本文已被 IEEE ICRA 2023 会议接收并将于会上发表 ©2023 IEEE。

面向纳米无人机精准视觉位姿估计的深度神经网络架构搜索

E. Cereda[1], L. Crupi[2], M. Risso[2], A. Burrello[3],L. Benini[34],A. Giusti[1],D. Jahier Pagliari[2],和 D. Palossi[14]

arXiv:2303.01931v1 [cs.RO] 3 Mar 2023

摘要— **微型自主无人机（UAV）是一个新兴且热门的研究主题。它们的尺寸仅与手掌相当，却能抵达大型机器人无法进入的地点，并能在有人活动的环境中安全运行。这类机器人的机载电子设备简单（功耗低于 100 mW），因此成本尤其低廉、颇具吸引力，但要实现机载复杂智能仍面临重大挑战。在这项工作中，我们采用一种新颖的神经网络架构搜索（NAS）技术，自动识别出多个用于视觉位姿估计任务的帕累托最优卷积神经网络（CNN）。我们的工作表明，真实环境中经过实地测试的机器人应用可以切实利用 NAS 技术，针对小型无人机的特定硬件约束，自动、高效地优化卷积神经网络。我们部署了多个经 NAS 优化的卷积神经网络，并在一架配备并行超低功耗片上系统的 27 克 Crazyflie 纳米无人机上以闭环方式运行。我们的结果将最先进水平提升了一个台阶：现场控制误差降低 32%，同时在 10 mW 和 90 mW 功耗下，机载推理频率分别达到∼10 Hz 和 ∼50 Hz。**

补充视频材料

实地测试：https://youtu.be/dVCScckvcg8。

I. 引言

纳米级无人机（UAV）凭借其小巧的体型（小于 10 cm）和轻盈的重量（小于 40 g），正日益受到关注。这些特点使其能够执行敏感任务，例如在GPS拒止狭窄空间以及近人环境中作业。此外，这些平台搭载的传感、机械和计算子系统较为简化，因此与体型更大的同类设备相比，成本尤其低廉，也更具吸引力。然而，要实现完全自主，即无需外部/远程基础设施或计算资源，仍面临挑战；其原因在于这些设备的电子系统较为简化，例如计算功耗低于 100 mW，算力比移动电话处理器低1至2个数量级。

这些限制阻碍了标准算法流程的采用，而这类流程通常依赖内存/计算密集型的复杂规划/定位/建图算法，以及庞大的预训练感知卷积神经网络（卷积神经网络）。因此，在纳米无人机上运行这些方法并不可行。在此背景下，小型卷积神经网络日益成为自主纳米无人机的现有技术水平 [1], [2]。首先，它们可以针对来自极其有限的传感器的数据进行训练，例如超低功耗的小型摄像头，而这类摄像头的分辨率/动态范围较差。其次，卷积神经网络在推理时的计算量和内存需求可预测且固定。第三，系统设计者可以通过定义合适的网络架构，调整这些需求以适应机器人可用的资源。

在更大型机器人中集成卷积神经网络的机器人专家，往往会忽略第三个方面，而依赖 ResNet50 [3] 等标准现有技术水平架构，因为这类架构很容易适配可用资源。相比之下，在开发纳米级机器人系统时，全面定义合适的定制卷积神经网络至关重要。神经网络架构的选择决定了模型能否在机器人上运行，并直接影响两个决定机器人行为的关键参数：*i*）预测性能，以及*ii*）实时吞吐量。

多年来，要满足这一复杂优化场景提出的要求（即获得准确且可部署的卷积神经网络），唯一的办法是依靠启发式方法和经验法则，手动进行迭代式超参数调优。这种做法既繁琐，又容易出错，还十分耗时。如今，执行此类优化的首选 方法基于神经架构搜索（NAS）范式 [4], [5]。神经架构搜索工具能够在任意规模的搜索空间内，自动探索不同的网络拓扑结构和超参数设置。此外，许多新颖的神经架构搜索方法还能够直接优化复杂的成本函数，将回归性能与网络的计算复杂度等不同目标结合起来 [5], [6], [7], 。

在本研究中，我们采用一种新颖且计算高效的神经架构搜索技术 [7], ，能够在准确率与模型大小之间生成帕累托最优卷积神经网络架构，以优化基于视觉的人体姿态估计任务。首先，我们通过增强现有神经架构搜索引擎的功能作出贡献，使其能够探索两种不同的种子卷积神经网络：PULP- Frontnet [2] 和 MobileNetv1 [8]。随后，我们对多个卷积神经网络进行全面分析，并将其部署在目标纳米无人机机器人平台上。最终，所得到的卷积神经网络在至少一项指标上优于现有技术水平基线 [2] ，这些指标包括模型大小（最多缩小5.6× ）、速度（最多提升 1.5× ）和准确率（水平位移误差最多降低 32%）。最后，通过一项具有挑战性的实地测试验证了准确率的提升；该测试采用了前所未见的环境。

本研究部分获得阿联酋技术创新研究所（TII）安全系统研究中心（SSRC）以及瑞士国家科学基金会（SNSF）通过 NCCR Robotics 提供的支持。[1]E. Cereda、A. Giusti 和 D. Palossi 任职于瑞士卢加诺 6962 的达莱莫勒人工智能研究所、USI 和 SUPSIname.surname@idsia.ch[2]L. Crupi、M. Risso 和 D. Jahier Pagliari 任职于意大利都灵 10129 的都灵理工大学控制与计算机工程系 name.surname@polito.it[3]A. Burrello 和 L. Benini 任职于意大利博洛尼亚 40136 的博洛尼亚大学电气、电子与信息工程系name.surname@unibo.it[4]L. Benini 和 D. Palossi 也任职于瑞士苏黎世 8092 的 ETH 苏黎世集成系统实验室¨urich，Z¨urich，8092，瑞士

II. 相关工作

神经架构搜索。神经架构搜索工具通过自动探索由不同层和/或超参数组合构成的大量架构空间，协助设计人员在深度神经网络的设计阶段开展工作。在资源受限的平台上，这些工具通常会最小化一个目标函数，该函数同时取决于任务准确率和非功能性成本指标（例如内存占用、延迟或能耗）。早期的神经架构搜索算法基于进化算法（EA） [4] 和强化学习（RL） [5]。这些方法能够探索任意搜索空间，并优化任何成本函数：它们会迭代地采样一个网络，将其训练至收敛以评估性能，然后利用这些信息指导下一次采样。

然而，其极高的灵活性意味着较差的可扩展性，因为其计算需求非常高，即使是简单任务也需要数千个图形处理器小时。相反，可微神经架构搜索（DNAS） [9], [6]已被提出，以缓解这一问题。早期的可微神经架构搜索方法利用超网络，即其层的输出是对多个应用于相同输入的备选子层（例如不同类型的卷积）的输出进行加权求和所得的网络 [9]。在训练期间，根据取决于准确率和复杂度的损失函数，对分配给各个备选项的权重进行优化；搜索结束时，通过为每一层选择权重最大的备选项来得到最终网络。与 EA/RL 方法相比，可微神经架构搜索牺牲了优化目标定义上的部分灵活性，因为该目标必须是可微的，以换取仅需单次训练的搜索过程。然而，超网络的收敛*i*）并不简单，*ii*）仍然代价高昂（就图形处理器内存而言），因为每一层都需要实例化多个备选项。

基于掩码的可微神经架构搜索 是迈向轻量级架构搜索的又一步。它们用单路径 DNN 取代超网，这种网络通常称为种子网络 [10], [11], [7]。其搜索空间由从种子网络通过删减 得到的子架构组成（例如，删除每个卷积中的部分通道）。实践中，训练时使用可训练掩码来模拟这些子网络，掩码会剪除种子网络的部分结构。与基于超网的可微神经架构搜索一样，掩码会在训练过程中进行优化；但由于种子网络远小于超网，与常规 DNN 训练相比，时间和内存开销极小。一方面，基于掩码的可微神经架构搜索只能生成从种子网络派生的网络；另一方面，这使搜索空间的探索能够细致得多。例如，它们可以轻松生成通道数任意的卷积（例如，对于具有 32 个通道的种子网络，可生成 17、25 或 31 个通道）：采用超网方法时，这几乎难以实现 [7]。

纳米无人机。由于外形尺寸、有效载荷和可用能源有限，在纳米无人机上部署用于自主导航的卷积神经网络模型，会面临严峻的传感、计算和内存限制。基于商用现货（COTS）微控制器单元（MCU） [12], [13]的解决方案只能采用极精简的神经网络架构，以实现实时性能，例如 [13]以每帧 27 kMAC（乘加运算）在 100 Hz 的速率运行。因此，这些方法仅适用于低维输入信号，不适合处理相机图像。通过精心的软硬件协同设计，利用通用超低功耗多核片上系统（SoC） [15]、集成的软件部署流程 [16]、[17]以及经过人工调优的 CNN 架构，更先进的视觉方法 [2], [14]得以实现。这些技术突破使 PULP-Frontnet [2], 这一计算量为 14.7 MMAC（比 [13]大 3 个数量级）的模型能够在商用现货 Crazyflie 2.1 纳米无人机上以 48 Hz 的推理速率运行，同时功耗仅为 96 mW。

一个尚未解决的研究问题仍然围绕着卷积神经网络的复杂度/内存削减展开，因为解决这一问题将为以下目标铺平道路：*i*）在电池供电设备上更高效地利用能源，以及*ii*）实现多任务系统，而这些目标在纳米无人机上尚未实现。近期研究 [18], [19] 旨在以尽可能小的准确率下降降低卷积神经网络的内存占用，在仅损失 4% 准确率的情况下，计算量最多可降低 27× [19]，但代价仍然是对架构进行大量手动微调。在本文中，我们利用神经架构搜索技术，在机器人领域高效、自动地解决这一问题。

人体姿态估计。我们研究的机器人问题是：利用自主纳米无人机对人体目标进行单目相对姿态估计。尽管最先进的计算机视觉方法具有出色的准确率 [20], [21]、[22]，但由于所需的计算资源量 [22] （∼10s GMAC），它们仍远非纳米无人机所能承受。迄今为止，PULP-Frontnet 卷积神经网络 [2] 是少数几个能够在 Crazyflie 纳米无人机上实时、完全执行人体姿态估计任务的实例之一。复杂的卷积神经网络能够估计人体的完整骨架 [21]，甚至是稠密的三维网格表示 [22]，而 PULP-Frontnet 则将预测结果简化为空间三维点（x， y， z）以及相对于重力 z 轴的旋转角（ϕ）。MobileNetv1 [8] 是另一种广泛用于基于视觉的分类任务的现有技术水平卷积神经网络；事实也证明，当其精简到分配给我们纳米无人机的相同功耗范围时，仍能保持较高准确率（例如，在 ImageNet 上达到 68.2% 的 Top-1 准确率） [23], [17]。基于这些原因，我们选择 PULP-Frontnet 和 MobileNetv1 作为神经架构搜索的种子模型，并将 PULP-Frontnet 模型称为现有技术水平基线（F_{SoA}），用于实地比较。

III. 方法

A. 种子网络深度神经网络

在这项工作中，我们使用基于掩码的可微神经架构搜索工具来优化三个种子模型：一个基于浅层 PULP-Frontnet 架构，另外两个基于更深的 MobileNetv1。PULP- Frontnet 由 6 个卷积层组成，需要 304 k 个参数，最多执行 14.7M 次乘加运算。该网络的设计最初在 [2]中提出，专为部署在纳米无人机上而设计。另一方面，MobileNetv1 [8]是一个更深的网络，由 27 个不同的卷积层构成，其与 PULP-Frontnet 的主要区别在于

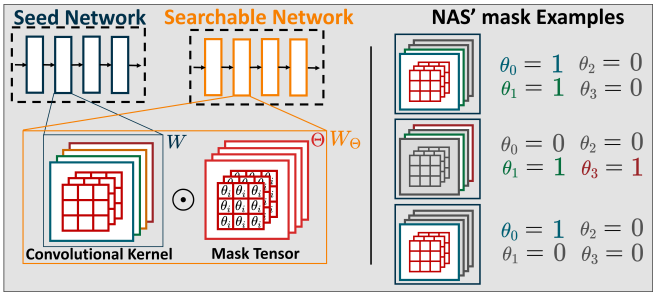

图 1。左：所提出的掩码方案。右：在包含四个滤波器的层上，通道搜索过程的三种可能结果。 $\theta_0 = \theta_1 = 1$ 和$\theta_2 = \theta_3 = 0$ 表示保留前两个通道，而移除后两个通道。在第二个示例中，仅保留第二个和最后一个通道；在最后一个示例中，仅保留第一个通道。

采用可分离的深度卷积层和逐点卷积层，而非传统卷积层。我们的两个 MobileNetv1 种子网络在宽度乘数上有所不同。宽度乘数是一种控制模型中各处特征图大小的超参数：我们采用 1.0（$M^{1.0}$）和 0.25（$M^{0.25}$）两种宽度乘数，对应的参数量分别为 3.21 M 和 204k。

B. 网络架构搜索

我们使用一种基于掩码的可微神经架构搜索工具来优化网络架构，即时间剪枝 （PIT） [7], ，其目标是搜索更小、更轻量的网络，同时保持与种子模型几乎相同的准确率。PIT 的内存和执行时间开销较低（仅占训练时间的 ~30 %），比其他神经架构搜索引擎低几个数量级 [4], [5], [9]。此外，我们的任务有可用的优质参考架构，可将其用作种子网络，这与基于掩码的方法十分契合；该方法能够高效地探索种子网络中包含的不同子网络。

我们提出的神经架构搜索会对二维卷积神经网络所有卷积层中的输出通道数进行细粒度搜索。该方法源自 [7] 中最初针对一维时序卷积网络提出的通道搜索方案。图 1 总结了我们的神经架构搜索功能：对于种子网络中的每个卷积层，提取其具有 C_{out} 个输出通道的权重张量 W，并构建相应的可搜索张量 W_Θ 如下：

$$W_\Theta = W \odot \mathcal{H}(\theta) \tag{1}$$

其中， $\odot$ 表示哈达玛积， θ 是一个包含 C_{out} 个元素的可训练掩码向量，其中每个分量 θ_i表示特定通道的掩码； $\mathcal{H}$ 是用于将 θ二值化的海维赛德阶跃函数。根据 $\mathcal{H}(\theta_i)$的二值化值， W 的第i个输出通道可能会被保留（$\mathcal{H}(\theta_i) = 1$）或移除（$\mathcal{H}(\theta_i) = 0$）。因此，PIT能够找到相对于种子网络仅在通道数量上有所差异的模型。与 [7], 中提出的原始版本相比，我们还支持联合探索逐点层和深度层的通道。这两种层是深度可分离卷积块中的主要层，也是MobileNet架构的基本组件。

为此，我们使用共享掩码来选择这两种层的通道数量，因为深度卷积的输出通道完全由其前面的逐点层决定。

所得网络会被置于常规训练循环中，其中 W 和 θ 协同训练，以解决以下优化问题：

$$\min_{W,\theta} \mathcal{L}(W;\theta) + \lambda\mathcal{R}(\theta) \tag{2}$$

其中， $\mathcal{L}$ 是任务特定损失函数（即与种子卷积神经网络所使用的损失函数相同）， $\mathcal{R}$ 是表示待优化成本指标的正则化损失项（例如参数数量、操作数量等）。 λ 是所谓的正则化强度，即用于平衡这两个损失函数的手动调节标量项。一旦定义了公式 2 中的损失项，λ 就成为设计者可以调节的主要旋钮，用于推动搜索趋向于精度更高或成本更低的网络。针对我们的使用场景，我们取 $\lambda \in [5 \cdot 10^{-11} : 5 \cdot 10^{-5}]$。作为正则化损失 $\mathcal{R}$，我们使用了网络参数数量关于神经架构搜索掩码值的可微估计。这样，我们便可以将探索阶段偏向于同时具备较小规模和较高精度的架构。

C. 系统设计

机器人平台。 我们的目标机器人平台是 Bitcraze Crazyflie 2.1[1], a，这是一款重 27 克的纳米四旋翼。该无人机配备一颗主 STM32 微控制器，负责状态估计和控制回路等基本功能。在实地部署中，我们为其扩展了一块重 5 克的商用 AI-deck 配套板 [1]。AI-deck 配备了另一颗微控制器：GreenWaves Technologies GAP8。该芯片基于 RISC-V 多核片上系统，体现了并行超低功耗范式 [15] 。这两颗处理器通过双向 UART 接口通信。GAP8 采用两个电源域：单核片上控制器(FC)负责协调与外部内存/传感器的交互，并将计算密集型内核卸载到另一个 8 核集群 (CL)域中。该片上系统的存储层次结构包括所有集群核心共享的 64 kB 低延迟 L1 内存，以及位于 FC 域内的 512 kB L2 存储器。GAP8 还配备了两个 DMA 引擎，可高效地自动执行内存与外部外围设备之间的数据传输；这些设备包括 8 MB 片外 DRAM 和 QVGA 单色摄像头，二者均搭载于 AI-deck 上。不过，它既不提供数据缓存，也不配备硬件浮点单元，因此分别要求显式进行数据管理，并采用整数定点量化运算。

部署工具。 为满足平台的要求，我们利用两种工具来*i*)训练纯整数神经网络，以及*ii*)为目标网络自动且最优地生成 C 代码。第一种工具是开源的 NEMO 库 [16], ，它基于 PyTorch 框架，用于通过三个连续步骤训练网络。首先，NEMO

[1]https://www.bitcraze.io/products/crazyflie-2-1/

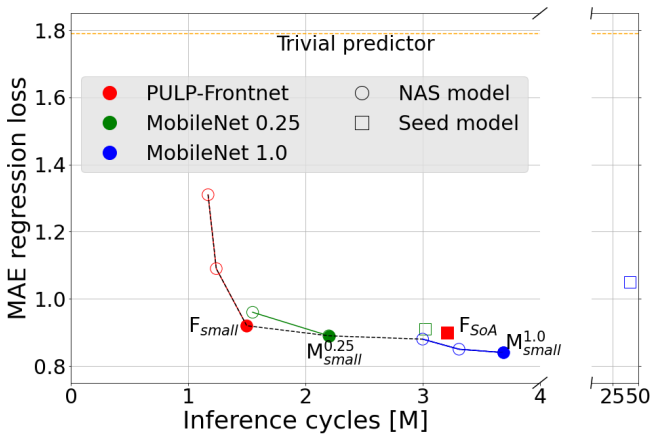

图2. 从神经架构搜索中提取的网络在时钟周期数与平均绝对误差空间中的帕累托曲线（越低越好）。

训练一个全精度 浮点网络，以最小化姿态估计向量的 L1 损失之和（x， y， z， ϕ）。我们使用随机梯度下降优化器，以 0.001 的学习率 训练 100 轮，并在最后选取验证损失最低的模型。随后，我们将该模型转换为伪量化 网络。在此阶段，权重和激活值仍以 float32 数值表示，但其大小只能取离散集合中的值（对于 8 位量化，该集合包含 256 个值）。鉴于 GAP8 的优化内核提供了相应支持，我们在此微调步骤中使用逐层线性量化和 PACT [24]。我们选择使用 8 位量化，因为 RISC-V 指令集架构扩展中的 SIMD 运算原生支持该量化方式，因此每个周期可执行四次 int8 乘加运算。为最大限度地提高网络的准确率，我们使用浮点权重初始化伪量化网络，并使用相同的参数额外训练 100 轮。最后一步是创建可部署的整数 网络。与伪量化网络相比，所有张量均由整数和浮点缩放因子表示。张量$\mathbf{t}$ 近似为

$$\mathbf{t} \approx \varepsilon_{\mathbf{t}} \cdot \mathbf{t}^{*}\,, \tag{3}$$

其中，$\mathbf{t}$ 是伪量化张量，$\mathbf{t}^{*}$ 是仅含整数的张量，$\varepsilon_{\mathbf{t}}$ 是缩放浮点因子。因此，该网络可以仅通过对整数值进行乘法和累加，完全在整数域中运行。

卷积神经网络部署基于开源 PULP-NN 库 [25]，，该库包含经过优化的 8 位定点算术内核。具体而言，它利用 GAP8 片上系统中集群的八个通用核心以及指令集架构专用优化，以最大限度地提升内核性能。理论峰值为每周期 32 次乘加运算（得益于 int8 SIMD，每个核心每周期可执行 4 次乘加运算，8 个核心合计），PULP-NN 对正方形图像可达到每周期 15.6 次乘加运算的峰值。需要注意的是，该库的峰值是在寄存器文件中完美复用数据的情况下测得的，从而减少了每次乘加运算所需的加载和存储操作。如果寄存器文件中没有存储可复用的数据，则需要两次加载乘加运算的输入，并将结果存储到内存中一次，从而使理论峰值降至每周期 8 次乘加运算。另一方面，PULP-NN 内核假设数据存储在共享的 64 kB L1 内存中，因此其适用范围仅限于较小的层。

表I 测试集实验结果

网络	MAE				R^2分数			
	x	y	z	ϕ	x	y	z	ϕ
简单基线	0.46	0.61	0.17	0.55	0.0	0.0	0.0	0.0
$\mathbf{F}_{SoA}$ [2]	0.19	0.18	0.09	0.44	0.80	0.65	0.55	0.26
$\mathbf{F}_{small}$	0.20	0.20	0.09	0.44	0.79	0.65	0.51	0.24
$\mathbf{M}^{0.25}_{small}$	**0.17**	0.20	0.08	0.44	**0.84**	0.63	0.62	0.22
$\mathbf{M}^{1.0}_{small}$	0.17	**0.18**	**0.07**	**0.42**	0.83	**0.67**	**0.62**	**0.28**

为应对这一限制，我们采用了另一款开源工具 DORY [17]。该工具会自动生成基于模板的 C 代码，用于封装 PULP-NN 内核、管理 GAP8 的不同层级内存（即 L1、L2 和外部随机存取存储器），并协调张量的移动。具体而言，DORY 采用分块方法，将层拆分成较小的节点，使其张量能够放入系统的 L1 内存中。因此，内核可以直接在这些节点上执行。DORY 实际生成的 C 例程会：i）执行小型节点的内核，且张量存储在 L1 中；ii）对张量从 L2 到 L1 的传输进行双缓冲，从而始终为内核执行准备好数据。请注意，由于直接存储器访问（DMA）不会阻塞，内核调用始终与异步 DMA 调用重叠执行。

闭环控制。我们将模型部署为与 [2]，相同的闭环控制系统的一部分，该系统能够使无人机保持在人员前方的期望三维姿态。系统涉及四个主要组件：*i*）卷积神经网络模型根据单张图像输出某一时刻的姿态估计；*ii*）卡尔曼滤波器生成平滑且时间上一致的姿态序列；*iii*）速度控制器根据受试者当前的位姿计算无人机的期望位姿，并生成速度设定值，使无人机移动到期望位姿。*iv*）Crazyflie 的底层控制，负责电机驱动和稳定控制。我们采用卡尔曼滤波器，并将其与模型的四个输出解耦，同时假设过程噪声协方差矩阵和观测噪声协方差矩阵均为对角矩阵。速度控制器也分为线速度控制和角速度控制两部分：前者旨在到达指定的目标位置，后者则使受试者保持在图像帧中央。

IV. 结果

A. 神经架构搜索帕累托分析

在图 2 中，我们展示并分析了神经架构搜索算法找到的架构。我们在一个坐标轴上比较推理延迟（时钟周期数），在另一个坐标轴上报告平均绝对误差（平均绝对误差），该误差表示网络预测值与真实值在 (x， y， z， ϕ) 上的 L1 误差之和

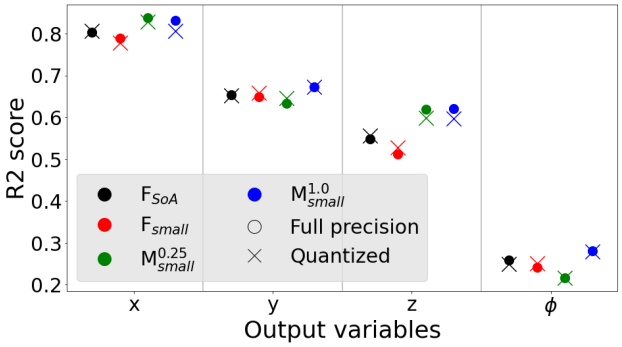

图 3. 四种部署模型的 R^2分数（越高越好）。

。简单基线预测器，即始终将每个输出预测为其在测试集中的均值的网络，代表了模型平均绝对误差的下界。从三个种子网络架构出发，神经架构搜索发现了八个新模型，其中大多数位于平均绝对误差与周期数构成的空间中的全局帕累托前沿上。找到的架构执行周期数介于 1.27M 至 3.69M 之间，对应的平均绝对误差值为 1.31 至 0.84。具体而言，PULP-Frontnet 模型位于帕累托曲线最左侧，体积非常小，但准确度较低。中间部分由源自 MobileNet 0.25× 种子网络的模型构成。最后，准确度最高的架构源自 MobileNet 1.0×。这一种子架构过大，无法装入 GAP8 的 L2 存储器。不过，我们的神经架构搜索算法能够将其缩小到足以找到出色的解：性能绝对领先的模型，其延迟相比 F_{SoA} [2] 仅增加 15%。

从全局帕累托前沿中，我们选择了四种模型进行部署和进一步分析。第一种是 F_{SoA}，对应于原始的 PULP-Frontnet，即当前最先进的模型，也是我们的基线模型。 F_{small} 架构是速度最快的模型，同时其平均绝对误差与F_{SoA}大致相当。 $M^{1.0}_{small}$ 是神经架构搜索找到的最准确架构，但在延迟方面的开销也最高。$M^{0.25}_{small}$ 代表了两项指标之间最均衡的权衡，其性能均优于或等同于 F_{SoA}。在接下来的章节中，我们将进一步分析所选架构在测试集上的性能，以及它们在闭环系统中的现场表现。

B. 回归性能

在表I和图3中，我们细分了模型在四个输出变量（x, y, z, ϕ）上的回归性能。对于每个输出，除了报告平均绝对误差（MAE）外，我们还报告决定系数 R^2，这是一种标准的无量纲指标，表示模型所解释的目标变量方差占比[2]。与平均绝对误差相比，R^2 分数能够衡量回归器的质量，而不受目标变量方差的影响，因此更适合

[2] $R^2 = 1 - \sum_i (y_i - \hat{y}_i)^2 / \sum_i (y_i - \bar{y})^2$，其中 y_i 表示真实输出，$\hat{y}_i$ 表示模型对每个测试样本的预测，i表示 $\bar{y}$ 真实输出的均值。完美的回归器对应于 $R2 = 1.0$，而简单基线回归器的取值为 $R2 = 0.0$。

表II 单帧推理的计算量和内存占用（**F**：PULP-FRONTNET，**M**：MOBILENET）。

网络	参数	MAC	周期数	内存	P [毫瓦]	T [帧/秒]
$\mathbf{F}_{SoA}$ [2]	304 千	14.7 百万	3.2 百万	499 千字节	92.2	45.3
$\mathbf{F}_{small}$	44 k	7.6 M	1.5 M	231 kB	81.3	71.6
$\mathbf{M}^{0.25}_{small}$	65 k	7.4 M	2.2 M	591 kB	86.9	51.2
$\mathbf{M}^{1.0}_{small}$	54 k	12.4 M	3.7 M	415 kB	88.3	32.7

用于比较不同变量之间的回归性能。所有输出上的性能趋势均与聚合损失一致，证实了 $M^{1.0}_{small}$ 在所有输出中表现最佳，唯独在 x上， $M^{0.25}_{small}$的表现略好。在图3中可以看到，所有模型对 x 和 y 的预测效果最好，而 ϕ 最难估计，这与背景工作中的发现一致。

C. 机载性能评估

我们评估了第IV-A节中选定的卷积神经网络的机载性能。为准确确定计算、功耗和内存需求，我们对模型进行了性能分析，在 GAP8 片上系统上运行模型，并使用 RocketLogger 数据记录仪 [26] （64 ksps）。在这些实验中，我们测试了三个片上系统工作点：最低功耗（FC @25 MHz CL@25 MHz VDD@1 V）、能效最高（ FC@25 MHz CL@75 MHz VDD@1 V）和性能最高（ FC@250 MHz CL@170 MHz VDD@1.2 V），如 PULP-Frontnet 基线所示 [2]。表II总结了对四种卷积神经网络的分析结果，显示神经架构搜索模型显著减少了参数量（相较于 F_{SoA}最多减少 85%）、乘累加运算次数、时钟周期数和内存占用（这三项均减少约 50%）。此外，还报告了实地实验所用工作点（最高性能）下的功耗和吞吐量数据。图4展示了三个工作点下功耗与吞吐量之间的关系。我们发现，两个较小的模型 $M^{0.25}_{small}$和 F_{small} 在这两方面均优于基线模型 F_{SoA}，因此能效更高。图5详细展示了整个系统的功耗构成。Crazyflie 电子设备加上 AI-deck 占总功耗预算的 4.8%，几乎耗尽了可分配给感知和计算的功耗 [27]。由于 GAP8 占其中的 24%，因此显然有必要尽可能优化其工作负载。

D. 实地实验评估

我们进一步通过一次闭环实地实验验证所提出的网络。我们请一名受试者沿预先设定的路径行走，同时要求无人机保持在其前方、眼睛高度的位置，并与其相距 $\Delta = 1.3$ 米。为保持一致，我们采用 [2], 提出的相同测试设置：路径全长 50 秒，包含 8 个难度逐渐增加的阶段（朝不同方向直线行走、沿曲线行走，以及原地旋转）。为确保实验不同轮次之间具有可重复性，我们请受试者完全忽略无人机的行为，改为按照节拍器的节拍同步迈步。

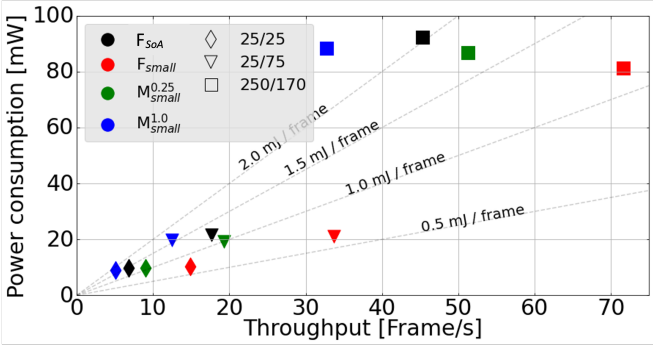

图 4. 四种模型在不同工作点的吞吐量与功耗。

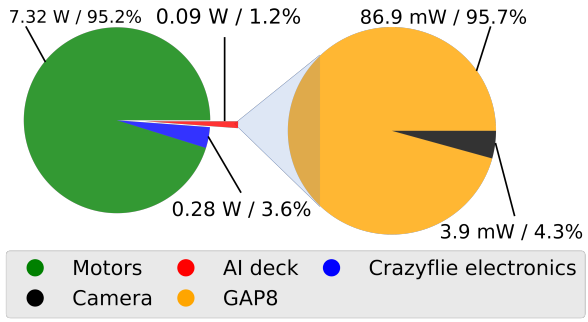

图 5. 纳米无人机运行 $\mathrm{M}^{0.25}_{small}$ 模型时的功耗分解。

实验中的受试者和环境均未出现在训练集中，以此考验模型的泛化能力。我们对四种模型分别进行了四轮实验，此外还进行了一轮基线运行：此时利用动作捕捉系统提供的受试者姿态精确信息来控制无人机。测试飞行总计 17 次。

表 III 总结了我们在本次实验中的结果。我们提供了系统性能三个方面的定量指标：整体路径完成度、推理准确率和控制准确率。对于每个模型，我们通过四次运行的总飞行时间和平均路径完成百分比来衡量路径完成度；一旦人员完全离开摄像机视场，我们便中断该次运行。在本次实验所用的、极具挑战且从未见过的环境中，两个 PULP-Frontnet 模型难以完成路径（尤其是 F_{small}，平均仅完成 35%），而两个 MobileNet 模型则始终能够持续跟踪直至结束。对于未完成的运行，我们用星号标记相应指标，因为这些指标无法与其他运行直接比较。

推理准确率随后单独评估各模型，衡量其相对于无人机准确估计受试者位置的能力。与参考研究一样，本文不考虑 z 分量，因为在我们的任务中，目标高度近似恒定。我们报告平均绝对误差（MAE），以便与第 IV-B 节报告的测试集离线性能直接比较。不出所料，绝对性能由于环境不同而显著下降，

表 III 现场实验结果

网络	飞行时间 [s]	已完成路径 [%]	MAE x	MAE y	MAE ϕ	平均姿态误差 e_{xy} [米]	平均姿态误差 e_θ [弧度]
动作捕捉	165	100	0.0	0.0	0.0	0.18	0.21
F_{SoA} [2]	140	85	0.33*	0.12*	0.77*	0.72*	0.78*
F_{small}	58	35	0.81*	0.53*	0.55*	0.65*	0.42*
$\mathbf{M}^{0.25}_{small}$	**165**	**100**	**0.25**	**0.11**	**0.52**	**0.49**	**0.59**
$\mathbf{M}^{1.0}_{small}$	**165**	**100**	0.31	0.13	0.52	0.58	**0.59**

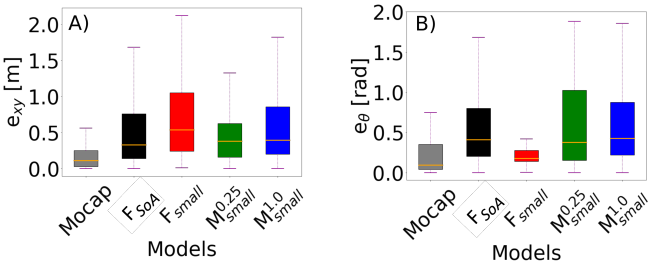

图6。实地控制误差分布（越低越好）。箱线图的须线标记数据的 5th 和 95th 百分位数。

同时，我们看到，最佳的现场回归性能来自 $M^{0.25}_{small}$，而不是测试集上的 $M^{1.0}_{small}$。一种解释是，前一个模型的参数数量较少，因此不易过拟合，从而具有更好的泛化能力。

最后，控制准确率评估整个系统在受试者沿路径移动时对其进行跟踪的精度。我们将无人机的实际位姿与期望位姿进行比较，测量两种误差：绝对位置误差 e_{xy}（即水平面上两个位姿之间的距离）和绝对角度误差 e_θ（即方向之间的差异）。$M^{0.25}_{small}$ 在这两项指标上均表现最佳，证明其在实地表现最为出色。在图6中，$M^{0.25}_{small}$ 是绝对位置误差方差最低的模型，进一步解释了其更优的性能。此外，通过直观观察无人机的行为可以发现，$M^{0.25}_{small}$ 模型明显比基线 F_{SoA} 更加准确。作为对结果的补充，我们提供了四个模型在实地实验中的行为视频。

V. 结论

本文介绍了一个实用案例：将NAS技术应用于纳米无人机上的一项具有挑战性的机器人视觉感知任务——人体姿态估计。我们从两个种子卷积神经网络（即 PULP-Frontnet和MobileNetv1）出发，选出四个帕累托最优模型，部署在资源受限（即功耗低于100 mW）的纳米四旋翼上。我们通过全面分析评估卷积神经网络的能力：从其在独立测试集上的回归性能、机载性能评估（功耗和吞吐量），到在前所未见的环境中进行的现场闭环测试。我们的最佳模型将现场控制误差降低了32%，以 ~50 Hz@90mW的实时推理速率提升了现有技术水平。

参考文献

[1] D. Palossi、F. Conti和L. Benini，“一种用于自主纳米无人机的开源、开放硬件深度学习视觉导航引擎”，载于《2019年第15届传感器系统分布式计算国际会议》（DCOSS）论文集，2019年，第604–611页。[2] D. Palossi、N. Zimmerman、A. Burrello、F. Conti、H. Muller、L. M. Gambardella、L. Benini、A. Giusti和J. Guzzi，“在超低功耗自主纳米无人机上完全机载的AI人体姿态估计”，《IEEE物联网期刊》，第9卷，第3期，第1913–1929页，2022年。[3] K. He、X. Zhang、S. Ren和J. Sun，“用于图像识别的深度残差学习”，载于《IEEE计算机视觉与模式识别会议》（CVPR）论文集，2016年6月。[4] E. Real、S. Moore、A. Selle、S. Saxena、Y. L. Suematsu、J. Tan、Q. V. Le和A. Kurakin，“图像分类器的大规模进化”，载于《国际机器学习会议》。PMLR，2017年，第2902–2911页。[5] M. Tan、B. Chen、R. Pang、V. Vasudevan、M. Sandler、A. Howard和Q. V. Le，“Mnasnet：面向移动设备的平台感知神经架构搜索”，载于《IEEE/CVF计算机视觉与模式识别会议》论文集，2019年，第2820–2828页。[6] H. Cai、L. Zhu和S. Han，“Proxylessnas：针对目标任务和硬件直接进行神经架构搜索”，arXiv预印本 arXiv:1812 00332，2018年。。[7] M. Risso、A. Burrello、F. Conti、L. Lamberti、Y. Chen、L. Benini、E. Macii、M. Poncino和D. J. Pagliari，“边缘端时间卷积网络的轻量级神经架构搜索”，《IEEE计算机学报》，2022年。[8] A. G. Howard、M. Zhu、B. Chen、D. Kalenichenko、W. Wang、T. Weyand、M. Andreetto和H. Adam，“MobileNets：面向移动视觉应用的高效卷积神经网络”，arXiv预印本 arXiv:1704.04861，2017年。[9] H. Liu、K. Simonyan和Y. Yang，“DARTS：可微分架构搜索”，arXiv预印本 arXiv:1806.09055，2018年。[10] A. Gordon、E. Eban、O. Nachum、B. Chen、H. Wu、T.-J. Yang和E. Choi，“MorphNet：快速简便的资源受限深度网络结构学习”，载于《IEEE计算机视觉与模式识别会议》论文集，2018年，第1586–1595页。[11] A. Wan、X. Dai、P. Zhang、Z. He、Y. Tian、S. Xie、B. Wu、M. Yu、T. Xu、K. Chen等，“FBNetV2：针对空间和通道维度的可微分神经架构搜索”，载于《IEEE/CVF计算机视觉与模式识别会议》论文集，2020年，第12 965–12 974页。[12] G. Shi、W. H¨onig、Y. Yue和S.-J. Chung，“Neural-Swarm：利用学习到的交互实现近距离多旋翼飞行器的分布式控制”，载于《2020年IEEE机器人与自动化国际会议》（ICRA）论文集，2020年，第3241–3247页。[13] W. Zhao、J. Panerati和A. P. Schoellig，“面向资源受限移动机器人的到达时间差超宽带定位的学习式偏差校正”，《IEEE机器人与自动化快报》，第6卷，第2期，第3639–3646页，2021年。[14] D. Palossi、F. Conti和L. Benini，“一种用于自主纳米无人机的开源、开放硬件深度学习视觉导航引擎”，载于《2019年第15届传感器系统分布式计算国际会议》（DCOSS）论文集，2019年，第604–611页。[15] M. Gautschi、P. D. Schiavone、A. Traber、I. Loi、A. Pullini、D. Rossi、E. Flamand、F. K. G¨urkaynak和L. Benini，“适用于可扩展物联网终端设备的带DSP扩展的近阈值RISC-V内核”，《IEEE超大规模集成电路系统学报》，第25卷，第10期，2017年10月。[16] F. Conti，“技术报告：用于部署模型的Nemo深度神经网络量化”，arXiv预印本 arXiv:2004.05930，2020年。[17] 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进行实际应用中的深度神经网络端到端部署”，《IEEE计算机学报》，第1–1页，2021年。[18] M. Navardi、A. Shiri、E. Humes、N. R. Waytowich和T. Mohsenin，“一种用于高效基于视觉的自主无人机导航的优化框架”，载于《2022年IEEE第4届人工智能电路与系统国际会议》（AICAS）论文集，2022年，第304–307页。[19] L. Lamberti、V. Niculescu、M. Barci´s、L. Bellone、E. Natalizio、L. Benini和D. Palossi，“Tiny-PULP-Dronets：压缩神经网络，以便在多任务自主纳米无人机”，载于*2022 IEEE第4*届人工智能电路与系统国际会议*(AICAS)*，2022年，第287–290页。[20] X. Sun、B. Xiao、F. Wei、S. Liang 和 Y. Wei，“积分人体姿态回归”，载于欧洲计算机视觉会议论文集*(ECCV)*，2018年9月。[21] D. C. Luvizon、D. Picard 和 H. Tabia，“使用多任务深度学习进行二维/三维姿态估计和动作识别”，载于*2018 IEEE/CVF*计算机视觉与模式识别会议，2018年，第5137–5146页。[22] R. A. G¨uler、N. Neverova 和 I. Kokkinos，“Densepose：野外密集人体姿态估计”，载于*IEEE*计算机视觉与模式识别会议论文集 *(CVPR)*，2018年6月。[23] A. Capotondi、M. Rusci、M. Fariselli 和 L. Benini，“Cmix-nn：面向内存受限边缘设备的混合低精度卷积神经网络库”，*IEEE*电路与系统 *II* 汇刊：简报，第67卷，第5期，第871–875页，2020年。[24] J. Choi、S. Venkataramani、V. Srinivasan、K. Gopalakrishnan、Z. Wang 和 P. Chuang，“精确高效的2位量化神经网络”，载于第*2*届*SysML*会议论文集，第2019卷，2019年。[25] A. Garofalo、M. Rusci、F. Conti、D. Rossi 和 L. Benini，“Pulp-nn：一种用于在基于RISC-V的并行超低功耗集群边缘端进行量化神经网络推理的计算库”，载于*2019*年第*26*届*IEEE*电子、电路与系统国际会议 *(ICECS)*。IEEE，2019年，第33–36页。[26] 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。[27] R. Wood、B. Finio、M. Karpelson、K. Ma、N. P´erez-Arancibia、P. Sreetharan、H. Tanaka 和 J. Whitney，“‘Pico’飞行器的研究进展”，国际机器人研究杂志，第31卷，第11期，第1292–1302页，2012年。