

基于 YOLOv5 的红外相机野生动物图像识别

杨铭伦^{1,2}, 张旭^{1,2}, 郭颖^{1,2*}, 于新文^{1,2}, 侯亚男^{1,2}, 高家军^{1,2}

¹中国林业科学研究院资源信息研究所, 北京 100091;

²国家林业和草原局林业遥感与信息技术重点实验室, 北京 100091

摘要 为在红外相机等资源受限平台上实时、准确地实现海量野生动物图像自动识别, 改善野生动物监测过程中数据传输负载重、时效性低等问题, 基于 YOLOv5 模型, 利用 5 类物种的红外相机图像构建数据集, 对 YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x 四种网络结构进行训练。通过对比不同网络结构的精度、检测速度、体积, 明确最优网络结构; 同时分析模型在复杂背景信息干扰下的识别效果, 评价 YOLOv5 在真实野外场景的适用性; 并与其他同类算法的比较, 明确 YOLOv5 用于野生动物识别的优势。实验结果表明: 四种网络结构的识别精度均较高, F1-score 和平均精度 (mAP) 均在 90% 以上, 其中 YOLOv5m 的综合性能最好; YOLOv5 在多种复杂背景信息干扰下识别效果仍较好, 能够很好地适应真实野外场景; 与其他算法相比, YOLOv5 同时具有精度高、鲁棒性强、资源占用低等优势。YOLOv5 是一种轻量化的模型且性能优越, 为在资源受限的平台上进行野生动物实时识别提供了新的契机。

关键词 机器视觉; 野生动物; 目标检测; YOLOv5; 图像识别; 深度学习

中图分类号 TP391.4

文献标志码 A

DOI: 10.3788/LOP202259.1215015

Recognition of Wild Animals Using Infrared Camera Images Based on YOLOv5

Yang Minglun^{1,2}, Zhang Xu^{1,2}, Guo Ying^{1,2*}, Yu Xinwen^{1,2}, Hou Yanan^{1,2}, Gao Jiajun^{1,2}

¹Research Institute of Forest Resource Information Techniques, Chinese Academy of Forestry,
Beijing 100091, China;

²Key Laboratory of Forestry Remote Sensing and Information Technology, State Forestry and
Grassland Administration, Beijing 100091, China

Abstract In this paper, we propose the construction of an extended YOLOv5 model using the infrared camera image datasets of five species to achieve the automatic recognition of massive wild animal images in real-time, accuracy on resource-limited platforms such as infrared cameras. Furthermore, we improve the negative load and low timeliness of data transmission in wildlife monitoring. Here, the dataset constructed is used to train four network structures, namely, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. By comparing the accuracy, detection speed, and volume of different network structures, the optimal network structure was determined. Simultaneously, we analyzed the recognition effect of the model under the interference of complex background information to evaluate the applicability of YOLOv5 in real-field scenes. Compared with similar algorithms, the advantages of YOLOv5 for wildlife recognition outweighed others. The experimental results show that the recognition accuracy of the four

收稿日期: 2021-07-28; 修回日期: 2021-08-23; 录用日期: 2021-08-31

基金项目: 中国林业科学研究院中央级公益性科研院所基本科研业务费专项资金、海南长臂猿智能感知技术研究与应用示范(CAFYBB2019ZB011)

通信作者: *guoying@ifrit.ac.cn

network structures was high. Moreover, F1-score and average accuracy (mAP) were more than 90%, and the comprehensive performance of YOLOv5m was the best. However, YOLOv5 still has a good recognition effect under the interference of several complex background information and can adapt to real-field scenes. Compared with other algorithms, YOLOv5 has the advantages of high precision, strong robustness, and low resource occupation. It is a lightweight model with superior performance, which provides a new opportunity for real-time wildlife identification on resource-constrained platforms.

Key words machine vision; wild animal; target detection; YOLOv5; image recognition; deep learning

1 引言

野生动物是构成生态系统的重要生物类群,具有调控生态系统结构和功能,维持生态系统健康平衡的作用^[1]。野生动物监测可为种群现状评估、区域规划和野生动物保护等宏观决策提供科学依据。红外相机是目前进行野生动物监测的主要工具^[2]。利用红外相机等监测设备获取动物视频和图像数据^[3],具有对动物干扰小、人为限制因素少、全天候、可持续等优势^[4],通过相机的多点布设,可以获取野生动物活动的时间及空间位置信息,能够更加全面地分析野生动物的活动规律,在一定程度上弥补传统人工调查方式的不足。然而,随着红外相机应用的不断普及,所获取的数据量急剧增加,也加大了人工识别的工作负担,因此亟待应用更先进的计算机自动识别方法对海量红外相机数据中的野生动物进行自动、准确、快速识别。

深度学习是人工智能技术的一个分支,近年来在图像识别领域的应用取得了重大突破,在一些大型哺乳动物^[5-7]、鸟类^[8-10]、昆虫^[11]等物种识别方面已经有了许多较为成功的应用案例。然而现已广泛应用的深度学习模型参数量大,对计算和存储资源需求高,难以直接部署在红外相机这类存储和计算资源有限的设备上,野生动物识别只能在将数据传回服务器端后才能进行。然而由于野外环境复杂,红外相机设备经常会出现误触发的现象,将此类图片回传至服务器端会增加网络负载,而且需要大量的人力成本完成误触发图片的筛选,增加了工作量;此外,新拍摄的图片需要回传至服务器端进行预测,在网络条件有限的情况下无法快速响应,这就降低了野生动物监测的时效性。

YOLO是一种经典的单阶段目标检测模型,于2016年由Redmon等^[12-15]提出,至今已迭代至第5代版本。最新的YOLOv5于2020年5月正式发布,与YOLOv1~YOLOv4相比体积更小、推理速度更快,适用于计算和存储资源有限的边缘设备。Hu等^[16]采

用YOLOv5模型进行了道路裂缝检测,准确率高达88.1%,同时该模型对于每张图片的检测时间最短可达到11.1 ms,这表明YOLOv5在进行道路裂缝检测时不仅能达到较高的准确率,而且检测速度较快。Liu等^[17]基于YOLOv5实现了对交通信号灯的实时准确识别,且该模型的单张图片检测时间为10 ms,准确率为97.2%。Yan等^[18]将改进的YOLOv5模型部署到苹果采摘机器人上完成了对苹果的实时检测与采摘,改进后的模型可实时准确地检测多个苹果目标,单张图片检测时间为15 ms,准确率为86.75%。以上研究表明,YOLOv5不仅能够适用于计算和存储资源有限的设备,而且能够快速准确地实现多目标检测。将YOLOv5应用于野生动物图像识别将有望解决红外相机野生动物监测中数据传输负载重、时效性低的问题。但目前YOLOv5在野生动物识别方面的研究还鲜有报道。

因此,本文基于YOLOv5模型和红外相机图像数据研究野生动物智能实时识别技术,利用布设在野外的红外相机获取到的图像数据构建野生动物图像数据集,据此建立基于YOLOv5的野生动物图像最优检测模型,探究不同网络结构对模型检测结果的影响,选取能够在体积、速度与检测精度之间达到良好平衡的最优网络架构;研究野外真实场景中复杂背景信息干扰的条件下模型的表现;同时探讨YOLOv5在野生动物识别中的优势所在。

2 YOLOv5模型简介

YOLOv5模型由Ultralytics公司在2020年5月正式提出,该模型在YOLOv1~YOLOv4的基础上进行了改进,并融入了近年来目标检测领域的一些新技术来提高模型检测性能,模型代码基于当前最流行的深度学习框架PyTorch实现。与YOLOv1~YOLOv4所使用的Darknet框架相比,生态较为成熟,软硬件支持相对完善,支持在多种设备上部署。

YOLOv5分为4种不同大小的模型,分别是

v5s、v5m、v5l、v5x, 其权重文件大小分别是 14, 42, 93, 170 MB, 随着模型体积的增大, 模型检测精度不断提高, 检测速度不断下降, 资源占用显著增加。YOLOv5 大体框架可分为输入端(input)、骨干网络(backbone)、neck 端、输出端(output)4 个部分。在输入端, YOLOv5 采用的技术有 mosaic 数据增强、自适应锚框估计、自适应图片缩放等。Mosaic 数据增强技术通过随机缩放、随机裁剪、随机排布等方式将 4 张图片融合成 1 张图片, 在丰富图片背景的同时提高模型对小目标的检测能力; 自适应锚框估计技术将训练集上的标注框进行聚类, 可自适应计算出最佳锚框值; 自适应图片缩放技术将原始图片缩放至输入尺寸时, 可以自适应地在图像周围添加

最少的黑边以减小计算量、加快推理速度。YOLOv5 的骨干网络中使用了 CSP 结构及 focus 模块。Neck 层使用 FPN+PAN 结构加强网络的特征融合能力。输出端使用 GIoU_Loss^[19] 代替传统的 IoU_Loss 作为 bounding box 损失函数, GIoU 克服了传统 IoU 存在的缺陷, 将 GIoU 用于目标检测模型可以提高模型检测性能, 其计算方法为

$$L_{GIoU} = L_{IoU} - \frac{|A_c - U|}{|A_c|}, \quad (1)$$

式中: L_{IoU} 为传统的交并比, 即预测框和真实框的交集面积与其并集面积的比值; A_c 为预测框与真实框的最小外包矩形的面积; U 为预测框与真实框并集的面积。YOLOv5 的总体结构如图 1 所示。

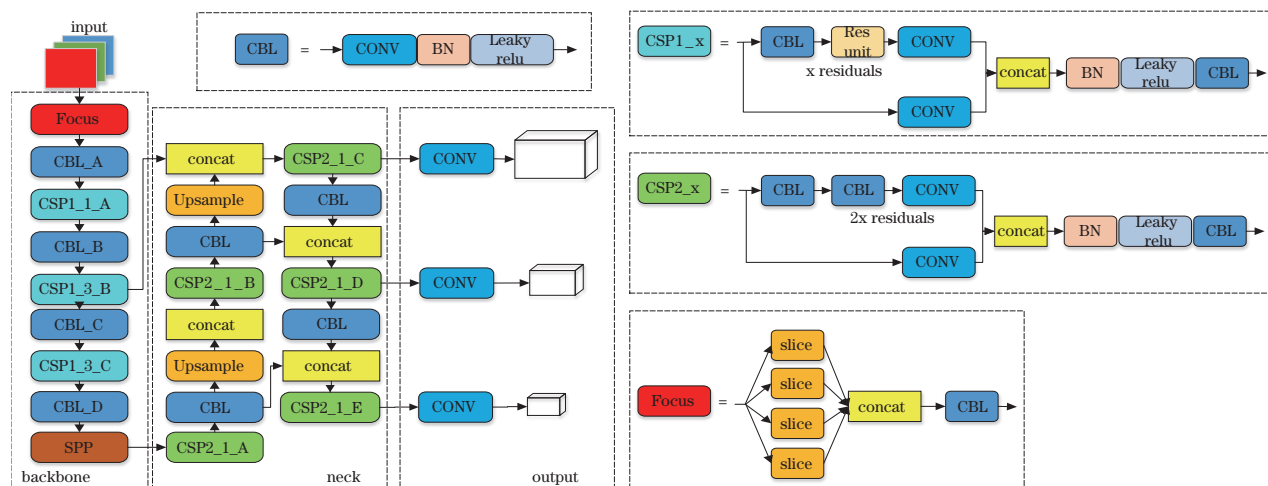


图 1 YOLOv5 网络结构图

Fig. 1 Network structure diagram of YOLOv5

YOLOv5 的 4 种模型总体结构相同, 其区别主要在于模型深度与宽度方面, 即模型层数与卷积层个数不同, v5s、v5m、v5l、v5x 模型层数与卷积核个数依次增加, 具体如表 1 和表 2 所示。网络层数和

表 1 四种网络结构深度对比

Table 1 Depth comparison of four network structures

Residual component	YOLOv5s	YOLOv5m	YOLOv5l	YOLOv5x
CSP1_1_A	1	2	3	4
CSP1_3_B	3	6	9	12
CSP1_3_C	3	6	9	12
CSP2_1_A	1	2	3	4
CSP2_1_B	1	2	3	4
CSP2_1_C	1	2	3	4
CSP2_1_D	1	2	3	4
CSP2_1_E	1	2	3	4

表 2 四种网络结构宽度对比

Table 2 Width comparison of four network structures

Filter model	YOLOv5s	YOLOv5m	YOLOv5l	YOLOv5x
Focus	32	48	64	80
CBL_A	64	96	128	160
CBL_B	128	192	256	320
CBL_C	256	384	512	640
CBL_D	512	768	1024	1280

卷积核个数的增加通常意味着模型具有更强的特征提取能力, 但同时也会加大存储和计算资源的负载。

3 材料与实验

3.1 数据采集与处理

所使用的数据是 2010—2014 年间布设在神农架保护区内的红外相机拍摄所得的, 多台红外相机

被布设在保护区内的不同地点并进行长期连续性监测,使获取到的数据具有复杂多样的背景信息、不同的光照条件及不同的季节特征,季节特征包括春夏季枝叶茂密的背景、秋季枯枝落叶背景、冬季

雪地背景等,地点涵盖植被茂密的林内及相对空旷的裸土地等,如图 2 所示。由于这些数据全部取自真实的野外场景,因此能够更好地探究模型在真实环境下的表现。



图 2 图像背景信息的复杂性

Fig. 2 Complexity of image background

所获取数据包含图像与视频两种,对于视频数据先按照 1 s 的采样间隔采样成图片,然后将部分相似度较高的图片剔除,保留有效图片。本实验组评估不同模型在进行野生动物分类识别方面的表现时,选取了 5 类较为常见的物种作为识别对象,分别是梅花鹿、毛冠鹿、斑羚、野猪、红腹角雉。使用图像标注软件 labeling 进行人工标注,框选出图像中的目标,最终得到 3368 张 5 类物种的图片,每张图片分辨率在 $1440 \times 1080 \sim 4000 \times 3000$ 不等。将数据集按照 8:1:1 的比例划分成训练集、验证集和测试集,其中训练集共 2727 张图片,验证集共 304 张图片,测试集共 337 张图片。数据集中各类物种图片数量如表 3 所示。

表 3 各类别图片数量统计

Table 3 Number of images in each class

Class	Train	Val	Test	Total
Sika deer	650	73	93	816
Tufted deer	538	54	63	655
Wild boar	571	74	72	717
Impala	567	71	69	707
Tragopan temminckii	401	32	40	473
Total	2727	304	337	3368

3.2 模型训练

模型训练使用的硬件环境: Intel i5-7 代/4 核 CPU, 12 GB 内存, NVIDIA GTX 1080 Ti GPU。软件环境: Windows 10 操作系统, Python 3.8 开发语言, PyTorch 深度学习框架, CUDA 10.1。

模型训练时,使用 COCO 数据集上预训练好的权重文件进行模型的初始化,模型输入图片尺寸需要为 32 的倍数,在多数情况下设置为 416×416 , 608×608 , 640×640 等几种,本实验组选择的输入图片尺寸为 608×608 。超参数设置上, batch size 设置为 16, 初始学习率为 0.01, 动量为 0.937, 权重衰减系数为 0.0005, 采用 SGD 优化器对参数训练过程进行优化。

3.3 评价指标

以精确率 (precision)、召回率 (recall)、平均精度 (mAP) 及 F1 分数 (F1-score) 作为评价指标来评估模型训练及预测的效果。精确率表示被检出的数据里有多少是正样本, 召回率表示数据集里的正样本有多少被检出。精确率和召回率的表达式为

$$R_{\text{precision}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}}}, \quad (2)$$

$$R_{\text{recall}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}}, \quad (3)$$

式中: N_{TP} 表示将正样本预测为正的个数; N_{FP} 表示将负样本预测为正的个数; N_{FN} 表示将正样本预测为负的个数。

$mAP^{[20]}$ 用于多类别检测任务的精度评价, 是通过对所有类别的检测精度求平均值得到的, 表达式为

$$P_{mAP} = \frac{1}{N} \sum_{n=1}^I P_{AP}(n), \quad (4)$$

式中: P_{AP} 是以 precision 为纵坐标, 以 recall 为横坐标绘制出的曲线的积分, 代表单个类别的平均精度。

F1 分数是精确率和召回率的调和平均值, 表达式为

$$S_{F1} = 2 \times \frac{R_{precision} \times R_{recall}}{R_{precision} + R_{recall}}. \quad (5)$$

4 结果与分析

4.1 基于 YOLOv5 的野生动物检测结果分析

4.1.1 训练轮数对总损失值的影响

为了得到 YOLOv5 最优的训练轮数, 首先研究了模型总损失值随训练轮数的变化趋势, 结果如图 3 所示。从图中可以看出, 随着训练轮数的增加, 总损失值呈稳步下降的趋势。各模型总损失值变化趋势基本一致, 均呈稳步下降趋势。在相同训练轮数的条件下各模型总损失值的差距相对较小, 说明模型深度与宽度的增加对模型总损失值的影响不大。训练轮数为 200 左右时, 各模型均开始逐渐趋于平稳, 300 时各模型的总损失值基本稳定在 0.02 上下, 此时各模型均已满足收敛的条件, 由此确定 300 为各模型的最佳训练轮数。

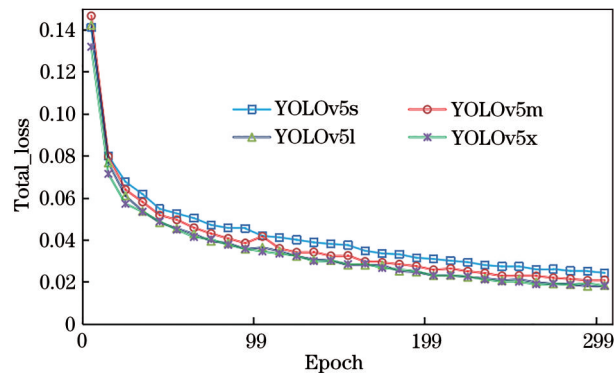


图 3 YOLOv5 总损失值变化趋势图

Fig. 3 Trend of total loss of YOLOv5

4.1.2 YOLOv5 四种网络结构的检测结果分析

YOLOv5s、YOLOv5m、YOLOv5l 和 YOLOv5x 四种型号模型的参数量依次增加, 为获取不同尺寸的模型对检测效果的影响, 以确定最优的模型, 对四种

型号模型的检测结果进行了对比分析, 结果如表 4 所示。从表中可以看出, YOLOv5 四种型号模型的 F1-score 均高于 0.9, mAP 均大于 0.96。检测效果最好的模型为 YOLOv5x, YOLOv5m 和 YOLOv5l 的检测结果相似, 但是 YOLOv5m 略优于 YOLOv5l。从推理时间来看, 运行速度最快的模型是 YOLOv5s, 每张 608×608 的图片平均耗时为 9.5 ms, 其次是 YOLOv5m 和 YOLOv5l, 运行最慢的是 YOLOv5x, 耗时为 21.7 ms, YOLOv5l 略快于 YOLOv5x, 但较 YOLOv5s 与 YOLOv5m 仍较慢。四种模型的运行时间与模型体积成正比, 体积较小的模型参数量较少、运行速度快、资源占用低; 相反, 体积较大的模型则需要更多的存储和计算资源以及更长的运行时间。

表 4 各模型在测试集上的实验结果对比

Table 4 Comparison of experimental results of each model on the test set

Model	F1-score	mAP	Model size /MB	Inference time /ms
YOLOv5s	0.917	0.965	14	9.5
YOLOv5m	0.91	0.961	42	12.9
YOLOv5l	0.908	0.96	93	17.9
YOLOv5x	0.918	0.965	170	21.7

在对检测效果定量分析的基础上, 本实验组进一步通过人工检查的方式, 对比了上述四个型号的 YOLOv5 模型在 337 张测试图片的检测效果。四种模型在 337 张测试图片上的错误检漏检数量分别是 28, 20, 21, 21。四种模型错误检测数量相当, YOLOv5m 略优于其他几种模型, YOLOv5l 与 YOLOv5x 效果相近, YOLOv5s 出错量相对较多, 主要是由于该模型容易将夜晚出现的斑羚错检为体型相似的梅花鹿和毛冠鹿, 出现这一现象的主要原因可能是 YOLOv5s 较其他几种模型网络层数较少, 难以有效地提取较为相似物种特征。

从模型对各类物种的检测结果来看, 检测结果中毛冠鹿被误检为梅花鹿的数量最多, 在 YOLOv5s 中有 6 例, v5m 有 4 例, v5l 与 v5x 各 5 例, 斑羚也容易被误检为梅花鹿, v5s 中有 7 例, v5m 有 3 例, v5l 与 v5x 各 2 例, 发生这些误检的情况主要集中于夜间图像, 白天图像的误检相对较少。另外, 梅花鹿被误检为斑羚和毛冠鹿的情况相对较少, 主要原因可能与训练样本数量有关, 该类别训练样本最多, 模型提取到的特征较为充分, 所以即使在夜间图像特征不明显的环境下也有较高的识别率。对于夜间图片检测率较低的问题, 可以考虑采用更多的训练样

本来达到更好的训练精度。各模型对于不同类别的检测效果统计如表 5 所示。

表 5 误检漏检数量统计

Table 5 Statistics of number of false and missed detections

Class	YOLOv5s	YOLOv5m	YOLOv5l	YOLOv5x
Sika deer	3	6	2	2
Tufted deer	12	9	12	10
Boar	0	0	0	2
Impala	12	4	4	3
Tragopan temminckii	1	1	3	1
Empty prediction box	0	0	0	2
Total	28	20	21	21

从以上结果来看, YOLOv5m 在具有较快检测速度和较小体积的同时还能保持较好的检测效果, 因此是 4 个模型中性能最佳的模型。

4.2 YOLOv5m 与 YOLOv3-tiny 对比分析

为进一步验证 YOLOv5m 模型在野生动物图像检测中的准确性, 对比了 YOLOv5m 与经典轻量

化目标检测模型 YOLOv3-tiny 的检测结果, 结果如表 6 所示。从表 6 可以看出, YOLOv5m 在 F1-score、mAP 的上均显著高于 YOLOv3-tiny。YOLOv3-tiny 虽具有较小的体积和较快的速度但同时精度损失严重, 说明相比于 YOLOv3-tiny 而言, YOLOv5m 更适用于对野生动物图像的检测。

表 6 YOLOv5m 与 YOLOv3-tiny 对比

Table 6 Comparison of YOLOv5m and YOLOv3-tiny

Model	F1-score	mAP	Weight / MB	Inference time / ms
YOLOv5m	0.91	0.961	42	12.9
YOLOv3-tiny	0.842	0.943	16.7	3.7

YOLOv3-tiny 在测试数据中共出现 46 例错误检测, 远高于 YOLOv5, 图 4 为 YOLOv5m 与 YOLOv3-tiny 的部分检测结果。从图 4 可以看出, 在夜间图像中, 动物特征不明显的情况下, YOLOv3-tiny 容易出现漏检误检, 而 YOLOv5m 依然可以准确检测目标。此外, YOLOv5 模型生成的检测框可以准确框



图 4 YOLOv3-tiny 与 YOLOv5m 检测效果。(a) YOLOv3-tiny; (b) YOLOv5m
Fig. 4 Detection effects of YOLOv3-tiny and YOLOv5m. (a) YOLOv3-tiny; (b) YOLOv5m

选出图像中的动物目标,而 YOLOv3-tiny 的结果中多出现检测框与实际物体不完全拟合的情况。由此可见,相比于 YOLOv3-tiny, YOLOv5m 检测效果较好,具有较高的准确性和良好的鲁棒性。

5 讨 论

5.1 模型对比评估

真实的野外场景中图像背景、光照条件、障碍物对动物的遮挡等都是影响到模型检测效果的重要因素。红外相机等数据采集设备具有较低的计算能力和较小的内存。因此在这类设备上部署动物图像识别功能时,模型既需要具有较小体积也需要保持合理的精度。

YOLO 作为单阶段目标检测模型,与 Faster-RCNN^[21-22] 等两阶段模型相比速度较快且资源占用较小,但即便如此,在运行时仍需要较大的计算和内存开销,难以适用于红外相机等设备资源受限平台。YOLOv5 则比以往的 YOLOv1-v4 更加轻量化,体积最大的 YOLOv5x 也仅为 YOLOv3 和 YOLOv4 的 1/2 左

右,这就为资源受限平台上的部署及应用带来了可能。已有研究表明, YOLOv5 在检测速度与精度上均优于 YOLOv3 与 YOLOv4 两者^[23]。在 YOLOv5 的四种网络结构中, YOLOv5m 的体积仅有 YOLOv5l 的 1/2 左右, YOLOv5x 的 1/4 左右,但在实验数据集上的检测效果已经与后者较为接近甚至高于后两者,能够在模型体积、检测速度和检测精度之间达到良好的平衡。此外,第 4.2 节实验结果表明,与同样适用于资源受限平台上的经典轻量化模型 YOLOv3-tiny 相比, YOLOv5m 不仅精度更高而且鲁棒性更强,优越性更加突出。因此, YOLOv5m 更适合被应用在红外相机等设备上进行野生动物识别和快速响应。

5.2 模型在复杂背景下的鲁棒性分析

在森林环境中获取的图像包含的背景信息往往较为复杂,树木枝叶遮挡以及动物身躯与背景色相混淆等情况也大量存在,这在一定程度上增加了野生动物识别的难度,因此在复杂的图像背景中准确识别野生动物信息也是衡量模型性能和实用性的重要因素。

所采用的 YOLOv5 模型的部分检测结果如图 5



图 5 复杂背景下检测结果示例

Fig.5 Examples of detection results in complex background

所示。从图中可以看出:图像中光线较暗或存在反光现象的条件下,模型仍可有效检测出目标;当动物自身的颜色与背景色十分相近时仍可有效检测;模型同时具有一定的抗遮挡能力,在一些场景中动物目标被较多的树木枝叶遮挡,也可以达到很好的效果;此外,在动物只有身体的某个部位出现在镜头中时,也可以准确的进行识别。这些检测结果说明,YOLOv5模型能够有效应对真实野外场景中各种复杂背景信息的干扰,具有良好的鲁棒性。

5.3 模型对白天与夜间数据的检测效果

在检测结果中,各模型检测夜间图像的正确率要低于白天图像,夜间图像为黑白图像,图像质量比白天RGB图像低,且图像上物体的纹理特征不明显,这是导致夜间图像判别率较低的主要原因。在各类物种的识别率方面,发现:各模型对梅花鹿、野猪和红腹角雉的识别准确率较高,出现的错误较少,但对于毛冠鹿和斑羚的误判数量较多。其主要原因是野猪和红腹角雉两个物种与其他物种的形态差异较大,而毛冠鹿、斑羚在形态上则具有一定相似性。这说明在夜间图像分辨率较低,且动物的纹理特征及色彩信息无显著差异时,形态及轮廓相似的动物发生误判的概率将会显著升高,这与文献[7]中基于YOLOv3的研究结果一致。此外,夜间图像存在的反光、拖影等因素也是影响检测效果的重要因素。

6 结 论

基于YOLOv5深度学习模型实现了红外相机野生动物图像识别,并得到了以下结论:

1) 总体来看,四种YOLOv5模型经过训练,F1-score和mAP均可达到0.9以上,具有较高的识别精度,对大部分图片均能准确识别,总体上能够较好地满足野生动物图像识别的需要。四种模型中,YOLOv5m不仅在精度上与YOLOv5l和YOLOv5x十分接近,而且同时具备体积小和检测速度快的优势,在精度、模型体积、检测速度之间达到了良好的平衡,因而在四种模型中的综合表现最好。

2) 真实场景下检测结果的表明,YOLOv5模型在真实场景中具有良好的鲁棒性,在野外光线较弱、反光、物体遮挡、前景色不明显、目标在图像中显示不完整等多种复杂情况下依然具有良好的检测效果,对野外环境具有较强的适应能力。但同时,模型对于夜间黑白影像以及形态较为相近的物

种图片也存在较多的误检和漏检,在今后研究中可通过扩充训练样本以及对夜间图像进行图像增强等方式提高模型的识别率。

3) YOLOv5与以往的YOLOv1-v4、Faster-RCNN等模型相比,具有轻量化的优势,能够相对容易地被部署在红外相机等资源受限的设备上;与其他轻量化模型相比具有精度高、鲁棒性好等优势,具有较高应用价值和应用潜力。

参 考 文 献

- [1] 肖文宏,周青松,朱朝东,等. 野生动物监测技术和方法应用进展与展望[J]. 植物生态学报, 2020, 44(4): 409-417.
Xiao W H, Zhou Q S, Zhu C D, et al. Advances in techniques and methods of wildlife monitoring[J]. Chinese Journal of Plant Ecology, 2020, 44(4): 409-417.
- [2] Li S, McShea W J, Wang D J, et al. A direct comparison of camera-trapping and sign transects for monitoring wildlife in the Wanglang National Nature Reserve, China[J]. Wildlife Society Bulletin, 2012, 36(3): 538-545.
- [3] Garcia-Sanchez A J, Garcia-Sanchez F, Losilla F, et al. Wireless Sensor Network deployment for monitoring wildlife passages[J]. Sensors, 2010, 10(8): 7236-7262.
- [4] 吴逊涛,卜书海,李先敏,等. 基于红外相机技术的太白山自然保护区野生动物监测[J]. 西北林学院学报, 2017, 32(3): 170-175.
Wu X T, Bu S H, Li X M, et al. Monitoring of wildlife with infrared camera in Taibaishan nature reserve [J]. Journal of Northwest Forestry University, 2017, 32(3): 170-175.
- [5] 孙蕊,张旭,郭颖,等. 基于Faster R-CNN金丝猴优化检测方法[J]. 激光与光电子学进展, 2020, 57(12): 121022.
Sun R, Zhang X, Guo Y, et al. Optimized detection method for snub-nosed monkeys based on faster R-CNN[J]. Laser & Optoelectronics Progress, 2020, 57(12): 121022.
- [6] Deb D, Wiper S, Gong S X, et al. Face recognition: primates in the wild[C]//2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems, October 22-25, 2018, Redondo Beach, CA, USA. New York: IEEE Press, 2018: 18619856.
- [7] 宫一男,谭孟雨,王震,等. 基于深度学习的红外相机动物影像人工智能识别:以东北虎豹国家公园为

- 例[J]. 兽类学报, 2019, 39(4): 458-465.
- Gong Y N, Tan M Y, Wang Z, et al. AI recognition of infrared camera image of wild animals based on deep learning: northeast tiger and leopard national park for example[J]. *Acta Theriologica Sinica*, 2019, 39(4): 458-465.
- [8] 肖理想, 罗泽. 基于 Android 和卷积神经网络的鸟类识别系统[J]. *计算机系统应用*, 2019, 28(9): 58-64.
- Xiao L X, Luo Z. Bird identification system based on android and convolutional neural network[J]. *Computer Systems & Applications*, 2019, 28(9): 58-64.
- [9] 邓旭冉, 闵少波, 徐静远, 等. 深度细粒度图像识别研究综述[J]. *南京信息工程大学学报(自然科学版)*, 2019, 11(6): 625-637.
- Deng X R, Min S B, Xu J Y, et al. A survey of deep fine-grained visual categorization[J]. *Journal of Nanjing University of Information Science & Technology (Natural Science Edition)*, 2019, 11(6): 625-637.
- [10] 刘坚. 基于深度神经网络的鸟类图像识别系统设计[J]. *农业装备与车辆工程*, 2019, 57(9): 113-116.
- Liu J. Design of bird image recognition system based on DNN[J]. *Agricultural Equipment & Vehicle Engineering*, 2019, 57(9): 113-116.
- [11] 董本志, 聂丽娜, 景维鹏, 等. 基于 Faster R-CNN 的榆紫叶甲虫识别方法研究[J]. *计算机工程与应用*, 2018, 54(23): 89-93, 108.
- Dong B Z, Nie L L, Jing W P, et al. Identification method of ambrostoma quadriimpressum motschlsky based on Faster R-CNN[J]. *Computer Engineering and Applications*, 2018, 54(23): 89-93, 108.
- [12] 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, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.
- [13] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6517-6525.
- [14] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08) [2021-01-15]. <https://arxiv.org/1804.02767>.
- [15] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23) [2021-01-15]. <https://arxiv.org/abs/2004.10934>.
- [16] Hu G X, Hu B L, Yang Z, et al. Pavement crack detection method based on deep learning models[J]. *Wireless Communications and Mobile Computing*, 2021, 2021: 5573590.
- [17] Liu W T, Wang Z Y, Zhou B, et al. Real-time signal light detection based on Yolov5 for railway[J]. *IOP Conference Series: Earth and Environmental Science*, 2021, 769(4): 042069.
- [18] Yan B, Fan P, Lei X Y, et al. A real-time apple targets detection method for picking robot based on improved YOLOv5[J]. *Remote Sensing*, 2021, 13(9): 1619.
- [19] Rezatofighi H, Tsoi N, Gwak J, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 658-666.
- [20] Chatzichristofis S A, Iakovidou C, Boutalis Y S, et al. Mean Normalized Retrieval Order (MNRO): a new content-based image retrieval performance measure [J]. *Multimedia Tools and Applications*, 2014, 70(3): 1767-1798.
- [21] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137-1149.
- [22] 文灿华, 李佳, 董雪. 基于 Faster RCNN 的生活垃圾智能识别[J]. *激光与光电子学进展*, 2020, 57(20): 201014.
- Wen C H, Li J, Dong X. Intelligent domestic garbage recognition based on faster RCNN[J]. *Laser & Optoelectronics Progress*, 2020, 57(20): 201014.
- [23] Liu K Y, Tang H T, He S, et al. Performance validation of yolo variants for object detection[C]// Proceedings of the 2021 International Conference on Bioinformatics and Intelligent Computing, January 22, 2021, Harbin China. New York: ACM Press, 2021: 239-243.