

分类号:  
学号: 20232208001

密级: 公开  
单位代码: 10759

石河子大学

硕士学位论文



面向边缘部署的 YOLO  
交通标志检测方法研究与系统实现

学位申请人	畅国文
指导教师	赵庆展 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

2026 年 06 月

分类号：  
学号：20232208001

密级：公开  
单位代码：10759

# 石河子大学

## 硕士学位论文



### 面向边缘部署的 YOLO 交通标志检测方法研究与系统实现

学位申请人	畅国文
指导教师	赵庆展 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

2026 年 06 月

**Research and System Implementation of YOLO  
Traffic Sign Detection Method for Edge Deployment**

A Dissertation Submitted to  
**Shihezi University**  
In Partial Fulfillment of the Requirements  
for the Degree of  
**Master of Engineering**

By

**Chang Guo-wen**  
**(Electronic Information)**


Dissertation Supervisor: Prof. Zhao Qing-zhan

June, 2026

## 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


本人所呈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： 

时间： 2026 年 5 月 15 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： 

时间： 2026 年 5 月 15 日

导师签名： 

时间： 2026 年 5 月 20 日

## 摘要

随着边缘计算技术的不断发展,交通标志检测在智能交通系统中的应用变得越来越重要。然而,现有方法在边缘设备部署过程中仍面临模型复杂度高、小目标检测性能不足以及推理效率受限等问题。针对上述挑战,本文设计并实现了一种适用于 Atlas 200I DK 平台的轻量级交通标志检测算法,并构建了完整的检测系统。

本研究首先分析了 GTSDB 和 TT100K 数据集特征,针对交通标志小目标检测特点改进 YOLO 网络结构,提出了 RSO-YOLO 算法。在模型结构方面,通过构建多层次特征融合机制并引入跨层连接,有效增强浅层细粒度信息的利用能力;设计深层下采样结构,在降低分辨率的同时缓解特征信息流失问题;结合重参数化策略与大核卷积结构,提升模型表达能力并优化推理效率,同时在颈部网络中引入通道注意力机制以增强特征表征能力。其次,针对 Atlas 200I DK 平台的硬件特性,对模型进行部署优化。通过参数重构与 BN 层融合降低计算开销,采用 ARQ 与 HFMG 量化算法对权重与激活进行低比特表示,并结合数据均衡预处理缓解离群值对量化精度的影响。同时,引入 AIPP 模块对输入数据预处理过程进行硬件级优化,以减少数据搬运与计算延迟。在系统实现方面,基于多线程机制设计上位机软件,实现模型管理、视频流处理及检测结果可视化等功能。

实验结果表明,在 GTSDB 数据集上,RSO-YOLO 模型的精度较基线模型提升 4.8 个百分点;在 TT100K 数据集上提升 7.5 个百分点,同时保持较低的推理延迟。经量化后模型体积减少 66.7%,精度仅下降 0.3%,推理延迟降低至 27ms。结合 AIPP 优化,系统整体延迟进一步降低 6.5ms。在 720P 视频流输入下,系统检测帧率达到 27.6 FPS,满足非密集型交通标志检测任务的需求。

综上,本文提出的 RSO-YOLO 方法在保证实时性的同时显著提升了小目标交通标志检测性能,并通过模型压缩与硬件协同优化实现高效边缘部署。研究成果可为边缘智能视觉系统的设计与实现提供参考。

**关键词:** 交通标志检测; 边缘计算; 深度学习; 模型量化; YOLO 算法

## Abstract

With the continuous development of edge computing technology, traffic sign detection has become increasingly important in intelligent transportation systems. However, existing methods still face challenges such as high model complexity, insufficient small object detection performance, and limited inference efficiency during edge device deployment. To address these challenges, this paper designs and implements a lightweight traffic sign detection algorithm suitable for the Atlas 200I DK platform and constructs a complete detection system.

This study first analyzes the characteristics of the GTSDB and TT100K datasets, and proposes the RSO-YOLO algorithm by improving the YOLO network structure specifically for small traffic sign detection. In terms of model architecture, a multi-level feature fusion mechanism with cross-layer connections is constructed to effectively enhance the utilization of shallow fine-grained information. A deep downsampling structure is designed to mitigate feature information loss while reducing resolution. By combining re-parameterization strategies with large kernel convolution structures, the model's expressive capability is enhanced and inference efficiency is optimized. Additionally, a channel attention mechanism is introduced in the neck network to strengthen feature representation capabilities. Second, deployment optimization is performed for the hardware characteristics of the Atlas 200I DK platform. Computational overhead is reduced through parameter re-structuring and batch normalization layer fusion. ARQ and HFMG quantization algorithms are employed for low-bit representation of weights and activations, while balanced data preprocessing mitigates the impact of outliers on quantization accuracy. Meanwhile, the AIPP module is introduced to perform hardware-level optimization of the input data preprocessing process, reducing data transfer and computational latency. For system implementation, host computer software is designed based on a multi-threading mechanism to realize functions including model management, video stream processing, and detection result visualization.

Experimental results demonstrate that on the GTSDB dataset, the RSO-YOLO model achieves a 4.8 percentage point improvement in accuracy compared to the baseline model, and a 7.5 percentage point improvement on the TT100K dataset, while maintaining low inference latency. After quantization, the model size is reduced by 66.7% with only a 0.3% accuracy drop, and inference latency is reduced to 27ms. Combined with AIPP optimization, the overall system latency is further reduced by 6.5ms. With 720P video stream input, the system achieves a detection frame rate of 27.6 FPS, meeting the requirements of non-intensive traffic sign detection tasks.

In summary, the RSO-YOLO method proposed in this paper significantly improves small object traffic sign detection performance while ensuring real-time capability, and achieves efficient edge deployment through model compression and hardware co-optimization. The research findings provide valuable reference for the design and implementation of edge intelligent vision systems.

**Key words:** Traffic sign detection; Edge computing; Deep learning; Model quantization; YOLO algorithm

# 目录

第一章 绪论.....	1
1.1 研究背景和意义.....	1
1.1.1 研究背景.....	1
1.1.2 研究意义.....	2
1.2 国内外研究现状.....	2
1.2.1 基于机器学习的交通标志检测算法.....	2
1.2.2 基于深度学习的交通标志检测算法.....	4
1.2.3 边缘计算在交通标志检测中的应用现状.....	5
1.3 交通标志检测技术的研究难点.....	6
1.4 主要研究内容及章节安排.....	7
第二章 相关理论与技术.....	9
2.1 数据集简介.....	9
2.2 目标检测技术概述.....	10
2.2.1 深度学习目标检测基本理论.....	10
2.2.2 YOLO 系列目标检测技术.....	15
2.2.3 目标检测任务的评估指标.....	19
2.3 模型量化技术概述.....	19
2.3.1 模型量化的基本原理.....	20
2.3.2 常见量化方法.....	23
2.3.3 量化精度损失与评估方法.....	26
2.4 边缘计算平台比较与选型.....	27
2.5 本章小结.....	28
第三章 交通标志检测算法实现.....	29
3.1 数据分析和处理.....	29
3.2 网络结构设计.....	31
3.2.1 网络结构总体设计.....	31
3.2.2 小目标检测网络设计.....	32
3.2.3 重参数化结构设计.....	36
3.3 实验结果与分析.....	38
3.3.1 实验环境与参数设置.....	38
3.3.2 实验结果与评价指标.....	40

3.3.3 消融实验.....	44
3.4 本章小结.....	45
第四章 Atlas 200I DK 硬件加速及模型部署.....	46
4.1 Atals 200I DK 开发者套件平台搭建.....	46
4.1.1 Atlas 200I DK 开发者套件介绍.....	46
4.1.2 实验平台环境配置.....	47
4.2 硬件加速方案.....	50
4.2.1 模型量化.....	50
4.2.2 模型前处理过程设计.....	54
4.3 实验结果与分析.....	55
4.3.1 量化效果对比.....	55
4.3.2 模型部署.....	58
4.3.3 推理速度对比.....	59
4.4 本章小结.....	61
第五章 交通标志检测系统设计.....	62
5.1 开发环境搭建.....	62
5.2 系统框架设计.....	63
5.2.1 系统需求分析.....	63
5.2.2 系统层次结构.....	64
5.3 系统软件实现.....	66
5.3.1 基于多线程技术的软件性能优化.....	66
5.3.2 软件界面设计.....	70
5.4 系统软件验证与评估.....	71
5.4.1 系统验证方案.....	71
5.4.2 系统功能验证与评价.....	72
5.5 本章小结.....	76
第六章 总结与展望.....	77
6.1 本文总结.....	77
6.2 未来展望.....	78
参考文献.....	79

# 第一章 绪论

## 1.1 研究背景和意义

### 1.1.1 研究背景

近年来，随着城市化进程加快和机动车保有量激增，全球道路交通问题日益突出。根据世界卫生组织（World Health Organization, WHO）的数据，去年约有 119 万人死于道路交通事故<sup>[1]</sup>，这是全球主要死亡原因之一。中国作为世界上最大的发展中国家，面临着尤为严峻的交通挑战。根据中国国家统计局的数据，截至 2024 年上半年，中国机动车保有量已超过 4.4 亿辆<sup>[2]</sup>，城市道路交通压力巨大。近期高德地图发布的 2023 年全年统计结果显示，对于高延误-高时间价值城市，单个路口高峰时段每小时经济损失达 1277 元<sup>[3]</sup>。为应对这一挑战，中国政府高度重视智能交通系统的发展。自 2011 年起，中国陆续出台了一系列促进智能交通技术创新的政策<sup>[4-6]</sup>。在这一背景下，基于人工智能和计算机视觉的交通标志检测和识别技术得到了迅速发展。

作为智能交通系统（Intelligent Transportation System, ITS）<sup>[7]</sup>的重要组成部分，交通标志检测和识别技术在提高道路安全性、减少交通事故方面发挥着越来越重要的作用。交通标志检测和识别系统能够自动从数字图像或视频流中检测和识别各类交通标志，为驾驶员或自动驾驶车辆提供及时、准确的道路状况信息，从而大幅降低交通事故风险。然而，由于交通场景的复杂性、标志的多样性以及环境因素的影响，实现高效准确的交通标志检测仍然面临诸多挑战。

近些年，基于深度学习的目标检测算法，尤其是 YOLO（You Only Look Once）系列算法，在交通标志检测领域展现出了卓越的性能。YOLO 算法以其高效的单阶段检测架构和实时处理能力，成为了交通标志检测研究的热点。但将这些复杂的深度学习模型部署到实际的交通系统中仍然面临着计算资源限制、实时性要求和能耗问题等挑战。

为了解决上述问题，边缘计算的引入为交通标志检测任务提供了新的思路。边缘计算能够将数据处理任务下移至网络边缘设备，从而显著减少数据传输延迟，提高系统响应速度，同时增强数据隐私和安全性。在交通标志检测领域，基于边缘计算的解决方案有望实现更快速、更可靠的实时检测和识别，为自动驾驶和智能交通系统的发展提供强有力的技术支持。

### 1.1.2 研究意义

基于边缘计算的交通标志检测系统设计具有重要的现实意义和应用价值。这项研究可以帮助我们打造更安全、更智能的道路交通环境。首先,它能够大大提高道路安全性。通过快速准确地识别各种交通标志,系统可以及时提醒驾驶员注意前方的路况,比如限速、让行或危险警告等,从而有效减少交通事故的发生。对于自动驾驶车辆而言,这一系统更是充当了关键的环境感知模块,为车辆的决策和控制提供必要的输入。其次,这项研究推动了边缘计算技术在交通领域的应用。与传统的云计算方式相比,边缘计算将数据处理放在离用户更近的地方,这样不仅能够更快地处理信息,还能节省网络带宽,提高整个系统的效率。在交通标志检测这样需要快速反应的场景中,边缘计算的优势尤为明显。

## 1.2 国内外研究现状

交通标志检测是智能交通系统和自动驾驶技术中的关键组成部分。近年来,随着人工智能技术的快速发展,交通标志检测领域取得了显著进展。本节将从三个主要方面对交通标志检测算法的研究现状进行综述。

### 1.2.1 基于机器学习的交通标志检测算法

早期的研究主要集中在利用传统机器学习方法进行交通标志识别。机器学习的交通标志检测算法一般先对图像进行预处理,如去噪和增强;然后提取特征,常用的特征包括颜色、形状和纹理信息;最后使用分类器,如支持向量机(Support Vector Machine, SVM)、k-近邻(k-Nearest Neighbor, k-NN)<sup>[8]</sup>和决策树等,对提取的特征进行分类,从而实现交通标志的检测和识别。传统的机器学习方法在交通标志检测和识别中仍然发挥着重要作用,特别是在计算资源有限的场景下。

Sugiharto 和 Harjoko<sup>[8]</sup>提出了一种结合 HOG(Histogram of Oriented Gradient)和 PHOG(Pyramid Histogram of Oriented Gradient)特征的方法,并比较了 SVM 和 k-NN 分类器的性能。他们的研究表明,HOG+SVM 组合在检测准确率上表现最佳,达到 82.01%,而 PHOG+k-NN 组合也表现出较高的准确性(81.19%)。这一研究揭示了特征提取方法与分类器选择的重要性,为后续研究奠定了基础。随着研究的深入,学者们开始探索更复杂的特征融合和分类策略。胡聪等人<sup>[9]</sup>提出了一种基于极大极稳定区域(MSERs)和 SVM 的交通标志检测方法。该方法通过简化 Gabor 滤波器增强边缘信息,利用 MSERs 算法

生成推荐区域,并结合 HOG 特征和 SVM 分类器进行识别。这种方法在 GTSDB 和 CSTD 数据集上都取得了良好的检测性能,显示了多种算法融合潜力。Li 等人<sup>[10]</sup>进一步推进了特征融合的研究,提出了一种基于颜色直方图和 HOG 特征的融合方法。他们创新性地使用 PCA (Principal Component Analysis)降维技术来提取高效的特征向量,减少了计算复杂度和时间成本。这种方法在 GTSRB 数据集上实现了 99.99%的准确率,每幅图像的处理时间仅为 16.91 毫秒,显著提高了交通标志识别的准确性和实时性。Wang<sup>[11]</sup>的研究聚焦于寻找最优的机器学习分类器。通过比较 SVM、MLP (Multilayer Perceptron)和 LR (Logistic Regression)在交通标志分类中的性能,发现 SVM 在所有测试的分类器中表现最佳。此外,研究还发现图像锐化预处理能显著提高分类准确率,这为后续研究提供了有价值的见解。为了应对交通标志识别中的尺度和旋转变化的问题,Yazdan 和 Varshosaz<sup>[12]</sup>提出了一种创新的方法。他们利用立体成像系统生成交通标志的正交视图,消除透视效应对识别的影响。这种方法仅需为每个交通标志提供一个参考图像,就能在存在尺度和旋转变化的情况下进行准确识别,将识别准确率提高到 93.1%,比传统方法提升了 4.9%。

随着深度学习技术的发展,研究者开始探索将传统机器学习方法与深度学习技术相结合的途径。Kerim 和 Efe<sup>[13]</sup>提出了一种新型数据集和混合人工神经网络(ANN)模型。他们将 GTSRB 数据集和 TSRD 合并,并通过类别合并和样本筛选,创建了一个更均衡和多样化的训练数据集。他们开发的混合 ANN 模型结合了 9 种不同的子模型,分别训练于颜色强度、HOG 和 LBP (Local Binary Pattern)特征,通过多数投票技术整合预测结果。这种方法在新提出的数据集上达到了 95%的 top-2 分类准确率,展示了混合模型在提高分类准确性方面的潜力。最近的研究进一步探索了更复杂的特征学习和分类策略。Khalid 等人<sup>[14]</sup>提出了一种基于混合特征的低复杂度交通标志检测方法,旨在平衡计算效率和检测精度。Zhou 和 Tian<sup>[15]</sup>则提出了一种基于多核学习(MKL)的交通标志识别方法(MKL-SING)。该方法能够自动从训练数据中学习最优的核组合,避免了传统方法中依赖人工选择核函数的不足。通过结合颜色、形状和纹理等多种特征,MKL-SING 方法显著提高了交通标志识别的准确性和鲁棒性。Chaghazardi 等人<sup>[16]</sup>的研究为这一领域带来了新的视角,强调了逻辑推理在提高模型鲁棒性和可解释性方面的潜力。

基于机器学习的交通标志检测算法研究呈现出从单一特征和分类器向多特征融合、混合模型以及结合深度学习技术的趋势。研究者们不断探索新的特征提取方法、分类策略和模型融合技术,以应对实际交通场景中的各种挑战,如光照变化、尺度和旋转变化、部分遮挡等问题。基于机器学习的方法在交通标志检测和识别任务中取得了显著成果,但这些方法通常需要手动设计特征,这可能无法充分捕捉复杂场景中的所有相关信息。而且在处理大规模数据集和复杂场景时可能会遇到性能瓶颈。这些限制促使研究人员转向更强大的深度学习方法,以寻求更高的精度和更好的泛化能力。

## 1.2.2 基于深度学习的交通标志检测算法

相比于传统的机器学习方法，基于深度学习的交通标志检测算法具有自动学习特征的能力，可以直接从原始图像数据中学习更加复杂和抽象的特征表示。这类方法使用卷积神经网络（Convolutional Neural Networks, CNN）等深度学习模型作为特征提取器；通过全连接层或其他特定的网络结构进行分类或回归，从而实现交通标志的检测和识别。随着深度学习技术的快速发展，特别是 CNN 在计算机视觉任务中的成功应用，基于深度学习的交通标志检测算法逐渐成为研究的主流方向。目前研究主要集中在网络结构优化、注意力机制应用、小目标检测、复杂环境适应、长尾分布问题解决等方面。

在网络结构优化方面，YOLO 系列算法因其高效的实时性能成为研究热点。Sun 等人<sup>[17]</sup>提出 LLTH-YOLOv5 算法，针对低光场景进行了优化，通过引入轻量级低光增强网络、BIFPN 和基于 Transformer 的检测头，显著提高了低光环境下的检测准确率。Zhou 等人<sup>[18]</sup>基于 YOLOv8 提出 MixChannel\_YOLO 算法，引入混合通道注意力机制和 Slim-Neck 结构，提高了模型的识别能力和计算效率。Zhao 等人<sup>[19]</sup>提出 YOLO-FLC 算法，通过 FasterNet Block、RepConv 结构和 EMA 注意力机制实现了模型轻量化。高良鹏等人<sup>[20]</sup>在 YOLOv8n 基础上提出 Faster-YOLOv8，采用 C2f-Faster 模块优化网络结构，显著减少了参数量和模型大小。

除 YOLO 外，研究者还探索了其他网络结构。Chen 等人<sup>[21]</sup>提出了结合 CNN 和多尺度 Transformer 的半监督学习框架。Wang 等人<sup>[22]</sup>提出 DK-Former 模型，结合卷积随机傅里叶特征和 Transformer 模型，同时捕获局部和全局特征。Mingwin 等人<sup>[23]</sup>探索了将 Vision Transformers (ViT) 应用于交通标志识别的潜力，提出了创新的 EATFormer 架构。Padhy 等人<sup>[24]</sup>提出了基于改进 LeNet-5 卷积神经网络的自动交通标志识别与分类模型，通过图像预处理、特征提取和数据增强等技术，在 GTSRB 数据集上达到了 99.84% 的准确率。

注意力机制在提升交通标志检测性能方面发挥了重要作用。An 等人<sup>[25]</sup>提出了基于级联注意力机制的交通标志识别算法，设计了级联注意力特征增强模块和互注意力增强模块。田鹏和毛力<sup>[26]</sup>在 YOLOv8 基础上引入 Bi-level Routing Attention (BRA) 注意力机制，增强网络对小目标的感知能力。Shi 等人<sup>[27]</sup>在 AAMNet 交通标志实例分割网络中设计了 Adaptive Focus Localizer Head，包含通道注意力模块。Ding<sup>[28]</sup>提出了改进的 YOLO-PAN 算法，引入 LSKA 注意力机制。

小目标检测一直是交通标志检测领域的难点。Tian 等人<sup>[29]</sup>提出 OFFR-YOLO 模型，引入 SOIoU 损失函数和 SOPANet 路径聚合网络，提高小目标检测性能。Mahaur 等人<sup>[30]</sup>提出 IS-YOLOv5 模型，通过组深度可分离卷积和注意力驱动的空洞 CSP 块，显著提高

了小目标的检测精度和速度。胡均平等人<sup>[31]</sup>改进 YOLOv5 算法，通过增设新的梯度路径和引入可学习的自适应权重，提高了小目标交通标志的检测精度。

针对复杂环境下的交通标志检测问题，吴攀超等人<sup>[32]</sup>提出 CF-YOLO 检测模型，设计了基于颜色衰减先验的自适应伽马变换图像预处理算法，提高了雾霾环境中的检测能力。项新建等人<sup>[33]</sup>提出 Retinex-Gamma-Mask R-CNN 算法，结合光照增强算法与 Mask R-CNN，有效提升了复杂光照条件下的检测与识别性能。

对于长尾分布问题，Gao 等人<sup>[34]</sup>提出了基于注意力融合和分层组 softmax 的交通标志检测方法。Guo 等人<sup>[35]</sup>提出梯度再平衡交通标志识别（GRTR）方法，引入梯度再平衡机制和双阶段注意力模块，有效缓解了类别不平衡问题。

为提高模型泛化能力，研究者采用了多种数据增强和模型优化技术。李厚杰等人<sup>[36]</sup>提出基于伪样本正则化 Faster R-CNN 的算法，通过生成有标注的伪样本数据来增强训练数据集。Lim 等人<sup>[37]</sup>提出了结合预训练模型和集成学习的方法，通过多数投票技术提高了模型的准确性和鲁棒性。Kumaravel 等人<sup>[38]</sup>关注于优化深度学习算法在自动驾驶车辆交通标志识别中的超参数，利用 Pelican Optimization Algorithm（POA）和 Cuckoo Search Algorithm（CSA）对 YOLOv5 和 Faster R-CNN 等模型进行优化。

此外，针对特定国家或场景的交通标志，也有相应的研究。Waziry 等人<sup>[39]</sup>利用迁移学习方法，将预训练模型适应于土耳其交通标志数据集。Manocha 等人<sup>[40]</sup>针对韩国交通标志开发了基于深度学习的检测模型，在复杂多变的道路环境中表现出较高的检测准确率和实时性。

尽管基于深度学习的交通标志检测算法取得了显著进展，但仍面临计算复杂度高、在资源受限的边缘设备上部署困难等挑战。这促使研究人员探索如何在边缘计算环境中高效地部署和运行交通标志检测算法，如 Padaria 等人<sup>[41]</sup>提出的基于分割学习和联邦学习的交通标志分类方法，为解决自动驾驶车辆在交通标志识别中的计算效率和数据隐私问题提供了新的思路。

基于深度学习的交通标志检测算法研究呈现多元化发展趋势，在提高检测精度和速度的同时，也更加关注算法的轻量化、鲁棒性和实际应用性能。未来研究方向可能会更加注重算法在复杂实际场景中的适应性、边缘计算环境下的高效部署，以及与其他智能交通系统的深度集成。然而，深度学习模型还存在计算复杂度高，在资源受限的边缘设备上部署可能面临困难等问题。这些挑战促使研究人员探索如何在边缘计算环境中高效地部署和运行交通标志检测算法。

### 1.2.3 边缘计算在交通标志检测中的应用现状

边缘计算作为一种将计算任务从云端转移到网络边缘的技术，为解决深度学习模型

在资源受限环境下的部署问题提供了新的思路。在交通标志检测领域，边缘计算的应用主要涉及模型压缩、量化和硬件加速等技术，以在保证检测精度的同时，降低计算复杂度和能耗。

Fang 等人<sup>[42]</sup>提出了轻量级网络 MicronNet-BF 用于交通标志分类。该网络通过融合深度可分离卷积和瓶颈特征，有效降低了模型复杂度，在 GTSRB 数据集上达到了 99.38% 的准确率，处理时间仅为 1.41 秒。这一研究为后续在边缘设备上部署高效交通标志检测模型奠定了基础。Wang 等人<sup>[43]</sup>开发了车辆载小型交通标志自适应检测器 (VATSD)。通过优化特征提取模块和引入图像增强网络，显著提高了小型交通标志的检测准确性。该系统在 Jetson Xavier NX 平台上实现了 21.6 FPS 的实时处理速度，展示了边缘计算在实际应用中的潜力。张上等人<sup>[44]</sup>和张倩等人<sup>[45]</sup>分别对 YOLOv5s 模型进行了改进。张上等人引入了注意力机制 (CBAM 和 CA 融合) 和 FPGM 剪枝技术，将模型大小压缩至 0.54MB，在 Jetson Nano 上达到 21 帧/秒的检测速度。张倩等人则融合了 Transformer 和改进的 PANet，在 Jetson AGX Xavier 平台上实现了 76FPS 的检测速度。这两项研究都在保持或提高检测精度的同时，显著降低了模型复杂度，为边缘设备上的实时交通标志检测提供了可行解决方案。成怡等人<sup>[46]</sup>提出了改进 CenterNet 的交通标志检测算法。通过引入 ResNeSt50 作为主干网络，设计多尺度感受野模块和特征增强模块，并改进损失函数，该算法在保证检测精度的同时，实现了每秒 91.09 帧的高速检测，为自动驾驶中的实时交通标志识别提供了有效支持。陈哲等人<sup>[47]</sup>基于 FCOS 算法提出了改进的交通标志检测方法。通过引入注意力模块 CBAM、Swish 激活函数和轻量级多尺度特征融合，在 TT100K 数据集上实现了 83.2% 的 F1-measure 和 24.39FPS 的检测速度，验证了算法在边缘设备上的可行性。Jayasinghe 等人<sup>[48]</sup>探讨了在嵌入式系统上实现实时交通标志和交通灯检测的方法。他们提出了一种基于深度学习的端到端检测框架，通过 TensorRT 优化检测模型，并与 Robot Operating System (ROS) 集成，部署在 Nvidia Jetson AGX Xavier 嵌入式设备上。该系统实现了 63 帧每秒的高推理速度，并能够应对不同天气和光照条件的挑战，展示了边缘计算在复杂道路场景中的应用潜力。

边缘计算在交通标志检测中的应用主要聚焦于模型轻量化、硬件加速和系统集成等方面。这些研究为在资源受限的边缘设备上实现高效、实时的交通标志检测提供了重要支持，推动了智能交通系统和自动驾驶技术的发展。然而，如何在进一步提高检测精度的同时保持模型的轻量化和实时性，仍是未来研究需要持续关注方向。

### 1.3 交通标志检测技术的研究难点

交通标志检测技术的研究面临着多方面的挑战和难点，这些难点不仅涉及技术层面，

还包括实际应用中的各种复杂情况：

复杂环境下的识别问题：如图 1-1 所示，交通标志往往处于复杂多变的环境中，这给识别系统带来了巨大挑战。首先是光照变化问题，不同时间、天气条件下，标志可能出现曝光不足、过度曝光或部分光照的情况，严重影响识别准确性。其次是遮挡问题，标志可能被树木、建筑物或其他车辆部分或完全遮挡，增加了识别的难度。再者，标志本身可能因长期暴露在外而出现形变、褪色或被污染，这些都会降低识别系统的性能。

小目标识别问题：如图 1-2 所示，在实际行车场景中，由于透视效应<sup>[12]</sup>，远处的交通标志在图像中往往呈现为小目标。这些小目标不仅在图像中占比小，而且细节信息有限，容易被忽视或误判。提高小目标交通标志的识别率是评估系统性能的关键指标之一，也是研究的重点难题。

实时性和轻量化需求：对于自动驾驶等应用场景，交通标志识别系统需要在高速移动的车辆上实时运行。这就要求算法不仅要准确，还要高效。如何在有限的车载计算资源下实现快速准确的识别，如何设计轻量化模型以减少参数量和计算复杂度，都是研究中需要考虑的重要问题。



图 1-1 复杂场景交通标志

Figure 1-1 Traffic signs in complex scenarios



图 1-2 交通标志小目标

Figure 1-2 Small targets of traffic signs

## 1.4 主要研究内容及章节安排

随着智能交通和自动驾驶技术的快速发展，交通标志检测系统在保障道路安全方面发挥着越来越重要的作用。然而，目前基于深度学习的交通标志检测算法在边缘设备上的部署仍面临着诸多挑战，主要表现在计算资源受限、实时性要求高、以及复杂环境下检测精度不稳定等方面。针对这些问题，本文提出了一种基于改进 YOLO 算法的交通标志检测方案，并在 Atlas 200I DK 边缘计算平台上实现了高效部署。主要开展以下研究工作：

### (1) 基于改进 YOLO 的交通标志检测算法设计

针对交通标志检测中小目标识别困难、特征提取不充分等问题,本研究对 YOLO 网络结构进行了针对性改进。首先,通过引入多层次特征提取网络和跨层特征融合机制,增强了模型对小尺寸交通标志的检测能力。其次,采用重参数化结构结合大核卷积技术,在保证特征提取能力的同时减少了计算延迟。通过在 GTSDb 和 TT100K 数据集上的实验验证,从 mAP 和 FPS 等多个维度对算法性能进行全面评估。

### (2) 基于 Atlas 200I DK 的模型量化与硬件加速设计

为实现算法在边缘设备上的高效运行,本研究从模型量化和硬件加速两个方面进行优化。在模型量化方面,采用混合精度量化策略,对网络中的权重和激活值进行精细化量化处理,显著减少模型参数量和计算开销。在硬件加速方面,深入分析 Atlas 200I DK 的计算特性,优化 AIPP 预处理参数配置,并在与上位机软件配合使用时执行高效的线程调度策略,充分发挥硬件算力。通过量化前后的性能对比实验,验证了优化方案的有效性。

### (3) 交通标志检测系统的实现

基于改进的检测算法和模型优化策略,本研究设计并实现了一套完整的交通标志检测系统。该系统采用多线程架构,实现了图像采集、参数控制和结果显示等功能模块的并行处理。通过合理的任务划分和线程同步机制,确保系统运行的实时性和稳定性。同时,开发了用户友好的交互界面,支持实时检测结果的可视化展示和系统参数的灵活配置。

本文共分为六章,具体章节安排如下:

第一章为绪论,介绍研究背景和意义,分析国内外研究现状,阐述本文的主要研究内容。

第二章介绍相关理论与技术基础,包括目标检测基本原理、YOLO 系列算法演进、模型量化技术,以及边缘计算平台的特点与选型分析。

第三章重点阐述改进的交通标志检测算法,详细描述网络结构设计、特征提取优化等关键技术,并通过实验验证算法性能。

第四章介绍算法在 Atlas 200I DK 平台上的优化与部署过程,包括模型量化、AIPP 预处理优化等关键技术的实现与效果分析。

第五章描述交通标志检测系统的软件实现,包括系统架构设计、多线程优化、界面开发等内容,并进行系统功能验证。

第六章总结全文工作,并对未来研究方向进行展望。