DOI:10.13203/j.whugis20200253 文章编号:1671-8860(2020)06-0895-09

多尺度空洞卷积的无人机影像目标检测方法

张瑞倩[1] 邵振峰[2] Aleksei Portnov[3] 汪家明[2]

1 武汉大学遥感信息工程学院,湖北 武汉,430079
2 武汉大学测绘遥感信息工程国家重点实验室,湖北 武汉,430079
3 莫斯科国立测绘大学,俄罗斯 莫斯科,105064

摘 要:无人机作为一种新型遥感传感器,越来越多地被应用在医疗、交通、环境监测、灾害预警、动物保护以及军事等领域。由于无人机飞行器飞行高度差异大、采集影像视角可变、飞行速度快,因此无人机影像上的目标具有尺度变化大、分布差异明显、背景复杂、存在大量遮挡等特点,这为无人机影像目标检测带来了一定的困难。针对此,提出一种多尺度空洞卷积的无人机影像目标检测方法,在现有的目标检测算法的基础上,增加多尺度的空洞卷积模块,加大视野感知域,提高网络对无人机影像中的目标分布情况、尺寸差异等特点的学习能力,进一步提升网络对无人机影像中多尺度、复杂背景下的目标的检测精度。实验结果表明,所提出的算法在不增加网络参数的情况下,提升了无人机影像上目标检测的精确度和召回率,具有一定的有效性和鲁棒性。

关键词:多尺度模型;空洞卷积;无人机影像;目标检测

中图分类号:P237 **文献标志码**:A

无人驾驶飞机(unmanned aerial vehicle,UAV)是利用无线电遥控设备和自备的程序控制装置操纵的不载人飞机,或者由车载计算机完全地或间歇地自主操作[1-3]。随着科学技术和遥感传感器设计的不断发展,无人机作为一种新型遥感传感器,越来越多地被应用在医疗、交通、环境监测、灾害预警、动物保护以及军事等领域。无人机遥感技术[4]作为一种重要的空间数据采集手段,具有寿命长、图像实时传输、成本低、灵活性强等多种优点。

作为无人机应用领域的核心技术之一,无人机影像目标检测技术[5]对于无人机影像的应用(如目标监督、管理和行为识别等)有着极为重要的意义。随着计算机视觉、图像与视频处理及模式识别技术的发展,自然图像中的目标检测技术也越来越成熟,检测结果不断提高,相关检测算法也逐渐被运用到无人机影像的目标检测任务中。然而,由于无人机飞行器飞行高度差异大、采集影像视角可变、飞行速度快,因此无人机影像上的目标具有尺度变化大、分布差异明显、背景复杂、存在大量遮挡等特点,这给无人机影像上目标检测特征的选取带来了难题。

图1展示了公开数据集VisDrone[6]中的4幅无人机影像和其标注的感兴趣目标位置。图1中,红色矩形框表示标记的目标所在位置。从图1可以看出,4幅影像的拍摄视角存在较大差异,导致目标分布和尺度等具有较大差异。整体来看,图1中的4幅图均为无人机影像,但是目标分布等差异明显,表现出无人机影像复杂多变的特点。同时,无人机影像上的目标相对遥感卫星影像和自然场景影像中的目标更加复杂,目标检测问题具有更大难度。

传统无人机影像目标检测方法主要通过融合视频背景差分或运动信息的方法进行,通过视频图像帧之间的差异、背景分割等,完成无人机影像中运动目标的检测[7-10]。尽管这些方法在目标检测上显示出一定的有效性,但是它们对于视频中的帧间差异具有很高的依赖性,对于有遮挡的目标、小尺寸目标、复杂背景条件下的目标,很难达到较好的检测效果。同时,结合目标特点的特征表达和背景差异相结合的方法需要对不同特征进行设计,增加了工作量。

收稿日期:2020-05-28

项目资助:国家重点研发计划战略性国际科技创新合作重点专项(2016YFE0202300);中国工程院咨询研究项目(2020ZD16);国家自然科学基金(41771454);湖北省自然科学基金计划创新群体项目(2018CFA007)。

第一作者:张瑞倩,博士,主要研究方向为遥感影像处理和目标检测。zhangruiqian@whu.edu.cn

通讯作者:邵振峰,博士,教授。shaozhenfeng@whu.edu.cn

(a) 俯视视角下的无人机影像 (b) 倾斜视角下的无人机影像 (c) 水平视角下的无人机影像 (d) 夜晚场景下的无人机影像

图1 不同视角、不同背景下的无人机影像目标位置示意图

Fig.1 An Illustration of Object Location in UAV Imagery with Different Perspectives, Backgrounds

随着计算机视觉技术的快速发展,自然场景中的目标检测技术不断提高,在ImageNet[11]、MSCOCO[12]等自然场景数据集上,目标检测方法的检测精度越来越高,其特征提取结果也更加鲁棒、高效。自然场景下的目标检测方法主要分为两阶段目标检测方法和单阶段目标检测方法两大类[13]。两阶段目标检测方法主要有两个阶段:首先网络生成一系列候选框,然后通过卷积神经网络对候选框的位置进行精准分类和回归,获得目标准确的位置和分类。早期典型的两阶段的目标检测方法有R-CNN[14]、SPP-Net[15]、Fast R-CNN[16]和Faster R-CNN[17]算法,获得了一系列较好的实验结果。为了进一步提高方法精度,其他文献又通过更精细准确的特征网络表达方法(HyperNet[18])、更精准的候选区域提取网络(FPN[19]、CRAFT[20])、更完善的感兴趣区域提取策略(R-FCN[21]、Cascade R-CNN[22])、样本后处理(OHEM[23]、Soft-NMS[24])等技术,获得了更准确优秀的目标检测结果。除了两阶段目标检测方法,单阶段目标检测方法也被广泛应用在自然场景影像检测中。单阶段的方法将整个目标检测网络看作一个整体,将目标位置的定位问题看作是单个回归问题,通过一个端到端的网络直接完成目标位置和类别的回归。最为知名的单阶段方法有YOLOv1、YOLOv2算法[25-26]、SSD算法[27]和一系列在SSD算法基础上进行增强改进的算法,如DSSD[28]、DSOD[29]等。单阶段的方法利用整幅影像作为网络的输入,直接将回归得到的矩形框位置和所属类别作为目标的预测位置和类别。单阶段的方法不需要进行候选区域提取,提升了检测的速度。但是相对两阶段的方法,单阶段的方法通常准确度不高。

自然图像目标检测技术发展迅速,一些文献开始尝试利用优秀的自然图像目标检测方法对无人机影像中的目标进行检测[30-33]。由于无人机图像存在大量的目标聚集分布现象,其目标相对自然场景中的目标尺度更加多变,对无人机影像仅采用适用于自然场景的目标检测方法,难以获得准确的检测结果。常见的改进方法需要加入预处理等数据处理操作,增加了人工成本和时间消耗。针对这些问题,文献[34]提出了一种结合深度学习和深度估计的微型无人机目标检测方法,对深度预测和目标检测进行融合,在一定程度上解决了目标分布带来的检测难度问题,但是其输入影像是立体像对,在普通场景下的应用受限。

总的来说,现有文献中无人机影像的目标检测方法主要有融合视频背景差分或运动信息的方法和利用类似自然影像目标检测的方法两种。然而,融合视频背景差分或运动信息的方法对于视频中的帧间差异具有很高的依赖性,而对有遮挡的目标、小尺寸目标、复杂背景条件下的目标,该方法很难达到较好的检测效果;类似自然影像目标检测的方法也很难对无人机影像中目标特有的分布情况进行特征表达,难以对小目标、聚集目标和有遮挡目标进行检测。因此,本文提出了一种多尺度空洞卷积的无人机影像目标检测方法,在现有的目标检测方法的基础上,增加多尺度的空洞卷积模块,加大视野感知域,提高网络对数据分布情况和数据尺寸差异的学习能力,提升网络对无人机影像中多尺度、复杂背景、存在遮挡情况的目标的检测能力。

1 空洞卷积的思路与设计

空洞卷积[35]又名扩张卷积,最早来源于小波变换。空洞卷积方法通过在卷积层引入一个新参数——扩张率来定义卷积操作时卷积核处理数据值的间距。普通的卷积层主要通过下采样操作获取特征图,如极大值池化方法,这使得采样得到的特征图尺寸越来越小,最后特征图上的每一像素点对应的原图是一个较大区域,这也意味着整张图的特征响应变得非常稀疏。为了使特征图的密集度增加,让特征表达更密集地响应原图情况,空洞卷积方法被应用在了卷积神经网络中。该方法不仅能获得密集的空间响应,而且不需要学习更多参数,已经被应用到了语义分割

等领域[36],取得了较好的实验效果。

空洞卷积是卷积神经网络中卷积层的一种改进结果。卷积层是深度卷积神经网络的重要组成部分,卷积操作[33]可视为求取图像对应位置附近区域像素的权重之和,并赋值到输出特征中心像素的过程。如图 2 所示,图 2(a)展示了一个 3×3 卷积核,图 2(b)展示了利用该卷积核对一张 9×9 的图像进行卷积时,中心坐标位置的像素卷积处理后的响应结果。卷积核直接与图 2(b)中的红色区域进行卷积操作,而该操作能够影响到的像素位置(考虑到图像的连续性,相邻像素与像素之间具有关联性)为黄色区域,黄色越深表示影响越大。通过图 2(b)可以看出,经过 3×3 卷积操作,输出图像的中心位置的像素主要由图像中心位置附近的一个 5×5 区域进行响应。当采用一个 3×3、扩张率为 2 的空洞卷积(如图 2(c)所示)进行操作时,输出图像中心位置的像素响应如图 2(d)所示。由图 2 可见,由于空洞卷积的感受野增加到 5×5,其完成了在计算量保持不变的情况下,对图像的一个 7×7 位置区域的响应过程。这充分展示了空洞卷积扩大感知域的作用。当每个像素点都通过空洞卷积进行操作时,可以极大地提高特征图的响应密集度,更好地完成特征提取操作,获得输入图像的更准确表达。

鉴于无人机影像的特点,对于无人机影像上的目标来说,处于不同飞行姿态下的目标具有较大差别,如果能增大感知域,能够在一定程度上对无人机影像中的目标分布情况进行分析,让网络更好地学习无人机影像中的目标分布情况,进一步促进网络对多尺度、复杂背景、存在遮挡情况的目标的检测能力进行改善和提高。

因此,本文提出了一种多尺度空洞卷积的无人机影像目标检测方法,通过空洞卷积和金字塔影像相结合的方法,增大无人机影像目标检测网络中的视野感知域,增强目标检测网络对背景、目标分布情况的分析能力,提升网络对无人机影像中多尺度、复杂背景、存在遮挡情况的目标的检测精度。

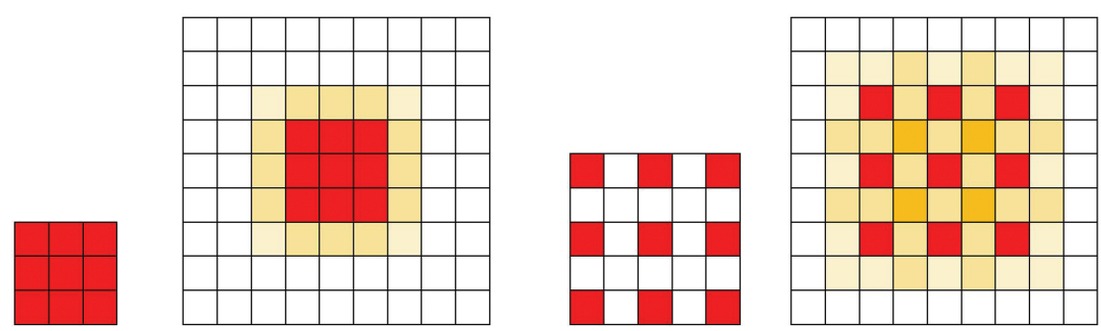


图 2　卷积操作和空洞卷积操作示意图

Fig.2　An Illustration of Convolutional Operator and Dilated Convolutional Operator

2　本文研究方法

针对无人机影像中目标检测存在的问题,本文提出了一种多尺度空洞卷积的无人机影像目标检测方法,旨在通过空洞卷积和金字塔影像相结合的方法提升目标检测网络提取特征的感知域大小,增强网络对于背景、目标分布情况的分析能力,提升目标检测网络在无人机影像中的目标检测精度。同时,因为空洞卷积不需要增加网络的参数,检测更方便、快捷,这对于无人机影像的后续应用有较大意义。

多尺度空洞卷积的无人机影像目标检测方法流程示意图如图 3 所示,其主要以两阶段目标检测方法作为基础网络,整个网络由骨干网络、多尺度空洞卷积特征提取、候选区域提取、分类与定位 4 部分组成:(1)骨干网络基于现有的特征提取网络(如 ResNet 网络[37]等),对输入图像进行常规特征抽象和提取,获得深度特征。(2)多尺度空洞卷积特征提取网络部分是通过对深度特征进行一系列空洞卷积来获取图像多个尺度特征。(3)候选区域提取网络根据选择的基础算法的结构进行设计,完成候选区域的提取。(4)对候选区域进行分类和回归,完成整个多尺度空洞卷积的无人机影像目标检测。除了骨干网络外,候选区域提取、分类与定位部分也可以应用在现有的目标检测方法中,也就是说,可以修改成任意网络结构作为本文方法的后两步。整个网络结构主要通过创新性的多尺度空洞卷积特征提取模块,在现有的

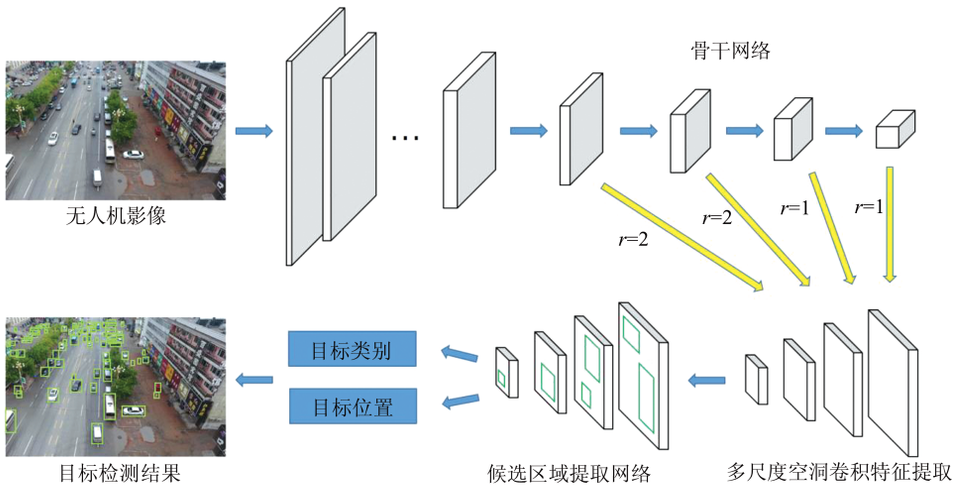


图3 多尺度空洞卷积的无人机影像目标检测方法流程示意图

Fig.3 Diagram of Multi-scale Dilated Convolutional Neural Network for Object Detection in UAV Image

方法基础上对特征的感知域进行扩大，进一步提升网络对无人机影像中目标的检测能力。

本文提出的多尺度空洞卷积特征提取网络以骨干网络提取得到的4层特征层作为输入，设第i个特征为$X_i(i=1,2,3,4)$，其特征高、宽和通道数分别为H_i、W_i、T_i。i越小，代表其特征越深，为高度、宽度更小、通道数更多的特征层，其具有更高维的特征表达；而i越大，特征层越浅，代表更低维的特征。经过多尺度空洞卷积后，得到特征X_i^d的计算公式为：

$$\begin{cases} X_i^d = C[D(X_i, r_i) \oplus L(X_{i-1}^d)], i>1 \\ X_1^d = X_1, i=1 \end{cases} \tag{1}$$

式中，$D(\cdot)$为扩张率为r_i的空洞卷积操作；$L(\cdot)$是空洞卷积第$i-1$层求出的特征的上采样操作函数，其通过最邻近值的采样方式，将特征的$\{H_{i-1}, W_{i-1}\}$扩大两倍；$\oplus$代表逐元素求和操作[38]；$C(\cdot)$是一系列卷积操作；将输出特征通道数T_i设为256通道，即$T_i^d=256$。

候选区域提取和分类与定位网络部分使用现有的方法，本文采用两阶段的检测网络，在实验部分采用了Faster R-CNN和Cascade R-CNN两种网络作为基础网络，分别加入了多尺度空洞卷积进行实验。虽然候选区域提取和分类与定位网络部分细节随着基础检测方法的不同有一定的差异，但都是以真实值的定位回归和分类为主体。对于每个目标，设该目标的位置预测值p由其中心点坐标(p_x, p_y)、预测矩形框高p_h、宽p_w和预测类别p_c表示，通过计算p和其对应的真值g的偏移量Δ进行网络损失函数的计算。对于每一个p，其所对应的真值为交集与并集的比(intersection over union，IoU)[13,16]最大的真值的位置。IoU的计算通过p和g所在区域面积的交集除以并集得到。同时，网络训练时的偏移量计算公式为：

$$\Delta_x = (g_x - p_x)/p_x \tag{2}$$

$$\Delta_y = (g_y - p_y)/p_y \tag{3}$$

$$\Delta_w = \ln(g_w/p_w) \tag{4}$$

$$\Delta_h = \ln(g_h/p_h) \tag{5}$$

在网络训练过程中，损失函数由类别损失函数和预测坐标损失函数两部分[17]组成，具体的损失函数定义由不同的基础算法类型决定。随着损失不断地进行反向传播，整个网络在训练中逐渐收敛，获得最终的目标检测结果。

3 实验与分析

3.1 实验数据

本文以公开数据集VisDrone[6]为实验数据，进行基础实验和多尺度空洞卷积的无人机影像目标检测实验。VisDrone数据集是无人机影像目标检测中使用最多、数据量最大的数据集之一，基准数据集包含265 228帧和10 209幅静态图像组成的400个视频片段。其针对目标检测设计的数据集包含安装在不同无人机型号上的摄影机采集的影像，涵盖了不同地理位置、不同地理环境、不同天气条件下的影像。同时，不同影像之间，采集时的无人机姿态、拍摄角度、拍摄高度等差异明显，存在大量不同角度、不同尺度和遮挡情况的目标，对实验有很好的指导意义。该数据集关注行人和车辆类别的目标，并采集了不同分布(包含稀疏和拥挤场景)情况下的目标。

在目标类别上,详细标注了10类目标,分别为:运行中的行人、静止状态的人、自行车、小轿车、货车、卡车、三轮车、遮阳蓬三轮车、公共汽车和摩托车。目标类别标记精细,便于目标检测实验的评价。

3.2　评价指标

在评价指标上,本文沿用目标检测领域中的常见评价指标[11-12],主要分为精确度指标和召回率指标两类。精确度指标计算了在不同IoU阈值θ设定下的精确度计算结果,包括P、P_{50}、P_{75}、P_s、P_m、P_l等6个指标。对于精确度指标,其计算公式为$P=[\mathrm{TP}/(\mathrm{TP}+\mathrm{FP})]\times 100\%$,其中TP表示检测正确的预测值个数,FP表示检测错误的预测值个数。同时,设定阈值θ,表示当预测的目标位置和真值所在区域的IoU大于θ时,判定该预测值为检测正确。在精度评价指标中,P_{50}和P_{75}分别设定$\theta=0.50$和$\theta=0.75$时的精确度计算结果,而P、P_s、P_m和P_l的阈值设定为θ为0.50～0.95,每隔0.05取值情况下的精确度计算结果的平均值。P_s、P_m和P_l则依据文献[12]中的目标尺寸分类,分别表示小尺寸目标、中等尺寸目标和大尺寸目标的精确度计算结果。在召回率指标中,本文主要采用R_1、R_{10}和R_{100}这3个评价指标,分别表示规定每张图片仅检测出1个、10个、100个目标时,实验的总召回率计算结果。对于召回率指标,其计算公式为$R=[\mathrm{TP}/(\mathrm{TP}+\mathrm{FN})]\times 100\%$,其中FN表示没有检测出来的真值数目。

3.3　实验设置

本文主要以Faster R-CNN和Cascade R-CNN两种网络作为基础网络(即表1中的FR-CNN实验和表2中的CR-CNN实验),分别加入多尺度空洞卷积进行实验(即表1中的FR-Ours实验和表2中的CR-Ours实验)。值得注意的是,基础实验Faster R-CNN和Cascade R-CNN均加入了FPN[16]网络作为对照,以更加有效地反映空洞卷积对于无人机影像目标检测的有效性。每组实验以与其对应的基础网络实验作为对照实验。在骨干网络上,统一选择了ResNet-50[37]。针对输入的VisDrone影像,实验统一将图片长宽分辨率设置为1 200×675像素,作为网络中图像的输入大小。在扩张率选择上,以图3所示的方式将r_1和r_2设置为1,将r_3和r_4设置为2进行空洞卷积操作。本文所进行的所有实验均在同一台式设备上完成,其主板配置为Intel(R) Core(TM) i7-9800X CPU @ 3.80 GHz,并配备2个11 GB显存的NVIDIA Geforce RTX 2080ti显卡,操作系统为Ubuntu 16.04版本,所有程序基于PyTorch平台上公开的Open MMLab Detection[38]框架实现。同时,在实验训练的过程中,网络在2台显卡上同时训练,每台显卡上的批量处理图片数量设置为4,网络优化器选择随机梯度下降方法,并设置学习率为0.02。

表1　以Faster R-CNN为基础网络的实验结果/%

Tab.1　Experimental Results Based on Faster R-CNN Method/%

实验	精确度						召回率		
	P	P_{50}	P_{75}	P_s	P_m	P_l	R_1	R_{10}	R_{100}
FR-CNN	17.4	31.3	17.4	7.9	27.1	34.3	7.8	23.5	28.4
FR-Ours	17.5	31.5	17.7	8.1	27.2	36.0	7.9	23.8	28.7

表2　以Cascade R-CNN为基础网络的对比实验结果/%

Tab.2　Experimental Results Based on Cascade R-CNN Method/%

实验	精确度						召回率		
	P	P_{50}	P_{75}	P_s	P_m	P_l	R_1	R_{10}	R_{100}
CR-CNN	18.4	30.9	19.5	8.4	28.5	36.1	8.2	23.8	28.2
CR-Ours	18.5	31.1	19.6	8.5	28.6	37.9	8.3	24.1	28.5

为了进一步验证网络对实验结果的精度和召回率的提高,本文还进行了一组对输入图像预处理后的基础实验和对比实验。对于无人机影像,由于存在大量目标在网络特征图中过小的情况,失去了重要的纹理结构特征,因此,对输入图像进行裁剪,能够在一定程度上提高实验结果,检测出更多目标。为了验证本文方法在输入影像为裁剪后图像(图像更小,目标区域所占比例更大)时的有效性,本文进一步进行了实验。实验同样在基础网络(即裁剪后的Faster R-CNN加入FPN网络,本文以Crop_Base表示该组实验)和加入了多尺度空洞卷积的方法(以Crop_Ours表示该组实验)下进行。裁剪过程将原有图片全部裁剪成为800×800像素大小,图片与图片之间的重叠度设置为最少100像素。裁剪前后的训练图像数量大大增加,由6 471幅训练图像增加到23 567幅。

3.4　实验结果与分析

以加入了FPN[19]网络的Faster R-CNN[15]和Cascade R-CNN[22]作为基础实验,以不加FPN而加入多尺度空洞卷积方法作为本文实验进行多组对比实验。同时,本文还进行了输入影像预处

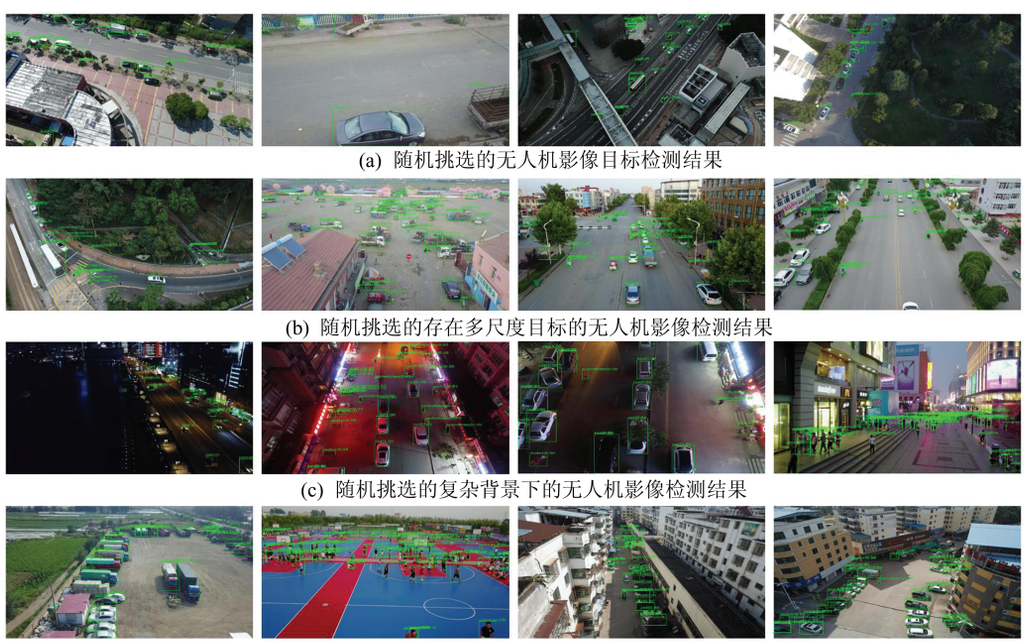
理裁剪后的附加实验，表1～3分别展示了3组定量实验结果。从3组结果可以发现，本文方法在公开数据集VisDrone数据集[6]上获得了相对对比实验更好的实验精确度和召回率，在3组实验中，几乎所有的评价指标下，本文方法的实验结果都相对基础实验有了一定的提升，这证明了本文方法的有效性和优越性。同时，对比不同尺寸影像上的评价结果P_s、P_m和P_l可以发现，本文提出的多尺度空洞卷积的方法对于不同尺寸影像中的目标检测精度均有一定程度的提升。

表3 设置裁剪过影像作为输入影像的实验结果/%

Tab.3 Experimental Results with Cropped Images/%

实验	精确度						召回率		
	P	P_{50}	P_{75}	P_s	P_m	P_l	R_1	R_{10}	R_{100}
Crop_Base	19.3	34.7	19.1	9.5	29.4	36.8	11.7	28.1	31.5
Crop_Ours	19.4	34.8	19.4	9.5	29.2	39.2	11.7	28.2	31.7

为了更好地理解本文方法在无人机影像目标检测中的意义，本文进行了定性实验。图4(a)中展示了一些随机挑选的VisDrone数据集上基于CR-Ours方法的目标检测结果。其中，绿色矩形框标记了网络预测的目标位置，每个矩形框上的标签展示了预测目标的分类结果和置信度。由于无人机影像存在多角度(鸟瞰视角、倾斜视角、水平视角)、不同飞行高度的特点，这导致了无人机影像上的目标出现分布情况差异、尺寸变化大、遮挡现象严重等问题。同时，因为无人机拍摄时间、地点的不同，无人机影像的背景情况和目标特征等具有较大差别。从检测结果(见图4(b)～4(d))可以发现，本文方法可以检测到无人机影像中的多尺度目标(见图4(b)，无人机视角倾斜导致目标尺寸差异较大)、复杂背景下目标(如图4(c)列举了夜晚复杂背景下的影像)和存在遮挡情况的目标(如图4(d)影像中的目标存在大量遮挡情况)，在人眼观察下具有较好的检测定位和分类结果，这进一步证实了本文方法的有效性。



(a) 随机挑选的无人机影像目标检测结果

(b) 随机挑选的存在多尺度目标的无人机影像检测结果

(c) 随机挑选的复杂背景下的无人机影像检测结果

(d) 随机挑选的存在遮挡情况的无人机影像目标检测结果

图4 多尺度空洞卷积的无人机影像目标检测可视化结果

Fig.4 Visualization Results of the Multi-scale Dilated Convolutional Neural Network for Object Detection in UAV Image

4 结 语

无人机飞行器具有多变的飞行高度、可变的视角，飞行速度快，无人机影像上的目标分布情况差异大，并且目标尺度多样，背景复杂，存在大量遮挡等，这为无人机影像上的目标检测带来了困难。针对以上难题，本文提出了一种多尺度空洞卷积的无人机影像目标检测方法。该方法在现有的目标检测算法的基础上，引入多尺度空洞卷积模块到无人机影像目标检测应用中，通过空洞卷积的设计，增大检测网络的视野感知域，提高网络对无人机影像中的目标分布情况、尺寸差异等特点的学习能力，进一步提升网络对无人机中多尺度、复杂背景、存在遮挡情况的目标的检测精度。实验结果证明，本文提出的方法在不增加网络参数的情况下提升了无人机影像上的目

标检测精确度和召回率,具有一定的有效性和鲁棒性。

参 考 文 献

[1] Zhang Liting. The Applications of UAV in the Field of Ship Pilotage[J]. *China Ports*, 2016(9):50-51(张立庭. 无人机在船舶引航领域的应用[J]. 中国港口, 2016(9):50-51)

[2] Fan Bangkui, Zhang Ruiyu. Unmanned Aircraft System and Artificial Intelligence [J]. *Geomatics and Information Science of Wuhan University*, 2017, 42(11):1 523-1 529(樊邦奎, 张瑞雨. 无人机系统与人工智能[J]. 武汉大学学报·信息科学版, 2017,42(11):1 523-1 529)

[3] Li Deren, Li Ming. Research Advance and Application Prospect of Unmanned Aerial Vehicle Remote Sensing System [J]. *Geomatics and Information Science of Wuhan University*, 2014, 39(5): 505-513(李德仁,李明. 无人机遥感系统的研究进展与应用前景[J]. 武汉大学学报·信息科学版, 2014, 39(5): 505-513)

[4] Jin Wei, Ge Hongli, Du Huaqiang, et al. A Review on Unmanned Aerial Vehicle Remote Sensing and Its Application [J]. *Remote Sensing Information*, 2009(1):88-92

[5] Saifa F M S, Prabuwono A S, Mahayuddin Z R. Real Time Vision Based Object Detection from UAV Aerial Images: A Conceptual Framework[J]. *Communications in Computer and Information Science*, 2013, 376:265-274

[6] Du Dawei, Zhu Pengfei, Wen Longyin, et al. VisDrone-DET2019: The Vision Meets Drone Object Detection in Image Challenge Results [C]. IEEE International Conference on Computer Vision (VisDrone Workshop), Seoul, Korea,2019

[7] Li Yansheng, Zhang Yongjun, Yu Jingang, et al. A Novel Spatio-Temporal Saliency Approach for Robust Dim Moving Target Detection from Airborne Infrared Image Sequences[J]. *Information Sciences*, 2016, DOI: 10.1016/j.ins.2016.07.042

[8] Wang Xiaohua, Zhang Cong, Li Cong, et al. Unmanned Aerial Vehicles Target Detection Based on Bio-inspired Visual Attention[J]. *Aeronautical Science and Technology*, 2015(11):78-82(王晓华, 张聪, 李聪,等. 基于仿生视觉注意机制的无人机目标检测[J]. 航空科学技术, 2015(11):78-82)

[9] Kalantar B, Mansor S B, Halin A A, et al. Multiple Moving Object Detection from UAV Videos Using Trajectories of Matched Regional Adjacency Graphs [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(9):5 198-5 213

[10] Ma Huangte, He Yongjie, Wang Chunmei, et al. Research on Human Detection and Tracking Technology Based on UAV Vision[J]. *Computer Technology and Development*, 2018, 28(10): 115-118(马皇特, 贺永杰, 王春梅,等. 基于无人机视觉的人体检测跟踪技术研究[J]. 计算机技术与发展, 2018, 28(10):115-118)

[11] Deng Jia, Dong Wei, Socher R, et al. ImageNet: A Large-Scale Hierarchical Image Database [C]. The IEEE Conference on Computer Vision and Pattern Recognition, Miami, Florida, USA, 2009

[12] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: Common Objects in Context [C]. The IEEE Conference on Computer Vision and Pattern Recognition, Columbus, Ohio, 2014

[13] Wang Caiyun. Research Progress in Object Detection[C]. The 23rd Annual Conference of New Network Technology and Application in 2019, China Computer Users Association Network Application, Hefei, Anhui, 2019(王彩云. 目标检测的研究进展[C]. 中国计算机用户协会网络应用分会2019年第23届网络新技术与应用年会,合肥, 2019)

[14] Girshick R, Donahue J, Darrell T, et al. Faster R-CNN: Region-Based Convolutional Networks for Accurate Object Detection and Segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 38(1):141-158

[15] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, 37(9):1 904-1 916

[16] Girshick R. Fast R-CNN [C]. IEEE International Conference on Computer Vision, Santiago, Chile, 2015

[17] Ren Shaoqing, He Kaiming, 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):1 137-1 149

[18] Kong Tao, Yao Anbang, Chen Yurong, et al. HyperNet: Towards Accurate Region Proposal Generation and Joint Object Detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, Nevada, USA, 2016

[19] Lin T Y, Dollar P, Girshick R, et al. Feature Pyramid Networks for Object Detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, USA, 2017

[20] Yang Bin, Yan Junjie, Lei Zhen, et al. CRAFT Objects from Images[C]. IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, Nevada, USA, 2016

[21] Dai Jifeng, Li Yi, He Kaiming, et al. R-FCN: Object Detection via Region-Based Fully Convolutional Networks[C]. Conference and Workshop on Neural Information Processing Systems, Barcelona, Spain, 2016

[22] Cai Zhaowei, Vasconcelos N. Cascade R-CNN: Delving into High Quality Object Detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, USA, 2017

[23] Shrivastava A, Gupta A, Girshick R. Training Region-based Object Detectors with Online Hard Example Mining[C]. IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, Nevada, USA, 2016

[24] Bodla N, Singh B, Chellappa R, et al. Soft-NMS: Improving Object Detection with One Line of Code[C]. The European Conference on Computer Vision, Zurich, Switzerland, 2014

[25] 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, Nevada, USA, 2016

[26] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger[C]. IEEE Conference on Computer Vision and Pattern Recognition, Hawaii,USA, 2017

[27] Liu Wei, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector[C]. The European Conference on Computer Vision, Amsterdam, Netherlands, 2016

[28] Fu Chengyang, Liu Wei, Ranga A, et al. DSSD: Deconvolutional Single Shot Detector[C]. IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, USA, 2017

[29] Shen Zhiqiang, Liu Zhuang, Li Jianguo, et al. DSOD: Learning Deeply Supervised Object Detectors from Scratch[C]. IEEE International Conference on Computer Vision, Venice, Italy, 2017

[30] Wang Xiaoliang, Cheng Peng, Liu Xinchuan, et al. Fast and Accurate, Convolutional Neural Network Based Approach for Object Detection from UAV[C]. The 44th Annual Conference of the IEEE Industrial Electronics Society, Washington D C, USA, 2018

[31] Bazi Y, Melgani F. Convolutional SVM Networks for Object Detection in UAV Imagery[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 56(6): 3 107-3 118

[32] Liang Dong, Gao Sai, Sun Han, et al. UAV Detection in a Motion Camera Combining Kernelized Correlation Filters and Deep Learning[J]. *Acta Aeronautica et Astronautica Sinica*, 2020, DOI: 10.7527/S1000-6893.2020.23733(梁栋, 高赛, 孙涵,等. 结合核相关滤波器和深度学习的运动相机中无人机目标检测[J]. 航空学报, 2020, DOI: 10.7527/S1000-6893.2020.23733)

[33] Kapania S, Saini D, Goyal S, et al. Multi Object Tracking with UAVs Using Deep SORT and YOLOv3 RetinaNet Detection Framework[C]. The 1st ACM Workshop on Autonomous and Intelligent Mobile Systems, Bangalore, India,2020

[34] Aguilar W G, Quisaguano F J, Rodríguez G A,et al. Convolutional Neuronal Networks Based Monocular Object Detection and Depth Perception for Micro UAVs[C]. International Conference on Intelligent Science and Big Data Engineering, Xiamen, China, 2018

[35] Yu F, Koltun V. Multi-Scale Context Aggregation by Dilated Convolutions[C]. The International Conference on Learning Representations, San Juan, Puerto Rico, 2016

[36] Qu Changbo, Jiang Siyao, Wu Deyang. Multiscale Semantic Segmentation Network Based on Cavity Convolution[J]. *Computer Engineering and Applications*, 2019, 55(24): 91-95(曲长波, 姜思瑶, 吴德阳. 空洞卷积的多尺度语义分割网络[J]. 计算机工程与应用, 2019,55(24): 91-95)

[37] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Deep Residual Learning for Image Recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, Nevada, USA, 2016

[38] Pang Jiangmiao, Chen Kai, Shi Jianping, et al. Libra R-CNN: Towards Balanced Learning for Object Detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, Los Angeles, USA, 2019

Multi-scale Dilated Convolutional Neural Network for Object Detection in UAV Images

ZHANG Ruiqian[1]　*SHAO Zhenfeng*[2]　*PORTNOV Aleksei*[3]　*WANG Jiaming*[2]

1　School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China
2　State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China
3　Moscow State University of Geodesy and Cartography, Moscow 105064, Russia

Abstract: As a new type of remote sensing sensor, unmanned aerial vehicle (UAV) has been used in various fields such as medical treatment, transportation, environmental monitoring, disaster warning, animal protection and military increasingly. Since UAV images are acquired from multiple flying altitudes, perspectives with high speed, objects in UAV images have various scales and perspectives with different distributions, which brings a series of problems to object detection in UAV images. To address these problems, we propose an object detection method based on multi-scale dilated convolutional neural network. It improves existing detection methods by a creative multi-scale dilated convolutional module which facilitates the whole network to learn deep features with increased field of view perception and further improves the performance of object detection in UAV images. We adopt three comparative experiments on base network and our proposed method. And experimental results show that our proposed network has a high precision and recall for object detection in UAV images. Moreover, objects are detected with high performance in multiple perspectives, various scales and complex backgrounds, which indicates the effectiveness and robustness of our method. Object detection in UAV image is significant in both civil and military fields. However, existing methods are limited with objects in multiple perspectives, scales and backgrounds. Our proposed method improves the performance of existing networks by dilated convolutional operator. Experimental results demonstrate the effectiveness and robustness of the proposed method.

Key words: multi-scale network; dilated convolutional neural network; UAV images; object detection

First author: ZHANG Ruiqian, PhD, specializes in remote sensing image processing and object detection. E-mail: zhangruiqian@whu.edu.cn
Corresponding author: SHAO Zhenfeng, PhD, professor. E-mail: shaozhenfeng@whu.edu.cn

Foundation support: Strategic Special Project of International Cooperation in Science and Technology Innovation, the National Key Research and Development Plan(2016YFE0202300); Chinese Academy of Engineering Consulting Research Project (2020ZD16); the National Natural Science Foundation of China(41771454); Hubei Province Natural Science Foundation Planned Innovation Group Project (2018CFA007).

引文格式:ZHANG Ruiqian, SHAO Zhenfeng, PORTNOV Aleksei, et al. Multi-scale Dilated Convolutional Neural Network for Object Detection in UAV Images[J]. Geomatics and Information Science of Wuhan University, 2020, 45(6):895-903. DOI:10.13203/j.whugis20200253
(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(6):895-903. DOI:10.13203/j.whugis20200253)