

项目详细方案

---

基于改进的 Cascade RCNN 的农田害虫图像识别方案

## 摘要

病虫害一直是农业生产中无法避免的问题，每年都会由此造成巨大的经济损失。传统害虫诊断过程依赖农业专家的知识，人工识别耗时且主观性较强、容易出错，难以达到现代农业的自动化需求。而相较于传统依附于人工主观识别分析的方法，深度学习是一种新兴的图像处理和数据分析技术，其中深度卷积神经网络在处理图像、视频等方面取得突破性进展，其在农业领域的应用引起广泛关注。深度学习识别方法是相对于传统图像识别方法而言的，该方法主要是进行底层特征提取，学习大量样本数据从而得到深层的数据特征，识别精度高、速度快。因此，使用人工智能实现自动化农田害虫检测是不可逆转的时代趋势，研究针对农田害虫图像检测模型具有非常重要的现实意义。

针对问题一，本文根据赛题提供的数据集作为数据源，选择了 Cascade RCNN、Faster RCNN、SSD 三大模型架构作为预选基础检测网络，以 ResNet50 作为骨干网络，进行模型训练，对比三种模型架构的性能。其中 Cascade RCNN 的 mAP 在 FPS 为 4.2 的情况下，mAP 高达 53.2；Faster RCNN 的 mAP 为 43.5，FPS 为 19.2；SSD 的 mAP 为 29.9，FPS 为 48.1。为了更好地解决害虫识别问题，本文最终选择了 mAP 最高的 Cascade RCNN 作为基础检测网络运用到实际目标检测中。

针对问题二，在进行害虫检测的时候，由于害虫的体积尺度不均衡，容易造成小目标漏检的情况，针对小体积尺度害虫的检测问题，在选择出最佳基础检测网络后，本文提出了一种基于 Cascade RCNN 改进模型对农田尺度不均衡的害虫进行检测，避免漏检的情况发生。选用 Cascade RCNN 作为基础检测网络，再选用残差神经网络 ResNetX(X=50, 101, 152) 作为骨干网络进行训练。通过模型训练，最终选出最优的骨干网络 ResNet152 进行最后的预测，并且针对害虫尺度不均衡的问题加入了 FPN，将 ResNet152 输出的具有大量位置信息的浅层特征层与和具有丰富语义信息的深层特征层进行均衡化融合，最终完美完成图像分类与位置预测。实验结果表明，采用这一基于 Cascade RCNN 改进模型，农田害虫图像识别检测的 mAP 达到 56.3，FPS 为 4.1，完美达到了农田害虫检测作业。

针对问题三，本文通过模型训练与评估，得到性能最佳的 Cascade RCNN-ResNet152-FPN 作为最终农田害虫检测模型。利用此网络模型结构对数据集中的各类害虫数量进行统计，并得到各类害虫的数量分布。农户及专家可根据此结果对实际农田培育时采用针对性化学药剂、防蛾灯等农田虫害防治方法，对实际农田害虫防范具有指导性意义。

关键词：害虫检测；目标检测；Cascade RCNN；FPN；ResNet

## 目录

一、引言 .....	5
1.1 研究背景与意义 .....	5
1.2 害虫识别研究现状 .....	5
1.3 研究目标 .....	6
1.4 研究内容与结构 .....	6
二、数据集分析 .....	8
2.1 数据集介绍 .....	8
2.2 数据集可视化分析 .....	8
2.3 数据预处理 .....	10
2.3.1 筛选错误标签 .....	10
2.3.2 数据重采样 .....	11
2.3.3 数据增广 .....	12
2.3.4 图像归一化 .....	13
2.3.5 数据集预处理总结 .....	13
2.4 划分数据集 .....	13
2.5 本章小结 .....	14
三、基础检测网络的确定 .....	15
3.1 相关技术 .....	15
3.1.1 ResNet50 .....	15
3.1.2 Cascade RCNN .....	16
3.1.3 Faster RCNN .....	16
3.1.4 SSD .....	18
3.2 实验结果 .....	18
3.3 本章小结 .....	19
四、基于 Cascade RCNN 的农田害虫检测改进算法 .....	21
4.1 相关技术 .....	21
4.1.1 FPN .....	21
4.1.2 ResNet .....	22

4.2 网络结构 .....	22
4.2 采用的 Tricks .....	23
4.3 实验步骤 .....	24
4.3.1 实验环境 .....	24
4.4 实验结果及分析 .....	25
4.4.1 Epoch-mAP 分析 .....	25
4.4.2 Epoch-Loss 分析 .....	25
4.4.3 mAP-FPS 分析 .....	26
4.5 本章小结 .....	26
五、模型检测效果以及农田害虫防范建议 .....	28
5.1 模型检测结果 .....	28
5.2 农田害虫防范建议 .....	28
六、总结 .....	30
七、未来展望及现实意义 .....	31

## 一、引言

### 1.1 研究背景与意义

2021 年 9 月，国务院公示批复原则同意《全国高标准农田建设规划(2021—2030 年)》。通过实施《规划》，到 2022 年建成高标准农田 10 亿亩，以此稳定保障 1 万亿斤以上粮食产能；到 2025 年建成 10.75 亿亩，并改造提升现有高标准农田 1.05 亿亩，以此稳定保障 1.1 万亿斤以上粮食产能。

为了能够实现这一目标，中国提出了以科学发展观为指导，深入贯彻“公共植保”和“绿色植保”理念，以“防灾减灾、提质增效、保障安全”为目标，以一贯务实作风和高度责任感，认真履行工作职责，扎实做好病虫灾害监测预警和防治指导公共服务，加快推进专业化防治和绿色防控，控制病虫危害，减少化学农药用量，保障农业增产增收和农产品质量安全。

农作物病虫害是世界性范围的农业问题。目前，害虫管理方式主要仍为使用化学农药，对害虫进行消杀，但由于没有做到通知防范、过度使用某一化学农药导致害虫产生抗体、过度使用农药等问题，导致无法有效治理害虫的同时，破坏了生态环境，与科学发展观理念相违背且影响产量。一旦害虫在田间传播，必须及时发现，以便农民及时提供治疗，防止害虫传播。

实时获取虫害信息并定向灭杀是一种高效且环境友好型的方式，在解决虫害问题的同时，降低了对环境的影响。

传统的害虫识别主要是为三点：①依靠人工识别的经验性判断；②根据害虫的颜色、纹理、形状等形态学特征查阅病虫害图谱确定；③询问专家。但这些方法存在着较大的弊端：人工识别的经验性判断和通过害虫图谱的方式由于存在较大主观性，可能会导致判断错误，并错误使用农药；通过询问专家获知害虫种类，再对其使用农药，效率较低。因此，传统的害虫识别方法费时、易出错、繁琐，不利用农户及时解决农田害虫的问题。

使用识别效率高、运行速度快的人工智能识别技术代替传统农田害虫识别，是不可逆转的时代趋势。

### 1.2 害虫识别研究现状

传统的害虫识别的算法，一般集中在提取害虫颜色、形状、纹理等图像特征上，首先对图像进行人工特征提取，再将提取到的特征作为输入，利用机器学习方法进行建模和识别。此前，国内早期大多采用的是对害虫图像的全局特征进行提取，如许振伟基于二阶灰度统计储粮害虫的纹理特征利用 BP 神经网络训练样本，得到害虫的种类和轮廓并且能实现对储粮害虫的快速检测和准确分类；邹修国等基于不变矩提取形状特征值的方法实现了对稻飞虱的分类等等；一些专家利用局部图像特征来判断害虫的类别，虽然局部图像特征提取相比全局特征提取更为精确，但由于两种特征提取的方式仍依赖于人工选取特征，但当需要人工提取多个特征时，其效率会变得很低，模型的优劣取决于特征的选择，存在较大的局限性，且泛化能力较差。

近年来，随着深度学习的发展与广泛使用，一些深度学习算法被应用在农田害虫图像识别中，并在许多害虫检测实际应用中的成绩超越了之前的方法，例如基于深度学习的番茄叶片病虫害分类框架、基于深度学习架构的果蝇识别方法等一系列

模型，其准确率高达 89%及 95.68%。通过分析以往的工作，深度学习方法已被证明可以显著改善害虫分类性能。

### 1.3 研究目标

本次研究主要针对赛方提供的数据集进行训练分析，以常见的 28 种害虫为例，建立常见农田害虫图像的数据集，并对深度学习模型中重要的网络结构和训练参数进行了优化和改进，构建一种精度高、效率快的基于深度学习的农田害虫图像检测模型检测系统替代传统人工识别方法。

### 1.4 研究内容与结构

本文主要内容是通过确定农田害虫图像识别模型结构，进行模型训练，努力提高模型推理准确率。进而可将模型应用于实际生活中，实现实时性强、精度高、效率快的农田害虫图像识别。研究内容如下：

(1) 通过分析原始数据集，更加深入了解原始数据集中各类害虫的属性，从而更好的了解数据集的害虫数量、大小、位置等信息，并对模型结构的确定有一定的指导意义。

(2) 基于深度学习的农田害虫识别与分类方法研究。针对农田害虫识别与分类方法进行研究，对比 Cascade RCNN、Faster RCNN、SSD 三种模型基础检测网络的性能，选择出最佳基础检测网络。

(3) 根据(2)中选择出的效果最佳检测网络，使其作为检测网络头，骨干网络采用 ResNet X 网络，X 分别取 50, 101, 152, 通过对比在其检测效果，确定最终的骨干网络。

(4) 农田害虫图像存在目标较小且重叠度较高的情况，进行图像检测时仍存在漏检的问题，本文使用 FPN 对其进行优化，效果良好。

(5) 根据模型训练效果，最终确定了 Cascade RCNN-ResNet152-FPN 模型作为识别害虫位置及种类的模型和算法。

本论文共分为七个章节，具体章节架构安排如下：

第 1 章：引言。主要阐述本课题的研究背景及意义，分析总结农田害虫图像检测方法的研究现状，说明本文的研究目标、内容及组织结构。

第 2 章：数据集分析。主要对数据集进行人工清洗、可视化分析及预处理。通过对数据集的分析与处理，对后续的确模型结构及训练具有较大的指导意义。

第 3 章：基础检测网络的确定。通过使用 Cascade RCNN、Faster RCNN、SSD 三个模型作为基础检测网络，对比其 mAP 及 FPS 等性能表现，选择最优的基础检测网络结构作为最终的基础检测网络结构。

第 4 章：基于改进的 Cascade RCNN 的农田害虫检测模型。主要使用 Cascade RCNN 作为检测网络头部，加入 FPN 作为颈部网络，解决农田害虫尺度不均衡问题。骨干网络使用 ResNet 50、101、152，通过对比其检测效果来确定最终模型的骨干网络。使其完美完成农田害虫检测任务。

第 5 章：通过模型训练与评估，得到性能最佳的 Cascade RCNN-ResNet152-FPN 作为最终农田害虫检测模型，对不同种类的害虫数量进行统计分析，对该农田的实际种植提出针对性防范建议，降低害虫造成的经济损失，提高种植效益。

第 6 章：总结。总结论文已完成的工作。

第 7 章：未来展望及现实意义。并分析未来工作的方向及对现实农业发展的指导性意义。

## 二、数据集分析

### 2.1 数据集介绍



图 2-1 28 类害虫图鉴

本文的数据集由官方提供图像文件与相应的 csv 标记信息构成，共 576 张有效图片，1019 个有效标记数据，其中包含 28 种农田害虫，原数据集未分为训练集和验证集，具体害虫种类见表 2-1，害虫图鉴见图 2-1。

表 2-1 害虫种类名称及其编号

编号	0	6	7	8	9	10
名称	不含虫子	大螟	二化螟	稻纵卷叶螟	白背飞虱	褐飞虱属
编号	25	41	105	110	115	148
名称	地老虎	蝼蛄	粘虫	草地螟	甜菜夜蛾	黄足猎蝽
编号	156	222	228	235	256	280
名称	八点灰灯蛾	棉铃虫	二点委夜蛾	甘蓝夜蛾	蟋蟀	黄毒蛾
编号	310	387	392	394	398	401
名称	稻螟蛉	紫条尺蛾	水螟蛾	线委夜蛾	甜菜白带野螟	歧角螟
编号	402	430	480	485	673	
名称	瓜绢野螟	豆野螟	石蛾	大黑鳃金龟	干纹冬夜蛾	

### 2.2 数据集可视化分析

为了清晰有效地展示数据集信息，本章将统计得到的数据进行可视化展示，对各类害虫数量、真实框长宽比、真实框长宽进行分析，对后续的模型结构选择具有一定的指导性意义。



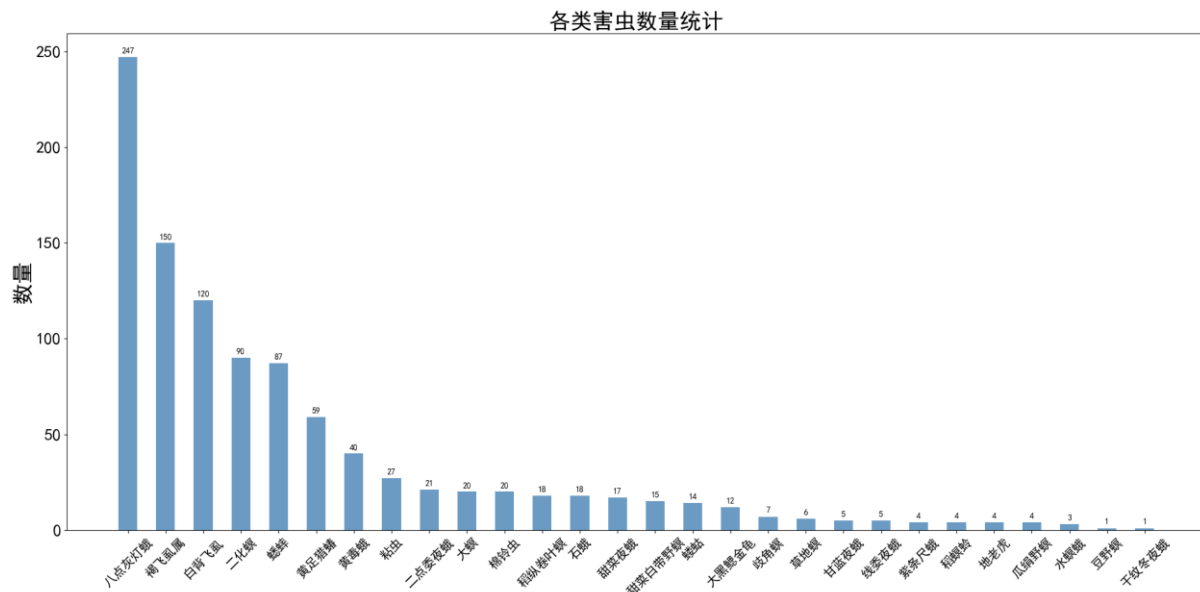


图 2-2 各类害虫数量

数据集中共出现 1019 个标签，各类标签数量统计如图 2-2 所示。从图中可以明显看出，八点灰灯蛾的数量最多，豆野螟、干绞冬夜蛾的数量最少，各类害虫在数量尺度上极度不均衡，可以在数据预处理阶段加强对豆野螟、干绞冬夜蛾等数量较少的害虫种类增加采样比例，使之农田各类害虫在数量尺度上更为均衡。

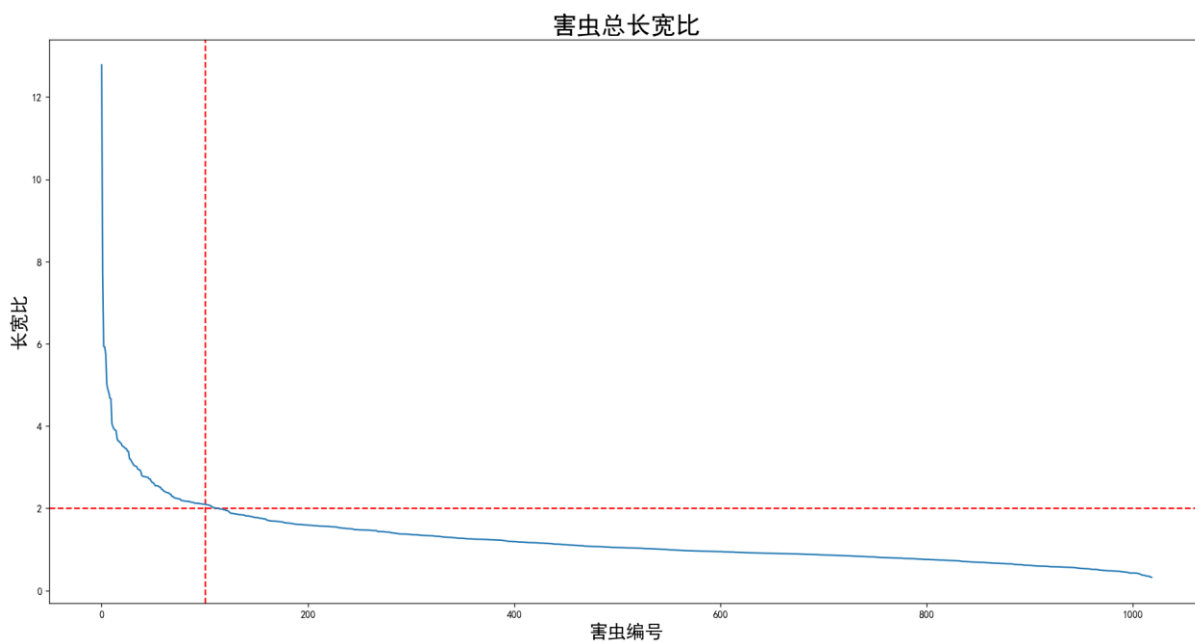


图 2-3 害虫总长宽比

通过分析数据集中各害虫的总长宽比，从而了解数据集中所有害虫的总长宽比范围，如图 2-3 所示，长宽比主要在 0~2 之间，此数据将有利于模型对于预选框长宽比的确定。

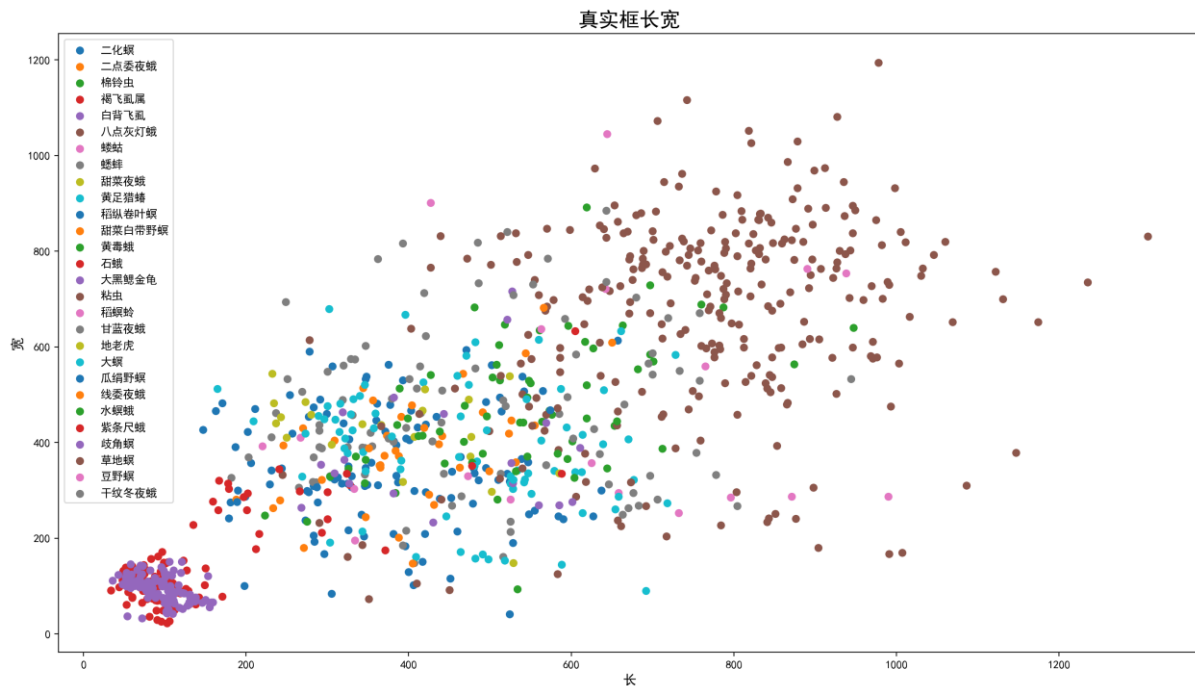


图 2-4 真实框长宽

为了更加深入地了解原始数据集中各类害虫的真实框长宽，通过对各类害虫长宽比进行可视化分析，从而更好地调整各类害虫的预选框，避免因害虫尺度不均衡导致漏检现象，真实框长宽如图 2-4 所示。

由图可知，歧角螟和紫条尺蛾的长宽比较集中，处于（20，20）与（200，200）之间；八点灰灯蛾的长宽分布最为不均衡，最小值约为（380，80）、最大值约为（1350，830），其跨度较大。

通过对各类害虫的真实框长宽分析，再使用 K-Means 聚类算法能够确定模型各预选框的长宽大小。

## 2.3 数据预处理

由于数据集中的数据存在漏标、错标、特征块大小不一等问题，需要经过处理提取出可供模型使用的数据，才能更加精准获取害虫标签信息，并建立能够准确确定害虫位置及种类的模型和算法，流程如图 2-5 所示。



图 2-5 数据预处理流程

本文的数据预处理主要分为以下步骤：

### 2.3.1 筛选错误标签

本文在数据集分析时，发现少量错误标签，便将其清洗出数据集中，防止错误标签对后续模型训练产生消极影响，如数据集中出现误标为黄足猎蝽、紫条尺蛾、褐飞虱属、二化螟的害虫标签，异常标签如图 2-6 所示。



图 2-6 存在问题的数据集

图 2-7 为正常标签与异常标签的对比图。



图 2-7 异常数据集对比

### 2.3.2 数据重采样

在数据集分析时，发现数据集存在数据不均衡问题。模型不均衡会较大程度影响模型训练的结果，若数据集存在各类目标数量尺度不均衡问题，会导致模型对数量较少的虫子识别效果较差，导致漏检的情况发生，此时模型的召全率低。因此调节各类害虫采样率，做到数据均衡，从而达到各类害虫数量之间处于较为平衡的状态，提高了模型的可靠性。

数据重采样结果如图 2-8 所示，由原先农田害虫数量尺度不均衡变为均衡状态，对模型的召全率有较大的提升。

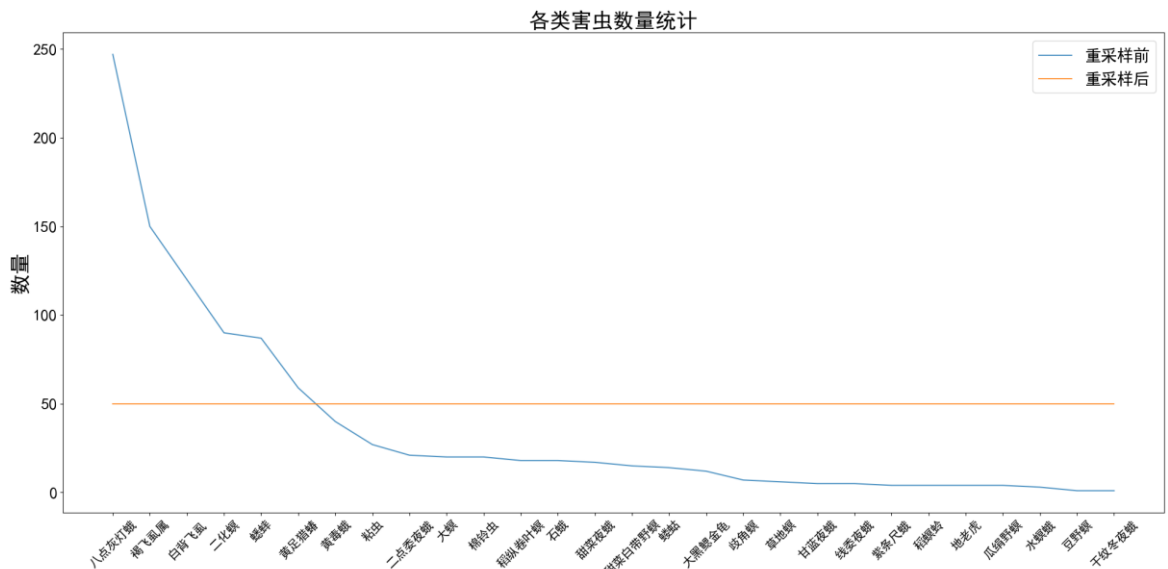


图 2-8 数据重采样

### 2.3.3 数据增广

在图像分类任务中，图像数据的数据增广是一种常用的正则化方法，常用于数据量不足或者模型参数较多的场景。通过扩大数据规模、增加训练数据集，让数据集尽可能的多样化，使得训练的模型具有更强的泛化能力。

深度学习模型需要有大量的标注数据进行支撑。真实应用情景中，经常会出现缺乏标注数据、数据分布不均衡导致模型鲁棒性差、模型性能不佳的问题，利用数据增广来增加模型的稳定性，能在一定程度上解决这些问题。

目前数据增广主要包括：水平、垂直翻转、旋转、缩放、锐化、亮度、裁剪、剪切、平移、对比度、色彩抖动、噪声等，本文主要选取了旋转、对比度、饱和度、亮度、锐化进行数据增广，将原本的 576 张图像增广到 2880 张。每次训练会随机取几种进行数据增广，而每次取几种方式后，其增广效果也是随机的。由于两个阶段都是随机选择增广方式和增广强度，故本文的数据增广有无限种可能。数据增广流程如图 2-9 所示。

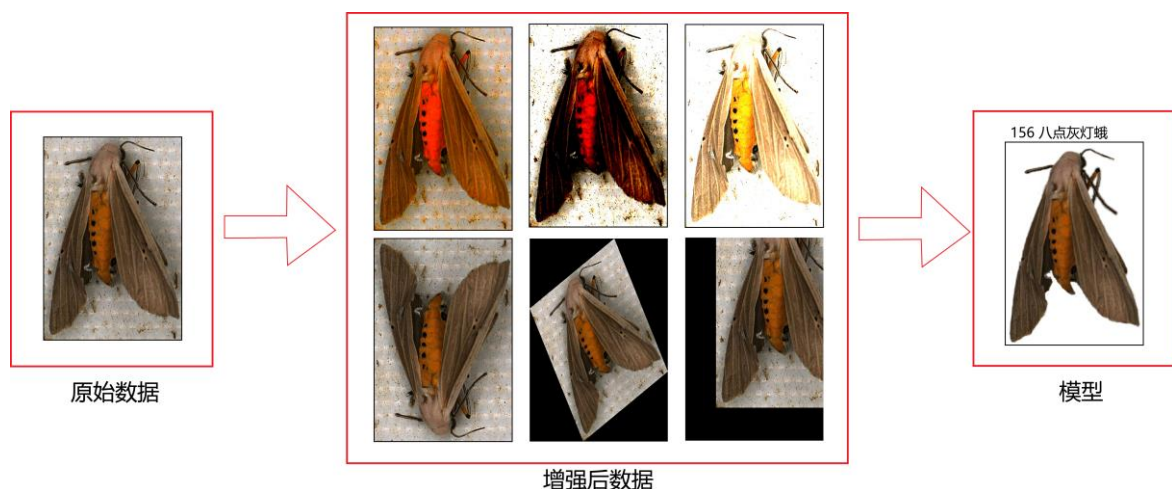


图 2-9 八点灰灯蛾数据增广流程图



## 2.5 本章小结

通过清洗错误标签、数据重采样、数据增广、图像归一化这四个步骤完成数据预处理，得到预处理后各类害虫数量，便于后续模型的选择与训练。最后将预处理完后的图像与标签数据按照 7:3 划分为训练集和验证集，完成数据集预处理。

### 三、基础检测网络的确定

#### 3.1 相关技术

##### 3.1.1 ResNet50

为了选择出最佳的农田害虫识别与分类模型，本文选取了 Cascade RCNN、Faster RCNN、SSD 三个模型进行训练，并在其中选择出能够准确确定害虫位置及种类的最佳模型。为保证单一变量原则，本文将三个模型对比时，骨干网络均选择了 ResNet50。

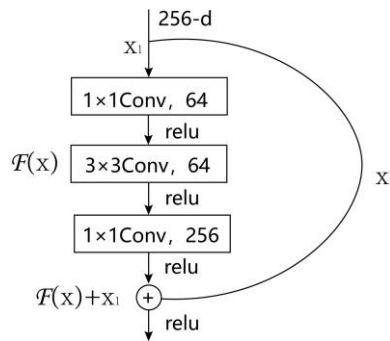


图 3-1 ResNet50 网络的残差单元结构图

ResNet50 是经典的运用残差单元结构组成的网络，是在 2015 年由何凯明等人共同提出，旨在解决因卷积神经网络不断加深而可能导致的梯度消失和网络退化的问题。其网络层次如图 3-2 所示。ResNet50 包括一个步长为 2 的  $7 \times 7$  卷积层、四个残差单元级联卷积层和一个全连接层。其中残差单元是由一个  $3 \times 3$  的卷积核、两个  $1 \times 1$  的卷积核和一个残差连接分支组成，其结构如图 3-1 所示。其中： $x_1$  为输入，其期望输出为  $h(x)$ ，但加入一个恒等映射  $x_1$  后原始学习特征变成  $h(x) = (f(x) + x_1)$ ，因此残差是指  $f(x) = h(x) - x_1$ ，使得网络拟合更容易，直接映射的加入也使得网络的下一层一定比前一层的图像信息多。ResNet50 不仅计算量比 VGG16 更少，而且能提取更多害虫的原始深层特征信息。

Layer name	Output size	50-Layer
Conv1	112×112	$7 \times 7, 64, \text{stride}=2$
Conv2_x	56×56	$3 \times 3 \text{ max pool, stride}=2$
		$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
Conv3_x	28×28	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$
Conv4_x	14×14	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$
Conv5_x	7×7	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	Average pool, 1000-d fc, softmax
FLOPs		$3.8 \times 10^9$

图 3-2 ResNet50 网络结构



### 3.1.2 Cascade RCNN

Cascade RCNN 整体检测流程是通过骨干网络提取特征再通过 FPN 组合特征图，生成更加易于分类与定位的特征图组，RPN 结构按尺度在对应的特征图上生成候选框，最终由多个 ROI Pooling 结构再对候选框进行多次框回归。

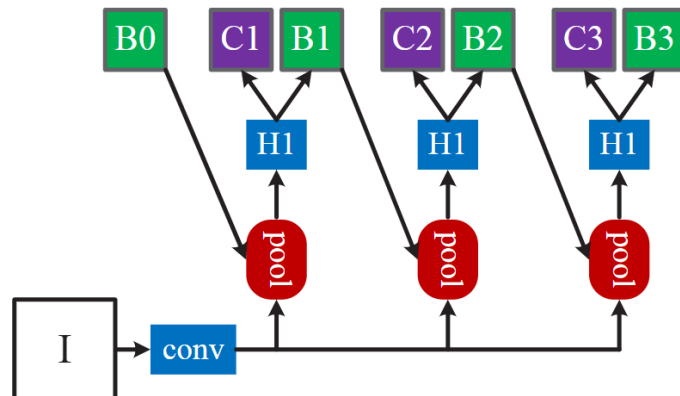


图 3-3 Cascade RCNN 网络结构

通过将 RPN 输出的检测框输入一个框回归模块 (如图 3-4 所示)，再将框回归模块的输出作为下一层的框回归模块的输入，如此重复提高 IOU 阈值，将使得预测框的定位效果获得提升。Cascade RCNN 可以提升检测效果，其关键点在于其采用了 4 次框回归。这些框回归模块通过生成  $d_x$ 、 $d_y$  调整矩形框的位置上，生成  $d_w$ 、 $d_h$  调整矩形框的长宽。但这些调整只具有线性调整能力，一旦生成的矩形框与真实框的差距过大，就很难通过框回归调整为真实框。因此 Cascade RCNN 通过将多个框回归模块级联来增强线性回归的能力。

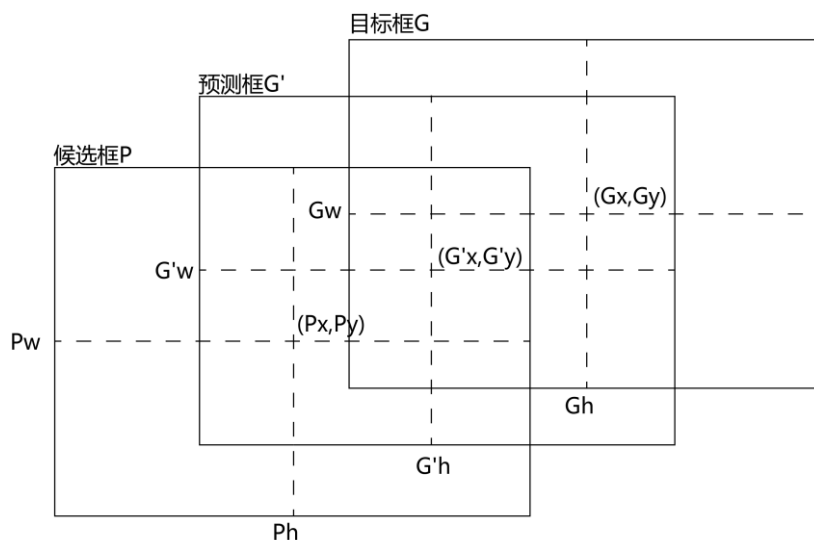


图 3-4 边框回归

### 3.1.3 Faster RCNN

Faster RCNN 是 Two-Stage 目标检测模型中的典型代表，Faster RCNN 检测部分主要可以分为四个模块：



(1) 卷积神经网络(conv layer), 用于提取特征。通过一组 conv+relu+pooling 层来提取图像的 feature maps, 用于后续的 RPN 层和推荐框选取。

(2) RPN 网络。在特征提取网络最后一层 feature maps 上滑动一个 3\*3 的窗口, 每次滑窗的中心处产生 9 个 anchor, 用来生成 RPN 网络训练所需要的 anchor 样本。将低维向量送入两个 1\*1 的卷积层, 一个是 bbox 的回归层, 一个是 bbox 的分类层; 滑动窗口的位置提供了物体的大体位置信息, 此外框的回归提供了框更精确的位置。

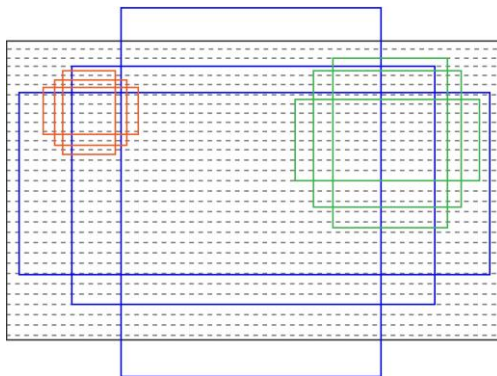


图 3-5 anchor 图

(3) 感兴趣区域池化(RoI Pooling), 用于收集 RPN 生成的推荐框中的每个框坐标, 并从 (1) 中 feature maps 中提取出来, 生成推荐特征图送入后续全连接层分类和回归。

(4) 分类与回归。利用推荐特征图计算出具体类别, 同时再进行一次边框回归获得检测框最终的精确位置。

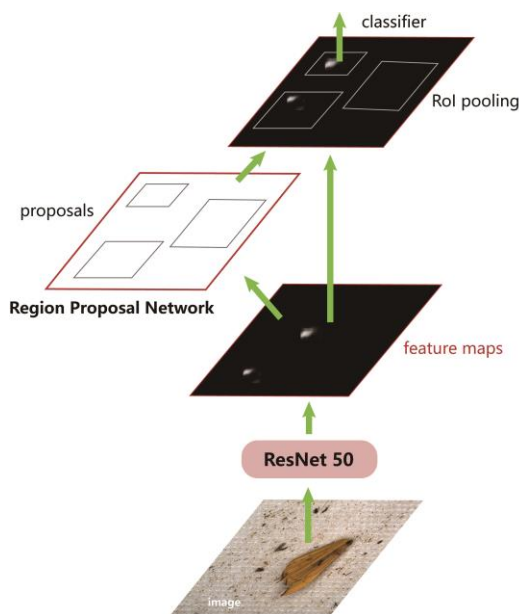


图 3-5 Faster RCNN 模型架构图

### 3.1.4 SSD

SSD(Single Shot MultiBox Detector) 是由 2016 年提出的典型的 One-stage 的多框预测方法。SSD 的提出主要解决了目标检测中速度慢、精度低的问题。SSD 的算法步骤主要如下：

- 1) 输入一幅图片 (300x300)，将其输入到预训练好的分类网络中来获得不同大小的特征映射，修改了传统的 VGG16 网络。将 VGG16 的 FC6 和 FC7 层转化为卷积层，如图 3-6 上的 Conv6 和 Conv7；去掉所有的 Dropout 层和 FC8 层；添加了 Atrous 算法 (hole 算法)；将 Pool5 从 2x2-S2 变换到 3x3-S1；
- 2) 抽取 Conv4\_3、Conv7、Conv8\_2、Conv9\_2、Conv10\_2、Conv11\_2 层的 feature map，然后分别在这些 feature map 层上面的每一个点构造 6 个不同尺度大小的 bbox，然后分别进行检测和分类，生成多个 bbox，如图 3-6 所示。

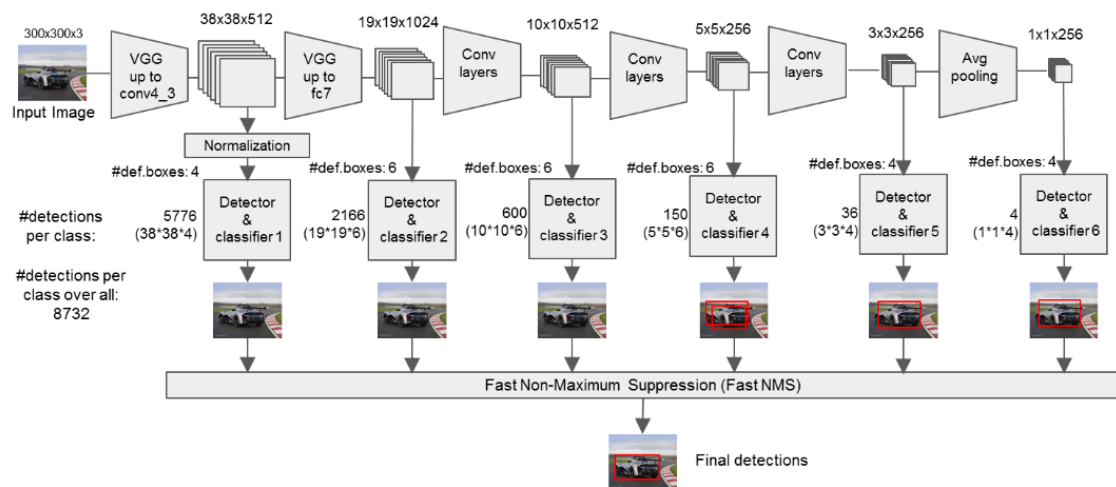


图 3-6 bbox 生成流程图

- 3) 将不同 feature map 获得的 bbox 结合起来，经过非极大值抑制方法来抑制掉一部分重叠或者不正确的 bbox，生成最终的检测结果，SSD 模型架构图如图 3-7 所示。

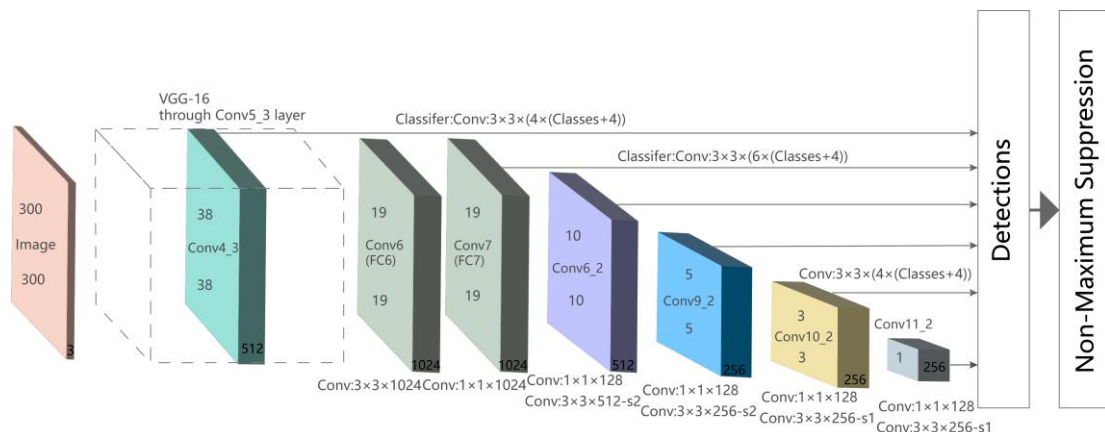


图 3-7 SSD 模型架构图

### 3.2 实验结果

表 3-1 各模型 FPS-mAP 对比表

模型类名	mAP	FPS
Cascade RCNN	53.2	4.2
Faster RCNN	43.5	19.2
SSD	29.9	48.1

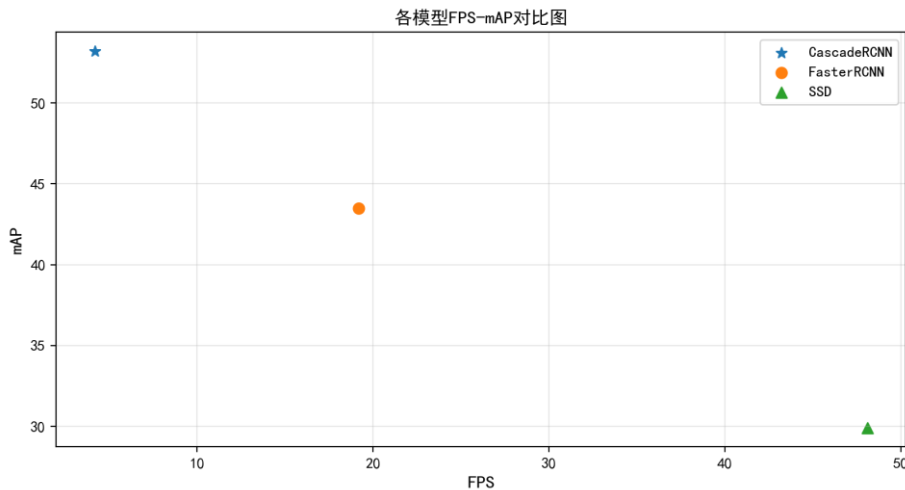


图 3-6 各模型 FPS-mAP 对比图

选用 Cascade RCNN、Faster RCNN、SSD 三个模型作为检测头，骨干网络统一选择 ResNet50 进行训练，为了验证三种模型结构的性能，本文对三者的 mAP 及 FPS 进行比较，模型相关结果指标为表 3-1 所示。可视化后如图 3-6，可以看出：

- (1) Cascade RCNN 的 mAP 在 FPS 为 4.2 的情况下高达 53.2。
- (2) Faster RCNN 的 mAP 为 43.5，FPS 为 19.2。
- (3) SSD 的 mAP 为 29.9 FPS 为 48。

通过对比模型的结果可以发现：

1. mAP: Cascade RCNN > SSD > Faster RCNN
2. FPS: SSD > Faster RCNN > Cascade RCNN

由于赛题最终评判标准为农田害虫识别的精确度，通过横向对比三类模型的 mAP，最终确定模型的基本检测网络为 Cascade RCNN。

### 3.3 本章小结

本章主要对农田害虫图像识别所需的理论知识进行了详细分析，为后文模型选择提供了理论基础，详细分析了 Cascade RCNN、Faster RCNN、SSD 三个检测网络对于数据集的检测效果，并选出了检测效果能力最强的 Cascade RCNN。下一章

将基于 Cascade RCNN 与第二章数据集分析得到的一些结论对农田害虫识别进行针对性改进。

#### 四、 基于 Cascade RCNN 的农田害虫检测改进算法

本章根据第二章分析数据集得到的结果，针对性改进 Cascade RCNN 算法。

由于第二章通过数据预处理已经解决了农田害虫数量尺度不均衡、数据分布不均匀等问题，本章主要目标是解决农田害虫形态大小不均衡问题，以及找到最适合 Cascade RCNN 算法的骨干网络。

##### 4.1 相关技术

###### 4.1.1 FPN

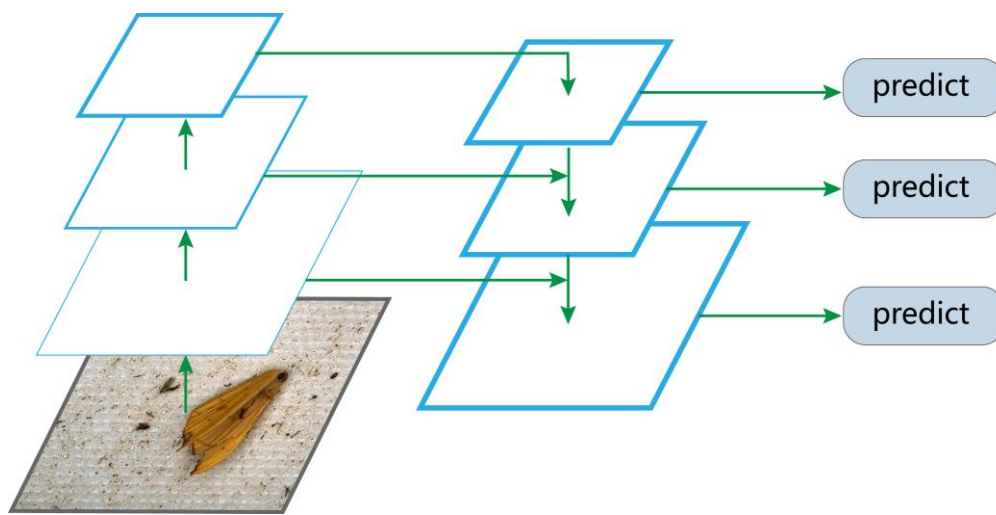


图 4-1 多尺度特征融合均衡化网络结构

2012 年之前，已经有了很多传统的方法来解决基于图像视觉的目标检测方法，但效果参差不齐；2012 年之后，随着 AlexNet 的出现，研究者发现卷积神经网络特征提取能力更强，便产生了新的诸如 RCNN 系列算法，对目标做初步定位，再对初步定位的目标做进一步精细。

而在这个背景下，FPN 应运而生。FPN (Feature Pyramid Networks) 特征图金字塔网络是 2017 年提出的一种网络，FPN 主要解决的是物体检测中的多尺度问题，通过简单的网络连接改变，在基本不增加原有模型计算量的情况下，大幅度提升了小物体检测的性能。

小目标难以识别的问题是目标检测的通病，而 FPN 推动了小目标的识别。当前，FPN 网络是目标检测算法中最常用的特征融合网络，具体实现方式如图 4-1 所示。

FPN 是在特征提取网络生成的多尺度特征图的基础上，将最后一层特征图做 2 倍上采样后与前一层特征图相加生成融合后的特征图，然后将新生成的特征图以同样的方式进行特征融合，逐层重复以上操作，形成特征金字塔。但是特征金字塔网络长距离信息流动会导致标签信息的流失和更多的关注相邻分辨率，这种融合方式会产生标签特征信息不平衡问题，所以本文基于 FPN 提出多尺度特征融合均衡化网络，利用相同深度融合的平衡语义特征来增强多层次的特征。首先将通过 FPN 网络形成的每层特征图通过上采样操作变成相同尺度，然后进行像素级的聚合均值，再将聚合均值化的特征图通过与之之前相反的采样操作形成不同尺度的特征图，从而得

到标签特征信息更平衡的多尺度特征图，实现图像特征的多层复用和融合均衡化，提高小目标图像的检测精度。

#### 4.1.2 ResNet

本文第二章以及简述过 ResNet50 的结构，本章需要使用到 ResNet50、ResNet101 以及 ResNet152，它们差别并不大，只是在每个 Block 重复的次数不一样。图 4-2 为 ResNet 结构表。

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		$1.8 \times 10^9$	$3.6 \times 10^9$	$3.8 \times 10^9$	$7.6 \times 10^9$	$11.3 \times 10^9$

图 4-2 ResNet 结构

#### 4.2 网络结构

在本文第三章，通过使用 ResNet50 为骨干网络，使用 Cascade RCNN、Faster RCNN、SSD 作为检测网络头，最终选出检测效果最好的 Cascade RCNN 检测网络头。在本文第二章，通过对数据集的分析，发现数据集存在目标形态尺度不均衡，且 Cascade RCNN-ResNet50 存在部分小目标漏检的情况，故本章对其引入 FPN 模块，作为其颈部网络，并且骨干网络选用 ResNet50、ResNet101、ResNet152 其中一种，即训练三个检测算法，最终从中筛选出检测效果最好的网络，其结构如下：

表 4-1 网络结构表

Head	Neck	Backbone
Cascade RCNN	FPN	ResNet50
Cascade RCNN	FPN	ResNet101
Cascade RCNN	FPN	ResNet152

确定三个检测算法结构后，将其连接起来，其主要连接思路为：

1. 将图像数据传入骨干网络后，经过 4 个大的残差块，每次宽高减半，通道数翻倍。
2. 删除 ResNet 最后的全连接层，将四个残差块输入到 FPN 层，并自顶向下进行融合并输出。
3. 将 FPN 层得到的 feature map 输入 RPN 层，其将在不同粒度上进行网格搜索，



并生成多个 Anchor。

4. 将 ResNet 最后一层的输出与 RPN 经过 Cascade RCNN，通过 3 次筛选后，最终输出图像分类与位置预测结果。
5. 最终构成图 4-3 的结构(以骨干网路以 ResNet50 为例)。

## CascadeRCNN-ResNet50-FPN

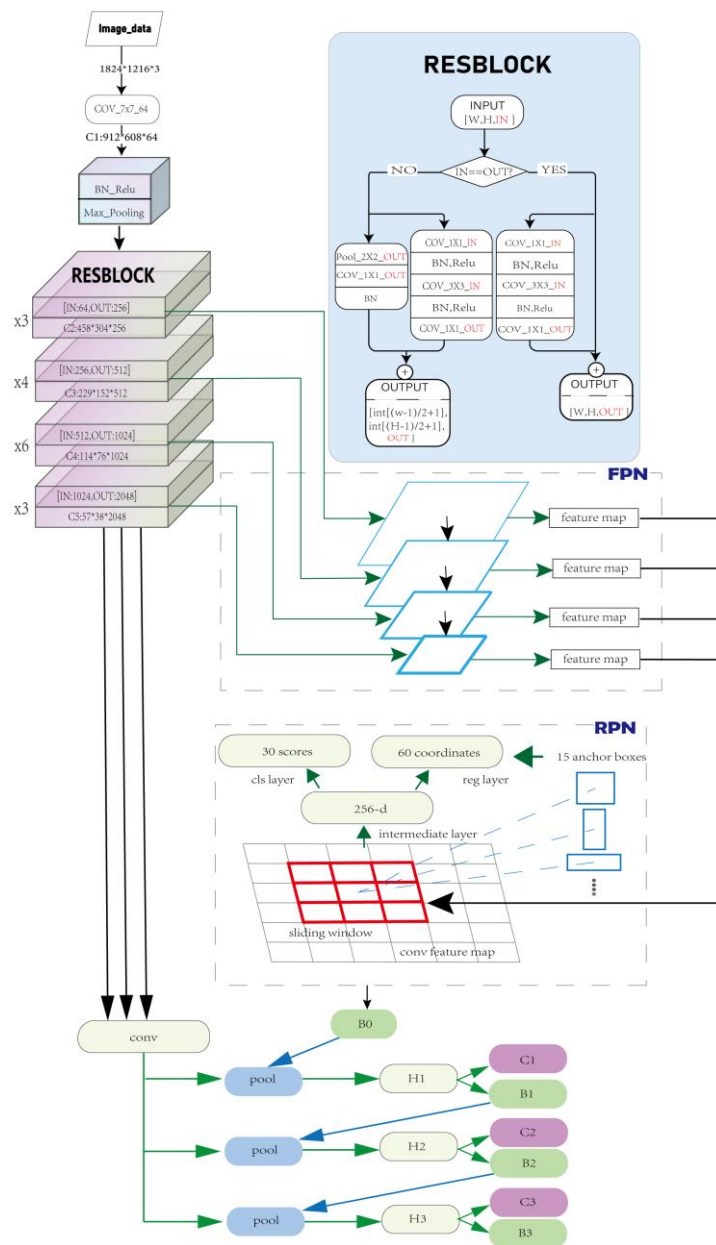


图 4-3 Cascade RCNN-ResNet50-FPN 结构图

### 4.2 采用的 Tricks

事实上，对于 Cascade RCNN-ResNet-FPN 的优化运用了很多 tricks，由于相对简单，本文只将采用的 tricks 与介绍列为下表：

- RandomResize：对图片进行 resize 用不同的制度训练可以提高鲁棒性；

- RandomFlip: 图片执行水平翻转;
- NormalizeImage: 归一化;
- Batch\_transforms: 因为需要生成不同框, 大小不一, 因此要跟随不同规格的图片填充 stride;
- Shuffle: 打乱图片;
- Drop\_last: 抛弃一些数据;

### 4.3 实验步骤

#### 4.3.1 实验环境

本文模型训练环境具体配置如下表 4-2 所示:

表 4-2 设备环境

CPU	GPU	操作系统
Intel Xeon Gold 5218 1TB×2	Tesla V100 32G×4	Ubuntu16.04.6

农田害虫检测实验中, 改进的 Cascade RCNN 农田害虫检测算法从训练到测试获取检测结果的具体实现步骤如下(骨干网络以 ResNet50 为例):

- (1) 将训练集作为网络的输入, 先通过骨干网络 ResNet50 进行特征提取, 将 ResNet50 的四个残差单元级联卷积层的输出作为多尺度特征图输出;
- (2) 将多尺度特征图通过多尺度特征融合均衡化网络进行特征融合, 输出信息更丰富且均衡的多尺度特征图;
- (3) 将多尺度特征图输入至共享的 RPN 网络, 产生存在害虫的预测框;
- (4) 把产生的农田害虫建议区域与 ResNet50 网络最终输出的特征图相结合, 经过池化运算后得到各类农田害虫的候选区域的特征图;
- (5) 将输出特征图送入 Cascade RCNN Head 后得到各类农田害虫的分类信息和位置信息, 通过非极大值抑制后选出最佳的预测框;
- (6) 重复以上步骤直至模型收敛, 则训练结束;
- (7) 提取出农田害虫检测模型的权重参数;
- (8) 将测试集通过农田害虫检测权重模型进行检测, 得到各类农田害虫的分类结果和定位结果;



#### 4.4 实验结果及分析

为了验证三种模型的性能与拟合程度，本文对其 PR 曲线、FPS、mAP、Loss 以及训练的 Epochs 进行比较。

##### 4.4.1 Epoch-mAP 分析

此步骤的主要目的是确定模型是否拟合完成，即查看 Epoch-mAP 图，随着训练的 Epoch 增加，mAP 时候逐渐平缓，若逐渐平缓，则证明模型接近训练完成。

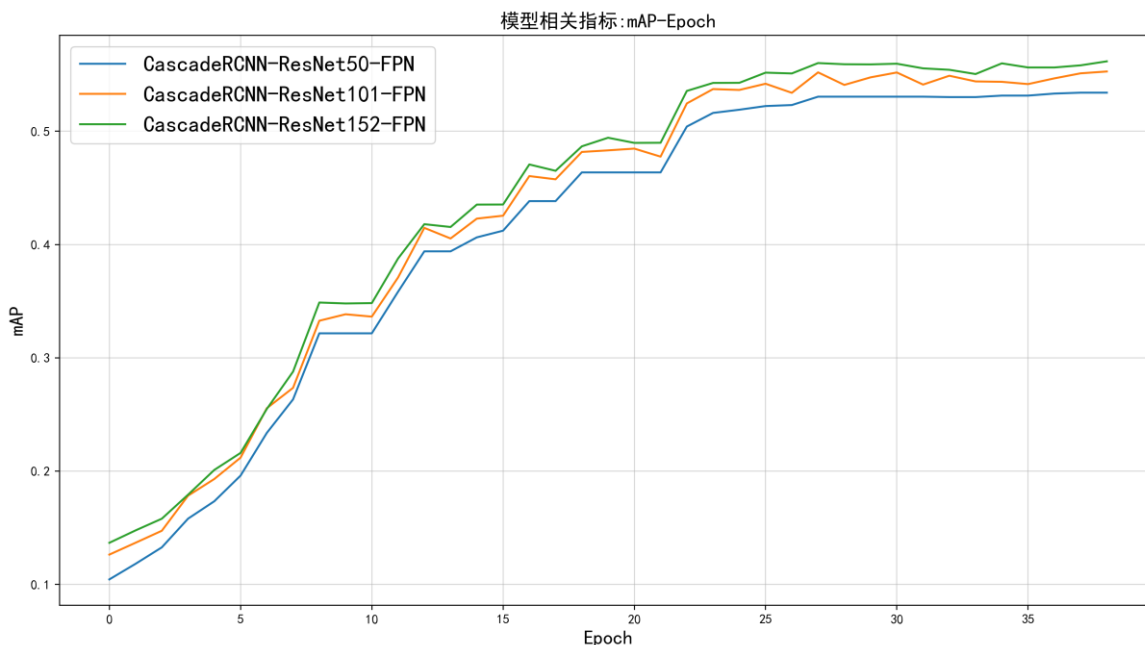


图 4-4 各模型 Epoch-mAP 图

通过监测各模型在整个训练过程中 mAP 的变化情况，可以直观的反映训练效果，具体如图 4-4 所示。其中，水平轴代表训练过程中的 Epoch，垂直轴代表 mAP 值，从图中可以看出：

- 三类模型均在训练到 35epoch 时 mAP 趋于平缓，证明模型已经拟合。
- 三类模型 mAP 均高于 54，证明其预测效果优异。

##### 4.4.2 Epoch-Loss 分析

此步骤的主要目的是确定模型是否存在过拟合或欠拟合情况，即查看 Epoch-Loss 图，随着 Epoch 增加：

- 模型的训练损失和验证损失越接近，则证明模型拟合效果好，
- 若训练损失大于验证损失，则证明模型在训练过程中出现了过拟合情况。
- 若训练损失小于验证损失，则证明模型在训练过程中出现了欠拟合情况

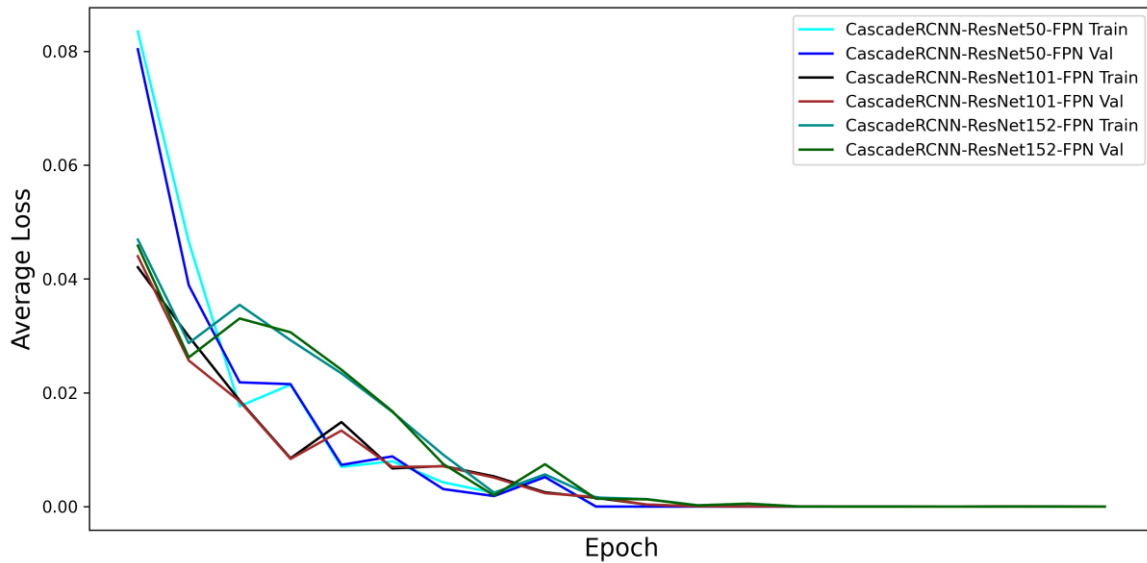


图 4-5 Epoch-Loss 图

通过监测模型在整个训练过程中的训练平均损失与验证平均损失可以直观的反应模型的拟合状况，如图 4-5 所示，其中水平轴代表训练过程中的 Epoch，垂直轴代表平均损失值，从图中可以看出：

- 三类模型拟合效果较好，不存在过拟合与欠拟合的情况。

#### 4.4.3 mAP-FPS 分析

通过前两步的分析，已经能够确定模型的拟合情况优秀，接下来通过 mAP-FPS 来分析模型检测性能，各模型 mAP-FPS 结果如表 4-3 所示：

表 4-3 各模型 mAP-FPS 表

模型	mAP	FPS
Cascade RCNN-ResNet50-FPN	54.8	10.2
Cascade RCNN-ResNet101-FPN	55.2	7.3
Cascade RCNN-ResNet152-FPN	56.3	4.1

通过表 4-3 可得如下结论：

- 当骨干网络为 ResNet50 时，FPS 最高为 10.2。
- 当骨干网络为 ResNet152 时，mAP 最高为 56.3。

#### 4.5 本章小结

本章针对 Cascade RCNN 算法仍存在对小目标和重叠度高的图像识别率低的问题，提出了一种基于 Cascade RCNN 的图像检测模型，能够完美运用在农田害虫检测中。通过采用 ResNet152 替换原骨干网络作为 Cascade RCNN 算法的特征提取网

络，完成图像分类与预测，以及提出基于 FPN 网络改进的多尺度特征融合均衡化网络，提高害虫图像检测精度。

## 五、模型检测效果以及农田害虫防范建议

### 5.1 模型检测结果

根据赛方第三题要求：检测 800 张图，检测每张图片并统计虫子编号以及数量。

本文使用 Cascade RCNN-ResNet152-FPN 对每张图片进行检测后，统计并可视化如下图 5-1：

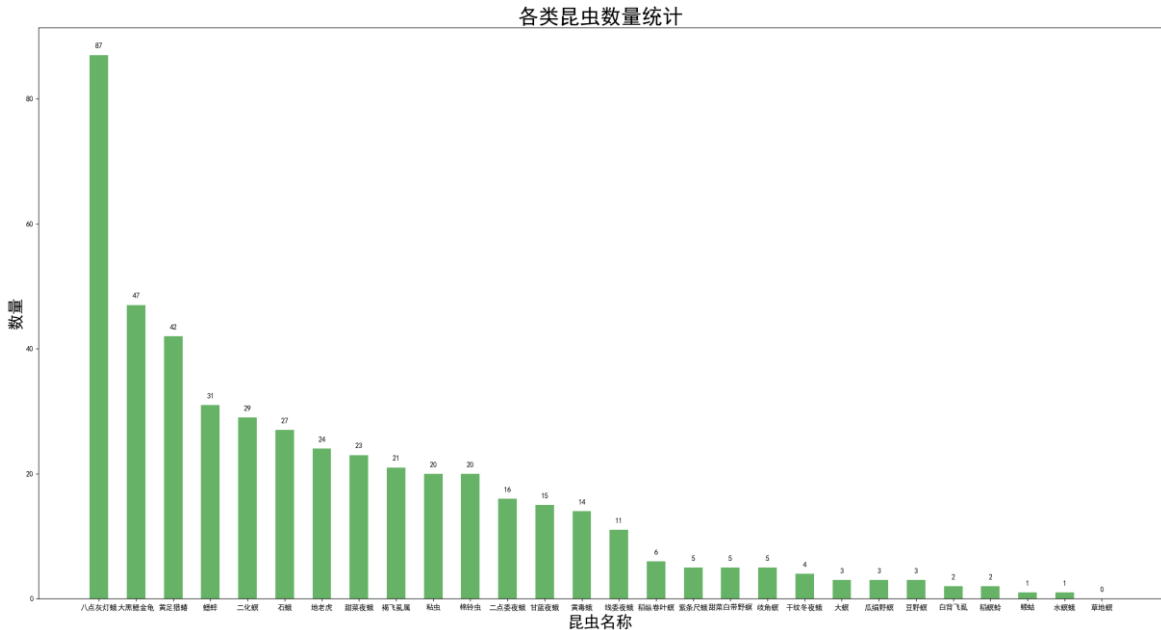


图 5-1 模型检测结果

通过对数据集中不同种类的害虫数量进行统计分析，本文得出以下结果：

- 1) 各类害虫的数量分布不均衡，存在较大的数量差；
- 2) 害虫数量排行前五的为八点灰灯蛾、大黑鳃金龟、黄足猎蝽、蟋蟀，其数量（单位：只）分别为 87、47、42、31、29；
- 3) 害虫数量排行倒数五种的为草地螟、水螟蛾、蝼蛄、稻螟蛉、白背飞虱，其数量（单位：只）分别为 0、1、1、2、2。

### 5.2 农田害虫防范建议

不同种类害虫的数量分布对实际种植过程中的防范具有重大意义。通过对数量分布的分析，本文发现八点灰灯蛾的数量较多，其数量比数量位列第二的大黑鳃金龟多一倍。因此，本文建议在后续该农田种植过程中加强对八点灰灯蛾的防范与消杀。

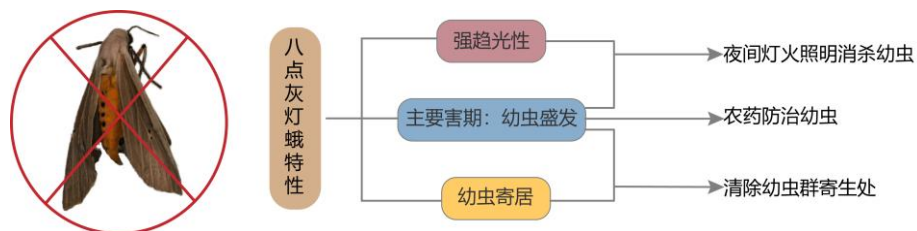


图 5-2 八点灰灯蛾消杀

八点灰灯蛾成虫有强趋光性，卵多于下午 14 时至此 13 零时孵化，以入夜 20—22 时孵化最盛，初孵化幼虫群集于寄主植物青叶上，多于玉米、蔬菜、红薯等作物地边的石缝、土隙，杂草根或泥块下面，越冬时幼虫尚未老熟便回将种植物的幼茎嫩叶咬断吃光，一头幼虫可在夜间将 4—6 株玉米苗全部吃光，其对农作物收成具有较大影响。

因此，本文针对八点灰灯蛾的生活习性提出以下几点防范建议：

1) 人工消杀。八点灰灯蛾幼虫盛发为最主要的害期，因八点灰灯蛾成虫有强趋光性，可在夜间利用灯火照明消杀其幼虫。

2) 农业防治，冬季时勤除农田中的杂草、清除田间乱石、填平土隙，清除初孵化幼虫群的寄生处。

3) 勤耕翻土地，消灭表土或枯叶残株内的越冬幼虫。

4) 化学药剂防治。在八点灰灯蛾成虫盛发期和幼虫 2 龄前喷 25%灭幼脲 3 号悬浮剂 500~600 倍液，或 40%氰戊菊酯乳油 3000 倍液，或 2.5%功夫乳油 2000 倍液，或 20%灭扫利乳油，2.5%王星乳油 3000 倍液，或 20%绿·马乳油 2000 倍液，以上化学药剂的实际效果都较为良好。在使用化学药剂防治八点灰灯蛾应注意在采收前 7 天停止用药。

## 六、 总结

本文针对农田害虫图像识别的需求，设计高效、准确率高、能够部署在实际农田种植场景中的算法模型，并对数据集的特点进行针对性优化，其主要工作如下：

查阅相关资料，了解农田害虫的图像识别方法的国内外现状。

通过对数据集分析得到，数据集中存在害虫的数量、大小尺度不均衡，且数据集较少、图像不均衡的现象。

在第二章通过数据预处理的方式，即通过筛选错误标签、数据重采样、数据增广、图像归一化的操作，解决了各害虫的数量尺度不均衡、图像不均衡问题，并增大了数据集。

在第三章，固定骨干网络 ResNet50，而检测头选用 SSD、Faster RCNN、Cascade RCNN 训练检测网络，得到 Cascade RCNN 作为检测头效果最好。

在第四章的目的是训练出一个效果优异的检测网络，在第二章分析结果中，只有害虫大小尺度不均衡这一个问题暂未解决，本文在检测头为 Cascade RCNN 的基础下，引入 FPN 网络来解决小尺度害虫检测的问题，并比较了骨干网络为 ResNet50、101 与 152 的检测性能，最终训练出效果最好的检测网络 Cascade RCNN-ResNet152-FPN，其 mAP 高达 56.3，FPS 为 4.1。

在第五章，使用 Cascade RCNN-ResNet152-FPN 对测试集预测结果，并统计害虫种类与数量，并以此数据为提出针对性消杀害虫建议。

## 七、 未来展望及现实意义

如今，人工智能的使用领域越来越广泛。基于深度学习的图像识别技术是现代化农业的一个重要领域，其有着提取图片特征客观、可以利用不同的数据架构应对不同的需求问题的优点。通过对害虫图像进行加工、分割，以便提取有用的信息并进行识别，有着广泛的应用前景，也是当前及未来农业现代化、工业化、信息化的研究重热点。利用现代信息技术进行植物病虫害防治是未来我国植保发展的方向，但基于深度学习的图像识别方法对机器配置、样例数量等要求较高。

因此，为了使图像识别技术更好的应用于病虫害防治工作中，可以着眼于以下几个方面：

(1) 深度学习+传统图像识别方式。农户或者农业技术人员利用田间发现拥有复杂背景的病虫害图像，人工积累学习样例，发挥深度学习的优势，更新特征数据集，提高模型检测准确率。

(2) 建立大型农业图像数据共享中心。共享中心将我国不同地区采集的图像数据收集到专用数据库中，研究人员可在数据库中根据个人需求查看，降低研究人员采集的时间，减少不必要的时间浪费提高工作效率。

(3) 引入高光谱技术，为农作物或者病虫害拍摄不同波段的图像信息，结合深度学习的图像识别技术进行植物无损检测。可见光成像技术已趋于成熟，为了弥补其只能采集农作物表面信息，引入红外成像和高光谱成像，提高农作物表面信息采集的完全度。

(4) 模型轻量化。提高算法执行速度，满足实时性要求，以实现在移动设备上的应用。

(5) 优化移动端应用程序。在满足模型轻量化的前提下，优化移动端应用程序也是重要研究方向之一。目前病虫害识别系统在使用端主要依赖于手机移动 APP，该系统具有使用简单，反应速度快，功能强等特点，但是手机端 APP 对于网络依赖性强，可以进一步优化数据库结构，降低软件整体所占空间，以便达到在无网络状态下的手机端离线识别农作物病虫害功能。

## 参考文献

- [1] 人民资讯. “中国提出到 2022 年建成高标准农田 10 亿亩”.  
<https://baijiahao.baidu.com/s?id=1711110158912860846&wfr=spider&for=pc>,  
2022. 4. 24
- [2] Yanfen Li, Hanxiang Wang, L. Minh Dang, Abolghasem Sadeghi-Niaraki, Hyeonjoon Moon. “Computers and Electronics in Agriculture”. 2021. 1-4
- [3] Happy. “深度学习数据增广技术一览”  
<https://zhuanlan.zhihu.com/p/144921458>. 2022. 4. 24
- [4] Shaoqing Ren, Kaiming He, Ross Girshick, & Jian Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks IEEE Transactions on Pattern Analysis and Machine Intelligence. 2015
- [5] Zhaowei Cai and Nuno Vasconcelos. “Cascade R-CNN: Delving into High Quality Object Detection” arXiv: Computer Vision and Pattern Recognition. 2017. 1-5. pag.
- [6] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu and Alexander C. Berg. “SSD: Single Shot MultiBox Detector” european conference on computer vision. 2016: 5-7. pag.
- [7] J. Dai, Y. Li, K. He, and J. Sun. R-FCN: object detection via region-based fully convolutional networks. In NIPS, pages 379-387, 2016. 2, 3, 5, 8
- [8] P. Dollár, P. Welinder, and P. Perona. Cascaded pose regression. In CVPR, pages 1078-1085, 2010. 5
- [9] C. Elkan. The foundations of cost-sensitive learning. In IJ-CAI, pages 973-978, 2001. 2
- [10] P. F. Felzenszwalb, R. B. Girshick, D. A. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. IEEE Trans. Pattern Anal. Mach. Intell., 32(9):1627-1645, 2010. 2, 3
- [11] S. Gidaris and N. Komodakis. Object detection via a multi-region and semantic segmentation-aware CNN model. In ICCV, pages 1134-1142, 2015. 3
- [12] R. B. Girshick. Fast R-CNN. In ICCV, pages 1440-1448, 2015. 1, 2, 3, 5
- [13] S. Han, J. Pool, J. Tran, and W. J. Dally. Learning both weights and connections for efficient neural network. In NIPS, pages 1135-1143, 2015. 6
- [14] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask r-cnn. In ICCV, 2017. 3, 6, 8
- [15] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, and K. Murphy. Speed/accuracy trade-offs for modern convolutional object detectors. CoRR, abs/1611.10012, 2016. 8



- [16]Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. B. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. In MM, pages 675-678, 2014. 5
- [17]H. Li, Z. Lin, X. Shen, J. Brandt, and G. Hua. A convolutional neural network cascade for face detection. In CVPR, pages 5325-5334, 2015. 3
- [18]T. Lin, M. Maire, S. J. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. Microsoft COCO: common objects in context. In ECCV, pages 740-755, 2014. 2, 5
- [19]Z. Cai, Q. Fan, R. S. Feris, and N. Vasconcelos. A unified multi-scale deep convolutional neural network for fast object detection. In ECCV, pages 354-370, 2016. 2, 3
- [20]X. Cao, Y. Wei, F. Wen, and J. Sun. Face alignment by explicit shape regression. In CVPR, pages 2887-2894, 2012. 5
- [21]J. Dai, K. He, and J. Sun. Instance-aware semantic segmentation via multi-task network cascades. In CVPR, pages 3150- 3158, 2016. 3
- [22]A Coarse-to-Fine Model for Rail Surface Defect Detection. YU H M,LI Q Y,TAN Y Q,et al. IEEE Transactions on Instrumentation and Measurement . 2019
- [23]An adaptive extraction method for rail crack acoustic emission signal under strong wheel-rail rolling noise of high-speed railway. HAO Q S,SHEN Y,WANG Y,et al. Journal of Mechanical Systems . 2021