



主持人: 任明迅

# 基于残差卷积神经网络模型的勺嘴鹬动作识别

杨雪珂<sup>1</sup>, 蒙金超<sup>1</sup>, 冯悦恒<sup>1</sup>, 林婷婷<sup>1,2</sup>, 王兆君<sup>2</sup>, 刘 辉<sup>1</sup>

(1. 海南大学 林学院, 海口 570228; 2. 中国科学院 动物研究所, 北京 100101)

**摘要:** 为开启海南热带地区鸕鹚类涉禽的动作识别以及其他野生鸟类行为学自动识别的研究, 建立了基于野外采集影像的勺嘴鹬 (*Eurynorhynchus pygmeus*) 动作图像数据集。该数据集由表达勺嘴鹬主要行为模式的9种动作标签组成; 同时利用 ResNet50、ResNet101 和 ResNet152 共3种残差卷积神经网络模型尝试对勺嘴鹬的动作进行自动识别。结果表明, ResNet50、ResNet101、ResNet152 测试集准确率分别为 96.90%、96.94% 和 96.90%, 说明3种模型都能对勺嘴鹬图像进行快速准确的动作识别。

**关键词:** 残差卷积神经网络; 鸟类图像; 动作识别; 勺嘴鹬

中图分类号: Q958 文献标志码: A 文章编号: 1674-7054(2023)05-0481-09

杨雪珂, 蒙金超, 冯悦恒, 等. 基于残差卷积神经网络模型的勺嘴鹬动作识别 [J]. 热带生物学报, 2023, 14(5): 481-489. doi: 10.15886/j.cnki.rdswwb.20220088

随着科技的发展, 鸟类图像的大量采集变得更加便捷, 可利用图像采集设备(如红外相机<sup>[1]</sup>、普通相机)采集, 也可利用智能手机采集。随着鸟类数据共享平台(eBird<sup>[2]</sup>、中国爱鸟网等)的建立, 来自世界各地的大量鸟类图像数据被上传至网络数据库。仅以 eBird 为例, 已有超过 2 000 万张鸟类图像被上传至该平台, 并且图像数量依旧在迅速增长。鸟类图像记录了鸟类形态学特征、生境信息<sup>[3]</sup> 及行为学特征, 对鸟类学研究有着重要的价值, 但面对如此大量的鸟类图像数据, 仅凭人工处理是无法满足需求的。为了能快速自动化处理大量的鸟类图像数据, 深度学习(Deep Learning)领域的研究人员已经开展了相关研究。图像识别技术应用于鸟类物种识别已有一些成功案例, 在标准鸟类图像数据库 CUB200-2011<sup>[4]</sup> 的技术报告中, Welinder 等<sup>[4]</sup> 就使用局部区域和基于传统特征的词包模型实现分类, Berg 等<sup>[5]</sup> 提出 POOF 特征, Yao<sup>[6]</sup> 和 Yang<sup>[7]</sup> 等均尝试使用模板匹配的方法替换定位算法中的滑动窗口, 以降低算法复杂度。基于图像识别技术的珍稀濒危鸟类的行为识别具

有重大应用价值<sup>[8]</sup>, 但整体看来, 目前的研究主要涉及计数<sup>[9]</sup>、密度估计<sup>[10]</sup>、生境因素识别<sup>[11]</sup> 等。虽然目前自动监测获得的影像数据量极大, 其中很大一部分行为表达数据对于无人干扰下鸟类行为的研究很有价值, 但是目前使用这种方式对野生鸟类行为学进行的研究少之又少。动物通常以身体运动和身体姿势来表达行为, 动物的行为是其对环境 and 生理状况的一种反应, 为研究动物健康水平提供了重要的线索<sup>[12]</sup>, 所以进行动物行为识别或动作识别是十分必要的。

卷积神经网络(Convolutional Neural Networks, CNN)为深度学习的代表算法之一, 其在图像识别领域展现了非常大的潜力和极佳的性能<sup>[13-15]</sup>。经典的神经网络模型包括 ResNet<sup>[16]</sup>、VGG Net<sup>[17]</sup>、ALexNet<sup>[18]</sup>、GoogLeNet<sup>[19]</sup> 等, 其中, 残差网络(Residual Network, ResNet)良好地解决了网络加深带来的学习退化问题, 该模型目前在各领域取得了广泛应用<sup>[20]</sup>。目前利用卷积神经网络来实现野生动物自动识别的研究较多。史春妹等<sup>[21]</sup> 运用单次多盒目标检测方法来进行东北虎的个体识

收稿日期: 2022-10-22

修回日期: 2023-01-24

第一作者: 杨雪珂(1997-), 女, 海南大学林学院 2020 级硕士研究生. E-mail: 15136910537@163.com

通信作者: 刘辉(1988-), 男, 博士. 讲师. 研究方向: 热带野生动物保护. E-mail: liuhui\_leen@163.com

别,达到 97.4% 的准确率。石鑫鑫等<sup>[22]</sup>提出了一种全连接算法与稀疏连接算法相结合的全卷积神经网络解决了蛙声识别问题,准确率达到 99.67%。还有使用基于感兴趣区域与卷积神经网络的野生动物物种自动识别方法实现了基于野生动物监测图像的物种识别研究,平均识别率均可达到 90% 左右<sup>[23]</sup>。残差网络模型的应用研究成果对于鸟类图像大数据的有效利用有着重要的辅助作用,但仍不能满足实际需要,仍缺乏应用卷积神经网络进行野生鸟类的动作自动化识别研究<sup>[24]</sup>。

勺嘴鹬隶属鹬科 (Scolopacidae) 滨鹬属 (Calidris), 是一种仅分布于东亚-澳大利西亚候鸟迁徙路线上的涉禽<sup>[25]</sup>, 被世界自然保护联盟 (IUCN) 红色名录列为极度濒危 (CR) 物种。其在我国的江苏、浙江、福建、广东、海南等省份均有被记录到, 其中, 海南儋州湾是海南岛目前已知的为数不多的勺嘴鹬的为数不多的越冬地。繁殖期以外的时期, 勺嘴鹬只在滨海滩涂湿地有分布, 觅食地主要为潮间带的滩涂<sup>[26]</sup>。目前, 国内外主要利用环志等回收数据, 开展勺嘴鹬的栖息地保护<sup>[27]</sup>、种群数量<sup>[28]</sup> 和分布区系<sup>[29]</sup> 等的研究, 未涉及其行为动作的识别。因此, 笔者尝试使用卷积神经网络模型 (ResNet50、ResNet101 和 ResNet152) 进行勺嘴鹬动作的自动识别研究, 以期开启海南热带地区鹬类<sup>[30]</sup> 涉禽的动作识别以及其他野生鸟类行为学自动识别的研究。

## 1 材料与方法

**1.1 研究区域概况** 儋州湾 (109°02' ~ 109°36'E, 19°33' ~ 20°01'N) 位于海南省儋州市中北部, 由北部湾伸入洋浦半岛进而形成的半封闭内湾, 面积约为 50 km<sup>2</sup>。该区域属于季风性气候, 冬季干燥、夏季湿润, 年平均降雨量约 1 426 mm, 年均气温 23.1 °C<sup>[31]</sup>。儋州湾于 1986 年被设立为市级自然保护区, 红树林面积约 133 hm<sup>2</sup>, 是黑脸琵鹭 (*Platalea minor*)、小青脚鹬 (*Tringa guttifer*)、勺嘴鹬等珍稀濒危迁徙涉禽的重要越冬地和停歇地<sup>[32]</sup>。

### 1.2 研究方法

**1.2.1 数据的采集与预处理** 数据采集于 2020-11-21—2021-03-20, 采集时间段主要集中在 9:00-17:30, 共获得 42 份用相机拍摄的勺嘴鹬视频 (佳










能 SX60HS 长焦数码相机、尼康 D500 单反相机和尼康 Z6 微单相机, 视频尺寸为 1920×1080 30p)。将采集的数据进行预处理: 1) 通过查找相关文献、咨询专家和快速浏览现有视频资料确定可以用于识别勺嘴鹬动作的标签 (表 1); 2) 将 42 份视频数据解帧 (每 5 秒解帧, 有重复的图片只保留 1 张, 再进行人工修正), 共获得 66 875 张勺嘴鹬图像; 3) 删除勺嘴鹬与其他鸟类同框的图像和因拍摄抖动或对焦失败形成的模糊图像, 手动筛选相应标签的图像 (每 1 张图像包含 1 种标签), 并在筛选结束后创建以标签命名的文件夹, 最后共获得 9 个标签文件夹; 4) 使用 Grad-Cam<sup>[33]</sup> 对 3 种模型的预测结果进行可视化处理 (图 1)。图 1 中橙色部分表示模型是基于这些区域判断出图像中的动作类别。

**1.2.2 构建数据集** 由于目前没有用于识别勺嘴鹬动作的公共数据集, 因此, 笔者建立了 1 个由猎食、觅食、休憩、理羽、洗浴、抖羽、振翅和踱步 9 种动作标签构成的共 2 174 张图片的数据集 (表 2), 该数据集按照 3 : 1 : 1 的比例随机划为训练集、验证集和测试集<sup>[34]</sup>。

**1.3 残差卷积神经网络和迁移学习** 卷积神经网络包含多个卷积层、池化层和全连接层<sup>[35]</sup>。卷积层和池化层是提炼图像特征关键部分的模型, 全连接层能够在高层次特征域内把图像分类作为主要实现的图像映射<sup>[36-38]</sup>。卷积神经网络也可以被认为是由特征提取器和分类器 2 个部分组成, 具有端到端特征提取和分类的特性。在卷积神经网络的结构中, 卷积的层次更深, 网络学习的能力也就更强, 那么特征图能得到的信息也会更全。然而, 随着网络层次和结构逐渐加深, 网络内的计算量也将随之增多, 进而导致网络也变得更为复杂, 同时可能会导致梯度消失和网络退化等问题<sup>[39]</sup>, 从而导致识别效果和稳定性都不理想。

残差网络 (ResNet) 是最近十多年以来相关领域研究人员提出的最新关于执行计算机领域视觉任务的一种典型的卷积神经网络, 因其加入了残差模块从而减少了随网络深度的增加而引起的梯度消失的问题<sup>[40]</sup>, 一方面减少了参数数量, 另一方面在网络中增加了直连通道, 增加了卷积神经网络对特征的学习能力<sup>[41]</sup>。鉴于此背景, 笔者使用 ResNet101<sup>[42]</sup>、ResNet50<sup>[43]</sup> 和 ResNet152<sup>[44]</sup> 3 种不

表 1 勺嘴鹬的动作分类及定义

动作标签	描述	图像
理羽	用喙部梳理、修饰身体不同部位羽毛的过程	
猎食	狩猎食物的一种警戒状态, 一般表现为不停转圈、直立伸长脖颈等观察食物位置的一系列动作	
抖羽	使羽毛微微鼓起或竖起, 抖动羽毛的过程	
进食	吞咽食物的一系列动作	
觅食	将喙部贴在水面或深入水下寻觅食物的过程	
踱步	脖子或身子略向前倾, 脚抬起往前的过程	
休憩	头颈后转、喙前端埋于翅下或头颈略为回缩、闭眼静止不动的行为	
洗浴	将身体浸入水中或在水体表面, 通过颈部伸缩、翅膀的快速抖动及身体摆动等动作将水遍布身体, 清洗身体的过程	
振翅	大幅度拍打翅膀或翅膀完全张开的过程	

同层数的残差卷积神经网络模型作为本研究的基本网络。

因本研究数据集的图像相对较少, 所以采用残差网络迁移学习<sup>[20]</sup>的方法, 将前人训练 ImageNet 图像数据集得到的不同深度残差网络模型的模型参数当作勺嘴鹬图像训练 3 种网络模型的初始化值, 然后把预训练网络内的最后一层的全连接层输出替换为本研究的勺嘴鹬图像数据集的类别数 9, 以此为基础, 再将勺嘴鹬图像识别模型进行下一步训练(图 2)。

**1.4 实验环境配置** 实验中所有代码均是在 PyTorch<sup>[45]</sup>神经网络框架下完成的。本研究选

用的是 PyTorch1.3.1, 框架环境为 GPU: Tesla V100Mem: 32 GB, 操作系统是 Linux Cento OS 7.2 服务器。

### 1.5 实验模型

**1.5.1 模型结构** 本研究所用模型是以经典的 ResNet50、ResNet101 和 ResNet152 模型为基础进行新的改善, 网络结构超参数具体设置分别为损失函数设置为交叉熵函数; 优化器设置为自适应矩估算法; 学习率设置为 0.0001; 训练轮数为 100; 批量为 64。

**1.5.2 模型优化与评价标准** 本研究采用自适应矩估计(Adam)方法<sup>[46]</sup>进行模型的优化。图像分



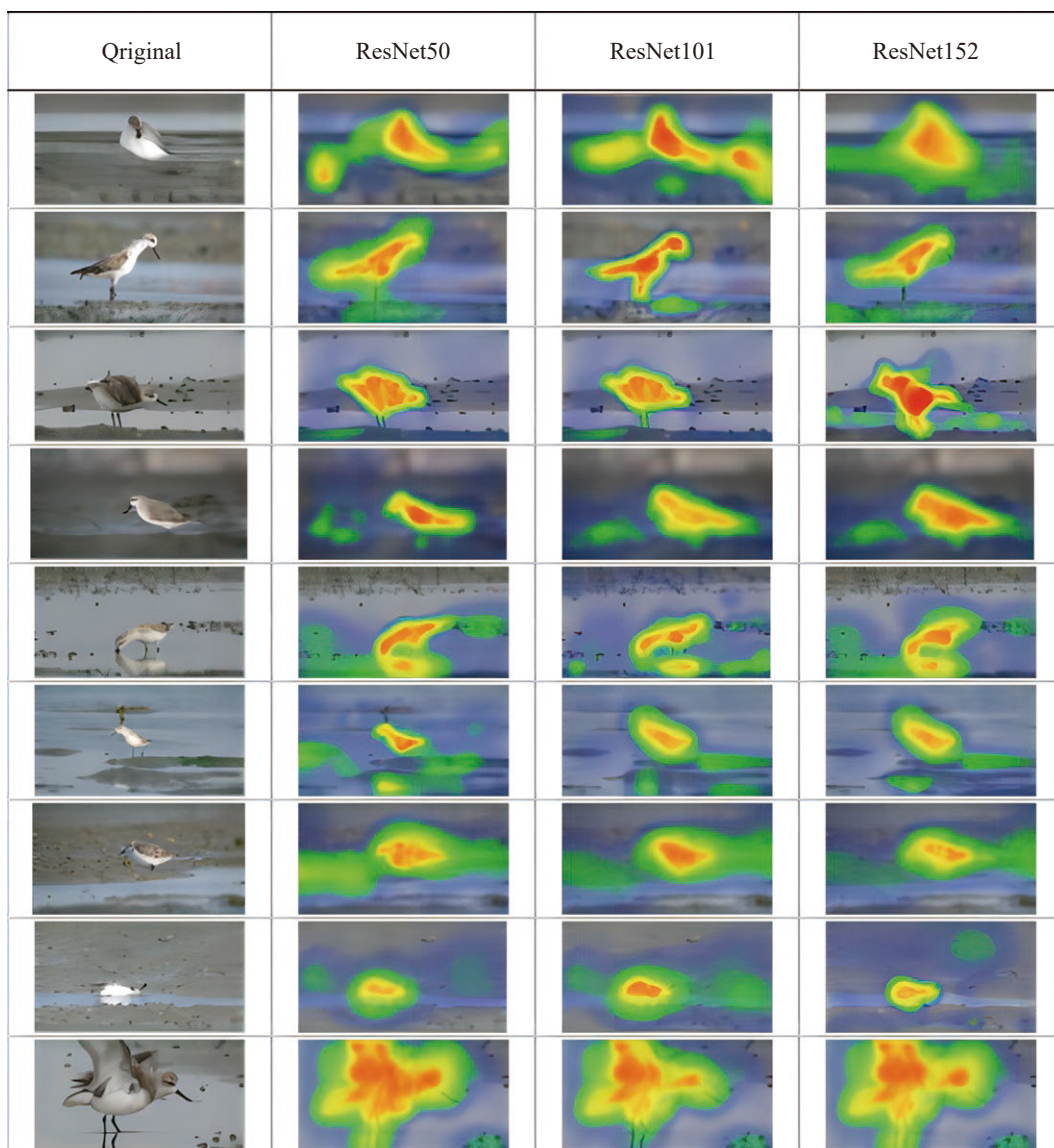


图1 Grad-CAM 热力图

注: 图像中各部分对预测结果的权重, 颜色越深红, 表示该区域对网络的响应越大, 对预测结果的贡献越大, 也即该模型对该区域的关注度越高。

表2 动作标签数据集

动作标签	数量/张	训练集/张	验证集/张	测试集/张
理羽	463	270	90	90
猎食	229	89	30	30
抖羽	126	74	25	25
进食	69	32	11	11
觅食	407	196	66	66
踱步	191	93	31	32
休憩	171	100	34	34
洗浴	401	240	80	81
振翅	90	52	17	18
总数	2174	1146	384	387

类评估中常用的评价指标有准确率(精度)、精确率、召回率和 F1-score<sup>[47]</sup>, 在此基础上本研究添加模型训练时长作为模型的评价指标。根据真实标签和预测结果, 将所有测试图像分为 4 类。总共有 4 个基本数: 真阳性( $TP$ )、真阴性( $TN$ )、假阳性( $FP$ )和假阴性( $FN$ )。正确预测的测试图像的数量由  $TP$  和  $TN$  表示, 错误预测的图像的数量由  $FN$  和  $FP$  表示。准确率就是正确分类样本所占总样本的比率, 能够衡量分类器对于总样本的判断能力<sup>[33]</sup>。精确率又称查准率, 值越高说明正确分类率越高。召回率又称查全率, 值越高说明识别得越全面。F1-score 表示的是算法的综合性能, 可以

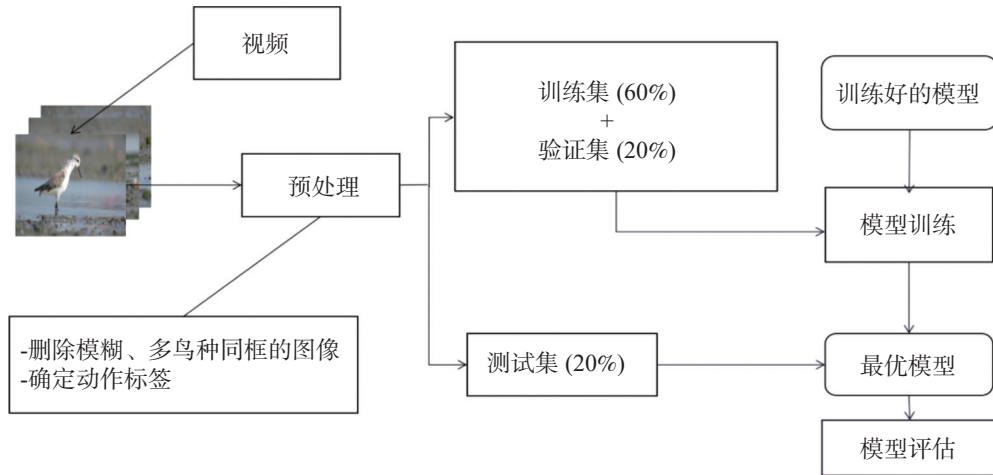


图 2 残差网络数据分析流程图

平衡召回率和准确率的影响,其取值范围为 0~1,值越大表示算法性能越好。以上评价标准形式化定义如下:

$$\text{准确率} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{精确率} = \frac{TP}{TP+FP}$$

$$\text{召回率} = \frac{TP}{TP+FN}$$

$$F1-score = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}$$

式中: 阳性与阴性是相对的,若阳性代表觅食,阴性代表踱步;  $TP$  为真阳性(true positive)图像数目,真阳性则证明这个图像在整个预测分类的结果和标记标签中属于觅食;  $FP$  为假阳性(false positive)图像数目,假阳性证明此图像标记标签是觅食,但是在预测分类结果中为踱步;  $TN$  为真阴性(true negative)图像数目,真阴性证明该图像标记标签为踱步,并且对其预测的结果也属于踱步;  $FN$ (false negative)是假阴性图像数目,假阴性是图像标记标签踱步,但在预测的分类结果里是觅食<sup>[26]</sup>。

## 2 结果与分析

**2.1 不同模型的损失和验证精度** 随着训练轮数的增加,每个模型的预测值与真实值之间的损失都呈下降趋势,验证集中的准确性都呈上升趋势(图 3)。最后,随着训练轮数增多,每个模型的验证准确率都趋于稳定,仅存在小幅波动,这表明网络得到了充分的训练。

**2.2 不同模型的识别准确率和训练时长** 从准确率来说,3 种网络模型准确率之间的差别较小,

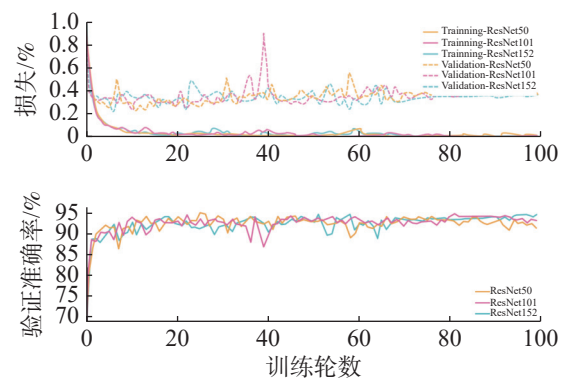


图 3 3 种模型的损失和准确率

不同模型的验证集准确率和测试集准确率都在 95% 以上,说明模型的泛化能力良好(表 3)。模型在每轮训练时长上都有较快的速度,时长由短到长依次为 ResNet50、ResNet101 和 ResNet152,其中,ResNet50 和 ResNet152 测试集准确率都是 96.90%,ResNet101 测试集准确率为 96.64%,低于其他 2 个模型。ResNet50 训练时长是 89.78 s·轮,ResNet152 训练时长是 101.87 s·轮,在同样准确率的情况下,ResNet50 训练时长低于 ResNet152。

**2.3 3 种模型的精度、召回率及 F1-score** 图 4 展示了不同模型在数据集上的评价指标结果。进食动作标签在 ResNet101 和 ResNet152 模型的

表 3 ResNet50、ResNet101 和 ResNet152 模型的准确率和训练时长

模型	训练集 准确率/%	验证集 准确率/%	测试集 准确率/%	训练时 长/(s·轮)
ResNet50	92.41	96.61	96.90	89.78
ResNet152	90.49	95.31	96.90	101.87
ResNet101	87.87	97.40	96.64	95.53

F1-score 值偏低, 除此之外的其他动作标签在不同的模型的 F1-score 值都高于 0.90, 表明本研究算法的综合性能较好。在模型精度方面, 除踱步和进食 2 种动作标签的精度在 ResNet50 和 ResNet101 上低于 0.90, 其余标签的精度都不低于

0.90, 说明 3 种模型在识别动作标签时都有较强的识别能力。在召回率方面, 进食标签在 ResNet152 上的召回率在 0.80 以下, 其余标签在不同的模型上召回率都在 0.80 以上, 且大部分高于 0.90, 说明进食动作识别得不够全面。

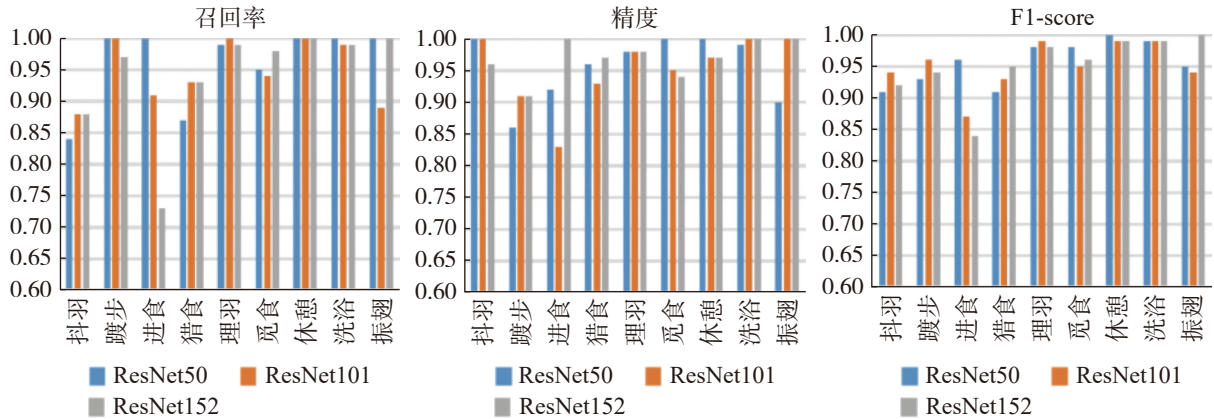


图 4 3 种模型的精度、召回率及 F1-score

2.4 勺嘴鹬 9 种动作标签的预测分类结果 图 5 对角线上表示在测试集中预测正确的图像数量, 其余为预测错误的数量。模型对一些动作的识别容易存在误判现象, 如 ResNet50 模型在识别猎食和踱步时错判最多, 有 4 张猎食被错判成踱步;

ResNet101 模型有 2 张抖羽被错判成理羽; ResNet152 模型有 3 张进食被错判成觅食。ResNet50、ResNet101 和 ResNet152 模型识别错误的图像总数分别是 12、13、20 张。

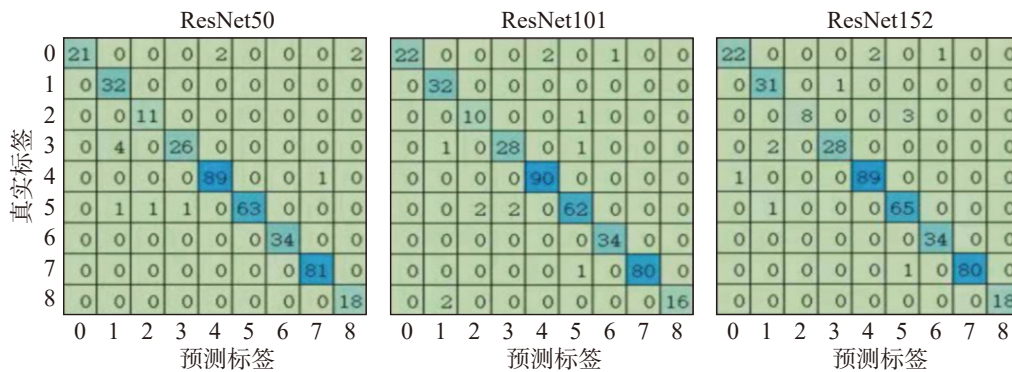


图 5 测试集勺嘴鹬图像识别后统计的混淆矩阵

注: 0~8 分别表示抖羽、踱步、进食、猎食、理羽、觅食、休憩、洗浴和振翅 9 种动作标签。

### 3 讨论

目前在家禽的动作和行为识别方面已经有了初步研究。劳凤丹等<sup>[48]</sup>基于人工设计的 10 种特征利用贝叶斯分类法对单只蛋鸡的行为进行了识别, 取得了不错的结果。但文献 [48] 中蛋鸡的行为识别是在人工设计的特征(如蛋鸡图像的质心点坐标、轮廓面积、移动距离等)的基础上进行的, 人工设计的特征往往对领域知识依赖度高, 还需要大量的实验测试, 可能只在特定的任务上才能获

得不错的效果<sup>[49]</sup>。相比而言, 本研究利用卷积神经网络实现“端到端”的方式进行勺嘴鹬动作识别, 特征提取和分类过程均由模型自动完成, 使用这种方法的门槛被大大降低。Wang 等<sup>[11]</sup>验证了利用深度卷积神经网络从鸟类图像中识别栖息环境元素的可行性, 最大识别率达到 95.52%, 所以笔者期望未来可以进行勺嘴鹬等稀濒危涉禽栖息地的因素识别, 进一步推进栖息地选择的研究。本研究的数据集样本相对较少, 只实现了 9 种动作识



别, 未实现勺嘴鹬所有可能动作的识别, 期望未来可以通过补充勺嘴鹬不同动作类别的图像进一步完善勺嘴鹬的其他动作的识别研究。因鸬鹚类涉禽可能在外观、体型或行为相似的情况(勺嘴鹬与红颈滨鹬), 所以模型未来可以推广到鸬鹚类涉禽的动作识别, 进一步推动有关濒危珍稀涉禽的保护研究。

本研究的不足主要有两个。第一, 只进行勺嘴鹬的动作研究, 所以模型在识别与勺嘴鹬体型、行为等方面差异较大的鸟类时, 识别效果可能不大理想。此外, 因一些不可避免的环境因素和人为因素, 采集的视频数据主要集中在光线较好的9:00-17:30, 所以模型可能更适用于在光线良好时间段拍摄图像的识别。本研究中踱步和进食的识别精度低于其他动作的主要原因是数据量不够充足, 导致识别某几个标签的时候容易混淆, 而且因为静态图像识别动作的缺陷, 导致踱步和进食的识别较差, 笔者会在后续工作中进行数据补充。第二, 使用的是单标签方法来标记勺嘴鹬的动作, 而在实际应用时可能会出现1张图像有多标签、部分分类照片较少的情况, 迁徙候鸟具有显著的集群行为<sup>[50]</sup>, 此类研究往往需要识别1张图像中不同鸟种的不同动作。如果未来通过结合目标检测、图像分割和物种识别等技术, 把一群鸟转换成单只鸟进行识别, 实现从多种鸟同框的图片中识别出多个不同的动作并用一个框将每个动作分割出来, 进一步再通过采集更多的图像数据进行训练, 可以使训练模型在实际应用时具有更强的适应性。

**致谢:** 新英湾红树林保护区陈正平同志和中国热带农业科学院橡胶研究所杨川助理研究员对本研究的野外工作提供了大力的支持, 在此深表感谢!

## 参考文献:

- [1] 王丞, 冉伟, 杨朝辉, 等. 梵净山保护区主要雉类的繁殖期栖息地选择与空间分布[J]. *林业科学*, 2020, 56(11): 134 - 142.
- [2] SULLIVAN B L, WOOD C L, ILIFF M J, et al. eBird: A citizen-based bird observation network in the biological sciences [J]. *Biological Conservation*, 2009, 142(10): 2282 - 2292.
- [3] 肖治术, 王学志, 黄小群. 青城山森林公园兽类和鸟类资源初步调查: 基于红外相机数据[J]. *生物多样性*, 2014, 22(6): 788 - 793.
- [4] WELINDER P, BRANSON S, MITA T, et al. Caltech-UCSD Birds 200[J]. California Institute of Technology. 2010:CNS-TR-2010-001. .
- [5] BERG T, BELHUMEUR P N. Poof: Part-based one-vs.-one features for fine-grained categorization, face verification, and attribute estimation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013: 955-962.
- [6] YAO B, BRADSKI G, Fei-Fei L. A codebook-free and annotation-free approach for fine-grained image categorization[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2012:3466-3473.
- [7] YANG S, BO L, WANG J, et al. Unsupervised template learning for fine-grained object recognition [J]. *Advances in neural information processing systems*, 2012,: 3122-3130.
- [8] 周智恒, 牛畅, 尚俊媛, 等. 一种基于结构保持零样本学习的鸟类濒危物种识别方法: CN110717512A[P]. 2023-04-07.
- [9] 李鹏博, 王向文. 基于深度特征融合生成的密集人群计数网络[J]. *计算机应用与软件*, 2021, 38(3): 153 - 158.
- [10] 李婧, 吴俊峰, 于红. 一种基于冗余裁剪的鱼群密度估计算法[J]. *计算机与数字工程*, 2020, 48(12): 2864 - 2868.
- [11] WANG Z, WANG J, LIN C, et al. Identifying habitat elements from bird images using deep convolutional neural networks [J]. *Animals*, 2021, 11(5): 1263.
- [12] 王鹏, 唐尚波, 陆舟, 等. 广西山心沙岛的春季水鸟群落[J]. *野生动物学报*, 2019, 40(4): 957 - 963.
- [13] 李亚召, 云利军, 叶志霞, 等. 基于卷积神经网络的霉变烟叶图像识别方法研究[J]. *计算机工程与科学*, 2021, 43(3): 473 - 479.
- [14] 刘仲博. 基于卷积神经网络的电选粉煤灰颗粒图像识别与烧失量预测模型[J]. *中国矿业*, 2021, 30(5): 125 - 129.
- [15] 陈恒晟, 王军, 毛毅, 等. 基于协同特征的显著性目标检测算法[J]. *通信技术*, 2021, 54(8): 1883 - 1890.
- [16] 郑秋梅, 谭丹, 王风华. 基于改进 ResNet 网络的交通标志识别研究[J]. *计算机与数字工程*, 2021, 49(5): 947 - 951.
- [17] JANGRA M, DHULL S K, SINGH K K, et al. ECG arrhythmia classification using modified visual geometry group network (mVGGNet)[J]. *Journal of Intelligent & Fuzzy Systems*. 2020,38(3) : 3151-3165.
- [18] WAGLE S A, HARIKRISHNAN R. Comparison of Plant Leaf Classification Using Modified AlexNet and Support Vector Machine [J]. *Traitement du Signal*, 2021, 38(1): 79 - 87.
- [19] BOBKOWSKA K, BODUS-OLKOWSKA I. Potential and Use of the Googlenet Ann for the Purposes of Inland Water Ships Classification [J]. *Polish Maritime Research*, 2020, 27(4): 170 - 178.
- [20] 柳天滋, 陈昕, 李想, 等. 基于深度残差神经网络迁移

- 学习的牙形刺图像识别[J]. *古生物学报*, 2020, 59(4): 512 – 523.
- [21] 史春妹, 谢佳君, 顾佳音, 等. 基于目标检测的东北虎个体自动识别[J/OL]. *生态学报*, 2021(12): 1-9.
- [22] 石鑫鑫, 鱼昕, 刘铭. FCNN 深度学习模型及其在动物语音识别中的应用[J]. *吉林大学学报(信息科学版)*, 2021, 39(1): 60 – 65.
- [23] 刘文定, 李安琪, 张军国, 等. 基于 ROI-CNN 的赛罕乌拉国家级自然保护区陆生野生动物自动识别[J]. *北京林业大学学报*, 2018, 40(8): 123 – 131.
- [24] 陈斌, 朱晋宁, 东一舟. 基于残差整流增强卷积神经网络的表情识别[J]. *液晶与显示*, 2020, 35(12): 1299 – 1308.
- [25] 曲方圆, 李淑芸, 赵林林, 等. 黄海生态区保护空缺分析[J]. *生物多样性*, 2021, 29(3): 385 – 393.
- [26] 彭鹤博, 蔡志扬, 章麟, 等. 勺嘴鹬在中国的分布状况和面临的主要威胁[J]. *动物学杂志*, 2017, 52(1): 158 – 166.
- [27] 马天, 张国钢, Syroechkovski E E, 等. 俄罗斯远东地区勺嘴鹬繁殖地夏季水鸟调查[J]. *动物学杂志*, 2018, 53(4): 507 – 518.
- [28] AUNG P P, MOSES S, CLARK N A, et al. Recent changes in the number of spoon-billed sandpipers *Calidris pygmaea* wintering on the Upper Gulf of Mottama in Myanmar [J]. *Oryx*, 2018, 54(1): 23 – 7.
- [29] CLARK N A., ANDERSON G Q A, LI J, et al. First formal estimate of the world population of the Critically Endangered spoon-billed sandpiper *Calidris pygmaea* [J]. *Oryx*, 2018, 52(1): 137 – 146.
- [30] 冯尔辉, 陈伟, 廖宝文, 等. 海南东寨港红树林湿地鸟类监测与研究[J]. *热带生物学报*, 2012, 3(1): 73 – 77.
- [31] 褚梦凡, 肖晓彤, 丁杨, 等. 海南儋州湾红树林区沉积有机质来源及碳储量[J]. *海洋科学*, 2021, 45(2): 22 – 31.
- [32] 马维, 王瑁, 王文卿, 等. 海南岛西海岸红树林软体动物多样性[J]. *生物多样性*, 2018, 26(7): 707 – 716.
- [33] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-cam: Visual explanations from deep networks via gradient-based localization[C]//2017 IEEE International Conference on Computer Vision (ICCV). October 22-29, 2017, Venice, Italy. 2017: 618-626.
- [34] 龚安, 姚鑫杰, 杜波, 等. 基于集成学习与生成对抗网络的皮肤镜图像分类方法[J]. *科学技术与工程*, 2021, 21(3): 1071 – 1076.
- [35] 王国伟, 刘嘉欣. 基于卷积神经网络的玉米病害识别方法研究[J]. *中国农机化学报*, 2021, 42(2): 139 – 145.
- [36] 孙海蓉, 潘子杰, 晏勇. 基于深度卷积自编码网络的小样本光伏热斑识别与定位[J]. *华北电力大学学报(自然科学版)*, 2021, 48(4): 91 – 98.
- [37] 何海明, 齐冬莲, 张国月, 等. 快速高效去除图像椒盐噪声的均值滤波算法[J]. *激光与红外*, 2014, 44(4): 469 – 472.
- [38] 魏书伟, 曾上游, 周悦, 等. 基于并行残差卷积神经网络的多种树叶分类[J]. *现代电子技术*, 2020, 43(9): 96 – 100.
- [39] 潘兵, 曾上游, 杨远飞, 等. 基于双网络级联卷积神经网络的设计[J]. *电光与控制*, 2019, 26(2): 57 – 61.
- [40] 张怡, 赵珠蒙, 王校常, 等. 基于 ResNet 卷积神经网络的绿茶种类识别模型构建[J]. *茶叶科学*, 2021, 41(2): 261 – 271.
- [41] 陆雅诺, 陈炳才. 基于注意力机制的小样本啤酒花病虫害识别[J]. *中国农机化学报*, 2021, 42(3): 189 – 196.
- [42] NASIRAHMADI A, STURM B, EEWARDS S, et al. Deep learning and machine vision approaches for posture detection of individual pigs [J]. *Sensors*, 2019, 19(17): 3738.
- [43] ELPELTAGY M, SALLAM H. Automatic prediction of COVID-19 from chest images using modified ResNet50 [J]. *Multimedia tools and applications*, 2021, 80(17): 26451 – 26463.
- [44] 李恒, 张黎明, 蒋美容, 等. 一种基于 ResNet152 的红外与可见光图像融合算法[J]. *激光与光电子学进展*, 2020, 57(8): 128 – 134.
- [45] RAUBER J, ZIMMERMANN R, BETHGE M, et al. Foolbox native: Fast adversarial attacks to benchmark the robustness of machine learning models in pytorch, tensorflow, and jax [J]. *Journal of Open Source Software*, 2020, 5(53): 2607.
- [46] KOIRO E, GIMPLE G, LAMMICH S, et al. Low cholesterol stimulates the nonamyloidogenic pathway by its effect on the  $\alpha$ -secretase ADAM 10 [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2001, 98(10): 5815 – 5820.
- [47] HUANG H, XU H, WANG X, et al. Maximum F1-score discriminative training criterion for automatic mispronunciation detection [J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2015, 23(4): 787 – 797.
- [48] 劳凤丹, 滕光辉, 李军, 等. 机器视觉识别单只蛋鸡行为的方法[J]. *农业工程学报*, 2012, 28(24): 157 – 163.
- [49] 李文博. 基于深度神经网络算法的眼底图像语义分割研究[D]. 上海: 上海应用技术大学, 2021. DOI: 10.27801/d.cnki.gshyy.2021.000212.
- [50] WILLIAMS H J, HOLTON M D, SHEPARD E L C, et al. Identification of animal movement patterns using tri-axial magnetometry [J]. *Movement ecology*, 2017, 5(1): 1 – 14.



## Action recognition of spoon-billed sandpipers (*Eurynorhynchus pygmeus*) based on residual convolutional neural network model

YANG Xueke<sup>1</sup>, MENG Jinchao<sup>1</sup>, FENG Yueheng<sup>1</sup>, LIN Tingting<sup>1,2</sup>, WANG Zhaojun<sup>2</sup>, LIU Hui<sup>1</sup>

(1. School of Forestry, Hainan University, Haikou, Hainan 570228, China; 2. Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China)

**Abstract:** With the widespread application of image acquisition equipment and data sharing platform, the amount of bird image data has been increasing at an unprecedented speed. How to effectively deal with such a large amount of data has become a major challenge. In recent years, convolutional neural network has shown strong practicability and effectiveness in the application of automatic bird image processing. However, there has been no research on automatic recognition of movements in wild birds. In view of this, a special action image dataset of the sandpiper was established based on field images. The dataset was composed of nine action tags representing the main behavior patterns of spoon-billed sandpipers (*Eurynorhynchus pygmeus*). At the same time, three residual convolutional neural network models, ResNet50, ResNetT101 and ResNet152, were used to automatically recognize the movements of the spoon-billed sandpipers. The experimental results showed that the three models achieved excellent results in action recognition with their accuracy rates of the test set being 96.90% (ResNet50), 96.94%(ResNet101) and 96.90% (ResNet152), respectively. This indicates that these three models have a rapid recognition of the movements of the spoon-billed sandpiper.

**Keywords:** residual convolutional neural network; bird image; movement recognition; spoon-billed sandpiper

(责任编辑:叶 静)