

基于多模态视觉自注意力机制的 机器人目标识别研究

马文静¹, 李亚哲¹, 高军亮²

(1. 郑州商业技师学院 商业贸易学院, 郑州 450121; 2. 郑州大学 网络空间安全学院, 郑州 450000)

摘要: 针对现有机器人目标识别算法识别精确率低、虚警率过高的不足, 在多模态条件下开展机器人目标识别研究, 设计视觉自注意力机制网络模型; 基于动态窗口构建机器人运动模型, 采集多模态数据并完成配准预处理, 通过评价函数评估机器人运动姿态精度; 将卷积神经网络 (CNN) 作为注意力模块的骨干网络, 嵌入注意力机制模块以优化特征提取效率, 结合位置编码挖掘关键特征信息, 引入双门控机制强化模型对长序列数据的处理能力与训练稳定性; 实验测试结果显示, 该识别算法在无遮挡场景下的测试集识别精确率为 99.1%、虚警率为 0.4%, 有遮挡场景下的测试集识别精确率为 98.0%、虚警率为 1.0%, 性能显著优于现有识别算法; 消融实验验证了模型多轮改进对性能提升的正向贡献率; 该方法可满足工业现场、户外作业等复杂场景下机器人目标识别的高精度、低误警需求, 为机器人精准执行任务提供可靠技术支持。

关键词: 多模态; 视觉注意力机制; 机器人目标识别; CNN; 门控机制

Research on Robot Target Recognition Based on Multimodal Visual Self-Attention Mechanism

MA Wenjing¹, LI Yazhe¹, GAO Junliang²

(1. School of Commercial Trade, Zhengzhou Commercial Technician College, Zhengzhou 450121, China;

2. Zhengzhou University, Zhengzhou 450000, China)

Abstract: Aiming at the shortcomings of existing robot target recognition algorithms, such as low recognition accuracy and excessively high false alarm rate, research on robot target recognition under multimodal conditions is conducted, with a visual self-attention mechanism network model presented. Based on a dynamic window, a robot motion model is constructed, which collects multimodal data, achieving registration preprocessing, thus evaluating the robot's motion posture accuracy through an evaluation function. A convolutional neural network (CNN) is adopted as the backbone network of the attention module, with the self-attention mechanism module embedded to optimize its feature extraction efficiency. The key feature information is mined by combining positional encoding, and a dual gating mechanism is introduced to enhance the model's ability to process long-sequence data and training stability. Experimental test results show that the recognition algorithm achieves a test set recognition accuracy of 99.1% and a false alarm rate of 0.4% in unoccluded scenarios, and a recognition accuracy of 98.0% and a false alarm rate of 1.0% in occluded scenarios. This algorithm is notably superior in performance to existing recognition ones. Ablation experiments verify that the model improves its performance iteratively. This method can meet the high-precision and low-false-alarm requirements of robot target recognition in complex scenarios such as industrial sites and outdoor operations, providing a reliable technical support for robots to accurately perform tasks.

Keywords: multimodality; visual attention mechanism; robot target recognition; CNN; gate control mechanism

0 引言

伴随着自动化控制技术、计算机科学技术、无线通

信技术和互联网技术的飞速发展, 机器人制造水平和智能化程度有了显著提升, 尤其是近些年机器学习^[1]、深度神经网络等技术的突飞猛进^[2], 赋予了可移动机器人

收稿日期: 2025-12-04; 修回日期: 2026-01-23。

基金项目: 郑州市社会科学调研课题(ZSJX20250794)。

作者简介: 马文静(1984-), 女, 大学本科。

引用格式: 马文静, 李亚哲, 高军亮. 基于多模态视觉自注意力机制的机器人目标识别研究[J]. 计算机测量与控制, 2026, 34(7): 261-267.

更多智能化和智慧化的发展空间。现阶段,移动机器人已经开始在工业制造、野外救援、高危作业、医疗卫生、家庭护理、餐饮服务等诸多领域崭露头角^[3-7]。在各类机器人的人机交互过程中,对标的物的目标识别是最关键的步骤和环节,目标识别也是机器人的核心感知功能之一,目标识别精度关系到机器人的运动、避障和任务的具体执行效果。当前,目标识别是可移动机器人设计、制造、控制领域的研究热点之一^[8],汇总分析现阶段国内外关于可移动机器人目标识别的方案,主要包括基于视觉图像、地图的采集与处理的识别方法,基于激光器定位与激光测量的检测方案,基于深度神经网络的目标识别与处理方案等。

文献 [9] 提出一种单目视觉即时定位与地图构建 (SLAM, simultaneous localization and mapping) (即时定位与地图构建) 的方法,双目视觉 SLAM 核心原理是通过融合视觉传感器、毫米波雷达等测量数据,同步估计可移动机器人自身运动轨迹变化和环境坐标的实时更新,进而实现对目标的精确识别。双目镜头在视觉图像定位方面有一定优势,但该方案仍旧未解决动态物体的干扰和漂移问题且算法的复杂度较高;文献 [10] 提出一种基于激光点云的目标识别方法,通过激光器发射光束并动态采集点云集合,激光点云目标识别的优势是基于点云匹配实现高精度建图,效率高且精度好。但该种方案的不足在于易受灰尘等颗粒物的干扰,且该种方案的实现成本较高;深度神经网络模型是当前数据处理、图像处理的最有效方法之一,以 YOLO 模型为代表的深度神经网络模型是最有效的方案之一,YOLO 模型的优势在于更新迭代速度较快,能够及时补偿原模型存在的不足。文献 [11] 提出一种基于 YOLOv8s 的目标识别方法,YOLOv8s 模型的优势在于小型化和轻量化,能够保持较高推理速度和图像处理速度,但 YOLOv8s 模型对于小目标的检测精度会显著下降且模型的损失函数有待改善。

针对现有方案的不足,本文提出一种多模态自注意力机制的移动机器人目标检测方案。自注意力机制是一种通过动态权重分配捕捉序列内部关联的深度学习机制,与传统的循环神经网络、前馈神经网络等网络类型相比,注意力机制无需受顺序处理数据的限制,且具有更好的并行数据训练和处理能力。本文针对机器人目标识别的多模态需求对自注意力机制进行了多重优化,具体包括在卷积网络框架嵌入注意力机制模块,提升对二维图像数据的处理能力,并引入了双门控机制改善模型对长序列数据的依赖情况,进一步通过提升注意力机制网络的性能,以改善机器人目标识别的效果。

1 机器人目标识别的框架设计

可移动机器人目标识别的基本框架包括系统的初始

化,机器人运动模型构建、姿态数据的采集、预处理与特征提取,目标识别的算法设计,如图 1 所示。

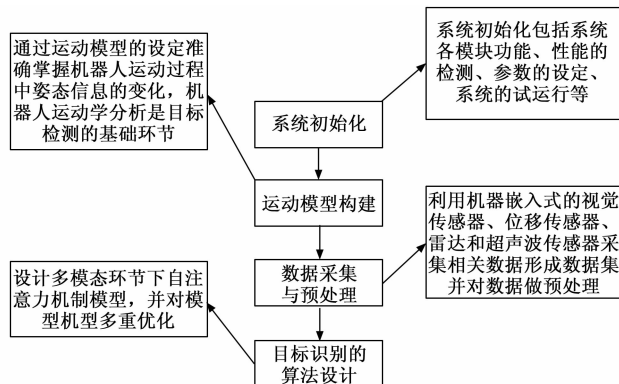


图 1 机器人目标识别系统总体框架设计

在机器人目标识别框架中运动学模型构建和算法设计是核心环节,且运动姿态的调整和运动数据的采集与反馈是目标识别算法应用的基础。

1.1 系统初始化及机器人目标识别运动学模型

可移动机器人在工作前先将系统初始化,验证各模块的功能和整个系统的功能,并确保软硬件在正常的状态下进行姿态数据的采集和目标识别。由于可移动机器人目标识别是在运动状态下的标的物目标识别^[12],因此本文设计了一种基于动态窗口的运动学模型,并在速度空间 (v, ω) 内做数据采样 (其中 v 为机器人运动线速度, ω 是机器人的运动角速度,线速度和角速度通过机器人内置的传感器测量)。可移动机器人在连续时刻的动态窗口运动模型平面图,如图 2 所示。

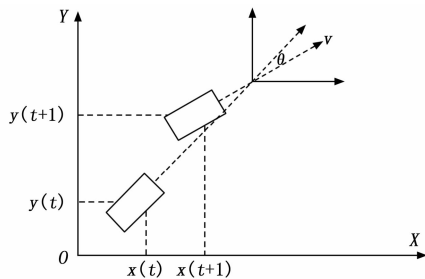


图 2 动态窗口的机器人运动模型

图 2 中,可移动机器人从 t 时刻到 $t+1$ 时刻的变化记为 Δt (根据本文机器人的运行速度 1 m/s,将时间窗口定义为 0.1 s),转向角度为 θ , t 时刻机器人的动态窗口运动表达式如下:

$$\begin{cases} \Delta x(t) = v(t) \Delta t \cos \theta \\ \Delta y(t) = v(t) \Delta t \sin \theta \\ \Delta \theta = \omega(t) \Delta t \end{cases} \quad (1)$$

式中, $\Delta x(t)$ 和 $\Delta y(t)$ 分别表示机器人在两个轴向的运动增量, $v(t)$ 表示在当前 t 时刻机器人的运动速度, $\omega(t)$ 表示当前 t 时刻的角速度,在下一个 $t+1$ 时刻机器人的

动态窗口运动表达式如下:

$$\begin{cases} x(t+1) = x(t) + v(t)\Delta t \cos\theta \\ y(t+1) = y(t) + v(t)\Delta t \sin\theta \\ \theta(t+1) = \theta(t) + \omega(t)\Delta t \end{cases} \quad (2)$$

式中, $x(t)$ 和 $x(t+1)$ 分别表示当前 t 时刻和下一 $t+1$ 时刻机器人在 x 轴向的运动情况, $y(t)$ 和 $y(t+1)$ 分别表示当前 t 时刻和下一 $t+1$ 时刻机器人在 y 轴向的运动情况。

1.2 设计机器人运动评价函数

受各种扰动和偏差的影响机器人运动中的轨迹可能会发生偏离^[13-14], 因此需要将机器人的运动速度 $v(t)$ 进行动态离散化解^[15], 从多种可能的运动轨迹中选择最优的轨迹。为此本文设计了一种关于线速度和角速度的评价函数 $G(v, \omega)$:

$$G(v, \omega) = G[\eta \cdot W(v, \omega) + \zeta \cdot D(v, \omega) + \xi \cdot V(v, \omega)] \quad (3)$$

式中, W 、 D 、 V 分别表示角速度、距离和线速度的子评价函数, η 、 ζ 、 ξ 分别为 W 、 D 、 V 对应的权重系数, η 、 ζ 、 ξ 取值在 $0 \sim 1$ 之间且三者的和为 1。 η 、 ζ 、 ξ 三者之间的取值是此消彼长的关系, 通过不同的赋权来控制总评价函数。合理的赋权后相当于对各子函数做归一化处理, 明确各不同子函数在机器人稳定性控制中的作用, 确保最终的评价结构更加科学与合理。合理的参数赋权还能够与子函数的取值范围和物理意义相对应, 例如, 在机器人直向运动时 D 子函数占主导地位, 响应的系数权重较大; 而在机器人转向时 W 、 V 子函数占主导地位, 对应系数取值权重较大, 通过不同的赋权比重使机器人运动更合理, 能够有效降低机器人目标识别产生的偏差。

1.3 数据采集预处理及模型更新

在可移动机器人对标的物进行目标跟随与识别过程中, 通过采集标的物的特征及融合自身的位置获取精确的数据之间关系。数据的采集需要重点关注同步性并通过空间位置信息的特征匹配进行数据集相关内容的配准与预处理, 这对于标的物的识别至关重要^[16]。数据预处理会结合采集数据的相似性度量方法确定目标位置, 且在数据配准中还能够融合视觉、雷达、超声波等多模态的数据^[17-19], 以提升目标识别的精确度及提升模型的鲁棒性。目标识别是一种综合识别方案, 尤其在较为复杂的场景中, 如低照度、有遮挡等, 通过多模态数据的采集与配准预处理可以显著提升图像的匹配精度, 有效避免了标的物漂移和误识别情况的发生, 为目标跟随的稳定性提供了可靠的支持。在配准环节本文采用主成分分析法对采集到的多模态数据进行线性变换 (当累计方差贡献率 $\geq 95\%$ 时, 确定主成分个数为 5), 并对高维数据进行降维处理 (将图像数据的维度从 5 维降至 3 维)。在空间范围内任何两组针对同一标的物的数据可

视为同一组基, 不同基的数据会有不同的特征表示, 而数据配准也是针对多模态数据的一种降维处理方案, 用协方差 $\text{cov}(A, B)$ 表示两组数据集 A 、 B 的相关性 (两种数据集是针对同一标的物不同模态下的采样数据):

$$\text{cov}(A, B) = \frac{1}{m} \sum_{i,j=1}^m (a_i - \bar{a})(b_j - \bar{b}) \quad (4)$$

式中, m 表示数据集的数据数量, $\bar{a}$ 和 $\bar{b}$ 分别表示数据集 A 、 B 中的数据均值。对两组数据集 A 、 B 进行配准后的得到的新数据集 C 表示如下:

$$C = \frac{1}{m} AB^T \quad (5)$$

式中, B^T 表示 B 的转置, 若令空间内任一点的坐标为 (X, Y, Z) , 多模态条件下的坐标数据配准过程如下所示:

$$C = \begin{bmatrix} \text{cov}(X, X) & \text{cov}(X, Y) & \text{cov}(X, Z) \\ \text{cov}(Y, X) & \text{cov}(Y, Y) & \text{cov}(Y, Z) \\ \text{cov}(Z, X) & \text{cov}(Z, Y) & \text{cov}(Z, Z) \end{bmatrix} \quad (6)$$

在目标跟踪与识别的过程中通过采集不同采样点的坐标进行相关性分析、配准, 并动态跟踪模型以更好地适应背景与环境的变化, 以提升机器人识别过程中的精确性和鲁棒性。多模态数据采样和配准后从当前帧中提取标的物的目标特征, 并结合动态窗口滑动与评价函数提升样本的可信度, 并增强模型的泛化能力。

1.4 多模态视觉注意力机制与目标识别

多模态目标检测是要将多种不同场景和不同模态采集到的数据, 表示为同一类语义并基于多个不同维度识别同一个标的物, 从中挖掘不同模态间的具有一定相关性的互补信息之间的内在联系, 使得单模态信息缺失的部分可以由其他模态来补充, 进而实现更精确的目标识别。多模态环境下, 可移动机器人采集的实时信息具有海量性特点, 且计算复杂度要求更高, 需要在较短时间内完成信息的处理与特征提取以便于实现机器人与后台上位机之间的交互。

1.4.1 视觉注意力机制骨干网络构建

自注意力机制最早用于自然语言的处理中, 自注意力机制是一种较为高效的深度学习工具, 但其计算量较大且模型的复杂度较高, 将注意力机制直接用于可移动机器人的目标识别会造成训练开销的暴增。针对该种情况本文的设计理念是借助 CNN 的基础架构, 在 CNN 结构中融入注意力机制模块 (将自注意力机制模块嵌入到 CNN 框架的卷积层), 实现两类深度模型的优化和互补, 模型的基础架构如图 3 所示。

多模态视觉注意力机制网络的输入数据先进入包含多个卷积层 (CONV, convolutional layer) 的卷积网络, 卷积层的数量选择视输入集的规模, 复杂程度, 模态数量等参数确定。CNN 网络的优势在下采样和特征

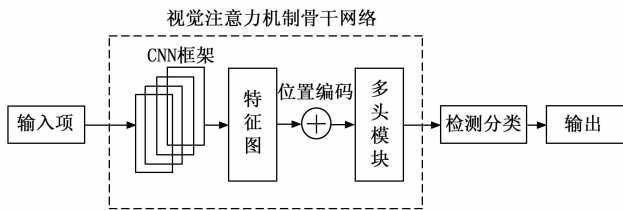


图 3 视觉自注意力机制的骨干网络构建

提取，能够降低对二维图像数据的处理难度并提升图像处理效率，可以有效减少注意力机制模块的计算量。CONV 卷积层卷积核选择 3×3 的结构（卷积核的尺寸与对应的数据维度 3 维相对应），卷积过程如图 4 所示。

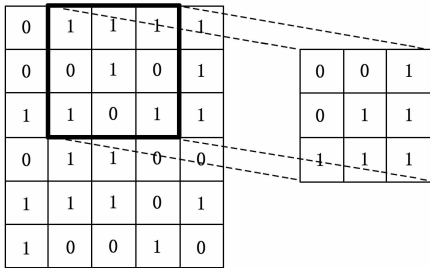


图 4 卷积层的卷积过程

CNN 网络的激活函数选择了具有非线性特征的带泄漏修正线性单元（L-ReLU, leaky rectified linear unit）函数，以更好地应对多模态环境下的数据训练特征，L-ReLU 激活函数图像如图 5 所示。

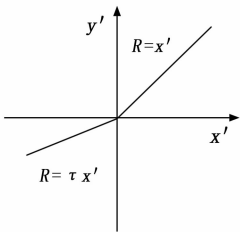


图 5 L-ReLU 激活函数的图像

L-ReLU 激活函数的数学表达式如下所示：

$$\begin{cases} R(x') = x', x' \geq 0 \\ R(x') = \tau x', x' < 0 \end{cases} \quad (7)$$

式中， τ 表示一个非常小的正数，用于调节在非线性条件下的激活函数斜率。CONV 卷积层也涵盖了池化功能，在该模块内对采集到的数据集进行归一化处理。

CNN 骨干网络全连接层也内嵌于 CNN 框架中，全连接层的层数设计取决于输入图像集的规模，如果当前层的神经元数量为 M ，前一层神经元的数量为 N ，则全连接层神经元的权重矩阵 W 的大小为 $M \times N$ ，表示为 w_{MN} 。用 x'_k 表示 CNN 骨干网络的全连接层第 k 个神经元的输入值，用 y'_k 表示对应的 CNN 网络卷积层和池化层的输出，针对多模态特征图的特征提取过程描述如下：

$$y'_k = R\left(\sum_{k=1}^N w_{MN} x'_k + b_k\right) \quad (8)$$

式中， R 表示 L-ReLU 激活函数， b_k 表示 L-ReLU 激活函数的偏置项。由于在骨干网络中增加了多头注意力机制模块，需要通过位置编码提升视觉识别的效果。这是因为多头注意力机制需要通过位置信息实现对序列的排序及对关键信息的位置编码，再利用多头注意力机制网络训练数据和提取关键特征，用于精确地目标识别。

1.4.2 多头注意力机制模块的优化与多模态融合改进

若 Q 表示为查询向量矩阵， K 为键向量矩阵， V 表示为值向量矩阵，用 W^Q 、 W^K 和 W^V 分别表示查询向量矩阵，在多头注意力机制模块中，注意力机制的头数由 K 为键向量矩阵， V 表示为值向量矩阵的维数及输入集的规模确定。键向量矩阵和值向量矩阵的映射矩阵，通过逐点卷积生成各向量矩阵的过程如下：

$$\begin{cases} Q = y(t) \cdot W^Q \\ K = y(t) \cdot W^K \\ V = y(t) \cdot W^V \end{cases} \quad (9)$$

注意力机制单元模块 A_w 由内容加位置、内容加内容等两个部分构成，如下所示：

$$A_w = qk^T + qr^T \quad (10)$$

式中， q 表示查询向量， k 表示键向量， r 表示位置向量。查询向量、键向量、位置向量等都可以视为一个独立的向量，并被赋予了多个头的子空间，在获取到单一的注意力机制模块后对其进行归一化处理，降低模型的复杂度并将各元素逐个相乘得到对应的输出，多头注意力机制的计算过程如下（归一化处理函数选择 softmax 函数）：

$$A_w(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{r}}\right)V \quad (11)$$

多头输入拼接后的线性组合与模型的最终输出保持一致，多头注意力机制的优势在能够捕捉多模态环境下目标物的不同特征，采用并行独立运算的方式也提升处理非线性、长序列数据或图像的效率。

1.4.3 双门控结构

包括注意力机制模型在内的深度神经网络在处理长序列数据时，不可避免地会出现梯度问题和依赖经验的情况，会影响到对目标物的识别，在多模态融合背景下计算过程更为复杂，为此本文设计了一种双门控用于解决建模依赖和模型梯度问题。本文设计的门控机制包括了重置门 g_1 和更新门 g_2 ，即在多头注意力机制模块中内置了门控机制，用于解决模型训练与目标识别中存在的各种问题：

$$g_1 = R\{w_1 \cdot [h_{t-1}, y(t)]\} \quad (12)$$

$$g_2 = R\{w_2 \cdot [h_{t-1}, y(t)]\} \quad (13)$$

式 (12) ~ (13) 中， R 表示激活函数， h_{t-1} 表示上一 t

-1 时刻注意力机制模型的状态， w_1 、 w_2 分别表示重置门和更新门的权重，初始状态下的权重比例分配采用随机正态分布方式确定。当前 t 时刻状态标识 h_t 如下：

$$h_t = [1 - y(t)] \odot h_{t-1} + y(t) \odot \tilde{h}_t \tag{14}$$

式中， $\tilde{h}_t$ 表示当前 t 时刻所期望的状态， $\odot$ 表示逐元素乘法：

$$\tilde{h}_t = \tanh\{w_h[g_2 \odot h_{t-1}, y(t)]\} \tag{15}$$

式中，选用的激活函数为双曲正切函数（ $\tanh$, hyperbolic tangent function）， w_h 是与当前状态 h_t 对应的权重比例。在引入门控机制后通过重置门 g_1 评估上一时刻状态对当前状态所产生的影响，通过更新 g_2 控制当前状态的融合，继而解决了对过去经验过度依赖的情况并形成最新的记忆单元，提升多头注意力机制的训练能力。基于多模态视觉注意力机制的目标识别总体流程，如图 6 所示。

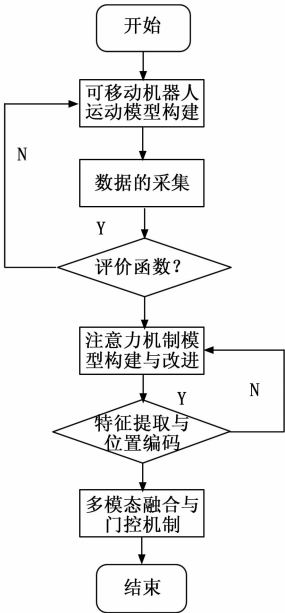


图 6 多模态视觉注意力机制的目标识别总体流程

- 可移动机器人目标识别的具体执行步骤如下：
- 1) 调试可移动机器人系统确保机器人硬件模块与软件流程有效，同步构建基于动态窗口的机器人运动模型。
 - 2) 基于多模态的视角针对同一目标采集不同维度的数据，并对数据做配准预处理，并通过评价函数评估数据集的质量。
 - 3) 构建一种多头注意力机制模型，并基于 CNN 网络框架提升对原始数据的处理能力并同步降低模型复杂度。
 - 4) 对多模态视觉多头注意力机制模型进行优化和改进，通过引入门控机制，以进一步提升注意力机制模型的二维图像数据训练能力，继而提高对标的物的识别精度和识别效果。

2 实验结果与分析

2.1 实验环境搭建与实验参数设定

实验在本市的某建筑工地进行，主要通过可移动机器人监控建筑人员的规范作业情况和安全生产情况（例如识别进入工地的人员是否佩戴安全帽，是否在安全通过通行等），当人员出现违规作业时机器人识别并发出警报提醒负责安全的人员予以纠正。实验选用的机器人为昊艾智能 3D 轮式巡检机器人，配备了 SICK 双目视觉智能监测系统、激光探测雷达和超声波雷达，能够基于多个维度获取到标的物的信息。

机器人的相关参数，机器人控制系统软硬件及算法参数设定如表 1 所示。

表 1 相关参数设定

序号	参数	数值值	序号	参数	数值值
1	重量	60 kg	9	最高主频	3.6 GHz
2	长	0.8 m	10	RAM	32 GB
3	宽	0.8 m	11	ROM	2 TB
4	高	1.3 m	12	神经网络层数	5 层
5	像素	1 080 p	13	卷积核	3 * 3
6	有效观测距离	200 m	14	w_1	0.1~0.6
7	有效探测距离	300 m	15	w_2	0.4~0.9
8	CPU	i7	16	最大迭代次数	200

有效观测时间为 12 小时（早 6 时~晚 18 时），引入本文算法与文献 [9]、文献 [10]、文献 [11] 等 3 种现有方法参与对比，其中文献 [9] 的双目镜头参数与本文方法下镜头保持一致，文献 [11] YOLO 相关神经网络参数设定也与本文方法一致，确保各算法在最优状态下和同一尺度下进行。选用同一型号的目标识别巡检机器人，在同一温度下进行实验，确保实验的可比性和公平性，以验证本文提出目标识别方法的优势。全部 4 台机器人采集到的无遮挡图像 1 552 张（训练集 1 164，测试集 388），有遮挡图像 756 张（训练集 567，测试集 189），其他相关的模态数据共 4 320 组（训练集 3 240，测试集 1 080，包括雷达数据、超声波数据等）。

2.2 实验指标的选择与实验结果分析

在实验过程中采用了高清摄像头和无人机高空采集数据，并将这些数据作为理论精确值，用于判断各算法下机器人对于操作人员违规行为的监测结果。实验在两种环境下进行（无遮挡和有遮挡（工地的遮挡物，如钢筋等）），实验测试选择的指标为精确率 P 和虚警率 FPR ，其中精确率 P （表示模型预测正样本能力，值越大性能越好），计算公式如下：

$$P = \frac{TP}{TP + FP} \tag{16}$$

虚警率 FPR （负样本被预测为正样本的比例，反映模型预测样本纯度的能力，值越小性能越好），计算

公式如下：

$$FPR = \frac{FP}{TN + FP}$$

(17)

式(16)～(17)中， TN 、 TP 、 FN 、 FP 分别表示真正、真负、假正和假负的样本类型。

2.2.1 无遮挡环境下的实验结果

将图像样本的训练集划分为3个子集，测试集不再划分，系统会自动区分有遮挡和无遮挡的图像并进行目标识别，无遮挡情况下的实验数据分析结果如表2～5所示。

表 2 无遮挡训练集的目标识别结果（精确率）

训练集的子集	目标识别方法/%			
	本文方法	文献[9]	文献[10]	文献[11]
1	99.1	97.5	97.0	95.7
2	99.2	96.3	95.9	95.2
3	99.1	97.1	96.8	96.7
均值	99.1	97.0	96.6	95.9

表 3 无遮挡测试集的目标识别结果（精确率）

测试集	目标识别方法/%			
	本文方法	文献[9]	文献[10]	文献[11]
数值	99.1	95.8	95.7	95.9

表 4 无遮挡训练集的目标识别结果（虚警率）

训练集的子集	目标识别方法/%			
	本文方法	文献[9]	文献[10]	文献[11]
1	0.4	2.4	3.6	3.2
2	0.7	2.6	3.0	3.3
3	0.5	3.1	2.1	1.9
均值	0.5	2.7	2.9	2.8

表 5 无遮挡测试集的目标识别结果（虚警率）

测试集	目标识别方法/%			
	本文方法	文献[9]	文献[10]	文献[11]
数值	0.4	3.1	3.0	3.2

无遮挡场景下，在训练集中本文算法下的精确率和虚警率分别为99.1%、0.5%，优于3种传统的目标识别算法，而在测试集的结果与训练集趋同，本文算法下的精确率和虚警率分别为99.1%、0.4%，由于本文算法采用动态窗口识别融合注意力机制网络，能够在机器人运动过程中相对准确地提取目标图像特征，实现对目标的精确识别，降低识别的虚警率并提升识别的准确率。

2.2.2 有遮挡环境下的实验结果

有遮挡环境下更注重算法对多模态数据融合处理能力并基于多个维度进行数据融合和推测，有遮挡情况下

各算法针对训练集和测试集的实验结果，如表6～9所示。

表 6 有遮挡训练集的目标识别结果（精确率）

训练集的子集	目标识别方法/%			
	本文方法	文献[9]	文献[10]	文献[11]
1	97.3	90.2	88.3	87.8
2	96.9	88.4	89.9	89.2
3	98.1	89.5	87.2	89.3
均值	97.4	89.4	88.5	88.8

表 7 有遮挡测试集的目标识别结果（精确率）

测试集	目标识别方法/%			
	本文方法	文献[9]	文献[10]	文献[11]
数值	98.0	87.2	86.3	87.7

表 8 有遮挡训练集的目标识别结果（虚警率）

训练集的子集	目标识别方法/%			
	本文方法	文献[9]	文献[10]	文献[11]
1	1.2	8.8	10.1	9.2
2	1.1	8.1	7.9	9.6
3	1.2	7.9	8.2	9.4
均值	1.2	8.3	8.7	9.4

表 9 有遮挡测试集的目标识别结果（虚警率）

测试集	目标识别方法/%			
	本文方法	文献[9]	文献[10]	文献[11]
数值	1.0	9.2	9.3	9.0

在有遮挡的场景下，遮挡物会对所有的算法产生不利影响，但本文提出目标识别算法在多模态数据补偿和融合数据训练方面有一定优势。在训练集中本文算法下的精确率和虚警率分别为97.4%、1.2%，优于3种传统的目标识别算法（传统方法的精确率均值已经低于90%），而在测试集的结果与训练集趋同，本文算法下的精确率和虚警率分别为98.0%、1.0%，也具有较为明显的比较优势。

2.3 消融实验

通过消融验证模型各步优化对模型总体性能的贡献率，本文对机器人目标识别算法的优化主要包括了动态窗口模型中评价函数的引用，CNN骨干网络与注意力机制模块的融合，门控循环机制的引入等。零未优化的模型为模型1，评价函数优化的模型为模型2，CNN框架下的模型为模型3，门控机制优化下的模型为模型4，针对无遮挡情况下目标识别精确率，无遮挡情况的目标识别虚警率，有遮挡情况的目标识别精确率，有遮挡情况的目标识别虚警率的消融实验结果，如图7和表10所示。

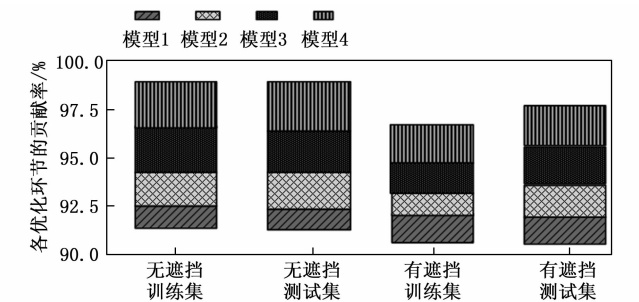


图7 各模型的精确率累计贡献率消融实验结果

从图7的消融实验结果可知模型3（CNN骨干网络架构）和模型4（门控机制）对于模型整体性能的贡献率较大。

表10 各模型的虚警率消融实验结果/%

指标	模型1	模型2	模型3	模型4
无遮挡训练集	7.3	5.5	2.6	0.5
无遮挡测试集	7.5	5.9	3.0	0.4
有遮挡训练集	8.7	6.8	4.1	1.2
有遮挡测试集	8.9	6.4	3.7	1.0

从表10的消融实验结果可以分析出，各模型的优化均对最终的结果产生了正向效应，但仍旧是模型3和模型4的贡献率相对较大。

3 结束语

机器人目标识别相对于传统人工、半人工的目标识别具有效率高、精度高、误警率低等诸多优势，尤其在较为复杂的场景下机器人目标识别的优势更加明显。针对现有目标识别方法存在的识别精确率低、虚警率偏高的不足，本文在多模态条件下设计了一种基于注意力机制网络模型的目标识别方法。在机器人运动学模型设计方面采用了一种动态窗口模式，并通过评价函数评估机器人的运动姿态。针对注意力机制网络模型算力与模型复杂度不匹配的问题，引入了CNN架构作为骨干网络内嵌了注意力机制模型；并通过引入门控机制进一步强化注意力机制模型的训练能力。实验结果显示，提出目标识别方案在有遮挡条件下测试集的精确率和虚警率分别为98%和1.0%，均显著低于传统的机器人目标识别方法。随着机器人技术的不断发展和成熟，其在目标识别中的应用越来越广泛，将深度学习算法融入机器人目标识别过程之中，加强了多模态环境下图像训练和处理能力。在未来，深度学习和深度神经网络算法在机器人目标识别领域将会发挥更大的作用。

参考文献：

[1] 杨 闰，李婧如，贾志昆，等. 基于可操作度的移动作业机器人多目标路径规划方法 [J]. 控制与决策，2024，39（10）：3243-3252.

[2] 孟广耀，王 哲，孙英暖，等. 轮腿可变式移动机器人越障研究 [J]. 计算机仿真，2023，40（3）：460-466.

[3] 陈艺文，江文松，杨 力，等. 基于运动约束的移动机器人路径规划 [J]. 计算机集成制造系统，2023，29（4）：1186-1193.

[4] 张 宇，李宇翔，张贺飞，等. 撼地者：基于遥操作与自主导航的移动救援机器人 [J]. 中国科学技术大学学报，2023，53（1）：32-43.

[5] 刘尚非，王军政，刘冬琛，等. 多功能并联式六轮足救援机器人 [J]. 机械工程学报，2024，60（21）：14-26.

[6] 孙鹏宇，李 君，刘宇飞，等. 基于车臂协同策略的移动机械臂柔顺开门技术 [J]. 机器人，2025：1-14.

[7] 王东署，高旭霖. 未知环境中移动机器人柔性行为决策的调节发育学习 [J]. 控制理论与应用，2023，40（4）：641-652.

[8] 王熙来，邓晓燕，郭晓婷. 基于改进卡尔曼滤波的移动机器人目标识别与定位研究 [J]. 机床与液压，2024，52（16）：26-31.

[9] 陈久朋，陈治帆，伞红军，等. 基于卷积神经网络的移动机器人自适应光照增强单目视觉SLAM算法 [J]. 农业机械学报，2024，55（12）：383-391.

[10] 黎晓昀，叶志清，周 慧. 自主移动机器人微型激光雷达点云数据滤波研究 [J]. 激光杂志，2023，44（6）：167-171.

[11] JEONG M, PARK S, KWON S M, et al. Rapid detection of soybean nutrient deficiencies with YOLOv8s for precision agriculture advancement [J]. Scientific Reports, 2025, 15（1）.

[12] 王欣宇，陈广锋，李 侠. 基于改进ViBe算法的运动目标检测 [J]. 东华大学学报：自然科学版，2023，49（1）：95-102.

[13] 李宗刚，李龙雄，杜亚江，等. 基于扰动观测器的铆接机器人轨迹跟踪控制 [J]. 振动与冲击，2023，42（14）：296-305.

[14] 易成群，吴佳晔，李嘉莉. 基于扰动环境下的轮式机器人路径跟踪控制 [J]. 组合机床与自动化加工技术，2023（4）：23-26.

[15] 李伟庆，王永娟，高云龙. 复杂环境下仿蛇机器人的路径规划策略 [J]. 兵器装备工程学报，2024，45（7）：28-37.

[16] 张 梅，傅艳霞，朱金辉，等. 基于改进水波算法的复杂多人共站装配线平衡研究 [J]. 计算机集成制造系统，2024，30（1）：129-143.

[17] 李佳益，马智亮，陈礼杰，等. 面向施工机器人定位的多模态数据融合方法研究综述 [J]. 计算机工程与应用，2024，60（15）：11-23.

[18] 邵 雷，石 峰. 生成式人工智能对社交机器人的影响与治理对策 [J]. 情报杂志，2024，43（7）：154-163.

[19] 王顺杰，蔡国永，吕光瑞. 对抗软对比调制动态图学习的多模态会话情绪识别方法 [J]. 计算机研究与发展，2025，62：1-19.