DOI:10.13203/j.whugis20210469 文章编号:1671-8860(2021)12-1819-13

# 类脑导航算法:综述与验证

郭 迟[1,2] 罗宾汉[1] 李 飞[2] 陈 龙[3] 刘经南[1]

1 武汉大学卫星导航定位技术研究中心,湖北 武汉,430079
2 武汉大学人工智能研究院,湖北 武汉,430079
3 中山大学数据科学与计算机学院,广东 广州,510275

**摘 要**:类脑导航算法是近年来的新兴研究热点,这类算法通过对生物导航能力的模仿实现自主导航,核心问题是如何提升泛化能力。介绍了类脑导航算法的研究背景与理论基础,经过调研总结出了其计算框架;以类脑导航算法计算框架为骨干对该领域的突出工作进行了讨论分析,并通过严格的控制变量实验验证了一些典型改进方法的效果。主要贡献有:全面地介绍并总结了类脑导航领域的理论基础与突出工作;总结出了类脑导航算法的计算框架,该框架科学定义了算法不同部分的职能,从而能解构具体的算法,完成细粒度的分类和对比;通过理论分析与实验验证,总结出了有价值的结论,并展望了未来的发展。

**关键词**:类脑导航;人工智能;自主导航;感知;记忆;策略

**中图分类号**:P228;TP242.6 **文献标志码**:A

自主导航作为移动机器人的基础能力,是机器人科学和人工智能的一大挑战。过去几十年间,基于精确结构化的、几何建模的传统方法产出了许多实用成果。随着心理学、脑科学和机器学习的不断发展[1-2],学术界涌现出一类全新的移动机器人导航算法,称为类脑导航算法。此类算法具有如下特点:①利用神经网络模拟生物导航能力,无需对环境精确建模[3];②通过与环境交互学习导航能力[4];③能够在从未见过的陌生环境中自主导航[5]。

传统的导航方法多基于几何的、结构化的地图,需要通过精确测量和计算完成模型的构建和利用,在难以测量场合的使用有所受限。而类脑导航方法并不依赖于精确的测量和几何建模,是通过神经网络形成对环境的某种认知并进行导航规划。目前的类脑导航算法主要是在目标驱动任务下研究,即需要在一个场景中以视觉为基础,导航到一个以某种形式指定的导航目标前。

类脑导航研究如何像生物一样导航,通过对生物导航能力的模仿探索生物导航能力的内在机理。笔者认为,生物导航的核心能力有:(1)从来自感官的高维信息中感知并提取有用信息;(2)将这些信息暂时保存下来,形成短期记忆,并基于当前观测的信息和短期记忆中的信息推理当前所处环境的全貌,形成某种关于环境的内在表示;(3)根据这种内在表示完成规划与决策。生物导航能力的获得主要依靠与环境的交互中积累对学习导航能力有益的经验;而导航能力的学习的主要依据有生物趋利避害的天性,以及模仿其他生物的交互行为的能力。

心理学和脑科学的进展帮助人们了解到获取这种导航能力需要的关键结构和学习机制;而深度学习的兴起为模仿这些结构和机制提供了基础:多种多样的神经网络可用于模仿生物大脑的不同结构,而深度强化学习、模仿学习的出现让人们可以在导航这样的序贯决策问题下完成网络模型的训练,最终希望网络模型获得像生物一样的导航能力,能够不依赖精确传感器、不构建精确几何地图和不依赖强定位手段,以视觉为基础,在陌生环境也能完成自主导航任务。

算法需要很强的泛化能力,才能在完全陌生的环境中成功导航,如何提升泛化能力是该领域研究的核心问题。而提升泛化能力的主要难点有过拟合、数据效率低、记忆容量小且不稳定等。新兴领域的一批起步工作有算法描述参差、细节差别大等特点,但在本文总结出的计算框架下分

**收稿日期**:2021-08-27

**项目资助**:国家重点研发计划(2016YFB0501801)。

**第一作者**:郭迟,博士,教授,主要从事北斗应用、无人系统导航与基于位置服务研究。guochi@whu.edu.cn

析，能清晰地判断某一个具体工作的不同部分对这些子问题的贡献，实现细粒度的分类和对比；另外，还对一些得到认可的改进工作在严格的控制变量法下进行了实验验证，有一定的参考意义；综合理论与实验产生的结论和预测能对该领域发展起到积极的推动作用。

## 1 类脑导航问题的数学描述

类脑导航算法从一种交互的、上层决策的角度来看待导航问题，如图1所示，可以将问题描述为马尔可夫决策过程（Markov decision process，MDP）。MDP将学习者和决策者称为智能体，将与之交互的系统称为环境。智能体基于环境的状态选择要执行的动作，环境基于该动作更新自己的状态，并反馈以奖励，从一次交互的开始到结束称为一次试验。一个MDP由以下几个要素构成：一个环境的状态集合$S$，一个智能体可以采取的动作集合$A$；状态转移概率$p(s'|s,a)$，它表示在状态$s$下采取动作$a$后转移到状态$s'$的概率；奖励函数$r(s,a,s')$，即在状态$s$采取动作$a$转移到状态$s'$后能够获得的奖励[6]。给出以下定义，从而方便将类脑导航问题描述为MDP。

定义1：定义状态$s\in S$包含智能体所处的具体物理场景和导航任务的全部信息，包括智能体自身的位姿、速度等变量。

定义2：定义观察$o_t\in O$是状态$s$的函数，可设$o_t=f(s_t)$，且$o_t\neq s_t$，主要包含在当前位姿下摄像机对物理场景的观测图像（记为$x$）、导航目标的描述（记为$g$）。可以进一步包括自身的速度（记为$v$）、动作（记为$a$）、姿态角（记为$\phi$）等信息。

定义3：定义动作集合$A$为离散的不同动作$a$的集合，动作$a$是上层运动动作，例如前进0.5 m，左转30°等。

定义4：定义一个特殊的动作，记为Done，表示智能体已经完成导航任务并终止本次交互，此时如果导航目标在智能体视野范围1 m内即判定导航成功，否则判定为失败。

定义5：定义智能体的策略$\pi$为一个以观察为输入，在动作空间$A$上的概率分布：

$$a_t\sim\pi(o_t) \tag{1}$$

概率越高的动作被选择的可能性越大。使用神经网络来近似该策略函数，记为策略模型$\pi(o_t;\theta)$。

定义6：定义交互轨迹$\tau_{t:T}$为智能体根据策略$\pi$与环境的交互产生的状态-动作-奖励链：

$$s_t,a_t,r_{t+1},s_{t+1},a_{t+1},r_{t+2}\cdots s_{T-1},a_{T-1},r_T,s_T$$

定义该轨迹的累积回报：

$$G_{t:T}=\sum_{i=t}^{T-1}r_{i+1} \tag{2}$$

定义7：定义智能体与环境的最大交互步数$T_{\max}$。当达到最大交互步数$T_{\max}$时，本次实验终止，且认为智能体导航失败。

定义8：定义状态价值函数为从某个状态开始，采取策略$\pi$后，能够最终获得的累积回报的期望：

$$v_\pi(s)=E_\pi\left[G_t|s_t=s\right] \tag{3}$$

在无法获得状态$s$时，我们可以基于观察$o_t$使用神经网络来近似状态价值，记为$V(o_t;\boldsymbol{\theta})$。

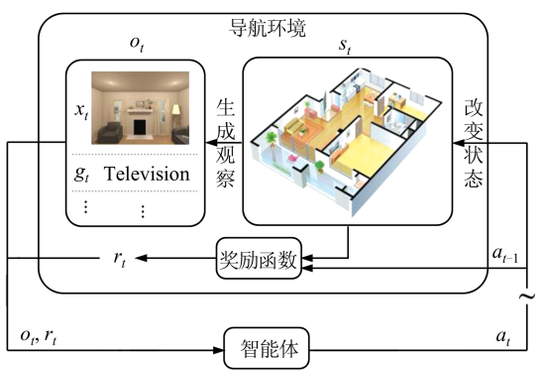



图1 导航智能体与环境的交互

Fig.1 Interaction of Navigation Agent and Environment

## 2 类脑导航问题的模拟环境

在现实世界中进行导航交互训练存在如下问题：效率低下并且有危险性，不容易控制和重现实验条件。因此，需要模拟的导航环境用于训练。模拟环境主要分为真实重建环境与仿真合成环境。真实重建环境基于真实环境的扫描数据集，如Habitat[7]、支持Gibson[8]、Matterport3D[9]（M3D）、Replica[10]等数据集；仿真合成环境是使用计算机图形技术构建的虚拟环境，如AI2-THOR[11]等。真实重建环境和现实环境的差距更小，运行速度更快，但无法进行物理交互；仿真合成环境虽然可以对环境进行修改布置和物理交互，但运行速度相对较慢，与现实的差距明显。表1列举了6个典型的模拟环境。

本文涉及实验都基于AI2-THOR模拟环境进行，该模拟环境拥有4类室内场景（厨房、客厅、卧室、浴室，如图2所示），每类场景各有30个实例，一共120个室内场景，包含各种各样的物体可作为导航目标。

**表 1　不同模拟环境简介**

Tab. 1　Summary of Different Simulated Environments

| 模拟环境 | 数据集 | 场景规模 | 场景修改 | 亮点 |
|---|---|---|---|---|
| DeepMind Lab[12] | 渲染合成 | 小型迷宫 | 支持 | 高度可定制化 |
| AI2-THOR[11] | 渲染合成 | 室内单个房间 | 支持 | 物体对象可交互模拟真实物理 |
| RoboTHOR[13] | 渲染合成以及真实场景 | 室内单个房间 | 支持 | 仿真环境对应真实存在的场景 |
| MINOS[14] | SUNCG[15] M3D | 多房间完整室内建筑 | 支持 | 多模态传感信息 |
| House3D[16] | SUNCG | 多房间完整室内建筑 | 支持 | 支持房间导航任务 |
| Habitat[7] | M3D Gibson,Replica | 多房间完整室内建筑 | 不支持 | 运行高速,可导入自定义数据集 |

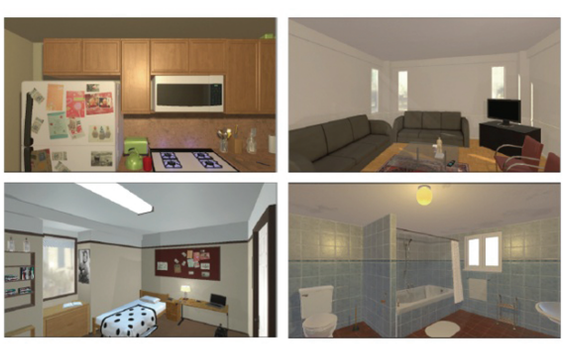


图 2　AI2-THOR 场景实例

Fig.2　Instances of AI2-THOR Scenes

## 3　类脑导航计算框架

对策略模型的训练与设计是类脑导航方法的核心。结合学术界现有的工作和上述生物导航能力的总结,本文设计类脑导航算法框架如图 3 所示。

策略模型的训练分为学习算法和经验生成两个部分,一个负责数据的使用,一个负责数据的获取。学习算法对应生物以趋利避害或者模仿为准则来学习的能力,主要分为在线学习算法和离线学习算法。在线学习算法包含强化学习算法和模仿学习算法(需要专家动作 $a_t^e$),模仿生物在交互或模仿中学习的过程,使用交互轨迹 $\tau$ 来计算损失函数并反向传播更新模型参数(如图 3 中 $L$ 与蓝色虚线)。离线学习算法主要用于预训练模型的某些部分,尤其是感知模型(如图 3 中绿色虚线)。

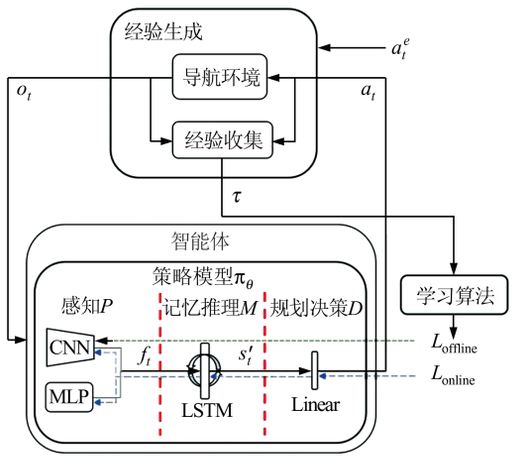



图 3　类脑导航计算框架

Fig. 3　Calculation Framework of Brain-like Navigation

经验生成对应生物在交互中积累有益于学习的经验的能力。该模块需要控制环境与相关超参数(如奖励函数),并收集环境与智能体交互所产生的交互轨迹 $\tau$ 交予学习算法。设计更优的奖励函数、生成高效率的经验对学习至关重要。异步优势函数的表演者-评论家(asynchronous advantage actor-critic, A3C)[17] 算法即是多线程并行多组环境与智能体,产生异步的经验用于训练,取得了很好的效果。

策略模型通常使用人工神经网络来构建,主要包含 3 个部分:感知模型 $P$、记忆推理模型 $M$ 以及规划决策模型 $D$,每个部分一般采用不同的神经网络结构来实现具体的功能。

感知模型对应生物从来自感官的高维信息中感知并提取有用信息的能力。接收到原始数据(观察 $o_t$)后,感知模型 $P$ 将其提取特征并降维得到初级表示 $f_t$:

$$f_t = P(o_t) \tag{4}$$

由于观察一般含有图像和以自然语言形式指定的目标,因此感知模型往往需要卷积神经网络(convolutional neural network, CNN)和多层感知器(multilayer perceptron, MLP)。

记忆推理模型对应生物将初级表示暂时保存下来,形成并更新短期记忆 $m_t$ 的能力;同时也对应生物基于当前观测的信息和短期记忆中的信息推理估计环境的全貌的能力。接收到初级表示 $f_t$ 后,记忆推理模型 $M$ 将更新记忆 $m_t$,并完成高级的推理行为,得到关于环境的某种高级内在表示,记为 $s_t'$:

$$s_t', m_t = M\left(f_t, m_{t-1}\right) \tag{5}$$

通常需要使用有记忆能力的神经网络，如图3中使用的长短时记忆神经网络(long short-term memory，LSTM)[18]。

规划决策模型对应生物根据所构建的内在表示完成规划决策的能力。规划决策模型$D$基于这种高级表示输出动作空间上的概率分布：

$$a_t \sim D(s_t') \tag{6}$$

通常使用线性层和softmax函数输出概率分布，网络图中一般省略softmax层，如图3所示。

根据上述归纳，可将不同的模型方法分为4栏，如图4所示，灰色的网络代表使用对应的损失函数离线训练。使用该框架梳理不同工作后，可以容易地比较不同模型的核心与差异所在。

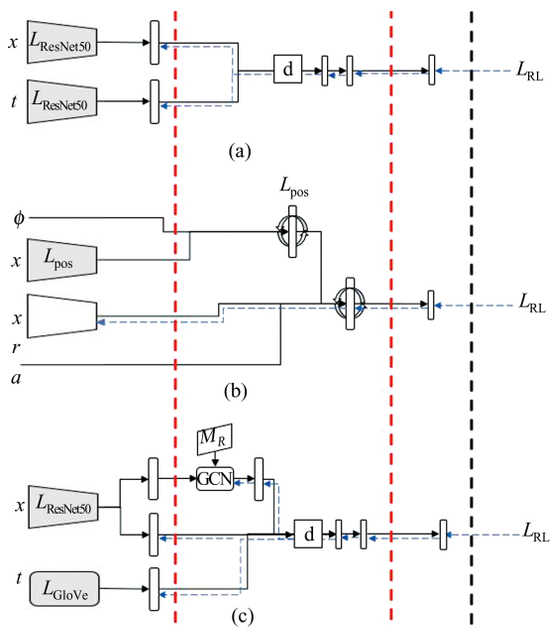



图4 模型设计与梯度传播图

Fig. 4 Model Design and Gradient Propagation

## 4 类脑导航的策略模型训练

### 4.1 学习算法

学习算法分为在线学习方法和离线学习算法，作用为模仿生物的趋利避害或者模仿的学习能力，完成对策略模型的参数更新。离线学习方法利用从导航环境中提取的离线数据或者外部数据集来完成训练，主要是用于模型的某个部分，尤其是感知模型。在线学习方法对应生物在趋利避害的交互或模仿中学习到导航能力的过程。其基础方法有模仿学习和强化学习两类。

模仿学习方法的基本方法是行为克隆，是一种以专家动作为标签的有时序的监督学习方法。在模拟环境中获得最优路径并反推出每一个状态应当采取的最佳动作是容易的，解决了专家动作的获得问题；但行为克隆方法容易过拟合，因此Gupta等[19]采用Ross等[20]提出的模仿学习算法，在保持高学习效率的同时提高模型泛化能力。

强化学习的方法主要有基于价值函数的方法和基于策略梯度的方法。强化学习的优势在于不依赖专家动作，通过不断地与环境交互试错来学习，但可能存在非常严重的数据效率问题。其核心在于对价值函数的近似，并通过价值函数影响决策。

为了在学习算法层面解决数据效率问题，研究者陆续提出了辅助任务方法、元强化学习方法和强化＋模仿学习方法。

辅助任务方法是一种结合在线学习与离线学习的方法，可用于缓解数据效率问题，即辅助任务方法[21]。辅助任务方法让智能体在在线训练的过程中，使用在线训练获得样本数据，同时让整个模型或部分模型在线地完成一些离线的静态任务，例如Kulhanek等[22]和Mirowski等[23]采用的辅助任务包括与目标间的姿态角预测、像素控制和奖励预测等，带来了额外的监督信号，一定程度上缓解了数据效率问题。

元强化学习方法是解决小样本训练问题的元学习方法和强化学习相结合的方法。如Wortsman等[24]受模型无关的元学习算法[25]的启发，在测试阶段使用交互损失函数$L_{int}$继续更新参数(如图5中橙色虚线)以继续学习具体的某一个导航场景，达成小数据下快速适应的效果。

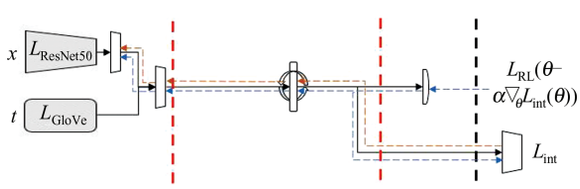



图5 元强化学习方法实例

Fig. 5 Instance of Meta-Reinforcement Learning Method

强化＋模仿学习方法的典型方法有Du等[26]结合强化学习与模仿学习算法，通过同时传播强化学习的梯度和模仿学习的梯度来训练模型。

此外，一些防止过拟合的技巧可应用在学习算法中：在损失函数中加入动作概率熵可以惩罚动作的单一性[18]；随机失活是一种有效地防止过拟合的方法，Banino等[3]证明随机失活方法对于形成六边形激活模式至关重要。各种防止过拟合的方法以及在强化学习中的效果可以在Cobbe等[27]的工作中查看。

### 4.2　经验生成

经验生成部分对应生物在交互中积累有益于学习的经验的能力,职能是控制环境与相关超参数(如奖励函数),并收集环境与智能体交互所产生的轨迹数据,主要是为了解决强化学习算法中数据效率低下的问题。数据效率低下的主要原因是交互数据的不稳定性,例如算法在训练之初产生的数据往往是随机的;其次是导航的奖励稀疏问题,即导航时智能体一般只有在交互的最后才能获得较大的奖励,其余时间都无法获得奖励甚至获得负奖励。

在经验生成层面解决强化学习的数据效率问题的方法主要有多线程并行采样、轨迹复用、课程学习和设计更优的奖励函数。

多线程并行采样:A3C 算法即是通过多线程并行多组环境与模型,每个线程都使用表演者-评论家强化学习算法,不同线程上的模型异步地更新梯度,有效提高了训练效率和稳定性。本文实验中也通过多环境并行获取经验,使用优势函数的表演者-评论家算法,但取消了异步梯度更新过程,完整流程如算法 1 所示。

算法 1:多环境并行的模型训练算法

设总训练帧数为 $F$,轨迹采样长度为 $n$,学习率为 $\alpha$,奖励折扣率为 $\gamma$。将策略函数参数记为 $\theta$,价值函数参数记为 $\theta_v$。

初始化 $N$ 个并行的导航环境,得到一组观察 $\boldsymbol{o}_1$。然后开始 $n$ 步经验的采集:根据策略 $\pi(\boldsymbol{o}_t;\theta)$ 采样执行一组动作 $\boldsymbol{a}_t$,收到一组回报 $\boldsymbol{r}_t$ 和一组新观察 $\boldsymbol{o}_{t+1}$。构建二进制向量 $\boldsymbol{m}_t=(m_1\ m_2\cdots m_N)$,每一个分量指示时间 $t$ 时对应的环境是否到达终止状态,如此循环 $n$ 次,每次令 $t=t+1$,$F=F-1$。采样完成后进行参数更新,令 $\boldsymbol{R}=V(\boldsymbol{o}_{n+1};\theta_v)$,对 $t\in\{n,n-1\cdots1\}$ 循环执行如下更新操作:

$$\boldsymbol{R}\leftarrow\boldsymbol{r}_t+\gamma\cdot\boldsymbol{m}_t\cdot\boldsymbol{R}$$

$$\boldsymbol{A}\leftarrow\boldsymbol{R}-V\left(\boldsymbol{o}_t;\theta_v\right)$$

$$\theta\leftarrow\theta+\alpha\nabla_\theta\log\pi\left(\boldsymbol{a}_t|\boldsymbol{o}_t;\theta\right)\cdot\boldsymbol{A}$$

$$\theta_v\leftarrow\theta_v+\alpha\partial\boldsymbol{A}^2/\partial\theta_v$$

再从经验采集开始继续循环,直到 $F=0$。

轨迹多目标复用:Lü 等[32]注意到一条轨迹可能经过多个目标,因此可以复用到不同目标训练过程中,从而提高训练效率。

课程学习:Mirowski 等[23,28]采用了课程学习[33]方法,模仿生物学习过程,在训练过程中逐步增加任务难度。例如随着实验次数增多,逐渐提高智能体和目标的距离,在训练初始阶段提高了有效数据的比例。

奖励函数设计:导航任务中奖励设计通常分为 3 个部分:任务完成时奖励、惩罚奖励和探索鼓励奖励。表 2 总结了具有代表性的奖励函数设计,发现鼓励智能体的探索行为能有效地提升导航效果。

**表 2　奖励函数代表性设计一览表**

Tab. 2　Representative Designs of Reward Function

| 文献 | 成功奖励 | 时间惩罚 | 碰撞惩罚 | 探索鼓励奖励 |
|---|---|---|---|---|
| Zhu 等[11] | 常量 (10) | −0.01 | −0.1 | — |
| Mirowski 等[28] | 常量 (10) | — | — | 地图上分布着奖励为 1 或 2 的“水果” |
| Mirowski 等[23] | 路径长度加权奖励 | — | — | — |
| Shi 等[29] | 常量 | — | 常量 | 通过预测下一个状态来衡量探索的程度,将探索程度作为额外的好奇心奖励 |
| Druon 等[30] | 常量(5) | −0.01 | — | 目标出现在视野中的外接矩形框是目前最大的时候能获得正比于矩形框大小奖励 |
| Ye 等[31] | — | — | — | 目标出现在视野中的外接矩形框是目前最大的时候能获得正比于矩形框大小奖励 |

## 5　类脑导航的策略模型设计

### 5.1　感知模型

感知模型的目的是对输入数据进行特征提取,去除数据中的冗余信息,保留有用信息,对应的是生物从来自感官的高维信息中感知并提取有用信息的能力。

感知模型的设计根据感知的数据对象和感知结果而异,例如图像数据需要卷积神经网络,独热码数据可使用线性网络。如果需要目标检测的外接矩形框作为显式的感知结果,则需要设计目标检测网络。

导航目标可以以多种方式表示,如单词、独热码和图像等。对不同目标表示的处理见表 3。使用单词来指定导航目标的方法一般采用离线训练好的词嵌入网络感知单词中的信息[24,34-35]。

由于可以借鉴深度学习在各种静态感知任

务的成功，感知模型的网络结构设计难度有所降低。但感知模型位置靠前、对图像的感知网络层数较深，以及强化学习算法的数据效率问题，更容易造成模型在在线学习算法中收敛缓慢甚至无法收敛。使用在线方式学习感知模型的方法往往只能在网络层数上做出妥协，如感知图像数据的卷积网络均不超过4层[3,16,23,28-29,39-41]。解决该问题的一个有效方法是离线学习。使用图像识别等静态感知任务来离线学习感知模型，可以回避在线学习的收敛问题。大部分工作[11,19,24,26,31-32,34-37,42]均使用了在图像识别任务上预学习的残差神经网络[43](ResNet50)来处理图像，如图4所示；Gordon等[38]不使用预学习模型，而使用模拟环境中提取出的数据重新离线学习感知模型。需要显式结果的网络一般都需要离线学习，如Ye等[31]的工作中需要目标检测的矩形框。离线学习使得人们可以使用更多样的网络来构建不同的初级表示，并在后续模型中进行显式或隐式地融合处理[36,44]。

**表3 不同目标表示处理**

Tab. 3 Different Treats of Target Representation

| 目标 | 对应网络 | 代表文献 |
|---|---|---|
| 独热码 | 线性 | [26,36] |
| 单词 | 词嵌入网络 | [24,34-35] |
| 物体图像 | 卷积网络 | [31] |
| 视点图像 | 卷积网络 | [11,37] |
| 相对位置 | 无处理 | [38] |
| 绝对位置 | 无处理 | [23] |

为感知模型设计专用的辅助任务也可以加速其收敛。Mirowski等[28]让智能体在导航过程中在线地完成一些与导航相关的静态感知任务，包括深度预测和回环检测，其中深度预测任务有两个，其中一个是专用于辅助图像感知网络的训练，如图6所示。这为网络训练带来了额外的监督信号，缓解了奖励稀疏问题，但没有放弃在线训练对感知能力的贡献。

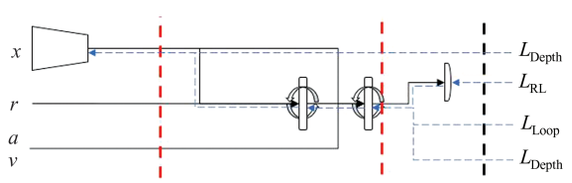



图6 辅助任务方法实例

Fig. 6 Instance of Auxiliary Tasks

类脑导航希望在一个模拟环境下训练得到的导航模型能够迁移到另一个模拟环境或者真实世界中，这要求感知模型能适应数据间的差异，继续提供有效的感知结果。Gordon等[38]在新数据下用离线学习的方式精调感知模型部分；Zhu等[41]通过一个对抗适应网络来精调在模拟环境下训练好的网络，使真实数据和仿真数据的特征向量有相同的概率分布和语义信息，如图7所示。

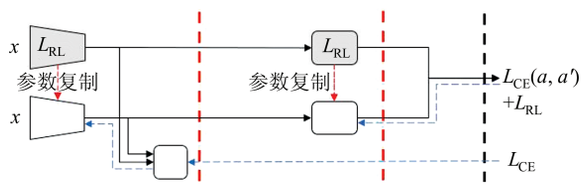



图7 对抗适应网络实现能力迁移

Fig. 7 Ability Transformation by Generative Adversarial Networks

### 5.2 记忆推理模型

记忆推理模型的目的是综合初级表示，形成记忆并基于记忆完成推理，得到对环境的较为全面而非结构化的高级内在表示。本部分对应生物对环境信息的记忆能力和基于观察与记忆推理建模环境全貌的能力。

记忆结构的容量问题是神经网络的共性问题。神经网络的网络参数可以看作一种记忆，但其容量非常有限，如果只使用线性层来完成本部分的网络设计，会导致智能体成为一种“反射式”的智能体。

一些早期工作[11,34]虽然使用线性层来完成该部分的工作，但它们通过延时堆叠输入的观察数据来满足对记忆的需求，即累积$n$步的观察作为一个新的有时间相关性观察。由于感知模型是静态的，为了保持结构划分一致，可以视为是对初级表示进行了延时处理，见图4(a)。若$n$值太小，那么智能体的记忆容量会偏小；若$n$值太大，则网络复杂度又会急剧增大。

循环神经网络作为拥有短期记忆能力的网络很自然地应用于模型中，例如将单层LSTM隐藏层看作智能体形成的对环境信息的记忆[24,26,36]。堆叠循环神经网络可完成更复杂的设计。Mirowski等[28]采用双层LSTM结构，第一层学习推理立即回报与图像之间的关系，得到情境信息供第二层LSTM使用，如图6所示。Banino等[3]构造的双层LSTM结构有明确的分工：一个通过路径积分离线学习，能够推理自身的位置并得到类似生物的表示(其实验证明最后线性层的激活情况类似于生物中的网格细胞激活模式)，

另一个LSTM层综合处理该类生物表示、图像特征、奖励和动作,推理构建完整的高级表示,见图4(b)。

基于循环神经网络的设计仍然存在容量小、记忆变化快的问题。为解决此问题,比循环神经网络更复杂的记忆神经网络开始应用于记忆推理中。Oh等[45]将记忆单元以队列形式储存,基于当前的情境,通过软注意力机制来读取。由于是只读的结构,智能体难以迭代形成对环境更充分的认识;Pritzel等[46]利用可微分神经字典[47]结构,拥有可学习的读写机制,因此能够迭代记忆,但学习参数的增加可能导致收敛时间变慢。

显式记忆方法可以利用有具体含义的记忆信息和计算结构来设计显式记忆方法。Gupta等[19]的工作为显式空间记忆的代表工作,其记忆数据为一张以智能体为中心的粗略认知地图$\mathrm{Map}_c$,并通过编解码网络,从视觉图像中得到感知地图$\mathrm{Map}_r$后,根据动作$a$的调整来更新,没有测量和监督过程,如图8所示。显式记忆也可以是拓扑形式,Yang等[34]采用了物体之间的关系拓扑图来帮助导航,关系图通过外部数据集得到并作为先验知识,用图卷积神经网络(graph convolutional network, GCN)[48]来提取特征向量用于决策,见图4(c);Lu等[35]使用显式机制在训练和测试中动态地将新关系写入关系图,并利用GCN得到的特征计算势函数估计两点之间的空间与语义距离。

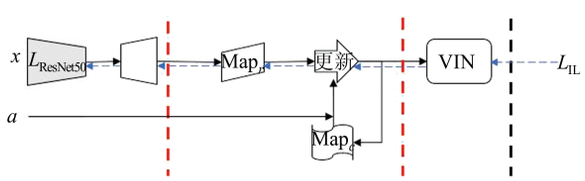



图 8　显式记忆推理与规划决策设计

Fig. 8　Explicit Designs of Memory and Plan Model

### 5.3　规划决策模型

规划决策模型的目的是基于推理得到的高级内在表示完成规划,并输出最终的决策动作。本部分对应生物基于对环境的认识和建模进行规划,并依据规划完成当前决策的过程。

由于规划过程在隐式过程中的不明显性以及决策过程可以用线性层简单实现,本部分工作量较少。Gupta等[19]使用值迭代网络[49](value iteration network, VIN)完成对路径的规划,VIN将迭代地对地图和其预测代价进行卷积计算,实际上是对值迭代公式的模仿,其卷积核最终能学习到状态的转移概率。经过多层多尺度的VIN计算,能够得到最小尺度上智能体的代价地图并指导当前动作的选择,见图8。

Shen等[44]利用25个视觉任务来预训练25个卷积网络并用模仿学习单独训练它们输出动作,然后用一个新的CNN来学习一种基于情境的注意力机制。该注意力为一个采样概率向量,在最终决策时基于该向量选择25个感知结果产生的动作中的其中一个,实现通过对具体情境的注意力来选择特定的策略,见图9。

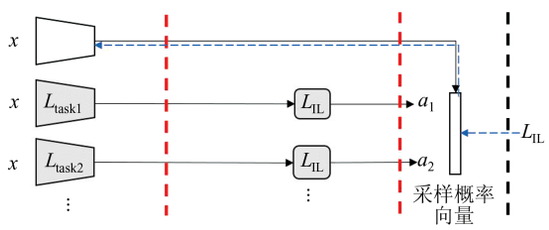



图 9　基于注意力的动作选择

Fig. 9　Action Selection Based on Attention Mechanism

Du等[26]提出尝试性策略网络,检测在短时间内是否出现重复的视觉表示,从而判断智能体是否进入锁死状态,并输出一个指导动作与当前选择的动作进行交叉熵计算,进行梯度反向传播,修正整个网络的策略。

Watkins-Valls等[50]通过监督学习训练了一个判断是否应该提出Done动作的策略网络,只有在该网络判定不应该提出Done动作时才进行其他动作的决策选择。

## 6　实验分析

前文对多种方法分模块进行了综述介绍,本节对其中被沿用较多的一些基本改进进行实验验证。在不同的工作中,这些基本改进是在不同的实验条件与设置下沿用的,人们无法直接对比它们的原始实验数据,因此需要通过严格的控制变量实验去确认这些方法在相同条件下的效果,以节省其他研究者的验证时间。本节将在严格的控制变量法下实现6个基础性的方法,每一个方法都在某个模块应用了某个基础改进,在完全相同的实验条件下验证这些方法在陌生环境下的导航效果。

值得一提的是,本文实验并不追求得到有较高绝对性能的算法,只需要算法间的相对性能足够判别出优劣即可。

### 6.1 实验设置

使用AI2-THOR环境训练和测试要对比的方法，4类房间中每类房间的1～15号房间，即共60个房间作为训练集；每类房间的16～20号房间，即共20个房间作为测试集。在测试集中的表现反映模型的泛化能力，显示了导航能力的优秀程度。

在每类房间各选择了4种目标作为导航目标，如表4所示，在训练和测试中使用相同的目标集合。

**表4 实验中选择的导航目标**

Tab. 4 Navigation Targets Chosen in Experiment

| 房间类型 | 目标 |
|---|---|
| 厨房 | Toaster, Microwave, Fridge, Coffee Maker |
| 客厅 | Pillow, Laptop, Television, Garbage Can |
| 卧室 | House Plant, Lamp, Book, Alarm Clock |
| 浴室 | Sink, Toilet Paper, Soap Bottle, Light Switch |

奖励函数和Zhu等[11]的工作相同，导航成功时得10，碰撞惩罚为－0.1，时间惩罚为－0.01。本实验的动作空间为$A=$｛前进，左转，右转，Done｝，共4个动作，其中左右转的角度为45°，Done动作对应定义4的描述。

### 6.2 评价指标

使用Anderson等[51]提出的两个评价指标用于评价模型的导航效果，即成功率（success rate，SR）和路径长度加权成功率（success weighted by path length，SPL），计算公式如下：

$$\mathrm{SR}=\frac{1}{N}\sum_{i=1}^{N}S_i \tag{7}$$

$$\mathrm{SPL}=\frac{1}{N}\sum_{i=1}^{N}S_i\frac{l_i}{\max\left(p_i,l_i\right)} \tag{8}$$

式中，$S_i$表示第$i$次实验是否成功导航到目标，成功为1，失败为0；$N$表示实验次数；$l_i$代表在第$i$次实验中智能体成功导航需要的最短路径长度；$p_i$代表智能体实际花费的路径长度，用采取的动作数量代替。

### 6.3 模型与算法细节

本实验一共对比了6种方法，如图10所示，6种方法形成两组对比实验，分别针对感知模型和记忆推理模型。规划决策模型目前工作量较少，也尚未受到高认可度的沿用，因此暂不专门验证其效果。控制模型的其他部分采用相同的设计：完全相同的学习算法、经验生成方法和超参数。

6种方法都使用单词作为目标的表示，用预训练好的词嵌入网络GloVe[52]处理得到300维的特征向量；输入图像大小为128×128×3像素；规划决策模型都使用简单线性层，都采用16线程的A2C算法，流程如§4.2中的算法1。训练超参数考虑了算法和任务的要求以及过往实验中的经验值，设置学习率$\alpha=0.0007$，采样步长$n=80$，奖励折扣率$\gamma=0.99$，总训练帧数$F=1\times10^8$，最大交互步数$T_{\max}=200$。

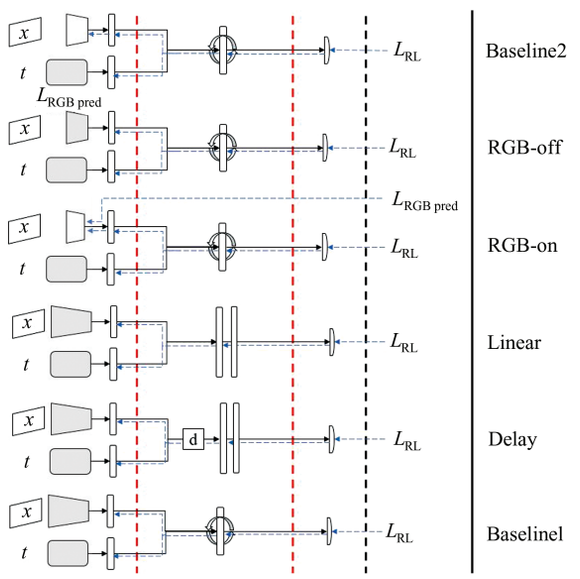



图10 实验网络结构与训练方法

Fig. 10 Model Designs and Training Methods for Experiment

感知模型对比组包含Baseline2、RGB-off和RGB-on，都采用LSTM作为记忆推理模型，仅有感知模型设计不同：Baseline2在线训练一个简单的4层卷积神经网络，RGB-off使用RGB预测任务离线训练的相同卷积网，RGB-on使用RGB预测任务作为简单卷积网的辅助任务。

记忆推理模型对比组包含Linear、Delay和Baseline1，都采用预训练的ResNet50作为感知模型，仅有记忆推理模型设计不同：Linear使用两层线性层，Delay堆叠4个时间步的观察，Baseline1使用单层LSTM。

其他细节包括：RGB预测任务要求网络将输入图像卷积处理成特征向量（编码）后再通过反卷积还原（解码）回图像。在RGB-on中RGB预测任务的损失函数乘以了0.0005，使其不至于完全覆盖强化学习损失函数的效果；离线预训练的RGB预测的效果如图11所示，第一行是模型的还原效果，第二行是原始图像。

### 6.4 实验结果

测试结果如表5所示，表5中列出了不同方

法在4种房间的SR与SPL值以及平均值。不同类型房间的性能差异较大,这是其面积大小和布局复杂度导致的。例如,卧室里有床和书桌使可移动区域变得复杂,客厅一般面积较大,找到目标需要更多的时间,因此所有算法在卧室和客厅的效果都较弱,而在面积小、布局简单的浴室效果最好。测试效果最优的Baseline1的实际导航效果如图12所示。

Baseline2、RGB-off和RGB-on 3个模型分别代表感知模型的在线训练、离线训练和辅助任务训练。从表5可以看出,在线训练的结果最差,辅助任务方法其次,而离线训练的结果最好。这说明了稳定的感知结果(即初级表示)更有利于导航能力的形成;但这3个方法的效果都低于使用ResNet50的另一组,可见使用离线训练的深层感知模型能够极大地提升导航性能。

(a) 模型的还原效果

(b) 原始图像

图 11　RGB还原效果图

Fig.11　Examples of RGB Reconstructions

**表 5　不同测试集测试结果(SR/SPL)**/%

Tab. 5　SR/SPL Results on Test Set/%

| 方法 | 厨房 | | 客厅 | | 卧室 | | 浴室 | | 平均值 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | SR | SPL | SR | SPL | SR | SPL | SR | SPL | SR | SPL |
| Baseline2 | 17.05 | 7.83 | 7.55 | 3.02 | 1.02 | 0.32 | 11.00 | 4.53 | 9.16 | 3.93 |
| RGB-off | 19.18 | 7.45 | 10.19 | 3.46 | 1.47 | 0.56 | 17.29 | 7.01 | 12.03 | 4.62 |
| RGB-on | 16.83 | 7.45 | 11.85 | 5.25 | 3.50 | 1.48 | 10.59 | 3.23 | 10.69 | 4.35 |
| Linear | 7.92 | 2.69 | 8.94 | 2.88 | 9.88 | 5.79 | 55.69 | 28.96 | 20.60 | 10.08 |
| Delay | 21.23 | 6.72 | 11.99 | 5.28 | 4.07 | 1.74 | 45.96 | 23.13 | 20.81 | 9.22 |
| Baseline1 | 27.52 | 9.59 | 9.10 | 3.32 | 9.95 | 4.51 | 54.47 | 26.00 | 25.26 | 10.85 |

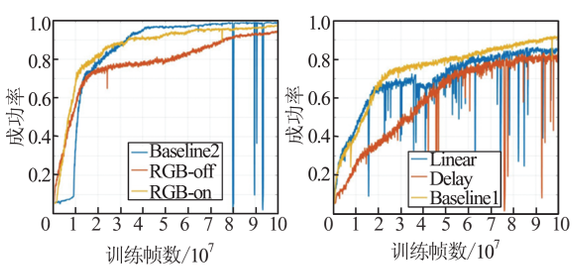

图 12　Baseline1在测试场景的导航的最后8帧图像观察与对应动作

Fig. 12　Last 8 Observations and Actions on Test Scenes by Baseline1

3个算法的训练阶段成功率曲线如图13(a)所示。在线训练方法在训练集上最终取得最优的成功率,但也变得非常不稳定,这可能是模型进入了过拟合状态。而辅助任务方法的收敛速度和成功率都优于离线训练方法。值得一提的是,本文并没有探索本实验条件下辅助任务方法的最优设置,例如辅助任务与强化学习损失函数之间的比例。

综上得出：(1)在线训练简单的卷积神经网络容易过拟合，无法完成有效的特征提取，从而泛化能力很差；(2)离线训练能够提供稳定的感知结果，这对导航来讲至关重要，但离线任务对导航问题的贡献度值得进一步探究；(3)辅助任务方法训练时收敛速度最快且稳定，能起到稳定训练、引导感知能力形成的作用，因此认为该方法非常值得进一步研究。

Linear、Delay和Baseline1对应记忆推理模型设计的基础方法，分别代表无记忆功能、延时堆叠和循环神经网络3种方法。从表5可以看到，使用LSTM的模型有相当的优势，而观察堆叠和纯线性方法的效果几乎相同，且由于观察堆叠后特征转换需要的线性层参数更多，导致需要更长的训练时间。训练过程中的成功率曲线如图13(b)所示。可以看出，采用LSTM的模型的收敛速度和稳定程度都是最优的，而纯线性方法和观察堆叠方法的稳定性非常差，这是由于作为记忆推理模型的线性层不稳定，且记忆机制不如LSTM有效。




(a) Baseline2、RGB-off和RGB-on训练成功率曲线　(b) Linear、Delay和Baseline1训练成功率曲线

图13　训练成功率曲线图

Fig. 13　Curves of Training Success Rate

综上得出：(1)记忆模块对泛化能力的形成至关重要，只使用线性层无法进一步提高导航性能，训练非常不稳定；(2)对观察进行延时堆叠的效果并不好，且同样不稳定；(3)LSTM能大幅提升收敛速度、稳定性与测试性能，证明在陌生环境中通过记忆构建对环境的认知是模型泛化能力的关键。

## 7　结　语

本文经过对类脑导航算法的调研，结合对生物的导航能力的总结，总结出了类脑导航算法的计算框架，分为感知模型、记忆推理模型、规划决策模型、经验生成和学习算法，将目前该领域的重要工作拆解归纳到了具体部分并进行了理论分析与实验验证，得到了如下结论：

1)感知模型应结合静态感知任务的优秀成果，通过辅助任务的方式辅助强化学习训练以获得综合的感知能力，从而更适合于导航这样的综合的非静态感知问题；同时希望在感知能力向真实世界迁移的问题上做出更多的尝试。

2)记忆推理模型应使用有记忆能力的神经网络，复杂的记忆结构或机制对导航至关重要，因此建议新的、更复杂的记忆结构与机制融入到记忆推理模型中，希望对模型所构建的环境的非结构化的内在表示有更深入的解读和研究。

3)除了简单的线性处理以外，规划决策模型尚无被广泛采用的更复杂的方法，希望未来该环节能够产生受认可的开创性工作。

4)在经验生成方面，推荐使用多线程采样的方法，应在奖励函数设计中进一步鼓励智能体的探索行为，探索新的数据采样或利用模式，缓解数据效率难题。

5)在学习算法方面，强化学习或模仿学习算法各有优势与劣势，都值得被进一步研究；希望将机器学习领域的新进展，例如元学习算法等引入到导航模型的训练中。

## 参　考　文　献

[1] Doeller C F, Barry C, Burgess N. Evidence for Grid Cells in a Human Memory Network[J]. *Nature*, 2010, 463(7 281): 657-661

[2] Kropff E, Carmichael J E, Moser M B, et al. Speed Cells in the Medial Entorhinal Cortex[J]. *Nature*, 2015, 523(7 561): 419-424

[3] Banino A, Barry C, Uria B, et al. Vector-based Navigation Using Grid-Like Representations in Artificial Agents[J]. *Nature*, 2018, 557(7 705): 429-433

[4] Smith L, Gasser M. The Development of Embodied Cognition: Six Lessons from Babies[J]. *Artificial Life*, 2014, 11(1-2):13-29

[5] Epstein R A, Patai E Z, Julian J B, et al. The Cognitive Map in Humans: Spatial Navigation and Beyond[J]. *Nature Neuroscience*, 2017, 20(11): 1 504-1 513

[6] Sutton R S, Barto A G. Reinforcement Learning: An Introduction[M]. Cambridge, MA: MIT Press, 1998

[7] Savva M, Kadian A, Maksymets O, et al. Habitat: A Platform for Embodied AI Research[C]//IEEE International Conference on Computer Vision, Seoul, Korea (South), 2019

[8] Xia F, Zamir A R, He Z, et al. Gibson Env: Real-World Perception for Embodied Agents[C]// IEEE

Computer Society Conference on Computer Vision and Pattern Recognition, Anchorage, Alaska,2018

[9] Chang A , Dai A , Funkhouser T , et al. Matterport3D: Learning from RGB-D Data in Indoor Environments[C]//International Conference on 3D Vision (3DV), Qingdao,China,2017

[10] Straub J, Whelan T, Ma L, et al. The Replica Dataset: A Digital Replica of Indoor Spaces[J/OL]. (2019-06-13)[2021-3-10]. http://arxiv.org/abs/1906.05797

[11] Zhu Y, Mottaghi R, Kolve E, et al. Target-Driven Visual Navigation in Indoor Scenes Using Deep Reinforcement Learning[C]//IEEE International Conference on Robotics and Automation, Singapore, Singapore, 2017

[12] Beattie C, Leibo J Z, Teplyashin D, et al. DeepMind Lab [J/OL]. (2016-12-13) [2021-3-10]. http://arxiv.org/abs/1612.03801

[13] Deitke M, Han W, Herrasti A, et al. RoboTHOR: An Open Simulation-to-Real Embodied AI Platform [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA,2020

[14] Savva M, Chang A X, Dosovitskiy A, et al. MINOS: Multimodal Indoor Simulator for Navigation in Complex Environments[J/OL]. (2017-12-11) [2021-3-10]. http://arxiv.org/abs/1712.03931

[15] Song S, Yu F, Zeng A, et al. Semantic Scene Completion from a Single Depth Image[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA,2017

[16] Wu Y, Wu Y, Gkioxari G,et al. Building Generalizable Agents with a Realistic and Rich 3D Environment[C]// The 6th International Conference on Learning Representations, Vancouver, Canada, 2018

[17] Mnih V, Badia A P, Mirza L, et al. Asynchronous Methods for Deep Reinforcement Learning [C]// The 33rd International Conference on Machine Learning, ICML, New York, USA, 2016

[18] Hochreiter S, Schmidhuber J. Long Short-Term Memory[J]. *Neural Computation*, 1997, 9(8):1 735-1 780

[19] Gupta S, Tolani V, Davidson J, et al. Cognitive Mapping and Planning for Visual Navigation[J]. *International Journal of Computer Vision*, 2020, 128(5): 1 311-1 330

[20] Ross S, Gordon G J, Bagnell J A. A Reduction of Imitation Learning and Structured Prediction to No-Regret Online Learning[C]//The 14th International Conference on Artificial Intelligence and Statistics, Fort Lauderdale, USA, 2011

[21] Jaderberg M, Mnih V, Czarnecki W M, et al. Reinforcement Learning with Unsupervised Auxiliary Tasks [C]//The 5th International Conference on Learning Representations, Toulon, France, 2017

[22] Kulhanek J, Derner E, De Bruin T, et al. Vision-based Navigation Using Deep Reinforcement Learning [C]//European Conference on Mobile Robots, Prague, Czech Republic, 2019

[23] Mirowski P, Grimes M K, Malinowski M, et al. Learning to Navigate in Cities Without a Map[J]. *Advances in Neural Information Processing Systems*, 2018, 2 018: 2 419-2 430

[24] Wortsman M, Ehsani K, Rastegari M, et al. Learning to Learn How to Learn: Self-Adaptive Visual Navigation Using Meta-Learning[C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 2019

[25] Finn C, Abbeel P, Levine S. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks [C]//The 34th International Conference on Machine Learning, Sydney, Australia, 2017

[26] Du H, Yu X, Zheng L. Learning Object Relation Graph and Tentative Policy for Visual Navigation [C]// The 16th European Conference on Computer Vision, Glasgow, UK, 2020

[27] Cobbe K, Klimov O, Hesse C, et al. Quantifying Generalization in Reinforcement Learning[C]//The 36th International Conference on Machine Learning, Long Beach, California, USA,2019

[28] Mirowski P, Pascanu R, Viola F, et al. Learning to Navigate in Complex Environments[C]//The 5th International Conference on Learning Representations, Toulon, France, 2017

[29] Shi H, Shi L, Xu M, et al. End-to-End Navigation Strategy with Deep Reinforcement Learning for Mobile Robots[J]. *IEEE Transactions on Industrial Informatics*, 2020, 16(4): 2 393-2 402

[30] Druon R, Yoshiyasu Y, Kanezaki A, et al. Visual Object Search by Learning Spatial Context [J]. *IEEE Robotics and Automation Letters*, 2020, 5(2): 1 279-1 286

[31] Ye X, Lin Z, Li H, et al. Active Object Perceiver: Recognition-Guided Policy Learning for Object Searching on Mobile Robots[C]//IEEE International Conference on Intelligent Robots and Systems, Madrid, Spain, 2018

[32] Lü Y, Xie N, Shi Y, et al. Improving Target-Driven Visual Navigation with Attention on 3D Spatial Relationships[J/OL]. (2020-4-29)[2021-3-10]. http://

arxiv.org/abs/2005.02153

[33] Bengio Y, Louradour J, Collobert R, et al. Curriculum Learning[C]//ACM International Conference Proceeding Series, Montreal, Quebec, Canada, 2009

[34] Yang W, Wang X, Farhadi A, et al. Visual Semantic Navigation Using Scene Priors[C]//The 7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, 2019

[35] Lu Y, Chen Y, Zhao D, et al. MGRL: Graph Neural Network Based Inference in a Markov Network with Reinforcement Learning for Visual Navigation[J]. *Neurocomputing*, 2020, 421: 140-150

[36] Mousavian A, Toshev A, Fišer M, et al. Visual Representations for Semantic Target Driven Navigation[C]//IEEE International Conference on Robotics and Automation, Montreal, QC, Canada, 2019

[37] Savinov N, Dosovitskiy A, Koltun V. Semi-parametric Topological Memory for Navigation[C]//The 6th International Conference on Learning Representations, Vancouver, BC, Canada, 2018

[38] Gordon D, Kadian A, Parikh D, et al. SplitNet: Sim2Sim and Task2Task Transfer for Embodied Visual Navigation[C]//IEEE International Conference on Computer Vision, Seoul, Korea (South), 2019

[39] Wu Y, Wu Y, Tamar A, et al. Bayesian Relational Memory for Semantic Visual Navigation[C]//IEEE International Conference on Computer Vision, Seoul, Korea (South), 2019

[40] Kahn G, Villaflor A, Ding B, et al. Self-Supervised Deep Reinforcement Learning with Generalized Computation Graphs for Robot Navigation[C]//2018 IEEE International Conference on Robotics and Automation, Brisbane, Australia, 2018

[41] Zhu F, Zhu L, Yang Y. Sim-real Joint Reinforcement Transfer for 3D Indoor Navigation[C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 2019

[42] Wu Q, Manocha D, Wang J, et al. NeoNav: Improving the Generalization of Visual Navigation via Generating Next Expected Observations[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(6):10 001-10 008

[43] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016

[44] Shen W, Xu D, Zhu Y, et al. Situational Fusion of Visual Representation for Visual Navigation[C]//IEEE International Conference on Computer Vision, Seoul, Korea (South), 2019

[45] Oh J, Chockalingam V, Singh S, et al. Control of Memory, Active Perception, and Action in Minecraft[C]//The 33rd International Conference on Machine Learning, New York , USA, 2016

[46] Pritzel A, Uria B, Srinivasan S, et al. Neural Episodic Control[C]//The 34th International Conference on Machine Learning, Sydney, NSW, Australia, 2017

[47] Graves A, Wayne G, Reynolds M, et al. Hybrid Computing Using a Neural Network with Dynamic External Memory[J]. *Nature*, 2016, 538(7 626): 471-476

[48] Kipf T N, WellinG M. Semi-Supervised Classification with Graph Convolutional Networks[C]//The 5th International Conference on Learning Representations, Toulon, France, 2017

[49] Tamar A, Wu Y, Thomas G, et al. Value Iteration Networks[C]// Advances in Neural Information Processing Systems, Barcelona, Spain, 2016

[50] Watkins-Valls D, Xu J, Waytowich N, et al. Learning Your Way Without Map or Compass: Panoramic Target Driven Visual Navigation[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems, Las Vegas, NV, USA, 2020

[51] Anderson P, Chang A, Chaplot D S, et al. On Evaluation of Embodied Navigation Agents[J/OL]. (2018-7-18)[2021-3-10]. http://arxiv.org/abs/1807.06757

[52] Pennington J, Socher R, Manning C D. GloVe: Global Vectors for Word Representation[C]// 2014 Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, 2014

# Review and Verification for Brain-Like Navigation Algorithm

*GUO Chi*[1,2]　*LUO Binhan*[1]　*LI Fei*[2]　*CHEN Long*[3]　*LIU Jingnan*[1]

1　GNSS Research Center, Wuhan University, Wuhan 430079, China

2　Artificial Intelligence Institute of Wuhan University, Wuhan 430079, China

3　School of Data and Computer Science, Sun Yat-sen University, Guangzhou 510275, China

**Abstract**: **Objectives:** In recent years, the brain-like navigation algorithm is a new research hotspot, which is expected to achieve autonomous navigation by imitating the ability of biological navigation. The core issue is how to improve generalization ability. **Methods:** This paper introduces the research background and theoretical basis of the brain-like navigation algorithm. After investigation, we propose a computational framework of brain-like navigation algorithm. The outstanding works in this field are discussed and analyzed under this framework, and we carried out experimental verification of some basic methods. **Results:** The main contributions of this paper are: (1) Comprehensively introduces and summarizes the theoretical basis and outstanding works in this field. (2) Proposes the computational framework of the brain-like navigation algorithm, which scientifically defines the functions of different modules of the algorithm. (3) Through theoretical analysis and experimental verification, we summarized valuable conclusions and expectations. **Conclusions:** In terms of model design, mature methods of deep learning can also be applied to this problem, but need more modifications to further improve navigation capabilities; in terms of model training, combining the advantages of multiple learning algorithms is hopeful to further improve the generalization ability.

**Key words**: brain-like navigation; artificial intelligence; autonomous navigation; perception; memory; policy

**First author:** GUO Chi, PhD, professor, specializes in BeiDou application, unmanned system navigation, and location-based services (LBS). E-mail:guochi@whu.edu.cn

**Foundation support:** The National Key Research and Development Program of China(2016YFB0501801).

**引文格式**:GUO Chi, LUO Binhan, LI Fei, et al. Review and Verification for Brain-Like Navigation Algorithm[J]. Geomatics and Information Science of Wuhan University, 2021, 46(12):1819-1831. DOI:10.13203/j.whugis20210469(郭迟,罗宾汉,李飞,等.类脑导航算法:综述与验证[J].武汉大学学报·信息科学版,2021,46(12):1819-1831.DOI:10.13203/j.whugis20210469)