不确定检测下结合蒙特卡洛规划的无监督主动视觉搜索

Francesco Taioli、Francesco Giuliari、Yiming Wang、Riccardo Berra、Alberto Castellini、Alessio Del Bue、*IEEE* 会员，Alessandro Farinelli、Marco Cristani、*IEEE* 会员，以及 Francesco Setti、*IEEE* 会员

arXiv:2303.03155v1 [cs.RO] 6 Mar 2023

摘要——我们提出了一种用于在环境中进行物体主动视觉搜索的解决方案，其中已知的唯一信息是其二维平面图。与最先进水平的方法相比，我们的解决方案具有三项关键特性，使其对检测器故障更具合理性和鲁棒性：*(i)* 它是无监督的，因为不需要任何训练阶段。*(ii)* 在探索过程中，根据一种直观机制更新二维平面图上的概率分布，而改进的信念更新提高了智能体探索的有效性。*(iii)* 我们通过利用特定检测器的成功统计数据，将物体检测器可能失效这一认知纳入上述概率建模中。我们的解决方案称为 POMP-BE-PD（基于 POMCP 的在线运动规划：通过探索和概率检测进行信念建模）。它利用智能体的当前位姿和 RGB-D 观测，通过由蒙特卡洛规划方法求解的 POMDP 来学习最优搜索策略。在主动视觉数据库基准上，与竞争方法相比，我们将所有环境中的平均成功率显著提高了 35%，同时将平均路径长度降低了 4%。因此，即使不使用任何训练程序，我们的结果仍达到了最先进水平。

索引词—主动视觉搜索、目标物体导航、部分可观测马尔可夫决策过程、部分可观测蒙特卡洛规划、在线策略学习

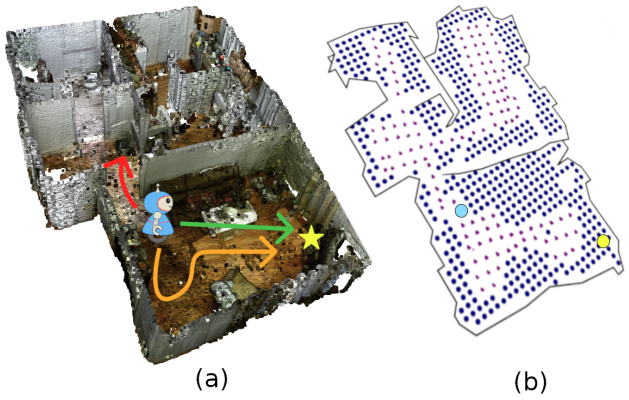


图 1. 智能体在已知环境中初始化，任务是通过视觉搜索目标物体，即定位该物体并靠近它。(a) 环境的三维重建；智能体必须避开较长的路线（橙色标示），沿可能的最短路径（绿色标示）前往目标（黄色星形），同时避免完全错过目标（红色标示）。(b) 我们在 POMCP 建模中对应的场景二维网格地图：蓝点表示可能的物体位置，紫色叉号表示可能的机器人位姿。

1 引言

AMONG 机器人学中最有趣的领域之一是主动视觉搜索（AVS）问题 [1], ，在这一问题中，智能机器人智能体必须自主寻找远处的物体，并通过自我中心视觉传感器移动和探索周围环境。AVS适用于许多不同的场景，例如智能家居领域 [2], [3], 、个人辅助 [4], 、搜救 [5], [6], 以及极具吸引力的火星探索 [7]。本文聚焦于室内环境中的AVS问题 [8],，其中唯一可用的信息是二维地图。我们提出一种综合运动规划策略的方法，该策略能够决定如何基于感知模块移动智能体，以目视检测并接近特定物体，即目标（见图 1）。

在具有自我中心视角的真实场景中执行 AVS 是一项极具挑战性的问题，这是由于观测质量难以预测——即远处的物体、运动模糊和低分辨率——以及场景杂乱导致的局部视野和遮挡，还有对新环境的泛化能力不足。这不仅会影响目标检测，也会影响规划策略。为应对这一挑战，近期的工作主要基于深度强化学习（RL），例如，以深度视觉嵌入 [9], [10] 为输入的深度循环 Q 网络（DRQN）。要训练此类 DRQN 模型，需要大量数据，即来自多个真实场景或模拟环境、涵盖成功和失败搜索回合的不同长度观测序列。

本文采取了不同的视角，提出了一种用于 AVS 的在线学习方法。其基本思路不是利用海量训练数据来学习一个完整的策略，而是采用一种先进的规划方法，根据从环境中收集的观测构建环境配置，并据此制定最佳行动。

• *F. Taioli*、*R. Berra*、*A. Castellini*、*A. Farinelli*、*M. Cristani* 和 *F. Setti* 任职于意大利维罗纳大学创新医学工程系。

• *F. Giuliari* 任职于意大利热那亚的意大利理工学院（*IIT*）模式分析与计算机视觉（*PAVIS*）研究组，并隶属于热那亚大学（*UniGe*）。

• *A. Del Bue* 任职于意大利热那亚的意大利理工学院（*IIT*）模式分析与计算机视觉（*PAVIS*）研究组。

• *Y. Wang* 任职于意大利特伦托的布鲁诺·凯斯勒基金会（*FBK*）深度视觉学习（*DVL*）部门。

而这些观测是在环境中收集的 t.

这种方法论上的根本转变是通过考量部分可观测蒙特卡洛规划（POMCP）方法 [11]实现的。POMCP 已应用于 rocksample、battleship 和*pocman* （部分可观测吃豆人）等基准问题，并取得了令人印象深刻的结果；然而，将其用于机器人应用仍是一个开放且具有挑战性的研究问题。

在我们之前的工作中 [12], ，我们提出了 POMP，这是一种在线策略学习方法，以智能体的当前位姿和一帧 RGB-D 帧作为输入，规划使智能体接近目标物体的下一步移动。我们将该问题建模为部分可观测马尔可夫决策过程，并通过蒙特卡洛规划方法求解，使我们能够同时探索环境并搜索物体。与强化学习方法不同，该方法的主要优势在于 POMP 不需要训练阶段，因此能够非常灵活地在小型和中等规模的真实场景中解决 AVS 问题。尽管在不使用任何训练数据的情况下取得了接近最先进水平的结果，但真实的物体检测器存在假阳性和漏检，准确性不足。因此，智能体可能会被检测器误导，在错误的位置终止探索，从而降低整体成功率。

为了解决这些问题，本文提出 POMP-BE-PD，这是 POMP 的一种扩展，它以概率形式描述观测模型，使我们能够更好地处理真实物体检测器产生的假阳性，并提升智能体探索的效果。

我们方法的可视化表示如图 2 所示。在每个时间步，我们将智能体在已知二维地图中的位姿——即位置和方向——以及传感器采集得到的一个 RGB-D 帧输入模型。我们对 RGB 图像使用现成的物体检测器，以识别目标物体（若存在）的边界框。随后利用目标候选区域的深度通道，获取其在地面地图中的候选位置。我们利用这些信息，构建目标物体所有候选位置上的概率分布。策略通过蒙特卡洛模拟在线学习，因此具有通用性，易于部署于任何环境中。当信念空间中的某个位置超过阈值时，POMCP 探索即告终止。

关键在于，我们的方法利用环境模型，考虑传感器的视场以及该区域内智能体所有允许的移动方式。对于我们的主动视觉搜索场景，可以通过构建环境地图轻松获得此类模型，其中包括障碍物、墙壁或家具等固定元素的位置。我们的运动策略明确利用环境知识进行可见性建模，而其他基于强化学习的策略 [9], [10] 则以数据驱动方式隐式编码此类环境知识。

探索阶段结束后，概率对接模块会引导智能体尽快接近目标位置——即与目标具有正面朝向视角的最近位姿。首先，我们在所有可能机器人位姿构成的图上估计最短路径 [13], 然后通过路径重规划来提高鲁棒性。

与我们之前的工作 [12], 相比，本文的主要贡献如下：

- 一种新策略，可在使用真实物体检测器时提升其对假阳性和漏检的鲁棒性：我们用概率检测取代确定性检测，并通过贝叶斯推断考虑所有可能物体位置上的概率分布；
- 一种新的POMCP信念更新策略，使我们能够降低探索的总路径长度，并提高其在大状态空间环境中的有效性；
- 一种新的对接方法，利用探索过程中收集的信息，提高应对上述问题的鲁棒性；
- 更深入的实验分析与结果讨论，提供主动视觉数据集中所有场景的结果；以及
- 对 POMCP视觉搜索方法的扩展描述，包含更多数学细节和讨论。

2 相关工作

与本研究相关的两个主要研究主题是主动视觉搜索和部分可观测马尔可夫决策过程中的规划。下文将简要综述这两个主题的主要研究，并重点介绍我们相较于当前最先进成果的原创贡献。

主动视觉搜索通常也称为物体目标导航，是具身人工智能研究领域中的一项特定任务。具身人工智能通过从自我中心视角与环境互动来学习，是一个新兴的研究领域。在这一领域中，AVS是一项专注于检测并接近特定物体 [14] 的任务。

早期方法利用中间物体——例如沙发与电视之间的关系——来缩小目标物体的搜索范围。尽管中间物体通常因尺寸较大而更容易检测，但它们与目标之间的空间关系可能并不固定。[15], 提出了一种概率方法：当检测到预期会共同出现的物体时，目标出现的可能性就会增加。

借助深度强化学习技术，基于深度学习的 AVS 得以实现 [9], [10], [16],，其中通常利用视觉神经嵌入进行动作策略训练。Han等人 [16]提出了一种深度Q网络（DQN），其智能体状态由描述当前 RGB 观测的 CNN特征以及已检测物体的边界框构成。然而，该方法假设物体必须先被检测到。为解决搜索任务，EAT [9] 从当前 RGB 观测以及由区域提议网络（RPN）生成的候选目标裁剪图像中提取特征，然后将这些特征输入动作策略网络。类似地，GAPLE [10] 利用深度通道中的深度信息丰富深度视觉特征，以进行策略学习。尽管 GAPLE 声称具有通用性，但其代价是训练成本高昂：GAPLE 使用基于合成 SUNCG数据集的模拟器 House3D 渲染出的 100 个场景进行训练。其他使用异步优势演员评论家（A3C）算法 [17], 、长短期记忆网络（LSTM）架构 [18], 以及与深度Q学习结合的 Transformer网络 [19]来学习最优策略的方法也面临同样的局限。

近期，社区也开始对 AVS 任务进行基准测试。包括 Habitat ObjectNav 在内的挑战 [20]旨在推动相关方法的发展，使智能体能够从随机初始位姿出发，在未知环境中仅凭感知输入进行导航，寻找给定对象类别的特定实例，

同时，研究者还开放了包含密集标注语义的三维真实世界空间大规模数据集，以便训练和测试模型 [21]。由于场景地图未知，大多数方法旨在通过结合语义探索与物体搜索的目标来学习策略 [22]。在此类语义探索的基础上，还可以进一步将关于观察位置的感知技能与导航相分离 [23] ，以提高成功率。此外，物体之间的空间关系也被表示为图，并通过图卷积网络进行嵌入，以指导导航策略 [24], ；外部常识知识也已显示出在利用空间图学习进行物体定位方面的优势 [3]。

一般来说，基于强化学习的策略依赖大量数据训练，以编码环境模型和运动策略。不同的是，我们提出的基于 POMCP 的方法会明确利用现有场景知识，并在线为智能体路径进行高效规划，无需额外的离线训练。

2.1 蒙特卡洛规划

在最优策略计算方面，部分可观测马尔可夫决策过程（部分可观测马尔可夫决策过程）是一种广泛使用的框架，用于表示不确定环境中的动态过程并解决相关的序贯决策问题 [25]。求解非平凡的部分可观测马尔可夫决策过程的精确解通常在计算上难以处理 [26], ，但近年来，开发近似求解器方面取得了显著进展。近年来，以近似方式求解部分可观测马尔可夫决策过程的一种高效策略是蒙特卡洛树搜索（MCTS） [27], [28], [29]。使用蒙特卡洛树搜索求解部分可观测马尔可夫决策过程的主要优势在于可扩展性。基于蒙特卡洛树搜索的策略会在线计算策略即仅针对智能体在其轨迹中访问的特定状态（或在部分可观测环境中的信念）进行计算。这对于状态空间非常大的领域以及信念空间维度无限的部分可观测环境至关重要，因为信念是状态上的概率分布。在蒙特卡洛树搜索系统中，系统状态表示为树的节点，动作/观测则表示为边。通过蒙特卡洛模拟生成搜索树，并采用特定的动作选择策略，例如应用于树的上置信界（UCT）算法 [30],，以高效平衡探索与利用。

利用蒙特卡洛树搜索（MCTS）的 POMDP 最具影响力的求解器是部分可观测蒙特卡洛规划（POMCP） [11] ，它结合了基于粒子滤波器的信念表示、基于 MCTS 的动作 Q 值计算策略，以及一种高效更新智能体信念的方法。POMCP 已衍生出多种扩展。BA-POMCP [31]将 POMCP 扩展至贝叶斯自适应部分可观测马尔可夫决策过程，使环境模型能够在执行过程中学习。在 [32], 中提出了一个用于多智能体部分可观测马尔可夫决策过程中可扩展规划的 POMCP 版本；该方法引入了 POMDP 中的模型学习，同时还在信念中考虑了模型参数的不确定性。在 [33] 中提出了一个用于处理成本约束的可扩展 POMCP 扩展。近期研究重点关注在 POMCP 中引入关于环境和策略的先验知识。在 [34] 中，利用已知的状态变量关系来提升 POMCP 的性能。在 [35]中，使用理论模可满足性（SMT）生成表示 POMCP 策略部分内容的逻辑规则，以提升策略的可解释性并识别由于参数调优错误而导致的异常动作选择。在将概率规划与符号方法相结合的研究方向中， [36]再次利用逻辑规则生成安全屏障，以提升 POMCP 的安全性。在 [37], [38]中进一步改进了该技术，其中采用主动方法和基于答案集程序归纳学习的方法来学习逻辑规则。

POMCP 已应用于多个领域。其中一些与部分已知环境的探索 [39] 以及机器人寻找与跟随人员 [40] 有关。另一些 [41], [42] 则涉及仅使用 POMCP 进行机器人导航，或采用分层方法，以 POMCP 负责高层控制、神经网络负责低层控制。近年来，广受欢迎的基于蒙特卡洛树搜索的方法也被用于开发在围棋中表现超人的智能体 [43], [44]。本文提出的方法与上述所有工作不同，因为它针对 AVS 领域对 POMCP 进行了专门化，并改进了信念更新、物体的概率检测和对接，而这些改进在现有文献中尚未出现。据我们所知，文献中唯一关于使用 POMCP 进行 AVS 的工作是[12], [45]。与这些工作的差异显著：[12] 采用标准的信念更新，假设目标检测完全准确，并使用一种朴素的对接流程；而 [45] 则聚焦于完全未知的环境。

3 方法

我们考虑一个智能体在已知环境中导航，以定位并接近特定物体为目标。智能体探索环境以识别目标物体，确定其在地图上的位置，然后向其移动。

为与相关文献保持一致，时间步 t 时智能体的位姿[1]为 $p_t = \{x_t, y_t, \theta_t\}$，其中 x 和 y 是地面平面上的坐标，θ 是方向。在每个时间步，智能体从预定义集合 A中采取一个动作 a_t ：具体而言，智能体可以_向前移动、_向后移动、_顺时针旋转，或_逆_时针旋转。旋转角度固定。

当智能体到达新的位姿 p_t时，会收到一项观测，该观测是将物体检测器应用于 RGB-D相机采集的图像后得到的输出。我们将搜索空间建模为栅格地图（见图 1(b)），其中每个单元格可能属于以下情况之一：*(i)*“视觉遮挡”，即单元格被障碍物占据，例如墙壁或一件家具，导致智能体无法穿过其视线；*(ii)* “空”，即智能体可以进入该单元格，因此其中不可能有物体；或者*(iii)* “候选位置”，即既非上述情况，因此该单元格可能是目标物体所在的位置。

3.1 部分可观测马尔可夫决策过程

我们将 AVS问题表述为部分可观测马尔可夫决策过程（ POMDP），这是一种用于对动态环境中不确定条件下的序贯决策过程进行建模的标准框架 [25]。POMDP 是一个元组 $(S, A, O, T, Z, R, \gamma)$，其中 S是有限的部分可观测状态集合，A 是有限的动作集合， Z 是有限的

1. 此处的位姿指二维机器人位姿，即位置和方向。本文不考虑与智能体结构相关的复杂运动学（例如使用人形机器人时的情况）。

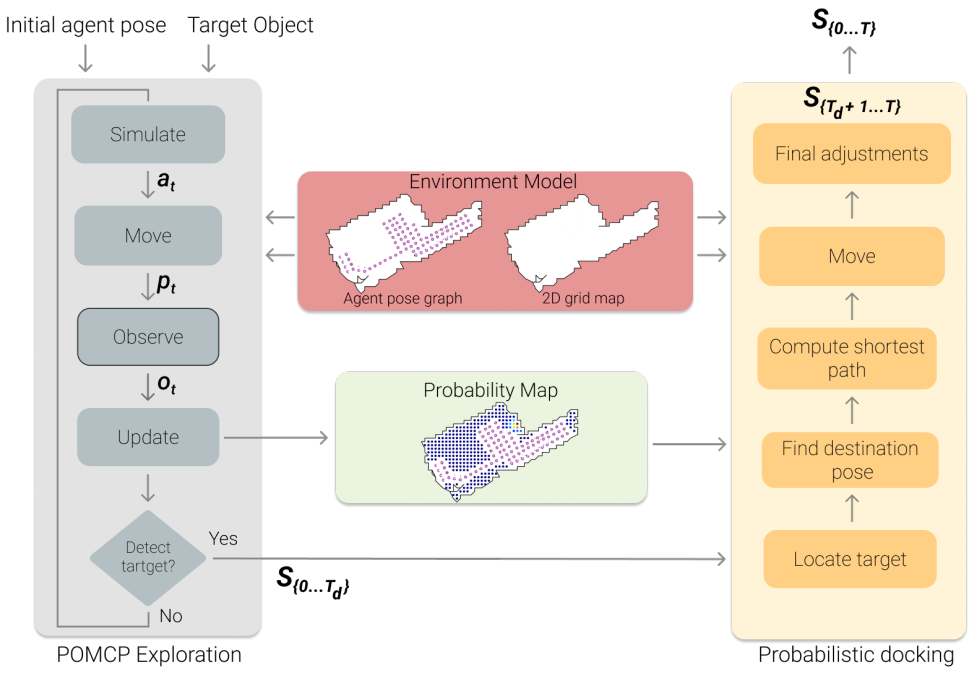


图 2. 我们提出的 POMP-BE-PD 方法的整体架构。红色方框表示输入 POMCP 模块的先验知识，灰色方框表示用于检测目标物体的探索策略，黄色方框表示用于到达目标位姿的概率对接策略，绿色方框表示位置上的概率分布。数学符号：状态 s_t，动作 a_t，位姿 p_t，观测 o_t，POMCP 状态序列 $s_{\{0..T_d\}}$，对接状态序列 $s_{\{T_{d+1}..T\}}$，完整状态序列 $s_{\{0..T\}}$。

观测集合， T: $S \times A \to \Pi(S)$ 是状态转移模型，O: $S \times A \to \Pi(Z)$ 是观测模型， R: $S \times A \to \mathbb{R}$是奖励函数， $\gamma \in [0,1)$ 是折扣因子。在部分可观测马尔可夫决策过程中运行的智能体，旨在通过在每个状态 s_t下选择最佳动作 a_t ，使其期望总折扣奖励 $E[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t)]$, 最大化；其中 t 是时间点； γ 会降低远期奖励的权重，并确保（无穷）级数收敛。通过在每个时间步考虑所有状态上的概率分布（称为信念 B），对状态的部分可观测性进行建模。POMDP求解器是能够以精确或近似方式为部分可观测马尔可夫决策过程计算策略 的算法，即一个将信念映射为动作的函数 π: $B \to A$ 。

3.2 部分可观测蒙特卡洛规划

部分可观测蒙特卡洛规划（POMCP） [11] 是一种用于部分可观测马尔可夫决策过程的在线基于蒙特卡洛的求解器。它使用蒙特卡洛树搜索（MCTS）在每个时间步选择一个近似最优的动作。给定由无权粒子滤波器表示的当前信念，通过从当前信念执行一定数量的模拟来生成蒙特卡洛树。这些模拟以高效的方式估计从当前信念出发所有动作的Q值。选择估计Q值最高的动作，并在真实环境中执行。POMCP的一大优势在于，它能够扩展到大型状态空间，因为它从不表示完整策略，而只生成与规划执行过程中实际观测到的信念状态相关的策略部分。此外，局部策略近似在线生成，所使用的是环境模拟器，即一个根据当前状态和动作，按照POMDP的转移模型和观测模型提供新状态及观测的函数。POMCP算法的一个重要步骤是信念更新。每当智能体在真实世界中移动时，POMCP都会更新信念，以反映通过观测环境获得的新信息。在智能体完成所有模拟并在 a 真实世界中执行单个动作后，它会接收到一个观测 o。新的信念被初始化为：从当前根节点出发，在树中选择动作 a对应的分支，再选择观测 o 对应的分支后，到达的节点的粒子滤波器。

3.3 探索、定位与接近

本文提出的方法是针对 AVS问题对 POMCP 的一种专门化。它基于以下定义的三个主要要素，POMCP 将这三个要素结合起来，在环境中搜索物体。我们假设 n是智能体在环境中可以采用的位姿数量， m 是环境中的物体数量， k是每个物体可被放置的位置数量。*(i)* 所提出框架的第一个要素是位姿图$\mathcal{G}$ ，其中节点表示智能体的 n 个可能位姿，边则只连接智能体通过单次动作能够到达的位姿。因此， $\mathcal{G}$ 用于约束在真实世界和仿真世界中无法执行的动作。*(ii)* 第二个要素是集合 $\mathcal{H} = \{1, \ldots, k\}$，其中包含每个物体在环境中可能占据的所有位置索引。 $\mathcal{H}$ 中的每个索引都对应搜索环境拓扑中的一个特定位置。*(iii)* 第三个要素是物体可观测性矩阵 $\mathbf{L} = (l_{i,j}) \in \{0,1\}^{n \times k}$，其中$l_{i,j} = 1$ 表示位置 j 可从位姿 i 处看到——即位置j 位于处于 i的智能体的视场内。矩阵 $\mathbf{L}$可通过可见性函数 f_L，根据环境的物理属性，从 $\mathcal{G}$ 和 $\mathcal{H}$ 确定性地推导得出；该函数会计算从智能体的每个位姿观察各个物体的可见性。该矩阵用于仿真中采用的观测模型，

该模型在模拟中使用。具体而言，如果从当前位姿观测到目标物体，观测模型就返回 1，否则返回 0。更ˆ形式化地说，当智能体位于位置 $i \in \mathcal{G}$ 且ˆ目标物体位于位置 $j \in \mathcal{H}$，并且满足 $l_{i,j} = 1$时，模型返回 1。请注意，每次具体模拟中的目标物体位置都是已知的，因为该位置由模拟开始时采样的粒子所定义。另一方面，真实世界中的观测基于物体检测器。在真实世界和仿真世界中，如果观测到物体，我们都会给予正奖励；否则，将给予负奖励（对应于执行移动所消耗的能量），并继续基于 POMCP 的搜索。

为避免智能体多次访问相同的位姿，智能体会维护一个内部记忆向量，记录当前运行过程中访问过的所有位姿。每当智能体再次访问某个位姿时，都会获得较高的负奖励。在真实世界中每执行一步后，智能体都会从物体检测器接收一个观测值1，表示已观测到目标物体；0 则表示未观测到。

智能体在每个时间步的信念，是对环境中所有候选物体位置的近似概率分布，用来表示POMCP隐藏状态。如果在固定步数内未观测到物体，该方法就会终止并报告搜索失败。

3.3.1 信念更新

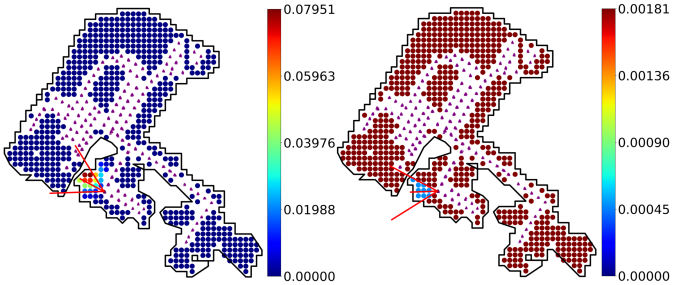
在我们最初的POMP [12], 表述中，信念采用标准POMCP策略进行更新。

这种方法的一个问题与状态空间的基数有关。在AVS中，状态空间同时描述智能体的位姿和目标的位置。由于物体可能位于任何位置，而模拟次数有限，因此某些状态可能不会在模拟阶段被考虑，只能在恢复活力阶段找回。如果这些状态未能找回，就会从信念中移除，并且无法再被恢复，即使它们是有效位置。这种方法的另一个问题是，在恢复活力阶段，新粒子是从先前的信念中采样得到的。这会形成一个反馈回路：在信念更新后存活下来的粒子，在恢复活力阶段被选中的概率更高。因此，在模拟次数有限的情况下，信念可能会局限于状态空间中的某个子空间。

在 POMP-BE-PD 中，我们更改了信念更新和重采样过程，以克服这些问题。首先从所有状态（即候选物体位置）的均匀分布中采样粒子，生成初始信念。辅助变量pp 存储当前尚未观测到的物体位置列表，其中 $pp = \{j \in \mathcal{H} \mid j$ 表示尚未观测到$\}$。该集合初始化为$pp = \mathcal{H}$。在每个时间步，智能体通过物体检测器获取当前视场内位置的观测，并通过移除其中不包含待搜索物体的已观测位置来更新 pp 。然后，从满足 pp 约束的位置均匀分布中采样新的信念——也就是说，待搜索物体位于属于$\mathcal{H} \setminus pp$的位置。如下文实验所示，这种信念更新方式有助于提升性能。

3.3.2 概率检测

我们为智能体配备了 [46], 中介绍的目标驱动实例检测器（TDID），这是一种旨在识别和分类物体类别中特定实例的架构。给定一幅图像，



（a）物体在视场内　　（b）物体在视场外

图 3. 创建向量 D 时考虑的两种情况。示例取自 Home 003 2。在情况（a）中，目标是确定物体的位置，并以多元正态分布的形式分配概率。在情况（b）中，我们为视场内的位置分配较低的概率，为视场外的位置分配较高的概率。注：为便于可视化，我们为颜色条设置了不同的刻度。

TDID 会返回一个坐标列表，表示相应的边界框（如果有）、一个分数 $s \in [0, 1]$ 以及相应的类别 c。在我们的工作中，我们仅考虑分数大于 0.9 的检测结果。此外，根据 真阳性 （ True Positive ）、假阳性 （ FalsePositive）和 假阴性 （FalseNegative）的比率，我们定义：

$$Precision = \frac{TP}{TP + FP} \qquad Recall = \frac{TP}{TP + FN}$$

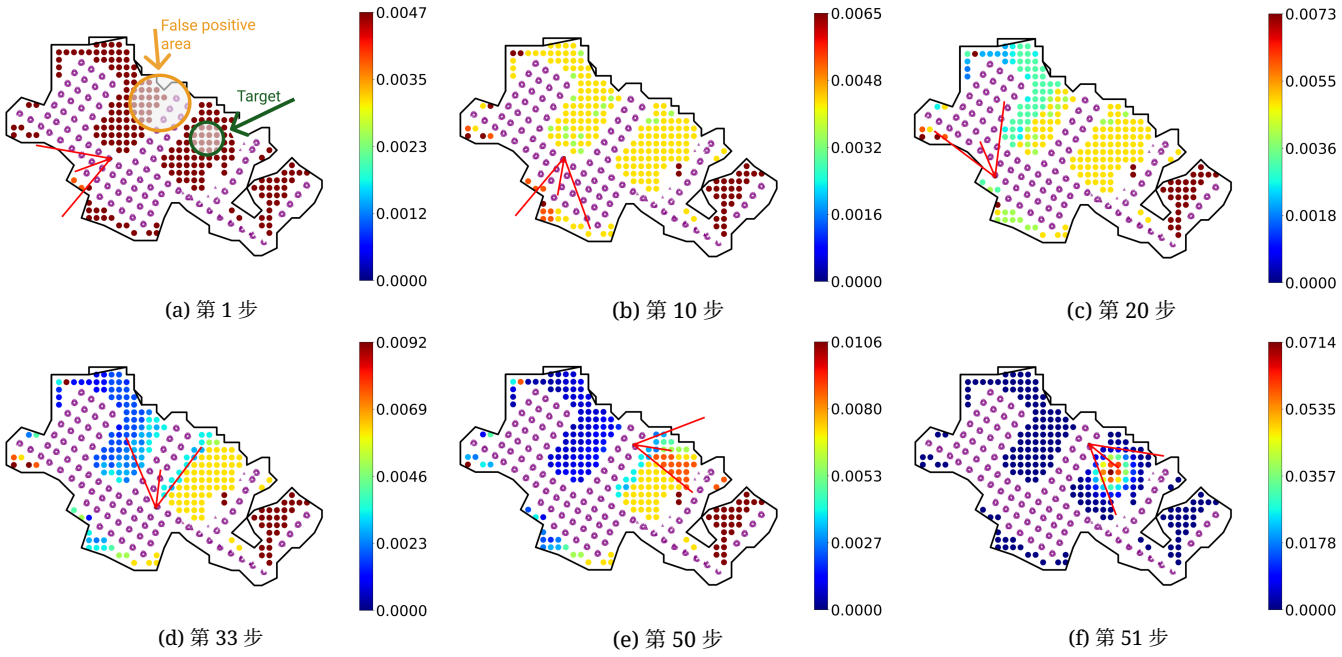
此外，我们将 F_1 分数定义为：

$$F_1\text{-score} = 2\,\frac{Precision \times Recall}{Precision + Recall},$$

其中， $F_1 \in [0, 1]$ 可解释为$Precision$ 与 $Recall$的调和平均值。

在 POMP [12]中，当物体检测器在视场内识别出目标物体时，规划器便会终止探索阶段。因此，物体检测器的假阳性会导致探索在物体并不存在的位置错误终止。在 POMP-BE-PD 中，我们旨在降低这一问题的影响。首先，我们定义一个向量 $\mathcal{D} = \{d_1, \ldots, d_k\}$，表示仅根据当前时间步的观测，目标物体位于位置 j 的概率。换言之，在真实环境中的每一步，我们都会重置所有 d_j，令其为 $j = 1, \ldots, k$，且仅考虑当前观测。如果物体检测器在当前视场内检测到物体，则将视场外位置 j 处的 $p_j = 0$设为相应值，并根据多元正态分布设置视场内位置的概率 p_j；该分布的均值位于检测器定位物体的位置 j （见图 3(a)）。如果物体检测器未在视场内检测到物体，则将视场内位置 j处的 $d_j = F_1$ 设为相应值，并将视场外位置 j处的 $d_j = 1 - F_1$ 设为相应值（见图 3(b)）。在这两种情况下，我们都会对 $\mathcal{D}$ 进行归一化，使得 $\sum_{j=1}^{k} d_j = 1$。请注意， F_1 具有类别特异性——即它反映了物体检测器对特定物体类别的检测性能。作为第二种辅助数据结构，我们定义一个向量$\mathcal{R} = \{p_1, \ldots, p_k\}$，表示综合全部观测历史后目标物体位于位置 j 的概率——即该向量表示一种全局概率，也利用了此前各步的信息。具体而言，数值 p_j 会在每个时间步t 按照以下规则更新：

$$p_j^t = \frac{p_j^{t-1} \cdot d_j^t}{\sum_{i=1}^{k} p_i^{t-1} \cdot d_i^t} \tag{1}$$



(a) 第 1 步 (b) 第 10 步 (c) 第 20 步

(d) 第 33 步 (e) 第 50 步 (f) 第 51 步

图 4. 使用所提出的 POMP-BE-PD 方法在 Home 016 1 中进行的实验。在步骤 (a) 中，我们在环境中初始化智能体；并标出目标位置和一个假阳性区域。从步骤 (b) 到 (c)，机器人探索上方区域；在步骤 (d) 中，我们展示了该方法对假阳性的鲁棒性；最后，在步骤 (e) 中，我们确定高概率位置，并在步骤 (f) 中找到目标。

对所有 $j \in \mathcal{H}$均成立。最后，我们定义一个阈值 $\tau = \frac{c}{n}$，其中 n 为候选物体位置的数量，$c \in \mathbb{N}$ 为一个常数，用于提高我们对概率检测的置信度。当当前位姿的视场内存在一个物体位置 j，且其概率 p_j超过该阈值时，我们便终止 POMCP 探索阶段。更正式地说，必须满足以下退出条件：

$$(p_j \geq \tau) \wedge (L_{i,j} = 1). \tag{2}$$

按照这一流程，如果物体不在当前视场内，我们就假设它一定在其他某个位置，因此提高相应位置的概率。相反，如果物体在当前视场内，我们就提高靠近物体三维位置的位置的概率，并降低其他位置的概率。此外，我们并非仅仅依赖物体检测器的输出，而是利用环境的旧状态和当前状态，随着时间推移不断积累信息。图 4 展示了一个过程，其中可以看到环境中各位置概率的变化。

3.3.3 概率对接

给定满足式2退出条件的物体位置 $j \in \mathcal{H}$，我们首先确定目标位姿——即最接近目标位置且朝向目标位置的智能体位姿 $i \in \mathcal{G}$。然后，我们使用 Dijkstra算法 [13] 计算当前位姿 $i \in \mathcal{G}$ 与估计目标位姿 $i \in \mathcal{G}$之间的最短路径。

智能体朝目标位姿移动时，不会使用物体检测器，因为我们足够确定目标物体位于位置 j。

该策略的性能优于 [12]中介绍的鲁棒视觉对接。一个关键区别在于，在鲁棒视觉对接中，物体检测器会沿路径运行，因此，当检测器性能不佳时，漏检和假阳性很容易使智能体偏离最终目标，并在接近阶段造成致命后果。

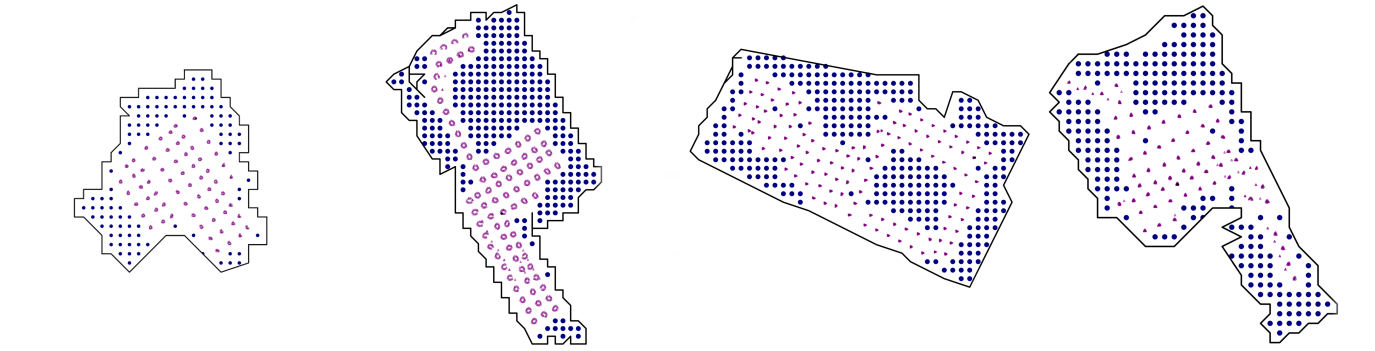
4 实验

我们在主动视觉数据集基准 [8], 上测试了我们的方法。这是一个用于主动视觉搜索的公开基准，包含超过30,000张RGBD图像，采集自15种不同的室内环境，涵盖33种不同的目标对象。与 [9], 一致，我们将数据集中的每个场景按照视觉搜索任务的难度分为简单、中等或困难：简单环境由单个小房间构成；中等难度的公寓包含一个大房间或一个额外的小房间——例如浴室或开放空间；困难公寓则包含多个大房间。实验中，我们使用了两个简单场景（Home 005 2和Home 015 1）、三个中等场景（Home 001 2、Home 016 1和Home 014 2）以及三个困难公寓（Home 003 2、Home 004 2和Home 013 1）。图5和图6展示了这些场景的一些示例。

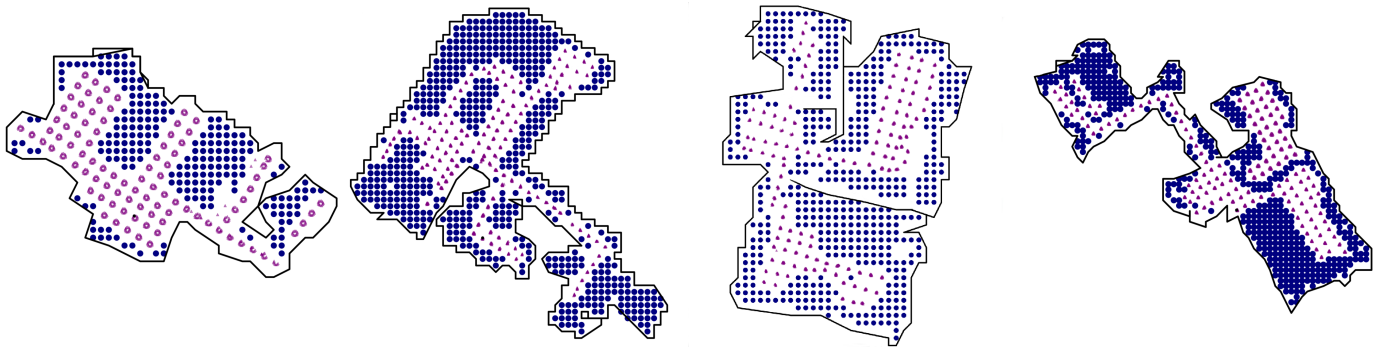
我们考虑三项指标：成功率 (SR) [49]，定义为智能体成功到达某个目标位姿（如 AVDB 中所提供）的次数占试验总次数的百分比（数值越大表示搜索效果越好）；平均路径长度 (APL)，定义为智能体在成功回合中访问的位姿总数除以成功回合数（数值越低表示绝对效率越高）；以及路径长度加权成功率 (SPL) [49]，定义如下：

$$SPL = \frac{1}{N} \sum_{i=1}^{N} S_i \frac{l_i}{\max(p_i, l_i)}, \tag{3}$$

其中，l_i 表示某一回合中目标与目标之间最短路径的长度，p_i 表示智能体在某一回合中走过的路径长度，S_i 表示该回合是否成功的二元指示变量



(a) 简单：主页_ 005_ 2 [97] (b) 简单：主页_ 015_ 1 [265] (c) 中等：主页_ 001_2 [289] (d) 中等：主页_ 014_2 [243]



(e) 中等：主页_ 016_1 [215] (f) 困难：主页_ 003_ 2 [560] (g) 困难：Home_ 004_2 [490] (h) 困难：Home_ 013_ 1 [490]

图 5. AVBD 中 3 个不同难度等级测试场景对应的二维平面图（未按比例绘制）（参见 [9]）。对于每个环境，我们报告其名称，并在括号中注明可能的物体位置数量。随着难度增加，可能的物体位置数量增多，空间布局也更加复杂。



(a) 简单：Home 005 2 (b) 简单：Home 015 1 (c) 中等：Home 001 2 (d) 中等：Home 014 2

(e) 中等：Home 016 1 (f) 困难：Home 003 2 (g) 困难：Home 004 2 (h) 困难：Home 013 1

图6. AVBD 中三个不同难度等级的测试场景（如 [9] 所示）。其中，我们将简单环境定义为由单个小房间组成的环境；中等难度的公寓包含一个大房间，或额外的小房间，例如浴室或开放空间；困难公寓则包含多个大房间。简单：Home 005 2 和 Home 015 1；中等：Home 001 2、Home 016 1 和 Home 014 2；困难：Home 003 2、Home 004 2 和Home 013 1。

表1 使用GT对象标注在AVDB的三个场景上进行的实验结果。所有方法均采用 [9]中定义的协议进行比较。星号(*) 表示评估是在不同的对象子集上进行的。

	简单（Home 005 2）			中等（Home 001 2）			困难（Home 003 2）			Avg.		
	SR ↑	平均路径长度 ↓	成功加权路径长度 ↑	SR ↑	平均路径长度 ↓	成功加权路径长度 ↑	SR ↑	平均路径长度 ↓	成功加权路径长度 ↑	SR ↑	平均路径长度 ↓	成功加权路径长度 ↑
随机游走	0.32	74.00	0.19	0.11	74.48	0.21	0.10	79.27	0.17	0.18	75.91	0.19
EAT [9]	0.77	12.20	-	0.73	16.20	-	0.58	22.10	-	0.69	16.80	-
DQN(*) [47]	*1.00*	*11.06*	-	*0.69*	*18.15*	-	-	-	-	-	-	-
DQN-TAM(*) [48]	*0.98*	*17.85*	-	*0.60*	*24.19*	-	-	-	-	-	-	-
POMP [12]	0.98	13.60	0.72	0.73	17.10	0.80	0.56	20.5	0.72	0.76	17.07	0.75
POMP-BE-PD	0.98	11.93	0.74	0.80	17.86	0.76	0.92	24.52	0.66	**0.90**	18.10	0.72

回合 i （数值越大表示绝对效率越高）。如果智能体在固定步数内（我们的实验中为 200 步）到达 AVDB 给出的目标位姿，则该回合视为成功。需要注意的是，在一些研究中，SPL 指标有时也称为 ASPPL [46]。

4.1 定量结果

我们将提出的方法 POMP-BE-PD 与随机游走基线进行比较——即允许智能体在每个时间步从所有可行操作中随机选择一个——并与四种最先进方法进行比较，分别是：EAT [9], 、DQN [47],、DQN-TAM [48],以及我们此前的工作 POMP [12]。其中，只有后者是无监督方法，而前三种方法需要使用训练数据来学习策略。由于已发表方法均未提供官方代码，我们只能按照 [9] 提出的协议与其结果进行比较。与基准论文 [8], 中定义的标准协议相比，该协议仅使用真实标注进行目标检测，并且只在有限数量的场景上提供结果。此外，该协议只使用一部分目标对象。DQN 和 DQN-TAM 只使用两个场景（一个简单场景和一个中等场景），因此平均值列不适用于公平比较。此外，DQN 和 DQN-TAM 在评估中使用了不同的对象子集。

从表1所报告的结果中可以清楚地看出，就SR而言，我们的方法POMP-BE-PD优于EAT，尽管APL略有增加；这是合理的，因为我们现在考虑的是更具挑战性的情况，正如我们将在第4.3节中深入探讨的那样。至于POMP-BE-PD与DQN之间的比较，我们注意到，DQN方法在简单场景中优于我们的方法，但在中等场景中，我们在SR方面超过了竞争方法，同时APL相当。

值得注意的是，为了取得这些结果，两个DQN方法都需要使用13个场景来训练出最佳策略，而我们的方法完全不需要训练。

使用 [46] 提供的物体检测器所得到的POMP和POMP-BE-PD结果报告于表3中。同样地，我们可以看到SR显著提升了35%，而SPL仅小幅下降0。03，这主要得益于我们提出的方法能够处理更复杂的情况。

4.2 消融研究

我们还提供了一项消融研究，以深入分析模型中不同项的贡献。下文将通过将所提出的方法与其若干部分版本进行比较来回答一些问题，这些版本仅采用新的信念更新（称为 POMP-BE），或仅采用概率检测（称为 POMP-PD）。

信念更新：新的信念更新是否能缩短回合长度？在导航困难场景时，新的信念更新有哪些优势？

在图 7 中，我们按照难度汇总了各回合，并根据回合到达目标所需的最短路径长度进行分组。在图 7(a) 中，我们汇总了简单场景（Home 005 2 和 Home 015 1）的结果；在图 7(b) 中，汇总了困难场景（Home 003 2、Home 004 2、Home 013 1）的结果；在图 7(c) 中，汇总了中等场景（Home 001 2、Home 014 2 和 Home 016 1）的结果；最后，在图 7(d) 中，我们汇总了 AVDB 中所有场景的结果。由这些报告可以得出，通过从信念更新中移除已经观测到的位置，我们可以优化探索阶段，从而缩短回合长度。

此外，表2通过隔离所有可能的误差来源来分析新的信念更新的影响，即，我们将物体检测器替换为真实标注，从而在规划和对接阶段均消除了假阳性和漏检的来源。在简单场景中，我们在保持相同成功率的同时，降低了平均路径长度并提高了效率（即成功加权路径长度）。我们推测，简单的布局（图 5(a)、5(b)）以及最短路径长度的集中分布（图 7(a)）使得信念更新无法带来益处。

然而，场景越困难，可选物体位置越多、空间布局越复杂，性能提升就越明显。从中等场景开始，我们将 SR 从 0.76 提高到 0.85，同时将 APL 从 29.64 降低到 25.50，总体效率也从 0.6 提高到 0.65。困难场景也呈现出类似的结果。

概率检测与对接：

概率检测能否减少误报？有没有办法改进对接，同时考虑规划过程中收集到的信息？

为回答第一个问题，我们分析回合中可能出现的不同失败类型。具体而言，我们考虑三种错误：定位，即在 POMCP 探索期间满足退出条件，但视场中实际上并不存在目标物体。对接，即在 POMCP 探索的最后一个位姿处，智能体正确检测到了目标，但未能到达 AVDB 定义的目标位姿；其他，例如，智能体未能在时限内检测到目标物体，或执行了路径中不允许的动作；

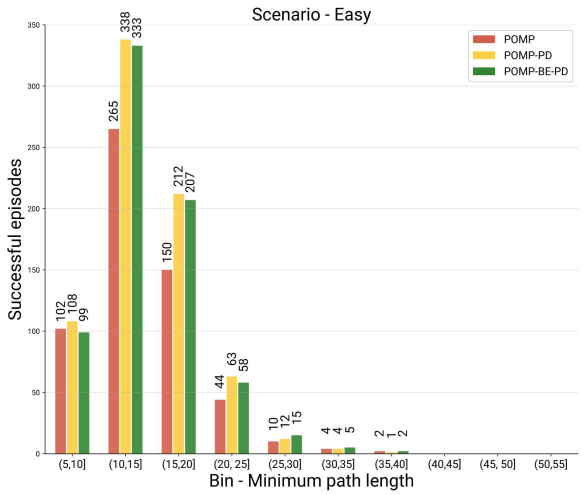
在图 8 中，我们给出了每个规划器在所有场景上的平均错误率。首先可以注意到，引入概率检测方法后，假阳性显著减少（降低 ∼32%），从而提高了我们方法的鲁棒性。此外，利用规划过程中收集的知识，可以为提高对接阶段的鲁棒性提供可靠机制。事实上，观察图 8 可发现，对接错误率降低了 ∼35.7%。

为了衡量概率检测的影响，我们在表 3 中进行了消融研究，将概率方法与信念更新分离开来，同时在规划和对接过程中使用检测器。在所有场景中，POMP-PD 的成功率均大幅高于我们之前的 POMP 方案（提高了 19% -25%）。然而，我们也注意到平均路径长度增加，总体效率（成功加权路径长度）有所下降。

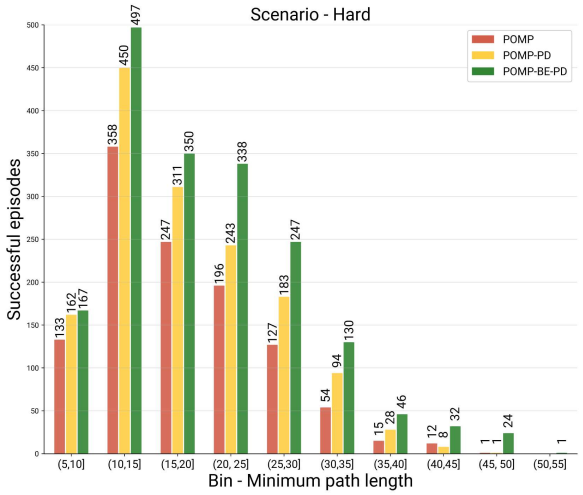
这并不令人意外：事实上，如果智能体需要对假阳性更加确信并具备更强的鲁棒性，就必须采取更多步数来提高目标位置的概率，并使其达到 $\geq$ 该阈值 τ。

4.3 定性结果

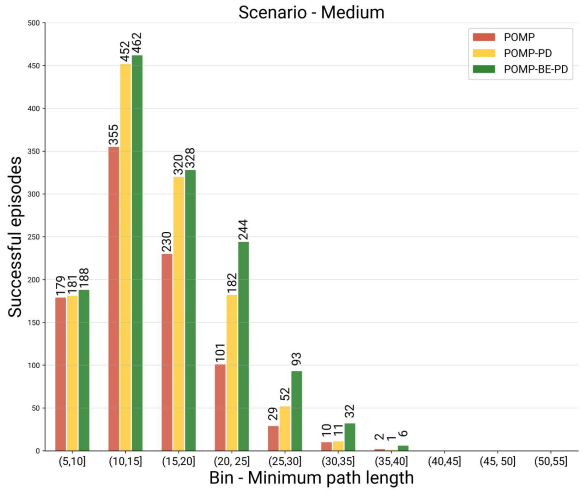
在图 4 中，我们将所提出方法的一个回合进行可视化，其中可以观察到位置概率分布的演变，以及对假阳性的鲁棒性。初始位姿定义于图 4(a) 中。从图 4(b) 到图 4(c)，智能体在环境的顶部区域进行探索，但未取得成功。在图 4(d) 中，机器人遇到了一个假阳性：式 1 中定义的更新规则与生成的概率图共同提供了一个鲁棒框架，使探索不会停止。



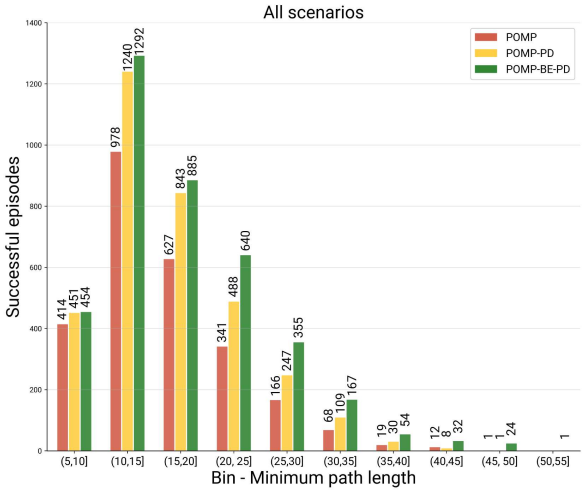
(a) 简单



(b) 困难



(c) 中等



(d) 全部

图 7。我们按照到达物体所需的最少步数对回合进行汇总，从而纳入回合的难度。在图 (a) 中为简单场景的结果：Home 005 2 和 Home 015 1；图 (b) 中为困难场景：Home 003 2、Home 004 2、Home 013 1；图 (c) 中为中等场景：Home 001 2、Home 014 2 和 Home 016 1；图 (d) 为所有场景的总和。规划和对接过程中均使用 [46]，提供的物体检测器所得的结果。

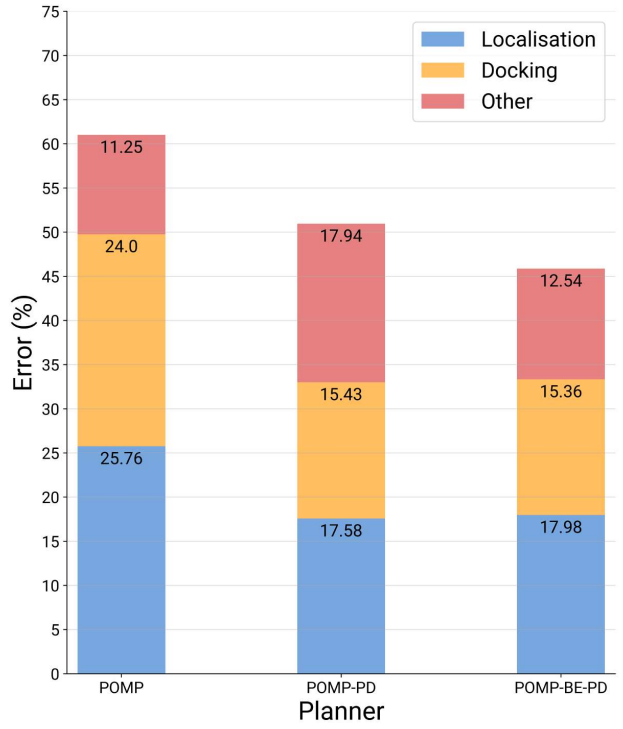
表2 不同版本改进型 POMP 的结果，其中 AVD 在每个难度级别包含更多场景。POMP-BE 是对信念更新进行修改后的 POMP。使用真实标注而非检测器，并通过 2^{10} 模拟所得的结果。

难度	场景	POMP[12]			POMP-BE		
		SR ↑	平均路径长度 ↓	成功加权路径长度 ↑	SR ↑	平均路径长度 ↓	成功加权路径长度 ↑
Easy	Home_005_2	0.94	12.96	0.78	0.93	12.26	0.78
	Home_015_1	0.75	23.66	0.61	0.73	17.04	0.72
	Avg.	**0.84**	18.31	0.70	0.83	**14.65**	**0.75**
中等	Home_001_2	0.80	18.20	0.72	0.81	19.95	0.69
	Home_014_2	0.76	41.07	0.51	0.90	19.99	0.66
	Home_016_1	0.71	29.64	0.56	0.83	36.55	0.60
	Avg.	0.76	29.64	0.60	**0.85**	**25.50**	**0.65**
Hard	Home_003_2	0.43	21.90	0.65	0.79	31.93	0.58
	Home_004_2	0.45	66.20	0.39	0.57	47.71	0.49
	Home_013_1	0.55	49.72	0.50	0.74	53.11	0.56
	Avg.	0.48	45.94	0.51	**0.70**	**44.25**	**0.54**
平均值		0.67	32.92	0.59	**0.79**	**29.82**	**0.63**

表 3 使用 [46] 提供的真实检测器，在 AVD 中每个难度级别包含更多场景时，POMP 及 POMP-BE-PD 的变体的结果。

难度	场景	POMP[12]			POMP-BE			POMP-PD			POMP-BE-PD（提出的方法）		
		SR ↑	平均路径长度 ↓	成功加权路径长度 ↑	SR ↑	平均路径长度 ↓	成功加权路径长度 ↑	SR ↑	平均路径长度 ↓	成功加权路径长度 ↑	SR ↑	平均路径长度 ↓	成功加权路径长度 ↑
Easy	Home_005_2	0.60	17.90	0.68	0.58	16.18	0.71	0.81	26.08	0.51	0.79	22.70	0.56
	Home_015_1	0.49	34.76	0.46	0.45	38.76	0.51	0.55	35.34	0.42	0.54	30.50	0.49
	Avg.	0.54	**26.33**	0.57	0.52	27.47	**0.61**	**0.68**	30.71	0.47	0.67	26.60	0.53
中等	Home_001_2	0.40	20.73	0.62	0.39	19.36	0.63	0.50	31.00	0.47	0.57	28.50	0.53
	Home_014_2	0.53	47.60	0.48	0.60	18.52	0.64	0.60	45.79	0.41	0.66	21.38	0.56
	Home_016_1	0.29	50.23	0.42	0.28	47.05	0.45	0.36	57.73	0.33	0.41	53.26	0.39
	Avg.	0.41	39.52	**0.51**	0.42	**28.31**	0.57	0.49	44.84	0.40	**0.55**	34.38	0.49
Hard	Home_003_2	0.19	26.60	0.53	0.33	30.53	0.55	0.39	62.36	0.34	0.48	42.86	0.42
	Home_004_2	0.42	69.84	0.36	0.55	47.31	0.47	0.44	70.26	0.31	0.54	61.93	0.36
	Home_013_1	0.25	61.41	0.46	0.31	77.09	0.44	0.26	62.80	0.37	0.34	54.38	0.46
	Avg.	0.29	52.62	**0.45**	0.40	**51.64**	0.49	0.36	65.14	0.34	**0.45**	53.06	0.41
平均值		0.40	41.13	**0.50**	0.44	**38.85**	0.55	0.49	48.92	0.40	**0.54**	39.44	0.47

图 8. POMP、POMP-PD 和 POMP-BE-PD 的错误百分比。错误分为三类：定位、对接和其他。在规划和对接过程中，我们使用了 [46], 提供的物体检测器。



事实上，在相同情况下， [12], 中定义的标准 POMP 会终止该回合。

此外，我们可以注意到，在智能体尚未探索的区域（环境的右侧部分），目标物体出现的概率正在上升：如果我们在其他地方都没有找到目标，那么它必定就在这一区域。最后，在图 4(e) 中，我们找到一个包含目标概率较高的区域；在图 4(f) 中，我们找到了被搜索的物体。

5 结论

本文介绍了我们提出的 POMP-BE-PD 方法，用于解决已知环境中的主动视觉搜索（AVS）问题。POMP-BE-PD 基于 POMCP 规划器，通过高效利用环境的二维平面图信息，在线学习策略；因此，我们的方法无需耗费大量时间或计算资源进行训练。为应对性能不完美的物体检测器——其假阳性率高且可能发生漏检，从而严重影响整体成功率——我们将确定性检测转变为概率性检测。在真实世界中每执行一步后，贝叶斯推断结合所有可能物体位置的概率分布，使我们能够将假阳性错误减少32%。此外，为了应对原始 POMCP 在 AVS 领域中受限的信念空间，我们引入了一种新的信念更新方法，在每个时间步都考虑所有尚未观测到的可能位置。我们按照 AVDB 基准对该方法进行了全面评估，并取得了最先进水平的结果。此外，通过多项消融研究，我们展示了该方法的优势。在所有环境中的平均表现上，与此前的 POMP 方案相比，我们将成功率显著提高了14%，同时将平均路径长度缩短了4%。

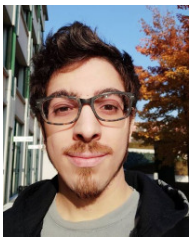
参考文献

[1] J. K. Tsotsos，“主动与被动视觉搜索的相对复杂度”，《国际计算机视觉期刊》，第7卷，第2期，第127–141页，1992年。[2] K. Sj¨o¨o、A. Aydemir 和 P. Jensfelt，“用于主动视觉搜索的拓扑空间关系”，Robotics and Autonomous Systems，第60卷，第9期，第1093–1107页，2012年。[3] F. Giuliari、G. Skenderi、M. Cristani、Y. Wang 和 A. Del Bue，“用于部分场景中物体定位的空间常识图”，载于《计算机视觉与模式识别》（CVPR），2022年。[4] J. Park、T. Yoon、J. Hong、Y. Yu、M. Pan 和 S. Choi，“零样本主动视觉搜索（zavis）：面向机器人助手的智能物体搜索”，arXiv预印本 arXiv:2209.08803，2022年。[5] A. Rasouli、P. Lanillos、G. Cheng 和 J. K. Tsotsos，“面向移动机器人的基于注意力的主动视觉搜索”，《自主机器人》，第44卷，第2期，第131–146页，2020年。[6] M. Leslie，“机器人迎战DARPA地下挑战赛”，第2–4页，2022年。[7] P. Pouya 和 A. Madni，“主动搜索以定位火星古代水源迹象：一种在线概率方法”，载于 ASCEND，2021年，第4024页。[8] P. Ammirato、P. Poirson、E. Park、J. Koˇseck´a 和 A. C. Berg，“用于开发和基准测试主动视觉的数据集”，载于 IEEE 国际机器人与自动化会议（ICRA），2017年。

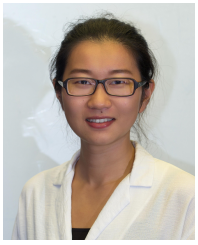
[9] J. F. Schmid、M. Lauri 和 S. Frintrop，“探索、接近与终止：基于深度强化学习评估主动视觉目标搜索中的子任务”，载于 IEEE/RSJ智能机器人与系统国际会议（IROS），2019年。[10] X. Ye、Z. Lin、J. Lee、J. Zhang、S. Zheng 和 Y. Yang，“GAPLE：面向室内环境机器人目标搜索的通用接近策略学习”，IEEE机器人与自动化快报，第4卷，第4期，第4003–4010页，2019年。[11] D. Silver 和 J. Veness，“大型POMDP中的蒙特卡洛规划”，载于《神经信息处理系统进展》（NeurIPS），2010年。[12] Y. Wang、F. Giuliari、R. Berra、A. Castellini、A. Del Bue、A. Farinelli、M. Cristani 和 F. Setti，“POMP：基于POMCP的室内环境主动视觉搜索在线运动规划”，载于《英国机器视觉会议》（BMVC）论文集，2020年。[13] E. W. Dijkstra，“关于图论中两个问题的一则注记”，Numerische mathematik，第1卷，第1期，第269–271页，1959年。[14] D. Batra、A. Gokaslan、A. Kembhavi、O. Maksymets、R. Mottaghi、M. Savva、A. Toshev 和 E. Wijmans，“重新审视ObjectNav：对导航至物体的具身智能体进行评估”，arXiv预印本 arXiv:2006.13171，2020年。[15] L. Kunze、K. K. Doreswamy 和 N. Hawes，“利用定性空间关系进行间接目标搜索”，载于 IEEE国际机器人与自动化会议（ICRA），2014年。[16] X. Han、H. Liu、F. Sun 和 X. Zhang，“利用深度Q网络进行多步动作预测的主动目标检测”，IEEE工业信息学汇刊，第15卷，第6期，第3723–3731页，2019年。[17] P. Mirowski、R. Pascanu、F. Viola、H. Soyer、A. Ballard、A. Banino、M. Denil、R. Goroshin、L. Sifre、K. Kavukcuoglu 等，“学习在复杂环境中导航”，载于《学习表征国际会议》（ICLR），2017年。[18] A. Mousavian、A. Toshev、M. Fišer、J. Košecká、A. Wahid 和 J. David-son，“面向语义目标驱动导航的视觉表征”，载于《国际机器人与自动化会议》（ICRA）。IEEE，2019年，第8846–8852页。[19] K. Fang、A. Toshev、L. Fei-Fei 和 S. Savarese，“面向长时域任务中具身智能体的场景记忆变换器”，载于 IEEE/CVF计算机视觉与模式识别会议论文集，2019年，第538–547页。[20] K. Yadav、S. K. Ramakrishnan、J. Turner、A. Gokaslan、O. Maksymets、R. Jain、R. Ramrakhya、A. X. Chang、A. Clegg、M. Savva、E. Un-dersander、D. S. Chaplot 和 D. Batra，“Habitat挑战赛2022”，https://aihabitat.org/challenge/2022/，2022年。[21] K. Yadav、R. Ramrakhya、S. K. Ramakrishnan、T. Gervet、J. Turner、A. Gokaslan、N. Maestre、A. X. Chang、D. Batra、M. Savva 等，“Habitat-Matterport三维语义数据集”，arXiv预印本 arXiv:2210.05633，2022年。[22] D. S. Chaplot、D. Gandhi、A. Gupta 和 R. Salakhutdinov，“使用目标导向语义探索进行目标物体导航”，载于《神经信息处理系统进展》论文集，2020年。[23] S. K. Ramakrishnan、D. S. Chaplot、Z. Al-Halah、J. Malik 和 K. Grauman，“PONI：通过无交互学习实现目标物体导航的势函数”，载于 IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集。IEEE，2022年。[24] D. A. Sasi Kiran、K. Anand、C. Kharyal、G. Kumar、N. Gireesh、S. Baner-jee、R. D. Roychoudhury、M. Sridharan、B. Bhowmick 和 M. Krishna，“用于目标物体导航的空间关系图与图卷积网络”，载于 IEEE自动化科学与工程国际会议（CASE），2022年，第1392–1398页。[25] L. Kaelbling、M. Littman 和 A. Cassandra，“部分可观测随机领域中的规划与行动”，《人工智能》，第101卷，第1–2期，第99–134页，1998年。[26] C. Papadimitriou 和 J. Tsitsiklis，“马尔可夫决策过程的复杂度”，Math. Oper. Res.，第12卷，第3期，第441–450页，1987年。[27] S. Thrun，“蒙特卡洛POMDP”，载于《神经信息处理系统进展》（NeurIPS），2000年。[28] R. Coulom，“蒙特卡洛树搜索中的高效选择性与备份算子”，载于国际计算机与博弈会议（CG），2006年。[29] C. Browne、E. Powley、D. Whitehouse、S. Lucas、P. Cowling、P. Rohlf-shagen、S. Tavener、D. Perez、S. Samothrakis 和 S. Colton，“蒙特卡洛树搜索方法综述”，IEEE Trans. Comp. Intell. AI Games，第4卷，第1期，第1–43页，2012年。[30] L. Kocsis 和 C. Szepesv´ari，“基于赌博机的蒙特卡洛规划”，载于《欧洲机器学习会议》（ECML），2006年。[31] S. Katt、F. A. Oliehoek 和 C. Amato，“利用蒙特卡洛树搜索在POMDP中学习”，载于第34届国际机器学习会议（ICML）论文集，系列：机器学习论文集，D. Precup 和 Y. W. Teh 编，卷 70。PMLR，2017 年，第1819–1827 页。[32] C. Amato 和 F. A. Oliehoek，《多智能体部分可观测马尔可夫决策过程的可扩展规划与学习》，载于 AAAI人工智能大会，2015 年。[33] J. Lee、G.-H. Kim、P. Poupart 和 K.-E. Kim，《受约束的部分可观测马尔可夫决策过程的蒙特卡洛树搜索》，载于神经信息处理系统进展（NeurIPS），2018 年。[34] A. Castellini、G. Chalkiadakis 和 A. Farinelli，《状态变量约束对部分可观测蒙特卡洛规划的影响》，载于国际人工智能联合会议（IJCAI），2019 年。[35] G. Mazzi、A. Castellini 和 A. Farinelli，《在部分可观测蒙特卡洛规划中识别意外决策：一种基于规则的方法》，载于第 20 届国际自主智能体与多智能体系统会议（AAMAS），F. Dignum、A. Lomuscio、U. Endriss 和 A. Now´e 编。ACM，2021 年，第 889–897 页。[36] ——，《部分可观测蒙特卡洛规划的基于规则的屏蔽》，载于第三十一届国际自动规划与调度会议论文集（ICAPS），S. Biundo、M. Do、R. Goldman、M. Katz、Q. Yang 和 H. H. Zhuo 编。AAAI Press，2021 年，第 243–251 页。[37] ——，《为 POMCP 屏蔽主动生成逻辑规则》，载于 AAMAS ’22：第 21 届国际自主智能体与多智能体系统会议。国际自主智能体与多智能体系统基金会（IFAAMAS），2022 年，第 1696–1698 页。[38] D. Meli、G. Mazzi、A. Castellini 和 A. Farinelli，《学习逻辑规范以实现 POMCP 软策略引导》，载于第 22 届国际自主智能体与多智能体系统会议（AAMAS），F. Dignum、A. Lomuscio、U. Endriss 和 A. Now´e 编，2023 年。[39] M. Lauri 和 R. Ritala，《基于前向仿真的机器人探索规划》，Robotics and Autonomous Systems，第 83 卷，第 15–31 页，2016 年。[40] A. Goldhoorn、A. Garrell、R. Alqu´ezar 和 A. Sanfeliu，《人形服务机器人通过连续实时 POMCP 寻找并跟随人员》，载于第 14 届 IEEE-RAS国际人形机器人会议（Humanoids）。IEEE，2014 年，第 741–747 页。[41] A. Castellini、E. Marchesini 和 A. Farinelli，《具有状态变量约束的部分可观测蒙特卡洛规划：用于移动机器人导航》，Engineering Applications of Artificial Intelligence，第 104 卷，第 104382 页，2021 年。[42] M. Zuccotto、M. Piccinelli、A. Castellini、E. Marchesini 和 A. Farinelli，《在 POMCP 中学习状态变量关系：一个面向移动机器人的框架》，机器人与人工智能前沿，第 9 卷，2022 年。[43] D. Silver、A. Huang、C. J. Maddison、A. Guez、L. Sifre、G. van den Driessche、J. Schrittwieser、I. Antonoglou、V. Panneershelvam、M. Lanc-tot、S. Dieleman、D. Grewe、J. Nham、N. Kalchbrenner、I. Sutskever、T. P. Lillicrap、M. Leach、K. Kavukcuoglu、T. Graepel 和 D. Hassabis，《借助深度神经网络和树搜索精通围棋》，《自然》第 529 卷第 7587 期，第 484–489 页，2016 年。[44] D. Silver、T. Hubert、J. Schrittwieser、I. Antonoglou、M. Lai、A. Guez、M. Lanctot、L. Sifre、D. Kumaran、T. Graepel、T. Lillicrap、K. Simonyan 和 D. Hassabis，《一种通过自我对弈精通国际象棋、将棋和围棋的通用强化学习算法》，《科学》第 362 卷第 6419 期，第 1140–1144 页，2018 年。[45] F. Giuliari、A. Castellini、R. Berra、A. D. Bue、A. Farinelli、M. Cristani、F. Setti 和 Y. Wang，《POMP++：基于 POMCP 的未知室内环境主动视觉搜索》，载于 IEEE/RSJ智能机器人与系统国际会议（IROS）。IEEE Press，2021 年，第1523–1530 页。[46] P. Ammirato、C.-Y. Fu、M. Shvets、J. Kosecka 和 A. C. Berg，《目标驱动的实例检测》，arXiv 预印本 arXiv:1803.04610，2018 年。[47] S. Liu 和 G. Tian，《一种用于机器人主动目标检测的深度Q学习网络高效训练策略》，载于自动化、控制与智能系统中的CYBER技术国际会议（CYBER），2022 年，第 202–206 页。[48] S. Liu、G. Tian、Y. Cui 和 X. Shao，《一种采用新型训练算法的服务机器人深度Q学习网络主动目标检测模型》，信息技术与电子工程前沿，第 23 卷第 11 期，第 1673–1683 页，2022 年 11 月。[49] P. Anderson、A. X. Chang、D. S. Chaplot、A. Dosovitskiy、S. Gupta、V. Koltun、J. Kosecka、J. Malik、R. Mottaghi、M. Savva 和 A. R. Zamir，《具身导航智能体的评估》，arXiv 预印本 arXiv:1807.06757，2018 年。



FrancescoTaioli 是都灵理工大学人工智能国家博士项目的博士生，该项目与维罗纳大学合作开展，由 Marco Cristani 教授和 Alessandro Farinelli 教授指导。他于 2022 年以优异成绩毕业于维罗纳大学，获得计算机工程理学硕士学位。他的主要研究兴趣是计算机视觉和深度学习，重点在于提升智能体的自主性。



Francesco Giuliari （IEEE 学生会员）是热那亚大学的博士生。目前，他在 Alessio Del Bue 博士的指导下，供职于意大利理工学院。他于 2018 年获得维罗纳大学计算机科学理学硕士学位。他的主要研究兴趣包括计算机视觉、场景理解和基于视觉的智能体导航。



Yiming Wang 是布鲁诺·凯斯勒基金会（FBK）深度视觉学习（DVL）部门的研究员。她的研究主要聚焦于基于视觉的场景理解，旨在促进有益于社会的自动化。Yiming 于 2018 年获得伦敦玛丽女王大学（英国）电气工程博士学位，研究方向为基于视觉的多智能体导航。自 2018 年起，她一直在意大利理工学院模式分析与计算机视觉（PAVIS）研究方向担任博士后研究员。

技术研究所（IIT），研究方向涉及主动三维视觉。她曾组织过几次相关主题的研讨会，并积极担任计算机视觉和机器人学领域顶级会议与期刊的审稿人。



Riccardo Berra 是维罗纳大学计算机科学专业毕业生，目前就读于机器人与智能工业计算机工程硕士学位项目。Riccardo 曾多次参与机器人学和计算机视觉领域的论文发表，师从 Marco Cristani 教授和 Francesco Setti 教授。



Alberto Castellini 自 2018 年起担任维罗纳大学计算机科学系助理教授。在此之前，他曾任职于波茨坦大学/马克斯·普朗克研究所和维罗纳大学，担任研究员。他的研究兴趣包括不确定性下的概率规划、强化学习、统计学习，以及面向智能系统（例如机器人/信息物理系统）的数据分析。他曾在人工智能领域的期刊和会议上发表论文（IEEE Int. Syst.、Eng. App. of Artificial Intelli-

智能、机器人自主系统、IJCAI、AAMAS、ICAPS、IROS）。



Alessio Del Bue （IEEE会员）是意大利热那亚意大利理工学院（IIT）终身制高级研究员，负责模式分析与计算机视觉（PAVIS）研究方向。他是100多篇经过同行评审的期刊论文和国际会议论文的合著者，研究主题涉及计算机视觉和机器学习。目前，他的研究兴趣包括利用多模态输入（图像、深度信息和音频）进行三维场景理解，以支持辅助人工智能的开发

智能系统。他是主要计算机视觉会议（CVPR、ICCV、ECCV 和 BMVC）技术委员会的成员。他担任《模式识别》和《计算机视觉与图像理解》期刊的副编辑。他是 ELLIS 的成员。



Alessandro Farinelli 自2019年12月起担任维罗纳大学计算机科学系的正教授。他的研究兴趣集中于开发应用于机器人学和信息物理系统的人工智能系统新方法。具体而言，他专注于信息物理系统中的多智能体协调、去中心化优化、强化学习和数据分析。Alessandro Farinelli 曾担任多个国家级和国际级人工智能领域研究项目的首席研究员

人工智能广泛领域中的项目。



MarcoCristani 是维罗纳大学计算机科学系的正教授（Professore Ordinario）、国家研究委员会（CNR）副成员，以及意大利理工学院（IIT）的外部合作者。他的主要研究兴趣是统计模式识别和计算机视觉，尤其关注深度学习和生成式建模，并将其应用于社会信号处理和时尚建模。在这些领域，他已发表超过200篇论文。他曾组织 11

国际研讨会，并共同创立了衍生公司 Humatics，专注于时尚电子商务。他曾担任多个国家级和国际项目的首席研究员，包括 PRIN 和 H2020 项目。他是 IAPR 会士，也是 IEEE 成员。



Francesco Setti 是维罗纳大学计算机科学系助理教授，也是认知科学与技术研究所（ISTC-CNR）的副研究员。他的研究兴趣包括应用于机器人学和制造业的计算机视觉、机器学习和人工智能。他是 50 多篇发表于国际同行评审期刊和会议论文的合著者。他是 IEEE、IAPR 和 BMVA 的成员，并担任所有主要机器

学习会议和期刊（包括 CVPR、ICCV、ECCV 和 TPAMI）的审稿人。