本文已被 IEEE ICRA 2023 会议接收发表 ©2023 IEEE。

超低功耗深度学习单目相对定位（用于机载纳米四旋翼）

S. Bonato[1], S. C. Lambertenghi[1],E. Cereda[1],A. Giusti[1],和 D. Palossi[12]

arXiv:2303.01940v1 [cs.RO] 3 Mar 2023

摘要— 精确的相对定位是群体机器人系统中的关键功能模块。本文提出一种新颖的自主端到端系统，通过深度神经网络（DNN）实现两架同伴纳米无人机之间的单目相对定位，即无人机重量低于 40 g、处理功耗低于 100 mW。为适应超受限的纳米无人机平台，我们提出一种垂直整合框架，涵盖从数据集采集到最终现场部署的全过程，包括数据集增强、量化和系统优化。实验结果表明，我们的 DNN 仅利用低分辨率单色图像，即可精确定位距离达 ∼2m 的 10 cm 大小目标纳米无人机。在独立测试数据集上，我们的模型取得了 0.42 的平均 R^2 分数和 18 cm 的均方根误差；在一次持续 ∼60s 的飞行测试中，实地预测误差平均为 15 cm，闭环控制误差为 17 cm。最终，所提出的系统通过展示长续航跟踪性能（持续跟踪最长达 2 分钟）、在前所未见环境中的部署泛化能力，以及仅需 95mW 功耗即可实现 48 Hz 机载实时推理速率，提升了当前最先进水平。

补充视频材料

实地测试：https://youtu.be/pUGL1qu3Z1k。

一、引言

纳米级四旋翼是一种新兴的机器人平台，重量不到 40 克，直径不到 10 厘米 [1], [2]。它们成本低廉，能够在受限空间内以及人类附近安全运行，因此在许多场景中开辟了颇具吸引力的新应用 [3]。当多架自主纳米无人机以群体（集群）的形式协同工作时，这一平台的潜力将进一步提升；然而，要实现这一雄心勃勃的目标并非易事，因为它们需要在飞行过程中可靠地相互定位，即相对定位。鉴于这类机器人在传感器类型和质量（例如 QVGA摄像头）、计算能力（单核微控制器单元上的算力仅为数百 MOps/s）、内存容量（片外存储器不足 10 MB，片上存储器不足 1 MB）和功耗（计算功耗仅为数百 mW）方面受到的极端限制，这无疑是一项重大挑战。

我们的工作通过设计、实现并在实际环境中（闭环）全面验证一种完全在机载运行的视觉相对定位方法，来应对这一挑战。

本研究部分得到了阿联酋技术创新研究院（TII）安全系统研究中心（SSRC）以及瑞士国家科学基金会（SNSF）通过国家机器人研究能力中心的资助。[1]S. Bonato、S. C. Lambertenghi、E. Cereda、A. Giusti 和 D. Palossi 任职于达勒莫勒人工智能研究所、意大利语区大学和瑞士南方应用科技大学，地址为瑞士卢加诺，邮编 6962。name.surname@idsia.ch[2]D. Palossi 还任职于苏黎世联邦理工学院集成系统实验室¨苏黎世，¨瑞士苏黎世，邮编 8092dpalossi@iis.ee.ethz.ch



图 1. 我们的实地系统：一架纳米无人机通过视觉估计另一架无人机的相对位姿。

该方法可在一架 27 g 的纳米无人机上以最高 48 帧/秒的速度运行，功耗不超过 95 mW。我们的系统如图 1 所示，使用来自低分辨率、低动态范围微型摄像头的输入帧；它采用定制的深度卷积神经网络（卷积神经网络），包含超过 304k 个参数，并以回归器的形式训练，用于返回同伴纳米无人机的相对位置。我们讨论并说明了贯穿整个系统的设计选择，包括*i*) 用于训练和评估数据集的数据采集设置；*ii*) 卷积神经网络架构的选择，以及其在目标平台上的训练、量化和部署；*iii*) 对回归性能、吞吐量和计算功耗的分析；*iv*) 将系统集成到控制回路中，以完成跟随同伴四旋翼无人机的任务；*v*) 在各种环境中验证现场控制性能。尽管在为大型机器人开发系统时，这些方面往往被忽视，但我们强调，其中一些方面对于面向纳米四旋翼无人机的应用至关重要。

本文的主要贡献在于讨论并通过大量定量实验验证了整个方法，相关视频演示作为补充材料发布。在第二节回顾相关工作后，第三节介绍系统的关键方面（硬件平台、数据集、卷积神经网络的架构与训练，以及机器人控制器）。在第五节总结全文之前，第四节概述了在静态数据集和实地环境中获得的实验结果，表明：*i*) 在独立测试数据集上，平均 R^2得分为 0.42，均方根误差（RMSE）为 18 厘米；*ii*) 在 ∼60秒飞行测试中，实地平均预测误差为 15 厘米，对应的闭环控制误差为 17 厘米；*iii*) 长续航跟踪性能可支持最长 2 分钟的飞行；以及*iv*) 机载实时推理速率达到 48 Hz，计算功耗为 95 mW。

II. 相关工作

在许多应用中，估计附近机器人的位姿至关重要，例如群体编队、编队飞行和导航过程中的避障等。深度学习领域近期有许多研究利用一张或多张相机图像来估计已知物体的位姿 [4], [5], [6], [3], [1]。在机器人学中，这项任务通常借助机器人的机载传感器来完成，估计物体相对于机器人参考坐标系的位置（有时也包括方向）；完成这项感知任务后，机器人便能执行朝向物体前进、避开物体或抓取物体等高级操作。我们的工作研究一种特定的视觉物体位姿估计场景，其中观测机器人和单个目标物体均为 Crazyflie 纳米无人机。

物体位姿估计子系统通常依赖机载立体相机 [7], [8], 激光雷达 [9], [10],红外传感器 [11], 超宽带（UWB）无线电 [12], [13] 或单目视觉 [4], [1], [3], 的数据，我们的系统也采用了单目视觉。尽管摄像头图像质量较差，我们仍证明，即使没有视觉基准标记 [4], [14], [15], 也足以定位观测到的 10 厘米级纳米无人机。

根据应用需求，可能需要估计物体的完整三维位姿，即其位置和三维方向 [8], [4]。在这种情况下，可以在机器学习背景下采用专门的参数化方法，明确考虑刚体变换所构成的 SE(3) 群的拓扑结构 [16]。在其他应用中，仅估计三维位置就足够了 [11], [17], [18], [5], [13]。本文遵循先前的工作 [3], ，通过无人机的三维位置来解决其位姿估计问题。

视觉位姿估计是一项具有挑战性的模式识别问题，并且会带来巨大的计算负担，尤其是考虑到我们的目标纳米无人机平台的能力有限。在某些情况下， [19],此类计算会被卸载到功能更强大的远程计算机上，由其接收传感数据并返回估计位姿。大型无人机 [20],[21]配备了功能强大的中央处理器（有时还包括图形处理器）和高质量传感器。相比之下，我们工作的一个关键方面是将所有计算集成到一台内存和计算能力均受限的 Crazyflie 纳米无人机 [22], 上；由于其采用微控制器级处理器（功耗低于 100mW），这带来了重大挑战。

纳米无人机的定位系统通常依赖 UWB 技术，采用额外的临时基础设施（UWB 固定锚点） [23] ，或机载多边定位算法 [13]。前一种方法的主要缺点是灵活性和可部署性极低，且成本较高；后一种方法由于 UWB 无线电通信，会显著增加机载功耗，在工作状态下最高可达 ∼300mW [13]。不同的是，Li 等人 [1]最近研究了一种基于视觉的位姿估计问题，与本文所述问题非常相似，且使用相同的硬件（即无人机、摄像头和微控制器）。他们的工作重点是提出一种新方法，为自监督卷积神经网络生成基于 UWB 的训练标签。这些基于 UWB 的标签用于训练卷积神经网络，随后仅在一个规模有限、包含 48 张图像的测试集上进行离线验证。相比之下，我们的贡献在于系统设计、集成和全面的实验验证：我们所选的架构规模更大（卷积层为 8 层而非 5 层，参数更多） 12×，但由于所需运算更少，运行速度反而快 30%。我们还研究了实时性能，在一个规模更大的测试集上进行了离线定量评估 15×，并报告了整个系统与控制器集成后在连续飞行期间的实地性能定量分析。归根结底，我们的主要贡献与具体应用无关，而实验验证则考察了一项简单的编队飞行任务。

III. 系统实现

A. 机器人平台

在本工作中，我们采用 Bitcraze Crazyflie 2.1 开源纳米无人机，其主要飞行控制器为一款 STM32 单核微控制器单元（MCU）。在我们的配置中，我们为其扩展了一块商用现成（COTS）Flow-deck，通过光流摄像头和测量距地距离的飞行时间传感器来改进机载状态估计。此外，我们还为纳米无人机配备了一个额外的并行超低功耗（PULP）微控制器、一台 QVGA 单色摄像头、一个 WiFi 无线电，以及 8/16MB 的动态随机存取存储器/闪存；这些组件均搭载于 AI-deck COTS 伴随板上。计入所有这些附加组件后，我们的纳米无人机重量为 33g，直径为 10 cm。

PULP 微控制器是 Greenwaves Technologies GAP8[1],a RISC-V 架构的 9 核处理器。GAP8 具有两个电源域：织构控制器 （FC），其配备一个内核，可与外部内存/接口（例如 AI 开发板上的动态随机存取存储器/闪存和 UART 通信通道）以及第二个域——并行集群 （CL）进行交互。CL 负责使用其 8 个内核执行密集计算。该处理器的整体存储层次结构分为两层：所有集群内核共享的 64 kB 低延迟 L1 存储器，以及位于 FC 域内的 512 kB L2 存储器。然而，GAP8 不提供数据缓存或硬件浮点单元（FPU），因此分别需要进行显式数据管理并采用整数定点量化运算。

B. 训练和测试数据集

我们的工作依赖于专门采集的带标签数据集；我们分别采集了两个互不重叠的数据集，用于训练和定量测试。数据集中的每个样本由以下内容组成：一张输入图像（320 × 320 灰度图）以及四个输出变量的真实标注数据，即 x、 y、 z 和 ϕ，也就是目标无人机相对于观测无人机的相对朝向。真实值由飞行无人机的绝对位姿生成，这些位姿通过一套由 18 台摄像机组成的 Optitrack 红外运动追踪系统记录。在数据采集过程中，无人机由 ROS 节点按照预定义的航点序列进行控制

[1]https://greenwaves-technologies.com/gap8 微控制器 人工智能

waypoints.ROS 同时负责精确匹配无人机的位姿与通过 WiFi 从纳米无人机传输的机载图像。

为了收集训练数据集，我们利用这一装置记录两架无人机沿不同路径飞行的多个航次，旨在最大限度地提高相对位姿各个分量的变化范围，并采集具有多种背景的图像。具体而言，我们采用两种运动模式。第一种模式中，两架无人机彼此相对，绕房间中心旋转；两架无人机同时记录数据，因此既充当观测无人机，也充当目标无人机。我们通过调节两架无人机所绕圆周的半径、两架无人机之间的相对高度，以及两架无人机沿圆周的角度偏移量，重复进行多个航次。第二种模式中，观察无人机保持静止，目标无人机沿位于竖直平面内的多个圆周飞行，这些平面与观察无人机的光轴正交；目标无人机沿着多个此类圆周飞行，各圆周与观察无人机的距离和半径各不相同。

数据采集完成后，所有目标无人机不在观察无人机视场内的样本均被剔除；最终共获得 21001 个样本，并随机划分为 19664 个训练样本和 1337 个验证样本。训练阶段会即时对图像进行数据增强。随机数据增强会改变图像曝光、伽马值和动态范围；此外，还会随机对每张图像施加噪声、模糊和/或渐晕。

对于测试数据集，我们记录了一段新的会话，其中包含训练数据集中均未出现的行为；这确保定量评估不会因训练数据上的潜在过拟合而获得虚高结果。具体而言，目标无人机沿螺旋状轨迹飞行，使相对位姿的所有组成部分都呈现显著变化。无人机最初在 x轴上对齐，相距0.2米，彼此相向。观察无人机保持静止，目标无人机则沿 x 轴缓慢飞离观察无人机。在此过程中，目标无人机沿该轴交替进行向内和向外的螺旋运动，如图2所示。该过程共生成754个样本。

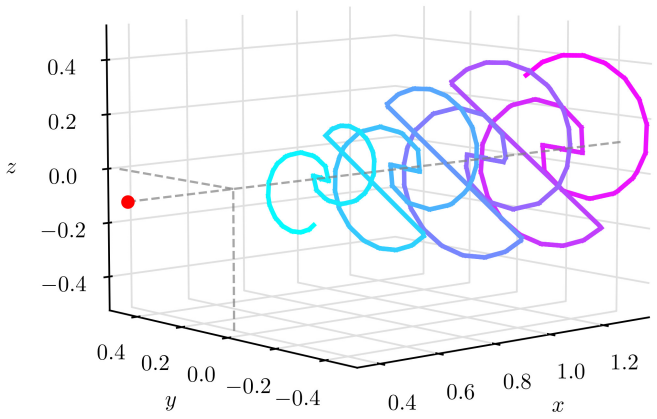


图2. 目标纳米无人机飞行的三维轨迹，用于采集我们的测试数据集。显示为红色的观察无人机位于 $(x,\ y,\ z) = (0,\ 0,\ 0)$，朝向正 x

C. 深度神经网络

我们测试了两种卷积神经网络：PULP-Frontnet，这是一种小型网络，适用于感知任务和纳米无人机；以及 MobileNetV2，这是一种常用于边缘设备的网络。对于 Frontnet，网络以 5×5 卷积层开始，随后是 2×2 最大池化层，然后重复三次由两个 3×3卷积组成的自定义模块。每个模块的第一层卷积步长为2，进一步缩小特征图尺寸并增加输出通道数。所有卷积层后均接批归一化层和 ReLU激活。

对于 MobileNetV2，我们考虑了16种在内存和计算需求方面均比原始架构更轻量的变体。这些网络由一个标准卷积层、四个倒置残差块、一个平均池化层和一个全连接层组成。图3展示了我们精简版 MobileNetV2 的模板。我们的搜索空间由 t 和 n参数化，二者分别表示两个中间块的扩展因子和重复次数。我们考虑由 $(t,\ n) \in \{6, 8, 10, 12, 14\} \times \{2, 3, 4\}$定义的15个模型，其中 $\times$ 表示笛卡尔积；此外，我们还纳入了 $(t,\ n) = (2,\ 2)$ 的情况，即在考虑步幅布局约束时可能的最小变体。

D. 系统部署

在 PyTorch 中完成模型训练后，由于我们的目标处理器缺少 FPU（第 III-A 节），因此需要将其数值表示从 32位浮点表示（即全精度）转换为 8位整数定点运算。我们使用 NEMO [24] 库，该库采用参数化裁剪激活技术 [25]，通过三个连续步骤对卷积神经网络的权重和激活值进行量化。第一步训练浮点卷积神经网络，以最小化位姿估计向量的 L1损失之和（$x,\ y,\ z,\ \phi$）。随后，我们将验证损失较低的模型转换为所谓的伪量化 版本，此时权重和激活值的幅值只能取离散的（0-255）值集合，但仍使用 float32数据类型。在进入最后阶段之前，我们进行为期十个训练轮次的基于 PACT 的微调量化阶段。最后进入可部署整数 阶段，此时所有张量均由整数和一个浮点缩放因子表示。此时，网络可以完全通过仅对整数值进行乘法和累加来执行。

所采用的第二个工具称为 DORY [26]，，它从量化模型出发，生成经过优化的 C代码。DORY 通过将优化后的基础内核（例如卷积、线性层、最大池化层等）组合到基于模板的封装代码中，还负责内存编排，以充分利用 GAP8 的 L2 和 L1存储器以及片外DRAM中的数据局部性。

为完成整个处理流程的开发，我们在 DORY 生成的推理例程基础上扩展了：*i*) 图像采集例程、*ii*) 图像裁剪（从 160×160 到 160×96像素），以及*iii*) 最后的 UART通信，以便

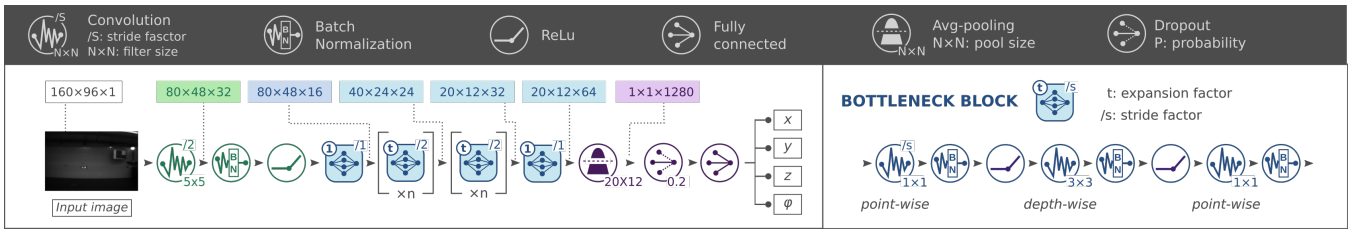


图3. 用于我们轻量化MobileNetV2变体的网络结构模板。

将卷积神经网络的输出发送到纳米无人机上的主 STM32 微控制器。最后，我们以双缓冲方式将推理时间（所有例程中最长的部分）与图像采集和裁剪过程（合计 $\sim$ 0.6M 个周期）重叠。借助这一机制，当集群正在计算上一张图像时，飞控可以开始采集并裁剪新图像，从而使整个处理流程的延迟仅受推理例程限制。

一旦新的预测值（x，y，z，以及 ϕ）从GAP8发送至STM32，该位姿便会通过卡尔曼滤波器进行滤波，从而平滑位姿序列。滤波器的输出被送入速度控制器，如 [3],所示，该控制器将这一时间上一致的位姿序列转换为速度设定值，以使目标纳米无人机保持0.8m的恒定距离，然后再转发至Crazyflie PID级联控制器。

IV. 结果

A. 网络架构的比较

本节中，我们将PULP-Frontnet模型与MobileNetV2卷积神经网络的16种变体进行比较。图4报告了每种架构的三项指标：内存占用（即8位权重的数量）；乘加（MAC）运算次数，其直接影响以每秒帧数衡量的可实现吞吐量；以及回归性能。

我们使用 R^2 分数（即决定系数）分别量化各输出变量的回归性能。这是一种标准指标 [27] ，用于衡量模型所解释的输出变化比例。始终输出测试样本平均值的虚拟预测器得分为 $R^2 = 0$；而理想回归器的得分为 $R^2 = 1$。图 4 报告了对 $x, y,$ 和 z 分量的平均 R^2 分数；该分数忽略了 ϕ 分量的性能。正如我们将在下文讨论的，所有模型在这一分量上的表现都较差。

最大的 MobileNetV2 模型具有最佳回归性能，紧随其后的是 PULP- Frontnet（分别为 $R^2 = 0.43$ 和 0.42）。我们观察到，MobileNetV2 架构呈现出一致的趋势：n值越大，回归性能越好。这一趋势表明，对于我们的问题，在参数数量相同的情况下，较深的 MobileNetV2 模型比宽度较大的模型更适用。

在存储权重所需的内存方面，MobileNetV2模型的大小为 111 kB 至 340 kB，而 PULP-Frontnet 模型需要 304 kB ；所有这些模型都能装入 GAP8 SoC 的 L2 存储器中。MAC 数量方面的差异更为显著：所有 MobileNetV2模型所需的运算都远多于 PULP-Frontnet，最多可达 10.4$\times$ 倍。假设所有模型的计算效率均为 4 MAC/周期，且集群频率为 170 MHz，则最小和最大的 MobileNetV2模型分别可达到 20 帧/秒和 4.6 帧/秒，而 PULP-Frontnet 可达到 48 帧/秒。因此，我们选择 PULP-Frontnet作为候选模型，将其部署到纳米无人机上进行实地测试。

在图5-A中，我们报告了 PULP-Frontnet CNN 各输出变量的 R^2 分数，其中包括全精度版本和量化版本。x、y和 z 输出的回归性能良好，最高可达 $R^2 = 0.53$；而ϕ 分量的得分接近 0，也就是说，其表现并不优于简单预测器。ϕ 分量的性能如此之差是意料之中的：要正确预测 ϕ，就必须通过低分辨率、单色且带噪声的图像理解纳米无人机的朝向；即使对人类观察者来说，这也是几乎不可能完成的任务，因为如图5-B/C所示，目标外观不会因其朝向而显著改变。图5-A还表明，训练后量化阶段对回归性能的影响可以忽略：R^2最多下降 3 个百分点（从 0.52 降至 0.49）。

B. 回归性能

图6进一步分析了 PULP-Frontnet 模型对各输出变量的预测值与真实值之间的对比；理想估计器生成的散点图中，所有点都应落在对角线上。

我们首先注意到，x （即目标距离）的预测颇具挑战。这是意料之中的，因为 x 只能通过目标图像的大小来推断，而目标图像仅占据少数几个像素，因此模型（以及人类观测者）难以准确评估。输出 y 和z 的估计准确度显著更高；这两个变量分别与目标图像在水平轴和垂直轴上的位置有关，因此更容易预测。我们还注意到，当目标较远时，预测往往更保守，即更接近 0。这是因为目标图像会变得更小、更难检测，模型因而无法确定目标位置。此外，相对位姿的 y 和 z 分量不仅取决于目标图像在帧中的位置，也取决于目标距离；因此，影响距离估计的不确定性也会反映在 y 和 z的预测中。

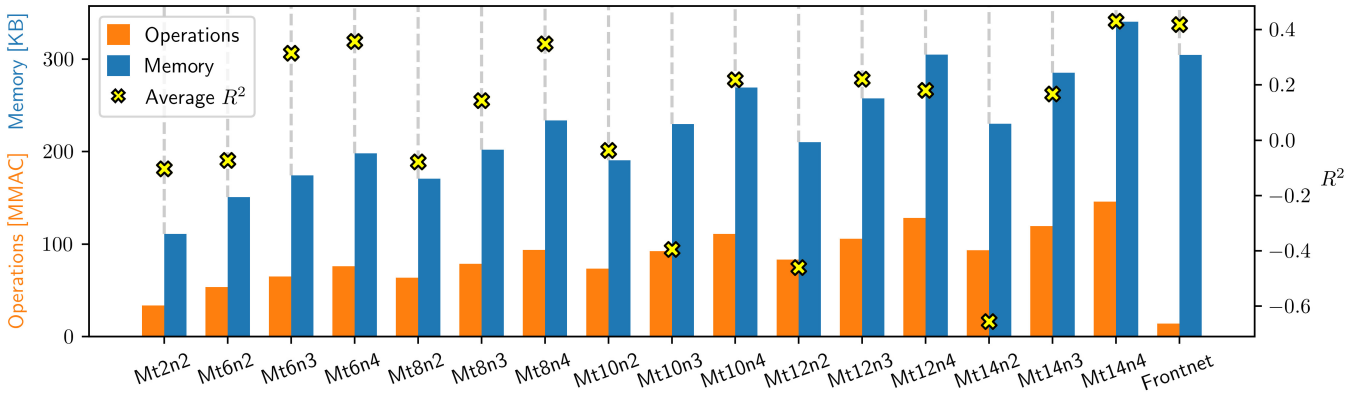


图4。所有测试架构的概览；柱状图对应左轴，表示乘加运算次数（橙色）和8位权重数量（蓝色）；交叉标记对应右轴，表示在 x、 y 和 z 输出上取平均的 R^2 得分。

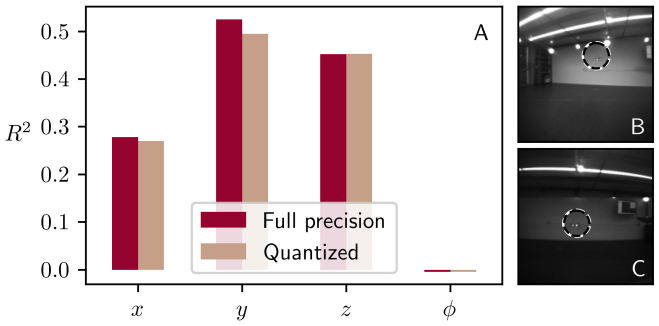


图5。A）PULP-Frontnet 针对每个输出变量的回归性能；全精度（红色）和量化（棕褐色）模型。B、C）具有不同 ϕ 的两幅图像（目标无人机以圆圈标出）。

对于 z 分量，即两架无人机之间的相对高度，我们观察到预测值略有但持续偏低（低了 ~5cm）。我们认为，这种现象与训练和测试数据集的数据分布差异有关，或者是四旋翼无人机上的相机俯仰角略有偏差所致。图6最右侧的图表表明，如前所述，模型无法为 ϕ提供任何有用信息，即相对位姿的方向分量。

C. 机载实时性能

本节评估了 PULP-Frontnet CNN 在不同工作点（即不同电压和频率）下运行时的机载性能。如 [3], 所述，我们探讨了三种配置，分别是最低功耗 （VDD@1.0 V FC@25MHz CL@25 MHz）、最节能 （VDD@1.0 V FC@25 MHz CL@75 MHz）和最高性能（VDD@1.2 V FC@250 MHz CL@175 MHz）。对于所有旨在延长系统续航时间的场景，最低功耗配置尤为重要，例如纳米无人机降落后开始充当无处不在的智能传感器。另一方面，只要吞吐量满足任务要求，就应始终优先选择最节能的配置。最后，在需要最大化 CNN 推理速率的情况下，例如

表1 PULP-FRONTNET 吞吐量和功耗。

频率 [MHz]	VDD [V]	帧率 [帧每秒]	功耗 [毫瓦]
飞控/集群 @ 25/25	1.0	6.8	9.9
飞控/集群 @ 25/75	1.0	19.7	25.1
飞控/集群 @ 250/175	1.2	48.3	95.4

敏捷性和响应速度至关重要。

表1报告了使用 RocketLogger 数据记录器 [28] (64 ksps) 对 GAP8 SoC 进行性能（帧每秒，fps）和平均功耗测量的结果。结果范围从最低的 6.8 fps@9.9mW 到最高的 48.3fps@95.4 mW。图7所示的纳米无人机整体功耗分解呈现出预期趋势：执行器的功耗占主导地位（占总功耗的95%），而电子设备仅占一小部分（5%）。最后，运行我们单目相对定位卷积神经网络的 GAP8 SoC，其功耗仅占总功耗的1.2%，与现有技术水平 [3], [29]非常一致。

D. 现场控制性能

我们使用两架飞行无人机进行实地测试，以验证系统。其中一架无人机充当目标无人机，执行预定义动作，并沿着与第 III-B 节所述相同的螺旋状轨迹飞行。另一架无人机（观察无人机）则尝试与目标无人机保持固定的相对位姿：距离为 1m（$x = 0.8$m）；高度相同（$z = 0$m）；同时保持目标位于帧的中心（$y = 0$m），并使观察无人机的偏航角保持不变。我们的卷积神经网络在观察无人机的 Nano 无人机上实时运行，其预测值用于闭环控制。在此实验中，两架无人机的位姿均由运动捕捉系统记录。观察无人机的朝向经过调整，使其 x 轴（即前进方向）与世界坐标的 x 轴对齐。为定量评估性能，我们将记录到的观察无人机位姿（以世界坐标表示）与预期位姿进行比较，如图 8 所示。我们还给出了平均绝对误差（MAE）评分，以便更轻松地比较三个分量。

如补充视频所示，观测无人机能够在整个测试过程中（~60秒）可靠地估计相对位置。

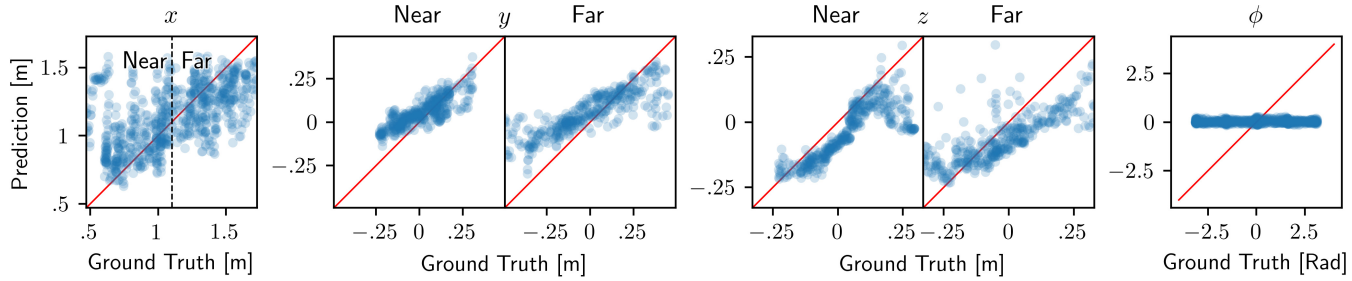


图6。 预测值（纵坐标轴）与整个测试集的真实值（横坐标轴）进行对比，每个输出变量对应一个样本点。For y、 z 和 ϕ 输出分别重新记录目标位于近处（x <1.1 m）和远处（x ≥1.1 m）的样本。

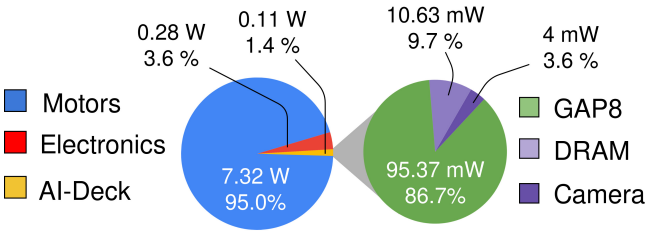


图7。机器人平台运行 PULP-Frontnet 时的功耗分解：VDD@1.2 V、FC@250MHz，以及集群@175 MHz。

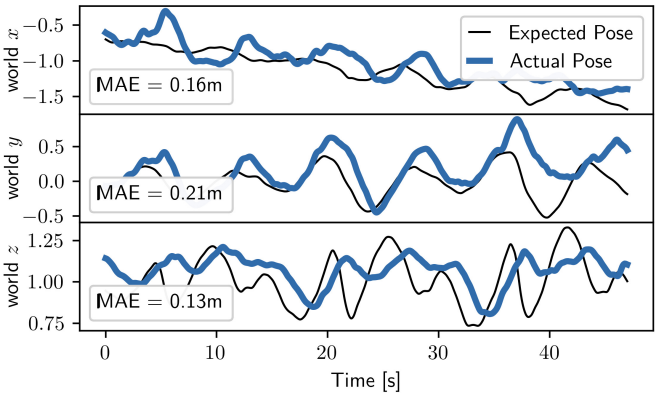


图8。实地测试期间，观察无人机绝对世界位姿的 x、 y 和 z 分量。粗蓝线表示无人机的真实位姿；细黑线表示无人机预期跟踪的理想位姿，即相对于移动目标的固定相对位姿。

尽管控制器计算带来了一定的延迟，但我们可以观察到所有分量的绝对坐标都高度吻合。 x 和 z 分量最为准确，平均绝对误差分别为16厘米和13厘米。相比之下，在 y 输出上，我们看到观测无人机每当需要向左移动时都会略微超调，因此平均绝对误差为21厘米。另一方面，我们可以观察到该分量的响应速度最高，无人机能够在需要时迅速反转 y 方向。纳米无人机位姿估计任务极具挑战性，因为我们需要估计一个微小物体的位姿：在0.4米距离处，该物体的大小仅为 $\sim 8 \times 18$ 像素；在1.5米处则仅为 $\sim 2 \times 5$像素，即分别占输入图像的0.9%和0.06%。将这些结果与面向人体姿态估计的现有技术水平 PULP-Frontnet 基线 [3], 进行比较可知，要达到同等复杂度，需要人体目标距离纳米无人机超过60米。

我们还在目标无人机执行随机路径而非预定义路径的场景中测试了该模型。在一次持续2分钟的实验中，观察无人机成功锁定并在整个实验期间保持其相对于目标无人机的相对位姿。该实验体现了集群作业中的一种情况：同伴无人机需要建立视线条件。随附的视频材料展示了该无人机持续保持视线的能力，时间约占无人机生命周期的一半。

最后，为了检验最终卷积神经网络的泛化能力，我们在一个前所未见的 室内环境中进行了额外的一次实地测试。该环境不同于前文各节实验所用的环境，且未出现在任何数据集中。目标无人机由人类飞行员远程操控，沿三维轨迹飞行，轨迹覆盖的空间体积约为$4\times4\times3$m。自主观察无人机开始执行任务时与目标无人机相距 1.2 m。实验结果见补充视频。由于该环境（即学生宿舍）未配备运动跟踪系统，我们无法计算位置跟踪性能的定量指标。观察无人机表现符合预期，成功跟踪目标，并在保持设定的相对距离 x=0.8 m 的同时，使目标持续处于视场内 53s。相比之下，根据具体飞行情况，静止不动的观察无人机将在 ~2 至 10 s 后失去目标。

V. 结论

本文研究了两架资源受限纳米无人机之间的单目相对位姿估计这一感知问题。我们提出了一种基于深度学习、仅使用视觉信息的垂直集成系统。在对多种（即 17 种）卷积神经网络进行表征后，我们选择 PULP-Frontnet 模型在纳米无人机上运行。我们完成了从数据集采集、数据增强、量化到系统优化的端到端开发与部署流程。总之，结果表明，仅使用低分辨率单色图像，就能对距离最远达 ~2 米、尺寸为 10 厘米的目标纳米无人机实现精确定位（平均控制误差为 17 厘米），并具有长续航实地性能，实验时长最长可达 2 分钟。

参考文献

[1] S. Li、C. De Wagter 和 G. C. H. E. De Croon，“采用高效深度神经网络的自监督单目多机器人相对定位”，载于 2022 年机器人与自动化国际会议（ICRA），2022 年 5 月，第 9689–9695 页。[2] D. Palossi、J. Singh、M. Magno 和 L. Benini，“纳米级无人机上的目标跟随”，载于 2017 年第七届 IEEE 传感器与接口进展国际研讨会（IWASI）。IEEE，2017 年，第 170–175 页。[3] D. Palossi、N. Zimmerman、A. Burrello、F. Conti、H. M¨ uller、L. M. Gambardella、L. Benini、A. Giusti 和 J. Guzzi，“超低功耗自主飞行纳米无人机上的全机载人工智能驱动人类位姿估计”，IEEE物联网期刊，第 1–1 页，2021 年。[4] A. Breitenmoser、L. Kneip 和 R. Siegwart，“一种基于单目视觉的六维机器人相对定位系统”，载于 2011 年 IEEE/RSJ智能机器人与系统国际会议，2011 年，第 79–85 页。[5] A. Carrio、S. Vemprala、A. Ripoll、S. Saripalli 和 P. Campoy，“使用深度图的无人机检测”，载于 2018 年 IEEE/RSJ智能机器人与系统国际会议（IROS），2018 年，第 1034–1037 页。[6] F. Schilling、F. Schiano 和 D. Floreano，“室外环境中的基于视觉的无人机编队”，IEEE机器人与自动化快报，第 6 卷，第 2 期，第 2954–2961 页，2021 年。[7] A. Milella、F. Pont 和 R. Siegwart，“使用立体视觉的协作机器人基于模型的相对定位”，-，第 8 页，2005 年。[8] M. J. Schuster、K. Schmid、C. Brand 和 M. Beetz，“多机器人团队的基于分布式立体视觉的六维定位与建图”，《现场机器人学》，第 36 卷，第 2 期，第 305–332 页，2019 年 3 月。 [在线]。可访问：https://onlinelibrary.wiley.com/doi/10 1002 21812 。/rob。T. [9] ¨ Schuppstuhl、K. Tracht 和 A. Raatz 编，《装配、搬运与工业机器人科学学会年刊 2021》。Cham：施普林格国际出版公司，2022 年。 [在线]。可访问：https://link.springer.com/10 1007 978 3 030 74032 0 。/ - - - - A. Wasik、[10] R. Ventura、J. N. Pereira、P. U. Lima 和 A. Martinoli，“多机器人系统基于激光雷达的相对位置估计与跟踪”，载于《Robot 2015：第二届伊比利亚机器人会议》，L. P. Reis、A. P. Moreira、P. U. Lima、L. Montano 和 V. Mu˜ noz-Martinez 编。Cham：施普林格国际出版公司，2016 年，第 3–16 页。[11] J. F. Roberts、T. Stirling、J.-C. Zufferey 和 D. Floreano，“用于室内飞行机器人的三维相对定位传感器”，《自主机器人》，第 33 卷，第 1 期，第 5–20 页，2012 年 8 月。 [在线]。可访问：https://doi.org/10 1007 10514 012 9277 0 。/s - - - F. Shan、H. Huo、J. [12] Zeng、Z. Li、W. Wu 和 J. Luo，“用于动态密集网络的超宽带群体测距协议”，IEEE/ACM网络学报，第 1–15 页，2022 年，会议名称：IEEE/ACM网络学报。[13] V. Niculescu 、 D. Palossi 、 M. Magno 和 L. Benini，“ 纳米无人机对传感器节点进行节能、精确的基于超宽带三维定位”，IEEE物联网期刊，第 1–1 页，2022 年。[14] F. Vanegas 和 F. Gonzalez，“在复杂无GPS环境中应对传感器与环境不确定性，实现无人机导航”，《传感器》，第 16 卷，第 5 期，2016 年。 [在线]。可访问：https://www.mdpi.com/1424 8220 16 5 666 - / / / D. Pickem、P. Glotfelter、[15] L. Wang、M. Mote、A. Ames、E. Feron 和 M. Egerstedt，“Robotarium：可远程访问的群体机器人研究测试平台”，载于 2017 年 IEEE机器人与自动化国际会议（ICRA），2017 年 5 月，第 1699–1706 页。[16] S. Mahendran、H. Ali 和 R. Vidal，“使用卷积神经网络进行三维姿态回归”，载于 IEEE计算机视觉国际会议研讨会论文集，2017 年，第 2174–2182 页。[17] M. e. a. Achtelik，“用于无GPS室内环境下自主直升机的立体视觉与激光里程计。”国际光学工程学会，2009 年。 [在线]。可访问：https://dspace.mit.edu/handle/1721 1 52660 。/ N. Kirchner 和 T. [18] Furukawa，“室内无人机的红外定位”，-，2005 年 1 月。[19] V. Walter、M. Saska 和 A. Franchi，“使用紫外LED标记物快速实现无人机相互相对定位”，载于 2018 年无人机系统国际会议（ICUAS），2018 年，第 1217–1226 页。[20] A. Carrio、S. Vemprala、A. Ripoll、S. Saripalli 和 P. Campoy，“使用深度图的无人机检测”，载于 2018 年 IEEE/RSJ智能机器人与系统国际会议（IROS），2018 年，第 1034–1037 页。

[21] F. Schilling、F. Schiano 和 D. Floreano，“室外环境中的基于视觉的无人机编队”，*IEEE*机器人与自动化快报，第 6 卷，第 2 期，第 2954–2961 页，2021 年。[22] D. Palossi、F. Conti 和 L. Benini，“面向自主纳米无人机的开源、开放硬件深度学习驱动视觉导航引擎”，载于*2019*年第*15*届传感器系统中的分布式计算国际会议（*DCOSS*），2019 年，第 604–611 页。[23] W. Zhao、A. Goudar、J. Panerati 和 A. P. Schoellig，“资源受限移动机器人超宽带定位的基于学习的偏差校正”，*arXiv*预印本 *arXiv:2003.09371*，2020 年。[24] F. Conti，“技术报告：用于部署模型的 NEMO 深度神经网络量化 ”，*CoRR*，第 abs/2004.05930 卷，2020 年。 [在线]。可访问：https://arxiv.org/abs/2004 05930 。[25] J. Choi、Z. Wang、S. Venkataramani、P. I.-J. Chuang、V. Srinivasan 和 K. Gopalakrishnan，“Pact：用于量化神经网络的参数化裁剪激活”，*arXiv*预印本 *arXiv:1805.06085*，2018 年。[26] A. Burrello、A. Garofalo、N. Bruschi、G. Tagliavini、D. Rossi 和 F. Conti，“Dory：在低成本物联网微控制器上自动端到端部署真实世界深度神经网络”，*IEEE*计算机学报，第 70 卷，第 8 期，第 1253–1268 页，2021 年。[27] N. J. Nagelkerke 等，“关于决定系数一般定义的说明”，*Biometrika*，第 78 卷，692，1991 3，691第 – 页。[28] L. Sigrist、A. Gomez、R. Lim、S. Lippuner、M. Leubin 和 L. Thiele，“RocketLogger：用于物联网设备原型开发的移动功耗记录器：演示摘要”，载于第*14*届*ACM*嵌入式网络传感器系统会议论文集 *CD-ROM*，2016 年，第 288–289 页。[29] V. Niculescu、L. Lamberti、F. Conti、L. Benini 和 D. Palossi，“通过自动化端到端优化和深度神经网络部署提升自主纳米无人机的性能”，*IEEE*电路与系统新兴和精选主题期刊，第 11 卷，第 4 期，第 548–562 页，2021 年。