

# 基于卷积神经网络下昆虫种类图像识别应用研究

魏甫豫<sup>1</sup>,张振宇<sup>2a</sup>,梁桂珍<sup>2b</sup>

(1.河北工程大学 园林与生态工程学院,河北 邯郸 056000;

2.新乡学院 a.美术学院;b.数学与统计学院,河南 新乡 453000)

**摘要:**昆虫种类图像识别是农业智能化识别虫害的重要方式,精准高效识别昆虫种类是进行针对性防治虫害的前提.利用昆虫数据集 ArTaxOr 及 Insect\_det,基于卷积神经网络下图像分类如 MobileNet,ResNet 及目标检测(FasterRCNN)、Yolo 技术,运用迁移学习进行模型训练,并对比分析训练结果,获取最优昆虫种类图像识别模型.将构建的最优模型采用 EasyEdge 平台进行部署,从而实现了模型到端的全流程开发模式,为后续昆虫种类图像识别场景化应用研究提供依据参考.

**关键词:**昆虫种类图像识别;卷积神经网络;图像分类;目标检测;模型场景应用

**中图分类号:**TP391.41

**文献标志码:**A

我国作为占世界近四分之一人口的大国,粮食问题直接影响人们的物质生活水平,关系到国家的稳定.农业是我国重要的基础产业,虫害是影响粮食生产的重要限制条件之一,因此虫害防治就显得尤为重要.但昆虫种类众多,目前已知的种类超过百万.对于不同种类的虫害也需要有相应针对性的防治措施.因此高效快速准确识别昆虫种类是有效防治虫害的前提.在农作物虫害识别及预测领域,我国传统方法且目前仍普遍使用的办法多是早期传统人工类识别方法,即对昆虫种类的识别都是基于人工调查,如目测法、拍打法、扫网法、诱捕法<sup>[1]</sup>等,大多通过昆虫专家或相关领域学者对昆虫进行人工识别,其准确性精准度取决于相关调查专家的专业水平,且现有昆虫学领域相关专家和专业人员无法满足分布广泛、各类场景的实际应用需求.因此这类方法局限性很大,效率低下,费时费力,很难做到高效率快速识别,不是最佳的监测识别方法.随着人工智能机器学习的快速发展,特别是在深度学习领域各种算法不断涌现,使得在昆虫种类图像领域方面进行高效快速准确的识别成为可能.

## 1 昆虫种类图像识别研究理论基础

### 1.1 卷积神经网络

在深度学习领域,CNN 即卷积神经网络是一种在图像分类识别领域极具优势的方法.其结构包含输入层、隐含层和输出层,且卷积神经网络的隐含层包含卷积层、池化层、全连接层,这个五层结构是一个标准的卷积神经网络模型.卷积层作为重要的核心模块,它包含多个卷积核,可以从输入图像中提取特征.池化层可以重新提取卷积层的特征,即多重滤波增强特征信息.全连接层将提取的特征结合起来,以便最终使用 softmax 进行目标分类.首个标准深度卷积神经网络是由 LeNet-5<sup>[2]</sup>开始,其模型在灰度图像手写体字符识别方面有着很高的识别率,其也含有标准 CNN 网络结构.随后在 2012 年 AlexNet<sup>[3]</sup>网络在 ImageNet 大赛有着突出表现,AlexNet 网络将激活函数 Sigmoid<sup>[4]</sup>替换成 ReLU<sup>[5]</sup>,ReLU 函数具有梯度稳定的特点,解决了神经网络中梯度弥散的问题.自此,基于卷积神经网络的算法和模型大量出现,并在计算机视觉领域迅速发展.

收稿日期:2021-12-22;修回日期:2022-03-25.

基金项目:国家自然科学基金(11871238);河南省科技厅科技攻关项目(132102310482);河南省高等学校重点科研项目(20B110014);国家级大学生创新训练项目(202011071015).

作者简介(通信作者):魏甫豫(1989—),男,河南新乡人,河北工程大学硕士研究生,研究方向为农业工程与信息技术, E-mail:895079499@qq.com.

随着模型深度越来越深,模型复杂度也不断增加。后续产生了一系列模型网络,如 Google 提出的 googLeNet<sup>[6]</sup> 网络模型,由牛津大学视觉几何计算组提出的 VGG<sup>[7]</sup> 模型等。而由微软提出的 ResNet<sup>[8]</sup> 残差网络解决了神经网络深度增加带来网络退化的问题,缓解了深度神经网络增加深度梯度消失的难题。其模型优化方便,性能有了大幅提升。卷积神经网络在图像识别领域之所以能被广泛应用并取得显著效果的关键在于它的两个特点,即局部连接和权值共享方式。局部连接指 CNN 与 ANN 神经元全连接不同,而是进行部分连接,好处是在特定视域下每个卷积核都可以专注于一种特征。权值共享意味着在同层中的卷积核参数共享,并且可以直接输入图像,较传统图像识别复杂图像特征提取方法更优。

## 1.2 迁移学习

使用卷积神经网络进行图像识别虽然有很大的优势,但其对数据集的要求较高,并且需要大量标记数据集进行训练。迁移学习是深度学习中一种非常重要的方法。在图像识别领域,采用预训练模型作为新任务的第一个模型来训练新任务的数据集。其目的是使用预训练模型的权重和参数重新训练新的数据集,加快新模型的收敛速度,减少对数据集中数据量的需求,并减少数据不足导致的模型过度拟合。简而言之,旧领域(即源领域)的学习模型将应用于新领域(即目标领域)上。迁移学习有多种类型,例如基于样本、特征、参数和关系的迁移学习。在实践中,使用迁移学习可以节省从头开始训练新的神经网络的时间。因为由个人创建像 ImageNet 这样庞大数据集是不现实的。在大多数情况下,在特定的实际应用场景中很少有标记的数据集。因此,使用预训练模型进行微调是一个非常好的选择。

## 2 昆虫种类图像识别研究现状

当前已有很多专家学者对不同种类的昆虫进行了研究,成果显著。彭明杰<sup>[9]</sup> 构建了蜻蜓识别系统,张瑞瑞等<sup>[10]</sup> 首次将基于神经网络方法运用于松材线虫病变色木识别方面,薛大暄等<sup>[11]</sup> 设计了一种识别模型 IH-CDM 结合 Inception\_v2 网络的 Faster R-CNN,对单一种类物体进行目标检测有良好的检测效果,对美国白蛾可达到 99.5% 识别率。王卫民等<sup>[12]</sup> 构建了昆虫图像分割的 Insect-Net 模型,基于 U-Net 可将图片背景与目标昆虫进行分离并计数。王茂林等<sup>[13]</sup> 设计的自动化计数软件运用 OpenCV 和 VGG 相结合,可以精准对蓟马进行预测预报。陈彦彤等<sup>[14]</sup> 提出了蝇类面部识别模型,此模型将蝇类面部轮廓和具体部位进行特征提取,结果表明对蝇类面部识别准确率可达 94.03%。张银松等<sup>[15]</sup> 设计了一种改进的目标检测算法,将 Faster-RCNN 主干网络由 ResNet50 代替 VGG16,可以更有效地提取小昆虫的特征,减少密集昆虫区域漏检。由此可见,在昆虫种类图像识别领域,已有许多学者开展了相关研究,且取得了较为显著的结果。但基于以上研究表明,目前已有研究大多为特定场景下优化阶段,未能在实际应用场景下落地部署发布。

深度学习领域的技术已日趋成熟,尤其在图像识别领域的卷积神经网络及目标检测算法在不断优化与迭代。在昆虫种类图像识别领域获得了显著的研究结果,并有了极大的发展。在实际应用领域,已有相关科研单位及高校与科技公司进行项目合作,探索开发出了多款昆虫图像识别 APP 产品,如慧植农当家、植保家等 APP。但仍有不足,一方面是由于日常实际应用场景下获取识别图像的干扰较大,在不同场景下其算法模型难以适用。另一方面,由于昆虫种类繁多,需要大量的昆虫图像数据集来对算法模型进行训练。而现有昆虫数据集大多是某一些昆虫种类或针对特定科研目的场景下昆虫集合的数据集,其种类及数量都相对有限,目前尚未有较为完善的昆虫种类数据集来支撑。因此其数据集训练的对应算法模型就不能满足实际应用场景下昆虫识别的普适性需求。由此可见,进行构建模型到端的场景化研发应用就显得非常重要且势在必行了。

## 3 昆虫图像识别应用研究

### 3.1 端到端场景化昆虫图像分类

为实现模型到端的基于场景化研发应用需求,本文基于 kaggle 开源昆虫数据集,利用开源工具 Paddle,基于 ImageNet 的预训练模型采用两种目前比较流行的卷积网络结构 MobileNetV2 和 ResNet50\_vd\_ssl 进行迁移学习训练。本实验模型训练环境为 Windows10 版 X64\_64bit 操作系统,Intel i5-4590CPU 处理器,Intel(R)HDGraphics 4600 显存,8 GB 内存。

### 3.1.1 数据准备

本实验模型到端场景化应用研究的数据集以 kaggle 平台上的昆虫开源数据集 arthropod-taxonomy-orders-object-detection-dataset 作为实验数据集,简称 ArTaxOr 数据集.kaggle 是 2010 年由 Anthony Goldbloom 在墨尔本创立的平台,作为一个为开发者和数据科学家提供机器学习竞赛与分享数据编程等功能的平台.ArTaxOr 数据集共 15 376 张图片,占 12.06 GB 大小.图像大多拍摄于自然条件下的场景,包含 7 类数据,Araneae,Coleoptera,Diptera,Hemiptera,Hymenoptera,Lepidoptera,Odonata,即蜘蛛目、鞘翅目、双翅目、半翅目、膜翅目、鳞翅目和蜻蜓目.每类昆虫包含 2 000 余张图片,共计 15 000 余张图片.

### 3.1.2 数据预处理

考虑 CNN 算法在图像视点变化及对数据波动变化处理能力的不足,如出现目标在图像中不同位置,有不同角度、包含不同背景、以不同光照强度的情况.为了使训练模型具有更好鲁棒性并提高拟合度,在进行数据集训练前,需要对数据进行预处理.对数据集进行数据增强优化策略,随机对图像进行水平翻转、垂直翻转进行位置改变,进行随机亮度、随机对比度调整,并进行图像归一化 Normalize 处理,进而扩充数据集,提高模型泛化能力.

利用 Paddle 内置工具命令对本实验数据集进行数据划分,将数据划分为训练集、验证集和测试集.执行命令需要提前在对应实验 PC 机上预安装 Paddle,通过 pip 安装命令可进行快速安装 Paddle.命令执行可将数据集随机划分为 70% 的训练集,20% 验证集,10% 的测试集.并自动在对应目录下生成 labels.txt,train\_list.txt,val\_list.txt,test\_list.txt 4 个文件,即标签文本、训练集文本、验证集文本和测试集文本.命令行如下:

```
paddlex--split_dataset--format ImageNet--dataset_dir ArTaxOr--val_value 0.2--test_value 0.1.
```

### 3.1.3 模型训练

Paddle 作为基于飞桨核心框架、开发套件及深度全流程的开发工具,内置了 20 余种图像分类模型网络,包括适用于移动端的 MobileNet 系列网络,及适用于服务器端的其他类型,如 ResNet 系列,DenseNet<sup>[16]</sup> 系列等内置模型网络.由微软提出的 ResNet 残差网络解决了神经网络深度增加导致的网络退化问题,缓解了深度神经网络增加深度梯度消失的难题.其模型易优化,性能有了大幅提升.至今 ResNet 都仍是非常稳定的 CNN 骨架网络,而其模型深度也已达 152 层之多.经典流行的 ResNet 虽然在识别效果上有足够精准度,但其模型参数量和训练计算量巨大,在某些实际场景下,这些过于庞大且复杂的模型难以应用于嵌入式设备或移动终端.因此更适合移动端的轻量级 CNN 模型设计就应运而生.MobileNet 系列是轻量级经典网络模型.谷歌在 2017 年提出了 MobileNetV1<sup>[17]</sup>,其最重要的创新是提出了深度可分离卷积(depthwise separable convolution),其主要作用是将传统网络中的卷积分离成两步来进行,分别为 depthwise 和 pointwise.这样的好处是比传统卷积神经网络减少了几倍的计算量,其缺点是会损失一定的精度.MobileNetV1 在网络深度上已经减少到 28 层,但其泛化能力较强,保持了较高的网络鲁棒性.2018 年谷歌提出 MobileNetV2<sup>[18]</sup>,较之 MobileNetV1 引入了 Inverted Residuals 和 Linear Bottlenecks,即反向残差和线性瓶颈,虽然网络为 54 层,但优化了网络结构,模型体积更小,速度更快.

适用于服务器端 ResNet 系列中的 ResNet50\_vd\_ssld 模型网络大小为 103.5 MB,在 ImageNet 数据集上 Top1 准确率有 82.4%,Top5 准确率有 96.1%,相比于此系列的其他模型网络在体积大小适合的同时也保持有更高精度.适用于移动端 MobileNet 系列中的 MobileNetV2 模型网络大小为 15.0 MB,在 ImageNet 数据集上 Top1 准确率有 72.2%,Top5 准确率有 90.7%,相比于此系列的其他模型网络在较小体积的同时也保持有较高精度.因此本实验采用了轻量级较具代表性 MobileNet 系列中的模型小预测快速的 MobileNetV2 与经典模型 ResNet 系列中模型精度高的 ResNet50\_vd\_ssld 两种模型网络进行数据集训练.实验 MobileNetV2 网络与 ResNet 网络采用在 ImageNet 上训练模型作为预训练模型,并对 MobileNetV2 网络不进行加载预训练模型作为参照组.针对本实验数据集,对 Paddle 训练接口的内置参数进行 fine-tune.最主要核心参数调整如下:

pretrain\_weights(str):即预训练模型.若为指定路径,则加载路径下预训练模型;若为字符串"IMAGENET",则自动下载 imagenet 数据集的预训练模型权重;若为 None,则不进行预训练模型加载.默认为"IMAGENET".本实验使用"IMAGENET"为预训练模型,并采用 None 作为不进行加载预训练模型的参照组.

num\_epochs(int):训练迭代轮数,即模型对所有样本数据训练集训练一轮就是一个 epoch.本实验参数设置为 8.

learning\_rate(float):本实验采用默认初始学习率 0.025.

train\_batch\_size(int):训练数据批大小,默认为 64.batch\_size 设定大小跟实验环境所需的 GPU 或 CPU 相关,设定值越大,对相应显存或内存消耗越大.本实验设置参数为 100.

save\_interval\_epochs(int):模型保存迭代轮数间隔,本实验采用默认值 1,即为每一轮进行一次模型保存.

lr\_decay\_gamma(float):默认优化器的学习率衰减率.默认为 0.1.

lr\_decay\_epochs(list):默认模型优化器的学习率衰减轮数,默认为[30,60,90].本实验采用[4,6,8],即学习率在第 4 个 epoch,第 6 个 epoch 和第 8 个 epoch 进行衰减.每一次衰减为上次学习率与 lr\_decay\_gamma 的乘积,即每次衰减为之前学习率的 0.1.

### 3.1.4 模型训练结果

基于不同预训练模型的训练预测值汇总如表 1 所示.

表 1 基于不同预训练模型的训练结果

Tab. 1 Training results based on different pre training models

Model	pretrain_weights	Top1(eval)	Top5(eval)
MobileNetV2	None	0.294 6	0.861 1
MobileNetV2	ImageNet	0.746 0	0.982 1
ResNet50_vd_sslid	ImageNet	0.885 2	0.995 8

对比不加载预训练模型 MobileNetV2 模型网络与加载 ImageNet 模型的 MobileNetV2 模型网络,Top1 预测精准度分别为 0.294 6 和 0.746 0,Top5 预测精准度分别为 0.861 1 和 0.982 1.训练结果可以明显看出,运用迁移学习加载预训练模型的网络能更高效对目标数据集进行模型训练,且结果更加收敛,鲁棒性高.对比加载 ImageNet 的 MobileNetV2 模型网络和 ResNet50\_vd\_sslid 模型网络,在 Top1 预测精准度上,ResNet50\_vd\_sslid 比 MobileNetV2 高近 14 个百分点,在 Top5 预测精准度上,ResNet50\_vd\_sslid 比 MobileNetV2 高 1.37 个百分点.训练的结果可以明显看出,ResNet50\_vd\_sslid 模型网络在同等条件下训练所得模型各个指标均高于 MobileNetV2 模型网络.但若考虑其二者的模型大小、模型训练预测时长及预测 Top5 准确率的差异,MobileNetV2 模型网络其综合性能比更高,相对于 ResNet50\_vd\_sslid 模型网络更适用于部署发布应用,即符合 MobileNetV2 适用于移动端开发部署的要求.

### 3.1.5 模型结果保存

Paddle 在训练时,可根据训练接口的 save\_interval\_epoch 参数进行保存模型,保存模型目录中的 model.pdparams 包含"\_model\_", "\_params\_"和"model.yml"3 个文件."\_model\_"即为网络结构的相关信息, "\_params\_"为包括网络模型的参数权重信息.这两个文件也是作为后续 EasyEdge 开源平台模型部署发布所必需的信息文件.

## 3.2 端到端昆虫图像目标检测

目标检测的产生源于传统 CNN 网络无法对图像内对应目标进行定位,只能进行相应的分类且不能对多目标物体进行检测.而目标检测算法的出现和发展,大大加速了图像目标定位的实现.目标检测实现了图像中目标对象的定位和分类两重性问题,且更进一步解决了多目标的分类及定位问题.使得图像识别应用领域更广泛,运用场景更加多元化.其算法核心在于对定位目标的锚点 Anchor 设定,即选取候选框.针对 Anchor 设定不同,大致可分为 Anchor-Based 算法和 Anchor-Free 算法两大类.Anchor-Based 算法中针对 Anchor 是否和目标分类同时进行检测,可分为两阶段算法和单阶段算法.两阶段算法指先在图像上选取候选框,再对候选框目标进行分类,其代表是 Faster R-CNN<sup>[19]</sup>,提出 RPN(Region proposal network 区域建议网络)代替 SS(selective search),提高精度同时还大大减少了目标候选框数量,提升了速度.单阶段算法的代表是 YOLO<sup>[20]</sup>算法,单阶段算法速度远大于两阶段算法,且对于目标定位的精准度也没有较大损失.

### 3.2.1 数据准备

采用百度开源 Insect\_det 数据集作为目标检测数据,此数据集涵盖共 6 大种类,包括 leconte,boerner,

armandi, linnaeus, coleoptera, acuminatus, 共计 217 张标注图像。

其数据标注采用 VOC 数据格式, 即每张图像含有同名 XML 文件, 其包括图像尺寸大小, 目标物体类别、坐标等信息, 具体字段如下所示:

```

图像名称:                <width>647</width>
  <filename>0001.jpg</filename>    <height>647</height>
图像尺寸, 包括宽度、高度和深度:    <depth>3</depth>
  <size>                            </size>
目标名称及坐标, 还有一些如目标对象姿态、目标被截断原因、和识别难易程度等非必须字段:
  <object>                            <xmin>362</xmin>
    <name>leconte</name>                <ymin>224</ymin>
    <pose>Unspecified</pose>            <xmax>399</xmax>
    <truncated>0</truncated>          <ymax>287</ymax>
    <difficult>0</difficult>          </bndbox>
    <bndbox>                            </object>

```

### 3.2.2 数据预处理

考虑目标检测数据集作为一个小样本集, 为使训练模型具有更好鲁棒性并提高拟合度, 在进行训练前, 也对其进行预处理。对数据集进行数据增强优化策略, 如随机对图像进行水平翻转、垂直翻转进行位置改变, 进行随机亮度、随机对比度调整, 进而扩充数据集, 并对数据集按照 70% 的训练集, 20% 验证集, 10% 的测试集划分。

### 3.2.3 模型训练

框架采用 Paddle, 其内置包含如 FasterRCNN 及 Yolo 系列等具有代表性的目标检测模型网络, 本文采用两阶段具有代表性的 FasterRCNN-ResNet50 网络, 及单阶段 Yolo 系列的 YoloV3-MobileNetV3, PPYOloTiny-MobileNetV3 网络进行训练。

其各自核心参数是 FasterRCNN-ResNet50: 运用两阶段 FasterRCNN 与以 ResNet50 为主干网络, 采取在 COCO 数据集上的预训练模型, 输入图像尺寸为  $800 \times 1\ 333$ , 迭代轮数为 12, 学习率为 0.002 5, Batch size 为 2, 学习衰减轮数为 [8, 11]; YoloV3-MobileNetV3: 运用单阶段 YoloV3 与以 MobileNetV3 为主干网络, 采取在 COCO 数据集上的预训练模型, 输入图像尺寸为  $608 \times 608$ , 迭代轮数为 100, 学习率为 0.004 167 7, Batch size 为 8, 学习衰减轮数为 [50, 90]; PPYOloTiny-MobileNetV3: 运用单阶段 PPYOloTiny 与以 MobileNetV3 为主干网络, 采取在 COCO 数据集上的预训练模型, 输入图像尺寸为  $608 \times 608$ , 迭代轮数为 100, 学习率为 0.004 167 7, Batch size 为 8, 学习衰减轮数为 [50, 90]。

与图像分类衡量网络准确率的 Top1 与 Top5 不同, 目标检测的衡量网络性能有几个常见指标: IoU 阈值: 目标检测的测量符合条件依据 IoU 阈值判定, 其表示预测框和真实标签框的交集与预测框和真实标签框的并集比值, 常规设定为 0.5,  $T_P$  表示正样本被识别正确的总量,  $F_P$  表示负样本被错误识别为正样本的总量即误报量,  $F_N$  表示正样本被识别为负样本的总量即漏报量, precision( $P$ ) 表示找出正确框数量概率, 即精准度, recall( $R$ ) 表示找出正确框的比例, 即召回率,  $P_A$  表示各类别目标平均精度值, 也是由 PR 曲线面积所表示。PR 曲线面积表示精准度和召回率所形成的面积, 用积分计算,  $m_{P_A}$  表示所有类别的  $P_A$  均值, 公式如下所示:

精准度 precision( $P$ ) 公式为

$$P = \frac{T_P}{T_P + F_P},$$

召回率 recall( $R$ ) 公式为

$$R = \frac{T_P}{T_P + F_N},$$

$$P_A = \int_0^1 P(R) dr, m_{P_A} = \frac{\sum_{n=1}^N AP(n)}{N}.$$

### 3.2.4 模型训练结果

采用 COCO 预训练模型的 FasterRCNN-ResNet50 网络训练结果如图 1 所示。

混淆矩阵							
类别 (Class)	leconte	boerner	armandi	linnaeus	coleoptera	acuminatus	back_ground
leconte	51.0	4.0	0.0	0.0	0.0	0.0	1.0
boerner	5.0	31.0	0.0	0.0	0.0	0.0	0.0
armandi	2.0	0.0	19.0	11.0	1.0	0.0	0.0
linnaeus	0.0	1.0	9.0	27.0	0.0	2.0	0.0
coleoptera	0.0	0.0	0.0	0.0	18.0	2.0	2.0
acuminatus	0.0	0.0	2.0	3.0	17.0	16.0	1.0
back_ground	4.0	4.0	0.0	0.0	1.0	0.0	0.0
平均精准率的均值 (mAP)							
0.8782							
整体检测评估结果							
类别 (Class)	精准率 (Precision)	召回率 (Recall)	平均精准率 (Average Precision)				
leconte	0.7903	0.9608	0.8890				
boerner	0.7750	0.9688	0.9694				
armandi	0.6333	0.9048	0.8633				
linnaeus	0.6585	0.9310	0.9001				
coleoptera	0.4865	0.9000	0.8803				
acuminatus	0.8000	0.6957	0.7672				

图1 FasterRCNN-ResNet50网络训练结果

Fig.1 FasterRCNN-ResNet50 training results

运用 FasterRCNN-ResNet50 网络训练在精准率上最高的一类为 acuminatus 的 0.800 0,最低的一类为 coleoptera 的 0.486 5;在召回率上最高的一类为 boerner 的 0.968 8,最低的一类为 acuminatus 的 0.695 7;在 AP 上最高一类为 boerner 的 0.969 4,最低一类为 acuminatus 的 0.767 2.6 大类别整体平均精准率即 mAP 为 0.878 2.

采用 COCO 预训练模型的 YoloV3-MobileNetV3 网络训练结果如图 2 所示。

混淆矩阵							
类别 (Class)	leconte	boerner	armandi	linnaeus	coleoptera	acuminatus	back_ground
leconte	47.0	1.0	0.0	0.0	0.0	0.0	4.0
boerner	2.0	28.0	0.0	0.0	0.0	0.0	2.0
armandi	0.0	0.0	17.0	5.0	0.0	0.0	0.0
linnaeus	0.0	0.0	7.0	26.0	0.0	0.0	0.0
coleoptera	0.0	0.0	0.0	0.0	17.0	2.0	2.0
acuminatus	0.0	0.0	1.0	5.0	8.0	13.0	2.0
back_ground	0.0	0.0	0.0	0.0	0.0	0.0	0.0
平均精准率的均值 (mAP)							
0.8772							
整体检测评估结果							
类别 (Class)	精准率 (Precision)	召回率 (Recall)	平均精准率 (Average Precision)				
leconte	0.9592	0.9216	0.9034				
boerner	0.9655	0.8750	0.9691				
armandi	0.6800	0.8095	0.8475				
linnaeus	0.7222	0.8966	0.8517				
coleoptera	0.6800	0.8500	0.9196				
acuminatus	0.8667	0.5652	0.7721				

图2 YoloV3-MobileNetV3网络训练结果

Fig.2 YoloV3-MobileNetV3 training results

运用 YoloV3-MobileNetV3 网络训练在精准率上最高的一类为 boerner 的 0.965 5,最低的为 coleoptera 与 armandi 两类,同为 0.680 0;在召回率上最高的一类为 leconte 的 0.921 6,最低的一类为 acuminatus 的 0.565 2;在 AP 上最高一类为 boerner 的 0.969 1,最低一类为 acuminatus 的 0.772 1.6 大类别整体平均精准率即 mAP 为 0.877 2.

采用 COCO 预训练模型的 PPYOloTiny-MobileNetV3 网络训练结果如图 3 所示。

运用 PPYOloTiny-MobileNetV3 网络训练在精准率上最高的一类为 leconte 的 0.770 5,最低的一类为 armandi 的 0.380 0;在召回率上最高的一类为 linnaeus 的 1.000 0,最低的一类为 acuminatus 的 0.478 3;在 AP 上最高一类为 boerner 的 0.932 0,最低一类为 armandi 的 0.398 2.6 大类别整体平均精准率即 mAP 为 0.712 1.

对比目标检测网络训练结果可知,由 FasterRCNN-ResNet50 及 YoloV3-MobileNetV3 训练所得模型网络 mAP 指标较好,考虑模型大小、训练时长及模型响应时间等性能指标,YoloV3-MobileNetV3 训练模型更适用于部署发布应用,即适用于边缘设备开发部署的要求。

### 3.2.5 模型结果保存

Paddle 在训练时,可根据训练接口的 `save_interval_epoch` 参数进行保存模型,与目标分类产生文件信息不同,网络结构信息文件名为 `model.pdmodel`,网络参数文件为 `model.pdiparams`.这两个文件也是作为后续 EasyEdge 开源平台模型目标检测部署发布所必需的信息文件.

混淆矩阵	类别 (Class)	leconte	boemer	armandi	linnaeus	coleoptera	acuminatus	back_ground
	leconte	48.0	11.0	2.0	1.0	0.0	0.0	3.0
	boemer	12.0	31.0	0.0	0.0	0.0	0.0	1.0
	armandi	0.0	0.0	19.0	18.0	2.0	0.0	1.0
	linnaeus	0.0	1.0	27.0	29.0	3.0	0.0	0.0
	coleoptera	0.0	0.0	0.0	0.0	18.0	17.0	2.0
	acuminatus	0.0	0.0	2.0	3.0	22.0	11.0	0.0
	back_ground	1.0	0.0	0.0	0.0	1.0	0.0	0.0
平均精准率的均值 (mAP)		0.7121						
整体检测评估结果								
	类别 (Class)	精准率 (Precision)	召回率 (Recall)	平均精准率 (Average Precision)				
	leconte	0.7705	0.9216	0.8815				
	boemer	0.7209	0.9688	0.9320				
	armandi	0.3800	0.9048	0.3982				
	linnaeus	0.5686	1.0000	0.7512				
	coleoptera	0.3913	0.9000	0.7996				
	acuminatus	0.3929	0.4783	0.5101				

图3 PPYOloTiny-MobileNetV3网络训练结果

Fig. 3 PPYOloTiny-MobileNetV3 training results

## 4 模型到端边缘的部署实现

训练完成的模型往往并不能直接在端的产品级应用,需要根据不同型号不同硬件配置的服务端或移动端来进行针对性的部署和发布,如较为便捷的方式利用 Docker 开源应用容器进行打包,将应用包到一个可移植的镜像中,其可在 Windows 或 Linux 系统上发布,但其最大一个局限性是不能在 32 bit 的 Linux 或 Windows 环境下使用.而开源 EasyEdge 可以解决所有关于模型最终发布为端产品的局限性问题.提供了统一方案和开源平台,让开发者能更高效将训练模型发布于 APP 产品或应用于服务端,嵌入式设备中.

本实验采用百度开源服务平台 EasyEdge,可将训练好的模型快速转化并发布成到端的模型应用.其产品包含模型中心,纯离线服务,端协同服务三大功能模块部分.模型应用部分包括自建模型和开源模型两类.开源模型涵盖近 10 多种开源模型,包括图像分类及目标检测等.本实验利用自建模型功能,训练模型发布使用方式如下,选择模型中心的上传本地模型模块.如图 4 所示.

图4 EasyEdge界面

Fig. 4 EasyEdge interface

模型类型内置图像分类、物体检测、人脸检测、姿态估计、语义分割 5 种主流方式,模型框架涵盖了几乎所有现今主流框架,如 caffe, TensorFlow, PyTorch 等,且其对应框架下也几乎涵盖目前主流所有模型网络。如图像分类中的 AlexNet, VGG 系列, MobileNet 系列, ResNet 系列等近 20 余种,物体检测中的 SSD, FasterRCNN, Yolo 系列, PP-Yolo 系列等。

部署昆虫图像分类模型网络:在自定义模型名称中将本实验模型名称定为 kaggle-Insect。本实验选择模型的类型为图像分类,采用 PaddlePaddle 框架,模型网络使用 MobileNet 系列的 MobileNetV2。网络结构采用 MobileNetV2 训练模型的\_model\_,网络参数使用 MobileNetV2 训练模型的\_params\_,模型标签使用对应实验数据集的标签文本即 label.txt。选好相应配置及上传相关内置参数文件后,点击校验并提交模型功能按钮。

部署昆虫目标检测模型网络:在自定义模型名称中将本实验模型名称定为 Yolo-Insect。本实验选择模型的类型为目标检测,采用 PaddlePaddle 框架,模型网络使用 Yolo 系列的 YoloV3。网络结构采用 YoloV3-MobileNetV3 训练模型的 model.pdmodel,网络参数使用 YoloV3-MobileNetV3 训练模型的 model.pdiparams,模型标签使用对应实验数据集的标签文本即 label.txt。选好相应配置及上传相关内置参数文件后,点击校验并提交模型功能按钮。

在平台进行系统校验后会分配模型 ID 及显示对应模型名称,模型类型以及后续 SDK 操作功能。应用生成端模型功能,会显示相应发布离线 SDK 界面(如图 5)。端模型提供模型网络相适用的各类芯片硬件、操作系统。其芯片硬件涵盖目前市面主流各类厂商产品,包括服务端与移动端,如通用 x86 芯片、通用 ARM 芯片、苹果 A-Bionic、高通骁龙 GPU 与 DSP、华为海思 NPU、英特尔系列处理器、英伟达 GPU 等。操作系统也包括目前主流的移动端 Android、IOS 与服务器端 Windows、Linux。并且可支持单模型多版本 SDK 发布。

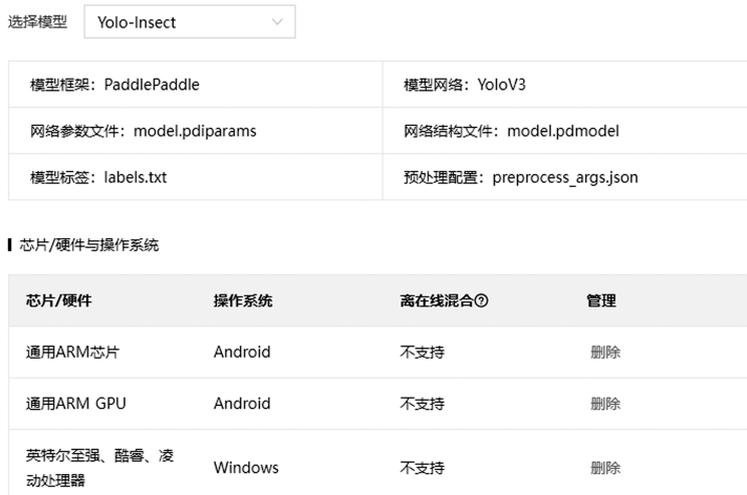


图5 发布离线SDK界面

Fig. 5 Publish offline SDK interface

在进行多版本设置进行发布后,平台系统会自动校验并生成对应版本 SDK。在全部模型界面下可以看到目前版本生成状态。已发布的模型版本可以进行获取 SDK, Demo 体验下载等操作。运用手机浏览器进行对应 SDK 二维码扫描即可下载对应版本 SDK 应用。最终实现模型到边缘产品的发布流程。

## 5 总 结

本文在昆虫种类识别领域技术基础方面对卷积神经网络 CNN 及其发展的脉络进行说明,并对目前深度学习应用于图像分类、目标检测方面的主流模型网络进行梳理,介绍了迁移学习相关的原理,结合目前昆虫种类研究识别现状,提出了一种模型到边缘的落地场景的全流程开发应用实现方法。以深度学习图像分类及目标检测两方面主流模型网络为主,本文在图像分类上采用 kaggle 开源昆虫数据集作为实验数据集,运

用 Paddle 开源框架工具,采用基于 ImageNet 的预训练模型并对比两种目前比较流行的卷积神经网络结构 MobileNetV2 和 ResNet50\_vd\_sslsd 进行迁移学习训练,在目标检测上采用百度开源昆虫数据集作为实验数据集,采用基于 COCO 的预训练模型并对比两阶段有代表性的 FasterRCNN 及单阶段具有代表性的 Yolo 系列两种网络进行迁移学习训练,并对其实验效果较好的图像分类中 MobileNetV2 训练模型及目标检测中 YoloV3-MobileNetV3 训练模型在 EasyEdge 上进行部署发布,生成了应用于移动端的 SDK 版本,完成了从模型构建到端的落地场景的全流程开发应用实现,为后续昆虫图像识别在场景化应用方面的研究提供依据参考,本实验也存在一定局限性,如在数据集上未采用细分类数据集,只是划分到目级,对于科级和种级未进行探究实验,对于像昆虫种类较为细化的图像识别领域训练模型不够细化,目标检测场景化数据性较弱,这也是今后需要继续深入研究的方向。

## 参 考 文 献

- [1] 廖肖依,肖芬.昆虫标本的采集、制作和保存方法[J].现代农业科技,2012(6):42-43.  
LIAO X Y, XIAO F. Collection, production and preservation of insect specimen [J]. Modern Agricultural Science and Technology, 2012(6):42-43.
- [2] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86(11):2278-2324.
- [3] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6):84-90.
- [4] CYBENKO G. Approximation by superpositions of a sigmoidal function [J]. Mathematics of Control, Signals and Systems, 1989, 2(4):303-314.
- [5] AGARAP A F. Deep learning using rectified linear units (ReLU) [EB/OL]. [2022-03-03]. <https://arxiv.org/abs/1803.08375>.
- [6] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. [2022-03-03]. <https://arxiv.org/abs/1409.1556>.
- [7] Computer Vision Foundation. Going deeper with convolutions [C]. Going deeper with convolutions. Boston: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015:1-9.
- [8] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C] // IEEE Conference on Computer Vision and Pattern Recognition. Washington D C: IEEE Computer Society, 2016:770-778.
- [9] 彭明杰. 基于多模态输入卷积神经网络的蜻蜓识别算法 [J]. 电子世界, 2021(2):100-102.  
PENG M J. Dragonfly recognition algorithm based on multimodal input convolutional neural network [J]. Electronics World, 2021(2):100-102.
- [10] 张瑞瑞, 夏浪, 陈立平, 等. 基于 U-Net 网络和无人机影像的松材线虫病变色木识别 [J]. 农业工程学报, 2020, 36(12):61-68.  
ZHANG R R, XIA L, CHEN L P, et al. Recognition of wilt wood caused by pine wilt nematode based on U-Net network and unmanned aerial vehicle images [J]. Transactions of the Chinese Society of Agricultural Engineering, 2020, 36(12):61-68.
- [11] 薛大暄, 张瑞瑞, 陈立平, 等. 基于 Faster R-CNN 的美国白蛾图像识别模型研究 [J]. 环境昆虫学报, 2020, 42(6):1502-1509.  
XUE D X, ZHANG R R, CHEN L P, et al. Faster R-CNN based image recognition research of Hyphantria cunea [J]. Journal of Environmental Entomology, 2020, 42(6):1502-1509.
- [12] 王卫民, 符首夫, 顾榕蓉, 等. 基于卷积神经网络的虫情图像分割和计数方法 [J]. 计算机工程与科学, 2020, 42(1):110-116.  
WANG W M, FU S F, GU R R, et al. An insect image segmentation and counting method based on convolutional neural network [J]. Computer Engineering & Science, 2020, 42(1):110-116.
- [13] 王茂林, 荣二花, 张利军, 等. 基于图像处理的蓟马计数方法研究 [J]. 山西农业科学, 2020, 48(5):812-816.  
WANG M L, RONG E H, ZHANG L J, et al. Study on counting Thripidae Thysanoptera based on image processing [J]. Journal of Shanxi Agricultural Sciences, 2020, 48(5):812-816.
- [14] 陈彦彤, 陈伟楠, 张献中, 等. 基于深度卷积神经网络的蝇类面部识别 [J]. 光学精密工程, 2020, 28(7):1558-1567.  
CHEN Y T, CHEN W N, ZHANG X Z, et al. Fly facial recognition based on deep convolutional neural network [J]. Optics and Precision Engineering, 2020, 28(7):1558-1567.
- [15] 张银松, 赵银娣, 袁慕策. 基于改进 Faster-RCNN 模型的粘虫板图像昆虫识别与计数 [J]. 中国农业大学学报, 2019, 24(5):115-122.  
ZHANG Y S, ZHAO Y D, YUAN M C. Insect identification and counting based on an improved Faster-RCNN model of the sticky board image [J]. Journal of China Agricultural University, 2019, 24(5):115-122.
- [16] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition. Washington D C: IEEE Computer Society, 2017:2261-2269.

- [17] HOWARD A G, ZHU M L, CHEN B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[EB/OL]. [2022-09-03]. <https://arxiv.org/abs/1704.04861>.
- [18] SANDLER M, HOWARD A, ZHU M L, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D C: IEEE Computer Society, 2018: 4510-4520.
- [19] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [20] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Washington D C: IEEE Computer Society, 2016: 779-788.

## Research on application of insect species image recognition based on convolutional neural network

Wei Fuyu<sup>1</sup>, Zhang Zhenyu<sup>2a</sup>, Liang Guizhen<sup>2b</sup>

(1. School of Landscape and Ecological Engineering, Hebei University, Handan 056000, China;

2. a. Academy of Fine Arts; b. School of Mathematics and Statistics, Xinxiang University, Xinxiang 453000, China)

**Abstract:** Image recognition of insect species is an important way of agricultural intelligent identification of pests, and accurate and efficient identification of insect species is the premise of targeted pest control. In this paper, we use the insect datasets ArTaxOr and Insect\_det, based on image classification under convolutional neural network, such as MobileNet and ResNet, and target detection under convolutional neural network, such as FasterRCNN and YOLO technology, uses transfer learning to train the model, compares and analyzes the training results to obtain the optimal image recognition model of insect species. The constructed optimal model is deployed on EasyEdge platform, thus realizing the whole process development mode from model to end, which provides a basis and reference for the follow-up research on the scene application of insect species image recognition.

**Keywords:** insect image recognition; convolutional neural network; image classification; target detection; model scene application

[责任编辑 陈留院 赵晓华]