

一种基于深度学习的机械臂分拣方法

田思佳¹, 顾强², 胡蓉¹, 李锐戈¹, 何顶新¹

(1. 华中科技大学, 湖北 武汉 430074; 2. 武汉精锋微控科技有限公司, 湖北 武汉 430000)

摘要: 针对分拣过程中视觉系统识别复杂物体时速度慢、对环境变化适应性不足的问题, 提出一种基于轻量型卷积神经网络的机械臂快速分拣方法。该方法首先使用基于轻量型卷积神经网络的 MobileNet-SSD 算法对图像中的目标物体进行检测, 获取目标类别和位置信息; 然后根据检测输出结果对图像进行预处理和边缘检测, 最终得到校正后的定位结果。在 PROBOT Anno 机械臂平台上进行分拣实验, 实验结果表明, 相比于传统的图像处理方法, 提出的方法能对复杂目标物体实现快速的检测和定位, 对于目标形态和环境的多样性来说具有更好的鲁棒性。

关键词: 机械臂分拣; 深度学习; 目标检测; 机器视觉

中图分类号: TP242

文献标识码: A

doi: 10.11959/j.issn.2096-6652.202029

A robot sorting method based on deep learning

TIAN Sijia¹, GU Qiang², HU Rong¹, LI Ruige¹, HE Dingxin¹

1. Huazhong University of Science and Technology, Wuhan 430074, China

2. Wuhan PS-Micro Technology Co., Ltd, Wuhan 430000, China

Abstract: A fast robot sorting method based on lightweight convolutional neural network was proposed to improve the recognition speed and environmental adaptability, especially for sorting complex objects. Firstly, the MobileNet-SSD algorithm was used to detect and classify the objects based on lightweight convolutional neural network. Secondly, image preprocessing and edge extraction were used to revise the object locations according to the above object detection results. The sorting experiments on PROBOT Anno robot arm show that the proposed method can achieve fast detection and location for complex objects. Compared with traditional image processing methods, the proposed method is robust to the diversity of target morphology and environment.

Key words: robot sorting, deep learning, object detection, machine vision

1 引言

随着工业自动化水平的不断提高, 越来越多的工业机器人参与到工业生产过程中。分拣作业是当前工业生产的重要环节。在传统的分拣流水线中, 机器人的运动控制采用示教或离线编程的方式实现, 操作人员需要严格设定目标位置和机器人的运动轨迹, 因此传统的分拣方法无法适应灵活的工作环境^[1]。

机器视觉技术与工业分拣系统的结合能够显著增强机器人的环境适应能力, 提高分拣作业的可靠性和灵活性^[2]。在基于机器视觉的工业机器人系统中, 目标检测与识别是实现分拣任务的关键技术。Wang J W 等人^[3]利用尺度不变特征变换 (scale-invariant feature transform, SIFT) 算法结合矩不变量, 提出一种目标识别和定位的方法, 实验表明, 该图像处理算法具有较强的鲁棒性。Wang S Y 等人^[4]在 Canny 算法框架中对输入图像进行多尺度滤波分

收稿日期: 2020-08-05; 修回日期: 2020-09-04

通信作者: 田思佳, 1360727243@qq.com

基金项目: 国家自然科学基金资助项目 (No.61672244)

Foundation Item: The National Natural Science Foundation of China (No.61672244)

析,提取其精确的边缘特征,并通过生成不同旋转角度的被检测对象模型建立模板库,克服了位置变化对传统模板匹配识别算法的影响。上述经典的视觉控制方法是基于特定的图像特征的,在目标对象较为复杂或环境多变的应用场景中难以取得很好的分拣效果。

Hinton G E 等人^[5]在 2006 年最早提出了深度置信网络的概念,还引入分层预训练技术,开启了深度学习的时代。与传统特征提取方法相比,深度学习通过数据驱动的方式可以自主学习到具有更好泛化能力的内在特征,并在目标分类识别、图像分割等任务中取得了突破性的进展。深度学习方法能够有效提高机械臂智能化分拣的能力。杜学丹等人^[6-7]使用深度学习中的基于区域的更快的卷积神经网络(faster region-based convolutional neural network, Faster R-CNN)等算法完成了物体的识别与定位任务,他们训练的目标检测模型具有很好的泛化能力和较高的准确率,但检测速度较慢,实时性较差。Kumar R 等人^[8]首先利用特征提取算法获取目标位置,然后训练神经网络对目标进行分类,达到了较高的分拣准确率。但该位置检测算法对光线等环境因素比较敏感,严重影响了分类网络的准确性。

基于以上问题,本文提出一种基于轻量型卷积神经网络的机械臂快速分拣方法,并在 PROBOT Anno 机械臂平台上完成了对 3 种动物模型的快速、无序分拣。本文提出的方法首先使用基于 MobileNet 基础网络的单次多框检测器(single shot

MultiBox detector, SSD) 算法^[9]对图像中的目标进行检测,获取目标的类别和边框位置;然后根据目标检测结果,对图像进行预处理和边缘提取,最终得到校正的定位结果。本文提出的方法具备较高的识别和定位准确率,并且能够满足工业系统中对形状不规则物品无序分拣的实时性要求,提高分拣的效率。

2 机械臂分拣系统及算法模型

虽然工业系统中不同分拣任务的被分拣对象不同,但是只要获得目标物体的定位坐标和类别信息,就能实现对复杂目标的快速准确分拣。机械臂分拣系统结构及算法流程如图 1 所示,算法流程分为以下两部分。

- 使用基于轻量型卷积神经网络的目标检测算法对图像中的目标物体进行检测,得到其类别和预测框位置信息。
- 根据检测结果对图像进行裁剪,对裁剪出的图像进行预处理和边缘检测,得到校正后的目标定位坐标。

本文提出的方法先采用深度学习方法进行目标检测,确定物体位置,然后进一步校正得到最终定位坐标,这样做有两个目的:一是深度学习的检测方法对于目标形态和环境的多样性来说具有很强的鲁棒性,能更准确地检测到物体的位置区域和类别;二是根据目标检测算法输出的结果对图像中包含物体的局部像素区域进行边缘检测操作,背景相对单一,可以实现更好的定位效果。

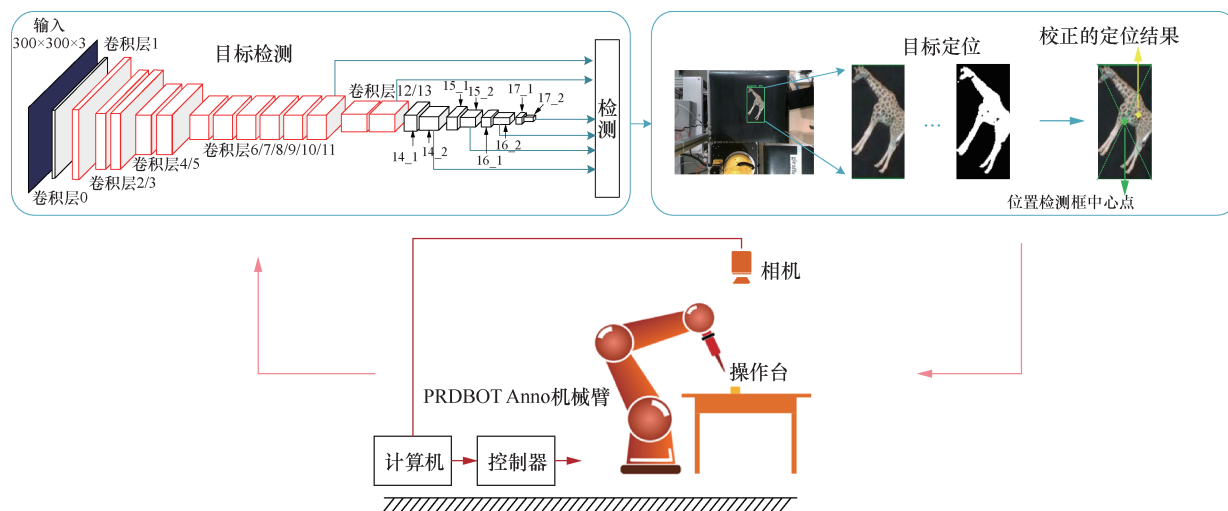


图1 机械臂分拣系统结构及算法流程

3 目标检测

3.1 目标检测方案

基于卷积神经网络的目标检测算法较好地解决了传统检测算法中存在的问题，其主要分为基于区域提名与卷积分类的算法（two-stage 算法）和基于回归的算法（one-stage 算法）两类^[10]。

two-stage 算法虽然检测精度较高，但检测速度较慢，难以满足工业需求，因此本文着重研究检测速度更快的 one-stage 算法。

SSD 算法是 one-stage 算法的代表算法之一。SSD 算法不需要进行区域提名，可直接从图片获得预测结果，其将目标检测任务简化成了回归问题。SSD 算法结合 YOLO（you only look once）算法的回归思想和 Faster R-CNN 算法的锚框机制实现了对目标的精准定位。它通过在多个特征图上设定多种长宽比的先验框，提高了对不同尺度目标的检测效果。

针对模型性能和实际应用需求，本文采用 SSD 算法实现机械臂分拣任务中的视觉检测。

3.2 目标检测模型

SSD 算法采用 VGG16 (visual geometry group 16) 模型进行特征提取，新增卷积层以获得多尺度特征图，用于最终的检测。为了提高检测速度，本文将基础网络调整为 MobileNet^[11]。MobileNet 是一种基于深度可分离卷积构建的轻量型深度神经网络，可以更好地满足机械臂在分拣任务中对检测速度的要求。

基于 SSD 算法的目标检测流程如图 2 所示，包含 3 个步骤：首先，将图像送入 MobileNet 卷积层和 SSD 卷积层进行特征提取；然后，采用两组 3×3 的卷积层，在选取的 6 个特征图上提取检测结果；最后，经过非极大值抑制（non-maximum suppression, NMS）输出最终的物体类别和边框位置。

先验框的设置是 SSD 算法的关键步骤。SSD 算法在 6 个特征图的每个单元（一个像素点为一个单元）设置了不同长宽比 a_r ($a_r \in \{1, 2, 3, 1/2, 1/3\}$) 的先验框。预测的边界框以先验框为基准。如图 3 所示，每个单元设置了 4 个不同的先验框，每个先验框都对应 loc 和 conf 两部分检测值。

loc 为边界框的位置， (cx, cy, w, h) 分别表示边界框的中心坐标以及宽和高。为了方便训练，通常要对边界框进行如下编码：

$$\begin{aligned} l^{cx} &= (g^{cx} - d^{cx}) / d^w \\ l^{cy} &= (g^{cy} - d^{cy}) / d^h \\ l^w &= \log(g^w / d^w) \\ l^h &= \log(g^h / d^h) \end{aligned} \quad (1)$$

其中， $(d^{cx}, d^{cy}, d^w, d^h)$ 表示先验框位置， $(g^{cx}, g^{cy}, g^w, g^h)$ 表示真实的标注框位置， l 表示边界框预测值。

conf 是各个类别的置信度，SSD 算法将背景也作为一个特殊的类别。

在模型的训练过程中，需要进行先验框的匹配，匹配规则如下：

- 对于图中的真实标注框，找与其交并比最大的先验框进行匹配；

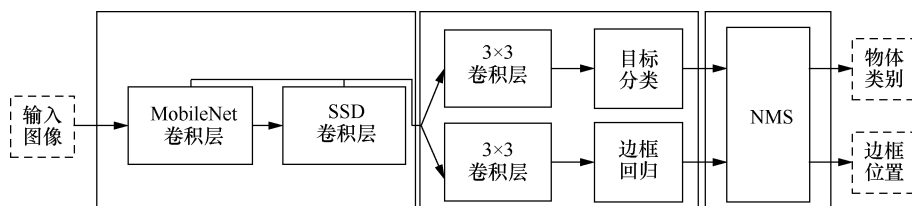


图2 基于 SSD 算法的目标检测流程

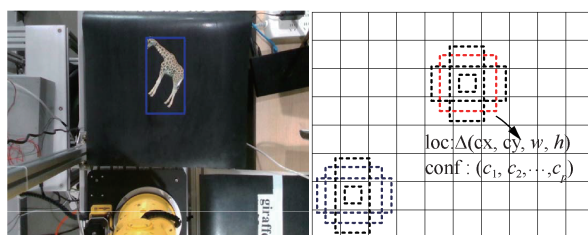


图3 先验框长宽比及对应输出

• 对于剩余未匹配的先验框，若其与某个真实标注框的交并比大于 0.5，则对应的真实标注框被匹配为正样本；否则该先验框被匹配为负样本。

由于负样本数一般远大于正样本数，故采用难分样本挖掘的方法对负样本进行抽取，以保证正负样本的比例为 1:3。基于以上定义，SSD 算法采用位置误差与置信度误差的加权和作为损失函数：

$$L(x, c, l, g) = \frac{1}{N} (L_{\text{conf}}(x, c) + \alpha L_{\text{loc}}(x, l, g)) \quad (2)$$

其中， c 为置信度， l 和 g 分别为边界框预测值和真实标注框的位置参数， N 为正样本数量，权重系数 α 通过交叉验证设置为 1。

位置误差 $L_{\text{loc}}(x, l, g)$ 采用 smooth L1 损失函数：

$$L_{\text{loc}} = \sum_{i \in \text{Pos}} \sum_{m \in \{cx, cy, w, h\}} x_{ij}^k \text{smooth}_{\text{L1}}(l_i^m - \hat{g}_j^m) \quad (3)$$

其中，Pos 为正样本集合； $x_{ij}^p \in \{1, 0\}$ 为指示参数； $x_{ij}^p = 1$ 时表示第 i 个先验框与第 j 个真实标注框匹配，且类别为 p ； $\hat{g}_j^m$ 为第 j 个真实标注框经过编码得到的位置映射，计算如下：

$$\begin{aligned} \hat{g}_j^{\text{cx}} &= (g_j^{\text{cx}} - d_i^{\text{cx}}) / d_i^w \\ \hat{g}_j^{\text{cy}} &= (g_j^{\text{cy}} - d_i^{\text{cy}}) / d_i^h \\ \hat{g}_j^w &= \log(g_j^w / d_i^w) \\ \hat{g}_j^h &= \log(g_j^h / d_i^h) \end{aligned} \quad (4)$$

置信度误差 $L_{\text{conf}}(x, c)$ 采用 softmax 损失函数：

$$\begin{aligned} L_{\text{conf}} &= - \sum_{i \in \text{Pos}} x_{ij}^p \log(\hat{c}_i^p) - \sum_{i \in \text{Neg}} \log(\hat{c}_i^0) \\ \hat{c}_i^p &= \frac{\exp(c_i^p)}{\sum_p \exp(c_i^p)} \end{aligned} \quad (5)$$

其中，Neg 为负样本集合， c_i^p 表示第 i 个边界框中包含的对象为类别 p 的置信度， $\hat{c}_i^p$ 表示第 i 个边界框中包含的对象为类别 p 的概率， $\hat{c}_i^0$ 表示第 i 个边界框中包含的对象为背景的概率。

4 目标定位

目标检测模型输出的位置是贴合物体外部轮廓的矩形，校正前后的定位点示意如图 4 所示。对于形状不均匀的物体，将矩形中心（如图 4 中的绿色点）作为最终定位点是不合适的。目标定位的目的是对目标检测的输出结果做进一步处理，对定位

点进行校正（如图 4 中的黄色点），以提高分拣的成功率。目标定位流程如图 5 所示。

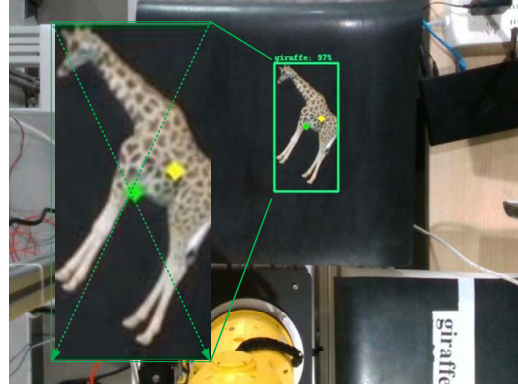


图 4 校正前后的定位点示意

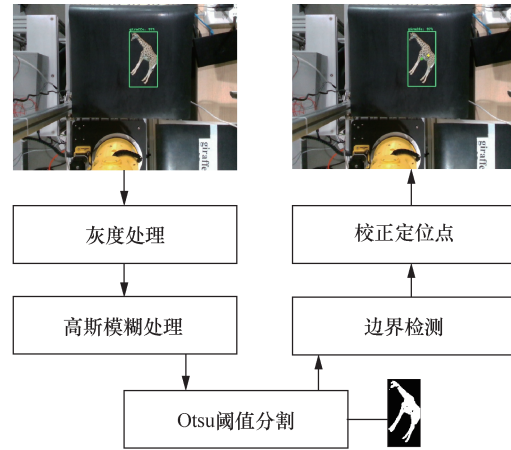


图 5 目标定位流程

目标定位流程如下。

步骤一 根据目标检测位置输出结果对图片进行裁剪，此时通过计算边框的中心点坐标，获得粗略的定位点位置（如图 5 中绿色点所示）。

步骤二 对彩色图像的三通道分量进行平均，获得灰度图，增强图片对比度。

步骤三 使用长宽均为 3 的高斯核对图像进行高斯模糊处理。

步骤四 使用自适应阈值分割算法（Otsu）最小化类间方差，分离前景和背景。

步骤五 通过边界检测提取物体轮廓，进而根据图像矩计算出轮廓的质心，将定位点位置由边框中心点位置校正为质心点位置。

5 实验结果与分析

5.1 目标检测定位实验

本文提出的方法基于深度学习框架 TensorFlow

与机器人软件平台 ROS，使用 Python 开发。

本文 MobileNet-SSD 模型由 COCO 预训练模型进行初始化。自建数据集按照 PASCAL VOC 数据集格式进行构建，包含 3 类共 1 500 张图片，每类图片 500 张。3 个类别分别为长颈鹿、鸭子、鹦鹉。随机抽取每类图片的 70% 用于模型训练，剩余的 30% 用于模型测试。

模型训练在 Ubuntu 系统中完成，通过 GTX1080 显卡实现训练加速。模型训练的超参数设置如下：批处理参数 `batch_size` 为 24，初始学习率 `init_learning_rate` 为 0.004 3，学习率衰减策略为指数衰减，衰减因子 `decay_factor` 为 0.95，最大迭代次数为 200 000 次。

目标检测在 Windows 系统中完成，PC 处理器为英特尔 Core i7-6700K@4.00 GHz 四核，内存为 16 GB。在 CPU 环境中，目标检测速度约为 0.080 6 s/张。目标检测及目标定位实验结果如图 6 所示。实验结果表明：算法具有较高的检测准确性和较快的检测速度；校正后的定位点（图 6 中的黄色点）更靠近物体中心位置，有助于提升分拣成功率。

为进一步验证所提方法的检测准确性与检测速度，本文进行了以下两组对比实验。

(1) 检测准确性对比验证

在相同的实验环境中，分别采用本文的检测方法和基于 Canny 边缘提取算法^[8]进行位置检测，实验结果如图 7 所示。图 7 (a) 和图 7 (c) 为采用本

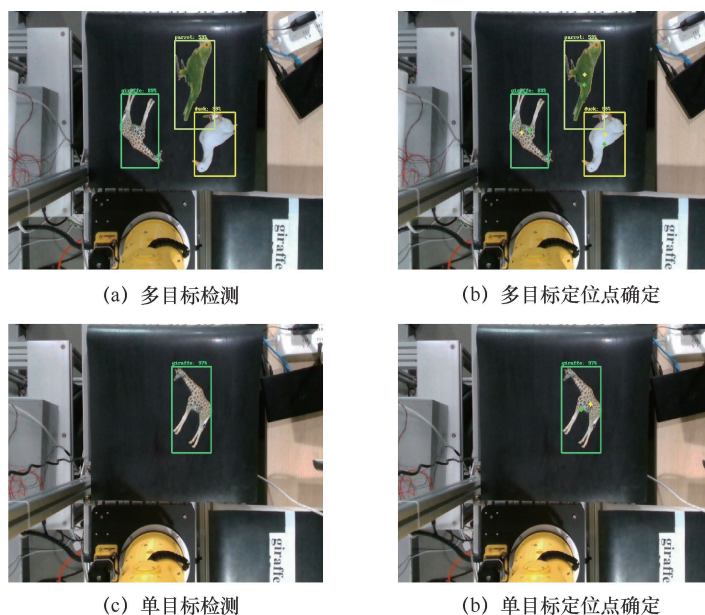


图6 目标检测及目标定位实验结果

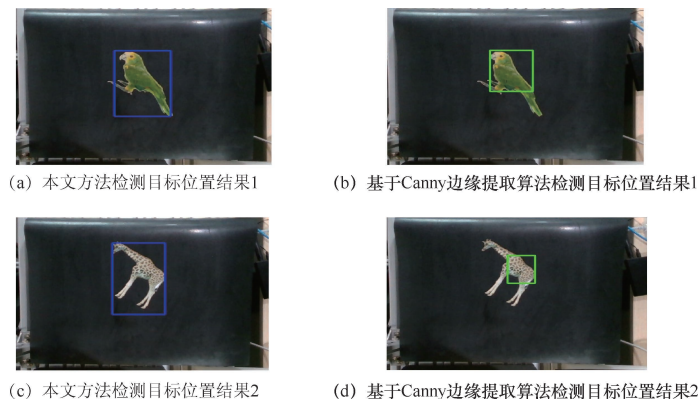
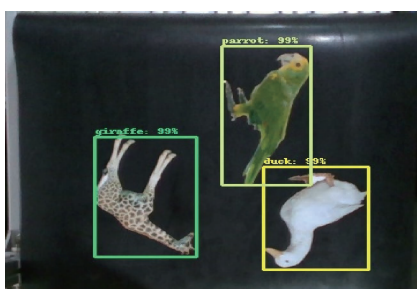


图7 本文检测方法结果与传统检测方法结果对比

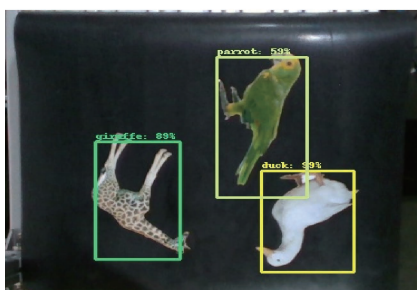
文基于深度学习的检测方法的检测结果。图 7 (b) 和图 7 (d) 为采用基于 Canny 边缘提取算法的检测结果^[8]，基于 Canny 边缘提取算法对细长形状的物体位置检测效果较差，且光线变化对检测的准确性有很大的影响。结果表明，对于光线、目标形状等，本文方法都具有较强的鲁棒性，可实现准确的位置检测。

(2) 检测速度对比验证

在相同的实验环境中，分别采用本文基于 MobileNet-SSD 的方法和基于 Faster R-CNN 的方法^[6]进行实验。检测结果如图 8 所示。实验结果表明，本文方法的目标识别与定位结果具有较高的准确性。但识别结果的置信度和目标位置预测框的精度较低。



(a) Faster R-CNN方法检测效果



(b) 本文方法检测效果

图 8 Faster R-CNN 方法结果与本文检测方法结果对比

上述两种方法的检测性能指标量化见表 1。结果表明，在保证较高检测准确率的前提下，本文目标检测方法的检测速度得到了显著提升。

表 1

本文方法与基于 Faster R-CNN 的方法的检测性能比较

类别	本文方法			基于 Faster R-CNN 的方法		
	检测时间/s	正确检测个数/个	检测准确率	检测时间/s	正确检测个数/个	检测准确率
长颈鹿	0.078	293	97.7%	3.734	300	100%
鸭子	0.080	300	100%	3.721	300	100%
鹦鹉	0.082	290	96.7%	3.738	299	99.7%

5.2 机械臂分拣实验

本节在基于视觉的 PROBOT Anno 机械臂平台上进行了 3 种动物模型的无序分拣实验，完成对所提方法的验证。以长颈鹿等动物为被抓取对象，能够保证被检测物体的复杂性。PROBOT Anno 机械臂分拣实验结果如图 9 所示，分拣测试结果见表 2。结果表明，本文提出的方法具有较高的分拣成功率。

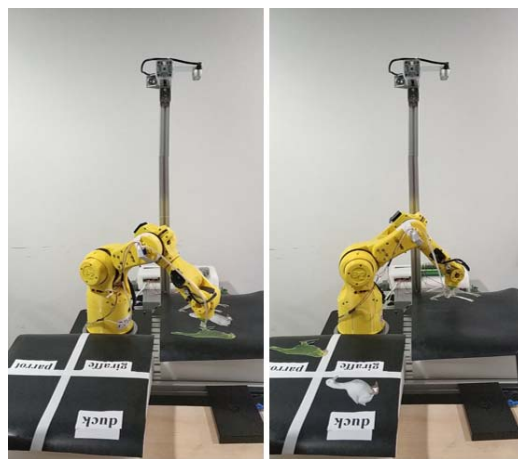


图 9 PROBOT Anno 机械臂分拣实验结果

表 2

分拣测试结果

类别	分拣实验次数/次	成功分拣次数/次	分拣成功率
长颈鹿	100	98	98%
鸭子	100	100	100%
鹦鹉	100	96	96%

6 结束语

本文提出了一种基于轻量型卷积神经网络的机械臂快速分拣方法，并通过 PROBOT Anno 机械臂平台上的分拣实验证明了该方法的有效性。在接下来的工作中，本方法将进一步融入三维视觉处理算法，增强场景和被分拣对象的复杂度，提升算法应对复杂场景的处理能力。

参考文献:

- [1] LI C Y, MA Y B, WANG S Y, et al. Novel industrial robot sorting technology based on machine vision[C]// International Conference on Modelling, Identification and Control. Piscataway: IEEE Press, 2017: 902-907.
- [2] HAZRAT ALI M, AIZAT K, YERHAN K, et al. Vision-based robot manipulator for industrial applications[J]. Procedia Computer Science, 2018, 133: 205-212.
- [3] WANG J W, DOU H Z, ZHENG S K, et al. Target recognition based on machine vision for industrial sorting robot[J]. Journal of Robotics, Networking and Artificial Life, 2015, 2(2): 100-102.
- [4] WANG S Y, LIN H, GAI R L, et al. An application of vision technology on intelligent sorting system by delta robot[C]// IEEE International Conference on E-health Networking. Piscataway: IEEE Press, 2017.
- [5] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507.
- [6] 杜学丹, 蔡莹皓, 鲁涛, 等. 一种基于深度学习的机械臂抓取方法[J]. 机器人, 2017, 39(6): 820-828, 837.
DU X D, CAI Y H, LU T, et al. A robotic grasping method based on deep learning[J]. Robot, 2017, 39(6): 820-828, 837.
- [7] CHEN Z H, ZOU H B, WANG Y B, et al. A vision-based robotic grasping system using deep learning for garbage sorting[C]// 2017 36th Chinese Control Conference. Piscataway: IEEE Press, 2017: 11223-11226.
- [8] KUMAR R, KUMAR S, LAL S, et al. Object detection and recognition for a pick and place robot[C]// Asia-Pacific World Congress on Computer Science & Engineering. Piscataway: IEEE Press, 2014: 1-7.
- [9] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[J]. European Conference on Computer Vision, 2016: 21-37.
- [10] LI Z M, PENG C, YU G, et al. Light-head R-CNN: in defense of two-stage object detector[J]. arXiv preprint, 2017, arXiv: 1711.07264.
- [11] HOWARD A G, ZHU M L, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint, 2017, arXiv: 1704.04861.

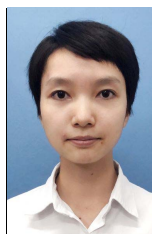
[作者简介]



田思佳（1994-），女，华中科技大学硕士生，主要研究方向为机器人与人工智能。



顾强（1987-），男，博士，武汉精锋微控科技有限公司总经理，主要研究方向为运动控制。



胡蓉（1996-），女，华中科技大学硕士生，主要研究方向为机器人与嵌入式运动控制。



李锐戈（1994-），男，华中科技大学硕士生，主要研究方向为软硬件协同设计。



何顶新（1966-），男，华中科技大学教授，主要研究方向为计算机控制技术、嵌入式系统、机器人控制技术等。