doi:10.6041/j.issn.1000-1298.2025.10.052

基于无人机图像和 MLL－YOLO v10s 的草原放牧羊只实时检测模型

张东彦[1,2]　叶佳炜[1,2]　郭阳阳[3]　胡根生[3]　李威风[1,2]　唐晶磊[1]　韩　冬[1,2]

（1.西北农林科技大学陕西省农业信息感知与智能服务重点实验室，陕西杨凌 712100；
2.西北农林科技大学机械与电子工程学院，陕西杨凌 712100；
3.安徽大学农业生态大数据分析与应用国家地方联合工程研究中心，合肥 230601）

摘要：为满足内蒙古草原牧场牧户对大批散养羊群实现精准实时监控和管理的需求，提出一种高精度、轻量化的实时无人机遥感目标检测模型 MLL－YOLO v10s（MobileNetV4 LSKA LSCD－YOLO v10s），实现无人机高空视角下大批羊群中羊只的实时检测。为解决羊群密集和相互遮挡造成的羊只检测困难以及实时性差的问题，基于 YOLO（You only look once）v10 模型，采用 MobileNetV4 作为主干网络，降低模型参数量并提升计算效率；引入大型可分离核注意力模块（LSKA），增强对小目标的特征捕获能力；设计轻量化共享卷积检测头（LSCD），通过权重共享减少计算冗余，提高模型的计算效率。与 YOLO 系列、Faster R－CNN（Faster regions with convolutional neural networks）及其他经典网络模型相比，改进后的 MLL－YOLO v10s 模型在测试集上的 mAP 达到了 93.6%，较基线模型提升 3.4 个百分点，平均帧率为 135 f/s，参数量仅 1.268×10^7，在密集遮挡场景下漏检率显著降低，模型体积与计算量优于主流单/双阶段目标检测算法。提出的 MLL－YOLO v10s 模型在无人机航拍场景下对高密度聚集及部分遮挡的羊只目标展现出更强的鲁棒性，在参数量和计算量上也有明显优势，为边缘计算设备与无人机结合应用提供支持，可为天然牧场中使用无人机巡检羊群提供一种有效的实时检测方法。

关键词：羊只检测；YOLO v10；无人机；小目标；MobileNetV4；轻量化

中图分类号：TP391.4　　**文献标识码**：A　　**文章编号**：1000-1298(2025)10-0575-10　　**OSID**：

Real-time Detection Model for Grazing Sheep in Grassland Based on UAV Imagery and MLL－YOLO v10s

ZHANG Dongyan[1,2]　YE Jiawei[1,2]　GUO Yangyang[3]　HU Gensheng[3]　LI Weifeng[1,2]　TANG Jinglei[1]　HAN Dong[1,2]

(1. *Shaanxi Key Laboratory of Agricultural Information Perception and Intelligent Service, Northwest A&F University, Yangling, Shaanxi* 712100, *China*
2. *College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi* 712100, *China*
3. *National Engineering Research Center for Agro-Ecological Big Data Analysis and Application, Anhui University, Hefei* 230601, *China*)

Abstract: Aiming to meet the demand of herdsmen in Inner Mongolia grassland pastures for accurate real-time monitoring and management of large flocks of free-range sheep, a high-precision and lightweight real-time unmanned aerial vehicle (UAV) remote sensing target detection model named MLL－YOLO v10s (MobileNetV4 LSKA LSCD－YOLO v10s) was proposed. This model enabled real-time detection of individual sheep in large flocks from the high-altitude perspective of UAVs. To address the challenges of difficult sheep detection and poor real-time performance caused by densely packed and mutually occluded sheep, the following improvements were made based on the YOLO (you only look once) v10 model. MobileNetV4 was employed as the backbone network to reduce the number of model parameters and enhance computational efficiency. The large separable kernel attention (LSKA) module was introduced to strengthen the model's ability to capture features of small targets. A lightweight shared convolutional detection head (LSCD) was designed to reduce computational redundancy through weight sharing and

收稿日期：2025－03－31　修回日期：2025－05－14

基金项目：内蒙古自治区科技计划项目(2022YFSJ0039)

作者简介：张东彦(1982—)，男，教授，博士，主要从事大数据与人工智能技术、智能传感器技术研究，E-mail：hello-lion@hotmail.com

通信作者：韩冬(1991—)，男，讲师，博士，主要从事大数据与人工智能技术、农业图像识别研究，E-mail：hanskyy@nwafu.edu.cn

improve the computational efficiency of the model. Compared with the YOLO series, faster regions with convolutional neural networks (Faster R-CNN), and other classic network models, the improved MLL-YOLO v10s model achieved a mean average precision (mAP) of 93.6% on the test set, which was 3.4 percentage points higher than that of the baseline model. It had an average frame rate of 135 f/s and only 1.268×10^7 parameters. In densely occluded scenarios, the false-negative rate was significantly reduced. The model's size and computational requirements were superior to those of mainstream single/dual-stage target detection algorithms. The proposed MLL-YOLO v10s model demonstrated stronger robustness in detecting densely aggregated and partially occluded sheep in UAV aerial photography scenarios. It also had obvious advantages in terms of the number of parameters and computational requirements. This model provided support for the combined application of edge computing devices and UAVs, offering an effective real-time detection method for UAV-based sheep flock inspections in natural pastures.

Key words: sheep detection; YOLO v10; UAV; small object; MobileNetV4; lightweight

0 引言

在现代农业和畜牧业中,动物目标检测技术已成为一种重要工具[1]。该技术在动物摄食行为监控、活动模式分析以及生长发育状态评估等方面,展现出卓越的潜力,为优化养殖管理、提升生产效率提供了重要支持[2]。此外,在智能监控领域,动物目标检测也已广泛应用于养殖场的实时监管与高效管理,实现了精准的个体跟踪与健康状态评估。在内蒙古草原地区,草原放牧是传统习惯的生产方式,草场面积广阔,牧户采用散养方式饲养羊群,但是随着放牧的无序化,对生态造成破坏,为了保护草场生态平衡,政府规定单只羊的放牧面积不得低于 20 km^2,这给羊群定位和管理带来了巨大挑战。

近年来,随着计算机硬件性能提升与深度学习技术迅猛发展,动物目标的实时传输检测逐渐成为计算机视觉领域的研究热点[3-4]。然而,动物目标检测仍然面临诸多挑战,包括动物外观和行为的多样性、复杂多变的生态环境(如光照变化、天气条件及遮挡物)等,这些因素严重影响了现有检测模型的鲁棒性和准确性。

无人机技术和深度学习技术的引入为草原牧场的监控和管理提供了新的解决方案,通过采集与分析航拍图像数据,可以实现羊群的快速定位与实时监控[5]。但在无人机高空航拍场景中面临诸多问题:由于拍摄距离远导致目标像素占比过低,传统卷积网络难以捕获有效的鉴别性特征。已有检测模型[6-9]只针对养殖场中群体养殖环境下的羊只,对草原放牧散养的羊只效果并不好。翁智等[10]改进 YOLO v8 模型,构建了 CSD-YOLO v8s 检测模型,进一步实现了模型轻量化,但是该模型依然无法满足在嵌入式设备上进行实时检测的要求。YOLO v8 结构仍采用分离式任务头和较重的解码过程,难以实现真正端到端的优化,导致在速度、统一性和多任务处理方面能力较弱;而 YOLO v10 通过引入轻量化稀疏检测头、可分离解码器和统一的架构设计,在保持高精度的同时大幅提升了推理效率与多任务表现。现有无人机羊群检测模型面临目标像素少、光照条件复杂、背景信息冗余及计算资源受限等问题,无法在精度、参数量、计算量和速度之间取得良好平衡,且难以在嵌入式设备上高效运行。

针对上述问题,本文提出一种轻量级 YOLO v10s 改进算法 MLL-YOLO v10s,用于解决无人机图像中的羊只目标检测问题。

1 材料与方法

1.1 数据集采集与构建

1.1.1 数据来源

本文构建的专用数据集实地拍摄于内蒙古自治区锡林郭勒盟锡林浩特市白音锡勒牧场黄花树特分场(116°37′59″ E,44°0′59″N),拍摄时间为 2023 年夏季。为获取羊群自然状态下的数据,选择天然放牧的羊群作为拍摄主体,运用大疆 Mavic3E 型无人机实施航拍数据采集。影像采集在晴朗天气条件下进行,以规避光照对图像的干扰,同时兼顾地形多样性,分别在牧场平地、缓坡等不同地貌区域获取影像数据。航拍过程中采用多高度层复合采集策略,飞行高度区间设置为 20 ~ 60 m,共获取 19 段原始视频数据。每段影像时长控制在 1 ~ 3 min,视频采用 MP4 格式存储,原始分辨率为 4K 级别(3 840 像素 × 2 160 像素),帧速率保持 30 f/s。图 1 展示了不同航拍高度下的画面截取效果。本文使用基于 Python 的影像分解系统,将连续视频流转换为独立图像帧,单段视频可生成 300 ~ 900 幅静态图像。在此基础上,通过图像旋转、尺寸变换、明暗度调节、局部裁剪及噪声叠加等处理手段进行数据增强,最终形成包含 12 942 幅高质量图像的专用数据集。

1.1.2 数据预处理

为确保数据图像的清晰度和多样性,通过双重

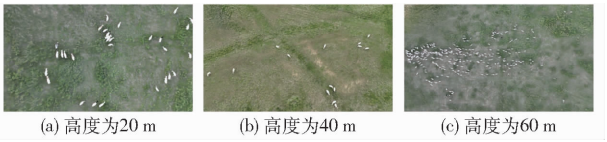


图 1　无人机不同高度采集的羊群视频示例帧图像

Fig. 1　Example frame images from sheep herd videos captured by drone at different altitudes

质量筛选机制优化数据样本。在图像清晰度评估阶段，基于拉普拉斯变换原理构建数字滤波器计算所得方差值作为清晰度量化指标。设定清晰度阈值为 70，筛除低于该标准的模糊影像。针对图像冗余问题，采用深度特征比对策略，设定相似性判别阈值为 0.8，对特征相似度超出该临界值的重复样本执行去重操作。经自动化筛选后最终获得 7 340 幅有效图像。数据集构建采用系统化编码方案，所有图像按 0001～7340 顺序进行唯一标识。图 2 呈现了数据集部分图像情况。

采用 Labelme 开源标注平台对图像数据进行目标标注。标注过程中，通过矩形框精确定位羊体轮廓，同时记录每幅影像的数字化特征参数，标注成果以 XML 格式存储，图 3 展示了标注样例。

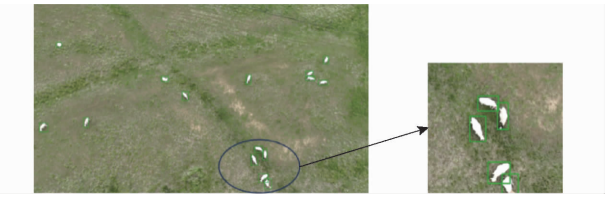


图 2　羊群目标检测数据标注样例

Fig. 2　Examples of annotated data for sheep herd target detection

1.2　基于 MLL－YOLO v10s 的羊只检测模型

YOLO 系列目标检测算法通过不断优化网络架构和改进算法设计，在目标检测领域取得了显著进展。从 YOLO v1 到 YOLO v10，每一代版本都在检测精度、处理速度和模型轻量化方面进行了改进。YOLO v10 网络架构在继承前代模型优点的基础上，进一步优化了主干网络的特征提取能力，引入了更高效的检测头设计，并结合了多尺度预测和注意力机制等技术，从而在目标检测精度与处理效率方面具有显著优势。这种优化使其在保持实时检测能力的同时，能够更好地适应复杂场景下的目标检测需求。YOLO v10 的主干网络整合了多层级卷积单元、C2f 特征融合组件及空间压缩下采样模块（SCDown），这些结构设计有效提升了特征表达能力与模型训练效果。但需要关注的是，多尺度卷积层在迭代过程中容易生成重复性特征信息，这种架构特性可能导致模型体积膨胀，并对实时检测性能产生制约，无法在嵌入式设备上高效运行。此外，在航拍影像中，羊只个体通常呈现目标尺寸微小、空间分布密集以及动态遮挡频繁等特征，这些特性导致漏检测和误检测严重。因此，加强模型对小目标羊只的特征学习能力至关重要。为解决上述问题，本文改进 YOLO v10s 模型以增强对无人机图像中羊只的检测能力，重点解决小目标辨识、遮挡干扰和群体密集问题，减少羊只误检测和漏检测数量。

首先，YOLO v10 的原有主干网络采用了较深的网络结构，包含多个卷积层和可能的注意力机制，以提高特征提取能力，但该结构的计算量和参数量较大。相比之下，MobileNetV4 采用了轻量化设计，如深度可分离卷积和倒置残差块，减少了计算量和参数量。通过比较，MobileNetV4 的总参数量显著减少，计算效率得到提升。因此采用 MobileNetV4 网络作为改进 YOLO v10 的特征提取主干网络，降低算法参数量，提升算法计算效率。相较于以往的 MobileNet 网络替代 YOLO v5 网络中的特征提取模块方法，更加注重效率与性能的平衡，以较少的计算代价增加网络的深度和感受野并最大化计算利用率，使其对具有轻量化需求的识别任务更具备优势。其次，在 SPPF（Spatial pyramid pooling-fast）模块引入创新性的大核可分离注意力模块（LSKA），该设计通过融合空间感知与通道注意力机制，显著强化对羊只关键体征的表征能力。这种混合注意力架构在保持计算效率的同时，提升密集场景下的目标区分度，同时将降低跨场景迁移学习的泛化误差。最后，为增加模型的检测速度，提升模型的可部署性，引入了一种 LSCD 轻量化目标检测头方法，通过使用共享卷积，可以大幅减少参数数量，这使得模型更轻便，更适用于资源受限的嵌入式设备中。改进后的 MLL－YOLO v10s 模型的结构如图 3 所示。

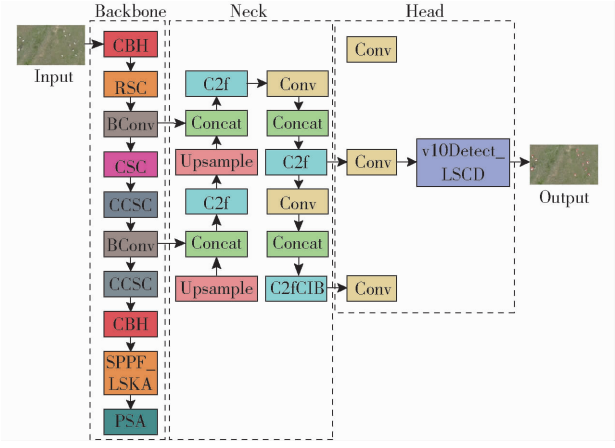


图 3　MLL－YOLO v10s 羊只检测模型结构

Fig. 3　Architecture of MLL－YOLO v10s sheep detection model

1.2.1 MobileNetV4 - backbone

MobileNetV4 是移动设备神经网络架构,旨在为移动生态系统提供通用且高效的模型。MobileNetV4 通过引入通用反向瓶颈(UIB)搜索块和 MobileMQA 注意力模块[11],实现了模型的轻量级化,UIBblocks 结构如图 4 所示。UIB 模块是一种统一且灵活的结构,扩展了 MobileNetV2 引入的倒置瓶颈(IB)块,融合了 IB、ConvNext、前馈网络(FFN)和额外深度可分(ExtraDW)变体。这种设计使得网络架构更加灵活多变,能够更好地适应不同的任务需求。MobileMQA 机制将简单性置于优化性能的首位,专为移动加速器设计,能带来显著的速度提升。这使得 MobileNetV4 在确保优异性能表现的基础上,大幅压缩了参数规模与运算负担。得益于创新设计的 NAS 方案,该模型将粗粒度和细粒度搜索相结合,使网络构建效率显著提升的同时,网络结构也得到深度优化。这种双重优化机制确保了该模型在各类计算设备上都能实现最佳运行效能。

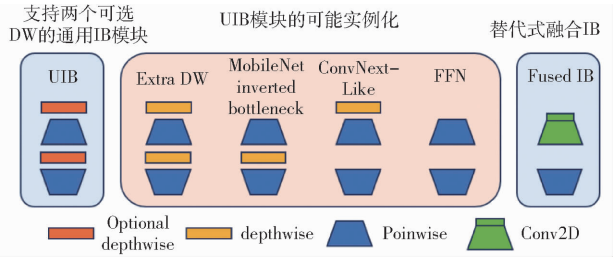


图 4 UIBblocks 结构

Fig. 4 UIBblocks structure

本文通过引入 MobileNetV4 作为 MLL - YOLO v10s 的主干网络,有效压缩了模型的参数规模与运算量,优化了架构的轻量化特性,使算法在移动端与嵌入式平台上展现出更强的部署适应性。

1.2.2 基于 LSKA 的 SPPF - LSKA 模块

在羊只目标检测场景中,复杂背景干扰、光照波动、多尺度目标及密集遮挡等环境变量,对微小目标的检测精度与处理效率提出更高的要求。随着网络深度增加,羊只小尺寸目标的特征表征能力逐渐衰减,导致检测模型出现漏判与误判现象。针对这些技术瓶颈,在经典 SPPF 模块中整合了大型可分离核注意力机制(Large separable kernel attention, LSKA)[12],通过强化特征信息的关键维度提取能力来优化模型性能。由此构建的 SPPF - LSKA 增强架构(图 5、6)有效提升了目标的判别性表征。

LSKA 大卷积核分解使用 2 个一维卷积 $\mathrm{Conv}_{k\times 1}$ 和 $\mathrm{Conv}_{1\times k}$ 代替二维卷积:

$$\mathrm{Conv}_{k\times k}\approx \mathrm{Conv}_{k\times 1}\circ \mathrm{Conv}_{1\times k} \tag{1}$$

这样可以将二维大卷积核(如 21 × 21)分解为 2 个一维卷积,大大减少参数和计算量。

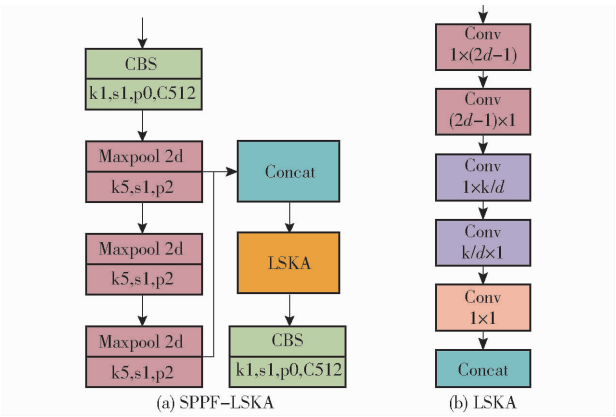


图 5 SPPF - LSKA 和 LSKA 结构

Fig. 5 SPPF - LSKA and LSKA structure

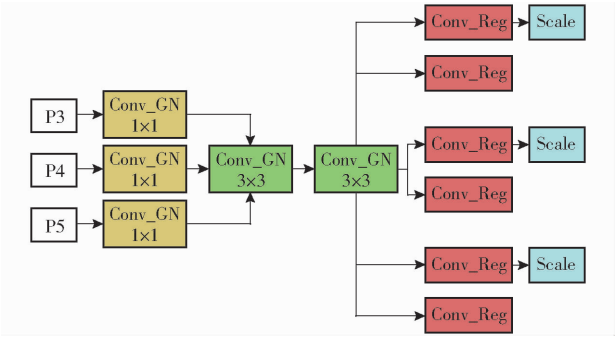


图 6 LSCD 检测头结构图

Fig. 6 LSCD detection head structure diagram

假设输入特征图 $X\in \mathbf{R}^{C\times H\times W}$,LSKA 的流程如下:

(1)对输入施加 2 个一维卷积

$$Y_1=\mathrm{Conv}_{1\times k}(X) \tag{2}$$

$$Y_2=\mathrm{Conv}_{k\times 1}(X) \tag{3}$$

记为组合卷积

$$Y=\mathrm{DWConv}_{k\times 1}(\mathrm{DWConv}_{1\times k}(X)) \tag{4}$$

其中 DWConv 表示 depthwise 卷积,每个通道独立处理。

(2)对卷积结果 Y 进行注意力加权

$$\mathrm{Attention}(Y)=\sigma(\mathrm{Conv}_{7\times 7}(Y))Y \tag{5}$$

其中 σ 为 sigmoid 函数。

(3)最终输出

$$Z=\mathrm{Attention}(Y) \tag{6}$$

得到 LSKA 的输出,具备长距离建模能力且计算高效。

在特征图输入 SPPF - LSKA 增强模块时,多尺度池化操作首先对特征信息进行空间压缩与维度精炼,有效抑制噪声干扰与冗余特征,同步扩展模型感受野。随后通过连接层将区域化特征进行合并,继而传入 LSKA 模块。该模块实施深度卷积核分解技术(将二维卷积拆分为水平/垂直方向的一维卷积)与多级卷积串联策略[13],在保持目标显著性特征的同时,强化模型对羊只图像特征的捕获能力,显著提

升羊只图像理解能力。其计算公式为

$$Z_1^C = \sum_{H,W} W^C_{(2d-1)\times 1} * (W^C_{(2d-1)} * F_1^C) \quad (7)$$

$$Z_2^C = \sum_{H,W} W^C_{\frac{k}{d}\times 1} * (W^C_{1\times\frac{k}{d}} * Z_1^C) \quad (8)$$

$$A^C = W_{1\times 1} * Z_2^C \quad (9)$$

$$F^C = A^C \otimes F_1^C \quad (10)$$

式中　H——特征图高度

W——特征图宽度

$*$——卷积　　$\otimes$——哈达玛乘积

d——扩张率　　k——内核大小

F_1^C——输入特征图

Z_1^C——深度卷积输出，由输出特征图 F_1^C 和大小为 $1\times(2d-1)$、$(2d-1)\times 1$ 的卷积得到

Z_2^C——全局空间信息，由 Z_1^C 和大小为 $\frac{k}{d}\times 1$、$1\times\frac{k}{d}$ 的扩张深度卷积得到

A^C——注意力图谱，由 Z_2^C 与 1×1 内核卷积得到

F^C——注意力图谱 A^C 和特征图谱 F_1^C 的哈达玛乘积

1.2.3　LSCD 轻量化检测头

共享卷积是卷积神经网络（CNN）中的一个核心概念。在 CNN 中，卷积层由一系列的卷积核（或称为过滤器、滤波器）组成，这些卷积核与输入图像的不同位置进行卷积运算，生成对应的特征图。共享卷积的核心优势在于其权重共享的机制，这一机制显著减少了网络需要训练的参数数量，从而提高了计算效率，并降低了过拟合的风险。通过保持卷积核对整个图像的权重一致性，共享卷积能够更有效地提取图像中的局部特征，同时保持空间结构信息。此外，其平移不变性使得模型对输入数据的平移变换具有鲁棒性，进一步增强了模型的泛化能力[14]。

本文设计了一种基于群归一化 GroupNorm[15] 和共享卷积的轻量化检测头（Lightweight shared convolutional detection，LSCD），其网络结构如图 6 所示。LSCD 检测头充分利用了 GroupNorm 和共享卷积的优势，能够在保持特征信息有效融合的情况下，减少计算量和运算复杂度。通过引入共享卷积，卷积层的重复计算模块得以精简，这不仅显著压缩了模型体积，更为移动端部署提供了可行性。针对多尺度检测任务中的特征适配，采用可调节的特征缩放层对输出结果进行缩放，克服不同检测头对应目标尺寸差异带来的干扰。这种复合优化策略在模型运算效率与识别准确率之间实现了较好平衡，使轻量级检测器在降低资源消耗的同时，仍能维持可靠的性能表现。

在图 6 所示的 LSCD 检测头网络结构中，GN_Conv 1×1 模块表示的是群归一化 GroupNorm + 卷积 Conv 操作，其中 1×1 表示卷积核大小为 1×1。2 个群归一化卷积模块 GN_Conv 3×3 共享权重，3 个预测框卷积模块 Conv_Box 共享权重，3 个分类卷积模块 Conv_Cls 共享权重。在每个 Conv_Box 模块后面连接着 Scale 模块（Scale 模块主要是一个缩放系数，用来匹配对不同尺度目标的检测）。

1.3　评价指标

在评估目标检测模型性能时，采用多维度的量化评价体系，主要包括 5 个指标：精度（Precision，P）、全类平均精度（Mean average precision，mAP）、平均帧率（Frames per second，FPS）、参数量（Params）以及计算量（FLOPs）。其中，FPS 反映系统的实时处理能力，具体指每秒可分析的图像帧数；Params 表征网络结构的复杂程度；FLOPs 则体现单次前向传播所需的浮点运算次数，用于衡量模型的计算量。AP 的计算依据精度和召回率（Recall，R）曲线围成的面积，通过积分运算获得该指标数值。而 mAP 作为综合性评价标准，考虑了检测算法的召回率和准确率，将所有类别的 AP 计算结果进行加权平均，形成全局性能评估。

2　结果分析

2.1　实验环境及参数设置

本实验所用的硬件设备及参数如表 1 所示。训练数据集包含 5 138 幅羊只图像，同时划分 734 幅样本构建验证集，用于模型性能评估和超参数调优。输入图像统一缩放至 640 像素 ×640 像素，每批次处理 32 幅图像，选择 SGD（Stochastic gradient descent）优化器，动量参数设为 0.937，初始学习率设为 0.001。在此硬件配置与训练方案下，模型进行 200 轮迭代训练。

表 1　无人机羊只目标检测实验硬件参数配置

Tab.1　Hardware configuration for UAV-based sheep object detection experiment

硬件配置	参数
CPU	AMD（R）Ryzen7 5800X CPU@3.80 GHz
GPU	NVIDIA RTX 3080 Ti（12 GB）
内存容量/GB	32
操作系统	Windows 10
开发工具	Microsoft Visual Studio Code 1.75.1
CUDA 版本	12.7
深度学习框架	PyTorch 2.2.0

2.2 消融实验

为评估算法改进方案的有效性,设计消融实验对 MobileNet 主干网络、SPPF - LSKA 模块和 LSCD 检测头进行验证。在保持实验环境一致性的前提下,以 YOLO v10s 作为基线模型,逐步引入各优化组件,评估各模块对网络性能的影响,评估结果如表 2 所示。

表 2 无人机羊只目标检测消融实验

Tab. 2 Ablation test on UAV-based sheep object detection

MobileNetV4	LSKA	LSCD	精度/%	全类平均精度/%	计算量	平均帧率/(f·s^{-1})	参数量
			92.3	90.2	2.16×10^{10}	110	7.200×10^{6}
√			93.0	91.7	2.27×10^{10}	137	9.940×10^{6}
	√		94.2	92.2	3.39×10^{10}	86	1.349×10^{7}
		√	92.7	90.9	2.05×10^{10}	118	6.870×10^{6}
√	√		95.8	93.4	3.06×10^{10}	114	1.486×10^{7}
√		√	93.2	91.5	2.49×10^{10}	141	1.191×10^{7}
	√	√	94.9	92.8	3.51×10^{10}	90	1.673×10^{7}
√	√	√	95.8	93.6	2.63×10^{10}	135	1.268×10^{7}

注:√表示采用此方式,下同。

从表 2 可知,在 YOLO v10s 模型上将 backbone 替换为 MobileNetV4 后,模型的平均帧率增加了 27 f/s,参数量增加了 2.74×10^{6},计算量增加了 1.1×10^{9},可见采用 MobileNetV4 网络作为模型的 backbone,尽管计算量有所增加,但通过其优化的网络结构和高效的计算策略,在处理速度上仍实现了提升。引入 LSKA 模块后,精度提升 1.9 个百分点,达 94.2%,全类平均精度增长 2 个百分点,达 92.2%,证明该组件能提高羊只特征捕获能力,特别是在通道维度增强小目标聚焦能力。但同时平均帧率下降 24 f/s,计算量增加 1.23×10^{10}。与之形成对比的是 LSCD 轻量化检测单元的应用,在精度值增加 0.4 个百分点的基础上,平均帧率提升 8 f/s,运算负荷减少 1.1×10^{9},直观验证了共享卷积结构在维持特征提取能力的同时优化计算效率的优势。当集成 MobileNetV4 主干网络与 LSKA 模块时,模型精度提升 3.5 个百分点,虽然模型计算量增加 3.73×10^{6} 参数量增加 2×10^{9},但平均帧率逆势提高 4 f/s。后续引入 LSCD 检测单元时,模型精度保持稳定,却成功缩减计算资源占用并加速推理过程。最终形成的MLL - YOLO v10s 架构在无人机羊只检测场景中精度达 95.8%、全类平均精度达 93.6%,较基准模型分别提升 3.5、3.4 个百分点,在保持 135 f/s 平均帧率的同时,将计算量压缩至 2.63×10^{10},为部署无人机边缘计算设备实现实时目标检测提供了可行性保障。

2.3 对比实验结果分析

2.3.1 不同检测模型对比

通过与多个典型的目标检测模型进行对比验证改进算法的有效性。选择单阶段(SSD、RetinaNet)、双阶段(Faster R - CNN)及 YOLO 系列(v8s/v10s)的典型检测架构,与提出的 MLL - YOLO v10s 进行无人机小目标羊只检测性能对比。图 7 展示了各模型在测试集上的综合表现,表 3 则量化呈现全类平均精度、平均帧率和训练生成模型文件大小等核心指标。实验数据显示,相较于 Faster R - CNN 双阶段检测器,改进模型全类平均精度提升 30.9 个百分点,平均帧率加快 122 f/s,模型文件大小缩减 303 MB。对比 RetinaNet 与 SSD 单阶段算法,全类

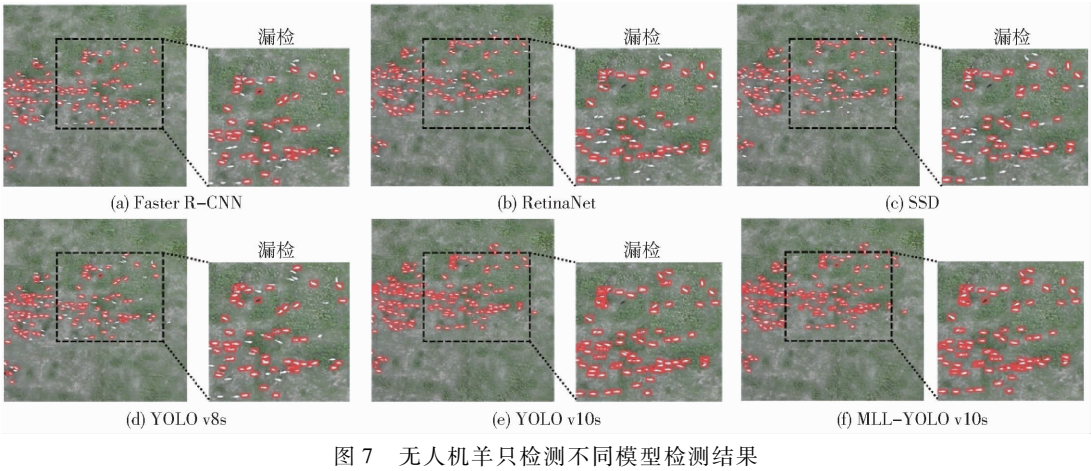


图 7 无人机羊只检测不同模型检测结果

Fig. 7 UAV sheep detection results with different models

表 3　不同网络模型的羊只检测实验结果

Tab. 3　Experimental results of sheep detection with different network models

模型	全类平均精度/%	平均帧率/(f·s^{-1})	模型文件大小/MB
Faster R－CNN	62.7	13	327
RetinaNet	75.4	17	239
SSD	57.1	56	103
YOLO v8s	87.9	105	23
YOLO v10s	90.2	110	28
MLL－YOLO v10s	93.6	135	24

平均精度优势分别达 18.2、36.5 个百分点，平均帧率提高 118、79 f/s，模型文件压缩 215、79 MB。与 YOLO v8s 以及 YOLO v10s 模型相比，虽然参数规模略有增加，但全类平均精度分别提升 5.7 个百分点和 3.4 个百分点，同时平均帧率也得到改善。分析表明，MLL－YOLO v10s 在检测精度（mAP 95.8%）与运算效率（135 f/s）之间实现了更优平衡，其 1.268×10^{7} 的轻量级参数规模显著优于传统双阶段方法，较单阶段算法也展现出明显优势。解决了高空拍摄场景下小目标检测精度与实时性难以兼顾的痛点，为无人机羊只检测提供了高性价比的技术路径。

根据图 7 的检测结果，本文提出的 MLL－YOLO v10s 模型在小目标聚集场景下未出现漏检现象，表现出较高的检测精度与鲁棒性。相比之下，单阶段目标检测算法（如 SSD 和 RetinaNet）、双阶段目标检测算法（如 Faster R－CNN）以及 YOLO v8s 模型均在该场景下出现了不同程度的漏检。而 YOLO v10s 模型虽成功检测出所有白色羊只，但在其他毛色羊只的检测中存在漏检，表明其在特殊场景下的鲁棒性有待进一步提升。

为了判断数据差异是否真实存在，而非随机波动，从而验证研究假设的可靠性，本文增加了显著性检验（如 t 检验）以验证性能提升的统计意义。重复进行 5 次相同的实验，将 5 次实验结果计算假定值（P 值）来验证结果是否具有显著性，显著性检验结果见表 4。由表 4 可知假定值均小于 0.05，表明研究结果具有统计显著性。

表 4　显著性检验结果

Tab. 4　Significance test results

模型	第 1 次实验		第 2 次实验		第 3 次实验		第 4 次实验		第 5 次实验		假定值	
	全类平均精度/%	平均帧率/(f·s^{-1})	全类平均精度/%	平均帧率/(f·s^{-1})	全类平均精度/%	平均帧率/(f·s^{-1})	全类平均精度/%	平均帧率/(f·s^{-1})	全类平均精度/%	平均帧率/(f·s^{-1})	全类平均精度	平均帧率
Faster R－CNN	62.7	13	61.3	15	63.5	11	62.4	16	63.8	14	0.018	0.012
RetinaNet	75.4	17	73.6	19	77.2	14	74.3	21	76.8	15	0.025	0.031
SSD	57.1	56	55.3	53	58.7	58	56.4	54	59.2	59	0.031	0.013
YOLOv8s	87.9	105	86.3	103	88.5	107	87.2	102	88.7	109	0.028	0.031
YOLOv10s	90.2	110	91.5	108	89.7	113	90.8	107	91.1	114	0.007	0.033
MLL－YOLOv10s	93.6	135	94.2	133	92.5	137	93.1	133	94.7	138	0.015	0.026

同时，为了便于与近年来在 VisDrone 数据集上发布的实时检测模型进行对比，本文也使用 VisDrone2019 数据集进行训练。为了确保各模型的深度和宽度一致，所有对比模型均选取后缀为“s”的 small 版本进行实验，实验结果如表 5 所示。

与该数据集上发布的其他改进算法相比，MLL－YOLO v10s 保持了最低的参数量和计算量，并且识别精度与主流先进算法处于同等量级。MLL－YOLO v10s 仅以不到 64% 的计算量，就达到了近似 DMFF－YOLO 的 mAP50，进一步验证了本文模型改进的有效性和技术优势。

表 5　改进模型的对比实验结果

Tab. 5　Comparison of experimental results with improved models

模型	全类平均精度/%	参数量	计算量
YOLO v8s－CEBI[16]	40.6	5.6×10^{6}	2.09×10^{10}
CA－YOLO v8[17]	48.8	7.4×10^{6}	4.0×10^{10}
DMFF－YOLO[18]	51.3	7.7×10^{6}	3.52×10^{10}
LW－YOLO[19]	42.9	5.58×10^{6}	2.36×10^{10}
CM－YOLO v8s[20]	45.9	3.49×10^{6}	3.12×10^{10}
改进 YOLO v8s[21]	43.1	1.21×10^{6}	
PECS－YOLO[22]	36.8	2.6×10^{6}	1.19×10^{10}
改进 YOLO v9c	39.3	3.0×10^{6}	1.31×10^{10}
RT－DETR	45.1	3.7×10^{6}	2.79×10^{10}
NanoDet	40.7	1.8×10^{6}	8.4×10^{9}
PP－YOLOE[23]	39.8	1.5×10^{6}	7.9×10^{9}
MLL－YOLO v10s（本文模型）	48.3	2.43×10^{6}	2.36×10^{10}

2.3.2　不同注意力模块对比

本文集成大型可分离核注意力 LSKA 模块，增强模型对对象形状的学习能力，提高鲁棒性和泛化能力。为验证 LSKA 注意力模块在无人机密集羊群检测中的优势，系统对比了多种主流注意力机制改

进方案(表6)。LSKA 在实现长距离依赖建模的同时,采用可分离卷积有效控制了计算复杂度,使其在精度上达到最高,同时保持良好的速度。相比之下,NAM 和 CBAM 虽有较好精度提升,但结构更复杂、推理速度受限;SimAM 与 NAM 虽然轻量快速,但性能提升较有限。因此,LSKA 模块在结构效率与精度表现上最为均衡,适用于对精度与速度均有要求的主流视觉任务。而 SPPF - LSKA 创新性地引入形状先验学习机制,通过注意力权重调整使网络更关注目标轮廓特征而非表面纹理,这一改进有助于提高模型对对象形状的学习能力,从而提高模型的鲁棒性和泛化能力。

表 6 不同注意力机制实验对比结果

Tab. 6 Experimental comparison of different attention mechanisms

模型	精度/%	全类平均精度/%	计算量	平均帧率/(f·s^{-1})
原始 YOLO v10s	87.3	85.1	2.92×10^{10}	93
SE	90.8	88.2	3.03×10^{10}	88
SimAM	90.2	87.7	2.99×10^{10}	95
NAM	90.1	87.3	2.97×10^{10}	97
CBAM	92.6	91.1	3.00×10^{10}	85
Dynamic Head	93.3	92.4	3.04×10^{10}	88
LSKA	93.7	91.5	3.16×10^{10}	89
SPPF - LSKA	94.2	92.2	3.39×10^{10}	86

2.3.3 改进 SPPF 模块的性能分析

本文针对 YOLO v10 的特征提取架构进行创新优化,重点改进 SPPF 模块的尺度感知能力。为验证 SPPF - LSKA 特征提取模块在无人机密集羊群检测中的优势,系统对比了多种主流改进方案(表7)。通过集成大核分离卷积注意力机制,构建的 SPPF - LSKA 复合架构使精度提升 2.2 个百分点,全类平均精度增长 1.9 个百分点。该方法采用特征融合机制整合不同层级的语义信息,在增加少量参数的前提下显著增强小目标识别能力。尤其在羊群密集和遮挡严重的情况下,增加了羊群检测的准确性,降低了羊群目标的漏检和误检率。对比实验显示:基于膨胀卷积的 ASPP 模块虽能扩展感受野,但对密集小目标易产生特征混淆;采用轻量激活函数的 SimSPPF 方案虽提升了推理速度,但关键特征捕获能力严重下降;而 SPPCSPC 架构因多分支池化操作导致计算冗余,运算效率低。本方案创新性地引入形状先验学习机制,通过注意力权重调整使网络更关注目标轮廓特征而非表面纹理,这一改进有助于提高模型对对象形状的学习能力,从而提高模型的鲁棒性和泛化能力。实际部署数据显示,优化后的特征提取架构在维持 86 f/s 实时处理能力的同时,将计算量控制在 3.39×10^{10} 以满足实时性的要求。

表 7 不同 SPPF 模块实验对比结果

Tab. 7 Comparison results of different SPPF modules in experiments

模型	精度/%	全类平均精度/%	计算量	平均帧率/(f·s^{-1})	参数量
SPPF	92.0	90.3	2.86×10^{10}	105	1.113×10^{7}
ASPP	92.7	90.9	3.50×10^{10}	84	1.979×10^{7}
SimSPPF	92.4	90.5	2.86×10^{10}	108	1.247×10^{7}
SPPCSPC	93.1	91.7	3.34×10^{10}	88	1.594×10^{7}
SPPF - LSKA	94.2	92.2	3.39×10^{10}	86	1.349×10^{7}

3 讨论

本文以 YOLO v10s 为基线模型,创新性地构建了 MLL - YOLO v10s 检测模型,该方案在维持轻量化的同时,显著提升了识别精度,有效改善了无人机在复杂环境及多变光照条件下执行地面羊只检测任务时出现的漏检与误判问题。实验表明,当目标距离较远、分布密集或存在相互遮挡时,原检测模型的识别性能显著下降,为此构建了大核分离卷积注意力模块,提高模型对对象形状的学习能力,从而提高模型的鲁棒性和泛化能力。融入注意力机制,关注模型的全局信息提升对小目标的检测能力,并使用轻量化检测头 LSCD 保证模型的检测速度。

为了进一步验证改进的有效性,使用热力图对比基线系统与各项改进后的模型,热力效果如图 8 所示,图像中共有 88 只羊,可以看到基线系统存在目标漏检,且注意力范围扩大未有效焦距被检测目标上;在引入 MobileNetV4 - backbone 后,检测性能并没有因为参数量与计算量的降低而大幅下降;在引入 LSKA 注意力机制后,模型对小目标感兴趣区域更为集中,能够识别到更多有效信息,特征提取能力大幅提高;替换检测头为 LSCD 后,在保持原有特征提取能力的前提下,模型进一步轻量化。定量分析结果见表 8。

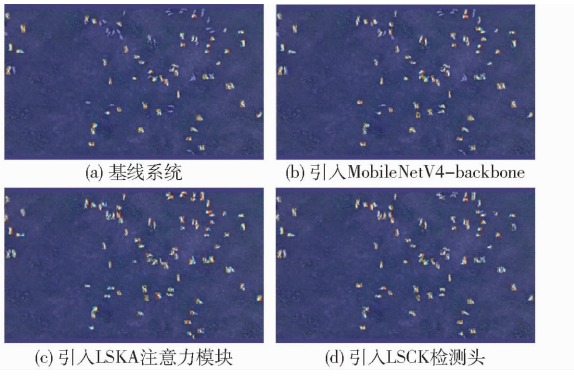


图 8 热力图对比

Fig. 8 Heatmap comparison

表 8　热力图结果分析

Tab. 8　Heatmap result analysis

MobileNetv4	LSKA	LSCD	聚焦羊只数	IoU
			54	0.73
√			61	0.79
√	√		83	0.93
√	√	√	88	0.97

虽然 MLL－YOLO v10s 在密集羊群与遮挡场景中展现出优异的识别性能，但当部署至岩石地貌等特殊环境时，模型误检率有所上升。经分析发现，当前训练数据集以白/灰色系羊只与绿色草原构成的强色差对比为主要特征，这种特定环境下的数据优势在提升常规场景检测精度的同时，也导致模型在非典型地貌中易受相似色块干扰。当白色羊只在牧草资源遭受过度采食的裸露土地背景下或深色羊只在牧草茂盛的深色背景下时，检测目标与背景颜色对比度降低，可能会有漏检的情况出现。针对这一局限，一方面可以扩展非典型地貌的样本采集，另一方面引入图像处理技术合成高仿真训练素材。通过结合实地数据获取与算法生成手段，能够有效增强数据集的场景多样性，从而提升模型在特殊环境下的鲁棒性。

除了对上述问题进行探索，未来仍有需改进的方面。首先，本文的实验对象为乌珠穆沁羊，还应通过增加不同品种羊的图像数据来增加模型的应用场景，提高模型的检测能力。其次，本文数据集只涵盖了白天晴天天气下的航拍羊群数据，当天气为雨雾天气，由于能见度降低，拍摄的图像对比度降低；当为夜晚环境时，光线较弱，拍摄的图像噪点较多，以上因素都会对检测结果有较大影响。针对这一局限性，可以对已有数据集进行降低对比度、降低色阶等后期处理增加数据集的多样性，从而提升模型在特殊环境下的鲁棒性。再者，本文只针对了牧区中的羊进行研究，并未对同在牧区放牧的牛、马等牲畜进行对象检测研究，应增加上述研究，提高模型在草原牧区的应用范围。最后，为进一步提高应用，实现牧场的智能化放牧，将模型部署到装有边缘计算设备的无人机上，通过无人机巡检实现牧场放牧牲畜的自动检测和跟踪。

4　结论

（1）本文针对无人机对地目标检测中存在目标小，密集以及遮挡严重导致漏检和误检以及模型计算量大、参数多、实时性差的问题，构建了基于无人机图像的羊群数据集，提出一种用于天然草原放牧羊只实时检测的深度学习模型 MLL－YOLO v10s，获得较好的检测结果。

（2）以 YOLO v10s 模型作为基线模型，引入大型可分离核注意力（LSKA）模块，提高模型的鲁棒性和泛化能力，增强模型的关键特征提取能力。利用 MobileNetV4 网络替代 YOLO v10 网络中的特征提取模块缩减模型的参数量和计算量。再利用轻量化检测头 LSCD，在显著压缩网络规模的前提下实现了精度可控性优化。

（3）基于自建数据集开展的验证测试显示，优化模型达 93.6% 的全类平均精度，在保持平均帧率 135 f/s 的同时，将参数量与计算量分别控制在 1.268×10^{7} 和 2.63×10^{10}。实际应用测试证实，该方案在无人机航拍场景下对高密度聚集及部分遮挡的羊只目标展现出更强的抗干扰检测能力，在参数量和计算量上也有明显优势。

（4）本文提出的模型有效地解决了无人机检测任务中羊只漏检和误检现象严重的问题，并且模型计算量小、参数少、速度快、实时性好，为天然牧场无人机巡检羊群提供了一种有效的实时检测方法，为边缘计算设备与无人机结合应用提供支持。

参　考　文　献

[1] LELIVELD L M C, PROVOLO G. A review of welfare indicators of indoor-housed dairy cow as a basis for integrated automatic welfare assessment systems[J]. Animals, 2020, 10(8): 1430.

[2] AQUILANI C, CONFESSORE A, BOZZI R, et al. Review: precision livestock farming technologies in pasture-based livestock systems[J]. Animals, 2022, 16(1): 100429.

[3] 毕文浩, 苍岩. 基于 YOLOv5 的轻量级猪只检测网络设计[J]. 应用科技, 2023, 50(1): 40－44.
BI Wenhao, CANG Yan. Design of lightweight pig detection network based on YOLOv5[J]. Applied Science and Technology, 2023, 50(1): 40－44. (in Chinese)

[4] ZHANG X, XUAN C, XUE J, et al. LSR－YOLO: a high-precision, lightweight model for sheep face recognition on the mobile end[J]. Animals, 2023, 13(11): 1824.

[5] ZHAO H K, MAO R, LI M, et al. SheepInst: a high-performance instance segmentation of sheep images based on deep learning[J]. Animals, 2023, 13(8): 1338.

[6] WANG Y, MA L, WANG Q, et al. A lightweight and high-accuracy deep learning method for grassland grazing livestock detection using UAV imagery[J]. Remote Sensing, 2023, 15(6): 1593.

[7] WANG X, HU Y, WANG M, et al. A real-time lightweight behavior recognition model for multiple dairy goats[J]. Animals,

2024, 14(24): 3667.
[8] 王旺, 王福顺, 张伟进, 等. 基于改进 YOLO v8s 的羊只行为识别方法[J]. 农业机械学报, 2024, 55(7): 325 - 35, 44.
WANG Wang, WANG Fushun, ZHANG Weijin, et al. Sheep behavior recognition method based on improved YOLO v8s[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(7): 325 - 35, 44. (in Chinese)
[9] WANG J, ZHAI Y H, ZHU L, et al. PD - YOLO: a study of daily behavioural detection in housed sheep[J]. Plos One, 2024, 19(11): e0313412.
[10] 翁智, 刘海鑫, 郑志强. CSD - YOLOv8s:基于无人机图像的密集小目标羊只检测模型[J]. 智慧农业(中英文), 2024, 6(4): 42 - 52.
WENG Zhi, LIU Haixin, ZHENG Zhiqiang. CSD - YOLOv8s: dense sheep small target detection model based on UAV images [J]. Smart Agriculture, 2024, 6(4): 42 - 52. (in Chinese)
[11] QIN D, LEICHNER C, DELAKIS M, et al. MobileNetV4-universal models for the mobile ecosystem[C]//Proceedings of the European Conference on Computer Vision, F, 2024.
[12] LAU K W, PO L M, REHMAN Y A U. Large separable kernel attention: rethinking the large kernel attention design in CNN [J]. Expert Systems with Applications, 2024, 236: 121352.
[13] 王福顺, 王旺, 孙小华, 等. 基于改进 YOLO v8n-seg 的羊只实例分割方法[J]. 农业机械学报, 2024, 55(8): 322 - 32.
WANG Fushun, WANG Wang, SUN Xiaohua, et al. Sheep instance segmentation method based on improved YOLO v8n-seg [J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(8): 322 - 32. (in Chinese)
[14] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84 - 90.
[15] WU Y X, HE K M. Group normalization[J]. International Journal of Computer Vision, 2020, 128(3): 742 - 755.
[16] 胡惠娟, 秦一锋, 徐鹤, 等. 面向无人机航拍图像的 YOLOv8 目标检测改进算法[J]. 计算机科学, 2025, 52(4): 202 - 211.
HU Huijuan, QIN Yifeng, XU He, et al. An improved YOLOv8 object detection algorithm for UAV aerial images[J]. Computer Science, 2025, 52(4): 202 - 211. (in Chinese)
[17] 付锦燚, 张自嘉, 孙伟, 等. 改进 YOLOv8 的航拍图像小目标检测算法[J]. 计算机工程与应用, 2024, 60(6): 100 - 109.
FU Jinyi, ZHANG Zijia, SUN Wei, et al. Improved YOLOv8 small target detection algorithm in aerial images[J]. Computer Engineering and Applications, 2024, 60(6): 100 - 109. (in Chinese)
[18] QIU X, CHEN Y, SUN C, et al. DMFF - YOLO: YOLOv8 based on dynamic multiscale feature fusion for object detection on UAV aerial photography[J]. Ieee Access, 2024, 12: 125160 - 9.
[19] 廖宁生, 曹天秀, 刘科言, 等. 复合特征与多尺度融合的无人机小目标检测算法[J]. 计算机工程与应用, 2025, 61(3): 111 - 120.
LIAO Ningsheng, CAO Tianxiu, LIU Keyan, et al. Small target detection algorithm for UAV based on composite feature and multi-scale dusion[J]. Computer Engineering and Applications, 2025, 61(3): 111 - 120. (in Chinese)
[20] 李姝, 李思远, 刘国庆. 基于 YOLOv8 无人机航拍图像的小目标检测算法研究[J]. 小型微型计算机系统, 2024, 45(9): 2165 - 2174.
LI Shu, LI Siyuan, LIU Guoqing. Research on small target detection algorithm based on YOLOv8 UAV aerial images[J]. Journal of Chinese Mini-Micro Computer Systems, 2024, 45(9): 2165 - 2174. (in Chinese)
[21] 李岩超, 史卫亚, 冯灿. 面向无人机航拍小目标检测的轻量级 YOLOv8 检测算法[J]. 计算机工程与应用, 2024, 60(17): 167 - 178.
LI Yanchao, SHI Weiya, FENG Can. Lightweight YOLOv8 detection algorithm for small object detection in UAV aerial photography[J]. Computer Engineering and Applications, 2024, 60(17): 167 - 78. (in Chinese)
[22] 王舒梦, 徐慧英, 朱信忠, 等. 基于改进 YOLOv8n 航拍轻量化小目标检测算法:PECS - YOLO[J/OL]. 计算机工程, doi: 10.19678/j.issn.1000 - 3428.0069353.
WANG Shumeng, XU Huiying, ZHU Xinzhong, et al. Lightweight small object detection algorithm based on improved YOLOv8n aerial photography: PECS - YOLO[J/OL]. Computer Engineering, doi: 10.19678/j.issn.1000-3428.0069353. (in Chinese)
[23] XU S, WANG X, LV W, et al. PP - YOLOE: an evolved version of YOLO[J]. arXiv Preprint, arXiv: 2203.16250, 2022.