计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2019,28(1):94−99 [doi: 10.15888/j.cnki.csa.006748]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于 SSD_MobileNet 模型的 ROS 平台目标检测①

童　星, 张　激

(中国电子科技集团公司第 32 研究所, 上海 201808)
通讯作者: 童　星, E-mail: chit_chat@foxmail.com

摘　要: 目标检测是机器人技术领域中重要的技术环节, 而作为机器人开发领域中最受欢迎的平台之一, ROS (Robot Operating System) 平台实现快速准确的目标检测功能是非常必要的. 目前深度学习方法是实现目标检测功能的核心技术, 但当前 ROS 平台自带的目标检测数据包实现原理仍是基于传统的局部图像特征描述方法, 目标检测鲁棒性差, 泛化能力弱. 本文就将针对以上问题, 提出一种基于 SSD_MobileNet 框架, 结合独立制作的图像数据集训练定制的目标检测模型, 并将模型集成到 ROS 平台实现快速准确的目标检测功能.

关键词: 目标检测; 机器人技术; ROS 平台; 深度学习; SSD_MobileNet

引用格式: 童星,张激.基于 SSD_MobileNet 模型的 ROS 平台目标检测.计算机系统应用,2019,28(1):94–99. http://www.c-s-a.org.cn/1003-3254/6748.html

Object Detection of ROS Platform Based on SSD_MobileNet Model

TONG Xing, ZHANG Ji

(The 32nd Research Institute of China Electronic Technology Group Corporation, Shanghai 201808, China)

Abstract: Object detection plays a key role in robotics, and as one of the most popular platforms for robot development, it is very necessary for ROS (Robot Operating System) platform to achieve fast and accurate object detection. Recently, the deep learning method is the core technology to realize the object detection function, but the object detection packets carried by the ROS platform are still based on the traditional local image feature description method, which with poor robustness and weak generalization ability. Aiming at the above problems, we propose a customized object detection model based on SSD_MobileNet framework, which combines the image dataset independently, and integrate the model to the ROS platform to achieve fast and accurate object detection function.

Key words: object detection; robot technology; ROS platform; deep learning; SSD_MobileNet

目标检测作为计算机视觉的重要研究方向, 被广泛应用于无人驾驶、智能摄像头、人脸识别等新兴科研领域. 它是以图像分类技术为基础, 对图像中目标对象进行识别分类, 并且在目标对象周围绘制适当大小的边界框对其进行定位实现的[1]. 从 2012 年深度学习算法 AlexNet 获得 ILSVRC (ImageNet Large Scale Visual Recognition Challenge) 图像分类比赛的冠军后, 深度学习算法在图像处理和目标检测技术应用中不断取得重大突破, 其中卷积神经网络 (Convolution Neural Network, CNN) 和候选区域 (Region Proposal) 算法起到了关键性作用. 目前很多成熟的目标检测深度学习算法在检测精度和速度上有着非常不错的表现, 以 SSD[2] (Single Shot MultiBox Detector) 算法为例, 使用 VOC2007 数据集在 NVIDIA Titan X 上测试, mAP (mean Average Precision) 可以达到 74.3%, 速度达到每秒 59 帧.

在机器人技术领域中, 目标检测同样具有重要作用. 它是机器人完成诸多智能行为的必要前提, 例如机

① 收稿时间: 2018-07-30; 修改时间: 2018-08-27; 采用时间: 2018-08-29; csa 在线出版时间: 2018-12-26

械手臂的智能抓取和无人车的智能避障. 而 ROS (Robot Operating System) 作为目前最受欢迎的机器人应用开发平台之一, 实现快速准确的目标检测功能具有十分重要的实际意义. ROS 应用开发平台提供了消息传递、分布式计算、代码重用等优势功能, 具有跨平台、模块化、集成度高和社区活跃等特点[3]. 目前 ROS 平台自带一些实现目标检测功能的数据包, 主要是基于传统的局部图像特征描述方法实现的. 这些传统方法与深度学习算法相比, 目标检测的鲁棒性和泛化能力上存在明显的差距.

本文将首先介绍 ROS 平台实现目标检测功能的原理, 并选取典型的目标检测数据包, 实现完整的目标检测过程, 分析实验结果; 然后通过制作用于目标检测的图像数据集, 结合 SSD_MobileNet 预训练模型, 在 NVIDIA JETSON TX2 开发板上训练定制的目标检测模型, 并将训练好的模型集成到 ROS 平台, 实现目标检测功能, 并与 ROS 平台自带数据包实现的目标检测效果进行对比, 得出结论.

1 ROS 平台目标检测功能

传统的 ROS 平台目标检测功能主要是基于局部图像特征描述方法实现的. ROS 平台通过摄像头等设备获取包含目标物体的图像场景, 运用局部图像特征描述方法提取目标对象的特征, 并根据提取的特征去目标场景中检测和识别目标对象. 本章节首先介绍局部图像特征描述的实现原理, 然后以 find_object_2d 数据包为例介绍 ROS 平台目标检测功能的实现过程, 并分析实验结果.

1.1 局部图像特征描述

局部图像特征描述主要功能是寻找图像中的对应点以及完成物体特征描述, 目前在三维场景构造、物体识别、图像拼接和配准等领域应用广泛[4]. 局部图像特征描述的核心有两点: 不变性和可区分性, 不变性是指对图像变化情况下的处理能力, 而可区分性指的是对图像中不同对象特征的区别能力[5]. 好的局部图像特征描述算法应同时具备好的不变性和可区分性, 代表性的算法有 SIFT (Scale Invariant Feature Transform)、SURF (Speeded Up Robust Features) 等. 由于图像位置变化多样性、光线变化多样性、视角多样性等因素, 使得通过传统的局部图像特征描述方法提取鲁棒的物体特征十分困难. 同时局部图像特征描述方法提取局部纹理特征时丢失的全局信息, 也让目标检测的泛化能力十分有限.

1.2 find_object_2d 目标检测数据包

find_object_2d 是 ROS 平台中实现目标检测功能的典型数据包, 它具有简单的 Qt 图形界面, 通过调用 OpenCV 库的 SURF、SIFT 等局部图像特征描述子实现目标检测功能[6]. find_object_2d 包目标检测功能的实现过程如图 1 所示. 首先启动 find_object_2d 包的目标检测节点, 节点会订阅由摄像头等图像获取设备发布的图像会话, 获得图像场景信息, 并打开 Find-Object 窗口. 在该窗口中可以直接提取图像场景中目标对象的特征, 然后在目标检测阶段, find_object_2d 包根据提取的对象特征在图像场景中检测目标对象, 并在检测到目标对象之后, 在图像场景中用适当大小的边界框标识出检测到的目标对象. find_object_2d 包实现目标检测的框架如图 2 所示. 调用 USB 摄像头驱动启动节点/usb_cam, 发布的/usb_cam/image_raw 会话由/find_object_2d 节点订阅, 进而实现目标检测功能.

(a) USB摄像头捕获的图像场景 (b) 自动提取对象特征 (c) 检测目标对象

图 1 find_object_2d 包目标检测功能的实现过程

1.3 实验结果分析

以 find_object_2d 包为代表的实现 ROS 平台目标检测功能, 在目标对象特征提取过程中不需要大量的图像数据集即可完成特征提取, 但是在目标对象检测

过程中, 对于目标物体因视角、光照强度等因素照成的图像变化影响鲁棒性差. 同时由于单张图像提取局部特征的限制, 对于外形近似的同类别对象也无法检测和识别, 实验效果如图 3 所示.

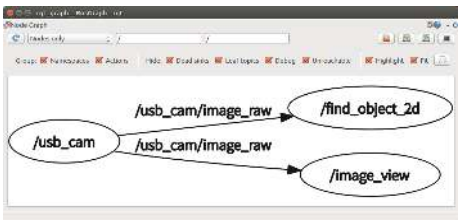


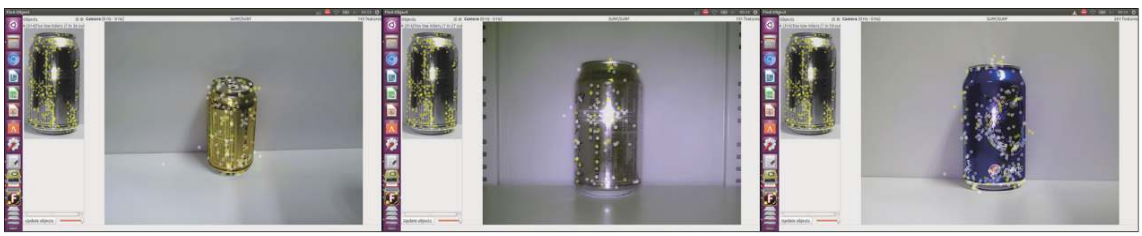
图 2 find_object_2d 包实现目标检测的框架

2 深度学习算法实现 ROS 平台目标检测

深度学习算法应用于目标检测技术尽管时间不长, 但是效果显著, 并且不断取得进步与突破. 从使用卷积神经网络处理多尺度滑窗的 OverFeat 算法[7], 到以 R-CNN[8−11]算法 (包括 R-CNN, SPP-NET, Fast R-CNN, Faster R-CNN 等) 为代表的结合卷积神经网络分类和候选区域的目标检测算法, 再到以 YOLO[12](You Only Look Once) 和 SSD 为代表的将目标检测转化为回归问题的算法, 深度学习已经成为目标检测的核心, 推动着目标检测技术的快速发展. 本次实验选择基于 SSD 框架结合 MobileNet 网络结构的目标检测模型, 本章将首先介绍 SSD 模型的基本结构和实现目标检测的原理, 再分析结合 MobileNet 网络作为特征提取模块的优势, 最后介绍完整的模型训练以及集成到 ROS 平台的过程.

2.1 SSD_MobileNet 模型

目前基于候选区域的深度学习目标检测算法效果令人满意, 以 Faster R-CNN 算法为例, 在 VOC2007 测试集测试 mAP 达到 73.2%, 但是测试速度仅为每秒 5 帧[11]. 这是因为预先获取候选区域, 然后对每个区域进行分类处理计算量非常巨大, 无法达到实时的目标检测效果. 于是便孕育了一类使用回归思想的目标检测算法, 既给定输入图像, 直接在图像的多个位置上回归出这个位置的目标边框以及目标类别. YOLO 算法是将目标检测任务转换成回归问题的代表性算法[12], 它将图像分割为若干网格, 每个网格输出若干包围盒, 包含目标置信度的值, 同时每个网格预测各自的类别信息, 测试时通过每个网格的类别概率和包围盒的置信度即可实现目标检测. 但是 YOLO 对于目标对象的尺度比较敏感, 对尺度变化较大的物体泛化能力较差, 即位置预测不够精确. 同时因为每个网格只检测一个目标对象, 所以容易造成误检. 针对以上问题, SSD 目标检测方法提供了很好的解决方案, 它结合 YOLO 的回归思想和 Faster R-CNN 算法中的 anchor 机制, 在保证了实时性前提下提高了目标检测的精度.



(a) 视角调整为俯角，无法对相同对象实现目标检测

(b) 光线调暗，无法对相同对象实现目标检测

(c) 更换外形近似的同类对象，无法实现目标检测

图 3 实验结果展示

SSD 模型是由 Wei Liu、Dragomir Anguelov 等人[2]提出的使用单个深层神经网络检测图像中对象的方法. SSD 模型的结构如图 4 所示: 前五层为 VGG-16 网络的卷积层, 第六和第七层全连接层转化为两个卷积层, 后面再添加三个卷积层和一个平均池化层. 基础网络结构为 VGG-16 卷积神经网络的主体, 并连接多层卷积层和池化层作为额外特征提取层. SSD 同样采用回归方法获取目标对象位置和类别, 不同的是 SSD 使用的是目标对象位置周围的特征而非全图的特征. SSD 各卷积层将特征图分割为若干相同大小的网格, 称为 feature map cell, 对每个网格设定一系列固定大小的包围盒, 称为 default boxes. 然后分别预测 default boxes 的偏移以及类别得分, 最终通过非极大值抑制方法得到检测结果. default boxes 的作用类似于 Faster R-

CNN 的 anchor 机制, 不同的是 default boxes 作用于不同层次的多个特征图上, 这样可以利用多层的特征以最佳尺度匹配目标对象的实际区域 (groudtruth).

SSD_MobileNet 模型是使用 MobileNet 网络代替 VGG 网络作为基础网络结构. MobileNet 是 Andrew G. Howard 等人提出的适用于嵌入式视觉应用的高效模型. MobileNet 的主要特点是用深度级可分离卷积替代传统网络结构的标准卷积来解决卷积网络的计算效率低和参数量巨大的问题. Andrew G. Howard 等人在实验中对比了基于 SSD 框架下, VGG 模型和 MobileNet 模型使用 COCO 数据集训练及测试的结果[13], 如表 1 所示. 可见 SSD 框架结合 MobileNet 网络结构实现目标检测尽管检测准确率略有下降, 但计算量和参数量大幅减少. 对于机器人等嵌入式平台应用来说, 硬件资源有限, 使用 MobileNet 这样的轻量级、低延迟的网络模型能够有效地提高目标检测的实时性.

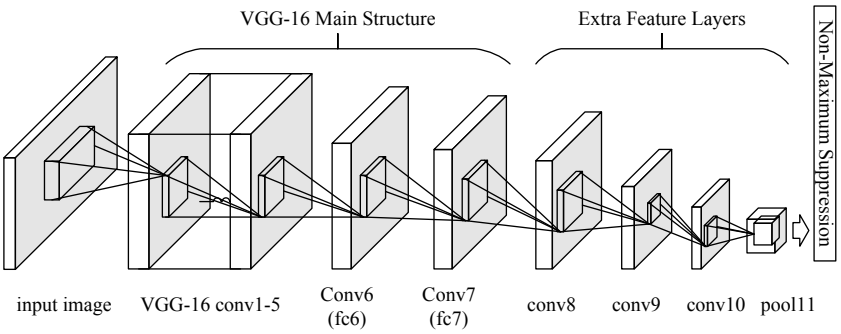


图 4 SSD 模型的结构图

表 1 SSD-VGG 和 SSD-MobileNet 测试结果对比

Framework resolution	Model	mAP (%)	Billion Mult-Adds	Million parameter
SSD	VGG	21.10	34.9	33.1
	MobileNet	19.30	1.2	6.8

2.2 模型的训练与集成

Google 推出的 TensorFlow[14]是目前最受欢迎的深度学习框架之一, 而开放的 Object Detection API 是基于 TensorFlow 构造的开源代码示例库, 通过 Object Detection API 能够对一些大型且成熟的目标检测模型进行高效的重复利用, 结合特定的图像数据集即可训练具有定制功能的目标检测模型. 本次实验将通过 Object Detection API, 基于 Google 预训练 SSD_MobileNet 模型, 结合自己搜集制作的图像数据集, 训练定制的目标检测模型, 并集成到 ROS 平台, 实现目标检测功能.

2.2.1 数据集制作

模型训练需要的图像数据集必须为 TFRecord 档案格式. 首先搜集包含目标对象的图片, 实验为了与 ROS 平台 find_object_2d 包实现目标检测实验进行对照, 选择的目标对象类别为易拉罐, 以及与易拉罐相近的杯子和瓶子, 每类图片各 20 张, 共计 60 张图片. 再使用图片标记软件对原始图片进行目标对象位置标记, 并转化为标准的 VOC 目标检测数据集格式: 包含 Annotations、ImageSets、JPEGImages 三个文件, 其中 Annotations 文件存放了每张图片的标注信息, 为.xml 格式; ImageSets 文件记录了分别用于训练、验证和测试的样本名称, 为.txt 格式; 而 JPEGImages 则存放了所有图片, 为.jpg 格式. 最后调用 Object Detection API 库中的数据格式转化程序 create_pascal_tf_record.py, 将数据集转化为 TFRecord 格式.

2.2.2 模型训练

实验中预训练模型版本选择的是 ssd_mobilenet_v1_coco_2017_11_17. 下载模型, 编写模型训练配置文件, 结合制作的图像数据集, 在 NVIDIA JETSON TX2 嵌入式开发板上训练定制的模型. 实验中设置初始学习率为 0.004, 衰减速度和系数分别为 800 720 和 0.95. 训练步数设定为 95 000 步, 使用 TensorFlow 自带的可视化工具 TensorBoard 可查看模型训练情况, 如图 5 所示, 随着训练步数的增加模型的损失率逐渐减小, 并最终接近 1.0. 训练结束之后, 调用 Object Detection API 库中的 export_inference_graph.py 脚本将包含模型结构和参数的临时文件转化为可独立运行的 PB 模型文件.

2.2.3 模型集成

ROS 平台通过定义节点来表示应用程序, 不同节点之间通过预先定义的会话、服务或行为来实现彼此

的通信[15]. 实现目标检测模型的 ROS 平台集成, 需建立如图 6 所示的目标检测节点以及网络连接方式. 首先创建 detect_ros 目标检测节点, 该节点中包含训练好的目标检测模型文件、目标对象的标签文件以及 TensorFlow 的 Object Detection API 库. 节点实时地订阅由 USB 摄像头驱动发布的包含图像场景信息的会话, 以获得原始图像信息, 之后通过调用 OpenCV 库的 cv_bridge 模块将图像信息转化 OpenCV 数据格式, 再使用目标检测模型对 OpenCV 格式的图像数据进行进行推理, 最后将推理的结果实时地发布会话, 订阅该会话即可查看目标检测结果.

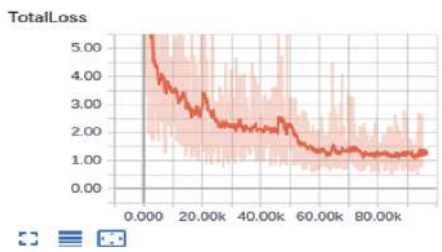


图 5 损失值随模型训练步数的变化

图 6 目标检测节点以及网络连接方式

3 实验结果对比与分析

连接外接 USB 摄像头, 运行目标检测节点, 使用集成到 ROS 平台的 SSD_MobileNet 目标检测模型可以识别出目标对象的类别, 并通过适当大小的包围盒标注目标对象实现定位, 如图 7 所示: 以 81% 的概率推测目标对象类别为 can (易拉罐), 并且以适当大小的包围盒圈住目标对象实现定位. 以该目标检测结果为标准, 设置相关对照实验如图 8 所示, 实验证明 SSD_MobileNet 模型对于不同视角、不同光线强度场景下, 都有着非常好的目标检测鲁棒性, 同时对于同类别对象检测具有一定的泛化能力. 并且可通过训练数据集标签类别的丰富, 使得目标检测算法在单个场景图像中实现多个不同类别目标对象的检测和识别. 实验中训练数据集中包含 can、cup、bottle 三种不同但外形近似的类别, 可以检测出三种类别的对象.

图 7 目标检测结果展示

通过对以上实验结果进行分析, 以及与 find_object_2d 包实现的目标检测结果进行对比, 可以发现深度学习算法, 对于层次信息丰富的图像数据, 具有更好的特征提取和特征表达能力. 使得基于深度学习的目标检测算法能够不受视角变化、光线强弱等因素的影响, 获得鲁棒性强的目标检测效果. 同时伴随着训练数据集的丰富, 深度学习目标检测算法对于同类别的目标对象检测以及单个场景多类别目标对象检测都有良好的表现.

(a) 视角调整为俯角，对相同对象实现目标检测 (b) 光线调暗，对相同对象实现目标检测 (c) 更换外形近似的同类对象，实现目标检测 (d) 实现单个场景多种类别对象的目标检测

图 8 对照实验结果展示

4 结论

本文首先介绍了 ROS 平台目标检测功能的实现原理, 并以 find_object_2d 数据包为例演示了目标检测过程, 通过分析实验结果发现传统的目标检测方法鲁棒性和泛化能力差. 然后介绍了目前流行的深度学习目标检测算法 SSD 的结构及其实现机制, 并分析了 SSD 框架结合 MobileNet 网络结构的实时性优势, 最后基于 Google 的 SSD_MobileNet 预训练模型, 结合制

作的图像数据集, 重新训练定制的目标检测模型, 并集成到 ROS 平台实现目标检测功能. 通过实验结果对比, 证明了深度学习目标检测算法对图像特征提取和表达具有更好的表现, 目标检测的鲁棒性和泛化能力更强.

参考文献

1 郑泽宇, 顾思宇. TensorFlow: 实战 Google 深度学习框架. 北京: 电子工业出版社, 2017.

2 Liu W, Anguelov D, Erhan D, *et al*. SSD: Single shot multibox detector. arXiv preprint arXiv:1512.02325, 2015.

3 Quigley M, Conley K, Gerkey B, *et al*. ROS: An open-source robot operating system. ICRA Workshop on Open Source Software. Kobe, Japan. 2009, 3. 2–4.

4 雷兰一菲, 郎海涛. 几种典型局部图像特征的比较. 计算机应用, 2010, 30(S2): 50–53.

5 樊彬. 局部图像特征描述概述. http://www.sigvc.org/bbs/thread-165-1-1.html. (2012-10-08)[2018-05-30].

6 Joseph L. ROS Robotics Projects. Birmingham: Packt Publishing Ltd, 2017. 171–180.

7 Sermanet P, Eigen D, Zhang X, *et al*. OverFeat: Integrated recognition, localization and detection using convolutional networks. arXiv preprint arXiv:1312.6229, 2013.

8 Girshick R, Donahue J, Darrell T, *et al*. Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA. 2014. 580–587.

9 He K, Zhang X, Ren S, *et al*. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904–1916.

10 Girshick R. Fast R-CNN. Proceedings of 2015 International Conference on Computer Vision (ICCV 2015). Santiago, Chile. 2015, 12, 1440–1448.

11 Ren SQ, He KM, Girshick R, *et al*. Faster R-CNN: Towards real-time object detection with region proposal networks. Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal, Canada. 2015. 91–99.

12 Redmon J, Divvala S, Girshick R, *et al*. You only look once: Unified, real-time object detection. Proceedings of 2016 IEEE Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 779–788.

13 Howard AG, Zhu ML, Chen B, *et al*. MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861, 2017.

14 Abadi M, Agarwal A, Barham P, *et al*. TensorFlow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467, 2016.

15 Joseph L. Mastering ROS for Robotics programming. Birmingham: Packt Publishing Ltd., 2015.