

分类号：TP391

单位代码：10110

学 号：S1707015



中北大學

North University Of China

硕士学位论文

(专 硕)

基于深度学习的林业害虫监测
系统设计与实现

硕士研究生 候瑞环
指导教师 杨喜旺
校外导师 张占一
学科领域 计算机技术

2021年5月27日

图书分类号 TP391 密级 _____
UDC ^{注 1} _____

硕士 学位 论文

基于深度学习的林业害虫监测系统设计与实现

候瑞环

校内指导教师（姓名、职称） 杨喜旺 副教授

校外指导教师（姓名、职称） 张占一 高级工程师

申请学位级别 工程硕士

专业名称 计算机技术

论文提交日期 2021 年 5 月 20 日

论文答辩日期 2021 年 5 月 25 日

学位授予日期 年 月 日

论文评阅人 李爱萍（教授） 蔺素珍（教授）

答辩委员会主席 蔡江辉（教授）

2021 年 5 月 27 日

注 1：注明《国际十进分类法 UDC》的分类

基于深度学习的林业害虫监测系统设计与实现

摘要

害虫爆发会对林业资源及生态安全造成极大威胁、给国家造成巨大经济损失。因此，实现林业害虫监测非常必要。目前针对林业害虫的研究比较少，多数基于储粮害虫与农作物害虫进行研究，且害虫监测防控手段以诱捕器诱捕、虫情测报仪诱测与人工识别计数为主，费时费力、易出错。本文使用改进的深度学习目标检测模型 YOLOv4-TIA 对林业昆虫图像进行自动特征提取及识别检测，并在此基础上设计并实现了林业害虫监测系统。本文主要研究工作如下：

首先，构建包含两种应用场景的害虫图像数据集。一种为从网络搜集整理的野外自然环境下的 5 种常见森林害虫图像，使用 LabelImg 进行害虫目标标注，共 2000 幅图像；另一种为诱捕器背景下的 7 种林业害虫图像，共 1973 幅图像。通过数据增强操作将两种数据集的图像数量分别扩充至 12000 张与 10758 张，可被应用于基于深度学习的林业害虫检测的相关研究。

其次，构建适用于林业害虫检测的 YOLOv4-TIA 网络。通过在 YOLOv4 的骨干网络中设计 Triplet Attention 模块来改进 CSPDarknet53 结构，捕获跨维交互，改善由标准 CSPDarknet53 生成的特征表示，实现在不增加网络参数量的前提下通过强化目标特征表示来提取高质量图像特征；通过增加跳跃连接与高效多方向的跨尺度连接相结合的特征融合方式，改进 YOLOv4 检测网络中 PANet 的特征融合方式，更好地平衡多尺度特征信息，获取更为丰富的语义信息和位置信息，同时利用 Focal loss 优化损失函数。将改进的 YOLOv4-TIA 网络与其他四种模型进行实验对比，结果表明改进的 YOLOv4-TIA 网络在本次研究的数据集上性能最佳，mAP 可达 98.8%。

最后，文章在 YOLOv4-TIA 网络的基础上设计并实现了林业害虫监测系统。系统主要包括害虫图像信息录入、读取图像文件、害虫检测、害虫计数、虫害预警与虫情分析等功能，为林业科学防治提供了一种新的思路与依据。

关键词: 林业害虫图像，害虫检测网络，YOLOv4-TIA，Triplet Attention，残差变换，特征融合，监测系统

The design and implementation of forestry pest monitoring system based on deep learning

ABSTRACT

pests outbreak will pose a great threat to forestry resources and ecological security, thus causing huge economic losses to the country. Therefore, it is necessary to monitor forest pests. Currently, there are few studies on forestry pests, and most studies are carried out based on stored grain insects and crop pests. Moreover, the main means of pest monitoring and control include trap trapping, pest situation measuring and reporting instrument forecast and manual identification and counting, which is time-consuming, laborious and error-prone. In this paper, the improved deep learning target detection model YOLOv4-TIA was used to extract features automatically and identify forest insect images, and on this basis, a forest pest monitoring system was designed and implemented. The main research work of this paper is as follows:

Firstly, pest image datasets were constructed under two application scenarios. One was to collect and sort out images of five common forest pests in the wild natural environment from the network; LabelImg was used to mark pest targets, with a total of 2,000 images; the other was the image of 7 kinds of forest pests under the background of trap, with a total of 1973 images. Both datasets were expanded through data enhancement operation, and the final number of images reached 12,000 and 10,758 respectively, which can be applied to the related studies of forestry pest detection tasks based on deep learning.

Secondly, the YOLOv4-TIA pest detection network suitable for forestry pest detection was constructed. The structure of CSPDarknet53 was improved by designing the Triplet Attention modules in the backbone network of YOLOv4, to capture cross-dimensional interaction and improve the feature representation generated by the standard CSPDarknet53, achieving to extract high-quality image features by strengthening the target feature representation without increasing the amount of network parameters; The feature fusion method of PANet in the YOLOv4 detection network was improved through skip connection and efficient multi-directional cross-scale connection, so as to better balance multi-scale feature information and obtain richer semantic information and location information, at the

same time, Focal loss was used to optimize the loss function. By comparing the improved YOLOv4-TIA network with the other four models, it was found that the improved YOLOv4-TIA network has the best performance on the dataset of this study, and the mAP could reach 98.8%.

Lastly, based on the YOLOv4-TIA network, a forest pest monitoring system was designed and implemented. The system mainly includes the functions of pest image information input, image file reading, pest detection, pest counting, pest early warning and pest situation analysis, which provides a new idea and basis for scientific forestry control.

Keywords: Forest Pest Image, Pest Detection Network, YOLOv4-TIA, Triplet Attention, Residual Transformation, Feature Fusion, Monitoring System

目 录

第 1 章 绪论	1
1.1 课题研究背景与意义	1
1.2 国内外研究现状	2
1.3 研究内容及组织结构	3
1.3.1 研究内容及技术路线	3
1.3.2 论文结构安排	5
第 2 章 基于深度学习的图像检测识别技术与基础理论	7
2.1 深度学习	7
2.1.1 深度学习概述	7
2.1.2 深度学习框架	7
2.2 卷积神经网络基本原理	8
2.2.1 卷积层	8
2.2.2 池化层	10
2.2.3 全连接层	11
2.2.4 激活函数	11
2.3 基于深度卷积神经网络的目标检测技术	13
2.3.1 两阶段目标检测方法	14
2.3.2 单阶段目标检测方法	17
2.3.3 检测算法模型的损失函数与优化器	24
2.4 本章小结	26
第 3 章 林业害虫检测网络的设计	27
3.1 基于 YOLOv4 改进的林业害虫检测网络	27
3.1.1 骨干网络的改进	27
3.1.2 融合改进 PANet 结构的 YOLOv4-TIA	29

3.1.3 损失函数的优化-Focal Loss 函数	30
3.2 数据集与数据预处理.....	30
3.2.1 数据集介绍.....	30
3.2.2 数据预处理.....	31
3.3 林业害虫候选区域参数聚类.....	36
3.4 实验与数据分析.....	38
3.5 本章小结.....	42
第 4 章 林业害虫监测系统的设计与实现	43
4.1 林业害虫监测系统的功能需求分析.....	43
4.2 林业害虫监测系统设计	44
4.2.1 系统总体设计.....	44
4.2.2 系统模块功能设计	46
4.2.3 数据库表设计	47
4.3 林业害虫监测系统实现.....	50
4.4 林业害虫监测系统测试	55
4.4.1 平台测试环境说明.....	55
4.4.2 平台功能测试.....	55
4.4.3 系统性能测试	57
4.5 本章小结.....	57
第 5 章 总结与展望	58
5.1 总结.....	58
5.2 展望	59
参考文献	60
攻读硕士学位期间发表的论文及所取得的研究成果	67
致谢	

第1章 绪论

1.1 课题研究背景与意义

森林被誉为“地球之肺”，是自然界的宝贵资源，覆盖地球表面的三分之一，具有经济、生态和社会三大效益^[1]。近 40 年来，天然林面积逐步减少，为了应对此问题，我国大力发展人工林，造林面积居世界首位。天然林发展时间长，物种丰富，抗虫害能力、稳定性更强；而人工林却因树种单一、长期使用化学药剂等因素更容易发生病虫害。目前，我国林业害虫约 2400 种，破坏性强且较为常见的有 10 余种，包括松才线虫、美国白蛾、黄脊竹蝗、粗梢双条杉天牛和光肩星天牛、油茶尺蠖、杨扇舟蛾、榆蓝叶甲、日本松干蚧、华山松大小蠹等。这些害虫的侵袭或寄生，会使林木在形态、组织或生理生态上发生一系列不正常变化，影响林木健康发展，致使林木产量和质量下降；且害虫本身繁殖能力强、传播速度快且危害周期长，一旦蔓延，势头难以遏制，发病面积广，疫木治理困难，耗时耗力，会造成严重的人力、物力与经济损失，更为严重者会致使树木枯死甚至导致整个林区生态环境恶化。据 2019 年统计报告，我国美国白蛾虫害发生面积为 12171.70 万亩，受损区域 1153.30 万亩，虫害范围涉及 13 个省，近 600 个县，不仅造成了巨大经济损失，还严重影响当地生态平衡^[2]。2021 年一月至今，我国在辽宁、四川、重庆、安徽、浙江、湖南、福建等省份连续开展了林业松才线虫病有害生物监测预报与疫木治理工作，建立除治队伍，落实专人监管，实行分区防治、分级管理。其中，广西疫区涉及 13 个市，除治小班 5704 个；浙江发病面积 35608 亩，投入劳动用工 18898 人次。由此可见，病虫害爆发有极端危害，为了降低林业害虫爆发的几率，推进防控信息化建设，建立虫情监测网络，进行准确及时的虫情监测，为科学防治提供理论依据十分必要。

传统的林业害虫监测方法以人工识别与计数为主，测报员和专家使用肉眼、显微镜来观察害虫，并运用经验来判断害虫类别并统计害虫数量^[3]。此方法虽然可以完成害虫的识别与统计，但无法应对季节性虫害，会因测报员与专家人手不足而造成检测识别效率低下，或因人的主观因素造成识别误差。

面对虫情测报与预警的需求，学者们在自动化与害虫监测预警系统方面开展了大量

研究，以自动识别害虫图像中目标为基础，通过统计害虫数目对虫情进行测报。因此，对害虫目标进行识别与计数的前提通常是先采集到害虫的图像、声音和雷达等信号。但这些方式中一些关键技术尚未得以突破，以致这些系统对目标的识别精确度低，在实际应用中难以推广^[4]。

随着计算机硬件、图像处理技术和机器学习的发展，害虫图像检测法得以应用于昆虫识别与计数，此方法通常将图像采集技术与机器学习方法相结合，与传统方法相比，它虽然借助了算法的计算分析，检测效率较高，但仍需要人工来设计图像中的害虫特征，因而模型鲁棒性差、泛化能力弱^[5]。近年来，深度学习飞速发展，在图像识别、目标检测、实例分割、疲劳驾驶等领域展现出了目前最先进的性能，且已经在人脸识别、行人检测、自动驾驶、医学成像、工业检测、机器人视觉、智能视频监控、遥感图像等任务上得到了广泛应用。现如今，图像采集技术已相对成熟，图形处理器（Graphics Processing Unit, GPU）芯片算力也更加强大，利用深度学习技术提高虫情监测水平已成为农林业工程领域的研究热点之一。基于深度学习的图像检测方法大大提升了对图像中害虫目标进行分类和定位的速度与准确率，对于提高虫情监测水平具有重要意义。因此，本文基于深度学习的基础理论与方法对林业害虫图像检测算法进行系统性的研究，并设计、实现可行的林业害虫监测系统。

1.2 国内外研究现状

早期的基于图像的昆虫检测识别方法主要依赖昆虫的翅膀、触角、足等外部形态学特征^[6]。早在 1995 年至 2005 年间，国内外专家学者便开始针对昆虫形态图像识别开展了许多研究工作^[7-8]。国外也率先将研究成果与实际应用相结合，相继落地了 ABIS^[9]、DAISY^[10]、DrawWing^[11]等昆虫识别软件。1990 年前后，国内沈佐锐教授也开始研究昆虫形态技术并开发了名为 BUGVISU 的昆虫识别系统，该系统可以基于图像中六个特征实现对昆虫目标的有效识别^[12]。但以上研究及其相应的软件主要应用于昆虫形态研究，无法应用于生产实践。

随着计算机视觉技术的发展，人工提取图像中害虫的形态学特征与机器学习相结合的分类方法被逐渐应用于昆虫识别任务。机器学习中被常用于进行昆虫特征分类的分类器包括支持向量机(Support Vector Machine, SVM)^[13]、KNN^[14]、K-means^[15]和神经网

络^[16]等。在利用机器学习方法进行昆虫识别分类的研究中，Saeed 和 Mohammad 等^[17]人利用 SVM 分类方法检测草莓植株上的害虫。PanChunhua 等^[18]人先使用网络搜索算法提取蔬菜害虫图像的目标区域，再使用 SVM 网络对目标区域中的害虫目标种类进行识别，识别准确率达到了 93%。2006 年，廉飞宇^[19]提出了基于支持向量机的淘汰法，实现了对近百张害虫图像中目标的识别。2008 年，周曼等^[20]人利用 BP 神经网络对害虫图像特征进行提取并在此基础上进行分类。Zhao Juan 等^[21]人将 Ada Boost 图像特征选择算法与 SVM 相结合，实现了对农业害虫天剑的检测识别。这些研究方法虽然可以达到不错的检测识别精度，但是需要手工设计特征，且只能应用于害虫图像样本量少及类别有限的情况下。在环境复杂或者害虫类别繁多的情况下，这些方法无法提取图像中的深层特征信息，模型鲁棒性相对较差、泛化能力弱、识别精度较低。

近年来，深度学习快速发展，在计算机视觉领域也取得了很大成就。许多专家学者也逐渐着眼于使用深度学习方法进行昆虫图像的识别^[22]。深度学习方法可以通过卷积神经网络自动提取图像的深层信息，最终学习到更为复杂的高层语义信息，在大规模复杂背景的昆虫图像识别中表现优异。如 Ding 和 Taylor^[23]利用卷积神经网络对性诱剂粘板上的害虫图像实现了害虫的检测识别与计数；2017 年，程尚坤^[24]利用 Caffe 框架设计出一种基于深度卷积神经网络的粮食害虫浅层检测模型。该模型采用 Softmax+cross-entropy 方式的损失函数对甲虫类的检测分类率高达 95%，同时提高了蝶类害虫的识别率。2018 年，Shen 等^[25]使用了训练较慢，但检测精度较高的 Faster-RCNN 方法，以候选区域网络(Region Proposal Network, RPN) 提取图像中可能包含储粮害虫的区域，并利用改进的 Inception 神经网络结构识别这些区域。此方法 mAP 值达到了 88%，能够有效地检测有杂质背景下的储粮害虫；程曦等^[26]使用深度卷积神经网络实现了储粮害虫图像的识别分类，并且达到了 97.61% 的准确率。

1.3 研究内容及组织结构

1.3.1 研究内容及技术路线

对上述国内外研究现状进行分析，可以发现，近年来针对昆虫图像的检测研究虽然较多，但关于林业害虫图像检测识别与监测系统方面的研究与应用相对较少，缺乏集林

业昆虫图像检测与监测预警为一体的系统。设计可以对林业昆虫图像进行精确识别的林业害虫监测系统，对于实现及时的虫情监测与提高林业害虫实时防治和监测水平具有重要意义。因此，本文通过对目标检测算法进行系统性的研究，最终以 YOLOv4 目标检测网络^[27]为基础，针对直接将 YOLOv4 网络应用于林业昆虫图像时检测精度不足以达到实际应用需求的问题，提出使用三分支注意力^[28]（Triplet Attention）对 CSPDarknet53 骨干网络进行改进，优化检测模块特征融合方式、优化损失函数的方法，提升模型对林业害虫的检测准确率，并且基于所提出方法设计并实现了林业害虫监测系统。本文主要研究内容如下：

首先，本文针对两种场景下共 12 种林业害虫进行检测识别，并且单独构建了一个包含破坏性强且较为常见的 5 种共计 2000 张的自然环境下的林业害虫图像样本集，并通过旋转、翻转、缩放、剪裁、调整亮暗、对比度等多种数据增强操作对两种数据集进行扩充，使其可供进行基于深度学习的林业害虫检测的相关研究。

其次，本文研究并实现了基于深度学习目标检测算法的昆虫目标检测识别方法，并针对网络在复杂环境下难以提取图像特征、检测效果不佳的问题，对 YOLOv4 目标检测模型进行了改进，使用 Triplet Attention 模块对骨干网络进行优化，使其可以重点关注图像中的目标特征；同时对颈部网络 PANet 进行改进，增加跳跃连接与高效多方向的跨尺度连接方式，通过改善特征融合方式平衡多尺度特征信息，最终提取更为丰富的害虫目标特征信息，提高网络检测识别精确度，通过与四种经典目标检测模型进行对比实验，结果表明本文所提方法在林业害虫检测任务上性能最佳。

最后，根据 YOLOv4-TIA 害虫检测网络设计并实现了林业害虫监测系统。系统通过构建昆虫信息数据库来进行数据存储，并设计软件功能，其软件功能包括昆虫图像信息录入、图像读取、害虫检测与计数、虫情分析和虫害预警。完成软件设计后对系统相关功能进行测试，测试表明系统基础功能运作正常，可以满足林业害虫监测预警任务的需求。本文技术路线如图 1-1 所示。

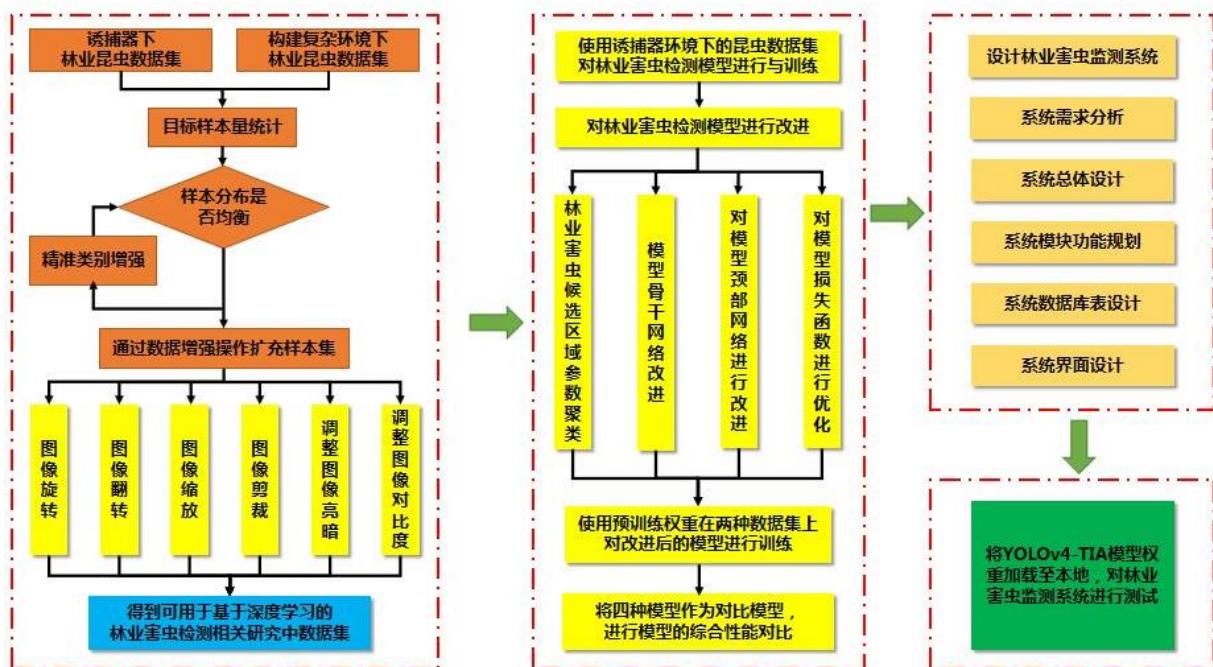


图 1-1 课题技术路线

Fig 1-1 Technical route of the subject

1.3.2 论文结构安排

论文共包含五个章节，各个章节内容如下：

第一章，绪论。首先阐明了监测林业害虫虫情、建立虫情监测网络的研究背景与意义，紧接着陈述了近年来常用的基于图像的害虫识别与软件监测方法的国内外研究现状，最后将论文的主要研究内容及结构安排进行了阐述。

第二章，基于深度学习的图像检测识别技术与基础理论。在对深度学习理念和框架进行介绍的基础上，阐述图像检测识别技术中卷积神经网络基本原理，并且介绍了目前比较流行的单阶段和两阶段目标检测算法，并且对检测算法模型的损失函数和优化器原理进行了解释说明。

第三章，基于改进 YOLOv4 模型的林业害虫检测研究。在 YOLOv4 网络的基础上，通过改进骨干网络、检测网络与优化损失函数提出 YOLOv4-TIA 害虫检测网络；整理制作两种场景下的林业害虫数据集，并通过多种数据增强方式对样本集进行扩充；通过优化锚框参数来使模型更适用于林业害虫检测任务，最后在经数据增强后的林业害虫数据集上，对比分析了本文所提检测网络和其他一些主流目标检测网络的检测准确率、召回

率及 FPS，探究优化后的林业害虫检测模型的可行性和性能提升效果。

第四章，林业害虫监测系统的设计与实现。基于改进的 YOLOv4-TIA 害虫检测网络设计并实现了林业害虫监测系统，并对系统进行了测试。

第五章，总结与展望。对基于深度学习的林业害虫监测系统的研究工作进行了总结，并且提出未来可提升的研究方向。

第 2 章 基于深度学习的图像检测识别技术与基础理论

2.1 深度学习

2.1.1 深度学习概述

现如今深度学习多指利用深度神经网络来训练模型，深度神经网络是神经网络的一种延续，神经网络早期为机器学习模型中的一种。多数机器学习模型由数学领域发展而来，而神经网络则是模仿生物神经网络发展而来。神经网络于 1940 年被提出，但发展受限。直到 2006 年，多伦多大学的 Hinton 教授等人发表了关于 deep learning 的论文《Reducing the dimensionality of data with neural networks》^[29]，得益于深度学习理论与计算机硬件 GPU 的发展与支持，深度学习快速发展。深度学习的本质是对数据进行特征学习，通过多层由神经元相连组成的网络提取输入数据的高质量表示，使数据在高维层面线性可分，从而可以通过简单的分类器进行分类与回归任务。深度学习使用低层特征来形成更抽象的高层特征，可以很好地模拟复杂的数据关系^[30]，并对数据进行分层表示，保存有用信息，减少需要处理的数据量。当神经网络结构过于简单，层数少、神经元少时，其表达能力便比较弱；当网络结构过于复杂时，会需要巨大的运算资源。2012 年至 2015 年间，在与深度学习有关的研究中，研究者们比较追求提高网络“深度”，通过增加网络层数来使深度模型变得更为复杂。在之后的研究中，会偏向增加神经元的数量、改变神经元的处理方式，最终形成了各种类型的深度神经网络来进行深度学习。

2.1.2 深度学习框架

在进行与深度学习有关的研究工作时，可选择的框架有很多，比较流行的有 Keras^[31]、Pytorch^[32]、Caffe^[33]、TensorFlow^[34]等。深度学习框架多采用多核计算的方式，表达能力比较强，具有移植性。一个合适的框架对进行目标检测研究至关重要。通过分析比较最终将 Keras 框架用于林业害虫检测识别模型的训练中。

Keras 为基于 Python 的深度学习库，不同于其他深度学习框架，Keras 是一个高级 API 规范，提供了一个简单和模块化的 API 来创建和训练神经网络，隐藏了大部分复杂

的细节，并且可以在 TensorFlow 、 Theano 或 CNTK 上运行，不仅可以作为用户界面，而且可以扩展它在其中运行的其他深度学习框架后端的功能。 Keras 一开始被作为 Theano 的简化前端，几经发展，成为 Google TensorFlow 的一部分，同时支持 Apache MXNet 及 Deeplearning4J。由于获得了广泛支持， Keras 作为框架间迁移工具的地位已不可撼动。开发人员不仅可以对深度学习神经网络算法和模型进行移植，还可以对预先训练的网络和权重进行移植。 Keras 框架主要有以下几点优势：

第一， 用户对深度学习应用程序将有更加良好的体验。 Keras API 经过精心设计，它面向对象、灵活易用，因此对用户很友好。研究人员在定义全新深度学习模型时无需使用十分复杂的后端，因此在进行代码实现时更加简洁明了；

第二， 无缝 Python 集成。 Keras 是本机 Python 包，因此可以便捷访问整个 Python 数据科学生态系统；

第三， 大型的可移植工作主体和强大的知识库。目前，研究人员多将 Keras 与 Theano 后端结合使用，在此基础上产生了大型的工作主体和强大的知识库，可以帮助深度学习研究者轻松进行后端移植，并且经过预训练的模型只需稍作调整便可轻松切换后端；

第四， Keras 免费提供了大量学习资源和代码样例。

2.2 卷积神经网络基本原理

1995 年， LeCun 等人提出了卷积神经网络（ Convolutional Neural Network,CNN ）的概念，其核心是卷积^[35]。 CNN 是一种前馈神经网络，主要由卷积层、池化层和全连接层组成^[36]。不同于之前人工设计特征的方式，卷积神经网络可以根据目标函数的损失值反向回传更新卷积核参数，自动学习可分辨的卷积特征^[37]，可以被看做一个函数或者黑箱，有输入和输出，输入一般为二维像素阵列或图片，输出为我们所需要的东西。目前常被用于处理图像、音频、文本等数据可以被处理为图像格式的文件。

2.2.1 卷积层

卷积层（ Convolution Layer ），也叫特征提取层，其主要任务是对输入数据进行卷积计算。卷积过程实质上可以描述为使用卷积核（ Kernel ）对图像进行扫描的过程。计

计算机看图时看到的是像素的阵列，如图 2-1 所示，输入为 5×5 的灰度图像像素阵列，卷积核如图 2-2 所示，预设为 3×3 的矩阵，使用预设卷积核对图 2-1 进行扫描，做卷积操作，去提取图 2-1 中包含图 2-2 特征的特征，其中步幅（Stride）为 1，即每次从左向右或从上向下移动的距离为一格，不做填充（Padding），即不在图像周围填充像素。

-1	1	1	1	-1
1	-1	1	-1	1
1	1	-1	1	1
1	-1	1	-1	1
-1	1	1	1	-1

图 2-1 灰度图像像素阵列

Figure 2-1 Grayscale image pixel array

-1	1	-1
1	-1	1
-1	1	-1

图 2-2 卷积核 filter 示例

Figure 2-2 Example of convolution kernel filter

图 2-1 和图 2-2 中，-1 表示暗，1 表示亮，具体卷积过程为使用卷积核从左上角到右下角对图像进行扫描，将卷积核预设值与图像中对应位置的像素值相乘并进行加和运算^[38]。以扫描图 2-1 左上角（绿色框区域）为例，具体计算方法如式 2-1 所示。

$$\frac{(-1) \times (-1) + 1 \times 1 + (-1) \times 1 + 1 \times 1 + (-1) \times (-1) + 1 \times 1 + 1 \times (-1) + 1 \times 1 + (-1) \times (-1)}{9} = 0.55 \quad (2-1)$$

式 2-1 得到的值为 0.55，代表着绿色框区域特征与图 2-2 特征不完全一致，但相似度较高为 0.55。所有计算所得的值会被放到新的被称作 feature map 的矩阵中，如图 2-3 所示，反映着卷积核从图片中提取出来的特征，数值越接近 1 表示原图中的特征与卷积核特征越相似，feature map 即是上一层的输出，也是下一层的输入。

0.55	-0.77	0.55
-0.77	1	0.33
0.55	-0.77	0.55

图 2-3 feature map 矩阵图

Figure 2-3 the matrix of feature map

上述卷积运算过程中， 5×5 的输入矩阵经 3×3 大小的卷积核进行卷积后，在步幅为 1 时，得到的输出 feature map 尺寸为 3×3 ，图片尺寸变小，卷积输出 feature map 的尺寸计算方式如式 2-2 和式 2-3 所示：

$$H_{out} = \frac{H + 2p_h - k_h}{s_h} + 1 \quad (2-2)$$

$$W_{out} = \frac{W + 2p_w - k_w}{s_w} + 1 \quad (2-3)$$

式 2-2 和式 2-3 中， H 和 W 表示输入图片的高和宽； p 表示 padding 填充； k 表示 kernel 卷积核； s 表示 stride 步幅； p_h 与 p_w 表示在输入图片宽度和高度的两侧采取等量填充，； k_h 和 k_w 表示卷积核的高和宽； s_h 与 s_w 表示卷积核在宽度与高度方向上滑动的步幅； H_{out} 和 W_{out} 表示输出特征图的高和宽^[39]。从公式可以发现，padding 与 stride 会影响卷积之后的特征图大小，合理的进行 padding 与 stride 的选择可以解决经过多次卷积之后图片尺寸不断减小的问题。

2.2.2 池化层

池化（Pooling），也叫做下采样。现实应用中，输入卷积神经网络的图像往往尺寸不一，且为了节省计算时间，按批量进行训练或者测试时，会产生大量的参数和运算量。为了在提高计算速度、降低计算量的同时能保留图像的主要特征并保证图像特征的平移不变性和旋转不变性，通常会在依次连接的卷积层之间插入池化层（Pooling Layer）对各区域特征进行整合，实现对特征图降维的目的^[40]。池化操作使用特征图中像素附近的统计摘要作为网络特定位置的输出，该输出由 $k \times k$ 大小的卷积核在特征图中滑动计算。池化通常有最大池化和平均池化两种。平均池化操作示意图如图 2-4 所示。

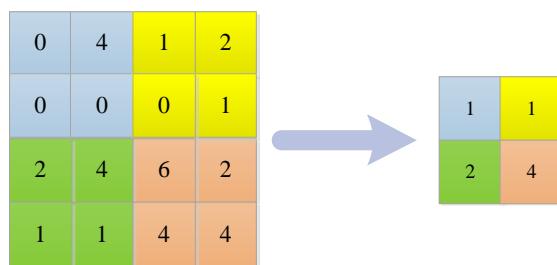


图 2-4 平均池化操作示意图

Figure 2-4 Schematic diagram of average pooling operation

图 2-4 中，平均池化操作使用各部位的平均值代替窗口覆盖的整片区域，图中数字表示像素值。

最大池化操作示意图如图 2-5 所示。最大池化操作保留窗口覆盖区域的最大数值。

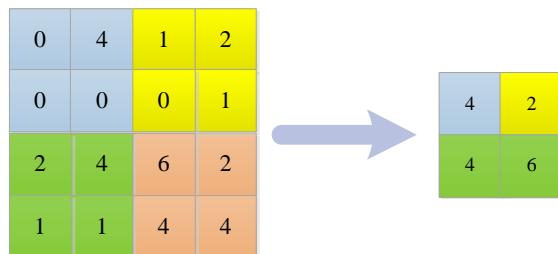


图 2-5 最大池化操作示意图

Figure 2-5 Schematic diagram of maximum pooling operation

2.2.3 全连接层

全连接层（Fully Connected Layers, FC）通常在卷积神经网络的末端，它通过使用本层每一个神经元连接上一层的所有神经元节点来将提取到的特征集合到一起，给出分类任务中目标所属类别的概率，非常适合处理分类任务^[41]。

2.2.4 激活函数

神经网络通常被看作通用函数的合拢器，被期望能计算任何函数（包括线性函数与非线性函数），而卷积计算是一种线性操作，仅通过堆叠卷积层来增加网络深度，逐层相加，最终得到的仍是一个线性函数，其特征表达能力有限，不足以进行复杂数据建模。因此，我们通常会在卷积层后增加非线性激活函数，一为提高网络的表征能力，二为使函数可微，以便执行反向传播的优化策略，提升网络的非线性建模能力^[42]。在深度学习中，有三种常见的激活函数：Sigmoid、Tanh、ReLU。

Sigmoid 激活函数是最先被使用的激活函数之一，它可以将神经元的输出值压缩到（0,1）的区间范围内，其表达式如式 2-4 所示：

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2-4)$$

Sigmoid 激活函数的图像如图 2-6 所示。

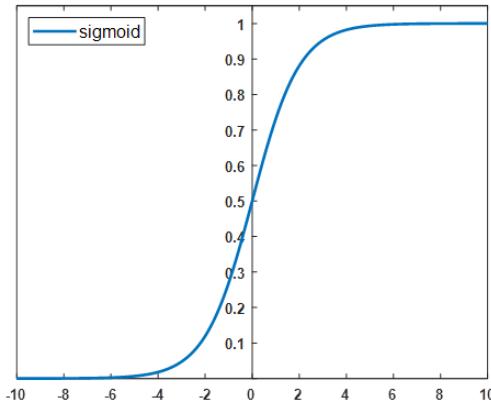


图 2-6 Sigmoid 激活函数图像示意图

Figure 2-6 Sigmoid activation function image diagram

Tanh 激活函数输出以 0 为基准，它会将输入的数字压缩到-1 到 1 的区间范围内。其表达式如式 2-5 所示：

$$f(x) = \frac{1-e^{-2x}}{1+e^{-2x}} \quad (2-5)$$

Tanh 激活函数的图像如图 2-7 所示。

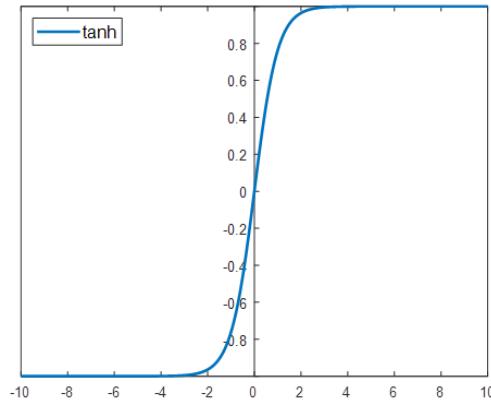


图 2-7 Tanh 激活函数图像示意图

Figure 2-7 Schematic diagram of Tanh activation function image

ReLU 激活函数表达式如式 2-6 所示：

$$f(x) = \max(0, x) \quad (2-6)$$

ReLU 激活函数的图像如图 2-8 所示。

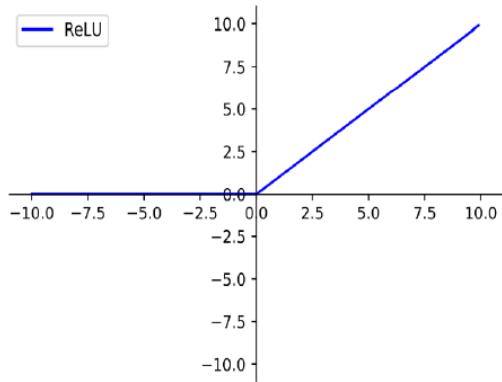


图 2-8 ReLU 激活函数图像示意图

Figure 2-8 Schematic diagram of ReLU activation function image

从图 2-6、图 2-7 可知，Sigmoid 和 Tanh 可以保持原点附近梯度，但因算力提升，研究者们多着眼于搭建多层神经网络模型，而层与层之间神经元的传递是将每一个神经元的输出值经过激励函数的处理，传递给下一个神经元，各层神经元之间为连乘关系，也就是说，当每一层神经元输出值被压缩至 $(0,1)$ 时，连乘结果会越来越小，直至为 0，即造成模型梯度消失问题。在 Sigmoid 中，当神经元的激活状态接近于 0 或 1 时，这些区域的梯度会非常接近于 0；在 Tanh 中，当神经元的激活状态接近于 -1 或 1 时，这些区域的梯度会非常接近于 0，趋近于没有信号通过神经元传递权重和数据^[43]，Sigmoid 和 Tanh 激励函数容易造成梯度消失问题，因此常被用在层数较少的神经网络模型中或用在分类模型输出层中来计算分类概率；而 ReLU 激活函数当 x 大于 0 时，梯度恒为 1，弥补了 Sigmoid 和 Tanh 的不足之处，不易造成梯度消失问题，且运算简单使得收敛速度更快，其收敛速度比 Tanh 增加了 6 倍。因此，ReLU 激活函数及其变种 Leaky ReLU 常被用在多层神经网络的中间层中，如 2.3.2.2 章节中所阐述的 YOLOv4 算法便使用了 Leaky ReLU 作为激励函数。

2.3 基于深度卷积神经网络的目标检测技术

目标检测的任务是在给定图像中确定是否存在待检测目标，主要分为定位任务和分类任务。传统的目标检测方法主要步骤有三：一通过滑动窗口生成候选区域^[44]；二将每个候选区域提取为相应的特征向量，这些向量往往是人工精心设计的；三通过支持向量机等传统的分类方法将提取到的特征向量进行具体分类。传统的目标检测方法虽然可以完成目标检测任务，但仍暴露出了许多问题。近年来，深度学习的引入大大提升了目

标检测任务的性能，尤其是在精度和速度方面，卷积神经网络通过提取图像深层的语义特征，省去了人工提取和设计特征的步骤，现已被广泛应用于目标检测。基于深度学习的目标检测方法流程如图 2-9 所示。

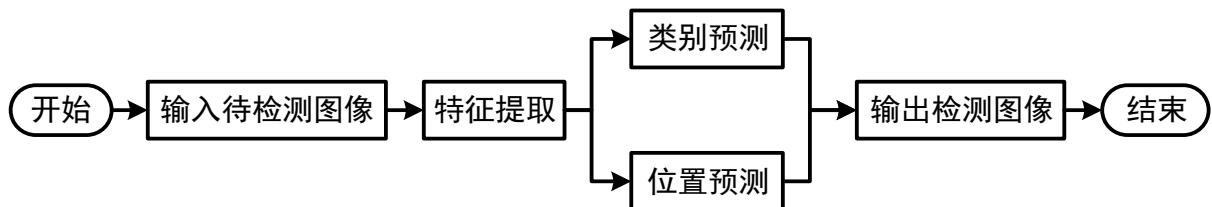


图 2-9 基于深度学习的目标检测算法流程图

Figure 2-9 Flow chart of target detection algorithm based on deep learning

与传统的检测算法相比，它们在多分类任务中具有更好的鲁棒性、准确性和速度^[45]。

2.3.1 两阶段目标检测方法

现代计算机视觉中，基于深度学习的目标检测方法主要分为基于区域建议的双阶段目标检测方法和基于回归的单阶段目标检测方法。双阶段目标检测算法以 R-CNN、Fast R-CNN、Faster R-CNN 为代表，单阶段目标检测算法以 SSD、YOLO 为代表^[46]。

两阶段的目标检测方法先从输入图像中生成可能存在目标的候选区域，再使用卷积神经网络对每一个候选区域进行预测，该类方法检测精度高，但是检测速度慢^[47]。目前主流的两阶段目标检测算法流程框图如图 2-10 所示。

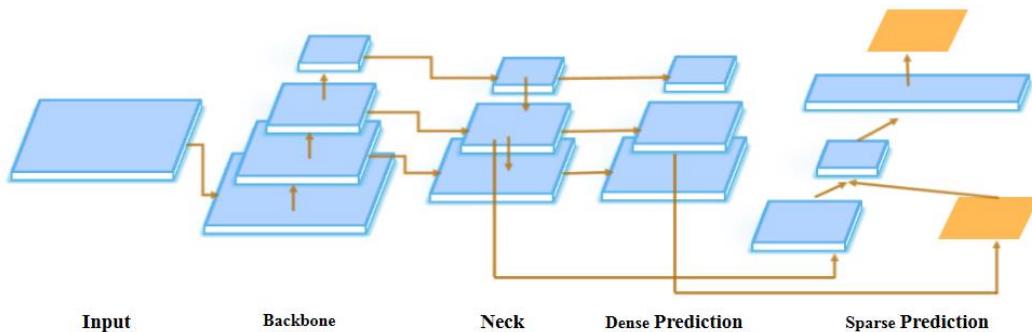


图 2-10 两阶段目标检测算法流程框图

Figure 2-10 Block diagram of the two-stage target detection algorithm flow

图 2-10 所示，该类算法通常包含 Backbone、Neck、Prediction 三大部分。Backbone 通常被称为骨干网络，是一个卷积神经网络，它可以将输入图像转化为特征映射。Neck 会将转化后得到的特征映射做进一步的处理，Prediction 首先在处理后的特征映射上继

续生成一些候选区域，也叫感兴趣区域 ROI(Region of Interest)，这些区域代表可能的目标位置，然后再对这些候选区域进行预测。

2014 年，Ross Girshick 提出了 R-CNN^[48] (Region-based Convolutional Neural Network) 模型，它是最早的基于深度学习的两阶段目标检测方法，其网络结构如图 2-11 所示。R-CNN 先使用传统的选择搜索 (selective search) 方法从输入图像中生成大约 2000 个候选区域，这些候选区域的尺寸往往不一致^[49]。因此，R-CNN 会在接下来的处理中将这些区域缩放到指定大小(通常为 224×224)，再通过卷积神经网络提取相应的特征，最终会由 SVM 等分类器来完成分类和定位。

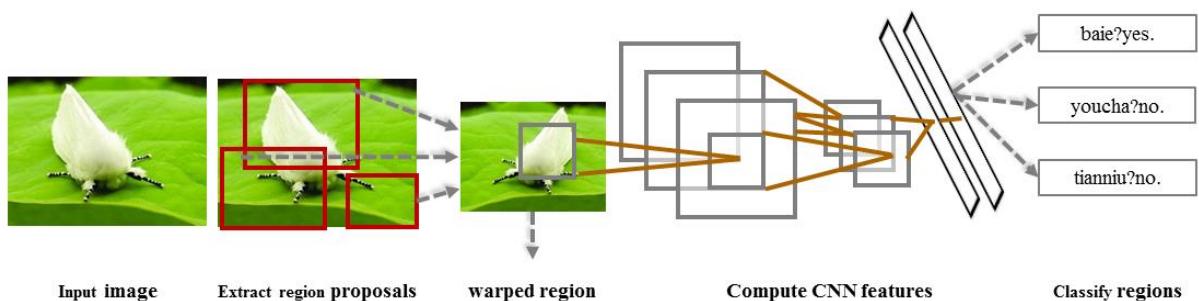


图 2-11 R-CNN 目标检测算法流程框图

Figure 2-11 Block diagram of the R-CNN target detection algorithm flow

R-CNN 方法虽然首次提出了两阶段的目标检测框架，但在选择感兴趣区域时耗费的时间比较长，且由于每个候选区域都需要一次 CNN 的前向传播处理，导致计算量大、检测单张图片耗时超 10 秒，同时对内存的要求比较高^[50]。Fast R-CNN^[51]在 R-CNN 的基础上进行了改进，其网络结构如图 2-12 所示。

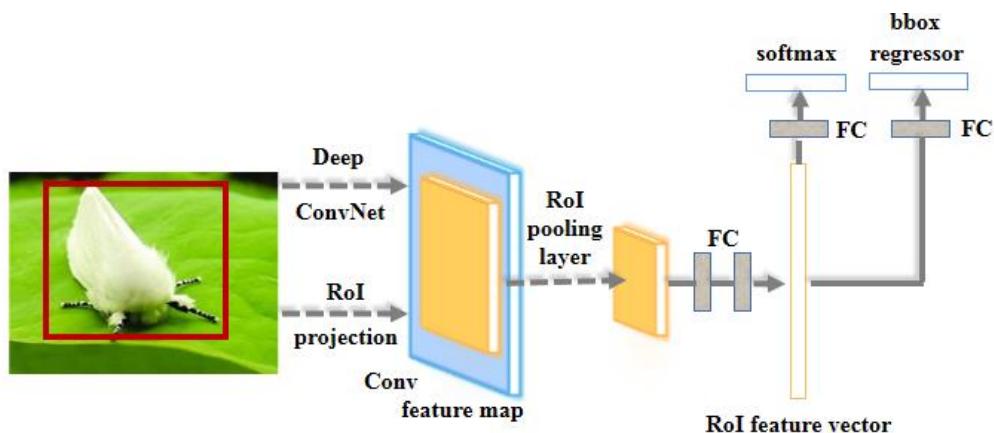


图 2-12 Fast R-CNN 目标检测算法流程框图

Figure 2-12 Block diagram of the Fast R-CNN target detection algorithm flow

Fast R-CNN 先将输入图像送入一个卷积神经网络（也就是骨干网络）提取特征，同时得到候选区域与其相应的特征映射，通过特征共享来减少卷积神经网络的运算量，提升网络的检测速度。与此同时，Fast R-CNN 提出使用 ROI 池化代替 R-CNN 的缩放，将每个候选区域特征映射调整为 7×7 大小，并对每个子区域进行最大池化操作，Fast R-CNN 还将用于目标框分类的分类损失和用于目标框回归的回归损失结合起来进行训练，以简化网络。

尽管 Fast R-CNN 的检测速度有所提升，但仍然无法达到实时检测的要求^[52]。这很大程度上是因为 R-CNN 与 Fast R-CNN 均采用选择搜索算法这样的传统方法选择候选区域，数量庞大，需要耗费大量的内存与计算量。Faster R-CNN 对此做出了改进，其网络结构如图 2-13 所示，它使用深度学习的方法代替传统搜索方法，提出区域提议网络 (Region Proposal Network, RPN)，并首次往两阶段目标检测框架中引入了基于 Anchor 的检测概念^[53]。Faster R-CNN 使用滑动窗口在特征图上学习到候选区域，在对候选区域提取特征后再进行分类和回归操作。

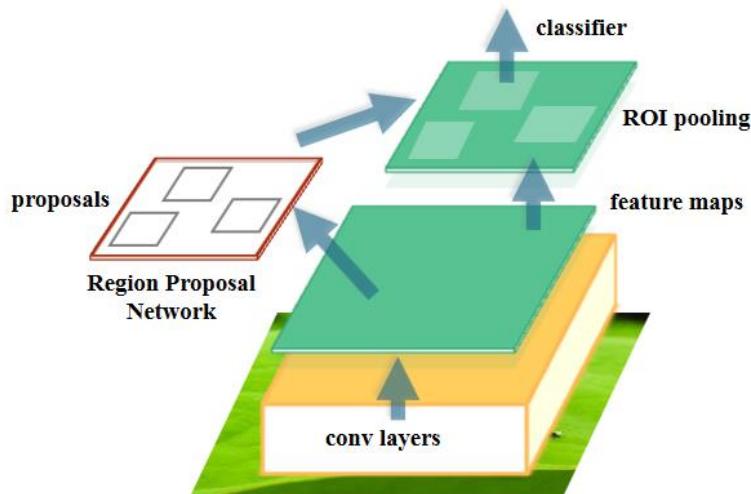


图 2-13 Faster R-CNN 目标检测模型图

Figure 2-13 Block diagram of the Faster R-CNN target detection algorithm flow

Faster R-CNN 中 RPN 结构如图 2-14 所示。它以输入图像的特征映射作为输入，使用 3×3 的滑动窗口搜索候选区域，每次窗口滑动时为其中心像素设置 k 个 anchor，每个 anchor 的形状和尺寸不一。对于每个 anchor，均包含两类参数：一类用来区分 anchor 中的是目标还是背景，共产生 $2k$ 个参数用于分类，通过 Sigmoid 函数判断该 anchor 位置包含目标物体的概率；一类用来对边界框的回归量进行预测，共产生 $4k$ 个参数用于

回归，在 anchor 位置存在目标的情况下，对人为定义的 anchor 进行尺度修正。

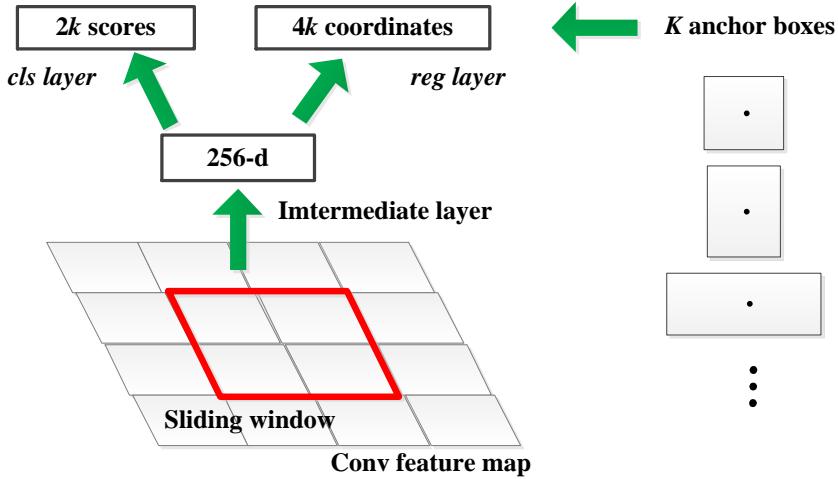


图 2-14 RPN 网络结构图

Figure 2-14 RPN network structure diagram

在 Faster R-CNN 中，每个边界框的位置基于先验框 anchor 进行预测，，，则预测框计算公式如式 2-7、式 2-8、式 2-9、式 2-10 所示。

$$x^* = (x - x_a) / w_a \quad (2-7)$$

$$y^* = (y - y_a) / h_a \quad (2-8)$$

$$w^* = \log(w / w_a) \quad (2-9)$$

$$h^* = \log(h / h_a) \quad (2-10)$$

式 2-7、式 2-8、式 2-9 和式 2-10 中， (x, y, w, h) 表示真实边框左上角顶点坐标以及宽和高， (x_a, y_a, w_a, h_a) 表示 anchor 的位置坐标， (x^*, y^*, w^*, h^*) 表示网络需要学习的目标位置相对于真实框的偏移量。

2.3.2 单阶段目标检测方法

单阶段方法采用端到端的网络结构，使用一个将分类与回归封装于一体的前向卷积神经网络，将图像的每个可能区域看作候选区域，并直接预测目标的位置信息和分类概率，计算效率高，检测速度快且检测精度也不断提升。且由于近年来，将目标检测任务应用到以无人机、机器人和摄像头等计算能力有限的终端设备上的需求越来越多，研究者们渐渐意识到，相比于双阶段目标检测模型而言，单阶段目标检测模型更易于落地。

2.3.2.1 SSD 系列算法

2016 年, Wei Liu 在 ECCV 2016 上提出了 SSD (Single Shot MultiBox Detector) 模型, 它属于基于深度学习的 one-stage 目标检测方法, 同 Yolo 一样, 在检测速度上有比较明显的优势^[54]。与 Yolo 不同的是, 该模型使用基于特征金字塔的多尺度预测结构对目标进行预测。SSD 模型输入图像大小有两种, 分别为 300×300 的分辨率和 512×512 的分辨率, 以 300×300 大小的输入为例, SSD 模型结构图如图 2-15 所示。

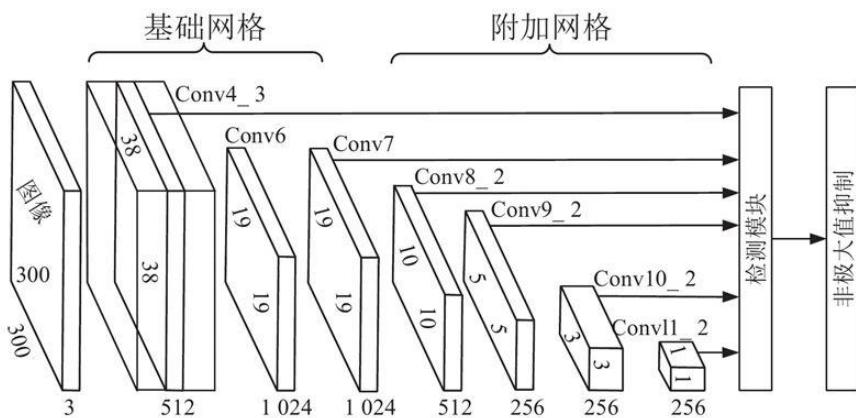


图 2-15 SSD 网络结构示意图

Figure 2-15 SSD network structure diagram

SSD 模型以去除全连接层的 VGG-16 卷积网络作为骨干网络, 提取输入图像特征, 将得到的特征图作为检测网络的特征输入, 在特征检测网络中, 引入特征金字塔结构的概念, 添加四组卷积层, 通过 1×1 和 3×3 这两组不同的卷积核对特征图进行降维与进一步的特征提取, 与骨干网络中的最后两层相结合, 共同组成多尺度特征检测网络来进行不同尺度下的特征检测任务。模型最终会在从 38×38 到 1×1 不等的六组不同尺度的特征图上分别进行目标分类概率预测与边框回归^[55]。SSD 同样通过 anchor 生成机制生成不同比例的先验框, 在此基础上来预测目标的预测框 (c_x, c_y, w, h)。每个先验框生成不同类别的置信值和 4 个坐标值 (x, y, w, h), 从而用于预测不同尺度的目标的类别和位置。与 YOLO 相比, 该模型少了全连接层, 采用特征金字塔式的多尺度预测结构, 较小的特征图因有较大的感受野而被用来对大目标进行检测, 反之, 较大的特征图被用来对小目标进行检测, 最终可以提升检测网络对目标的检测能力。

2.3.2.2 YOLO 系列算法

You Only Look Once (YOLO) 系列算法^[56-59]是 Facebook 公司提出最早的单阶段目标检测框架，有 v1、v2、v3、v4 等多个版本。YOLOv4 是该系列网络中检测性能最优的一代，实现了速度和精度的最佳平衡。YOLOv1 由 Redmon J 等于 2016 年提出，当图像被输入 YOLOv1 后，通常会被划分成 $S \times S$ 个网格（ S 一般为 7），每个小方格负责预测 B (B 一般取 2) 个边界框 (bounding box) 及其对应的位置信息 $[x, y, w, h]$ 和置信度 C (confidence)，并针对每个小方格预测 C 个类别的分数 (Score)^[60]。每个 bounding box 都由 (x, y, w, h, c) 这 5 个预测值组成， (x, y) 表示预测边界框的中心相对于所在真实框位置的偏移量， (w, h) 表示预测边界框相对于整个图像的宽和高。置信度 C 包含两层信息，一为目标 Object 是否存在于当前网格；二为该网格是否能准确预测目标^[61]。置信度 C 计算公式如式 2-11 所示。

$$C_i^j = P_{i,j}(\text{Object})^* IOU_{\text{pred}}^{\text{truth}} \quad (2-11)$$

式中, C_i^j 表示第 i 个网格单元中第 j 个边界框的目标置信度得分; $P_{i,j}(\text{Object})$ 表示网格中目标存在的概率, 取值范围为 0 到 1; truth 表示真实值, pred 表示预测值, $IOU_{\text{pred}}^{\text{truth}}$ 表示预测框与真实框之间的交并比。

YOLOv1 网络结构如图 2-16 所示。

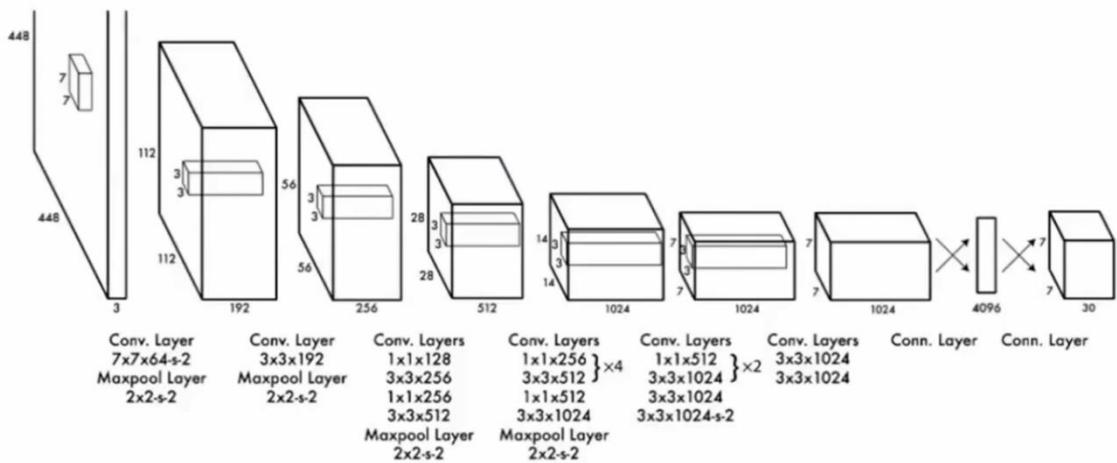


图 2-16 YOLOv1 网络结构图

Figure 2-16 YOLOv1 network structure diagram

YOLOv1 的检测效率可达 45 帧每秒，但却不如两阶段目标检测方法的检测精度高，

且存在定位不够准确，召回率较低的问题。在 CVPR2017 上，YOLOv2 被提出，主要从七个方面对 v1 进行改进：一是通过在每个卷积层后增加批标准化（Batch Normalization, BN）处理来加速模型收敛、使 mAP 提升了 2%、并起到了防止过拟合的作用；二是使用 448×448 （YOLOv1 采用的为 224×224 ）的图像输入来训练 backbone 使 mAP 提升了 4%；三是使用锚定框来预测边界框，使召回率从 81% 提高到 88%；四是在训练集上使用 k-means 聚类来找到合适的 anchor，可以使网络学习、预测更为简单、准确；五是通过限制预测坐标信息，直接预测边界框中心坐标，使网络训练更加稳定；六是通过融合高层特征图与低层特征图，使模型检测效果有了 1% 的提升；七是采用多尺度训练来提升网络的鲁棒性。YOLOv2 中使用 anchor 预测时需要回归的是真实框与 anchor 的偏置量，计算方法如式 2-12、式 2-13、式 2-14、式 2-15 所示。

$$b_x = \sigma(t_x) + c_x \quad (2-12)$$

$$b_y = \sigma(t_y) + c_y \quad (2-13)$$

$$b_w = p_w e^{t_w} \quad (2-14)$$

$$b_h = p_h e^{t_h} \quad (2-15)$$

式 2-12、式 2-13、式 2-14 和式 2-15 中，每个边界框的位置基于 t_x , t_y , t_w , t_h 进行预测， (t_x, t_y, t_w, t_h) 表示网络需要回归的真实边框与 anchor 的偏置量； (c_x, c_y, p_w, p_h) 表示 anchor 的位置坐标； $\sigma()$ 是一个 sigmoid 函数； (b_x, b_y, b_w, b_h) 表示最终预测边界框的位置坐标。

为进一步提高 YOLO 算法对小目标的检测精度，Redmon 和 Farhadi 经多种尝试提出了 YOLOv3 方法。YOLOv3 使用包含 53 层卷积层的 Darknet-53 网络作为特征提取网络，与 YOLOv2 中的 Darknet-19 网络相比，Darknet-53 中使用了残差单元来构建更深的网络结构^[62]。在目标预测方面，YOLOv3 采用 FPN 架构（Feature Pyramid Networks for Object Detection）来进行多尺度检测，当输入为 416×416 时，YOLOv3 预测 3 个尺度的特征图，大小分别 (13×13) , (26×26) , (52×52) ，从而提高了对小目标的检测精度^[63]。

2020 年，Alexey 提出了 YOLOv4 目标检测模型，是 YOLO 系列模型的最新版本，时至今日，YOLOv4 已成为目标检测的首选方案之一，并且相对于上述其他算法实现了速度与精度的最佳平衡。其网络结构如图 2-17 所示。不同于 YOLO v3，YOLOv4 在数

据处理中引入了 mosaic 数据增强，使用 CSPDarknet53 替代 YOLOv3 中的 Darknet53 作为主干网络，加入了能够显著地改善感受域大小，且对速度影响不大的 SPP block，使用 PANet 替代 FPN 作为颈部网络来进行多通道特征融合，同时将 WRC、CSP、CmBN、SAT、Mish-activation、Mosaic data augmentation、DropBlock regularization、CioU loss 等多种训练技巧进行组合论证，以提高 Convolutional Neural Network(CNN)的精度，并对激活函数与损失函数均进行了优化，这些方式使得 YOLOv4 目标检测器可以在具有较小 mini-batch 批处理大小的单 GPU 上进行训练，使得使用单 1080 Ti 或 2080 Ti GPU 训练出超快速和精确的目标检测器成为可能^[59]。

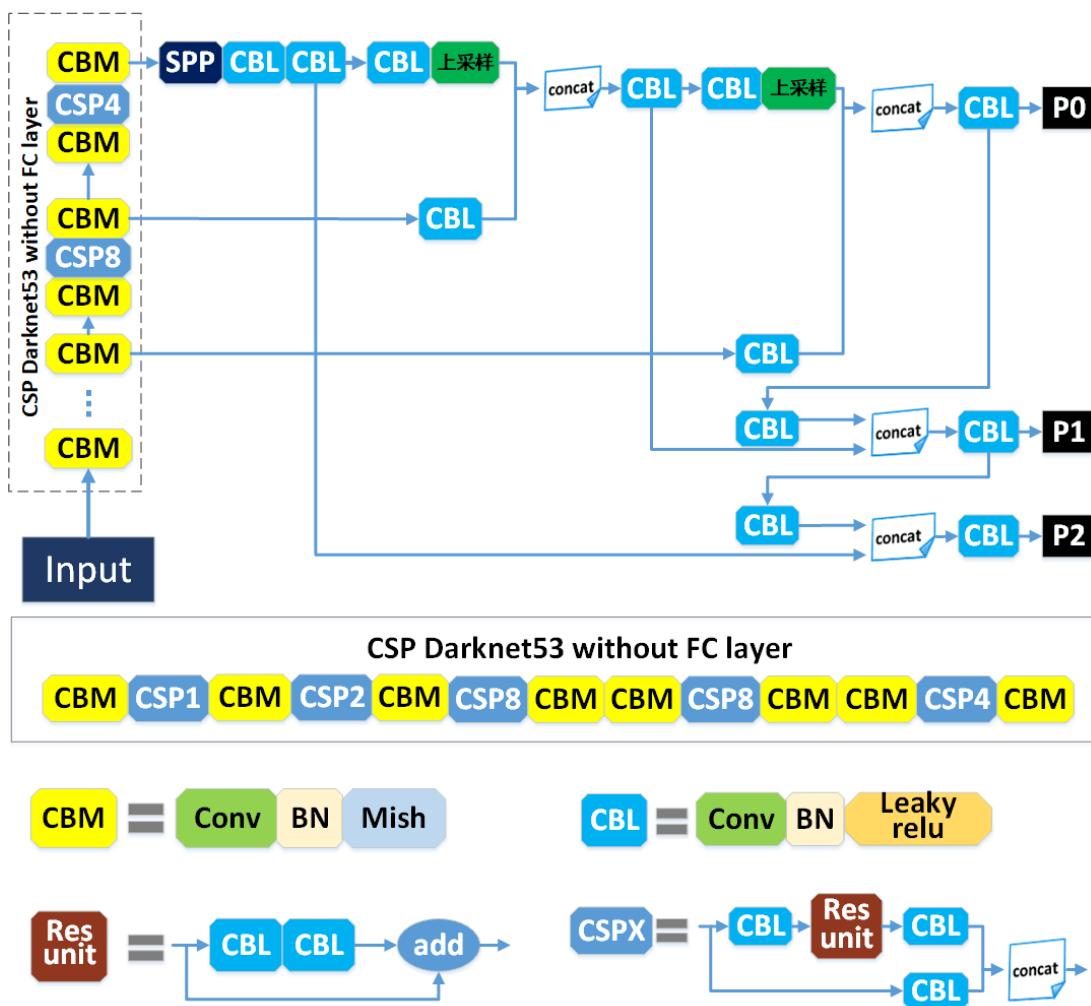


图 2-17 YOLOv4 网络结构示意图

Figure 2-17 YOLOv4 network structure diagram

(1) CSPDarknet53

YOLOv4 模型的第一大改进点是使用 CSPDarknet53 作为其骨干网络，CSPDarknet53

主要由 CBM 模块和 CSP 模块构成，CBM 模块由卷积层（Conv）、批归一化层（Batch Normalization, BN）和 Mish 激活功能组成，CSP 模块包含两条支路，一条是主干部分的卷积，一条用来生成一个大的残差边，通过对两条支路的跨级拼接与通道整合增强 CNN 的学习能力，CSP 结构以 CBL 模块为基础单元，其中 CBL 模块主要由卷积层（Conv）、批归一化层（Batch Normalization, BN）和 Leaky ReLu 激活功能组成^[64]。

（2）SPP-NET

YOLOv4 的另一大改进点是，在检测部分使用了 SPP（spatial pyramid pooling）模块，可以使得任意大小的特征图都能够转换成固定大小的特征向量^[65]。这是由于在 YOLOv3 中，预测部分的卷积神经网络需要固定尺度的输入。因此，当需要检测不同大小的图片或者对不同大小的特征图进行拼接时，就需要经过 crop 裁剪、缩放等操作变换图片尺寸，这会在一定程度上导致图片信息丢失，降低识别精度。SPPNet 方法提出了整体学习的思路。其基本思想是，先提取出整张图像的特征图，再在空间金字塔池化层利用空间关系从整张图像的特征图中提取出各个区域建议网络(region proposal)的特征。空间金字塔池化层可以将任意大小的输入向量转换为固定长度的特征向量，不需要考虑输入尺寸的大小，并且可以做到在图像变形情况下表现稳定。SPP 层通常作用最后一层卷积之后，不仅允许测试的时候输入不同大小的图片，训练的时候也允许输入不同大小的图片，通过不同尺度的训练同时可以防止模型过拟合（overfit）。

SPPNet 的网络结构图如图 2-18 所示。深蓝色部分代表卷积之后的特征图，空间金字塔池化层以不同大小的块来提取特征，首先将卷积之后的特征图像按照预先给定的 ($S \times S$) 划分成不同大小的子区域，图中 S 的取值分别为 1、2、4，划分后的区域大小分别为 1×1 , 2×2 , 4×4 ，可以得到 21 种不同的块(Spatial bins)，紧接着对每一个块进行最大池化，最终提取到 21 维的特征向量。

（3）PANet

YOLOv4 采用 PANet 的思想对不同层级的特征进行融合，PANet（Path aggregation network）指路径聚合网络，其作用是将来自骨干网络不同层级的参数进行聚合，它是由 Mask R-CNN 改进而来，具体改进有三^[66]。首先，原始 Mask R-CNN 没有很好地利用低层信息，高层的 Feature maps 关注物体整体，低层的 Feature maps 关注物体的纹理图案，使用低层的信息可以对物体进行更好地定位。对此 PANet 增加了 Bottom-up Path

Augmentation 将低层的信息传导到高层中去，同时减少了高层到低层的信息流通需要穿过的卷积层数。其次，原 ROI Pooling 只在最后一层上提取信息，而 PANet 则使用 Adaptive Feature Pooling 同时对多个层级进行 ROI Pooling，将多层级的信息整合后进行预测。最后，Mask 预测分支将 FCN 式的预测和 fully-connected 式的预测相融合，形成了 PANet 的特征融合结构，同时关注局部与整体信息，从而提升对特征信息的利用率。

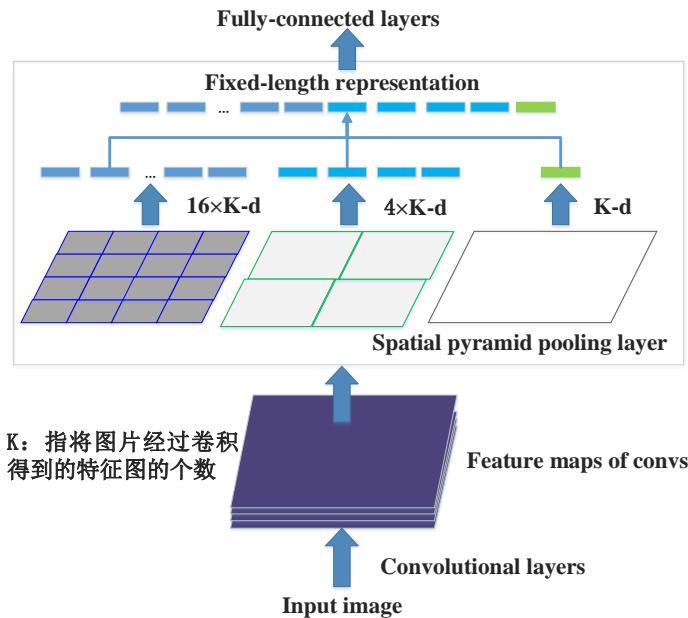


图 2-18 SPPNet 的网络结构示意图

Figure 2-18 SPPNet network structure diagram

YOLOv3 在进行多通道特征融合时借鉴的是 FPN 的思想，FPN 网络结构如图 2-19 所示，它只通过 Up-bottom Path Augmentationd 的方式将高层信息传导到低层去，同 Mask R-CNN 一样，不能很好地利用特征图低层位置信息。

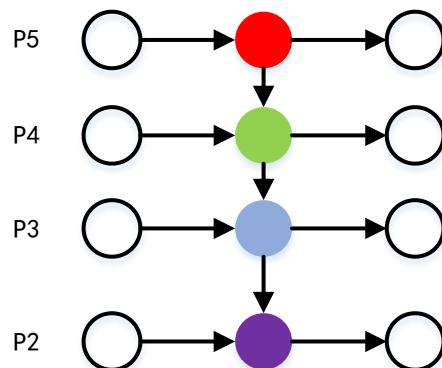


图 2-19 FPN 网络结构示意图

Figure 2-19 FPN network structure diagram

YOLOv4 使用 PANet 替换 FPN 来实现双向特征融合，PANet 网络结构如图 2-20 所示。

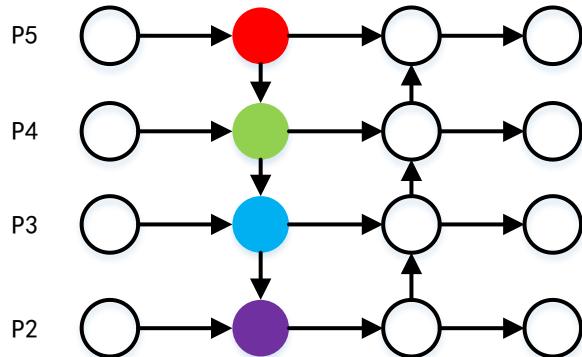


图 2-20 PANet 网络结构示意图

Figure 2-20 PANet network structure diagram

对比 FPN 和 PANet 网络结构可以看出，FPN 结构只能实现顶层向底层的特征融合，而 PANet 可以实现顶层向底层和底层向顶层的双向特征融合，通过自底向上的路径增强，利用准确的低层定位信号增强整个特征层次，从而缩短了低层与顶层特征之间的信息路径。

2.3.3 检测算法模型的损失函数与优化器

(1) 损失函数

损失函数是指导一个网络训练的标准，其控制着网络学习的方向，监督着训练的正确性^[67]。一般来说，网络最后一层输出其识别结果，通过损失函数衡量将该结果与标签之间的距离，优化损失值以训练整个网络。损失函数的设计需与评估指标相匹配。目标检测是一个多任务学习的问题，通常有分类和回归两个子任务用于衡量目标框识别和定位的准确性。

对于分类分支，一般采用 softmax 交叉熵(Cross-Entropy)损失，即先使用 softmax 来将分类输出标准化为概率分布，在 softmax 激活函数下，分类输出结果被抑制为[0,1]范围内的概率向量，其每个维度代表样本为该类别的概率；紧接着使用 Cross-Entropy 来刻画预测值与真实结果之间的相似度，交叉熵计算公式如式 2-16 所示。

$$L = -\sum_{j=1}^C y_j \log s_j \quad (2-16)$$

式 2-16 中, L 表示损失; C 表示类别数; s_j 表示 softmax 输出向量 S 的第 j 个值, 即当前预测样本属于第 j 个类别的概率; y_j 表示 label 即正确的类别标签。

由于利用 softmax 交叉熵方法来计算损失时要求传入的 labels 必须经过 one-hot 编码, 即 y 是一个 $1 \times C$ 的向量, 共有 C 个值, 且只有真实标签对应位置的值为 1, 其他位置值均为 0。因此, 式 2-16 中, 交叉熵损失可以简化为 $L = -\log s_j$ 。

对于回归分支, 在目标检测中一般采用 smoothL1 损失函数, 该损失是 L1 损失(绝对误差)和 L2 损失(均方误差)的结合^[68], 计算公式如式 2-17 所示。

$$L_{smoothL1} = \frac{1}{n} \sum_{i=1}^n \begin{cases} 0.5 * (y_i - \hat{y}_i)^2, & \text{if } |y_i - \hat{y}_i| < 1 \\ |y_i - \hat{y}_i| - 0.5, & \text{if } |y_i - \hat{y}_i| \geq 1 \end{cases} \quad (2-17)$$

式中, n 表示类别数; L 表示 Loss 损失, $L_{smoothL1}$ 表示使用 smoothL1 方法计算的损失值; y_i 表示目标值, 即期望输出的正确类别标签; \hat{y}_i 表示模型输出值。

(2) 优化器

卷积神经网络一般用反向传播算法计算参数的梯度, 对每一个可学习参数, 基于其梯度进行更新学习, 在第 t 次迭代(训练)时权重参数 w 的更新方式计算方法如式 2-18 所示。

$$w_t = w_t - \alpha \bullet \nabla w_t \quad (2-18)$$

式中, α 表示学习率, 一般与批次大小成正相关。对于学习率的设置, 一般是在迭代初期设置一个较小的初始值(一般为 0.001), 称为热启动, 该学习率称为 warm-up 学习率, 用于使网络能够在训练初期适应新数据集的分布。之后, 以一个较大的学习率(0.01)开始, 训练时观察损失值, 当损失在几代训练都不发生变化时, 将当前学习率成比例下降(一般下降比例为 0.1)重复 2 次达到模型收敛。此外, 还可以通过训练中途重新启用较大学习率的方式来跳出局部最优。

在目标检测问题中, 每次迭代的批次会较小, 一般为 1-2 个样本, 此时每次迭代的梯度都只是通过小批次数据计算得到的, 这样的方式称为随机梯度下降(Stochastic

Gradient Descent,SGD)。由于随机性问题，网络的训练方向每次可能会有较大偏差导致收敛太慢。因此，通常在 SGD 中引入一个积攒历史梯度信息动量 η (momentum)来加速 SGD,momentum SGD 梯度下降的训练方式如式 2-19 和式 2-20 所示。

$$\Delta w_t' = \eta \Delta w_{t-1} + (1-\eta) \Delta w_t \quad (2-19)$$

$$w_t = w_t - \alpha \bullet \Delta w_t' \quad (2-20)$$

其中 η 为动量系数，一般取 0.9，表示 90% 的梯度来自于迭代前积攒的历史梯度，剩下 10% 来自于本批次。Momentum SGD 主要解决 SGD 的两个问题：(1) 是随机梯度的方法对噪声和外点敏感；(2) SGD 在收敛过程中网络损失值震荡较大，训练不平滑。在 Momentum SGD 的基础上，近年来出现了 RMSprop 和 Adam 等优化方法，通过引入二阶动量进一步优化学习曲线。

2.4 本章小结

本章首先介绍了深度学习的由来以及框架，并且介绍了深度学习中最重要的卷积神经网络的基本原理，对其基本结构分开进行了阐述。在第三小节中介绍了当前比较流行且先进的几种基于深度学习的目标检测算法及其发展历程，重点分析 YOLOv4 算法的优越性，为后续改进与研究提供理论支持，其他算法在后续章节中将被作为对比模型以验证本次研究中所提出方法的优越性。

第3章 林业害虫检测网络的设计

3.1 基于 YOLOv4 改进的林业害虫检测网络

3.1.1 骨干网络的改进

YOLOv4 网络是平等对待每个通道特征的，这在一定程度上限制了算法的检测性能。为进一步提升模型精度，本研究利用 Triplet Attention 对 YOLOv4 中的 CSPDarknet53 特征提取网络进行改进，改进后的模型命名为 YOLOv4-TIA。Triplet attention 模块是一种几乎没有参数、且不涉及维数降低的廉价且有效的注意力机制，是一个额外的神经网络，网络结构如图 3-1 所示。

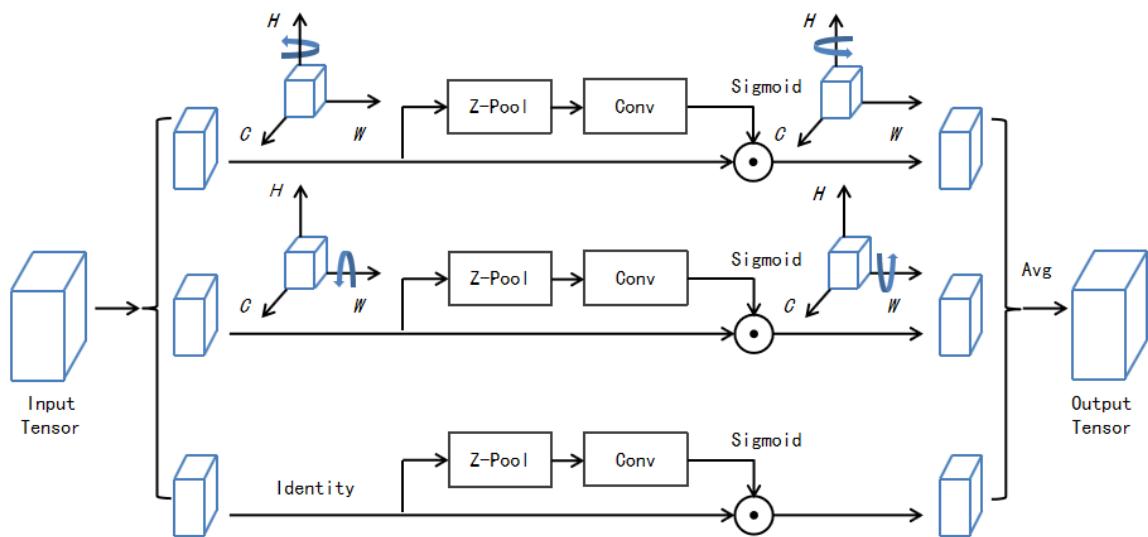


图 3-1 Triplet Attention 网络结构图

Fig.3-1 Network structure diagram of Triplet Attention

其原理是通过旋转操作和残差变换建立维度间的依存关系，主要包含三个分支，其中两个分支分别用以捕获通道 C 维度和空间维度 W/H 之间的跨维度交互，剩下的一个分支用来捕捉(H,W)维间的依赖关系，通过对 3 个 branch 的输出进行平均来实现跨维度交互。

传统的通道注意力计算会选取高层特征图进行全局池化，以具有全局的感受野，但这种计算方式会丢失大量的空间信息，也无法体现通道维数和空间维数之间的相互依赖

性，传统的空间注意力计算同理。其原理如图 3-2 所示。

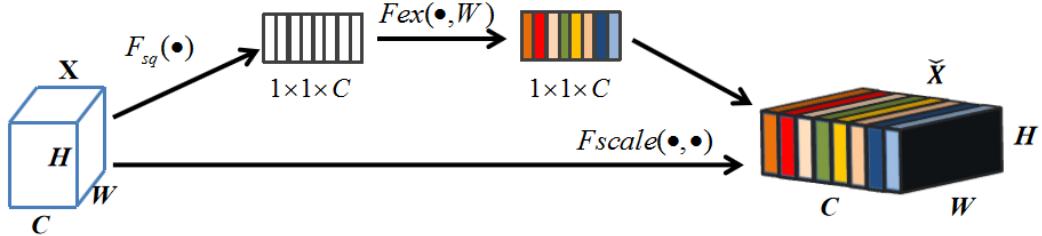


图 3-2 通道注意力机制原理图

Fig.3-2 Schematic diagram of channel attention mechanism

基于 Spatial 和 Channel 的 CBAM 虽然可以体现通道与空间之间的相互依赖性，但通道注意与空间注意的计算是分离的，且在混合维度进行添加会带来非常巨大的计算量的消耗。Triplet attention 的意义在于：一是不会丢失大量的空间与通道信息；二是可以进行跨维度交互；三是不像 CBAM 一样带来极大的算力消耗。

本文将 Triplet attention 模块引入 CSPDarknet53，使网络可以通过自动学习获取跨维度交互，提升有效的特征通道权重，从而让网络重点关注重要的特征通道。结合三分支注意力模块改进的 YOLOv4-TIA 模型骨干网络结构如图 3-3 所示。

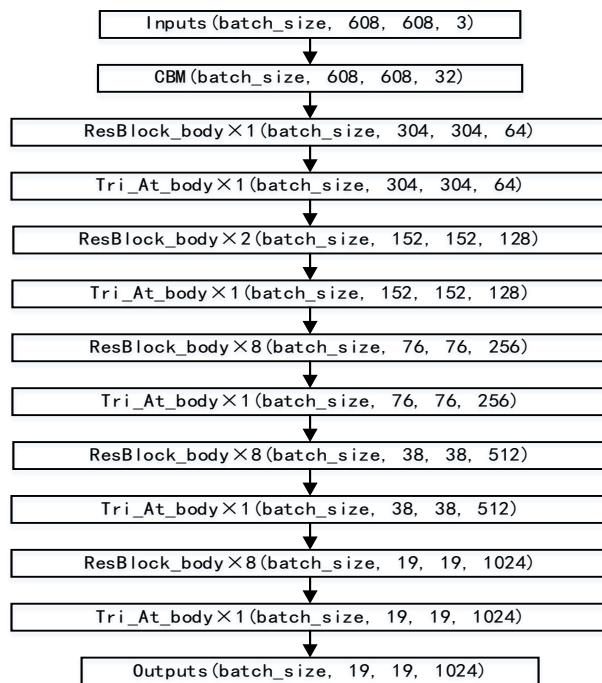


图 3-3 YOLOv4-TIA 骨干网络结构图

Fig.3-3 the backbone network structure diagram of YOLOv4-TIA

3.1.2 融合改进 PANet 结构的 YOLOv4-TIA

输入图片经 YOLOv4-TIA 网络中的骨干网络提取特征后，需要经过检测颈部网络进行预测输出。YOLOv4 采用 PANet 作为颈部网络，其网络结构如图 3-4 所示。

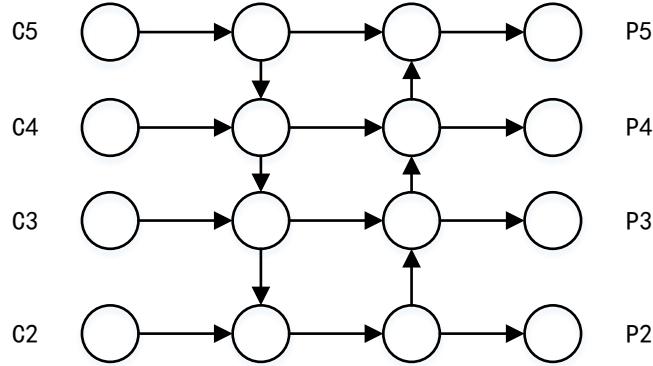


图 3-4 PANet 网络结构图

Fig.3-4 Network structure diagram of PANet

图 3-3 中 C 表示输入特征图，Pre 表示输出特征图。PANet 结构虽然可以将高层特征图的强语义信息与低层特征图的定位信息进行双向融合，但在融合时是将不同层级的特征直接进行相加，并未考虑它们对最后输出贡献不等的问题。为了使网络更为高效，并让网络可以学习不同输入特征的权重，本文提出使用改进的 PANet 结构，即 BiFPN 的特征融合方式，去除掉 YOLOv4 特征融合网络结构中只有一条输入边和输出边的结点，并且在输入和输出结点是同一层级时，使用残差的方式添加一条额外的边，在不增加计算开销的同时融合更多的特征^[66]。改进的 YOLOv4 特征融合方式示意图如图 3-5 所示。

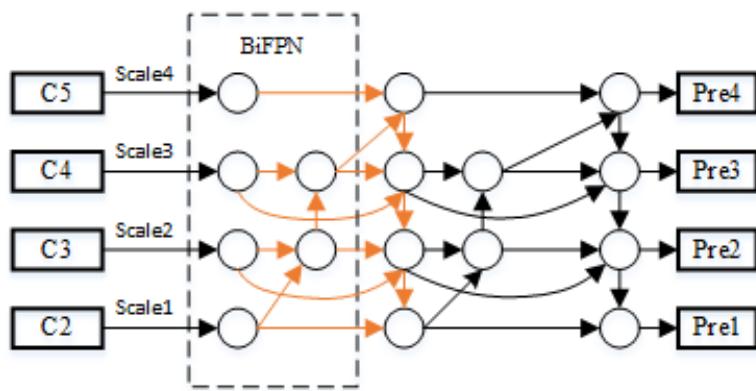


图 3-5 改进的 YOLOv4 特征融合方式示意图

Fig.3-5 Network Schematic diagram of improved YOLOv4 feature fusion method

输入图片经过 YOLOv4-TIA 的骨干网络得到各层级特征图后，通过跨级连接和同级跳跃连接这样高效的多尺度特征融合方式，最终获得尺度分别为 152、76、38 和 19 的四个层级的特征图，并在此基础上进行预测。

3.1.3 损失函数的优化-Focal Loss 函数

由于林业监测中采集到的昆虫数据存在样本不均衡且不易进行分类的问题，在 YOLO v4-TIA 中使用 Focal loss 来优化分类损失，Focal loss 可以用来解决一阶段模型中背景分类不平衡和正负样本严重失衡的问题^[69]。YOLOv4 中损失函数分为位置损失、置信度损失和类别损失三部分，将 Sigmoid 与交叉熵损失函数相结合来进行计算，在多分类任务中，交叉熵表示的是激活函数实际输出与期望输出值的距离，交叉熵越小表示二者的概率分布越近^[70]，计算公式如式 3-1 所示。

$$L = -\sum_{j=1}^C y \log y' \quad (3-1)$$

式中， L 表示交叉熵损失； C 表示分类类别数； j 的取值范围为 $[1, C]$ ； y 表示期望输出， y' 表示经过激活函数的实际输出。改进后 Focal Loss 计算如式 3-2 所示， α 取 0.25， β 取 2：

$$L_f = -\alpha [1 - y']^\beta \log y' \quad (3-2)$$

式中， L_f 表示焦点损失； α 因子的作用是平衡样本数量， $\beta > 0$ 的作用是减少易分类样本的损失。

3.2 数据集与数据预处理

3.2.1 数据集介绍

在林业害虫检测任务中，北京林业大学采集的诱捕器环境下的昆虫数据集场景复杂度不够，数据集规模小，应用场景有限。而目前现有的一些公开数据集以农作物害虫目标为主，林业害虫检测数据集稀少，且多旨在解决害虫识别问题，但在实际应用中，害虫监测与预警防治的需求不仅在于害虫种类的识别，更需要对害虫进行精确的定位。因为准确的检测与识别性能可以方便害虫的计数统计，可以在此基础上进行进一步的虫情

分析，这对于当前害虫监测预警工作至关重要，同时可以为实时防治提供技术支持，准确的定位信息可以为科学防治提供理论依据。因此，本此研究构建包含两种场景的林业害虫检测图像数据集，为目前林业害虫检测研究提供数据实验平台。

（1）诱捕器环境下的昆虫数据集

对于以诱捕器环境为主的较简单背景下的林业害虫检测场景，本文以北京林业大学采集的诱捕器环境下的多类别昆虫数据集作为数据基准。该数据集中共有 1973 张图像，分辨率在 1300×1300 左右，以.jpg 的格式存储。图像标注格式遵循 PASCAL VOC 标注格式，以 xml 文件保存。每张图像中均包含多类昆虫，该数据集共涵盖 Boerner、Leconte、Linnaeus、acuminatus、armandi、coleoptera 和 linnaeus 七类昆虫目标。

（2）复杂背景下的昆虫数据集

本文不仅关注诱捕器环境下的害虫检测任务，还关注复杂场景下的害虫检测问题。为此，本文尝试建立复杂场景下的数据集，为自然场景中的害虫检测任务提供可行的数据基准。本文从网络搜集整理了自然环境下的 5 种常见森林害虫图像，图像中目标类别包含松材线虫、美国白蛾等，共整理 2000 幅有效害虫样本，使用 LabelImg 对害虫目标边界框进行标注，相关信息以一维向量 (x, y, w, h, c) 表示，其中 x, y 为边界框的左上角点坐标， w, h 为边界框的宽和高， c 表示该害虫的类别。

复杂背景下的昆虫数据集中各类样本各包含 400 张图片，所选取图片中昆虫大小、姿态各不相同，以成年后害虫为主，以满足模型对检测分类的复杂度要求。由于时间有限，所标注样本量不足以构成深度学习中害虫检测任务的要求。因此，本文通过数据增强操作对两种数据集均进行了扩充，最终图像数量分别达到了 10758 张与 12000 张，可被应用于基于深度学习的林业害虫检测的相关研究，具体的数据增强内容将在 3.2.2 小节中进行详细阐述。

3.2.2 数据预处理

在解决深度学习目标检测任务的过程中，往往需要花费大量的精力去处理数据。对收集到的数据进行整理、分析，挖掘出数据中有用的信息，对训练神经网络来说十分重要，良好的数据处理不仅会加速模型训练，也会提高模型效果^[71]。因此，数据分析是必不可少的环节。本次研究中，首先对诱捕器环境和自然环境下的昆虫原始数据集中每类

目标样本数量进行统计，以明确各类别的数量分布，诱捕器环境下的统计结果如表 3-1 所示。

表 3-1 诱捕器环境下的昆虫数据集中每类目标样本统计结果

Table 3-1 The statistical results of each type of target sample in the insect data set in the trap environment

昆虫目标	Boerner	Leconte	Linnaeus	acuminatus	armandi	coleoptera	linnaeus
样本数量	1595	2216	818	953	1765	2091	909

根据表 3-1 的统计结果可以观测到诱捕器环境下的昆虫数据集中各类别的昆虫样本分布并不均衡。“Linnaeus”、“acuminatus”和“linnaeus”三类目标的样本数较少。因此为了使每类样本数量分布更为均衡，对包含这三类目标的图片进行数据增强操作，对此采用的增强方式是将数目较少的类别目标区域复制，后经过放大、旋转、平移等方式粘贴回原图，达到精准类别增强。以“Linnaeus”类为例进行数据增强，增强前后对比如图 3-6 所示。



图 3-6 精准数据增强前后效果

Fig.3-6 The effect of data before and after accurate enhancement

对诱捕器环境下的昆虫数据集进行精准类别增强处理后每类目标样本数目如表 3-2 所示。

表 3-2 精准数据增强后诱捕器环境下的昆虫数据集中每类目标样本统计结果

Table 3-2 The statistical results of each type of target sample in the insect data set in the trap environment
after the enhancement of the accurate data

昆虫目标	Boerner	Leconte	Linnaeus	acuminatus	armandi	coleoptera	linnaeus
样本数量	1595	2216	1480	1604	1765	2091	1598

复杂环境下的昆虫数据集统计结果如表 3-3 所示。

表 3-3 复杂环境下的昆虫数据集中每类目标样本统计结果

Table 3-3 The statistical results of each type of target sample in the insect data set in a complex environment

昆虫目标	样本数量
yousha	400
baie	430
tianniu	410
zhuhuang	530
songcai	400

根据表 3-3 的统计结果可以观测到复杂环境下的昆虫数据集中各类别的昆虫样本分布较为均衡，共有“yousha”、“baie”、“tianniu”、“zhuhuang”、“songcai”五类目标样本，数量均在 450 左右。因此不对该数据集进行精准类别增强操作。

虽然对两种场景下的数据集均进行了样本分布情况统计，实现了使样本均衡分布的目的。但是深度学习以数据为驱动。在深度学习中，充足的样本数量才可以提升模型的泛化能力。原始的两种数据集分别包含 1973 和 2000 幅图像，样本数量不足，因此本次研究通过数据旋转、数据翻转、图像缩放、图像剪裁、调整亮暗、对比度等多种数据增广操作来进行图像预处理工作，最终达到扩充数据集、提高样本质量最终使图片特征更容易提取的目的。

对诱捕器环境下图像进行顺时针旋转 180 度操作前后对比如图 3-7 所示。



(a) 图像进行顺时针旋转操作前

(b) 图像进行顺时针旋转操作后

图 3-7 对诱捕器环境下图像进行数据旋转操作前后效果

Fig.3-7 Before and after data rotation operation on the image in the trap environment

对复杂环境下图像进行顺时针旋转 180 度操作前后对比如图 3-8 所示。



(a) 图像进行顺时针旋转操作前



(b) 图像进行顺时针旋转操作后

图 3-8 对复杂环境下图像进行数据旋转操作前后效果

Fig.3-8 Before and after data rotation operations on images in complex environments

对诱捕器环境下图像进行类似于镜面翻折数据的翻转操作前后对比如图 3-9 所示。



(a) 图像进行数据翻转操作前



(b) 图像进行数据翻转操作后

图 3-9 对诱捕器环境下图像进行数据翻转操作前后效果

Fig.3-9 Before and after data flipping operation on the image in the trap environment

对复杂环境下图像进行类似于镜面翻折数据的翻转操作前后对比如图 3-10 所示。



(a) 图像进行数据翻转操作前



(b) 图像进行数据翻转操作后

图 3-10 对复杂环境下图像进行数据翻转操作前后效果

Fig.3-10 Before and after data flipping operation on the image in complex environments

对诱捕器环境下图像进行随机裁剪操作前后对比如图 3-11 所示。随机从图像中选择一部分，将这部分图像裁剪出来，之后调整为原图像的大小。



(a) 图像进行裁剪操作前 (b) 图像进行裁剪操作后

图 3-11 对诱捕器环境下图像进行裁剪操作前后效果

Fig.3-11 The effect of data before and after the image is cropped in the trap environment
对复杂环境下图像进行随机裁剪操作前后对比如图 3-12 所示。



(a) 图像进行裁剪操作前 (b) 图像进行裁剪操作后

图 3-12 对复杂环境下图像进行裁剪操作前后效果

Fig.3-12 The effect of data before and after the image is cropped in complex environments
对诱捕器环境下图像进行亮暗调整操作前后对比如图 3-13 所示。



(a) 图像进行亮度操作前 (b) 图像进行亮度调整操作后

图 3-13 对诱捕器环境下图像进行亮暗调整操作前后效果

Fig.3-13 Before and after the effect of adjusting the brightness and darkness of the image
in the trap environment

对诱捕器环境下图像进行对比度调整操作前后对比如图 3-14 所示。

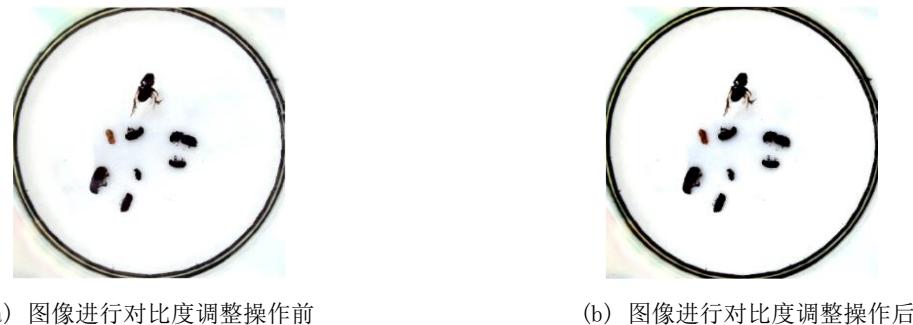


图 3-14 图像进行对比度调整操作前后效果

Fig.3-14 The effect of data before and after image contrast adjustment operation in the trap environment

3.3 林业害虫候选区域参数聚类

YOLOv4 模型检测的基础是在输入图片上生成一系列 anchor box, anchor box 是 w 和 h 固定的初始候选框, 是生成预测框的前提, anchor 的选用会直接影响模型的性能^[72]。YOLO v4 模型使用的 anchor boxes 是在 coco 数据集上统计出来的比较合适的锚框大小, 部分 anchor 过大, 并不适用于我们的昆虫检测任务。因此, 本文通过对林业昆虫数据集聚类得到合适的 anchor 大小。在聚类方法上, 本文使用 K-means++代替 YOLO v4 模型中使用的 K-means 算法。K-means 算法流程图如图 3-15 所示。

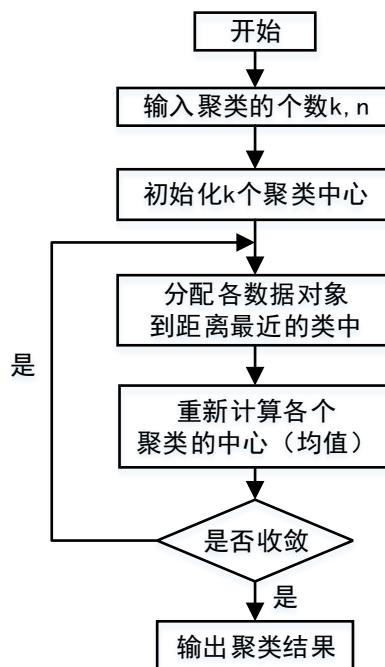


图 3-15 K-means 算法流程图

Fig.3-15 Algorithm flow chart for K-means

K-means 算法原理、易实现，但在确定初始聚类中心时需要人为设定，而不同的初始聚类中心对聚类结果的影响较大。因此，本次研究使用 K-means++ 对林业昆虫数据集中昆虫对象的 w 和 h 进行聚类。kmeans++ 在对初始聚类中心的选择上进行了优化，具体实现步骤为：首先随机选择一个点作为第一个初始类簇中心点，然后选择距离该点最远的那个点作为第二个初始类簇中心点，最后再选择距离前两个点距离最大的点作为第三个初始类簇的中心点，以此类推，直至选出 k 个初始类簇中心点^[73]。初始点选出后，继续使用标准的 K-means 算法进行聚类。

聚类评估标准常用轮廓系数来表示，计算公式如式 3-3 所示，轮廓系数取值范围为 [-1,1]，取值越接近 1 则说明分类效果越好：

$$s_i = \frac{b_i - a_i}{\max(b_i, a_i)} \quad (3-3)$$

式中：对于每个点 i 为已聚类数据中的样本， b_i 为样本 i 到其他各个族群的所有样本的距离平均值中取的最小值， a_i 为样本 i 到本身簇的距离平均值。

在诱捕器环境下的昆虫数据集上进行聚类与轮廓系数的计算，可以得到聚类类别数与轮廓系数对应关系如图 3-16 所示。

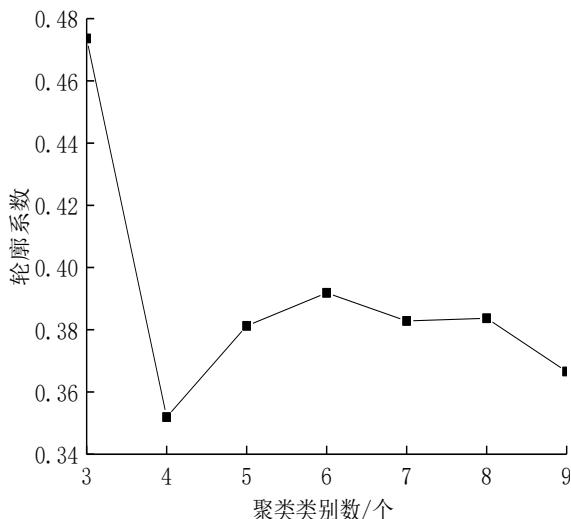


图 3-16 不同聚类类别数下的轮廓系数对比

Fig.3-16 Comparison of contour coefficients under different numbers of clustering categories

图中，类别数的取值范围为 3~9，经分析可得，以 k 表示聚类类别数， k 为 3 时，轮廓系数最高，聚类效果最好； k 为 6 时，聚类效果次之； k 为 7 时，聚类效果比较好。本文将三者相结合，最终得到候选区域参数分别为 (51, 52)、(68, 69)、(63, 91)、

(88, 64)、(92, 137)、(99, 140)、(136, 91)、(140, 100)、(146, 144)。同理，可以得到复杂环境下的昆虫数据集上候选区域参数分别为(27, 43)、(48, 54)、(53, 68)、(76, 59)、(84, 113)、(91, 127)、(114, 89)、(126, 108)、(131, 139)。

3.4 实验与数据分析

使用调整后的数据集作为训练集分别对改进前后的模型进行训练，经数据精准增强与数据增广后两种数据集中分别有10758和12000幅图像。从数据集中随机选取80%作为训练集，其余20%作为测试集。本次实验操作系统为windows server 2012 R2,CPU为Intel(R) Xeon(R) Gold 6145 CPU @ 2.00GHz,内存128GB, GPU为英伟达T4,显存15G。训练时batchsize设置为16,学习率为0.001,优化器为Adam。完成模型训练后，将SSD、Faster-RCNN、YOLOv4、YOLOv3作为对比模型，以诱捕器环境下的昆虫数据集为例，在测试集上采用以下评定指标分别对4个模型进行综合性能测试。

Precision: 精确率，通过计算正确检测目标数与整体检测到的目标数量的比例，衡量模型的分类能力，计算公式如式3-4所示；

$$precision = \frac{\text{正样本预测为真}}{(\text{正样本预测为真} + \text{负样本预测为真})} = \frac{TP}{(TP+FP)} \quad (3-4)$$

Recall: 召回率，通过计算正确检测目标数与验证集目标总数的比例，衡量模型的查全能力，计算公式如式3-5所示；

$$recall = \frac{\text{正样本预测为真}}{(\text{正样本预测为真} + \text{负样本预测为假})} = \frac{TP}{(TP+FN)} \quad (3-5)$$

IoU: 面积交并比，通过计算模型预测目标的矩形区域与验证集中目标标定的矩形区域的面积交并比，衡量模型的位置预测能力。

在诱捕器环境下的昆虫数据集上，YOLOv3模型的P-R曲线如图3-17所示。

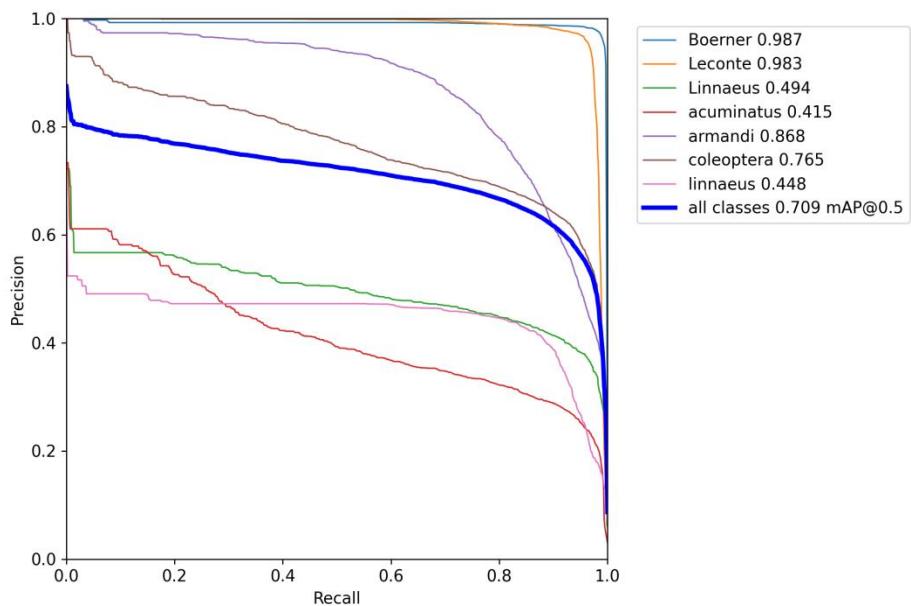


图 3-17 YOLOv3 模型的 Precision-Recall 变化曲线图（诱捕器数据集）

Figure 3-17 Precision-Recall change curve of YOLOv3 model in Trapper dataset

在诱捕器环境下的昆虫数据集上，YOLOv4 模型的 P-R 曲线如图 3-18 所示。

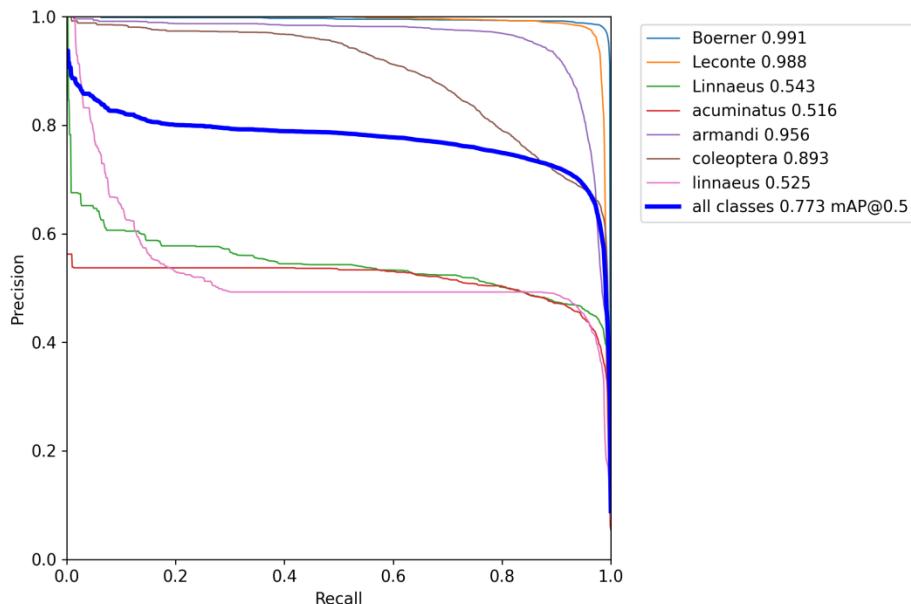


图 3-18 YOLOv4 模型的 Precision-Recall 变化曲线图（诱捕器数据集）

Figure 3-18 Precision-Recall change curve of YOLOv4 model in Trapper dataset

在诱捕器环境下的昆虫数据集上，YOLOv4-TIA 模型的 P-R 曲线如图 3-19 所示。

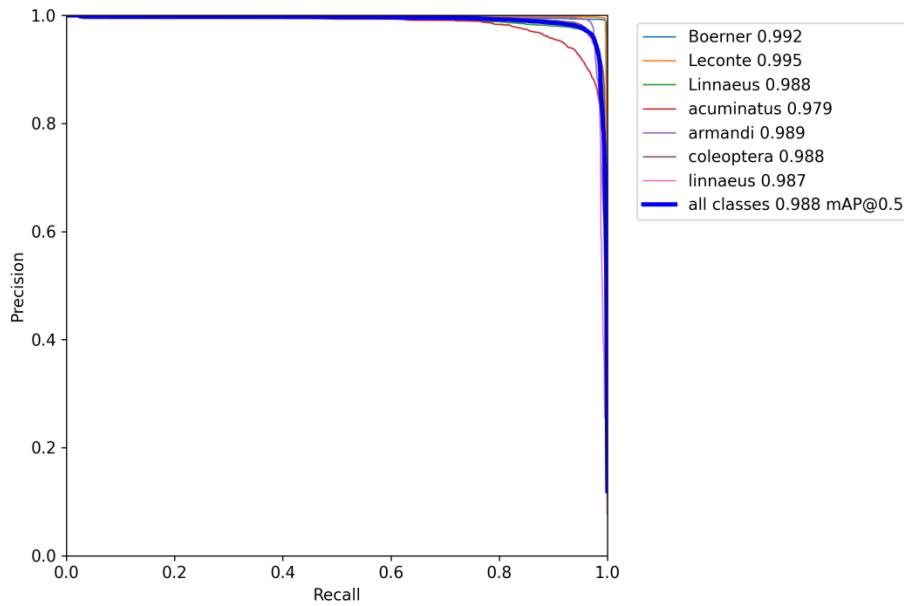


图 3-19 YOLOv4-TIA 模型的 Precision-Recall 变化曲线图（诱捕器数据集）

Figure 3-19 Precision-Recall variation curve of YOLOv4-TIA model in Trapper dataset

从三种模型的 P-R 曲线图对比可以观察到，YOLOv4-TIA 模型的效果最好，且 mAP 值最高，详细的对比结果如表 3-4 所示。

表 3-4 5 种模型进行林业害虫检测的综合性能对比(诱捕器下昆虫数据集)

Table 3-4 Comparison of comprehensive performance of five models for forestry pest detection in Trapper dataset

检测方法	Precision/%	Recall/%	FPS	mAP/%
SSD	78.3	81.5	19	74.3
Faster-RCNN	80.6	82.3	9	73.2
YOLOv3	73.4.	82.8	41	70.9
YOLOv4	91.5	93.1	64	77.3
YOLOv4-TIA	94.4	96.7	75	98.8

分析上表数据可得，改进的 YOLO v4-TIA 模型检测速度与 YOLOv4 相差不大，比 SSD 快 4 倍，比 Faster-RCNN 快 8 倍，比 YOLOv3 快 1.8 倍。准确率比其他模型高，且 mAP 值最高，达到了 98.8%。可见 YOLOv4-TIA 通过捕捉跨维度交互增强了网络特征表达能力，通过 Focal loss 进一步增强了模型的检测精确率；同时关注空间信息的检测器能够检测到更小的目标，YOLOv4-TIA 的召回率相比于 YOLOv4 提升了 7.6%。

在 CSPDarknet53 骨干网络相同位置上添加 SE、CBAM、Triplet Attention 三种不同的注意力机制时，模型参数增加量如表 3-5 所示。

表 3-5 CSPDarknet53 添加不同注意力模块参数增加量对比

Table 3-5 Comparisons of various attention modules based on parameter increased using CSPDarknet53

Method	Parameters increased
SE	2.303M
CBAM	2.318M
Triplet Attention	0.0012M

综合分析表 3-4 与表 3-5 数据可得，结合 Triplet Attention 结构改进的 YOLOv4-TIA 网络仅需增加很小的计算开销便可以给模型带来显著的性能收益。在测试集上对昆虫图像的实际检测效果对比如图 3-20 和图 3-21 所示，图中所有昆虫对象检测结果用彩色矩形框标示，图 3-20 为使用 YOLOv4 算法的检测结果，图 3-21 为使用 YOLOv4-TIA 算法的检测结果。



(a) 测试图像 1



(b) 测试图像 2

图 3-20 YOLOv4 算法检测结果

Fig.3-20 YOLOv4 algorithm test results



(a) 测试图像 1



(b) 测试图像 2

图 3-21 YOLOv4-TIA 算法检测结果

Fig.3-21 YOLOv4-TIA algorithm test results

可以观察到，直接使用 YOLOv4 算法进行检测存在一些昆虫对象漏检的情况，如图 3-20 中(a)图和(b)图分别漏检三只、四只昆虫，图 3-21 中(a)图和(b)图均无漏检情况。从

上图两组对比结果，可以直观的看出 YOLOv4-TIA 对于目标的检测精确率和召回率都要高于 YOLOv4，对于小目标的检测能力也更为优秀。

3.5 本章小结

本章从目标检测网络、数据集、和候选区域参数优化三个方面对林业害虫检测模型进行了优化，在 YOLOv4 算法方面，从骨干网络、检测网络中的特征融合方式以及损失函数三个方面对算法进行改进；在数据集方面，本文通过多种数据增强方式对图像样本集进行了扩充，使图像样本数达到了 22758，并且通过精准数据增强实现了样本均衡分布；最后通过对数据集进行 anchor 调整，得到适合进行林业害虫检测的参数，以提升改进后模型对林业害虫目标的检测识别率和检测速度。

最后在经数据处理后的林业害虫数据集上，通过实验对 SSD、Faster-RCNN、YOLOv4、YOLOv3 和 YOLOv4-TIA 这五种模型进行综合性能对比，最终验证了该模型对林业害虫检测任务的适用性。

第4章 林业害虫监测系统的设计与实现

林业害虫的有效治理是林木健康生长的关键。为了能通过对林区害虫的有效识别，实现为林区测报人员提供虫情及时预警，降低虫害爆发率，并为针对性防治提供理论依据的目标。本章在构建的两种不同场景林业害虫数据集和 YOLOv4-TIA 害虫检测网络的基础上，设计并实现了林业害虫监测系统。将昆虫图像信息录入、林业害虫图像读取、害虫检测与计数、虫情分析与虫害预警功能集成为一体，来满足林区工作人员对林业害虫进行监测的需求。

4.1 林业害虫监测系统的功能需求分析

需求分析是系统开发的出发点，是整个开发过程中的核心部分，详细的需求分析可以为系统设计起到指导性作用，为实现对林业害虫虫情的智慧监测，本文针对目前多数林区所面临的问题进行分析总结，并在此基础上对系统功能需求进行分析。

目前多数林区在进行林业害虫监测时仍以人工测报和诱捕器测报为主，多数面临以下问题：

(1) 数据管理问题。人工测报需人为进行照片采集，且信息记录方式仍以纸质和电子版为主；而诱捕器摄像头所采集照片往往被保存在独立的文件夹中，且图片信息仍需单独记录，数据之间的互动性和联系不足，不利于进行高效数据管理。基于此确定图像信息录入功能，使用数据库对图像及其录入信息进行管理，同时实现信息化目标。

(2) 数据读取问题。由于在系统中使用深度学习的方法对林业害虫图像进行检测，图像需从数据库或文件夹进行读取，基于此确定图像读取功能。

(3) 检测结果管理问题。传统的人工检测识别所采用的记录方式以纸质手册为主，不利于进行统计分析，且不够直观，基于此确定检测、计数与预警功能，实现对检测结果的可视化，并且通过计算机来对检测结果中各类别的害虫数目及占比进行统计。

(4) 检测结果分析问题。为了从历史检测结果中挖掘更有价值的信息，以可视化的方式更直观的观察虫情趋势，基于此确定虫情分析功能。

(5) 系统管理问题。在林区管理中，林区管理员与测报员的职责与角色存在差异，

基于此确定用户角色和管理功能，根据不同角色对访问系统人员分配相应权限的账户。

通过对系统功能需求进行分析明确了系统主要功能如下：

(1) 图像信息录入功能

本文设计的林业害虫监测系统中的图像信息录入功能可以将昆虫图片与图片相关信息添加并保存到数据库。

(2) 林业害虫图像读取功能

根据林区管理人员或者测报员的需求导入需要进行检测识别的昆虫图片。

(3) 害虫检测与计数功能

该功能通过第三章中论述的林业害虫检测算法对待检测图片中的害虫进行检测识别，并对检测结果中各类别的害虫数目进行计数与统计分析，并对识别与统计结果进行展示。

(4) 虫情分析功能

根据月份对不同昆虫类别的数量变化情况进行分析展示，可以观察虫情变化趋势。

(5) 虫害预警功能

根据检测与计数功能中得到的统计结果，提出预警，为林区管理人员判断虫害状况以及治理策略的选取提供科学依据。

(6) 用户管理功能

本文设计的林业害虫监测系统主要面向林区管理员和测报员，因此将用户按照类型划分为管理员和林区测报员，林区管理员对应的权限为能够使用林业害虫监测系统的所有基本功能，包括昆虫图像信息录入、读取昆虫图像、害虫检测与计数、虫情分析和虫害预警功能。测报人员可以对昆虫图像信息进行录入并查看虫情分析结果。

4.2 林业害虫监测系统设计

4.2.1 系统总体设计

本节在 4.1 节中阐述的林区管理人员和测报员对监测系统的需求分析和第三章中论述的改进后的适用于林业害虫检测的 YOLOv4-TIA 模型的基础上，设计、实现了一个林业害虫监测系统。系统主要包含六个功能模块，分别为：用户管理模块、昆虫图像信息

录入模块、害虫图像读取模块、害虫检测与统计模块、虫害预警模块、虫情分析模块。系统功能结构图如图 4-1 所示。

其中图像读取之后的害虫检测与统计将在高性能的云服务器中完成，并将检测与统计结果存储到数据库中。害虫监测系统的交互界面将被布置在本地 Windows 系统中，它会通过读取数据库中保存的检测结果数据来获得统计分析结果。

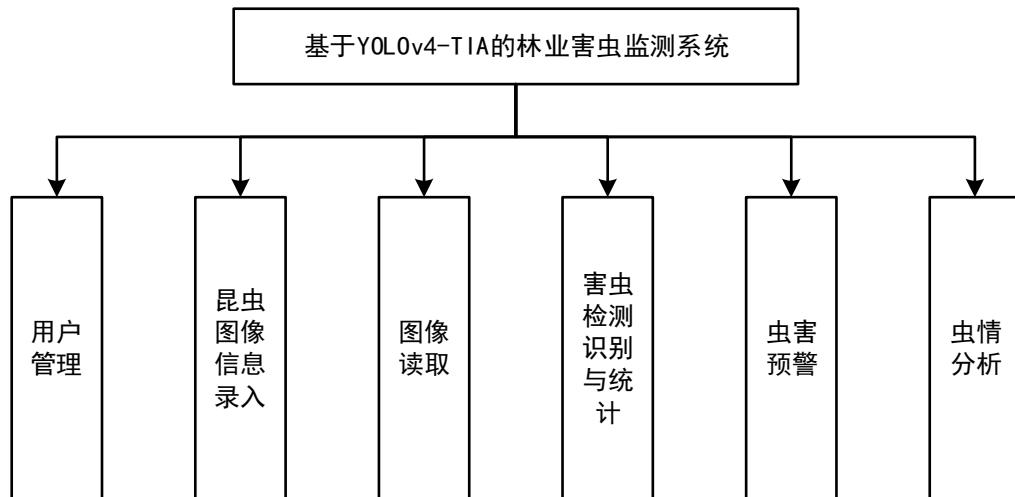


图 4-1 系统功能结构图

Fig.4-1 System function structure diagram

系统软件流程图如图 4-2 所示。用户经由本地电脑，选取用户身份类型，输入账号密码登录林业害虫监测系统。如果是林区管理员角色，可以对林区测报员的账号进行增加、删除和修改操作。如果是测报员的角色，他可以将需要保存的昆虫图像及其相关信息保存到创建的昆虫数据库中，也可以从本地或者数据库中读取昆虫图像，并通过相关功能模块对图像进行检测处理和统计分析，并从分析结果得知虫情预警信息，从而判断虫害发生的情况，并且可以根据月份和害虫类别对历史检测结果进行对比分析，以更直观的方式观察虫情趋势，判断林区状况并做出合理的防治策略。

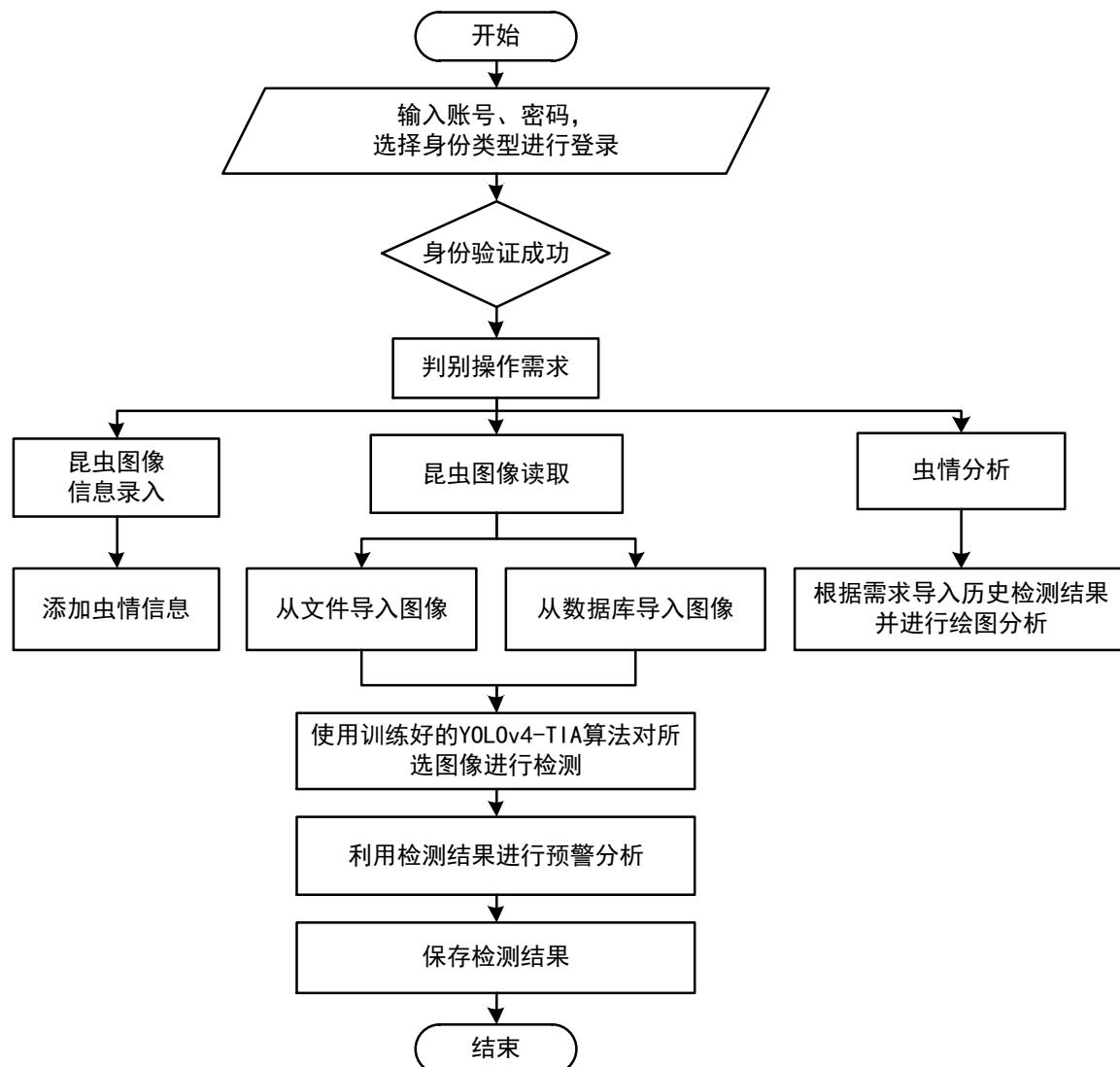


图 4-2 系统软件流程图

Fig.4-2 System software flow chart

4.2.2 系统模块功能设计

由系统总体设计来看,基于深度学习模型 YOLOv4-TIA 的林业害虫监测系统的两大核心功能模块为: 图像信息录入模块和害虫检测与统计预警模块, 本节将详细介绍这两部分内容。

(1)图像信息录入模块

害虫信息采集工作由测报员登录系统后, 点击图像信息录入功能按钮, 可以在弹出的界面上上传需要录入的图片, 同时可以通过填写相关采集信息对虫情信息进行记录。虫

情信息采集的流程如图 4-3 所示。

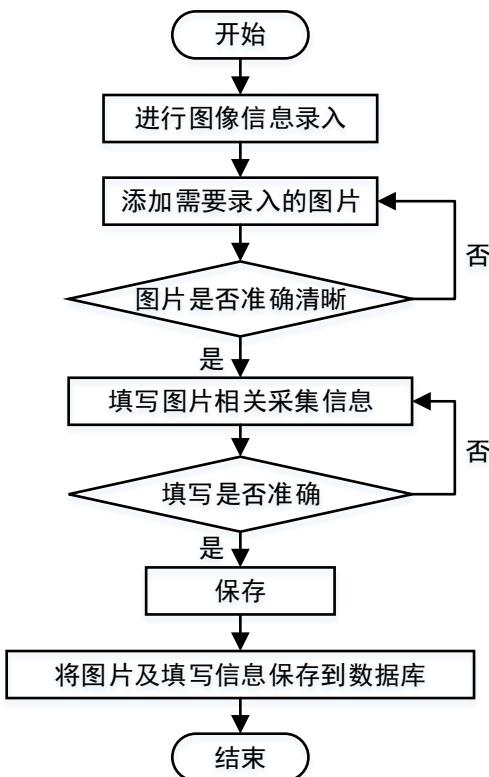


图 4-3 虫情信息采集流程图

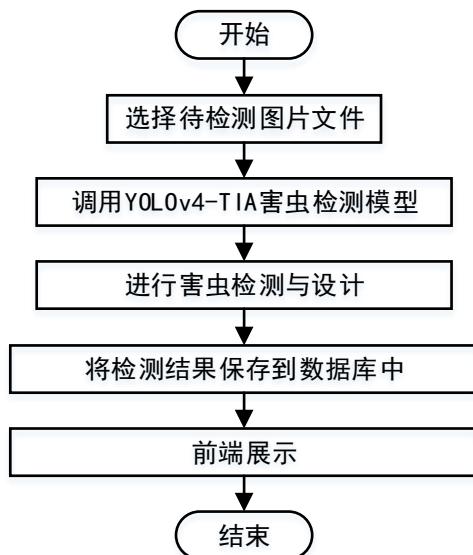


图 4-4 害虫检测与统计流程

Fig.4-3 Insect situation information collection flow chart

Fig.4-4 Pest detection and statistics flow chart

(2) 害虫检测与统计功能模块

测报员可以通过单张读取或批量读取图片功能从本地或者昆虫数据库中读取不同类型的昆虫图片，并调用 YOLOv4-TIA 害虫检测模型对图像中的害虫进行检测，完成所有待检测图片的检测识别后，检测与统计结果会保存到数据库中，并且可以在前端进行展示。害虫检测与统计的流程如图 4-4 所示。

4.2.3 数据库表设计

本系统所有的表都放在 insect 数据库下，共包含 8 张表，分别为 Administrator 表、Tester 表、Insect 表、MonitoringStation 表、Insectlib 表、InsectInformation 表、InsectPicture 表、TestInformation 表。

Administrator 表如表 4-1 所示，存储林区管理员的用户名和密码。

表 4-1 管理员信息表 (Administrator 表)

Table 4-1 Administrator information table (Administrator table)

字段名	字段描述	数据类型	主键	外键	非空	唯一	默认值	自增
id	管理员 ID	INTEGER	是	否	是	是	无	是
username	用户名	VARCHAR (20)	否	否	是	是	无	否
password	密码	VARCHAR (20)	否	否	是	否	无	否
type	类型	VARCHAR	否	否	否	否	无	否

Tester 表如表 4-2 所示，存储测报员的用户名、密码及账户状态。

表 4-2 测报员信息表(Tester 表)

Table 4-2 Tester information table (Tester table)

字段名	字段描述	数据类型	主键	外键	非空	唯一	默认值	自增
testerid	测报员 ID	INTEGER	是	否	是	是	无	是
testername	测报员姓名	VARCHAR (20)	否	否	是	否	无	否

Insect 表如表 4-3 所示，用于存储昆虫的编号和名称信息。

表 4-3 昆虫信息表(Insect 表)

Table 4-3 Insect Information Sheet (Insect Sheet)

字段名	字段描述	数据类型	主键	外键	非空	唯一	默认值	自增
InId	昆虫号	INTEGER	是	否	是	是	无	是
Inname	昆虫名称	VARCHAR (20)	否	否	是	否	无	否

MonitoringStation 表如表 4-4 所示，用于存储监测站的站号和名称信息。

表 4-4 监测站信息表(MonitoringStation 表)

Table 4-4 Monitoring Station Information Table (MonitoringStation Table)

字段名	字段描述	数据类型	主键	外键	非空	唯一	默认值	自增
msno	监测站号	VARCHAR (20)	是	否	是	是	无	否
msname	监测站名称	VARCHAR (20)	否	否	是	是	无	否

Insectlib 表如表 4-5 所示，用于存储昆虫所属昆虫库的信息，每个昆虫库都必须有昆虫库号、昆虫库名称以及库类型。

表 4-5 昆虫库表(Insectlib 表)

Table 4-5 Insect library table (Insectlib table)

字段名	字段描述	数据类型	主键	外键	非空	唯一	默认值	自增
libno	昆虫库号	INTEGER	是	否	是	是	无	否
Libname	昆虫库名称	VARCHAR(20)	否	否	是	是	无	否
createtime	Datetime	DATE	否	否	否	否	无	否
creator	库创建者	VARCHAR(255)	否	否	否	否	无	否
memo	库备注信息	TEXT	否	否	否	否	无	否
type	库类型	VARCHAR(255)	否	否	是	否	GENERAL	否

InsectPicture 表如表 4-6 所示，用于存储需要记录的昆虫图片信息，以及对应填写的信息表编号。

表 4-6 昆虫图片信息表(InsectPicture 表)

Table 4-6 Insect picture information table (InsectPicture table)

字段名	字段描述	数据类型	主键	外键	非空	唯一	默认值	自增
picid	图片 ID	INTEGER	是	否	是	是	无	是
filerpath	文件读取路径	VARCHAR(255)	否	否	否	否	无	否
InfoId	信息表编号	INT(8)	否	是	是	否	无	否

InsectInformation 表如表 4-7 所示，用于存储需要记录的昆虫相关信息，这个表中需要记录监测站编号、测报员的编号、昆虫库编号以及信息的采集时间。

表 4-7 昆虫信息采集表(InsectInformation 表)

Table 4-7 Insect Information Collection Form (InsectInformation Form)

字段名	字段描述	数据类型	主键	外键	非空	唯一	默认值	自增
InfoId	信息表编号	INTEGER	是	否	是	是	无	是
msno	监测站号	INT(4)	否	是	是	否	无	否
testerid	测报员 ID	INT(8)	否	是	是	否	无	否
libno	昆虫库号	INT(4)	否	是	是	否	无	否
collect_time	信息采集时间	DATE	否	否	否	否	无	否

TestInformation 表如表 4-8 所示，用于存储昆虫图片的检测结果信息，主要包含结果路径所在信息，以及以备注形式保存的各类别昆虫数目以及统计信息。

表 4-8 检测结果信息表(TestInformation 表)

Table 4-8 Test result information table (TestInformation table)

字段名	字段描述	数据类型	主键	外键	非空	唯一	默认值	自增
id	编号	INTEGER	是	否	是	是	无	是
libno	昆虫库号	INTEGER	否	是	是	否	无	否
InId	昆虫号	INTEGER	否	是	是	否	无	否
filepath	文件结果路径	VARCHAR(255)	否	否	否	否	无	否
num	昆虫数量	INTEGER	否	否	否	否	无	否

Insect 数据库中 8 张表之间的相互联系图如图 4-5 所示。

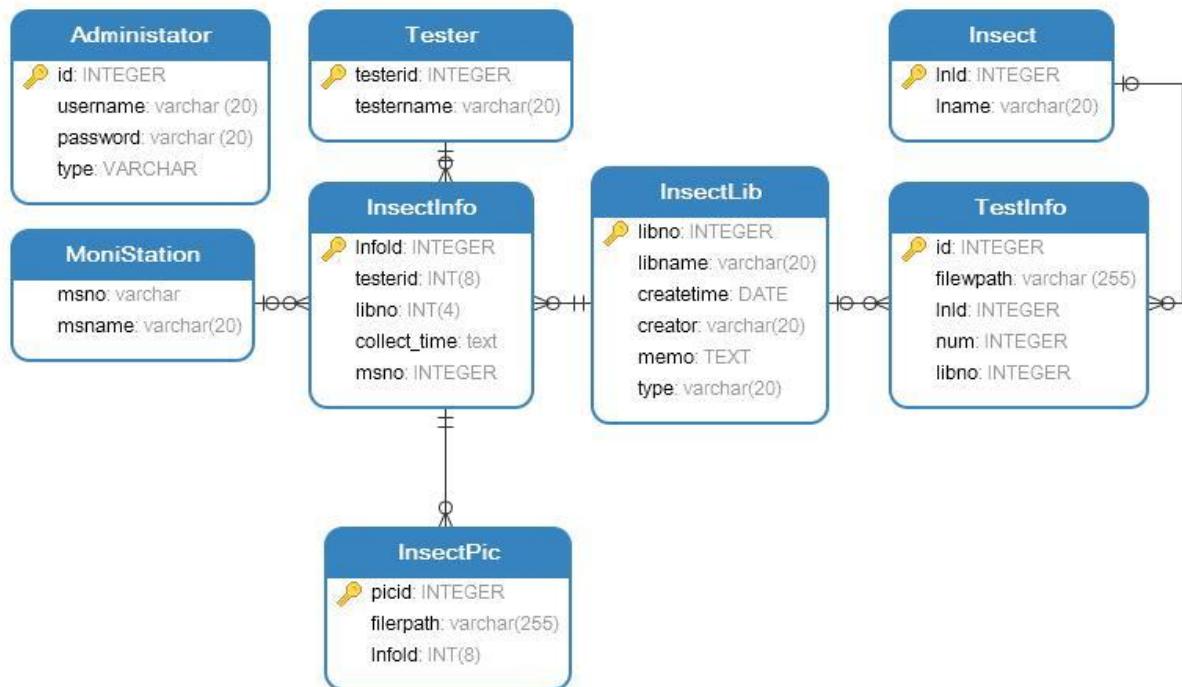


图 4-5 Insect 数据库中 8 张表之间的相互联系图

Fig.4-5 the correlation diagram between the eight tables in the Insect database

4.3 林业害虫监测系统实现

林业害虫监测系统其主要功能包括用户登录、虫情信息录入、从文件夹或数据中读取图像、害虫图像检测结果的显示、统计与预警、虫情分析等。

系统登录界面如图 4-6 所示，林区管理员或者测报员在输入正确的用户名、密码，选择用户类型后即可登入系统主功能界面。



图 4-6 系统登陆界面

Fig.4-6 System login interface

林业害虫监测系统主功能界面如图 4-7 所示，通过登录界面登录后，会进入系统主功能界面，包含图像信息录入、害虫图像读取、害虫检测与统计预警和虫情分析四个功能。



图 4-7 系统主功能界面

Fig.4-7 System main function interface

图像信息录入界面如图 4-8 所示，主要包含图片信息的录入以及图像采集相关说明信息的录入，包含虫情监测站信息、测报员信息、图片上传库信息以及采集时间。



图 4-8 害虫图像信息录入界面

Fig.4-8 Pest image information input interface

害虫图像读取界面主要提供图像单张导入或批量导入的功能，如图 4-9 所示。



图 4-9 害虫图像读取界面

Fig.4-9 Pest image reading interface

点击导入数据后，可以从文件夹选择待检测的图片，界面如图 4-10 所示。

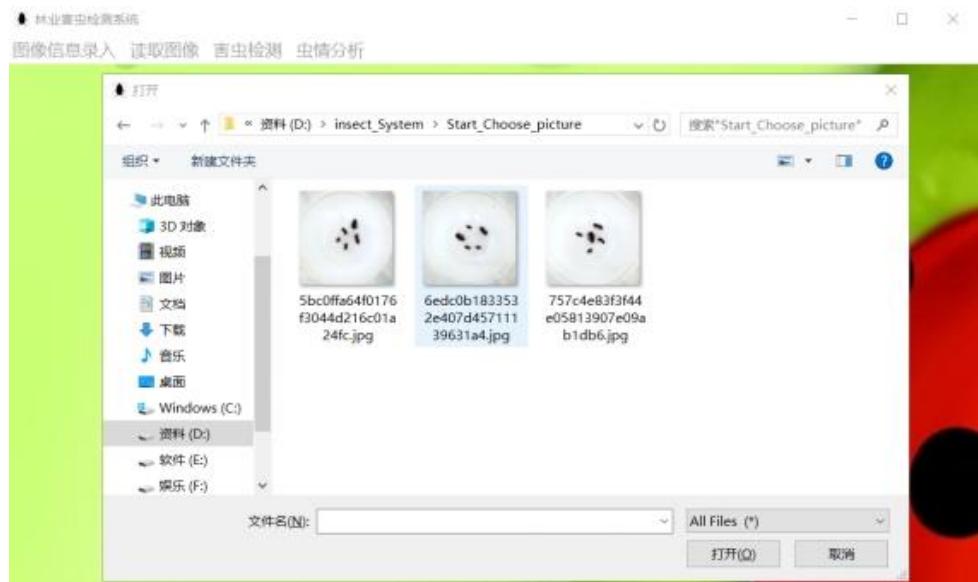


图 4-10 导入数据展示界面

Fig.4-10 Import data display interface

将图像导入后，其展示界面如图 4-11 所示。



图 4-11 图像导入展示界面

Fig.4-11 Image import display interface

害虫检测与统计界面对检测前后的图像进行展示，检测后的图像中会使用不同颜色的矩形框标注出检测到的害虫，并给出所检测图片中包含的害虫各类别数目的统计结果，如图 4-12 所示。



图 4-12 害虫检测识别结果与统计结果界面

Fig.4-12 Pest detection and identification results and statistical results interface

虫情分析界面可以按照月份对各害虫类别的数目统计量进行对比分析，帮助林区管理员、测报员根据虫情趋势及时察觉有异常的虫情信息。其界面如图 4-13 所示。

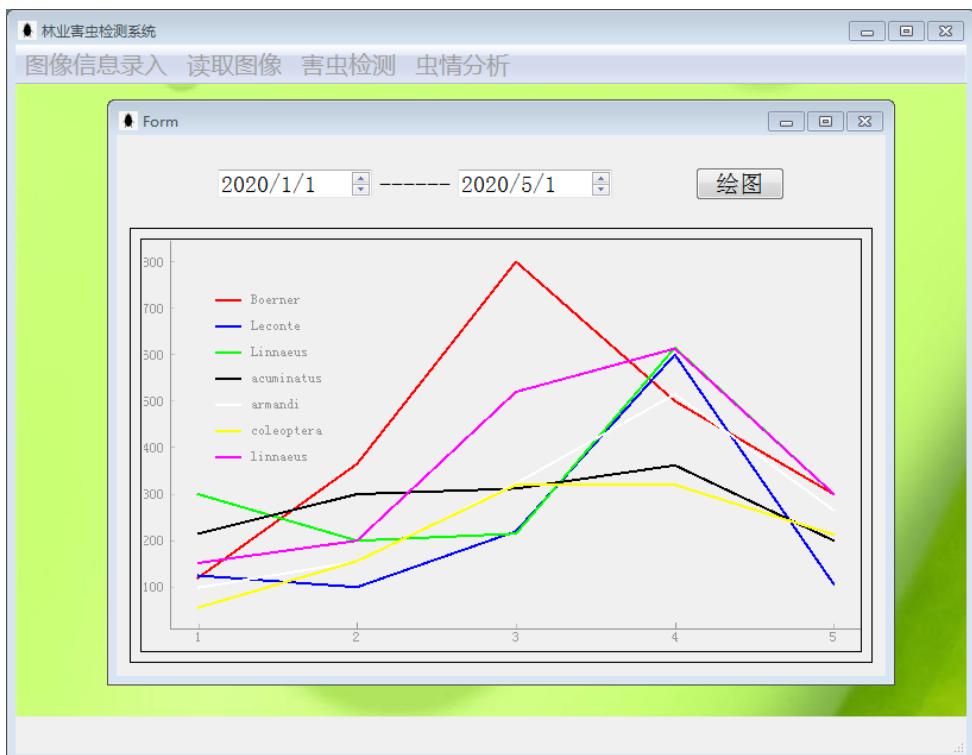


图 4-13 虫情分析结果界面

Fig.4-13 Insect situation analysis result interface

害虫检测与统计结果和虫情分析结果经预警分析可以起到提前预警的效果。

4.4 林业害虫监测系统测试

系统测试是系统软件开发过程中的一个重要组成部分。对系统进行严谨、全面的测试，可以及时发现同需求分析时不相符的功能、性能指标以及运行时存在的问题并立即处理，确保系统的质量和可靠性。

4.4.1 平台测试环境说明

林业害虫监测系统运行在 windows7 上，CPU 为英特尔 Core i5-6300U @2.4GHz；运行内存：4GB；使用 Pycharm 集成开发环境进行开发，并安装 PYQT5、SQLite3 数据库、Python3.8 语言运行所需的环境以及深度学习所需的环境。

4.4.2 平台功能测试

林业害虫监测系统服务主体为林区管理员和测报员，主要功能是为使用者提供一种便捷的虫情监测方式，按照功能可以主要划分为图像信息录入、读取图像、害虫检测以及虫情分析功能。

图像信息录入功能，主要为测报员录入图像信息提供便利，允许用户在虫情信息录入界面填写所采集图片对应的监测站信息、负责人信息、采集时间、昆虫库编号信息以及选择需要上传的图片。该功能详细测试情况如表 4-9 所示。

表 4-9 图像信息录入功能测试

Table 4-9 Image information inputting function test

测试功能	操作步骤	预期目标	测试结果
图像信息录入	进入系统主功能界面，点击图像信息录入按钮，继续点击虫情信息录入按钮；	系统界面正常显示出虫情信息录入填写界面及界面中需要填写内容及操作按钮图标等页面组件信息；	通过
虫情信息录入	填写虫情信息录入界面信息，点击选择文件选择需上传的图片信息，点击保存；	点击选择文件按钮后显示出文件选择窗口；点击保存后弹出框，保存至目标文件夹后，点击确定显示保存成功；	通过

读取图像功能，供用户选取待检测图像，用户通过选择导入方式从文件夹或数据库中选择目标图像，并在图像展示位置进行展示。该功能详细测试情况如表 4-10 所示。

表 4-10 读取图像功能测试

Table 4-10 Reading image function test

测试功能	操作步骤	预期目标	测试结果
读取图像	在系统主功能界面，点击读取图像按钮；	系统界面正常显示出读取图像界面，正常显示出操作按钮等页面组件信息；	通过
文件导入数据	点击文件导入数据按钮；	点击文件导入数据按钮后弹出文件选择窗口，从目标文件中选择需检测的图片，界面图像显示位置正常显示图片，底部信息栏动态显示待检测图像的数量；	通过
数据库导入数据	点击数据库导入数据按钮；	点击文件导入数据按钮后显示出文件选择窗口，从目标文件中选择需检测的图片，界面图像显示位置正常显示图片，底部信息栏动态显示待检测图像的数量；	通过
进入检测	从文件夹或数据库选择待检测图片后，点击进入检测按钮；	进入害虫图像检测界面，系统界面正常显示出图像检测界面及操作按钮图标等页面组件信息；	通过

害虫检测功能，为用户查看检测结果和预警分析结果提供便利，主要负责检测结果的展示和各类别数目分析统计与预警中的百分比统计。该功能详细测试情况如表 4-11 所示。

表 4-11 害虫检测功能测试

Table 4-11 Pest detection function test

测试功能	操作步骤	预期目标	测试结果
检测	在害虫图像检测界面点击检测按钮；	在检测展示窗口正常显示检测结果图；在底部信息栏动态显示所检测图像中各类别昆虫数量的统计信息；	通过
预警分析	点击预警分析按钮；	在底部信息栏动态显示所检测图像中各类别昆虫数量的占比信息；	通过

虫情分析功能，为用户按月份对各类别昆虫数目通过绘图进行更直观的展示。该功能详细测试情况如表 4-12 所示。

表 4-12 虫情分析功能测试

Table 4-12 Insect situation analysis function test

测试功能	操作步骤	预期目标	测试结果
虫情分析	在系统主功能界面，点击虫情分析按钮；	系统界面正常显示出虫情分析界面及操作按钮等页面组件信息；	通过
绘图	选择需要进行虫情分析的时间范围，点击绘图按钮；	在虫情分析展示窗口正常显示分析结果图。	通过

4.4.3 系统性能测试

本系统应用服务部署在 ThinkPad 笔记本电脑上，对系统中图片检测速度进行测试。测试中，在 CPU 上，单张检测图片耗时约在 0.08 秒左右，满足林业害虫系统实时检测的需求。

4.5 本章小结

本章主要介绍了林业害虫监测系统的功能需求分析、系统总体设计、模块功能设计、数据库表设计、界面实现以及系统测试。首先确定了系统的功能需求，之后分析系统的总体设计架构，并对各功能模块进行设计，同时对 insect 数据库中的表进行了设计，并对主要功能进行图文展示，最终对系统进行了功能与性能测试。系统可以满足林区工作人员对常见林业害虫的识别与监测预警需求，为林区管理人员与测报员的工作提供了便捷性。

第 5 章 总结与展望

5.1 总结

林业害虫一直是威胁林木健康生长、造成虫害爆发的重要因素。对森林害虫实施精准检测识别与及时监测预警，对提高林业科学防治与研究水平极其重要，同时可以为制定宏观决策提供科学依据。

计算机视觉技术与目标检测技术的发展，促使基于图像的害虫检测技术不断进步，本文采用基于深度学习的目标检测算法对林业害虫图像进行检测识别，并在此基础上设计林业害虫监测系统，使得林业害虫监测预警工作不再繁冗，可以经由高效智能化的计算机分析来完成。论文的主要工作如下：

第一，针对所使用的北京林业大学害虫数据集为诱捕器下采集的害虫图片，背景较为单一，会导致模型对自然背景下的林业昆虫识别能力不足，本文通过网上查找并整理了 10 种共 2000 幅常见林业害虫样本集，均为自然环境下森林昆虫图像，并通过对原始样本集进行旋转、翻转、缩放、剪裁、亮暗、对比度变换等多种数据增广操作来扩充数据集、提高样本质量，最终提高网络检测精度，提升模型的泛化能力。

第二，研究了基于深度学习的目标检测相关技术，结合林业害虫监测的应用需求，在 YOLOv4 网络的基础上，提出了 YOLOv4-TIA 林业害虫检测网络。通过结合三分支注意力在不增加网络参数量的基础上跨维度提取注意力特征图，使网络可以更好地关注图像中的目标特征信息，抑制无用信息。通过改进 PANet 网络，以跳跃连接与高效多方向的跨尺度连接相结合的特征融合方式，来更好的平衡多尺度特征信息，并利用 Focal loss 来优化分类损失，最终提升模型对林业害虫检测的整体性能。将 YOLOv4-TIA 网络与 YOLOv4、YOLOv3、SSD、和 Faster RCNN 这四种模型进行了对比试验，结果表明：本文提出的 YOLOv4-TIA 网络在本次研究的林业害虫数据集上性能最佳，检测准确率高达 99.8%、检测单张图片耗时 0.79 秒，具有良好的林业应用前景。

第三，在 YOLOv4-TIA 林业害虫检测网络的基础上设计并实现了林业害虫监测系统。该系统实现了昆虫图像信息录入、害虫图像读取、害虫检测与统计、虫情分析和虫害预警等功能。最后，对系统功能进行测试，测试结果符合设计目标。

5.2 展望

首先，本文虽然考虑了复杂环境下的林业害虫检测环境，但所制作的野外环境下的图像数据集主要包含一种昆虫类别，未来可以考虑向此环境下的数据集中补充多类别昆虫图像。

其次，本文在进行复杂环境下的害虫样本整理与标注工作时会耗费巨大的人力和时间成本。在未来的工作中，可以尝试采用更加先进的技术，如 GAN(生成对抗网络)等生成高质量的训练样本作为真实数据的有效补充，以新的视角解决样本集的需求问题。

最后，本文设计的林业害虫监测系统只能处理上传至服务器或者本地数据库的图像文件，但是在偏远地区，林区测报员往往在野外进行采集工作，所采集的图像与相关信息无法实时上传至本地数据库或服务器，这是非常不方便的。因此，在后续研究中可以进一步探索将移动端信息采集、实时检测分析与 PC 端相结合的监测方式，进一步提高监测系统的实用性，突出研究的价值与可行性。

参考文献

- [1] 左彦哲. 基于深度学习的害虫识别系统[D].北京林业大学,2018.
- [2] 李孝华. 基于卷积神经网络的蝴蝶识别系统设计与实现[D].西南大学,2020.
- [3] 马彬,金志明,蒋旭初,等. 储粮害虫在线监测技术的研究进展 [J]. 粮食储藏,2018,47(02):27-31.
- [4] 符首夫. 基于卷积神经网络的农作物昆虫识别与计数[D].江苏科技大学,2020.
- [5] 潘春华,肖德琴,林探宇,等. 于 SVM 和区域生长结合算法的南方主要蔬菜害虫分类识别 (英文) [J].农业工程学报,2018,34(08):192-199.
- [6] 封洪强 , 姚青 . 农业害虫自动识别与监测技术 [J]. 植物保护,2018,44(05):127-133+198.
- [7] R.J. White,L. Winokur. Quantitative description and discrimination of butterfly wing patterns using moment invariant analysis[J]. Bulletin of Entomological Research,2003,93(4):361-376.
- [8] Ziyi Liu,Junfeng Gao,Guoguo Yang, et al. Localization and Classification of Paddy Field Pests using a Saliency Map and Deep Convolutional Neural Network[J]. Scientific Reports,2016,6(1).
- [9] Schrder S , Drescher W , Steinhage V , et al. An automated method for the identification of bee species (Hymenoptera: Apoidea)[C]. Proc. Intern. Symp. on Conserving Europe's Bees. 1995:6-7.
- [10]P. J. D. Weeks,M. A. O'Neill,K. J, et al. Automating insect identification: exploring the limitations of a prototype system[J]. Journal of Applied Entomology,1999,123(1) :1-8.
- [11]Adam Tofilski. DrawWing, a program for numerical description of insect wings[J]. Journal of Insect Science,2004, 4(1):17.
- [12]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[J].中国农业大学学报,2003(03):47-50.

- [13]Zhao Juan,Cheng Xiao-Ping. Field pest identification by an improved Gabor texture segmentation scheme[J]. New Zealand Journal of Agricultural Research,2007,50(5):719-723.
- [14]Chenglu Wen,Daniel E. Guyer,Wei Li. Local feature-based identification and classification for orchard insects[J]. Biosystems Engineering,2009,104(3):299-307.
- [15]王志彬,王开义,张水发,等. 基于 K-means 聚类和椭圆拟合方法的白粉虱计数算法[J].农业工程学报.2014,30(01):105-112.
- [16]Chenglu Wen,Daniel Guyer. Image-based orchard insect automated identification and classification method[J]. Computers and Electronics in Agriculture,2012,89:110-115.
- [17]Ebrahimi M A , Khoshtaghaza M H , Minaei S , et al. Vision-based pest detection based on SVM classification method[J]. Computers and Electronics in Agriculture, 2017, 137(1):52–58.
- [18]林探宇,肖德琴,曾基业,等. 基于区域生长和SVM结合的黄金大蚊快速检测算法[J]. 广东农业科学,2016,43(07):172-177+4.
- [19]廉飞宇,张元. 基于小波变换压缩和支持向量机组的储粮害虫图像识别[J].河南工业大学学报(自然科学版),2006(01):21-24+28.
- [20]周曼,周明全. 基于BP神经网络的水稻害虫自动识别[J].北京师范大学学报(自然科学版),2008(02):165-167.
- [21]Naka,Kubota,Kuchiki,et al. Identification and field attraction of the female sex pheromone of a kiwifruit pest, Nokona feralis (Lepidoptera: Sesiidae)[J]. Bioscience, Biotechnology, and Biochemistry,2018,82(9).
- [22]Yuanhong Zhong,Junyuan Gao,Qilun Lei,et al. A Vision-Based Counting and Recognition System for Flying Insects in Intelligent Agriculture[J]. Sensors,2018,18(5): 1489-1508.
- [23]Ding W, Taylor G. Automatic moth detection from trap images for pest management [J]. Computers & Electronics in Agriculture, 2016,123(6):17-28.

- [24]程尚坤. 基于深度学习的储粮害虫检测方法研究[D].河南工业大学, 2017.
- [25]Yufeng Shen,Huiling Zhou,Jiangtao Li, et al. Detection of stored-grain insects using deep learning[J]. Computers and Electronics in Agriculture,2018,145:319-325.
- [26]程曦,吴云志,张友华.等. 基于深度卷积神经网络的储粮害虫图像识别[J].中国农学通报, 2018,34(1):154-158.
- [27]Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection[J].arXiv preprint arXiv:2004.10934, 2020.
- [28]Song H, Willi M, Thiagarajan J J, et al. Triplet Network with Attention for Speaker Diarization[J]. arXiv preprint arXiv:1808.01535, 2018.
- [29]G. E. Hinton,R. R. Salakhutdinov. Reducing the Dimensionality of Data with Neural Networks[J]. Science,2006,313(5786):504-507.
- [30]Bengio Yoshua,Courville Aaron,Vincent Pascal. Representation learning: a review and new perspectives[J]. IEEE transactions on pattern analysis and machine intelligence,2013,35(8):1798-1828.
- [31]Gulli A . Deep learning with Keras : implement neural networks with Keras on Theano and TensorFlow[J].2017.
- [32]Ketkar N . Introduction to PyTorch[J]. Deep Learning with Python, 2017:195-208.
- [33]Jia Y , Shelhamer E , Donahue J , et al. Caffe: Convolutional Architecture for Fast Feature Embedding[J].2014:675-678.
- [34]Abadi, Martin, Isard M, et al. Proceedings of the 1st ACM SIGPLAN International Workshop on Machine Learning and Programming Languages[C]. Acm Sigplan International Workshop on Machine Learning & Programming Languages. ACM, 2017:1-7.
- [35]周俊. 基于语义分割的人脸篡改检测研究[D].电子科技大学,2020.
- [36]唐洋. 基于深度学习的手语识别算法研究[D].江苏科技大学,2019.
- [37]Yu D, Deng L. Deep Learning and Its Applications to Signal and Information

Processing [Exploratory DSP][J]. IEEE Signal Processing Magazine, 2011, 28(1): 145-154.

[38] 邓宇. 基于深度学习的电路版图光刻热点检测技术[D]. 湖南大学, 2019.

[39] 钟德建. 基于深度学习模型迁移的鲁棒性识别关键问题研究[D]. 电子科技大学, 2019.

[40] 邵康. 基于深度学习的分布式遥感图像检索技术研究[D]. 华中科技大学, 2019.

[41] 邱隆庆. 基于特征学习的行人重识别研究[D]. 华南理工大学, 2017.

[42] 张洁, 赵红东, 李宇海, 等. 复杂背景下车型识别分类器[J]. 激光与光电子学进展, 2019, 56(04): 166-173.

[43] 刘入韶. 非接触成像下手纹识别方法研究[D]. 沈阳工业大学, 2019.

[44] 赵佳悦. 基于卷积神经网络的番茄病害检测研究[D]. 山东师范大学, 2020.

[45] 李茂鹏. 基于改进 SSD 的目标检测算法及剪枝优化研究[D]. 南京邮电大学, 2020.

[46] 陈康. 基于机器视觉的布匹瑕疵检测方法研究及软件开发[D]. 浙江工业大学, 2020.

[47] 裴嘉欣. 夜间无人车的周边目标信息感知研究[D]. 东华大学, 2019.

[48] Girshick R , Donahue J , Darrell T , et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[J]. IEEE Computer Society, 2013.

[49] 汪家亮. 基于 Kinect 的盲人室内智能辅助行走系统设计与开发[D]. 中国矿业大学, 2018.

[50] 胡鹏. 基于单目摄像头的密集人群跨线计数及其嵌入式系统的实现[D]. 重庆理工大学, 2020.

[51] Girshick R. Fast r-cnn[C]. Proceedings of the IEEE international conference on computer vision, 2015: 1440-1448.

[52] 刘永斌. 基于深度学习的 BGA 焊球空洞缺陷检测技术研究[D]. 沈阳建筑大学, 2020.

[53] 罗壮. 基于目标的图像检索关键技术研究及应用[D]. 浙江大学, 2018.

[54] 唐聪, 凌永顺, 郑科栋, 等. 基于深度学习的多视窗 SSD 目标检测方法[J]. 红外与激光工程, 2018, 47(01): 302-310.

- [55]冀树伟. 基于 SSD 的交通场景下多目标检测算法研究[D].中北大学,2019.
- [56]Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 779-788.
- [57]Redmon J , Farhadi A . YOLO9000: Better, Faster, Stronger[J]. IEEE, 2017:6517-6525.
- [58]Redmon J , Farhadi A . YOLOv3: An Incremental Improvement[J]. arXiv e-prints, 2018.
- [59]Bochkovskiy A , Wang C Y , Liao H . YOLOv4: Optimal Speed and Accuracy of Object Detection[J]. 2020.
- [60]袁浩期. 基于图像融合的流动人群体温检测系统设计[D].广东工业大学,2019.
- [61]李育胜. 一种基于 2D 图像驱动的稀疏点云分割方法[D].深圳大学,2019.
- [62]丁福光. 基于深度结构化学习的语义图像分割方法研究[D].浙江工业大学,2019.
- [63]吴倩. 基于视频分析的奶牛关键部位提取及跛行检测研究[D].西北农林科技大学,2019.
- [64]王亮,翟志强,朱忠祥,等.基于深度图像和神经网络的拖拉机识别与定位方法[J].农业机械学报,2020,51(S2):554-560.
- [65]邰旭鸣. 基于深度学习的多标签图像分类系统设计与实现[D].南京师范大学,2019.
- [66]Liu S, Qi L, Qin H, et al. Path Aggregation Network for Instance Segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE Press, 2018: 8759-8768.
- [67]苏婷婷. 基于迁移学习的农作物叶部病害图像识别的研究[D].山东农业大学,2020.
- [68]钱嘉伟. 基于双目异质机器视觉的检测方法及应用研究[D].浙江大学,2014.
- [69]Yu Zhenwei,Shen Yonggang,Shen Chenkai. A real-time detection approach for bridge cracks based on YOLOv4-FPM[J]. Automation in Construction,2021,122.
- [70]YANG Mingxin, ZHANG Yaoguang, LI Tao. Research on Small Sample Recognition

of Corn Disease Based on Convolutional Neural Network[J]. Chinese Journal of Eco-Agriculture, 2020,28(12):1924-1931.

[71] 刘天保. 基于卷积神经网络的人脸检测算法研究[J]. 软件导刊, 2020, 19(10):66-70.

[72] 鞠默然, 罗海波, 王仲博, 等. 改进的 YOLO V3 算法及其在小目标检测中的应用[J]. 光学学报, 2019, 39(07):253-260.

[73] 蔡云骐. 基于蓝牙信标的 k-means 指纹定位算法研究 [J]. 电子世界, 2017(02):177-178+181.

攻读硕士学位期间发表的论文及所取得的研究成果

(一) 发表的学术论文（已录用）

[1] 候瑞环,杨喜旺,王智超,高佳鑫. 一种基于 YOLOv4-TIA 的林业害虫实时检测方法[J]. 计算机工程.

(二) 申请的软著和专利

[1] 中北大学,候瑞环,杨喜旺,杨航,王智超,何宗博. 基于深度学习的昆虫检测系统 V1.0[软件专著]. 中华人民共和国国家版权局(2020SR0854139), 2020.07.30.

[2] 中北大学, 杨航, 候瑞环, 何宗博, 王智超, 高佳鑫. 机械零件标准化设计系统 V1.0[软件专著]. 中华人民共和国国家版权局(2020SR0854132), 2020.07.30.

致谢

行文至此，落笔为终，也意味着我在中北大学大数据学院三年的研究生生活已接近尾声。这里校园环境优美、资源丰富、平台广阔、学术氛围浓厚，为我提供了新的起点、新的人生方向。在这段宝贵的时光里，我学习知识的同时，也收获了珍贵的友情，这些时光记录着我不断成长与进步的步伐，给予了我继续迎难而上的勇气。在本文完成之际，请容许我对读研期间帮助过我的人致以深深的谢意。

首先，我要感谢我的导师杨喜旺老师与黄晋英老师。杨老师和蔼可亲、黄老师严厉负责，两位老师为课题组提供了良好的科研环境与科研设备，并给了我继续学习与研究的机会，支持我的论文选题，修正我的研究思路，悉心指导论文写作，一次次审思与斧正我的论文并督促我认真修改，培养我勤奋和严谨的治学态度；除了科研上的谆谆教导，两位老师在生活上给予我师辈般的包容与关怀，让我有了继续前行的信心与勇气，我很幸运在研究生阶段有这样宽厚仁爱的导师们。两位老师严谨的治学态度、精益求精的工作作风和深厚的专业造诣深深地影响着我。正因有老师的耐心指导，才使得我不断修正、完善、拓宽思路，不断提高论文写作与PPT制作能力，最终在科研上取得了进步。

其次，我要感谢实验室的同门。感谢王智超和高佳鑫师弟，在论文研究的理论学习与创新方面给予了我很大帮助，给我分享了许多国内外可以借鉴的前沿论文，并且在他们的帮助下，我在编程方面也提升了许多；感谢蔡波、张辉、杨凯和何宗博四位同学，不论在学习问题还是生活上，他们都给予了我极大的帮助和支持；同时感谢实验室其他所有师兄师姐师弟师妹们，感谢你们的陪伴，与你们在一起的时光快乐而又难忘。

然后，我要感谢家人对我的辛勤培育，感谢父母给予我莫大的支持与鼓励，感谢好朋友们的陪伴，你们是我最坚强的后盾。今后，我将继续努力，不辜负你们的期望。

最后，衷心感谢我的老师、同学以及家人朋友们，愿你们身体安康！对百忙之中抽出时间来进行论文评审的老师们致以深深的谢意，感谢你们对论文提出的宝贵意见。



中北大学研究生院监制