

基于改进 Faster R-CNN 的空中目标检测

冯小雨, 梅卫, 胡大帅

陆军工程大学石家庄校区电子与光学工程系, 河北 石家庄 050003

摘要 相比传统图像目标检测算法, 基于大数据和深度学习的检测算法无须人工设计特征, 且检测性能更稳健。在防空应用背景下, 自建了空中目标静态和视频图像数据集用于训练和测试, 改进了基于深度学习的目标检测框架 Faster R-CNN, 将其专用于空中目标检测。结合空中目标检测任务的特点和需求, 提出膨胀积累、区域放大、局部标注、自适应阈值、时空上下文等改进策略, 弥补了 Faster R-CNN 对弱小目标和被遮挡目标不敏感的缺陷, 提高了检测速度和精度。实验表明, 改进后的 Faster R-CNN 在应对弱小目标、多目标、杂乱背景、光照变化、模糊、大面积遮挡等检测难度较大的情况时, 均能获得很好的效果。数据集上测试结果的平均准确率均值较改进之前提高了 16.7%, 检测速度提高了 3 倍。

关键词 机器视觉; 深度学习; 卷积神经网络; 空中目标检测

中图分类号 TP391.41

文献标识码 A

doi: 10.3788/AOS201838.0615004

Aerial Target Detection Based on Improved Faster R-CNN

Feng Xiaoyu, Mei Wei, Hu Dashuai

Department of Electronics and Optics Engineering, Army Engineering University, Shijiazhuang, Hebei 050003, China

Abstract Compared with the traditional detectors, the detectors based on large data and deep learning do not require manually designed features and are more robust. Under the background of air defense, we build the images and videos dataset of aerial target for training and test, improve the deep learning-based detector Faster R-CNN, and specialize it in aerial target detection. Aiming at the peculiarities and requirements of aerial target detection, we propose the strategies such as accumulation of dilation, regional amplification, local tagging, adaptive threshold and spatio-temporal context to make up the shortage of Faster R-CNN that small weak or occluded targets can not be detected and improve the detection speed and accuracy. Experimental results show that the improved Faster R-CNN performs well under circumstances such as small weak or multiple targets, clutter, illumination changes, blur and large-area occlusion. Compared to the original Faster R-CNN, the mean average precision is improved by 16.7% on the built dataset, and the speed is 3 times faster.

Key words machine vision; deep learning; convolutional neural network; aerial target detection

OCIS codes 150.0155; 150.1135; 100.4996

1 引 言

在防空武器光电跟踪过程中, 空中目标检测的精度和速度直接影响跟踪效果。相比地面目标, 空中目标所处空域背景较为简单平缓, 但成像较小, 且易受光照、云雾影响。近处的一些目标(如战斗机)快速灵活机动, 难以捕捉, 镜头晃动容易造成画面模糊失真, 而且出于跟踪要求, 检测速度需满足实时性。

非深度学习的传统检测算法主要有帧差法^[1]、光流法^[2]、特征匹配法^[3]、背景建模法^[4]、阈值分割法^[5]几类。帧差法快捷且主动性好, 但只能检测静态背景中的移动物体; 光流法和特征匹配法检测精度和准确度高, 但需要手动初始化目标, 缺乏自主性且计算量大; 背景建模法有利于目标和背景的自动分割, 但模型的建立和更新较为耗时, 动态背景容易对结果产生干扰; 阈值分割法在背景简单而目标突

收稿日期: 2017-12-06; **收到修改稿日期:** 2018-01-22

基金项目: 国防科技预研项目(40405070102)

作者简介: 冯小雨(1993—), 男, 硕士研究生, 主要从事深度学习与计算机视觉方面的研究。E-mail: 826782445@qq.com

导师简介: 梅卫(1971—), 男, 博士, 副教授, 硕士生导师, 主要从事目标跟踪与人工智能方面的研究。

E-mail: meiwei@sina.com

出的情况下便捷高效,但在复杂背景下的效果并不理想。

传统检测算法存在着各种局限性,无法满足复杂的现实应用场景下的需求,而且无法对外观各异的目标进行细致分类,人工干涉较多,缺乏实用性。基于深度学习^[6]的检测算法可用于各种检测场景,建模灵活方便,综合性强,能够同时检测和识别多类目标,主动性好,目前已应用于车辆和行人的监控、识别等任务中,但尚缺乏专用于空中目标检测任务的框架。通过理论分析和实验验证可知,基于深度学习的空中目标检测算法不受动静态场景的使用限制,不易被杂乱背景、光照变化、目标外观等因素影响,对多目标、模糊目标、轻微遮挡或较小的目标均有良好的检测效果。深度学习可以直观理解为深层人工神经网络的学习方法,因具有强大的非线性表达能力、灵活可靠的数据驱动建模方式,以及加速运算平台的支撑,取得了许多重大突破^[7]。各种类型的人工神经网络结构中,卷积神经网络因其参数共享、局部连接和下采样结构的仿视觉处理方式,在图像识别、图像分割、目标检测、场景分类等视觉任务中,效果远超传统算法^[8]。甚快区域卷积神经网络(Faster R-CNN)^[9]是2015年微软研究院提出的基于区域推荐和卷积神经网络、可以进行端到端学习的目标检测框架。以快速检测著称的单次多框检测器(SSD)^[10]、YOLOv2^[11]等框架也借鉴了Faster R-CNN的诸多思想。

深度卷积网络基于图像的颜色、边缘等信息来检测目标,因此对于弱小目标和大面积遮挡目标等缺乏图像信息的目标难以发挥作用,而且Faster R-CNN的检测速度仍无法满足实时性。针对空中目

标检测任务的特点和需求,对Faster R-CNN的不足之处提出改进,主要包括4个方面:1)建立专用于空中目标检测的数据集,确保数据充足,具有代表性,利于提高模型质量;2)针对Faster R-CNN对弱小目标检测效果差的问题,提出膨胀积累和区域放大策略,降低漏警率;3)针对Faster R-CNN对大面积遮挡目标检测效果差的问题,提出局部标注和时空上下文策略,降低遮挡对检测的影响;4)针对Faster R-CNN固定检测阈值不够灵活的缺陷,提出自适应阈值策略,降低漏警率和虚警率。在4点改进之上,通过压缩Faster R-CNN候选区域的数量,在确保检测准确度的同时提高了速度。

2 数据集建立

相同结构和训练参数的深度学习模型,其质量关键在于数据集。PASCAL VOC数据集内容丰富,格式规范,图像和标注质量高,常被用作评测数据集。针对空中目标检测任务,按照VOC数据集格式标准构建空中目标检测数据集。数据集中每张图对应一个标签,标签标明了图片名称、图片中目标的类别,以及目标外接矩形的坐标和宽高。数据集包含38段约2.5万帧的彩色视频图像,以及约1500幅彩色静态图像,分为战斗机(军用固定翼飞机)、直升机(旋翼飞机)、客机(大型客机和小型私人飞机)、鸟类(空中飞行的鸟)、其他目标(导弹以及难以分辨的弱小目标)5大类目标,具体构成如表1所示。需要说明的是,在其他目标这一类别中,由于弱小目标并非背景又无法辨别,故单独成类。导弹纵截面细长而横截面很小,远距离正面拍摄难以发现和识别,故也归为弱小目标。

表1 空中目标数据集

Table 1 Dataset of aerial targets

Content	Fighter	Helicopter	Aerobus	Bird	Others
Video	11 sequences (about 7000 frames)	7 sequences (about 6000 frames)	10 sequences (about 9000 frames)	4 sequences (about 1000 frames)	6 sequences (about 4000 frames)
Image	300	300	300	300	300

数据集中大多目标都处于飞行状态,背景除天空外,也包括了地面和海面的场景(如飞行表演),在增加检测难度的同时,也对算法实用性提出了更高的要求。视频画质有高有低,分辨率范围从240 pixel×180 pixel到960 pixel×540 pixel,拍摄视角多为对空视角,持续时间从几秒到几分钟不等,既有晃动也有失焦,既有远景渐入也有近景抓拍,既有背景造成的被动遮挡也有目标间主动遮挡,全面

模拟了实战瞄准过程中可能出现的各种情况。

本文设计的数据集有两个特点。1)通常目标检测数据集多为静态图像,而本文以视频图像为主、静态图像为辅。这是因为视频图像包含时序信息,能够对目标的姿态、大小、位置进行连续采样,利于训练出结果更稳定的检测模型;另外,视频图像的场景单一,而静态图像场景丰富,便于为训练增加数据的多样性,利于训练出适应性更强的模型。2)通常目

标检测数据集用于表示物体位置的矩形边缘框只有一个。本文为了检测被大面积局部遮挡的目标,借鉴可变形部件模型^[12]的思想,提出了局部标注策略,即将目标的一些局部用矩形框标注起来(被标注的部位主要集中在各类飞机的机头、机翼、机尾)。由于目标运动时大面积局部遮挡一般时间持续很短,所以局部标注使用不宜过多,否则对没有遮挡的正常目标会出现局部检测得分较高、而整体目标由于检测得分较低被非极大抑制而排除的情况。本数据集中只有5%的图像被局部标注过,这一比例在没有出现上述副作用的情况下,很好地检测到了被大面积遮挡目标的剩余部分。

3 Faster R-CNN 原理及改进

3.1 Faster R-CNN 基本原理及其流程

Faster R-CNN 是 R-CNN 系列检测框架的最终版。R-CNN^[13]是第一个基于候选区域的卷积神经网络目标检测框架,经历了空间金字塔池化网络(SPP-Net)^[14]、Fast R-CNN^[15],以及 Faster R-CNN 后续三个版本的改进,检测速度和精度得到大幅度提升。这些检测框架的共同点是都分为这三个基本环节进行:首先进行候选区域提取(从待检测的图片中生成可能为目标区域),然后将候选区域送到训练好的卷积网络模型中提取特征,最后利用特征对候选区域进行分类和边缘框回归(即位置精调)。Faster R-CNN 在 Fast R-CNN 的基础上提出一种可以生成候选区域的区域推荐网络(RPN),并将二者整合为一个可以端到端学习的完整网络,既保证了精度又提高了速度。

如图1所示,将 Faster R-CNN 分为 RPN 候选区域提取和 Fast R-CNN 对候选区域进行分类和边缘框回归两阶段。步骤 a(卷积网络提取特征)为两阶段共用,步骤 b(anchor 机制生成大量候选区域)、c(Softmax 层将候选区域分为目标和背景两类)、d(全连接层映射得到候选区域位置)属于 RPN 阶

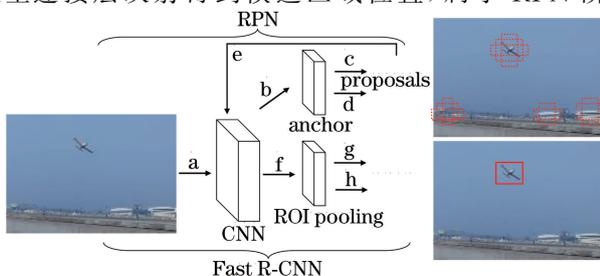


图1 Faster R-CNN 框架

Fig. 1 Frame of Faster R-CNN

段,步骤 e(最后一层卷积特征图上找到候选区域的映射)将 RPN 结果传递给 Fast R-CNN,步骤 f(感兴趣区域(ROI)池化层将不同尺寸候选区域的特征图映射规范到同一大小)、g(Softmax 层利用 RoI 池化特征得到候选区域属于各类别的概率)、h(通过全连接层实现边缘框回归,对候选区域位置进行精修)为 Fast R-CNN 检测阶段。详细原理和技术细节参照文献^[4]。

3.2 针对弱小目标的两点改进

目标从远处进入画面时,面积小,信号能量弱,被称作弱小目标,最初只包含少量像素点,很难与背景和噪声区分开,而且缺乏形状、纹理等信息,须等到目标像素增加到一定数目才能辨别。在白天,深色、不发光物体在远处常呈现为暗点,浅色、发光或反光物体则呈现为亮点,与背景中天空和云的亮度存在差异。本文基于亮度差异分割出可能属于目标的区域,再利用形态学运算去除噪声,最后将候选区域插值放大。

单一的二值化阈值无法有效分割不同明暗状态下的背景和背景,对大量不同光照条件下的图像进行二值化处理,针对亮目标和暗目标的分割问题,分别选取了 180 和 120 两个灰度阈值,可适用于大多数日间光照情况。分别对灰度图像进行二值化处理后,为综合两个阈值下的分割结果,将高阈值下的二值化图像取反后与低阈值下的二值化结果相加。之后对相加结果进行多次膨胀处理,使得原本只有几个像素的目标被较大的零值区域覆盖,在完整提取目标的同时消去背景中孤立的零值区域,以达到突出目标、区分平缓的空域背景和杂乱的地面景物的效果。图2展示了低阈值下的膨胀积累操作,图2(a)包含了一个弱小目标,对图2(a)进行阈值为120的图像二值化得到图2(b),之后进行两次全方位膨胀处理,分别如图2(c)和(d)所示。由图2(d)可知,覆盖目标的零值区域在保持孤立的情况下面积增加,且地面的孤立点基本消失,较好地实现了图像分割。

图像经过膨胀积累后得到可能包含目标和背景的多个区域块,计算各区域的面积,保留不大于 $10 \text{ pixel} \times 10 \text{ pixel}$ 面积的区域,用这些区域的外接矩形表示候选目标的位置,将对应的原图部分送到 Fast R-CNN 分类。深度卷积网络在对逐层提取图像特征的同时,各层生成的特征图尺寸也越来越小。图3(a)经过卷积和池化操作后得到图3(b),通过放大包含目标的部分可见,图像经过卷积和池化操作会损失细

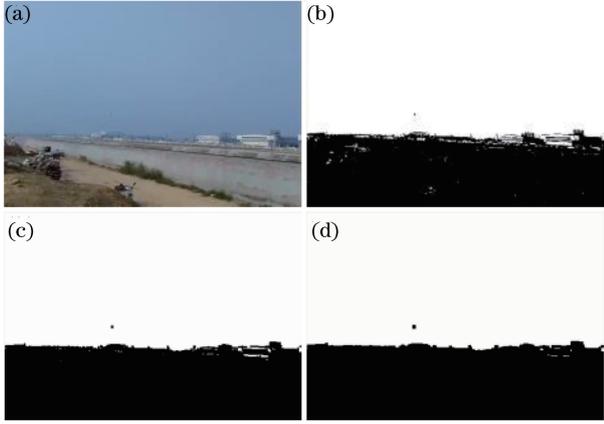


图 2 膨胀积累

Fig. 2 Accumulation of dilation



图 3 特征图的池化

Fig. 3 Pooling of feature map

节信息。对本身携带信息很少的弱小目标来说,很容易在深度卷积网络对图像的层层抽象中丢失。

为避免细节丢失,采用区域放大策略,即在把膨胀积累阶段所得候选区域的原图部分送到 Fast R-CNN 分类之前,先进行双线性插值,放大至 Fast R-CNN 候选区域的最小尺寸(本文设置的最小候选区域为 $10 \text{ pixel} \times 10 \text{ pixel}$)。这一策略简单有效地在卷积池化过程中保持了弱小目标的图像细节,利于将弱小目标和背景区分开,并尽早确定目标类别,其过程如图 4 所示。

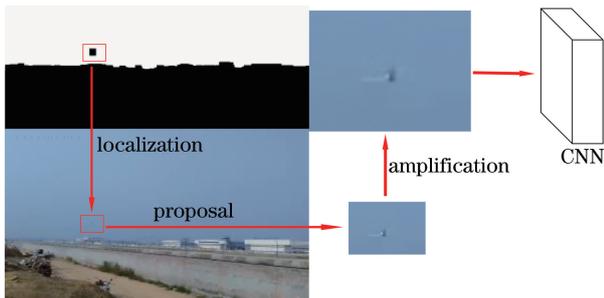


图 4 区域放大

Fig. 4 Regional amplification

将膨胀积累和区域放大两个策略结合使用,代替 RPN 提取候选区域,有效降低了 Faster R-CNN 弱小

目标检测的漏警率,提高了对弱小目标的识别能力。

3.3 针对大面积遮挡的改进

训练有素的卷积神经网络可以应对一定程度的遮挡,但无法应对大面积遮挡造成的目标图像信息严重缺失。对此,提出时空上下文策略,即之前的检测结果中获取有用的先验信息,合理预测少量候选区域,增加目标被检测的几率。这一思路借鉴了目标跟踪的方法来优化检测结果。卡尔曼滤波^[16]和粒子滤波^[17]常被用于跟踪算法中,选择卡尔曼滤波作为在上一帧和当前帧之间传递目标信息的工具,结合目标检测任务设计卡尔曼滤波模型。

系统的状态转移方程为

$$\mathbf{X}_{k+1} = \mathbf{A}\mathbf{X}_k + \mathbf{w}_k. \quad (1)$$

系统的测量方程为

$$\mathbf{Z}_k = \mathbf{H}\mathbf{X}_k + \mathbf{v}_k. \quad (2)$$

系统的滤波方程为

$$\hat{\mathbf{X}}_k = \mathbf{A}_{k-1}\hat{\mathbf{X}}_{k-1} + \mathbf{K}_k(\mathbf{Z}_k - \mathbf{H}_k\mathbf{A}_{k-1}\hat{\mathbf{X}}_{k-1}). \quad (3)$$

卡尔曼增益方程为

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}^T(\mathbf{H}\mathbf{P}_{k|k-1}\mathbf{H}^T + \mathbf{V}_k)^{-1}. \quad (4)$$

预测误差协方差矩阵方程为

$$\mathbf{P}_{k|k-1} = \mathbf{A}\mathbf{P}_{k-1}\mathbf{A}^T + \mathbf{W}_{k-1}. \quad (5)$$

修正误差协方差矩阵方程为

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k\mathbf{H})\mathbf{P}_{k|k-1}, \quad (6)$$

式中 \mathbf{I} 为 10×10 的单位矩阵,状态向量 $\mathbf{X}_k = (x, \Delta x, y, \Delta y, a, \Delta a, b, \Delta b, c, \Delta c)$ 。状态向量中 x, y, a, b 分别为第 k 帧某一目标 T 外接矩形框的左上角坐标和宽高, c 为目标置信度, $\Delta x, \Delta y, \Delta a, \Delta b, \Delta c$ 为目标 T 在第 k 帧与第 $k-1$ 帧中坐标、宽高和置信度的变化值。

状态转移矩阵为

$$\mathbf{A} = \begin{bmatrix} \mathbf{L} & 0 & 0 & 0 & 0 \\ 0 & \mathbf{L} & 0 & 0 & 0 \\ 0 & 0 & \mathbf{L} & 0 & 0 \\ 0 & 0 & 0 & \mathbf{L} & 0 \\ 0 & 0 & 0 & 0 & \mathbf{L} \end{bmatrix}, \quad (8)$$

式中 $\mathbf{L} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ 。

观测矩阵为

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}. \quad (9)$$

状态噪声 \mathbf{w}_k 和观测噪声 \mathbf{v}_k 均为高斯白噪声,

二者的协方差矩阵分别为

$$\mathbf{W} = \begin{bmatrix} \mathbf{E} & 0 & 0 & 0 & 0 \\ 0 & \mathbf{E} & 0 & 0 & 0 \\ 0 & 0 & \mathbf{E} & 0 & 0 \\ 0 & 0 & 0 & \mathbf{E} & 0 \\ 0 & 0 & 0 & 0 & \lambda \mathbf{E} \end{bmatrix}, \quad (10)$$

$$\text{式中 } \mathbf{E} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \lambda = 0.2, \mathbf{V} = \begin{bmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0.1 \end{bmatrix}.$$

$\mathbf{P}_{k|k-1}$ 和 \mathbf{X}_k 的初始化值分别为 $\mathbf{P}_{k=1} = \mathbf{W}$ 和 $\mathbf{X}_1 = \hat{\mathbf{X}}_1$, $\hat{\mathbf{X}}_1$ 为目标 T 出现的第一帧检测结果的状态向量, 作为第一帧的修正值传递给第二帧进行滤波, 其中 5 个变化值初始化为 0。从目标 T 出现第二帧开始, 取当前帧的预测值 $\mathbf{A}\hat{\mathbf{X}}_{k-1}$ 和修正值 $\hat{\mathbf{X}}_k$ 中的 (x, y, a, b, c) 作为该帧图像的两个候选区域, 连同 RPN 提取的候选区域一并提取 RoI 池化特征, 此时应将 c 作为该区域为前景目标的概率。该帧检测结束后, 将结果作为该帧滤波值送入下一帧滤波。当出现多个目标时则分别进行滤波, 目标个数增加时增加相应个数的滤波器。此外, 设定当目标连续 10 帧滤波值对应的候选区域没有作为检测结果时, 取消该滤波器。

需要说明的是, 卡尔曼滤波存在 3 个前提假设: 1) 被建模的系统是线性的; 2) 影响测量的噪声属于白噪声; 3) 噪声本质上是高斯分布的。很显然, 由于摄像机的运动和神经网络本身复杂的非线性映射, 目标在视频中的位置和置信度并非线性变化的。但本文只是将滤波作为提高候选区域质量的辅助手段, 而且在短时间内目标可以近似看成线性运动。实验表明, 时空上下文降低了检测结果的噪声, 提高了前后帧间的连贯性。

3.4 针对置信度固定阈值的改进

在 Faster R-CNN 用 Softmax 为候选区域进行分类的最后阶段, 候选区域会得到属于各个类别的置信度(即属于各个类别的概率), 当属于某类的置信度高于设定阈值时则将此候选区域判为该类目标。若同一候选区域有多个类别置信度高于阈值, 取最高者。

分别比较图 5(a) 和 (b), 图 5(c) 和 (d) 可以发现, 目标尺度较小或被遮挡时置信度相对较低。若采用固定阈值, 设置过高会排除许多真目标, 过低会混入许多假目标。通常的做法是不断调整阈值, 对

数据集进行多次测试, 计算出不同阈值下的平均准确率, 取平均准确率最大的阈值作为模型的最终阈值。但这种做法有适应数据集的倾向, 再庞大的数据集也无法涵盖现实中的所有情况。本文采用自适应阈值, 在避免适应数据集的同时提高模型的决策能力。

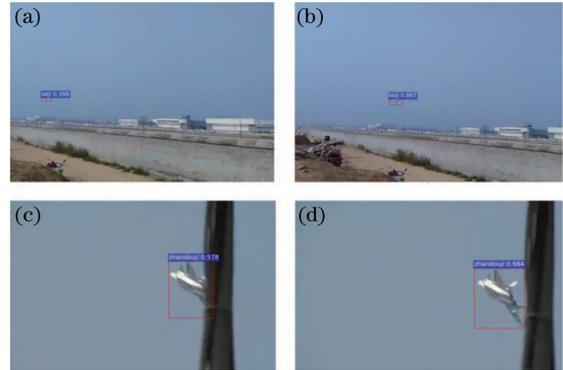


图 5 尺度和遮挡对置信度的影响

Fig. 5 Influences of scale and occlusion on confidences

无论固定或自适应, 阈值的设定都需要参考数据集中目标的得分情况。检测模型训练较好的情况下, 正确的检测结果中真目标和假目标置信度常常相差一两个数量级, 且真目标置信度通常在 0.1 以上。虽然与真目标的置信度存在差距, 但假目标也会因为某些特征与目标类似而取得 0.1 以上的高置信度(如图 6 右边的树叶), 单纯采用固定阈值无法将目标与背景区分开。注意到真假目标之间有着明显的差距, 可以通过各个候选区域得分差距设置自适应阈值, 以有效区分低置信度真目标和高置信度的假目标, 降低虚警率。



图 6 高置信度假目标

Fig. 6 False positive with high confidence

二阶差分可以表现离散数组的变化趋势大小, 可用于在一组置信度中确定阈值。检测一张图像 RPN 默认得到 300 个候选区域送入 Fast R-CNN, 最后每个候选区域都得到 5 个用来表示属于 5 个类别的置信度, 故共可以得到的 300 个 5×1 的数组。

比例为 4:1)。其中,训练集局部标注的比例约 5%,测试集不使用局部标注。在测试集 1 中选出 100 张包含小目标(本文设定小目标尺寸小于 10 pixel×10 pixel)的图像构成测试集 2。

采用训练集训练原始 Faster R-CNN,将此模型记为 M0。在 M0 基础上加入膨胀积累和区域放大策略,生成模型 M1。在 M0 基础上加入时空上下文策略,生成模型 M2。在 M0 基础上加入自适应阈值策略,生成模型 M3。最后将 M0 与所有策略结合在一起,生成模型 M4。使用测试集 1 对 M0、M2、M3、M4 进行测试和对比。为突出小目标检测效果,使用测试集 2 分别对 M0 和 M1 进行测试和对比。

此外,分别选取背景建模法中性能稳健的混合高斯模型(GMM)^[19]、光流法中的均衡速度和准确性的卢卡斯光流(KLT)法^[20],以及阈值分割法中的最大类间方差(OTSU)法^[21]作为传统对比算法,与 M4 对比测试集 1 上的检测效果。另选取 SSD 系列和 YOLO 系列检测框架作为深度学习对比算法,与 M4 对比 VOC 数据集飞机类别子集上的检测效果。

4.3 实验结果分析

实验结果如表 2、3 所示,分别对比模型在测试集 1 和测试集 2 上的 AP 和 mAP 值。

表 2 测试集 1 结果

Model	Fighter	Helicopter	Aerobus	Bird	Others	mAP
M0	90.76	91.04	92.35	63.71	72.76	82.12
M2	93.42	96.67	94.22	75.81	95.76	91.18
M3	91.53	92.64	93.19	69.37	78.07	84.96
M4	98.72	99.37	97.42	98.61	99.76	98.78

表 2 中 M0 和 M4 结果对比,各类的 AP 提高了 5%~35%,mAP 提高了约 16.7%,表明了本文策略总体对于弥补 Faster R-CNN 缺陷的有效性。

表 2 中 M0 和 M2 结果对比,各个类别的 AP 值提升 2%~23%,mAP 提高了约 9%,表明时空上下文策略弥补了卷积神经网络无法提取时序信息的缺陷,降低了检测结果的噪声,提高了前后帧间的连贯性。M0 和 M3 结果对比,各个类别的 AP 值提升 1%~6%,其中鸟类和其他目标两个类别中尺寸相对较小的目标较多,效果较为明显,表明自适应阈值发挥了区分低置信度真目标和高置信度假目标的作用。

表 3 中 M0 和 M1 结果对比说明了膨胀积累和区域放大对小目标检测的有效性。由于弱小目标类别难以确定,M1 的错误检测多为分类错误造成的,而 M0 没有检测出目标,表明深度卷积网络逐层抽

取特征的同时导致弱小目标信息被忽略。

表 3 测试集 2 结果

Model	Fighter	Helicopter	Aerobus	Bird	Others	mAP
M0	0	0	0	0	0	0
M1	54.05	68.66	62.26	39.61	85.36	61.99

由于 GMM、KLT 和 OTSU 算法不具有分类功能,无法计算 AP 值,故通过对比测试集 1 中有代表性的检测结果以直观表明各算法的检测性能。三种对比算法中使用限制较少,只有 KLT 算法需要人工初始化目标,通过优化参数后,三种算法在简单情况下对空中目标的检测效果都很好,故只展示特殊情况下的检测结果。

图 8 展示了 4 帧检测结果:图 8(a)到图 8(b)出现了镜头晃动,目标产生失焦模糊,图 8(b)到图 8(c)环境产生了光照变化,图 8(d)为目标在远处着陆时与地面背景相连。由图 8(a)到图 8(b)可知,M4 具备检测模糊目标的能力,GMM 和 OTSU 无法分割出目标与背景灰度接近的部分,而 KLT 无法找到可以匹配的特征点,导致检测失败;由图 8(b)~(c)可知,KLT 和 M4 的检测能力不受光照变化影响,而 GMM 和 OTSU 则出现大量干扰噪声,不利于之后将目标从背景中分割出;由图 8(d)可知,只有 M4 的检测能力不受杂乱背景影响。

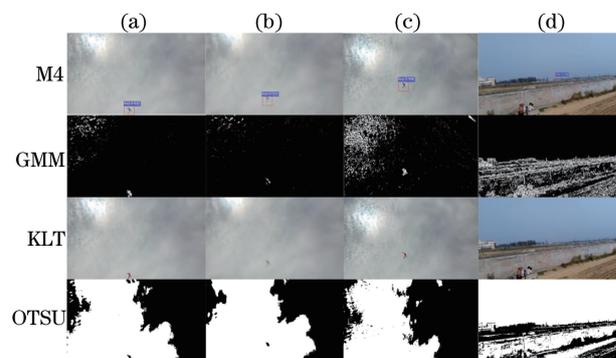


图 8 检测结果示例

Fig. 8 Examples of detection results

GMM 在画面小范围晃动时目标周围会产生阴影;KLT 法无法适应目标明显的姿态和尺度变化,对新出现的目标依然需要人工干预;OTSU 法只基于当前帧的灰度信息,受背景影响大,缺乏时域信息。当出现遮挡时,由于各算法没有预测能力,目标信息缺失,导致检测到的目标完整性降低。图 9 所示为 M4 在其他检测难度较大的画面上的检测结果,包含了大面积遮挡、目标弱小、复杂背景、光照影响、运动模糊等多种影响检测效果的因素,直观说明

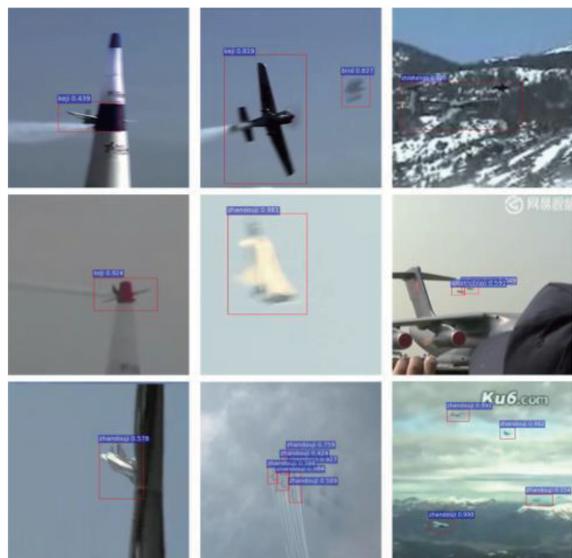


图 9 M4 检测结果示例

Fig. 9 Examples of detection results of M4

了本文改进算法的优越性。传统的检测算法由于无法兼顾各种情况而导致检测失败,原始FasterR-

表 4 VOC 数据集飞机类别检测性能

Table 4 Detection performances on class aero of VOC

Detector	Faster R-CNN ^[9]	SSD299 ^[10]	SSD443 ^[10]	YOLO ^[11]	YOLOv2 544 ^[11]	M4
AP /%	84.9	71.2	78.6	77.0	86.3	90.7%
Speed / (frame · s ⁻¹)	8.3	46	19	45	40	25

5 结 论

针对 Faster R-CNN 在弱小目标检测、大面积遮挡等方面的缺陷,结合空中目标检测任务的特点和需求,提出膨胀积累、区域放大、局部标注、自适应阈值、时空上下文等辅助策略加以改进。在自建的空中目标图像和视频数据集上进行训练和测试,结果表明,相比传统的目标检测算法和原始的 Faster R-CNN,本文算法不仅提高了 Faster R-CNN 的检测速度和精度,而且在弱小目标、多目标、大面积遮挡、模糊、光照变化、复杂背景等各种情况下均能获得更好的检测效果,稳健性远超其他空中目标检测算法,为深度学习在特定目标检测的应用提供了实例和新的思路。

由于卷积神经网络的结构不适合处理时序信息,未来可尝试结合递归神经网络^[22](一类具有记忆功能的神经网络)来解决视频目标检测和跟踪问题。此外,由于神经网络常常被当成黑箱处理,对网络模型本身的理解和优化,不仅有利于网络参数的选取和网络结构的设计^[23],而且可以压缩模型,去除冗余信息,精简提取的特征,提高算法效率,为深

CNN 对弱小和大面积遮挡目标检测效果也不好,而改进算法结合了深度学习和传统算法的一些思想,优势互补,实现了对空中目标检测效果的优化。

VOC 数据集是深度学习目标检测的公认测评数据集,其 20 类目标中包括飞机类别,本文使用 VOC2007 的测试集来测试模型 M4 和其他深度学习算法。由表 4 可知,模型 M4 对 VOC 飞机类别目标检测 AP 值最高,单个类别超越 YOLO 改进版 YOLOv2,表明本文改进发挥了作用,但由于 VOC 为静态图像数据集,时空上下文策略无法生效,改进效果不如在自建数据集上的效果明显。此外,候选区域的数量是影响 Faster R-CNN 耗时的重要因素。由于本文策略计算量小,耗时小,在提高候选区域质量的前提下压缩其数量,在确保准确性的同时大大减少了耗时。原始 Faster R-CNN 的检测速率为 8.3 frame · s⁻¹,M4 为 25 frame · s⁻¹,将原始模型的速度提高了 3 倍,虽仍不及 SSD 和 YOLO,但满足了实时性的基本要求。

度学习硬件的实现打下基础。

参 考 文 献

- [1] Wang Y X, Shi D L, Yang N. Fusion algorithm of frame difference and mean shift for high-speed UAVs' tracking [J]. Journal of Naval Aeronautical and Astronautical University, 2016, 31(4): 437-441. 王宜贤, 石德乐, 杨宁. 帧差法和 Mean shift 算法融合的高速无人机目标跟踪 [J]. 海军航空工程学院学报, 2016, 31(4): 437-441.
- [2] Chu J. Detection of moving target in dynamic background based on optical flow [J]. Journal of Nanchang University of Aeronautics and Astronautics, 2011, 25(3): 1-6. 储珺. 基于光流的动态背景运动目标检测算法 [J]. 南昌航空大学学报, 2011, 25(3): 1-6.
- [3] Li S S, Zhao G P, Wang J Y. Distractor-aware object tracking based on multi-feature [J]. Acta Optica Sinica, 2017, 37(5): 0515005. 李双双, 赵高鹏, 王建宇. 基于特征融合和尺度自适应的干扰感知目标跟踪 [J]. 光学学报, 2017, 37(5): 0515005.
- [4] Tan Y, Wang Y J. On a combination of background difference improved CamShift target tracking method

- [J]. Journal of Southwest China Normal University (Natural Science Edition), 2016, 41(9): 120-125.
谭艳, 王宇俊. 一种结合背景差分的改进 CamShift 目标跟踪方法[J]. 西南师范大学学报(自然科学版), 2016, 41(9): 120-125.
- [5] Yin L, Gao K, Bai T Z. Research on oilcan target segmentation in remote sensing image based on improved Otsu algorithm [J]. Optical Technique, 2012, 38(2): 197-201.
殷亮, 高昆, 白廷柱. 基于改进 Otsu 法的遥感图像油罐目标分割研究[J]. 光学技术, 2012, 38(2): 197-201.
- [6] Yu K, Jia L, Chen Y Q, *et al.* Deep learning: yesterday, today, and tomorrow[J]. Journal of Computer Research and Development, 2013, 50(9): 1799-1804.
于凯, 贾磊, 陈宇强, 等. 深度学习的昨天, 今天和明天[J]. 计算机研究与发展, 2013, 50(9): 1799-1804.
- [7] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527-1554.
- [8] Lecun Y, Bengio Y, Hinton G. Deep learning [J]. Nature, 2015, 521(7553): 436-444.
- [9] Ren S Q, He K M, Girshick R, *et al.* Faster R-CNN: towards real-time object detection with region proposal networks [C] // Proceedings of Advances in Neural Information Processing Systems, 2015: 91-99.
- [10] Liu W, Anguelov D, Erhan D, *et al.* Ssd: single shot multibox detector [C]. European Conference on Computer Vision, 2016: 21-37.
- [11] Redmon J, Farhadi A. YOLO9000: better, faster, stronger [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2017: 6517-6525.
- [12] Hare S, Golodetz S, Saffari A, *et al.* Struck: structured output tracking with kernels [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(10): 2096-2109.
- [13] Girshick R, Donahue J, Darrell T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation [C] // Proceedings of 2014 IEEE International Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [14] He K M, Zhang X Y, Ren S Q, *et al.* Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [15] Girshick R. Fast R-CNN [C] // Proceedings of 2015 IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [16] Qu S R, Yang H H. Multi-target detection and tracking of video sequence based on Kalman_BP neural network [J]. Infrared and Laser Engineering, 2013, 42(9): 2553-2560.
曲仕茹, 杨红红. 采用 Kalman_BP 神经网络的视频序列多目标检测与跟踪 [J]. 红外与激光工程, 2013, 42(9): 2553-2560.
- [17] Li S J, Zhu Z F. Track-before-detect algorithm via particle filtering [J]. Infrared and Laser Engineering, 2009, 38(2): 352-357.
李少军, 朱振福. 采用粒子滤波的先跟踪后检测算法 [J]. 红外与激光工程, 2009, 38(2): 352-357.
- [18] Jia Y, Shelhamer E, Donahue J, *et al.* Caffe: convolutional architecture for fast feature embedding [C] // Proceedings of the 22nd ACM International Conference on Multimedia, 2014: 675-678.
- [19] Wang D F, Wang Y D, Wang J W. Moving target detection algorithm based on improved Gaussian mixture model [J]. Laser Technology, 2014, 38(6): 776-779.
王东方, 王玉德, 王景武. 基于改进的混合高斯模型的运动目标检测方法 [J]. 激光技术, 2014, 38(6): 776-779.
- [20] Yang C C, Gu G H, Qian W X, *et al.* Hardware implementation of infrared image registration based on the Harris corner of KLT tracking [J]. Infrared Technology, 2013, 35(10): 632-637.
杨陈晨, 顾国华, 钱惟贤, 等. 基于 Harris 角点的 KLT 跟踪红外图像配准的硬件实现 [J]. 红外技术, 2013, 35(10): 632-637.
- [21] Jiang Q Y, Li P, Sun L. Application of Otsu method in motion detection system [J]. Journal of Computer Applications, 2011, 31(1): 260-262.
江亲瑜, 李平, 孙兰. 最大类间方差算法在运动检测系统中的应用 [J]. 计算机应用, 2011, 31(1): 260-262.
- [22] Cui Z, Xiao S, Feng J, *et al.* Recurrently target-attending tracking [C] // Proceedings of 2016 IEEE International Conference on Computer Vision, 2016: 1449-1458.
- [23] Cai Y Z, Yang D D, Mao N, *et al.* Visual tracking algorithm based on adaptive convolutional features [J]. Acta Optica Sinica, 2017, 37(3): 0315002.
蔡玉柱, 杨德东, 毛宁, 等. 基于自适应卷积特征的目标跟踪算法 [J]. 光学学报, 2017, 37(3): 0315002.