arXiv:2602.01040v1 [cs.RO] 1 Feb 2026

基于对比提示编排的自适应跨具身视觉运动策略学习

Yuhang Zhang，*IEEE* 学生会员，Chao Yan，*IEEE* 会员，Yu Jiaxi，Jiaping Xiao，*IEEE* 会员，以及 Mir Feroskhan，*IEEE* 会员

摘要——为具身智能体学习自适应视觉运动策略仍是一项艰巨挑战，尤其是在面临跨具身变化时，例如传感器配置和动态属性各不相同。传统学习方法往往难以将任务相关特征与特定于领域的变化（如光照、视场和旋转）区分开来，导致样本效率低下，并在未见环境中发生灾难性失败。为弥合这一差距，我们提出了对比提示编排（CAPO），这是一种学习视觉运动策略的新方法，融合了对比提示学习和自适应提示编排。在提示学习方面，我们设计了一种混合对比学习策略，结合视觉、时序动作和文本目标，建立一个可学习提示库，其中每个提示都会生成一种视觉表示，用以表征细粒度领域因素。基于这些学习得到的提示，我们提出了一种自适应提示编排机制，根据当前观测动态聚合这些提示。这使智能体能够即时识别主导领域因素，并自适应地构建最优状态表示。因此，策略优化有效地避免了无关干扰，防止了过拟合源域这一常见问题。大量实验表明，CAPO 在样本效率和渐近性能方面显著优于最先进基线方法。尤为重要的是，CAPO 在未见目标域上展现出卓越的零样本适应能力，这些目标域具有显著的环境变化（如照明）和物理变化（如视场和旋转），验证了其作为跨具身视觉运动策略适应可行解决方案的有效性。

索引词—**具身人工智能、视觉运动策略学习、跨具身适应、对比学习、提示编排**

I. 引言

具身人工智能（AI）的最新进展为机器人学习奠定了坚实基础，使其能够
具身智能体在复杂环境中执行各种实时决策任务，例如导航 [1],操纵 [2], 和运动 [3]。作为一种颇具前景的范式，基于视觉的具身智能体近年来受到越来越多的关注，它们依靠机载摄像头感知周围环境并生成动作指令。然而，如图1所示，视觉观测本质上与环境因素



(a) 光照敏感性

(b) 具身敏感性

(c) 光照与具身联合适应

图 1：视觉运动策略学习中的域敏感性，以及所提出的 CAPO 方法的适应能力示意图。(a) 基于传统学习的策略对光照高度敏感。在标准环境中训练的智能体面对剧烈的光照变化时无法适应，即使场景的底层布局保持不变。(b) 同样，标准方法也难以应对具身变化。在特定机器人配置上训练的策略部署到不同具身上时（例如，视场变化），便无法正常工作。(c) 相比之下，CAPO 通过自适应编排对比提示，将表征学习与策略优化协同起来。这种设计带来了出色的零样本适应能力，使智能体能够在光照与具身均发生变化的未见目标域中成功导航。

条件和智能体的物理配置，给感知和决策带来了显著的不确定性。具体而言，智能体的感官输入对环境和物理因素高度敏感，例如照明、视场（FOV）和步长，这些因素会大幅改变智能体的视觉观测。这些因素称为域因素，给学习自适应视觉运动策略带来了重大挑战。这引出了一个根本问题：我们如何开发一种策略，使智能体能够适应

Y. Zhang、J. Yu、J. Xiao 和 M. Feroskhan 任职于新加坡南洋理工大学机械与宇航工程学院，地址为新加坡 639798（电子邮箱：yuhang004@e.ntu.edu.sg；jiap-ing001@e.ntu.edu.sg；jiaxi002@e.ntu.edu.sg；mir.feroskhan@ntu.edu.sg）。C. Yan 任职于中国南京南京航空航天大学自动化学院，地址为 211106（电子邮箱：yan-chao@nuaa.edu.cn）。（通讯作者：*Mir Feroskhan* 和 *Jiaping Xiao*。）

变化的域因素?

视觉运动策略学习的一种主流范式依赖大规模专家演示和模仿学习（IL），通过训练策略来复现专家在自动驾驶等任务中的行为 [4] 以及机器人操作 [5]。然而，收集高质量的真实世界演示成本高昂、耗费人力，且难以扩展。强化学习（RL）提供了一种可扩展的替代方案：通过在高保真模拟器中交互来训练策略，从而实现高效的数据收集和快速迭代[6]。许多强化学习方法以端到端方式学习策略，直接将原始视觉输入映射为动作 [4],[7], [8]。尽管这些方法效果显著，但往往需要数百万次交互才能收敛，并且通常存在样本效率低下的问题。此外，直接从原始视觉输入学习，使这些方法对域因素导致的像素分布变化高度敏感，这种现象通常称为域差距。因此，在特定视觉条件下学习到的策略往往无法泛化到未见环境，即使底层任务结构保持不变。一些研究采用域随机化 [9]–[11], 在仿真中合成多样的视觉和物理变化，以促进适应。尽管这类方法有效，但通常需要仔细手动设计随机化分布，而且往往会以牺牲最优性为代价换取适应性，导致策略在特定环境中的表现欠佳 [12]。

为应对这些挑战，先前的研究日益转向解耦架构 [13]–[15]，用于域偏移下的策略适应，其中视觉表征学习与策略优化相分离。具体而言，近期的研究工作主要聚焦于视觉表征学习：一方面通过训练通用视觉编码器 [16]–[18]，另一方面利用大规模预训练模型 [19]–[21] 提取域不变特征。尽管这些方法提升了对分布偏移的鲁棒性，但它们通常会产生静态表征 [22],，限制了其在未见域因素下自适应地强调任务相关特征的能力。这种自适应灵活性的缺失往往会导致视觉嵌入不一致，并进一步造成策略退化。

本文提出对比提示编排（**Contr**A**s**tive **P**rompt **O**rchestration，**CAPO**），这是一种自适应跨具身视觉运动策略学习方法，旨在提升模型在未见领域中的零样本适应能力。我们的方法结合端到端范式与解耦范式的互补优势，既保留高样本效率的学习能力，又克服静态状态表示的僵化问题，从而在域偏移下实现更有效的适应。为此，CAPO 将对比提示学习策略与自适应编排机制相结合，兼顾表示的鲁棒性与策略的适应能力。在对比提示学习中，我们采用 CLIP [23]作为视觉语言骨干网络，并引入可学习提示，在不微调编码器的情况下，从同一观测中提取多个感知不同域的视觉表征。这些由提示生成的表征通过混合对比策略进行训练，其中视觉、时序动作和文本对比目标在不同提示间应用，以引导其优化。随后，为了适应动态域因素，我们设计了一种基于双分支注意力架构的自适应编排机制，根据当前观测动态聚合提示嵌入。随后，聚合后的表示被输入策略网络，并通过强化学习联合优化编排参数与策略网络。该设计使训练好的策略能够从源域（策略训练所在的域）无缝适应到目标域（策略部署所在的域）。通过结合预训练提示的稳定性与自适应编排的灵活性，我们的方法显著提高了样本效率和适应性能。

本文的贡献总结如下。

1）我们提出了 CAPO，这是一种新颖的视觉运动策略学习方法，通过保留样本高效的学习能力，同时借助集成提示学习与自适应编排克服静态表征的局限，解决了端到端强化学习和解耦强化学习存在的问题。该方法能够动态构建视觉嵌入，从而无需微调即可有效适应未见域因素。2）我们设计了一种混合对比学习策略，将视觉、时序动作和文本监督结合起来，用于对比提示学习。与标准的单目标方法相比，这种混合形式能够使编码器提取更丰富、更具适应性的特征，促使编码器封装细粒度领域因素。3）我们引入了一种双分支、基于注意力的编排机制，根据当前观测动态重新加权提示诱导嵌入，从而有效适应未见目标域和具身变化。4）我们开展了大量实验以评估 CAPO。结果表明，在样本效率和渐近性能方面，CAPO 优于最先进基线方法。此外，在涉及未见具身形态或环境因素的目标域中部署时，CAPO 展现出卓越的零样本策略适应能力。

本文其余部分安排如下。第二节回顾视觉运动策略学习、具身智能体的策略适应以及提示学习方面的相关工作。第三节介绍问题表述和预备知识。第四节详细描述所提出的 CAPO 方法。第五节概述实验设置。第六节报告模拟环境中的实验结果，并讨论消融研究。最后，第七节总结全文并探讨未来研究方向。

II. 相关工作

本节将相关工作分为三个主要领域：视觉运动策略学习、具身智能体的策略适应以及提示学习。

A. 视觉运动策略学习

视觉运动策略学习是具身智能体的一项热门任务，其目标是将高维

视觉输入映射为低维、可执行的动作。现有方法大致可分为模仿学习和强化学习范式。模仿学习方法 [24]–[27]通常依赖于从真实世界环境中收集的大规模高质量数据集，在这些数据集中，深度神经网络经过训练以模仿专家示范，并近似像素到动作映射。例如，先前的研究利用逼真模拟器中的人类轨迹 [26], 事故数据集 [24], 以及远程操控机器人示范 [27] 来训练模仿类人导航策略的智能体。然而，收集此类数据集耗时、成本高昂，并且难以有效扩展。此外，模仿学习方法常常受到分布偏移的困扰 [28], 即使与专家分布存在细微偏差，也会导致误差累积，从而使模型难以适应未见环境。

为缓解这些局限，基于强化学习的方法日益受到关注 [7], [8], [29]–[31], ，这类方法大致可分为端到端和解耦范式。通过与环境交互并从试错反馈中学习，强化学习使智能体能够学习出更适应多样且动态场景的策略。端到端学习方法直接将视觉输入映射为动作指令。例如，Kaufmann 等人 [29] 开发了一种用于自主无人机竞速的深度强化学习框架，能够直接从原始像素输入中学习敏捷的视觉运动策略，并达到冠军级性能。Zhang 等人 [8] 通过将可微物理学融入强化学习训练循环，进一步提升了样本效率和控制精度。尽管取得了成功，这类方法通常仍受制于高得令人望而却步的样本复杂度，因为仅从稀疏的标量奖励中学习丰富的视觉表征一直被认为效率极低 [14]。此外，当具身或传感器配置发生变化时，这些耦合架构难以适应。

为缓解这种低效，另一条独立的研究路线主张采用解耦视觉运动学习 [13]–[15]。在该范式中，视觉编码器通常通过自监督目标进行预训练，随后被冻结，以为下游策略优化提供稳定的状态表征。尽管这些解耦架构通过将表征学习与控制动力学相隔离而显著提升了样本效率，但它们往往产生僵化且静态的表示。由于编码器对下游任务动力学并不感知，且在部署期间保持固定，当环境或具身配置发生剧烈变化时，它缺乏适应的灵活性。

与这些工作不同，我们的方法通过自适应编排机制，将视觉表征学习与策略优化衔接起来，从而有效融合解耦强化学习架构的稳定性与端到端优化的灵活性。我们旨在提取以域因素为条件的自适应视觉特征，并将其动态组合用于策略学习。该设计有助于提升样本效率，并能够在先前未见的视觉域或具身变化之间实现更好的适应。

B. 具身智能体的策略适应

对于具身智能体，策略适应技术旨在使学习到的策略适应域因素各异的环境早期方法 [9]–[11] 广泛采用域随机化进行端到端学习，以在仿真中合成视觉纹理和物理参数。这些方法让智能体接触各种随机化场景，试图促进可泛化行为，但通常需要大量手动调优，并且在特定场景中往往会采取过于保守的行为 [12]。

因此，越来越多的研究开始转向视觉表征学习的解耦范式，旨在离线提取域不变特征，以抵御分布偏移。当前的方法通常分为两类：学习通用视觉编码器 [16]–[18], 或利用大规模预训练模型 [19]–[21]。第一类方法涉及训练定制编码器，通常采用对比学习范式。其核心机制是通过对齐语义相关样本（正样本对）的嵌入，同时拉远不相关样本（负样本对）的嵌入，优化潜在空间，从而增强表征的判别能力。另一方面，近期研究尝试直接利用大规模预训练模型（例如 CLIP[23]）丰富的语义先验。与从头训练定制编码器相比，这些大型模型具有更强的零样本鲁棒性和数据效率。例如，Zhang 等人 [21] 利用 CLIP 作为目标检索模块，将智能体观测与人类语言指令对齐。这种对齐使智能体能够在复杂、非结构化的环境中识别与任务相关的目标，从而有效支持开放词汇导航任务，而无需进行大量特定任务的微调。

尽管这两种范式都很有效，但在动态适应方面均存在局限。自定义编码器和冻结的预训练骨干网络通常会产生静态表征 [22],, 而在策略学习过程中，无论遇到何种域偏移，其特征提取过程都保持不变。相比之下，我们的方法融合了两种方法各自的优势。在利用大规模预训练模型的稳健基础之上，我们引入了通过对比学习进行优化的可学习提示。这种设计使我们的系统能够自适应地生成表示，从而弥补静态编码器在面对显著域差距时适应能力不足的问题。

C. 提示学习

提示学习通过引入轻量、可学习的提示，为全模型微调提供了一种参数高效的替代方案，使大型预训练模型能够适应下游任务。最初，人们在语言模型中探索了这一方法，通过优化一小组文本提示向量 [32]。此后，这一方法也扩展到视觉Transformer：在 Transformer 层中插入可学习提示，无需修改骨干网络权重即可实现广泛的任务适配 [33], , 这被称为视觉提示。近年来，这一概念进一步拓展至多模态嵌入空间中的视觉-文本提示，研究表明提示能够编码互补的任务特定线索 [34], [35]。例如，Khattak 等人 [34]提出了多模态提示学习，在视觉和语言

分支中引入提示，以确保二者相互协同并改善多模态表示之间的对齐。同样，Yang等人[35] 研究了用于实例级感知的细粒度视觉提示，并设计了一种模糊反向掩码策略，在保留空间连贯性的同时抑制背景噪声，从而增强零样本理解。尽管这些方法在静态视觉任务中取得了成功，但提示学习在具身智能体中的应用仍缺乏充分研究。

本文采用视觉提示技术，明确应对具身智能体面临的感知挑战，从而弥合这一差距。不同于以往工作，我们利用多样化的混合对比提示，从根本上重构表示空间，使复杂跨具身视觉运动策略学习场景中的零样本适应成为可能。

III. 预备知识与问题表述

本节介绍 CAPO 的预备知识，包括问题表述以及旨在促进视觉运动策略训练的对比学习目标。

A. 问题表述

我们将具身智能体的视觉运动任务表述为部分可观测马尔可夫决策过程（POMDP），其定义为一个元组 $\mathcal{M} = (\mathcal{S}, \mathcal{O}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \gamma)$。其中，$\mathcal{S}$ 表示环境的底层状态空间，智能体无法直接访问该空间。$\mathcal{O}$ 表示高维观测空间，智能体在每个时间步 t 接收自我中心视觉观测 $o_t \in \mathcal{O}$，该观测由给定隐藏状态 $s_t \in \mathcal{S}$ 时的条件观测概率 $\mathcal{P}(o_t \mid s_t)$ 决定。动作空间 $\mathcal{A}$ 由智能体可执行的导航指令构成。转移动态 $\mathcal{T}(s_{t+1} \mid s_t, a_t)$ 描述在给定当前状态和已执行动作 $a_t \in \mathcal{A}$ 的情况下环境状态的演变。目标由奖励函数 $\mathcal{R}: \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ 定义，该函数提供反映任务进展的标量反馈 r_t，而 $\gamma \in [0, 1)$ 是折扣因子。

不同于从头开始训练视觉编码器 f_θ 的主流端到端方法 [7], [30], [31], ，我们利用 CLIP 作为图像编码器（记为 Φ），将视觉表征学习与策略优化解耦。为了使这一固定骨干网络适应具身导航中的动态域因素，我们引入一组 K 可学习视觉提示，记为 $\mathbf{P} = \{\boldsymbol{p}^1, \boldsymbol{p}^2, \ldots, \boldsymbol{p}^K\}$。每个提示 $\boldsymbol{p}^k$ 均定义为一系列连续的可学习向量：

$$\boldsymbol{p}^k = [\boldsymbol{v}_1^k, \boldsymbol{v}_2^k, \ldots, \boldsymbol{v}_L^k], \quad \boldsymbol{v}_i^k \in \mathbb{R}^D, \tag{1}$$

其中，L 表示提示长度，D 表示嵌入维度。这些提示会被添加到输入观测 o_t 之前，并由编码器 Φ 处理，以提取具备上下文感知能力的视觉特征。

为了从这些多个提示嵌入中构建最终的状态表示，我们采用基于注意力的提示编排机制 $\mathcal{G}_{\text{attn}}$。令 $\mathbf{z}_t^v = \Phi(o_t, \varnothing)$ 表示使用冻结的 CLIP 编码器 Φ 且不含提示时得到的原始嵌入，$\boldsymbol{z}_t^k = \Phi(o_t, \boldsymbol{p}^k)$ 表示由第 k 个提示导出的特征嵌入。最终融合特征表示 $\boldsymbol{z}_t^f$ 通过对这些嵌入进行加权组合计算得出：

$$\boldsymbol{z}_t^f = \mathbf{z}_t^v + \mathcal{G}_{\text{attn}}(\{\boldsymbol{z}_t^1, \ldots, \boldsymbol{z}_t^K\}), \tag{2}$$

其中，注意力权重根据当前视觉观测动态确定（详情将在下一节介绍）。随后，这一融合表示 $\boldsymbol{z}_t^f$ 作为策略网络 π_ψ 的输入，该网络由 ψ 参数化。智能体的目标是学习最优策略 π_ψ^*，以最大化期望折扣累积奖励：$J(\pi_\psi) = \mathbb{E}_{\pi_\psi}[\sum_{t=0}^{T} \gamma^t r_t]$。

B. 对比学习

对比学习已成为自监督表示学习中的主流范式，这对于有效的策略适应至关重要。其目标是通过最大化语义相似实例之间的互信息，提取判别特征。它不依赖人工标注，而是将实例判别视为字典查找任务。具体而言，给定一批采样得到的观测，我们使用查询编码器 f_q 和键编码器 f_k 将输入映射到共享的潜在空间，生成特征向量 q 和 k。训练目标是优化正样本对（q_i，k_i）之间的对齐，同时使其与负样本形成对比。因此，通过最小化对比损失 [13]来进行优化：

$$\mathcal{L}_{\text{NCE}} = -\mathbb{E}\left[\log \frac{\text{sim}(q_i, k_i)}{\sum_{j=1}^{N} \text{sim}(q_i, k_j)}\right], \tag{3}$$

其中，相似度函数定义为 $\text{sim}(u, v) = \exp(u^T v)$，其中 u 和 v 分别表示 ℓ_2 归一化的特征向量。这里，我们通过采用混合对比策略（见第IV-B节），将这一表述扩展到导航任务。

IV. 方法

本节详细介绍所提出的对比提示编排（CAPO）方法，包括对比提示学习、混合对比学习、自适应提示编排以及视觉运动策略学习。

A. 概述

我们提出 CAPO，这是一种面向具身智能体的新型视觉运动策略学习方法，旨在实现对未见领域因素和跨具身配置的零样本适应。如图 2 所示，整体训练流程分为两个紧密协同的阶段：（a）对比提示学习；（b）自适应提示编排。在第 IV-B 节所示的提示学习阶段，我们采用预训练 CLIP 模型 [23]作为视觉特征提取器；我们不对整个骨干网络进行微调，而是冻结其图像编码器，同时引入一组可学习视觉提示。提示学习通过一种混合对比学习策略实现，该策略整合了视觉、时序动作和文本对比学习目标。这种混合对比学习形式构成了一个多样化的提示池。在第 IV-C 节介绍的提示编排阶段，我们

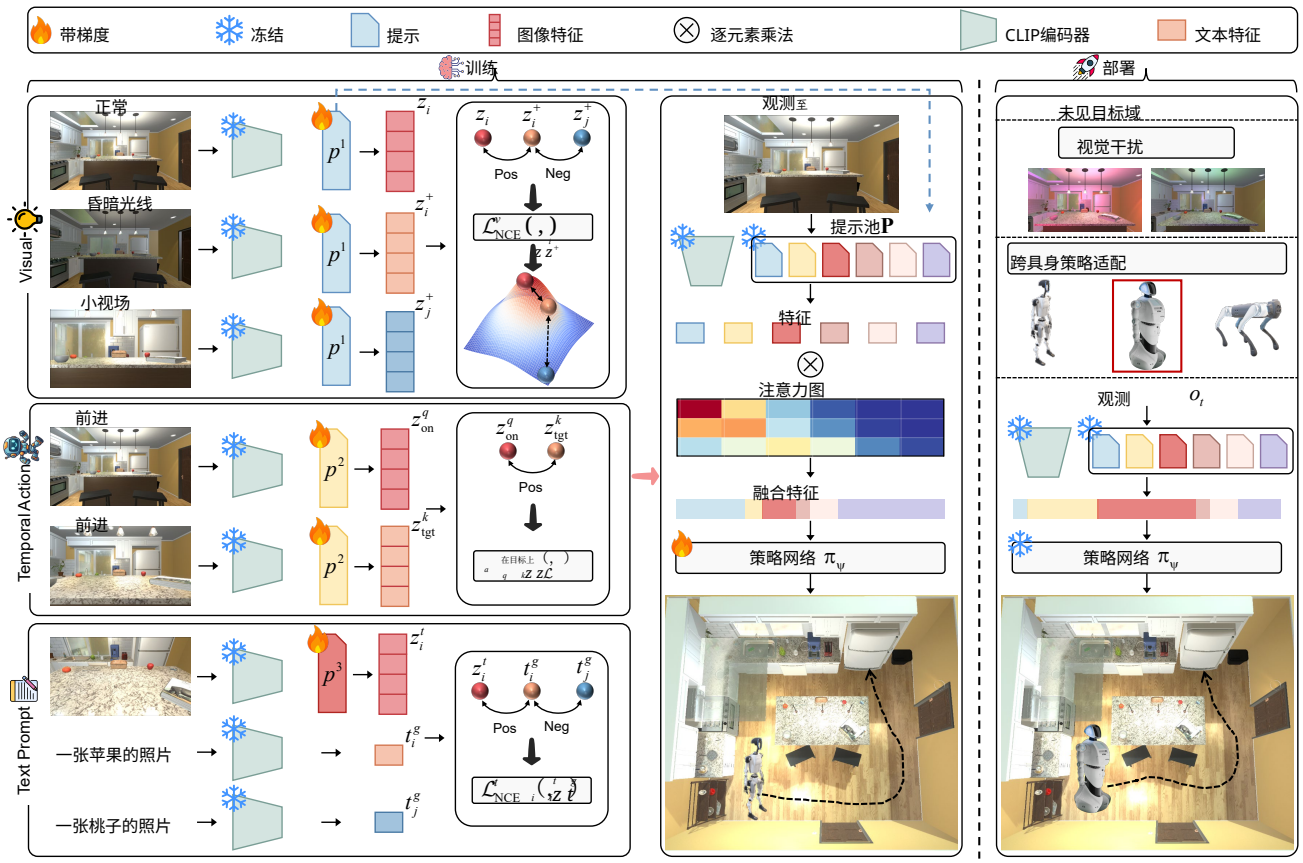


图 2：对比提示编排（CAPO）框架。(a) 对比提示学习。我们采用混合对比学习策略（结合视觉、时序动作和文本目标），为冻结的 CLIP 编码器 Φ学习一组特定领域的视觉提示$\mathbf{P}$。(b) 自适应提示编排。我们引入一种自适应提示编排机制，根据观测 o_t动态聚合这些冻结的提示，为策略网络π_ψ生成上下文感知特征。训练后的策略能够在未见领域实现零样本跨具身适应：通过自动调整提示权重来应对视觉干扰和具身变化，而无需微调。

冻结学习得到的提示池，并引入一种与策略学习相结合的自适应提示编排机制。给定当前观测，该机制通过自适应地为提示条件特征加权，动态构建状态表示。随后，将所得表示输入策略网络，并通过强化学习算法训练该网络以执行视觉运动任务。训练完成后，学习得到的策略可直接部署，通过根据观测动态重新加权提示诱导嵌入，实现零样本跨具身视觉运动适应。因此，策略能够调整融合后的视觉表征，以适应领域特定变化，而无需额外微调。完整算法及其理论分析概要见第 IV-D 节。

B. 对比提示学习

这里，我们详细介绍方法中的提示学习阶段，该阶段通过混合对比学习学习特定领域视觉提示。此阶段会生成一组视觉提示$\mathbf{P} = \{\boldsymbol{p}^1, \boldsymbol{p}^2, \ldots, \boldsymbol{p}^K\}$，每个提示都专门用于实现对某一特定域因素的不变性。这些提示随后会被冻结，并用于自适应提示编排阶段。我们从三个方面描述这一过程：数据收集、视觉提示学习和混合对比学习。这里，对比提示学习侧重于优化提示参数，而混合对比学习则作为训练目标，通过计算由不同提示生成的视觉嵌入之间的对比损失来实现。

数据收集。我们通过在 AI2-THOR 仿真环境 [36] 中，针对多种域因素和具身配置执行专家导航策略，来收集训练数据集 $\mathcal{D}$ 。每个域因素 $\mathcal{F}_i$ 均由一组环境参数和具身参数构成：

$$\mathcal{F}_i = (\phi_i, \psi_i, \theta_i, \delta_i, \ell_i^b, \ell_i^c, \ell_i^s, \ell_i^h), \tag{4}$$

其中，ϕ_i 表示水平视场参数，ψ_i 表示旋转角度，θ_i 表示向上/向下视角，δ_i 表示平移步长，而 ℓ_i^b、ℓ_i^c、ℓ_i^s、ℓ_i^h则表示控制亮度、对比度、饱和度和色调的光照参数。我们系统地改变这些因素，从而在多样的视觉条件和具身条件下生成轨迹。

具体而言，对于每个配置 $\mathcal{F}_i$，都会部署一个预训练专家策略 π_e ，以导航至随机采样的

目标物体。每个导航回合都会产生一条按时间排序的轨迹 $\tau_i^j = \{ o_t^{ij}\ a_t^{ij}\ r_t^{ij}\ g^j \}_{t=1}^{T_{ij}}$ $o_t^{ij} \in (\ ,\ ,\ ,\ ,\ )$，其 $\mathbb{R}^{C\times H\times W}$ t a^{ij} 中，表示时间步 的 RGB观测， $\mathcal{A}$ $_t{}^{j} \in$ $r_t^{ij} \in \mathbb{R}$ 表示执行的动作，指定奖励值，而 $g^j \in \mathcal{C}$ 则是与第 j 条轨迹关联的目标类别。

为促进跨域对比提示学习，我们在共享相同导航目标、于不同因素下采集的轨迹之间建立对应关系。给定参考因素 $\mathcal{F}_{\mathrm{b}}$ 与替代因素 $\mathcal{F}_i$，我们识别对应于同一目标物体 g^j 的对齐轨迹对（$\tau^j{}_{\mathrm{b}}\ \tau_i^j$）。对齐依据初始观测的语义相似度确定。对于每条轨迹，我们按如下方式计算物体存在向量：

$$\mathbf{m} = (m_1, m_2, \ldots, m_{|\mathcal{C}|}) \in \{0,1\}^{|\mathcal{C}|}, \tag{5}$$

其中每个元素表示智能体初始语义观测中相应物体类别是否存在。若两条轨迹的物体存在向量之间的F1分数超过预定义阈值κ，则认为它们对齐，从而确保配对轨迹来自不同域因素下语义相似的状态。该对齐过程产生一个按如下方式组织的精炼数据集

$$\mathcal{D} = \bigcup_{i=1}^{K} \mathcal{D}_i, \quad \mathcal{D}_i = \{\tau_i^1, \tau_i^2, \ldots, \tau_i^N\}, \tag{6}$$

其中每个子集 $\mathcal{D}_i$ 包含在同一域因素下采集的轨迹，且 N 为轨迹数量。

视觉提示学习。 我们利用预训练 CLIP 模型中蕴含的丰富语义先验 [23]从RGB观测中提取视觉特征。为以参数高效的方式将此预训练视觉编码器适配到特定域的视觉运动任务，我们不更新预训练模型的全部参数，而是引入一组可学习提示嵌入 [33] 这些嵌入在保持CLIP图像编码器冻结的同时调制视觉表示。

形式上，如前所述，我们定义一组视觉提示 $\mathbf{P} = \{\boldsymbol{p}^1, \boldsymbol{p}^2, \ldots, \boldsymbol{p}^K\}$，其中每个提示 $\boldsymbol{p}^k$ 都是由可学习向量 $\boldsymbol{p}^k = [\boldsymbol{v}_1^k, \boldsymbol{v}_2^k, \ldots, \boldsymbol{v}_L^k]$ 构成的序列$\boldsymbol{v}_i^k \in \mathbb{R}^D$。给定与第 k 个域因素对应的 RGB 观测 $o_t \in \mathbb{R}^{C\times H\times W}$，CLIP 编码器 Φ首先对图像执行图块嵌入操作，将其映射为由 N_p 个视觉词元组成的序列：

$$\mathbf{x}_{\text{patch}} = \text{PatchEmbed}(o_t) \in \mathbb{R}^{N_p\times d}, \tag{7}$$

其中 d 表示隐藏特征维度。将类别标记$\mathbf{x}_{\text{cls}} \in \mathbb{R}^{1\times d}$ 前置到图块序列中，并添加位置嵌入$\mathbf{E}_{\text{pos}} \in \mathbb{R}^{(N_p+1)\times d}$ 以纳入序列结构：

$$\mathbf{X}_0 = [\mathbf{x}_{\text{cls}}; \mathbf{x}_{\text{patch}}] + \mathbf{E}_{\text{pos}} \in \mathbb{R}^{(N_p+1)\times d}. \tag{8}$$

接着，我们将与当前域因素对应的提示嵌入 $\boldsymbol{p}^k$ 插入到类别标记之后的词元序列中。具体而言，首先通过可学习线性变换 $\mathbf{W}_p \in \mathbb{R}^{D\times d}$ 将提示词元从维度 D 投影到隐藏维度 d：

$$\tilde{\boldsymbol{p}}^k = \boldsymbol{p}^k \mathbf{W}_p \in \mathbb{R}^{L\times d}, \tag{9}$$

然后将其与视觉词元拼接，得到增强输入序列：

$$\mathbf{X}_0^{\mathcal{P}} = [\mathbf{x}_{\text{cls}}; \tilde{\boldsymbol{p}}^k; \mathbf{x}_{\text{patch}}] + \mathbf{E}_{\text{pos}}^{\mathcal{P}} \in \mathbb{R}^{(1+L+N_p)\times d}, \tag{10}$$

其中$\mathbf{E}_{\text{pos}}^{\mathcal{P}}$ 表示扩展后的位置嵌入，以容纳插入的提示长度 L。

增强后的词元序列由 Trans- former 编码器 ξ 处理，该编码器由 J 层组成，并生成一系列表示：

$$\mathbf{X}_j^{\mathcal{P}} = \xi_j(\mathbf{X}_{j-1}^{\mathcal{P}}), \quad j = 1, \ldots, J. \tag{11}$$

最后，从类别标记位置提取输出表示，经由 CLIP 投影头$\mathbf{W}_{\text{proj}} \in \mathbb{R}^{d\times d_{\text{out}}}$进行投影，并进行 ℓ_2归一化，得到适应域的视觉嵌入：

$$\mathbf{z}_t^k = \frac{\mathbf{X}_J^{\mathcal{P}}[0]\mathbf{W}_{\text{proj}}}{\|\mathbf{X}_J^{\mathcal{P}}[0]\mathbf{W}_{\text{proj}}\|_2} \in \mathbb{R}^{d_{\text{out}}}. \tag{12}$$

嵌入$\mathbf{z}_t^k$ 作为后续混合对比学习的输入。具体而言，它用于计算对比损失，其梯度会反向传播以更新相应的提示嵌入 $\boldsymbol{p}^k$。通过这一优化过程，提示逐步捕获跨域因素的任务相关不变性，形成一组稳健的基表示。

混合对比学习。 为了构建能够应对多种域变化的多样化提示池$\mathbf{P}$ ，我们提出一种混合对比学习策略，包含三个不同的目标，分别是视觉对比学习、时间动作对比学习和文本对比学习。

(i) 视觉对比学习。 其旨在使具身智能体能够抵御视觉外观变化，例如亮度、对比度、饱和度和色调的变化；这些变化不同于物理具身变化。具体而言，我们优化与光照域因素相关的提示子集，记为 $\{\boldsymbol{p}^{\ell_i^b}, \boldsymbol{p}^{\ell_i^c}, \boldsymbol{p}^{\ell_i^s}, \boldsymbol{p}^{\ell_i^h}\}$，以便在视觉扰动下提取不变表示。这里， $\ell_i^b, \ell_i^c, \ell_i^s, \ell_i^h$表示连续向量，用于表征第 i个域因素的光照配置。

对于与特定光照因素（例如亮度） $\boldsymbol{p}^{\ell_i^b}$ 相对应的提示，我们展示对比样本对的构建过程；相同的步骤也适用于其他光照因素ℓ_i^c， ℓ_i^s， ℓ_i^h。给定一个锚点观测 o_i，从数据集中 $\mathcal{D}$采样后，我们根据$o_i^+ = \mathcal{V}_{\ell_i^b}(o_i)$ 的参数空间对图像进行扰动，从而生成一个增强视图 ℓ_i^b。随后，我们使用以提示 Φ 为条件的冻结 CLIP 编码器，根据公式（7）–（12）提取锚点观测和正样本的特征。提取的特征可表示如下：

$$\boldsymbol{z}_i = \Phi(o_i, \boldsymbol{p}^{\ell_i^b}), \quad \boldsymbol{z}_i^+ = \Phi(o_i^+, \boldsymbol{p}^{\ell_i^b}), \tag{13}$$

其中， $\boldsymbol{z}_i$ 和 $\boldsymbol{z}_i^+$ 分别表示原始观测 o_i 及其增强视图 o_i^+ 的视觉嵌入，二者均使用相同的提示 $\boldsymbol{p}^{\ell_i^b}$ 进行编码。

为优化提示，我们采用对称 InfoNCE（信息噪声对比估计）目标，遵循式（3），并将增强视图视为正键

并将批次中的其他样本 $\mathcal{B}$ 视为负样本。对称 InfoNCE 损失通过对两个方向取平均来定义：

$$\mathcal{L}_{\text{NCE}}^{v} = \frac{1}{2}\mathbb{E}_{i\in\mathcal{B}}\left[\mathcal{L}_{\text{NCE}}^{\text{uni}}(\boldsymbol{z}_i, \boldsymbol{z}_i^{+}) + \mathcal{L}_{\text{NCE}}^{\text{uni}}(\boldsymbol{z}_i^{+}, \boldsymbol{z}_i)\right], \tag{14}$$

其中，单向 InfoNCE 损失定义为

$$\mathcal{L}_{\text{NCE}}^{\text{uni}}(\boldsymbol{z}_i, \boldsymbol{z}_i^{+}) = -\log \frac{\text{sim}(\boldsymbol{z}_i, \boldsymbol{z}_i^{+})}{\sum_{j=1}^{|\mathcal{B}|}\text{sim}(\boldsymbol{z}_i, \boldsymbol{z}_j^{+})}. \tag{15}$$

此外，为了实现更紧密的特征对齐，我们加入均方误差（MSE）正则项：

$$\mathcal{L}_{\text{MSE}}^{v} = \frac{1}{|\mathcal{B}|}\sum_{i=1}^{|\mathcal{B}|}\|\boldsymbol{z}_i - \boldsymbol{z}_i^{+}\|_2^2. \tag{16}$$

提示 $\boldsymbol{p}^{\ell_i^b}$ 的最终视觉对比目标是加权和：

$$\mathcal{L}_{\text{visual}} = \mathcal{L}_{\text{NCE}}^{v} + \lambda_v \mathcal{L}_{\text{MSE}}^{v}, \tag{17}$$

其中，λ_v 是一个缩放超参数，用于平衡对齐正则化强度。

（ii）时序动作对比学习。该目标旨在学习智能体配置参数变化下的行为不变性，具体包括视场、旋转角度、俯仰角和步长。因此，我们优化与动作执行因素相关的提示子集，记为 $\{\boldsymbol{p}^{u_i} \mid u_i \in \{\phi_i, \psi_i, \theta_i, \delta_i\}\}$，其中这些变量对应于上述具身参数。不同于依赖负样本的视觉对比学习方法，我们采用 BYOL 自蒸馏范式 [37]，以避免在时序动作空间中定义语义上有意义的负样本这一难题。

对于特定的动作条件提示 $\boldsymbol{p}^{u_i}$，我们从 $\mathcal{D}$ 中采样正样本对（o_q，o_k），这些样本对应相同的动作类型，但观测时所处的时间上下文和具身配置不同。为引入视角多样性，我们应用随机图像增强，生成增强视图 $\tilde{o}_q$ 和 $\tilde{o}_k$。该学习框架由两个并行网络组成：由 ω 参数化的在线网络，以及由 ν参数化的目标网络。在线网络包含三个组件：提示编码器Φ、投影器 ρ_ω和预测器 f_ω。目标网络采用与编码器和投影器相同的架构，但不包含预测器。目标网络参数 ν 根据在线网络参数 ω，按照动量更新规则 [13] 进行更新：

$$\nu \leftarrow \beta\nu + (1-\beta)\omega, \tag{18}$$

其中，$\beta \in [0,1]$ 是衰减率。

我们提取查询视图和键视图的投影表示。在线网络预测跨视图样本的目标表示。具体而言，令 $\mathbf{z}_{\text{on}}^{q} = f_\omega(\rho_\omega(\Phi(\tilde{o}_q, \boldsymbol{p}^{u_i})))$ 为输入查询视图时的在线预测，$\mathbf{z}_{\text{tgt}}^{k} = \text{SG}(h_\nu(\Phi(\tilde{o}_k, \boldsymbol{p}^{u_i})))$ 为键视图的目标投影，其中 SG(·) 表示停止梯度算子。非对称回归损失定义为归一化预测与目标之间的均方误差（MSE）：

$$\mathcal{L}^{a}(\tilde{o}_q, \tilde{o}_k) = \left\| \frac{\mathbf{z}_{\text{on}}^{q}}{\|\mathbf{z}_{\text{on}}^{q}\|_2} - \frac{\mathbf{z}_{\text{tgt}}^{k}}{\|\mathbf{z}_{\text{tgt}}^{k}\|_2} \right\|_2^2. \tag{19}$$

为强制对称，我们显式地沿两个方向计算损失：其一是在线网络处理 $\tilde{o}_q$ 以预测 $\tilde{o}_k$ 的目标表示，其二是交换二者角色（即在线网络处理 $\tilde{o}_k$）：

$$\mathcal{L}_{\text{action}} = \frac{1}{2}\mathbb{E}_{(\tilde{o}_q, \tilde{o}_k)\sim\mathcal{B}}\left[\mathcal{L}^{a}(\tilde{o}_q, \tilde{o}_k) + \mathcal{L}^{a}(\tilde{o}_k, \tilde{o}_q)\right], \tag{20}$$

这一对称目标鼓励提示 $\boldsymbol{p}^{u_i}$ 捕获一致的动作相关特征，使其不受时间偏移和增强噪声的影响。

(iii) 文本对比学习。尽管上述视觉和时序动作目标解决了环境和具身变化问题，但它们并未显式地强制实现视觉运动任务所需的语义对齐。为弥合视觉观测与语义目标规范之间的模态差距，我们引入文本对比学习目标。我们定义了一个独立的可学习文本提示 $\boldsymbol{p}^{t}$，其结构与视觉提示相同，用于调节冻结的 CLIP 编码器 Φ，以提取具有语义基础的特征。与适应环境变化的特定领域提示不同，$\boldsymbol{p}^{t}$ 旨在捕获与域偏移正交的任务级语义不变性。

具体而言，对于每个导航目标类别 $g \in \mathcal{C}$，我们使用模板“a photo of a [CLASS]”生成文本描述，并通过冻结的 CLIP文本编码器 Ψ对其进行编码，以获得类别特定文本嵌入$\boldsymbol{t}_i^{g} = \Psi(\text{template}(g)) \in \mathbb{R}^{d_{\text{out}}}$。这些嵌入预先计算并固定下来，作为语义锚点。给定一批观测 $\mathcal{B} = \{(o_i, g_i)\}_{i=1}^{|\mathcal{B}|}$，我们提取文本提示视觉嵌入 $\boldsymbol{z}_i^{t} = \Phi(o_i, \boldsymbol{p}^{t})$。此外，为了增强所学习表示的适应能力，我们引入了一种语义正则化机制。具体而言，我们向视觉嵌入 $\boldsymbol{z}_i^{t}$ 注入高斯噪声，以获得扰动表示：

$$\tilde{\boldsymbol{z}}_i^{t} = \boldsymbol{z}_i^{t} + \boldsymbol{\zeta}, \quad \boldsymbol{\zeta} \sim \mathcal{N}(0, \sigma^2\mathbf{I}). \tag{21}$$

我们强制要求该扰动表示与相应文本锚点保持较高的语义相似度 $\boldsymbol{t}_i^{g}$。这一策略可防止特征坍缩，并确保模型对特征级扰动具有鲁棒性。随后，我们按照公式（3）最大化视觉嵌入与相应文本嵌入之间的互信息：

$$\mathcal{L}_{\text{NCE}}^{t}(\tilde{\boldsymbol{z}}_i^{t}, \boldsymbol{t}_i^{g}) = -\log \frac{\text{sim}(\tilde{\boldsymbol{z}}_i^{t}, \boldsymbol{t}_i^{g})}{\sum_{j=1}^{|\mathcal{B}|}\text{sim}(\tilde{\boldsymbol{z}}_i^{t}, \boldsymbol{t}_j^{g})}. \tag{22}$$

与视觉对比学习类似，我们还在匹配的图像-文本对之间加入了 MSE正则化项：

$$\mathcal{L}_{\text{MSE}}^{t} = \frac{1}{|\mathcal{B}|}\sum_{i=1}^{|\mathcal{B}|}\|\tilde{\boldsymbol{z}}_i^{t} - \boldsymbol{t}_i^{g}\|_2^2. \tag{23}$$

最终的文本目标是加权和：

$$\mathcal{L}_{\text{text}} = \mathcal{L}_{\text{NCE}}^{t} + \lambda_t \mathcal{L}_{\text{MSE}}^{t}, \tag{24}$$

其中 λ_t 用于平衡正则化强度。通过在优化该正则化的同时优化式（24），$\boldsymbol{p}^{t}$ 学习引导视觉编码器突出目标相关语义，从而提供稳定的接地信号，以补充自适应的领域特定特征。

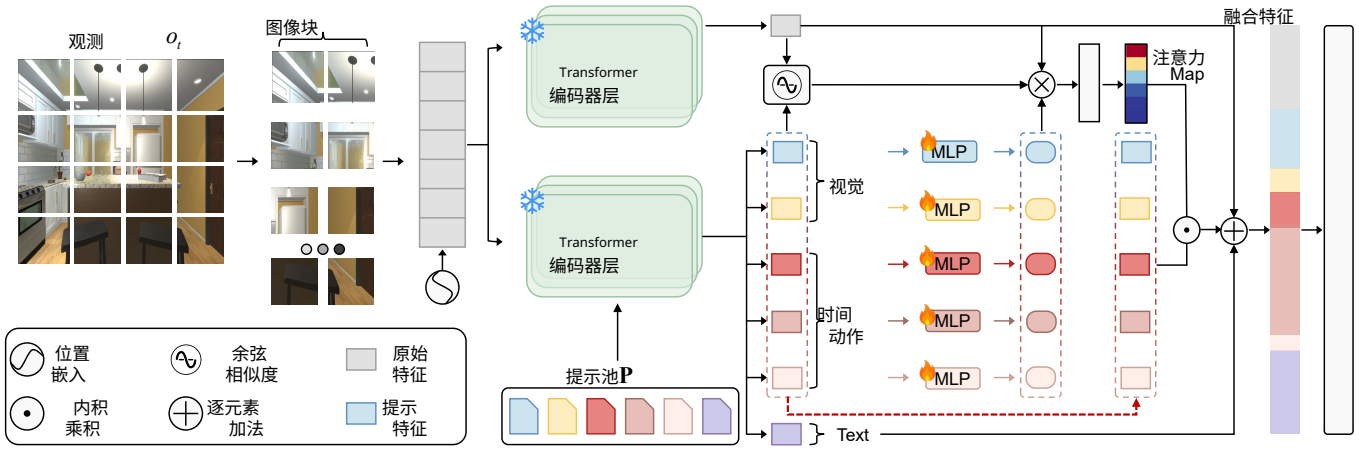


图3：自适应提示编排机制的架构。冻结的 CLIP 编码器使用提示池 $\mathbf{P}$从观测 o_t 中提取特征。双分支注意力机制（结合可学习的MLP投影和余弦相似度）根据上下文动态重新加权领域特定特征。关键在于，文本提示特征绕过此注意力模块，直接进行融合。该设计在将特征输入策略网络π_ψ之前，明确保留了不受域因素影响的目标导向语义。

因此，通过联合优化这一混合对比学习策略，我们得到一组领域特定的视觉提示，记为$\mathbf{P} = \{\boldsymbol{p}^1, \boldsymbol{p}^2, \ldots, \boldsymbol{p}^K\}$。每个提示 $\boldsymbol{p}^k$ 都经过优化，以表征环境或具身因素方面的不同不变性。随后，该学习得到的提示池以冻结状态用于促进策略学习。

C. 自适应提示编排

在本小节中，我们详细介绍方法中的自适应提示编排阶段：冻结的领域特定视觉提示$\mathbf{P} = \{\boldsymbol{p}^1, \boldsymbol{p}^2, \ldots, \boldsymbol{p}^K\}$经过自适应组合，并用于训练视觉运动策略。我们从两个方面描述这一过程：双分支注意力机制和策略学习。

双分支注意力机制。 虽然每个提示$\boldsymbol{p}^k$都擅长捕捉单个域因素特有的不变性，但稳健的具身视觉运动策略需要能够同时适应多种不断变化的条件。简单地聚合提示特征会忽略不同域因素对于给定观测的相关性差异。为此，我们提出一种自适应提示编排机制，记为 $\mathcal{G}_{\text{attn}}$，根据语义相关性动态调整提示的权重，如图 3所示。

给定时间步 t处的RGB观测 o_t ，我们首先使用不带提示的冻结的 CLIP 编码器 Φ 提取一个普通嵌入，记为$\mathbf{z}_t^v = \Phi(o_t, \varnothing)$。同时，我们利用学习得到的提示获得一组领域自适应嵌入 $\{\mathbf{z}_t^k\}_{k=1}^K$ ，其中$\mathbf{z}_t^k = \Phi(o_t, \boldsymbol{p}^k)$。值得注意的是，文本提示嵌入$\boldsymbol{z}_t^t = \Phi(o_t, \boldsymbol{p}^t)$被排除在编排机制之外，因为它编码了任务级语义不变性。为了融合这些表示，我们采用由可学习投影和语义一致性共同驱动的双引导注意力机制。为了融合这些表示，我们采用由可学习投影和语义一致性共同驱动的双引导注意力机制。

一方面，我们通过将嵌入投影为注意力键k_k，为每个提示计算可学习注意力分数 s_k^a：

$$s_k^a = \frac{(\mathbf{z}_t^v)^\top k_k}{\sqrt{d_{\text{out}}}}, \quad k_k = f_p(\mathbf{z}_t^k), \tag{25}$$

其中， $f_p(\cdot)$ 是一个由两个线性层和一个 SiLU激活 [38]，组成的投影网络， d_{out} 表示带有提示的 Φ的输出维度。

另一方面，为了防止编排机制关注无关特征，我们引入一个基于经ℓ_2归一化的原始嵌入与提示嵌入之间余弦相似度的正则化项，以确保语义保真。

$$s_k^c = (\mathbf{z}_t^v)^\top \mathbf{z}_t^k. \tag{26}$$

最终的注意力权重 α_k 通过对这些分数进行乘法融合，再经过 softmax归一化得到：

$$\alpha_k = \frac{\exp(s_k^a \cdot s_k^c)}{\sum_{j=1}^K \exp(s_j^a \cdot s_j^c)}. \tag{27}$$

聚合上下文表示$\mathbf{z}_t^f$ 通过$\mathcal{G}_{\text{attn}}$ 计算为提示嵌入的加权和，并借助与原始特征的残差连接加以增强，以保留原始语义信息：

$$\mathbf{z}_t^f = \mathbf{z}_t^v + \boldsymbol{z}_t^t + \mathcal{G}_{\text{attn}}(\{\boldsymbol{z}_t^1, \ldots, \boldsymbol{z}_t^K\}) = \mathbf{z}_t^v + \boldsymbol{z}_t^t + \sum_{k=1}^K \alpha_k \mathbf{z}_t^k. \tag{28}$$

该融合表示$\mathbf{z}_t^f$ 自适应地整合领域特定不变性，同时保持任务相关语义，从而作为策略学习的稳健输入。

策略学习。给定融合表示$\mathbf{z}_t^f$ ，我们通过融入任务特定的语义和历史信息，构建全面的观测嵌入 o_t 。具体而言，目标物体类别 $g \in \mathcal{C}$ 被编码为

算法 1： CAPO 训练流程

输入： 轨迹数据集 $\mathcal{D}$、冻结的 CLIP 编码器 Φ，可学习的提示池 $\mathbf{P}$

输出： 优化后的导航策略 π_ψ

// **对比提示学习**
对于每个迭代**执行**
  采样小批次 $\mathcal{B} \sim \mathcal{D}$
  // **视觉对比学习：**
  生成增强视图 o_i, o_i^+ 用于 $o_i \in \mathcal{B}$
  计算 $\mathcal{L}_{\text{visual}}$ 通过公式（17）
  $\mathbf{P} \leftarrow \mathbf{P} - \eta_v \nabla_{\mathbf{P}} \mathcal{L}_{\text{visual}}$
  // **时序动作对比学习：**
  从 $\mathcal{B}$采样（o_q， o_k）
  生成增强视图（$\tilde{o}_q$， $\tilde{o}_k$）
  计算 $\mathbf{z}_{\text{on}}^q = f_\omega(\rho_\omega(\Phi(\tilde{o}_q, \boldsymbol{p}^{u_i})))$
  计算 $\mathbf{z}_{\text{tgt}}^k = \text{SG}(h_\nu(\Phi(\tilde{o}_k, \boldsymbol{p}^{u_i})))$
  通过公式（ $\mathcal{L}_{\text{action}}$ 20）计算
  $\mathbf{P} \leftarrow \mathbf{P} - \eta_a \nabla_{\mathbf{P}} \mathcal{L}_{\text{action}}$， $\omega \leftarrow \omega - \eta_a \nabla_\omega \mathcal{L}_{\text{action}}$
  $\nu \leftarrow \beta\nu + (1-\beta)\omega$
  // **文本对比学习：**
  从 $\mathcal{B}$ 采样 (o_i, g_i)
  计算 $\mathcal{L}_{\text{text}}$ 通过公式（24）
  $\mathbf{P} \leftarrow \mathbf{P} - \eta_t \nabla_{\mathbf{P}} \mathcal{L}_{\text{text}}$
结束

// **自适应提示编排**
冻结 $\mathbf{P}$ 和 Φ，初始化 $\mathcal{G}_{\text{attn}}$ 和 π_ψ
对于每个迭代 **执行**
  采样展开轨迹 $\tau = \{o_t, a_t, r_t, g\}_{t=1}^T$
  计算 $\boldsymbol{z}_t^v = \Phi(o_t, \varnothing)$、 $\boldsymbol{z}_t^t = \Phi(o_t, \boldsymbol{p}^t)$、 $\{\boldsymbol{z}_t^k\}_{k=1}^K$
  计算 $\boldsymbol{z}_t^f$ 通过公式（28）
  计算 $\hat{o}_t = [\boldsymbol{z}_t^f, \mathbf{g}, \mathbf{e}_{a_{t-1}}]$， $h_t = \text{GRU}(\hat{o}_t, h_{t-1})$
  输出动作 $a_t \sim \pi_\psi(\cdot|h_t)$
  计算近端策略优化目标 $\mathcal{L}_{\text{PPO}}$ 通过公式（29）
  $\psi \leftarrow \psi - \eta_\psi \nabla_\psi \mathcal{L}_{\text{PPO}}$
结束

独热向量$\mathbf{g}$，并将上一个时间步$a_{t-1} \in \mathcal{A}$ 采取的动作映射为可学习嵌入$\mathbf{e}_{a_{t-1}}$。这些组成部分与视觉特征拼接，构成策略输入：$\hat{o}_t = [\mathbf{z}_t^f, \mathbf{g}, \mathbf{e}_{a_{t-1}}]$。

为缓解具身导航中固有的部分可观测性，我们采用门控循环单元（GRU） [39]来建模时序依赖。GRU 将当前观测嵌入 $\hat{o}_t$ 与前一隐藏状态h_{t-1} 一同处理，以生成当前隐藏状态 h_t，该状态作为演员-评论家架构的共享表示。演员头预测动作分布 $\pi_\psi(a_t|h_t)$，而评论家头估计状态价值 $V(h_t)$。我们使用近端策略优化算法（ PPO ） [40]，优化策略 π_ψ ，该算法通过以下目标训练整个网络架构（包括 GRU 和嵌入）：

$$\mathcal{L}_{\text{PPO}} = -\mathbb{E}_t \left[\min \left(r_t \hat{A}_t,\ \text{clip}(r_t, 1-\epsilon, 1+\epsilon)\hat{A}_t \right) \right], \quad (29)$$

其中， r_t 表示当前策略与先前策略之间的概率比率， A_t 是优势估计在时间步 t， ϵ是裁剪超参数。

D. 总结与理论论证

算法概要。我们在算法1中总结了 CAPO 的整体训练流程。在对比提示学习中，智能体通过混合对比学习策略学习一组多样化的特定领域视觉提示，从而捕捉视觉、时间和语义维度上的不变性。在自适应提示编排中，这些学习得到的提示会被冻结，并使用自适应提示编排机制优化视觉运动策略，使其能够动态适应未见域因素。

理论论证。CAPO 的核心目标是构建视觉运动策略 π_ψ ，使其在域因素分布 $\mathcal{F}$ 上的期望风险最小化。我们假设，对于任意上下文 $c \in \mathcal{F}$，都存在一个最优预言机表示 $\boldsymbol{z}_c^\star$。我们的目标是通过自适应提示编排机制 $\mathcal{G}_{\text{attn}}$ 对该预言机表示进行近似。

令 $R_E(\pi) = \mathbb{E}_c[\mathcal{L}_{\text{task}}(\pi(\boldsymbol{z}^f(c)), y_c^\star)]$ 表示期望风险，其中 $\mathcal{L}_{\text{task}}$ 表示任务特定损失函数（例如近端策略优化目标），$y_c^\star$ 表示给定上下文 c 下的最优动作或价值。令 $R^\star$ 表示预言机可达到的最小风险。我们将超额风险 定义为学习得到的策略与预言机性能之间的差异。

定理1. 设 $\mathcal{H}_{\mathbf{P}}$ 为由学习得到的提示池$\mathbf{P}$的凸包张成的假设空间。假设任务损失是 L_ℓ-利普希茨的，且策略是 λ_π-利普希茨的，则以至少 $1-\delta$的概率，超额风险的上界为：

$$R_E(\hat{\pi}) - R^\star \le \tilde{O}\left(\sqrt{\frac{1}{N}}\right) + L_\ell \lambda_\pi \cdot \mathbb{E}_c \left[\min_{\alpha \in \Delta_K} \left\| \sum_{k=1}^K \alpha_k \boldsymbol{z}^k - (\boldsymbol{z}_c^\star - \boldsymbol{z}^v) \right\|_2 \right]. \quad (30)$$

证明. 我们将超额风险分解为估计误差和近似误差，并同时对它们进行界定 [41]。假设损失是 L_ℓ-利普希茨的，且 π_ψ 是 λ_π-利普希茨连续的 [42]，则有：

$$\begin{aligned} R_E(\hat{\pi}) - R^\star &= \underbrace{(R_E(\hat{\pi}) - \inf_{\pi \in \mathcal{H}_{\mathbf{P}}} R_E(\pi))}_{\text{Estimation Error}} \\ &\quad + \underbrace{(\inf_{\pi \in \mathcal{H}_{\mathbf{P}}} R_E(\pi) - R^\star)}_{\text{Approximation Error}} \\ &\le \underbrace{\left(2\mathcal{R}(\Pi) + \sqrt{\frac{\log(1/\delta)}{2N}}\right)}_{\text{Bound via Uniform Convergence}} \\ &\quad + \underbrace{L_\ell \lambda_\pi \cdot \mathbb{E}_c \left[\min_\alpha \|\boldsymbol{z}^f - \boldsymbol{z}_c^\star\|_2 \right]}_{\text{Bound via Lipschitz Continuity}}. \end{aligned} \quad (31)$$

将融合表示 $\boldsymbol{z}^f = \boldsymbol{z}^v + \sum \alpha_k \boldsymbol{z}^k$ 的定义代入第二项，即可得到所述界。详细推导请参见**补充材料** 。 □

备注。 定理 1 将超额风险分解为估计项和近似项。该界表明，风险受与预言机的特征距离所控制（缩放系数为 $L_\ell \lambda_\pi$）。这为我们的设计提供了依据：混合式对比学习确保提示池 $\mathbf{P}$ 具有足够的多样性以缩小近似差距，而注意力机制 $\mathcal{G}_{\text{attn}}$则动态优化 α ，以实现精确对齐。

V. 实验设置

本节介绍 CAPO 的实验设置，涵盖仿真环境和领域配置、评估指标和基线方法，以及网络架构和超参数设置。

A. 仿真设置

我们在 AI2-THOR [36] 环境中进行实验。我们专门使用 FloorPlan21 对 ObjectNav任务进行训练和评估，其中数据划分由域因素的不同组合加以区分。成功定义为在最大时间步上限$T_{\max} = 500$内到达 12 个目标类别之一的 1.0 米范围内。对于对比提示学习，我们收集了一个专家数据集，涵盖由 $\mathcal{F} = (\phi, \psi, \theta, \delta, \ell^b, \ell^c, \ell^s, \ell^h)$ 定义的多样化领域配置。我们将光照参数$\ell^b, \ell^c, \ell^s, \ell^h$ （亮度、对比度、饱和度、色调）视为连续变量，而旋转 ψ、视角 θ和步长 δ为离散变量。视场 ϕ 被分为三组以缓解训练不稳定性，从而共得到 10 个不同领域因素。这些域因素从均匀分布中随机采样，以构建用于训练和评估的源域、已见目标领域和未见目标领域设置。所收集的数据集包含 1,468 条专家轨迹、共 22,754 个样本，其中跨领域轨迹对基于语义相似度阈值 $\kappa \geq 0.7$进行对齐。智能体在由前进、向左旋转、向右旋转、向上看、向下看和结束组成的离散动作空间中运行。

B. 评估指标

为定量评估我们的方法，我们报告了若干广泛采用的标准指标 [36]:

- 成功率（*SR*）（%）：智能体成功导航至距目标 1.0 米以内并执行终止动作的评估回合所占的百分比。
- 路径长度加权成功率（*SPL*）：一种综合指标，用于平衡任务完成情况与轨迹效率，定义为 SPL $= \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}_i \cdot \frac{d_i^\star}{\max(d_i^\star, p_i)}$，其中 $\mathbb{I}_i \in \{0,1\}$表示评估回合 i是否成功， $d_i^\star$表示到达目标的最优测地距离，p_i表示智能体实际行进的路径长度。
- 导航误差（*NE*）（米）：评估回合终止时，智能体最终位置与最近目标物体之间的平均欧几里得距离。
- 回合长度（*EL*）：每个回合平均消耗的时间步数。

C. 基线方法

为全面验证 CAPO 的有效性，我们将其与五种具有代表性的基线方法进行比较:

- CURL [13]：一种具有代表性的视觉对比学习方法，通过对同一观测进行数据增强来构造正样本对和负样本对，从而学习状态表示。
- ACO [43]：一种动作条件对比学习基线方法，将与相同已执行动作关联的观测视为正样本对，同时将其他观测视为负样本，以捕获动作相关特征。
- ATC [14]：一种时间对比学习方法，将相邻帧视为正样本对、远距离帧视为负样本，以增强时序一致性。
- ConPE [44]：一种最先进的基线方法，利用 CLIP 促进视觉对比学习和动作条件对比学习，同时采用单模态注意力机制进行策略适应。
- PPO [40]：一种标准的强化学习基线方法，通过将视觉骨干网络替换为冻结的 CLIP 编码器来修改原始PPO算法，以进行公平比较。

D. 参数与架构配置

对比提示学习。 我们采用CLIP ViT-B/32模型作为冻结骨干网络 Φ，其输入为尺寸为 $3 \times 224 \times 224$ 的RGB图像。可学习提示池$\mathbf{P}$ 由 $K = 10$ 个特定领域提示组成（4个用于外观，5个用于动作属性，1个用于文本）。每个提示$\boldsymbol{p}^k$ 的长度为 $L = 8$ ，嵌入维度为$D = 768$，并通过Xavier均匀初始化进行初始化。

对于视觉对比学习，我们使用 $|\mathcal{B}|$ =256的批量大小，并训练500个训练轮次。损失缩放超参数λ_v 和 λ_t 均设为1.0。我们使用动量为0.9的SGD优化器、权重衰减 10^{-3}，以及初始学习率为0.001的多项式衰减调度。对于时序动作对比学习，我们采用64的批量大小，并训练500个训练轮次。在线网络由提示编码器 Φ组成，其后接一个两层MLP投影器ρ_ω （$d_{\text{out}} = 512 \to 4096 \to 16$）和一个两层MLP预测器$f_\omega$ （$16 \to 512 \to 16$），二者均使用ReLU激活和批归一化。目标网络通过对在线网络进行指数移动平均来更新，动量系数为 $\beta = 0.99$。优化器设置遵循与视觉对比学习相同的调度。对于文本对比学习，我们采用256的批量大小，并训练500个训练轮次。我们使用Nesterov动量为0.9的SGD优化器、权重衰减 10^{-3}，以及初始学习率为0.002的多项式衰减调度。

自适应提示编排。 此处，提示池$\mathbf{P}$ 与编码器 Φ被冻结。注意力机制 $\mathcal{G}_{\text{attn}}$利用投影网络 f_p 计算注意力键，该网络实现为带SiLU激活的两层多层感知机（$512 \to 128 \to 512$）。策略网络π_ψ 接收融合视觉特征$\mathbf{z}_t^f \in \mathbb{R}^{512}$ 与独热目标向量$\mathbf{g} \in \mathbb{R}^{12}$的拼接输入。时序依赖由单层门控循环单元建模，其隐藏状态大小为

512。演员头和评论家头均为线性层，分别将门控循环单元的输出映射为动作对数几率与价值估计。

我们使用近端策略优化（PPO）和 48 个并行环境来优化策略。训练共进行 3×10^6 个环境步。在每个更新周期中，我们收集包含 4,800 个状态转移的轨迹缓冲区（每个环境的轨迹长度为 100）。关键超参数如下：学习率设为 3×10^{-4}，并线性衰减至 10^{-5}；PPO 裁剪范围为 $\epsilon = 0.2$，折扣因子为 $\gamma = 0.99$。广义优势估计（ GAE ）参数设为 0.95，价值损失系数设为 0.5。在奖励函数设计方面，我们采用复合奖励函数。具体而言，智能体成功到达目标物体后获得 +10.0 的奖励，每个时间步则受到 -0.01 的惩罚，以提高效率。此外，我们还采用基于距离的奖励塑形：智能体在每一步获得 $r = d_{t-1} - d_t$，其中 d_t 表示到目标的测地距离；为防止利用奖励漏洞，奖励幅度被限制为实际行进距离。所有训练和评估实验均在单块 NVIDIA GeForce RTX 4090 GPU 上进行。

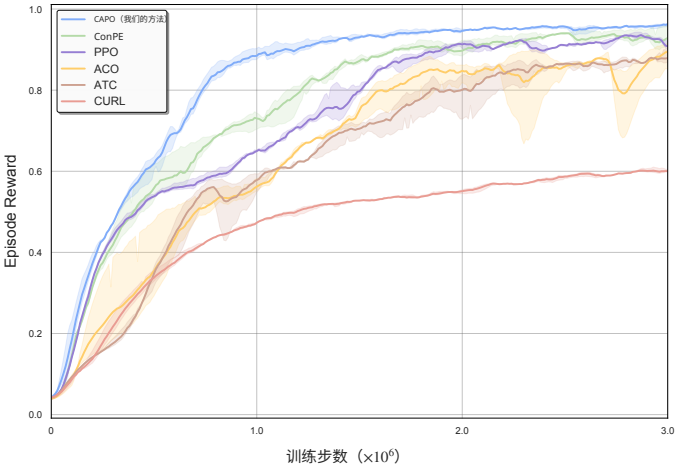


图 4：所有基线方法的回合奖励训练曲线。所有结果均为三个随机种子的平均值，阴影区域表示置信区间。CAPO 展现出更高的样本效率和渐近性能，收敛速度更快，且取得的奖励高于所有基线方法。

VI. 实验结果与分析

本节介绍实验结果，并对验证 CAPO 有效性的关键方面进行分析。我们在仿真中评估其在多种域因素和具身配置下的性能与适应能力，并将其与多种基线方法进行比较。此外，我们还开展消融研究，以考察各个核心组件的作用。

A. 与基线方法的比较

训练过程在 4 个源域上进行，每个源域都由一组独特的域因素组合所定义。评估时，我们采用一套全面的数据集，包括 4 个源域、30 个已见目标域和 10 个未见目标域，以严格评估适应能力。二者的主要区别在于：这些领域配置是否在表征学习阶段出现过。

训练过程的奖励曲线如图所示4。如图所示，我们的方法（CAPO）收敛速度最快，渐近性能也最强，始终优于所有基线方法。这一显著优势源于三项关键设计选择：(i) 一种协同训练范式，其中提示基底被冻结，而编排权重与策略则联合

经过优化，有效结合了解耦表示的稳定性与端到端学习的灵活性。这种混合范式绕过了端到端方法通常存在的样本效率低下问题，同时克服了完全解耦流水线所导致的静态表征的僵化；（ii）文本对比学习：与仅依赖视觉特征的基线方法不同，CAPO 利用了 CLIP 模型中蕴含的丰富语义先验。通过引入可学习文本提示，我们明确地将视觉表征与目标导向语义对齐。这种对齐不仅提高了样本效率，还提供了更稳健的特征空间，使策略能够收敛到更高的最优性能；（iii）自适应提示编排：我们不再使用静态特征表示，而是

引入一种动态注意力机制，根据当前上下文自适应地聚合特定领域的提示。这使策略能够快速识别并提高最相关特征的权重，从而有效加速学习过程并取得更好的性能。

在各基线方法中，ConPE 的表现位居第二。然而，由于依赖单模态表征学习，且缺乏明确的跨模态语义锚定，它未能达到 CAPO 的表现，这限制了其充分捕捉与目标相关线索的能力。值得注意的是，仅使用冻结的 CLIP 编码器的 PPO，表现优于部分对比学习基线方法。我们认为，这是因为预训练的 CLIP 编码器具有稳健的适应能力，与从头训练的模型相比，能够提供更优的初始特征。相反，尽管 ATC 和 ACO 试图捕捉时间或动作相关特征，但它们收敛较慢，且表现不及 PPO。这一反直觉的结果表明，它们的粗粒度建模以及缺乏语义引导，无法捕捉实现精确导航所必需的细粒度环境动态。CURL 的表现最差，这表明在有限的交互数据上进行实例级对比学习会导致预训练通用知识发生灾难性遗忘，并且在微调过程中会受到分布偏移的显著影响。

所有领域的定量结果汇总于表I。结果表明，CAPO在所有域因素上始终保持强劲的适应能力。相比之下，大多数基线方法出现显著的性能下降，尤其是在未见领域中，这表明它们未能有效应对分布偏移。尽管ConPE展现出一定的适应能力，但仍不及CAPO。这一性能差距进一步凸显了仅依赖单模态视觉线索的局限性。通过将文本对比学习与动态自适应提示编排相结合，CAPO有效弥合了视觉特征

表 I：仿真环境中的基线方法性能

方法	源域				已见目标域				未见目标域			
	SR↑	SPL↑	NE↓	EL↓	SR↑	SPL↑	NE↓	EL↓	SR↑	SPL↑	NE↓	EL↓
CURL [13]	52.0±1.9	0.32±0.07	0.48±0.08	32±6	15.5±7.0	0.12±0.07	0.93±0.15	55±13	10.3±4.5	0.07±0.03	1.06±0.05	70±11
ACO [43]	79.1±22.8	0.51±0.16	0.15±0.05	40±21	59.5±17.1	0.37±0.12	0.17±0.07	49±11	42.6±28.1	0.27±0.20	0.17±0.11	53±9
ATC [14]	82.1±8.1	0.53±0.10	0.12±0.04	27±6	70.4±8.8	0.44±0.08	0.15±0.09	41±16	51.1±11.2	0.31±0.09	0.15±0.08	46±11
ConPE [44]	96.1±2.3	0.64±0.06	0.03±0.01	20±4	83.3±3.9	0.53±0.09	0.06±0.08	33±6	80.5±6.2	0.52±0.09	0.07±0.07	36±4
近端策略优化 [40]	86.5±5.9	0.53±0.10	0.17±0.08	25±7	77.7±3.6	0.48±0.09	0.15±0.07	38±16	59.8±8.1	0.35±0.10	0.19±0.06	43±8
CAPO（我们的方				**18±3**	**90.9±3.4**			**29±5**	**86.4±5.7**			**32±3**

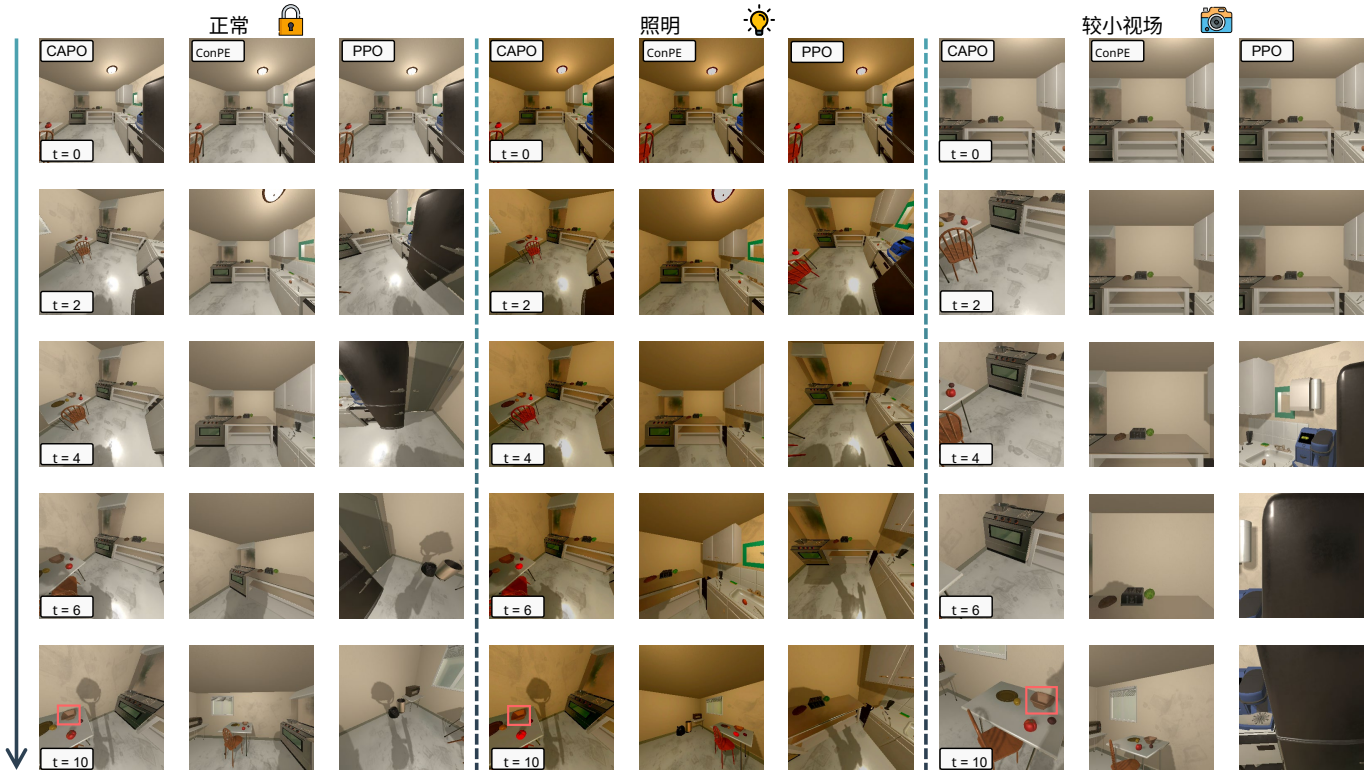


图 5：在多种域偏移下，目标为“碗”的导航回合可视化。各列展示了 CAPO、ConPE 和 PPO 在三种情形下的自我中心视角帧序列：正常、光照变化和跨具身（较小视场）。CAPO 在所有条件下都能迅速识别并接近目标，展现出强大的零样本适应能力。ConPE 表现出一定的适应能力，但锁定目标有所延迟，因此需要更多步数才能完成任务。相比之下，PPO 无法适应环境干扰，导致方向迷失并最终任务失败。

与目标语义之间的差距，同时有选择地聚合特定于域的特征。这种多模态对齐使策略能够在视觉干扰下保持稳健，并灵活适应未见的具身配置，而受限于固定或单模态表征的基线方法则无法泛化。

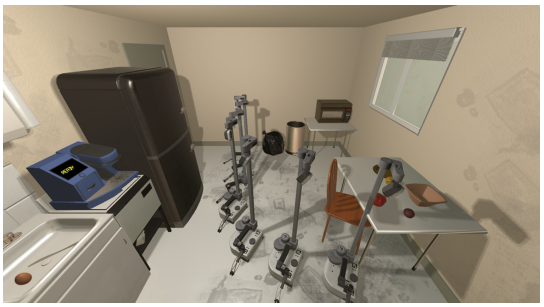
B. 跨具身策略适应

图中展示了智能体在不同域因素和具身配置下的第一人称视角序列。5可以看出，CAPO 在各评估域中表现出最果断的导航行为。具体而言，即使光照或视场发生剧烈变化，我们的智能体仍能稳定地将视觉焦点保持在目标“碗”上，而基线方法则常常执行冗余动作或失去跟踪。这表明我们的文本对比学习和多模态提示编排机制是有效的。通过将视觉特征明确锚定于不变的语言语义，并动态过滤域噪声，CAPO 确保策略始终与任务目标保持稳健一致，避免了缺乏这种跨模态锚定的方法中出现的性能退化。

此外，为了严格验证 CAPO 在传感器参数偏移（例如视场）之外的有效性，我们将评估扩展至物理跨具身适应。如图6所示，我们将仅在 ManipulaTHOR 来源智能体上训练的策略直接部署到两个异构机器人平台上：Stretch RE1 和 LoCoBot。这些具身形态的相机安装高度和运动学约束存在巨大差异，通常会扰乱传统智能体的状态分布。然而，



（a）ManipulaTHOR



(b) Stretch RE1



(c) LoCoBot

图 6：在不同的未见过的机器人具身形态上进行零样本策略适应。该策略在标准智能体 (a) ManipulaTHOR 上训练，随后直接部署到 (b) Stretch RE1 和 (c) LoCoBot 上。尽管与源域相比，具身配置存在显著差异，CAPO 仍成功适应了这些异构具身形态，从而无需任何微调即可实现有效导航。

在这一极具挑战性的场景中，CAPO 展现出卓越的适应能力。借助学习得到的提示，编排机制有效弥补了不同物理配置所造成的自我中心视角差异。这使智能体能够在不同具身形态上执行连贯的导航动作，进一步验证了 CAPO 跨具身形态部署的潜力。

为进一步分析内部表示，我们将 t-SNE 嵌入可视化 [45] 于图7中。可视化结果表明，由对应相同域因素的提示提取出的视觉表征紧密聚类，而与不同域因素相关的视觉表征则彼此明显分离。这种清晰的聚类结构证实了我们的混合对比学习策略有效。此外，图8展示了域与提示的相关性矩阵，说明自适应提示编排

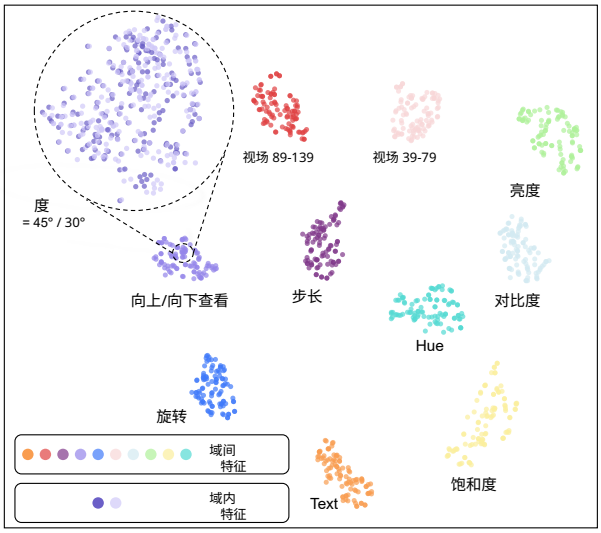


图 7：学习到的提示表示的 t-SNE 可视化。图中显示，不同域的提示彼此明显分离，而同一域内的特征则紧密聚类。这表明 CAPO 能够有效解耦视觉特征与具身特征。

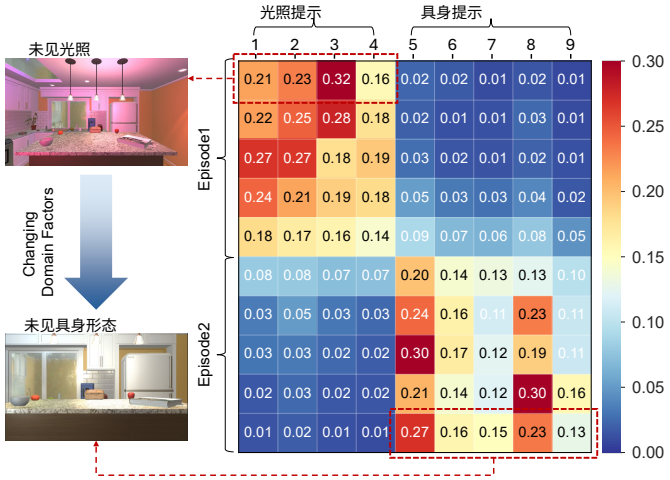


图 8：不同导航回合中的注意力权重可视化。请注意，此处未显示文本提示，因为它直接进行了融合。热力图显示注意力分布随时间动态变化，表明随着域因素的变化，智能体会自适应地重新分配提示权重，以优先关注相关特征。这验证了我们的自适应提示编排机制在部署期间的有效性。

该机制能够实时正确激活与主导领域因素相对应的特定提示。

C. 消融实验

为更好地阐释 CAPO 中各组件的贡献，我们开展了一系列消融研究与分析。

(1) 混合对比学习策略。我们拆解训练目标，以验证各个对比损失组成部分的必要性。我们评估了未采用

表 II：视觉运动策略学习中混合对比学习策略的消融实验

方法	源域				已见目标域				未见目标域			
	SR↑	SPL↑	NE↓	EL↓	SR↑	SPL↑	NE↓	EL↓	SR↑	SPL↑	NE↓	EL↓
CAPO（无视觉）	92.1±2.8	0.62±0.04	0.04±0.02	20±4	84.5±4.6	0.55±0.05	0.06±0.03	31±5	81.2±6.8	0.50±0.07	0.09±0.07	36±6
CAPO 不含动作	89.8±3.0	0.59±0.04	0.04±0.03	23±5	83.1±3.9	0.53±0.06	0.07±0.03	34±5	78.2±5.7	0.47±0.06	0.08±0.07	41±10
CAPO 不含文本	87.4±3.3	0.56±0.06	0.07±0.04	26±5	80.5±3.8	0.50±0.06	0.08±0.06	37±5	72.3±6.8	0.45±0.07	0.08±0.07	45±12
CAPO	**97.9±1.2**			**18±3**	**90.9±3.4**			**29±5**	**86.4±5.7**			**32±3**

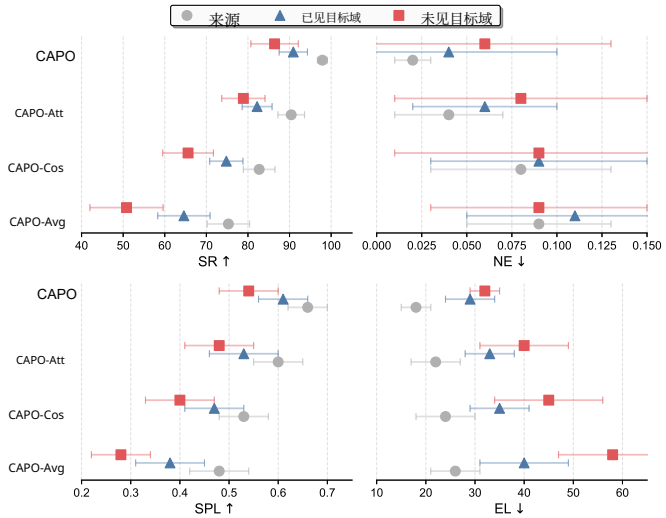


图 9：注意力机制的消融实验。我们将完整模型与均匀平均（CAPO-Avg）及单分支变体（CAPO-Att、CAPO-Cos）进行比较。结果表明，完整 CAPO 始终优于所有替代方案，证实了结合可学习投影与余弦相似度进行特征融合的必要性。

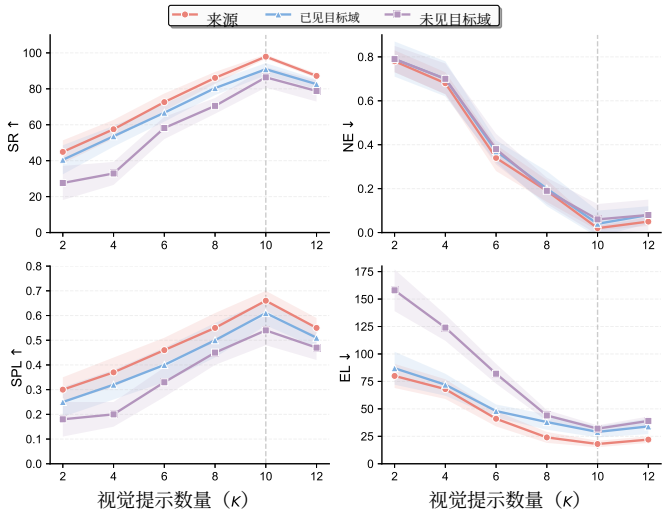


图10：提示池大小 K 的敏感性分析。随着 K 增加，性能持续提升，并在 $K = 10$ 时达到峰值，这与不同领域因素相吻合。将 K 进一步增加到 12 会导致性能下降，表明过多的提示会引入冗余和过拟合。

文本对比损失（w/o text）、动作时序损失（w/o action）和视觉对比损失（w/o visual）训练的 CAPO 变体。如表 II所示，w/o text 导致的性能下降最为显著，这凸显了语言先验所提供的语义对齐对于适应任务至关重要。w/o action 导致的性能降幅位居第二。鉴于我们的评估涵盖多种跨具身配置，这种性能下降证实，对动作条件时序一致性进行建模对于适应不同的智能体动态必不可少。值得注意的是，w/o visual 所导致的性能下降最小。我们认为，这是因为另外两个目标在训练期间也采用了适度的数据增强，即使没有显式的视觉对比损失，也能隐式地赋予模型一定程度的视觉不变性。不过，完整的混合目标仍能取得最佳性能，这表明这三个组成部分具有互补性。

(2) 注意力机制的有效性。我们通过将自适应提示编排机制与三种替代融合策略进行比较，分析其贡献：(i) CAPO-Avg，其中最终状态表示是所有提示嵌入的简单均匀平均；(ii) CAPO-
注意力仅利用可学习投影网络来计算注意力权重，而不考虑余弦相似度计算；(iii) CAPO-Cos 则仅依靠提示嵌入与原始嵌入之间的余弦相似度来确定权重，不使用可学习投影。对比结果如图9所示。我们观察到，CAPO-AVG 的表现明显落后于完整方法。与完整 CAPO 相比，CAPO-Att 和 CAPO-Cos 的表现均有所下降。这表明，单靠可学习投影或语义相似度计算都不足以取得理想效果。完整 CAPO 的优异表现证实了混合设计的必要性：可学习注意力提取任务相关特征，而余弦相似度则充当语义锚点。

D. 参数敏感性分析

为评估 CAPO 在不同超参数设置下的性能，我们进一步进行了敏感性分析。

(1) 对提示数量的敏感性。我们研究视觉提示数量对 K 策略成功率的影响。具体而言，我们使用不同的提示池大小 $K \in \{2, 4, 6, 8, 10, 12\}$训练 CAPO 方法，其中 $K = 12$ 的设置通过将视场参数细分为更小的区间得到。评估结果汇总于图10中。如图所示，性能持续提升

表 III：视觉运动策略学习中语义正则化的敏感性分析

方法	源域				已见目标域				未见目标域			
	SR↑	SPL↑	NE↓	EL↓	SR↑	SPL↑	NE↓	EL↓	SR↑	SPL↑	NE↓	EL↓
$\sigma = 0$	92.6±1.9	0.63±0.04	0.03±0.02	20±3	85.4±3.8	0.56±0.05	0.06±0.03	32±5	81.8±5.9	0.51±0.07	0.08±0.08	35±5
$\sigma = 0.1$	**97.9±1.2**	**(**		**18±3**	**90.9±3.4**	**(**		**29±5**	**86.4±5.7**	**(**		**32±3**
$\sigma = 0.3$	87.1±2.1	0.56±0.07	0.04±0.04	27±5	78.7±4.2	0.47±0.06	0.08±0.08	38±6	70.5±6.6	0.42±0.08	0.10±0.08	45±8
$\sigma = 0.5$	83.5±3.0	0.53±0.06	0.08±0.06	25±6	73.2±5.2	0.46±0.08	0.09±0.08	39±6	65.6±6.9	0.38±0.08	0.11±0.08	47±11
CAPO-Reg	80.2±3.9	0.49±0.07	0.13±0.06	37±6	69.8±6.3	0.41±0.08	0.11±0.09	47±8	57.4±7.7	0.32±0.09	0.15±0.09	72±13

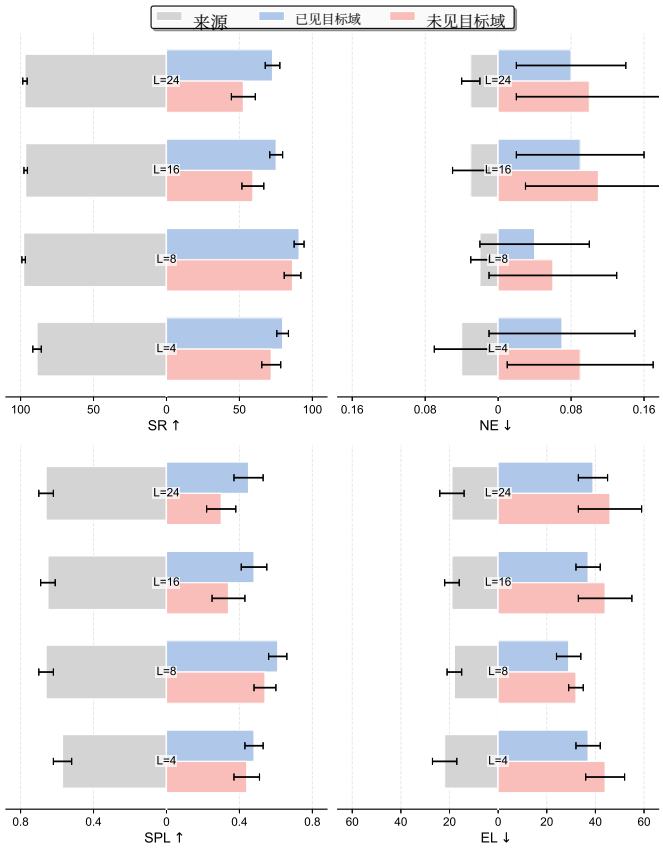


图 11：视觉提示长度敏感性分析L。我们评估了不同长度下的性能 $L \in \{4, 8, 16, 24\}$。结果表明，$L = 8$ 在特征表达能力与适应能力之间实现了最佳权衡。较短的提示（$L = 4$）会导致欠拟合，而较长的提示（$L = 24$）则会对源域过拟合，从而降低在未见目标域中的适应能力。

随着 K 增加。提示数量较少时，性能提升有限，因为模型容量不足以捕捉环境中存在的多种域因素。然而，当数量达到 $K = 10$左右时，性能趋于饱和；这与我们在数据收集阶段定义的不同域因素总数一致。这表明，将提示池大小与底层域变化相匹配是最佳选择，而过多的提示可能会造成冗余，却无法带来明显的性能提升。因此，我们采用$K = 10$ 作为默认设置。

（2）对提示长度的敏感性。我们考察提示长度对 CAPO 的影响——即 CAPO 对可学习视觉提示长度的敏感性L，并测试 $L \in \{4, 8, 16, 24\}$。图11 中的结果表明 $L = 8$ 能提供最佳平衡。较短的提示（$L = 4$）缺乏编码复杂领域特定特征所需的表达能力，导致欠拟合。相反，较长的提示（$L = 24$）往往会过拟合源域的特定视觉纹理，从而妨碍对未见目标域的零样本适应。这表明，适中的提示长度足以有效引导冻结的 CLIP 编码器。

(3) 对语义正则化的敏感性。我们考察了 CAPO 对语义正则化机制中所用噪声尺度的敏感性 $\sigma \quad \zeta \sim \mathcal{N}(0, \sigma^2 I)$。如表 III所示，引入适量噪声（$\sigma = 0.1$）可获得最高性能，优于其他设置。这表明，适当的正则化有助于防止对源域过拟合。然而，当 σ 从 0.3 增加到 0.5 时，性能会下降，这说明过度扰动会破坏特征完整性。此外，移除文本提示而仅依赖正则化（CAPO-Reg）会导致性能显著下降，证实噪声注入能够补充文本提示提供的显式语义引导，但无法取而代之。

VII. 结论

在这项工作中，我们提出了 CAPO，这是一种新颖的方法，将对比提示学习与自适应提示编排相结合，以提高导航策略的样本效率和跨具身适应能力。在对比提示学习中，我们采用融合视觉、动作和文本目标的混合对比策略，学习一个多样化的特定领域视觉提示池。这些提示能够有效捕捉环境和具身的细粒度变化。随后，我们冻结视觉编码器，并引入动态提示编排机制，根据当前观测上下文自适应地聚合这些提示。CAPO 使智能体能够在域变化下动态调整视觉特征的权重，从而快速适应未见过的配置。大量仿真实验表明，CAPO 显著提高了样本效率和渐近性能，同时能够在多种未见领域中实现稳健的零样本适应。然而，该方法的适应能力受源域多样性的限制，在与学习得到的提示池正交的极端分布偏移下，性能可能因此受限。未来的工作将通过在线提示适应来解决这一问题。此外，尽管当前方法主要将语义先验用于 ObjectNav任务，我们计划将这种语义对齐能力扩展到更复杂的视觉-语言导航任务。

将这种语义对齐能力扩展到更复杂的视觉-语言导航任务。

参考文献

[1] J. Xiao、R. Zhang、Y. Zhang 和 M. Feroskhan，“无人机的基于视觉的学习：综述”，IEEE Transactions on Neural Networks and Learning Systems，第 36 卷，第 9 期，第 15 601–15 621 页，2025年。[2] C. Kong、A. Luo、S. Wang、H. Li、A. Rocha 和 A. C. Kot，“用于图像篡改定位的像素不一致性建模”，IEEE Transactions on Pattern Analysis and Machine Intelligence，第 47 卷，第 6 期，第 4455–4472 页，2025年。[3] A. Srisuchinnawong 和 P. Manoonpong，“一种具有自适应在线学习能力、用于样本高效的机器人运动学习的可解释神经控制网络”，IEEE Transactions on Neural Networks and Learning Systems，第 36 卷，第 9 期，第 16 143–16 155 页，2025 年。[4] J. Wu 、 Y. Zhou 、 H. Yang 、 Z. Huang 和 C. Lv，“结合仿真到现实迁移、用于自主导航的人类引导的强化学习”，IEEE Transactions on Pattern Analysis and Machine Intelligence，第 45 卷，第 12 期，第 14 745–14 759 页，2023年。[5] S. Belkhale、Y. Cui 和 D. Sadigh，“Hydra：用于模仿学习的混合机器人动作”，载于 Conference on Robot Learning (CoRL)。PMLR，2023年，第 2113–2133 页。[6] M. Zare、P. M. Kebria、A. Khosravi 和 S. Nahavandi，“模仿学习综述：算法、最新进展与挑战”，IEEE Transactions on Cybernetics，第 54 卷，第 12 期，第 7173–7186 页，2024年。[7] Y. Zhang、C. Yan、J. Xiao 和 M. Feroskhan，“Npe-drl：通过非专家策略引导的强化学习，提升感知约束的避障能力”，IEEE Transactions on Artificial Intelligence，第 6 卷，第 1 期，第 184–198 页，2024年。[8] Y. Zhang、Y. Hu、Y. Song、D. Zou 和 W. Lin，“通过可微物理学习基于视觉的敏捷飞行”，Nature Machine Intelligence，第 1–13 页，2025年。[9] J. Tobin、R. Fong、A. Ray、J. Schneider、W. Zaremba 和 P. Abbeel，“通过域随机化将深度神经网络从仿真迁移到现实世界”，载于 2017年 IEEE/RSJ 智能机器人与系统国际会议 (IROS)。IEEE，2017年，第 23–30 页。[10] J. Tobin、L. Biewald、R. Duan、M. Andrychowicz、A. Handa、V. Ku- mar、B. McGrew、A. Ray、J. Schneider、P. Welinder 等，“用于机器人抓取的域随机化与生成模型”，载于 2018 IEEE/RSJ 智能机器人与系统国际会议 (IROS)。IEEE，2018年，第 3482–3489 页。[11] D. Horv´ath、G. Erd˝os、Z. Istenes、T. Horv´ath 和 S. F¨oldi，“使用 sim2real 域随机化进行目标检测及其在机器人应用中的应用”，IEEE Transactions on Robotics，第 39 卷，第 2 期，第 1225–1243 页，2022年。[12] A. Ku mar、Z. Fu、D. Pathak 和 J. Malik，“Rma：足式机器人的快速运动适应”，Robotics: Science and Systems (RSS)，第 1–11 页，2021年。[13] M. Laskin、A. Srinivas 和 P. Abbeel，“Curl：用于强化学习的对比无监督表征”，载于国际机器学习会议 (ICML)。PMLR，2020年，第 5639–5650 页。[14] A. Stooke、K. Lee、P. Abbeel 和 M. Laskin，“将表征学习与强化学习解耦”，载于国际机器学习会议 (ICML)。PMLR，2021年，第 9870–9879 页。[15] Y. Zhang、J. Xiao 和 M. Feroskhan，“学习用于自主无人机导航的跨模态视觉运动策略”，IEEE Robotics and Automation Letters，第 10 卷，第 6 期，第 5425–5432 页，2025年。[16] J. Zhao、Y. Wang、Z. Cai、N. Liu、K. Wu 和 Y. Wang，“通过对比世界模型学习用于自主无人机导航的视觉表征”，IEEE Transactions on Artificial Intelligence，第 5 卷，第 3 期，第 1263–1276 页，2023年。[17] J. Xing、L. Bauersfeld、Y. Song、C. Xing 和 D. Scaramuzza，“通过对比学习增强基于视觉的敏捷飞行中的稳健场景迁移”，载于 2024 IEEE 机器人与自动化国际会议 (ICRA)。IEEE，2024年，第 5330–5337 页。[18] Y. Zhang、J. Xiao、C. Yan 和 M. Feroskhan，“用于视觉运动策略的预言机引导掩码对比强化学习”，arXiv 预印本 arXiv:2510.05692，2025年。[19] A. Majumdar、G. Aggarwal、B. Devnani、J. Hoffman 和 D. Batra，“Zson：使用多模态目标嵌入的零样本目标导航”，神经信息处理系统进展 (NeurIPS)，第 35 卷，第 32 340–32 352 页，2022年。[20] D. Shah、B. Osi´nski、S. Levine 等，“Lm-nav：利用语言、视觉和动作大型预训练模型进行机器人导航”，载于 Conference on Robot Learning (CoRL)。PMLR，2023年，第 492–504 页。[21] Y. Zhang、H. Yu、J. Xiao 和 M. Feroskhan，“面向无人机、具有开放词汇目标理解能力的基于视觉-语言的接地导航”，arXiv 预印本 arXiv:2506.10756，2025年。[22] V. S. Dorbala、J. F. Mullen 和 D. Manocha，“具身智能体能找到你的猫咪造型马克杯吗？基于大语言模型的零样本目标导航”，IEEE Robotics and Automation Letters，第9卷，第5期，第4083–4090页，2023年。[23] A. Radford、J. W. Kim、C. Hallacy、A. Ramesh、G. Goh、S. Agarwal、G. Sastry、A. Askell、P. Mishkin、J. Clark 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（ICML）。PMLR，2021年，第8748–8763页。[24] A. Loquercio、A. I. Maqueda、C. R. Del-Blanco 和 D. Scaramuzza，“DroNet：通过驾驶学习飞行”，IEEE Robotics and Automation Letters，第3卷，第2期，第1088–1095页，2018年。[25] Y. Pan、C.-A. Cheng、K. Saigol、K. Lee、X. Yan、E. A. Theodorou 和 B. Boots，“敏捷自动驾驶的模仿学习”，《国际机器人研究杂志》，第39卷，第2-3期，第286–302页，2020年。[26] R. Ramrakhya、E. Undersander、D. Batra 和 A. Das，“Habitat-Web：从大规模人类示范中学习具身物体搜索策略”，载于 IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，2022年，第5173–5183页。[27] T. Z. Zhao、V. Kumar、S. Levine 和 C. Finn，“利用低成本硬件学习细粒度双手操作”，Robotics: Science and Systems（RSS），第1–18页，2023年。[28] S. Ross、G. Gordon 和 D. Bagnell，“将模仿学习和结构化预测归约为无悔在线学习”，载于第十四届国际人工智能与统计学会议（AISTATS）论文集。JMLR研讨会与会议论文集，2011年，第627–635页。[29] E. Kaufmann、L. Bauersfeld、A. Loquercio、M. M¨uller、V. Koltun 和 D. Scaramuzza，“利用深度强化学习实现冠军级无人机竞速”，《自然》，第620卷，第7976期，第982–987页，2023年。[30] J. Xiao、P. Pisutsin 和 M. Feroskhan，“视觉无人机集群协同目标搜索：一种嵌入自适应课程的多阶段强化学习方法”，IEEE Transactions on Neural Networks and Learning Systems，第36卷，第1期，第313–327页，2023年。[31] C. Yan、C. Wang、X. Xiang、K. H. Low、X. Wang、X. Xu 和 L. Shen，“障碍物密集环境中多架固定翼无人机的避碰集群飞行：一种基于任务专属课程的多智能体深度强化学习方法”，IEEE Transactions on Neural Networks and Learning Systems，第35卷，第8期，第10 894–10 908页，2023年。[32] K. Zhou、J. Yang、C. C. Loy 和 Z. Liu，“为视觉-语言模型学习提示”，《国际计算机视觉杂志》，第130卷，第9期，第2337–2348页，2022年。[33] L. Yang、Y. Wang、X. Li、X. Wang 和 J. Yang，“细粒度视觉提示”，神经信息处理系统进展（NeurIPS），第36卷，第24 993–25 006页，2023年。[34] M. U. Khattak、H. Rasheed、M. Maaz、S. Khan 和 F. S. Khan，“MAPLE：多模态提示学习 ”，载于 IEEE/CVF 计算机视觉与模式识别会议（CVPR）论文集，2023年，第19 113–19 122页。[35] L. Yang、X. Li、Y. Wang、X. Wang 和 J. Yang，“细粒度视觉文本提示”，IEEE Transactions on Pattern Analysis and Machine Intelligence，第47卷，第3期，第1594–1609页，2024年。[36] E. Kolve、R. Mottaghi、W. Han、E. VanderBilt、L. Weihs、A. Herrasti、M. Deitke、K. Ehsani、D. Gordon、Y. Zhu 等，“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。[37] J.-B. Grill、F. Strub、F. Altch´e、C. Tallec、P. Richemond、E. Buchatskaya、C. Doersch、B. Avila Pires、Z. Guo、M. Gheshlaghi Azar 等，“自举潜在表示：一种自监督学习新方法”，神经信息处理系统进展（NeurIPS），第33卷，第21 271–21 284页，2020年。[38] P. Ramachandran、B. Zoph 和 Q. V. Le，“搜索激活函数”，载于国际学习表征会议（ICLR），2018年，第1–13页。[39] M. O. Turkoglu、S. D’Aronco、J. D. Wegner 和 K. Schindler，“重新审视门控机制：可训练的深层多层循环神经网络”，IEEE Transactions on Pattern Analysis and Machine Intelligence，第44卷，第8期，第4081–4092页，2021年。[40] J. Schulman、F. Wolski、P. Dhariwal、A. Radford 和 O. Klimov，“近端策略优化算法”，arXiv 预印本 arXiv:1707 06347，2017年。[41] C. Feng 和 I. Patras，“MaskCon：面向粗粒度标注数据集的掩码对比学习”，载于 IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，2023年，第19 913–19 922页。

[42] Y. Lei、T. Yang、Y. Ying 和 D.-X. Zhou，“对比表征学习的泛化分析”，载于国际机器学习会议（*ICML*）。PMLR，2023年，第19 200–19 227页。[43] Q. Zhang、Z. Peng 和 B. Zhou，“通过观看 YouTube 视频学习驾驶：动作条件对比策略预训练”，载于欧洲计算机视觉会议（*ECCV*）。施普林格，2022年，第111–128页。[44] W. Choi、W. K. Kim、S. Kim 和 H. Woo，“通过对比提示集成实现具身智能体的高效策略适应”，神经信息处理系统进展（*NeurIPS*），2024年。[45] L. Van der Maaten 和 G. Hinton，“使用 t-SNE 可视化数据。”《机器学习研究期刊》，第9卷，第11期，2008年。



Yuhang Zhang （IEEE研究生会员）于2021年获得中国哈尔滨哈尔滨工程大学飞行器推进工程工学学士学位，并于2023年获得新加坡南洋理工大学（NTU）机械与航空航天工程学院工程硕士学位。目前，他正在南洋理工大学（NTU）新加坡校区机械与航空航天工程学院攻读博士学位。

他的研究主要关注无人机、深度强化学习以及视觉语言导航。



Chao Yan （IEEE会员）于2017年获得中国徐州中国矿业大学电气工程及其自动化工学学士学位，并分别于2019年和2023年获得中国长沙国防科技大学控制科学与工程硕士和博士学位。2021年至2022年，他曾在新加坡南洋理工大学机械与宇航工程学院访问，担任博士生。

他目前是中国南京南京航空航天大学自动化学院副教授。研究兴趣包括深度强化学习和无人机集群协同控制。



俞佳曦 于2025年获得中国济南山东大学机械设计制造及其自动化专业工学学士学位。目前，她正在新加坡南洋理工大学（NTU）机械与宇航工程学院攻读理学硕士项目。

她的研究兴趣主要集中于强化学习。



Jiaping Xiao （IEEE会员）于2014年和2017年分别获得中国北京北京航空航天大学飞行器设计与工程专业工学学士学位和飞行动力学与控制专业理学硕士学位，并于2024年获得新加坡南洋理工大学（NTU）智能系统专业博士学位。他目前是南洋理工大学（NTU）机械与宇航工程学院的研究员。

他的研究兴趣包括信息物理系统、强化学习、机器视觉以及空中机器人技术。



Mir Feroskhan （IEEE会员）于2011年获得新加坡南洋理工大学航空航天工程荣誉学士学位，并于2016年获得佛罗里达州墨尔本佛罗里达理工学院航空航天工程博士学位。他目前是南洋理工大学机械与航空航天工程学院的助理教授。

他的研究兴趣包括非线性控制系统、多智能体系统、飞行动力学与控制以及空中机器人技术。