本文已被 IEEE ICRA 2026 会议接收发表 ©2026 IEEE。

arXiv:2603.01850v1 [cs.RO] 2 Mar 2026

Tiny-DroNeRF：搭载于支持联邦学习的纳米无人机上的微型神经辐射场

Ilenia Carboni[13*]、Elia Cereda[1*]、Lorenzo Lamberti[12*]、Daniele Malpetti[1],Francesco Conti[3],Daniele Palossi[12]

摘要——重量不到 30 克的纳米级空中机器人可利用其灵活性和小巧外形，自主探索工业检查、搜救任务等杂乱狭窄的环境。然而，体型微小也意味着其资源受到严格限制：功耗低于 100 mW 的微控制器单元（微控制器）性能最高仅为 ∼100 GOps/s，内存预算也远低于 100 MB。尽管面临这些严格限制，我们仍致力于在纳米无人机上实现复杂的视觉任务，例如密集的三维场景重建——这是一项关键的机器人任务，也是空间感知和运动规划等基本能力的基础。性能领先的三维重建方法采用神经辐射场（NeRF）模型，需要数 GB 内存和大量计算，通常由功耗达数百瓦的高端图形处理器提供。本文提出 Tiny-DroNeRF，这是一种基于 Instant-NGP 的轻量级神经辐射场模型，并针对在我们纳米无人机搭载的 GAP9 超低功耗（ULP）微控制器上运行进行了优化。随后，我们进一步利用协作式联邦学习方案增强 Tiny-DroNeRF，将模型训练分配给多架纳米无人机。实验结果表明，与 Instant-NGP 相比，Tiny-DroNeRF 的内存占用减少了 96%，而重建精度仅下降 5.7 dB。最后，我们的联邦学习方案使 Tiny-DroNeRF 能够利用原本无法存储在单架无人机内存中的大量数据进行训练，从而提高整体重建精度。最终，我们的工作首次将神经辐射场模型在超低功耗微控制器上的训练与纳米无人机上的联邦学习相结合。

补充视频材料

补充视频可在以下网址观看：
https://youtu.be/-frFPUBGa0c

I. 引言

自主飞行的微型纳米无人机重量仅数十克，直径不足 10 厘米，能够在危险的工业厂房 [11] 或灾后环境 [12]等杂乱狭窄的空间中行动。得益于其自主性，即所有处理均在机载完成，无需依赖外部基础设施，它们不受视距操作条件或与基站最大距离的限制。配备小型相机和机载智能系统的纳米无人机已在简单的巡检和检测任务中得到验证，包括异常检测 [13], 语义定位 [14],和目标检测 [15]。在集群配置下，其效能会进一步提升：多架无人机可同时覆盖更大的区域，或协同实现共同目标 [16], ，例如搜寻幸存者。然而，对于这类微型机器人而言，更复杂的自主工作负载仍然难以实现。主要原因是它们缺乏空间感知能力，这源于其可搭载的计算和感知资源有限。与千克级飞行无人机相比，纳米无人机可用的内存和计算能力大约少 1000×， [17], 传感器也较为简单，例如低分辨率单目黑白相机。

神经辐射场（NeRF） [1] 是稠密三维重建任务中成熟的范式。在纳米无人机上实现NeRF，将使其能够自主执行新视角合成 [1],同步定位与建图 [18], 3D表面重建 [19], 或三维语义场景重建 [20]。然而，当前NeRF算法所需的内存与计算量超出纳米无人机能力数个数量级 [5], ，原因在于其复杂性以及针对特定场景重新训练的需求。因此，它们通常仅限于在拥有数GB可用内存、功耗达数百瓦的图形处理器上离线运行。

在本工作中，我们在适用于微型化纳米无人机的超低功耗微控制器的有限资源约束下，研究使用神经辐射场进行新视角合成的任务。借助纳米无人机的微小尺寸，我们聚焦于集群作业：每架纳米无人机从不同的已知位姿拍摄目标物体的少量图像，并在机上训练各自独立的NeRF实例。随后，借助联邦学习方法 [21], ，我们迭代聚合各个NeRF实例，构建全局NeRF，从而生成比任何单架无人机所能提供的更为丰富的三维重建。总体而言，我们的工作实现了场景连续三维表示的协同构建，从而能够利用从互不重叠的无人机视点采集的图像，生成先前未见相机位姿下的照片级真实视图。

具体而言，本文的贡献如下：

- 一种轻量级神经辐射场算法，称为 *Tiny-DroNeRF*，基于最先进水平（SoA）的 Instant-NGP [5]。我们通过超参数优化，在三维重建质量（即峰值信噪比（PSNR））、内存和计算量之间寻求最佳权衡。
- 一种内存高效的 Tiny-DroNeRF 实现，面向搭载于纳米无人机上的 GreenWaves Technologies（GWT）GAP9 超低功耗多核微控制器。
- 对*i*）在标准合成数据集 [1] 以及一个新颖的真实世界纳米无人机数据集上的重建精度，以及*ii*）执行-的详细实验分析

本研究部分受 SNSF RoboMix2 项目（资助编号：10004854）支持。

* 表示贡献相同。[1]I. Carboni、E. Cereda、L. Lamberti、D. Malpetti 和 D. Palossi 任职于瑞士 USI-SUPSI 的 Dalle Molle 人工智能研究所。通讯作者：ilenia.carboni@supsi.ch。[2] L. Lamberti 和 D. Palossi 还任职于瑞士苏黎世联邦理工学院集成系统实验室。[3] I. Carboni 和 F. Conti 任职于意大利博洛尼亚大学电气、电子与信息工程系。

TAB表 I：Synthetic NeRF 360° 数据集上的新视角合成方法比较 [1].

方法	年份	神经网络	主要贡献	PSNR [dB]	SSIM [-]	参数量 [MB]	训练 time	功耗 类别 [W]	设备
神经辐射场 [1]	2021	✓	基于坐标的多层感知机	31.01	0.947	5	24 h	10^2	GPU
SNeRG [2]	2021	✓	神经网格	30.38	0.950	1772	15 h	10^2	GPU
PlenOctrees [3]	2021	✓	八叉树 +球谐函数	31.71	0.958	1930	15 h	10^2	GPU
Plenoxels [4]	2022	×	体素网格 + 无神经网络	31.71	0.958	778	11 分钟	10^2	GPU
Instant-NGP [5]	2022	✓	参数化哈希网格	**33.18**	0.959	46	5分钟	10^2	GPU
TensoRF [6]	2022	×	带张量分解的体素网格	33.14	**0.963**	72	17分钟	10^2	GPU
K-Planes [7]	2023	×	多平面	32.21	0.960	385	38 分钟	10^2	GPU
MobileNeRF [8]	2023	✓	光栅化网格（烘焙）	30.90	0.947	538	27 h	10^2	GPU
BiRF [9]	2023	✓	3D-2D 二值化特征哈希编码网格 32.03		0.954	0.5	>5 分钟	10^2	GPU
CNC [10]	2024	✓	基于上下文的压缩框架	33.19	0.964	0.418	>5 分钟	10^2	GPU
Tiny-DroNeRF	2025	✓	基于 Instant-NGP，专为微控制器定制	26.94	0.892	**0.284**	**28s**	10^2	GPU
（我们的方法）				26.94	0.892	**0.284**	97 分钟	$<$ **0.10**	**MCU**

时间以及在GAP9微控制器上的内存占用。

- 一种用于在无人机集群中对纳米无人机上的 Tiny-DroNeRF 进行分布式训练的联邦学习方案。

在我们的实验中，Tiny-DroNeRF 大幅降低了资源需求，同时保持了实用的重建质量。与 Instant-NGP 相比，[5]，我们的内存使用量减少了 96%（从 527 MB 降至 21.4 MB），而峰值信噪比仅损失 5.7 dB。我们在 73 ms 内完成一个完整的前向—反向传播步骤，也就是说，在单台纳米无人机上完成全部训练（10k 步）需要 97 分钟。借助高效的硬件感知算法，我们成功使工作负载受计算量限制。我们的实现展现出出色的内存局部性，每个计算步骤与片外存储器之间的数据传输量仅为 3.6 MB。在实地测试中，我们从纳米无人机机载的灰度图像重建真实场景，峰值信噪比最高达到 21.2。

此外，我们证明了所提出的联邦学习方案能够让每架仅携带有限本地图像的纳米无人机借助对等节点的知识，在不交换原始训练图像的情况下合成场景中未见的部分。与能够访问所有图像的集中式训练相比，这种联邦方法的峰值信噪比仅低 0.6 dB，而每次合并（每进行 1 k 本地训练步数后执行一次）的通信开销仅为 2.24 秒。据我们所知，这是首次展示专为计算能力极其受限的设备设计的设备端 NeRF，也是联邦学习方案在资源受限的纳米无人机集群中的首次应用。

II. 相关工作

A. 新视角合成方法

神经辐射场 [1]提出了一种将三维场景表示为连续函数的范式，从而实现了逼真的新视角合成和高保真场景重建。这种方法确立了隐式神经表示，其中多层感知机（MLP）编码一个连续的体积场景，可在任意三维坐标处进行查询，并使其成为一种广泛采用的技术。然而，原始 NeRF 的训练和渲染速度较慢，因此推动了旨在加速优化和推理的大量研究。

为解决这些局限，后续研究探索了显式场景参数化，以提高效率。例如，PlenOctrees [3] 预先计算一种稀疏八叉树结构，用于存储辐射场参数，从而实现快速查询；Plenoxels [4] 和 TensoRF [6] 则以体素网格或张量分解方案取代神经网络，以降低计算开销。Instant-NGP [5] 提出了多分辨率哈希网格位置编码，并将其集成到 CUDA/C++ 库中，以加速神经图形基元的训练与渲染，使其成为高效、高性能神经辐射场的事实标准。

互补研究探讨了在保持重建保真度的同时降低内存需求。BiRF [9] 和 CNC [10] 提出了紧凑的网络架构，以额外计算换取更小的内存占用。尽管它们更能适应内存受限的设备，但更复杂的编码策略所引入的额外计算会增加执行时间。虽然文献中已有在商用 GPU 上进行渲染的实例，例如 SNeRG [2]和 MobileNeRF [8],，但现有方法的训练仍需要强大的图形处理器，这对嵌入式机器人设备上基于神经辐射场的方法构成了重大障碍。表 I 总结了新视角合成方法，报告了它们的主要贡献、在 Synthetic NeRF 360° 数据集 [1], 上以峰值信噪比和结构相似性指数（结构相似性）衡量的平均性能、参数量、训练时间以及功耗。

B. 面向神经辐射场的联邦学习

联邦学习（FL） [21]是一种协作式机器学习框架，多个参与方在不共享原始数据的情况下共同训练一个共享模型。参与者不交换数据，而是将模型更新发送给协调器；协调器可以是专用服务器，也可以是同时承担协调职责的对等节点。训练以迭代方式进行，在本地训练阶段与

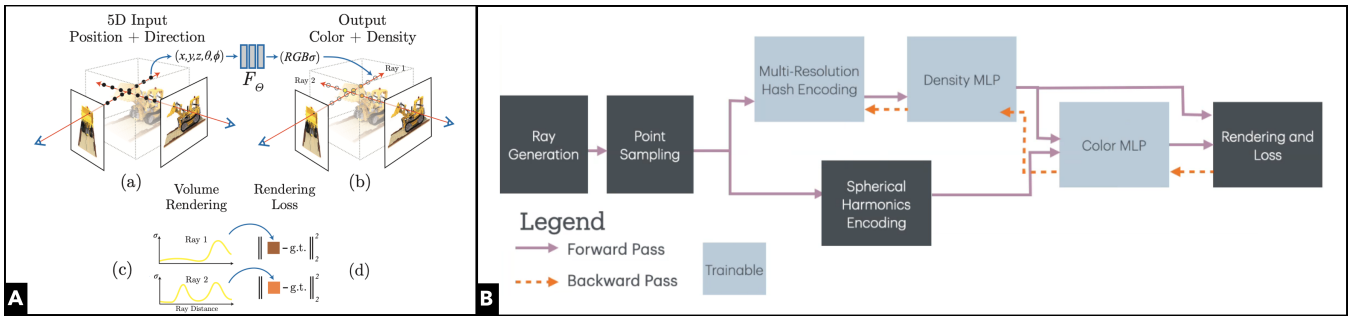


图 1：A）基于神经辐射场和体渲染的新视角合成 [1]。B）InstantNGP 算法 .

通信阶段之间交替进行：在本地训练阶段，每个参与者使用自己的私有数据更新模型；在通信阶段，协调器将收到的模型更新聚合为全局模型，并将其重新分发给所有参与者，以初始化下一轮训练。

首个 NeRF 联邦实现由 FedNeRF 提出 [22], , 随后有研究探索了用于大规模户外场景建模的基于联邦学习的方法 [23], 、结合联合位姿优化的分布式多机器人训练 [24],以及行星多智能体测绘 [25]。这些贡献仍处于探索阶段，尚无实地部署报告，也未进行详细的资源可行性分析。

在这些研究中，采用联邦学习的主要动机是扩展至更大规模的数据集 [23], [25], [26],、维持动态过程 [23], 以及提高通信效率 [24], [25]。相比之下，在本研究中，我们主要将联邦学习用作一种机制，使资源受限的无人机集群能够协同学习一个性能优于任何单架无人机独立学习所得模型的模型，从而为下游任务提供更强大的模型。

III. 背景

A. NeRF 概述

辐射场通过一个隐式函数表示场景，该函数将空间点 $\mathbf{x}$ 和视角方向 $\mathbf{d}$映射为体密度 σ 和发射颜色 $\mathbf{c}$:

$$f_\theta(\mathbf{x}) = (\mathbf{c}, \sigma) \tag{1}$$

神经辐射场 通过两个由 $\theta = \{\tau, \phi\}$ 参数化的多层感知机网络来近似该函数。图1所示的 Instant-NGP [5], 采用密度多层感知机，根据三维位置$\mathbf{x}$预测体密度 σ 以及隐藏特征向量$\mathbf{h}$:

$$f_\tau(\gamma(\mathbf{x})) = (\mathbf{h}, \sigma), \tag{2}$$

其中，三维位置通过位置编码 $\gamma(\mathbf{x})$表示。

颜色多层感知机根据隐藏特征和视角方向预测颜色 $\mathbf{c}$ ，

$$f_\phi(\mathbf{h}, \gamma(\mathbf{d})) = \mathbf{c}. \tag{3}$$

与三维位置一样，视角方向也通过 $\gamma(\mathbf{d})$编码。

Instant-NGP 的主要贡献是通过多分辨率哈希编码 [5] (MRHE)将三维位置映射为 $\mathbf{x}$ 高维参数化特征向量 $\gamma(\mathbf{x})$。这种可训练编码被组织为 L 个独立层级，每个层级都将空间细分为分辨率逐渐提高的体素网格。每个体素都映射到T 个特征向量的哈希表中，存储于 $\mathbb{R}^B$。给定位置 $\mathbf{x}$ 通过每一层的三线性插值编码为 $\gamma(\mathbf{x}) \in \mathbb{R}^{L\times F}$。随着更深层级的分辨率提高，多个体素会映射到固定大小哈希表中的同一个特征向量，但后续的密度多层感知机会学习消除由此产生的混叠。这一关键洞见使 T 成为可扩展的参数，我们可以对其进行扩展，以满足纳米无人机严格的内存要求。方向 $\mathbf{d}$也会进行编码，不过采用的是基于坐标的（即不可训练）编码函数$\gamma(\mathbf{d})$，即球谐函数。

训练期间，系统从训练集的随机图像中选取随机像素，沿着穿过这些像素的一定数量的射线 R进行光线步进，并沿射线采样点，从而生成由位置和视角方向组成的样本。将 B 个样本组成的批次输入多层感知机，得到样本颜色和体密度，再通过体渲染进行聚合，得到每条射线的颜色值 $\hat{\mathbf{c}}$。训练时使用 Adam 优化器，通过最小化射线颜色与真实像素颜色之间的 Huber 损失来优化模型。为减少在空白空间中的无效计算，Instant-NGP借助额外的数据结构占用网格，在光线步进过程中快速跳过空白区域。占用网格通过查询重建体积中每个网格单元内由密度多层感知机预测的密度，与网络训练交替更新。

B. 机器人平台

我们的机器人平台包括 Bitcraze Crazyflie 2.1 Brushless 四旋翼无人机（图 2-A）。这是一款市售纳米无人机，重量为 27 g，直径为 10cm。该无人机由一块 350 mAh 锂聚合物电池供电。我们通过开源*GAP9Shield* [27]，对该平台进行了扩展：这是一块可插拔的印刷电路板，集成了 GWT GAP9 片上系统（SoC）、32MB 外部 L3 HyperRAM、ublox NINA-W102 Wi-Fi 收发器（基于 ESP32，可提供高达 15Mbit/s 的带宽），以及一台低功耗微型摄像头。

图 2-B 所示的 GAP9片上系统分为两个功耗与频率域：片上系统域包含一个称为 Fabric Controller（片上控制器）（FC）的主核心；集群则包含 9 个通用核心、4 个浮点

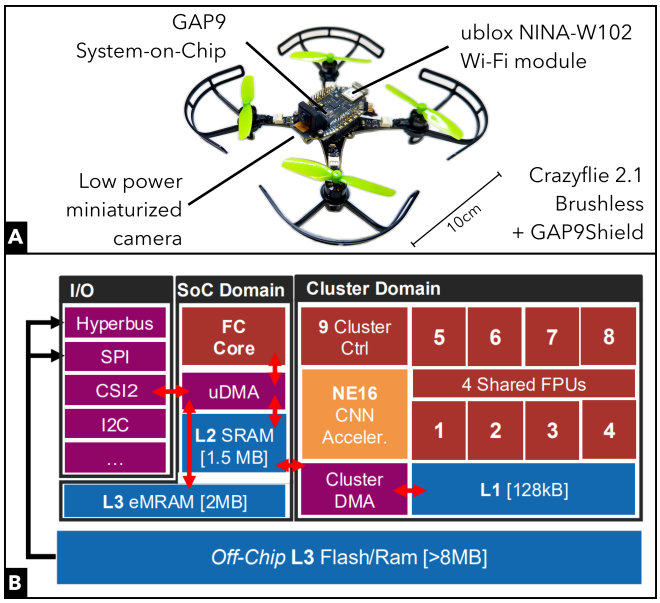


图 2：目标机器人平台。A）配备 GAP9Shield 扩展板的 Bitcraze Crazyflie 2.1 Brushless。B）GWT GAP9 片上系统架构。

运算单元（FPU），支持 16 位和 32 位数据类型，以及一个混合精度人工智能硬件加速器（NE16）。集群核心支持单指令多数据（SIMD）运算：9 个通用核心提供四通道 8 位整数单指令多数据运算，而 FPU 支持双通道 16 位单指令多数据运算，并可执行标量 32 位浮点运算。NE16 人工智能硬件加速器的峰值吞吐量为 150 整数 GOp/s。然而，该加速器专为采用整数数据类型的卷积神经网络前向传播而设计，而我们的工作负载需要反向传播和浮点运算。在所有实验中，我们均以峰值性能配置运行 GAP9，时钟频率为 370 MHz，供电电压为 0.8 V，对应的平均功耗为 60 mW。

该片上系统采用分层存储系统：集群共享 128 kB 的 L1 内存，飞控（FC）和集群均可访问 1.5 MB 的 L2 随机存取存储器。GAP9Shield 进一步通过 32 MB 的片外 L3 OctaSPI 随机存取存储器扩展了系统。集群核心访问 L1 约需 1 个周期，而直接访问 L2 则会额外产生约 100 个周期的延迟。两个专用的直接内存访问（DMA）引擎支持异步数据传输，使片上系统核心能够腾出时间执行计算：主 DMA 负责 L3–L2 传输（370 MB/s），集群 DMA 负责 L2–L1 传输（13.3 GB/s）。通过将 DMA 传输与计算重叠，可在后台将 L2 数据预取至 L1，使集群仅使用 L1 执行操作，并隐藏 L2 访问延迟。

IV. 方法

A. 硬件感知内存高效*NeRF*

由于我们的机器人平台（第 III-B 节）依赖于资源受限的微控制器，其片上和片外存储器容量都非常有限，因此必须大幅降低 SoA 新视角合成算法的内存和计算量需求。在基线配置下，Instant-NGP 每个训练步需要 527 MB 内存和 17.5 GOps。尤其是，内存占用包括模型可学习权重的存储，以及反向传播所需的中间张量。为减少峰值内存占用和每步运算量，我们对 Instant-NGP [5],进行了全面的超参数优化，目标包括*i*) 每个训练步的图像数量（img/step）、*ii*) 输入图像分辨率、*iii*) 批大小（B）以及 *iv*) MRHE 每层的最大参数数量（T）。

我们首先分析每个训练步的图像数量（img/step）。在每个训练步中，基线会从 NeRF Synthetic 数据集的全部 100 张训练图像中随机采样 [1]。在我们的微控制器上，存储 100 张图像会超出片上容量，导致频繁进行片外访问，且每次访问的延迟都高得多（见第 III-B 节）。因此，将每个训练步的图像数量（即 img/step）降至 1 张，有助于降低内存传输延迟。接下来，我们研究输入图像分辨率的影响。由于我们的无人机使用低分辨率相机，我们评估了分辨率低于 NeRF Synthetic 数据集中 800×800px 图像的输入。虽然 Instant-NGP 的运算和峰值内存并不直接取决于输入分辨率，但较大的图像会增加存储需求，因为在整个训练过程中都必须将其存储在内存中（通常是片外内存）。将输入尺寸降至 160×160 px 可减少 61.44MB 的 L3 内存占用。

随后，我们对不同的批次大小（B）进行了消融实验。Instant-NGP 的峰值内存占用主要来自多层感知机的中间特征图。减小 B 会线性降低每个训练步的运算次数和输出特征图的大小，从而显著降低峰值内存需求。最后，我们评估了 T 的影响。由于每个坐标的哈希查询次数保持不变，该参数对运算次数影响甚微。不过，它会直接影响内存使用量，因为较小的 T 值会减小哈希表的大小。总体而言，批大小 B 是决定每步运算次数、进而影响总延迟的主要因素，而 B 和 T 都会显著影响 Instant-NGP 的总内存占用。

B. 神经辐射场部署

通过 IV-A 节介绍的收缩方法优化 Instant-NGP，以减少其运算量和内存占用（在资源受限的微控制器上部署模型时，内存上限为 32MB）后，我们还需要为 Tiny-DroNeRF 算法实现经过优化的嵌入式 C 语言版本。平台严格的资源限制要求我们谨慎安排计算和内存传输，以充分发挥 GAP9 的计算能力。

我们的嵌入式实现由以下部分组成：*i*) 为每个算法模块手写内核；*ii*) 一种分块策略，即将待处理数据划分为适合放入本地内存的数据块；以及 *iii*) 批处理，从而在多核集群上实现高效的并行执行。我们将 Instant-NGP 流水线（图 1）中的每个模块映射到一个定制的手写

内核，并将其并行化运行于 GAP9 集群上。基线 Instant-NGP 使用 float16 算术（权重更新除外，为保证稳定性，权重更新采用 float32 累加 [28]），这使我们能够利用 GAP9 优化过的 SIMD 指令，高效计算多层感知机。

虽然收缩方法确定的配置能够放入纳米无人机的总可用内存中，但我们必须确保每个内核的输入、输出和中间张量都驻留在 128kB 的 L1 内存中。否则，计算会因内存传输带宽不足而停滞。为此，我们设计了一种分块策略：将内核数据细分为能够放入 L1 的数据块，并安排内存传输与计算重叠，从而确保核心持续获得数据。

所有内核都独立地对样本批次或射线批次进行运算，因此，沿这些维度进行内存访问，为并行化和分块提供了直接的机会。然而，MRHE 和占用网格仍然无法放入可用的 L2 存储器；因此，我们利用它们的层次结构，一次将一个层级移至 L2 存储器。

由于 L3 内存带宽约比 L2 低 $35\times$，尽量减少 L3 访问至关重要。由于大多数输入数据（射线和样本）都是随机生成的，我们将采样量限制在能够完全放入片上 L2 内存的最大值。然而，使用过小的批次大小进行训练 B 会阻碍收敛：这需要更多的步，且通常会导致较低的重建质量。为此，我们采用批次累积，将计算批次块的大小与有效训练批次大小解耦。实际上，我们处理适合放入 L2 的批次块，同时在多个步中累积权重梯度，直到达到所需的有效批次大小 B。

C. 联邦实现

对于 Tiny-DroNeRF 的联邦训练，我们采用一种联邦学习方案：其中一架无人机身兼两职，既利用自己的本地数据参与训练，同时又担任协调器。协调器首先在其本地数据集上进行 N_{pt} 步的短暂预训练阶段，随后将得到的初始全局模型分享给集群中的其他无人机。

在每个通信轮次中，所有无人机（包括协调器）都会进行 N_ℓ 步的本地训练，并将更新后的模型参数发送给协调器。我们研究了两种参数交换方式：(i) 每轮都传输占用网格、哈希表和多层感知机参数；或 (ii) 仅传输哈希表和多层感知机参数，由每架无人机在本地更新其占用网格。

模型聚合采用 FedAvg 算法 [21]，，该算法对接收到的参数进行加权平均。在我们的实验中，由于所有无人机拥有相同数量的训练图像，加权平均可简化为算术平均。随后，将所得全局模型广播给所有无人机，并在多个通信轮次中重复这一过程。

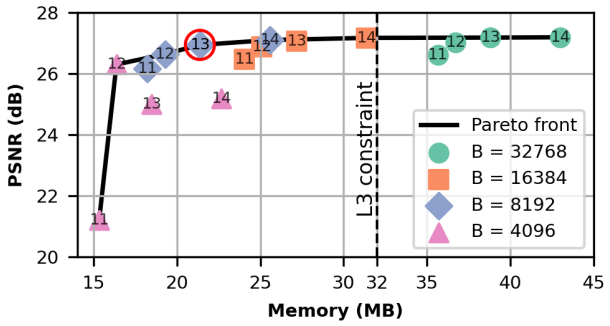


图 3：内存与重建质量的帕累托曲线。每个点对应一种配置：批次大小为 B（标记）且编码大小为 $\log_2(T)$（数字）。红圈标出的是为部署选定的配置。

我们在两种数据划分设置下评估联邦训练。第一种设置为独立同分布（IID）：训练图像经过随机打乱后，均匀分配给各客户端，因此每架无人机都能获得整个数据集的代表性样本。第二种设置为非独立同分布（non-IID）：我们按视点扇区划分训练图像，使每架无人机只能观察场景的特定部分，导致某些侧面完全不可见。

V. 实验结果

A. 实验设置

我们在 NeRF合成360°数据集 [1], 上验证模型。该数据集包含 8 种不同的物体，每种物体有 100 张训练图像和 200 张测试图像，分辨率为800×800。为匹配无人机相机的分辨率，我们将图像分辨率降低至 160×160。在评估时，我们遵循先前的工作，报告峰值信噪比和结构相似性 [1]。

B. 硬件感知内存高效*NeRF*

我们评估通过超参数优化实施的收缩方法如何影响系统的关键参数，例如总内存占用、每更新步操作数，以及峰值信噪比和结构相似性重建指标。表 II 考察了 Instant-NGP 的多种配置，逐一调整每步图像数、图像分辨率、批大小（B）以及哈希编码单个层级的最大参数数量（T）。表中每一行均保留五次独立运行中的最高峰值信噪比和结构相似性。正如预期，基线 Instant-NGP [5]（第一行）所需的内存远超纳米无人机的总内存，即 527 MB，而无人机最多只有 32 MB。类似地，由于每步需要执行 17.52 GOps，其训练时间也将超过两天。

第一项优化来自减少每步存储在片上存储器中的图像数量：从100张减少到1张，可将内存从2.56MB降至0.03 MB，峰值信噪比仅略微下降3.6%。为了在我们片上系统的L2存储器中至少保留哈希网格的一个层级，尽管已经缩减了所有图像分辨率和批大小，我们仍需将 T 至少降至 2^{14}，。然后，从这一

表 II：从 Instant-NGP 模型 [5] 开始进行超参数优化的结果。粗体显示的是选定用于部署的配置。

图像 / step	Res [px]	B	$\log_2(T)$	Mem [MB]	计算量 [GOp/步]	PSNR [dB]	SSIM [−]
100	800	264k	19	527	17.52	32.59	0.96
1	160	32k	14	43	2.64	27.19	0.89
			13	39	2.62	27.18	0.90
			12	37	2.61	27.03	0.89
			11	36	2.61	26.61	0.89
1	160	16k	14	31	1.65	27.17	0.89
			13	27	1.63	27.07	0.89
			12	25	1.62	26.87	0.89
			11	24	1.62	26.46	0.89
1	**160**	**8k**	14	26	1.15	27.10	0.89
			13	**21**	**1.14**	**26.94**	**0.89**
			12	19	1.13	26.64	0.89
			11	18	1.11	26.17	0.88
1	160	4k	14	23	0.91	25.18	0.78
			13	18	0.89	25.00	0.78
			12	16	0.88	26.30	0.89
			11	15	0.88	21.20	0.56

配置（第二行）出发——与基线相比，其峰值信噪比和结构相似性分别下降23%和7%——我们对 B 和 T进行了更细粒度的探索。

所有正在研究的配置都需要使用片外L3存储器。因此，我们认为采用 $B = 8$k和 $T = 2^{13}$ 的配置是最佳折中方案，如图3中的圆圈所示。我们将此配置命名为*Tiny-DroNeRF*。

C. 部署

我们分析了在GAP9片上系统上训练Tiny-DroNeRF的执行时间。我们手工编写了所有计算内核，并根据第IV-B节设计并行化和分块策略，然后在周期精确的GAP9 GVSOC仿真器上分析各内核的执行时间。我们采用8× 批次累积，每次处理包含1024个样本的数据块，并累积梯度，直至达到选定的$B = 8$k有效批次大小。整个流水线的峰值片上存储器使用量始终处于GAP9的严格限制范围内。最大分配量为L2中的1.3 MB和L1中的112 kB，符合片上存储器限制（分别为1.5 MB和128 kB）。每步的L3传输总量仅为3.6 MB，证实了出色的内存局部性，并确保传输延迟（9.7 ms）不会主导每步的计算时间（73 ms）。

图 4 展示了在单个包含 1024 个样本的批次图块上，各内核的执行延迟（蓝色）和 L2存储器使用量（红色）。前向传播的延迟为 12.8 ms，反向传播的延迟为 21.7 ms，而 Adam 优化器每 8 步更新一次，每次耗时 22.4 ms。颜色多层感知机的反向传播占据了延迟和 L2存储器峰值使用量的主导地位，决定了全局存储器上限。其他阶段（例如 MRHE 和密度多层感知机反向传播）的贡献较小

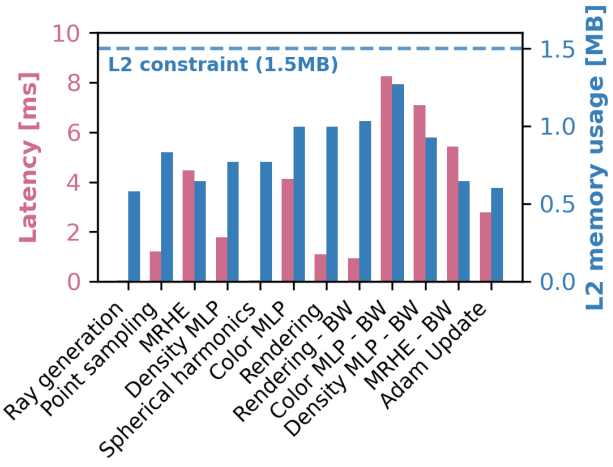


图4：在1024样本批次数据块上，每个训练步的内核延迟与片上存储器峰值用量。

但仍不可忽略。我们还会 [5] 每 256 步更新一次占用网格，延迟为 6.4 s（图 4 中未显示）。

因此，一个训练步的均摊延迟为73毫秒，相当于在单架纳米无人机上完成一次完整重建需要97分钟（8个批次数据块、 × 10k个训练步）。另一方面，渲染场景的新视角（即推理）仅需要前向传播。渲染延迟随取决于场景的每像素采样数 [5]而变化。在我们对NeRF合成数据集360o的实验中，渲染一幅160 × 160 像素图像平均需要13个样本/像素，相当于在1024样本批次数据块上执行325次前向传播。鉴于每个批次数据块的前向传播延迟为12.8毫秒，总渲染时间约为4.1秒。

总体而言，我们的性能剖析凸显了以内存为中心的优化效果，例如分块和批次累加，这些优化使延迟保持计算受限，而非受数据移动限制。

D. 现场*NeRF*

我们在一个真实部署场景中评估 Tiny-DroNeRF：使用纳米无人机在室内环境中采集新的数据集。该环境包含狭窄的墙壁和两个交通锥体，如图 6-A 所示。无人机飞越该空间，同时通过机载相机录制视频。我们以 30Hz 采集160×160px 单色相机图像，通过 Wi-Fi 传输，并使用 COLMAP [29] 进行后处理，以提取真实相机位姿。最终的数据集包含 1954 帧，其中 1709 帧用于训练，245 帧为测试图像。与之前在 NeRF Synthetic 360° 上进行的实验相比，这个现场数据集带来了两项额外挑战。首先，图像为灰度而非 RGB，因此重建难度更高。其次，环境杂乱，纵深可达 5 米；而合成数据集只包含一个背景透明的物体。因此，模型需要重建更大的场景体积。

我们在现场数据集上训练 Tiny-DroNeRF，并评估

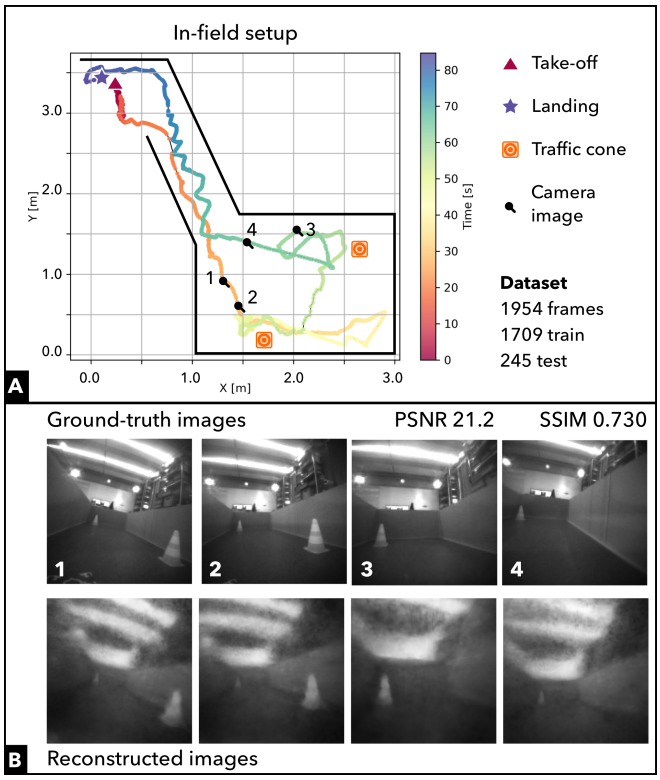


图 5：A）实地数据采集设置。B）真实图像样本及其相应的重建结果。

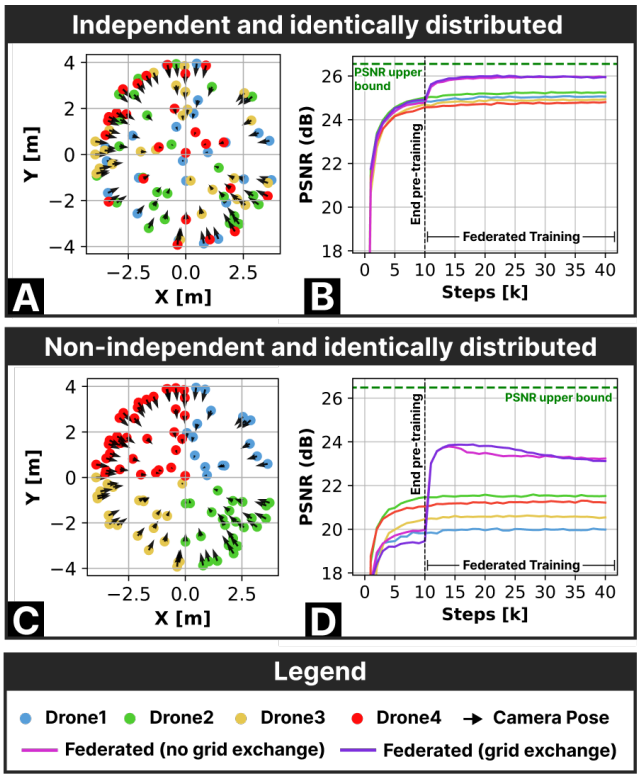


图 6：A-C）独立同分布与非独立同分布数据集分布的二维俯视图；B-D）独立同分布与非独立同分布数据集训练的联邦学习峰值信噪比测试曲线。

其重建质量，采用峰值信噪比（PSNR）和结构相似性（SSIM）进行衡量。我们的模型取得了 21.2 的峰值信噪比和 0.73 的结构相似性。图 5-B 展示了基于补充视频的定性结果。顶部为真实图像，底部为 Tiny-DroNeRF 的重建结果。重建视图捕捉到了走廊和环境中两个锥体的整体几何形状，表明 Tiny-DroNeRF 能够根据搭载在纳米无人机上拍摄的灰度图像，重建真实世界室内场景。

E. 联邦学习

我们使用 NeRF 360° Lego 数据集 [1]，在两种不同设置下评估了 Tiny-DroNeRF 的联邦实现。对于独立同分布设置和非独立同分布设置，我们均遵循该数据集的官方训练–测试划分，将 100 张训练图像平均分配给四架无人机（每架 25 张），并仅使用 200 张测试图像进行评估。相应的数据分布如图 6-A-C 所示。

在两种设置下，我们都探索了两种预训练配置，即 $N_{pt} = 5\text{k}$ 步和 $N_{pt} = 10\text{k}$ 步，以确定一个能够提供足够良好初始化的预训练时长，同时避免延迟联邦训练的开始或造成不必要的能耗。我们还改变了每轮的本地训练步数，因为过于频繁的通信会阻止本地模型取得有意义的进展，而通信过于不频繁则会导致本地过拟合，二者最终都会减缓全局收敛。基于这项探索，我们选择了预训练步数为 $N_{pt} = 10\,\text{k}$，每轮的本地训练步数为 $N_{\ell} = 1\text{k}$。本节中的所有结果均基于此配置。

随后，如图 6-B-D 所示，我们在独立同分布和非独立同分布条件下，将分别在每架无人机上使用其自身图像子集独立训练的模型，与两种联邦变体（进行和不进行占用栅格交换）进行了比较。作为参考，该图还包括一个在所有本地数据集的并集上训练的集中式模型，它代表了可实现性能的 上界。

在独立同分布和非独立同分布设置下，无论是否进行占用栅格交换，两种联邦变体的结果都相当，表明交换网格并非必要。这一发现尤其值得注意，因为占用网格是一种高维数据结构，约占总通信量的 88%。包含网格时，每轮需要传输 4.7 MB，导致每轮通信开销为 20.16 s；而省略网格后，传输数据量降至 0.52 MB，开销则降至每轮 2.24s。

此外，联邦模型的表现始终优于单独训练的模型。在独立同分布设置下，联邦训练取得的峰值信噪比比表现最好的单无人机模型高 0.7 dB，仅比集中式上界低 0.6 dB。在更具挑战性的非独立同分布设置下，联邦模型的表现比上界低 3.25 dB，但仍比表现最好的单无人机

模型高 1.7 dB，这表明无人机借助联邦模型，能够推断出物体中它们未曾直接观测过的部分的空间特征。

VI. 结论

稠密三维重建任务在纳米无人机上一直不可行，因为现有方法无法在可用资源下承受，即低于100mW的微控制器和几十MB的内存。我们提出Tiny-DroNeRF，一种面向微控制器的轻量级神经辐射场，相对于Instant-NGP将内存减少96%，峰值信噪比仅下降5.7 dB。我们从机载灰度图像重建真实场景，峰值信噪比为21.2。我们进一步提出一种联邦学习方案，使无人机群能够从有限的本地数据合成未见视角，相比单无人机训练将峰值信噪比提高1–2 dB，且每1 k步仅需2.24秒通信。据我们所知，这是首次在资源受限的微控制器上，结合联邦学习方法演示神经辐射场。

参考文献

[1] 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》，《*ACM*通讯》，第65卷，第1期，第99–106页，2021年12月。[2] P. Hedman、P. P. Srinivasan、B. Mildenhall、J. T. Barron 和 P. De- bevec，《为实时视图合成烘焙神经辐射场》，载于*IEEE/CVF*国际计算机视觉会议论文集，2021年，第5875–5884页。[3] A. Yu、R. Li、M. Tancik、H. Li、R. Ng 和 A. Kanazawa，《用于神经辐射场实时渲染的PlenOc-trees》，载于*2021*年*IEEE/CVF*国际计算机视觉会议（*ICCV*）。美国加利福尼亚州洛斯阿拉米托斯：IEEE计算机学会，2021年10月，第5732–5741页。[4] S. Fridovich-Keil、A. Yu、M. Tancik、Q. Chen、B. Recht 和 A. Kanazawa，《Plenoxels：无神经网络的辐射场》，载于*2022*年*IEEE/CVF*计算机视觉与模式识别会议（*CVPR*）。美国加利福尼亚州洛斯阿拉米托斯：IEEE计算机学会，2022年6月，第5491–5500页。[5] 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》，第41卷，第4期，第102:1–102:15页，2022年7月。[6] A. Chen、Z. Xu、A. Geiger、J. Yu 和 H. Su，《TensoRF：张量辐射场》，载于计算机视觉——*ECCV 2022*：第*17*届欧洲会议，以色列特拉维夫，*2022*年*10*月*23*日至*27*日，论文集，第*XXXII*部分。柏林、海德堡：施普林格出版社，2022年，第333–350页。[7] S. Fridovich-Keil、G. Meanti、F. R. Warburg、B. Recht 和 A. Kanazawa，《K-planes：空间、时间与外观中的显式辐射场》，载于*IEEE/CVF*计算机视觉与模式识别会议论文集，2023年，第12 479–12 488页。[8] Z. Chen、T. Funkhouser、P. Hedman 和 A. Tagliasacchi，《MobileNeRF：利用多边形光栅化管线，在移动架构上高效渲染神经场》，载于*2023*年*IEEE/CVF*计算机视觉与模式识别会议（*CVPR*）。美国加利福尼亚州洛斯阿拉米托斯：IEEE计算机学会，2023年6月，第16 569–16 578页。[9] S. Shin 和 J. Park，《二进制辐射场》，载于第*37*届神经信息处理系统国际会议论文集，丛书：NIPS '23。美国纽约州雷德胡克：Curran Associates Inc.，2023年。[10] Y. Chen、Q. Wu、M. Harandi 和 J. Cai，《基于Instant-NGP的神经辐射场能压缩到什么程度？》，载于*2024*年*IEEE/CVF*计算机视觉与模式识别会议（*CVPR*）。美国加利福尼亚州洛斯阿拉米托斯：IEEE计算机学会，2024年6月，第20 321–20 330页。[11] M. J. Anderson、J. G. Sullivan、J. L. Talley、K. M. Brink、S. B. Fuller 和 T. L. Daniel，《"Smellicopter"：一种生物混合式气味定位纳米飞行器》，载于*2019*年*IEEE/RSJ*智能机器人与系统国际会议（*IROS*），2019年，第6077–6082页。[12] A. Birk、B. Wiggerich、H. B¨ulow、M. Pfingsthorn 和 S. Schwertfeger，《使用无人机（UAV）执行安全、防护与救援任务》，《智能与机器人系统期刊》，第64卷，第1期，第57–76页，2011年10月。[13] M. Zauli 、 M. Pirazzi 、 F. Zonzini 和 L. De Marchi，《利用微型飞行器作为无线振动监测的传感器节点》，载于2024年IEEE传感器应用研讨会（SAS），2024年7月，第1–6页。[14] N. Zimmerman、H. M¨uller、M. Magno 和 L. Benini，《纳米无人机上基于楼层平面图和语义传感器融合的全机载低功耗定位》，载于2024年IEEE机器人与自动化国际会议（ICRA），2024年5月，第11 913–11 919页。[15] L. Lamberti、L. Bompani、V. J. Kartsch、M. Rusci、D. Palossi 和 L. Benini，《纳米无人机自主探索策略的生物启发式设计：基于CNN的目标检测》，载于2023年欧洲设计、自动化与测试会议（DATE），2023年4月，第1–6页。[16] M. Pourjabar、M. Rusci、L. Bompani、L. Lamberti、V. Niculescu、D. Palossi 和 L. Benini，《面向自主纳米无人机群探索的多传感器防碰撞设计》，载于2023年第30届IEEE电子、电路与系统国际会议（ICECS），2023年12月，第1–5页。[17] L. Lamberti、E. Cereda、G. Abbate、L. Bellone、V. J. K. Morinigo、M. Barci´s、A. Barci´s、A. Giusti、F. Conti 和 D. Palossi，《一种用于自主纳米无人机竞速的仿真到现实深度学习框架》，《IEEE机器人与自动化快报》，第9卷，第2期，第1899–1906页，2024年2月。[18] H. Wang、J. Wang 和 L. Agapito，《Co-SLAM：用于神经实时同步定位与建图的联合坐标与稀疏参数编码》，IEEE计算机视觉与模式识别国际会议（CVPR）论文集，第13 293–13 302页，2023年。[19] Z. Li、T. M¨uller、A. Evans、R. H. Taylor、M. Unberath、M.-Y. Liu 和 C.-H. Lin，《Neuralangelo：高保真神经表面重建》，载于IEEE/CVF计算机视觉与模式识别会议论文集，2023年，第8456–8465页。[20] L. Rold˜ao、R. de Charette 和 A. Verroust-Blondet，《三维语义场景补全：综述》，《国际计算机视觉期刊》，第130卷，第8期，第1978–2005页，2022年8月。[21] B. McMahan、E. Moore、D. Ramage、S. Hampson 和 B. A. y. Arcas，《基于去中心化数据的深度网络通信高效学习》，载于第20届人工智能与统计学国际会议论文集，丛书：机器学习研究论文集，A. Singh 和 J. Zhu 编，第54卷。PMLR，2017年4月20–22日，第1273–1282页。[22] L. Holden、F. Dayoub、D. Harvey 和 T.-J. Chin，《联邦神经辐射场》，2023年。[23] T. Suzuki，《使用神经辐射场进行大规模场景建模的联邦学习》，2024年。[24] M. Asadi、K. Zareinia 和 S. Saeedi，《Di-NeRF：用于协同学习并采用相对位姿优化的分布式神经辐射场》，《IEEE机器人与自动化快报》，第9卷，10 534，2024 11，10 527 no. pp. – 。[25] T.-I. Szatmari 和 A. Cauligi，《面向行星探索的联邦多智能体建图》，2025年。[26] Z. Tasneem、A. Dave、A. Singh、K. Tiwary、P. Vepakomma、A. Veer-araghavan 和 R. Raskar，《DecentNeRFs：基于众包图像的去中心化神经辐射场》，载于计算机视觉——ECCV 2024：第18届欧洲会议，意大利米兰，2024年9月29日至10月4日，论文集，第LIX部分。柏林、海德堡：施普林格出版社，2024年，第144–161页。[27] H. M¨uller、V. Kartsch 和 L. Benini，《GAP9Shield：一种用于纳米无人机视觉与测距应用的150 GOPS AI能力超低功耗模块》，载于欧洲机器人论坛。施普林格出版社，2024年，第292–297页。[28] P. Micikevicius、S. Narang、J. Alben、G. F. Diamos、E. Elsen、D. Garc´ıa、B. Ginsburg、M. Houston、O. Kuchaiev、G. Venkatesh 和 H. Wu，《混合精度训练》，载于第6届国际学习表征会议（ICLR 2018），加拿大不列颠哥伦比亚省温哥华，2018年4月30日至5月3日，会议论文集，2018年。[29] J. L. Sch¨onberger 和 J.-M. Frahm，《重新审视运动恢复结构》，载于IEEE计算机视觉与模式识别会议论文集，2016年，第4104–4113页。