

采用置信度分层协同决策的目标检测方法

靳苗苗^{1,2}, 余序宜^{1,2}, 赵英海³, 张会鹏^{1,2}, 尚颖慧^{1,2}, 梅魁志^{1,2}

(1. 西安交通大学人机混合增强智能全国重点实验室, 710049, 西安; 2. 西安交通大学人工智能与机器人研究所, 710049, 西安; 3. 北京华航无线电测量研究所, 100013, 北京)

摘要: 针对实时目标检测模型置信度校准不足以及后处理决策机制单一所引发的误检与漏检问题,提出了一种置信度分层协同决策架构。该架构以 YOLOv5 模型为基础检测网络,在推理阶段构建置信度分层机制,根据预测置信度分布,将候选框划分为高置信度直通路径和低置信度精细化判别路径。与此同时,引入极端梯度提升(XGBoost)模型,对低置信度样本进行二次分类与置信度修正,实现检测结果的差异化处理与计算资源的分配。在融合阶段设计统一的决策整合模块,对不同路径输出进行结构化融合与筛选,在不改变主干网络结构和基本不增加计算复杂度的前提下,提升了整体检测性能与预测可靠性。研究结果表明:在 2080Ti 测试环境下、COCO 数据集上,改进后 YOLOv5s 和 YOLOv5m 模型的平均精度均值 $m_{AP,50-95}$ 分别提升了 1.98%、1.99%;泛化至 YOLOv8m 时,模型的 $m_{AP,50-95}$ 为 55.07%,推理速度为 165.4 帧/s。该方法构建了深度神经网络与传统机器学习模型协同决策的新型融合范式,为高可靠领域目标识别系统提供了一种兼顾精度、效率与可信度的技术路径。

关键词: 目标检测;极端梯度提升;YOLOv5 模型;置信度分层;机器学习

中图分类号: TP399 **文献标志码:** A

DOI: 10.7652/xjtuxb202610002 **文章编号:** 0253-987X(2026)10-00022-12

An Object Detection Method Based on Confidence-Stratified Collaborative Decision-Making

JIN Miaomiao^{1,2}, YU Xu yi^{1,2}, ZHAO Yinghai³, ZHANG Huipeng^{1,2},
SHANG Yinghui^{1,2}, MEI Kuizhi^{1,2}

(1. State Key Laboratory of Human-Machine Hybrid Augmented Intelligence, Xi'an Jiaotong University, Xi'an 710049, China; 2. Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an 710049, China; 3. Beijing Huahang Institute of Radio Measurement, Beijing 100013, China)

Abstract: To address the issues of false positives and false negatives caused by insufficient confidence calibration and single post-processing decision mechanisms in real-time object detection models, a confidence-stratified collaborative decision-making framework is proposed. Using the YOLOv5 model as the base detection network, the framework constructs a confidence stratification mechanism during the inference stage. Based on the predicted confidence distribution, candidate bounding boxes are divided into a high-confidence passthrough path and a low-confidence refined discrimination path. Meanwhile, an extreme gradient boosting (XGBoost) model is introduced to

perform secondary classification and confidence recalibration on low-confidence samples, achieving differentiated processing of detection results and optimized allocation of computational resources. In the fusion stage, a unified decision integration module is designed to perform structured fusion and filtering on the outputs from different paths. This improves overall detection performance and predictive reliability without altering the backbone network structure or significantly increasing computational complexity. Experimental results show that, validated on the COCO dataset in a 2080Ti testing environment, the mean average precision $m_{AP,50-95}$ of the improved YOLOv5s and YOLOv5m models increased by 1.98% and 1.99%, respectively. When generalized to YOLOv8m, the model achieved an $m_{AP,50-95}$ of 55.07% with an inference speed of 165.4 frame/s. This method establishes a novel fusion paradigm for collaborative decision-making between deep neural networks and traditional machine learning models, providing a technical pathway that balances accuracy, efficiency, and credibility for object recognition systems in high-reliability domains.

Keywords: object detection; extreme gradient boosting; YOLOv5 model; confidence stratification; machine learning

随着深度神经网络的飞速发展,其表现出的优越性能极大程度上推动了计算机视觉领域的研究与落地应用,同时在图像分类、语义分割、自动驾驶和工业质检等众多场景中取得了卓越进展^[1-2]。目标检测作为一项基础且关键的核心任务,旨在从复杂图像当中精准定位并且去识别感兴趣的目标,其技术水平直接关系到下游应用的效能以及可靠性。开展高效目标检测模型的设计工作始终是学术界以及工业界关注的焦点。基于此,涌现出以 YOLO 系列以及 DETR 系列为代表的主流技术路线^[3-4]。

以 YOLO 为代表的单阶段检测器,由于在速度以及精度间有着出色的平衡,已成为处理实时检测任务的首选。从经典的 YOLOv5 模型^[5]出发,研究者们开展了全方位的优化工作。为提高小目标检测性能,Tang 等^[1]提出了 HIC-YOLOv5,通过改进网络结构,增强了对小尺度特征的感知能力。邱天衡等^[2]从特征融合以及注意力机制角度,对 YOLOv5 模型开展了系统性的改进工作。在模型轻量化方向,Song 等^[6]设计了 YOLO-ELWNet,在保持精度的同时降低了计算复杂度。张凡等^[7]针对朱鹮检测的特定应用,实现了对 YOLOv5s 模型的轻量化改造。Feng 等^[8]把超图计算引入 YOLO 框架,提出了 Hyper-YOLO,借助对目标间高阶关系进行建模来提高检测性能。Ji 等^[9]专注于微小目标,提出了高效的轻量化模型即 YOLO-TLA。Ge 等^[4]提出了 YOLOX,通过引入解耦头、先进的标签分配策略等,进一步推动了 YOLO 系列的演进。与此同时,

以 DETR 模型为代表的基于 Transformer 的端到端检测器,因其能够彻底消除对非极大值抑制(NMS)等手工后处理的依赖而受到重视。Zhao 等^[10]的研究表明,优化后的 DETR 模型在实时检测任务上可超越 YOLO 系列。Huang 等^[11]通过引入噪声多锚点策略,稳定训练并提升了 DETR 模型的性能。

为了不断提高检测性能,研究者们从多个维度对 YOLO 等模型进行了持续改进。在缓解训练过程当中的前景以及背景类别极端不平衡问题上,Lin 等^[12]提出的 Focal Loss 损失函数具有开创性的意义。其借助重塑标准交叉熵损失,为前景目标当中难以分类的样本分配更高的权重,从而有效地引导模型关注困难样本,提高了单阶段检测器的精度。开展精确定位是检测的核心挑战,因此 Zhu 等^[13]提出了软锚点目标检测方法,凭借设计一种软性分配策略,使得每个特征点可通过连续权重关联多个目标,从而增强了模型在目标密集以及遮挡等复杂场景下的定位鲁棒性和准确性。此外,针对特定航拍影像等应用场景当中普遍存在的小目标检测难题,Betti 等^[14]专门设计了 YOLO-S 模型,借助优化特征金字塔结构、引入高效的注意力模块并且把检测头予以精简,实现了对航拍影像当中细小目标的高精度以及轻量化检测。Yoon 等^[15]围绕低置信度预测展开了深入研究。在无源域场景下,提出运用低置信度伪标签来开展知识蒸馏的方法,表明低置信度区域蕴含了大量有价值但是模糊的信息,把它们进行精细化处理而非简单丢弃,是提高模型泛化能

力以及性能的关键,这就与当前主流检测器高置信度采纳以及低置信度抑制的简单处理逻辑形成了鲜明对比。

尽管上述研究均取得了进展,但当前主流实时目标检测模型在后处理阶段仍面临着两个关键挑战。一是对 NMS 的严重依赖。无论是 YOLO 系列还是部分 DETR 变体,推理流程仍普遍依赖 NMS 或者其改进版本如 Soft-NMS^[16],这种抑制策略难以适应复杂场景当中目标尺度以及密度的变化,会带来额外的计算开销和延迟^[4,10]。二是决策机制的单一性以及置信度校准的不足。现有模型通常选用单一的检测头来输出分类置信度以及定位框,置信度往往不能准确反映预测的可靠性^[17]。这个误校准现象在高置信度以及低置信度区间均可能会发生,导致模型要么漏检过于保守,要么产生大量假阳性过于激进^[18]。为了实现后处理效果的优化,一些研究尝试引入动态置信度阈值^[19-21]或关注置信度校准^[14]来缓解上述问题。Dave 等^[22]把 YOLOv4 模型以及极端梯度提升树(XGBoost)^[23]相结合用于交通管理,借助 XGBoost 对 YOLO 的初步输出开展二次分类工作,提升了整体精度。Babayomi 等^[24]提出的卷积 XGBoost,也展示了传统机器学习模型以及深度学习特征结合在特定任务像脑瘤检测上的潜力。然而,如何在后处理阶段开展一种轻量、决策机制的设计工作,使其能够依据预测本身的可靠性去进行差异化处理,从而实现精度以及效率的权衡,仍是一个值得探索的问题。在这个特定的领

域当中,已有部分研究尝试去引入专门的机器学习算法,来实现对目标检测的后处理以及置信度校准阶段的优化。例如,部分学者运用支持向量机(SVM)对候选框的几何特征以及初步置信度开展超平面划分工作过滤假阳性样本;另有研究选用随机森林(RF)等基于 Bagging 策略来开展的集成学习方法,借助构建多棵独立决策树来平滑预测置信度。然而,SVM 在处理目标检测当中不平衡的正负样本分布时难以找到最优的边界。随机森林虽然具备较好的抗过拟合能力,但其并行建树机制对低置信度区间内难以区分的困难样本缺乏针对性的纠错以及强化能力。相比之下,极端梯度提升作为一种基于 Boosting 策略开展的先进算法,借助二阶泰勒展开以及动态拟合残差,能够捕捉到低置信度样本当中的特征差异。为此,本文提出一种基于置信度分层协同决策的目标检测方法,以期实现检测精度与效率的优化权衡。

1 基于置信度分层协同决策的目标检测方法

为了实现目标检测任务中精度与效率的平衡,提出基于置信度分层的两阶段检测方法。图 1 给出了该方法的网络框架。图中, c 为候选框的置信度, τ_1 和 τ_2 分别为预设的低、高置信度阈值, S_{high} 和 S_{low} 分别为高、低置信度样本集, B_{high} 和 B_{low} 分别为两条处理路径输出的检测框集合, B 为最终融合后的预测框集合。

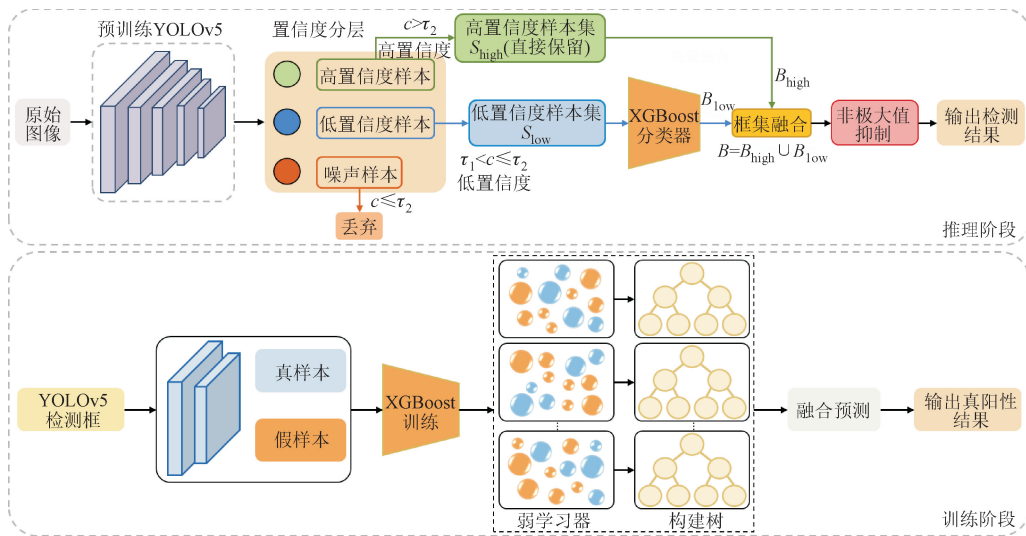


图 1 置信度分层协同目标检测方法框架

Fig. 1 Confidence-based stratified collaborative object detection framework

1.1 现有模型的局限性分析

现有的单阶段检测器虽然开展推理较为高效,但在复杂场景当中普遍存在置信度校准不足的问题。具体表现为:模型所输出的高置信度预测当中可能会夹杂假阳性,且低置信度区间内往往蕴含着大量由于遮挡、模糊以及小尺度所导致的真阳性目标。在不干预主干网络深层特征提取的前提下,为了实现对低置信度模糊样本的精细化重分类,本文引入极端梯度提升树作为二次决策模块。

1.2 置信度协同决策设计

本节详细阐述置信度分层协同决策目标检测方法的实现。整体依靠高效的主干检测网络(1.2.1 小节),生成初步预测以及置信度分布;针对置信度分布不均衡的特性,进行置信度分层以及修正策略(1.2.2 小节),对低置信度预测开展精细化判别;借助决策融合模块(1.2.3 小节),整合高、低置信度路径输出,形成最终检测结果,在降低后处理冗余的同时提高模型对困难样本的判别能力。

如图 1 所示,所提出的置信度决策融合设计方案在整体架构上被划分为了推理阶段以及训练阶段。在推理阶段,原始图像经过预训练的 YOLOv5 主干网络处理后,输出初步检测框。置信度分层模块根据所设定的低置信度阈值以及高置信度阈值,对候选框开展路径划分工作。若样本的预测置信度大于高阈值,则视为高置信度样本,直接依靠直通路保留下来以此保障推理效率;若样本的置信度低于低阈值,就会被视为背景噪声而直接予以丢弃。此时,处于两者之间的低置信度样本集会被送入决策融合模块,由 XGBoost 分类器开展二次判别以及置信度的修正工作。最终,高、低置信度路径的输出

经过整合以及非极大值抑制之后,给出最终模型检测结果。在训练阶段,为了训练能够有效区分模糊目标的二次判别器,先将 YOLOv5 模型在训练集上所输出的所有候选检测框提取出来,通过计算这些检测框以及真实标注框的交并比(IoU),将 IoU 大于所设定正样本阈值的检测框标记为真样本, IoU 低于负样本阈值或者被判定为背景的检测框标记为假样本,并将这些样本输入到 XGBoost 模型开展训练。模型通过贪心算法开展特征分裂以及学习器的构建工作,不断地去优化目标函数并且拟合残差。在融合预测环节,模型结合所有叶子节点的权重输出修正之后的置信度得分,从而提高模型对低置信度区间内真阳性结果的输出以及判别能力,进而对检测框架开展优化。

1.2.1 基准检测模型 YOLOv5

尽管 YOLOv8、YOLOv11 等较新版本的模型在结构上不断迭代,本文仍选取 YOLOv5 模型作为核心验证基准^[6]。其主要原因在于:YOLOv5 是目前工业界高可靠性领域落地最广泛、跨平台部署工具链最成熟的经典架构,验证该架构上的改进具有极强的工程指导价值。YOLOv5 作为典型的单阶段检测器,其置信度校准不足的问题^[17-18]在学术界已被充分定义与研究,这为本文所提的二次修正方法提供了对比基线。更重要的是,本文的核心贡献在于提出一种深度神经网络结合机器学习分类器的分层决策新范式,该范式可迁移至其他版本的 YOLO 架构中。YOLOv5 模型的结构如图 2 所示,其核心网络结构包括主干网络、颈部层和检测头,通过多尺度特征提取与融合提升检测性能。

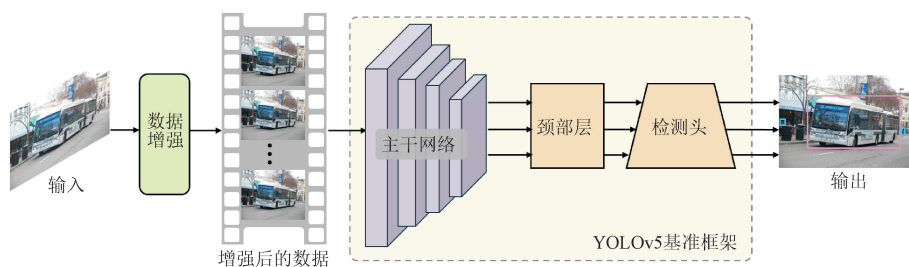


图 2 YOLOv5 模型结构和训练流程

Fig. 2 Structure and training process of YOLOv5 model

主干网络模块包含 Focus、Cross Stage Partial (CSP)和空间金字塔池化(SPP)组件。Focus 模块通过切片与拼接实现无信息损失的下采样,将输入尺寸 $3 \times H \times W$ 转换为 $12 \times H/2 \times W/2$,高度 H 和

宽度 W 均为 640^[5]; CSP 模块通过残差单元拆分特征流并融合,在降低计算量的同时保持特征表达能力; SPP 模块^[1]通过 5×5 、 9×9 、 13×13 多尺度最大池化融合不同感受野特征,增强对不同尺度目标

的适配性。颈部层模块采用特征金字塔网络与路径聚合网络(FPN-PAN)结构^[2,9],通过上采样与特征拼接实现跨尺度特征融合,弥补了单尺度特征对小目标检测的局限性。检测头模块输出检测框坐标、置信度与类别概率,其损失函数由边界框损失、置信度损失与类别损失加权组成,表达式如下

$$\mathcal{L}_{\text{YOLOv5}} = \lambda_{\text{box}} \mathcal{L}_{\text{Clou}} + \lambda_{\text{obj}} \mathcal{L}_{\text{BCE}}^{\text{obj}} + \lambda_{\text{cls}} \mathcal{L}_{\text{BCE}}^{\text{cls}} \quad (1)$$

式中: $\lambda_{\text{box}} = 0.05$ 、 $\lambda_{\text{obj}} = 1.00$ 、 $\lambda_{\text{cls}} = 0.50$ 为基于 YOLOv5 官方默认超参数配置的损失权重^[5]; $\mathcal{L}_{\text{BCE}}^{\text{obj}}$ 、 $\mathcal{L}_{\text{BCE}}^{\text{cls}}$ 分别为目标置信度损失^[5] 和类别分类损失^[5]; $\mathcal{L}_{\text{Clou}}$ 为完全交并比 Clou 损失,用于衡量预测框与真实框的重叠、距离及长宽比一致性,其表达式为

$$\mathcal{L}_{\text{Clou}} = 1 - R_{\text{IoU}} + \frac{\rho^2(p, p_g)}{d^2} + \alpha v \quad (2)$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w_g}{h_g} - \arctan \frac{w_b}{h_b} \right)^2 \quad (3)$$

$$\alpha = \frac{v}{1 - R_{\text{IoU}} + v} \quad (4)$$

其中,下标 g 表示真实框, R_{IoU} 为预测框与真实框的交并比, $\rho(p, p_g)$ 为预测框中心 p 与真实框中心 p_g 的欧氏距离, d 为包围两者最小矩形的对角线长度, v 为衡量长宽比一致性指标的惩罚项, α 为权重系数, w_g 、 h_g 分别为真实框的宽和高, w_b 、 h_b 分别为预测框的宽和高。

上述基准检测器虽然能够高效生成丰富的候选预测,但其输出置信度存在固有的校准问题^[13],即分数值并不能总是精确反映预测的真实可靠性,这为本文提出的协同决策方法提供了改进空间。

1.2.2 置信度分层模块

针对基准检测器输出置信度校准不足的问题,引入 XGBoost^[18-19] 构建轻量级置信度判别模块,专门处理模糊或困难的低置信度样本。该策略的核心是对检测结果进行置信度分层,并实现资源的分配。通过阈值 $\tau_1 = 0.1$ 和 $\tau_2 = 0.2$,将检测结果 $C(c)$ 划分为 3 类,表示如下

$$C(c) = \begin{cases} S_{\text{high}}, & c > \tau_2 \\ S_{\text{low}}, & \tau_1 < c \leq \tau_2 \\ S_{\text{discard}}, & c \leq \tau_1 \end{cases} \quad (5)$$

式中: c 为候选框中样本集合的置信度; S_{high} 为高置信度样本集, YOLOv5 检测精度已足够高,直接保留; S_{low} 为低置信度样本集,存在误检或漏检风险,需 XGBoost 修正; S_{discard} 为置信度过低的噪声样本,直接丢弃以降低冗余。

该模块采用标准 XGBoost 作为基础学习器,其核心改进在于构建了连接深度神经网络输出与树模型输入的结构化特征空间。针对进入 S_{low} 的低置信

度候选框,并未直接使用图像像素作为 XGBoost 的输入,而是从 YOLOv5 的预测头部提取了一组低维且信息密集的特征向量,其包含了预测框的几何特征、前景背景的客观置信度分数以及所有类别的条件概率分布向量。XGBoost 实现高精度的置信度修正与重打分,其训练目标是 minimized 正则化损失函数,第 t 轮的目标函数为

$$\mathcal{L} = \sum_{i=1}^n l(y_i, \hat{y}_{i,t-1} + f_t(x_i)) + \Omega(f_t) \quad (6)$$

式中: i 为样本索引; n 为训练总样本数; l 为二值交叉熵损失; t 为当前迭代的决策树索引; $\hat{y}_{i,t-1}$ 为前 $t-1$ 轮的集成预测结果; f_t 为第 t 棵决策树; $\Omega(f_t)$ 为正则项,用以控制模型复杂度防止过拟合^[19],表达式为

$$\Omega(f_t) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (7)$$

其中, $\gamma = 0.10$ 、 $\lambda = 0.01$ 为正则化系数,以平衡模型拟合能力与泛化性能, T 为树的叶子节点数, w_j 为第 j 个叶子节点的输出权重。

通过贪心算法生成每棵树后,最终集成决策树的输出,并经 Sigmoid 映射到 $[0, 1]$ 区间作为修正后的置信度,计算式为

$$\hat{y}_i = \sigma \left(\sum_{k=1}^{100} f_k(x_i) \right) \quad (8)$$

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (9)$$

式中: k 为集成决策树的索引; $\sigma(\cdot)$ 为 Sigmoid 激活函数; z 为 Sigmoid 激活函数的输入。

XGBoost 在每一轮迭代中拟合当前模型的负梯度方向,通过贪心算法生成新的决策树,并基于结构分数选择最优分裂点,该过程可形式化地表示为

$$\Delta J = \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma \quad (10)$$

式中: ΔJ 为该节点分裂前后的信息增益; G_L 、 H_L 分别为左子节点的一阶梯度和二阶梯度之和, G_R 、 H_R 分别为分裂后右子节点的一阶梯度和二阶梯度之和。

该分裂准则使得 XGBoost 能够自动识别具有判别性的特征组合,尤其适用于高维、稀疏或非线性的特征空间。由于树模型的数据驱动特性,尽管输入特征维度相同,针对不同规模基准模型训练的 XGBoost,其内部决策树结构也各自独立。在训练阶段, XGBoost 需针对每一种规模的主干网络,利用其自身在训练集上产生的低置信度预测结果,提取特征并训练对应的模型,以确保梯度提升树能够精准拟合并修正特定主干网络的预测。

1.2.3 决策融合模块

为将上述高置信度直通路径与低置信度修正路径的输出进行有效整合,设计了决策融合模块。对于 S_{high} 中的样本,直接保留其原始检测框 $B_{\text{high}} = (x_b, y_b, w_b, h_b, c_{\text{high}}, p_{\text{high}})$, 其中 x_b, y_b 为该框的中心坐标, c_{high} 为该框的置信度, p_{high} 为类别概率分布; 对于 S_{low} 中的样本,经 XGBoost 分类器修正后,得到检测框 $B_{\text{low}} = (x_b, y_b, w_b, h_b, \hat{c}_{\text{low}}, p_{\text{low}})$, 其中 $\hat{c}_{\text{low}}$ 为修正后的置信度, p_{low} 为该路径下输出的类别概率分布。框集融合模块将 S_{high} 与 S_{low} 的检测框合并为候选框集 $B = B_{\text{high}} \cup B_{\text{low}}$, 在当前框架下,采用轻量化 NMS^[12] 作为后处理步骤,以去除高度重叠的冗余框。其操作可具体表示为

$$B_{\text{final}} = \text{NMS}(B, \tau_{\text{NMS}}) \quad (11)$$

式中: B_{final} 为最终有效预测实例集合; τ_{NMS} 为非极大值抑制所设定的交并比阈值。

上述优化不仅能促使基准检测器学习如何检测目标,还能学习如何评估自身预测的可靠性,从而更合理地将样本分配给不同的决策路径,实现高性能检测^[10,25]。

2 实验结果与讨论

2.1 数据集选取与预处理

选取计算机视觉领域公认的通用目标检测基准数据集 COCO 数据集^[26] 开展所有实验。COCO 训练集共包含 118 287 张图像,其中 117 266 张为含标注的有效图像,1 021 张为无目标的背景图像,其标注体系覆盖 80 类日常生活常见目标,累计生成超过 86 万个目标边界框的精确标注。验证集共包含 5 000 张图像,其中 4 952 张为含标注的有效图像,48 张为背景图像,标注目标总数约 36 万个,且类别分布与训练集完全一致。对 COCO 数据集图像实施标准化预处理,所有图像统一调整为 640 像素 $\times$ 640 像素的固定尺寸,并采用 letterbox 填充策略,有效避免图像拉伸引发的目标形变,同时保障模型输入维度的一致性。实验中,特征图输出为 3 个尺度,锚框采用 YOLOv5 原生预聚类配置,具体参数如表 1 所示。

表 1 不同特征图的锚框尺寸

Table 1 Anchor box size for different feature maps

特征图尺寸/像素	锚框尺寸/像素
80 $\times$ 80	10 $\times$ 13, 10 $\times$ 30, 33 $\times$ 23
40 $\times$ 40	30 $\times$ 61, 62 $\times$ 45, 59 $\times$ 119
20 $\times$ 20	116 $\times$ 90, 156 $\times$ 198, 373 $\times$ 326

为了验证所提方法在小样本以及不同场景特征当中的泛化能力以及有效性,引入 PASCAL VOC 数据集来开展交叉验证。实验选用 VOC 2007 以及 VOC 2012 的训练验证集作为联合训练集,共包含 16 551 张图像;评估测试选用 VOC 2007 的测试集,共包含 4 952 张图像。该数据集的标注体系涵盖了 20 个常见目标类别。在此类已有较高检测精度的数据集上测试,能够更好地去评估本文分层决策机制以及极端梯度提升模型在维持决策稳定性方面的表现。

2.2 训练环境

硬件平台采用搭载 NVIDIA GeForce RTX 2080Ti 显卡的计算设备,软件环境基于 Python 3.8.10 构建,核心依赖库包括 PyTorch 2.4.1 深度学习框架、CUDA 12.1 并行计算工具及 XGBoost 1.7.6 算法库,模型学习率设置为 0.1。采用 100 棵基学习器集成 XGBoost 超参数,每棵决策树的最大深度设定为 6。选取 YOLOv5^[5] 作为基准模型,所有模型均加载 Ultralytics 官方发布的预训练权重。为确保基础模型的一致性,未对 YOLOv5 的网络结构进行任何修改,仅调用其推理阶段的特征输出与检测结果,训练出与其规模相匹配的 XGBoost 模型。其核心优势在于具备优异的延迟和精度权衡特性,且支持多尺度模型规格选型的灵活性。

2.3 性能指标

采用目标检测核心评估指标对模型检测性能进行全面量化评估。各指标如表 2 所示。由于精确率与召回率通常呈负相关关系,因此引入 F_1 分数作为平衡二者的综合度量指标。其中, T_p 为真正例, F_p 为假正例, F_n 为假负例。

表 2 目标检测模型评估指标

Table 2 Evaluation metrics for object detection models

性能指标	计算公式
精确率	$P = \frac{T_p}{T_p + F_p}$
召回率	$R = \frac{T_p}{T_p + F_n}$
F_1 分数	$F_1 = \frac{2PR}{P + R}$
$m_{AP,50}$	$m_{AP,50} = \frac{1}{Q} \sum_{q=1}^Q A_{P,q}$

表 2 中,平均精度均值 m_{AP} 为核心性能评估指标, $m_{AP,50}$ 指将交并比阈值严格设定为 0.50 时,该单一阈值下所有类别的平均精度 A_P 的均值; $m_{AP,50-95}$ 为以 0.05 为步长,从 0.50 递增至 0.95,取各阈值

下 m_{AP} 的均值,可全面反映模型在不同定位精度要求下的性能稳定性。 Q 、 q 分别为目标类别总数和第 q 个具体的类别。此外,引入浮点运算量来衡量模型的计算成本,并采用每秒传输帧数(帧率)来严格评估模型的实时推理性能。

2.4 实验结果与对比分析

2.4.1 不同规模 YOLOv5 模型的样本分布统计分析

将各模型在相同训练策略与置信度筛选阈值的条件下进行对比,结果如图 3 所示。由图可知,引入 XGBoost 非线性机器学习模型进行二次精细化分类是极其必要的。对于 YOLOv5x 模型,其总收集样本数较 YOLOv5n 模型减少约 33.9%,在低置信度下真阳性样本数降幅达 54.7%,假阳性样本数亦减少 28.7%,表明随着模型规模与复杂度的提升,其整体上呈现出样本数与构成随模型增大而变化的系统性趋势。模型规模的增大虽有助于提升特征表达能力与预测可靠性,但同时也会带来参数量与计算复杂度的增加。因此,实际应用中,需要在检测性能与模型开销之间取得合理权衡。基于上述考虑,选取小规模和中等规模的 YOLOv5 模型作为代表开展对比实验分析,以评估所提方法在不同复杂度模型下的性能表现。

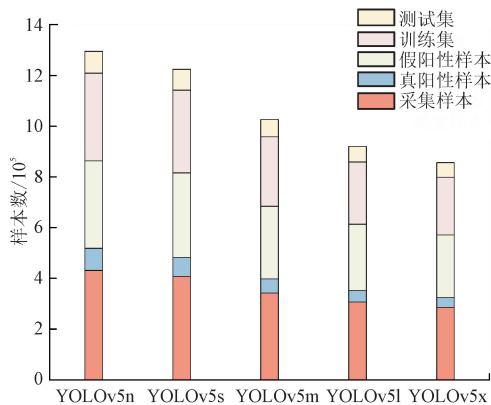


图 3 COCO 数据集中不同规模 YOLOv5 模型的样本数统计

Fig. 3 Sample count statistics of YOLOv5 models of different scales in the COCO dataset

2.4.2 小规模模型的精度与计算成本对比分析

在明确不同规模模型的样本分布特性后,选取小规模模型对本文方法的有效性进行验证。表 3 展示了 COCO 数据集上小规模目标检测模型在精度与计算成本上的性能对比。由表可知,优化后的小规模模型精度为 40.90%。添加了 XGBoost 模块

后,在保持与 YOLOv5s 相同计算成本的前提下, $m_{AP,50-95}$ 进一步提升至 42.88%,进一步验证了引入 XGBoost 进行后处理优化的有效性。本文方法不涉及大规模矩阵乘法运算,在整体浮点运算统计中对浮点运算量指标影响极小;而在常规模型复杂度统计中,该部分通常不计入卷积计算图内,因此整体浮点运算量可视为保持不变。此外,YOLOv5s 模型的推理速度由 249.70 帧/s 下降至 76.80 帧/s。这种处理速度的衰减主要源于树模型集成推理时串行的条件判断耗时。然而,76.80 帧/s 的处理速度依然远高于常规视频流的实时处理基准线。

表 3 COCO 数据集中本文方法在小规模模型上的精度与计算成本对比

Table 3 Comparison of accuracy and computational cost of the method proposed in this article on small-scale models in the COCO dataset

模型	浮点 运算量/ 10^9	参数量/ 10^6	帧率/ (帧 $\cdot$ s $^{-1}$)	$m_{AP,50-95}$ / %
本文(YOLOv5s w/o XGBoost)	16.44	7.23	249.70	40.90
本文(YOLOv5s w/ XGBoost)	16.45	7.23	76.80	42.88

注:表内“w/o”表示移除该模块,“w/”表示添加该模块。

表 4 给出了 VOC 数据集的评估结果。可以看出,在相对简单的小样本数据分布下,所提架构依然表现出稳定的优化能力。对于小规模 YOLOv5s 基线模型,在引入 XGBoost 后处理模块后,模型在完全不增加主干网络参数量与浮点运算量的条件下,整体 $m_{AP,50-95}$ 由 62.84% 稳步提升至 63.25%。这表明尽管 VOC 数据集的目标混叠程度较低,分层决策融合模块依然能够敏锐地捕捉并修正位于边缘置信度区间的困难样本。

表 4 VOC 数据集中本文方法在小规模模型上的精度与计算成本对比

Table 4 Comparison of accuracy and computational cost of the method proposed in this article on small-scale models in the VOC dataset

模型	浮点 运算量/ 10^9	参数量/ 10^6	帧率/ (帧 $\cdot$ s $^{-1}$)	$m_{AP,50-95}$ / %
本文(YOLOv5s w/o XGBoost)	15.90	7.06	269.30	62.84
本文(YOLOv5s w/ XGBoost)	15.90	7.06	100.60	63.25

注:表内“w/o”表示移除该模块,“w/”表示添加该模块。

此外,受限于 XGBoost 树模型集成的处理机制,模型的推理速度虽由 269.30 帧/s 下降至 100.60 帧/s,但仍能满足绝大多数高可靠应用领域的实时性检测指标。根据小规模模型的特征表达能力,其性能存在局限性。基于上述分析,有必要进一步在中等规模模型上验证该方法的适用性与性能情况。

2.4.3 中等规模模型的精度与计算成本对比分析

由表 5 可知,YOLOv5m 模型在 COCO 数据集上的精度为 48.53%;添加 XGBoost 模块后,YOLOv5m 模型在保持相同浮点运算量与参数量的前提下, $m_{AP,50-95}$ 提升至 50.52%。在算法处理效率方面,YOLOv5m 模型在融合 XGBoost 模块之后,帧率从 162.20 帧/s 降至 69.00 帧/s。伴随着基础网络规模的增大,前端所产生的边界候选框分布特性会发生变化,使得二次分类器的计算也随之去进行调整。尽管检测帧率较原始模型有所下降,但是 69.00 帧/s 的速度依然处在开展实时检测的高效区间内。这表明本文改进方法在维持与 YOLOv5m 主干网络相一致的计算成本的前提下,能够实现检测精度以及推理速度的良好平衡。

表 5 COCO 数据集中本文方法在中等规模模型上的精度与计算成本对比

Table 5 Comparison of accuracy and computational cost of the method proposed in this article on medium-scale models in the COCO dataset

模型	浮点 运算量/ 10^9	参数量/ 10^6	帧率/ (帧·s ⁻¹)	$m_{AP,50-95}$ / %
本文(YOLOv5m w/o XGBoost)	48.88	21.17	162.20	48.53
本文(YOLOv5m w/ XGBoost)	48.88	21.19	69.00	50.52

注:表内“w/o”表示移除该模块,“w/”表示添加该模块。

由表 6 可知,在 VOC 数据集上,集成极端梯度提升模块后,中等规模模型 YOLOv5m 的 $m_{AP,50-95}$ 从 69.13% 小幅提升至 69.28%,且参数量与计算量保持恒定不变。相较于 YOLOv5s 模型,YOLOv5m 模型自身更强的深层特征提取能力使得其在相对简单的 VOC 验证集上产生的模糊低置信度样本数更小,因此二次判别模块能够贡献的绝对提升空间有所收窄。在推理速度方面,帧率由 168.70 帧/s 降至 83.20 帧/s,但总体检测性能与实时性的综合权衡依然处于较高水平。这表明本文设计的决策融合架构在面临不同规模数据集和不同容量主干网络

时,均具有鲁棒性与有效性。

表 6 VOC 数据集中本文方法在中等规模模型上的精度与计算成本对比

Table 6 Comparison of accuracy and computational cost of the method proposed in this article on medium-scale models in the VOC dataset

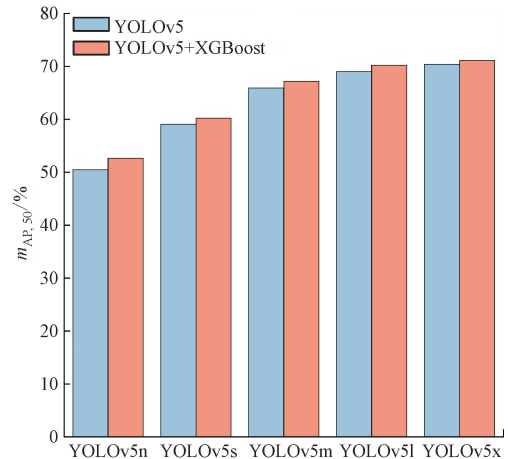
模型	浮点 运算量/ 10^9	参数量/ 10^6	帧率/ (帧·s ⁻¹)	$m_{AP,50-95}$ / %
本文(YOLOv5m w/o XGBoost)	48.10	20.93	168.70	69.13
本文(YOLOv5m w/ XGBoost)	48.10	20.93	83.20	69.28

注:表内“w/o”表示移除该模块,“w/”表示添加该模块。

2.4.4 不同规模 YOLOv5 模型整体性能提升趋势分析

为进一步分析本文方法在不同模型规模下的整体性能变化趋势,图 4 在 COCO 数据集上对比了原始模型与改进模型的检测性能。从图 4(a)可见,原始 YOLOv5 各模型的 $m_{AP,50}$ 随模型规模增大呈稳定递增趋势,从 YOLOv5 模型的 50.46% 逐步提升至 YOLOv5x 模型的 70.12%。在引入 XGBoost 模块后,所有模型的 $m_{AP,50}$ 均实现了提升,优化后依次为 52.64%、60.22%、67.20%、70.24%、71.15%,且模型规模越大, $m_{AP,50}$ 绝对值越高。

图 4(b)的对比结果显示,原始 YOLOv5 模型的 $m_{AP,50-95}$ 同样呈现随模型规模增大递增的现象。小尺度 YOLOv5 模型的指标相对提升幅度更大,在不改变主干网络结构与计算复杂度的前提下提升了整体检测性能,表明本文设计的 XGBoost 后处理模块对低精度基础模型的优化增益更为突出,且大尺度模型在优化后仍保持了精度绝对值的领先优势。



(a) $m_{AP,50}$ 指标

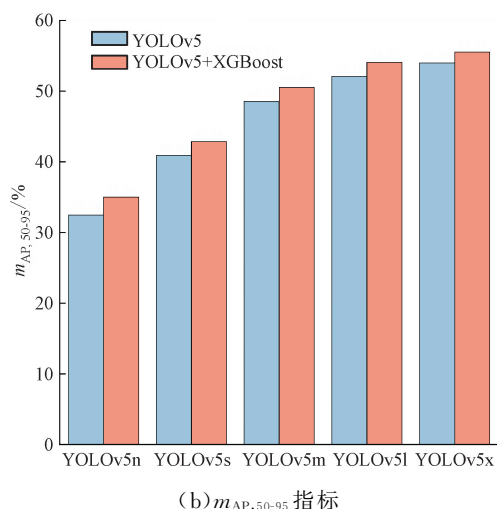


图4 COCO数据集中不同规模YOLOv5模型在添加极端梯度提升模块后的性能结果对比

Fig. 4 Comparison of performance results of YOLOv5 models of different scales after adding extreme gradient boosting module in the COCO dataset

此外,不同规模的模型在检测性能以及置信度分布方面呈现出差异,表明单一的置信度阈值策略难以兼顾不同模型的决策稳定性。因此,有必要引入分层决策机制的,针对不同的置信度区间去开展差异化的处理工作,从而使得检测系统的稳定性以及可靠性得以提升。

2.4.5 先进基线模型与同类机器学习算法的对比分析

为了进一步验证分层决策融合架构在更先进基准模型上的泛化能力,以及与其他传统机器学习分类器以及后处理优化的性能差异,选用YOLOv8m作为基础检测模型,在COCO数据集上开展对比实验。为了验证XGBoost在这种融合范式当中的优越性,实验还引入了随机森林、支持向量机以及改进的后处理算法Soft-NMS进行横向对比。为了更直观地反映算法的实际处理效率,同步引入帧率作为评估指标,具体的对比结果如表7所示。由表可见,在同领域专门算法以及其他机器学习方法的对比当中,本文方法展示出了更优的精度以及效率平衡,具体如下。

(1) YOLOv8m+Soft-NMS算法:虽然保持了高达325.00帧/s的推理速度,但其综合精度提高有限,难以满足对决策可靠性要求较高的场景。

(2) YOLOv8m+SVM算法:虽然 $m_{AP,50-95}$ 达到54.68%,但其推理速度大幅下降到了139.90帧/s。

表7 COCO数据集中采用不同后处理算法的YOLOv8m模型与机器学习模型的性能对比
Table 7 Performance comparison between YOLOv8m model and machine learning model using different post-processing methods in the COCO dataset

优化算法	帧率/ (帧·s ⁻¹)	$m_{AP,50}$ / %	$m_{AP,50-95}$ / %
YOLOv8m + Soft-NMS	325.00	69.27	54.22
YOLOv8m + SVM	139.90	69.35	54.68
YOLOv8m + RF	82.60	67.56	53.06
本文(YOLOv8m w/o XGBoost)	313.70	69.29	54.08
本文(YOLOv8m w/ XGBoost)	165.40	69.49	55.07

注:表内“w/o”表示移除该模块,“w/”表示添加该模块。

(3) YOLOv8m+RF算法:引入RF进行二次判别时,推理速度大幅衰减,这主要是因为随机森林这个的决策树群带来了密集的条件分支运算,难以像SVM的矩阵乘法那样得以实现高效的底层向量化加速。且由于随机森林在去处理此类特征的时候易受噪声影响,精度反而出现了退化。

(4) 本文方法:相比之下,本方法运用的XGBoost协同策略在各项对比当中取得了最优的检测精度,同时仍维持165.40帧/s的高效处理能力。这充分证明了XGBoost在低置信度样本精细化判别任务当中,较其他传统机器学习方法拥有更强的非线性建模能力以及更优的计算执行效能。

2.5 消融实验

2.5.1 引入XGBoost模块后的效果分析

为了定量分析XGBoost模块对检测性能的具体影响,表8给出了在COCO数据集上引入该模块后,不同规模YOLOv5模型在 F_1 分数、召回率以及检测精度的变化。由表可知,XGBoost增强模块对模型性能的影响呈现出明显的指标特异性与模型尺寸依赖性,具体如下。

(1)在所有模型尺寸下,召回率均得到提升。这验证了通过有效开展低置信度预测框的重分类工作,XGBoost模型漏检的问题得以有效缓解。

(2)精度增强效果呈现出分化。在YOLOv5l等较大规模模型上观察到微幅增益,而在YOLOv5n以及YOLOv5s等较小规模模型上则出现轻微损失。这表明,较大规模模型由于特征表达能力较强,置信度分布拥有着更好的可分性,因此二

次判别模块能够有效地把边界样本进行修正;而在轻量化模型当中,低置信度样本的噪声比例相对较高,极大程度上制约了二次判别模块开展纠错的能力。

(3)除了受限于特征表达能力的小规模模型外,其他中型和大型模型变体在融合 XGBoost 模块后, F_1 分数均获得了稳定提升。

表 8 COCO 数据集中采用 XGBoost 模块增强的 YOLOv5 模型性能分析

Table 8 Performance analysis of YOLOv5 model enhanced by XGBoost module in the COCO dataset

模型	精确率/%	召回率/%	F_1 分数/%
YOLOv5n	58.62	43.32	49.82
YOLOv5n w/ XGBoost	56.46	44.18	49.57
YOLOv5s	65.41	51.28	57.49
YOLOv5s w/ XGBoost	64.12	51.61	57.19
YOLOv5m	71.14	58.20	64.02
YOLOv5m w/ XGBoost	70.72	58.57	64.07
YOLOv5l	74.00	61.24	67.02
YOLOv5l w/ XGBoost	74.43	61.34	67.25
YOLOv5x	73.03	63.57	67.97
YOLOv5x w/ XGBoost	72.78	63.99	68.10

注:表内“w/”表示添加该模块。

2.5.2 置信度阈值对检测性能的影响分析

为确定最优阈值,在 COCO 验证集上对置信度区间 $[0.001, 0.500]$ 开展 20 个离散点的全采样,并绘制了精确率、召回率、 F_1 分数以及每张图像平均检测框数的变化曲线,如图 5 所示。分析发现,当阈值为极低值 0.001 时,虽然召回率维持在 80%,但每张图像会产生超过 150 个无效背景框。相反,当阈值为较高值 0.250 时,YOLO 自身的 F_1 分数虽达 0.618,但其召回率已降至 57.5%,大量困难正样本被提前截断,使得后续 XGBoost 失去了修正与召回的机会。

从量化指标来看,置信度等于 0.1 为数学意义上的帕累托最优拐点,此时每张图像的平均检测框数从阈值为 0.001 时的 154 骤降至 11,而真实目标召回率仍维持在 67.3%,几乎未受影响。同时,精确率(43.3%)与 F_1 分数(0.527)也进入了稳定的可优化区间。基于此,在实际模型中,本文采用的策略为:针对置信度大于 0.2 的检测框,直接输出为 YOLO 结果;针对置信度位于 0.1~0.2 的框,送入极端梯度提升进行二次精筛。该设计有效保留了困难

样本的召回潜力,实现了整体检测精度与效率的最优平衡。

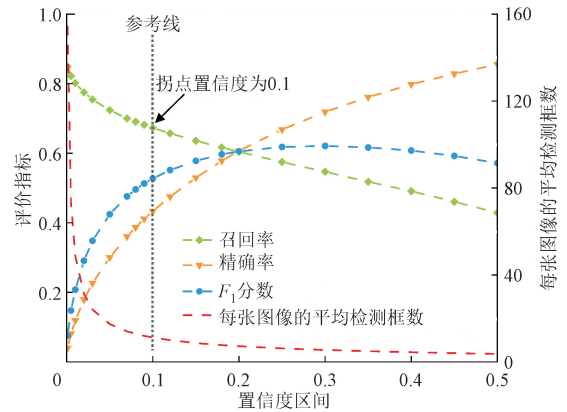


图 5 不同置信度区间对模型性能的影响

Fig. 5 Impact of different confidence intervals on model performance

3 结 论

针对检测模型置信度分布失衡问题,提出了一种置信度分层协同决策的目标检测方法。该方法在不改变主干网络的前提下,借助优化后处理阶段,有效提升了检测性能以及系统稳定性。研究得到的主要结论如下。

(1)将深度检测网络以及极端梯度提升树相协同,对低置信度样本开展二次判别以及置信度校准,实现检测结果的精细化修正以及后处理流程的结构化优化。

(2)在 COCO 以及 VOC 数据集上,各级模型在引入融合策略后,均实现了 1% 以上的绝对精度提升;通过引入二次判别机制,改进后各规模模型的推理速度仍能维持在 69~165 帧/s,在成功在提高预测可靠性的同时,保障了高实时性应用场景的落地需求。

(3)在维持模型原有计算复杂度的条件下,实现了检测精度以及稳定性的协同提升,并且验证了该融合框架在不同规模模型上的适用性。

考虑到不同检测框架在结构设计以及特征表达能力上的差异,后续研究可以在更多类型的目标检测模型上开展系统验证,以进一步评估该融合机制的泛化能力以及结构适配性。同时,可结合动态阈值调节机制以及多场景实验分析,增强方法在复杂环境以及多样化应用需求下的稳定性以及鲁棒性。

开源代码:

<https://github.com/jin2175/YOLO-XGBoost>

参考文献:

- [1] Tang Shiyi, Zhang Shu, Fang Yini. HIC-YOLOv5: improved YOLOv5 for small object detection [C]//2024 IEEE International Conference on Robotics and Automation (ICRA). Piscataway, NJ, USA: IEEE, 2024: 6614-6619.
- [2] 邱天衡, 王玲, 王鹏, 等. 基于改进 YOLOv5 的目标检测算法研究 [J]. 计算机工程与应用, 2022, 58(13): 63-73.
Qiu Tianheng, Wang Ling, Wang Peng, et al. Research on object detection algorithm based on improved YOLOv5 [J]. Computer Engineering and Applications, 2022, 58(13): 63-73.
- [3] 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 Yanhua, Zhang Duo, Chu Hongyu, et al. A review of YOLO object detection based on deep learning [J]. Journal of Electronics & Information Technology, 2022, 44(10): 3697-3708.
- [4] Ge Zheng, Liu Songtao, Wang Feng, et al. YOLOX: exceeding YOLO series in 2021 [PP/OL]. arXiv (2021-07-18) [2025-07-18]. <https://arxiv.org/abs/2107.08430>.
- [5] Jocher G, Chaurasia A, Stoken A, et al. ultralytics/yolov5: v7.0-YOLOv5 SOTA realtime instance segmentation [EB/OL]. (2022-11-22) [2025-12-10]. <https://github.com/ultralytics/yolov5/releases/tag/v7.0>.
- [6] Song Baoye, Chen Jianyu, Liu Weibo, et al. YOLO-ELWNet: a lightweight object detection network [J]. Neurocomputing, 2025, 636: 129904.
- [7] 张凡, 张鹏超, 王磊, 等. 基于 YOLOv5s 的轻量化朱鹮检测算法研究 [J]. 西安交通大学学报, 2023, 57(1): 110-121.
Zhang Fan, Zhang Pengchao, Wang Lei, et al. Research on lightweight crested ibis detection algorithm based on YOLOv5s [J]. Journal of Xi'an Jiaotong University, 2023, 57(1): 110-121.
- [8] Feng Yifan, Huang Jiangang, Du Shaoyi, et al. Hyper-YOLO: when visual object detection meets hypergraph computation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025, 47(4): 2388-2401.
- [9] Ji Chunlin, Yu Tao, Gao Peng, et al. Yolo-tla: an efficient and lightweight small object detection model based on YOLOv5 [J]. Journal of Real-Time Image Processing, 2024, 21(4): 141.
- [10] Zhao Yian, Lü Wenyu, Xu Shangliang, et al. DETRs beat YOLOs on real-time object detection [C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2024: 16965-16974.
- [11] Huang Zixin, Tao Xuesong, Liu Xinyuan. NAN-DETR: noising multi-anchor makes DETR better for object detection [J]. Frontiers in Neurorobotics, 2024, 18: 1484088.
- [12] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection [C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2017: 2999-3007.
- [13] Zhu Chenchen, Chen Fangyi, Shen Zhiqiang, et al. Soft anchor-point object detection [C]//Computer Vision-ECCV 2020. Cham, Germany: Springer International Publishing, 2020: 91-107.
- [14] Betti A, Tucci M. YOLO-S: a lightweight and accurate YOLO-like network for small target detection in aerial imagery [J]. Sensors, 2023, 23(4): 1865.
- [15] Yoon I, Kwon H, Kim J, et al. Enhancing source-free domain adaptive object detection with low-confidence pseudo label distillation [C]//Computer Vision-ECCV 2024. Cham, Switzerland: Springer Nature Switzerland, 2025: 337-353.
- [16] Bodla N, Singh B, Chellappa R, et al. Soft-NMS: improving object detection with one line of code [C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2017: 5562-5570.
- [17] Guo Chuan, Pleiss G, Sun Yu, et al. On calibration of modern neural networks [C]//Proceedings of the 34th International Conference on Machine Learning. Chia Laguna Resort, Sardinia, Italy: PMLR, 2017: 1321-1330.
- [18] Dussert G, Chamaillé-Jammes S, Dray S, et al. Being confident in confidence scores: calibration in deep learning models for camera trap image sequences [J]. Remote Sensing in Ecology and Conservation, 2025, 11(1): 88-99.
- [19] Rai A, Athya D. AdaptiVision: dynamic confidence thresholding for enhanced object detection [C]//2025 International Conference on Engineering Innovations and Technologies (ICoEIT). Piscataway, NJ, USA: IEEE, 2025: 592-598.
- [20] Ding Zhicheng, Lai Zhixin, Li Siyang, et al. Confidence trigger detection: accelerating real-time tracking-by-detection systems [C]//2024 5th International Conference on Electronic Communication and Artificial

- Intelligence (ICECAI). Piscataway, NJ, USA: IEEE, 2024: 587-592.
- [21] Thatikonda M, Pk M K, Amsaad F. A novel dynamic confidence threshold estimation AI algorithm for enhanced object detection [C]//NAECON 2024-IEEE National Aerospace and Electronics Conference. Piscataway, NJ, USA: IEEE, 2024: 359-363.
- [22] Dave P, Chandarana A, Goel P, et al. An amalgamation of YOLOv4 and XGBoost for next-gen smart traffic management system [J]. PeerJ Computer Science, 2021, 7: e586.
- [23] Chen Tianqi, Guestrin C. XGBoost: a scalable tree boosting system [C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: ACM, 2016: 785-794.
- [24] Babayomi M, Olagbaju O A, Kadiri A A. Convolutional XGBoost (C-XGBOOST) model for brain tumor detection [EB/OL]. arXiv (2023-01-05) [2025-11-05]. <https://arxiv.org/abs/2301.02317>.
- [25] Chen Yuming, Yuan Xinbin, Wang Jiabao, et al. YOLO-MS: rethinking multi-scale representation learning for real-time object detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025, 47(6): 4240-4252.
- [26] Reddy K G, Basha S S. Real time object identification: a study on COCO dataset [C]//Proceedings of the International Conference on Advanced Materials, Manufacturing and Sustainable Development (ICAMMSD 2024). Dordrecht, the Netherlands: Atlantis Press, 2025: 860-870.
- (编辑 李慧敏)