引用格式：某，某某，，等．中文题名中文题名中文题名中文题名[J]．航空学报, 2019, 40(X):XXXXX.ZHANG M, LYU M M, ZHU M,et al. Title title title title title title[J].Acta Aeronautica et Astronautica Sinica,2019,40(X):XXXXX(in Chinese). doi:

修改稿

基于深度学习的无人机航拍影像目标检测研究综述

摘　要：目标检测是提高无人机感知能力的关键技术之一，其研究对于无人机的应用有着重要意义。与基于手工特征的传统方法相比，基于卷积神经网络的深度学习方法具有强大的特征学习和表达能力，成为目前目标检测任务的主流算法。近年来，目标检测技术已经在自然场景图像上取得了一系列突破性进展，在无人机领域的研究也逐渐成为热点。首先系统阐述了基于深度学习的目标检测算法的研究进展，并总结了相关算法的优缺点。对常见的航空影像数据集进行了梳理并介绍了迁移学习的方法；从无人机影像背景复杂、目标较小、视场大、目标具有旋转性的特点出发，对无人机目标检测在近期的研究进行了归纳和分析。最后讨论了存在的问题和未来可能的发展方向。

关键词：目标检测；无人机影像；卷积神经网络；计算机视觉；深度学习；迁移学习

中图分类号：V279;TP181　**文献标识码：**A　**文章编号：**1000-6893（2019）XX-XXXXX-XX

无人机具有成本低、灵活性高、操作简单、体积小等优点，可以弥补卫星和载人航空遥感技术的不足，催生了更加多元化的应用场景。无人机影像的智能化分析处理不仅可以快速高效地提取地物信息，还能拓展无人机的场景理解能力。目标检测技术能够自动化识别和定位图像中目标，这种技术可以增强弱人机交互下无人机的感知功能[1]，为其自主探测和飞行提供基础的技术支持。

无人机航拍由于成像视角不同于自然场景图像，一般有以下特点：

1）背景复杂。无人机的拍摄视角和更大的幅宽可以获取到更丰富的地物信息，但这种无法突显目标的拍摄方式也给检测任务带来了噪声干扰。同时，由于无人机的飞行高度相对较低，空域环境较为复杂，因此遮挡现象在无人机航拍影像中较为常见，导致无人机对目标的观测往往具有不连续性和不完整性。

2）小目标。无人机图像中的目标尺度变化大，且小目标的比例远高于自然场景图像。

3）大视场。大幅宽下的影像往往包含着稀疏不均的目标分布，搜索目标需要花费更高的成本。

4）旋转。目标的朝向是任意的，同一类别目标的朝向角度也不相同。

目前的目标检测任务主要面向自然场景图像，在相应的应用问题，如人脸识别、行人检测等领域已经相对成熟[2]。但由于成像视角不同且

收稿日期：2019-xx-xx；退修日期：2019-xx-xx；录用日期：2019-xx-xx；网络出版时间：
网络出版地址：
基金项目：四川省科技计划重点研发项目（2019YFG0308）；四川省 2018-2020 年高等教育人才培养质量和教学改革项目（JG2018-325）；四川省大学生创新创业训练计划项目（S202010624029）；中国民用航空飞行学院科研项目（J2008-78）；中国民航飞行学院面上项目（J2020-078）
*通讯作者. qurk0226@qq.com

缺乏有效样本的训练，直接将现有算法应用于无人机领域效果较差。因此，研究适用于无人机的目标检测算法对其应用有着重大意义。

近年来，基于深度学习的无人机目标检测方法的研究在学术界备受关注，相关文献逐渐增多但对整体研究现状总结的综述性文献较少。本文第一章阐述了基于深度学习的目标检测算法研究进展；第二章第一节介绍了现有的航空影像数据集，第二节讨论了迁移学习的主要方式及其效果，第三节从无人机影像的特点出发，分析了相关的改进算法；第三章总结了现有无人机目标检测算法的不足并对未来的发展方向进行了展望。

1 基于深度学习的目标检测算法

卷积神经网络（Convolutional Neural Networks，CNN）是最具代表的深度学习算法之一，它的网络为权值共享结构，与输入图像的契合度高，可以更好的完成图像特征的提取和分类工作。

1.1 计算机视觉中的卷积神经网络

卷积神经网络通常由输入层、卷积层、池化层、全连接层和输出层组成，如图1所示。

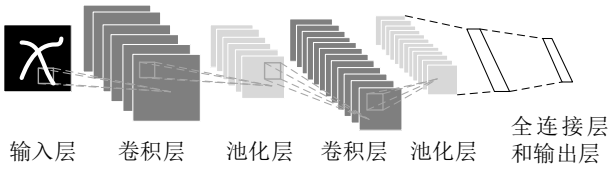

图 1 卷积神经网络的典型结构

Fig. 1 Typical structure of convolutional neural network

卷积神经网络的训练目的是损失函数（Loss Function）的最小化，即预测值Y与真实标记值Y_t的误差最小。对于分类问题和回归问题，交叉熵（Cross Entropy Error，CEE）和均方差（Mean Squared Error，MSE）损失函数是常见的损失函数：

$$CEE = -\frac{1}{n}\sum_{i=1}^{N}[Y\ln Y_t + (1-Y)\ln(1-Y_t)] \tag{1}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{N}(Y_t - Y)^2 \tag{2}$$

预测值与真实标记值间的误差通过反向传播逐层更新网络的各层的参数。对于卷积层X，权值向量$\boldsymbol{W}$和偏移向量$\boldsymbol{b}$的梯度为：

$$\frac{\partial \mathrm{E}(\boldsymbol{W},\boldsymbol{b})}{\partial \boldsymbol{W}_i} = \frac{\partial \mathrm{E}(\boldsymbol{W},\boldsymbol{b})}{\partial X_i}\frac{\partial X_i}{\partial \boldsymbol{W}_i} \tag{3}$$

$$\frac{\partial \mathrm{E}(\boldsymbol{W},\boldsymbol{b})}{\partial \boldsymbol{b}_i} = \frac{\partial \mathrm{E}(\boldsymbol{W},\boldsymbol{b})}{\partial X_i}\frac{dX_i}{d\boldsymbol{b}_i} \tag{4}$$

早期的目标检测算法利用人工几何特征来实现特征表达，如SIFT[3]、HOG[4]等。1998年，Lecun等[5]提出相对简单的卷积神经网络模型LeNet-5，利用神经网络抽取的到的特征替代人工提取的特征。AlexNet[6]在网络规模上变得更宽更深，使用了ImageNet[7]提供的大规模数据集和多GPU来训练。Network In Network[8]和ZFNet[9]的成功证明了对卷积神经网络结构的改动是可以大胆尝试的，表1列出了一些卷积神经网络的模型参数。

2014年Simonyan K等[10]提出了VGG网络，证明了网络在一定程度的加深有助于提升其图像分类性能。2015年Szegedy C等[11]从网络结构优化的角度出发，提出了由Inception模块构建的GoogLeNet。为了解决因网络深度增加而带来的梯度弥散问题，2016年He K等[12]提出了ResNet网络结构。2017年Huang G等[13]提出了密集连接的卷积神经网络模型DenseNet。2017年Hu J等[14]提出了SENet，从特征通道间的关系出发来提升整个网络的性能。

表 1 典型卷积神经网络的模型参数对比

Table 1 Comparison of model parameters of typical convolutional neural networks

卷积神经网络	参数量/亿	卷积层数	卷积核大小	全连接层数	TOP 5 错误率/%	年份
AlexNet	0.60	5	11，5，3	3	15.32	2012
VGG	1.38	13	3	3	7.32	2014
GoogLeNet	0.05	21	7，5，3，1	1	6.67	2014
ResNet152	0.60	151	7，3，1	1	3.57%	2015
DenseNet201	0.20	200	7，3，1	1	-	2017
SE ResNet50	0.28	49	7，3，1	17	-	2017

1.2 基于两阶段方法的目标检测

基于两阶段方法的目标检测又被称为基于候选区域（Region Proposal）的方法。图2给出了常见的两阶段目标检测算法结构。

2014年Girshick R等[15]尝试在AlexNet的基础上将Region Proposal和CNN结合起来，提出了检测性能有着大幅提升的R-CNN算法He K等[16]在卷积神经网络中使用了空间金字塔池化（Spatial Pyramid Pooling，SPP）模块，解决了输入固定大小图片的限制且避免了重复提取图像特征。2015年，Grishick R等[17]在R-CNN和SPP-Net算法结构的基础上的提出了Fast R-CNN，实现了端到端的检测。Ren S等[18]提出了Faster R-CNN改进了候选区域的生成方法，使用候选区域生成网络（Region Proposal Network，RPN）代替了Selective Search算法，实现了整个网络共享卷积特征，进一步提高了检测速度。

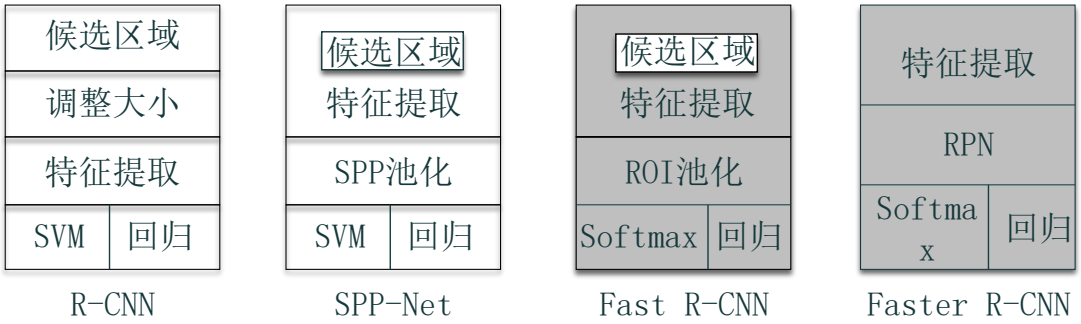

图 2 两阶段目标检测算法结构图

Fig. 2 Structure of two-stage object detection algorithm

1.3 基于单阶段方法的目标检测

单阶段方法只需对图片处理一次就能获得目标的分类和位置信息，运行速度较快，可以应用于对实时性要求较高的场景。

2015年Redmon J等[19]提出YOLO（You Only Look Once）算法，使用一个单独的神经网络，完成从图片输入到目标位置和类别信息的输出。2016年Liu W等[20]提出了SSD（Single Shot Multi-Box Detector）算法，进行多尺度特征的提取。图3为YOLO和SSD算法结构对比。

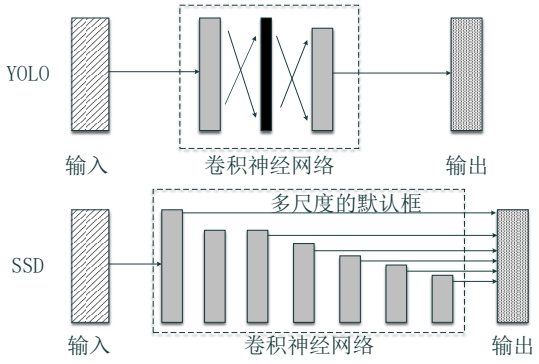

图 3 YOLO 和 SSD 算法结构对比

Fig. 3 A comparison between YOLO and SSD algorithm structure

2 无人机目标检测研究进展

无人机航拍因成像幅宽大、不受地理条件约束等优点，在工业巡检、交通管理、应急救援、安防等领域发挥着重要的作用。目标检测作为计算机视觉中的一项重要技术，正不断提高无人机感知能力和图像数据分析能力，促进无人机航拍在民用和军事领域转化出更多的应用。近年来基于深度学习的目标检测方法在无人机领域也取得了一些重要的研究进展。

2.1 航空影像数据集

无人机航拍图像自身有着显著的特点，使用自然场景图像数据集（如MSCOCO[21]和VOC[22]）来完成前者目标检测的训练任务难以取得令人满意的效果。

一些研究针对这一问题提出了航空影像数据集，相关图像数据集对比如表2所示。

UC Merced Land-Use[23]是一个用于土地利用研究的可见光遥感图像数据集，图片取自 USGS National Map Urban Area Imagery系列。NWPU VHR-10[24]数据集有10个类别的对象，这些图像

是从Google Earth和Vaihingen数据集裁剪而来的，并且由专家进行了类别标注。VEDAI[25]（Vehicle Detection in Aerial Imagery）数据集用于多种类车辆检测任务，该数据集的航空图像取自犹他州AGRC。UCAS-AOD[26]数据集用于航空图像下车辆和飞机的目标检测，图像采集于Google Earth，车辆数据集共210张图片。DOTA[27]数据集是一个用于航空图像中目标检测的大型数据集，图像的采集来自不同的传感器和平台，包含了不同比例、方向和形状的目标对象。

使用无人机作为拍摄平台而制作的数据集出现较晚。Stanford Drone Dataset[28]在2016年被提出，由无人机拍摄而制作的图像/视频数据集，为了提高挑战性，影像采集于校园中目标较为拥挤的场景。Okutama Action Dataset[29]同样使用无人机拍摄，是一个用于检测人体动作的视频数据集。CARPK[30]是一个停车场数据集，包含近9万辆车，用于无人机对车辆的检测和计数任务。VisDrone[31]数据集包含了不同天气和光照条件下的288个视频和1万余张图像，用于无人机图像目标检测、视频目标检测、单目标跟踪和多目标跟踪四种任务挑战。DroneVehicle[32]是一个面向车辆检测和车辆计数任务的数据集，包含了RGB图像和红外图像，采集还涵盖了昼夜时段以及目标的遮挡和尺度变化。

数据集对于深度学习来说有着至关重要的作用，然而对于无人机影像的目标检测任务，目前缺少ImageNet、MSCOCO和VOC这些类型的数据集。现有的航空影像数据集类别数量和标注的目标数量较少，大多数数据集关注的类别为车辆、飞机、船舶和建筑，与ImageNet多达200个目标类别以及近90万个标注数量相比，这些难以反映真实世界的复杂程度；此外，同一目标的尺度变化和旋转特性不够丰富，相比于卫星和传统航空遥感平台，无人机有着较高的灵活性，目标较为丰富的变化才能贴近无人机的实际航拍场景。因此，在采集和制作无人机影像数据集时应作如下考虑：

1）数据集具有较大的规模。目标类别、目标标注在数量上要足以支撑基于深度学习的方法，类别的选择除了满足实际应用还要平衡正负样本的比率，从而进一步提高无人机影像目标检测的技术水平。

2）数据集应具有较好的泛化性，淡化数据集本身的特征[33]。使用不同传感器进行航拍，保证相同类别目标具有不同的分辨率；拍摄时段和天气应多样化，从而确保影像信息之间具有偏差更加贴合实际。

3）数据集应充分表征无人机影像的特点。背景信息足够丰富，不能刻意排除模糊、有遮挡或难以分辨的目标；采集数据时应注意同类目标的多样性和相似性，包括尺度和形状的变化、旋转特性等。

表 2 不同航空图像数据集对比

Table 2 Comparison of different aerial image datasets

数据集	发布时间	图片数量	类别数量	图片宽度	包围盒	目标总数
UC Merced Land-Use	2010	2100	21	256	-	-
NWPU VHR-10	2014	800	10	~1000	水平	3775
VEDAI	2015	1210	9	1024	有向	3640
UCAS-AOD	2015	910	2	1280	水平	6029
DOTA-v1.0	2017	2806	15	800-4000	有向	188282
DOTA-v1.5	2019	2806	16	800-4000	有向	400000
CARPK	2017	1448	1	1280	水平	89777
VisDrone	2019	10209	10	2000	水平	2600000
DroneVehicle	2020	31064	5	840	有向	441642

2.2 迁移学习

使用迁移学习从较为成熟的大规模数据集进行相似性学习后应用于新领域是另一种解决问题的有效方案。

按照文献[34]对迁移学习的定义，可以将自然场景图像的数据集定义为源域数据，无人机影像数据集定义为目标域数据，其对应的学习任务

分别为源域任务和目标区任务。迁移学习可分为如表3所示的3种类型。自然图像上训练好的模型应用于无人机影像主要的任务为不同领域之间知识的迁移，其源域和目标域数据有相关性，学习的任务相同。

表 3 迁移学习的类型

Table 3 Types of transfer learning

学习类型	源域和目标域数据	源域和目标域数据任务
传统的机器学习	相同	相同
迁移学习	不同但相关	相同
	相同	不同但相关
	不同但相关	不同但相关

对于深度学习而言，迁移学习的过程为利用相关领域的知识通过深度神经网络进行目标域中模型的训练。按照学习的方法不同可将深度迁移学习分为三类：基于实例的迁移学习、基于特征的迁移学习和基于模型的迁移学习。

基于实例的迁移学习通过调整权重来增加度目标域有用样本的权重，以此补充目标域训练数据，这种方法目前使用较少；基于特征的迁移学习通过挖掘源域数据中可以覆盖目标域数据的部分，实现不同特征空间之间的知识迁移，在目标检测任务比较常见的是特征提取器的迁移；基于模型的迁移学习需要找到源域和目标域模型之间可以共享的参数或知识。

迁移学习在目标检测领域取得了一些不错的成果。Pan B等[35]提出了一种基于迁移学习和几何特征约束的级联CNN网络模型，在较少的遥感图像样本下实现了高精度的检测；Yuan G L等[36]在大中小三种规模的数据集之间使用两次迁移学习，实现了较高精度的夜间航拍车辆检测；Wang Z L等[37]使用深度迁移学习的方法成功地将在规模的仿真SAR图像数据集上学习到的知识迁移到实测SAR图像上，提高了数据稀缺下目标识别的准确率。但迁移学习的顺利进行是有条件限制的，需要保证源域和目标域之间具有共同点，具有一定的相似性和关联性[38][39]。

2.3 无人机目标检测算法

计算机视觉领域中基于深度学习的目标检测方法在自然场景领域取得了巨大的成功，这对于无人机目标检测任务是值得借鉴和参考的，很多国内外研究提出了效果显著的改进算法。本文将从无人机影像的四个特点出发，分析比较了一些具有代表性的方法。

2.3.1 无人机影像中的复杂背景问题

无人机影像中目标密集区域往往存在着大量形似物体容易导致目标混淆，从而导致检测中的漏检或误报增加。此外，无人机影像背景中大量噪声信息，还会使目标被弱化或遮挡，难以被连续和完整地检测。

近年来，国内外提出了一些效果显著的改进算法来抑制影响背景中的噪声信息。Audebert N等[40]在航拍图像中利用深度全卷积网络对车辆精确分割，通过连通分量的提取实现车辆检测，证明了航拍图像中语义分割和目标检测的结合，可以提高检测性能，尤其是在目标边界信息的提取上。。Mask R-CNN[41]、MaskLab[42]、HTC[43]等算法兼顾了分割和目标检测，并在两个任务上都取得了很好的效果。受此启发，Li C等[44]构建了一个语义分割指导下的RPN（semantic segmentation-guided RPN，sRPN）模块来抑制航拍图像中的背景杂波。这个模块将多层金字塔特征集成为一个新的特征后，进行空洞空间金字塔池化（Atrous Spatial Pyramid Pooling，ASPP[45]）和卷积运算，得到掩膜和语义特征，它们分别可以帮助指导RPN和得到更精准的回归结果。sRPN对检测精度有一定的提升作用，但获取的特征在尺度上较为稀疏，上下文信息联系不够紧密，容易造成信息丢失。文献[46]使用改进的多尺度空洞卷积来提取特征，扩大了对特征的感受野，提高了复杂背景下和有遮挡的目标检测效果。

受人类感知机制的启发，注意力机制被应用于深度学习中，其目的是聚焦并选择对任务有用的信息。注意力机制可以使检测器对目标进行“有区别”地检测，提升网络模型运行的效率。Yang X等[47]将注意力机制引入目标检测中，提出了SCRDet。使用一个有监督的多维注意力网络（Multi-Dimensional Attention leaner，MDA-NET）来突出目标特征，弱化背景特征。该网络由像素注意力和通道注意力网络两部分组成，像素注意力网络将初始的特征图进行卷积得到带有

前景和背景分数的显著图；通道注意力网络使用SENet为各特征通道的重要性赋权。图4为sRPN和MDA-NET算法结构。

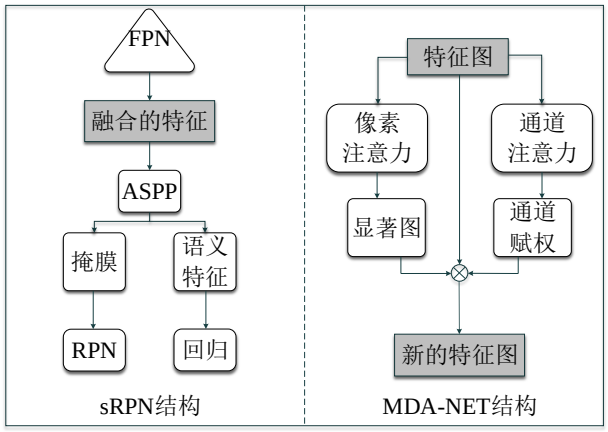

图 4 sRPN和MDA-NET算法结构

Fig. 4 Structure of sRPN and MDA-NET algorithm

在解决无人机影像中的复杂背景问题时，上下文信息可以帮助模型对目标与背景的理解，从而从提取更好的目标特征，但上下文信息需要进行筛选，通常只有部分信息是对模型有用的；空洞卷积在增加感受野的同时保留细节信息，为了适应无人机影像中的目标分布和遮挡情况，多尺度空洞卷积中提取的特征大小和数量显得尤为重要；注意力机制可以有效过滤背景中的无用信息，不过在无人机目标检测这种特定的场景下需要合理地分配权重，避免小目标的漏检或误报。

复杂背景中目标的精细检测算法在交通检测和城市规划中有着广泛的应用前景[40]，随着交通量的日益增长和城市规模的不断扩大，航拍影像中非目标噪声也越来越多，同时由于航拍中难以避免的遮挡问题也会导致目标信息不完整，因此，如何在复杂的环境中提取目标特征的研究具有重要的应用意义。

2.3.2 无人机影像中的小目标问题

无人机影像中目标的尺度范围大，建筑与行人、山川与动物经常出现在同一图片中。小目标在图片中占比极小，提供的分辨率有限，从而造成检测困难。

较早的一些研究中，Sevo I等[48]证明了卷积神经网络可以有效地融入到航空图像的目标检测算法中。Sommer L等[49]将Fast R-CNN和Faster R-CNN 用于航空图像中的车辆检测，通过调整锚定框的大小和特征图的分辨率，来适应小目标检测。虽然卷积神经网络具有一定的泛化能力，但网络中的卷积和池化操作使特征图细节信息丢失过多，这对小目标检测来说是十分不利的。

为了实现不同尺度的特征提取，提升小目标检测的性能，Lin T Y等[50]设计了一个特征金字塔网络（Feature Pyramid Networks，FPN），实现了细节较多的底层特征和语义信息丰富的顶层特征的融合。FPN算法使用内网特征金字塔来代替特征化图像金字塔，大幅减少了运算量，解决了训练与测试时间不一致问题，图5列举了计算机视觉中的一些金字塔结构。Azimi S M等[51]提出了一种图像级联网络（Image Cascade Network，ICN），使图像金字塔模型和FPN的结合成为可能。此外，为了克服固定卷积核对几何变换建模的局限性，使用DIN（Deformable Inception Network[11][52]）代替在FPN中使特征输出减少的1x1卷积核，来增强对小目标的定位能力。在DOTA数据集上的实验表明，使用ICN和DIN后mAP均有着明显提升。基于FPN的方法是高效的，无人机影像的特点决定了对其检测需要更多的细节特征。Yang X等[53]将DenseNet中的密集连接用于FPN算法，在自上而下的网络中通过横向连接和密集连接来获取更高的分辨率特征；Wang J等[54]使用改进的Inception模块来代替FPN中的横向连接来加强特征传播。这些算法在一定程度上提升了小目标检测的效果，但新的模块的加入增加了计算成本，算法的速度难以得到保证，实验条件下的高性能算法如何应用到实时性较强的场景中值得进一步研究。文献[55]提出了一个轻量化的深度残差网络模型（LResnet）将底层特征信息融合到高层中，检测速度较FPN有明显提升。

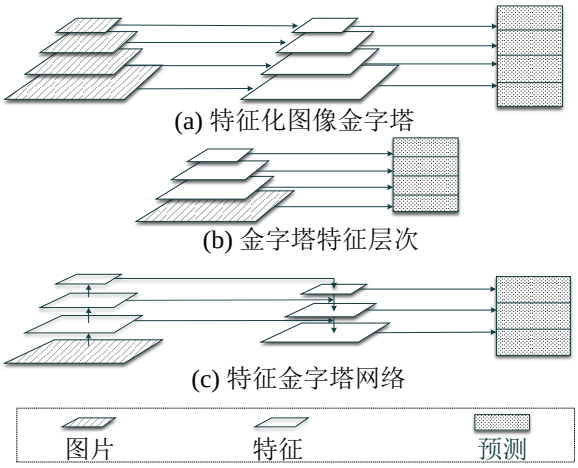

图 5　计算机视觉中的金字塔结构

Fig. 5 Pyramid structure in computer vision

He K等[56]和Zhu R等[57]对深度学习中的训练问题进行了研究：自然场景图像数据集下预训练好的模型对无人机影像的目标检测帮助有限，但从头开始的训练又增加了时间成本。Wang T等[58]将使用预训练的SSD算法和随机初始化训练的辅助网络结合，兼顾了训练时间成本和定位的精确性。辅助网络为标准的预训练网络提供包含准确轮廓边缘信息的低层、中层特征，补偿在预训练中准确轮廓边缘信息的丢失，使定位更准确。该算法在UAVDT数据集上的检测效果较好，提高了精度的同时保证了速度。在网络架构中增强预训练模型的泛化能力是直接有效的方法，并且可以保持原有网络模型的特性，但辅助网络本身也需要训练增加了运行和时间成本。Yu X等[59]将研究重点放在了预训练数据集和任务数据集的关系处理上，提出了尺度匹配方法，为远距离、大背景下小目标的目标检测带来性能提升。预先训练的数据集足够大时能在一定程度上提升检测效果，但预训练数据集与指定任务数据集差别很大时，得到的预训练模型帮助并不大。尺度匹配是一种尺度变换方法，使用于预训练的数据集和检测器学习的数据集之间的特征分布保持一致。实验以Faster R-CNN-FPN为基准并在MS COCO中进行预训练，结果表明该算法可以显著提升检测器性能。

在解决无人机影像中的小目标问题时，特征融合的方法可以结合多层特征来进行预测，提高对多尺度目标尤其是小目标的检测效果。根据不同场景下无人机目标检测任务的需求，具有相应特性的CNN模型或模块与FPN结合都取得了较好的检测效果，但却增加了时间成本。轻量化的网络模型是一种解决方法，另一种思路则从训练深度学习模型的角度出发，在已有数据集的情况下改进训练质量，具有很高的实际工程适用性。

小目标检测算法在小人脸检测、军事侦察、交通标志检测、安防等领域有着广泛应用需求[60]。在无人机的自主飞行中，应对突发紧急情况时主要依赖自身的传感器和控制器来完成对起落点的感知和紧急着陆，在较高的飞行高度完成对地面起落点的识别和空间定位对小目标检测的精度也有着很高的要求。

2.3.3　无人机影像中的大视场问题

无人机探测范围极广，且不受地理因素等限制，因此得到的图像视场往往很大。大视场下的目标检测面临着目标分布不均、目标稀疏等问题，如车辆总是在道路上密集存在、草原上的羊群也经常聚集在某一处、城市里的广场人群很密集而旁边的道路上的行人却相对稀疏。

直接将卷积检测器应用于这些图像所带来的处理成本很高，滑动窗口法虽然可以裁剪图片，但效率较低，因此很多算法通过减少搜索区域来提高效率。LaLonde R等[61]在研究广域运动图像（Wide Area Motion Imagery，WAMI）下的目标检测时，提出了一个两阶段的卷积神经网络，第一阶段改进了Faster R-CNN中的RPN，使之可以覆盖更多的潜在对象；第二阶段设计了一个基于神经元有效感受域的算法进行目标检测，只对高于设定阈值的一阶段输出进行高分辨率分析。实验使用WPAFB 2009数据集并与其他13种检测方法进行了对比，该算法的检测效果更好。Yang F等[62]针对航拍图像下目标分布不均匀的问题提出了一个面向聚集区域的目标检测算法。首先，将一个改进的RPN模块放置在特征提取网络的顶部来获取更大的感受野。第二步对提取的目标聚集区域进行尺度估计，对于偏移量过大的区域进行填充或分区操作，处理后的提取区域需要分别进行目标检测任务。最后，对所有检测结果通过NMS（Non Maximum Suppression）操作融合到全局图片上。在VisDrone、UAVDT和DOTA数据集上的实验表明，该算法检测性能和效率均有着

显著提升。

针对候选区域生成算法的缺点，一些研究将强化学习用于大视场图片的目标搜索中，如图6所示。Gao M等[63]将强化学习和卷积神经网络结合，用于大图像中小目标的检测。卷积神经网络对图片进行精细和粗略检测，所得的结果与真实值求差值来计算精度收益，经过回归生成精度收益图，指示不同区域的潜在放大精度增益。强化学习的任务是找到使奖励最大的动作，即更加高效地找到图像中的小目标。该算法在一些行人数据集上进行了实验，并未在无人机影像数据集上实验。Uzkent B等[64]的研究与前者类似，但面向的是视场较大的可见光遥感图像。该算法分别对图片进行粗略和精细搜索。两种搜索都是强化学习与卷积深度网络结合的级联算法。在粗略搜索中，先将低分辨率图片分割为相同大小的子图片，计算它们各自放大后的收益。在精细搜索中，对粗略搜索模块选择的子图片进行进一步的搜索空间优化，以最终决定要放大哪些子图片。在xView数据集上的实验表明，该方法提高了运行效率2.2倍，同时减少了对高分辨率图像的依赖性约70%。

在解决无人机影像中的大视场问题时，首先要考虑的是减少目标搜索的成本，常见的方法为区域特征编码方法的优化，如增加ROI输出的数量或增加ROI生成模块感受野；对子图片进行检测时，目标尺度的估计对检测精度有着较大影响。减少搜索区域的方法本质上还是两阶段的目标检测方法，需要遍历整张图片，效率较低。强化学习与CNN的结合实现了大视场影像中的自适应搜索，增加效率的同时保证了子图片检测的精度。

无人机在很多领域中的应用需要对较大的地理空间或场景进行监测和数据采集，如行人检测、遥感测绘、农业监测等。在大视场影像中进行快速准确的目标搜索和检测不仅可以减少运算时间、降低对硬件的要求，还有着重要的现实意义，克罗地亚搜救队在2019年提出了一种基于无人机航拍影像的搜索方法，用以寻找荒野中的失踪人员或有用的痕迹信息[65]。

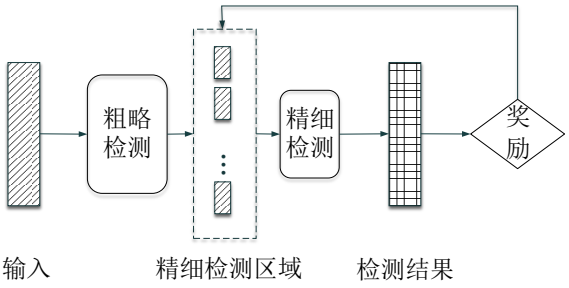

图 6 强化学习在目标检测中的应用

Fig. 6 Application of Reinforcement Learning in object detection

2.3.4 无人机影像中的旋转问题

无人机影像中的物体可能在任意位置和方向上出现，同一类物体的角度变化也不尽相同。无人机目标检测任务因此变得困难，旋转的物体使位置回归变得困难，因而大量的目标被漏检。文本检测也有着同样的特点，一些改进的目标检测的研究是在文本检测的启发下进行的，近年来有很多创新性的算法来解决目标的旋转问题。

常见的检测方法按照候选区域和包围盒的形式，可分为水平检测和旋转检测，如图7所示。

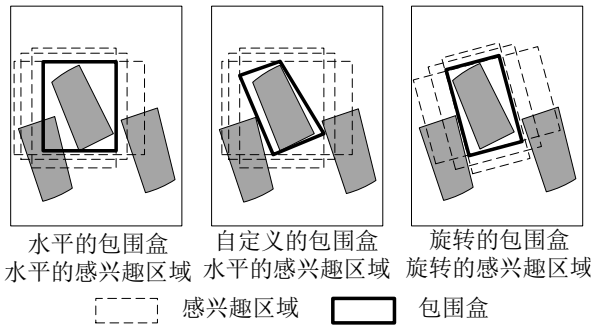

图 7 水平检测和旋转检测

Fig. 7 Rotated and horizontal object detection

Jiang Y等[66]改进了Faster R-CNN算法，用于检测任意方向的文本内容：使用两点坐标和盒高来描述包围盒，通过多尺度的ROI池化来更好的提取水平和竖直方向的特征。该方法提出的包围盒较好地适应了文字检测，但无人机影像中的目标存在分布密集的情况且旋转角度是任意的，需要新的包围盒形式来对其定位。Xu Y等[67]使用Faster R-CNN算法的分类结果，对于回归预测，引入旋转因子和水平包围盒顶点偏移量参数来对得到的水平包围盒进行偏移改进，用四边形来回

归定位。该算法在DOTA数据集上取得了73.39% mAP的结果，但由于仍是基于水平候选区域下的检测，位置回归的过程存在一些与真实值不匹配的情况。

Ma J等[68]提出了使用旋转的候选区域来进行文本检测。在Faster R-CNN算法中引入角度参数，生成带有角度信息的锚定框，进而得到任意方向的候选区域，并将此称为RRPN（Rotation Region Proposal Networks），相应的RROI（The Rotation Region of Interest）池化过程是将旋转的候选区域与特征图关联后再进行的池化操作。该方法提升了包围盒回归的精度，但由于产生更多的旋转锚定框，计算量较大。为了避免增加锚点数量，Jian Ding等[69]使用水平的锚定框，在RPN阶段通过全连接学习得到旋转ROI。具体来说，在有向包围盒（Oriented Bounding Box，OBB）注释的监督下，对RoI进行空间变换并学习变换参数（ROI-Transformer，RT）。之后，从旋转ROI中提取旋转不变特征，用于后续的分类和定位。由于避免了大量旋转锚定框的生成，该算法减少了计算量，在DOTA和HRSC数据集上的检测性能也有显著提升。

Zhou X等[70]提出了CenterNet，使用无锚点（Anchor-Free）的回归方法。用包围盒的中心点来表示目标，目标的大小尺寸则直接从中心点位置进行回归。Pan X等[71]在CenterNet的基础上增加了旋转角度预测，提出DRN（Dynamic Refinement Network）：根据物体的形状和旋转方向来自适应调整感受野，同时对目标的分类和回归进行动态修正。

同类物体的旋转同时会对目标检测中的分类任务造成困扰，CHENG G等[72]提出的算法用以解决旋转后带来的检测困难。通过在全连接层或ROI池化层加上正则约束项来优化一个新的目标函数，将训练样本在旋转前后的特征表示紧密地映射在一起，以确保旋转前后相似特征的分享，从而实现旋转不变性。实验表明该算法在航拍车辆检测和遥感图像分类任务上都有着性能提升。

在解决无人机影像中的旋转问题时，较为直接而简便的方法为保持水平的ROI不变，自定义包围盒的形状来适应目标旋转特性；使用旋转的RROI生成的区域特征与目标旋转特性较为匹配，可以有效避免大量的回归错位，但旋转的锚定框的生成增加了计算量；通过默认的水平锚定框转换得到RROI，避免了计算量的增加，且仍有着较高的回归精度。而相对于无锚点的回归摆脱了锚定框对包围盒的限制，增强了模型的实时性和精度，不过回归的稳定性需要进一步研究。

旋转问题是无人机航拍目标检测中的一大瓶颈问题，高精度的旋转检测的实现极大地拓展了无人机的应用场景，特别是航拍影像中密集区域中目标的定位，如停车场的车辆、港口中停泊的舰船、航空港中的航空器以及由此衍生的计数任务[30]。

表 4　不同无人机目标检测算法在 DOTA-v1.0 数据集上的有向目标检测结果对比

Table 4 Results comparison of OBB task on DOTA-v1.0 dataset of different UAV object detection algorithms

算法	飞机	棒球场	桥梁	田径场	小型车辆	大型车辆	船只	网球场	篮球场	储藏罐	足球场	环岛	港口	游泳池	直升机	mAP
sRPN	88.20	86.40	59.40	80.00	68.10	75.60	87.20	90.90	85.30	84.10	73.80	77.50	76.40	73.70	69.50	**76.60**
SCRDet	89.98	80.65	52.09	68.36	68.36	60.32	72.41	90.85	87.94	86.86	65.02	66.68	66.25	68.24	65.21	**72.61**
ICN	81.40	74.30	47.70	70.30	64.90	67.80	70.00	90.80	79.10	78.20	53.60	62.90	67.00	64.20	50.20	**68.20**
ICL-FPN	89.56	85.95	54.21	72.90	76.52	74.16	85.63	89.85	83.81	86.48	54.89	69.64	73.94	69.06	63.32	**75.33**
RRPN	88.52	71.20	31.66	59.30	51.85	56.19	57.25	90.81	72.84	67.38	56.69	52.84	53.08	51.94	53.58	**61.01**
RT	88.64	78.52	43.44	75.92	68.81	73.68	83.59	90.74	77.27	81.46	58.39	53.54	62.83	58.93	47.67	**69.56**
DRN	88.91	80.22	43.52	63.35	73.48	70.69	84.94	90.14	83.85	84.11	50.12	58.41	67.62	68.60	52.50	**70.70**

表 5　不同无人机目标检测算法的优缺点对比

Table 5 Comparison of the advantages and disadvantages of different UAV object detection algorithms

问题	算法	原理	优点	缺点	年份
复杂背景	Segment before detect	用语义分割得到的语义图来做车辆检测	语义图对车辆的分割精度高	-	2017
	sRPN	使用 ASPP 来指导 RPN，结合上下文信息进行检测	感受野增大，实现多尺度检测	ASPP 提取的特征在尺度上较为稀疏	2019
	特征提取模块	改进的多尺度空洞卷积	扩大了对特征的感受野，提高了复杂背景下和有遮挡的目标检测效果	-	2020
	MDA-NET	有监督的双通道注意力机制	弱化背景信息	检测小目标或形似物时容易引入干扰	2019
小目标	ICN	将图像金字塔和特征金字塔模型结合	多尺度提取强弱语义特征	网络较深或复杂时，增加的计算量较大	2018
	DIN	在 FPN 中使用可变性卷积模块	提高网络模型的几何变换建模能力		2018
	ICL-FPN	将 FPN 中的横向连接用 Inception 模块代替	加强特征传播，更好地处理目标尺度变化	网络实时性值得进一步研究	2018
	DFPN	在 FPN 自上而下的网络中使用密集连接	可以获取更高分辨率的特征		2018
	LResnet	使用残差结构的深度卷积和点卷积	融合底层和高层特征且速度优于 FPN	-	2020
	LSN	在主网络中加入初始化训练的轻量的辅助网络	兼顾预训练和初始化训练的优点	增加运行和时间成本	2019
	Scale Match	对预训练数据集和任务数据集进行尺度匹配	只改进数据集，不增加网络模型开销	-	2020
大视场	ClusterNet	增加 RPN 网络输出，利用神经元感受域原理筛选出待检目标	减少了高分辨率下的搜索空间	检测没有考虑到目标的尺寸；搜索效率较低	2018
	ClusDet	增加 RPN 感受野，对目标进行尺度估计后分别进行检测	对目标进行尺度估计，提高了检测精度	搜索效率较低	2019
	CPNet&CFNet	强化学习与 CNN 结合，不同目标使用不同精度的检测	自适应搜索，大幅提高了检测效率	-	2020
旋转	Gliding the vertex of HBB	用四边形包围盒代替矩形包围盒来定位目标	改进的包围盒可以适应旋转的目标	存在水平检测中的回归错位问题	2019
	RRPN	使用带角度信息的锚定框，得到任意方向的候选区域	有效避免了因旋转带来的回归错位	大量旋转锚定框的生成增加了计算量	2018
	ROI Transformer	使用水平锚定框，通过学习得到旋转 ROI	避免产生大量旋转锚定框，较少计算量	-	2019
	DRN	自适应目标形状和旋转的无锚点回归	运行速度快，泛化性较好	检测结果不稳定	2020
	RIFD-CNN	使用正则约束项来优化一个新的目标函数	目标旋转前后分享相似的特征，实现旋转不变性	-	2019

3　总结与展望

目前，无人机目标检测算法的受关注程度与日俱增，现有的算法也取得了不错的检测效果，但还有很大的改进空间。复杂背景给目标检测任务带来的干扰得到了有效抑制，但现有的算法仍存在虚警和漏检问题，如果目标处于过于密集或大量形似物的环境中，检测效果不太理想；基于两阶段方法的目标检测算法在分类和回归的精度上有优势，大部分小目标检测方法都是基于此来进行改进，加之新模块和网络的引入，使得检测速度仍然较慢；多数算法都是基于现有算法的改进，很多适用于自然场景下目标检测方法和思想被保留下来，增加了检测的局限性，如锚定框的

保留限制了目标位置的确定，需要有新的方法来提高定位精度。

针对上述问题和近几年的研究趋势，本文对无人机目标检测未来研究的方向做出如下讨论：

1）在增大感受野的同时，密集地生成不同尺度的特征。无人机影像的分辨率较高，ASPP可以在保证特征分辨率的同时，增大感受野，但随着扩张率的增长，空洞卷积会失效。与CNN一样，空洞卷积的网络结构是可以进一步优化的，从而在合理的扩张率内获得与无人机影像相匹配的感受野和尺度分布[73]。

2）自适应地融合特征和生成ROI。无人机因应用场景的不同而获取不同特性的影像，为了避免有用信息的丢失，在特征融合和生成ROI时可以给不同的特征层赋权，通过加权融合得到相应的上下文特征和高质量的ROI，进而提高目标检测模型的泛化性[74]。

3）深度学习方法与其他方法的结合。深度学习方法在目标检测领域有着显著的优势，也取得了极大的成功，其他算法的加入将会弥补单一方法的局限性：a.数据预处理算法。深度学习方法的效果依赖于输入数据的质量又无法筛选数据，可以从数据增强和减少数据中的冗余特征两方面出发，来提高深度学习模型的计算效率。b.模型的优化算法。深度学习通过调节学习率可以实现自适应优化，但数据或样本规模较大，或对收敛性有较高要求时，可以选择合适的算法来优化网络的结构和参数，提高检测效果。c.功能性算法。作为数据驱动的方法，深度学习并不是解决某些特定问题的最佳方案，可以选择针对性较强的算法并合理分配权重，灵活高效完成任务。

4）减少进行位置回归时的限制。基于锚点的回归中锚定框的设置需要与待检测的目标形状相吻合，但在无人机影像中，目标的形状和朝向多变，预设的锚定框限制了位置回归。无锚点的方法通过预测目标关键点来获得包围盒，不预设目标形状，以一种灵活的方式进行位置回归，更适合无人机航拍目标检测任务。对于关键点重合而导致的检测结果不稳定问题，可以对关键点进行二次预测和匹配来提高检测的精确性[75]。

4 结束语

本文总结了深度学习在目标检测领域的研究成果，分析了相关算法的优缺点。对航空影像数据集和迁移学习进行了详细介绍，重点从无人机影像的特点出发，对该领域的目标检测取得的进展进行了分析，指出了存在的问题和发展的方向。目前，无人机技术正处于快速发展时期，无人机目标检测具有广阔的研究前景。

参 考 文 献

[1] 朱华勇, 牛轶峰, 沈林成, 等. 无人机系统自主控制技术研究现状与发展趋势[J].国防科技大学学报,2010,32(03):115-120.
ZHU H Y, NIU Y F, SHEN C L, et al. State of the Art and Trends of Autonomous Control of UAV Systems [J]. Journal of National University of Defense Technology, 2010,32(03):115-120 (in Chinese).

[2] 宋闯, 赵佳佳, 王康, 等. 面向智能感知的小样本学习研究综述[J]. 航空学报, 2020, 41(S1):723756.
SONG C, ZHAO J J, WANG K, et al. Few shot learning based intelligent perception： A survey [J]. Acta Aeronautica et Astronautica Sinica, 2020, 41(S1):723756 (in Chinese).

[3] LOWE D G. Distinctive Image Features from Scale-Invariant Keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.

[4] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]// CVPR 2005: Proceedings of the 2005 IEEE conference on computer vision and pattern recognition. Washington, DC: IEEE Computer Society, 2005: 886-893.

[5] LECUN Y, BOTTOU L. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[6] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet Classification with Deep Convolutional Neural Networks[J]. Advances in neural information processing systems, 2012, 25(2): 1097-1105.

[7] DENG J, DONG W, SOCHER R, et al. Imagenet: A large-scale hierarchical image database[C]// CVPR 2009: Proceedings of the 2009 IEEE conference on computer vision and pattern recognition. Washington, DC: IEEE Computer Society, 2009: 248-255.

[8] LIN M, CHEN Q, YAN S. Network In Network[EB/OL]. （2014-03-04）[2020-05-11]. https://arxiv.org/pdf/1312.4400.pdf.

[9] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks[C]// ECCV 2014: 2014 European conference on computer vision. Berlin: Springer, 2014: 818-833.

[10] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[11] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]// CVPR 2015: Proceedings of the 2015 IEEE conference on computer vision and pattern recognition. Washington, DC: IEEE Computer Society, 2015: 1-9.

[12] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// CVPR 2016: Proceedings of the 2016 IEEE conference on computer vision and pattern recognition. Washington, DC: IEEE Computer Society, 2016: 770-778.

[13] HUANG G, LIU Z, DER MAATEN L V, et al. Densely Connected Convolutional Networks[C]. CVPR 2017: Proceedings of the 2017 IEEE conference on computer vision and pattern recognition. Washington, DC: IEEE Computer Society, 2017: 2261-2269.

[14] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-Excitation Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019: 1-1

[15] GIRSHICK R, DONAHUE J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// CVPR 2014: Proceedings of the 2014 IEEE conference on computer vision and pattern recognition. Washington, DC: IEEE Computer Society, 2014: 580-587.

[16] 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.

[17] GIRSHICK R. Fast R-CNN [C]// CVPR 2015: Proceedings of the 2015 IEEE conference on computer vision and pattern recognition. Washington, DC: IEEE Computer Society, 2015: 1440-1448.

[18] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[19] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]// CVPR 2016: Proceedings of the 2016 IEEE conference on computer vision and pattern recognition. Washington, DC: IEEE Computer Society, 2016: 779-788.

[20] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector[C]// ECCV 2016: 2016 European conference on computer vision. Berlin: Springer, 2016: 21-37.

[21] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft coco: Common objects in context[C]// ECCV 2014: 2014 European conference on computer vision. Berlin: Springer, 2014: 740-755.

[22] EVERINGHAM M, ESLAMI S M, VAN GOOL L, et al. The Pascal Visual Object Classes Challenge: A Retrospective[J]. International Journal of Computer Vision, 2015, 111(1): 98-136.

[23] 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.

[24] CHENG G, ZHOU P, HAN J, et al. Learning Rotation-Invariant Convolutional Neural Networks for Object Detection in VHR Optical Remote Sensing Images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(12): 7405-7415.

[25] RAZAKARIVONY S, JURIE F. Vehicle detection in aerial imagery[J]. Journal of Visual Communication and Image Representation, 2016: 187-203.

[26] ZHU H, CHEN X, DAI W, et al. Orientation robust object detection in aerial images using deep convolutional neural network[C]//2015 IEEE International Conference on Image Processing. Piscataway, NJ: IEEE, 2015: 3735-3739.

[27] XIA G S, BAI X, DING J, et al. DOTA: A large-scale dataset for object detection in aerial images[C]// CVPR

2018: Proceedings of the 2018 IEEE conference on computer vision and pattern recognition. Washington, DC: IEEE Computer Society, 2018: 3974-3983.

[28] ROBICQUET A, SADEGHIAN A, ALAHI A, et al. Learning Social Etiquette: Human Trajectory Understanding In Crowded Scenes[C] // ECCV 2016: 2016 European conference on computer vision. Berlin: Springer, 2016: 549-565.

[29] BAREKATAIN M, MARTI M, SHIH H, et al. Okutama-Action: An Aerial View Video Dataset for Concurrent Human Action Detection[C]// CVPR 2017: Proceedings of the 2017 IEEE conference on computer vision and pattern recognition. Washington, DC: IEEE Computer Society, 2017: 2153-2160.

[30] HSIEH M, LIN Y, HSU W H, et al. Drone-Based Object Counting by Spatially Regularized Regional Proposal Network[C]// ICCV 2017: Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2017: 4165-4173.

[31] ZHU P, WEN L, BIAN X, et al. Vision Meets Drones: A Challenge[EB/OL]. （ 2018-04-23 ） [2020-05-11]. https://arxiv.org/pdf/1804.07437.pdf

[32] ZHU P, SUN Y, WEN L, et al. Drone Based RGBT Vehicle Detection and Counting: A Challenge[EB/OL]. （ 2020-03-05 ） [2020-07-25]. https://arxiv.org/pdf/2003.02437.pdf

[33] TORRALBA A, EFROS A A. Unbiased look at dataset bias[C]// CVPR 2011: Proceedings of the 2011 IEEE conference on computer vision and pattern recognition. Washington, DC: IEEE Computer Society, 2011: 1521-1528.

[34] PAN S J, YANG Q. A Survey on Transfer Learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345-1359.

[35] PAN B, TAI J, ZHENG Q, et al. Cascade Convolutional Neural Network Based on Transfer-Learning for Aircraft Detection on High-Resolution Remote Sensing Images[J]. Journal of Sensors, 2017: 1-14.

[36] 袁功霖，侯静，尹奎英. 基于迁移学习与图像增强的夜间航拍车辆识别方法[J].计算机辅助设计与图形学学报,2019,31(03):467-473.
YUAN G L, HOU J, YIN K Y. Night-Time Aerial Image Vehicle Recognition Technology Based on Transfer Learning and Image Enhancement [J]. Journal of Computer-Aided Design & Computer Graphics, 2019,31(03):467-473 (in Chinese).

[37] 王泽隆，徐向辉，张雷. 基于仿真 SAR 图像深度迁移学习的自动目标识别[J].中国科学院大学学报,2020,37(04):516-524.
WANG Z L, XU X H, ZHANG L. Study of deep transfer learning for SAR ATR based on simulated SAR images [J]. Journal of University of Chinese Academy of Sciences, 2020,37(04):516-524(in Chinese).

[38] ZAMIR A R, SAX A, SHEN W B, et al. Taskonomy: Disentangling Task Transfer Learning[C]// CVPR 2018: Proceedings of the 2018 IEEE conference on computer vision and pattern recognition. Washington, DC: IEEE Computer Society, 2018: 3712-3722.

[39] YOSINSKI J, CLUNE J, BENGIO Y, et al. How transferable are features in deep neural networks?[C]// International Conference on Neural Information Processing Systems, Cambridge, MA: MIT Press, 2014: 3320-3328.

[40] AUDEBERT N, SAUX B L, LEFEVRE S, et al. Segment-before-Detect: Vehicle Detection and Classification through Semantic Segmentation of Aerial Images[J]. Remote Sensing, 2017, 9(4):368.

[41] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 386-397.

[42] CHEN L C, HERMANS A, PAPANDREOU G, et al. MaskLab: Instance Segmentation by Refining Object Detection with Semantic and Direction Features [C]// CVPR 2018: Proceedings of the 2018 IEEE conference on computer vision and pattern recognition. Washington, DC: IEEE Computer Society, 2018: 4013-4022.

[43] CHEN K, PANG J, WANG J, et al. Hybrid task cascade for instance segmentation[C]// CVPR 2019: Proceedings of the 2019 IEEE conference on computer vision and pattern recognition. Washington, DC: IEEE Computer Society, 2019: 4974-4983.

[44] LI C, XU C, CUI Z, et al. Learning Object-Wise Semantic Representation for Detection in Remote Sensing

Imagery[C]// CVPR 2019: Proceedings of the 2019 IEEE conference on computer vision and pattern recognition. Washington, DC: IEEE Computer Society, 2019: 20-27.

[45] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking Atrous Convolution for Semantic Image Segmentation[EB/OL]. （2017-12-05）[2020-05-11]. https://arxiv.org/pdf/1706.05587.pdf

[46] 张瑞倩, 邵振峰, ALEKSEI PORTNOV, 等. 多尺度空洞卷积的无人机影像目标检测方法[J].武汉大学学报(信息科学版),2020,45(06):895-903.
ZHANG R Q, SHAO Z F, ALEKSEI PORTNOV, et al. Multi-scale Dilated Convolutional Neural Network for Object Detection in UAV Images[J]. Geomatics and Information Science of Wuhan University, 2020,45(06):895-903(in Chinese).

[47] YANG X, YANG J, YAN J, et al. SCRDet: Towards more robust detection for small, cluttered and rotated objects[C]// ICCV 2018: Proceedings of the 2018 IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2019: 8232-8241.

[48] SEVO I, AVRAMOVIC A. Convolutional Neural Network Based Automatic Object Detection on Aerial Images[J]. IEEE Geoscience and Remote Sensing Letters, 2016, 13(5): 740-744.

[49] SOMMER L W, SCHUCHERT T, BEYERER J. Fast deep vehicle detection in aerial images[C]// WACV 2017: 2017 IEEE Winter Conference on Applications of Computer Vision. Washington, DC: IEEE Computer Society, 2017: 311-319.

[50] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]// CVPR 2017: Proceedings of the 2017 IEEE conference on computer vision and pattern recognition. Washington, DC: IEEE Computer Society, 2017: 2117-2125.

[51] AZIMI S M, VIG E, BAHMANYAR R, et al. Towards multi-class object detection in unconstrained remote sensing imagery[C]//Asian Conference on Computer Vision. Berlin: Springer, 2018: 150-165.

[52] DAI J, QI H, XIONG Y, et al. Deformable convolutional networks[C]// ICCV 2017: Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2017: 764-773.

[53] YANG X, SUN H, FU K, et al. Automatic Ship Detection in Remote Sensing Images from Google Earth of Complex Scenes Based on Multiscale Rotation Dense Feature Pyramid Networks[J]. Remote Sensing, 2018, 10(1):132.

[54] WANG J, DING J, GUO H, et al. Mask OBB: A Semantic Attention-Based Mask Oriented Bounding Box Representation for Multi-Category Object Detection in Aerial Images[J]. Remote Sensing, 2019, 11(24):2930.

[55] 刘芳, 吴志威, 杨安喆, 等. 基于多尺度特征融合的自适应无人机目标检测[J]. 光学学报,2020,40(10):133-142.
LIU F, WU Z W, YANG A Z, et al. Multi-Scale Feature Fusion Based Adaptive Object Detection for UAV[J]. Acta Optica Sinaica, 2020,40(10):133-142(in Chinese).

[56] HE K, GIRSHICK R, DOLLÁR P. Rethinking ImageNet Pre-training[C]// ICCV 2019: Proceedings of the 2019 IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2019: 4918-4927.

[57] ZHU R, ZHANG S, WANG X, et al. ScratchDet: Training single-shot object detectors from scratch[C]// CVPR 2019: Proceedings of the 2019 IEEE conference on computer vision and pattern recognition. Washington, DC: IEEE Computer Society, 2019: 2268-2277.

[58] WANG T, ANWER R M, CHOLAKKAL H, et al. Learning rich features at high-speed for single-shot object detection[C]// ICCV 2019: Proceedings of the 2019 IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2019: 1971-1980.

[59] YU X, GONG Y, JIANG N, et al. Scale Match for Tiny Person Detection[C]// WACV 2020: 2020 IEEE Winter Conference on Applications of Computer Vision. Washington, DC: IEEE Computer Society, 2020: 1257-1265.

[60] 刘颖, 刘红燕, 范九伦, 等. 基于深度学习的小目标检测研究与应用综述[J].电子学报,2020,48(03):590-601.
LIU Y, LIU H Y, FAN J L, et al. A Survey of Research and Application of Small Object Detection Based

on Deep Learning [J]. ACTA ELECTRONICA SINICA, 2020,48(03):590-601 (in Chinese).

[61] LALONDE R, ZHANG D, SHAH M. ClusterNet: Detecting Small Objects in Large Scenes by Exploiting Spatio-Temporal Information[C]// CVPR 2018: Proceedings of the 2018 IEEE conference on computer vision and pattern recognition. Washington, DC: IEEE Computer Society, 2018: 4003-4012.

[62] YANG F, FAN H, CHU P, et al. Clustered object detection in aerial images[C]// ICCV 2019: Proceedings of the 2019 IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2019: 8311-8320.

[63] GAO M, YU R, LI A, et al. Dynamic zoom-in network for fast object detection in large images[C]// CVPR 2018: Proceedings of the 2018 IEEE conference on computer vision and pattern recognition. Washington, DC: IEEE Computer Society, 2018: 6926-6935.

[64] UZKENT B, YEH C, ERMON S. Efficient object detection in large images using deep reinforcement learning[C]// WACV 2020: 2020 IEEE Winter Conference on Applications of Computer Vision. Washington, DC: IEEE Computer Society, 2020: 1824-1833.

[65] BOŽICSTULIC D, MARUSIC Ž, GOTOVAC S, et al. Deep Learning Approach in Aerial Imagery for Supporting Land Search and Rescue Missions[J]. International Journal of Computer Vision, 2019, 127(9): 1256-1278.

[66] JIANG Y, ZHU X, WANG X, et al. R2CNN: Rotational Region CNN for Orientation Robust Scene Text Detection [EB/OL]. （ 2017-06-30 ） [2020-05-11]. https://arxiv.org/ftp/arxiv/papers/1706/1706.09579.pdf

[67] XU Y, FU M, WANG Q, et al. Gliding vertex on the horizontal bounding box for multi-oriented object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020: 1-1.

[68] MA J, SHAO W, YE H, et al. Arbitrary-Oriented Scene Text Detection via Rotation Proposals[J]. IEEE Transactions on Multimedia, 2018, 20(11): 3111-3122.

[69] DING J, XUE N, LONG Y, et al. Learning ROI transformer for oriented object detection in aerial images[C]// CVPR 2019: Proceedings of the 2019 IEEE conference on computer vision and pattern recognition. Washington, DC: IEEE Computer Society, 2019: 2849-2858.

[70] ZHOU X, WANG D, KRAHENBUHL P, et al. Objects as Points [EB/OL]. （ 2019-04-25 ） [2020-07-25]. https://arxiv.org/pdf/1904.07850v1.pdf.

[71] PAN X , REN Y , SHENG K , et al. Dynamic Refinement Network for Oriented and Densely Packed Object Detection [EB/OL].（2020-06-10）[2020-07-25]. https://arxiv.org/pdf/2005.09973.pdf.

[72] CHENG G, HAN J, ZHOU P, et al. Learning Rotation-Invariant and Fisher Discriminative Convolutional Neural Networks for Object Detection[J]. IEEE Transactions on Image Processing, 2019, 28(1): 265-278.

[73] YANG M, YU K, ZHANG C, et al. DenseASPP for Semantic Segmentation in Street Scenes[C] // CVPR 2018: Proceedings of the 2018 IEEE conference on computer vision and pattern recognition. Washington, DC: IEEE Computer Society, 2018: 3684-3692.

[74] GUO C, FAN B, ZHANG Q, et al. AugFPN: Improving Multi-Scale Feature Learning for Object Detection.]// CVPR 2020: Proceedings of the 2020 IEEE conference on computer vision and pattern recognition. Washington, DC: IEEE Computer Society, 2020: 12595–12604.

[75] DONG Z, LI G, LIAO Y, et al. CentripetalNet: Pursuing High-Quality Keypoint Pairs for Object Detection.]// CVPR 2020: Proceedings of the 2020 IEEE conference on computer vision and pattern recognition. Washington, DC: IEEE Computer Society, 2020: 10519–10528.

Survey of object detection in unmanned aerial vehicle imagery based on deep learning

*

Abstract: To improve the autonomous sensing ability of Unmanned Aerial Vehicle(UAV), object detection is one of the key technologies. Research on object detection is of great significance in UAV applications. Compared with traditional methods based on manual features, deep learning based on convolutional neural network has powerful ability of feature learning and expression. Deep learning becomes the mainstream algorithm in object detection. In recent years, object detection has made a series breakthrough research result in the field of natural scene and the research in UAV has gradually become a hotspot simultaneously. Research progress of object detection algorithm based on deep learning was reviewed and the advantages and disadvantages of them were summarized. Then, some typical aerial image datasets and the method of Transfer Learning were introduced. Aiming at the complex background, small and rotating object, large field of view in UAV imagery, relevant algorithms were analyzed. At last, the existing problems and possible future development directions were discussed.

Keywords: object detection; UAV imagery; convolution neural network; computer vision; deep learning; transfer learning

---

Received：2019-xx-xx；Revised：2019-xx-xx；Accepted：2019-xx-xx；Published online:
URL:
Foundation item：Sichuan Province Science and Technology plan-Key Research and Development (2019YFG0308); Sichuan Province Talents Fostering Quality and Teaching Reform program of Higher Education in 2018-2020 (JG2018-325); Sichuan province College Students' Innovative Entrepreneurial Training Plan Program (S202010624029); Research Program of Civil Aviation Flight University of China (J2008-78); General Program by Civil Aviation Flight University of China (J2020-078)
*Corresponding author. E-mail: qurk0226@qq.com