http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
**DOI**: 10.13700/j.bh.1001-5965.2024.0420

# 无人机视角下目标检测的YOLO系列算法研究进展

陈金林，吴一全*，苑玉彬

(南京航空航天大学 电子信息工程学院，南京 211106)

**摘 要**：在诸多基于深度学习的目标检测算法中，YOLO系列算法因具有检测速度快、检测精度高、适应性强等优点，成为无人机(UAV)航拍图像目标检测的优先选择。UAV航拍图像与视频复杂多变的特点使该视角下YOLO系列目标检测算法成为一个具有挑战性的难题，因此，有必要将近期UAV视角下目标检测的YOLO系列算法研究与应用成果进行全面总结与分析。归纳10种YOLO基础版本的发展、联系、优化特点，重新绘制YOLO系列结构图，指出各种YOLO版本优化策略的作用与特点。综述在UAV目标检测应用领域中YOLO的优化方法数据训练前的预处理、训练与推理优化、注意力模块的应用、卷积优化、检测头及损失函数的优化；梳理UAV视角下的应用数据集及数据扩展方法。通过实验对UAV视角下数据集上YOLO系列算法的性能及卷积块注意力模块(CBAM)对YOLO不同改进的作用进行评估。指出了UAV视角下目标检测YOLO系列算法的研究还存在的一些问题与可能发展趋势。

**关 键 词**：无人机系统；目标检测；深度学习；YOLO系列算法；数据集；性能对比
**中图分类号**：V279; TP391.41
**文献标志码**：A **文章编号**：1001-5965(2026)10-3509-26

无人机(unmanned aerial vehicle, UAV)起飞灵活，操作简单，受地形障碍物影响较小，在恶劣环境下也具有较强的适应性，这使其可作为一种获取低空高分辨率遥感图像与视频的专用遥感系统平台。与传统平台相比，UAV航拍成本低、效率高，且能够弥补由于环境、时间、成本等限制造成的信息损失。UAV采集的图像、视频与音频等数据无论是内容上还是时间上都具有丰富的信息量。UAV通过搭载的摄像机与智能系统，能够实现实时处理拍摄的图像与视频，使其具有目标检测、识别、定位与追踪等功能[1]。

现阶段目标检测已经成为UAV应用领域的一项重要技术，尤其是UAV与深度学习的紧密融合，使UAV航拍目标的智能检测与识别技术开始逐步应用到各个行业中[2]。UAV视角下的目标检测在军事侦察、精准打击、智慧农林业、交通监控、电力巡检、灾害监测、搜索救援等军事与民用领域发挥着越来越重要的作用[3]。

UAV的飞行姿态大体分为起飞、悬停、巡航和降落，图像与视频的拍摄通常处于悬停与巡航状态下[4]。受UAV的飞行稳定性、载荷能力、飞行高度、巡航能力与飞行环境等因素的影响，UAV视角下的目标物体多数呈现出尺度多样、数量多、目标易变形、背景和场景复杂等特点，可提取的单个目标特征信息较少，这使目标检测变得更加困难[5-6]。UAV机载遥感传感器需要实时准确地调整飞行速度、飞行高度、飞行姿态、航拍角度，使目标物体始终处于UAV的监控范围内，因此，优秀的目标检测算法可以促进UAV飞行控制技术的进步。

现阶段UAV目标检测算法有基于两阶段的快速区域卷积神经网络(fast region-based convolutional neural network, FAST-RCNN)[7]、级联区域卷积神

**收稿日期**：2024-06-12；**录用日期**：2024-07-26；**网络出版时间**：2024-08-26 11:10
**网络出版地址**：link.cnki.net/urlid/11.2625.V.20240824.1220.001
**基金项目**：国家自然科学基金(61573183)；安徽省高校科研计划项目(2022AH051584)
* **通信作者**. E-mail：nuaaimage@163.com

*引用格式：陈金林，吴一全，苑玉彬. 无人机视角下目标检测的YOLO系列算法研究进展[J]. 北京航空航天大学学报，2026，52(10)：3509-3534. Chen J L，Wu Y Q，Yuan Y B. Research progress on YOLO series algorithms for target detection from UAV vision[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(10)：3509-3534 (in Chinese).*

经网络(cascade region-based convolutional neural network, Cascade RCNN)[8]、基于区域的全卷积网络(region-based fully convolutional network, R-FCN)[9]等;基于一阶段目标检测算法有YOLO(you only look once)[10]、单次多框检测器(single shot multi-box detector, SSD)[11]、RetinaNet[12]和全卷积单阶段目标检测(fully convolutional one-stage object detection, FCOS)[13]等。在众多的UAV目标检测模型中,YOLO具有版本多、相对准确度高、检测速度快、适应性强、网络结构简单等优点,使其备受关注。

近几年,该研究内容的相关综述较多,但多数仅从YOLO系列算法,或者侧重于UAV视角下目标检测模型角度论述。文献[14]于2022年介绍了YOLOv2～YOLOv4及各种改进,并对比相关改进算法的性能及优缺点。文献[15]于2022年对当时8个版本的YOLO网络结构、损失函数与应用场景进行了梳理。文献[16]于2023年梳理了YOLOv1～YOLOv7算法发展历程及改进研究,综述了典型数据集、评价指标整体框架和发展历程,重点论述了在输入阶段、特征提取阶段及预测阶段3个方面的改进方法,也阐述了YOLO算法一些应用领域。文献[17]于2024年全面综述了当前YOLO的多个不同版本,从YOLO的发展历程、网络结构、改进特点、性能对比等方面进行论述。上述几篇综述文献的主要贡献是对YOLO细致的总结与分析,可以帮助研究人员更清晰、全面地认识YOLO算法及其改进方法,但没有侧重分析UAV视角下YOLO目标检测系列算法的改进与应用。

在UAV视角下目标检测研究方面,文献[18]于2022年全面分析了基于深度学习的UAV视角下目标检测与跟踪方法的研究进展和前景,较详细地分析了基于UAV目标检测的一些数据集。文献[19]于2024年对UAV视角下的目标特点、数据集、目标检测算法及目标检测算法的性能做了全面概括。文献[20]于2024年综述了UAV航拍影像目标检测与语义分割方法的发展历程,对各种网络模型进行分类,分别总结了其改进策略、应用场景、贡献和局限性,同时,梳理了近些年UAV航拍影像的数据集。文献[21]于2023年综述了基于深度学习的UAV航拍视频多目标检测与跟踪算法,总结了该领域的最新进展,包括多目标检测、多目标跟踪,分别阐述了2类算法的原理并分析其优缺点。上述文献的主要贡献是综述了UAV航拍影像目标检测、语义分割、目标跟踪算法,并对后继相关研究的可能方向做出了合理的展望。这些算法也包括了YOLO目标检测算法,但这些综述在UAV视角下YOLO目标检测的改进策略、改进原因、改进优势分析较少。

近些年来,在UAV视角下目标检测的众多算法中,基于YOLO的研究与应用成果较多、发展快,许多UAV视觉目标检测模型是基于YOLO开发的。为应对UAV目标检测应用的行业需求,需对YOLO系列算法进行优化与改进。虽然文献[22]于2020年综述了当时UAV目标识别和检测的YOLO系列算法研究进展,并进行了YOLO应用算法的性能分析。文献[23]于2022年综述了UAV视角下目标检测的YOLOv2～YOLOv4算法。上述算法综述的YOLO版本偏少,仅经过几年的发展,YOLO就出现了多个版本,且在UAV视角下目标检测任务的应用中发挥着重要作用。因此,为进一步研究UAV视角下YOLO目标检测算法,有必要全面总结该领域的研究成果,分析算法的优化与改进特点。

本文对UAV视角下YOLO目标检测改进算法做更系统、更深入的总结。概述10种YOLO目标检测基础版本的发展历程及其改进方法;分别综述UAV视角下YOLO目标检测应用数据集的采集、扩展、增强等;总结UAV目标检测的YOLO算法改进,包括特征提取模块、检测头模块、注意力机制、模型轻量化优化、损失函数等;从UAV视角下YOLO目标检测方法和数据集2个维度阐述了该领域当前存在的有关问题,并对下一步的可能研究趋势进行了展望。

# 1 YOLO目标检测版本的发展历程

自YOLO问世以来,仅经过不到10年的发展,已经出现了16个版本。其中,YOLOv1～YOLOv9与YOLOX在UAV应用领域是较有影响力的目标检测算法,主要体现YOLO的基本设计思想、创新特点、应用研究,图1是10种版本的发展时间线及其改进基础。除了本文列举的版本以外,YOLO还有很多优秀的改进算法,且被应用到各个领域。

## 1.1 YOLO的10种版本

### 1.1.1 YOLO目标检测算法

YOLO是在2015年被提出的一阶段目标检测算法,其主干网络基于GoogLeNet,网络结构如图2所示,其中,$k$为每个网格单元预测的边界框数量,$c$为目标类别数。特征提取模块将输入图像划分为等大的特征网格。检测头模块中每个特征网格预测同一类别的边界框和不同类别的置信度。每个边界框的预测由5个参数组成:目标的置信度分数、网格单元的边界框中心坐标($x$、$y$)、高度和宽度。检测头可选择性地经过非极大值抑制(non-maximum suppression, NMS)来消除多余的边界框,直到找出与真实框损失最小的边界框。

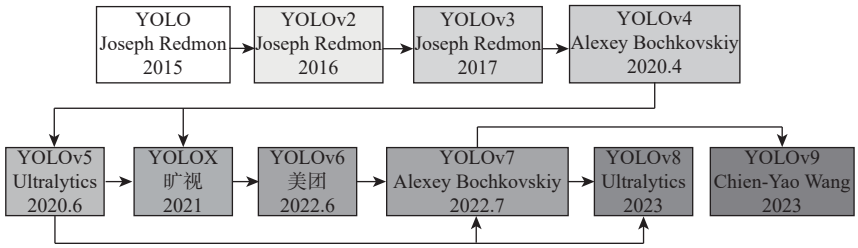



图 1　10 种 YOLO 版本发展时间线

Fig. 1　10 YOLO versions development timelines

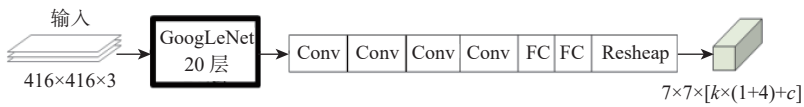



图 2　YOLOv1 检测网络结构

Fig. 2　YOLOv1 detection network structure

### 1.1.2　YOLOv2

YOLOv2[24] 网络结构如图 3 所示, 算法主要改进有 3 个方面: 锚框机制、特征提取、特征融合。YOLOv2 提前设定固定高宽的锚框, 使每个网格单元都有多个维度来预测目标。在训练过程中, 模型学习的是锚框和真实框之间的偏移; 在推理阶段, 模型学习锚框的偏移, 再逆推回最终的真实框预测值。YOLOv2 应用轻量级网络 Darknet-19 替代了 GoogLeNet, 改进了卷积层模块 CBL(convolution + batch normalization + leaky ReLU), 使用步幅为 2 的卷积层替代池化层实现特征图的降采样, 这样可以有效地降低因池化层引起的低层级特征损失。批归一化(batch normalization, BN)层可以使学到的数据达到同分布, 较大地提高训练速度。YOLOv2 将图像细粒度特征提取模块添加到转移层, 使浅层特征与深层特征相融合, 从而提升了模型对小目标的检测能力。

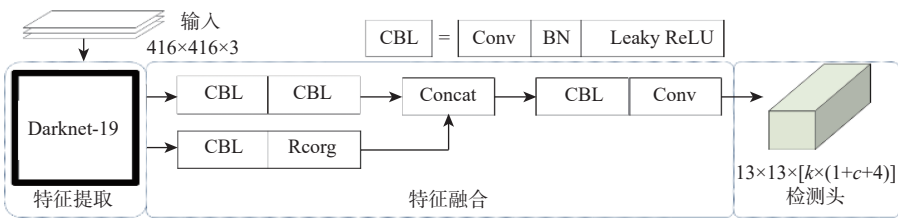



图 3　YOLOv2 检测网络结构

Fig. 3　YOLOv2 detection network structure

### 1.1.3　YOLOv3

YOLOv3[25] 网络结构如图 4 所示, 为提高特征提取能力、加深网络深度, 特征提取模块改进为堆叠多个残差模块的 Darknet53。图中相连的深色框表示自上而下, 浅色框表示自下而上, 后续不再赘述。为实现多尺度目标检测、提升小目标检测的性能, YOLOv3 在特征融合模块中使用了特征金字塔网络 (feature pyramid network, FPN), 如图 5 所示。FPN 强化了浅层网络特征信息, 使浅层特征图局部信息更易被提取到。模型采用了 3 个不同尺度检测大小的目标, 有效提高了多尺度目标检测能力。

### 1.1.4　YOLOv4

YOLOv4[26] 于 2020 年被提出, 其网络结构如图 6 所示。YOLO 发展到此时已经形成了比较完善的模型架构: 特征提取、特征融合、检测头。YOLOv4 的特征提取应用跨阶段部分连接(cross stage partial, CSP)模块[27] 替换了 YOLOv3 的残差模块。CSP 通过跨层连接进行合并, 增强了卷积神经网络的学习能力, 减少了计算量的同时保证了更高

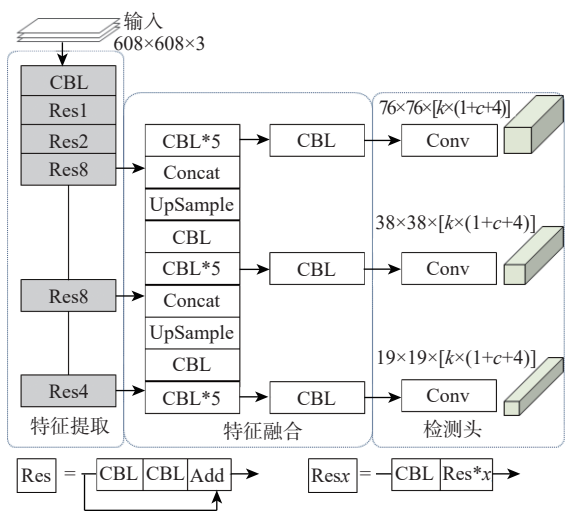



图 4　YOLOv3 检测网络结构

Fig. 4　YOLOv3 detection network structure

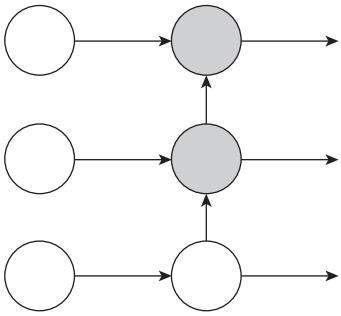


图5 FPN网络结构

Fig. 5 FPN network structure

的检测准确率。YOLOv4在特征融合前新增了空间金字塔池化(spatial pyramid pooling, SPP)网络，如图7所示，运用多尺度的池化拼接丰富了目标特征，使每个特征图能包含多个感受野信息，突出了目标物体的轮廓信息。

在特征融合中，添加了一个信息自浅层向深层融合的特征金字塔路径聚合网络(path aggregation network, PAN)，增强了特征图对不同感受野信息的感知，显著提升了模型的性能，如图8所示。为提高检测召回率，YOLOv4改进了锚框偏移机制，通

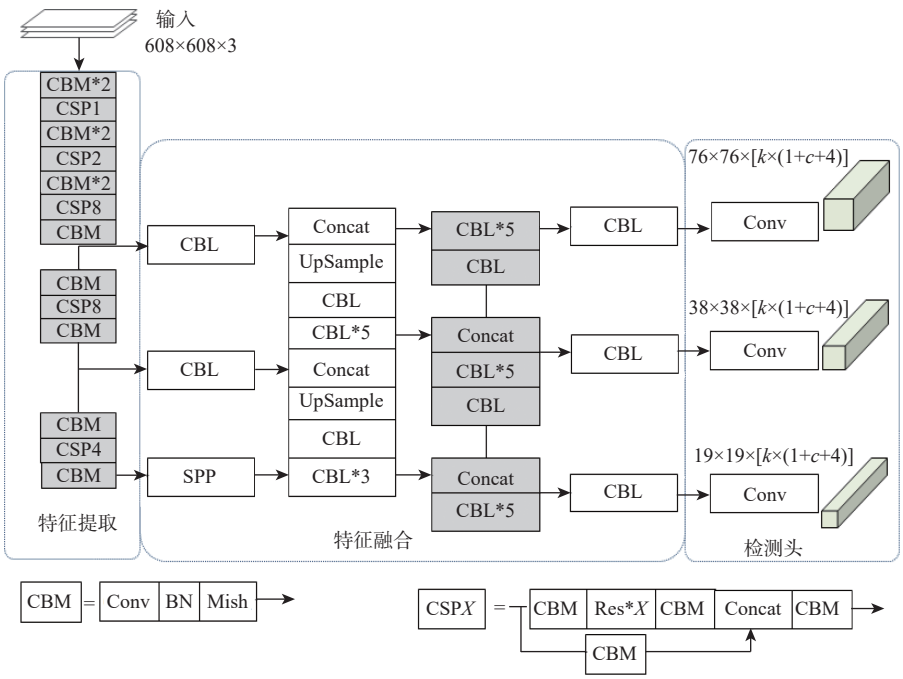



图6 YOLOv4检测网络结构

Fig. 6 YOLOv4 detection network structure

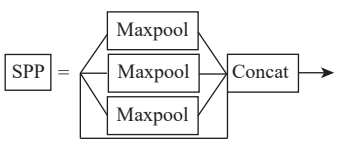



图7 SPP网络结构

Fig. 7 SPP network structure

过中心点偏移数值扩展，使正样本更多，增强了准确性。在训练阶段，对输入的数据进行了Masaic增强处理，丰富了物体的背景与小目标特征信息。

### 1.1.5 YOLOv5

Ultralytics公司于2020年公开发布的YOLOv5网络结构如图9所示。YOLOv5的特征提取模块包括Foucs模块、Conv模块、C3模块、快速空间金字塔池化(spatial pyramid pooling fast, SPPF)模块[28]。Foucs模块对图片进行切片操作，实现特征的无损下采样。采用C3模块替换YOLOv4的CSP模块，解决了大型卷积神经网络框架特征提取中梯度信息重复问题。SPPF模块可以将尺度不固定的特征图转化为统一尺度，经过局部特征与全局特征相融合，丰富了特征图的表达能力，有利于检测复杂多目标图像。在获得数据集的锚框时，YOLOv5采用了自动学习的方式，提高了模型对遮挡物体的检测，最终模型的召回率得到了提高。为满足对目标检测的不同需求，YOLOv5的开发团队为其设计了4种不同尺寸的版本。

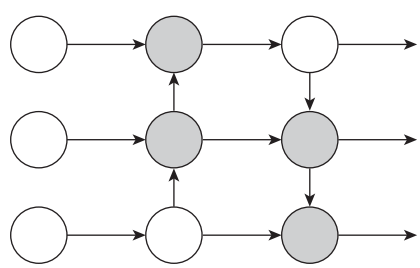


图8 FPN与PAN的融合网络结构

Fig. 8 Fusion network structure of FPN and PAN

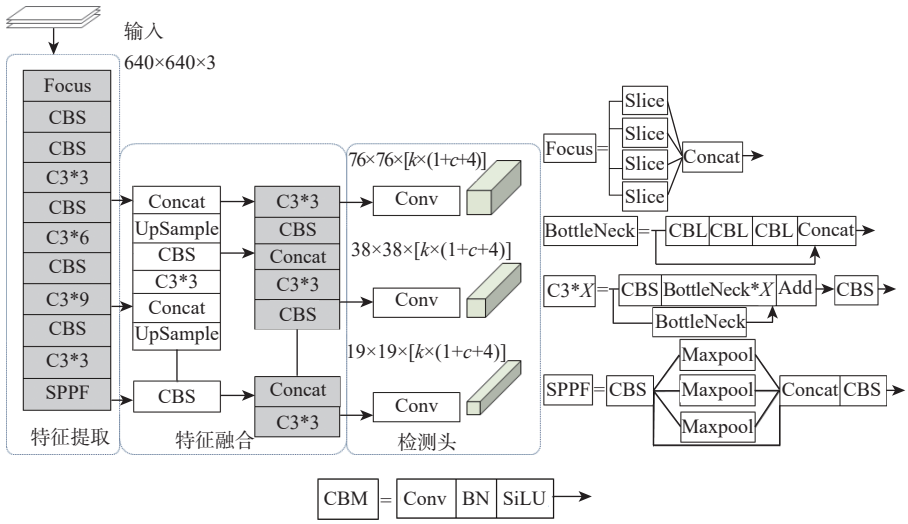



图 9　YOLOv5 检测网络结构

Fig. 9　YOLOv5 detection network structure

### 1.1.6　YOLOX

YOLOX[29] 的网络结构如图 10 所示。较早期的 YOLO 检测头是使用一层卷积同时完成分类和定位 2 个任务的耦合头。YOLOX 的解耦头是将分类和定位分为 2 个不同的卷积来完成, 如图 11 所示。这种做法虽然会增加模型的参数量, 但提高了目标检测的精度, 后期的版本多采用了解耦头结构。为解决因固定锚框导致训练过程中正负样本不均衡的问题, YOLOX 采用无锚框的方法, 直接预测网格偏移量及预测框的高度和宽度。YOLOX 选用简化最优传输分配(simplified optimal transport assignment, SimOTA)作为正负样本匹配策略, 解决了特殊情况下的样本分配不均衡问题, 但该方法增加了额外的训练时间。

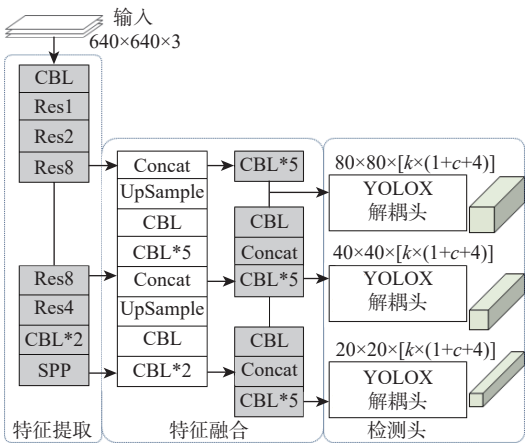



图 10　YOLOX 检测网络结构

Fig. 10　YOLOX detection network structure

### 1.1.7　YOLOv6

YOLOv6[30] 是美团视觉于 2022 年研发的, 其网络结构如图 12 所示, 其中, SConv 表示对称卷积模块。YOLOv6 的特征提取将重参数化模块(reparameterized block, RepBlock)作为小型网络的构建块, 对于大型模型使用跨阶段部分堆叠重参数化(cross stage partial stacked reparameterized, CSPStackRep)模块。为减少计算量并取得更低推理延迟, YOLOv6 采用混合通道策略重新设计了更高效的解耦头结构。为获得更多高质量的正样本, YOLOv6 引入了任务对齐的单阶段目标检测(task-aligned one-stage object detection, TOOD)动态分配正负样本策略, 并结合无锚框进一步提高检测精度。

### 1.1.8　YOLOv7

2022 年 YOLOv4 开发团队提出 YOLOv7[31] 的网络结构, 如图 13 所示, 其中, SPPCSPC 为空间金字塔池化与跨阶段局部网络。在特征提取与特征融合 2 个部分使用高效层聚合网络(efficient layer aggregation network, ELAN)模块, 其中, ELAN_H 表示改进 ELAN。ELAN 模块通过控制最短和最长的梯度路径, 使网络能够学习到更多的特征, 让更深的网络可以更加高效地学习和收敛, 模块结构如图 14 所示。在训练过程中, YOLOv7 的检测头引入了辅助头, 通过增加训练成本提升精度, 但不影

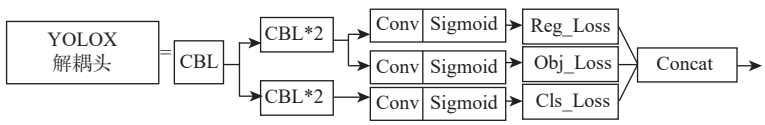



图 11　YOLOX 解耦头

Fig. 11　Decoupled head of YOLOX

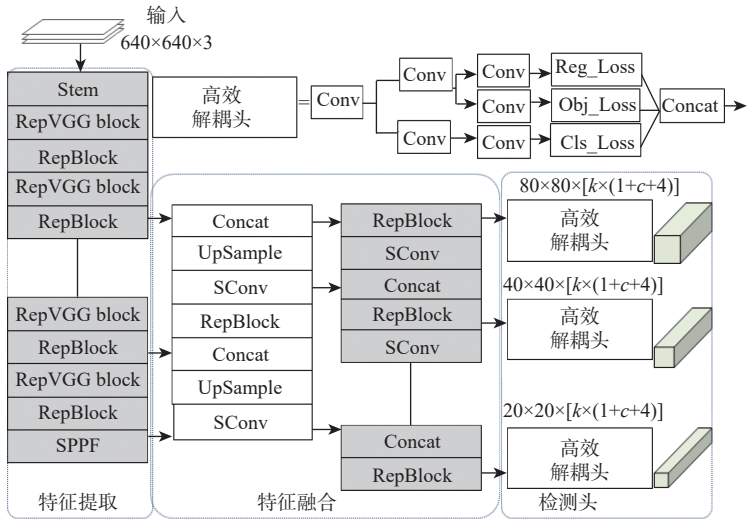



图 12 YOLOv6 检测网络结构

Fig. 12 YOLOv6 detection network structure

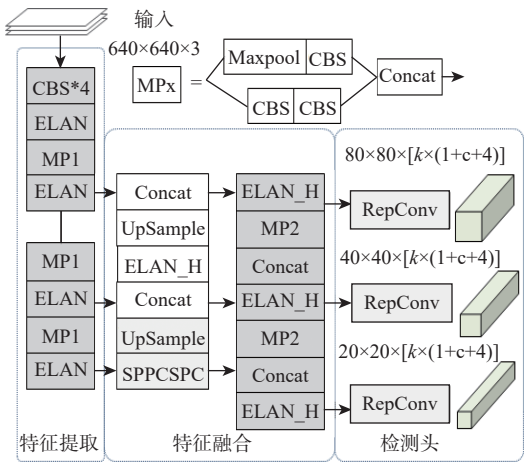



图 13 YOLOv7 检测网络结构

Fig. 13 YOLOv7 detection network structure

响推理的时间，辅助头结构如图 15 所示。YOLOv7 的检测头经过重参数化卷积（reparameterized convolution, RepConv）调整通道数，再使用 1×1 卷积预测坐标、类别、置信度，训练时有 3 个分支相加输出，推理时会将分支的参数重参数化到主分支上，这种重参数化技术在推理阶段将多个计算模块合并为一个模块，提高模型的效率和性能。

### 1.1.9 YOLOv8

Ultralytics 公司 2023 年发布了 YOLOv8，其继承了早期版本优良的改进策略，如无锚框、解耦头部、训练数据增强等，检测网络结构如图 16 所示。YOLOv8 将 YOLOv5 的 C3 模块替换成拥有更多的残差连接的 C2f 模块，提高了检测的精度。YOLOv8

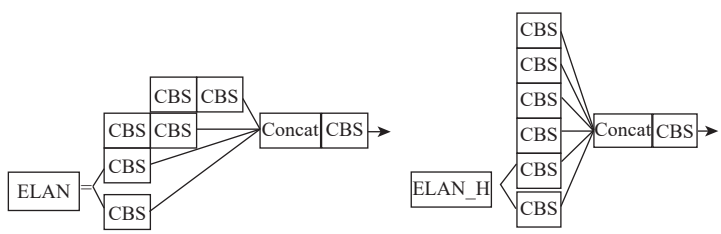



图 14 ELAN 网络结构

Fig. 14 ELAN network structure

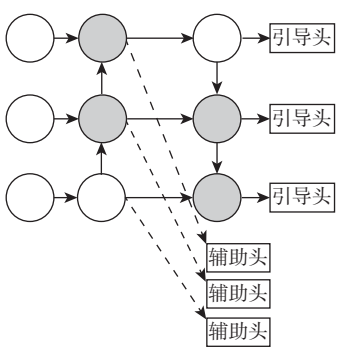



图 15 辅助检测头结构

Fig. 15 Auxiliary detection head structure

改变了交并比 (intersection over union, IOU) 匹配和单边比例的分配方式，运用分类与回归的分数作为加权分数选择正负样本。同时，YOLOv8 提供了更加丰富的超参数和模型结构选项，使其在应用中可以更加方便地进行模型调整和优化。模型开发团队提供了更加丰富的超参数和模型结构选项，使 YOLOv8 在应用中可以更加方便地进行模型调整和优化。

### 1.1.10 YOLOv9

YOLOv9[32] 由中国台湾 Academia、台北科技大

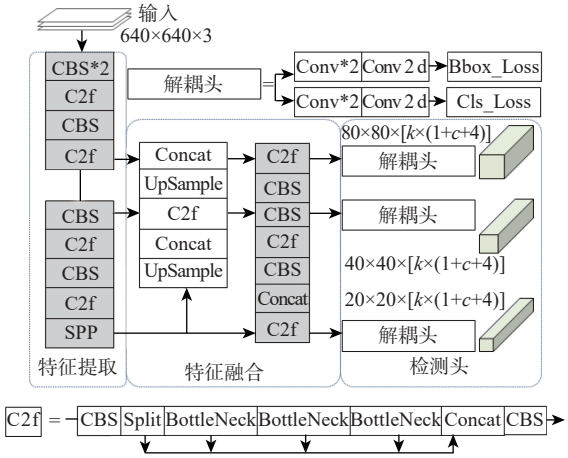



图 16　YOLOv8 检测网络结构

Fig. 16　YOLOv8 detection network structure

学等机构于 2023 联合开发, 网络结构如图 17 所示。研究者在 YOLOv9 中应用可编程梯度信息(programmable gradient information, PGI)学习目标特征。PGI 主要包括 3 个部分: 主分支用于推理的架构, 不需要额外的推理成本; 辅助可逆分支是为了解决网络加深造成的信息瓶颈问题; 多级辅助信息旨在处理深度监督带来的误差累积问题。在 YOLOv9 中, 广义高效层聚合网络(generalized efficient layer aggregation network, GELAN)使用传统的卷积实现了更好的参数利用, 该卷积兼顾了模型复杂度、准确性和推理速度。

除了上述 YOLO 系列算法的网络结构、训练策略、主要改进以外, YOLO 不同的版本在激活函数、损失

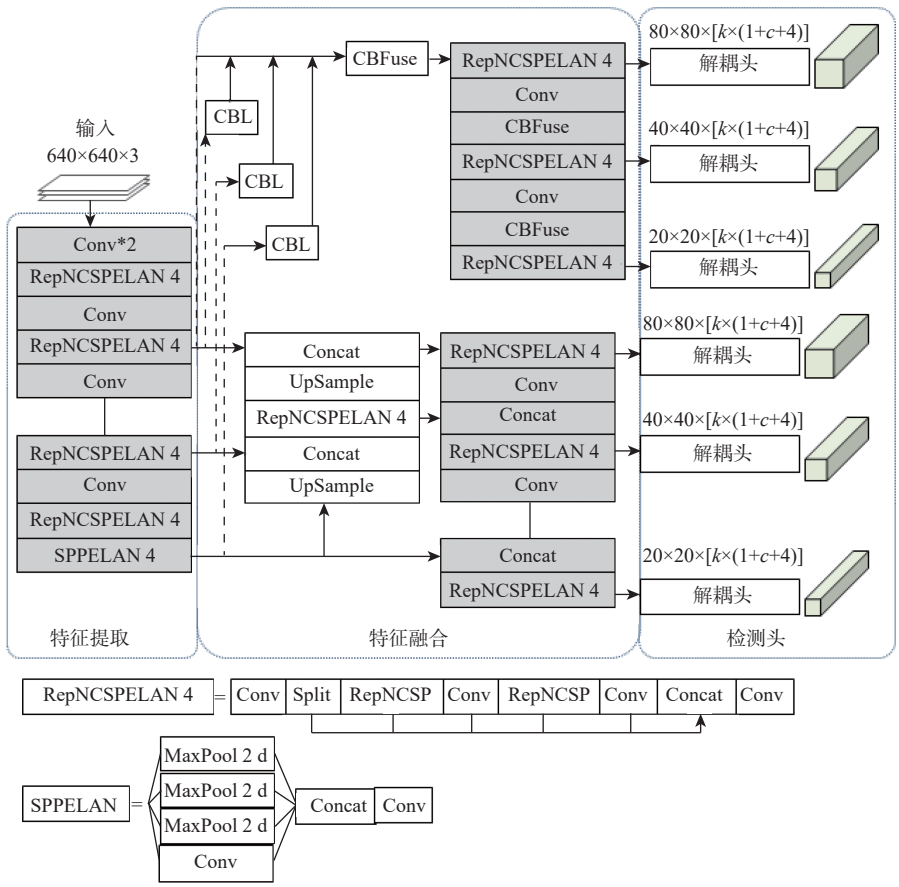



图 17　YOLOv9 检测网络结构

Fig. 17　YOLOv9 detection network structure

函数、Anchor 机制等方面也略有变化, 如表 1 所示。

### 1.2　YOLO 的性能

目前, YOLO 目标检测算法的性能评价指标大体可分为 2 类: 检测能力指标、算法复杂度指标。检测能力指标中有检测准确性指标与检测速度指标, 具体指标项目如表 2 所示, 其中, 样本中被正确识别的正样本数称为真阳性 (true positives, TP), 记为 $n_{TP}$; 样本中被正确识别的负样本数称为真阴性 (true negatives, TN), 记为 $n_{TN}$; 样本中被错误识别为负样本的正样本数称为假阴性 (false negatives, FN), 记为 $n_{FN}$; 样本中错误识别为正样本的负样本数为假阳性 (false positives, FP), 记为 $n_{FP}$。

在模型的训练过程中, 正样本帮助模型学习如何准确地定位和分类目标物体, 而负样本则负责模型学习如何区分目标物体和背景, 增强了模型的鲁棒性和准确性。在性能评价中, 正样本和负样本作为计算目标检测模型准确性的指标基础。正负样本的划分直接影响模型检测能力指标的优劣。

表1 YOLO不同的优化方案

Table 1 YOLO different optimization schemes

| YOLO | 锚框机制 | 正负样本分配 | 检测头 | 激活函数 | 损失函数 | | |
|---|---|---|---|---|---|---|---|
| | | | | | Cls_Loss | Obj_Loss | Bbox_Loss |
| YOLO | 无锚框 | Max IOU | 耦合头 | Leaky ReLU | 均方差 | 均方差 | IOU |
| YOLOv2 | 锚框 | Max IOU | 耦合头 | Leaky ReLU | 均方差 | 均方差 | IOU |
| YOLOv3 | 锚框 | Max IOU | 耦合头 | Leaky ReLU | 2元交叉熵 | 2元交叉熵 | IOU |
| YOLOv4 | 锚框 | Max IOU | 耦合头 | Leaky ReLU、Mish | 2元交叉熵 | 2元交叉熵 | DIOU |
| YOLOv5 | 锚框 | 跨网格搜索 | 耦合头 | Leaky ReLU | 2元交叉熵 | 2元交叉熵 | CIOU |
| YOLOX | 无锚框 | SimOTA | 解耦头 | SiLU | 2元交叉熵 | 2元交叉熵 | CIOU |
| YOLOv6 | 无锚框 | TOOD | 解耦头 | SiLU | 2元交叉熵 | 2元交叉熵 | CIOU |
| YOLOv7 | 无锚框 | 跨网格搜索 | 解耦头 | SiLU | 2元交叉熵 | 2元交叉熵 | CIOU |
| YOLOv8 | 无锚框 | TOOD | 解耦头 | SiLU | 2元交叉熵 | VFL | CIOU |
| YOLOv9 | 无锚框 | TOOD | 解耦头 | SiLU | 2元交叉熵 | VFL | CIOU |

注: DIOU和CIOU分别为距离交并比和完全交并比。

表2 YOLO性能评价指标

Table 2 Performance evaluation indicators of YOLO

| 类别 | 指标 | 公式 |
|---|---|---|
| 准确性 | 召回率$R$ | $n_{\mathrm{TP}}/(n_{\mathrm{TP}}+n_{\mathrm{FP}})$ |
| | 精确率$P$ | $n_{\mathrm{TP}}/(n_{\mathrm{TP}}+n_{\mathrm{FN}})$ |
| | 平均精度$P_k$ | $\int P(R)\mathrm{d}R$ |
| | 平均精度均值 | $\frac{1}{K}\sum_{k=1}^{K}P_k$ |
| 速度 | 帧率 | |

表2中参数量与浮点运算量的计算主要包含卷积层和全连接(fully connected, FC)层,卷积层的参数量$P_{\mathrm{J}}$和浮点运算量$F_{\mathrm{J}}$计算式分别为

$$P_{\mathrm{J}}=\sum_{i=1}^{N}P_i \tag{1}$$

$$P_i=c_{\mathrm{in},i}(h_i w_i c_{\mathrm{out},i}+1) \tag{2}$$

$$F_{\mathrm{J}}=\sum_{i=1}^{N}F_i \tag{3}$$

$$F_i=P_i H_i W_i \tag{4}$$

式中: $c_{\mathrm{in},i}$、$h_i$、$w_i$、$c_{\mathrm{out},i}$分别为第$i$个卷积的输入通道数、核长、核宽、输出通道数; $H_i$和$W_i$分别为输出特征图的长度和宽度; $N$为卷积核个数。

FC层的参数量$P_{\mathrm{Q}}$和浮点运算量$F_{\mathrm{Q}}$计算式分别为

$$P_{\mathrm{Q}}=\sum_{j=1}^{M}(I_j O_j+O_j) \tag{5}$$

$$F_{\mathrm{Q}}=2\sum_{j=1}^{M}(I_j O_j) \tag{6}$$

式中: $I_j$和$O_j$分别为第$j$个FC层输入和输出神经元个数; $M$为FC层个数。

YOLO是由数据驱动的深度学习算法,因此,训练与测试数据集对性能的影响是不可忽略的。表3

表3 YOLO性能对比[31-34]

Table 3 YOLO performance comparison[31-34]

| 版本 | 框架 | 数据集 | 图像尺寸/像素 | 骨干网络 | mAP/% | 帧率/(帧·s$^{-1}$) | 参数量 | 浮点运算量/次 |
|---|---|---|---|---|---|---|---|---|
| YOLO | Darknet | VOC | 448 | GoogLeNet | 63.4 | 45 | $2.629\times10^8$ | $3.37\times10^{10}$ |
| YOLOv2 | Darknet | VOC | 416 | Darknet24 | 76.8 | 67 | $4.87\times10^7$ | $3.06\times10^{10}$ |
| YOLOv3 | Darknet | COCO | 608 | Darknet53 | 33.0 | 78 | $6.3\times10^7$ | $1.18\times10^{11}$ |
| YOLOv4 | Darknet | COCO | 608 | CSPDarknet53 | 43.5 | 65 | $2.76\times10^7$ | $5.20\times10^{10}$ |
| YOLOv5s | Pytorch | COCO | 640 | CSPDarknet | 37.4 | 95 | $7.2\times10^6$ | $1.65\times10^{10}$ |
| YOLOXs | Pytorch | COCO | 640 | CSPDarknet | 40.5 | 156 | $9.0\times10^6$ | $2.68\times10^{10}$ |
| YOLOv6 | Pytorch | COCO | 640 | EfficientRep | 43.1 | 109 | $1.72\times10^7$ | $4.42\times10^{10}$ |
| YOLOv7s | Pytorch | COCO | 640 | ELANblock | 51.4 | 161 | $6.03\times10^6$ | $1.32\times10^{10}$ |
| YOLOv8s | Pytorch | COCO | 640 | CSPDarknet | 53.0 | 101 | $1.12\times10^7$ | $2.86\times10^{10}$ |
| YOLOv9 | Pytorch | COCO | 640 | ELANblock | 56.0 | | $7.1\times10^6$ | $2.64\times10^{10}$ |
| SSD | Pytorch | COCO | | ResNet101+FPN | 33.2 | 35.5 | | |
| ConterNet | Pytorch | COCO | | Hourglass104+FPN | 45.1 | 1.42 | | |
| RetinaNet | Pytorch | COCO | | ResNet101+FPN | 39.1 | 5.4 | | |
| FCOS | Pytorch | COCO | | ResNet101+FPN | 43.2 | 60 | | |

给出了 10 种 YOLO 版本的部分性能指标[31-34]。对比可以发现, 经过几代版本的改进, YOLO 目标检测算法的准确性、推理速度、参数量等性能都得到了不同程度的优化; 在同等的 Pytorch 框架下, 相比较于同类的一阶段目标检测算法, 较高版本的 YOLO 目标检测准确性、速度是比较有优势的。

## 2　UAV 视角下 YOLO 目标检测数据集与预处理

UAV 视角下的图像与视频因采集设备、采集方式、现场环境等因素的影响, 呈现出多样性的特点, 使 UAV 视角下 YOLO 目标检测模型的研究与应用更依赖于特定应用数据集与任务需求。研究人员在运用 YOLO 算法开发 UAV 视角下目标检测应用模型的过程中, 为提高模型对目标的检测能力, 需花费较多的精力收集、处理相关数据集以适应应用需求。

### 2.1　目标检测数据集

将 UAV 视角下目标检测模型部署到复杂且未知的现实场景时不仅要考虑模型的运算速度、检测准确性, 还要兼顾模型对目标物体的泛化能力。数据集中图像质量、大小、数量直接影响到模型检测的运算速度、准确性与泛化能力, 因此, 一个优秀深度学习模型数据集中的图像应该具有代表性、多样性、完整性和可靠性[33]。在基于 UAV 视角下的 YOLO 目标检测模型的研究与应用中, 常使用一些公共数据集, 如视觉对象类别(visual object classes, VOC)数据集[34]、微软常见对象上下文(microsoft common objects in context, COCO)数据集[35]、航空图像目标检测数据集(dataset for object detection in aerial images, DOTA)[36]、无人机视觉(vision meets drone, VisDrone)数据集[37]、光学遥感图像目标检测(object detection in optical remote sensing images, DIOR)数据集[38]、无人机基准目标检测与跟踪(unmanned aerial vehicle benchmark object detection and tracking, UAVDT)数据集、无人机 123 目标跟踪(unmanned aerial vehicle 123, UAV123)数据集、奥尔巴尼大学检测与跟踪(university at Albany detection and tracking, UA-DETRAC)数据集, 表 4 给出了这些公共数据集的性质及特点。DIOR、DOTA 是航空视角下的数据集, 其中包含大量小目标、丰富的场景, 可以提高检测模型的性能与泛化能力。然而, 这些数据集图像的拍摄视角是垂直地面的, 且图像背景色彩多以灰色、黑色、深绿色为主。因此, 这些数据集与 UAV 视角下的场景有所区别, 不利于 UAV 视角下 YOLO 检测模型的实验论证。

上述公共数据集可以用于模型预训练、测试、验证, 提高模型的泛化能力[39]。与自然拍摄视角不同, UAV 航拍图像由于视野更宽, 可能包含更多的目标物体, 且物体的尺寸比自然图像中小得多[38]。UAV 视角下的常用数据集包括:

表 4　常用公共数据集

Table 4　Common public datasets

| 类别 | 大小/(像素×像素) | 类别 | 数量 | 拍摄视角 |
|---|---|---|---|---|
| VOC | 300×500 | 20 | 33 000 | 自然视角 |
| COCO | 600×800 | 80 | 330 000 | 自然视角 |
| DIOR | 800×800 | 20 | 23 463 | 航空视角 |
| DOTA | 800×4 000 | 15 | 11 268 | 航空视角 |
| VisDrone | 2 000×1 500 | 12 | 10 209 | UAV视角 |
| UAVDT | 1 080×540 | 8 | 83 000 | UAV视角 |
| UAV123 | 1 280×720 | 12 | 110 000 | UAV视角 |
| UA-DETRAC | 960×540 | 4 | 140 000 | UAV视角 |
| Staford Drone | 1 417×2 019 | 10 | 19 000 | UAV视角 |
| AU-AIR | 1 920×1 080 | 9 | 32 000 | UAV视角 |

1) VisDrone 数据集, 包含 288 个视频片段, 由 261 908 帧视频和 10 209 幅静态图像组成。

2) UAVDT 数据集, 包含 100 个视频序列, 80 000 多帧, 大约 2 700 辆汽车。

3) UAV123 数据集, 由阿卜杜拉国王科技大学于 2016 年发布, 共包含 123 个视频序列和 110 000 多帧图像。

4) UA-DETRAC 数据集, 有 24 个不同场景的视频, 超过 140 000 帧, 包括不同天气环境下的车辆图像。

5) Staford Drone 数据集, 通过 UAV 收集斯坦福大学校园的航拍图像数据集, 含约 60 个航拍视频, 包含 10 种目标类型。

6) Aarhus 大学航空影像与遥感 (Aarhus university-aerial imagery and remote sensing, AU-AIR) 数据集包含 8 段视频, 共计 3 282 个提取帧。

这些公共数据集中的图像都是从 UAV 搭载的相机拍摄的视频中按帧采集的, 视频的数量少, 因此, 很多的图像差异性不大, 场景相似的图像较多, 这些相似图像不利于目标检测模型的应用与实验。

表 5 给出了部分 YOLOv5～YOLOv9 的 4 个不同版本检测模型在上述 4 种 UAV 视角下数据集上的性能指标。可以看出, 版本越高, 检测性能平均精度均值(mean average precision, mAP)也越高。每次实验数据从 UAV 视角下数据集中随机选取, 其中, 训练集数据是 4 000 张图像, 验证集为 1 000 张。

实验配制如表 6 所示, 本文后续实验也使用了相同的配制。除公共数据集外, 各个行业的具体需求催生了很多特定应用数据集。表 7 中展示了近

表5 YOLO检测模型在UAV视角下的实验性能

Table 5 Experimental performance of the YOLO detection model under UAV vision

| 算法 | 数据集 | mAP/% |
|---|---|---|
| YOLOv5s | VisDrone2019 | 34.5 |
| YOLOv5s | UA-DETRAC | 40.3 |
| YOLOv5s | UAVDT | 57.8 |
| YOLOv7s | VisDrone2019 | 38.9 |
| YOLOv7s | UA-DETRAC | 45.9 |
| YOLOv7s | UAVDT | 60.1 |
| YOLOv8s | VisDrone2019 | 39.8 |
| YOLOv8s | UA-DETRAC | 45.1 |
| YOLOv8s | UAVDT | 65.1 |
| YOLOv9c | VisDrone2019 | 45.3 |
| YOLOv9c | UA-DETRAC | 54.6 |
| YOLOv9c | UAVDT | 66.3 |

表6 实验运行环境

Table 6 Experimental environment configuration

| 云服务器 | GPU | 内存/GB | Python | CUDA |
|---|---|---|---|---|
| https://www.mistgpu.com/ | RTX3080 Ti 12 GB | 30 | 3.8 | 11.3 |

些年来基于UAV视角下YOLO目标检测模型使用的训练与测试集[38,40-50]。

## 2.2 UAV视角下目标检测数据集的扩充

UAV目标检测应用数据集中的图像多是从视频中采集,且过滤了特征差异较小的连续帧。这就导致数据集的数据量小、多样性不足、目标数量较少、背景多样性差。为提升目标检测模型的检测准确性,增强目标检测模型的泛化能力,可以应用各种扩充方法增加数据量。扩充数据集增加了数据量,但场景相似程度也相对较高,而YOLO系列算法将对于图像的特征提取与特征融合相结合,有利于相似图像特征的区分,因此,许多数据扩展方法的出现提高了YOLO目标检测算法的检测准确性。

扩充数据集的方式主要有2种:在已有的数据集上对图像进行各种变换处理;使用不同的拍摄设备以多种形式收集图像与视频。

### 2.2.1 基于图像变换的扩充

Andrew等[51]在研究奶牛检测的YOLOv2算法过程中,对UAV获取的图像进行随机合成,使扩充后的数据集包含11 384个合成和非合成图像。

Luo等[52]将1 978张UAV航拍图像应用随机旋转、镜像翻转、增加噪声与颜色变换方法扩充了4倍,再使用YOLOv3算法进行车辆检测模型训练与测试。Han等[40]使用UAV搭载的数码相机拍摄200张尺寸较大的钢结构表面裂纹图像构成原始数据集。由于钢结构表面裂缝区域在原始图像中的比例相对较小,很难对裂缝识别,因此,利用超像素分割方法对UAV图像进行预处理,将原始数据集扩充为4 000幅图像用于YOLOv3算法的训练与测试。

鲍文霞等[42]研究UAV自主着陆系统时,运用二维码图像作为着陆识别标志,标志识别算法使用YOLOv5m。为增加数据集的泛化性并提高检测准确性,在拍摄过程中,通过改变同一二维码的拍摄光照、摆放场景来丰富数据集。

### 2.2.2 基于数据收集的扩充

现阶段UAV可搭载的拍摄设备有数码相机、红外相机、多光谱相机、全景相机等,如图18所示。成像技术有可见光成像、红外成像、多光谱成像、全景摄像头成像等,其中,可见光成像是UAV最常使用的。

表7 UAV视角下的应用数据集

Table 7 Application datasets of UAV vision

| 文献 | 图像大小/(像素×像素) | 类别 | 数量 | 成像特点 | UAV拍摄高度/m |
|---|---|---|---|---|---|
| [41] | 5 472×3 648 | 地面控制点 | 1 745 | 可见光 | 40～220 |
| [40] | 5 472×3 078 | 钢结构裂纹 | 200 | 可见光 | 10～25 |
| [42] | 640×640 | 快速响应码 | 3 260 | 可见光 | <12 |
| [42] | 800×600 | 小麦 | 300 | 可见光 | <3 |
| [43] | 4 000×3 000 | 输电线路 | 2 600 | 可见光 | 50～260 |
| [44] | 1 330×1 000 | 枯死树 | 3 084 | 可见光 | 75～100 |
| [45] | 500×500 | 松材线虫病 | 4 208 | 可见光 | <525 |
| [46] | 1 152×864 | 绝缘子 | 860 | 可见光 | <100 |
| [38] | 640×512 | 行人、车辆 | 2 898 | 红外热图 | 60～130 |
| [47] | 1 920×1 680 | 行人 | 8 416 | LED光源 | 15～65 |
| [48] | 960×5 184 | 松材线虫病 | 4 500 | 多光谱图像 | 100～350 |
| [49] | 800×800 | 行人 | 7 092 | 鱼眼摄像头 | <6 |
| [50] | 2 448×2 048 | 伪装目标 | 2 215 | 偏振成像 | 10 |



(a) 数码相机　(b) 红外相机　(c) 多光谱相机　(d) 全景相机

图 18　UAV 可搭载的拍摄设备

Fig. 18　UAV capable shooting equipment

不同的拍摄设备丰富了 UAV 视角下图像的表现形式, 使图像具有多种模式。因此, 目标检测模型要适应不同模态的图像数据。为适应不同类别、不同模态的 UAV 图像数据集, 通常 YOLO 检测模型会有所改进。

1) 红外成像。与传统光学图像相比, 红外热图像具有空间分辨率低、信噪比小、纹理特征少的特点, 可以在白天和晚上拍摄。在夜间、极端天气条件下, 因为光线太暗、物体遮挡、飞行高度太高, 相机很难捕捉到目标物体, 因此, 基于可见光条件下采集的数据集被应用于训练目标检测模型, 通常性能表现不佳。

Hinzmann 等[53] 将可见光图与热红外图像结合训练改进的 YOLOv3, 最终降低了 UAV 检测模型对行人的错误识别率, 提高夜间检测性能。

UAV 在夜间或在恶劣天气飞行中检测目标物体时, 易产生误检、漏检。为解决该问题, Wang 等[44] 使用红外图像集训练 YOLOv5s 改进版, 提升了模型对夜间目标的检测能力。

Mokayed 等[54] 使用蛋白化技术对 UAV 航拍的晴朗天气条件下车辆图像进行变换, 用以模拟雪、雨、雾等天气条件下的车辆。结合蛋白化车辆图像与原图像训练 YOLOv5 与 YOLOv8, 最终使 2 个模型在恶劣天气检测车辆的准确性分别提高了 30% 和 17%。

2) LED 光源成像。为提高夜间检测性能, 也可以使用特殊光源来增强图像中目标物体的特征信息。硅基金黄光 LED 光源只含有长波长的红光和黄光成分, 在雨、雾、沙尘、雾霾等恶劣天气条件下具有较好的穿透能力, 这使物体即使在恶劣的天气条件下也能保持较高的清晰度和亮度。因此, Zhu 等[47] 在 UAV 上搭载硅基金黄光 LED 光源, 扩充了原始训练集, 设计了 UAV 视角下的 YOLOv3 行人检测模型, 实验验证了该模型在夜间环境下检测行人具有显著优势。

3) 多光谱成像。单一的基于可见光图像或基于多光谱图像的识别方法, 会受到同物异谱、同谱异物现象的影响, 导致各种误检和漏检, 影响识别精度。多光谱相机可以获取多光谱图像, 包括更多的光谱波段, 如绿色、红色、近红外和红外边缘图像信息。多光谱成像将特征光谱信息划分为多个光谱波段, 可以从多个光谱通道获取更多信息, 揭示隐藏在较窄光谱范围内的特征。Qin 等[48] 在 UAV 上搭载了多光谱相机和可见光相机拍摄松木树林, 通过将可见光图像与多光谱图像相结合构成数据集。由于健康树和感染树在红色波长区域表现出最显著的差异, 因此, 多光谱成像更容易区分目标物体与背景。基于该数据集提出了一种松树萎蔫病 YOLOv5 检测模型, 能够排除树种、植被、光秃秃的地面和阴影的干扰, 准确地检测到受感染的松树。

4) 全景摄像头成像。UAV 在着陆时要避免着陆路径上的潜在危险, 而全景摄像头可为 UAV 在选择着陆点时提供比较安全的视角, 但全景摄像头采集的鱼眼图存在着陆点目标变形与扭曲, 增加了目标检测难度。Thuan 等[49] 改进了目标检测 YOLOv7 算法, 在 COCO 数据集进行预训练, 在 UAV 全景数据集上对模型微调, 微调前对全景图像做旋转处理增加训练图像数量, 最终提高了模型对着陆点检测的性能。

5) 偏振光成像。偏振成像不仅可以获得目标的强度信息, 而且可以获取附加的偏振信息[50]。偏振可以自适应增强图像中目标的对比度, 同时, 背景杂波会导致成像性能严重下降, 利用偏振对图像中的背景信息进行抑制, 可以有效地保留目标信息。当目标与周围环境相似时, 基于可见光图像的检测方法难以区分。为提高模型对伪装目标的检测能力, 沈英等[55] 在 YOLOX 的特征提取部分添加 3D 卷积, 提高了偏振特征提取能力, 引入特征增强模块和跨层级特征聚合网络, 将含有 UAV 拍摄的伪装目标偏振图像作为模型的输入, 最终改进的 YOLOX 在识别伪装目标检测中的 mAP 提高了 11.0%。

上述 UAV 视角下数据扩充方法的主要特点与优势如表 8[40-42,45-50,52-56] 所示。

### 2.3　UAV 视角下目标检测训练图像的预处理

虽然 UAV 视角下目标检测模型的数据集有很多, 在训练模型过程中, 这些数据集的大小仍然是有限的。第 2.2 节的方法可以增加数据量, 但还不足以解决问题。在模型训练前期, 对现有数据进行预处理成为提高模型检测能力的主要方法之一。训练中数据预处理能生成与现有数据同分布的训练输入数据, 用以解决数据量少、数据场景单一的问题, 提高模型泛化能力。现有的图像预处理方法有针对单幅图像的, 也可以利用多幅图像样本生成新的图像以丰富目标与场景性能, 其预处理作用如表 9[41,53,57-61] 所示。

表8 UAV视角下的数据扩充方法特性

Table 8 Advantages of data augmentation methods from UAV vision

| 文献 | 扩充方法 | 优势 | 作用 |
|---|---|---|---|
| [52] | 随机合成 | 增加数据 | 提高泛化能力 |
| [41] | 随机旋转、镜像翻转、增加噪声与颜色变换方法 | 增加数据、改变目标形态、增加背景 | 提高泛化能力与检测准确性 |
| [40] | 超像素分割方法 | 突出目标 | 提高检测准确性 |
| [42] | 改变拍摄光照、地点 | 丰富背景 | 提高泛化能力与检测准确性 |
| [45-50,53-56] | 改变成像模式 | 丰富目标背景 | 提高泛化能力;减少漏检、误检 |
| [48] | 多光谱成像 | 增加数据、突出目标 | 减少漏检、误检 |

表9 图像预处理方法特性

Table 9 Characteristics of image preprocessing methods

| 文献 | 方法 | 作用 |
|---|---|---|
| [41,53,57] | 数据增强 | 增强模型鲁棒性、提高准确性和召回率 |
| [58] | SAHI | 提高小目标的可见性 |
| [59] | 生成对抗网络 | 突出目标特征 |
| [60] | 图像模式变换 | 提高模型检测准确率、召回率 |
| [61] | 图像压缩 | 提升检测速度 |

在YOLOv4～YOLOv8版本的训练过程中都有数据增强模块,数据增强在早期UAV目标检测的应用已经出现过。

2019年,Andrew等[51]随机对训练图像进行不同增强处理,包括水平和垂直翻转、裁剪、仿射变换、高斯模糊、平均或中位数模糊、噪声添加、背景变化、对比度变化等任意组合,再使用扩充数据集训练YOLOv2。该方法是在小幅度改变数据特征的情况下,允许有限的数据产生更多的目标前景与背景,使模型对UAV视角下图像具有更强的鲁棒性。赵爽等[56]为解决UAV航拍图像下的多尺度车辆检测问题,将大尺寸航拍图像切分成若干有重叠区域的小图像块输入YOLOv2进行训练与检测,使检测模型相比原YOLOv2检测算法的平均精度(average precision, AP)提高了约8%。

Wang等[58]应用改进的生成对抗网络来融合可见光和红外图像,增强了UAV图像中行人目标信息,在此基础上训练YOLOv3行人检测模型,模型的召回率与准确性都有所提高。

为提高网络检测速度,可以使用压缩后的图像,但压缩后的图像必然会损失一部分目标物体的特征,降低识别的准确性,尤其是针对图像中的小目标物体。Coletta等[61]在ImageNet数据集上训练YOLOv4目标检测模型任务时,提出了以模型检测精度为变量的图像JPEG压缩级别函数。模型根据性能的要求预先设定好对应的JPEG压缩级别,不仅可以满足精度的需求,也可以实现检测速度的提升。在该理论基础上,设计了UAV视角下YOLOv4野生动物快速检测模型。

鲍文霞等[42]分析了UAV视角下的麦穗分布稠密、重叠现象严重、背景信息复杂等特点,基于Retinex算法[57]提出一种图像反射光增强方法,弱化入射图像,可以增强图像中的目标区域特征,提高了YOLOv5改进算法检测物体的准确性和召回率。

Wang等[36]利用切片辅助超推理(slicing aided hyper inference, SAHI)算法对训练数据集进行预处理,使小目标具有相对较大的可见性。同时,使用数据增强策略增加小物体的训练权值,提高模型对小物体信息的感知能力。Tian等[35]在研究UAV视角下小目标检测问题过程中,基于VisDrone数据集,比较了SAHI对于YOLOv5与YOLOX目标检测模型性能提升的作用。说明了SAHI可以大幅提高目标检测模型的性能,也指出了运用该方法,YOLOv5目标检测模型性能优于YOLOX。

Chen等[60]用自适应裁剪算法对高分辨率训练数据集进行增强,用以提高目标检测模型对小物体的检测能力。HSV色彩空间的亮度在区分饱和度或色调信息方面能起到重要的作用,同时,HSV图像模式比RGB提供了更多的可感知信息。Wang等[36]在目标检测模型训练过程中将图像由RGB转成HSV,有效提高了目标物体的检测准确率和召回率。改进后的YOLOX目标检测模型在VisDrone数据集的检测性能指标mAP0.5提高了8%。

### 2.4 模型的训练与推理

YOLO是应用数据训练出来的模型,再应用于实际任务进行推理,因此,训练策略可以提高模型的性能。深度学习模型通常假设训练数据和测试数据来自同一领域,并遵循相同的分布。然而,这种假设在现实场景中并不总是成立,如自动驾驶、UAV跟踪识别任务中。UAV视角下目标检测模型的实际应用研究中,训练数据和测试数据之间普遍存在着场景分布不同,这种差异性必将影响模型的检测能力[62]。YOLO系列算法中很多模型的训练与推理策略可以有效地解决这些问题,如迁移学习、多尺度学习、多分辨率训练。

将学习到的预训练模型泛化到未知的目标领域, 用预训练模型中初始化良好的权重作为基础, 在特定的训练集中对其微调, 可以提高目标检测能力。陈蕊等[63]先在仿真数据集上训练 YOLOv5 目标检测模型, 随后利用真实数据集对模型进行微调训练, 使 UAV 视角下目标检测的性能提升了 2.7%。鲍文霞等[42]在麦穗检测模型训练过程采用迁移学习, 先用田间采集小麦图像数据集对基于 Transformer 预测头的 YOLO(Transformer prediction head-YOLO, TPH-YOLO)模型进行训练得到预训练模型, 再冻结 TPH-YOLO 骨干网络预训练模型参数, 使用 UAV 采集的小麦图像数据集对模型进行参数更新和优化训练。该种训练方法在保证运算速度的前提下提高了检测精度, TPH-YOLO 的 mAP 比 YOLOv5 提高 4.1%。

使用 UAV 摄像机获取的图像分辨率通常较大, 且不同摄像机获得的图像分辨率也不同。因此, 可使用一些大型公共数据集进行目标检测模型的预训练, 这些数据一般比 UAV 视角下图像分辨率小, 例如, COCO 数据集的分辨率是 640 像素×360 像素。如果数据集中包含的图像大小是单一的, 则模型提取的特征大小也相应固定。为增强目标检测模型的鲁棒性, 需要增加多尺度数据集训练来弥补[64]。

Rančić等[65]认为在训练中使用多尺度图像能够使模型更容易学习到目标物体的差异特征, 保证模型能够学习到目标物体的不同尺度特征。魏湧明等[66]先采用 ImageNet 数据集对 Darknet-19 进行预训练, 再采用低分辨率的 UAV 航拍图像分类数据集, 微调 Darknet-19, 使网络调整每层权重以适应高分辨率输入。经过多尺度训练的 YOLOv2 网络对不同尺度的检测数据集具有很强的适应性。Keles 等[67]也提出了相似的结论, 并提出了 UAV 视角下鹿种群识别模型 YOLOv4-tiny, 为能够达到较好的检测效果, 建议将训练时的图像分辨率设为 512 像素×512 像素与 608 像素×608 像素。

UAV 视角下的目标检测模型多数是面向视频的, 视频的每一帧是一个静态图像, 视频则是由这些静态图像按照一定的帧顺序播放而成, 因此, 目标检测时可以逐帧推理或间隔帧推理。目标物体可能会出现在连续几帧图像内, 因此, 视频内目标检测与定位可以应用多帧图像的相近信息。Afifi 等[68]根据当前帧中包含物体的位置, 提出下一帧中的裁剪位置。将彼此靠近的框合并在一起, 形成一个单独的裁剪。通过仅在包含对象的区域位置进行推理, 可减少对图像背景部分的计算推理。该做法不仅可以提高检测的准确性, 而且可以提高推理速度。Yavariabdi 等[69]设计了一种多尺度 UAV 检测卷积神经网络 YOLOv3-tiny, 该架构应用 Inception Black[70]从输入视频中提取目标物体的局部和全局特征。每 6 帧检测一次, 将检测到的物体作用于中间帧, 提高了视频检测的能力。

## 3　UAV 视角下 YOLO 网络结构优化

YOLO 的版本多、可扩展性、易移植性使其更适用于 UAV 视角下目标检测的不同任务需求。在 UAV 视角下目标检测各个应用领域, 为适应不同场景与行业需求, 会选择不同的 YOLO 版本, 并基于该 YOLO 进行网络结构优化来提高 UAV 目标检测模型的性能。

### 3.1　特征提取优化策略

YOLO 目标检测算法的性能极大地依赖于主干网络对图像中目标物体特征的提取与融合能力, 因此, 在研究 UAV 视角下 YOLO 目标检测模型的过程中, 主干网络的优化成为主要改进方向。UAV 视角下图像中小目标特征信息少, 大多数现有算法的小目标识别与定位精度都不高[71]。

#### 3.1.1　特征提取模块优化

为深化目标特征的提取, Liu 等[72]运用时空交互模块替代 YOLOv5 的 C3 模块, 并在 PAN 结构中引入递归门控卷积, 通过浅层信息挖掘和特征融合来提高小目标检测性能。改进后的 UAV 视角下 YOLO 目标检测模型在 VisDrone 数据集上的检测 mAP 较原来提高了 3.4%。陈范凯等[73]利用 C2f 模块替换 YOLOv5 的所有 C3 模块。优化后的目标检测模型 YOLOv5s_UAV 在增加少量参数的条件下, 较大地提升了模型的识别能力。该目标检测模型相较于 YOLOv5s 在数据集 VisDrone 上的 mAP 提高了 10.7%。冒国韬等[74]提出一种基于多尺度特征提取模块的 UAV 图像目标检测模型 多尺度自适应 YOLO(multi-scale adaptive YOLO, MSA-YOLO)。多尺度特征提取模块如图 19 所示, 其中, 4 个卷积运用不同大小的卷积核, 在 YOLOv5 的特征提取模块中应用多尺度卷积方式来获取不同尺度的特征信息。该目标检测模型在 VisDrone 数据集上的 mAP 较 YOLOv5 提升了 3.8%。

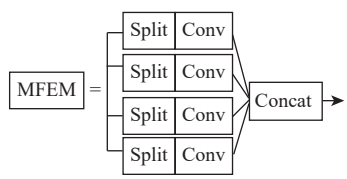



图 19　多尺度特征提取模块

Fig. 19　Multi-scale feature extraction module

注意力机制的核心思想是模仿人类视觉系统的选择性感知机制，对输入图像的特征权重进行动态调整，使模型集中在图像中最重要的区域，并抑制特征不相关部分[75]。

在YOLO目标检测模型中增加注意力模块必定增加模型的大小，但可以提高模型的目标检测能力。Yang等[76]改进了YOLOv4中的跨阶段局部X（cross stage partial X, CSPX）模块，在CSPX模块中增加了通道注意力机制压缩与激励网络 (squeeze and excitation networks, SENet) 与空间注意力机制空间变换网络 (spatial Transformer networks, STNet) 相结合的双注意力网络 (dual attention network, DANet) 结构。改进后的YOLOv4在识别UAV视角下输电线路绝缘子、鸟巢、间隔器和冲击锤较原YOLOv4模型的mAP值提高了2%。

为提高模型对相似目标的区分能力，减少物体的检测错误，李校林等[43]提出并行选择性注意力机制 (parallel selective attention mechanism, PSAM)，动态融合了特征的空间信息和通道信息。将这些模块增加到YOLOv5的特征提取中，虽然模型检测速度下降了15.0%，但检测精度提高了11.5%。Thuan等[49]为提高目标检测模型对图像特征的跨通道信息、位置信息的感知能力，将卷积块注意力模块 (convolutional block attention module, CBAM) 和坐标注意力 (coordiate attention, CA) 机制相结合，再应用于YOLOv5的特征提取模块；为减少模型参数量、提高检测速度，运用GhostNet[77]改进了C3，改进后的YOLOv5目标检测模型在自制UAV视角下松材线虫树数据集上的mAP提高到0.98%，与原YOLOv5模型相比，检测速度提高了1倍。奉志强等[78]为提高UAV目标检测模型对小目标特征的提取能力，提出一种融合了空间注意力与通道注意力的空间通道注意力模块 (space channel attention module, SCAM)，该模块相比于CBAM参数量更少，更关注小目标空间维度特征。在YOLOv5的C3后增加SCAM，改进后目标检测模型在VisDrone数据集上的mAP增加了12.1%，参数量只增加了约$1\times10^4$。

Wang等[44]将高效的动态稀疏关注机制BiFormer[79]集成到YOLOv8骨干网络模型的C2f之前，提高了模型对特征图中关键信息的关注，进一步优化了检测性能。在VisDrone数据上训练后，模型的mAP增加了2.2%。

### 3.1.2 特征融合模块优化

张宏鸣等[80]在YOLOv3的特征融合模块后加入了STNet用于增强玉米幼苗的空间位置特征，提高UAV视角下目标检测模型对玉米幼苗的检测能力。

林文树等[81]将跨阶段部分网络（cross stage partial network, CSPNet）应用于YOLOv4模型的特征融合模块，且在计算卷积前加入SENet以增加感受野信息改变网络的深度、宽度，最终实现模型的缩放，提高了检测精度。改进的YOLOv4-CSP模型在UAV视角下松树木数据集上检测病虫害树的正确率为97.5%，相比于YOLOv4的平均检测正确率提高3.4%，检测速度提高了1.5倍。

Zhang等[82]为突出风电叶片表面长条缺陷的特征提取能力，在YOLOv5特征提取模型中增加了CBAM，改进的目标检测模型在自制的UAV视角下风电叶片数据集上的mAP提高了9.5%，参数量增加了17.3%。

Wang等[36]在YOLOX的PAN模块增加了超轻量级子空间注意力模块 (ultra-lightweight subspace attention module, ULSAM)，增强了不同感受野特征图之间信息的相关性，抑制背景信息，有效提升了Neck模块对小目标特征的融合，增强了网络局部感知目标的能力。

Zhao等[83]在YOLOv7的最大池化卷积（max pooling convolution, MPConv）层后增加了无参数的注意力机制 (parameter-free attention mechanism, PFAM)，突出浅层网络中的目标特征，弱化不相关信息。改进的YOLOv7目标检测模型被应用于UAV视角下海面人员、船只和漂浮物检测时，模型能有效地检测出海面上的小型目标。Zhao等[84]为提升UAV视角下小目标物体的检测能力，在YOLOv7的PAN模块中加入CBAM注意力机制，并引入Swin Transform模块[85]，改进后的目标检测模型在VisDrone数据集上的mAP提高了9.5%，但参数量增加了11.6%。

### 3.1.3 特征提取与特征融合的混合优化

丛玉华等[86]改进了YOLOv5的主干网络，在CSP模块后增加了CBAM注意力机制，同时，在Neck中使用加权双向FPN，如图20所示，同一尺寸的特征图上增加了跳跃连接，使特征融合时能够获得更多的信息。上采样和下采样之后均引入CBAM模块增加模型的收敛速率，同时，增强了不同目标的特征。韩俊等[87]将混合域注意力模块增加到YOLOv5的主干网络中，该模块关注特征图通道间的关系与空间位置信息。改进后的YOLOv5目标检测模型在VisDrone数据集上训练后，相比原模型的mAP提高了13.8%。李子豪等[88]为提高UAV对密集小目标检测的准确性，提出了一种自适应协同注意力机制 (adaptive coordinate attention module, ACAM) 嵌入到YOLOv5的主干网络中。该模块将输入特征通过并联的空间和通道注意力

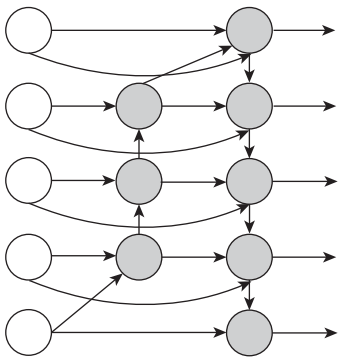


图 20　加权双向 FPN 结构

Fig. 20　Weighted bidirectional FPN structure

模块, 再将 2 个注意力权重自适应加权得到最终的注意力权重, 增强了特征的细节信息和语义信息。在 VisDrone 数据集上训练的改进 YOLOv5 与原模型相比, mAP 提升 1.5%。

表 10 归纳了在 YOLO 目标检测中增加注意力机制后的改进算法较改进前目标检测精度提升情况[36,44,48,76,78,80,82-84,86-89]。表中数据说明, 增加注意力机制可以有效提高 UAV 视角下目标检测的准确性。

在特征提取与特征融合模块中增加注意力机制或应用注意力机制替换原模型中某一个模块是 2 种重要的 YOLO 检测模型改进方法。为验证这些方法对于 YOLO 检测模型性能的改进能力, 共做了 12 次的实验, 实验中应用 CBAM 改进原 YOLO 检测模型, 实验数据集是从 VisDrone2019 中随机选取的 5 000 张图像, 训练与测试数据量为 4∶1, 结果可参考表 11。由表 11 的 mAP 指标可知, 在 3 种 YOLO 检测模型中增加注意力机制要优于替换原模型某个模块, 但增加注意力机制对于模型的大小会有一定的提高。因此, 本节中很多文献应用注意力机制优化 YOLO 检测模型时, 要结合自身的应用需求与 UAV 视角下数据集特点选择适当的被替换模块、增加位置、优化改进注意力机制或多种方法的融合。

### 3.2　YOLO 检测头结构优化策略

YOLO 采用多尺度检测可以提高不同大小目标的检测能力, 使用 UAV 航拍收集到的数据集通常包含许多非常小的目标, 使现有的网络检测非常困难。通常可以增加 1 个或 2 个预测头, 使其具有更小的感受野。因此, YOLO 的 FPN-PAN 多头结构可以缓解对象尺度变化的影响, 降低小目标检测网络的优化难度, 减少严重的物体尺度方差带来的负面影响。文献 [37,90-91] 在 UAV 视角下目标检测模型采用 4 个检测输出的检测头结构如图 21 所示; 文献 [81] 提出了有 5 个检测输出的检测头结构。

**表 10　UAV 视角下增加注意力机制的 YOLO 算法性能提升度**

**Table 10　Performance improvement of YOLO algorithm with an added attention mechanism from UAV vision**

| 算法 | 数据集 | AP提升量/% | 注意力机制 | 位置 |
|---|---|---|---|---|
| YOLOv4[76] | 高压输电线数据集 | 2 | DANet | Backbone |
| YOLOv4[82] | 虫害数据集 | 3.4 | SENet | Backbone |
| YOLOv5[78] | VisDrone | 11.5 | PSAM | Backbone |
| YOLOv8[44] | VisDrone | 2.2 | BiFormer | Backbone |
| YOLOv3[80] | 幼苗数据集 | 0.1 | SAM | Neck |
| YOLO3[89] | 棕榈树数据集 | 1.06 | CBAM | Neck |
| YOLOv5[82] | 风叶缺陷 | 4 | CBAM | Neck |
| YOLOX[36] | VisDrone | 5 | ULSAM | Neck |
| YOLOv7[83] | ODv2 | 7 | PFAM | Neck |
| YOLOv7[84] | VisDrone | 9.5 | CBAM | Neck |
| YOLOv5s[86] | VisDrone | 2 | CBAM | Backbone、Neck |
| YOLOv5[87] | VisDrone | 7.9 | SENet | Backbone、Nec |
| YOLOv5[78] | VisDrone | 4 | SCAM | Backbone、Neck |
| YOLOv5[48] | 松木数据集 | 5 | CBAM | Backbone、Neck |
| YOLOv5[88] | VisDrone | 1.5 | ACAM | Backbone、Neck |

**表 11　CBAM 对 YOLO 算法性能影响**

**Table 11　Impact of CBAM on the performance of YOLO algorithm**

| 算法 | 位置 | 置入方式 | mAP/% |
|---|---|---|---|
| YOLOv5s | Backbone+Neck | 替换C3 | 31.6 |
| YOLOv5s | Backbone | 替换C3 | 32.4 |
| YOLOv5s | Neck | 替换C3 | 33.8 |
| YOLOv5s | Backbone | 增加 | 37.6 |
| YOLOv7s | Backbone+Neck | 替换ELAN | 32.7 |
| YOLOv7s | Backbone | 替换ELAN | 37.9 |
| YOLOv7s | Neck | 替换ELAN | 39.0 |
| YOLOv7s | Backbone | 增加 | 39.8 |
| YOLOv8s | Backbone+Neck | 替换C2f | 33.8 |
| YOLOv8s | Backbone | 替换C2f | 34.9 |
| YOLOv8s | Neck | 替换C2f | 38.6 |
| YOLOv8s | Backbone | 增加 | 40.1 |

多个检测头会增加模型的计算成本、增加模型尺寸, 不利于模型的部署。如果目标检测模型在不受小目标的影响下, 可以减少检测头, 使模型尺寸变小[89]。Wu 等[92] 设计了轻型 YOLOv4-tiny 用于 UAV 在高空观测火焰和烟雾场景。模型采用了 13 像素×13 像素和 26 像素×26 像素这 2 个尺度 FPN-PAN 双头结构, 如图 22 所示。

自 YOLOX 版本后, 检测头多采用解耦头, 可能会导致 2 个任务之间的不对称, 即分类得分最高的样本和位置得分最高的样本, 其定位分数可能相差较大, 导致目标检测中潜在的不准确[33]。

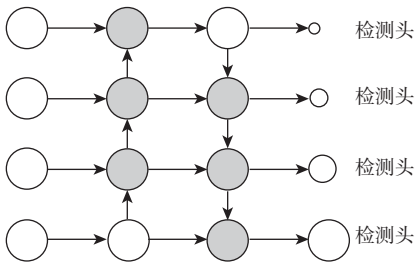



图21 FPN-PAN 4头结构

Fig. 21 Four headed structure of FPN-PAN

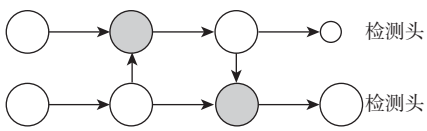



图22 FPN-PAN 双头结构

Fig. 22 Double headed structure of FPN-PAN

吴明杰等[93]改进了YOLOX的解耦头，直接输出目标边界框的位置信息及目标类别的概率分布。应用该解耦头改进YOLOv5，在VisDrone数据集上训练后，mAP提高了8.8%。陈旭等[94]提出一个新的解耦头模块替换YOLOv5的检测耦合头，新的解耦头可分阶段地执行不同的任务，第1阶段完成回归任务，第2阶段借助跨阶段卷积模块协助完成分类任务，防止因检测头特征共用引起回归与分类任务的混淆，提高了细类别的分类精度。

在UAV视角下，图像中目标长宽比与检测模型中的边界框长宽比可能达不到有效近似，从而使得目标检测模型精度下降。为解决该问题，Zhou等[95]设计了基于位置编码与像素重排的轻量化检测头PixSF-Head，如图23所示，其中，IFCOUS为特征图的FCOUS结构逆运算。以PixSF-Head为基础重新设计了YOLOX的辅助头与引导头，其中，引导头是一个蒸馏层，使用辅助头中的标签分配结果替换引导头的标签分配结果，以提高引导头整体的表达能力，而辅助头只在训练过程中执行，没有推理成本。改进后的YOLOX参数量增加了$6\times10^5$，模型在VisDrone数据集上训练后的mAP较原版本提高了9.8%。

Zhao等[96]设计了一种新的重参数化对齐(reparameterization align, RepA)解耦头，通过分类和回归之间信息连接来对齐2个任务，可以有效缓解2个分支任务的不对齐，最终获得更高的检测精度。改进后的YOLOX-s在COCO数据集上的mAP比原版本提高了2.3%，参数量只增加了$6\times10^4$。

为解决UAV航拍图像中交通目标检测时存在的水平包围框与目标真实轮廓贴合度较差、目标的水平包围框重叠度高导致相互抑制的问题，曲优等[97]提出一种基于锚框变换的单阶段旋转目标检测模型，改进YOLOv3的检测头，应用2次回归将水平锚框变换为旋转锚框，表示旋转检测框是5个参数张量，即$(x, y, w, h, \theta)$。在原有的边界框上增加一个角度参数。

### 3.3 UAV视角下YOLO网络结构尺寸优化

为提高YOLO运算速度，减小模型的尺寸，通常可使用网络结构量化、网络剪枝、替换卷积、知识蒸馏和轻量级网络等。

#### 3.3.1 YOLO网络结构的量化

网络结构量化是指减少模型参数的代码长度。Wang等[98]提出一种低比特量化方法来构建高效的单级检测模型，该方法是基于单边直方图的激活量化方法，通过最小化均方误差量化模型改进了YOLOv5s，将该模型参数量化为8位时。在COCO数据集进行测试，平均精度值几乎没有下降，改进的模型计算速度显著加快，量化后模型的尺寸从$5.76\times10^7$降低到$1.44\times10^7$。

#### 3.3.2 基于BN层的网络剪枝

网络剪枝是指对网络结构进行剪枝，其主要目的是调整模型的一些属性，生成不同尺度的模型，以满足不同推理速度与部署的需要。YOLO多个版本卷积后使用了BN层，如下所示：

$$\hat{x}_i=\frac{x_i-\mu}{\sqrt{\sigma^2+\varepsilon}}\gamma+\beta \tag{7}$$

式中：$x_i$为输入样本；$\mu$为样本均值；$\sigma^2$为样本方差，$\varepsilon>0$为一个极小项，$\gamma$和$\beta$为需要学习的参数，$\hat{x}_i$为输出样本。

BN层虽然起到了积极的作用，但增加了网络层数，占用了大量的计算时间。针对BN层缺点，Liu等[99]提出了一种简单有效的通道剪枝方法。直接采用BN层尺度因子$\gamma$作为剪枝指标，如果一个通道包含了对目标分类很重要的信息，其对应的$\gamma$系数受到损失函数的刺激而变大；如果一个通道包含与分类无关的信息，则在损失函数的影响下，其系数$\gamma$会变小。因此，根据剪枝率对具有较小$\gamma$的卷积通道进行相应数量的剪枝，最终得到一个紧凑的网络。

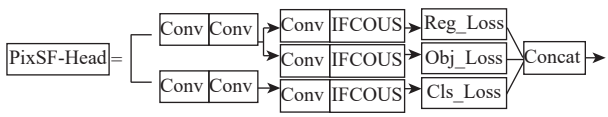



图23 PixSF-Head结构

Fig. 23 PixSF-Head structure

Shao 等[100]运用这种网络剪枝方法改进了 YOLOv3 模型, 获得更紧凑、参数更少的目标检测模型, 在此基础上设计了航空红外 YOLOv3(aerial infrared YOLOv3, AIR-YOLOv3)目标检测算法。在自制的红外航拍数据集上训练后, AIR-YOLOv3 与 YOLOv3 相比, mAP 仅下降了 1.7%, 但体积压缩接近 35.5%, 计算量减少了约 67%。

Qin 等[89]在运用 YOLOv3 算法检测 UAV 视角下单一目标棕榈树的过程中, 为提高检测速度, 减小模型尺寸, 也使用上述剪枝网络改进了 YOLOv3 模型。

Zhang 等[82]在检测 UAV 视角下风电叶片缺陷的研究过程中, 为降低模型的复杂性和参数量, 应用同样的 BN 尺度因子剪枝网络, 改进的 YOLOv5 检测模型推理计算少, 检测速度快。上述层剪枝与通道剪枝方法可以减少参数量, 但经多次剪枝后, 剪掉的层与通道可能也包含有用的特征信息, 影响检测准确性。因此, 侯鑫等[101]在训练中将迭代稀疏层剪枝与迭代稀疏通道剪枝相结合对模型进行剪枝处理, 剪枝后再通过微调来补偿剪枝带来的精度损失, 最终使模型的参数量更少、精度损失更小。

解决 BN 层缺点另一种有效的方法是 Xia 等[90]提出的卷积层和 BN 层融合, 即将卷积层和 BN 层融合为一个卷积层。如下所示:

$$\hat{x} = wx_i + b \tag{8}$$

$$y_i = \frac{\hat{x}_i - \mu}{\sqrt{\sigma^2 + \varepsilon}}\gamma + \beta = w_{\text{new}}x_i + b_{\text{new}} \tag{9}$$

$$w_{\text{new}} = \frac{w\gamma}{\sqrt{\text{Var}[\hat{x}] + b + \varepsilon}} \tag{10}$$

$$b_{\text{new}} = \frac{b - E[\hat{x}]}{\sqrt{\text{Var}[\hat{x}] + b + \varepsilon}} + \beta \tag{11}$$

式中: $w$为原始卷积权重; $b$为原始卷积层的偏置参数; $E[\hat{x}]$和 $\text{Var}[\hat{x}]$分别为均值和方差; $w_{\text{new}}$和$b_{\text{new}}$分别为新卷积层的权重和偏置参数。

融合后的网络结构可以删除 BN 层, 并使用新的权值和偏置进行卷积运算, 从而获得与原来相同的批归一化效果。这种将卷积层和 BN 层融合的方法在 YOLOX、YOLOv7 中也都有应用。

### 3.3.3　基于卷积的网络剪枝

深度学习模型大量地使用了卷积运算, 卷积运算增加了模型的尺寸与运算速度, 选择不同的卷积也可以改变 YOLO 检测模型的尺寸。每个卷积核对应于一个通道, 可以通过拆分空间维度和通道深度的相关性, 减少卷积计算所需要的参数量[102]。

为减少目标检测模型的参数量, 彭继慎等[103]使用深度可分离卷积(depthwise separable convolution, DSConv)改进了 YOLOv4 的 CBL 模块, 改进的目标检测模型 D-YOLOv4 检测速度提高了约 15%, 模型大小减少约 7.3%。

Wu 等[92]在研究 UAV 视角下火焰和烟雾区域检测任务过程中, 为减少算法的参数量和计算量, 使用 DSConv 替换 YOLOv4 特征融合模块中的标准卷积, 改进后的 YOLOv4 算法在自制火焰和烟雾数据集上较原算法的 mAP 增加了 3.14%, 参数量约为原算法的 10.1%。

Sharma 等[104]使用八度卷积替换了 YOLOv4 的标准卷积, 且在特征提取模块中加入了 CBAM 模块, 改进的目标检测模型提高了复杂背景下交通标志的检测速度、准确率。在自制的 UAV 视角下交通标志数据集上训练后, 模型检测的平均精度较 YOLOv4 提高了 3.5%, 检测速度提高到 12.5 帧/s。

陈卫彪等[105]使用 DSConv 替代 YOLOv5 标准卷积, 虽然网络层数增加了, 但整个网络的体积及参数量却降低了。改进的 YOLOv5 目标检测模型在 VisDrone 数据集上训练后, 模型的大小只有原来的 12.5%。

陈旭等[94]运用混合残差空洞卷积模块优化了 YOLOv5 骨干网络结构, 该卷积模块与普通卷积相比, 具有较少的计算量与较大的特征感受野。在 VisDrone 数据集上训练后的改进 YOLOv5 目标检测模型与原模型比较, 检测平均精度提升了 1.4%, 推理速度提升了 21.4%。

DSConv 在运算过程中会损失一部分特征信息, 从而降低模型的准确性, 而分组混洗卷积(group shuffle convolution, GSConv)结合了标准卷积和 DSConv 的优点。因此, Saydirasulovich 等[106]在 YOLOv8 模型中采用了 GSconv 与 DSConv 并行处理, 该方法没有完全切断通道间的连接, 而是选择在很大程度上保留这些连接, 从而保证了模型的检测准确性。在自定义的 UAV 视角下野火烟雾数据集训练后模型对野火烟雾目标检测 AP 提高了 4.3%。

### 3.3.4　基于知识蒸馏的轻量化

网络结构量化、网络剪枝、卷积替换对于模型的检测精度有一定的影响, 因此, 张可等[107]采用知识蒸馏对经过剪枝后的 YOLOv4 网络进行微调, 如图 24 所示。使用知识蒸馏可以进一步实现 1.6% 的精度增益, 这表明直接训练学生网络时, 学生网络不可避免地会学习到部分冗余特征, 而利用教师网络指导训练可以在一定程度上起到强迫作用, 从而使学生网络学习到更多的重要知识。引入知识蒸馏后的 YOLOv4 模型有助于损失的降低, 加快网

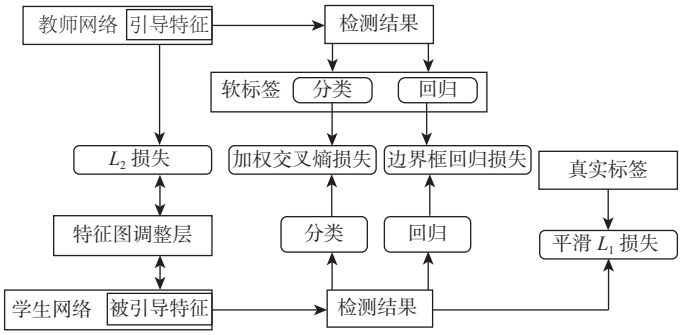



图 24 知识蒸馏架构

Fig. 24 Knowledge distillation structure

络收敛速度。改进后的YOLOv4模型在UAV视角下防振锤缺陷检测mAP提高了165.6%,视频检测速度由204帧/s提高到了217帧/s。

3.3.5 基于轻量级网络的优化

YOLO的骨干网络是算法核心部分,其对检测模型的检测速度、模型体积影响很大,因此,轻量级的骨干网络可以减小YOLO模型的体积,提高检测速度。

姜尚洁等[45]在研究UAV视角下玉米幼苗实时检测模型时,先对玉米幼苗图像进行增强处理;再使用MobileNetV3[108]改进YOLOv3的特征提取模块;检测头使用2个尺度检测玉米幼苗,改进后的目标检测模型特征增强YOLO(feature enhancement YOLO, FE-YOLO)在自制的玉米幼苗数据集上的计算量和推理时间较原YOLOv3模型分别提高了7.91%和33.76%。

罗潇等[46]将UAV拍摄的输电高压线的绝缘子分为正常与缺陷2类。考虑到目标检测任务中类别较少,选用ResNet18作为主干网络改进YOLOv3,能够避免网络出现过拟合,同时,对特征融合模块的特征金字塔执行上采样操作,将其与主干网络中的不同尺度的残差块进行融合,实现绝缘子检测的实时与快速处理。

为提取更多的目标有效特征,提高检测的精度与速度,黄丽明等[109]应用MobleNetV2[110]替换了YOLOv4的CSPDarkNet53,其中,模型的卷积使用DSConv。优化后的模型参数更少、结构尺寸变小,改进的YOLOv4在UAV视角下松木数据集上训练后,虽然检测精度增加不大,但模型的参数量仅为原模型的20%,模型检测速度提高了29.1%。

丛玉华等[86]为提升检测能力与检测速度,使用CSPDarkNet53-tiny替换YOLOv4的CSPDarkNet53,如图25所示。检测头使用通道缩减解耦头结构,改进的YOLOv4在VisDrone数据集上训练后,模型的推理时间较原模型减少了8 ms, mAP提升了7%。表12给出了UAV视角下一些YOLO目标检测




图 25 CSPDarkNet53-tiny网络结构

Fig. 25 CSPDarkNet53-tiny network structure

**表 12 网络结构尺寸优化与作用**

**Table 12 Optimization and role of network structure size**

| 文献 | 方法 | 方法作用 |
|---|---|---|
| [98] | 参数量化 | 减小模型大小 |
| [101-103] | 剪枝BN层 | 剪枝层、剪枝通道大小 |
| [105-108] | 替换卷积 | 剪枝卷积参数量 |
| [107] | 知识蒸馏 | 提高推理速度 |
| [110-114] | 替换小型网络结构 | 减小模型大小、提高推理速度 |

模型的网络结构优化方法与作用[98,101-103,105-108,110-114]。

甄然等[111]使用单个ConvNeXt Block结构[112]替换YOLOv7主干网络的ELAN模块,提出一种更加轻量化的主干网络,改进后的主干网络既节省计算资源,又提升了模型检测精度。

### 3.4 UAV视角下YOLO目标检测模型的损失函数

在UAV拍摄的图像中,很多目标尺寸较小,这些小目标在图像中所占面积比例远小于背景,而YOLO新的版本会侧重于对小目标检测能力的提高,使YOLO成为UAV视角下目标检测模型的主要选择。小目标物体特征信息少导致了数据集中正、负样本的不平衡,因此,损失函数成为UAV视角下YOLO目标检测模型的主要改进之一。

Yu等[113]为增强YOLOv4在复杂背景下检测裂缝的能力,利用焦损失(focal loss, FL)作为置信度损失函数。FL是一个动态缩放的交叉熵损失,通过动态缩放因子,可以自适应地降低训练过程中易区分样本的权重,从而聚焦难区分的样本,可以解决模型中目标与背景分类不平衡和正负样本分配不平衡的问题。FL求解如下:

$$\varphi_{\mathrm{FL}}=\begin{cases}-\alpha(1-y')^{\eta}\ln y' & y=1\\ -(1-\alpha)y'^{\eta}\ln(1-y') & y=0\end{cases} \tag{12}$$

式中:$y=1$为正样本,$y=0$为负样本;$y'$为激活函数

的输出; $\alpha$因子用于平衡正负样本比例不均匀; $\eta$因子的作用是减少易分类样本的损失, 使损失函数更加关注难分类和错误分类的样本。

引入 FL 替换损失函数后, 改进 YOLOv4 的 mAP 比原模型提高了 6.6%。

陈范凯等[73] 在研究 UAV 视角下绝缘子检测问题中, 由于绝缘子目标尺寸较小, 导致数据集中正、负样本的不平衡, 采用 FL 代替 YOLOv4 置信度损失中的标准交叉熵函数, 将类别损失函数替换为平衡交叉熵损失函数, 真实框回归损失函数保持不变。改进后模型的绝缘子检测平均准确率有了明显提高。

为提高 UAV 遥感图像的小目标检测性能, 张上等[114] 在式 (11) 的基础上进行了改进, 得到分类得分函数变焦损失 (varifocal loss, VFL):

$$\varphi_{\mathrm{VFL}}=\begin{cases}-q[q\ln p+(1-q)\ln(1-q)] & q>0\\ -\alpha p^{\eta}\ln(1-p) & q=0\end{cases} \quad (13)$$

式中: $p$为目标分数的分类预测值; $q$为分类条件。

若为目标类, 则将正样本类的$q$ 值设置为预测框和真实框之间的 IOU 值, 否则设置为 0; 若为背景类, 则将$q$值设置为 0。通过$p^{\eta}$缩放因子对负样本进行了处理, 突出正样本的贡献。将 VFL 应用于 YOLOv5 的分类损失函数, 在 VisDrone 数据集上训练后模型检测 mAP 提高了 1.4%。

在目标检测中, IOU 是预测框和真实框之间的 IOU, 其决定了预测框的选取, 因此, IOU 对于检测准确性起到了非常重要的作用。

Xia 等[90] 应用广义交并比 (generalized intersection over union, GIOU) 对 YOLOv3 的损失函数做了改进, 在 COCO 数据集上训练改进后的 YOLOv3 较原模型的 mAP 提高了约 2.5%。

谢椿辉等[115] 采用压缩卷积层与通道数改进解耦头替换了 YOLOv5 耦合头, 在提升精度的同时, 尽量减少模块带来的推理延时, 同时, 使用 Alpha-IoU[116] 预测框位置的损失, 增强在数据较少及存在数据噪声的情况下模型检测效果的鲁棒性。改进的 YOLO 算法在 VisDrone 数据集上训练后, 浮点运算量增加了约 13%, 推理时间减少了 0.7 ms。当检测框与真实框不相交时, GIOU 仍然可以计算梯度, 梯度不会为零, 因此, GIOU 在一定程度上克服了 IOU 作为损失函数的缺点。

吴明杰等[93] 为解决 YOLOv5 中完整交并比损失 (complete intersection over union loss, CIOU Loss) 计算回归结果不准确的问题, 使用聚焦式高效交并比 (focal and efficient intersection over union, Focal-EIOU) 损失函数, 将预测框的宽高拆分, 分别与最小外界框的宽高作差值运算, 从而提高模型对小型目标的检测精度。使用 Focal-EIOU 的改进检测模型与原 YOLOv5 相比, 虽然在浮点运算量的大小上有小幅度增加, 但在 VisDrone 数据集上的检测精度提高了 6.1%。

甄然等[111] 用 SIOU (SCYLLA-IoU)[117] 替换 YOLOv7 模型中的 CIOU, SIOU 加入角度向量来描述纵横比相同时预测框的坐标损失, 在训练阶段实现了更快的收敛, 且在推理方面具有更好的性能。改进的 YOLOv7 检测模型在 UAV 图像数据集中训练后, 相较于原模型的检测精度变化较小, 但参数量减少了 23.6%, 运算速度降为原来一半。

表 13 总结了这些方法的特点与作用[90,93,111,113-114,116], 可为后续 YOLO 目标检测模型的改进研究提供有价值的参考。

**表 13　损失函数的优化方法特点**

**Table 13　Characteristics of optimization methods for loss functions**

| 文献 | 解决问题 | 改进方法 | 作用 | 缺点 |
|---|---|---|---|---|
| [113] | 正负样本分配不平衡 | FL | 自适应地区分样本的权重 | 增加运算量 |
| [114] | 提高小目标检测能力 | VFL | 突出正样本的贡献 | 增加运算量 |
| [116] | 提升检测速度 | Alpha-IOU | 减少梯度为−1的风险, 增加收敛速度 | 增加运算量 |
| [90] | 提升检测精度 | GIOU | 减少梯度为0的风险 | 增加数据量 |
| [111] | 提升检测速度 | SIOU | 加入角度向量来描述纵横比, 增加收敛速度 | 精度没有得到有效提高 |
| [93] | 提高小目标检测能力 | Focal-EIOU | 拆分预测框的宽高 | 降低了推理速度 |

### 3.5　UAV 视角下 YOLO 目标检测模型的性能评价

模型训练过程中, 主要参数包括训练轮次、优化器中的学习率动量参数、权值衰减 (weight decay, WD)、初始学习率 (initial learning rate, ILR); 及实验使用的设备参数。表 14 给出了上述一部分 UAV 视角下 YOLO 目标检测模型实验性能[34-36,38,49,74,76,78,82,84,94-95,101,103,106,113]。可以看出, 在公共数据集上测试的目标检测准确性不是很高, 但在特定的应用数据集上对单个目标检测 mAP 平均约为 94%, 这说明现阶段的 YOLO 应用于 UAV 视角下目标检测模型中的目标检测准确性相对较好。

表14 UAV视角下YOLO目标检测模型的实验性能

Table 14 Experimental performance of YOLO object detection model from UAV vision

| 模型 | 数据集 | mAP/% | mAP0.5/% | 参数量 | epoch | 动量 | WD | ILR | 设备 | 帧率/(帧·s$^{-1}$) |
|---|---|---|---|---|---|---|---|---|---|---|
| YOLOv4[38] | COCO | 48 | 66 | | 500 | 0.9 | 0.000 5 | 0.001 | RTX2080Ti | 19.4 |
| YOLOv5[34] | COCO | 45 | 65 | $2.48\times10^{7}$ | 300 | 0.9 | 0.000 5 | 0.01 | TX1080Ti | 82.2 |
| YOLOX[95] | COCO | 43 | 61 | $1.45\times10^{6}$ | 300 | 0.9 | 0.000 5 | 0.01 | RTX3090 | 27.7 |
| YOLOv3[103] | VisDrone | 27 | | $6.15\times10^{7}$ | 300 | 0.9 | 0.000 5 | 0.002 5 | GTX1080Ti | 139.7 |
| YOLOv4[94] | VisDrone | 32 | 65 | $6.17\times10^{6}$ | 300 | 0.9 | 0.000 5 | 0.01 | 2070Super | 18.0 |
| YOLOv5[74] | VisDrone | 35 | | $1.08\times10^{8}$ | | | | | RTX3060 | 146.0 |
| YOLOv7[78] | VisDrone | 36 | 53 | $7.97\times10^{7}$ | 400 | 0.9 | 0.000 5 | 0.01 | TESLA P100 | |
| YOLOv3[101] | 行人 | 91 | | $2.28\times10^{8}$ | 500 | 0.9 | 0.000 5 | 0.001 | GTX1660Ti | 51.5 |
| YOLOv3[36] | 棕榈树 | 92 | | $8.13\times10^{7}$ | 100 | 0.9 | 0.000 5 | 0.001 | GTX1060 | 26.9 |
| YOLOv4[76] | 输电线路 | 90 | | | 300 | 0.949 | 0.000 5 | 0.001 | RTX3090 | 50.0 |
| YOLOv4[84] | 松树病虫害 | 96 | | $2.01\times10^{8}$ | 175 | 0.9 | 0.1 | 0.005 | i7_10750H | 4.2 |
| YOLOv4[113] | 钢裂纹 | 97 | | $1.16\times10^{7}$ | 500 | 0.9 | 0.000 5 | 0.001 | RTX1080Ti | 8.4 |
| YOLOv5[82] | 风电叶片 | 95 | | $3.78\times10^{7}$ | 300 | 0.9 | 0.000 5 | 0.003 2 | GTX1650 | 41.3 |
| YOLOv5[49] | 松线虫树 | 98 | 98 | $4.67\times10^{7}$ | 500 | 0.9 | 0.000 5 | 0.01 | RTX3060 | 67.0 |
| YOLOv5[35] | 枣椰树 | 92 | 90 | | 300 | 0.9 | 0.000 5 | 0.001 | RTX2080Ti | 11.3 |
| YOLOv8[106] | 烟雾 | 92 | 85 | | 200 | 0.9 | 0.000 5 | 0.001 | GTX1080Ti | 167.0 |

## 4 结束语

本文总结了YOLO系列算法的发展历程、相互关系、主要的优化策略,且运用统一流程绘制了其框架结构;梳理了近几年来UAV视角下目标检测YOLO系列算法的研究与应用成果,重点分析了目标检测模型的优化策略。从YOLO系列算法与UAV视角下目标检测应用2个角度进行了总结,有助于将YOLO算法的理论与应用进行比较;分析了UAV视角下目标检测YOLO算法的应用数据集、数据集扩充方法、预处理方法。尽管YOLO的版本越来越多,不同版本性能较之前有一定的提升,但仍无法满足UAV视角下的多元化任务需求。通过梳理,能为YOLO算法与UAV视角下目标检测技术的研究提供有价值的参考。

目前,UAV目标检测技术发展迅速、应用广泛,但其检测感知能力还不足以满足广泛应用与行业需求,还需要性能更为优越的YOLO目标检测模型与优化策略。根据当前UAV视角下的YOLO目标检测模型研究与应用现状,做出如下问题总结与研究展望:

1)数量更大、场景更丰富的UAV目标检测公共数据集。现有的UAV目标检测公共数据集中图像数量少,目标类别少,场景不够丰富,且YOLO系列算法在一些公共数据上的性能表现不佳,在VisDrone数据集上,YOLO目标检测的一些改进版本性能mAP的平均值约为0.35,而YOLO的10种基础版本在COCO数据集上mAP平均值约为0.43。使用这些公共数据集改进模型的测试与验证,很难充分检测模型的性能。如果使用迁移训练策略训练UAV目标检测模型,在这些公共数据集训练得到的低精度模型参数还需要更多地优化才能满足需求。因此,UAV视角下目标检测公共数据集的制作与完善需要研究人员或组织机构给予足够的关注与投入。

2)UAV视角下多目标检测YOLO算法。在UAV视角下YOLO目标检测应用分析中,多数应用是针对单一目标检测,检测精度较高。然而,多目标物体检测的精度往往不会很高,目标类别越多,检测模型平均精度会越低,因此,基于UAV视角下YOLO目标检测模型的多目标检测、多目标跟踪的能力有待进一步提升。

3)UAV视角下YOLO检测模型的多模态数据处理。由上述分析和总结可知,UAV采集的图像与视频易受到光源、环境、天气、设备等干扰,使图像数据具有同物异谱、同谱异物、遮挡、噪声干扰等多模态。因此,UAV视角下YOLO目标检测模型的改进要适应于不同模态的数据,确保目标检测的准确性与实时性。

4)轻量级UAV视角下YOLO算法与检测速度提升。提高小目标检测精度,通常在原有模型基础上增加更多的提取小目标特征的模块,这必将增加模型的复杂度与检测速度,因此,很多YOLO检测模型改进是权衡检测速度和检测精度。后续UAV视角下YOLO目标检测模型研究之一是寻求模型复杂度与检测精度的平衡。

5) YOLO 的特征信息共享机制。在 UAV 视角下 YOLO 目标检测模型使用的数据集中, 多数图像是在同一段视频中按帧采集, 相邻帧间的图像信息可能会有相关性。在目标检测过程中, 图像间的时间与空间信息共享可以提高目标检测模型的检测性能与检测速度。如何在 UAV 目标检测模型中应用图像空间或时间的共享信息来提高检测性能也是改进的研究方向。

6) 小目标检测能力提升。YOLO 现有的小目标检测性能难以满足各种现实应用, 如车辆无人驾驶系统、UAV 自动导航巡航系统, 这些应用对于小目标检测的准确性、实时性要求较高。从上述分析可见, UAV 视角下 YOLO 目标检测模型主要改进之一是提升模型对小目标的检测能力。因此, 今后 UAV 视角下 YOLO 检测模型的研究还会延续小目标检测性能的提升。

7) 多种特征提取技术融合。单纯依赖于 YOLO 解决 UAV 视角下目标检测任务可能达不到应用需求, 为提高模型的检测准确性与减小模型, 需将多种特征提取与融合技术和 YOLO 相结合。这些技术必将给 YOLO 的改进提供更好的研究方向。

8) UAV 视角下目标检测的多种感知技术融合。对目标的检测与识别仅仅运用图像识别技术可能达不到预期的效果, 因此, 可以结合目标所发出的声音, 使用音频识别技术提高对目标的定位、识别能力。

9) UAV 视角下旋转目标检测。YOLO 的基础版目标检测主要是找出物体在图像中的真实框, 真实框是一种矩形框, 矩形框的边平行于图像的边界。但实际上, UAV 在飞行过程中的姿态不一定能保证拍摄到的目标始终保持方向不变, 因此, YOLO 算法检测到的图像矩形框通常大于物体本身。用旋转后矩形框可能会更好地贴近目标, 有利于计算机视觉测量、目标定位、目标分类等任务。因此, 旋转矩形框检测正是当前 YOLO 目标检测的主要改进研究之一。

## 参考文献 (References)


[1] 蔡志浩, 陈文军, 赵江, 等. 基于动态视觉传感器的无人机目标检测与避障[J]. 北京航空航天大学学报, 2024, 50(1): 144-153.
Cai Z H, Chen W J, Zhao J, et al. Object detection and obstacle avoidance based on dynamic vision sensor for UAV[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(1): 144-153(in Chinese).

[2] 薛康, 刘琨, 江俊峰, 等. 基于 YOLOv5s 模型的光纤振动传感事件精准检测研究[J]. 光学学报, 2023, 43(2): 0228001.
Xue K, Liu K, Jiang J F, et al. Optical fiber vibration sensing detection with high accuracy based on YOLOv5s model[J]. Acta Optica Sinica, 2023, 43(2): 0228001(in Chinese).

[3] 刘莉, 曹潇, 张晓辉, 等. 轻小型太阳能/氢能无人机发展综述[J]. 航空学报, 2020, 41(3): 1-28.
Liu L, Cao X, Zhang X H, et al. Review of development of light and small scale solar/hydrogen powered unmanned aerial vehicles[J]. Acta Aeronautica et Astronautica Sinica, 2020, 41(3): 1-28 (in Chinese).

[4] 张继贤, 刘飞, 王坚. 轻小型无人机测绘遥感系统研究进展[J]. 遥感学报, 2021, 25(3): 708-724.
Zhang J X, Liu F, Wang J. Review of the light-weighted and small UAV system for aerial photography and remote sensing[J]. Journal of Remote Sensing, 2021, 25(3): 708-724(in Chinese).

[5] 刘芳, 孙亚楠. 基于自适应融合网络的无人机目标跟踪算法[J]. 航空学报, 2022, 43(7): 359-369.
Liu F, Sun Y N. UAV target tracking algorithm based on adaptive fusion network[J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(7): 359-369(in Chinese).

[6] 张博恒, 柴栋栋, 孟令博, 等. 基于多重检测的无人机抗遮挡目标跟踪算法[J]. 北京航空航天大学学报, 2023, 49(9): 2442-2454.
Zhang B H, Chai D D, Meng L B, et al. Anti-occlusion target tracking algorithm of UAV based on multiple detection[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(9): 2442-2454(in Chinese).

[7] Liu J H, Wang Z P, Wu Y P, et al. An improved faster R-CNN for UAV-based catenary support device inspection[J]. International Journal of Software Engineering and Knowledge Engineering, 2020, 30(7): 941-959.

[8] 刘英杰, 杨风暴, 胡鹏. 基于 Cascade R-CNN 的并行特征金字塔网络无人机航拍图像目标检测算法[J]. 激光与光电子学进展, 2020, 57(20): 201505.
Liu Y J, Yang F B, Hu P. Parallel FPN algorithm based on Cascade R-CNN for object detection from UAV aerial images[J]. Laser & Optoelectronics Progress, 2020, 57(20): 201505(in Chinese).

[9] Peng X, Zhong X G, Zhao C, et al. A UAV-based machine vision method for bridge crack recognition and width quantification through hybrid feature learning[J]. Construction and Building Materials, 2021, 299: 123896.

[10] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognitio. Piscataway: IEEE Press, 2016: 779-788.

[11] Li Y D, Dong H, Li H G, et al. Multi-block SSD based on small object detection for UAV railway scene surveillance[J]. Chinese Journal of Aeronautics, 2020, 33(6): 1747-1755.

[12] Liu J, Jia R, Li W, et al. High precision detection algorithm based on improved RetinaNet for defect recognition of transmission lines[J]. Energy Reports, 2020, 6: 2430-2440.

[13] Kannadaguli P. FCOS based human detection system using thermal imaging for UAV based surveillance applications[C]//Proceedings of the 2020 IEEE Bombay Section Signature Conference. Piscataway: IEEE Press, 2021: 79-83.

[14] 耿创, 宋品德, 曹立佳. YOLO 算法在目标检测中的研究进展[J]. 兵器装备工程学报, 2022, 43(9): 162-173.
Geng C, Song P D, Cao L J. Research progress of YOLO algorithm in target detection[J]. Journal of Ordnance Equipment En-

gineering, 2022, 43(9): 162-173(in Chinese).

[15] 邵延华, 张铎, 楚红雨, 等. 基于深度学习的YOLO目标检测综述[J]. 电子与信息学报, 2022, 44(10): 3697-3708.
Shao Y H, Zhang D, Chu H Y, et al. A review of YOLO object detection based on deep learning[J]. Journal of Electronics & Information Technology, 2022, 44(10): 3697-3708(in Chinese).

[16] 王琳毅, 白静, 李文静, 等. YOLO系列目标检测算法研究进展[J]. 计算机工程与应用, 2023, 59(14): 15-29.
Wang L Y, Bai J, Li W J, et al. Research progress of YOLO series target detection algorithms[J]. Computer Engineering and Applications, 2023, 59(14): 15-29(in Chinese).

[17] Terven J, Cordova-Esparza D. A comprehensive review of YOLO architectures in computer vision: from YOLOv1 to YOLOv8 and YOLO-NAS[EB/OL]. (2024-02-04)[2024-06-10]. https://arxiv.org/abs/2304.00501.

[18] Wu X, Li W, Hong D F, et al. Deep learning for unmanned aerial vehicle-based object detection and tracking: a survey[J]. IEEE Geoscience and Remote Sensing Magazine, 2022, 10(1): 91-124.

[19] 冷佳旭, 莫梦竟成, 周应华, 等. 无人机视角下的目标检测研究进展[J]. 中国图象图形学报, 2023, 28(9): 2563-2586.
Leng J X, Mo M J C, Zhou Y H, et al. Recent advances in drone-view object detection[J]. Journal of Image and Graphics, 2023, 28(9): 2563-2586(in Chinese).

[20] 罗旭东, 吴一全, 陈金林. 无人机航拍影像目标检测与语义分割的深度学习方法研究进展[J]. 航空学报, 2024, 45(6): 235-264.
Luo X D, Wu Y Q, Chen J L. Research progress on deep learning methods for object detection and semantic segmentation in UAV aerial images[J]. Acta Aeronautica et Astronautica Sinica, 2024, 45(6): 235-264(in Chinese).

[21] 苑玉彬, 吴一全, 赵朗月, 等. 基于深度学习的无人机航拍视频多目标检测与跟踪研究进展[J]. 航空学报, 2023, 44(18): 1-31.
Yuan Y B, Wu Y Q, Zhao L Y, et al. Research progress of UAV aerial video multi-object detection and tracking based on deep learning[J]. Acta Aeronautica et Astronautica Sinica, 2023, 44(18): 1-31(in Chinese).

[22] 邢姗姗, 赵文龙. 基于YOLO系列算法的复杂场景下无人机目标检测研究综述[J]. 计算机应用研究, 2020, 37(增刊2): 28-30.
Xing S S, Zhao W L. Review of UAV target detection in complex scenarios based on YOLO series algorithms[J]. Application Research of Computers, 2020, 37(Sup 2): 28-30(in Chinese).

[23] 杨欣, 王刚, 李椋, 等. 基于深度卷积神经网络的小型民用无人机检测研究进展[J]. 红外技术, 2022, 44(11): 1119-1131.
Yang X, Wang G, Li L, et al. Civil drone detection based on deep convolutional neural networks: a survey[J]. Infrared Technology, 2022, 44(11): 1119-1131(in Chinese).

[24] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 6517-6525.

[25] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08)[2024-06-10]. https://arxiv.org/abs/1804.02767.

[26] Bochkovskiy A, Wang C Y, Liao H M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2024-06-10]. https://arxiv.org/abs/2004.10934.

[27] Wang C Y, Mark Liao H Y, Wu Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE Press, 2020: 1571-1580.

[28] 陈宇梁, 董绍江, 孙世政, 等. 改进YOLOv5s的弱光水下生物目标检测算法[J]. 北京航空航天大学学报, 2024, 50(2): 499-507.
Chen Y L, Dong S J, Sun S Z, et al. Improved YOLOv5s low-light underwater biological target detection algorithm[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(2): 499-507(in Chinese).

[29] Ge Z, Liu S T, Wang F, et al. YOLOX: exceeding YOLO series in 2021[EB/OL]. (2021-08-06)[2024-06-10]. https://arxiv.org/abs/2107.08430.

[30] Li C Y, Li L L, Jiang H L, et al. YOLOv6: a single-stage object detection framework for industrial applications[EB/OL]. (2022-09-07)[2024-06-10]. https://arxiv.org/abs/2209.02976.

[31] Wang C Y, Bochkovskiy A, Liao H M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[EB/OL].(2022-07-06)[2024-06-10].https://arxiv.org/abs/2207.02696.

[32] Wang C Y, Yeh I H, Liao H M. YOLOv9: learning what you want to learn using programmable gradient information[EB/OL]. (2024-02-29)[2024-06-10]. https://arxiv.org/abs/2402.13616.

[33] 纪守领, 杜天宇, 邓水光, 等. 深度学习模型鲁棒性研究综述[J]. 计算机学报, 2022, 45(1): 190-206.
Ji S L, Du T Y, Deng S G, et al. Robustness certification research on deep learning models: a survey[J]. Chinese Journal of Computers, 2022, 45(1): 190-206(in Chinese).

[34] Wang C X, Wang H R. Cascaded feature fusion with multi-level self-attention mechanism for object detection[J]. Pattern Recognition, 2023, 138: 109377.

[35] Tian G Y, Liu J R, Yang W Y. A dual neural network for object detection in UAV images[J]. Neurocomputing, 2021, 443: 292-301.

[36] Wang X, He N, Hong C, et al. Improved YOLOX-X based UAV aerial photography object detection algorithm[J]. Image and Vision Computing, 2023, 135: 104697.

[37] Nepal U, Eslamiat H. Comparing YOLOv3, YOLOv4 and YOLOv5 for autonomous landing spot detection in faulty UAVs[J]. Sensors, 2022, 22(2): 1-15.

[38] Suo J S, Wang T Y, Zhang X Z, et al. HIT-UAV: a high-altitude infrared thermal dataset for Unmanned Aerial Vehicle-based object detection[EB/OL]. (2023-03-31)[2024-06-10]. https://arxiv.org/abs/2204.03245.

[39] 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 Computing, 2022, 26(1): 361-373.

[40] Han Q H, Liu X, Xu J. Detection and location of steel structure surface cracks based on unmanned aerial vehicle images[J]. Journal of Building Engineering, 2022, 50: 104098.

[41] Cheng C X, Yang J, Wang C, et al. Automatic detection of aerial survey ground control points based on YOLOv5-OBB[EB/OL]. (2023-03-06)[2024-06-10]. https://arxiv.org/abs/2303.03041.

[42] 鲍文霞, 谢文杰, 胡根生, 等. 基于TPH-YOLO的无人机图像麦穗计数方法[J]. 农业工程学报, 2023, 39(1): 155-161.
Bao W X, Xie W J, Hu G S, et al. Wheat ear counting method in UAV images based on TPH-YOLO[J]. Transactions of the Chinese Society of Agricultural Engineering, 2023, 39(1): 155-161

(in Chinese).

[43] 李校林, 刘大东, 刘鑫满, 等. 改进 YOLOv5 的无人机航拍图像目标检测算法[J]. 计算机工程与应用, 2024, 60(11): 204-214.
Li X L, Liu D D, Liu X M, et al. Target detection algorithm of UAV aerial image based on improved YOLOv5[J]. Computer Engineering and Applications, 2024, 60(11): 204-214(in Chinese).

[44] Wang G, Chen Y F, 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.

[45] 姜尚洁, 罗斌, 贺鹏, 等. 利用无人机多源影像检测车辆速度[J]. 测绘学报, 2018, 47(9): 1228-1237.
Jiang S J, Luo B, He P, et al. Vehicle speed detection by multi-source Images from UAV[J]. Acta Geodaetica et Cartog raphica Sinica, 2018, 47(9): 1228-1237(in Chinese).

[46] 罗潇, 於锋, 彭勇. 基于深度学习的无人机电网巡检缺陷检测研究[J]. 电力系统保护与控制, 2022, 50(10): 132-139.
Luo X, Yu F, Peng Y. UAV power grid inspection defect detection based on deep learning[J]. Power System Protection and Control, 2022, 50(10): 132-139(in Chinese).

[47] Zhu L, Xiong J H, Xiong F, et al. YOLO-Drone: airborne real-time detection of dense small objects from high-altitude perspective[EB/OL].(2023-10-11)[2024-06-10].https://arxiv.org/abs/2304.06925.

[48] Qin B X, Sun F G, Shen W X, et al. Deep learning-based pine nematode trees' identification using multispectral and visible UAV imagery[J]. Drones, 2023, 7(3): 183.

[49] Thuan P N, Westerlund T, Queralta J P. Vision-based safe autonomous UAV docking with panoramic sensors[EB/OL]. (2023-10-06)[2024-06-10]. https://arxiv.org/abs/2305.16008.

[50] Zhang J C, Shao J B, Chen J L, et al. Polarization image fusion with self-learned fusion strategy[J]. Pattern Recognition, 2021, 118: 108045.

[51] Andrew W, Greatwood C, Burghardt T. Aerial animal biometrics: individual Friesian cattle recovery and visual identification via an autonomous UAV with onboard deep inference[C]//Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway: IEEE Press, 2020: 237-243.

[52] Luo X, Tian X Y, Zhang H J, et al. Fast automatic vehicle detection in UAV images using convolutional neural networks[J]. Remote Sensing, 2020, 12(12): 1994.

[53] Hinzmann T, Stegemann T, Cadena C, et al. Deep learning-based human detection for UAVs with optical and infrared cameras: system and experiments[EB/OL]. (2020-08-10)[2024-06-10]. https://arxiv.org/abs/2008.04197.

[54] Mokayed H, Nayebiastaneh A, De K, et al. Nordic vehicle dataset (NVD): performance of vehicle detectors using newly captured NVD from UAV in different snowy weather conditions[EB/OL]. (2023-04-27)[2024-06-10]. https://arxiv.org/abs/2304.14466.

[55] 沈英, 刘贤财, 王舒, 等. 基于偏振编码图像的低空伪装目标实时检测[J]. 兵工学报, 2024, 45(5): 1374-1383.
Shen Y, Liu X C, Wang S, et al. Real-time detection of low-altitude camouflaged targets based on polarization encoded images[J]. Acta Armamentarii, 2024, 45(5): 1374-1383(in Chinese).

[56] 赵爽, 黄怀玉, 胡一鸣, 等. 基于深度学习的无人机航拍车辆检测[J]. 计算机应用, 2019, 39(S2): 91-96.
Zhao S, Huang H Y, Hu Y M, et al. Vehicle detection in satellite imagery based on deep learning[J]. Journal of Computer Applications, 2019, 39(S2): 91-96(in Chinese).

[57] Sun J, Yang Y, He X F, et al. Northern maize leaf blight detection under complex field environment based on deep learning[J]. IEEE Access, 2020, 8: 33679-33688.

[58] Wang C Q, Luo D, Liu Y, et al. Near-surface pedestrian detection method based on deep learning for UAVs in low illumination environments[J]. Optical Engineering, 2022, 61(2): 023103.

[59] Cui Q Q, Liu M, Huang X Y, et al. Coarse-to-fine visual autonomous unmanned aerial vehicle landing on a moving platform[J]. Biomimetic Intelligence and Robotics, 2023, 3(1): 100088.

[60] Chen Z W, Cao L J, Wang Q H. YOLOv5-based vehicle detection method for high-resolution UAV images[J]. Mobile Information Systems, 2022, 2022(1): 1828848.

[61] Coletta A, Giorgi F, Maselli G, et al. A2-UAV: application-aware content and network optimization of edge-assisted UAV systems[EB/OL]. (2023-07-24)[2024-06-10]. https://arxiv.org/abs/2301.06363.

[62] Koh P W, Sagawa S, Marklund H, et al. WILDS: a benchmark of in-the-wild distribution shifts[EB/OL]. (2021-07-16)[2024-06-10]. https://arxiv.org/abs/2012.07421.

[63] 陈蕊, 郑华飞, 蒋鸿宇, 等. 结合仿真迁移学习和自适应融合的无人机小目标检测[J]. 小型微型计算机系统, 2023, 44(8): 1743-1749.
Chen R, Zheng H F, Jiang H Y, et al. Combination of simulation-based transfer learning and adaptive fusion for UAV small object detection[J]. Journal of Chinese Computer Systems, 2023, 44(8): 1743-1749(in Chinese).

[64] Singh B, Davis L S. An analysis of scale invariance in object detection-SNIP[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 3578-3587.

[65] Rančić K, Blagojević B, Bezdan A, et al. Animal detection and counting from UAV images using convolutional neural networks[J]. Drones, 2023, 7(3): 179.

[66] 魏湧明, 全吉成, 侯宇青阳. 基于 YOLOv2 的无人机航拍图像定位研究[J]. 激光与光电子学进展, 2017, 54(11): 111002.1-111002.10.
Wei Y M, Quan J C, Hou Y Q Y. Aerial image location of unmanned aerial vehicle based on YOLOv2[J]. Laser & Optoelectronics Progress, 2017, 54(11): 111002.1-111002.10(in Chinese).

[67] Keles M C, Salmanoglu B, Guzel M S, et al. Evaluation of YOLO models with sliced inference for small object detection[EB/OL]. (2022-03-09)[2024-06-10]. https://arxiv.org/abs/2203.04799.

[68] Afifi M, Ali Y, Amer K, et al. Robust real-time pedestrian detection in aerial imagery on jetson TX2[EB/OL]. (2019-05-16)[2024-06-10]. https://arxiv.org/abs/1905.06653.

[69] Yavariabdi A, Kusetogullari H, Celik T, et al. FastUAV-NET: a multi-UAV detection algorithm for embedded platforms[J]. Electronics, 2021, 10(6): 1-9.

[70] Misbah M, Khan M U, Yang Z H, et al. TF-net: deep learning empowered tiny feature network for night-time UAV detection[EB/OL]. (2022-11-29)[2024-06-10]. https://arxiv.org/abs/2211.16317.

[71] 王兆魁, 方青云, 韩大鹏. 成像卫星在轨智能处理技术研究进展[J]. 宇航学报, 2022, 43(3): 259-270.

Wang Z K, Fang Q Y, Han D P. Research progress of on-orbit artificial intelligence technology for imaging satellite[J]. Journal of Astronautics, 2022, 43(3): 259-270(in Chinese).

[72] Liu H Y, Duan X H, Lou H T, et al. Improved GBS-YOLOv5 algorithm based on YOLOv5 applied to UAV intelligent traffic[J]. Scientific Reports, 2023, 13: 9577.

[73] 陈范凯, 李士心. 改进 YOLOv5 的无人机目标检测算法[J]. 计算机工程与应用, 2023, 59(18): 218-225.
Chen F K, Li S X. UAV target detection algorithm with improved YOLOv5[J]. Computer Engineering and Applications, 2023, 59(18): 218-225(in Chinese).

[74] 冒国韬, 邓天民, 于楠晶. 基于多尺度分割注意力的无人机航拍图像目标检测算法[J]. 航空学报, 2023, 44(5): 268-278.
Mao G T, Deng T M, Yu N J. Object detection in UAV images based on multi-scale split attention[J]. Acta Aeronautica et Astronautica Sinica, 2023, 44(5): 268-278(in Chinese).

[75] 祁宣豪, 智敏. 图像处理中注意力机制综述[J]. 计算机科学与探索, 2024, 18(2): 345-362.
Qi X H, Zhi M. Review of attention mechanisms in image processing[J]. Journal of Frontiers of Computer Science & Technology, 2024, 18(2): 345-362(in Chinese).

[76] Yang Z, Xu X F, Wang K K, et al. Multitarget detection of transmission lines based on DANet and YOLOv4[J]. Scientific Programming, 2021, 2021(1): 6235452.

[77] Han K, Wang Y H, Xu C, et al. GhostNets on heterogeneous devices via cheap operations[J]. International Journal of Computer Vision, 2022, 130(4): 1050-1069.

[78] 奉志强, 谢志军, 包正伟, 等. 基于改进 YOLOv5 的无人机实时密集小物体检测算法[J]. 航空学报, 2023, 44(7): 246-260.
Feng Z Q, Xie Z J, Bao Z W, et al. Real-time dense small object detection algorithm for UAV based on improved YOLOv5[J]. Acta Aeronautica et Astronautica Sinica, 2023, 44(7): 246-260(in Chinese).

[79] Zhu L, Wang X J, Ke Z H, et al. BiFormer: vision transformer with bi-level routing attention[EB/OL]. (2023-03-15)[2024-06-10]. https://arxiv.org/abs/2303.08810.

[80] 张宏鸣, 付振宇, 韩文霆, 等. 基于改进 YOLO 的玉米幼苗株数获取方法[J]. 农业机械学报, 2021, 52(4): 221-229.
Zhang H M, Fu Z Y, Han W T, et al. Detection method of maize seedlings number based on improved YOLO[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(4): 221-229(in Chinese).

[81] 林文树, 张金生, 何乃磊. 基于改进 YOLOv4 的落叶松毛虫侵害树木实时检测方法[J]. 农业机械学报, 2023, 54(4): 304-312.
Lin W S, Zhang J S, He N L. Real-time detection method of dendrolimus superans-infested Larix gmelinii trees based on improved YOLOv4[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(4): 304-312(in Chinese).

[82] Zhang R, Wen C B. SOD-YOLO: a small target defect detection algorithm for wind turbine blades based on improved YOLOv5[J]. Advanced Theory and Simulations, 2022, 5(7): 2100631.

[83] Zhao H Y, Zhang H P, Zhao Y Y. YOLOv7-sea: object detection of maritime UAV images based on improved YOLOv7[C]//Proceedings of the 2023 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops. Piscataway: IEEE Press, 2023: 233-238.

[84] Zhao L L, Zhu M L. MS-YOLOv7: YOLOv7 based on multi-scale for object detection on UAV aerial photography[J]. Drones, 2023, 7(3): 171-188.

[85] Liu Z, Lin Y T, Cao Y, et al. Swin Transformer: hierarchical vision Transformer using shifted windows[C]//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2021: 9992-10002.

[86] 丛玉华, 何啸, 邢长达, 等. 基于无人机的轻量化小目标检测网络[J]. 弹箭与制导学报, 2022, 42(6): 6-12.
Cong Y H, He X, Xing C D, et al. Lightweight small target detection network based on UAV[J]. Journal of Projectiles, Rockets, Missiles and Guidance, 2022, 42(6): 6-12(in Chinese).

[87] 韩俊, 袁小平, 王准, 等. 基于 YOLOv5s 的无人机密集小目标检测算法[J]. 浙江大学学报 (工学版), 2023, 57(6): 1224-1233.
Han J, Yuan X P, Wang Z, et al. UAV dense small target detection algorithm based on YOLOv5s[J]. Journal of Zhejiang University (Engineering Science), 2023, 57(6): 1224-1233(in Chinese).

[88] 李子豪, 王正平, 贺云涛. 基于自适应协同注意力机制的航拍密集小目标检测算法[J]. 航空学报, 2023, 44(13): 239-249.
Li Z H, Wang Z P, He Y T. Aerial-photography dense small target detection algorithm based on adaptive cooperative attention mechanism[J]. Acta Aeronautica et Astronautica Sinica, 2023, 44(13): 239-249(in Chinese).

[89] Qin Z W, Wang W S, Dammer K H, et al. A real-time low-cost artificial intelligence system for autonomous spraying in palm plantations[EB/OL]. (2021-03-06)[2024-06-10]. https://arxiv.org/abs/2103.04132.

[90] Xia Y, Ye G X, Yan S S, et al. Application research of fast UAV aerial photography object detection and recognition based on improved YOLOv3[J]. Journal of Physics: Conference Series, 2020, 1550(3): 032075.

[91] Liu X H, Wang C B, Liu L. Research on pedestrian detection model and compression technology for UAV images[J]. Sensors, 2022, 22(23): 9171.

[92] Wu R N, Hua C C, Ding W L, et al. Flame and smoke detection algorithm for UAV based on improved YOLOv4-tiny[C]//Proceedings of the PRICAI 2021: Trends in Artificial Intelligence. Berlin: Springer, 2021: 226-238.

[93] 吴明杰, 云利军, 陈载清, 等. 改进 YOLOv5s 的无人机视角下小目标检测算法[J]. 计算机工程与应用, 2024, 60(2): 191-199.
Wu M J, Yun L J, Chen Z Q, et al. Improved YOLOv5s small object detection algorithm in UAV view[J]. Computer Engineering and Applications, 2024, 60(2): 191-199(in Chinese).

[94] 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.
Chen X, Peng D L, Gu Y. Real-time object detection for UAV images based on improved YOLOv5s[J]. Opto-Electronic Engineering, 2022, 49(3): 210372(in Chinese).

[95] Zhou W, Min X L, Hu R, et al. FasterX: real-time object detection based on edge GPUs for UAV applications[EB/OL]. (2022-09-07)[2024-06-10]. https://arxiv.org/abs/2209.03157.

[96] Zhao Z P, He C, Zhao G M, et al. RA-YOLOX: re-parameterization align decoupled head and novel label assignment scheme based on YOLOX[J]. Pattern Recognition, 2023, 140: 109579.

[97] 曲优, 李文辉. 基于锚框变换的单阶段旋转目标检测方法[J]. 吉

林大学学报 (工学版), 2022, 52(1): 162-173.
Qu Y, Li W H. Single-stage rotated object detection network based on anchor transformation[J]. Journal of Jilin University (Engineering and Technology Edition), 2022, 52(1): 162-173(in Chinese).

[98] Wang M Z, Sun H X, Shi J, et al. Q-YOLO: efficient inference for real-time object detection[EB/OL]. (2023-07-01)[2024-06-10]. https://arxiv.org/abs/2307.04816.

[99] Liu Z, Li J G, Shen Z Q, et al. Learning efficient convolutional networks through network slimming[C]//Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 2755-2763.

[100] Shao Y H, Zhang X P, Chu H Y, et al. AIR-YOLOv3: aerial infrared pedestrian detection via an improved YOLOv3 with network pruning[J]. Applied Sciences, 2022, 12(7): 3627.

[101] 侯鑫, 曲国远, 魏大洲, 等. 基于迭代稀疏训练的轻量化无人机目标检测算法[J]. 计算机研究与发展, 2022, 59(4): 882-893.
Hou X, Qu G Y, Wei D Z, et al. A lightweight UAV object detection algorithm based on iterative sparse training[J]. Journal of Computer Research and Development, 2022, 59(4): 882-893(in Chinese).

[102] Guo X J, Li S Y, Yu J K, et al. PFLD: a practical facial landmark detector[EB/OL]. (2019-03-03)[2024-06-10]. https://arxiv.org/abs/1902.10859.

[103] 彭继慎, 孙礼鑫, 王凯, 等. 基于模型压缩的 ED-YOLO 电力巡检无人机避障目标检测算法[J]. 仪器仪表学报, 2021, 42(10): 161-170.
Peng J S, Sun L X, Wang K, et al. ED-YOLO power inspection UAV obstacle avoidance target detection algorithm based on model compression[J]. Chinese Journal of Scientific Instrument, 2021, 42(10): 161-170(in Chinese).

[104] Sharma V K, Dhiman P, Rout R K. Improved traffic sign recognition algorithm based on YOLOv4-tiny[J]. Journal of Visual Communication and Image Representation, 2023, 91: 103774.

[105] 陈卫彪, 贾小军, 朱响斌, 等. 基于 DSM-YOLOv5 的无人机航拍图像目标检测[J]. 计算机工程与应用, 2023, 59(18): 226-233.
Chen W B, Jia X J, Zhu X B, et al. Target detection for UAV image based on DSM-YOLOv5[J]. Computer Engineering and Applications, 2023, 59(18): 226-233(in Chinese).

[106] Saydirasulovich S N, Mukhiddinov M, Djuraev O, et al. An improved wildfire smoke detection based on YOLOv8 and UAV images[J]. Sensors, 2023, 23(20): 8374.

[107] 张可, 乐全明, 黄文礼, 等. 联合剪枝和知识蒸馏的 PCSA-YOLOs 防振锤缺陷检测算法[J]. 电工电能新技术, 2022, 41(11): 59-69.
Zhang K, Le Q M, Huang W L, et al. PCSA-YOLOs vibration damper defect detection algorithm combined with pruning and knowledge distillation[J]. Advanced Technology of Electrical Engineering and Energy, 2022, 41(11): 59-69(in Chinese).

[108] Howard A, Sandler M, Chen B, et al. Searching for MobileNetV3[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 1314-1324.

[109] 黄丽明, 王懿祥, 徐琪, 等. 采用 YOLO 算法和无人机影像的松材线虫病异常变色木识别[J]. 农业工程学报, 2021, 37(14): 197-203.
Huang L M, Wang Y X, Xu Q, et al. Recognition of abnormally discolored trees caused by pine wilt disease using YOLO algorithm and UAV images[J]. Transactions of the Chinese Society of Agricultural Engineering, 2021, 37(14): 197-203(in Chinese).

[110] Sandler M, Howard A, Zhu M L, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 4510-4520.

[111] 甄然, 刘雨涵, 孟凡华, 等. 基于改进 YOLOv7 的低空飞行物目标检测方法[J]. 无线电工程, 2024, 54(3): 633-643.
Zhen R, Liu Y H, Meng F H, et al. Low altitude flying target detection method based on improved YOLOv7[J]. Radio Engineering, 2024, 54(3): 633-643(in Chinese).

[112] Liu Z, Mao H Z, Wu C Y, et al. A ConvNet for the 2020s[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2022: 11966-11976.

[113] Yu Z W, Shen Y G, Shen C K. A real-time detection approach for bridge cracks based on YOLOv4-FPM[J]. Automation in Construction, 2021, 122: 103514.

[114] 张上, 张岳, 王恒涛, 等. 轻量化无人机遥感图像小目标检测算法[J]. 无线电工程, 2023, 53(10): 2329-2336.
Zhang S, Zhang Y, Wang H T, et al. Lightweight small target detection algorithm in UAV remote sensing image[J]. Radio Engineering, 2023, 53(10): 2329-2336(in Chinese).

[115] 谢椿辉, 吴金明, 徐怀宇. 改进 YOLOv5 的无人机影像小目标检测算法[J]. 计算机工程与应用, 2023, 59(9): 198-206.
Xie C H, Wu J M, Xu H Y. Small object detection algorithm based on improved YOLOv5 in UAV image[J]. Computer Engineering and Applications, 2023, 59(9): 198-206(in Chinese).

[116] 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]. (2022-06-22)[2024-06-10]. https://arxiv.org/abs/2110.13675.

[117] Elfwing S, Uchibe E, Doya K. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning[J]. Neural Networks, 2018, 107: 3-11.

# Research progress on YOLO series algorithms for target detection from UAV vision

CHEN Jinlin，WU Yiquan*，YUAN Yubin

(College of Electronic and Information Engineering，Nanjing University of Aeronautics and Astronautics，Nanjing 211106，China)

Abstract：The you only look once (YOLO) series of deep learning-based object identification algorithms has several advantages, including great adaptability, high detection accuracy, and quick detection speed. Make it the preferred choice for object detection in drone aerial images. Due to the complex and ever-changing nature of aerial images and videos captured by drones, the YOLO series object detection algorithms from the UAV vision have become a challenging problem. Therefore, it is necessary to comprehensively summarize and analyze the recent research and application achievements of YOLO series algorithms for object detection from the UAV vision. Firstly, the YOLO structural diagram is redrawn and the development, connection, and optimization features of the 10 fundamental versions of YOLO are summarized. The roles and objectives of various YOLO optimization strategies are pointed out, indicating the direction for further theoretical research. Secondly, the optimization methods of YOLO in the application field of drone object detection are summarized. This includes preprocessing before data training, optimization of training and inference, application of attention modules, convolutional optimization, detection heads, and optimization of loss functions. At the same time, some application datasets and data expansion methods from the UAV vision are sorted out. Through studies, the effectiveness of YOLO series algorithms on the UAV vision dataset and the influence of convolutional block attention module (CBAM) on various YOLO improvements are assessed. Finally, it is pointed out that there are still some problems and challenges in the research of YOLO series algorithms for object detection from the UAV vision. This article looks forward to possible development trends and research directions in this field.

Keywords：UAV system；object detection；deep learning；YOLO series algorithms；datasets；performance comparison

**Received**：2024-06-12；**Accepted**：2024-07-26；**Published Online**：2024-08-26 11：10

**URL**：link.cnki.net/urlid/11.2625.V.20240824.1220.001

**Foundation items**：National Natural Science Foundation of China (61573183); Anhui Provincial University Research Plan Project (2022AH051584)

* **Corresponding author**. E-mail：nuaaimage@163.com