

DOI:10.16356/j.1005-2615.2019.06.001

基于改进 Faster-RCNN 的机场场面小目标物体 检测算法

韩松臣¹ 张比浩¹ 李 炜¹ 汤新民² 付道勇¹

(1. 四川大学空天科学与工程学院, 成都, 610065; 2. 南京航空航天大学民航学院, 南京, 211106)

摘要: 针对目前应用于机场视频监控中的卷积神经网络方法存在小目标物体识别准确率较低的问题, 本文提出了一种基于 Faster-RCNN 并结合多尺度特征融合与在线难例挖掘的机场场面小目标检测算法。该算法采用 ResNet-101 作为特征提取网络, 并在该网络基础上建立了一个带有上采样的“自顶向下”的特征融合模块, 以生成语义信息更加丰富的高分辨率特征图。并在网络训练过程中, 采用在线难例挖掘的训练策略使模型更加鲁棒地对小目标样本进行定位。最后, 手工构建了一个包含 5 982 张图片的机场数据集, 用于检测模型的训练和测试。结果表明, 本文所提出算法显著提升了机场场面小目标物体检测的准确率, 且使整体平均检测准确率达到 80.8%, 该结果高于其他先进的目标检测模型。

关键词: 机场场面监视; 多尺度特征融合; 在线难例挖掘; 小目标物体检测

中图分类号: V351 **文献标志码:** A **文章编号:** 1005-2615(2019)06-0735-07

Small Target Detection in Airport Scene via Modified Faster-RCNN

HAN Songchen¹, ZHANG Bihao¹, LI Wei¹, TANG Xinmin², FU Daoyong¹

(1. School of Aeronautics and Astronautics, Sichuan University, Chengdu, 610065, China; 2. College of Civil Aviation, Nanjing University of Aeronautics & Astronautics, Nanjing, 211106, China)

Abstract: For the low precision of small target detection in the existing convolution neural network methods used in airport video surveillance, in this paper, a small target detection algorithm based on Faster-RCNN combined with multi-scale feature fusion and online-hard-example-mining (OHEM) is proposed. First of all, ResNet-101 is adopted as the feature extraction backbone, and a top-down multi-scale feature fusion pathway is established based on the ResNet-101 to generate richer semantic feature maps of a fine resolution. During the network training, OHEM is adopted to make the network more robust to locate the region of small target objects. At last, an airport dataset containing 5 982 pictures is constructed manually, which is used to verify the training and testing of the model. The results show that our modified Faster-RCNN algorithm significantly improves the accuracy of small target detection under airport situation. Besides, the mean average precision reaches 80.8%, which is higher than other advanced object detection models.

Key words: airport surface surveillance; multi-scale feature fusion; online-hard-example-mining (OHEM); small target detection

基金项目: 国家重点研发计划专项(2018YFC0809500)资助项目。

收稿日期: 2019-10-27; **修订日期:** 2019-11-30

作者简介: 韩松臣, 四川大学空天科学与工程学院教授、博导, 教育部高等学校教学指导委员会委员。1997年毕业于哈尔滨工业大学飞行器设计专业, 获得博士学位。先后任教于哈尔滨工业大学、南京航空航天大学、四川大学, 从事与航空航天领域的教学与科研工作。发表论文 120 余篇, 其中被 SCI/EI/ISTP 收录 50 余篇。申请并获得授权专利 10 余项, 多次获得省部级科技进步奖励。研究方向: 空管系统自动化、面向飞行器控制的信息工程与人工智能技术。E-mail: hansongchen@scu.edu.cn。

通信作者: 李炜, 男, 博士, 副教授, E-mail: li.wei@scu.edu.cn。

引用格式: 韩松臣, 张比浩, 李炜, 等. 基于改进 Faster-RCNN 的机场场面小目标物体检测算法[J]. 南京航空航天大学学报, 2019, 51(6): 735-741. HAN Songchen, ZHANG Bihao, LI Wei, et al. Small Target Detection in Airport Scene via Modified Faster-RCNN[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2019, 51(6): 735-741.

传统的机场场面监控通常依靠雷达监视系统进行,但其对于非协作目标无法有效监视,且随着场面非协作目标的增多将导致检测效率的下降。依靠序列图像的自动检测,从而对机场场面中的进离场飞机,地面管制人员与车辆进行有效的定位和识别,不仅能有效减少由繁杂而枯燥的人工监控工作带来的潜在人为疏忽,还能消除机场场面环境下可能存在的监控盲点,因而成为保障机场场面安全与有序运行的有效手段^[1],其重要性日渐突出。

近年来,基于卷积神经网络的目标检测算法 RCNN^[2]系列,尤其是 Faster-RCNN^[3],其目标识别的准确率在公开数据集 Pascal VOC^[4]的高分辨率大目标物体检测领域中取得了突破性的提升,已经被证明有利于复杂图像特征的提取^[5]。然而,这一系列的检测模型对像素面积较小的物体的检测效果仍不理想,因而并不适合直接应用于小目标物体的检测任务中。根据公开数据集 MS-COCO^[6]的定义,像素数量低于 32×32 面积的目标即为小目标物体,也有文献^[7]将目标所占区域相对图像尺寸大小低于 1% 的物体认定为小目标。在实际的机场场面监控环境中,由于机场场面监控摄像机往往固定在较高处的地面塔台中^[8],其监控视频中往往存在大量的低像素面积物体。图 1 展示了对所标定的机场场面数据集进行的物体尺寸的统计结果,从中可以观察到,无论是基于绝对尺寸或是相对尺寸对于小目标物体的定义,机场监控环境下都存在着大量的小目标物体。因此,提升机场场面环境中小目标检测的准确率成为保障场面安全与提升场面整体目标检测率的关键因素。

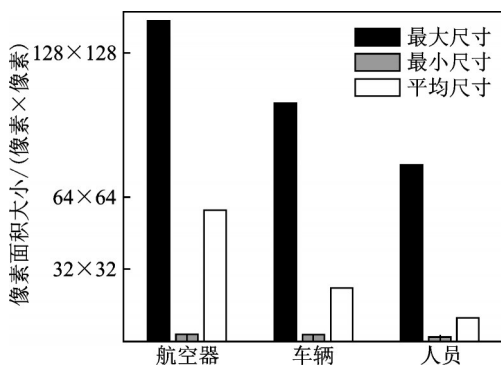


图1 机场数据集中各类物体像素面积统计
Fig. 1 Histogram of pixel area class by class

然而目前大部分先进的目标检测模型由于两个主要原因的限制并不适合直接应用于小目标物体的检测任务中^[9]。首先,无论是单阶段检测算法 SSD^[10]和 YoloV3^[11],或是双阶段检测算法 Faster-RCNN,其都依靠一组预定义尺寸大小和数量的锚箱对可能存在物体的位置提取区域候选框,然

而该组锚箱通常不能有效地覆盖小目标物体的大小比例,因而无法对其进行有效的定位。其次,现有的基于卷积神经网络架构的目标检测算法模型在对输入图像进行特征提取的过程中,在不同层级的特征图之间均采用池化卷积^[12]的操作对特征进行下采样以得到语义丰富的特征,但该操作也导致最后用于区域建议生成和物体分类的顶层特征图的分辨率远远小于输入图像^[13]。这意味着若一个场面管制车辆在监控视频中的像素面积大小为 32×32 ,其通过特征提取网络 VGG-16^[14]进行 5 次池化下采样操作后,该车辆在顶层特征图中仅只能被一个像素所表征,检测模型只能依靠该像素对物体的类别和位置进行预测,这往往会造成最终的错检或漏检。

在本文中,针对上述小目标物体检测的难点问题,改进了先进的目标检测模型 Faster-RCNN 使其更适用于机场场面的监控环境中的小目标检测。首先进行了更精细化地锚箱设计,结合在线难例挖掘^[15]的训练策略,使模型的训练能充分覆盖机场数据集中小目标物体的大小范围。其次,不同于传统的序列图像特征图融合方式^[16],采用自顶向下的多尺度特征图融合方法生成同时具有丰富语义信息与高分辨率的特征图。该方法依靠上采样操作放大顶层特征图,通过卷积进行不同尺度特征图的融合,能有效结合底层与高层特征图的特点,从而弥补了池化操作所带来的小目标物体特征丢失,这也是本文改进的 Faster-RCNN 能有效检测小目标物体的关键手段。

1 研究方法

1.1 网络整体架构概述

本文算法模型框架主要分为 Faster-RCNN 和自顶向下的特征图融合模块两部分,如图 2 所示。

首先采用 ResNet-101^[17]残差网络作为 Faster-RCNN 的特征提取骨干网络用于提取语义信息更加丰富的特征图,在此基础上利用一个具有水平连接和上采样操作的自顶向下的特征图融合模块将深,浅层特征图融合。并将所得的特征图作为 RPN 的输入。Fast-RCNN^[18]在对 RPN 输出的候选框进行进一步的校正,得到最终的检测结果。

1.2 自顶向下的多尺度特征融合网络

通常而言,对输入图像进行特征提取的卷积神经网络通过逐层计算的方式得到特征图层次结构,由于在特征层次结构中每层特征图之间会存在池化层,池化卷积操作在增强特征图语义信息的同时也会降低特征图分辨率,最终得到的图像特征图层

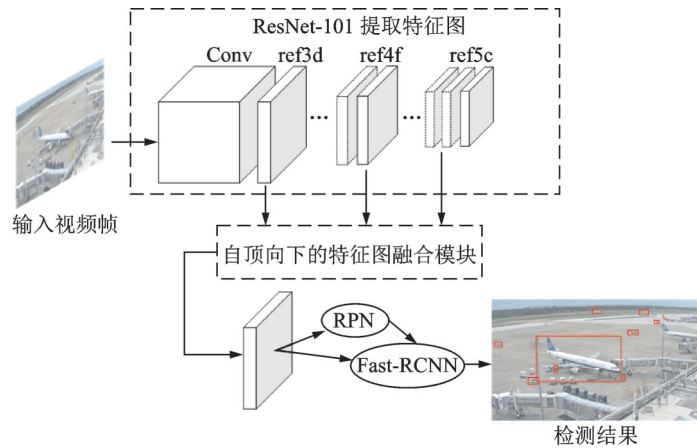


图 2 改进版Faster-RCNN的模型整体架构图

Fig. 2 Overall architecture of the proposed modified Faster-RCNN

级呈现为多尺度的特征图金字塔结构。在Faster-RCNN目标检测算法中,其仅依靠特征金字塔中的最顶层特征图,部署区域建议生成网络(RPN)来得到物体候选框,并进行后续的分类任务。特征金字塔的顶层特征虽然具有高度抽象的语义信息,但由于进行了多次的池化子采样操作,其对物体的边缘或几何细节信息等特征不敏感,小目标物体难以被表征^[19]。这给现有算法对小目标物体检测带来了诸多挑战。

值得注意的是,在特征图的金字塔结构中,浅层的特征图往往具有高分辨率和丰富的几何细节特征的特点,相反的是,顶层的特征图通常具有更强的语义抽象信息且对物体的姿态、位置变化等具有鲁棒性,但其分辨率较低。因此,一个有效的做法是在自低向上的前馈神经网络结构基础上,采用一个具有水平连接的自顶向下的特征图融合结构,将顶层特征图通过上采样手段放大分辨率,通过卷积将特征金字塔中深层、浅层特征图结合起来,生成同时具有高分辨率且语义信息更加丰富的特征图。本文采用的自顶向下的特征图融合结构如图3所示。

由于采用ResNet-101结构对输入图像进行特征提取,将提取到的特征图金字塔中的第3层,第4层,和第5层称为{ref3d, ref4f, ref5c}。考虑到该3层特征图的特征通道维度不同,因此在其进入自顶向下的特征融合通道之前,分别对每层特征图采用了 1×1 的卷积层以调整特征图通道维数为一个合适的固定大小。与此同时,要进行特征图融合的上述3层特征图分别相较于输入图像具有{8, 16, 32}像素大小的卷积步长,而最后通过特征融合网络输出的特征图则只具有8像素大小的卷积步长。因此,在采用逐像素相加的操作以融合不同层级特征图之前,对高层特征图进行了双线性插值运算以放

大其空间尺寸,使其与要融合的特征图的空间分辨率相一致。通过从最顶层特征图迭代进行上述操作,得到不同层级的融合特征图。最后,在融合上述3层的特征图之后,在融合生成的最后层特征图后额外添加了一个 3×3 的卷积层,用于降低由上采样操作所可能带来的特征混淆效应,并最终得到同时具有高分辨率与丰富语义信息的特征图,如图3黄色特征图所示。

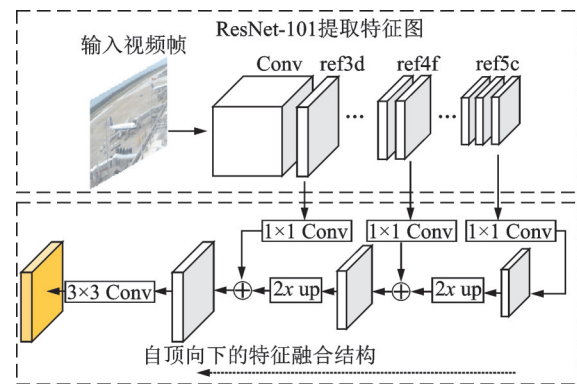


图 3 自顶向下的特征融合网络模块

Fig. 3 Top-down module of feature map fusion

1.3 在线难例挖掘

考虑到机场场面目标存在类别少,尺寸不一,监控视频的分辨率低等特点,若直接以正负样本比例极不平衡的感兴趣目标区域(Region of interest, RoI)^[20]集合进行网络训练,将导致训练收敛速度慢,检测模型容易对视频监控中的小目标误检,漏检。为此,Faster-RCNN在网络训练过程中采用随机采样策略,限定RoI集合中的正负样本比例,在一定程度上来缓解上述问题。

在本文的改进的Faster-RCNN的训练过程中,采用在线难例挖掘(Online hard example mining, OHEM)方法,取消人为预设定的正负样本比例,充分挖掘值得用于训练的“高质量”训练样本,

从而使网络能充分学习到容易发生错误判别的物体的特征,提高网络的训练收敛速度和算法的检测精确度。本文的具体做法步骤如下:

(1) 对于每张输入的训练图像,首先利用 ResNet-101 和特征融合网络计算出用于输入 RPN 的特征图。

(2) 利用(1)中得到的特征图和 RPN 网络计算得到 2 000 个 RoI 并作为 Fast-RCNN 的输入。

(3) 利用 Fast-RCNN 中的网络前向传播计算得到各 RoI 的损失函数值,并根据损失函数值对这 2 000 个 RoI 进行降序排列。

(4) 由于在同一张图中的 RoI 之间可能存在很大部分相互重叠情况。因此,如果一个 RoI 所带来的损失函数值较大,那么和它重叠较大的 RoI 很大程度上也会较大,这样在反向传播的时候就等于进行了一次重复的梯度计算。为避免这种情况,对降序排列的 RoI 进行一次非极大值抑制,过滤掉重叠面积较大的 RoI。最后在从剩余 RoI 中选择前 128 个 RoI 进行反向传播。实验中,非极大值抑制的重叠阈值选取为 0.7。

2 实验和数据分析

2.1 机场场面监控数据集

由于基于图像的机场目标监视领域缺乏公开数据集,因此本文构建了一个包含航空器、车辆和地面管制人员这 3 类机场场面目标的目标检测数据集,以供深度学习网络训练与模型测试使用,部分数据集图片示例如图 4 所示。其中所有的图像数据均来自贵阳、成都和海口机场的日间监视视频。



图 4 机场数据集部分样本图片

Fig. 4 Samples from the proposed airport dataset

首先对获取的视频进行预处理,“清洗”,剪切掉视频中非机场场面画面和一些由摄像头旋转得到的无关画面。每隔 80 帧视频图像提取一张图片,得到图片的分辨率为 $1\,920 \times 1\,080$ 和 960×540 像素两种,共计 5 982 张有效图片,27 320 个物体标

注实例。最后,这 5 982 张图片被严格按照公开数据集 Pascal VOC2007 的格式对每张图像中的物体实例进行相应手工标注。将所得数据集按照 3:1 的比例分为训练集与测试集,并计算训练集中的像素均值,在进行网络训练时,输入的每张训练图片会减去该均值,再进行训练,这个技巧会加快网络模型训练收敛的速度。数据集部分样例图片如图 4 所示,每张图片的停港飞机周围都存在着大量的小目标物体,且远处正在离港或进港的飞机在图像中也具有非常小的像素面积。

2.2 实施细节与评价标准

本文实验基于 TensorFlow 深度学习框架进行,选用一块 NVIDIA GTX-1080Ti 显卡和 Intel (R)Core I7-8700K CPU 作为硬件平台,程序运行环境为 Ubuntu 16.04LT 操作系统。基于深度学习领域中迁移学习的思想,使用在 ImageNet 分类任务中预训练的 ResNet-101 模型,初始化网络权重参数。对于 RPN 网络与 Fast-RCNN 网络的训练过程,采用 Faster-RCNN 原论文中的交替训练方式,对其分别训练。本文网络优化方法采用随机梯度下降(Stochastic gradient descent SGD),对网络进行 4 万次迭代训练,其计算式如式(1)。其中每一批量的训练数据包含 128 张训练样本图像,初始学习率为 0.01,在 2 万次迭代后学习率衰减为 0.000 1,最后设置动量因子为 0.9 的进行误差的反向传播。除此之外,还采用了图像水平旋转,随机裁剪,尺度缩放等数据增强技巧来提高模型网络的收敛速度

$$\theta_i = \theta_i - \alpha \sum_{j=1}^m (h_\theta(x^{(j)}) - y^{(j)}) x_i^{(j)} \quad (1)$$

式中: θ_i 为算法中的第*i*个参数; α 为手动设置的学习率; m 代表训练样本的数量; $h_\theta(x)$ 表示为参数的函数; $x^{(j)}$ 表示第*j*组的*x*; $y^{(j)}$ 表示第*j*组的*y*。

对于实验结果的评价,采用 AP(Average precision)平均精度对检测精确度进行衡量

$$J = \sum_{n=1}^M P/M \quad (2)$$

式中: J 为要求的某一类的平均精度, M 为总体样本中的正例个数, P 为在不同正例个数下对不同的召回值(recall)中最大的准确率(precision)。AP 值可以衡量通过训练得到的模型在检测每个类别上的表现程度,mAP(mean average precision)是模型在检测所有类别上的好坏,通过将每个类别所对应的 AP 值算术平均后得到。

2.3 实验结果与分析

为了更好的理解本文方法的对检测结果的影响,通过在机场数据集中采用不同的实验策略进行实验。首先探究了特征提取网络 ResNet-101 对于小目标检测效果的影响。对比了分别基于 ZF-Net, VGG-16 以及 ResNet-101 的 Faster-RCNN 检测模型。为了公平对比不同实验,除使用的特征网络不同以外,参数设置,训练步骤均保持一致,实验结果见表 1。

表 1 不同特征提取网络的检测效果比较

Tab. 1 Comparison of parameters between different feature extraction Networks

特征网络	飞机	人员	车辆	mAP
ZF-Net	90.6	7.1	69.9	55.9
VGG-16	90.7	19.1	79.8	63.2
ResNet-101	90.7	27.2	79.6	65.8

由表 1 可见,通过替换不同的特征提取网络, ResNet-101 网络在 mAP 结果上相比 ZF-Net, VGG-16 分别提升了 9.9%, 2.6%, 虽然在航空器, 车辆这两类具有较大尺寸的目标上检测结果几乎持平,但在人员类别上的检测结果提升了 20.1% 与 8.1%。这充分证明了 ResNet-101 残差对于小目标检测的优越性。此外,考虑到原始的 Faster-RCNN 中所使用的人为预定义的锚箱尺寸大小为 $\{128^2, 256^2, 512^2, \}$, 该尺寸大小不能很好地覆盖机场数据集中的物体尺度范围。因此研究对比了不同锚箱尺寸大小在机场数据集的检测性能结果。值得注意的是,训练与测试模型时所采用的锚箱尺寸保持一致,锚箱尺寸大小分布分为 A 组 $\{128^2, 256^2, 512^2, \}$, B 组 $\{64^2, 128^2, 256^2, \}$, C 组 $\{16^2, 32^2, 64^2, 128^2, 256^2, \}$ 。实验结果见表 2。

表 2 不同锚箱尺寸的检测效果比较

Tab. 2 Comparison of parameters between different pre-defined anchor sizes

锚箱尺寸	特征网络	飞机	人员	车辆	mAP
A 组	ResNet-101	90.7	27.2	79.6	65.8
B 组	ResNet-101	90.6	49.7	88.3	76.2
C 组	ResNet-101	90.7	52.1	88.8	77.2

由表 2 可知,通过将锚箱尺寸大小调整为 $\{64^2, 128^2, 256^2\}$ 以及继续额外添加更小的锚箱 $\{16^2, 32^2\}$, 使其尺寸分布更适应于特征图感受野同时更好地覆盖数据集中物体的大小范围,以 ResNet-101 为特征网络的 Faster-RCNN 在人员与车辆上的检测 AP 值相对于原始 Faster-RCNN 的设置检测结果

分别提升了 24.9% 与 9.2%。这证明了较小的预定义锚箱能有效的提升小目标的检测结果与检测边框回归的性能。最后,以表 2 中最后一行所示模型为基准,继续添加特征融合网络与在线难例挖掘进行训练与测试,实验结果见表 3。

表 3 特征融合网络与 OHEM 对检测结果影响

Tab. 3 Ablation experiments of top-down module and OHEM

特征融合网络	OHEM	飞机	人员	车辆	mAP
✓		92.5	55.4	90.2	79.4
	✓	92.4	53.8	89.1	78.4
✓	✓	94.3	56.7	91.4	80.8

由表 3 可以看到,向基准网络添加多尺度特征融合网络后,其对所有类别的检测结果均有促进作用。其中平均像素面积最小的人员类别的检测结果提升了 3.3%,这证明了通过特征融合策略增强特征图语义信息的方法对于小目标检测的有效性。在模型训练过程中,通过加入在线难例挖掘方法,同样得到了 1.2% 的 mAP 提升,这说明其使得小目标难例样本能更好地被网络模型学习。最后,将两种有效提升小目标检测结果的方法相结合,本文方法在机场场面检测的类别平均准确度最终达到了 80.8%。除上述实验外,为更好地证明本文方法的优越性,本文方法与其他先进的基于深度学习的目标检测算法进行了对比,其结果见表 4。

表 4 本文方法与主流目标检测方法检测结果比较

Tab. 4 Results comparison between the proposed method and other methods on the airport dataset

方法	飞机	人员	车辆	mAP
原始 Faster R-CNN	90.7	19.1	79.8	63.2
SSD300	89.4	13.6	85.6	62.9
YOLOv3	92.6	23.6	90.3	68.8
本文方法	94.3	56.7	91.4	80.8

由表 4 可知,在各类物体的检测结果上,本文的方法都优于其他先进的深度学习目标检测算法。通过一系列改进,本文的方法相比原始的 Faster-RCNN 在 mAP 中提升了 17.6%,尤其是在小目标类别人员中提升了 37.6% 的 AP 值,该提升证明了本文方法的成功。最后,图 5 展示了利用本文方法与原始 Faster-RCNN 进行测试的一些实例效果图。可以看到,本文方法对于机场场面小目标检测的漏检,检测结果低置信度的情况有明显改善。

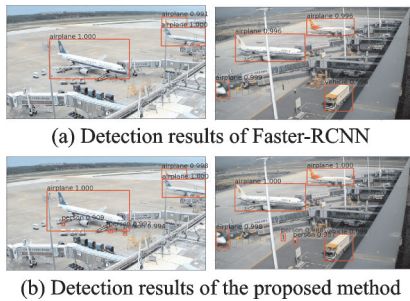


图5 两种检测方法效果对比

Fig. 5 Results comparison between the Faster-RCNN and the proposed method

3 结 论

本文提出了一种改进的Faster-RCNN目标检测模型来解决机场场面的小目标物体检测问题。本文的创新点主要总结为:

(1)提出了一种自顶向下的多尺度特征图融合方法。该操作通过上采样放大顶层特征图,通过 1×1 卷积融合ResNet-101特征提取网络结构中的不同尺度大小的特征图以生成同时具有高分辨率与高语义信息的特征,用于区域建议网络的部署以及后续物体类别的识别。实验证明,该模块能有效弥补小目标检测存在的池化“细节信息”丢失问题,提升小目标检测结果。

(2)通过分析机场场面监控数据集,进行预定义锚箱的精细化设计,使其大小、比例能更好覆盖机场场面的物体尺寸范围,并增加额外的小尺寸锚箱参与网络的训练与测试。并在训练过程中,采用在线难例挖掘的策略,加快模型训练的收敛过程,并使网络能更鲁棒的学习小目标的特征。实验表明,与原始Faster-RCNN模型对比,本文的锚箱设计与训练策略能进一步提高机场环境下的目标检测率,且对像素面积较小的类别提升更为明显。

(3)为了评估本文目标检测算法的整体效果,本文制作了基于国内各大机场监控视频的机场数据集。在实验中,通过与原始Faster-RCNN和近年来先进的目标检测算法SSD,YOLOv3对比。结果表明本文方法在机场环境下的目标检测的整体精度上优于其他检测算法,且在像素面积较小的类别中远远优于直接应用其他算法。

参考文献:

[1] SCHMIDT M, RUDOLPH M, WERTHER B, et al. Remote airport tower operation with augmented vision video panorama HMI [C]//2nd International

Conference Research in Air Transportation. [S.l.]: DLR, 2006: 221-230.

- [2] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]// 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA: IEEE, 2014.
- [3] 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.
- [4] EVERINGHAM M, van GOOL L, WILLIAMS C K I, et al. The pascal visual object classes (VOC) challenge [J]. International Journal of Computer Vision, 2010, 88(2): 303-338.
- [5] 胡彬,杨铖,邵叶秦,等.基于视频的行人再识别[J].南京航空航天大学学报,2019,51(5): 669-674.
HU Bin, YANG Cheng, SHAO Yeqin, et al. Video-based person re-identification [J]. Journal of Nanjing University of Aeronautics & Astronautics, 2019, 51(5): 669-674.
- [6] LIN T, MAIRE M, BELONGIE S, et al. Microsoft COCO: Common objects in context [C]//European Conference on Computer Vision. [S.l.]: Springer, 2014: 740-755.
- [7] HAN C, GAO G, ZHANG Y. Real-time small traffic sign detection with revised Faster-RCNN [J]. Multimedia Tools and Applications, 2019, 78(10): 13263-13278.
- [8] 廖智麟,陆晓安,黄一敏,等.直升机着舰引导与控制研究进展[J].南京航空航天大学学报,2018,50(6): 745-753.
LIAO Zhilin, LU Xiaolan, HUANG Yimin, et al. Research progress of landing guidance and control for carrier-based helicopter [J]. Journal of Nanjing University of Aeronautics & Astronautics, 2018, 50 (6): 745-753.
- [9] WU X, SAHOO D, HOI S C H. Recent advances in deep learning for object detection [EB/OL]. (2019-08-10) [2019-10-20]. <https://arxiv.org/abs/1908.03673?context=cs.CV>.
- [10] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multi-box detector [EB/OL]. (2015-12-08) [2019-10-20]. <https://arxiv.org/abs/1512.02325>.
- [11] REDMON J, FARHADI A. YOLOV3: An incremental improvement [EB/OL]. (2018-04-08) [2019-10-20]. <https://arxiv.org/abs/1804.02767>.

- [12] LIU L, OUYANG W, WANG X, et al. Deep learning for generic object detection: A survey[EB/OL]. (2018-09-06) [2019-10-20]. <https://arxiv.org/abs/1809.02165>.
- [13] BOSQUET B, MUCIENTES M, BREA V M. STDNET: A convnet for small target detection[C]//BMVC. [S.l.]: [s.n.], 2018: 1-12.
- [14] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. (2014-09-04) [2019-10-20]. <https://arxiv.org/abs/1409.1556>.
- [15] SHRIVASTAVA A, GUPTA A, GIRSHICK R. Training region-based object detectors with online hard example mining [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. [S.l.]: IEEE, 2016: 761-769.
- [16] LENG Xuefei, GONG Zhe, FU Ruizhe, et al. A sequence image matching method based on improved high-dimensional combined features[J]. Transactions of Nanjing University of Aeronautics and Astronautics, 2018, 35(5): 820-828.
- [17] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. [S.l.]: IEEE, 2016: 770-778.
- [18] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE international conference on computer vision. [S.l.]: IEEE, 2015: 1440-1448.
- [19] BOSQUET B, MUCIENTES M, BREA V M. STDnet: A convNet for small target detection[C]//BMVC. [S.l.]:[s.n.], 2018: 253.
- [20] ZHANG Zhouyu, CAO Yunfeng, ZHONG Peiyi, et al. An edge-boxes-based intruder detection algorithm for UAV sense and avoid system[J]. Transactions of Nanjing University of Aeronautics and Astronautics, 2019, 36(2): 253-263.

(编辑:张蓓)