基于全卷积网络的目标检测算法

施泽浩,赵启军
(四川大学 计算机学院 视觉合成图形图像技术国防重点实验室,四川 成都 610065)

摘 要: 目标检测是计算机视觉的一项重要任务,其主要内容是定位图像中出现的目标,并对其进行分类。主流算法普遍基于卷积加全连接的结构,存在模型参数巨大、检测效率低下等问题。而在现实应用中,比如自动驾驶车载系统、智能监控系统中对行人、车辆等目标的检测,往往对目标检测算法的实时性具有较高要求。为此,提出一种基于全卷积神经网络的目标检测算法。网络结构完全采用卷积层实现,不仅用卷积进行特征提取,而且用卷积层进行预测,采用多任务学习,大大提高了检测效率并降低了模型复杂度。相比主流深度学习目标检测算法,如 YOLO、Faster RCNN,该算法速度更快,模型参数更少,且保持相当的精度,在 PASCAL VOC2007 权威目标检测库上的平均准确率(mAP)达到 64.5。

关键词: 目标检测; 深度学习; 全卷积神经网络; 回归; 计算机视觉
中图分类号: TP301.6 **文献标识码:** A **文章编号:** 1673-629X(2018)05-0055-04
doi: 10.3969/j.issn.1673-629X.2018.05.013

Object Detection Algorithm Based on Fully Convolutional Neural Network

SHI Ze-hao,ZHAO Qi-jun
(National Key Laboratory of Fundamental Science on Synthetic Vision,School of Computer Science,Sichuan University,Chengdu 610065,China)

Abstract: Object detection is a primary mission in computer vision that focuses on localization and classification of objects in an image. Almost all the mainstream algorithms use convolution structure followed by fully connected layer,which lead to huge number of model parameters and poor efficiency.In real application like automatic driving,intelligence CCTV and so on,high inference efficiency is required.For this,we propose a fully convolutional based object detection algorithm.The network structure is implemented by convolution layer,both feature extraction and prediction through convolution.Multi-task learning is utilized to improve detection efficiency greatly and reduce the complexity of model.Compared to YOLO and Faster RCNN,the proposed algorithm is faster,less in model parameters and maintain its precision.The average accuracy on PASCAL VOC2007 dataset reaches to 64.5.

Key words: object detection; deep learning; convolutional neural network; regression; computer vision

0 引 言

在迈向更为复杂的图像理解中,需要的不仅是图像里有什么物体,更需要知道物体的具体位置,因此目标检测就显得尤为重要[1]。相比于特定目标的检测,如人脸检测、行人检测、车辆检测,通用目标检测需要检测的物体类别众多,类别之间距离大,难度大大增加,以至于传统的滑动窗口加分类器的一般检测流程难以驾驭。近年来,深度学习[2]不断在图像识别上取得突破,受到国内外研究人员的高度关注。自 2013 年以来深度学习开始应用到目标检测领域。相比于图像识别任务,目标检测任务更为复杂。首先,一幅图像中通常不只一个目标出现。其次,目标检测需要精确的包围框(bounding box)定位目标并对其进行分类。现有基于卷积神经网络[3]的目标检测算法普遍网络结构臃肿,要实际应用还需克服速度慢、模型参数巨大等不足。为此,提出一种基于全卷积结构的目标检测算法。

1 相关工作

早期的 HOG+SVM 或者 DPM 等传统算法[4-8],都是采用手工设计特征、滑动窗口加简单分类器的设计,

收稿日期: 2017-05-09 **修回日期:** 2017-09-12 **网络出版时间:** 2017-12-05
基金项目: 国家自然科学基金(61202161)
作者简介: 施泽浩(1991-),男,硕士,CCF 会员(51825G),研究方向为目标检测; 赵启军,博士,副教授,研究方向为模式识别、智能视频监控、生物特征识别。
网络出版地址: http://kns.cnki.net/kcms/detail/61.1450.TP.20171205.1434.116.html

在行人检测、人脸检测等单目标检测中效果较好，但是对多目标检测则比较局限。卷积神经网络由于其强大的表达能力，近年来在目标检测领域表现大大超越了传统算法。基于深度学习的目标检测算法可以分为两大类，一类是基于可能区域，另一类是基于直接回归。

2013 年 Girshick R 等提出基于可能区域的 RCNN 算法[9]，先采用其他算法提取可能的目标区域，再用卷积网络对每个区域进行特征提取与边框回归，由于可能区域数目较大，每个区域都需要进行一次前向传播，算法效率十分低下。2015 年 Girshick R 等针对 RCNN 的这一缺点，提出了 Fast RCNN[10]，使用感兴趣区域池化层(ROI pooling)，在特征图(feature map)上对每个区域进行特征选择，得到统一长度的特征后合并，统一送入后续网络。2015 年 Ren S 等提出了 Faster RCNN[11]，进一步将可能区域的提取集成到网络中，设计了预定义框(anchor)机制，使得检测任务变成端到端，不需要额外的可能区域的提取过程。

2016 年 Redmon J 等提出了基于回归的 YOLO 算法[12]，针对基于可能区域的方法速度慢这一问题，采用回归的方法，直接回归出目标的 BBox 与类别，牺牲了一部分精度，但是速度更快。由于使用了全连接层输出预测结果，YOLO 模型的参数巨大。

2015 年 Long J 等提出基于全卷积网络的图像分割算法[13]，证明了全卷积网络在图像分割中的有效性。

基于可能区域的方法精度较高，但是网络复杂臃肿，速度慢。基于直接回归的方法虽然牺牲了部分精度，但是速度快，更能满足实际应用的实时性需求。文中提出的算法是采用直接回归的方法，不同于 YOLO 的是采用了全卷积结构，减小了模型的参数量和过拟合的风险，同时借鉴了 Faster RCNN 的 anchor 机制，设计了一个多任务的损失函数，减小直接回归的难度。

2 算法原理

2.1 算法概述

算法采用基于回归的方法，直接以图像为输入，通过优化给定的目标函数，网络可以预测输出图像中目标的类别与 BBox，是一种端到端的结构。这种结构的主要优点是速度快。

不同于 YOLO 的是用卷积层代替全连接层做预测输出。如图 1(a)，YOLO 采用全连接层(FCDet 层)进行回归预测，通过对下层信息的融合直接输出目标的 BBox 和类别信息。而文中算法采用 anchor 机制，用卷积层(ConvDet 层)输出目标信息，如图 1(b)。

2.2 全卷积网络

卷积神经网络是人工神经网络的一种，在图像领域具有广泛的应用。一般的卷积神经网络包括卷积层、池化层、全连接层。而全卷积网络，只包含卷积层和池化层。全卷积网络的优点包括:

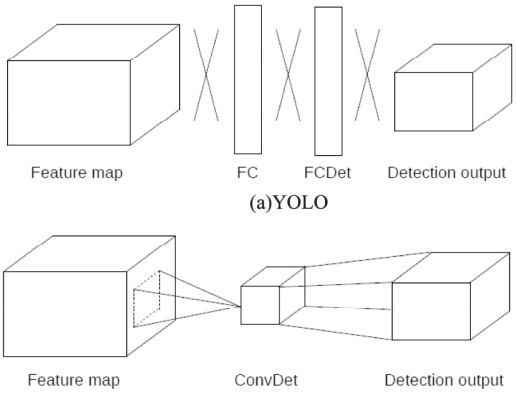

图 1 模型对比

(1) 全连接层对特征图上每个像素点同等对待，提取了全图信息，包括背景信息。而全卷积的特征图上每个像素只提取了其对应的图像感受野内的信息，减少了无关背景的干扰。

(2) 全连接层参数多，而卷积层参数少，不容易过拟合。

(3) 全卷积网络能适应不同的输入大小。

2.3 预定义框机制

预定义框机制是在卷积输出的特征图上的每个像素位置上设置一组不同大小、不同长宽比的预定义框，如图 2 所示。通过选择与目标 IOU 最大的一个预定义框进行目标函数优化。ConvDet 层将在每个像素位置同时输出各个预定义框的位置和尺寸的调整量 $(\delta_x, \delta_y, \delta_w, \delta_h)$，是否含有目标的分数 Score，还有属于各个类别的概率 Class。最后通过非极大值抑制(non-maximum suppression)得到最终的目标检测框。文中

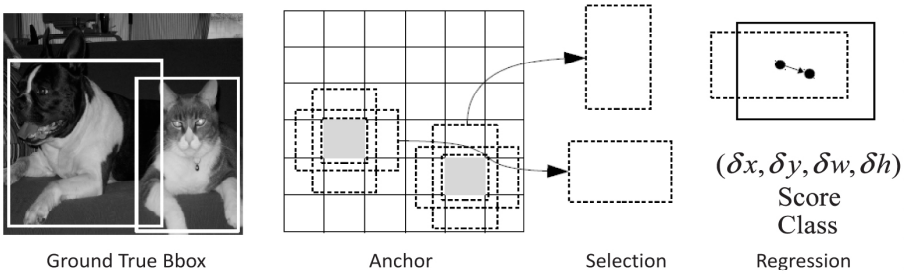

图 2 预定义框机制

算法通过对训练集的标签框做 K 均值聚类(k-means clustering) 得到 9 个预定义框。

2.4　损失函数

算法的损失函数是一个多任务损失函数,包括边框回归、anchor 得分回归,还有分类的交叉熵损失。

$$\begin{cases} \text{Loss} = L_{\text{bbox}} + L_{\text{score}} + L_{\text{class}} \\ L_{\text{bbox}} = \dfrac{\lambda_{\text{bbox}}}{N_{\text{obj}}} \sum_{i=1}^{K} I_k \left[\delta X_k - \delta X_k^G \right)^2 \right] \\ L_{\text{score}} = \sum_{i=1}^{K} \left[\dfrac{\lambda_{\text{score}}^{+}}{N_{\text{obj}}} I_k \left(\Upsilon_k - \Upsilon_k^G \right)^2 + \dfrac{\lambda_{\text{score}}^{-}}{K - N_{\text{obj}}} \bar{I}_k \Upsilon_k^2 \right] \\ L_{\text{class}} = \dfrac{1}{N_{\text{obj}}} \sum_{i=1}^{K} I_k l_c^G \log\left(p_c \right) \end{cases} \tag{1}$$

L_{bbox} 部分处理边框回归。其中,λ_{bbox} 是该部分损失的权重系数; N_{obj} 是出现的目标个数; K 是 anchor 的个数; $I_k \in \{0,1\}$ 指示了与目标 IOU 最大的 anchor,只取 IOU 最大的 anchor 参与 loss 计算; X 是一个四维向量 $(\delta_x, \delta_y, \delta_w, \delta_h)$,是 anchor 的修正量。

L_{score} 回归每个 anchor 是否含有目标的可能性。Υ_k^G 等于 anchor 与目标 BBox 的 IOU。若一个 anchor 与目标 BBox 的 IOU 越高,表示该 anchor 含有目标的可能性越高。

L_{class} 是交叉熵损失函数。其中,$l_c^G \in \{0,1\}$,$l_c = 1$ 表示该 anchor 属于第 c 类; $p_c \in [0,1]$ 是通过 Softmax 后的类别分布。

3　实　验

3.1　实验数据库

实验采用 VOC2007 和 VOC2012 目标检测数据库[14],含有二十类物体。VOC 数据库分为训练集、验证集和测试集,只允许用训练集和验证集做训练,不允许使用外部数据。训练集总共有 16 541 张图片,并在 VOC2007 的测试集上进行测试。

3.2　评价标准

IOU(intersection over union,交并比) ,AP(average precision,平均准确率) ,mAP(mean average precision,平均准确率均值) 是评价目标检测算法的三个主要参数。IOU 表示目标的检测框与目标的标签框交面积与并面积的比率。AP 和 mAP 的计算公式如下:

$$\text{AP} = \frac{1}{N} \sum_{i=1}^{n} \text{IOU}\left(b_i, b_i^{\text{gt}} \right) \tag{2}$$

$$\text{mAP} = \frac{1}{C} \sum_{c=1}^{20} \text{AP}_c \tag{3}$$

其中,b 为检测框; b^{gt} 为标签框; N 为测试集的标签框总数; C 为类别数 20。

3.3　结果分析

在 VOC2007 测试集上的测试结果见表 1。可见,文中算法的速度约是 Faster RCNN 的 7.5 倍,平均准确率和速度均超过了基于回归框架的 YOLO。检测结果见图 3。

表 1　实验结果

算法	训练集	mAP	FPS
Faster-RCNN	0712train+val	73.2	4.6
YOLO	0712train+val	63.4	30
文中算法	0712train+val	64.5	35

3.4　模型参数分析

Faster RCNN 与 YOLO 的网络结构均采用了全连接层,其中全连接层分别占了约 80% 与 72% 的参数。而文中算法采用全卷积结构,模型约为 Faster RCNN 的四分之一,YOLO 的十分之一,仅为 103 MB。表 2 对比了各个算法的模型大小。

表 2　模型大小

算法	基础网络	模型大小/MB
Faster-RCNN	VGG16[15]	485
YOLO	24Conv+2Fc	1 126
文中算法	Res50[16]	103

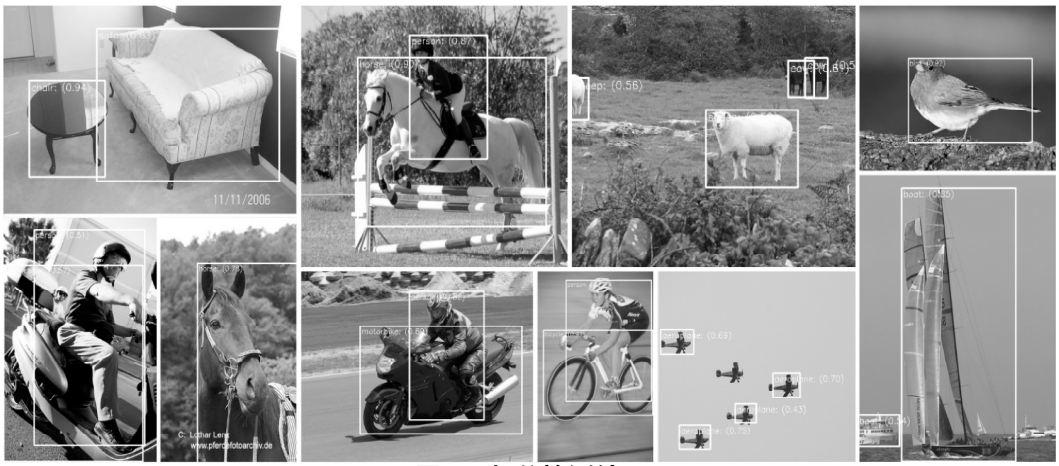

图 3　部分检测结果

4 结束语

为解决现有目标检测算法模型参数大、速度慢等缺点,提出一种基于全卷积网络的目标检测算法。该算法利用预定义框机制,用卷积层代替全连接层进行结果预测,大大降低了模型参数数目,提高了检测效率。下一步的工作可以设计更佳精简的基础网络,进一步提高模型的预测速度。

参考文献:

[1] 尹宏鹏,陈 波,柴 毅,等.基于视觉的目标检测与跟踪综述[J].自动化学报,2016,42(10):1466-1489.

[2] 孙志军,薛 磊,许阳明,等.深度学习研究综述[J].计算机应用研究,2012,29(8):2806-2810.

[3] 李彦冬,郝宗波,雷 航.卷积神经网络研究综述[J].计算机应用,2016,36(9):2508-2515.

[4] 赵丽红,刘纪红,徐心和.人脸检测方法综述[J].计算机应用研究,2004,21(9):1-4.

[5] 贾慧星,章毓晋.车辆辅助驾驶系统中基于计算机视觉的行人检测研究综述[J].自动化学报,2007,33(1):84-90.

[6] 李文波,王立研.一种基于Adaboost算法的车辆检测方法[J].长春理工大学学报:自然科学版,2009,32(2):292-295.

[7] FELZENSZWALB P, GIRSHICK R, MCALLESTER D,et al.Visual object detection with deformable part models[C]//Computer vision & pattern recognition. Washington, DC, USA: IEEE Computer Society,2010: 2241-2248.

[8] 曾接贤,程 潇.结合单双行人DPM模型的交通场景行人检测[J].电子学报,2016,44(11):2668-2675.

[9] GIRSHICK R,DONAHUE J,DARRELL T,et al.Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Computer vision and pattern recognition. Washington,DC,USA: IEEE Computer Society,2014: 580-587.

[10] GIRSHICK R.Fast R-CNN[C]//International conference on computer vision.Washington,DC,USA: IEEE Computer Society,2015: 1440-1448.

[11] REN S,HE K,GIRSHICK R,et al.Faster R-CNN: towards real-time object detection with region proposal networks[C]//Proceedings of the 28th international conference on neural information processing systems. Cambridge, MA, USA: MIT Press,2015: 91-99.

[12] REDMON J,DIVVALA S,GIRSHICK R,et al.You only look once: unified, real-time object detection[C]//IEEE conference on computer vision and pattern recognition.Washington,DC,USA: IEEE Computer Society,2016: 779-788.

[13] LONG J,SHELHAMER E,DARRELL T.Fully convolutional networks for semantic segmentation[J].IEEE Transactions on Pattern Analysis & Machine Intelligence,2017,39(4):640-651.

[14] EVERINGHAM M,VAN GOOL L,WILLIAMS C K I,et al.The PASCAL visual object classes challenge[J].International Journal of Computer Vision,2010,88(2):303-338.

[15] SIMONYAN K,ZISSERMAN A.Very deep convolutional networks for large-scale image recognition[EB/OL].(2014-04-10)[2017-06-13].https://arxiv.org/abs/1409.1556.

[16] HE K,ZHANG X,REN S,et al.Deep residual learning for image recognition[C]//Computer vision and pattern recognition.Washington,DC,USA: IEEE Computer Society,2016: 770-778.

(上接第54页)

ar.pdf.

[2] 张 引,陈 敏,廖小飞.大数据应用的现状与展望[J].计算机研究与发展,2013,50:216-233.

[3] 李建中,刘显敏.大数据的一个重要方面:数据可用性[J].计算机研究与发展,2013,50(6):1147-1162.

[4] 李建中,王宏志,高 宏.大数据可用性的研究进展[J].软件学报,2016,27(7):1605-1625.

[5] MILLER D W,YEAST J D,EVANS R L.Missing prenatal records at a birth center: a communication problem quantified[C]//Proceedings of AMIA annual fall symposium.Maryland: American Medical Informatics Association,2005: 535-539.

[6] SWARTZ N.Gartner warns firms of 'dirty data'[J].Information Management Journal,2007,41(3):6-12.

[7] KORN F,MUTHUKRISHNAN S,ZHU Y.Checks and balances: monitoring data quality problems in network traffic databases[C]//Proceedings of the 29th international conference on very large data bases.[s.l]:[s.n.],2003: 536-547.

[8] XIONG Hui,PANDEY G,STEINBACH M,et al.Enhancing data analysis with noise removal[J].IEEE Transactions on Knowledge & Data Engineering,2006,18(3):304-319.

[9] 李聪颖,王瑞刚,于金良.大数据分布式全文检索系统的设计与实现[J].计算机与数字工程,2016,44(12):2426-2430.

[10] 李卫榜,李战怀,陈 群,等.分布式大数据不一致性检测?[J].软件学报,2016,27(8):2068-2085.

[11] 维克托·迈尔-舍恩伯格,肯尼斯·库克耶.大数据时代[M].杭州:浙江人民出版社,2013.

[12] 曹黎侠,冯孝周.新的改进AHP算法研究及应用[J].计算机技术与发展,2010,20(12):115-117.

[13] 王 磊,黄梦醒.云计算环境下基于灰色AHP的供应商信任评估研究[J].计算机应用研究,2013,30(3):742-744.

[14] 赵焕臣,许树柏,和金生.层次分析法[M].北京:科学出版社,1986: 22-26.

[15] 魏翠萍.层次分析法中和积法的最优化理论基础及性质[J].系统工程理论与实践,1999,19(9):113-115.