ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 39,No. 4,Jul. 2024,pp. 908－921
DOI:10. 16337/j. 1004-9037. 2024. 04. 011

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

高分辨率特征增强的无人机航拍小目标检测

周　璇，葛　琦，邵文泽

（南京邮电大学通信与信息工程学院，南京 210003）

摘　要：针对无人机航拍图像背景复杂、小尺寸目标分布密集等造成的检测精度低等问题，提出一种高分辨率特征增强的无人机航拍小目标检测算法。首先，提出了高分辨率特征增强网络，通过减少主干网络的下采样倍数来扩大输出特征图的尺度，同时引入双线性插值法来减少采样后特征信息的丢失，从而保留更多语义特征与细节特征。其次，在主干网络嵌入一种结合局部跨阶段结构的快速空间金字塔池化（Spatial pyramid pooling fast cross stage partial construction，SPPFCSPC）模块，增强局部与全局特征的信息融合，从而获得更大的感受野。最后，通过马赛克混合数据增强方法来增强图像背景的复杂度，提高模型的泛化能力。在公开数据集 VisDrone 2019 上的实验结果表明，与“你只需看一次”（You only look once，YOLO）系列等其他主流算法相比，本文算法的平均精度均值有显著的提高，在不同场景下均验证了本文算法的优越性，表明本文算法对无人机航拍图像的密集小目标检测任务有较强的实用性。

关键词：小目标检测；无人机航拍图像；空间金字塔池化

中图分类号：TN911.73　　**文献标志码：**A

Small Target Detection in UAV Aerial Images Based on High Resolution Feature Enhancement

ZHOU Xuan, GE Qi, SHAO Wenze

(College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: Aiming at the problem of low detection accuracy caused by complex background and dense distribution of small size targets in unmanned aerial vehicle (UAV), this paper proposes a small target detection algorithm based on high resolution feature enhancement. Firstly, a high-resolution feature enhancement network is proposed, which expands the scale of the output feature map by reducing the sub-sampling times of the backbone. At the same time, the bilinear interpolation is introduced to reduce the loss of feature information after up-sampling, thereby preserving more semantic and detailed features. Secondly, the spatial pyramid pooling-fast module combined with the cross stage partial structure is embedded in the backbone to enhance the information fusion of local and global features, so as to obtain a larger receptive field. Finally, the mosaic-mixup data enhancement method is used to enhance the complexity of image background and improve the generalization ability of the model. Experimental results on the public dataset VisDrone 2019 show that compared with other mainstream algorithms such as the “ you only look once ”

基金项目：国家自然科学基金(61972213)。
收稿日期：2023-04-03；**修订日期：**2023-05-29

(YOLO) series, the mean average precision of the proposed algorithm has significantly improved. The advantages of the proposed algorithm have been verified in different scenarios, indicating that the algorithm has strong practicality for dense small target detection tasks in UAV aerial images.

Key words: small target detection; unmanned aerial vehicle (UAV) aerial image; spatial pyramid pooling (SPP)

引　言

随着无人机在军事、民用领域的飞速发展和广泛应用,其航拍图像数据量呈井喷式增长,基于深度学习的无人机航拍图像检测成为了计算机视觉的一个研究重点。但与其他常规的目标检测任务相比,航拍图像检测面临图像视野广、目标密集和小目标占比多等难题[1],而解决问题的突破口在于如何从小目标的有限特征中挖掘出足够多的特征信息。

近年来基于深度学习的目标检测算法主要分为两类:两阶段检测算法和一阶段检测算法,其中两阶段检测算法先是生成候选区域,再对候选区域进行分类和定位,如区域卷积神经网络(Region-based convolutional neural networks,R-CNN)系列算法[2-4]、Mask R-CNN算法[5]和Cascade R-CNN算法[6]等。但这些方法在检测过程中存在候选区域,通常会产生较大的计算开销。一阶段检测算法是通过初始锚框对目标进行类别预测和定位,可以在不进行候选区域的情况下完成端到端的目标检测。因而与两阶段检测算法相比,一阶段检测算法的网络结构更加简单、运行速度更快,具有更好的实时性,如"你只需看一次"(You only look once,YOLO)系列算法[7-10]、单次发射多边框检测器(Single shot multibox detector,SSD)算法[11]、RetinaNet算法[12]等。然而,它们对于小目标仍存在检测精度较低的问题。

为解决小目标检测精度低的问题,研究者们提出了大量的解决方案,这些方法的途径主要可分为4类:(1)提高输入特征的分辨率[13-14];(2)超采样和数据增强[15-16];(3)上下文学习[17-18];(4)多尺度学习[19-20]。Zhan等[21]通过在YOLOv5上额外添加小目标检测层,有效增强了小目标的检测效果,但这种做法使模型复杂度上升、计算量增大,造成模型的检测速度下降。Lim等[22]提出了一种基于上下文与注意力的小目标检测算法,通过增强网络对小目标的注意力,从而在一定程度上提高模型对小目标的识别能力,但对检测精度的提高有限,不能较好地完成现实场景下的无人机航拍图像检测任务。此外,Song等[23]提出了一种基于多尺度特征融合的小目标检测算法,借鉴特征金字塔的思想,实现低分辨率层的语义特征和高分辨率层的细节特征的充分融合,从而减少模型对小目标的漏检和误检。但该方法的缺点也很明显,主干网络层数的加深导致模型参数大量增加,模型复杂度升高,从而造成模型的检测速度变慢。

因为无人机数据集中小目标分布密集,部分图像因小目标占比多导致检测精度低以及远距离拍向造成目标外观模糊等问题,采用常规的目标检测算法很难达到理想效果。因此本文提出了一种高分辨率特征增强的无人机航拍小目标检测算法,首先设计高分辨率特征增强网络H-YOLOv5,对原YOLOv5的主干网络和颈部进行重新设计,通过减少主干网络下采样的倍数来扩大输出特征图的尺度,同时在颈部引入双线性插值法来提高上采样后特征图的清晰度,从而减少小目标细节特征及位置信息的损失,降低模型对小目标的漏检和误检。其次为增大感受野,在主干网络嵌入一种结合了局部跨阶段结构的快速空间金字塔池化(Spatial pyramid pooling fast cross stage partial construction,SPPFCSPC)模块,增强模型对小目标的感知能力,降低漏检率。最后设计马赛克混合(Mosaic-mixup,MM)数据增强方法,从数据集中任选4张训练样本进行随机裁剪及缩放,拼接成一张新的训练图像,同时选取同一批次内的两张不同训练图像按比例混合,通过增加图像背景复杂度的方式来增强模型的学习能力,提升模型的泛化性和鲁棒性。

1 相关工作

1.1 YOLOv5网络结构

本文以YOLOv5 6.0版本为基准介绍YOLOv5,它的网络结构主要可分为4部分:输入端(Input)、主干网络(Backbone)、颈部(Neck)和检测头(Head),模型网络结构如图1所示。

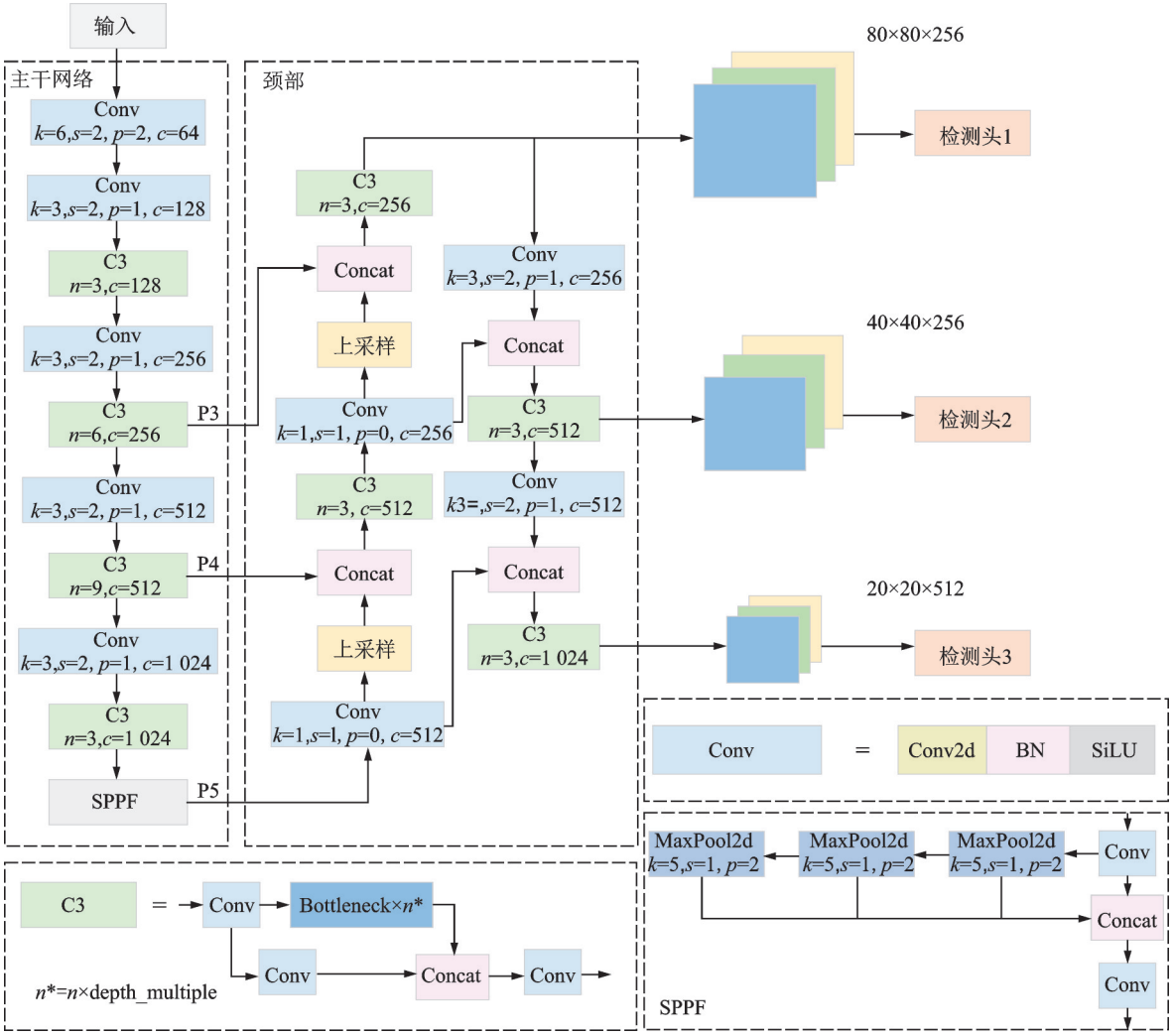


图1 YOLOv5网络结构图

Fig.1 YOLOv5 network structure

主干网络主要由卷积(Conv)模块、局部跨阶段结构(Cross stage partial,CSP)和快速空间金字塔池化(Spatial pyramid pooling-fast,SPPF)等模块组成。卷积模块由卷积层(Conv2d)、批标准化层(BN)及激活函数层(SiLU)组成。主干网络主要通过卷积模块来提取输入图像的特征,其中卷积核通过在输入图像上不断进行滑动得到更深的特征图。特征层矩阵各个点的计算公式为

$$y^t=\sum_{i,j\in U}x_{ij}^{t-1}K_{ij}^t+b^t \tag{1}$$

式中:$U=\{1,2,\cdots,k\}$;y为特征层矩阵中单个点的值;t为当前层级;i和j为在卷积核及对应感受野中对

应的位置坐标;x为感受野对应的原图区域;K为卷积核;b为偏置。因此主干网络的卷积模块相当于2倍下采样,每经过一次卷积模块,特征图的通道数增加一倍。

颈部采用特征金字塔(Feature pyramid networks,FPN)和路径聚合网络(Path aggregation network,PAN)[24]相结合的结构,将主干网络经多次下采样后输出的特征,分别与FPN自底向上的特征图进行融合。这种将高层的语义特征与浅层的位置信息融合的操作有利于多尺度目标的检测。但对小目标而言,多次下采样后特征图所含的有效特征信息减少,导致小目标的检测效果不理想。尤其是尺度小于8像素的目标,经8倍下采样后,整个目标都会在特征图中消失。为此,本文提出高分辨率特征增强网络H-YOLOv5,通过减少主干网络的下采样倍数来丰富小目标浅层的特征信息,同时在颈部引入双线性插值法来减少上采样后特征图的模糊,从而减少小目标轮廓和纹理特征的丢失,降低模型对小目标的漏检率。

1.2　空间金字塔池化

空间金字塔池化(Spatial pyramid pooling,SPP)可以将任意大小的特征向量转换成固定大小的特征向量,从而使卷积神经网络适应不同比例、大小的输入图像,并且可以帮助网络提取到多尺度目标的有效特征信息。如图2所示,SPP模块主要包含4个并行的分支,分别是3个不同池化核大小的最大池化层以及一个跳跃连接层。在SPP模块中,输入的特征先经过一个1×1的卷积模块,输出的通道数减半;再经过3个池化核大小分别为5、9、13的最大池化操作,输出的特征连同原始输入特征一起进行融合;最后通过1×1卷积模块调制通道数与输入通道数相同,从而得到新的特征图。SPP模块利用3个不同尺度的最大池化,在几乎不影响模型检测速度的前提下增大感受野,增强多尺度特征的信息融合。

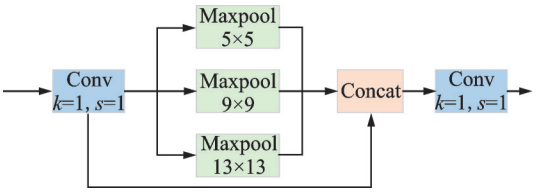


图2　SPP模块

Fig.2　SPP module

SPPF模块是YOLOv5在SPP模块基础上改进的空间金字塔池化,其结构如图3所示,SPPF模块使用3个5×5的最大池化来代替SPP模块中3个池化核大小分别为5、9、13的最大池化。相较于SPP模块,SPPF模块在保证检测精度的同时减少了模型的参数量和计算量,加快了模型的推断速度。

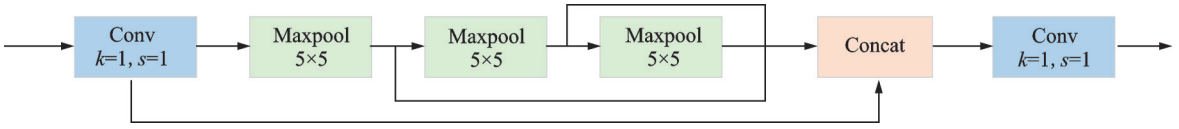


图3　SPPF模块

Fig.3　SPPF module

2　小目标检测算法

本文以YOLOv5为基础提出了针对无人机航拍小目标的检测算法,其网络结构如图4所示,其中标红区域显示了本文算法的主要改进之处。首先提出高分辨特征增强网络H-YOLOv5,减少主干网络下采样倍数的同时使用双线性插值法来实现特征图的上采样操作,使小目标包含更多的细节和位置信息,有利于模型对小目标的定位与识别;其次在主干网络添加SPPFCSPC模块,并且为轻量化主干网络取消尾部的C3(CSP Bottleneck with 3 convolutions)结构,在增大感受野的同时丰富特征的细节信息,

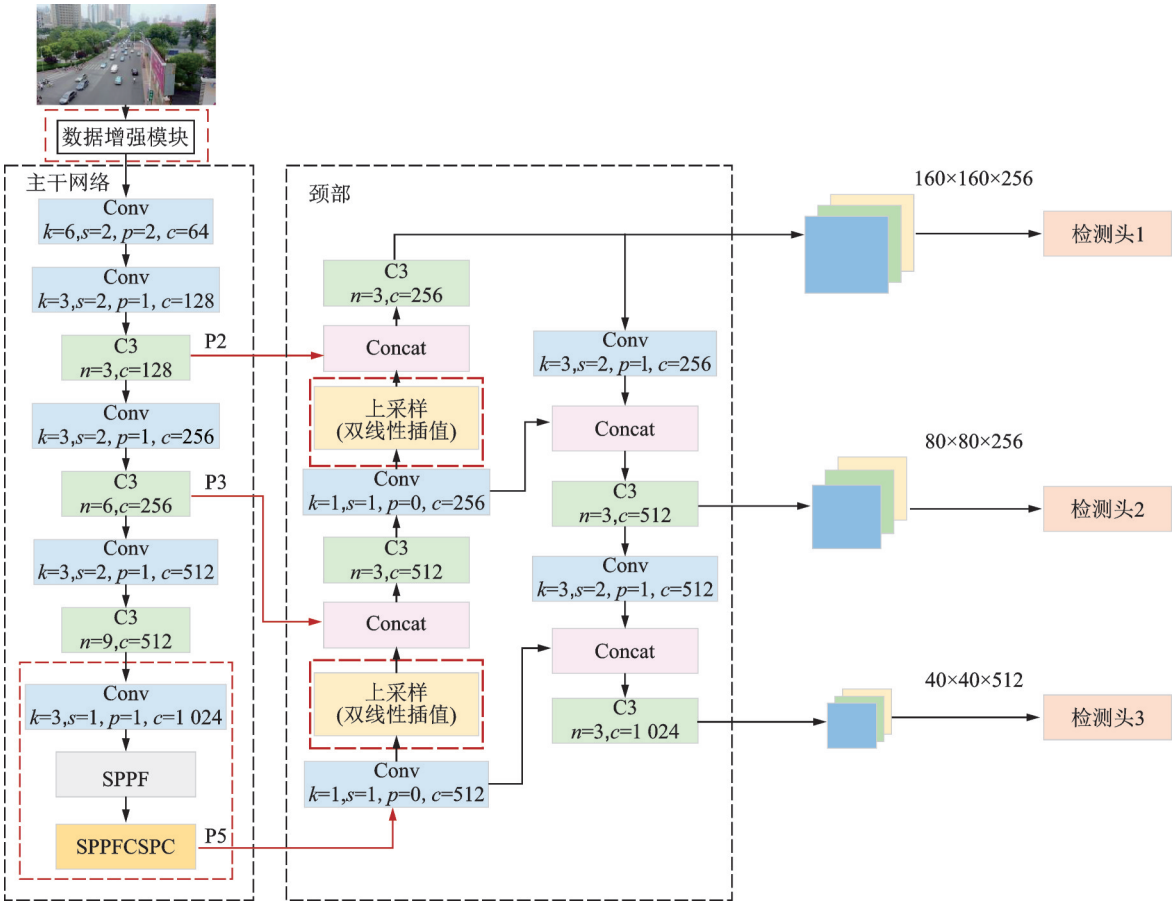


图 4 高分辨率特征增强的无人机航拍小目标检测模型结构

Fig.4 Structure of small target detection in UAV aerial image based on high resolution feature enhancement

降低小目标的漏检、误检;最后设计马赛克混合数据增强方法来丰富数据集,从而提升航拍图像的检测精度。下面将详细介绍网络结构中各个模块的信息。

2.1 高分辨率特征增强网络 H-YOLOv5

在YOLOv5网络中,输入图像在主干网络经8倍、16倍、32倍下采样后输出{P3,P4,P5}特征,分别与FPN自底向上的特征图进行融合,如图5(a)所示。输入图像经过多次下采样逐步映射为不同尺度的

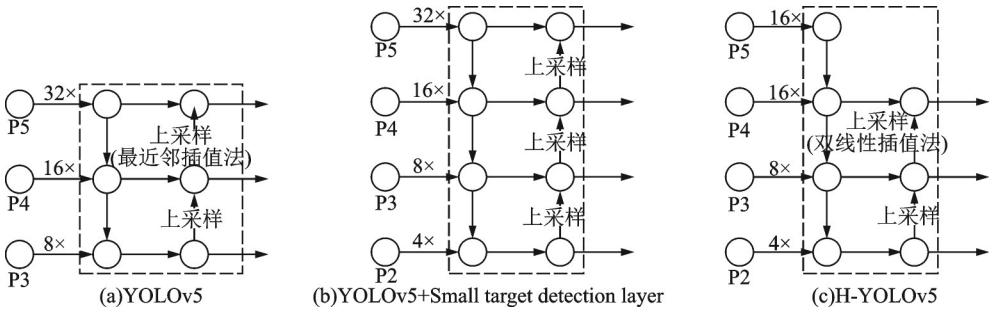


图 5 各网络的颈部结构示意图

Fig.5 Schematic diagram of neck structure of each network

特征图,而特征图所包含的有效像素信息也随之减少。表1直观地反映了将原始图像映射到不同特征层后所占像素情况。结合实际数据集中目标尺度的分布情况,将小目标进一步划分为检测尺度小于16像素×16像素的极小目标。

表1　原始图像映射到不同特征层后的分辨率

Table 1　Resolution of original image mapping to different feature layers　像素×像素

特征层	P2/4	P3/8	P4/16	P5/32
16×16	4×4	2×2	1×1	0.5×0.5
32×32	8×8	4×4	2×2	1×1
64×64	16×16	8×8	4×4	2×2

从表1中可以看到,小目标(目标尺寸小于32像素×32像素)特征映射到P5层后所包含的像素分辨率为1像素×1像素,极小目标甚至不到1像素×1像素,这对无人机航拍小目标来说检测效果甚微。针对上述问题,本文提出了高分辨率特征增强网络H-YOLOv5,在主干网络、颈部和检测头3部分适配高分辨率特征图,同时在特征融合部分引入双线性插值法来提高上采样后特征图的清晰度,保证传递更多的小目标细节信息并输出对小目标表征能力更强的特征图。

与{P3,P4,P5}相比,P2大尺度检测层中含有较为丰富的细节特征和位置信息更有利于无人机图像中小目标的检测,故有算法提出在原YOLOv5的基础上添加小目标检测层(P2层)来丰富特征信息,从而提升小目标的检测效果,如图5(b)所示。额外添加P2检测层,虽然能有效改善小目标的检测效果,但网络的卷积和上采样次数也随之增加,提高了模型的计算复杂度,导致模型的检测速度下降。为更好地平衡模型的精度和速度,本文在主干网络中去除了32倍的下采样层,使用步长为1的3×3卷积层进行替换,同时仍采用3个不同尺度的检测头结构,但调整检测头为{P2,P3,P4}所对应的检测分支,如图5(c)所示。

H-YOLOv5通过用步长为1的3×3卷积替代用于下采样的步长为2的3×3卷积来减少网络的下采样倍数,与直接添加小目标检测层相比,H-YOLOv5减少了颈部的卷积和上采样次数,确保高分辨率特征输入到检测头的同时提高了模型的推理速度。虽然上述通过减少下采样倍数保证了更多纹理和轮廓信息的传递,但原模型采用最近邻插值法进行上采样操作(如图5(a)所示),易破坏特征图的像素关系,导致图像特征模糊,从而造成小目标特征及位置信息的损失。因此本文在颈部引入双线性插值法来实现特征图的上采样(如图5(b)所示),从而提高上采样后特征图的清晰度。

双线性插值法在计算新特征图像素点值时,从输入图像中选择4个点并分别在两个方向进行3次单线性插值,假设输入图像上选择的4个点分别为$Q_{11}(x_1,y_1)$,$Q_{21}(x_2,y_1)$,$Q_{12}(x_1,y_2)$,$Q_{22}(x_2,y_2)$,设$R_1(x,y_1)$,$R_2(x,y_2)$,$P(x,y)$。首先在x方向上做两次单线性插值得

$$f(R_1)\approx\frac{x_2-x}{x_2-x_1}f(Q_{11})+\frac{x-x_1}{x_2-x_1}f(Q_{21}) \tag{2}$$

$$f(R_2)\approx\frac{x_2-x}{x_2-x_1}f(Q_{12})+\frac{x-x_1}{x_2-x_1}f(Q_{22}) \tag{3}$$

然后在y方向进行插值,得

$$f(P)\approx\frac{y_2-y}{y_2-y_1}f(R_1)+\frac{y-y_1}{y_2-y_1}f(R_2) \tag{4}$$

整理可得最终结果

$$f(x,y)\approx\frac{f(Q_{11})}{(x_2-x_1)(y_2-y_1)}(x_2-x)(y_2-y)+\frac{f(Q_{21})}{(x_2-x_1)(y_2-y_1)}(x-x_1)(y_2-y)+\frac{f(Q_{12})}{(x_2-x_1)(y_2-y_1)}(x_2-x)(y-y_1)+\frac{f(Q_{22})}{(x_2-x_1)(y_2-y_1)}(x-x_1)(y-y_1) \tag{5}$$

双线性插值法是单线性插值的拓展，通过两个方向上的3次插值，在保证原特征图像素关系的同时对特征图进行上采样，可提高采样后特征图的清晰度，减少小目标细节信息的丢失，减少小目标的漏检、误检。

2.2 结合局部跨阶段结构的快速空间金字塔池化模块

YOLOv5利用SPPF模块实现了不同尺度的局部特征与全局特征的信息融合，具有较好的特征提取能力，但对目标尺度差异大的图像仍存在检测精度低的问题。为此，本文在主干网络加入SPPFCSPC模块，进一步增强局部与全局特征信息的融合，有效提升网络对多尺度目标的特征提取能力。

CSP模块是YOLOv5根据CSPNet[25]的分割梯度流理念提出的，其结构如图6所示。本文将CSP模块运用到SPPF模块，通过拆分和合并策略，使梯度路径的数量翻倍，从而优化梯度信息，强化模型的特征聚合能力。

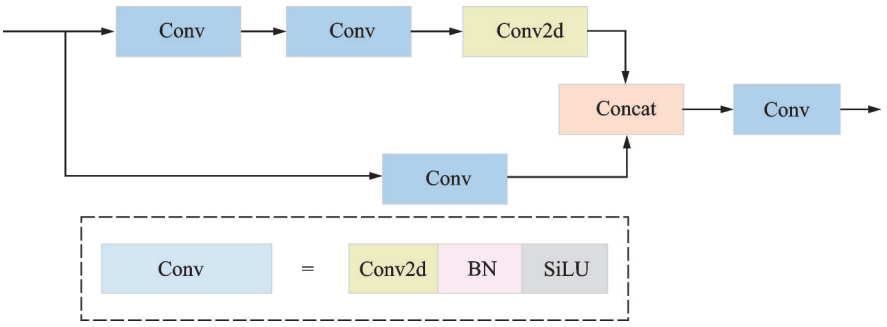


图6 CSP模块结构图

Fig.6 CSP module structure

图7是YOLOv5与本文算法的主干网络结构对比图。如图7所示，在SPPFCSPC模块中，输入的特征图被分别送入左右两个并列的分支，右侧分支为一个简单的1×1卷积模块，输出通道数减半。为保证最终输出的通道数不变，特征图在左侧分支时先经过一个1×1卷积模块，压缩输出通道数为原来的一半，再经过SPPF模块，利用3个不同大小的池化核进行池化，保留了不同尺度的局部特征。最后将左右跨阶段分层的特征图融合在一起，通过CSP结构使得梯度流在不同的网络路径中传播，避免过多地重复梯度信息，实现多尺度的局部特征与全局特征信息的更好融合，增强了网络对特征的聚合能力。

虽然上述通过添加SPPFCSPC模块增强了局部与全局的特征信息融合，但SPPFCSPC模块的引入使得模型参数增加，导致网络的检测速度下降明显。为平衡模型的精度与速度，本文通过减少主干网络的C3模块来减少网络深度，实现主干网络轻量化。从图7可以看出，与原始YOLOv5网络结构不同的是，本文算法去掉了原始主干网络中的最后一个C3模块，同时在SPPF模块后加入了SPPFCSPC模块。与YOLOv5相比，本文模型在平衡精度与速度的同时，增大了感受野，增强了模型对多尺度目标的特征提取能力，从而进一步提升模型的检测性能。

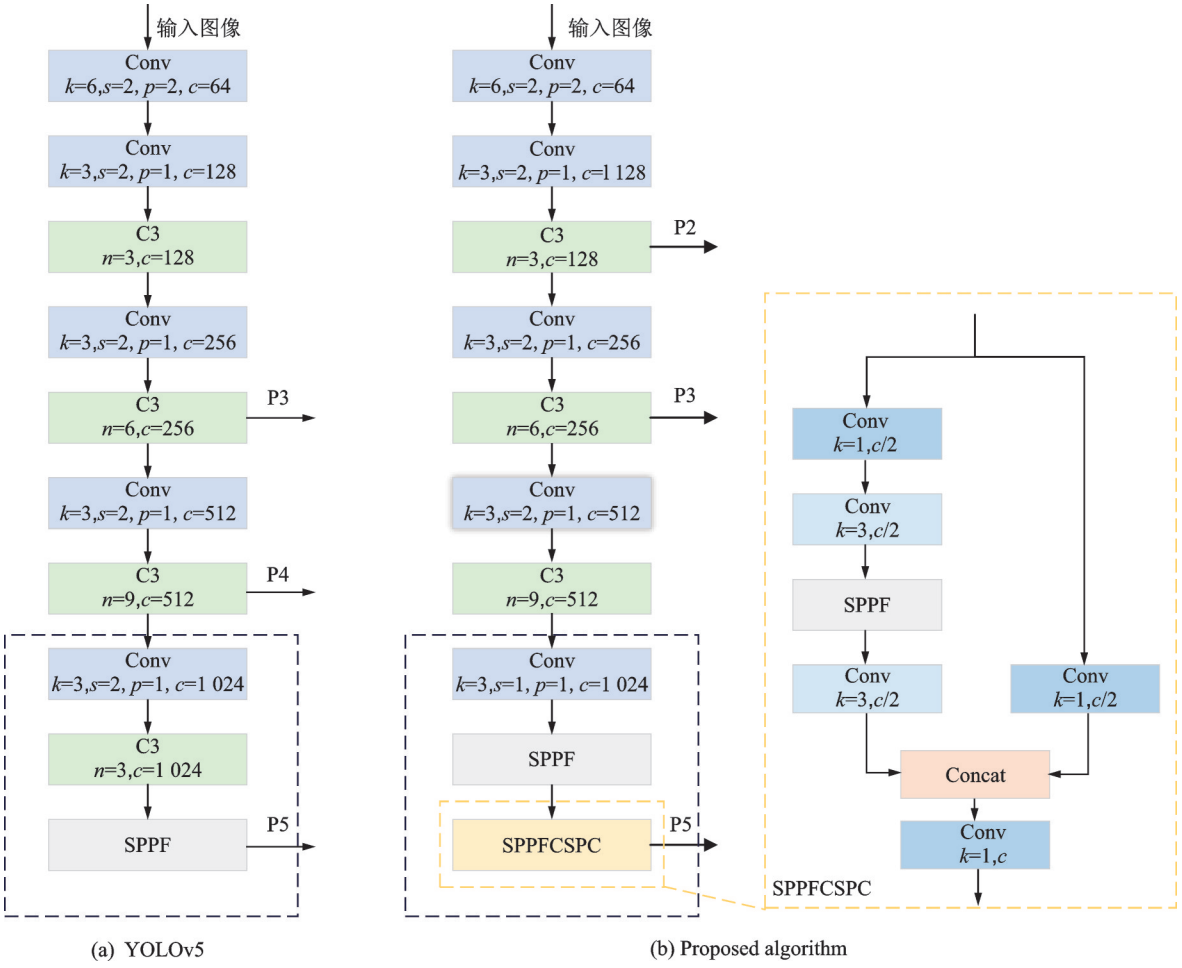


图 7　YOLOv5和本文算法主干网络的结构对比图

Fig.7　Structure comparison diagram of the backbone network of YOLOv5 and the proposed algorithm

2.3　数据增强模块

为提升复杂背景下的小目标检测精度,本文在YOLOv5原有的Mosaic数据增强方法的基础上加入了Mixup[26]数据增强方法,称为MM数据增强方法。

首先确定输出图像的固定尺寸为高h、宽w,在高和宽两个方向随机生成两条分割线,同时从数据集中选取4张图像进行随机裁剪、缩放,再按固定尺寸进行拼接,拼接后的图像形成新的训练样本。该操作主要通过随机裁剪和缩放丰富了数据集中目标的特征,增强了模型的学习能力,同时通过拼接的方式保留了图像的目标特征,极大程度地丰富了检测目标的背景,有效改善了因图像背景相似而造成的模型泛化能力低的问题。然后将按上述操作生成的两个不同训练样本通过逐像素线性相加的方式进行混合,具体混合过程如图8所示。该混合过程是通过Beta分布生成的混合系数对训练样本进行图像混合,生成的混合样本尺寸与原始训练样本相同。本文通过预先设定的阈值对生成的混合样本进行控制,实验中阈值设置为0.4。混合过程为

$$\lambda = \mathrm{Beta}(\alpha, \beta) \tag{6}$$

$$\tilde{x} = \lambda x_i + (1-\lambda) x_j \tag{7}$$

$$\tilde{y}=\lambda y_i+(1-\lambda)y_j \tag{8}$$

式中：x_i、x_j分别为同一批次中随机选择的两个不同训练样本；y_i、y_j表示训练样本分别对应的标签；λ为混合系数，服从Beta(α, β)分布；$\tilde{x}$为混合后的批次样本；$\tilde{y}$为混合后的批次样本对应的标签。通过该方法生成的训练样本计算量小且扩展了训练数据的空间分布，在几乎不影响检测速度的情况下，提升了模型在复杂背景下的小目标检测精度。

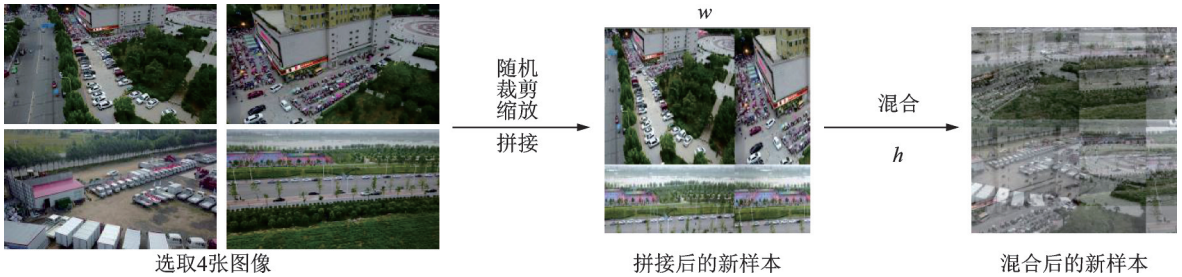


图 8 数据增强过程

Fig.8 Data enhancement procedure

3 实验与分析

本文使用天津大学于2019年发布的VisDrone 2019目标检测数据集，其中训练集6 471张图片，验证集548张图片，测试集1 610张图片，训练类别包含行人、汽车和摩托车等共10个类别，其中的目标标注框超过260万个。图9是Visdrone 2019数据集的目标标注框占原始图像尺寸的比例分布图，可以看出绝大多数目标的占比小于0.3，同时占比小于0.05的极小目标分布密集。

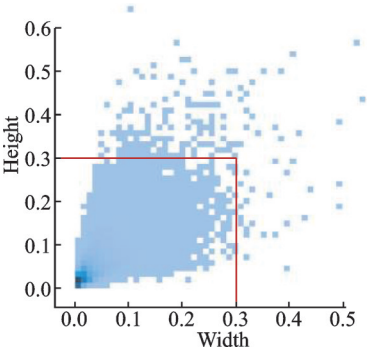


图 9 Visdrone 2019目标尺寸分布情况

Fig.9 Distribution of target size in Visdrone 2019

3.1 实验环境配置与评价指标

本次实验采用的硬件环境GPU为NVIDIA GeForce RTX 2080 Ti，软件环境为Ubuntu 18.04操作系统，选用Pytorch为深度学习框架。本文以YOLOv5S为基准，分别在训练集和验证集进行训练和测试。输入图像的大小为640像素×640像素，批次为16，训练总轮数为120，初始学习率为0.01，采用带动量的随机梯度下降(Stochastic gradient descent，SGD)法作为优化器，动量设置为0.937，权重衰减系数为0.000 5。

本文使用精确度P(Precision)、召回率R(Recall)、平均精度均值(Mean average precision，mAP)和每秒处理图像帧数(Frames per second，FPS)作为衡量模型性能的评价指标。平均精度(Average precision，AP)是通过计算P-R曲线与坐标轴包围的面积得到的单类别检测精度，而mAP为各个类别的AP值相加后再除以类别总数的均值，其中mAP 0.5指交并比(Intersection over union，IoU)阈值为0.5时的mAP值。文中若无特殊说明，mAP默认为0.5。

3.2 不同模型规格的YOLOv5在VisDrone 2019数据集上的对比实验

本文选取YOLOv5的3种模型规格进行对比实验，由小到大分别为s、m、l版本，它们的网络结构相同，不同之处在于模型的深度和宽度不同，实验结果如表2所示。实验表明，随着网络的深度和宽度增加，模型的检测精度逐渐提升，但模型的参数量也随之增大，模型的复杂度升高，加大了模型过拟合的

风险,同时模型的检测速度也随之下降。因此网络规模的增加会加大模型的计算量与复杂度,从而导致网络的实时性变差。在3个网络模型中,YOLOv5s模型的深度与宽度最小,模型的参数量最少,检测速度最快,但检测精度最差。而YOLOv5l的检测精度最高,但检测速度最慢。与其他规格的模型相比,YOLOv5s的参数量少且检测速度快,能较好地平衡检测的精度与速度。同时考虑到计算机有限的空间资源,本文选择轻量化的YOLOv5s提升基线网络检测效果。

表2　YOLOv5不同模型规格结果的比较

Table 2　Comparison of YOLOv5 model specification results

模型	深度	宽度	参数量/10^6	mAP/%	FPS/(帧·s^{-1})
YOLOv5s	0.33	0.50	**7.037**	34.4	**82.2**
YOLOv5m	0.67	0.75	20.889	37.6	60.5
YOLOv5l	1.00	1.00	46.157	**41.1**	52.7

注:加粗数据为最优值。

3.3　消融实验

为了验证本文提出的模型有效性,针对不同模型进行了消融实验,结果如表3所示,其中"√"表示使用该模块。表3第1组为基准网络YOLOv5s的检测结果。第2组为高分辨率特征增强网络H-YOLOv5的实验结果。在VisDrone 2019数据集上H-YOLOv5的mAP值相比第1组网络提升了5.7%,且精确度和召回率均有提升。这是因为H-YOLOv5相比基线网络减少了下采样倍数,改善了小目标在提取特征过程中特征信息丢失的问题,保留了小目标更多的特征信息。说明多尺度的信息融合,特别是充分利用浅层的轮廓位置信息可以提高网络对小目标的定位能力,进而增强模型对小目标的检测效果。而双线性插值法的引入提高了采样后特征图的清晰度,减少了由特征图模糊造成的目标信息损失,从而进一步提升了模型的检测性能。第3组模型在第2组的基础上在主干网络尾部引入SPPFCSPC模块,在验证集上相比第2组网络的mAP值提升了2.2%。这是因为在Visdrone 2019数据集中目标的图像占比小且尺度变化大,加入SPPFCSPC模块后能够有效加强局部与全局特征信息的融合,增强模型对多尺度目标的特征提取能力,从而提高模型的检测精度。第4组模型基于第3组模型添加了数据增强模块,在VisDrone 2019数据集上模型的精确度和召回率均有提升,同时mAP值提升了1.5%,验证了数据增强模块的有效性。加入数据增强模块后,模型利用马赛克混合数据增强方法来提高图像背景复杂度和丰富训练样本,从而提升模型的鲁棒性。第4组网络的精确度、召回率和检测精度均值均达到了最优值,验证了本文算法的合理性和有效性。

表3　消融实验

Table 3　Ablation experiments

实验序号	H-YOLOv5	SPPFCSPC	MM	*P*/%	*R*/%	mAP/%	FPS/(帧·s^{-1})
1				47.3	34.7	34.4	82.2
2	√			52.1	39.5	40.1	50.7
3	√	√		54.2	40.6	42.3	31.9
4	√	√	√	56.8	41.5	43.8	31.7

为了更直观地展示本文算法在不同真实场景下的检测效果,本文从VisDrone 2019数据集中分别挑选目标密集、光线昏暗、目标遮挡场景下的图像进行检测,并将YOLOv5s和本文算法的检测视觉效果进行了对比,检测效果如图10所示。可以看出,在目标密集场景下,本文算法与YOLOv5s相比在各个类别的检测精度上均有提升,可以更好地检测出互相遮挡的行人和远距离的汽车。本文算法由于引入了SPPFCSPC模块,网络在特征提取过程中考虑到目标的尺度变化,提取的特征能更好地适应图像

因拍摄距离远近造成的目标尺度变化大的情况。在光线昏暗且目标密集的场景下，YOLOv5s能较好地检测出分布密集的行人小目标，但相比本文算法，仍存在很多漏检。这是因为本文算法通过减少下采样倍数来扩大输出特征图的尺度，有效改善了小目标在多次下采样过程中产生的信息丢失问题，通过充分挖掘小目标浅层的细节特征和位置信息，提升了模型对小目标的定位能力。在目标被遮挡的场景下，本文算法可以更好地检测出被遮阳伞遮挡的行人和摩托车，与YOLOv5s相比，小目标的漏检和误检均得到减少。这是因为在遮挡场景下目标的可利用信息大幅减少，而本文模型通过双线性插值法来提高采样后特征图的清晰度，充分挖掘目标的可用信息，提高了模型对目标的识别能力，从而提高模型的检测精度。通过以上不同场景下检测效果的对比分析，可以发现本文算法在无人机航拍的密集小目标检测上有明显的优势。

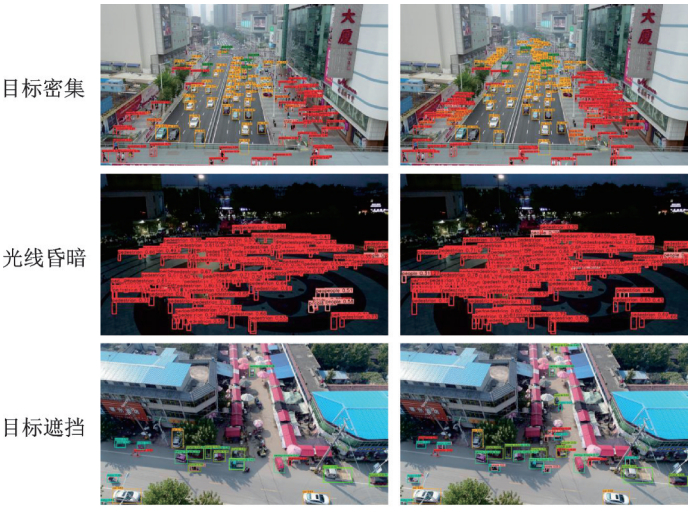


图 10 YOLOv5和本文算法在不同场景下的检测效果

Fig.10 Detection effect in different scenarios using YOLOv5 and the proposed algorithm

3.4 不同目标检测算法在VisDrone 2019数据集上的检测结果对比

为进一步验证本文算法的潜在优势，本文选取多种目标检测算法进行对比实验，囊括了近几年在小目标检测领域常用框架和最新的改进方法，实验结果见表4。通过对比分析VisDrone 2019验证集上各个类别mAP值的结果，可以看出与各算法模型相比，本文模型的综合性能最优。一方面在自行车、汽车、摩托车和行人等目标类别中检测精度最佳，另一方面与最新的YOLOv8模型相比，平均精度均值提

表 4 不同检测算法的对比实验

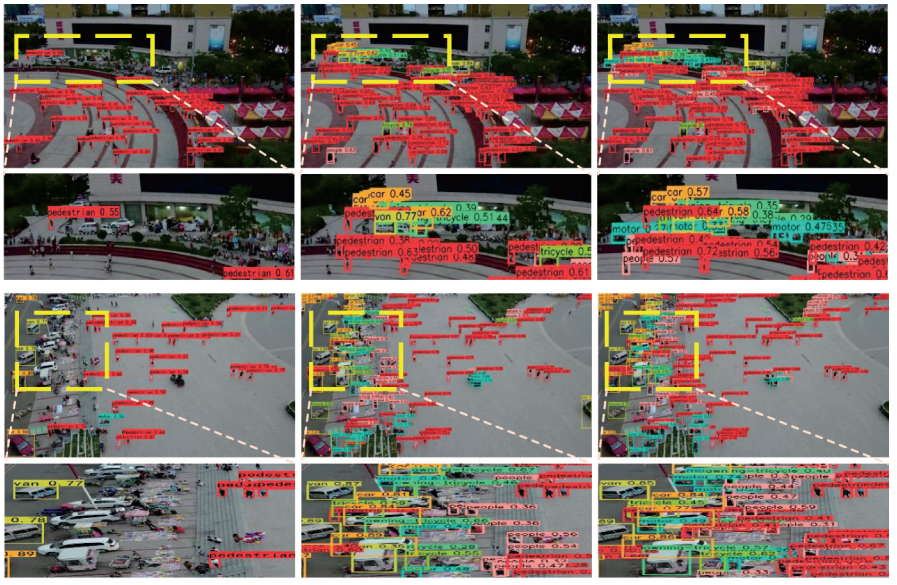
Table 4 Comparative experiment of different detection algorithms %

算法	主干网络	不同目标类别的相应检测精度										mAP
		Awn-tr	Bicycle	Bus	Car	Motor	Pedestrian	People	Tricycle	Truck	Van	
Faster R-CN	ResNet-50	5.9	0.6	37.3	31.6	3.0	4.6	0.9	5.5	29.5	24.5	14.3
RetinaNet	ResNet-50	4.5	2.1	40.8	29.3	2.6	4.2	0.9	6.4	35.0	24.3	15.0
SSD	VGG-16	15.5	5.0	47.2	63.2	19.1	18.7	9.0	11.7	33.1	30.0	25.3
Cascade R-CNN	ResNet-50	8.6	7.6	34.9	54.6	21.4	22.2	14.8	14.8	21.6	31.5	23.2
YOLOv5	CSPDarknet	11.9	12.0	46.4	74.6	39.3	40.2	32.6	21.8	30.9	37.7	34.4
YOLOv7[27]	CSPDarknet	**23.2**	8.5	**66.2**	74.8	31.3	45.8	7.4	19.1	**56.7**	**51.3**	38.4
YOLOX[28]	CSPDarknet	22.1	7.1	55.5	73.0	30.7	38.5	12.1	15.2	55.3	51.0	36.1
QueryDet[29]	ResNet-50	4.1	10.1	35.7	44.8	14.7	16.9	10.3	11.4	19.1	27.0	33.9
RFLA[30]	ResNet-50	5.6	4.6	25.9	42.2	12.6	12.3	9.5	8.9	15.7	20.4	29.6
YOLOv8	CSPDarknet	15.8	16.4	60.5	80.4	47.8	45.3	34.5	**29.9**	39.2	44.8	41.5
本文算法	CSPDarknet	17.9	**18.9**	58.9	**83.4**	**49.3**	**49.3**	**40.7**	29.2	40.3	48.0	**43.8**

注：加粗数据为最优值。

升了2.3%。用于加速高分辨率小目标检测的级联稀疏查询(Cascaded sparse query for accelerating high-resolution small object detection,Querydet)[29]和基于高斯感受野的标签分配策略(Gaussian receptive field based label assignment,RFLA)[30]是近年提出的针对小目标的检测算法。QueryDet提出级联稀疏查询来减少使用高分辨率特征的模型的计算量,而RFLA提出基于高斯感受野的标签分配策略来提高小目标的检测精度。对比小目标的检测算法QueryDet和RFLA,本文算法仍然保持了检测精度上的优势。

为更直观地展示本文方法的优势,本文将YOLOv7、YOLOv8和本文算法的检测视觉效果进行了对比。图11为YOLOv7、YOLOv8和本文算法在小目标密集图像上的对比结果。从图11可以看出,本文方法与YOLOv7、YOLOv8相比,在图像背景复杂、目标相互遮挡、小目标分布密集的场景下都取得了较好的检测效果,能更好地识别出远处的极小目标和相互遮挡的行人目标。本文算法因为采用了高分辨率特征增强网络,能在小目标密集区域充分挖掘小目标浅层的轮廓纹理信息,提取的特征图能更精准地覆盖到目标物体上,增强模型对小目标的定位能力,对小目标检测性能的提升尤为明显。同时,本文算法由于添加了SPPFCSPC模块,增大了特征图的感受野,加强了局部与全局特征信息的融合,提高了模型对多尺度目标的特征提取能力。数据增强模块的引入,提高了图像背景的复杂度,使模型能更好地应对复杂背景下的小目标检测,进一步提升了模型的整体检测性能。



(a) YOLOv7　(b) YOLOv8　(c) Proposed algorithm

图11　3种模型检测效果对比图

Fig.11　Detection effect comparison of three models

4　结束语

针对无人机图像中小尺寸目标的检测精度低的问题,提出一种高分辨率特征增强的无人机航拍小目标检测算法。首先提出高分辨率特征增强网络H-YOLOv5,通过减少主干网络的下采样倍数来提高特征图的分辨率,使模型在训练过程中能够充分挖掘小目标浅层的特征信息;其次将SPPFCSPC模块嵌入主干网络的尾部,通过增强感受野适应航拍图像目标的多尺度变化;最后提出了马赛克混合数据增强方法,有效改善模型由于训练背景相似导致的泛化能力低的问题,进一步提升了模型的鲁棒性。在VisDrone 2019数据集上的实验结果表明,本文算法有效提升了无人机航拍图像中小目标的检测精

度,但在大目标上与YOLOv7和YOLO*X*相比检测精度有所下滑。下一步将继续研究高精度目标后检测算法,保证小目标精准度的同时,进一步提升模型对大目标的检测精度。

参考文献:

[1] 江波, 屈若锟, 李彦冬, 等. 基于深度学习的无人机航拍目标检测研究综述[J]. 航空学报, 2021, 42(4): 524519.
JIANG Bo, QU Ruokun, LI Yandong, et al. Object detection in UAV imagery based on deep learning: Review[J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(4): 524519.

[2] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2014: 580-587.

[3] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision. [S.l.]: IEEE, 2015: 1440-1448.

[4] 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.

[5] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision. [S.l.]: IEEE, 2017: 2961-2969.

[6] CAI Z, VASCONCELOS N. Cascade R-CNN: Delving into high quality object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2018: 6154-6162.

[7] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2016: 779-788.

[8] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2017: 7263-7271.

[9] REDMON J, FARHADI A. Yolov3: An incremental improvement[EB/OL]. (2018-04-08)[2023-04-03]. https://arxiv.org/abs/1804.02767.

[10] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: Optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2023-04-03]. https://arxiv.org/abs/2004.10934.

[11] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]//Proceedings of the European Conference on Computer Vision. Amsterdam, Netherlands: Springer, 2016: 21-37.

[12] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of IEEE International Conference on Computer Vision. [S.l.]: IEEE, 2017: 2980-2988.

[13] CAI Z, FAN Q, FERIS R S, et al. A unified multi-scale deep convolutional neural network for fast object detection[C]//Proceedings of the European Conference on Computer Vision. Amsterdam, Netherlands: Springer, 2016: 354-370.

[14] FU C Y, LIU W, RANGA A, et al. DSSD: Deconvolutional single shot detector[EB/OL]. (2017-01-23)[2023-04-03]. https://arxiv.org/abs/1701.06659.

[15] KISANTAL M, WOJNA Z, MURAWSKI J, et al. Augmentation for small object detection[EB/OL]. (2019-02-19)[2023-04-03]. https://arxiv.org/abs/1902.07296.

[16] ZOPH B, CUBUK E D, GHIASI G, et al. Learning data augmentation strategies for object detection[C]//Proceedings of the European Conference on Computer Vision. Glasgow, UK: Springer, 2020: 566-583.

[17] CHEN C, LIU M Y, TUZEL O, et al. R-CNN for small object detection[C]//Proceedings of Asian Conference on Computer Vision. [S.l.]:[s.n.], 2017: 214-230.

[18] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(4): 834-848.

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

[20] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2017: 2117-2125.

[21] ZHAN W, SUN C, WANG M, et al. An improved YOLOv5 real-time detection method for small objects captured by UAV [J]. Soft Computing, 2022, 26: 361-373.

[22] LIM J S, ASTRID M, YOON H J, et al. Small object detection using context and attention[C]//Proceedings of International Conference on Artificial Intelligence in Information and Communication (ICAIIC). [S.l.]: IEEE, 2021: 181-186.

[23] SONG Z, ZHANG Y, LIU Y, et al. MSFYOLO: Feature fusion-based detection for small objects[J]. IEEE Latin America Transactions, 2022, 20(5): 823-830.

[24] LI H, XIONG P, AN J, et al. Pyramid attention network for semantic segmentation[EB/OL]. (2018-05-25)[2023-04-03]. https://arxiv.org/abs/1805.10180.

[25] WANG C Y, LIAO H Y M, WU Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. [S.l.]: IEEE, 2020: 390-391.

[26] ZHANG H, CISSE M, DAUPHIN Y N, et al. Mixup: Beyond empirical risk minimization[EB/OL]. (2017-10-25)[2023-04-03]. https://arxiv.org/abs/1710.09412.

[27] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[EB/OL]. (2022-07-06)[2023-04-03]. https://arxiv.org/abs/2207.02696.

[28] GE Z, LIU S, WANG F, et al. Yolox: Exceeding yolo series in 2021[EB/OL]. (2021-07-18)[2023-04-03]. https://arxiv.org/abs/2107.08430.

[29] YANG C, HUANG Z, WANG N. QueryDet: Cascaded sparse query for accelerating high-resolution small object detection [C]//Proceedings of the IEEE International Conference on Computer Vision. [S.l.]: IEEE, 2022: 13668-13677.

[30] XU C, WANG J, YANG W, et al. RFLA: Gaussian receptive field based label assignment for tiny object detection[C]//Proceedings of the European Conference on Computer Vision. Tel Aviv, Israel: Springer, 2022: 526-543.

作者简介:

周璇(1999-),女,硕士研究生,研究方向:深度学习、小目标检测,E-mail: xuan990302@163.com。

葛琦(1984-),**通信作者**,女,副教授,研究方向:图像处理与视觉理解,E-mail: geqi@njupt.edu.cn。

邵文泽(1981-),男,教授,研究方向:计算成像、计算机视觉、黑箱优化。

(编辑:陈珺)