

分类号: TP391.41

学校代码: 10363

密 级: 公开

学 号: 2220320122



安徽工程大学

Anhui Polytechnic University

# 硕士学位论文

题目 六安瓜片嫩芽视觉识别研究

论 文 作 者 李鸿飞

校 内 导 师 孔敏（教授）

校 外 导 师 孙益博

专业学位类别 电子信息

研究方向（领域） 智能信息处理及应用

论文提交日期: 2025 年 5 月 15 日

分类号: TP391.41

单位代码: 10363

密 级: 公 开

学 号: 2220320122

题 目 六安瓜片嫩芽视觉识别研究

题 目 Study on Visual Recognition of  
Lu'an Gua Pian Shoots

学生姓名: 李鸿飞

校内导师: 孔敏（教授）

校外导师: 孙益博

专业学位类别: 电子信息

研究方向(领域): 智能信息处理及应用

论文答辩日期: 2025 年 5 月 10 日

## 安徽工程大学学位论文原创性声明

本人郑重声明：我恪守学术道德，崇尚严谨学风。所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已明确注明和引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品及成果的内容。论文为本人亲自撰写，我对所写的内容负责，并完全意识到本声明的法律后果由本人承担。

学位论文作者签名：李鸿飞

日期：2025 年 5 月 15 日

## 安徽工程大学学位论文版权使用授权书

学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅或借阅。本人授权安徽工程大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

保密 ☐，在 \_\_\_\_\_ 年解密后适用本版权书。

本学位论文属于

不保密 ☒。

学位论文作者签名：

李鸿飞

指导教师签名：

孔

日期：2025 年 5 月 15 日

日期：2025 年 5 月 15 日

## 六安瓜片嫩芽视觉识别研究

### 摘 要

作为全球最大的茶叶生产国和消费国，我国茶叶种类繁多，在世界名优茶提供国中具有很高的地位。但目前茶叶嫩芽采摘以人工为主，面临采摘效率低、成本高等问题。为解决这些问题，本文基于深度学习技术开展自然环境下六安瓜片嫩芽视觉识别研究，主要工作包括以下四个方面：

首先，构建复杂自然环境下生长的茶叶嫩芽图像数据集。通过采集多角度、多光照、多天气条件下的六安瓜片图像，结合 LabelImg 工具对茶叶图像进行标注。采用翻转、Gamma 变换、添加噪声实现对茶叶图像离线数据增强，并在训练过程中引入 HSV 色彩空间变换、尺度缩放、Mosaic 等在线数据增强技术，有效解决茶芽样本过少问题，并提升数据集泛化能力。

其次，针对 YOLOv5 模型在公开茶芽数据集上特征提取能力不足、边界框偏移导致茶叶嫩芽漏检、误检问题做出改进，搭建了 YOLO-CBN 网络。在主干网络中嵌入 CBAM 混合注意力机制，增强模型对复杂环境下对茶芽目标的关注能力以及背景的抗干扰能力；在颈部网络中采用 BiFPN 替换原有的 PAN 结构，融合不同尺度的特征信息；选用归一化 NWD 损失函数，用 Wasserstein 距离来更准确地测量预测边界框和实际边界框之间的差异，提高模型对小目标茶芽检

测的准确度，同时更好地检测重叠茶芽。实验后得到的数据显示，改进后的模型平均精度达到 70%，较基准模型提升了 5.7%。

然后，为实现移动端部署需求，基于自制高质量数据集及 YOLO-CBN 基础上设计了轻量化且精度高的模型 YOLO-Tea。采用 SimSPPF 模块替代标准 SPPF，通过激活函数优化在保证精度的同时提升了推理速度；复用 YOLO-CBN 中的 BiFPN 结构，加强特征融合；将颈部网络的 Conv 卷积模块替换为 ODCConv 全维动态卷积，从空间大小、输入通道、输出通道和卷积核四个不同维度的权重进行学习，来提高网络模型的特征表达能力。实验结果显示最后模型权重文件仅为 14.9MB，平均精度为 83.7%，相较于 YOLOv8 等主流算法在模型大小和精度上具有显著优势。

最后，开发基于 PyQt5 的茶芽检测可视化用户界面。构建包含图像处理、视频流分析和实时检测功能的 GUI 平台，实现检测结果的可视化展示，处理后的结果可以保存到指定文件，方便对比调整模型结构。经过测试，能够满足茶园现场作业的工程化应用需求。

**关键词：**茶芽检测，YOLOv5，特征融合，全维动态卷积，PyQt5

# STUDY ON VISUAL RECOGNITION OF LU'AN GUA PIAN SHOOTS

## ABSTRACT

As the world's largest producer and consumer of tea, China has a wide variety of tea varieties and holds a high position among the world's top tea suppliers. But currently, the picking of tea buds is mainly done manually, facing problems such as low picking efficiency and high costs. In order to solve these problems, this study conducts research on visual recognition of Lu'an Gua Pian buds in natural environment based on deep learning technology. The main work includes the following four aspects:

Firstly, construct a dataset of tea sprout images grown in complex natural environments. The tea image was annotated with LabelImg tool by collecting Lu'an Gua Pian images under multi angle, multi light and multi weather conditions. We use flipping, Gamma transformation, and adding noise to enhance offline data of tea images, and introduce online data augmentation techniques such as HSV color space transformation, scaling, and Mosaic during the training process to effectively solve the problem of insufficient tea bud samples and improve the generalization ability of the dataset.

Secondly, in response to the insufficient feature extraction ability of

YOLOv5 model on the public tea bud dataset, as well as the problem of missed and false detection of tea buds caused by boundary box offset, improvements were made by building YOLO-CBN network. Embedding CBAM hybrid attention mechanism in the backbone network to enhance the model's ability to focus on tea bud targets in complex environments and resist background interference; In the neck network, BiFPN is used to replace the original PAN structure and fuse feature information of different scales; Using normalized NWD loss function and Wasserstein distance to more accurately measure the difference between predicted bounding boxes and actual bounding boxes, improve the accuracy of the model for detecting small target tea buds, and better detect overlapping tea buds. The data obtained after the experiment shows that the improved model has an average accuracy of 70%, which is 5.7% higher than the benchmark model.

Then, to meet the deployment requirements of mobile devices, a lightweight and high-precision model YOLO Tea was designed based on a self-made high-quality dataset and YOLO-CBN. Using SimSPPF module instead of standard SPPF, the activation function optimization improves inference speed while ensuring accuracy; Reuse the BiFPN structure in YOLO-BBN to enhance feature fusion; Replace the Conv convolution module of the neck network with ODConv full dimensional dynamic convolution, and learn weights from four different dimensions: spatial size, input channel, output channel, and convolution kernel to improve the



feature expression ability of the network model. The experimental results show that the final model weight file is only 14.9MB, with an average accuracy of 83.7%. Compared with mainstream algorithms such as YOLOv8, it has significant advantages in model size and accuracy.

Finally, develop a visual user interface for tea bud detection based on PyQt5. Build a GUI platform that includes image processing, video stream analysis, and real-time detection functions to achieve visual display of detection results. The processed results can be saved to designated files for easy comparison and adjustment of model structure. After testing, it can meet the engineering application requirements of on-site operations in tea gardens.

Li Hongfei (Electronic Information)

Supervised by Kong Min

**KEY WORDS:** Tea sprout detection, YOLOv5, Feature fusion, Full dimensional dynamic convolution, PyQt5

## 目 录

|                                   |     |
|-----------------------------------|-----|
| 摘 要.....                          | I   |
| ABSTRACT.....                     | III |
| 第 1 章 绪论.....                     | 1   |
| 1.1 研究背景及意义.....                  | 1   |
| 1.2 国内外研究现状.....                  | 2   |
| 1.2.1 传统茶芽目标检测现状.....             | 2   |
| 1.2.2 基于深度学习茶芽目标检测研究现状.....       | 2   |
| 1.3 存在问题.....                     | 3   |
| 1.4 本文的主要研究内容.....                | 3   |
| 第 2 章 深度学习基础理论与茶叶数据集构建.....       | 6   |
| 2.1 卷积神经网络.....                   | 6   |
| 2.2 卷积神经网络经典模型.....               | 8   |
| 2.3 YOLO 系列算法介绍.....              | 11  |
| 2.4 茶叶嫩芽数据集构建.....                | 13  |
| 2.4.1 图像采集地点.....                 | 14  |
| 2.4.2 茶芽图像数据采集.....               | 15  |
| 2.4.3 数据标注.....                   | 16  |
| 2.5 数据增强.....                     | 17  |
| 2.5.1 离线数据增强方法.....               | 18  |
| 2.5.2 在线数据增强方法.....               | 20  |
| 2.6 本章小结.....                     | 20  |
| 第 3 章 基于 YOLO-CBN 网络茶芽检测模型.....   | 21  |
| 3.1 YOLO-CBN 网络结构.....            | 21  |
| 3.1.1 CBAM 注意力机制.....             | 22  |
| 3.1.2 BiFPN 结构.....               | 23  |
| 3.1.3 归一化 NWD 损失函数.....           | 24  |
| 3.2 结果分析.....                     | 26  |
| 3.2.1 实验评价指标.....                 | 26  |
| 3.2.2 实验环境配置.....                 | 27  |
| 3.2.3 消融对比实验.....                 | 28  |
| 3.2.4 注意力模块选择及嵌入位置.....           | 29  |
| 3.2.5 改进损失函数前后实验对比.....           | 30  |
| 3.3 本章小结.....                     | 31  |
| 第 4 章 基于 YOLO-Tea 的轻量化茶芽检测模型..... | 32  |
| 4.1 YOLO-Tea 网络结构.....            | 32  |
| 4.1.1 SimSPPF 模块.....             | 33  |
| 4.1.2 ODConv 卷积模块.....            | 34  |
| 4.2 实验结果与分析.....                  | 37  |
| 4.2.1 实验细节.....                   | 37  |
| 4.2.2 消融对比实验.....                 | 37  |
| 4.2.3 与其他目标检测网络对比.....            | 38  |
| 4.3 本章小结.....                     | 39  |

|                                  |    |
|----------------------------------|----|
| 第 5 章 基于 PyQt5 的茶芽目标检测系统设计 ..... | 40 |
| 5.1 系统概述.....                    | 40 |
| 5.1.1 PyQt5 简介 .....             | 40 |
| 5.1.2 架构设计.....                  | 41 |
| 5.2 人机交互界面设计.....                | 41 |
| 5.2.1 环境配置.....                  | 41 |
| 5.2.2 主界面设计.....                 | 42 |
| 5.3 功能测试.....                    | 43 |
| 5.4 本章小结.....                    | 45 |
| 第 6 章 总结与展望.....                 | 46 |
| 6.1 总结.....                      | 46 |
| 6.2 展望.....                      | 47 |
| 参考文献.....                        | 48 |
| 攻读学位期间发表的学术成果.....               | 55 |
| 致 谢.....                         | 56 |

## 第 1 章 绪论

### 1.1 研究背景及意义

茶叶的起源众说纷纭，根据《神农百草经》记载，“神农尝百草，日遇七十二毒，得茶而解之”，早在公元 2700 多年就已经记载了茶叶的药用价值<sup>[1]</sup>。随着时间推移，茶叶中含有丰富的活性化合物被逐渐认知，茶叶也逐渐从药用价值发展到食用<sup>[2]</sup>，再到现在的饮用，茶也逐渐成为比较热门的饮品，销量仅次于水。茶叶发展过程不仅传承了茶的实用价值，也承载了悠久丰富的历史文化内涵<sup>[3]</sup>。茶艺，一种泡茶与饮茶的技术，承载数千年智慧与深厚感情的文化被传承下来。据不完全统计，全球茶叶种类大约有 3000 多种，主要分为六大类：绿茶、红茶、乌龙茶、白茶、黄茶和黑茶，每一类又有多个品种和变种<sup>[4]</sup>。

中国是世界上最大的茶叶生产国和消费国，在世界名优茶提供国中具有很高的地位，茶叶种类超过 1000 种。如图 1-1 所示，我国茶园面积和茶叶产量逐年增加<sup>[5]</sup>。2023 年茶叶产量达到了 355 万吨，同比增长 6.2%。我国 18 个主要产茶省（自治区、直辖市）的茶园总面积为 5149.76 万亩，同比增长 3.09%，占世界产茶的四成，位居世界首位<sup>[6]</sup>。随着脱贫攻坚实施，茶叶产业在农村经济发展中占据着越来越重要的作用，甚至在一些地区成为了支柱性产业<sup>[7]</sup>，如六安瓜片、霍山黄芽、舒城小兰花、祁门红茶、太平猴魁等。这些茶叶品种不仅是当地的特殊优势产业，能够促进当地经济发展，也对当地文化具有宣传作用，提高知名度<sup>[8]</sup>。

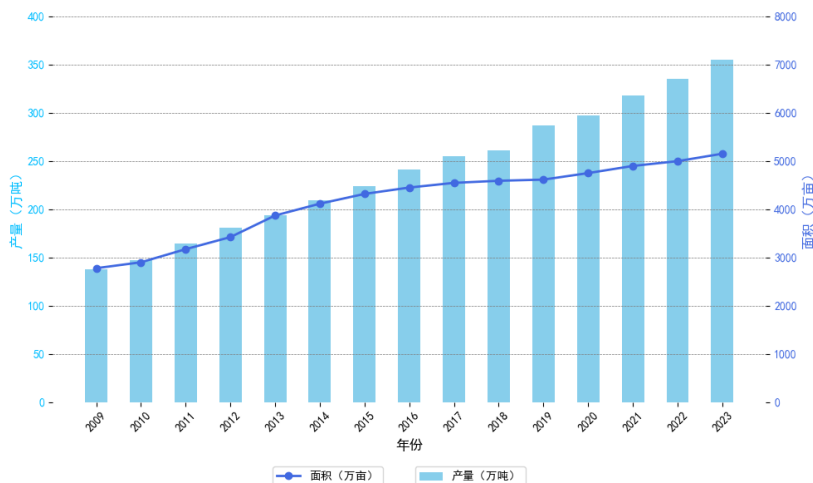


图 1-1 2009-2023 年我国茶园面积及茶叶产量变化趋势

茶叶与其它季节性作物一样，主要采摘时间根据茶树品种固定收获，目前茶叶采摘分为人工采摘和机械采摘两种方式。茶叶种植环境多为山区，主要采摘方式仍为传统的人工采摘，因此采摘效率低、成本高<sup>[9]</sup>，如图 1-2(a)所示。而现有的机械采茶主要为螺旋滚刀式、往复式及水平勾刀式，不仅无法精确定位到茶叶嫩芽，会导致茶树树枝和叶片折断，而且对茶树后续生长造成影响<sup>[10]</sup>，降低采摘出的茶芽总体质量，如图 1-2(b)所示。更重要的是只能适用于平整的茶园种植地，对于高低不平的山区茶园难以适用，因此具有极大的局限性<sup>[11]</sup>。随着茶叶产销量的不断提高，为了解决上述问题，基于深度学习及目标检测算法被逐渐用于机械设备上进行茶叶采摘<sup>[12]</sup>。本文针对当前算法问题，提出一种高精度的轻量化算法，保证检测精度的同时，能够方便部署到采摘设备上<sup>[13]</sup>。



(a)手工采茶



(b)机械采茶

图 1-2 手工采茶和机械采茶

## 1.2 国内外研究现状

目前茶芽目标检测的方法主要可以分为基于传统方法的目标检测和基于深度学习的目标检测两大类。传统目标检测是在给定图像上选择多个候选区域，在候选区域进行特征提取，最后再用训练分类器进行分类，输出检测目标的位置和类别信息<sup>[14]</sup>。而基于深度学习的目标检测是利用卷积神经网络的卷积层、池化层、全连接层来自动进行特征提取和分类<sup>[15]</sup>。

### 1.2.1 传统茶芽目标检测现状

传统图像处理研究主要基于颜色差异、纹理、形状等特征来区分茶芽与背景。唐仙等人<sup>[16]</sup>对 RGB 颜色模型的色差法进行分析，以该算子作为分割量，对比多种阈值分割法，得出 Ostu 和迭代阈值法茶芽识别效果最好。彭红星等人<sup>[17]</sup>提出

多类色彩目标识别，基于两次 Otsu 分割算法，提高分割质量和正确率。姜苗苗等[18]利用 MRFO 优化算法训练得到 ExG 系数，用该系数对嫩芽 RGB 图像进行灰度化，通过 Otsu 获得图像阈值计算，对阈值偏移矫正得到最佳阈值。周礼赞<sup>[19]</sup>等针对茶叶嫩芽分割问题提出了基于 G-B 分量的方法，通过分析茶叶嫩芽在 RGB 色彩空间下各分量信息，利用灰度拉伸扩大 G-B 色差图对比度，通过迭代法获取全局阈值，将嫩芽图像分割。Karunasena 等人<sup>[20]</sup>提取茶芽 RGB 图像，通过 LBP/C 算法，对茶叶嫩芽纹理和形状进行提取，结合支持向量机的茶芽定位识别完成模型训练。Zhang 等人<sup>[21]</sup>先将图像进行高斯滤波，通过最小误差确定最优阈值，并使用分段线性变换增强 G-B 分量，提高老叶与茶芽区分度，最后用分水岭算法对嫩芽进行分割。张博等人<sup>[22]</sup>基于 LabVIEW 图像采集模块获取图像，选择 HSL 色彩空间进行多灰度阈值分割，去除过大面积和过小面积区域，留下合适区域进行检测。以上方法都是通过计算机视觉中的图像处理技术人为获得茶叶嫩芽特征，来区分老叶和嫩芽的特征，由于人为干预过大，许多特征容易忽略。因此在实际采茶场景中识别速度慢、精度差，而且泛化能力较弱<sup>[23]</sup>。

### 1.2.2 基于深度学习茶芽目标检测研究现状

近年来，深度学习技术不断发展迭代，通过语义分割和目标检测算法能够对农作物进行检测和识别，并应用到相关农业领域。在茶芽目标检测方面将输入图像的目标类别和位置信息送入卷积神经网络进行模型训练，减少了茶叶嫩芽特征提取人为干预。相比较传统目标检测有较高的鲁棒性和泛化能力<sup>[24]</sup>。

Chen 等人<sup>[25]</sup>首先训练了一个更快的基于区域的 Faster R-CNN 神经网络来检测图像中的目标区域，随后再训练一个完全卷积网络，完成目标区域的采摘点的定位。Qian 等人<sup>[26]</sup>提出了一种基于改进的深度卷积编码解码网络的茶芽分割新方法，通过对比中心损失函数和跳跃连接进行改进，利用类间可分性更准确的识别茶芽特征。Yan 等人<sup>[27]</sup>改进 Mask R-CNN 模型，通过扩展网络设计中的 mask 分支和计算连通域的面积来识别，在非结构化环境下对重叠茶芽的识别效果较好。Xu 等人<sup>[28]</sup>将拥有快速检测优点的 YOLOv3 和拥有高精度分类能力的 DeniseNet201 结合实现对茶芽的精准定位。Li 等人<sup>[29]</sup>在 YOLOv4 网络中嵌入 CBAM 注意力模块，通过重新定义 SIoU 损失函数惩罚指标，在提高检测精度的同时，减小模型大小，降低机器人视觉模块的部署成本和难度。Shuai 等人<sup>[30]</sup>，在

YOLOv5 网络中，用 Bottleneck Transformers 作为残差模块，结合更大感知域的上采样算子 CARAFE 和注意力机制，来创建对茶芽特征图像的远距离依托关系，提升定位能力。Xie 等人<sup>[31]</sup>，提出的 Tea-YOLOv8s 模型结合了可变卷积、注意力机制和改进的空间金字塔池化，减少无关因素干扰，从而提高了检测精度。

以上研究在结果上表明，深度学习技术在茶芽目标检测结果上表现的更好，检测效率更高。因此现阶段茶芽目标检测领域中心已经从传统方法转向深度学习方法<sup>[32]</sup>。如何将卷积神经网络运用到实际场景、高效检测茶叶嫩芽是当前研究的主要方向<sup>[33]</sup>，对于林下作物检测具有重要推动。

### 1.3 存在问题

目前，茶叶嫩芽目标检测领域仍然存在以下几个难点：

(1) 由于复杂自然条件下茶叶图像采集困难，且茶叶种类过多，导致模型训练的茶芽图像样本数量有限。这种局限性下训练模型容易产生过拟合的问题，训练出的模型泛化能力差，无法运用到实际采茶场景。

(2) 茶叶多数生长在山区，在实际采茶场景中，受光照、遮挡情况影响，导致不同区域茶树长势差异较大，茶芽尺度大小不一，模型容易出现漏检的情况<sup>[34]</sup>。同时茶芽背景复杂，老叶、嫩芽颜色相近，互相遮挡情况严重，山区天气多变等因素都会影响模型检测能力，使检测精度降低。

(3) 通过不断加深神经网络层数，能够提高模型特征表达能力，提高检测性能。但会导致模型参数不断增加，需要更大的存储空间，不利于模型部署。

### 1.4 本文的主要研究内容

本文以自然条件下茶叶嫩芽为研究对象，为了实现更高效的智能化采摘，结合实际情况，精准定位茶芽位置，并不断优化改进茶芽识别算法，为茶叶嫩芽农业化、现代化采摘任务提供解决方案，确保机械采茶准确迅速的完成任务。本文技术路线如图 1-3 所示，各章节安排如下：

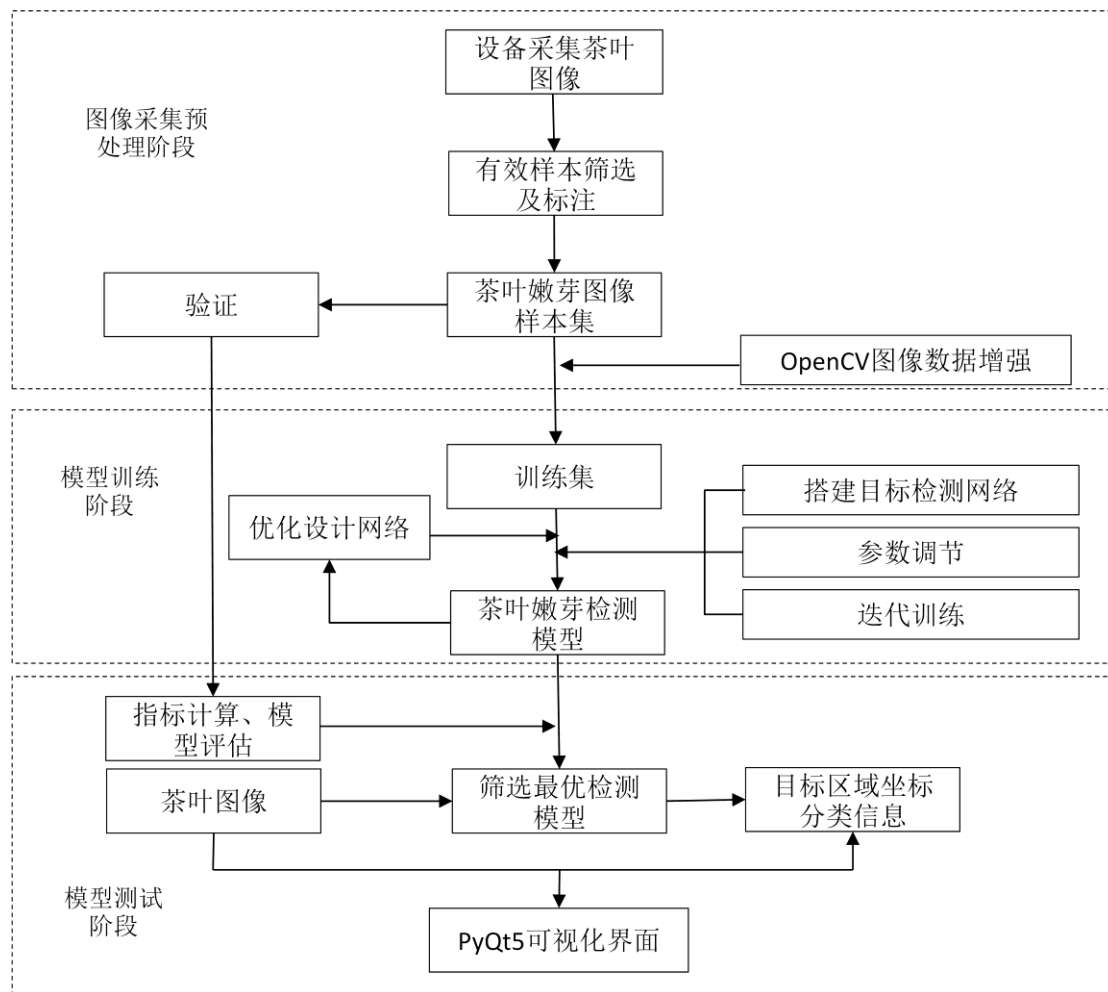


图 1-3 技术路线

第 1 章：绪论。首先介绍了茶叶嫩芽视觉识别算法研究的背景及意义，简单阐述了现阶段茶芽识别研究现状。其次分析了传统目标检测和基于深度学习目标检测的优劣，以及现阶段茶芽检测主要存在的问题。最后对本文技术路线有简单介绍。

第 2 章：深度学习相关知识与茶叶数据集构建。首先介绍了卷积神经网络如何特征提取，和现阶段主要卷积神经网络发展路线。其次，详细介绍了六安瓜片图像数据集采集和数据标注。最后说明了本文数据集主要进行的在线数据增强和离线数据增强的几个主要方式。

第 3 章：基于 YOLO-CBN 网络茶芽检测模型。搭建基于 YOLOv5 的 YOLO-CBN 深度学习模型，首先在主干网络中嵌入结合通道和空间的 CBAM 混合注意力机制，提高模型对复杂环境下对茶芽目标的特征提取能力。其次在颈部网络中使用更高效的 BiFPN 结构，融合不同尺度的特征信息。最后结合归一化 NWD 损



失函数，解决边界框偏移问题，这提高了模型对小目标茶芽的检测能力，同时能够更好地处理重叠茶芽。

第 4 章：基于 YOLO-Tea 的轻量化茶芽检测模型。为了方便将网络移植到移动设备上，基于自制数据集和第 3 章模型的基础上设计了轻量化且精度较高的模型，首先将主干网络部分使用 SimSPPF 模块，提高模型精度和检测速度。然后复用 YOLO-CBN 中的 BiFPN 结构，加强多尺度融合。最后将颈部网络的 Conv 卷积模块替换为 ODConv 全维动态卷积，从空间大小、输入通道、输出通道和卷积核四个不同维度的权重进行学习，来提高网络模型的识别性能，提高特征适应性。最终目的减轻模型大小，增加模型可移植性。

第 5 章：基于 PyQt5 的茶芽目标检测系统设计。载入模型训练好的模型文件，通过图形界面对图像、视频和实时场景的茶芽目标进行识别，使结果可视化，最后将处理后的结果可以保存到指定文件。

第 6 章：总结与展望。总结了本文使用的研究方法和取得的成果，针对本文工作存在的不足，对茶叶嫩芽目标检测的未来研究进行了展望。

## 第 2 章 深度学习基础理论与茶叶数据集构建

基于卷积神经网络目标检测算法在农业智能采摘领域已经得到了广泛的运用。本章首先介绍卷积神经网络如何进行特征提取，其次对目前主流的目标检测 YOLO 算法迭代发展过程，最后介绍茶自制茶叶嫩芽数据集采集、标注和数据增强的方式。

### 2.1 卷积神经网络

卷积神经网络（Convolutional Neural Networks, CNN）是由卷积层、激活函数层、池化层和全连接层组成，主要用于图像分类、目标检测和图像分割，依赖于卷积操作，实现对目标的特征提取<sup>[35]</sup>。与传统神经网络相比，CNN 有以下优势：(1)参数共享，同一卷积核在输入数据上滑动，可以减少参数量。(2)特征提取，避免手动设计特征，自动学习图像特征。(3)特征降维，通过其中的池化层可以降低目标特征的维度，既能降低网络参数量，又能保存目标的有用特征信息<sup>[36]</sup>。

#### (1) 卷积层

这是 CNN 的第一个组成模块，也是最重要的模块，主要执行数学任务中的卷积，用于提取目标有效特征并生成对应的特征图。卷积层通过滑动卷积核在输入图像上进行卷积运算，具体是通过对对应特征像素相乘之后将结果求和，从而得到图像的局部特征<sup>[37]</sup>。从左到右、从上到下重复该过程，处理完原特征图的所有特征点后，输出新的特征图，这称为一次卷积操作，结构如图 2-1 所示。

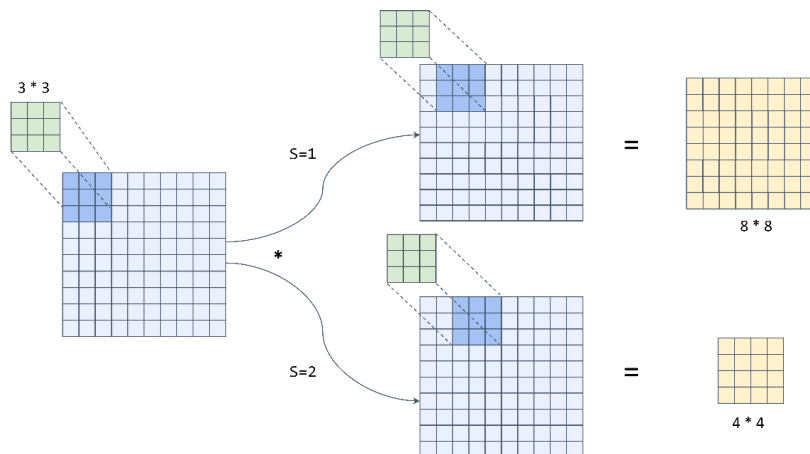


图 2-1 卷积操作

输出特征图的大小根据输入特征图尺寸、卷积核大小、步长 $s$ 和填充 $p$ 来确定，当输入特征图尺寸为 $M \times N$ 时，卷积核大小为 $k \times k$ 时，卷积后的目标特征图的大小为公式（2-1）和（2-2）所示，输出特征图的大小就为 $M' \times N'$ 。

$$M' = \frac{M + 2 \times p - k}{s} + 1 \quad (2-1)$$

$$N' = \frac{N + 2 \times p - k}{s} + 1 \quad (2-2)$$

## （2）激活函数

经过卷积层后的模型是不同线性模型的叠加，所以无论卷积层叠加多少，框架有多复杂，都是相当于单层的线性模型。为了保证有效卷积，每次卷积操作后都要使用激活函数，实际上这是一类简单的数学函数。该函数可帮助网络学习图像特征之间的非线性关系，从而帮助模型更好的逼近任何非线性函数，同时还能够缓解梯度消失问题<sup>[38]</sup>。

激活函数结构类似人脑神经元，激活函数相当于一个神经单元，可以决定特征信息被传到哪个单元中，具体就是前一层的神元收集和处理特征信息，通过激活函数对这些信息进行加工处理，然后传递给下一层的神元<sup>[39]</sup>，来实现对原模型的逼近，激活函数结构图如 2-2 所示。

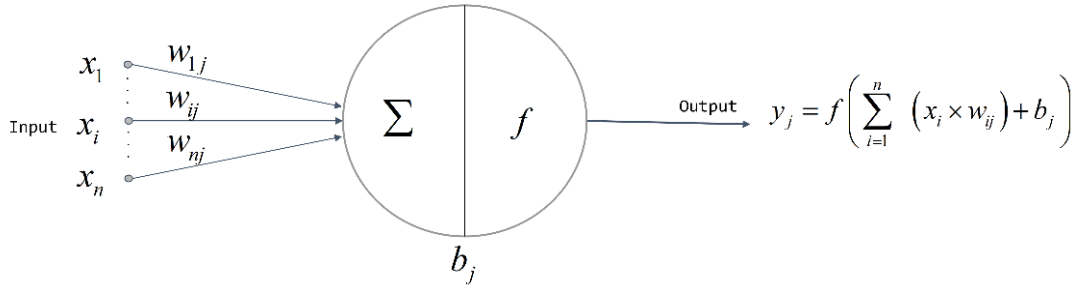


图 2-2 激活函数结构图

在 2-2 中， $x_i$  就是上一层神经元传来的输出特征， $w_{ij}$  代表对应特征的权值信息，将前一层的  $n$  个特征结合权值信息传到神经元中，加入  $b_j$  大小的偏置，经由激活函数  $f$ ，得到最终输出。常见的激活函数有 sigmoid、tanh 和 Relu 激活函数。

## （3）池化层

池化层是卷积神经网络中用于降维操作，其核心功能是通过局部区域进行下采样，在保留主要特征的同时降低数据维度和参数量。其主要池化操作包括最

大池化（Max Pooling）和平均池化（Average Pooling）。最大池化选择窗口内像素的最大值，适用于保留纹理细节；平均池化计算窗口内的平均值，平滑特征响应，适用于抑制噪声<sup>[40]</sup>。池化层通过设置步长 $s$ 逐个区域滑动，池化层降低了后续卷积层的计算开销，是构建高效分层特征金字塔的基础<sup>[41]</sup>。以最大池化为例，当步长 $s$ 为 2，填充  $p$  为 0，池化核大小为  $2 \times 2$ ，计算过程如图 2-3 所示。

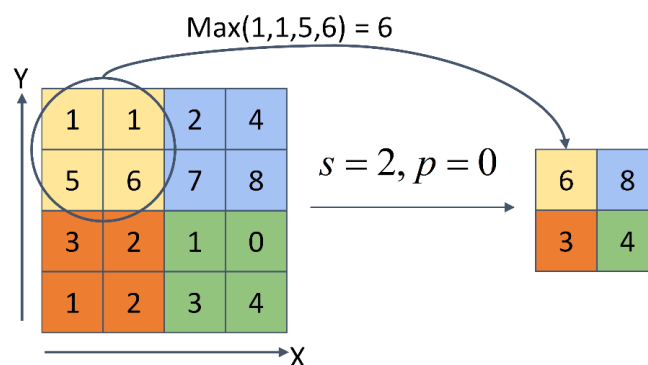


图 2-3 最大池化操作

#### (4) 全连接层

全连接层（Fully Connected Layer）是神经网络中最后一层，其功能是将前一层的特征图展平为一维向量后，通过全连接权重矩阵将其映射到样本标记空间的作用<sup>[42]</sup>。在实际使用中，全连接层可由卷积操作实现，具体是对前层是全连接的全连接层可以转换为卷积核为  $1 \times 1$  的卷积，而前层是卷积层的全连接层可以转换为卷积核为前层卷积输出结果的高和宽一样大小的全局卷积<sup>[43]</sup>。

## 2.2 卷积神经网络经典模型

自 1998 年 LeNet 被提出后，神经网络迎来了迅速发展，各种经典的网络被提出来，本节将重点对以下几种比较经典的神经网络结构进行介绍，分别为 LeNet 网络、AlexNet、VGG 以及 ResNet 网络。

#### (1) LeNet 网络

LeNet 是首个成功应用于实际任务的卷积神经网络，其用于 MNIST 手写数字数据集识别错误率低于 1%，首次系统化整合卷积层、池化层和全连接层，奠定现代 CNN 的基本框架<sup>[44]</sup>。该网络包含两个  $5 \times 5$  的卷积核、两个  $2 \times 2$  的平均池

化和三个全连接层，将输入  $28 \times 28$  的黑白图像，经由网络后输出 10 维输出，分别对应 0 到 9 这 10 个输出类别，LeNet 网络结构如图 2-4 所示。

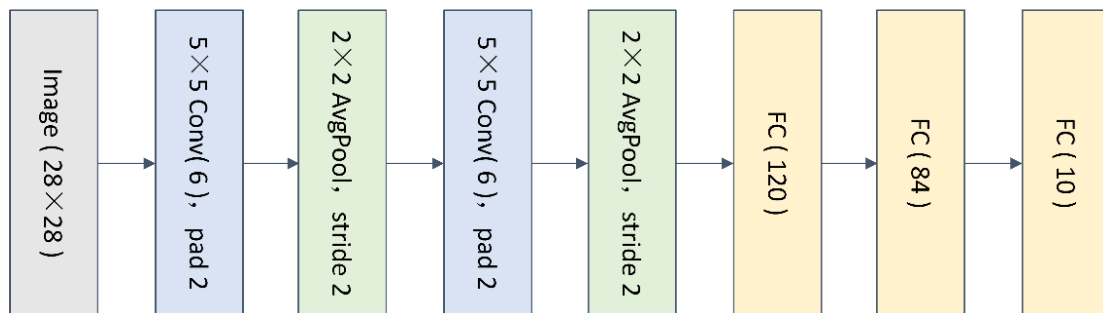


图 2-4 LeNet 网络结构图

### (2) AlexNet 网络

AlexNet 相较于 LeNet 将网络深度扩展至 8 层，由 5 个卷积层、3 个最大池化层和 4 个全连接层组成，显著提升了特征提取能力。其中第一层采用  $11 \times 11$  的大卷积核，用于捕捉全局特征，然后逐步缩小至  $5 \times 5$  和  $3 \times 3$  来提取局部细节，并通过步幅为 2 的  $3 \times 3$  最大池化层逐渐压缩特征尺寸。使用 Relu 激活函数替换 Sigmoid，减少了幂运算，加速收敛并缓解梯度消失问题<sup>[45]</sup>，结构图如 2-5 所示。

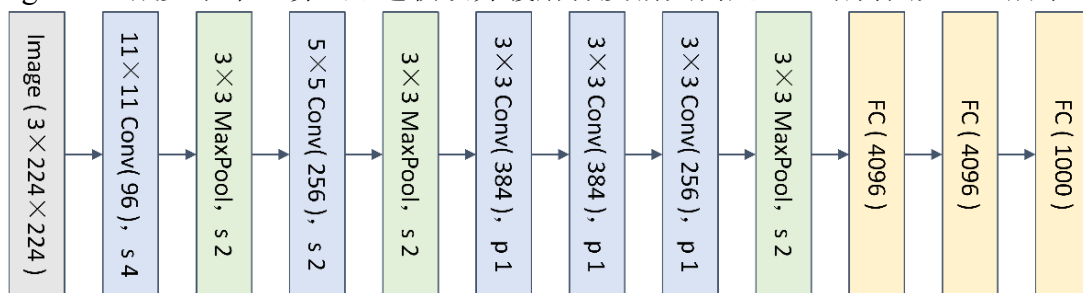


图 2-5 AlexNet 网络结构图

### (3) VGG 网络

在 VGG 提出之前，AlexNet 就已经使 CNN 走进大众视野，相比较于 AlexNet，VGG 卷积核、池化层更深，通道数也更多，而 VGG 的出现，正式开启了在 CNN 中使用小卷积核代替大卷积核的方法来进行特征提取，从图 2-6 所示，VGG 使用的卷积核全为  $3 \times 3$ 。摒弃 AlexNet 中的  $11 \times 11$  和  $5 \times 5$  大卷积核，采用  $3 \times 3$  小卷积核堆叠特征提取，能够降低参数量，证明了更小卷积核和更深的神经网络深度能够有效提高模型的性能。尽管 VGG 的理念是通过更深的网络结构提高提取特征的能力，但通过统一的卷积和池化结构可以使网络更加简洁高效，在 VGGNet 训练中，也可以先训练浅层的简单的 VGG，复用之前 VGG 权重，如此反复训

练，能够使更深的 VGG 收敛速度更快<sup>[46]</sup>。

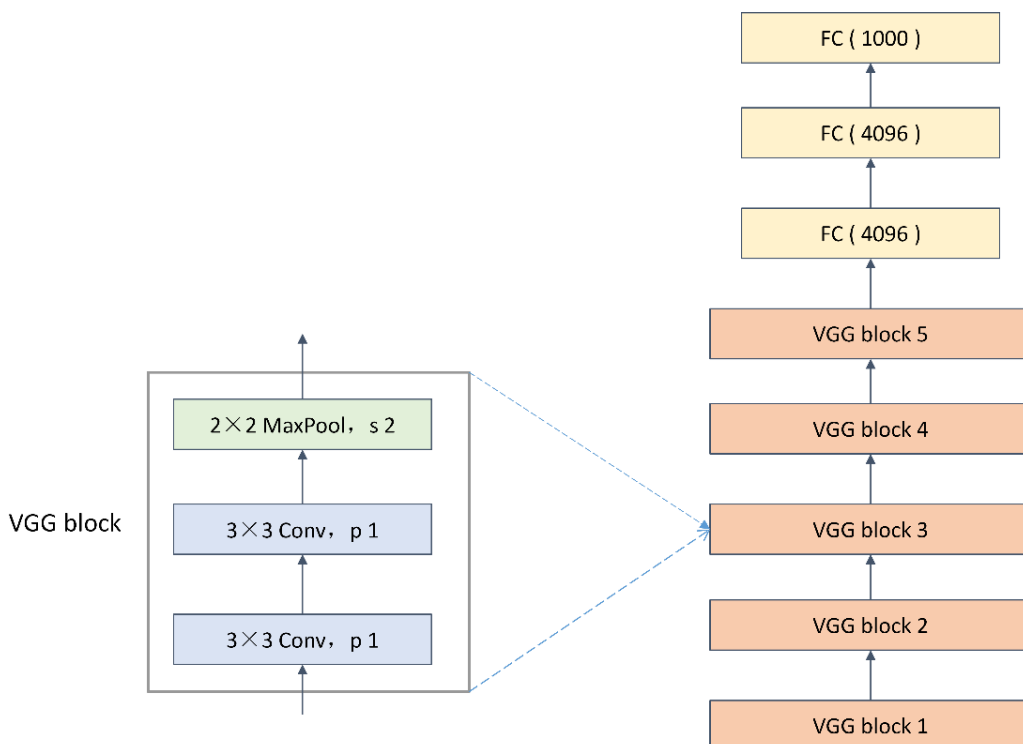


图 2-6 VGG 网络结构图

#### (4) ResNet 网络

从理论上讲，加深神经网络的深度就可以使每一层学习到更加细化的特征，提升模型表达能力和性能。但随着网络不断堆叠，梯度呈现指数级增长，虽然每次都会和激活函数的导数相乘减少这个值，仍会抑制不住指数增长，这就会引发梯度爆炸和梯度消失问题。另外一个问题就是，当网络层数过多时，非线性激活函数的存在会造成很多不可逆的信息损失，网络加深到一定程度，过多的信息损失就会造成网络的退化。在 AlexNet 开始，神经网络将 Sigmoid 替换为 Relu 是为了减少梯度爆炸和梯度消失的可能性。而 ResNet 引入残差学习的思想，通过在网络中加入跳跃连接，使得每一层不仅学习当前的特征，还可以直接跳过一部分层级信息，从而进一步缓解了梯度消失和信息丢失问题。残差结构的目的是使  $F(x)$  逼近于 0，在网络深度不断增加时，找到最优的浅层网络，保证网络不会因为深度而发生网络退化<sup>[47]</sup>。残差模块的应用解决了深层网络训练中的实际问题，使神经网络能够更加高效、稳定的训练，其结构如图 2-7 所示。

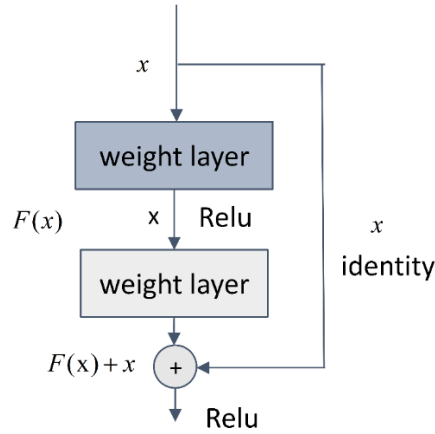


图 2-7 残差模块结构图

随着卷积神经网络发展,基于区域的 CNN 检测图像中对象的算法演化出来。基于区域的 CNN 的基本思想是首先在图像或视频中生成一组候选区域,然后使用 CNN 将每个区域分类为包含对象或不包含对象。然后对被分类为包含对象的区域进行进一步处理,以优化对象的位置并将其归类为特定类别。此概念也称为 two-stage 检测。例如 Fast R-CNN、Faster R-CNN 和 Mask R-CNN 它们都是为了提高目标检测过程的速度和准确性而提出的。这些模型结合一些方法来生成候选区域来对包含对象的区域进行分类和细化。

尽管基于区域的 CNN 能够以相对令人满意的准确度解决物体检测任务,但它们也存在一些缺点。其中一些缺点是:计算成本高、效率低、泛化能力有限以及缺乏上下文。它们计算成本高且效率低,因此不适用于低功耗设备上的实时物体检测。此外,还需要大量标记数据集,而获取这些数据集的成本很高且耗时,这些因素使它们不适合应用到实际场景。

## 2.3 YOLO 系列算法介绍

One-stage 是另外一类物体检测算法,它能够仅需一次通过模型检测图像或视频中的物体,而无需分多个阶段或区域检测,例如 YOLO(You Only Look Once, YOLO)和 SSD(Single Shot MultiBox Detector, SSD),使用单个卷积神经网络直接预测图像或视频中物体的类别标签和边界框。这些模型使用大量带标签图像及其相关物体边界框的数据集进行端到端训练。单次物体检测比两次物体检测方法更快、更高效,因为它无需多个阶段,可以在较低功率设备上实时运行。

YOLOv1 的基本思想是将输入图像划分为单元格网格,并针对每个单元格预测物体存在的概率以及物体的边界框坐标。YOLOv1 的过程可以分为几个步骤:输入图像通过 CNN 从图像中提取特征;然后将这些特征传递给一系列完全连接的层,这些层可以预测类概率和边界框坐标;将图像分成一个个单元格网格后,每个单元格负责预测一组边界框和类概率,网络的输出是每个单元的一组边界框和类概率;然后使用非最大抑制的后处理算法过滤边界框,以删除重叠的框并选择具有最高概率的框;最终的输出是图像中每个对象的一组预测边界框和类标签<sup>[48]</sup>。YOLOv1 的一个主要优势是它可以一次性处理整个图像,这使得它比 R-CNN 及其变体等两阶段物体检测器更快、更高效。

YOLOv2 是 YOLOv1 检测算法改进版。YOLOv2 与原始 YOLO 之间的主要区别之一是使用多尺度方法,在 YOLOv2 中,输入图像以多种尺度输入 CNN,这使得模型能够检测不同大小的物体。这是通过使用特征金字塔网络 FPN (feature pyramid network, FPN) 实现的,该网络允许模型从同一图像中提取不同尺度的特征;YOLOv2 使用与原始 YOLO 不同的损失函数,称为平方误差(sum-squared error, SSE) 损失函数, SSE 损失函数更加稳定,有助于模型更快地收敛;在架构方面,YOLOv2 网络更深,特征提取能力更强<sup>[49]</sup>。

YOLOv3 采用 Darknet-53 网络架构,结合上采用和维度拼接的多尺度方法,对于小目标识别效果更好;使用 logistic 多标签分类替代 softmax 单标签分类的问题<sup>[50]</sup>。

YOLOv4 是 YOLO 重要版本之一,相比较之前采用更先进的网络架构 SPP (Spatial Pyramid Processing, SPP),能够从不同尺度和分辨率的图像中提取特征,使模型能够检测不同大小的物体;此外,YOLOv4 还使用了 CSP (Cross-stage partial connection, CSP) 的结构来提高模型的准确性<sup>[51]</sup>。

YOLOv5 由 Ultralytics 团队于 2020 年提出。基于 EfficientNet 架构在许多数据集上取得最好的性能, EfficientDet 架构旨在提高计算和内存使用效率,同时实现高精度;相比较于其它 YOLO 算法, YOLOv5 直接使用单个卷积层直接预测边界框,使模型更加灵活轻量;在数据集上采用 mosaic 数据增强;训练方面能够使用迁移学习,这使得它可以在大型数据集上进行预训练,然后在较小的数据集上进行微调。这使得模型可以从广泛的数据中学习,并更好地推广到新数据



[52]。本文基于 YOLOv5 逐步改进模型提高需求，YOLOv5 网络结构如图 2-8 所示。

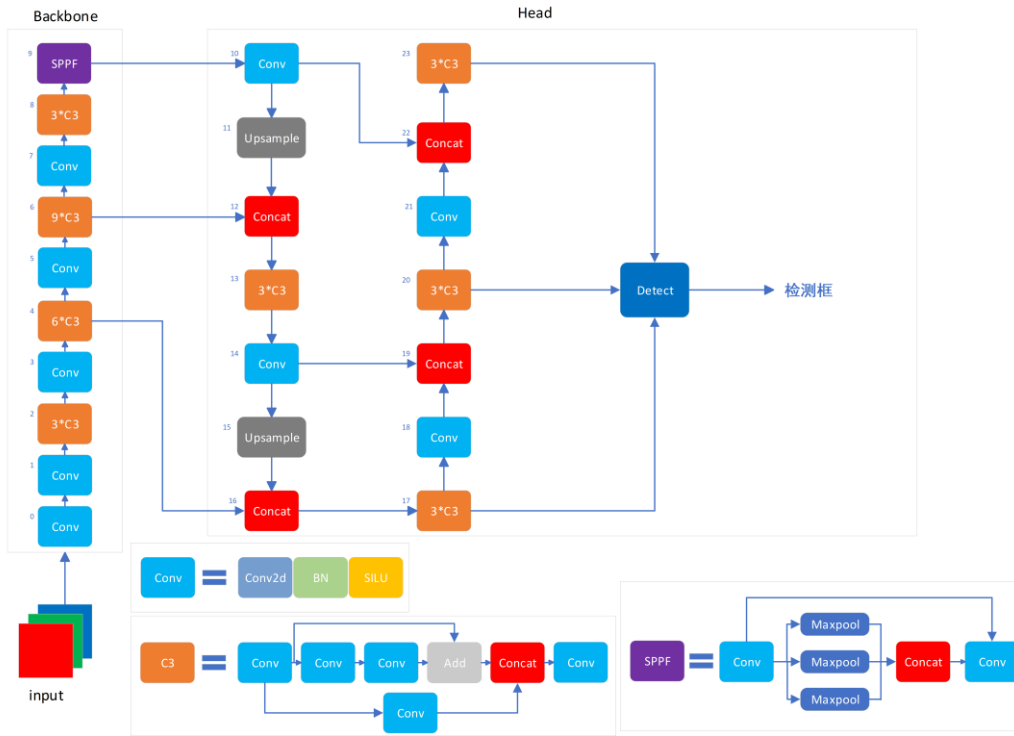


图 2-8 YOLOv5 网络结构

后续 YOLO 算法均是在之前结构上进行改进，YOLO 在实时物体检测逐渐被广泛应用。不过在特殊场景例如复杂背景、微小目标检测的准确性的表现不佳，本文也是根据这些缺陷在 YOLOv5 的基础上进行改进，满足复杂场景下茶叶嫩芽检测。

## 2.4 茶叶嫩芽数据集构建

图像采集和预处理是计算机视觉任务的基础，茶叶嫩芽目标检测模型需要建立在茶叶嫩芽数据集基础上进行训练。在农业场景中，数据集的质量直接影响模型的效果，例如茶叶嫩芽在不同光照条件下呈现不同颜色、茶芽被枝叶遮挡、数据多样性，不同品种、生长阶段的茶叶需要不同的处理，数据采集时需要覆盖这些情况。预处理是根据数据集特性，进行相应的数据转换和处理，从而提高模型的准确性和效率。系统化的图像采集与预处理设计，不仅能为模型提供高质量输入，更重要的是建立农业视觉任务的物理和数字映射桥梁。特别是在农业应用中，现实环境中的挑战使得这些步骤更加关键，这对后续特征学习、模型泛化具有决

定性影响。

#### 2.4.1 图像采集地点

中国作为茶叶原产国，拥有数千年的茶文化积淀，多元的茶类体系构筑了丰富的产业生态。在中国茶业版图中，六安茶始终占据独特地位。作为全国重要农业基地，这里不仅孕育了位列十大名茶的六安瓜片，更在明代《茶疏》中就被誉为“大江以北茶之冠”。依托大别山生态优势，截止至 2024 年六安市茶叶种植面积超过 80 万亩、产量 4.75 万吨、一产产值 55 亿元、综合产值 192.6 亿元。现有的标准化茶叶加工厂 70 家，规上茶叶加工企业 25 家，成为安徽省内第二大茶产区<sup>[53]</sup>。

六安市位于安徽省西部，地处大别山北麓与江淮丘陵过渡带，地理坐标介于北纬 31°06'~32°48'之间，东经 115°22'~116°11'之间，是长江与淮河的分水岭区域。其西南部的金寨县，作为大别山腹地核心产茶区，拥有大量山区和丘陵地带，凭借独特的自然地理位置成为六安茶产业的“黄金心脏”，县域内平均海拔 500 米以上，年均气温 15.2℃，年降水量 1400 毫米，昼夜温差显著，形成了“晴时云雾绕山、雨则溪涧润土”的立体微气候。这种高海拔云雾滋养、多漫射光照的条件，配合富含有机质的酸性土壤<sup>[54]</sup>，显著缓了茶树生长周期，促使茶叶氨基酸含量提升 20%~30%，非常适合茶叶的种植，这使金寨成为我国绿茶地理标志保护的标杆产区。

本文主要选择金寨瓜片为研究对象，六安瓜片是六安市种植面积最大的绿茶，也是六安的代表性茶叶。尤其是核心产区蝙蝠洞一带，火山岩风化土中硒、锌等微量元素的富集，造就了六安瓜片叶片肥厚、兰香高锐、滋味鲜醇的顶级品质。2019 年，六安金寨入选第一批国家农产品质量安全县名单；2020 年，被列入国家园林县城；同年，入选第一批全国农作物病虫害“绿色防控示范县”创建推评名单<sup>[55]</sup>。因此选择在金寨县内以六安瓜片为采集对象，对六安瓜片的茶叶嫩芽进行图像和数据采集工作。图像采集地理位置与环境如图 2-9 所示。

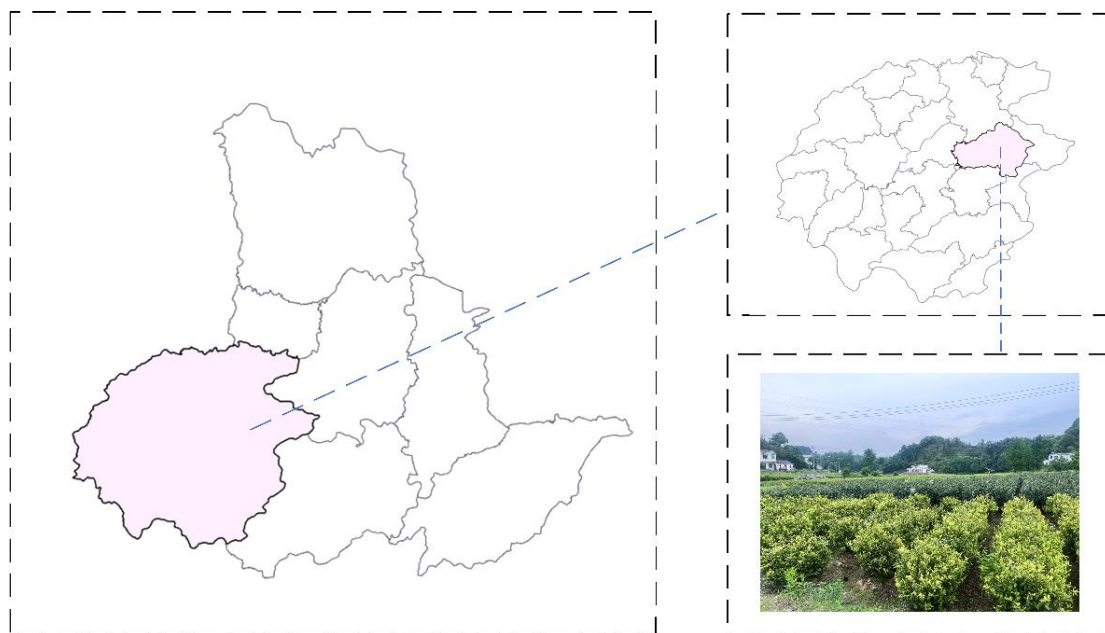


图 2-9 六安市金寨县地理位置与图像采集环境

#### 2.4.2 茶芽图像数据采集

目前网上还未有公开的六安茶叶数据集，在公开的茶叶数据集较少，且茶树生长环境均为地形平整的茶园采摘场景，因此茶树生长环境相似，茶芽特征单一，对复杂环境下的茶芽检测效果差。因此需要实地采集高质量的茶叶嫩芽图像。

本文所采用茶叶嫩芽数据集采集地点为中国安徽省六安市金寨县油坊的西茶谷，采集对象为六安瓜片，一共采集了 2130 张原始图像，采用 JPG 格式进行存储。数据采集时间为春茶生长阶段，拍摄时间为上午 8 点到下午 6 点。在对茶叶嫩芽图像进行采集时，拍摄设备距离茶树距离为 10~50cm，拍摄角度相对于地面夹角保持一定范围，尽可能模仿山区采茶机器的视角进行场景拍摄。为了保证数据集的多样性，提高模型泛化能力和鲁棒性，特意选择在不同拍摄角度、范围、光照亮度和茶芽数目等方面拍摄图像<sup>[56]</sup>，如图 2-10 所示，其中图(a)所展示的为侧面和顶部条件下的拍摄视角度，图(b)展示的为近角和广角条件下的拍摄角度，图(c)为亮和暗条件下的拍摄角度，图(d)为茶叶嫩芽少目标和多目标条件下的拍摄角度。

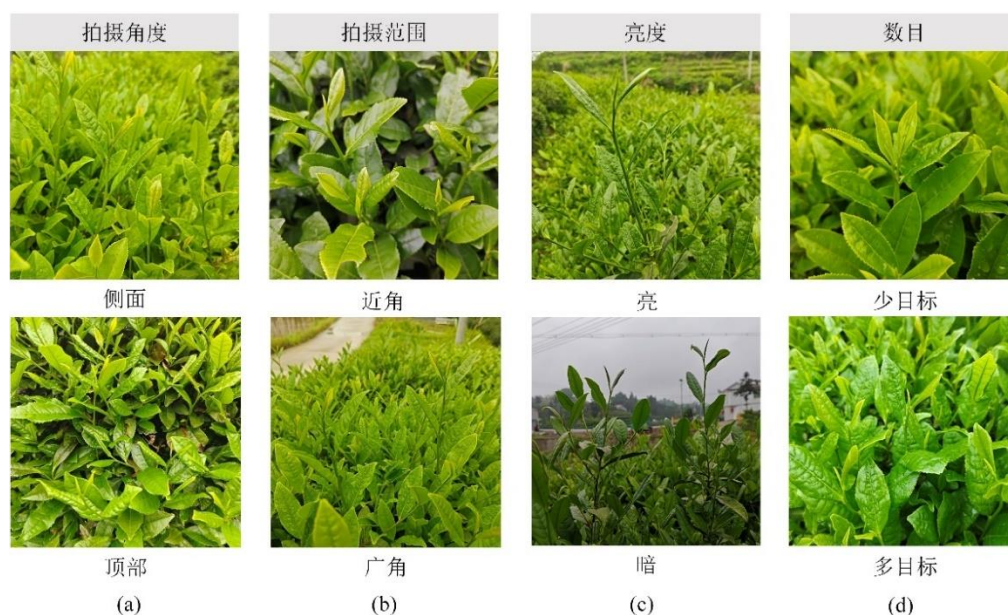


图 2-10 各种条件下的茶叶拍摄图像

### 2.4.3 数据标注

茶叶嫩芽数据集的构建是深度学习模型训练的关键环节，数据集质量的好坏直接影响模型性能与检测精度。由于当前网络上缺乏复杂情况下茶叶嫩芽标准数据集，需要自主构建数据集。采用 LabelImg 开源标注工具，重点针对单芽、一芽一叶和一芽两叶进行标注，同时在标注前建立茶芽目标标准。茶叶生长环境密集，茶芽会与老叶、树枝之间会有遮挡，因此对于被遮挡超过 70% 的茶芽和出现在图像边界框的目标不进行数据标注。为了防止训练后的模型过拟合，对图像种聚焦模糊的目标及低像素区域不进行数据标注。对于出现在图像边缘的图像，如果茶芽采摘点位于图像内就进行标注，如果不在就舍弃这类茶芽目标。标注界面如图 2-11 所示。



图 2-11 茶芽标注界面

每个标注框包含目标名称、类别和归一化后的坐标信息，每个标注框在图像中的种类信息和位置信息可以用五个变量所表示，第一列数字 0 代表目标类别为第一类，第二列为目标中心位置归一化后横坐标  $x$ ，第三列为目标中心位置归一化后纵坐标  $y$ ，第四列为归一化后标注框宽度  $w$ ，第五列为归一化后标注框高度  $h$ 。具体位置信息计算如公式 (2-3)、(2-4) 和 (2-5)，由于本研究采用 YOLO 格式的数据集，因此这些信息最终以 txt 文件格式进行保存，保存形式如 2-12 所示，并将所有目标类别均归为 tea bud。

$$x_c = \frac{x_{\max} + x_{\min}}{2}, y_c = \frac{y_{\max} + y_{\min}}{2} \quad (2-3)$$

$$x = \frac{x_c}{width}, y = \frac{y_c}{height} \quad (2-4)$$

$$w = \frac{x_{\max} - x_{\min}}{width}, h = \frac{y_{\max} - y_{\min}}{height} \quad (2-5)$$

其中， $x_c$  为归一化前的标注框横坐标， $y_c$  为归一化前标注框纵坐标，这两个参数均分别取标注框横纵坐标平均值。 $width$  和  $height$  分别为图形的宽和高。

```
0 0.206543 0.182943 0.127279 0.237630
0 0.458659 0.270833 0.121745 0.179036
0 0.679525 0.545085 0.107747 0.081055
0 0.579590 0.578125 0.118164 0.092448
0 0.935872 0.642253 0.077474 0.139323
```

图 2-12 YOLO 格式数据集保存形式

## 2.5 数据增强

在目标识别网络种，当数据量少时，机器学习和深度学习模型容易出现过拟合、泛化能力差、性能不稳定等问题，这些问题会导致无法准确预测。茶叶采摘时间短暂，而且茶叶嫩芽图像获取代价过大。因此，非常有必要对茶芽数据集进行数据增广，以提升模型的鲁棒性和泛化能力。数据增强分为离线增强和在线增强。离线增强指在模型训练前对原始数据集进行一系列增广操作，其核心优势在于计算资源前置化，通过预先执行耗时操作，避免训练时重复计算，特别适用于计算能力受限的边缘设备<sup>[57]</sup>。例如，在茶叶嫩芽检测中，可模拟强光、晨雾、天



气等环境，批量生成包含不同光照条件、气象干扰的增强样本，并保存为固定数据集。在线增强在训练过程中实时动态生成增强样本，在每轮训练下对同一原始数据施加随机变换，例如 Mosaic、MixUp、CutMix，其优势能够迫使模型学习本质特征而非固定数据特征，在茶芽目标检测场景中，能有效模拟茶叶生长动态变化，提升模型的泛化性<sup>[58]</sup>。

### 2.5.1 离线数据增强方法

#### (1) 水平、垂直翻转图像增强

水平翻转能够模拟自然场景中茶芽左右对称生长特性，符合茶芽生长规律，针对小目标嫩芽，能够使模型关注全局而非局部，无论目标出现在左边还是右边，模型都能正确识别。进行垂直翻转是考虑到枝条弯曲、风向或采茶机的视角不同，出现上下颠倒的情况，这时候垂直翻转可以有助于模型应对这些情况。如图 2-13 所示，(a)为原图，(b)为水平翻转后的图像，(c)为上下反转后的图像。



图 2-13 茶叶水平垂直翻转前后对比图

#### (2) Gamma 变换图像增强

采集数据集中，茶叶嫩芽在不同时间和天气下被拍摄，光照条件差异大。Gamma 变换符合人眼视觉特性，可以用于调整图像的亮度，能够通过非线性变换来增强或减弱图像的对比度。在茶叶嫩芽采摘场景下，使用 Gamma 变换可以生成各种光照情况的图像，帮助模型适应这种变化<sup>[59]</sup>。当 Gamma 常数等于 1 时，图形无变换，对应图 2-13 中(a)部分；当 Gamma 常数小于 1 时，提高亮度，图像整体灰度值变大，增强早上和晚上的嫩芽细节，对应图 2-13 中(b)部分；当 Gamma 常数大于 1 时，降低亮度，图像灰度值变小，增强中午强光下的嫩芽细节，对应图 2-14 中(c)部分。数学公式如式 (2-6) 所示。

$$V_{out} = \left( \frac{V_{in}}{V_{max}} \right)^{\gamma} \times V_{max} \quad (2-6)$$

其中，式（2-6）中  $V_{in}$  为输入像素初始值， $V_{out}$  为变换后的像素值， $V_{max}$  为像素最大值， $\gamma$  为 Gamma 常数。主要就是先对图像进行归一化处理，再应用 Gamma 变换，最后还原像素范围，调整对比度后的效果图如下 2-14 所示。



图 2-14 Gamma 亮度变换前后对比图

### (3) 添加噪声数据增强

在数据集中加入噪声的意义在于提升模型对真实场景干扰的鲁棒性，相较于其它数据增强，加入噪声能够补偿拍摄成像时的电子噪声，模拟识别过程中模拟传感器的缺陷问题；防止训练数据集过于干净的数据分布，防止过拟合，提高模型泛化能力；模拟田间茶叶生长环境多雨雾天气等条件，提高物理环境适配<sup>[60]</sup>。加入高斯噪声用于模拟电子设备传感器和传输过程中的随机误差，表现为图像各像素点的亮度值在真实值附近呈连续、平滑的波动，整体看起来图片的噪杂程度较小，如图 2-15(b)所示；加入椒盐噪声用于模拟雨雾天和信号传输错误的较大误差，表现会有明显的噪杂程度，如图 2-15(c)所示。



图 2-15 添加噪声数据前后对比图

### 2.5.2 在线数据增强方法

在线数据增强是在模型训练期间进行的,增强方式不限于旋转、缩放、剪切、翻转、颜色调整等,最主要的优点是增加了输入数据的多样性,不需要提前生成所有增强数据,使训练流程更高效,特别是在使用 GPU 加速时,数据加载和增强可以并行处理。最终结果能够明显增强模型对不同环境和情景的适应能力,提高最终模型的检测精度。不过在线数据增强会增加训练时计算压力,对计算机算力有较高要求。本文主要采用 HSV 色彩变换、图像缩放和 Mosaic 数据增强。

HSV 色彩变换通过随机调整图像的颜色空间参数实现数据增强。具体而言,算法对输入图像的色调(H)、饱和度(S)和亮度(V)分别施加随机比例的偏移,用于模拟不同光照、季节或环境下的颜色变化。HSV 色彩变换能够提升模型对颜色差异的鲁棒性,使其能适应茶叶嫩芽在不同光照条件下的外观变化<sup>[61]</sup>;图像缩放是将输入图像拉伸或压缩到不同尺寸,并通过填充保持与输入分辨率一致,这种方法能让模型学习不同尺度的目标特征,增强对茶叶嫩芽大小变化的适应能力,同时不会减少像图像裁剪导致的信息丢失;Mosaic 数据增强算法是将四张训练图像随机裁剪、缩放后拼接为一张合成图像,并同步调整标注框的位置,图像内容大了,而且标签信息更多,训练一张图像基本等同于训练四张小图,相比较于 CutMix 两张图像拼接的效果更好。Mosaic 提升了模型对小目标、密集分布目标的检测能力,同时通过批量归一化多图数据,优化训练效率,尤其适用于茶叶嫩芽这类形态多样且分布不规则的场景。

最终,经过筛选图像和离线数据增强后数据集达到 4100 张,并将数据集按照 8:2 划分送入网络进行在线数据增强。为了减少在线数据增强带来训练速度上的影响,本文选择多线程和 GPU 来提升训练效率。

## 2.6 本章小结

在本章节中,首先详细介绍了卷积神经网络的基本构成,包含了卷积层、激活函数、池化层和全连接层。其次通过介绍 LeNet、AlexNet、VGG、ResNet 几个经典网络,讲述了卷积神经网络如何解决实际问题、发展成熟的过程以及 YOLOv5 网络演变的过程。最后阐述了茶叶嫩芽数据制作思路 and 过程,包括采茶地址的选择、图像采集的细节、数据集的扩充级茶叶嫩芽数据的标注。



### 第3章 基于 YOLO-CBN 网络茶芽检测模型

针对复杂环境下，茶芽目标检测精度普遍较低的问题，本文基于 YOLOv5 提出了检测精度更高的 YOLO-CBN 模型。在 YOLOv5 模型主干网络部分添加 CBAM 注意力机制，抑制无效信息，提取有效信息，提高模型抗干扰能力；将原模型中颈部网络结构替换为 BiFPN 结构，增强网络特征表达能力，提高模型检测精度；通过 NWD 损失函数改善茶芽位置偏差敏感的问题，并在公开数据集上进行实验比较。

#### 3.1 YOLO-CBN 网络结构

图 3-1 显示了改进 YOLOv5 的模型结构，将其命名为 YOLO-CBN。其中 C3 结构作用和 CSP 架构基本相同，不过包含了三个标准卷积层，是对残差特征学习的主要模块，由两个分支拼接而成，一部分经过 CBS 堆叠，另一部分仅经过一个基本卷积模块<sup>[62]</sup>。SPPF 模块是在 SPP 模块基础上增加更多的区域尺度和卷积核大小，方便提取更多尺度的信息。

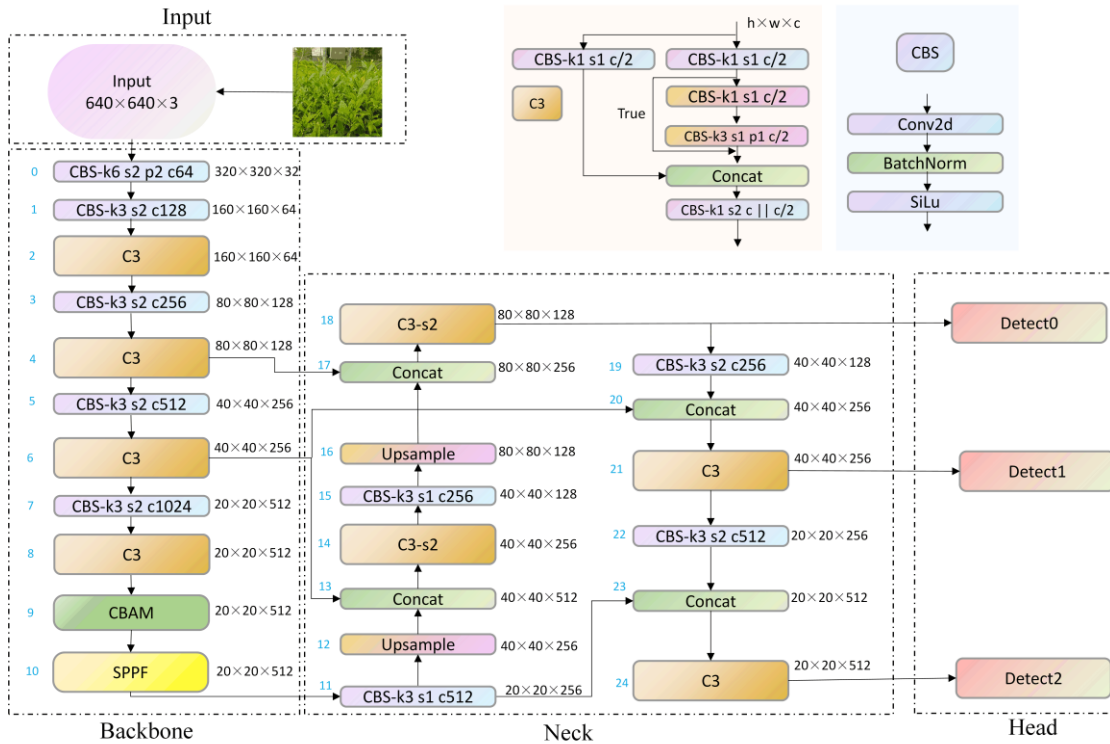


图 3-1 YOLO-CBN 网络结构图

### 3.1.1 CBAM 注意力机制

注意力机制是深度学习中非常重要的技术，在计算机视觉、自然语言处理等领域被广泛应用。通过模拟人类观察事物时的注意力分配方式，帮助模型更加关注输入数据中的重要部分，进而提升模型的性能和效果。核心原理就是在于，并非所有输入信息都对模型的输入结果同等重要，通过引用注意力权重，模型可以学习将注意力集中在关键信息上，并忽略无关信息。目前主流注意力机制主要分为空间注意力机制、混合注意力机制、通道注意力机制和自注意力机制。

通道注意力机制使用平均池化和最大池化将特征图空间信息聚合，然后将它们共享全连接层，处理得到后的信息加和取 Sigmoid 激活函数，此时得到输入特征层每个通道 0~1 的权值。通道注意力机制用于处理图像数据，在处理每个通道时，根据其它通道相关性动态分配注意力权重，用于强调最具信息量的通道<sup>[63]</sup>，对于背景复杂的茶芽检测有很好的效果，如图 3-2 所示。

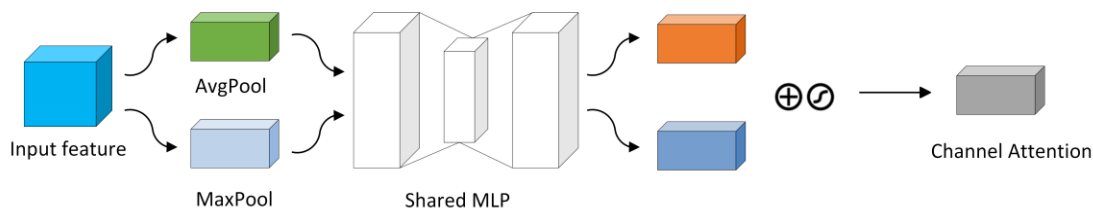


图 3-2 通道注意力模块

空间注意力机制对输入进来的特征层在每个特征点的通道上取最大池化和平均池化，并将两个结果堆叠进行卷积来调整通道数，最后经过 Sigmoid 函数获得输入特征层每一个特征点 0~1 的权值。空间注意力机制在用于处理图像数据，在处理每个像素或图像区域时，根据其它像素或区域信息动态地分配注意力权重，以便集中处理包含重要信息的区域部分<sup>[64]</sup>，对于提高茶芽这种易受遮挡的小目标检测有很大帮助，如图 3-3 所示。

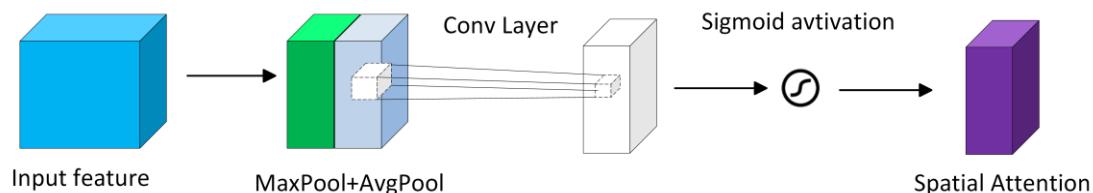


图 3-3 空间注意力模块

对于自注意力机制，通过注意力得分函数来计算序列中每个位置与其他位置

之间的相似度，得到注意力权重，然后对输入序列进行加权求和，可以减少模型对输入信息的过度依赖，但对设备计算能力要求较高，不适合部署到像采茶机这种移动设备上。

本文使用的 CBAM (Convolutional block attention module, CBAM) 是一种混合注意力机制，结合了空间注意力机制和通道注意力机制，并串联形成的注意力模块，在拥有空间和通道注意力优点的同时，对设备要求远低于自注意力机制<sup>[65]</sup>，非常适合复杂背景下的茶叶嫩芽识别，CBAM 结构如图 3-4 所示。

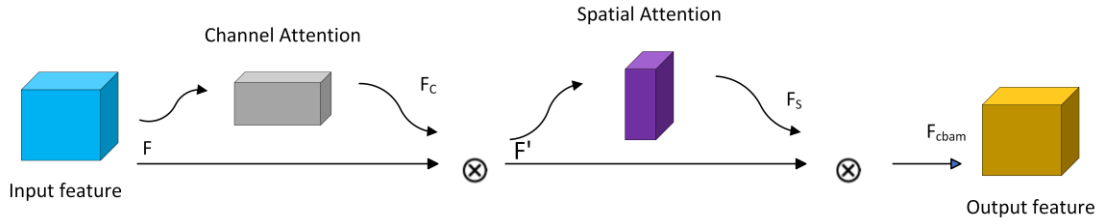


图 3-4 CBAM 注意力机制

### 3.1.2 BiFPN 结构

在目标检测这种计算机视觉任务中，CNN 虽然能够提取特征，但会出现低层的特征图具有高分辨率但缺少语义信息，而高层的特征图具有丰富的语义信息但分辨率低。为了结合不同层次的特征图，引入了特征金字塔网络 FPN，如图 3-5(a)所示。FPN 从自下而上生成特征图序列的顶部开始，通过上采样将分辨率加倍，再与特征图序列中高分辨率低语义的特征图加和，重复上述步骤生成多尺度特征金字塔<sup>[66]</sup>。其中每个特征图都与不同分辨率的输入图像对应，这就是为什么 FPN 能够同时对不同尺度目标进行检测。

深层的特征图虽然有高语义特征，但由于分辨率低，定位信息较弱，PAN (Path Aggregation Network, PAN) 在不同于 FPN 增强多尺度表达，它选择将浅层定位信息传到深层，增强多尺度定位能力，如图 3-5(b)所示。PAN 首先将低分辨率的特征图上采样到与高分辨率特征图相同的分辨率，并将其拼接而不是加和，得到更加丰富的特征图，增强网络位置信息，缺点是会增加计算量。

BiFPN 相比于 PAN 有以下几点改进，如图 3-5(c)所示。在跨尺度连接方面，如果一个节点只有输入而缺少特征融合，那么对于一个特征网络的意义就太小，所以 BiFPN 删除了只有一个输入边的节点；处于同一级的输入节点与输出节点建立新的路径，能够特征融合的同时减少计算成本。跨尺度连接在融合更多特征

的同时，不需要增加太多计算量，对于茶叶嫩芽目标检测模型提升精度的同时，对模型的负荷影响较小。在加权特征融合方面，PAN 网络对不同分辨率图像特征融合时，会将图像转换为相同分辨率，然后进行级联，但不同分辨率输入特征对于输出特征贡献不同。为了避免特征信息的缺失，BiFPN 网络将输入特征乘上不同权重，让网络学习每个输入特征的重要性。特别是和 SPPF 模块共同作用网络，一个用于提取多尺度信息，一个用于多尺度融合，对于网络性能提升有非常显著的作用<sup>[67]</sup>。

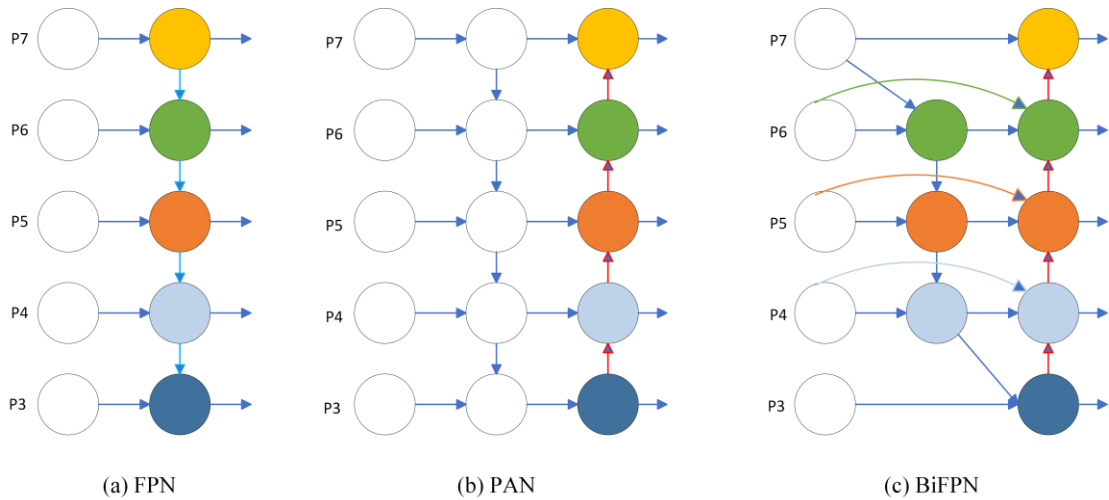


图 3-5 不同特征融合网络结构

采用 BiFPN 结构进行特征融合可以解决以下问题：一是解决了茶叶嫩芽较小的检测问题，有助于茶芽浅层特征的检测；二是解决了茶芽被遮挡的问题，利用跨尺度连接的方式，可以对不同特征按照跨尺度权重进行抑制或增强特征表达，来减少茶芽被遮挡导致识别不出来的问题。

### 3.1.3 归一化 NWD 损失函数

在 YOLO 损失函数中，定位损失通常依赖于交集与并集（Intersection of Union, IoU）来测量预测边界框和实际边界框之间的匹配程度。然而，IoU 对边界框中的微小偏差特别敏感，尤其是对于图像中较小的对象，这可能会导致损失函数计算不稳定。为了解决这些问题，本文引入了 NWD 损失函数。它将边界框建模为高斯分布，并使用 Wasserstein 距离来更准确地测量预测边界框和实际边界框之间的差异<sup>[68]</sup>。这提高了模型对小目标茶芽的检测能力，同时能够更好地处理重叠茶芽，并提供了更稳定和平衡的训练过程，最后通过归一化计算损失。

对于图像中的目标，其边界框中会有除目标外的背景像素，像茶芽目标不是矩形形状，在边界框的非中心处会集中有背景像素。为了区分目标像素和背景像素，对边界框进行二维高斯分布建模，并根据信息重要性分配权重。对用中心处目标像素的权重高，从中心到边界框权重逐渐降低，对于水平边界框  $R=(c_x, c_y, w, h)$ ， $(c_x, c_y)$  为边界框中心坐标， $w$  和  $h$  分别为边界框宽度和高度，椭圆方程为 (3-1)：

$$\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} = 1 \quad (3-1)$$

其中： $(\mu_x, \mu_y)$  为椭圆的中心坐标， $\sigma_x$  和  $\sigma_y$  分别为椭圆沿  $x$  轴和沿  $y$  轴长度。可以推出  $\mu_x = cx$ ， $\mu_y = cy$ ， $\sigma_x = \frac{w}{2}$ ， $\sigma_y = \frac{h}{2}$ ，二维高斯分布的密度函数如式 (3-2)：

$$f(x|\mu, \Sigma) = \frac{\exp(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu))}{2\pi |\Sigma|^{\frac{1}{2}}} \quad (3-2)$$

其中： $x$ 、 $\mu$  和  $\Sigma$  分别为高斯分布坐标  $(x, y)$ 、均值向量和协方差矩阵。当满足  $(x-\mu)^T \Sigma^{-1}(x-\mu) = 1$  时，式 (3-1) 表示的椭圆方程为是二维的高斯分布密度等值线。针对边界框  $R=(c_x, c_y, w, h)$  建模为二维高斯分布  $N(\mu, \Sigma)$ ，如式 (3-3)：

$$\mu = \begin{bmatrix} c_x \\ c_y \end{bmatrix}, \Sigma = \begin{bmatrix} \frac{w^2}{4} & 0 \\ 0 & \frac{h^2}{4} \end{bmatrix} \quad (3-3)$$

因此，可以通过 Wasserstein 距离判断两个二维高斯分布相似度，进一步判断这两个边界框的相似度。对于两个二维高斯分布  $\mu_1 = N(m_1 + \Sigma_1)$  和  $\mu_2 = N(m_2 + \Sigma_2)$  之间的二位 Wasserstein 距离定义为式 (3-4)：

$$W_2^2(\mu_1, \mu_2) = m_1 - m_2^2 + \text{Tr}\left(\Sigma_1 + \Sigma_2 - 2(\Sigma_1^{1/2} \Sigma_2 \Sigma_1^{1/2})^{1/2}\right) \quad (3-4)$$

进一步简化为式 (3-5)：

$$W_2^2(\mu_1, \mu_2) = \|m_1 - m_2\|_2^2 + \|\Sigma_1^{\frac{1}{2}} - \Sigma_2^{\frac{1}{2}}\|_F^2 \quad (3-5)$$

其中： $\|\cdot\|_F$  为 Frobenius 范数。

因此对于两个边界框  $A=(cx_a, cy_a, w_a, h_a)$  和  $B=(cx_b, cy_b, w_b, h_b)$  的二维高斯分布  $N_a$  和  $N_b$  的二阶 Wasserstein 距离为式 (3-6)：

$$W_2^2(N_a, N_b) = \left\| \left[ cx_a, cy_a, \frac{\omega_a}{2}, \frac{h_a}{2} \right]^T, \left[ cx_b, cy_b, \frac{\omega_b}{2}, \frac{h_b}{2} \right]^T \right\|_2^2 \quad (3-6)$$

为了将  $W_2^2(N_a, N_b)$  的距离度量转化为相似度量，将结果限制在 0~1 之间，将其以指数方式归一化处理，如式 (3-7)，并统称为归一化 Wasserstein 距离。

$$NWD(N_a, N_b) = \exp \left( -\frac{\sqrt{W_2^2(N_a, N_b)}}{C} \right) \quad (3-7)$$

其中：C 为数据密切相关的常数。

将模型预测最终得到边界框高斯分布模型为  $N_p$ ，标签高斯分布模型为  $N_g$ ，最终归一化 Wasserstein 距离损失函数表示为式 (3-8)：

$$L_{NWD} = 1 - NWD(N_p, N_g) \quad (3-8)$$

将 NWD 损失与 IoU 相结合的好处在于综合考虑两者的优势。NWD 损失和 IoU 损失分别关注边界框的位置和形状。因此，综合考虑可以更全面地评价边界框的匹配度，从而提高目标检测的准确率。IoU 损失对边界框的位置和形状很敏感，而 NWD 损失可以在一定程度上提供更稳定的匹配度量，即使边界框位置有轻微的偏差，也可以通过 NWD 损失进行校正。通过修改损失函数，该模型有效提高了茶芽目标的检测精度。

## 3.2 结果分析

### 3.2.1 实验评价指标

在目标检测任务中，预测框与真实框的位置决定了对其评价，一共可以分成以下四种情况：实际为正类，预测也为正类（True Positive, TP）；实际为负类，预测为正类（False Positive, FP）；实际为正类，预测为负类（False Negative, FN）；实际为负类，预测为负类（True Negative, TN）。当真实框与预测框的交集比上真实框与预测框的并集是否大于设定的阈值来判定预测框是否归为正样本，否者就为负样本。

在实验过程中，评价指标是衡量模型性能的依据。常用的评价指标包含评准确率（Accuracy）、精确率（Precision, P）、和召回率（Recall, R）、平均精度（Average Precision, AP）、平均准确率（Mean Average Precision, mAP）以及帧

数 (Processing time per frame, FPS) 等。

其中, 精确率是指模型正确预测样本数量占总样本数量的比例。当精确率越大时, 说明预测结果误检的越少, 计算公式如 (3-9) :

$$P = \frac{TP}{TP + FP} \quad (3-9)$$

召回率则反映了真实为正的样本中被正确预测出来的比例。当召回率越大, 说明预测结果漏检率越少, 计算公式如 (3-10) :

$$R = \frac{TP}{TP + FN} \quad (3-10)$$

AP 值是计算某一类别 PR 曲线与坐标轴围成的面积, 如式 (3-11) :

$$AP = \int_0^1 P(R) dR \quad (3-11)$$

mAP 是针对多类别 AP 求和求平均, 是综合考虑精确率和召回率的评价指标, 适合全面评价一个模型检测效果的好坏, 由于被检测对象只有茶芽一类, 故类别 N 为 1, AP 等于 mAP, 如式 (3-12) :

$$mAP = \frac{\sum_1^N AP}{N} = \frac{\sum_1^N \int_0^1 P(R) dR}{N} \quad (3-12)$$

而 FPS 是衡量图像处理或模型推断熟读的指标, 表示一秒内处理图像的帧数, 本文 FPS 计算指标如式 (3-13) :

$$FPS = \frac{1000ms}{(pre\_process + inference + NMS\_per)} \quad (3-13)$$

其中: pre\_process 是图像预处理时间, 包括升维、降维、padding 填充、通道变换等; inference 是推理速度, 指经过预处理后, 图像输入模型到输出模型的时间; NMS\_per 是非极大值抑制处理, 用于对输出模型结果进行转换。

### 3.2.2 实验环境配置

本章为验证改进效果, 在公开数据集上进行实验, 训练前均不加载预训练权重, 并采用相同配置超参数, 确保实验结果的可靠性。实验全部内容均在同一台计算机下完成, 采用的是 PyCharm 作为集成开发环境, 表 3-1 列出实验所用计算机硬件配置, 及软件开发环境。网络训练图像的输入大小为  $640 \times 640$ , 训练过程循环 300 个 epoch, batchsize 设置为 32, 初始学习率设置为 0.01。

表 3-1 计算机软硬件参数信息

| 名称   | 配置                                |
|------|-----------------------------------|
| CPU  | AMD EPYC 7H12 64-Core Processor   |
| GPU  | NVIDIA GeForce RTX 3090           |
| 操作系统 | Windows 10                        |
| RAM  | 512 GB                            |
| 开发环境 | Python3.8, pytorch1.9.0, CUDA11.1 |

### 3.3.3 消融对比实验

为了验证改进后的 YOLOv5 网络模型对茶芽目标检测的有效性，将上述每个改进模块分别进行消融对比，结果如表 3-2 所示。

表 3-2 消融实验结果对比

| Algorithm             | FLOPs(G) | P(%) | R(%) | mAP(%) |
|-----------------------|----------|------|------|--------|
| YOLOv5                | 15.8     | 70.6 | 61.2 | 67.5   |
| YOLOv5_CBAM           | 15.8     | 68.9 | 63.9 | 68.7   |
| YOLOv5_BiFPN          | 16.4     | 71.1 | 64.2 | 69.3   |
| YOLOv5_NWD            | 15.8     | 73.5 | 60.3 | 68.1   |
| YOLOv5_CBAM_BiFPN     | 16.4     | 72.7 | 62.3 | 68.6   |
| YOLOv5_NWD_CBAM       | 15.8     | 74.8 | 62.3 | 69.1   |
| YOLOv5_NWD_BiFPN      | 16.4     | 67.8 | 64.9 | 68.7   |
| YOLOv5_NWD_CBAM_BiFPN | 16.4     | 76.3 | 61.0 | 70.0   |

从表 3-2 中可以看出，相比于 YOLOv5 网络，单独在 CBAM、BiFPN 和 NWD 基础上改进都使模型的 mAP 值增大。每两个改进作用在同一个 YOLOv5 上时，比单独加入初始网络模型上提升更大。当三个改进同时作用于原始模型时，精确率和 mAP 的值最高，此时模型的 GFLOPs 增大了 3.8%，这是由于归一化 NWD 损失函数计算相较之前更为复杂，使模型的计算略微增大。改进后的 YOLO-CBN 网络精确率提高了 5.7%，召回率降低了 0.2%，最终 mAP 的值提高了 2.5%，图 3-6 为公开数据集下验证集随机图像检测结果，由结果可以看出 YOLO-CBN 对遮挡的茶芽检测效果更好，同时对漏检和误检也有改善。





图 3-6 改进前后实验对比

### 3.2.4 注意力模块选择及嵌入位置

注意力机制种类多样,在检测不同目标时,不同注意力机制带来的效果会有很大不同,对于茶叶嫩芽这种小目标,本文选用主流的 CBAM、SE (Squeeze-and-Excitation Networks) 和 SimAM (A Simple, Parameter-Free Attention Module) 三种不同的注意力模块,在相同配置文件和超参数下嵌入到 YOLOv5 模型的相同位置,进行实验对比。实验结果如表 3-3 所示,SE 模块和 SimAM 模块,虽然精确率都有一定幅度的增加,但 recall 会大幅下降, mAP 评价指标都低于原始模型。只有 CBAM 模块嵌入到原始模型时, mAP 的提升最明显。故选择 CBAM 注意力机制嵌入到 YOLOv5 模型中。

表 3-3 不同注意力机制的检测实验结果对比

| Algorithm    | P(%) | R(%) | mAP(%) |
|--------------|------|------|--------|
| YOLOv5       | 70.6 | 61.2 | 67.5   |
| YOLOv5-CBAM  | 68.9 | 63.9 | 68.7   |
| YOLOv5-SE    | 72.8 | 60.1 | 67.3   |
| YOLOv5-SimAM | 80.4 | 55.2 | 67.1   |

CBAM 注意力机制是一种即插即用的模块,可以直接嵌入到网络框架的任何部位中去,但嵌入不同位置对模型的检测效果不同,为了确定 CBAM 嵌入的最佳位置,在 YOLOv5 网络结构的不同位置分别加入 CBAM,使用相同配置文件和超参数进行实验。实验结果如表 3-4 所示,只有 YOLOv5-1 的精确率、召回率和 mAP 相比于原始模型全部有提升。至于为什么没有选择多层引用注意力机

制导致，这是由于添加多个 CBAM 增大模型参数，对模型提升不大，难以从输入数据中学习更多的关键目标信息。反而选择单个 CBAM 更容易捕捉到目标的关键特征。综合考虑选择将 CBAM 模块加入到 YOLOv5 模型的 SPPF 模块前。

表 3-4 CBAM 注意力机制嵌入网络中不同位置检测实验结果对比

| Algorithm | P(%) | R(%) | mAP(%) |
|-----------|------|------|--------|
| YOLOv5    | 70.6 | 61.2 | 67.5   |
| YOLOv5-1  | 68.9 | 63.9 | 68.7   |
| YOLOv5-2  | 70.2 | 59.2 | 67.2   |
| YOLOv5-3  | 74.5 | 60.5 | 68.3   |
| YOLOv5-4  | 68.4 | 62.9 | 68.0   |

其中：YOLOv5-1 表示 CBAM 嵌入到 SPPF 模块前；YOLOv5-2 表示 CBAM 嵌入到主干网络 C3 模块后；YOLOv5-3 表示 CBAM 嵌入到检测头之前；YOLOv5-4 表示 CBAM 嵌入到颈部网络 C3 模块后。

### 3.2.5 改进损失函数前后实验对比

为了分析损失函数对模型提升，将 YOLOv5 初始损失函数 CIOU 替换为归一化 NWD 损失函数，使用这两个损失函数训练模型时的 mAP 曲线如图 3-7 所示。使用归一化 NWD 损失函数的收敛速度快于 CIOU，对两种损失函数训练时的 mAP 曲线析发现，归一化 NWD 损失函数相比较于 CIOU 的平均准确率有明显提高如图。归一化 NWD 损失函数的网络模型的准确率和 mAP 分别提高了 2.9%和 0.6%，实验表明，使用 NWD 损失函数能够改善茶芽偏差的问题。

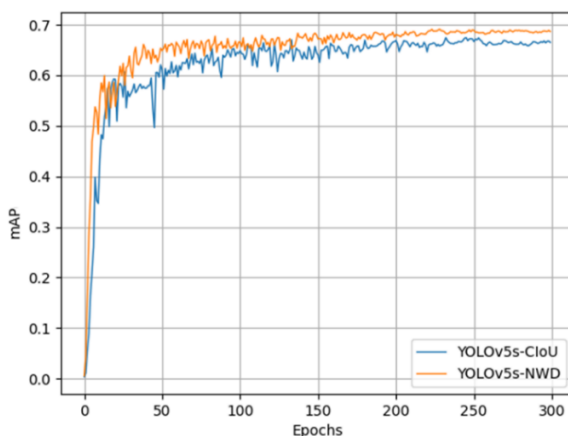


图 3-7 CIOU 和 NWD 损失函数 mAP 曲线图

### 3.3 本章小结

本章主要介绍了改进 YOLOv5 算法对茶芽的检测。针对茶芽检测中目标茶芽存在精确度低，容易漏检误检的问题，提出了一种 YOLO-CBN 的高性能茶叶嫩芽检测模型。首先在改进后的模型在主干网络 SPPF 模块前加入 CBAM 注意力机制增强网络特征表示，提高对茶叶嫩芽目标检测的精度和泛化能力。其次在颈部网络中，使用 BiFPN 结构来代替原有的结构，聚合不同分辨率的特征，增加模型的多尺度融合。最后使用归一化 Wasserstein 距离结合的损失函数，改善 IoU 指标对茶芽的位置偏差敏感的问题，提高对茶芽目标检测的鲁棒性。实验结果表明，改进后的模型相比于原始模型，精确率提高了 5.7%，平均准确率提高了 2.5%，有效减少了茶叶嫩芽的误检现象，在智能化采摘的检测任务中具有显著优势。

## 第4章 基于 YOLO-Tea 的轻量化茶芽检测模型

从上一章实验结果分析可知, 基于 YOLOv5 的改进网络在准确率上有了一定提高, 但对于需要更高准确率、召回率, 同时又方便部署到移动机械设备上的茶芽检测模型, 难以满足现阶段的要求。针对这一些问题, 本章在第3章的网络模型基础进一步改进, 保证模型轻量化的同时, 进一步提高模型准确率和召回率, 以满足实际场景下实时性检测。改进后的模型 YOLO-Tea 有以下几点改进。首先, 对网络上公开的茶芽数据集替换为第2章自制数据集, 其中包含了多种复杂情况下的茶叶图像, 数据质量和标注情况更准确, 对提高模型泛化能力有重要作用。其次, 在主干网络中引入更轻量化的 SimSPPF (Simplified Spatial Pyramid Pooling Fast) 模块, 通过替换激活函数的方式, 有效缓解训练过程梯度消失的问题。然后, 复用第三章颈部网络中的 BiFPN, 通过聚合不同分辨率下的特征, 来增强模型的多尺度融合能力。最后, 将颈部网络中传统的卷积模块替换为 ODConv (Omni-Dimensional Dynamic Convolution) 卷积模块, 使模型关注更多维度的信息, 进一步提高小目标茶芽检测精度。

### 4.1 YOLO-Tea 网络结构

根据残差结构的数量、模型的深度和卷积通道宽度, YOLOv5 主要分为 YOLOv5s、YOLOv5m、YOLOv5l 和 YOLOv5x 四个版本。将第2章构建的数据集用于以上四个版本, 来检测不同版本性能, 方便选择适合研究内容的版本参数, 测试结果如表 4-1 所示。其中 YOLOv5s 的 Parameters 和 FLOPs 最小, mAP 相对于其它版本没有差距太大; YOLOv5x 和 YOLOv5l 虽然 mAP 最高, 但 Parameters 和 FLOPs 远超其它版本, 不利于部署到采茶机器设备上去<sup>[69]</sup>, 而 YOLOv5m 的 parameters 是 YOLOv5s 的三倍, 但 mAP 却只提高 1.7%。为了平衡模型大小和精度, 本研究选择模型深度为 0.33, 卷积通道宽度为 0.5 的 YOLOv5s 为基础模型, 在能够快速推理和保证精度的同时, 方便部署到移动设备上去, 后续模型的改进也将在 YOLOv5s 的基础上。

表 4-1 不同型号 YOLOv5 性能及参数

| Algorithm | Depth multiple | Width multiple | Parameters(M) | FLOPs(G) | mAP (%) |
|-----------|----------------|----------------|---------------|----------|---------|
| YOLOv5s   | 0.33           | 0.50           | 7.01          | 15.8     | 80.5    |
| YOLOv5m   | 0.67           | 0.75           | 20.85         | 47.9     | 82.2    |
| YOLOv5l   | 1.0            | 1.0            | 46.11         | 107.6    | 84.2    |
| YOLOv5x   | 1.33           | 1.33           | 86.17         | 203.8    | 84.9    |

为了保证改进后模型 Parameters 和 FLOPs 较小变化的同时，提高模型精度指标、降低模型大小，对这些问题做出进一步改进，使模型对聚焦模糊茶芽和小目标茶芽有更好的检测效果，图 4-1 显示了改进后的 YOLO-Tea 的网络结构。

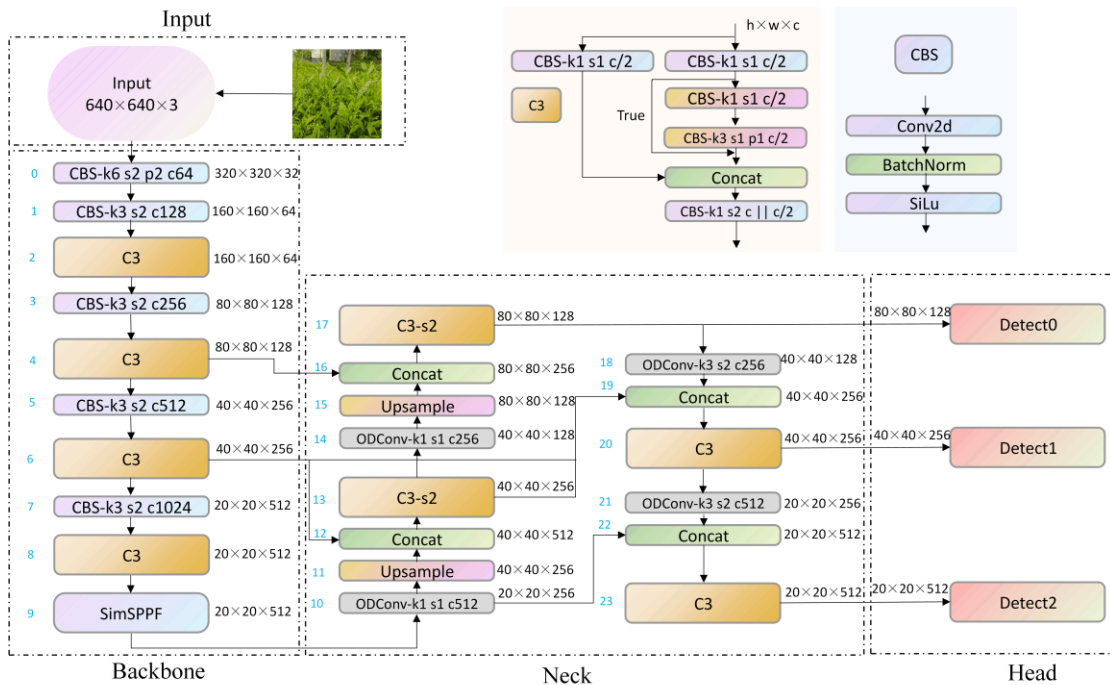


图 4-1 YOLO-Tea 网络结构

#### 4.1.1 SimSPPF 模块

He 等人<sup>[70]</sup>提出 SPP (Spatial Pyramid Pooling, SPP) 模块，在模块中将三个最大池化核并联方式实现局部特征和全局特征的融合，这种方法可以避免重复计算卷积特征，提高生成候选框的速度，来节省计算成本。在 SPP 模块的基础上，YOLOv5 提出 SPPF，主要区别在于将 SPP 中池化核并联的方式变为串联，这种方式通过多层池化的方式极大减少模型计算量，还能够提高模型速度<sup>[71]</sup>。本文所使用的 SimSPPF 相较于 SPPF 模块主要区别在于模块内部使用的激活函数。结构如图 4-2 所示，SPPF 所采用的激活函数为 Silu 激活函数，而 SimSPPF 采用为

Relu 激活函数，由于只采用池化核为 5 的二维最大池化，相比于传统模块的 5、9 和 13 的池化核最大池化，可以减少计算量，同时通过并联的残差结构，将池化后的部分特征进行拼接，并继续进行下一步池化，进而得到不同层次的特征，从而提高模型对输入图像特征提取能力。

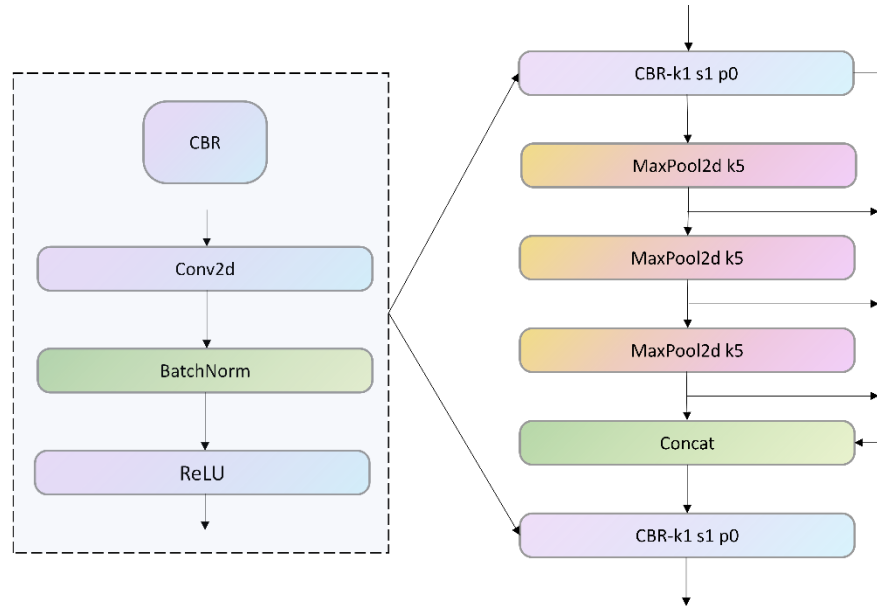


图 4-2 SimSPPF 网络结构图

为了验证 SimSPPF 的引入对茶芽检测有更好的优势，对 YOLOv5 和替换为 SimSPPF 模块后的 YOLOv5-SimSPPF 网络进行对比实验，结果如表 4-2 所示，替换 SimSPPF 模块后，模型的 FPS 和精度均有提高，并且参数量基本保持不变。

表 4-2 加入 SimSPPF 模块实验结果

| Algorithm      | Parameters(M) | FLOPs(G) | Speed(FPS) | Size(M) | mAP (%) |
|----------------|---------------|----------|------------|---------|---------|
| YOLOv5         | 7.01          | 15.8     | 42.92      | 14.5    | 0.805   |
| YOLOv5-SimSPPF | 7.01          | 15.8     | 43.86      | 14.5    | 0.818   |

#### 4.1.2 ODConv 卷积模块

在深度学习领域中，为了提升网络模型表现，一般通过扩展网络宽度、深度、分辨率等增大模型的方法，而且在卷积时使用的卷积核都与输入数据集样本无关，这无疑对网络收敛性能造成极大阻碍。作者 Yang 首次提出动态卷积 (Conditionally parameterized convolutions, Condconv) [72]，区别于特征图上的加权方式，他对多个卷积核进行线性加权，使权值与输入有关，提高动态卷积对输入数据的依赖性。简而言之，就是对不同输入使用不同卷积核，再对不同卷积核



进行注意力加权。

CondConv 卷积依赖使用卷积核个数来赋予卷积核动态属性, 为了避免上述的问题, 在模型中使用 ODConv, 它从空间大小、输入通道、输出通道和卷积核四个不同维度的权重进行学习, 来提高网络模型的识别性能<sup>[73]</sup>, 如图 4-3 所示。图(a)是 ODConv 沿空间维度的卷积注意力运算,  $a_{si}$  对卷积核  $W_i$  的每个空间维度赋予注意力权重, 使模型更加关注茶叶图像的空间位置信息; 图(b)是 ODConv 沿输入通道维度的卷积注意力运算,  $a_{ci}$  对卷积核  $W_i$  每个输入通道的卷积滤波器赋予不同的注意力权重, 让模型更加专注不同通道之间的特征关系, 使模型更好地提取信息和组合信息, 进一步提高模型对茶叶图像中的特定颜色通道或纹理特征的关注; 图(c)是 ODConv 沿输出通道维度的卷积注意力运算,  $a_{fi}$  对卷积核  $W_i$  每个输出通道的卷积滤波器赋予不同的注意力权重, 让模型在不同输出通道之间进行更有针对性的进行特征选择, 从而增强模型对于嫩芽特征提取能力; 图(d)是 ODConv 沿卷积核维度的卷积注意力运算,  $a_{wi}$  对整个卷积核赋予不同的注意力权重, 使模型可以全局地调节卷积操作, 从而能够更好地捕获输入数据的整体特征。四种不同注意力机制相互补充, 使模型在不同维度上对各个输入数据产生不同维度的影响, 从而更好地捕获数据的关键特征和上下文信息, 提高模型对细节的捕捉能力和模型整体的泛化能力。ODConv 对动态卷积的定义如公式 (4-1) 所示。

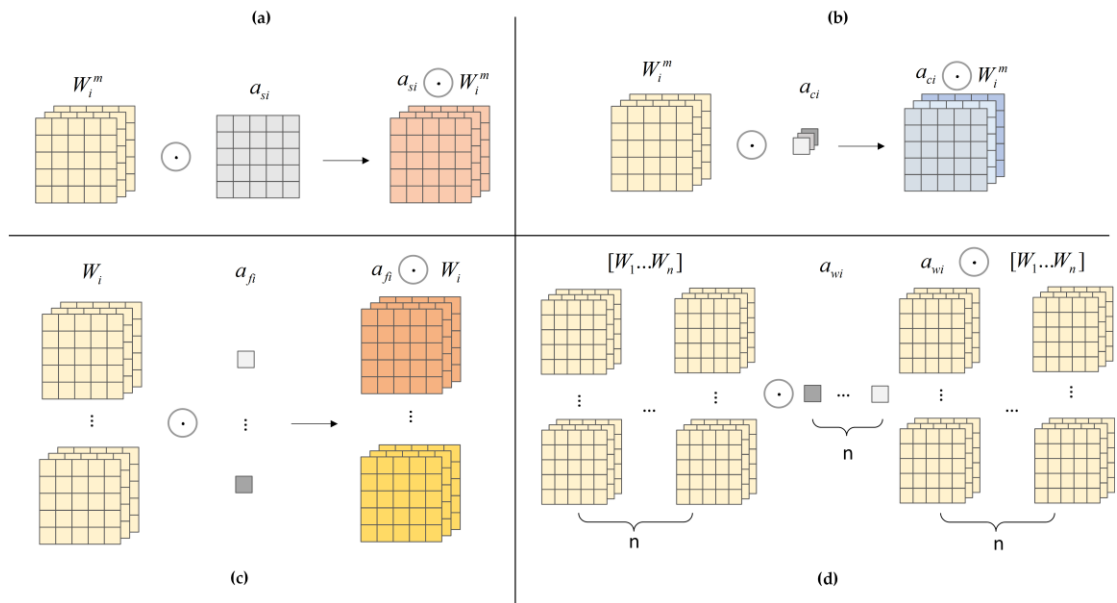


图 4-3 ODConv 四种不同注意力权重作用机制

$$y = (a_{w1} \odot a_{f1} \odot a_{c1} \odot a_{s1} \odot W_1 + \dots + a_{wn} \odot a_{fn} \odot a_{cn} \odot a_{sn} \odot W_n) * x \quad (4-1)$$

ODConv 卷积结构图如图 4-4 所示，输入特征先经过 GAP (Global Average Pooling) 层中进行处理，送入全连接层和激活函数 Relu 层，再由全连接层和 Sigmoid 层之后生成的输入送入图中四个不同权重进行处理。经过这一系列处理后，ODConv 能够提高茶芽特征提取多样化，极大增强网络特征提取能力。选择 GAP 层处理输入特征是由于，GAP 层没有参数用于优化，可以减少或避免过拟合，虽然在捕捉空间能力上受到妨碍，但平均池化对空间信息进行求和，对输入的空间变化的鲁棒性更强，适用于复杂的茶芽采摘场景。

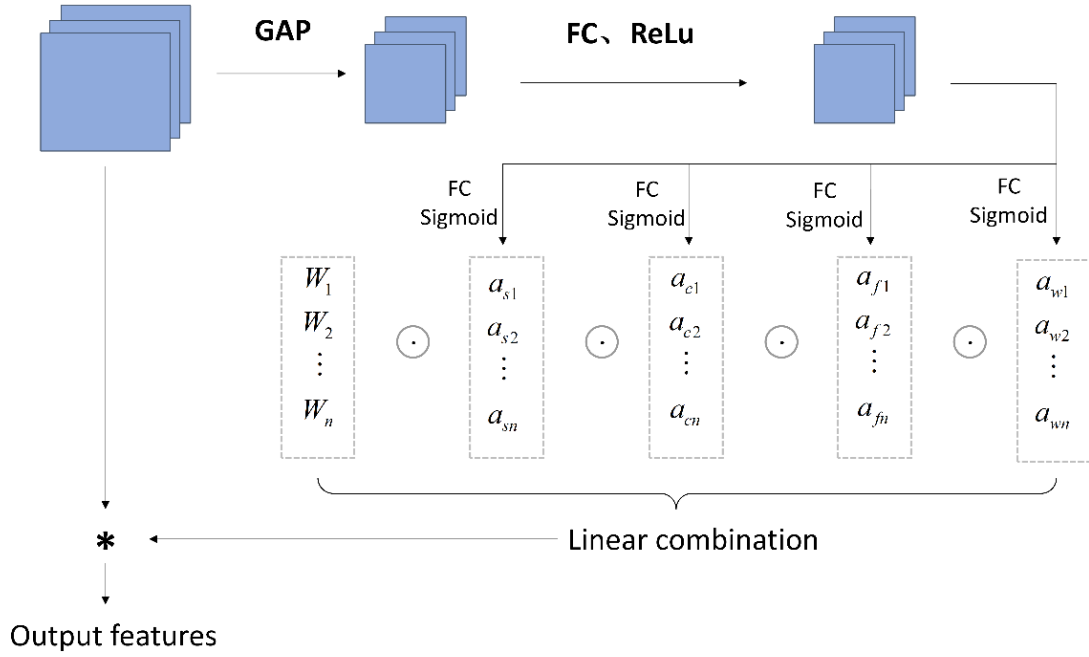


图 4-4 ODConv 卷积结构

对于茶叶嫩芽检测，初始的 CSB 卷积模块更关注浅层边缘特征信息，它也是区分茶芽的关键因素，而 ODConv 更侧重于茶芽纹理、颜色等更深度特征，因此在模型中应取两种模块特征提取的结合，选择在主干网络部分使用初始的 CSB 卷积模块进行浅层特征提取，在颈部网络部分使用 ODConv 卷积，结合使用第 3 章的 BiFPN 结构进行特征融合，兼顾了不同阶段的特征，对茶芽检测精度提高有明显作用。如表 4-3 所示，YOLOv5-all 表示 YOLOv5 模型中独立的 CSB 卷积模块全部替换为 ODConv。而 YOLOv5-ODConv 表示只替换颈部网络中的独立 CSB 模块。对比两数据可以发现，将网络中的全部 CSB 模块替换为 ODConv，会忽略太多了浅层网络的信息，导致检测精度会下降，因此对原始网络改进时，本文选择以 YOLOv5-ODConv 的方式进行改进。



表 4-3 加入 ODConv 模块实验结果

| Algorithm     | Parameters(M) | FLOPs(G) | Size(M) | Speed(FPS) | mAP (%) |
|---------------|---------------|----------|---------|------------|---------|
| YOLOv5-all    | 7.12          | 11.0     | 14.7    | 31.65      | 81.7    |
| YOLOv5-ODConv | 7.09          | 14.7     | 14.6    | 33.90      | 82.4    |

## 4.2 实验结果与分析

### 4.2.1 实验细节

实验软硬件配置环境与第 3 章环境相同。将输入图片像素缩放至网络指定好的  $640 \times 640$ ，采用 SGD 优化器，初始学习率设置为 0.01，动量设置为 0.937，采用余弦退火调度器逐渐降低学习率，帮助模型稳定收敛，提高训练效率，衰减条件如公式 (4-2)。同时，为了保证实验结果的准确性，YOLO-Tea 及对比网络均未载入预训练权重。

$$lr_{new} = lr_{initial} * ((1 + \cos(\frac{epoch_{cur}}{epoch_t} * \pi)) / 2) \quad (4-2)$$

其中： $lr_{new}$  为新的学习率， $lr_{initial}$  为初始学习率， $epoch_{cur}$  为当前训练到的多少轮值， $epoch_t$  为一共训练的轮数。

### 4.2.2 消融对比实验

为了验证改进后 YOLO-Tea 网络模型对茶芽目标检测的有效性，对各模块进行消融对比，结果如表 4-4 所示。从表中可以看出，与 YOLOv5 模型相比较，单独在初始模型的基础上替换 SimSPPF、BiFPN 和 ODConv 都使模型的 mAP 有明显提高。将不同改进以不同方式组合后，评估指标达到最高，可以看出本文 YOLO-Tea 相比于原始模型在 FLOPs 下降的同时，评价指标 P、R、mAP 分别提高了 4.4%、2.3%和 3.2%。从消融对比实验结果来看，与原始模型相比，本研究基于 YOLOv5s 模型的改进都有一定提升，图 4-5 数据集中随机图像检测结果。

表 4-4 消融实验对比结果

| Algorithm             | FLOPs(G) | P(%) | R(%) | mAP(%) |
|-----------------------|----------|------|------|--------|
| YOLOv5                | 15.8     | 80.1 | 71.8 | 80.5   |
| YOLOv5_SimSPPF        | 15.8     | 80.6 | 74.4 | 81.8   |
| YOLOv5_BiFPN          | 16.4     | 81.4 | 73.4 | 82.4   |
| YOLOv5_ODConv         | 14.7     | 83.0 | 71.3 | 82.4   |
| YOLOv5_SimSPPF_BiFPN  | 16.4     | 84.1 | 72.6 | 82.6   |
| YOLOv5_SimSPPF_ODConv | 14.7     | 83.1 | 72.8 | 82.6   |
| YOLOv5_BiFPN_ODConv   | 14.8     | 84.1 | 73.9 | 82.7   |
| YOLO-Tea              | 14.8     | 84.5 | 74.1 | 83.7   |



图 4-5 改进前后实验对比

#### 4.2.3 与其他目标检测网络对比

为了验证基于 YOLOv5 改进的 YOLOv5-Tea 模型的优越性，与其他主流算法模型进行性能指标比较，鉴于两阶段算法网络框架模型太大，既不方便部署到轻量化设备，也达不到实时检测的效果，本文主要对比单阶段的 YOLO 算法系列，对比模型选择 YOLOv5、YOLOv6、YOLOv7 和 YOLOv8，结果如表 4-5 所示。

表 4-5 不同算法实验结果对比

| Algorithm | Parameters(M) | FLOPs(G) | Size (M) | mAP(%) |
|-----------|---------------|----------|----------|--------|
| YOLOv5    | 7.01          | 15.8     | 14.5     | 80.5   |
| YOLOv6    | 18.50         | 45.17    | 38.7     | 80.2   |
| YOLOv7    | 9.32          | 26.7     | 19.0     | 81.1   |
| YOLOv8    | 11.13         | 28.4     | 22.5     | 81.6   |
| YOLO-Tea  | 7.24          | 14.8     | 14.9     | 83.7   |

从结果上看 YOLO-Tea 拥有最高检测精度，模型计算量最小，模型参数量和大小仅高于 YOLOv5 模型，方便部署到移动设备上。图 4-6 为了模型训练过程中指标随训练轮数变化，代表红色线条的 YOLO-Tea 在模型收敛后 mAP 明显高于其它模型，而且收敛速度也排在中间位置。说明通过本文改动，模型多项指标均有改善。

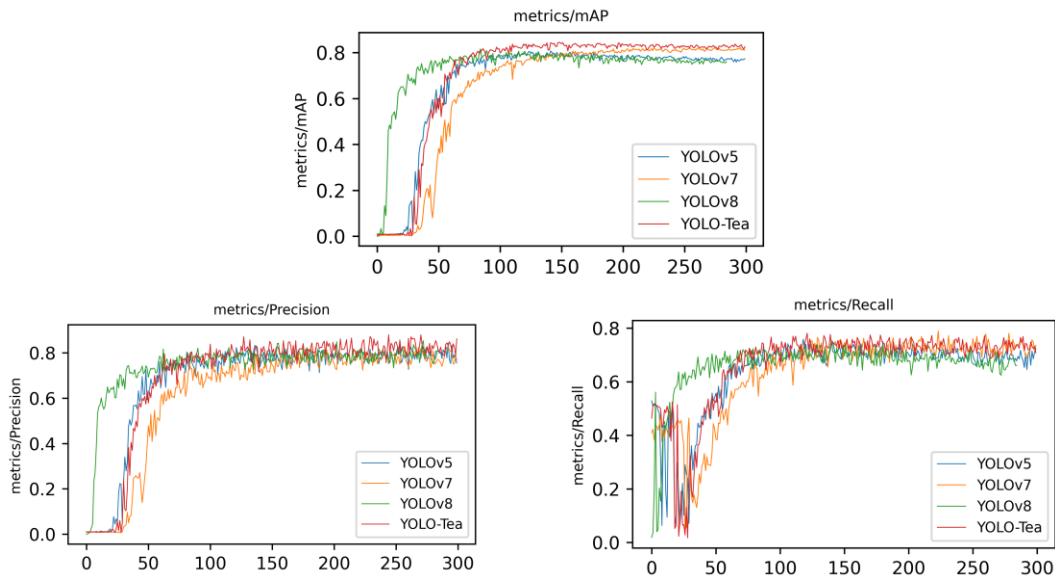


图 4-6 与其他目标检测网络对比结果图

### 4.3 本章小结

本章通过替换为自制数据集和网络模型优化，进一步提高模型性能。同时，控制了模型参数及大小，适合模型部署到移动设备上，避免在硬件设备算力低、性能差导致识别效果不好的劣势，改进后的 YOLO-Tea 识别茶芽的准确率为 84.5%，召回率为 74.1%，平均检测精度为 83.7%，模型最终大小仅为 14.9MB，能够适用于实际采茶场景。

## 第 5 章 基于 PyQt5 的茶芽目标检测系统设计

基于前两章训练出的目标检测模型，为了给检测系统提供一套操作便捷、反馈直观的智能化工具，本章基于 PyQt5 开发了一个能够方便高效进行检测任务的系统界面，通过图形用户界面（Graphical User Interface, GUI），可以省去繁杂的代码调试过程，实现输入图像、视频实时检测跟踪，提高用户的操作效率<sup>[74]</sup>。本章将围绕系统架构设计、功能实现及交互优化展开详细阐述。

### 5.1 系统概述

#### 5.1.1 PyQt5 简介

PyQt5 作为 Python 语言与 Qt 应用框架的深度融合工具，为开发者提供了构建跨平台桌面应用程序的高效解决方案。Qt 框架自 1991 年由挪威 Trolltech 公司发布以来，经过三十多年的迭代演进，逐渐成为工业级 GUI 开发的标准。PyQt5 通过将 Qt 的 C++ 类库封装为 Python 模块，既保留了 Qt 强大的图形渲染能力与事件处理机制，又充分发挥了 Python 语言的简洁性与快速开发优势，这种优势结合使其在科学计算、数据可视化、工业控制等领域得到广泛应用<sup>[75]</sup>。相较于传统 GUI 开发工具，PyQt5 的核心竞争力体现在三个维度：

（1）它继承了 Qt 框架一次编写，多平台编译的基因，开发者无需针对 Windows、macOS、Linux 等不同操作系统进行重复适配，能够降低了跨平台部署的复杂度。

（2）其内置的样式表机制（Qt Style Sheets, QSS），允许开发者像设计网页一样通过 CSS 语法定制界面元素的外观，这种声明式的风格定义方式大幅提升了 UI 设计的灵活性与可维护性。

（3）独特的信号与槽通信模型，以松耦合的方式实现了界面组件与业务逻辑的交互，能够避免了传统回调函数带来的代码混乱问题。

PyQt5 通过 SIP（Simplified Wrapper and Interface Generator）工具实现了 Python 与 C++ 的高效互操作，使得开发者既能使用 Python 的快速原型开发优势，

又能调用 Qt 底层优化的图形渲染引擎，这种实现方式在茶芽检测系统中得到充分体现<sup>[76]</sup>。检测算法基于 OpenCV 处理图像数据，通过 PyQt5 的 QImage 类实现像素级渲染控制，最终在界面层呈现出带有动态标注框的检测结果。

### 5.1.2 架构设计

界面端负责主要功能模块的显示及结果可视化，通过简洁的可视化界面实现高效人机互动功能，平台的整体功能框架图如 5-1 所示。

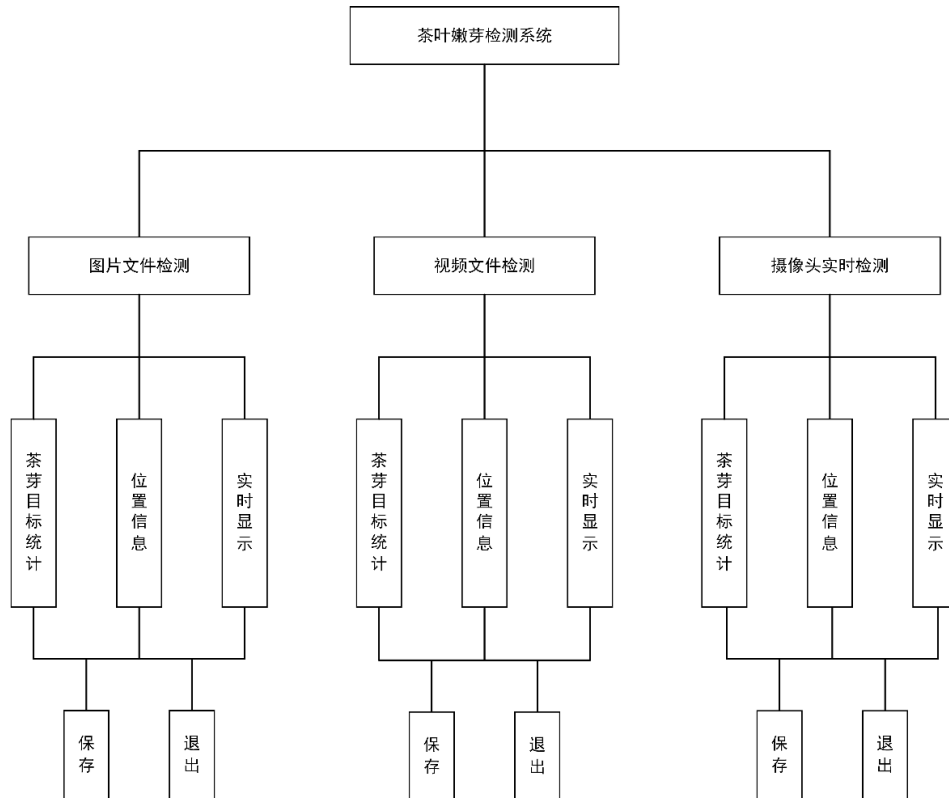


图 5-1 平台功能框架图

## 5.2 人机交互界面设计

### 5.2.1 环境配置

本章设计的茶芽用户检测界面，使用的是 PyCharm 为集成开发环境，在 Windows 操作系统上搭建基于 PyTorch 深度学习框架的茶芽用户检测界面，配备 AMD EPYC 7H12 64-Core Processor 硬件平台。通过 PyCharm 导入 PyQt5、Numpy、OpenCV、PyTorch 等工具包，为用户检测界面功能和设计提供支持，开发环境使用工具包如表 5-1 所示。

表 5-1 PyQt5 开发环境工具包

| 工具包                | 功能                  |
|--------------------|---------------------|
| Numpy 1.26.1       | 用于处理多维数组和矩阵运算       |
| Pytorch 2.6.0      | 深度学习框架，支持网络构建、张量计算等 |
| PyQt5 5.15.2       | 创建桌面应用程序和图形用户界面     |
| Opencv 4.11.0      | 提供丰富图形处理及计算机视觉算法    |
| PyQt5-tools 5.15.2 | Python 转换工具包        |

### 5.2.2 主界面设计

为了让检测结果直观高效，通过图形用户界面，用户可以轻松地在图片、视频和摄像头实时检测之间切换，本章设计了图 5-2 茶芽检系统主界面。简洁的界面不仅提升了系统的可用性，还使得检测过程更加直观透明，便于结果的实时观察和分析。此外，GUI 还可以集成其他功能，如检测结果的保存与导出、检测参数的调整，从而为用户提供一个全面、综合的检测方式，促进智能检测技术的广泛应用。系统主界面包含以下几个核心功能区：



图 5-2 茶芽检系统主界面

(1) 文件导入区域：位于茶芽检测系统主界面的右上部分，主要功能是让用户选择检测目标的方式，分为图片文件、视频文件和摄像头实时检测三种。

(2) 图像显示区域：位于茶芽检测系统主界面左侧，主要显示文件导入后茶芽检测框的情况，可以根据导入的文件实时刷新界面。



(3) 检测结果及位置信息区域：分为两个部分，第一部分位于主界面下方的位置，用于展示检测文件的每个检测框位置信息、类别、置信度和文件路径，方便了解文件的整体信息。第二部分位于主界面右侧中间位置，显示检测文件的用时和目标数目，同时可以根据目标选择来展示每一个检测框信息。

(4) 操作区域：可以将检测结果的图像文件和视频文件保存在主文件夹下的 save\_data 文件里，方便实验结果的对比。也可以直接退出茶芽检测系统

由此可以看出，茶芽图像检测主界面 能够集多功能于一体，确保了用户在茶芽图像检测中能够接收到清晰的视觉信息、直观的操作体验和准确的数据展示，这些功能的共同协作，确保了用户能够高效地识别目标并结果分析，提高工作效率。界面的设计观念保证简洁，并在茶芽检测与分析的过程中提供更多信息反馈，满足检测需求。

### 5.3 功能测试

本节根据茶叶嫩芽检测系统的功能区进行综合性功能测试，先将训练好的模型放入主程序的指定 model 文件夹里，运行程序。然后将需要检测的图片、视频导入进去，就能自动将茶芽标注出来，图 5-3 为第 4 章训练出的 best.pt 权重文件下测试集的检测结果示意图，检测结果与位置的总体信息显示在指定区域上。



图 5-3 检测结果示意图



根据目标选择可以查看各检测框的详细信息，以 tea bud\_1 为例，图 5-4 展示了单个茶芽显示信息，用户可以根据哪种情况下茶芽检测效果较差针对性修改网络模型。



图 5-4 单个检测框信息查看

点击操作区域的保存按钮，可以将本次检测结果的所有文件保存到 save\_data 文件中。以检测整个文件夹结果为例，保存结果如图 5-5 所示。

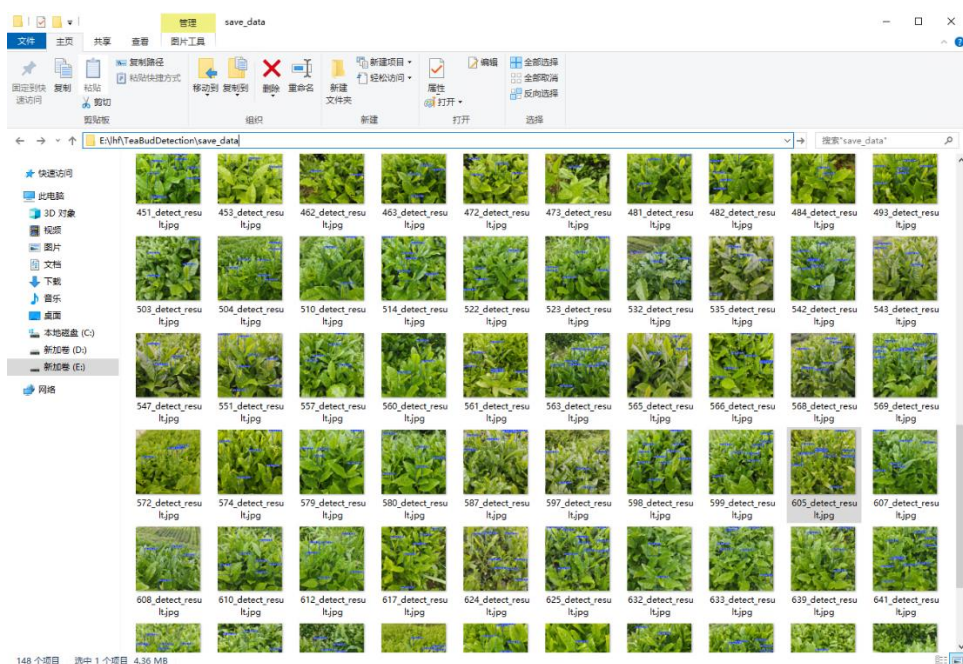


图 5-5 检测结果保存

## 5.4 本章小结

茶叶嫩芽检测平台的界面设计并非简单的功能堆砌,而是一场人机协作逻辑的深度重构。PyQt5 框架的引入,在算法黑箱与用户之间架起一座透明桥梁,将抽象的卷积神经网络输出转化为用户的可视化检测结果。这种将结果具象化,本质上是对农业智能化技术以最朴素的形式融入生产场景。本章的 PyQt5 构建的用户界面迎合上面观念,简洁高效的界面设计,确保检测任务的高效性和易操作性。

## 第 6 章 总结与展望

### 6.1 总结

中国作为全球最大的茶叶生产国，每年春茶采摘季需动员数百万劳动力，但人工采摘效率低、标准不一的问题导致茶叶采摘产能受限，茶农收益难以突破。尤其在名优茶产区，嫩芽的适时采摘直接决定茶叶品质与经济效益，然而产茶区环境的复杂性使得机械化采摘始终面临许多技术问题。为解决复杂自然环境下茶叶嫩芽识别，本文围绕农业生产中实际面临的复杂场景需求，研究了数据采集、图像处理、改进算法、图像用户界面等，实现对茶叶嫩芽高效准确的检测。本文主要内容如下：

（1）数据采集与图像数据预处理：利用手机模仿采茶机的视角在不同自然光照强度、不同角度、不同拍摄范围等条件下对六安瓜片进行了图像采集，将拍摄图像在 LabelImg 进行数据标注，并将其以 YOLO 数据集格式进行存储。为了提高数据集多样性，对数据集进行水平上下翻转、Gamma 变换、添加噪声等离线数据增强，以及在线数据增强。最后将数据集的训练集和验证集按照 8:2 的比例随机划分，来完成数据集的构建。

（2）通过改进 YOLOv5 算法，在公开数据集上进行算法改进，实现对茶芽检测精度的提高。在主干网络中引入了混合注意力机制 CBAM 来提高复杂背景下的茶芽检测能力；在颈部网络采用高效的 BiFPN 结构，提高网络模型的多尺度融合能力；最后使用归一化的 NWD 损失函数改善茶芽位置偏差敏感的问题。最后通过消融实验对比，体现改进后的 YOLO-CBN 模型更优的检测能力。

（3）YOLO-Tea 的轻量化模型在自制数据集上实现了模型权重文件仅为 14.9MB 大小的高精度检测。在主干网络部分采用计算量更少的 SimSPPF 模块提高模型特征提取能力；在颈部网络采用与 YOLO-CBN 相同的 BiFPN，提高多尺度融合能力；最后将颈部网络的 Conv 卷积模块替换为 ODConv 全维动态卷积提高特征适应性。经过与现阶段主流 YOLO 算法进行对比，可知 YOLO-Tea 的模型更小，在复杂场景下对茶芽检测精度更高，同时能够达到实时检测的效果。

（4）使用 PyQt5 工具集成 YOLO-Tea 模型，开发茶叶嫩芽检测系统的人机

交互界面，通过界面能够对图像、视频和实时场景的茶芽进行识别，并将结果可视化，处理后的结果可以保存到指定文件，方便对比调整模型。

## 6.2 展望

本文研究的对茶芽的目标检测在实验上取得良好的表现效果，其中涉及多个方面，例如如何采集优质数据集、数据集数据增强、设计及改进网络模型、构建茶叶嫩芽交互界面等。并验证了实验结果的可行性，对茶叶智能化采摘的发展提供一定帮助。但由于个人能力和时间有限在运用到实际采茶场景前任有许多优化空间：

（1）本文研究使用的自制数据集通过人工标注完成，由于茶芽目标密集，且与老叶极为相似，标注时带来的工作量巨大，十分耗费精力，因此标注结果可能会有一定主观性，存在漏标的可能性。这种不确定性会对实验结果造成一定影响。因此，有必要寻找替代人工标注方法，最后再人为补充标注，来提高数据集质量。

（2）本文中主要关注的为六安瓜片，且采摘茶叶图像为春茶，因此在未来研究中依托林下作物平台，将针对不同类别茶树、不同季节进行数据采集，丰富研究课题深度，同时关注更多的评价指标，避免不平衡数据集的情况。

（3）本文现阶段只能实现在实验室对茶叶嫩芽识别和定位，未能运用到实际机械采茶场景上，因此针对茶芽定位问题需要结合茶园真实情况进一步改善。

## 参考文献

- [1] 李煜东.神农尝百草篇——茶[J].科技视界,2023,13(20):49-50.
- [2] 尹惠玲. 茶叶对人体的保健作用[J]. 饮食科学, 2017, 10(20): 109.
- [3] 杨蕊,刘家欣,唐永蕊,等.功能性茶饮料的研究进展[J].粮食与食品工业,2025,32(01):20-22.
- [4] 茶叶分类 ISO 国际标准正式公布[J].茶业通报,2023,45(02):64.
- [5] 袁国凤.我国茶叶市场规模持续增长[N].中国食品报,2024-10-16(007).
- [6] 梅宇,梁晓.2023 年我国茶叶产销及进出口形势分析[J].中国茶叶,2024,46(04):18-26.
- [7] 朱永兴.最近 5 年世界各主要产茶国茶园面积和茶叶年产量[J].茶叶科学,2015,35(04):362.
- [8] 肖宏儒,张萌,李德权,等.我国茶叶生产机械化发展现状及战略分析[J].茶叶科学技术,2008,(02):21-23.
- [9] 喻世龙.割刀自适应调节的采茶机优化设计与试验研究[D].福建农林大学,2024.
- [10] 吕孝石,杨文凯,蒋宇辰,等.国内外主要茶叶采摘机械概述[J].现代园艺,2025,48(07):101-102+106.
- [11] 倪康,石元值,郑生宏,等.四种采茶机的名优茶采摘效果综合比较研究[J].中国茶叶,2024,46(04):40-43+51.
- [12] 段凯,蔡克桐,梅军,等.丘陵山区茶园中耕机械化研究进展[J].安徽农业科学,2017,45(22):159-161.
- [13] Xu W, Zhao L, Li J, et al. Detection and classification of tea buds based on deep learning[J]. Computers and Electronics in Agriculture, 2022, 192: 106547.
- [14] Tan H R, Zhou W. Metabolomics for tea authentication and fraud detection: recent applications and future directions[J]. Trends in Food Science & Technology, 2024: 104558.
- [15] Springenberg J T, Dosovitskiy A, Brox T, et al. Striving for simplicity: The all convolutional net[J]. arXiv preprint arXiv:1412.6806, 2014.

- [16] 唐仙, 吴雪梅, 张富贵, 等. 基于阈值分割法的茶叶嫩芽识别研究[J]. 农业装备技术, 2013 (6): 10-14.
- [17] 彭红星, 邹湘军, 陈丽娟, 等. 基于双次 Otsu 算法的野外荔枝多类色彩目标快速识别[J]. 农业机械学报, 2014, 45(4): 61-68.
- [18] 姜苗苗, 问美倩, 杨芷羽. 基于 MRFO 的茶叶嫩芽图像分割方法[J]. 农业装备与车辆工程, 2021, 59(5): 29-32.
- [19] 周礼赞, 方梦瑞, 吕军. 低对比度茶叶嫩芽图像自动分割方法的研究[J]. 蚕桑茶叶通讯, 2019, 2: 22-25.
- [20] Karunasena G, Priyankara H. Tea bud leaf identification by using machine learning and image processing techniques[J]. International Journal of Scientific & Engineering Research, 2020, 11(8): 624-628.
- [21] Zhang L, Zou L, Wu C, et al. Method of famous tea sprout identification and segmentation based on improved watershed algorithm[J]. Computers and Electronics in Agriculture, 2021, 184: 106108.
- [22] 张博, 张禹, 白广东. 基于 LabVIEW 的茶叶嫩芽识别技术研究[J]. 工业控制计算机, 2020, 33(03): 61-63.
- [23] Bao W, Fan T, Hu G, et al. Detection and identification of tea leaf diseases based on AX-RetinaNet[J]. Scientific reports, 2022, 12(1): 2183.
- [24] Cao M, Fu H, Zhu J, et al. Lightweight tea bud recognition network integrating GhostNet and YOLOv5[J]. Mathematical biosciences and engineering: MBE, 2022, 19(12): 12897-12914.
- [25] Chen Y T, Chen S F. Localizing plucking points of tea leaves using deep convolutional neural networks[J]. Computers and Electronics in Agriculture, 2020, 171: 105298.
- [26] Qian C, Li M, Ren Y. Tea sprouts segmentation via improved deep convolutional encoder-decoder network[J]. IEICE TRANSACTIONS on Information and Systems, 2020, 103(2): 476-479.
- [27] Yan L, Wu K, Lin J, et al. Identification and picking point positioning of tender tea shoots based on MR3P-TS model[J]. Frontiers in plant science, 2022, 13: 962391.
- [28] Xu W, Zhao L, Li J, et al. Detection and classification of tea buds based on deep

- learning[J]. Computers and Electronics in Agriculture, 2022, 192: 106547.
- [29] Li J, Li J, Zhao X, et al. Lightweight detection networks for tea bud on complex agricultural environment via improved YOLO v4[J]. Computers and Electronics in Agriculture, 2023, 211: 107955.
- [30] Shuai L, Mu J, Jiang X, et al. An improved YOLOv5-based method for multi-species tea shoot detection and picking point location in complex backgrounds[J]. Biosystems Engineering, 2023, 231: 117-132.
- [31] Xie S, Sun H. Tea-YOLOv8s: A tea bud detection model based on deep learning and computer vision[J]. Sensors, 2023, 23(14): 6576.
- [32] Wang J, Zhang C, Chang M, et al. Optimization of electronic nose sensor array for tea aroma detecting based on correlation coefficient and cluster analysis[J]. Chemosensors, 2021, 9(9): 266.
- [33] Xiaoxiao S U N, Shaomin M U, Yongyu X U, et al. Detection algorithm of tea tender buds under complex background based on deep learning[J]. Journal of Hebei University (Natural Science Edition), 2019, 39(2): 211.
- [34] Zhang F, Sun H, Xie S, et al. A tea bud segmentation, detection and picking point localization based on the MDY7-3PTB model[J]. Frontiers in Plant Science, 2023, 14: 1199473.
- [35] Li Z, Liu F, Yang W, et al. A survey of convolutional neural networks: analysis, applications, and prospects[J]. IEEE transactions on neural networks and learning systems, 2021, 33(12): 6999-7019.
- [36] Gu J, Wang Z, Kuen J, et al. Recent advances in convolutional neural networks[J]. Pattern recognition, 2018, 77: 354-377.
- [37] Aloysius N, Geetha M. A review on deep convolutional neural networks[C]//2017 international conference on communication and signal processing (ICCSP). IEEE, 2017: 0588-0592.
- [38] Niepert M, Ahmed M, Kutzkov K. Learning convolutional neural networks for graphs[C]//International conference on machine learning. PMLR, 2016: 2014-2023.
- [39] Xu J, Li Z, Du B, et al. Reluplex made more practical: Leaky ReLU[C]//2020 IEEE



- Symposium on Computers and communications (ISCC). IEEE, 2020: 1-7.
- [40] Boureau Y L, Ponce J, LeCun Y. A theoretical analysis of feature pooling in visual recognition[C]//Proceedings of the 27th international conference on machine learning (ICML-10). 2010: 111-118.
- [41] Sun M, Song Z, Jiang X, et al. Learning pooling for convolutional neural network[J]. Neurocomputing, 2017, 224: 96-104.
- [42] Basha S H S, Dubey S R, Pulabaigari V, et al. Impact of fully connected layers on performance of convolutional neural networks for image classification[J]. Neurocomputing, 2020, 378: 112-119.
- [43] Petersen P, Voigtlaender F. Equivalence of approximation by convolutional neural networks and fully-connected networks[J]. Proceedings of the American Mathematical Society, 2020, 148(4): 1567-1581.
- [44] 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.
- [45] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Advances in neural information processing systems, 2012, 25.
- [46] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.
- [47] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770–778).
- [48] 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.
- [49] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7263-7271.
- [50] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.
- [51] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of

- object detection[J]. arXiv preprint arXiv:2004.10934, 2020.
- [52] 马鸽,李洪伟,严梓维,等.基于多注意力的改进 YOLOv5s 小目标检测算法[J].工程科学学报,2024,46(09):1647-1658.
- [53] 黄佳惠,施贵安,童婉婷,等. “互联网+” 背景下六安瓜片助农新模式研究[J].数字农业与智能农机,2024,(05):102-105.
- [54] 向娇,宋超,石迎春,等.安徽省六安地区土壤 pH 空间变异特征及其影响因素[J].土壤通报,2021,52(01):34-41.
- [55] 孙鲁云,殷曼玲.推动六安瓜片茶产业高质量发展研究[J].新疆农垦科技,2023,46(02):28-31.
- [56] Xu W, Zhao L, Li J, et al. Detection and classification of tea buds based on deep learning[J]. Computers and Electronics in Agriculture, 2022, 192: 106547.
- [57] 高友文, 周本君, 胡晓飞. 基于数据增强的卷积神经网络图像识别研究[J]. 计算机技术与发展, 2018, 28(8): 62-65.
- [58] Shimada M, Ohtaki T. Generating large-scale high-quality SAR mosaic datasets: Application to PALSAR data for global monitoring[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2010, 3(4): 637-656.
- [59] Kamalov F, Denisov D. Gamma distribution-based sampling for imbalanced data[J]. Knowledge-Based Systems, 2020, 207: 106368.
- [60] 孙海英. 图像高斯噪声及椒盐噪声去噪算法研究[D]. 上海: 复旦大学, 2012.
- [61] Ajmal A, Hollitt C, Freen M, et al. A comparison of RGB and HSV colour spaces for visual attention models[C]//2018 International conference on image and vision computing New Zealand (IVCNZ). IEEE, 2018: 1-6.
- [62] Wang C Y, Liao H Y M, Wu Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2020: 390-391.
- [63] Wang Q, Wu B, Zhu P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 11534-11542.
- [64] Zhu X, Cheng D, Zhang Z, et al. An empirical study of spatial attention mechanisms in deep networks[C]//Proceedings of the IEEE/CVF international

- p>conference on computer vision. 2019: 6688-6697.
- [65] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.
- [66] Ghiasi G, Lin T Y, Le Q V. Nas-fpn: Learning scalable feature pyramid architecture for object detection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 7036-7045.
- [67] Chen J, Mai H S, Luo L, et al. Effective feature fusion network in BIFPN for small object detection[C]//2021 IEEE international conference on image processing (ICIP). IEEE, 2021: 699-703.
- [68] Wang J, Xu C, Yang W, et al. A normalized Gaussian Wasserstein distance for tiny object detection[J]. arXiv preprint arXiv:2110.13389, 2021.
- [69] Gui Z, Chen J, Li Y, et al. A lightweight tea bud detection model based on Yolov5[J]. Computers and Electronics in Agriculture, 2023, 205: 107636.
- [70] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 37(9): 1904-1916.
- [71] Jia Y, Huang C, Darrell T. Beyond spatial pyramids: Receptive field learning for pooled image features[C]//2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2012: 3370-3377.
- [72] Yang B, Bender G, Le Q V, et al. Condconv: Conditionally parameterized convolutions for efficient inference[J]. Advances in neural information processing systems, 2019, 32.
- [73] Li C, Zhou A, Yao A. Omni-dimensional dynamic convolution[J]. arXiv preprint arXiv:2209.07947, 2022.
- [74] Zhang H, Huang W, Qi J G. Design and implementation of object image detection interface system based on PyQt5 and improved SSD algorithm[C]//2022 IEEE 10th Joint International Information Technology and Artificial Intelligence Conference (ITAIC). IEEE, 2022, 10: 2086-2090.

- [75] 任安虎,李珊.集成 Vissim-Python 和 Qt 的信控交叉口 DRL 配时仿真系统设计[J].计算机应用与软件,2024,41(11):53-59+122.
- [76] 刘建丰,陈烨,焦良葆,等.基于 YOLOv5 的深度学习样本管理和训练自动化平台的设计与实现[J].信息化研究,2024,50(06):66-72.

## 攻读学位期间发表的学术成果

- [1] Li H, Kong M, Shi Y. Tea Bud Detection Model in a Real Picking Environment Based on an Improved YOLOv5[J]. Biomimetics, 2024, 9(11): 692.

## 致 谢

从 2022 年入学至今，转眼间三年的研究生生涯已至尾声。站在此刻回望实验室里此起彼伏的键盘声、雨后校园飘落的花瓣，都成了记忆里闪着微光的碎片。这一路虽偶有焦虑迷茫，却也满载成长与温暖，而所有的感动与收获，皆因身后有许多人的帮助。

感谢导师在我求学道路上各个方面的帮助，感谢各位评审老师能在百忙中抽出时间参与我论文评审，指出我论文中的问题。感谢实验室同学给我提供良好的实验环境，也祝你们顺利毕业，找到满意的工作。

此外，感谢国家资助和培养，以后一定好好回报社会，为国家建设贡献一份力量。