

基于卷积神经网络的无人机遥感影像水鸟自动识别

林海^① 高大中^① 张童^② 崔国发^{①*}

① 北京林业大学生态与自然保护学院 北京 100083; ② 国家林业和草原局华东调查规划院 杭州 310019

摘要: 水鸟监测是了解水鸟种群和分布动态、开展水鸟和湿地保护的基础,但该活动耗时耗力。近年来,随着无人机遥感技术的发展,使用小型无人机获得高分辨率的水鸟遥感影像已经成为可能;与此同时,卷积神经网络提供了一种快速识别无人机遥感图像中的鸟类的方法。我们尝试结合两种技术,使用卷积神经网络 Mask R-CNN 与 YOLOv3 识别湖南西洞庭湖国家级自然保护区无人机遥感影像中的大型水鸟,取得了良好的效果:模型检测拍摄到的鸭属鸟类,包括绿翅鸭 (*Anas crecca*) 和罗纹鸭 (*A. falcata*) 的结果平均精度达到 0.93,精度达到 90.83%,召回率达到 93%;检测小天鹅 (*Cygnus columbianus*) 的结果平均精度达到 0.91,精度达到 84.38%,召回率达到 84.00%。结果表明,将无人机遥感技术与卷积神经网络结合,可以快速统计水鸟数量,在种群监测工作中具有应用潜力。

关键词: 水鸟; 无人机; 种群监测; 卷积神经网络

中图分类号: Q958 **文献标识码:** A **文章编号:** 0250-3263 (2024) 03-450-10

Waterbirds Auto-Detection in Unmanned Aerial Vehicle Remote Sensing Image Base on Convolutional Neural Network

LIN Hai^① GAO Da-Zhong^① ZHANG Tong^② CUI Guo-Fa^{①*}

① School of Ecology and Nature Conservation, Beijing Forestry University, Beijing 100083;

② East China Forest Inventory and Planning Institute, Hangzhou 310019, China

Abstract: [Objectives] Waterbird monitoring plays a crucial role in understanding population dynamics and guiding conservation efforts, but it has traditionally been a time-consuming process. In this study, our objective is to integrate unmanned aerial vehicle (UAV) remote sensing with convolutional neural networks (CNN) to achieve rapid and accurate estimation of waterbird populations. **[Methods]** We employed the DJI Mavic 2 Zoom UAV to capture high-resolution remote sensing images in the West Dongting Lake National Nature Reserve in Hunan. The UAV was flown at an altitude of 75 m, with its camera positioned in a vertically downward-facing orientation. We obtained images with a ground resolution of 1.2 cm/pixel, Table 1

基金项目 国家自然科学基金项目 (No. 32171545);

* 通讯作者, E-mail: fa6716@163.com;

第一作者简介 林海, 男, 硕士研究生; 研究方向: 自然保护区学; E-mail: sclh45@bjfu.edu.cn.

收稿日期: 2023-06-14, 修回日期: 2023-11-27 DOI: 10.13859/j.cjz.202423115

displays the waterbirds captured in the images. We selected 503 images to construct a dataset, including two categories: *Anas crecca/A. falcata* and *Cygnus columbianus*, with 3 778 and 395 samples respectively. The dataset has several training sets of different sizes (Table 2) and a validation set of 3 032 samples. For each training set, we independently developed Mask R-CNN and YOLOv3 models, evaluating their performance using the validation set. Evaluation metrics include average precision, recall, precision, and *F1*-score. **[Results]** When identifying *A. crecca/A. falcata*, Mask R-CNN model achieved a recall rate of 93.00% and a precision of 90.83% (Table 4, Fig. 4), while the YOLOv3 model achieved a recall rate of 93.00% and a precision of 88.79% (Table 5, Fig. 5). After reaching 178 ind for *A. crecca/A. falcata* in the training set, further augmentation did not result in a significant improvement in the performance of both models. When identifying *C. columbianus*, the performance of both models improved with an increase in the size of the training set. The Mask R-CNN model achieved a recall rate of 84.00% and a precision of 84.38% (Table 6, Fig. 6), while the YOLOv3 model achieved a recall rate of 90.00% and a precision of 81.69% (Table 7, Fig. 7). The Mask R-CNN model detected images at a speed of approximately 12 images/s, while the YOLOv3 model detected images at a speed of 20 - 30 images/s. **[Conclusion]** Our study proposes a potential solution for efficient and accurate waterbird population monitoring in natural habitats. Our models demonstrated high accuracy in identifying *A. crecca/A. falcata*, the recognition accuracy difference between Mask R-CNN and YOLO was minimal. Remarkably, by integrating UAV remote sensing with CNN, our approach demonstrates the potential for training highly efficient and accurate waterbird identification models with minimal annotated data—perhaps requiring fewer than 250 ind per waterbird species, as suggested by our results.

Key words: Waterbird; Unmanned aerial vehicle; Population monitoring; Convolutional neural network

水鸟监测工作中, 传统鸟类多样性监测多依赖于人为定点观测及样线调查, 对于活动范围较大或夜间活动的鸟类等监测局限性较大, 且耗时耗力, 计数容易产生主观偏差 (雷倩等 2018), 且存在部分区域难以抵达、对部分集群鸟类调查效率低的问题。卫星遥感影像分辨率不足, 难以识别鸟类, 使用有人驾驶飞行器进行监测成本高昂又有着不小的安全隐患。近年来, 无人机凭借低成本、高分辨率、灵活方便及较有人飞行器安全等优点, 逐渐成为生态学研究的重要手段之一 (李杰等 2019)。随着无人机遥感图像分辨率的不断提高, 该技术逐渐被运用于鸟类种群监测中 (Díaz-Delgado et al. 2017, Korcak-Abshire et al. 2019, Shafer et al. 2019)。使用无人机可以大大提高种群监测工作的效率, 同时无人机可前往人类难以到达的地区, 获得凭地面调查难以获得的数据。

目前, 无人机遥感图像水鸟识别依赖人工

解译。为降低水鸟监测工作的人力和时间成本, 加快图像解译的速度并实现自动解译, 我们尝试使用卷积神经网络 (convolutional neural network, CNN) 模型掩膜区域卷积神经网络 (mask region-based convolutional neural network, Mask R-CNN) (He et al. 2017) 与 “只需看一次 3” 模型 (you only look once version 3, YOLOv3) (Redmon et al. 2018), 探索一种自动识别的方法。

2012 年以来, 卷积神经网络取得了极大的发展 (Krizhevsky et al. 2012)。卷积神经网络模仿动物视觉皮层设计, 是深度学习的代表模型之一 (Gu et al. 2018), 在计算机视觉和自然语言处理等领域获得了广泛应用。卷积神经网络可以提取图像的特征, 并通过重复训练学习如何通过提取出的特征识别图像中的目标。

卷积神经网络精度远较传统目标识别系统高, 这已被 ImageNet 等计算机视觉领域的竞赛

结果证明。此外,训练卷积神经网络不需要在每一个识别任务中都手动设计和调整复杂的特征识别算法,需要的计算机视觉、数学等专业知识远较传统目标识别系统少,更利于推广应用。目前,已有一些将卷积神经网络运用于动物野外监测工作的尝试,如使用该技术分辨不同鸟类的鸣叫(Kahl et al. 2017),自动识别东北虎(*Panthera tigris* ssp. *amoyensis*) (史春妹等 2021)、川金丝猴(*Rhinopithecus roxellanae*) (胡旭 2019)、大熊猫(*Ailuropoda melanoleuca*) (Hou et al. 2020) 个体,均取得了良好的效果。也有克服训练样本少(张毓等 2021)和类别不平衡(蔡前舟 2022)等野外动物监测场景中常见问题的尝试。目前在动物野外监测工作中应用卷积神经网络的研究大多使用地面水平拍摄的数据,特别是红外相机拍摄的数据,但也有一些研究采用了从高处向下采集的数据,如运用固定架设相机采集的高原鼠兔(*Ochotona curzoniae*) 图像进行图像检测与分割的方法研究(陈刚琦 2020)。

使用卷积神经网络识别无人机遥感影像中野生动物的尝试于 2017 年开始,卷积神经网络被用于识别纳米比亚 Kuzikus 保护区中的大型野生动物,这两项技术的结合被认为可以实现对大型野生动物的实时监测(Kellenberger et al. 2017)。随后,更多结合这两项技术的研究出现。在鸟类监测方面,采用卷积神经网络模型 Mask R-CNN 识别无人机红外遥感影像中的野火鸡(*Meleagris gallopavo*) (Kassim et al. 2020) 和采用卷积神经网络模型 YOLO 识别鸟巢(Li et al. 2022) 均取得了良好的效果。目前结合卷积神经网络与无人机遥感两项技术进行动物识别的研究大多集中在大型野生动物识别,而有关鸟类识别的研究较少。

本研究使用我们先前在湖南西洞庭湖国家级自然保护区运用小型无人机获取的遥感影像(高大中等 2021) 训练 Mask R-CNN 模型与 YOLOv3 模型,尝试使用卷积神经网络自动识别无人机遥感影像中的水鸟。

1 材料与方法

1.1 图像采集

本研究区域位于湖南西洞庭湖国家级自然保护区内。该保护区位于湖南省汉寿县境内,被列入国际重要湿地名录,以水鸟、水鸟栖息地以及湿地生态系统为主要保护对象。

通过前期踏查和无人机试飞,我们选取了水鸟集中分布的四个固定调查点作为本次研究的遥感图像采集区域,分别位于保护区内半边湖、龙王湖、青山湖和安乐湖地区(图 1)。采集区域内人为干扰弱,少有遮蔽物,地表主要为浅水和滩涂地形。

本研究选用大疆“御” Mavic 2 Zoom 无人机。其主要参数为:对角线轴距 354 mm,续航时间 31 min,传感器尺寸 1/2.3 in CMOS,图像分辨率 4 000 像素 × 3 000 像素,镜头等效焦距 24 ~ 48 mm。

本研究的遥感图像均在 10 ~ 16 时之间拍摄,拍摄时天气晴朗,能见度高,无人机相机倾角为 90°,航高 75 m。感光度、快门速度、白平衡、焦距等参数由无人机控制。拍摄同时通过地面观测记录被拍摄鸟类的类别。图像的地面分辨率为 1.2 cm/像素(图 2)。

1.2 数据集概况

将前期研究获得的遥感影像裁切成多张 1 000 × 750 像素的图片,并从裁切后的图片中选取了 503 张图片进行标注,标注采用 VGG Image Annotator 软件。共计拍摄和标注了 4 312 只鸟,标注框包含目标鸟类与鸟类的阴影,包括 3 778 只绿翅鸭(*Anas crecca*) 和罗纹鸭(*A. falcata*)、395 只小天鹅(*Cygnus columbianus*)、45 只大白鹭(*Ardea alba*)、29 只苍鹭(*A. cinerea*) 及 65 只凤头麦鸡(*Vanellus vanellus*) (表 1)。

参考地面观测记录,无法凭肉眼区分图像中的绿翅鸭和罗纹鸭,但可将二者与其他种类区分开,其他种类之间的区分没有困难。因此在本研究中将绿翅鸭和罗纹鸭作为一类进行处



图 1 样地位置示意图

Fig. 1 Location diagram of sample area



图 2 无人机遥感影像与手动标注示例

Fig. 2 Example of remote sensing images with manual annotation of one bird

理。由于包含大白鹭、苍鹭或凤头麦鸡的图片数量过少，难以分割数据集进行模型训练与验证，本次研究以小天鹅和绿翅鸭/罗纹鸭作为主要研究对象（表 2）。

绿翅鸭/罗纹鸭体棕黑色，颈短，体型呈纺锤形，图片中长度约 33 像素。小天鹅体白色，嘴黄色，体型呈纺锤形，图片中长度约 92 像素。

通常情况下，提高训练数据量可以改善神经网络模型的识别准确度，但这会导致标注工作量的增加。为探索训练数据量与识别准确度的关系，将数据集随机分为两部分，第一部分包含

表 1 遥感图像中的各种水鸟

Table 1 Birds in remote sensing images

物种 Species	大小 (± 15 cm) Size (± 15 cm)	示例图 Picture
大白鹭 <i>Ardea alba</i>	90	
苍鹭 <i>A. cinerea</i>	90	
小天鹅 <i>Cygnus columbianus</i>	110	
凤头麦鸡 <i>Vanellus vanellus</i>	32	
绿翅鸭/罗纹鸭 <i>Anas crecca/A. falcata</i>	40	

表 2 各训练集中小天鹅和绿翅鸭/罗纹鸭数量 (只)

Table 2 Number of *Cygnus columbianus* and *Anas crecca/A. falcata* in each training dataset (ind)

训练集大小 Size of training dataset	绿翅鸭/罗纹鸭 <i>Anas crecca/A. falcata</i>	小天鹅 <i>Cygnus columbianus</i>
22	10	12
34	16	18
59	37	42
180	99	81
339	178	161
373	249	124
610	375	235

1 280 个样本, 第二部分包含 3 032 个样本。从第一部分随机抽取图片, 建立不同大小的训练集用于训练神经网络模型, 并使用第二部分作为验证集, 比较训练集大小不同时的识别准确率。

1.3 卷积神经网络模型

本次研究采用 Mask R-CNN 与 YOLOv3 两种模型进行鸟类识别。Mask R-CNN 是 2017 年提出的一种卷积神经网络模型, 可用于目标识

别、目标分类和语义分割等任务。Mask R-CNN 主要由以下三部分组成。

第一部分将整张图片输入特征提取网络 (feature proposal network, FPN) 生成整张图片的特征图 (feature map), 将特征图交由区域生成网络 (region proposal network, RPN), 区域生成网络根据特征图包含的图片特征生成候选框, 同时对候选框的位置和尺寸进行修正和筛选, 生成候选区域 (region proposal)。这些候选区域对应着特征图中最可能出现目标的区域。

第二部分输入前一步生成的候选区域, 由于候选区域尺寸不同, 为统一表示这些候选区域的图像特征, 使用 ROI Align 方法对各个候选区域提取相同大小的特征图。

第三部分则是根据第二步提取出的特征对每个候选区域进行分类, 同时进行边界回归, 修正输出的候选框的大小, 完成目标识别和分类的工作。

与 Mask R-CNN 相比, YOLOv3 不使用区域生成网络生成候选区域, 直接将整张图片划分为大小相同的数个部分, 通过特征提取网络同时预测物体分类和位置, 这大大简化了神经网络模型。在 ImageNet 等计算机视觉领域的竞赛中, YOLOv3 相比 Mask R-CNN 表现出了更快的检测速度, 但精度稍低。

本次研究使用 MMDetection 2.28.1 目标检测框架 (Chen et al. 2019)。Mask R-CNN 的特征提取网络选用 ResNet-101, YOLOv3 的特征提取网络选用 DarkNet-53, 训练批次大小均为 9, 训练步数 2 000。训练环境信息见表 3。

1.4 识别结果评价方法

本研究选用人工标注后的鸟类数据作为评价数据。训练完成的神经网络模型会对每一张图片给出一系列识别结果, 每一个识别结果由三部分构成 (图 3), 分别为候选框 (识别结果的轮廓, 理想状态下应为鸟类轮廓)、类别和置信度 (神经网络模型对该识别结果可信度的自我评价)。

表 3 模型训练环境

Table 3 Model training environment

名称 Name	属性 Property
操作系统 Operating system	Ubuntu 20.04 x64
中央处理器 Central processing unit	Intel(R) Xeon(R) Gold 6242R CPU
内存 Random access memory	64 GB
显卡 Graphics card	NVIDIA GeForce RTX 3090
机器学习框架 Machine learning framework	PyTorch 1.10.0

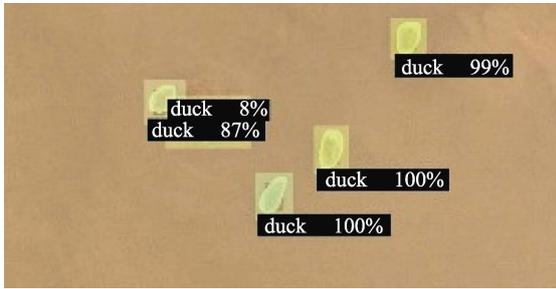


图 3 识别结果示例

Fig. 3 Example of detection results

duck 是该识别结果的分类，对应绿翅鸭/罗纹鸭，其后的百分数为置信度，最大 100%。

“duck” is the category of the recognition result, corresponding to the *Anas crecca/A. falcata*. The following percentage is confidence score, up to 100%.

1.4.1 置信度阈值设置 神经网络模型的识别结果中包括一些置信度不高的识别结果，通常的做法是设置一个置信度阈值，排除所有置信度不高于阈值的识别结果。若该阈值设置太高，则会造成大量漏检，若该阈值过低，则会造成识别结果正确率低下。本研究调整置信度阈值，在不同置信度阈值的条件下分析神经网络模型的识别结果。

1.4.2 识别结果正误判断 我们以神经网络模型给出的候选框与人工标注候选框的重合情况判断神经网络模型预测结果正确与否，以交并比（intersection over union, IoU）作为判断识别结果是否正确的依据。考虑到统计遥感图

像中的鸟类数量并不需要在图像中精确地标出鸟类的边界，本研究设置交并比的阈值为 0.2，即交并比大于等于 0.2 时认为该判断正确，否则认为该判断错误。交并比的计算公式如下：

$$V_{IoU} = \frac{S(C \cap G)}{S(C \cup G)}$$

式中， V_{IoU} 指交并比， C 指神经网络模型识别图像产生的候选框（candidate bound）， G 指测试数据中给出的原标记框（ground truth bound）， S 表示给定区域的面积。

1.4.3 精度和召回率 为比较使用不同训练数据量训练出的神经网络模型之间识别效果的差距，我们使用召回率（recall, R ）和精度（precision, P ）对这些模型的识别结果进行评估。相关公式如下：

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

式中， T_p 为模型预测结果为正确的正例数，即成功识别出的鸟类数量， F_n 表示模型预测结果为错误的正例数，即未被识别出的鸟类数量， F_p 表示模型预测结果为正确的负例数，即识别结果中与实际情况不符的结果数量。

1.4.4 模型总体评价 一个好的模型应同时有高精度和高召回率，但不同置信度阈值情况下，识别结果会有所不同，精度和召回率也会随之改变。为在不同置信度阈值条件下全面评价神经网络模型，我们采用平均精度（average precision, AP）对模型进行评价。相关公式如下： $F_{AP} = \int_0^1 p(R) dR$ ，式中， F_{AP} 指平均精度， R 为召回率， p 为一个函数，指调整置信度阈值，使召回率达到给定值 R 时，对应的精度 P 的值。

为直观地表示精度（ P ）、召回率（ R ）和置信度阈值三者的关系，绘制 P - R 曲线。即给定一组识别结果，对每一个不同的置信度阈值，都有与其对应的精度 P 和召回率 R 。以精度和召回率为坐标轴建立坐标系，在其中标出不同置信度阈值条件下识别结果的精度和召回率所对应的点，将这些点相连，所得曲线即为 P - R 曲线。

使用 F_1 分数（以下简称为 F_1 ）的最大值来表示在置信度阈值最好地兼顾了召回率和精度时，某一模型的表现： $F_1 = \frac{2PR}{P+R}$ ，式中， P 为精度， R 为召回率。

2 结果

2.1 识别绿翅鸭/罗纹鸭

使用 Mask R-CNN 时，当训练集中绿翅鸭/罗纹鸭个体数达到 178 后，识别绿翅鸭/罗纹鸭结果的召回率达到 92.00%，精度达到 91.35%，增加训练量至 375，模型达到了 93.00% 的召回率和 90.83% 的精度。检测图片的速度约为 12 张/s（图 4，表 4）。

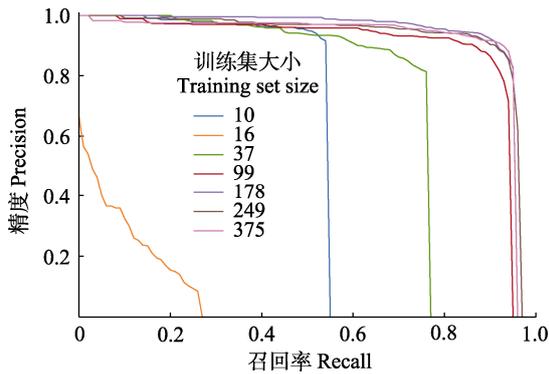


图 4 不同大小训练集所得 Mask R-CNN 模型识别绿翅鸭/罗纹鸭精度-召回率曲线

Fig. 4 Recall-precision (P - R) curves of Mask R-CNN models for identifying *Anas crecca/A. falcata* with different training sets sizes

使用 YOLOv3 时，模型在训练集仅有 10 只绿翅鸭/罗纹鸭时就达到了 82.00% 的召回率及 87.26% 的精度，增加训练量至 375，模型达到了 93.00% 的召回率和 88.79% 的精度，检测图片速度为 20 ~ 30 张/s（图 5，表 5）。

使用 YOLOv3 与 Mask R-CNN 时，训练集中绿翅鸭/罗纹鸭个体数从 99 增大到 375 时，所得神经网络模型的表现均无显著提升，说明对于本次研究所使用数据集，训练集中绿翅鸭/

表 4 不同大小训练集所得 Mask R-CNN 模型识别绿翅鸭/罗纹鸭结果平均精度、 F_1 分数、召回率和精度

Table 4 Average precision, F_1 -score, recall, and precision of Mask R-CNN models for identifying *Anas crecca/A. falcata* detection with different training dataset

训练集中绿翅鸭/罗纹鸭数量 Number of <i>Anas crecca</i> and <i>A. falcata</i> (ind)	平均精度 Average precision	F_1 分数 F_1 -score	召回率 (%) Recall	精度 (%) Precision
10	0.54	0.69	55.00	91.57
16	0.09	0.19	16.00	23.48
37	0.73	0.79	77.00	81.37
99	0.90	0.89	90.00	87.81
178	0.94	0.92	92.00	91.35
249	0.93	0.91	92.00	90.21
375	0.93	0.92	93.00	90.83

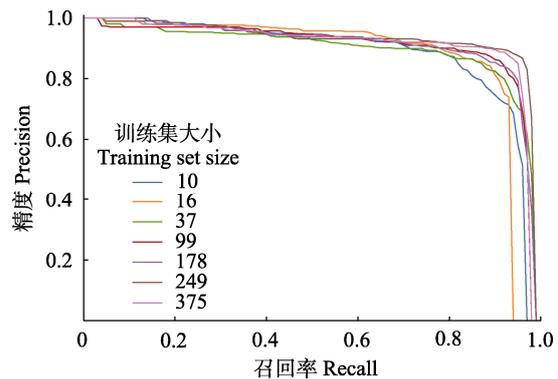


图 5 不同大小训练集所得 YOLOv3 模型识别绿翅鸭/罗纹鸭精度-召回率曲线

Fig. 5 Recall-precision (P - R) curves of YOLOv3 models for identifying *Anas crecca/A. falcata* with different training set sizes

罗纹鸭个体数达到 100 ~ 250 只即可有效识别绿翅鸭/罗纹鸭，即依此法统计开阔水面上规模 4 000 只左右的绿翅鸭/罗纹鸭种群时，只需标注约 3% ~ 6% 的个体即可达到良好的识别效果。

2.2 识别小天鹅

识别小天鹅时，模型表现随着训练集大小增加而改善，在训练集大小达到 235 时，Mask R-CNN 模型达到了 84.00% 的召回率和 84.38%

表 5 不同大小训练集所得 YOLOv3 模型识别绿翅鸭/罗纹鸭结果平均精度、F1 分数、召回率和精度

Table 5 Average precision, F1-score, recall, and precision of YOLOv3 models for identifying *Anas crecca*/*A. falcata* detection with different training dataset

训练集中绿翅鸭/罗纹鸭数量 Number of <i>Anas crecca</i> and <i>A. falcata</i> (ind)	平均精度 Average Precision	F1 分数 F1-score	召回率 (%) Recall	精度 (%) Precision
10	0.89	0.85	82.00	87.26
16	0.89	0.87	90.00	84.07
37	0.90	0.87	91.00	82.49
99	0.91	0.89	92.00	85.96
178	0.92	0.88	94.00	83.28
249	0.93	0.92	96.00	87.62
375	0.92	0.91	93.00	88.79

的精度，检测速度约为 12 张/s 图片（图 6，表 6）；YOLOv3 模型达到了 90.00%的召回率和 81.69%的精度，检测速度为 20 ~ 30 张/s 图片（图 7，表 7）。

3 讨论

本研究以湖南西洞庭湖保护区的 3 778 只绿翅鸭/罗纹鸭与 395 只小天鹅无人机遥感影像数据作为研究对象，首先通过人工标注遥感

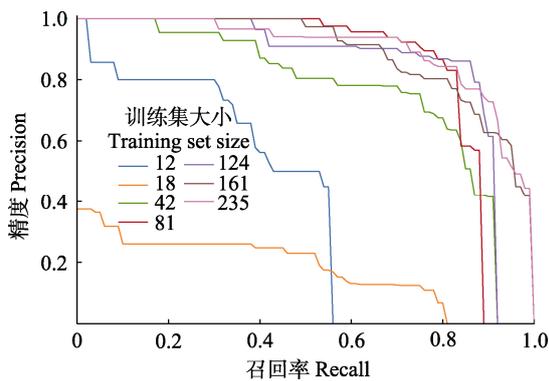


图 6 不同大小训练集所得 Mask R-CNN 模型识别小天鹅精度-召回率曲线

Fig. 6 Recall-precision (P-R) curves of Mask R-CNN models for identifying *Cygnus columbianus* with different training set sizes

表 6 不同大小训练集所得 Mask R-CNN 模型识别小天鹅结果平均精度、F1 分数、召回率和精度

Table 6 Average precision, F1-score, recall, and precision of Mask R-CNN models for identifying *Cygnus columbianus* detection with different training dataset

训练集中小天鹅数量 Number of <i>Cygnus columbianus</i> (ind)	平均精度 Average Precision	F1 分数 F1-score	召回率 (%) Recall	精度 (%) Precision
12	0.41	0.52	54.00	50.00
18	0.18	0.32	46.00	24.79
42	0.77	0.76	76.00	75.38
81	0.85	0.84	79.00	89.47
124	0.86	0.87	87.00	86.15
161	0.90	0.81	82.00	80.30
235	0.91	0.84	84.00	84.38

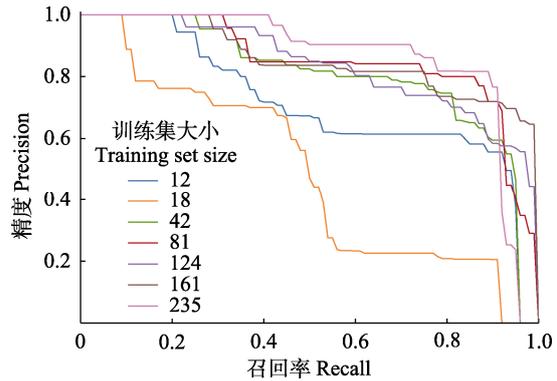


图 7 不同大小训练集所得 YOLOv3 模型识别小天鹅精度-召回率曲线

Fig. 7 Recall-precision (P-R) curves of YOLOv3 models for identifying *Cygnus columbianus* with different training set sizes

影像中的水鸟建立数据集，其次训练卷积神经网络模型 Mask R-CNN 与 YOLOv3，最终实现遥感影像中水鸟的自动识别并对结果进行评价。

目前使用卷积神经网络识别无人机遥感影像中野生鸟类的技术仍处于起步阶段，相似研究较少，如使用 Mask R-CNN 模型在无人机红外遥感影像中自动识别野火鸡，召回率达到 94%，精度为 86.6%（Kassim et al. 2020）。在

表 7 不同大小训练集所得 YOLOv3 模型识别小天鹅结果平均精度、F1 分数、召回率和精度

Table 7 Average precision, F1-score, recall, and precision of YOLOv3 models for identifying *Cygnus columbianus* detection with different training dataset

训练集中小天鹅数量 Number of <i>Cygnus columbianus</i> (ind)	平均精度 Average Precision	F1 分数 F1-score	召回率 (%) Recall	精度 (%) Precision
12	0.72	0.71	84.00	61.36
18	0.50	0.54	46.00	66.67
42	0.81	0.78	82.00	74.65
81	0.86	0.83	87.00	80.00
124	0.84	0.77	87.00	68.29
161	0.86	0.81	93.00	71.43
235	0.87	0.86	90.00	81.69

识别牲畜或大型野生动物方面,相似研究采用 YOLO 模型自动识别无人机遥感影像中的牛 (*Bos taurus*),达到了 95.7%的精度和 94.6%的召回率 (Shao et al. 2020)。使用特别设计的卷积神经网络识别 Kuzikus 保护区内的野生动物,达到了 83.7%的精度和 78.4%的召回率 (Kellenberger et al. 2019)。使用 Mask R-CNN 模型识别无人机遥感影像中的牛和羊 (*Ovis aries*),达到了 96%的精度和 92%的召回率 (Xu et al. 2020)。

与其他结合无人机遥感与卷积神经网络技术识别鸟类的研究或是识别无人机遥感图像中大型牲畜的相关研究结果相比,本研究所得模型识别绿翅鸭/罗纹鸭的精度与召回率较高,达到了其他研究中卷积神经网络模型识别大型动物的水平。在训练数据集大小相近时,卷积神经网络识别小天鹅的效果稍差,我们认为这可能是由于照片中存在的白色干扰物比黑色干扰物更多,或是由于小天鹅体型较大,导致其在照片中展现不同的姿势对模型造成了更大的影响。但由于本次研究未获得足够数量和分辨率的数据,尚无法深入研究该技术在区分不同种类的水鸟时的表现,该技术在区分遥感影像中不同种类的鸟类,特别是区分外形及大小相近的鸟类(如区分各种鸭类)时的效果仍需进一

步探索。

在本研究中,当使用最大的训练集时,我们观察到 Mask R-CNN 和 YOLO 在水鸟识别任务上的表现差距不大,Mask R-CNN 有着稍高的平均精度,但 F1 分数略低于 YOLO, YOLO 的检测速度高于 Mask R-CNN,接近实时检测的需求,当训练集相对较小的情况下, YOLO 表现出更佳的性能。这表明,在背景简单且无需精确标出目标边界的水鸟识别任务中,两种神经网络模型都表现出了相当的能力。但需要强调的是,本次实验的水鸟样本相对较少,水鸟之间的间距较大,它们与背景之间的颜色差异显著,且本次实验的水鸟外形较为规则,没有如长伸的脖颈等复杂多变的外形。在实际应用中,可能会面临更为复杂的水鸟识别场景,因此,实际运用中仍需要更多的评估和研究,以确定在不同的水鸟识别场景中应选择哪种模型以满足特定需求。我们根据实验结果推测,在可以获得足够的训练数据量时(参考罗纹鸭和绿翅鸭的识别结果,每种水鸟所需训练量可能少至 250 只),结合无人机遥感与卷积神经网络技术可以用于水鸟自动监测。此外,该技术同样可以在极地、平静海面 and 裸露崖壁等遮蔽物较少的生境识别鸟类,也可用于识别其他动物。未来将卷积神经网络技术与云计算、无人机编队飞行、多光谱遥感及激光雷达遥感等技术综合运用,可以进一步提高识别的精度和速度,有望应用于大范围的水鸟同步监测。

参 考 文 献

- Chen K, Wang J Q, Pang J M, et al. 2019. MMDetection: open MMLab detection toolbox and benchmark. [EB/OL]. [2023-05-05]. <http://arxiv.org/abs/1906.07155>.
- Díaz-Delgado R, Mañez M, Martínez A, et al. 2017. Using UAVs to map aquatic bird colonies // Díaz-Delgado R, Lucas R, Hurford C. The Roles of Remote Sensing in Nature Conservation. Cham: Springer, 277–291.
- Gu J X, Wang Z H, Kuen J, et al. 2018. Recent advances in convolutional neural networks. *Pattern Recognition*, 77(C):

- 354–377.
- He K M, Gkioxari G, Dollár P, et al. 2017. Mask R-CNN // Institute of Electrical and Electronics Engineers. 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy. IEEE, 2980–2988.
- Hou J, He Y X, Yang H B, et al. 2020. Identification of animal individuals using deep learning: a case study of giant panda. *Biological Conservation*, 242: 108414.
- Kahl S, Wilhelm-Stein T, Hussein H, et al. 2017. Large-scale bird sound classification using convolutional neural networks. *CLEF (working notes)*, 1866.
- Kassim Y M, Byrne M E, Burch C, et al. 2020. Small object bird detection in infrared drone videos using Mask R-CNN deep learning. *Electronic Imaging*, 32(8): 85-1–85-8.
- Kellenberger B, Marcos D, Lobry S, et al. 2019. Half a percent of labels is enough: efficient animal detection in UAV imagery using deep CNNs and active learning. *IEEE Transactions on Geoscience and Remote Sensing*, 57(12): 9524–9533.
- Kellenberger B, Volpi M, Tuia D. 2017. Fast animal detection in UAV images using convolutional neural networks // Institute of Electrical and Electronics Engineers. 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Fort Worth, TX, USA. IEEE, 866–869.
- Korczak-Abshire M, Zmarz A, Rodzewicz M, et al. 2019. Study of fauna population changes on Penguin Island and Turret Point Oasis (King George Island, *Antarctica*) using an unmanned aerial vehicle. *Polar Biology*, 42(1): 217–224.
- Krizhevsky A, Sutskever I, Hinton G E. 2012. ImageNet classification with deep convolutional neural networks // Proceedings of the 25th International Conference on Neural Information Processing Systems: Volume 1. Red Hook, NY, USA: Curran Associates, 1097–1105.
- Li H, Dong Y Q, Liu Y X, et al. 2022. Design and implementation of UAVs for bird's nest inspection on transmission lines based on deep learning. *Drones*, 6(9): 252.
- Redmon J, Farhadi A. 2018. YOLOv3: an incremental improvement. [EB/OL]. [2023-05-05]. <http://arxiv.org/abs/1804.02767.pdf>.
- Shafer M W, Vega G, Rothfus K, et al. 2019. UAV wildlife radiotelemetry: system and methods of localization. *Methods in Ecology and Evolution*, 10(10): 1783–1795.
- Shao W, Kawakami R, Yoshihashi R, et al. 2020. Cattle detection and counting in UAV images based on convolutional neural networks. *International Journal of Remote Sensing*, 41(1): 31–52.
- Xu B B, Wang W S, Falzon G, et al. 2020. Livestock classification and counting in quadcopter aerial images using Mask R-CNN. *International Journal of Remote Sensing*, 41(21): 8121–8142.
- 蔡前舟. 2022. 基于卷积神经网络的野生动物种类识别中长尾问题研究. 南充: 西华师范大学硕士学位论文.
- 陈刚琦. 2020. 基于卷积神经网络的高原鼠兔图像检测与分割方法研究. 兰州: 兰州理工大学硕士学位论文.
- 高大中, 林海, 林乐乐, 等. 2021. 利用小型无人机监测西洞庭湖水鸟的可行性探讨. *动物学杂志*, 56(1): 100–110.
- 胡旭. 2019. 基于注意力机制的金丝猴面部识别研究与实现. 西安: 西安电子科技大学硕士学位论文.
- 雷倩, 李金亚, 马克明. 2018. 遥感技术在鸟类生态学研究中的应用. *生物多样性*, 26(8): 862–877.
- 李杰, 刘强. 2019. 无人机水禽监测模式的设立原则探讨. *热带地理*, 39(4): 546–552.
- 史春妹, 谢佳君, 顾佳音, 等. 2021. 基于目标检测的东北虎个体自动识别. *生态学报*, 41(12): 4685–4693.
- 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.