

分类号：
学号：20222108041

密级：公开
单位代码：10759

石河子大学

硕士学位论文



基于边缘计算的片式滴头质量检测系统研究与实现

学位申请人	赵泽宇
指导教师	李志刚 教授
申请学位类别	专业硕士
专业名称	电子信息
研究领域	计算机技术
所在学院	信息科学与技术学院

中国·新疆·石河子

2025年5月

分类号：
学号：20222108041

密级：公开
单位代码：10759

石河子大学

硕士学位论文



基于边缘计算的片式滴头质量检测系统研究与实现

学位申请人	赵泽宇
指导教师	李志刚 教授
申请学位类别	专业硕士
专业名称	电子信息
研究领域	计算机技术
所在学院	信息科学与技术学院

中国·新疆·石河子

2025年5月

**Research and implementation of chip dripper quality detection
system based on edge computing**

A Dissertation Submitted to

Shihezi University

In Partial Fulfillment of the Requirements

for the Degree of

Master of Engineering

By

Zhao Ze-Yu

(Electronic Information)

Dissertation Supervisor **Li Zhi-gang**

May, 2025

石河子大学学位论文独创性声明及使用授权声明

学位论文独创性声明

本人所提交的学位论文是在我导师的指导下进行的研究工作及取得的研究成果。据我所知，除文中已经注明引用的内容外，本论文不包含其他个人已经发表或撰写过的研究成果。对本文的研究做出重要贡献的个人和集体，均已在文中作了明确的说明并表示谢意。

研究生签名：赵泽宇

时间：2025年5月26日

使用授权声明

本人完全了解石河子大学有关保留、使用学位论文的规定，学校有权保留学位论文并向国家主管部门或指定机构送交论文的电子版和纸质版。有权将学位论文在学校图书馆保存并允许被查阅。有权自行或许可他人将学位论文编入有关数据库提供检索服务。有权将学位论文的标题和摘要汇编出版。保密的学位论文在解密后适用本规定。

研究生签名：赵泽宇

时间：2025年5月26日

导师签名：李正刚

时间：2025年5月26日

摘要

随着边缘计算和深度学习的广泛应用，其在智慧农业领域的潜力逐步显现。片式滴头作为灌溉系统中的关键组件，其质量检测对于提高农业灌溉精准性具有重要意义。然而，由于边缘设备计算资源有限，传统深度学习模型难以直接应用于此类场景。针对这一问题，本文基于边缘计算硬件框架，围绕片式滴头检测任务，对 YOLOv8 模型进行了轻量化与精度优化改进，并在实际边缘设备上完成了系统部署与测试。本文主要研究内容如下：

(1) 片式滴头检测算法模型轻量化设计研究。为降低 YOLOv8 模型的计算复杂度，本文首先将主干网络替换为 StarNet 网络，将 C2f 模块中的 Bottleneck 结构优化为 StarBlock 模块，此外，对检测头部分进行了参数复用设计，通过减少冗余参数进一步降低资源需求。对比原 YOLOv8 模型浮点计算量下降了 27.6%，参数量下降了 45.3%。上述优化策略使得改进后的模型在满足边缘设备计算能力的同时，具备良好的轻量化特性，为资源受限的实时检测任务提供了可行方案。

(2) 轻量级模型算法优化研究。引入 TA 注意力机制强化模型的特征提取能力，提升对片式滴头关键部位的检测准确性。同时，采用 WIoU v3 损失函数代替传统损失函数，改善模型对小目标和复杂目标的回归效果。通过实验验证，平均精度均值相比原轻量化模型提升 5.1%，使模型在精度和效率之间达到了较好的平衡。

(3) 片式滴头智能边缘检测平台部署研究。本文选用 RK3588 边缘计算平台，将 YOLOv8 改进模型进行跨框架转换和部署优化。具体流程包括将 PyTorch 模型转换为 ONNX 格式，再通过 RKNN 工具优化生成适配 RK3588 的模型文件。改进模型在 RK3588 边缘设备上模型平均加载时间相比原 YOLOv8 模型缩短 7.9 秒，平均推理时间缩短 2.4 秒，本文设计的边缘检测平台不仅能够实时检测片式滴头，还具有低延迟优势，充分验证了该方法在农业装备生产场景中的可行性与应用价值。

关键词：边缘计算；目标检测；YOLOv8；RK3588

Abstract

With the widespread application of edge computing and deep learning, their potential in the field of smart agriculture is gradually emerging. As a key component in the irrigation system, the quality inspection of the chip dripper is of great significance to improving the accuracy of agricultural irrigation. However, due to the limited computing resources of edge devices, traditional deep learning models are difficult to directly apply to such scenarios. To address this problem, this thesis, based on the edge computing hardware framework, has improved the YOLOv8 model by lightweighting and optimizing the accuracy of the chip dripper detection task, and completed the system deployment and testing on the actual edge device. The main research contents of this thesis are as follows:

(1) **Research on lightweight design of chip dripper detection algorithm model.** In order to reduce the computational complexity of the YOLOv8 model, this thesis first replaces the backbone network with the StarNet network and optimizes the Bottleneck structure in the C2f module to the StarBlock module, thereby significantly reducing the number of parameters and computation of the model. In addition, the detection head part is designed with parameter reuse to further reduce resource requirements by reducing redundant parameters. Compared with the original YOLOv8 model, the floating point calculation amount is reduced by 27.6% and the number of parameters is reduced by 45.3%. The above optimization strategy enables the improved model to have good lightweight characteristics while meeting the computing power of edge devices, providing a feasible solution for resource-constrained real-time detection tasks.

(2) **Research on lightweight model algorithm optimization.** The TA attention mechanism is introduced to enhance the feature extraction ability of the model and improve the detection accuracy of key parts of the chip dripper. At the same time, the WIoU V3 loss function is used instead of the traditional loss function to improve the regression effect of the model on small and complex targets. Through experimental verification, the average precision mean is improved by 5.1% compared with the original lightweight model, so that the model achieves a good balance between accuracy and efficiency.

(3) **Research on the deployment of the intelligent edge detection platform for chip drippers.** This thesis uses the RK3588 edge computing platform to convert and deploy the improved YOLOv8 model across frameworks. The specific process includes converting the PyTorch model into ONNX format, and then optimizing the RKNN tool to generate a model file adapted to RK3588. The average loading time of the improved model on the RK3588 edge device is 7.9 seconds shorter than that of the original YOLOv8 model, and the average inference time is shortened by 2.4 seconds. The edge detection platform designed in this thesis can not only detect the sheet dripper in real time, but also has the advantage of low latency, which fully verifies the feasibility and application value of this method in the agricultural equipment production scenario.

Key words: edge computing; target detection; YOLOv8; RK3588

目录

摘要.....	I
Abstract	II
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 传统机器视觉检测方法.....	3
1.2.2 深度学习目标检测方法.....	3
1.2.3 边缘计算研究现状.....	5
1.3 研究内容与技术路线.....	7
1.3.1 研究内容.....	7
1.3.2 技术路线.....	7
1.4 本文组织结构.....	9
第 2 章 相关技术及理论介绍	10
2.1 卷积神经网络.....	10
2.1.1 卷积层.....	10
2.1.2 池化层.....	11
2.1.3 激活网络.....	12
2.1.4 批标准化.....	13
2.2 YOLO 系列网络模型.....	14
2.3 目标检测算法评价指标.....	19
2.3.1 交并比.....	19
2.3.2 精确度和召回率.....	19
2.3.3 平均精度值.....	20
2.4 本章小结.....	20
第 3 章 面向片式滴头检测算法的轻量化设计研究	21
3.1 数据集构建.....	21
3.2 骨干网络轻量化设计.....	22
3.2.1 星操作.....	23
3.2.2 StarNet 轻量级主干网络及其模块.....	24
3.3 特征融合模块与检测头轻量化设计.....	26

3.3.1 特征融合模块轻量化设计	26
3.3.2 检测头轻量化设计	27
3.4 实验与分析	30
3.4.1 实验环境与参数设置	30
3.4.2 基于轻量化模型的消融试验	31
3.4.3 检测算法对比试验	32
3.5 本章小结	33
第 4 章 轻量化模型的算法优化研究	34
4.1 基于注意力机制的 S-YOLO 算法优化	34
4.1.1 CBAM 注意力机制	34
4.1.2 TA 注意力机制	37
4.1.3 对比实验	38
4.2 损失函数优化	39
4.2.1 SIOU 损失函数	40
4.2.2 WIOU 损失函数	42
4.2.3 对比实验	43
4.3 实验设置与结果分析	43
4.4 本章小结	45
第 5 章 片式滴头检测模型边缘计算硬件部署	46
5.1 总体方案设计	46
5.2 边缘平台选择	47
5.3 软件部署	48
5.4 平台测试及结果分析	51
5.5 本章小结	52
第 6 章 结论与展望	53
6.1 结论	53
6.2 未来展望	54
参考文献	55
致谢	59
作者简介	60

第1章 绪论

1.1 研究背景及意义

新疆深居亚欧大陆腹地，远离海洋的封闭地形使水资源成为该地区最重要的资源之一。新疆水资源总量约 878 亿立方米，其中地表水占比超 90%，主要依赖天山、昆仑山等高山冰川和季节性积雪融水补给，形成春旱、夏洪、秋缺、冬枯的独特水文特征。冰川融水贡献了约 40% 的河川径流量，地下水方面，平原区可开采量为 85 亿立方米，水资源与人口、耕地分布错位问题突出。通过多年持续发展，新疆已建成全球规模最大的节水农业示范区。全疆高效节水灌溉面积突破 6000 万亩，智能水肥一体化系统覆盖率达 65%，其中棉花作物机械化采收率超过 90%。农业装备制造企业自主研发的智能化滴灌设备，成功实现地下渗灌与气象数据联动的精准调控，较传统漫灌节水 50%、节肥 30% 的同时，推动特色林果业单产提升 25%。这种将水资源保护与农业生产效能提升相结合的实践模式，为干旱半干旱地区农业可持续发展提供了重要示范样本。

当前我国制造业正加速向创新引领发展模式转型，明确将工业智能化转型作为工业体系转变的目标之一。以智能技术重构制造工艺制造流程的智能化控制与质量监控体系，正成为重塑产业形态的革新路径。伴随消费市场对产品精细化需求攀升，工艺环节复杂度提升使得生产过程中易出现质量偏差，产品缺陷不仅降低产品性能参数，更直接影响企业经济效益与市场地位，制约行业整体发展。

片式滴头是一种小型的滴灌装置，通常由塑料制成。它们被安装在滴灌管道上，通过微小的孔或开口控制水滴的流动。每个片式滴头可以提供一定数量的水滴，以滴灌植物根部所需的水分，从而实现高效的植物灌溉。片式滴头作为滴灌系统的关键组件，可以提供精确的水分控制和节水效果，因此在现代农业中得到广泛应用。

随着工业飞速发展，人工智能等先进技术与传统工业的不断融合，工业信息化成为工业发展的趋势，对于滴灌产品生产企业来说，其中的重要一环就是滴头的质量检测，在视觉识别技术与硬件算力发展的背景下，以高精度光学感知为核心的工业检测系统取得突破性进展。这类技术依托可编程逻辑控制器与高速图像采集装置对产品进行质量检测，推动生产线向无人化方向演进。本研究聚焦片式滴头质量在线监测，设计开发集成边缘计算的片式滴头质量检测方法，该技术不仅有助于完善滴灌装备制造产业链质量保障体系，同时也有一定的市场前景。基于上述的需求和背景，将智能化引入企业中是势在必行的事情，这可以大大提升效率。但工厂的设备不能承受对设备符合较大的智能化检测模型，这就为工厂的智能化推进增加了阻碍。

因此研发基于边缘计算设备的滴头检测系统具有现实的应用价值和研究价值。该轻量级系统能够布置在工厂的边缘设备上，并且能够在生产过程中利用深度学习训练滴头图片中的信息，使用提取的图片和标准产品进行匹配，最终得出该滴头的检验结果，工作人员在工作过程中可以根据结果对滴头进行选取。

已有学者针对深度学习检测行为进行了大量的研究，多数利用传统的机器学习算法，但是忽视了各物体之间的物理属性区别。近年来发展火热的深度学习方法用来解决检测问题，但是对滴头的特征及相关属性的理解有限。基于上述问题，本文以边缘计算平台为基础，引入注意力机制来实现对滴头的检测，有助于完善深度学习与边缘计算在滴头检测方面的研究。

如何检测工业生产中的质量具有重要的研究意义。本研究根据构建的边缘计算平台结合新疆某滴灌公司采集的实际生产数据，将检测到的异常数据及时反馈给工作人员，能够精准把握工厂的实时状态并为做出科学决策，有助于石河子地区的滴灌产业建设，响应国家号召。

1.2 国内外研究现状

在现代化工业体系中，品质管控的核心环节在于产品质量检测，产品质量检测需系统检测产品表面的裂纹、残疵、异物遮蔽、色斑渗透、腔体破损等异常缺陷，完成缺陷类型识别、几何特征解析、空间定位标记及量化参数提取等多维特征参数采集^[1]。工业制造智能化发展趋势明显，传统的质量检测方法难以满足动态化检测需求。基于深度神经网络质量检测方法相较于传统模式识别方法展现出显著技术优势：其通过端到端训练机制规避了繁琐的特征工程优化流程；借助多层非线性变换构建层级化特征抽象机制，有效解决了人工特征描述子泛化能力不足的瓶颈问题。目前，基于深度学习的表面质量评估模型已在多个工业场景实现技术落地。然而初期构建的复杂网络架构对计算资源的依赖性，严重制约了其在产线边缘计算单元的适配性。技术实用化的关键突破点在于实现模型结构与部署环境的协同优化，确保算法性能与硬件资源的动态平衡。最新技术演进中，研究者聚焦于模型压缩与加速技术的创新，着力构建推理效率与识别精度的均衡优化策略，推动工业检测系统向高精度低功耗方向持续迭代。但在早期的深度学习模型往往对算力有着较高的要求，这一特点使得早期的深度学习模型难以适应工厂边缘设备的要求。模型本身的成功与否取决于模型能否成功在适宜设备上有效部署并发挥模型本身的作用。

近年来深度学习理论不断更新，研究人员将关注视角转向深度学习模型中如何兼顾模型准确度和轻量化，这一转变使减少深度学习模型参数量技术得到发展。

1.2.1 传统机器视觉检测方法

传统机器视觉检测方法是基于手工特征的算法的视觉检测方法，这些传统算法通过提取手工特征来识别目标物体。2001 年，Viola-Jones^[2]检测器采用积分图像技术，在 AdaBoost^[3]算法的基础上使用 Haar-like^[4]小波特征来进行人脸检测。通过多级检测机制优化识别效率，使图像的非关键部分被迅速过滤，进而将运算资源集中于潜在目标范围，有效缩减整体处理负荷。该技术采用层级筛选策略，优先排除低概率区域，确保算力聚焦于高可能性对象，实现效能与精度的平衡提升。之后，R.Lienhart 和 J.Maydt 用对角特征对 Haar 特征库进行了扩展^[5]，在深度学习技术出现之前，Viola-Jones 一直是人脸检测算法的主流框架。HOG^[6]（Histograms of Oriented Gradients）是一种基于图像梯度的特征提取方法，被广泛应用于计算机视觉和机器学习领域。由 N.Dalal 和 B.Triggs 在 2005 年提出。核心思想是通过捕捉图像局部区域的梯度方向分布来描述物体形状特征，从而实现目标的识别。输入图像通常转换为灰度图，再将图像划分为若干小区域，统计每个小区域内所有像素的梯度方向，将其离散化为 9 个方向区间，形成 9 维的方向直方图。以减少计算复杂度。将相邻的多个小区域组合成一个模块，将模块内的所有小区域直方图进行串连，使所有的模块进行串联，形成整幅图像的 HOG 描述子。DPM^[7]（Deformable Part Model）算法由 P.Felzenszwalb 于 2008 年提出，主要用于解决物体在图像中因姿态、视角变化导致的形变问题。将目标分解为多个可变形部件，通过建模部件间的空间关系和局部特征来提高检测鲁棒性。SVM^[8]（Support Vector Machines）是一种监督学习算法，将不同类别的数据在特征空间中最大化间隔寻找一个最优超平面。在经典框架中，SVM 被用于部件过滤器训练以区分目标部件与背景。

综上所述，传统目标检测方法高度依赖人工定义特征的构建，导致特征表征的泛化性能和场景迁移能力存在局限性，在面对多目标的复杂场景时，识别精度随着检测对象数量的增加而呈现衰减趋势，此类算法由于模型结构设计缺陷导致算法复杂度提高，占用大量硬件资源和计算资源，相较而言，深度神经网络通过自主提取多层次语义特征突破了传统方法的性能瓶颈，更在模型适应度和计算资源分配方面实现了跨越性提升。

1.2.2 深度学习目标检测方法

在深度学习出现之前，经典目标检测方法在复杂环境中的识别准确率存在显著局限。2012 年，AlexNet^[9]于 ImageNet 竞赛中取得突破性成果，研究者证实卷积神经网络具备提取高阶特征的能力，推动深度学习技术在视觉领域全面应用，并促使检测技术进入高速发展阶段。基于深度学习的检测框架主要分为两类：以 R-CNN^[10]（Regions with Convolutional Neural Network Features）算法为代表的二阶段检测算法，其优势在于高精度但存在计算延迟。以及 YOLO^[11]（You Only Look Once）系列为代表的单阶段检测算

法。一阶段检测算法相比二阶段算法检测精度有所降低，但检测速度相比二阶段算法有所提升，侧重实时性优化。对于追求实时结果的工业界相对应用较广。

2014 年，双阶段代表性算法 R-CNN 算法问世，R-CNN 使用选择性搜索等方法来生成一组候选区域，这些候选区域可能包含目标对象。对每个候选区域，通过裁剪和缩放形成固定大小的图像块，这些图像块被称为区域建议。对每个区域建议，通过一个预先训练好的卷积神经网络提取特征。Fast R-CNN^[12]是一种快速的目标检测算法，是 R-CNN 算法系列的改进版本。Fast R-CNN 在 R-CNN 的基础上进行了优化，与 R-CNN 类似，Fast R-CNN 也使用选择性搜索等方法生成候选区域，但不同的是，Fast R-CNN 将整张图像作为输入，而不是每个候选区域单独输入，从而减少了冗余的卷积计算。Faster R-CNN^[13]是一种改进的目标检测算法，旨在提高 R-CNN 系列算法的速度和准确性。它由 Microsoft Research 团队于 2015 年提出。Faster R-CNN 引入了一个称为区域建议网络的模块，用于快速而准确地生成候选区域。在卷积特征图上滑动窗口，同时预测候选框的边界框回归和目标性得分，从而生成高质量的区域建议。Mask R-CNN^[14]是一种在目标检测任务中具有重要影响的算法，由 Kaiming He 等人于 2017 年提出，是 Faster R-CNN 的扩展版本。Mask R-CNN 不仅可以进行目标检测，还可以实现实例分割，即在检测到的对象周围生成精确的对象掩模。

SSD^[15] (Single Shot MultiBox Detector) 是一种流行的单阶段目标检测算法，由 Wei Liu 等人于 2016 年提出。SSD 的设计旨在实现高效的目标检测，能够在单个前向传播中同时预测目标的位置和类别，同时处理不同尺度的目标。SSD 借助多层次特征图实施检测，以此高效识别各类尺寸的物体。这些特征图来自于基础卷积神经网络的不同层级。在每个特征图上，SSD 使用多个预定义的先验框来预测目标的位置和类别。这些先验框覆盖了不同尺度和长宽比的目标。对于每个先验框，SSD 通过卷积神经网络预测目标的类别概率和边界框的位置调整值。这些预测通过分类器和回归器实现。在输出的目标框中，SSD 使用非极大值抑制 (Non-Maximum Suppression) 来去除重叠较多的框，以获得最终的目标检测结果。

YOLO 把视觉检测视为回归问题进行处理，YOLOv1 将输入图像均匀划分成大小相等的若干个网格，一个网格只能预测一个物体，并生成两个预测框，预测框的大小不限，每一个框都包含五个元素，预测框中心坐标与子表格左上角间的偏移量，预测框的宽和高，以及类别数目，端到端的结构使得 YOLO 在速度上有较大优势。

在 YOLOv1 版本上进行创新后得到了 YOLOv2^[16]，YOLOv3^[17]等改进版本，研究人员根据改进版本进行实际应用创新，Xie^[18] 针对工业场景下的缺陷识别任务，开发了一种融合特征增强机制的 YOLO 架构创新方案。该研究引入深度解耦卷积运算模块替代标准卷积层，实现网络参数压缩与计算强度优化。通过重构层级特征融合结构强化多层次特征图间的空间关联特性优化跨尺度信息传递机制。采用基于密度分布的 k-means++

聚类策略以重新设计预设锚框参数的合理性并有效改善模型训练过程的收敛效率。在标准工业缺陷基准数据集上的实验表明,该方案在检测速度与定位精度方面均展现出显著优势。Liu^[19]研究团队开发了 LF-YOLO 架构,专门针对工业无损检测领域的焊缝缺陷识别任务。该方案通过集成高效特征编码组件优化网络表征能力,并创新性改进网络拓扑结构以增强工业场景适配性。实测数据显示,该框架在焊缝 X 光图像解析任务中实现 92.9% 的平均精度均值,同时保持 61.5 帧每秒的实时处理速度。通过跨领域验证测试,在通用目标检测基准 COCO 数据集上的评估结果进一步证实了该架构的泛化性能提升。

医疗领域, Wu^[20]等人基构建了基于 YOLO 框架的 Cell_YOLO 架构,其核心算法采用中心距优化型 NMS 抑制策略与增强型池化层设计,在降低运算复杂度的同时提升细胞分割精度,该方法可高效处理宫颈细胞重叠区域的精细分割问题。农业领域, Li^[21]等人开发了 MG-YOLO 框架用于灰霉孢子动态捕捉,通过融合 BiFPN^[22]特征金字塔与 GhostCSP^[23]轻量级架构,使模型检测准确率达到 98.3%,为农作物病害预警系统提供了可靠技术支撑。当前单阶段检测领域除 YOLO 系列外,还涵盖 RetinaNet^[24], EfficientDet 等。

综上所述,深度神经网络方法展现出出色的跨场景适应能力,其通过海量训练样本的自适应过程即可实现对多样化工业品缺陷的泛化性识别。相较于传统图像处理技术,该框架直接从原始输入信息的复杂映射关系中自主构建特征空间,这种数据驱动的特征表示机制相比人工预设特征展现出显著优势。

1.2.3 边缘计算研究现状

边缘计算(Edge Computing)是一种分布式计算范式,其核心思想是将数据处理、存储和分析能力从传统的云端下沉到靠近数据源的网络边缘,为用户提供最邻近的支持。程序在边缘端启动,能够迅速响应网络请求,符合各行业对即时操作、智能应用、数据安全与隐私保障的需求。边缘计算位于物理对象与工业系统连接的中间层,或者处于物理对象的最高层。而云计算依然能够获取边缘计算所存储的过往信息。

边缘计算的概念最早可以追溯到 20 世纪 90 年代末的内容分发网络。内容分发网络通过在靠近用户的地理位置部署缓存服务器,减少数据传输的距离和时间,优化了互联网内容的交付效率。真正意义上的边缘计算是在云计算兴起后,作为其补充而逐渐成型。2009 年,学术界提出了“边缘计算”的概念,强调将计算任务从集中式云端下沉到网络边缘,以应对物联网设备激增和数据处理需求的变化。

2014 年,边缘计算开始进入产业视野。作为边缘计算的一种早期形式,强调在网络边缘设备上进行处理。随后,边缘计算的概念被进一步扩展,涵盖了从终端设备到边缘服务器的多种计算形态。

边缘计算的技术发展经历了几个关键阶段:初期阶段主要是雾计算和移动边缘计算

时期。边缘计算主要服务于移动网络优化，解决带宽瓶颈和延迟问题。欧洲电信标准化协会在 2014 年成立了工作组，推动边缘计算在 4G 网络中的应用，满足移动网络的低时延高质量网络要求。

随着 5G 技术的商用化和物联网设备的普及，区块链技术的发展，边缘计算进入快速发展阶段，5G 的高带宽、低延迟特性为边缘计算提供了理想的网络基础。人工智能的边缘化部署成为技术热点，推动边缘设备智能化发展。

中国学者在边缘计算领域的研究多聚焦实践导向的探索，结合多样化情境需求，开发适配的智能模型或解决方案。百度飞桨框架发布的轻量级推理引擎 PaddleLite 具备 NVIDIA EGX 边缘计算平台兼容性，冯原^[25]等将其部署于自主导航 AGV 设备并完成稳定性验证。AI 推理框架 TensorFlow Lite^[26]将云端服务部署的 TensorFlow 模型优化为 TensorFlow Lite，使其适配边缘计算节点，实现对有限数据流的即时分析。谷歌工业物联网平台验证 TensorFlow 框架可适配 EdgeTPU 硬件，促使 AI 功能向边缘计算设备迁移，执行边缘侧数据实时处理任务。Caffe^[27]也可进行轻量化后部署到边缘设备上。微软推出的边缘协同方案 Azure IoT Edge，将云端的机器学习、认知服务、分析服务下沉到边缘设备，加快对异常事件的预警反应^[28]。

边缘计算通过分布式节点在近端网络层完成数据解析与计算任务，有效突破了移动设备面临的算力瓶颈与传输时延双重制约。该模式依托本地化数据处理机制，不仅规避了云端回传路径中的安全隐患，更通过数据驻留原则强化了隐私保护合规性。在集成机器视觉的智能终端场景中，边缘计算支持从图像采集、预处理到特征提取、决策生成的全链路实时处理，相较于传统云端依赖模式中终端-云中心-终端的环状传输结构，其采用的计算卸载策略可提升感应速度，同时降低网络流量负载。这种端侧智能化重构既满足工业质检、动态安防等高时效场景需求，也为智慧城市等大规模物联部署提供了可持续的通信资源优化方案。

当对国内外研究现状进行分析后，可以得出以下结论：

传统目标检测方法高度依赖人工标注数据的构建，导致算法应用存在局限性，目标检测方法需要新的技术支撑。

深度学习方法通过大量数据训练样本自适应过程实现对多种物体实现目标检测，相比较传统方法有明显优势，成为主流目标检测学习方法。

边缘计算通过将数据在数据源附近进行处理，无需通过数据传输，节省时间，使数据处理效率得到提升。

综上所述，当前目标检测模型和边缘计算存在一些关键问题，首先，近年来模型对精度的要求越来越高，但精度的提升也使得模型变大，推理时间和模型加载时间变长，已有的轻量化模型检测精度太低，无法满足实时目标检测需求，其次，工业生产中并没有合适的边缘计算设备来满足工厂生产中所需要的数据处理等需求。当前目标检测模型和

边缘计算现状需要技术对其进行优化。

1.3 研究内容与技术路线

1.3.1 研究内容

本文在模型轻量化、模型检测性能优化两个方面对片式滴头检测算法进行设计，优化并验证模型效果。最后将模型在 RK3588 开发板上进行部署和模型验证，为农业装备生产提供边缘端质量检测思路，具体内容如下：

(1) 片式滴头检测算法轻量化设计研究，本文选择 YOLOv8n^[29]作为标注模型进行轻量化改进，使用 StarNet^[30]主干网络替换原 YOLOv8n 模型中的主干网络，以减少原网络中的参数量和计算量。通过共用检测头参数减少参数量，对原检测头进行轻量化设计。使用 StarNet 中 StarBlock 模块对 YOLOv8 中 C2f 模块进行改进，保证模型检测精度在合理下降范围内实现模型轻量化。

(2) 轻量级模型算法优化研究，对比 CBAM^[31]注意力机制和 (Convolutional Block Attention Module) TA 注意力机制^[32] (Triplet Attention) 对模型参数量，计算量和准确度的平衡能力，通过对比后选择引入 TA 注意力机制并添加在主干网络末端。同时选择 Wise-IoU^[33]损失函数对原损失函数进行替换并进行对比试验，在未增加参数量和计算量的同时，对模型训练过程进行精度提升。

(3) 片式滴头智能边缘检测平台部署研究。将模型的 pt 格式模型转化为 ONNX (Open Neural Network Exchange) 格式模型，再将 ONNX 格式模型转化为 RNNK (Rockchip Neural Network) 格式模型，在选择的边缘硬件设备上对改进后的模型进行部署试验。完成能在 RK3588 端进行片式滴头缺陷检测的轻量化系统，为农机装备生产管理者提供直观质量检测方案。

1.3.2 技术路线

本文技术路线主要分为数据准备、理论研究和系统设计三个阶段：

查阅相关文献和资料，分析片式滴头检测的研究现状和应用需求，明确检测任务的目标与边缘计算设备的约束条件。在数据准备过程中，针对片式滴头的特性，构建了具有代表性的数据集，确保样本多样性和检测任务的覆盖面。数据准备阶段为后续模型的改进与系统的实现提供了充足的支持。

在理论研究阶段，围绕片式滴头检测的实际需求，分析主流深度学习目标检测方法的特点及其在边缘设备上的应用可行性，设计了一种更适合片式滴头检测的优化模型。结合检测任务特点，进行了模型结构的轻量化调整，并设计了改进策略以提升模型在有

限计算资源下的性能表现。通过对优化后的模型进行实验验证和对比分析，评估了模型在精度和效率方面的适用性，探索了其在片式滴头检测场景中的表现优势。

在系统设计与整合阶段，结合理论研究阶段的优化成果，针对边缘设备的特点，完成了模型的部署与系统功能的实现。系统整体架构设计注重模块化和实用性，涵盖实时检测及结果输出等功能模块。最终完成的系统在满足边缘计算场景实时性需求的同时，具备良好的稳定性和资源利用效率，为智慧农业片式滴头检测提供了高效解决方案。

技术路线整体结构图如图1-1所示。

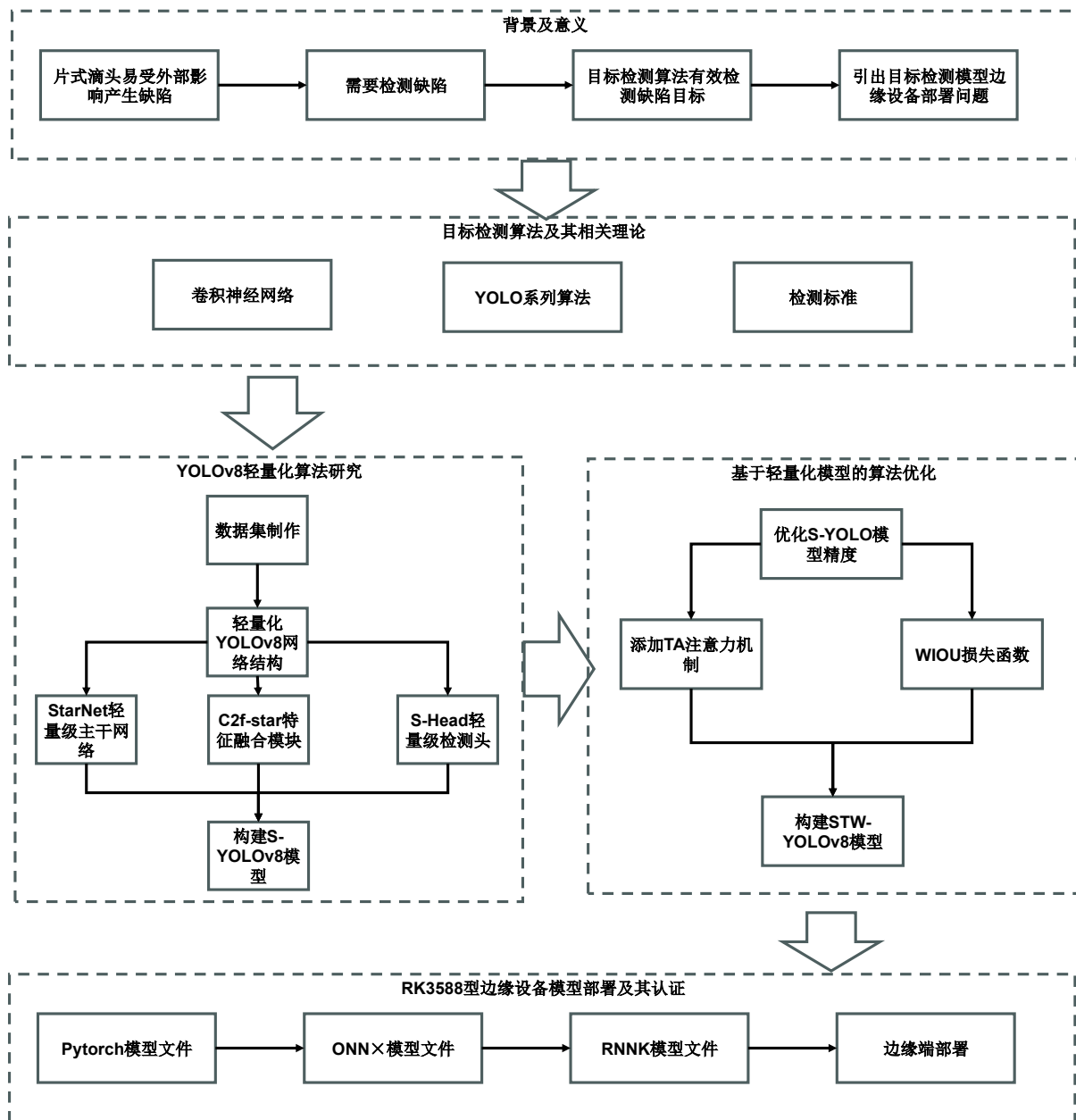


图1-1 研究总体路线图

Fig. 1-1 Overall roadmap of research