

基于卷积神经网络计算机视觉的物体识别设备

欧阳天成；李泽宇；杜忻怡

(吉林大学 仪器科学与电气工程学院, 长春 130022)

摘要：近年来，深度学习已成为机器学习和人工智能领域的热门研究方向，以其强大的特征提取能力和高精度的识别性能脱颖而出。作为一个快速发展且充满活力的领域，深度学习在过去的十余年中吸引了越来越多的研究者关注。特别是在卷积神经网络（CNN）方面，其在人工智能的多个领域中展现出了卓越的性能和广泛的应用前景。本文提出了一种基于卷积神经网络的物体识别系统，旨在通过运用计算机视觉、传感器、通信技术等实现对视觉传感器捕获数据的高效处理和分析。该系统采用卷积神经网络进行物体识别，使用 YOLO 算法对视频流进行逐帧处理，并将置信度高的结果显示在屏幕上对应物品的上方。该系统具有较高的检测精度和较快的检测速度，可广泛应用于物联网下的智能化识别系统。

关键词：深度学习； 物体识别； 卷积神经网络

Object Recognition Device Based on Convolutional Neural Network and Computer Vision

OUYANG Tiancheng; LI Zeyu; DU Xinyi

(College of materials Science and Engineering, Jilin University, Changchun 130022, China)

Abstract: Deep learning is the latest trend in machine learning and artificial intelligence research, which has attracted more and more attention from researchers because of its strong ability to extract features and high recognition accuracy. Convolutional neural networks (CNN) have demonstrated superior performance in several areas of artificial intelligence [1]. This paper proposes an object recognition system based on convolutional neural networks, aiming to efficiently process and analyze data captured by visual sensors through the use of computer vision, sensors, communication technology, etc. The system uses convolutional neural networks for object recognition and processes video streams frame by frame using the YOLO algorithm, displaying the results with high confidence on the screen above the corresponding items. The system has high detection accuracy and fast detection speed and can be widely applied to intelligent recognition systems under the Internet of Things.

Key words: Deep learning; Object recognition; Convolutional neural networks

0 前言

深度学习是机器学习和人工智能研究的最新趋势。由于它提取特征能力强、识别精度高等优点，正在受到越来越多研究者的关注^[1]。卷积神经网络（CNN）模型是深度学习模型中最重要的一种经典结构，它可以自动学习样本数据的特征表示，因此已经广泛应用于各个领域，在物体识别领域效果尤其显著^[2]。

各个应用领域将不同 CNN 作为特征提取的

主干网络添加不同的功能单元结构形成了新的应用模型。Vaillant 等人于 1994 年提出将 CNN 应用于人脸检测，训练模型检测图像中是否有人脸，并在所有可能位置扫描整个图像^{[3][4][5][6][7]}。

本项目旨在通过运用卷积神经网络、计算机视觉、传感器、通信技术、无线传输技术等多种学科技术，设计出一款物联网下的智能化识别系统，实现对视觉传感器捕获（图像、视频）数据的高效处理和分析。操作人员可以便携地手持终端分析眼前画面，进行对象识别，辨识物品的名称。

1 总体设计方案

本设计的研究内容包括深度学习物品识别原理、CSI 摄像头的驱动与调用、I2C 通信以及 3.5 寸屏显示、识别系统的软硬件设计，总体的设计方案如图 1 所示：

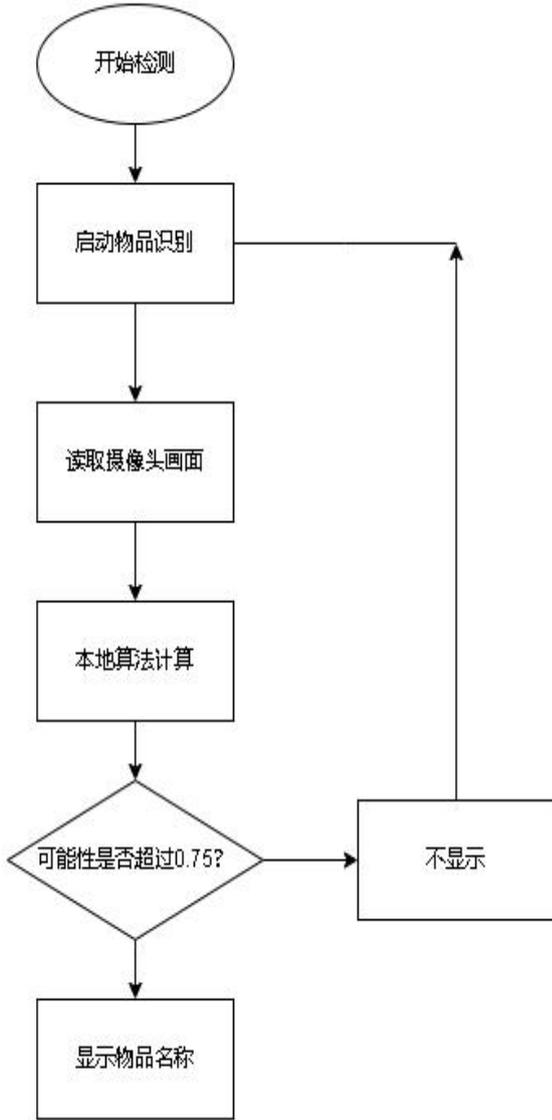


图 1 总体方案设计图

Fig. 1 Overall scheme design

本设计需要实现的目标如下：

- (1) 将 CSI 摄像头的视频流捕捉并显示在屏幕上；
- (2) 采用 YOLO V8 算法，对视频流进行逐帧处理并将结果输出；
- (3) 将置信度高的结果显示在屏幕画面中对应物品的上方。

2 系统实现的关键技术及难点

2.1 使用 CMAKE 手动编译 OPENCV

OpenCV，全称为 Open Source Computer Vision Library，是一个开源的计算机视觉和机器学习软件库，其广泛的应用范围涵盖了图像处理 and 计算机视觉的多个任务。这个库提供了丰富的函数和算法，以满足各种复杂和精细的视觉处理需求，包括图像滤波、特征检测、目标跟踪、图像分割、3D 重建等，支持多种编程语言接口，如 C++、Python、Java 等。OpenCV 在图像处理、机器学习、人机交互、机器人导航等领域有广泛的应用，是计算机视觉领域的重要工具之一^[8]。

在 raspberrypiOS(bullseye) 上，并没有合适的 opencv 预编译包，所以需要手动编译 opencv，如图 2 所示。安装好 CMAKE 后从中科大镜像源手动拉取部署。

```

jlu@jlu:~/Desktop/yolo $ sudo pip install ultralytics -i https://pypi.mirrors.ustc.edu.cn/simple
Looking in indexes: https://pypi.mirrors.ustc.edu.cn/simple
Collecting ultralytics
  Using cached https://mirrors.bfsu.edu.cn/pypi/web/packages/57/e4/762a11cb12f751eda6fa6716a97c06
Requirement already satisfied: matplotlib<=3.3.0 in /usr/lib/python3/dist-packages (from ultralytic
Collecting numpy>=1.22.2 (from ultralytics)
  Using cached https://mirrors.bfsu.edu.cn/pypi/web/packages/65/6e/09db70a523a96d25e115e71cc56a6f96
Installing build dependencies ... done
Getting requirements to build wheel ... done
Installing backend dependencies ... done
Preparing metadata (pyproject.toml) ... done
Collecting opencv-python<=4.6.0 (from ultralytics)
  Using cached https://mirrors.bfsu.edu.cn/pypi/web/packages/25/72/da7c69a3542071bf1e8f65336721b8b2
Installing build dependencies ...
  
```

图 2 编译过程

Fig. 2 Compilation process

2.2 使用深度学习算法 YOLO 对视频流进行帧处理

(1) 进行滤波

中值滤波是一种非线性信号处理技术，它通过用像素或其邻域内像素值的中值来代替该像素的值，从而消除噪声或平滑图像。这种方法特别适用于消除“椒盐”噪声，即随机出现的亮或暗像素。中值滤波的效果通常比线性滤波（如均值滤波）更为自然，因为它不会过度平滑图像的边缘和细节。

从 CSI 摄像头读取到待识别图片（视频帧率）后，先对图片进行降噪处理，使用中值滤波去除噪点，如图 3 所示。

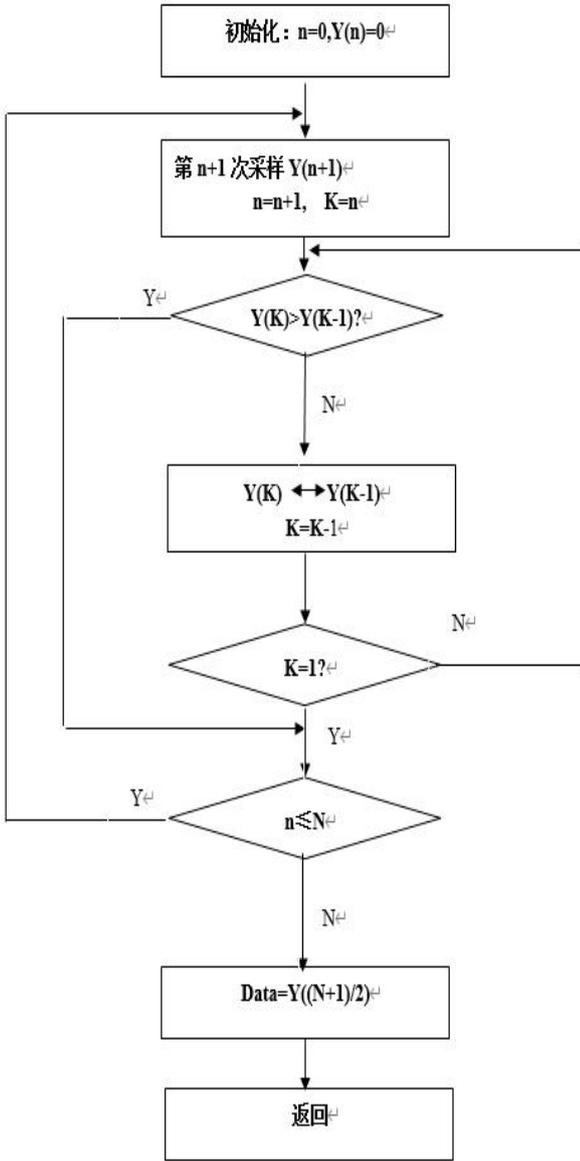


图 3 中值滤波流程图

Fig. 3 Median filter flow chart

(2) 特征提取

特征提取是图像处理和分析中的核心步骤，它旨在从图像中提取出有意义的信息，以便进行后续的认识、分类或理解。在具体应用中，我们可能关注的是图像中的特定特征，如边缘、角点或纹理。在物体识别任务中，我们可能关注物体的轮廓和纹理特征。

为了实现这些特征的提取，我们使用 Canny 边缘检测算法，通过计算图像中每个像素点的梯度强度和方向来检测边缘。除此之外，还有 SIFT 和 SURF 等算法则用于提取图像中的关键点和生成相应的特征描述子，这些描述子对于图像旋转、缩放和光照变化等具有一定的鲁棒性（健壮性）。

在平滑处理之后，我们利用这些算法对图像进行特征提取，随后使用 YOLO 算法进行处理。

3 测试结果

3.1 设备整体结构

基于树莓派控制器的物品识别终端整体实物图如图 4 和图 5 所示。此设备由供电系统、控制系统、摄像头、通用串行总线人机交互控制器、保护外壳等组成。



图 4 设备正面图

Fig. 4 Front view of equipment



图 5 设备背面图

Fig. 5 Equipment back drawing

3.2 功能测试

基于树莓派控制器的物品识别终端最终实现的功能有：调用 CSI 摄像头并将 CSI 摄像头的视频

流捕捉并显示在屏幕上；采用 YOLO V8 算法，对视频流进行逐帧处理并将结果输出；将置信度高的结果显示在屏幕画面中对应物品的上方。

(1) 摄像头测试：

通过远程 ssh 连接到树莓派，运行编写的 sh 脚本自动补全需要环境，然后调用摄像头测试画面，可以看到成功调用了摄像头，并且可以输出画面相关信息，如图 6 所示。

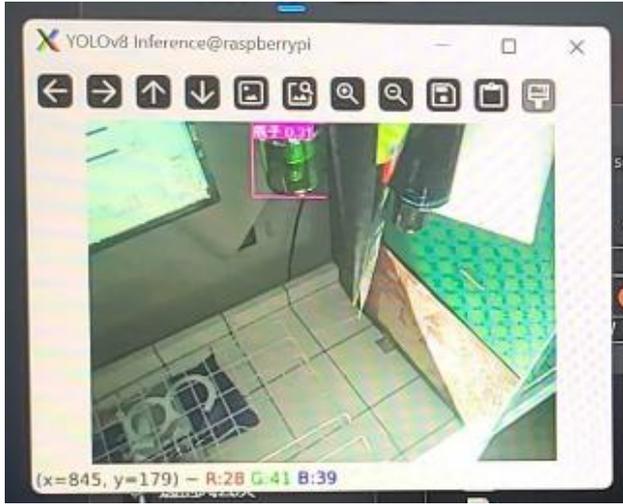


图 6 摄像头测试图

Fig. 6 Camera test diagram

(2) YOLO 算法测试（预训练模型测试）：

将 videocapture(0)改成 videocapture(tk.mp4)对提前准备好的卡车视频进行测试。如图 7 所示，识别延迟在 24-32 毫秒之间，符合期望。

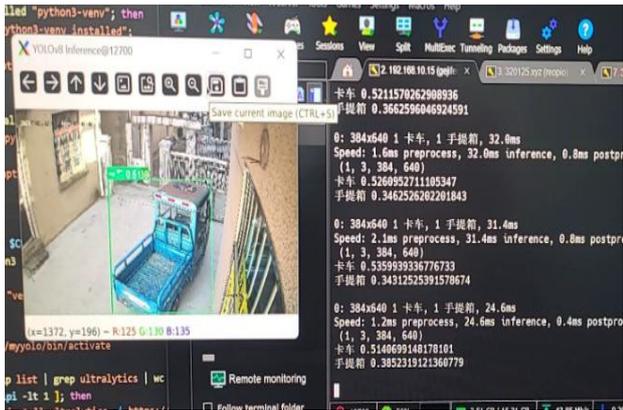


图 7 预训练模型测试

Fig. 7 Pre-trained model testing

(3) YOLO 识别摄像头视频流测试：

调用摄像头进行识别，在物体距离相对远的情况下置信度可以达到 0.76，延迟也在 700-800 毫秒左右，可以看到调用摄像头比处理视频延迟要高上一些，但是识别延迟总体依然远小于目标（3s），置信度达到了要求，并且成功在屏幕上显示了识别物体的位置信息与中文名，如图 8 所示。



图 8 调用摄像头识别测试

Fig. 8 Call the camera recognition test

4 结论

通过使用 YOLO V8 的调用本地预训练模型，实现了较短延迟、较高可信度的对 CSI 摄像头的视频流进行逐帧识别，并可以在屏幕上对应物品上方显示其中文名称与具体置信度。

参考文献

- [1] 严春满,王铖.卷积神经网络模型发展及应用[J].计算机科学与探索,2020,(15):27-46.
- [2] 王禹嫣.深度学习在图像识别中的应用与发展[J]中国战略新兴产业,2017,(48):115-116.
- [3] 兰宁.基于 Android 平台的图像识别设计方法与实现[J]电子技术与软件工程,2021,(19):61-64.
- [4] 李超波,李洪均,徐晨.深度学习在图像识别中的应用[J].南通大学学报(自然科学版),2018,(17):1-9.
- [5] PS Janardhanan. Project repositories for machine learning with TensorFlow[J]. ProcediaComputer Science, 2020, 171(C): 188-196.
- [6] Wang Yufei and Qi Chengchao and Du Qinglei. Target recognition of heterogeneous fusionimages based on improved YOLO model deep learning algorithm[J]. Journal of Physics: ConferenceSeries, 2022, 2170(1)
- [7] Ji Chen. The Absence and Reconstruction of Game Spirit in Early Childhood Education[J].International Journal of Higher Education Teaching Theory, 2021, 2(2)
- [8] 马祖鑫,崔允贺,秦永彬等.融合深度强化学习的卷积神经网络联合压缩方法[J/OL]. 计算机工程与应用 :1-11[2024-03-01].http://kns.cnki.net/kcms/detail/11.2127.TP.20240228.1635.025.html.