基于Jaco机械臂的抓取行为研究

摘要：

1 引言

对于现实生活中的很多物体，人只需要用眼睛去看就可以准确的判断它的位置进行抓取，而对于机器人在抓取前需要给出物体的准确位姿，因此视觉是抓取前需要解决的首要问题，尤其机器视觉目标检测技术中的准确性和实时性以及仿人行为的目标抓取是机器人亟待解决的问题。

本文以Jaco机械臂构建的服务机器人系统为对象，通过对目标检测技术和机械臂抓取技术的研究，采用了YOLOv3算法进行实时检测，控制机械臂对目标物体进行抓取。服务机器人系统主要内容如下：

1. 比较了当前目标检测中几种不同的算法，最终采用YOLOv3实现本次实验的目标检测。通过对常见物体的实时目标检测实验，验证了YOLOv3算法目标检测的实时性和准确性。
2. 构建了服务机器人平台，设计了实时目标检测算法和抓取算法软件模块。通过实际抓取实验及数据分析，验证了服务机器人系统抓取的实时性与有效性。
3. 目标检测

目标检测的最终目的是从一幅场景或者图片中找出需要的目标物体，主要过程包括目标检测和目标识别两个部分，主要任务是要检测出图片上的物体，标注出来并且分析判断该物体具体是什么。目标检测技术的难点在待检测区域中目标物的提取与识别。目标检测技术目前已经由传统目标检测算法，经过基于区域提名的深度学习目标检测算法，发展至基于回归的深度学习目标检测算法。

深度学习已经广泛的应用到目标检测中，一个经典例子就是基于区域的卷积神经网络（R-CNN）。该算法首先使用选择性搜索（SS）方法在整个图像中搜索候选区域，然后使用CNN（卷积神经网络）模型在每个候选区域上提取特征并将其转换为固定长度的特征向量，解这将特征向量输入到特定种类的线性向量机中分类并得到与候选区域对应的类别分数，最后根据类别分数的高低和非极大值抑制方法选择出最佳边界框。然而，R-CNN的不足之处在于模型对每个候选区域都要输入到CNN中提取特征，并且物体类别和边界框不是同时生成，因此模型计算量较大，不能达到实施效果。

Fast-RCNN在RCNN的基础上，采用自适应尺度池化对整个网络进行优化，从而规避了R-CNN中冗余的特征提取操作，提高了网络识别的准确率。它首先将整个图像输入到卷积层和池化层中生成特征图，接着使用ROI（感兴趣区域）

池化层处理特征图并使其生成一个固定长度的特征向量，然后将特征向量输入到全连接层分别进行2个不同的处理：一个是输入到softmax分类器中对候选区域分类，另一个是进行边界框的回归，以此产生对应此区域的边界框参数。最后，通过非极大值抑制产生最终结果。Fast-RCNN的优点是对整个图像只进行一遍卷积操作，无需存储训练和测试过程产生的中间值,因此其速度相较于 R-CNN 大为提升，并且目标类别和边界框是一同输出的。

虽然Fast-RCNN的训练过程有所简化，检测效果有所提升，但是仍达不到实时要求，原因在于提取候选区域的效率仍旧不高，因而Faster R-CNN模型被提出。Faster R-CNN通过构建区域建议网络(Region Proposal Network，RPN)提取候选框,取代时间开销大的选择性搜索方法，区域提名、分类、回归等操作一起共用卷积特征,进一步提升了速度。

由于机械臂的闭环动态抓取对物体检测的实时性要求比较高，而faster-rcnn不能满足我们的实验要求，因此我们采用YOLOv3算法来解决此问题。YOLO系列目标检测算法不同于RCNN系列,它是一个one-stage的目标检测算法 。YOLO直接在输出层回归bounding box的位置和bounding box所属类别，因此，我们不需要额外的分支来提取候选区域，实现了one-stage，运行速度有了极大的提高。YOLO的优势在于检测速度快、背景误检率比 R-CNN 等要低、支持对非自然图像的检测。而YOLOv3在YOLOV1,YOLOV2的基础上，又做了改进。主要的改进有：调整了网络结构；利用多尺度特征进行对象检测；对象分类用logistic取代了softmax，因此YOLOV3相比于YOLOv1,YOLOv2运行速度又有了提高，采用此算法，能很好地满足本次实验的实时性要求。

* 1. 目标检测实验

Kinect v2 标定后，检验YOLOv3算法对常见物体的实时目标检测情况。首先，在实验桌上放置一个橙子，用YOLOv3目标检测算法实时检测桌面情况，随机截图，如图1所示，然后在实验台上放置一个橙子、一个苹果、一根香蕉、一瓶瓶装水，使用YOLOv3模型检测实验桌上的这些物体，如图2所示。

After the Kinect v2 calibration, the real-time target detection of the common object by the YOLOv3 algorithm is examined. First, place an orange on the experimental table, use the YOLOv3 target detection algorithm to detect the desktop situation in real time, random screenshots, as shown in Figure 1, then place an orange, an apple, a banana, a bottle of water on the bench. The objects on the experimental table were tested using the YOLOV3 model, as shown in Figure 2.

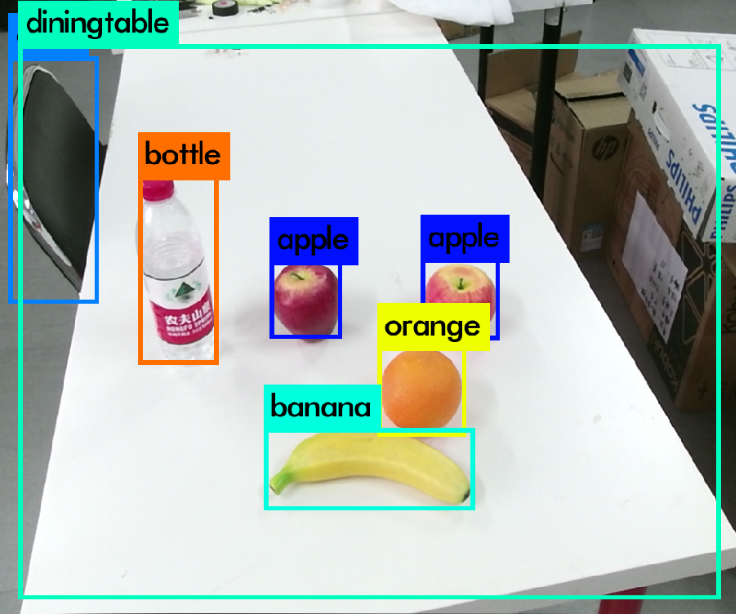
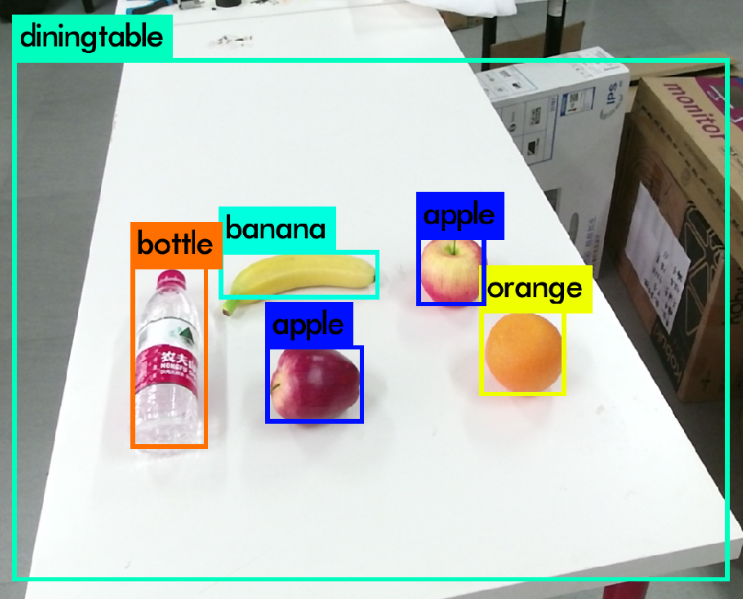
 

图1 单一物体识别(Single object recognition)

图2 多个物体识别(Multiple object recognition)

由实验结果可知，YOLOv3算法能准确使用不同颜色的边界框框出目标物体，识别了桌面小型常见物体的种类，并在边界框的上方显示物体正确的英文名称。

According to the experimental results, the YOLOv3 algorithm can accurately use the bounding boxes of different colors to frame the target object, identify the types of common objects on the desktop, and display the correct English name of the object above the bounding box.

（二）准确率结果分析

在实验场景下，反复移动实验所选用的五个物体各50次，观察其识别的准确率。当对其中某个物体进行测试时，任意移动目标物体，而保持其他剩下的物体的位置不变。YOLO算法能够成功框出目标物体且正确检测其名称。物品识别率见表1。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 框正确次数 | 框正确率 | 名称正确次数 | 名称正确率 |
| 苹果1 | 48 | 96% | 48 | 96% |
| 苹果2 | 46 | 92% | 46 | 92% |
| 橙子 | 48 | 96% | 48 | 96% |
| 香蕉 | 49 | 98% | 49 | 98% |
| 农夫山泉 | 46 | 92% | 45 | 90% |

表1 目标检测成功率实验结果

由实验结果分析可知，基于YOLOv3算法的目标检测对常见物体的苹果、香蕉、瓶子等的边界框与识别准确率较高。后续可以继续训练神经网络，进一步提高算法的识别准确率。

1. 系统平台搭建及抓取实验(System platform construction and Grasping experiment)

机器人系统平台由Kinect、Jaco机械臂和底盘组成，抓取实验是基于机器人实验平台，在ROS系统上进行相关功能模块的设计，并根据Kinect v2摄像头采集的信息完成机械臂抓取的测试。

The robot system platform consists of Kinect, Jaco robot arm and chassis. The grasping experiment is based on the robot experiment platform. The relevant function modules are designed on the ROS system, and the robot arm grabbing test is completed according to the information collected by the Kinect v2 camera.

* 1. ROS系统下的物体抓取实验

实验前，相机与机械臂的坐标关系已标定，即可获得实际抓取点在机械臂基坐标系下的三维位置信息。首先，计算抓取框重心的像素坐标（在相机坐标系下对应的坐标()坐标转换为

，

其中为相机的 内参数，为坐标点对应的深度值。其次，计算相机坐标系下的坐标在机械臂基坐标系下对应的坐标，坐标转换为



其中为标定得到的手眼关系矩阵，即为目标物体在机械臂基坐标系下的位置。

* + 1. 单一物体识别抓取实验(Single Object Recognition Grasp Experiment)

选择一个苹果放置于实验桌上，将物体随意摆放在桌面上。搭建该场景的实验平台，成功标定Kinect v2后，在系统中输入抓取苹果放置于指定置物筐的命令，服务机器人系统立即执行抓取放置任务。服务机器人系统识别并抓取苹果的过程分解图如图3所示，以视屏每2秒截图展示服务机器人的连续抓取过程。图3（a）所示，机械臂开始运动，图3（d）中显示机械臂第7秒成功抓取苹果，图3（g）所示，第13秒机械臂将苹果放于置物筐中。

Select an apple on the lab table and place the object on the table. After setting up the experimental platform of the scene, after successfully calibrating Kinect v2, input the command to grab the apple placed in the specified storage basket in the system, and the service robot system immediately performs the crawling placement task. The process decomposition diagram of the service robot system to identify and capture Apple is shown in Figure 3. The screen captures the continuous crawling process of the service robot every 2 seconds. As shown in Fig. 3(a), the robot arm starts to move. Figure 3(d) shows that the robot arm successfully grabs the apple in the 7th second. As shown in Fig. 3(g), the robot arm puts the apple in the storage basket in the 13th second.

图片包含 室内, 地板, 餐桌

描述已自动生成 图片包含 室内, 餐桌, 地板, 墙壁

描述已自动生成 图片包含 室内, 地板, 餐桌

描述已自动生成

1. 第1秒 (b) 第3秒 (c) 第5秒 (d) 第7秒

图片包含 地板, 室内, 人员

描述已自动生成 图片包含 地板, 室内, 人员, 男士

描述已自动生成图片包含 地板, 室内

描述已自动生成

(e) 第9秒 （f）第11秒 （g） 第13秒 （h）第15秒

图片包含 地板, 室内

描述已自动生成图片包含 地板, 室内

描述已自动生成图片包含 室内, 地板, 餐桌

描述已自动生成

图3单个物体抓取实验(Single object grasp experiment)

* + 1. 复杂环境下目标物体抓取实验 ( Target object grasp experiment in complex environment)

选择不同物品放置于实验桌上，将物体随意摆放在桌面上，使物体不要相互紧挨着。因为机械臂手有一定大小，如果中间碰撞到其他的水果，将视为抓取失败。

Select different items to place on the experiment table, and place the objects on the table at random so that the objects do not cling to each other. Because the robot arm has a certain size, if the middle collides with other fruits, it will be regarded as a crawl failure.

搭建此场景的实验平台，成功标定Kinect v2后，先选择桌面上苹果作为目标抓取物，在系统中输入抓取苹果放置于指定筐的命令，服务机器人系统立即执行抓取苹果任务。成功抓取苹果置于置物筐后，然后在系统中输入抓取橙子放置于指定筐的命令，服务机器人系统立即执行抓取橙子任务。

After setting up the experimental platform of this scene, after successfully calibrating Kinect v2, first select Apple on the desktop as the target grab, input the command to grab the apple placed in the specified basket in the system, and the service robot system immediately executes the task of grabbing the apple. After successfully grabbing the apple in the storage basket, and then inputting the command to grab the orange in the specified basket in the system, the service robot system immediately performs the task of grabbing the orange.

在抓取过程中，实时读取Kinect v2图像数据，可以看出，YOLOv3算法准确识别桌面上苹果、橙子、香蕉与瓶装水，即使在抓取过程中，目标检测算法仍能够检测出机械臂抓取的物体以及被遮挡的物体，进一步验证了该目标检测算法对于本次实验抓取的实时性要求。服务机器人系统先抓取苹果再抓取橙子的过程分解图如图3所示，下面截图展示服务机器人的连续抓取过程。

In the process of crawling, the Kinect v2 image data is read in real time. It can be seen that the YOLOv3 algorithm accurately recognizes apples, oranges, bananas and bottled water on the desktop. Even during the crawling process, the target detection algorithm can detect the mechanical arm grasping. The object taken and the occluded object further verify the real-time requirements of the target detection algorithm for this experiment. The process decomposition diagram of the service robot system first grabs the apple and then grabs the orange is shown in Figure 3. The screenshot below shows the continuous crawling process of the service robot.

图片包含 室内, 地板, 人员, 墙壁

描述已自动生成图片包含 室内, 餐桌, 墙壁

描述已自动生成图片包含 室内

描述已自动生成图片包含 室内, 地板, 餐桌

描述已自动生成图片包含 室内, 餐桌, 地板, 墙壁

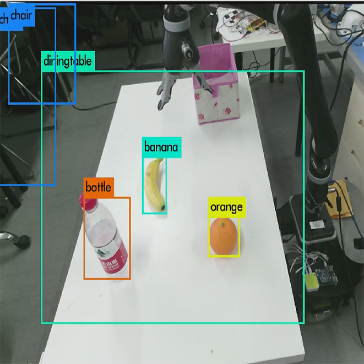
描述已自动生成图片包含 室内, 墙壁, 餐桌

描述已自动生成

图片包含 室内, 地板, 墙壁, 餐桌

描述已自动生成图片包含 室内, 地板, 餐桌, 墙壁

描述已自动生成

图片包含 室内, 地板, 餐桌, 墙壁

描述已自动生成图片包含 室内, 地板, 物品

描述已自动生成

图片包含 室内, 地板, 餐桌, 用具

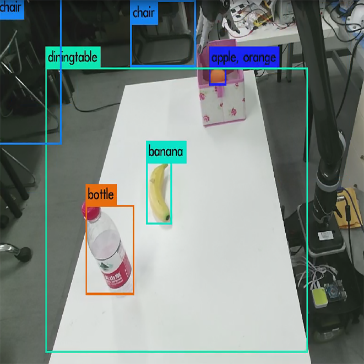
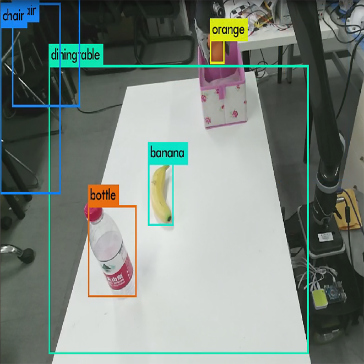
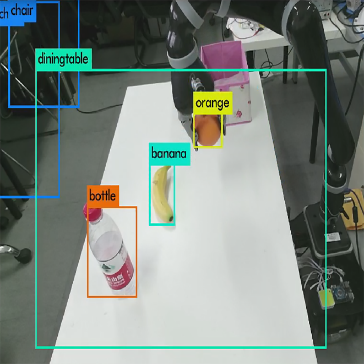
描述已自动生成

图4 复杂场景下的连续抓取全过程视频截图

Continuous grasp full-process video screenshot in complex scenes

复杂环境下的目标抓取实验先将苹果视为目标物体，再将橙子视为目标物体，实验干扰物为香蕉与瓶装水。通过前后两次实验验证可知，本文设计的服务机器人系统不仅能够满足单一场景下的物体识别和抓取，也可以满足在复杂场景下的目标物体的准确识别与实时抓取。

Target capture experiments in complex environments first consider apples as target objects, then oranges as target objects, and experimental interferences are bananas and bottled water. It can be seen from the two experiments before and after. The service robot system designed in this paper can not only meet the object recognition and capture in a single scene, but also meet the accurate recognition and real-time capture of target objects in complex scenes.

\*

* + 1. 抓取成功率实验

选择同样场景，选择各种常见物体进行抓取实验，针对同一个物体，随机摆放于桌面上任意位置，对机器人系统下达抓取并放置于置物筐的指令。对所选五种物体任意位置实验50次，判断本文设计的算法对不同物体的抓取成功概率，以及不同物体定位的准确性。

表2 抓取成功率实验

|  |  |  |  |
| --- | --- | --- | --- |
|  | 实验次数 | 成功次数 | 成功率 |
| 橙子 | 50 | 46 | 92% |
| 苹果 | 50 | 45 | 90% |
| 青椒 | 50 | 44 | 88% |
| 梨 | 50 | 41 | 82% |
| 香蕉 | 50 | 33 | 66% |

通过表2可知，橙子、苹果和青椒的抓取准确度良好，梨抓取成功率相对较低，香蕉的准确率未达到预计状态，分析由于所选用的梨为上小下大形状，体积较小，重心低。而香蕉为长条状物体，只能横置于桌面，由于机械臂指尖的局限性，抓取的成功性较低，本实验所设计的抓取策略并不太适用于重心较低或狭长型物体，要提高这类型物品抓取准确性，需要考虑优化抓取策略或采用抓取位姿。

1. 结论

本文搭建了一个机械臂抓取系统，可以在多物体场景下实现机械臂的动态闭环抓取。通过实验验证了本文方法能对图像中的目标物体进行较为准确的分类，在jaco机械臂上进行了抓取实验，实验结果证明了目标检测方法的有效性。

未来的研究方向是用一个统一的架构实现物体分类和抓取检测，能够将2种网络模型更加有效地结合成一个整体，最终同时进行目标检测和抓取位置检测。

In this paper, a mechanical arm grabbing system is built, which can realize the dynamic closed-loop grasping of the robot arm in a multi-object scene. The experimental results show that the proposed method can accurately classify the target objects in the image, and the grabbing experiment is carried out on the jaco robot arm. The experimental results demonstrate the effectiveness of the target detection method.

   The future research direction is to realize object classification and grab detection with a unified architecture, which can combine the two network models into one whole more effectively, and finally perform target detection and capture position detection at the same time.

相关工作

目标检测的任务是找出图像中所有感兴趣的目标，确定它们的位置和类别。由于各类物体有不同的外观、形状、姿态，加上成像时光照，遮挡等因素的干扰，目标检测一直是计算机使君领域最具有挑战的问题。With the advantages of deep learning, object detection study has yielded many great achievements.

基于深度学习的目标检测方法的关键一个组成是基于region-based 网络，如rcnn, fast r-cnn and faster r-cnn，原始的rcnn使用滑动窗口方法检测图像中的物体的类别和位置，由于神经网络需要处理大量的图像块，这种方法需要很大的计算量，为了提高检测速度，Fast R-CNN

利用共享卷积层，现在不是每一个候选框都当做输入进入CNN了，而是输入一张完整的图片，只卷积一次，每个框找自己的对应映射。Although it can significantly decrease the computation cost, it still needs to process a great many candidates of object detection. Ren et al. [22] proposed the faster R-CNN based on region proposal network to generate the region proposals in an efficient and accurate way. 这种方法可以极大的减少region proposal阶段的计算时间. However, the region-based detection methods cannot detect objects by processing an image only once.

最近，one-stage 目标检测网络such as you only look once (YOLO) [23] and single shot multibox detector (SSD) [13]，在计算效率和物体检测的准确率上面都取得很好的表现。不同于RCNN系列，YOLO[23]再也不需要类似的region proposal提取的步骤，所有的一切过程是单流程的，直接在一张图上面进行处理（前向处理），将图像的位置设计为一个可以通过回归输出的参数作为最终网络输出的的一部分，同类别向量一起输出，极大的减少了运行时间。在这些优势的基础上以及本次实验的要求，we utilize the yolo approach in our work for object detection.

accuracy of the target detection algorithm of the design service robot and the effectiveness of the human-like behavior capture strategy are verified.

Regardless of a single scene or a complex interference scene, the YOLO algorithm can quickly frame the target object with the bounding box, detect the target object, and use the

Finally, in order to verify the probability of successful capture of different objects by the algorithm designed in this paper, the same scene is selected, and the selected five objects are tested 50 times at any position. After experimental statistics, the algorithm has good grasping accuracy for oranges, apples and green peppers[19]. The success rate of pear grabbing is relatively low, and the accuracy of bananas is not as expected. The analysis is based on the shape of the pears. The volume is small and the center of gravity is low. The banana is a long object, which can only be placed on the tabletop. Due to the limitation of the fingertip of the robotic arm, the success of the grasping is low. The grip designed by this experiment is not suitable for the lower center of gravity or the narrow objects, in