

3we: 面向仿真到现实具身智能研究的开放基础设施

3we Contributors

<https://github.com/telleroutlook/3we-robot-platform>

摘要

具身智能研究需要仿真、真实硬件与学习算法的深度协同——然而现有平台往往迫使研究者在昂贵的商业系统、纯仿真环境，或具有陡峭 ROS2 学习曲线的硬件平台之间做出取舍。本文提出 **3we**，一套完全开源的基础设施，使相同的 Python 代码无需任何修改即可在轻量级 mock 仿真器、Gazebo、NVIDIA Isaac Sim 及真实硬件 (Raspberry Pi 5 + Hailo-8L) 上完全一致地运行。该平台提供 AI 优先的 Python API，可将典型 ROS2 导航代码从 60 余行压缩至 5 行；同时提供 Gymnasium 兼容环境、原生 VLM/VLA 集成，以及覆盖 7 个评估场景的标准化基准测试。在物料清单 (BOM) 成本不超过 300 美元的可复现硬件上，初步验证表明点导航任务的仿真到现实迁移率超过 0.6，社区基线在结构化仿真环境中达到 82% 的成功率 (SPL 0.65)。所有软件 (Apache 2.0 协议)、硬件设计 (CERN-OHL-P v2 协议) 及文档 (CC-BY-SA 4.0 协议) 均已公开发布。

1 引言

具身智能——研究能够在物理环境中感知与行动的智能体——已成为人工智能领域的核心挑战之一 [1,2]。视觉-语言模型 (VLM) [3] 与视觉-语言-动作 (VLA) 模型 [4,5] 的近期进展表明，基础模型可以通过自然语言指令直接控制机器人。然而，将上述进展转化为可复现的研究成果面临三大根本性障碍：

(1) **硬件成本与可及性**。主流移动机器人平台 (TurtleBot 4: 1,200 美元以上; Stretch RE1: 25,000 美元) 将大多数研究团队——尤其是发展中国家及教学机构——拒于门外。

(2) **仿真到现实的迁移鸿沟**。研究者通常在仿真环境 (Habitat [2]、Isaac Sim [6]) 中开发算法，但部署到真实硬件时往往需要重写全部代码，工程代价极高。

(3) **软件复杂性**。ROS2 [7] 提供了健壮的中件，但对于专注于算法研究而非通信基础设施的 AI 研究者而言，其学习曲线十分陡峭。

为此，本文提出 **3we** (“three-way Embodied”，三向具身)，一套同时解决上述三大障碍的开放基础设施：

- **BOM 不超过 300 美元的开放硬件**，提供完整可复现的 PCB 设计 (KiCad)、机械图纸及选型指南，遵循 CERN-OHL-P v2 协议。
- **零代码 Sim2Real**: 统一的 Python API, `Robot(backend="gazebo")` 与 `Robot(backend="real")` 执行完全相同的用户代码。
- **AI 优先 API**: 从安装到首次导航仅需 5 行代码, 原生支持 VLM/VLA 集成及 Gymnasium 接口。
- **标准化基准测试**: 覆盖 7 个场景的 3 类任务, 提供可复现的基线结果与社区排行榜。

2 相关工作

2.1 移动机器人平台

TurtleBot [8] 自 2010 年起成为 ROS 教育领域的事实标准。TurtleBot 4 (1,200 美元以上) 集成了 Nav2, 但缺乏 AI 优先 API、仿真与硬件一致性保证, 且成本为本平台的 4 倍。Linorobot2 [9] 提供开放设计, 但面向爱好者, 缺乏研究级基准测试体系。

2.2 仿真环境

Habitat [2] 与 AI2-THOR [10] 提供高保真视觉导航仿真，但缺乏物理机器人部署路径。NVIDIA Isaac Sim [6] 支持 GPU 加速强化学习训练，但需要昂贵的算力且不提供开放硬件参考。上述系统仅解决仿真或现实的问题，而非两者之间的桥梁。

2.3 AI 机器人框架

LeRobot [11] (HuggingFace) 为操作任务提供 VLA 模型部署能力，但缺乏 ROS2 集成、自主导航栈及移动平台的 Sim2Real 一致性保证。其近期的移动端扩展 (LeKiwi、EarthRover) 仍处于早期阶段，尚未达到 Nav2 级别的自主能力。

2.4 平台定位

3we 占据独特的生态位：在保证 Sim2Real API 一致性的前提下，同时提供完全开放的硬件与软件，并具备 ROS2 原生导航与 AI 优先设计。表 ?? 总结了与主要平台的关键差异。

3 系统架构

3.1 三层设计

3we 采用分层架构，将 AI 研究与机器人工程解耦：

1. **AI 优先 Python API** (用户层): 研究者使用标准 Python async/await 编写代码，无需接触 ROS2、launch 文件或消息类型。
2. **3we-core** (中间件层): 负责后端调度、传感器数据标准化、坐标变换与安全检查。
3. **ROS2 / micro-ROS** (基础设施层): Nav2 路径规划、SLAM、micro-ROS 到 ESP32 固件的桥梁，对用户完全透明。

3.2 后端多态性

核心抽象为 `BackendBase`，由五个后端实现：

- **Mock**: 零依赖的 2D 运动学仿真，支持可配置噪声。可在任意机器上运行，无需 GPU 或 ROS2。
- **Gazebo**: 基于 Gazebo Harmonic 的完整物理仿真，对 CPU 友好，支持无头模式 (CI 兼容)。
- **Isaac Sim**: GPU 加速渲染与物理仿真，支持大规模强化学习训练及域随机化。
- **Real**: 通过 ROS2 话题驱动物理硬件，API 接口完全一致。

所有后端均返回冻结数据类 (`Pose2D`、`LaserScan`、`RGBDImage`)，字段与语义完全统一，由共享类型定义强制保证。

3.3 API 设计

以下示例展示了 VLM 驱动导航的完整工作流：

```
from threewe import Robot

async with Robot(backend="gazebo") as
    robot:
        image = robot.get_camera_image()
            # (H,W,3) uint8
        scan = robot.get_lidar_scan()
            # LaserScan
        result = await robot.
            execute_instruction(
                "find the red door and move
                toward it"
            )
        print(f"Success: {result.success}")
    )
```

切换至真实硬件仅需修改后端参数，无需任何代码改动、重新编译或配置变更。

3.4 硬件规格

表 ?? 列出了标准 SKU 的参考硬件配置。

所有设计遵循 CERN-OHL-P v2 协议发布，提供 KiCad 8 源文件、Gerber 制造文件及逐步装配文档。

4 仿真到现实一致性

4.1 传感器噪声模型

Mock 与 Gazebo 后端注入了与参考硬件物理特性匹配的噪声:

- **激光雷达 (LD06):** 高斯距离噪声 $\sigma = 8 \text{ mm}$
- **IMU (BNO055)** : 陀螺仪噪声 $\sigma = 0.0014 \text{ rad/s}$
- **里程计:** 麦克纳姆轮打滑率 5–15% 随机变化

4.2 迁移验证协议

本文定义了五项具有明确通过标准的标准迁移测试:

总体通过标准要求 $\geq 80\%$ 的测试项目通过。表 ?? 中的迁移率数据来自基于记录噪声模型的初步硬件验证; 大规模验证正在进行中。

5 基准测试套件

5.1 任务定义

基准测试套件定义了三类难度递增的导航任务:

- **点导航 (PointNav):** 导航至 (x, y) 坐标。评估指标: 成功率 (SR)、SPL。
- **目标导航 (ObjectNav):** 寻找目标物体类别。评估指标: SR、SPL。
- **探索 (Exploration):** 在时间预算内最大化区域覆盖率。评估指标: 覆盖率。

5.2 评估环境

七个标准化场景覆盖多样化环境: office_v2 (20×15m)、apartment_v1 (8×10m)、corridor_v1 (50m 走廊)、warehouse_v1 (30×20m)、cluttered_v1 (5×5m)、outdoor_v1 (50×30m) 及 dynamic_v1 (15×15m, 含动态障碍物)。

每个场景提供 10–50 对预定义起点/目标点以保证可复现性, 使用种子 [0, 99] 进行 100 轮评估。

5.3 基线结果

表 ?? 展示了 Nav2 DWB 规划器在代表性场景中的基线结果。

上述基线结果作为社区排行榜的参考基准, 可通过以下命令完全复现:

```
threewe benchmark run --task pointnav
\
--scene office_v2 --episodes 100
```

5.4 Gymnasium 集成

面向强化学习研究者, 3we 提供标准 Gymnasium 环境接口:

```
import gymnasium as gym
env = gym.make("3we/Navigation-v1",
               scene="office_v2",
               backend="mock")
obs, info = env.reset()
obs, reward, term, trunc, info = env.
step(action)
```

6 结论与未来工作

本文提出了 3we, 一套完全开源的具身智能研究基础设施, 通过统一的 Python API 弥合了仿真与现实之间的鸿沟。核心贡献包括: (1) 跨四个后端的零代码 Sim2Real 迁移; (2) BOM 不超过 300 美元的可复现开放硬件; (3) 将导航代码复杂度降低 12 倍的 AI 优先 API 设计; (4) 附社区基线的标准化基准测试体系。

未来方向包括: 基于 Isaac Sim 域随机化的大规模强化学习、兼容 LeRobot/Open-X 格式的社区轨迹数据共享平台、支持第三方机器人的硬件抽象层, 以及 VLA 模型在边缘硬件上的系统化部署评估。

本平台已在 <https://github.com/telleroutlook/3we-robot-platform> 公开发布, 分别遵循 Apache 2.0 (软件)、CERN-OHL-P v2 (硬件) 及 CC-BY-SA 4.0 (文档) 协议。

参考文献

- [1] Peter Anderson, Angel Chang, Devendra Singh Chaplot, Alexey Dosovitskiy, Saurabh Gupta, Vladlen Koltun, Jana Kosecka, Jitendra Malik, Roozbeh Mottaghi, Manolis Savva, et al. On evaluation of embodied navigation agents. In *arXiv preprint arXiv:1807.06757*, 2018.
- [2] Manolis Savva, Abhishek Kadian, Oleksandr Maksymets, Yili Zhao, Erik Wijmans, Bhavana Jain, Julian Straub, Jia Liu, Vladlen Koltun, Jitendra Malik, et al. Habitat: A platform for embodied ai research. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9339–9347, 2019.
- [3] OpenAI. Gpt-4v(ision) system card. *OpenAI Technical Report*, 2023.
- [4] Anthony Brohan, Noah Brown, Justice Carbajal, Yevgen Chebotar, Xi Chen, Krzysztof Choro-manski, Tianli Ding, Danny Driess, Avinava Dubey, Chelsea Finn, et al. Rt-2: Vision-language-action models transfer web knowledge to robotic control. *arXiv preprint arXiv:2307.15818*, 2023.
- [5] Kevin Black, Noah Brown, Danny Driess, Adnan Esber, Michael Suber, Brian Ichter, Pierre Sermanet, et al. π_0 : A vision-language-action flow model for general robot control. *arXiv preprint arXiv:2410.24164*, 2024.
- [6] NVIDIA Corporation. Isaac sim: Scalable robotics simulation. *NVIDIA Developer*, 2023.
- [7] Steven Macenski, Tully Foote, Brian Gerkey, Chris Lalancette, and William Woodall. Robot operating system 2: Design, architecture, and uses in the wild. *Science Robotics*, 7(66):eabm6074, 2022.
- [8] Open Robotics. Turtlebot. <https://www.turtlebot.com/>, 2023.
- [9] Linorobot. Linorobot2: Autonomous mobile robot framework. <https://linorobot.org/>, 2023.
- [10] Eric Kolve, Roozbeh Mottaghi, Winson Han, Eli VanderBilt, Luca Weihs, Alvaro Herrasti, Daniel Gordon, Yuke Zhu, Abhinav Gupta, and Ali Farhadi. Ai2-thor: An interactive 3d environment for visual ai. In *arXiv preprint arXiv:1712.05474*, 2017.
- [11] Remi Cadene, Simon Alibert, Pierre Sermanet, et al. Lerobot: State-of-the-art machine learning for real-world robotics. *GitHub repository*, 2024.