

# 面向具身目标导航的序列扩散规划器研究

张思贤<sup>1),2)</sup> 余欣瑶<sup>1),2)</sup> 宋新航<sup>1),2)</sup> 王翼遥<sup>1),2)</sup> 蒋树强<sup>1),2)</sup>

<sup>1)</sup>(中国科学院计算技术研究所智能信息处理重点实验室 北京 100190)

<sup>2)</sup>(中国科学院大学计算机科学与技术学院 北京 100049)

**摘 要** 物体目标导航是具身智能的基础,其要求智能体根据视觉观测和用户指定的目标在一个未知环境中导航到指定的物体。在导航决策中,决策过程是序列化的,即决策结果是一系列最可能的从当前位置到目标的行动序列。然而,现有的物体导航方法,无论是端到端学习方法还是模块化方法,仅为单步规划,即仅根据当前模型输入预测下一步的单步行为,这导致导航规划结果缺乏在时间上的一致性,从而导致决策短视,进而损失任务的总体效率。为此,本文针对物体目标导航任务,提出序列扩散规划器,用于在导航的过程中根据实时的观测结果以及预先设定的目标类别,生成从当前时刻到未来的序列化规划结果,用以指导智能体高效且快速地导航到目标物体。序列扩散规划器基于扩散概率生成模型(DDPM)来实现,通过自动收集的成功导航过程的行为序列片段来进行训练,从而学习给定导航当前状态到未来状态序列的条件分布。当序列化决策器训练完成,它能够基于智能体的历史语义地图和给定的目标物体,为智能体规划出未来应该采取的行为序列。序列扩散规划器在 Gibson 和 MP3D 的公开虚拟模拟器以及真实实现环境中进行了评估,实验结果表明,相比于基线单步规划器,序列扩散规划器在 Gibson 和 MP3D 模拟环境下在导航成功率(SR)上带来 7.6%和 8.7%的提升,在导航效率(SPL)上带来 4.2%和 3.1%的提升。此外在真实环境下,相比于基线单步规划器,序列扩散规划器在导航成功率的提升更为明显,带来 15.7%的性能提升。

**关键词** 扩散模型;目标导航;具身智能;深度学习

中图法分类号 TP18 DOI号 10.11897/SP.J.1016.2025.02030

## Sequential Diffusion Planner for Embodied Object Goal Navigation

ZHANG Si-Xian<sup>1),2)</sup> YU Xin-Yao<sup>1),2)</sup> SONG Xin-Hang<sup>1),2)</sup> WANG Yi-Yao<sup>1),2)</sup> JIANG Shu-Qiang<sup>1),2)</sup>

<sup>1)</sup>(Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190)

<sup>2)</sup>(School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing 100049)

**Abstract** Object goal navigation is a fundamental task for embodied intelligence, which requires an agent to navigate to a specific object in an unknown environment based on visual observations and a user-specified goal. In human navigation decision-making, the process is sequential, i. e., the decisions made are a series of probable actions from the current location to the target. However, existing object navigation methods, whether end-to-end learning methods or modular methods, typically rely on single-step planning, i. e., only single-step action is predicted based solely on the current state, which results in a lack of temporal consistency in navigation plans, leading to myopic decisions and low task efficiency. To address this, we propose a Sequential Diffusion Planner (SDP) for object goal navigation. SDP plans a sequence of actions from the current time

收稿日期:2025-02-07;在线发布日期:2025-07-15。本课题得到国家自然科学基金杰出青年基金(62125207)、国家自然科学基金面上项目(62272443)、国家自然科学基金重点项目(62032022)、国家自然科学基金联合重点项目(U23B2012)、国家自然科学基金重大项目(62495084,62495085)、北京市自然科学基金(Z190020,L242020)、博士后创新人才支持计划(BX20250391)资助。张思贤,博士,助理研究员,主要研究领域为具身智能、视觉目标导航。E-mail:sixian.zhang@vipl.ict.ac.cn。余欣瑶,硕士研究生,主要研究领域为具身智能、目标导航等。宋新航,博士,副研究员,主要研究领域为具身智能、多媒体内容分析和理解。王翼遥,硕士研究生,主要研究领域为具身智能、目标导航等。蒋树强(通信作者),博士,研究员,中国计算机学会(CCF)高级会员,主要研究领域为图像/视频等多媒体信息的分析、理解与检索技术和多模态智能技术等。E-mail:sqjiang@ict.ac.cn。

to the future, based on real-time observations and target objects, to efficiently guide the agent. Our SDP is implemented based on a diffusion probabilistic model (DDPM). SDP is trained to learn the conditional distribution of the optimal sequential decision given a navigation state by using pre-collected trajectory segments. During navigation, the trained SDP plans a trajectory sequence for the agent based on its historical semantic map and the specified target object. Our SDP is evaluated in the virtual simulators of Gibson and MP3D, as well as in real-world environments. Experimental results show that, compared to the baseline single-step planner, the SDP achieves a 7.6% and 8.7% improvement in success rate (SR) in the Gibson and MP3D simulation environments, respectively, and a 4.2% and 3.1% improvement in navigation efficiency (SPL). Moreover, in the real-world environment, Our SDP demonstrates a more significant improvement in success rate, achieving a 15.7% performance gain over the baseline single-step planner.

**Keywords** diffusion model; object goal navigation; embodied AI; deep learning

## 1 引言

具身智能(Embodied AI)是指智能体在与真实环境的不断交互中,基于获得的来自环境的反馈而获得智能的过程。具身智能要求智能体具有对所处环境全面的理解,并且能够与环境中的事物、人类、或其他智能体进行交互。而交互的基础是智能体可以定位并导航到指定的交互对象。因此具身导航是具身智能的基础任务。其他领域导航(如图像或地图导航)通常假设环境已知且静态不变,只涉及基于完整地图进行路径规划,而具身导航则需要智能体实时感知和理解周围环境并根据这些信息动态调整导航策略,不仅要考虑目标位置,还需要处理与物理环境的交互(如避障、动力学限制),其更侧重对环境的理解和对动态环境的适应能力。

具身视觉物体目标导航(ObjectNav)是具身智能的重要研究分支之一,这一任务定义为:将智能体置于一个未知环境中,此时环境完整的全局地图是不可知的,智能体需要根据实时视觉信息与先前所学习的知识(例如,“椅子的附近可能存在桌子”),在导航的每一步决策要采取的动作,然后逐步导航到用户指定的物体目标(例如,“找到盘子”)。为了高效地完成物体目标导航任务,智能体首先需要构建记忆系统,用于存储之前见过的历史观测到的信息以避免冗余的搜索。此外,它还需要总结导航经验知识来学习实时导航规划,以规划未来应该要采取的动作,来快速导航到指定目标。

人类天生具备卓越的导航能力,生理学研究表明人脑中会构建“认知地图”<sup>[1]</sup>,它能够将导航中的

短期记忆与日常生活中的积累的长期记忆结合,进而推断目标的可能位置。此外,人类在导航过程中的决策过程是基于认知信息寻找最可能的行动序列的过程<sup>[2]</sup>,即人类的导航规划是一个序列化的过程,这一特性保证了导航规划结果的时序一致性和全局最优性。然而,现有的物体目标导航方法大都仅基于单步规划,而非序列化的规划。例如,现有的端到端学习方法<sup>[3-6]</sup>将规划器设计为端到端网络,并通过强化学习(RL)<sup>[3-5]</sup>或模仿学习(IL)<sup>[6]</sup>进行训练。这些规划器在每个时刻根据当前的自我中心视角输出一个单步的行动规划,这种规划器由于使用端到端的网络,其规划的结果可解释性差,此外每一步的规划结果只依赖于以当前观测为主的主视图,并没有约束相邻步规划动作在运动趋势上的一致性,这使得规划结果缺乏时序一致性,从而导致智能体存在频繁摆动走重复路线的问题。此外,这些方法隐式编码了所有的过去观测,导致环境的几何空间信息缺失,限制了在复杂环境中的泛化能力<sup>[7-8]</sup>。另一方面,模块化方法<sup>[9-11]</sup>尝试通过在导航过程中构建语义地图来存储历史信息,语义地图记录了场景的几何空间结构的细节,从而显式地记录已探索的区域并有效避免了冗余探索。然而,模块化方法的决策器依然是单步的,这类方法的决策器通常通过监督学习训练获得,其根据每时刻下语义地图和指定目标在每时刻不断预测航点,航点的预测同样是单步的,忽视了规划动作时序的一致性。

受人类的认知地图与序列化决策的生理学研究启发,本文提出序列扩散规划器。序列扩散规划器通过收集的导航演示进行训练来学习导航经验(长期记忆),然后在导航的过程中根据导航过程中构建

的语义地图(短期记忆)和目标,为智能体规划未来应该采取的行为序列。相比于先前工作仅输出每一时刻的单步预测,本文的方法是输出未来连续时间片段下的序列化的规划结果。

序列规划器以每时刻积累的语义地图和指定目标为输入,输出规划的行为序列。由于语义图和行为序列都是高维的,因此直接学习这种从语义地图到行为序列的条件分布是困难的<sup>[12]</sup>。近年来,扩散模型在表示和学习高维复杂分布方面取得了显著的成功<sup>[13-14]</sup>。扩散模型在扩散过程中逐渐向数据中添加噪声,将复杂数据分布逐步转化为简单分布(例如,高斯分布)。在反向过程(去噪过程)中,逐步预测包含在输入中的噪声,并不断迭代去除噪声,从而实现从简单的随机分布生成复杂的数据分布。这种迭代学习方式增强了学习高维数据分布的稳定性和可控性。因此,为了学习从导航状态(记录在语义地图中的已观测信息和给定目标物体)到最优规划序列的高维分布,序列扩散规划器采用扩散模型进行构建,并且基于收集的最优规划序列通过扩散和去噪过程来学习这种分布,从而在测试过程的未知环境中为智能体规划未来应当采取的最优行为序列。

本文针对视觉目标导航任务提出了序列扩散规划器(Sequential Diffusion Planner, SDP)。序列扩散规划器旨在根据智能体的构建的语义地图中存储的历史观察和目标物体规划出未来应采取的行为序列。具体来说,本文首先在训练环境中使用精确地图收集成功导航过程的行为序列示例,然后驱动配备语义图模块的智能体沿着这些轨迹前进,并在每时刻记录语义地图、动作和位姿数据。由于基于精确的地图,这些行为序列是最优的完成导航任务的行为序列。这些收集到的最优行为序列被进一步按照固定间隔拆分训练数据对,数据对包含序列片段和片段起始位置的导航状态两个部分,其中序列片段是从最优行为序列中拆分出的序列片段,导航状态则是记录了每个片段起始时刻的语义图和导航目标。基于收集到的训练数据对,本文使用 DDPM<sup>[13]</sup>来训练序列扩散规划器。训练过程中通过对序列片段不断叠加噪声逐步将序列片段转化成高斯噪声,序列扩散规划器以导航状态的编码作为条件,以加噪潜变量作为输入,训练目标是以导航状态为条件预测包含在输入带噪序列中的噪声。在导航的过程中,智能体处于未知的环境中,不可获得完整的精确全局地图,智能体需要根据先前学习的导航经验进行自主规划导航路径。本文所提出的序列扩散规划

器在导航过程的每一时刻基于智能体当前的导航状态(累积的语义地图以及导航目标),为智能体规划出应当采取的行为序列,然后驱动智能体沿着规划路径进行移动,直到找到目标。本文基于 Gibson<sup>[15]</sup>和 MP3D<sup>[16]</sup>公开虚拟模拟数据集评估了所提出的序列扩散规划器的性能,同时搭建了真实的评测平台,并在真实环境对所提出的方法评估。实验结果表明,与基线单步规划器相比,序列扩散规划器在 Gibson 和 MP3D 仿真环境中,分别提高了导航成功率(SR)7.6%和 8.7%,以及导航效率(SPL)4.2%和 3.1%。此外,在真实环境评估中,序列扩散规划器相比于基线单步规划器,表现出更显著的导航成功率提升,取得了 15.7%的性能增幅。

## 2 相关工作

### 2.1 目标导航

目标驱动导航是具身智能体在现实环境执行交互任务的前置基础,本文面向物体目标导航,其中目标是以用户指定的特定物体语义(例如,椅子)为导航目标。当前的物体目标导航的工作可以分成端到端学习方法和模块化方法这两种方法。端到端学习方法通过与环境进行交互获取来自环境的反馈来训练导航模型,采用强化学习(RL)<sup>[3,17-18]</sup>或模仿学习(IL)<sup>[6,19]</sup>来进行训练。这些方法通常将目标物体类别、视觉表示<sup>[5,20-21]</sup>和物体关系先验信息<sup>[4,22-24]</sup>进行联合编码输入到导航模型中,导航模型需要根据这些隐式编码在每个时间步预测下一步的单步动作。例如,EmbCLIP<sup>[5]</sup>利用视觉语言预训练大模型将实时的视觉观测进行编码,策略网络基于隐编码通过强化学习训练来学习动作预测。而 SP<sup>[22]</sup>通过集中统计外部数据来预先构建场景图以捕捉物体与物体之间的关联关系,然后在导航过程中将视觉观测和场景图进行编码,基于隐式编码预测需要执行的动作。端到端方法通常采用隐式的编码来记录所有过去的观察结果,但这会导致环境的结构空间信息的缺失。因此端到端学习的方法在复杂布局环境和真实环境中的泛化能力相对薄弱。

模块化方法<sup>[9,25-26]</sup>通常会在导航的过程中构建一个语义地图,来记录从导航开始以来的所有观测到的物体语义以及它们的位置信息。基于构建的语义地图,模块化方法会进一步训练一个航点规划器,规划器基于构建的语义地图预测下一时刻的目标航点。当前的模块化方法通过监督学习来训练航点规



划器,例如,PONI<sup>[10]</sup>提出了一种网络,该网络基于语义地图来预测两个互补的势函数,并利用这些势函数来决定在何处寻找一个未知物体,通过监督学习的方式对势函数网络进行训练,并将其整合到一个模块化框架中,以执行目标导航任务。而 PEANUT<sup>[27]</sup>则直接预测目标的可能位置,通过预测出概率的大小来确定下一步的航点。然而当前航点规划器仍旧是单步规划器,需要每个时刻重新预测规划,且并不考虑与先前规划的一致性。

总结来说,当前的端到端学习方法是带有隐式记忆的单步规划器,而模块化方法则是带有显式记忆的单步规划器。相比之下,本文的序列扩散规划器是一种序列规划的模块化方法,其根据导航状态(语义地图和导航目标)来预测未来应该采取的行为序列,因此它是一种带有显式记忆的序列化规划器。显式记忆确保了环境中位置空间信息的保留,而序列规划确保了决策的时序一致性和可解释性。

## 2.2 序列规划方法

序列化行为规划能够帮助智能体考虑整个任务的长期目标,从而提高决策的连贯性,在具身智能领域中的物体操作<sup>[28-29]</sup>、目标导航<sup>[6,20,30]</sup>、自动驾驶<sup>[31-32]</sup>等任务有广泛的应用。在目标导航领域,Habitat-web<sup>[6]</sup>通过在虚拟环境中,人类遥控智能体进行导航,来收集人类的导航演示的行为序列,然后通过模仿学习的方式来训练导航策略拟合收集的演示序列。ENTL<sup>[20]</sup>通过自监督学习的方式,基于Transformer的结构,以收集的演示序列在某时刻的前段演示序列输入,模型通过自监督训练,学习预测在相同时刻下后段演示序列。SPOC<sup>[30]</sup>基于从模拟器采集的最短路径进行训练,训练序列采用错位预测的方式,即输入序列与输出序列相同位置上在时序错位一个时间间隔。因此尽管 SPOC 使用序列化的预测,但其预测结果仅为后一个时刻的行为。相比于现有的序列规划方法,本文所提出的序列扩散规划器基于模拟器中收集的最短路径进行训练,相比于人类演示需要昂贵的人工成本,降低了数据收集的成本。此外序列扩散规划器的输出是未来一段时间序列的行为规划,相比于仅预测未来一个时刻的行为,增强了结果可解释性与决策的连贯性。

## 2.3 扩散模型

扩散模型(DDPMs)<sup>[13]</sup>是一种通过迭代的方式预测并去除随机分布中包含的噪声从而来学习复杂的数据分布的生成模型。扩散模型已成功应用于与图像相关的领域,包括图像生成<sup>[33-35]</sup>、超分辨率<sup>[36-37]</sup>、

图像修复<sup>[38-39]</sup>和图像编辑<sup>[40-41]</sup>。扩散模型主要使用基于 U-Net<sup>[34,36]</sup>和基于 Transformer<sup>[33,35,37]</sup>的架构。U-Net 架构通过编码器和解码器之间的跳跃连接,能够在多尺度特征之间进行高效融合,有助于恢复图像细节并保持结构信息。而 Transformer 架构利用自注意力机制能够建模全局依赖关系并捕捉长程上下文,使得它在处理复杂任务时具有更强的表达能力。近年来,扩散模型逐渐应用于机器人领域。一些工作<sup>[42-45]</sup>利用扩散模型进行数据增强,来解决现实世界的真实数据面临稀缺或缺乏多样性的问题。扩散模型还被多个研究<sup>[42-45]</sup>作为自然数据合成器用于数据增强。此外,Diffuser<sup>[46]</sup>、Crossway Diffusion<sup>[47]</sup>和 Diffusion policy<sup>[28]</sup>等方法,利用扩散模型学习智能体的多模态行为数据,从而为智能体学习更好的规划器以完成具身任务。本文所提出的序列扩散规划器旨在学习基于语义图和目标条件下的行为序列分布,主要用于解决物体目标导航任务。T-Diff<sup>[48]</sup>基于扩散模型学习智能体的轨迹预测,与其相比差异主要在于(1)模型结构,T-Diff 使用 Transformer 架构,而本文方法使用 UNet 架构,UNet 结构的计算复杂度低,便于面向实体部署,同时可以更高效地捕捉序列间多种尺度的关系;(2)输出序列,T-Diff 仅输出位置序列,而本文的方法不仅包含位置还具有动作,通过位置一动作联合学习可以增强不同模态信息直接的互补性,增强预测结果的准确性。

## 3 目标导航规划器

物体目标导航任务要求将智能体在一个未知的环境中导航到用户指定的特定类型的物体(例如“椅子”)。在每个导航回合的开始,智能体会被初始化在一个随机位置,在导航过程中的每一时刻  $t$ ,智能体通过视觉传感器接收到主视 RGB-D 观察图像  $I_t$ 、目标物体  $o$  和传感器位姿  $p_t$ ,其中位姿包括智能体的空间坐标和朝向方向。智能体可以执行的离散动作包括前进、左转、右转和停止。当智能体判断任务已完成时,会自动执行停止动作。一个成功的导航回合定义为:如果智能体在规定的步数内停止,并且目标物体位于其视野范围内,且距离小于阈值(如小于 1 米),则该回合被视为成功。

现有的物体目标导航的方法可以分为端到端学习方法和模块化方法。端到端学习方法<sup>[3,5-6,19]</sup>在每时刻进行单步规划,可以形式化地表示为  $\pi(a_t | I_t, o)$ ,其中  $a_t$  表示在  $t$  时刻应该采取的动作。这些方

法通常使用强化学习(RL)或模仿学习(IL)来训练规划器  $\pi(a_t | I_t, o)$ 。强化学习的训练目标是最大化奖励函数的期望总和。以  $\chi$  表示在回合中基于以  $\pi$  为规划器执行导航任务过程中得到的动作和奖励二元组序列,则训练目标为

$$\operatorname{argmax}_{\pi} \mathbb{E}_{\chi \sim \pi} [R_{\chi}], R_{\chi} = \sum_{t=1} \beta^{t-1} r_t \quad (1)$$

其中,  $\beta$  是衰减因子,  $r_t$  表示奖励函数,通常是稀疏奖励值,即当且仅当智能体成功完成一次导航时获得正向(如 5 分)的奖励,而没有其他获得正向奖励的方式。这是因为稠密奖励可能会抑制智能体的探索行为<sup>[7]</sup>,从而降低在未知环境中的泛化能力<sup>[19]</sup>。然而,由于智能体在训练的过程中奖励值的稀疏,大多数导航回合难以获得正向奖励,导致训练效率低。此外,随着策略参数的更新,先前导航回合采集的数据过时,需要收集额外的新数据,这导致了采样低效的问题。因此,端到端学习方法存在的样本效率低、训练计算成本高的问题。而对于模仿学习(IL)的方法<sup>[6,19]</sup>,其训练目标是最小化规划器输出与人类演示之间的差异(即行为克隆),公式表示为

$$\operatorname{argmin}_{\pi} \sum_{I_t^d, a_t^d \sim \mathcal{D}} -a_t^d \log(\pi(a_t | I_t^d, o)) \quad (2)$$

其中,  $\mathcal{D}$  是人类演示数据集。策略网络以人类演示的视觉观测  $I_t^d$  为输入,训练目标是要求策略网络的输出应对齐人类演示中人类采取的行为  $a_t^d$ 。然而由于收集人类演示是昂贵的,这使得模仿学习的训练成本较高。此外,端到端学习方法均将所有的历史观察通过隐式的方式进行编码,导航规划器仅依赖隐式编码进行决策而忽视导航环境的空间几何关系,因此制约限制了这些方法在复杂环境和真实环境下的导航泛化性。

模块化方法<sup>[9-10,27,49]</sup>通常在导航过程中构建局部语义地图来存储导航中观测的环境空间几何信息,帮助智能体充分记录已探索环境以避免冗余探索。模块化的方法基于语义地图学习航点规划器,用于规划智能体下一步的运动方向。航点规划器可以形式化地表示为  $f(\Omega | m_t, o)$ ,其中  $\Omega$  是  $m_t$  中的像素点的子集合,例如,  $\Omega$  可以定义为  $m_t$  的边界<sup>[10]</sup>或  $m_t$  外的未知区中目标的潜在位置<sup>[27]</sup>。 $f(\Omega | m_t, o)$  通常利用监督回归的方式进行训练,训练损失为

$$\operatorname{argmin}_f \sum_{\Omega^*} -\Omega^* \log(f(\Omega | m_t, o)) \quad (3)$$

其中,  $\Omega^*$  是根据目标的实际位置计算得到的用于训练的真实标签,  $\Omega^*$  的值越高表示距离目标的概率越大。基于  $f$  的预测,智能体选择距离目标概率

最大的点作为航点来引导智能体。由于模块化方法的规划器  $f$  需要在每时刻不断地重新预测,因此其也只能进行单步的运动规划。尽管监督学习相比于强化学习的训练效率更高,然而由于模拟器可以提供的训练<sup>[15-16,50]</sup>房间种类有限,例如 Gibson、MP3D 和 HM3D 等包含的房间个数均小于 1000 个,这导致与位置相关的监督约束会使得规划器  $f$  在训练房间的布局上过拟合<sup>[8]</sup>,从而影响方法的泛化性。

总结来说,当前端到端学习方法通常采用带有隐式记忆的单步规划器,其训练过程受到样本效率低和训练成本高的影响。另一方面,模块化方法采用带有几何记忆的单步规划器,由于其训练过程存在有限的、与位置相关的强监督约束,其泛化性受到限制。在本文的工作中,本文的轨迹扩散模型是一个带有几何记忆的序列规划器,其目的是学习条件分布  $p(\tau_t | m_t, o)$ ,其中  $\tau_t$  表示计划的未来行为序列。几何记忆则确保了高效的探索,序列规划确保了决策的时序一致性,并且可以显式地看出智能体的运动趋势,使得模型决策具有可解释性。此外,在模型训练方面,本文利用从不同起点到目标位置的多样化行为序列数据进行训练,同时通过分割行为序列数据进一步丰富数据的多样性,并通过 DDPM 来学习这个条件分布,其优势在于:相比于端到端强化学习方法,提升了训练样本采样的效率;相比于端到端模仿学习方法,降低了使用人类示范高成本数据的依赖;相比于模块化方法,避免了使用单一绝对位置信息导致过拟合训练环境布局,提升了模型泛化性。

## 4 序列扩散规划器

### 4.1 扩散模型

扩散模型(DDPMs<sup>[13]</sup>)是一种概率生成模型,是一种通过迭代的方式不断地去除从高斯分布中采样的变量中的噪声来学习数据分布的方法。扩散模型的前向过程被定义为扩散过程,这个过程通过马尔可夫链逐渐将高斯噪声应用于真实数据,从而将复杂分布逐步转化成高斯分布,扩散过程可以形式化表达为

$$q(x_{1:T} | x_0) = \prod_{\tau=1}^T q(x_{\tau} | x_{\tau-1}) \quad (4)$$

$$q(x_{\tau} | x_{\tau-1}) = \mathcal{N}(x_{\tau}; \sqrt{\alpha_{\tau}}, (1 - \alpha_{\tau})I)$$

其中,  $\alpha_{\tau}$  是超参数。 $x_0$  是真实数据,而  $x_1, \dots, x_T$  是带噪声的隐状态数据。隐状态数据和真实数据的维度相同,即  $x_{0:T} \in R^d$ 。通过应用重参数化技巧,即

将噪声数据分解成一个确定性部分和一个随机部分,将复杂采样过程转化为较简单的操作,以便在训练过程中能够通过反向传播来优化模型。采样带噪声的数据的过程为

$$x_\tau = \sqrt{\bar{\alpha}_\tau} x_0 + \sqrt{1 - \bar{\alpha}_\tau} \epsilon_\tau \quad (5)$$

其中,  $\epsilon_\tau \sim \mathcal{N}(0, I)$  为高斯噪声。通过扩散过程,原始数据被拓展成了更多的带噪数据。扩散过程结束后,扩散模型通过反转扩散马尔可夫链来学习数据分布,该过程表示为  $p(x_{\tau-1} | x_\tau)$ , 称为去噪过程。基于理论证明<sup>[13]</sup>,这个过程可以简化等效为预测添加到  $x_\tau$  上的噪声。噪声预测网络  $\epsilon_\theta$  使用预测噪声与真实的高斯噪声  $\epsilon_\tau$  之间的均方误差进行训练。此外,扩散模型可以引入其他额外输入(例如文本<sup>[35,40]</sup>或图像<sup>[28,51]</sup>)进行条件化的生成训练,此时训练目标可以表述为

$$\mathcal{L}_\theta = \mathbb{E}_{x, c, \epsilon, \tau} [\|\epsilon_\tau - \epsilon_\theta(x_\tau, c, \tau)\|_2^2] \quad (6)$$

其中,  $c$  是输入条件的编码。当噪声预测模型  $\epsilon_\theta$  训练完成,真实数据就可以从随机噪声开始通过不断地对输入噪声进行迭代去噪,进行逐步生成。

#### 4.2 训练数据

序列扩散规划器的目标是根据智能体的导航观测为智能体规划出从当前状态到未来的行为序列,从而帮助智能体高效地导航至目标物体。

为了训练序列扩散规划器,本文收集一系列训练数据对  $((m_t, o), \xi_t)$ , 其中  $m_t$  是时间步  $t$  时的语义地图,  $o$  表示目标物体,语义地图  $m_t \in R^{(4+n) \times h \times w}$  包含  $(4+n) \times h \times w$  个地图通道,其中 4 表示基础的地图信息通道数,包括障碍物地图、已探索区域、智能体的当前位置地图和所有先前的行为与位置地图,  $n$  表示语义类别的数量,  $h, w$  为地图的尺寸。而  $\xi_t \in R^{3 \times k} = [X_{t+1}, \dots, X_{t+k}]$  表示从时间  $t+1$  到  $t+k$  的行为序列,其中序列中的每个位姿  $X_t$  包含三维数据,其中一个维度记录了智能体的动作  $a_t$ , 另外两个维度则记录了智能体的坐标位置  $(x_t, y_t)$ 。为了获得高效训练数据,即获得成功导航回合的数据对,首先在训练房间内随机初始化智能体的一个起始位置,然后基于 FMM 算法<sup>[52]</sup> 根据当前训练房间的精确碰撞地图计算从起始位置到特定目标  $o$  的最优路径。其中 FMM 算法是一种基于数值模拟的高效最短路径计算方法,它通过模拟波的传播,能够根据碰撞地图计算两点之间的最短路径。注意这种数据收集仅在训练环境中进行,测试环境是未知的,无法获取对应的精确地图。随后驱动配备语义地图模块的智能体沿着这条最优路径前进,

并记录每个时间步的语义地图  $m_t$  和智能体位姿  $X_t$ 。进一步地,收集到的数据序列被分割成长度为  $k$  的子序列  $\xi_t$ , 这些数据对  $((m_t, o), \xi_t)$  以及对应的语义地图  $m_t$  和目标物体  $o$  最终构成了扩散模型的训练数据,收集过程如图 1 所示。

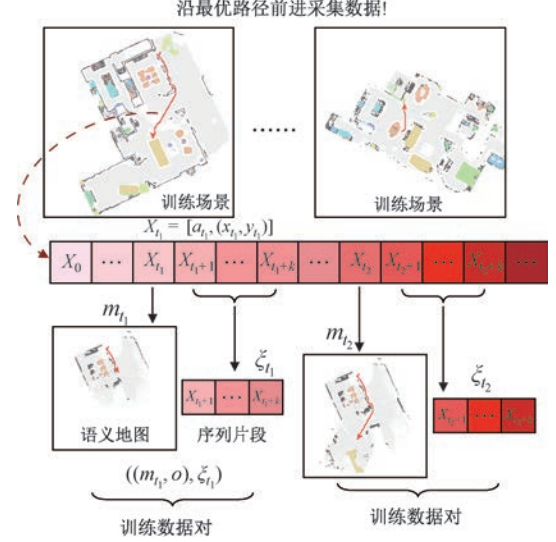


图 1 训练数据收集

#### 4.3 序列扩散规划器训练

本文提出的序列扩散规划器基于 DDPM 进行训练以估计条件分布  $p(\xi_t | m_t, o)$ , 即基于 4.1 节中介绍的 DDPM 通过扩散过程与去噪过程来使得序列扩散规划器能够在导航过程的每个  $t$  时刻,根据当前的导航状态  $m_t$  以及导航目标  $o$ , 规划出从当前时刻到未来  $k$  步的行为。在扩散过程中,通过如下公式为原始收集到的真实序列数据增加噪声:

$$\xi_t^\tau = \sqrt{\bar{\alpha}_\tau} \xi_t^0 + \sqrt{1 - \bar{\alpha}_\tau} \epsilon_\tau \quad (7)$$

其中,  $\epsilon_\tau \sim \mathcal{N}(0, I)$  表示高斯噪声,而  $\xi_t^0$  是真实数据  $\xi_t$ , 即在收集的数据对  $((m_t, o), \xi_t)$  中的序列片段。 $\bar{\alpha}_\tau$  和  $\tau$  是用来控制扩散过程调度的超参数,它控制着扩散过程的强度。注意这里涉及两个时间步 ( $t$  和  $\tau$ ), 其中  $t$  表示导航中的某一时刻,而  $\tau$  表示扩散或去噪过程中的过程调度参数。序列扩散规划器的训练目标是准确预测出添加在真实数据中的噪声,基于公式 (6), 序列扩散规划器的训练损失函数可以形式化地表示为

$$\mathcal{L}_\theta = \mathbb{E}_{\xi, s, \epsilon, \tau, t} [\|\epsilon_\tau - \epsilon_\theta(\xi_t^\tau, s_t, \tau)\|_2^2] \quad (8)$$

其中,  $\theta$  表示序列扩散规划器的模型参数,而  $s_t$  表示序列扩散规划器的规划条件,  $s_t$  是语义地图  $m_t$  和目标物体  $o$  的编码的联合表征。语义地图  $m_t$  由 ResNet18 网络进行编码,取全局池化后的尺度为  $1 \times 512$  的编码,而目标物体  $o$  首先转化成 one-hot



向量编码,然后将两种编码进行拼接,拼接后的编码通过用线性层网络进行编码作为序列扩散规划器的生成条件。

序列扩散规划器是基于 U-Net 扩散模型构建,如图 2 所示。U-Net 扩散模型将 U-Net 和扩散模型结合起来,具有对称的结构,由下采样路径和上采样路径组成,下采样(编码阶段)来捕捉多层次的噪声轨迹特征,然后通过上采样(解码阶段)逐步恢复输入的空间尺寸,最终实现生成未来行为序列。在下采样阶段,编码器逐步降低带噪输入序列特征的空间维度,同时增加特征通道数,具体通道数依次为 256、512 及 1024。每一采样层采用步幅卷积操作以减小特征图尺寸并增加通道数,从而增强轨迹特征的表达能,其中步幅卷积是指在进行卷积操作时,使用了大于 1 的步幅,导致卷积核跳过某些位置而不是逐一处理每个位置,使得每经过一层卷积时,输入特征图的空间尺寸(如高度和宽度)逐渐减小。在上采样阶段,解码器负责恢复带噪输入序列特征的原始空间维度。通过卷积操作进行上采样,逐步恢复特征图的空间尺寸,并采用跳跃连接的方式,在通道维度上将对下采样阶段的特征图进行拼接,以融合多层次的特征信息。此过程逐步恢复轨迹的空间细节与精度,确保生成轨迹的平滑性和连续性。此外,为了引入导航目标以及语义地图的编码作为序列规划生成的控制条件,并将其通过交叉注意力层嵌入至 U-Net 的各个层级中,交叉注意力层将导航目标和语义地图的信息融入生成的行为序列中,通过将来自不同输入源的信息进行融合,控制扩散过程中的行为序列的生成与预测。

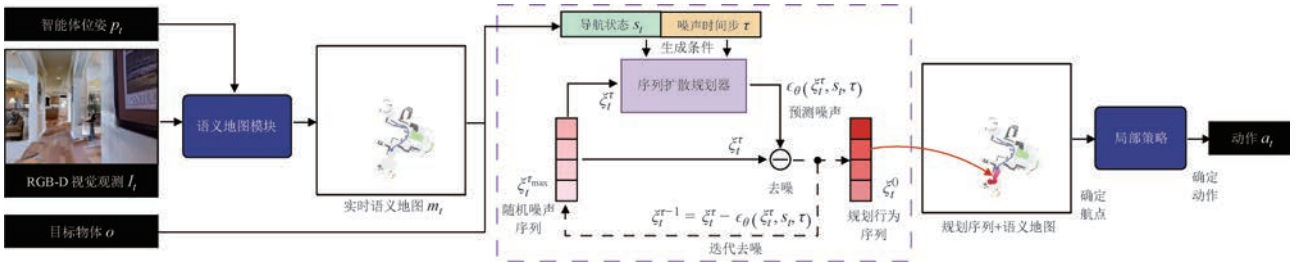


图 3 基于序列扩散规划器的导航框架图

物体  $o$  通过图 2 所示的方式进行编码,然后编码特征拼接在一起,形成导航状态  $s_t$ 。在训练阶段,训练好的序列扩散规划器以初始化的高斯噪声  $\xi_t^{\tau_{\max}}$  作为初始输入,以导航状态和噪声时间步为条件,预测包含在带噪声潜在数据  $\xi_t^\tau$  中的噪声  $\epsilon_\theta(\xi_t^\tau, s_t, \tau)$ ,带噪声的潜在行为序列通过以下公式进行

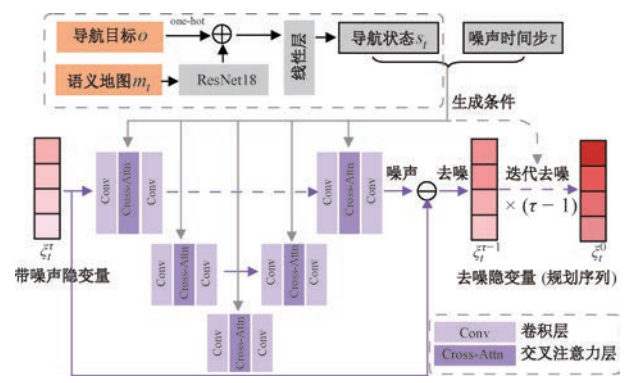


图 2 序列扩散器模型结构图

#### 4.4 基于序列扩散规划器的导航

本节将介绍智能体在未知环境中,借助序列扩散规划器规划应该采取的行为,逐步完成目标导航任务的过程。在导航过程中的每时刻  $t$ ,智能体通过 RGB-D 视觉传感器接收当前主视 RGB-D 观察图像  $I_t$ 、当前智能体位姿  $X_t$ ,以及目标物体  $o$ ,基于序列扩散规划器的导航框架如图 3 所示,其中智能体通过语义地图模块在导航的过程中构建一个逐步增长的语义地图  $m_t$  用以记录已探索过的区域布局 and 结构信息。语义地图模块接收在导航过程的每时刻的 RGB-D 图像为输入,然后将图像通过预训练的分割模型<sup>[53]</sup>获得语义分割结果。随后,基于深度图像中的深度信息,将每个 RGB 像素及其语义标签投影到三维点云空间中。此外,低于智能体高度的点云视为障碍物。然后,这些点云被从上到下投影转换为体素占用网格,从而创建一个以自我为中心的地图。根据智能体的位姿,以自我为中心地图与现有的全局地图进行聚合。语义地图  $m_t$  和目标

去噪:

$$\xi_t^{\tau-1} = \xi_t^\tau - \epsilon_\theta(\xi_t^\tau, s_t, \tau) \quad (9)$$

去噪过程重复进行  $\tau_{\max}$  步,逐步迭代生成出最终的规划序列  $\xi_t^0$ 。

当规划序列  $\xi_t^0$  (即  $\xi_t$ ) 生成完毕,基于局部策略<sup>[9,26]</sup>驱动智能体沿着该规划序列移动。在应用该

规划序列时,实验发现,若选择  $\xi_t$  上过于靠近  $t$  的点会导致序列规划频率过高,使得计算复杂度大;而选择过于靠近  $t+k$  的点会导致出现规划失误时错误引导时间过长,使得智能体陷入错误导航方向太长,因此为了平衡两种影响,本文选择  $\xi_t$  的  $1/2$  处作为导航阶段性航点。当阶段性航点确定后,以语义地图中的障碍物和可通行区域地图为基础,通过 FMM 算法计算出由当前位置到航点的无碰撞路径,然后按照智能体的动作尺度将路径转化为智能体需要执行的动作序列,智能体将执行这些动作直到行进到航点附近。之后序列扩散规划器会重新生成一个新的轨迹,而局部策略则在导航的每一步重新规划确定性动作,直到驱动智能体导航到目标。

## 5 实验

### 5.1 实验设置

#### 5.1.1 数据集

为了与其他方法进行公平比较,本文在公开标准的物体目标导航数据集上进行测试。本文基于 Habitat 模拟器平台的 Gibson<sup>[15]</sup> 和 Matterport3D (MP3D)<sup>[16]</sup> 数据集上评估序列扩散规划器的性能。对于 Gibson 数据集,本文使用来自 tiny-split<sup>[10]</sup> 的 25 个训练场景和 5 个验证场景进行实验验证,其中测试验证场景中包含 1000 个验证导航回合,涵盖 6 种类型目标物体。对于 MP3D 数据集,本文使用 56 个训练场景和 11 个验证场景,其中包含 2195 个验证导航回合,涵盖 21 类目标物体。此外,本文还构建了真实的现实环境,对本文方法部署在实体机器人中进行评估。

#### 5.1.2 实现细节

在训练轨迹扩散模型时,本文分别从 Gibson 和 MP3D 的训练场景中采样了 84k 和 465k 组语义地图-轨迹数据对(其中数据中 85% 用于训练,15% 用于验证)。为了输入到序列扩散规划器中,语义地图被放缩为  $224 \times 224$  的尺寸,然后通过 ResNet-18<sup>[54]</sup> 编码,其中第一个卷积层的输入通道被调整以适应语义地图的维度。本文基于 U-Net<sup>[55]</sup> 结构实现序列扩散规划器。训练使用 AdamW 优化器<sup>[56-57]</sup>,基础学习率为  $1e-4$ ,前 1000 步使用线性预热步骤,在预热步骤后扩散模型的学习率衰减因子为  $1e-3$ ,语义地图编码器的学习率衰减因子为  $1e-6$ 。模型总训练轮次为 200 次。

扩散训练过程参考常见的扩散模型的训练设置

过程,训练过程中使用指数移动平均(EMA),其中最大衰减为 0.9999,最大噪声调度  $\tau_{\max}$  设置为 100。预测序列的长度  $k=32$ 。遵循先前导航算法的实验设置,智能体的转弯角度设置为 30 度,直行步长的距离设置为 25 cm。导航过程中最大时间步数限制为 500,超过这个时间步智能体会被强制停止,并被判定为失败的导航回合;导航过程中序列扩散规划器的规划时间间隔  $t_{T-\text{diff}}$  设置为 5。实验中所汇报的实验结果均是 3 次实验评估的性能平均值,此外由于多次重复实验的结果中的导航性能波动并不明显,因此实验结果中没有标注方差。

#### 5.1.3 评测指标

为了量化在物体目标导航任务中的性能,本文参考先前导航工作<sup>[9-10,27]</sup>,使用了以下指标进行性能评估。

SR(成功率)表示成功回合占总导航回合的比例。其定义为

$$SR = \frac{1}{N} \sum_{i=1}^N S_i \quad (10)$$

其中,  $N$  是验证试验的总次数,  $S_i$  是一个指示变量,表示第  $i$  次试验是否成功。

SPL(路径长度加权成功率)表示成功回合的成功率按路径最优路径与实际路径比值作为系数进行加权平均,其定义为

$$SPL = \frac{1}{N} \sum_{i=1}^N S_i \frac{l_i^*}{\max(l_i, l_i^*)} \quad (11)$$

其中,  $l_i^*$  表示由模拟器计算出的最短路径长度,  $l_i$  指的是第  $i$  次试验的实际路径长度。这一指标相比于成功率,更进一步地反映了智能体的路径相对于最短路径的差距,反映了导航轨迹的效率。

DTS(到目标的距离)表示回合结束时,智能体停止的位置与目标之间的距离,其定义为

$$DTS = \frac{1}{N} \sum_{i=1}^N \max(L_{i,g} - \xi, 0) \quad (12)$$

其中,成功阈值  $\xi = 1m$ 。当某次试验成功时,到目标的距离值为 0。

此外,本文使用 MSE(均方误差),来衡量生成的序列与最优序列真值之间的距离,其定义为

$$MSE = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{x}_i)^2 \quad (13)$$

其中,  $n$  是收集的标准规划序列数据对的验证集数量。  $x_i$  指的是真实轨迹,  $\hat{x}_i$  是生成的轨迹。MSE 用于评估本文所提出的序列扩散规划器输出序列的准确性。



## 5.2 虚拟环境下测试结果

### 5.2.1 地图放缩尺寸的消融实验

实时的语义地图被放缩到指定大小以输入到序列扩散规划器中,由于地图的大小反映了对环境记录的精度,即地图尺寸越大记录环境的细节越精细,本文探究了输入到序列扩散规划器中不同地图放缩大小对导航性能以及计算复杂度的实验,结果如图4所示。实验结果表明,除了地图过小的情况以外,序列扩散规划器可以兼容不同尺寸的语义地图,同时均具有良好的导航性能。此外,随着地图大小的增加,记录信息的粒度提升,对应的导航性能也提升,这说明了越精细的地图可以帮助取得更好的导航性能,然而与此同时模型的计算复杂度也随之增加,根据实验结果,当地图大小超过300时,导航性能趋于平稳,而计算复杂度持续上升。为了平衡性能和计算复杂度,在模拟器评测实验中本文选择了 $224 \times 224$ 的地图放缩尺度大小。需要注意的是,当序列扩散规划器需要适应大规模场景时,该规划器可以兼容更大尺寸的地图输入,以适应更复杂环境。

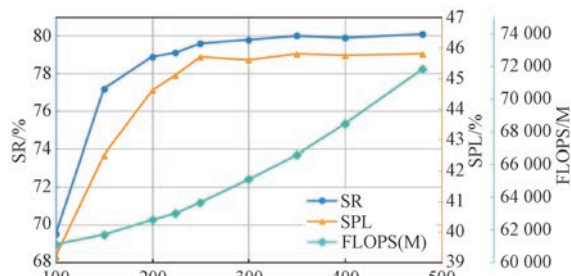


图4 不同地图尺度大小的导航性能与计算复杂度

### 5.2.2 规划目标切换频率的实验评估

由于本文提出序列扩散规划器的出发点是为了确保了规划的时序一致性,这里的时序一致性是指,规划出的下一步动作应该考虑到先前的规划结果,保持规划的一致性,防止规划的目标在空间中大幅度地切换,导致智能体由于规划目标大幅度频繁切换而来回摆动,使智能体运动趋势频繁变化,最终使得导航效率较低。为了衡量运动趋势的频繁变化,本文采用平均轨迹曲率作为统计指标来评估这些变化。形式化地,给定智能体在一次导航回合的轨迹 $\tau = [p_0, p_1, \dots, p_T]$ ,其中 $p_t = (l_t, \theta_t)$ , $l_t$ 是2D位置坐标, $\theta_t$ 是智能体的朝向角度。平均轨迹曲率定义为

$$\kappa = \mathbb{E}_{\tau} \left[ \sum_{t=1}^T \frac{\theta_t - \theta_{t-1}}{1 + \|l_t - l_{t-1}\|_2} \right] \quad (14)$$

其中, $\kappa$ 的单位为 rad/m。本文在 Gibson 的测试集

上使用相同的初始位置和目标对端到端学习方法(DD-PPO<sup>[58]</sup>和 EmbCLIP<sup>[5]</sup>)和序列扩散规划器(SDP)进行评估,其中 DD-PPO 采用分布式的强化学习训练策略,端到端训练导航模型,EmbCLIP 采用 CLIP 模型进一步改进模型中的视觉编码器,同样采用强化学习端到端训练导航模型。实验结果如表1所示:统计结果表明,端到端方法表现出更大的运动趋势变化,支持了本文关于端到端方法比 SDP 更频繁地切换目标的研究动机,而本文所提出的序列扩散规划器能够在规划结果上表现出更好的时序一致性。

表1 规划器规划目标切换的实验评估

方法	$\kappa$ (rad/m)	SR (%)	SPL (%)
DD-PPO <sup>[58]</sup>	179.07	15.0	10.7
EmbCLIP <sup>[5]</sup>	121.78	68.1	39.5
SDP	63.98	81.2	45.2

### 5.2.3 轨迹生成的可视化结果

本文可视化了扩散模型以循环迭代的方式生成规划序列的过程,如图5所示。最初,规划序列被初始化为随机的坐标点,随着去噪过程的进行,这些点逐渐聚集,最终形成了在 $\tau=0$ 时的规划序列。需要注意的是,尽管展示的可视化是在 Gibson 的验证集上进行,即场景布局对是未知的,但可视化结果表明,序列扩散规划器(SDP)生成出的规划序列是准确的,即规划出的路径是朝向目标物体的,即使目标物体当前是并不可见的。在可视化图中真实的目标位置轨迹扩散方法仍然能够生成最有效的目标路径,如真实轨迹图中标记的目标位置所示。

### 5.2.4 规划序列长度的消融实验

本文针对序列扩散规划器输出的规划序列的长度进行了消融实验。当使用最小轨迹长度,即设置为1时,序列扩散规划器的输出为单步规划,在这种情况下序列扩散规划器退化成为一般模块化方法的规划方式,即单步规划方式。实验结果如表2所示,当长度设置较短时,例如1或4,相应的导航性能较差,这是由于相比于单步或者较短长度的规划,长度合适的预测序列可以让每个预测点获得来自邻近点的上下文信息。这有助于纠正和抚平单个点的预测误差,减少结果对单点误差的敏感性,从而保证了更稳定的预测,并提高了轨迹预测的整体准确性。当规划序列过长时,一方面规划序列的精度下降,如实验结果表明在长度为40时,MSE指标下降,表明规划序列导致出现误差;另一方面规划序列过长时,如果出现错误预测,则对智能体导航过程影响时间会

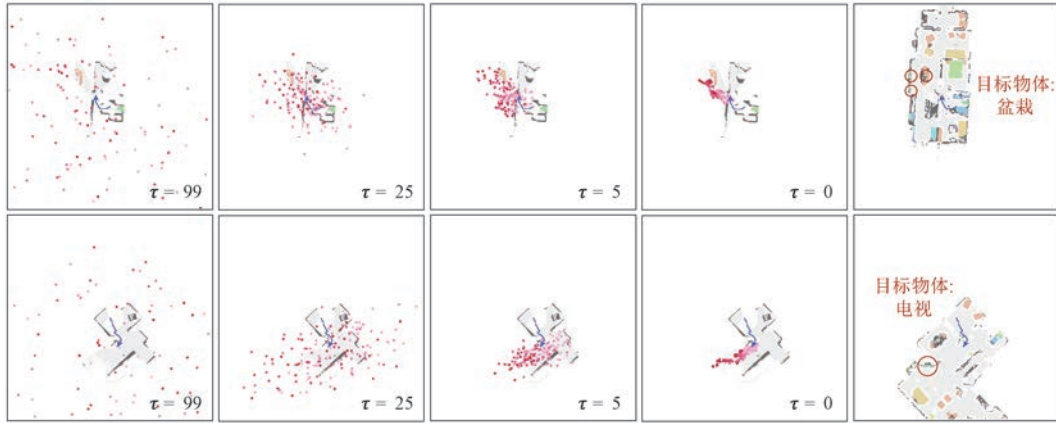


图5 规划序列的迭代去噪生成过程

表2 Gibson (val)上序列扩散规划器(SDP)规划序列长度消融实验

ID	方法	规划序列质量		导航性能(Gibson)		
		序列长度	MSE	SR(%)	SPL(%)	DTS(m)
1	单步规划方法(PONI <sup>[10]</sup> )	—	—	73.6	41	1.25
2	SDP	1	0.2289	74.0	42.2	1.41
3	SDP	4	0.1937	75.7	42.6	1.29
4	SDP	8	0.1876	76.2	43.1	1.26
5	SDP	16	0.1565	78.5	44.1	1.17
6	SDP	24	0.1022	79.7	44.8	1.09
7	SDP	32	0.0835	<b>81.2</b>	<b>45.2</b>	<b>0.99</b>
8	SDP	40	0.0993	79.8	44.7	1.21

增加,这会导致导航系统对于序列扩散规划器的容错程度降低,导致导航性能的下降。因此实验结果表明,过小或者过大的预测序列均不合适,通过实验整定,本文最终选择序列长度为32。

#### 5.2.5 训练方式的消融实验

本研究使用扩散模型的出发点是语义地图与规划序列均是高维的,直接学习这种分布是困难的,因而选择使用扩散模型来学习从语义地图到规划序列的分布。为了证明这一动机的合理性,本文考虑了以下简单的解码器来学习轨迹生成  $P(\tau_t | m_t, o)$  作为对比,它采用了与序列扩散规划器(SDP)相同的参数网络,并且使用相同的条件输入。然而,与通过 DDPM 训练的 SDP 不同,这个对照模型仅使用 MSE 损失进行训练,即单步直接生成结果,而非 DDPM 的逐步生成。实验结果如表3所示,通过监督训练直接学习  $P(\tau_t | m_t, o)$  的方法表现较差。本文中由于  $m_t$  和  $\tau_t$  都是高维的,目标分布  $P(\tau_t | m_t, o)$  的维度会更复杂,此外鉴于训练房间数量有限(少于100),  $P(\tau_t | m_t, o)$  是稀疏的,难以通过单步直接生成。相比之下,扩散模型(DDPM)通过其扩散和去噪过程,逐渐将复杂的分布简化为多个较简单的分布的叠加,模型的学习过程从直接学习复

杂分布转化为仅需要学习噪声预测,这使得能够更好地学习  $P(\tau_t | m_t, o)$  分布。

表3 与更简单的模型进行轨迹生成比较

方法	序列质量		导航性能	
	MSE	SR ( % )	SPL ( % )	DTS ( m )
SDP(监督训练)	0.5993	59.4	33.6	2.04
SDP(扩散生成)	0.0349	81.2	45.2	0.99

综上实验和分析验证了本文使用扩散模型学习序列规划生成的必要性。

#### 5.2.6 目标位置多样性实验

由于在真实场景中,物体的位置是灵活多变的,本文验证了目标物体位置打乱,特别是处于并不好预测的情形下时,所提出序列扩散规划器的鲁棒性。本文选择可编辑的 ProcTHOR<sup>[59]</sup> 进行实验,实验中选择了全新的10个测试场景进行验证,这些场景对于智能体来说是未知的。

实验结果如表4所示,由于初始场景中物体的摆放比较便于寻找,因此在物体打乱程度为0时,序列扩散规划器取得了良好的导航表现。而随着物体打乱程度的增加,由于其位置是随机打乱的,其原本的布局规律被破坏,因此导航的性能逐步下降。但如图7的序列扩散规划器失败样例可视化结果(具

体分析见 5.2.10 节)展示的,序列扩散规划器可以及时调整,因此导航成功率下降并不明显,而导航效率指标考虑了路径长度,由于序列扩散规划器需要进行导航调整,导致路径效率降低。总体来说当打乱程度小于 60% 时,实验结果性能较好,验证了序列扩散规划器的鲁棒性。

表 4 目标位置多样性实验

方法	SR/%	SPL/%
目标位置打乱 0%	69.3	36.8
目标位置打乱 20%	66.1	35.2
目标位置打乱 40%	62.3	32.9
目标位置打乱 60%	57.0	24.8
目标位置打乱 80%	42.7	19.9

### 5.2.7 跨模拟器泛化性测试

为验证序列扩散规划器针对不同环境布局的泛化性,本文设置了训练和测试在不同模拟环境的验证实验。实验结果如表 5 所示,其中第一行实验结果中智能体在 Gibson 上进行训练而在 MP3D 中进行测试,而第二行实验结果中,智能体的训练和测试均在 MP3D 中进行。实验结果表明,尽管在不同模拟环境中进行训练和测试时,序列扩散规划器的性能略低于训练和测试在相同模拟环境时的性能,但性能损失较少。由于训练过程中,智能体的初始位置的多样,以及通过对行为序列的裁剪,丰富了数据的多样性,这些增强模型的泛化能力。实验表明序列扩散规划器在跨模拟器的实验设置中依然可以保持良好的泛化性。

表 5 序列扩散规划器跨模拟器泛化性测试

训练	测试	SR/%	SPL/%
Gibson	MP3D	38.7	14.4
MP3D	MP3D	40.5	15.2



图 6 基于序列扩散规划器的导航过程

### 5.2.10 序列扩散规划器失败样例可视化

为了探究序列扩散规划器在出现预测失误时对

### 5.2.8 在更大规模模拟器中泛化性评估

为评估序列扩散规划器在更大规模、更多样场景的泛化性,本文选择 ProcTHOR<sup>[59]</sup> 进行进一步泛化性评估,实验结果如表 6 所示。所有的方法均只在 Gibson 数据集上进行训练,测试的 100 个场景均是先前没有见过的,实验结果表明,单步预测的方法,如 SemExp、PONI、SGM,在更多样的环境中相比于表 9 (具体分析见 5.2.14 节)展示的性能较低,表明单步预测方法的泛化性略差,而本文所提出的序列扩散规划器 SDP 展现出更好的泛化性,在表 6 所示的测试中,性能在 SR 指标提升 2.3%,在 SPL 指标提升 0.3%。

表 6 在大规模模拟器 ProcTHOR 上泛化性测试

方法	SR/%	SPL/%
SemExp <sup>[9]</sup>	57.7	29.5
PONI <sup>[10]</sup>	60.1	33.8
SGM <sup>[11]</sup>	63.0	33.9
SDP	65.3	34.2

### 5.2.9 导航过程可视化

基于序列扩散规划器(SDP)的导航过程,如图 6 所示,其中第一行展示了智能体的第一人称 RGB 观测图像,第二行则展示了在导航过程中构建的语义地图及通过序列扩散规划器(SDP)生成的规划序列;最右侧图像展示了测试场景的真实地图(需要注意这个地图对智能体而言是不可知的),其中标出了智能体的实际轨迹和目标位置。根据右侧的真实地图和目标位置,可以看出,在导航过程中,由序列扩散规划器生成的规划序列始终是正确和高效的(生成的序列是从当前位置到目标的最优轨迹),即使在目标物体在初期不可见( $T < 55$ )的情况下也可以预测出朝向目标位置的规划序列,从而帮助智能体在生成的规划序列的引导下高效地找到目标。

智能体导航的影响,本文提供了序列扩散规划器做出错误预测并导致探索错误方向的可视化样例,如



图 7 所示。如图所示,在导航过程中,当序列扩散规划器规划出错误的方向时,智能体可能会进入错误的房间。然而,由于本文选择通过控制规划长度来

防止智能体在错误的路径上停留过长时间,并设置适当的更新频率,根据持续更新的局部地图生成新的轨迹,从而纠正之前的错误规划。



图 7 序列扩散规划器失败样例可视化

如图 7 所示,在导航的初始步骤中,由于环境观察有限,序列扩散规划器规划出了错误的方向。然而,随着更多环境信息通过语义地图被观测到,新的正确轨迹会通过更新频率覆盖先前的错误轨迹。同时由于规划序列长度控制在合理的范围内,因此,智能体可以迅速调整其方向(见图中的标注)避免长时间陷入错误规划方向的引导。可视化结果证明了本文的序列扩散规划器对于单步预测错误的鲁棒性。

5.2.11 与其他序列规划方法的比较

本文选择 Habitat-web<sup>[6]</sup> 和 ENTL<sup>[20]</sup> 这两种在目标导航领域同样采用序列规划的方法进行比较。序列扩散规划器与他们的差异在于以下几个方面:

(1)训练数据。Habitat-web 和 ENTL 都依赖于人类示范轨迹进行训练,这些数据的获得需要耗费昂贵的人力物力。相比之下,序列扩散规划器使用自动收集的轨迹,可以降低数据收集成本。

(2)模型结构。序列扩散规划器的预测是基于语义地图进行,然而 Habitat-web 和 ENTL 仅依赖于从几个相邻步骤的自我中心视角的隐式编码来进行序列规划。语义地图几何地保留了所有的历史观

察数据,而仅依赖自我中心视角的隐式编码会导致几何空间信息的丢失,从而限制了它们在复杂环境中的表现。

(3)导航性能。表 7 展示了与其他序列规划方法在 MP3D 上的导航性能比较结果。结果表明,序列扩散规划器(SDP)的导航性能在与其他序列规划方法的比较中表现更优,特别是相较于 ENTL,有显著的提升。

表 7 与其他序列规划方法的比较

方法	SR/%	SPL/%
ENTL <sup>[20]</sup>	17.0	5.0
Habitat-web <sup>[6]</sup>	35.4	10.2
SDP	40.5	15.2

5.2.12 计算复杂度和训练数据量比较

由于序列扩散规划器采用扩散生成进行训练,为评估其训练开销以及数据量的大小,本文比较了端到端学习方法(如 DD-PPO<sup>[58]</sup> 和 Habitat-web<sup>[6]</sup>)、模块化方法(如 PONI<sup>[10]</sup>)以及序列扩散规划器,结果如图 8 所示。根据对比结果可以看出,序列扩散规划器的训练消耗和训练样本数量和与模块化方法

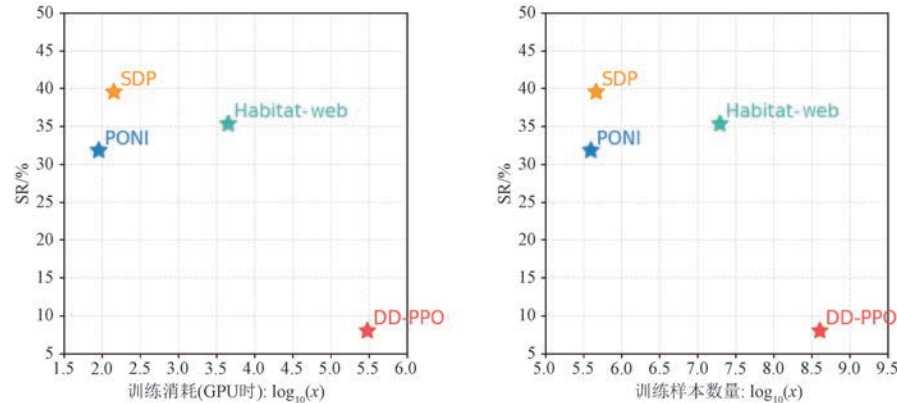


图 8 计算复杂度和训练数据量比较

PONI 相似,而显著低于端到端方法如 DD-PPO 和 Habitat-web 的消耗。

### 5.2.13 比较人类演示数据与本文采样数据

本文方法使用在仿真器中的基于精确地图采样的最短路径数据,其路径轨迹相比于人类演示数据的轨迹更短。而由于人类在收集阶段中也缺少全局地图,人类演示轨迹中包含更多探索过程,使得其轨迹更丰富和多样。

对于基于模仿学习的 Habitat-web<sup>[6]</sup> 方法,其训练过程模拟从初始坐标到结束的完整马尔科夫链,并且每一步的输入均是一个包含之前时刻所有信息的同维度隐式编码向量。因此,对于模仿学习方法,人类演示数据的轨迹更长,且包含更多探索信息,可以帮助学习更多样的行为序列,并且可以弥补隐式编码的有限状态表征,相较于仿真采集可以获得更好的实验效果,如表 8 所示。而对于本文所提出的序列扩散规划器方法,其学习的是由局部语义地图和目标到轨迹片段的分布,一方面由于局部语义地图提供结构化的显式描述,因此模型更易学习什么时候偏向探索(当局部语义地图信息较少时)、什么时候偏向导航。其次,由于序列扩散规划器输出是轨迹的片段,因此受收集数据完整链路长短影响小。综合上面两个原因,如表 8 所示,基于不同数据训练时,本文方法差异不明显,但考虑到人为演示的收集成本,因此本文方法在数据收集方面更具优势。

表 8 人类演示数据与本文采样数据的比较

方法	人类演示	仿真采样	SR/%	SPL/%
Habitat-web <sup>[6]</sup>	✓		35.4	10.2
		✓	20.7	6.9
SDP	✓		41.2	15.0
		✓	40.5	15.2

### 5.2.14 与相关目标导航工作的比较

本文将所提出的序列扩散规划器与当前目标导航最先进的方法在标准公开模拟器中进行性能比较,所选取的方法包括端到端的<sup>[5-7,20,58,60-62]</sup>和模块化的<sup>[9-10,25,63-67]</sup>方法。需要注意的是,一些方法通过引入额外信息<sup>[6-7,59]</sup>或辅助任务<sup>[58,68]</sup>,比如在原模拟基础上对物体的布局进行删改,以扩充数据,或者通过引入除目标导航之外的如探索、图像导航等额外任务以获得更多监督信号。相比于本文仅利用训练模拟器并且仅使用目标导航训练监督,这些方法引入额外信息会导致不公平的比较。因此,本文主要

关注以下基准:SemExp<sup>[9]</sup>,PONI<sup>[10]</sup>,OVG-Nav<sup>[62]</sup>,L3MVN<sup>[67]</sup>,L2M<sup>[64]</sup>和 SSCNav<sup>[61]</sup>。

PONI 通过引入监督学习来预测与目标相关的信息,从而增强了 SemExp 的能力。OVG-Nav 使用语义拓扑地图为智能体规划高层次的子目标节点。L3MVN 利用大语言模型(LLM)基于当前边界的语义信息推断未知区域。L2M 和 SSCNav 通过学习构建单时间步的自我中心地图来提升导航性能。由于 L2M 和 SSCNav 仅在其自定义数据集上报告结果,本文使用其他工作中的复现结果<sup>[66]</sup>或本文复现的结果进行比较。

本文在 Gibson 和 MP3D 的验证集上将序列扩散规划器与这些方法进行了比较,结果如表 9 所示。在 Gibson 数据集上,序列扩散规划器相较于当前最先进的方法,在 SR、SPL 和 DTS 指标上分别提高了 2.7%、6.1%和-0.01m。在 MP3D 数据集上,序列扩散规划器在相同指标上相较于当前最先进的方法,分别提高了 3.8%、2.9%和-0.53m。实验结果验证了序列扩散规划器在目标导航任务中的性能优势。

### 5.2.15 与基于大模型的导航方法的比较

本文选择 ESC<sup>[69]</sup>、VLFM<sup>[70]</sup>和 SG-Nav<sup>[71]</sup>这三种基于大模型的导航方法进行比较,如表 10 所示。其中,ESC 和 SG-Nav 通过提问大语言模型(Deberta v3、LLaMA-7B)当前观察来确定下一步探索方向来进行导航,而 VLFM 利用视觉语言预训练大模型(BLIP)计算每时刻视觉观察与目标语义相似度来确定下一步探索方向来进行导航。这些方法依赖于大模型作为知识,但每一步的行为规划是单步的,而本文所提方法 SDP 通过预测序列化行为规划,因此在导航性能方面,本文的导航性能与基于大模型的性能接近,甚至取得了更好的导航性能。而对于计算效率,本文以 VLFM 的在导航每步的平均浮点运算次数(FLOPs)为基准,比较其他方法与其相比在计算复杂度的比率,结果如表 10 所示。本文的方法在计算效率上与基于大模型的方法相比具有显著的更高效率。

## 5.3 真实环境下的测试结果

### 5.3.1 环境布局

除了虚拟仿真器,本文还在真实环境中评估了序列扩散规划器的性能。如图 9 所示,本文基于 140 平方米的开放空间进行实验,该空间通过一些可移动的墙体进行分隔,这些墙体由多个带有轮子的隔板组成。通过这些可移动墙体,空间被划分为

表 9 与相关工作的比较

ID	方法	Gibson			MP3D		
		SR/%	SPL/%	DTS/m	SR/%	SPL/%	DTS/m
I	Random	0.4	0.4	3.89	0.5	0.5	8.05
	DD-PPO <sup>[58]</sup>	15.0	10.7	3.24	8.0	1.8	6.94
	Red-Rabbit <sup>[60]</sup>	—	—	—	34.6	7.9	—
	THDA <sup>[7]</sup>	—	—	—	28.4	11.0	5.62
	SSCNav <sup>[61]</sup>	—	—	—	27.1	11.2	5.71
	EmbCLIP <sup>[6]</sup>	68.1	39.5	1.15	29.2	10.1	5.40
	Habitat-web <sup>[6]</sup>	—	—	—	35.4	10.2	—
	ENTL <sup>[20]</sup>	—	—	—	17.0	5.0	—
	OVG-Nav <sup>[62]</sup>	—	—	—	35.8	12.3	5.69
II	FBE <sup>[63]</sup>	64.3	28.3	1.78	22.7	7.2	6.70
	ANS <sup>[26]</sup>	67.1	34.9	1.66	27.3	9.2	5.80
	SemExp <sup>[9]</sup>	71.1	39.6	1.39	28.3	10.9	6.06
	PONI <sup>[10]</sup>	73.6	41.0	1.25	31.8	12.1	5.10
	L2M <sup>[64]</sup>	—	—	—	32.1	11.0	5.12
	Stubborn <sup>[65]</sup>	—	—	—	31.2	13.5	5.01
	3D-aware <sup>[66]</sup>	74.5	42.1	1.16	34.0	14.6	<b>4.78</b>
	L3MVN <sup>[67]</sup>	76.9	38.8	1.01	—	—	—
	SGM <sup>[11]</sup>	78.0	44.0	1.11	37.7	14.7	4.93
	T-Diff <sup>[48]</sup>	79.8	45.0	1.02	39.7	15.0	5.19
序列扩散规划器(SDP)		<b>81.2</b>	<b>45.2</b>	<b>0.99</b>	<b>40.5</b>	<b>15.2</b>	4.98

表 10 与基于大模型的导航方法进行比较

方法	导航性能			计算效率
	SR/%	SPL/%	DTS/m	FLOPs
ESC <sup>[69]</sup>	70.3	39.8	1.41	6.45×
VLFM <sup>[70]</sup>	77.6	42.7	1.12	1.00×
SG-Nav <sup>[71]</sup>	78.1	44.3	1.06	10.37×
SDP	81.2	45.2	0.99	0.05×

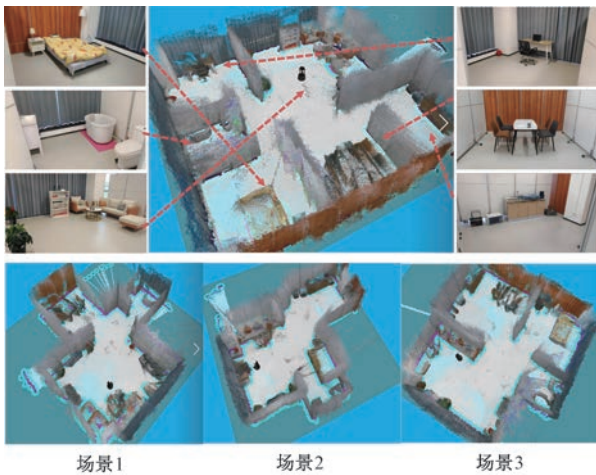


图 9 真实室内导航环境

多个房间,并且通过布置常见的家具,这些房间被安排成多个场景,例如客厅、卧室、厨房和浴室。

每个房间的面积大约为 20 平方米,每个场景中包含了各种家居物品,包括可移动墙体、电视、盆栽植物、时钟、微波炉、杯子、盘子、篮球等。总共有 35 种不同类型的物品。通过改变可移动墙体和物品的

位置,本文构建了 3 种具有不同布局的环境,目标物体在不同的场景布置不同,智能体随机初始在不同位置,本文这个真实环境中评估序列扩散规划器性能。

### 5.3.2 实现细节

为了实际部署导航算法,本文使用开源实体机器人 Locobot 作为实体智能体。Locobot 的型号为 Locobot-wx250s,其配备了 Kobuki YMR-K01-W1、Intel NUC 迷你 PC 和 Intel RealSense 深度摄像头 D435。除实体机器人外,本文还利用一台远程计算机通过无线局域网实时监控机器人状态。Locobot 和远程计算机均运行 ROS(noetic,适用于 Ubuntu20.04)。在导航的每一时刻,机器人接收来自 RealSense 摄像头的 RGB-D 图像和每个时刻的传感器位姿。传感器位姿通过开源的 RGB-D SLAM 解决方案 RT-AB-Map 进行估计。RGB-D 图像和每时刻的位姿通过建图模块来构建语义地图。当目标不可见时,机器人采用导航算法,如 PONI<sup>[10]</sup>或序列扩散规划器,进行导航探索,直到目标被观察到;当目标可见时,模型以语义地图中目标的位置设定为目标位置,使用 Dijkstra 算法规划从当前位置到目标的路径。

本文选择了 6 个物品类别(即马桶、时钟、电视、瓶子、沙发和盆栽)作为目标。对于每个场景中的目标物品,将机器人放置在 5 个不同的起始点。当机



机器人执行“完成”(Done)动作时,记录以下数据:

(1)目标物品是否出现在机器人的最终视图中;(2)目标物品与机器人最终位置之间的距离;基于这些数据,计算每次导航回合是否成功,从而计算成功率(SR)指标。

### 5.3.3 实验结果

如表 11 所示,本文选择 PONI 作为对照方法进行实验比较。PONI 是模块化的方法,通过在每一时刻预测单步的可能的边缘以及目标的潜在位置来引导智能体进行导航。

表 11 真实环境下测试结果

场景	方法	马桶	时钟	电视	瓶子	沙发	盆栽	平均值
场景 1	PONI	40.00	60.00	20.00	20.00	40.00	30.00	37.14
	SDP	60.00	70.00	50.00	40.00	50.00	50.00	54.29
场景 2	PONI	50.00	30.00	30.00	10.00	20.00	40.00	31.34
	SDP	50.00	40.00	40.00	30.00	40.00	50.00	42.86
场景 3	PONI	30.00	30.00	20.00	10.00	30.00	40.00	27.14
	SDP	40.00	40.00	50.00	30.00	50.00	60.00	45.71
平均值	PONI	40.00	40.00	23.33	13.33	30.00	36.67	31.90
	SDP	50.00	50.00	46.67	33.33	46.67	53.33	47.62

结果表明,相比于单步预测的方式,序列扩散规划器采用序列化的规划方式,在位置灵活的小的物体,如盆栽、瓶子上的性能提升明显。而总体上,相比于对照方法序列扩散规划器的成功率提升了 15.72%,实验结果展示了序列扩散规划器在现实环

境导航的优良性能。

此外本文还提供了基于序列扩散规划器在真实环境下的导航可视化结果如图 10 所示,可视化结果显示,生成的轨迹与实际目标位置对齐良好,展示了序列扩散规划器在真实的现实场景中的鲁棒性。



图 10 真实环境下导航可视化结果

## 6 总结与展望

针对视觉目标导航任务,本文提出序列扩散规划器,旨在为智能体在导航过程规划行为序列,帮助智能体可以准确且快速地接近目标。序列扩散规划器利用智能体的历史观测和目标物体作为条件,通过迭代去噪的方式生成规划序列。基于规划序列智能体通过局部策略将序列转化成每一步需要执行的动作,直到找到目标物体。本文采用公开虚拟 3D 模拟器平台 Gibson 和 MP3D 对所提出的方法进行评估。实验结果表明,与基准方法相比,序列扩散规划器显著提高了导航性能。可视化实验结果显

示,序列扩散规划器生成的规划序列为智能体提供了有效的导航指导。此外本文还在真实环境中搭建了导航环境,基于 Locabot 机器人平台验证评估序列扩散规划器在真实环境下的性能,实验结果与可视化结果均表明序列扩散规划器相比于基准模型表现出良好的可扩展性和泛化能力。

尽管序列扩散规划器在面向有限目标物体集合的导航中展现了优良的性能,然而真实环境的物体集合是无穷尽的,因此后续的工作中将探究如何将面向有限物体的导航能力拓展到未见到过的新类别的物体中,以增加序列扩散规划器的泛化性和现实世界的应用性。

## 参 考 文 献

- [1] Tolman Edward C. Cognitive maps in rats and men. *Psychological Review*, 1948, 55(4): 189
- [2] Wang Yingxu, Ruhe Günther. The cognitive process of decision making. *International Journal Of Cognitive Informatics And Natural Intelligence*, 2007, 1(2): 73-85
- [3] Zhu Yuke, Mottaghi Roozbeh, Kolve Eric, et al. Target-driven visual navigation in indoor scenes using deep reinforcement learning//*Proceedings of the 2017 IEEE International Conference on Robotics and Automation*. Singapore, 2017: 3357-3364
- [4] Zhang Sixian, Song Xinhang, Bai Yubing, et al. Hierarchical object-to-zone graph for object navigation//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021: 15130-15140
- [5] Khandelwal Apoorv, Weihs Luca, Mottaghi Roozbeh, et al. Simple but effective: CLIP embeddings for embodied AI//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans, USA, 2022: 14809-14818
- [6] Ramrakhya Ram, Undersander Eric, Batra Dhruv, et al. Habitat-web: Learning embodied object-search strategies from human demonstrations at scale//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans, USA, 2022: 5163-5173
- [7] Maksymets Oleksandr, Cartillier Vincent, Gokaslan Aaron, et al. THDA: Treasure hunt data augmentation for semantic navigation//*Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision*. Montreal, Canada, 2021: 15354-15363
- [8] Zhang Sixian, Song Xinhang, Li Weijie, et al. Layout-based causal inference for object navigation//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver, Canada, 2023: 10792-10802
- [9] Chaplot Devendra Singh, Gandhi Dhiraj, Gupta Abhinav, et al. Object goal navigation using goal-oriented semantic exploration//*Proceedings of the Advances in Neural Information Processing Systems*. Virtual, 2020
- [10] Ramakrishnan Santhosh Kumar, Chaplot Devendra Singh, Al-Halah Ziad, et al. PONI: Potential functions for object-goal navigation with interaction-free learning//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans, USA, 2022: 18868-18878
- [11] Zhang Sixian, Yu Xinyao, Song Xinhang, et al. Imagine before go: Self-supervised generative map for object goal navigation//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA, 2024: 16414-16425
- [12] Bellman Richard. *Adaptive Control Processes-A Guided Tour* (Reprint from 1961). Princeton Legacy Library. Princeton, USA; Princeton University Press, 2015
- [13] Ho Jonathan, Jain Ajay, Abbeel Pieter. Denoising diffusion probabilistic models//*Proceedings of the Annual Conference on Neural Information Processing Systems*. Virtual, 2020
- [14] Ye Mao, Wu Lemeng, Liu Qiang. First hitting diffusion models for generating manifold, graph and categorical data//*Proceedings of the Annual Conference on Neural Information Processing Systems*. New Orleans, USA, 2022
- [15] Xia Fei, Zamir Amir Roshan, He Zhi-Yang, et al. Gibson Env: Real-world perception for embodied agents//*Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA, 2018: 9068-9079
- [16] Chang Angel X, Dai Angela, Funkhouser Thomas A, et al. Matterport3D: Learning from RGB-D data in indoor environments//*Proceedings of the 2017 International Conference on 3D Vision*. Qingdao, China, 2017: 667-676
- [17] Wortsman Mitchell, Ehsani Kiana, Rastegari Mohammad, et al. Learning to learn how to learn: Self-adaptive visual navigation using meta-learning//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Long Beach, USA, 2019: 6750-6759
- [18] Zhang Sixian, Li Weijie, Song Xinhang, et al. Generative meta-adversarial network for unseen object navigation//*Proceedings of the 17th European Conference on Computer Vision*. Tel Aviv, Israel, 2022: 301-320
- [19] Ramrakhya Ram, Batra Dhruv, Wijmans Erik, et al. PIRL-Nav: Pretraining with imitation and RL finetuning for object NAV//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver, Canada, 2023: 17896-17906
- [20] Kotar Klemen, Walsman Aaron, Mottaghi Roozbeh. ENTL: Embodied navigation trajectory learner//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023: 10863-10872
- [21] Bar Mayo, Tamir Hazan, Ayellet Tal. Visual navigation with spatial attention//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Virtual, 2021: 16898-16907
- [22] Yang Wei, Wang Xiaolong, Farhadi Ali, et al. Visual semantic navigation using scene priors//*Proceedings of the 7th International Conference on Learning Representations*. New Orleans, USA, 2019
- [23] Du Heming, Yu Xin, Zheng Liang. Learning object relation graph and tentative policy for visual navigation//*Proceedings of the 16th European Conference on Computer Vision*. Glasgow, UK, 2020: 19-34
- [24] Ye Xin, Yang Yezhou. Hierarchical and partially observable

- goal-driven policy learning with goals relational graph//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Virtual, 2021; 14101-14110
- [25] Chaplot Devendra Singh, Gandhi Dhiraj, Gupta Saurabh, et al. Learning To explore using active neural SLAM//Proceedings of the 8th International Conference on Learning Representations. Addis Ababa, Ethiopia, 2020
- [26] Chaplot Devendra Singh, Salakhutdinov Ruslan, Gupta Abhinav, et al. Neural Topological SLAM for visual navigation//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2020; 12872-12881
- [27] Zhai Albert J., Wang Shenlong. PEANUT: Predicting and navigating to unseen targets//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2023; 10926-10935
- [28] Chi Cheng, Feng Siyuan, Du Yilun, et al. Diffusion policy: Visuomotor policy learning via action diffusion//Proceedings of the Robotics: Science and Systems XIX. Daegu, Republic of Korea, 2023
- [29] Brohan Anthony, Brown Noah, Carbajal Justice, et al. Rt-2: Vision-language-action models transfer web knowledge to robotic control. arXiv Preprint arXiv:2307.15818, 2023
- [30] Ehsani Kiana, Gupta Tanmay, Hendrix Rose, et al. SPOC: Imitating shortest paths in simulation enables effective navigation and manipulation in the real world. arXiv preprint arXiv:2312.02976, 2023
- [31] Le Mero Luc, Yi Dewei, Dianati Mehrdad, et al. A survey on imitation learning techniques for end-to-end autonomous vehicles. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(9): 14128-14147
- [32] Hu Yihan, Yang Jiazhi, Chen Li, et al. Planning-oriented autonomous driving//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada, 2023; 17853-17862
- [33] Ramesh Aditya, Dhariwal Prafulla, Nichol Alex, et al. Hierarchical text-conditional image generation with clip latents. CoRR, 2022, abs/2204.06125
- [34] Saharia Chitwan, Chan William, Saxena Saurabh, et al. Photorealistic text-to-image diffusion models with deep language understanding//Proceedings of the Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022. New Orleans, USA, 2022
- [35] Peebles William, Xie Saining. Scalable diffusion models with transformers//Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris, France, 2023; 4172-4182
- [36] Rombach Robin, Blattmann Andreas, Lorenz Dominik, et al. High-resolution image synthesis with latent diffusion models//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA, 2022; 10674-10685
- [37] Wu Chanyue, Wang Dong, Bai Yunpeng, et al. Hsr-diff: Hyperspectral image super-resolution via conditional diffusion models//Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris, France, 2023; 7060-7070
- [38] Lugmayr Andreas, Danelljan Martin, Romero Andrés, et al. RePaint: Inpainting using denoising diffusion probabilistic models//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA, 2022; 11451-11461
- [39] Xie Shaoan, Zhang Zhifei, Lin Zhe, et al. Smartbrush: Text and shape guided object inpainting with diffusion model//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada, 2023; 22428-22437
- [40] Brooks Tim, Holynski Aleksander, Efros Alexei A. Instructpix2pix: Learning to follow image editing instructions//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada, 2023; 18392-18402
- [41] Sheynin Shelly, Polyak Adam, Singer Uriel, et al. Emu edit: Precise image editing via recognition and generation tasks. CoRR, 2023, abs/2311.10089
- [42] Chen Zoey Qiuyu, Kiani Shosuke C., Gupta Abhishek, et al. Genaug: Retargeting behaviors to unseen situations via generative augmentation//Proceedings of the Robotics: Science and Systems XIX. Daegu, Republic of Korea, 2023
- [43] Yu Tianhe, Xiao Ted, Tompson Jonathan, et al. Scaling robot learning with semantically imagined experience//Proceedings of the Robotics: Science and Systems XIX. Daegu, Republic of Korea, 2023
- [44] Lu Cong, Ball Philip J., Teh Yee Whye, et al. Synthetic experience replay//Proceedings of the Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023. New Orleans, USA, 2023
- [45] He Haoran, Bai Chenjia, Xu Kang, et al. Diffusion model is an effective planner and data synthesizer for multi-task reinforcement learning//Proceedings of the Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023. New Orleans, USA, 2023
- [46] Janner Michael, Du Yilun, Tenenbaum Joshua B., et al. Planning with diffusion for flexible behavior synthesis//Proceedings of the International Conference on Machine Learning. Baltimore, Maryland, USA, 2022; 9902-9915
- [47] Li Xiang, Belagali Varun, Shang Jinghuan, et al. Crossway diffusion: Improving diffusion-based visuomotor policy via



- self-supervised learning. CoRR, 2023, abs/2307.01849
- [48] Yu Xinyao, Zhang Sixian, Song Xinhang, et al. Trajectory diffusion for objectgoal navigation//Proceedings of the 38th Annual Conference on Neural Information Processing Systems. Vancouver, Canada, 2024
- [49] Zhu Minzhao, Zhao Binglei, Kong Tao. Navigating to objects in unseen environments by distance prediction//Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Kyoto, Japan, 2022; 10571-10578
- [50] Karmesh Yadav, Ram Ramrakhya, Santhosh Kumar Ramakrishnan, et al. Habitat-matterport 3D semantics dataset//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada, 2023; 4927-4936
- [51] Ye Hu, Zhang Jun, Liu Sibo, et al. IP-Adapter: Text compatible image prompt adapter for text-to-image diffusion Models. CoRR, 2023, abs/2308.06721
- [52] Sethian James A. Fast Marching Methods. SIAM Rev., 1999, 41(2): 199-235
- [53] He Kaiming, Gkioxari Georgia, Dollár Piotr, et al. Mask R-CNN//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy, 2017; 2980-2988
- [54] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Deep residual learning for image recognition//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA, 2016; 770-778
- [55] Ronneberger Olaf, Fischer Philipp, Brox Thomas. U-net; Convolutional networks for biomedical image segmentation//Proceedings of the Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015. Munich, Germany, 2015; 234-241
- [56] Kingma Diederik P., Ba Jimmy. Adam: A method for stochastic optimization//Proceedings of the 3rd International Conference on Learning Representations. San Diego, USA, 2015
- [57] Loshchilov Ilya, Hutter Frank. Decoupled weight decay regularization//Proceedings of the 7th International Conference on Learning Representations. New Orleans, USA, 2019
- [58] Wijmans Erik, Kadian Abhishek, Morcos Ari, et al. DD-PPO: Learning near-perfect pointGoal navigators from 2.5 billion frames//Proceedings of the 8th International Conference on Learning Representations. Addis Ababa, Ethiopia, 2020
- [59] Deitke Matt, VanderBilt Eli, Herrasti Alvaro, et al. ProcTHOR: Large-scale embodied AI using procedural generation//Proceedings of the Conference on Neural Information Processing Systems. Virtual, 2022
- [60] Ye Joel, Batra Dhruv, Das Abhishek, et al. Auxiliary tasks and exploration enable objectGoal navigation//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada, 2021; 16097-16106
- [61] Liang Yiqing, Chen Boyuan, Song Shuran. SSCNav: Confidence-aware semantic scene completion for visual semantic navigation//Proceedings of the IEEE International Conference on Robotics and Automation. Xi'an, China, 2021; 13194-13200
- [62] Yoo Hwiyeon, Choi Yunho, Park Jeongho, et al. Commonsense-aware object value graph for object goal navigation. IEEE Robotics and Automation Letters, 2024, 9(5): 4423-4430
- [63] Yamauchi Brian. A frontier-based approach for autonomous exploration//Proceedings of the IEEE International Symposium on Computational Intelligence in Robotics and Automation CIRA'97 - Towards New Computational Principles for Robotics and Automation. Monterey, USA, 1997; 146-151
- [64] Georgakis Georgios, Bucher Bernadette, Schmeckpeper Karl, et al. Learning to map for active semantic goal navigation//Proceedings of the Tenth International Conference on Learning Representations. Virtual, 2022
- [65] Luo Haokuan, Yue Albert, Hong Zhang-Wei, et al. Stubborn: A strong baseline for indoor object navigation//Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Kyoto, Japan, 2022; 3287-3293
- [66] Zhang Jiazhao, Dai Liu, Meng Fanpeng, et al. 3D-aware object goal navigation via simultaneous exploration and identification//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023; 6672-6682
- [67] Yu Bangguo, Kasaei Hamidreza, Cao Ming. L3MVN: Leveraging large language models for visual target navigation//Proceedings of the International Conference on Intelligent Robots and Systems., 2023; 3554-3560
- [68] Chen Shizhe, Chabal Thomas, Laptev Ivan, et al. Object goal navigation with recursive implicit maps. Cornell University Library, 2023, abs/2308.05602
- [69] Zhou Kaiwen, Zheng Kaizhi, Pryor Connor, et al. Esc: Exploration with soft commonsense constraints for zero-shot object navigation//Proceedings of the International Conference on Machine Learning. Honolulu, USA, 2023; 42829-42842
- [70] Yokoyama Naoki, Ha Sehoon, Batra Dhruv, et al. VLFM: Vision-language frontier maps for zero-shot semantic navigation//Proceedings of the International Conference on Robotics and Automation (ICRA). Yokohama, Japan, 2024
- [71] Yin Hang, Xu Xiuwei, Wu Zhenyu, et al. Sg-nav: Online 3d scene graph prompting for llm-based zero-shot object navigation. Advances in Neural Information Processing Systems. Vancouver, Canada, 2024, 37: 5285-5307



**ZHANG Si-Xian**, Ph. D. His research interests include embodied AI, scene recognition and object goal navigation.

**YU Xin-Yao**, B. E. candidate. Her research interests include embodied AI, scene recognition and object goal navigation.

**SONG Xin-Hang**, Ph. D. , associate professor. His re-

search interests include embodied AI, scene recognition, and multimedia content analysis.

**WANG Yi-Yao**, B. E. candidate. His research interests include embodied AI, scene recognition and object goal navigation.

**JIANG Shu-Qiang**, Ph. D. , professor. His research interests include multimedia content analysis and retrieval, image/video understanding, and multimodal intelligence.

## Background

Object goal navigation is a fundamental task for embodied agents interacting with the world. It requires the agent to navigate to a specified object in an unknown environment based on visual observations and a user-defined goal. Existing object navigation methods, whether end-to-end learning or modular approaches, typically perform single-step planning, i. e. , predicting the next action based solely on the current model input. This leads to a lack of temporal consistency in navigation planning, resulting in myopic decisions and ultimately reducing overall task efficiency.

In human navigation, decision-making is sequential, where the results form a series of the most likely actions from the current position to the goal. Inspired by human navigation abilities, this paper introduces a Sequential Diffusion Planner (SDP), which generates a sequence of actions from the current moment to the future based on real-time observations and predefined goal categories. The SDP is designed to guide the

agent to efficiently and rapidly navigate to the target object. Evaluations on the public virtual simulators, Gibson and MP3D, as well as in real-world environments, demonstrate that the sequential actions planned by the SDP enable the agent to navigate more precisely and efficiently.

This work was supported by National Natural Science Foundation of China under Grant 62125207, 62272443, 62032022, U23B2012, 62495084 and 62495085, in part by Beijing Natural Science Foundation under Grant Z190020 and L242020, and supported by the Postdoctoral Fellowship Program and China Postdoctoral Science Foundation under Grant Number BX20250391. This paper focuses on facilitating efficient goal-oriented navigation tasks for agents by proposed sequential diffusion planner. Our team, which has been engaged in research on embodied AI, ObjectNav tasks, and scene recognition for several years, has contributed a number of publications in this field.