DOI:10.19652/j.cnki.femt.2304663

单聚合YOLO航拍小目标检测算法

杨辉羽 李海明
(上海电力大学计算机与科学学院 上海 201306)

摘 要:使用无人机采集的航拍图中存在背景复杂、目标密集、目标重叠等诸多问题,这都对现有的目标检测网络提出了挑战。以YOLOv5为基础进行改进,修改原有的BackBone网络,嵌入改进后的单聚合(OSA)模块,解决因为网络深度造成的梯度衰减问题;针对原网络结构对小目标的定位不准确,获得的信息不充分问题,增加一个160×160的小目标检测层应对小目标难以检测问题,同时修改特征融合网络丰富语义信息;最后改进原有的损失函数CIoU,长宽不再是一个统一的整体计算损失,而是分开优化,提高预测方框的准确度。算法在VisDrone2019无人机航拍数据集上实验结果表明,平均精度均值(mAP)与原算法相比提升了5.2%,检测帧率达到了45 fps,训练模型大小为18.9 MB。

关键词:YOLOv5;改进OSA模块;小目标检测层;CIoU
中图分类号:TN2 **文献标识码**:A **国家标准学科分类代码**:520.604

Single aggregation YOLO algorithm for airborne small target detection

Yang Huiyu Li Haiming
(School of Computer and Science, Shanghai University of Electric Power, Shanghai 201306, China)

Abstract: There are many problems in the aerial photos collected by UAV, such as complex background, dense targets, overlapping targets, which pose a challenge to the existing target detection network. Based on YOLOv5, the original BackBone network is modified and the improved OSA module is embedded to solve the gradient attenuation problem caused by network depth. In view of the inaccurate positioning of small targets in the original network structure and the insufficient semantic information obtained, a 160×160 small target detection layer is added to deal with the problem of difficult detection of small targets, and the feature fusion network is modified to enrich semantic information. Finally, the original loss function CIoU is improved. The length and width are no longer a unified whole to calculate the loss, but are optimized separately to improve the accuracy of the prediction box. The experimental results of this algorithm on VisDrone 2019 UAV aerial photography data set show that compared with the original algorithm, mAP has improved by 5.2%, the detection frame rate has reached 45 fps, and the training model size is 18.9 MB.

Keywords: YOLOv5; improved OSA module; small target detection layer; CIoU

0 引 言

无人机逐渐成为航拍的主流,实现无人机航拍图像的目标检测与目标跟踪是在安防巡逻、预防山林火灾、电力检修、检查高层建筑安全等相关领域都有着重要地位。与普通的地面目标检测不同,无人机航拍因为其高空视角,拍摄的图像拍摄视角大,图像中存在小目标密集重叠、背景繁杂、不同目标类型尺寸差异巨大等问题。这些多重因素下会严重造成到目标检测精度低的问题,特别是小目标的检测精度过低。

如今,以深度学习为基础的目标检测方法得到了飞速发展。基于特征与基于分割的传统检测方法在性能与泛化方面远远逊于深度学习方法。当下主流基于深度学习的检测方法有两个发展方向,一个是以提取局部特征为主的区域卷积神经网络[1](region-CNN,RCNN)方法作为代表的两步检测模型,另一个是将检测任务作为一个回归任务处理的单步检测模型。在两步检测模型中,第一步是提取图像中检测目标的高置信度帧,提取识别帧区域采用的

收稿日期:2023-01-19

方法有局部区域建议网络[2](region proposal network,RPN)、选择性搜索[3](selective search,SS)等方法。第二阶段是检测是根据待确定识别区域的特征再进行进一步区分。单步检测模型是将检测问题转化为一个回归问题,直接在原始图像上做检测与分类问题。其中以 RCNN、Faster-RCNN[4]等为代表的两步检测模型;SSD[5](single shot multibox detector)、YOLO[6](you only look once)作为单步检测模型代表。

目前针对航拍图像的改进检测模型主要在特征提取方式上进行改进。文献[7]在主干网络一种引入可增大感受野的残差空洞卷积模块提高空间特征的利用率,以提高检测精度,这导致网络规模更为巨大,对实验训练的设备提出更高的要求。文献[8]在主干网络建立深层语义信息与浅层语义信息多尺度检测信息的依赖关系,增加浅层网络特征层的权重,提高对微小目标的检测能力,虽然在小目标上得到了比较好的检测效果,但是在模型推理速度上有所下降,达不到无人机检测实时性的要求。文献[9]通过在主干网络深度级联的方法构建瓶颈连接注意力模块,将其嵌入至主干特征提取网络,强化对基础特征的提取达到对小目标的更加精准的定位,然而此方法实在 YOLOv4[10]上进行的改进实验,训练得到的模型权重巨大,对于嵌入设备有更高的要求。这些方法都是通过单一改进主干网络对于小目标的检测能力来提高检测精度,并且在改进过程中仅仅是考虑检测精度的提升。本文实验不仅从检测精度上评估训练模型,还从模型大小,检测速率上考虑模型的优劣。

航拍图中小目标在主干特征提取网络卷积因为网络深度的问题中容易出现梯度衰减、特征信息消失现象,造成小目标的漏检。考虑到原主干网络 CSP-Dark53[10]在深层次的卷积过程导致梯度消失而且残差网络并不能很好的利用提取的特征等问题,采用改进后的 VOVNet[11]网络替换原 CSPDarkNet 网络,从而消除梯度衰减,加强特征的利用。本文不是单单对主干网络进行特征提取的改进,同时在特征融合网络的进行重新设计。考虑到无人机航拍图像中目标尺寸差距过大、小目标数量过多、小目标漏检率高等问题,修改原网络的检测层,增加一个 160×160 的小目标[12]检测层以提高对小目标的检测精度。利用修改后的特征融合网络将尺寸不同的特征进行融合拼接,采用 K-means+[13]与遗传算法[14]的方法重新标定先验框,加速训练过程。解决小目标信息缺失、目标尺度相差过大等问题。改进 CIoU[15]损失函数,CIOU 损失函数将长宽损失作为一个整体没有考虑当预测方框与真实方框出现长宽比相同时的情况。因此本文分开计算预测方框中的长宽损失,防止出现预测方框与真实方框相似的情况,提升预测方框的精准度。

1 改进 YOLOv5 算法模型

1.1 原 YOLOv5 模型介绍

YOLOv5s[15]目标检测模型主要是由输入端、主干网络、特征融合网络、预测头 4 个部分组成。在主干网络以 CSPDarknet 为主干网络,主要包括了 Focus[15]切片操作、跨阶段局部网络[15](cross stage partial networks,CSP-Net)、空间池化金字塔结构[15](spatial pyramid pooling-fast,SPPF),在特征融合部分采用沿用 YOLOv4 中的双向尺度特证融合结构,之后再将提取到的特征信息传入检测头,输出结构为 80×80、40×40、20×20 尺度大小的预测结果,以实现大中小目标的类别与位置。最后通过非极大抑制[15](non maximum suppression,NMS)的方法输出置信度最大的类别预测。

1.2 改进 YOLOv5 整体结构

现有的 YOLOv5 的 BackBone 主干网络因为网络深度与卷积层池化层的特性问题造成特征信息丢失问题;有航拍图像中存在大量的小目标,背景复杂等图像本身就存在的问题,这些多重因素下导致现有的检测网络对航拍图像检测提出了挑战。

本文实验在原 YOLOv5s 上进行改进的核心思想在主干网络与特征融合网络两个部位对加强小目标的关键特征提取能力。改进 YOLOv5 的整体架构如图 1 所示。首先重新设计 BackBone 主干特征提取网络与多尺度融合网络提高模型对小目标的提取能力。之后通过对损失函数[16](efficient IoU,EIoU)训练改进 YOLO,使得对目标精度有更好的回归效果。

改进 YOLO 首先是对主干网络的替换,原使用的 CSPDarkNet 因为网络深度的问题,在特征提取的过程中造成一定程度上特征信息丢失。为此重新设计主干网络,采用单聚合(one shot aggregation ,OSA)[17]模块克服梯度衰减的问题。并且在检测速度上 OSA 模块与其他的模块相比有一定的优势。

在特征融合部分新建一个 160×160 的特征层,因为浅层的特征图中有更加丰富的位置信息,有利于小目标的检测。再增加 160×160 检测层,使得浅层位置信息与深层的特征信息在特征融合部分通过双通道融合,使得网络适应小目标的检测。

最后修改网络的损失函数 CIoU,由于航拍图像存在的目标存在重叠,占比小等问题,对于一个好的定位损失函数有这较高的要求。修改原 CIoU 中的定位损失函数,长宽不在作为一个整体损失计算,而是分开计算,这样有利于小目标的预测框的回归更加精准,以此来达到预测框的定位精确。

1.3 改进 VoVNet 主干特征提取网络

现有研究表明,随着深度学习网络变得越来越深,输入信息、梯度信息经过许多卷积池化层,当它到达网络的

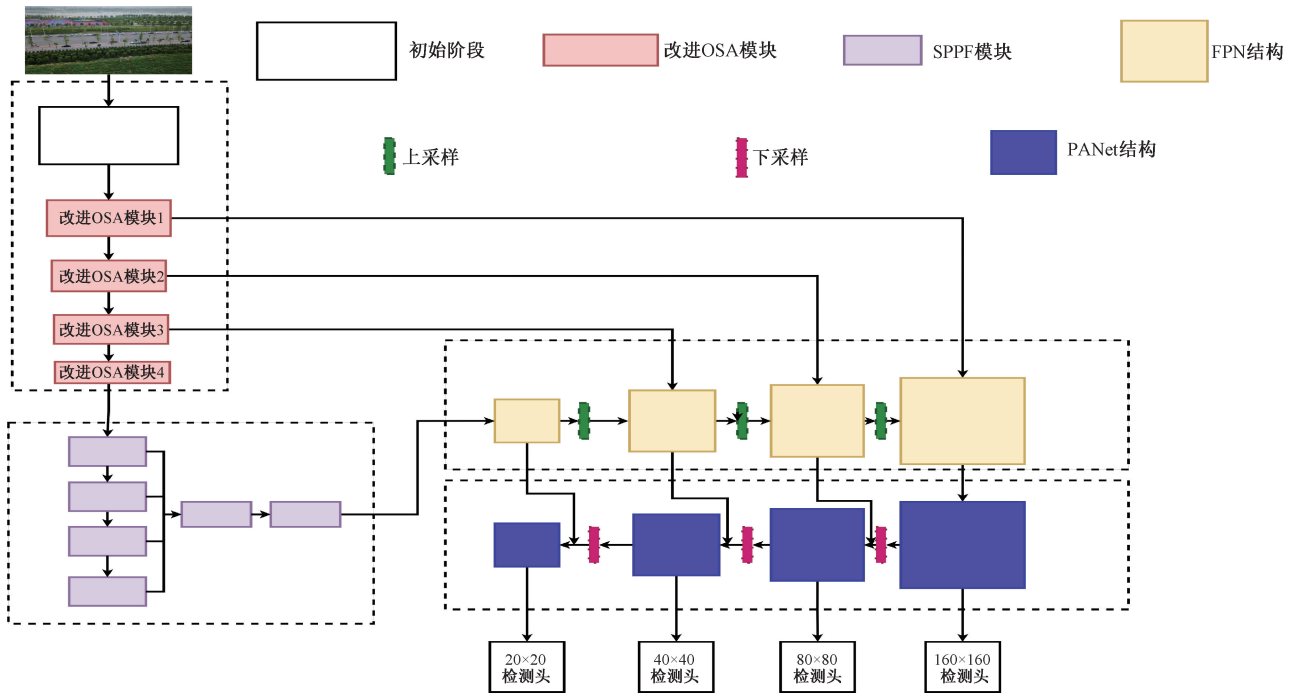


图 1　改进 YOLOv5 的整体结构

末端时，存在一定程度的衰减。在原主干特征提取网络CSPDarkNet53，因为网络的深度问题存在因为梯度衰减造成的特征信息不完全的问题。为了克服梯度衰减带来的问题，本文采用一种改进的 OSA 模块以此来解决该问题，OSA 模块是在 Dense 模块基础上研究的。首先，与CSPDarkNet 不同是 OSA 模块在最后一层融合前面所有的特征信息，这种方式解决了因为网络深度而造成的梯度消失的问题，与之对比的是密集连接模块[18]（Dense Block）每层都融合之前的特征信息，这样会导致中间的特征信息冗余，也会减慢模型的推理速度。与此同时在设计高效的网络时，每秒浮点计算数（floating-point operations per second，FLOPs）与模型的参数是两个主要考虑方面。但是仅仅减少模型的参数大小与 FLOPs 不等同于减少推理时间。在相同的 FLOPs 的条件下，ShuffleNetv2[19] 比 MobileNetv2[20] 在 GPU 上的推理速度更快，所以除了 FLOPs 与模型规模要考虑外还要考虑其他重要因素，如内存访问成本（memory access cost，MAC），计算公式如下：

$$MAC = hw(c_i + c_o) + k^2 c_i c_o \tag{1}$$

式中：h、w 是特征图的长和宽；c_i、c_o 是输入、输出的通道数；k 是卷积核的大小。

$$MAC \geqslant 2\sqrt{\frac{hwA}{k^2}} + \frac{A}{hw} \tag{2}$$

其中，$A = k^2 hwc_i c_o$。根据均值不等式式(2)可以知道，当输入与输出的通道数一致时，MAC 取到最小值，这样的设计才最高效。通过图 2、3 所示可以知道，OSA 模块输入输出的通道数一致，这使得 MAC 取到了下界，MAC 最小。解释了在相同大小的 FLOPs 与模型规模上，本文实验的推理速度更快。

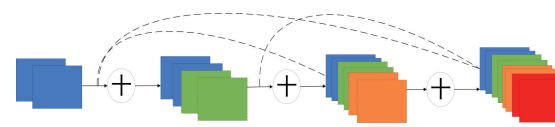


图 2　Dense 模块示意图

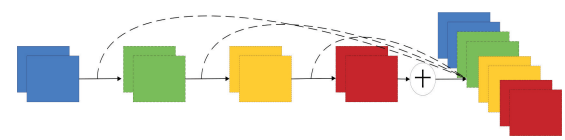


图 3　OSA 模块示意图

图 3 是没有改进的 OSA 模块，输入经过 5 次相同的 3×3 卷积核，每次都将输出保留，最后与输入进行拼接合并操作，再经过个 1×1 卷积得到最终输出。伴随 OSA 模块在 VOVNet 中的堆叠，造成精度饱和或者降低的问题，这是由于 Conv 卷积等变换函数的不断增加，使得梯度的反向传播出现困难。

改进的 OSA 模块借鉴残差网络（ResNet）[21] 的原理机制，将原始输出特征直接拼接到输出端，实现端到端的连接。这样解决了梯度反向传播的困难，实现每个阶段都可以实现梯度的反向传播。也可以使得 VoVNet 可以扩大其网络深度。除此之外还通过一个改进的 eSE[22]（effective-squeeze-and-excitation，Block）模块用于显示建模特征图之间的相互依赖性以达到增强特征表达能力的

目的。

SE[22]模块的计算公式为：

$$A_{ch}(\boldsymbol{X}_i)=\alpha(\boldsymbol{W}_2(\beta(\boldsymbol{W}_1(\kappa(\boldsymbol{X}_i)))))\tag{3}$$

式中：α 为非线性 RELU 的激活函数；β 为 sigmoid 激活函数；$\boldsymbol{W}_1$、$\boldsymbol{W}_2$ 分别是两个全连接层的权值矩阵；$\boldsymbol{X}_i$ 是输入进 SE(Squeeze-and-Excitation Block)模块的特征图矩阵；$\kappa(\boldsymbol{X}_i)$是通道全局平均操作。

式(4)是 eSE 模块的计算过程，在 eSE 模块中减少一个信道压缩的过程，将 $\boldsymbol{W}_1$ 这个全连接的压缩矩阵去除，这是因为在压缩过程完成之后再用 $\boldsymbol{W}_2$ 进行激发会造成一定特征信息的丢失，在完成式(4)的操作后与输入进来的特征图进行融合，如式(5)所示。

$$A_{eSE}(\boldsymbol{X}_i)=\alpha(\boldsymbol{W}_2(\kappa(\boldsymbol{X}_i)))\tag{4}$$

$$X_{refine}=A_{eSE}(\boldsymbol{X}_i)\otimes\boldsymbol{X}_i\tag{5}$$

原始的 OSA 模块与改进 OSA 模块内部对特征信息的具体操作步骤如图 4、5 所示。

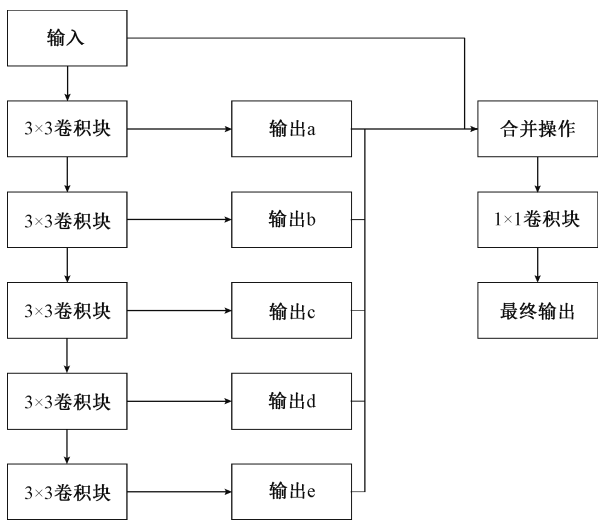


图 4　OSA 模块

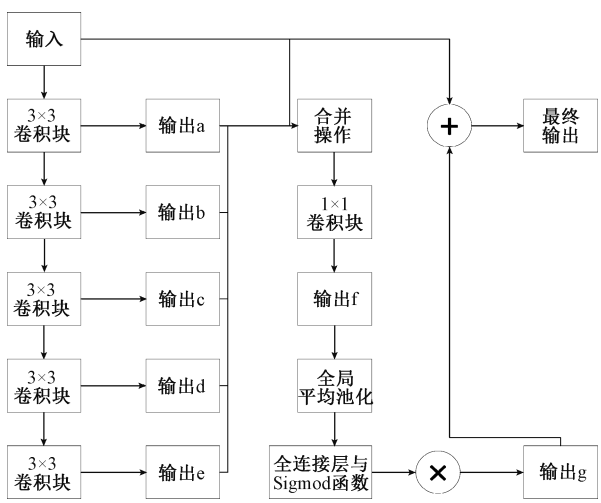


图 5　改进 OSA 模块

1.4　加入小目标检测层

原始图像经过主干网络特征提取得到 20×20、40×40、80×80 尺寸大小的特征图。考虑到无人机航拍图像的中存在大量小目标，密集而且遮挡严重的检测目标，这些小目标需要增大网络的检测尺度，这样才能更好的检测出小目标，所以增加一个小目标的检测层 160×160 尺度的浅层检测层，充分利用浅层的位置信息，以提高小目标的检测效果。但是增加一个小目标检测层不可避免的导致模型规模增大。为此考虑将 20×20 尺度大小的深层检测层删除以达到对模型规模的剪枝操作。但是结果显示，删除了 20×20 检测层会导致大目标的特征信息提取不足，检测效果精度大幅度的降低；另外，本文在删除 20×20 大目标检测层后，突破了 160×160 尺度大小限制的检测层，增加一个 320×320 尺度大小的检测层，4 个检测层尺度大小为 40×40、80×80、160×160、320×320。但是实验结果显示小目标上检测精度并没有提高太多，而且部分大目标如卡车，面包车等的检测精度有所下降。除此之外网络的推理速度大幅度降低，训练得到的模型规模也大幅度增加。综合考虑下为了不影响整体算法的检测精度，保留 20×20 大目标检测层，只增加 160×160 尺度的小目标检测层(图 6)。这样网络规模与推理速度也不会被影响太多，而导致算法退化严重。

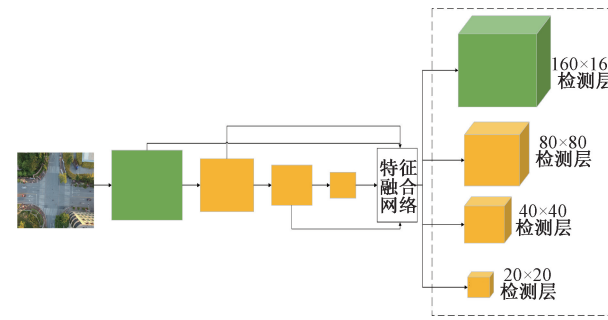


图 6　改进后的检测层

通过 4 倍的下采样的得到的浅层信息具有相对更加精确的位置信息，这是在小目标预测过程中需要的。在特征融合网络中深层具有充足特征信息的特征图与浅层具有相对精确定位信息的特征图进行融合。如图 7 所示，特征金字塔网络[23](feature pyramid networks，FPN)与路径聚合网络[24](path aggregation network，PANet)相结合。特征金字塔网络从上到下传达深层语义特征，而路径聚合网络从下到上传达目标位置信息。通过双向传输的语义信息的融合有助于模型的特征学习能力的提高与提高模型对位置信息的检测能力。

1.5　改进损失函数

小目标因为在图像中密集、占比小、重叠等因素，对定位损失函数更为敏感。原 YOLOv5 中 CIoU 损失函数在预测过程中采用长宽作为统一的损失考虑在预测大目标可以作为定位损失函数，但在小目标检测过程中对位置更

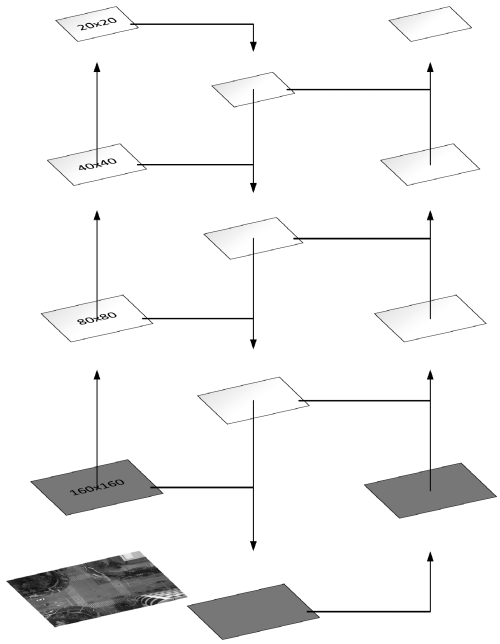


图 7　增加一层的改进特征融合网络结构

加敏锐，再把长宽作为一个统一的整体考虑就导致小目标的位置回归不准确。因此，改进损失函数的核心思想就是拆分长宽，让长宽分别回归，而不是作为一个整体考虑。原 YOLOv5 使用的损失函数如式(6)所示，总损失由 3 部分组成，置信度损失、定位损失和类别损失。置信度损失与类别损失采取的是二元交叉熵作为损失函数。

$$Loss_{obj} = Loss_{loc} + Loss_{conf} + Loss_{cla} \tag{6}$$

$$Loss_{loc} = 1 - CIOU \tag{7}$$

$$Loss_{conf} = -\sum_{i=0}^{B\times B}\sum_{j=0}^{F} I_{ij}^{obj}[C'_i \log C_i^j + (1-C'_i)\cdot \log(1-C_i^j)] - a_{noobj}\sum_{i=0}^{B\times B}\sum_{j=0}^{F} I_{ij}^{noobj}[C'^j_i \log C_i^j + (1-C'^j_i)\cdot \log(1-C_i^j)] \tag{8}$$

$$Loss_{cla} = -\sum_{i=0}^{B\times B} I_{ij}^{obj}\sum_{c\in cla}[\hat{P}_j^i \log P_j^i + (1-\hat{P}_j^i)\cdot \log(1-P_j^i)] \tag{9}$$

式中：B 代表模型将图像分为 $B\times B$ 个小格子；F 代表每个子网格的对应的 anchor box；c 代表被检测的目标所属种类；I_{ij}^{obj} 表示有目标的锚方框；反之 I_{ij}^{noobj} 表示没有目标的锚方框；α_{noobj} 是没有目标方框的置信度损失权重系数。

如图 8 所示，实线方框代表真实方框，表示为 T，虚线方框代表预测方框，表示为 P，预测方框的中心点与真实方框的中心点距离表示为 d，可以包裹预测方框与真实方框的最小方框表示为外方框 C，外方框对角线距离表示为 c。原 YOLOv5 以 CIoU 为定位损失。

$$IoU = \frac{T \cap P}{T \cup P} \tag{10}$$

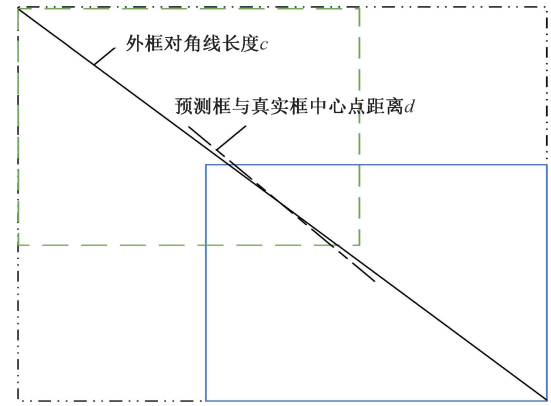


图 8　预测框与真实框交并示意图

$$CIoU = IoU - \frac{b^2}{c^2} - \alpha v \tag{11}$$

$$\alpha = \frac{v}{(1-IoU)+v} \tag{12}$$

$$v = \frac{4}{\pi}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2 \tag{13}$$

式中：IoU 表示的是真实方框 T 和预测方框 P 的交集与并集的比值；b 为真实方框中心点与预测方框中心点的欧几里得距离；α 是一个不参与梯度计算的平衡参数；v 是一个衡量长宽比的一致性参数。

CIoU 把长宽作为一个整体考虑，结合式(1)，当预测方框与真实方框的满足 $\{(w = kw^{gt}, h = kh^{gt}), k = R\}$ 时，CIoU 中此项的惩罚就失去作用。如图 9 所示，其中实线方框是真实方框，虚线是预测方框，当虚线预测方框长宽与实线真实方框长宽比相同时，CIoU 中的长宽损失就此失去了惩罚作用预测框不再进行优化调整。

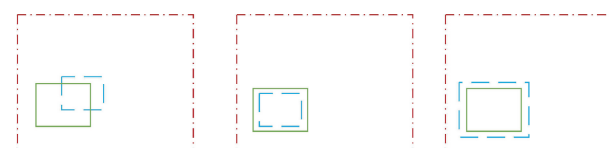


图 9　CIoU 演示图

为此，改进 CIoU 中的长宽损失，不在将长宽作为统一的损失考虑。将长宽分开计算损失，如式(14)所示。

$$EIoU = IoU - \frac{b^2}{c^2} - \frac{d^2(w^p, w^{gt})}{w^2} - \frac{d^2(h^p, h^{gt})}{h^2} \tag{14}$$

式中：IoU 代表预测方框和真实方框的交并比，真实方框和预测方框中心点欧氏距离为 b，外方框对角线为 c，外方框宽为 w，外宽高是 h，预测方框宽 w^p，真实方框 w^{gt}，预测方框高 h^p，真实方框高 h^{gt}。真实方框与预测方框的平方差为 $d^2()$。

如图 10 所示，长宽分别作为损失，不再作为一个整体考虑，这样长宽分别回归，这样回归使得长宽各自计算损失，有更好的定位效果，也更符合小目标在图像中的特

性。这样就解决了CIoU在水平与垂直方向上的误差，并且提高了定位回归的精度。定位损失训练结果如图11所示，在模型刚开始训练阶段，CIoU收敛速度比EIoU快，但是在最终的定位损失上CIoU明显比EIoU高出0.01左右。这表现出EIoU在定位损失上有更好的优势。

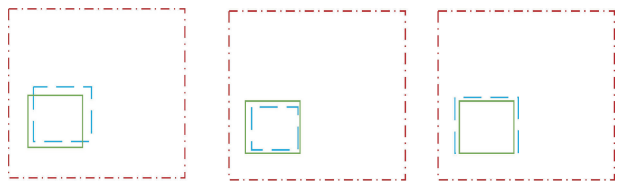


图10 EIoU演示图

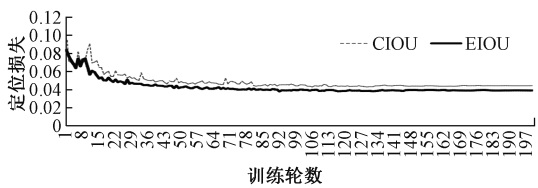


图11 定位损失训练结果比较

1.6 预测框框回归

预测框回归其实质是寻找一种映射函数关系，使得原始方框通过映射函数得到一个和真实方框几乎吻合的回归方框。先验方框和预测方框的在训练过程关系如图12所示，实线方框表示先验方框，虚线方框表示预测框。预测方框是由先验方框经过缩放、平移一系列操作得到。原始图片依据$S\times S$个特征图尺寸划分成个$S\times S$个网格，每个网格会有3个预测方框，每个预测方框中含有4个位置坐标信息和1个类别置信度信息。如果真实方框中有某个目标中心点坐标在某个网格中，该网格负责预测这个目标信息。

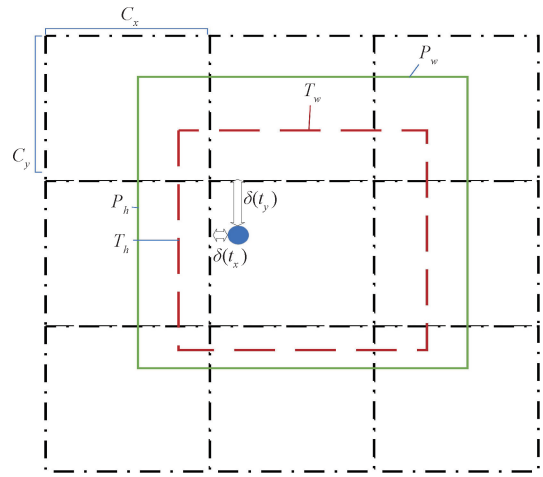


图12 先验方框示意图

目标方框的坐标计算公式为：

$$\begin{aligned} l_x &= 2\sigma(t_x)-0.5+c_x \\ l_y &= 2\sigma(t_y)-0.5+c_y \\ l_w &= p_w(2\sigma(t_x))^2 \\ l_h &= p_h(2\sigma(t_h))^2 \\ P_r(obj)\times IoU(l,obj) &= \sigma(t_0) \end{aligned} \tag{15}$$

式中：c_x、c_y是先验方框所在网格相较于左上角的偏移量；预测方框中心点偏移量为t_x、t_y；预测方框长宽偏移量t_w、t_h；σ是Sigmoid函数用于预测值t_x、t_y、t_w、t_h投影到[0，1]；p_w、p_h是先验方框的宽与高；l_x、l_y是预测宽的中心点坐标，l_w、l_h是预测方框的宽与高，点坐标$\sigma(t_0)$是预测方框的置信度，由预测方框的几率与预测方框和真实方框的交并比IoU相乘得到。$\sigma(t_0)$的设定相对应的阈值可以减少低于阈值的预测方框，从而进一步减少计算量，之后采取非极大值抑制算法NMS输出最终高置信度预测方框。

本文实验因为增加了一个小目标检测层，对于先验框要重新进行计算规划，这里采用K-means+聚类算法与遗传算法对先验框进行回归聚类以达到网络的快速收敛，在进行回归预测过程中发现使用IoU作为评价指标比用欧氏距离作为评价指标效果更好，网络更早的达到收敛。因此在本文采用IoU作为评估指标。

本文使用的先验框与原YOLOv5使用的先验框如表1所示。

表1 特征图大小与先验方框大小关系

特征图大小	本文先验框	YOLOv5先验框
20×20	(116,90),(156,198),(373,326)	(44,25),(36,53),(88,70)
40×40	(30,61),(62,45),(59,119)	(11,12),(23,13)(21,28)
80×80	(10,13),(16,30),(33,23)	(4,5),(6,13),(11,10)
160×160	(5,6),(8,14),(15,11)	

2 实验结果分析

2.1 数据集介绍

实验的数据集采用VisDrone2019。数据集是由天津大学的机器学习与数据挖掘实验室采集制作而成的。数据集包含了10 209张(6 471张用于训练，548张用于验证，3 190张用于测试)图片来自各种无人机拍取，覆盖了中国14个城市，拍摄背景涵盖范围也非常广泛，日常的环境都几乎包含在里面，同时考虑了不同天气气候的因素。

数据集使用LabelImg进标注，标注格式为txt格式。数据集共有10个类别，分别是行人、人、汽车、面包车、巴士、卡车、摩托车、自行车、遮阳篷三轮车和三轮车，标签数目统计如图13所示。其中汽车、面包车、卡车、巴士为大目标，其余的类别是小目标。txt文件中的每一行代表一

个标记目标，共5列，其中坐标信息通过归一化操作使其在[0,1]。如图14所示，第1列为标签类别、第2列标记方框中心点横坐标与原始图片宽度的比值，第3列为标记方框中心的纵坐标与原始图片高度的比值，第4列bbox宽度与图片宽度比值，第5列bbox高度与图片高度比值。

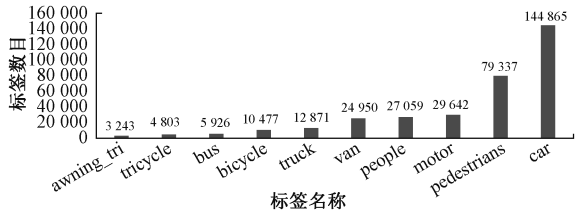


图13 标签数目统计

图14 标注信息样本

2.2 实验环境与参数配置

实验配置环境如下：操作系统Ubuntu18.04.5；GPU服务器为GeForce GTX 3090，显存大小24 GB，CPU处理器为Intel(R) Xeon(R) Gold 6330 CPU @ 2.00 GHz；使用语言为Python 3.8；深度学习方框架为Pytorch 1.7。

本文实验在单卡上进行训练与验证，实验迭代次数为200，每次迭代的batch-size大小设置为32，优化器是SGD随机梯度下降算法。在开始训练时为缓解出现过拟合现象，采取训练预热方法，在训练过程中，采取weight decay的策略防止出现过拟合现象，其他的配置不变，保持默认。

2.3 实验评估指标与实验分析

实验中使用评价指标是平均精度均值(mean average precision，mAP)、平均精度(average precision，AP)、帧率以及模型规模这4种。上述衡量指标计算公式分别是：

$$AP = \int_0^1 P \mathrm{d}R \tag{16}$$

$$mAP = \frac{\sum_{i=1}^{N} AP_i}{N} \tag{17}$$

式中：越大的AP值表示该算法对该类别的检测效果越好；mAP对评价一个算法模型具有全局性，越高的mAP值表示该模型的准确度越好，对这个数据集的效果越好。帧率是每秒检测的帧数越多，模型规模越小表明对硬件设备的要求越低、拥有更好的应用价值。

2.4 消融实验对比

为了更好的验证不同引进模块对改进网络的对网络影响，设计了一系列的消融实验。以YOLOv5s作为BaseLine为了验证改进算法模型的有效性。通过设计7组消融实验，如表2所示每次只加入一个模块，添加各个不同的模块方法为A、B、C、D、E、F、G。从表2可以看出，修改主干网络提取特征网络对于检测精度上是有有提升的，检测精度与BaseLine相比较提升了2.8%。也可以得知在与为改进的OSA模块相比较而言，检测检测精度与检测速度分别提升了0.5%，17 fps，但是模型权重规模上升了0.6 MB。加入小目标检测层与BaseLine相比较，检测精度提升了3%，检测速率下降了10 fps，模型规模上升了2.7 MB；加入EIoU损失函数检测速率与模型权大小没有明显的变化，但是在检测精度上有0.2%的提升。之后的实验加入小目标检测层与改进的VoVNet可以知道检测精度与BaseLine相比较而言提升了检测精度提升了5%，但是检测速率与模型权重都不及BaseLine，检测速率下降了16 fps，模型权重增大5.1 MB。

表2 消融实验分析

方法	模型	mAP/%	帧率/fps	模型大小/MB
A	YOLOv5s	32.8	61	13.8
B	YOLOv5s+OSA模块	35.1	34	15.9
C	YOLOv5s+改进OSA模块	35.6	51	16.5
D	YOLOv5s+小目标检测层	35.8	52	16.4
E	YOLOv5s+EIoU	33.0	61	13.8
F	C+小目标检测层	37.8	45	18.9
G	F+EIoU	37.8	45	18.9

2.5 不同模型的检测效果比较

为了验证本文改进后的模型与其他的目标检测算法具有效性与先进性。本文与原有的算法YOLOv5s、SSD、Faster-RCNN、YOLOv3[25]、YOLOv4、YOLOv5m[26]、YOLOX[27]、PP-YOLOv1[28]、PP-YOLOv2[28]、TPH-YOLO[28]等在数据集VisDrpne2019进行实验对比，实验结果如表3所示。

实验结果表明，在数据集VisDrpne2019上本文提出的改进模型有着相较于其他主流的目标检测算法有着最高得检测精度，相较于作为基线的YOLOv5s检测精度提高了5.2%，与TPF-YOLO改进的检测模型相比较，在检测速率相差不多的情况下，检测精度比其高3%，模型权重规模仅仅是其的28.1%。与改进的模型PP-YOLO系列相比较，在检测精度上与V1版本相比较提升了2.4%，模型规模也仅仅是其的22.6%，检测速率比其高5 fps；与V2版本相比较检测精度提高了1.7%，模型权重规模是其的21.8%，检测速率比其高出7 fps；与YOLOX相比较检测精度比其高出4.3%，检测速率比其底7 fps，模型规模是其的38.7%。表4结果表明，在小目标上，pedestrian、people、bicycle、tricycle、awning、motor这些小目标上

表3 不同模型检测效果比较

模型	主干网络	mAP/%	帧率/fps	模型权重/MB
FasterRCNN	ResNet50	24.6	21	106
SSD	VGG16	21.6	35	96.2
YOLOv3	DarkNet-53	28.6	58	118
YOLOv3-SPPF	DarkNet-53	28.9	58	118.3
YOLOv4	CSPDarkent53	31.4	56	245
YOLOv5s	CSPDraknet53	32.6	**61**	**13.8**
TPH-YOLOv5	改进 CSPDraknet53	34.8	48	69.5
YOLOX	改进 CSPDraknet53	33.5	52	50.4
YOLOv5m	改进 CSPDraknet53	34.6	55	42.2
PP-YOLOv1	ResNet50－vd-dcn	35.4	40	86.1
PP-YOLOv2	改进 CSP v6	36.1	38	89.4
本文模型	VoVNet39	37.8	45	18.9

注：黑体表示实验在 VisDrone2019 数据集上的最优结果

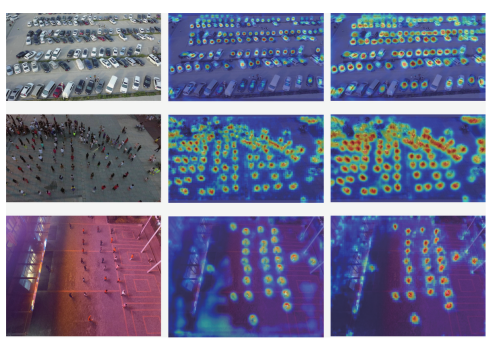
提升的检测精度是 6.6%、6.3%、4.3%、7.7%、6.1%、7.3%，大目标 car、van、truck、bus 提升的检测精度为 2.9%、3.4%、3.2%、3.9%，可以看出本文改进的模型确实在小目标上的检测精度与 BaseLine 相比较提升明显。综上实验结果可以知道本文改进的模型有着最高的检测精度，并且保持了较好的实时性，同时在训练模型权重大小上面保留了一定的优势，因此证明了本文算法模型有一定的优越性。

2.6 模型检测效果对比分析

为了更好的体现出改进模型在小目标上的检测提升效果，采用梯度加权的类激活映射（gradient-weighted class activation mapping，Grad-CAM）技术对特征图进行可视化，如图 15 所示。通过对比两组可视化的结果可以知道，本文可视化特征图相较于 YOLOv5s 的亮度更加大，颜色也更加深，这种现象意味着在主干特征网络与特征融合网络的阶段捕获了更多的特征信息，尤其是在凸图像中的小目标信息。相较于 YOLOv5s 的热力图，捕获了更多的小目标信息，充分捕特征图中的细小差异。原 YOLOv5as 算法模型在远端密集的小目标与近端目标获得注意力不够多，在改进算法模型中，远端小目标与近端目标都获得了比原算法更多关注。

表4 本文模型与 YOLOv5s 各个类别的检测精度 （%）

模型	pedestrian	people	bicycle	car	van	truck	tricycle	awning	bus	motor
YOLOv5s	48.4	9.2	12.1	75.5	39.5	37.4	15.2	15.0	48.3	25.7
本文模型	55.0	15.5	16.4	78.4	42.9	40.6	22.9	21.1	52.2	33.0



（a）原图 （b）原热力图 （c）改进热力图

图15 原 YOLO 与改进 YOLO 热力图

（a）原算法 （b）改进算法

图16 检测效果对比

图 16 所示为两种模型的检测效果，图 16A 检测背景相对简单，目标相对单一且分散的条件下，原 YOLO 与本文算法模型都可以较好地检测出相应的目标。但是在图 16B 和 C 中，图像背景复杂，目标密集并且重叠较多，可以看出原 YOLOv5s 在远端与密集条件下出现检测乏力的情况，但是本文模型在远端与密集情况下检测出来的效果较好。这在一定程度佐证了改进算法的优势。

3 结 论

本文实验在原有的 YOLOv5s 模型的基础上更换主干特征提取网络，消除了因为卷积网络深度导致的梯度衰减与特征信息丢失的问题并且改进 OSA 模块，直接将 OSA 模块中的原始输入连接上输出同属加入一个 eSE 模块增加输出特征图之间的表达能力；在特征融合网络部分重新构造一个新的 160×160 小目标检测层以适应小目标在航拍图像中占比小、密集且遮挡多难以检测特性；最后

修改损失函数CIoU，原YOLov5s模型中长宽作为一个整体计算损失会导致当真实框与预测框长宽比相同时出现长宽损失失去惩罚作用导致预测框的回归不够精确的问题，因此拆开长宽分别计算这样长宽损失也能够适应小目标的回归。在实验的结果可以知道，检测精度提升了5.2%，相比于原模型提升了16.56%；检测速率降低了16 fps；模型规模增大了5.1 MB。虽然本次实验在检测精度上有所提升，但是牺牲了检测速率与模型大小。下一步研究方向是降低模型大小与进一步提高模型的检测速率。

参 考 文 献

[1] CHENG B, WEI Y, SHI H, et al. Revisiting RCNN: On awakening the classification power of faster-RCNN[C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 453-468.

[2] CHEN Y P, LI Y, WANG G, et al. A multi-strategy region proposal network[J]. Expert Systems with Applications, 2018, 113(113): 1-17.

[3] UIJLINGS J R R, VAN DE SANDE K E A, GEVERS T, et al. Selective search for object recognition[J]. International Journal of Computer Vision, 2013, 104(2): 154-171.

[4] SUN X, WU P, HOI S C H. Face detection using deep learning: An improved faster RCNN approach[J]. Neurocomputing, 2018, 299(299): 42-50.

[5] LIU W, ANGUELOV D, ERNAN D, et al. SSD: Single shot multi-box detector[C]. Computer Vision-ECCV 2016. Springer, 2016: 21-37.

[6] REDOMON J, DIVVVALA S, GIRSHICK R, et al. You only look once: Unified real-time object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.

[7] 陈旭，彭冬亮，谷雨. 基于改进YOLOv5s的无人机图像实时目标检测[J]. 光电工程，2022，49(3)：69-81.

[8] 王恒涛，张上，陈想，等. 轻量化无人机航拍目标检测算法[J]. 电子测量技术，2022，45(19)：167-174.

[9] 王鼎山，贾世杰. 基于目标感知增强的无人机航拍目标检测[J]. 计算机工程与设计，2022，43(7)：2071-2077.

[10] 吕辉，董帆. 基于YOLOv4的复杂交通状况下多目标检测算法[J]. 国外电子测量技术，2022，41(12)：41-47.

[11] SU K, YAN W, WEI X, et al. Stereo VoVNet-CNN for 3D object detection[J]. Multimedia Tools and Applications, 2022, 12(11): 1-11.

[12] 李壮飞，杨风暴，郝岳强. 一种基于残差网络优化的航拍小目标检测算法[J]. 国外电子测量技术，2022，41(8)：27-33.

[13] 单明陶，高玮玮. 改进YOLOv4的内丝接头密封面缺陷检测算法[J]. 电子测量与仪器学报，2022，36(5)：120-127.

[14] 石欣，卢灏，秦鹏杰，等. 一种远距离行人小目标检测方法[J]. 仪器仪表学报，2022，43(5)：136-146.

[15] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020: 12993－13000.

[16] 马晓东，魏利胜，刘小珲. 基于新型YOLO v5算法的磁悬浮球精确识别[J]. 电子测量与仪器学报，2022，36(8)：204-212.

[17] LEE Y, PARK J. Centermask: Real-time anchor-free instance segmentation[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 13906－13915.

[18] HUANG G, LIU Z, VAN D M L. Densely connected convolutional networks [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4700-4708.

[19] MA N, ZHANG X, ZHENG H T, et al. Shufflenet V2: Practical guidelines for efficient CNN architecture design[C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 116-131.

[20] MA R, WANG J, ZHAO W, et al. Identification of maize seed varieties using MobileNetV2 with improved attention mechanism CBAM[J]. Agriculture, 2022, 13(1): 1-17.

[21] WU Z, SHEN C, VAN DEN HENGEL A. Wider or deeper: Revisiting the resnet model for visual recognition [J]. Pattern Recognition, 2019, 90 (90): 119-133.

[22] GAO R, WANG T. Motion deblurring algorithm for wind power inspection images based on Ghostnet and SE attention mechanism[J]. IET Image Processing, 2023, 17(1): 291-300.

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

[24] LI S, LI Y, LI Y, et al. YOLO-firi: Improved YOLOv5 for infrared image object detection [J]. IEEE Access, 2021, 9(9): 141861－141875.

[25] HASSAN N I, TAHIR N M, ZAMAN F H K, et al. People detection system using YOLOv3 algorithm[C]. 2020 10th IEEE International Conference on Control System, Computing and Engineering (ICCSCE). IEEE, 2020, 45(8): 131-136.

[26] WU T H, WANG T W, LIU Y Q. Real-time vehicle

and distance detection based on improved yolov5 network[C]. 2021 3rd World Symposium on Artificial Intelligence (WSAI). IEEE, 2021: 24-28.

[27] LIU B, HUANG J, LIN S, et al. Improved YOLOX-S abnormal condition detection for power transmission line corridors[C]. 2021 IEEE 3rd International Conference on Power Data Science (ICPDS). IEEE, 2021: 13-16.

[28] LI Y, HUANG H, CHEN Q, et al. Research on a product quality monitoring method based on multi scale PP-YOLO[J]. IEEE Access, 2021, 9(45): 80373-80387.

作 者 简 介

杨辉羽，硕士，主要研究方向为深度学习，目标检测。
E-mail：vyanghuiyu@163.com

李海明，博士，教授，主要研究方向为深度学习，智能信息处理等。
E-mail：lhm@shiep.edu.cn