改进 RetinaNet 特征融合方式的无人机检测方法

马田源,孙 涵
(南京航空航天大学 计算机科学与技术学院,江苏 南京 211106)

摘 要:常见的目标检测方法如 R-CNN 系列方法、YOLO 系列方法和 RetinaNet 等,虽然在通用数据集上有着不俗表现,但是在无人机小目标检测任务中的表现却不尽如人意。分析原因是这些方法采用的传统特征金字塔融合网络 FPN 存在着上采样失真、语义信息衰减以及深层语义差异的问题,导致目标检测网络未能获取足够有辨识度的特征,致使其在无人机小目标检测任务中表现不佳。对此,该文提出了一种基于 RetinaNet 网络的多尺度特征融合方法。该方法采用像素洗牌上采样模块构建了像素洗牌融合网络,并且引入了深层语义增强模块,可在多尺度特征融合阶段提升无人机小目标在网络浅层的特征表示效果,进而提升深度神经网络对无人机小目标的检测性能。最后在自建蜂群无人机数据集上的实验结果显示,引入新的特征融合方法之后,网络对无人机的检测精度达到 91.2%,提升了 1.7%。

关键词:小目标检测;无人机检测;RetinaNet;特征融合;深度学习
中图分类号:TP301 **文献标识码**:A **文章编号**:1673-629X(2022)12-0103-07
doi:10.3969/j.issn.1673-629X.2022.12.016

An Improved Feature Fusion Method for Drone Detection Based on RetinaNet Extraction

MA Tian-yuan, SUN Han
(School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

Abstract: Although most object detection methods like R-CNN, YOLO and RetinaNet have outstanding performance in general datasets, but their performance in small drone detection task is not satisfactory. The reason is that the traditional feature pyramid fusion network FPN adopted by these methods has problems of up-sampling distortion, semantic information attenuation and deep semantic difference, which leads to the failure of object detection network to acquire enough identifiable features, resulting in its poor performance in small drone detection task. Therefore, we propose a multi-scale feature fusion method based on RetinaNet network. In this method, the up-sampling module of pixel shuffle unit is used to construct the pixel shuffle feature fusion network, and the high-level semantic enhancement module is introduced to improve the feature representation effect of small drone in the shallow layer of the network in the stage of multi-scale feature fusion. Then the deep neural network can improve the detection performance of small drone. Finally, experiments on the self-built drone dataset show that the network detection accuracy of drone can reach 91.2%, which increases by 1.7%.

Key words: small object detection; drone detection; RetinaNet; feature fusion; deep learning

0 引 言

近年来,随着无人机制造成本和使用难度的不断降低,无人机得以迅速应用到各行各业,普通人也可以很容易操控无人机。这给人们的生产和生活带来了很大方便,但与此同时,无人机“黑飞”也给空中交通、公共安全等领域的监管带来了巨大挑战。因此,作为无人机监管的重要环节,对无人机小目标的检测就成为了目前亟待解决的关键问题。

目前无人机检测的方案有很多种,基于雷达、声、无线电、光电设备等。这些利用无人机的物理属性对无人机进行定位的技术非常常见,但是这些检测方案往往需要非常昂贵的设备和严格的配置。基于视觉的方法则成本较低,配置相对简单,且易于部署。

而今,随着深度学习技术的发展,其在计算机视觉领域取得了大量的突破,优秀的工作不断涌现出来,早期的工作例如 Fast R-CNN[1]、Faster R-CNN[2]、

收稿日期:2021-12-08 **修回日期**:2022-04-12
基金项目:国防科技创新特区项目(XX)
作者简介:马田源(1996-),男,硕士,CCF 会员(B5362G),通讯作者,研究方向为计算机视觉;孙 涵(1978-),男,博士,副教授,CCF 南京秘书长(33361M),研究方向为计算机视觉。

YOLO[3]等,在各个通用数据集上均取得了不错的成绩。后来何恺明等人提出了特征金字塔融合网络(Feature Pyramid Network,FPN)[4],该网络由自底向上和自顶向下两条通路以及一条横向连接三部分构成。首先,自底向上进行特征提取;然后,横向连接对齐通道数目;接着,自顶向下进行特征融合;最后,得到具有多尺度特征的特征图进行预测。该方法通过将语义信息从深层传播到浅层来实现多尺度的特征融合,使深层丰富的语义信息得以传递到浅层,解决了浅层语义信息缺失的问题,提高了网络对不同尺度目标,尤其是对小目标的检测性能。因此 FPN 在各种通用的目标检测方法中普遍被采用,例如 RetinaNet[5]、FCOS[6]等工作。然而,无人机小目标与通用尺度的目标不同,它尺度极小,通常占整张图像的像素比低于 0.1%,且小型无人机镂空的外观,使其与背景融为一体,特征极不明显,检测难度很高。

该文总结了 FPN 方法在无人机目标检测任务中表现不佳的三点原因:(1)原始 FPN 上采样所使用的最近邻插值法会造成特征偏移失真,这些偏移会对无人机小目标检测造成很大的影响;(2)无人机小目标通常在网络浅层被检测出来,而网络浅层特征的语义信息缺失问题导致网络对小目标检测性能不佳。因此原始的 FPN 方法将深层的语义信息传递到浅层来缓解这个问题,但是在 FPN 的实际操作过程中,通道维度对齐的做法仍会产生大量通道信息衰减,造成深层语义信息向浅层传递不足的问题。对于更需要深层语义信息补充的无人机这类极小目标来说,语义信息不足会产生更大的影响;(3)在 FPN 特征融合时,网络最深层特征只有本层的语义信息,同其他层产生了语义差异,这种差异在一定程度上会影响后续的推理。

为了解决上述问题,基于 RetinaNet 提出了新的多尺度特征融合方式,具体贡献如下:

(1)受到超分辨率领域中效果十分优秀的像素洗牌(Pixel Shuffle,PS)[7]上采样方式的启发,设计了像素洗牌上采样模块(Pixel Shuffle Unit,PSU),将其作为特征融合网络中的上采样方法,解决特征失真问题,并且使深层的语义信息得到充分利用。

(2)设计了像素洗牌特征融合网络(Pixel Shuffle Feature Fusion,PSFF)。采用 PSU 改进了上采样方式,并且重构了特征通道融合流程。降低了深层语义信息向浅层传递过程中的衰减,使深层的语义信息能够有效地传递到浅层,进而增强浅层无人机小目标的特征表示。

(3)增加了深层语义增强模块(High-level Semantic Enhancement,HSE)。在网络深层追加特征提取模块,使网络可以提取到更深层的语义信息,提高网络整体的特征提取能力,同时消除 FPN 最深层同其他层之间的语义差异。

实验结果验证,在自建蜂群无人机数据集上,与其他类似特征融合方式相比,提出的新特征融合方式具有明显优势,基于 RetinaNet 的检测效果提升了1.7%,精度达到了 91.2%。

1 研究现状

1.1 目标检测

通用目标检测方法主要分为两个大方向:一是以 Faster R-CNN[2]、Cascade R-CNN[8]、Grid R-CNN[9]、Sparse R-CNN[10]为代表的二阶段目标检测方法。这类方法首先预测出目标可能出现的候选区域(Region Proposal),再对该区域内的目标进行目标检测和类别预测[11],这类方法精度相对较高,但是模型相对复杂,且推理速度较慢;二是以 YOLO[3,12-13]系列、SSD[14-16]系列、RetinaNet 为代表的一阶段目标检测方法。这类方法没有复杂的预测候选区域的过程,直接根据输入图像预测出目标的类别和检测框的位置,因此这类方法具有速度快、开销低的优点,但是检测精度往往不如二阶段检测方法。

1.2 多尺度特征融合

为了应对不同尺度的目标检测,何恺明等人首先提出了特征金字塔融合网络 FPN[4]。后续对 FPN 进行改进的多种多尺度特征融合方法相继提出。例如 PAFPN[17]采用双向融合,在原有自顶向下融合的基础上,增加了反向传递,缩短了浅层与深层特征之间的信息路径,在一定程度上提高了信息的利用率,让融合效率提高。NAS-FPN[18]在多尺度特征融合网络中,将神经网络搜索出来的不规则拓扑结构作为融合网络,效果优异,但训练成本极高。文献[19]提出了 HRFPN,使特征金字塔在进行下采样的过程中保留尽可能多的细节信息。BiFPN[20]中重复堆叠简化的 PAFPN,同时引入权重参数,来平衡不同尺度的特征信息。文献[21]在显著性检测领域提出了多尺度特征金字塔网格 MFPG 来丰富语义信息。

1.3 无人机目标检测

基于传统方法的无人机检测工作如下:文献[22]采用基于卡尔曼模型的方法动态地对无人机进行检测和跟踪。文献[23]利用无人机本身的移动性、振动性、空间性这三个本身的物理特性来发现并定位无人机。基于深度学习的无人机检测工作如下:文献[24]在 YOLOv3 中引入多尺度的特征图融合方法,以此提升无人机检测精度。文献[25]基于 RetinaNet 构建了无人机目标检测网络,对多旋翼无人机进行检测识别,取得了较好的效果。文献[26]基于 SSD,将高层的特

征引入到浅层,以此增强浅层无人机小目标的特征表示,来提高网络对无人机的检测效果。文献[27]受到人脸检测方法的启发,设计了轻量级迭代的无人机检测网络 TIB-Net,取得了不错的检测效果。

2　模型和方法

2.1　网络整体框架

原始 RetinaNet[5] 在采用 FPN 的基础之上,又引入聚焦损失 Focal Loss 的概念,解决了类不平衡问题,提高了一阶段目标检测方法的精度,同时也保持了较高的检测速度。其网络由三个主要部分构成:一是特征提取的主干网络;二是多尺度特征融合网络;三是分类和边界框回归的检测网络。因为该方法具有较高的检测速度,且结构清晰易于扩展,对小目标有较好的检测效果,所以该文采用该框架作为基本的检测框架。

总体结构如图 1 所示。框架基于 RetinaNet,该文重新设计了新的多尺度特征融合方法。其中 Backbone 部分采用 ResNet50[28] 进行自底向上的特征提取。多尺度特征融合部分采用像素洗牌融合网络 PSFF 进行自顶向下的多尺度特征融合,深层语义增强模块 HSE 继续提取更深层次的语义信息,并且注入到 PSFF 中,让深层的语义信息可以更有效地传递到浅层。最后为两个 FCN 全卷积网络(Fully Convolutional Networks,FCN)[29] 分类分支和检测框回归分支。

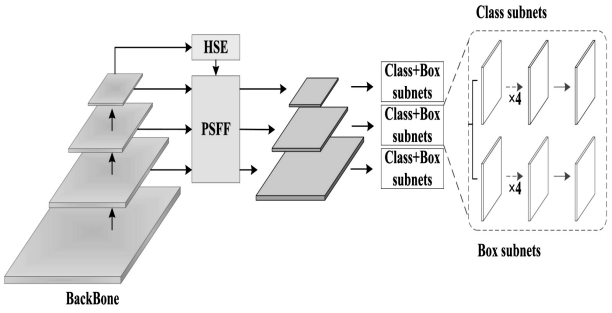

图 1　总体结构

2.2　基于像素洗牌的特征融合网络 PSFF

原始 FPN 网络在自底向上的通路中,随着网络的不断加深,特征通道数会逐层增加,特征图尺寸会逐层减小。在网络的最深层,通道数量最多。例如 ResNet50 的{C2,C3,C4,C5}特征层分别对应{256,512,1 024,2 048}个特征通道,最深层的 C5 通道数量为 2 048 包含最丰富的语义信息。在横向连接中,由于特征融合需要对齐特征维度,所以该步骤将得到的特征层的通道维数统一压缩到 256。在自顶向下的通路中,从顶层开始,进行 2 倍最近邻上采样的同时,与横向连接对齐通道数量的特征层进行特征融合。

可见,原始 FPN 上采样的过程造成了特征失真;在进行横向连接时将所有层的通道维数统一压缩到 256 再进行自顶向下的特征融合的做法,造成了深层特征没有充分地传递到浅层的问题;最深层特征 F5 直接来自 C5 层,没有融合其他层的特征,只具备一层的语义信息,造成了深层同其他层之间的语义差异。

考虑以上因素,该文对特征融合网络进行了重新设计,其中包括两大块:其一,使用像素洗牌上采样模块 PSU 构造了像素洗牌特征融合网络 PSFF。该网络既解决了原始 FPN 上采样方法的缺陷,又可以让深层特征充分地参与到向浅层传递的过程中去,使深层语义信息更有效地传递到浅层,以此提升浅层小目标的特征表示。其二,引入深层语义增强模块 HSE,既解决深层语义差异问题,又可以提取更深层的语义信息,提高网络整体的特征提取能力。

为此,该文首先引入深层语义增强模块 HSE,来补充深层缺失的语义信息;接着,采用像素洗牌上采样模块 PSU,利用深层大量的通道信息进行上采样,在完成上采样的同时,将特征的维度信息转化到空间尺度信息中去,以此将深层的语义信息由低分辨率的特征图带入到高分辨率特征图,既对齐了维度又对齐了尺度;然后,进行逐像素相加和,即特征融合;最后进行去除混叠和通道压缩。具体实现如图 2 所示,其中{C3,C4,C5}为来自主干网络 Backbone 也就是 ResNet50 特征提取网络的第 3、第 4 和第 5 层的特征,{F3,F4,F5}为对应的输出。PSU 为像素洗牌上采样模块。C5 * 为深层语义增强模块 HSE 的输出,Conv-3×3 用来去除混叠效应。图中“64x”代表相对于原始网络输入图像下采样 64 倍的尺度,“64x -> 32x”代表该过程图像尺度由 64x 上采样到 32x,“2 048 -> 256”表示通道数从 2 048 变为 256。“⊕”表示特征图之间逐像素相加。

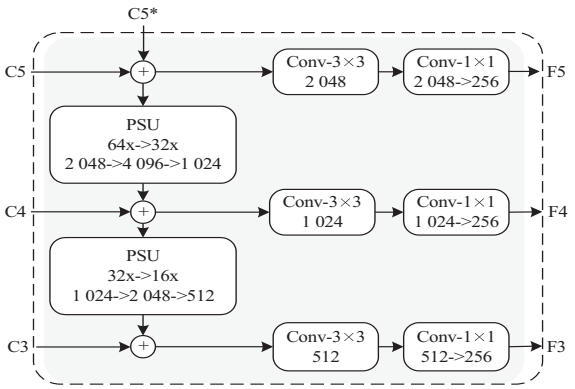

图 2　像素洗牌特征融合网络 PSFF

2.3　像素洗牌上采样模块 PSU

原始的特征金字塔融合网络 FPN 采用最近邻插值的上采样法,该方法在上采样的过程中使用邻近像素值进行填充,简单且不需要复杂计算。但该方法在计算上采样图像像素点对应位置时,对非整数的计算

结果直接向下取整的做法会造成像素偏移,使某一区域的像素值相同,让像素变化不连续,产生锯齿状失真。考虑以上因素,该文采用了像素洗牌[7]上采样方法,该方法具有高效、快速、无参的特性,更重要的是,它可以充分利用主干网络最深层大量的通道信息,将其作为像素洗牌所需的像素信息,进行像素洗牌的上采样。简言之,就是可以把通道维度信息转换成空间信息保留下来。如此,不仅避免了简单最近邻差值特征失真的缺陷,还可以最大程度地让深层丰富的语义信息传递到浅层,让浅层获得到更加丰富的深层语义信息,提升浅层无人机小目标特征表示,进而提升网络对无人机小目标的检测性能。

具体做法如图 3 所示。设上采样倍数为 u ,低分辨率图像的通道数为高分辨率图像通道数的 u^2 倍。默认上采样倍数为 2,最左侧低分辨率图像的通道数为 4c,得到的高分辨率图像的通道数为 1c。该过程通过像素重组,将原始低分辨率图像中的像素按照特定方式重新排列(洗牌),最终组合成大小为原图 u 倍的高分辨率图像,通道数为原图的 $\frac{1}{u^2}$ 。

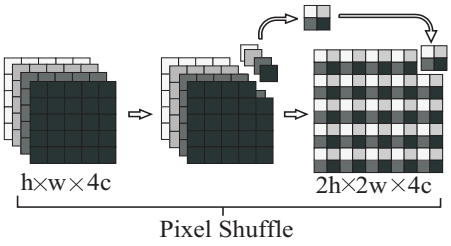

图 3 像素洗牌上采样原理 PS

将该做法整理为公式(1),其中 PS 代表 Pixel Shuffle,是像素洗牌英文名简称,默认扩张因子(Scale)为 2,代表使用像素洗牌方法进行 2 倍上采样,其中输入为 $x' \in R^{h\times w\times 4c}$ 其通道数量为 $4c$,输出为 $y' \in R^{2w\times 2h\times c}$ 其通道数量为 c ,在这个过程中通道数由 $4c$ 减少到 c ,图像尺寸由 $h \times w$ 扩增到 $2h \times 2w$ 。

直观来看,可以在原始 FPN 方法的基础上直接替换上采样方式。首先,将统一压缩到 256 维的特征进行 4 倍通道扩增;接着,采用像素洗牌上采样,通道数还原为 256;最后,与上层特征进行融合。以上作法虽然解决了原始 FPN 上采样特征失真问题,但是并没有解决深层特征大量丢失的问题。因为横向连接的过程特征首先被统一压缩到 256 维,这一造成深层特征大量丢失问题的根源,没有被改变。所以该文未予采用。

考虑到像素洗牌上采样方法具有将通道维度信息转换成空间信息的特质,如此就可以将深层的大量通道作为像素洗牌进行上采样的像素来源,与此同时,解决上采样失真和深层特征大量丢失的问题。该文设计了像素洗牌上采样模块,具体流程如图 4 所示。首先,将输入特征通过卷积核为 1 的卷积进行 2 倍通道扩增;接着,经过像素洗牌进行 2 倍上采样,此时通道维数缩减到 $\frac{1}{2}c$;最后,输出上采样的结果。

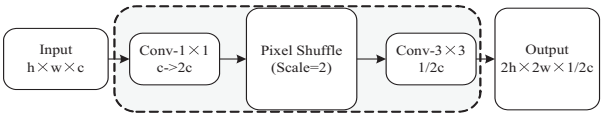

图 4 像素洗牌上采样模块 PSU

整理如公式(2)所示,输入记作 $x \in R^{h\times w\times c}$,输出特征图记作 $y \in R^{2h\times 2w\times \frac{1}{2}c}$,卷积 Conv-1×1 大小是 $1 \times 1 \times 2c$,卷积 Conv-3×3 大小为 $3 \times 3 \times \frac{1}{2}c$ 。

$$y' = \mathrm{PS}(x') \tag{1}$$

$$y = \mathrm{Conv} - 3 \times 3(\mathrm{PS}(\mathrm{Conv} - 1 \times 1(x))) \tag{2}$$

2.4 深层语义增强模块 HSE

在原始的 FPN 中,融合后网络的最深层 F5 层仅包含它本层一个尺度的语义信息,而其他层则融合了多层语义信息,同其他层之间产生了语义差异。对此,该文设计了深层语义增强模块 HSE,旨在向 F5 中注入更深层次的语义信息,消除该层同其他层之间的语义差异问题,同时提高网络整体的特征提取能力。

HSE 具体流程如图 5 所示,采用并行分支结构。第一个分支首先进行全局平均值池化,接着用广播的形式还原特征尺度,以获取全局信息;第二个分支采用瓶颈结构结合 3 × 3 卷积获取局部信息;第三个分支利用最大值池化进行下采样,结合 3 × 3 卷积,然后用 1 × 1 卷积进行通道维度扩增,再通过像素洗牌还原特征尺度,以再次增大感受野来获取更丰富的语义信息。

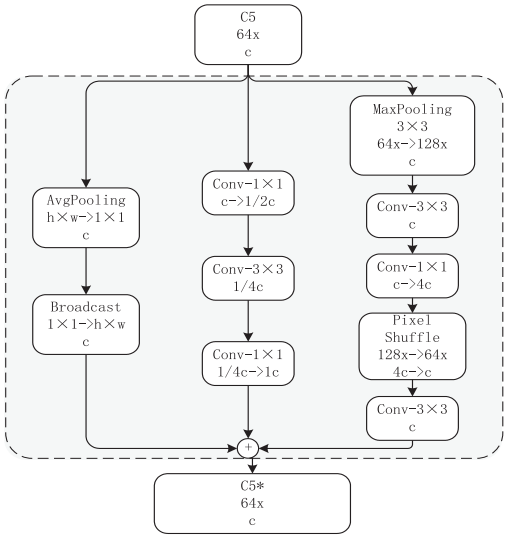

图 5 深层语义增强模块 HSE

最后 HSE 模块以残差形式加入特征融合网络,防止网络退化。在特征融合网络中引入该深层语义增强模块,首先消除了最后一层 F5 与其他层之间的语义差异;其次,在层内形成了多层级的特征提取,提高了网

络整体的特征提取能力;最终为自顶向下的特征融合过程提供了更加丰富的深层语义信息。因此使浅层小目标可以获得更多深层的语义信息,进而提升网络对无人机小目标的特征提取能力。

2.5　损失函数

在训练设置部分,模型的损失是两个部分损失的加和,第一部分为回归子网络的检测框回归损失,采用标准的 Smooth L1 Loss 损失;第二部分为分类子网络的分类损失,采用带平衡因子的聚焦损失(Focal Loss)[5],其形式如公式(3)。

$$FL(p_t) = -\alpha_t (1 - p_t)^{\gamma} \log(p_t) \tag{3}$$

FL 代表 Focal Loss,其中 α_t 为正负样本的加权平衡参数,值越大正样本的权重越大,这里设置为 0.25。γ 为聚焦系数,控制样本权重更新速率,是一个大于 0 的超参数,该文设置为 2。p_t 表示样本属于正样本的概率,$(1 - p_t)^{\gamma}$ 是权重表达式。以上超参数设置均来自 RetinaNet[5] 原文。

Focal Loss 聚焦损失函数就是给简单样本和困难样本分别加上一组权重系数。这个系数跟模型预测的样本属于真实类别的概率相关。对于简单样本,如果模型预测该样本属于真实类别的概率很大,那该样本对于模型来说就是简单样本,此时的 p_t 接近于 1,权重系数接近于 0,如此就会降低简单样本的损失权重。对于困难样本,如果模型预测某样本属于真实样本的概率很小(也可以说这很可能是错误分类),那么该样本对于模型来讲就属于困难样本,此时的 p_t 很小,权重系数接近于 1,如此就会让困难样本的损失得以最大限度地保留。无人机小目标尺度小、特征少、区分度不高导致其难以识别,对于网络来说属于困难样本。而聚焦损失又能够让网络聚焦于困难样本,因此采用该损失函数,有利于网络学习无人机小目标的特征,提高对无人机小目标的检测性能。

3　实验过程与结果

3.1　数据集描述

实验采用自建的蜂群无人机数据集。以往的无人机数据集都是单架次的,即每张图像中只存在一个无人机目标。但真实场景中会有多架无人机同时出现的情况,目前还没有这类公开的数据集。而在图像中剪切粘贴很多架次无人机,这种方式生成的无人机群数据集又有失样本的真实性。因此本次研究专门采集了真实的无人机群数据,并且为保证数据的多样性,涵盖了不同天气条件、不同时间段、不同地点、不同距离以及不同视角的数据。如图 6,其中(a,b,c)分别为晴天、阴天、雾霾天,(a,b,c)分别为中午、下午和傍晚,(b/c,d,a/e/f)分别为城市、乡村、净空,(a/e/f,b/c/d)分别为远距离和近距离图像,(e,f)分别为下面视角和侧面视角,且以上每个场景关注的空域中都有包含 1 ~ 10 架次无人机的无人机编队。所有数据均通过手持摄像头和固定机位的云台摄像头以视频形式采集,然后抽帧获得,最后对其中 2 843 张图像进行标注,标注采用 VOC 数据集[30] 的标注格式。

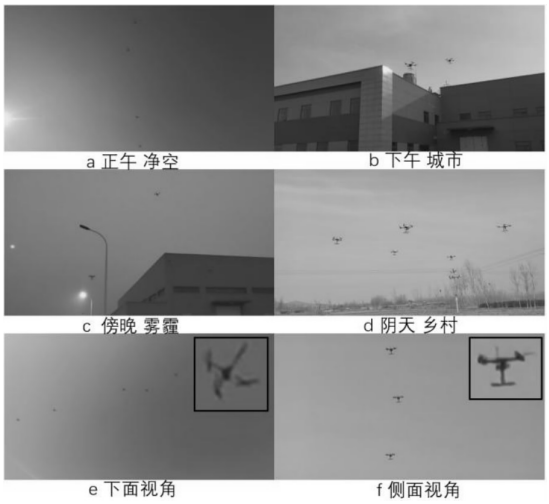

图 6　数据集中多场景无人机样本

3.2　训练过程

实验环境如下:Ubuntu20.04 LTS 系统、Intel i9-9900KCPU×16、内存为 64 GB、两张 NVIDIA 2080Ti 显卡(12 GB 显存)、Pytorch1.7.1、CUDA10.2、CUDNN7.6.5,实验采用的预训练模型是由 Pytorch 官方提供的在 Imagenet 上进行了预训练的模型。

为公平起见,所有的实验结果都是基于以下训练参数:迭代次数为 12、动量为 0.9、批大小为 4、初始学习率为 0.000 5、每训练 4 个 epoch 权重衰减为原来的 0.3 倍。为了降低模型计算开销,所有图像在送入模型前尺寸都被统一调整为 1 333×800。

对于 RetinaNet,主干网络全部采用 ResNet50 作为特征提取网络,其余配置与 RetinaNet 原文配置相同。

3.3　实验结果

3.3.1　评估指标

为了评价该方法的有效性,该文采用 VOC[30] 数据集的评价指标,使用 mAP(mean Average Precision)来评价检测器的性能。mAP 是由精确率和召回率得到的。其计算公式如公式(4)和公式(5):

$$AP = \int_0^1 p(r)\,\mathrm{d}(r) \tag{4}$$

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{5}$$

其中,p(Precision)表示精确率、r(Recall)表示召回率,AP(Average Precision)表示平均精度,是计算在不同召回率下精度的平均值,N 表示类别数目,由于该文无人机检测任务目标类别数为 1,因此这里的 N 值

设为 1。

3.3.2 实 验

为验证该方法的有效性,表 1 展示了模块消融实验。分别使用深层语义增强模块 HSE 和像素洗牌融合网络 PSFF。可以看出在只引入深层语义增强模块 HSE 时模型性能提升 0.7%,只引入像素洗牌融合网络 PSFF 时,模型性能提升 0.9%,在两者都加入网络时,模型精度提升到 91.2%,提升了 1.7%。

表 1 模块消融实验

Methods	PSFF	HSE	mAP/%
RetinaNet			89.5
RetinaNet		√	90.2
RetinaNet	√		90.4
RetinaNet	√	√	91.2

如表 2 所示,其中 Neck 表示该方法的多尺度特征融合部分,Backbone 表示主干网络部分。第 1 行为基准实验,第 2、3 行分别为使用 HRFPN 和 PAFPN 特征融合方法的实验结果,最后一行中的 Ours 代表该文提出的多尺度特征融合方法。可以看出,同其他特征融合方法相比,使用该文提出的新的特征融合方式具有更好的结果,对无人机小目标检测的精度更高。

表 2 对比实验

Methods	Backbone	Neck	mAP/%
RetinaNet	Resnet50	FPN	89.5
RetinaNet	Resnet50	HRFPN[19]	90.2
RetinaNet	Resnet50	PAFPN[17]	90.4
Cascade R-CNN[8]	Resnet50	FPN	90.5
Sparse R-CNN[10]	Resnet50	FPN	90.6
FCOS[6]	Resnet50	FPN	90.0
YOLOv3[12]	DarkNet53	FPN	86.2
YOLOv4[13]	CSPDarkNet53	FPN	87.1
YOLOv5	YOLOv5-x	FPN	89.0
TIB-Net[27]	-	-	90.2
TIB-Net++[31]	-	-	90.5
RetinaNet(ours)	Resnet50	Ours	91.2

另外,该文也与使用其他方法进行无人机检测的方法进行对比。这里由于缺乏公开的代码,使用当前最优秀的通用目标检测方法进行对比实验。表 2 下半部分展示的对比实验包括:二阶段的方法 Cascade R-CNN、Sparse R-CNN,一阶段的方法 FCOS、不同版本的 YOLO,同样专门进行无人机检测的 TIB-Net、TIB-Net 的改进方法(表 2 中的 TIB-Net++)。从结果上看,文中方法都优于它们。为了展示模型的性能,图 7 将部分场景的检测结果进行展示,其中右上角为图中无人机目标的放大图。可以看出,RetinaNet 在引入该文提出的新特征融合方式之后,能够在不同场景、不同天气、不同光照条件下得到较好的检测结果,网络具有鲁棒性,并且检测结果十分精准。

新的特征融合方法在特征融合的过程中缓解了上采样造成的失真、深层语义信息衰减以及深层语义差异问题。所以,网络可以提取到更加丰富的语义信息,并且深层的语义信息可以更好地丰富浅层小目标的特征表示。因此,该网络模型可以根据无人机所处的背景调整其置信度,有效地过滤掉一些不合理的误检情况,进而提升无人机小目标在各种场景的检测性能。

对于检测失败的情况,例如,图 7 误检中误将电线杆的一部分误检为无人机,从展示的放大细节来看,电线杆支架部分的外观同无人机确实十分接近,所以网络将其误检为无人机目标。图 7 漏检中,漏检了一架无人机,原因可能是该无人机由于距离较远,且与背景墙的颜色和线条较为接近,导致漏检。对于这类背景十分复杂的情况,网络的检测效果仍有待提高。

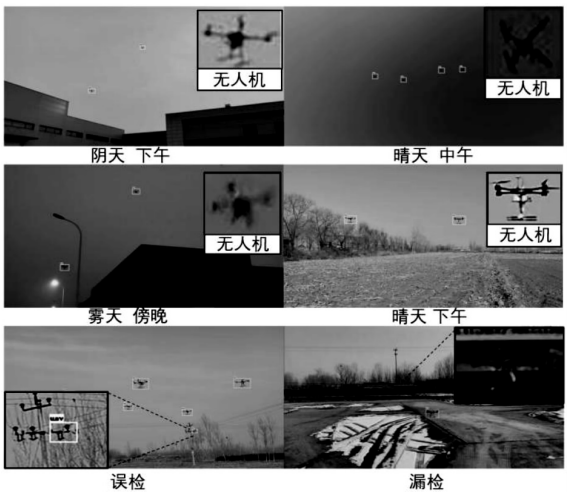

图 7 检测结果

4 结束语

基于 RetinaNet,根据无人机小目标的特点进行了有针对性地改进。首先,通过引入深层语义增强模块 HSE,解决深层语义差异问题,同时又能提高网络整体的特征提取能力;接着,基于像素洗牌上采样模块 PSU 设计了像素洗牌特征融合网络 PSFF,改进了上采样方式造成特征失真的问题,使网络能够将深层的语义信息更有效地向浅层传递,以此增强浅层小目标的特征表示;最后,在自建蜂群无人机数据集上进行实验,验证了提出的方法可以提升无人机小目标的检测效果。

对于接下来的工作,准备从目标之间的关系入手,尝试建立无人机与背景之间、无人机群之间的关联性,以此解决复杂场景中漏检和误检的问题。

参考文献:

[1] GIRSHICK R B. Fast R-CNN[C]//Proceedings of 2015 IEEE international conference on computer vision. Santiago:

IEEE,2015:1440-1448.

[2] REN S,HE K,GIRSHICK R,et al. Faster R-CNN:towards real-time object detection with region proposal networks [J]. Advances in Neural Information Processing Systems, 2015,28:91-99.

[3] REDMON J,DIVVALA S K,GIRSHICK R B,et al. You only look once:unified,real-time object detection[C]// Proceedings of 2016 IEEE conference on computer vision and pattern recognition. Las Vegas:IEEE,2016:779-788.

[4] LIN T,DOLLÁR P,GIRSHICK R B,et al. Feature pyramid networks for object detection[C]//Proceedings of 2017 IEEE conference on computer vision and pattern recognition. Honolulu:IEEE,2017:936-941.

[5] LIN T,GOYAL P,GIRSHICK R B,et al. Focal loss for dense object detection[C]//Proceedings of IEEE international conference on computer vision. Venice:IEEE,2017: 2999-3007.

[6] TIAN Z,SHEN C,CHEN H,et al. FCOS:fully convolutional one-stage object detection[C]//Proceedings of 2019 IEEE/CVF international conference on computer vision. Seoul: IEEE,2019:9626-9635.

[7] SHI W,CABALLERO J,HUSZAR F,et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]//Proceedings of 2016 IEEE conference on computer vision and pattern recognition. Las Vegas:IEEE,2016:1874-1883.

[8] CAI Z,VASCONCELOS N. Cascade R-CNN:delving into high quality object detection[C]//Proceedings of 2018 IEEE conference on computer vision and pattern recognition. Salt Lake City:IEEE,2018:6154-6162.

[9] LU X,LI B,YUE Y,et al. Grid R-CNN[C]//Proceedings of IEEE conference on computer vision and pattern recognition. Long Beach:IEEE,2019:7363-7372.

[10] SUN P,ZHANG R,JIANG Y,et al. Sparse R-CNN:end-to-end object detection with learnable proposals[C]//Proceedings of IEEE conference on computer vision and pattern recognition.[s. l.]:IEEE,2021:1445-1446.

[11] 张泽苗,霍　欢,赵逢禹. 深层卷积神经网络的目标检测算法综述[J]. 小型微型计算机系统,2019,40(9):1825-1831.

[12] FARHADI A,REDMON J. Yolov3:an incremental improvement[C]//Computer vision and pattern recognition. Berlin: Springer,2018.

[13] BOCHKOVSKIY A,WANG C Y,LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[J]. arXiv: 200410934,2020.

[14] LIU W,ANGUELOV D,ERHAN D,et al. SSD:single shot multibox detector[C]//European conference on computer vision.[s. l.]:Springer,2016:21-37.

[15] FU C Y,LIU W,RANGA A,et al. DSSD:deconvolutional single shot detector[J]. arXiv:170106659,2017.

[16] LI Z,ZHOU F. FSSD:feature fusion single shot multibox detector[J]. arXiv:171200960,2017.

[17] LIU S,QI L,QIN H,et al. Path aggregation network for instance segmentation[C]//Proceedings of 2018 IEEE conference on computer vision and pattern recognition. Salt Lake City:IEEE,2018:8759-8768.

[18] GHIASI G,LIN T,LE Q V. NAS-FPN:learning scalable feature pyramid architecture for object detection[C]//Proceedings of IEEE conference on computer vision and pattern recognition. Long Beach:IEEE,2019:7036-7045.

[19] SUN K,XIAO B,LIU D,et al. Deep high-resolution representation learning for human pose estimation[C]//Proceedings of IEEE conference on computer vision and pattern recognition. Long Beach:IEEE,2019:5693-5703.

[20] TAN M,PANG R,LE Q V. EfficientDet:scalable and efficient object detection[C]//Proceedings of 2020 IEEE/CVF conference on computer vision and pattern recognition. Seattle:IEEE,2020:10778-10780.

[21] 张卫明,史彩娟,任弼娟,等. 多尺度特征金字塔网格的显著性目标检测[J/OL]. 小型微型计算机系统:1-10 [2022-03-02]. http://kns.cnki.net/kcms/detail/21.1106.TP.20210517.1534.013.html.

[22] WU Y,SUI Y,WANG G. Vision-based real-time aerial object localization and tracking for UAV sensing system[J]. IEEE Access,2017,5:23969-23978.

[23] 周　伟. 基于 CSI 的无人机检测与 3D 定位研究[D]. 大连:大连理工大学,2018.

[24] 甘雨涛. 卷积神经网络在低空空域无人机检测中的研究[D]. 成都:电子科技大学,2019.

[25] 胡　焱,徐志强,刘文劲,等. 基于 RetinaNet 的低小慢无人机目标识别[J]. 现代计算机,2021(5):66-70.

[26] 刘朋飞,周　海,冯水春,等. 基于改进 SSD 的多尺度低空无人机检测[J]. 计算机工程与设计,2021,42(11):3277-3285.

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

[28] HE K,ZHANG X,REN S,et al. Deep residual learning for image recognition[C]//Proceedings of 2016 IEEE conference on computer vision and pattern recognition. Las Vegas: IEEE,2016:770-778.

[29] LONG J,SHELHAMER E,DARRELL T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2015,39 (4):640-651.

[30] EVERINGHAM M,VAN GOOL L,WILLIAMS C K,et al. The pascal visual object classes (voc) challenge[J]. International Journal of Computer Vision,2010,88(2):303-338.

[31] 杨　健,孙　涵. 基于深度特征提取的无人机检测算法[J]. 计算机技术与发展,2021,31(11):71-75.