

基于机器学习的目标检测算法综述

李名波

长春理工大学计算机科学技术学院

摘要: 目标检测在计算机视觉和机器学习领域是比较热门的技术。经典的目标识别方向有提取目标的HOG特征, 训练SVM分类器进行识别等初级特征提取方法, 这类方法的准确率早已被机器学习超过。随着机器学习在目标检测技术上的应用, 各种各样的优秀的目标检测算法出现, 但是缺少对这些算法进行一个系统的总结和比较。本文比较了基于区域提议类的目标检测算法R-CNN, Fast-RCNN, Faster-RCNN; 端到端的检测算法如YOLO系列, SSD算法; 以及最近提出的CascadedRCNN算法, RefineDet算法, SNIP算法, R-FCN-3000算法, DES算法, STDN算法。分析了各个目标识别算法在各大公开数据集中的识别准确率和速度, 以及算法各自的优点和缺陷。并对目标检测算法进行了一个总结。

关键词: 机器学习 目标检测 区域提议

由于机器学习技术的飞速发展, 机器学习在各个领域的应用也越来越多。尤其是在图像处理方向, 越来越多的人开始使用机器学习进行目标检测。层出不穷的算法准确率和速度在不断的提高, 需要对各个算法进行一个系统的整理和比较。目标检测技术的目的是在所给定的图片中检测到感兴趣的目标, 锁定目标所在的矩形区域和种类。目标检测技术具有挑战性的问题是目标的不同形态、姿势、光照的影响、遮挡的干扰等。这些问题使得目标检测任务变得更加艰难。传统的检测方法比较倾向于使用滑动窗口的方式, 其主要包括: 使用尺寸不同的窗口选中图中的部分区域; 分析选中区域的特征, 如对行人的姿态效果较好的HOG特征, 对人脸特征识别效果较好的Haar特征; 使用如SVM分类器, Adaboost分类器对目标进行分类。其中多尺度形变部件模型DPM算法是这类方法中表现比较出色的。DPM将物体看成多个部件, 用部件之间的关系对整体的物体进行描述, 这个性质很好的符合大多数物体的非刚性性质, 取得了较好的成绩。随即Ross Girshick提出的R-CNN算法将深度学习的目标检测算法准确率提升了一个档次。MSRA何恺明等人提出SPP-net算法去掉了R-CNN中的crop/warp操作, 使用空间金字塔池化层 (SPP) 解决了CNN全连接层要求输入图片大小必须一致的问题。紧接着速度更快的Fast-RCNN算法和Faster-RCNN算法出现了。Fast-RCNN算法解决了RCNN算法在区域提议阶段需要进行大量重复的待检测框运算, 很大程度上提高了算法的识别速度。而Faster-RCNN算法是直接区域提议放在了网络模型中, 称为区域提名网络 (Region Proposal Networks) 来进行待检测框的计算, Faster-RCNN舍弃了原本独立的区域提议步骤, 使得能够和分类网络, 回归网络公用卷积特征, 检测速度得到进一步的提升。之后的R-FCN算法使用位置敏感的卷积网络来代替了最后的全连接层, 进一步让计算共享, 提升效率。

基于端到端的方法有以速度闻名的YOLO系列算法, 最初的YOLO算法将目标的识别行为和定位行为合二为一, 简化了检测流程, 实现了端到端的检测, 速度能够达到45FPS。SSD算法针对YOLO算法检测小目标

不佳的弱点进行改进, 使用了anchor机制和多尺度, 提升了算法准确率。之后又出现了改进的YOLOV2算法, 对YOLO算法的多目标框架进行了调整, 提升了准确率。YOLOV3算法^[1]提升了识别准确率, 尤其是针对小目标。

Cascade RCNN算法区别于普通级联, Cascade R-CNN的检测网络是通过在不同的IOU阈值确定的正负样本上训练得到的, 算法的准确率比较优秀。RefineDet算法^[2]结合了SSD算法和RPN算法的有点, 在保证算法速度的前提下提高了检测准确率。R-FCN-3000算法引入了位置敏感得分图来解决原来R-FCN算法重复的计算, 提升了算法速度。DES算法时基于SSD算法改进的, 解决了原算法中对小目标对象不敏感的问题, 提升了算法的准确率。STDN算法^[3]使用DenseNet网络作为特征提取网络, 提高了目标检测算法对不同尺度目标的检测效果。

一、准确率比较

各个目标检测算法在各大公开数据集上的准确率比较如表1所示, Pascal VOC challenge数据集和COCO数据集中包含大量包含目标的复杂场景, 多种类别的目标, 是比较权威的衡量目标检测算法性能的公开数据集。

表1 目标检测算法准确率比较

算法名称	VOC2007mAP (%)	VOC2012mAP (%)	COCO mAP (%)
Fast RCNN	70.0	68.4	—
Faster-RCNN	73.2	70.4	34.9
R-FCN	79.5	77.6	29.9
YOLOV1	66.4	57.9	—
SSD	76.8	74.9	30.3
YOLOV2	78.6	73.4	—
YOLOV3	79.4	—	33.0
Cascade RCNN	—	—	42.8
RefineDet	83.8	83.5	41.8
R-FCN-3000	83.6	—	34.9
DES	81.7	80.3	32.8
STDN	80.9	—	31.8

算法整体准确率表现最佳的是Shifeng Zhang等人的RefineDet算法, 在VOC2007和VOC2012中均表现最佳, 在COCO数据集中表现仅次于Cascade RCNN算法。RefineDet算法一方面引入了目标检测算法中由粗到细的回归思想, 即先通过RPN网络粗粒度的矩形框信息, 在通过一般的回归支路进行回归得到更加精确的矩形框信息。另一个方面引入特征融合的手段进行网络检测, 很有效的提高了对小目标的检测效果。

在COCO数据集中表现最佳的Cascade RCNN算法是由多个检测模型组成, 基于各不相同的IOU阈值的正负样本训练得到这些模型, 之前模型的输出作为之后模型的输入, 检测模型越往后界定正负样本的IOU阈

值不断上升。使用Cascade RCNN这种级联结构能够让每一步的检测器都专注于检测IOU在一定范围的提议区域, 由于输出IOU普遍都会比输入IOU大, 所以检测效果会越来越好。

R-FCN-3000算法是将R-FCN算法进行了改进, 适应更多类别的检测。R-FCN算法的局限在于一旦类别数目较多时, 生成分类支路的位置敏感得分图时计算量比较大, 因此R-FCN-3000通过解耦分类支路的生成过程, 将原本的方法变成了两个支路。以此提升了准确率和速度。识别准确率在VOC2007上的表现比较理想, 在COCO数据集上表现也不错。

在VOC2007和VOC2012上表现不错的DES算法是基于SSD算法进行的改进, 解决了SSD算法针对小目标时检测效果不理想的问题。DES增加了识别小目标时的高级语义特征, 引入分割模块提高了低级别时特征图的语义信息。引入全局激活模块提高了高级别时的特征图语义信息。同样表现不错的STDN算法主网络使用的DenseNet网络, 并且引入了规模转移层, 在几乎不怎么增加参数数量和计算量的前提下生成较大尺度的特征图, 同样提升了识别准确率。

YOLOV3算法在保证较高检测速度的同时仍然能够在公开数据集中达到几乎80%的准确率。YOLOV3算法时YOLO系列最新的算法, 采用端到端的方式进行训练, 一个损失函数就解决了训练问题, 只需要关注输入端到输出端即可。同时YOLOV3进行了多尺度的训练, 针对小目标的检测准确率更高。

其他算法如Fast-RCNN算法, Faster算法, SSD算法在准确率上已经被最近的算法超过, 并且在速度上也不具备优势。

二、速度比较

算法的速度也是衡量算法实用性的重要指标, 检测速度太慢在很多应用场景没有实用价值。目标检测算法的速度比较如表2所示:

表2 目标检测算法速度比较

算法名称	速度 (FPS)
Fast RCNN	3
Faster-RCNN(low)	5
Faster-RCNN(high)	17
R-FCN	6
YOLOV1	45
SSD	19
YOLOV2	40
YOLOV3-320	45.4
YOLOV3-416	34
Cascade RCNN	8.6
RefineDet	24.1
R-FCN-3000	30
DES	31.7
STDN321	40.1
STDN513	28.6

表2中在保证较高准确率的前提下，速度最快的是以速度闻名的YOLO系列算法，YOLOV3在输入为320×320分辨率下能够达到45.4FPS的速度，这在目标检测算法中算是非常高的速度了。在这样的速度下仍然能够有较高的识别准确率，实用性非常高。YOLOV2和YOLOV1算法的速度都比较令人满意，但是在准确率上已经被YOLOV3超越了很多。

其次是输入为321×321大小的STDN算法能够达到40.1的FPS，即是调整输入为513×513提升了较高准确率，仍然能够保证实时性能达到28.6FPS。在速度和YOLOV3算法不分上下的情况下准确率比YOLOV3算法更加优秀。

准确率较高的DES算法检测速度也达到了31.7FPS的速度。准确率更高的R-FCN-3000也达到了30FPS，其次准确率最高的RefineDet算法能够达到不错的24.1FPS的检测速度。这三个准确率最高的算法都达到了实时性能，实用性非常高。

相比之下勉强能够实时的SSD算法和Faster-RCNN算法在准确率上也没有任何优势，和Fast-RCNN，YOLOV1，YOLOV2等算法一样，被强大的新算法逐渐淘汰在后面。

三、结论

本文对基于机器学习的目标检测算法进行了一系列的比较。在算法准确率上，如今的目标检测算法准确率越来越高，经典的Faster-RCNN算法，YOLOV2算法，SSD等算法准确率已经被超越。在算法速度上，机器学习目标检测整体算法速度已经越来越快，实时性几乎是目标检测算法必须满足的调价，并且都保证了较高准确率的前提。同时又很多算法能够通过调整输入分辨率和网络结构，在准确率和速度之间调节，能够根据具体的使用需求在准确率和速度之间做出更好的平衡。

当更注重算法实时性能时, STDN算法是最值得推荐的算法, 其次的DES算法, R-FCN-3000算法, YOLOV3算法, RefineDet也非常值得推荐, 在这些实时算法中, RefineDet算法的准确率最高。如果需要的目标检测算法更加追求准确率, 那RefineDet算法, Cascade RCNN算法是准确率令人比较满意的算法。

我的笔记⁰

参考文献³

引证文献¹