

DOI:10.3969/j.issn.1001-4551.2017.06.026

基于 SSD 网络模型的多目标检测算法

蔡汉明¹, 赵振兴¹, 韩 露¹, 曾祥永²

(1. 青岛科技大学 机电工程学院, 山东 青岛 266042; 2. 北京盛开互动科技有限公司, 北京 100089)

摘要:针对现代化工厂中视觉机器人或智能终端处理多目标检测算法的计算任务繁重、运算速度较慢等问题,将网络通信技术应用到算法处理中进行了在线检测。对 TCP/IP 协议进行了研究,建立了智能终端和云端之间的关系,提出了将智能终端采集到的图像数据进行预处理然后使用基于 TCP 的 Socket 多线程通信方式将图像数据送入云端,在云端的多台计算机上同时使用 SSD 网络模型的多目标检测算法进行了并行处理,并将结果传回智能终端。利用计算机单机与智能终端在线检测在处理时间上进行了对比试验。试验结果表明:在线检测速度稍慢,但已满足实际需求;智能终端在线检测降低了对智能机器人终端硬件的要求,回收的数据可以再利用,并且可以实现算法动态升级。

关键词:目标检测;卷积神经网络;SSD;智能机器人;Socket 网络通信

中图分类号:TP242.6

文献标志码:A

文章编号:1001-4551(2017)06-0685-04

Multiple target detection algorithm based on SSD networks

CAI Han-ming¹, ZHAO Zhen-xing¹, HAN Lu¹, ZENG Xiang-yong²

(1. College of Mechanical and Electrical Engineering, Qingdao University of Science and Technology, Qingdao 266061, China; 2. Beijing Interjoy Technology Limited, Beijing 100089, China)

Abstract: Aiming at the problems of heavy computational tasks and slow speed of operation during visual robots or intelligent terminal handling multiple target detection algorithm in modern factory, network communications technology was investigated. In order to establish the relationship between the intelligent terminal and the cloud, TCP/IP protocol were researched, and this method was proposed that the image data collected from the terminals should be preprocessed and sent to the cloud by Socket multi-thread communication based on the TCP, then the image data should be processed in the cloud with multi-target detection of SSD network model, and the results were returned to the terminal at last. This algorithm was tested on the processing time by computer LAN or intelligent terminal online on the processing time. The results indicate that on-line measuring is slower, but it meets the practical needs. Intelligent terminal online detection reduces the requirements for the robot terminal hardware, and the data recycled can be reused, besides, the method can also make the algorithm dynamically be upgraded.

Key words: target detection; convolutional neural network; single shot multibox detector; intelligent robot; Socket network communication

0 引 言

视觉感知信息量在智能机器人获得所有信息量中占有很高的比重^[1],是智能机器人不可或缺的一部分^[2-3]。因此,处理好计算机视觉方面的问题^[4],精确提取图像中的有效信息对智能机器人的智能化程度起

着至关重要的作用。

由于智能机器人所在的视觉环境可能包含多个不同的目标,需要对其进行多目标检测。目前许多多目标检测算法都是逐步提取图像特征,并用这些特征表示图像信息,通过一些学习方法来训练算法模型^[5]。虽然多目标检测算法应用比较广泛,但还是存在一些

收稿日期:2016-11-13

作者简介:蔡汉明(1960-),男,湖北黄陂人,教授,硕士生导师,主要从事机械制造自动化、计算机辅助设计与制造方面的研究。E-mail: cadm60@163.com

不足之处,尤其是对计算机单机硬件要求比较高。

为了降低对机器人终端硬件的要求,本研究将基于 SSD 网络模型的多目标检测算法和基于 TCP 的 Socket 多线程通信技术结合起来实现在线检测^[6],以解决对机器人终端硬件要求高的问题。

1 卷积神经网络与多目标检测算法

卷积神经网络是人工神经网络的一种^[7-8],目前在字符识别、人脸识别等领域已广泛应用^[9-10],它具有和生物神经网络极其相似的权值共享结构^[11],通过逐步抽取图像中的信息,最终提取图像的有效特征并对图像进行分类或检测。

卷积神经网络的每一层都由多个二维平面组成,每一个二维平面都由上层输入经过卷积或者池化得到^[12]。卷积层是对输入做卷积运算,每一个卷积核与前一层的输入局部相连,并提取该局部区域的特征。池化层是对卷积特征图做下采样(又叫子采样),也就是提取特征图中局部区域中具有代表性的信息来代表整个局部区域。池化层一般跟在卷积层的后面,其特

征图和上一个与其相连的卷积层中的特征图是一一对应。全连接层与卷积层相似,其卷积核的大小和输入的特征图大小相同,故得名全连接。

目前应用较广泛的多目标检测算法的网络模型有 Faster R-CNN^[13]、YOLO^[14]、Single Shot MultiBox Detector 等^[15]。Faster Rcn 采用 anchor 机制,在最后一层卷积特征图上进行滑窗,最终得到 $40 \times 60 \times 9$ 个候选窗口,取前面的一部分(300 个)窗口进行检测,在 GPU 加速的情况下能达到 5 fps,速度上与实际应用还有一定距离;YOLO 将输入图像划分为 $S \times S$ 个网格,然后分别对每个网格进行预测,虽然速度较快,但对于离得较近的物体和小物体的检测效果不好,并且泛化能力偏弱。本研究使用利用不同层上的卷积特征进行多尺度检测的 SSD 网络模型。

2 多目标检测算法的实现

2.1 网络结构

SSD 网络模型是 Faster R-CNN、YOLO 的改进版本,SSD 结构如图 1 所示。

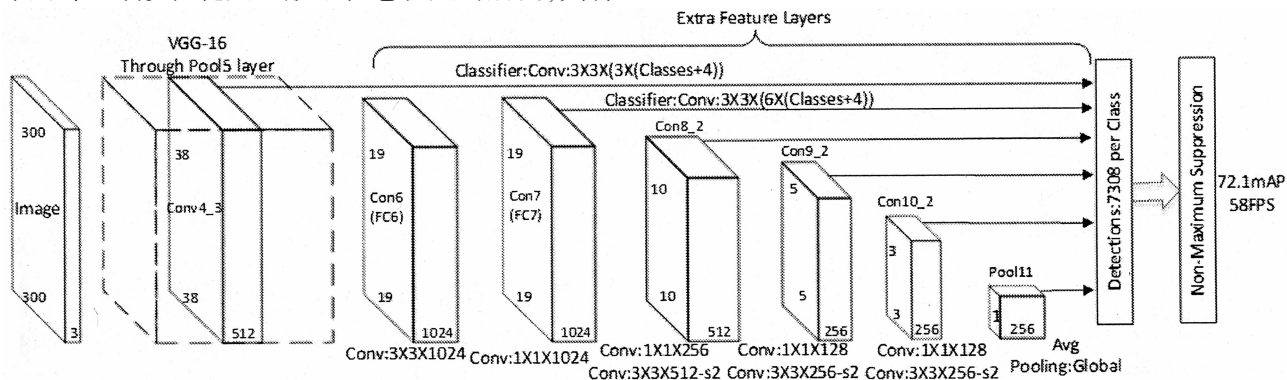


图 1 SSD 结构图

其最左端输入的是一张 $300 \times 300 \times 3$ 的图像,其中 3 代表其 R、G、B 3 个通道。网络中的主分支就是把卷积层、池化层按照一定的顺序串联起来,逐步抽取图片的有效信息。增加的这些额外特征层是把主分支上面的部分特征层组合在一起作为输入,然后用一组卷积核采用滑动窗口的方式对这些输入做卷积运算。其中每一次卷积运算都会得到一组类别的得分或者是相对于默认候选区域的坐标偏移量。最后通过综合这些得到的检测和分类结果,即可得到每个物体在图像中的位置以及图像中的物体类别。

2.2 训练过程

卷积神经网络其实就是一种输入和输出的映射关系。由于该映射关系比较复杂,人为规定的精确数学表达关系不能够满足其需求。因此,只需要把海量的

已知数据对设定好的网络模型进行训练,网络就能够得到这种符合人们需求的映射关系。

SSD 把检测和分类一体化,实现端对端的训练。为了验证该模型的可行性,采用 MSCOCO 数据集,训练过程包括以下 5 个步骤:

- (1) 输入的图像经过前向传播得到图像的基本特征;
- (2) 抽取多层级的特征图并在这些特征图中的各个位置上选取不同大小、不同长宽比的候选区域;
- (3) 计算每个候选区域的坐标位置偏移量和类别得分;
- (4) 根据候选区域和坐标位置偏移量计算出最终区域,再根据类别得分计算候选区域的损失函数,累加得到最后的损失函数;

(5)由最后的损失函数经反向传播过程来修正各个层的权值。

2.3 测试效果

为了测试 SSD 的速度和精度,本次测试使用 MSCOCO 数据集训练的模型,用摄像头进行视频测试,效果截图如图 2 所示。

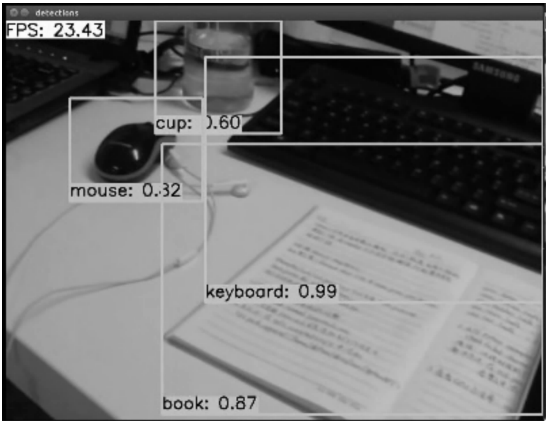


图 2 测试效果图

3 多目标检测算法的应用

由于卷积神经网络网络结构比较复杂,运算量巨大,将该算法放在机器人或智能终端上,需要价格昂贵的硬件。为了降低其对机器人终端硬件的要求,可以采用以下步骤:①图像数据的预处理;②使用基于 TCP 的 Socket 多线程通信将数据送入云端,在云端采用 SSD 网络模型的多目标检测算法得到结果后,返回给智能终端实现在线检测。

3.1 图像数据的预处理

为了减少图片上传所用时间,就要对图片在上传到服务器之前进行预处理,即是对图像进行降维操作^[16],分为裁边和缩放,如图 3 所示。



图 3 图片预处理

由于图像检测的网络对图片要求长宽比为 1:1,而机器人采集到的图片数据长宽比不满足该要求,需要裁掉图像中外围边框部分,如图 3 中 crop 所示。一般以图片的中心为基准裁剪尽可能大的 1:1 的图片区域。

另外,机器人视野中远距离的物体对机器人的行为影响不大,加上高清图片的数据在上传过程中对网

络要求比较高,在一定范围内对裁剪后的图像进行缩放处理(例:缩放后大小为 300 × 300)。

3.2 在线检测

Socket 是 Internet 网络编程的通用 API,在 windows 和 linux 等平台下都有相应的实现版本。Client/Server 模型是最基本的 Socket 通讯模型^[17-18],本研究在服务器上同时开两个 Socket(R_Socket 和 W_Socket),R_Socket 供机器人终端通过公共网络来访问连接,W_Socket 供服务器端的处理计算机通过局域网来访问连接,为了加快处理速度,每个 Socket 都采用多线程非阻塞模式处理。所有的 Client 端采用实时连接,当机器人或处理计算机开始运行后,立即连接到服务器。

图片经过预处理后,机器人终端通过 R_Socket 连接把图片上传到服务器,把该图片添加到服务器的任务列表中,并间隔检测该任务的处理状态。

对于服务器端来说,当检测到任务列表里面的新任务之后,在保存处理计算机的状态列表里检索一个处于空闲状态的处理计算机,并把该任务发送给该处理计算机。处理计算机接收到该任务后,通过局域网共享读取该图片文件并处理。处理完成后把处理结果通过 W_Socket 发送给服务器。服务器接收到处理结果后,把该结果放到任务列表中对应的位置,并改变该任务的处理状态。当检测该任务的线程检测到该任务已经处理完成时候,取出该任务结果,并通过 R_Socket 发送到智能机器人终端。在线检测流程图如图 4 所示。

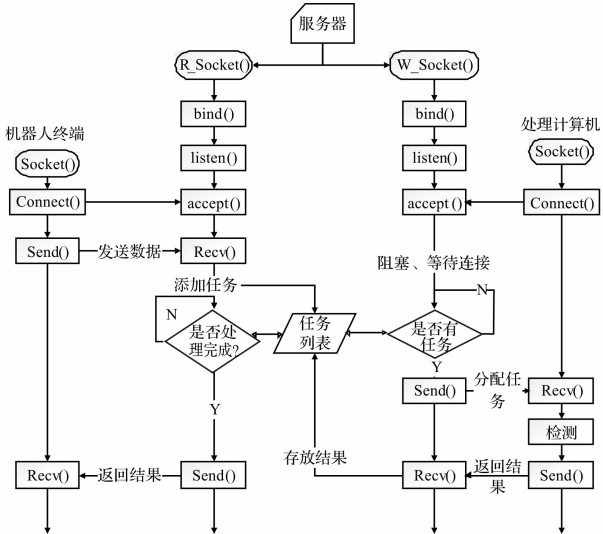


图 4 在线检测流程图

4 实验结果分析

本研究将基于计算机单机处理结果和机器人终端通过基于 TCP 的 Socket 多线程通信的在线检测结果

进行对比。其中计算机单机采用操作系统为 Ubuntu14.04,CPU 为 Intel i5-3450m,主频为 3.10 GHz,显卡为 GTX Titan X,使用 GPU 加速。机器人终端所用的网络为 4G 移动宽带连接,3 台机器人终端同时连接服务器请求在线处理,服务器端使用一台处理计算机,该处理计算机配置与计算机单机相同。

结果对比如表 1 所示。

表 1 不同平台上单张相同图片处理时间对比

平台	处理时间/ms
计算机单机	92.4
计算机终端+后台服务器(在线检测)	127.8

由实验结果可以看出,机器人终端+后台服务器进行在线检测比计算机单机的处理速度慢,但在实际应用中,机器人终端(通常为 ARM 处理器)计算机配置比计算机单机配置低很多,并且不能使用 GPU 进行加速处理,处理单张图片的时间远远超过计算机单机所用时间。

5 结束语

为了解决成本和运算速度问题,本研究提出对于每台终端采用图像数据的预处理并使用 TCP 的 Socket 多线程通信方式将图像送入云端,在云端采用 SSD 网络模型的多目标检测算法,解决运算速度的问题,计算完成后传回终端。

通过这种方法在很大程度上降低了对智能机器人终端硬件的要求,可以使得智能机器人终端的系统平台多样化,大大降低了机器人终端的硬件配置要求,同时回收的数据可以再利用,可以实现算法动态升级。

参考文献(References):

[1] 张菁,沈兰荪,DAVID DAGAN FENG. 基于视觉感知的图像检索的研究[J]. 电子学报,2008,36(3):494-499.

[2] 蔡自兴,刘健勤. 面向 21 世纪的智能机器人技术[J]. 机器人技术与应用,1998(6):2-3.

[3] 赵建伟,李国晗,唐兵,等. 智能助老服务机器人开闭环控制[J]. 兵工自动化,2015,34(5):81-83.

[4] 郑南宁. 计算机视觉与模式识别[M]. 北京:国防工业出版社,1998.

[5] 蔡强,刘亚奇,曹健,等. 图像目标类别检测综述[J]. 计算机科学与探索,2015,9(3):257-265.

[6] 罗亚非. 基于 TCP 的 Socket 多线程通信[J]. 电脑知识与技术,2009(3):563-565.

[7] VINCENT P, LAROCHELLE H, BENGIO Y, et al. Extracting and composing robust features with denoising autoencoders[C]// Proceedings of the 25th International Conference on Machine learning, New York: ACM,2008.

[8] LECUN Y, KAVUKCUOGLU K, FARABET C. Convolutional networks and applications in vision[J]. **International Symposium On Circuits & Systems**, 2010, 14(5):253-256.

[9] CIRESAN D C, MEIER U, MASCI J, et al. High-performance neural networks for visual object classification[R]. Switzerland: Dalle Molle Institute for Artificial Intelligence, 2011.

[10] LECUN Y, BENGIO Y. Convolutional networks for images, speech, and time series[J]. **The handbook of brain theory and neural networks**, 1995, 3361(10):1995.

[11] 胡健,汪庆宝. 多层前向神经网络在手写体数字识别应用中的研究[J]. 北京工业大学学报,1996(4):127-133.

[12] 乔林,张雄伟,史海宁,等. 第 4 讲深度学习应用中的常见模型[J]. 军事通信技术,2016(1):98-104.

[13] 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 and Machine Intelligence**, 2017, 39(6):1137-1149.

[14] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas:IEEE,2016.

[15] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multiBox detector[J]. **Computer Vision-ECCV 2016**, 2016(9905):21-37.

[16] 徐志节,杨杰,王猛. 一种新的彩色图像降维方法[J]. 上海交通大学学报,2004,38(12):2063-2067.

[17] 徐爱华,全书海. Socket 网络通信及其在电梯监控系统中的应用[J]. 武汉理工大学学报:信息与管理工程版,2006,28(11):56-59.

[18] 刘松,钟子发. 基于对等模式的 Socket 网络通信[J]. 微处理机,2008,29(5):46-47.

[编辑:李辉]

本文引用格式: