

# 基于图像的野生动物检测与识别综述<sup>①</sup>

柯 澳, 王宇聪, 胡博宇, 林 琦, 李 勇, 双 丰

(广西大学 电气工程学院, 南宁 530004)

通信作者: 李 勇, E-mail: [youlong197157@163.com](mailto:youlong197157@163.com)



**摘 要:** 野生动物监测对于野生动物保护和生态系统维护至关重要, 而野生动物的检测与识别是实现监测的核心技术. 近年来, 随着计算机视觉技术的迅速发展和广泛应用, 基于图像的非接触式方法在野生动物监测领域引起了广泛的关注, 研究人员提出了各种方法来解决该领域的不同问题. 然而, 野外环境的复杂性使得对野生动物进行精确检测和识别仍具有一定的挑战. 为了推动该领域的研究, 本文对现有的基于图像的野生动物监测方法进行了综述, 主要包括 3 个部分: 野生动物图像获取方法、野生动物影像预处理方法以及野生动物检测与识别算法. 文章按照图像数据集和野生动物检测与识别算法的不同处理机制对这些方法进行了探讨和分类. 最后, 本文对基于深度学习的野生动物监测研究热点与存在问题进行了分析和总结, 并对未来的研究重点提出了展望.

**关键词:** 野生动物; 监测; 目标检测; 图像分类; 综述; 机器视觉

引用格式: 柯澳, 王宇聪, 胡博宇, 林琦, 李勇, 双丰. 基于图像的野生动物检测与识别综述. 计算机系统应用, 2024, 33(1): 22-36. <http://www.c-s-a.org.cn/1003-3254/9369.html>

## Review on Image-based Wildlife Detection and Recognition

KE Ao, WANG Yu-Cong, HU Bo-Yu, LIN Qi, LI Yong, SHUANG Feng

(School of Electrical Engineering, Guangxi University, Nanning 530004, China)

**Abstract:** Wildlife monitoring is essential for wildlife conservation and ecosystem maintenance, and wildlife detection and identification is the core technology to achieve monitoring. In recent years, with the rapid development and widespread application of computer vision technology, image-based non-contact methods have attracted extensive attention in the field of wildlife monitoring, and researchers have proposed various methods to solve different problems in this field. However, the complexity of wild environment still poses challenges for accurate detection and identification of wildlife. In order to promote research in this field, the existing image-based wildlife monitoring methods are reviewed in this study, which mainly include three sections: wildlife image acquisition methods, wildlife image preprocessing methods, and wildlife detection and recognition algorithms. These methods are discussed and classified according to the different processing mechanisms of image datasets and wildlife detection and recognition algorithms. Finally, the research hotspots and existing problems of wildlife monitoring based on deep learning are analyzed and summarized, and the prospect for future research priorities is proposed in the study.

**Key words:** wildlife; monitoring; object detection; image classification; review; machine vision

野生动物资源是生物资源的重要组成部分, 但全球野生动物数量在不到 50 年内减少了 68%<sup>[1]</sup>, 保护与

合理利用野生动物资源对可持续发展至关重要. 科学有效地野生动物监测能够实时获取野生动物的信息,

① 收稿时间: 2023-07-24; 修改时间: 2023-08-21; 采用时间: 2023-08-24; csa 在线出版时间: 2023-11-24  
CNKI 网络首发时间: 2023-11-28

包括图像、地理位置和环境条件,通过分析这些信息可以了解野生动物的种类、数量和栖息地状况,为野生动物保护提供依据。然而,野生动物具有分布范围广、种群密度低、行为不定以及对干扰的敏感性高等特点,给监测带来了很大的困难<sup>[2]</sup>。

野生动物监测方法主要分为接触式法与非接触式法。接触式法指在监测过程中人会与野生动物产生接触,该类方法主要包括人工调查法(包括诱捕计数法、集群计数法等)和全球定位系统(global positioning system, GPS)项圈技术,非接触法主要包括基于声音检测的方法与基于影像分析的方法。人工调查法是一种传统的调查方法,通过样线、样方调查等方法搜集动物实体或动物出现过的证据,由于多数动物善于奔走,活动隐秘性高,往往需要投入大量的人力才能保证结果的可靠性。另外,大规模的人力调查不可避免地会干扰动物的正常生活,而且给动物佩戴GPS项圈也难免会对动物的生活造成一定影响。对于非接触式法,声音检测包含的信息有限,基于声学处理的监测方法应用局限较多,而基于影像的野生动物监测具有全天候监测、成本较低且信息量大等明显优势。近年来,随着计算机视觉、机器学习和深度学习等技术的快速发展,基于图像的野生动检测成为研究的热点问题。

野生动物的监测需要获取广域森林中长时间的数据,因此基于图像的野生动物监测存在图像与视频数据量大且图像质量不高等问题。要实现野生动物的自动监测,需解决快速自动检测、识别和高鲁棒性的问题。虽然目前已存在大量目标检测的算法,但针对野生

动物检测设计的算法较少,并且在基于图像的野生动物检测与识别领域缺乏综述和相关方法的总结归纳。为了推动未来的研究,本文结合最新的研究成果,对不同类型野生动物监测方式进行了对比分析,总结了现有的野生动物图像数据集,并介绍了与野生动物影像预处理相关的方法。然后对现有基于图像的野生动物检测与识别相关研究进行了系统的总结归纳,主要包括:基于机器学习和基于深度学习的检测和识别方法。最后,针对基于图像的野生动物检测与识别方法存在的问题进行了总结和展望。

## 1 野生动物影像获取方法及数据集对比

图1展示了野生动物监测的各种方法。其中,野生动物影像的获取方式主要包括遥感卫星、红外触发相机、车载无人机系统(unmanned aircraft system, UAS)和无线图像传感器网络等。表1给出了4种野生动物影像获取方法的对比分析。



图1 常用的野生动物监测的方法

表1 常用的野生动物监测方法对比分析

监测方法	图像获取平台	相机类型	优点	缺点
人工调查 <sup>[3,4]</sup>	—	—	具有获取信息较为全面,数据来源可靠,依靠研究人员在调查过程中收集粪便、羽毛等与野生动物直接相关的信息;能够实现区域的经度、纬度、海拔高度和生境特征等信息的记录	危险性高、难以全天候监测、劳动强度大、监测周期长、主观性强、监测范围有限等
GPS项圈技术 <sup>[5]</sup>	—	—	精确度较高、监测方便、实时追踪动物的地理坐标位置、便于研究各种生活习性、具有警示作用	需给动物带上项圈,对动物并不友好,工作量较大,不能获取动物影像信息,电池具有不稳定性,信号易受环境影响,佩戴GPS项圈会给动物的生活带来不便
声学处理 <sup>[6]</sup>	—	—	能在长期和多变的气候条件下对野生动物实时监测,特别适合对于研究昆虫的密度分布和多样性的研究	声音包含的信息有限,且自然环境中噪声较多,难以从声音中提取足够信息
影像处理	遥感卫星	遥感相机	无接触式,覆盖范围广、目标数量多,更便于分析种群居变化等,包含更丰富的全局信息	卫星监测的灵活性较差、数据精细度较低;卫星/无人机远距离拍摄的照片区域大,只有少数包含动物,动物区域较小,且其颜色容易与背景植物和土壤平滑融合。在大量图像中识别和计数单个动物非常复杂且费时,难以精确测量局部微观信息
	车载 <sup>[7]</sup> / 无人机(UAS) <sup>[8-12]</sup>	红外相机/ CCD(charge-coupled device)	无人机可以在短时间内覆盖大面积区域,且可以在人迹罕至的偏远地区使用,成本相对低,更易部署,对于采集信息的人类更加安全	

表1(续) 常用的野生动物监测方法对比分析

监测方法	图像获取平台	相机类型	优点	缺点
影像处理	固定式红外相机 <sup>[13,14]</sup>	红外触发相机	可近距离细粒度监测, 通过和红外补光灯的协作, 可以实现全天候自主检测, 监测时间长、监测信息广等	数据量大, 需大量时间区分误触发图像和野生动物图像、成本较高、周期较长
	无线图像传感器网络 <sup>[15,16]</sup>	红外相机/CCD	具有红外相机的特地, 采集的图像、视频等丰富的媒体信息可以通过无线远程实时传输, 实现监测目标的无线远程、实时、精准监测	其具体应用受自身处理器、电源续航等限制, 拍摄时会受环境遮挡的影响

在卫星遥感图像中, 个体动物通常没有明显的外形特征, 只表现为一些具有不同亮度或灰度值的像素点, 如图1(c)所示. 此外, 卫星图像的空间尺度极大, 因此主要用于估计野生动物的种群规模和空间分布特征<sup>[17]</sup>.

车载相机固定在车身上, 而野外的路面状况不利于车辆行驶, 且车辆的颠簸会使相机震动而影响图像质量, 如图1(d)所示. 由于无人机不受地形地貌的影响, 且拍摄视野较地面更为开阔, 因此目前应用较多的是基于无人机和无线图像传感器网络的监测方法, 如图1(e)所示, 可实现对野生动物更准确地检测. 无人机通常搭载光电或红外相机, 能提供较高分辨率的数据, 且相比于人工驾驶的飞机, 具有更低的成本和风险, 但无人机也有一定局限, 例如续航时间较短. 基于无人机的监测系统, Linchant 等人<sup>[18]</sup>给出了多种系统的详细参数和对比分析.

无线图像传感器网络和红外相机因其显著的优势而广泛应用于野生动物的监测, 如图1(f)所示, 无线图像传感器网络由多个具有无线通信功能的传感器组成. 理论上, 卫星和无人机传感器也可以构建无线传感器

网络, 但在野生动物监测中, 组成该网络的传感器大部分是部署在地面的红外触发相机或运动触发相机等.

近年来, 机器学习和深度学习在图像领域得到了广泛的应用, 此类方法都需要有标签的图像作为训练集对模型进行训练, 因此, 基于图像的野生动物检测与识别需构建有效的野生动物图像数据集. 研究人员构建了大量的 RGB 图像和红外图像的数据集用于野生动物的检测、识别和分类等模型学习与算法训练, 常见的野生动物图像数据集对比及相关信息如表2所示.

这些数据集由不同的传感器获取, 包括红外触发相机、CCD相机、红外相机、运动触发相机以及各种型号的高清相机. 其中, 红外触发相机和运动触发相机在有动物进入相机视野内时会自动触发拍照, 也被称为相机陷阱(camera-trap), 因此常被用于野外长周期的监测任务, 红外相机能相对清晰地拍摄到夜视场景. 数据集的来源主要有3种: 研究人员专门拍摄的数据集; 纪录片或者其他视频资料或从中截取的图像; 网络中的图像资料.

表2 常见的野生动物图像数据集对比分析

数据集名称	图像数量	种数	图像获取方法	说明
Snapshot serengoti <sup>[19,20]</sup>	120万	48	红外触发相机和CCD	图像源于坦桑尼亚塞伦盖蒂国家公园(DLC Covert Reveal和SG565混合拍摄), 物种的样本数高度不平衡, 哺乳动物种类丰富. <a href="http://www.snapshotserengeti.org">www.snapshotserengeti.org</a>
DS1数据集 <sup>[21]</sup>	30167	6	无线图像传感器和红外相机	图像源于内蒙古赛罕乌拉保护区, 彩色图像大小大多为2560×1920, 包含黑夜和白天的图像, 类别相对均衡
Costa rica camera trap dataset <sup>[22]</sup>	300000	12	运动触发相机	图像源于Costa Rica地区, 数据包括10种野生动物、人与误触发图像, 类别分布非常不均衡. <a href="https://mtimm100.github.io/cameraTrap.html">mtimm100.github.io/cameraTrap.html</a>
JaguarID <sup>[22]</sup>	176	16	CCD	图像一些从动物园拍摄, 一些来源于Flickr, 图像的质量从低到高不等, 图像的颜色、灰度、运动模糊程度有所不同
iNaturalist <sup>[23]</sup>	670000	5089	CCD	图像的质量较高, 图像聚焦在感兴趣的动物上, 动物通常位于图像的中心, 光线良好, 并且没有任何遮挡. <a href="https://iNaturalist.org">iNaturalist.org</a>
CUB-200-2011 <sup>[24]</sup>	11788	200	CCD	图像源于北美洲, 加州理工学院鸟类数据集包含15个部件位置, 312个二进制属性, 1个边界框. 图像的背景区域较少, 图片质量相对较高. <a href="http://www.vision.caltech.edu/visipedia/CUB-200-2011.html">http://www.vision.caltech.edu/visipedia/CUB-200-2011.html</a>
Wildlife surveillance <sup>[25]</sup>	8368	6	红外相机和CCD	图像拍摄于英国, 包含农场建筑中的野生和家养动物: 猫、鸟、狗、狐狸、大鼠和兔子, 图像的质量相对较高

表 2 (续) 常见的野生动物图像数据集对比分析

数据集名称	图像数量	种数	图像获取方法	说明
Kuzikus UAV datasets <sup>[26]</sup>	3098	超过 20种	佳能PowerShot S110 (UAV)	图像由紧凑型相机安装在eBee上的相机航拍摄于Botswana, South Africa and Namibia, 图像大小大于4000×3000, 图像中包含多种目标, 存在大量小目标, 可以用于检测有无动物的区域, SAVMAP consortium. <a href="https://lasig.epfl.ch/savmap">https://lasig.epfl.ch/savmap</a>
WVDDS (wildlife video detection datasets) <sup>[27]</sup>	6601	12	高清相机	来源于网络的视频图像, 具有时间以及空间序列关系, 同一图像中存在包含多个目标的情况, 用于目标检测
WLD (wildlife documentary) dataset <sup>[28]</sup>	747000	65	高清相机	来源于YouTube上的15个纪录片, 超过 4000个目标, 具有视频影像的优势. <a href="https://github.com/hellock/WLD">https://github.com/hellock/WLD</a>
AWA2数据集 <sup>[29,30]</sup>	37322	50	CCD	来源于Flickr, Wikipedia等, 85个二进制和连续类属性, 平均每个类别包括746个图像. <a href="http://cvml.ist.ac.at/AwA2">http://cvml.ist.ac.at/AwA2</a>
NACTI-64 <sup>[31]</sup>	29899	22	高清相机和红外相机	来源于NACTI (North American camera trap images) <sup>[32]</sup> , 图像大小归一化为64×64的大小. <a href="https://pan.baidu.com/s/1-q-TLXrecGxHfY_6gksz3Q">https://pan.baidu.com/s/1-q-TLXrecGxHfY_6gksz3Q</a>
Wildlife spotter dataset <sup>[33]</sup>	125621	15	高清相机和红外相机	图像源于澳大利亚中南部的维多利亚, 彩色和灰度设置均为1920×1080或2048×1536分辨率, 具有15种类别的单标签, 类别分布不均衡. <a href="https://wildlifespotter.net.au">https://wildlifespotter.net.au</a>
Far-eco dataset <sup>[34]</sup>	20600	14	高清相机和红外相机	图像拍摄于中国西南地区的4个自然保护区, 包含大熊猫、金丝猴等多种珍稀动物, 图片质量相对较高

## 2 野生动物影像预处理方法

基于图像的目标检测算法发展迅速, 其中数据集的质量常常影响算法在实际中的应用效果, 由于野生动物检测图像数据来源于无人控制的相机陷阱, 而野外的环境复杂多变, 例如由光照变化引起的图像照度不足或过度曝光、大雾天气下能见度较低等因素, 监测图像的质量经常受到严重影响, 此外, 一些珍贵动物数量稀少, 很难拍摄足够数量的样本用于训练.

当数据集质量较差或数据量有限时, 可采用数据增强方法来改善图像质量和扩展数据集. 图 2 总结了几种常用的数据增强方法, 包括旋转、裁剪、缩放

等, 可对采集的图像进行空间层面的变换. 在数据集较小的情况下, 可使用数据集扩充的方法, 此类方法通常包括 3 种: 模拟器构造虚拟数据集、将旋转后的图像拼接成新的数据集, 以及利用生成对抗网络进行无监督学习生成新样本. Ferreira 等人<sup>[35]</sup>设计了一个野生动物 3D 模拟器, 能够产生 4 种野生动物的三维模型, 并将其放置在不同的场景中, 此外, 该模拟器还可以设置不同程度的遮挡和动物运动状态的信息, 研究人员将模拟器生成的数据与真实数据混合训练, 以提高模型在分类与分割方面的性能表现, 如图 3(a) 所示.

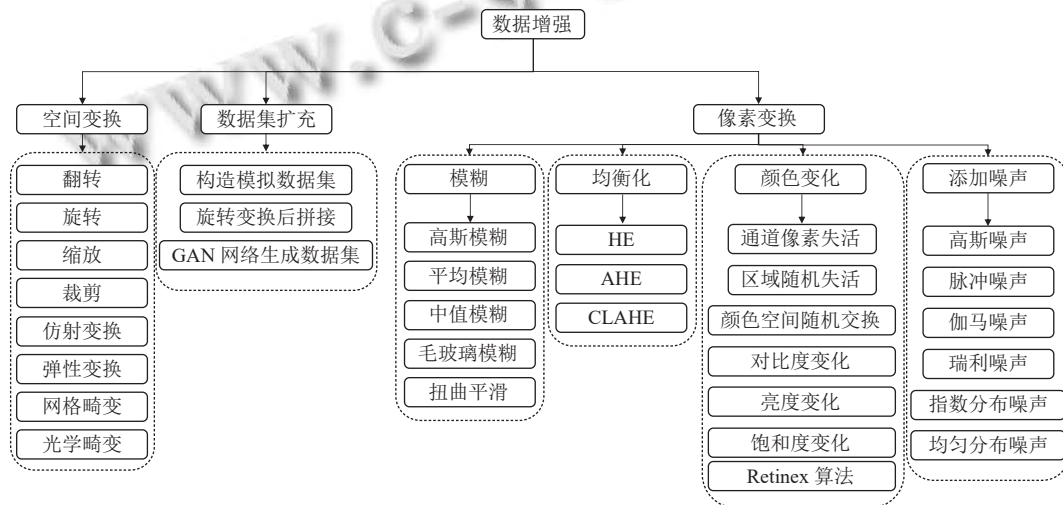


图 2 野生动物检测识别数据增强的方法

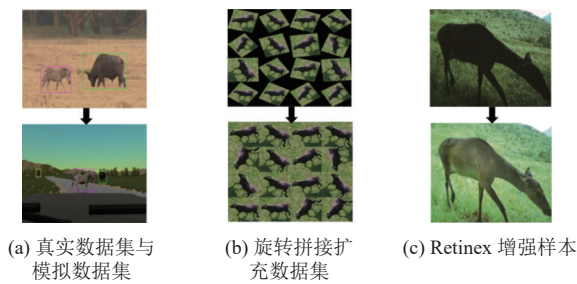


图3 图像增强实例

模拟器的方法能够便捷生成大量数据,从而降低研究成本。然而,模拟图像通常在光线和纹理细节方面与实际照片存在差异,因此需要进一步提高模拟数据集的实用性。Okafor等人<sup>[36]</sup>提出了一种在操作分类阶段的方法,将一张图像转换为包含多组原始图像旋转副本的新图像,步骤如下:首先,创建一个 $n \times n$ 单元的网格,每个单元包含一组不同的随机旋转图像;然后将这些旋转图像组合到新创建的图像中,并添加自然背景。该算法用于生成新的训练和测试图像,以增强图像中的信息量,如图3(b)所示。

生成对抗网络是一种非监督学习的网络,通过生成器和判别器之间的对抗学习过程,此方法可带来更多丰富的数据变化,并提高图像质量。像素变换通常包括4种算法:图像模糊处理、均衡化、颜色变化和添加噪声,这些方法可以处理图像,引入各种变换,并改善图像的视觉效果。曾陈颖<sup>[37]</sup>利用先验知识确定图像中的颜色集,生成颜色分布直方图,再进行差分归一化增强图像。Hong等人<sup>[38]</sup>提出的DAGC方法采用基于加权累积密度函数估计自适应参数Gamma-correction方法来处理HSV颜色空间值带上的雾。张军国等人<sup>[39]</sup>提出基于Retinex理论的光照自适应图像增强方法,该方法首先利用基于复合梯度的引导滤波估计图像照度分量,以克服光照突变造成的伪光晕现象;并采用基于Otsu阈值的对比度自适应拉伸方法对照度分量进行校正,以解决传统算法过度增强的问题;最后,利用照度分量单通道图像计算反射分量图像,以保证色彩的真实性,如图3(c)所示。

### 3 野生动物检测与识别算法

野外监测图像存在多种挑战,如复杂背景、遮挡情况不明、野生动物位置和大小不定、不同物种图像数量不均、以及视觉特征相似等情况,给野生动物的检测与识别带来了困难,随着图像数据量不断增长,研

究人员迫切需要研究出高效准确的检测识别算法用于实际应用中。

针对野生动物的检测算法,按照特征的类型分为低层、中层和高层特征。在基于图像的野生动物检测与识别中,最关键的步骤是特征的提取,常用的特征包括灰度分布直方图<sup>[7]</sup>、HSV颜色分量<sup>[40]</sup>、几何特征、边缘特征、SIFT (scale-invariant feature transform)、cLBP (compound local binary pattern)<sup>[41]</sup>、HOC (histogram of colors)<sup>[38]</sup>和 Haar-like<sup>[42]</sup>等低层特征; BoVW (bag-of-visual-words)<sup>[43]</sup>、HOG (histogram of oriented gradients)<sup>[44]</sup>特征等中层特征;以及基于卷积神经网络的高层特征。

野生动物检测与识别算法也可以分为有学习和无学习两种方法。早期的图像处理大多是无学习机制的方法。如Gonzalez等人<sup>[45]</sup>基于无人机获取的图像,利用阈值和二值模板匹配的方法进行动物检测和追踪等任务,这类方法通常基于先验条件或模板匹配等方式来判断野生动物区域及类别,但鲁棒性较差,对于复杂多变的环境效果并不理想。有学习机制的方法按照学习机制的不同分为基于机器学习与深度学习的方法。此外,由于野生动物监测的特殊性,其样本图像较少,有研究者提出了零样本学习 (zero-shot learning, ZSL) 和广义零样本学习 (generalized zero-shot learning, GZSL) 的方法,用来对训练集中未出现过的类别进行预测。

#### 3.1 基于机器学习的方法

在基于图像的野生动物检测机器学习方法中,常见方法主要分为基于无监督学习和有监督学习的方法。

##### 3.1.1 基于无监督机器学习的方法

如图4所示,基于无监督学习的目标检测和识别方法与基于先验条件或模板匹配的方法相似,两种方法都涉及数据处理流程,不同的是无监督学习方法使用了Mean Shift等分类器,让模型自动在没有标签的数据集中找到规律。无监督学习中常用的图像分割方法包括阈值分割、边缘分割、直方图法和小波变换法等。阈值分割法通过度量图像灰度值,设置不同类别的阈值来实现图像分割;边缘分割法通过检测图像中灰度值发生跳变的边缘来实现分割;直方图法通过建立图像的颜色直方图,并利用直方图的波峰和波谷来表示颜色值的范围,来达到分割的目的;小波变换法则通过在小波域中最小化关于变量 $u$ 和 $v$ 的变分泛函,达到分割图像的目的。

在图像分割的同时,通常还会进行特征提取,如非负矩阵分解(NMF),然后采用主成分分析(principal component analysis, PCA)或t-SNE对特征进行降维,最后采用聚类方法对图像进行分类,常见的聚类算法包括K均值聚类(K-means)、K近邻聚类(K-nearest neighbor, KNN)、凝聚聚类和DBSCAN(density-based spatial clustering of applications with noise)等。例如,Christiansen等人<sup>[46]</sup>使用热特征参数和KNN分类器

对机载红外图像进行动物区域和非动物区域的检测。Seymour等人<sup>[10]</sup>将热成像图作为输入,选中图像中温度超过一定阈值的点标记为海豹,然后对选中的一些聚集点进行扩充,形成多边形,根据多边形的大小和内部平均温度,可区分成年海豹和幼年海豹群体,并使用高通滤波器将点簇分割成海豹个体,以统计海豹的数量。这种无监督方法无需训练数据,算法相对简单,但鲁棒性较差,精度并不理想,并且适用范围有限。

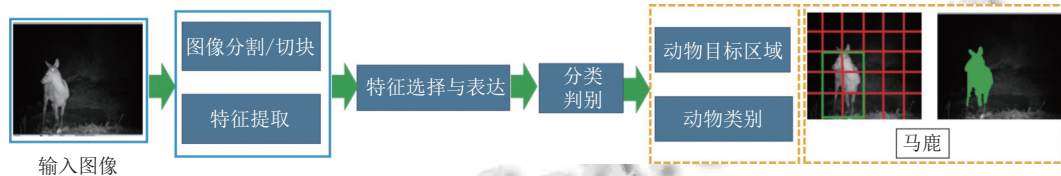


图4 基于无监督学习的野生动物检测与识别框架

### 3.1.2 基于有监督机器学习的方法

有监督学习用已知标签样本进行训练或构建识别模型,识别精度和鲁棒性均优于无监督学习方法。在基于有监督学习的野生动物目标检测与识别算法中,训练阶段需要将一定数量带有标注的图像输入到模型中。首先,提取图像的低中层特征,再利用BoVW、LDA(linear discriminant analysis)等特征处理方法提取更高层次的特征,进一步将特征输入到分类器中,通过将分类器输出结果与标签对比来训练分类器。在测试阶段,

分类器根据输入图像提取的特征直接输出结果,其算法框架如图5所示。有监督学习方法中常见的分类器有SVM(support vector machine)、RF(random forest)和AdaBoost等。

根据野生动物在图像中的目标大小,可以将处理的图像分为野生动物区域占比较大的图像和较小的图像,此方法有助于针对不同尺寸的目标采取不同的处理方法和算法,以更好地进行野生动物的检测与识别。

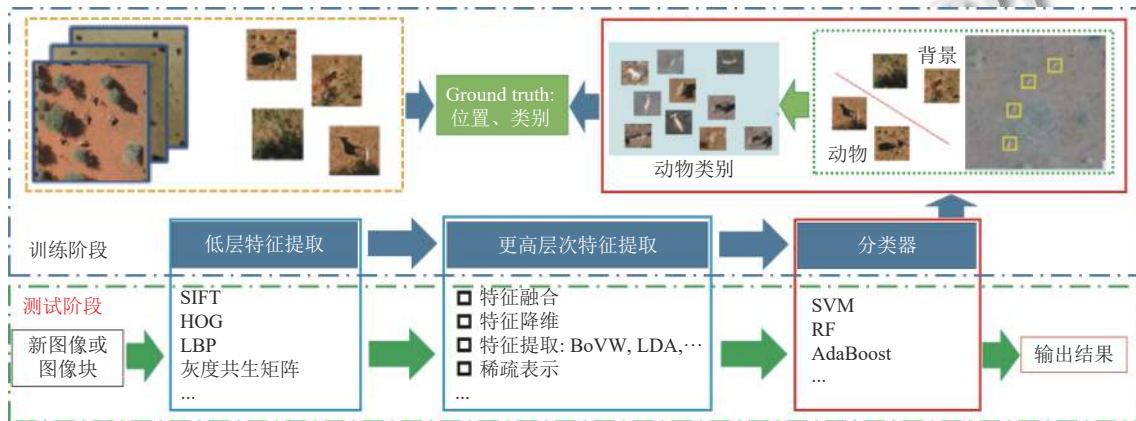


图5 基于有监督机器学习的野生动物目标检测与识别算法框架

#### (1) 野生动物区域占比较大的图像

对于图像中野生动物目标区域较大的图像,通常需要提取图像的大量特征,其中包括低层特征(如颜色、纹理、形状等)和高层特征(如卷积神经网络中提取的深层特征),并利用提取的特征及其对应的标签训练分类器,如支持向量机(SVM)、随机森林(RF)和深

度学习模型。训练过程中,分类器会学习图像特征与其对应的类别之间的关系,从而对新的图像进行分类。最后,利用训练好的分类器对提取的特征进行分类,实现野生动物的目标检测与识别。这种方法可以根据图像特征的多样性和对应的标签进行较准确的分类,提高野生动物检测与识别的精度和鲁棒性。例如, Kamencay

等人<sup>[47]</sup>通过对图像进行预处理,然后基于 PCA、LDA 和 LBPH (local binary patterns histograms) 等方法提取图像的特征,最后利用分类器进行分类. Qiao 等人<sup>[40]</sup>利用 HSV 颜色空间的特征和鸟类的眼睛与鸟喙之间距离以及鸟喙宽度距离的比率等融合特征,利用 SVM 和决策树实现 15 种鸟类的分类. 这些方法利用一些低层特征和有效的分类器进行分类,但精度还有待提高. 此外,这些算法主要适用于图像中只包含一个目标且背景区域较少的情况,且通过相机获取的野生动物监测图像中野生动物往往不处于图像中心位置,针对这类情况, Yu 等人<sup>[42]</sup>首先通过人工裁剪图像中的野生动物目标区域,并利用 ScSPM (sparse coding spatial pyramid matching) 的方法获取更高层次的特征,进而实现 18 种野生动物的分类.

### (2) 野生动物区域占比较小的图像

在航拍或卫星图等遥感图像中,画面覆盖范围大,而野生动物占比较小,如图 6(d) 中所示. 在该类图像中提取野生动物的特征有一定的困难,通常的做法是通过裁剪图像中多余部分来限定野生动物的有效区域,而人工裁剪需要耗费大量人力资源. 针对此类问题,研究人员提出了野生动物区域检测和兴趣域提取的方法,此类方法通常先将图像分割成多个图像块,然后以每个图像块为处理单元进行分类任务. 例如, Rey 等人<sup>[26]</sup>提取图像的 HOC 特征,并构造更高层次的 BoVW 特征,然后基于 EESVM (ensemble of exemplar support vector machine) 分类器实现野生动物的有效检测,从而能够有效地获取野生动物的区域. Feng 等人<sup>[48]</sup>利用图像平滑和降噪处理,通过边缘检测等方法获取目标的显著性区域. Alshawa 等人<sup>[44]</sup>在均匀间隔的密集网格上计算 HOG 特征,并使用重叠的局部对比度归一化来提高精度,然后用 SVM 分类器进行分类,从而区分出每个单元是背景还是动物.

## 3.2 基于深度学习的方法

传统的基于人工特征提取的方法需要设计特征并使用分类器进行分类,但存在特征不全面、判别性差和设计难的问题. 深度学习在图像领域的广泛应用,通过卷积神经网络自动提取强大的特征,因此成为野生动物检测与识别主要研究方向. 本节将重点讨论基于深度学习的方法在目标区域相对较大与较小情况下的应用. 此外,还介绍了零样本学习 (zero-shot learning,

ZSL) 和广义零样本学习 (generalized zero-shot learning, GZSL) 的方法.



图 6 不同类型的网络输入图像

### 3.2.1 目标区域相对较大的情况

#### (1) 基于整幅图像的野生动物分类/识别网络

针对目标区域较大的图像,如图 6(a) 所示,其中野生动物区域在图像中占比较大,而且背景区域复杂度低,此类情况通常直接使用网络提取的深度特征作为野生动物的特征,并利用这些特征进行野生动物的识别. 例如, Okafor 等人<sup>[49]</sup>利用灰度和颜色特征与卷积神经网络 (如 AlexNet 和 GoogLeNet) 构建了分类模型,实现了野生动物的分类. Norouzzadeh 等人<sup>[50]</sup>使用 AlexNet、GoogLeNet、VGG 和 ResNet 系列网络模型,实现了野生动物物种的识别和计数等任务. Trammell 等人<sup>[23]</sup>在 iNaturalist 数据集上采用多种卷积神经网络进行了野生动物分类. 此外, Tang 等人<sup>[51]</sup>提出了一种基于半监督学习的方法,利用卷积神经网络提取特征,并通过视觉和语义知识迁移的方法整合了有关目标相似性方法,实现了目标的有效分类. Favorskaya 等人<sup>[31]</sup>设计了一种名为 DANAS 的新型轻量级网络,该网络在具有低分辨率和保持纵横比的相机陷阱图像上进行训练,使用损失函数进行搜索引导,其超参数根据理论推导得出. 这些网络主要是针对图像中背景区域较少的情况,其分类精度会受到背景的复杂程度影响.

#### (2) 基于感兴趣区域的野生动物分类/识别网络

图像中野生动物监测系统获取的图像通常含有较多的背景区域,直接对这些图像进行分类的精度并不理想,如图 6(b) 所示. 针对复杂背景环境的影响,为提高分类精度,研究者首先采用特征提取的方法来截取包含野生动物的感兴趣区域 (region of interest, ROI),然后使用分类网络对 ROI 实现了野生动物分类,图 7 展示了这类方法的一般过程.

常见的 ROI 提取方法有 RPN (region proposal

network), GA-RPN 和 Adaptive-RPN 等. RPN 是使用全卷积网络实现与检测网络共享整幅图像的卷积特征, 从而快速生成候选区域. GA-RPN 通过特征指导锚点框中心位置和宽高比的生成, 提高了生成锚点的效率, 对于一些具有极端长宽比的物体非常有效. Adaptive-RPN 选取交并比损失 (IoU Loss) 作为损失函数, 重点优化交并比, 以对抗对尺度敏感性. 在检测到目标区域后, 使用 CNN 网络对目标区域进行分类, 背景分割采

用编码译码的结构, 常见的背景分割网络有 FCN<sup>[52]</sup>, U-Net<sup>[53]</sup>, SegNet<sup>[54]</sup>等. 为了提高识别准确率并实现端到端的野生动物识别, Xie 等人<sup>[15]</sup>提出了一种名为 SE-ResNet 网络, 该网络基于 SENet 的注意力机制和 ResNet 的特征提取网络, 用于对 5 种野生动物进行分类. Timm 等人<sup>[22]</sup>使用 InceptionV3 模型在 costa rica camera trap dataset 上实现了监测图像中是否包含野生动物的判断和对 10 种野生动物的分类, 其准确率达到了 79%.

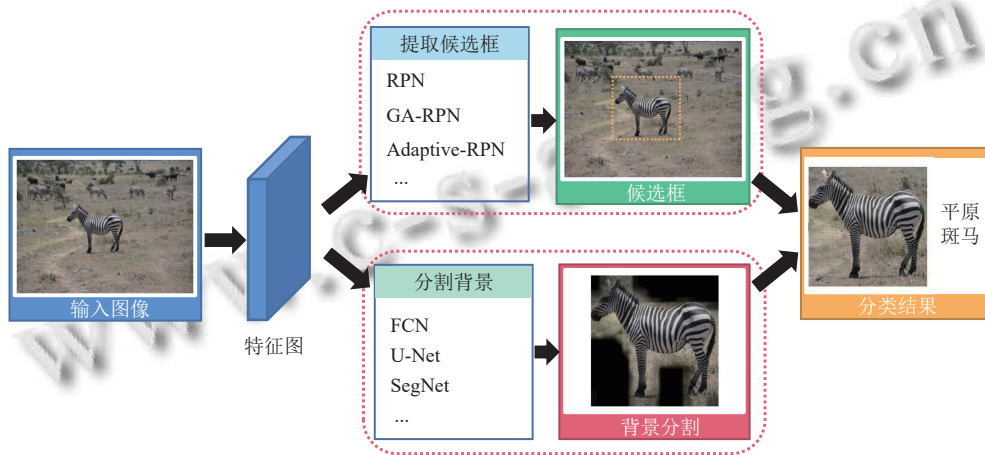


图7 基于兴趣域的野生动物分类与识别过程

由于陷阱相机经常会被误触发, 导致收集的数据集中的部分图像不包含所需要检测的野生动物种类. 对于此类问题, Parham 等人<sup>[55]</sup>首先设计了一个 CNN 作为图像过滤器, 用于过滤掉不包含兴趣域的图片. 在分割出动物区域后, 通过设计一些判决条件 (例如个体动物是否存在遮挡、是否位于图像中心等), 解决 annotations of interest (AoI) 任务, 以提高某些图像处理的优先级, 通过以上两个方法, 可以减少整体处理时间. 最后, 使用 YOLO 对图中的感兴趣的单个动物进行分类. 在野生动物监测中常使用无线图像传感器网络来实现远程实时监测. Feng 等人<sup>[56]</sup>通过 CNN 网络提取图像中的显著区域, 并基于 SPIHT 编码将显著区域置于传输的最高优先级. 在接收到图像后, 采用添加了自注意力机制的 SA-ResNet152 网络对野生动物进行框选和标注, 实现对 6 种野生动物的分类.

### (3) 基于图像的野生动物检测与识别网络

如图 6(a) 和图 6(b) 所示, 野生动物研究的重点是关注野生动物所在的区域. 因此, 在进行野生动物位置检测的同时, 还需要对野生动物进行识别, 实现端到端的目标区域获取和野生动物分类. 这类网络通常包含

回归和分类两个任务, 基于深度学习的目标检测与识别网络主要分为两种框架: Two-stage 系列和 One-stage 系列. One-stage 框架直接预测物体的类别概率和位置坐标值, 准确度相对较低但速度更快; Two-stage 框架通过 RPN 网络生成区域建议, 再对区域位置进行精细调整, 以得到更准确的目标框, 但速度相对较慢. 图 8 展示了这两类框架的流程图, 其中, conv 代表卷积, RPN 为区域生成网络, feature map 为卷积后的特征图. 近年来 One-stage 框架的算法如 SSD 和 YOLO 等系列在精度和速度上均取得了不错的效果, 因此这些网络得以改进并应用到野生动物的检测. 例如, 宫一男等人<sup>[57]</sup>采用了基于 DarkNet-53 骨干网络的 YOLOv3, 通过微调方法实现了东北虎豹国家公园区域的 RGB 图像和红外图像的 8 种野生动物物种的检测与识别; 刘文定等人<sup>[58]</sup>提出了基于全局-局部的双通道野生动物自动识别方法, 首先利用 DarkNet-19 将野生动物所在区域从红外监测图像中分割出来, 建立 ROI 图像数据库, 再将原始图像和 ROI 图像输入基于 VGG16 的双通道卷积神经网络进行特征提取和融合, 同时实现了野生动物区域的检测和识别, 并且获得了较高的准确度. Song



等人<sup>[34]</sup>通过在将由相机捕获的原始图像数据制作成数据集的过程中,在统一的数据维度上应用简单线性迭代聚类(SLIC)超像素分割方法,将图像快速分割成多个具有语义意义的子区域.与像素相比,超像素更有利

于提取结构信息和局部特征.为了解决长尾数据问题,蔡前舟等人<sup>[59]</sup>提出了一种基于两阶段学习和重加权相结合的长尾数据解决方法,采用基于迁移学习的两阶段方法来训练网络.

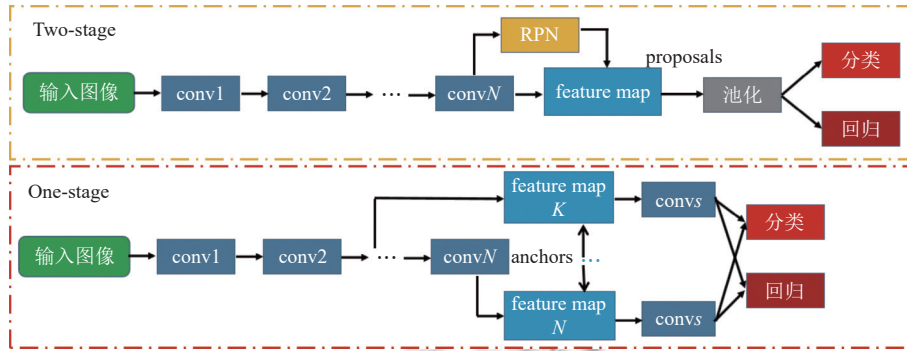


图8 不同的目标检测与识别网络框架

#### (4) 基于视频的野生动物检测与识别网络

由于一些野生动物监测的图像是由视频帧构成的,为了提高野生动物检测与识别的准确率并实现野生动物的实时动态识别,研究者提出了基于视频的野生动物检测与识别方法.如图6(c)所示,根据野生动物监测的视频中,相关联的两帧图像可以便于去除背景区域、获取更多目标信息,从而更有利于野生动物的检测和识别.例如,陈建促等人<sup>[27]</sup>利用互信息熵来表征连续图像帧之间的关系,并构造图像帧的相关因子.然后利用YOLOv3模型提取每帧图像的特征图,从而构建了Context-aware YOLO模型,实现了视频中野生动物的检测与识别;杜科<sup>[21]</sup>则基于监测节点拍摄图像之间的相关性,利用注意力机制、InceptionV4和GRU构建了一种基于Attention-CNN-GRU的野生动物分类网络.这些网络均实现了较高精度的野生动物检测与识别.此外,为了解决野生动物监测图像中常见的遮挡问题,王越等人<sup>[60]</sup>在视频监测图像中利用视频帧的时间序列关系,利用熵、运动偏移量和YOLOv3模型实现了野生动物的检测与识别.

#### 3.2.2 目标区域较小的情况

上述的研究大多是针对图像较小、目标占比较大的图像.然而,基于无人机/遥感卫星的野生动物监测图像通常具有较大的图像尺寸,且野生动物目标较小且数量较多,目标的精细程度较低,如图6(d)所示.这使得在这类图像中进行目标标注和识别变得困难,例如,针对遥感图像的覆盖范围较大、目标区域较小且目标数量较多等问题,Kellenberger等人<sup>[61]</sup>利用如图9所示的端到端网络进行大型野生动物的检测.

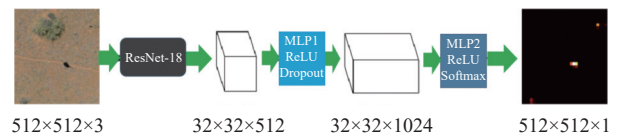


图9 文献[62]中的动物目标检测网络框架

该网络首先利用 ResNet-18 进行特征提取,并通过微调的方式进行训练.此外,网络添加了2个多层感知机(MLP),该感知机模块包括非线性激活函数 ReLU、Dropout 正则化和 Softmax 激活函数;此外,Kellenberger等人<sup>[63]</sup>提出了一种基于主动学习策略的深度卷积神经网络,该网络基于半监督的思想,通过少量标注图像实现了对野生动物目标的高准确率检测.尽管小目标的检测与识别面临许多的挑战,但Li等人<sup>[64]</sup>和Tong等人<sup>[65]</sup>提供了在具有畸变和小目标多的大场景图像中的多种算法实验结果和网络框架的对比分析,为小目标区域较小的情况下的野生动物目标检测和识别提供了参考方案.针对野生动物分类和检测的深度学习基础网络的总结如表3所示.

#### 3.2.3 零/小样本学习

在野生动物检测与识别过程中,对于某些珍稀动物样本,其样本数量较少,甚至可能涉及从未见过的物种.对于这些样本,传统的CNN网络依赖于大量数据反复学习特征,往往难以取得良好的分类结果.零样本学习(ZSL)是一种大规模概念学习方法,可以在没有标记训练数据的情况下学习新概念.它通过从可见样本中学习特征,然后识别出未出现过的类别(不可见类),因此近年来受到了广泛的关注.

表3 深度学习算法总结

任务	图像类型	预处理	常见的基础网络
野生动物分类/识别	红外图像和彩色图像	(1) 手动提取ROI (2) 传统算法的ROI提取 (3) ROI提取网络	AlexNet <sup>[49]</sup> 、GoogLeNet <sup>[50]</sup> 、VGG系列 <sup>[56]</sup> 、ResNet系列 <sup>[63]</sup> 、SqueezeNet <sup>[66]</sup> 、MobileNet <sup>[67]</sup> 、DenseNet <sup>[68]</sup> 、ShuffleNet <sup>[69]</sup> 、Xception <sup>[70]</sup> 等
野生动物检测与识别	航拍图像(红外和彩色) 地面拍摄图像(红外和彩色)	— —	Two-stage: Faster R-CNN <sup>[71]</sup> 、R-FCN <sup>[72]</sup> 、Mask R-CNN <sup>[73]</sup> 、Cascade R-CNN <sup>[74]</sup> 等; One-stage: YOLO系列 <sup>[10,26,45,53]</sup> 、SSD <sup>[75]</sup> 、RetinaNet <sup>[76]</sup> 、M2Det <sup>[77]</sup> 、RefineDet <sup>[78]</sup> 、TridentNet <sup>[79]</sup> 、DSSD <sup>[80]</sup> 、DSOD <sup>[81]</sup> 等

为了解决样本不足的情况, Xian 等人<sup>[30]</sup>在对象检测中引入了属性层的概念, 通过定义人类指定的对象的高维特征描述(例如形状、颜色甚至地理信息), 而不仅依赖于训练图像来训练相应的特征. 这种方法在一定程度上实现了迁移学习的目的. 受标签嵌入的启发, Akata 等人<sup>[82]</sup>提出了一个结构化的联合嵌入框架(SJE), 利用多种辅助语义信息, 如传统属性、文本语义信息(如词向量等)和等级标签嵌入等, 以替代手动标注属性的过程. 为了解决ZSL中对源类的过拟合的问题, Jiang 等人<sup>[83]</sup>提出可转移对比网络(TCN), 明确地将知识从源类转移到目标类. 该方法通过自动将一个图像与不同类别进行对比来判断它们是否一致, 从而利用类别的相似性将知识从源图像转移到相似的目标类, 提高了对目标图像的分类鲁棒性. 大多数零样本学习方法需要预先提供语义属性作为中间信息, 以发现观察到的类别与未观察到的类别之间的内在关系. 然而, 在实际应用中, 对样本对象进行丰富的属性标注往往是困难的. 为了克服这一障碍, Xu 等人<sup>[84]</sup>开发了一种低秩语义分组(low-rank semantics grouping, LSG)方法, 以半监督的方式进行零样本学习,

该方法利用视觉和语义信息之间的内在关系, 并从可见类中恢复缺失的标签信息.

ZSL中的测试集由训练集中未出现过的类别样本组成. 然而, 考虑到现实数据的分布情况, 可见类通常比不可见类要多, 因此在测试阶段不出现可见类的情况非常罕见. 为了使ZSL模型更符合野生动物识别需求, Chao 等人<sup>[85]</sup>提出了广义零样本学习任务(generalized zero-shot learning, GZSL), 其中测试数据包括可见类和不可见类, 并将它们分类到所有类别的联合标记空间中. Li 等人<sup>[86]</sup>提出了一种乘积量化零样本学习(PQZSL)方法, 用于学习嵌入和量化器, 将视觉特征压缩成紧凑的编码用于近似神经网络(ANN)搜索. 该方法首先将视觉特征投影到正交语义空间中, 然后利用产品量化(PQ)对单个属性进行量化. 针对AWA数据集和CUB数据集上的ZSL和GZSL任务, 对比结果总结如表4所示, 其中, mrAcc表示平均识别精度; Top-1表示Top-1准确率; U表示不可见类的平均Top-1准确率; S表示可见类的平均Top-1准确率; H表示U和S的调和平均数; “—”表示该项无结果.

表4 野生动物检测算法在AWA和CUB上的比较结果(%)

Methods	ZSL				GZSL					
	AWA		CUB		AWA			CUB		
	mrAcc	Top-1	mrAcc	Top-1	U	S	H	U	S	H
DEWISE (2013) <sup>[87]</sup>	56.7	54.2	33.5	52.0	13.4	68.7	22.4	23.8	53.0	32.8
SJE (2015) <sup>[82]</sup>	73.9	65.6	51.7	53.9	11.3	74.6	19.6	23.5	59.2	33.6
ALE (2016) <sup>[85]</sup>	61.9	59.9	40.3	54.9	16.8	76.1	27.5	23.7	62.8	34.4
f-xGAN (2018) <sup>[88]</sup>	—	68.2	—	57.3	57.9	61.4	59.6	43.7	57.7	49.7
PQZSL (2019) <sup>[86]</sup>	—	—	—	—	31.7	70.9	43.8	43.2	51.4	46.9
TVN (2019) <sup>[89]</sup>	—	68.8	—	58.1	27.0	67.9	38.6	26.5	62.3	37.2
TCN (2020) <sup>[83]</sup>	70.3	—	59.5	—	49.4	76.5	60.0	52.6	52.0	52.3
MCGM-VAE (2020) <sup>[90]</sup>	—	65.02	—	63.54	74.78	57.45	64.98	58.00	51.06	54.31
E-PGN (2020) <sup>[91]</sup>	—	—	—	—	62.1	83.4	71.2	52.0	61.1	56.2
BB-OOD (2020) <sup>[92]</sup>	—	—	—	—	55.9	94.9	72.6	53.8	94.6	68.6
DRN (2020) <sup>[93]</sup>	—	—	—	—	50.1	81.4	62.1	46.9	58.8	52.2
LSG (2021) <sup>[84]</sup>	—	—	—	—	—	—	—	49.6	50.4	50.0
SE-OOD (2021) <sup>[94]</sup>	—	—	—	—	67.60	93.83	78.58	54.67	95.74	69.60

## 4 问题分析与展望

目前虽然提出了一些基于机器学习和深度学习的野生动物检测与识别算法,但提出的模型框架相对较为简单,且效果还不够理想.根据上述分析,总结出目前基于图像的野生动物检测与识别研究存在的问题及未来研究的展望,具体如下.

### (1) 野生动物物种识别

由于现有的野生动物数据集中,只有 iNaturalist 和 CUB-200-2011 的背景影响较小,且图像的质量较高.然而,在野外采集的野生动物监测系统中,图像通常含有较多的背景,且受到各种因素的影响较大,此外,不同物种样本的数量也存在不均衡的情况.因此,需要研究能够充分提取有效区域特征、抗噪性强、能够解决样本不均衡问题的专门针对野生动物物种识别的网络.此外,根据文献[95]的研究可知,动物的声音也可以判别动物的种类,因此,在采集动物图像的同时获取野生动物的声音,并采用多视角学习的方法,结合图像和声音信号的信息,可以提高野生动物物种识别的准确率和可靠性.

### (2) 野生动物检测与识别网络

目前提出的野生动物检测与识别网络对于复杂的场景、图像质量较差和目标不完整等情况的效果并不理想,这些网络主要针对地面获取的野生动物图像,这些图像中目标数量相对较少且目标较大,便于检测和识别,但针对无人机拍摄的图像中野生动物的检测与识别同时实现的端到端网络仍然较少.因此,针对复杂情况的野生动物检测与识别的网络,尤其是针对包括无人机图像和无线传感器网络图像中小目标、目标不完整、图像质量较差等情况下的识别,是野生动物检测与识别网络研究的重点.

### (3) 轻量级的检测识别网络

在移动端或配置相对较低的处理器上实现野生动物目标检测与识别对于野生动物监测系统具有重要的意义,快速而自动地在移动设备上进行野生动物检测与识别需要采用轻量级的网络结构.然而,目前大多数网络模型的参数较多,对于系统的硬件等要求较高,这使得在移动端上实现高效野生动物检测与识别成为一个挑战.

### (4) 小样本或零样本学习的检测识别网络

深度学习方法在野生动物的检测和识别任务中通常需要大量标记的数据集进行训练,以获得良好的性能表现,但对于某些珍稀野生动物的检测和识别,收集

和标注大量的数据集是非常困难的.相比之下,人类具备从少样本中学习和推理的能力,能识别出从未见过的类别,解决这类问题的研究重点是开发零样本学习模型,即能够在没有训练样本的情况下预测出训练集中未出现过的类别.

### (5) 野生动物的视频跟踪和检测

目前,基于视频的野生动物跟踪和检测研究相对较少,但这是野生动物研究中的一个重要方向.野生动物的运动轨迹和状态监测对于了解它们的行为和生态具有重要意义.在目标跟踪方法方面,关键问题是如何有效地获取动态移动的野生动物影像,其中可能涉及视频采集技术、摄像机部署策略以及图像处理技术等方面的研究.同时,实现野生动物目标的实时跟踪还要求能够将目标进行准确分类,以进一步了解野生动物的行为和特征.因此,研究能够实现实时野生动物跟踪和检测,并具有强大抗噪性的端到端网络模型具有重要意义.

## 参考文献

- 1 于华. 近 50 年全球野生动物减少 2/3. 生态经济, 2020, 36(11): 5-8.
- 2 梁成陆. 野生动物保护的意義及举措. 现代农业科技, 2021(10): 184-185. [doi: 10.3969/j.issn.1007-5739.2021.10.075]
- 3 陈梦霄, 王晓玲, 穆丽光, 等. 内蒙古赛罕乌拉国家级自然保护区豹的日活动节律及其季节变化研究. 四川动物, 2019, 38(2): 214-219. [doi: 10.11984/j.issn.1000-7083.20180273]
- 4 何金权. 广东东源康禾自然保护区陆生野生动物资源现状及保护对策. 内蒙古林业调查设计, 2018, 41(5): 33-35.
- 5 Allan BM, Arnould JPY, Martin JK, *et al.* A cost-effective and informative method of GPS tracking wildlife. *Wildlife Research*, 2013, 40(5): 345-348. [doi: 10.1071/WR13069]
- 6 Xu WT, Zhang X, Yao LN, *et al.* A multi-view CNN-based acoustic classification system for automatic animal species identification. *Ad Hoc Networks*, 2020, 102: 102115. [doi: 10.1016/j.adhoc.2020.102115]
- 7 Steen KA, Villa-Henriksen A, Therkildsen OR, *et al.* Automatic detection of animals in mowing operations using thermal cameras. *Sensors*, 2012, 12(6): 7587-7597. [doi: 10.3390/s120607587]
- 8 Cilulko J, Janiszewski P, Bogdaszewski M, *et al.* Infrared thermal imaging in studies of wild animals. *European Journal of Wildlife Research*, 2013, 59(1): 17-23. [doi: 10.1007/s10344-012-0688-1]

- 9 Pavlíček J, Jarolimek J, Jarolimek J, *et al.* Automated wildlife recognition. *AGRIS On-line Papers in Economics and Informatics*, 2018, 10(1): 51–60. [doi: [10.7160/aol.2018.100105](https://doi.org/10.7160/aol.2018.100105)]
- 10 Seymour AC, Dale J, Hammill M, *et al.* Automated detection and enumeration of marine wildlife using unmanned aircraft systems (UAS) and thermal imagery. *Scientific Reports*, 2017, 7(1): 45127. [doi: [10.1038/srep45127](https://doi.org/10.1038/srep45127)]
- 11 Chamoso P, Raveane W, Parra V, *et al.* UAVs applied to the counting and monitoring of animals. *Proceedings of the 5th International Symposium on Ambient Intelligence*. Springer, 2014. 71–80.
- 12 程跃红, 张卫东, 彭晓辉, 等. 卧龙国家级自然保护区热水河温泉周边有蹄类野生动物红外相机监测初报. *农业与技术*, 2016, 36(1): 178–180.
- 13 刘雪华, 武鹏峰, 何祥博, 等. 红外相机技术在物种监测中的应用及数据挖掘. *生物多样性*, 2018, 26(8): 850–861. [doi: [10.17520/biods.2018053](https://doi.org/10.17520/biods.2018053)]
- 14 陈善安, 胡春鹤, 王远, 等. 基于无线传感器网络的野生动物图像监测系统. *现代制造技术与装备*, 2017(3): 64–66. [doi: [10.3969/j.issn.1673-5587.2017.03.030](https://doi.org/10.3969/j.issn.1673-5587.2017.03.030)]
- 15 Xie JJ, Li AQ, Zhang JG, *et al.* An integrated wildlife recognition model based on multi-branch aggregation and squeeze-and-excitation network. *Applied Sciences*, 2019, 9(14): 2794. [doi: [10.3390/app9142794](https://doi.org/10.3390/app9142794)]
- 16 Zhao P, Liu SM, Zhou Y, *et al.* Estimating animal population size with very high-resolution satellite imagery. *Conservation Biology*, 2021, 35(1): 316–324. [doi: [10.1111/cobi.13613](https://doi.org/10.1111/cobi.13613)]
- 17 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. *植物生态学报*, 2020, 44(4): 409–417. [doi: [10.17521/cjpe.2019.0165](https://doi.org/10.17521/cjpe.2019.0165)]
- 18 Linchant J, Lisein J, Semeki J, *et al.* Are unmanned aircraft systems (UASs) the future of wildlife monitoring? A review of accomplishments and challenges. *Mammal Review*, 2015, 45(4): 239–252. [doi: [10.1111/mam.12046](https://doi.org/10.1111/mam.12046)]
- 19 Swanson A, Kosmala M, Lintott C, *et al.* Snapshot serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. *Scientific Data*, 2015, 2: 150026. [doi: [10.1038/sdata.2015.26](https://doi.org/10.1038/sdata.2015.26)]
- 20 Villa AG, Salazar A, Vargas F. Towards automatic wild animal monitoring: Identification of animal species in camera-trap images using very deep convolutional neural networks. *Ecological Informatics*, 2017, 41: 24–32. [doi: [10.1016/j.ecoinf.2017.07.004](https://doi.org/10.1016/j.ecoinf.2017.07.004)]
- 21 杜科. 基于 Attention-CNN-GRU 的野生动物监测图像分类 [硕士学位论文]. 北京: 北京林业大学, 2019. 1–70.
- 22 Timm M, Maji S, Fuller T. Large-scale ecological analyses of animals in the wild using computer vision. *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. Salt Lake City: IEEE, 2018. 1977–19772.
- 23 Trammell MJ, Oberoi P, Egenrieder J, *et al.* Contextual label smoothing with a phylogenetic tree on the iNaturalist 2018 challenge dataset. *Journal of the Washington Academy of Sciences*, 2019, 105(1): 23–46.
- 24 Wah C, Branson S, Welinder P, *et al.* The caltech-UCSD birds-200-2011 dataset. *Technical Report*, Pasadena: California Institute of Technology, 2011.
- 25 Chen RL, Little R, Mihaylova L, *et al.* Wildlife surveillance using deep learning methods. *Ecology and Evolution*, 2019, 9(17): 9453–9466. [doi: [10.1002/ece3.5410](https://doi.org/10.1002/ece3.5410)]
- 26 Rey N, Volpi M, Joost S, *et al.* Detecting animals in African savanna with UAVs and the crowds. *Remote Sensing of Environment*, 2017, 200: 341–351. [doi: [10.1016/j.rse.2017.08.026](https://doi.org/10.1016/j.rse.2017.08.026)]
- 27 陈建促, 王越, 朱小飞, 等. 融合多特征图的野生动物视频目标检测方法. *计算机工程与应用*, 2020, 56(7): 221–227. [doi: [10.3778/j.issn.1002-8331.1901-0087](https://doi.org/10.3778/j.issn.1002-8331.1901-0087)]
- 28 Chen K, Song H, Loy CC, *et al.* Discover and learn new objects from documentaries. *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu: IEEE, 2017. 1111–1120.
- 29 Lampert CH, Nickisch H, Harmeling S. Attribute-based classification for zero-shot visual object categorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, 36(3): 453–465. [doi: [10.1109/TPAMI.2013.140](https://doi.org/10.1109/TPAMI.2013.140)]
- 30 Xian YQ, Lampert CH, Schiele B, *et al.* Zero-shot learning—A comprehensive evaluation of the good, the bad and the ugly. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, 41(9): 2251–2265. [doi: [10.1109/TPAMI.2018.2857768](https://doi.org/10.1109/TPAMI.2018.2857768)]
- 31 Favorskaya M, Pakhirka A. Animal species recognition in the wildlife based on muzzle and shape features using joint CNN. *Procedia Computer Science*, 2019, 159: 933–942. [doi: [10.1016/j.procs.2019.09.260](https://doi.org/10.1016/j.procs.2019.09.260)]
- 32 Tabak MA, Norouzzadeh MS, Wolfson DW, *et al.* Machine learning to classify animal species in camera trap images: Applications in ecology. *Methods in Ecology and Evolution*, 2019, 10(4): 585–590. [doi: [10.1111/2041-210X.13120](https://doi.org/10.1111/2041-210X.13120)]
- 33 Nguyen H, Maclagan SJ, Nguyen TD, *et al.* Animal recognition and identification with deep convolutional neural networks for automated wildlife monitoring. *Proceedings of*

- the 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA). Tokyo: IEEE, 2017. 40–49.
- 34 Song YL, Wang HP, Li S, *et al.* CNN based wildlife recognition with super-pixel segmentation for ecological surveillance. Proceedings of the 8th IEEE Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER). Tianjin: IEEE, 2018. 132–137.
- 35 Ferreira BA, De Villiers JP, De Freitas A. Dashcam based wildlife detection and classification using fused data sets of digital photographic and simulated imagery. Proceedings of the 23rd IEEE International Conference on Information Fusion (FUSION). Rustenburg: IEEE, 2020. 1–8.
- 36 Okafor E, Smit R, Schomaker L, *et al.* Operational data augmentation in classifying single aerial images of animals. Proceedings of the 2017 IEEE International Conference on Innovations in Intelligent Systems and Applications (INISTA). Gdynia: IEEE, 2017. 354–360.
- 37 曾陈颖. 面向珍稀野生动物保护的图像监测与识别技术研究[硕士学位论文]. 北京: 北京林业大学, 2015. 1–69.
- 38 Hong NM, Thanh NC. A single image dehazing method based on adaptive gamma correction. Proceedings of the 6th NAFOSTED Conference on Information and Computer Science (NICS). Hanoi: IEEE, 2019. 558–562.
- 39 张军国, 程浙安, 胡春鹤, 等. 野生动物监测光照自适应 Retinex 图像增强算法. 农业工程学报, 2018, 34(15): 183–189. [doi: 10.11975/j.issn.1002-6819.2018.15.023]
- 40 Qiao BW, Zhou ZF, Yang HT, *et al.* Bird species recognition based on SVM classifier and decision tree. Proceedings of the 1st International Conference on Electronics Instrumentation & Information Systems (EIIS). Harbin: IEEE, 2017. 1–4.
- 41 Noviyanto A, Arymurthy AM. Beef cattle identification based on muzzle pattern using a matching refinement technique in the SIFT method. Computers and Electronics in Agriculture, 2013, 99: 77–84. [doi: 10.1016/j.compag.2013.09.002]
- 42 Yu XY, Wang JP, Kays R, *et al.* Automated identification of animal species in camera trap images. Eurasip Journal on Image and Video Processing, 2013, 2013(1): 52. [doi: 10.1186/1687-5281-2013-52]
- 43 Uddin MS, Akhi AY. Horse detection using Haar like features. International Journal of Computer Theory and Engineering, 2016, 8(5): 415–418. [doi: 10.7763/IJCTE.2016.V8.1081]
- 44 Alsahwa B, Maussang F, Garelo R, *et al.* Marine life airborne observation using HOG and SVM classifier. Proceedings of the OCEANS 2016 MTS/IEEE Monterey. Monterey: IEEE, 2016. 1–5.
- 45 Gonzalez LF, Montes GA, Puig E, *et al.* Unmanned aerial vehicles (UAVs) and artificial intelligence revolutionizing wildlife monitoring and conservation. Sensors, 2016, 16(1): 97. [doi: 10.3390/s16010097]
- 46 Christiansen P, Steen KA, Jørgensen RN, *et al.* Automated detection and recognition of wildlife using thermal cameras. Sensors, 2014, 14(8): 13778–13793. [doi: 10.3390/s140813778]
- 47 Kamencay P, Trnovszky T, Benco M, *et al.* Accurate wild animal recognition using PCA, LDA and LBPH. Proceedings of the 2016 ELEKTRO. Strbske Pleso: IEEE, 2016. 62–67.
- 48 Feng WZ, Zhang JG, Hu CH, *et al.* A novel saliency detection method for wild animal monitoring images with WMSN. Journal of Sensors, 2018, 2018: 3238140.
- 49 Okafor E, Pawara P, Karaaba F, *et al.* Comparative study between deep learning and bag of visual words for wild-animal recognition. Proceedings of the 2016 IEEE Symposium Series on Computational Intelligence (SSCI). Athens: IEEE, 2016. 1–8.
- 50 Norouzzadeh MS, Nguyen A, Kosmala M, *et al.* Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. Proceedings of the National Academy of Sciences of the United States of America, 2017, 115(25): E5716–E5725.
- 51 Tang YX, Wang J, Gao BY, *et al.* Large scale semi-supervised object detection using visual and semantic knowledge transfer. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016. 2119–2128.
- 52 Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640–651. [doi: 10.1109/TPAMI.2016.2572683]
- 53 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. Proceedings of the 18th International Conference on Medical Image Computing and Computer-assisted Intervention. Munich: Springer, 2015. 234–241.
- 54 Badrinarayanan V, Kendall A, Cipolla R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481–2495. [doi: 10.1109/TPAMI.2016.2644615]

- 55 Parham J, Stewart C, Crall J, *et al.* An animal detection pipeline for identification. Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). Lake Tahoe: IEEE, 2018. 1075–1083.
- 56 Feng WZ, Hu CH, Wang Y, *et al.* A novel hierarchical coding progressive transmission method for WMSN wildlife images. *Sensors*, 2019, 19(4): 946. [doi: [10.3390/s19040946](https://doi.org/10.3390/s19040946)]
- 57 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. *兽类学报*, 2019, 39(4): 458–465.
- 58 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. *北京林业大学学报*, 2018, 40(8): 123–131.
- 59 蔡前舟, 郑伯川, 曾祥银, 等. 结合长尾数据解决方法的野生动物目标检测. *计算机应用*, 2022, 42(4): 1284–1291.
- 60 王越, 陈建促, 朱小飞, 等. 基于熵和运动偏移量的野生动物遮挡检测方法. *计算机应用研究*, 2019, 36(12): 3869–3873, 3877.
- 61 Kellenberger B, Marcos D, Tuia D. Detecting mammals in UAV images: Best practices to address a substantially imbalanced dataset with deep learning. *Remote Sensing of Environment*, 2018, 216: 139–153. [doi: [10.1016/j.rse.2018.06.028](https://doi.org/10.1016/j.rse.2018.06.028)]
- 62 Ren XB, Han TX, He ZH. Ensemble video object cut in highly dynamic scenes. Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland: IEEE, 2013. 1947–1954.
- 63 Kellenberger B, Marcos D, Lobry S, *et al.* 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*, 2019, 57(12): 9524–9533. [doi: [10.1109/TGRS.2019.2927393](https://doi.org/10.1109/TGRS.2019.2927393)]
- 64 Li Y, Tong GF, Gao HS, *et al.* Pano-RSOD: A dataset and benchmark for panoramic road scene object detection. *Electronics*, 2019, 8(3): 329. [doi: [10.3390/electronics8030329](https://doi.org/10.3390/electronics8030329)]
- 65 Tong GF, Chen HR, Li Y, *et al.* Object detection for panoramic images based on MS-RPN structure in traffic road scenes. *IET Computer Vision*, 2019, 13(5): 500–506. [doi: [10.1049/iet-cvi.2018.5304](https://doi.org/10.1049/iet-cvi.2018.5304)]
- 66 Real E, Aggarwal A, Huang YP, *et al.* Regularized evolution for image classifier architecture search. Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Honolulu: AAAI, 2019. 4780–4789.
- 67 Howard AG, Zhu ML, Chen B, *et al.* MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv:1704.04861, 2017.
- 68 Huang G, Liu Z, van der Maaten L, *et al.* Densely connected convolutional networks. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 2261–2269.
- 69 Zhang XY, Zhou XY, Lin MX, *et al.* ShuffleNet: An extremely efficient convolutional neural network for mobile devices. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 6848–6856.
- 70 Chollet F. Xception: Deep learning with depthwise separable convolutions. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 1800–1807.
- 71 Ren SQ, He KM, Girshick R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137–1149. [doi: [10.1109/TPAMI.2016.2577031](https://doi.org/10.1109/TPAMI.2016.2577031)]
- 72 Dai JF, Li Y, He KM, *et al.* R-FCN: Object detection via region-based fully convolutional networks. Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona: Curran Associates Inc., 2016. 379–387.
- 73 He KM, Gkioxari G, Dollár P, *et al.* Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(2): 386–397. [doi: [10.1109/TPAMI.2018.2844175](https://doi.org/10.1109/TPAMI.2018.2844175)]
- 74 Wang XL, Cheng P, Liu XC, *et al.* Focal loss dense detector for vehicle surveillance. Proceedings of the 2018 International Conference on Intelligent Systems and Computer Vision (ISCV). Fez: IEEE, 2018. 1–5.
- 75 Jeong J, Park H, Kwak N. Enhancement of SSD by concatenating feature maps for object detection. Proceedings of the 2017 British Machine Vision Conference. London: BMVC, 2017.
- 76 Lin TY, Goyal P, Girshick R, *et al.* Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(2): 318–327. [doi: [10.1109/TPAMI.2018.2858826](https://doi.org/10.1109/TPAMI.2018.2858826)]
- 77 Zhao QJ, Sheng T, Wang YT, *et al.* M2Det: A single-shot object detector based on multi-level feature pyramid network. Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Honolulu: AAAI, 2019. 9259–9266.
- 78 Zhang SF, Wen LY, Bian X, *et al.* Single-shot refinement neural network for object detection. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern

- Recognition. Salt Lake City: IEEE, 2018. 4203–4212.
- 79 Li YH, Chen YT, Wang NY, *et al.* Scale-aware trident networks for object detection. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE, 2020. 6053–6062.
- 80 Fu CY, Liu W, Ranga A, *et al.* DSSD: Deconvolutional single shot detector. arXiv:1701.06659, 2017.
- 81 Shen ZQ, Liu Z, Li JG, *et al.* DSOD: Learning deeply supervised object detectors from scratch. Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE, 2017. 1937–1945.
- 82 Akata Z, Reed S, Walter D, *et al.* Evaluation of output embeddings for fine-grained image classification. Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015. 2927–2936.
- 83 Jiang HJ, Wang RP, Shan SG, *et al.* Transferable contrastive network for generalized zero-shot learning. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE, 2019. 9764–9773.
- 84 Xu BR, Zeng ZG, Lian C, *et al.* Semi-supervised low-rank semantics grouping for zero-shot learning. IEEE Transactions on Image Processing, 2021, 30: 2207–2219. [doi: [10.1109/TIP.2021.3050677](https://doi.org/10.1109/TIP.2021.3050677)]
- 85 Chao WL, Changpinyo S, Gong BQ, *et al.* An empirical study and analysis of generalized zero-shot learning for object recognition in the wild. Proceedings of the 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 52–68.
- 86 Li J, Lan XG, Liu Y, *et al.* Compressing unknown images with product quantizer for efficient zero-shot classification. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019. 5458–5467.
- 87 Akata Z, Perronnin F, Harchaoui Z, *et al.* Label-embedding for attribute-based classification. Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland: IEEE, 2013. 819–826.
- 88 Xian YQ, Lorenz T, Schiele B, *et al.* Feature generating networks for zero-shot learning. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 5542–5551.
- 89 Zhang HF, Long Y, Guan Y, *et al.* Triple verification network for generalized zero-shot learning. IEEE Transactions on Image Processing, 2019, 28(1): 506–517. [doi: [10.1109/TIP.2018.2869696](https://doi.org/10.1109/TIP.2018.2869696)]
- 90 Shao J, Li XR. Generalized zero-shot learning with multi-channel Gaussian mixture VAE. IEEE Signal Processing Letters, 2020, 27: 456–460. [doi: [10.1109/LSP.2020.2977498](https://doi.org/10.1109/LSP.2020.2977498)]
- 91 Yu YL, Ji Z, Han JG, *et al.* Episode-based prototype generating network for zero-shot learning. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 14032–14041.
- 92 Chen XY, Lan XG, Sun FC, *et al.* A boundary based out-of-distribution classifier for generalized zero-shot learning. Proceedings of the 16th European Conference on Computer Vision. Glasgow: Springer, 2020. 572–588.
- 93 Huang S, Lin JK, Huangfu LW. Class-prototype discriminative network for generalized zero-shot learning. IEEE Signal Processing Letters, 2020, 27: 301–305. [doi: [10.1109/LSP.2020.2968213](https://doi.org/10.1109/LSP.2020.2968213)]
- 94 Ding JY, Hu X, Zhong XR. A semantic encoding out-of-distribution classifier for generalized zero-shot learning. IEEE Signal Processing Letters, 2021, 28: 1395–1399. [doi: [10.1109/LSP.2021.3092227](https://doi.org/10.1109/LSP.2021.3092227)]
- 95 Zhou FT, Huang S, Xing Y. Deep semantic dictionary learning for multi-label image classification. Proceedings of the 35th AAAI Conference on Artificial Intelligence. AAAI, 2021. 3572–3580.

(校对责编: 孙君艳)