的相机视频流、运行优化算法以及进行 NeRF 重建。我们提出的模块可实时处理四路分辨率为 960x720 的图像流。这种高效率得益于对各架无人机分别并行运行优化算法。每架无人机拍摄六张图像，因此根据分别优化后的无人机位姿，共使用 24 张图像进行重建。标准均匀无人机配置也采用相同步骤。

IV. 结果与分析

A. 评估指标

为了评估光度准确度，我们使用峰值信噪比（PSNR）和结构相似性指数（SSIM），评估通过传统的 360 度位姿以及我们方法优化后的位姿生成的 NeRF 渲染图像质量。

峰值信噪比衡量原始图像与重建图像之间的差异。它计算像素的最大可能值与原始图像和重建图像之间的均方误差之比。峰值信噪比越高，表示重建图像中的失真或噪声越低。

另一方面，SSIM 用于评估原始图像与重建图像之间的结构相似性。它考虑图像的亮度、对比度和结构，并计算一个介于 -1 到 1 之间的分数，其中 1 表示完全相似。与 PSNR 不同，SSIM 的设计更符合人类感知准确度，因为它考虑到人类视觉系统对结构信息变化的敏感度高于对像素值变化的敏感度。

B. 定性与定量结果

我们将本方法与一种基线无人机布置方法进行比较，该方法将无人机均匀布置在目标物体周围。我们在四个场景中开展实验：人、吉他、机器狗和回收箱。我们为每个场景重建两个 NeRF 模型：一个基于常规无人机布置，另一个基于我们的优化布置。图1 展示了我们的方法与基线方法的定性结果对比。左侧图像由采用常规无人机布置构建的 NeRF 模型渲染而成。右侧图像则由同一场景的 NeRF 模型渲染而成，但该模型是通过 DroNeRF 优化重建的。与基线方法构建的模型相比，我们的方法构建的 NeRF 模型质量更高。两者的区别在于，均匀的无人机布置没有考虑到有利的相机位置，导致信息丢失。DroNeRF 通过检测最大的感兴趣区域、计算相应的边界框，再据此将无人机调整至所需位置，从而解决这一问题。这样，无人机便能拍摄到包含中心物体重要细节的图像，进而构建出质量更高的 NeRF 模型。更多结果请访问我们的网站：ideas.cs.purdue.edu/research/dronerf。

表 I 比较了我们的方法与基线方法的定量结果。为计算峰值信噪比和 SSIM 分数，我们将无人机拍摄的图像分为训练集和测试集。训练图像用于优化 NeRF 模型。测试集包含具有新颖视角（或相机位姿）的未见图像。随后，我们使用这些图像以及优化后的 NeRF 模型在相同相机角度下生成的渲染图像，计算峰值信噪比和 SSIM 分数。

	PSNR		**SSIM**	
场景	标准	优化后	标准	优化后
人	11.30	18.93	0.43	0.66
吉他	8.89	17.67	0.29	0.56
机器狗	10.18	22.72	0.40	0.70
回收箱	9.26	20.13	0.34	0.67

表 I：不同场景下标准（均匀放置）和优化（*DroNeRF*）无人机位置的光度结果。

结果表明，对于全部四个场景，使用 DroNeRF 优化后的无人机位置所获得的峰值信噪比和 SSIM 分数均高于标准均匀布置。具体而言，对于 机器狗 场景，基线方法的峰值信噪比为 10.18，SSIM 为 0.40，而 DroNeRF 的峰值信噪比达到 22.72（提升 123%），SSIM 达到 0.70（提升 75%）。对于其他场景，与基线方法相比，DroNeRF 的峰值信噪比至少提升了 68%，SSIM 至少提升了 53%（对于 人场景）。总体而言，结果表明，使用 DroNeRF 优化后的无人机位置能够获得更高的图像

质量，适用于光度重建任务，相较于标准均匀布置表现更佳。不过需要注意的是，该评估仅在数量有限的场景上进行，因此还需要进一步评估，以验证这些发现的普适性。

V. 结论与局限性

我们提出了一种名为 DroNeRF 的新方法，用于优化无人机在光度三维重建任务中的布置。该方法考虑了从不同无人机视点观察场景表面时的遮挡和可见性，以提高重建场景的准确度。独特的相机位置显著影响三维模型的质量，尤其是对于非对称物体。通过优先考虑物体的几何形状，我们证明，与均匀布置无人机相比，DroNeRF 在光度准确度方面表现更佳，评估指标为峰值信噪比和 SSIM。此外，结果表明，该方法适用于不同场景，并能显著提升光度质量。总之，我们提出了一种利用无人机成像提高光度三维重建质量的有效方法，可应用于考古学、建筑学和环境监测等领域。

局限性：该算法仅依据物体几何形状来确定无人机的位置。这可能导致位姿估计不准确，尤其是在动态或不断变化的环境中。此外，该算法假设感兴趣物体处于静止状态，且不考虑运动物体。如果场景中存在物体运动，这可能导致三维模型出现噪声。这为未来的改进提供了机会。

参考文献

[1] J.-J. Gonzalez-Barbosa, T. Garcia-Ramirez, J. Salas, J.-B. Hurtado- Ramos 和 J.-d.-J. Rico-Jimenez, 《实现完全覆盖的最优相机布置》, 载于*2009*年*IEEE*国际机器人与自动化会议, 2009年。[2] R. Chabra、A. Ilie、N. Rewkowski、Y.-W. Cha 和 H. Fuchs, 《针对已知三维动态场景捕获优化商用深度相机的布置》, 载于*2017*年*IEEE*虚拟现实会议, 2017年。[3] B. Mildenhall、P. P. Srinivasan、M. Tancik、J. T. Barron、R. Ramamoor- thi 和 R. Ng, 《NeRF：将场景表示为用于视图合成的神经辐射场》, *CoRR*, 第abs/2003.08934卷, 2020年。[4] T. M¨uller、A. Evans、C. Schied 和 A. Keller, 《采用多分辨率哈希编码的即时神经图形基元》, 《*ACM*图形学汇刊》, 2022年。[5] S.-F. Chng、S. Ramasinghe、J. Sherrah 和 S. Lucey, 《GARF：用于高保真重建和位姿估计的高斯激活辐射场》, 2022年。[6] C. Lin、W. Ma、A. Torralba 和 S. Lucey, 《束调整神经辐射场》, *CoRR*, 第abs/2104.06405卷, 2021年。[7] H. Kong、H. C. Akakin 和 S. E. Sarma, 《用于斑点检测的广义高斯拉普拉斯滤波器及其应用》, *IEEE*控制论汇刊, 2013年。[8] K. Tarabanis、P. Allen 和 R. Tsai, 《计算机视觉中的传感器规划综述》, *IEEE*机器人与自动化汇刊, 1995年。[9] A. Mavrinac 和 X. Chen, 《相机网络中的覆盖建模：综述》, 国际计算机视觉杂志, 第101卷, 第205–226页, 2012年。[10] J. Zhao、D. Haws、R. Yoshida 和 S.-c. S. Cheung, 《求解最优摄像机布置问题的近似技术》, 载于*2011*年国际计算机视觉大会研讨会, 2011年。[11] Y.-G. Fu、J. Zhou 和 L. Deng, 《采用三维布设相机监控二维平面区域》, 《传感器》, 第14卷, 第2期, 第1988–2011页, 2014年。[12] A. J. Watras、J.-J. Kim、H. Liu、Y. H. Hu 和 H. Jiang, 《实现最大视场视频拼接的最优相机位姿与布置配置》, 《传感器》, 第18卷, 2018年。[13] S. Fleishman、D. Cohen-Or 和 D. Lischinski, 《用于基于图像建模的自动相机布置》, 载于论文集：第七届太平洋计算机图形学与应用会议（目录号*PR00293*）, 1999年。

[14] A. State、G. Welch 和 A. Ilie, 《面向基于图像的虚拟现实应用的交互式相机布置与可见性模拟器》, 载于《立体显示与虚拟现实系统 XIII》, 系列丛书：国际光学工程学会（SPIE）会议系列, 第6055卷, 2006年。[15] A. Bera、S. Kim 和 D. Manocha, 《面向数据驱动人群仿真与内容生成的在线参数学习》, 《计算机与图形学》, 第55卷, 第68–79页, 2016年。[16] N. L. Williams、A. Bera 和 D. Manocha, 《利用可见性多边形在静态与动态场景中实现重定向行走》, IEEE 可视化与计算机图形学汇刊, 第27卷, 第11期, 第4267–4277页, 2021年。[17] T. Groueix、M. Fisher、V. G. Kim、B. C. Russell 和 M. Aubry, 《Atlasnet：一种通过纸浆塑形学习三维表面生成的方法》, CoRR, 第abs/1802.05384卷, 2018年。[18] N. Wang、Y. Zhang、Z. Li、Y. Fu、W. Liu 和 Y. Jiang, 《Pixel2mesh：从单张RGB图像生成三维网格模型》, CoRR, 第abs/1804.01654卷, 2018年。[19] C. R. Qi、H. Su、K. Mo 和 L. J. Guibas, 《Pointnet：基于点集进行三维分类与分割的深度学习》, CoRR, 第abs/1612.00593卷, 2016年。[20] L. M. Mescheder、M. Oechsle、M. Niemeyer、S. Nowozin 和 A. Geiger, 《占据网络：在函数空间中学习三维重建》, CoRR, 第abs/1812.03828卷, 2018年。[21] S. Peng、M. Niemeyer、L. M. Mescheder、M. Pollefeys 和 A. Geiger, 《卷积占据网络》, CoRR, 第abs/2003 04618卷, 2020年。[22] A. Yu、S. Fridovich-Keil、M. Tancik、Q. Chen、B. Recht 和 A. Kanazawa, 《Plenoxels：无需神经网络的辐射场》, CoRR, 第abs/2112.05131卷, 2021年。[23] C. Sun、M. Sun 和 H. Chen, 《直接体素网格优化：辐射场重建的超快速收敛》, CoRR, 第abs/2111.11215卷, 2021年。[24] E. R. Chan、C. Z. Lin、M. A. Chan、K. Nagano、B. Pan、S. D. Mello、O. Gallo、L. J. Guibas、J. Tremblay、S. Khamis、T. Karras 和 G. Wetzstein, 《高效的几何感知三维生成对抗网络》, CoRR, 第abs/2112.07945卷, 2021年。[25] A. Cao 和 J. Johnson, 《Hexplane：一种用于动态场景的快速表示方法》, arXiv:2301.09632, 2023年。[26] S. Fridovich-Keil、G. Meanti、F. Warburg、B. Recht 和 A. Kanazawa, 《K-planes：空间、时间与外观中的显式辐射场》, 2023年。[27] A. Chen、Z. Xu、A. Geiger、J. Yu 和 H. Su, 《Tensorf：张量辐射场》, 载于欧洲计算机视觉会议, 2022年。[28] J. L. Sch¨onberger 和 J.-M. Frahm, 《运动恢复结构再探》, 载于2016年计算机视觉与模式识别会议, 2016年, 第4104–4113页。[29] X. Wang、R. Xian、T. Guan、C. M. de Melo、S. M. Nogar、A. Bera 和 D. Manocha, 《Aztr：结合自动变焦与时序推理的航拍视频动作识别》, arXiv预印本 arXiv:2303.01589, 2023年。[30] B. Xia、I. Mantegh 和 W.-F. Xie, 《由多智能体强化学习驱动的无人机导航与冲突消解智能方法》, 载于2023年国际无人机系统会议（ICUAS）。IEEE, 2023年, 第713–722页。[31] E. Cheung、A. Bera 和 D. Manocha, 《基于驾驶员行为分类的高效安全车辆导航》, 载于IEEE计算机视觉与模式识别会议研讨会论文集, 2018年, 第1024–1031页。[32] A. Bera、T. Randhavane、R. Prinja 和 D. Manocha, 《Sociosense：在考虑社会与心理约束的情况下于行人中进行机器人导航》, 载于2017年IEEE/RSJ智能机器人与系统国际会议（IROS）。IEEE, 2017年, 第7018–7025页。[33] E. Cheung、T. K. Wong、A. Bera、X. Wang 和 D. Manocha, 《Lcrowdv：生成用于基于模拟的人群行为学习的带标签视频》, 载于《计算机视觉——ECCV 2016研讨会：阿姆斯特丹，荷兰, 2016年10月8日至10日及15日至16日，论文集，第二部分14》。施普林格, 2016年, 第709–727页。[34] A. Bera、S. Kim 和 D. Manocha, 《面向数据驱动人群仿真的高效轨迹提取与参数学习》, 载于图形界面会议, 第70卷, 2015年。[35] J. Snape 和 D. Manocha, 《在三维工作空间中导航多架简易飞机》, 载于2010年IEEE机器人与自动化国际会议。IEEE, 2010年, 第3974–3980页。[36] S. Chen、Y. Zhang、Y. Xu 和 B. Zou, 《利用极线约束实现无需已知位姿相机的结构感知NeRF》, 2022年。[37] R. J. Holt、H. Man、R. Martini、I. Mukherjee、R. Netravali 和 J. Wang, 《边界监控最优相机布置结果综述》, 载于2007年数据挖掘、入侵检测、信息保障与数据网络安全会议, 2007年。[38] T. Lindeberg, 《通过自动尺度选择进行特征检测》, 国际计算机视觉期刊, 第30卷, 第77–116页, 1998年9月。