

面向无人机航拍图像小目标检测方法

吴海斌, 张亚, 胡鹏

(安徽理工大学人工智能学院, 安徽淮南 232001)

摘要: 针对航拍图像目标检测中小目标特征模糊问题, 提出一种改进 YOLO_v5x 的目标检测算法。通过在 YOLO_v5x 的主干和颈部网络中添加空间到深度 (space-to-depth, SPD) 模块来减少细粒度信息丢失; 在检测输出端添加 1 个小目标预测头, 提高算法学习低分辨率特征的效率; 引入协调注意力 (coordinate attention, CA) 机制, 将横向和纵向的位置信息编码到通道注意中, 增强网络对不同维度特征的提取能力; 在完整交并比 (complete-intersection over union, CIoU) 损失函数的基础上引入 Alpha 交并比 (α -IOU) 损失函数, 获得更准确的边界框回归, 实现图像中目标更精确的定位。通过在 Visdrone 数据集上对改进 YOLO_v5x 算法进行训练和对比实验, 结果表明: 相比于原 YOLO_v5x, 改进目标检测算法的平均检测精度提升了 7.8%, 小目标检测的平均精度达 23.9%, 能够有效识别无人机航拍图中的小目标; 相比于 RetinaNet、YOLOX-S、Grid-RCNN 等目标检测算法, 改进目标检测算法的小目标检测平均精度最高, 在当前主流检测小目标算法中达到先进水平。

关键词: 无人机; 目标检测; 航拍图像; 注意力机制

中图分类号: TP 391.4 **文献标志码:** A **doi:** 10.12415/j.issn.1671-7872.23087



A Small Target Detection Method for Unmanned Aerial Vehicle Aerial Photography Images

WU Haibin, ZHANG Ya, HU Peng

(School of Artificial Intelligence, Anhui University of Science and Technology, Huainan 232001, China)

Abstract: Aiming at the problem of fuzzy features of small targets in aerial image detection, an improved YOLO_v5x target detection method was proposed. A space-to-depth (SPD) module was added to the backbone and neck network of YOLO_v5x to reduce the loss of fine-grained information, and a small target prediction head was added to the detection output to improve the efficiency of the algorithm in learning low-resolution features. At the same time, the coordinate attention (CA) mechanism was introduced to encode the horizontal and vertical position information into the channel attention to enhance the ability of the network to extract different dimensional features. In order to improve the target positioning accuracy, the Alpha intersection over union (α -IOU) loss function was introduced based on the complete-intersection over union (CIoU) loss function. To obtain more accurate bounding box regression, to achieve more accurate target positioning in the image. Through training and comparative experiments on the improved YOLO_v5x algorithm on the Visdrone datasets. The results show that compared with

收稿日期: 2023-05-29

基金项目: 安徽省高校自然科学基金项目 (2022AH050801); 淮南市科技计划项目 (2021005); 安徽理工大学校级重点项目 (QNZD2021-02)

作者简介: 吴海斌 (1997—), 男, 安徽安庆人, 硕士生, 主要研究方向为计算机视觉。

通信作者: 张亚 (1978—), 男, 安徽淮南人, 副教授, 主要研究方向为信号处理、机器视觉等。

引文格式: 吴海斌, 张亚, 胡鹏. 面向无人机航拍图像小目标检测方法 [J]. 安徽工业大学学报 (自然科学版), 2024, 41(1):65-73.

the original YOLO_v5x, the average detection accuracy of the improved target detection algorithm was increased by 7.8%, and the average detection accuracy of small target detection was up to 23.9%, which can effectively identify small targets in unmanned aerial vehicle aerial photos. Compared with other target detection algorithms such as RetinaNet and YOLOX-S, the average precision of small target detection was the highest in the improved target detection algorithm, reaching the advanced level among the current mainstream small target detection algorithms.

Keywords: unmanned aerial vehicle; target detection; aerial image; coordinate attention mechanism

随着深度学习技术的快速发展,基于深度学习的无人机航拍图像检测在城市交通、环境监测和灾害检测等领域的应用越来越广泛^[1]。由于无人机飞行高度和拍摄角度的独特性,导致捕获图像中存在大量小目标,准确识别无人机图像中的小目标是提升无人机航拍目标检测的关键。

目前,基于深度学习的目标检测算法一般被分为两类。第一类是以区域卷积神经网络^[2](region-convolutional neural network, R-CNN)、快速的区域卷积神经网络^[3](faster-region-convolutional neural network, Faster-R-CNN)为代表的二阶段算法,包含候选框的提出和分类器的构建。此类方法在目标检测过程中精度较高,但检测速度较慢,难以满足现代化生产对高速实时检测的需求;易丢失图像中小目标的位置信息,导致小目标特征的学习效率较低。另一类是由 Joseph 等^[4]提出的以 YOLO 为代表的一阶段目标检测算法,将目标检测转化为回归问题,通过网格划分特征图,得到目标的边界框位置及所属类别。此类方法检测速度较高,能够实现端到端的检测,但在准确性上表现不佳。随着 YOLO_v2^[5], YOLO_v3^[6] 和 YOLO_v4^[7] 相继提出,在保持较高检测速度的同时检测精度也在逐步提高。YOLO_v5^[8] 是 YOLO 系列的一个延伸,通过加入 Focus 结构和自适应图片缩放,可减少算法的参数数量和计算量;相对于 YOLO_v4 等系列的 YOLO 算法,YOLO_v5 的检测速度和精度有显著提高,但在检测小物体上由于缺乏对小目标特征的有效提取,检测精度有待提升。

针对无人机航拍图像中小目标多的问题,Zhu 等^[9] 提出基于 transformer 的可变形检测转换器(deformable detection transformer, DDETR),利用可变形注意力模块来提高小目标的检测精度,但由于采用残差网络提取特征,导致低层次特征信息丢失,小目标检测精度依然不够理想。田港等^[10] 将三尺度检测模型改进为双尺度检测模型,并通过通道裁剪减小模型内存占用和计算量,但导致小目标检测的精度降低。张智等^[11] 通过密集高级组合模式连接双 swin-transformer 骨干网络,并结合特征金字塔构

建小目标特征聚合网络,在不损失语义信息的同时得到对小目标更好的特征描述,但由于改进算法在不经训练就直接在 CARPK 数据集上进行测试,导致小目标的检测精度仍较低、泛化性能较差。郝立等^[12] 提出复杂环境下无人机作战的小目标检测方法,可改善复杂光照下小目标的识别,但没有考虑实际应用场景,导致模型的泛化能力较低。陈映雪等^[13] 提出根据相邻帧信息预测当前帧丢失目标的位置范围,并利用检测结果进行修正,实现丢失目标的再召回,但该网络训练针对性强,属于专用型目标检测模型。上述小目标检测方法分别考虑了检测的速度、精度,但提高速度会导致精度下降,同时小目标种类的单一、环境的固定导致小目标检测方法分类和定位的局限性。YOLO_v5x 是 YOLO_v5 中网络结构最深、精度最高的版本,针对无人机航拍图目标检测存在的问题,提出一种基于 YOLO_v5 小目标检测方法,通过对 YOLO_v5x 的结构进行改进,以增强图像中目标定位的准确性,提高图像分辨率低且目标较小的检测精度。

1 YOLO_V5x 算法改进

YOLO_v5 网络结构是在 YOLO_v4 的基础上进行改进的,具有极强的网络部署能力^[7]。YOLO_v5x 是深度、宽度最大的网络,结构包含输入端、主干网络、颈部网络和输出端 4 个部分,如图 1。

输入端主要对图像进行灰度处理等预操作,生成满足输入主干网络的特征图像。在网络训练时,为进一步提高检测的精度和速度,使用 Mosaic 数据增强方法丰富被检测对象的背景,采用自适应锚选框对图像进行缩放。主干网络也称特征提取网络,将预处理图像输入 Focus 模块,为每隔一个点获得相应的特征值,并连接 4 个独立的特征层以获得最终结果,将图像的宽度和高度信息集中到通道中,解决下采样带来的信息丢失问题^[14]。同时使用效果与跨级网络(cross stage partial, CSP)结构基本相似的 C3 模块替换 BottleneckCSP^[15] 模块,减少算法的参数数量,提高处理速度。Conv 模块是对输入的特征图进行

卷积操作, 目的是调整特征图的通道数。快速空间金字塔池化 (fast spatial pyramid pooling, SPPF) 模块有效避免了对特征图裁剪、缩放导致的图像失真, 可提高检测精度。颈部网络进一步提取图像的特征, 采用特征金字塔^[16] (feature pyramid network, FPN) 和路径聚合结构^[17] (path aggregation network, PANet),

前者通过自上向下传达语义特征, 后者则自下向上传递定位信息, 通过融合上下层的特征信息, 加快信息的传播, 提高目标检测的性能。输出端是对特征图不同尺度的目标进行最后的预测, 包括 1 个分类分支和 1 个回归分支, 判断特征图中是否存在对应的对象。

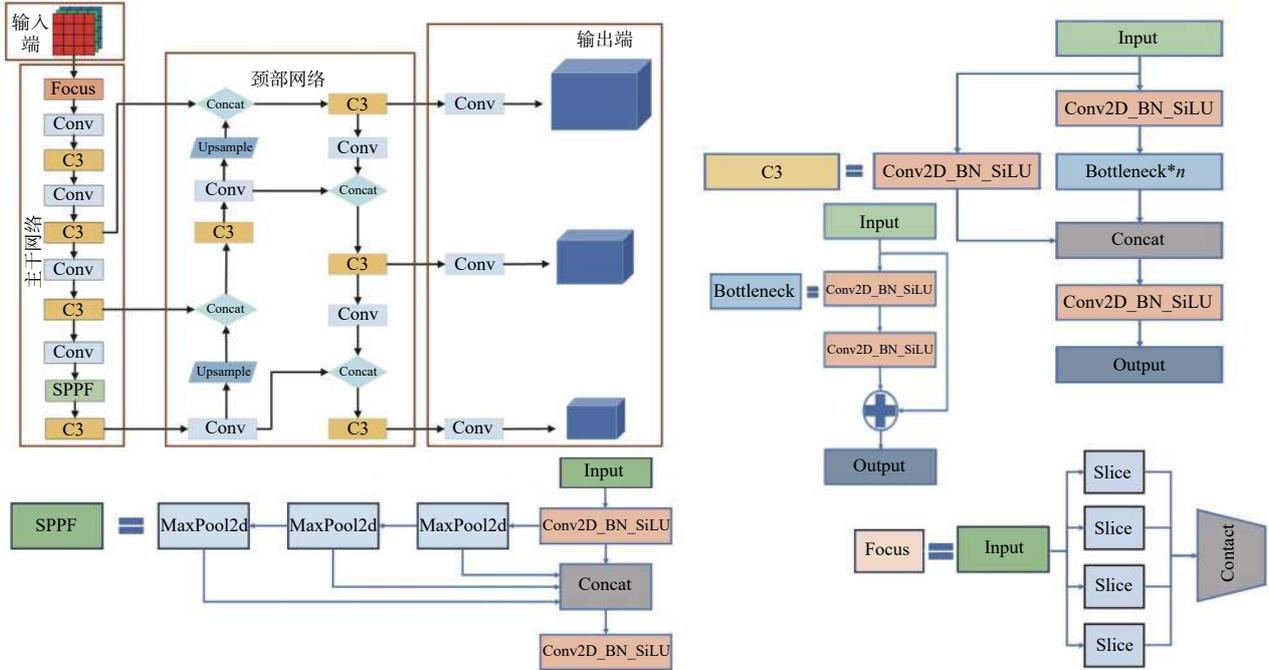


图1 YOLO_v5x 网络结构

Fig. 1 YOLO_v5x network structure

1.1 空间到深度模块

在计算机视觉的目标检测任务中, 卷积神经网络表现突出, 但输出结果的精度与图像分辨率成正比。在多数研究场景中, 输入特征图分辨率较高, 图像中存在大量的冗余信息充足, 跨步卷积和池化造成的特征丢失可忽略。当图像中目标较小、分辨率较低时, 卷积神经网络结构受到细粒度特征信息丢失和特征模糊的影响, 性能明显降低。文中在此基础上引入空间到深度 (space-to-depth, SPD) 模块, SPD 是一个空间到深度的层结构^[18]。基于图像转换技术, 在原始图像输入神经网络前对图像进行缩放, 且将其应用在网络内部和下采样特征映射中, 在不丢失信道维度中信息的同时对特征图进行下采样^[19]。由于无人机航拍的特殊性, 获得的图像大部分分辨率低及目标较小。

对于 SPD 机构, 当特征图 X 大小为 $S_1 \times S_2 \times C_1$, 子序列特征图为

$$f_{i,j} = X[i : S_1 : a, j : S_2 : a], (i = 0, \dots, a-1; j = 0, \dots, a-1) \quad (1)$$

通常, 给定特征图 X , $S_1 = S_2 = S$, S 为特征图的

空间维度, C_1 为特征图的通道维数, a 为特征图裁剪因子, 子图 $f(x,y)$ 由所有的向量 $X(i,j)$ 组成, 每个子特征图按裁剪因子对特征图 X 进行等间隔下采样。当 $a=2$ 时, 得到 $f_{0,0}, f_{1,0}, f_{0,1}, f_{1,1}$ 4 个子特征图, 每个子特征图的大小为 $(\frac{S}{2}, \frac{S}{2}, C_1)$ 。然后沿通道维数将子特征图连接, 得到空间维度降低 2 倍、通道维数增加 4 倍的特征图 X' , 即将特征图 $X(S, S, C_1)$ 转换为特征图 $X'(\frac{S}{2}, \frac{S}{2}, 4C_1)$, 如图 2。

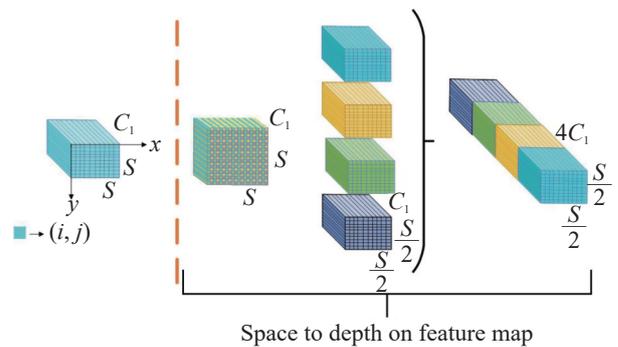


图2 当 $a=2$ 时 SPD 的图解

Fig. 2 Diagram of SPD at $a=2$

1.2 协调注意力机制

注意力机制是通过给输入的特征图赋予不同的权重,选择有用的信息,从而得到更好的特征。通道注意力机制可提高算法的性能,但由于忽略位置信息,导致空间信息的缺失。文中在此基础上引入协调注意力机制,将位置信息嵌入通道注意力机制。CA机制将通道注意力分解为沿2个空间方向聚类特征的编码过程,在捕获远程依赖关系的同时保持精确的位置信息,最后获得1对方向感知和位置敏感的注意图,将其应用在输入特征图,增加目标的特征表达能力^[20]。无人机捕获的航拍图中,多数目标较小且覆盖区域较大,使用CA机制可使算法更关注有用的对象,对提高网络模型的收敛具有较优的效果。CA机制包含协调信息嵌入和协调注意力生成,如图3。

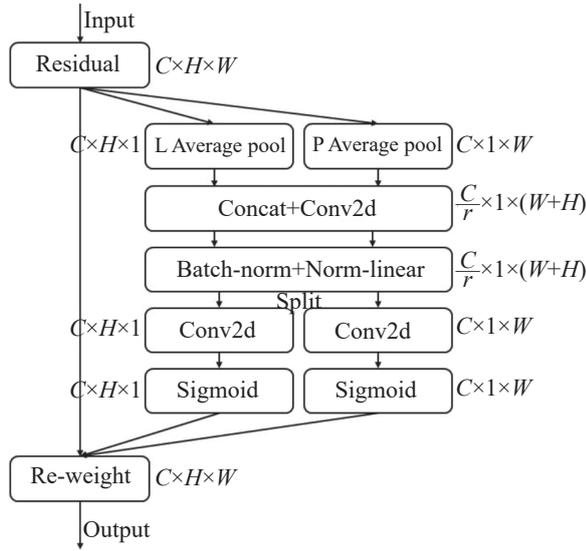


图3 协调注意力机制

Fig. 3 Coordinate attention mechanism

1.2.1 协调信息嵌入

在视觉任务中,位置信息对捕获空间结构十分重要。全局平均池化是将空间信息压缩到通道中,导致位置信息难以保存。因此,CA机制将全局平均池化分解为一对一维特征操作。具体是将输入特征图 X 分别沿水平(level, L)和垂直(perpendicularity, P)方向对每个通道进行全局平均池化(average pool)操作。

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i \leq W} x_c(h, i) \quad (2)$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j \leq H} x_c(j, w) \quad (3)$$

式中: h 为高度,表示特征图 X 第 c 通道处的高度,大

小为 $0 \sim H$; x_c 表示坐标的灰度值; $z_c^h(h)$ 表示高度为 h 的第 c 通道的灰度总值; w 为宽度,表示特征图 X 第 c 通道处的宽度,大小为 $0 \sim W$; $z_c^w(w)$ 表示宽度为 w 的第 c 通道的灰度总值; W 为特征图 X 的宽度; H 为特征图 X 的高度; C 为特征图 X 的通道数,分别生成尺寸为 $C \times H \times 1$ 和 $C \times 1 \times W$ 的特征图。

1.2.2 协调注意力生成

为更好地利用协调信息嵌入变换产生的表征,提出简单捕获感兴趣区域以及有效捕捉通道间关系的协调注意力生成变换。

$$f = \delta(F_1([z^h, z^w])) \quad (4)$$

式中: $[,]$ 表示空间维度的串联(concat); F_1 表示利用 1×1 卷积函数进行降维; δ 为非线性激活函数(non-linear),生成特征图 $f \in \mathbf{R}_r^{C \times (H+W) \times 1}$ 。

$$g^h = \sigma(F_h(f^h)) \quad (5)$$

$$g^w = \sigma(F_w(f^w)) \quad (6)$$

其中: σ 为Sigmoid型函数,将 f 沿空间维度分为(split) $f^h \in \mathbf{R}_r^{C \times H \times 1}$ 和 $f^w \in \mathbf{R}_r^{C \times 1 \times W}$, 1×1 卷积变换 F_h 和 F_w 表示将 f^h 和 f^w 进行升维操作,结合 σ 得到最后的注意力向量 $g^h \in \mathbf{R}^{C \times H \times 1}$ 和 $g^w \in \mathbf{R}^{C \times 1 \times W}$ 。最后,协调注意力的输出可写为

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (7)$$

1.3 边界框回归损失函数

边界框回归是通过预测框来定位图像中的目标,提高检测器学习特征信息的效率。YO-LO_v5的边界框损失采用完整交并比(complete-intersection over union, CIOU)损失函数,公式如下所示:

$$L_{\text{CIOU}} = 1 - U + \frac{\rho^2(b, b')}{d^2} + \beta v \quad (8)$$

$$v = \frac{4}{\pi^2} (\arctan \frac{w^t}{h^t} - \arctan \frac{w}{h})^2, \beta = \frac{v}{(1-U)+v} \quad (9)$$

其中: L_{CIOU} 为CIOU函数的损失定义; b 和 b' 代表预测框和真实框的中心点, t 表示真实框; ρ 为预测框和真实框之间的欧氏距离; d 为同时包含预测框和真实框的最小封闭区域的对角线距离; β 为权衡的参数; v 为衡量长宽比一致性的参数; w^t 和 h^t 为真实框的宽度和高度; U 为预测框和真实框的交并比。CIOU损失函数考虑了预测框的重叠面积、中心点距离和长宽比,但描述的是相对值,没有考虑难易样本的平衡问题,导致算法在检测小目标的定位精度较低。

文中在CIOU损失函数的基础上引入 α -IOU损失函数^[21]。通过调节 α 值实现不同水平的回归精度, α 对不同算法或数据集的敏感程度不同。 α -IOU是

通过提高 IOU 目标的损失和梯度的权重来获得更好更快的回归精度, 特别是在小目标的定位和检测性能上更加突出。 α -IOU 可广泛用于改进最先进的检测算法, 且无需引入额外的参数, 同样无需增加训练及推理时间, 其表达式如下:

$$L_{\alpha\text{-IOU}} = \frac{1 - U^\alpha}{\alpha}, \alpha > 0 \quad (10)$$

$$L_{\alpha\text{-CIoU}} = 1 - U^\alpha + \frac{\rho^{2\alpha}(b, b')}{c^{2\alpha}} + (\beta v)^\alpha \quad (11)$$

式 (11) 为基于 α -IOU 损失函数的 α -CIoU 损失函数。

1.4 改进 YOLO_v5x 算法的网络结构

无人机航拍图中的目标尺寸较小, 目标特征不够明显, 导致算法的检测精度降低。文中在此基础上添加 1 个小目标预测头, 结合其他 3 个预测头, 在 YOLO_v5x 算法中添加 SPD 模块、CA 机制、 α -IOU 损失函数和小目标预测头。改进 YOLO_v5x 算法的网络结构如图 4。

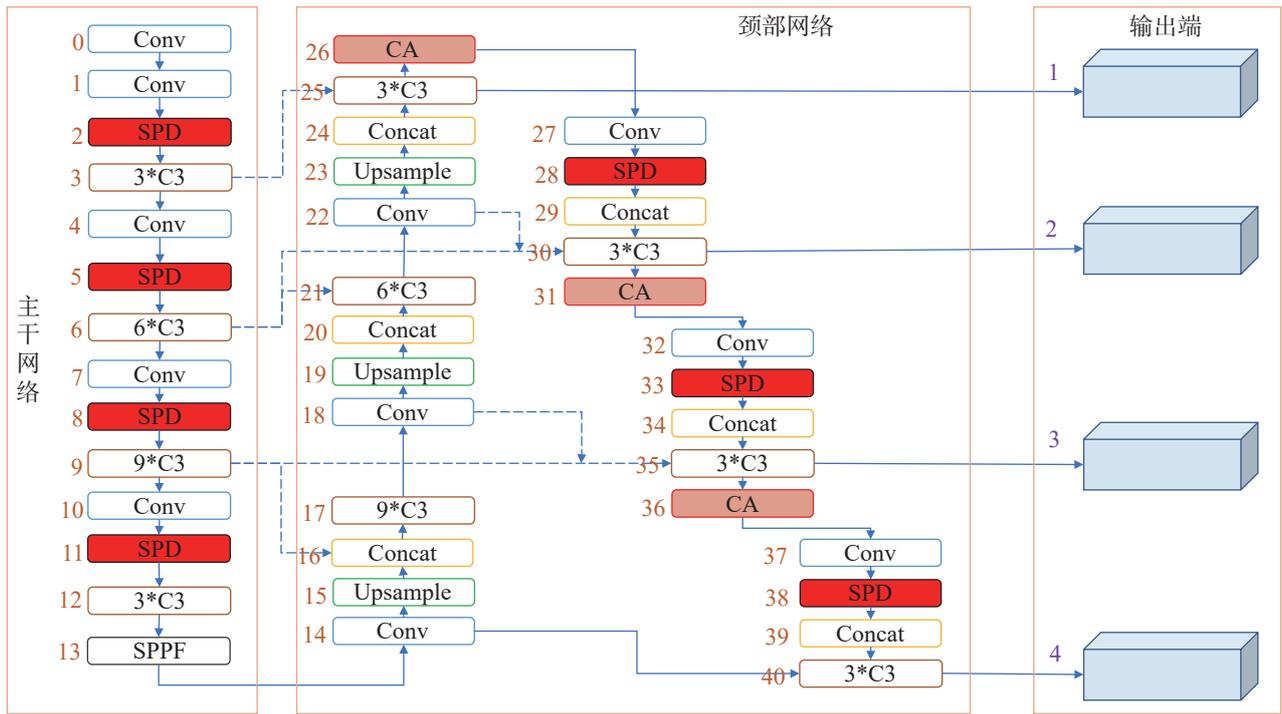


图4 改进 YOLO_v5x 算法的网络结构

Fig. 4 Network structure of improved YOLO_v5x algorithm

2 实验与结果分析

为评估改进算法的结构性能, 采用 Visdrone^[22] 无人机目标数据集对改进前后的算法进行目标检测实验。使用 COCO 数据集^[23] 的评价指标 (AP, AP₅₀, AP₇₅, AP_S, AP_M, AP_L) 评价目标检测算法的检测精度。其中: AP 表示平均精度, 指精度-召回曲线的面积; AP₅₀, AP₇₅ 分别表示计算平均精度时, 以 0.50, 0.75 为 IOU 的阈值; AP_S 表示小目标即像素面积小于 32×32 的目标框检测的平均精度, 是本文改进算法的核心评价指标; AP_M 表示中等目标即像素面积在 32×32~96×96 的目标框检测的平均精度; AP_L 表示大目标即像素面积大于 96×96 的目标框检测的平均精度。

2.1 实验过程

采用中国移动云操作系统, 搭配 8core、50 GB

RAM CPU、Tesla V100-32 GB GPU, 使用 CUDA11.6、Python 3.9 框架; 设置训练次数 (epoch)200、批量大小 (batch-size)4, 衰减系数 0.000 5、学习率 0.01。Visdrone 无人机目标检测数据集共 10 209 张图像, 其中训练集 6 471 张、验证集 548 张、测试集 3 190 张, 包括 10 个类别, 分别是行人、人、自行车、汽车、货车、卡车、三轮车、带篷三轮车、公交车和摩托车。实验算法训练过程包括 5 个步骤, 分别是数据集的处理、模型的搭建、参数设置、训练和验证, 如图 5。

实验过程包括: 在相同参数、数据集和环境条件下, 通过在 YOLO_v5x 算法的基础上分别引入 SPD 模块、CA 机制、 α -IOU 损失函数以及添加小目标预测头等模块, 验证改进 YOLO_v5x 算法的检测效果; 同时, 将改进 YOLO_v5x 算法与目前主流

的目标检测算法进行对比实验,验证其可靠性和泛化性。

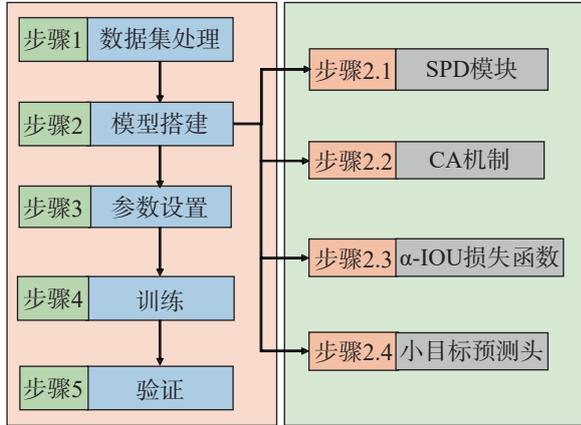


图5 算法训练过程

Fig. 5 Model training process

2.2 实验结果分析

2.2.1 α 不同取值时的小目标检测精度

为验证 α -IOU中 α 值对改进算法检测小目标精度的影响,修改 α -IOU中 α 值为 $[1/2, 10]$,在Visdrone无人机目标检测数据集上选取部分图像进行训练测试,结果如图6。由图6可看出: $\alpha = [1/2, 3]$ 时,AP

和 AP_{50} 随 α 的增大而逐渐提升; $\alpha = (3, 10]$ 时,AP和 AP_{50} 随 α 的增大而逐渐下降。因此,文中后续实验均取 $\alpha = 3$ 。

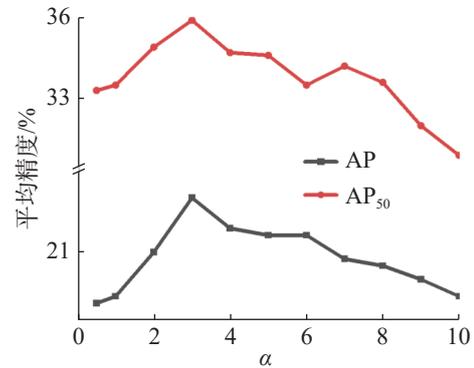


图6 α 不同取值时改进算法对应的平均精度

Fig. 6 Average accuracy of the improved algorithm with different values of α

2.2.2 添加不同模块时的小目标检测精度

为验证YOLO_v5x算法在相同参数和环境下引入 α -IOU损失函数、CA机制及添加SPD模块、小目标预测头(Head)等模块时的小目标检测效果,对其进行消融实验,结果如表1。并将改进前后YOLO_v5x算法的实验结果可视化展示,如图7。

表1 不同模块组合对应的小目标检测精度

Tab. 1 Small target detection accuracy corresponding to different modules

| α -IOU | SPD | CA | Head | AP/% | AP_{50} /% | AP_{75} /% | AP_S /% | AP_M /% | AP_L /% | FPS/(帧/s) |
|---------------|-----|----|------|------|--------------|--------------|-----------|-----------|-----------|-----------|
| × | × | × | × | 25.5 | 42.6 | 18.4 | 12.1 | 28.2 | 33.7 | 37.3 |
| × | × | √ | × | 26.1 | 42.3 | 20.9 | 13.0 | 29.9 | 33.9 | 44.2 |
| √ | × | × | × | 26.5 | 42.3 | 23.4 | 13.7 | 34.9 | 49.4 | 46.2 |
| √ | √ | × | × | 32.3 | 51.5 | 25.3 | 15.2 | 36.0 | 44.7 | 54.3 |
| √ | √ | √ | √ | 33.3 | 52.3 | 32.2 | 23.9 | 42.5 | 44.9 | 8.23 |

注: ×表示未添加该列模块; √表示添加该列模块。

表1中第一行作为消融实验的基线,为YOLO_v5x算法;第二行引入CA机制模块,通过将目标位置信息编码到通道注意中,算法更关注有用的特征信息,小目标的检测平均精度 AP_S 提高0.9%,同时检测速度也提高了6.9帧/s;第三行引入 α -IOU损失函数,通过获得更准确的边界框回归精度,实现算法对目标更精确的定位,小目标检测平均精度 AP_S 提高1.6%,检测速度也提高了8.9帧/s;第四行在 α -IOU的基础上添加SPD模块,在减少细粒度信息丢失的同时进行下采样,提高算法对低分辨率特征的提取能力,小目标检测平均精度 AP_S 提高3.1%,检测速度也提高了17帧/s。由于无人机拍摄高度和角度的变化,航拍图像中小目标较多并且特征不明显,最后一行为添加CA机制、 α -IOU

损失函数、SPD模块和1个预测头,算法的小目标检测平均精度 AP_S 得到较大提高,比原基线算法提高了11.8%,但是由于参数量的增加,检测速度降低比较明显。

图7中左侧黄色标注的内容为观察对象,将检测结果进行局部放大显示于右下角方框。其中:第二列为小目标密集环境改进前后YOLO_v5x算法的检测结果;第三列为在弱光环境改进前后YOLO_v5x算法的检测结果。由图7可看出:在小目标密集的情况下,改进算法依然能够有效检测出标识框中的行人和自行车;在弱光环境导致的小目标模糊的情况下,改进算法能够有效检测出标识框中的汽车和公交车,改进的YOLO_v5x算法表现出较强的可靠性和泛化性。

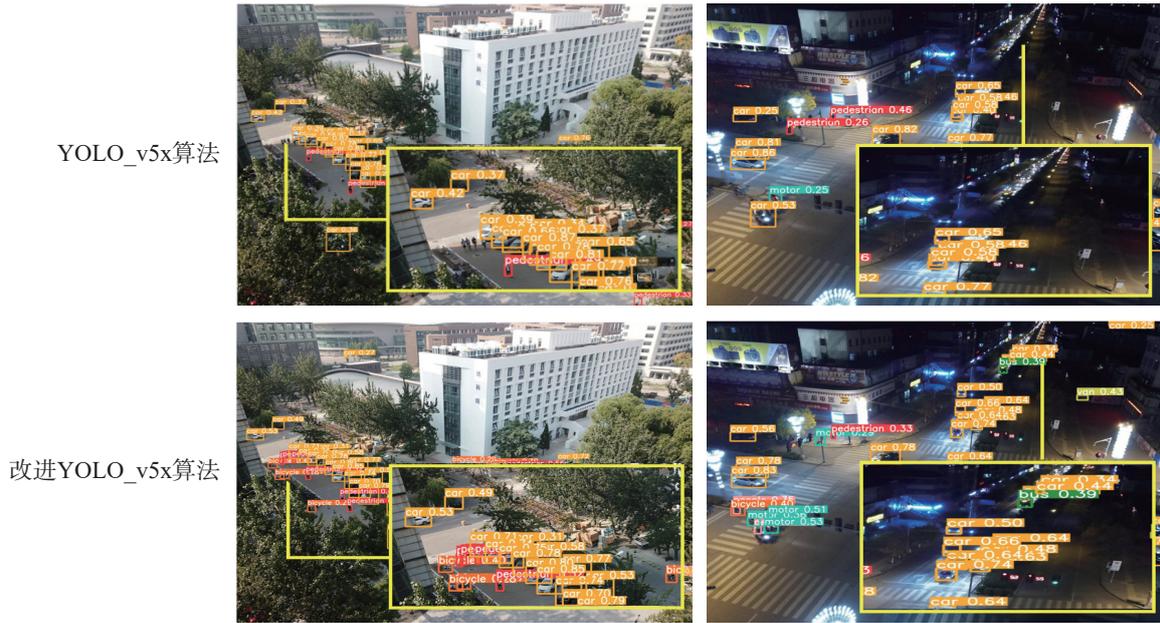


图7 改进前后算法的检测结果

Fig. 7 Detection results of algorithms before and after improvement

2.2.3 不同算法的小目标检测效果

为验证改进后算法的检测性能,同时选取当前较流行的算法,如 RetinaNet^[24]、YOLOX-S^[25]、Grid-RCNN^[26]和改进 Double-Head RCNN^[27],在 Visdrone

数据集上进行目标检测对比实验。为突出改进的 YOLO_v5x 算法在复杂环境下对小目标的检测性能,选取白天、黑夜和小目标密集 3 种环境进行可视化,结果如图 8。

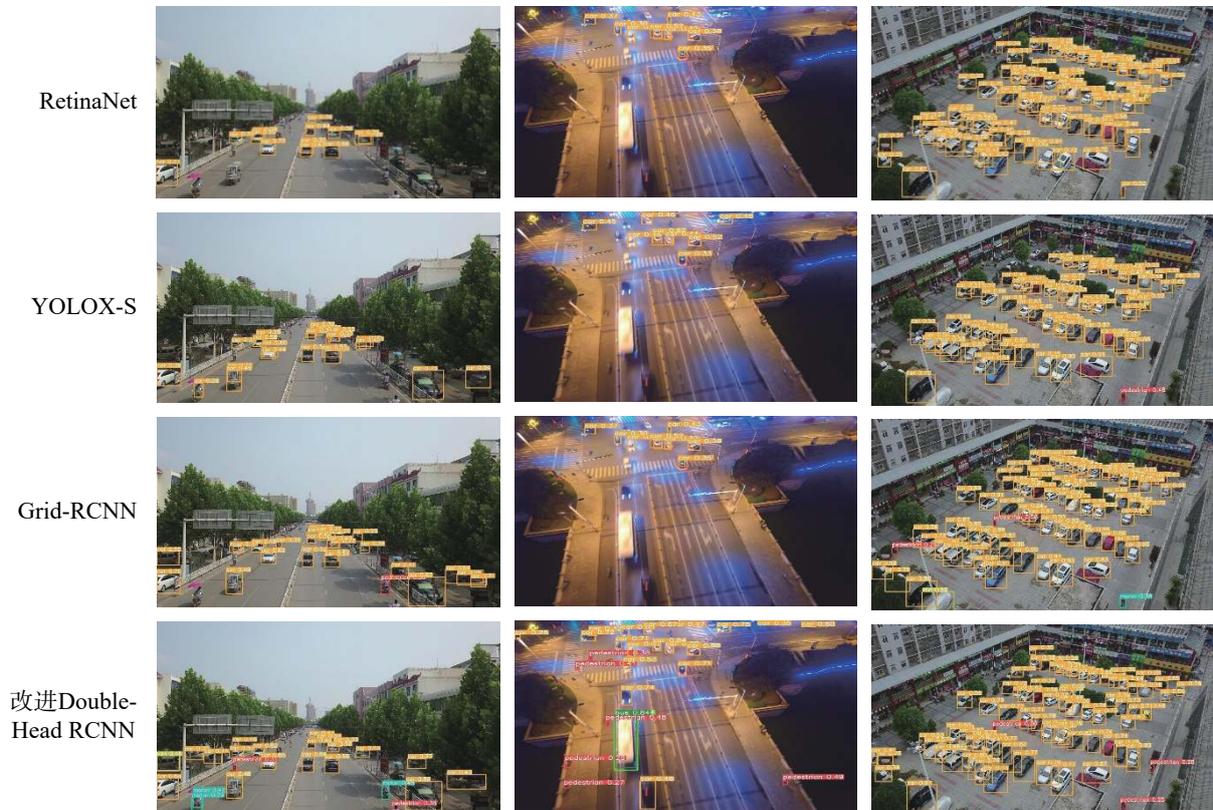




图8 不同算法的检测结果

Fig. 8 Detection results of different algorithms

由图8可看出:由于RetinaNet^[24]、YOLOX-S^[25]和Grid-RCNN^[26]对底层特征的关注度较低,导致小目标的检测性能减弱,误检率和漏检率较高;尽管改进Double-Head RCNN^[27]通过引入transformer、可变形卷积及CARAFE-FPN模块^[28],一定程度上解决小目标特征丢失的问题,但仍会出现小目标漏检误

检情况;本文算法通过减少下采样的方式防止细粒度信息丢失,并引入注意力机制和额外的小目标检测头,增强网络对不同维度位置信息的提取能力,降低了小目标的漏检率和误检率,在不同场景均获得较好的检测结果,充分体现所提改进算法的优越性。5种算法的检测精度如表2。

表2 不同检测算法的对比实验结果

Tab. 2 Comparative experimental results of different detection algorithms

| 算法 | 主干 | AP/% | AP ₅₀ /% | AP ₇₅ /% | AP _s /% | AP _M /% | AP _L /% | FPS/(帧/s) |
|------------------------------------|---------------|------|---------------------|---------------------|--------------------|--------------------|--------------------|-----------|
| RetinaNet ^[24] | ResNet-50 | 18.5 | 30.1 | 19.3 | 8.2 | 31.7 | 48.0 | 16.6 |
| YOLOX-S ^[25] | CSPDarkNet | 19.9 | 34.6 | 19.6 | 10.8 | 30.9 | 42.6 | 53.1 |
| Grid-RCNN ^[26] | ResNet-50 | 25.1 | 39.3 | 26.9 | 15.8 | 37.8 | 47.7 | 10.4 |
| 改进Double-Head RCNN ^[27] | R50-Attention | 26.4 | 44.5 | 26.9 | 18.1 | 36.1 | 48.7 | 5.9 |
| 本文算法 | Darknet53 | 33.3 | 52.3 | 32.2 | 23.9 | 42.5 | 44.9 | 8.2 |

根据表2可得:本文算法的AP, AP₅₀, AP₇₅, AP_s, AP_M, AP_L分别为33.3%, 52.3%, 32.2%, 23.9%, 42.5%, 44.9%, 检测速度(FPS)达到8.2帧/s;相较于一阶段算法RetinaNet和YOLOX-S,本文算法在AP_s上分别高出15.7%和13.1%,但检测速度略低;改进Double-HeadRCNN算法AP_s为18.1%,且AP_L为48.7%,最大,但AP_s相比本文算法低5.8%,且本文算法的小目标检测平均精度即AP_s最大,在当前主流检测小目标算法中达到先进水平。

3 结论

针对无人机航拍图像小目标特征模糊导致难以检测的问题,提出一种改进的YOLO_v5x方法,且利用Visdrone数据集进行目标检测实验。通过在YOLO_v5x算法中添加SPD模块以及小目标检测头,减少细粒度信息的丢失,能够有效提高小目标检测的效率;引入CA注意力机制、且在CIoU损失函数的基础上引入 α -IOU损失函数,能够明显提高小目标的定位精度。结果表明:本文算法在Visdrone无人机目标检测数据集上可达到33.3%的平均精度,相比于原始YOLOv5x提高了7.8%;与其他算法相

比,本文算法的平均精度和小目标检测平均精度最高,能够更好地提取小目标中有用的特征信息,提高了小目标检测定位的准确性,解决了小目标特征模糊导致难以检测的问题;同时实现了不同类别小目标的检测,可增强小目标检测算法的泛化能力。

参考文献:

- [1] 刘树东, 刘业辉, 孙叶美, 等. 基于倒置残差注意力的无人机航拍图像小目标检测[J]. 北京航空航天大学学报, 2023, 49(3):1-13.
- [2] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE, 2014:580-587.
- [3] 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.
- [4] 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, NV, USA: IEEE, 2016:779-788.

- [5] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. NY: IEEE, 2017: 7263–7271.
- [6] REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB/OL]. arXiv: 1804.02767 [cs. CV]. (2018-04-08)[2023-09-21].<https://doi.org/10.48550/arXiv.1804.02767>.
- [7] BOCHKOVSKIY A, WANG C Y, LIAO H M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. arXiv: 2004.10934[cs. CV]. (2020-04-23)[2023-09-21].<https://doi.org/10.48550/arXiv.2004.10934>.
- [8] Ultralytics: YOLOv5[EB/OL]. (2021-09-20)[2023-09-21].<https://github.com/ultralytics/YOLOv5>.
- [9] ZHU X, SU W, LU L, et al. Deformable DETR: deformable transformers for end-to-end object detection[EB/OL]. arXiv: 2004.10934 [cs. CV]. (2021-03-18) [2023-09-21].<https://doi.org/10.48550/arXiv.2010.04159>.
- [10] 田港, 张鹏, 邹金霖, 等. 基于改进YOLOv4的无人机目标检测方法[J]. 空军工程大学学报(自然科学版), 2021, 22(4):9–14.
- [11] 张智, 易华挥, 郑锦. 聚焦小目标的航拍图像目标检测算法[J]. 电子学报, 2023, 51(4):944–955.
- [12] 郝立, 张皓迪. 面向无人机作战的复杂光照环境下小目标检测方法研究[J]. 系统仿真技术, 2022, 18(2):85–89,95.
- [13] 陈映雪, 丁文锐, 李红光, 等. 基于视频帧间运动估计的无人机图像车辆检测[J]. 北京航空航天大学学报, 2020, 46(3):634–642.
- [14] BIE T, FAN K, TANG Y. UAV recognition and tracking method based on YOLOv5[C]//17th IEEE Conference on Industrial Electronics and Applications (ICIEA). Chengdu, China: IEEE, 2022:1234–1239.
- [15] WANG C Y, LIAO H Y M, WU Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Seattle, WA, USA: IEEE, 2020:1571–1580.
- [16] LIN T Y, HE K, GIRSHICK R, et al. Feature pyramid networks for object detection[C]// IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017:936–944.
- [17] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]//IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018:8759–8768.
- [18] SUNKARA R, LUO T. No more strided convolutions or pooling: a new CNN building block for low-resolution images and small objects[C]//Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Cham: Springer Nature Switzerland, 2022:443–459.
- [19] SAJJADI M S M, VEMULAPALLI R, BROWN M. Frame-recurrent video super-resolution[C]// IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 6626–6634.
- [20] HOU Q, ZHOU D Q, FENG J. Coordinate attention for efficient mobile network design[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, TN, USA: IEEE, 2021: 13708–13717.
- [21] HE J B, ERFANI S, MA X J, et al. Alpha-IoU: a family of power intersection over union losses for bounding box regression[EB/OL]. arXiv: 2110.13675 [cs. CV]. (2022-01-22)[2023-09-23]. <https://doi.org/10.48550/arXiv.2110.13675>.
- [22] ZHU P, WEN L, DU D, et al. Vision meets drones: past, present and future[EB/OL]. arXiv: 2001.06303 [cs. CV].(2021-10-04) [2023-09-21].<https://doi.org/10.48550/arXiv.2001.06303>.
- [23] LIN T Y, MARIE M, BELONGIE S, et al. Microsoft COCO: common objects in context[C]//European Conference on Computer Vision. Zurich: Springer International Publishing, 2014:740–755.
- [24] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[EB/OL]. arXiv: 1708.02002[cs. CV]. (2017-08-07) [2023-09-21].<https://doi.org/10.48550/arXiv.1708.02002>.
- [25] GE Z, LIU S, WANG F, et al. YOLOX: exceeding yolo series in 2021[EB/OL]. arXiv: 2107.08430 [cs. CV]. (2021-08-06) [2023-09-21].<https://doi.org/10.48550/arXiv.2107.08430>.
- [26] LU X, LI B, YUE Y, et al. Grid R-CNN[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE, 2019:7355–7364.
- [27] 王殿伟, 胡里晨, 房杰, 等. 基于改进Double-Head RCNN的无人机航拍图像小目标检测算法[J/OL]. 北京航空航天大学学报, (2023-01-10)[2023-09-21].<https://doi.org/10.13700/j.bh.1001-5965.2022.0591>.
- [28] 郭艺轩, 贾向东, 曹胜男. 基于3维泊松点过程理论的无人机协助的异构网络模型及其性能[J]. 安徽大学学报(自然科学版), 2023, 49(4):68–75.