

基于 YOLO 模型的机器人电梯厅门装箱状态快速识别方法

赵海文, 李锋, 赵亚川, 齐兴悦
(河北工业大学 机械工程学院, 天津 300130)

摘要: **目的** 针对电梯厅门柔性生产线机器人装箱后厅门状态识别问题, 提出一种基于 YOLO 模型的电梯厅门装箱状态快速识别方法。 **方法** 采用工业相机采集装箱后厅门图像信息, 并制作成样本训练集, 然后将训练集输入到目标识别分类检测模型中, 通过调整网络结构参数进行迭代训练。 **结果** 经过测试验证, 文中提出的识别方法对装箱后厅门的状态分类识别成功率在 99% 以上, 而且识别速度明显优于传统机器视觉处理算法。 **结论** 文中提出的厅门装箱状态快速识别方法, 可有效解决工业环境中复杂多变光照因素对传统机器视觉处理算法造成的识别效率低、误判率高等问题, 并能满足生产系统节拍要求。

关键词: 电梯厅门; 机器人装箱; YOLO 模型; 状态识别

中图分类号: TB486; TP391 文献标识码: A 文章编号: 1001-3563(2019)07-0180-06

DOI: 10.19554/j.cnki.1001-3563.2019.07.027

Rapid Recognition Method for Loading State of Robot Elevator Hall Door Based on YOLO Model

ZHAO Hai-wen, LI Feng, ZHAO Ya-chuan, QI Xing-yue

(School of Mechanical Engineering, Hebei University of Technology, Tianjin 300130, China)

ABSTRACT: The work aims to propose a method for rapidly recognizing the state of the elevator hall door based on the YOLO model for the problem of the elevator hall door packing state recognition in the flexible production line robot of elevator hall door. The industrial camera was used to capture the container image and make a sample training set. Then the training set was input into the target recognition classification detection model, and iterative training was performed by adjusting the network structure parameters. After testing and verification, the recognition method proposed had a success rate of more than 99% for hall door state recognition, and the recognition speed was obviously superior to the traditional machine vision processing algorithm. The rapid recognition method for hall door packing state proposed can effectively solve the problems of low recognition efficiency and high misjudgment rate of traditional machine vision processing algorithms due to complex and variable illumination factors in industrial environment, and can meet the beat requirements of production system.

KEY WORDS: elevator hall door; industrial robot packing; YOLO model; state recognition

工件的识别和分类是工业机器人在生产线上应用的主要形式^[1-2]。目前广泛应用于工业生产线上机器人很多都是通过示教编程或离线编程完成预设的固定动作和功能^[3]。一旦工件发生与预设位置不符的情况, 就会导致机器人无法完成任务, 甚至有被破坏的可能。

机器视觉是利用视觉传感器和计算机来代替人眼完成判断和决策, 其综合了人工智能、图像处理、模式识别等理论与方法, 来完成对目标的检测、判断、识别和测量等功能^[4-5]。利用深度学习的自学习能力, 主要是卷积神经网络和候选区域算法, 使目标检测识别分类取得了巨大的突破。ROSS 等提出了

收稿日期: 2018-12-04

作者简介: 赵海文 (1974—), 男, 博士, 河北工业大学副教授, 主要研究方向机电一体化、智能工厂技术等。

RCNN 检测算法,在此基础上经过进一步研究,又提出了 Fast R-CNN 和 Faster R-CNN^[6]算法,在准确率和精度上都有所提高。REDMON^[7]等提出 YOLO (you only look once) 为一种全新的目标检测算法,把目标判定和目标识别合二为一,采用回归的方式做到了真正的端到端检测,极大地提高了检测实时性。随后对 YOLO 模型进行优化,提出了 YOLOv2^[8]和 YOLOv3^[9]算法,实现快速检测的同时还达到了较高的准确率。

电梯厅门自动装箱工序是电梯厅门机器人柔性生产线的最后工序,由于采用了木质箱体,箱体加工制造误差较大,这样在机器人装箱过程中,可能会出现厅门倾倒现象。如果不能及时发现,势必会对倾倒的厅门造成损坏,同时也会造成生产线宕机。由于机器人作业环境受光照影响大,如车间投入的户外光、车间照明灯以及车间顶棚悬挂的大风扇所产生的阴影等,传统的机器视觉处理算法在厅门装箱后对厅门的状态识别受到了极大挑战,甚至无法完成指定工作。针对该问题,文中提出基于 YOLO 模型的快速目标状态识别方法,完成对装箱厅门状态的识别,极大地提高了厅门包装入箱的效率和质量。

1 机器人厅门装箱系统简介

电梯厅门机器人装箱系统主要由箱体定位输送机、门板翻转机和装箱机器人 3 部分组成,见图 1。

箱体定位输送机具有对包装箱体定位和输送功能。门板翻转机实现对厅门的输送、定位、翻转和检测。其中,翻转目的是为了使得厅门呈立式姿态便于机器人的抓取和装箱。装箱机器人作为装箱系统的核心部分,其作用是抓取翻转后的厅门,将厅门装入已定位完成的空箱中。机器人手爪到达指定位置后,通过相机采集装箱后厅门的图像;将采集的图像制作成样本训练集。然后将采集的图像输入到经过离线训练的分类模型中进行识别分类。判断已完成装箱的厅门是否发生倾倒情况,将识别为发生倾倒的厅门通过上位机报警提示,以人工介入处理的方式对倾倒厅门进行处理。



图 1 电梯厅门机器人装箱系统

Fig.1 Robot packing system for elevator hall door

图像采集系统是由工业相机、光源、图像采集卡、上位机、工控机和工业机器人组成的光机电一体化视觉识别分类检测装置,其结构简图见图 2。

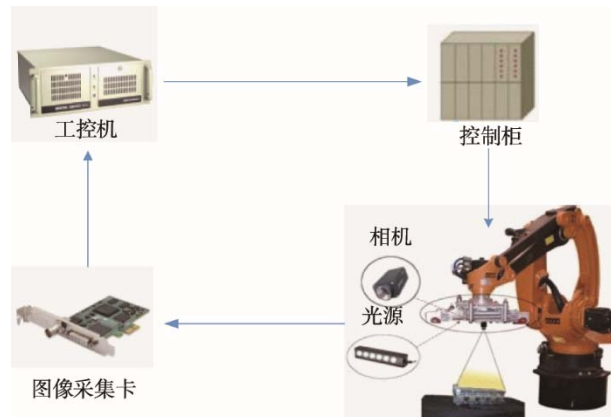


图 2 图像采集系统结构简图

Fig.2 Structure of the image acquisition system

2 卷积神经网络

深度学习是在传统神经网络上运用各种机器学习算法来解决识别图像、文本等问题的算法集合^[10]。深度学习核心是通过特征学习利用分层网络分层次地进行训练,通过自学习获取特征信息从而解决需要人工设计特征的难题。卷积神经网络是一种多层神经网络,擅长处理图像识别分类等问题^[11]。

CNN 最早由 Yann LeCun 提出并应用在手写字体识别上 (MINST)。卷积神经网络是一种前馈式神经网络,前向传播是特征提取的过程,反向传播是权值修正的过程。典型的卷积网络由卷积层、池化层、全连接层组成^[12]。卷积神经网络结构见图 3,其中卷积层与池化层配合,组成多个卷积组合逐层提取特征,最后通过若干个全连接层完成分类。CNN 通过卷积来模拟特征区分,并且通过卷积的权值共享及池化来降低网络参数的数量级,然后通过传统神经网络完成分类等任务^[13—15]。



图 3 卷积神经网络结构

Fig.3 Structure of convolutional neural network

2.1 YOLO 算法原理

首先通过特征提取网络对输入图像提取特征,将图像分成 $S \times S$ 的网格,如果图像中某个物体的中心坐标落在哪个网格单元中,那么就由该网格单元来预测

该物体。每个网格单元预测固定数量的 bounding box 和 bounding box 置信度分数,在 bounding box 中只有和 ground truth 的交并比 (IOU) 最大的 bounding box 才用来预测该物体的。每个 bounding box 包含 5 个预测值: x, y, w, h 和置信度 (C)。(x, y) 坐标表示边框的中心相对网格单元格的边界。如式 (1) 所述, t_x, t_y, t_w, t_h 为模型的预测输出。 C_x 和 C_y 表示网格单元的坐标。 P_w 和 P_h 表示预测前 bounding box 的宽高。 b_x, b_y, b_w 和 b_h 表示预测的 bounding box 的中心坐标和宽高。

$$\begin{aligned} b_x &= \sigma(t_x) + c_x \\ b_y &= \sigma(t_y) + c_y \\ b_w &= p_w e^{t_w} \\ b_h &= p_h e^{t_h} \end{aligned} \quad (1)$$

置信度 C 表示是否包含物体和包含物体情况下检测位置的准确性, 定义为:

$$C = Pr(\text{object}) * IOU_{\text{pred}}^{\text{truth}} \quad (2)$$

$Pr(\text{object})$ 表示预报边界框内存在目标的概率, 若有物体落在网格单元中, 则置信度分数为 1, 反之则为 0。 $IOU_{\text{pred}}^{\text{truth}}$ 为预测边界框与真实边界框之间的交并比系数。

YOLO 算法模型通过非极大值抑制过滤边框输出最终判断结果。为了优化模型, YOLO 算法模型使用了均方和误差作为损失函数 (loss function) 的参数。损失函数由 3 部分组成: 定位误差、置信度误差和识别误差, 公式见式 (3)。

$$\begin{aligned} l_{\text{loss}} &= \lambda_{\text{corr}} \sum_{i=0}^s \sum_{j=0}^B 1_{ij}^{\text{obj}} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] + \\ &\lambda_{\text{corr}} \sum_{i=0}^s \sum_{j=0}^B 1_{ij}^{\text{obj}} [(\sqrt{\omega_i} - \sqrt{\hat{\omega}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2] + \\ &\sum_{i=0}^s \sum_{j=0}^B 1_{ij}^{\text{obj}} (c_i - \hat{c}_i)^2 + \lambda_{\text{noobj}} \sum_{i=0}^s \sum_{j=0}^B 1_{ij}^{\text{obj}} (c_i - \hat{c}_i)^2 + \\ &\lambda_{\text{corr}} \sum_{i=0}^s \sum_{j=0}^B 1_{ij}^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c_i) - \hat{p}_i(c))^2 \end{aligned} \quad (3)$$

式中: $(\hat{x}, \hat{y}, \hat{\omega}, \hat{h}, \hat{C}, \hat{p})$ 为预测值; 1_{ij}^{obj} 为物体落入格子 i 的第 j 个边界框做出预测, 如果某个单元格中没有目标, 则不对分类误差进行反向传播; $\lambda_{\text{corr}}=5$, $\lambda_{\text{noobj}}=0.5$ 。

2.2 改进的 YOLO 网络结构

YOLOv3-tiny 模型在 YOLOv3 模型基础上改进而来。由于文中涉及的数据集远远小于 PASCAL VOC 等大型数据集, 且文中厅门装箱状态识别分类种类少, 只有正常和异常 (厅门倾倒) 2 种类型。较深的提取网络容易造成模型的过拟合, 使得检测性能下降, 因此考虑对网络进行适当简化, 以适应文中的状

态识别对象。

在特征提取部分, 针对文中目标识别对象特征, 设计的网络结构比较浅, 没有 res 层, 只包含卷积层和最大池化层。所有卷积层采用零填充, 不改变输入图像尺寸, 池化层采用最大池化函数。故文中网络模型包含 10 个卷积层、6 个最大池化层。在分类和检测回归部分, 采用 2 次 route 操作将输入层连接在一起, 分别是将 13 层 feature map 进行卷积操作生成 $13 \times 13 \times 128$ 的 feature map, 然后进行上采样生成 $26 \times 13 \times 128$ 通道的 feature map。然后再与 8 层的 $26 \times 26 \times 256$ 通道的 feature map 合并, 再进行卷积操作。最后输出 $26 \times 26 \times 21$ 大小的 feature map, 在此基础上进行分类和位置回归。文中改进后的网络模型结构见图 4。

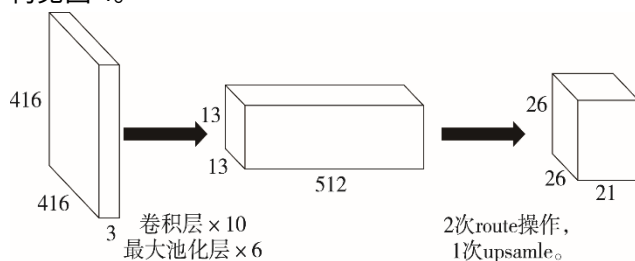


图 4 YOLOv3-tiny 网络结构模型

Fig.4 Structure of YOLOv3-tiny network model

3 实验验证与结果分析

3.1 训练目标检测网络

利用现场工业机器人末端安装的工业相机采集目标图像作为训练集样本, 采用增加图像对比度、增加图像曝光度以及旋转图像等方式实现对数据集的扩充。采集原始图像样本数量为 300 张, 经过数据集扩充后样本图像数量为 500 张, 其中训练集 300 张, 验证集 200 张。部分训练样本集见图 5, 按照标准 VOC 数据集格式进行构建。

1) 按照 VOC 数据集格式来构建数据集, 用 Python 读取文件夹下所有训练样本图像然后统一命名并修改格式, 并按照文件夹结构放置图像文件 (.xml 文件及 .txt 文件), 见图 6。

2) 使用 Labellmg 软件标记样本图像中的目标区域以便训练使用, 通过鼠标框选目标区域完成标记。

3) 算法模型训练。实验软硬件配置见表 1。设置配置文件参数: Batch size 是每次训练数据的样本数量; Subdivisions 是将 batch 分割为 Subdivisions 个子 batch, 每个子 batch 的大小为 batch/subdivisions; Class 表示类别数, 文中样本训练集中只有 2 类, 正常和异常, 故为 2; filter 表示输出特征图的数量, 计算公式见式 (4)。

$$\text{Filter} = (\text{Class} + 5) \times 3 \quad (4)$$

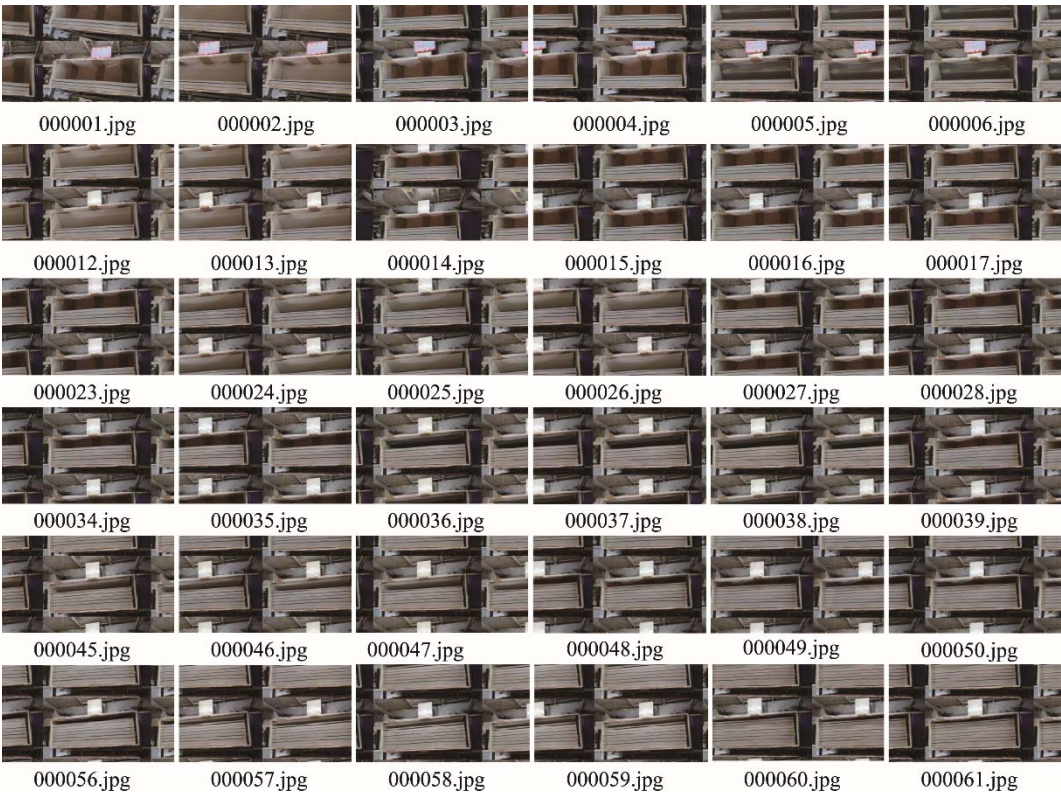


图 5 部分训练样本集
Fig.5 Partial training sample set

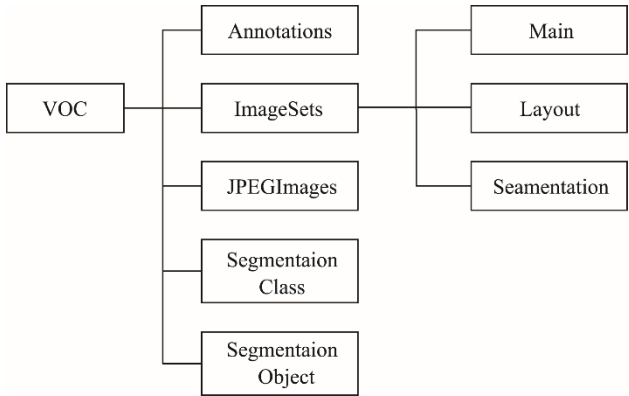


图 6 VOC 数据集格式
Fig.6 VOC data set format

表 1 软硬件配置
Tab.1 Software and hardware configuration

名称	相关配置
操作系统	Windows 7
CPU	Intel (R) Core (TM) i5-6500
GPU	NVIDIA GeForce GTX750Ti
深度学习框架	Darknet
安装库	cuda 9.1, Python 3.5, cuDNN 7.0

Learning_rate 表示学习率，学习率决定了参数达到最优速度的快慢，学习率过大，可能会导致函数无

法收敛，甚至发散；反之学习率过小，优化效率可能过低，算法长时间无法收敛。合适的学习率应该是在保证收敛的前提下，能尽快收敛。训练参数的配置见表 2。

表 2 训练参数配置
Tab.2 Training parameter configuration

Batch size	subdivisions	Class	filter	Learning_rate
64	16	2	2	0.001

3.2 实验结果分析

在文中，状态识别算法需要判断装箱后厅门的状态正常还是异常，即检测装入箱体內的厅门是否倾倒，是一个二分类问题。为了验证该算法的性能，使用 200 张图片进行分类识别实验测试，以便对目标识别算法的准确率和效率进行综合分析性能评估。设置了对比实验，分别用 YOLOv3 算法和 Faster R-CNN 算法进行对比实验，测试算法的识别效果。

采用准确率、错检数以及每秒检测帧数来对比算法的优劣。准确率是指正确识别出装箱厅门状态（正常与倾倒）的次数占总识别次数的比例；错检数为无法正确识别厅门状态的次数。目标识别算法对比分析见表 3。

由表 3 对比分析结果可知，3 种算法分别在相同的数据集进行检测，都能够规定的生产节拍（该生

表 3 检测结果对比
Tab.3 Comparison of test results

算法模型	准确率/%	错检数	每秒检测帧数
Faster R-CNN	95.0	10	2
YOLOv3	97.0	6	60
YOLOv3-tiny	99.5	1	60

产线系统的生产节拍为 45 s) 内完成任务。YOLOv3-tiny 算法在复杂的工业环境下准确率较 Faster R-CNN 算法有较大优势, 相对于 YOLOv3 算法准确性也略高。实时性方面 YOLOv3-tiny 算法每秒检测帧数远大于比 Faster R-CNN 算法, 完全能够满足实时性要求。不同光照强度、对比度和添加干扰情况工业现场实验环境下检测见图 7—8, 其中在亮度正常、变亮、减弱, 对比度正常、增强、减弱, 添加干扰情况下部分识别结果见图 7a—j。在亮度正常、变亮、减弱, 对比度正常、增强、减弱, 添加干扰情况下部分识别结果见图 8a—j。

实验结果充分表明, 文中算法在电梯厅门装箱过程中对厅门的状态识别具有准确率高、识别速度快等特点, 能够满足实际生产要求。对于作业环境受诸多因素干扰的工业环境, 基于 YOLO 模型的目标分类识别算法采用端到端的网络结构设计, 利用回归算法对目标类别以及目标边界框位置进行拟合, 能够更好地克服复杂多变的光照环境对于目标特征提取时所带来的影响。

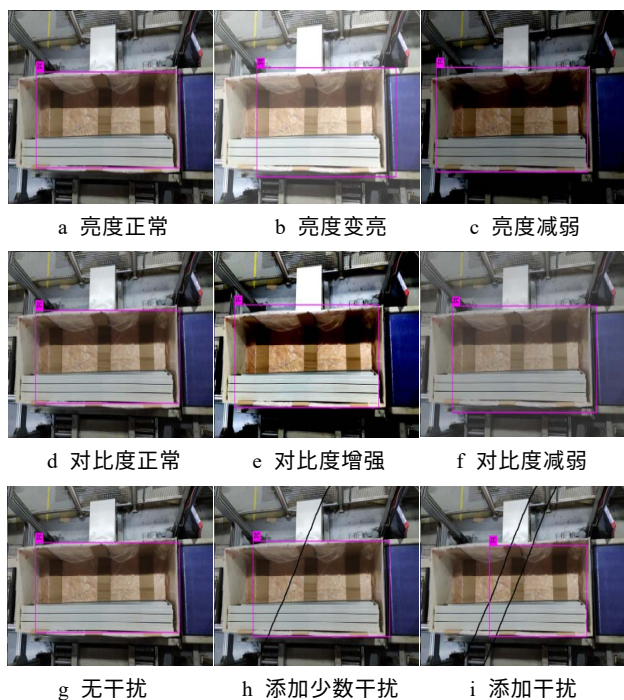


图 7 厅门装箱在正常情况下的部分识别结果
Fig.7 Partial identification result of hall door packing under normal conditions

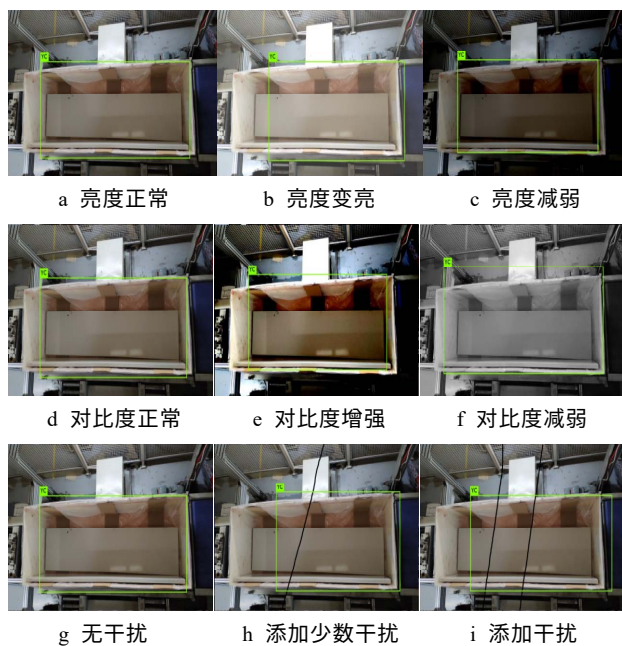


图 8 厅门装箱在异常情况 (倾倒) 下的部分识别结果
Fig.8 Partial identification result of hall door packing (dumping) under abnormal conditions

4 结语

文中将基于 YOLO 模型的目标状态识别分类方法, 应用于电梯厅门机器人柔性生产线中装箱工序对厅门装箱状态的识别, 有效地解决了实际工业工况环境中各种复杂多变光照因素对于传统视觉图像处理算法所造成识别效率低、误判率高等问题。

实验结果表明, 文中提出的目标识别分类方法可快速、准确地识别判断装箱厅门的状态, 并适合在工业环境中应用。后续还将对在线训练和样本数据集的自动标注、数据集的自扩充等相关工作进行深入研究, 使得该视觉系统能达到使用时间越长, 识别速度越快, 识别结果越准确的目标。

参考文献:

- [1] 朱良, 林浒, 吴文江. 基于机器视觉的工业机器人工件定位[J]. 小型微型计算机系统, 2016, 37(8): 1873—1877.
ZHU Liang, LIN Hu, WU Wen-jiang. Workpiece Positioning of Industrial Robot Based on Machine Vision[J]. Journal of Chinese Computer Systems, 2016, 37(8): 1873—1877.
- [2] 解则晓, 陈文柱, 迟书凯, 等. 基于结构光视觉引导的工业机器人定位系统[J]. 光学学报, 2016(10): 162—169.
XIE Ze-xiao, CHEN Wen-zhu, CHI Shu-kai, et al. Industrial Robot Positioning System Based on the Guidance of the Structured-Light Vision[J]. Acta Optica

- Sinica, 2016(10): 162—169.
- [3] 倪自强, 王田苗, 刘达. 基于视觉引导的工业机器人示教编程系统[J]. 北京航空航天大学学报, 2016, 42(3): 562—568.
NI Zi-qiang, WANG Tian-miao, LIU Da. Vision Guide Based Teaching Programming for Industrial Robot[J]. Journal of Beijing University of Aeronautics and Astronautics, 2016, 42(3): 562—568.
- [4] 罗时光. 基于机器视觉的玻璃瓶口缺陷检测方法[J]. 包装工程, 2018, 39(3): 183—187.
LUO Shi-guang. Glass-bottle Defect Detection Method Based on Machine Vision[J]. Packaging Engineering, 2018, 39(3): 183—187.
- [5] 杨仁民, 郑洲, 陈斌, 等. 基于机器视觉的零件特征尺寸提取算法[J]. 包装工程, 2017, 38(9): 151—156.
YANG Ren-min, ZHENG Zhou, CHEN Bin, et al. Extraction Algorithm of Part Feature Sizes Based on Machine Vision[J]. Packaging Engineering, 2017, 38(9): 151—156.
- [6] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 39(6): 1137—1149.
- [7] REDMON J, DIVVALA S, GIRSHICK R, et al. You Only Look Once: Unified, Real-time Object Detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779—788.
- [8] REDMON J, FARHADI A. YOLO9000: Better, Faster, Stronger[C]// IEEE Conference on Computer Vision & Pattern Recognition, 2017.
- [9] REDMON J, FARHADI A. Yolov3: An Incremental Improvement[C]// IEEE Conference on Computer Vision and Pattern Recognition, 2018.
- [10] 张慧, 王坤峰, 王飞跃. 深度学习在目标视觉检测中的应用进展与展望[J]. 自动化学报, 2017, 43(8): 1289—1305.
ZHANG Hui, WANG Kun-feng, WANG Fei-yue. Advances and Perspectives on Applications of Deep Learning in Visual Object Detection[J]. Acta Automatica Sinica, 2017, 43(8): 1289—1305.
- [11] 刘海龙, 李宝安, 吕学强, 等. 基于深度卷积神经网络的图像检索算法研究[J]. 计算机应用研究, 2017, 34(12): 3816—3819.
LIU Hai-long, LI Bao-an, LYU Xue-qiang, et al. Image Retrieval Based on Deep Convolutional Neural Network[J]. Application Research of Computers, 2017, 34(12): 3816—3819.
- [12] 常亮, 邓小明, 周明全, 等. 图像理解中的卷积神经网络[J]. 自动化学报, 2016, 42(9): 1300—1312.
CHANG Liang, DENG Xiao-ming, ZHOU Ming-quan, et al. Convolutional Neural Networks in Image Understanding[J]. Acta Automatica Sinica, 2016, 42(9): 1300—1312.
- [13] 胡安翔, 李振华. 基于Faster R-CNN改进的数粒机系统[J]. 包装工程, 2018, 39(9): 141—145.
HU An-xiang, LI Zhen-hua. Improved Capsule Counting Machine Based on Faster R-CNN[J]. Packaging Engineering, 2018, 39(9): 141—145.
- [14] 伍锡如, 黄国明, 孙立宁. 基于深度学习的工业分拣机器人快速视觉识别与定位算法[J]. 机器人, 2016, 38(6): 711—719.
WU Xi-ru, HUANG Guo-ming, SUN Li-ning. Fast visual Identification and Localization Algorithm for Industrial Sorting Robots Based on Deep Learning[J]. ROBOT, 2016, 38(6): 711—719.
- [15] 訾斌, 尹泽强, 李永昌, 等. 基于YOLO模型的柔索并联机器人移动构件快速定位方法[J]. 机械工程学报, 2019, 55(3): 64—72.
ZI Bin, YIN Ze-qiang, LI Yong-chang, et al. A Fast Mobile Component Location Method for Cable-driven Parallel Robots Based on YOLO Model[J]. Journal of Mechanical Engineering, 2019, 55(3): 64—72.