

光电工程

Opto-Electronic Engineering

中文核心期刊 中国科技核心期刊
Scopus CSCD

改进YOLOv7的无人机视角下复杂环境目标检测算法

张润梅, 肖钰霏, 贾振楠, 陈中, 陈梓华, 袁彬, 曹炜威, 宋妮妮

引用本文:

张润梅, 肖钰霏, 贾振楠, 等. 改进YOLOv7的无人机视角下复杂环境目标检测算法[J]. 光电工程, 2024, 51(5): 240051.

Zhang R M, Xiao Y F, Jia Z N, et al. Improved YOLOv7 algorithm for target detection in complex environments from UAV perspective[J]. *Opto-Electron Eng*, 2024, 51(5): 240051.

<https://doi.org/10.12086/oe.2024.240051>

收稿日期: 2024-03-06; 修改日期: 2024-04-21; 录用日期: 2024-04-24

相关论文

基于改进YOLOv5s的无人机图像实时目标检测

陈旭, 彭冬亮, 谷雨

光电工程 2022, 49(3): 210372 doi: 10.12086/oe.2022.210372

协同随机森林方法和无人机LiDAR空谱数据的盐沼植被“精灵圈”识别

韩江涛, 谭凯, 张卫国, 周若彤, 刘帅

光电工程 2024, 51(3): 230188 doi: 10.12086/oe.2024.230188

PAW-YOLOv7: 河道微小漂浮物检测算法

栾庆磊, 常昕昱, 吴叶, 邓从龙, 史艳琼, 陈梓华

光电工程 2024, 51(4): 240025 doi: 10.12086/oe.2024.240025

更多相关论文见光电期刊集群网站 



光电工程
Opto-Electronic Engineering

<http://cn.ojournal.org/oe>



 OE_Journal



Website

DOI: 10.12086/oe.2024.240051

改进 YOLOv7 的无人机视角下复杂环境目标检测算法

张润梅^{1,2,3,4}, 肖钰霏¹, 贾振楠¹, 陈 中^{1,2}, 陈梓华^{1,2},
袁 彬^{1,2,3,4}, 曹炜威⁴, 宋妮妮^{3*}

¹安徽建筑大学机械与电气工程学院, 安徽 合肥 230601;

²安徽省工程机械智能制造重点实验室, 安徽 合肥 230601;

³安徽省仿真设计与现代制造工程技术研究中心, 安徽 黄山 242700;

⁴民航飞行技术与飞行安全重点实验室, 四川 广汉 618300

摘要: 针对无人机在航拍过程中容易受到恶劣环境的影响, 导致航拍图像出现辨识度低、被障碍物遮挡、特征严重丢失等问题, 提出了一种改进 YOLOv7 的无人机视角下复杂环境的目标检测算法 (SSG-YOLOv7)。首先从 VisDrone2019 数据集和 RSOD 数据集中分别抽取图片进行五种环境的模拟, 将 VisDrone 数据集扩充至 12803 张, RSOD 数据集扩充至 1320 张。其次, 聚类出更适合数据集的锚框尺寸。接着将 3D 无参注意力机制 SimAM 引入主干网络和特征提取模块中, 增加模型的学习能力。然后重构特征提取模块 SPPCSPC, 融合不同尺寸池化通道提取的信息同时引入轻量级的卷积模块 GhostConv, 在不增加模型参数数量的同时提高算法对密集多尺度目标检测精度。最后使用 Soft NMS 优化锚框的置信度, 减少算法的误检、漏检率。实验结果表明, 在复杂环境的检测任务中 SSG-YOLOv7 检测效果优良, 性能指标 VisDrone_mAP@0.5 和 RSOD_mAP@0.5 较 YOLOv7 分别提高了 10.45% 和 2.67%。

关键词: 无人机; 复杂环境; YOLOv7; simAM 注意力机制; SPPCSPC; 数据增强

中图分类号: TP391

文献标志码: A

张润梅, 肖钰霏, 贾振楠, 等. 改进 YOLOv7 的无人机视角下复杂环境目标检测算法 [J]. 光电工程, 2024, 51(5): 240051
Zhang R M, Xiao Y F, Jia Z N, et al. Improved YOLOv7 algorithm for target detection in complex environments from UAV perspective[J]. *Opto-Electron Eng*, 2024, 51(5): 240051

Improved YOLOv7 algorithm for target detection in complex environments from UAV perspective

Zhang Runmei^{1,2,3,4}, Xiao Yufei¹, Jia Zhennan¹, Chen Zhong^{1,2}, Chen Zihua^{1,2},
Yuan Bin^{1,2,3,4}, Cao Weiwei⁴, Song Weiwei^{3*}

¹ School of Mechanical and Electrical Engineering, Anhui Jianzhu University, Hefei, Anhui 230601, China;

² Key Laboratory of Intelligent Manufacturing of Construction Machinery, Hefei, Anhui 230601, China;

³ Anhui Simulation Design and Modern Manufacturing Engineering Technology Research Center, Huangshan, Anhui 242700, China;

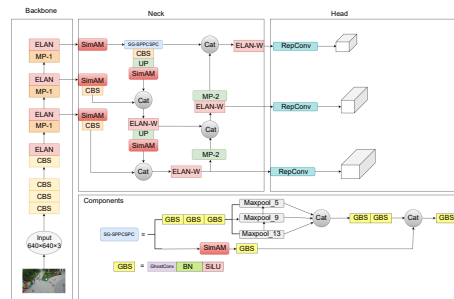
⁴ Key Laboratory of Civil Aviation Flight Technology and Flight Safety, Guanghan, Sichuan 618300, China

收稿日期: 2024-03-06; 修回日期: 2024-04-21; 录用日期: 2024-04-24

基金项目: 安徽省仿真设计与现代制造工程技术研究中心开放基金 (SGCXZD2101); 基于知识图谱的无人机安全知识库的构建 (FZ2021KF10)

*通信作者: 宋妮妮, swwahjzu11@163.com。

版权所有©2024 中国科学院光电技术研究所



Abstract: To address the challenges faced by drones during UAV (unmanned aerial vehicle) photography in adverse conditions, such as low image recognition, obstruction by obstacles, and significant feature loss, a novel algorithm named SSG-YOLOv7 was proposed to enhance object detection from the perspective of drones in complex environments. Firstly, 12803 images were augmented from the VisDrone2019 dataset, and 1320 images were augmented from the RSOD dataset to simulate five different environments. Subsequently, anchor box sizes suitable for the datasets were clustered. The 3D non-local attention mechanism SimAM was integrated into the backbone network and feature extraction module to enhance the model's learning capabilities. Furthermore, the feature extraction module SPPCSPC was restructured to integrate information extracted from channels with different pool sizes and introduce the lightweight convolution module GhostConv, thereby improving the precision of dense multi-scale object detection without increasing the model's parameter count. Finally, Soft NMS was employed to optimize the confidence of anchor boxes, reducing false positives and missed detections. Experimental results demonstrate that SSG-YOLOv7 exhibits superior detection performance in complex environments, with performance metrics VisDrone_mAP@0.5 and RSOD_mAP@0.5 showing improvements of 10.45% and 2.67%, respectively, compared to YOLOv7.

Keywords: UAV; complex environment; YOLOv7; simAM attention mechanism; SPPCSPC; data enhancement

1 引言

近年来, 随着科技的快速发展, 使用无人机 (unmanned aerial vehicle, UAV) 结合深度学习完成目标检测任务正在兴起, 无人机在安防巡警、搜索救援、电力检修等方面得到了广泛的应用^[1], 使用低成本的无人机摄影技术结合深度学习可以在各个领域创造出很高的价值。对于无人机摄影图像而言, 航拍图像中存在大量小目标并且拍摄物体尺度变化剧烈, 高速和低空飞行会使密集的目标产生运动模糊, 这使现有的目标检测方法受到了广泛的挑战。

目标检测是计算机视觉的一个重要分支, 旨在将图像或视频中感兴趣的目标区域进行定位与识别。随着计算机算力的提升和摄影设备技术的发展, 深度学习以其强大的特征学习能力被广泛地应用于图像处理和目标检测。基于深度学习的目标检测算法主要分为两种, 分别是以 RetinaNet^[2]、YOLO^[3-4] 系列为代表的单阶段 (one-stage) 检测方法和以 SPPNet^[5]、R-CNN^[6] 系列为代表的双阶段 (two-stage) 检测方法。单阶段检测算法主要是通过生成密集的锚框 (anchor-boxes) 来覆盖图像, 通过神经网络对每个锚框进行分类和回归; 而双阶段检测算法是使用一些策略, 比如选择性搜索 (selective search)^[7] 或 RPN^[8] 生成一组候选框然后对生成的候选框进行分类和回归。2015年由 Joseph Redmon 等人首次提出基于单个神经网络的目标检测算法 YOLO (you only look once)^[9], 只需要浏览一次就可以识别图中物体的类别和位置, 这使 YOLO 系

列算法在相同的识别类别范围中具有识别速度最快的特点, 随着 YOLO 系列的不断更新迭代, YOLO 模型的准确率在不断上升, 越来越多的专家学者开始研究如何在 YOLO 模型上进行改进使其更加适应于目标领域。其中 Yin 等^[10] 通过设计模块 AF-SSD, 将网络提取到的深层特征与浅层特征相融合, 以此来增加待检测目标的特征信息。Qi 等^[11] 将 SPPCSPC 结构中的两个卷积层更改为 SimAM 注意力机制, 从而提高网络对于密集目标的检测精度。Liu 等^[12] 将注意力机制 CA 加入主干网络, 同时将 BiFPN 结构引入颈部网络, 有效地利用了不同层次的特征信息, 理解目标和背景之间的关系, 使得检测精度大幅度提高。

上述研究在提高小目标检测精度和检测速度上具有良好的检测效果, 在航拍图片质量较低时效果不明显, 而无人机视角下的目标具有密集度高、种类多、多尺度等特点, 检测精度及效率易受天气影响, 航拍图像质量较低, 使得检测难度及计算负荷增加。因此, 本文以 YOLOv7^[13] 为基线网络, 做出以下四个部分的改进: 1) 使用 K-means++ 算法对 Visdrone 数据集和 RSOD 数据集进行聚类分析, 生成适合数据集的四个不同尺寸目标的锚框大小, 聚类后生成的锚框更贴合本数据集的目标尺寸。2) 将无参注意力机制 simAM 融入模型, 加入注意力机制的模型表现出明显的准确率和检测速度上的提升。3) 重构 SPPCSPC 模块, 使用 GhostConv 代替普通卷积, 减少了冗余特征的提取, 实现特征提取网络的轻量化。4) 使用 Soft NMS 代替 NMS, 使用高斯加权重构 IOU 大于阈值的

候选框的置信度, 大大减少了模型在检测过程中误检、漏检小尺寸目标的情况。同时对 Visdrone2019 数据集和 RSOD 数据集中的部分图片进行复杂天气模拟, 验证改进后的网络能够有效地应用于无人机视角下复杂环境的目标检测。

2 YOLOv7 算法介绍

随着 YOLO 系列算法的不断更新迭代, YOLOv7 在 5 FPS 至 160 FPS 范围内速度和精度均超过目前已知的所有物体检测器, 在 V100 的 GPU 上, 30 FPS 的情况下实时检测精度高达 56.8% AP^[10]。YOLOv7 模型在 YOLOv5 模型的基础上将模型重参数化引入到网络架构中、采用跨网格搜索和匹配的标签分配策略、提出一种高效的远程注意力网络 ELAN 以及包含辅助头的训练方法, 这使 YOLOv7 能够在不增加推理时间的情况下提高准确率^[14]。YOLOv7 的网络结构可以分为输入 (Input)、骨干网络 (Backbone)、头部 (Head)。

输入模块 (Input) 将输入进来的图像执行几个预处理步骤, 包括 Mixup、Mosaic 和随机缩放等数据增强, 其目的是丰富数据集并添加小目标, 使网络的检测效果更好。完成数据增强后将图像统一缩放到标准尺寸 (640×640×3), 输入到骨干网络中。

骨干网络 (Backbone) 通过深度卷积来提取不同尺度的特征信息, 并且将图像的特征信息进行深度的融合, 主要由 CSB 模块、ELAN 模块和 MP 模块组成。与之前 YOLO 系列的主干网络的区别主要在于 ELAN 模块和 MP 模块, ELAN 模块由两个分支组成, 第一个分支通过一个 1×1 的卷积模块, 第二个分支首先经过一个 1×1 的卷积模块做通道数的变化, 然后再经过四个 3×3 的卷积模块进行特征提取, 并将四个特征合并到一起, 通过控制最长和最短梯度路径来增强网络的鲁棒性。E-ELAN 是基于 ELAN 的设计的, 通过合并基数、洗牌和扩展, 实现保留初始梯度路径的情况下不断提高网络的学习能力。MP 模块也由两个分支构成, 其作用是进行下采样。第一个分支进行最大池化 (Maxpool), 再经过一个 1×1 的卷积做通道数的变化; 第二个分支先进行一个 1×1 的卷积再经过一个 3×3 的卷积。最后将两个分支的结果合并到一起。

头部网络 (Head) 用于模型预测, 由 SPPCSPC 模块, ELAN-W、MP-2 和 Repconv 组成, SPPCSPC 模块是在 SPP 结构中引入了 CSP 结构, 使其具有一个

残差边可以进行辅助优化, 减少重复提取相同特征网络的结构, 节省了计算成本。

3 本文算法

3.1 锚框的优化

YOLOv7 网络默认使用 K-means 算法对 COCO 数据集中得到的锚框进行聚类, 在后续的训练过程中通过遗传算法调整锚框的尺寸^[15]。K-means 聚类算法使用欧式距离作为衡量数据对象间的相似度, 其聚类效果非常依赖于初始中心点的选择, 为了更好的运算结果往往需要多次进行运算选取最优的聚类结果, 并且 K-means 算法只能迭代收敛到局部最优解, 无法获得全局最优解。为了解决这一问题, 本文使用 K-means++^[16] 算法代替 YOLOv7 中的 K-means 算法, K-means++ 算法是在传统的 K-means 算法上进行改进, 通过概率分布的方式来选择中心点, 一次只选择一个聚类中心, 然后通过距离权重来选择下一个中心点, 并且重复执行此过程直至选择 K 个中心点。这种聚类方式使得中心点的选择能够既趋向于局部最优解, 也趋向于全局最优解。

无人机视角下的目标检测具有目标尺寸大小不一、分布密集且背景复杂的特点, 传统的 K-means 算法聚类的锚框不能很好地适应无人机视角下的目标尺寸。通过采用 K-means++ 算法可以获得与实际数据集匹配度更高的锚框。如表 1 所示, 为 K-means++ 对本文使用的数据集中不同尺寸的目标进行聚类, 生成更加合适的锚框尺寸。

表 1 K-means++ 生成的锚框尺寸
Table 1 Anchor frame size generated by K-means++

特征图尺寸	感受野	锚框
20x20	Big	[33,49],[63,73]
40x40	Medium	[14,35],[27,23]
80x80	Small	[20,8,8,15,14]
160x160	Tiny	[2,5,4,11]

3.2 SimAM 注意力模块

注意力机制 (Attention mechanism)^[17] 具有选择性地聚焦于任务相关特定部分的特点, 可以从大量复杂的信息中筛选出重要的信息, 因此被广泛地应用于深度学习模型中。对于无人机视角下的复杂环境的目标

检测, 在网络中融合注意力机制可以提高网络中被测物体的权重, 使模型的学习更具有目的性。Yang 等人受神经科学理论的启发, 提出在注意力模块中使用能量函数来评估每个神经元的重要性, 提出 3D 无参注意力模块 SimAM^[18], 其结构图如图 1 所示。

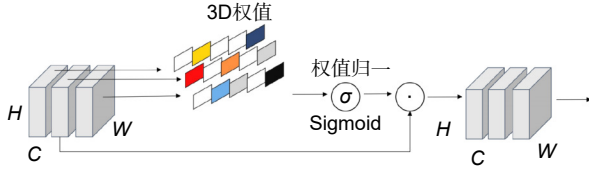


图 1 SimAM 注意力模块
Fig. 1 SimAM attention module

该注意力模块将能量函数与注意力机制相结合, 通过对每个神经元分配唯一的权重, 评估神经元的重要程度从而增强模型对于重要信息的关注程度, 与目前主流的注意力模块相比, SimAM 可以在模型中同时兼顾通道和空间维度的注意力, 并且结构相对简单, 推导出 3D 权重的同时保持模块的轻量化, 有效地避免了模型计算量增加的问题。目标神经元与其他神经元的区别越大, 该神经元表现的越重要。能量函数 Energy 定义为

$$e_i(w_i, b, y, x_i) = \frac{1}{M-1} \sum_{i=1}^{M-1} [-1 - (w_i x_i + b_i)]^2 + (1 - (w_i t + b_i))^2 + \lambda w_i^2, \quad (1)$$

其中: i 表示空间上的维度索引, $M=H \times M$ 表示某通道上神经元的个数, w_i 和 b_i 表示神经元变换时的权重和偏差。 w_i 和 b_i 的计算公式如下:

$$w_i = -\frac{2(t - u_i)}{(t - u_i)^2 + 2\sigma_i^2 + 2\lambda}, \quad (2)$$

$$b_i = -\frac{1}{2}(t + u_i)w_i, \quad (3)$$

其中: t 和 u_i 分别为输入特征的目标神经元和其他神经元, σ_i^2 表示所有神经元在每个通道上的方差, 使用 u_i 和 σ_i^2 减少计算量。其计算过程如下:

$$\sigma_i^2 = \frac{1}{M-1} \sum_{i=1}^{M-1} (x_i - u_i)^2, \quad (4)$$

$$u_i = \frac{1}{M-1} \sum_{i=1}^{M-1} x_i. \quad (5)$$

$\tilde{X}$ 表示计算的 3D 权重, E 表示每个通道上的能量函数, 通过添加 Sigmoid 函数来限制 E 中过大的值。 $\tilde{X}$ 与 E 的计算公式如下:

$$\tilde{X} = \text{Sigmoid}\left(\frac{1}{E}\right) \odot X, \quad (6)$$

$$E = \frac{4(\sigma_i^2 + \lambda)}{(t - u_i)^2 + 2\sigma_i^2 + 2\lambda}. \quad (7)$$

3.3 SPPCSPC 模块的改进

YOLOv7 中使用的 SPPCSPC 模块由空间金字塔池化结构 (SPP)^[19] 和分支结构 (CSP) 组成, 其中 SPP 模块主要通过最大池化扩大网络的感受野进行特征提取, 使模型更加具有泛化能力, 有效地捕获不同尺度和比例的目标信息。CSP 结构通过对特征图进行级联更好地结合上下文信息, 提升模型的鲁棒性。但 SPP 模块在进行特征提取的过程中易忽略与池化通道不匹配的特征图, 丢弃一些小目标具有的特征, 进而导致小目标无法被有效检测。考虑到无人机视角下的目标具有小尺度目标多, 密集度高的特点, 丢弃小目标的特征会使检测精度不断降低, 本文提出一种新的特征提取模块 SG-SPPCSPC 代替 SPPCSPC 模块。不同于 SPPCSPC 结构的并行池化过程, SG-SPPCSPC 模块将不同的池化通道进行整合, 极大地丰富了待测目标的特征信息, 显著减少池化过程中小目标特征丢失。尤其在小目标密集的检测情况下, 融合不同的池化通道能够显著地提高小目标检测的精确度。航拍过程中无人机处于低空高速飞行状态, 当风速较高或者天气恶劣时, 机体本身的抖动频繁发生, 航拍过程中的待测目标更易产生运动模糊。本文提出将 SimAM 注意力机制融入特征提取模块, 使模型能够有效应用于复杂环境的检测任务中。在实际应用中, 无人机设备不具有太高的算力, 使用更加轻量级的 GhostConv 代替 SPPCSPC 中的常规卷积, 实现模型轻量化的同时不影响检测精度。改进前后的特征提取模块参数量和 GFLOPs 对比结果见表 2, 改进后的网络 SSG-YOLOv7 结构如下图 2。

表 2 SPPCSPC 与 SG-SPPCSPC 模块参数量和 GFLOPs 对比结果

Table 2 Comparison of SPPCSPC and SG-SPPCSPC

模块类型	Parameters/M	GFLOPS
SPPCSPC 模块	12.8	16.2
SG-SPPCSPC 模块	3.69	4.9

GhostConv 的卷积结构包含常规卷积和 Ghost 卷积两个部分^[20]。首先使用常规卷积来提取图片中丰富的特征, 这些丰富的特征中包含了大量的冗余特征, 大大占用了模型的计算资源, 因此 GhostConv 在常规卷积之后对特征图采用廉价的线性变化进行运算, 通过合并生成最终的特征图, 作用过程如图 3 所示。

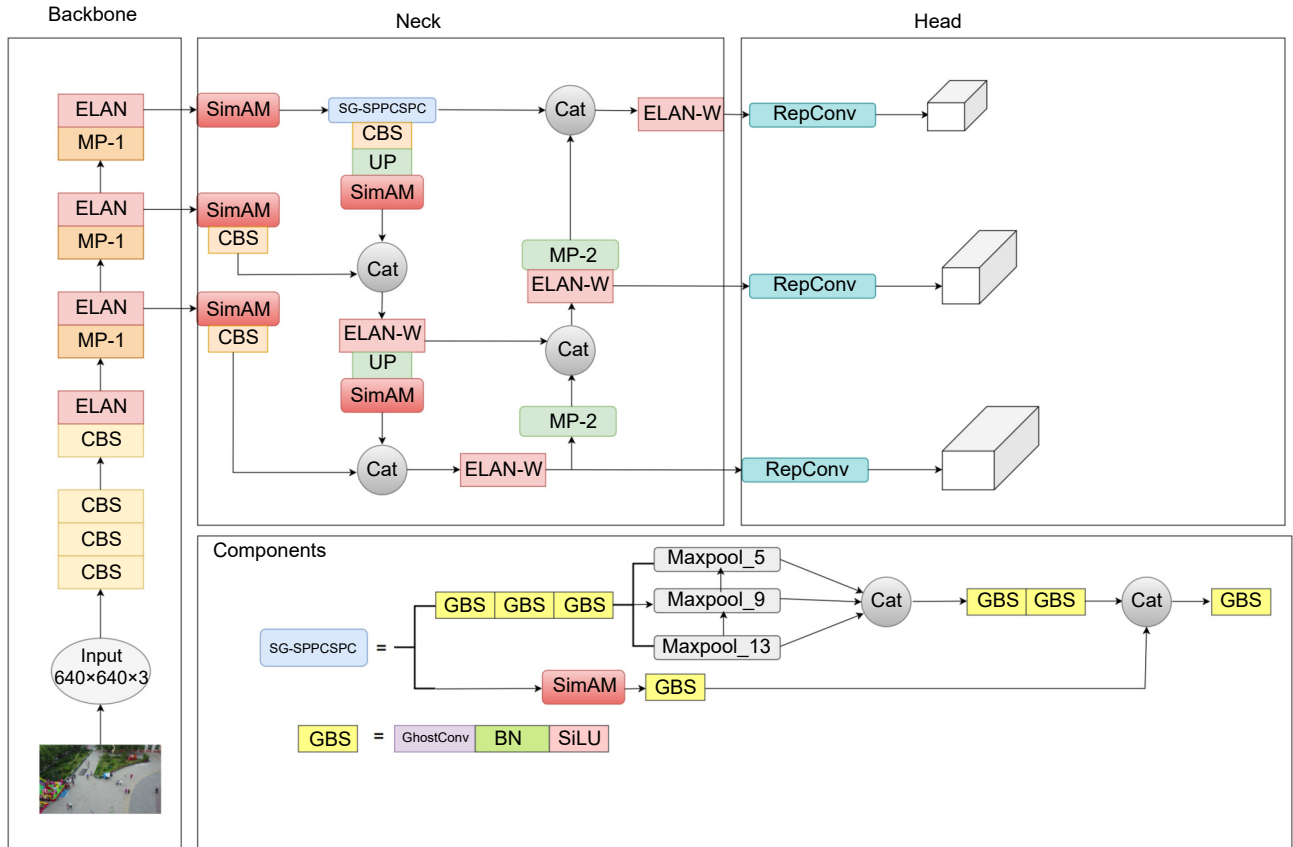


图2 SSG-YOLOv7 整体结构

Fig. 2 SSG-YOLOv7 overall structure

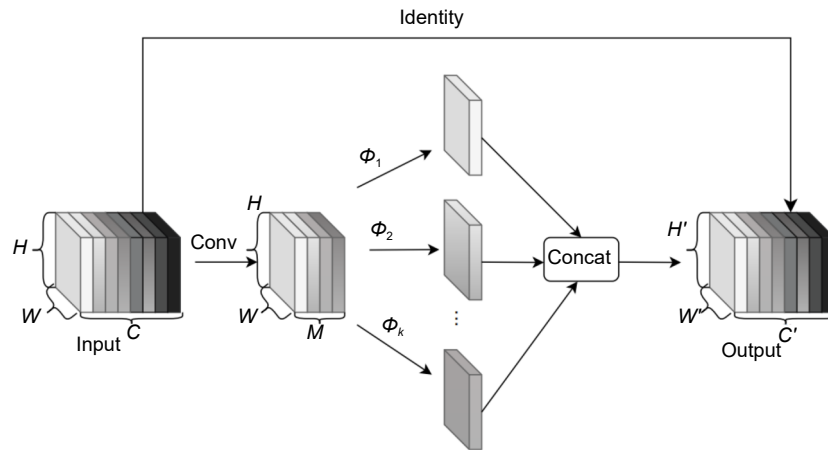


图3 GhostConv 结构图

Fig. 3 GhostConv structure

假设输入数据的张量为 $W \times H \times C$, 经过第一次常规卷积后的输出数据的张量为 $W' \times H \times M$, 其运算参见式 (8) ~ (10):

$$r_1 = M \times W' \times H' \times C \times k \times k, \quad (8)$$

$$r_2 = (S - 1) \times n \times H' \times W' \times d \times d, \quad (9)$$

$$r_s = \frac{r_1}{r_2} \approx S, \quad (10)$$

其中, $S (S > 1)$ 为第一次常规卷积后输出通道产生的总映射数, 即单位映射数 (Φ_i) 用 $S-1$ 表示。经过单位映射后生成 Ghost 特征图, Ghost 特征图与第一次常规卷积产生的特征图合并成通道数为 n 的最终特征图。 r_1 表示常规卷积的计算总量, k 表示常规卷积核的大小; r_2 表示 GhostConv 的计算总量, d 表示线性变化

卷积核大小。 r_s 大于 1 进而说明使用 GhostConv 代替常规卷积能够有效地减少运算参数量。

3.4 Soft NMS

基于无人机视角下目标检测的特殊性, 航拍过程中高楼、大树以及地面建筑容易对待测目标产生视角上的遮挡。原始的 YOLOv7 使用的非极大值抑制算法 NMS (non-maximum suppression) 会对候选框的置信度进行排序。与最具代表性的候选框的 IOU (intersection over union) 相对比, 删除 IOU 大于一定阈值的候选框, 在无人机视角下; 复杂环境的目标检测任务中, 采用 NMS 算法易造成误检漏检。Soft NMS 是对 NMS 的一种改进, Soft NMS 不会直接删除 IOU 大于阈值的候选框而是降低置信度^[21], 尽可能保留更多的候选框, 有利于任务中小目标的检测, 如图 4 所示对比了 NMS 与 Soft NMS 对于被遮挡物体的检测效果图。本文采用 Soft NMS 高斯加权重置任务中的置信度, 高斯加权公式为

$$S_i = S_i e^{-\frac{IoU(M, b_i)^2}{\sigma}}, \forall b_i \notin D, \quad (11)$$

式中, 对于每一个候选框 b_i , 计算其与最具有代表性的候选框 M 的交并比, 当 IOU 大于阈值时使用高斯惩罚系数降低该候选框的得分 S_i , 有效地避免了直接删除候选框造成的漏检。

4 实验

4.1 实验环境及参数的设置

本文所有实验均在 Windows10 平台上完成, CPU 采用 Intel Core i7-13700KF, GPU 采用 NVIDIA GeForce RTX 4090(24G 显存), 使用 Python3.7、Pytorch1.9.1、CUDA11.1 框架实现模型的搭建以及训练。

实验中使用的预训练权重均来自官方, 输入图像的尺寸为 640×640, 批处理量 (batch_size) 设置为 8, 迭代次数 (epoch) 设置为 300, 选用 SGD 作为优化器, 初始学习率设置为 0.001, 最大学习率设置为 0.01,

权重衰减参数设置为 5×10^{-4} , 在训练工程中使用余弦退火算法进行学习率的调整。

4.2 数据集的处理

本文选择使用的基础数据集分别是 VisDrone2019-DET^[22] 和 RSOD, 其中 VisDrone2019 数据集由天津大学机器学习与数据挖掘实验室团队进行收集和标注, 该数据集包含不同的城市和乡村环境、密集与稀疏的场景以及十种不同类别的物体, 目标类别有货车、汽车、卡车、公共汽车、行人、人、摩托车、自行车、三轮车和遮篷三轮车, 总共 8018 张图片。RSOD 数据集包含飞机、油箱、操场和立交桥四个类别, 总共 976 张图片包含 780 张训练集、98 张测试集、98 张验证集。将两种数据集按照 8: 1: 1 进行训练集、测试集和验证集的划分, 划分后 VisDrone2019 数据集包含 6414 张训练集、802 张测试集、802 张验证集。

为了模拟无人机在多种复杂环境下的航拍效果, 本文使用 Imgaug 库中的变换函数分别对 VisDrone2019 数据集和 RSOD 数据集进行数据增强, 从两种数据集的训练集、验证集和测试集中分别抽取多张图片进行了薄雾、雨天、浓雾、运动模糊、雨雾五种状态的模拟, 扩充后 VisDrone2019 数据集达到 12803 张, 其中 10241 张训练集, 1281 张验证集, 1281 张测试集。RSOD 数据集达到 1320 张, 其中 1056 张训练集, 132 张验证集, 132 张测试集。从两种数据集中各选取一张航拍图像 (每一小组的左侧原始图片选取自 VisDrone2019 数据集, 右侧原始图片选取自 RSOD 数据集) 进行五种效果模拟的数据增强, 数据增强效果对比如图 5 所示。

4.3 评价指标

为评估文中模型的性能, 本文采用的评价指标有平均精度 (AP)、平均精度均值 (mAP)、参数量 (Parameters)、准确率 (Precision)、召回率 (Recall)、模型每秒处理的帧数 (FPS)、模型算力 (GFLOPs),



图 4 (a) NMS 与 (b) Soft NMS 检测效果对比示例图

Fig. 4 Comparison of (a) NMS and (b) soft NMS detection effect sample chart

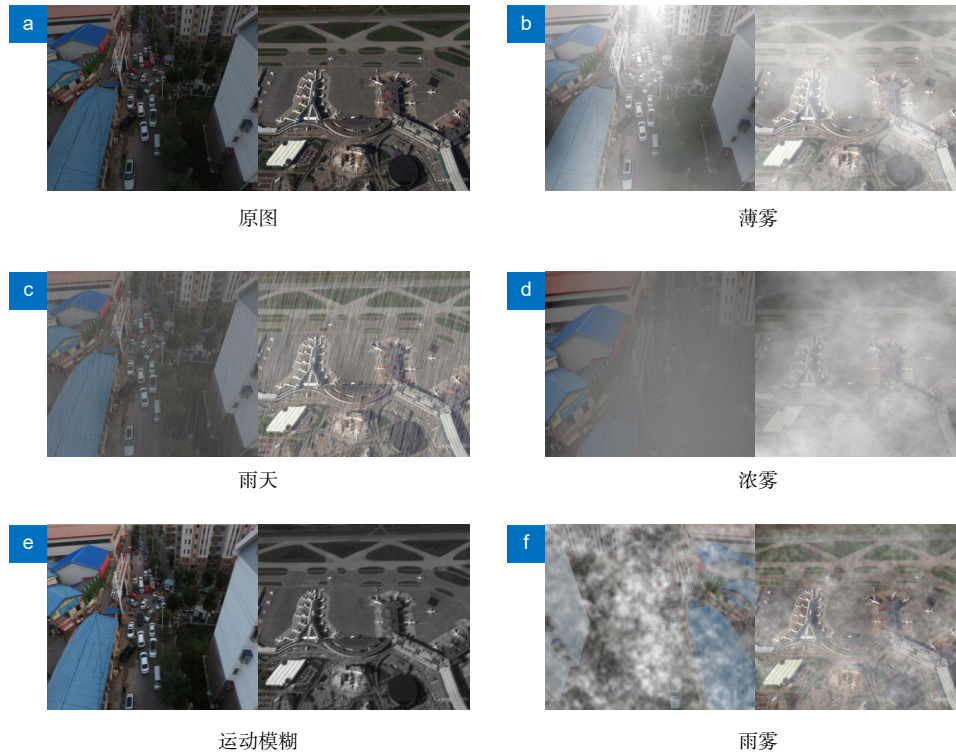


图 5 两类数据集数据增强效果对比

Fig. 5 Data augmentation comparison of two kinds of datasets

1GFLOPs 等于 10 亿次浮点运算用来描述模型的复杂度。主要的评价指标为 mAP ，相应的计算公式为

$$mAP = \frac{\sum_{i=1}^N AP}{N}, \quad (12)$$

其中： AP 表示每个类别的平均精度， AP 的值是 P-R 曲线与坐标轴围成的面积， mAP 为所有类别的平均精度 AP 相加权重取平均值， N 代表数据集中的类别数。 AP 与 mAP 越大代表模型的检测性能越好。

$$precision = \frac{TP}{TP + FP}, \quad (13)$$

$$Recall = \frac{TP}{TP + FN}, \quad (14)$$

$$AP = \int_0^1 P(R) dR, \quad (15)$$

式中： TP 表示真正例，即正样本识别为正类； FP 表

示假正例，即负样本识别为正类； FN 表示假负例，即正样本识别为负类。此外，本实验中采用的 mAP 包括 $mAP@0.5$ 和 $mAP@0.5:0.95$ ， $mAP@0.5$ 表示在 IOU 阈值设置为 0.5 时的 mAP ， $mAP@0.5:0.95$ 表示在 IOU 阈值处于 0.5~0.95 之间，步长为 0.05 的 10 个 IOU 阈值的平均 mAP 。

4.4 消融实验及结果分析

为了验证本文提出的各个模块的改进措施对于无人机视角下的目标检测效果有提升作用，本文设置了 A~E 总共五个组别的消融实验，将改进后的模型分别在两个数据集上进行验证，实验结果对比如表 3 所示。其中“√”代表该组实验引入的方法，Vis_mAP0.5 表示使用 VisDrone 数据集训练后模型的平均精度均值，RSOD_mAP0.5 表示使用 RSOD 数据集训练后模型的平均精度均值。在消融实验中逐步将锚框

表 3 消融实验结果

Table 3 Results of ablation experiments

Model	K-means++	SimAM	SG-SPPCSPC	Soft NMS	Vis_mAP@0.5/%	RSOD_mAP@0.5/%	Parameters/M	FPS	GFLOPs
A					40.89	95.60	37.6	82	106.5
B	√				44.15	96.91	37.6	82	106.5
C	√	√			46.40	97.22	37.6	87	107.2
D	√	√	√		48.61	97.91	28.5	93	95.9
E	√	√	√	√	51.34(+10.45)	98.27(+2.67)	28.5	93	95.9

聚类算法 K-means++、注意力机制 simAM、改进后的特征提取模块 SG-SPPCSPC 和软化的非极大值抑制算法 Soft NMS 嵌入到基础模型 YOLOv7 中。通过对比嵌入模块前后模型的检测精度、模型参数量、每秒处理帧数和模型复杂度, 评价每个模块对于原算法的优化效果。A 组实验使用 VisDrone 数据集和 RSOD 数据集分别对原始的 YOLOv7 模型进行训练, 使用 VisDrone 数据集训练后的 mAP0.5 只有 40.89%, RSOD 数据集训练后的 mAP0.5 达到 95.60%; B 组使用 K-means++ 进行聚类生成了 9 个更适合本数据集的候选框, Vis_mAP0.5 与 RSOD_mAP0.5 分别提升 3.26% 和 1.31%。C 组实验引入 3-D 无参注意力机制后, 模型的参数量不增加, 同时 Vis_mAP0.5 提升了 2.25%, RSOD_mAP0.5 提升 0.31%。D 组实验在 C 组的基础上进一步引入 SG-SPPCSPC 模块, Vis_mAP0.5 提升 2.21%, RSOD_mAP0.5 提升 0.69%, 同时参数量和 GFLOPs 明显降低, FPS 显著提升。E 组实验融合四项改进策略, 改进后的模型 Vis_mAP0.5 达到 51.34%, 较改进之前提高 10.45%, RSOD_mAP0.5 达到 98.27%, 较改进之前提高 2.67%, 同时参数量、FPS 和 GFLOPs 比原始模型分别下降 24.2%、13.4% 和 10.0%, 有效地证明了 SSG-YOLOv7 可以在无人机视角下复杂环境的目标检测任务中取得出色的成果。

表 4 展示了使用 Imgaug 库对两种数据集进行数据增强前后, YOLOv7 和 SSG-YOLOv7 的检测结果对比。可以看到, 数据增强后, YOLOv7 和 SSG-YOLOv7 在数据集上的检测效果均高于原始数据集。其中, YOLOv7 在 VisDrone 数据集中提升了 4.13%, 在 RSOD 数据集中提升了 3.59%; SSG-YOLOv7 在 VisDrone 数据集中提升了 8.71%, 在 RSOD 数据集中提升了 4.45%。这有效地证明了本文提出的数据增强方式能够使模型学习到复杂环境下图像中更多的细节, 提升模型在复杂环境下的检测精度。

为了更加直观地看到 SSG-YOLOv7 的检测效果, 本文从两种数据集中各抽取一张图片 (第一、三列原始图片选取自 VisDrone2019 数据集, 第二、四列原始图片选取自 RSOD 数据集) 对其进行运动模糊、浓雾、雨天、薄雾、雨雾五种环境的模拟。使用

YOLOv7 与 SSG-YOLOv7 对原图和变换后的图片分别进行检测, 检测结果可视化对比如图 6 所示。其中左边 (第一、二列) 为使用 YOLOv7 进行检测的结果, 右边 (第三、四列) 为使用 SSG-YOLOv7 进行检测的结果, 通过对比可以发现在各类复杂环境下, 原始 YOLOv7 网络存在大量误检漏检的情况, 尤其是对于一些被建筑物遮挡了部分的目标, YOLOv7 往往不能正确地判断。而本文提出的 SSG-YOLOv7 特征提取模块被进一步改进, 融合不同的池化通道与 3D 权重相结合, 使 SSG-YOLOv7 在无人机视角下的目标检测中具有明显优势。

4.5 与主流网络进行对比

为了进一步验证本文改进的算法在无人机视角下复杂环境的目标检测任务中更具有优势, 本文设置了一系列对比实验, 实验过程中确保实验环境、实验数据集与实验超参数的一致性, 实验结果如表 5 所示。通过对比主流目标检测实验结果可以发现针对无人机视角下复杂环境的目标检测中, Faster R-CNN 与 SSD 网络检测精度明显低于 YOLO 系列。本文改进算法 SSG-YOLOv7 的 Visdrone_mAP0.5 达到 51.34%, RSOD_mAP0.5 达到 98.3%, Visdrone_mAP0.5:0.95 达到 29.21%, RSOD_mAP0.5:0.95 达到 70.0%, 与 YOLO 系列算法相比, SSG-YOLOv7 在两个数据集上的精度均达到最优。同时模型参数量较低, 帧处理速度相较原始 YOLOv7 有明显提升, 并且优于模型参数量相近的 YOLOv8l 与 YOLOv5l, 提升网络精度的同时没有牺牲推理速度。综上可以得出, SSG-YOLOv7 网络在无人机视角下的目标检测任务中, 性能明显优于其他网络。

5 结 论

针对无人机视角下复杂天气环境的目标检测任务, 以及小尺度目标误检漏检问题, 本文采用 Imgaug 变换函数分别对 Visdrone2019 数据集和 RSOD 数据集进行数据增强, 模拟了五种恶劣天气下的航拍效果, 以 YOLOv7 为基线网络, 提出一种新的目标检测算法 SSG-YOLOv7。使用 K-means++ 聚类算法聚类出九个不同尺寸的锚框, 使其更适合本文数据集。在颈

表 4 数据增强前后 mAP(%) 对比
Table 4 Comparison of mAP(%) before and after data enhancement

Model	原始VisDrone	增强后VisDrone	原始RSOD	增强后RSOD
YOLOv7	36.76	40.89	92.01	95.60
SSG-YOLOv7	42.63	51.34	93.82	98.27

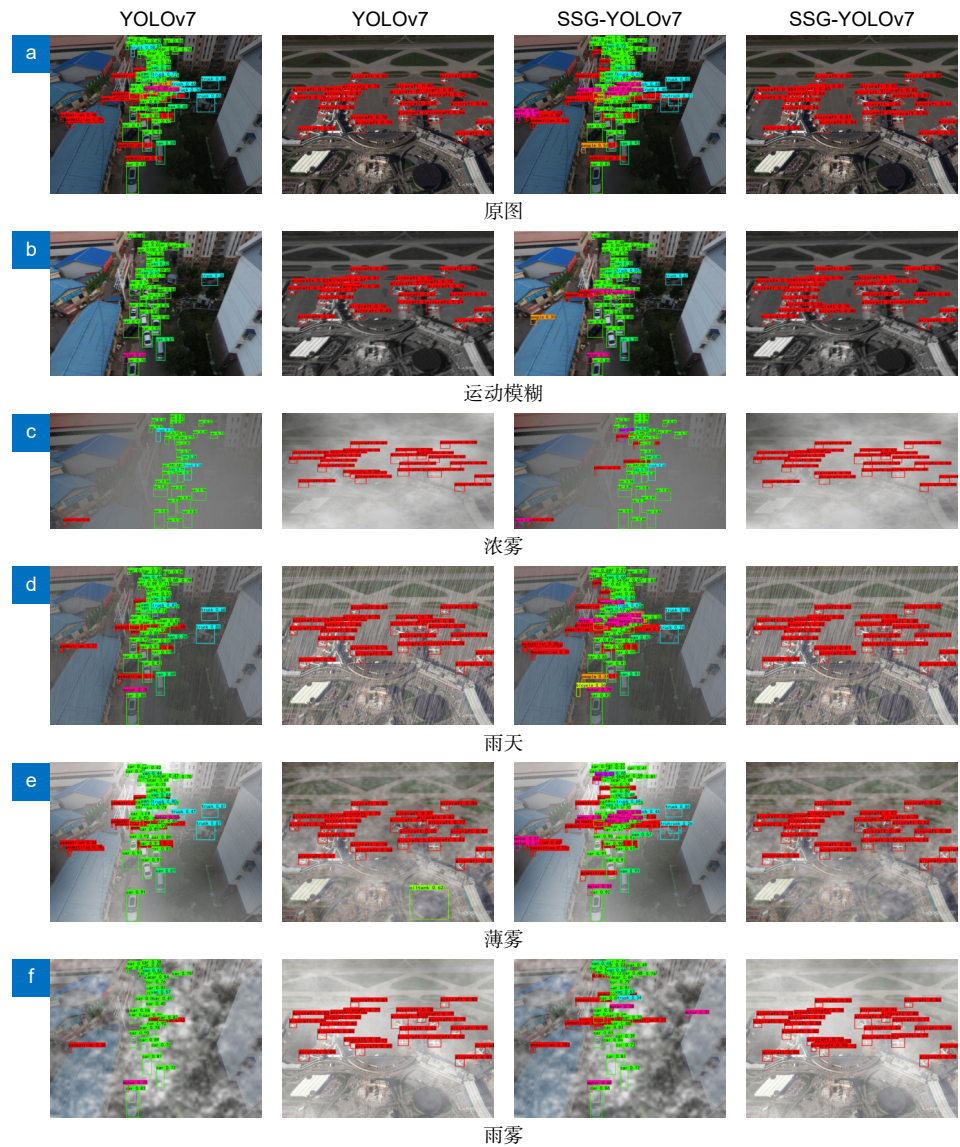


图 6 YOLOv7 与 SSG-YOLOv7 检测效果可视化对比

Fig. 6 Visual comparison of YOLOv7 and SSG-YOLOv7 detection effect

表 5 对比实验结果

Table 5 Comparison of experimental results

Method	Visdrone_mAP@0.5 /%	Visdrone_mAP@0.5:0.95 /%	RSOD_mAP@0.5 /%	RSOD_mAP@0.5:0.95 /%	FPS	Parameters/M
Faster R-CNN ^[6]	20.0	8.91	85.6	54.1	43	137.10
SSD ^[23]	10.2	5.1	87.4	52.6	249	26.29
YOLOv5s	27.4	15.6	94.0	59.5	126	7.28
YOLOv5m	32.0	18.8	95.2	66.4	98	21.38
YOLOv5l	36.5	21.5	95.1	68.3	75	47.10
YOLOv7 ^[13]	40.8	24.0	95.6	69.6	82	37.62
YOLOv8s	43.1	25.0	94.1	63.0	160	11.17
YOLOv8m	39.6	22.8	94.1	68.7	122	25.90
YOLOv8l	43.7	25.1	96.0	68.9	98	43.69
本文算法	51.3	29.2	98.3	70.0	93	28.49

部网络和特征提取模块中引入无参注意力机制 simAM, 使网络更好地感受到目标的空间位置信息, 并且不增加模型参数量。重构特征提取模块, 将多尺

寸池化通道相融合同时引入 GhostConv 层, 使网络提取到的目标特征更加丰富, 达到检测精度与检测速度的平衡。使用 Soft NMS 代替 NMS 有效减少了模型

对于小目标的误检、漏检率。采用改进后的算法在两种数据集上进行大量实验, 证明了本文算法能够更好地应用于无人机视角下复杂环境的目标检测任务中。

在进一步的工作中, 我们致力于在降低模型参数的同时保证较高精度, 使模型更加容易部署在算力不高的设备上。

利益冲突: 所有作者声明无利益冲突

参考文献

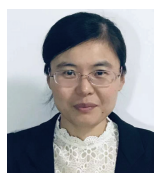
- [1] Chen X, Peng D L, Gu Y. Real-time object detection for UAV images based on improved YOLOv5s[J]. *Opto-Electron Eng*, 2022, 49(3): 210372.
陈旭, 彭冬亮, 谷雨. 基于改进YOLOv5s的无人机图像实时目标检测[J]. *光电工程*, 2022, 49(3): 210372.
- [2] Yang S, Wang J, Hu L, et al. Research on occluded object detection by improved RetinaNet[J]. *Comput Eng Appl*, 2022, 58(11): 209–214.
阳珊, 王建, 胡莉, 等. 改进RetinaNet的遮挡目标检测算法研究[J]. *计算机工程与应用*, 2022, 58(11): 209–214.
- [3] Zhan W, Sun C F, Wang M C, et al. An improved YOLOv5 real-time detection method for small objects captured by UAV[J]. *Soft Comput*, 2022, 26(6): 361–373.
- [4] Liu W, Quijano K, Crawford M M. YOLOv5-tassel: detecting tassels in RGB UAV imagery with improved YOLOv5 based on transfer learning[J]. *IEEE J Sel Top Appl Earth Obs Remote Sens*, 2022, 15: 8085–8094.
- [5] Purkait P, Zhao C, Zach C. SPP-Net: deep absolute pose regression with synthetic views[Z]. arXiv: 1712.03452, 2017. <https://doi.org/10.48550/arXiv.1712.03452>.
- [6] Girshick R. Fast R-CNN[C]//*Proceedings of 2015 IEEE International Conference on Computer Vision*, 2015: 1440–1448. <https://doi.org/10.1109/ICCV.2015.169>.
- [7] Uijlings J R R, van de Sande K E A, Gevers T, et al. Selective search for object recognition[J]. *Int J Comput Vis*, 2013, 104(2): 154–171.
- [8] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE Trans Pattern Anal Mach Intell*, 2017, 39(6): 1137–1149.
- [9] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016. <https://doi.org/10.1109/CVPR.2016.91>.
- [10] Yin R H, Zhao W, Fan X D, et al. AF-SSD: an accurate and fast single shot detector for high spatial remote sensing imagery[J]. *Sensors*, 2020, 20(22): 6530.
- [11] Qi X M, Chai R, Gao Y M. Algorithm of reconstructed SPPCSPC and optimized downsampling for small object detection[J]. *Comput Eng Appl*, 2023, 59(20): 158–166.
齐向明, 柴蕊, 高一萌. 重构SPPCSPC与优化下采样的小目标检测算法[J]. *计算机工程与应用*, 2023, 59(20): 158–166.
- [12] Shang J C, Wang J S, Liu S B, et al. Small target detection algorithm for UAV aerial photography based on improved YOLOv5s[J]. *Electronics*, 2023, 12(11): 2434.
- [13] 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[C]//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023: 7464–7475. <https://doi.org/10.1109/CVPR52729.2023.00721>.
- [14] Tang F, Yang F, Tian X Q. Long-distance person detection based on YOLOv7[J]. *Electronics*, 2023, 12(6): 1502.
- [15] Huang T Y, Cheng M, Yang Y L, et al. Tiny object detection based on YOLOv5[C]//*Proceedings of the 2022 5th International Conference on Image and Graphics Processing*, 2022: 45–50. <https://doi.org/10.1145/3512388.3512395>.
- [16] Ismahan H. I-k-means+: an iterative clustering algorithm based on an enhanced version of the k-means[J]. *Pattern Recognit*, 2018, 79: 402–413.
- [17] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//*Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017: 6000–6010. <https://doi.org/10.5555/3295222.3295349>.
- [18] Yang L X, Zhang R Y, Li L D, et al. SimAM: a simple, parameter-free attention module for convolutional neural networks[C]//*Proceedings of the 38th International Conference on Machine Learning*, 2021: 11863–11874.
- [19] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 936–944. <https://doi.org/10.1109/CVPR.2017.106>.
- [20] Han T, Wang Y H, Tian Q, et al. GhostNet: more features from cheap operations[C]//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 1577–1586. <https://doi.org/10.1109/CVPR42600.2020.00165>.
- [21] Bodla N, Singh B, Chellappa R, et al. Soft-NMS - improving object detection with one line of code[C]//*Proceedings of 2017 IEEE International Conference on Computer Vision*, 2017: 5562–5570. <https://doi.org/10.1109/ICCV.2017.593>.
- [22] Du D W, Zhu P F, Wen L Y, et al. VisDrone-DET2019: the vision meets drone object detection in image challenge results[C]//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision Workshop*, 2019: 213–226. <https://doi.org/10.1109/ICCVW.2019.00030>.
- [23] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[C]//*Proceedings of the 14th European Conference on Computer Vision*, 2016: 21–37. https://doi.org/10.1007/978-3-319-46448-0_2.

作者简介



张润梅 (1971-), 女, 博士, 教授, 主要研究方向为机器人技术、图像处理。

E-mail: zrmahjzu11@163.com



【通信作者】宋妮妮 (1984-), 女, 博士, 副教授, 主要研究方向为搅拌摩擦表面改性技术、现代加工技术。

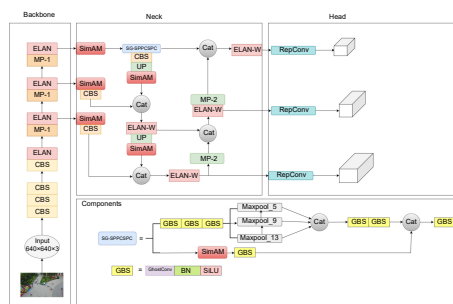
E-mail: swwahjzu11@163.com



扫描二维码, 获取PDF全文

Improved YOLOv7 algorithm for target detection in complex environments from UAV perspective

Zhang Runmei^{1,2,3,4}, Xiao Yufei¹, Jia Zhennan¹, Chen Zhong^{1,2},
Chen Zihua^{1,2}, Yuan Bin^{1,2,3,4}, Cao Weiwei⁴, Song Weiwei^{3*}



SSG-YOLOv7 overall structure

Overview: Using low-cost unmanned aerial vehicle (UAV) photography technology combined with deep learning can create significant value in various fields. Targets captured from a UAV perspective often exhibit drastic scale variations, uneven distribution, and susceptibility to obstruction by obstacles. Moreover, UAVs typically fly at low altitudes and high speeds during the capture process, which can result in low-resolution aerial images affected by weather conditions or the drone's own vibrations. Maintaining high detection accuracy in such complex environments is a crucial challenge in UAV-based target detection tasks. Therefore, this paper proposes a new target detection algorithm, SSG-YOLOv7, based on YOLOv7. Firstly, the algorithm utilizes the K-means++ clustering algorithm to generate four different-scale anchor boxes suitable for the target dataset, effectively addressing the issue of large-scale variations in targets from the UAV perspective. Next, by introducing the SimAM attention mechanism into the neck network and feature extraction module, the model's detection accuracy is improved without increasing the model's parameter count. Subsequently, the pooling layers at different scales of the feature extraction module are fused to enable the model to learn richer target feature information in complex environments. Additionally, GhostConv is used to replace traditional convolutional modules to reduce the parameter count of the feature extraction module. Finally, Soft NMS is employed to reduce the false detection and missed detection rates of small-scale targets during the detection process, thereby enhancing target detection effectiveness from the UAV perspective. In the experimental process, the original VisDrone dataset and RSOD dataset are simulated under five complex environments using transformation functions from the Imgaug library. SSG-YOLOv7 is validated against the original algorithm. Compared to the original algorithm, the proposed algorithm improves the average precision (mAP@0.5) of the model by 10.45% in the VisDrone dataset and by 2.67% in the RSOD dataset, while reducing the model's parameter count by 24.2%. This effectively demonstrates that SSG-YOLOv7 is better suited for target detection tasks in complex environments from the UAV perspective. Additionally, the experiment compares the detection accuracy of YOLOv7 and SSG-YOLOv7 before and after data augmentation on both datasets. In the VisDrone dataset, YOLOv7 improves by 4.13%, while SSG-YOLOv7 improves by 8.71%. In the RSOD dataset, YOLOv7 improves by 3.59%, while SSG-YOLOv7 improves by 4.45%. This effectively proves that SSG-YOLOv7 can learn more target features from samples in complex environments, accurately locate the targets, and is suitable for multi-target detection tasks in complex environments from the UAV perspective.

Zhang R M, Xiao Y F, Jia Z N, et al. Improved YOLOv7 algorithm for target detection in complex environments from UAV perspective[J]. *Opto-Electron Eng*, 2024, 51(5): 240051; DOI: [10.12086/oee.2024.240051](https://doi.org/10.12086/oee.2024.240051)

Foundation item: Project supported by Open Fund of Anhui Simulation Design and Modern Manufacturing Engineering Technology Research Centre (SGCZXZD2101), and Construction of UAV Safety Knowledge Base Based on Knowledge Graph (FZ2021KF10)

¹School of Mechanical and Electrical Engineering, Anhui Jianzhu University, Hefei, Anhui 230601, China; ²Key Laboratory of Intelligent Manufacturing of Construction Machinery, Hefei, Anhui 230601, China; ³Anhui Simulation Design and Modern Manufacturing Engineering Technology Research Center, Huangshan, Anhui 242700, China; ⁴Key Laboratory of Civil Aviation Flight Technology and Flight Safety, Guanghan, Sichuan 618300, China

* E-mail: swwahjzu11@163.com