

基于深度信息融合的航拍车辆检测

陶晓力,刘宁钟,沈家全

(南京航空航天大学 计算机科学与技术学院,江苏 南京 211106)

摘要:随着汽车数量的快速增长以及无人机飞控技术的迅速发展,基于无人机航拍的车辆检测技术越来越有用武之地。传统的基于滑动窗口以及手工设计特征的车辆检测不仅计算量巨大,鲁棒性也不够好。卷积神经网络在目标检测方面发挥了显著的优势,但是常见的网络对于航拍遥感图像中的小目标检测效果一般。文中基于 faster-RCNN 在 VGG16 网络上使用通道合并融合的方式设计了超特征图,通过结合浅层特征以及深层特征的方式提取小目标的特征以提高检测的召回率。同时修改 RPN 层的包围框的大小以提高检测的准确性。在慕尼黑车辆数据集以及自己收集的数据上进行了测试,通过对比实验可知,该方法使得车辆检测的效果有了明显提升,在两个数据集上分别达到了 72.3% 和 80.5% 的 mAP。

关键词:车辆检测;无人机;卷积神经网络;超特征图;小目标检测

中图分类号:TP31

文献标识码:A

文章编号:1673-629X(2019)09-0117-05

doi:10.3969/j.issn.1673-629X.2019.09.023

Aerial Vehicle Detection Based on Depth Information Fusion

TAO Xiao-li, LIU Ning-zhong, SHEN Jia-quan

(School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

Abstract: With the rapid development of automotive and UAV technology, vehicle detection technology based on UAV aerial photography is becoming more and more useful. Traditional vehicle detection based on sliding windows and manual design features is not only computationally intensive but also not robust enough. Convolutional neural network plays a significant advantage in target detection, but common networks have a general effect on small targets in aerial remote sensing images. In this paper, we design the hyper feature map on the VGG16 network using channel merge and fusion based on faster-RCNN. The features of small targets are extracted by combining shallow features and deep features to improve the recall rate. At the same time, the size of the bounding box of the RPN layer is modified to improve the accuracy of detection. We test the model on the Munich vehicle dataset and the data collected. The experiment shows that the proposed method has significantly improved the effectiveness of vehicle detection compared to existing methods, reaching 72.3% and 80.5% of mAP on the two datasets.

Key words: vehicle detection; UAV; convolutional neural network; hyper feature map; small target detection

0 引言

随着无人机技术的快速发展,无人机在遥感图像拍摄中的应用越来越频繁。无人机技术的优点在于低成本,设备尺寸小,安全,最重要的是能够快速和按需完成任务。现阶段,无人机技术的发展已经能够为人们提供具有极高分辨率的遥感图像,同时能够包含空间信息如 GPS 定位等。这使得基于无人机图像的分析应用场景大大扩展,包括植被检测、电气设备检查、

城市交通监测等等。

近年来,随着汽车数量的快速增长,在给人们的生活和工作带来便利的同时,交通事故、道路拥挤等问题也变得更加频发。因此,结合无人机检测技术的智能交通越来越受到交通部门以及学者的关注。而车辆检测作为智能交通领域中绕不开的点,更是研究的重点。然而,由于无人机航拍图像中的车辆尺寸相对较小且方向可变,自动车辆检测存在着很多困难与挑战。此

收稿日期:2018-11-09

修回日期:2019-03-12

网络出版时间:2019-04-24

基金项目:国家自然科学基金(61375021);南京航空航天大学研究生创新基地(实验室)开放基金(kfj20171608);中央高校基本科研业务费专项资金

作者简介:陶晓力(1993-),男,硕士,CCF 会员(89015G),研究方向为计算机视觉和模式识别;刘宁钟,教授,博导,研究方向为计算机视觉和模式识别。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20190424.1055.066.html>

外,在复杂背景下的目标实时检查也是一大难点。

1 相关工作

根据现有的目标检测方法,检测任务主要分为三个部分:特征提取,候选区域的生成以及分类^[1]。在大部分已经存在的车辆检测方法中,手工设计的特征应用广泛。文献[2-3]使用类哈尔特征(Haar-like)和局部二值化模式(LBP)进行特征提取。文献[2]使用了尺度不变特征变换描述符(SIFT)。文献[4]中使用了方向梯度直方图。这些算法通过滑动窗口的方法,以预设的窗口遍历整张特征图,然后使用 SVM 或 adaboost 等分类器对提取的特征进行预测,以判断窗口中是否存在可能的车辆^[5-6]。

目标检测的性能表现主要依赖于特征提取。然而上述方法并不能很好地适用于所有的应用场景,尤其是复杂背景下的车辆检测。同时滑动窗口的方式会导致过多的重复计算,在检测效率上也不尽如人意。

最近随着深度学习的兴起,卷积神经网络(CNN)在目标检测领域取得了重大的进步。多种检测模型包括 RCNN^[7]、Fast-RCNN、Faster-RCNN^[8]、YOLO^[9]、SSD^[10]等都有着良好的表现,在识别精确率以及检测时间上都优于传统的人工设计特征的检测算法。

RCNN 使用选择性搜索算法来生成候选区域,之后对每一个区域通过卷积神经网络提取深度特征,并使用 SVM 分类器进行分类。RCNN 是最早的基于 CNN 的目标检测算法,尽管识别准确率有了极大的提升,但由于大量的重复计算以及候选区域的生成,CNN 的训练和 SVM 都不在一个网络中,导致训练的繁琐和检测时间的过长。Faster RCNN 使用区域建议网络(RPN)来生成候选区域,并与特征提取的卷积神经网络共享参数,之后通过全连接层实现目标分类,真

正意义上实现了端到端的网络结构,加快了网络的训练和检测速度。

目前,Faster RCNN 在目标检测方面达到了一个领先的水平,但是,在航拍车辆检测方面仍然存在缺陷。由于粗糙的特征图设计,RPN 网络不适合小目标检测。同时,在航拍图像的复杂大背景下,误识别也影响着模型的识别效果^[11-12]。

2 模型和方法

2.1 模型

ZFNet^[13]和 VGG16^[14]是两类典型的基于 CNN 的目标检测网络。ZFNet 是一种轻量型的模型,它的模型参数较少,主要包括 5 层卷积层以及 2 层全连接层。VGG 模型是一系列不同深度的神经网络模型,包括 VGG13、VGG16、VGG19 等。这一系列模型显示了随着网络结构的加深,模型的表现将会得到提升。其中 VGG16 的综合表现最好,因此,文中将在 VGG16 的基础上进行优化,结合 Faster RCNN 模型,构建出可行的车辆检测网络。

VGG16 网络共包含 13 层卷积网络以及紧接其后的三层全连接网络,若按照网络的输出大小划分,也可以将前面 13 层卷积网络分为 5 部分。不同于传统的将第五部分的输出特征图直接传递给 RPN 网络,在文中网络,通过添加 concat 层将后面的三部分结合形成新的超特征层,具体的网络结构如图 1 所示。串联模型可以通过学习权重的融合获得目标体系结构信息,减少背景噪声对检测性能的影响。同时,相关研究表明,更深的网络可以得到更高的召回率,而浅层的网络可以得到更好的准确率。因此,结合浅层特征和深层特征的方式可以有效地提高模型的检测效果。

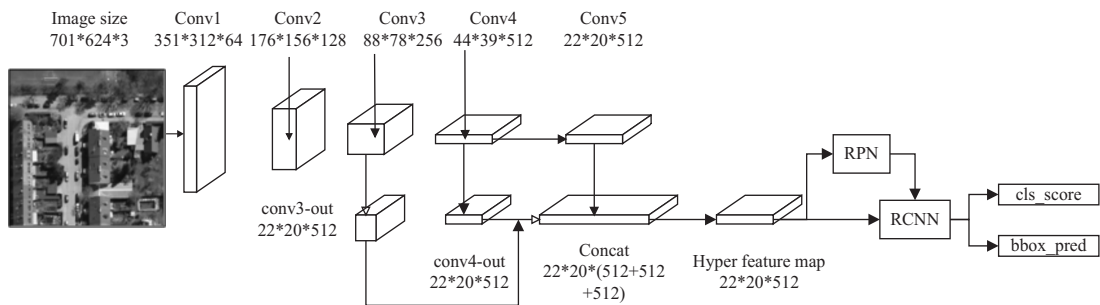


图 1 网络结构

超特征图如图 1 所示,这是基于 VGG16 的卷积网络,输入为 701 * 623 * 3 的 jpg 格式图片。因为要将 conv3,conv4,conv5 三部分的输出通过 concat 层链接,而由于这三部分的输出结果的长宽不一样,所以具体的做法是使用 512 个大小为 3 * 3 * 256 的卷积核对 conv3 的输出进行卷积,步幅为 2,然后再接一层

maxpool 层,得到的结果为 22 * 20 * 512。使用 512 的大小为 3 * 3 * 512 的卷积核对 conv4 的输出进行卷积,步幅为 2,得到的结果为 22 * 20 * 512。接着使用 concat 层将 conv3_out,conv4_out 和 conv5 进行通道合并,使用 512 个 1 * 1 * 1 536 卷积核进行卷积,得到所需的超特征层,大小为 22 * 20 * 512。

2.2 训练过程

输入图片经过一系列卷积之后变成超特征图传入 RPN 层。RPN 层的目的是预测所有可能的候选区域,并给出前背景的概率^[9]。通过 3 * 3 的滑动窗口,RPN 层将得到的超特征图的对应部分变成一个 512 维的特征向量,然后对其进行前背景的分类以及包围框的回归。因为对于每一个滑动窗口预测 k 个物体可能存在的位置,所以共有 $2k$ 个前背景的分值以及 $4k$ 个包围框的值。如图 2 所示,文中 k 的值是 9。

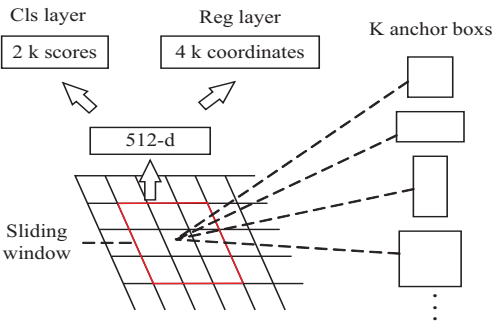


图2 RPN 层

由于车辆数据有限,使用基于 ImageNet 得到的 VGG16 预训练模型初始化该网络。在训练的每一次迭代过程中,如果预测的区域 B_p 与真实的包围框 B_g 的交并比 (IOU) 大于 0.7,那么设置一个正标签 ($p_i^* = 1$),如果 IOU 小于 0.3,那么设置一个负标签 ($p_i^* = 0$),忽略 IOU 在 0.3 ~ 0.7 之间的包围框。交并比的定义如下:

$$IOU = \frac{\text{area}(B_p \cap B_g)}{\text{area}(B_p \cup B_g)} \tag{1}$$

其中, $\text{area}(B_p \cap B_g)$ 表示 B_p 与 B_g 的交集的面积; $\text{area}(B_p \cup B_g)$ 表示 B_p 与 B_g 的并集的面积。

使用多任务损失函数来反向传播用以更新网络参数,目的是最小化分类和回归的错误率。多任务损失函数定义如下:

$$L(p_i, loc_i) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{bbr}} \sum_i p_i * L_{bbr}(loc_i, loc_i^*) \tag{2}$$

其中, i 表示一个训练批次中包围框的索引; p_i 和 p_i^* 分别表示每一个包围框是否是目标的预测概率以及真实标签; N_{cls} 和 N_{bbr} 分别表示每个训练批次的图片区域的个数以及预测区域的个数; L_{cls} 表示车辆和背景分类的对数损失函数; L_{bbr} 表示包围框的回归损失函数:

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

$$L_{bbr}(loc_i, loc_i^*) = \text{smooth}_{L_i}(loc_i - loc_i^*) \tag{4}$$

其中, $\text{smooth}_{L_i}(x) = \begin{cases} 0.5x^2 & |x| < 1 \\ |x| - 0.5 & \text{其他} \end{cases}$; loc_i 是一个向量,表示预测的包围框的 4 个参数化坐标; loc_i^* 是与正包围框对应的坐标向量。

RPN 网络中最重要的是预测包围框的生成。上文提到过卷积之后图片大小成为了原来的 1/32,即特征图上的一点对应于原图中 32 * 32 大小的一个区域。在 VGG16-faster-RCNN 中,包围框的大小设置:基础大小为 16,比例变换为 (0.5, 1, 2),尺度变换为 (8, 16, 32),因此最终得到三类尺度的包围框 (256, 512, 1 024),对应比例变换后得到 9 种包围框,如图 3 所示。

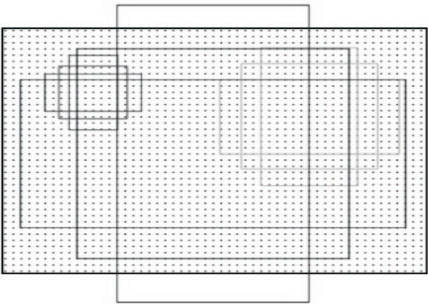


图3 9 种包围框

从包围框的大小可知,如此之大的包围框只适合正常大小的目标,不适用小目标检测。根据航拍车辆的大小,在包围框基础大小和比例变换不变的情况下,修改尺度变换为 (2, 4, 8),最终得到的包围框大小如表 1 所示。

表 1 各包围框大小			
模型	比例=2 : 1	比例=1 : 1	比例=1 : 2
VGG16	360 * 176	256 * 256	184 * 368
	720 * 352	512 * 512	368 * 736
	1 440 * 704	1 024 * 1 024	736 * 1 472
文中模型	90 * 44	64 * 64	46 * 92
	180 * 88	128 * 128	92 * 184
	360 * 176	256 * 256	184 * 368

3 实验结果与分析

3.1 数据集

使用了两个数据集进行实验对比,一是慕尼黑航拍车辆数据集,另一个是自己使用无人机收集的南京石杨路和东麟路的航拍车辆数据。

DLR 3K 慕尼黑航拍车辆数据收集的车辆地点是德国城市慕尼黑。这些图像是通过 Canon Eos 1Ds Mark III 相机从飞机上拍摄的,分辨率为 5 616 * 3 744 像素,焦距为 50 mm,并以 JPEG 格式存储。图像拍摄自地面 1 000 米的高度,地面采样距离约为 13 厘米。

数据集中共包括 20 张图片,其中 10 张为训练集,10 张为测试集,共细分为 7 种类型的车。将 pkw_car 和 pkw_van 合并为 car,将 cam 和 truck 合并为 truck,这样共得到 car 8 381 辆, truck 130 辆。将每张 5 616 * 3 744 的原始图片裁剪成 11 * 10 的有重合的图像块,每个图像块大小为 702 * 624。最后将每个图像块通过旋转 90°,180°,270°的方式增加 4 倍,共得到 8 800 张图。

自己收集的车辆数据包括石杨路和东麟路两部分,拍摄点据地面约为 100 米,分辨率为 5 472 * 3 078。其中石杨路 376 张,包含 3 565 辆汽车以及 127 辆卡车共 3 692 辆,东麟路 101 张,包含 1 930 辆汽车以及 318 辆卡车。将原图像的长宽缩小 1/4 到 1 368 * 770,同时通过镜像变换将车辆样本增加 4 倍。

3.2 结果对比

采用 mAP 指标来衡量模型的好坏,定义如下:
$$mAP = \int_0^1 P(R) dR \tag{5}$$
其中, R 表示召回率; P 表示该召回率下的准确率。

$$P = \frac{TP}{TP + FP} \tag{6}$$

$$R = \frac{TP}{TP + FN} \tag{7}$$

其中,TP 表示将正样本预测为正的数量;FP 表示将负样本预测为正的数

量;FN 表示将正样本预测为负的数量。

实验结果如表 2 所示。分别在慕尼黑数据集和自

己的数据上运行了 ZF 模型,改进后的 ZF 模型, VGG16 模型以及改进后的 VGG16 模型。可以看到,改进后的模型相较于原始模型,其 mAP 都有了较大的提升,VGG16 模型的效果又好于 ZF 模型。而最终基于 VGG16 改进的方案得到了最好的效果。同时,在自己数据集上的效果比慕尼黑数据集上好的原因在于,自己收集的车辆数据集分辨率更高、更清晰,特征相较于慕尼黑车辆数据集更为明显。部分实验结果如图 4 和图 5 所示。

表 2 不同模型的 mAP

模型	慕尼黑数据集	自己收集的数据集
ZF	0.617	0.676
改进的 ZF	0.673	0.732
VGG16	0.694	0.774
文中方法	0.723	0.805

4 结束语

基于 Faster RCNN 设计了一个具有较好精确率和鲁棒性的航拍车辆检测模型。建立一个超特征图,将深度网络的浅层特征和深层特征相结合,进行目标车辆的特征提取。实验结果表明,该方法在小目标检测上具有良好的效果。同时,在 RPN 层上,通过先验知识,根据车辆的大小设计了合适的 anchor box,也提升了检测模型的性能。另外,该方法也存在不足,仍旧会有小部分的车辆漏检。因此,将会继续深入研究网络模型,在提升准确率和召回率的同时,减少检测时间。



图 4 慕尼黑车辆数据结果



图5 自己收集的车辆数据结果

参考文献:

- [1] TANG Tianyu, ZHOU Shilin, DENG Zhipeng, et al. Vehicle detection in aerial images based on region convolutional neural networks and hard negative example mining[J]. Sensors, 2017, 17(2): 336–352.
- [2] SHAO Wen, YANG Wen, LIU Gang, et al. Car detection from high-resolution aerial imagery using multiple features [C]//IEEE international geoscience and remote sensing symposium. Munich, Germany; IEEE, 2012: 4379–4382.
- [3] 朱 彬, 王少平, 梁华为, 等. 基于 Haar-like 和 MB-LBP 特征分区域多分类器车辆检测[J]. 模式识别与人工智能, 2017, 30(6): 569–576.
- [4] KLUCKNER S, PACHER G, GRABNER H, et al. A 3D teacher for car detection in aerial images[C]//IEEE 11th international conference on computer vision. Rio de Janeiro, Brazil; IEEE, 2007: 1–8.
- [5] TUERMER S, KURZ F, REINARTZ P, et al. Airborne vehicle detection in dense urban areas using HoG features and disparity maps[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2013, 6(6): 2327–2337.
- [6] 陈拥权, 陈 影, 陈学三. 基于 Adaboost 分类器的车辆检测与跟踪算法[J]. 计算机技术与发展, 2017, 27(10): 165–168.
- [7] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. [s. l.]: IEEE, 2014: 580–587.
- [8] REN Shaoqing, HE Kaiming, 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. Montreal; MIT Press, 2015: 91–99.
- [9] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. [s. l.]: IEEE, 2016: 779–788.
- [10] LIU Wei, ANGUELOV D, ERHAN D, et al. Ssd: single shot multibox detector [C]//European conference on computer vision. Cham: Springer International Publishing, 2016: 21–37.
- [11] 阮 航, 王立春. 基于特征图的车辆检测和分类[J]. 计算机技术与发展, 2018, 28(11): 39–43.
- [12] 孙秉义, 文珊珊, 吴 昊, 等. 基于深度学习的高分辨率遥感图像车辆检测[J]. 东华大学学报: 自然科学版, 2018, 44(4): 520–525.
- [13] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks[C]//European conference on computer vision. Cham: Springer International Publishing, 2014: 818–833.
- [14] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [C]//International conference on learning. [s. l.]: [s. n.], 2015: 1150–1210.