

融入注意力机制的东北虎和东北豹图像识别算法研究

李雪冬^① 韩姝^① 杨拂晓^② 费龙^① 闫泰辰^{③④}

① 长春师范大学地理科学学院 长春 130032; ② 吉林高分遥感应用研究院有限公司 长春 130021; ③ 吉林省林业勘察设计研究院 长春 130022; ④ 东北虎豹国家公园管理局 长春 130000

摘要: 为解决自动感应红外相机在野外环境监测中遇到树木遮挡、背景干扰、夜间识别困难等问题,本研究通过构建融入注意力机制的深度学习模型作为目标识别的基础框架,实现更高效、更精准的野生动物识别方法。本研究选取东北虎豹国家公园内的东北虎 (*Panthera tigris altaica*)、东北豹 (*P. pardus orientalis*)、野猪 (*Sus scrofa*)、梅花鹿 (*Cervus nippon*) 和狍子 (*Capreolus pygargus*) 五物种作为研究对象,提出融入注意力机制模块并实现局部跨通道交流的卷积神经网络模型,以实现降低复杂背景环境对目标识别的影响,完成区分昼夜、不同角度及不同场景下的动物精准识别。结果表明:本数据集下 YOLO_v5m 算法平均精度均值为 86.67%,引入迁移算法后平均精度均值为 91.16%,提高了 4.49%,有较好的识别效果,且训练时长缩短了 106 min;在迁移算法的基础上融入 CA、CBAM、SE 和 ECA 四类注意力机制后,CA 注意力具有良好的性能,平均精度为 93.72%,相比另外三种注意力机制分别提高了 1.85%、1.78%、1.05%。此外,融入注意力机制的深度学习模型还具有精度高、稳定性强等优势,更适用于复杂背景下的东北虎与东北豹识别。

关键词: 野生动物;深度学习;红外相机;迁移算法;注意力机制

中图分类号: Q958 **文献标识码:** A **文章编号:** 0250-3263 (2025) 06-814-11

An Image Recognition Algorithm Integrated With Attention Mechanism for Amur Tiger and Amur Leopard

LI Xue-Dong^① HAN Shu^① YANG Fu-Xiao^② FEI Long^① YAN Tai-Chen^{③④}

① School of Geographic Sciences, Changchun Normal University, Changchun 130032; ② Jilin Institute of GF Remote Sensing Application, Changchun 130021; ③ Jilin Forestry Survey and Design Research Institute, Changchun 130022; ④ Northeast China Tiger and Leopard National Park Administration, Changchun 130000, China

Abstract: [Objectives] The Amur Tiger and Amur Leopard are endangered protected animals, and using efficient means to identify and monitor them is of great significance for the conservation of species diversity.

基金项目 吉林省科技发展计划项目 (No. YDZJ202301ZYTS221), 长春师范大学自然科学基金项目 (长师大自科合字 (2019) 第 08 号);

第一作者介绍 李雪冬, 男, 讲师; 研究方向: 智能影像识别; E-mail: lixuedong@ccsfu.edu.cn。

收稿日期: 2024-05-06, 修回日期: 2024-11-29 DOI: 10.13859/j.cjz.202524108 CSTR: 32109.14.cjz.24108

In order to solve the problems of tree occlusion, background interference, and difficulty in nighttime identification encountered in infrared camera monitoring in the wild, this study builds a deep learning model incorporating attention mechanism as a basic framework for target recognition, providing an efficient and accurate method for wildlife identification. **[Methods]** This study captured video images of wild animals by deploying automatic infrared cameras in the Northeast China Tiger and Leopard National Park. Eight hundred videos were selected for keyframe extraction. After noise removal, image enhancement, and image calibration, a dataset composed of 11 020 images was constructed for five species: Amur Tiger, Amur Leopard, Wild Boar, Sika Deer, and Roe Deer. This study proposed a convolutional neural network model integrating an attention mechanism module and realizing local cross-channel communication to reduce the impact of complex background environments on target recognition. This model achieved precise identification of animals in different scenarios, including day and night, different angles, and different scenes. The recognition performance of the model was evaluated via metrics such as average precision, recall, accuracy, and F1 score. **[Results]** The mean average precision value of the YOLO_v5m algorithm was 86.67%. After introduction of transfer learning, the mean average precision value was increased to 91.16%, and the time consumption was shortened by 106 min. Among the four types of attention mechanisms: CA, CBAM, SE, and ECA, the CA attention mechanism exhibited the best performance, achieving the average accuracy of 93.72%, which was 1.85%, 1.78%, and 1.05% higher than the other attention mechanisms, respectively (Fig. 5). **[Conclusion]** The deep learning model proposed in this study, which integrates transfer learning and attention mechanism, has the advantages of high accuracy and strong robustness, balancing training speed and recognition accuracy. By deploying infrared cameras to capture images of wild animals, this study can better test the potential of the model under the real living conditions of wild animals. The improved model in this study is more suitable for the identification of Amur Tigers and Amur Leopards in complex backgrounds.

Key words: Wild animal; Deep learning; Infrared camera; Transfer learning; Attention mechanism

东北虎 (*Panthera tigris altaica*) 与东北豹 (*P. pardus orientalis*) 位于森林生态系统食物链顶端, 是国家一级重点保护野生动物, 被世界自然保护联盟 (International Union for Conservation of Nature, IUCN) 分别评估为濒危级与极危级物种, 具有非常重要的保护与科研价值。借助红外相机技术开展野生动物监测是目前野生动物保护最直接和有效的手段, 红外相机具有对动物干扰小、背景真实以及全天候等优势 (Norouzzadeh et al. 2018, 丰大林等 2022)。随着红外相机的广泛应用, 监测所获得的图像数量迅速增长, 面对如此庞大的数据量, 传统的人工筛选鉴别方式表现不足, 耗费周期长且误判率高 (王劭文等 2023, 柯澳等 2024), 亟需运用新的方法实现对海量数据集的准确分

析。韩家臣 (2021) 分析当前野生动物识别所面临的问题, 通过反向传播及引入代价敏感算法, 增强了对野生动物局部特征的识别准确度。阎婧宇等 (2024) 借助无人机实时跟踪野生动物获取监测图像。大型猫科动物倾向于夜间狩猎, 但夜间目标图像与背景相近, 辨识度较低。何嘉 (2019) 针对如何提升红外图像的识别效率和准确度提出了红外图像的域迁移机制。宫一男等 (2019) 则利用迁移学习的方法提升模型的学习效率, 从而提高识别准确率。李仁杰 (2018) 建立了大规模的灵长类动物面部数据集, 在大规模训练样本下弥补学习效率的不足。

随着深度学习技术的不断发展, 越来越多的学者尝试将深度学习应用到图像识别研究领域 (Hinton et al. 2006, Cai et al. 2018)。

Krizhevsky 等 (2012) 提出的 AlexNet 卷积神经网络模型引发了目标检测研究的热潮, Girshick 等 (2014) 提出的基于区域的卷积神经网络 (region-based convolutional neural networks, RCNN) 为以后各大经典算法提供了基础思路, He 等 (2015) 提出的空间金字塔池化卷积网络 (spatial pyramid pooling convolutional networks, SPPNet) 经典结构克服了 RCNN 中重叠卷积难题。Ren 等 (2017) 将目标检测真正融入到卷积神经网络中, 提出了经典的双阶段目标检测算法 Faster RCNN 模型, 至今仍被广泛应用于目标检测领域 (孙蕊等 2020)。同年经典的单阶段目标检测模型 YOLO 算法凭借其轻量型迅速引起热议 (Redmon et al 2015, 林典敏等 2023), 该算法的出现提高了目标检测的实时性与实用性, 使动物识别变得更加方便快捷。杨铭伦等 (2022) 对比 YOLO 算法的 v5s、v5m、v5l、v5x 四种网络模型, 验证 YOLO_v5m 的综合性能最好。利用深度学习模型可以有效地完成图像识别任务。野生动物活动轨迹复杂, 且存在复杂背景信息干扰, 深度学习在提取红外图像信息时特征丢失严重, 如何实现复杂背景下动物的特征识别成为亟待解决的问题。

为解决此类问题, 程浙安 (2019) 以 SA 注意力机制作为基础改进了 Faster RCNN 动物识别模型, 注意力机制能够模拟人类视觉, 有选择性地关注较为重要的信息, 改进后的识别模型平均精确率值达到 92.2%, 结果表明注意力机制能够有效地提取图像特征, 从而提高识别精度, 但更深的网络结构也增加了训练与检测时间。李安琪 (2020) 结合 SE-ResNeXt 模型用于坦桑尼亚塞伦盖蒂国家公园公开图像识别, 郭子豪等 (2023) 为解决节肢动物识别率不高的问题, 设计了有监督的并行空间与通道注意力机制, 该方法在 ArTaxOr 数据集上测试平均精度均值 (mean average precision, mAP) 达到 0.717, 在小目标中检测良好, 在红外图像数据是否保持优势仍需进一步验证。上述文章

表明, 结合注意力机制的深度学习模型具有较好的识别效果, 但模型结构的加深要求计算机的算力较高, 因此, 需要一种算法兼顾识别精度与检测速度, 实现对野外环境中目标的实时跟踪。

针对以上问题, 本研究提出基于迁移学习与 CA-YOLO_v5m 结合的东北虎与东北豹目标识别方法, 该方法适用于复杂自然生态环境下野生动物的识别任务。本研究设计在昼夜背景下算法的识别实验, 并与其他深度学习模型对比, 验证算法的优势性, 实现了东北虎与东北豹的监测, 同时为解决图像存在目标局部特征不明显与红外相机实时监测的问题提供了新的思路。

1 材料与方法

1.1 研究区域概况

东北虎豹国家公园 ($42^{\circ}38'45'' \sim 44^{\circ}18'36''$ N, $129^{\circ}05'01'' \sim 131^{\circ}18'52''$ E) 是我国境内野生东北虎与东北豹的唯一共同分布区 (蒋亚芳等 2021), 地处中国吉林、黑龙江两省交界的老爷岭南部区域 (图 1), 东起吉林省珲春林业局青龙台林场, 西至吉林省大興沟林业局岭东林场, 南自吉林省珲春林业局敬信林场, 北到黑龙江省东京城林业局三道林场, 总面积为 140.65 万 hm^2 。东北虎豹国家公园以中低山、峡谷和丘陵地貌为主, 森林覆盖率达到 93.32%, 良好的自然环境不仅满足了不同类型食草动物的需求, 也为食肉动物提供了丰富的食物来源, 形成了完整的食物链和丰富的生物多样性, 得天独厚的生存环境使得东北虎豹等野生动物得以稳定栖息和繁衍。东北虎豹国家公园的建立对这两种大型濒危猫科动物的繁育与保护具有十分重要的意义。

1.2 数据采集与处理

本研究通过在研究区内布设的红外相机获取真实的研究区内动物视频影像, 实现多角度与多方位的物种监测。真实的原始数据影像能够反馈出野生动物在东北虎豹国家公园内生存

的真实情况。本研究获取了 2013 至 2023 年 10 年间东北虎、东北豹、野猪 (*Sus scrofa*)、梅花鹿 (*Cervus nippon*) 和孢子 (*Capreolus pygargus*) 等丰富的陆生脊椎动物视频资源。

红外相机捕捉的视频图像数据存在多种问题：模糊、遮挡、白天或夜间、背景变化、空图像等。图像背景信息的复杂性与数据预处理结果如图 2 所示。

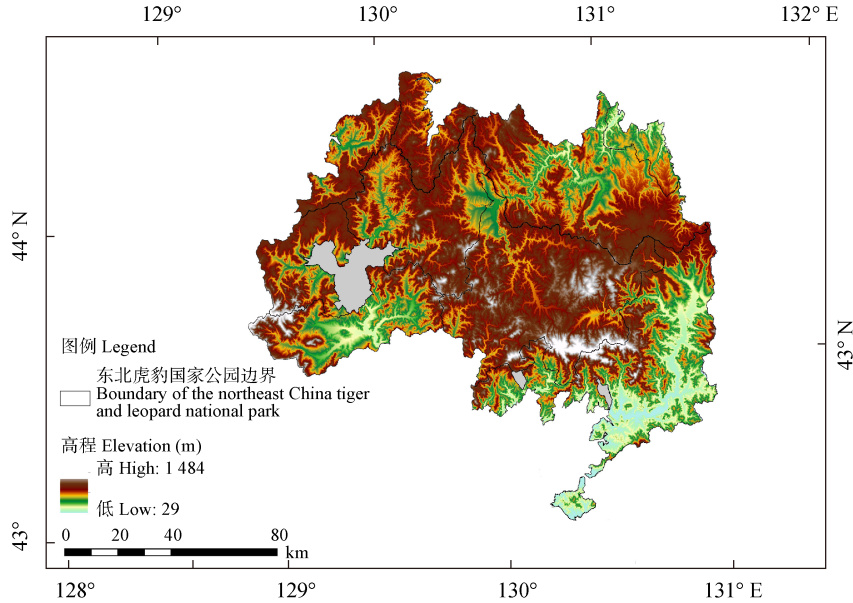


图 1 研究区位置

Fig. 1 Location of the study area

底图审图号 GS (2024) 0650 号, 底图无修改。

The approval number of the base map is GS (2024) 0650, and the base map remains unchanged.

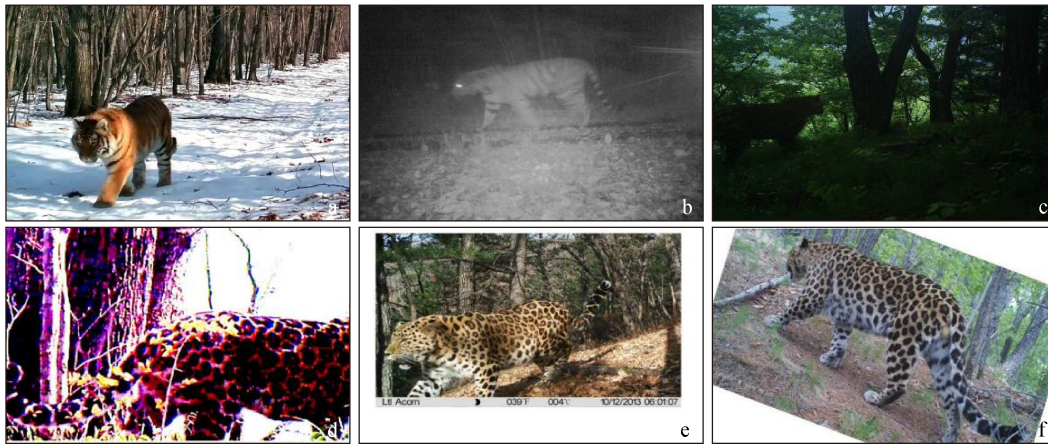


图 2 图像背景信息的复杂性及数据预处理结果

Fig. 2 Complexity of the image background information and data pre-processing results

a. 理想状态下拍摄动物图像; b. 夜间拍摄的模糊图像; c. 拍摄的背景复杂且光线较差的图像; d. 图像增强颜色抖动处理后图像; e. 尺寸变换处理后图像; f. 随机翻转处理后图像。

a. An animal image taken under ideal conditions; b. A blurred image taken at night; c. An image with a complex background and poor lighting conditions; d. The image after image enhancement color dithering; e. The image after size transformation; f. The image after random flipping.

本研究所采用的东北虎、东北豹数据包含部分高清的局部特征图像，有利于模型学习东北虎与东北豹的关键外形特征。本研究选取其中 800 个具有明显动物特征的视频作为原始文件，基于图像和视频处理工具箱实现数据的关键帧提取，将文件后缀统一为 MP4 格式。遍历文件所在位置依次获取每个视频文件的路径，记录文件名称并按照 25 帧为间隔抽取图片，记录文件并进行相似度对比，最终以“文件名 + 序号 + .jpg”构造完整文件名，完成红外相机采集视频流文件转换为原始图像文件。

为了防止分类模型出现过拟合的情况，采用增强图像的方法增加样本数量（张毓等 2021），常用的增强图像方法包括添加脉冲噪声、尺寸变换、伸缩旋转、网格分割和色彩抖动等。经过初步筛选，剔除大量重复误触发与效果不佳图像，最终获得 11 020 张动物图像。通过上述数据扩充方法构建出的数据集图像数量见表 1。

2 深度学习模型建立

2.1 模型框架

目前常用的深度学习分为以 YOLO 系列为代表的 One-stage 检测算法（赵凯琳等 2021）和以 Faster RCNN 为代表的 Two-stage 检测算法（马光凯等 2022）。Faster RCNN 可高效地

识别物体的类别，其骨干网络为 VGG-16 网络，通过加深网络模型来提高检测精度。但当网络加深到一定程度时，在梯度反向传播的过程中会产生连乘效应，出现梯度消失或梯度爆炸的情况，模型的性能会急速下降。虽然使用 ResNet 结构替换原始骨干网络可以解决这个问题，但更深的模型结构仍延长了训练时长与速度。

由于 YOLO 系列算法在训练速度与检测效率上都占有优势，可以同时处理更多的图像数据，且程序内存较小，因此受到了广泛的关注（于文华等 2023）。YOLO_v5 根据模型复杂度由低到高分 s、m、l、x 四种不同大小的模型，其权重文件大小分别是 14、42、93 和 170 MB，随着模型规模的不断加深，检测精度随之提升，但资源占用显著增加，检测速度也随之下降。模型框架可以分为四个部分：输入端（input）、骨干特征层（backbone）、预测层（neck）和回归层（head）。模型采用 mosaic 数据增强技术，大大提高了其对复杂背景物体的检测能力。骨干特征网络中使用了 CSP 结构及 focus 模块，能够提取到更精细的特征并且不用担心网络退化。预测层（neck）使用 FPN + PAN 结构加强网络的特征融合能力。输出端使用 GIoU_Loss 作为 bounding box 损失函数，克服了传统 IoU 存在的缺陷。

表 1 野生动物数据集分布

Table 1 Distribution of the wildlife dataset

数据集 Dataset	训练集 Training set		测试集 Test set		合计 Total
	昼间 Day	夜间 Night	昼间 Day	夜间 Night	
东北虎 <i>Panthera tigris altaica</i>	1 280	960	320	240	2 800
东北豹 <i>P. pardus orientalis</i>	1 280	960	320	240	2 800
野猪 <i>Sus scrofa</i>	1 280	704	320	176	2 480
梅花鹿 <i>Cervus nippon</i>	672	512	168	128	1 480
狍子 <i>Capreolus pygargus</i>	576	592	144	148	1 460
合计 Total	5 088	3 728	1 272	932	11 020

2.2 注意力机制

注意力机制 (attention mechanism) 是一种模仿人类视觉与认知系统的方法, 引入注意力机制的深度学习模型可以有选择性地关注输入图像的重要信息, 提高模型的性能及泛化能力。在深度学习中, 常见的注意力机制实现方式有 SENet、CBAM、ECA 和 CA 等。

SENet (Hu et al. 2020) 是一种典型的通道注意力机制。SENet 关注图像每一个通道的权值, 对于输入的特征层先进行全局平均池化后, 再进行两次全连接, 能够获取输入特征层每一个通道的权值。由于 SENet 结构简单, 不需要引入新的函数层, 使模型在项目部署与复杂度上具有良好的特性。CBAM 将通道注意力机制和空间注意力机制进行结合, 对输入的特征层分别进行全局平均池化与全局最大池化, 并且利用共享全连接层处理获取激活函数 (付江涛 2022)。Wang 等 (2019) 优化 SENet 构建出一种高效通道注意力模块 ECANet, 提出了一种不降维的局部跨通道交互策略。ECANet (Wang et al. 2019) 去除了原来 SE 模块中的全连接层, 直接在全局平均池化之后的特征上通过一个 1D 卷积进行学习。为了避免通过交叉验证手动调整 k 参数, 使用一种自适应方法确定 k 参数。CA 注意力机制则总结了 SENet 与 CBAM 的优势 (Hou et al. 2021), 在通道注意力的基础上兼顾其位置关系, 将通道注意力与空间注意力联合起来, 在保证模型灵活与轻量的前提下提升了视觉任务的性能。

目标检测模型包含 Backbone、Neck 和 Head 三个组件, 分别作用于特征提取、特征融合、特征分类和回归预测, 研究者们习惯于将注意力机制添加到 Backbone 中, 从而提取到更好的特征。本研究考虑迁移算法能够发挥更好的优势, 如果在主干特征提取层放置注意力机制会导致网络的预训练权重无法使用, 因此将注意力机制应用加强特征提取网络上。如图 3 所示, 在主干网络提取出来的两个有效特征层上采样后增加注意力机制。CSP 模块包含多个

卷积层和残差连接, 本研究在 CSP_1-3 结构上添加第一次注意力; SSPF 模块是一个空间金字塔池化模块, 通过对输入特征图进行不同尺度的池化操作, 可以获取到更丰富的特征信息, 本研究在 SSPF 模块内添加第二次注意力; CONCAT 层用于融合来自不同来源的特征图, 在该位置添加第三次注意力可以进一步突出关键特征。三次注意力机制模型能够更准确地关注到图像中的关键信息, 从而提高模型的预测准确率和性能。

2.3 模型训练

本研究使用 Python 3.8 作为开发语言, 选用简洁高效的开源深度学习框架 Pytorch 1.10 进行实验, CUDA 版本为 11.0。本研究的硬件平台为 Intel(R) Xeon(R) CPU E5-2696 v3, 一块 NVIDIA GeForce RTX 4060 Ti 工作站, 运行内存为 64 GB, 研究所采用的软件为 windows 11 操作系统。

在训练目标检测网络的超参数设置上, 本研究设置批量大小 (batch size) 为 8, 选择的学习率 (lr) 是 0.001, 置信度 (confidence) 为 0.001, 门限值取 0.75, 当训练世代 (epoch) 为 100 次, 损失函数趋向于稳定且模型收敛, 采用 SGD 优化器对参数训练过程进行优化。模型训练时, 训练集与测试集比为 8:2, 并且引入迁移学习的策略。

2.4 评价指标

本实验以前述 5 物种平均精度的平均值 (mean average precision) 评价检测任务的精度, 以 F1 分数 (F1-score) 来衡量二分类模型精确度, 用召回率 (recall) 表示标注目标被检测出的比例。精确度 (precision) 即判断为正类且真实类别也为正类的样本数量占模型判断为正类的样本数量的比例。计算公式如下。

$$P_{mA} = \frac{\sum_{c=1}^C P_A(c)}{C}, \quad F_1 = 2 \times \frac{T_p}{(2 \times T_p + F_p + F_N)},$$

$$R = \frac{T_p}{T_p + F_N}, \quad P = \frac{T_p}{T_p + F_p},$$

式中, P_{mA} 为物种平均精度的平均值, P_A 为物种检测的平均准确

率 (average precision), C 为物种总数量; F_1 为 $F1$ 分数, P_T 为被模型预测为正的正样本, P_F 为被模型预测为正的负样本, F_N 为被模型预测为负的正样本; R 为召回率, P 为精确度。

3 结果

3.1 基于深度学习模型的识别结果

为验证复杂背景下 YOLO_v5m 算法与 Faster RCNN 识别算法的适用性, 本研究设计两组对比实验分析识别精度, 数据集选取东北虎与东北豹图像测试, 测试结果见表 2。

表 2 YOLO 与 Faster RCNN 算法性能对比

Table 2 Comparison of the performance between YOLO and Faster RCNN

	F1 分数 F1-score	平均精度均值 (%) Mean average precision, mAP	每秒传输帧数 Frames per second, FPS
YOLO_v5m	0.89	88.46	33.59
Faster RCNN	0.88	91.15	12.93

通过性能对比实验可知, Faster RCNN 算法在识别精度上优于 YOLO_v5 目标检测算法, 东北虎与东北豹的 mAP 值提升了 3%, 但每秒帧数 (frames per second, FPS) 较低。Faster

RCNN 目标检测算法具有高准确率的优势, 并基于此优势一直处于目标检测领域的核心位置。但由于其模型结构较深, 超参数设置较多, 检测速率较低, 无法适应不断提高的实时性要求。YOLO_v5 的实质是将检测问题转化为回归问题, 能够快速获得目标类别和位置信息, 因此具有较高的检测速度, 可达到实时检测。

本研究继续以 Faster RCNN 目标检测算法训练 5 类动物数据集, 检测结果见图 3。东北虎与东北豹的 mAP 值分别为 93.75% 和 88.55%。由于野猪、梅花鹿、狍子图像数据均为昼夜数据集, 且 3 类物种图像总数占总图像的 49%, 训练样本数量进一步影响了图像的识别精度。

3.2 昼夜条件下模型的检测效果

本研究针对昼夜条件下东北虎与东北豹的识别效果进行了三组对比实验。第一组实验仅对夜间灰度图像进行训练, 第二组实验将昼间与夜间图像放进同一训练批次进行训练, 第三组实验采用第二组实验的预训练权重进行 Fine-tune 训练图像, 减轻模型训练所需要的时间, 三组实验均以 YOLO_v5m 模型为基础架构。

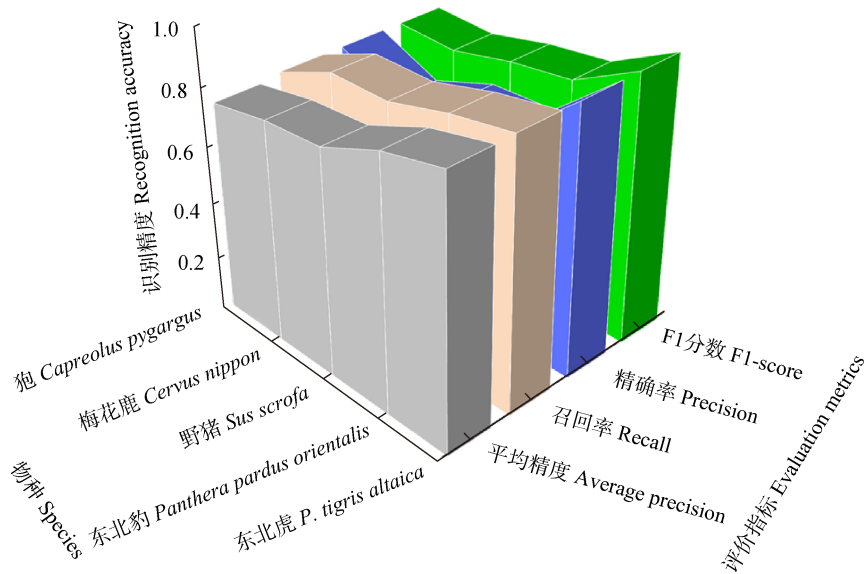


图 3 Faster RCNN 不同种类动物指标对比

Fig. 3 Faster RCNN Comparison of the indicators of different animal species

由于夜间图像较少，为满足迭代次数及梯度下降，3 组实验采用不同的迭代次数。随着训练世代的增加，总损失值先快速下降后逐渐趋向稳定。不同数据集下变化趋势基本一致，说明数据集的差异对模型总损失值的影响不大。训练世代为 100 左右时，变化曲线由迅速下降走向平稳，此时各模型均已满足收敛的条件（表 3）。

当模型停止训练且训练完毕后，Loss 均已达到了模型收敛的条件，因此可以认为模型是有效的。根据 3 组实验对比结果可知：基于 YOLO_v5m 模型对夜间自然状态下的东北虎与东北豹识别 mAP 值为 83.15%，昼夜间的 mAP

值为 86.67%，在小样本识别时效果较差。第三组实验结果表明，迁移学习机制在小样本数据集上能进一步提高识别准确度，其 mAP 值为 91.16%，模型收敛较好，且训练时长缩短了 106 min。在识别结果中，模型识别夜间图像的准确率要低于日间图像，夜间图像为灰度图像，图像质量较低，且东北虎与东北豹的纹理特征不明显，背景干扰大，这些因素综合导致夜间图像识别率较低。实验三识别结果如图 4 所示。

3.3 融入注意力机制的 YOLO_v5m 对比分析

在数据集上对比融入不同注意力机制的目标检测模型，验证融入注意力机制的深度学习模型方法的普适性。

表 3 三组实验测试结果

Table 3 Test results of the three experimental groups

实验 Test	目标类别 Target species	F1 分数 F1-score	召回率 (%) Recall	精确度 (%) Precision	平均精度 (%) Average precision	训练时长 (h) Training length
1	东北虎 <i>Panthera tigris altaica</i>	0.80	86.07	75.00	83.49	3.35
	东北豹 <i>P. pardus orientalis</i>	0.80	85.11	75.47	82.80	
2	东北虎 <i>P. t. altaica</i>	0.86	90.00	83.08	87.43	4.27
	东北豹 <i>P. pardus orientalis</i>	0.85	88.33	81.54	85.90	
3	东北虎 <i>P. t. altaica</i>	0.86	92.22	80.19	89.17	2.50
	东北豹 <i>P. pardus orientalis</i>	0.89	93.83	84.92	93.14	

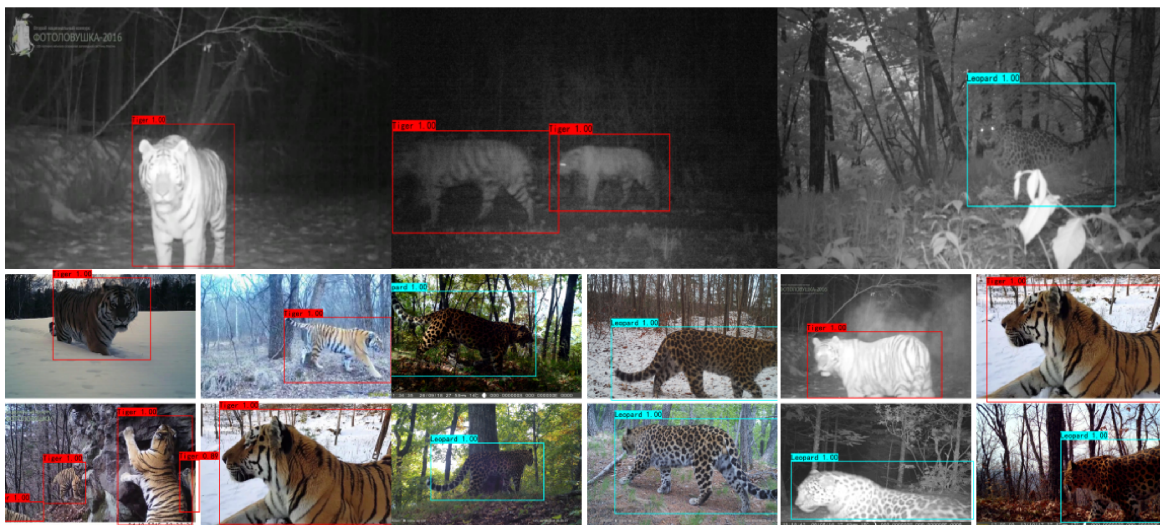


图 4 检验输出正确识别结果

Fig. 4 Example of the correct identification results of the test output

样本数量的不同也会影响最终的识别精度，本次设计的对比实验采用包含 11 020 张昼夜图像的野生动物数据集，并引入迁移学习策略。本研究选择 SE、ECA、CBAM 三类注意力机制与 CA 注意力机制对比，以识别精确度以及 mAP 值作为评价指标，采用 YOLO_v5m 模型框架，为控制不同注意力机制模型的有效对比，实验所采用训练参数、训练世代、引入注意力位置均保持一致。识别结果的对比如图 5 所示。此外，进一步生成了梯度类激活热力图，便于直观查看原始图像上模型的关注区域（图 6），热力图中高亮区域即为注意力模型关注区域。模型可以正确区分复杂昼夜背景下

背景与前景，对于东北虎豹物种类别也有较好的关注度。

与引入 SE、ECA、CBAM 注意力机制相比，融入 CA 注意力机制的 YOLO_v5m 模型精确率分别提高了 1.25%、0.05%和 1.73%，物种平均精度均值分别提高了 1.78%、1.05%和 1.85%（图 5）。在 YOLO_v5m 模型中添加 CA 注意力机制能够提升其对东北虎与东北豹的识别效果。CA-YOLO_v5m 在检测精度上与 Faster RCNN 算法（表 2）相比，东北虎豹的 mAP 值提升了 2.57%。融入 CA 注意力机制模型兼顾了检测速度与识别精度，大大提高了模型对野外红外目标的检测性能。

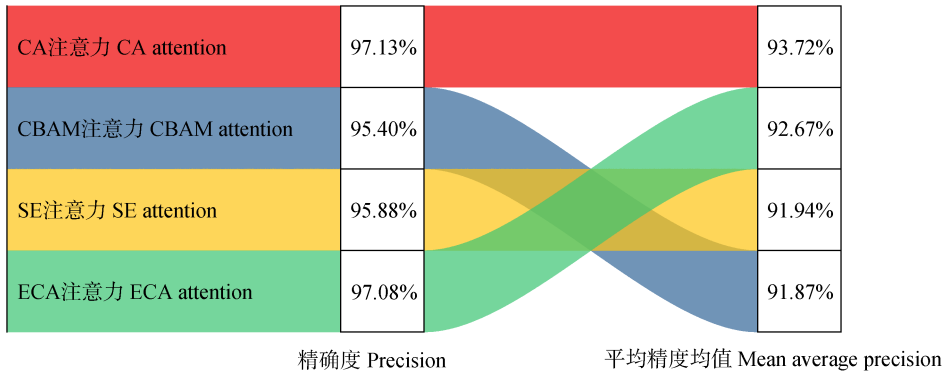


图 5 不同注意力机制评价指标

Fig. 5 Evaluation indicators of different attention mechanisms

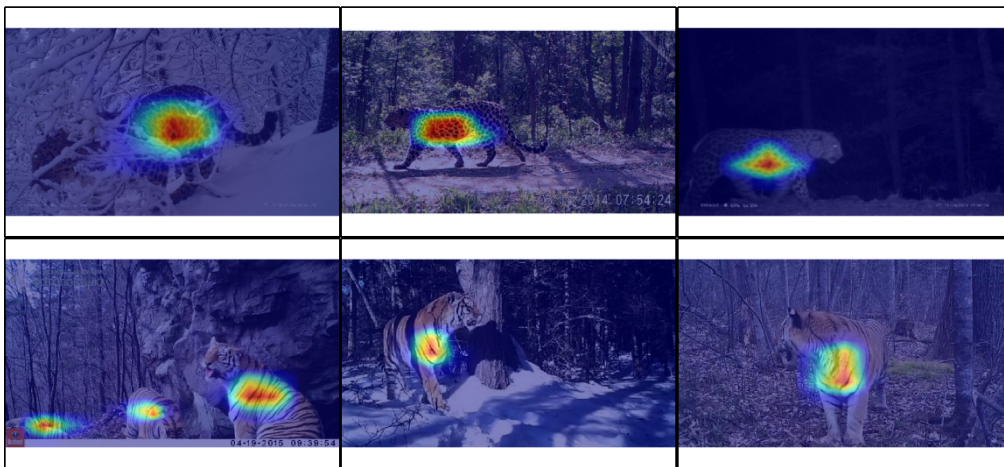


图 6 融入 CA 注意力机制的模型对正样本的识别热力图

Fig. 6 Identification heat map of the model integrating the CA attention mechanism for positive samples

4 结论

东北虎与东北豹物种数量稀少, 且活动轨迹复杂, 利用高效手段识别与监测东北虎豹对物种多样性保护具有重要意义。本研究基于 CA-YOLO_v5m 的深度学习模型实现了对真实背景下的东北虎与东北豹图像识别。

在创建的数据集样本中训练, Faster RCNN 算法网络结构深, 但精度较高, 野生动物的 mAP 值为 91.15%, 而轻量型的 YOLO_v5m 算法 mAP 值为 88.46%, 降低了 2.69%。以东北虎与东北豹单独作为研究对象, 训练后 mAP 值为 86.67%, 引入迁移学习后的 YOLO_v5m 算法 mAP 值为 91.16%, 提升了 4.49%, 训练时长较改进前缩短了 106 min, 迁移算法能够更多地节省训练时间。为得到兼顾检测速度与精度的识别算法, 本研究提出 CA-YOLO_v5m, 并引入迁移学习的策略, 将三次注意力机制添加在 Neck 层, 改进后 YOLO_v5m 算法的 mAP 值为 93.72%, 与 SE、CBAM、ECA 相比, mAP 分别提高了 1.78%、1.05% 和 1.85%, 比原始 YOLO_v5m 模型提高了 7.05%, 实现了更高效的东北虎豹识别。

由于东北虎与东北豹属于濒危动物, 样本数量在模型训练时表现不足, 且模型在识别形态相近物种图像时也存在较多的误检与漏检。增加物种类别或扩充数据后可以获得更好的检测效果, 能够提升模型的适用性与识别率。今后可以针对数据集与模型改进两方面进一步研究, 采用神经网络的方式实现数据增强; 研究更深的网络模型, 优化参数提高模型对东北虎与东北豹的检测效果。

参 考 文 献

- Cai Z W, Vasconcelos N. 2018. Cascade R-CNN: delving into high quality object detection // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 6154–6162.
- Girshick R, Donahue J, Darrell T, et al. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation // 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, 580–587.
- He K M, Zhang X Y, Ren S Q, et al. 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition // Fleet D, Pajdla T, Schiele B, et al. Computer Vision-ECCV 2014. Cham: Springer, 8691: 346–361.
- Hinton G E, Salakhutdinov R R. 2006. Reducing the dimensionality of data with neural networks. *Science*, 313(5786): 504–507.
- Hou Q B, Zhou D Q, Feng J S, et al. 2021. Coordinate Attention for Efficient Mobile Network Design // 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, 13708–13717.
- Hu J, Shen L, Albanie S, et al. 2020. Squeeze-and-excitation networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(8): 2011–2023.
- Krizhevsky A, Sutskever I, Hinton G E. 2012. ImageNet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25(6):1097–1105.
- Norouzzadeh M S, Nguyen A, Kosmala M, et al. 2018. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Science of the United States of America*. 115(25): E5716–E5725.
- Redmon J, Divvala S, Girshick R, et al. 2015. You only look once: unified, real-time object detection. *CoRR*, abs/1506.02640.
- Ren S Q, He K M, Girshick R, et al. 2017. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6): 1137–1149.
- Wang Q, Wu B, Zhu P, et al. 2019. ECA-Net: Efficient channel attention for deep convolutional neural networks. *CoRR*, abs/1910.03151.
- 程浙安. 2019. 基于深度卷积神经网络的内蒙古地区陆生野生动物自动识别. 北京: 北京林业大学硕士学位论文.
- 丰大林, 于晶晶. 2022. 基于红外触发相机技术对猕猴西藏亚种的冬季生境选择研究. *野生动物学报*, 43(3): 629–635.
- 付江涛. 2022. 基于改进的 Faster R-CNN 目标检测研究与应用.

- 银川: 宁夏大学硕士学位论文.
- 宫一男, 谭孟雨, 王震, 等. 2019. 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. *兽类学报*, 39(4): 458–465.
- 郭子豪, 董乐乐, 曲志坚. 2023. 基于改进 Faster RCNN 的节肢动物目标检测方法. *计算机应用*, 43(1): 88–97.
- 韩家臣. 2021. 基于深度学习的野生动物图像识别方法研究. 兰州: 西北师范大学硕士学位论文.
- 何嘉. 2019. 基于深度学习的野生动物智能检测与识别. 深圳: 深圳大学硕士毕业论文.
- 蒋亚芳, 田静, 赵晶博, 等. 2021. 国家公园生态系统完整性的内涵及评价框架: 以东北虎豹国家公园为例. *生物多样性*, 29(10): 1279–1287.
- 柯澳, 王宇聪, 胡博宇, 等. 2024. 基于图像的野生动物检测与识别综述. *计算机系统应用*, 33(1): 22–36.
- 李安琪. 2020. 基于卷积神经网络的野生动物监测图像自动识别方法研究. 北京: 北京林业大学硕士学位论文.
- 李仁杰. 2018. 灵长类动物面部数据集构建及识别方法研究. 西安: 西北大学硕士学位论文.
- 林典敏, 吴可廷. 2023. 基于 YOLOv5 算法的小动物识别技术研究. *信息技术与信息化*, (8): 194–198.
- 马光凯, 吴伟, 刘丹, 等. 2022. 圈养东北虎幼崽群体活动的自动跟踪方法研究. *野生动物学报*, 43(3): 605–613.
- 孙蕊, 张旭, 郭颖, 等. 2020. 基于 Faster R-CNN 金丝猴优化检测方法. *激光与光电子学进展*, 57(12): 259–268.
- 王劭文, 王东亮, 凌成星, 等. 2023. 基于无人机遥感的黑龙江南瓮河国家级自然保护区驼鹿种群数量及空间分布调查. *野生动物学报*, 44(3): 486–493.
- 阎婧宇, 谢永华. 2024. 基于 YOLOv7-Tiny 算法的无人机实时跟踪野生动物方法. *野生动物学报*, 45(2): 251–261.
- 杨铭伦, 张旭, 郭颖, 等. 2022. 基于 YOLOv5 的红外相机野生动物图像识别. *激光与光电子学进展*, 59(12): 372–380.
- 于文华, 谢永华. 2023. 基于轻量化 YOLOX 算法的东北虎检测. *野生动物学报*, 44(4): 760–767.
- 张毓, 高雅月, 常峰源, 等. 2021. 小样本条件下基于数据扩充和 ResNeSt 的雪豹识别. *北京林业大学学报*, 43(10): 89–99.
- 赵凯琳, 靳小龙, 王元卓. 2021. 小样本学习研究综述. *软件学报*, 32(2): 349–369.