

基于视觉相似性度量的机器人目标识别研究

上官琚¹

SHANGGUAN Jun

摘要

针对已有方法存在的视觉相似性度量误差大、目标误识率高等问题,文章提出基于视觉相似性度量的机器人目标识别研究。应用 Transformer 及其核心组件多头注意力机制进行机器人视觉图像特征提取,融合余弦相似度与马氏距离构建双分支视觉相似性度量模型,获取视觉相似性度量结果,基于目标库中样本分布实际情况对其进行重标定,采用加权 K-NN 分类器匹配与识别机器人目标,从而实现研究目标。消融实验结果显示:所提方法的机器人目标识别准确率达到 99.56%,充分证实此方法具有一定的有效性与可行性。对比实验结果显示:应用后视觉相似性度量误差最大值仅为 1%,机器人目标误识率最小值达到了 4%。

关键词

机器人; 目标匹配与识别; 视觉图像; 视觉相似性度量; 特征提取; 消融实验

doi: 10.3969/j.issn.1672-9528.2026.06.013

0 引言

机器人技术是现代科技发展的重要里程碑,其发展历程为单一功能机械臂→智能体(具备自主决策能力),对于人类生产与生活方式有着巨大的影响^[1]。早期工业机器人由于固定场景、预设程序等因素的限制,只能完成重复性流水线作业,但是随着计算能力、传感器技术、人工智能算法等的不断发展及其突破,现代机器人已具备路径规划、环境感知、动态交互等多种能力,在多个重要领域展现出不可替代的价值,例如医疗领域、农业领域、物流领域等^[2]。已有方法均存在特征提取鲁棒性不足、计算效率与精度平衡问题,无法满足现今机器人的应用需求。其中,文献[3]针对无序目标与环境噪声导致的识别精度下降问题,提出一种融合强化学习的无序目标识别框架。通过双边滤波将目标图像分解为低照度与反射分量,分别采用压缩策略处理照度分量(抑制噪声)与增强策略处理反射分量(提升细节),合并后生成优化图像;将深度学习与强化学习结合,将增强图像输入深度强化学习模型,通过智能体与环境交互动态调整识别策略,实现无序目标的自适应识别。但是,在设计方法实际应用过程中,分解后的照度与反射分量可能存在信息损失,尤其在强噪声或动态光照下,合并图像易产生伪影,影响后续特征提取精度,进而影响目标识别的准确性;文献[4]针对服务机器人视觉抓取中密集遮挡导致的识别困难,提出基于 YOLOv7 的改进模型。构建深度过参数化高效聚合网络

(DOP-Net),通过多卷积核并行运算增强特征提取能力,聚焦目标未遮挡区域。多卷积核并行运算虽增强特征,但可能引入冗余计算,尤其在轻量化需求下,需权衡特征丰富度与计算成本,降低了设计方法的普遍适用性;文献[5]针对采摘机器人对水果目标的识别需求,设计双目视觉系统与 Softmax 分类模型。Softmax 分类器假设类别互斥,但水果目标可能存在重叠或相似类别,导致分类边界模糊。与此同时,立体匹配易受光照、遮挡干扰,深度信息误差可能影响目标定位,尤其在枝叶遮挡下,三维重建精度下降,最终导致水果目标识别性能不足。为了解决上述问题或缺陷,提出基于视觉相似性度量的机器人目标识别研究。

1 机器人目标识别方法设计

1.1 机器人视觉图像特征提取

在机器人目标识别任务中,已有方法主要采用传统卷积神经网络的权重共享机制与局部感受野提取机器人视觉图像特征,但其在全局信息整合、长距离依赖建模等方面存在着较大的局限性,致使视觉图像特征提取效果较差^[6]。设计方法应用 Transformer 及其核心组件多头注意力机制进行机器人视觉图像特征提取,其具有全局信息交互、动态权重分配等优势,能够更精准的提取机器人视觉图像特征,为后续研究奠定坚实基础。

Step1: 视觉图像分块与线性嵌入。将机器人视觉图像(二维)转换为序列化特征向量,适配 Transformer 输入格式。设定原始机器人视觉图像为 I , 将其划分为不重叠的 N 个小块,通过全连接层将每个块展平为向量 x_i , 用公式表示为:

1. 闽西职业技术学院 福建龙岩 364021

$$\mathbf{x}_i = \mathbf{W}_e \cdot \zeta(I_i) + \mathbf{b}_e \quad (1)$$

式中： $\mathbf{W}_e$ 与 $\mathbf{b}_e$ 代表权重矩阵与偏置向量， $\zeta(\cdot)$ 代表机器人视觉图像块展平操作函数， I_i 代表第 i 个机器人视觉图像块。

依据公式(1)即可获得适配Transformer输入格式的机器人视觉图像，记为 $X = \{\mathbf{x}_i, i=1, 2, \dots, N\}$ 。

Step2: 位置编码注入。为弥补Transformer对空间结构的天然缺失，为每个机器人视觉图像块分配可训练的位置向量 $\mathbf{p}_i$ ，将其与图像向量进行相加处理，即可获得含位置信息的视觉图像，用公式表示为：

$$\begin{cases} X' = \{\mathbf{x}'_i, i=1, 2, \dots, N\} \\ \mathbf{x}'_i = \mathbf{x}_i + \mathbf{p}_i \end{cases} \quad (2)$$

式中： X' 代表含位置信息的视觉图像， $\mathbf{x}'_i$ 代表第 i 个含位置信息的视觉图像向量。

Step3: 多头注意力机制计算。通过多组注意力头并行计算，捕捉视觉图像不同子空间的特征交互，获取视觉图像的多头注意力特征^[7]，用公式表示为：

$$\begin{cases} \mathbf{Z} = \delta(h_1, h_2, \dots, h_N) \cdot \mathbf{W}_o \\ h_i = \text{soft max} \left(\mathbf{x}'_i \cdot \frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}} \right) \mathbf{V} \end{cases} \quad (3)$$

式中： $\mathbf{Z}$ 代表视觉图像的多头注意力特征， $\delta(\cdot)$ 代表多头注意力输出结果拼接函数， h_i 代表第 i 头注意力的输出结果， $\mathbf{W}_o$ 代表输出投影矩阵， $\mathbf{Q}$ 、 $\mathbf{K}$ 与 $\mathbf{V}$ 代表注意力机制中的查询向量、键向量与值向量， d_k 代表键向量维度。

Step4: 残差连接与层归一化。为了缓解梯度消失，加速Transformer训练收敛，对多头注意力特征进行残差连接与层归一化处理，用公式表示为：

$$\begin{cases} \mathbf{Z}' = \mathbf{Z} + X' \\ \mathbf{Z}'' = \frac{\mathbf{Z}' - \mu}{\sigma} \odot \gamma + \beta \end{cases} \quad (4)$$

式中： $\mathbf{Z}'$ 代表多头注意力特征残差连接结果， $\mathbf{Z}''$ 代表多头注意力特征层归一化处理结果， μ 与 σ 代表多头注意力特征的均值和标准差， γ 与 β 代表可学习参数。

Step5: 视觉图像特征聚合与输出。通过全局平均池化对视觉图像特征 $\mathbf{Z}''$ 进行聚合与输出，用公式表示为：

$$\mathbf{F} = \text{AvgPool}(\mathbf{Z}'') \quad (5)$$

式中： $\mathbf{F}$ 代表机器人视觉图像特征提取结果， $\text{AvgPool}(\cdot)$ 代表全局平均池化函数。

1.2 视觉相似性度量模型构建

现有的方法主要采用单一方式进行视觉相似性度量，其难以应对复杂场景变化^[8]。因此，设计方法有效地融合余弦相似度与马氏距离，兼顾特征方向一致性与空间分布适配性，构建双分支视觉相似性度量模型。其中，余弦相似度主要是计算特征向量之间的夹角余弦值，消除模长影响，度量视觉

图像特征的方向相似性，用公式表示为：

$$S_{\cos}(\mathbf{F}, \mathbf{F}_i) = \frac{\mathbf{F} \cdot \mathbf{F}_i}{\|\mathbf{F}\|_2 \cdot \|\mathbf{F}_i\|_2} \quad (6)$$

式中： $S_{\cos}(\mathbf{F}, \mathbf{F}_i)$ 代表待检测目标视觉图像特征向量 $\mathbf{F}$ 与第 i 个已知目标视觉图像特征向量 $\mathbf{F}_i$ 之间的余弦相似度， $\|\cdot\|_2$ 代表 L_2 范数。

马氏距离主要是引入协方差矩阵表征特征向量的空间分布，度量视觉图像特征的空间相似性，用公式表示为：

$$D_{\text{Mah}}(\mathbf{F}, \mathbf{F}_i) = \sqrt{(\mathbf{F} - \mathbf{F}_i)^T \Sigma^{-1} (\mathbf{F} - \mathbf{F}_i)} \quad (7)$$

式中： $D_{\text{Mah}}(\mathbf{F}, \mathbf{F}_i)$ 代表待检测目标视觉图像特征向量 $\mathbf{F}$ 与第 i 个已知目标视觉图像特征向量 $\mathbf{F}_i$ 之间的马氏距离。为了统一两种度量方式的量纲与极性（余弦越大越相似，马氏距离越小越相似），对余弦相似度与马氏距离进行归一化处理^[9]，用公式表示为：

$$\begin{cases} \hat{S}_{\cos}(\mathbf{F}, \mathbf{F}_i) = \frac{S_{\cos}(\mathbf{F}, \mathbf{F}_i) + 1}{2} \\ \hat{S}_{\text{Mah}}(\mathbf{F}, \mathbf{F}_i) = e^{-\lambda \cdot D_{\text{Mah}}(\mathbf{F}, \mathbf{F}_i)} \end{cases} \quad (8)$$

式中： $\hat{S}_{\cos}(\mathbf{F}, \mathbf{F}_i)$ 与 $\hat{S}_{\text{Mah}}(\mathbf{F}, \mathbf{F}_i)$ 代表余弦相似度与马氏距离的归一化处理结果， λ 代表缩放因子，取值范围为 $[0, +\infty)$ 。根据特征熵动态分配权重系数，用公式表示为：

$$\begin{cases} \omega_{\cos} = \frac{1}{1 + e^{-H(\mathbf{F})}} \\ \omega_{\text{Mah}} = 1 - \omega_{\cos} \end{cases} \quad (9)$$

式中： $\omega_{\cos}$ 与 ω_{Mah} 代表余弦相似度与马氏距离的权重系数， $H(\mathbf{F})$ 代表特征熵。

基于式(8)与式(9)输出结果确定待检测目标视觉图像特征向量 $\mathbf{F}$ 与第 i 个已知目标视觉图像特征向量 $\mathbf{F}_i$ 之间的视觉相似性度量结果，用公式表示为：

$$S_{\text{fusion}} = \omega_{\cos} \cdot \hat{S}_{\cos}(\mathbf{F}, \mathbf{F}_i) + \omega_{\text{Mah}} \cdot \hat{S}_{\text{Mah}}(\mathbf{F}, \mathbf{F}_i) \quad (10)$$

式中： S_{fusion} 代表视觉相似性度量结果。

1.3 机器人目标匹配与识别

以1.2节计算的视觉相似性度量结果 S_{fusion} 为依据，针对已知目标库中样本分布不均衡问题，对其进行重标定，采用加权K-NN分类器匹配与识别机器人目标，为机器人应用提供有效的帮助。

常规情况下，已知目标库中样本分布很难均衡，故引入类别权重对视觉相似性度量结果 S_{fusion} 进行修正，用公式表示为：

$$\begin{cases} \hat{S}_{\text{fusion}} = S_{\text{fusion}} \cdot \omega_{y_i} \\ \omega_{y_i} = \frac{N_{\text{total}}}{|K| \cdot N_{y_i}} \end{cases} \quad (11)$$

式中： $\hat{S}_{\text{fusion}}$ 代表视觉相似性度量修正结果， ω_{y_i} 代表第 i 个样

本目标类别标签 y_i 对应的权重系数, N_{total} 代表总样本数量, K 代表总目标类别数量, N_{y_i} 代表第 i 个样本目标类别标签 y_i 对应的样本数量。

基于式 (11) 计算待识别目标视觉图像特征向量 F 与已知目标库中所有样本视觉图像特征向量 F_i 之间的视觉相似性度量结果, 按照从高到低的顺序对 $\hat{S}_{\text{fusion}}$ 进行排列, 提取前 k 个样本构成近邻集合 Ω_k , 采用加权 K-NN 分类器对 Ω_k 中样本类别标签 y_i 进行加权投票, 从而获得机器人目标匹配与识别结果^[10], 用公式表示为:

$$\begin{cases} P(y=c) = \frac{\sum_{i \in \Omega_k} \hat{S}_{\text{fusion}} \cdot \Pi(y_i=c)}{\sum_{i \in \Omega_k} \hat{S}_{\text{fusion}}} \\ \hat{y} = \arg \max [P(y=c)] \end{cases} \quad (12)$$

式中: $P(y=c)$ 代表样本类别标签 $y_i=c$ 的加权投票结果, $\Pi(\cdot)$ 代表指示函数。类别匹配时取 1, 否则取 0, $\hat{y}$ 代表机器人目标最终的匹配与识别结果。

综上所述, 在双分支视觉相似性度量模型的构建与应用基础上, 实现了机器人目标的精准匹配与识别, 为机器人未来发展提供一定的帮助。

2 实验

选取基于强化学习的机器人目标识别方法与基于改进 YOLOv7 的机器人目标识别方法作为对照方法, 简称为方法 1 与方法 2, 联合本文方法共同进行对比实验, 通过实验数据处理与计算, 分析本文方法的应用性能。

2.1 实验场景搭建

依据本文方法应用性能测试需求, 搭建实验场景, 如图 1 所示。

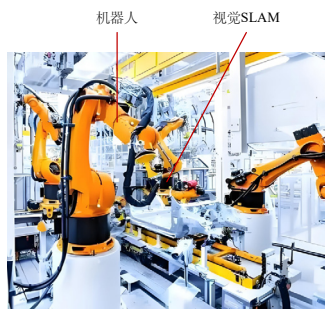


图 1 实验场景示例图

如图 1 所示, 在机器人末端搭载视觉 SLAM, 主要通过视觉传感器对未知环境中的目标进行感知与识别。视觉 SLAM 主要由两个摄像头、图像采集设备与计算设备构成。摄像头采用 Intel RealSense D435i, 帧率 30 帧/s; 图像采集器为 PCIe-6328 型; 显卡为 NVIDIA Jetson AGX Xavier, 运行内存 16 GB。

2.2 混淆矩阵

在实验对象 (机器人) 数据库中随机抽取 900 个实验样本 (每个目标类别实验样本数量为 100 个), 以此为基础, 对本文方法机器人目标识别性能进行测试, 并绘制机器人目标识别结果的混淆矩阵, 如图 2 所示。

1	100	0	0	0	0	0	0	0
2	0	99	0	0	0	0	0	1
3	0	0	100	0	0	0	0	0
4	0	0	0	100	0	0	0	0
5	0	0	0	0	100	0	0	0
6	0	0	2	0	0	98	0	0
7	0	0	0	0	0	0	100	0
8	0	0	0	0	0	0	0	100
9	0	0	0	1	0	0	0	99
	1	2	3	4	5	6	7	8

检测目标类别编号

真实目标类别编号

图 2 混淆矩阵示意图

如图 2 所示, 机器人目标类别主要为齿轮、轴承、障碍物、连接器、芯片、喷枪、焊枪、夹具与 PCB 板, 将其编号为 1~9。混淆矩阵中行代表真实目标类别编号, 列代表检测目标类别编号, 对角线单元格内部数字代表目标类别正确检测的样本数量, 非对角线单元格内部数字代表目标类别错误检测的样本数量。图 2 中目标类别 2、6 与 9 被错误检测的样本数量分别为 1、2 与 1, 其他目标类别的检测结果完全正确, 则本文方法的机器人目标识别准确率达到了 99.56%, 充分证实本文方法具有一定的有效性与可行性。

2.3 视觉相似性度量误差分析

应用本文方法、方法 1 与方法 2 度量视觉相似性, 计算其与实际视觉相似性之间的误差 (单位为 %), 如图 3 所示。

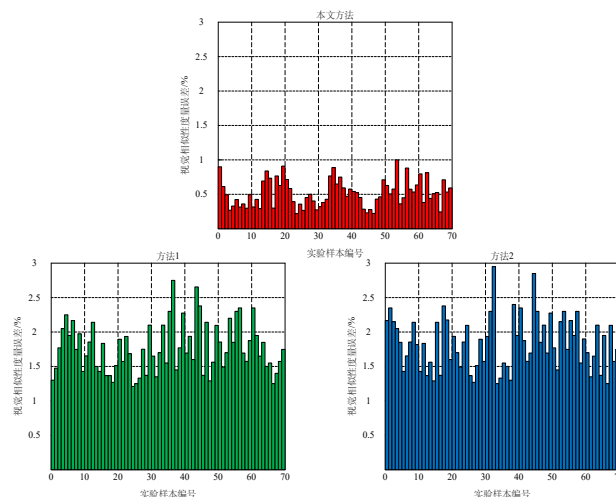


图 3 视觉相似性度量误差示意图

相较于方法1与方法2,本文方法应用后视觉相似性度量误差显著较小,其最大值仅为1%。因为本文方法以余弦相似度与马氏距离为基础,构建了双分支视觉相似性度量模型,能够同时兼顾特征方向一致性与空间分布适配性,从而精准地度量视觉图像特征向量之间的方向相似性与空间相似性,再根据特征熵动态分配权重系数,确定最终的视觉相似性度量结果,极大地缩短了视觉相似性度量结果与实际视觉相似性之间的距离,进而显著降低了本文方法的视觉相似性度量误差。

2.4 机器人目标误识率分析

应用本文方法、方法1与方法2识别机器人目标,计算误识率,如图4所示。

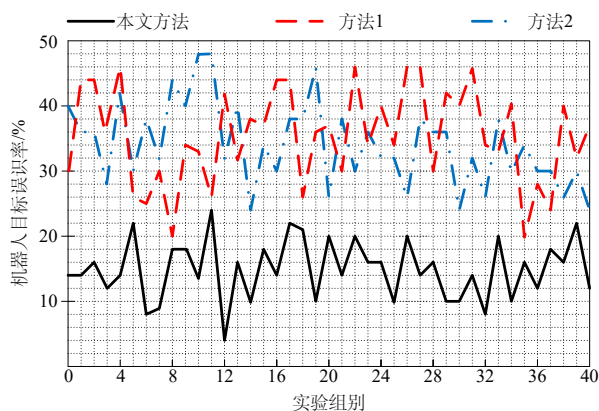


图4 机器人目标误识率示意图

本文方法应用后机器人目标误识率显著低于方法1与方法2,其最小值达到了4%。这主要是因为本文方法通过双分支视觉相似性度量模型精准地度量了视觉图像特征向量之间的视觉相似性,并根据目标库中样本分布情况对视觉相似性进行了重标定,进一步增加了视觉相似性度量的精准性。与此同时,以重标定视觉相似性为基础,采用加权K-NN分类器完成了机器人目标的匹配与识别,从而大幅度地降低了本文方法的机器人目标误识率。

3 结语

机器人目标识别是其实现其他功能的基础与前提,但是,在实际应用过程中,由于环境光照变化无法预测,其会导致目标表观特征失真,反光表面、动态阴影等的存在,还可能引发目标误检现象。与此同时,目标遮挡问题在复杂场景中尤为突出,部分可见的物体难以通过局部特征准确匹配,致使目标识别准确性无法得到保障。此外,机器人实际应用场景对于目标识别的实时性要求较高,这对应用方法复杂度与计算资源之间的平衡提出了较大的挑战。为了满足未来机器人的发展需求,提出基于视觉相似性度量的机器人目标识别研究。实验结果显示:设计方法有效地降低了视觉相似性度

量误差与机器人目标误识率,为相关研究提供一定的借鉴。

参考文献:

- [1] 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.
- [2] 王茂森,鲍久圣,鲍周洋,等.融合金字塔结构与注意力机制的煤矿井下巡检机器人PT目标检测算法[J].煤炭科学技术,2024,52(6):206-215.
- [3] 徐楚原,何健.基于强化学习的机器人无序目标智能识别仿真[J].计算机仿真,2023,40(10):440-444.
- [4] 陈仁祥,邱天然,杨黎霞,等.改进YOLOv7的服务机器人密集遮挡目标识别[J].光学精密工程,2024,32(10):1595-1605.
- [5] 冯治广,董佳佳,王茂英.基于Softmax的采摘机器人目标识别技术研究[J].农机化研究,2023,45(2):184-188.
- [6] 戴楚舒,张选德,熊静.基于规范化注意力机制的孪生单目标视觉追踪[J].陕西科技大学学报,2023,41(1):166-173.
- [7] 李兴,王志勇,王子豪,等.一种基于权重自适应相似性度量改进的InSAR-BM3D算法[J].遥感信息,2023,38(2):71-78.
- [8] 赵玉清,贾奥莹,王白娟,等.柑橘果实采摘机器人目标识别算法研究进展[J].江苏农业学报,2024,40(8):1552-1560.
- [9] 冯高峰.农业机器人采摘目标识别技术研究:基于FCM模糊聚类算法[J].农机化研究,2024,46(3):30-33.
- [10] 刘鹏,袁啸林,侯吉磊,等.真空检漏机器人目标识别技术研究[J].电子测量与仪器学报,2023,37(8):128-135.

【作者简介】

上官琨(1981—),女,福建龙岩人,本科,讲师,研究方向:计算机。

(收稿日期:2025-11-26 修回日期:2026-06-10)