DOI:10.19651/j.cnki.emt.2315250

# 基于 CS-YOLOv5s 的无人机航拍图像小目标检测

翁俊辉[1] 成 乐[2] 黄曼莉[1] 隋 皓[1] 朱宏娜[1]
(1.西南交通大学物理科学与技术学院 成都 610031; 2.西南交通大学信息科学与技术学院 成都 610031)

**摘 要**: 无人机航拍图像存在小目标分布密集且目标尺度变化大等检测难点,本文提出一种面向无人机航拍图像小目标的跨尺度目标检测模型—CS-YOLOv5s。首先,在 YOLOv5s 基础上,引入小目标检测器,提高模型对小目标的捕捉能力;进一步,将最大池化分支嵌入上下文增强模块,提取并增强骨干网络尾部的深层特征,再注入 PANet,实现深浅层特征有效融合和模型跨尺度检测能力的提升;同时采用 SPDConv 模块替换下采样卷积模块,实现无人机航拍图像中密集目标高效检测。实验表明,CS-YOLOv5s 在数据集 VisDrone2019 达到 42.0% mAP0.5,较基准模型提升 9.8%,有效增强网络模型对无人机航拍图像小目标的识别能力,为无人机目标智能识别提供支撑。

**关键词**: 无人机航拍图像;YOLO;小目标检测器;上下文增强模块;SPDConv 模块
**中图分类号**: TP391.4;TN919.8 **文献标识码**: A **国家标准学科分类代码**: 520.6040

# Small target detection for UAV aerial images based on CS-YOLOv5s

Weng Junhui[1] Cheng Le[2] Huang Manli[1] Sui Hao[1] Zhu Hongna[1]
(1. School of Physical Science and Technology, Southwest Jiaotong University, Chengdu 610031, China;
2. School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031, China)

**Abstract**: To address the challenges in detecting small targets with dense distribution and large-scale variations in UAV aerial images, a cross-scale target detection model for UAV aerial images, named CS-YOLOv5s, is proposed. Firstly, based on YOLOv5s, micro-object detector is utilized to improve the model ability for capturing small targets. Then, the max-pooling branch is embedded into the context augment model, extracting and enhancing deep feature maps at the tail of the backbone network. The PANet is injected to achieve effective fusion of deep and shallow features with enhancing the cross-scale detection capability. Furthermore, the down-sampling convolution module is replaced with the SPDConv module to achieve efficient detection of dense objects in UAV aerial images. Experiments demonstrate that CS-YOLOv5s achieves 42.0% mAP0.5 on the VisDrone2019 dataset, which is increased by 9.8% than that of the baseline model. Our model enhances the network ability to recognize small targets in UAV aerial images effectively, which provides a new way for intelligent targets recognition of UAV.

**Keywords**: UAV aerial images;YOLO;micro-object detector;context augmentation module;SPDConv module

## 0 引 言

近年来,因具有小巧灵活、搭载能力强、运动速度快等优势,无人机在精准农业[1]、电力巡检[2]、灾害救援[3]等领域得到广泛应用。然而,无人机航拍图像中的目标分布多样性高,图像中小目标作为主要的信息来源和检测对象,存在像素值占比小、易被遮挡、目标尺度变化大和分布密集等检测难点。同时人工处理无人机航拍图像信息成本高,无法实现大规模高效批量化处理,然而传统的目标检测算法及其优化方案大多是针对自然图像且中大尺度的目标,通常并不适用于无人机航拍场景[4]。因此,针对无人机航拍图像的目标检测具有重要研究意义和实际应用价值。

深度学习技术由于其强大的处理能力,被广泛用于自然图像目标检测领域[5]。基于深度学习的目标检测算法主要包含两大类,一类是基于锚框预设的网络框架,包含单阶段检测模型,如单次多框检测器(single shot multibox detector)[6]和 YOLO[7];两阶段检测模型,如区域建议卷积神经网络(region with convolutional neural networks feature,R-CNN)[8]、Faster R-CNN[9]等。另一类是基于无预设锚框的网络框架,如 CornerNet[10]、CenterNet[11]、FCOS[12]。因单阶段目标检测算法检测速度快、精度较高,诸多学者基于单阶段目标检测算法开展了无人机航拍小目标识别的研究。Hu 等[13]在路径增强网络(path aggregateon net-work,

收稿日期:2023-12-29

PANet)[14]和特征金字塔网络(feature pyramid net-work, FPN)[15]的瓶颈结构基础上,提出新型上采样特征增强结构,增强瓶颈网络特征融合能力;Wang 等[16]针对不同尺度目标,设计条形瓶颈(strip bottleneck,SPB)模块,并提出了无人机图像目标检测器 SPB-YOLO,并结合 PANet 网络结构提高检测器在无人机图像密集检测任务中的表现;赵耕彻等[17]采用引入 MobileNetv3 替换 YOLOv4 的主干网络,一方面结合深度可分离卷积奖模型轻量化,一方面增加小感受野检测头,提升网络模型检测精度;陈朋磊等[18]在 RetinaNet 基础上引入 Swin Tran-sformer,增强了网络对特征提取的能力;Zhu 等[19]将 YOLOv5 与 Transformer[20]相结合,在检测头和卷积模块进行修改,提高对无人机航拍图像的检测和识别能力;Zhao 等[21]在 TPH-YOLOv5 的基础上提出了 TPH-YOLOv5++,改进了与 Transformer 的结合方式,显著降低了参数量;Tang 等[22]采用卷积注意力模块(convolutional blo-ck attention module,CBAM)和 Involution[23]模块增强骨干网络和瓶颈网络之间的耦合关系,提升模型对无人机航拍图像的敏感度。实际中,无人机航拍图像除了小目标占比多的特点外,还存在目标跨尺度范围大且小目标分布密集的情况,因此进一步提升复杂条件下的小目标识别能力至关重要。

综上所述,本文针对无人机航拍图像中小目标分布密集且目标跨尺度范围大的检测识别难点,提出了基于上下文增强和空间转深度卷积的 YOLOv5s(context augment model and space to depth YOLOv5s, CS-YOLOv5s)模型,在保证将检测速度满足实时性的前提下,实现较高的检测准确率,本文主要贡献有:

1) 提出了一个多尺度空间上下文增强模块,通过多尺度空洞卷积与最大池化在深层特征上提取深层语义,并传递到瓶颈网络中,增强模型对不同尺度目标的泛化性和准确性。

2) 提出了一个简洁高效的小目标检测器(micro-object detector, MOD),构建深层瓶颈网络结构,实现精确的小目标识别,同时保持中大尺度的检测性能不变。

3) 采用空间转深度卷积(space to depth Conv, SPDconv)[24]模块替换下采样卷积层,降低卷积聚合过程中的信息丢失,提高网络模型对无人机航拍数据密集分布小目标的检测能力。

## 1　算法模型

### 1.1　YOLOv5s 网络框架

YOLOv5s 网络如图 1 所示,主要由 3 部分组成,分别是提取图像特征的骨干网络、融合特征的瓶颈网络以及用于目标类别和位置回归的头部检测网络。该网络架构遵循单一职责原则,有效降低模型的开发成本。该框架在训练阶段使用了众多数据增强方法,包括裁减、翻转和缩放等几何变换,以及混合增强和马赛克增强等方法。

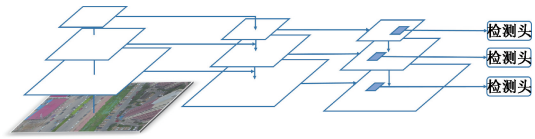


图 1　YOLOv5s 的网络框架

### 1.2　CS-YOLOv5 网络框架

本文针对无人机航拍图像小目标检测任务,在 YOLOv5s 框架的基础上提出了 3 方面改进。首先,在浅层特征上增设小目标检测器,提取小目标的空间信息,增强网络模型对小目标的捕捉能力。接着,采用改进的上下文增强模块(context augment model, CAM),提取深层语义信息并注入瓶颈网络结构,增强网络模型的跨尺度检测能力。最后,采用 SPDConv 模块替换传统的卷积模块,加强网络模型对密集分布目标的信息提取能力。

网络结构如图 2 所示,图 2 中 BotteneckCSP 模块是特征提取的模块,此模块不改变特征的大小,采用了 ResNet[25]里的残差结构,其内部 BottleNeck 结构的数量可以进行调节。此外,Upsample 为上采样操作,Concat 为拼贴操作,Detect 为目标检测头。

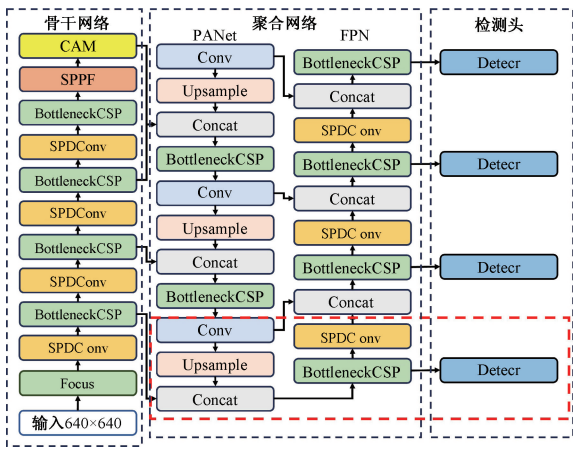


图 2　CS-YOLOv5s 的网络结构

1) 小目标检测器

如图 2 所示,红色矩形框选中的部分是 MOD,该主要检测对象是小目标。原网络结构针对中层或深层特征设置 3 个检测头,未针对小目标进行结构设计,而 MOD 结构保留第 1 个 BottleneckCSP 模块提取的特征,将丰富的小目标信息和空间信息传入 PANet 结构,构建更完整的特征传递回路,结合自下而上的 FPN 结构,在不同尺度聚合特征并传递给对应尺度的检测头,降低小目标漏检误检概率,并提高网络模型对其他尺度目标的检测能力,实现每个检测分支对特定尺度范围目标检测能力的提升,弥补原网络结构在细小目标检测方面的不足。

在锚框预设方面,选择采用 K-means 聚类树算法对数据进行聚类分析,获取对网络模型和数据集耦合性较强的锚框预设数据,计算结果为 4 个尺度共 12 个 anchor box,

分别为(5,6)(7,9)(12,10)、(10,13)(16,30)(33,23)、(47,52)(75,79)(96,102)和(116,90)(148,187)(372,324)。

2) 上下文增强结构

CAM[26]采用 3 个不同扩张率的空洞卷积提取并增强骨干网络尾部特征,再从上到下注入到 FPN 结构中,实现深层特征融合。本文在此基础上,将池化分支嵌入 CAM 结构中,对比最大池化和平均池化对检测结果的影响,实验结果如表 1 所示,最终选择最大池化方式嵌入 CAM 结构,网络结构如图 3 所示,不同扩张卷积的卷积核大小都为 3×3。

**表 1 不同池化方式对比结果**

| 池化方式 | mAP0.5 | 模型参数量/M | FPS |
|---|---|---|---|
| 最大池化 | 32.8% | 9.86 | 75.19 |
| 平均池化 | 32.6% | 9.86 | 75.19 |

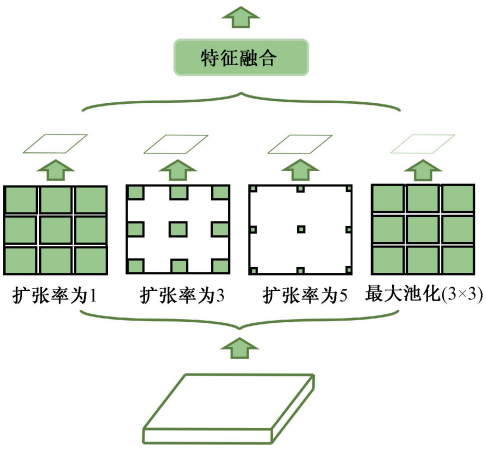


图 3 上下文增强模块网络结构

在特征融合部分比较两种方法,如图 4 所示,图 4(a)表示深度上的特征融合,图 4(b)表示像素级别的特征融合。试验结果如表 2 所示,结果表明,深度上的特征融合方式效果优于像素级别的特征融合方式,后续将都采用深度上的特征融合方式进行特征融合。同时将提取的特征注入瓶颈

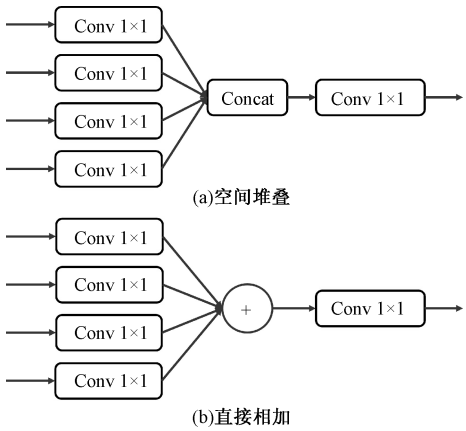


图 4 两种特征融合方式

**表 2 两种特征融合方式对比结果**

| 特征融合方式 | mAP0.5 | 模型参数量/M | FPS |
|---|---|---|---|
| 空间堆叠 | 32.8% | 9.86 | 75.19 |
| 直接相加 | 32.4% | 9.73 | 75.76 |

网络结构中,增强不同尺度特征的有效信息,提升网络模型的跨尺度检测能力。

3) 空间到深度的转换模块

SPDConv 模块是 Foucs 切片方法的推广,具体结构图如图 5 所示。实现原理是将卷积或池化的下采样过程用特征重排替代,并采用深度压缩卷积压缩特征的通道数,达到增强特征的目的。本文采用 SPDConv 模块替换下采样卷积模块,基于 SPDConv 抽样式提取特征,密集分布小目标的信息被有效保留,提升网络模型对密集分布小目标的检测能力。若特征的维度为 $W\times H\times C$,其数学表达式如下:

$$\begin{aligned} \boldsymbol{f}_{0,0} &= \boldsymbol{X}[0:W:2,0:H:2] \\ \boldsymbol{f}_{1,0} &= \boldsymbol{X}[1:W:2,0:H:2] \\ \boldsymbol{f}_{1,1} &= \boldsymbol{X}[1:W:2,1:H:2] \\ \boldsymbol{f}_{0,1} &= \boldsymbol{X}[0:W:2,1:H:2] \end{aligned} \tag{1}$$

式中:$W$、$H$ 分别代表图像的宽度和长度,$C$ 代表图像的通道数,$\boldsymbol{f}_{0,0}$、$\boldsymbol{f}_{1,0}$、$\boldsymbol{f}_{1,1}$ 和 $\boldsymbol{f}_{0,1}$ 是特征子图,抽取的尺度为 2。

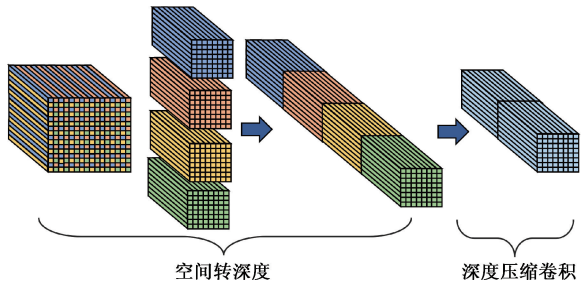


图 5 Space to Depth Conv 网络结构

### 1.3 损失函数

模型的损失函数由 3 部分组成,计算公式如下:

$$L_{all} = \lambda_1 L_{obj} + \lambda_2 L_{box} + \lambda_3 L_{cls} \tag{2}$$

式中:$\lambda_1$、$\lambda_2$ 和 $\lambda_3$ 是各项损失的比重,分别取 0.4、0.3 和 0.3,$L_{cls}$、$L_{obj}$ 是分类损失和置信度损失,采用带逻辑回归的二元交叉熵函数。$L_{box}$ 是预测框回归的损失,本文采用 CIoU (class intersect of union)[27],其计算公式为:

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(\boldsymbol{b},\boldsymbol{b}^{gt})}{c^2} + \alpha v \tag{3}$$

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

$$\alpha = \frac{v}{(1-IoU)+v} \tag{5}$$

式中:$\rho^2(\boldsymbol{b},\boldsymbol{b}^{gt})$ 表示预测框和标签框的欧几里得距离,$c$ 代表两个坐标框最小外接矩阵的对角线长度,$w$、$h$、$w^{gt}$、$h^{gt}$ 分别代表预测框的长宽和真实框的长宽,IoU (intersect of union)为预测框和真实框交并比。CIoU 在 IoU 的基础

上引入预测框和真实框的惩罚项，更精确定位目标真实位置，有效反应两者真实的重叠情况，所以本文选用 CIoU 作为边界框的损失函数。

## 2　实验与结果分析

本文采用的实验数据集为 VisDrone2019[28]，该数据集是由天津实验室 AISKYEYE 团队制作，所有航拍图像资料均由无人机采集，包含 288 个视频片段，由 261 908 帧和 10 209 幅静态图像组成。整个数据包含 10 类目标，共有 260 万标注框。图 6 是训练集所有类别目标尺寸的分布情况，横坐标代表标签框的宽度，纵坐标代表标签框的高度。小目标的定义是像素点小于等于 32×32 的目标[29]，经计算，VisDrone2019 的小目标占比为 48.01%，表明应用该数据集开展小目标检测研究是合理的。

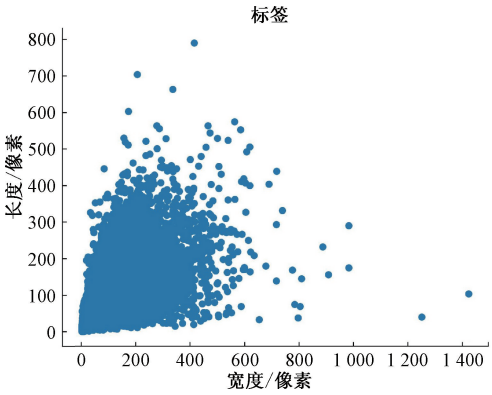


图 6　VisDrone2019 数据集标签分布图

本文采用的系统为 Ubuntu22.04，实验环境为 python3.9、pytorch2.0 和 cuda12.1。所有的实验均在 RTX4070 GPU 上进行，且设置相同的超参数。其中，训练 epochs 设置为 150 次，batch size 设置为 4，初始学习率 0.01，分辨率设置为 640 × 640，预训练模型采用 YOLOv5s.pt。

### 2.1　评价指标

评价指标采用 mAP0.5（mean average precision，IoU＝0.5）、Parameters 和 FPS（frames per second）。mAP0.5 为所有目标类别的预测框和真实框的 IoU 超过 0.5 的平均检测精度；Parameters 为模型参数量，反应模型复杂程度；FPS 为每秒检测图片数量，用于评估模型的检测速度。其中，检测相同的测试图片并计算每个模型的 FPS。

### 2.2　消融实验

本文设置消融实验来证明各个模块均可增强模型对无人机航拍目标的检测识别能力，消融实验结果如表 3 所示。实验数据表明，无论是单模块改进还是多模块结合，都有效提高无人机航拍小目标的检测和识别效果。

实验 4 结果显示，MOD 模块通过提取浅层特征中小目标的空间信息，较基准模型 mAP0.5 提升了 5.4%，有效提升了网络模型对小目标的检测能力。仅考虑单模块改进时，MOD 模块对网络模型性能的提升最大。

**表 3　在基准模型上的消融实验**

| 序号 | MOD | 改进的 CAM | SPDConv | mAP0.5/% |
|---|---|---|---|---|
| 1 | | | | 32.2 |
| 2 | | ✓ | | 32.8 |
| 3 | | | ✓ | 33.2 |
| 4 | ✓ | | | 37.6 |
| 5 | ✓ | ✓ | | 40.9 |
| 6 | ✓ | | ✓ | 40.3 |
| 7 | ✓ | ✓ | ✓ | 42.0 |

实验 2、3 结果显示，改进的 CAM 结构提升了 0.6% mAP0.5，SPDConv 结构提高了 1.1% mAP0.5。实验 5、6 表明，当两个模块分别和 MOD 进行结合后，将分别提高 8.7%和 8.1% mAP0.5，实验验证了两种结构都有效提高网络模型对小目标的检测效果，但瓶颈网络的深度影响两种结构的性能。在增设 MOD 后，SPDConv 模块有效防止浅层特征信息的丢失，为不同尺度检测头提供丰富的空间信息，提高网络模型对密集分布目标的检测能力；改进的 CAM 将增强后的深层特征注入瓶颈网络，促进深浅层特征有效融合，有效增强网络模型对多尺度目标的检测能力。

第 7 组实验 mAP0.5 较基线增加了 9.8%，实验数据表明，各个模块针对特定的需求，不管是独立存在还是共同存在，都有效提升网络模型对无人机航拍图像小目标的检测效果。

### 2.3　对比实验

为验证本文算法的有效性，将本文算法跟其他目标检测相关算法进行对比，具体结果如表 4 所示。

**表 4　不同网络模型的对比结果**

| 检测模型 | mAP0.5/% | 模型参数量/M | FPS |
|---|---|---|---|
| RetinaNet | 27.7 | — | 20.9 |
| Faster-RCNN | 37.3 | 136.7 | 13.2 |
| YOLOv4s | 26.5 | 7.02 | 90 |
| YOLOv5s | 32.2 | 7.02 | 90 |
| YOLOv5m | 34.7 | 21.80 | 24.02 |
| YOLOv8s | 30.9 | 11.1 | 86.7 |
| CDNet | 35.8 | — | 61.8 |
| HIC-YOLOv5 | 39.7 | 7.95 | 59.93 |
| TPH-YOLOv5 | 41.5 | 20.04 | 25.04 |
| CS-YOLOv5s(ours) | 42.0 | 14.42 | 51.02 |

实验结果表明，在保证检测实时性的前提下，本文提出的改进算法在 VisDrone2019 数据集上的检测精度比近年来大部分参数量相近的模型算法更有优势，mAP0.5 达到了 42.0%。对比基线模型 YOLOv5s，在模型大小扩大一倍的基础上增加了 9.8% mAP0.5，牺牲一部分检测速度

换取更高的检测精度。与 TPH-YOLOv5 算法相比,检测精度提高了 0.5%,且本文的检测速度更快;与 HIC-YOLOv5 算法相比,准确率提高了 2.3%;与 YOLOv5m 相比,在参数量约为 YOLOv5m 的 2/3 时 mAP0.5 提高了 7.3%,验证了本文算法的有效性。从总体结构出发,本文算法提出改进的 CAM 模块在骨干网络末尾提取并增强上下文信息,再注入 PANet 结构,采用 SPDConv 模块替换下采样卷积模块,同时添加小目标检测器,有效提高网络模型对无人机航拍图像小目标的检测识别效果。对比实验结果表明,本文模型在参数量相近甚至略小于的情况下有更高的检测效果,表明本文算法模型有效性和优越性。

### 2.4 可视化结果分析

为验证本文算法的实际检测效果,选取 VisDrone2019 测试集中不同环境条件下、小目标分布密集等场景作为测试对象。如图 7 所示,图 7(a)、(d)和(g)分别是夜景、存在遮掩目标和高空俯视视角的原始输入图像,其中用黄色矩形框标注出两者检测效果差异明显的部分,右侧两列为关键区域的放大对比图。

图 7(b)和(c)是夜晚行人密集场景下检测结果,基准模型未检测出 13 位行人目标,而本文模型检测出 14 位行人目标,表明本文模型显著提升密集分布小目标的检测效果。图 7(e)和(f)是存在遮掩目标场景下检测结果,基准模型未检测出被遮掩的小车,本文模型不仅检测出被遮掩的小车,还检测出摩托车上的人类,表明本文算法对遮掩目标


图 7 CS-YOLOv5s 和基准模型结果对比

的检测更有效。图 7(h)和(i)是垂直视角场景下检测结果,基准模型将厂房识别成面包车且部分目标漏检,本文模型检测出摩托和部分小车且没有将厂房误识别为面包车,表明本文算法较基准模型的泛化性更好,漏检误检率降低。

表 5 是准模型和本文模型在不同类别的 mAP0.5 对比,结果显示所有类别均有提高,表明网络模型的检测精度在所有尺度范围内较基准模型均有提高,其中行人和公交车两类显著提升,分别提升了 12%和 18.9% mAP0.5。从实际检测效果图和不同类别检测数据对比两方面,本文模型性能都高于基准模型,特别在小目标的检测能力方面得到显著提升。

**表 5 CS-YOLOv5s 模型和基准模型在不同类别的检测结果对比情况**

| 模型 | AP0.5/% | | | | | | | | | | 所有/% |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 行人 | 人 | 自行车 | 车辆 | 面包车 | 货车 | 三轮车 | 带蓬三轮车 | 公交车 | 摩托车 | |
| YOLOv5s | 37.9 | 30.8 | 10.4 | 72.2 | 34.0 | 29.4 | 19.3 | 11.4 | 39.3 | 37.2 | 32.2 |
| CS-YOLOv5s | **49.9** | **38.8** | **18.0** | **82.5** | **45.2** | **37.2** | **26.4** | **15.0** | **58.2** | **48/2** | **42.0** |

## 3 结 论

针对无人机航拍图像目标检测,本文提出针对小目标样本高效检测 CA-YOLOv5s 模型。通过添加小目标检测器,实现浅层特征中小目标的空间信息提取,并进一步增强模型对小目标的捕捉能力。同时将最大池化分支嵌入上下文增强模块,对骨干网络尾部的深层特征进行增强,充分解析深层特征的语义信息并注入到瓶颈网络结构中,提高模型对多尺度目标的检测能力。最后,基于 SPDConv 结构替换下采样卷积模块,将空间信息转变为深度信息,降低特征在卷积过程的信息丢失,提高模型对密集分布目标的检测能力。

在本文实验设定的条件下,本文提出的算法在 VisDrone2019 数据集上达到了 42.0% mAP0.5,检测速度 51.02 FPS,满足对无人机航拍图像实时检测的需求;在同等条件下对比,本文算法性能优于 RetinaNet、Faster R-CNN、YOLOv4s、YOLOv5s、YOLOv5m、YOLOv8s、CDNet、HIC-YOLOv5 和 TPH-YOLOv5,因此,在面向无人机航拍图像小目标时,本文算法模型具有较高的鲁棒性和检测效果。

## 参考文献

[1] 陈鹏飞.无人机在农业中的应用现状与展望[J].浙江大学学报(农业与生命科学版),2018,44(4):399-406.

[2] 叶翔,孙嘉兴,甘永叶,等.改进 YOLOv3 模型在无人机巡检输电线路部件缺陷检测中的应用研究[J].电测与仪表,2023,60(5):85-91.

[3] BOŽIĆ-ŠTULIĆ D, MARUŠIĆ Ž, GOTOVAC S. Deep learning approach in aerial imagery for supporting land search and rescue missions [J]. International Journal of Computer Vision, 2019, 127(9): 1256-1278.

[4] 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.

[5] ZOU Z, CHEN K, SHI Z, et al. Object detection in

20 years: A survey[J]. Proceedings of the IEEE, 2023, 113(3): 257-276.
[6] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector [C]. European Conference on Computer Vision, 2016: 21-37.
[7] TERVEN J R, CORDOVA-ESPARZA D. A comprehensive review of YOLO: From YOLOv1 and beyond[J]. ArXiv Preprint, 2023, ArXiv: 2304.00501.
[8] 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, 2014: 580-587.
[9] 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.
[10] LAW H, DENG J. Cornernet: Detecting objects as paired keypoints[C]. Proceedings of the European Conference on Computer Vision, 2020, 128(3): 642-656.
[11] DUAN K, BAI S, XIE L, et al. Centernet: Keypoint triplets for object detection[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 6569-6578.
[12] TIAN Z, SHEN C, CHEN H, et al. Fcos: Fully convolutional one-stage object detection [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 9627-9636.
[13] LIN H, ZHOU J, GAN Y, et al. Novel up-scale feature aggregation for object detection in aerial images[J]. Neurocomputing, 2020, 411: 364-374.
[14] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.
[15] LIN T Y, DOLLR 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.
[16] WANG X, LI W, GUO W, et al. SPB-YOLO: An efficient real-time detector for unmanned aerial vehicle images[C]. 2021 International Conference on Artificial Intelligence in Information and Communication, 2021: 099-104.
[17] 赵耘彻,张文胜,刘世伟.基于改进 YOLOv4 的无人机航拍目标检测算法[J].电子测量技术,2023,46(8):169-175.
[18] 陈朋磊,王江涛,张志伟,等.基于特征聚合与多元协同特征交互的航拍图像小目标检测[J].电子测量与仪器学报,2023,37(10):183-192.
[19] ZHU X K, LYU S C, WANG X, et al. TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios[C]. 18th IEEE/CVF International Conference on Computer Vision, 2021: 2778-2788.
[20] PARMAR N, VASWANI A, USZKOREIT J, et al. Image transformer[C]. International Conference on Machine Learning, 2018: 4055-4064.
[21] ZHAO Q, LIU B, LYU S, et al. TPH-YOLOv5++: Boosting object detection on Drone-Captured Scenarios with cross-layer asymmetric transformer[J]. Remote Sensing, 2023, 15(6): 1687-1687.
[22] TANG S, FANG Y, ZHANG S. HIC-YOLOv5: Improved YOLOv5 for small object detection[J]. ArXiv Preprint, 2023, ArXiv: 2309.16393.
[23] WANG A, PENG T, CAO H, et al. TIA-YOLOv5: An improved YOLOv5 network for real-time detection of crop and weed in the field[J]. Frontiers in Plant Science, 2022, 13: 1091655.
[24] 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, 2022: 443-459.
[25] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
[26] XIAO J, ZHAO T, YAO Y, et al. Context augmentation and feature refinement network for tiny object detection [C]. International Conference on Learning Representations, 2021: 1-11.
[27] 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, 34(7): 12993-13000.
[28] ZHU P, WEN L, DU D, et al. Detection and tracking meet drones challenge [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(11): 7380-7399.
[29] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft coco: Common objects in context [C]. European Conference on Computer Vision, 2014: 740-755.

## 作者简介

**翁俊辉**,硕士研究生,主要研究方向为航拍目标检测、信号识别。
E-mail: wjh@my.swjtu.edu.cn

**成乐**,博士研究生,主要研究方向为深度学习,信号处理。
E-mail: cancokecoke@gmail.com

**黄曼莉**,硕士研究生,主要研究方向为通信辐射源个体识别,信号识别。
E-mail: cerehhh@163.com

**隋皓**,博士研究生,主要研究方向为光纤非线性效应、深度学习。
E-mail: 516856048@qq.com

**朱宏娜**(通信作者),教授,主要研究方向为机器学习及智能信号处理,光纤通信。
E-mail: hnzhu@home.swjtu.edu.cn