



基于高效特征提取和大感受野的无人机航拍图像目标检测

沈联宇, 朱凤华, 王知学, 沈震, 熊刚

引用本文:

沈联宇, 朱凤华, 王知学, 等. 基于高效特征提取和大感受野的无人机航拍图像目标检测[J]. *智能系统学报*, 2025, 20(4): 813-821.

SHEN Zhenyu, ZHU Fenghua, WANG Zhixue, et al. Uav aerial image target detection based on high-efficiency feature extraction and large receptive field[J]. *CAAI Transactions on Intelligent Systems*, 2025, 20(4): 813-821.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202405001>

您可能感兴趣的其他文章

舰载机位姿实时视觉测量算法研究

Research on real-time vision measurement algorithm of shipborne aircraft pose
智能系统学报. 2021, 16(6): 1045-1055 <https://dx.doi.org/10.11992/tis.202103014>

动态云台摄像机无人机检测与跟踪算法

Drone detection and tracking in dynamic pan-tilt-zoom cameras
智能系统学报. 2021, 16(5): 858-869 <https://dx.doi.org/10.11992/tis.202103032>

基于改进FCOS的拥挤行人检测算法

Crowded pedestrian detection algorithm based on improved FCOS
智能系统学报. 2021, 16(4): 811-818 <https://dx.doi.org/10.11992/tis.202010012>

融合视觉显著性再检测的孪生网络无人机目标跟踪算法

Siamese network combined with visual saliency re-detection for UAV object tracking
智能系统学报. 2021, 16(3): 584-594 <https://dx.doi.org/10.11992/tis.202101035>

采用相关滤波的水下海参目标跟踪

Underwater sea cucumber target tracking algorithm based on correlation filtering
智能系统学报. 2019, 14(3): 525-532 <https://dx.doi.org/10.11992/tis.201711037>

基于显著性检测的双目测距系统

Binocular distance measurement system based on saliency detection
智能系统学报. 2018, 13(6): 913-920 <https://dx.doi.org/10.11992/tis.201712005>

DOI: 10.11992/tis.202405001

网络出版地址: <https://link.cnki.net/urlid/23.1538.tp.20250226.0803.002>

基于高效特征提取和大感受野的 无人机航拍图像目标检测

沈朕宇¹, 朱凤华², 王知学¹, 沈震², 熊刚²

(1. 山东交通学院 轨道交通学院, 山东 济南 250300; 2. 中国科学院自动化研究所, 多模态人工智能系统全国重点实验室, 北京 100190)

摘要: 针对无人机航拍图像中存在小目标、目标遮挡、背景复杂的问题, 提出一种基于高效特征提取和大感受野的目标检测网络 (efficient feature and large receptive field network, EFLF-Net)。通过优化检测层架构降低小目标漏检率; 在主干网络融合新的构建模块以提升特征提取效率; 引入内容感知特征重组模块和大型选择性核网络, 增强颈部网络对遮挡目标的上下文感知能力; 采用 Wise-IoU 损失函数优化边界框回归稳定性。在 Vis-Drone2019 数据集上的实验结果表明, EFLF-Net 较基准模型在平均精度上提高了 5.2%。与已有代表性的目标检测算法相比, 该方法对存在小目标、目标相互遮挡和复杂背景的无人机航拍图像有更好的检测效果。

关键词: 无人机航拍图像; 小目标检测; 特征提取; 多尺度变化; YOLOv8; 上下文信息; 感受野; 损失函数

中图分类号: TP391.4 **文献标志码:** A **文章编号:** 1673-4785(2025)04-0813-09

中文引用格式: 沈朕宇, 朱凤华, 王知学, 等. 基于高效特征提取和大感受野的无人机航拍图像目标检测 [J]. 智能系统学报, 2025, 20(4): 813-821.

英文引用格式: SHEN Zhenyu, ZHU Fenghua, WANG Zhixue, et al. Uav aerial image target detection based on high-efficiency feature extraction and large receptive field[J]. CAAI transactions on intelligent systems, 2025, 20(4): 813-821.

Uav aerial image target detection based on high-efficiency feature extraction and large receptive field

SHEN Zhenyu¹, ZHU Fenghua², WANG Zhixue¹, SHEN Zhen², XIONG Gang²

(1. School of Rail Transit, Shandong Jiaotong University, Ji'nan 250300, China; 2. National Key Laboratory of MultiModal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China)

Abstract: Aiming at the problems of small targets, target occlusion and complex background in UAV aerial images, a target detection network based on high-efficiency feature extraction and large receptive field (EFLF-Net) was proposed. Firstly, the missed detection rate of small targets was reduced by optimizing the detection layer architecture. Then, the new building blocks were integrated in the backbone network to improve the efficiency of feature extraction. Then, a content-aware feature recombination module and a large selective kernel network were introduced to enhance the context-aware ability of the neck network for occluded targets. Finally, the Wise-IoU loss function was used to optimize the bounding box regression stability. Experimental results on the VisDrone2019 dataset show that EFLF-Net improves the average precision by 5.2% compared with the basic algorithm. Compared with the existing representative target detection algorithms, the proposed method has better detection effects for UAV aerial images with small targets, mutual occlusion of targets and complex backgrounds.

Keywords: drone aerial images; small target detection; feature extraction; multi-scale variation; YOLOv8; context information; receptive field; loss function

收稿日期: 2024-05-03. 网络出版日期: 2025-02-26.

基金项目: 国家自然科学基金项目 (U24A20277); 北京市自然科学基金项目 (L241016); 重庆市交通科技项目 (CQJT-CZKJ2024-04).

通信作者: 熊刚. E-mail: gang.xiong@ia.ac.cn.

随着无人机技术的迅速发展, 其航拍能力在农业、环境监测、城市规划及灾害应对等多个领域展现出广泛的应用潜力, 成为现代科技应用的重要组成部分^[1]。然而, 无人机航拍的图像往往

具有大视角,导致目标外观变化大,并且目标可能具有不同大小和严重遮挡,进一步增加了目标检测的复杂性^[2]。在无人机航拍图像目标检测领域,以卷积神经网络(convolutional neural network, CNN)为基础的算法已成为主流方法,其中,YOLO(you only look once)系列^[3]、SSD(single shot multibox detector)^[4]、RetinaNet(retina network)^[5]、Fast R-CNN(fast region-based convolutional neural network)^[6]和 Faster R-CNN(faster region-based convolutional neural network)^[7]等代表性算法得到了广泛应用。

为解决无人机航拍图像目标检测出现的问题,秦振等^[8]提出了一种改进的 SSD 方法,通过结合相邻特征和不同尺度间的非局部关系,提高了小目标检测的精度。Lee 等^[9]提出了一种增强的 RetinaNet 模型,该模型将可变形卷积纳入 ResNet-50 主干、更深的特征金字塔层和多头配置中的交叉联合感知分支,以提高检测性能。邓姗姗等^[10]提出了一种基于 Faster R-CNN 的多层频域特征融合目标检测算法,针对总像素面积较小的目标,修改了区域提案网络(region proposal network, RPN)中的锚点尺度,并采用多尺度卷积特征融合的方法,将来自不同特征层的特征进行融合,从而解决了小目标在深层特征图中信息丢失的问题。吴明杰等^[11]提出了一种改进的 YOLOv5s 小目标检测算法,采用双层路由注意力提升特征提取效果,并引入结合注意力机制的动态目标检测头,以增强特征表达能力。Wang 等^[12]在基线网络 YOLOX-X 的基础上,利用 SAHI(slicing aided hyper inference)算法对训练集进行了预处理和数据增强;在路径聚合网络(path aggregation network, PAN)阶段,引入具有丰富空间信息的浅层特征图,并添加目标检测头;最后,优化了边界框回归的损失函数,以提高图像检测的精度。牛为华等^[13]在改进 YOLOv7 时,将骨干网络中的低层小目标检测层融入聚合网络结构,并新增了一个检测头,提高对小目标的检测能力。但是这些方法仍存在不足,在目标尺寸小、背景信息复杂、目标遮挡的场景中会出现误检、漏检和检测精度不高的问题。

针对以上问题,本文在 YOLOv8n 架构的基础上,提出了一种基于高效特征提取和大感受野的目标检测网络(efficient feature and large receptive field network, EFLF-Net)算法,并在 VisDrone-2019 数据集上验证了该方法的有效性。

1 YOLOv8 算法

在各种目标检测算法中,YOLOv8^[14] 凭借其在速度、准确性、适用性和易用性等方面的优势脱颖而出。YOLOv8 算法是 YOLO 系列目标检测算法的改进版本,采用了一种更深、更宽的网络结构,以增加对目标的感知能力和特征提取能力。本文选用 YOLOv8n 作为基础模型,优化该模型以解决小目标、目标遮挡和背景复杂的问题。该模型的网络架构由主干网络(Backbone)、颈部网络(Neck)和检测头(Head)共 3 部分组成。Backbone 通常采用 DarkNet^[15]、ResNet^[16] 等卷积神经网络,用于提取输入图像的特征信息。Neck 采用特征金字塔网络(feature pyramid network, FPN)^[17]和 PAN^[18]的结构,以获得不同尺度的特征信息,从而提高对不同尺寸目标的检测能力。Head 则负责预测目标的类别、位置和置信度等信息^[19]。

2 EFLF-Net 网络

2.1 EFLF-Net 网络结构

为了解决无人机航拍图像目标检测中小目标检测精度较低的问题,本文提出了以 YOLOv8n 为基线的 EFLF-Net 目标检测算法,网络结构如图 1 所示。首先,通过优化检测层架构,提高网络对小目标检测精度;然后在主干网络融合一个新的 CNN 构建模块^[20],避免信息的丢失,使网络获得了更高的检测精度;此外,在颈部网络引入内容感知特征重组(content-aware reassembly of features, CARAFE)模块^[21]和大型选择性核网络(large selective kernel network, LSKNet)模块^[22],使模型以较大的感受野根据输入特征进行特征重组,提高网络关注检测目标的最相关的空间背景区域的能力;最后,引入 WIoU(Wise-IoU)损失函数^[23]替换原始的 CIoU(Complete-IoU)^[24],提高网络的泛化能力和检测性能。

2.2 优化检测层架构

为提升对无人机航拍图像中小目标的检测能力,在原有检测架构基础上引入了分辨率为 160 像素×160 像素的 P2 小目标检测层,以增强高分辨率浅层特征的表达能力。同时,移除原有 P5 (20 像素×20 像素)检测层的检测头,以控制模型复杂度并降低计算开销。检测层级由原始的 P3、P4、P5 调整为 P2、P3 和 P4,分别用于检测小、中、大尺度目标,详细改进如图 2 所示。其中,图 2(a)为原始检测层架构,图 2(b)为优化后的检测层架构。

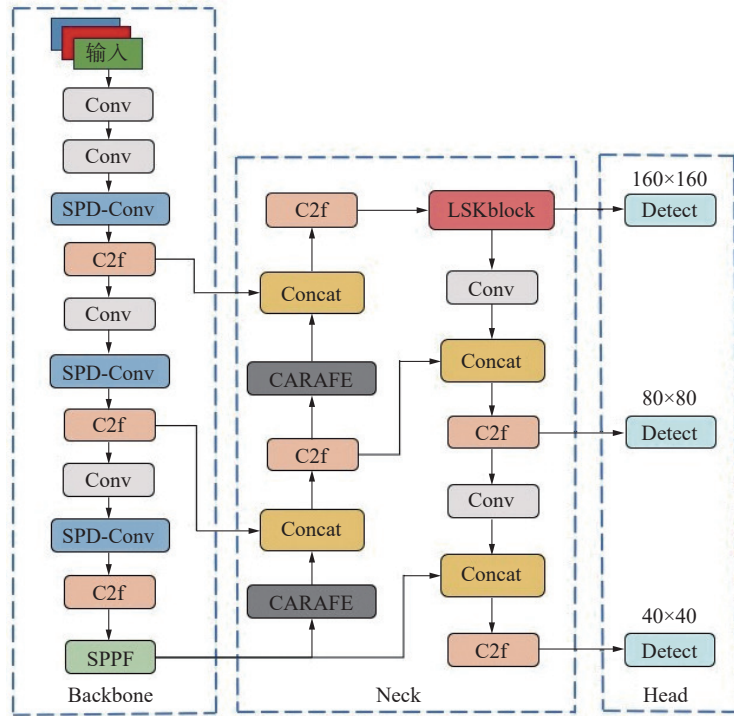


图 1 EFLF-Net 网络结构

Fig. 1 EFLF-Net network structure

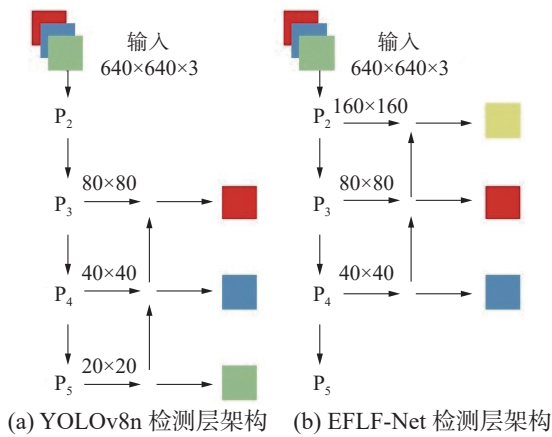


图 2 检测层架构对比

Fig. 2 Detect layer architecture comparison

2.3 大型选择性核网络

为了解决无人机航拍图像中存在背景复杂的问题,引入 LSKNet。该网络通过动态调整感受野的特征提取模块,更有效地处理了不同目标所需的背景信息差异。该动态感受野通过空间选择机制实现,对 Depth-wise^[25] 卷积核处理的特征进行加权 and 空间融合。卷积核的权重依据输入动态调整,使得模型能够自适应地选择合适的核大小,针对不同目标调整感受野。

LSKNet 块是主干网络中的一个可重复堆叠的块,其结构如图 3 所示。每个 LSKNet 块包括大核选择 (large kernel selection, LK Selection) 子块和前馈网络 (feed-forward network, FFN) 子块共

2 个残差子块。LK Selection 子块根据实际需求调整网络感受野的大小; FFN 子块则负责特征的通道混合和优化,包含 1 个全连接层、深度卷积操作、GELU 激活函数以及第 2 个全连接层。LSK 模块由大核卷积序列和空间核选择机制组成,嵌入到了 LSKNet 块的 LK Selection 子块中。

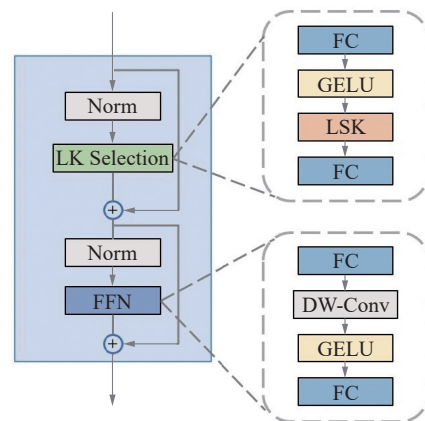


图 3 大型选择性核网络模块

Fig. 3 A block of LSKNet

2.4 新的 CNN 构建模块

卷积神经网络在图像分类和目标检测等计算机视觉任务中得到了广泛应用。然而,当面对小目标或低分辨率图像时,传统的网络结构常常无法提供理想的性能。为解决这一挑战,提出了一个称为 SPD-Conv(space-to-depth convolution) 的新的 CNN 构建模块,它替代了常规的跨步卷积和池

化层。该模块由空间-深度 (space-to-depth, SPD) 层与非跨步卷积层组成, 能够灵活处理多尺度特征图, 并按比例因子对其进行有效下采样。SPD 层将特征图的空间维度转换成深度维度, 通过增加通道数保留了更多信息。非跨步卷积层保持了空间维度, 减少了通道数。通过引入 SPD-Conv 模块不但避免了信息的丢失, 而且使网络获得了更高的检测精度、更低的误检率以及更好的目标定位能力, SPD-Conv 模块如图 4 所示。

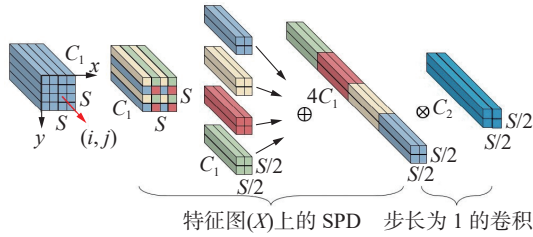


图 4 SPD-Conv 模块
Fig. 4 SPD-Conv module

假设中间特征图 X 的大小为 $S \times S \times C_1$, 则切出的子特征图为

$$\begin{cases} f_{0,0} = X[0 : S : t, 0 : S : t] \\ f_{1,0} = X[1 : S : t, 0 : S : t] \\ \vdots \\ f_{t-1,0} = X[t-1 : S : t, 0 : S : t] \\ \vdots \\ f_{t-1,t-1} = X[t-1 : S : t, t-1 : S : t] \end{cases}$$

图 4 为 $t=2$ 时的 SPD-Conv 下采样模块, 输入尺寸为 $S \times S \times C_1$ 的原始特征图 X 。原始特征图 X 首先经过空间-深度层, 按照设定系数 t 对 X 进行下采样, 经采样后原始特征图 X 切出了 4 个子特征图, 其中每个子特征图的空间维度变为 $S/2$, 通道维度仍为 C_1 。然后 4 个子特征图沿通道维度进行拼接, 得到大小为 $(S/2, S/2, 4C_1)$ 的特征图 X' 。在空间-深度层之后, 再经过一个带有 C_2 滤波器的非跨步 (即步长为 1) 卷积层, 其中 $C_2 < 4C_1$, 进一步变换得到大小为 $(S/2, S/2, C_2)$ 的特征图 X'' 。

2.5 内容感知特征重组模块

内容感知特征重组 (content-aware reassembly of features, CARAFE) 模块是自适应感受野特征增强模块, 通过像素点的空间位置来决定上采样核的方式, 能够有效地提高特征图的分辨率和语义信息, 并且对小目标的检测效果更好, 同时只带来轻量计算量。CARAFE 模块的核心原理是通过逐通道的自适应滤波来捕获更广泛的上下文信息, 并且通过插值操作将细节信息与原始特征图融合, 以生成高质量的上采样结果。

CARAFE 主要分为预测上采样核和特征重组

模块。当输入图像大小为 $C \times H \times W$ 且上采样率为 σ 时, 预测上采样核通过 1×1 卷积将通道数减少至 C_m , 然后进行内容编码和上采样核的预测, 使用 $k_{\text{encoder}} \times k_{\text{encoder}}$ 的卷积层操作, 其中输入通道数和输出通道数分别为 C_m 和 $\sigma^2 k_{\text{up}}^2$ 。接着通道在空间维度上展开并生成 $\sigma H \times \sigma H \times k_{\text{up}}^2$ 的上采样核, 随后进行归一化处理, 确保卷积核权重的总和为 1。特征重组模块在每个输出特征图位置进行映射, 提取以该位置为中心的 $k_{\text{up}} \times k_{\text{up}}$ 的区域, 并使用预测的上采样核进行点积运算, 以计算最终输出值。在不同位置但同一通道的特征图共享相同的上采样核。最终, 生成尺寸为 $C \times \sigma H \times \sigma W$ 输出特征图。

2.6 损失函数

在无人机航拍目标检测中损失函数能够指导模型正确的定位和分类目标, 抑制背景干扰, 以及平衡不同任务之间的优化目标, 从而提高模型的检测性能和泛化能力。本文引用损失函数 WIoU v3, 用于无人机目标检测任务中取代传统的 CIoU 评估方法。WIoU v3 考虑了目标的形状和大小差异性, 通过引入权重参数, 可以根据具体的检测场景进行调整, 提高了评估的灵活性和准确性。相比传统方法, WIoU v3 对边界框的位置和大小进行了更精细的计算, 使得在目标存在遮挡或者重叠部分时, 匹配结果更加准确, 有助于提高检测算法的鲁棒性和可靠性。通过 WIoU v3 评估, 可以更直观地了解检测算法在不同场景下的表现, 指导算法的优化和改进, 进一步提升无人机目标检测的性能和效果。

通过距离度量方法构建了距离注意力机制, 从而设计了具有双重注意力机制的 WIoU v1。其具体表达式为

$$\mathcal{R}_{\text{WIoU}} = \exp\left(\frac{(x-x_{gt})^2 + (y-y_{gt})^2}{(W_g^2 + H_g^2)^*}\right)$$

$$\mathcal{L}_{\text{WIoUv1}} = \mathcal{R}_{\text{WIoU}}(1 - I_{\text{IoU}})$$

式中: x 和 y 为预测框中心点的坐标, x_{gt} 和 y_{gt} 为真实框中心点的坐标, W_g 和 H_g 为预测框和真实框之间的最小包围矩形的宽度和高度, I_{IoU} 为预测框和真实框的交并比。

通过利用 β 构造了一个非单调聚焦系数并将其应用于 WIoU v1, 从而实现了具有动态非单调特征映射的 WIoU v3。采用了明智的梯度增益分配策略, 使得 WIoU v3 在性能上获得了显著提升。WIoU v3 具体表达式为

$$\beta = \frac{\mathcal{L}_{\text{IoU}}^*}{\mathcal{L}_{\text{IoU}}} \in [0, +\infty)$$

$$r = \frac{\beta}{\delta\alpha^{\beta-\delta}}$$

$$\mathcal{L}_{\text{WIoUv3}} = r\mathcal{L}_{\text{WIoUv1}}$$

式中: $\mathcal{L}_{\text{WIoUv3}}$ 为定位损失函数值, 采用了动态非单调评估锚盒质量的机制, 使模型更专注于锚定普通质量的框, 并提升了模型对目标的定位能力; α 和 δ 为超参数; β 为离群度; $\mathcal{L}_{\text{IoU}}^*$ 为单调聚焦系数; $\bar{\mathcal{L}}_{\text{IoU}}$ 为动量的滑动平均值。在无人机航拍场景中的目标检测任务中, 小物体的比例较大, 增加了检测的难度。WIoU v3 能够动态地优化对小物体的加权, 以提升模型的检测性能。

3 实验验证与对比分析

3.1 数据集

为评估所提出算法的性能, 实验选择了 VisDrone2019^[26] 数据集进行验证。该数据集由天津大学 AISKEYEYE 团队收集, 使用无人机摄像设备拍摄, 包含 10 种目标类别。VisDrone2019 数据集共包含 10209 张静态图片, 其中训练集 6471 张图片, 验证集 548 张图片, 测试集 3190 张图片。

3.2 实验环境及参数配置

本文使用的实验环境配置如下: 操作系统为 64 位 Windows11 系统; CPU 为 13th Gen Intel(R) Core(TM) i5-13400F, 工作频率为 2.50 GHz; GPU 为 NVIDIA GeForce RTX 3090, 显存大小为 24 GB; 实验环境为 Python3.8.18, PyTorch2.0.0, CUDA11.8。相关参数配置: 将输入图片的分辨率统一设置为 640×640, 缩放至统一尺寸, 以提升处理速度; 设置训练批次 (batch-size) 为 8, 确保 GPU 能够全功率运行; 训练迭代次数 (epoch) 为 150, 通常在此时网络能够收敛; 随机种子设为 1, 固定模型训练产生的随机数, 最大程度复现模型效果; 使用随机梯度下降 (stochastic gradient descent, SGD) 优化器, 并采用梯度下降法调整学习率; 初始学习率为 0.01, 逐步减小学习率至最小值 0.0001, 初始设定较大的学习率有助于加快训练速度, 而设定最小学习率则有助于避免网络陷入局部最小值。

3.3 参数指标

本研究在 VisDrone2019 数据集上评估算法性能时, 选用了精确率 P (precision)、召回率 R (recall)、平均精度 (average precision, AP)、平均精度均值 (mean average precision, mAP)、计算量 (floating point of operations, FLOPs) 和参数量 (parameters, Params) 等指标进行比较分析, 旨在量化模型的表现, 并为后续训练与调优提供依据。mAP50 表示交并阈值为 0.5 时的平均精度均值, 其计算公式为

$$P = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}}}$$

$$R = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}}$$

$$A_{\text{AP}} = \int_0^1 P(R) dR$$

$$M_{\text{mAP}} = \frac{1}{N} \sum_{i=1}^N A_{\text{AP}}^i$$

式中: N_{TP} 为模型正确预测为正类的正样本数量, N_{FP} 为模型错误地将负样本预测为正类的数量, N_{FN} 为模型错误地将正样本预测为负类的数量, A_{AP}^i 为每个类 i 的平均精度, N 为数据集中的类别数, M_{mAP} 为所有类 i 的平均精度的均值。

3.4 实验结果与分析

3.4.1 LSKNet 模块对网络性能的影响

为了评估 LSKNet 模块在不同位置引入对网络性能的影响, 以 YOLOv8n 为基准模型, 设计了如表 1 所示的对比实验, 旨在全面分析模块位置对模型表现的具体作用。实验中, 4 种不同的配置分别进行了测试: LSK-B 表示在骨干网络的第 4 个 C2f 模块后添加 LSKNet 模块, LSK-S 表示在小目标层后添加 LSKNet 模块, LSK-L 表示在大目标层后添加 LSKNet 模块, LSK-N 表示在颈部网络的 3 个输出端后都添加 LSKNet 模块。通过该实验深入分析不同模块位置的引入如何影响网络的特征提取、目标检测以及整体性能。

表 1 LSKNet 不同位置对比结果
Table 1 LSKNet different position comparison results

模型	mAP50/%	P/%	R/%	Params/10 ⁶	FLOPs/10 ⁹
YOLOv8n	31.3	42.6	31.9	3.00	8.1
LSK-B	31.9	42.3	32.1	3.16	8.3
LSK-S	32.7	45.1	32.5	3.02	8.2
LSK-L	32.0	43.1	32.3	3.13	8.2
LSK-N	32.0	42.7	32.6	3.17	8.4

注: 加黑代表最优结果, 下同。

由实验结果可知, 在 YOLOv8n 模型的小目标层后添加 LSKNet 模块, 计算成本增加最小的同时检测精度得到了最大的提升。因此, 选择在小目标层后添加 LSKNet 模块可以获得最佳性能。

3.4.2 主流算法对比实验

为进一步评估 EFLF-Net 算法对无人机航拍图像中小目标、目标遮挡和背景复杂目标的检测优势, 本研究在 VisDrone2019 数据集上开展了与其他主流目标检测算法的对比实验, 主要算法有 Faster-RCNN、CenterNet、RetinaNet、CornerNet、Cascade-RCNN、YOLOv5s、YOLOX-s 和 YOLOv8n, 各类目标的平均精度如表 2 所示。

表 2 mAP50 对比实验结果
Table 2 Results of mAP50 compare experiment

%

模型	目标类别										平均
	行人	人员	自行车	汽车	面包车	卡车	三轮车	带棚三轮	巴士	摩托车	
Faster-RCNN	21.4	15.6	6.7	51.7	29.5	19.0	13.1	7.7	31.4	20.7	21.7
CenterNet ^[27]	28.0	11.6	9.0	51.0	36.5	27.9	20.1	19.9	37.7	21.0	26.0
RetinaNet	13.0	7.9	1.4	45.5	19.9	11.5	6.3	4.2	17.8	11.8	13.9
CornerNet ^[28]	20.4	6.6	4.6	40.9	20.2	20.5	14.0	9.3	24.4	12.1	17.4
Cascade-RCNN	19.9	12.3	8.4	54.1	35.3	26.4	17.4	9.2	42.2	19.6	24.5
YOLOv5s ^[29]	39.1	31.8	10.2	73.3	35.4	31.7	19.5	11.7	38.9	37.0	32.9
YOLOX-s ^[30]	15.0	10.9	4.2	49.3	24.4	15.7	10.5	6.1	25.6	16.9	17.9
YOLOv8n	33.6	26.4	7.2	75.0	37.1	26.9	19.7	11.7	42.0	33.8	31.3
EFLF-Net	43.9	34.7	11.5	80.4	41.7	28.0	22.3	13.2	46.6	42.9	36.5

实验结果中 EFLF-Net 算法在行人、人员、自行车、汽车、面包车、卡车、三轮车、带棚三轮、巴士和摩托车上的 mAP50 值分别为 43.9%、34.7%、11.5%、80.4%、41.7%、28.0%、22.3%、13.2%、46.6% 和 42.9%，取得了最优效果。与基准模型 YOLOv8n 的 mAP50 值进行对比，EFLF-Net 算法在行人、人员、自行车、三轮车和带棚三轮小目标类别上的提升分别为 10.3%、8.3%、4.3%、2.6% 和 1.5%；在汽车、面包车和摩托车中目标类别上提高了 5.4%、4.6% 和 9.1%，在卡车、巴士大目标类别上提高了 1.1% 和 4.6%。同时比次优算法 YOLOv5s 的 mAP50 值高 3.6%，但是在大目标 Truck 类别上精度较低。由此可见，EFLF-Net 算法在进行无人机目标检测的平均精度较好，效果更好。这一结

果主要由于 EFLF-Net 合理优化了目标检测层，加强了小目标特征信息的提取和保留，并通过引入 LSKNet 模块进一步加强了应对复杂背景的检测能力。同时，引入 SPD-Conv 和 CARAFE，提升了检测精确度的同时大大降低了漏检率，使得本文算法在小目标检测方面具有巨大优势，同时也显著提升了中、大目标检测效果，表现出良好的鲁棒性。

3.4.3 消融实验结果及分析

为评估 EFLF-Net 算法各模块的效果，本研究以 YOLOv8n 作为基准模型，在 VisDrone2019 数据集上进行了消融实验，旨在分析不同模块对无人机航拍图像检测性能的影响。实验首先单独测试每个改进模块，然后逐步集成其他模块进行实验，最终的结果如表 3 所示。

表 3 消融实验结果
Table 3 Results of ablation experiment

YOLOv8n	优化层	SPD-Conv	CARAFE	LSKNet	WIoU v3	mAP50/%	P/%	R/%	Params/M	FLOPs/G
√						31.3	42.6	31.9	3.01	8.1
√	√					34.1	44.6	33.5	0.98	9.4
√		√				34.2	44.6	34.5	3.27	9.6
√			√			32.3	43.5	32.0	3.14	8.6
√				√		32.7	45.1	32.5	3.02	8.2
√					√	32.2	43.7	32.2	3.01	8.1
√	√	√				36.2	45.9	35.7	1.10	10.4
√	√	√	√			36.3	46.5	35.8	1.13	10.2
√	√	√	√	√		36.4	46.6	36.0	1.08	10.2
√	√	√	√	√	√	36.5	46.7	36.3	1.08	10.2

由表 3 消融实验结果可知，添加不同的改进策略对模型的检测性能影响不同。单独进行目标层优化，mAP50 增加了 2.8%，P、R 分别增加了 2%、1.6%，Params 降低了 2.03×10⁶，对整体模型轻量化起重要作用；加入 SPD-Conv，mAP50 增加了

2.9%，P、R 分别增加了 2%、2.6%；引入 CARAFE 模块后，mAP50 增加了 1%，P、R 分别增加了 0.9% 和 0.1%，添加 SPD-Conv 后可以有效降低目标的漏检率；引入 LSKNet 模块后，mAP50 增加了 1.4%，P、R 分别增加了 2.5%、0.6%；修改 CIoU 损

失函数为 WIoU 损失函数, mAP50 增加了 0.9%, P 、 R 分别增加了 1.1% 和 0.3%, 有效降低冗余计算。

在初始模型中, 依次引入目标检测层优化、SPD-Conv、CARAFE、LSKNet 和 WIoU 模块, 实验结果显示, 这些改进逐步提升了模型的性能, 得到的 mAP50 依次比基础模型增加了 2.8%、4.9%、5%、5.1% 和 5.2%, P 依次比基础模型增加了 2%、3.3%、3.9%、4% 和 4.1%, R 依次比基础模型增加了 1.6%、3.8%、3.9%、4.1% 和 4.4%。

通过消融实验结果显示, 优化的每一个模块

都展现出了良好的性能。最终优化后的模型相较于基准算法, P 值提升了 4.1%, R 值提升了 4.4%, mAP50 提升了 5.2%, 同时减少了参数量。虽然计算量增加了 2.1×10^9 , 但检测准确率有明显的提升, 整体网络性能显著提升, 证实了本文所提出算法的可行性。

3.4.4 实验可视化与分析

为更加直观地展示 EFLF-Net 算法在实际场景中的检测效果, 本研究在 VisDrone2019 数据集的典型场景中进行了可视化对比, 相关结果如图 5 所示。

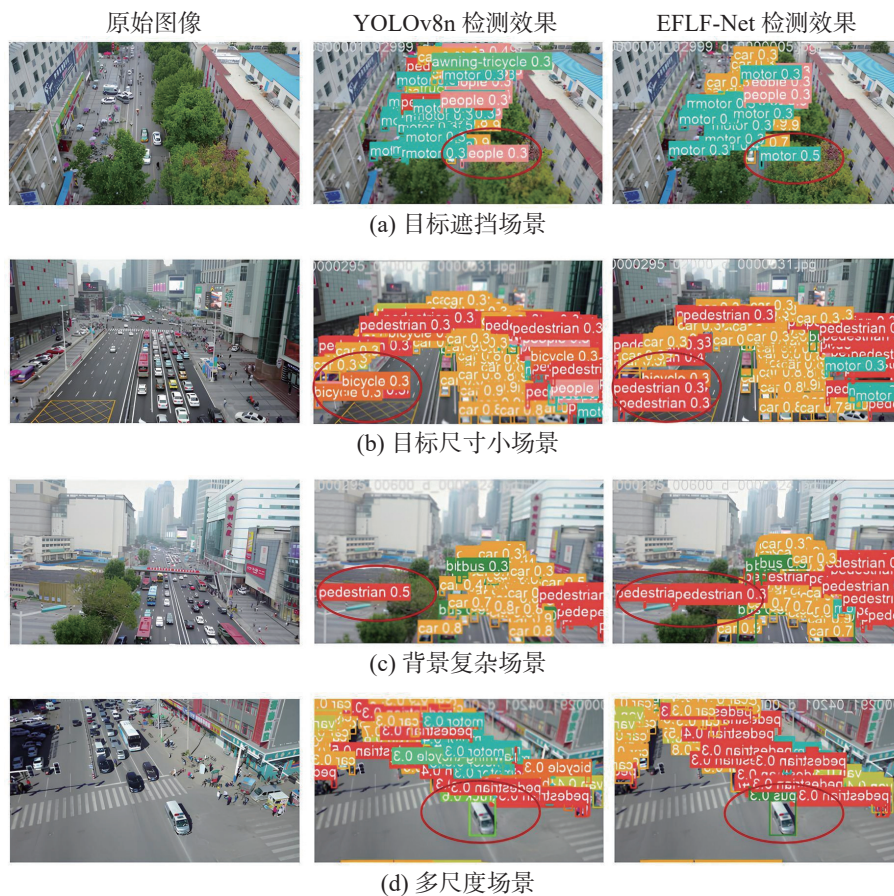


图 5 检测效果对比

Fig. 5 Comparison of detection effect

图 5(a) 出现了检测目标被遮挡的情况, 在图片中下方有被树荫遮挡的摩托车, YOLOv8n 算法将摩托车误检为人, 而本文改进算法准确检测出了摩托车。图 5(b) 出现在目标尺寸过于小的情况, 在图片左下方有人在行走, 但是因为行人尺寸过小, 使得 YOLOv8n 算法将行人误检为自行车, 而本文改进算法正确识别了行人。图 5(c) 出现了背景复杂的情况, 在图片街道、广场和建筑的复杂背景中, 左侧有多个行人, YOLOv8n 算法只识别出了一个行人目标, 而本文改进算法未出现漏检。图 5(d) 出现了巴士因视角不同而尺寸不

同的情况, YOLOv8n 算法误将巴士检测为卡车, 而本文改进算法正确识别。从比较结果可以看出, 相较于原始算法, 改进后的方法在无人机航拍图像的目标检测中展现了更低的误检率和漏检率, 且能更好地适应复杂背景、目标遮挡以及多尺度场景, 显著提升了检测性能。

3.4.5 模型泛化性实验对比

为了验证本文提出算法的适用性和鲁棒性, 首先将 VisDrone2019 数据集中的图像添加椒盐噪声; 然后, 将得到的噪声数据集与基础算法进行对比实验, 这样做可以更真实地模拟现实世界中

的数据情况,从而提高模型在实际应用中的泛化能力,并帮助模型更好地学习和适应真实世界中的变化和噪声,增强其鲁棒性和泛化能力。椒盐噪声处理还可视为一种数据增广技术,通过引入噪声使得训练数据更加多样化,有助于防止模型在训练过程中过拟合,并提升其对未知数据的适应能力。改进前后模型在噪声数据集上的实验结果如表 4 所示。

表 4 改进前后模型在噪声数据集上的实验结果

Table 4 Experimental results of the models before and after improvement on the noisy dataset

模型	mAP50/%	P/%	R/%	Params/10 ⁶
YOLOv8n	30.0	41.0	30.4	3.01
EFLF-Net	34.3	45.3	33.8	1.08

根据表 4 数据显示,改进后的算法在检测精度、查准率和召回率上分别提升了 4.3%、4.3% 和 3.4%,同时减少了参数量。这些结果充分证明了本文提出的改进算法具有良好的泛化性和鲁棒性,进一步验证了其在航拍图像目标检测中的适用性。

4 结束语

针对无人机航拍图像存在的小目标、背景干扰和目标重叠的问题,在 YOLOv8n 的基础上提出了基于高效特征提取和大感受野的目标检测(EFLF-Net)算法。该算法通过优化检测层架构提高了对小目标的检测精度。在主干网络中引入 SPD-Conv 模块,避免了信息丢失,提高了检测精度。同时,在颈部网络中引入 CARAFE 和 LSKNet 模块,使模型能够更好地关注检测目标及其周围背景区域。此外,引入 WIoU 损失函数,提高了网络的泛化能力和检测性能。在 VisDrone2019 数据集上的实验表明,EFLF-Net 相比 YOLOv8n 初始算法的平均精度提高了 5.2%,效果较好,并且算法精度超越了其他主流算法,证明了其准确性,具有广泛的应用前景。虽然改进后的算法在提升检测精度的同时也有效减小了参数量,但在参数优化方面仍有进一步提升的空间。与初始算法相比,改进算法虽然提高了精度,但计算量有所增加,导致计算时间和成本的上升。未来的研究将聚焦于在提高精度的基础上,进一步减少模型的计算量和参数量,从而实现更轻量化的设计,便于在无人机航拍实时检测任务中应用。

参考文献:

[1] 何宇豪,易明发,周先存,等.基于改进的 Yolov5 的无

人机图像小目标检测[J].智能系统学报,2024,19(3):635–645.

HE Yuhao, YI Mingfa, ZHOU Xiancun, et al. UAV image small-target detection based on improved Yolov5[J]. *CAAI transactions on intelligent systems*, 2024, 19(3): 635–645.

[2] 刘威,靳宝,周璇,等.基于特征融合及自适应模型更新的相关滤波目标跟踪算法[J].智能系统学报,2020,15(4):714–721.

LIU Wei, JIN Bao, ZHOU Xuan, et al. Correlation filter target tracking algorithm based on feature fusion and adaptive model updating[J]. *CAAI transactions on intelligent systems*, 2020, 15(4): 714–721.

[3] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 779–788.

[4] TAN Ling, WU Hui, XU Zifeng, et al. Multi-object garbage image detection algorithm based on SP-SSD[J]. *Expert systems with applications*, 2025, 263: 125773.

[5] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2020, 42(2): 318–327.

[6] CUI Jian, ZHANG Xinle, ZHANG Jiahuan, et al. Weed identification in soybean seedling stage based on UAV images and Faster R-CNN[J]. *Computers and electronics in agriculture*, 2024, 227: 109533.

[7] KARAKO K, MIHARA Y, ARITA J, et al. Automated liver tumor detection in abdominal ultrasonography with a modified faster region-based convolutional neural networks (Faster R-CNN) architecture[J]. *Hepatobiliary surgery and nutrition*, 2022, 11(5): 675–683.

[8] 秦振,李学伟,刘宏哲.基于改进 SSD 的鲁棒小目标检测算法[J].东北师大学报(自然科学版),2023,55(4):59–66.

QIN Zhen, LI Xuewei, LIU Hongzhe. Robust small target detection algorithm based on improved SSD[J]. *Journal of Northeast Normal University(natural science edition)*, 2023, 55(4): 59–66.

[9] LEE S S, LIM L G, PALAIAHNAKOTE S, et al. Oil palm tree detection in UAV imagery using an enhanced RetinaNet[J]. *Computers and electronics in agriculture*, 2024, 227: 109530.

[10] 邓姗姗,黄慧,马燕.基于改进 Faster R-CNN 的小目标检测算法[J].计算机工程与科学,2023,45(5):869–877.

DENG Shanshan, HUANG Hui, MA Yan. A small object detection algorithm based on improved Faster R-CNN[J]. *Computer engineering & science*, 2023, 45(5): 869–877.

[11] 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,2019,60(2):191–199.

WU Mingjie, YUN Lijun, CHEN Zaiqing, et al. Im-

- proved YOLOv5s small target detection algorithm in UAV view[J]. *Computer engineering and applications*, 2019, 60(2): 191–199.
- [12] WANG Xin, HE Ning, HONG Chen, et al. Improved YOLOX-X based UAV aerial photography object detection algorithm[J]. *Image and vision computing*, 2023, 135: 104697.
- [13] 牛为华, 魏雅丽. 基于改进 YOLOv7 的航拍小目标检测算法[J]. *电光与控制*, 2024, 31(1): 117–122.
NIU Weihua, WEI Yali. Small target detection in aerial photography images based on improved YOLOv7 algorithm[J]. *Electronics optics & control*, 2024, 31(1): 117–122.
- [14] TERVEN J, CORDOVA-ESPARZA D M, ROMERO-GONZÁLEZ J A. A comprehensive review of YOLO architectures in computer vision: from YOLOv1 to YOLOv8 and YOLO-NAS[J]. *Machine learning and knowledge extraction*, 2023, 5(4): 1680–1716.
- [15] KRIEGEL J, DEJAM J, DURTH H, et al. Zur strafbarkeit von datenfunden im darknet[J]. *Datenschutz und datensicherheit-DuD*, 2024, 48(12): 769–774.
- [16] SHEN Kenan, ZHAO Dongbiao. Fault analysis and fault degree evaluation *via* an improved ResNet method for aircraft hydraulic system[J]. *Scientific reports*, 2025, 15: 4132.
- [17] FENG Dapeng, ZHUANG Xuebin, CHEN Zhiqiang, et al. Position information encoding FPN for small object detection in aerial images[J]. *Neural computing and applications*, 2024, 36(26): 16023–16035.
- [18] LIU Shu, QI Lu, QIN Haifang, et al. Path aggregation network for instance segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 8759–8768.
- [19] 伍麟, 郝鸿宇, 宋友. 基于计算机视觉的工业金属表面缺陷检测综述[J]. *自动化学报*, 2024, 50(7): 1261–1283.
WU Lin, HAO Hongyu, SONG You. A review of industrial metal surface defect detection based on computer vision [J]. *IEEE/CAA journal of automatica sinica*, 2019, 50(7): 1261–1283.
- [20] SUNKARA R, LUO Tie. No more strided convolutions or pooling: a new CNN building block for low-resolution images and small objects[EB/OL]. (2022–08–07)[2024–04–01]. <https://arxiv.org/abs/2208.03641>.
- [21] DING Zhipeng, WANG Ben, SUN Shuifa, et al. Improved SmapGAN remote sensing image map generation based on multi-head self-attention and carafe[J]. *Journal of applied remote sensing*, 2024, 18(1): 014526.
- [22] HOU Yue, ZHANG Zhihao, DU Lixia, et al. A fully locally selective large kernel network for traffic video detection[J]. *Measurement*, 2025, 242: 115779.
- [23] WANG Chenghao, LUO Zhongqiang, QI Ziyuan. Transformer oil leakage detection with sampling-WIoU module[J]. *The journal of supercomputing*, 2024, 80(6): 7349–7368.
- [24] HUANG Zixin, TAO Xuesong, LIU Xinyuan. NAN-DETR: noising multi-anchor makes DETR better for object detection[J]. *Frontiers in neurorobotics*, 2024, 18: 1484088.
- [25] MARAPATLA A D K, ILAVARASAN E. An effective attack detection framework using multi-scale depth-wise separable 1DCNN via fused grasshopper-based lemur optimizer in IoT routing system[J]. *Intelligent decision technologies*, 18(3): 1741–1762.
- [26] WANG Xin, HE Ning, HONG Chen, et al. YOLO-ERF: lightweight object detector for UAV aerial images[J]. *Multimedia systems*, 2023, 29(6): 3329–3339.
- [27] ZHU Xingfei, WANG Qimeng, ZHANG Bufan, et al. An improved feature enhancement CenterNet model for small object defect detection on metal surfaces[J]. *Advanced theory and simulations*, 2024, 7(8): 2301230.
- [28] NAWAZ M, NAZIR T, MASOOD M, et al. Analysis of brain MRI images using improved CornerNet approach[J]. *Diagnostics*, 2021, 11(10): 1856.
- [29] WANG Zhaodi, YANG Shuqiang, QIN Huafeng, et al. CCW-YOLO: a modified YOLOv5s network for pedestrian detection in complex traffic scenes[J]. *Information*, 2024, 15(12): 762.
- [30] ZHANG Hongtao, ZHENG Li, TAN Lian, et al. YOLOX-S-TKECB: a Holstein cow identification detection algorithm[J]. *Agriculture*, 2024, 14(11): 1982.

作者简介:



沈朕宇, 硕士研究生, 主要研究方向为图像处理与目标检测。E-mail: 2216825930@qq.com。



朱凤华, 副研究员, 博士, 主要研究方向为智能交通、云计算与大数据分析。E-mail: fenghua.zhu@ia.ac.cn。



熊刚, 研究员、博士生导师, 主要研究方向为人工智能、智能控制与管理。获吴文俊人工智能奖、中国自动化学会科技奖等 10 余项。发表学术论文 450 余篇, 出版专著共 3 部, 授权 PCT 6 项, 授权专利 90 余项, 登记软著 90 余项。E-mail: gang.xiong@ia.ac.cn。