2025 年 第 25 卷 第 19 期
2025, 25(19):08151-07
科 学 技 术 与 工 程
Science Technology and Engineering
ISSN 1671—1815
CN 11—4688/T

DOI:10.12404/j.issn.1671-1815.2405983
引用格式:张强, 张灿智, 曹恒, 等. 融合 ViT 与多尺度注意力的改进 YOLOv8 飞鸟识别算法[J]. 科学技术与工程, 2025, 25(19): 8151-8157.
Zhang Qiang, Zhang Canzhi, Cao Heng, et al. Improved YOLOv8 bird recognition algorithm integrates ViT and multi-scale attention[J]. Science Technology and Engineering, 2025, 25(19): 8151-8157.

融合 ViT 与多尺度注意力的改进 YOLOv8 飞鸟识别算法

张强, 张灿智, 曹恒, 员腾蛟
(中国民用航空飞行学院空中交通管理学院, 德阳 618300)

摘 要 针对飞鸟识别中存在密集目标识别不准确、小目标检测困难等问题,提出一种基于改进 YOLOv8 的飞鸟识别算法。首先,针对密集目标识别难度大的问题,采用多尺度线性注意力机制 EfficientViT 替换骨干网络,实现全局感受野和多尺度学习,提升模型性能和效率的同时提高密集目标识别效果。然后,针对小目标飞鸟检测困难、容易出现漏检的问题,引入高效多尺度注意力(efficient multi-scale attention, EMA)机制,通过通道重组实现跨维度聚合特征,从而更好地捕捉全局信息,实现多尺度特征融合,减少漏检概率。实验结果表明,改进模型在鸟类识别基准数据集 CUB-200-2011 和自制数据集 birds28 上的 mAP50 分别达到 77.1% 和 88.4%,较原始 YOLOv8 模型分别提高了 4.5 和 5.4 个百分点,验证了改进模型的有效性。

关键词 飞鸟识别; 多尺度注意力; 密集目标识别; YOLOv8; EfficientViT; EMA
中图法分类号 TP391; **文献标志码** A

Improved YOLOv8 Bird Recognition Algorithm Integrates ViT and Multi-scale Attention

ZHANG Qiang, ZHANG Can-zhi, CAO Heng, YUAN Teng-jiao
(College of Air Traffic Management, Civil Aviation Flight University of China, Deyang 618300, China)

[**Abstract**] In order to solve the problems of inaccurate dense target recognition and difficult detection of small targets in bird recognition, a bird recognition algorithm based on improved YOLOv8 was proposed. Firstly, in order to solve the problem of difficult dense object recognition, the multi-scale linear attention mechanism EfficientViT was used to replace the backbone network to realize the global receptive field and multi-scale learning, improve the performance and efficiency of the model, and improve the dense object recognition effect. Then, in order to solve the problem that it is difficult to detect small target birds and is prone to missed detection, an efficient multi-scale attention EMA (efficient multi-scale attention) mechanism was introduced to realize cross-dimensional aggregation features through channel recombination, so as to better capture global information, realize multi-scale feature fusion, and reduce the probability of missed detection. The experimental results show that the mAP50 of the improved model on the benchmark dataset CUB-200-2011 and birds28 reaches 77.1% and 88.4%, respectively, which is 4.5 and 5.4 percentage points higher than the original YOLOv8 model, respectively, which verifies the effectiveness of the improved model.
[**Keywords**] bird recognition; multi-scale attention; dense target recognition; YOLOv8; EfficientViT; EMA

飞鸟撞击航空器是影响航空安全的世界性难题[1],中国民航局航空安全办公室和民航科学技术研究院联合发布的《中国民航安全信息统计报告》显示,鸟击是民航第一大事故征候来源,占比达 43.26%。近年来随着生态环境的改善以及航班增多,鸟击风险正在逐渐提高。大多数鸟击事件都发生在机场和机场附近空域,为有效防止鸟击事件发生,需要对机场及其周边的鸟类情况有充分的了解[2]。但在鸟情调查工作中,人工识鸟效率低,且难度大。因此可以利用机场光电设施获取鸟类图片,再通过深度学习算法对鸟类进行自动识别[3],帮助机场更好地开展鸟防工作。

利用计算机视觉实现鸟类识别的难点[4-6]在于鸟类特征的特殊性,不同子类之间往往具有相似的外观,而同一子类也可能因为不同的姿态和背景遮挡而造成较大差距。随着深度学习和卷积神经网

收稿日期: 2024-08-09 **修订日期**: 2024-12-23
基金项目: 中央高校基本科研业务费专项(ZJ2023-007);教育部产学合作协同育人项目(202101199029)
第一作者: 张强(1986—),男,汉族,四川广安人,博士,教授。研究方向:民航通信导航监视、低空监视。E-mail:271198043@qq.com。

络(convolutional neural networks,CNN)的发展[7-8],目标分类识别技术也取得了重大突破[9]。深度学习以其卓越的计算能力和深度特征提取技术,在飞鸟识别领域展现出了非凡的潜力。通过自动学习并识别鸟类之间复杂且微妙的特征差异,深度学习模型能够显著提升鸟类识别的精度,有效减少重复性工作[10-11]。针对飞鸟目标易被遮挡、类间区别度小等难点,研究人员已经采取多种方法来改进目标检测和鸟类图像识别任务。姿态归一化 CNN,依赖边界框,先完成局部定位检测并对图片裁剪,将不同的部位姿态对齐,再获得不同部位的卷积层特征,并对这些特征进行处理后实现分类。采用姿态对齐操作和多尺度特征融合方式,减轻了姿态问题的干扰,增加了识别准确率。但是依赖标注信息,对于复杂姿态的目标,姿态归一化可能会增加额外的计算和模型复杂度。Bellafkir 等[12]利用深度学习模型解决由于环境条件变化所导致的识别效果差的问题,利用可靠性分数和集成预测的组合实现了较理想的识别效果,但是该模型采用主动学习的方式,依赖人类专家反馈,需要足够的专家反馈示例才能达到较好的识别效果。YOLO(yolo only look once)算法广泛应用于各种计算机视觉任务[13],而且至今仍在快速迭代更新中[14]。陈天华等[15]改进了 YOLOv5 算法去识别鸟类,引入双层路由注意力提升了模型效果,但其仅在自制数据集中实验,无法验证改进模型的泛化能力。YOLOv8 集成了此前一系列 YOLO 算法的优点,在网络中大量使用了残差连接[16],性能更优秀,但在飞鸟识别任务中也存在一些不足,因此对 YOLOv8 进行改进,以更好地适应飞鸟识别任务。主要改进方式为采用多尺度线性注意力机制(EfficientViT)替换主干网络,使用更轻量和高效的硬件操作实现多尺度学习和全局感受野,以提升对目标密集图像的检测识别效果,同时降低计算成本,实现性能和速度的提升。以及引入高效多尺度注意力(efficient multi-scale attention, EMA)模块,通过跨空间重塑信道的方式提高参数的利用率,同时对于捕获图像边缘目标以及小尺度鸟类具有更好的效果。

1 YOLOv8 算法

YOLO 系列算法自提出就受到业界广泛关注,YOLOv8 集成了前几代算法的优势,具有更快的检测速度和识别精度,在各种图像处理任务都表现出色[17-18]YOLOv8 又被细分为五个版本以适应各种图像处理任务的要求,分别为 YOLOv8n、YOLOv8s、YOLOv8m、YOLOv8l、YOLOv8x。YOLOv8 的具体结构由主干网络(Backbone)、颈部网络(Neck),以及头部(Head)组成。主干网络提取特征信息,颈部网络负责特征融合,头部则输出结果。虽然 YOLOv8 的性能已经十分出色,但对于飞鸟识别任务来说还略有不足,存在识别精度不高、小目标检测困难、密集目标预测不准确等问题。因此,针对飞鸟的一些特性,提出了改进的 YOLOv8 模型。

2 飞鸟特性分析

飞鸟识别较一般图像识别任务更为困难,因为飞鸟具有高度相似的从属类别,也就是类间差异小,而不同的姿势、拍摄角度等因素又可能会造成较大的类内差异。如图 1 所示,金腰燕因飞行姿态和拍摄角度问题,呈现出较大差别。而金腰燕、灰椋鸟、白头鹎之间的差异很小。因此,引入 EfficientViT 对飞鸟特征进行多尺度学习,提高模型的飞鸟识别能力。

图 1 类间差异与类内差异

Fig. 1 Inter-class and intra-class differences

此外,由于飞鸟具有群聚特性,飞鸟数据集中也常出现目标密集的情况。YOLOv8 的主干网络采用的 Darknet-53 结构框架利用残差连接和 Split 操作实现了多尺度特征信息的整合,但无法有效处理密集目标。而 EfficientVIT 采用的多尺度线性注意力方法可以有效识别密集目标,因此将其加入 YOLOv8 主干网络,以加强密集目标的识别能力。

最后,部分飞鸟体积较小,占整幅图像不足十分之一,属于小目标范畴,YOLOv8 在识别时会出现漏检和识别不准确的情况。高效多尺度注意力(EMA)可以避免特征降维,可以捕获更多的特征信息,有利于小目标的识别,因此将 EMA 引入改进模型,以减少漏检的概率,提升模型对小目标飞鸟的识别性能。

3 YOLOv8 算法改进

3.1 EfficientViT 改进骨干网络

Transformer 模型在处理自然语言方面表现出色,

自注意力机制允许模型同时关注 Input 部分的所有位置,可以捕捉全局语义信息。Dosovitskiy 等[19]将 Transformer 引入计算机视觉方面,提出了 ViT(vision Transformer),通过将图像拆分为像素块的方法将图像嵌入模型,用处理自然语言的方法处理图像。在大规模预训练之后,VIT 在图像处理任务中取得了能与最先进的卷积神经网络相比的良好效果。但密集目标的检测识别效果仍然较差,Xie 等[20]研究证明对密集目标的处理需要全局感受野,Guo 等[21]使用大核注意机制实现了大感受野,但大卷积核对硬件要求较高,该方法存在一定的局限性。Cai 等[22]提出的 EfficientViT,进一步提高模型效率,且在高分辨率的密集预测任务中表现出色,EfficientViT 利用轻量级线性注意力实现了全局感受野和多尺度学习,并且对硬件没有特殊要求,在实验中验证了其对密集目标检测的有效性。因此将 EfficientVIT 的核心模块应用于 YOLOv8 网络架构中,以提高鸟类识别时密集检测的准确率以及模型性能。

图 2 为多尺度线性注意力模块,输入的图像经过线性投影层,生成 $\boldsymbol{Q}/\boldsymbol{K}/\boldsymbol{V}$ 矩阵,再经过 ReLU 注意力模块获取全局感受野,最后再经线性投影层实现特征融合。

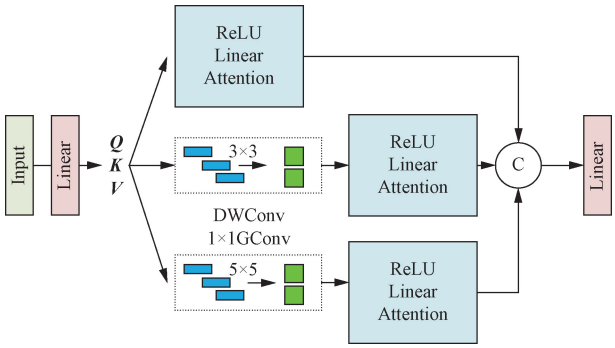


图 2　多尺度线性注意力模型示意图

Fig. 2　Schematic diagram of a multiscale linear attention model

ReLU 注意力的相似度函数为

$$\mathrm{Sim}(\boldsymbol{Q},\boldsymbol{K}) = \mathrm{ReLU}(\boldsymbol{Q})[\mathrm{ReLU}(\boldsymbol{K})]^{\mathrm{T}} \tag{1}$$

当输入为 $\boldsymbol{x} \in \mathbf{R}^{N\times f}$ 时,ReLU 注意力可表示为

$$O_i = \sum_{j=1}^{N} \frac{\mathrm{ReLU}(\boldsymbol{Q}_i)[\mathrm{ReLU}(\boldsymbol{K}_j)]^{\mathrm{T}}}{\sum_{j=1}^{N}\mathrm{ReLU}(\boldsymbol{Q}_i)[\mathrm{ReLU}(\boldsymbol{K}_j)]^{\mathrm{T}}}V_j \tag{2}$$

式中: $\boldsymbol{Q} = \boldsymbol{x}\boldsymbol{W}_Q$, $\boldsymbol{K} = \boldsymbol{x}\boldsymbol{W}_K$, $\boldsymbol{V} = \boldsymbol{x}\boldsymbol{W}_V$,且 $\boldsymbol{W}_Q$、$\boldsymbol{W}_K$、$\boldsymbol{W}_V \in \mathbf{R}^{f\times d}$,都是线性投影矩阵。利用矩阵乘法的可结合性,可将式(2)变形得

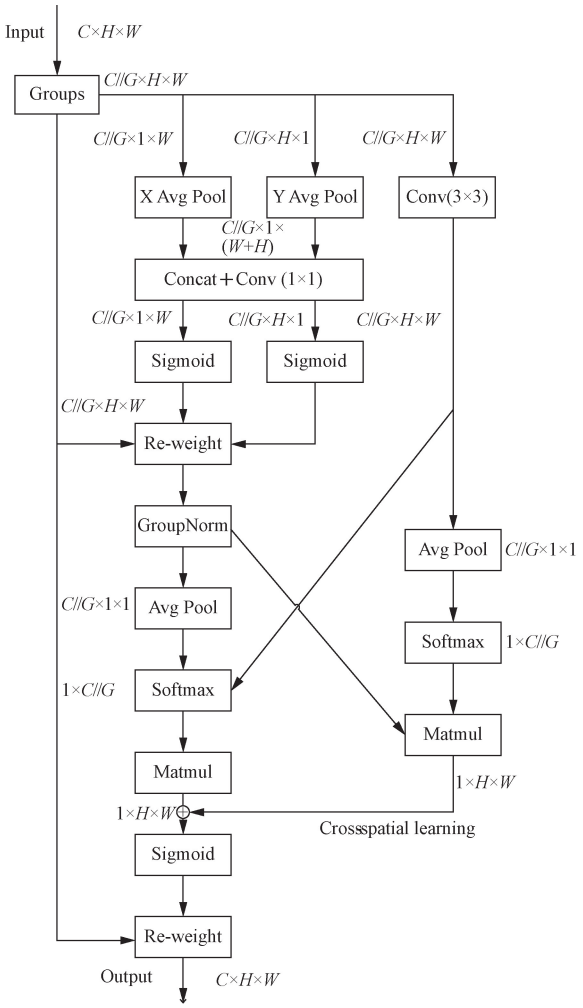
$$O_i = \frac{\mathrm{ReLU}(\boldsymbol{Q}_i)\left\{\sum_{j=1}^{N}[\mathrm{ReLU}(\boldsymbol{K}_j)]^{\mathrm{T}}\boldsymbol{V}_j\right\}}{\mathrm{ReLU}(\boldsymbol{Q}_i)\left\{\sum_{j=1}^{N}[\mathrm{ReLU}(\boldsymbol{K}_j)]^{\mathrm{T}}\right\}} \tag{3}$$

此时就实现了对 $\left\{\sum_{j=1}^{N}[\mathrm{ReLU}(\boldsymbol{K}_j)]^{\mathrm{T}}V_j\right\} \in \mathbf{R}^{d\times d}$ 和 $\left\{\sum_{j=1}^{N}[\mathrm{ReLU}(\boldsymbol{K}_j)]^{\mathrm{T}}\right\} \in \mathbf{R}^{d\times 1}$ 的重复使用,将其引入主干网络后,降低了硬件成本,使得模型更为高效。

3.2　引入 EMA 注意力机制

鸟类数据集常存在目标尺寸变化大的问题,这就使得检测识别时会出现漏检问题。为使模型能更好的关注到小目标鸟类,同时也具备良好的对大尺寸目标检测识别能力,引入了高效多尺度注意力(EMA)[23]。

如图 3 所示,EMA 具有灵活的结构特征,能够融入 CNN 主干架构中,它通过跨空间重塑通道的方式,避免了特征降维,从而能捕获更多的特征信息。因此该模块对于区别不同鸟种的细微特征更具效果。EMA 采用了特征分组的方法,将输入的特征分为 G 个子特征组,分别学习特征语义。这种方法允许模型在多 GPU 资源上部署,同时减弱了噪声的干



C 表示输入通道的数量;H 和 W 分别表示输入特征的空间维度

图 3　EMA 结构示意图

Fig. 3　Schematic diagram of the EMA structure

扰。同时整体采用并行结构,分为 1×1 卷积分支和 3×3 卷积分支,避免了更大的网络深度,提高了参数的利用率。在 1×1 分支中采用了平均池化操作,实现空间编码通道信息,以赋值通道不同的重要性。在 3×3 卷积分支实现多尺度特征捕捉。最后将信息通过 Sigmoid 函数聚合,此时输出结果与输入结果的大小相同,因此可以更方便的将 EMA 加入到卷积神经网络的架构中。通过不同空间维度方向的跨空间信息聚合方法,以实现更丰富的特征聚合,同时这种方法可以处理短程和长程依赖性。与形成的有限感受野的渐进行为相反,并行利用 3×3 和 1×1 卷积捕获中间特征之间的更多上下文信息。因此,引入 EMA 模块可以给模型带来更高的性能增益。

4 实验与结果分析

4.1 数据集

在飞鸟识别的实验中,最常用和权威的数据集是加州理工大学制作的 CUB-200-2011,它包含 200 种鸟类,11 788 张尺寸不一的图片,每张图片都包含细节标注信息,但本文中仅采用其标注框与属性类别信息实现识别。同时,为验证模型适用性,也为使研究更具实际意义,结合《机场常见鸟类防范指南》,构建了常见于中国本土的鸟种数据集 birds28,包含 28 种鸟类,共 5 040 张图片。

图 4、图 5 分别为 CUB-200-2011 和 birds28 的数据集情况。CUB-200-2011 每幅图只有一个目标,数据量与种类更为整齐。而自制数据集中,特意选取了更复杂的密集目标以更贴合实际情况,因此数据量与种类之前会呈现参差状态。

4.2 实验环境与评价指标

实验采用的操作系统为 Windows 11 专业版,处理器为 Intel(R) Core(TM) i7-12700KF,3.6 GHz,32 GB 运行内存,图像处理器为 GeForceRTX3080。实验框架环境为 Pytorch2.3.1,CUDA 版本为 12.2。具体的实验环境参数设置如表 1 所示。

为验证引入多尺度线性模型在鸟类识别任务中

表 1 实验环境参数

Table 1 Experimental environmental parameter

参数	设置
epoch	300
batch	16
Imgsz	640
workers	4
lr0	0.01
momentum	0.937
weight_decay	0.000 5

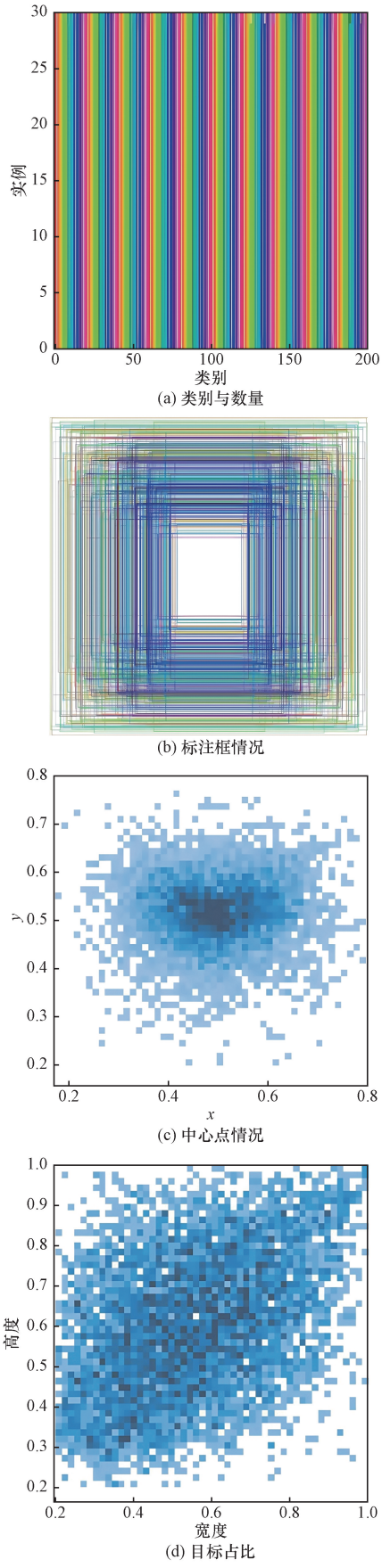


图 4 CUB-200-2011 数据集情况

Fig. 4 Data set situation of CUB-200-2011

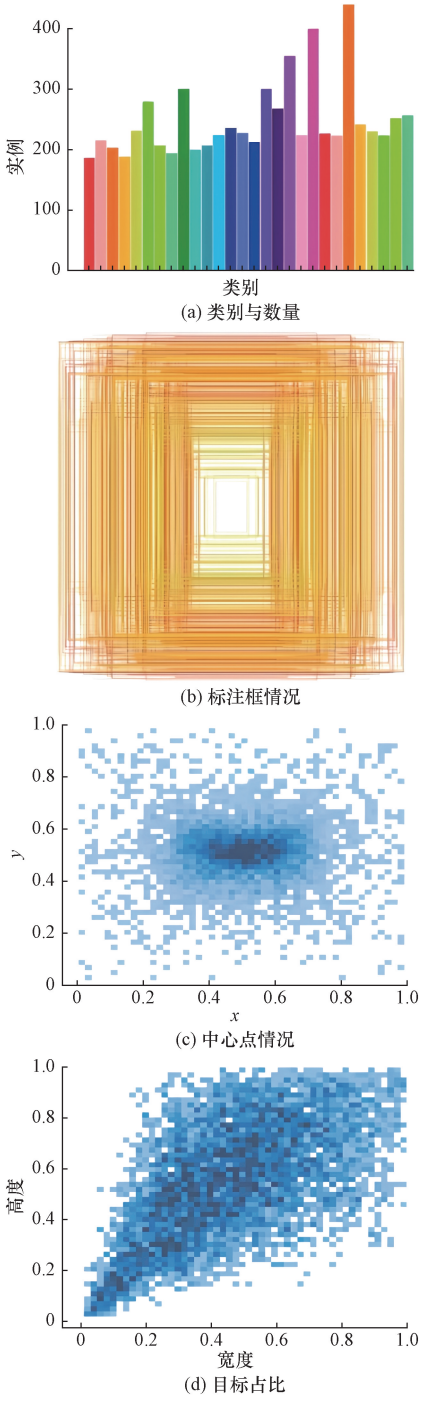


图 5　birds28 数据集情况

Fig. 5　Data set situation of birds28

的有效性,在两个数据集中都对多种 YOLOv8 变体模型进行了对比试验,在所有的比较实验和消融实验中均未采用预训练权值。采用 mAP50 和 mAP50-95 两个评价指标来衡量模型效果,同时确保参数和计算需求处于同一量级。

4.3　消融实验

为验证改进方法的有效性,以原始的 YOLOv8 模型为参考,逐步引入改进策略,分别在 CUB-200-2011 和 birds28 数据集上进行消融实验。在同等的实验条件下,比较 YOLOv8、引入多尺度线性注意力模块的模型(YOLOv8-EfficientViT)、引入跨空间高效多尺度注意力模块的模型(YOLOv8-EMA)、同时引入多尺度线性注意力和跨空间高效多尺度注意力模块的模型(YOLOv8-EfficientViT-EMA)之间的性能。

如表 2、表 3 所示,在引入 EMA 机制后,模型捕捉多尺度变化的能力提升,在 CUB-200-2011 数据集中,mAP50 和 mAP50-95 分别提升 1.6% 和 2%,在 birds28 中 mAP50 和 mAP50-95 分别提升 0.3% 和 0.6%。在引入 EfficientViT 之后,模型对于密集目标的检测识别和性能都有所提升,在 CUB-200-2011 数据集中,mAP50 和 mAP50-95 分别提升 3% 和 2.2%,在 birds28 中 mAP50 和 mAP50-95 分别提升 5.2% 和 5.4%。消融实验证明了模型改进的有效性,最终改进模型相比原始模型在 CUB-200-2011 数据集中 mAP50 和 mAP50-95 分别提升 4.5% 和 3.4%,在 birds28 中 mAP50 和 mAP50-95 分别提升 5.4% 和 6.3%。同时模型的参数量和计算量(GFLOPs)都与原模型处于同一数量级。图 6、图 7 为识别效果对比。

图 6 为原始 YOLOv8 模型的识别效果,可以看到存在漏检和准确度不高的问题,图 7 为改进后的模型识别效果,明显改善了漏检问题,识别的准确性也获得了提高。

表 2　CUB-200-2011 数据集消融实验

Table 2　CUB-200-2011 data set ablation experiment

模型	mAP50	mAP50-95	参数量/10^6	GFLOPs
YOLOv8	0.726	0.62	3.37	9.9
YOLOv8-EMA	0.742	0.64	3.38	10
YOLOv8-EfficientViT	0.756	0.642	4.37	11.2
YOLOv8-EfficientViT-EMA	0.771	0.654	4.39	11.3

表 3　birds28 数据集消融实验

Table 3　birds28 data set ablation experiment

模型	mAP50	mAP50-95	参数量/10^6	GFLOPs
YOLOv8	0.830	0.592	3.01	8.2
YOLOv8-EMA	0.827	0.586	3.02	8.4
YOLOv8-EfficientViT	0.882	0.646	4.01	9.5
YOLOv8-EfficientViT-EMA	0.884	0.655	4.02	9.6

4.4　对比试验

为进一步验证改进模型的优势,将改进后的算法与其他算法进行实验比较,分别在 CUB-200-2011

和 birs28 数据集上验证,以 mAP50 和 mAP50-95 为评价指标。选择了 YOLOv5、YOLOv6、resnet18 以及

图 6 原始 YOLOv8 模型识别效果

Fig. 6 Original YOLOv8 model recognition effect

图 7 改进后的 YOLOv8 模型识别效果

Fig. 7 The improved YOLOv8 model recognition effect

表 4 CUB-200-2011 数据集模型对比实验

Table 4 CUB-200-2011 data set model comparison experiment

模型	mAP50	mAP50-95	参数量/10^6	GFLOPs
YOLOv5	0.687	0.580	2.87	8.9
YOLOv6	0.666	0.572	4.68	14.0
YOLOv8-ghost	0.716	0.615	2.08	6.8
YOLOv8-ghost-EMA	0.718	0.611	2.68	8.2
YOLOv8-BiFPN	0.749	0.639	3.15	9.9
Resnet18	0.770	0.642	13.69	36.9
YOLOv8-EfficientViT-EMA	0.771	0.654	4.39	11.3

表 5 birds28 数据集模型对比实验

Table 5 birds28 dataset model comparison experiment

模型	mAP50	mAP50-95	参数量/10^6	GFLOPs
YOLOv5	0.827	0.577	2.51	7.2
YOLOv6	0.751	0.523	4.24	11.9
YOLOv8-ghost	0.804	0.574	1.72	5.2
YOLOv8-ghost-EMA	0.808	0.569	1.73	5.2
YOLOv8-BiFPN	0.836	0.598	2.79	8.3
Resnet18	0.795	0.539	13.33	35.1
YOLOv8-EfficientViT-EMA	0.884	0.655	4.02	9.6

YOLOv8 的其他变体算法进行对比,在均未采用预训练权重的前提下,模型性能均优于其他算法。如表 4、表 5 所示,虽然 resnet18 在 CUB-200-2011 数据集中取得了与改良后模型相近似的结果,但其参数量和计算需求却远大于改进模型,而且其在 birds28 数据集中表现欠佳,说明其泛化性不好。同模型大小下,改进后的模型性能优于其他算法。

5 结论

针对鸟类识别任务中,由于目标密集导致的识别不准确问题,和由于目标尺寸变化大而引起的漏检问题,提出了基于 YOLOv8 的改进模型。实验结果表明改进的模型在 CUB-200-2011 数据集上,mAP50 和 mAP50-95 分别提升了 4.5% 和 3.4%,验证了模型的有效性,同时在自制数据集 birds28 上,mAP50 和 mAP50-95 分别提升了 5.4% 和 6.3%,验证了模型的泛化性。

参 考 文 献

[1] Zhang C, Shi F, Zhang X, et al. Airport near-altitude flying birds detection based on information compensation multi-scale feature fusion[J]. IEEE Sensors Journal, 2023, 23(19): 22867-22879.

[2] Metz I C, Ellerbroek J, Mühlhausen T, et al. The bird strike challenge[J]. Aerospace, 2020, 7(3): 26.

[3] Ferreira A C, Silva L R, Renna F, et al. Deep learning-based methods for individual recognition in small birds[J]. Methods in Ecology and Evolution, 2020, 11(9): 1072-1085.

[4] Liu H, Zhang C, Deng Y, et al. TransIFC: invariant cues-aware feature concentration learning for efficient fine-grained bird image classification[J]. IEEE Transactions on Multimedia, 2023, 27: 1677-1690.

[5] Wang K, Yang F, Chen Z, et al. A fine-grained bird classification method based on attention and decoupled knowledge distillation [J]. Animals, 2023, 13(2): 264.

[6] Chakraborti T, McCane B, Mills S, et al. CoCoNet: a collaborative convolutional network applied to fine-grained bird species classification[C]//2020 35th International Conference on Image and Vision Computing New Zealand (IVCNZ). New York: IEEE, 2020: 1-6.

[7] 任书杰, 胡勇, 何文祥, 等. 基于深度学习的砂岩组分显微图像识别[J]. 科学技术与工程, 2024, 24(9): 3727-3736.
Ren Shujie, Hu Yong, He Wenxiang, et al. Microscopic image recognition of sandstone components based on deep learning[J]. Science Technology and Engineering, 2024, 24(9): 3727-3736.

[8] 林开颜, 牛程远, 张浩平, 等. 基于深度学习的景观植物颜色特征提取方法[J]. 科学技术与工程, 2024, 24(17): 7059-7065.
Lin Kaiyan, Niu Chengyuan, Zhang Haoping, et al. A method for extracting color characteristics of landscape plants based on deep learning[J]. Science Technology and Engineering, 2024, 24(17): 7059-7065.

[9] 谢威宇, 张强. 基于深度学习的图像中无人机与飞鸟检测研究综述[J]. 计算机工程与应用, 2024, 60(8): 46-55.

Xie Weiyu, Zhang Qiang. Review on detection of drones and birds in photoelectric images based on deep learning convolutional neural network[J]. Computer Engineering and Applications, 2024, 60(8): 46-55.

[10] Won C S. Multi-scale CNN for fine-grained image recognition[J]. IEEE Access, 2020, 8: 116663-116674.

[11] Tan M, Zhou J, Peng Z, et al. Fine-grained image classification with factorized deep user click feature[J]. Information Processing & Management, 2020, 57(3): 102186.

[12] Bellafkir H, Vogelbacher M, Schneider D, et al. Edge-based bird species recognition *via* active learning[C]//International Conference on Networked Systems. Cham: Springer Nature Switzerland, 2023: 17-34.

[13] 宣以国,余成波, 蒋启超, 等. 基于改进 YOLOv7 的道路裂缝和坑洞检测算法[J]. 科学技术与工程, 2024, 24(17): 7205-7213.
Xuan Yiguo, Yu Chengbo, Jiang Qichao, et al. Improved YOLOv7 road crack and pothole detection algorithm[J]. Science Technology and Engineering, 2024, 24(17): 7205-7213.

[14] Kumar A, Das S D. Bird species classification using transfer learning with multistage training[C]//Computer Vision Applications: Third Workshop, WCVA 2018. Singapore: Springer, 2019: 28-38.

[15] 陈天华, 朱家煊, 印杰. 基于注意力机制的鸟类识别算法[J]. 计算机应用, 2024, 44(4): 1114-1120.
Chen Tianhua, Zhu Jiaxuan, Yin Jie. Bird recognition algorithm based on attention mechanism[J]. Journal of Computer Applications, 2024, 44(4): 1114-1120.

[16] Cohen A S, Cont R, Rossier A, et al. Scaling properties of deep residual networks [C]//International Conference on Machine Learning. New York: PMLR, 2021: 2039-2048.

[17] Li Y, Fan Q, Huang H, et al. A modified YOLOv8 detection network for UAV aerial image recognition[J]. Drones, 2023, 7(5): 304.

[18] Wu T, Dong Y. YOLO-SE: improved YOLOv8 for remote sensing object detection and recognition[J]. Applied Sciences, 2023, 13(24): 12977.

[19] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16 × 16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv: 2010. 11929, 2020.

[20] Xie E, Wang W, Yu Z, et al. SegFormer: simple and efficient design for semantic segmentation with transformers[J]. Advances in Neural Information Processing Systems, 2021, 34: 12077-12090.

[21] Guo M H, Lu C Z, Hou Q, et al. Segnext: rethinking convolutional attention design for semantic segmentation[J]. Advances in Neural Information Processing Systems, 2022, 35: 1140-1156.

[22] Cai H, Li J, Hu M, et al. Efficientvit: lightweight multi-scale attention for high-resolution dense prediction[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. New York: IEEE, 2023: 17302-17313.

[23] Ouyang D, He S, Zhang G, et al. Efficient multi-scale attention module with cross-spatial learning[C]// 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New York: IEEE, 2023: 1-5.