arXiv:2601.03520v1 [cs.NE] 7 Jan 2026

一种基于强化学习的模型：利用多尺度位置场进行建图与目标导向导航

Bekarys Dukenbaev、Andrew Gerstenslager、Alexander Johnson、Ali A. Minai

摘要—在复杂且部分可观测的环境中实现自主导航，仍是机器人学面临的核心挑战。研究者已提出若干基于哺乳动物海马体中位置细胞的仿生建图与导航模型。本文提出一种新的鲁棒模型，该模型采用多个空间尺度上的并行位置场层、基于重放的奖励机制以及动态尺度融合。仿真结果表明，与单尺度基线相比，该模型能够提高路径效率并加快学习，凸显了多尺度空间表征对于自适应机器人导航的价值。

索引词—自主导航、位置细胞、多尺度表征、认知地图、强化学习、仿生机器人学

I. 引言

自主机器人导航的一个共同目标，是使智能体能够从经验中学习，并像动物一样灵活适应地到达特定目标。其挑战在于构建能够在复杂环境中高效探索、建图和规划的系统，而无需大量先验知识。传统的同步定位与建图（SLAM）方法依赖对几何结构的算法重建，但在具有部分可观测性、稀疏感知和尺度问题的非结构化环境中往往表现不佳，并且会产生大量内存和计算成本 [1]。相比之下，哺乳动物能够在复杂环境中快速、灵活地学习导航 [2]，，并在经验有限的情况下形成内部空间表征并到达目标；这一过程由海马复合体的神经回路引导，涉及具有位置特异性活动的位置细胞 [3]、具有空间周期性活动的网格细胞 [4]，以及支持快速价值赋值与规划的重放／预重放机制 [5]–[8]。

海马系统的一个重要特征是位置场沿背腹轴呈现空间尺度变化 [9]。计算模型通常采用单一空间分辨率，因此多尺度组织结构的功能意义尚未得到充分探索。本文介绍了一种基于强化学习的新模型，通过实现多尺度表征来解决现有方法的关键局限，并支持动态尺度整合。在该模型中，不同空间分辨率下的位置场被构建为并行层，并与一个轻量级价值学习器相耦合。在运行时，系统融合不同尺度的预测，并通过自适应融合机制，在每个决策点选择能够提供最清晰、最可靠方向性奖励结构的空间分辨率。该架构提出了两项关键创新：（i）并行多尺度神经群体；（ii）动态尺度加权机制，利用不同尺度下价值图结构的差异来稳定策略更新，而不预设任何特定尺度具有固定作用。在不同大小和复杂程度的环境中进行的模拟表明，多尺度模型在路径效率、学习速度和总体目标达成表现方面均优于单尺度基线。综合而言，结果表明，多尺度空间表征能够提升基于强化学习的导航性能，并为海马尺度多样性如何支撑适应性行为提供了计算层面的解释。

B. Dukenbaev、A. Gerstenslager 和 A. Johnson 任职于美国辛辛那提辛辛那提大学计算机科学系。A.A. Minai 任职于美国辛辛那提辛辛那提大学电气与计算机工程系。

II. 背景与研究动机

A. 空间表征的神经基础

理解动物如何表征空间并在其中导航，一直是神经科学的核心问题，也为仿生机器人学提供了重要灵感。海马体及其周围的内侧颞叶结构对这一能力至关重要；病灶研究表明，海马受损的啮齿动物在执行空间记忆和导航任务时会出现严重障碍 [10]。

O’Keefe 和 Dostrovsky 发现位置细胞，证实当动物处于特定位置时，海马神经元会选择性地放电，这支持了环境内部存在一幅认知地图 的观点 [11], [12]。后续研究发现了其他具有空间调谐特性的细胞群：头方向细胞 编码异中心定向 [13]，而边界向量细胞 则会对墙壁或障碍物等环境边界的距离和角度作出响应 [14]，[15]。这些神经系统共同整合感觉线索与自我运动线索，生成结构化的多模态表征，从而支撑灵活的空间记忆与导航。

B. 多尺度空间表征

海马结构的一个显著特征是沿背腹轴呈现出空间尺度的系统性梯度。背侧海马细胞表现出较小且高分辨率的位置场，能够实现精确定位；而腹侧细胞则呈现更大、更弥散的位置场，能够捕捉整体空间背景 [9]，[16]。这种多尺度 组织支撑层级化的空间编码，其中细尺度表征支持局部

导航，而粗尺度表征则能够实现高效的远程规划。

C. 序列编码与可塑性

位置细胞编码的不仅是位置，还包括经历的时间顺序。脉冲时序依赖可塑性（STDP）会增强依次活跃的细胞之间的连接，将轨迹串联成连贯的状态空间表征 [17], [18]。这一机制解释了为何位置场会遵循环境边界——障碍物两侧的细胞不会按顺序被激活，因此连接仍然较弱。STDP 与神经调节信号相结合时，也有助于强化学习，使时间顺序排列的状态序列能够在事后与奖励相关联 [5], [19]。

D. 重放与预演

海马体会在尖波涟漪期间离线重新激活空间序列，此时先前经历过的轨迹或潜在的未来轨迹会被短暂重放。在这些事件中，网络会短暂重放位置细胞活动模式，这些模式代表过去或潜在的未来轨迹。重放 特指对先前经历过的路径所对应序列的重新激活。这些序列可以按正向或反向出现；正向重放有助于记忆巩固，而反向重放则沿着经历过的路径向后传播奖励信息，增强较早状态与成功结果之间的关联 [5], [7], [20], [21]。

相比之下，预演指动物移动之前出现的轨迹序列激活。这些有结构、面向未来的激活表明，海马会模拟可能的未来路径，从而支持规划和决策 [6]。预演与替代性试错行为密切相关：啮齿动物会在决策点停下，并短暂评估不同路线 [22], [23]。

总之，海马重放和预演对应于现代强化学习中的若干核心机制，包括蒙特卡洛树搜索（MCTS） [24]、复演或“做梦” [25], 以及资格迹 [26]。

E. 计算与机器人模型

海马体的计算模型长期以来一直在研究位置细胞和边界细胞如何支撑空间记忆与导航，通常采用由吸引子动力学、赫布学习、边界向量细胞输入和网格细胞驱动的单尺度表征 [15], [27]–[31]。基于这些模型，人们已开发出多种成功的方法，用于在存在障碍物的复杂环境中导航 [32]–[35]。这些方法都以某种方式结合了记忆重放、奖励传播和路径规划，以寻找高效路径，不过具体做法存在很大差异。然而，它们都基于单一分辨率尺度的位置细胞。Erdem 和 Hasselmo 提出了一种采用多尺度位置场的分层导航模型 [36]，该模型随后通过 RatSLAM模型 实现并应用于一台由视觉驱动的实体机器人 [37]。该模型利用多尺度正向重放进行导航，具体做法是在每个回合中搜索明确的目标导向线性轨迹，利用较大位置场更远的支撑范围，寻找通往目标的线性航向。这种方法在小型开放环境中效果良好，但无法推广到有障碍物的大型和／或复杂环境。为解决这一问题，人们提出了另一种采用多尺度位置场的方法，研究表明，较大的位置场有助于快速粗略探索，而较小的位置场则能改善障碍物和目标附近的轨迹细化 [39]–[41]。然而，这种方法使用的是由固定位置场构成的规则网格。两种方法都不支持多尺度信息的自适应利用。本模型通过将自组织并行多尺度位置场群体与奖励学习网络相结合，克服了上述所有局限；该网络构建多尺度奖励地图，并在动作选择过程中执行在线尺度选择和自适应融合。

III. 模型架构

该模型由四个核心层组成——头方向细胞（HD）、边界向量细胞（BVC）、位置细胞（PC）和奖励网络——并以并行方式部署于多个空间尺度上。每种尺度都包含一套完整的 BVC–PC–奖励层，各自具有调谐参数，能够生成从精细到粗略分辨率的空间编码和价值估计。这些组件的行为大体类似于其生物学对应物：头方向细胞提供全局定向信号，边界向量细胞编码边界几何，位置细胞形成位置特定表征，而奖励层则为已访问状态赋予价值。该模型建立在本研究组此前开发的一个较简单的单尺度模型之上 [34], [35], [42]。

不同于采用固定位置场形状（通常为高斯形）或经过设计的布局的方法，这里的所有位置场都是通过由感觉驱动、与环境交互而动态形成的。通过为每个尺度指定不同的边界向量细胞调谐宽度 σ_r （径向）和 σ_θ（角向），实现多尺度结构；这些宽度决定了位置场大小以及各尺度奖励地图的空间粒度。图 1 总结了感觉输入、各尺度的神经处理，以及后续的决策。

A. 模型层

1）头方向细胞层：头方向（HD）层对智能体的自我中心无关航向进行编码。每个头方向细胞在自我中心无关坐标中都有一个偏好方向；当智能体的航向与其偏好方向一致时，该细胞的放电率达到最大值，而随着航向偏离偏好方向，放电率则对称地下降。因此，该层的活动通过群体编码表示智能体相对于固定外部参照的方向航向，使其能够在环境中保持全局方向感。

形式上，头方向细胞 i 的放电率为

$$v_i^h = \mathbf{x}' \cdot \begin{bmatrix} \cos\left(\theta_i^h + \theta_0\right) \\ \sin\left(\theta_i^h + \theta_0\right) \end{bmatrix}, \tag{1}$$

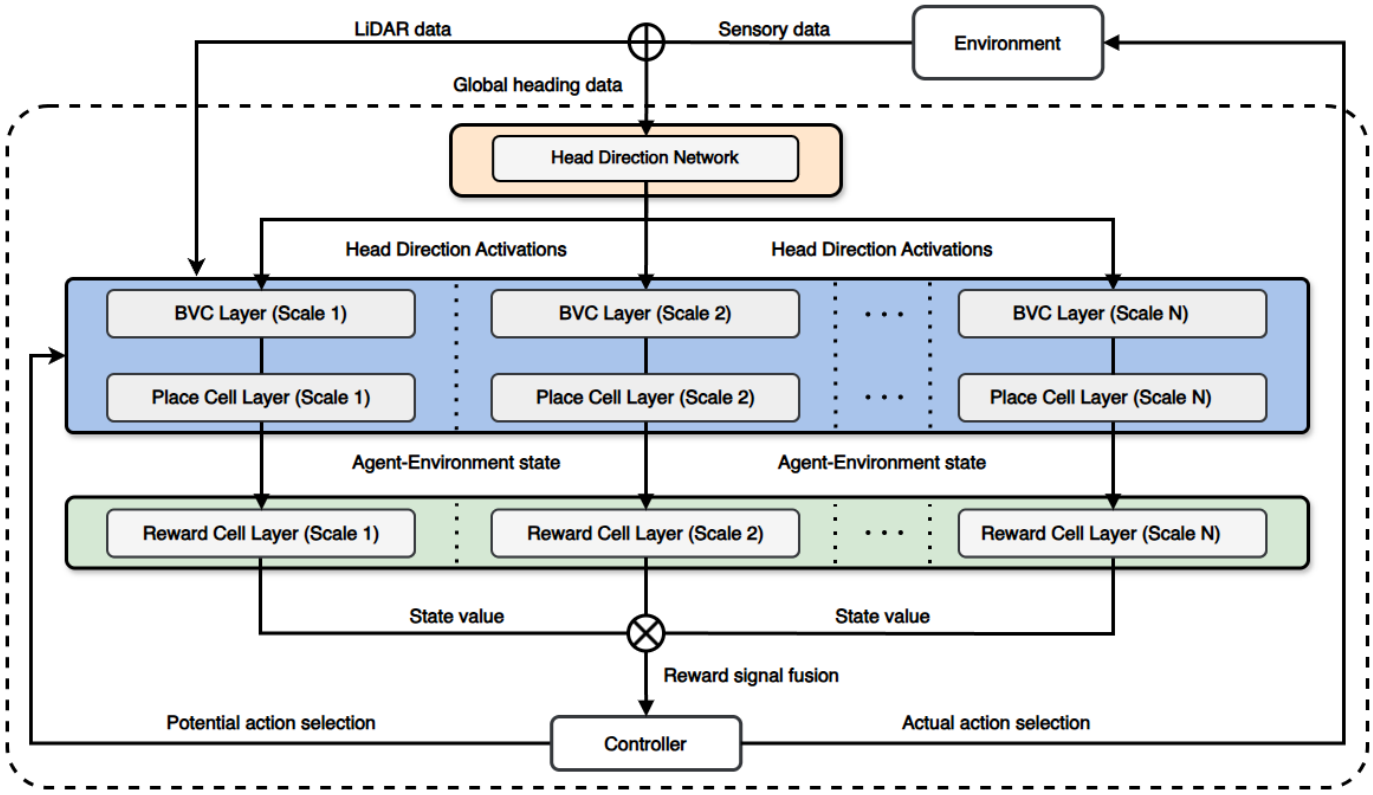


图1. 多尺度系统架构，整合了感觉输入、神经处理层和决策组件。头方向网络处理全局朝向数据，并调制多个尺度上的边界向量细胞层和位置细胞层的激活。随后，各尺度的位置细胞层激活被传递至相应的奖励细胞层，由其为每个可能的航向返回一个潜在奖励值。接着，融合模块对这些奖励值进行汇总，并据此选择并执行单个最优动作。

其中， θ_0 表示锚定线索（一个固定的外部参照）的朝向角， θ_i^h 是第 i 个头方向细胞的首选异中心方向， $\mathbf{x}' = [x_0', x_1']$ 表示智能体在笛卡尔坐标 [36]中的瞬时速度。模型使用 $n_{\mathrm{hd}} = 8$ 个首选方向为

$$\{\theta_d\}_{d=0}^{n_{\mathrm{hd}}-1} = \{0^\circ, 45^\circ, \ldots, 315^\circ\},$$

我们将其称为规范基准航向集合 $\boldsymbol{\Theta}$ 。

2）边界向量细胞层：边界向量细胞层采用了Barry等人 [15], 提出的模型，其中每个边界向量细胞 i 通过结合距离和角度上的两条高斯调谐曲线，对位于偏好距离 d_i 和方向 ϕ_i的边界产生响应。设

$$\mathbf{r} = [r_1, \ldots, r_{n_{\mathrm{res}}}], \qquad \boldsymbol{\theta} = [\theta_0, \ldots, \theta_{n_{\mathrm{res}}}]$$

表示由 n_{res} 激光雷达光束检测到的最近障碍物的距离和方位角。放电率为

$$v_i^b = \frac{1}{N_{\mathrm{BVC}}} \sum_{j=0}^{n_{\mathrm{res}}} \left(\frac{\exp\left[-\frac{(r_j - d_i)^2}{2\sigma_r^2}\right]}{\sqrt{2\pi}\,\sigma_r} \times \frac{\exp\left[-\frac{(\theta_j - \phi_i)^2}{2\sigma_\theta^2}\right]}{\sqrt{2\pi}\,\sigma_\theta} \right), \tag{2}$$

其中， σ_r 和 σ_θ 分别控制距离和方向调谐，而 N_{BVC} 对整个群体的活动进行归一化。变量包括： r_j （距离）、第 j束光束的 θ_j （方位）； d_i、 ϕ_i （首选距离和方向），边界向量细胞 i 的参数；以及 n_{res}（传感器分辨率）。

已有多种生成多尺度位置场的机制被提出 [43], [44]，但边界向量细胞输入最直接地控制了场大小 [15], [41]。在该模型中，尺度差异仅由边界向量细胞调谐宽度 σ_r 和 σ_θ 决定；这些宽度决定了边界向量细胞响应的空间平滑度，进而决定下游位置场的分辨率。较大的宽度会产生更宽广的粗尺度场，而较小的宽度则会产生空间上更精确的场。

3）位置细胞层：位置细胞（PC）层是模型中空间表征的主要所在，由具有局部调谐活动的位置细胞组成，其活动形式为位置场。每个位置细胞都接收来自边界向量细胞的加权兴奋，并同时受到边界向量细胞总活动的抑制（前馈抑制）以及其他位置细胞的抑制（循环抑制）。这些相互作用共同形成稳定且均匀分布的位置场 [34]，[35], [42]。

a）活动模型：位置细胞 i 的膜电位 s_i^p 按以下规律变化：

$$\tau_p \frac{ds_i^p}{dt} = -s_i^p + \sum_{j=0}^{n_b} W_{ij}^{pb} v_j^b - \Gamma_{pb} \sum_{j=0}^{n_b} v_j^b - \Gamma_{pp} \sum_{j=0}^{n_p} v_j^p, \tag{3}$$

其中， τ_p 为膜时间常数； W_{ij}^{pb} 为边界向量细胞→位置细胞突触（采用稀疏初始化以促进多样化的感受

野）； v_j^b 和 v_j^p 分别为边界向量细胞和位置细胞的放电率；Γ_{pb}，Γ_{pp}尺度前馈抑制和循环抑制。位置细胞 i 的放电率为

$$v_i^p = \tanh\Big(\big[\psi\, s_i^p\big]_+\Big), \tag{4}$$

通过整流确保输出非负，增益 ψ则用于设定响应的尖锐程度。

b) 位置场的自组织：当智能体探索新的环境时，位置场会通过竞争学习逐渐形成：受到强烈驱动的位置细胞会增强其边界向量细胞输入，而抑制作用则会压制竞争细胞，从而确保每个位置细胞都获得一个独特的位置场，且所有位置场共同覆盖整个环境。突触适应遵循 Oja 规则的一种变体 [45]:

$$\tau_{w^{pb}} \frac{dW_{ij}^{pb}}{dt} = v_i^p\Big(v_j^b - \tfrac{1}{\alpha_{pb}}\, v_i^p\, W_{ij}^{pb}\Big), \tag{5}$$

其中， $\tau_{w^{pb}}$ 为学习率， α_{pb} 为归一化因子；每个突触以 p_{pb} =0。25 的概率初始化为 1。与实际海马体中观察到的表征相比，这种机制产生的位置表征冗余度较低；在实际海马体中，位置场可能会显著重叠。模型中的这一选择体现了冗余度与效率之间的权衡，并使模型能够使用数量较少的位置细胞。若在含有更多神经元的层中采用局部而非全局抑制性投射，模型便能轻松容纳更高的冗余度。

c) 位置细胞邻接关系学习：为了根据位置编码创建环境的拓扑表示，3D 张量W^{pp} = $[W_{kij}^{pp}]$对位置细胞之间的方向性邻接关系进行编码。每个位置细胞 j与其他每个位置细胞 i 之间有 8 个突触，每个突触分别对应 8 个基准朝向之一，并由 k进行索引。为捕捉时间顺序，模型在一个较短的时间窗口内，对突触前位置细胞 j 、突触后位置细胞 i 以及每个头方向细胞k 的活动进行积分。积分后的活动变化如下：

$$\tau_m \frac{d\Upsilon_j^p}{dt} = -\Upsilon_j^p + v_j^p, \tag{6}$$

$$\tau_m \frac{d\Upsilon_i^p}{dt} = -\Upsilon_i^p + v_i^p, \tag{7}$$

$$\tau_m \frac{d\Upsilon_k^h}{dt} = -\Upsilon_k^h + v_k^h, \tag{8}$$

其中， τ_m 为积分常数。这些指数轨迹保留了近期激活逐渐衰减的记忆。

邻接权重按照以下方式更新

$$\tau_{w^{pp}} \frac{dW_{kij}^{pp}}{dt} = \Upsilon_k^h\left(v_i^p \Upsilon_j^p - v_j^p \Upsilon_i^p\right), \tag{9}$$

其中， $\tau_{w^{pp}}$ 控制学习速度， Υ_k^h 控制每个航向的更新。差分项的符号编码运动方向：当智能体沿 k方向从 j 航向移动至 i 航向时，W_{kij}^{pp} 随之增大；反向移动时则减小，从而实现脉冲时序依赖可塑性（STDP）的时间平滑版本 [17]。因此，PC层学习到一个反映环境拓扑结构和导航流的有向邻接图。

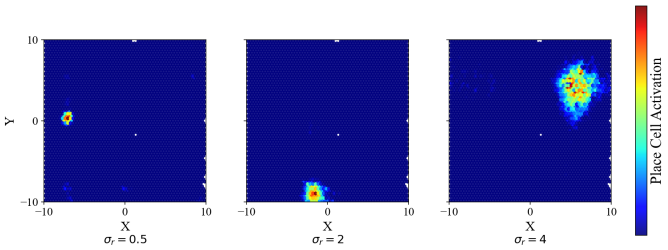


图 2. 在 20 × 20米环境中，三个具有代表性的位置细胞在不同空间尺度（σ_r = 0.5、2。0、和 4。0 米）下的激活模式。更大的 σ_r 值会产生更大的感受野，表明空间表示更为粗略。颜色表示激活强度。

d) 多尺度位置细胞实例化：每个空间尺度都维护其自身的位置细胞群体，具有相同的动力学特性，但上游边界向量细胞调谐宽度不同。图2展示了实验中使用的三个尺度的感受野示例。

这些尺度依赖型感受野自然源自边界向量细胞的调谐参数；所有位置场形成和邻接学习在各个尺度上均按照公式 4–9 进行。

4）奖励细胞层：奖励细胞会在受奖励目标 [46]，[47] 所在的位置激活。在模型中，每个奖励细胞只对一个目标作出响应。每个尺度都有独立的奖励细胞层，因此，对于任一目标，激活的奖励细胞数量等于模型中的空间尺度数量。每个尺度上的奖励细胞都会接收该尺度所有位置细胞的突触输入。学习完成后，奖励细胞的活动会以高度压缩的价值信号提供目标距离的标量估计（类似于腹侧纹状体或下托的价值编码），从而编码出一幅奖励地图：其峰值位于目标处，并向各个方向随距离增加而单调递减。由于各位置细胞尺度的空间分辨率各不相同，所得奖励图在平滑度和空间范围上也有所差异：较粗的尺度会产生支撑范围更广的价值曲面，而较精细的尺度则会形成空间细节更丰富的结构。

奖励单元根据位置细胞的活动计算归一化激活值。令 $\mathbf{v}^p \in \mathbb{R}^{n_p}$ 为位置细胞放电率向量，$\mathbf{w}^r \in \mathbb{R}^{n_p}$ 为从位置细胞到奖励单元的突触权重。奖励单元的原始激活值为：

$$a^r(v^p) = \frac{\mathbf{w}^r \cdot \mathbf{v}^p}{\max\left(\|\mathbf{v}^p\|_1, \varepsilon\right)}, \tag{10}$$

其中 $\varepsilon = 10^{-4}$ 用于防止除以零。最终放电率经过整流并限制在一定范围内：

$$v^r = \min\left(\max(a^r(v^p), 0), B\right), \tag{11}$$

其中 B 是作为参数设置的较大上限，以确保活动非负且符合生物学上的合理范围。

a) 通过重放进行奖励学习： 学习过程中，奖励单元不会利用自身的前向激活 a^r 来驱动突触可塑性。相反，重放会增强那些能够预测奖励或在奖励出现之前活跃的位置细胞所形成的突触。当智能体首次到达目标时，在目标处活跃的位置细胞会增强其与奖励单元之间的连接。随后，模型进入离线反向重放 阶段，此时先前活跃的位置细胞会按照时间压缩的逆序重新激活，这与

海马尖波涟漪动力学 [5]一致。突触更新完全由重放的位置细胞活动驱动，具体如下所述。

该机制将较早的状态与最终获得的奖励联系起来，并支持长程价值赋值。在重放开始前，为所有 i 将权重更新累加器 Δw_i^r 初始化为零。在重放期间，突触更新逐步累积：

$$\Delta w_i^r \leftarrow \Delta w_i^r + \frac{\overline{v}_i^p}{\|\overline{v}^p\|_\infty} \exp\left(-\frac{t_r}{\tau_r}\right), \tag{12}$$

其中，$-v_i^p$ 是位置细胞 i 的重放放电率，并通过 $\|v^p\|_\infty$ 进行归一化，以防止较大的激活值占据主导。可变 t_r 是离散的重放步索引（$t_r = 0,\ 1,\ .....$），其中 $t_r = 0$ 对应目标状态，并沿重放轨迹递增。衰减时间常数 τ_r 确保距离目标越远的状态，对累积权重更新的贡献逐渐越小。

重放结束后，应用归一化的权重更新：

$$w_i^r \leftarrow w_i^r + \frac{\Delta w_i^r}{\|\Delta \mathbf{w}^r\|_\infty}, \tag{13}$$

这既实现了突触竞争，也确保距离较远或预测能力较弱的状态获得较小的权重变化。由此得到的权重编码了奖励地图：一个理想情况下在目标处达到峰值、并沿经历轨迹随距离平滑下降的曲面。

b) 时序差分（*TD*）学习： 重放建立了长程奖励传播机制，但模型还会利用时序差分规则在线优化奖励预测。给定观测到的奖励 R_{next}，预测误差为

$$\delta = R_{\text{next}} - \mathbf{w}^r \cdot \mathbf{v}^p, \tag{14}$$

并且突触根据以下规则更新

$$\mathbf{w}^r \leftarrow \mathbf{w}^r + \eta \delta \mathbf{v}^p, \tag{15}$$

学习率为 η。TD 学习会提高目标附近的预测准确性，并在重复回合中稳定奖励剖面。

c) 多尺度奖励地图：上游边界向量细胞调谐宽度（σ_r, σ_θ）的差异会产生具有互补空间特性的奖励地图：

- **小尺度：** 高分辨率、快速变化的奖励结构，适用于障碍物或目标附近的精细操控，但远离已经历轨迹时空间泛化能力有限，且无法在远离目标处提供充分引导。
- **中尺度：** 结构适中的奖励剖面，适用于中等复杂度，在局部细节与空间泛化之间实现平衡。
- **大尺度：** 平滑、缓慢变化的奖励剖面，能够支撑远距离引导、广泛的空间泛化，以及在开放环境中的导航。

这些地图（图 3）构成规划过程中多尺度价值融合的基础：粗尺度地图推动智能体朝远处的目标前进，而细尺度地图可防止碰撞，并在局部细化轨迹。

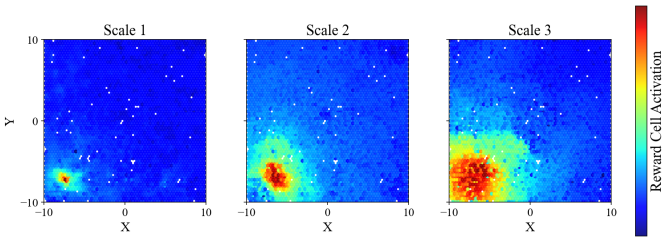


图 3. 三种空间尺度下的奖励细胞激活模式，对应于位于一个 20×20 米环境左下角的奖励。颜色表示细胞激活强度

B. 运行模式：探索与利用

模型在两种交替模式下运行，以控制其整体行为：用于构建空间表征的探索 模式，以及利用学习得到的地图进行目标导向导航的利用 模式。

a) 探索： 在探索过程中，智能体在环境中进行随机游走，通过以下过程构建其内部地图：

1) **位置场形成：** 位置细胞通过竞争学习（公式 5）形成局部化感受野，逐渐铺满环境。

2) **邻接学习：** 当智能体在位置场之间移动时，利用方向门控的赫布学习（公式 6、7、8、9）更新循环突触，形成编码空间连通性和运动方向的邻接权重张量。

3) **奖励地图初始化：** 首次遇到目标时，奖励单元被激活，与其共同激活的位置细胞之间的突触得到增强，从而将奖励位置锚定在网络中，并通过反向重放（公式 12）和时序差分学习生成初始奖励地图。

b) 利用：空间地图和奖励地图建立后，智能体便切换到目标导向导航。每一步中，它都会进行内部模拟（预演），设想每种可能航向所产生的后果，评估各个尺度下的预测奖励，融合不同尺度的预测结果，最后选择最优的连续运动方向。

1) **预演：** 每一步中，智能体都会利用邻接关系驱动的预测，对所有基准航向进行一步内部模拟：

$$\hat{v}_i^p(\theta) = \tanh\left(\left[\sum_{j=1}^{n_p} W_{aij}^{pp}\, v_j^p - v_i^p\right]_+\right), \tag{16}$$

其中，W_{aij}^{pp} 是邻接张量中与基准航向 θ_a 对应的切片，该航向最接近候选方向 θ。向量$\hat{\mathbf{v}}^p(\theta)$表示智能体预期在朝 θ方向移动时会产生的位置细胞活动。该想象模式仅用于评估不同尺度下的预测奖励；预演步骤本身不会计算奖励。用强化学习术语来说，这一操作相当于一步基于模型的展开 [26]。

2) **通过多尺度整合进行奖励评估与动作选择：** 针对每个航向，奖励预测——

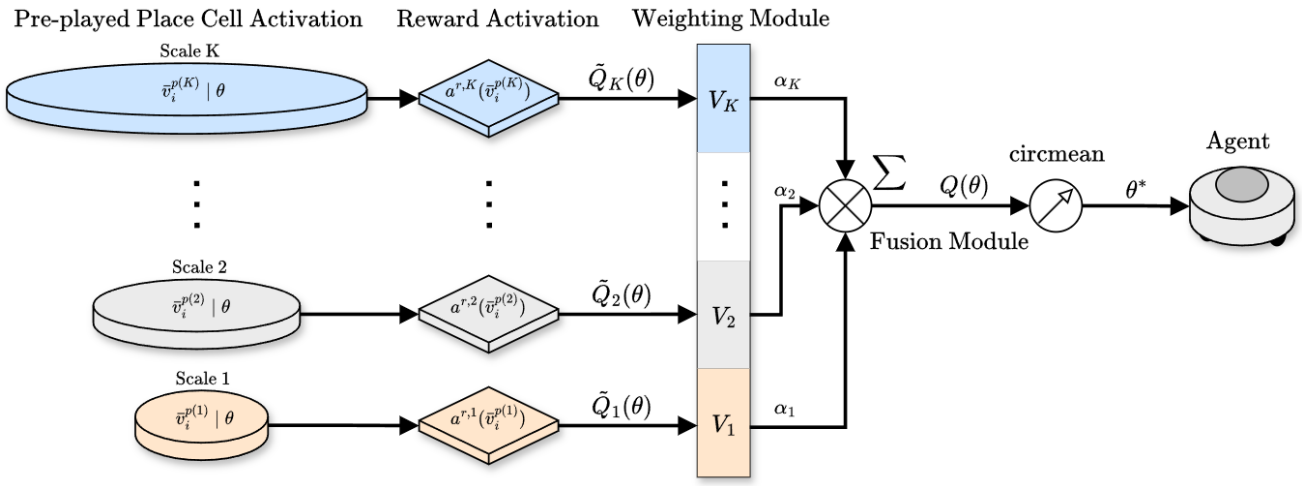


图 4. 奖励融合架构。预演为每个航向和尺度生成预测的位置细胞活动；奖励图中的方向性变化决定加权因子 α_k，加权和则确定最佳动作 a^*。

将所有尺度的奖励预测通过动态尺度加权机制进行整合，并由此得到最优航向，详情见第 III-C 节。无法产生足够强奖励信号的尺度将被排除。需要指出的是，得到的最优航向不一定是 8 个基准航向之一，实际运动发生在连续空间中。

3)循环预防。 智能体通过检测在没有向前移动时过度转向的情况来避免紧密的旋转循环，并暂时恢复探索。

C. 尺度的动态整合

特定尺度的奖励预测： 对于每个尺度 k，航向 θ 的奖励估计通过将奖励细胞激活（第 III-A4 节）应用于预演预测的位置细胞活动 $\bar{\mathbf{v}}^{p,k}|\theta$ （来自公式 16）而获得。对于奖励权重$\mathbf{w}^{r,k}$,

$$Q_k(\theta) = \frac{\mathbf{w}^{r,k} \cdot \left(\bar{\mathbf{v}}^{p,k} \mid \theta\right)}{\max\left(\|\bar{\mathbf{v}}^{p,k} \mid \theta\|_1,\ \varepsilon\right)}. \tag{17}$$

为根据方向性奖励结构而非幅值比较各个尺度，需要对每个剖面进行归一化：

$$\tilde{Q}_k(\theta) = \frac{Q_k(\theta)}{\max_{\theta'} Q_k(\theta') + \varepsilon}. \tag{18}$$

基于变化的尺度加权： 为优先考虑信息丰富的尺度，智能体根据其归一化剖面$Q_k(\theta)$，测量每个有效尺度 k 上的总方向性奖励变化：

$$V_k = \sum_{d=0}^{n_{\mathrm{hd}}-2} \left|\tilde{Q}_k(\theta_{d+1}) - \tilde{Q}_k(\theta_d)\right|, \tag{19}$$

其中，$\{\theta_d\}_{d=0}^{n_{\mathrm{hd}}-1}$ 是在预演期间使用的离散基准航向。随后，对这些变化值在有效尺度范围内进行归一化，以获得式 21 中使用的混合权重：

$$\alpha_k = \frac{V_k}{\sum_{j\in\mathcal{V}} V_j + \varepsilon}, \tag{20}$$

使具有更大方向变化的尺度 V_k 在融合奖励 $Q(\theta)$中获得更高权重。

多尺度融合： 在利用过程中，每个候选航向的归一化奖励估计会在各个空间尺度之间进行融合：

$$Q(\theta) = \sum_{k\in\mathcal{V}} \alpha_k\, \tilde{Q}_k(\theta), \tag{21}$$

其中，$\mathcal{V}$ 仅包含最大预测奖励超过有效性阈值的尺度。系数 α_k （定义于式 20）表示各有效尺度对融合奖励剖面的相对贡献。未达到阈值的尺度会从动作选择中排除。不过，这些尺度仍会继续更新其位置细胞权重和奖励细胞权重，因此，$\mathcal{V}$ 仅决定哪些尺度参与决策，而不决定哪些尺度继续学习。如果没有任何有效尺度（$\mathcal{V} = \varnothing$），智能体会短暂切换到探索状态，然后重新评估奖励信号。

带障碍物规避的动作选择： 为避免不安全的移动，距离障碍物过近时会屏蔽相应的候选航向。对于每个航向 θ_d，根据测距仪估算其到边界的最小距离 $d\ (\theta_d)$ 。如果 $d\ (\theta_d) < d_{\mathrm{safe}}$，则在融合之前，将所有尺度上的相应奖励值设为零：

$$\tilde{Q}_k(\theta_d) \leftarrow 0 \quad \text{for all } k \text{ if } d(\theta_d) < d_{\mathrm{safe}}. \tag{22}$$

随后，将最优移动方向计算为融合奖励剖面的圆周均值：

$$\theta^* = \arctan2\left(\sum_{\theta} Q(\theta)\sin\theta,\ \sum_{\theta} Q(\theta)\cos\theta\right), \tag{23}$$

即使 $Q(\theta)$ 的最大值较为宽泛或呈多峰分布，也能在连续空间中得到唯一的航向。

IV. 实验设置与性能分析

我们在两种设置下评估了多尺度模型：(i)实验 *1*，用于评估经过预训练的导航效率；以及 (ii) 实验*2*，用于考察从朴素初始化开始的在线策略收敛。表 I 总结了目标和流程。

表 I 两项实验概要

	实验 1：路径效率	**实验2：策略学习**
目标	评估单尺度与 多尺度策略的路径效率	评估在线策略收敛和学习 动力学
环境	环境 1–4	环境 1（开放场地）
指标	到达目标的步数	每回合到达目标的步数，回合收敛 每个回合的趋同率
流程	无奖励映射后进行一次重 播以形成奖励地图；固定起点	无预训练；智能体学习位置场、相邻关系和奖励地图；固定起点，并在各回合中持续学习 点
测试运行	每种策略在每个环境中进行 20 次试验	每种策略运行 51 个回合 $\times$ 5
主要差异	使用预训练的空间地图进行评估	从第0集开始完全在线学习

表II 导航策略

尺度索引	**策略**	σ_r (米)	**# 位置细胞**	**# 位置细胞层**
1	小尺度	0.5	2000	1
2	中尺度	2.0	500	1
3	大尺度	4.0	250	1
–	多尺度	可变	可变	3

A. 实验设置

仿真在 Webots R2025a [48] 中进行，使用了一台配备罗盘的差速驱动机器人，用于更新航向；该机器人还配备了平面测距仪，每次 360° 扫描提供 720 条射线，并将数据直接输入边界向量细胞层。

1) 实验组与对照组：在两项实验中，多尺度策略作为实验条件，其与三个单尺度控制策略的唯一区别在于对空间尺度的整合。所有策略均接收相同的感知数据和奖励数据，从而确保任何性能差异都可直接归因于多尺度整合的影响。

2) 环境设计：使用了四个 20×20 米的场地（环境 1–4），这些场地设有边界墙和不同的内部障碍物布局，用于改变导航复杂度。障碍物形成开放区域和狭窄走廊，目标是一个固定的、半径为 0.5 米的区域，只有在抵达后才能被检测到。

3) 导航策略：实验采用了三种单尺度导航策略——小、中、大——以及一种整合全部三种空间分辨率的多尺度策略（表II）。

B. 实验 *1*：路径效率评估

1) 概述：本实验使用预训练空间模型评估目标导向导航，以分离位置场尺度的内在影响。所有学习过程均在评估前完成，导航期间不更新任何权重。

2）训练流程：训练分为两个阶段。在地图构建阶段，不提供奖励，通过学习位置场和方向邻接关系来建立连接：边界向量细胞→位置细胞突触 W_{ij}^{pb} 在探索过程中进行调整，并根据前文所述的基于尖峰时序依赖可塑性的规则学习方向邻接权重 W_{kij}^{pp}，而奖励权重 w_i^r 保持冻结。在随后的目标搜寻阶段，智能体首先遇到目标，触发一次反向重放事件，通过更新 w_i^r建立奖励地图。重放结束后，所有权重均被冻结以供评估。由于三个空间尺度均基于相同的感觉输入并行训练，因此，多尺度策略的差异仅在于导航过程中如何组合各尺度的预测。

3) 评估流程：我们在环境1–4中测试了四种策略（小尺度、中尺度、大尺度、多尺度），每个环境进行了20次试验。性能通过步数衡量；在移动增量固定的情况下，步数完全决定路径长度。

4) 结果与图示：图6展示了所有策略在各环境中的步数性能。

5) 分析与讨论：图6汇总了各环境中的步数。

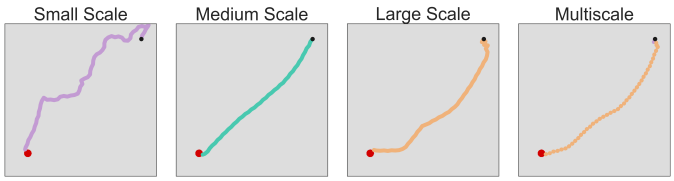
环境1（开放）。中尺度策略所需步数最少（480步）。小尺度策略表现不佳（1605.35步），原因是出现了振荡式修正。大尺度与多尺度策略表现相近（547.45步与532步），其中多尺度策略产生了更平滑的轨迹，且以粗尺度活动为主。

环境 2（单个障碍物）。单个障碍物放大了尺度差异。小尺度和大尺度需要的绕行步数相近（2326.95 和 2273 步），而多尺度策略大幅减少了所需步数（686 步）。

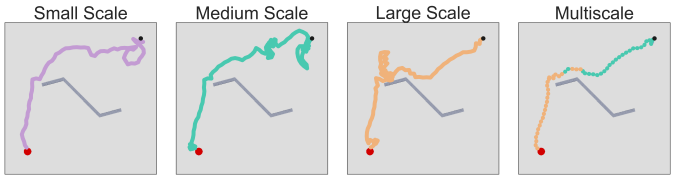
环境 3（两个障碍物）。小尺度和多尺度的性能几乎相同（838.90 对 847.90 步），但多尺度的方差较低。中尺度和大尺度需要的步数显著更多（1522.90 和 1815.55 步）。

环境 4（被遮挡的目标）。目标被边界墙遮挡时，多尺度再次表现最佳（780 步）。其次是小尺度（995 步），而中尺度和大尺度需要绕行更长的路线（2616 和 3752.35 步）。

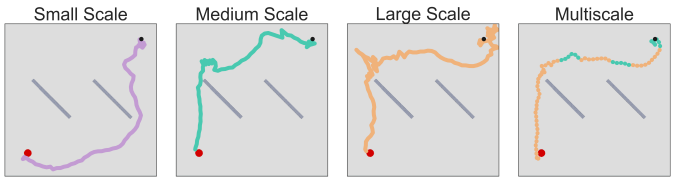
在所有环境中，多尺度策略达到或超过了最强的单尺度基线。单尺度结果揭示了互补优势：粗尺度



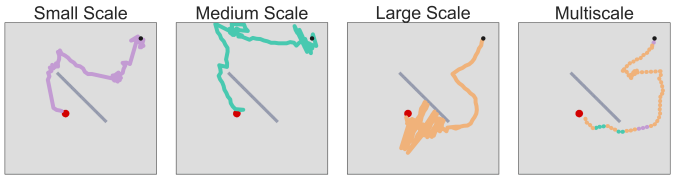
(a) 环境 1



(b) 环境2



(c) 环境3



(d) 环境4

图5. 随机选择的轨迹。在多尺度运行中，每个路径段上的颜色表示主导决策的尺度，但未显示为主导尺度的尺度仍可能处于激活状态。

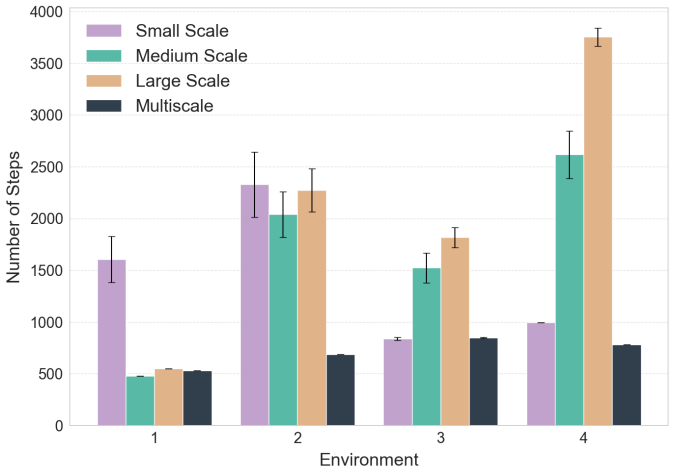


图6. 按策略和环境划分的路径效率（步数）。误差条表示均值标准误。

支持开放空间中的高效全局引导，而细尺度则能在杂乱环境中实现精确的局部机动。多尺度策略通过在每一步根据各尺度的方向性奖励变化对尺度进行加权来利用这些特性，从而持续产生更短、更可靠的轨迹，并具有更低的跨试验方差。

C. 实验2：策略学习评估

1) 概述：本实验以莫里斯水迷宫任务[49]–[51],为范本，在无障碍环境中单独考察学习动力学。在生物学范式中，啮齿动物起初在没有平台位置先验知识的情况下探索环境，并在连续的试验中逐渐形成稳定的目标导向轨迹。类似地，本实验中的智能体从一无所知开始，通过探索、奖励和重放来改进其导航策略。

与使用预训练空间表征评估导航表现的实验 1 不同，实验2衡量这些表征如何在线形成并趋于稳定。因此，早期轨迹预计会有所变化，本实验的重点是收敛行为，而非理想的路径效率。

2) 实验程序：每种策略（小、中、大和多尺度）均在每次试验中评估 51 个回合（0–50），共进行 5 次试验，因此每种策略共计 255 个回合。所有评估均在环境1中进行。该环境是一个开放场地，旨在尽量减少障碍物带来的混杂效应，并单独考察策略学习。该流程分为两个阶段：

- **初始朴素探索（第0集）**。智能体一开始没有先验位置场、空间邻接关系或奖励关联。导航最初是随机的，所有突触连接都保持可塑性，使空间结构能够通过经验逐渐形成。此阶段在每次试验开始时进行一次。

- **奖励引导导航与策略优化（第1–50回合）**。首次遇到目标后，反向重放事件会沿经历轨迹向后传播奖励信息，从而建立内部奖励地图。之后的回合（从第1回合开始）利用该地图引导智能体向高价值区域移动，同时继续优化空间表征和导航策略。

为评估学习效率，我们记录了每回合结束时到达目标所需的步数。由于第0回合和最初几个获得奖励的回合存在暂态变化，因此分析中排除了第0至第5回合。我们计算了第6至第50回合（45个回合）的平均步数，以此稳定地衡量收敛情况。如果智能体在模拟时间的120分钟内未能到达目标，则终止该回合并记录相应的步数。

3) 结果与讨论：图 7 展示了每种策略的轨迹如何随回合推进而演变。第0集反映了在奖励地图尚不存在时，智能体进行的朴素目标寻求探索。首次遇到目标后，奖励单元的重放机制会迅速传播价值信息，轨迹也变得越来越直接。小尺度导航仍呈现振荡（尤其是在智能体远离目标时），中尺度和大尺度的改进则更为渐进，而多尺度智能体收敛到平滑、高效路径的速度最快。

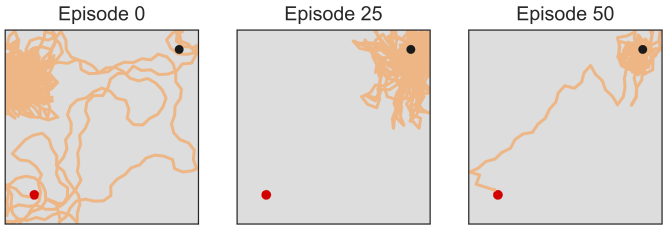
图 8a 和图 8b 对这些效果进行了量化。图 8a 展示了每种导航策略下，智能体在各回合中到达目标所需的步数，并对所有试验的数据进行了汇总。



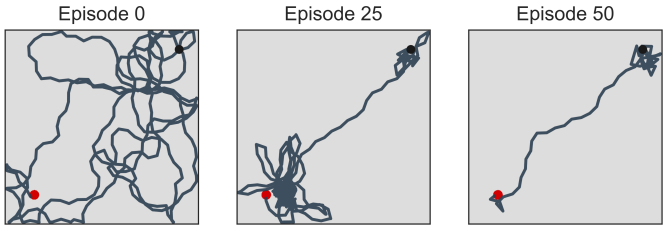
(a) 小尺度



(b) 中尺度

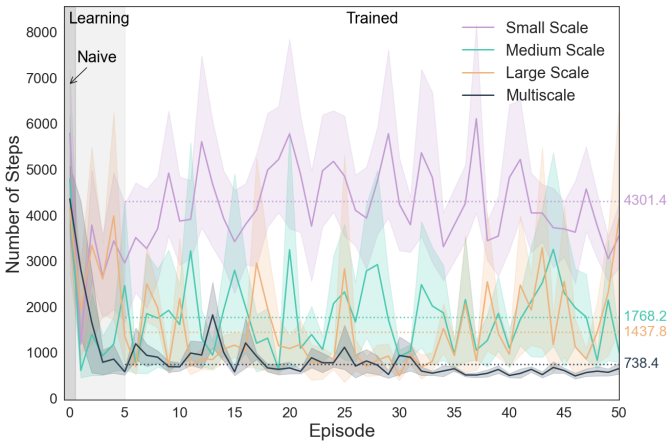


(c) 大尺度

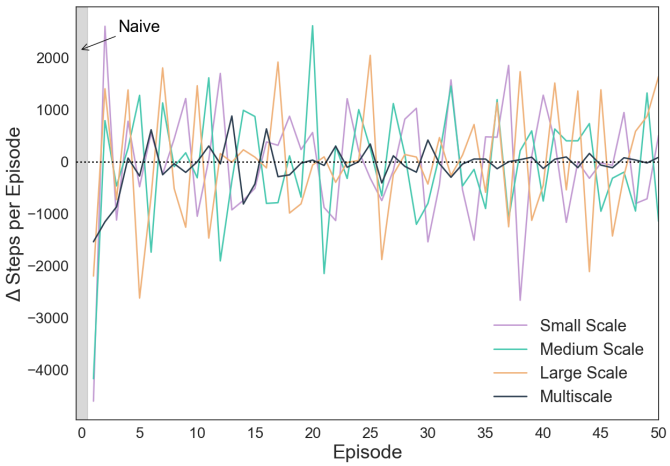


(d) 多尺度

图 7. 展示策略随回合优化的路径（左→右）。



(a) 步数收敛



(b) 步数变化率

图 8. 不同导航策略下的收敛动态。(a) 带有均值标准误阴影的步数收敛情况；虚线表示最终 45 个回合的均值。Naive、学习和训练回合以垂直阴影标示。(b) 连续回合之间步数（ΔSteps）的变化率；虚线表示 ΔSteps $= 0$（无变化）。

图 8b 展示了图 8a 所报告的步数变化率（Δ步数），该数值通过计算所有试验中相邻回合平均步数之差得出。低于零基线的值表示有所改善（平均而言，步数少于上一回合），高于零基线的值则表示有所退步。低于水平基线（Δ步数 < 0）的值表示智能体在这些回合中有所改善（步数少于上一回合），高于该基线的值则表示有所退步。

小尺度。小尺度策略产生了最高的最终步数（~4301），反映出获取高效全局导航策略的困难。尽管细粒度位置场提供了精确的局部线索，但也使智能体对感知状态中的微小变化高度敏感。这种敏感性在 ΔSteps 轨迹中表现得尤为明显，该轨迹呈现出交替出现的大幅负向和正向波动；这种不稳定性表明经常发生过度校正，反复抵消先前取得的改进。

中尺度。中尺度策略实现了显著更低的最终步数（~1788）。其 ΔSteps 曲线呈现中等程度的波动——在学习早期主要为负值，随后出现幅度较小、间歇性的反转。这一模式表明，中尺度表征能够在保留细粒度优化中一定不一致性的同时，支持有效的全局导航。

大尺度。大尺度策略进一步收敛（~1438），这与粗粒度位置场在开放环境中提供强大全局引导的情况一致。ΔSteps 信号包含持续的负向低谷，对应于显著的改进，其间穿插着偶发的正向尖峰，这些尖峰可能源于最终接近目标时的不精确性。与较小尺度相比，大尺度智能体表现出较少的 destabilizing 调整，但回合后期的变异性仍然存在。

多尺度。多尺度策略最终步数最低（~738），并展现出最快、最平稳的

表III 步数的方差分析和 Tukey诚实显著差异检验结果。

策略比较	均值差	p值	显著性
大尺度 vs. 中尺度	330.4	0.322	No
大尺度与多尺度	-699.4	0.0018	**Yes**
大尺度与小尺度	2879.9	0.000	**Yes**
中尺度与多尺度	-1029.8	0.000	**Yes**
中尺度与小尺度	2549.5	0.000	**Yes**
多尺度与小尺度	3579.4	0.000	**Yes**

收敛。其 Δ步数轨迹始终紧密围绕零基线，在初始学习后方差明显降低，反映出早期快速提升，随后策略巩固趋于稳定。大多数 Δ步数值接近或略低于零。

a) 统计与定量分析：单因素方差分析证实，策略对步数有显著影响（$F = 127.36$，$p = 1.22 \times 10^{-68}$）。Tukey诚实显著差异检验事后比较（表III）显示，多尺度策略的表现显著优于所有单尺度策略（$p < 0.01$）；多尺度与小尺度之间的差异最大（3579.4 步；Cohen's $d = 0.94$，效应量较大）。与多尺度条件（37）相比，小尺度条件的均值标准误最高（412），反映出不稳定性和局部过拟合，这类似于短视强化学习策略中的高方差。大尺度与中尺度之间的差异不显著（$p = 0.322$），这表明在开放环境中二者的覆盖范围有所重叠，粗尺度表征在全局引导中占主导地位。总之，这些结果表明，尺度的动态整合通过减轻细尺度振荡和粗尺度的不精确性，提高了收敛效率与稳定性。

D. 实证发现总结

在各项任务中，多尺度策略的表现达到或超过了所有单尺度基线，且收敛速度更快、方差更低。其自适应加权提升了在障碍物密集环境中的鲁棒性，这一点得到了方差分析（$p = 1.22 \times 10^{-68}$）和图基检验（$p < 0.01$）的支持。这些发现表明，动态尺度整合能够有效地将远程规划与精确的局部调整结合起来，实现鲁棒导航。

V. 讨论

A. 偏差—方差与奖励剖面结构

结果表明，每种空间尺度都会为价值景观带来不同的统计特性。较粗尺度产生平滑、低方差的奖励剖面，有助于可靠地进行长距离移动，但在障碍物附近提供的细节有限。较细尺度产生更陡峭的方向梯度，从而能够进行精确的局部调整，但更容易受到噪声影响并出现振荡行为。基于方差的加权机制通过提升在当前步骤表现出最清晰方向结构的尺度，平衡了这些影响，而无需在各尺度之间进行固定的分工。这种动态选择解释了为何多尺度策略比任何单尺度方案都能产生更一致的轨迹，并降低跨试验方差。

B. 参数敏感性与尺度配置

模型性能取决于少数几个可解释的超参数，其中最重要的是边界向量细胞调谐宽度 σ_r，σ_θ，它们决定位置场大小，进而决定奖励图的空间频率成分。过小的 σ_r会放大噪声，并可能导致之字形运动；而过大的 σ_r 则会过度平滑梯度，使狭窄通道变得难以辨识。式（22）中的融合与门控设置（例如奖励有效性阈值和障碍物掩码）决定哪些尺度参与动作选择，以及对不安全方向的抑制程度。预演分辨率则进一步体现了计算成本与角度精度之间的权衡。

VI. 结论与未来工作

本文提出了一种导航模型，该模型在多个空间尺度上并行运行位置场群体，并通过基于变异的加权整合其价值估计。在四个场地中进行预训练路径效率评估时，多尺度策略在每种环境下均达到或超过表现最佳的单尺度基线，通常能够产生更短、更可靠的轨迹，且跨试验方差更低。在另一项独立的学习实验中，与任何单尺度对照相比，同一多尺度方案收敛得更快、变异性也更低，这一点从步数轨迹及其逐回合变化率中均可看出。总体而言，这些发现支持了这样一种观点：多尺度空间表征具有互补优势，而通过基于变异的加权将其融合，有助于提高路径效率并使价值更新更加稳定。

未来工作的主要方向包括真实世界评估、自动尺度发现、用于融合的学习得到的尺度比例、多目标导航、三维导航，以及与网格细胞的整合。

致谢

作者衷心感谢 Adedapo Alabi 在实现方面给予的指导和建议，并感谢 Dieter Vanderelst 与我们进行的有益讨论。

参考文献

[1] S. Thrun、W. Burgard 和 D. Fox，《概率机器人学》（智能机器人与自主智能体）。麻省理工学院出版社，2005年。[2] M. Rosenberg、T. Zhang、P. Perona 和 M. Meister，“迷宫中的小鼠展现出快速学习、顿悟式突破和高效探索”，*eLife*，第10卷，文章编号e66175，2021年。[3] J. O'Keefe 和 J. Dostrovsky，“海马体作为空间地图：来自自由活动大鼠单单位活动的初步证据”，《脑研究》，第34卷，第1期，第171–175页，1971年。[4] M. Moser、T. Hafting、M. P. Witter、E. I. Moser 和 M.-B. Moser，“小鼠的网格细胞”，海马体，第18卷，第12期，第1230–1238页，2008年。[5] D. J. Foster 和 M. A. Wilson，“清醒状态下海马位置细胞中行为序列的反向重放”，《自然》，第440卷，第7084期，第680–683页，2006年。[6] G. Dragoi 和 S. Tonegawa，“海马细胞集合对未来位置细胞序列的预演”，《自然》，第469卷，第397–401页，2011年。[7] H. F. ´Olafsd´ottir、F. Carpenter 和 C. Barry，“休息期间网格细胞与位置细胞的协调重放”，《自然—神经科学》，第19卷，第6期，第792页，2016年。[8] H. F. ´Olafsd´ottir、D. Bush 和 C. Barry，“海马重放在记忆与规划中的作用”，《当代生物学》，第28卷，第1期，第R37–R50页，2018年。

[9] K. B. Kjelstrup、T. Solstad、V. H. Brun、T. Hafting、S. Leutgeb、M. P. Witter、E. I. Moser 和 M.-B. Moser，“海马体空间表征的有限尺度”。《科学》，第321卷第5885期，第140–143页，2008年。[10] J. B. Hales、M. I. Schlesiger、J. K. Leutgeb、L. R. Squire、S. Leutgeb 和 R. E. Clark，“内侧嗅周皮层病变仅部分破坏海马位置细胞及依赖海马的位置记忆”，《细胞报告》，第9卷，第3期，第893–901页，2014年11月。 [在线]。网址：https://doi.org/ 10 1016 2014 10 009 。/ [11] j.celrep. . . E. C. Tolman，“大鼠与人类的认知地图”。《心理学评论》，第55卷，第4期，第189–208页，1948年。[12] J. O’keefe 和 L. Nadel，《作为认知地图的海马体》。牛津：克拉伦登出版社，1978年。[13] J. S. Taube、R. U. Muller 和 J. B. Ranck，“在自由活动大鼠后下托记录的头方向细胞。I. 描述与定量分析”，《神经科学杂志》，第10卷，第2期，第420–435页，1990年。[14] J. O’Keefe 和 N. Burgess，“海马神经元位置场的几何决定因素”，《自然》，第381卷，第6581期，第425–428页，1996年。[15] C. Barry、C. Lever、R. Hayman、T. Hartley、S. Burton、J. O’Keefe、K. Jeffery 和 N. Burgess，“ 位置细胞放电与空间记忆的边界向量细胞模型 ”，《神经科学评论》，第17卷，第1–2期，第71页，2006年。[16] M. Jung、S. I. Wiener 和 B. L. Mcnaughton，“大鼠背侧与腹侧海马单位空间放电特征的比较”，载于《神经科学杂志》，1994年。 [在线]。网址：https://api.semanticscholar.org/CorpusID:18951767 [17] G.-q. Bi 和 M.-m. Poo，“培养的海马神经元中的突触修饰：对脉冲时序、突触强度和突触后细胞类型的依赖性”，《神经科学杂志》，第18卷，第24期，第10 464–10 472页，1998年。 [在线]。网址：https://doi.org/10.1523/JNEUROSCI.18-24-10464.1998 [18] K. L. Stachenfeld、M. M. Botvinick 和 S. J. Gershman，“海马体作为预测地图”，《自然—神经科学》，第20卷，第11期，第1643–1653页，2017年。 [在线]。网址：https://doi.org/10. 1038/nn.4650 [19] J. Epsztein，“心理重放助力灵活导航”，2022年。[20] A. K. Lee 和 M. A. Wilson，“慢波睡眠期间海马体对序列经验的记忆”，《神经元》，第36卷，第6期，第1183–1194页，2002年。[21] H. F. ´Olafsd´ottir、D. Bush 和 C. Barry，“海马重放在记忆与规划中的作用”，《当代生物学》，第28卷，第1期，第R37–R50页，2018年。[22] A. E. Papale、M. C. Zielinski、L. M. Frank、S. P. Jadhav 和 A. D. Redish，“海马尖波涟漪事件与决策中的替代性试错行为之间的相互作用”，《神经元》，第92卷，第5期，第975–982页，2016年。[23] A. D. Redish，“替代性试错”，《自然—神经科学评论》，第17卷，第3期，第147–159页，2016年。[24] D. Silver、A. Huang、C. J. Maddison、A. Guez、L. Sifre、G. Van Den Driessche、J. Schrittwieser、I. Antonoglou、V. Panneershelvam、M. Lanctot、S. Dieleman、D. Grewe、J. Nham、N. Kalchbrenner、I. Sutskever、T. Lillicrap、M. Leach、K. Kavukcuoglu、T. Graepel 和 D. Hassabis，“借助深度神经网络和树搜索掌握围棋”，《自然》，第529卷，第7587期，第484–489页，2016年。 [在线]。网址：https://doi.org/ 10 1038 16961 。/nature D. Hafner、J. Pasukonis、J. Ba 和 T. Lillicrap，“通过世界模型掌握多样化领域”，arXiv预印本 arXiv: *2301 04104*, 。版本最后修订于4月。在线。网址：https://arxiv.org/abs/ 2301 04104 。R. S. Sutton 和 A. G. Barto，《强化学习：导论》。麻省理工学院出版社，2018年。[27] N. Burgess、M. Recce 和 J. O’Keefe，“ 海马功能模型 ”，《神经网络》，第 7 卷，第 1065–1081 页，1994年。[28] A. Samsonovich 和 B. McNaughton，“连续吸引子神经网络模型中的路径积分与认知地图”，《神经科学杂志》，第17卷，第5900–5920页，1997年。[29] M. Tsodyks，“海马空间地图的吸引子神经网络模型”，《海马体》，第9卷，第4期，第481–489页，1999年。[30] S. Doboli、A. A. Minai 和 P. J. Best，“潜在吸引子：海马体中依赖情境的位置表征模型”，《神经计算》，第12卷，第5期，第1009–1043页，2000年。[31] E. I. Moser、E. Kropff 和 M.-B. Moser，“位置细胞、网格细胞与大脑的空间表征系统”，《神经科学年度评论》，第31卷，第69–89页，2008年。[32] U. M. Erdem 和 M. Hasselmo，“一种基于网格细胞和前向轨迹规划的目标导向空间导航模型”，《欧洲神经科学杂志》，第 35 卷，第 6 期，第 916–931 页，2012 年。[33] V. Edvardsen、A. Bicanski 和 N. Burgess，“在杂乱环境中利用网格细胞和位置细胞进行导航”，《海马体》，第 30 卷，第 220–232 页，2020 年。[34] A. Alabi、A. A. Minai 和 D. Vanderelst，“通过重放，在海马启发式强化学习模型中实现一次性空间学习”，载于 2020 年国际神经网络联合会议（IJCNN）。IEEE，2020 年，第 1–8 页。[35] A. Alabi、D. Vanderelst 和 A. A. Minai，“基于一种新型海马位置场模型的目标导向导航空间表征快速学习”，《神经网络》，第 161 卷，128，2023 年，第 116 页。– .[36] U. M. Erdem 和 M. E. Hasselmo，“一种受生物学启发的分层目标导向导航模型”，《巴黎生理学杂志》，第 108 卷，第 1 期，第 28–37 页，2014 年。[37] M. J. Milford、G. F. Wyeth 和 D. Prasser，“RatSLAM：一种用于同步定位与建图的海马模型”，载于 IEEE 国际机器人与自动化会议论文集（ICRA’04），2004 年，第 1 卷。IEEE，2004 年，第 403–408 页。[38] U. M. Erdem、M. J. Milford 和 M. E. Hasselmo，“一种分层目标导向导航模型在视觉环境中选择轨迹”，《学习与记忆神经生物学》，第 117 卷，第 109–121 页，2015 年。[39] M. Llofriu、G. Tejera、M. Contreras、T. Pelc、J. M. Fellous 和 A. Weitzenfeld，“利用多尺度空间表征进行目标导向机器人导航学习”，《神经网络》，第 72 卷，第 62–74 页，2015 年。[40] P. Scleidorovich、M. Llofriu、J.-M. Fellous 和 A. Weitzenfeld，“一种结合背侧和腹侧海马位置场图的空间认知计算模型：多尺度导航”，《生物控制论》，第 114 卷，第 17 期，第 187–207 页，2020 年。[41] P. Scleidorovich、J. M. Fellous 和 A. Weitzenfeld，“在杂乱环境中调整海马多尺度位置场分布，以优化空间导航和学习”，《计算神经科学前沿》，第 16 卷，2022 年。[42] A. A. Alabi，“基于一种新型海马位置场模型的目标导向导航自组织空间表征快速学习”，博士论文，辛辛那提大学，俄亥俄州辛辛那提，2022 年，委员会主席：Ali A. Minai，博士。[43] T. Neher、A. H. Azizi 和 S. Cheng，“从网格细胞到具有逼真场大小的位置细胞”，《公共科学图书馆·综合》，第 12， 17， 0181618, 卷，第 2017 [ ] p. e . 在线。可查阅：https://journals.plos.org/plosone/article?id 10 1371 0181618 。/ journal.pone.[44] D. Lyttle、B. Gereke、K. K. Lin 和 J.-M. Fellous，“海马体背腹轴网格细胞到位置细胞模型中的空间尺度与位置场稳定性”，《海马体》，第 23， 8， 729 744, 卷，第 2013 [ ] 10 期，第 页–。在线。可查阅：https://onlinelibrary.wiley.com/doi/abs/ 。1002 22132 /hipo. E. Oja，“作为主成分分析器的简化神经元模型”，[45]《数学生物学杂志》，第 15 卷，第 3 期，第 267–273 页，1982 年。[46] J. L. Gauthier 和 D. W. Tank，“海马体中专门用于奖励编码的细胞群”，《神经元》，第 99 卷，第 7 期，2018 年，第 179–193 页，编号 1，e .[47] Z. Xiao、K. Lin 和 J.-M. Fellous，“背侧远端CA1区海马细胞的奖励—位置联合编码特性”，《生物控制论》，第 114 卷，第 285–301 页，2020 年。[48] Webots，“http://www.cyberbotics.com”，开源移动机器人仿真软件。 [在线]。可查阅：http://www.cyberbotics.com [49] R. Morris，“用于研究大鼠空间学习的水迷宫程序的改进”，《神经科学方法杂志》，第 11 卷，第 1 期，第 47–60 页，1984 年。[50] C. V. Vorhees 和 M. T. Williams，“莫里斯水迷宫：评估空间学习及相关学习与记忆形式的流程”，《自然·实验手册》，第 1 卷，第 2 期，第 848 页，2006 年。[51] R. G. Steele，R J Morris，“动物学习因果因素的初步实验。II”，《比较与生理心理学杂志》，第 9 卷，第 118–136 页，1999 年。