

DOI: 10.13382/j.jemi.B2407993

基于空间多尺度连续性特征提取的 轮式机器人经验地图构建*

申传艳¹ 牛晶¹ 高光浩¹ 郑佳豪¹ 张利鹏² 刘世锋¹

(1. 天水师范大学机电与汽车工程学院 天水 741001; 2. 燕山大学车辆与能源学院 秦皇岛 066004)

摘要:针对传统同步定位与地图构建(SLAM)算法建图过程复杂、调参难度大和泛化能力差等问题,提出了一种基于空间多尺度连续性特征提取的轮式机器人经验地图构建方法。首先,在 ResNet18 网络模型中添加金字塔切分注意力模块(PSA)和空洞空间卷积池化金字塔(ASPP)模块,PSA 对中间层的特征进行分组处理,通过计算不同通道的注意力权重捕获多尺度信息,提高特征的表达能力,ASPP 利用不同扩张率的空洞卷积和全局平均池化整合全局上下文信息,进一步强化空间多尺度连续性特征的表征;其次,利用改进的 ResNet-PSA-ASPP 网络,在 donkey_sim 仿真模拟器和机器人实际运行跑道上对采集的数据集进行深度学习训练,获取优化后的机器人转向角度预测模型;最后,利用 donkey_sim 仿真模拟器和机器人操作系统(ROS)分别在仿真环境和实际场景下进行模型性能测试实验。实验结果表明,提出的模型对转向角度预测的误差分别减少了 38.47%、44.34%、35.51%,相比经典的 ResNet18、ResNet50、VGGNet 等网络模型在特征提取能力、计算效率和建图准确度上均获得显著提升。

关键词: 轮式机器人;经验地图;空间多尺度连续性特征;PSA;ASPP;转向角度预测模型

中图分类号: TP391.4;TN911.7 **文献标识码:** A **国家标准学科分类代码:** 510.80

Experience map construction for wheeled robots based on spatial multi-scale continuity feature extraction

Shen Chuanyan¹ Niu Jing¹ Gao Guanghao¹ Zheng Jiahao¹ Zhang Lipeng² Liu Shifeng¹

(1. School of Mechanical, Electrical and Automotive Engineering, Tianshui Normal University, Tianshui 741001, China;

2. School of Vehicle and Energy, Yanshan University, Qinhuangdao 066004, China)

Abstract: To address the issues of complex mapping processes, difficult parameter tuning, and poor generalization in traditional simultaneous localization and mapping (SLAM) algorithms, this paper proposes a wheeled robot experience map construction method based on spatial multi-scale continuity feature extraction. First, PSA and ASPP modules are integrated into the ResNet18 architecture. PSA groups intermediate features and calculates attention weights across channels to capture multi-scale information, thereby enhancing feature representation. ASPP incorporates dilated convolutions with varying dilation rates and global average pooling to aggregate global contextual information, further strengthening the representation of spatial multi-scale continuity features. Then, the improved ResNet-PSA-ASPP model is trained on datasets collected in both the donkey_sim simulator and real-world robot racetrack scenarios. Finally, model performance is evaluated in both simulated and real-world environments using the donkey_sim simulator and the robot operating system (ROS). Experimental results show that the proposed model reduces steering angle prediction errors by 38.47%, 44.34%, and 35.51%, respectively, and significantly outperforms classical networks such as ResNet18, ResNet50, and VGGNet in feature extraction capability, computational efficiency, and mapping accuracy.

Keywords: wheeled robots; experience map; spatial multi-scale continuity features; PSA; ASPP; steering angle prediction model

收稿日期: 2024-11-28 Received Date: 2024-11-28

* 基金项目: 甘肃省重点研发计划(23YFFE0001)、甘肃省高校教师创新基金(2023A-1143)、天水师范大学产业支撑引导项目(CYZ2023-05)资助

0 引言

随着人工智能技术的飞速发展,轮式机器人在军民领域的应用日趋广泛。轮式机器人在自主导航过程中,需要依赖高精度的环境地图实现路径规划与避障。目前,主流的经验地图构建方法主要依赖同步定位与地图构建(simultaneous localization and mapping, SLAM)技术,结合里程计和多种传感器数据进行环境感知与地图生成。然而,传统方法通常需要大量的时间和精力,过程复杂冗余,耗时较长,泛化能力不足,且受环境因素干扰较强。

深度学习算法是人工智能技术的重要组成部分,近年来在理论知识和应用实践上都发展迅速,尤其在图像识别和分类领域表现优异。姬晓飞等^[1]提出一种姿势引导和特征融合的行人重识别算法, DukeMTMC-ReID 数据集的 Rank-1、mAP 各达到了 91.2%、81.8%, 具有较佳的实用性,模型虽融合了全局和局部特征,但对极端相似外观或细微差异的区分能力有限。肖湘等^[2]提出了一种多分支融合的卷积神经网络(convolutional neural network, CNN),基于多分支的特征提取模块在特征空间对道路上下文、道路空间细节和道路边界建模,模型具有较好的特征提取能力,但多分支结构计算量大,难以部署到低功耗设备。张勃兴等^[3]针对因行人图像背景差异大、人体外观相似导致的行人再识别准确率低的问题,提出了一种利用特征融合与多尺度信息的行人重识别方法,在 Market-1501 上, mAP 和 Rank-1 分别达到了 86.77% 和 94.83%。吕宜生等^[4]提出一种新的时空信息融合模型,在双流卷积网络的基础上引入门控循环单元(gated recurrent unit, GRU)网络来实现端到端自动驾驶车辆转向角预测,提出的时空模型在驾驶转向角预测的准确度和平稳性方面效果显著,但 CNN+GRU 的计算开销大,难以满足高帧率实时控制需求。

深度学习算法结合多尺度特征提取、定位和建图模块同时也推动了地图构建向高精度、语义化、自适应方向发展。侯远韶^[5]提出了基于深度学习的多源信息融合巡检机器人 SLAM 技术,在深度神经网络框架下将多个传感器信息进行融合,同时结合深度神经网络对局部特征的提取能力,估算巡检机器人的运动路径实现地图重建,但该模型中注意力机制仅处理空间维度,忽视了对传感器时序异步性的建模。房立金等^[6]提出了一种利用深度学习提高动态环境下视觉语义即时定位与地图构建的方法,消除了动态物体对建图的影响,提高了建图精度,但该方法存在对语义分割精度依赖高、实时性不强及场景适应性有限等问题。

尽管现有工作在多尺度空间特征提取和语义化建图

上取得了显著进展,但仍存在以下不足:1)多分支融合、双流 CNN+GRU 等结构往往参数量和计算量大,难以满足实时性和低功耗设备的要求;2)对语义分割精度、预训练模型依赖度高,面对新场景时泛化能力不足;3)多为分步处理,缺少能够同时兼顾高精度、低延迟、自适应性的统一连续特征提取方法。

针对上述现有研究中存在的问题,以轮式机器人的转向角预测控制为研究目标,提出了一种基于融合空间多尺度连续性特征提取的改进 ResNet18 神经网络模型的经验地图构建方法。该方法使用 ResNet18 网络模型作为基础架构,通过将金字塔切分注意力模块(pyramid split attention, PSA)和空洞空间卷积池化金字塔模块(atrous spatial pyramid pooling, ASPP)与残差模块相结合提升网络对空间多尺度连续性特征的关注度。相较于其他的特征提取网络,该模型结构设计上融合了轻量化残差学习、多尺度注意力机制与上下文建模能力,在准确率与模型复杂度之间取得了更优平衡。

1 机器人转向角识别模型构建

利用 ResNet18 网络模型作为基础架构,并结合 PSA 注意力机制和 ASPP 模块,针对机器人转向角特征进行识别预测训练。

1.1 ResNet18 基础架构

网络层数越深,卷积神经网络获取更深层次特征的能力越强,然而往往会伴随出现梯度消失或爆炸、性能下降等问题,ResNet 利用加入残差模块的方法有效缓解了此类问题^[7]。ResNet 利用恒等映射引入了快速连接通道,使得某一层所产生的输出越过一个或多个层级,将输入数据直接传导至输出部分,借助快速连接通道,堆叠层的输出可以拟合所必需的底层映射,以此推动这些层经拟合操作构建相应的残差映射^[8],残差映射定义为:

$$H(x_l, W_l) = F(x_l, W_l) + x_l \quad (1)$$

式中: x_l 为第 l 个残差单元的输入值,即上一层输出的特征映射; W_l 为第 l 个残差单元的权重; $H(x_l, W_l)$ 为底层映射; $F(x_l, W_l)$ 为残差函数。

残差模块结构如图 1 所示,通过快速连接通道,模型的参数量和计算量大幅降低,使得训练时的效率和收敛速度得到提升,同时克服了深度神经网络因样本量不足而导致的过拟合问题。底层的特征信息通过快速连接通道可以直接传输给高层,从而帮助网络更有效地学习到多尺度的特征表示。残差单元可表示为:

$$y_l = h(x_l) + F(x_l, W_l) \quad (2)$$

$$x_{l+1} = f(y_l) \quad (3)$$

式中: x_{l+1} 、 y_l 分别为第 $l+1$ 、 l 个残差单元的输出; $h(x_l)$ 为第 l 个残差单元的底层映射; f 为 ReLU 激活函数。

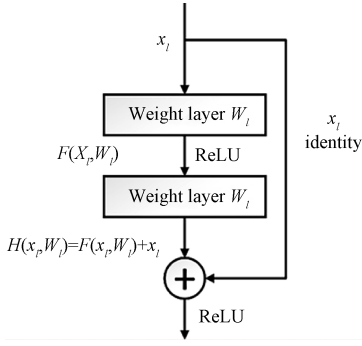


图1 残差模块结构

Fig. 1 Residual block

忽略激活函数,令 $x_{l+1} = f(y_l) = y_l$, 则浅层 l 到深层 L 的学习特征可表示为:

$$x_{l+1} = x_l + F(x_l, W_l) \quad (4)$$

$$x_L = x_l + \sum_{i=1}^{L-1} F(x_i, W_i) = x_0 + \sum_{i=0}^{L-1} F(x_i, W_i) \quad (5)$$

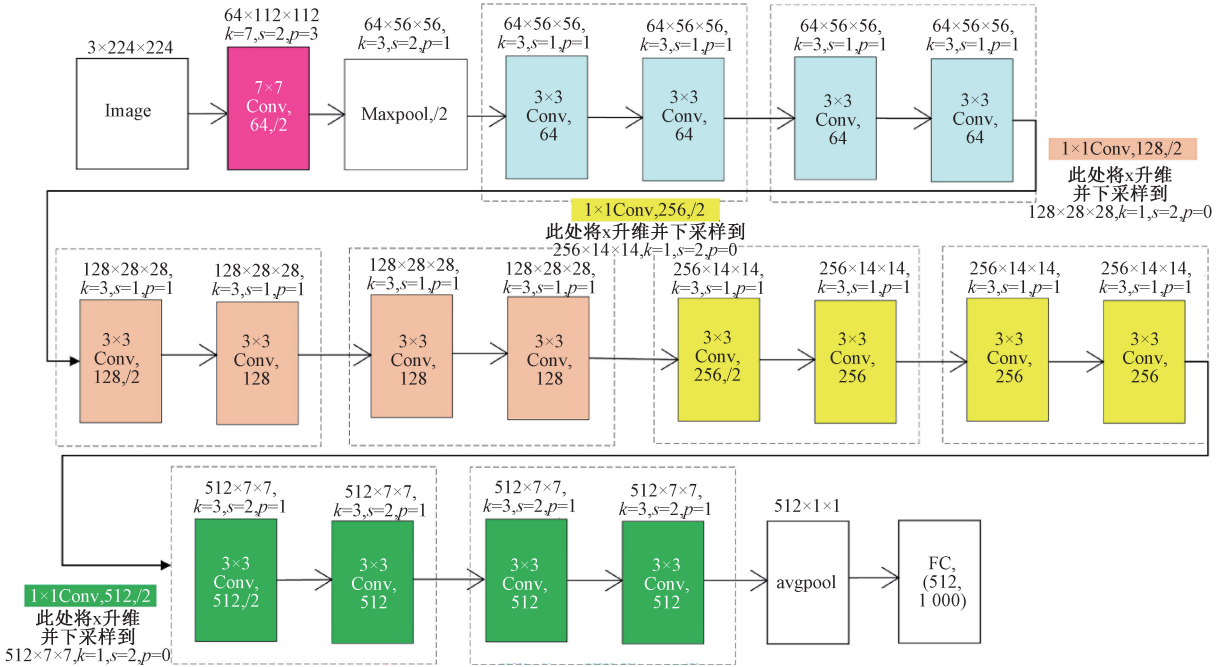
$$\frac{\partial \varepsilon}{\partial x_l} = \frac{\partial \varepsilon}{\partial x_L} \frac{\partial x_L}{\partial x_l} = \frac{\partial \varepsilon}{\partial x_L} \left[1 + \frac{\partial}{\partial x_l} \sum_{i=1}^{L-1} F(x_i, W_i) \right] \quad (6)$$

式中: $\frac{\partial \varepsilon}{\partial x_l}$ 为第 l 个残差单元不通过权重层传递的损失函数; $\frac{\partial}{\partial x_l} \sum_{i=1}^{L-1} F(x_i, W_i)$ 为第 l 个残差单元通过权重层传递的损失函数; x_i 为第 i 个残差单元的输入; W_i 为第 i 个残差单元的系列权重。

由式(2)~(4)可知,模型在前向传播时通过引入短路机制能更快速地获取浅层信息,同时反向传播的梯度消失问题也得到了有效避免,进而大幅度提升了网络模型的模型性能。

ResNet 系列模型主要包括多种网络层数各异的残差网络,常见的有 ResNet18、ResNet34、ResNet50、ResNet101 和 ResNet152。

考虑到模型实时推理计算的要求和难度,本文使用 ResNet18 网络模型为基础网络架构,其网络结构如图2所示。模型包含1个 7×7 卷积层、16个 3×3 卷积层(构成8个残差块,每个块包含2个卷积层)以及一个全连接层。考虑到模型泛化能力和多尺度特征提取的要求,在 ResNet18 基础架构基础上引入了 PSA 注意力模块和 ASPP 空洞空间卷积池化金字塔。



Conv 为卷积; s 为步长; k 为卷积核大小; p 为填充; Maxpool 为最大值池化; avgpool 为平均池化

图2 ResNet18 网络架构

Fig. 2 ResNet-18 architecture

1.2 金字塔切分注意力模块

注意力模块的嵌入可以实现 CNN 识别性能的显著提升,常见的注意力模块有^[9] 挤压-激励网络 (SENet)、瓶颈注意力模块 (BAM)、卷积块注意力模块 (CBAM)、

高效通道注意力网络 (ECANet)、全局上下文网络 (GCNet)、频率通道注意力网络 (FcaNet) 等。但目前这些模块在以下两个方面的表现仍有不足: 多尺度的空间特征信息不能高效获取,特征空间的表达不够丰富;通

道注意力或者空间注意力无法形成持久的依赖关系,只能有效提取局部特征。基于此,将轻量且高效的金字塔切分注意力 PSA 模块插入到 ResNet18 基础网络模型中。

PSA 通常由 3 部分组成:多尺度特征提取,这一部分通过不同尺度的卷积或池化操作进行特征提取。例如,可能会使用不同的卷积核大小或者池化窗口来捕获局部和全局信息;挤压-激励(squeeze and excitation, SE)模块,SE 模块在 PSA 中的作用通常是作为通道注意力机制来提升特征图中各通道的权重,它通过压缩全局信息并激活通道特征的方式,增强模型对通道信息的敏感性;挤压-拼接(squeeze and concat, SPC)模块,SPC 模块通过对特征图进行全局池化得到全局描述,并将这个全局描述与原始特征图拼接在一起,从而结合了局部空间信息和全局信息。这一操作为 PSA 模块带来了更丰富的特征表达能力,帮助模型同时关注空间和通道特征。

大多数注意力机制仅关注局部区域,忽略了全局空间注意力的变化。而金字塔切分注意力 PSA 模块能够同时关注通道注意力和空间注意力,捕获全局信息,并有效建立不同尺度通道注意力之间的持久依赖关系。将 PSA 模块替代 ResNet Bottleneck^[10]中的 3×3 卷积核,其他部分保持不变,从而形成了新的高效金字塔切分注意力(efficient pyramid split attention, EPSA)模块。基于此模块,构建了新的骨干网络 EPSANet,它具备强大的多尺度特征表示能力。与此同时,EPsANet 在图像识别任务中的评估分类模型性能指标 Top-1 准确率大幅度优于现有技术,效果如图 3 所示,且在计算参数量上更加高效,在 ResNet18 中引入该模块可以在降低计算量的同时提升网络性能。

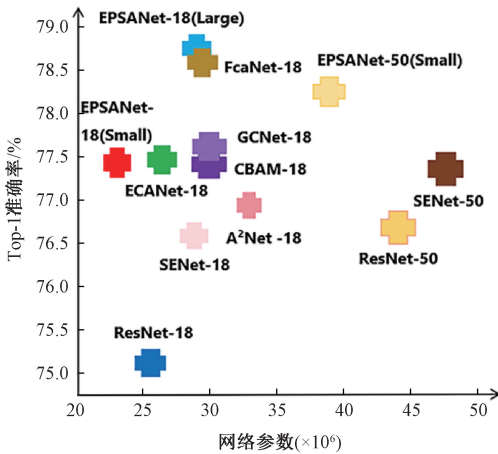


图 3 不同注意力模块准确率对比

Fig. 3 Accuracy comparison of different attention modules

1) SPC 模块

SPC 模块的主要作用是进行多尺度特征提取,通过

多分支合作的方式提取输入信息的空间特征信息,SPC 的结构如图 4 所示。各分支互不干扰,学习不同尺度的空间特征信息,通过局部化方式构建跨通道信息交互。该设计能够并行处理不同尺度空间信息,有利于丰富空间信息的捕捉。多尺度卷积因金字塔结构能更好的捕捉多维度空间特征,同时利用减少通道维数,提高空间特征的提取效率^[11]。假设输入为 X , SPC 模块首先将其分为 S 部分,依次为 X_0, X_1, \dots, X_{S-1} , 且每个部分的通道数均为 $C' = \frac{C}{S}$, 对不同部分提取不同尺度特征,最后将提取的多尺度特征通过 Concat^[12] 进行合并。随着内核尺寸的加大,SPC 模块能够在不增加计算量的前提下,通过分组卷积策略应对具有不同卷积核尺寸的输入向量,同时可以灵活修改群组大小而不添加额外的参数。卷积核 K 的大小和群组 G 关系如下:

$$G = 2^{\frac{K-1}{2}} \quad (7)$$

用于生成多尺度特征图的函数可表示为:

$$F_i = \text{Conv}(k_i \times k_i, G_i)(X) \quad i = 0, 1, 2, \dots, S-1 \quad (8)$$

式中:第 i 个卷积核尺寸 $k_i = 2 \times (i+1) + 1$, 第 i 个分组

大小为 $G_i = 2^{\frac{k_i-1}{2}}$, $F_i \in R^{C' \times H \times W}$ 。

据此融合后的完整多尺度特征图可表示为:

$$F = \text{Cat}([F_0, F_1, \dots, F_{S-1}]) \quad (9)$$

式中: $F \in R^{C \times H \times W}$ 。多尺度特征图得到后,需要重新标定各尺度特征图 F_i 的通道注意力权重,公式如下:

$$Z_i = \text{SEWeight}(F_i) \quad i = 0, 1, 2, \dots, S-1 \quad (10)$$

式中: $Z_i \in R^{C' \times 1 \times 1}$, 整个多尺度通道注意力权重向量为:

$$Z = Z_0 \oplus Z_1 \oplus \dots \oplus Z_{S-1} \quad (11)$$

为构建持久通道注意力的依赖关系,促进不同尺度通道注意力信息的相互作用,采用 Softmax 函数重新调整通道注意力信息权重,计算公式如下:

$$\text{att}_i = \text{Softmax}(Z_i) = \frac{\exp(Z_i)}{\sum_{i=0}^{S-1} \exp(Z_i)} \quad (12)$$

式中: $Z_i \in R^{C' \times 1 \times 1}$ 。将相匹配尺度的特征图 F_i 与重新调整后的注意力向量在通道级别进行逐元素相乘,即:

$$Y_i = F_i \odot \text{att}_i \quad i = 1, 2, 3, \dots, S-1 \quad (13)$$

对赋予权重后的不同尺度通道注意力特征图实施维度融合,生成更具信息量的特征图,计算公式如下:

$$\text{Out} = \text{Cat}([Y_0, Y_1, \dots, Y_{S-1}]) \quad (14)$$

2) SE Weight 模块

SE 模块^[13]聚焦于通道注意力机制,通过对每个通道的重要性进行权重调整,提升通道维度上的特征表达,结构如图 5 所示。包含 Squeeze 压缩和 Excitation 激励两部分,各应用于提取全局特征和通道关系的动态调整。

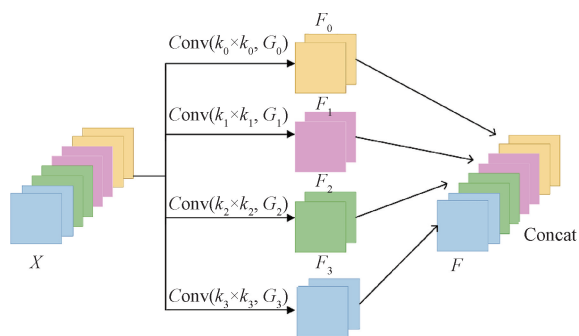


图4 SPC 模块结构

Fig. 4 SPC module structure

通过全局平均池化 (global average pooling, GAP) 获得通道数据,进而把全局特征信息融入到通道表示中。

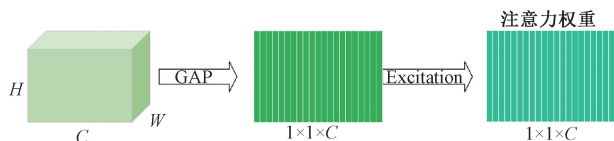


图5 SE Weight 模块结构

Fig. 5 SE weight module structure

第 c 个通道的输入特征映射为 X , 全局池化计算公式为:

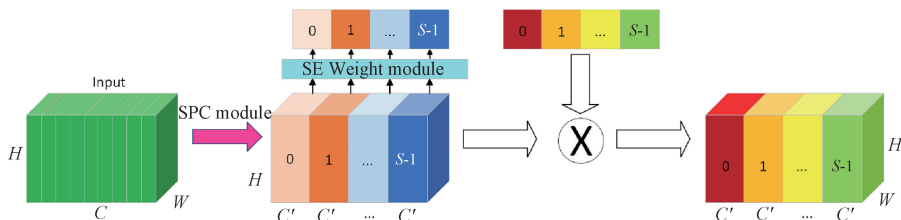


图6 PSA 模块结构

Fig. 6 PSA module structure

PSA 模块通过金字塔式的分割策略,将空间特征分成多个尺度,能够提取不同感受野下的特征信息,使模型具备多尺度感知能力,这样可以更好地捕捉图像中的不同大小、不同空间位置的信息。在多尺度特征上应用注意力机制,通过加权融合不同尺度的特征,使网络能够聚焦于更重要的空间位置和通道,进而提升模型的表示能力^[16]。由于 PSA 模块能够自适应地学习到多尺度特征的权重,在不同的任务中能有效提升网络性能。

1.3 ASPP 模块

ASPP 空洞空间卷积池化金字塔,它是空洞卷积和金字塔池化 (spatial pyramid pooling, SPP) 的结合。在面对特定输入时,ASPP 通过并行应用不同扩张率的空洞卷积

$$g_c = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W X_c(i, j) \quad (15)$$

式中: $X \in R^{C \times H \times W}$, C 、 H 、 W 分别表示特征图像的通道数、高度和宽度。

输入的特征信息首先进入全局池化层处理,然后依次进入全连接层、ReLU 函数层、第 2 个全连接层和 Sigmoid 激活函数层,最终计算出第 i 个通道特征的注意力权重,具体计算公式如下:

$$w_i = \sigma(W_1 \delta(W_0(g_i))) \quad (16)$$

式中: δ 表示 ReLU 激活函数; σ 表示 Sigmoid 激励函数;

W_0 和 W_1 代表两个全连接层, $W_0 \in R^{C \times \frac{C}{r}}$, $W_1 \in R^{\frac{C}{r} \times C}$, 前一个降维,后一个维度恢复。不同通道的线性特征信息能够被两个全连接层高效融合,促进高维和低维信息之间的相互影响与协同作用。随后各通道权重信息由激励函数重新调整分配,以便更高效地提取特征信息。

3) PSA 模块

PSA 模块^[14]同时对通道和空间维度进行注意力建模,结合了 SPC 模块的多尺度上下文特性和 SE 模块的通道加权能力。如图 6 所示,PSA 的工作原理为使用 SPC 模块切分各通道,对各通道特征信息进行多尺度空间信息提取,并使用 SE Weight 模块提取各通道注意力,生成各通道注意力向量;采用 Softmax 函数标定通道注意力向量特征,输出多尺度空间特征信息注意力加权特征图,生成的特征图含有更具多样性的空间特征信息^[15]。

来进行多尺度采样^[17]。如图 7 所示,ASPP 本质上由一个 1×1 的卷积层,多个具有不同扩张率的 3×3 卷积层和一个池化层组成。根据需要自行调节不同层的膨胀因子,可以更加灵活的提取不同尺度下的特征信息^[18-21]。

2 改进的 ResNet18 网络

构建的机器人转角识别图像数据集中的图片由于具备较高相似度,环境复杂多变等因素,不易于提取特征,为实现神经网络能提取到更具影响力和深层次的机器人转向过程的场景空间连续性变化特征信息,在 ResNet18 网络中引入了 PSA 注意力机制模块和 ASPP 模块,改进后的 ResNet18 网络结构如图 8 所示,layer1 ~ layer4 代表

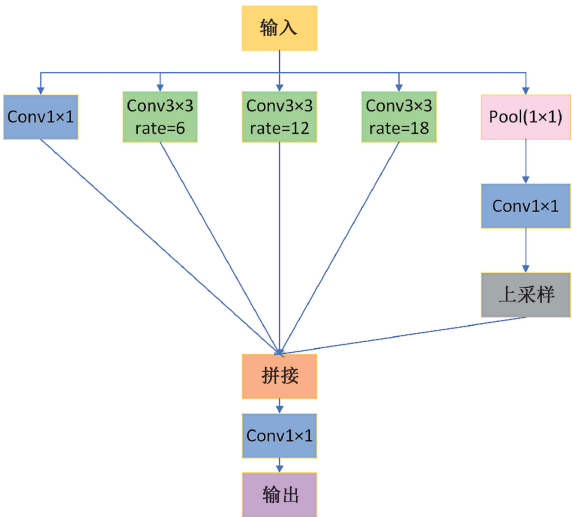


图 7 ASPP 模块结构
Fig. 7 ASPP module structure

4 个卷积层组,每个组包含两个残差块。

ResNet18 的初始卷积和池化层保持不变,包括第一个卷积层、批归一化 (BN)、ReLU 激活函数和最大池化层。layer1 和 layer2 保持不变,继续作为浅层模块提取特征。在 layer3 之后,将 PSA 注意力模块插入,PSA 接收 layer3 提取的特征,进行多尺度特征划分和注意力加权,输出跳跃连接相加前,特征图将先通过 PSA 模块进行多尺度注意力处理。PSA 模块主要通过通道注意力机制对特征进行再加权,使网络对不同位置和通道的信息具有自适应性。PSA 增强后的特征再输入到下一个卷积层组 layer4,layer4 完成高层特征提取后,将 ASPP 模块插入。ASPP 模块利用多尺度扩张卷积对 layer4 的输出特征进行多尺度融合,进一步捕捉更大的上下文信息,ASPP 增强后的特征将送往分类器。

与传统的 ResNet18 相比,引入 PSA 注意力机制,可以为 ResNet18 提供更丰富的多尺度信息,在处理包含不

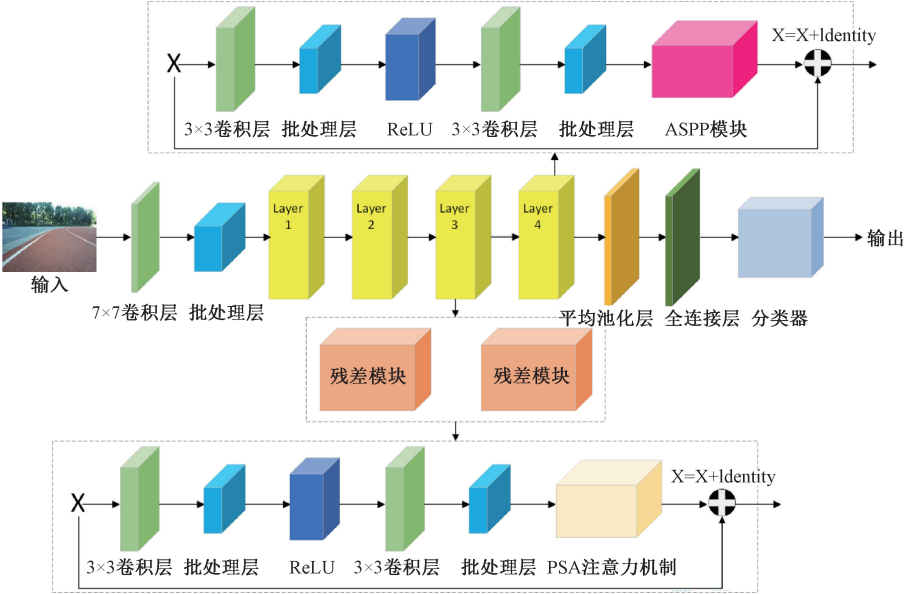


图 8 改进后模型的结构
Fig. 8 Improved models structure

同尺度目标的图像时表现更佳。能够有效提升模型在细节和整体结构上的平衡,能够更好处理跑道上的转角信息。通过 PSA 模块的通道注意力和空间注意力机制,ResNet18 可以自适应地选择更有用的特征,使得网络更具表达力。在图像分类、语义分割等任务中,这种增强的特征选择能力可以显著提高转向角度识别性能,通过对不同尺度的特征加权融合,增加了 ResNet18 在不同场景下的鲁棒性,这对于复杂的户外场景或多样化数据集,尤其是自动驾驶等任务中的场景分类、物体识别等任务非常有用。

ASPP 模块利用不同的扩张率进行卷积操作,能够捕

捉不同尺度的信息。对于场景理解和角度检测等任务,ASPP 可以提取从局部到全局的特征,提高模型对物体不同尺度的感知能力。通过空洞卷积,ASPP 可以在不增加计算量的情况下扩大卷积核的感受野,使模型能够聚合更多空间上下文信息。通过多尺度的信息融合,可以增加特征表达的丰富性,这对于捕捉场景中的细微差别和特殊模式有很大帮助。尤其是结合 ResNet18 的残差结构,ASPP 可以有效补充特征提取的不足之处,提升模型的泛化能力^[22]。

PSA 注意力机制和 ASPP 模块的插入位置不会影响 ResNet18 的原有残差连接,它们只是作为附加模块,不改

变核心残差路径。PSA 模块增强的特征进入 layer4, ASPP 模块增强的特征直接送入全局池化和分类层,因此残差结构的跳跃连接和加法操作依然有效,保持了残差网络的梯度流动优势。

3 试验与分析

3.1 数据与环境

本文图像数据在学校西操场采集。操控轮式机器人通过奥比中光 Gemini Pro 双目深度相机采集以跑道、天空为背景的机器人转向图像,轮式机器人的硬件架构如图 9 所示。该轮式机器人由共轴摆式悬挂系统小车底盘、底层主控、一个 M10P 激光雷达和一个 Gemini Pro 双目深度相机等组成。相机视场角为水平 71.5°、垂直 56.7°,景深范围 0.25 m 至无穷远,最大帧率 30 fps 以及分辨率为 1 920 pixel×1 080 pixel。移动机器人最大速度为 0.63 m/s。



图 9 机器人结构
Fig.9 Robot structure

图像数据集包含了不同光照条件下机器人在操场跑道上运行的空间场景图像,共计 16 435 张。图 10(a)~(i)所示分别为机器人在晴天、阴天和下雨天时左转、右转、直行情况下所采集的图像。

1) 数据集预处理

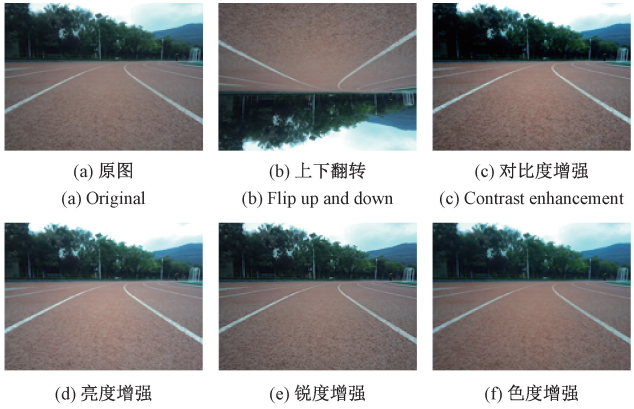
为增强卷积神经网络模型的泛化能力和准确性,避免过拟合,因此将采集后的图片通过锐度调整、亮度调整、翻转、色度调整、对比度调整等方法进行数据集增强扩充,如图 11(a)~(f)所示。按照 8 : 2 的比例划分训练集和测试集,用于网络预训练以获得新权重,具体信息如表 1 所示。图像标签在采集时命名为转向角度大小。为实现改进后的网络模型训练效率更高和收敛速度更快,训练前将图片尺寸调整为 224 pixel×224 pixel。

2) 实验环境与参数设置

实验在 Windows 10 操作系统下进行,配备了 13 代 Intel Core i5-13600F 处理器(主频 3.50 GHz),并搭载



图 10 训练集图像示例
Fig. 10 Examples of images from the training set



(d) Brightness enhancement (e) Sharpness enhancement (f) Chromaticity enhancement

图 11 数据增强样本

Fig. 11 Augmented samples

NVIDIA GeForce RTX 4060 Ti 显卡,16 GB 显存。深度学习框架采用 PyTorch 1.12.1,编程语言为 Python,代码开发和运行环境使用 PyCharm。超参数的大小对网络模型

表 1 转向角图像数量分布
Table 1 Distribution of the number of images per steering angle

类别	数据总量	训练集数量	测试集数量
左转	4 680	3 744	936
右转	6 491	5 193	1 298
直行	5 264	4 211	1 053
总计	16 435	13 148	3 287

训练的效率和收敛速度影响显著,经过多次调参,最优参数为将学习率调整为 1×10^{-4} ,批量大小设置为 32,训练迭代次数为 100,采用 Adam 优化器。

3.2 性能评价指标

采用均方误差(mean squared error, MSE)、平均绝对误差(mean absolute error, MAE)和均方根误差(root mean squared error, RMSE)^[23] 3 个评价指标来评估改进后模型的转向角预测性能。

MSE 是计算预测值与真实值之间误差的平方的平均值,其公式为:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (17)$$

式中: y_i 是第 i 个样本的真实值; \hat{y}_i 是第 i 个样本的预测值; n 是样本的数量。

MAE 是计算预测值与真实值之间误差的绝对值的平均值,其公式为:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (18)$$

式中: y_i 是第 i 个样本的真实值; \hat{y}_i 是第 i 个样本的预测

值; n 是样本的数量。

RMSE 是均方误差的平方根,其计算公式为:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (19)$$

式中: y_i 是第 i 个样本的真实值; \hat{y}_i 是第 i 个样本的预测值; n 是样本的数量。

3.3 改进网络的性能测试

1) ResNet18 网络改进前后的模型性能对比

通过测试集中的样本对改进后的网络模型进行性能评估,改进前与改进后的模型性能由图 12、13 和表 2 所示,与 ResNet18 模型相比,改进后的模型 MAE 减少了 38.47%,为 0.147 8; MSE 降低了 44.34%,仅为 0.048 7; RMSE 减小了 35.51%,为 0.237 5,模型性能提升的同时,迭代 100 次所用时长减少了 0.199 h,仅为 1.907 h,测试结果表明,本文针对 ResNet18 网络模型的改进能有效提高转向角预测的准确率,增强模型的识别性能,减少模型的迭代时长。

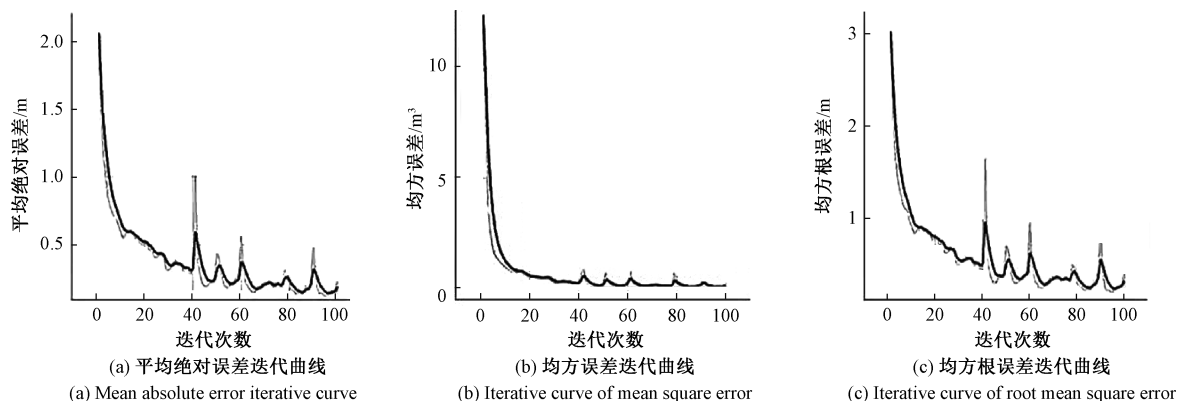


图 12 模型改进前的误差曲线

Fig. 12 Error curve prior to model improvement

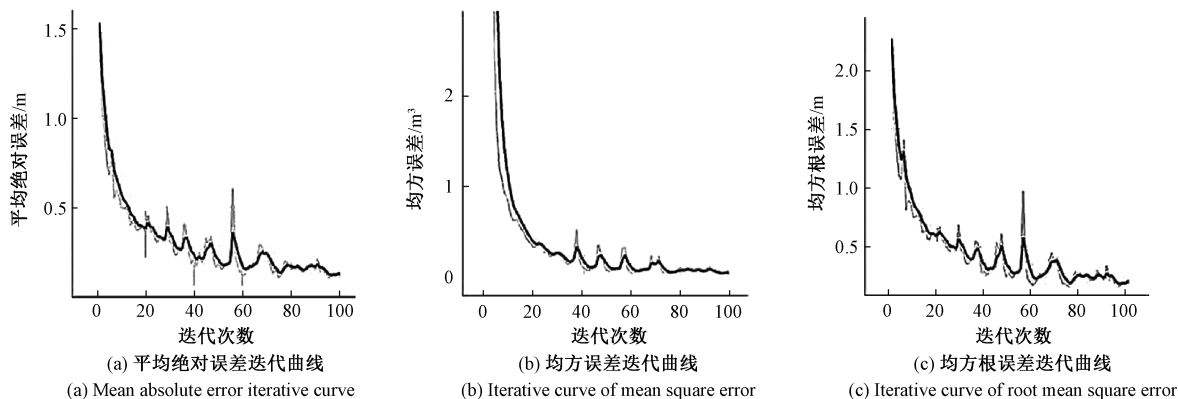


图 13 模型改进后的误差曲线

Fig. 13 Error curve following model improvement

表 2 改进前后的模型性能对比

Table 2 Model performance comparison before and after improvement				
模型	MAE	MSE	RMSE	迭代 100 次用时/h
ResNet18	0.240 2	0.087 5	0.368 3	2.106
改进 ResNet18	0.147 8	0.048 7	0.237 5	1.907

2) 消融实验结果分析

为了检验不同优化策略对模型识别误差的影响,开展了若干消融实验,结果如表 3 所示。在 ResNet18 模型架构的基础上,分别引入了两项优化策略(只引入 PSA 注意力模块、同时引入 PSA 注意力模块和 ASPP 模块),模型的均方误差、平均绝对误差、均方根误差都有效降低,尤以同时引入 PSA 注意力模块和 ASPP 模块的模型性能最好,较原模型误差分别减少了 38.47%、44.34%、35.51%。

表 3 消融实验结果
Table 3 Ablation study results

引入 PSA 注意力机制	引入 ASPP 模块	MAE	MSE	RMSE
×	×	0.240 2	0.087 5	0.368 3
√	×	0.236 2	0.071 1	0.325 8
√	√	0.147 8	0.048 7	0.237 5

注:“×”表示未进行此项改进;“√”表示进行了此项改进

3) 改进 ResNet18 模型与其他优秀网络模型的性能对比

为深入测试所提出的改进模型性能,将其与其他优秀的神经网络模型 ResNet18、ResNet34、ResNet50、VGGNet、GoogLeNet 进行对比实验,实验结果如表 4 所示。由表 4 可知,超参数设定和其他实验条件相同的情况下,由于参数量较少、数据集规模不大,ResNet18 模型在 3 种不同深度 ResNet 模型中迭代 100 次所用时长最短,均方误差、平均绝对误差和均方根误差均低于 ResNet34 和 ResNet50。所改进的网络模型迭代完成所用时长低于上述 3 种 ResNet 模型以及 VGGNet 和 GoogLeNet 模型,均方误差、平均绝对误差和均方根误差分别降低显著。实验结果表明,提出的模型在识别误差上显著降低,提高了空间多尺度连续性特征识别的准确性和精度,为模型在机器人系统中的实际应用和部署奠定了基础,此外也证明了该优化方法的有效性。

3.4 公开数据集验证

为了验证所提出 ResNet-PSA-ASPP 网络模型的先进性,在 TUM 公开数据集上进行试验,TUM 数据集是由慕尼黑工业大学计算机视觉组提供的标准 RGB-D 数据集,主要用于视觉 SLAM 和机器人导航算法的开发与评估。

表 4 改进 ResNet18 与其他模型性能对比

Table 4 Performance comparison between the improved ResNet18 and other models				
模型	MAE	MSE	RMSE	迭代 100 次用时/h
ResNet18	0.240 2	0.087 5	0.368 3	2.106
ResNet34	0.266 7	0.089 3	0.396 2	2.357
ResNet50	0.271 2	0.089 9	0.397 8	2.395
VGGNet	0.938 4	0.156 2	0.778 3	3.987
GoogLeNet	0.536 9	0.113 5	0.657 9	2.489
改进 ResNet18	0.147 8	0.048 7	0.237 5	1.907

表 5 为本文模型与其他对比模型利用 TUM 数据集预测转向角度的 MSE、模型运行速度及模型参数量大小。只对不同模型 MSE 误差最低情形下的运行速率及模型参数量大小进行对比,运行速度不包括数据处理时间。由表 5 可知,虽然 ResNet18 模型均方误差为 0.093 1,没有达到最高,但模型参数量为 11.69×10^6 ,运行速度 530 f/s,是所有模型中运行速度最快、参数量最小的,对边缘部署任务友好。所提出的改进 ResNet18 模型均方误差为 0.053 8,在对比模型中均方误差最低,模型参数量为 14.84×10^6 ,运行速度为 490 f/s,在精度与性能之间取得了良好的平衡。

表 5 各模型运行速度及大小对比

Table 5 Comparison of inference speed and model size for each model			
模型	MSE	速率/fps	参数量/($\times 10^6$)
ResNet18	0.093 1	530	11.69
ResNet34	0.093 3	380	21.80
ResNet50	0.093 8	320	25.56
改进 ResNet18	0.053 8	490	14.84

3.5 轮式机器人虚拟平台仿真实验

为进一步验证文中提出的改进 ResNet18 网络的整体性能,在 Unity 开发的 donkey_sim 模拟器上对该网络进行了测试。

采集的图片命名标签为转向角度,将采集的数据集拆分为训练集和测试集,使用改进的 ResNet18 网络对训练集进行训练,在训练过程中使用数据可视化工具 tensorboard 进行损失函数变化监控。在该模拟器中,控制小车的主要是油门和转向角度两个参数。训练结束后将模拟器中小车转向值设置为 0,油门值设置为 0.1 匀速行驶,通过代码将改进的神经网络训练出的模型导入到模拟器中,运动过程中根据获取的图像自动推理输出转向角度,控制小车运动,构建出的经验地图部分场景如图 14 所示。

从图 14(a)~(l) 可以看出,无论在何种场景地图下,模型推理出的转向角度均能够很好地引导小车行驶,

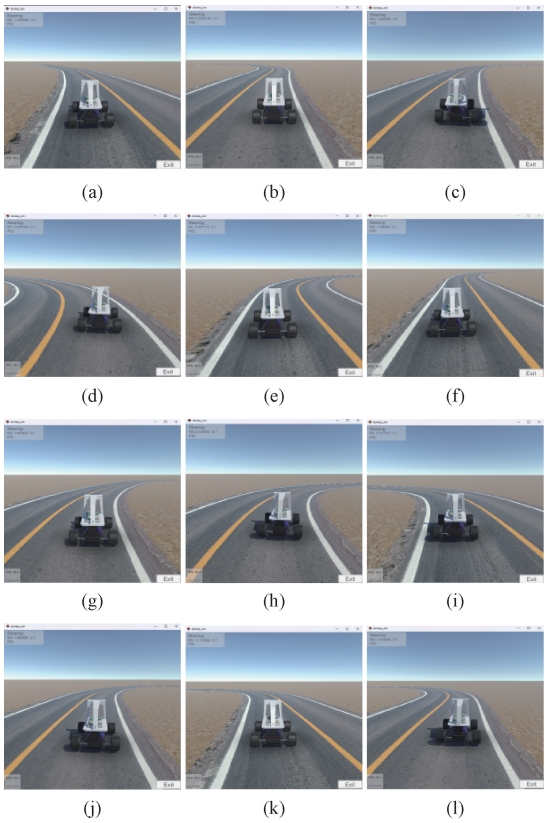


图 14 虚拟平台经验地图构建

Fig. 14 Construction of the virtual platform experience map

使其保持在车道线的中间位置。这表明模型具有较高的泛化能力和鲁棒性,能够在不同的环境下精准识别小车行驶轨迹,输出合理的转向角度,并进行合理的路径规划与控制。通过这一特性,模型在复杂场景中不仅能够保持小车稳定行驶,还可以构建出适应环境变化的经验地图。这种能力使得模型在面对复杂多变的场景时,能够根据实时推理结果调整驾驶策略,从而为小车提供更加智能和高效的导航支持,进一步验证了模型在端到端自动驾驶任务中的适用性和可靠性。

3.6 实景实验

为验证所提模型的有效性,并更精确地分析该模型在轮式机器人上构建经验地图的性能与精度,本实验将改进的 ResNet18 神经网络训练得到的模型部署到采集图像的轮式机器人操作系统 (robot operating system, ROS) 中,在真实操场环境下开展经验地图构建实验,如图 15 所示。任选其中一条跑道,起点任意设置,将机器人放置在两条跑道线中间。

导入小车的运动库和其他相关的库函数,同时打开相机指令,定义小车初始油门为 0.1 m/s,初始转向角为 0°,设置改进网络模型路径和推理环境。小车在运动过程中通过获取图像和信息预处理,自动计算输出转向角

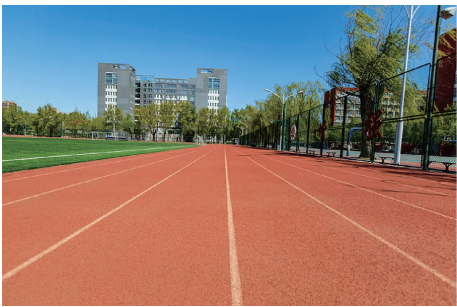


图 15 实验场景图片

Fig. 15 Images of the experimental setup

度并执行相应的动作。通过惯性测量单元 (inertial measurement unit, IMU) 中的加速度计和陀螺仪测量加速度和角速度并结合里程计数据,对这些数据进行积分来获得相对运动轨迹,不同模型的相对运动轨迹对比如图 16 所示,各模型的转向角偏差均值如表 6 所示。

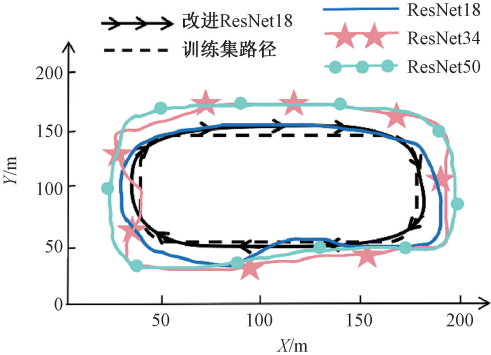


图 16 训练集与真实路径对比

Fig. 16 Comparison of the training set and real path

表 6 不同模型转向角偏差均值

Table 6 Mean steering angle deviation of different models

模型	ResNet18	ResNet34	ResNet50	改进 ResNet18
转向角偏差均值/(°)	0.59	1.23	1.52	0.15

从图 16 和表 6 可知, ResNet34 和 ResNet50 神经网络模型推理出的路径相差不大,均偏离训练集路径较多,转向角偏差均值都大于 1°, ResNet18 推理出的路径优于 ResNet34 和 ResNet50,转向角偏差均值为 0.59°,改进的模型推理出的转向角度与训练集路径大致相同且优于其他模型,转向角偏差均值仅为 0.15°。这充分验证了文章提出模型的有效性。模型不仅在推理速度和稳定性方面表现出色,同时在角度预测的精度上也达到了实际应用的要求,完全能够满足轮式机器人构建经验地图的需求。这一结果表明,该模型在复杂场景中表现出较强的适应性和鲁棒性,为自主导航任务提供了有力的技术支持。

4 结 论

本文基于 ResNet18 进行改进,提出了一种空间多尺度连续性特征提取的网络模型—ResNet-PSA-ASPP。核心思路在于借助高效的特征提取和多尺度融合,在不增加计算量的基础上全面提升模型在经验地图构建中的表现力。ResNet-PSA-ASPP 采用 ResNet18 作为主干网络,引入 PSA 注意力模块和 ASPP 模块,改进的网络模型不仅显著提升了空间多尺度连续性特征提取效率,还进一步捕捉更大的上下文信息,扩大了卷积核的感受野,在转向角预测准确率与模型复杂度之间取得了更优平衡。在实验测试中,较 ResNet18、ResNet50、VGGNet 等优秀网络模型,提出模型具有更高的预测精度,预测的误差分别减少了 38.47%、44.34%、35.51%,在真实场景的构建经验地图实验中也验证了模型的有效性。未来进一步探索不同尺度的卷积核或池化方式,以适配具体任务。比如,对于自动驾驶的场景识别,可能需要更大尺度的感受野来捕捉道路环境,未来将进一步优化模型结构和增强数据多样性,以提高模型的鲁棒性。

参考文献

- [1] 姬晓飞,赵帅,宋京浩,等. 基于姿势估计和特征融合的行人重识别算法[J]. 电子测量与仪器学报, 2024, 38(4): 187-194.
JI X F, ZHAO SH, SONG J H, et al. Person re-recognition algorithm based on pose estimation and feature fusion[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(4): 187-194.
- [2] 肖湘,金立左. 基于多分支融合网络的实时道路识别算法[J]. 工业控制计算机, 2022, 35(10): 136-137, 140.
XIAO X, JIN L Z. Real-time road recognition algorithm based on multi-branch fusion network[J]. Industrial Control Computer, 2022, 35(10): 136-137, 140.
- [3] 张勃兴,马敬奇,张寿明,等. 利用全局与局部关联特征的行人重识别方法[J]. 电子测量与仪器学报, 2022, 36(6): 205-212.
ZHANG B X, MA J Q, ZHANG SH M, et al. Person re-identification method using global and local association features[J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(6): 205-212.
- [4] 吕宜生,刘雅慧,陈圆圆,等. 融合时空特征的端到端自动驾驶车辆转向角预测[J]. 中国公路学报, 2022, 35(3): 263-272.
LYU Y SH, LIU Y H, CHEN Y Y, et al. End-to-end steering angle prediction of autonomous vehicles based on spatiotemporal characteristics[J]. China Journal of Highway and Transport, 2022, 35(3): 263-272.
- [5] 侯远韶. 基于深度学习的多源信息融合巡检机器人 SLAM 技术研究[J]. 信息技术与信息化, 2023(8): 217-220.
HOU Y SH. Research on SLAM technology of multi-source information fusion inspection robot based on deep learning[J]. Information Technology and Informatization, 2023(8): 217-220.
- [6] 房立金,刘博,万应才. 基于深度学习的动态场景语义 SLAM[J]. 华中科技大学学报(自然科学版), 2020, 48(1): 121-126.
FANG L J, LIU B, WAN Y C. Dynamic scene semantic SLAM based on deep learning[J]. Journal of Huazhong University of Science and Technology (Natural Science Edition), 2020, 48(1): 121-126.
- [7] 潘仁勇,张欣,陈孝玉龙,等. 基于 DTS-ResNet 的苹果叶片病害识别方法[J]. 国外电子测量技术, 2022, 41(9): 142-148.
PAN R Y, ZHANG X, CHEN X Y L, et al. Identification method of apple leaf disease based on DTS-ResNet[J]. Foreign Electronic Measurement Technology, 2022, 41(9): 142-148.
- [8] 李韬,朱文忠,车璇. 基于改进 ResNet-50 与迁移学习的苹果叶片病害的图像识别[J]. 科学技术与工程, 2024, 24(24): 10370-10381.
LI T, ZHU W ZH, CHE X. Image recognition of apple leaf diseases based on improved ResNet-50 and transfer learning[J]. Science Technology and Engineering, 2024, 24(24): 10370-10381.
- [9] 李文涛. 基于改进卷积神经网络的小目标检测方法研究[D]. 无锡:江南大学, 2022.
LI W T. Research on small object detection method based on improved convolutional neural network[D]. Wuxi: Jiangnan University, 2022.
- [10] 刘熠翕,江旻珊,张学典. 融合金字塔切分注意力模块的视杯视盘分割[J]. 上海理工大学学报, 2022, 44(6): 532-539, 545.
LIU Y X, JIANG M SH, ZHANG X D. Fusion of pyramid segmentation attention module of the optic cup disc segmentation[J]. Journal of University of Shanghai for Science and Technology, 2022, 44(6): 532-539, 545.
- [11] LI T, LV X Y, WANG G, et al. UAV image object recognition method based on small sample learning[J]. Multimedia Tools and Applications, 2023, 82(17): 26631-26642.
- [12] MA C C, MI J Q, GAO W L, ET AL. SSGAN: A semantic similarity-based GAN for small-sample image

- augmen- tation [J]. Neural Processing Letters, 2024, 56:149.
- [13] 陈诗瑶,孔淳,冯峰,等. 一种基于改进 ResNet18 神经网络的苹果叶片病害识别方法[J]. 山东农业科学, 2024, 56(10):174-180.
- CHEN SH Y, KONG CH, FENG F, et al. A method for identifying apple leaf diseases based on improved ResNet18 neural network [J]. Shandong Agricultural Sciences, 2024, 56(10):174-180.
- [14] LI G Q, YU L, FEI SH M. A deep-learning real-time visual SLAM system based on multi-task feature extraction network and self-supervised feature points[J]. Measurement, 2021, 168: 108403.
- [15] YANG X, ZHANG Y F, GUO Y Q, et al. An image super-resolution deep learning network based on multi-level feature extraction module[J]. Multimedia Tools and Applications, 2021, 80(5):7063-7075.
- [16] KIRTI K, NAVIN R, JYOTSNA Y, et al. Multi-resolution analysis and deep neural network architecture based hybrid feature extraction technique for plant disease identification and severity estimation [J]. Evolutionary Intelligence, 2022, 17(2): 1163-1183.
- [17] DAI G W, TIAN ZH M, FAN J CH, et al. DFN-PSAN: Multi-level deep information feature fusion extraction network for interpretable plant disease classification[J]. Computers and Electronics in Agriculture, 2024, 216:108481.
- [18] GURUPRAKASH K S, SIVA K P, RAMACHANDRAN A, et al. Crop pest identification using deep network based extracted features and MobileENet in smart agriculture [J]. Land Degradation & Development, 2024, 35(11): 3642-3652.
- [19] 王芳,崔丹丹,李林. 基于深度学习的采摘机器人目标识别定位算法[J]. 电子测量技术, 2021, 44(20): 162-167.
- WANG F, CUI D D, LI L. Target recognition and localization algorithm for picking robot based on deep learning[J]. Electronic Measurement Technology, 2021, 44(20): 162-167.
- [20] 石昌友,孙强,卢建平,等. 多尺度卷积神经网络的图像边缘检测[J]. 电子测量技术, 2022, 45(8): 121-128.
- SHI CH Y, SUN Q, LU J P, et al. Image edge detection based on multi-scale convolutional neural network [J]. Electronic Measurement Technology, 2022, 45(8): 121-128.
- [21] 李伟,张旭东. 基于卷积神经网络的深度图像超分辨率重建方法 [J]. 电子测量与仪器学报, 2017, 31(12): 1918-1928.
- LI W, ZHANG X D. Super-resolution reconstruction method of deep image based on convolutional neural network [J]. Journal of Electronic Measurement and Instrumentation, 2017, 31(12): 1918-1928.
- [22] 李梅梅,胡春海,龙平,等. 基于 MultiRes+UNet 网络的车道线检测算法[J]. 电子测量与仪器学报, 2020, 34(9): 117-122.
- LI M M, HU CH H, LONG P, et al. Lane line detection algorithm based on MultiRes+ UNet network[J]. Journal of Electronic Measurement and Instrumentation, 2020, 34(9): 117-122.
- [23] 熊焕,俞建峰,钱陈豪,等. 基于自适应特征融合的抓取检测方法研究 [J]. 电子测量与仪器学报, 2024, 38(10):118-127.
- XIONG H, YU J F, QIAN CH H, et al. Research on grasping detection method based on adaptive feature fusion [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(10): 118-127.

作者简介



申传艳, 2023 年于临沂大学获得学士学位, 现为天水师范大学硕士研究生, 主要研究方向为机器人移动路径规划及深度学习。

E-mail: 1339733752@qq.com



Shen Chuanyan received her B. Sc. degree from Linyi University in 2023. Now she is a M. Sc. candidate in Tianshui Normal University. Her main research interests include robot movement path planning and deep learning.

牛晶(通信作者), 2008 年于中国农业大学获得学士学位, 2010 年于北京理工大学获得硕士学位, 现为天水师范大学副教授, 主要研究方向为自动驾驶、机器人路径导航规划及深度学习。

E-mail: sensily@163.com

Niu Jing (Corresponding author) received her B. Sc. degree from China Agricultural University in 2008 and M. Sc. degree from Beijing Institute of Technology in 2010. Now she is an associate professor at Tianshui Normal University. Her main research interests include autonomous driving, robot path navigation planning and deep learning.