文章编号:1007-1423(2021)20-0074-06 DOI:10.3969/j.issn.1007-1423.2021.20.015

基于改进的 Faster R-CNN 的车辆目标检测

单志勇,宫加辉

(东华大学,上海 201600)

摘要:

车辆目标检测作为交通管理系统的重要组成部分具有重要的研究意义。为了解决传统车辆目标检测带来的准确率低的问题,提出了基于改进的 Faster R-CNN 算法的车辆目标检测。改进后的 Faster R-CNN 算法在原始 Faster R-CNN 算法的基础上随机选取 960×540 、900×500 、800×480 三种尺寸的训练图片进行训练,同时对 RPN(Region Proposal Network)中的候选区域比例进行了扩展,增加了 1:3、3:1 两种比例。改进后的目标检测的 mAP 达到了 95.56%,比基于 Faster R-CNN 算法的车辆目标检测的 mAP 提高了 0.06%。

关键词:

Faster R-CNN 算法; 目标检测; 深度学习

0 引言

随着城市化进程的不断加快,交通管理系统是当前交通领域需要研究的热点问题。车辆目标检测是交通管理系统的重要组成部分之一,广泛应用在智能监控系统和智能停车系统等领域中。车辆目标检测应用在智能交通监控系统中可以加快救援速度和分析马路拥堵情况,极大的缓解了交通压力。车辆目标检测应用在智能停车系统中,通过对停车场内的停车情况进行分析,可以缓解停车难的问题。可见,车辆目标检测在交通管理系统中的应用极大地提高了交通管理的效率。所以,优化车辆目标检测问题对增强交通管理系统具有重要意义和应用价值。

传统的目标检测算法流程图如图 1 所示。首先输入需要检测的图片,采用滑动窗口的方式进行候选框的选取,其次,采用基于人工设计特征提取[1]的算法,如 HOG(Histogram of Oriented Gradient)[2]对候选框进行特征提取;然后使用分类器,如 SVM[3]等进行判定,在经过分类器判定后获得的候选框会有一部分重叠,因此,需要在经过 NMS[4]将重叠的部分筛选合并找到目标物体的最佳位置,得到最终的检测结果。

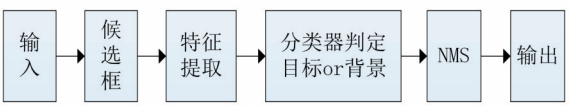


图1 传统目标检测算法流程图

传统目标检测的不足主要体现在候选框和特征提取部分。选取候选框采用的是滑动窗口的方式,若滑动窗口的大小和步长的变化不合理,则会产生大量冗余的候选框,降低了目标检测的速度;特征提取的方法是基于人工设计的,特征提取过程中极易受到个人主观性的影响,降低了目标检测的准确率。

2014 年,Ross Girshick 等人提出了 R-CNN[5]算法,这一算法明显提高了目标检测的速度。R-CNN 算法在传统算法模型的基础上进行两个方面的调整:一是采用选择性搜索(Selective Search)[6]代替滑动窗口的方式进行候选框的选取,大约选取 1000~2000 个候选框,减少了计算量;其二是采用 CNN 代替人工设计特征提取特征图,有效避免了人为主观性的影响。

2015 年 Ross Girshick 提出了 Fast R-CNN[7],弥补了 R-CNN 算法中输入图像需固定尺寸的不足。Fast R-CNN 算法的输入图像的尺寸是任意的,并且是直接

对整张图像卷积，减少了重复计算；采用 ROI Pooling 层固定特征的尺寸，以便让特征图以合适的尺寸输入到全连接层；采用 Softmax 层输出目标的类别。Fast R-CNN 算法只需单个模型即可完成目标检测，极大地提高了检测效率。

2016 年 Ross Girshick 在 R-CNN 和 Fast R-CNN 的基础上提出了 Faster R-CNN[8]。Faster R-CNN 算法采用 RPN（Region Proposal Network）代替选择性搜索法提取候选框，并把提取候选框，选取特征图和判定目标类别和位置放置在同一个网络，解决了模型结合慢的问题，检测效率和准确率都满足了目标检测的要求，极大地提高了目标检测的综合性能。

本文主要内容就是研究改进的 Faster R-CNN 算法在实际车辆目标检测的场景中应用，主干网络选取了 VGG16[9]预训练模型，训练和测试的样本集来自 UA-DETRAC 数据集，使用训练集训练目标检测模型，使用测试的样本集对得到的模型进行测试。实验结果表明使用改进的 Faster R-CNN 算法进行车辆目标检测具有较高的泛化能力，同时也提高了车辆目标检测的准确度和效率。

1 基于改进的Faster R-CNN的车辆目标检测

Faster R-CNN 网络主要由以下四个内容组成：卷积层、RPN 层、RoI Pooling 层和坐标回归层。其网络结构图如图 2 所示。卷积层采用 VGG16 预训练模型来提取特征，其网络结构由 13 个 Conv 层、13 个 ReLU 层和 4 个 Pooling 层组成。在整个 VGG16 模型中，每经过一次 Pooling 层的输出都是输入的 1/2。生成的 feature map 被共享于 RPN 层和 RoI Pooling 层；PRN 层利用 Softmax 和 bounding box 判断 anchors 所属类别和修正 anchors 的边框，以便获得更准确的区域建议框；ROI Pooling 层的输入来自卷积层生成的 feature map 和 RPN 层的区域建议框，该层的作用就是将输入固定到统一大小并输出至坐标回归层的全连接层；坐标回归层则是用来判断目标的类别和坐标位置。

Faster R-CNN 模型进行训练时采用的训练图片的尺寸是恒定的，这在一定程度上降低了目标检测的效率。本文提出了多尺度训练，在 960×540 、900×500 、800×480 三种不同尺寸的训练图片中随机选取进行训练，能有效提高车辆目标检测的鲁棒性。

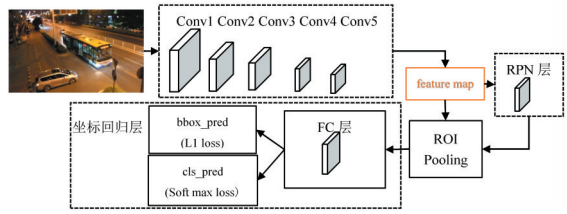
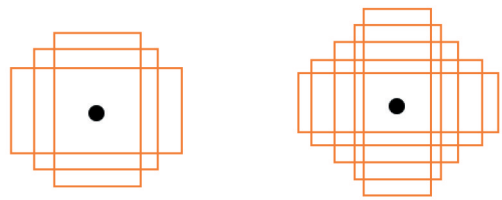


图2 Faster R-CNN 的网络结构图

1.1 RPN 网络

Anchor 是 RPN 网络中的一个重要概念，它与滑窗法设置的窗口大小类似，都是预设图像的参照框。原始 Faster R-CNN 网络中 Anchor 是在 RPN 网络的 3×3 卷积层生成的，每个锚点对应三种尺度缩放比和三种宽高比分别为[128,256,512]、[1:2, 1:1, 2:1] 的 9 种 Anchor。改进后的 RPN 增加了 1:3、3:1 两种候选区域的比例。改进前后的对比图如图 3 所示。



（a）原始比例　　（b）改进后比例

图3 改进前后对比图

RPN 是 Faster R-CNN 网络的核心部分，是一个完整的卷积神经网络。RPN 作用在特征提取网络输出的特征图像上，首先是经过一个卷积层选取锚点，生成 anchor；然后分成两路，一路输入分类层利用 Softmax 判断 anchor 的类别，在进行 Softmax 的前后都对输出做了 Reshape，以便 Softmax 进行二分类和减少计算的复杂度；一路输入边框回归层用来判断 anchors 的边框位置，计算 anchor 的位置与真实框位置的偏移量并进行调整获取位置更准确的 anchor；分类层和边框回归层的输出共同输入 Proposal，Proposal 根据 Img_info 提供的信息和 NMS 算法筛选大约 2000 个更准确的 anchor。

1.2 RoI Pooling

RoI Pooling 使得原始图片的大小可以是任意的，减少了原始图片在缩放成固定大小时带来的信息损

失；RoI Pooling 可以将不同尺寸的图片生成的特征图转换为固定尺寸的特征图，帮助全连接层以及分类层更好地吸收。

RoI 利用 Max Pooling 将宽和高为（$w\times h$）的 RoI 窗口用 $(W\times H)$ 的子窗口进行分割，得到约等于 $\frac{h}{H}\times\frac{w}{W}$ 个子窗口。利用 Max Pooling 计算每个子窗口的最大像素点，形成一个 $W\times H$ 的 Feature map。

1.3 损失函数

Faster R-CNN 中的损失分为回归损失和分类损失两大类，分类损失是 Softmax，回归损失是 Smooth L1 造成的损失。其总的损失函数表达式如下：

$$L\left(\{p_i\},\{t_i\}\right)=\frac{1}{N_{cls}}\sum_i L_{cls}\left(p_i,p_i^*\right)+\lambda\frac{1}{N_{reg}}\sum_i p_i^* L_{reg}\left(t_i,t_i^*\right) \tag{1}$$

RPN 和 ROI 的分类损失表达式相同，均是交叉熵损失。但 RPN 损失是二分类交叉熵损失，ROI 损失是多分类交叉熵损失。其表达式如公式(2)所示。

$$Loss_{cls}=\frac{1}{N_{cls}}\sum_i L_{cls}\left(p_i,p_i^*\right) \tag{2}$$

其中，N_{cls} 为总的 anchor 的数量；在 RPN 分类损失函数中，p_i 为第 i 个 anchor 的预测分类概率；p_i^* 为标签，当 anchor 为 positive 时，$p_i^*=1$，当 anchor 为 negative 时，$p_i^*=0$；$L_{cls}\left(p_i,p_i^*\right)$ 是二分类交叉熵损失函数，其表达式如公式(3)所示。

$$L_{cls}\left(p_i,p_i^*\right)=-\log\left[p_i^* p_i+\left(1-p_i^*\right)\left(1-p_i\right)\right] \tag{3}$$

RPN 和 ROI 的回归损失函数均是由 Smooth L1 Loss 计算的，Smooth L1 Loss 解决了在预测值和真实值很接近的时候发生梯度爆炸以及函数在 0 点不可导影响收敛的问题。Smooth L1 Loss 的表达式如公式(4)所示，函数图像如图 4 所示。回归损失函数的表达式如公式(5)所示。

$$Smooth_{L1}(x)=\begin{cases}0.5x^2 & ,|x|<1\\ |x|-0.5\sigma^2 & ,\text{otherwise}\end{cases} \tag{4}$$

$$Loss_{reg}=\lambda\frac{1}{N_{reg}}\sum_i p_i^* L_{reg}\left(t_i,t_i^*\right)=\lambda\frac{1}{N_{reg}}\sum_i p_i^* R\left(t_i-t_i^*\right) \tag{5}$$

其中，N_{reg} 由 anchor 位置的数量决定；$p_i^* L_{reg}\left(t_i,t_i^*\right)$ 表示只有正样本才有边框回归损失；$L_{reg}\left(t_i,t_i^*\right)=R\left(t_i-t_i^*\right)$，$R\left(t_i-t_i^*\right)$ 是 Smooth L1 函数；λ 为权重平衡参数；t_i 是第 i 个 anchor 预测的边框回归的参数化坐标；t_i^* 是第 i 个 anchor 的真实框的参数化坐标。

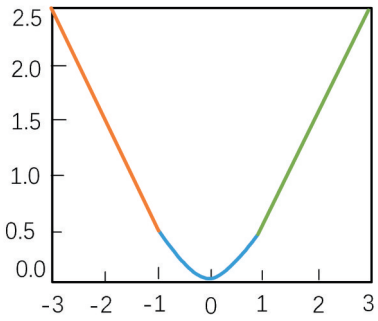


图4 Smooth L1 函数图像

2 实验结果与分析

2.1 数据集介绍

随着车辆目标检测受到了国内外学者的深入研究，如今可供选择的车辆检测数据集也很多，主要包括 KITTI、N-CARS、CompCars 和 UA-DETRAC 等。KITTI 数据集采主要是行驶在卡尔斯鲁厄的乡村公路和高速公路上的车辆，数据采集场景在国外，而本实验研究的车辆目标检测的应用场景主要是在国内的街道上，与本实验的实际场景存在偏差。N-CARS 和 CompCars 数据集主要应用在汽车分类，缺少本实验中需要的在道路拥堵、光照太强等不同条件下的稀有数据集。UA-DETRAC 数据集既是在国内 24 座城市拍摄的，也具有道路拥堵、光照强度不同的稀有数据集，与本实验所需实际场景基本相符。

因此，本实验所用数据集均截取至 UA-DETRAC 数据集。其中训练集有 10572 张，测试集有 6705 张图片。部分图像如图 5 所示。本实验所选取的 10572 张训练集和 6705 张测试集均有相应的标注文件。

图5 UA-DETRAC 数据集

2.2 实验结果

本实验中训练所用的硬件配置如下：采用 NVIDIA GeForce GTX 1080 TI GPU 加速，搭建运行环境 PyTorch 1.6.0、CUDA 10.1 以及 cuDNN 7，编程语言选择了 Python，用 Python 3.6.9 运行程序。由于所用程序数据集标注格式采用的是 VOC2007 数据集所用标注格式，因此，在训练之前本实验先将所用数据集标注文件的格式改为了 VOC2007 数据集标注文件的格式。训练过程中采用 VGG16 预训练模型，预训练模型的参数权重是 VGG16 在 ImageNet[10]下训练好的，初始学习率设置为 0.001，当迭代次数达到 9 之后，就将学习率设置为原来的 0.1，即 0.0001；本实验在 GPU 加速下训练迭代 14 次，得到车辆目标检测的模型。

本实验的检测结果均是在 Visdom 下的可视化窗口呈现的，图 6 是原始图片的标签图、图 7 是原始图片的预测图、图 8 是算法改进后的 5 种 loss、图 9 是 Faster R-CNN 算法改进前后的评估指标 mAP 的对比图。由图 7 可以看出车辆检测的概率均在 0.9 以上，但存在漏检的情况。该情况可能是受到天气条件、遮挡以及光照变化等因素的影响导致的，也可能是数据集在进行预处理时导致的原始图片的信息损失以至于无法识别出来车辆，还有可能是因为原始图片是在监控视频中截取出来的虚化严重无法清晰识别出车辆。由图 8 可知，roi_loc_loss 的损失最大，rpn_cls_loss 的损失最小，总损失函数最终控制在 0.176 左右，图 8 表明了适当的增加样本数据集有利于降低车辆目标检测的损失和提高车辆目标检测的准确率。由图 9 可知，基于改进后的 Faster R-CNN 算法的车辆目标检测的 map 最终稳定在 90.56%，而基于 Faster R-CNN 算法的车辆目标检测的 map 最终稳定在 90.50%，mAP 提高了 0.06%。

图6 标签图

图7 检测效果图

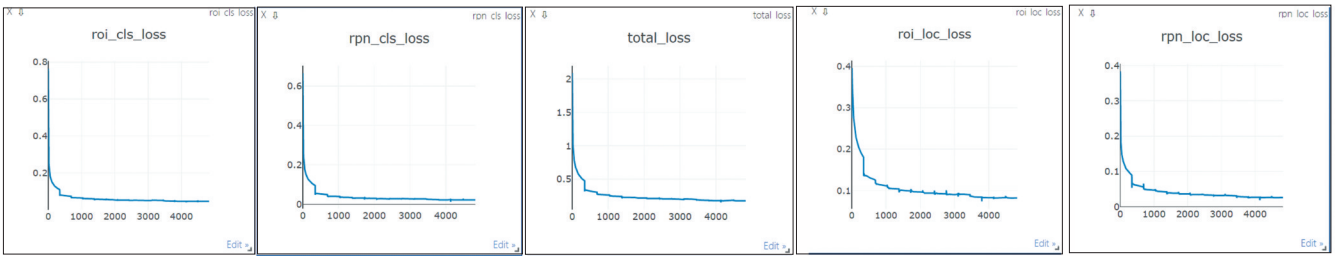
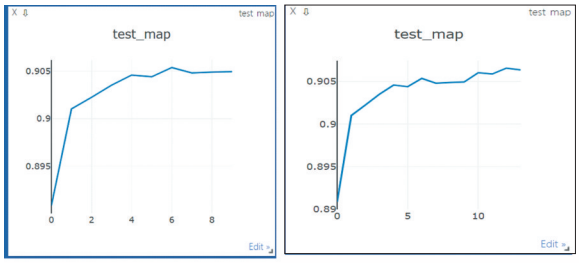


图8 5种损失函数



(a)原始 mAP　　(b)改进后 mAP

图9 mAP对比图

3 结语

本文利用了公开 UA-DETRAC 数据集，基于改进的 Faster R-CNN 算法完成了对实际场景中的车辆目标检测，有效避免了传统目标检测中出现的过于依赖人工特征提取的问题，有效地提高了车辆目标检测的准确率和速度，为车辆目标检测在智能监控和无人驾驶等领域的应用奠定了基础。虽然检测的准确率和速度已经有了很大的提高，但仍然存在错检、漏检以及无法做到实时检测的情况。因此，未来车辆目标检测的难点就在于如何尽可能减小光照、天气等外部因素的干扰，以及如何做到实时检测。

参考文献：

[1]WANG X Y, YANG M, ZHU S H, et al. Regionlets for generic object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(10): 2071-2084.

[2]TAIGMAN Y, YANG M, RANZATO M A, at al. DeepFace: closing the gap to human-level performance in face verification[J]. IEEE Conference on Computer Vision and Pattern Recognition. [S.I.]: IEEE Press, 2014: 1701-1708.

[3]KAZEMI F M, SAMADI S, POORREZA H R, at al. Vehicle recognition using curvelet transform and SVM[J]. the 4th International Conference on Information Technology[S.I.]: IEEE Press, 2007: 516-521.

[4]A. NEUBECK, L. VAN GOOL. Efficient non-maximum suppression[J]. 18th International Conference on Pattern Recognition (ICPR'06) (2006), 850-855, 10.1109/ICPR.2006.479.

[5]R.GIRSHICK. Rich feature hierarchies for accurate object detection and semantic segmentation[J]. Image Net Large-Scale Visual Recognition Challenge Workshop[S.I.]: ICCV Press, 2013: 10-15.

[6]J. R. R. UIJLINGS, K. E. A. SANDE,T. GEVERS, A. W. M. SMEULDERS. Selective Search for Object Recognition[J]. International Journal of Computer Vision.2013 (2).

[7]R.GIRSHICK. Fast R-CNN[J]. IEEE International Conference on Computer Vision (ICCV), 2015.

[8]REN S, HE K, GIRSHICK R.et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. International Conference on Neural Information Processing Systems, 2015.

[9]JAZAYERI A, CAI H, ZHENG J Y, et al. Vehicle detection and tracking in car video based on motion model[J]. IEEE Transactions on Intelligent Transportation Systems, 2011, 12(2): 583-595.

[10]ZHOU Y, LIU L, SHAO L, et al. Fast automatic vehicle annotation for urban traffic surveillance[J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 19(6): 1973-1984.

作者简介：

单志勇(1972-)，男，上海松江人，博士，副教授，研究方向为电磁场与微波技术、电磁计算与电磁兼容、天线理论与工程设计、通信信号与信息处理

通信作者：宫加辉(1997-)，女，山东济宁人，硕士研究生，研究方向为图像处理、深度学习，E-mail: 1466870362@qq.com

收稿日期：2021-03-02　　修稿日期：2021-03-27

Vehicle Detection Method Based on Fast R-CNN

SHAN Zhiyong, GONG Jiahui

(Donghua University, Shanghai 201600)

Abstract:
As an important part of traffic management system, vehicle target detection has important research significance. In order to solve the problem of low accuracy brought by traditional vehicle target detection, a vehicle target detection method based on improved Faster R-CNN algorithm is proposed. Based on the original Faster R-CNN algorithm, the improved Faster R-CNN algorithm randomly selects 960×540, 900×500 and 800×480 training images for training, and expands the proportion of candidate regions in RPN(Region Proposal Network) by 1:3 and 3:1. The mAP of improved target detection is 95.6%, which is 0.1% higher than that of vehicle target detection based on Faster R-CNN algorithm.

Keywords:
Faster R-CNN Algorithm; Target Detection; Deep Learning

(上接第 73 页)

Confusion Matrix Classification Performance Evaluation and Python Implementation

YU Ying[1,2], YANG Tingting[1,2], YANG Boxiong[1,2]

(1. School of Information and Intelligent Engineering, University of Sanya, Sanya 572011;
2. Academician Chen Guoliang's Workstation, University of Sanya, Sanya 572011)

Abstract:
This paper discusses the confusion matrix, and how to use Scikit-learn to learn and classify confusion matrix. Then introduces how the accuracy, precision and recall are calculated, and how they relate to evaluating deep learning models.

Keywords:
Confusion Matrix; Scikit-Learn; Accuracy; Precision; Recall