

基于深度学习的摄像头与机械臂融合视觉感知与目标识别算法研究

熊少军

武汉商学院信息工程学院, 湖北 武汉

收稿日期: 2025年10月12日; 录用日期: 2025年11月12日; 发布日期: 2025年11月24日

摘 要

目前, 随着深度学习、机器视觉以及机械臂控制技术的迭代更新, 使用深度学习对目标物体进行检测、识别、控制以发展成为新兴研究领域。本文主要研究一种基于深度学习的摄像头与机械臂视觉感知与目标识别算法。通过研究深度学习算法和ROS2操作系统的基本原理和应用, 针对机械臂、摄像头视觉感知与目标识别典型应用场景, 提出了先进行相机标定, 保证摄像头采集图像的成像质量和精度, 采用YOLOv8目标检测算法实现对目标物体的准确识别, 最后使用机械臂PID控制算法实现机械臂的精确运动控制抓取、移动目标物。通过反复实验, 表明该算法能够实现对目标物体的快速精准识别, 并通过机械臂实现目标物体精确运动控制。

关键词

深度学习, 神经网络, 视觉感知, 目标检测, 机械臂

Research on the Fusion of Camera and Robot Arm Visual Perception and Object Recognition Algorithm Based on Deep Learning

Shaojun Xiong

College of Information Engineering, Wuhan Business University, Wuhan Hubei

Received: October 12, 2025; accepted: November 12, 2025; published: November 24, 2025

Abstract

Currently, with the iterative updates of deep learning, machine vision, and robotic arm control

文章引用: 熊少军. 基于深度学习的摄像头与机械臂融合视觉感知与目标识别算法研究[J]. 计算机科学与应用, 2025, 15(11): 156-161. DOI: 10.12677/csa.2025.1511293

technologies, the use of deep learning for detecting, recognizing, and controlling target objects has become an emerging research field. This article mainly studies a deep learning based visual perception and object recognition algorithm for cameras and robotic arms. By studying the basic principles and applications of deep learning algorithms and ROS2 operating system, this paper proposes to first calibrate the camera to ensure the imaging quality and accuracy of the captured images for typical application scenarios of robotic arms, camera visual perception, and object recognition. The YOLOv8 object detection algorithm is used to achieve accurate recognition of the target object, and finally the PID control algorithm of the robotic arm is used to achieve precise motion control of the robotic arm for grasping and moving the target object. Through repeated experiments, it has been shown that the algorithm can achieve fast and accurate recognition of target objects, and achieve precise motion control of target objects through a robotic arm.

Keywords

Deep Learning, Neural Networks, Visual Perception, Object Detection, Robotic Arm

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着物联网、图像处理、深度学习、人工智能等相关技术的飞速发展，智能化的传感器、摄像头等图像采集设备的功能更加丰富、集成度高、通信效率高、延迟低、成本低，应用场景十分广泛。随着深度学习的不断迭代更新，在目标检测与识别在计算机视觉领域的研究逐步深入[1]。卷积神经网络(Convolutional Neural Networks, CNN)适用于图像处理、目标检测，卷积神经网络是一种前馈神经网络，适用于训练多层神经网络的学习算法模型[2]。

随着深度学习的深入研究，尤其是卷积神经网络(CNN)的提出，带来了更为准确和高效的目标检测与识别算法。但在实际应用过程中，如何将这些深度学习算法有效地集成融合到机械臂视觉系统中，实现快速且准确的目标识别是目前面临的一个挑战[3]。

本文通过研究一种基于深度学习的摄像头与机械臂视觉感知与目标识别算法，通过使用 YOLOv8 结合 ROS2 系统设计系统化算法，使机械臂融合感知环境能力，提升在复杂工程场景中的自主识别与操作物体能力。本文研究有助于推动视觉感知与机械臂技术的深度融合，提高智能化生产效率具有一定的实践价值和社会意义。

2. 国内外研究现状

随着深度学习与 ROS 系统的迅速发展，深度学习技术在摄像头图像采集与机械臂的视觉感知和目标识别领域应用广泛。国外的一些研究机构在机械臂与计算机视觉中的实际场景应用中，是通过卷积神经网络(CNN)来实现物体识别与自动化控制。如何在保证检测精度的同时提高算法的运行效率，如何处理遮挡和相似目标带来的检测困难，以及如何适应不同光照条件和复杂背景下的目标检测等问题，都需要未来的研究去解决。此外，随着深度学习技术的不断发展，新的算法和模型不断涌现，伴随着生成式人工智能与大模型的迭代更新，当下目标分类、目标检测、图像检索等计算机视觉的研究也在快速发展。在国内，越来越多的高校和科研机构开始关注并投入到运动目标检测与识别的研究中，有学者使用深度学习与神经网络等相关知识来实现振动平台上的目标检测任务[4]。有学者使用引入了改进的卷积神经网络

算法以提高电力线路小目标物料识别的准确率[5]。有学者研究并改进运动目标检测方法,进行移动阴影算法的研究与改进[6]。有学者利用卷积神经网络的池化层对目标特征的损失函数,处理目标特征模糊的情况,以此提高检测的精度[7]。

通过研究深度学习在智能制造中的应用,特别是结合 ROS2 系统,基于 YOLO (You Only Look Once) 系列的目标检测模型,通过改进摄像头标定技术和 PID 控制策略可实现更加精准和稳定的视觉识别场景中的机械臂操作[8]。

目前国内外通过的摄像头或深度相机与机械臂融合,通过将深度学习算法与 ROS2 操作系统的进行研究及工程实践,可以有效提升机械臂执行目标指令的准确性,推动智能机器人技术的发展[9]。

3. 深度学习算法和 ROS2 操作系统

3.1. 深度学习算法

深度学习算法核心是构建一个由多个隐藏层组成的神经网络模型。训练过程中,算法通过反向传播算法调整这些权重和偏置,以最小化预测输出与实际标签之间的误差。具体公式(1)如下所示。

$$L = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (1)$$

其中, N 是样本数量, y_i 是第 i 个样本的真实标签, $\hat{y}_i$ 是模型预测的标签。本文使用 PyTorch 框架实现全连接神经网络的 Python 代码。

3.2. 卷积神经网络

一个卷积层通过对输入向量,设计可学习的滤波器(卷积核),通过训练学习生成特征图。输入一般是二维向量,如高度、图像。在图形图像处理方面,卷积神经网络具有适配度,特征提取和模式分类同时进行,网络具有较强适应性。

假设输入数据是一个二维向量矩阵 X , 其大小为 $m \times n$, 表示图像的高度和宽度。一个卷积核 K 的大小为 $F \times F$, 其参数包括权重和偏置项。公式(2)所示卷积操作。

$$Y_{i,j} = \sum_m \sum_n X_{m,n} K_{i+m,j+n} + b \quad (2)$$

其中, $Y_{i,j}$ 是输出特征图的第 i, j 个元素, b 表示偏置项, $X_{m,n}$ 表示输入数据矩阵 X 中的第 m, n 个元素, $K_{i+m,j+n}$ 是卷积核的第 $i+m, j+n$ 个元素。

目前,主要是在连续的卷积层之间使用激活函数,为了减少模型参数的数量并防止过早拟合,需要进行池化层的相关操作。

3.3. ROS2 操作系统

ROS2 (Robot Operating System 2) 是机器人、机械臂行业广泛使用的开源操作系统框架,通过对节点库函数 API 进行设计、编码和测试用于操作控制机械臂。机械臂包含电机、减速器、控制器等核心组件,可通过 Python 或 C++ 编程语言进行 ROS2 编程。相比于 ROS 版本,ROS2 版本在性能、可扩展性和安全性方面显著提升,适合于实时系统和分布式计算环境。使用 ROS2 和 Gazebo 搭建机械臂虚拟仿真实验环境,进行可视化工作空间和路径轨迹规划。通过 ROS 的节点编程控制和深度学习算法实现在指定区域内的机械臂对目标物体的抓取控制操作。在深度学习算法与机械臂视觉融合感知的研究过程中,实现摄像头的图像数据采集、处理以及与机械臂控制系统的数据交互与控制。在不同硬件配置下 ROS2 节点响应时长差异,硬件配置的优劣对 ROS2 节点的响应时间有显著影响。高性能的 CPU 及 RAM 可以显著提高 ROS2 运行

效率。

4. 机械臂摄像头视觉感知与目标识别

4.1. 相机标定

相机标定是视觉感知系统中的一个关键步骤。通过拍摄多个不同角度和位置的棋盘格图像，收集到足够样本数据来设定相机的内参和外参。相机的内参矩阵 K 可以表示为：

$$K = \begin{bmatrix} f_x & 0 & c_x & 0 & f_y & c_y & 0 & 0 & 1 \end{bmatrix} \quad (3)$$

其中， f_x 和 f_y 分别是 x 轴和 y 轴方向的焦距， c_x 和 c_y 是主点坐标。为了实现相机标定，使用 OpenCV 库中的函数。通过循环捕获视频流中的每一帧，并查找棋盘格的角点，使用 “cv2.calibrateCamera” 函数进行相机标定，输出相机的内参矩阵和畸变系数。

4.2. 机械臂 PID 控制

在机械臂的视觉感知与目标识别中，实现精确的位置控制是至关重要。PID (比例 - 积分 - 微分) 控制器因其简单有效而被广泛应用于机械臂的运动控制中。PID 控制器通过调整三个参数：比例系数 K_p 、积分系数 K_i 和微分系数 K_d 来优化控制性能。

PID 控制器的基本公式为：

$$u(t) = K_p e(t) + K_i \int_0^t e(\tau) d\tau + K_d \frac{de(t)}{dt} \quad (4)$$

其中， $e(t)$ 是设定点与实际位置之间的误差， $u(t)$ 是控制器输出，用于驱动机械臂达到期望位置。PID 参数的调整对控制效果有显著影响，可通过试错法来优化输入参数。以下是使用 Python 语言实现 PID 控制器实现代码。在机械臂的视觉感知与目标识别系统中，PID 控制器不仅需要处理机械臂的运动控制，还要考虑到摄像头的实时反馈。因此，PID 控制器的设计和参数调整需要根据具体的机械臂结构和实际工作场景进行优化调整。

4.3. 基于 YOLOv8 的目标检测

本文在机械臂视觉感知与目标识别的研究中使用 YOLO 算法。YOLOv8 采用单次前向传播完成目标检测任务。YOLOv8 的损失函数由三个部分组成：边界框坐标损失、置信度损失和类别损失。具体公式如下：

$$L_{total} = \lambda_{coord} L_{coord} + \lambda_{obj} L_{obj} + \lambda_{cls} L_{cls} \quad (5)$$

其中， L_{coord} 是边界框坐标的损失， L_{obj} 是置信度损失， L_{cls} 是类别损失， λ_{coord} , λ_{obj} , λ_{cls} 分别是这三个部分的权重系数。

5. 实验结果与分析

5.1. 数据集

为保证实验的可重复性与结果的可信度，本研究使用了自建数据集与公开基准相结合的策略。自建数据集和采集流程如下：

(1) 采集环境与设备：采集在实验室内完成，室内光照以日光白光为主(色温约 5500 K)，并在若干实验组中加入强光与弱光条件以考察模型鲁棒性。图像采集设备为工业相机(1920 × 1080)与 Intel RealSense D435 深度相机。机械臂型号为 XX-6DOF，控制通过 ROS2 节点与真实关节驱动器通讯。

- (2) 样本数量与类别：训练集约 4000 张 RGB 图像，验证集 800 张，测试集 1000 张。端到端物理测试共 1000 次抓取实验。
- (3) 标注方法：目标检测标注采用 COCO 风格边界框与类别标签，人工双重标注并经一致性检查。定位精度实验中使用棋盘格标定并记录真实目标质心位置以计算误差。
- (4) 数据增强与预处理：随机缩放、翻转、颜色扰动与裁剪，统一尺寸 640×640 并标准化。

5.2. 实验结果与分析

YOLOv8 采用单次前向传播完成目标检测任务。对于每个边界框，YOLOv8 定义了 5 个参数： x, y, w, h 和 $conf$ 。其中， x, y 表示边界框中心相对于网格左上角的坐标， w, h 表示边界框的宽度和高度，而 $conf$ 表示该边界框包含目标的概率以及背景的概率。

本研究基于 YOLOv8 的目标检测算法对机械臂摄像头图像采集进行目标检测与识别分析，通过调整参数和优化器，可以进一步提高模型的性能。硬件实验环境配置为 32 GB RAM 和 NVIDIA RTX 3090 GPU 的服务器工作站，通过对其他算法进行对比分析评估模型性能。

Table 1. Analysis of experimental results
表 1. 实验结果分析

序号	算法	成功率	定位 RMSE (mm)	IoU	总耗时(s)
1	YOLOv8	92%	6.2 ± 2.1	0.85	2.8
2	YOLOv5	88%	7.5 ± 2.8	0.80	3.1
3	SSD	85%	9.0 ± 3.4	0.78	3.4

通过表 1 实验结果表明，基于 YOLOv8 设计的算法在目标检测任务上具有较高的效率和准确性，且端到端系统的抓取成功率超过 90%，重复定位精度约 6.2 mm 平均任务耗时约小于 3 s，证明系统在实时抓取任务中的可行性与鲁棒性。

6. 总结

本文研究一种基于深度学习的机械臂视觉感知与目标识别算法，该算法结合 YOLOv8 与 ROS2 进行算法设计，实现对指定区域内的不同大小、形状和颜色的目标物体进行高效识别、选择及自动化控制。本文通过对深度学习算法和 ROS2 操作系统的研究，将卷积神经网络应用于机械臂的摄像头视觉感知与目标识别领域，实现对复杂环境下的目标准确识别、定位及控制等自动化任务操作。通过相关实验，实验结果表明该算法具有较高的稳定性、准确性和抗干扰性，验证了研究方法的有效性。未来的研究可以探索更多优化策略，以根据实际应用场景构建相匹配的目标检测与识别算法。深度学习算法具有较高的计算复杂性，而机械臂控制系统计算资源是有限的，因此还需要进一步进行算法模型优化和硬件配置进行优化。

参考文献

[1] 朱龙闯. 一种基于深度学习的图像盲去运动模糊算法[J]. 现代计算机, 2021(4): 69-73.

[2] 卢宏涛, 张秦川. 深度卷积神经网络在计算机视觉中的应用研究综述[J]. 数据采集与处理, 2016, 31(1): 1-17.

[3] 王凯军. 基于深度强化学习的机械臂视觉伺服抓取[D]: [硕士学位论文]. 杭州: 浙江大学, 2022.

[4] 周睿豪. 微小运动目标智能检测技术研究[D]: [硕士学位论文]. 成都: 电子科技大学, 2023.

-
- [5] 杨亚萍, 刘军, 王洪亮. 基于改进卷积神经网络的智能物料识别技术研究[J]. 电子设计工程, 2024, 32(8): 191-195.
 - [6] 陈嵘. 智能监控系统中运动目标检测算法的研究与实现[D]: [硕士学位论文]. 湘潭: 湘潭大学, 2018.
 - [7] 梁丽香, 王贵文, 张翠翠. 基于深度学习的网络图像目标检测方法[J]. 信息与电脑(理论版), 2022, 34(9): 56-58.
 - [8] 陈璟. 基于机器视觉的机械臂目标稳准抓取研究[J]. 兰州职业技术学院学报, 2023(4): 70-72+86.
 - [9] 李龙岩. 基于视觉的机械臂智能抓取技术研究[D]: [硕士学位论文]. 郑州: 华北水利水电大学, 2023.