

# 去中心化视觉基于自主空中野生动物监测

Makram Chahine<sup>\*,†,1</sup>, William Yang<sup>\*,1</sup>, Alaa Maalouf<sup>1</sup>, Justin Siriska<sup>2</sup>,  
Ninad Jadhav<sup>2</sup>, Daniel Vogt<sup>2</sup>, Stephanie Gil<sup>2</sup>, Robert Wood<sup>2</sup>, and Daniela  
Rus<sup>1</sup>

\* Equal contribution,

<sup>1</sup>CSAIL MIT, Cambridge MA 02139, USA,

<sup>2</sup>SEAS Harvard University, Boston MA 02134, USA,

<sup>†</sup>Corresponding author: [chahine@mit.edu](mailto:chahine@mit.edu)

**摘要** 野生动物野外作业需要高效的并行部署方法来识别和与特定个体互动，从而实现同时的群体行为分析以及健康和安全干预。之前的机器人解决方案是从畜群的角度来解决问题，或者是由人工操作且规模有限。我们提出了一种用于野生动物监测的去中心化视觉多旋翼系统，该系统可扩展、带宽低且传感器最少（单个机载 RGB 相机）。我们的方法能够在自然栖息地中稳健地识别和跟踪大型物种。我们开发了新颖的基于视觉的协调和跟踪算法，这些算法设计用于动态无结构环境，并不依赖于集中式通信或控制。通过真实世界的实验验证了我们的系统，在各种野外条件下展示了可靠的部署。

**Keywords:** 分散式机器人技术，野生动物监测

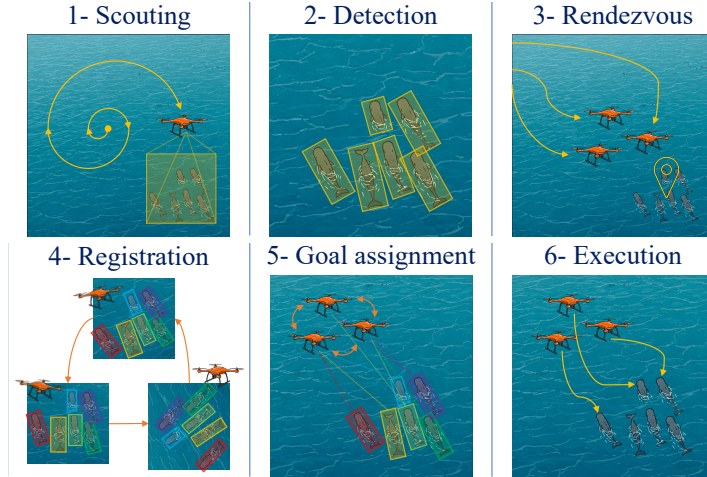


图 1. 用于野生动物监测的协同多旋翼系统。

## 1 介绍

大规模监测野生动物需要能够在一个大型环境中进行并行、个体级观察的系统。特别是，理解物种行为和交流能力要求能够在同一时间识别和追踪多个动物——这对行为分析、健康监控及干预至关重要。传统的监测策略通常是手动的、集中的或在规模和适应性上受到限制。

我们解决了利用自主空中代理在野外检测和追踪个别动物的问题。我们的解决方案针对诸如海洋哺乳动物监测等场景，其中远程检测和实时协调至关重要。以抹香鲸为例，我们介绍了一个完全自主的、分散化的空中系统，该系统仅依赖于机载 RGB 视觉来检测物种、跨代理注册身份、分配目标并执行监控行为——完全不需要集中控制或 GPS。

该系统引入了新的模块，用于分散式视觉注册 (Box-ICP)、基于图神经网络的任务分配以及在通信和感知约束下的端到端现实世界部署。通过实际和模拟实验，我们验证了其在非结构化、动态环境中执行大型动物群体级监控的能力。

## 2 相关工作

**基于视觉的检测与导航。**目标检测在过去十年中取得了显著进展，最初是通过两阶段框架如 Faster R-CNN [1]，该框架将区域提议和分类分离，随后是诸如 SSD [2] 和 YOLO [3] 等高效单阶段模型。更近一些的检测器集成了变换器架构，如 DETR [4]，或通过改进卷积设计提高推理速度和鲁棒性，例如在 YOLOv8 和 YOLOv12 [5,6] 中的应用。这些进展使得高性能的车载检测成为可能，适用于无人机上的实时部署。

同时，基于视觉的导航已从经典模仿学习系统 [7,8] 发展到现代端到端深度学习方法 [9]，这些方法在结构化环境中实现了稳健的导航。然而，这些模型在领域变化时常常表现不佳，并且缺乏可解释性。为了解决这些问题，最近的研究探索了安全意识学习 [10] 和鲁棒泛化技术，如液体神经网络 [11]、高斯散射 [12] 和模型驱动适应 [13,14,15]。此外，新兴范式现在整合基础模型以实现灵活的、指令条件下的策略 [16]，以及可扩展的互联网规模感知骨干网络以支持实时部署 [17]。

**目标分配。**线性和分配 (LSA) 问题 [18] 是多智能体匹配的核心公式，其中心化解决方案如匈牙利算法在机器人领域被广泛使用。可扩展的扩展

已被探索用于团队协调 [19]，而分散式替代方案则通过基于 Lyapunov 的方法 [20]、分布式匈牙利近似 [21] 和迭代共识协议 [22] 来解决有限的通信问题。

最近，基于学习的方法被开发出来以近似 LSA 解决方案。深度神经网络显示出了早期的潜力 [23]，而图神经网络（GNNs）为分散分配提供了一种结构化且可扩展的替代方案。GLAN [24] 将问题表述为二分边选择，而后继的工作 [25,26] 展示了在不同团队规模和条件下近乎最优的分散目标分配能力，使它们非常适合于现实世界的集群部署。

**可视注册。**注册技术对于在分布式代理之间对齐观察结果是基础的。经典方法包括迭代最近点（ICP）算法 [27]，该算法通过最近邻匹配和刚性变换来对齐点云。概率扩展如高斯混合模型（GMMs）[28] 和一致点漂移（CPD）[29] 结合了不确定性和软匹配。近期方法利用学习的表示 [30] 或结构约束，例如图或几何特征 [31]。这些方法为我们的 Box-ICP 模块的设计提供了信息，在该模块中，它在一个去中心化的视觉注册设置下在边界框边缘集之间结合了结构感知匹配。

**共识与可扩展协调。**分散式协调框架解决了鲁棒一致性协议 [32]、分散式编队飞行 [33] 和可扩展的局部博弈论规划 [34,35]。这些方法使代理能够以最少的通信计划安全轨迹，为我们设计低带宽、鲁棒的多代理控制提供了参考。

### 3 技术方法

提出的去中心化多代理野生动物监测管道由图 1 中描绘的六个算法组件构成，具体如下所示：

1. **侦察：**识别潜在的 [36] 区域，并使用螺旋轨迹进行搜索。
2. **检测：**使用视觉模型 [5] 进行扫描并分享小组的 GPS 坐标。
3. **会合点：**引导群集通过分散式编队控制形成群体 [33]。
4. **注册：**达成检测边界框的视觉 ID 共识 [5]。
5. **目标分配：**推断目标去中心化 GNN-LSA[26]。
6. **执行：**通过分割和跟踪 [17] 执行视觉监控

我们的主要贡献是设计了一个新颖、统一且连贯的基于视觉的系统，该系统涵盖了从侦查到个体任务执行的所有内容。管道的一个基石是视觉检测模块，它是在一个经过仔细整理的真实鲸鱼图像数据集上训练而成的，这些图像来自海洋中的实地镜头以及陆地上印刷海报的俯视图，从而增强了在不同

**Algorithm 1** 箱-ICP

---

**Require:** 箱子  $\mathcal{B}_1 = \{B_1^i\}_{i=1}^{N_1}, \mathcal{B}_2 = \{B_2^j\}_{j=1}^{N_2}$   
**Ensure:** 变换  $T \in \mathbb{R}^{4 \times 4}$ , 匹配  $\sigma: \{1, \dots, N_1\} \rightarrow \{1, \dots, N_2\}$

- 1: 初始化  $T \leftarrow I_4$
- 2: **while** not converged **do**
- 3:    $C \leftarrow 0 \in \mathbb{R}^{N_1 \times N_2}$
- 4:   **for**  $i = 1$  to  $N_1$  **do**
- 5:     **for**  $j = 1$  to  $N_2$  **do**
- 6:       计算  $D \in \mathbb{R}^{4 \times 4}$ :  $B_1^i$  和  $B_2^j$  之间的边距离
- 7:       在  $D$  上求解 LSA, 设置  $C_{ij} \leftarrow$  总匹配成本
- 8:     **end for**
- 9:   **end for**
- 10:  在  $C$  上求解 LSA 以获得对应关系  $\sigma$
- 11:  构成匹配框边的点集  $\mathcal{P}_1, \mathcal{P}_2$
- 12:  通过最小二乘对齐在  $(\mathcal{P}_1, \mathcal{P}_2)$  上估计  $T'$
- 13:   $T \leftarrow T' \cdot T$ , 更新  $\mathcal{B}_1 \leftarrow T \cdot \mathcal{B}_1$
- 14: **end while**
- 15: **return**  $T, \sigma$

---

条件下进行鲁棒检测的数据集。侦查时依赖于该检测来寻找场景中的鲸鱼，但更重要的是，代理群使用生成的边界框作为核心个体注册算法的输入，这使得细粒度的目标分配成为可能。

**侦察、检测、会合:** 通过部署侦察代理按照预定义的搜索模式（例如增长螺旋轨迹）在高潜力区域（由实地团队猜测或传感器信息提供 [36]）周围进行侦查。侦察运行检测模型，并在获得一个鲁棒的鲸鱼存在信号后（最近图像缓存中 80% 的帧具有检测结果），会引导其视觉追踪所有边界框的过滤质心。其当前位置用于作为去中心化编队飞行群的目标。

我们的设计关键在于注册和目标分配模块，这两个模块都在严格的通信限制下仅使用视觉输入进行操作。为了确保可扩展性和可靠性，我们将代理之间的通信限制为环形拓扑结构，其中每个无人机只与其两个直接相邻的无人机进行通信。实际上，我们的技术贡献集中在基于视觉的分散式注册和目标分配。

**单独注册:** 我们设计了一个方案来对齐连续代理之间检测到的边界框，并在动物身份上达成视觉共识。我们的 Box-ICP 算法源自 ICP 方法，并进行了适应以纳入将框分配给各个代理的约束条件，从而对配对代理的身份进行对



齐 (算法 1)。事实上, 在接收到来自其前驱者的边界框坐标后, 一个代理迭代运行以下过程直到收敛: (i) 构建矩阵  $D$  表示每对边界框之间的距离, 通过确定每对边框边缘之间的对应关系来解决边缘距离矩阵上的 LSA。(ii) 紧接着在  $D$  上解决另一个 LSA 以最小化匹配成本, 并计算边界框对应关系 (与经典 ICP 不同, 后者不利用结构约束而直接使用所有点计算距离)。(iii) 通过点云中的所有点的最小二乘法来计算最优变换。

沿环形通信图, 代理对相继对其身份表示。当返回到环中的第一个代理时, 可以验证身份共识。我们假设所有无人机都配备了朝下的摄像头, 并且图像中包含所有个体, 这可以通过控制蜂群的编队和高度来确保。

**目标分配:** 我们基于目标分配的去中心化图神经网络 (DGNN-GA) 框架 [26], 对其进行适应, 以提高在低通信密度设置下的性能, 并适用于目标数量超过代理的情况。训练数据通过在单位正方形上均匀采样代理和目标生成, 根据代理相对于中间点的角度位置循环连接代理。每个代理可获得的目标数量受到限制, 并从其最近邻中选择。使用欧几里得距离作为定义目标成本的度量标准, 并采用集中式匈牙利算法获取精确解来标记训练数据。对于每种  $n_a$  和  $n_g$  的组合, 我们生成 50k 个图。我们将适用范围扩展到目标数量超过代理的情况, 通过替换要求每个目标都分配的约束条件, 改为使用惩罚重复目标分配的损失函数, 如下文更正式地讨论。

设  $A \in \mathbb{R}^{n_a \times n_g}$  为软分配矩阵 (行  $i$ , 记作  $A_i$ , 对应于代理  $i$  的目标分配概率)。令  $\hat{A} \in \{0, 1\}^{n_a \times n_g}$  为通过逐行  $\arg \max$  获得的硬分配矩阵  $A$ 。

任务有效性损失。鼓励每个代理被分配到恰好一个目标:

$$\mathcal{L}_{\text{validity}} = \left\| \mathbf{1} - \sum_{i=1}^{n_a} \hat{A}_i \right\|_2 + \left\| \mathbf{1} - \left\| \hat{A}_i \right\|_2 \right\|_2$$

2. 分配多样性损失。鼓励智能体通过最小化它们的分配向量 (与  $\epsilon > 0$ ) 之间的余弦相似性来选择不同的目标:

$$\mathcal{L}_{\text{div}} = \sum_{i=1}^{n_a} \sum_{j=i+1}^{n_a} \frac{A_i \cdot A_j}{\max(\|A_i\|_2, \epsilon) \cdot \max(\|A_j\|_2, \epsilon)}$$

3. 监督优化损失。基于真实标签的二元交叉熵损失  $Y \in \{0, 1\}^{n_a \times n_g}$ :

$$\begin{aligned} \mathcal{L}_{\text{pos}} &= -0.9 \cdot \mathbb{E}[Y \cdot \log(A + \epsilon)] \\ \mathcal{L}_{\text{neg}} &= -0.1 \cdot \mathbb{E}[(1 - Y) \cdot \log(1 - A + \epsilon)] \\ \mathcal{L}_{\text{CE}} &= \mathcal{L}_{\text{pos}} + \mathcal{L}_{\text{neg}} \end{aligned}$$

最终损失。结合结构和监督的总损失：

$$\mathcal{L}_{\text{total}} = \alpha \cdot (\mathcal{L}_{\text{validity}} + \mathcal{L}_{\text{div}}) + (1 - \alpha) \cdot \mathcal{L}_{\text{CE}}, \quad \text{with } \alpha = 0.5$$

**执行：**每个代理使用其分配目标对应的最新可用帧的中心点来初始化分割掩码，如 [17] 中所述。以约 20 赫兹的实时运行频率跟踪获得的掩码可以实现对各个目标的稳健监控。

## 4 实验

我们设计了一系列实验，从受控离线评估到实际的多无人机部署，以单独验证系统中的每个模块以及在完整管道内的组合。

**侦察与检测：**我们的视觉检测模型是在多米尼加海岸收集的真实无人机捕鲸录像以及在鲸鱼海报上进行的飞行测试数据上训练的，增强了数据集的多样性。训练图像由手工标注，总数据集包含 1407 帧，其中 90% 用于训练，10% 用于验证以评估离线性能。现实世界的测试涉及四旋翼无人机沿足球场上鲸鱼海报的螺旋轨迹飞行。我们验证了侦察无人机自主搜索、检测海报、在其上方保持位置并实时向基站传输其坐标的能力。我们还在来自实地的真实世界数据上离线运行该模块。

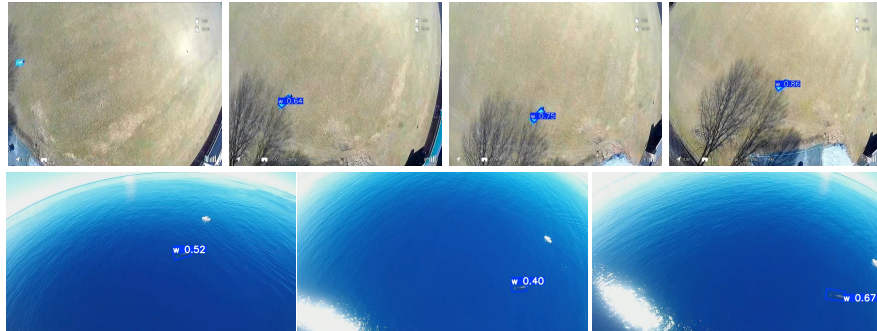
**检测与注册：**为了模拟各种鲸鱼的视角，我们从训练数据集外的空中镜头中合成 39 帧图像，并在三个不同的时间点拍摄，应用了平移、旋转和剪切变换，并叠加了一张海洋图片以处理空像素的问题。场景中共有 9 只鲸鱼。我们进行检测，得到的结果边界框作为 Box-ICP 匹配的点云。此测试重点考察算法对噪声边界框数据以及帧间不同重叠比例的处理能力。

**目标分配：**GNN 模型评估使用大规模运行，并通过随机化代理和目标初始化来定义图条件。衡量最优成本接近度和冗余分配的指标在不同的代理/目标配置（包括比代理更多的目标）下评估模型。由于这是一个仅基于成本矩阵（像素距离）的高级决策，因此在这一抽象级别上对模型进行评估，独立于飞行数据。此评估特别考察了 GNN 在稀疏通信图中的性能，反映了真实世界无人机群的限制。

**去中心化的注册和分配在现实世界中：**使用一架无人机从多个视角捕捉鲸鱼图像，我们在现实世界和实时环境中测试了分散式视觉检测、注册和目标分配组件的有效性。

**完整管道在现实世界中的应用：**关键测试涉及在无人机编队上执行整个管道的真实环境设置。为了初步验证，我们在一个受控环境中部署两架无人机，该环境中包含六张鲸鱼海报分布在足球场上。此设置涵盖了系统中的所有模块——从自主侦察和检测到分散式注册、基于 GNN 的目标分配以及多智能体跟踪——提供了对方法的完整闭环评估。

## 5 结果



**图 2.** 侦察检测运行。顶部一行展示了一个带有鲸鱼海报的完整测试，从左到右的列分别对应：鲸鱼出现、首次检测、无人机定位平均边界框像素位置、传输坐标。底部一行是检测器在野外手动飞行数据上的离线运行。

**侦察与检测：**检测器实现了 98.4% 的召回率和 94.7% 的精度，证明了其在准确识别鲸鱼的同时最小化误报方面的有效性。我们在现实高度下测试了自主搜索，并展示了足以应对实际搜寻任务的强大性能。事实上，相对于海报尺寸而言，测试相当于在 100 多米的高度进行真实飞行（视相机规格而定）。检测不仅足够精确以定位鲸鱼，而且还作为初始跟踪的视觉参考。图 2 显示了来自检测运行的画面。

已经训练了包括 YOLOv8、YOLOv11 和 DETR 在内的多个架构，我们确定前者在我们的数据集上的性能占优势（参见表 1）。

在一个由包含 10 头鲸鱼的单帧构建的自定义评估数据集上，我们使用图像编辑工具移除额外的鲸鱼来生成各种增强和目标较少的例子（图 4 中的

表 1. 我们在鲸鱼数据集上对检测模型的比较。OA: 整体准确率, FPR: 假正率, FNR: 假负率。

| 模型       | OA (%)       | FPR (%)     | 假阴性率 (%)    |
|----------|--------------|-------------|-------------|
| YOLOv8s  | <b>98.29</b> | <b>1.05</b> | <b>0.66</b> |
| YOLOv8n  | 96.88        | 1.95        | 1.17        |
| YOLOv11n | 96.25        | 2.46        | 1.29        |
| YOLOv11s | 95.81        | 1.18        | 3.01        |
| DETR     | 58.08        | 9.70        | 32.22       |

$n_g = 5$  示例)。我们评估检测模型在这些图像中准确检测出鲸鱼数量的能力, 并在图 3 中提供了不同数量的鲸鱼以及三个检测置信度阈值的成功率。

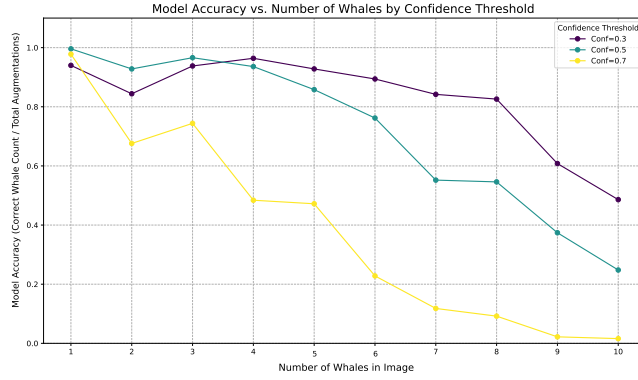


图 3. 检测成功率与鲸鱼数量的关系 (每个数据点  $N = 200$ )。

**检测与注册:** 在所有 9 头鲸鱼场景的 39 个视图中, 我们的检测模块正确地检测到了数量正确的框, 并且检查时位置合适。我们评估了 Box-ICP 对 200 对帧中的个体进行匹配的一对一成功率, 这些帧来自 741 种可能的组合。每一个分配都是正确的, 示例如图 4 所示。在一个  $n$ -无人机环形网络中, 总体成功率为  $s(n) = s_p^n$ , 其中一对一的成功率是  $s_p = s_{\text{det}}^2 \cdot s_{\text{reg}}$ 。检测器在多鲸鱼设置中的可靠性以及注册算法的完美准确性确保了可扩展性, 尽管几何成功率有所下降, 但系统已经准备好进行现实世界的原位测试。

**目标分配:** 我们使用最优得分 (% 最优分配) 和多样性得分 (% 非重叠解决方案) 来评估模型。结果汇总在表 2 中, 基于 5000 次随机初始化的平均值, 其中 GNN 实例是通过训练 5 个智能体、10 个目标和 5 轮消息传递得到的。我们通过对  $n_g < 10$  的情况添加  $10 - n_g$  高成本幽灵目标来处理这些

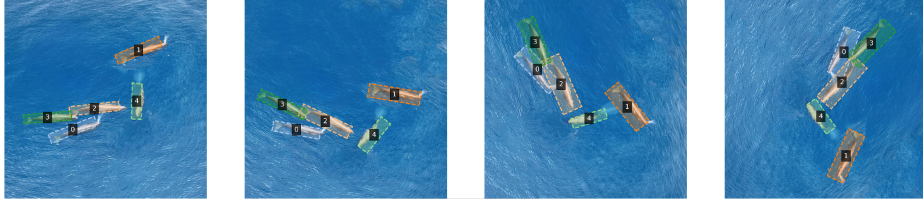


图 4. 检测并针对最左侧帧进行个体注册。

情形。我们在目标数量保持接近训练设置的情况下（直到  $n_g = 7$ ）实现了高度非重叠且接近最优的分配。比智能体更多的目标是一种理想的去中心化解决方案，因为即使是单个额外的目标也能在大约 80% 的时间里帮助避免重叠。通过已知的车队规模，我们可以训练几个合适的 GNN 来提供稳健的性能，覆盖可能遇到的各种个体数量。

表 2. 目标分配结果  $n_a = 5$

| $n_g$          | 5    | 6    | 7    | 8    | 9    | 10   |
|----------------|------|------|------|------|------|------|
| Optimality (%) | 64.1 | 67.0 | 73.2 | 84.9 | 91.3 | 92.2 |
| Diversity (%)  | 34.3 | 78.8 | 91.2 | 93.3 | 93.6 | 94.6 |

**分散式注册和分配在现实世界中的应用：**我们使用来自 3 个视角的帧来模拟 3 个代理的存在，以证明该系统的可行性。去中心化的基于视觉的解决方案实时运行，用于检测鲸鱼、对齐表示并分配目标，如图 5 所示。

**完整的真实世界管道：**我们成功地在两架飞越足球场的无人机上部署了整个管道，六张鲸鱼海报被放置为目标。一段演示完整步骤执行任务的视频包含在补充材料中，并且可以通过这个链接在线查看。

**通信带宽：**每头鲸鱼都被编码为一个四坐标框，由每个 8 值组成，每个值大小为 2 字节。每头鲸鱼还有一个相应的索引，可以用 1 字节表示，因为我们不期望在一个给定的拍摄中有超过  $2^8$  头鲸鱼。因此，对于  $N_w$  头鲸，我们预计使用  $9N_w$  字节。

无人机还交换隐藏状态向量以促进目标分配。假设我们的隐藏状态维度为  $d_h$ ，其大小则是  $4 \cdot d_h$  字节。

在我们的 GNN 中，我们使用  $d_h = 32$ 。由于我们期望  $N_w < 20$ ，我们的总消息大小上限为  $32 \cdot 4 + 9 \cdot 20 = 308$  字节。假设无线电连接至少为 1MHz，我们的延迟将大约为 0.002 秒。

## 6 讨论

我们提出的管道涵盖了从自主侦察、鲁棒检测、群聚收集、个体注册、目标分配到监控执行的所有任务方面，代表了基于视觉的实时野生动物监测的一种新方法。在去中心化注册和 GNN 驱动的任务分配方面的技术进步，在真实世界数据上的验证表明，这对于生态应用中的可扩展且高效的多无人机协调是一个重要进展。此外，完整的任务硬件验证证明了该系统作为研究野外动物物种的现场科学家的有效而高效的现实解决方案的可行性。





图5. 现实世界中的实时部署，包括注册和目标分配，涉及3个代理（每列一个）和3个目标。第一行显示原始帧，第二行描绘了从每个代理的角度检测到的内容及其初始身份，第三行展示了 Box-ICP 对齐后的表示形式，最后一行以红色显示分配给每个代理的目标。

## 参考文献

1. S. Ren, K. He, R. Girshick, J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks (2016)
2. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.Y. Fu, A.C. Berg, *SSD: Single Shot MultiBox Detector* (Springer International Publishing, 2016), p. 21 – 37. DOI 10.1007/978-3-319-46448-0\_2
3. J. Redmon, S. Divvala, R. Girshick, A. Farhadi. You only look once: Unified, real-time object detection (2016)
4. N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, S. Zagoruyko. End-to-end object detection with transformers (2020)
5. D. Reis, J. Kupec, J. Hong, A. Daoudi. Real-time flying object detection with yolov8 (2024)
6. Y. Tian, Q. Ye, D. Doermann. Yolov12: Attention-centric real-time object detectors (2025)
7. D.A. Pomerleau, Alvin: An autonomous land vehicle in a neural network, *Advances in neural information processing systems* **1** (1988)
8. M. Bojarski, D. Del Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L.D. Jackel, M. Monfort, U. Muller, J. Zhang, et al., End to end learning for self-driving cars, *arXiv preprint arXiv:1604.07316* (2016)
9. P.S. Chib, P. Singh, Recent advancements in end-to-end autonomous driving using deep learning: A survey, *IEEE Transactions on Intelligent Vehicles* (2023)
10. W. Xiao, T.H. Wang, R. Hasani, M. Chahine, A. Amini, X. Li, D. Rus, Barriernet: Differentiable control barrier functions for learning of safe robot control, *IEEE Transactions on Robotics* (2023)
11. M. Chahine, R. Hasani, P. Kao, A. Ray, R. Shubert, M. Lechner, A. Amini, D. Rus, Robust flight navigation out of distribution with liquid neural networks, *Science Robotics* **8**(77), eadc8892 (2023)
12. A. Quach, M. Chahine, A. Amini, R. Hasani, D. Rus. Gaussian splatting to real world flight navigation transfer with liquid networks (2024). URL <https://arxiv.org/abs/2406.15149>
13. T.H. Wang, W. Xiao, M. Chahine, A. Amini, R. Hasani, D. Rus, in *Learning for Dynamics and Control Conference* (PMLR, 2023), pp. 1099–1111
14. L. Yin, M. Chahine, T.H. Wang, T. Seyde, C. Liu, M. Lechner, R. Hasani, D. Rus, in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (2023), pp. 6334–6341. DOI 10.1109/IROS55552.2023.10342229



15. E. Kaufmann, L. Bauersfeld, A. Loquercio, M. Mueller, V. Koltun, D. Scaramuzza, Champion-level drone racing using deep reinforcement learning, *Nature* **620**, 982 (2023). DOI 10.1038/s41586-023-06419-4
16. M. Chahine, A. Quach, A. Maalouf, T.H. Wang, D. Rus. Flex: End-to-end text-instructed visual navigation with foundation models (2024)
17. A. Maalouf, N. Jadhav, K.M. Jatavallabhula, M. Chahine, D.M. Vogt, R.J. Wood, A. Torralba, D. Rus, Follow anything: Open-set detection, tracking, and following in real-time, *IEEE Robotics and Automation Letters* **9**(4), 3283 (2024). DOI 10.1109/LRA.2024.3366013
18. H.W. Kuhn, The hungarian method for the assignment problem, *Naval Research Logistics Quarterly* **2**(1-2), 83 (1955). DOI <https://doi.org/10.1002/nav.3800020109>
19. H. Aziz, A. Pal, A. Pourmiri, F. Ramezani, B. Sims. Task allocation using a team of robots (2022)
20. D. Panagou, M. Turpin, V. Kumar, in *2014 IEEE International Conference on Robotics and Automation (ICRA)* (2014), pp. 6757–6762. DOI 10.1109/ICRA.2014.6907857
21. S. Ismail, L. Sun, in *2017 International Conference on Unmanned Aircraft Systems (ICUAS)* (2017), pp. 23–28. DOI 10.1109/ICUAS.2017.7991447
22. S. Chopra, G. Notarstefano, M. Rice, M. Egerstedt, A distributed version of the hungarian method for multirobot assignment, *IEEE Transactions on Robotics* **33**(4), 932 (2017). DOI 10.1109/TRO.2017.2693377
23. M. Lee, Y. Xiong, G. Yu, G.Y. Li, Deep neural networks for linear sum assignment problems, *IEEE Wireless Communications Letters* **7**(6), 962 (2018). DOI 10.1109/LWC.2018.2843359
24. H. Liu, T. Wang, C. Lang, S. Feng, Y. Jin, Y. Li. Glan: A graph-based linear assignment network (2022)
25. C. Aironi, S. Cornell, S. Squartini, in *Applied Intelligence and Informatics*, ed. by M. Mahmud, C. Ieracitano, M.S. Kaiser, N. Mammone, F.C. Morabito (Springer Nature Switzerland, Cham, 2022), pp. 90–101
26. M. Goarin, G. Loianno, Graph neural network for decentralized multi-robot goal assignment, *IEEE Robotics and Automation Letters* **9**(5), 4051 (2024). DOI 10.1109/LRA.2024.3371254
27. Y. Chen, G. Medioni, Object modelling by registration of multiple range images, *Image and vision computing* **10**(3), 145 (1992)

28. A. Joshi, C.H. Lee, On the problem of correspondence in range data and some inelastic uses for elastic nets, *IEEE transactions on neural networks* **6**(3), 716 (1995)
29. A. Myronenko, X. Song, Point set registration: Coherent point drift, *IEEE transactions on pattern analysis and machine intelligence* **32**(12), 2262 (2010)
30. Y. Wang, J.M. Solomon, in *Proceedings of the IEEE International Conference on Computer Vision* (2019), pp. 3523–3532
31. P.K. Agarwal, J.M. Phillips, in *Proceedings of the Eighteenth Canadian Conference on Computational Geometry*, Kingston, Canada (2006), pp. 143–146
32. W.M. Haddad, M. Chahine, A hybrid thermodynamic control protocol for semistability and consensus of network systems with intermittent information, *IEEE Transactions on Automatic Control* **66**(6), 2754 (2021). DOI 10.1109/TAC.2020.3008301
33. S. Leung, O. Montenbruck, Real-time navigation of formation-flying spacecraft using global-positioning-system measurements, *Journal of Guidance, Control, and Dynamics* **28**(2), 226 (2005). DOI 10.2514/1.7474
34. M. Chahine, R. Firoozi, W. Xiao, M. Schwager, D. Rus, in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (2023), pp. 880–887. DOI 10.1109/IROS55552.2023.10341677
35. M. Chahine, R. Firoozi, W. Xiao, M. Schwager, D. Rus, Intention communication and hypothesis likelihood in game-theoretic motion planning, *IEEE Robotics and Automation Letters* **8**(3), 1223 (2023). DOI 10.1109/LRA.2023.3238131
36. N. Jadhav, S. Bhattacharya, D. Vogt, Y. Aluma, P. Tønnesen, A. Prabhakara, S. Kumar, S. Gero, R.J. Wood, S. Gil, Reinforcement learning – based framework for whale rendezvous via autonomous sensing robots, *Science Robotics* **9**(95), eadn7299 (2024). DOI 10.1126/scirobotics.adn7299