arXiv:2302.01520v1 [cs.RO] 3 Feb 2023

多条细线　　实现物体导航元能力解耦　　n

Ronghao Dang [1] Lu Chen [1] Liuyi Wang [1] Zongtao He [1] Chengju Liu [1] Qijun Chen [1]

摘要

我们提出了元能力解耦（MAD）范式，将各种物体导航方法整合到一个架构系统中，使它们能够相互增强、共同演进。基于元能力解耦（MAD）范式，我们设计了多重思维（MT）模型，利用不同的思维抽象出各种元能力。我们的方法从输入、编码和奖励三个方面对元能力进行解耦，同时采用多重思维协作（MTC）模块，促进不同思维之间的相互协作。MAD 为物体导航引入了一套全新的定性与定量可解释性系统。通过在 AI2-Thor 和 RoboTHOR 上进行大量实验，我们证明了，与最先进水平（SOTA）方法相比，我们的方法在典型物体导航任务和零样本物体导航任务上均表现更优。

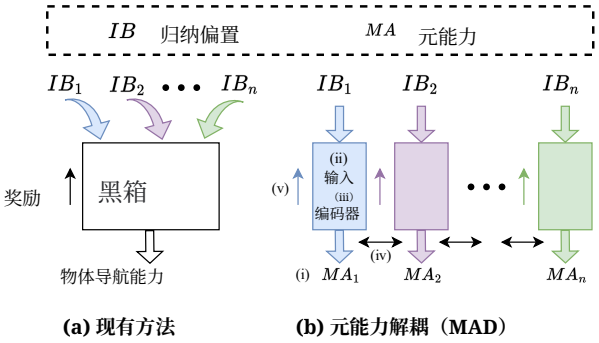


图*1.* (a) 现有方法通过在黑箱模型中引入各种归纳偏置，直接提升整体物体导航能力。(b) 我们提出的元能力解耦（MAD）范式将整体物体导航能力分解为多种元能力，并为每种元能力设计特定的输入、思维编码器和奖励。

1. 引言

物体导航（Zheng 等人，2022年；Moghaddam 等人，2022年；Du 等，2021年；Wang 等人，2022年）是一项具有挑战性的任务，要求智能体通过第一人称视觉观测，在未知环境中寻找目标物体。由于任务的复杂性，研究人员通过引入不同的归纳偏置（图1（a））开发了大量技术，以推动该领域的发展。然而，遗憾的是，物体导航领域尚未形成类似于计算机视觉（Cao 和 Wu，2022年；d'Ascoli 等人，2021年）或自然语言处理（Levine 等人，2022年；Kharitonov 和 Chaabouni，2021年）领域中那样统一的归纳偏置范式。受这一缺陷启发，我们通过对当前主流方法进行归纳与升华，提出了元能力解耦（MAD）范式，希望统一并连接各种物体导航方法。

本文涉及两个重要的新概念：元能力和思维。**元能力**指完成复杂任务所需的每一种基本能力。例如，解决一道数学问题需要整合多种元能力，如文本理解、逻辑推理和概念抽象。没有这些元能力，仅凭直觉不足以完成复杂任务。**思维**指针对某种能力进行信息抽象。通常，这种抽象通过神经网络进行端到端建模。

根据元能力和思维的定义，我们总结了当前主流的物体导航方法，并指出其局限性。如图 2 所示，物体导航方法分为四类：关联方法（Dang等人，2022a；Zhang等，2021年）、记忆方法（Chen 等，2022；Fukushima 等，2022）、死锁专用方法（Du 等，2020；Lin 等，2021年）以及 SLAM 方法（Ravichandran 等，2022；Liang 等，2021年）。这四类方法引入的不同归纳偏置，决定了它们侧重哪些元能力，又忽视了哪些元能力。因此，现有方法都试图用带有偏向性的思维来抽象物体导航所需的终极能力（图1 (a)）。然而，由于奖励信号稀疏且含糊不清，带有偏向性的思维难以隐式地解耦出物体导航中至关重要的完整元能力。

为解决上述问题，我们提出了元能力解耦（MAD）范式（图1（b）），通过五个阶段解决具身人工智能任务：（i）基于先验知识选择元能力；（ii）根据每种元能力的特征确定其对应思维的输入特征

；（iii）为每种思维设计合适的编码网络；（iv）根据任务特征设计不同思维之间的协作模块；（v）为每种元能力设计奖励与惩罚。在此过程中，元能力在三个方面实现解耦：输入、编码和奖励信号。本文主要研究物体导航任务，但我们相信，MAD范式可以扩展到其他类似的具身人工智能任务。

受 MAD 范式启发，我们为物体导航任务设计了一种多重思维（MT）模型。首先，我们选取五种元能力（详见第 3 节）：直觉、搜索、导航、探索和障碍物。随后，针对这五种元能力，我们分别采用整体图像特征、目标检测特征、目标导向记忆、历史状态记忆和障碍物位置记忆作为相应思维的输入。每种思维都使用一个具有必要归纳偏置的简单编码网络。此外，我们设计了一个多重思维协作（MTC）模块，以促进不同元能力之间的协作。最后，我们设计了元能力奖励，以引导每种思维抽象地理解相应的元能力。

在 AI2-Thor（Kolve 等人，2017）和 RoboTHOR（Deitke 等人，2020）数据集上进行的大量实验表明，我们的 MAD 范式不仅在典型的物体导航任务上优于最先进方法，在零样本物体导航任务上也表现出色。此外，基于 MAD 对 MT 模型进行的可解释性分析表明，我们的方法显著提升了物体导航任务的可解释性和灵活性。我们的贡献总结如下：

- 我们提出一种通用的元能力解耦（MAD）范式，以归纳并统一当前各种物体导航方法。
- 遵循MAD范式，我们为物体导航任务设计了一种多重思维（MT）模型，该模型在典型物体导航任务和零样本物体导航任务中的表现均优于现有模型。
- 我们的元能力可解释性框架为未来研究者提供了一种新颖的分析模式。

2. 相关工作

2.1. 物体导航

目标特定典型物体导航任务要求智能体在未知环境中导航至已知目标物体。近期的一些方法致力于改进网络或引入先验知识，以解决物体导航任务中的各种问题。我们将这些方法分为四类（图 2）：（i）关联方法（Wu 等，2019年；Gao 等，2021年；Yang 等，2019年），利用物体关联或区域关联，使智能体能够构建场景的关系图模型；（ii）记忆方法（Zhu 等，2021年；Kwon 等，2021年），依靠长期显式记忆，更全面地考虑历史信息并据此做出决策；（iii）SLAM 方法（Chaplot 等，2020b；Ramakrishnan 等，2022），实时构建以智能体为中心的语义地图；（iv）死锁专用方法（Wortsman 等，2019；Du 等，2020），采用特殊机制帮助智能体摆脱局部死锁状态。由于缺乏元能力解耦视角，每类方法都只强调部分元能力，因此缺乏解决复杂任务的综合能力。

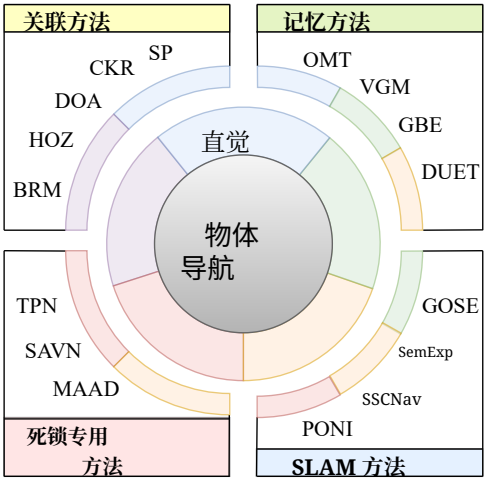


图 2。**各类物体导航方法概述**。我们将主流的物体导航方法分为四类，这些方法通过改进神经网络来增强特定的元能力。

目标无关零样本物体导航任务随着多模态对比学习的发展，正受到越来越多的关注（Radford 等人，2021年）。这项任务要求训练环境在测试时屏蔽目标物体。Al-Halah 等人（2022）将各种模态映射到图像目标嵌入空间，从而适配图像目标导航智能体。Zhao 等人（2022）将物体与目标的关系表示为余弦相似度，以缓解过拟合。这些零样本物体导航方法本质上是对典型物体导航架构的扩展：将离散类别输入映射到连续语义空间。

2.2. 物体导航中的解耦思路

解耦是人工智能领域的一个常见概念（Duan 等人，2021年），在物体导航任务中也十分常见。SemExp（Ravichandran 等人，2022）将连续决策过程解耦为两个离散步骤："该去哪里寻找物体？"以及"如何导航到 (x, y)？"。AVSW（Park 等人，2022）将环境探索与前往目标的导航解耦。ANS（Chaplot 等人，2020a）利用实时构建的语义地图，将环境建模与端到端网络解耦。上述解耦方法都将端到端物体导航模型拆分为若干独立训练的模型，因此失去了灵活性和简洁性方面的优势。我们的 MAD 范式在保持端到端学习的同时，提高了模型的可解释性和泛化能力。

3. 元能力解耦（MAD）

物体导航是现实世界中一项复杂的远距离决策任务。要高效、准确地导航至目标物体，需要多种元能力的协助。基于现有的物体导航方法和人类经验，我们从物体导航任务中解耦出以下五种元能力：

1) 直觉能力：直接根据原始图像特征得出动作决策的能力。**2) 搜索能力：**通过知识关联寻找目标物体的能力。**3) 导航能力：**根据记忆中的目标方位信息导航至目标位置的能力。**4) 探索能力：**高效、全面地获取场景信息的能力。**5) 避障能力：**避免与障碍物发生碰撞的能力。

这五种元能力在当前方法中均有所体现（图2），然而，单一方法往往只侧重于某些元能力。通过明确识别并解耦这些能力，研究人员可以更轻松地结合多种方法的优势。

4. 多重思维（MT）网络

多重思维（MT）网络是在 MAD范式的指导下设计的。如图 3所示，我们为每种元能力设计了合适的输入（第 4.2节）、编码网络（第 4.3节）和奖励（第 4.6节）。此外，我们还设计了一个多重思维协作（MTC）模块（第 4.4节），用于在不同类型的思维之间交互信息。

4.1. 任务定义

智能体被初始化为随机状态 $s = \{x, y, \theta, \beta\}$和随机目标物体 p。在每个时间戳 t，智能体根据单视角 RGB图像 o_t 和目标 p学习导航策略 $\pi(a_t|o_t, p)$，其中 $a_t \in A = \{MoveAhead;\ RotateLeft;\ RotateRight;\ LookDown;\ LookUp;\ Done\}$ 并且，当智能体认为自己已导航至目标位置时，$Done$ 为其输出。如果在输出 Du 等，2020）时，智能体位于目标一定阈值范围内（即 1.5 米），且能正确检测到目标，则该导航回合被视为成功。$Done$ 为其输出时，导航回合被视为成功。

零样本物体导航任务将目标物体划分为训练集 $P_{train} = \{p_1, p_2, \cdots, p_n\}$ 和测试集 $P_{test} = \{p_{n+1}, p_{n+2}, \cdots, p_{n+m}\}$。测试集中的物体仅在测试过程中可用。

4.2. 思考输入

每种思维的输入都是相应元能力最重要的归纳偏置。在满足元能力要求的前提下，输入特征应尽可能简洁。如图 3 中的思考框所示，我们根据元能力的特点，从智能体可用的信息中选取了五种专门化输入。

1) **直觉思维输入** $IT_i \in \mathbb{R}^{7\times7\times512}$ 通过固定权重的 ResNet18（He等人，2016）从第一人称视角图像中提取。**2) 搜索思维输入** $ST_i \in \mathbb{R}^{N\times262}$ 是使用 DETR（Carion等人，2020）从图像中提取的物体视觉与位置特征。**3) 导航思维输入** $NT_i \in \mathbb{R}^{D_n\times9}$ 遵循（Dang等人，2022b）提出的目标导向记忆图（TOMG）。导航思维仅关注目标相关信息；因此，TOMG由目标边界框以及智能体在已访问目标可见节点上的坐标组成。**4) 探索思考输入** $ET_i \in \mathbb{R}^{D_e\times4}$ 是智能体的历史位置和摄像机角度。**5) 障碍物思维输入** $OT_i \in \mathbb{R}^{D_o\times2}$ 是已知不可到达节点的位置。当智能体尝试到达某个节点但失败时，会将该节点记录为不可到达。

N 是物体的数量。D_n、D_e、D_o 分别表示已访问的目标可见节点、已访问节点和已知不可达节点的数量。

4.3. 思考嵌入

思考嵌入将思维输入抽象为用于决策的语义空间。以往的工作（Gao 等，

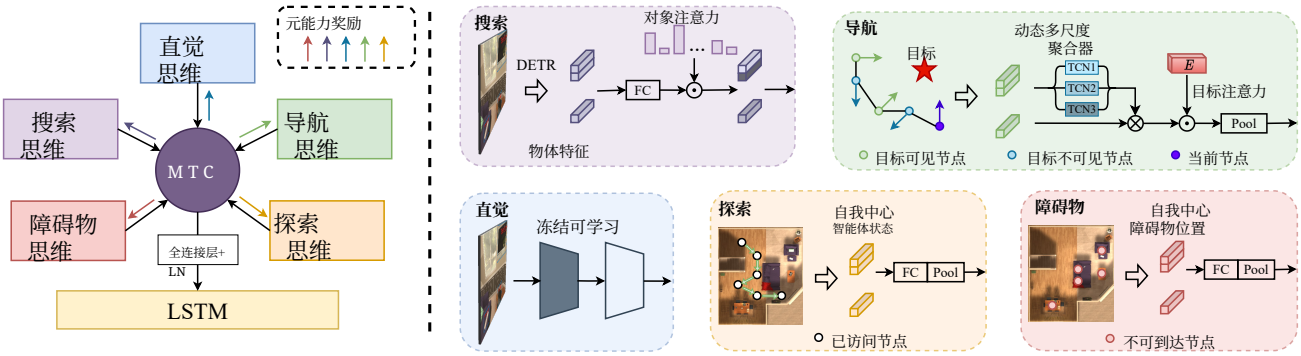


图*3.* **模型概览**。MTC：多重思维协作。我们的多重思维（MT）模型主要由五个思维模块和一个MTC模块组成，位于LSTM网络之前。右侧的彩色方框与左侧思考嵌入的细节相对应。(1) 直觉思维以图像ResNet18编码作为输入。(2) 搜索思维以DETR提取的物体特征作为输入。(3) 导航思维以目标可见节点处的目标方位信息作为输入。(4) 探索思维以智能体历史状态作为输入。(5) 障碍物思维以已知不可到达节点的位置作为输入。

2021年；Chen 等，2022）引入了各种先验知识到思维编码网络中，以引导模型的注意力。然而，我们的 MT 方法仅根据每种元能力的特点使用少量编码技术，凸显了 MAD 范式本身的优势。

直觉思维 通过一个简单的可学习逐点卷积，直接对输入 ResNet 特征进行编码：

$$IT_o = \delta(Conv(IT_i)) \tag{1}$$

其中，$Conv$ 表示逐点卷积，δ 表示 ReLU 非线性函数（Nair 和 Hinton，2010）。

搜索思维 搜索思维旨在当目标不在视野中时，使智能体能够以最少的步骤快速捕捉目标。为了具备目标关联能力，我们采用 $G_t \in \mathbb{R}^{N\times N}$（Dang等人，2022a）提出的无偏有向对象注意力（DOA）图，为每个物体分配权重。我们基于目标 G_t^p 从 G_t 中提取物体的注意力权重向量 p，并将其分配给每个编码后的目标特征：

$$ST_o = \delta(ST_i W^{ST}) \odot G_t^p \tag{2}$$

W^{ST} 是一个可学习的参数矩阵，$\odot$ 使每个物体特征都能乘以其对应的注意力系数。

导航思维 导航思维需要具备记忆、定位和导航至目标的能力。我们借鉴了（Dang等人，2022b）提出的目标感知多尺度聚合器（TAMSA），将不同位置的目标观测信息映射为相对于当前状态的目标方向。

首先，决策（例如向右转）是相对于智能体的当前状态做出的（x_c，y_c，θ_c，β_c），因此我们对存储在 TOMG 中的智能体状态（x_i，y_i，θ_i，β_i）进行自我中心化处理。

$$\begin{aligned}(\widetilde{x}_i, \widetilde{y}_i) &= (x_i, y_i) - (x_c, y_c) \\ (\widetilde{\theta}_i^x, \widetilde{\beta}_i^x) &= sin((\theta_i, \beta_i) - (\theta_c, \beta_c)) \\ (\widetilde{\theta}_i^y, \widetilde{\beta}_i^y) &= cos((\theta_i, \beta_i) - (\theta_c, \beta_c)) \quad i \in \Delta_M\end{aligned} \tag{3}$$

其中，Δ_M 表示目标可见节点的索引集合。为确保角度坐标与位置坐标的数量级相同，我们使用 sin 和 cos 将角度坐标归一化到 $[-1, 1]$。经过这种自我中心坐标变换后，我们得到自我中心 TOMG 特征 $\widetilde{NT_i} \in \mathbb{R}^{D_n\times 11}$。同样，探索思维和障碍物思维中的智能体状态也需要按照上述方法进行自我中心变换。

后续编码过程表示如下：

$$NT_o = H^T F_{NT}(\widetilde{NT_i}) \odot F_E(E) \tag{4}$$

$$H = \sum_{j=1}^{3} TCN_j(\widetilde{NT_i}) \tag{5}$$

$H \in \mathbb{R}^{D_n\times 1}$ 由不同尺度的卷积核求和得到，这些卷积核通过使用多尺度时间卷积网络（TCNs）处理 NT_i 生成。$F_{NT}(\cdot)$ 将 NT_i 映射到高维特征空间。由于搜索不同目标时，导航思维需要适应性地改变，因此独热目标索引 E 经由两个全连接层（FC）$F_E(\cdot)$ 进行编码，从而生成通道激活向量，用于重新校准通道特征响应。

探索思维 我们希望智能体能够利用探索思维更高效地探索环境，并避免重复探索。在对智能体的状态进行自我中心化处理后

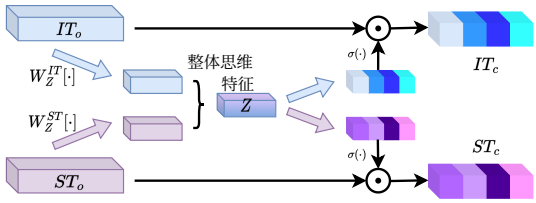


图 4。**两个思维之间的多重思维协作（MTC）模块**。思维数量的增加不会影响整体结构。我们首先从多重思维的输出中提取整体思维特征。随后，为每种思维生成通道激活向量，以重新校准思维特征。

我们还通过极坐标化坐标引入了两种归纳偏置。(i)借助极坐标的距离，网络可以学习到距离当前节点越近的历史状态越重要。(ii)借助极坐标的角度，网络可以学习到朝探索较少的方向行进能够获得更多场景信息。随后，我们使用两个全连接层 $F_{ET}(\cdot)$ 和一个全局平均池化层，得到探索思维的输出。

$$ET_o = \frac{1}{D_e}\sum_{l=1}^{D_e} F_{ET}(g(ET_i\langle l\rangle)) \tag{6}$$

其中，$\langle l\rangle$ 从历史记忆图中检索第 l 个节点的特征，而 $g(\cdot)$ 表示将笛卡尔坐标系转换为极坐标系的过程。

障碍物思维 以往的方法通常会反复与同一障碍物发生碰撞，从而陷入死锁。我们的障碍物思维通过记忆发生碰撞的障碍物，帮助智能体迅速摆脱死锁状态。整体编码过程与探索思维的编码过程相似。

$$OT_o = \frac{1}{D_o}\sum_{l=1}^{D_o} F_{OT}(g(OT_i\langle l\rangle)) \tag{7}$$

在上述五个思维编码网络中，较为复杂的直觉思维、搜索思维和导航思维之后均添加了 Dropout 层。

4.4. 多重思维协作（MTC）

尽管我们将物体导航所需的元能力解耦，但不同元能力思维之间仍有必要协作。例如，当搜索思维发现目标位于智能体右侧时，障碍物思维需要对右侧的障碍物给予更多关注。因此，我们设计了一个多重思维协作（MTC）模块（图 4），用于在不同思维之间传递共享信息。

MTC模块主要利用来自全部思维的浓缩信息，重新校准每种思维的通道权重。首先，我们将多个思维的输出压缩为整体思维特征表示 Z:

$$Z = \delta(W_Z\,[IT_o, ST_o, NT_o, ET_o, OT_o] + b_Z) \tag{8}$$

然后，生成每种思维的激励信号，以重新校准各思维的输出 $\mathcal{X}T_o$:

$$\mathcal{X}T_c = \mathcal{X}T_o \odot \sigma(W_{\mathcal{X}T}Z + b_{\mathcal{X}T}) \quad \mathcal{X} \to I, S, N, E, O \tag{9}$$

其中 $\sigma(\cdot)$ 表示 Sigmoid 激活函数。

4.5. 策略学习

经过整体思维校准后，需要将所有思维特征整合为一个统一的表示向量:

$$G = \delta(W_G[IT_c, ST_c, NT_c, ET_c, OT_c]_{LN} + b_G) \tag{10}$$

其中 $[\cdot]_{LN}$ 将每种思维的最终输出特征拼接起来，并使用层归一化来稳定前向输入分布和反向传播梯度。最后，多重思维联合表示 G 用于学习一个长短期记忆网络（Hochreiter & Schmidhuber，1997）动作策略 $\pi(a_t|G_t,\ p)$。沿用先前的研究（Yang等人，2019；Dang等人，2022a），我们将此任务视为强化学习问题，并采用异步优势演员-评论家算法（异步优势演员-评论家算法）（Mnih等人，2016）。

4.6. 元能力奖励

我们的模型由两种奖励进行监督：基础奖励 R_B 和元能力奖励 R_{MA}。与先前的研究（Zhang等人，2021年）类似，R_B 由三部分组成。(i) 每执行一步，我们都会施加 -0.01 的小额负奖励。(ii) 为鼓励移动，如果智能体输出 $MoveAhead$，则给予 0.01 的正奖励。(iii) 如果在一定步数内到达目标物体类别中的任一物体实例，智能体将获得 5.0 的大额正奖励。元能力奖励旨在实现每种元能力所需达成的目标。

$$R_{MA} = R_s + R_n + R_e + R_o \tag{11}$$

搜索奖励 R_s 如果在视野中正确识别出目标物体，则为 $R_s = 0.01$，否则为 $R_s = 0$。

导航奖励 R_n 一旦找到目标，如果所选动作能让智能体更接近目标，$R_n = 0.01$；否则 $R_n = 0$。

探索奖励 R_e 如果智能体反复到达同一状态， $R_e = -0.01$; 否则 $R_e = 0$。

障碍物奖励 R_o 如果智能体与障碍物发生碰撞， $R_o = -0.01$, 否则 $R_o = 0$。

R_{MA} 使每种元能力思维都能更快地把握学习方向。因此，在最初的 C 训练回合中，为引导模型的学习方向，模型同时接受元能力奖励 R_{MA} 和基础奖励 R_B的监督。之后，为了获得更好的任务性能指标，模型仅接受基础奖励 R_B的监督。

5. 实验

5.1. 实验设置

数据集AI2-Thor（Kolve 等人，2017）和 RoboTHOR（Deitke 等人，2020）是我们的主要实验平台。AI2-Thor 针对 4 种房间布局分别包含 30 张不同的平面图：厨房、客厅、卧室和浴室。对于每种场景类型，我们使用 20 个房间进行训练、5 个房间进行验证，另用 5 个房间进行测试。RoboTHOR 由 89 套公寓组成，其中 75 套可供使用。我们使用其中 60 套进行训练，15 套进行验证。RoboTHOR 是 AI2-Thor 环境的复杂版本，其地面面积大 2.4 倍，路径长度长 5.5 倍。

对于零样本物体导航，我们将广泛使用的 22 个目标类别（Pal 等人，2021年；Zhao 等人，2022）重新划分为 18/4 和 14/8 个已见/未见类别。我们以已见物体类别作为目标来训练模型，并以未见物体类别作为目标来测试模型。

评估指标 我们使用成功率（SR）和按路径长度加权的成功率（SPL）（Anderson 等人，2018）指标来评估我们的方法。SR 表示智能体完成任务的成功率，公式为 $SR = \frac{1}{F}\sum_{i=1}^{F} Suc_i$，其中 F 是回合数， Suc_i 表示第 i个回合是否成功。SPL 更全面地考虑路径长度，定义为 $SPL = \frac{1}{F}\sum_{i=1}^{F} Suc_i \frac{L_i^*}{max(L_i, L_i^*)}$，其中 L_i 是智能体走过的路径长度， L_i^* 是理论上的最短路径。

实现细节 仅采用直觉思维（IT）和基础奖励 R_B 的模型是我们的基线。我们使用 2 块 RTX 2080Ti Nvidia GPU 和 18 个工作进程训练模型，共进行 3M 个导航回合。丢弃率设为 0.3，且元能力奖励 R_{MA} 仅用于最初的 0.2M （ C ）个回合。我们报告全部目标（ ALL ）以及最优轨迹长度大于 5 的部分目标（$L \geq 5$）的结果。

表 *1*。各项元能力的消融实验。移除某项元能力意味着移除相应的思维和元能力奖励。

IT	ST	NT	ET	OT	全部 (%) SR↑	全部 (%) SPL↑	$L \geq 5$(%) SR↑	$L \geq 5$(%) SPL↑
✓					43.48	18.91	30.36	14.34
✓	✓				75.94	42.77	68.24	43.28
✓	✓	✓			79.62	46.13	72.96	45.93
✓	✓	✓	✓		81.97	48.75	75.54	48.95
✓	✓	✓	✓	✓	**83.14**	**50.23**	**77.03**	**50.88**

表 *2*。 多重思维协作（MTC）模块和元能力奖励的消融实验。

ID	方法	全部 (%) 成功率↑	全部 (%) SPL↑	$L \geq 5$(%) 成功率↑	$L \geq 5$(%) SPL↑
1	完整 MT	83.14	50.23	77.03	50.88
2	MT →无 MTC	82.26	50.14	76.17	50.52
3	MT → 否 R_{MA}	81.96	49.54	76.22	49.93
4	$R_B + R_{MA}$ (全部回合)	81.31	48.29	75.36	48.44

5.2. 消融实验

元能力消融 物体导航任务被分解为五种元能力，并在表1中对其进行了消融。根据 MAD 结构，搜索能力是最重要的元能力，其次是导航能力，而探索能力和避障能力则起到辅助作用。

MTC与元能力奖励消融表2展示了消融结果：在第二行和第三行中分别移除了MTC模块和元能力奖励 R_{MA}。我们观察到，MTC模块对成功率的影响更大，而元能力奖励则提升了成功率和SPL。表2第四行的结果 R_B 表明，在整个训练过程中使用基础奖励 R_{MA} 和元能力奖励，会导致在训练后期使用元能力奖励时，模型偏离对最终目标（通过最短路径找到物体）的追求，并造成奖励与实际性能脱节。

5.3. 不同目标的比较分析

图 5 比较了我们的MT方法与当前最先进水平的方法（DOA（Dang等人，2022a））在每个目标物体上的成功率。以往方法在处理小型物体以及复杂环境（例如卧室）中的物体时表现不佳。受益于我们方法最多的五个目标物体（以红色标注）大多是此前尚未解决的目标。受益于我们方法最少的五个目标物体（以蓝色标注）大多常见于较简单的厨房场景。从饼图中可以观察到，与简单场景（例如厨房、浴室）相比，我们的MT模型对复杂场景（例如卧室、客厅）中的成功率提升做出了更大的总体贡献

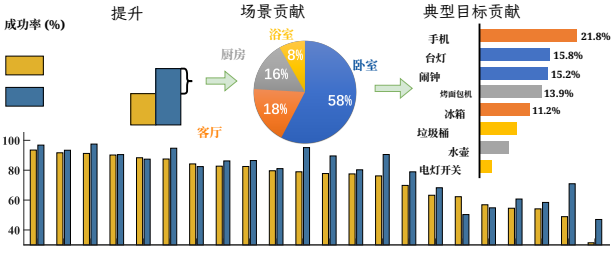


图*5*。**针对每个目标，比较我们的 MT 方法与 DOA 方法在 SR 指标上的表现**。红色和蓝色标记分别表示 MT 方法性能提升最高和最低的目标。饼图展示了各场景对总体 SR 提升的贡献。随后，将每个场景中贡献最高的两个物体绘制成条形图。

表 *3*。在 AI2-Thor / RoboTHOR 数据集上与特定目标的最先进方法进行比较。

ID	方法	全部 (%) SR↑	全部 (%) SPL↑	$L \geq 5$(%) SR↑	$L \geq 5$(%) SPL↑	回合时间 (s)↓
I	SSCNav	77.14/38.12	31.09/14.10	71.73/33.46	34.33/11.04	1.34/4.14
	PONI	78.58/38.42	33.78/16.30	72.92/34.72	36.40/13.22	1.59/4.58
II	OMT	71.13/32.17	37.27/20.09	61.94/25.33	38.19/18.16	0.64/2.01
	VGM	73.95/35.82	40.69/23.71	64.07/27.22	40.73/19.54	0.73/2.46
III	TPN	67.32/30.51	37.01/18.62	58.13/23.89	35.90/14.91	0.24/0.77
IV	HOZ	68.53/31.67	37.50/19.02	60.27/24.32	36.61/14.81	0.28/0.81
	VTNet	72.24/33.92	44.57/23.88	63.19/26.77	43.84/19.80	0.32/1.33
	DOA	74.32/36.22	40.27/22.12	67.88/30.16	40.36/18.32	0.33/1.25
V	**MT**	**83.14/42.80**				

）。这些发现表明，我们的MT方法能够通过灵活的元能力，应对复杂环境中的多方面决策挑战。更多实验见附录D 和E。

5.4. 与最先进方法的比较

目标特定典型物体导航 在表 3中，我们的 MT 模型与四类最先进模型进行了比较。 **(I) SLAM 方法**。 实时构建语义地图并进行子目标路径规划，增强了这些方法的可解释性。然而，由于探索环境的成本高昂，每个回合所需的时间是其他方法的数倍。 **(II) 记忆方法**。对长期历史信息的显式记忆增强了模型的探索和导航能力。尽管如此，现有记忆方法存在大量冗余，导致泛化能力较差。 **(III) 死锁专用方法**。 导航过程中经常出现死锁状态

表 *4*. 在 AI2-Thor 数据集上与目标无关零样本最先进方法的比较。

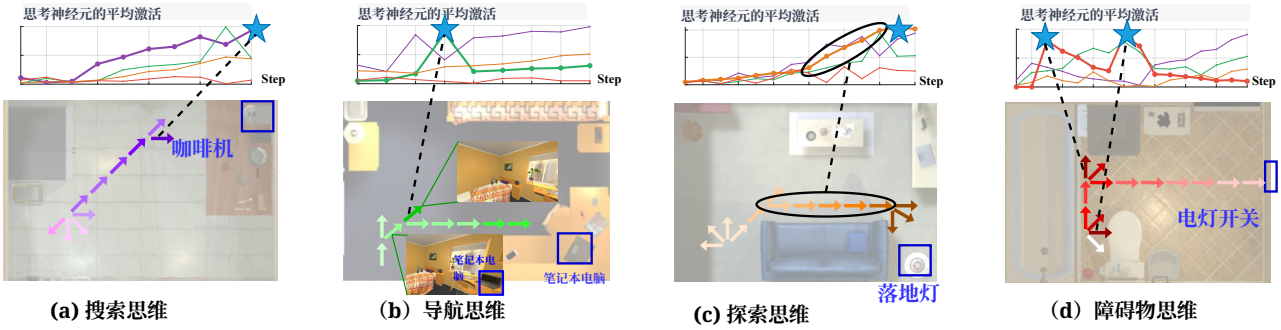
方法	已见/未见 划分	未见类别 全部 (%) SR↑	全部 (%) SPL↑	$L \geq 5$(%) 成功率↑	$L \geq 5$(%) SPL↑
随机	18/4	9.76	2.03	0.82	0.27
ZER	18/4	31.28	15.06	25.74	14.60
ZSON	18/4	57.32	20.94	46.43	21.78
MT-ZS	18/4	**68.61**	**27.95**	**57.53**	**29.28**
随机	14/8	7.70	3.19	0.44	0.08
ZER	14/8	24.62	10.42	14.33	8.99
ZSON	14/8	52.74	18.11	33.53	14.38
MT-ZS	14/8	**62.40**	**24.08**	**46.58**	**23.76**

。尽管引入死锁专用模块可以解决一些问题，却会破坏模型训练的整体连贯性。 **(IV) 关联方法**。 物体或区域关联是最常见的归纳偏置。严格来说，关联方法仅用于增强搜索能力，对其他能力影响甚微。在 MAD 范式的指导下，我们的 MT 方法显式解耦物体导航任务中的各种元能力，从理论上统一了上述四类模型。与计算复杂度相近的最先进方法（DOA（Dang等人，2022a））相比，我们的 MT 方法使成功率和 SPL（AI2-Thor / RoboTHOR，%）整体分别提升了 8.82/6.58 和 9.96/6.95。

目标无关零样本物体导航 与典型的物体导航任务不同，零样本任务中的目标物体在训练期间不可见。因此，我们将MT模型中的物体独热编码替换为目标的Glove嵌入余弦相似度（Zhao 等人，2022），由此得到MT-ZS模型。与基于CLIP的ZER模型（Khandelwal 等人，2022）和类别无关的ZSON模型（Zhao 等人，2022）相比，我们基于MAD范式的MT-ZS模型展现出明显优势（表4）。MAD范式在零样本任务中的成功，也证明了其对于泛化难度较高的具身人工智能任务同样有效。更多实验见附录 C。

5.5. 元能力定性分析

我们的MT模型在导航过程中基于各种元能力的综合作用做出决策。在图6中，我们将思维输出神经元在每个 $\mathcal{X}T_c$ 步骤的均值可视化，以探索每种思维如何在不同场景下影响模型推理。直觉思维没有表现出明显规律，因此在此仅展示其余四种思维。(a) 当观察到目标或与目标相关的物体时，搜索思维会被激活；随着接近目标，其活跃程度逐渐增强。(b)



(a) 搜索思维 (b) 导航思维 (c) 探索思维 (d) 障碍物思维

图6。 **我们将导航过程中每种思维对应神经元的平均激活可视化**。 箭头颜色的深浅表示当前思维对应输出神经元的平均值，与上方折线图相对应。蓝色五角星表示路径中该思维最活跃的步骤。

表 5。不同模型元能力的定量比较。

方法	全部 (%) SSR (S) ↑	NSNLP (N) ↑	REP (E) ↓	碰撞惩罚 (O) ↓
基线	91.35	23.15	5.28	12.74
DOA	$95.82_{(+4.47)}$	$44.11_{(+20.96)}$	$7.14_{(+1.86)}$	$10.26_{(-2.48)}$
MT	$97.76_{(+6.14)}$	$51.39_{(+28.24)}$	$4.03_{(-1.25)}$	$4.93_{(-7.81)}$

当目标物体突然从智能体的视野中消失时，导航思维的水平会显著增强，使智能体能够迅速重新找到目标物体。(c) 持续向前移动会迅速激活探索思维。(d) 当智能体遇到障碍物时，障碍物思维的激活程度达到最大；随着其与障碍物之间的距离增加，激活程度逐渐降低。每种思维的激活都为基于元能力的模型决策过程提供了更清晰的解释。更多分析见附录F。

5.6. 元能力定量分析

元能力指标 为了定量评估各模型的元能力，我们定义了四项元能力指标：(i) 搜索成功率 (SSR)：找到目标的成功率；(ii) 按导航路径长度加权的导航成功率 (NSNPL)：找到目标后，在导航阶段的 SPL；(iii) 重复探索概率 (REP)：重复到达相同状态的概率；(iv) 碰撞概率 (CP)：导致与障碍物发生碰撞的动作所占比例。SSR 和 NSNPL 值越大，表明搜索和导航能力越强；REP 和 CP 值越小，则表明探索和避障能力越强。更多详细说明见附录 B。

分析 如表 5所示，我们的 MT 模型在各项元能力上的表现都显著优于其他模型。当前最先进水平的 DOA方法主要利用物体关联来增强搜索能力；然而，REP 表明，相较于基线模型，DOA方法的探索能力有所下降。这一现象说明，如果不对元能力进行解耦，增强一种元能力可能会削弱其他元能力。值得注意的是，我们的 MT模型仅使用了 DOA 模型中一小部分用于增强搜索能力的归纳偏置（第 4.3 节），但 MT模型的搜索能力仍优于 DOA 模型。这一发现使我们认为，MAD范式所促进的思维特异性可以放大每种归纳偏置对相应元能力的影响。

6. 局限性

本文仍存在一些局限性。(i) 元能力的选择依赖于人类经验，因此如何解耦更抽象的元能力仍是一个开放问题。(ii) 本文仅将 MAD 应用于物体导航任务并进行了实验，我们希望研究人员能将其拓展到更多具身人工智能任务。(iii) 元能力思维如何影响模型的决策，仍有许多值得探索的方向。

7. 结论

本文提出了元能力解耦（MAD）范式，引导研究人员从元能力的角度设计和分析物体导航模型。基于 MAD，我们设计了多重思维（MT）模型，该模型在典型物体导航任务和零样本物体导航任务中的表现均显著优于最先进方法。此外，我们从元能力层面对 MT模型进行了定性和定量的可解释性分析。除物体导航外，其底层原理在理论上也可推广至其他具身人工智能任务。

参考文献

Al-Halah, Z., Ramakrishnan, S. K., and Grauman, K. 无需经验：用于语义视觉导航的即插即用模块化迁移学习。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第17031–17041页，2022年。

Anderson、P.、Chang、A.、Chaplot、D. S.、Dosovitskiy、A.、Gupta、S.、Koltun、V.、Kosecka、J.、Malik、J.、Mottaghi、R.、Savva、M. 等。具身导航智能体的评估。*arXiv*预印本 *arXiv:1807.06757*，2018。

Cao、Y.-H. 和 Wu、J. 随机卷积神经网络能够识别物体：卷积神经网络的一种归纳偏置及其应用。载于*AAAI*人工智能大会论文集，第 1 卷，第 194–202 页，2022。

Carion、N.、Massa、F.、Synnaeve、G.、Usunier、N.、Kirillov、A. 和 Zagoruyko、S. 使用 Transformer 进行端到端目标检测。载于欧洲计算机视觉会议，第 213–229 页，2020。

Chaplot, D. S., Gandhi, D., Gupta, S., Gupta, A., 和 Salakhutdinov, R. 使用主动神经SLAM学习探索。发表于国际学习表征会议，*ICLR*，2020a。

Chaplot, D. S., Gandhi, D. P., Gupta, A., 和 Salakhutdinov, R. R. 使用目标导向的语义探索进行物体目标导航。神经信息处理系统进展，33:4247–4258，2020b。

Chen, S., Guhur, P., Tapaswi, M., Schmid, C., 和 Laptev, I. 放眼全局，立足局部：用于视觉语言导航的双尺度图Transformer。发表于*IEEE/CVF*计算机视觉与模式识别会议，第16516–16526页，2022。

Dang, R., Shi, Z., Wang, L., He, Z., Liu, C., 和 Chen, Q. 用于物体导航的无偏有向物体注意力图。发表于*MM '22*：第*30*届*ACM*国际多媒体会议，第3617–3627页，2022a。

Dang, R., Wang, L., He, Z., Su, S., Liu, C., 和 Chen, Q. 搜索还是导航？用于物体导航的双重自适应思考。*arXiv*预印本 *arXiv:2208.00553*，2022b。

d'Ascoli, S., Touvron, H., Leavitt, M. L., Morcos, A. S., Biroli, G., 和 Sagun, L. ConViT：通过软卷积归纳偏置改进视觉Transformer。发表于国际机器学习会议，第 2286–2296 页，2021年。

Deitke, M., Han, W., Herrasti, A., Kembhavi, A., Kolve, E., Mottaghi, R., Salvador, J., Schwenk, D., Vander-Bilt, E., Wallingford, M., 等。RoboTHOR：一个开放的仿真到现实具身人工智能平台。在*IEEE/CVF*计算机视觉与模式识别会议论文集，第3164–3174页，2020年。Du, H., Yu, X., 和 Zheng, L. 学习用于视觉导航的物体关系图和暂定策略。在欧洲计算机视觉会议，第19–34页，2020年。Du, H., Yu, X., 和 Zheng, L. VTNet：用于物体目标导航的视觉Transformer网络。在国际学习表征会议，2021年。

Duan, D., Wu, X., 和 Si, S. 基于网络解耦方法的神经网络新型可解释机制。《工程管理前沿》，8(4)：572–581，2021年。

Fukushima, R., Ota, K., Kanezaki, A., Sasaki, Y., 和 Yoshiyasu, Y. 用于物体目标导航的物体记忆Transformer。发表于国际机器人与自动化会议，第 11288–11294 页，2022年。

Gao, C., Chen, J., Liu, S., Wang, L., Zhang,Q., 和 Wu, Q. 用于远程具身指代表达的房间与物体感知知识推理。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第 3064–3073 页，2021年。

He, K., Zhang, X., Ren, S., 和 Sun, J. 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论文集，第 770–778 页，2016 年。

Hochreiter, S.和 Schmidhuber, J. 长短期记忆。神经计算，9(8):1735–1780，1997 年。

Khandelwal, A., Weihs, L., Mottaghi, R., 和 Kembhavi, A. 简单而有效：面向具身人工智能的 CLIP嵌入。载于*IEEE/CVF*计算机视觉与模式识别会议，第 14809–14818 页，2022 年。

Kharitonov, E. 和 Chaabouni, R. 犹豫时会怎么做：对序列到序列学习器中归纳偏置的研究。载于国际学习表征会议，2021年。

Kolve, E., Mottaghi, R., Han, W., VanderBilt, E., Weihs, L., Herrasti, A., Gordon, D., Zhu, Y., Gupta, A. 和 Farhadi, A. Ai2-thor：用于视觉人工智能的交互式三维环境。*arXiv*预印本 *arXiv:1712.05474*，2017 年。

Kwon, O., Kim, N., Choi, Y., Yoo, H., Park, J. 和 Oh, S. 用于视觉导航的无监督表征视觉图记忆。载于*IEEE/CVF*国际计算机视觉会议，第15870–15879页，2021年。

Levine, Y.、Wies, N.、Jannai, D.、Navon, D.、Hoshen, Y. 和 Shashua, A. 上下文学习的归纳偏置：

重新思考预训练样例设计。在国际学习表征会议，2022 年。

Liang, Y., Chen, B. 和 Song, S. SSCNav：面向视觉语义导航的置信度感知语义场景补全。载于*IEEE*国际机器人与自动化会议，第 13194–13200 页，2021年。

Lin, X., Li, G. 和 Yu, Y. 用于远程具身视觉定位的场景直觉智能体。载于*IEEE*计算机视觉与模式识别会议，第 7036–7045 页，2021年。

Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., Silver, D. 和 Kavukcuoglu, K. 深度强化学习的异步方法。载于国际机器学习会议，第 1928–1937 页，2016年。Moghaddam, M. M. K., Abbasnejad, E., Wu, Q., Shi, Q. J. 和 van den Hengel, A. Foresi：成功感知视觉导航智能体。载于*IEEE/CVF*冬季计算机视觉应用会议，第 3401–3410 页，2022年。Nair, V. 和 Hinton, G. E. 修正线性单元改进受限玻尔兹曼机。载于国际机器学习会议，2010年。Pal, A., Qiu, Y. 和 Christensen, H. 学习用于目标物体导航的层次关系。载于机器人学习会议，第 517–528 页，2021年。Park, J., Yoon, T., Hong, J., Yu, Y., Pan, M. 和 Choi, S. 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。Pennington, J., Socher, R. 和 Manning, C. D. Glove：用于词表示的全局向量。载于*2014*年自然语言处理实证方法会议论文集（*EMNLP*），第 1532–1543 页，2014年。Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J. 等。通过自然语言监督学习可迁移视觉模型。载于国际机器学习会议，第 8748–8763 页，2021年。Ramakrishnan, S. K., Chaplot, D. S., Al-Halah, Z., Malik, J. 和 Grauman, K. PONI：通过无交互学习，为目标物体导航构建势函数。载于*IEEE/CVF*计算机视觉与模式识别会议，第 18868–18878 页，2022年。Ravichandran, Z., Peng, L., Hughes, N., Griffith, J. D. 和 Carlone, L. 层次表示与显式记忆：学习有效的导航策略，应用于

3d 场景图 hs usi$^{ng\ grap}$ h $_{neura}$ l networks. In *Int* 机器人与自动化国际会议，第 9272–9279 页，2022年。Savva, M., Kadian, A., Maksymets, O., Zhao, Y., Wijmans, E., Jain, B., Straub, J., Liu, J., Koltun, V., Malik, J., 等。Habitat：具身人工智能研究平台。载于*IEEE/CVF*国际计算机视觉会议论文集，第 9339–9347 页，2019年。Wang, L., He, Z., Dang, R., Chen, H., Liu, C., 和 Chen, Q. Res-STS：通过结合评分器的自训练，为目标导向的视觉语言导航生成指代表达。*IEEE*电路与系统视频技术汇刊，2022年。Wortsman, M., Ehsani, K., Rastegari, M., Farhadi, A., 和 Mottaghi, R. 学会如何学习：基于元学习的自适应视觉导航。载于 *IEEE*计算机视觉与模式识别会议，第 6750–6759 页，2019年。Wu, Y., Wu, Y., Tamar, A., Russell, S., Gkioxari, G., 和 Tian, Y. 用于语义视觉导航的贝叶斯关系记忆。载于*IEEE/CVF*国际计算机视觉会议论文集，第 2769–2779 页，2019年。Yang, W., Wang, X., Farhadi, A., Gupta, A., 和 Mottaghi, R. 利用场景先验进行视觉语义导航。载于国际学习表征会议，2019年。Zhang, S., Song, X., Bai, Y., Li, W., Chu, Y., 和 Jiang, S. 用于物体导航的分层对象到区域图。载于*IEEE/CVF*国际计算机视觉会议论文集，第 15130–15140 页，2021年。Zhao, Q., Zhang, L., He, B., Qiao, H., 和 Liu, Z. 零样本物体目标视觉导航。*arXiv*预印本 *arXiv:2206.07423*，2022年。Zheng, K., Chitnis, R., Sung, Y., Konidaris, G., 和 Tellex, S. 朝着最优相关性对象搜索迈进。载于机器人与自动化国际会议，第 7313–7319 页，2022年。Zhu, F., Liang, X., Zhu, Y., Yu, Q., Chang, X., 和 Liang, X. SOON：基于图探索的场景导向对象导航。载于*IEEE*计算机视觉与模式识别会议，第 12689–12699 页，2021年。

A. 相关工作

正文将典型的物体导航方法分为四类，并概述了这四类方法所解决的问题。本章将更详细地介绍每一类中的代表性模型。

A.1. 关联方法

关联方法大致可分为三类：物体关联、区域关联和房间关联。具有代表性的物体关联方法包括 SP（Yang 等，2021年）、DOA（Dang 等，2022a）和 CKR（Gao 等，2021年）。SP 和 DOA 仅依赖环境场景中的数据，对已知物体之间的空间相关性进行建模。CKR 则将来自大规模外部知识图谱的物体间语义相关性纳入模型。HOZ（Zhang 等，2021年）提出了区域关联，以由粗到细的方式引导智能体。BRM（Wu 等，2019）采用概率房间关系图的形式来捕捉布局先验。

A.2. 记忆方法

记忆方法会显式存储大量历史信息，例如视觉特征、坐标特征、物体特征等。VGM（Kwon 等，2021年）基于观测图像无监督表征之间的相似性逐步构建，而这些表征则从无标签图像数据集中学习得到。OMT（Fukushima 等，2022）使用 Transformer 处理存储在记忆中的显著物体。DUET（Chen 等，2022）提出联合长期动作规划，以实现全局动作空间中的高效探索。

A.3. SLAM 方法

传统的已知环境导航方法都依赖于 SLAM地图，因此，通过实时建图来探索未知环境也是一种可行的方法。由于建图成本高昂，目前大多数方法选择使用较粗略的语义地图。GOSE（Chaplot 等，2020b）构建情景语义地图，并根据目标物体类别利用该地图高效探索环境。SSCNav（Liang 等，2021年）通过置信度感知语义场景补全模块显式建模场景先验，以补全场景并指导智能体进行导航规划。PONI（Ramakrishnan 等，2022）提出了一种网络，该网络以语义地图为条件，预测两个互补的势函数，并利用它们决定到哪里寻找未见物体。

A.4. 死锁专用方法

死锁专用模块通常是整体方法的一部分，用于帮助智能体摆脱循环状态。TPN（Du 等，2020）使用预训练主模型探索环境，并为死锁状态提供专家动作。SAVN（Wortsman 等，2019）将观测数据的相似性作为判断动作是否成功的依据，并将其纳入损失函数。

B. 元能力指标

在正文中，我们介绍了用于评估四种元能力的新指标。本章将更详细地说明这四项指标。为了区分搜索能力和导航能力，我们以首个目标可见帧为界，将整个回合划分为“搜索”阶段和“导航至”阶段。在“搜索”阶段，智能体主要依靠搜索能力来定位目标物体。一旦观察到目标物体，智能体便进入“导航至”阶段，主要依靠导航能力前往目标物体所在位置。

S搜索能力指标 SSR 是“搜索”阶段的成功率，其公式为

$$SSR = \frac{1}{F}\sum_{i=1}^{F} Nav_i \tag{12}$$

其中，Nav_i 表示第 i 个回合是否进入“导航至”阶段。

导航能力指标 **c** NSNPL 考虑“导航至”阶段的导航效率，其定义如下：

$$NSNPL = \frac{1}{F_{Nav}} \sum_{i=1}^{F} Suc_i Nav_i \frac{L_i^{*Nav}}{max(L_i^{Nav}, L_i^{*Nav})} \tag{13}$$

其中， Suc_i 表示第 i个回合是否成功， F_{Nav} 表示进入“导航至”阶段的回合数。 L_i^{Nav}表示“导航至”阶段的路径长度， L_i^{*Nav}表示“导航至”阶段的理论最短路径长度。在测试过程中，我们根据每条任务路径中“导航至”阶段的起始位置（即智能体首次识别出目标的位置），实时计算 L_i^{*Nav}。直观而言，NSNPL 可视为“导航至”阶段的 SPL。

探索能力指标 REP 利用智能体返回先前访问过的状态的概率，反映了探索环境的效率。

$$REP = \frac{1}{F} \sum_{i=1}^{F} \frac{L_i - RS_i}{L_i} \tag{14}$$

其中， RS_i 表示第 i个回合中遇到的不同智能体状态的数量。

障碍物能力指标 在现实世界中，应尽可能避免与障碍物发生碰撞。CP反映了整个回合中导致与障碍物碰撞的动作所占的比例。

$$CP = \frac{1}{F} \sum_{i=1}^{F} \frac{OA_i}{L_i} \tag{15}$$

其中， OA_i 是第 i个回合中发生的障碍物碰撞次数。

C. 与最先进方法的比较

C.1. 目标特定典型物体导航

正文中，我们仅比较成功率和 SPL 指标，但对各种方法的元能力指标分析不足。表 6和表7 分别全面展示了各种方法在 AI2-Thor 和 RoboTHOR 数据集上的性能指标。SLAM 方法（I）和死锁专用方法（III）属于模块化方法，而记忆方法（II）和关联方法（IV）属于端到端方法。我们的 MAD 范式在确保端到端训练的同时，融合了上述方法的优势，并为未来研究提供了清晰的理论框架。

（I）SLAM 方法 基于 AI2-Thor 和 RoboTHOR 平台的 SLAM 方法相对较少，因此，我们将 Habitat（Savva 等，2019年）平台上的最先进方法（SSCNav（Liang 等，2022年）和 PONI（Ramakrishnan 等，2019年）迁移到 AI2-Thor 和 RoboTHOR 数据集。SLAM 方法通常采用航点预测的方式来引导智能体导航。这种离散导航模式大幅延长了搜索目标的路径，从而降低了整体 SPL。更严重的是，构建精确地图需要消耗大量计算资源和探索时间，导致其回合时间比其他方法长数倍。不过，显然，SLAM 方法具备较强的导航能力（NSNPL）。这是因为，一旦 SLAM 方法正确建立目标及其周围环境的语义地图，导航至目标位置就会容易得多。SLAM 方法受到一些研究者青睐的另一个重要原因是其具有较强的可解释性。我们希望，在 MAD 范式的基础上，端到端方法在导航任务中的可解释性能够逐步提高。

(II) 记忆方法 目前，大多数记忆方法都只是对历史记忆进行粗略建模。尽管从所有可用的历史信息中挖掘元能力可能会提升模型的整体能力，尤其是探索能力（REP），但冗余信息结构会降低泛化能力，甚至影响其他元能力。我们的 MT模型中的探索思维借鉴了记忆方法中的历史记忆结构，但我们的记忆特征更加精简，从而减轻了模型的学习负担。

表6. 与 AI2-Thor 中特定目标的最先进方法进行比较。 ✗表示不可接受的资源消耗。

ID	方法	全部 (%) SR↑	SPL↑	SSR↑	NSNPL↑	REP↓	CP↓	$L \geq 5$(%) SR↑	SPL↑	SSR↑	NSNPL↑	REP↓	CP↓	回合时间（秒）↓
I	SSCNav	77.14	31.09	89.14	51.72	5.14	**4.58**	71.73	34.33	89.02	50.73	7.30	5.41	1.342 ✗
	PONI	78.58	33.78	89.48	**52.39**	5.29	4.90	72.92	36.40	89.13	**51.82**	7.64	5.75	1.591 ✗
II	OMT	71.13	37.27	93.17	41.36	4.62	9.88	61.94	38.19	92.23	42.63	6.81	10.74	0.645
	VGM	73.95	40.69	93.20	44.21	4.51	9.30	64.07	40.73	92.14	45.97	6.62	10.14	0.714
III	TPN	67.32	37.01	91.07	40.24	5.83	5.22	58.13	35.90	90.27	38.69	8.06	6.89	0.241
IV	HOZ	68.53	37.50	91.44	40.83	8.32	10.77	60.27	36.61	90.31	39.82	11.54	11.36	0.283
	VTNet	72.24	44.57	94.18	46.74	7.91	10.71	63.19	43.84	92.85	46.15	10.88	11.52	0.321
	DOA	74.32	40.27	95.82	44.11	7.14	10.26	67.88	40.36	93.92	44.03	10.39	10.95	0.334
V	**MT**	**83.14**	**50.23**	**97.76**	51.39	**4.03**	4.93	**77.03**	**50.88**	**96.47**	51.49	**6.25**	**5.26**	0.352

表 7. 与 RoboTHOR 中特定目标的最先进方法进行比较。 ✗表示不可接受的资源消耗。

ID	方法	全部 (%) SR↑	SPL↑	SSR↑	NSNPL↑	REP↓	CP↓	$L \geq 5$(%) SR↑	SPL↑	SSR↑	NSNPL↑	REP↓	CP↓	回合时间（秒）↓
I	SSCNav	38.12	14.10	61.37	35.14	8.93	10.83	33.46	11.04	60.91	33.92	10.17	13.44	4.145 ✗
	PONI	38.42	16.30	58.46	**39.83**	8.32	11.22	34.72	13.22	58.11	**38.44**	11.64	14.27	4.582 ✗
II	OMT	32.17	20.09	61.77	24.51	7.72	16.45	25.33	18.16	57.35	23.82	9.70	18.83	2.011
	VGM	33.95	22.74	62.10	25.96	8.21	15.81	26.82	19.44	57.51	24.77	10.66	18.25	1.984
III	TPN	30.51	18.62	59.64	20.64	9.76	10.47	23.89	14.91	54.64	19.51	12.28	13.51	0.769
IV	HOZ	31.67	19.02	60.11	21.02	12.49	18.55	24.32	14.81	54.23	20.38	15.79	22.02	0.808
	VTNet	33.92	23.88	63.29	28.26	11.26	17.04	26.77	19.80	57.72	27.50	14.63	21.10	1.325
	DOA	36.22	22.12	64.18	25.88	11.33	17.14	30.16	18.32	61.39	25.11	14.82	21.52	1.247
V	**MT**	**42.17**	**29.14**	**68.05**	36.68	**7.62**	**9.91**	**37.98**	**23.80**	**66.93**	36.50	**9.25**	**12.48**	1.225

(III) 死锁专用方法 在（Wortsman 等，2019）中，研究者开始关注导航过程中的死锁问题。TPN（Du 等，2020）利用监督训练的死锁逃逸模块，使 REP 和碰撞惩罚（CP）分别达到 5.83/9.76 和 5.22/10.47（AI2-Thor/RoboTHOR，%）。然而，这种外部的死锁专用模块需要大量人工标注的逃逸动作。因此，如果发生数据集迁移，就必须再次承担高昂的标注成本。我们的 MT方法在确保模型可扩展性的同时，其 REP 和 CP 指标分别比 TPN 方法低 1.80/2.14 和 0.29/0.56（AI2-Thor/RoboTHOR，%）。

（IV）关联方法 关联方法主要学习物体之间的内在关联，以加快对目标物体的视觉捕获，从而增强搜索能力（SSR）。然而，过度关注物体层级的语义信息可能会忽视导航细节，表现为 REP 和碰撞惩罚过高。我们的 MT 方法通过解耦更多元能力，帮助关联方法优化环境探索和避障，从而使成功率和 SPL 分别提升 8.82/5.95 和 9.96/7.02 （ AI2-Thor/RoboTHOR，% ），且几乎没有增加额外参数。

C.2. 目标无关零样本物体导航

任务定义 我们将 22 种物体分为两类，即已见和未见。根据类别比例的不同，有两种实验设置。（i）18/4：从每个场景（卧室、客厅、厨房和浴室）中各抽取一种物体，并将其归入未见物体类别。（ii）14/8：从每个场景中各抽取两种物体，并将其归入未见物体类别。目标集合划分完成后将不再更改。训练期间，以已见物体作为待寻找的目标，而检测器（如目标检测或实例分割检测器）无法识别未见物体。测试期间，指示智能体导航至给定目标集合中的任意目标。

MT $\rightarrow$ MT-ZS MT模型不适用于零样本物体导航任务，因为它使用独热编码来表示物体语义，并采用固定大小的物体注意力矩阵，从模型的角度限制了物体类别的数量。零样本物体导航任务要求智能体能够定位任意数量的目标物体。因此，我们使用 GloVe 编码（Pennington 等人，2014），并基于相对于目标的语义余弦相似度进行物体关联。以目标为中心的连续语义空间

*Ta*表 *8.* 使用目标无关零样本最先进方法比较已见和未见物体上的性能 s.

方法	已见/未见划分	未见类别					已见类别				
		全部 (%)		$L \geq 5$(%)		回合长度↓	全部 (%)		$L \geq 5$(%)		回合长度↓
		成功率↑	SPL↑	成功率↑	SPL↑		成功率↑	SPL↑	成功率↑	SPL↑	
随机	18/4	9.76	2.03	0.82	0.27	37.523	9.36	2.81	1.12	0.35	38.146
ZER	18/4	31.28	15.06	25.74	14.60	18.492	25.17	10.02	21.83	10.57	19.447
ZSON	18/4	57.32	20.94	46.43	21.78	15.746	58.72	18.47	38.44	18.72	16.046
MT-ZS	18/4	**68.61**	**27.95**	**57.53**	**29.28**	13.215	**66.00**	**27.41**	**50.78**	**27.45**	14.742
随机	14/8	7.70	3.19	0.44	0.08	36.032	8.17	2.94	0.37	0.16	38.512
ZER	14/8	24.62	10.42	14.33	8.99	20.917	32.77	17.25	30.20	13.28	19.033
ZSON	14/8	52.74	18.11	33.53	14.38	16.680	59.91	23.56	34.84	20.08	15.892
MT-ZS	14/8	**62.40**	**24.80**	**46.58**	**23.76**	14.514	**70.21**	**28.48**	**57.63**	**30.46**	13.791

使智能体能够接受寻找任意目标的请求。

结果分析在正文中，我们仅分析以未见物体为目标的测试指标。在表8中，我们补充了以已见物体为目标的实验结果。为了实现零样本物体导航，ZER（Khandelwal 等人，2022）和 ZSON（Zhao 等人，2022）在已见目标集合上的性能均大幅下降。结合多种元能力的 MT-ZS 模型具有更强的导航鲁棒性。因此，我们的方法在搜索已见目标和未见目标方面都展现出明显优势。

D. 目标层面实验

总体指标显示，我们的MT模型明显优于其他方法。更具体地说，我们希望了解该方法在不同目标物体和场景条件下的性能。这些发现不仅有助于我们深入了解该方法的优势，也揭示了它的不足之处，可为未来的研究提供参考。

D.1. 实验设置

在AI2-Thor的测试平面图中，我们随机初始化了8000个任务（场景、初始位置和目标物体）。当智能体完成这8000个任务后，我们分别统计每种目标物体对应的指标（SR、SPL、SSR、NSNPL、REP、碰撞惩罚CP）。为了观察智能体执行长路径任务时的表现，我们还提取了路径长度大于5的回合。DOA（Dang等人，2022a）方法和我们的MT方法的实验结果分别列于表9和表10中。

D.2. 分析

搜索能力如图7的 SSR 子图所示，智能体寻找目标（忽略距离）的成功率已经很高，大多数目标的搜索成功率都超过了 90%。其中，炉灶、水壶和冰箱等目标的搜索成功率甚至可达 100%。这表明，引入关联机制后，算法搜索目标物体的能力确实变得非常强大。我们的 MT 方法在搜索思维的编码过程中，借鉴并简化了 DOA 中的一些物体注意力分配技术。令人意外的是，MT 的搜索能力不但没有下降，在寻找某些目标（如书和手机）时反而有所提升。这一现象有两个原因：（i）尽管我们在模型层面将元能力解耦，但完成任务时，各种元能力之间能够相互促进。在搜索阶段，智能体经常因陷入局部死锁状态而无法脱身，导致搜索失败。因此，如果智能体具备更强的环境探索能力和避障能力，搜索目标就会更容易。（ii）以往的模型通常试图通过某种思维隐式地抽象出各种元能力。这会分散思维的注意力，从而削弱元能力的归纳偏置效果。元能力解耦后，思维专用于某种元能力，学习方向也更加明确，因此归纳偏置的效果更为显著。

导航能力 如图7中的 NSNPL 子图所示，与 SSR 指标相比，NSNPL 指标在搜索不同目标时表现出显著差异。尤其是在复杂环境中搜索闹钟和手机等小型物体时，DOA方法的导航能力不足。MT模型的优势主要体现在这些具有挑战性的目标上。尽管如此，当目标物体是咖啡机、碗和平底锅时，MT方法的

表 *9.* 在 AI2-Thor 上将每个物体作为目标应用 DOA 方法的结果。

MT	全部 (%)						$L \geq 5$(%)						回合
	SR↑	SPL↑	SSR↑	NSNPL↑	REP↓	CP↓	SR↑	SPL↑	SSR↑	NSNPL↑	REP↓	CP↓	长度↓
闹钟	31.42	17.65	99.28	17.77	15.74	21.74	25.58	15.67	99.22	14.87	15.84	23.85	37.35
Book	54.54	27.31	84.84	31.74	16.03	16.54	44.16	25.75	80.00	31.15	16.92	17.91	25.43
Bowl	88.33	51.55	100.00	58.33	5.82	8.46	84.84	53.59	100.00	57.48	6.72	9.26	12.85
手机	48.90	25.05	86.26	32.24	7.91	13.91	35.24	20.28	79.50	23.95	8.94	15.27	33.28
椅子	63.24	33.08	98.02	31.59	7.07	9.88	53.00	32.28	97.26	31.08	7.81	11.66	26.19
咖啡机	91.66	53.39	100.00	53.82	8.94	4.70	87.23	57.87	100.00	54.80	10.54	5.44	12.38
台灯	78.94	42.74	93.42	49.79	5.70	10.24	69.81	42.06	90.56	46.24	7.33	11.37	17.42
落地灯	54.13	27.89	95.48	29.98	8.99	11.93	49.56	28.90	94.78	30.35	11.45	12.87	27.31
冰箱	77.77	42.66	100.00	40.59	6.02	12.38	72.09	45.06	100.00	44.12	8.31	14.07	21.18
垃圾桶	69.83	44.00	96.83	45.65	8.10	6.07	65.67	44.58	96.34	45.30	10.75	8.94	21.69
水壶	87.50	55.95	100.00	64.39	2.94	6.28	85.71	59.42	100.00	64.07	4.77	7.81	18.00
笔记本电脑	82.49	44.33	95.72	46.84	2.83	14.11	74.28	42.12	93.71	44.59	4.87	16.88	18.07
电灯开关	82.71	47.95	97.95	53.07	3.11	4.37	75.14	50.06	98.22	50.24	5.27	5.27	17.22
微波炉	91.23	46.97	100.00	46.50	3.62	5.10	87.50	52.92	100.00	51.16	5.12	6.26	15.49
Pan	62.26	35.40	94.34	40.34	4.85	6.18	56.82	30.99	93.18	38.83	4.91	7.31	24.92
盘子	79.63	41.12	96.29	43.41	6.32	3.01	72.97	42.23	94.59	44.35	7.04	3.77	17.57
Pot	56.86	30.42	92.15	33.91	5.23	4.23	41.93	23.56	87.09	25.75	6.15	4.92	25.98
遥控器	77.52	43.35	93.02	45.75	4.80	12.30	69.76	40.91	90.69	41.89	5.82	13.24	17.07
Sink	90.17	43.43	94.57	44.80	3.57	4.83	80.34	44.86	90.34	47.74	5.08	5.71	13.46
炉灶	93.49	58.89	100.00	62.92	2.83	4.97	89.19	60.55	100.00	60.34	4.56	6.24	13.11
电视	84.21	44.84	99.12	43.88	7.40	15.69	79.10	48.43	98.76	49.81	8.22	15.31	18.07
烤面包机	76.19	46.00	100.00	51.56	3.36	5.88	72.72	46.83	100.00	50.43	4.63	7.36	17.16

表 *10.* 在 AI2-Thor 上将各物体分别作为目标应用 MT 方法的结果。

MT	ALL (%) $L \geq$(%)						5						回合
	SR↑	SPL↑	SSR↑	NSNPL↑	REP↓	CP↓	SR↑	SPL↑	SSR↑	NSNPL↑	REP↓	CP↓	长度↓
闹钟	46.02	31.14	98.66	32.22	8.22	9.03	41.21	29.10	96.83	30.17	11.92	12.15	22.41
Book	59.71	34.07	89.21	37.38	7.13	7.84	45.92	32.65	84.25	34.94	11.73	11.61	28.52
Bowl	86.40	57.26	98.52	57.24	2.80	4.42	82.11	59.16	100.00	51.61	3.85	4.37	10.05
手机	69.93	34.21	90.03	40.93	5.97	5.61	59.66	29.88	87.37	35.80	6.93	6.94	27.33
椅子	67.25	36.55	98.21	35.91	6.88	4.25	58.79	39.02	97.92	36.91	6.71	6.66	19.35
咖啡机	92.36	49.72	99.14	51.02	5.16	2.04	91.63	57.54	95.03	59.88	5.26	2.65	16.84
台灯	94.13	54.93	94.92	57.85	4.11	5.28	91.42	58.51	95.17	61.85	6.10	6.83	11.97
落地灯	57.43	35.05	92.70	39.21	5.72	6.31	54.91	36.37	94.92	41.32	6.93	8.04	26.42
冰箱	88.55	48.32	100.00	51.28	4.91	5.16	91.05	53.54	100.00	56.01	6.15	6.82	16.26
垃圾桶	77.91	51.21	97.58	52.61	3.42	4.10	74.83	51.73	96.68	52.17	4.27	3.97	19.23
水壶	93.72	61.83	100.00	68.72	1.81	3.54	92.99	62.04	100.00	69.43	2.05	3.85	18.84
笔记本电脑	85.47	48.10	96.76	52.92	2.03	6.33	79.20	49.67	93.11	51.64	3.22	7.46	13.62
电灯开关	85.21	52.69	97.89	59.43	1.97	2.84	82.78	55.32	97.02	57.82	3.07	3.64	16.72
微波炉	96.49	55.08	100.00	55.13	2.14	3.34	96.25	61.47	99.78	59.55	3.14	3.16	11.62
Pan	49.30	29.94	93.12	35.90	3.83	2.87	46.74	28.81	93.62	31.02	4.17	3.15	23.24
盘子	80.05	45.33	100.00	45.23	3.16	3.02	73.85	42.13	100.00	43.00	3.24	3.00	18.01
Pot	53.81	31.75	94.10	39.04	4.02	2.91	36.19	83.42	84.10	26.31	4.38	2.54	28.74
遥控器	79.27	46.22	93.17	49.95	3.92	5.83	72.71	47.17	90.25	48.79	5.06	6.75	17.13
Sink	89.45	48.79	93.92	48.77	2.41	2.72	78.34	55.69	91.00	55.67	4.85	4.07	10.62
炉灶	95.80	63.21	100.00	68.84	1.70	2.80	93.51	69.31	100.00	63.24	2.89	3.14	10.94
电视	81.41	45.98	99.43	48.96	3.80	4.96	75.06	48.33	96.20	51.99	5.74	5.95	13.62
烤面包机	89.51	51.82	100.00	56.83	1.55	3.44	89.19	54.48	100.00	59.10	2.37	3.62	15.52

NSNPL 指标略有下降。这三种物体常见于厨房环境，而厨房环境通常布局简单、障碍物较少。由此可推断，与复杂场景相比，MT在简单场景中的优势并不那么突出。

探索能力 如图 7的 REP 子图所示，由于缺乏历史记忆，采用 DOA 模型的智能体在寻找书和闹钟等物体时，经常会反复探索同一区域。这种重复探索不仅会浪费时间，还可能扰乱模型的时序推理逻辑。解耦探索能力在解决这一问题上取得了显著改善。

避障能力 如图 7 的碰撞惩罚子图所示，DOA方法的碰撞障碍问题非常严重，甚至会反复撞上同一个障碍物。笔记本电脑、书和手机等物体通常位于复杂的场景布局中，导致智能体面临许多约束严格、难以导航的空间。在这类场景中，MT方法的避障能力对于优化避障效果非常明显。相反，盘子、锅和电灯开关等物体旁边很少有密集的障碍物，因此对避障能力的需求较低。

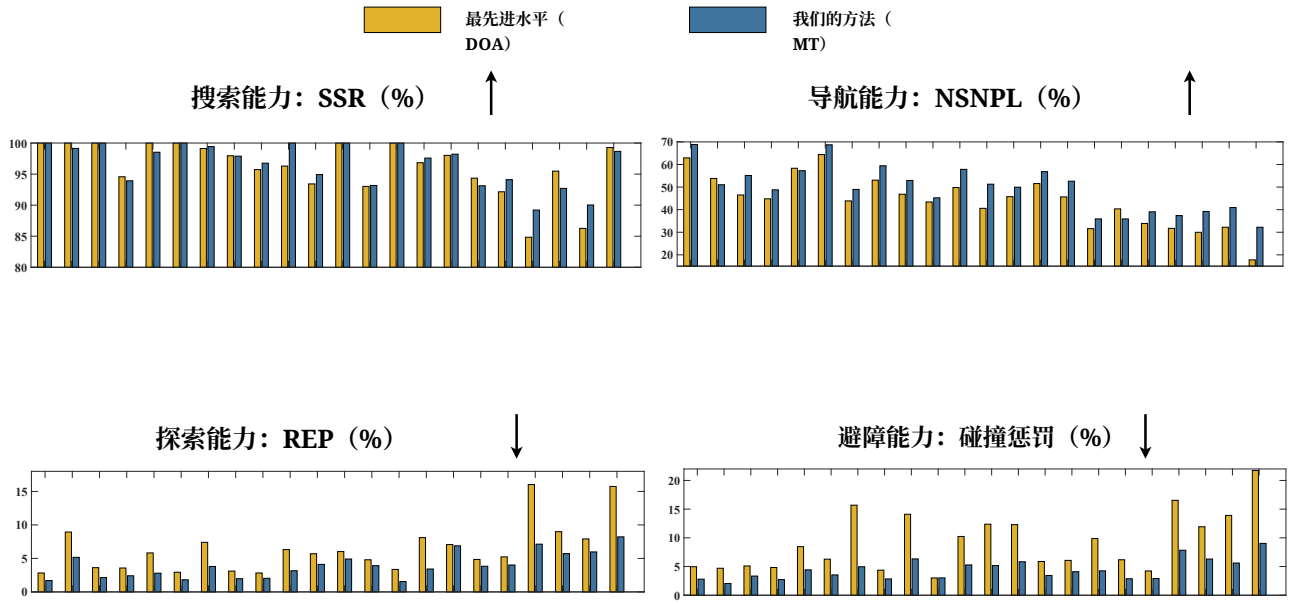


图 7. 比较DOA方法和MT方法在目标层面的元能力指标。红色和蓝色标记rs分别表示MT方法性能提升最好和最差的目标。

E. 场景级实验

通过分析每个目标物体，我们发现不同场景之间的差异相当明显。因此，我们在不同场景中对DOA方法和我们的MT方法进行实验。从每个场景中随机选取1000个任务（平面图、目标和初始状态），并让这两种方法同时完成这1000个任务。结果如表 11 所示。

E.1. 结果分析

智能体会根据场景的变化依赖不同的导航元能力。在厨房和浴室等简单环境中，仅凭直觉和搜索能力，智能体就足以高效地进行目标导航。客厅的面积更大，而且定位目标后需要走更长的导航路径。因此，智能体需要更强的导航能力，以免在丢失目标时反复搜索；我们的 MT 模型在 NSNPL 中取得了 6.38/7.79（全部/$L \geq 5$，%）的提升，体现了这一点。此外，卧室更具挑战性，因为搜索区域不仅更大，障碍物也更加复杂。面对这些情况，智能体需要更强的探索和避障能力，这一点也体现在 REP 和碰撞惩罚（CP）指标相较其他场景显著提升。

受计算机视觉和自然语言处理领域的影响，具身人工智能任务中的许多方法往往过度依赖直觉能力来解决所有问题。为反驳这一观点，我们必须认识到，复杂的长期决策任务（例如视觉语言导航（VLN）、具身问答（EQA））比简单的直觉型任务（例如图像分类、目标检测）更需要逻辑推理。如果我们只专注于提升直觉能力，算法将局限于简单环境。只有具备多种认知推理能力的智能体才能解决复杂环境中的问题。我们的 MAD范式为智能体的逻辑推理提供了理论基础和实施指导。

表11 . 我们在每个场景中将 MT方法与 DOA方法进行比较。每个场景测试 1000 个回合 s.

场景	方法	全部 (%)						$L \geq 5$(%)						回合长度↓
		SR↑	SPL↑	SSR↑	NSNPL↑	REP↓	CP ↓	SR↑	SPL↑	SSR↑	NSNPL↑	REP↓	CP ↓	
客厅	DOA	69.15	38.14	95.88	39.35	5.14	7.83	61.81	37.45	94.76	38.31	6.49	8.31	23.024
	MT	73.47	42.96	95.27	45.73	3.81	5.07	67.88	44.38	93.72	46.10	4.83	5.80	20.394
	MT - DOA	4.32	4.82	-0.61	6.38	-1.33	-2.76	6.07	6.93	-1.04	7.79	-1.66	-2.51	-2.630
厨房	DOA	83.89	49.30	98.78	52.41	4.01	4.93	78.76	50.96	98.34	51.72	4.62	4.77	16.817
	MT	87.62	55.84	98.51	55.30	3.37	3.91	82.75	56.66	97.94	55.91	3.65	3.72	16.172
	MT - DOA	3.73	6.54	-0.27	2.89	-0.64	-1.02	3.99	5.70	-0.40	4.19	-0.97	-1.05	-0.645
卧室	DOA	62.10	34.51	93.60	38.17	11.68	17.91	51.17	31.92	91.31	33.73	15.04	19.27	24.418
	MT	75.70	44.07	96.35	45.81	6.29	6.53	65.04	43.62	93.65	44.22	8.54	7.84	20.548
	MT - DOA	**13.60**	**9.56**	**2.75**	**7.64**	**-5.39**	**-11.38**	**13.87**	**11.70**	**2.34**	**10.49**	**-6.50**	**-11.43**	**-3.870**
浴室	DOA	87.10	45.64	95.00	48.94	3.41	3.22	78.72	48.54	92.84	51.80	5.13	5.61	13.848
	MT	89.05	51.17	95.43	51.76	3.62	3.28	82.33	55.03	93.51	56.07	5.66	5.56	13.250
	MT - DOA	1.95	5.53	0.43	2.82	0.21	0.06	3.61	6.49	0.67	4.27	0.53	0.05	-0.598

F. 元能力可解释性分析

在正文第5.5 节中，我们通过可视化每个步骤的思维激活情况，介绍了对导航中元能力行为特征的观察。然而，一些现象既有趣又出乎意料。我们将在下文详细解释这些现象的成因。

F.1. 搜索思维与导航思维之间主导权的转换

在大多数时间里，搜索思维和导航思维在决策中占据主导地位，因此二者之间的主导权转换会产生重大影响。如图 6 （a、b）所示，看到目标物体后，只要目标物体仍处于视野内，搜索思维就会继续占据主导地位。然而，一旦目标物体消失在视野中，导航思维便会接替它，成为决策的主导者。这一现象源于模型的长期学习经验。当目标处于视野内时，搜索思维能够直接依据目标边界框定位目标，比利用导航思维的记忆知识定位目标更加准确。目标一旦从视野中消失，搜索思维便无法有效定位目标，此时导航思维的能力就至关重要。

F.2. 探索思维对向前动作的偏好

如图 6 (c) 所示，在探索思维中，旋转的激活强度低于向前移动。我们从以下两个方面解释这一现象：

1）使用基线模型时，智能体通常会沉迷于旋转，很少向前移动，这导致其在远处物体上的成功率较低。因此，我们添加了一项奖励，以鼓励智能体向前移动，从而缓解这一问题。神经网络倾向于选择最简单的方式来拟合奖励信号。探索思维会明确记录智能体的历史路径，因此，学习提高输出向前动作的概率是最容易实现的。由此，探索思维偏好向前动作，是因为其任务就是增强智能体的向前动作。2）另一方面，我们通常认为旋转是获取环境信息最有效的方法，但向前移动以获取更详细、更准确的信息同样至关重要。