专栏：〈红外目标检测〉

前言：红外目标探测具有工作距离远、抗干扰能力强、测量精度高、不受天气影响、能昼夜工作等特点，在军事和民用领域得到了广泛的应用。近年来，在以深度学习技术为代表的智能化浪潮推动下，目标探测领域取得了长足的发展与进步。基于此，《红外技术》面向研究人员推出“红外目标检测专栏”，力图展示目标检测技术的最新研究成果，为从事相关研究的读者提供参考。

通过广泛征集和严格评审，本期专栏收录了来自南京工业大学、西安电子科技大学、苏州大学等从事红外目标检测团队的 8 篇论文。论文内容既有对小型无人机检测等热门研究方向的综述与分析，也有针对弱小目标检测、抗遮挡目标跟踪、三维目标识别等人工智能最新应用技术的研究。

然而，红外目标的多样性、探测环境的复杂性、应用场景的开放性等都对红外目标检测技术的发展和应用提出了更严峻的挑战。本期专栏只是一个起点，希望能够启发广大读者作出更多更精彩的研究。

最后，感谢各位审稿专家和编辑的辛勤工作。

——王卫华

基于深度卷积神经网络的小型民用无人机检测研究进展

杨 欣[1,2]，王 刚[2,3]，李 椋[2]，李邵港[1,2]，高 晋[4]，王以政[2]

（1. 南华大学，湖南 衡阳 421001；2. 军事科学院军事认知与脑科学研究所，北京 100850；
3. 北京脑科学与类脑研究中心，北京 102206；4. 中国科学院自动化研究所，北京 100190）

摘要：小型民用无人机预警探测是公共安全领域的热点问题，也是视觉目标检测领域的研究难点。采用手工特征的经典目标检测方法在语义信息的提取和表征方面存在局限性，因此基于深度卷积神经网络的目标检测方法在近年已成为业内主流技术手段。围绕基于深度卷积神经网络的小型民用无人机检测技术发展现状，本文介绍了计算机视觉目标检测领域中基于深度卷积神经网络的双阶段算法和单阶段检测算法，针对小型无人机检测任务分别总结了面向静态图像和视频数据的无人机目标检测方法，进而探讨了无人机视觉检测中亟待解决的瓶颈性问题，最后对该领域研究的未来发展趋势进行了讨论和展望。

关键词：计算机视觉；目标检测；视频目标检测；无人机检测；深度卷积神经网络；

中图分类号：TP391.4 **文献标识码**：A **文章编号**：1001-8891(2022)11-1119-13

Civil Drone Detection Based on Deep Convolutional Neural Networks: a Survey

YANG Xin[1,2]，WANG Gang[2,3]，LI Liang[2]，LI Shaogang[1,2]，GAO Jin[4]，WANG Yizheng[2]

(1. *University of South China, Hengyang* 421001, *China*; 2. *Institute of Military Cognition and Brain Sciences, Academy of Military Sciences, Beijing* 100850, *China*; 3. *Chinese Institute for Brain Research, Beijing* 102206, *China*; 4. *Institute of Automation, China Academy of Sciences, Beijing* 100190, *China*)

Abstract: Vision-based early warnings against civil drones are crucial in the field of public security and are also challenging in visual object detection. Because conventional target detection methods built on handcrafted features are limited in terms of high-level semantic feature representations, methods based on deep convolutional neural networks (DCNNs) have facilitated the main trend in target detection over the past several years. Focusing on the development of civil drone-detection technology based on DCNNs, this paper introduces the advancements in DCNN-based object detection algorithms, including two-stage and one-stage

收稿日期：2021-09-03；**修订日期**：2021-10-13.

作者简介：杨欣（1997-），女，硕士研究生，研究方向为视频目标检测。E-mail: yangxinioi@163.com。

通信作者：王刚（1988-），男，副研究员，研究方向为类脑视觉感知。E-mail: g_wang@foxmail.com。

基金项目：北京市自然科学基金（4214060）；国家自然科学基金（62102443）。

algorithms. Subsequently, existing drone-detection methods developed for still images and videos are summarized separately. In particular, motion information extraction approaches to drone detection are investigated. Furthermore, the main bottlenecks in drone detection are discussed. Finally, potentially promising solutions and future development directions in the drone-detection field are presented.

Key words: computer vision, object detection, video object detection, civil drone detection, deep convolutional neural networks

0 引言

随着无人航空技术的快速发展，小型民用无人机一方面被广泛应用于安全巡查、农业监测、抗灾救援等任务中，为人类生产和生活带来极大的便利和帮助；另一方面，无人机凭借其价低便携、易于部署、隐蔽性强等特性，也成为违禁品走私、间谍测绘、抵近侦察等违法行为的重要手段，对公共安全造成巨大威胁。因此，开发面向低空近程小型无人机的预警探测系统具有重要意义。由于小型无人机雷达反射面小、飞行高度低、运动速度慢，而且常隐藏在楼宇、山坳或树林等背景中，传统雷达探测易受地杂波干扰难以辨别目标，因而光电传感器（包括红外和可见光等频段）相比于雷达更适于复杂背景下的低空近程无人机目标探测。光电传感器获得图像视频数据后，需要进一步采用视觉目标自动识别技术输出无人机检测结果。

视觉目标检测是指在图像中发现、识别并标记特定目标的过程[1]，与物体分类、目标跟踪和图像分割技术密切相关。经典目标检测方法[2-4]通常采用滑动窗口策略，即采用一系列的滑动窗口遍历整个图像来判断图像中目标可能存在的位置，然后在图像窗口上提取一些手工设计的特征，例如尺度不变特征变换[5]，方向梯度直方图[6]和局部二值模式[7]等，再使用支持向量机（support vector machine，SVM）[8]或 AdaBoost[9]分类器对提取的特征进行分类。由于分类后仍然可能存在许多冗余窗口，还需要再使用非极大值抑制[10]技术排除冗余窗口，实现目标检测。由于经典目标检测算法采用滑动窗口策略来生成目标候选区域，窗口冗余计算量大，时间复杂度高，目标检测效率有限。同时，采用手工设计的特征来进行检测，可移植性差，难以应对目标形态和背景的变化，而且每次对新类别目标检测都要花费大量时间来设计手工特征。

为了解决经典目标检测方法存在的上述瓶颈性问题，研究人员在近年来将最初应用于物体分类的深度卷积神经网络（deep convolutional neural networks，DCNNs）引入到目标检测领域[11]，将特征学习和模式判别统一到同一模型框架下，同时借助大规模标注数据和高性能计算资源，实现了低阶图像特征和高阶语义特征的层次化表征，在多个大型公开数据集取得了可观的目标检测精度。因此，基于 DCNN 的方法已成为目标检测领域的主流手段之一[12-13]。在通用目标检测技术的基础上，业内已经提出了一些面向小型无人机的目标检测算法，在检测精度和实时性方面取得了一定的积极进展。本文对业内现有的无人机目标检测算法进行了归纳总结，探讨了现有算法在实际应用中尚存在的瓶颈性问题，并对基于 DCNN 的无人机目标检测未来发展方向进行了展望。

1 基于 DCNN 的视觉目标检测

基于深度卷积神经网络的目标检测算法[14]主要可以分为基于候选区域的双阶段算法和端到端的单阶段算法，表 1[15-43]对该类代表性算法进行了归纳。这些工作重塑了计算机视觉领域中目标检测的架构和思路，对无人机目标检测算法的开发具有重要的支撑作用和借鉴意义。

1.1 双阶段方法

深度卷积神经网络最初用于物体分类，识别图片中是否包含某个感兴趣的目标，即主要回答“what”的问题，而目标检测还需要对目标进行定位，解答“what is where”的问题。针对经典目标检测方法存在的局限性，R-CNN[15]将 DCNN 从图像分类引入目标检测，采用 DCNN 代替手工设计来自动提取和表征特征。R-CNN 首先从输入图片中选择性搜索选出约 2000 个候选区域，将每个候选区域缩放到固定大小再输入到类似 AlexNet[16]的网络模型，提取一个维度为 4096×1 的特征向量，然后分别对每个类别训练一个 SVM 分类器，判断每个候选区域是否包含某个类别的目标，进而训练回归器来修正候选区域中目标的位置，最后用训练好的模型对新输入的图片做预测。这种将目标检测分为候选区域提取和目标分类的方法一般被称为双阶段方法（如图 1 所示）。R-CNN 在 VOC2012 数据集上取得了 53.3%的按类均值平均精度（mAP），相对于之前的经典目标检测算法提升了 30%左右，展示出 DCNN 在目标检测领域的巨大潜力。然而，该算法对生成 2000 个候选区域提取特征，候选区域之间重叠多，提取特征时存在着大量的冗余

计算，影响检测速度，同时每一个候选区域提取特征前要缩放到固定尺寸，这会导致区域内目标发生几何形变，影响目标检测的性能。

针对上述问题，2015 年 He 等人提出了空间金字塔池化（Spatial Pyramid Pooling）的 SPPNet[17]模型，空间金字塔池化能够在输入任意大小的情况下产生固定大小的输出，只需一次性提取整张图片的特征，然后在特征图中找到每个候选区域对应的特征图，在每个候选区域的特征图上应用空间金字塔池化，形成这个候选区域的一个固定长度的特征向量，再用 SVM 分类器分类。该方法与 R-CNN 相比速度提升了 100 倍，但是由于 SPP 的结构阻断了梯度下降的反向传播，网络难以对卷积层参数进行有效更新，导致检测准确度降低。

此外，R-CNN 训练中需要将提取到的特征进行保存，然后为每个类训练单独的 SVM 分类器和边界框回归器，需要耗费大量的存储空间。2016 年提出的 Fast R-CNN[18]将物体分类与检测框回归在同一网络框架下训练，不需额外存储特征。Fast R-CNN 还借鉴了 SPPNet 中的空间金字塔池化层，将网络的最后一个池化层替代为 ROI pooling，用 softmax 全连接层来代替 SVM 分类器。Fast R-CNN 极大地缩短了训练时间和预测时间，基于 VGG16 的 Fast R-CNN 模型在 VOC2012 数据集上获得了 66%的 mAP 值，在训练速度上比 R-CNN 提升近 9 倍，比 SPPNet 提升近 3 倍，测试速度比 R-CNN 快大约 213 倍，比 SPPNet 快大约 10 倍。

表 1 视觉目标检测领域代表性算法归纳

Table 1 Summary of representative algorithms in the visual object detection field

	Model	Year	Backbone	Characteristics
Two-stage	R-CNN[15]	2014	AlexNet[16]	Integrate CNN classification and proposal generation; need multi-stage training; time-consuming and space-consuming.
	SPPNet[17]	2015	ZFNet[19]	Introduce the spatial pyramid pooling (SPP) into CNNs.
	Fast R-CNN[18]	2015	AlexNet、VGG16[20]	Introduce regions of interest (RoIs) pooling layer; difficult to achieve real-time detection.
	Faster R-CNN[21]	2015	ZFNet、VGG	Introducing region proposal network (RPN) to generate high-quality proposals; complex training procedures and poor real-time performance.
	ION[22]	2016	IRNN[23]	Improve performance on small object detection by employing context and multi-scale skip pooling.
	R-FCN[24]	2016	ResNet101[25]	Apply the fully convolutional neural network (FCN) to Faster R-CNN to share the computation of the entire network, improving detection speed.
	FPN[26]	2017	ResNet101	Propose a feature pyramid model to handle scale variation issues in object detection.
	Mask R-CNN[27]	2018	ResNeXt[28]、FPN	Add parallel branches to extend Faster R-CNN to achieve object segmentation, which cannot be detected in real-time.
	PANet[29]	2018	FPN	Bottom-up enhancement path and adaptive feature pooling are introduced.
	TridentNet[30]	2019	ResNet101	Elucidating the effect of receptive field on objects of different sizes in object detection tasks.
	CPNDet[31]	2020	Hourglass104[32]	Generate anchor-free proposals; two-step classification for filtering proposals.
One-stage	YOLOv1[33]	2016	GoogLeNet[34]	End-to-end real-time detection does not produce proposals but has poor detection accuracy and difficult to detect small cluster objects.
	SSD[35]	2016	VGG16	Combined with CNN and YOLOv1 model, SSD detects on multi-scale layers, which is faster and more accurate than YOLOv1.
	YOLOv2[36]	2016	DarkNet19	Propose DarkNet19 to achieve high precision and high speed, but it is still difficult to detect small objects.
	RetinaNet[37]	2018	ResNeXt101+FPN	Proposed focal loss function to solve the extreme foreground-background class imbalance problem.
	YOLOv3[38]	2018	DarkNet53	Improving performance on small objects by multi-scale detection.
	STDN[39]	2018	DenseNet169[40]	Resolve multi-scale objects by employing a scale transformation module.
	CornerNet[41]	2019	Hourglass104	Regard the object detection task as a key point detection problem， by inferencing two key points (upper left and lower right corners) as the prediction box.
	YOLOv4[42]	2020	CSPDarknet53	Faster and more accurate object detection in terms of mosaic data augmentation and self-adversarial training tips.
	DETR[43]	2020	ResNet101	Introduce transformer structure to object detection field, but the performance for small targets needs to be improved.

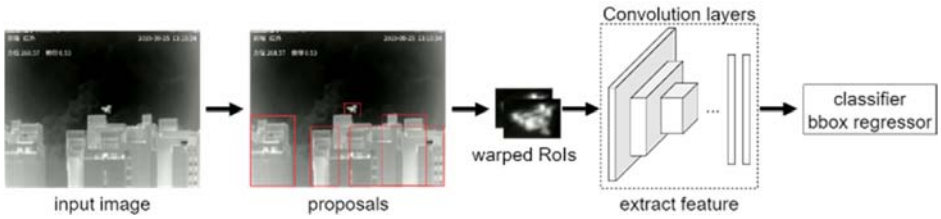

图 1 以 R-CNN 算法[15]为例的双阶段目标检测算法示意图

Fig.1 Flowchart of the two-stage object detection algorithm, taking R-CNN [15] as an example

上文介绍的 R-CNN、SPPNet 和 Fast R-CNN 都是用选择性搜索来生成候选区域，计算效率低，没有实现端到端的目标检测。针对该问题，Faster R-CNN[21]提出了区域候选网络来代替选择性搜索，而且区域候选网络与检测网络共享卷积特征，同时引入锚框（Anchor box）适应目标外形的变化，提升了检测精度和速度。

大多数目标检测算法输出的结果是目标的类别标签及其矩形外接框（bounding box），在外接框中既包括目标本身也包含局部背景。但在一些任务中需要输出像素级的检测结果，即输出实体分割结果。Mask R-CNN[27]在原有 Faster R-CNN 的基础上，在每个感兴趣区域上添加基于全卷积网络的掩模（mask）预测分支，用于判断给定像素是否属于目标，还添加了原始图像与特征图对齐的模块，进而同时得到像素级别的图像分割和目标检测结果。

1.2 单阶段方法

相比于双阶段算法，单阶段目标检测算法同时预测目标类别和位置信息，不需要显式地生成候选框（如图 2 所示），因此检测速度通常较快。

2016 年提出的 YOLO[33]（You Only Look Once）实现了端到端的模型训练和目标检测，在单阶段目标检测的发展过程中具有里程碑的意义。该模型以 GoogLeNet 为骨干网络，将输入图片分为 $s\times s$ 个网格，每个网格负责预测 B 个检测框和 C 个类别概率，相应地，每个网格输出的目标预测框包含 5 个参数，即 x，y，w，h，confidence；其中，(x, y)表示预测框中心相对当前网格的偏移量，(w, h)表示预测框相对整张图像的大小，confidence 表示预测框包含某类目标的置信度。YOLO 算法的损失函数由坐标误差、置信度误差和分类误差 3 个部分构成，通过调整坐标误差和分类误差的权重，进而提高坐标误差的比重，适当降低分类误差权重，可以防止网络过早收敛，提高网络的稳定性。YOLO 算法不需要生成一系列候选框，直接在整张图像上做回归和分类，能够大幅度提升检测速度。然而，由于该算法假定每个网格内只有 1～2 个目标，极大地限定了预测目标数量的上限，因此检测小型目标和群簇目标时极易出现漏检。

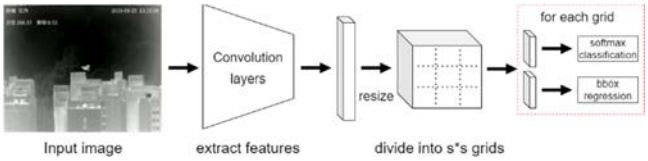

图 2 以 YOLO 算法[33]为例的单阶段目标检测算法流程示意图

Fig.2 Flowchart of the one-stage object detection algorithm, taking YOLO [33] as an example

鉴于浅层网络通常可以学习和表征图像更多的细节信息，针对多尺度目标检测任务，Liu[35]等人以 VGG16 为基础提出了 SSD（Single Shot MultiBox Detector）模型，将 VGG16 网络中的全连接层改为卷积层，并在末端增加了 4 个卷积层，同时使用 5 个层次的卷积特征图进行检测；借鉴 Faster RCNN 算法的思想，在特征图上设置不同几何尺寸的先验检测框，并直接在特征图上进行密集采样提取候选框，检测准确度和速度相比 YOLO 均有提升。但是由于浅层特征在目标表征方面存在局限性，SSD 在检测小目标时仍然存在一定困难。

2017 年提出的 YOLOv2 算法[36]采用了若干改进策略来提升初版 YOLO 算法的准确度和召回率。YOLOv2 在卷积网络中加入批归一化（Batch normalization），加快了模型收敛；通过添加 passthrough 层，将浅层特征与深层特征联系起来，改进神经网络模型对细节特征的提取和表征能力；借鉴 Fast R-CNN 方法的 anchor box 思想，用 k-means 聚类算法生成更具代表性的先验检测框；进行多尺度输入分辨率训练，使得网络在检测时能适应不同分辨率。YOLOv2 虽然解决了 YOLO 模型召回率低和定位准确性差的问题，但在小目标检测方面的改进仍然有限。

2018 年 Redmon 等人提出了 YOLOv3 算法[38]。该算法借鉴了残差网络中捷径连接架构，有效缓解了网络退化的问题；采用了类似特征金字塔的思想，面向 3 个尺度进行目标检测；通过特征图上采样和特征融合，使网络能够从早期特征映射中的上采样特征和

更细粒度的信息中获得更精细的语义信息，从而提升小目标的检测效果；通过优化卷积核尺寸提高了计算效率。在后续的 YOLOv4[42]中，作者比较不同训练技巧和算法，设计了一个能够应用于实际工作环境中的快速目标检测，而且能够在单块 GPU 上训练的模型。

2 基于 DCNN 的小型无人机视觉检测研究

2.1 无人机目标检测数据集

基于 DCNN 的目标检测算法通常需要依靠较大规模的数据集进行模型训练和性能评估。然而，当前业内仍然缺乏公开的大型无人机检测数据集。现有的无人机检测国际挑战赛数据集和公开发表文献中的自建数据集介绍如下。

2.1.1 Anti-UAV2020 数据集

Anti-UAV2020[44]数据集包含 160 段较高质量的双模态（可见光＋近红外）视频序列，其中 100 段视频用于训练和验证，60 段视频用于测试。该数据集涵盖了多种场景、多种尺度和多种机型（包括 DJI-Inspire、DJI-Phantom 4、DJI-Mavic Air、DJI-Mavic PRO）的商用无人机。该数据集中的示例图片如图 3 所示。可见光与近红外视频数据分别由固定于地面的可见光和红外光电传感器采集获得。已公开的标注数据真值由专业数据标注员给出，其中标注信息包括：检测框位置和大小、目标属性（大、中、小型目标，白天、夜晚、云雾、楼宇、虚假目标、速度骤变、悬停、遮挡、尺度变化）以及表示当前帧是否存在目标的标志位。在第二届 Anti-UAV2021[45]反无人机挑战大赛中，数据集已扩展到 280 段高清红外视频数据，涵盖多种复杂场景下无人机目标的快速运动，使无人机探测任务更具挑战性。

图 3 Anti-UAV2020 数据集示例图片（左列为可见光图像，右列为红外图像）

Fig.3 Example images in the Anti-UAV2020 dataset (The left column shows RGB images, and the right column shows infrared images)

2.1.2 Drone-vs-Bird Detection Challenge 数据集

Drone-vs-Bird Detection Challenge[46]数据集包含 11 个在不同时间拍摄的 MPEG4 格式视频，每个视频文件对应有 XML 格式的标注文件。如图 4 所示，场景中的无人机呈现出多尺度、多视角和亮度异质性。特别地，数据集中包含大量远距离的小尺寸无人机和飞鸟，很多无人机的面积小于 20 像素，有 300 多个无人机的目标标注检测框边长甚至低至 3～4 个像素，对这些微小目标的检测非常具有挑战性。

图 4 Drone-vs-Bird Detection Challenge[46]数据集示例图片

Fig.4 Example images in the Drone-vs-Bird Detection Challenge[46] Dataset

2.1.3 未开源自建数据集

除了上述公开数据集外，许多研究人员通过自建数据集来训练网络，并在其公开发表的论文中进行了相应的介绍。

文献[47]建立的 Anti-Drone Dataset 包含 449 个视频，所拍摄的无人机机型包括 Mavic pro，Phantom 2 和 Phantom 等，视频帧分辨率为 2048×1536 和 1024×768，帧速率为 24 FPS。如图 5 所示，该数据集中的视频画面涵盖了不同的相机角度、放大倍率、天气、白天或黑夜等情况，反映出无人机目标检测任务的复杂性。

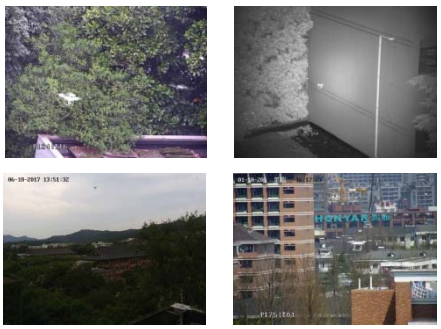

图 5 Anti-drone Dataset[47]中示例图片

Fig.5 Example images in the Anti-drone dataset[47]

UAV data[48]采集了 20 款无人机的图像，其中包括 15 种旋翼无人机、3 种固定翼无人机和 2 种无人直升机。该数据集还特别突出了背景的复杂性和多样性，如图 6 所示，画面中的无人机背景包括居民建筑、

商业中心、山地、林木、河流、工厂、海岸等 30 个不同的地点，较好地反映了无人机探测系统在实际部署时可能会遇到的多种场景。该数据集包含 200000 张图像，其中包括 140000 张训练集图像和 60000 张测试集图像以及每张图像对应的标注真值，图像分辨率为 1920×1080。

图 6 UAV dataset[48]示例图片

Fig.6 Example images in the UAV dataset[48]

2.2 面向静态图像的无人机检测

围绕无人机探测预警任务，业内学者基于主流目标检测的算法开发了相当数量的无人机目标检测算法。这些算法主要解决的问题包括：基于通用目标检测算法的多尺度无人机目标检测、少样本无人机目标检测和红外图像无人机目标检测等。

2.2.1 基于通用目标检测算法的无人机目标检测

无人机目标检测算法按照是否显式生成候选区域，同样可大致分为双阶段和单阶段算法，两种类型的算法各具优势。在相同的数据集中，不采用任何优化算法的情况下，双阶段的 Faster R-CNN 算法有较高的检测准确率，单阶段的 YOLO 系列算法处理速度较快。当前计算机视觉领域提出的面向静态图像的无人机目标检测算法介绍如下。

针对远距离无人机在成像视野中尺寸小的问题，Vasileios[49]通过在 Faster R-CNN 训练中加入深度超分辨率模型提出了新型无人机目标检测算法。如图 7 所示，该算法中的超分辨率模型[50]采用深度残差网络来提取特征并重构图像，提升输入图像中无人机小目标的分辨率，进而提升基于 Faster R-CNN 目标检测模型的召回率。Celine Craye[51]等人将无人机的检测分为两个步骤，首先将视频图像的时空序列输入 U-Net[52]模型中来获取无人机候选区域，再使用 ResNet101 模型对其进行分类，该算法与双阶段算法 R-CNN 相似，能够提升对小目标无人机的检测效果。然而，采用基于 Faster R-CNN 等双阶段的检测方法在计算实时性方面存在一定局限性。

鉴于 YOLO 系列算法计算效率方面存在优势，文献[53]开发了基于 YOLOv2 的无人机目标检测算法。然而，由于 YOLOv2 算法在工作时需要在图像上划分网格，而且每个网格最多只能预测单个目标，因此多个目标落入同一个网格时就会出现漏检。此外，传统深度卷积网络在所学特征对方向和尺度变化鲁棒性差，因此对于小物体和重叠物体检测效果不佳。

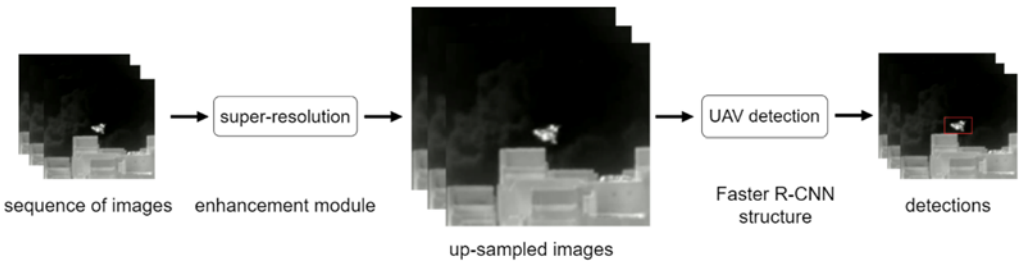

图 7 超分辨率增强模块结合 Faster R-CNN 模型的无人机检测算法流程图[49]

Fig.7 Flowchart of the UAV detection algorithm[49] combined with the super-resolution enhancement module and the Faster R-CNN model

文献[54]基于 YOLOv3 的 Darknet53 骨干网络采用 Gabor 滤波器调制 DCNN 中的卷积核，借以增强特征对方向和尺度变化的鲁棒性，并在数据集上进行了验证，性能超过了基于尺度不变特征变换（Scale-invariant feature transform, SIFT）特征和局部特征聚合描述符、词袋和费舍尔向量等分类模型相结合的方法。但是该算法尚未与 YOLOv3 等基于 DCNN 的目标检测方法进行对比，Gabor 滤波器调制 DCNN 算法的优势没有得到验证。

由于无人机目标在成像视场中的尺度变化较大，YOLOv3 中在 3 个尺度层面的检测难以有效覆盖无人机尺度变化范围。针对该问题，文献[55]在 YOLOv3 模型中加入多尺度的特征融合，来检测尺度变化显著的无人机。文献[48]同样基于 YOLOv3 模型提出了针对无人机目标检测的 UAVDet 模型（如图 8 所示），将 YOLOv3 扩展为 4 个尺度进行预测，而且在第二个下采样后增加两个残差模块来获得更多定位信息。需要指出的是，由于单阶段算法没有显式生成候选框的过程，YOLO 系列算法需要事先使用 k-means[56]聚类算法根据数据集生成先验框，因此在使用 YOLO 系列

算法进行目标检测时，同样需要使用 k-means 对特定的无人机数据集聚类生成更适合无人机的先验框。同时，为了解决图像中存在的运动模糊问题，对数据集用高斯模糊和运动模糊的方法进行数据增强，有效提升检测准确度和召回率。

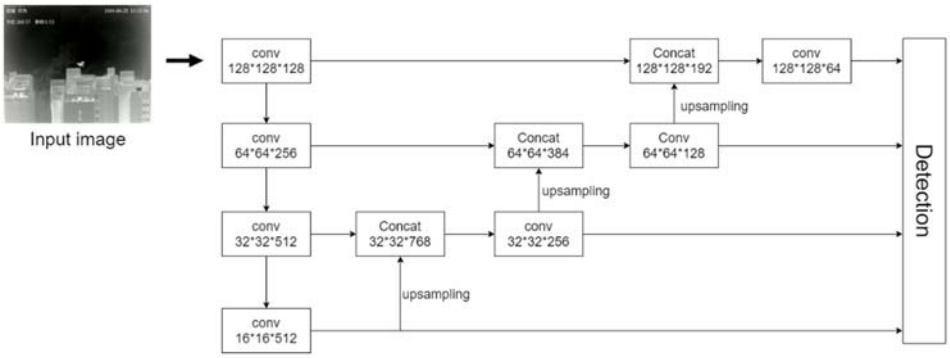

图 8 基于多尺度 YOLOv3 的 UAVDet 算法[48]流程示意图

Fig.8 Flowchart of the UAVDet algorithm[48] that is based on the multi-scale YOLOv3 structure

2.2.2 迁移学习和数据增广在无人机检测中的应用

如前文所述，基于 DCNN 目标检测算法通常是数据驱动的监督学习算法，需要依靠较大规模的数据集进行模型训练和性能评估，但是目前业内缺乏公开的大型无人机检测数据集，基于少样本数据集训练 DCNN 模型容易造成过拟合问题，因此研究人员通过迁移学习和数据增广来缓解这个矛盾。

迁移学习是一种机器学习领域常用的技术，通常指将一个预训练的模型被重新用在另一个任务中的过程，能够将模型在一种数据集中学到的知识迁移应用在另一个数据集中，进而提高模型的泛化性能。具体在无人机检测任务上，可以首先在其他类型（如通用目标检测）的大规模数据集中对模型进行比较充分的训练，然后将预训练的网络在特定的相对较小规模无人机检测数据集上进行微调。Muhamma 等人[57]将经过 ImageNet 数据集预训练过的模型在 Drone-vs-Bird Detection Challenge 数据集上进行微调，进而使模型能够更好地检测无人机。作者采用 Faster R-CNN 算法，对比了 ZFNet，VGG16 和 VGG_CNN_1024 三种特征提取网络的检测性能，结果显示 VGG16 模型在该数据集取得相对更好的性能。在 2019 年的 Drone-vs-Bird Detection Challenge 挑战赛中，竞赛数据引入了更复杂的目标背景、更丰富的光照条件以及更多变的画面缩放，甚至还有很多低对比度画面和多种鸟类存在的场景。Nalamati 等人[58]采用了类似的迁移学习技术路线，并且对比了 Faster R-CNN 和 SSD 算法，其实验结果表明基于 ResNet101 网络的 Faster R-CNN 算法检测准确度较好，但是在实时性方面存在局限性。

数据增广是另外一种缓解模型训练过拟合问题的常用手段，通过变换现有数据或根据现有数据创建新的合成数据来增加样本数量。常用的数据增广方法有图像几何变换、翻转、颜色修改、裁剪、旋转、添加噪声、随机遮挡、透明度混叠、裁剪混叠等。这些方法都可以引入到无人机目标检测中来缓解少样本的问题。例如，针对大规模无人机目标检测数据获取困难的问题，文献[59]将鸟和无人机的图像块拼接到不同的背景图片中，最终得到了 676534 张图片，进而可以更好地训练无人机目标检测模型。

2.2.3 红外图像无人机检测

可见光图像分辨率高，通常具有较好的纹理和形状信息，非常利于 DCNN 模型进行特征学习和表征，进而实现无人机检测。但是，在雾天或夜间等光照条件差的情况下，可见光传感器获得的图像数据能见度差，难以捕获无人机目标。相比之下，红外成像传感器具有探测距离远、全天候工作、光照条件适应性强等优势，但同时也存在分辨率小、对比度差、信噪比低、纹理形状信息缺乏等缺点，因此面向红外图像的无人机目标检测更具挑战性。文献[60]对红外图像进行倒置，直方图均衡，去噪和锐化预处理后，在 YOLOv3 模型的基础上引入 SPP 模块和 GIOU（Generalized Intersection over Union）损失函数，改善了模型对近距离大目标和边缘目标的检测能力。文献[61]使用全卷积神经网络对红外图像进行分割，利用视觉显著性机制对小目标进行增强，抑制背景和虚警，检测结果优于典型的红外目标检测算法。文献[62]利用红外图像与可见光图像的互补特性进行多尺度显著特征融合，使用改进的 YOLOv3 模型进行检测，采用注意机制对辅助网络和骨干网络的特征信息融合，增强有效信息通道，抑制无效信息通道，提升小目标检测效果。

当红外图像中的无人机目标尺寸非常小时（例如小于 9×9 像素），需要将无人机目标看作红外小目

标进行检测。基于手工特征的红外小目标检测典型方法包括高斯差分滤波器、局部对比度算法[63]、二维最小均方滤波器[64]、形态学 Top-hat 变换[65-66]算法、非线性图像块处理[67]模型等。针对基于手工特征的方法自适应能力有限的问题，近来有学者将 DCNN 引入红外小目标检测领域。文献[68]将小目标检测问题转化为小目标位置分布分类问题，利用全卷积网络对红外小目标进行背景抑制和目标增强，同时获得目标潜在区域；然后将原始图像和目标潜在区域同时输入分类网络，进而输出目标检测结果。在 50000 张图片上的训练和测试结果表明，该方法能够有效检测复杂背景和低信噪比甚至存在运动模糊的小目标。但是，该方法仍然存在虚警率较高的问题，这是因为在很多情况下，仅仅依赖静态外观特征难以区分真实小目标和背景中的非目标点状物体。因此，在复杂背景和低信噪比情况下有效利用时空上下文信息进行红外小目标检测仍然是一项具有挑战性的任务[69]。

2.3 面向视频数据的无人机检测

面向视频数据的无人机检测是无人机检测的核心任务，一方面是因为基于光电传感器的无人机探测数据通常为视频数据（即图像序列），另一方面在单帧静态图像上无法辨识目标时需要借助视频数据中的上下文时空信息进行目标增强和检测识别。然而，基于视频数据实现无人机检测也存在若干难点。一是视频序列中的连续帧之间存在大量冗余信息；二是复杂运动模态的背景会对目标检测造成极大干扰；三是无人机剧烈运动或者传感器镜头失焦会造成目标外观模糊。因此，面向视频数据的无人机检测需要联合静态外观信息和目标特异性运动信息（即空域和时域的上下文信息）进行判别。如前文所述，计算机视觉领域已经提出了相当数量的面向静态图像的目标检测方法，但是面向视频数据的目标检测特别是无人机检测的研究还相对较少，已有的工作主要借助光流和时序特征来表征运动信息，进而更好地实现视频数据中的目标检测任务。

2.3.1 基于光流场的视频目标检测

视频运动目标检测是在视频的连续图像序列中将运动物体检测出来的过程，运动目标检测方法包括两帧/多帧差分法、背景抑制法和光流法等，其中光流法对运动信息的表征最为有效。光流的概念通常是指空间中的运动物体在成像平台上像素运动的瞬时速度（包含速率和方向）。如果图像中没有运动目标时，整幅图像中的光流是连续变化的；如果存在运动目标，那么运动目标形成的光流场与背景的光流场就会存在差异，进而可以将运动目标与背景进行区分。光流场的有效计算方法最初是由 Horn 和 Schunck[70]于 1981 年提出，该方法假设物体的瞬时灰度值不变且在整个图像上平滑变化来求解光流。Lueas 和 Kanade[71]提出了改进光流算法，假设在一个小空间领域上运动矢量保持恒定，然后使用加权最小二乘法估计光流。但是以上方法需要通过迭代的方式计算光流，通常计算量比较大。更重要的是，该类方法对图像连续帧亮度恒定的假设过于严格，因而在复杂光照条件下的光流计算准确度有限。2015 年 Fischer 将光流计算转化为监督学习问题，提出了基于深度学习的 FlowNet[72]方法。如图 9 所示，FlowNet 模型的输入为连续的两帧图像（支持 RGB 图像），网络分为卷积下采样和反卷积上采样两部分，其中下采样网络负责分层提取特征和编码高级语义信息，反卷积网络利用高级语义信息解码和分层提取的特征进行光流预测，借助大量数据的训练，显著提升了光流计算性能。后续的 FlowNet2.0[73]模型和 RAFT[74]模型进一步提高了基于 DCNN 的光流计算能力。

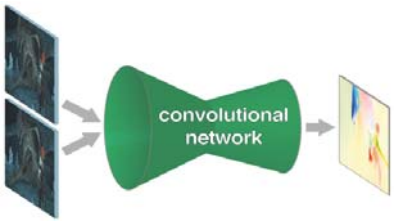

图 9 FlowNet[72]模型计算光流过程示意图

Fig.9 Diagram of the FlowNet[72] for calculating optical flow

鉴于光流场在目标运动信息表征方面存在许多优良特性，可以预期将光流信息引入视频运动目标检测将显著提升视频目标检测的性能。一种思路是利用光流信息消除图像连续帧之间的冗余信息。例如，文献[75]发现 DCNN 模型提取的相邻帧图像的特征图通常非常相似，因此利用 DCNN 模型逐帧处理视频将消耗大量的非必要计算资源，因此可以在处理视频时按固定时间间隔仅选取和处理关键帧，而非关键帧的特征可以由关键帧的特征借助光流信息迁移获得。由于光流计算速度远高于 DCNN 特征提取速度，因此该方法大幅减少了视频处理的计算量，从而提升了视频目标检测速度。然而，该方法主要适用于运动物体和背景在相邻帧之间连续变化的情况。另一种利用光流信息进行视频运动目标检测的思路是将光流信息与静态外观信息进行叠加，从而进一步增加目标与背景之间的差异性。文献[76]采用 DCNN 模型获得当前帧和参考帧的外观特征图，同时采用 FlowNet 模型预测当前帧和参考帧的光流场，然后将对应帧的外观特征图与光流信息叠加为时空混合特征图，进而根据当前帧

和参考帧的时空混合特征图获得目标检测结果。这种方法有效地利用了视频数据的时空信息，而且有助于解决运动目标模糊的问题，因此显著提升了目标检测性能。但是该方法对目标强度和局部信噪比有一定的要求，而且主要适用于离线视频目标检测，在实时在线目标检测方面还需要改进。借助无人机视频及其标注数据，这些基于光流场的目标检测模型可以有效迁移到无人机检测任务中。

2.3.2 基于多帧相关特征的无人机检测

光流法通常在视频图像质量较高时能够有效表征目标运动信息，但在目标模糊或者极端弱隐的情况下容易失效。针对该问题，Rozantsev 等人[77]利用时序维度上的多个连续帧对目标能量进行累积进而达到目标增强的目的。如图 10 所示，首先用不同尺度的滑动窗口在图像序列中获取时空图像立方体（Spatio-Temporal Image Cube）；然后对每个 cube 进行运动补偿得到时空稳像立方体，这个操作能够极大地增强候选目标的能量，增强潜在目标的局部信噪比；最后再采用分类器判断该时空稳像立方体是否包含目标，并通过非极大值抑制技术优化目标检测结果。该方法与基于光流的方法相比，抗复杂背景干扰和抗目标运动模糊的能力显著提高。

由于卷积神经网络训练过程丢失时间维度信息，无法保证特征的时空一致性的问题，除了上述用运动补偿来获得时空稳定特征的方法外，有研究者提出输入图像序列到神经网络中来提取隐含的运动信息，主要包括 Siamese[78]和循环神经网络（Recurrent Neural Network, RNN）[79]网络。文献[80]提出了基于全卷积神经网络的目标检测框架，该框架通过使用 Siamese 网络来提取时序信息，同时，RNN 作为一种时间序列模型也能够提供时序信息，在循环神经网络中，当前层的输出不仅与输入有关，还取决于前一时刻的输入，使得神经网络具有“记忆”功能，RNN 主要应用于自然语言处理领域。

面向视频数据的无人机检测在实际应用中通常会遇到树枝、飞鸟等动态的非目标干扰物，单纯利用帧间光流信息难以将其与真实目标区分开来。针对该问题，文献[81]发现无人机作为一种人工设计的飞行器，其飞行动力学具有一定的特异性规律，因此提出一种基于多帧目标形态变化特性和航迹规律的无人机目标检测方法，能够一定程度上降低目标检测虚警率。但是该方法的目标分割过程建立在背景差分法之上，因此对背景运动复杂度以及传感器运动（包括移动、转动和扰动）幅度具有较高的要求。

2.4 无人机检测的难点问题及解决思路

2.4.1 无人机检测的难点问题

如图 11 所示，小型民用无人机目标检测的难点主要包括目标特性复杂性和背景复杂性两个方面。

无人机检测的目标特性复杂性主要体现在：①无人机的型号、颜色、外形、运动特性等复杂多变；②无人机数量较多时，在成像视场中有时会出现相互重叠、遮挡等情况；③无人机距离传感器较远时，在成像视场中尺寸较小，缺乏形状和纹理等信息；④无人机快速机动或者传感器失焦时会造成目标模糊；⑤无人机运动或者传感器变焦时会造成目标尺度变化。

无人机检测的背景复杂性主要体现在：①无人机的天空背景有时会存在云朵、强光等干扰；②无人机飞行高度较低时，其背景会出现建筑物、塔吊、山坳等静态物体或者树枝、旗帜、海浪等动态物体；③无人机飞行时背景中会出现飞鸟、风筝等干扰物。

此外，图像噪声和成像过程扰动也会显著降低深度卷积网络的模式判别正确率。而且，业内目前缺乏大型公开无人机数据集，为高容量模型的训练和评估造成一定困难。若干已有工作[47-48]虽然通过自建数据集来缓解数据需求矛盾，但是难以用于算法性能的横向对比。

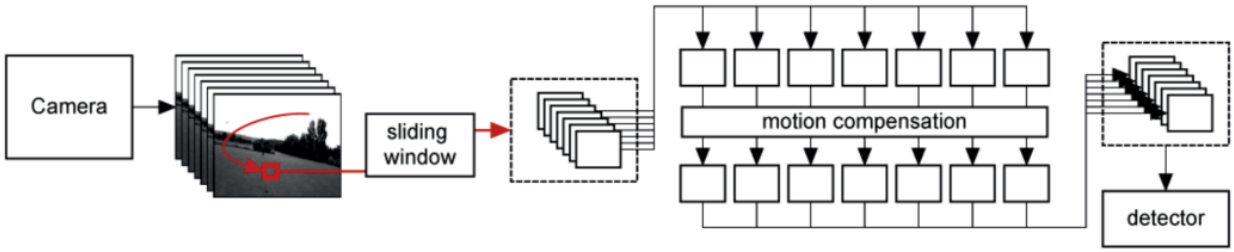

图 10 运动补偿的目标检测算法流程[77]

Fig.10 Flow chart of the object detection algorithm incorporating motion compensation[77]

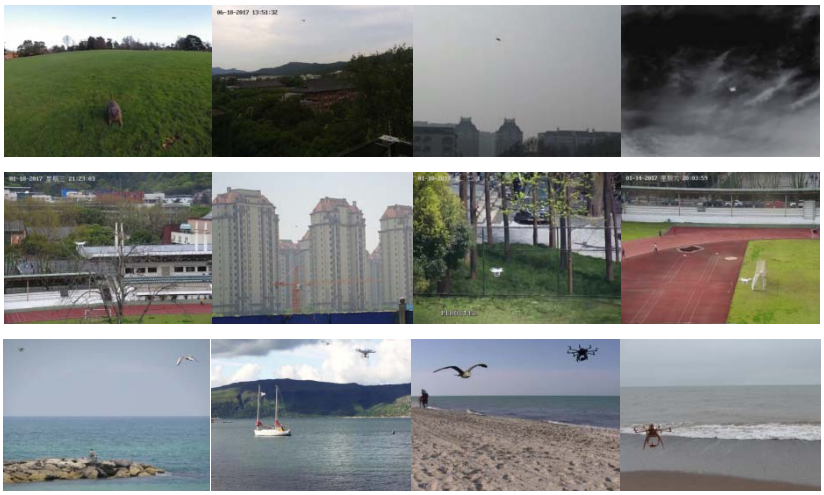

图 11 无人机检测的难点和瓶颈性问题示例图像

Fig.11 Image examples to demonstrate difficulties and bottlenecks in drone detection

注：第一行：目标小尺寸且缺乏外观信息[47,55,62]；第二行：背景复杂多样[47-48]；第三行：目标尺度异质性问题[53]

Note: Row 1: Targets that are small and weak in appearance information[47,55,62]; Row 2: Targets in complex and diverse backgrounds[47-48]; Row 3: Targets that have heterogeneous scales [53])

2.4.2 突破小型无人机检测瓶颈的若干思路

通过前文对视觉目标检测文献的梳理可以发现，当前算法虽然已经初步实现了小型民用无人机的自动化检测，但是在复杂条件下实现低虚警率、高召回率、强鲁棒性的无人机检测仍然是一项极具挑战性的任务。针对基于深度卷积神经网络的小型民用无人机检测系统存在的瓶颈性问题，未来工作在以下几个方面值得深入研究。

一是更合理地根据静态图像中上下文信息搜索和辨别目标特性复杂的无人机目标。人类在目标发现和识别过程中通常伴随眼跳现象，即反映眼动规律的注视点会按照无意注意和任务驱动有意注意的规律跳跃性感知语义要素，并通过高级推理快速完成目标价值判定。与通用目标检测和显著性检测等视觉任务不同，小型无人机目标的尺寸、纹理、形状等信息的特异性较低。因此探究如何利用空间上下文（Spatial Context）信息进行任务驱动的推理式快速搜索以及根据关键语义要素实现无人机目标模式判别具有重要的理论及应用意义。

二是更有效地提取和表征目标运动信息，并将其作为关键特征用于无人机目标判别。从小型无人机检测的人类行为实验结果显示，在很多复杂场景下即使是人类也很难仅凭小型无人机的静态表观特性完成目标检测任务，而视频数据的时间上下文（Temporal Context）信息是准确检测目标的重要基础。人脑视觉信息加工过程中，同样需要借助背侧通路和腹侧通路分别处理运动和静态表观信息，并在多个层次上进行横向信息投射和跨层交互融合。因此，探究无人机目标运动信息提取和表征方法，利用目标视觉运动信息辅助目标定位和识别，进而通过消除相邻视频帧的冗余信息增加目标检测效率，具有重要的研究价值。

三是更好地融合目标静态表观特征和运动特征，综合利用时空上下文信息进行无人机目标检测。人脑视觉系统中存在并行信息处理的大细胞通路和小细胞通路，在脑区架构方面存在背侧通路和腹侧通路，分别处理视觉运动和静态表观信息，并在多个层次上进行有效融合。因此，综合利用时空上下文信息进行无人机目标检测将是未来解决小型无人机目标检测瓶颈问题的关键。

四是建立大规模公开小型无人机数据集。由于目前业内基于深度卷积神经网络的先进算法大多是基于数据驱动的算法，需要依赖标注数据进行模型训练、验证和测试。业界现有的若干数据集在反映多类型复杂背景和多样化无人机目标方面还存在一定差距，因此建立大规模公开无人机数据集对促进小型民用无人机检测技术的研究和发展具有重要意义。此外，引入自监督学习、无监督学习等机器学习技术也是缓解无人机数据不足矛盾的一个重要思路。

3 总结与展望

小型民用无人机为人类社会带来便利的同时也给公共安全造成了较大威胁。面向高准确性和高鲁棒性的无人机目标检测，计算机视觉领域已经提出了相当数量的算法。本文首先介绍了目标检测领域中基于深度卷积神经网络的主流算法，然后针对小型无人机检测任务分别总结了面向静态图像和视频数据的无

人机检测方法，进而归纳了造成无人机检测困难的主要原因。

业内现有工作虽然已经初步实现了小型民用无人机自动目标检测，但是在复杂条件下实现低虚警率、高召回率、强鲁棒性、低能耗性的无人机检测仍然是一项极具挑战性的任务。目标特性复杂性和目标背景复杂性都会对无人机检测算法的性能造成严重影响，图像噪声和对抗性扰动也会显著降低深度卷积网络的模式判别正确率。此外，业内目前缺乏大型公开无人机数据集，为高容量模型的训练和评估造成一定困难。虽然有研究人员通过自建数据集来缓解数据需求矛盾，但是难以用于算法性能的横向对比。针对基于深度卷积神经网络的小型民用无人机检测系统存在的瓶颈性问题，预期未来工作将围绕图像空间上下文信息提取与表征、视频时间上下文信息提取与表征、视觉时空上下文信息融合和大规模数据集的建立等方面展开。

值得指出的是，深度卷积神经网络模型已经在通用目标检测和物体分类等视觉任务中取得了较好的性能，然而在复杂背景下的低慢小目标检测任务中依然无法达到人类甚至非人灵长类的识别水平。深度卷积神经网络虽然符合神经可塑性、非线性整合和分层加工等机制，但仍然是对生物神经系统高度抽象化的模型，关于深度卷积网络的可解释性、小样本泛化性、对抗鲁棒性等方面的研究还处于初始阶段，人工智能和计算机视觉领域还比较缺乏能够有效模拟灵长类认知推理、学习记忆、反馈调节等机制的算法和模型。因此，通过借鉴和模拟灵长类视知觉和学习记忆等神经机制提出更符合生物视觉特性的视觉计算模型[82]，对于突破小型无人机视觉检测在可解释性、鲁棒性、可迁移性和低功耗等方面存在的瓶颈性问题具有重要的理论研究价值和良好的应用前景。

参考文献:

[1] WANG J, LIU Y, SONG H. Counter-unmanned aircraft system (s)(C-UAS): State of the art, challenges, and future trends[J]. *IEEE Aerospace and Electronic Systems Magazine*, 2021, **36**(3): 4-29.

[2] LI Xiaoping, LEI Songze, ZHANG Boxing, et al. Fast aerial UAV detection using improved inter-frame difference and SVM[C]//*Journal of Physics: Conference Series. IOP Publishing*, 2019, **1187**(3): 032082.

[3] WANG C, WANG T, WANG E, et al. Flying small target detection for anti-UAV based on a Gaussian mixture model in a compressive sensing domain[J]. *Sensors*, 2019, **19**(9): 2168.

[4] Seidaliyeva U, Akhmetov D, Ilipbayeva L, et al. Real-time and accurate drone detection in a video with a static background[J]. *Sensors*, 2020, **20**(14): 3856.

[5] ZHAO W, CHEN X, CHENG J, et al. An application of scale-invariant feature transform in iris recognition[C]//*Proceedings of the IEEE/ACIS* 12th *International Conference on Computer and Information Science, IEEE*, 2013: 219-222.

[6] SHU C, DING X, FANG C. Histogram of the oriented gradient for face recognition[J]. *Tsinghua Science and Technology*, 2011, **16**(2): 216-224.

[7] SHEN Y K, CHIU C T. Local binary pattern orientation based face recognition[C]//*Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE*, 2015: 1091-1095.

[8] YUAN Xiaofang, WANG Yaonan. Parameter selection of support vector machine for function approximation based on chaos optimization[J]. *Journal of Systems Engineering and Electronics*, 2008, **19**(1): 191-197.

[9] FENG J, WANG L, Sugiyama M, et al. Boosting and margin theory[J]. *Frontiers of Electrical and Electronic Engineering*, 2012, **7**(1): 127-133.

[10] WEI L, HONG Z, Gui-Jin H. NMS-based blurred image sub-pixel registration[C]//*Proceedings of the International Conference on Image Analysis and Signal Processing. IEEE*, 2011: 98-101.

[11] 罗会兰, 陈鸿坤. 基于深度学习的目标检测研究综述[J]. **电子学报**, 2020, **48**(6):1230-1239.
LUO Huilan, CHEN Hongkun. Survey of object detection based on deep learning[J]. *Acta Electronica Sinica*, 2020, **48**(6): 1230-1239.

[12] Bosquet B, Mucientes M, Brea V M. STDNet: exploiting high resolution feature maps for small object detection[J]. *Engineering Applications of Artificial Intelligence*, 2020, **91**: 103615.

[13] SUN H, YANG J, SHEN J, et al. TIB-Net: Drone detection network with tiny iterative backbone[J]. *IEEE Access*, 2020, **8**: 130697-130707.

[14] LIU L, OUYANG W, WANG X, et al. Deep learning for generic object detection: a survey[J]. *International Journal of Computer Vision*, 2020, **128**(2): 261-318.

[15] 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*, 2014: 580-587.

[16] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//*Proceedings of the Advances in Neural Information Processing Systems*, 2012, **25**: 1097-1105.

[17] HE K, ZHANG X, REN S, 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.

[18] Girshick R. Fast R-CNN[C]//*Proceedings of the IEEE International Conference on Computer Vision*, 2015: 1440-1448.

[19] Zeiler M D, Fergus R. Visualizing and understanding convolutional networks[C]//*Proceedings of the European Conference on Computer Vision*, 2014: 818-833.

[20] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J/OL]. *arXiv preprint arXiv*:1409.1556, 2014.

[21] 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*, 2016, **39**(6): 1137-1149.

[22] Bell S, Lawrence Zitnick C, Bala K, et al. Inside-outside net: detecting objects in context with skip pooling and recurrent neural

networks[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 2874-2883.

[23] LE Q V, Jaitly N, Hinton G E. A simple way to initialize recurrent networks of rectified linear units[J/OL]. *arXiv preprint arXiv*: 1504.00941, 2015.

[24] DAI J, LI Y, HE K, et al. R-FCN: Object detection via region-based fully convolutional networks[J/OL]. *arXiv preprint arXiv*:1605.06409, 2016.

[25] 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*, 2016: 770-778.

[26] LIN T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 2117-2125.

[27] He K, Gkioxari G, Dollár P, et al. Mask R-CNN[C]//*Proceedings of the IEEE International Conference on Computer Vision*, 2017: 2961-2969.

[28] XIE S, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 1492-1500.

[29] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 8759-8768.

[30] LI Y, CHEN Y, WANG N, et al. Scale-aware trident networks for object detection[C]//*Proceedings of the IEEE International Conference on Computer Vision*, 2019: 6054-6063.

[31] DUAN K, XIE L, QI H, et al. Corner proposal network for anchor-free, two-stage object detection[C]//*European Conference on Computer Vision. Springer, Cham*, 2020: 399-416.

[32] Newell A, YANG K, DENG J. Stacked hourglass networks for human pose estimation[C]//*Proceedings of the European Conference on Computer Vision*, 2016: 483-499.

[33] 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*, 2016: 779-788.

[34] Szegedy C, LIU W, JIA Y, et al. Going deeper with convolutions [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015: 1-9.

[35] LIU W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector[C]//*Proceedings of the European Conference on Computer Vision. Springer*, 2016: 21-37.

[36] Redmon J, Farhadi A. YOLO9000: Better, faster, stronger[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 7263-7271.

[37] LIN T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//*Proceedings of the IEEE International Conference on Computer Vision*, 2017: 2980-2988.

[38] Redmon J, Farhadi A. YOLOv3: An incremental improvement[J/OL]. *arXiv preprint arXiv*: 1804.02767, 2018.

[39] ZHOU P, NI B, GENG C, et al. Scale-transferrable object detection[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 528-537.

[40] HUANG G, LIU Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 4700-4708.

[41] LAW H, DENG J. Cornernet: Detecting objects as paired keypoints[C]//*Proceedings of the European Conference on Computer Vision*, 2018: 734-750.

[42] Bochkovskiy A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection[J/OL]. *arXiv preprint arXiv*: 2004.10934, 2020.

[43] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers[C]//*European Conference on Computer Vision. Springer, Cham*, 2020: 213-229.

[44] JIANG N, WANG K, PENG X, et al. Anti-UAV: A large multi-modal benchmark for UAV tracking[J]. *arXiv preprint arXiv*:2101.08466, 2021.

[45] ZHAO J, WANG G, LI J, et al. The 2nd Anti-UAV Workshop & Challenge: Methods and results[J]. *arXiv preprint arXiv*:2108.09909, 2021.

[46] Coluccia A, Fascista A, Schumann A, et al. Drone-vs-Bird detection challenge at IEEE AVSS2019[C]//*Proceedings of the IEEE International Conference on Advanced Video and Signal Based Surveillance. IEEE*, 2019: 1-7.

[47] WU M, XIE W, SHI X, et al. Real-time drone detection using deep learning approach[C]//*Proceedings of the International Conference on Machine Learning and Intelligent Communications*, 2018: 22-32.

[48] ZHAO W, ZHANG Q, LI H, et al. Low-altitude UAV detection method based on one-staged detection framework[C]//*Proceedings of the International Conference on Advances in Computer Technology, Information Science and Communications IEEE*, 2020: 112-117.

[49] Magoulianitis V, Ataloglou D, Dimou A, et al. Does deep super-resolution enhance UAV detection?[C]//*Proceedings of the IEEE International Conference on Advanced Video and Signal Based Surveillance IEEE*, 2019: 1-6.

[50] Kim J, Kwon Lee J, Mu Lee K. Accurate image super-resolution using very deep convolutional networks[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 1646-1654.

[51] Craye C, Ardjoune S. Spatio-temporal semantic segmentation for drone detection[C]//*Proceedings of the IEEE International Conference on Advanced Video and Signal Based Surveillance. IEEE*, 2019: 1-5.

[52] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation[C]//*Proceedings of the International Conference on Medical Image Computing and Computer-assisted Intervention*, 2015: 234-241.

[53] Aker C. End-to-end Networks for Detection and Tracking of Micro Unmanned Aerial Vehicles[D]. Ankara, Turkey: Middle East Technical University, 2018.

[54] 张锡联, 段海滨. 一种基于 Gabor 深度学习的无人机目标检测算法[J]. **空间控制技术与应用**, 2019, **45**(4): 38-45.
ZHANG X, DUAN H. A target detection algorithm for UAV based on Gabor deep learning[J]. *Aerospace Control and Application*, 2019, **45**(4): 38-45.

[55] 马旗, 朱斌, 张宏伟, 等. 基于优化 YOLOv3 的低空无人机检测识别方法[J]. **激光与光电子学进展**, 2019, **56**(20): 279-286.
MA Q, ZHU B, ZHANG H, et al. Low-Altitude UAV detection and

recognition method based on optimized YOLOv3[J]. *Laser & Optoelectronics Progress*, 2019, **56**(20): 279-286.

[56] Cohen M B, Elder S, Musco C, et al. Dimensionality reduction for k-means clustering and low rank approximation[C]//*Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*, 2015: 163-172.

[57] Saqib M, Khan S D, Sharma N, et al. A study on detecting drones using deep convolutional neural networks[C]//*Proceedings of the IEEE International Conference on Advanced Video and Signal Based Surveillance. IEEE*, 2017: 1-5.

[58] Nalamati M, Kapoor A, Saqib M, et al. Drone detection in long-range surveillance videos[C]//*Proceedings of the IEEE International Conference on Advanced Video and Signal Based Surveillance. IEEE*, 2019: 1-6.

[59] Aker C, Kalkan S. Using deep networks for drone detection[C]//*Proceedings of the IEEE International Conference on Advanced Video and Signal Based Surveillance. IEEE*, 2017: 1-6.

[60] 张汝榛, 张建林, 祁小平, 等. 复杂场景下的红外目标检测[J]. **光电工程**, 2020, **47**(10): 128-137.
ZHANG R, ZHANG J, QI X, et al. Infrared target detection and recognition in complex scene[J]. *Opto-Eletronic Engineering*, 2020, **47**(10):128-137.

[61] 刘俊明, 孟卫华. 融合全卷积神经网络和视觉显著性的红外小目标检测[J]. **光子学报**, 2020, **49**(7):46-56.
LIU J, MENG W. Infrared small target detection based on fully convolutional neural network and visual saliency[J]. *Acta Photonica Sincia*, 2020, **49**(7): 46-56.

[62] 马旗, 朱斌, 程正东, 等.基于双通道的快速低空无人机检测识别方法[J]. **光学学报**, 2019, **39**(12): 105-115.
MA Q, ZHU B, CHENG Z, et al. Detection and recognition method of fast low-altitude unmanned aerial vehicle based on dual channel[J]. *Acta Optica Sinica*, 2019, **39**(12): 105-115.

[63] CUI Z, YANG J, JIANG S, et al. An infrared small target detection algorithm based on high-speed local contrast method[J]. *Infrared Physics & Technology*, 2016, **76**: 474-481.

[64] ZHAO Y, PAN H, DU C, et al. Bilateral two-dimensional least mean square filter for infrared small target detection[J]. *Infrared Physics & Technology*, 2014, **65**: 17-23.

[65] Lange H. Real-time contrasted target detection for IR imagery based on a multiscale top hat filter[C]//*Signal Processing, Sensor Fusion, and Target Recognition VIII. International Society for Optics and Photonics*, 1999, **3720**: 214-226.

[66] BAI X, ZHOU F, ZHANG S, et al. Top-Hat by the reconstruction operation-based infrared small target detection[C]//*Proceedings of the International Conference in Electrics, Communication and Automatic Control Proceedings*, 2012: 867-873.

[67] 王刚, 陈永光, 杨锁昌, 等. 采用图像块对比特性的红外弱小目标检测[J]. **光学精密工程**, 2015, **23**(5): 1424-1433.
WANG G, CHEN Y, YANG S, et al. Infrared dim and small target detection using image block contrast characteristics[J]. *Optics and Precision Engineering*, 2015, **23**(5):1424-1433.

[68] 吴双忱, 左峥嵘. 基于深度卷积神经网络的红外小目标检测[J]. **红外与毫米波学报**, 2019, **38**(3): 371-380.
WU S, ZUO Z. Infrared small target detection based on deep convolutional neural network[J]. *Journal of Infrared and Millimeter Waves*, 2019, **38**(3): 371-380.

[69] 李俊宏, 张萍, 王晓玮, 等. 红外弱小目标检测算法综述[J]. **中国图象图形学报**, 2020, **25**(9): 1739-1753.
LI J, ZHANG P, WANG X, et al. A survey of infrared dim target detection algorithms[J]. *Journal of Image and Graphics*, 2020, **25**(9): 1739-1753.

[70] Horn B K P, Schunck B G. Determining optical flow[C]//*Techniques and Applications of Image Understanding. International Society for Optics and Photonics*, 1981, **281**: 319-331.

[71] Lucas B D, Kanade T. An iterative image registration technique with an application to stereo vision[C]//*Proceedings of the International Joint Conference on Artificial Intelligence*, 1981: 674-679.

[72] Dosovitskiy A, Fischer P, Ilg E, et al. Flownet: Learning optical flow with convolutional networks[C]//*Proceedings of the IEEE International Conference on Computer Vision*, 2015: 2758-2766.

[73] Ilg E, Mayer N, Saikia T, et al. FlowNet 2.0: Evolution of optical flow estimation with deep networks[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 2462-2470.

[74] Teed Z, Deng J. Raft: Recurrent all-pairs field transforms for optical flow[C]// *Proceedings of the European Conference on Computer Vision*, 2020: 402-419.

[75] ZHU X, XIONG Y, DAI J, et al. Deep feature flow for video recognition[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 2349-2358.

[76] ZHU X, WANG Y, DAI J, et al. Flow-guided feature aggregation for video object detection[C]//*Proceedings of the IEEE International Conference on Computer Vision*, 2017: 408-417.

[77] Rozantsev A, Lepetit V, Fua P. Flying objects detection from a single moving camera[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015: 4128-4136.

[78] Bertinetto L, Valmadre J, Henriques J F, et al. Fully-convolutional siamese networks for object tracking[C]//*Proceedings of the European Conference on Computer Vision*, 2016: 850-865.

[79] Stewart R, Andriluka M, Ng A Y. End-to-end people detection in crowded scenes[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 2325-2333.

[80] ZHAO B, ZHAO B, TANG L, et al. Deep spatial-temporal joint feature representation for video object detection[J]. *Sensors*, 2018, **18**(3): 774.

[81] 刘宜成, 廖鹭川, 张劲, 等. 基于轨迹和形态识别的无人机检测方法[J]. **计算机工程**, 2020, **46**(12): 283-289.
LIU Y, LIAO L, ZHANG J, et al. UAV detection method based on trajectory and shape recognition[J]. *Computer Engineering*, 2018, **18**(3): 774.

[82] 吴飞, 阳春华, 兰旭光, 等. 人工智能的回顾与展望[J]. **中国科学基金**, 2018, **32**(3): 243-250.
WU F, YANG C H, LAN X, et al. Retrospect and prospect of artificial intelligence[J]. *Bulletin of National Natural Science Foundation of China*, 2018, **32**(3): 243-250.