

一网打尽：四旋翼赛车中的领域随机化跨不同平台应用

Robin Ferede, Till Blaha, Erin Lucassen, Christophe De Wagter, Guido C.H.E. de Croon

Abstract—在高速四旋翼飞行器竞速中，找到一个能够在不同平台之间表现良好的单一控制器仍然具有挑战性。本研究展示了首个适用于无人机竞速的神经网络控制器，该控制器可以在物理上不同的四旋翼飞行器之间通用。我们演示了一个经过领域随机化训练的单个网络可以稳健地控制各种类型的四旋翼飞行器。该网络仅依赖当前状态来直接计算电机指令。这种通用控制器的有效性通过在两种截然不同型号（3 英寸和 5 英寸竞速四旋翼飞行器）上的实际测试得到验证。我们进一步将这种通用控制器的性能与专门为 3 英寸和 5 英寸无人机训练的控制器进行了比较，使用了它们识别模型参数的不同水平领域随机化（0%，10%，20%，30%）。虽然通用控制器的速度略慢于微调模型，但它在不同平台上的适应性表现突出。我们的结果显示没有随机化的模拟到现实转换失败，而增加随机化可以提高鲁棒性但会降低速度。尽管存在这种权衡，我们的研究结果突显了领域随机化对于推广控制者的潜力，为通用 AI 控制器铺平了道路，这些控制器能够适应任何平台。

Index Terms—无人机竞速，强化学习，领域随机化，真实差距，模拟到现实迁移

I. 介绍

无人机市场正在经历快速的增长 [1]，然而大多数应用仍然依赖于人类飞行员。鉴于电池寿命限制了飞行时间，提高无人机的速度和自主性提供了显著的优势 [2]。因此，对自主无人机竞速的研究是一个不断增长的领域，它有助于推动技术进步并提升无人机的能力。

无人机竞速中一个重要的挑战是开发能够跨不同四旋翼平台良好泛化的控制器。虽然人类飞行员可以轻松地对各种无人机进行微调以适应，但当前的赛车 AI 系统通常会过度拟合其特定平台，限制了它们的泛化能力。这一局限性阻碍了现有无人机竞速技术向其他实际应用 [2] 的转移。

最近四旋翼控制方面的进展主要集中在强化学习 (RL) 技术上 [3]–[7]。这些方法在无人机竞速中展示

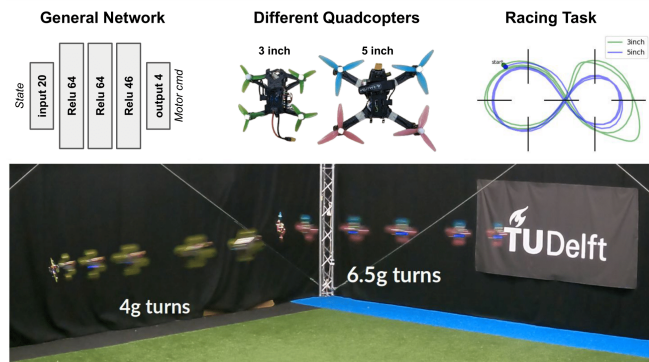


Fig. 1. 我们训练了一个单独的神经网络来控制物理上不同的无人机。定时摄影图像展示了强化学习网络从仿真到现实的成功迁移，使其能够使用 3 英寸和 5 英寸四旋翼无人机进行无人机竞速。

了令人印象深刻的结果，神经控制器超越了人类冠军 [6]。然而，这些成就依赖于对无人机平台的精确建模，并且难以推广到不同的无人机。

领域随机化 (DR) 已成为跨不同环境转移 RL 策略的显著方法。该技术涉及在具有不同参数的模拟环境中训练策略，以提高它们在现实世界场景中部署时的鲁棒性 [8]。例如，DR 促进了机器人操作的一次性从仿真到现实的迁移 [9]、使用合成数据进行视觉导航 [10] 以及对象检测 [11]。此外，在端到端四旋翼控制中，DR 结合实时适应对于弥合现实差距至关重要 [12]–[14]。然而，尽管 DR 增强了鲁棒性，但也伴随着权衡，包括更长的训练时间和性能最优性的降低 [8]，[15]。

当前对四旋翼飞行器的广义控制研究仅集中在低空悬停或恢复至悬停状态下的低级任务上，对于需要在显著更高的速度下进行引导和控制的比赛场景则探索较少。最近的研究探讨了各种方法：一种开发了一个网络用于稳定低级控制（位置控制和轨迹跟踪），该网络通过领域随机化 [16] 技术推广至多种四旋翼飞行器上。另一种方法 [17] 提出了一种单次实时技术，通过对未知四旋翼飞行器的执行机构效能参数进行迭代确定来控制其动作，扩展了 [18] 的工作成果。此外，最近的工作训练了一个神经网络来估算四旋翼飞行器参

¹The authors are with the Micro Air Vehicle Lab of the Faculty of Aerospace Engineering, Delft University of Technology, 2629 HS Delft, The Netherlands robinferede@tudelft.nl, t.m.blaha@tudelft.nl, e.lucassen@student.tudelft.nl, c.deWagter@tudelft.nl, g.c.h.e.deCroon@tudelft.nl

数的潜在表示，并根据另一个条件神经网络控制器实现接近悬停状态下的飞行，从而在现实世界部署时实现在不同硬件和扰动情况下的在线自适应调整 [19]。

在这项工作中，我们通过使用领域随机化来开发适用于高速竞速的神经网络控制器，从而将这些结果扩展到了无人机竞速中。我们的控制器直接将状态信息映射到电机命令，实现了端到端的引导和控制。我们的主要贡献包括：

- 1) 介绍我们所知的第一个能够在不同无人机之间进行高速竞速操作的神经网络控制器
- 2) 对不同程度领域随机化对仿真到现实迁移能力和飞行速度性能影响的调查。

本文组织如下：第 II 节介绍了无人机模型和控制公式；第 III 节描述了实验设置；第 IV 节分析了一般控制网络的性能并考察了不同程度领域随机化的影响；第 V 节提供了结论，并概述了未来工作的方向。

II. 方法论

A. 四旋翼模型

为了训练我们的神经策略，我们使用一个类似于 [14] 中概述的参数模型来模拟我们的四旋翼动力学。四旋翼的状态和控制由下式给出：

$$\mathbf{x} = [\mathbf{p}, \mathbf{v}, \boldsymbol{\lambda}, \boldsymbol{\Omega}, \boldsymbol{\omega}]^T \quad \mathbf{u} = [u_1, u_2, u_3, u_4]^T$$

其中 $\mathbf{p}$ 是无人机的位置， $\mathbf{v}$ 是速度， $\boldsymbol{\lambda}$ 是欧拉角，是机体速率， $\boldsymbol{\omega}$ 是螺旋桨的转速（单位：弧度/秒）。 $\mathbf{u} \in [0, 1]^4$ 表示发送到电机控制器的标准命令。运动方程表示如下：

$$\dot{\mathbf{p}} = \mathbf{v} \quad \dot{\mathbf{v}} = g\mathbf{e}_3 + R(\boldsymbol{\lambda})\mathbf{F} \quad (1)$$

$$\dot{\boldsymbol{\lambda}} = Q(\boldsymbol{\lambda})\boldsymbol{\Omega} \quad \dot{\boldsymbol{\Omega}} = \mathbf{M}$$

$$\dot{\omega}_i = (\omega_{ci} - \omega)/\tau \quad (2)$$

其中， R 是旋转矩阵， Q 将体速率转换为欧拉角导数。稳态电机响应建模为： $\omega_{ci} = (\omega_{\max} - \omega_{\min})\sqrt{k_i u_i^2 + (1 - k_i)u_i} + \omega_{\min}$ 。具体力 F 和力矩 M ¹ 被建模为：

$$\mathbf{F} = \left[-\sum_{i=1}^4 k_x v_x^B \omega_i, -\sum_{i=1}^4 k_y v_y^B \omega_i, -\sum_{i=1}^4 k_\omega \omega_i^2 \right]^T$$

¹ M 表示角加速度。通过 $k_\square$ 间接估计了惯性矩，并忽略了陀螺效应。

TABLE I

识别出的参数用于 3 英寸和 5 英寸无人机

param	3 inch	5 inch	param	3 inch	5 inch
$\hat{k}_\omega$	14.3	27.1	$\omega_{\min}$ (rad/s)	305.4	238.49
$\hat{k}_x$	0.16	0.16	$\omega_{\max}$ (rad/s)	4887.57	3295.5
$\hat{k}_y$	0.18	0.24	k_l	0.84	0.95
			τ (sec)	0.04	0.04
$\hat{k}_{p1}$	615.0	711.0	$\hat{k}_{r1}$	47.1	35.2
$\hat{k}_{p2}$	598.0	718.0	$\hat{k}_{r2}$	47.1	35.2
$\hat{k}_{p3}$	650.0	691.0	$\hat{k}_{r3}$	47.1	35.2
$\hat{k}_{p4}$	479.0	724.0	$\hat{k}_{r4}$	47.1	35.2
$\hat{k}_{q1}$	217.0	573.0	$\hat{k}_{r5}$	5.57	6.49
$\hat{k}_{q2}$	238.0	637.0	$\hat{k}_{r6}$	5.57	6.49
$\hat{k}_{q3}$	280.0	548.0	$\hat{k}_{r7}$	5.57	6.49
$\hat{k}_{q4}$	196.0	640.0	$\hat{k}_{r8}$	5.57	6.49

$$\mathbf{M} = \begin{bmatrix} -k_{p1}\omega_1^2 - k_{p2}\omega_2^2 + k_{p3}\omega_3^2 + k_{p4}\omega_4^2 \\ -k_{q1}\omega_1^2 + k_{q2}\omega_2^2 - k_{q3}\omega_3^2 + k_{q4}\omega_4^2 \\ -k_{r1}\omega_1 + k_{r2}\omega_2 + k_{r3}\omega_3 - k_{r4}\omega_4 + \\ \dots - k_{r5}\dot{\omega}_1 + k_{r6}\dot{\omega}_2 + k_{r7}\dot{\omega}_3 - k_{r8}\dot{\omega}_4 \end{bmatrix}$$

参数 $k_\square, \omega_{\min}, \omega_{\max}$ 是通过对 3 英寸和 5 英寸四旋翼飞行器的手动飞行数据使用线性回归进行估算的。我们根据最大角速度 $\omega_{\max}$ 对参数进行归一化。归一化参数为：

$$\hat{k}_\omega, \hat{k}_{pi}, \hat{k}_{qi} = (k_\omega, k_{pi}, k_{qi}) \times \omega_{\max}^2$$

$$\hat{k}_{xi}, \hat{k}_{yi}, \hat{k}_{ri} = (k_{xi}, k_{yi}, k_{ri}) \times \omega_{\max}$$

这种缩放考虑了 $\omega_{\max}$ 的变化，并确保模型的一致性。估计值可以在表 I 中找到。I 这些参数突出了每个平台的独特控制特性。例如，3 英寸平台的推重比大约是 5 英寸平台的一半（与 $\hat{k}_\omega$ 成比例），电机转速快 50%，且螺距效率（ $\hat{k}_{pi}$ ）不到 5 英寸平台的一半。

B. 强化学习问题定义

比赛设置由七个边长为 1.5x1.5 米的正方形门组成，排列成一个类似数字 8 的赛道，如图 1 所示。无人机从随机选择的一个门前 1 米处开始，方式类似于 [3]。其他初始状态变量定义如下： $\mathbf{v} \in [-0.5, 0.5]^3$ ， $\phi, \theta \in [-\frac{\pi}{9}, \frac{\pi}{9}]$ ， $\psi \in [-\pi, \pi]$ ， $\boldsymbol{\Omega} \in [-0.1, 0.1]^3$ 和 $\boldsymbol{\omega} \in [\omega_{\min}, \omega_{\max}]$ 。类似于 [5]，待优化的奖励函数包括进度奖励、速率惩罚和碰撞惩罚：

$$r_k = \begin{cases} -10, & \text{if collided} \\ \|\mathbf{p}_{k-1} - \mathbf{p}_{gk}\| - \|\mathbf{p}_k - \mathbf{p}_{gk}\| - c\|\boldsymbol{\Omega}\| & \text{otherwise} \end{cases}$$

这里， $\mathbf{p}_{gk}$ 表示当前目标门中心的位置，而 $\mathbf{p}_k$ 和 $\mathbf{p}_{k-1}$ 分别表示无人机的当前位置和前一位置。速率惩罚系

TABLE II
一般策略的随机化方案

参数	分布	参数	分布
$\omega_{\min}$	$\mathcal{U}(0, 500)$	$\hat{k}_{\omega}$	$\mathcal{U}(10, 30)$
$\omega_{\max}$	$\mathcal{U}(3000, 5000)$	$\hat{k}_x$	$\mathcal{U}(0.1, 0.3)$
k_l	$\mathcal{U}(0, 1)$	$\hat{k}_y$	$\mathcal{U}(0.1, 0.3)$
τ	$\mathcal{U}(0.01, 0.1)$		
$\hat{k}_p$	$\mathcal{U}(200, 800)$	$\hat{k}_{p1}, \hat{k}_{p2}, \hat{k}_{p3}, \hat{k}_{p4}$	$\hat{k}_p \pm \mathcal{U}(50)$
$\hat{k}_q$	$\mathcal{U}(200, 800)$	$\hat{k}_{q1}, \hat{k}_{q2}, \hat{k}_{q3}, \hat{k}_{q4}$	$\hat{k}_q \pm \mathcal{U}(50)$
$\hat{k}_r$	$\mathcal{U}(20, 80)$	$\hat{k}_{r1}, \hat{k}_{r2}, \hat{k}_{r3}, \hat{k}_{r4}$	$\hat{k}_r$
$\hat{k}_{rd}$	$\mathcal{U}(2, 8)$	$\hat{k}_{r5}, \hat{k}_{r6}, \hat{k}_{r7}, \hat{k}_{r8}$	$\hat{k}_{rd}$

数设置为 $c = 0.001$ 。当无人机与地面接触或飞出一个 $10 \times 10 \times 7$ 米的边界框时，记录一次碰撞。当发生碰撞时，回合结束。此外，当错过门（即无人机穿过门平面但位于 1.5×1.5 米范围之外）时，回合也结束。

该奖励函数与上一节中的四旋翼模型一起，用于创建一个使用 PPO 算法 [20] 进行训练的 gym 环境，利用了 Python 库 Stable-Baselines3 [21]。我们的实现模拟了 100 个无人机并行运行，使用折扣因子 $\gamma = 0.999$ 和最大集长度为 1200 时间步（12 秒）。训练过程在总共 1 亿时间步后结束。我们实现的代码可以在 GitHub² 上公开获取。

C. 随机化

1) 一般政策：表中概述的随机化 II 训练了通用策略来控制 3 英寸和 5 英寸的四旋翼飞行器。均匀分布被设计为涵盖两种尺寸的所有参数。

2) 微调政策：我们通过对表中的值应用随机化，训练了五个针对 3 英寸和 5 英寸四旋翼飞行器的策略。I 引入 0%、10%、20% 和 30% 的变化。具体来说，每个参数乘以 $\mathcal{U}(1 - p, 1 + p)$ ，其中 p 表示随机化百分比。推力线性化常数 k_l 是一个例外，因为它被限制在确保其保持低于 1 的范围内。

D. 政策

选定的神经策略是一个三层全连接网络，具有 ReLU 激活函数，每层有 64 个神经元（详见第 IV-A 节以了解动机）。该政策接收 20 个观察值，包括四旋翼飞行器的状态以及当前和未来门的信息，类似于先前的工作 [14]：

$$\mathbf{x}_{obs} = [\mathbf{p}^{g_i}, \mathbf{v}^{g_i}, \boldsymbol{\lambda}^{g_i}, \boldsymbol{\Omega}, \boldsymbol{\omega}, \mathbf{p}_{g_{i+1}}^{g_i}, \boldsymbol{\psi}_{g_{i+1}}^{g_i}]^T \quad (3)$$

这里， g_i 表示第 i 个门的参考帧，而 $\mathbf{p}_{g_{i+1}}$ 和 $\boldsymbol{\psi}_{g_{i+1}}$ 分别表示下一个门的位置和方向。网络输出四个电机命令 ($\mathbf{u}$)。由于模型参数应用了领域随机化，这些参数成为马尔可夫决策过程 (MDP) 状态的一部分。因此，当前的观测形式下状态并非完全可观测 [15]。一些方法试图通过将状态或动作历史纳入观测来解决这个问题 [6], [22]，而其他方法则使用参数输入 [19]。在第 IV-A 节中，我们探讨了这些方法并证明了选择方程 3 的合理性。

III. 实验设置

我们在两架四轴飞行器上测试了我们的策略。它们之间的共同点是计算硬件和实验基础设施：我们使用运行 INDIfight³ 的 STM32H743 微控制器，这是 Betaflight 的一个分支。机载 EKF 融合来自 TDK InvenSense ICM-42688-P IMU 的数据以及来自外部光学运动捕捉系统 (Optitrack) 的位置/姿态观测结果。然后该状态以 1000Hz 的更新率由机载的神经网络使用。

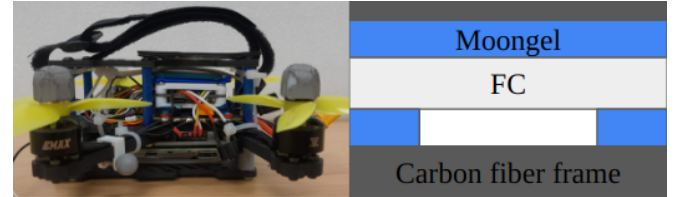


Fig. 2. 这张图片显示了用于测试的无人机。在基架顶部放置了四块 Moongel 角件，而不是螺栓，在其上放有一个装有飞行控制器的 3D 打印外壳。然后在外壳上放置第二层 Moongel，并通过两条扎带稍微向下压缩。

在初始测试中，高速时螺旋桨振动导致加速度计饱和，超过了加速度计 $\pm 16g$ 的限制。我们希望通过添加如图 2 所示的减震垫来解决这个问题。对 Sorbothane、胶带和 Moongel Damper Pads [23] 进行了比较，但后者被证明能够最有效地控制高频共振，同时仍能传递无人机的真实低频运动。这一结果与之前关于减震垫性能的研究发现一致，其中 Moongel 确实被证明在消除噪声方面是最优的 [24]。虽然 3D 打印外壳不是必需的，但它有助于更均匀地分布力到 FC 上，而连接到底座框架的扎带提供了预加载压缩，以保持稳定性和提高减震性能。

²https://github.com/tudelft/optimal_quad_control_RL/tree/icra2025

³<https://github.com/tudelft/indiflight>

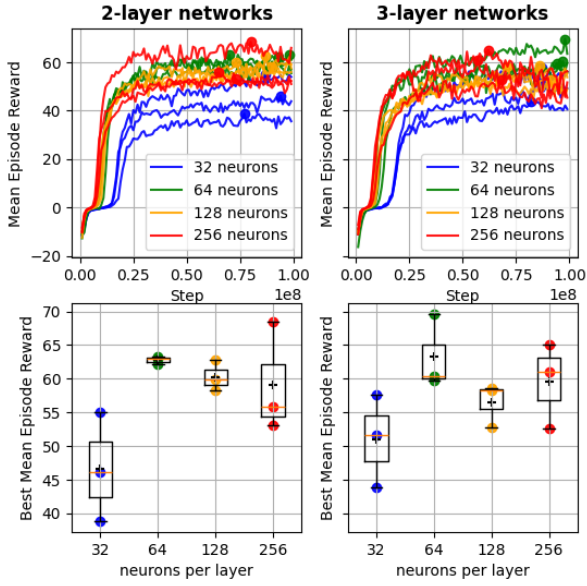


Fig. 3. 网络规模比较: 在评估的架构中, 64,64,64 架构基于每次运行 1 亿时间步长的三次独立运行, 获得了最高的平均剧集奖励。

IV. 结果

A. 选择架构

我们使用随机模拟器评估了几种架构, 并基于三次独立运行中获得的最高平均剧集奖励选择了 3 层 64 神经元的架构, 每次运行包含 1 亿个时间步。网络大小比较如图 3 所示。接下来, 我们修改了选定的架构以解决部分可观测性问题, 增加了各种输入。修改包括:

- 将最后的 1、2 或 3 个动作添加到观察中
- 添加最后的 1、2 或 3 个动作, 步长为 10 (过去以 0.1 秒为单位的步骤)
- 添加状态历史 (最后 1、2 或 3 个状态, 有或无步长 10)
- 添加模型参数 (真实值) 或带有 10% 和 20% 噪声

每种变体遵循相同的训练设置, 并在 1000 次展开中进行评估。参见图 4 以获取结果。所有修改后的输入均未显示出明显的改进, 导致我们将其排除在进一步使用之外。这一意外的结果表明需要进一步调查, 这可能涉及更长的训练周期、超参数调整或替代参数表示方法。在本文的其余部分中, 我们将重点放在没有额外输入的网络上。

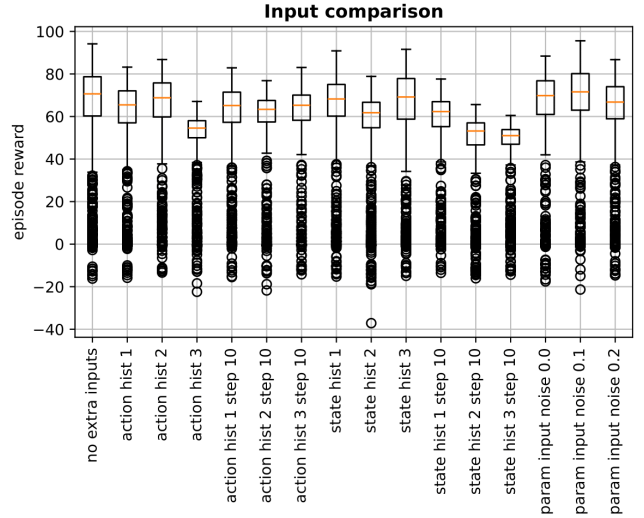


Fig. 4. 具有各种输入修改的 64,64,64 ReLU 网络的训练性能比较。

TABLE III

模拟结果基于 1000 次政策滚动的平均值

网络	3 英寸 sim				5 英寸 sim			
	ep rew	ep len	#gates	%crash	ep rew	ep len	#gates	%crash
一般	64.67	1140	40	5.4	76.78	1186	52	1.2
3 英寸 30%	59.97	1200	41	0.0	3.95	204	4	99.8
3 英寸 20%	69.02	1200	43	0.0	-6.24	127	3	100
3 英寸 10%	77.59	1197	46	0.3	-2.24	70	2	100
3 英寸 0%	78.55	1193	46	0.6	-1.78	89	2	100
5 英寸 30%	4.23	195	3	100	74.03	1199	49	0.1
5 英寸 20%	-0.35	125	1	100	88.33	1197	58	0.3
5 英寸 10%	4.57	218	4	100	92.76	1183	55	1.5
5 英寸 0%	-0.69	143	1	100	99.35	1179	59	2.0

B. 模拟

我们在 9 个不同的环境中训练了选定的架构: 第 II-C.1 节中描述的一般随机化和第 II-C.2 节中描述的微调模型, 这些模型包括具有 0%、10%、20% 和 30% 随机化水平的 3 英寸和 5 英寸无人机。训练遵循上述相同的程序。性能通过在 3 英寸和 5 英寸无人机环境 (从 II-B 节中描述的范围采样的相同初始条件) 上进行 1000 次滚动评估。参见表. III。

如所示, 通用模型在两个四旋翼飞行器模拟中均表现出色, 分别以 3 英寸环境中的 5.4% 和 5 英寸环境中的 1.2% 的坠毁率成功穿越门框。然而, 与每个特定无人机优化过的模型相比, 其表现略逊一筹。有趣的是, 优化过的模型无法在不同类型的无人机之间迁移, 导致由于参数范围不重叠而出现 100% 的坠毁率, 即使进行了 30% 的随机化也是如此。在这些优化过的模型

中, 较低的随机化水平与较高的平均回合奖励相关联, 尽管这会导致坠毁率略有增加。

C. 实际飞行

我们现在在各自的平台上对模拟中表现良好的网络进行测试。具体来说, 微调模型在其指定的平台进行评估, 而通用模型则在 3 英寸和 5 英寸四旋翼上进行测试。所有飞行都从悬停状态开始, 在 1 号门前 1 米处, 并重复三次 (每次 12 秒, 与 RL 框架中使用的最大剧集长度相匹配)。第一次飞行使用完全充电的电池, 随后的飞行随着电池电量减少而进行, 到第三次运行时几乎耗尽电量。飞行日志以 100Hz 采样, 以便与模拟器的时间步长相匹配, 通过剧集奖励指标实现直接比较。此外, 在相同的初始条件下, 在模拟器中复现每段飞行。真实和模拟飞行的轨迹如图??所示。可以看到, 通用网络成功在两个无人机上导航比赛路线, 尽管速度略慢。所有飞行的性能指标见表 IV-B 和 IV-B, 显示随着随机化从 30% 减少到 10%, 性能提高, 在 10% 时奖励最高。在 0% 随机化下, 无人机不再通过门。图??提供了剧集奖励的箱形图, 清楚地说明了现实差距如何随领域随机化的减少而扩大。

D. 时间最优性

框架中的奖励设计用于学习时间最优飞行。不仅在训练过程中将实现的每集奖励与最优值进行比较是有用的, 还将实际圈速与时间最优飞行进行比较作为一种验证和上下文手段也是有用的。最优的 8 字形路径 (见图??) 被制定为一个边值问题, 并使用 [25] 解决了 5 英寸模型的问题, 得到理论圈速为 1.07 秒, 相比之下, 在模拟中我们的最快 RL 策略的圈速是 1.6 秒, 而在现实生活中则是 1.9 秒。尽管时间最优程序不包括阻力项, 我们仍将大部分差异归因于 RL 策略采取了次优路径; 远离门框边缘相对较远。然而, 最优控制方法在强制通过门中心时返回 1.53s。这表明我们的奖励函数并不完全与优化圈速对齐。如表 Tab. IV-B 所示, 经过 10% 微调的网络在回合奖励方面实现了最高的实际性能, 但它通过的门比 20% 变体少。因此我们得出结论, 在狭窄赛道上实现更多时间优化性能时, 奖励函数应包括允许更好地调整风险与回报平衡的项, 鼓励代理更靠近门边缘飞行。

V. 结论

在这项工作中, 我们介绍了一种适用于多款四旋翼平台的高速无人机竞速的神经网络控制器。我们在 3 英寸和 5 英寸四旋翼上以高达 10 m/s 的速度进行了实际飞行演示其有效性。我们的基准测试包括将这种通用控制器与专门针对 3 英寸和 5 英寸四旋翼进行不同级别域随机化训练的神经策略进行比较。尽管我们的一般模型在性能上略低于微调控制器, 但它表现出跨平台的强大迁移能力。我们的实验表明, 增加域随机化最初会增强鲁棒性并有助于缩小现实差距。然而, 过度随机化最终会影响飞行速度。我们探索了将在线适应纳入网络中, 如 [17] 和 [19] 所讨论的那样, 但未能达到预期改进。潜在因素可能包括训练时长、超参数调整或参数表示的选择。尽管如此, 即使没有额外输入, 我们通用方法的重大成功也彰显了其潜力并为未来研究铺平道路。解决已识别的挑战可能会带来显著性能提升, 使之成为进一步探索的一个令人兴奋的方向。

REFERENCES

- [1] M. Hassanalain and A. Abdelkefi, "Classifications, applications, and design challenges of drones: A review," *Progress in Aerospace Sciences*, vol. 91, pp. 99–131, 2017.
- [2] D. Hanover, A. Loquercio, L. Bauersfeld, A. Romero, R. Penicka, Y. Song, G. Cioffi, E. Kaufmann, and D. Scaramuzza, "Autonomous Drone Racing: A Survey," *IEEE Transactions on Robotics*, vol. 40, pp. 3044–3067, Jan. 2024.
- [3] Y. Song, M. Steinweg, E. Kaufmann, and D. Scaramuzza, "Autonomous drone racing with deep reinforcement learning," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 1205–1212.
- [4] R. Penicka, Y. Song, E. Kaufmann, and D. Scaramuzza, "Learning minimum-time flight in cluttered environments," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 7209–7216, 2022.
- [5] Y. Song, A. Romero, M. Müller, V. Koltun, and D. Scaramuzza, "Reaching the limit in autonomous racing: Optimal control versus reinforcement learning," *Science Robotics*, vol. 8, no. 82, p. eadg1462, 2023. [Online]. Available: <https://www.science.org/doi/abs/10.1126/scirobotics.adg1462>
- [6] E. Kaufmann, L. Bauersfeld, A. Loquercio, M. Müller, V. Koltun, and D. Scaramuzza, "Champion-level drone racing using deep reinforcement learning," *Nature*, vol. 620, no. 7976, pp. 982–987, Aug. 2023. [Online]. Available: <https://doi.org/10.1038/s41586-023-06419-4>
- [7] J. Eschmann, D. Albani, and G. Loianno, "Learning to Fly in Seconds," Nov. 2023, arXiv:2311.13081.

TABLE IV

性能指标: 3 英寸飞行器

网络	实际飞行				模拟			
	ep rew	#gates	V_{mean}	V_{max}	ep rew	#gates	V_{mean}	V_{max}
Gen	49.73	31	6.38	10.58	69.76	42	8.59	10.99
	49.38	31	6.38	10.33	69.77	42	8.59	10.96
	48.78	31	6.31	10.4	69.97	42	8.6	10.94
30%	53.92	37	6.4	9.03	60.44	41	6.86	9.76
	53.48	37	6.34	8.73	60.61	41	6.86	9.68
	53.25	37	6.33	8.85	60.6	41	6.86	9.67
20%	55.97	36	6.61	10.73	69.67	43	7.56	11.58
	55.4	35	6.54	10.71	69.67	43	7.55	11.58
	55.32	35	6.49	10.48	69.72	43	7.56	11.58
10%	59.86	39	7.13	12	77.74	46	8.45	13.15
	59.99	39	7.13	11.56	78.17	46	8.46	13.24
	59.34	38	7.05	11.39	78.11	46	8.46	13.25
0%	7.08	5	6.58	11.46	80.82	46	8.79	12.8
	7.11	5	6.49	11.25	81.09	46	8.78	12.49
	10.18	7	6.51	10.87	80.93	46	8.77	12.38

TABLE V

性能指标: 5 英寸飞行器

网络	实际飞行				模拟			
	ep rew	#gates	V_{mean}	V_{max}	ep rew	#gates	V_{mean}	V_{max}
Gen	61.51	46	7.83	9.88	77.14	53	9.09	11.03
	61.22	46	7.84	9.83	77.5	53	9.12	11.37
	60.65	46	7.8	9.84	77.74	53	9.14	11.26
30%	63.22	42	7.01	8.82	74.38	48	8.01	10.22
	62.87	42	7.01	8.73	74.1	48	8.01	10.15
	62.59	41	6.98	8.7	73.7	48	8	10.1
20%	71.08	48	8.13	11.8	89.44	58	9.99	13.8
	71.44	49	8.16	11.74	89.56	58	9.99	13.9
	71.3	48	8.13	11.52	89.84	58	10	14.09
10%	71.83	46	8.63	12.45	95.68	58	10.53	12.81
	72.67	47	8.73	12.18	94.93	58	10.52	12.81
	71.98	46	8.63	12.07	95.67	58	10.54	13.05
0%	10.62	8	6.96	11.84	99.9	60	11.34	15.4
	17.16	13	7.46	13.13	100.55	59	11.34	15.53
	11.34	8	7.43	12.39	97.48	59	11.27	15.37

- [8] J. Josifovski, M. Malmir, N. Klarmann, B. L. Žagar, N. Navarro-Guerrero, and A. Knoll, “Analysis of randomization effects on sim2real transfer in reinforcement learning for robotic manipulation tasks,” in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 10 193–10 200.
- [9] O. M. Andrychowicz, B. Baker, M. Chociej, R. Jozefowicz, B. McGrew, J. Pachocki, A. Petron, M. Plappert, G. Powell, A. Ray *et al.*, “Learning dexterous in-hand manipulation,” *The International Journal of Robotics Research*, vol. 39, no. 1, pp. 3–20, 2020.
- [10] A. Loquercio, E. Kaufmann, R. Ranftl, A. Dosovitskiy, V. Koltun, and D. Scaramuzza, “Deep drone racing: From simulation to reality with domain randomization,” *IEEE Transactions on Robotics*, vol. 36, no. 1, pp. 1–14, 2019.
- [11] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, “Domain randomization for transferring deep neural networks from simulation to the real world,” in *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2017, pp. 23–30.
- [12] R. Ferede, G. de Croon, C. De Wagter, and D. Izzo, “End-to-end neural network based optimal quadcopter control,” *Robotics and Autonomous Systems*, vol. 172, p. 104588, 2024.
- [13] S. Origer, C. De Wagter, R. Ferede, G. C. de Croon, and D. Izzo, “Guidance & control networks for time-optimal quadcopter flight,” *arXiv preprint arXiv:2305.02705*, 2023.
- [14] R. Ferede, C. De Wagter, D. Izzo, and G. C. De Croon, “End-to-end reinforcement learning for time-optimal quadcopter flight,” in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 6172–6177.
- [15] G. Tiboni, P. Klink, J. Peters, T. Tommasi, C. D’Eramo, and G. Chalkatzaki, “Domain randomization via entropy maximization,” *arXiv preprint arXiv:2311.01885*, 2023.
- [16] A. Molchanov, T. Chen, W. Hönig, J. A. Preiss, N. Ayanian, and

- G. S. Sukhatme, "Sim-to-(multi)-real: Transfer of low-level robust control policies to multiple quadrotors," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 59–66.
- [17] T. Blaha, E. Smeur, and B. Remes, "Control of Unknown Quadrotors from a Single Throw," *arXiv*, vol. arXiv:2406.11723, 2024.
- [18] E. J. J. Smeur, Q. Chu, and G. C. H. E. de Croon, "Adaptive Incremental Nonlinear Dynamic Inversion for Attitude Control of Micro Air Vehicles," *Journal of Guidance, Control, and Dynamics*, vol. 39, no. 3, pp. 450–461, Mar. 2016.
- [19] D. Zhang, A. Loquercio, X. Wu, A. Kumar, J. Malik, and M. W. Mueller, "Learning a Single Near-hover Position Controller for Vastly Different Quadcopters," May 2023.
- [20] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.
- [21] A. Raffin, A. Hill, A. Gleave, A. Kanervisto, M. Ernestus, and N. Dormann, "Stable-baselines3: Reliable reinforcement learning implementations," *Journal of Machine Learning Research*, vol. 22, no. 268, pp. 1–8, 2021. [Online]. Available: <http://jmlr.org/papers/v22/20-1364.html>
- [22] I. Geles, L. Bauersfeld, A. Romero, J. Xing, and D. Scaramuzza, "Demonstrating agile flight from pixels without state estimation," *arXiv preprint arXiv:2406.12505*, 2024.
- [23] RTOM Corporation, "Moongel damper pads," <https://rtom.com/moongel-damper-pad/>, accessed: August 20, 2024.
- [24] S. F. ul Haq Gilani, M. H. bin Mohd Khir, R. Ibrahim, E. ul Hassan Kirmani, and S. I. ul Haq Gilani, "Modelling and development of a vibration-based electromagnetic energy harvester for industrial centrifugal pump application," *Microelectronics Journal*, vol. 66, pp. 103–111, 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0026269216305948>
- [25] J.-B. Caillaud, O. Cots, J. Gergaud, P. Martinon, and S. Sed, "OptimalControl.jl: a Julia package to model and solve optimal control problems with ODE's." [Online]. Available: <https://control-toolbox.org/OptimalControl.jl>

