

融合多尺度特征的轻量级航拍目标检测算法^{*}

党宏社 高宛蓉 孟饶晨 张选德

(陕西科技大学电气与控制工程学院 西安 710021)

摘要: 针对无人机航拍图像中目标尺寸变化大、相互遮挡等原因导致的目标误检漏检问题,以 YOLOv8s 为基础,提出了一种融合多尺度特征的轻量级目标检测算法。在主干网络中利用轻量级多尺度卷积 EMSC 替换 C2f 模块中的 Bottleneck,增强多尺度特征表达能力;颈部网络中引入轻量级上采样算子 Dysample,捕捉图像细微特征;优化 Task Aligned Assigner 超参数,解决训练过程中样本不平衡问题。最后,设计系统可视化界面,借助可视化界面完成对航拍目标的检测。在数据集 VisDrone2019 上的仿真表明,该算法精准率和召回率较基准算法分别提升了 2.4% 和 3.3%, mAP_{0.5} 提升了 3.5%,有效提高了航拍目标检测效果。在 UAVDT 数据集上进行模型泛化性实验验证,效果优于其他经典算法。

关键词: 无人机;小目标检测;多尺度卷积;动态上采样;正样本分配

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

Target detect algorithm of lightweight in aerial images
based on multi-scale feature fusion

Dang Hongshe Gao Wanrong Meng Raochen Zhang Xuande

(College of Electrical and Control Engineering, Shanxi University of Science and Technology, Xi'an 710021, China)

Abstract: Aiming at the problem of target misdetection and missing detection caused by large size changes and mutual occlusion in UAV aerial images, a lightweight target detection algorithm based on YOLOv8s is proposed by integrating multi-scale features. In the backbone network, lightweight multi-scale convolutional EMSC is used to replace Bottleneck in C2f module, which enhances the expression ability of multi-scale features. The lightweight upsampling operator Dysample is introduced into the neck network to capture the fine features of the image. Task Aligned Assigner hyperparameters are optimized to solve the problem of sample imbalance during training. Finally, the system of visual interface is designed, and the object of aerial photography is detected by visual interface. The simulation on the data set VisDrone2019 shows that the accuracy and recall rate of the algorithm are improved by 2.4% and 3.3% respectively compared with the benchmark algorithm, and mAP_{0.5} is improved by 3.5%, effectively improving the effect of aerial photography target detection. The model generalization experiment is carried out on UAVDT data set, and the effect is better than other classical algorithms.

Keywords: UAV; small target detection; multi-scale convolution; dysample; Task Aligned Assigner

0 引言

无人机在军事、民用、商业等领域的应用日益广泛^[1]。其高速移动和拍摄角度的变化,目标呈现多尺度、相互遮挡、背景复杂等特点,使得目标检测具有挑战性。

目前,在无人机视角下的目标检测研究中,深度学习方法已成为主流^[2]。这些方法主要分为两大类:一类是基于区域提议的两阶段检测算法,如 Fast R-CNN 和 Faster

R-CNN 等;另一类是回归型的单阶段检测算法,如 SSD 系列以及 YOLO 系列等。尽管两阶段算法在准确性上表现优异,但其检测速度较慢,不适用于实时响应的无人机航拍场景。相比之下,单阶段算法以其快速的检测速度和较高的精度,更适合这一应用需求。因此,许多研究者选择基于单阶段算法进行无人机航拍目标检测,并展开了一系列研究。雷帮军等^[3]通过设计跨层特征融合模块,有效整合了小目标在不同层次上的语义信息,提升了小目标的检测能

力,但此方法涉及多层次数据交互和计算,增加了整体网络的计算复杂度。胡峻峰等^[4]在优化 YOLOv8 时,通过部分卷积层取代 3×3 卷积层以减少网络冗余,并采用均衡焦点损失作为类别损失函数来平衡网络训练中尾部类别的梯度权重,从而增强了模型的类别检测能力,然而部分卷积层在特征提取过程中可能会忽略一些关键的局部信息,限制了模型对细节的捕捉能力。陈佳慧等^[5]设计了等级制残差结构的多尺度特征提取模块 C3Res2Block,旨在提高网络的多尺度特征表征能力并抑制冲突,此外采用软化非极大值抑制算法优化候选框的置信度,提高了模型对密集小目标的检测精度。Sui 等^[6]将双向加权特征金字塔网络(bidirectional feature pyramid network, BiFPN)与 YOLOv8 主干网络结合,以增强多层次特征融合,从而获得更多的语义信息,同时引入了结合自注意力的检测头,以改善检测性能,然而此算法的训练收敛时间较长,不能满足快速检测的要求。颜豪男等^[7]采用改进的拉普拉斯算子与自适应图像增强网络(illumination adaptive transformer, IAT)进行融合,以增强目标边缘特征,并使用双分支结构对正常图像和增强图像进行并行学习,构建了双分支特征级融合自适应增强网络,能够自动适应光照并增强图像信息,但是过度依赖拉普拉斯算子来增强边缘特征可能导致模型对局部特征过于敏感,从而忽略全局上下文信息。殷波^[8]在 YOLOv8 中引入注意尺度序列融合机制(attentional scale sequence fusion, ASF),旨在融合不同尺度的特征图以提取更丰富的特征信息,然而 ASF 的使用增加了模型的计算量,不适应实时检测需求。潘玮等^[9]引入了大型可分离卷积注意力以改进空间金字塔池化(spatial pyramid pooling and feature fusion, SPPF)模块,增强不同层级特征之间的信息交融,从而丰富感受野信息,但这种方法增加了网络的层数,导致网络计算量的增加。Wang 等^[10]融合了 BiFormer 注意力机制优化骨干网络,并引入了焦点特征处理模块(focal fasterNet block, FFNB),增加了两个新的检测尺度,使得检测能覆盖五个不同尺度,降低了漏检率,然而对于小型物体的检测精度仍然不高。陈朋磊等^[11]设计了一种高效的小目标特征聚合网络(small target feature aggregation network, SFANet),旨在高效整合浅层特征图中小目标的细节信息,并且引入了多元协同特征交互模板(multiple collaborative feature interaction module, MCFIM),增强网络在多尺度目标检测方面的性能,然而此设计导致了网络计算量的显著增加。

综上所述,尽管现有的航拍目标检测算法在一定程度上提高了检测性能,但仍存在一些不足:1)样本分布不均衡,影响模型对少数类别目标的识别能力;2)目标存在多尺度变化,难以实现准确检测;3)特征提取过程中,容易丢失细节信息;4)小目标定位困难,容易导致误检和漏检。

针对上述存在的难点,本文以 YOLOv8s^[12]为基础,提出一种融合多尺度特征的轻量级 YOLOv8s 算法,主要贡

献如下:

1)设计轻量级多尺度卷积(efficient multi-scale conv, EMSC),替代主干网络 C2f 模块中的 Bottleneck。这一设计不仅减少了网络参数和计算量,还能有效捕获不同尺度下的特征信息,提升多尺度特征的表达能力。

2)引入轻量级上采样算子(upsampling by dynamic sampling, Dysample),替换颈部网络中的上采样模块。该算子根据输入特征图动态调整偏移量,有助于保留图像的细节信息。

3)优化正样本分配策略的 Task Aligned Assigner 超参数,通过合理调整参数以解决正负样本分布不均衡的问题。

1 融合多尺度特征的轻量级 YOLOv8s 算法

1.1 算法框架

本文算法整体流程如图 1 所示,航拍图像经过预处理被调整为 640×640 的尺寸,然后输入到算法网络中进行处理,得到 3 个尺寸的特征张量,接着对预测框进行调整并执行非极大值抑制,最终输出检测结果图。

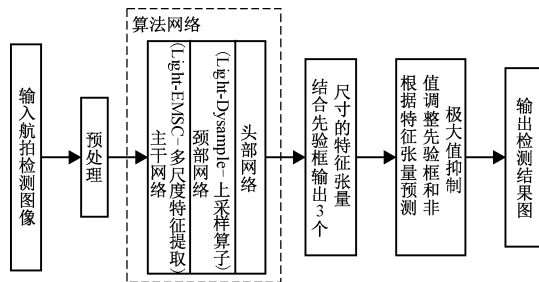


图1 算法整体流程

Fig. 1 Overall flow of algorithm

算法网络如图 2 所示,以 YOLOv8s 为基准算法。主要由主干网络、颈部网络和头部网络组成。图中虚线方框所框的为本文的改进模块,图像输入到网络后进入主干网络中提取特征,主干网络中设计轻量级多尺度卷积 EMSC 替换 C2f 模块中的 Bottleneck,通过采用不同大小的卷积核,捕获不同尺度下的特征信息,并在不同特征之间进行信息融合,以便能同时处理全局和局部特征。然后特征图进入颈部网络进行特征增强,在颈部网络中引入 Dysample 上采样算子,捕捉图像中的细微特征,有效提高上采样效率。最后优化算法损失函数中正样本匹配 Task Aligned Assigner 的超参数,解决网络正负样本分配不均衡的问题,使模型更好地学习目标之间的对应关系,从而提高整体模型的训练效果和性能。

1.2 轻量级多尺度卷积 EMSC

YOLOv8s 主干网络中大量采用常规卷积,导致处理多尺度物体时缺乏足够的上下文信息,限制复杂场景下的特征表征能力,同时还存在参数量和计算量过大的问题。因此引入一种轻量级多尺度卷积 EMSC 替换主干网络 C2f

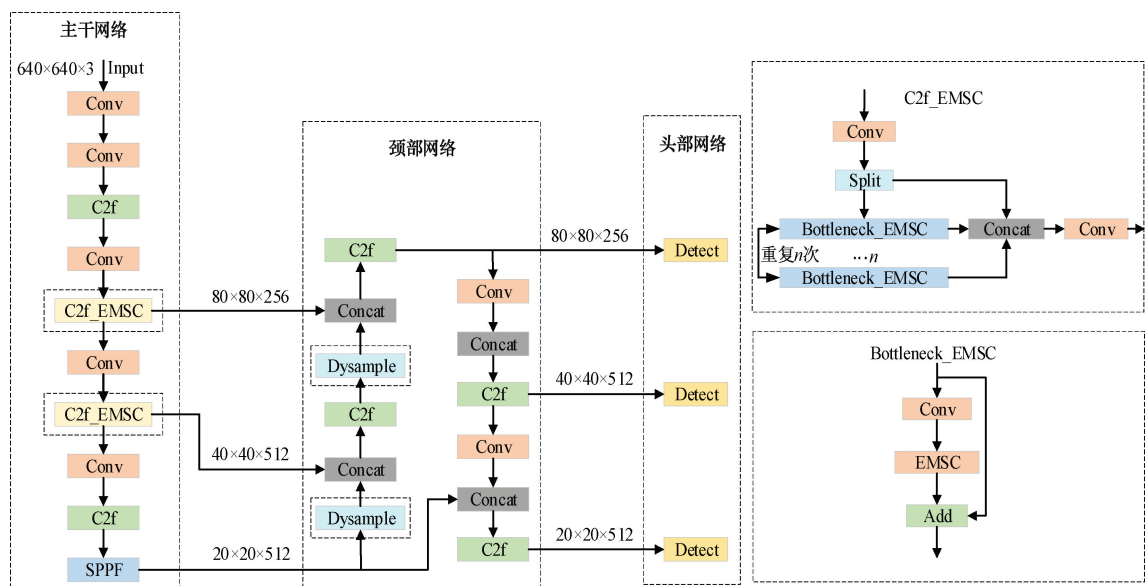


图 2 算法整体框架

Fig. 2 Overall framework of algorithm

中 Bottleneck 里的第一个常规卷积,轻量级多尺度卷积 EMSC 采用了 Mobilenetv1、Mobilenet2^[13] 和 Ghostconv^[14] 的分组思想,减少了网络的参数量和计算量。

EMSC 结构如图 3 所示。首先,输入张量后获取通道数,通过 split 操作将输入张量分为两部分,其中一部分直接连接到下一层,而另外一部分执行 rearrange 操作,对张量进行维度重新排列操作^[15],然后对每个组的特征图进行相应的卷积操作,可以并行地提取多尺度特征。将卷积后的特征图再通过 rearrange 操作重新排列成原来的维度,然后将处理后的结果和另外一部分拼接到一起,得到新的特征图,最后再利用一个 1×1 的逐点卷积去交换它的通道信息,进行特征融合,从而减少网络参数数量并增强特征表达能力,以获得最终的输出特征图。rearrange 操作的目的是将输入张量的维度进行重新排列,以适应多尺度卷积的操作。在这里,rearrange 操作将通道维度与尺度维度进行了转置,确保了网络能够充分利用不同尺度下的特征信息,提升了特征提取的效率和准确性。

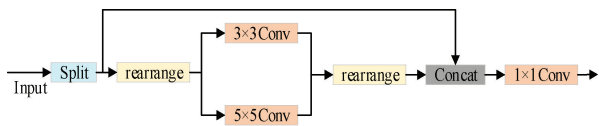


图 3 EMSC 结构

Fig. 3 EMSC structure

1.3 轻量级上采样算子 Dysample

目前常用的上采样方法包括插值上采样和转置卷积上采样。YOLOv8s 使用最近邻插值上采样,其是简单复制最近邻像素的值;而转置卷积上采样是使用转置卷积层来对特征图进行上采样,通过学习可训练的卷积核来实现尺寸的放大。它们都利用固定的采样方式,无法根据特征

图的内容自适应地选择合适的采样点,造成语义信息丢失的问题。因此引入轻量级上采样算子 Dysample^[16]到颈部网络,捕捉图像中的细微特征,更好地保留图像的细节信息。

轻量级动态上采样算子 Dysample 采用了点采样的思想,通过引入偏移量来实现上采样,根据输入特征图的内容动态地调整偏移量,从而使得上采样的结果更加具有位置感知性,更好地保留和重建特征图中的细节信息,提高模型的性能。Dysample 主要包括采样点生成器和网格采样模块。输入形状大小为 $C \times H_1 \times W_1$ 的特征映射 $\mathbf{X}$ 和一个大小为 $2 \times H_2 \times W_2$ 的采样集 $\mathbf{S}$,其中 C 表示通道数, H 表示高, W 表示宽度,采样集中第一维的 2 表示 x 和 y 坐标,使用网格采样函数对输入 $\mathbf{X}$ 进行双线性插值采样,得到大小为 $C \times H_2 \times W_2$ 的上采样特征图 $\mathbf{X}'$ 。这一上采样过程如式(1)所示。

$$\mathbf{X}' = \text{grid_sample}(\mathbf{X}, \mathbf{S}) \quad (1)$$

其中, $\mathbf{X}$ 为输入特征, $\mathbf{X}'$ 为上采样特征, $\mathbf{S}$ 为采样集。采样点生成器生成采样集 $\mathbf{S}$, 给定上采样比例因子 s 和形状大小为 $C \times H \times W$ 的特征图 $\mathbf{X}$, 使用输入和输出通道数为 C 和 $2s^2$ 的线性层来生成大小为 $2s^2 \times H \times W$ 的偏移量 $\mathbf{O}$, 然后通过像素重组将其重塑为 $2 \times sH \times sW$ 。采样集 $\mathbf{S}$ 就是偏移量 $\mathbf{O}$ 与原始网格采样 $\mathbf{G}$ 之和, 运算过程如式(2)所示。

$$\mathbf{S} = \mathbf{G} + \mathbf{O} \quad (2)$$

动态范围操作过程如式(3)所示。

$$\mathbf{O} = 0.5 \text{sigmoid}(\text{linear}_1(\mathbf{X})) \times \text{linear}_2(\mathbf{X}) \quad (3)$$

接收输入特征图 $\mathbf{X}$, 并经过卷积操作将输入映射为输出, 生成的输出通过 sigmoid 函数进行激活, 然后乘以 0.5, 用于缩放范围参数的值。这个范围参数通过与偏移量相

乘来调整目标的尺度,并进行归一化操作,然后利用像素重组函数对坐标进行重新排序,最后与初始位置相加来获得最终的位置坐标。为了增加偏移的灵活性,通过线性投影输入特征进一步生成逐点动态范围因子,动态范围因子用于根据输入数据的特性动态地调整偏移量,以适应不同的特征分布和数据情况。Dysample 动态上采样相对于传统上采样能更好地保留图像的细节信息,避免因简单复制像素而导致的模糊或锯齿效应。Dysample 结构如图 4 所示。

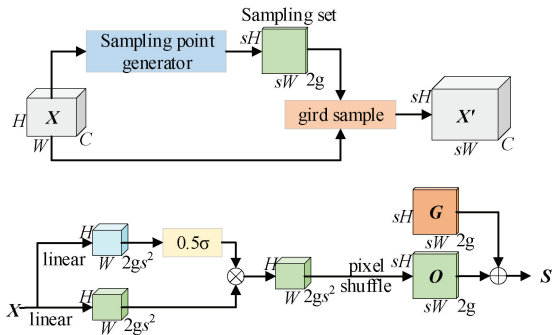


图4 Dysample 结构

Fig. 4 Dysample structure

1.4 Task Aligned Assgner 超参数

YOLOv8s 在损失函数计算方面,使用了 Task Aligned Assgner 正样本分配策略,其策略为根据分类与回归分数加权的分数选择正样本。首先,计算真实框和预测框的匹配程度,计算过程如式(4)所示:

$$align_metric = s^\alpha \times u^\beta \quad (4)$$

其中, s 是预测类别分值, u 是预测框和真实框的 CIOU 值, α 和 β 为权重超参数。其次,对于每个真实框,对 $align_metric$ 的匹配程度进行排序,选取 $topK$ 个预测框作为正样本。最后,对于一个预测框与多个真实框匹配的情况进行处理时,保留 CIOU 最大的真实框。

在密集小目标检测中,过高的 $topK$ 值可能会增加低质量预测框的数量,特别是由于小目标在特征图上占据的空间较小。在一个目标遮挡另一个的情况下,两个目标可能会竞争同一个预测框,导致其中一个目标缺乏高质量的预测框来进行准确预测。因此,适当降低 $topK$ 的值,调整匹配时考虑的预测框数量,确保每个真实目标框都能够被合适的预测框所匹配。此外,调整 α 和 β 的值可以控制正负样本之间的匹配关系,避免模型训练过程中出现样本不平衡问题。

2 实验结果与分析

2.1 实验设计

1)数据集介绍

本文使用 VisDrone2019^[17]数据集,其包含 10 类标签。共 10 209 张静态图像:6 471 张用于训练、548 张用于验证、

1 610 张用于测试。这些图像涵盖了各种城市和乡村场景、不同背景和天气条件,为研究提供了丰富的样本多样性。

为了验证本文算法的有效性,还使用了 UAVDT 数据集^[18],该数据集包含 50 个视频片段。本文按照通用标准,将前 30 个视频片段中的 22 919 张图像用于训练,后 20 个视频片段中的 17 457 张图像用于测试。

2)实验配置与评价指标

实验配置如表 1 所示。为了全面客观地评价算法性能,选取平均精度均值(mean average precision, mAP)、精准率 P (Precision)、召回率 R (Recall)、模型参数量 $Params$ 、模型计算量 $GFLOPs$ 、检测速度 (frame per second, FPS)作为模型性能的评价指标。其中精准率 P 、召回率 R 、平均精度均值 mAP 、计算速度 FPS 的计算公式如下:

$$P = \frac{TP}{TP + FP} \quad (5)$$

$$R = \frac{TP}{TP + FN} \quad (6)$$

$$mAP = \frac{1}{n} \sum_{i=1}^n \int_0^1 P(R) dR \quad (7)$$

$$FPS = \frac{frameNum}{elapsedTime} \quad (8)$$

式(5)和(6)中的 TP 为正样本被预测为正样本的数量, FP 为负样本被预测为正样本的数量, FN 为正样本被预测为负样本的数量。式(7)中的 mAP 衡量模型在所有类别上的性能。式(8)中的 $frameNum$ 为处理的总帧数, $elapsedTime$ 为处理所用的总时间。

表1 实验软硬件环境配置

Table 1 Experimental software and hardware environment configuration

实验配置	版本/型号
CPU	12 vCPU Intel(R) Xeon(R) Silver 4214R CPU@2.40 GHz
GPU	RTX 3080 Ti
操作系统	Linux 操作系统
深度学习框架	Pytorch1.10
显存	12 G
Cuda	11.3

3)模型训练

训练样本采用 640×640 , 进行 150 轮训练, Batchsize 设置为 8, workers 设置为 2, 初始学习率设置为 0.01。如果连续 50 轮内模型性能未提升, 将启动早停机制结束训练。在最后 10 轮中停用数据增强策略, 选择 SGD 优化器, 并利用梯度下降法动态调整学习率。

2.2 结果分析

1)消融实验

为了确定多尺度卷积 EMSC 替换 C2f 模块的数量,以 YOLOv8s 为基础,分别在主干网络中替换 1 个 C2f,2 个 C2f 和 3 个 C2f,进行了对比实验,由表 2 可知,替换越多的 C2f 精度逐渐降低,但速度越快。而相较于替换 3 个 C2f,替换 2 个 C2f 在速度没有下降多少的情况下,精度提升了很多。因此,选择替换主干网络中 2 个 C2f 模块,以在速度和精度之间取得较好平衡。

表 2 替换不同数量 EMSC 对比实验

Table 2 Comparative experiments to replace different number of EMSC				
模型	$mAP_{0.5}$ / %	$Params$ / M	$GFLOPs$	FPS
YOLOv8s	40.3	11.12	28.5	113.4
YOLOv8s+1EMSC	41.4	11.06	27.7	112.9
YOLOv8s+2EMSC	41.3	10.82	26.9	115.0
YOLOv8s+3EMSC	40.5	10.35	26.7	115.2

由表 3 的超参数对比实验结果可知,控制 α 和 β 固定不变,通过调整 $topK$,可以看出当 $topK=5$ 时,各项指标均有所提升;接着固定 $topK$ 和 β ,调整 α ,发现当 $\alpha=1.0$ 时,除了召回率略有降低外,其他指标表现最佳;最后固定 $topK$ 和 α ,调整 β ,发现当 $\beta=6.0$ 时,除了召回率稍有下降外,其他指标表现最佳。因此,选择将 $topK$ 设定为 5、 α 设定为 1.0、 β 设定为 6.0 时,模型效果最佳。

表 3 超参数对比实验

Table 3 Comparative experiments of hyperparameters				
TAL Para	P / %	R / %	$mAP_{0.5}$ / %	$mAP_{0.5,0.95}$ / %
$topK=10, \alpha=0.5, \beta=6.0$	52.1	38.9	40.3	24.0
$topK=13, \alpha=0.5, \beta=6.0$	51.6	38.4	40.0	23.7
$topK=7, \alpha=0.5, \beta=6.0$	51.4	39.8	40.9	24.1
$topK=5, \alpha=0.5, \beta=6.0$	52.9	40.0	41.3	24.3
$topK=5, \alpha=0.75, \beta=6.0$	52.2	39.6	40.9	23.9
$topK=5, \alpha=1.0, \beta=6.0$	53.8	39.8	41.8	24.4
$topK=5, \alpha=1.0, \beta=5.0$	52.6	39.9	41.2	24.2
$topK=5, \alpha=1.0, \beta=4.0$	51.8	39.6	41.0	24.0

由表 4 的消融实验结果可知,引入轻量级多尺度卷积 EMSC 替换主干网络中 2 个 C2f 模块中的 Bottleneck 后精准率提升了 0.4%,召回率提升了 1.7%, $mAP_{0.5}$ 提升了 1.0%,参数量下降了 0.3 M, FPS 提升了 1.6。该效果源于融合了主干网络的特征和多尺度卷积 EMSC 的优势,使得网络更有效地捕获特征之间的复杂关系,从而提升模型的性能和泛化能力。

表 4 消融实验
Table 4 Ablation experiments

改进策略				
2EMSC		✓	✓	✓
Dysample			✓	✓
超参数				✓
P / %	52.10	52.50	53.20	54.50
R / %	38.90	40.60	41.50	42.20
$mAP_{0.5}$ / %	40.30	41.30	42.50	43.80
$Params$ / M	11.12	10.82	10.85	10.85
$GFLOPs$	28.50	26.90	26.90	26.90
FPS	113.4	115.00	111.70	111.50

使用 Dysample 轻量级上采样算子后, $mAP_{0.5}$ 达到 42.5%,提升了 1.2%。该效果源于 Dysample 动态上采样算子有效改善了上采样后图像的视觉效果,确保了细节信息的完整性,为后续的图像处理任务提供清晰、准确的输入数据。

优化正样本匹配 Task Aligned Assigner 的超参数后,精准率为 54.5%,召回率为 42.2%, $mAP_{0.5}$ 为 43.8%,分别提升了 1.3%、0.7%、1.3%。该效果源于适当调整 $topK$ 值,确保每个真实目标框都能够被合适的预测框所匹配。而调整 α 和 β 超参数可以控制正负样本之间的匹配关系。

相比于 YOLOv8s 使用 3 种改进策略后,模型的精准率提升了 2.4%,召回率提升了 3.3%。在保持检测速度几乎不变时, $mAP_{0.5}$ 也提升了 3.5%。

2)模型对比实验

改进效果对比如图 5 所示,改进后的算法与基准算法的收敛速度相近,改进后的算法在性能方面表现出色。经过 150 轮训练后,算法的各项性能指标趋于稳定,达到了收敛状态。

由表 5 的不同检测算法性能对比可知,本文所提算法与 Faster-RCNN、RetinaNet、SSD 算法相比,在拥有更低参数量和更快检测速度的同时, $mAP_{0.5}$ 分别提升了 10%、21.9%、17.3%。虽然 YOLOv7-tiny、YOLOv5s 算法参数量较少,但检测性能各项指标不及所提算法。而相比于 YOLOv8s 基准算法,所提算法在保持检测速度基本持平

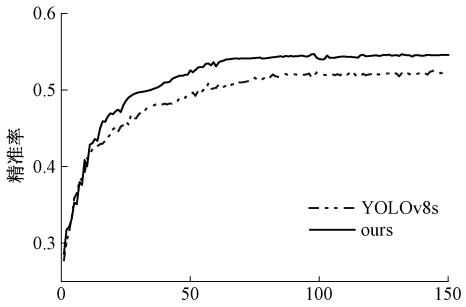
表 5 不同算法性能对比

Table 5 Performance comparison of different algorithms

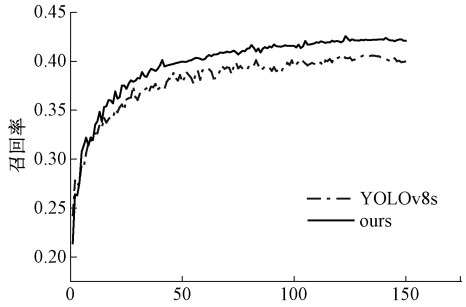
网络结构	$Params /$ M	$mAP_{0.5} /$ %	$GFLOPs$	FPS
Faster-RCNN	41.19	33.8	206.73	15.2
SSD	24.50	26.5	87.90	40.3
RetinaNet	19.80	21.9	93.70	23.5
YOLOv5s	7.20	34.2	16.50	105.0
YOLOv7-tiny	6.03	36.2	13.10	90.5
YOLOv8s	11.12	40.3	28.50	113.4
Ours	10.85	43.8	26.90	111.5

3) 实验效果分析

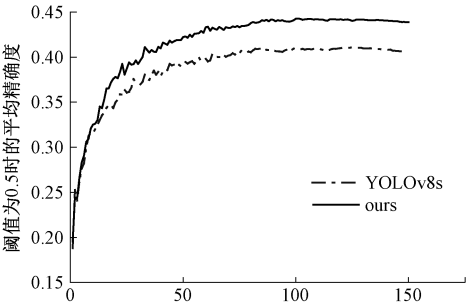
为了验证本文所提算法在实际场景中的效果,选取 VisDrone2019 测试集中难度较大的图像进行检测。检测效果对比如图 6 所示,图 6(a)为基准算法 YOLOv8s 检测效果图,图 6(b)为本文算法的检测效果图。部分明显变化在图中标出,方框表示漏检,椭圆框表示误检。在第 1 组夜间低光照背景复杂场景下,YOLOv8s 将草丛错误识别为车,将卡车误判为公共汽车,而本文算法避免了这些误检;在第 2 组多尺度变化场景下,基准算法由于目标大小变化,对三轮车、面包车和行人出现了漏检,而本文算法显著降低了漏检的风险。在第 3 组和第 4 组部分遮挡和几乎完全遮挡场景下,YOLOv8s 对于小部分遮挡目标能够正确检测,但对于大部分和几乎完全遮挡的目标容易漏检。



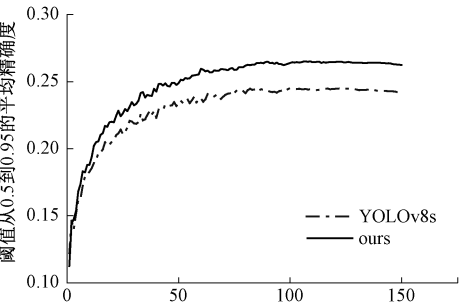
(a) 精确率对比
(a) Precision comparison



(b) 召回率对比
(b) Recall comparison



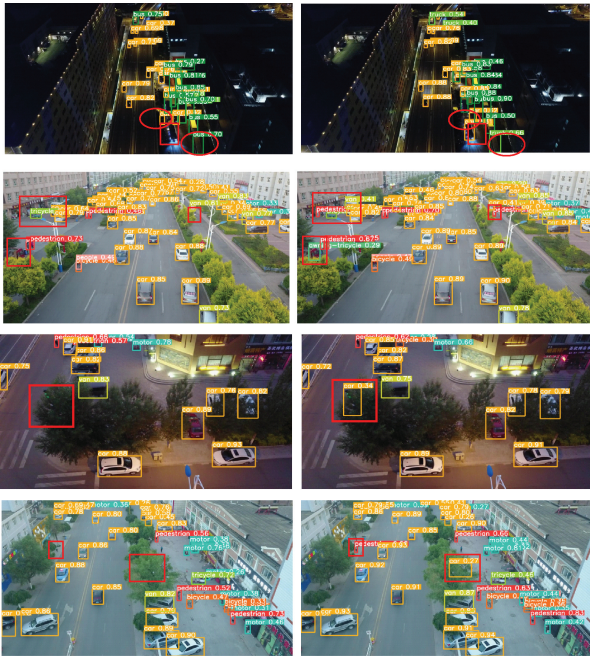
(c) 阈值为0.5时的平均精确度对比
(c) Average accuracy comparison when the threshold is 0.5



(d) 阈值从0.5到0.95的平均精确度对比
(d) Average accuracy comparison of thresholds from 0.5 to 0.95

图 5 改进效果对比

Fig. 5 Comparison of improvement effects



(a) 基准算法
(a) Reference algorithm
(b) 本文算法
(b) Textual algorithm

图 6 不同场景下检测效果对比

Fig. 6 Comparison of detection effects in different scenarios

的情况下,参数量和计算量均有所减少, $mAP_{0.5}$ 提升了 3.5%。

相比之下,本文算法能够更精确地检测出这些目标。因此,与基准算法 YOLOv8s 相比,本文算法在面对环境变化时表现更为稳定,具有较强的抗干扰性,有效降低了误检和漏检的可能性。

4)模型泛化性实验验证

由表 6 的不同算法在 UAVDT 数据集上的性能对比可知,本文提出算法相较于基准算法 YOLOv8s,在保持检测速度几乎不变、且参数量减少的情况下,mAP_{0.5} 提升了 3.7%。同时与 YOLOv7-tiny、YOLOv5s 相比,本文算法在 mAP_{0.5} 上分别提高了 7.7%、9.6%,FPS 分别提升了 21、6.8。因此,本文算法在其他航拍数据集上展现出了通用性和良好的泛化能力。

表 6 不同算法在 UAVDT 数据集上的性能对比

Table 6 Performance comparison of different algorithms on UAVDT dataset

网络结构	Params/M	mAP _{0.5} /%	FPS
YOLOv5s	7.30	21.3	103.4
YOLOv7-tiny	6.05	23.2	89.2
YOLOv8s	11.13	27.2	111.9
Ours	10.86	30.9	110.2

5)系统可视化

为了更加直观地展示航拍目标检测结果且更好地应用到实际中,根据本文所提算法开发了一款航拍目标检测性能验证系统,检测界面如图 7 所示。



图 7 检测界面

Fig. 7 Interface of detection

系统界面包括权重配置文件的选择、打开图片、保存路径和开始检测等功能。配置文件可选择 pt 或 yaml 文件,提供预训练的参数,使系统可以直接被加载到模型中进行预测。系统数据源的输入包含图片、文件夹、视频和摄像头等多种输入方式,输入后的数据源将在系统页面上方显示,开始检测后,页面下方将实时显示处理后的结果图,包括检测到的目标类别及其置信度。最后设置固定的保存路径,将检测结果保存到本地文件中。

3 结 论

针对无人机视角下目标尺寸变化大、相互遮挡等特点导致的目标误检漏检问题,以 YOLOv8s 为基础,提出了一种融合多尺度特征的轻量级目标检测算法,设计轻量级多尺度卷积 EMSC,捕获不同尺度下的特征信息;引入 Dysample 上采样算子,保留图像细节信息;优化 Task Aligned Assigner 超参数,解决正负样本分配不均衡问题。最后,开发系统可视化界面,以展示航拍目标检测的效果。在 VisDrone2019 数据集上的实验结果表明,所提算法在保证检测速度几乎不变的同时,精准率、召回率、mAP_{0.5} 均有所提高,参数量和计算量降低。在 UAVDT 数据集上的验证进一步证实了本文算法的有效性。

参考文献

[1] 刘雨田. AR 无人机在安防领域的应用探讨[J]. 中国安防, 2024, 19(3): 39-44.
LIU Y T. Application of AR UAV in security field[J]. China Security & Protection, 2024, 19(3): 39-44.

[2] 刘宸,李士心,孟范润,等. 基于深度学习的无人机目标检测研究综述[J]. 计算机科学与应用, 2023, 13(5): 1092-1099.
LIU CH, LI SH X, MENG F R, et al. Survey of UAV target detection based on deep learning[J]. Computer Science and Application, 2023, 13(5): 1092-1099.

[3] 雷帮军,余翱,吴正平,等. 基于位置感知和跨层特征融合的航拍小目标检测算法[J]. 电子测量技术, 2024, 47(5): 112-123.
LEI B J, YU AO, WU ZH P, et al. Small target detection algorithm for aerial photography based on location awareness and cross-layer feature fusion[J]. Electronic Measurement Technology, 2024, 47(5), 112-123.

[4] 胡峻峰,李柏聪,朱昊,等. 改进 YOLOv8 的轻量化无人机目标检测算法[J]. 计算机工程与应用, 2024, 60(8): 182-191.
HU J F, LI B C, ZHU H, et al. Improved YOLOv8 lightweight UAV target detection algorithm [J]. Computer Engineering and Applications, 2024, 60(8): 182-191.

[5] 陈佳慧,王晓虹. 改进 YOLOv5 的无人机航拍图像密集小目标检测算法[J]. 计算机工程与应用, 2024, 60(3): 100-108.
CHEN J H, WANG X H. Dense small object detection algorithm based on improved YOLOv5 in UAV aerial images[J]. Computer Engineering and Applications, 2024, 60(3): 100-108.

- [6] SUI J, CHEN D, ZHENG X, et al. A new algorithm for small target detection from the perspective of unmanned aerial vehicles[J]. IEEE Access, 2024, 11(12): 29690-29697.
- [7] 颜豪男,吕伏,冯永安. 特征级自适应增强的无人机目标检测算法[J]. 计算机科学与探索, 2024, 18(6): 1566-1578.
- YAN H N, LYU F, FENG Y AN. Feature-level adaptive enhancement for UAV target detection algorithm[J]. Journal of Frontiers of Computer Science and Technology, 2024, 18(6): 1566-1578.
- [8] 殷波. 基于 ASF-WIoU-YOLOv8 的无人机航拍图像多目标检测算法[J]. 计算机科学与应用, 2024, 14(5): 137-146.
- YIN B. Multi-object detection algorithm for UAV aerial images based on ASF-WIoU-YOLOv8[J]. Computer Science and Application, 2024, 14(5): 137-146.
- [9] 潘玮,韦超,钱春雨,等. 面向无人机视角下小目标检测的 YOLOv8s 改进模型[J]. 计算机工程与应用, 2024, 60(9): 142-150.
- PAN W, WEI CH, QIAN CH Y, et al. Improved YOLOv8s model for small object detection from perspective of drones[J]. Computer Engineering and Applications, 2024, 60(9): 142-150.
- [10] WANG G, CHEN Y, AN P, et al. UAV-YOLOv8: A small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios[J]. Sensors, 2023, 23(16): 7190.
- [11] 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.
- CHEN P L, WANG J T, ZHANG ZH W, et al. Small object detection in aerial images based on feature aggregation and multiple cooperative features interaction[J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(10): 183-192.
- [12] LOU H T, DUAN X H, GUO J M, et al. DC-YOLOv8: Small-size object detection algorithm based on camera sensor[J]. Electronics, 2023, 12(10): 2323.
- [13] PU H W, YI K Y. A comparative analysis of efficientNet and mobileNet models' performance on limited datasets: An example of american sign language alphabet detection[J]. Highlights in Science: Engineering and Technology, 2024, 20(4): 558-564.
- [14] WANG T, ZHANG SH Q. DSC-Ghost-Conv: A compact convolution module for building efficient neural network architectures[J]. Multimedia Tools and Applications, 2023, 83(12): 36767-36795.
- [15] 唐建军,梁浩,朱张勤,等. 基于重排频谱时频脊的海面小目标检测[J]. 电光与控制, 2021, 28(4): 53-57.
- TANG J J, LIANG H, ZHU ZH Q, et al. Detection of small target on sea surface based on time-frequency ridge of rearranged spectrum[J]. Electronics Optics & Control, 2021, 28(4): 53-57.
- [16] LIU W Z, LU H, FU H T, et al. Learning to upsample by learning to sample[C]. IEEE/CVF International Conference on Computer Vision, 2023: 6027-6037.
- [17] DU D W, ZHU P F, WEN L Y, et al. VisDrone-DET2019: The vision meets drone object detection in image challenge results[C]. IEEE/CVF International Conference on Computer Vision Workshops. Seoul, South Korea: IEEE Press, 2019: 213-226.
- [18] DU D W, QI Y K, YU H Y, et al. The unmanned aerial vehicle benchmark: Object detection and tracking[C]. European Conference on Computer Vision (ECCV), 2018: 370-386.

作者简介

党宏社,博士,教授,主要研究方向为数字图像处理、工业机器人、无线传感网络。

E-mail: 919021914@qq.com

高宛蓉(通信作者),硕士研究生,主要研究方向为计算机视觉。

E-mail: 2239155747@qq.com

孟饶辰,硕士研究生,主要研究方向为深度学习。

E-mail: 2272557457@qq.com

张选德,博士,教授,主要研究方向为图像处理、模式识别、视觉追踪。

E-mail: 1621778143@qq.com