

文章编号:1674-8190(2023)01-013-14

无人机航拍野生动物智能检测与统计方法综述

祝宁华¹, 郑江滨², 张阳³

(1. 西北工业大学 工程实践训练中心, 西安 710072)

(2. 西北工业大学 软件学院, 西安 710072)

(3. 西安现代控制技术研究所, 西安 710065)

摘要: 近年来无人机航拍技术逐步应用于野生动物保护,在很大程度上提高了考察效率。由于航拍图像与地面拍摄图像的特征差异较大,加之野生动物生存环境背景复杂,目前没有通用的方法可直接应用于野生动物航拍图像的检测与统计。本文回顾了智能检测和统计技术近年来的发展,针对无人机航拍野生动物图像的大场景、小目标、多尺度、复杂背景等特点,介绍了无人机航拍动物群数据集的选取与建立方法,以及基于深度学习的检测与统计方法,并进行了深层次地分析,归纳了各类方法的优势和可应用场景,总结了各方法的特点和适用范围,同时针对存在的问题给出了改进方向。

关键词: 无人机航拍;生态保护;深度学习;迁移学习;目标检测;数量统计

中图分类号: S863; V19

文献标识码: A

DOI: 10.16615/j.cnki.1674-8190.2023.01.02

Review of intelligent detection and statistical methods of wild animals in UAV aerial photography

ZHU Ninghua¹, ZHENG Jiangbin², ZHANG Yang³

(1. Engineering Practice Training Center, Northwestern Polytechnical University, Xi'an 710072, China)

(2. School of Software, Northwestern Polytechnical University, Xi'an 710072, China)

(3. Xi'an Modern Control Technology Research Institute, Xi'an 710065, China)

Abstract: Recently, UAV aerial photography technology has been gradually applied to wildlife protection, which has greatly improved the investigation efficiency. Due to the great difference in characteristics between aerial images and ground images, and the complex background of wildlife living environment, there is no general method that can be directly applied to the detection and statistics of UAV aerial wildlife photography. In this paper, firstly, the development of intelligent detection and statistics technology in recent years is reviewed. Then, according to the characteristics of large scene, small target, multi scale and complex background of UAV aerial wildlife photography, the selection and establishment methods of UAV aerial wildlife dataset is introduced, and the detection and statistics methods based on deep learning as well. Finally, the advantages and applicable scenes of these methods are summarized, and the improvement direction is given.

Key words: UAV aerial photography; ecological protection; deep learning; transfer learning; object detection; statistics

收稿日期: 2022-03-24; 修回日期: 2022-07-28

基金项目: 陕西省重点研发计划(2021ZDLGY09-08)

通信作者: 祝宁华, znh@nwpu.edu.cn

引用格式: 祝宁华, 郑江滨, 张阳. 无人机航拍野生动物智能检测与统计方法综述[J]. 航空工程进展, 2023, 14(1): 13-26.

ZHU Ninghua, ZHENG Jiangbin, ZHANG Yang. Review of intelligent detection and statistical methods of wild animals in UAV aerial photography[J]. Advances in Aeronautical Science and Engineering, 2023, 14(1): 13-26. (in Chinese)

0 引言

无人机是航空技术和自动化技术发展的产物,正在向智能化方向发展,尤其是和航拍图像检测技术的结合,使得基于无人机航拍的自动统计技术逐渐成为当今的研究热点之一。在野生动物保护领域,采用无人机航拍、跟踪、检测与统计等技术,拥有人工考察和保护所无法比拟的优势,是近年来兴起的一种全新尝试。无人机航拍受环境和地形等自然条件限制小,机动、灵活,便于获得野生动物活动轨迹和栖息地范围等信息,还能及时了解野生动物的生存现状、迁徙途径,监控是否有野生动物被猎杀等情况^[1],具有监测面积广、多角度覆盖、效率高、成本低等优势。无人机航拍还可实现对野生动物的大尺度、多角度观察,进一步鉴别野生动物种类、性别、健康状况等。总的来说,无人机航拍的应用是野生动物考察和保护中方便快捷的一种全新手段。

然而,从航拍获取的图像中提取检测信息并进行统计,目前还依赖于人工识别,导致航拍的大量图像数据不能直接快速服务于生态保护。其原因主要归结于两点:第一,野生动物栖息在复杂的地形与多样的气候带,无人机航拍获取的野生动物群图像与地面拍摄不同,具有大场景、小目标、多尺度、复杂背景、遮挡重叠等特点,使得要从无人机航拍图像中智能、准确地检测、识别、统计出所关心的动物群的数量难度大;第二,模式识别和大数据等技术的发展,为地面拍摄人群与航拍车辆的检测与统计^[2-4]提供了极大的便利,但在动物群的统计方面应用还很少,这主要是因为动物的特征相对人脸和车辆特征少,还常常受野外背景环境与阳光照射影响而发生变化,并且在无人机航拍中,图像会因高度、视角等因素,出现目标动物特征的损失,使得检测统计难度更大。

随着深度学习技术的深入发展,其与航拍动物群的智能检测结合,大幅推进了无人机航拍用于生态监测的成效。深度学习非常依赖有效的数据集,尽管近年来数据集的研究和基于深度学习的检测方法较多,但是航拍的特殊性约束了这些方法的使用。为此,本文通过对文献资料的深层分析,归纳和分析了有效的航拍动物数据集^[5]的选取与建立方法,以及无人机航拍动物群的智能检测和数量统计方法;在此基础上剖析了各种方法的特点和适用场景,并针对存在的问题给出了改

进方向。

1 航拍动物群数据集的选取与建立

应用深度学习进行无人机航拍动物群数量统计时,动物数据集的选取与建立至关重要,会直接影响统计结果的精确度。而航拍野生动物最难的是获取直接可供网络训练的样本,为此,本节围绕航拍动物群数据集的选取与建立开展研究。

1.1 动物类与航拍类数据集现状和存在的问题

现有的较为常用的动物类数据集有:MSCOCO^[6]、PASCAL VOC^[7]、AWA^[8]、kaggleanimals-10等。这些数据集均为自然场景下的图像,视角以地面正视近景为主。而无人机航拍图像与地面相机等拍摄的自然场景图像不同,无人机从空中往地面拍摄时,其视角通常为斜视和俯视,并且视场广泛,目标体积相对变小,同时目标会呈现多形态的特点。这使得航拍图像具有大视场、小目标的特点,与地面自然场景的近景、清晰大目标的特点完全不同。因此,若直接使用地面自然场景数据集训练的神经网络来检测、统计航拍动物群数量,将会带来很大误差。

航拍图像数据集研究近几年才起步,数据集类别和标注量较少,同时航拍动物领域的数据集有所欠缺,常见的航拍数据集:DOTA^[9]、UCAS-AOD^[10]、NWPU VHR-10^[11]以及RSOD-Dataset^[12]等的研究目标为飞机、轮船、车辆、球场、桥、建筑等,基本上都是应用于交通工具的检测识别。另有少数的数据集,如UC Merced Land-Use^[13]和LEVIR^[14],为土地图像遥感数据集与人类居住环境类地面特征数据集。虽然都是航拍图像,但不包含动物数据特征。因此,这些数据集不能应用于生态监测的动物数量统计研究中。

1.2 航拍动物类数据集的建立方法

用于深度学习的数据集在采集和处理过程中,需满足目标标注量和目标类别的大规模性,来实现充分的网络训练,避免过拟合现象,以及平衡正负样本比例等。同时,图像还需要具备多样性,符合无人机航拍影像特点,与实际情景更贴合,如动物不同视角的不同形态和姿态,不同高度或远近距离产生的多尺度,树木和阴影的遮挡等。但

现有数据集多无法直接使用,为此需要搭建符合应用场景的航拍动物数据集(正样本)。但航拍的有效动物数据集非常有限,经常存在样本特征缺失的情况,如何建立有效的网络训练数据集,本文从样本单一和特征缺失两个方面介绍样本扩充与样本均衡的处理方法。

(1) 样本单一情况下的数据集扩充处理方法

我们所能获取的野生动物航拍图片通常是某一个角度、某一个清晰度、某一个姿态的,如何利用有限样本进行特征扩充,可以采用信息增强处理方法(如图1所示)。图1(a)是一张羊群图像,而实际中,动物的相互位置、背景清晰度、可视尺度都是不同的,为了表达不同特征,在获取的图像样本上,通过缩放、翻转和增加模糊度等方法模拟出更多特征,添加进样本集,如此从单一样本扩充到了四副代表不同特征的图像数据,来实现有效训练数据集的扩充。



(a) 原图 (b) 翻转 (c) 缩小 (d) 添加噪声

图1 数据集的扩充

Fig. 1 Expansion of data set

这类样本增强^[15]处理方法还可采用调节图像亮度、对比度等参数,马赛克拼接,雾化背景环境,旋转图像等手段,利用有限的样本数据表达出更多的样本特征,来扩充有效样本集。

扩充与均衡处理后的样本集增强了神经网络模型的鲁棒性,使数据集包含更全面的语义信息和更丰富的特征信息。其中,调节图片的亮度、饱和度、对比度及色调等,可体现不同光照和阴影下航拍的背景特征,以利于学习在不同的光照条件下或在阴影遮挡环境下动物的特征;通过不同比例缩放、裁剪等处理,可模拟不同位置、不同尺寸、不同分辨率的环境下动物的不同比例尺度;运用旋转的方式,可模拟不同角度拍摄的动物形态;增加随机噪声可以模拟不同分辨率相机拍摄的图像。

(2) 目标特征残缺情况下的数据集扩充处理方法

在航拍视角下,动物目标特征容易缺失。马赛克拼接技术如图2所示,图2(a)是一只大象,它

的主要识别特征如耳朵、鼻子、腿都浸没在水中,信息特征缺失,如何使用该数据样本,可以尝试“移花接木”的马赛克拼接方法,如图2(b)所示,如此处理后该图像信息就完整有效了。

同时找到可选用的自然场景中完整的学习图像,如图2(c)所示,进行局部特征学习,补充进航拍动物群数据集中,用于斜视图的近处大目标训练。再将UC Merced Land-Use和LEVIR中的山地山区类数据集处理为负样本,也补充进所搭建的数据集中,以此来提高航拍动物类数据集的丰富性。



(a) 水中大象 (b) 马赛克拼接 (c) 自然场景图像

图2 马赛克拼接

Fig. 2 Mosaic splice

(3) 动物群样本稀少情况下的数据集扩充处理方法

野生动物种类很多,在搭建数据集过程中,容易出现某类别动物的航拍图像少。在这种正样本较为欠缺的情况下,可以尝试采用直推式迁移学习方法来解决该问题。直推式迁移学习方法^[16]不同于传统的机器学习,它在源域、目标域任务相同的情况下,不要求源域、目标域数据相同,只需相关即可,如此可以选择相关的场景图像来作为补充。相关场景图像如图3所示,目标任务是统计羊群的数量(图3(a)),可以选择同样具有群特征的人群(图3(b))作为补充数据集,再将成熟的、训练好的神经网络模型以及可获得的大规模人群数据集进行相似性训练学习,得到充分的训练样本后,再应用到羊群的统计。这样就解决了动物群样本稀少情况下的数据集扩充问题。



(a) 密集羊群

(b) 密集人群

图3 相关场景图像

Fig. 3 Related scene image

2 无人机航拍动物群的智能检测与统计方法

无人机航拍具有大视场的特点,所拍摄的动物群分布不均,存在不同的密集度,可大体分为稀疏型与密集型两类,如图4所示。



(a) 稀疏型动物群



(b) 密集型动物群

图4 不同密集度动物群

Fig. 4 Different density fauna

若想要机器智能地统计出航拍动物群数量,则需要选用合适的神经网络来学习所搭建的数据集。因此,首先要分析动物群特点并分类,再针对各类的特点采用合适的统计方法。

2.1 稀疏型动物群的智能检测与统计

对于稀疏型的动物群,检测即可统计。但航拍图像具有多视角的特点(如图5所示),俯视拍摄和斜视拍摄的目标形态不同,群体分布也有差异,因此对于稀疏型的动物群,成功检测成为关键。



(a) 俯视拍摄



(b) 斜视拍摄

图5 多视角航拍

Fig. 5 Multi-angle aerial photography

在训练检测网络时,一般需要依据应用场景针对性地挑选合适的训练目标视角形态。然而无论哪种视角,都具有相同的航拍特点,即俯视时由于无人机飞行高度不同,拍摄的目标大小不一,具有多尺度特点,同时高空拍摄形成了小目标特点;斜视时,除了“近大远小”造成的多尺度特点外,远处的动物尺度也都很小,同样存在着小目标特点。因此,除了训练集的选取要适用于应用场景外,还需要研究合适的检测方法,以考虑多尺度的同时提高小目标的检测率和精准度。

基于深度学习的目标检测共有两条发展主线,如图6所示。

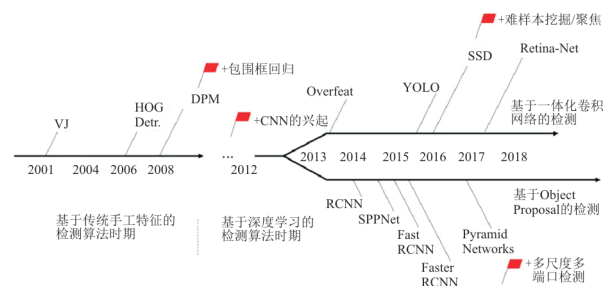


图6 目标检测方法的发展历程图

Fig. 6 Development history of target detection methods

一条是基于候选区域(region-based)的目标检测(也称两阶段(two-stage)方法),主要的算法有R-CNN^[17]系列、SPP-Net(Spatial Pyramid Pooling Network)^[18]等。另外一条是基于回归的目标检测(也称单阶段(one-stage)方法),主要的算法有YOLO(You Only Look Once)系列^[19]、SSD(Single Shot Multibox Detector)^[20]、Retina-Net(Focal Loss)^[21]等。这两类方法在航拍数据集中均有应用,但通过案例应用发现其适应性不同,各方法深层解决问题的思路也不一样。

2.1.1 基于候选区域的目标检测方法

在应用 CNN (卷积神经网络^[22]) (包括 R-CNN、SPP-Net (Spatial Pyramid Pooling Network)、Fast R-CNN^[23]以及 Faster R-CNN^[24]等算法)实现基于候选区域检测的初期,虽然可获得较好的目标检测效果,但这些网络只有单一的预测结构,无法满足航拍动物群这类多尺度、小目标的需求,因此所需的卷积神经网络应为多层输出,并能提高小目标的检测率。主要从以下 3 个方面进行了改进。

(1) 网络结构改进法

为了实现多尺度的目标检测,最直接的方法是在 Faster R-CNN 网络上进行改进。Cai Z 等^[25]利用 CNN 网络中的池化层 (Pooling) 扩大感受野,可与不同尺度映射的特点,在池化层前面的一个卷积层上 (conv4-3、conv5-3、conv6) 引出分支 (如图 7 所示),进行特征提取检测,形成了一种多尺度

分类器,即 MS-CNN (Multi Scale CNN) 模型。

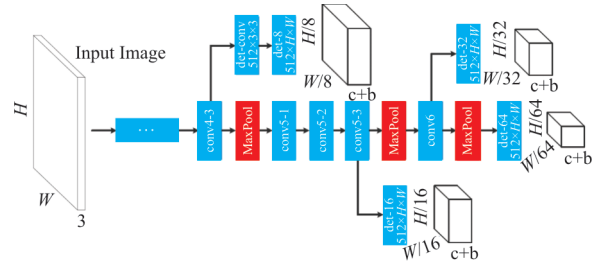


图 7 MS-CNN 网络模型^[25]
Fig. 7 MS-CNN network model^[25]

Li J 等^[26]提出的 SAF R-CNN (Scale-aware Fast R-CNN) 模型 (如图 8 所示) 采用另一种结构改进思路,以 Fast R-CNN 网络为基础,共享底层卷积提取的特征图,设置两个不同尺度的子网络 (双路结构) 对大、小目标进行检测,再通过加权层将两个子网络的结果融合输出。

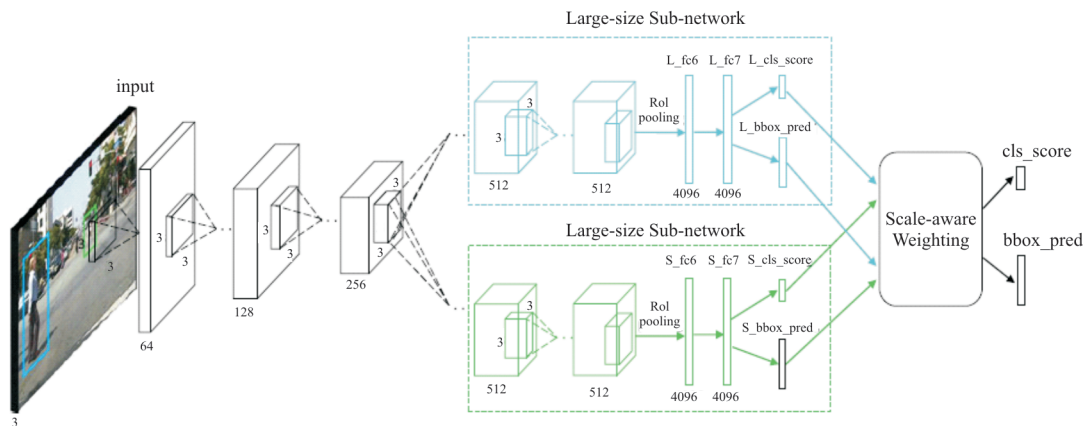


图 8 SAF R-CNN 网络模型^[26]
Fig. 8 SAF R-CNN network model^[26]

文献[25-26]还给出了对多尺度目标检测效果有所提升的验证,并提高了对小目标的检测率。但这些改进都依赖于网络分支的尺度,从而出现了网络无法较为全面地覆盖所有尺度目标、无平滑过渡等问题。对于能够悬停或过顶俯拍的无人机,因拍摄的目标尺度分类相对较少,这种基于候选区域的两个方法就都可行。但对于固定翼无人机,前飞拍摄形成尺度变化大的斜视图,该方法会受到局限。

(2) 特征金字塔方法

为解决航拍中尺度变化大的斜视拍摄问题,可采用处理图像分辨率的方法来实现多尺度特征的提取,但这样又容易造成大量目标特征的流失。

针对该问题,Lin T Y 等^[27]提出了特征金字塔网络 (Feature Pyramid Networks, 简称 FPN) 模型 (如图 9 所示),这是一种自顶向下 (Top-down) 和横向连接的融合方式,不仅利用特征采样映射展现了更高分辨率的特征,还实现了每层特征图的独立预测。

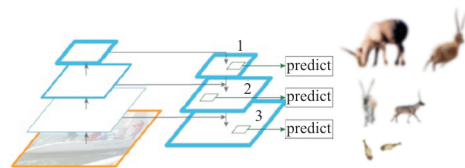


图 9 FPN 模型示意图^[27]
Fig. 9 FPN model schematic diagram^[27]

受到该思想和多尺度分类器的启发,黄继鹏等^[28]进一步对 Faster R-CNN 网络进行改进(如图 10 所示),让不同卷积层提取不同尺度特征图后进行融合,使用最后一层的特征图完成检测,形成了多尺度 Faster R-CNN 网络模型。这样的改进在对小目标检测时,比 Faster R-CNN 网络提升了 5% 的检测精度。类似地,Liu S 等^[29]提出了 PANet(Path Aggregation Network)模型(如图 11 所示),其思想是横向连接合并自底而上路径和自顶向下路径的特征图,提高低层信息(精准定位)的利用率,实现特征增强,还能有效缩短信息传播路径,减少信息丢失。

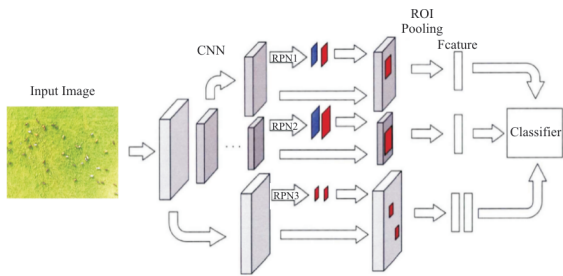


图 10 多尺度 Faster R-CNN 网络模型^[28]
Fig. 10 Multi-scale Faster R-CNN network model^[28]

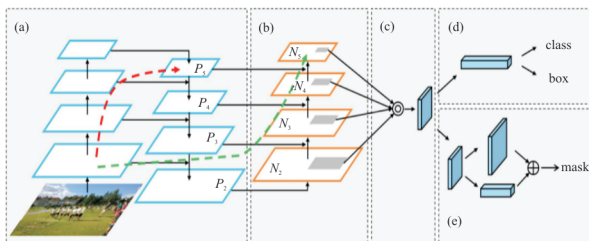
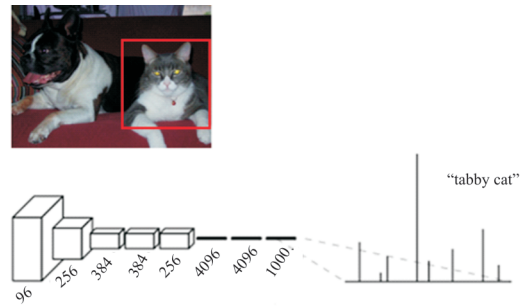


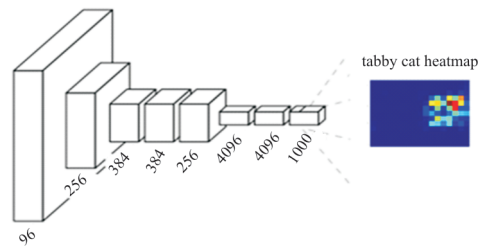
图 11 PANet 网络模型^[29]
Fig. 11 PANet network model^[29]

(3) 语义分割方法

语义分割方法如图 12 所示。上述模型中的全连接层会将输入的图片(二维矩阵)压缩成一维向量(如图 12(a)所示),从而造成空间信息丢失。为此将图片中的语义类别进行分区(即语义分割^[30]),输出具有像素语义标注的分割图,该图同为二维矩阵,能清晰地标识出各类别的空间位置(如图 12(b)所示)。对于语义分割的处理,J. Long 等^[31]提出了将全连接层替换为全卷积层的 FCN(Fully Convolutional Networks)模型,实现了图像级分类到像素级分类的细化,从而让空间位置信息更加清晰直接,还可接收任意尺寸的图像输入,更有利于多目标的精准定位。



(a) 全连接层输出一维向量



(b) 卷积层输出二维矩阵

图 12 语义分割^[30]

Fig. 12 Semantic segmentation^[30]

He K 等^[32]将语义分割与 FPN 融合,提出了 Mask R-CNN 方法(如图 13 所示)。其中,FPN 用于构建多尺度特征;并行 mask 分支(FCN)用于预测分割掩码,同时判断像素类别;ROI Align 用双线性内插法获取像素点上的图像值,实现像素级别的对齐。这种网络设计,将多尺度特征提取与像素级分类集于一体,具有更高的精确度和较强的灵活性,稍加改动就可适用于各项任务。

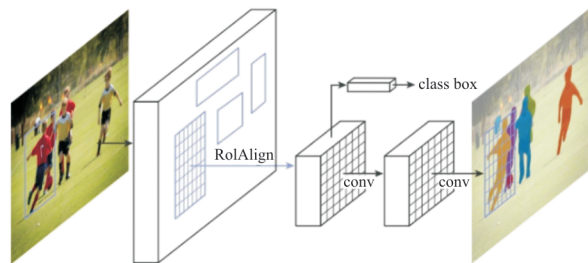


图 13 Mask R-CNN 网络模型^[32]

Fig. 13 Mask R-CNN network model^[32]

由此可见,若将语义分割方法应用于航拍检测,动物与背景就可直观地区分开来,这样在一定程度上就解决了无人机航拍图像中复杂背景干扰和遮挡情况下的检测问题。

综上,无人机航拍俯视图的目标检测一般的方法都可以适应,斜视图的目标检测可以采用特征金字塔方法或语义分割方法进行综合。

2.1.2 基于回归的目标检测方法

基于候选区域的两阶段算法实时性较差,且算法中间需要存储内存大,不适用于轻便型机载仪器设备,因此另一个目标检测的分支——基于回归的单阶段目标检测方法逐渐发展起来。

(1) 基于回归的目标检测方法的起步

2016年,J. Redmon等^[19]借助回归的思想,将分类与回归合并在一个CNN网络中,先利用小网格(grid cell)对目标中心点进行检查,再进行局部预测,获取目标边界框(Bounding Box,简称BBBox)坐标、置信度及类别概率。这种端到端的方式,可看到全局图像,获得上下文信息,使得检测速度更快。这就是最早的YOLO(You Only Look Once: Unified, Real-Time Object Detection)方法。文献[19]还给出了fast-YOLO的构架,检测速度可达155 fps。越快的检测速度,意味着无人机航拍中实时性能越好。

(2) 解决多尺度、多种类、小目标的回归检测方法

在单阶段方法中,实现多尺度的检测,重点在于锚框(anchor)。Liu W等^[20]提出的SSD(Single Shot MultiBox Detector)方法,利用两个卷积层提取不同尺度特征图(金字塔网络结构如图14所示),再通过划分单元格获取不同尺度锚框(如图15所示),从而实现多尺度检测。

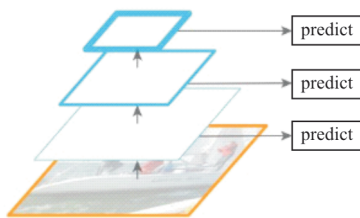
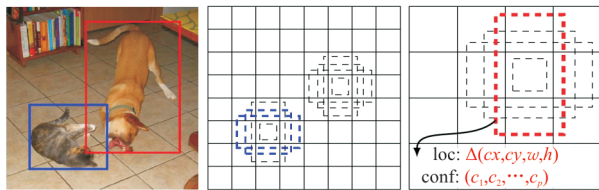


图 14 特征金字塔结构

Fig. 14 Pyramidal feature hierarchy



(a) Image with GT boxes (b) 8×8 feature map (c) 4×4 feature map

图 15 SSD的锚框^[20]

Fig. 15 Anchor for SSD^[20]

所用的锚框尺度是可计算的,比Faster R-CNN中用固定的锚框进行多尺度检测的效果更

好。同时,参数共享的方式让其检测速度比YOLO快,且mAP(mean average precision)指标也有所提升。

SSD方法的思想促使J. Redmon等^[33]提出了以DarkNet-19为骨干的YOLOv2(YOLO9000: Better, Faster, Stronger)方法。采用k-means聚类法产生锚框,使得IOU值更高;运用直接映射来提高网络训练速度;对高分辨率图像进行迁移学习,提升了网络的响应能力;采用BN(Batch Normalization)方法,可加速收敛,降低过拟合。这些改进能有效提升网络训练效果,提高网络运算速度和精度。文献[33]还结合检测与分类联合训练的方法(WordTree方法),让网络可检测出9000种类别的目标。这些方法的引入对多尺度和动物的多类别检测很有效果,能较好地满足航拍较大目标群的检测统计任务,但对于小目标的检测还有待加强。

为此,在YOLOv2方法的特征提取模型(Darknet-19)中融入残差网络(Residual Network,简称ResNet),以降低池化的梯度负面效果,减少特征丢失;再借鉴Faster R-CNN的RPN方法和FPN方法的思想提高网络空间表征能力,提高对小目标检测的精确度,形成了YOLOv3(YOLOv3: An Incremental Improvement)方法^[34]。YOLOv3方法在输出时采用3种尺度(scale)(如图16所示)来完成定位与分类,可更有效地实现多尺度目标检测,尤其是对小目标的检测。

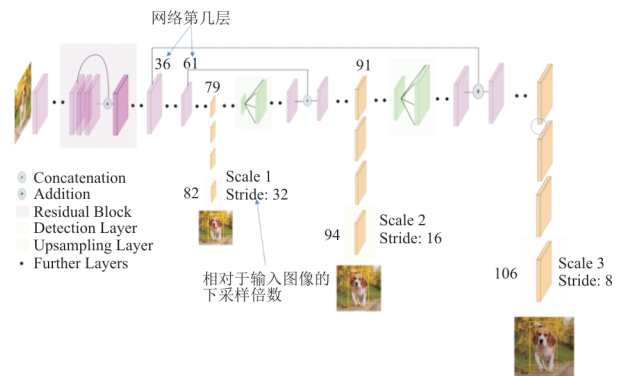


图 16 YOLOv3网络模型^[34]

Fig. 16 YOLOv3 network model^[34]

(3) 解决遮挡、高效的回归检测方法

YOLOv3对小目标的检测效果良好,但在无人机航拍大视场下,动物群总会出现遮挡现象,此时将出现漏检。为此,A. Bochkovskiy等^[35]提出了

YOLOv4 (YOLOv4: Optimal Speed and Accuracy of Object Detection)方法,即在 YOLOv3 的基础上采用 PANet 结构和 SPP 网络 (Spatial Pyramid Pooling Network)来扩大感受野,提高特征提取能力,获取上下文信息,有效提升遮挡目标的检测性能。另外,在训练过程中还采用多种数据增强技术,尤其是创新性地使用了如 Cutout^[36]、Hide and Seek^[37]、Grid Mask^[38]、CutMix^[39]、Mosaic、style transfer GAN^[40]等图像遮挡技术,这也让网络对遮挡目标的检测能力有所提升。

在实际检测中,还经常面临设备的限制,甚至希望在手机上使用。为提高检测效率,发展了 YOLOv5 方法^[41],其使用了 PyTorch 框架,数据集训练方便且快速。采用了自适应锚框计算和图片缩放,提高了网络训练效果。使用两种 CSPNet(跨阶段局部网络)架构^[42],增强了网络学习能力并促进了特征融合。通过控制网络深度和宽度参数,使网络量级可调,方便用于不同监测器和应用场景,具有较好的灵活性,对野外无人机搭载器的实时动物检测统计作业很有帮助。

总之,上述两类检测统计方法适用于稀疏型动物群体。其中,对于航拍俯视图中目标尺度变化较小、目标特征较为统一的,可选用基于候选区域的方法完成检测统计;而对于航拍斜视图中目标尺度变化范围大、目标形态变化多的,可利用回归方法筛选中心点后进行检测统计。同时,在面对复杂背景和遮挡时,可参考语义分割方法与上下文信息提取法实现检测统计。

但面对数量庞大的密集群体,采用这些检测方法会由于群体密度大、小目标多、遮挡严重、分辨率低等因素影响而出现漏检率高、检测精度低等问题。因此,对于无人机航拍密集动物群的统计还需研究密集型统计方法。

2.2 密集型动物群的智能统计

目前对于密集群体的统计研究,大多是针对人群的统计(Crowd Counting)^[43]。而从图像视角来说,密集人群的特点与密集动物群的特点是相似的(如图 17 所示),因此可迁移引用人群统计方法,结合动物群数据集训练,完成无人机航拍密集型动物群的统计。



(a) 密集人群

(b) 密集羊群

图 17 密集群体

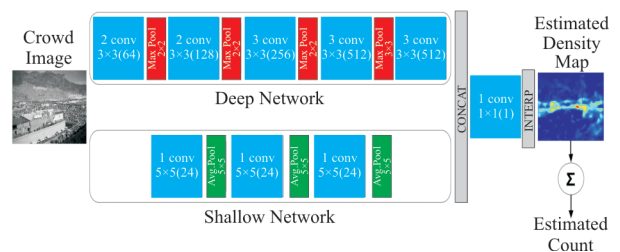
Fig. 17 Dense group

密集群体和稀疏群体的目标特点差异较大,神经网络搭建的思想也不同。目前针对密集群体的统计方法,是吸取 V. Lempitsky 等^[44]提出的密度图估计思想,再仿照 Wang C 等^[45]和 Fu M 等^[46]利用 CNN 网络来实现对人群数量的统计。可见网络结构和密度图的生成是密集型动物群统计的关键。

CNN 与密度图的首次结合是 Cong Z 等^[47]提出 Crowd CNN 模型,是采用相似场景微调法来解决跨场景统计精确度以及密度图质量;采用透视法(视角归一化)来解决多尺度问题。但还是存在航拍和监控位置多变,相似样本难以全面覆盖等问题。因此,后续为提高统计精确度,实现各类密集群体的统计,开展了从网络结构和密度图质量两方面的改进研究。

2.2.1 网络结构的改进

L. Boominathan 等^[48]提出 CrowdNet 方法,用深、浅层网络融合输出的双分支结构(如图 18 所示)来预测密度,深层网络获取高级语义信息(脸部和身体),浅层网络识别底层像素语义信息(远处头斑图案)。该方法不仅延续了 CNN 与密度图融合的思想,还让目标特征在被提取的过程中实现了互补效果,有助于大规模变化群体的统计;采用图像金字塔(如图 19 所示)进行数据增强,帮助识别不同规模群体。

图 18 CrowdNet 网络模型^[48]Fig. 18 Crowdnet network model^[48]

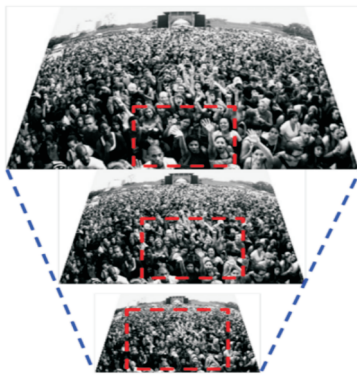


图 19 图像金字塔^[48]
Fig. 19 image pyramid^[48]

同样采用多分支结构提高多尺度统计的还有 Zhang Y 等^[49]提出的 MCNN(multi-column CNN)方法,其由三列结构相同的卷积网络搭建而成,通过改变卷积核尺度和数量,来适应不同尺度目标;利用目标大小与相邻目标中心间距的关系来确定扩散参数,用几何自适应高斯核生成密度图,来实现动物间距较小的群体数量统计。但这种方法在动物间距较远时误差较大。

延续图像金字塔结构,O. R. Daniel 等^[50]搭建了新型统计卷积网络(Counting CNN,简称 CCNN),用 6 个卷积层来解决分割的图像块到密度图的非线性映射,实现密度估计问题向回归问题的转化。将获得的密度图送入 Hydra CNN(尺度感知计数网络,如图 20 所示)中,各分支依据不同尺度金字塔的特定比例,训练相应的特征表达,然后组合各分支的输出,计算密度估计值。由于该方法无需透视图和几何信息,多尺度的比例完全依赖于图像金字塔,因此存在网络性能与统计精确度受图像金字塔影响较大的问题。

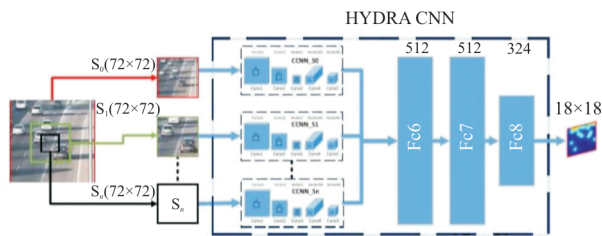


图 20 Hydra CNN 网络模型^[50]
Fig. 20 Hydra CNN network model^[50]

为了减小图像金字塔对网络性能的影响,又能兼顾各种密度的区域,D. B. Sam 等^[51]设计了 Switch 分块处理模块(如图 21 所示),将输入图片分割成 9 个区域块(patch),设定各区域内的人群密

度与尺度特征一致,依次送入 Switch 模块中,该模块主要用来判断分割块的密度与尺度等级,并传送给相应的回归分支中,共有三个回归分支网络,R1~R3,感受野逐渐减小,检测密度也由小变大,R3 适用于高密度像素特征图。这就是 Switch-CNN 网络^[51],然而该网络也存在弊端,如面对密度变化大的场景时,笼统的分类使得统计效果变差。

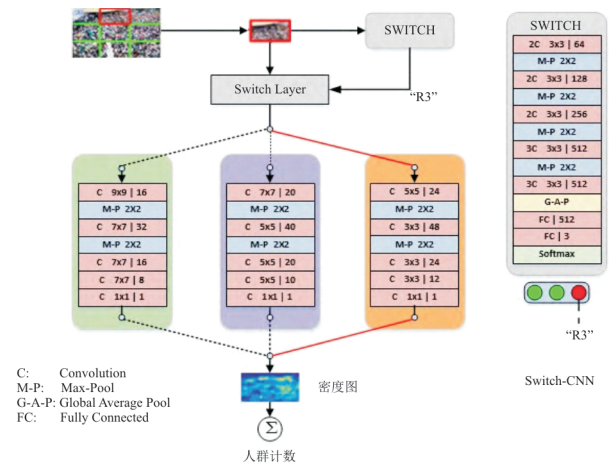


图 21 Switch-CNN 网络模型^[51]
Fig. 21 Switch-CNN network model^[51]

为此,D. B. Sam 等^[52]吸取生长网络(Growing Network)的思想,提出 IG-CNN (incrementally growing CNN)方法,以模型增长(二叉树)的方式让网络自动划分密度等级,其等级数量由测试验证集的误差值(MAE)决定,当误差值不再变化时,模型停止生长。这种改进方法极大地减小了笼统分类带来的误差,让密度图更接近实际群体分布,统计数量更精准。

2.2.2 密度图质量的提升

上述几种方法从等级分类上减小了统计误差,但并没有提高密度图质量,且网络中所采用的像素级欧几里得损失函数还会影响密度图的清晰度。关于密度图质量提升的研究有以下两个方面。

(1) 借鉴对抗网络思想

V. A. Sindagi 等^[53]借鉴生成对抗网络(Generative Adversarial Nets,简称 GAN)思想,提出了新方法 CP-CNN(Contextual Pyramid CNNs)。在网络的 F-CNN(Fusion-CNN)结构中,将预测的密度图与全局、局部上下文信息进行融合,再采用像素级欧氏损失和对抗损失的加权组合进行训练,生成高质量密度图。其中,网络中融入的上下文

金字塔可实现密度等级划分。

S. Zan 等^[54]也在 GAN 思想的启发下提出了 ACSCP (Adversarial Cross-Scale Consistency Pursuit Network) 方法,采用 U-net 所构造的背对背编码器—解码器(back-to-back encoder-decoder)双分支结构,来解决之前多分支网络存在的各分支独立、缺少标准的问题;并用跨尺度一致性正则化方法,使全图与子图之间的残差最小化,同时实现子网络的协作功能,减小跨尺度误差,由此提高输出密度图质量。

(2) 引入空洞卷积和尺度聚合模块

Li Y 等^[55]提出了另一种提高密度图质量的方法——CSRNet (Congested Scene Recognition Networks) 模型。引入六层空洞卷积代替池化层作为后端网络,在保证分辨率不降低的前提下,扩大感受野,提高输出密度图的质量;并验证了在相同大小的感知域中,小卷积核、多卷积层模型的输出质量优于大卷积核、少卷积层模型。

多尺度可以很好地体现群体密度,但密度规模的分类又受到子网络数量的限制。为此,Cao X 等^[56]提出了一种尺度聚合型编码—解码器网络 SANet (Scale Aggregation Network),借鉴 Inception structure 的思想^[57],将尺度聚合模块引入编码器,卷积与反卷积的交替配合组成解码器,其中的每个卷积层都是采用多尺度卷积核来提取特征,这就可以在满足尺度多样性的同时,生成高分辨率和高质量密度图。

密集动物群与密集人群在分布上具有相似性,以上针对人群研究的方法都可迁移应用于密集动物群。

2.3 稀疏和密集混合型动物群的智能统计

上述方法是分别针对稀疏或密集型分布进行的研究,而实际的航拍动物群体是多变的,稀疏型与密集型常常并存,需要进一步研究一种方法来解决这种混合型的统计问题。

(1) 解决密度图融合的问题

Liu J 等^[58]综合低密度群体的检测方法 with 高密度群体的密度图回归方法,提出了 DecideNet 方法,利用高斯卷积层将检测图转换为密度图,再用注意力模块,自适应地分配检测与回归的权重,实现两种密度图的融合,这种合并式的方法可有效提高密度图统计的应用范围。

M. Hossain 等^[59]提出的 SAAN (Scale-Aware Attention Networks) 方法同样引入了注意力机制,但侧重点在于尺度感知。用全局上下文信息和局部细节信息感知图像密度,获得注意力分数;再以该分数为依据,对多种尺度感受野获取的特征图进行加权,得到全局密度图,实现密集群体的数量统计。

(2) 解决背景干扰、透视失真、遮挡情况下的密度图生成问题

上述密度图融合的方法只能应用于群体分布不规律的统计,但大视场图像还易于出现背景噪声、遮挡、透视失真等干扰。

2019 年,Liu N 等^[60]提出了 ADCrowdNet (Attention-injective Deformable Convolutional Network for Crowd) 方法,在注意力机制中融入可变形卷积网络,开创了用独立的模型去生成注意力图像,并过滤背景干扰的先河。该方法由两个级联网络构成,AMG (Attention Map Generator) 用来处理多样化群体分布,生成注意力图像,提取低层次特征,不仅能减少各类干扰,还可提供上下文信息;DME (Density Map Estimator) 为可变形卷积的多尺度网络,用于生成密度图,解决了遮挡、多样化群体分布以及透视失真等问题。

此外,不同高度视角拍摄时,透视失真和个体像素值都会受相机位置的影响。为此,Shi M 等^[61]针对性地提出了透视感知网络 (Perspective-Aware Convolutional Neural Network, 简称 PACNN),将透视信息集成到密度回归中,即用模型直接预测透视图,提供目标尺度变化的信息;再通过透视感知加权编码后,自适应地与多尺度密度图融合输出新的密度图,以此提高对透视失真的鲁棒性。

3 总结与展望

综上所述,基于深度学习的检测和统计方法很多,但各有其特点,且适用的场景也不同(如表 1 所示),可以看出:航拍人群、车辆与航拍动物群都有相似之处,已有针对人群、车辆的方法都可进行借鉴和迁移应用,但还会出现一些特殊干扰,如动物的严重遮挡、地面保护色、动物与物体形态相似、动物随机运动等,对这些问题可以从以下 3 个方面进一步研究:

(1) 对于动物群重叠、植被遮挡等情况,可以

参考加入上下文信息和局部特征提取的方法,再加权融合各层特征图,以获得更精准的候选区域,提高神经网络的检测性能;

(2) 面对低对比度图像,如动物的保护色、动物轮廓与石头、土包相似,光线因素等,可以结合红外热力图实现目标定位,丰富动物数据集的样

本特征;

(3) 由于航拍视角与动物群活动都是动态的,导致时间坐标上的形态特征和位置坐标上的轨迹变化都发生改变,这种情况可以通过光流法等进行处理,来获得并预测动物群的行动轨迹,以减少重复计数和漏检的概率。

表 1 不同航拍动物群体统计方法的优势与应用场景

Table 1 Advantages and application scenarios of different aerial animal population statistics methods

类型	方法	原理	优点	应用场景
稀疏群体	MS-CNN	中间卷积层提取特征	实现多尺度检测	俯视图简单背景动物群体
	SAF R-CNN	双路检测结构	扩大检测尺度范围	俯视图简单背景动物群体
	多尺度 Faster R-CNN	融合特征金字塔与多尺度分类器	提高小目标检测精度	俯视图无遮挡、多尺度动物群体
	PANet	连接自底而上和自顶向下双路径	特征增强,缩短传播路径,减少信息丢失	俯视图多尺度动物群体
	FCN	采用全卷积层进行图像语义分割	输出二维矩阵,标识空间位置信息	俯视图遮挡、复杂背景动物群体
	Mask R-CNN	FPN 提取多尺度特征、FCN 预测分割掩码	解决小目标、复杂背景和遮挡问题的检测	俯视图多尺度、小目标、遮挡、复杂背景动物群体
	YOLOv2	聚类法产生锚框,残差网络直接映射, WordTree 方法	实现多尺度、多类别的检测	斜视图较大、多尺度、多类别动物群体
	YOLOv3	采用三种 scale 输出	实现多尺度检测,提高小目标检测精确度	斜视图无遮挡、多尺度、小目标动物群体
	YOLOv4	使用图像遮挡技术,加入上下文信息	提升遮挡目标的检测性能	斜视图遮挡、多尺度、小目标动物群体
	YOLOv5	采用自适应锚框计算和图片缩放,双 CSP 结构	增强网络学习能力,网络量级灵活可变	斜视图动物群体,可搭载到各类仪器设备
密集群体	CrowdNet	深、浅层双分支网络结构	有助于不同规模群体统计	大规模变化动物群体
	MCNN	利用几何自适应高斯核获取相邻目标间距参数	适应多尺度目标,提高密度图质量	间距较小的动物群体
	IG-CNN	模型增长(二叉树)的方式,网络自动划分密度等级	减小笼统分类,密度图更接近实际群体分布	密度变化大的动物群体
	CP-CNN	对抗网络思想,上下文金字塔划分密度等级	提高密度图质量	密集动物群体
	ACSCP	背对背编码器-解码器双分支结构,对抗损失结构,跨尺度一致性正则化	子网络协作,减小跨尺度误差	密集动物群体
	CSRNet	引入空洞卷积	保障分辨率,扩大感受野	密集动物群体
	SANet	多尺度聚合型编码-解码器	获取多尺度特征,提高密度图分辨率	密集动物群体
	DecideNet	加入高斯卷积层,自适应分配权重	提高密度图应用范围,适用稀疏群体	密集型与稀疏型并存的动物群体
混合群体	SAAN	引入注意力机制	尺度感知图像密集度	密集型与稀疏型并存的动物群体
	ADCrowdNet	注意力机制融入可变形卷积网络	过滤背景干扰、透视失真	带有背景噪声的动物群体
	PACNN	将透视信息集成到密度回归中	解决透视失真问题	多种高度拍摄的动物群体

4 结束语

(1) 直接用于网络训练的航拍野生动物类数据集还远远不够,可以采用样本扩充、样本均衡、样本迁移等方法来建立有效的数据集。

(2) 对于稀疏型的动物群,检测即可统计。航拍为俯视图时,可采用基于候选区域的方法;航拍为斜视图时,可采用基于回归的方法;面对复杂背景和遮挡时,可参考语义分割方法与上下文信息提取法来实现检测统计。

(3) 对于密集型的动物群统计,可以采用密度图估计法。当动物群体规模变化大时,可采用多分支网络结构来获取密度图,还可采用对抗网络和空洞卷积等方法来提高密度图质量。

(4) 对于稀疏和密集混合型动物群的统计,同样可以采用密度图法。通过引入注意力机制法来生成融合型的密度图;对于有背景干扰、透视失真的情况,可采用变形卷积网络或透视失真法来获取密度图。

(5) 目前已有的方法,可以在一定程度上实现航拍野生动物的智能检测和统计,但对于严重遮挡、低对比度地面保护色、动物与物体形态极其相似、动物随机运动等情况,还需要进一步研究,以提高检测与统计的准确率。

随着无人机航拍技术的发展,野生动物保护的考察周期明显缩短,人工成本显著降低。无人机航拍与深度学习技术相结合,又进一步提高了动物群数量检测与统计的效率。无人机目标检测技术在野生动物保护方面具有广阔的研究前景。

参考文献

- [1] 吴昊, 马永胜, 王天晖, 等. 野生动物保护现状及对生态环境的影响[J]. 农村实用技术, 2021(11): 128-130.
WU Hao, MA Yongsheng, WANG Tianhui, et al. Current situation of wildlife protection and its impact on ecological environment[J]. Rural Practical Technology, 2021(11): 128-130. (in Chinese)
- [2] 吴懿. 基于无人机平台的人群计数和异常检测方法研究[D]. 南京: 南京邮电大学, 2021.
WU Yi. Research on crowd counting and anomaly detection method based on UAV platform[D]. Nanjing: Nanjing University of Posts and Telecommunications, 2021. (in Chinese)
- [3] 王朝辉, 王润哲, 郭震冬, 等. 基于YOLO的街景影像中行人车辆检测方法[J]. 北京测绘, 2021, 35(11): 1452-1455.
WANG Zhaohui, WANG Runzhe, GUO Zhendong, et al. The method of pedestrian and vehicle detection in street view image based on YOLO[J]. Beijing Surveying and Mapping, 2021, 35(11): 1452-1455. (in Chinese)
- [4] 孙妍. 基于无人机交通视频的车辆检测与跟踪算法研究[D]. 秦皇岛: 燕山大学, 2021.
SUN Yan. Research on vehicle detection and tracking algorithm based on UAV traffic video[D]. Qinhuangdao: Yanshan University, 2021. (in Chinese)
- [5] 陈建促. 基于视频的野生动物目标检测算法研究[D]. 重庆: 重庆理工大学, 2019.
CHEN Jiancu. Research on video-based wild animal object detection algorithm[D]. Chongqing: Chongqing University of Technology, 2019. (in Chinese)
- [6] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft coco: common objects in context[C]// 2014 European Conference on Computer Vision. Berlin: Springer, 2014: 740-755.
- [7] EVERINGHAM M, ESLAMI S M, GOOL L V, et al. The pascal visual object classes challenge: a retrospective[J]. International Journal of Computer Vision, 2015, 111(1): 98-136.
- [8] XIAN Y, LAMPERT C H, SCHIELE B, et al. Zero-shot learning—a comprehensive evaluation of the good, the bad and the ugly[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 99: 1-11.
- [9] XIA G, BAI X, DING J, et al. DOTA: a large-scale dataset for object detection in aerial images[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 3974-3988.
- [10] ZHU H, CHEN X, DAI W, et al. Orientation robust object detection in aerial images using deep convolutional neural network[C]// IEEE International Conference on Image Processing. US: IEEE, 2015: 3735-3739.
- [11] CHENG G, HAN J, ZHOU P, et al. Multi-class geospatial object detection and geographic image classification based on collection of part detectors[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2014, 115: 119-132.
- [12] XIAO Z, LIU Q, TANG G, et al. Elliptic fourier transformation-based histograms of oriented gradients for rotationally invariant object detection in remote-sensing images[J]. International Journal of Remote Sensing, 2015, 36(2): 618-644.
- [13] YANG Y, NEWSAM S. Bag-of-visual-words and spatial extensions for land-use classification[C]// Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems. New York: Association for Computing Machinery, 2010: 270-279.
- [14] ZOU Zhengxia, SHI Zhenwei. Random access memories: a new paradigm for target detection in high resolution aerial remote sensing images[J]. IEEE Transactions on Image Processing, 2018, 27(3): 1100-1111.
- [15] 王俊杰. 应用于深度学习的样本采集自动标注及数据增强

- 系统[D]. 济南: 山东大学, 2021.
- WANG Junjie. Sample acquisition automatic labeling and data enhancement system applied to deep learning[D]. Jinan: Shandong University, 2021. (in Chinese)
- [16] 张超. 迁移学习分类技术及应用研究[D]. 西安: 西安电子科技大学, 2018.
- ZHANG Chao. Research and application of transfer learning [D]. Xi'an: Xidian University, 2018. (in Chinese)
- [17] 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, OH, USA: IEEE, 2013: 580-587.
- [18] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 37(9): 1904-1916.
- [19] REDMON J, DIVVALA S, GIRSHICK R, et al. You Only Look Once: unified, real-time object detection[C]// IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 779-788.
- [20] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]// The 14th European Conference on Computer Vision. [S. l.]: Springer, 2016: 21-37.
- [21] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]// IEEE Transactions on Pattern Analysis & Machine Intelligence. [S. l.]: IEEE, 2017: 2999-3007.
- [22] 李炳臻, 刘克, 顾佼佼, 等. 卷积神经网络研究综述[J]. 计算机时代, 2021(4): 8-12, 17.
- LI Bingzhen, LIU Ke, GU Jiaojiao, et al. Review of the researches on convolutional neural networks [J]. Computer Era, 2021(4): 8-12, 17. (in Chinese)
- [23] GIRSHICK R. Fast R-CNN[C]// IEEE International Conference on Computer Vision Santiago. [S. l.]: IEEE, 2015: 1440-1448.
- [24] 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 & Machine Intelligence, 2017, 39(6): 1137-1149.
- [25] CAI Z, FAN Q, FE R S, et al. A unified multi-scale deep convolutional neural network for fast object detection [J]. Springer International Publishing, 2016, 125: 1-15.
- [26] LI J, LIANG X, SHEN S M, et al. Scale-aware fast R-CNN for pedestrian detection [J]. IEEE Transactions on Multimedia, 2015, 99: 1-9.
- [27] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE, 2017: 936-944.
- [28] 黄继鹏, 史颖欢, 高阳. 面向小目标的多尺度 Faster-RCNN 检测算法[J]. 计算机研究与发展, 2019, 56(2): 319-327.
- HUANG Jipeng, SHI Yinghuan, GAO Yang. Multi-scale Faster-RCNN algorithm for small objects detection [J]. Journal of Computer Research and Development, 2019, 56(2): 319-327. (in Chinese)
- [29] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation [C] // IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE, 2018: 8759-8768.
- [30] 程擎, 范满, 李彦冬, 等. 无人机航拍图像语义分割研究综述[J]. 计算机工程与应用, 2021, 57(19): 57-69.
- CHENG Qing, FAN Man, LI Yandong, et al. Review on semantic segmentation of UAV aerial images[J]. Computer Engineering and Applications, 2021, 57(19): 57-69. (in Chinese)
- [31] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640-651.
- [32] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN [C]// IEEE International Conference on Computer Vision. [S. l.]: IEEE, 2017: 386-397.
- [33] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C] // IEEE Conference on Computer Vision and Pattern Recognition. US: IEEE, 2017: 6517-6525.
- [34] REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-21)[2022-03-24]. https://www.researchgate.net/publication/324387691_YOLOv3_An_Incremental_Improvement.
- [35] BOCHKOVSKIY A, WANG C Y, LIAO H M. YOLOv4: optimal speed and accuracy of object detection [EB/OL]. (2020-04-01)[2022-03-24]. https://www.researchgate.net/publication/340883401_YOLOv4_Optimal_Speed_and_Accuracy_of_Object_Detection.
- [36] DEVRIES T, TAYLOR G W. Improved regularization of convolutional neural networks with cutout[EB/OL]. (2017-08-09)[2022-03-24]. https://www.researchgate.net/publication/319135616_Improved_Regularization_of_Convolutional_Neural_Networks_with_Cutout.
- [37] KRISHNA K S, HAO Y, ARON S, et al. Hide-and-peek: a data augmentation technique for weakly-supervised localization and beyond[EB/OL]. [2022-03-24]. <https://arxiv.org/pdf/1811.02545.pdf>.
- [38] CHEN P, LIU S, ZHAO H, et al. GridMask data augmentation [EB/OL]. (2020-01-20)[2022-03-24]. https://www.researchgate.net/publication/338569713_GridMask_Data_Augmentation.
- [39] YUN S, HAN D, CHUN S, et al. CutMix: regularization strategy to train strong classifiers with localizable features [C]// Proceedings of the IEEE International Conference on Computer Vision. US: IEEE, 2019: 6023-6032.
- [40] ROBERT G, PATRICIA R, CLAUDIO M, et al. ImageNet-trained cnns are biased towards texture, increasing shape bias improves accuracy and robustness[C]// Interna-

- tional Conference on Learning Representations. [S. l. : s. n.], 2019: 1-17.
- [41] 江磊, 崔艳荣. 基于YOLOv5的小目标检测[J]. 电脑知识与技术(学术版), 2021, 17(26): 131-133.
JIANG Lei, CUI Yanrong. Small target detection based on YOLOv5[J]. Computer Knowledge and Technology (Academic Edition), 2021, 17(26): 131-133. (in Chinese)
- [42] WANG C Y, LIAO H, WU Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. [S. l.]: IEEE, 2020: 1571-1580.
- [43] 钮嘉铭, 杨宇. 基于CNN的人群计数与密度估计研究综述[J]. 软件导刊, 2021, 20(8): 247-252.
NIU Jiaming, YANG Yu. Overview of crowd counting and density estimation based on CNN [J]. Software Guide, 2021, 20(8): 247-252. (in Chinese)
- [44] LEMPITSKY V, ZISSERMAN A. Learning to count objects in images[C]// The 24th Annual Conference on Neural Information Processing Systems 2010. US: Curran Associates Inc. , 2010: 1159-1163.
- [45] WANG C, ZHANG H, YANG L, et al. Deep people counting in extremely dense crowds[C]// The 23rd ACM International Conference. Shanghai: ACM, 2015: 1299-1302.
- [46] FU M, XU P, LI X, et al. Fast crowd density estimation with convolutional neural networks[J]. Engineering Applications of Artificial Intelligence, 2015, 43(11): 81-88.
- [47] CONG Z, LI H, WANG X, et al. Cross-scene crowd counting via deep convolutional neural networks[C]// IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015: 833-841.
- [48] BOOMINATHAN L, KRUTHIVENTI S, BABU R V. CrowdNet: a deep convolutional network for dense crowd counting [C]// The 24th ACM Int Conf on Multimedia. New York: ACM, 2016: 640-644.
- [49] ZHANG Y, ZHOU D, CHEN S, et al. Single-image crowd counting via multi-column convolutional neural network[C]// IEEE Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE, 2016: 589-597.
- [50] DANIEL O R, LOPEZ-SASTRE R J. Towards perspective-free object counting with deep learning[C]// European Conference on Computer Vision. Boston: Springer, 2016: 615-629.
- [51] SAM D B, SURYA S, BABU R V. Switching convolutional neural network for crowd counting[C]// Computer Vision & Pattern Recognition. Honolulu: IEEE, 2017: 4031-4039.
- [52] SAM D B, SAJIAN N N, BABU R V. Divide and grow: capturing huge diversity in crowd images with incrementally growing CNN[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 3618-3626.
- [53] SINDAGI V A, PATEL V M. Generating high-quality crowd density maps using contextual pyramid CNNs[C]// IEEE International Conference on Computer Vision. Berlin: IEEE, 2017: 1-11.
- [54] ZAN S, YI X, NI B, et al. Crowd counting via adversarial cross-scale consistency pursuit [C] // IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 5245-5254.
- [55] LI Y, ZHANG X, CHEN D. CSRNet: dilated convolutional neural networks for understanding the highly congested scenes[C]// IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 1091-1100.
- [56] CAO X, WANG Z, ZHAO Y, et al. Scale aggregation network for accurate and efficient crowd counting[C]// European Conference on Computer Vision. US: Springer, 2018: 734-750.
- [57] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C] // IEEE Conference on Computer Vision and Pattern Recognition. Berlin: IEEE, 2015: 1-9.
- [58] LIU J, GAO C, MENG D, et al. DecideNet: counting varying density crowds through attention guided detection and density estimation[C]// IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 5197-5206.
- [59] HOSSAIN M, HOSSEINZADEH M, CHANDA O, et al. Crowd counting using scale-aware attention networks [C]// IEEE Winter Conference on Applications of Computer Vision. Waikoloa: IEEE, 2019: 1280-1288.
- [60] LIU N, LONG Y, ZOU C, et al. ADCrowdNet: an attention-injective deformable convolutional network for crowd understanding [C] // IEEE/CVF Conference on Computer Vision and Pattern Recognition. Berlin: IEEE, 2019: 3225-3234.
- [61] SHI M, YANG Z, XU C, et al. Revisiting perspective information for efficient crowd counting [C] // IEEE/CVF Conference on Computer Vision and Pattern Recognition. Berlin: IEEE. 2019: 7271-7280.

作者简介:

祝宁华(1994—),女,硕士,研究实习员。主要研究方向:无人智能控制,图像识别。

郑江滨(1971—),男,博士,教授。主要研究方向:计算机视觉与虚拟现实,人工智能和图像处理,嵌入式信息计算技术等。

张阳(1991—),男,博士,工程师。主要研究方向:飞行器设计,空气动力学。

(编辑:马文静)