

文章编号: 2095-2163(2023)11-0077-05

中图分类号: TP242.6

文献标志码: A

深度视觉手势识别协作机械臂物体抓取系统

黄旭阳¹, 韩建海^{1,2}, 李向攀^{1,2}

(1 河南科技大学 机电工程学院, 河南 洛阳 471023; 2 河南科技大学 河南省机器人与智能系统重点实验室, 河南 洛阳 471023)

摘要: 为了提高手势识别的准确率以及机械臂抓取过程中控制手爪的夹持方向,实现对机械臂的灵活控制,研发基于深度视觉的手势识别模块。本文采用深度学习模型 YOLOv4 对手势识别进行研究。针对 YOLOv4 模型存在的输出内容单一,无法控制手爪抓取过程中夹持方向的问题,对 YOLOv4 算法的模型进行改进,将 3 个特征层的输出 $(32, 24, 75)$, $(16, 12, 75)$, $(8, 6, 75)$ 增加到 $(32, 24, 87)$, $(16, 12, 87)$, $(8, 6, 87)$, 增加 2 个手势关键点信息 (x, y) 的输出,通过这 2 个手势关键点的位置信息判断手势的水平方向,实现抓取过程中控制手爪的朝向,从而实现机械臂的灵活控制。实验表明,通过改进 YOLOv4 算法可以实现实时手势图像的精确识别以及机械臂的灵活控制。

关键词: 手势识别; 深度学习; YOLOv4

Cooperative robot arm object grab system of depth visual gesture recognition

HUANG Xuyang^{1,2}, HAN Jianhai^{1,2}, LI Xiangpan^{1,2}

(1 College of Mechanical and Electrical Engineering, Henan University of Science and Technology, Luoyang Henan 471023, China;

2 Henan Key Laboratory of Robot and Intelligent Systems, Henan University of Science and Technology, Luoyang Henan 471023, China)

Abstract: In order to improve the accuracy of gesture recognition and control the grasping direction of the gripper during the grasping process of the manipulator, a gesture recognition module based on depth vision is developed to realize the flexible control of the manipulator. In this paper, the deep learning model YOLOv4 is used to study gesture recognition. In view of the problem that the YOLOv4 model has a single output content and can't control the clamping direction in the process of grasping, the YOLOv4 algorithm model is improved by increasing the outputs of three feature layers $(32, 24, 75)$, $(16, 12, 75)$ and $(8, 6, 75)$ to $(32, 24, 87)$, $(16, 12, 87)$ and $(8, 6, 87)$. The horizontal direction of the gesture is judged by the position information of the two key points of the gesture, and the orientation of the gripper is controlled in the grabbing process, so as to realize the flexible control of the manipulator. Experiments show that the improved YOLOv4 algorithm can realize the accurate recognition of real-time gesture images and the flexible control of the manipulator.

Key words: gesture recognition; deep learning; YOLOv4

0 引言

手势不仅是人与人之间的一种重要交流方式,同时也改变了人与机器之间的交流方式^[1],手势识别目前被广泛应用在多种领域^[2]。在控制领域,手势识别系统取代了按键或者触摸屏控制,通过手势实现机器人运动过程中的目标辨识使用户感到更加自然^[3]。在计算机交互领域,手势取代了鼠标、键盘对计算机的控制,使人与计算机之间的交互更加智能化^[4]。随着卷积神经网络的不断发展,基于深度学习的视觉手势识别系统的应用范围也会越来越广泛。

最初的手势识别方法主要有数据手套和视觉识

别。第一种方法主要利用穿戴设备获取手部各个关节的角度信息,这些设备将手势信息传入识别系统中,精确度比较高,但是需要大量传感器辅助,佩戴繁琐且设备成本较高^[5]。视觉识别方法应用更加方便,通过图像处理、图像分割完成特征提取,然后通过分类算法对手势特征进行识别。但是对数据集的依赖性大,容易受到环境光照、背景干扰,算法的局限性比较大。

随着卷积神经网络的发展,手势识别中的目标分类问题被转换成了目标检测的问题。Girshick 等学者提出 R-CNN 用于目标检测,随后又提出 Fast R-CNN, Faster R-CNN 等基于区域建议的目标检测算法,提高了识别的速度和精度^[5]。Redmon 等学

基金项目: 河南省科技攻关项目(212102310890, 192102210065)。

作者简介: 黄旭阳(1997-),男,硕士研究生,主要研究方向:计算机视觉、机器人技术及应用。Email: 571250316@qq.com; 韩建海(1961-),男,博士,教授,研究生导师,主要研究方向:机器人技术。

收稿日期: 2022-11-07

者^[6]在2016年的CVPR大会提出了基于回归的YOLO算法, YOLO算法不生成候选区域, 直接利用检测网络生成目标的位置、类别信息, 进一步提高了检测的速度。

YOLOv1自2016年提出, 因其检测速度较快而广受关注, 但对目标的定位不够准确, 随后又相继提出了YOLOv2, YOLOv3。在2020年, Bochkovskiy等学者^[7]提出了YOLOv4, 该网络在YOLOv3的基础上结合了很多提高精度的技巧, 包括加权残差链接(WRC)、跨阶段部分连接(CSP)、交叉小批量标准化(CmBN)、自对抗训练(SAT)等, 最终在COCO的数据集上的AP值为43.5%, 在TeslaV100上的实时监测速度达到了65 FPS, 有效提高了识别速度和精度。

1 系统方案设计

基于深度视觉手势识别的协作机械臂物体抓取系统整体设计方案的设计共由2部分组成, 包括手势识别系统和协作机械臂物体抓取系统(图1)。

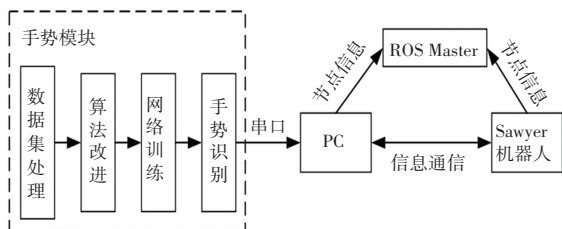


图1 系统整体方案设计

Fig. 1 Overall scheme design of the system

手势识别系统是指通过数据集处理, 将数据通过改进后的算法模型进行训练, 训练后的网络模型对单目摄像机采集到的数据进行预测, 再将预测的手势信息通过串口发送到PC端。

在机械臂抓取物体的系统中, PC端和Sawyer机器人会把自身的节点信息发送到ROS Master中保存起来, ROS Master会根据节点信息进行匹配, 匹配成功后PC端可以直接和Sawyer机器人进行通信, PC端将通过串口获得的手势信息进行处理后, 转换成相应的指令通过节点以话题、服务、动作的通讯方式发出, 机器人节点接收指令后完成相应的动作。

2 手势识别系统设计

手势识别系统主要分为如下步骤。

(1) 数据采集: 采集不同背景, 保存手势的数据信息, 并对采集后的数据进行数据标定。

(2) 算法设计: 为了更加方便地控制机械臂的运动, 基于YOLO v4算法进行改进, 增加输出信息

的种类, 可以实现机械臂的更全面地控制。

(3) 网络模型训练: 将处理后的图片通过卷积、池化、全连接进行特征提取和分类, 通过不断地更新网络权值使预测值和真实值之间的loss小于某一期望值。

(4) 手势识别: 通过对相机采集到的数据进行识别, 将手势的分类信息进行输出, 获得有效的手势识别结果。

2.1 YOLOv4 检测原理

YOLOv4算法是在原有YOLOv3的基础上, 结合了很多提升精度的技巧, 优化了数据处理、基础网络的算法, 使用了Mish激活函数, 提高了识别的速度和精度, YOLOv4的改进如图2所示。这里, 可对图2中的重要设计内容, 展开阐释分述如下。

(1) YOLOv4的主干网络使用了CSPDarkNet53网络, CSPNet的核心思想是输入切分, 通过分割梯度流, 将基础层的特征图分为2部分, 然后利用跨阶段特征融合策略将其合并起来。CSPNet不仅减少了计算量, 而且实现了更丰富的梯度组合。YOLOv4的主干网络使用Mish激活函数代替了Leaky_ReLU激活函数, 平滑的激活函数取代了硬零边界, 允许更好的信息深入神经网络, 提高准确性和泛化能力^[8]。

(2) YOLOv4在CSPDarkNet53结构后添加了SPP模块, SPP用最大池化来满足最后输入特征一致, 解决输入图像尺寸不统一的问题, 可以增加感受视野, 分离出最显著的上下文特征, 使算法适应不同分辨率的图像。

(3) YOLOv4采用了FPN+PAN模块, FPN+PAN一个非常重要的特点就是特征反复提取。FPN将高特征层通过上采样和低特征层做融合得到预测的特征图, 增强了语义信息; PAN自底向上传递定位特征, 通过对FPN模块的补充, 进一步提高了特征提取的能力, 从而加强了对目标的检测能力。

2.2 改进的YOLOv4算法

YOLOv4在YOLOv3的基础上融合了提高精度的技巧, 包括数据增强、加权残差连接、跨阶段部分连接、交叉小批量标准化、Mish激活等, 使得YOLOv4在识别精度和识别速度方面有了很大的提升, 但在最后网格的特征通道上只有先验框的位置、置信度和种类的信息, 不能满足对机械臂灵活控制时的数据要求。针对这种情况, 本文在YOLOv4算法的基础上优化YOLOv4的输出模型结构, 将网格特征通道上加入2个手势关键点的分析, 通过对这2个关键点的训练和分析, 得到手的水平方向, 进而控制机械臂手爪的夹持方向, 提高机械臂控制的灵活性。

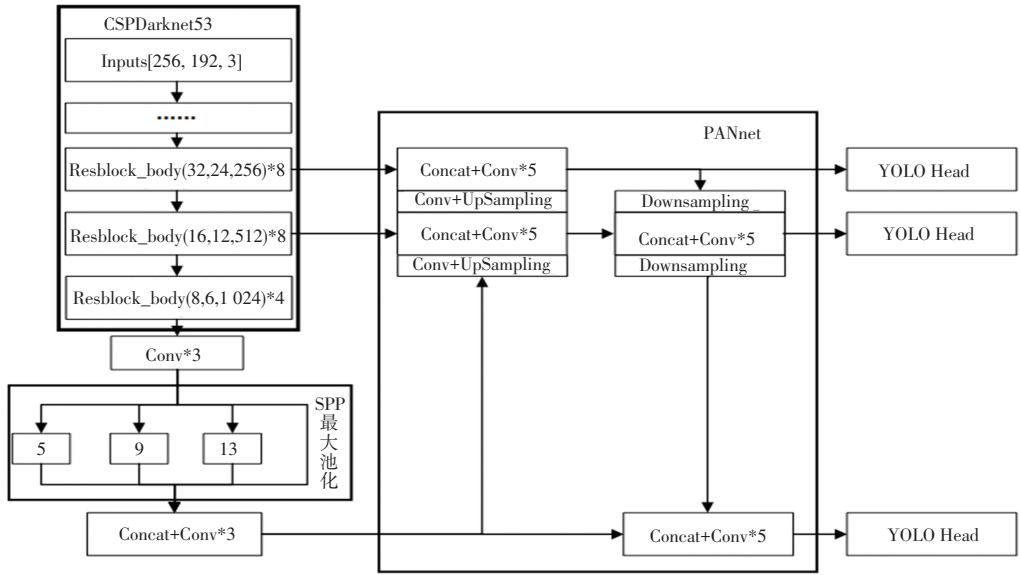


图 2 YOLOv4 网络结构图

Fig. 2 YOLOv4 network structure diagram

YOLOv4 通过 CSPDarkNet53 网络完成对输入图片特征的提取,并对输入图片提取 3 个有效特征层,将 3 个有效特征层分成大小不同的网格,比如特征层为 $(N, 8, 6, 75)$ 的格式就是将输入图像分成了 $8 * 6$ 的网格;接下来在这些网格中心建立 3 个先验框,再通过网络预测结果来判断哪些先验框中有所需要检测的物体,并判断出所检测到物体的种类。

以特征层 $(N, 8, 6, 75)$ 为例,由于每个网格中共同有 3 个先验框,所以 75 可以看作 $3 * 25$,25 中包括 4 个先验框调整参数 (x, y, h, w) , 1 个置信度信息用来确定框内有没有检测到物体,20 是网络检测物体的种类。

为了更加灵活地控制机械臂,将特征层输出内容做出调整,如图 3 所示,通过增加 1, 10 两个关键点的位置信息来判断手的水平方向,为此将特征层的输出结果改变为 $(N, 8, 6, 3, 29)$,新增的 4 组数据表示 2 个手势关键点的坐标信息,随即推得最后特征层的输出结果为 $(N, 8, 6, 87)$ 。

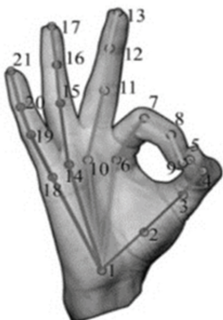


图 3 手势关键点

Fig. 3 Gesture key points

2.3 网络模型训练

神经网络中的损失是衡量训练出来的模型泛化能力好坏的重要指标。损失是指训练后网络模型的预测值和目标值之间的差值,预测值和目标值之间的差值越小,损失值也就越小,说明训练的模型也就越准确,泛化能力也就越好。当损失值比较高时,就需要通过优化函数不断地更新网络权值,使差值达到最小或者小于某一期望值,从而提高模型的准确度和泛化能力^[8-9]。

2.3.1 loss 的计算

损失一般由 3 部分组成。第一部分是预测框 x, y, w, h 带来的误差,用 $CIoU$ loss 计算;第二部分是置信度带来的误差,用 BCE loss 来计算;第三部分是类别和关键点带来的误差,用 BCE loss 来计算。

Intersection over Union (IoU) 是目标检测中的一种重要的评价价值。 IoU 通过计算这 2 个框 A, B 间的 Intersection Area I (相交的面积)和 Union Area U (总的面积)的比值来获得:

$$IoU = \frac{|A \cap B|}{|A \cup B|} = \frac{|I|}{|U|} \quad (1)$$

IoU 是比值的概念,对尺度不敏感,而且当 2 个框的重叠为 0 时, IoU 无法进行反向传播。 $CIoU$ 将预测框和真实框距离、重叠率和惩罚项都考虑进去了,使目标回归更趋稳定^[10]。 $CIoU$ 的公式如下:

$$CIoU = IoU - \frac{\rho^2(b, b^{gt})}{c^2} - \alpha v \quad (2)$$

其中, $\rho^2(b, b^{gt})$ 表示预测框和中心框的距离, c 表示能同时包含预测框和真实框的对角线的距

离。 α 和 v 的公式如下:

$$\alpha = \frac{v}{1 - IoU + v} \quad (3)$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \quad (4)$$

把 $1 - CloU$ 就得到了 $loss_{box}$, 所以 $loss_{box}$ 的公式如下:

$$loss_{box} = 1 - IoU - \frac{\rho^2(b, b^{gt})}{c^2} - \alpha v \quad (5)$$

$$loss_{obj} = - \lambda_{cls} \sum_{i=0}^{s^2} \sum_{j=0}^B I_{i,j}^{obj} \lambda_c [\hat{C}_i^j \log(C_i^j) + (1 - \hat{C}_i^j) \log(1 - C_i^j)] - \lambda_{cls} \sum_{i=0}^{s^2} \sum_{j=0}^B I_{i,j}^{noobj} \lambda_c * [\hat{C}_i^j \log(C_i^j) + (1 - \hat{C}_i^j) \log(1 - C_i^j)] \quad (6)$$

$$loss_{cls} = - \lambda_{class} \sum_{i=0}^{s^2} \sum_{j=0}^B I_{i,j}^{obj} \sum_{c \in classes} \lambda_c * [\hat{p}_i(c) \log(p_i(c)) + (1 - \hat{p}_i(c)) \log(1 - p_i(c))] \quad (7)$$

$$loss_{total} = loss_{box} + loss_{obj} + loss_{cls} \quad (8)$$

其中,带“^”的变量表示标签真值; s 表示特征层的长宽; B 表示每个网格点先验框的个数; $I_{i,j}^{obj}$ 表示预测框是否为正样本,是、为1,否、为0; $I_{i,j}^{noobj}$ 表示该检测框是否是负样本,是、为1,否、为0,如果2个值都为0,说明该检测框为忽略样本。

2.3.2 手势识别结果分析

使用改进后的YOLOv4算法对自制数据集进行检测,手势检测结果如图如4所示,各类手势的置信度见表1。在图4中,预测框在原图上的位置就是手势所在的位置,手势关键点的连线代表着各个手势点的坐标信息,预测框上面左边的标签表示检测出来的手势种类,右边的表示所识别手势的置信度。在表1中,可以看到各类手势的置信度达到了98%,置信度表示该手势被正确识别的概率,根据实验数据显示,改进的YOLOv4算法可以准确地识别出手势种类以及各个关键点的坐标信息。

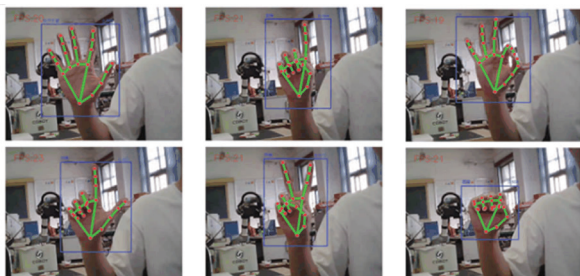


图4 手势检测结果

Fig. 4 Gesture detection results

表1 各类手势置信度

Tab. 1 Gestures confidence

手势	张开手掌	食指	OK	空鼠	双指	握拳
置信度	0.988	0.984	0.979	0.981	0.975	0.988

3 基于改进YOLOv4改进算法的协作机械臂物体抓取实验

3.1 实验环境

实验所用平台来自实验室搭建的基于ROS系统的Sawyer协作机械臂控制系统,操作系统及版本:Ubuntu20.04;手势识别模块:PaddlePi-K210 AI开发套件。

3.2 机械臂抓取物体系统步骤

在手势识别系统中,根据识别到的不同手势信息,来实现机械臂的物体抓取动作,手势识别系统的步骤流程如图5所示。在图5中,首先将机械臂处于使能状态,进入准备阶段,随后打开摄像头获取手势种类、手势关键点和手势水平方向的信息,通过手势的种类控制机械臂执行相应的动作,利用手势的关键点和水平方向的信息来实现机械臂跟踪手势以及抓取物体方向的控制。

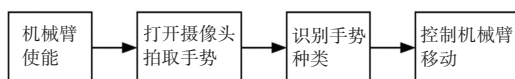


图5 机械臂抓取物体系统步骤

Fig. 5 Steps of mechanical arm grabbing object system

3.3 手势识别应用结果

将手势识别模块应用在机械臂控制中,本文主要设计了2种应用模式,如图6所示。第一种为指令控制模式,也就是检测到触发类动作时输出该指令,随即控制机械臂做出相应的动作。第二种为跟随模式,指检测到手势状态变化时输出该指令,接着使机械臂手爪跟随手势中心点移动,再通过检测手势水平方向控制跟随移动过程中机械臂的手爪夹持方向。

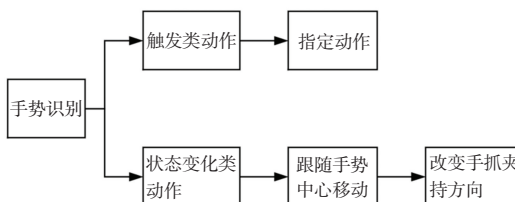


图6 应用模式图

Fig. 6 Application pattern diagram

3.3.1 指令控制模式

在该控制模式中,由手势模块将手势信息通过串口的方式发送到 PC 端,在 PC 端对手势信息进行分析,通过手势分析将指令发送到 Sawyer 机器人。当识别的手势为两指时,机械臂的手爪夹紧;当检测的手势为空鼠时候,机械臂的手爪张开,如图 7 所示。当检测的手势为张开手掌时,机械臂进行摇头操作。指令模式下手势对应指令见表 2。

表 2 指令模式下手势对应指令

Tab. 2 Instructions gesture correspond to gestures in instruction mode

手势	握拳	张开手掌	空鼠	两指
机械臂	回归零点	摇头	手爪张开	手爪夹紧

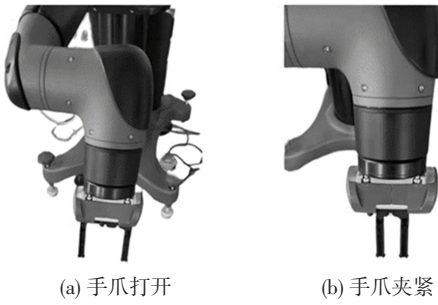


图 7 指令控制实验结果图

Fig. 7 Command control experiment result chart

3.3.2 跟随控制模式

在这种控制模式下,机械臂不仅实现指令控制,而且可以跟随手势中心点移动,完成准确的物体夹取动作。跟随模式实验结果如图 8 所示。



图 8 跟随模式实验结果图

Fig. 8 Experiment result diagram of in follow-up mode

当检测的手势为手掌变 OK 时,进入跟随模式,手爪跟随手势中心点移动同时根据手势水平方向的夹角来调整手爪的夹持方向。当检测的手势为 OK 变手掌时,退出跟随控制,稍后就可以利用指令控制进行物体的夹取。跟随模式下手势对应指令见表 3。表 3 中,GHA (Gesture horizontal angle) 表示手势中 1, 10 两个关键点与水平方向的夹角。

表 3 跟随模式下手势对应指令

Tab. 3 Instructions corresponding to gestures in follow-up mode

手势	手掌变 OK	OK 变手掌	GHA < 90°	GHA > 90°
机械臂	进入跟随模式	退出跟随模式	手爪水平向右	手爪水平向左

4 结束语

本文利用视觉识别的方法实现了手势对机械臂的控制,最终的效果为机械臂跟随手掌移动到指定位置后完成抓取动作。本文设计了多种手势动作,进行了机械臂的控制实验,而后通过对 YOLOv4 算法的改进,在训练结果中引入了手势方向的训练,从而有效实现了在跟随手掌移动过程中手爪夹持方向的问题,让机械臂的控制更加灵活。

参考文献

- [1] 刘亚玲,徐琨,杨紫琦. 基于深度卷积神经网络的手势识别[J]. 信息通信,2019(1):32-33.
- [2] 徐文杰. 基于 Kinect 深度信息的动态手势识别[D]. 沈阳:东北大学,2014.
- [3] 谭典雄. 基于视觉的动态手势识别技术研究[D]. 哈尔滨:哈尔滨工业大学,2014.
- [4] 杨凯. 基于计算机视觉的无标记手势识别技术研究[D]. 北京:华北电力大学(北京),2018.
- [5] 王粉花,黄超,赵波,等. 基于 YOLO 算法的手势识别[J]. 北京理工大学报,2020,40(8):873-879.
- [6] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]// Computer Vision and Pattern Recognition. Las Vegas :IEEE, 2016:779-788.
- [7] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.
- [8] 郭紫嫣,韩慧妍,何黎刚,等. 基于改进的 YOLOV4 的手势识别算法及其应用[J]. 中北大学学报(自然科学版),2021,42(3):223-231.
- [9] 唐进民. 深度学习之 pytorch 实战计算机视觉[M]. 北京:电子工业出版社,2018.
- [10] 李少博,董世浩,夏宇航,等. 基于卷积神经网络的头盔佩戴检测模型的研究[J]. 科学技术创新,2021(25):149-150.