

DOI: 10.13382/j.jemi.B2205862

结合拆分注意力机制和下一次预期观察的视觉导航*

刘紫燕^{1,2} 杨模¹ 袁浩¹ 梁静¹ 梁水波¹ 孙昊堃¹

(1. 贵州大学大数据与信息工程学院 贵阳 550025; 2. 贵州大学大数据国家重点实验室 贵阳 550025)

摘要:针对深度强化学习视觉导航算法因导航场景变化而导致导航精度下降,影像匹配的实时性和可靠性降低的问题,提出一种融合拆分注意力机制和下一次预期观测(NEO)的视觉导航模型。首先使用 ResNest50 骨干网络提取当前状态和目标状态的特征以降低网络冗余,利用跨阶段部分连接 CSP 强化捕获浅层目标特征信息以增强模型的学习能力。然后提出改进的损失函数,使得推理网络更加接近于真实后验,从而智能体能在当前环境下做出最佳决策,进一步提升不同场景下模型的导航精度。在 AVD 数据集和 AI2-THOR 场景进行训练及测试,实验结果表明,本文算法导航精度高达 96.8%,平均 SR 提升约 3%,平均 SPL 提升约 6%,可以满足导航精度和实时匹配的要求。

关键词: 视觉导航;深度强化学习;拆分注意力机制;下一次预期观测

中图分类号: TP242;TP391.41 **文献标识码:** A **国家标准学科分类代码:** 510.40;510.80

Visual navigation combining split attention mechanism and next expected observation

Liu Ziyang^{1,2} Yang Mo¹ Yuan Hao¹ Liang Jing¹ Liang Shuibob¹ Sun Haokun¹

(1. College of Big Data and Information Engineering, Guizhou University, Guiyang 550025, China;

2. State Key Laboratory of Public Big Data, Guizhou University, Guiyang 550025, China)

Abstract: A visual navigation model incorporating split attention mechanism and next expected observation (NEO) is proposed to address the problem that deep reinforcement learning visual navigation algorithm degrades navigation accuracy, real-time and reliability of image matching due to navigation scene changes. The features of current and target states are first extracted using the ResNest50 backbone network to reduce network redundancy. The shallow target feature information is captured intensively using a cross-stage-partial-connections CSP to enhance the learning ability of the model. Then an improved loss function is proposed to make the inference network closer to the true posterior so that the agent can make the best decision in the current environment and further improve the navigation accuracy of the model in different scenarios. The training and testing are conducted on AVD dataset and AI2-THOR scenes, and the experimental results show that the navigation accuracy of the algorithm in this paper is as high as 96.8%, with an average SR improvement of about 3% and an average SPL improvement of about 6%, which meets the requirements of navigation accuracy and real-time matching.

Keywords: visual navigation; deep reinforcement learning; split attention mechanism; next expected observations (NEO)

0 引言

随着智能体技术的快速发展,智能体在各行各业中的使用愈发普遍。其中,智能体的自主导航能力十分关

键,根据不同的目标位置生成任意的无碰撞路径^[1]。传统智能体导航主要为基于地图的策略,包括同时定位和建图^[2](simultaneous localization and mapping, SLAM)及路径规划^[3],通过 SLAM 方法建模环境地图,再利用路径规划方法生成起点到终点的无碰撞路径。其中,若

收稿日期: 2022-09-27 Received Date: 2022-09-27

* 基金项目:贵州省科学技术基金(黔科合基础[2016]1054)、贵州省联合资金(黔科合 LH 字[2017]7226 号)、贵州大学 2017 年度学术新苗培养及创新探索专项(黔科合平台人才[2017]5788)、贵州省科技计划项目(黔科合 SY 字[2011]3111)资助

SLAM 对环境认知存在偏差,在路径规划阶段极可能出现错误甚至失败的情况。因此,在无地图情况下实现智能体的自主导航尤为重要。

在未知、不确定的环境中智能体实现各类复杂任务,视觉导航是重要前提。传统路径规划算法无法有效实现无地图完成独立导航,特别是基于激光和视觉传感器构建地图的情况^[4],由于激光传感器造价高,在雨雪等恶劣条件下效果较差;视觉传感器造价低,但在光线条件差时出现纹理特征缺失。而随着深度强化学习的快速发展——深度 Q 学习(deep Q network, DQN)的推出^[5],使用原始高维环境输入,为无地图导航实现奠定基础。随后,基于值函数与策略函数的有关深度强化学习算法被不断提出。基于值函数的方法例如 Double DQN、Dueling DQN 等算法,基于策略函数的方法例如异步优势评论者算法(asynchronous advantage actor, A3C)^[6],下一步预期观察(next expected observations, NEO)^[7]等算法。

基于深度强化学习算法的智能体在无地图情况下的视觉导航已经成为目前的研究热点^[8]。最早提出基于深度强化学习的异步优势评论家算法的目标驱动视觉导航模型,只需将目标图像及智能体当前预测的第一视角图像输入到视觉导航框架中,即可输出智能体行动直达目标位置,实现端到端的无地图视觉导航。文献[9]使用图卷积网络对先验知识进行编码,通过先验信息的引入,提高了对未知场景和未经过训练的目标进行导航的泛化能力。文献[10]利用注意力机制和 3D 空间关系编码,在深度强化学习算法 A3C 的基础上进行了改进,提高了机器人在未知环境中的导航能力;文献[11]提出了一种上下文网格(context grid)来表示当前看到的场景和目标(target)之间的相似度,该方法具有更好的泛化性,更适合宽阔的场景;文献[12]为每个 target 提供一个祖元素(parent)集合,即定义了一个状态空间,强化学习(reinforcement learning, RL)模型可以从中学习去探索目标间的关系,平衡了先验知识与当前观测,保证了充分的训练与泛化。文献[13]提出通过学习一个智能体(agent)来改进视觉导航的跨目标和跨场景泛化,该 agent 通过设想它期望看到的下一次观测来引导,被称为下一次预测观察(next expected observations, NEO)模型,该模型相较于之前的模型有在导航性能上有较大的提升^[14]。

本文在文献[14]的框架上进行改进,出了一种结合拆分注意力机制和下一次预期观察的视觉导航网络模型,该模型以智能体第一视角获得的 RGB-D 图像及其在智能体坐标系下目标点的极性坐标为输入,结合算法改进网络结构和奖励函数来训练目标,以智能体的连续动作值作为输出,实现无地图的端到端视觉导航任务,使智能体在获取观察到的图像时能够更多的关注积聚信息,提高主要特征的利用率,从而提升智能体在不同的场景

下的导航精度。

1 注意力机制

注意力机制在人类感知中起着至关重要的作用。注意力机制允许人们选择性地关注关键信息,而忽略其他无关信息。卷积神经网络(convolutional neural network, CNN)中的注意力模块可以加速网络的学习过程,对目标任务提取更关键和有区别的特征,增强网络模型的鲁棒性,更适应小的训练数据集。谷歌 DeepMind 团队在执行图像分类任务时首次提出了注意力机制,掀起了研究注意力机制的热潮^[15]。Kaul 等^[16]提出了一种将注意力纳入 FCN 的方法 FocusNet,该方法从一个单独的卷积自动编码器生成的特征图中执行医学图像分割。Hu 等^[17]提出 SENet 通过显式建模通道之间的相互依赖关系,自适应地重新校准通道特征响应。Woo 等^[18]进一步扩展了 SENet 中的挤压-激励模块。本文使用的卷积注意力模块(convolutional block attention module, CBAM)是一种轻量级通用模块,几乎不使用计算资源,并且能够基于给定的中间值执行自适应特征细化特征映射,图 1 为其主要结构,按照不同维度(如通道、空间、时间、类别等)定义,主要分为通道注意力(channel attention, CA)和空间注意力(spatial attention, SA)模块。通道注意力旨在显示的建模出不同通道之间的相关性,学习自动获取每个特征通道的重要程度并赋予不同的权重系数,从而强化重要的特征抑制非重要特征。空间注意力旨在提升关键区域的特征表达,用过空间转换保留关键信息并为每个位置生成权重掩码加权输出,从而增强感兴趣的特定目标区域同时弱化不相关的背景区域。

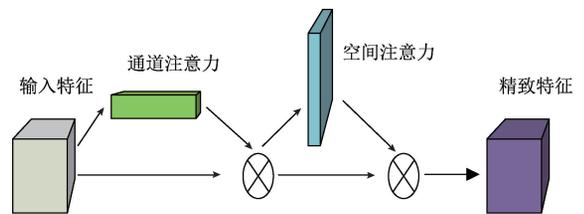


图 1 CBAM 模块结构

Fig. 1 CBAM module structure

1.1 空间注意力模块

空间注意模块(SAM)作为注意力模块的一部分被引入卷积神经网络,并在分类和检测任务中表现出良好的性能。空间注意力关注的是图像之间的位置信息,它描述了输入特征之间的空间关系。形式上,输入特征 $F \in R^{H \times W \times C}$,其中 H 、 W 、 C 分别表示入口空间注意路径图像的高度、宽度和通道数。归一化的 1×1 卷积和 1×1 卷积构成空间注意模块。第 1 次卷积后信道数减少 $1/2$,后

续卷积将信道数减少到 1。\$F'\$ 是通过 Sigmoid 函数将单个通道中的像素值映射到 \$[0, 1]\$ 范围得到的。为了对来自通道注意路径输出的维度 \$F_c\$ 和来自空间注意路径输出的维度 \$F_s\$ 进行元素相乘, \$F'\$ 在通道上堆叠 \$C\$ 次得到 \$F_s\$, 空间注意模块的输出特征计算为:

$$F^s = \sigma(f^{1 \times 1}(\text{ReLU}(\text{BN}(f^{1 \times 1}(F)))))) \quad (1)$$

其中, \$f^{1 \times 1}(\cdot)\$ 表示滤波器大小为 \$1 \times 1\$ 卷积运算, \$\text{BN}(\cdot)\$ 表示批处理归一化, \$\sigma(\cdot)\$ 表示激活函数, 空间注意力模块原理如图 2 所示。

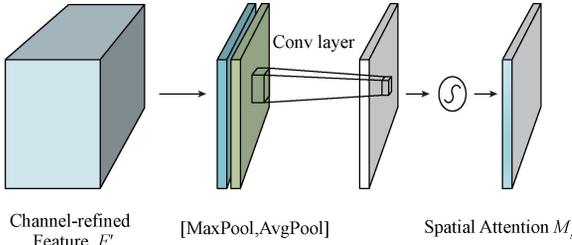


图 2 空间注意力模块原理

Fig. 2 principle of the spatial attention module

1.2 通道注意力模块

通道注意力获取图像上不同通道之间的相关性, 为了有效计算通道注意力, 需要对输入特征图的空间维度进行压缩, 对于空间信息的聚合, 常用的方法是平均池化。该模块通过使用全局平均池(global average pooling, GAP)生成通道描述符, 并在整个通道上下文中将特征映射输入聚合到 CAM。如果使用 \$\gamma_d^{mp} = [y_1^{mp}, y_2^{mp}, \dots, y_F^{mp}]\$, 其中 \$y_f^{mp} \in \mathbb{R}^{W \times H}\$, 作为通道注意模块的输入数据。全局平均池化的方法如下:

$$s_f = F_{sq}(y_f^{mp}) = \frac{1}{H \times W} \sum_m \sum_n y_f^{mp}(m, n) \quad (2)$$

其中, \$F_{sq}\$ 为空间压缩(GAP)函数, \$y_f^{mp}\$ 为 \$f^{th}\$ 信道的空间位置, \$H \times W\$ 分别对应该通道的高度尺寸和宽度尺寸。通过 GAP 压缩, 每个二维特征映射生成 \$s_f\$。通道注意力模块的第 2 步是激励, 它根据第 1 步中嵌入的全局信息捕获通道之间的依赖关系。通过这个函数可以学习通道之间的非排他性关系和非线性相互作用。SE block 包含两个全连接层(fully connected layers, FC), 其中输入向量依次编码为 \$1 \times 1 \times \frac{F}{r}\$ 和 \$1 \times 1 \times F\$ 的形状, 最后一个 SE 块的输出表示为:

$$C_s = F_{se}(z; S) = \sigma(S_2 \delta(S_1 z)) \quad (3)$$

其中, \$S_1\$ 和 \$S_2\$ 分别为第 1FC 层和第 2FC 层的参数。\$\delta\$ 为蒸馏线性单元(ReLU), \$\sigma\$ 为 S 型函数, \$r\$ 为还原比, 通道注意力模块原理如图 3 所示。

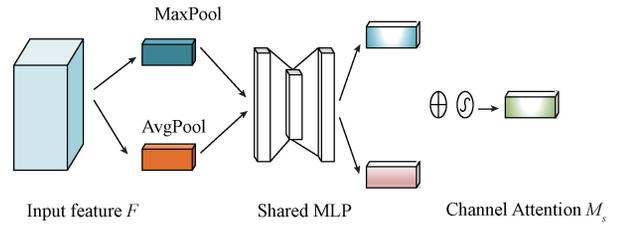


图 3 通道注意力原理

Fig. 3 Principle of channel attention

2 方法

智能体的目标是通过最少的步骤导航到目标位置。经过训练, 智能体能够在新的场景下导航, 从而验证模型的泛化能力^[19-21]。通过深度强化学习来学习目标驱动策略函数, 并在时间 \$t\$ 的动作 \$a\$ 可以如式(4)表示:

$$a \sim \pi(S_t, g | u) \quad (4)$$

目标驱动视觉导航智能体的目的是学习一个随机策略 \$\pi\$, 其中 \$S_t\$ 是当前状态的表示, \$g\$ 是目标状态的表示。策略 \$\pi\$ 的输出是动作的概率分布。

2.1 变分贝叶斯导航

在得到当前观测到的图像 \$x\$ 后, 文献[14]提出并不需要取直接预测下一时刻的最佳动作 \$a\$, 而是在假设在下一时刻的最佳动作是已知的, 并且已经执行, 就可以选择用生成模型来生成下一时刻的预期观测值, 如式(5)所示:

$$p_\theta(\hat{x}, z | x, a) = p_\theta(\hat{x} | z) p_\theta(z | x, a) \quad (5)$$

其中, \$p_\theta(\hat{x}, z | x, a)\$ 是下时刻预期观测 \$\hat{x}\$ 和一个潜在变量 \$z\$ 联合分布的参数模型。可以通过所选择的下一个动作 \$a\$ 来获取下一个观测 \$x\$。

为了学习生成式模型, 通常会最大化边际对数似然 \$\log p_\theta(\hat{x} | x, a)\$。但是, 当模型由神经网络参数化时, 由于边际似然的难解性, 优化可能会很困难。此外, 一个最好的行动是未知先验和本质上是由目标 \$g\$。为此, 用变分推理和引入一个推理网络 \$\lambda(z | x, g)\$ 来近似真实的后验概率 \$p(\theta | x, z)\$, 特别是要优化下界的边缘相似性:

$$\log p_\theta(\hat{x} | x, a) \geq E_{z \sim q_\lambda(z | x, g)} [\log \frac{p_\theta(\hat{x}, z | x, a)}{q_\lambda(z | x, g)}] = L(\hat{x}) \quad (6)$$

这个下界构成了其目标函数:

$$J = -E_{z \sim q_\lambda(z | x, g)} [\log p_\theta(\hat{x} | z)] + \text{KL}[q_\lambda(z | x, g) || p_\theta(z | x, a)] = -L(\hat{x}) \quad (7)$$

其中, KL 表示 Kullback-Leibler 散度。

在训练过程中, \$p_\theta(z | x, a)\$ 可以被估计为一个高斯分布, 条件是当前观测 \$x\$ 和地面真实动作 \$a\$, 导致混合后

验先验强加于潜在分布。为实现智能体导航,本文学习了一个导航动作分类器 $q_\varphi(a|x,\hat{x},\hat{a})$,它根据当前的观测 x 、生成的 NEO \hat{x} 以及之前的动作属于 \tilde{a} ,预测下一个最佳动作 a 。结合动作预测,目标函数为:

$$J = -\alpha E_{z \sim q_\lambda(z|x,g)} [\log p_\theta(\hat{x}|z)] + \beta KL[q_\lambda(z|x,g) || p_\theta(z|x,a)] + \gamma E_{a \sim p(a)} [-\log q_\varphi(a|x,\hat{x},\hat{a})] \quad (8)$$

其中, $a \sim \text{cat}(1/C)$ 。式(8)中的目标函数由重构损失、KL 散度损失和交叉熵损失组成。在整个实验过程中,3 个超参数分别设为 $\alpha = 0.01, \beta = 0.0001$ 和 $\gamma = 1$ 。图 4 所示为导航模型的概率转移模型。

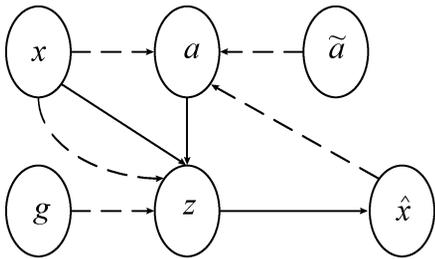


图 4 导航概率转移模型

Fig. 4 Navigating the probability transfer model

2.2 基于下一次预期观察视觉导航

变分贝叶斯模型目标所对应网络体系结构包括 3 个子网络,如图 5 所示。变分推理模块需要完整的观察视图当前机器人位置以及目标视图作为输入^[22],使用 ResNet-50 提取 2048-d 中的特征向量。输入图像分辨率为 64×64 。然后使用这些输出的 2048-D 特征向量,用 MLP 推断出一个维度为 400 的潜在变量向量。这里,KL 散度损失被最小化,以使潜在变量的分布与当前观测和地面真实行动的先验估计相匹配。然后,NEO 生成模块使用两层 MLP 和 5 层卷积网络,从潜在矢量中生成正视图中的 NEO。这项任务是在实地观察的监督下训练的。动作预测模块使用 4 层 MLP,将 NEO 生成模块(2048-D)的最后一层特征、当前观测的特征(2048-D)和之前动作(7-D 单热向量)提取的特征(1024-D)的串联映射到预测的下一动作(7-D)。子网络用地面真实行动来训练。

2.3 结合注意力机制和下一次预期观察的视觉导航

1) 融合注意力的强化学习网络

目标驱动的视觉导航最大问题是泛化性问题,即对于已经训练好的导航模型是否能适应新的场景,在新的场景下导航的保持与原来相近性能。

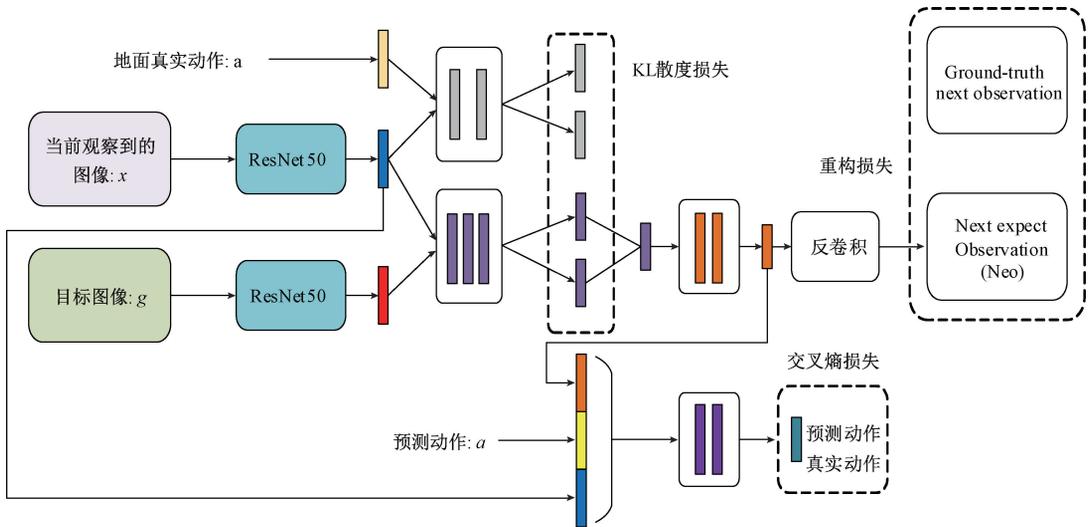


图 5 基于 Neo 模型的视觉导航网络框架

Fig. 5 A framework for visual navigation network based on Neo model

本章在基于下一次预期观察进行视觉导航的基础上融合拆分注意力,并改进了连接方式及损失函数,使得智能体能够把握输入图像的主要信息,结合 NEO 模型让智能体更加了解导航目标之间的上下文信息,更能适应新的场景,融合 CSP-ResNeSt50 网络改进视觉导航模型如图 6 所示。

该网络主要由 3 部分组成,图像输入:当前观察到的

图像 x ;共享孪生网络:特征提取过程;策略生成网络:融合 SLAM 策略生成网络。图像输入的是智能体当前观察到的及目标所在地的 RGB 图像,经由共享孪生网络提取特征并投影到特征空间判断两幅图像的空间位置关系后输入有 LSTM 的策略生成网络,因为使用的是 A3C 算法,最终输出的是四维的策略分布和单一的值函数。

2) 基于 ResNeSt50 的特征提取网络

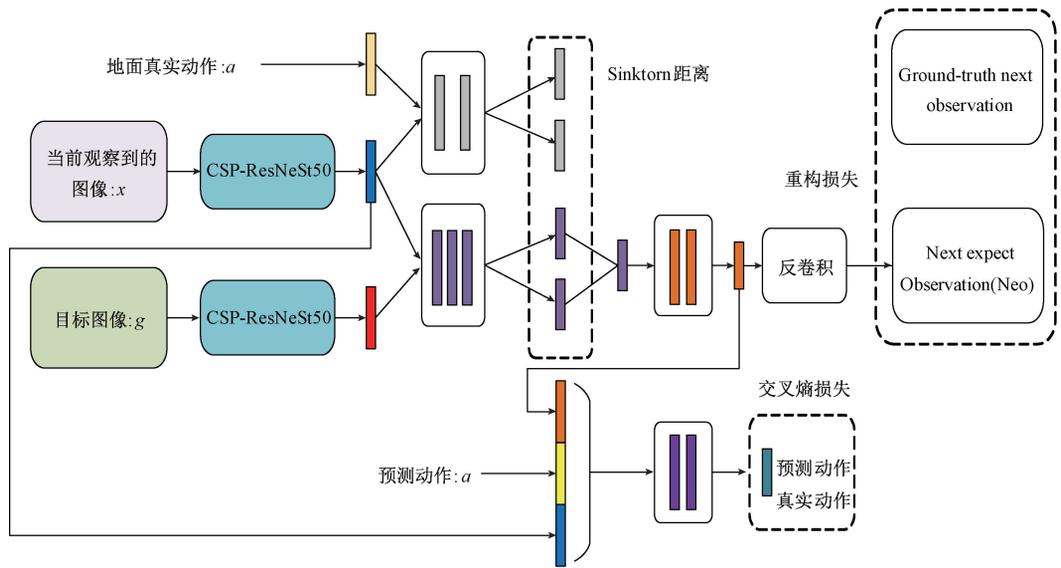


图 6 融合 CSP-ResNeSt50 的改进视觉导航模型

Fig. 6 Improved visual navigation model incorporating CSP-ResNeSt50

图像分类任务的研究,由于其简单且模块化的结构,大部分目标检测等任务仍将 ResNet 或者其变体作为骨干网络(Backbone),但由于 ResNet 模型最初是为图像分类而设计的,可能不适合各种下游应用,它的接收野大小有限,并且缺乏跨通道交互^[20]。故本文使用 ResNet50 来进行特征提取,其主要框架如图 7 所示。

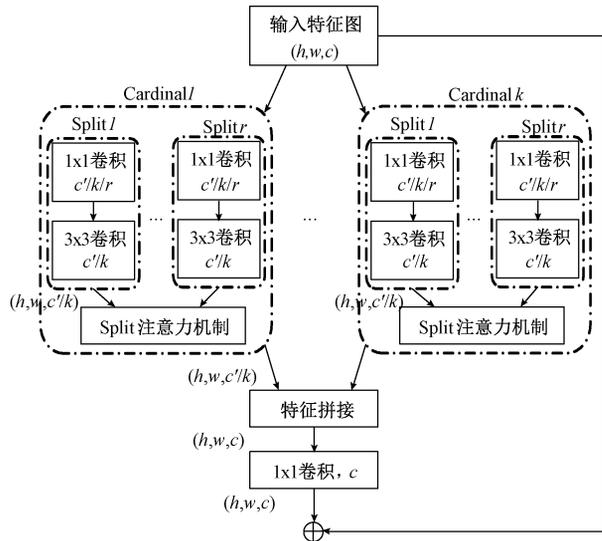


图 7 ResNeSt block 的网络结构

Fig. 7 Network structure of the ResNeSt block

如图 7 所示, Split-Attention block 计算单元由两部分组成:特征图组和拆分注意力。特征图组的数量由超参数 K 决定,基数组里面的拆分数由参数 R 决定,特征组的总数计算如式(9)所示:

$$G = KR \tag{9}$$

输入特征图经过两次分组处理,第 1 次分为 2 个基数组,第 2 次将每个主分组中又拆分为 2 个部分,拆分的每个部分经过 l 和 k 的卷积之后汇入拆分注意力模块,再将这 2 个基数组输出的特征做特征拼接操作,在整个过程中,输入和输出的尺寸一致。

拆分注意力模块(split-attention block)借鉴 SENet 思想,即基于通道注意力的机制,其主要思想是对不同的通道赋予权重以描述通道的重要程度,其结构如图 8 所示。

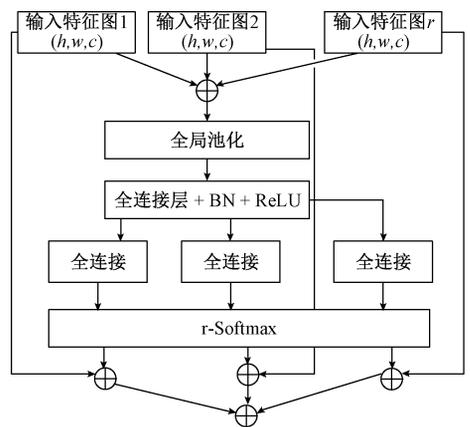


图 8 拆分注意力通道结构

Fig. 8 Split attention channel structure

首先,代表第 k 个基数组的输入 U_k 如式(10)所示:

$$U_k = \sum_{i=R(k-1)}^{Rk} U_i \tag{10}$$

其中, $U_k \in \mathbb{R}^{H \times W \times C/K}$, $k \in 1, 2, \dots, K$ 。通过跨空间维度的全局平均池,可以收集具有嵌入式通道统计信息的

全局上下文信息。输入特征图经过全局平均池化后,可以计算出通道权重统计量,第 c 个分量的计算方式如式 (11) 所示:

$$s_c^k = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W U_c^k(i, j) \quad (11)$$

其中, $c = C/K$ 。经过 Softmax 激活函数之后的权重 $a_i^k(c)$ 为:

$$a_i^k(c) = \begin{cases} \frac{\exp(\delta_i^c(s^k))}{\sum_{j=0}^R \exp(\delta_j^c(s^k))}, R > 1 \\ \frac{1}{1 + \exp(-\delta_i^c(s^k))}, R = 1 \end{cases} \quad (12)$$

加入权重后的输出为 $V^k (V^k \in \mathbb{R}^{H \times W \times C/K})$, 则第 c 个 Cardinal 的输出为:

$$V_c^k = \sum_{i=1}^R a_i^k(c) U_{R(k-1)+r} \quad (13)$$

接着将每个 Cardinal 的输出拼接起来,得到最后的输出 V 为:

$$V = \text{Concat} \{V^1, V^2, \dots, V^K\} \quad (14)$$

最后将 ResNeSt block 按 ResNet50 的形式类似堆叠,就可以得到最终的 ResNeSt50。

3) 跨连接 CSP-ResNeSt50

ResNeSt50 特征提取网络相较于 ResNet50 获得更好性能,在增加一些参数但不增加计算量的情况下实现了更好的效果,但从特定的智能体视觉导航任务并考虑模型的复杂程度问题,需进一步减少计算量,使得模型在不降低或者少降低精度的情况下达到轻量化。跨阶段部分网络 CSPNet 的思想优化 ResNeSt50,提出融合 CSPNet 的 CSP-ResNeSt50 特征提取网络,其结构如图 9 所示。

如图 9 所示,输入特征图在进行通道分割之后分为两个部分,第 1 部分通过多个阶段的 ResNeSt 模块,第 2 部分只有 1/2 的通道数通过 ResNeSt 模块,经过卷积和滤波操作之后与第 1 部分进行特征拼接,总的通道数为 $3c/2$,这样做的目的是截断梯度流,防止过多的重复梯度,减少计算量,更利于视觉导航模型的迁移和部署。

4) 损失函数

为了测量推理网络 $q_\lambda(z|x, g)$ 和真实后验概率 $p_\theta(z|x, a)$ 的概率分布差异,Neo 模型使用 Kullback-Leibler 散度或简称 KL 散度的度量,这个概念起源于概率论和信息论,KL 散度与相对熵、信息散度和判别信息密切相关,是两个概率分布之间差异的非对称度量,其表达式如式 (15) 所示:

$$D_{KL}(p_\theta(z|x, a) || q_\lambda(z|x, g)) = \int_{-\infty}^{\infty} p_\theta(z|x, a) \ln \frac{p_\theta(z|x, a)}{q_\lambda(z|x, g)} d(x) \quad (15)$$

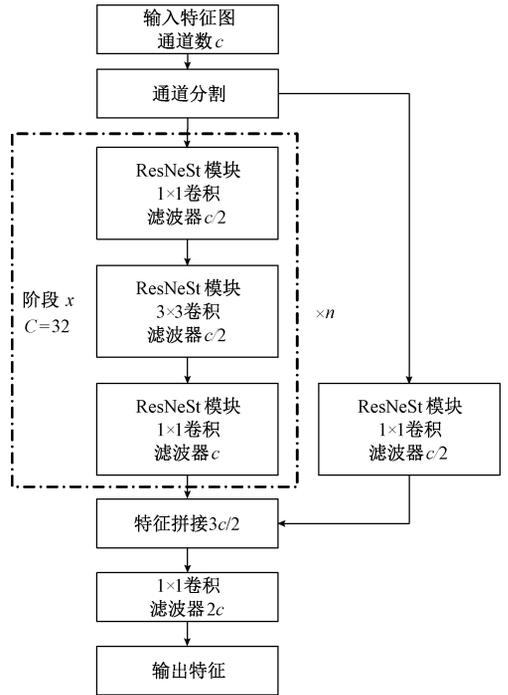


图 9 跨连接的 CSP-ResNeSt50

Fig. 9 Cross-connected CSP-ResNeSt50

其中, $p_\theta(z|x, a)$ 表示未知分布, $q_\lambda(z|x, g)$ 表示近似分布。

尽管散度最小化方法在经验上取得了成功,但由于对不同超参数的敏感性以及根据评估的分布进行训练的难度,仍然难以在高维空间中评价。而且 KL 散度是不对称的,即 $p_\theta(z|x, a)$ 和 $q_\lambda(z|x, g)$ 的相似程度在不同参考的标准下不同, $D_{KL}(p_\theta(z|x, a) || q_\lambda(x, g)) \neq D_{KL}(q_\lambda(x, g) || p_\theta(z|x, a))$,对某些场景下的视觉导航任务来说,预测观察值和真实观察值的最好能够不根据参考度量变化而改变,为了解决这些问题,借鉴最优传输理论中的思想,模型采取 Sinkhorn 距离代替 KL 散度来计算损失。

最优传输理论通过考量基础度量空间,提供了比较退化分布的有效方法。将 $P_k(\Gamma)$ 看作是有限 k 次矩的 Polish 度量空间 (Γ, d) 上的 Borel 概率测度集。给定两个概率测度 $p, q \in (\Gamma, d)$, k -Wasserstein 度量定义如式 (16) 所示:

$$w_k(p, q) = \left(\inf_{\xi \in \Omega(p, q)} \int_{\Gamma} c(x, y)^k d\xi(x, y) \right)^{\frac{1}{k}} \quad (16)$$

其中, $\Omega(p, q)$ 表示联合概率分布集,其边缘分别为 $p, q, c(x, y)$ 代表 x 到 y 的转移成本, x, y 分别是 p, q 的抽样。

但对于高斯分布,在实践中通常难以求解,于是将其正则化就可以得到 Sinkhorn 距离表达式如式 (17) 所示:

$$w_s^\beta(p, q)_c = \inf_{\xi_\beta \in \Omega_\beta(p, q)} E_{x, y \in \xi_\beta} [c(x, y)] \quad (17)$$

其中, $\Omega_\beta(p, q)$ 表示在 $\Omega(p, q)$ 中所有熵大于 $H(p) + H(q) - \beta$ 的联合分布集合。

于是就可以更新整个网络中的目标函数如式 (18) 所示, 并将其系数 β 的值调整为 0.001。

$$J = -\alpha E_{z \sim q_\lambda(z|x, g)} [\log p_\theta(\hat{x}|z)] + \beta w_s^\beta(p_\theta(z|x, a), q_\lambda(z|x, g)) + \gamma E_{a \sim p(a)} [-\log q_\varphi(a|x, \hat{x}, \tilde{a})] \quad (18)$$

5) 奖励设置

本文设置奖励函数如式 (19) 所示, 如果智能体到达指定目标, 则给予 10 的正奖励; 如果智能体发生碰撞, 则给予 0.1 的负奖励; 如果在行进状态, 则给予 0.01 的负奖励, 以刺激智能体不断进行探索, 而不是为了避免碰撞停在原地。

$$R = \begin{cases} 10, & \text{agent reach the target} \\ -0.1, & \text{agent collides} \\ -0.01, & \text{其他} \end{cases} \quad (19)$$

3 实验

实验环境为 Ubuntu16.04、GPU 为 NVIDIA2080Ti、深度学习框架为 TensorFlow, 采用的深度强化学习算法为 A3C。

3.1 数据集介绍

训练和测试所用的数据集为主动视觉数据集 (active vision dataset, AVD) 中的真实场景和 AI2-THOR 仿真环境, 该两个数据集之间的差异、特点与获取环境等对比, 如表 1 所示。

表 1 数据集对比

Table 1 Comparison of dataset

AVD 数据集				
特点	场景/个	频繁实例	2D 边界框	RGBD 图
真实场景	15	30+	70 000+	30 000+
AI2-THOR 数据集				
特点	场景/个	频繁实例	获取环境	操作属性
仿真场景	120	115	Ubuntu	openobject/ closeobject

1) 主动视觉数据集

AVD 是由真实场景下密集的 RGB-D 图像集合而成, 这样, 通过简单地取样就可以捕获视图来模拟机器人在环境中的移动。数据集包括: 厨房、客厅、餐厅、办公室、浴室等。一个典型的房间的数据将包括数以千计的 RGB-D 图像, 以及用于几十个实例对象的数以千计的对象预选框。由于视图受采集装置位置或角度的影响, 因此 API 可以允许智能体在指定场景中的相对运动, 并正

确选择下一个视图, AVD 中的实例如图 10 所示。



图 10 厨房场景和调味辣椒酱

Fig. 10 Kitchen scenes and seasoned chili sauce

2) AI2-THOR 仿真场景

AI2-THOR 是由艾伦人工智能研究所 (AI2)、斯坦福大学、卡耐基梅隆大学、华盛顿大学、南加州大学合作完成的。它为智能体提供了一个高度仿真的世界, 用于降低强化学习在真实环境下训练的试错成本, 智能体可以和该场景中的各种家具家电交互, 例如打开冰箱、推倒椅子、把电脑放在桌子上等。

它由 120 张逼真的平面图组成, 可以将其分为 4 种不同的房间布局: 厨房、客厅、卧室和浴室, 因此每种布局有 30 张平面图。将每个场景类型的前 20 个房间作为训练集, 其余 10 个房间作为评估, 图 11 所示为其中一个场景下的 4 个视角。

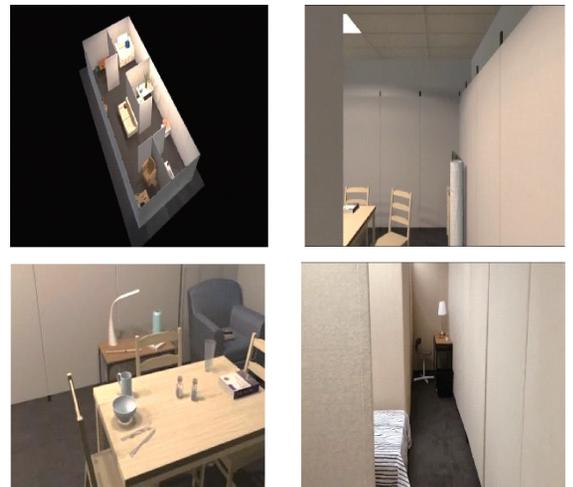


图 11 AI-THOR 仿真环境

Fig. 11 AI-THOR simulation environment

3.2 对比的模型和指标

Random Walk: 在每个时间步骤中, 智能体随机选择一个动作。

TD-A3C: 由 Zhu 等^[3]提出的最早的目标驱动视觉导航模型, 采用连续 4 帧作为输入, 这是目标驱动的视觉导航领域最先提出的模型。

Gated-LSTM-A3C: 文献 [23] 提出的基于 LSTM 的

A3C 模型的变体,导航目标为指定的一张图像。

HAUSR:该模型提出了一种通用异步后继特征表示方法,与 A3C 算法结合使得模型有更好的泛化能力。

NeoNav:文献[7]提出的先进模型,也是改进的基础模型。

本文评价的指标选择导航成功率(success rate, SR)和路径长度加权的成功率(success weighted by path length, SPL),SR 为成功导航的概率,SPL 的计算的方法如式(20)所示:

$$SPL = \frac{1}{N} \sum_{i=1}^N S_i \frac{L_i}{\max\{P_i, L_i\}} \quad (20)$$

其中, N 为导航任务的个数, S_i 为第 i 个任务成功与否的二元指标, P_i 和 L_i 分别表示第 i 个任务的实际路径长度和最短路径距离。

3.3 实验结果与分析

首先在 AI-THOR 四个场景:bathroom_02, bedroom_04, kitchen_02, livingroom_08 的一部分场景进行训练,并取剩下的一部分场景进行测试,测试结果如表 2 所示。

表 2 AI2THOR 环境下各个模型的 SR 和 SPL 对比

Table 2 Comparison of SR and SPL for each model in AI2THOR environment

Method	Kitchen	Living	Bed	Bath	Avg.
Random walk	7.0/3.5	1.8/1.0	2.6/1.5	17.9/8.0	7.3/3.5
TD-A3C	11.4/1.6	5.6/0.4	5.3/0.7	24.3/2.3	11.7/1.3
Gated-LSTM-A3C	13.1/3.2	4.9/1.1	5.1/1.2	19.3/7.9	10.6/3.4
NeoNav	19.8/10.6	11.5/5.3	13.6/5.9	21.9/9.6	16.7/7.9
Ours	20.7/11.1	11.2/5.5	14.8/7.0	22.6/10.1	17.325/8.475

由表 2 数据可以看到,改进模型的 SR 在 Living 场景下略低于 NeoNav,在其他两个场景下有明显的提升,并且在 4 个场景下的平均值相较于 NeoNav 有部分提升,提升约 3%;对比 SPL,改进模型在 4 个场景下的表现均好

于 NeoNav,提升约 6%。除此之外,本文在主动视觉数据集 AVD 上进行了训练及测试,测试结果如表 2、3 所示,其中表 2 各个模型的 SR 和 SPL 比较,表 3 为 Neo 模型和 Neo-attention 模型的总体对比。

表 3 AVD 数据集中各个模型的 SR 和 SPL 对比

Table 3 Comparison of SR and SPL of each model in the AVD dataset

Method	Table	Exit	Couch	Refrigerator	Sink	Avg.
Random walk	34.8/12.9	29/11.3	29.8/10.8	27.4/10.7	23/10.2	28.8/11.2
TD-A3C	45.8/5.8	37.6/6.3	37.2/5.0	16.8/4/4	23.4/4.7	32.2/5.2
Gated-LSTM-A3C	31/8.6	31.1/13.7	25.3/5.8	31.4/12.9	23/8.3	28.4/9.9
NeoNav	0.55/ 0.28	0.4/0.24	0.295/0.17	0.505/ 0.31	0.42/0.23	0.434/0.246
Ours	0.535/0.3	0.355/0.24	0.355/0.2	0.49/0.33	0.49/0.25	0.445/0.264

将 AVD 中 5 类物体作为智能体导航的目标,测试智能体在 200 步之内是否能够达到对应目标,从 5 个模型的测试结果如表 3 所示,改进模型在其中 3 个目标的 SR 略低于 NeoNav,在另外两个目标的 SR 远好于 NeoNav,在 1 000 步内的平均 SR 相较于 NeoNav 提升约 2%;而对于 SPL,则在每个目标导航过程中都好于 NeoNav,在 1 000 步内相较于 NeoNav 提升约 7%,证明改进模型在导航到不同目标时,兼顾成功率和路径轨迹长度的同时,获得了更好的泛化能力。

为了进一步验证改进模型的整体导航性能提升,本文测试了 NeoNav 模型和改进模型在 1 000 步以内的 SPL 和 SR,从两个模型导航的 1 000 步以内总的 SPL 和 SR 来看,改进模型的 SPL 相较于 NeoNav 提升约 6%,SR 相较于 NeoNav 提升约 2%,证明改进模型在应对不同导航目标是有良好的泛化性,如表 4 所示。

表 4 Neo 模型和改进模型在 1 000 步内的比较

Table 4 Comparison of Neo model and improved model within 1 000 steps

	Spl(total)	Success(total)
NeoNav	249.49	434/1 000
NeoNav-attention	264.88	445/1 000

4 结 论

本文在已有的基于深度强化学习视觉导航框架下,提出一种结合注意力机制和下一次预期观察的视觉导航模型,根据智能体第一视角观察到的 RGB-D 图像和目标点在智能体坐标系下的极性坐标,通过改进网络结构和跨连接结构训练和仿真,能够有效的提取观察图像的特征信息,使智能体能够更加关注视图中的有用信息。实

验结果表明,算法精度满足导航实时配合的要求。未来的工作将考虑将视觉信息和语义信息等融合后跨模态导航,智能化的视觉导航要想真正应用到实际机器人上,需要进一步提升模型在实际场景中的泛化能力,人机交互也是未来研究的一个重要方向。

参考文献

- [1] NWAONUMAH E, SAMANTA B. Deep reinforcement learning for visual navigation of wheeled mobile robots[C]. 2020 SoutheastCon, 2020: 1-8.
- [2] KHAIRUDDIN A R, TALIB M S, HARON H. Review on simultaneous localization and mapping (SLAM)[C]. 2015 IEEE International Conference on Control System, Computing and Engineering (ICCSCE). IEEE, 2015: 85-90.
- [3] ZHU K, ZHANG T. Deep reinforcement learning based mobile robot navigation: A review[J]. Tsinghua Science and Technology, 2021, 26(5): 674-691.
- [4] 孙龙龙, 江明, 焦传佳. 基于运动矢量的改进视觉 SLAM 算法[J]. 电子测量与仪器学报, 2020, 34(9): 23-31.
SUN L L, JIANG M, JIAO CH J. Improved visual SLAM algorithm based on the motion vector [J]. Journal of Electronic Measurement and Instrumentation, 2020, 34(9): 23-31.
- [5] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning [J]. Nature, 2015, 518(7540): 529-533.
- [6] ZHU Y, MOTTAGHI R, KOLVE E, et al. Target-driven visual navigation in indoor scenes using deep reinforcement learning [C]. 2017 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2017: 3357-3364.
- [7] 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): 10001-10008.
- [8] MNIH V, HEESS N, GRAVES A. Recurrent models of visual attention [J]. Advances in Neural Information Processing Systems, 2014, 27.
- [9] YANG W, WANG X, FARHADI A, et al. Visual semantic navigation using scene priors [J]. arXiv preprint arXiv:1810.06543, 2018.
- [10] LYU Y, SHI Y, ZHANG X. Improving target-driven visual navigation with attention on 3D spatial relationships[J]. Neural Processing Letters, 2022: 1-20.
- [11] DRUON R, YOSHIYASU Y, KANEZAKI A, et al. Visual object search by learning spatial context [J]. IEEE Robotics and Automation Letters, 2020, 5(2): 1279-1286.
- [12] QIU Y, PAL A, CHRISTENSEN H I. Target driven visual navigation exploiting object relationships [J]. arXiv preprint arXiv:2003.06749, 2020.
- [13] ANDERSON P, CHANG A, CHAPLOT D S, et al. On evaluation of embodied navigation agents [J]. arXiv preprint arXiv:1807.06757, 2018.
- [14] WU Q, MANOCHA D, WANG J, et al. NeoNav: Improving the generalization of visual navigation via generating next expected observations[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(6): 10001-10008.
- [15] MNIH V, HEESS N, GRAVES A. Recurrent models of visual attention [J]. Advances in Neural Information Processing Systems, 2014, 27.
- [16] KAUL C, MANANDHAR S, PEARS N. Focusnet: An attention-based fully convolutional network for medical image segmentation [C]. 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). IEEE, 2019: 455-458.
- [17] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.
- [18] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3-19.
- [19] 于鲲, 丛明煜, 戴文聪. 基于三维模型的空间目标视觉位姿测量[J]. 仪器仪表学报, 2019, 40(4): 179-188.
YU K, CONG M Y, DAI W C. 3D model-based visual attitude measurement of spatial targets [J]. Chinese Journal of Scientific Instrument, 2019, 40(4): 179-188.
- [20] 汪刚志, 王新华, 陈冠宇, 等. 基于视觉的无人机空中加油目标识别技术研究 [J]. 电子测量技术, 2020, 43(13): 89-94.
WANG G ZH, WANG X H, CHEN G Y, et al. Research on vision-based aerial refueling target identification technology for UAVs [J]. Electronic Measurement Technology, 2020, 43(13): 89-94.
- [21] 吉丰, 胡江湖. 导航自定位辐射源衰落信号智能识别算法[J]. 国外电子测量技术, 2020, 39(7): 52-57.
- [22] JI F, HU J H. Intelligent identification algorithm of navigation self-localization radiation source fading signal [J]. Foreign Electronic Measurement Technology, 2020, 39(7): 52-57.
- [22] WU Q, GONG X, XU K, et al. Towards target-driven visual navigation in indoor scenes via generative imitation

learning [J]. IEEE Robotics and Automation Letters, 2020, 6(1): 175-182.

- [23] WANG T, WU Z, WANG D. Visual perception generalization for vision-and-language navigation via meta-learning [J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, DOI: 10.1109/TNNLS.2021.3122579.

作者简介



刘紫燕 (通信作者), 1997 年于贵州大学获得学士学位, 2000 年于贵州大学获得硕士学位, 现为贵州大学硕士生导师, 主要研究方向为无线通信系统、移动机器人、大数据挖掘分析。

E-mail: Leizy@sina.com

Liu Ziyen (Corresponding author), received her B. Sc. degree from Guizhou University in 1997, M. Sc. degree from Guizhou University in 2000, respectively. Now she is a master supervisor. Her main research interests include wireless communication system, mobile robot, big data mining and analysis.



杨模, 2021 年于贵州大学获得学士学位, 现为贵州大学硕士研究生, 主要研究方向为计算机视觉、目标检测。

E-mail: yangmo97@126.com

Yang Mo received his B. Sc. degree from Guizhou University in 2021. Now he is a M. Sc. candidate at Guizhou University. His main research interests include computer vision and object detection.