

# 基于图像识别的有害生物检疫鉴定探索研究

孙佳佳<sup>1</sup> 吕飞<sup>1</sup> 雷晨曦<sup>2</sup> 尚岩峰<sup>2</sup> 熊惠霖<sup>2</sup>

(1.太仓海关 江苏太仓 215400; 2.上海交通大学电子信息与电气工程学院)

**Exploring research on pest quarantine identification basing on image recognition.** Sun Jiajia<sup>1</sup>, Lü Fei<sup>1</sup>, Lei Chenxi<sup>2</sup>, Shang Yanfeng<sup>2</sup>, Xiong Huilin<sup>2</sup> (1. Taicang Customs House, Taicang 215400, China; 2. School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University)

**Abstract** The identification of exotic pest has always been the focus of quarantine work at ports. And identifying the various pests in species level or genera level accurately is a big challenge. Image recognition technology is a feasible way to solve this issue. In this paper, a hierarchical classification model based on CNN was used to identify 9 681 pictures of 70 typical species which were intercepted from imported wood, at family level, genus level and species level. In the experiment on test dataset with 1936 images, the average classification accuracy at family, genus and species level of the model achieved 97.71%, 95.85% and 86.92% respectively. The experimental results demonstrated the ability of our model to identify biological images and the feasibility of using biological images to identify port pests.

**Keywords** pest; image recognition; deep learning; convolutional neural network; identification

**摘要** 外来有害生物的鉴定一直是口岸检疫工作的重点,而对品种繁多的有害生物在种属层级进行准确识别则是检验检疫工作的难点,最新的计算机图像识别技术提供了解决这一问题的一种可行的途径。本文利用深度卷积网络的层次分类模型,对进口木材中经常截获的 70 种典型物种共 9 681 张图片在科、属、种这 3 个分类层级上进行了识别鉴定,在 1 936 张图片的测试中,模型在科、属、种上对 70 类有害生物的平均识别精度分别为 97.71%、95.85% 和 86.92%。实验结果证明了模型对生物图像的学习能力,以及利用生物图像识别口岸有害生物的可行性。

**关键词** 有害生物;图像识别;深度学习;卷积神经网络;鉴定

中图分类号 S41 文献标识码 A DOI:10.19662/j.cnki.issn1005-2755.2020.05.002

近年来,随着我国对外贸易的持续发展,植物检疫工作在国家经济、社会领域的关注度不断提高,国门生物安全得到了社会的广泛关注,国家对植物检疫工作越来越重视,口岸检疫把关能力逐年提升。2017 年,全国各口岸在进境植物检疫中共截获有害生物 5 956 种、1 053 450 种次。在全国口岸通关时长不断压缩的情况下,植物检疫工作者面对数量庞大的有害生物种类,尤其是以形态鉴定为主的昆虫和杂草籽,必须快速、准确鉴定出这些有害生物。传统上,一般通过形态特征对有害生物进行鉴定<sup>[1-2]</sup>。这就要求一线检疫人员熟悉有关物种的形态特征,能够迅速准确地判断。然而,受限于海量的有害生物种类,且同科同属的物种生物视觉特征相

似度较高,需要很强的专业技能才能进行鉴别,这无疑对一线检疫技术人员的专业素质提出了相当高的要求。加上技术人员难以获得外来物种的相关资料和信息,使得鉴定工作更为困难。因此,现场执法和检疫鉴定难度高,效率低,直接影响了植物疫情截获率。而利用有害生物的图像识别方法来对有害生物进行鉴定,可以在一定程度上缓解这一问题。在人工智能和图像识别技术不断突破,大数据和云计算日益成熟的技术背景下,形成大规模有害生物的智能图像识别成为可能。通过建设有害生物图像资源库,构建有害生物识图鉴图服务体系,可以显著提升有害生物检疫鉴定、监测和防控工作效率。

利用生物图像信息进行生物分类的研究大致

基金项目:南京海关科技计划项目(2018KJ10)

第一作者 E-mail: ntsun2010@163.com

收稿日期:2019-12-20

可以分为 3 个阶段。早期使用的方法大多采用将人工提取的生物图像形态特征与传统模式识别相结合,进行生物分类识别。比如,于新文等<sup>[3]</sup>利用区域面积、偏心率、形状参数、周长、似圆度、叶状性等 6 个昆虫图像几何形态特征,对棉铃虫(*Helicoverpa armigera*)、玉米螟(*Pyrausta nubilalis*)、黑咬猎蝽(*Ectomocoris atrox*)等 3 种昆虫的 145 张图片进行了识别,准确率为 100%。李正等<sup>[4]</sup>提取了躯干偏心率、占空比、空虚度、整体偏心率、似圆度、球状性、圆形性、叶状性和质心距离比等 9 个昆虫图像形态特征,并利用支持向量机模型(Support Vector Machine)对桔小实蝇(*Bactrocera dorsalis*)、番石榴实蝇(*Bactrocera correcta*)、辣椒实蝇(*Bactrocera latifrons*)、锈红寡鬃实蝇(*Bactrocera rubiginus*)等 20 种云南省常见实蝇进行分类,准确率为 85.336%。由于采用生物图像几何形态特征进行识别要求被识别的生物形态完好,拍摄图片质量良好,且人工选择的形态特征大多受具体生物类别的影响较大,故后来逐渐转向使用图像的颜色和纹理等信息作为特征采用机器学习的方法进行识别。如娄定风等<sup>[5]</sup>基于昆虫图像的形状和纹理算法对 6 个目下的 14 种昆虫的视频图像进行识别,测试结果表明该算法对不同体长、各种长宽比,以及深色的昆虫识别效果良好。竺乐庆等<sup>[6]</sup>利用昆虫图像的颜色特征和 OpponentSIFT 特征结合词袋模型和 SVM 分类器对鳞翅目下的 10 种共 576 个样本的昆虫进行了识别,准确率达 100%。近年来,随着深度学习的发展,深度卷积神经网络(CNN)被逐渐引入昆虫图像识别这一领域。如马梦园等<sup>[7]</sup>采用深度学习算法对鳞翅目下的 70 种共 3 972 张昆虫图像进行识别,准确率达 99.2%。然而,不同环境下获取的生物图像差异较大,不同的光照、角度、尺度等都对识别模型提出了挑战。因此,如何提升识别模型的稳健性(即学习样本的精准度),减少环境因素对其识别准确率的影响,也是图像识别需要研究的重点。

本文利用深度卷积神经网络的层次分类模型,对 70 种典型物种共 9 681 张图片,在科、属、种 3 个分类层级上进行了识别鉴定。这些种类是近 5 年来各口岸较常截获的有害生物,包括昆虫、杂草两大类。经过多次测试模型对生物图像的学习能力,验证了利用生物图像来识别口岸有害生物具有可行性,以期在现场检疫人员快速准确识别有害生物提供技术支持。

## 1 材料与方法

### 1.1 实验材料

70 种口岸较常截获外来生物(昆虫成虫 52 种,杂草种子 18 种)的图片,分别属于小蠹科、天牛科、长蠹科、芎科、菊科、旋花科、大戟科及马齿苋科,每种图片数量从 23 张到 1 087 张不等,共 9 681 张,由中国检科院和南京海关鉴定提供。

### 1.2 方法

#### 1.2.1 模型设置

深度学习的模型有很多种,包括深度信念网络、卷积神经网络、自动编码器等,其中深度卷积神经网络与传统模式识别方法相比具有更强大的特征学习和特征表达能力。本文采用基于深度卷积神经网络的层次分类模型进行研究,模型层次与有害生物分类层次一致,即整个识别模型分为科分类模型、属分类模型和种分类模型三大部分。每一层次节点都有识别的相似性;层次越高,分类地位越高,相似性越高。

#### 1.2.2 模型训练

模型采用在 ImageNet 上预训练的 VGG160 作为层次分类网络的基础网络,将整个数据集以各类别的图片数量为基准,按照 8:2 的比例,随机地将每个生物类别的图片分为训练集和测试集两个部分。其中训练集包含 7 745 张图片,测试集 1 936 张。

模型训练时,用训练集中各类别的图片作为模型的输入,用每张图片所对应的科、属、种名作为标签,来分别对科、属、种 3 个层级的模型进行训练。训练结束后,使用对应的测试集来对模型的识别结果进行评估。

## 2 结果与分析

模型训练完成后,利用测试集中的图片对模型进行测试,模型对 70 种有害生物的识别结果如表 1 所示。科、属、种的平均识别率如表 2 所示。

从表 1 可以看出,在科和属这两个分类层级上,模型对 70 种有害生物的识别率均达到了较高的水平,而在种这一分类层级下,模型对大部分物种也有较好的识别效果。由于同一属不同种的物种之间较为相似,且深度卷积神经网络的训练需要较多的样本,因此部分样本量少的物种识别效果不佳。

从表 2 来看,随着生物分类层级的降低,物种之间变得愈加相似,故模型的识别结果也呈现下降趋势,但是总体来看,还是保持着较高的识别精度。而在 1 936 张测试图片的验证,也在一定程度上反映了算法的稳健性。

表 1 模型对 70 种有害生物识别的测试结果

科	中文名	拉丁名	图片数量 / 张	种识别率 / %	属识别率 / %	科识别率 / %
小蠹科	六齿小蠹	<i>Ips acuminatus</i>	205	81.30	97.56	97.56
	美东最小齿小蠹	<i>Ips avulsus</i>	28	50.00	100.00	100.00
	美雕齿小蠹	<i>Ips calligraphus</i>	224	79.10	100.00	100.00
	锡特加云杉齿小蠹	<i>Ips concinnus</i>	40	50.00	100.00	100.00
	重齿小蠹	<i>Ips duplicatus</i>	117	81.40	87.50	87.50
	南部松齿小蠹	<i>Ips grandicollis</i>	408	89.00	100.00	100.00
	加州松齿小蠹	<i>Ips integer</i>	28	50.00	83.33	83.33
	北美西部松齿小蠹	<i>Ips latidens</i>	52	100.00	100.00	100.00
	美松齿小蠹	<i>Ips pini</i>	248	82.00	100.00	100.00
	松十二小蠹	<i>Ips sexdentatus</i>	455	82.40	98.90	98.90
	落叶松八齿小蠹	<i>Ips subelongatus</i>	411	84.30	100.00	100.00
	大云齿小蠹	<i>Ips tridens</i>	26	60.00	100.00	100.00
	云杉八齿小蠹	<i>Ips typographus</i>	1 087	93.10	99.08	99.08
天牛科	灰长角天牛	<i>Acanthocinus aedilis</i>	56	75.00	83.33	100.00
	小灰长角天牛	<i>Acanthocinus griseus</i>	41	66.70	88.89	100.00
	黄腹锥鞘天牛	<i>Acyphoderes abdominalis</i>	50	100.00	100.00	100.00
	爪哇木棉棘胸天牛	<i>Ancylonotus tribulus</i>	47	80.00	90.00	100.00
	星天牛	<i>Anoplophora chinensis</i>	70	78.60	92.86	100.00
	光肩星天牛	<i>Anoplophora glabripennis</i>	60	83.30	83.33	91.67
	桑天牛	<i>Apriona germari</i>	35	71.40	85.71	100.00
	褐梗天牛	<i>Arhopalus rusticus</i>	72	86.70	86.67	93.33
	暗梗天牛	<i>Arhopalus tristis</i>	34	71.40	85.71	100.00
	赤梗天牛	<i>Arhopalus unicolor</i>	31	100.00	100.00	100.00
	松幽天牛	<i>Asemum amurense</i>	29	83.30	83.33	100.00
	榕八星天牛	<i>Batocera rubus</i>	47	90.00	80.00	100.00
	栎天牛	<i>Cerambyx scopolii</i>	56	83.30	100.00	100.00
	斑胸蜡天牛	<i>Ceresium sinicum</i>	33	71.40	85.71	100.00
	竹绿虎天牛	<i>Chlorophorus annularis</i>	74	80.00	86.67	100.00
	刺角楝天牛	<i>Cordylomera spinicornis</i>	39	75.00	87.50	100.00
	北美家天牛	<i>Hylotrupes bajulus</i>	39	50.00	75.00	100.00
	松墨天牛	<i>Monochamus alternatus</i>	56	100.00	91.67	100.00
	樟子松墨天牛	<i>Monochamus galloprovincialis</i>	30	80.32	100.00	100.00
	云杉小墨天牛	<i>Monochamus sutor</i>	46	70.00	90.00	100.00
	云杉大墨天牛	<i>Monochamus urusovi</i>	39	75.00	87.50	100.00
	折斑尼胸天牛	<i>Nyssodrysternum signiferum</i>	24	60.00	71.43	85.71
	箭丽虎天牛	<i>Plagionotus arcuatus</i>	60	100.00	83.33	100.00
	截尾丽虎天牛	<i>Plagionotus detritus</i>	71	100.00	100.00	100.00
	四带刺尾虎天牛	<i>Placosternus difficilis</i>	69	92.90	92.93	100.00
	黄褐棍腿天牛	<i>Phymatodes testaceus</i>	39	75.00	75.00	100.00
	栎红天牛	<i>Pyrrhidium sanguineum</i>	63	100.00	100.00	100.00
	密毛皮花天牛	<i>Rhagium mordax</i>	51	100.00	100.00	100.00
	红足楔天牛	<i>Saperda discoidea</i>	36	50.00	75.00	100.00
	白桦楔天牛	<i>Saperda scalaris</i>	31	100.00	100.00	100.00
	冷杉杉天牛	<i>Semanotus litigious</i>	24	57.10	100.00	100.00
	长角栎天牛	<i>Stromatium longicorne</i>	23	80.00	80.00	100.00
	光胸断眼天牛	<i>Tetropium castaneum</i>	76	81.30	83.75	100.00
	暗褐断眼天牛	<i>Tetropium fuscum</i>	45	88.90	77.78	100.00
	麻天牛	<i>Thyestilla gebleri</i>	27	100.00	100.00	100.00
	叉脊虎天牛	<i>Xylotrechus buqueti</i>	33	100.00	85.71	100.00
	宽斑脊虎天牛	<i>Xylotrechus colonus</i>	55	60.90	100.00	100.00
	巨胸脊虎天牛	<i>Xylotrechus magnicollis</i>	26	83.30	100.00	100.00
长蠹科	双钩异翅长蠹	<i>Heterobostrychus aequalis</i>	545	89.90	88.99	88.99
大戟科	铁苋菜	<i>Acalypha australis</i>	129	82.90	69.23	69.23
菊科	豚草	<i>Ambrosia artemisiifolia</i>	500	88.00	98.00	98.00
	三裂叶豚草	<i>Ambrosia trifida</i>	952	97.40	100.00	100.00

(续表 1)

科	中文名	拉丁名	图片数量 / 张	种识别率 / %	属识别率 / %	科识别率 / %
苋科	白苋	<i>Amaranthus albus</i>	111	82.60	95.65	95.65
	银丁菜	<i>Amaranthus ascendens</i>	36	80.70	85.71	100.00
	野苋	<i>Amaranthus blitum</i>	50	40.00	100.00	100.00
	凹头苋	<i>Amaranthus lividus</i>	82	87.50	100.00	100.00
	长芒苋	<i>Amaranthus palmeri</i>	458	96.70	96.74	96.74
	繁穗苋	<i>Amaranthus paniculatus</i>	41	68.90	100.00	100.00
	反枝苋	<i>Amaranthus retroflexus</i>	325	86.20	95.38	96.92
	西部苋	<i>Amaranthus rudis</i>	496	91.00	99.00	99.00
	刺苋	<i>Amaranthus spinosus</i>	253	81.60	100.00	100.00
	苋	<i>Amaranthus tricolor</i>	76	68.80	100.00	100.00
	糙果苋	<i>Amaranthus tuberculatus</i>	132	80.80	96.15	96.15
	皱果苋	<i>Amaranthus viridis</i>	232	91.50	95.74	95.74
	大花马齿苋	<i>Portulaca grandiflora</i>	24	80.00	100.00	100.00
马齿苋科	马齿苋	<i>Portulaca oleracea</i>	80	93.30	93.33	93.33
旋花科	圆叶牵牛	<i>Pharbitis purpurea</i>	93	89.50	94.74	94.74

表 2 模型对 70 种有害生物科、属、种的平均识别率

分类层级	科	属	种
平均识别率 / %	97.71	95.85	86.92

### 3 讨论

本文采用层次结构的深度卷积神经网络模型对 70 种口岸经常截获的典型物种在科、属、种这 3 个维度下进行了识别,并在由 1 936 张图片构成的测试集中,分别达到了 97.71%、95.85% 和 86.92% 的科、属、种分类准确率。与以往对生物识别的研究相比,本文使用的方法特别之处在于:(1)对于输入的有害生物图片,我们在科、属、种 3 个层级上对其进行识别,并输出其科、属、种的信息。与以往的研究相比,我们的分类层级更低,输出结果更加丰富,更有助于帮助一线检疫人员识别外来生物。(2)我们采用目前最新人工智能方法,即基于深度学习的方法来对这 70 种物种进行建模,有害生物的图片数量达 9 681 张,基于大量样本训练的深度卷积神经网络其识别结果更为稳健,识别精度更高。(3)不同场景下的图像(光照、角度、尺度等)对识别模型会产生较大的影响,我们采用层次结构的深度卷积神经网络模型对有害生物的图片在科、属、种 3 个层级进行识别,层次结构模型采用至顶而下的方法,即使在低层级(比如种)上分类错误,也能保证可能在较高层级(比如科、属)的分类准确性,稳健性更强。(4)我们识别的物种包括昆虫和植物种子,更为丰富多样。

通过研究发现,识别模型的准确率在 3 个层级上各不相同,即识别准确率在科、属、种层级上逐层

下降,这符合我们对有害昆虫、杂草分层识别的预期。由于物种的层级越低,物种之间的差异越小,这也增加了识别该层级的难度,所以低层级的物种识别精度会有所降低。另外,由于同一属不同种的物种之间较为相似,且样本数量差异较大,模型在样本较少种类的识别准确率会有所降低,但总体识别准确率能够达到实用水平。这在一定程度上证明了利用基于深度学习的图像识别技术对口岸有害生物进行识别的可行性。以后的工作将围绕算法对检疫现场手机拍摄图片进行识别的准确性开展,以进一步推动有害生物图像智能识别技术的发展。

### 参考文献

- [1] 王新国,李凯兵,江志海,等. 中南美洲 5 种犀天牛的检疫鉴定. 植物检疫,2018,32(2):33-36.
- [2] 张俊华,陆军,刘露希,等. 进口原木携带的切刺小蠹属昆虫的检疫鉴定. 植物检疫,2018,32(1):54-58.
- [3] 于新文,沈佐锐,高灵旺,等. 昆虫图像几何形状特征的提取技术研究. 中国农业大学学报,2003(3):47-50.
- [4] 李正,倪远平,刘迪,等. 实蝇图像识别中的形态特征提取研究. 计算机仿真,2011,28(7):254-257.
- [5] 姜定风,章桂明,焦懿,等. 基于形状与纹理算法的通用昆虫图像模式识别研究. 植物检疫,2012,26(4):10-15.
- [6] 竺乐庆,张大兴,张真. 基于颜色名和 OpponentSIFT 特征的鳞翅目昆虫图像识别. 昆虫学报,2015,58(12):1331-1337.
- [7] 马梦园. 基于深度学习的鳞翅目昆虫图像处理研究 [D]. 杭州:浙江工商大学,2018.
- [8] Wang D Q, Shen Z Q, Shao J, et al. Multiple granularity descriptors for fine-grained categorization. Computer Vision (ICCV), 2015, 1:2399-2406.
- [9] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [EB/OL]. (2014-09-04) [2019-11-26]. <https://arxiv.org/pdf/1409.1556.pdf>.