分类号：TP391　　学校代码：10363
密　级：公开　　学　　号：2220910116

安徽工程大学
Anhui Polytechnic University

硕士学位论文

题目 基于机器视觉的轮胎关键信息实时检测与识别系统

论 文 作 者　李萌阳
校 内 导 师　严楠（教授）
校 外 导 师　刘冬（正高级工程师）
专业学位类别　计算机技术
研究方向（领域）　智能系统与应用

论文提交日期：2025 年 5 月 9 日

分类号：TP391　　单位代码：10363

密　级：公开　　学　　号：2220910116

题　　目 基于机器视觉的轮胎关键信息实时检测与识别系统

题　　目 A Real-time Detection and Recognition System for Key Tire Information Based on Machine Vision

学　生　姓　名：　李萌阳

校　内　教　师：　严　楠（教授）

校　外　教　师：　刘　冬（正高级工程师）

专业 学位 类别：　计算机技术

研究方向（领域）：　智能系统与应用

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：

日期：2025 年 5 月 9 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。

本学位论文属于 保密 □，在 ____ 年解密后适用本版权书。

不保密 ☑。

学位论文作者签名：  　　指导教师签名： 

日期：2025年5月9日　　日期：2025年5月9日

基于机器视觉的轮胎关键信息实时检测与识别系统

摘 要

本文将研究重点聚焦于轮胎胎侧关键信息检测与识别这一极具研究价值且意义重大的课题，深入且系统地剖析了当前字符识别技术的发展现状，并着重探讨了在轮胎胎侧字符识别领域所面临的一系列复杂挑战。随着汽车产业朝着智能化方向的迅猛发展，对于轮胎关键信息的精准检测与识别的需求愈发迫切，这已成为保障汽车行驶安全、提升车辆管理效率以及推动汽车后市场服务智能化的关键环节。然而，经过对现有研究成果的梳理发现，目前的研究大多集中于轮胎的 DOT 信息识别，而对于其他同样重要的关键信息，如轮胎尺寸信息的检测与识别，尚存在诸多不足之处，且缺乏一种能够直接、快速且易于使用的检测系统，以满足实际应用场景中的多样化需求。针对上述问题，本文提出了一种基于机器视觉的轮胎关键信息实时检测与识别系统，该系统创新性地集成了多阶段的检测方法，能够实现对轮胎胎侧的 DOT 信息以及轮胎尺寸信息的全面、精准检测与识别。

首先，提出改进 YOLOv8n-Seg 的实例分割算法用于精准定位轮胎区域，通过优化网络结构，显著降低了轮胎区域定位算法对计算资源的需要；其次，提出改进 YOLOv8s 的目标检测算法用于检测多个轮胎胎侧的关键信息区域，包括轮胎尺寸信息与轮胎DOT信息所在区域，有效提高了不同关键信息区域检测的召回率和精确率；然后，提出基于 YOLOv8s-Cls 改进的图像分类算法用于对关键区域图片进行角度分类，为后续的文本检测与字符识别提供了准确的角度信息，从而降低了因角度差异导致的误检率；并且提出改进 YOLOv8s 的

文本检测算法和基于 SVTR 改进的文字识别算法，分别用于轮胎胎侧关键信息的文本检测与字符识别。这些改进算法在提升检测精度的同时，通过优化网络结构和计算流程，有效提高了检测与识别的准确率，实现了快速准确的检测与识别，显著提高了系统的实时性和实用性。

实验结果表明，本文所提出的改进算法在检测精度和计算效率方面均优于现有的主流算法，能够有效解决轮胎胎侧关键信息检测与识别中的漏检、误检问题，显著提升了检测的准确性和效率。这一成果为轮胎关键信息检测领域提供了技术思路和方法，具有重要的理论和实践意义。

最后，完成了对不同检测阶段的系统设计以及模型部署，提出了一个快速且易于使用的基于机器视觉的轮胎胎侧关键信息实时检测与识别系统的原型设计。该系统通过RTSP视频流抽取关键帧实现实时检测，通过 Python 构建后端逻辑，利用 PyQt 进行界面设计，实现了从图片输入到结果输出的端到端检测流程，操作简便且易于扩展。该系统不仅能够为轮胎关键信息的自动化检测提供有效的解决方案，而且具有较高的应用价值和广阔的推广前景，有望在汽车制造、轮胎售后管理以及智能交通等领域得到广泛应用，为相关行业的智能化发展提供有力的技术支持。

关键词：深度学习，目标检测，文字识别，YOLOv8，SVTR

A REAL-TIME DETECTION AND RECOGNITION SYSTEM FOR KEY TIRE INFORMATION BASED ON MACHINE VISION

ABSTRACT

Research focuses on the highly valuable and significant topic of tire sidewall key information detection and recognition. It conducts an in-depth and systematic analysis of the current state of character recognition technology and particularly explores the complex challenges faced in the field of tire sidewall character recognition. With the rapid development of the automotive industry towards intelligence, the demand for precise detection and recognition of tire key information has become increasingly urgent. It has become a crucial link in ensuring automotive driving safety, improving vehicle management efficiency, and promoting the intelligent development of automotive aftermarket services. However, after reviewing the existing research findings, it is found that most of the current research is concentrated on the recognition of DOT information of tires, while there are still many deficiencies in the detection and recognition of other equally important key information, such as tire size information. Moreover, there is a lack of a direct, fast, and easy-to-use detection system to meet the diverse needs of practical application scenarios.

To address the above issues, this paper proposes a real-time detection and recognition system for tire key information based on machine vision. The system innovatively integrates a multi-stage detection method, which can achieve comprehensive and precise detection and recognition of DOT

information and tire size information on the tire sidewall. Firstly, an improved YOLOv8n-Seg instance segmentation algorithm is proposed for accurate positioning of the tire area. By optimizing the network structure, the computational resource requirements of the tire area positioning algorithm are significantly reduced. Secondly, an improved YOLOv8s object detection algorithm is proposed for detecting multiple key information areas on the tire sidewall, including the areas where tire size information and tire DOT information are located. This effectively improves the recall and precision of detecting different key information areas. Then, an improved image classification algorithm based on YOLOv8s-Cls is proposed for angle classification of key area images, providing accurate angle information for subsequent text detection and character recognition, thereby reducing the mis-detection rate caused by angle differences. In addition, an improved YOLOv8s text detection algorithm and an improved character recognition algorithm based on SVTR are proposed for text detection and character recognition of key information on the tire sidewall. These improved algorithms not only enhance detection accuracy but also effectively improve the accuracy of detection and recognition by optimizing the network structure and computational process, achieving fast and accurate detection and recognition, and significantly improving the real-time performance and practicality of the system.

Experimental results show that the improved algorithms proposed in this paper are superior to the existing mainstream algorithms in terms of detection accuracy and computational efficiency. They can effectively solve the problems of missed detection and mis-detection in the detection and recognition of key information on the tire sidewall, and significantly improve the accuracy and efficiency of detection. This achievement provides a new technical idea and method for the field of tire key

information detection, and has important theoretical and practical significance.

Finally, the system design and model deployment of different detection stages are completed, and a fast and easy to use prototype design of a real-time detection and recognition system for tire side key information based on machine vision is proposed. The system realizes real-time detection by extracting key frames from RTSP video streams, builds back-end logic by Python, and designs the interface by PyQt. The system realizes an end-to-end detection process from image input to result output, which is simple to operate and easy to expand. The system can not only provide effective solutions for the automatic detection of tire key information, but also has high application value and broad promotion prospects. It is expected to be widely used in automobile manufacturing, tire after-sales management and intelligent transportation and other fields, and provide strong technical support for the intelligent development of related industries.

KEY WORDS: deep learning,object detection, text-recognition,yolov8, svtr

目 录

第 1 章 绪论

1.1 研究背景与研究意义

随着中国经济的迅猛增长，汽车产业与交通领域正朝着智能化方向迅速发展，中国已成为汽车智能化发展的枢纽。机器视觉，作为智能化进程中的一个前沿技术，正推动着新工业革命的到来[1]，并在汽车与交通领域中得到广泛应用。目前，智能化技术已在汽车行业的车牌识别、车道线检测以及车辆缺陷检测等多个项目中发挥着重要作用。汽车轮胎作为汽车的关键组成部分，承载着整车重量，缓解冲击力，确保行驶平稳性和乘坐舒适性；同时，轮胎亦关乎提升汽车的牵引力和制动性，对保障汽车的使用寿命和安全行驶至关重要[2]-[3]。在汽车使用过程中，必须选用与车辆匹配且处于安全使用期限内的轮胎，而这些关键信息均编码于轮胎侧壁的尺寸标识和 DOT 信息编码中。

轮胎尺寸信息涵盖轮胎宽度、高宽比、类型及轮毂尺寸等关键参数，而 DOT 信息则代表美国交通部认证的轮胎安全标准[4]。轮胎侧壁的 DOT 信息中进一步包含了产地、制造工厂及生产年份等细节[5]。鉴于轮胎作为橡胶制品存在使用期限，超期使用的轮胎可能引发安全隐患，因此在汽车使用过程中，对轮胎的关键参数进行持续监测和精确识别显得尤为关键，这对于及时发现潜在问题、确保行车安全具有重要意义。目前，多数厂商和工厂仍依赖于人工目视检查和手动记录的方式，这种方法效率低下且易出错。因此，迫切需要开发一系列智能算法，以实现对轮胎字符的快速、智能化检测与识别，从而提高效率和准确性。

1.2 字符识别现状

1.2.1 传统字符识别研究现状

光学字符识别（Optical Character Recognition, OCR）技术是文本信息提取领域的关键技术之一。该技术通过电子设备扫描纸质或电子文档中的文字图像，检测字符的形状特征，并将其转换为计算机可识别的文字编码，实现文本内容的数字化处理[6]。在 OCR 技术的发展历程中，印刷体文字的识别是最早被研究和开发的领域。德国科学家 Taushek 于 1929 年便取得了 OCR 技术的专利。自 20 世纪 50 年代起，为了将报刊杂志、文件资料和单据报表等文字材料输入计算机进

行处理，西文 OCR 技术成为研究的重点，旨在替代传统的人工键盘输入。经过 40 余年的发展和计算机技术的快速进步，西文 OCR 技术已广泛应用于多个领域，实现了信息处理的“电子化”转型，极大地促进了 OCR 技术在实际应用中的普及和推广。与西文 OCR 技术相比，印刷体汉文 OCR 技术的研究起步较晚，其发展是建立在印刷体数字识别和英文识别技术基础之上的。印刷体汉字识别的研究最早可追溯至 20 世纪 60 年代。1966 年，BIM 公司的 Casey 和 Nagy 发表了首篇关于印刷体汉字识别的学术论文，他们采用简单的模板匹配方法，成功识别了 1000 个不同的印刷体汉字。此后，日本学者在该领域进行了大量的研究工作，取得了显著成果，如 1977 年东芝综合研究所研制的可识别 2000 个单体印刷汉字的系统，以及 80 年代初期日本武藏野电气研究所开发的可识别 2300 个多体汉字的印刷体汉字识别系统，这些成果代表了当时汉字 OCR 技术的最高水平。

我国的 OCR 技术起步较晚，始于 20 世纪 70 年代末，当时国内少数研究者开始探索汉字识别技术，发表了若干学术论文，并开发了一些模拟识别软件和系统。这些初期成果虽然有限，但为后续研究打下了坚实的基础。到了 1986 年，OCR 研究在国内掀起了一股热潮，众多研究机构和企业相继研发出实用化的系统产品。其中，清华大学电子工程系、中国科学院计算所智能中心、北京信息工程学院、沈阳自动化研究所等单位在该领域做出了显著贡献。特别是清华大学电子工程系开发的“清华 TH-OCR”产品和汉王集团研发的“尚书 OCR”产品，一直处于印刷体汉字识别技术的领先地位，并在国内市场中占据主导份额。这些标志性产品不仅彰显了国内学术界和企业界在前沿技术领域的强大研发实力，也为印刷体汉字识别技术的发展动向和应用前景提供了积极的示范。

1.2.2 基于深度学习的字符识别研究现状

近些年，随着深度学习的兴起 ，在基于深度学习的 OCR 算法上诞生了例如 EAST[7]、SAST[8]、DBNet[9]、DBNet++[10]等文本检测算法与 CRNN[11]、SVTR[12] 文本识别算法等等，并且通用的目标检测算法如 YOLOv5、YOLOv8、Faster R-CNN[13]等算法也可以完成文本检测任务，现在已有大量的研究人员将这些算法单独使用或者组合在一起使用共同完成对文字的检测与识别。

李卓璇等人[14]通过改进 DBNet 模型，实现了对电商图像中文字的高效检测。该研究首先在 DBNet 网络的特征金字塔网络（Feature Pyramid Networks，FPN）中集成了 SimAM 注意力模块，该模块能够在不增加额外参数的前提下，捕获更

多包含文字的区域特征。其次，引入了迭代自选择特征融合模块，以融合来自不同层级的特征信息，增强模型对文字特征的识别能力。最后，通过双边上采样模块，实现了粗粒度特征的恢复以及细粒度特征的精细化修复。这些改进显著提升了模型在电商图像文字检测任务中的表现。

Men J 等人[15]对 DBNet 进行了优化，以提升文本检测的性能。他们通过实施数据增强技术扩展了训练数据集，从而增强了模型的泛化能力。此外，他们还改进了损失函数，使得模型能够更精确地预测文本的位置。这些改进有助于提高模型对文本检测任务的准确性和鲁棒性。

刘彦希等人[16]针对电气设备铭牌文字检测任务，提出了一种改进的 EAST 算法。该研究将 ResNet50 残差网络作为模型的主干网络，替代了原有的 VGG 网络，以此增强模型的特征提取能力。此外，研究中引入了特征金字塔和空洞卷积技术，旨在强化特征图的细节表达以及上下文信息的整合，并扩展模型的感受野。这些改进使得模型在针对电气设备铭牌的文本检测任务上，相较于原始的 EAST 算法，展现出更优越的性能。通过这些技术革新，模型能够更有效地捕捉铭牌文字的特征，提高了检测的准确性和可靠性。

汪嘉等人[17]对 DBNet 文本检测算法进行了改进，以提高文本检测的效率和准确性。他们将 ShuffleNetV2 网络作为新的主干网络，替换了原有的网络结构，有效减少了模型的参数量。此外，他们引入了 ECA（Efficient Channel Attention）注意力机制模块，增强了模型对通道特征信息的提取能力。结合 CRNN（Convolutional Recurrent Neural Network）文字识别算法，该改进方案显著提升了声级计校准工作的自动化水平和准确性。这些技术改进使得模型在文本检测任务中表现出更高的性能，尤其是在声级计校准等应用场景中。

1.2.3 轮胎胎侧字符识别研究现状

汽车轮胎胎侧上印有关键信息，其中包括轮胎的尺寸规格和生产日期，这些信息是确保轮胎安全使用的基础[2]。轮胎侧壁上的文字属于压印文字，压印文字的形成原理在于背景与字符区域的高度差异导致光反射的不一致性，从而使得字符内容得以显现[18]。针对轮胎侧壁上的关键压印文字进行识别，属于字符识别技术的一个分支。目前，已有国内外学者对此领域进行了一定的研究探索。

Kazmi W[19]等人提出了一种基于深度学习与传统机器视觉结合进行轮胎胎侧关键信息压印文字检测和识别的高效工业系统。首先通过运动相机采集轮胎图

片，对采集到的图片进行霍夫圆变换来检测轮胎位置，然后对检测到的轮胎进行解翘曲，将其矫正为一个矩形，然后通过基于深度学习的级联卷积神经网络（Convolutional Neural Networks, CNN）进行文本检测与文字识别，从而完成对轮胎胎侧压印文字的检测与识别。

李嘉豪[20]通过采用多阶段深度学习方法实现了轮胎胎侧 DOT 信息压印文字的检测与识别。研究首先应用基于 Advanced EAST 的文本检测算法精准定位轮胎侧壁上的 DOT 信息文本区域。针对检测图像中文本方向的多样性对后续字符识别造成的影响，研究引入了一种方向角度分类器，对检测到的文本角度进行分类并执行矫正，以确保文本方向的一致性。最终，利用 CRNN（Convolutional Recurrent Neural Network）算法对矫正后的文本进行字符识别，完成了从文本检测到字符识别的全过程。这一方法论为轮胎胎侧 DOT 信息的自动化识别提供了一种有效的技术途径。

张晨梦[21]采用传统机器视觉技术与深度学习检测算法相结合的方法，对轮胎侧壁的 DOT 信息文字进行检测。研究首先通过图像处理技术实现同心圆检测，从而分割出轮胎侧壁区域，并通过霍夫圆变换矫正翘曲，将待检测区域转换为矩形形态。随后，利用改进的 Faster R-CNN 算法进行 DOT 信息的文本检测，该算法在双线性插值法辅助下进行图像下采样，以增强边缘信息的检测精度。双线性插值法在像素间进行插值，有效保持了图像边缘细节，而阈值处理技术则对下采样后的图像进行二值化，以精确识别目标对象。最终，通过大津法对单个字符进行分割，并运用 VGGNet[22]模型对字符进行识别，从而成功提取轮胎侧壁的 DOT 信息。这一方法论的提出，为轮胎 DOT 信息的自动化检测提供了一种有效的技术路径。

曲星华[23]通过采用改进的单阶段 YOLOv4 目标检测算法针对文本检测任务进行了优化，并结合改进的 LeNet-5[24]识别算法用于文字识别。在模型改进中，引入了批量归一化层 BN 和 Inception 模块，这些改进显著提升了模型的识别精度。最终，研究者利用 Qt 软件平台将检测与识别模块进行集成，提出了一套完整的轮胎 DOT 信息检测系统。

经过对国内外轮胎胎侧文字识别研究的综合分析，可以发现已有部分学者对轮胎胎侧文字识别领域进行了初步探索。然而，现有研究的焦点大多集中在轮胎

胎侧的 DOT（Department of Transportation）信息检测上，而忽略了轮胎其他关键信息（如品牌标识、生产日期、规格参数等）的识别与提取。这种研究倾向在一定程度上限制了轮胎胎侧文字识别技术在更广泛场景中的应用潜力。并且在现有的检测算法中，有一部分需要与传统图像处理算法结合使用，传统图像处理方法在轮胎胎侧文字识别中的主要作用包括图像去噪、边缘增强、对比度调整等，这些处理步骤对于提高后续文字识别算法的准确性和效率至关重要。然而，传统图像处理方法往往需要针对特定参数进行大量调整，例如在不同的光照条件、背景干扰以及轮胎表面磨损程度下，都需要重新优化图像处理算法的参数。这种对特定参数的依赖性严重限制了传统图像处理方法在不同应用场景中的泛化能力，导致其在面对复杂多变的实际工况时，难以实现稳定可靠的轮胎胎侧文字识别效果。

综上所述，尽管在轮胎胎侧文字识别领域已取得了一定的研究成果，但现有研究仍存在诸多局限性。需要更加全面地考虑轮胎胎侧的各种关键信息，并探索更加高效、鲁棒且具有泛化能力的图像处理与文字识别算法，以推动该技术在轮胎行业的广泛应用。

1.3 目前轮胎胎侧关键信息检测与识别待解决的问题

目前文本检测与识别技术在多个领域已经取得了较为成熟的进展，其检测与识别的精确度也达到了较高的水平。在轮胎胎侧文字检测与识别这一特定领域，研究亦取得了诸多成果。然而，这些研究领域仍然存在一些亟待解决的问题。

首先，与常见的文字检测与识别相比，轮胎胎侧的文字属于压印文字，其与背景的对比度较低，缺乏明显的边界，这使得传统的文本检测方法难以有效应用。此外，轮胎胎侧背景较深，容易造成误识别与漏识别。其次，在目前关于轮胎胎侧的文字检测与识别研究中，大多数研究主要集中在轮胎的 DOT 信息的识别上，而忽略了其他关键信息的检测与识别，例如轮胎的尺寸信息。这种研究的局限性可能导致在实际应用中无法全面满足轮胎检测的需求。

综上所述，尽管在文本检测与识别领域已经取得了一定的成果，但在轮胎胎侧文字检测与识别方面，仍需进一步研究以解决上述问题，从而提高检测与识别的准确性和任务的全面性。

针对轮胎胎侧文字识别领域的现存问题，本文构建了一种多阶段检测系统。该系统融合了多种改进算法，能够对轮胎胎侧的 DOT 信息及尺寸信息进行全面

且高效的检测与识别。借助多样化的改进方法，该系统在检测效率与检测准确度方面取得了显著提升，为轮胎质量检测与安全监管提供了有力的技术支持。

1.4 主要研究内容及研究路线

针对轮胎胎侧背景颜色较深并且胎侧关键信息文字为压印文字，特征提取困难且特征信息较少的问题，现有的研究集中在研究 DOT 信息的检测与识别，而忽略了其他关键信息的检测与识别的局限性上以及缺乏易于使用的集成系统，本文提出了一种基于机器视觉的轮胎关键信息实时检测与识别系统，集成多阶段的检测方法实现对轮胎胎侧的 DOT 信息以及轮胎尺寸信息进行检测与识别。本文的研究主要工作如下：

(1)简述本研究的背景和研究意义，介绍字符识别以及轮胎胎侧压印文字识别的国内外发展现状。

(2)阐述了所涉及的相关理论与技术研究，旨在为后续的实验设计、数据分析以及结论推导提供坚实的理论基础与技术支撑。通过对现有文献的深入梳理与分析，详细介绍了相关领域的前沿理论进展。

(3)为了精确检测轮胎胎侧上的关键信息区域，本文提出了轮胎胎侧关键信息所在区域定位算法：

首先，为了实现轮胎区域的精准定位与分割，我们采用了基于深度学习的单阶段实例分割模型 YOLOv8-Seg，并进行针对性改进。该模型能够有效地检测并分割出图像中的轮胎区域。改进后的 YOLOv8-Seg 以其高效的目标识别能力，从复杂的背景中准确提取轮胎区域，为进一步的文本检测和文字识别提供了坚实的基础。

其次，采用了改进的 YOLOv8s 目标检测算法对轮胎胎侧的压印文字进行关键信息区域检测，利用 RFMLNet 网络以替换原有的主干网络，该网络通过多重注意力机制提高对关键区域特征的提取。此外，提出了 DyFCGSlimNeck 轻量化颈部结构，这一结构在降低模型计算复杂度的同时，增强了颈部结构对多尺度特征的融合能力，加入了上下文锚点注意力机制 CAA，旨在从主干网络中提取更多的上下文特征，以此提升检测精度。

然后，对检测到的关键区域进行角度分类与校正是确保后续文字检测与识别准确性的关键步骤。当文本图像存在非正常阅读角度时，可能会导致后续的文字

识别出现错误或无法识别的情况。因此，本文提出了基于改进 YOLOv8-Cls 的图像分类算法，用于对图片进行角度分类，从而为后续的图像处理提供准确的角度信息，以便将文本图像校正至可检测方向。在改进算法中采用 StarNet 作为主干网络，通过优化网络结构，显著降低了模块的计算资源需求，实现了模型的轻量化设计。

(四)在完成对轮胎胎侧关键信息所在区域定位任务后，本文提出了轮胎胎侧关键信息 OCR 识别算法：

首先，由于轮胎胎侧的关键信息包括不同的类别，本研究对检测到的文本进行进一步的细化检测与分类，提出改进的 YOLOv8s 目标检测算法，检测出轮胎的宽度、高宽比、轮胎类型、轮毂大小以及轮胎的生产日期的压印文本区域，在改进的模型中改进了一种新型轻量化特征提取模块 RepNCSPELAN4_CAA，该模块的设计宗旨是在降低模型计算负担的同时，通过整合上下文信息注意力机制，维持模型的高检测精度；提出一种动态采样的轻量化大感受野特征融合网络 DyLKVHSPAN，与原有的颈部结构相比，DyLKVHSPAN 实现轻量化且未对特征融合造成显著损失；设计了一种新型低计算量检测头 DWLHead。该检测头的设计理念是减轻模型在输出阶段的计算负荷，实现检测模块的轻量化，同时优化检测效果。

然后，对检测到的不同类别的文本区域进行文字识别，利用 SVTR 文字识别算法进行压印文字字符识别，SVTR 是一种基于视觉的单模型场景文本识别算法，旨在通过简化的架构实现高效且准确的文本识别。，专门用于解决光学字符识别（OCR）任务，能够高效地从图像中提取文字信息，并以文本形式输出，在处理中文、英文、数字以及多语言混合场景时，SVTR 展现出了卓越的性能，使其在多种实际应用场景中具有广泛的应用潜力。

(五)完成对不同检测阶段的系统设计以及模型部署，实现检测系统的后端构建，利用 PyQt 进行界面设计，完成对检测系统的前端以及 UI 设计，最终实现基于机器视觉的轮胎胎侧关键信息实时检测与识别系统。

(六)总结与展望。本章对本文的研究工作进行了全面总结与深入分析，并对未来进一步的研究方向提出了展望。通过系统梳理本文的研究内容、方法及实验结果，清晰呈现了本研究的核心贡献与创新点。同时，基于当前研究的成果与不

足，对未来可能的研究方向进行了深入探讨，旨在为后续研究提供有益的启示与方向指引。

第 2 章 相关理论与技术研究

本节对本文所涉及的相关技术与理论进行了详细的综述。综述内容包括深度学习理论、目标检测技术、注意力机制、特征金字塔结构，以及 YOLOv8 目标检测、分类、分割算法、文本检测算法和文字识别算法等技术原理。通过对这些技术与理论的系统分析，为研究文本检测与识别技术提供了全面的理论支撑，为后续研究的深入展开提供了坚实的理论保障。

2.1 基于回归的一阶段实例分割算法

其设计理念源于单阶段目标检测的理论框架。此类算法能够同时并行地执行目标的分割与检测任务，展现出模型结构简洁、运算速度迅捷、检测精度高以及具备强大的实时处理能力等显著优势[25]。这种高效的算法设计不仅优化了计算资源的利用，而且在实时性要求较高的应用场景中，如农作物检测[26]和工业自动化检测[27]等领域，提供了关键的技术支持，下面介绍 YOLACT[28]、以及 YOLOv8-Seg 实例分割算法。

2.1.1 YOLACT 实例分割算法

YOLACT（You Only Look At Coefficients）是一种经典的基于回归的一阶段实例分割算法。该算法通过在传统的一阶段目标检测模型架构中引入一个掩码分支，与检测分支并行运行，从而实现了高效的一阶段实例分割检测。在 YOLACT 中，检测分支负责输出每个检测到的物体的类别、边界框信息以及 k 个掩码的置信度。与此同时，掩码分支生成 k 个掩码原型。通过将这些掩码原型与相应的掩码置信度相乘并求和，最终生成实例分割的结果。YOLACT 算法具有高掩码质量、高检测精度以及快速的推理速度的优点。然而，在处理图像中存在多个重叠目标的情况时，原型掩码在定位上面临挑战，容易导致定位误差的出现。YOLACT 算法的结构图如图 2-1

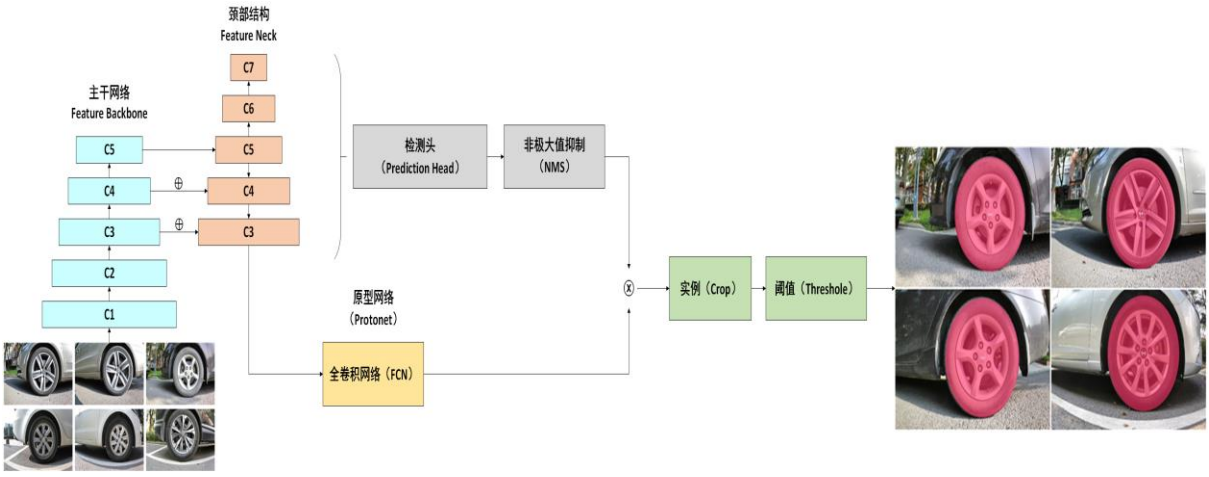

图 2-1 YOLACT 算法结构

Figure 2-1 YOLACT Algorithm Structure

2.1.2 YOLOv8-Seg 实例分割算法

YOLOv8-Seg 实例分割算法在整体流程上与 YOLACT 相同，同样是采用掩码分支与检测分支并行运行实现实例分割的检测方法，与 YOLACT 不同的是，YOLOv8 的整体网络结构更加高效。与 YOLACT 的主干网络相比，YOLOv8 的主干网络 CSPDarkNet 拥有更深的网络结构并更加高效；颈部结构相较与 YOLACT 的颈部结构，采用了更复杂与高效的颈部结构，特征融合能力更强。YOLOv8-Seg 实例分割算法结构如图 2-2

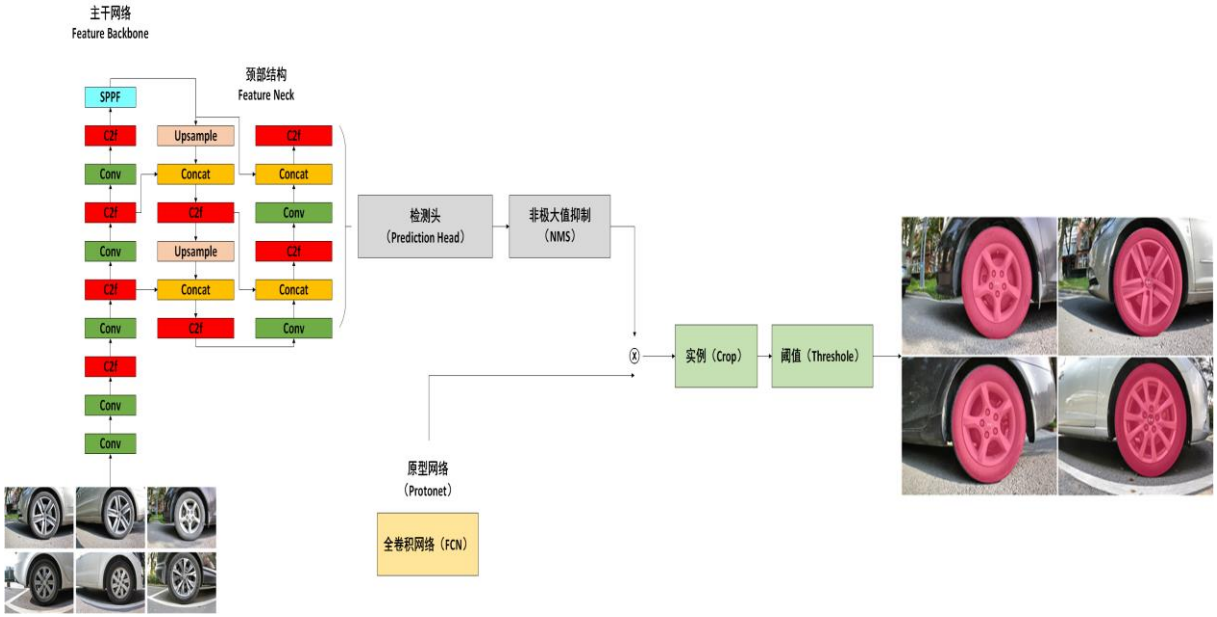

图 2-2 YOLOv8-Seg 实例分割算法结构

Figure 2-2 YOLOv8-Seg Instance Segmentation Algorithm Structure

2.2 基于深度学习的目标检测算法

2.2.1 基于二阶段的目标检测算法

基于二阶段的目标检测算法首先通过区域建议网络（Region Proposal Network, RPN）从输入图像中生成一组候选区域，这些区域被认为具有较高概率包含目标对象。此为算法的第一阶段，其目的在于初步筛选出可能包含目标的图像区域。随后，在第二阶段，算法对这些候选区域进行细致的分类和边界框回归分析，以精确确定目标的类别及其在图像中的具体位置。此类算法的经典代表包括 R-CNN、Fast R-CNN、Faster R-CNN 等。

2.2.1.1 R-CNN 目标检测算法

R-CNN（Region-based Convolutional Neural Networks）[29]是由 Ross Girshick 等人在 2014 年首次提出，在此之前目标检测通常基于手工设计的特征和传统的机器学习算法，如支持向量机（SVM）[30]和随机森林[31]。这些方法在复杂的场景中往往无法提供准确的检测结果。R-CNN 通过引入深度学习的卷积神经网络（Convolutional Neural Networks, CNN），利用其强大的特征学习能力，极大地改进了目标检测的准确性和性能。

首先，R-CNN 算法采用选择性搜索算法（Selective Search）对输入图像进行区域建议生成。在保证召回率的前提下生成约 2000 个候选区域（Region Proposal）。随后，每个候选区域需进行标准化预处理。由于卷积神经网络（Convolutional Neural Networks, CNN）对输入尺寸的敏感性，算法将各候选区域统一缩放至固定维度。经预处理的图像块随后馈入预训练的 CNN 模型进行深度特征提取，最终输出高维特征向量。在分类阶段，算法采用多类别支持向量机（SVM）分类器对特征向量进行目标判别。通过构建一对多分类策略，计算每个候选区域包含特定类别目标物体的概率分布，建立初始检测置信度评分。为消除冗余检测结果，算法引入非极大值抑制（Non-Maximum Suppression, NMS）机制。该技术基于交并比（IoU）[32]阈值对重叠候选框进行迭代筛选，保留各局部区域内置信度最高的检测框，有效解决密集检测带来的重复预测问题。最后阶段通过边界框回归（Bounding Box Regression）进行空间坐标优化，从而提升目标定位精度，输出具有最高置信度的候选框作为最终预测。R-CNN 结构如图 2-3 所示。

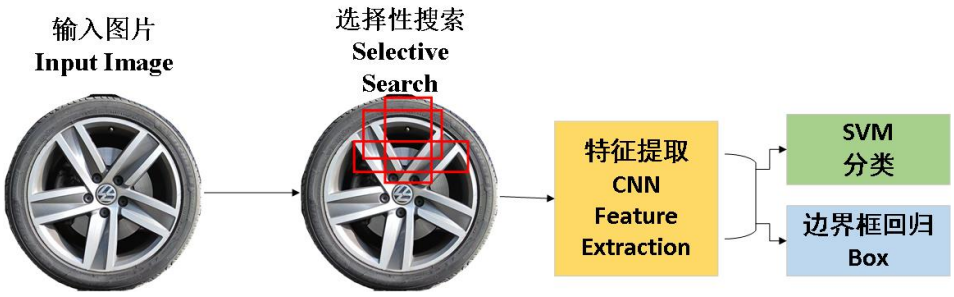

图 2-3 R-CNN 目标检测算法结构

Figure 2-3 R-CNN Object Detection Algorithm Structure

尽管二阶段检测算法在检测精度方面展现出显著优势，但其网络结构相对复杂，对计算资源的需求较高，且推理速度相对较慢，这在一定程度上限制了其在实时性要求较高的应用场景中的适用性。因此，尽管此类算法在准确性方面表现卓越，但在追求高效实时处理的场景下，其性能表现可能并不理想。

2.2.1.2 Fast R-CNN 目标检测算法

Fast R-CNN[33]相较于 R-CNN 展现出显著的改进。该算法通过共享特征图以及采用 RoI（Region of Interest）池化技术，有效避免了重复计算的问题，从而显著提升了计算效率。此外，Fast R-CNN 采用端到端的训练方式，将目标分类与边界框回归任务整合至同一网络架构中，不仅简化了训练流程，还进一步提升了检测精度。这些创新性改进使得 Fast R-CNN 在目标检测任务中表现出了更高的性能和效率。Fast R-CNN 的结构如图 2-4 所示

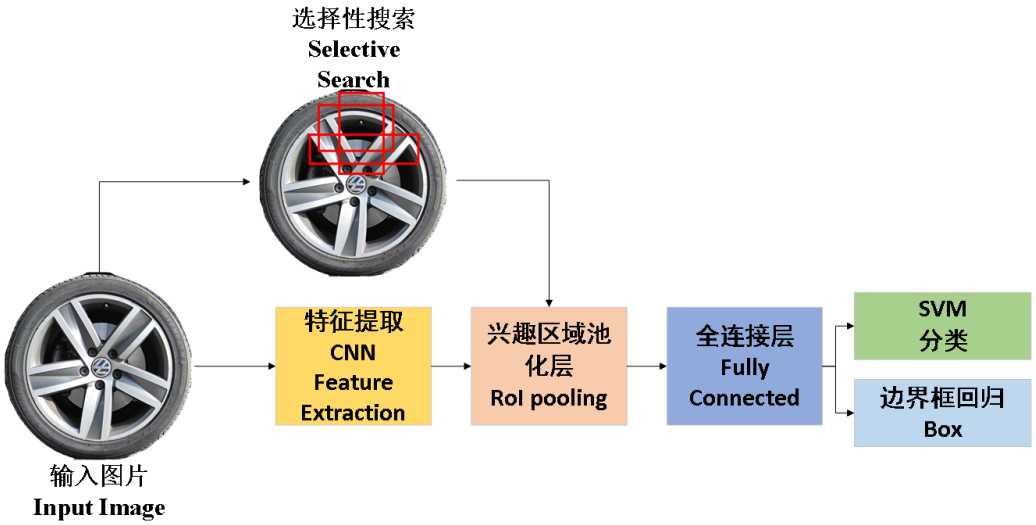

图 2-4 Fast R-CNN 目标检测算法结构

Figure 2-4 Fast R-CNN Object Detection Algorithm Structure

2.2.1.3 Faster R-CNN 目标检测算法

Faster R-CNN 相较于 Fast R-CNN 的主要优势在于引入了区域提议网络（Region Proposal Network, RPN）。这一创新性改进彻底解决了 Fast R-CNN 在生成候选区域时所面临的速度瓶颈问题。RPN 能够与检测网络共享卷积特征图，从而实现端到端的训练，进一步提升了检测速度和精度。此外,PRN 和检测网络共享同一组卷积特征，避免重复计算，Faster R-CNN 的网络结构图如图 2-5 所示。

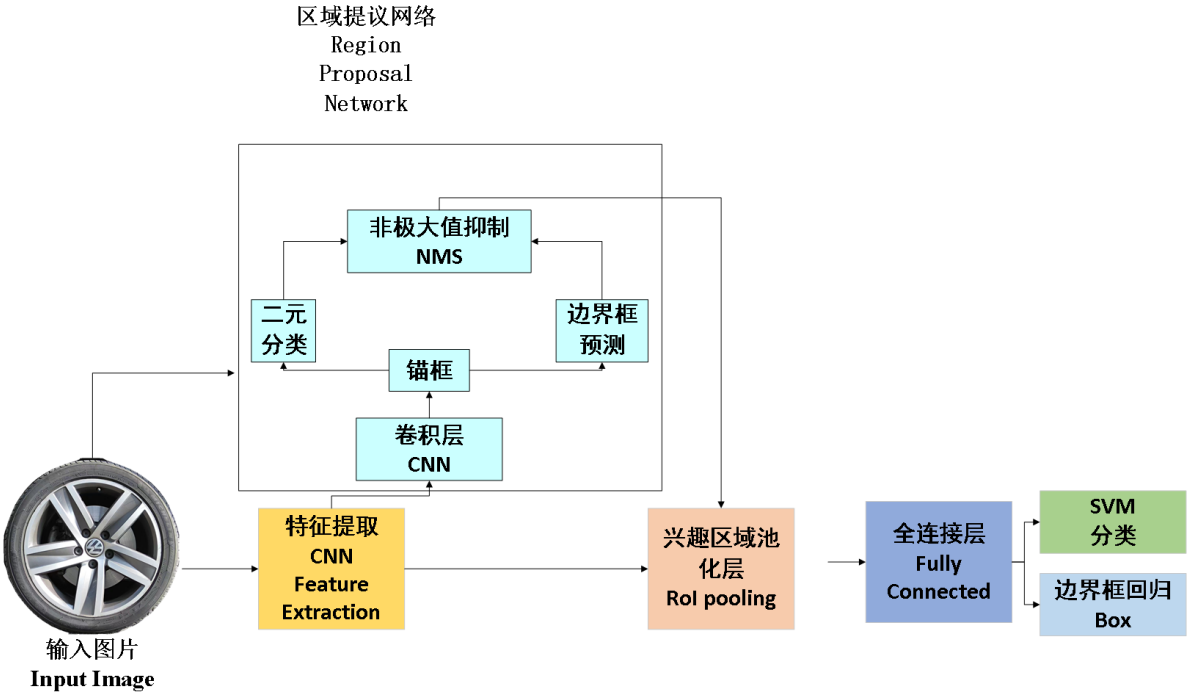

图 2-5 Faster R-CNN 目标检测算法结构

Figure 2-5 Faster R-CNN Object Detection Algorithm Structure

这些改进显著增强了 Faster R-CNN 在复杂场景下的目标检测能力，使其在多种应用领域中表现出色。

2.2.2 基于一阶段的目标检测算法

一阶段目标检测算法通过在单次前向传播过程中直接完成目标的分类与定位，省略了复杂的候选区域生成以及多阶段处理流程，从而显著降低了计算复杂度与推理时间。其模型结构简洁，适合在轻量级设备以及对实时性要求较高的场景中应用。此外，一阶段目标检测算法采用端到端的训练方式，进一步提升了检测效率。因此，一阶段目标检测算法在实时性与资源利用效率方面表现出色，尤其适用于自动驾驶、安防监控等需要快速响应的应用场景。

目前基于基于一阶段的 YOLO 系列检测算法同时兼顾精度与速度，成为了目前主要的研究方法[34]。

2.2.2.1 YOLOv5 目标检测算法

YOLOv5 目标检测算法作为一种广泛应用于一阶段目标检测的先进算法，通过将目标检测任务转化为回归问题，显著提升了检测效率[35]。其主干网络以 C3 模块为核心特征提取单元，C3 模块由三个卷积块构成，旨在通过增加网络深度和扩大感受野来增强特征提取能力；在颈部结构设计中，YOLOv5 采用了特征金字塔网络（FPN）[36]与路径聚合网络（PAN）[37]的结合方案，相较于单独使用 FPN 或 PAN，这种设计能够更有效地融合多尺度特征图的信息，从而提升模型对不同尺度目标的检测性能；

YOLOv5 的检测头采用了基于 Anchor 的检测方法，通过预定义的 Anchor 框来适应不同大小和形状的目标。该算法利用 K-means 聚类算法生成 Anchor 框，并在三个不同尺度的特征图上进行预测，每个网格位置配备三个 Anchor 框。这种设计不仅提高了模型对目标形状和大小的适应性，还进一步优化了检测精度和效率，YOLOv5 的结构图如图 2-6 所示：

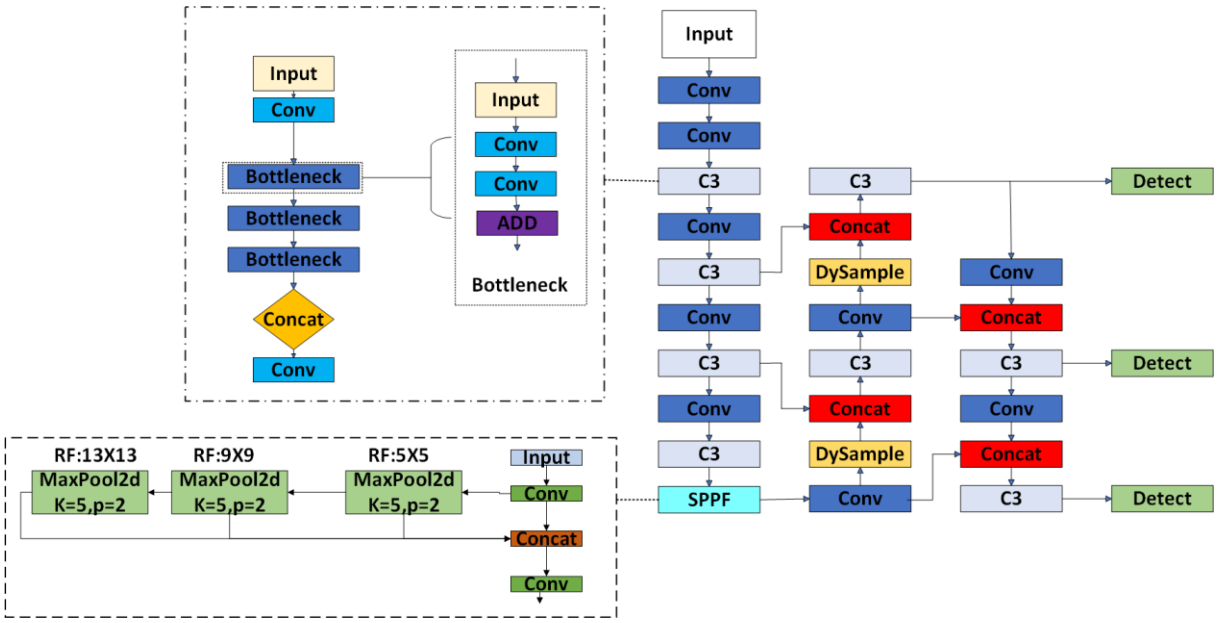

图 2-6 YOLOv5 目标检测算法结构

Figure 2-6 YOLOv5 Object Detection Algorithm Structure

YOLOv5 采用 CIoU Loss（Complete Intersection over Union Loss）[38]来作为定位损失（Bounding Box Loss），CIoU 能更全面地优化边界框回归，提升定位精度，其计算公式如下：

$$L_{CIoU} = 1 - IOU + \frac{p^2(b, b^{gt})}{C^2} + \alpha v \tag{2.1}$$

$p^2(b,b^{gt})$分别代表了预测框和真实框的中心点的欧式距离，C 代表的是能够同时包含预测框和真实框的最小闭包区域的对角线距，α 是权重参数，ν 是用来衡量宽高比一致性的参数。

置信度损失和分类损失都采用了二元交叉熵（BCEWithLogitsLoss），二元交叉熵是一种用于二分类问题的损失函数，其计算公式如下：

$$BCE = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot log(p(y_i)) + (1 - y_i) \cdot log(1 - p(y_i)) \tag{2.2}$$

N 是样本数量，y_i 是第 i 个样本的真实标签，p_i 是模型预测的第 i 个样本为正类的概率。

2.2.2.2 YOLOv8 目标检测算法

YOLOv8 目标检测算法在 YOLOv5 的基础上引入了一系列创新性改进，从而在性能和灵活性方面实现了显著提升。与 YOLOv5 相比，YOLOv8 采用了更轻量级的 C2f 模块，取代了原有的 C3 模块，这一改动不仅优化了计算效率，还进一步增强了模型的适应性。在颈部结构设计中，YOLOv8 移除了 1×1 卷积的降采样层，这一调整有助于减少计算冗余，同时保留更多细节信息。在检测头的设计上，YOLOv8 摒弃了 YOLOv5 中耦合的检测头，转而采用了解耦头设计，将分类任务与回归任务分离，从而提高了模型对不同任务的针对性处理能力，进一步提升了检测精度，YOLOv8 的网络结构图如图 2-7 所示：

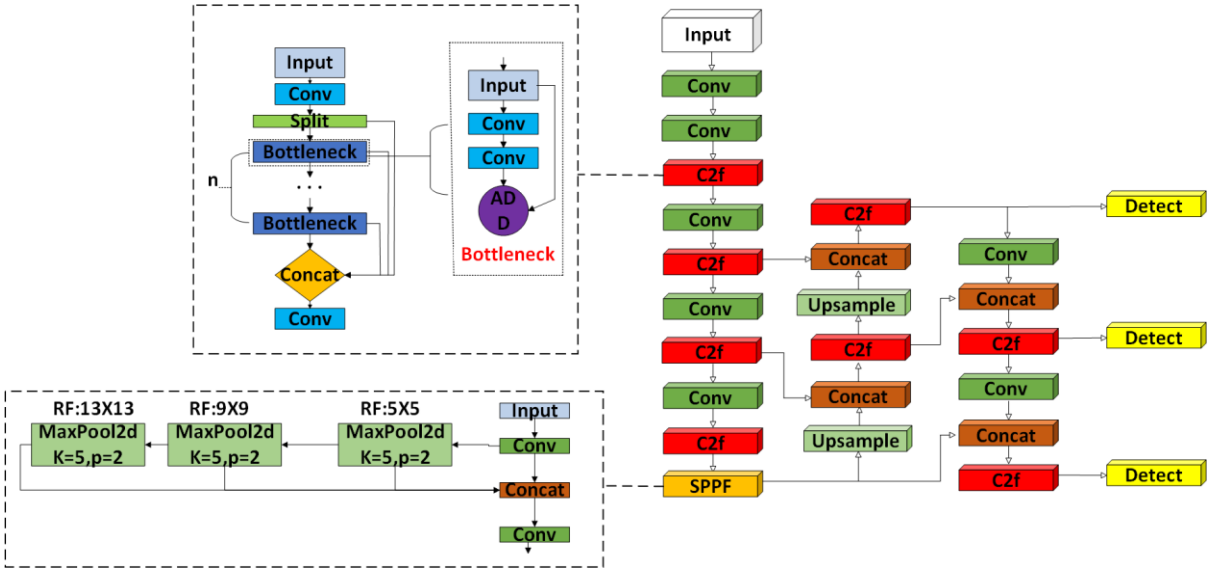

图 2-7 YOLOv8 目标检测算法结构

Figure 2-7 YOLOv8 Object Detection Algorithm Structure

YOLOv8 还引入了 Anchor-Free 机制，取代了 YOLOv5 中基于 Anchor 的检测方法。这一改进消除了对预定义 Anchor 框的依赖，使得模型能够更加灵活地适应不同大小和形状的目标，显著增强了模型在复杂场景中的检测能力。

并且，YOLOv8 在 CIoU Loss 的基础上引入了 Distribution Focal Loss(DFL)，将连续的坐标值离散化为概率分布，通过交叉熵优化目标值及其邻近值的概率，提高定位的精确度，其计算公式如下：

$$DFL(S_i,S_{i+1})=-((y_{i+1}-y)log(S_i)+(y-y_i)log(S_{i+1})) \tag{2.3}$$

S_i 表示每个点的概率，y 表示目标点。

另外，YOLOv8 取消了单独的置信度损失，与分类损失合并，同样采用二元交叉熵损失，并且在分类损失中结合了任务对齐（Task-Aligned）策略[39]，在训练过程中动态调整正负样本的权重，使分类与回归任务更协同。

综上所述，一阶段目标检测算法在兼顾检测精度与检测速度方面表现出显著优势，同时其模型结构经过轻量化设计，能够有效降低计算资源消耗，提升运行效率。鉴于轮胎胎侧关键区域目标检测任务对实时性和准确性的双重需求，本文选择基于一阶段目标检测算法的 YOLOv8 作为基础框架，并针对轮胎胎侧检测任务的特定需求对其进行相应的改进。通过这些改进，旨在进一步提升 YOLOv8 在轮胎胎侧关键区域检测任务中的性能表现，以满足实际应用场景中对快速、准确检测的要求。

2.3 图像分类算法

2.3.1 传统特征提取图像分类算法

传统特征提取图像分类方法是一种以人工设计特征为核心的图像分类技术。该方法首先利用多种特征提取算子，如边缘检测、纹理分析、颜色直方图等，借助图像处理手段从图像中提取出具有代表性的低级特征，这些特征通常涉及图像局部或全局的颜色、纹理、形状等信息。随后，采用传统机器学习算法，如支持向量机（SVM）、k 近邻（kNN）等，对提取到的特征进行分类，以确定图像的类别。然而，当面对复杂图像数据时，这种特征提取图像分类方法往往难以获取足够具有区分性的特征，并且对图像的尺度、旋转、光照等变化较为敏感，使得分类性能受到限制。此外，手工设计特征的过程需要大量的先验知识和经验，特征提取算子的调参过程通常较为复杂，且特征的表达能力有限，难以适应图像数据的多样性和复杂性。轮胎胎侧通常背景较深，特征信息较少，并且通常出现在不同光线下的场景，因此传统特征提取图像分类方法不适用于轮胎胎侧关键信息文本角度分类任务。

2.3.2 基于深度学习的图像分类算法

2.3.2.1 基于深度学习神经网络的图像分类算法

鉴于传统基于手工特征提取的图像分类方法存在诸多固有局限性，例如对图像数据内在特性的表征能力有限、分类精度易受图像复杂度影响以及泛化性能不足等问题，研究者们逐渐将目光转向更为先进的解决方案。

为有效克服人工特征提取的瓶颈，基于深度学习的图像分类算法应运而生，并迅速成为该领域的主流研究方法。这类方法通过构建具有多层非线性变换的神经网络结构，能够自适应地从原始图像数据中学习到具有高度判别性的多层次特征表示，从而避免了传统方法中依赖人工设计特征提取算子的局限性。这种端到端的学习范式不仅显著减少了繁琐的参数调优过程，而且在多个公开基准数据集上取得了突破性的性能提升，为图像分类任务提供了更为强大且灵活的解决方案。目前，例如 LeNet-5、AlexNet[40]、VGGNet 与 ResNet[41]等基于深度学习的神经网络在图像分类的应用上取得了成功，其中，在 ResNet 中引入了跳跃式连接，即残差连接的设计，在每个模块中增加一个跨层连接，让信息可以直接传递到后面的层次，从而解决深层网络训练中的梯度消失问题，在 ImageNet 竞赛中取得了

3.57% 的 top-5 错误率的优秀成绩，成为了目前常用的基于深度学习的图像分类算法的主要网络结构。

2.3.2.2 基于 YOLOv8-Cls 的 图像分类算法

YOLOv8-Cls 是 YOLOv8 算法系列中专门用于图像分类任务的模型，继承了 YOLOv8 在目标检测领域的高效性和灵活性，在特征提取网络上与 YOLOv8 相同采用了 CSPDarknet，CSPDarknet 是在传统 Darknet 架构的基础上引入 CSPNet（Cross Stage Partial Networks）理念而形成的，旨在提高特征提取的效率和准确性，同时减少计算量，Darknet 作为 ResNet 的变体，依然采用了残差连接从而复用特征。并且，在 YOLOv8-Cls 中的主要特征提取模块 C2f 采用了 ELAN 理念与残差设计相结合的架构，既增加了梯度流又通过残差设计解决退化问题与梯度问题。

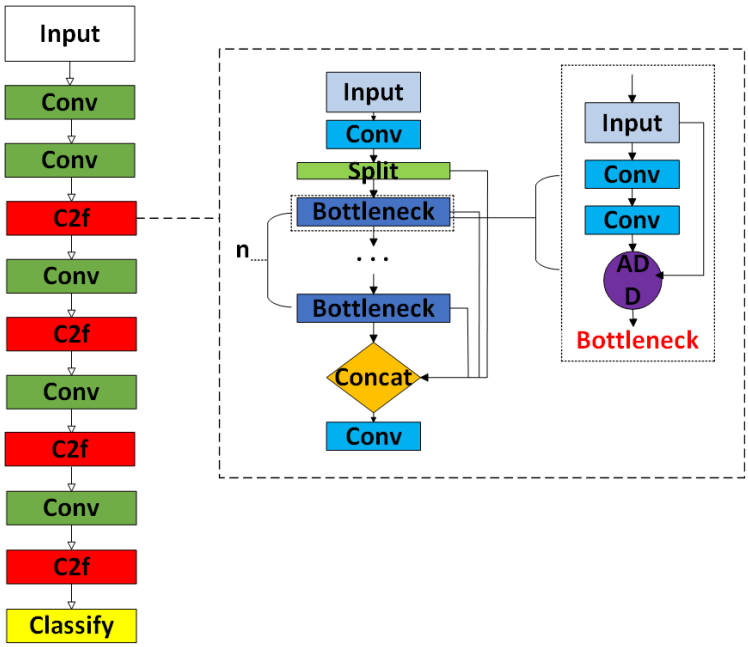

图 2-8 YOLOv8-Cls 分类算法结构

Figure 2-8 YOLOv8-Cls Classification Algorithm Structure

与 ResNet 相比，YOLOv8-Cls 的特征提取网络 CSPDarknet 在检测精度与轻量化上实现了平衡，在检测精度与 ResNet-50 相当的情况下，CSPDarknet 的参数量与计算量更低，并且在小目标分类和复杂场景下，YOLOv8-Cls 表现更优。

综述所述，与其他的分类网络相比，YOLOv8-Cls 在速度、灵活性和多任务支持方面更具竞争力，尤其适用于实时应用和资源受限的场景，因此，本文采用

YOLOv8-Cls 作为轮胎胎侧文本角度分类算法并进行特征提取网络的改进，提高模型的分类精度。YOLOv8-Cls 的网络结构图如图 2-8 所示。

2.4 基于深度学习的文本检测算法

2.4.1 基于深度学习的二分类文本检测算法

本节首先对目前常见的基于深度学习的二分类文本检测算法进行了综述，包括 EAST、SAST、DB 以及 DB++。这些算法在文本检测领域得到了广泛应用。

2.4.1.1 EAST 文本检测算法

基于深度学习的二分类文本检测算法主要分为基于回归的方法与基于分割的方法。其中，基于回归的算法通常利用锚框（anchor box）机制来实现文本区域的定位。Zhou 等人基于 PVANet 设计了 EAST（Efficient and Accurate Scene Text detection）网络模型。在该模型中，首先采用类似于语义分割的方法生成几何图，即对图像中的每个像素进行分类，判断其是否属于文本区域，从而得到文本区域的概率图。接着，根据几何形状生成四边形锚框，这些锚框能够适应不同方向和形状的文本。最后，通过非极大值抑制（Non-Maximum Suppression, NMS）筛选出最终的目标框，去除冗余框，从而实现文本区域的精确检测。

2.4.1.2 SAST 文本检测算法

SAST（Single-Shot Arbitrarily-Shaped Text Detector）文本检测算法由 Wang 等人提出，作为 EAST 文件检测算法的衍生版本，旨在高效且准确地检测自然场景中的任意形状文本。该算法基于上下文关注的多任务学习框架，通过全卷积网络（FCN）实现文本区域的语义分割和几何特征学习。SAST 的核心在于其多任务学习机制，同时预测文本中心线响应图（TCL）、中心线像素与四角点偏移量（TVO）、中心线像素与文字中心点偏移量（TCO）以及中心线到文字上下边界偏移量（TBO），从而充分利用文本的上下文信息，增强文本特征表示。此外，SAST 提出了 Context Attention Block（CAB）模块，通过注意力机制聚合上下文信息，进一步提升文本特征的表达能力。在实例分割方面，SAST 采用 Pixel-to-Quad 方法，结合高层检测信息和底层分割信息，通过像素级别的精细操作，将文本区域准确地分割出来，并以多边形的形式表达。通过这些创新设计，SAST 在多个公开数据集上取得了优异的性能，证明了其在任意形状文本检测任务中的有效性和可靠性。

2.4.1.3 DB 文本检测算法

DB（Differentiable Binarization）文本检测算法是一种基于分割的文本检测方法，其核心创新在于将传统的二值化过程转化为网络可学习、可微分的模块，从而实现了文本检测的端到端训练。该算法通过网络预测每个像素点的阈值，而不是使用固定的阈值，使得二值化过程可以随着网络的训练而不断优化。DB 算法的网络结构通常包括三个主要部分：Backbone 网络、特征金字塔网络（FPN）和 DB 模块。其中，Backbone 网络负责提取图像的多尺度特征，特征金字塔网络（FPN）对 Backbone 网络输出的特征进行增强，提高特征图的表达能力。DB 模块则以概率图和阈值图的差值作为输入，通过可微二值化技术得到近似二值图。在训练阶段，DB 算法同时对概率图、阈值图和近似二值图进行监督，以优化整个网络。此外，DB 算法引入了自适应阈值机制，每个像素点都有一个独立的阈值，能够更好地区分文本区域和背景。这种自适应阈值机制在处理复杂场景和不规则文本时表现出色。

2.4.1.4 DB++文本检测算法

DB++文本检测算法作为 DB 算法的改进版本，通过引入自适应多尺度特征融合（Adaptive Scale Fusion，ASF）模块以及空间注意力机制，进一步提升了场景文本检测的性能。其网络结构在继承 DB 算法可微二值化（Differentiable Binarization）模块的基础上，利用 ASF 模块对特征金字塔网络（FPN）输出的多尺度特征进行加权融合，从而更好地捕捉文本区域的上下文信息，增强特征表示的鲁棒性。这种改进不仅提高了对多尺度文本的检测精度，还在一定程度上保持了较高的检测效率，使其在文本检测的实际应用场景中具有广泛的应用前景。

2.4.2 基于 YOLOv8 的文本检测算法

上述常见的文本检测算法大多基于深度学习的二分类方法，仅能区分文本与背景，而无法对文本中的不同语义进行分类。在轮胎胎侧的场景中，不同区域的文字所蕴含的意义存在显著差异，因此需要更细致的分类，并从中筛选出具有重要价值的文本区域。现有的基于深度学习的二分类文本检测算法在处理轮胎胎侧文字区域的检测任务时存在局限性，难以满足实际需求，因此需要采用带有分类能力的目标检测算法进行轮胎胎侧文本检测任务。

鉴于此，为了更有效地完成轮胎胎侧文字区域的检测任务，有必要采用兼具分类与检测能力的通用目标检测算法。在此方面，已有部分学者开展了相关研究。例如上文中提及的张晨梦提出了一种基于 Faster R-CNN 的目标检测算法，专门用于检测轮胎胎侧的 DOT 信息文本区域；曲星华则利用改进的 YOLOv4 网络进行轮胎 DOT 信息的检测，然而二阶段目标检测算法 Faster R-CNN 在推理效率与轻量化上有所不足，一阶段目标检测算法 YOLOv4 虽然在推理效率与轻量化上有所改进，但是其算法性能有所不足。

综上所述，基于对任务需求的分析以及对现有算法的综合考量，本文决定采用 YOLOv8s 目标检测算法，并对其进行针对性改进，以构建适用于轮胎胎侧关键信息文本检测的算法框架。

2.5 基于深度学习的文字识别算法

本节对目前常见且广泛应用的两种文本识别算法进行了详细阐述，分别是 CRNN 与 SVTR。CRNN 作为一种经典的端到端文字识别算法，通过 CNN 提取图像特征，利用双向 LSTM 捕获上下文信息，并借助 CTC 模块实现序列解码，其在处理长文本序列时表现出色。而 SVTR 作为一种新兴的视觉模型，仅依靠视觉特征即可完成文字识别任务，其在多个基准数据集上达到了 SOTA 性能，同时在速度和精度上取得了良好的平衡。

2.5.1 CRNN 文字识别算法

CRNN（Convolutional Recurrent Neural Network）是一种端到端的文字识别算法，由卷积神经网络（Convolutional Neural Networks, CNN）、循环神经网络（Rerrent Neural Network, RNN）和连接时序分类（Connectionist Temporal Classification, CTC）三个核心模块构成。其创新性在于将图像特征提取、序列建模与标签对齐整合为一个统一的框架。该算法通过卷积神经网络（Convolutional Neural Networks, CNN）提取图像的空间特征，其典型结构采用 VGG 或 ResNet 等骨干网络。输入图像被规范化为固定高度（如 32 像素）而宽度可变的格式，以确保适应不同长度的文本序列。卷积层输出的特征图按列切分为特征向量序列，每个向量对应原图中的局部区域，由此实现从图像到序列的特征转换。

在循环层中，双向长短期记忆网络（Bi-LSTM）对特征序列进行时序建模，利用其双向结构捕获上下文依赖关系。这种设计弥补了 CNN 在长距离语义关

联上的不足，尤其适用于字符间存在强相关性的文本识别任务。最终，连接时序分类（CTC）层通过动态规划算法解决序列预测中的标签对齐问题。它允许模型在无需显式标注字符位置的情况下直接输出预测序列，有效处理了输入输出长度不一致的挑战。

相较于传统 OCR 方法，CRNN 摒弃了字符切割步骤，实现了从图像到文本的直接映射，显著提升了识别效率。其优势体现在三个方面：一是通过 CNN-RNN 联合架构融合空间与时序特征；二是利用 CTC 的柔性对齐机制降低标注成本；三是支持任意长度文本识别，泛化性强。该算法在场景文本识别、车牌识别及工业编号检测等领域得到广泛应用。后续改进工作多基于其框架进行轻量化设计或结合注意力机制优化长文本性能。在训练过程中，模型通过反向传播同步优化卷积、循环网络参数及 CTC 损失函数，形成端到端的学习范式，CTC 损失函数的公式如下：

$$CTCLoss = -logp(r \mid x) \tag{2.4}$$

其中，r 是目标标签序列，x 是由 RNN 层输出的特征序列。$p(r|x)$是在给定特征序列 x 的条件下，目标序列 r 的条件概率。

CRNN 文字识别算法结构图如图 2-9 所示：

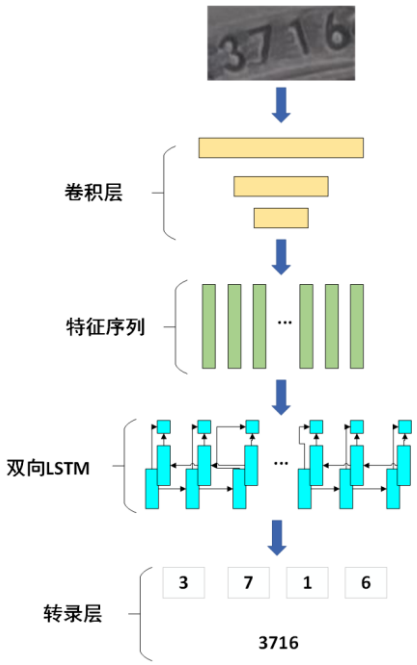

图 2-9 CRNN 文字识别算法结构

Figure 2-9 CRNN Text Recognition Algorithm Structure

2.5.2 SVTR 文字识别算法

SVTR（Scene Text Recognition with a Single Visual Model）是一种基于单一视觉模型的高效场景文本识别算法，旨在解决传统两阶段模型（特征提取 + 序列建模）的复杂性和低效性问题。该算法通过分块图像标记化框架，将图像文本分解为细粒度的字符组件（Character Component），并利用多粒度特征感知实现端到端识别。

在 SVTR 文字识别网络中，首先通过渐进式重叠补丁嵌入（Progressive Overlapping Patch Embedding），输入 $H\times W\times 3$ 尺寸的图像通过两个级联的 3×3 卷积层进行降采样，生成尺寸为 $\frac{H}{4}\times\frac{W}{4}\times D_0$ 的字符组件，其中 D_0 为初始通道数。每个组件对应图像中的局部区域，通过逐步增加特征维度，在降低计算复杂度的同时增强局部特征融合能力，随后通过三个 Stage 实现多尺度特征提取。

在 Stage1-2 中通过混合模块（Mixing Blocks）交替使用全局混合与局部混合，其中全局混合通过自注意力机制建模所有字符组件间的全局依赖关系，削弱非文本区域的干扰，而局部混合采用 $k\times k$ 的滑动窗口机制聚焦字符笔划级局部特征，例如字符的结构，增强对形变文本的鲁棒性。其中全局混合的自注意力机制的计算公式如式所示：

$$Attention(Q,K,V)=Softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{2.5}$$

在通过每个混合模块后通过卷积层降采样将特征图高度压缩为原尺寸的 1/2，宽度维度扩展为 D_1(Stage 1)，D_2(Stage 2)构建金字塔式特征表达，最终输出尺寸为 $\frac{H}{16}\times\frac{W}{4}\times D_2$。

在完成 Stage 1 与 Stage 2 后，进入 Stage 3 序列生成的流程，Stage3 通过全局平均池化将特征图高度降至 1，输出维度为 $1\times\frac{W}{4}\times D_3$ 的序列化特征，最终经线性分类器完成逐位置字符预测。相较于传统 CTC 解码，SVTR 直接通过视觉特征解码，消除序列建模的计算冗余。

SVTR 的结构图如图 2-10 所示：

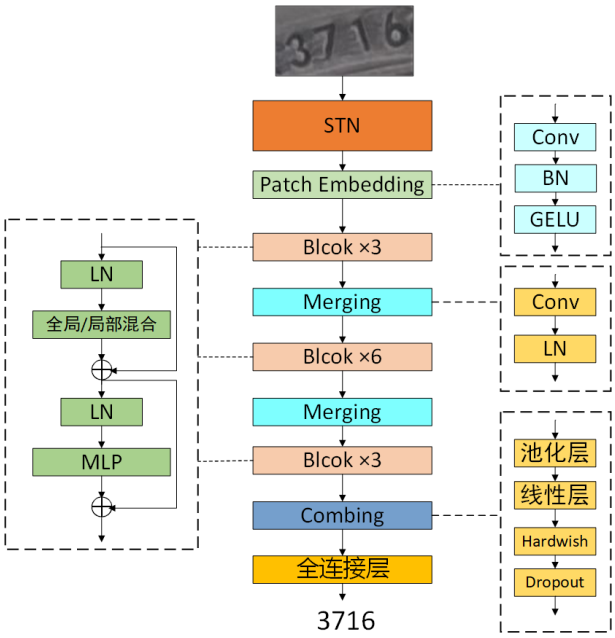

图 2-10 SVTR 文字识别算法

Figure 2-10 SVTR Text Recognition Algorithm

2.6 本章小结

本章对论文所涉及的相关技术与理论进行了全面且深入的综述。首先，对深度学习理论进行了系统介绍，并详细阐述了目标检测技术中的各类算法，包括基于二阶段和一阶段的目标检测算法。对这些算法的优缺点及其适用场景进行了深入分析。随后，重点介绍了 YOLOv8 目标检测、分类、分割算法、文本检测算法以及 SVTR 文字识别算法等技术原理，并分析了它们在不同任务中的性能表现和改进空间。通过对这些技术与理论的系统分析，为后续研究中针对轮胎胎侧关键信息检测与识别的算法设计、模型选择及优化提供了坚实的理论基础与技术支撑，确保所构建的检测系统能够高效、准确地实现对轮胎胎侧关键信息的实时检测与识别。

第 3 章 轮胎胎侧关键信息所在区域定位

本节提出了一种分阶段的轮胎胎侧关键信息所在区域定位算法，该算法将复杂的任务细分为多个相互关联且具有明确目标的子任务。每个阶段均配备专门设计的深度学习模型，这些模型根据各自任务的独特需求进行了优化，涵盖了图像分割、目标检测、分类等多个关键领域。具体而言，图像分割模型负责从复杂的背景中精确提取车辆轮胎的整体区域；目标检测模型则用于检测与分类轮胎胎侧的关键信息所在区域；分类模型用于对检测到的区域图片进行准确的角度分类，以便对区域图片进行角度矫正，使其达到可正常阅读的角度。通过这种多阶段的协同工作机制，各个模型在完成自身任务的同时，相互传递信息并整合结果，最终实现对轮胎胎侧关键信息的全面、准确检测。

3.1 轮胎区域分割

在轮胎关键信息检测系统的初步阶段，对从自然环境中获取的图像进行初步处理是至关重要的步骤。本文采用基于回归的一阶段实例分割算法进行车辆的轮胎区域的实例分割，以实现快速从复杂背景中精准分离轮胎区域的目标。此方法有效地剔除了图像中无关背景的干扰，从而显著提升了后续检测流程的准确性和效率。通过这种预处理策略，系统能够更加专注于轮胎区域，进而优化整个检测系统的性能表现。

3.1.1 改进 YOLOv8n-Seg 的实例分割算法

鉴于车辆轮胎颜色较深，与周围环境形成鲜明对比，故将轮胎区域从无关背景中分割出来无需耗费过多计算资源。基于此，本研究选用 YOLOv8 实例分割系列中规模最小的 YOLOv8n - Seg 网络模型对轮胎区域进行分割，并创新性地提出 C2f - Ghost 模块。该模块巧妙利用 GhostConv（Ghost Convolution）[42]的轻量化设计，有效降低 YOLOv8s-Seg 对计算资源的需求，从而实现网络模型的整体轻量化，在保证分割效果的同时，显著提升了模型的运行效率和实用性，改进后的 YOLOv8n-Seg 的实例分割算法的结构如图 3-1 所示：

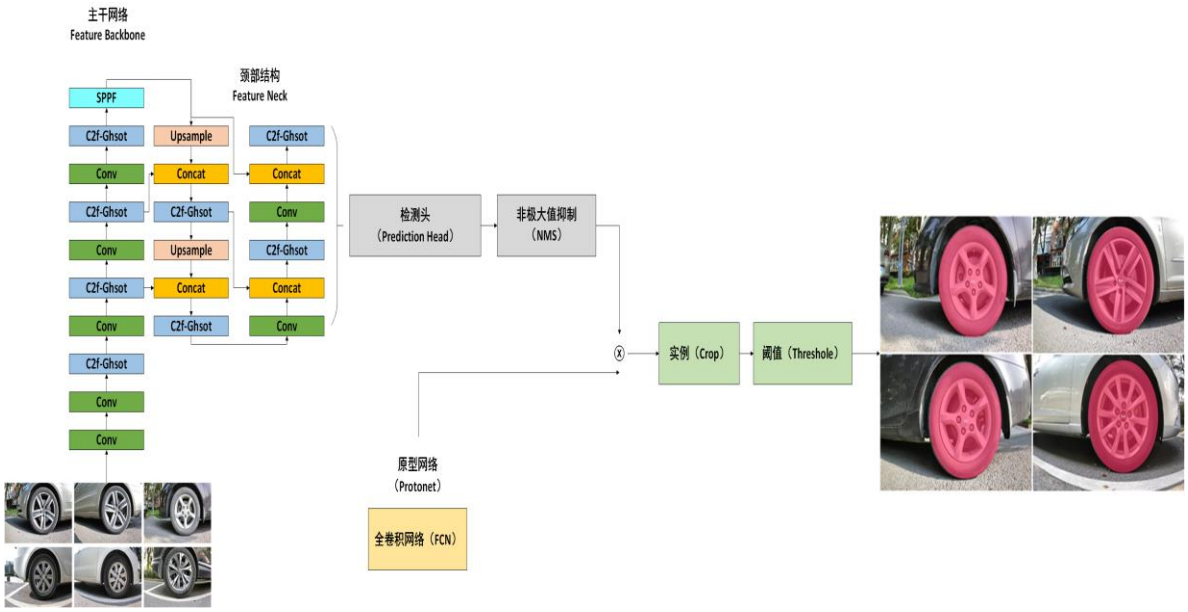

图 3-1 改进 YOLOv8n-Seg 的实例分割算法结构

Figure 3-1 Improved YOLOv8n-Seg Instance Segmentation Algorithm Structure

3.1.1.1 轻量化模块 C2f-Ghost

GhostConv 是由 Han 等人提出的一种高效卷积操作创新方法。其核心思想是利用已有的特征图（feature maps）通过低成本的线性变换生成更多的“幽灵”特征图（ghost feature maps），从而显著降低深度神经网络的计算复杂度。该方法基于特征图冗余现象，采用分阶段生成策略。首先，通过少量传统卷积核提取内在特征图（intrinsic features），随后对每个内在特征施加廉价操作（cheap operations），如深度卷积（Depthwise Convolution），生成具有相似语义信息的幽灵特征图（ghost features）。这种机制能够在保持特征表达能力的同时，减少模型参数约 30%-50%，GhostConv 的结构如图 3-2 所示：

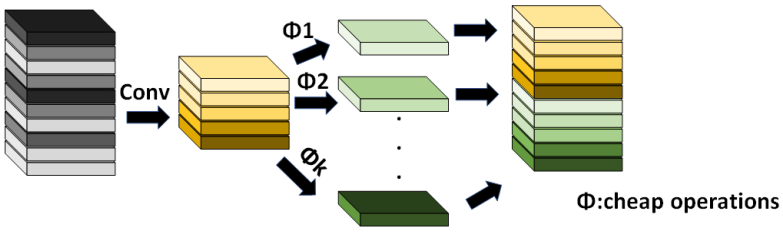

图 3-2 GhostConv 结构图

Figure 3-2 GhostConv Structure

因此将 Ghost 与 C2f 相结合，可以降低 YOLOv8s-Seg 对计算资源的需要，同时减少在网络对特征提取过程中产生的冗余，C2f-Ghost 轻量化模块的结构如图 3-3 所示：

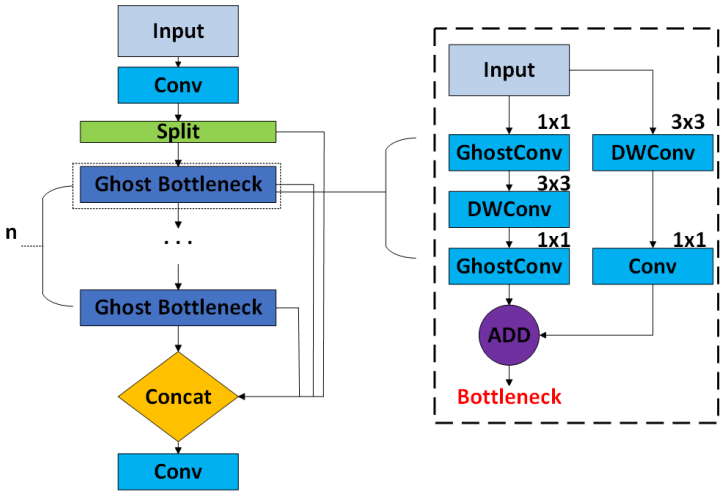

图 3-3 C2f-Ghost 轻量化模块结构

Figure 3-3 C2f-Ghost Lightweight Module Structure

3.1.1.2 实验数据分析

为验证改进后的 YOLOv8n-Seg 实例分割算法在引入 C2f-Ghost 轻量化模块后的性能提升效果，本节精心设计了一系列实验，旨在全面展示该改进措施的有效性。

本节的实验训练参数均采用 YOLOv8s-Seg 的默认训练参数，部分主要训练参数如下表 3-1 所示：

表 3-1 分割算法实验参数

Table 3-1 Experimental Parameters of the Segmentation Algorithm

参数名称	参数设置
批量大小	16
动量	0.937
权重衰弱	0.0005
学习率	0.01
迭代次数	300

为了客观评价本节基于改进 YOLOv8s-Seg 的实例分割算法模型性能，本文采用平均精度 Mean Average Precision(mAP)来评估模型的精度，其中包括回归框的 mAP 与分割掩码 mAP,共同评估模型的检测精度。通过模型的参数量 Params 来衡量模型的对计算资源的需要。mAP 的计算公式如下：

$$mAP = \frac{\sum_{i=1}^{k} AP_i}{k} \tag{3.1}$$

其中，k 代表类别数量，AP 代表平均精度。

为了对比本文所提出的改进 YOLOv8s 的轮胎胎侧关键信息检测算法与目前已有算法的性能，在采用同样数据集与训练策略的情况下进行了对比实验，实验结果如表 3-2 所示：

表 3-2 改进 YOLOv8s-Seg 的实例分割算法对比实验结果

Table 3-2 Comparison of Instance Segmentation Algorithm Results for the Improved YOLOv8s-Seg

名称	mAP@50：95(Bbox)	mAP@50：95(Mask)	参数量 /10^6
YOLACT(ResNet50)	0.898	0.99	30.7
YOLOv8n-Seg	0.987	0.99	12.0
Ours	0.987	0.99	9.7

由实验数据可知，由于轮胎在图像中占据较大比例且与背景的区分度较高，分割掩码的准确度极为精准，能够高效地完成将轮胎从背景中分割出来的任务。然而，由于不同网络主干架构的差异，模型在定位精度上存在一定的差距。YOLOv8s-Seg 实例分割算法以及本文提出的改进 YOLOv8s-Seg 实例分割算法均采用基于 DarkNet53 的主干网络，相较于 YOLACT 所采用的 ResNet50，其性能更为优越。因此，YOLOv8s-Seg 实例分割算法与本文提出的改进 YOLOv8s-Seg 实例分割算法在定位精度上均处于较高水平，二者表现相当。值得注意的是，本文提出的改进 YOLOv8s-Seg 实例分割算法引入了轻量化模块 C2f-Ghost，这使得该算法在计算资源的需求上显著降低，综合性能表现最为出色。

3.1.1.3 算法检测结果示例

图 3-4 实例分割算法检测结果示例

Figure 3-4 Example of Instance Segmentation Algorithm Detection Results

通过对图 3-4 结果示例的分析，可以清晰地观察到，本文提出的改进 YOLOv8s-Seg 实例分割算法能够以极高的精度从原始图像中定位并分割出轮胎区域。这一成果不仅体现了算法在目标分割任务中的卓越性能，更为后续的检测

流程提供了极为可靠的基础。准确的轮胎区域分割能够有效减少背景噪声的干扰，使得后续检测算法能够更专注于轮胎胎侧的关键信息，从而显著提升整个检测系统的效率与准确性。此外，改进算法在保持高精度的同时，还通过轻量化设计优化了计算资源的使用，进一步增强了其在实际应用场景中的可行性和实用性。

3.1.2 本节小结

本节选择 YOLOv8n-Seg 实例分割算法作为基线算法。在此基础上，本文提出了一种改进的 YOLOv8s-Seg 轻量化实例分割算法。该改进算法通过引入 GhostConv 与 C2f 结合的方式，构建了一种轻量化特征提取模型 C2f-Ghost。这一模型不仅有效降低了模型的参数量，而且在保持较高检测精度的同时，显著减少了计算资源的消耗。与 YOLACT 实例分割算法以及原始的 YOLOv8-Seg 实例分割算法相比，改进后的实例分割算法在总体性能上达到了最优，展现出更高的效率和更好的适应性。

3.2 轮胎胎侧关键信息区域检测算法

轮胎的关键信息检测与识别涉及三个阶段，首先对轮胎胎侧的关键信息区域检测，提取到轮胎的尺寸数据以及包含在轮胎生产日期的 DOT 信息代码，然后对区域内进行关键信息文本检测分类和字符识别。

鉴于轮胎胎侧包含了多种具有不同含义的文本，为了精确地从复杂的胎侧文本中识别并分类出轮胎尺寸文本和 DOT 信息文本所在的具体区域位置，本节深入研究了基于深度学习的目标检测算法。通过对前文综述中的目标检测算法的细致比较与评估，本研究最终选择了基于回归的一阶段目标检测算法 YOLOv8s 作为关键信息区域检测算法的核心检测框架并进行针对性的改进。

3.2.1 基于改进 YOLOv8s 的轮胎胎侧关键信息区域检测算法

尽管 YOLOv8 相较于两阶段检测算法 Faster R-CNN 在计算资源需求上有所降低，其模块复杂度依然较高。虽然 YOLOv8 相较于 YOLOv5 在检测性能上有所提升，但由于轮胎胎侧背景颜色较深且胎侧文字为压印文字，特征提取难度较大且特征信息较少，YOLOv8 的特征提取能力仍存在不足。因此，针对上述问题，本节提出了一种基于 YOLOv8 改进的轮胎胎侧关键信息区域检测算法。在综合考虑 YOLOv8 中五种目标检测模型（n、s、m、l、x）的计算量及参数量后，本文选择 YOLOv8s 作为进一步优化的模型算法。

首先，本文利用一种新的主干网络 RFMLNet。该网络通过引入高效通道注意力机制 MLCA（Mixed Local Channel Attention）[43]提取通道特征，并结合 RFAConv[44]感受野卷积模块，不仅关注感受野空间特征，还为大尺寸卷积核提供了有效的注意力权重，从而显著提升了网络性能。

在颈部结构方面，本文提出了一种轻量化的 DyFCGSlimNeck。该结构结合了 FasterNet[45]中的 FasterBlock 模块与 CGLU（Convolutional GLU）[46]注意力机制，提出了 VoVFCG（VoV FasterBlock CGLU）特征提取模块。该模块在保证轻量化的同时，能够融合更多的特征信息。此外，引入了 DySample[47]上采样模块，以实现更高效的上采样操作。

同时，本文引入了上下文注意力机制 CAA（Context Anchor Attention）[48]，以增强网络对特征上下文信息的提取能力，从而提高网络对不同关键信息所在区域的分类能力。

改进后的 YOLOv8s 的轮胎胎侧关键信息区域检测算法的网络结构如图 3-5 所示：

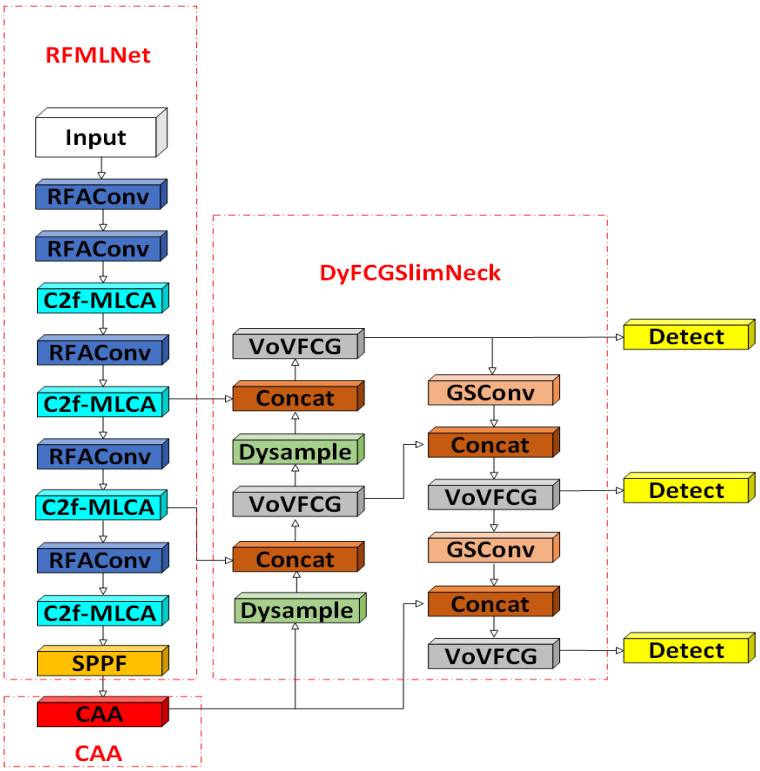

图 3-5 改进 YOLOv8s 的轮胎胎侧关键信息区域检测算法结构

Figure 3-5 Improved YOLOv8s Tire Sidewall Key Information Region Detection Algorithm Structure

3.2.1.1 RFMLNet 网络结构

在 YOLOv8 主干网络中，普通卷积承担着特征提取的重要任务。然而，普通卷积在特征提取过程中存在明显的局限性，其卷积核大小固定，无法根据特征所在位置的差异进行灵活调整，同时，也未对不同特征的重要程度进行区分，这无疑限制了检测网络对特征提取的能力[49]。轮胎胎侧关键信息所在区域的特征相对匮乏，为有效解决这一问题，提升主干网络的特征提取能力，进而优化模型性能，本文在主干网络中引入了 RFAConv。RFAConv 不仅完全解决了卷积核共享参数的问题，还充分考虑了感受野中每个特征的重要性，其结构如图 3-6 所示。

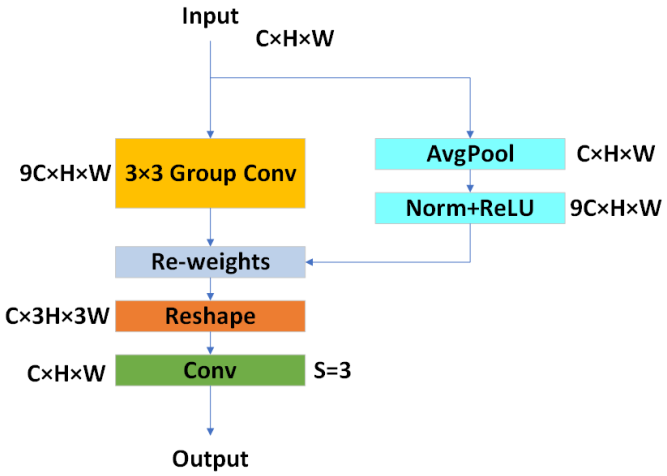

图 3-6 RFAConv 模块

Figure 3-6 RFAConv Module

RFAConv 的设计包含两个分支。在第一个分支中，平均池化层被用于提取每个感受野内的特征信息，随后通过分组卷积对特征进行融合。在此过程中，对特征的重要程度进行分级与排序，从而确定特征的权重值。在第二个分支中，采用 3×3 的分组卷积进行特征提取。根据感受野的大小，动态选择相应的分组卷积尺寸以生成展开特征。最终，将两个分支提取到的特征进行融合，依据权重值确定最终提取的特征，并通过调整大小得到最终的输出。其计算公式如下：

$$Output = Conv(Reshape(GroupConv(x) \times ReLU(Norm(AvgPool(x)))))$$

(3.2)

在主干网络的特征提取过程中，大量特征提取不足，这最终导致关键信息的丢失。通常的解决办法是引入注意力机制以提取更多的特征。然而，现有的注意力机制，如 Squeeze-and-Excitation（SE）[50]和 Efficient Channel Attention（ECA）

[51]，将整个通道特征图压缩为单个值，忽略了每个通道内的空间信息。因此，本文在主干网络中引入了混合局部通道注意力（Mixed Local Channel Attention, MLCA），并与 C2f 相结合，组成 C2f-MLCA 特征提取模块，其结构如图 3-7 所示。

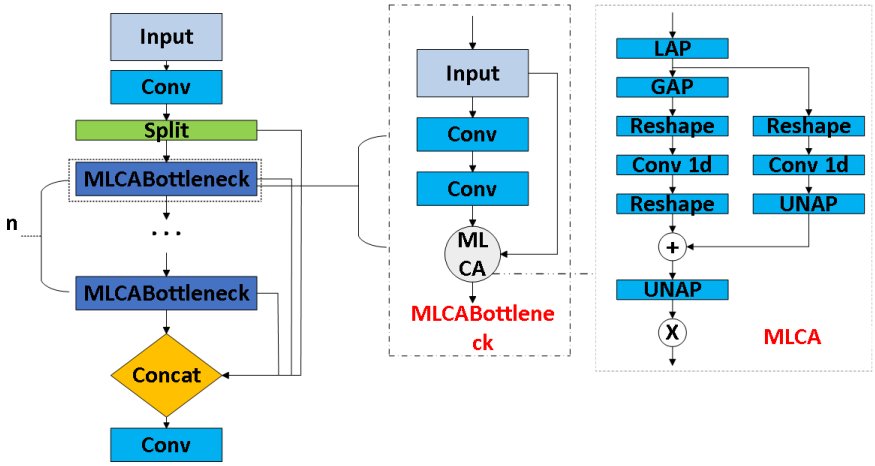

图 3-7 C2f-MLCA 模块

Figure 3-7 C2f-MLCA Module

在混合局部通道注意力（Mixed Local Channel Attention, MLCA）模块中，首先采用局部平均池化（Local Average Pooling, LAP）提取局部空间信息。随后，将特征分为两个分支进行计算。第一个分支通过全局平均池化（Global Average Pooling, GAP）提取全局特征信息；第二个分支则保留局部空间信息。两个分支在经过一维卷积处理后，通过平均反池化（Anti Average Pooling, UNAP）操作恢复至原始分辨率，并进行信息融合，以实现混合局部注意力的效果。将该模块与 C2f 相结合，使得 C2f-MLCA 模块在特征提取过程中具备混合局部注意力的能力，从而提升模型的整体性能。

3.2.1.2 上下文锚点注意力机制 CAA

在特征提取过程中，上下文信息蕴含着丰富的语义化信息。然而，轮胎胎侧存在大量文字，但并非所有文字区域均为关键信息区域。因此，需要加强语义信息的提取，以增强模型的检测与分类能力。基于此，本文引入上下文锚点注意力机制（Contextual Anchor Attention, CAA），以强化上下文信息的提取，提升模型的泛化性能，从而有效提高模型的推理能力，其结构图如图 3-8 所示。

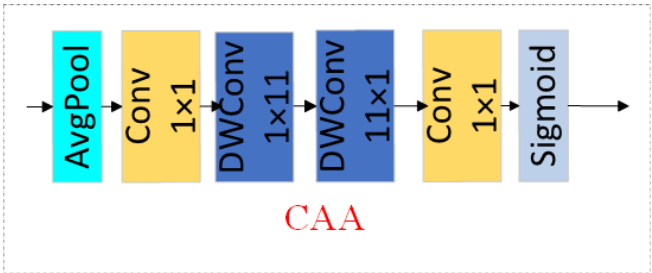

图 3-8 CAA 注意力机制模块

Figure 3-8 CAA Attention Mechanism Module

在 CAA 注意力机制中，首先通过平均池化操作与 1×1 的卷积操作提取局部区域特征。接着，利用 1×11 和 11×1 的深度卷积近似实现大核卷积，以充分提取上下文特征。最终，通过 1×1 的卷积操作与 Sigmoid 函数对特征进行融合并输出。

3.2.1.3 DyFCGSlimNeck 轻量化颈部网络

在 YOLOv8 的颈部结构中，采用了特征金字塔网络（Feature Pyramid Network, FPN）与路径聚合网络（Path Aggregation Network, PAN）相结合的设计。相较于单独使用 FPN 或 PAN，这种设计能够融合更多的特征图信息，然而，由于其结构更为复杂，导致计算资源的需求显著增加。基于此，Li 等[52]在 YOLOv5 以及 YOLOv8 颈部结构的基础上提出了 SlimNeck，通过采用 GSConv 与 VoVGSCSP 特征提取模块来降低参数量与计算量。然而，这种方法在特征提取方面仍存在一定的局限性。此外，在 YOLOv8 颈部网络的上采样操作中，采用了最近邻插值法，但这种方法会导致图像像素的丢失，进而产生类似马赛克的效果，从而在特征提取过程中造成信息的缺失与错误。

为了解决上述问题，本文在 SlimNeck 的基础上改进提出了 DyFCGSlimNeck 轻量化颈部网络。在 DyFCGSlimNeck 网络中，在 VoVGSCSP 模块的基础上加入 Convolutional GLU（CGLU）注意力机制与 FasterBlock 模块，构建了 VoVFCG 模块。同时，引入了 Dysample 上采样模块进行上采样操作，以避免图像特征的丢失。

（1）轻量化特征融合模块 VoVFCG

在 YOLOv8 中，采用 C2f 模块进行特征融合，但其结构包含大量的卷积操作，导致计算资源的需求较高。在 SlimNeck 中，采用 VoVGSCPS 模块进行特征融合，相较于 C2f 模块，其表现更为轻量化，但其特征融合能力仍有待提高。因

此，本文为了在更好地融合特征的基础上加强轻量化设计，提出了 VoVFCG 轻量化特征融合模块。该模块采用轻量化的 FasterBlock 与 CGLU 注意力机制相结合，构建了 Faster_CGLU 模块，以实现这一目标。GLU 是一种通道混合器，在许多深度学习任务中相较于多层感知机（MLP）展现出了更优的性能。GLU 由两个线性操作组成，其中一个映射通过门控机制激活。与 SE（Squeeze-and-Excitation）机制不同，每个 Token 的信号均来自 Token 本身。Shi Dai 等通过研究发现，在 GLU 的门控分支的激活函数之前添加一个 3×3 深度卷积，可以使其结构具备注意力机制的特点，并将其转换为基于最近邻特征的门控通道注意力机制，即卷积门控单元（Convolutional GLU，CGLU）。FasterBlock 作为 FasterNet 的组成模块，由一个 3×3 的部分卷积（PConv）与两个 1×1 的普通卷积构成。PConv 的设计方法是对部分通道进行卷积以提取空间特征，同时保持剩余通道的不变状态，这种方法显著减少了计算资源的需求。VoVFCG 轻量化特征融合模块的结构图如图 3-9 所示。

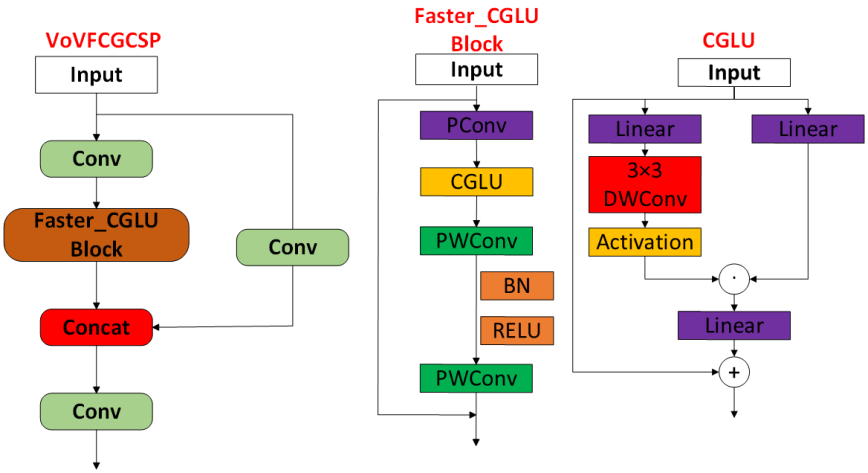

图 3-9 VoVFCG 模块

Figure 3-9 VoVFCG Module

（2）动态上采样模块 Dysample

在 YOLOv8 中采用的最近邻插值法会导致图像出现马赛克化现象，进而造成特征的丢失。为了解决这一问题，本文引入了基于动态点采样的 Dysample 上采样方法。相较于目前常用的基于内核的上采样改进方法，例如 CARAFE 上采样方法[53]，Dysample 上采样方法在提升融合网络对图像的处理效果方面表现得更加轻量化、简洁且高效。

在 Dysample 方法中，首先通过线性操作将输入特征映射为逐点的位置偏置，随后将其与输入特征的点坐标相加，从而获得新的采样位置。这种设计确保了上采样后的点在输入特征图上呈均匀分布，但同时也可能引发采样位置重叠的问题[54]。因此，因此，在经过 linear 操作生成最初的偏移量后，在 pixel shuffle 方法前，乘以 0.25 的参数用来缓解这个重影的负面效果，生成最终的偏移量 O。而采样集 S 为偏移量 O 与初始采样网格 G 的和，其结构如图 3-10 所示。

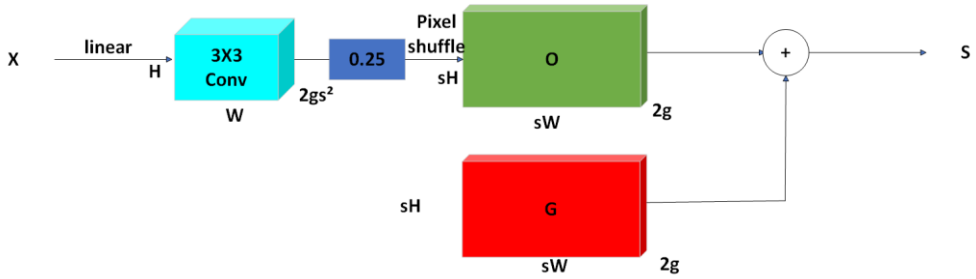

图 3-10 DySample 结构图

Figure 3-10 DySample Structure

Dysample 的计算公式如式所示：

$$O = Pixelshuffle(0.25(\mathrm{Conv}(linear(x)))), S = G + O \tag{3.3}$$

（3）轻量化卷积模块 GSConv

普通标准卷积(Conv)模块采用的是密集通道卷积操作，需要的计算资源较大，目前常见的轻量化卷积模块，如深度卷积（DWConv）采用的是利用分组通道的方式减少计算量与参数量，但是这种办法降低了模型对特征的提取能力和对特征的融合能力。为了解决上述问题，Li 等人提出了一种由普通标准卷积与深度可分类卷积共同构成的轻量化卷积模块 GSConv,其对特征的提取能力可以和普通标准卷积相媲美，但是需要的计算资源仅仅为普通标准卷积的一半。

在 GSConv 中，输入的特征图通道数为 C1，首先采用一个 Conv 模块进行初步的特征提取，输出一个通道数为 C2/2 的特征图 T1，然后将其送入一个 DWConv 中进行进一步的特征提取，得出一个新的特征图 T2，并且保持通道数 C2/2 不变，然后对特征图 T1 与特征图 T2 进行连接（Concat）得到连接的特征图 T3，最后对特征图 T3 进行混合洗牌（shuffle）操作得到通道数为 C2 的最终输出特征图。GSConv 的结构如图 3-11 所示。

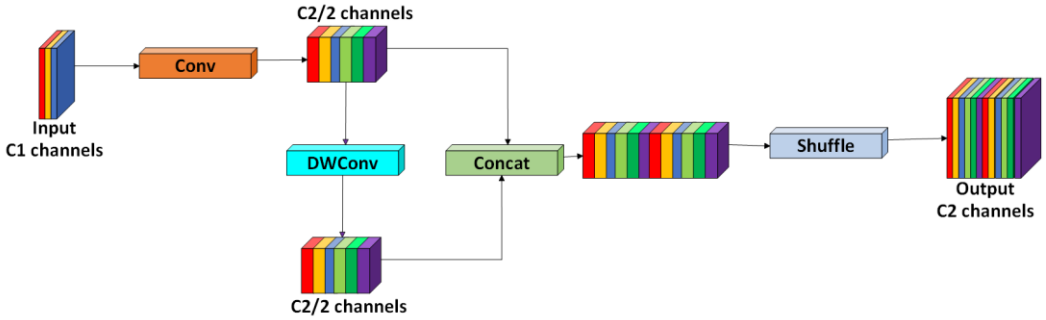

图 3-11 GSConv 模块

Figure 3-11 GSConv Module

3.2.1.4 实验数据分析

为了系统地评估 YOLOv8s 算法在引入 RFMLNet 改进主干网络、上下文锚点注意力机制以及 DyFCGSlimNeck 轻量化颈部网络后的性能提升效果，本节设计了一系列实验。这些实验包括了对比试验与消融实验，分别验证各个改进点的有效性及其对整体算法性能的具体影响。

本节使用的训练参数为 YOLOv8s 的初始化参数，其主要实验参数如表 3-3 所示。

表 3-3 关键区域检测算法实验参数

Table 3-3 Experimental Parameters of the Key Area Detection Algorithm

参数名称	参数设置
批量大小	16
动量	0.937
权重衰弱	0.0005
学习率	0.01
迭代次数	500
训练图片大小	640

为了客观评价本节基于 YOLOv8s 改进的轮胎胎侧关键信息区域检测算法模型性能，本文采用平均精度 Mean Average Precision(mAP)来评估模型的精度，mAP 计算标准结合了 Precision (P)和 Recall (R)，是目前常用的评价模型检测效果的指标。通过模型的参数量 Params 和模型的计算量 GFLOPs 来衡量模型的大小。mAP、P、R 的计算公式如下所示：

$$R = \frac{TP}{(TP + FN)} \tag{3.4}$$

$$P = \frac{TP}{(TP + FP)} \tag{3.5}$$

$$\mathrm{m}AP = \frac{\sum_{i=1}^{k} AP_i}{k} \tag{3.6}$$

其中：TP 为真正例；FP 为假正例；TN 为真负例；FN 为假负例。

为了对比本文所提出的改进 YOLOv8s 的轮胎胎侧关键信息检测算法与目前已有算法的性能，在采用同样数据集与训练策略的情况下进行了对比实验，实验结果如表 3-4 所示：

表 3-4 轮胎胎侧关键信息区域检测算法对比实验结果

Table 3-4 Comparison of Detection Algorithm Results for Key Information Areas on Tire Sidewalls

名称	P	R	mAP@50	mAP@50：95	参数量 /10^6	计算量 (GFLOPs)
YOLOv3-Tiny	0.712	0.503	59.2%	29.8%	8.68	12.9
YOLOv5s	0.556	0.363	43.1%	27.4%	7.02	15.8
YOLOv7-Tiny	0.441	0.396	40.3%	24.3%	6.02	13.1
YOLOv8s	0.718	0.563	62.8%	42.5%	11.13	28.6
YOLOv9s	0.742	0.635	70.6%	51.1%	7.73	25.8
YOLOv10s	0.653	0.483	56.6%	35.6%	8.04	24.4
YOLO11s	0.707	0.589	60.8%	41.8%	9.41	21.3
YOLO12s	0.591	0.608	60.4%	40.3%	9.07	19.3
Ours	0.743	0.675	72.9%	51.2%	12.1	26.5

通过与其他算法对比实验可以看出，本文所提出的改进 YOLOv8s 的轮胎胎侧关键信息检测算法的综合性能达到了最佳，与 YOLOv5s 与 YOLOv8s 相比，本文所提出的改进 YOLOv8s 的轮胎胎侧关键信息检测算法的 mAP@50 提高了 29.8 个百分点与 10.1 个百分点，mAP@50：95 分别提高了 23.8 个百分点与 8.7 个百分点，计算量相较 YOLOv8s 下降了 7%；与 YOLOv3-Tiny 与 YOLOv7-Tiny 相比，本文所提出的改进 YOLOv8s 的轮胎胎侧关键信息检测算法的 mAP@50 提高了 29.8 个百分点与 10.1 个百分点，mAP@50：95 分别提高了 23.8%与 8.7%；与 YOLOv9s 与 YOLOv10s 相比，在 mAP@50 分别提高了 2.3 个百分点与 16.3 个百分点，mAP@50：95 分别提高了 14.9 个百分点与 0.1 个百分点；与目前 YOLO 的最新发展 YOLO11s 与 YOLO12s 相比，本文所提出的改进 YOLOv8s 的轮胎胎侧关键信息检测算法的 mAP@50 分别提高了 12.1 个百分点与 12.5 个百分点，mAP@50：95 分别提高了 9.4 个百分点与 10.9 个百分点，证明了本节提出的改进算法在精确度上的优越性。

对比实验结果表明，本节提出的 RFMLNet 主干网络在采用多种注意力机制后，相较于 YOLO 系列算法的原主干网络，利用不同注意力机制模块对特征提取能力的差异化增强，能够更有效地提取轮胎胎侧关键信息所在区域的特征信

息。此外，网络在特征金字塔模块后引入的 CAA 注意力机制，显著提升了模型对特征上下文信息的提取能力。最终，DyFCGSlimNeck 结构通过采用 SlimNeck 的结构、动态上采样以及轻量化特征融合模块 VoVFCG，在降低计算量的基础上，优化了特征融合效能，其融合能力优于 YOLO 系列算法的 Neck 结构，从而为模型精度的提升提供了有力支撑。

为了验证模型以及各个改进的有效性，在保持各个参数不变的情况下，在 YOLOv8s 的基础上添加各个改进点，分别对模型进行训练与验证，消融实验的结果如表 3-5 所示：

表 3-5 轮胎胎侧关键信息区域检测算法消融实验结果

Table 3-5 Ablation Study Results of Detection Algorithm for Key Information Areas on Tire Sidewalls

RFMLNet	CAA	DyFCGSlimNeck	P	R	mAP@50	mAP@50：95	参数量 /10^6	计算量 (GFLOPs)
			0.718	0.563	62.8%	42.5%	11.13	28.6
√			0.653	0.65	68%	44.6%	11.17	28.9
	√		0.636	0.562	62.8%	43.6%	11.66	28.9
		√	0.616	0.609	62.9%	43.7%	9.92	24.4
√	√		0.839	0.537	70.5%	46.7%	11.71	29.3
√		√	0.65	0.675	73.1%	49.8%	9.97	24.8
	√	√	0.686	0.615	67.5%	46.5%	10.46	24.8
√	√	√	0.743	0.675	72.9%	51.2%	12.10	26.5

通过消融实验可以看出，在添加了 RFMLNet 后，模型的 mAP@50 与 mAP@50：95 分别提高了 5.2 个百分点与 2.1 个百分点，但是参数量与计算量带来了小幅的上升，其原因是在于 RFMLNet 相较于 YOLOv8 原有的主干网络相比，RFAConv 与 C2f-MLCA 的多种注意力机制模块的组合可以提取到更多的特征，从而提高了模型的性能，同时由于加入了注意力机制，带来了额外的计算代价；在添加了 CAA 注意力机制模块后，mAP@50：95 提高了 1.1 个百分点，其原因是在于通过添加上下文注意力机制 CAA，检测模型对上下文信息的获取能力有所提高，而上下文信息对模型的分类能力至关重要，因此提高了模型的检测能力，但是同样带来了少许计算资源的需要；在添加了 DyFCGSlimNeck 后，模型的 mAP@50 与 mAP@50：95 分别提高了 0.1 个百分点与 1.2 个百分点，同时参数量与计算量分别下降了 10.9%与 14.7%，这是由于在加入大感受模块 VoVFCG 与 Dysample 上采样模块后，更好的融合了主干网络提取到的特征，从而提高了模型的性能，同时因为添加了轻量化模块 FasterBlock，从而降低了计算资源的需要。当同时添加 RFMLNet 与 CAA 后，mAP@50 与 mAP@50：95 分别

提高了 7.7 个百分点与 4.2 个百分点；当添加 RFMLNet 与 DyFCGSlimNeck 的组合时，mAP@50 与 mAP@50：95 分别提高了 10.3 个百分点与 7.3 个百分点；当 DyFCGSlimNeck 组合 CAA 注意力机制模块使用时 mAP@50 与 mAP@50：95 分别提高了 4.7 个百分点与 4 个百分点，在同时添加了三个改进点组成模型后，模型的综合性能达到了最佳，mAP@50 与 mAP@50：95 分别提高了 10.1 个百分点与 8.7 个百分点，计算量下降了 7%。

表 3-6 主干网络对比

Table 3-6 Comparison of Backbone Networks

名称	mAP@0.5	mAP@0.5:0.95	参数量 Params	计算量 GFLOPs
YOLOv8s+RFMLNet	68%	44.6%	11.17M	28.9
YOLOv8s+FasterNet	41.9%	22.6%	8.62M	21.7
YOLOv8s+EMO	66.1%	41.7%%	10.63M	20.1

表 3-7 特征融合网络对比

Table 3-7 Comparison of Feature Fusion Networks

名称	mAP@0.5	mAP@0.5:0.95	参数量 Params	计算量 GFLOPs
YOLOv8s+DyFCGSlimNeck	62.9%	43.7%	9.92M	24.4
YOLOv8s+ASF	55.6%	37.5%	11.29M	30.1
YOLOv8s+BiFPN	62.6%	42.8%	7.37M	25.0

如表 3-6 与表 3-7 所示，相较于轻量化主干网络 FasterNet 以及基于 Transformer 的主干网络 EMO，本节所提出的 RFMLNet 在精确度方面表现最为优异。与 FasterNet 相比，RFMLNet 的 mAP@50 和 mAP@50:95 分别提升了 26.1 个百分点和 22 个百分点；与 EMO 相比，RFMLNet 的 mAP@50 和 mAP@50:95 分别提升了 1.9 个百分点和 2.9 个百分点。这一结果表明，FasterNet 作为一种轻量化网络，在深色背景下对轮胎胎侧关键区域特征的提取能力存在不足；而 EMO 主干网络虽然基于 Transformer 架构，检测精度优于 FasterNet，但在特征提取方面，相较于本节提出的融合多种注意力机制的 RFMLNet，仍存在一定的局限性。

与目前应用较为广泛的特征融合网络 ASF 和 BiFPN 相比，本文提出的 DyFCGSlimNeck 在动态采样和大核心模块的支持下，精度实现了显著提升，达到了最佳水平。同时，DyFCGSlimNeck 通过使用 GSConv，在一定程度上有效控制了计算资源的增长。与 ASF 和 BiFPN 相比，DyFCGSlimNeck 的 mAP@50 分别提升了 7.3 个百分点和 0.3 个百分点，mAP@50:95 分别提升了

6.2 个百分点和 0.9 个百分点。这些数据进一步证实了 DyFCGSlimNeck 在轮胎胎侧关键信息所在区域的特征融合能力方面具有显著优势。

3.2.1.5 算法检测效果对比

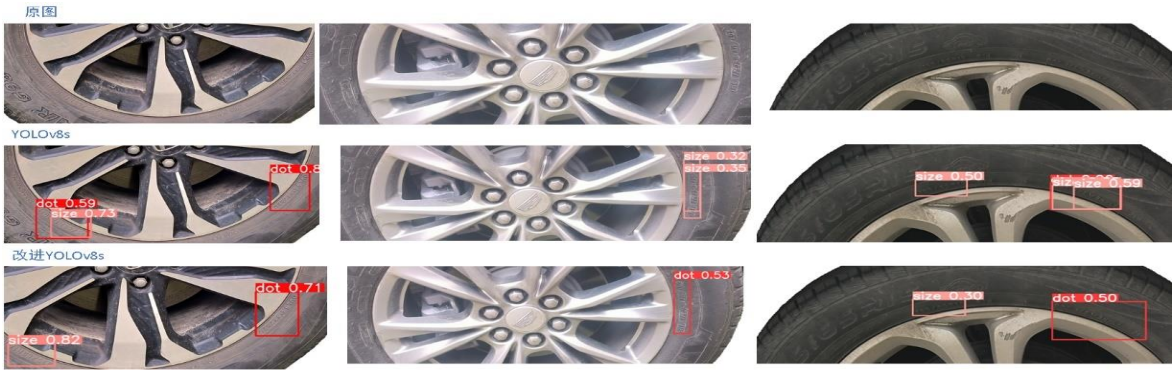

图 3-11 轮胎胎侧关键区域检测算法效果对比

Figure 3-11 Comparison of Tire Sidewall Key Area Detection Algorithm Effects

由图 3-11 检测结果对比可以看出，原始算法 YOLOv8s 容易出现漏检误检的现象，并且出现两个类别框相重合的现象，原因是对轮胎胎侧关键信息区域的特征提取不充分导致的；本文提出的轮胎胎侧关键信息区域检测算法在轮胎胎侧关键信息区域检测中表现优秀，解决了上述算法无法解决的漏检误检问题，准确的检测与分类出相应的关键信息所在区域。

3.2.2 本节小结

本节提出的基于 YOLOv8s 改进的轮胎胎侧关键信息区域检测算法，在原有 YOLOv8s 架构的基础上，引入了多种创新机制，包括 RFMLNet 主干网络（融合多种注意力机制）、DyFCGSlimNeck 轻量化点采样颈部融合网络以及上下文锚点注意力机制（Contextual Anchor Attention, CAA）。相较于原始的 YOLOv8s 目标检测算法，本节所提出的改进算法在检测效果上表现出显著的优越性。在与其他目标检测算法的对比中，本文算法在检测精度方面达到了最佳，能够有效减少漏检和误检的问题，同时并未引入过多的计算代价，与 YOLOv9 与 YOLOv10 相比，本节提出的改进算法同样具有最优秀的综合性能。

3.3 轮胎胎侧关键信息所在区域角度分类

轮胎胎侧关键信息文本角度分类是文本检测与识别处理环节中的重要步骤。在诸多实际应用场景中，轮胎图像的获取角度往往并非理想的从左到右的正常阅读方向，这使得文本信息呈现出非正常的倾斜状态。文本角度分类算法旨在针对此类场景，对文本检测算法所定位出的文本区域图片进行精准的角度分类。通过

这一操作，能够为系统提供关键的角度信息，以便依据分类结果对那些处于非正常阅读角度的文本图片实施有效的矫正措施，从而确保后续文本识别、信息提取等环节能够在符合常规阅读习惯的图像基础上顺利开展，极大地提升了整个轮胎图像处理流程的准确性和可靠性。

在图像处理与模式识别领域，图像分类算法主要划分为两大类别：传统特征提取分类方法与基于深度学习的图像分类算法。本节通过对各类算法的性能、适用场景以及优缺点的全面分析，经综合考量与深入研究，本节最终确定选用 YOLOv8-Cls 作为针对轮胎胎侧关键信息文本角度分类的算法基础，并结合轻量化特征提取网络 StarNet[55]提高模型对角度的分类准确度。

为实现高效精准的文本角度分类，本节选定基于一阶段的图片分类算法 YOLOv8-Cls 作为基础算法框架，并对其进行了针对性的改进优化。用 StarNet 作为特征提取网络。StarNet 凭借其独特的网络结构设计，能够深度挖掘图片中的关键特征信息，在显著提高分类精度的同时，有效降低了对计算资源的需求。

3.3.1 基于 YOLOv8s-Cls 改进的轮胎关键信息区域角度分类算法

为准确分类轮胎胎侧关键信息文本的角度，以便对文本进行方向矫正，同时提高模型的执行效率，降低模型对计算资源的需要，本文选择 YOLOv8s-Cls 分类算法对文本角度进行分类，引入星形运算网络 StarNet。StarNet 的主要模块 Star Blcok 采用星形运算结构，能够在低维输入空间中高效捕捉复杂、高维且非线性的特征。高维特征与低维特征之间的相互映射能够增强特征的判别能力，进而提高模型的分类能力，因此将 StarNet 加入 YOLOv8s-Cls 分类模型中，可以有效提高模型对轮胎胎侧关键信息所在区域图片的角度分类能力，改进后的 YOLOv8s-Cls 的结构图如图 3-12 所示：

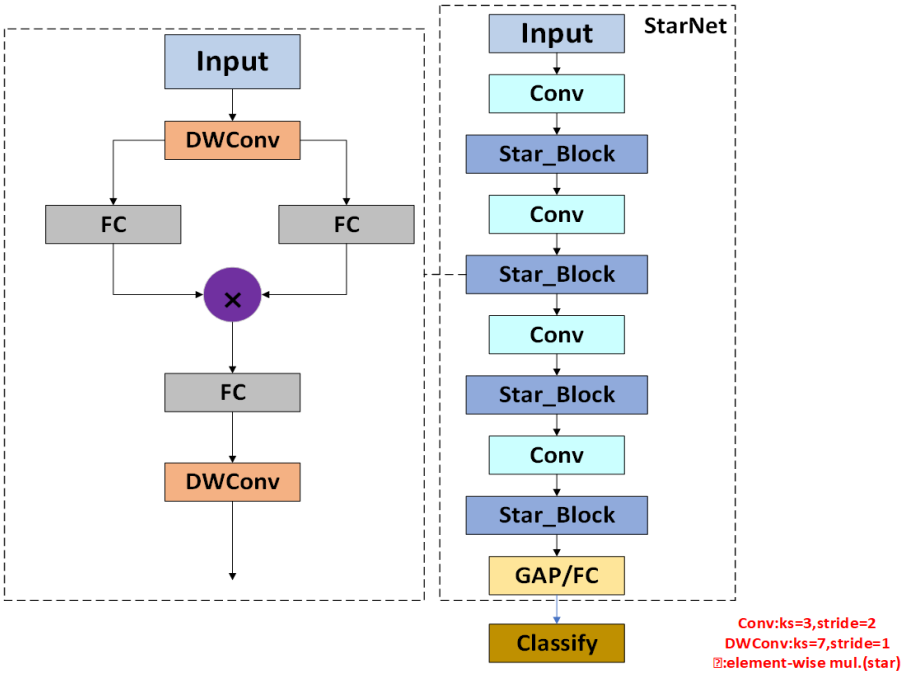

图 3-12 YOLOv8s-Cls 结构图

Figure 3-12 YOLOv8s-Cls Structure

3.3.1.1 StarNet

相较于神经网络中传统的通道数变换操作，元素乘法操作能够在不同通道之间执行成对的特征乘法。这种操作能够在保持特征空间紧凑性的同时，实现维度的近乎无限扩展。Star Block 模块通过利用元素乘法将输入特征映射至一个高维非线性特征空间，从而显著增强网络的表达能力和计算效率。首先星形运算可以表示：

$$\begin{aligned}&(W_1^T X)*(W_2^T X)\\&=(\sum_{i=1}^{d+1} w_1^i x^i)*(\sum_{j=1}^{d+1} w_2^j x^j)\\&=\sum_{i=1}^{d+1}\sum_{j=1}^{d+1} w_1^i w_2^j x^i x^j\\&=a_{(1,1)}x^1x^1+\cdots+a_{(d+1,d+1)}x^{d+1}x^{d+1}\end{aligned} \tag{3.7}$$

$$a_{(i,j)}=\begin{cases} w_1^i w_2^j & i=j \\ w_1^i w_2^j + w_1^j w_2^i & i\neq j \end{cases} \tag{3.8}$$

通过公式可以清晰地看出，星形运算能够实现不同维度特征之间的相互映射，且并未引入复杂的结构与运算过程。Star Block 的处理流程首先通过卷积核大小为 7 的深度卷积进行特征提取。大卷积核在充分捕获特征的同时，深度卷积有效

降低了计算成本。随后，通过两个独立的卷积层模拟全连接层的功能，以实现特征的整合。这两个卷积层的输出特征通过星形运算进行多维映射，从而增强特征的非线性表达能力。映射后的特征进一步通过卷积全连接层进行整合，最终通过一个大核深度卷积层进行深层次的特征提取，以进一步提升特征的表达程度。而将 Star Block 进行堆叠，并在每个 Star Net 前加入一个 3×2 的 Conv 模块进行降采样，通过这样的堆叠构成了星形网络 StarNet，并将其替换 YOLOv8s-Cls 的主干网络。

3.3.1.2 实验数据分析

为了系统评估 StarNet 网络在引入 YOLOv8s-Cls 分类算法后的性能提升效果，本节采用了对比实验。通过对比实验，从而验证改进的有效性。

与上文相同，本节的训练参数仍然采用 YOLOv8s 的初始化训练参数，部分实验参数如表 3-8 所示。

表 3-8 分类算法实验参数

Table 3-8 Experimental Parameters of the Classification Algorithm

参数名称	参数设置
批量大小	16
动量	0.937
权重衰弱	0.0005
学习率	0.01
迭代次数	500
训练图片大小	640

本节采用用于验证分类模型的指标 Acc(Accuracy)来进行评估，Acc 是指模型对图片进行正确预测的结果在全部预测结果中的占比，可以直接反应模型对图片进行准确分类的能力，Acc 的计算公式如下：

$$ACC = \frac{TP + TN}{TP + FP + FN + TN} \tag{3.9}$$

为了展示改进后的分类算法的性能优越性，本节同样采用相同的数据集与训练参数进行不同分类检测模型的对比实验，实验结果如表 3-9 所示：

表 3-9 分类算法对比实验结果

Table 3-9 Comparison of Classification Algorithm Results

名称	Acc	参数量/10^6	计算量(GFLOPs)
YOLOv8s-Cls	86.6%	5.08	12.4
YOLOv8s-Cls-RestNet50	82.1%	26.1	68.8
YOLOv8s-Cls-FasterNet	85.8%	2.63	5.9
Ours	91%	1.09	3.3

经对比实验验证，上文中提到的 ResNet50 主干网络无论是在分类精确度上还是轻量化程度上都不及 YOLOv8s-Cls 的主干网络，而本节提出的基于 StarNet

主干网络改进的 YOLOv8s-Cls 角度分类算法在精确度上较原始 YOLOv8s-Cls 提高了 4.4%，较采用 FasterNet 轻量化改进的 YOLOv8s-Cls 分类模型提高了 5.2%，改进后的 YOLOv8s-Cls 分类模型在精确度上达到了最佳状态。在模型轻量化程度方面，改进后的 YOLOv8s-Cls 分类模型相较于原始 YOLOv8s-Cls 在参数量上降低了 79%，计算量上降低了 73%。与采用 FasterNet 轻量化改进的 YOLOv8s-Cls 分类模型相比，本节提出的基于 StarNet 改进的 YOLOv8s-Cls 分类模型在参数量上降低了 59%，计算量上降低了 44%。总体而言，本节提出的改进 YOLOv8s-Cls 分类模型在轮胎胎侧关键区域图片角度分类任务的性能上达到了最优水平。

3.3.1.3 分类结果示例

图 3-13 分类算法结果示例

Figure 3-13 Example of Classification Algorithm Results

根据分类结果示例图 3-13 可以看出，改进后的 YOLOv8s-Cls 分类模型可以非常优秀的完成对轮胎胎侧关键信息所在区域图片角度的分类任务，从而提高后续阶段检测与识别的准确度。

3.3.2 本节小结

本节首先对图片角度分类与矫正对后续检测与识别的影响进行了深入阐释，随后对多种基于深度学习的图像分类算法及其架构进行了全面综述与深入剖析。在此基础上，本研究选择以 YOLOv8s-Cls 分类算法作为基础模型，并针对该模

型进行了针对性的改进，采用星形运算网络（StarNet）替换了原有的主干网络，从而在提升分类精度的同时降低了模型对计算资源的需求。通过对比实验以及分类结果示例可以清晰地看到，改进后的 YOLOv8s-Cls 分类模型能够高效地完成对轮胎胎侧关键区域图片角度的分类任务，为后续对图片进行非阅读角度的矫正提供了有力支持。

第 4 章 轮胎胎侧关键信息 OCR 识别

在轮胎胎侧关键信息所在区域的检测任务圆满完成之后，便正式进入轮胎胎侧关键信息的光学字符识别（OCR）阶段。在本章中，提出了一种经过优化的基于 YOLOv8s 的文本检测算法。该算法能够在检测文本区域的同时，对不同语义信息的文本内容进行精准的分类处理，为后续的字符识别环节提供高质量的输入数据。紧接着，提出了一种基于 SVTR 改进的文字识别算法，该算法能够以极高的效率和准确度对文本进行识别，从而准确地提取出轮胎胎侧关键信息的字符内容。

4.1 轮胎胎侧关键信息文本检测

在完成对轮胎胎侧关键信息所在区域的检测并获取其位置信息后，下一步是对该区域图像中的文本进行定位检测，以确定文本的确切位置。

EAST、SAST、DB 以及 DB++这些算法在文本检测领域得到了广泛应用，但它们主要针对二分类文本检测任务，只能分类出文本与背景，无法满足本文轮胎胎侧关键信息检测任务中对不同语义文本进行分类的需求。基于此，本节提出采用能够进行分类的目标检测算法来执行轮胎胎侧关键信息的文本检测任务，并进一步提出了一种基于 YOLOv8s 改进的轮胎胎侧关键信息文本检测算法。该改进算法旨在更好地完成文本检测任务，提升检测的准确性和效率。

4.1.1 基于 YOLOv8s 改进的轮胎胎侧关键信息文本检测算法

为了更好的完成轮胎胎侧关键信息文本检测任务，从轮胎胎侧关键信息区域中准确的检测到文本区域并根据不同语义信息进行准确分类，同时为了加快系统的执行速度以及降低系统整体对计算资源的需要，以便于在应用边缘设备的场景中使用，本节设计了一种基于 YOLOv8s 改进的轮胎胎侧关键信息文本检测算法，该算法在确保检测精度的同时，显著降低了计算复杂度，实现了轮胎胎侧文字检测的快速准确检测与分类。为进一步优化模型性能，本文改进了一种轻量化特征提取模块 RepNCSPELAN4_CAA，以替代 YOLOv8 主干网络中的 C2f 模块，从而在减少网络参数量和计算量的同时，增强特征提取能力。此外，针对颈部（Neck）结构的轻量化设计，本文提出了一种动态采样的轻量化大感受野特征融合网络

DyLKVHSPAN，以解决传统轻量化卷积可能导致的特征丢失问题，确保多层级特征的高效融合。最后，在检测头的设计中，基于 YOLOv8s 的检测头架构，本文提出了一种低计算量检测头 DWLHead，通过轻量化模块设计和扩大感受野的策略，在降低计算开销的同时进一步提升检测性能。实验结果表明，所提出的改进方案在计算效率和检测精度之间实现了良好的平衡，为复杂场景下的文本检测任务提供了有效的解决方案。基于 YOLOv8s 改进的轮胎胎侧关键信息文本检测算法的结构图如图 4-1 所示：

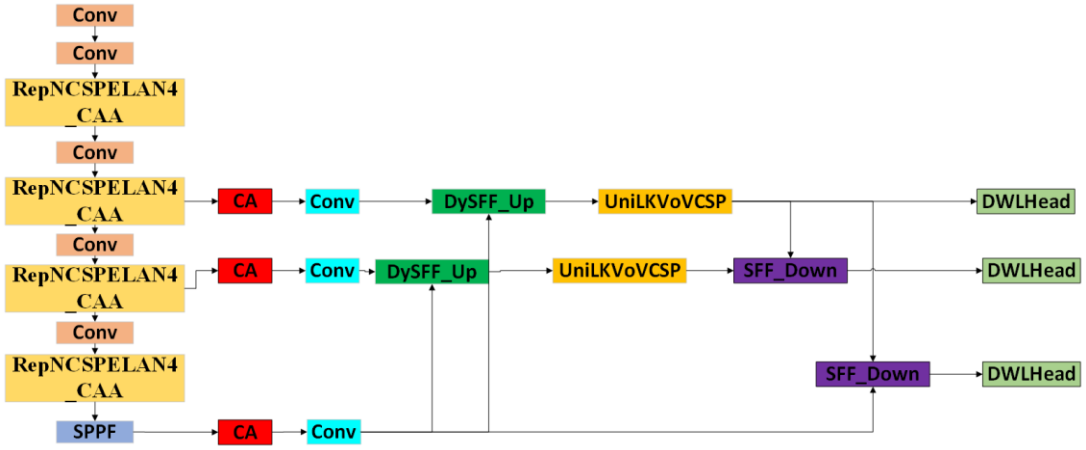

图 4-1 基于 YOLOv8s 改进的轮胎胎侧关键信息文本检测算法网络结构

Figure 4-1 Network Structure of Tire Sidewall Key Information Text Detection Algorithm Based on Improved YOLOv8s

4.1.1.1 轻量化特征提取模块 RepNCSPELAN4_CAA

尽管 C2f 模块在设计上融入了 ELAN 理念，但其仅由常规卷积（Conv）模块构成的结构限制了其在复杂场景下的特征提取能力。为了克服这一局限性，Wang 等[56]在 YOLOv9 中提出了一种创新的重参数化 ELAN 结构模块 RepNCSPELAN4。该模块通过引入重参数化技术，旨在提升模型的特征提取性能，同时优化网络的计算复杂度和参数量，从而实现更高的计算效率和检测精度。

RepNCSPELAN4 模块的核心设计包括四个关键组件：Conv 模块、RepConv 模块[57]、RepNCSP 模块和 RepNBottleneck 模块。其中，RepConv 模块采用多分支结构，在训练过程中通过重参数化技术实现参数的高效复用，从而在减少模型参数量的同时，提取更为丰富的特征信息。RepNCSP 模块和 RepNBottleneck 模块则通过融合多种卷积操作，进一步增强了特征提取的多样性和鲁棒性。

在特征提取过程中，上下文信息对于构建丰富且语义化的特征表示至关重要，能够显著提升模型的泛化能力和推理性能。然而，传统的特征提取方法往往难以有效保留上下文信息，导致特征表示的局限性。为此，该算法引入了上文中提到的上下文锚点注意力机制（Contextual Anchor Attention, CAA）模块，以增强上下文信息的捕获能力，在处理细长目标区域（如桥梁、裂缝以及轮胎胎侧关键信息区域）时表现出更优的特征提取效果。

基于上述分析，本文改进了一种上下文感知特征提取模块 RepNCSPELAN4_CAA，该模块将上下文锚点注意力机制 CAA 与重参数化 ELAN 结构相结合，旨在提取更多特征的同时保留丰富的上下文信息。如图 4-2 所示，RepNCSPELAN4_CAA 模块的特征提取过程首先在 RepNCSPELAN4 模块内完成，随后通过 Concat 操作对提取的特征进行整合拼接。接着，利用 CAA 模块进一步提取上下文信息，其中小核卷积用于捕获局部特征，而并行的深度卷积则用于捕获多尺度上下文信息。最后，通过一个 Conv 模块整合这些特征并调整输出维度，以满足后续处理的需求。这一设计不仅提升了特征提取的多样性，还显著增强了模型在复杂场景下的检测性能。

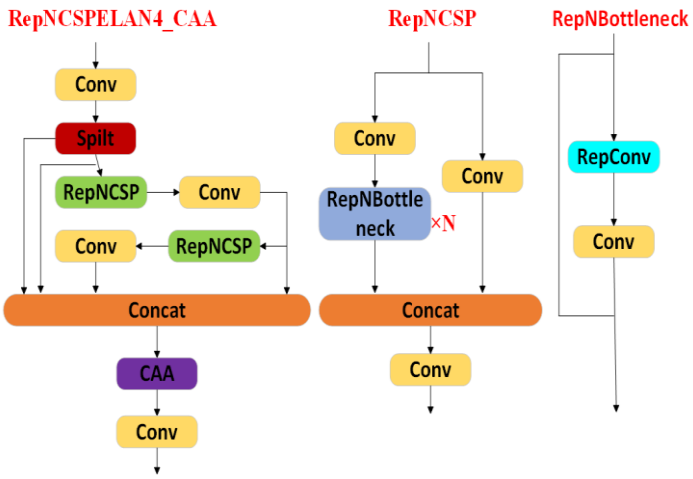

图 4-2 RepNCSPELAN4_CAA 模块

Figure 4-2 RepNCSPELAN4_CAA Module

4.1.1.2 动态采样的轻量化大感受野特征融合网络 DyLKVHSPAN

正如上文中提及的，YOLOv8 的颈部结构存在特征融合不足以及在上采样过程中造成特征丢失的问题，在本节中提出的基于 YOLOv8s 改进的轮胎胎侧关键信息文本检测算法中提出了一种针对性的特征融合网络。新的特征融合网络结构以高效轻量化的特征融合网络 HSPAN 为框架结构，在其中提出了一种大感受野

特征融合模块 UniLKVoVCSP，并在 SFF_Up 模块中结合上文提及的动态上采样器 DySample 模块，共同构建一种新型的动态采样轻量化高效特征融合网络 DyLKVHSPAN，其结构如图 4-3 所示。

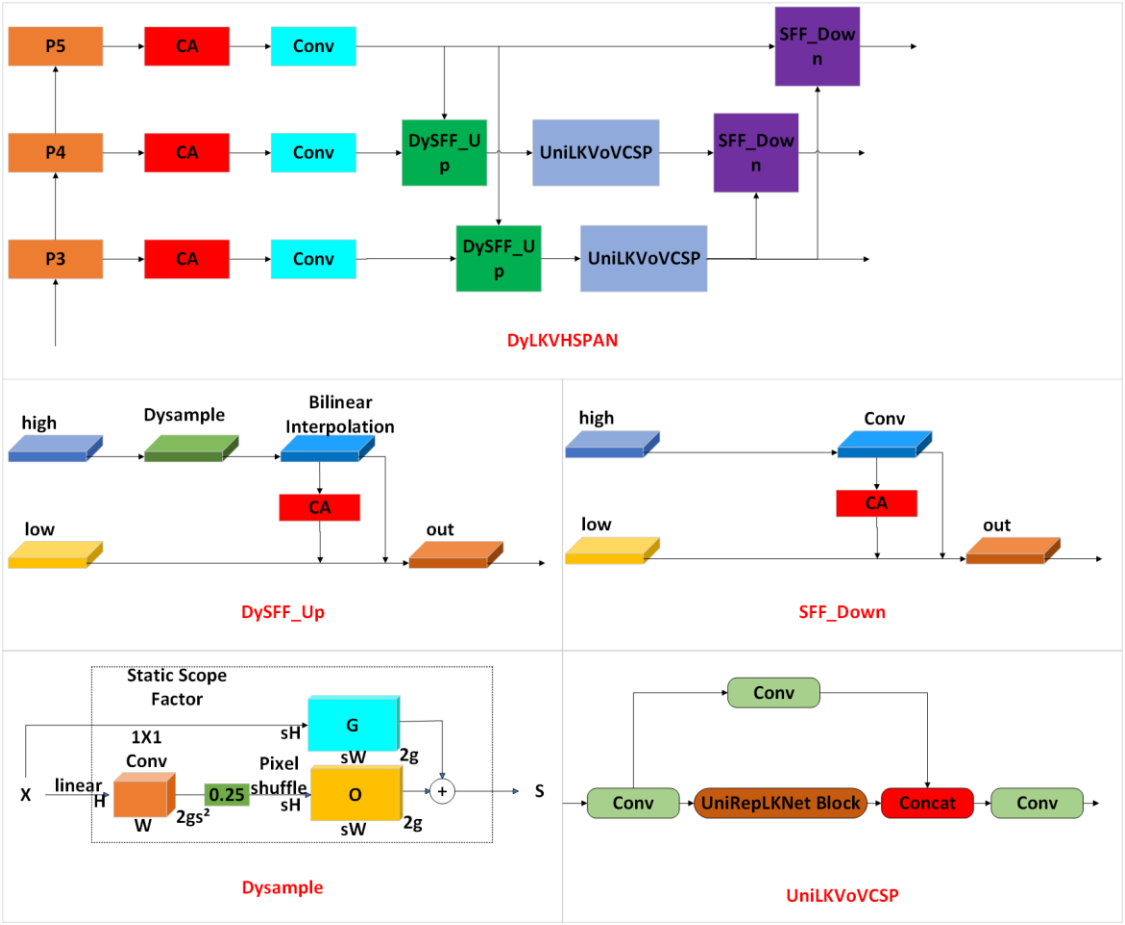

图 4-3 DyLKVHSPAN 结构

Figure 4-3 DyLKVHSPAN Structure

（1）高效轻量化特征融合网络 HSPAN

鉴于轮胎胎侧文字区域特征的稀缺性，实现对提取特征的高效多尺度融合对于提升检测性能至关重要。Chen 等提出了一种基于层次尺度的特征金字塔网络（HSFPN）[58]，旨在实现多尺度特征的深度融合。HSFPN 的基本结构依然是基于特征金字塔网络（FPN）的网络结构，尽管其在特征提取和融合方面表现出了一定的性能，但仍有进一步提升的空间。HSPAN 的结构如图 4-4 所示

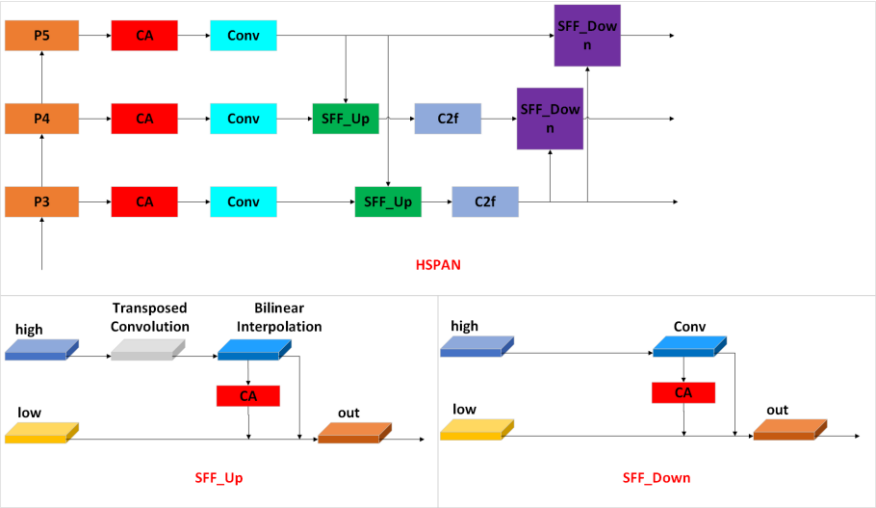

图 4-4 HSPAN 结构

Figure 4-4 HSPAN Structure

（2）大感受野的特征融合模块 UniLKVoVCSP

在构建特征融合网络的特征融合模块时，面临着保持模型轻量化与整合多尺度特征信息的双重挑战。轮胎胎侧背景颜色较深，区域特征稀缺且难以提取，这进一步增加了特征融合的难度。因此，本文旨在通过特征融合模块整合更多多尺度特征信息，以提高检测性能。为此，本文在轻量化特征融合模块 VoVGSCSP 的基础上，引入大核重参数化模块 UniRepLKNet Block[59]，以替换原有的 GS Bottleneck。这一改进旨在扩大模块的感受野，从而提出一种新型的大感受野特征融合模块——UniLKVoVCSP。

UniRepLKNet Block 由结构重参数化的 Dilated Reparam Block、SE 注意力机制、前馈神经网络（FFN）以及一个批归一化（BN）层构成。Dilated Reparam Block 结合了大核非膨胀卷积与多个小核膨胀卷积，并采用共享参数的重参数化设计，从而实现等效于一个超大核卷积的效果，显著扩大了感受野。通过这种设计，UniLKVoVCSP 能够在保持模型轻量化的同时，有效整合多尺度特征信息，从而提高特征融合模块的性能。UniLKVoVCSP 的结构如图 4-5 所示。

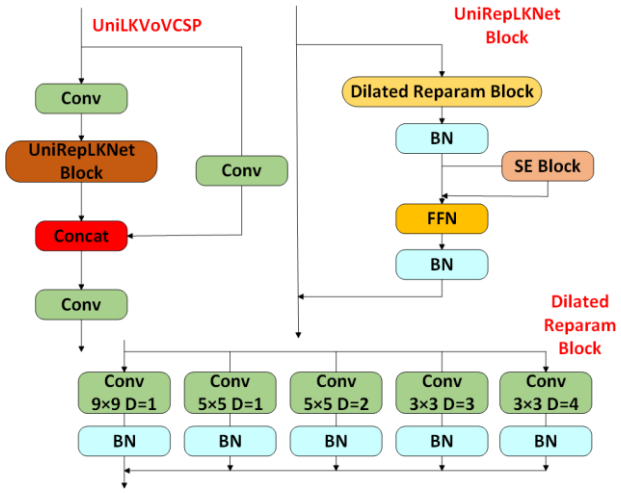

图 4-5 UniLKVoVCSP 模块

Figure 4-5 UniLKVoVCSP Module

4.1.1.3 低计算量解耦检测头 DWLHead

YOLOv8 检测头采用解耦设计，通过两个独立的分支分别处理回归和分类任务。每个分支由两个卷积核尺寸为 3 的一维卷积模块和一个卷积核尺寸为 1 的二维卷积组成。相较于耦合设计，这种结构显著提高了检测精度，但不可避免地增加了计算负担。为了降低计算量并减少对计算资源的依赖，本文提出了一种低计算量的解耦检测头——DWLHead。

研究发现，原始检测头中两个卷积核尺寸为 3 的一维卷积操作的计算成本较高。因此，本文对检测头进行了优化。具体而言，保留了一个步长为 2 的卷积核尺寸为 3 的一维卷积操作，以减少计算量并实现下采样，同时保留关键特征信息。第二个卷积操作则采用卷积核尺寸为 7 的深度卷积，以扩大感受野并提取更多特征信息，同时有效控制计算成本。这种新型检测头被命名为 DWLHead，其结构如图 4-6 所示。

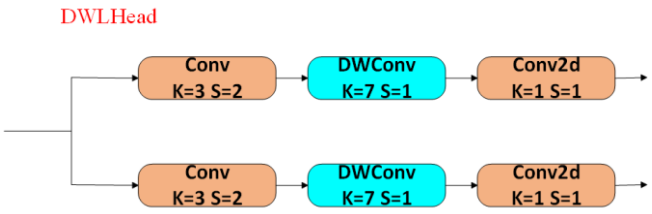

图 4-6 DWLHead 模块结构

Figure 4-6 DWLHead Module Structure

4.1.1.4 实验数据分析

为了验证 YOLOv8s 算法在加入轻量化特征提取模块 RepNCSPELAN4_CAA、动态采样的轻量化大感受野特征融合网络 DyLKVHSPAN 以及低计算量解耦检

测头 DWLHead 后，在轮胎胎侧关键信息文本检测任务中的性能提升效果，本节采用了对比实验与消融实验。通过对比实验，旨在展示改进后的算法在检测精度、效率以及资源消耗等方面的综合优势；而消融实验则用于深入分析各个改进模块对整体性能的改进。

本节实验仍然采用 YOLOv8s 的初始化训练参数。通过训练对模型参数进行微调，以达到最佳的训练效果。部分实验参数如表 4-1 所示。

表 4-1 文本检测实验参数

Table 4-1 Experimental Parameters of Text Detection

参数名称	参数设置
批量大小	16
动量	0.937
权重衰弱	0.0005
学习率	0.01
迭代次数	500
训练图片大小	640

本节实验仍然采用平均精度 Precision (P)、Recall (R)和 Mean Average Precision(mAP)来评估模型的精度，通过模型的参数量 Params 和模型的计算量 GFLOPs 来衡量模型的大小。

在采用同样数据集与训练策略的情况下进行了对比实验，将已有算法与本节提出的基于 YOLOv8s 的轮胎胎侧关键信息文本检测算法进行对比，实验结果如表 4-2 与表 4-3 所示：

表 4-2 与 YOLO 系列算法对比实验结果

Table 4-2 Comparison of Experimental Results with YOLO Series Algorithms

名称	P	R	mAP50：95	参数量 /10^6	计算量 (GFLOPs)
YOLOv3-Tiny	93.8%	93.5%	66.0%	8.68	12.9
YOLOv5s	96.0%	94.7%	71.0%	7.02	15.8
YOLOv7-Tiny	94.5%	97.9%	70.2%	6.02	13.1
YOLOv8s	97.2%	94.8%	73.0%	11.13	28.6
YOLOv10s	94.9%	90.2%	71.2%	8.04	24.4
YOLO11s	97.5%	94.5%	75.6%	9.41	21.3
YOLOv12s	96.0%	95.3%	75.2%	9.07	19.3
Ours	97.4%	98.2%	75.5%	5.74	14.3

表 4-3 与其他优秀算法对比实验结果

Table 4-3 Comparison of Experimental Results with Other State-of-the-Art Algorithms

名称	P	R	mAP50：95	参数量 /10^6	计算量 (GFLOPs)
文献[60]	96.2%	94.0%	72.8%	7.30	22.2
文献[61]	93.6%	94.0%	73.4%	5.92	20.8
文献[62]	97.4%	96.6%	73.5%	9.82	25.5
文献[63]	94.9%	96.6%	73.5%	7.99	29.4
Ours	97.4%	98.2%	75.5%	5.74	14.3

本节提出的轮胎胎侧文本检测算法采用的新型特征提取模块能够捕获更全面的特征信息，其颈部结构在特征融合效率上超越了 YOLO 系列中的传统颈部网络，而 DWLHead 检测头在减少计算量的同时，有效增强了检测性能。由表可知，与 YOLOv8s 及基于 YOLOv8s 的 YOLOv10s 相比，本节提出的轮胎胎侧文本检测算法在检测精度和计算成本上均表现更优，mAP@50：95 分别提升 2.5 和 4.3 个百分点，参数量分别减少 48.4%和 28.6%，计算量分别下降 49.6%和 41.6%。与 YOLO 的最新模型 YOLO11s 相比，本文改进的模型和 YOLO11s 与 YOLOv12s 的检测效果相当，但是参数量与计算量相较于 YOLO11s 分别下降了 39%与 28%和 32.9%与 26%，证明了本节提出的基于 YOLOv8s 改进的算法需要更小的计算资源，可以更好的在需要边缘计算的场景中部署。

由表 4-3 可知：与文献[60]相比，本节提出的降低检测头计算量的方法更为精细，避免了单纯使用分组卷积可能导致的检测精度下降；相较于文献[61]中采用的 BiFPN，本节提出的 DyLKVHSPAN 在结构设计上更为高效，得益于其大感受野模块和先进的上采样技术，使得特征融合效果更为出色；与文献[62]中采用的 CPCA 注意力机制相比，本节选择了更适合检测轮胎胎侧细长文本区域的上下文锚点注意力机制 CAA，从而在轮胎胎侧关键信息区域的检测任务中展现出更优越的性能；文献[63]采用 FasterBlcok 改进 C2f 模块，虽然实现了模型的轻量化，但是这种轻量化方法容易造成特征提取不足，因为在处理本节任务时精度并不优秀。本节提出的 RepNCSPELAN4_CAA 在轻量化设计的同时，也确保了检测精度的提升。

为了验证改进算法的优越性，采用消融实验对 YOLOv8s 模型以及各个改进方案进行消融实验对比，具体实验结果如表 4-4 所示。

表 4-4 文本检测算法消融实验结果

Table 4-4 Ablation Study Results of Text Detection Algorithm

RepNCSPELAN4_CAA	DYLKVHSPAN	DWLHead	P	R	mAP50:95	参数量 /10^6	计算量 (GFLOPs)
			97.2%	94.8%	73.0%	11.13	28.4
√			95.3%	97.0%	73.8%	10.19	23.0
	√		97.0%	96.4%	74.0%	7.41	21.8
		√	97.5%	95.1%	74.0%	12.21	21.8
√	√		97.7%	96.0%	74.6%	6.49	18.5
√		√	95.6%	96.1%	74.6%	11.29	23.0
	√	√	96.5%	96.4%	74.6%	6.67	15.5
√	√	√	97.4%	98.2%	75.5%	5.74	14.3

消融实验结果表明，每个改进点都有助于提高综合性能。在 YOLOv8s 中加入 RepNCSPELAN4_CAA 用于主干网络时，RepNCSPELAN4_CAA 增强了模型对输入特征提取的能力，mAP@50:95 增加 0.8 个百分点，参数量降低 8.4%，计算量降低 19%；在 YOLOv8s 加入 DYLKVHSPAN 特征融合网络时，相较于原有的特征融合网络，结构更简单高效，不仅提升了精度，也降低了参数量与计算量；相较于原有的 YOLOv8s 检测头，DWLHead 在结构的设计上有着更好的性能，mAP@50:95 提升 1 个百分点，计算量降低 23.2%，但是由于采用大核深度卷积的原因，参数量提升 9%。相较于单独加入改进点，将改进点组合同时加入使用时效果更加优秀，加入 RepNCSPELAN4_CAA 与 DYLKVHSPAN 时 mAP@50:95 提升 1.6 个百分点，参数量降低 41.7%，计算量降低 34.9%；在 RepNCSPELAN4_CAA 的基础上加入 DWLHead 时，mAP@50:95 提高 1.6 个百分点，参数量几乎持平，计算量降低 19%；在 DYLKVHSPAN 和 DWLHead 结合时，mAP@50:95 提升 1.6 个百分点，参数量降低 40%，计算量降低 45%；在加入所有改进点共同构成本节提出的改进模型时，无论是在精度还是在轻量化程度上都实现了很好的效果，在 mAP@50:95 上提高 2.5 个百分点，在轻量化的效果上更为出色，参数量降低 48.4%，计算量降低 49.6%。

4.1.1.5 检测效果对比

为了更直观地展示本节提出的轮胎胎侧文本检测算法比 YOLOv8s 目标检测算法带来的提高，从数据集中随机选择若干图片进行推理，推理结果如图所示。可以看出，由于轮胎的胎侧背景较深，原始的 YOLOv8s 目标检测算法容易出现漏检问题，而本节提出的轮胎胎侧文本检测算法在检测效果上可以看出比起原始的 YOLOv8s 目标检测算法在漏检问题上表现更好。文本检测算法对比结果如图 4-7 所示。

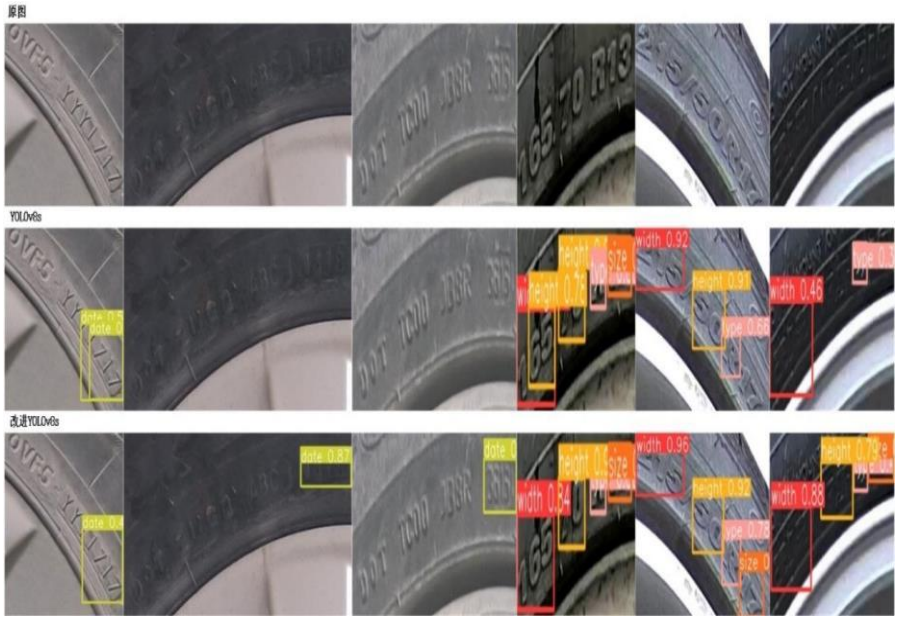

图 4-7 文本检测算法效果对比

Figure 4-7 Comparison of Text Detection Algorithm Effects

4.1.2 本节小结

本节首先综述了当前主流的基于深度学习的二分类文本检测算法，并深入分析了其检测原理及其在轮胎胎侧文本检测任务中的适用性。研究发现，尽管这些算法在通用文本检测任务中表现优异，但在轮胎胎侧文本检测场景中存在一定的局限性，无法根据不同的语义信息对不同语义的文本进行分类。基于此，本文决定采用具有分类功能的 YOLOv8s 目标检测算法作为基础框架，以更好地适应轮胎胎侧文本检测任务的特点。

本文提出的改进算法在 YOLOv8s 的基础上，引入了轻量化特征提取模块 RepNCSPELAN4_CAA、动态采样的轻量化大感受野特征融合网络 DyLKVHSPAN 以及低计算量解耦检测头 DWLHead。实验结果表明，相较于原始 YOLOv8s 算法及其他主流目标检测算法，本文提出的改进算法在检测精度方面显著提升，同时有效减少了漏检和误检问题。此外，该算法在计算效率方面表现出色，能够在较低的计算代价下实现高效的文本检测，从而为轮胎胎侧关键信息检测任务提供了一种更为优越的解决方案。

4.2 轮胎胎侧关键信息文字识别

在轮胎胎侧关键信息检测与识别系统中，文字识别作为最终环节，肩负着对检测并矫正后的文本区域进行字符识别的关键任务。

4.2.1 基于 SVTR 改进的轮胎胎侧关键信息文字识别算法

在综合考虑 CRNN 与 SVTR 两种算法的性能和特点后，本文选择 SVTR 文字识别算法作为轮胎胎侧关键信息文字识别的核心算法。SVTR 不仅在识别精度上表现卓越，而且在处理速度上也具有显著优势，能够保证较高的 FPS（Frames Per Second），从而满足实际应用场景中对实时性的要求。并且本节通过改进 SVTR 算法，加入了空间变换网络 STN（Spatial Transformer Networks）[64]，增强网络对几何变换的适应能力，并且提出了新的损失函数 FocalCTC Loss，进一步提高识别精确度，这些改进旨在进一步提升轮胎胎侧关键信息检测与识别系统的整体性能，为轮胎关键信息的准确提取提供有力支持。改进后的 SVTR 网络模型如图 4-8 所示：

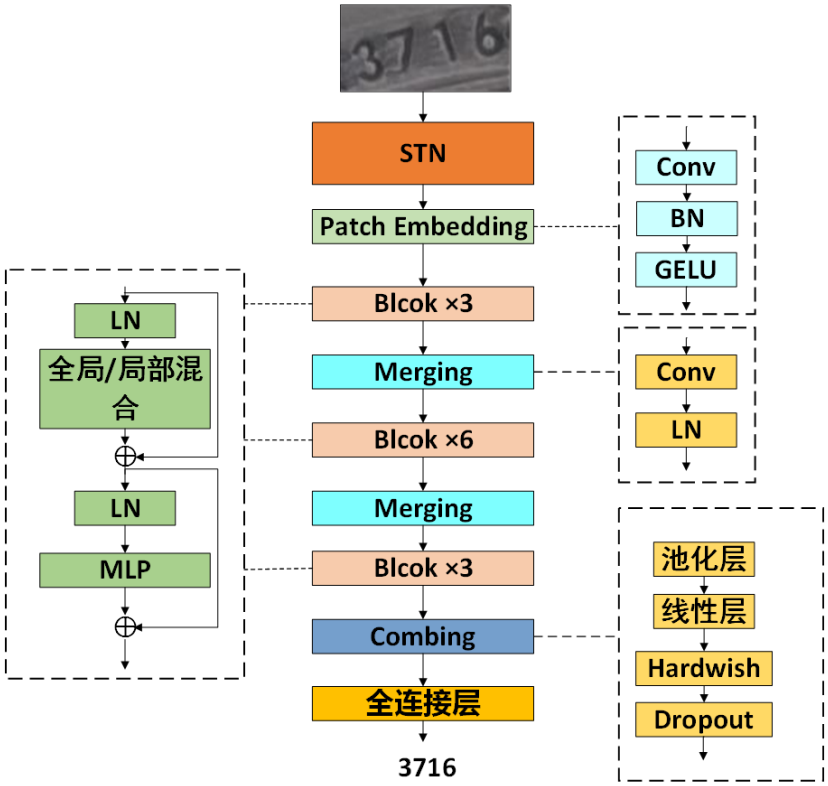

图 4-8 改进 SVTR 文字识别算法结构

Figure 4-8 Improved SVTR Text Recognition Algorithm Structure

4.2.1.1 STN

是一种可嵌入传统卷积神经网络（Convolutional Neural Network, CNN）的可学习网络模块。该模块能够显式地对输入图像实施空间变换，显著提升网络对输入图像几何变形的适应能力，进而赋予模型空间不变性。即使在字符识别任务中，待识别字符发生一些的几何变换，模型依然能够输出准确的识别结果。

空间变换网络（Spatial Transformer Networks, STN）模块由三个关键组成部分构成，分别是定位网络（Localization Network）、网格生成器（Grid Generator）与采样器（Sampler）。定位网络作为 STN 的核心组件，其主要任务是学习输入图像的空间变换参数。它接收输入图像，并输出空间变换所需的参数 θ，这些参数定义了一个变换矩阵，用于调整图像的空间位置。网格生成器的作用是接收定位网络输出的变换参数 θ，并生成一个对应于输出图像的坐标网格。该网格定义了从输入特征图中采样的位置。采样器的功能相对简洁，即利用生成的采样网格从输入特征图中进行采样，从而生成变换后的特征图。

STN 网络模块的结构图如图 4-9 所示：

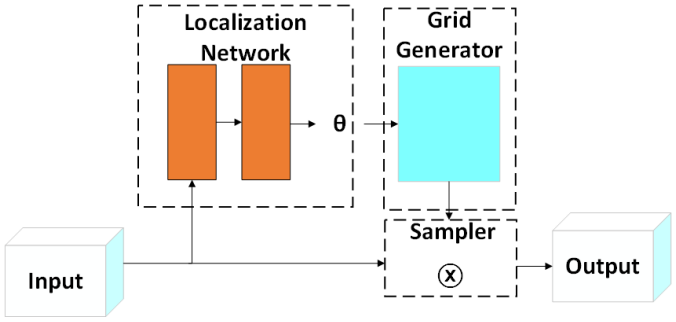

图 4-9 STN 网络模块图

Figure 4-9 STN Network Module

4.2.1.2 Focal CTC Loss

在目标检测领域，Focal Loss[65]被提出以应对正负样本比例严重失衡的问题。传统的交叉熵损失函数在处理此类不平衡数据时，往往因多数类别的主导作用而使模型对少数类别的分类性能显著下降。在序列数据处理任务中，CTC Loss 被广泛应用于诸如文字识别等场景，然而，它同样面临着样本不平衡以及对困难样本挖掘不足的挑战。因此，将 Focal Loss 与 CTC Loss 相结合，能够有效解决文字识别过程中样本不均衡以及对难例挖掘不足的问题，从而提升模型在复杂场景下的鲁棒性和准确性。

Focal Loss 通过引入调制因子$(1-p_t)^\gamma$和平衡因子 α，降低易分类样本的权重，增大难分类样本的权重，Focal Loss 的计算公式如下式 4-1：

$$FocalLoss(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t) \tag{4.1}$$

其中，p_t 是模型对样本的预测概率，α_t 用于调节正负样本权重，γ 是调节难易样本权重的参数，当 p_t 趋向于 1，即说明该样本是易区分样本，此时调制因子

$(1-p_t)^{\gamma}$ 是趋向于 0，说明对损失的贡献较小，即减低了易区分样本的损失比例，反之则加大难区分样本的损失比例。

在序列数据处理任务中，CTC Loss 的主要目标是解决模型输出序列与目标序列长度不一致的问题。传统的损失函数（如交叉熵损失）要求输入和输出的长度严格一致，这在处理变长序列时会带来显著的困难。为了解决这一问题，CTC Loss 引入了“空白”字符（blank），允许模型输出比目标序列更长的序列。通过特定的路径映射和概率计算，CTC Loss 能够有效地将模型输出序列映射到目标序列，并据此确定最终的损失值。这种方法不仅提高了模型对变长序列的适应性，还增强了模型在序列生成任务中的灵活性和鲁棒性。

而在 CTC Loss 中同样存在对困难样本挖掘不足的问题，因此本节将 Focal Loss 与 CTC Loss 相结合，提出了 Focal CTC Loss，提高 CTC Loss 对困难样本的损失比例，Focal CTC Loss 的计算公式如下式 4-2：

$$FocalCTCLoss = \alpha * (1 - p(r \mid x))^{\gamma} * CTCLoss \tag{4.2}$$

4.2.1.3 实验数据分析

为更全面地呈现引入空间变换网络 STN 与 Focal CTC Loss 后的改进 SVTR 文字识别算法的性能提升，本节采用对比实验的方法，展示其在文字识别任务中的效果。

本节实验采用 PaddleOCR 的模型架构与初始化的训练参数，通过调整训练参数以达到最佳的训练效果，部分实验参数如表 4-5 所示：

表 4-5 识别算法实验参数

Table 4-5 Experimental Parameters of Recognition Algorithm

参数名称	参数设置
训练批量大小	128
验证批量大小	1
学习率	0.0005
迭代次数	500
训练图片大小	3,32,100

本节与分类模型相同，采用准确度（Accuracy,Acc）来评估文字识别模型的精度，并且本节额外采用帧率（FPS）来评估模型的推理速度，帧率（FPS）的计算公式如下式 4-3 所示，其中 *Total Frames* 代表总帧数，*Total Time* 代表总时间。

$$FPS = \frac{Total\ Frames}{Total\ Time} \tag{4.3}$$

为了更直观地呈现不同文字识别算法的性能差异，本节通过对比实验对各算法的精确度与帧率（FPS）进行了量化评估。实验结果如下：

表 4-6 文字识别算法对比实验结果

Table 4-6 Comparison of Text Recognition Algorithm Results

模型名称	Acc	FPS
CRNN	0.77	121
SVTR	0.89	122
SVTR+STN+Focal CTC Loss	0.91	117

由对比实验的结果可以看出，改进后的 SVTR 文字识别算法对轮胎胎侧关键信息字符的综合性能最佳，在精度上远高于 CRNN，而 FPS 也达到了 100 以上。这是由于 SVTR 摒弃了传统两阶段（特征提取 + 序列建模）的范式，直接通过视觉 Transformer 实现端到端的特征建模。其核心模块采用分层式混合块（Mixing Blocks），通过并行化的全局注意力（Global Mixing）和局部卷积（Local Mixing）操作，能够同时捕获轮胎表面文字的全局语义特征（如字符间排列规律）和局部细节特征（如磨损字符的边缘纹理），并且 SVTR 通过分阶段的高度合并（Merging）和宽度池化（Combing）机制，动态调整特征图分辨率，并且在 SVTR 网络之前加入了 STN 变化网络模块，有效矫正了畸变的图片，并且提出了 Focal CTC Loss，有效平衡了难例在损失函数中的占比，综上所述，改进后的 SVTR 算法通过视觉 Transformer 的创新架构、变化网络模块以及专注难例的 Focal CTC Loss，在轮胎胎侧文字特有的低对比度、曲面畸变、字符磨损等复杂场景下，展现出更优的准确性与实时性。

4.2.2 本节小结

本节提出了一种改进的 SVTR（Scene Vision Text Recognition）文字识别算法。该算法在 SVTR 原有架构的基础上，引入了空间变换网络（STN）模块以及 Focal CTC Loss，从而显著提升了对轮胎胎侧关键信息的文字识别效果。STN 模块通过显式地对输入图像进行空间变换，增强了模型对图像几何变形的适应性；而 Focal CTC Loss 则有效解决了序列数据处理中样本不平衡以及对难例挖掘不足的问题。实验结果表明，改进后的 SVTR 算法在识别准确率和帧率（FPS）这两个关键指标上均展现出显著优势，综合性能优于其他对比算法。因此，本研究最终选定改进后的 SVTR 算法作为本文文字识别任务的核心模型。

第 5 章 轮胎胎侧关键信息检测与识别系统原型设计

本节致力于将前文所提及的各类先进算法进行有机整合，构建了一个基于 Python 语言的综合性检测系统。该系统旨在实现端到端的轮胎胎侧关键信息检测与识别任务，以满足日益增长的自动化检测需求。通过将多种算法的优势相互融合，该系统能够高效地处理轮胎胎侧图像数据，精准地识别出关键信息，从而为后续的质量评估和决策提供可靠依据。为了进一步提升系统的应用性与交互性，本研究采用了 PyQt5 框架来构建前端用户界面（UI）。PyQt5 以其强大的功能和灵活的定制能力，为用户提供了直观、便捷的操作体验，使得该检测系统不仅在技术层面具备高效性，在实际应用中也能够更好地适应不同用户的需求，极大地增强了系统的实用性和可扩展性。

5.1 系统需求分析

轮胎胎侧关键信息检测与识别系统的需求分析需围绕功能性与交互性需求展开。具备针对不同类型的轮胎的关键信息检测识别能力，并且可以适应不同的检测场景。核心功能包括基于深度学习的轮胎胎侧关键信息（规格、DOT 信息中的生产日期）自动检测与识别，并通过日志功能记录在后台文件中；用户交互层面需设计可视化界面，提供实时检测反馈、历史数据查询等功能，确保操作便捷性。

5.2 系统总体设计

本系统的总体设计遵循端到端的设计理念，旨在为用户提供高效且便捷的使用体验。用户仅需选择开始检测，随后系统后端将自动启动推理引擎，对输入图片进行深度分析和处理，并快速得出准确的识别结果。这一过程无需用户进行额外的干预或复杂的操作，识别结果将直接输出至系统前端界面，用户可在第一时间直观地查看最终结果。这种无缝衔接的流程设计不仅确保了系统检测的整体性和连贯性，还显著提升了用户的使用体验，使用户能够以极低的学习成本和操作负担完成复杂的检测任务。系统的工作流程如图所示，清晰地展示了从用户输入

到结果输出的完整流程，进一步凸显了本系统在操作便捷性和自动化程度方面的优势，检测系统的总体设计如图 5-1 所示：

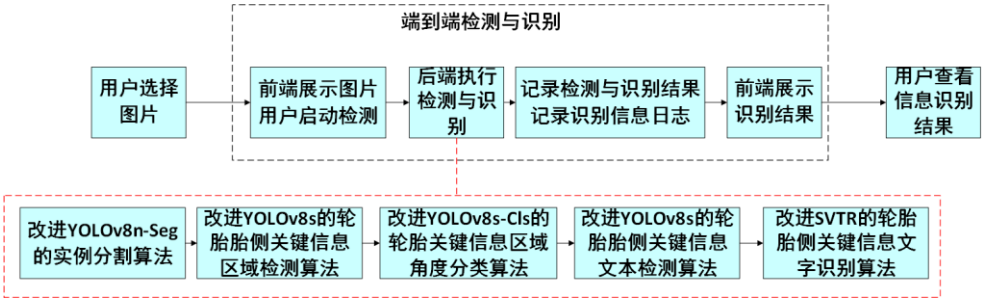

图 5-1 系统总体设计

Figure 5-1 Overall System Design

5.3 系统功能设计

本系统的功能设计分为系统实时性设计、后端功能设计与前端功能设计。

5.3.1 实时性设计

在轮胎胎侧关键信息检测与识别系统的实际应用中，实时检测功能发挥着至关重要的作用。为了满足系统的实时性要求，本研究在系统设计中引入了实时流协议（Real Time Streaming Protocol, RTSP）视频流抽帧处理机制。通过从 RTSP 视频流中合理抽取关键帧进行处理，该机制在保证检测与识别精度的同时，显著提升了系统的处理速度，从而有效满足了实时性需求。

RTSP 是一种网络控制协议，用于在客户端和服务器之间建立实时流媒体会话。本系统通过 RTSP 协议从摄像头或其他视频源获取实时视频流。视频流通常以较高的帧率（如 30 帧/秒）连续获取图像数据。直接对每一帧进行处理会导致较大的计算负担，难以满足实时性要求。

因此，本系统采用动态抽帧策略，根据实际需求和场景动态调整抽帧间隔。例如，在检测场景中，若轮胎的运动速度相对较慢，可以适当增大抽帧间隔；而在运动速度较快时，减小抽帧间隔以确保检测的完整性。通过这种方式，系统能够在保证关键信息不丢失的前提下，显著减少需要处理的图像数量，从而提高处理效率。

由于抽帧减少了处理的图像数量，整个检测与识别过程能够在更短的时间内完成，从而实现对视频流的近实时处理。这一机制不仅优化了系统的实时性能，还确保了检测与识别的准确性，使其适用于各种动态场景。

5.3.2 系统后端功能设计

系统接收上传的关键帧图片后，首先通过改进 YOLOv8s-Seg 的实例分割算法模型定位轮胎主体，利用图像分割技术提取轮胎 ROI 区域；随后采用专门训

练的基于改进 YOLOv8s 的轮胎胎侧关键信息区域检测算法，精准定位胎侧信息区（如 DOT 信息编码、规格参数）；并且在完成关键信息区域检测后，针对可能倾斜的文本区域，使用基于 YOLOv8s-Cls 改进的轮胎胎侧文本角度分类算法判断检测到的轮胎胎侧关键区域图片的角度（0°/90°/180°/270°），并根据检测到的角度将图像矫正至 0°；矫正后的图片输入基于 YOLOv8s 改进的轮胎胎侧关键信息文本检测算法，采用多种改进方法，精确的定位文本区域，最后将文本行区域送入基于 Transformer 架构的 SVTR 文字识别模型完成字符 OCR 识别，最终系统将 OCR 的检测结果，写入检测日志系统记录，实现端到端的文字检测与识别机制。整个处理链采用微服务架构，通过队列实现顺序处理，并运用模型量化技术优化 GPU 资源利用率。后端功能的设计示例如下图 5-2：

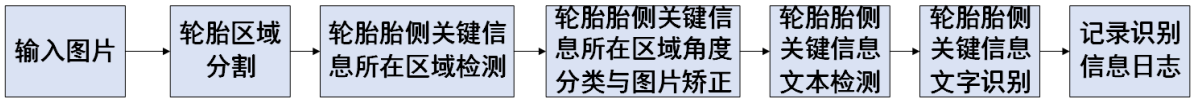

图 5-2 后端设计示例

Figure 5-2 Example of Backend Design

5.3.3 系统前端功能设计

系统的前端设计基于 Pyqt5 与 Python 实现，前端功能共分为四个部分，分别是图片显示区域、识别结果区域、选择待检测图片按钮以及关键信息检测按钮，前端设计如图 5-3 所示。

图 5-3 前端设计示例

Figure 5-3 Example of Frontend Design

在用户点击选择输入视频流按钮后，系统将触发连接 RTSP 视频流，开始抽取关键帧。当抽取到关键帧后，将关键帧图片展示在图片显示区域，供用户确认其选择是否准确无误。在用户确认图片无误并点击关键信息检测按钮后，系统后端将被触发并启动，随即开始执行检测与识别任务。检测完成后，系统将读取识别日志，并将检测结果呈现于识别结果区域，以便用户记录和查阅。

5.4 系统测试

通过利用测试集里未参与模型训练的原始数据对系统进行测试，测试结果如图 5-4 所示：

图 5-4 系统测试示例

Figure 5-4 Example of System Testing

5.5 本节小结

本节提出了一种基于 RTSP、Python 与 PyQt5 的轮胎胎侧关键信息检测与识别系统。该系统通过 RTSP 视频流抽取关键帧，实现实时检测。并且通过 Python 构建后端逻辑，利用其强大的计算能力和丰富的库资源，实现对轮胎胎侧图像的高效处理。同时，采用 PyQt5 构建前端架构，充分发挥其在图形用户界面（GUI）设计方面的优势，为用户提供直观、便捷的操作体验。系统的工作流程遵循端到端的设计理念，用户通过前端界面触发事件，后端随即启动检测与识别逻辑，依次调用多个不同的检测与识别模型。这一过程不仅实现了轮胎胎侧关键信息的精准检测与识别，还确保了系统的高效性和稳定性。最终，检测结果将实时返回至前端，并在结果区域清晰展示，便于客户记录和查阅。

第 6 章 总结与展望

6.1 研究工作总结

本研究聚焦于轮胎胎侧关键信息检测与识别这一重要领域，深入分析了字符识别技术的发展现状以及轮胎胎侧字符识别所面临的挑战。随着汽车产业的智能化发展，对轮胎关键信息的精准检测与识别愈发重要。然而，现有的研究多集中于轮胎的 DOT 信息识别，对于其他关键信息如尺寸信息的检测与识别尚存在不足，且缺乏直接、快速且易于使用的检测系统。针对这些问题，本研究提出了一种基于机器视觉的轮胎关键信息实时检测与识别系统，集成多阶段的检测方法，实现对轮胎胎侧的 DOT 信息以及轮胎尺寸信息的全面检测与识别。具体研究工作及成果如下：

（1）系统性地对现有理论与技术进行了全面的综述与深入分析，涵盖了图像分割、目标检测、类别分类以及文字识别等多个关键领域。通过梳理各领域的发展脉络与技术进展，揭示了当前技术的优势与局限，为后续研究提供了坚实的理论基础与技术参考。

（2）提出了一种多阶段的基于深度学习的轮胎胎侧关键信息检测与识别算法，包括改进 YOLOv8n-Seg 的实例分割算法用于精准定位轮胎区域，改进 YOLOv8s 的目标检测算法用于检测轮胎胎侧的关键信息区域，基于 YOLOv8s-Cls 改进的图像分类算法用于对关键区域图片进行角度分类，以及改进 YOLOv8s 的文本检测算法和基于 SVTR 改进的文字识别算法，分别用于轮胎胎侧关键信息的文本检测与字符识别。

（3）在各个检测与识别阶段，对模型进行了针对性的改进与优化，如引入轻量化模块、注意力机制等，以提升模型的检测精度、降低计算复杂度，实现快速准确的检测与识别。

（4）完成了对不同检测阶段的系统设计以及模型部署，实现了基于机器视觉的轮胎胎侧关键信息实时检测与识别系统的原型设计，该系统通过 Python 构建后端逻辑，利用 PyQt 进行界面设计，实现了从图片输入到结果输出的端到端检测流程。

本研究的成果为轮胎关键信息的自动化检测提供了有效的解决方案，具有较高的应用价值和推广前景。

6.2 研究成果意义

本研究的成果在理论与实际应用层面均具有重要意义。

在理论层面，本研究提出了一系列针对轮胎胎侧关键信息检测与识别的创新性算法与模型改进，这些成果不仅丰富了字符识别领域的技术方法体系，还为复杂场景下低对比度、曲面畸变等困难条件下字符识别的研究提供了新的思路与参考。通过引入轻量化模块、注意力机制等技术手段，本研究在提升模型检测精度的同时，有效降低了计算复杂度，实现了快速准确的检测与识别，为相关领域的研究提供了有益的借鉴。

在实际应用层面，本研究所开发的基于机器视觉的轮胎胎侧关键信息实时检测与识别系统，能够有效满足轮胎生产、销售、维修等环节中对轮胎关键信息自动化检测的需求。该系统通过 Python 构建后端逻辑，并利用 PyQt 进行界面设计，实现了从图片输入到结果输出的端到端检测流程，具有较高的实用性和可操作性。其应用能够显著提高工作效率、减少人工错误，对于保障轮胎产品的质量与使用安全具有积极作用，具有较高的应用价值和推广前景。

6.3 局限性与未来工作

尽管本研究在轮胎胎侧关键信息检测与识别领域取得了一定成果，但仍存在若干局限性。首先，尽管所构建的数据集在一定程度上体现了多样性，但样本数量相对有限，尤其在极端光照条件或特殊场景下的样本覆盖不足，这可能限制了模型在这些特殊情况下的泛化能力。其次，虽然本研究提出了多阶段的检测与识别算法，但在实际应用中，整个系统的处理速度仍有待进一步提升，以更好地满足实时性要求较高的在线检测场景。此外，对于轮胎胎侧字符识别中的一些特殊情况，如字符磨损严重、变形复杂等，模型的识别准确率仍有提升空间。

针对上述局限性，未来的研究工作可从以下几个方面展开：一是进一步扩充和完善数据集，增加更多样化的样本，尤其是针对特殊场景和极端条件下的样本采集，以提升模型的泛化能力和鲁棒性；二是继续优化算法和模型结构，探索更高效的计算策略和轻量化模型设计，提高系统的实时处理性能；三是深入研究字

符识别中的特殊情况，结合更多的先验知识和图像处理技术，进一步提升模型在复杂情况下的识别准确率。

6.4 展望

随着人工智能与机器视觉技术的持续演进，字符识别领域正迎来更为广阔的发展前景，同时也面临着更高的技术要求。展望未来，我们有望见证更加智能化与高效化的轮胎胎侧关键信息检测与识别系统的诞生。这些系统将具备更强的适应性，能够精准且迅速地应对复杂多变的实际应用场景，实现对轮胎关键信息的精准、快速检测与识别。

此外，相关技术的突破不仅将推动轮胎检测领域的进步，还将为其他领域中类似复杂场景下的字符识别问题提供宝贵的借鉴与启示。这将有力地促进字符识别技术在更多领域的广泛应用与深入发展，为相关行业的智能化转型提供坚实的技术支撑。

参考文献

[1] 宋春华,彭泫知.机器视觉研究与发展综述[J].装备制造技术,2019,(06):213-216.

[2] 贾雪峰,冯启章,刘献栋等.爆裂轮胎精确建模及其动态特性仿真方法研究[J].汽车工程,2023,45(05):854-864+872.

[3] 王岩,梁冠群,危银涛.基于支持向量机的智能轮胎路面辨识算法[J].汽车工程,2020,42(12):1671-1678+1717.

[4] 邓海燕.轮胎产品 DOT 认证概述[J].轮胎工业,2002,(07):394-397.

[5] 李明珊,孙文文,齐林,等.全钢载重子午线轮胎 DOT 识别编码编制细则[J].橡胶科技,2019,17(11):635-637.

[6] 卢畅畅,宁少文,唐德昌.光学字符识别技术(OCR)的研究于应用[J].中国战略新兴产业,2018(28):1-3.

[7] Zhou X, Yao C, Wen H, et al. East: an efficient and accurate scene text detector[C]//Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. 2017: 5551-5560.

[8] Wang P, Zhang C, Qi F, et al. A single-shot arbitrarily-shaped text detector based on context attended multi-task learning[C]//Proceedings of the 27th ACM international conference on multimedia. 2019: 1277-1285.

[9] Liao M, Wan Z, Yao C, et al. Real-time scene text detection with differentiable binarization[C]//Proceedings of the AAAI conference on artificial intelligence. 2020, 34(07): 11474-11481.

[10] Liao M, Zou Z, Wan Z, et al. Real-time scene text detection with differentiable binarization and adaptive scale fusion[J]. IEEE transactions on pattern analysis and machine intelligence, 2022, 45(1): 919-931.

[11] Shi B, Bai X, Yao C. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 39(11): 2298-2304.

[12] Du Y, Chen Z, Jia C, et al. Svtr: Scene text recognition with a single visual model[J]. arXiv preprint arXiv:2205.00159, 2022.

[13] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 39(6): 1137-1149.

[14] 李卓璇,周亚同.改进 DBNet 的电商图像文字检测算法研究[J].计算机工程与科学,2023,45(11):2008-2017.

[15] Men J .Multiple Improvement Strategies for High-Density Text Detection Based on DBNet[C]//Department of Mathematics, Purdue University, Department of Computer and Information Sciences, University of Strathclyde.Proceedings of the 2nd International Conference on Computing Innovation and Applied Physics（part4）.Department of Mathematics,New York University;,2023:8.DOI:10.26914/c.cnkihy.2023.109814.

[16] 刘彦希,吴浩,蔡源,等.基于改进 EAST 算法的电气设备铭牌文字检测[J].四川轻化工大学学报(自然科学版),2024,37(03):42-50.

[17] 汪嘉,祝海江,王寅初,等.基于改进 DBNet-RNN 的声级计读数识别方法[J].计量学报,2024,45(08):1191-1199.

[18] 薛璨.基于机器视觉的药品包装质量检测[D]. 徐州：中国矿业大学,2022.

[19] Kazmi W, Nabney I, Vogiatzis G, et al. An efficient industrial system for vehicle tyre (tire) detection and text recognition using deep learning[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 22(2): 1264-1275.

[20] 李嘉豪.基于深度学习的轮胎 DOT 信息识别[D]. 广州：广东工业大学,2022

[21] 张晨梦.基于深度学习的轮胎压印字符识别研究[D].南京：南京邮电大学,2021.

[22] Simonyan K. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[23] 曲星华.汽车轮胎字符识别方法研究及分组系统设计[D].沈阳：沈阳工业大学,2023.

[24] 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.

[25] 苏丽,孙雨鑫,苑守正.基于深度学习的实例分割研究综述[J].智能系统学报,2022,17(01):16-31.

[26] 倪纪鹏,朱立成,董力中,等.基于 SwinS-YOLACT 的番茄采摘机器人实时实例分割算法研究[J].农业机械学报,2024,55(10):18-30.

[27] 刘敬敬,孙凤乾,张皓翔,等.基于 YOLOv8 的轻量化煤屑颗粒群实例分割方法[J/OL]. 煤 炭 学 报 ,1-14[2025-01-13].https://doi.org/10.13225/j.cnki.jccs.2024.0142.

[28] Bolya D, Zhou C, Xiao F, et al. Yolact: Real-time instance segmentation[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 9157-9166.

[29] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.

[30] Cortes C. Support-Vector Networks[J]. Machine Learning, 1995.

[31] Breiman L. Random forests[J]. Machine learning, 2001, 45: 5-32.

[32] Yu J, Jiang Y, Wang Z, et al. Unitbox: An advanced object detection network[C]//Proceedings of the 24th ACM international conference on Multimedia. 2016: 516-520.

[33] Girshick R. Fast r-cnn[J]. arXiv preprint arXiv:1504.08083, 2015.

[34] 沈凌云,郎百和,宋正勋,等.基于 DCS-YOLOv8 模型的红外图像目标检测方法[J].红外技术,2024,46(05):565-575.

[35] Zhang K; Zhang P; Chen J; Long M; Lin W. Dangerous Goods Detection based on improved YOLOv5s X-ray image [J]. Journal of Shaanxi University of Sci ence and Technology, 2023, 41 (06): 176-183+200. doi:10.19481/j.cnki.issn2096 398x.2023.06.006.

[36] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2117-2125.

[37] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8759-8768.

[38] Zheng Z, Wang P, Liu W, et al. Distance-IoU loss: Faster and better learning for bounding box regression[C]//Proceedings of the AAAI conference on artificial intelligence. 2020, 34(07): 12993-13000.

[39] Feng C, Zhong Y, Gao Y, et al. Tood: Task-aligned one-stage object detection[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE Computer Society, 2021: 3490-3499.

[40] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Advances in neural information processing systems, 2012, 25.

[41] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[42] Han K, Wang Y, Tian Q, et al. Ghostnet: More features from cheap operations[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 1580-1589.

[43] Wan D, Lu R, Shen S, et al. Mixed local channel attention for object detection[J]. Engineering Applications of Artificial Intelligence, 2023, 123: 106442.

[44] Zhang X, Liu C, Yang D, et al. RFAConv: Innovating spatial attention and standard convolutional operation[J]. arXiv preprint arXiv:2304.03198, 2023.

[45] Chen J, Kao S, He H, et al. Run, don't walk: chasing higher FLOPS for faster neural networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 12021-12031.

[46] Shi D. TransNeXt: Robust Foveal Visual Perception for Vision Transformers[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 17773-17783.

[47] Liu W, Lu H, Fu H, et al. Learning to upsample by learning to sample[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 6027-6037.

[48] Cai X, Lai Q, Wang Y, et al. Poly kernel inception network for remote sensing detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 27706-27716.

[49] 刘云翔,马海力,朱建林,等.基于感受野注意力卷积的自动驾驶多任务感知算法[J/OL].计算机工程与应用,1-11[2024-09-13].

[50] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.

[51] 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.

[52] Li H, Li J, Wei H, et al. Slim-neck by GSConv: A better design paradigm of detector architectures for autonomous vehicles[J]. arXiv preprint arXiv:2206.02424, 2022.

[53] Wang J, Chen K, Xu R, et al. Carafe: Content-aware reassembly of features[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 3007-3016.

[54] 田青,王颖,张正,等.改进 YOLOv8n 的选通图像目标检测算法[J/OL].计算机工程与应用,1-12[2024-09-16].

[55] Ma X, Dai X, Bai Y, et al. Rewrite the Stars[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 5694-5703.

[56] Wang C Y, Yeh I H, Liao H Y M. YOLOv9: learning what you want to learn using programmable gradient information [J]. arXiv e-Print ,2024,arXiv:2402.13616.

[57] Ding X, Zhang X, Ma N, et al. Repvgg: Making vgg-style convnets great again[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 13733-13742.

[58] Chen Y F, Zhang C Y, Chen B, et al. Accurate leukocyte detection based on deformable-DETR and multi-level feature fusion for aiding diagnosis of blood diseases[J]. Computers in Biology and Medicine, 2024, 170: 107917.

[59] Ding X H, Zhang Y Y, Ge Y X, et al. UniRepLKNet: a universal perception large-kernel ConvNet for audio, video, point cloud, time-series and image recognition [J]. arXiv e-Print ,2023,arXiv:2311.15599.

[60] 杨琚钱,胡平,戴家树.基于 Yolov8-scG 神经网络的电动车头盔佩戴检测算法[J/OL].重庆工商大学学报(自然科学版):1-10[2024-08-14].

[61] 李茂, 肖洋轶, 宗望远, 等. 基于改进 YOLOv8 模型的轻量化板栗果实识别方法[J]. 农业工程学报, 2024, 40(1): 201-209

[62] 罗亮,郎霄,祖国庆,等.一种基于改进 YOLOv8 的气缸套缺陷检测方法[J/OL].中国机械工程:1-12[2024-08-14].

[63] 杨红欣,陈越,裴国权,等.基于轻量化 YOLOv8-FasterBlock 模型的云南小粒咖啡生豆分级方法[J].食品科学,2025,46(04):268-277.

[64] Jaderberg M, Simonyan K, Zisserman A. Spatial transformer networks[J]. Advances in neural information processing systems, 2015, 28.

[65] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2980-2988.

攻读学位期间获得的成果

[1] Li M, Yan N. The Tire Sidewall Key Information Region Detection Algorithm based on the Improvement of YOLOv5[C]//Proceedings of the 2024 7th International Conference on Image and Graphics Processing. 2024: 13-18. (第一作者, EI)

[2] 李萌阳,严楠.TireYOLO：轮胎胎侧压印文字检测算法[J/OL].南京信息工程大学学报,1-13[2025-03-11]. (第一作者, 北大中文核心)

攻读学位期间参与的科研项目

[1] 安徽未来技术研究院企业合作项目（2023qyhz11）
[2] 安徽工程大学校企合作项目（HX-2023-09-050）

致谢

写了太多的学术话，太累了，致谢写点直白的吧，虽然可能不太通顺。

三年的时间说过去就过去了，当年看到自己录取信息的时候的场景还在眼前，在从老家回市里的路上，我看到自己梦想的成真，当年差点考不上本科的我也成了一名实实在在如假包换的研究生，对于我来说，实现了人生真正的跨越。在这一方面，我感谢我的爸妈，你们本有机会让我去读花钱少的专科，但是还是在四年期间花了 20 万元左右让我读了一个本科，并且在我一战没上岸后让我能安心在家再考了一年，谢谢你们!

读研期间，首先我要感谢我的导师严楠老师，能成为您的学生是我读研来第一个幸运的事情，您对我的培养让我真正的从一个本科生的思维变成了能自己去提出论点，验证论点，实现论点的研究生思维，这对我来说是真正的成长，让我自身完成了一次提升，有了一个更加具体的，解决问题的能力。而且您的专业知识也让我受益匪浅，并且您严谨的治学态度、务实的工作作风和宽厚的人格魅力，都深深地影响着我；同时，也谢谢您把我带到 MVIC 实验室中，谢谢您带我进到了一个自己非常感兴趣，愿意深耕的一个领域和方向，而且让我的跨越不止是在学历上，而是在领域上，让我可以进入到一个我不读研就很难进入的领域，也让我现在能找到这么舒服的工作，再次对您表示由衷的感谢。并且在生活中，您对我的帮助也很大，只要能让我参加的学术会议您都让我去了，对比起来我过得真的很幸福!同时也感谢刘冬老师在本人研究生期间的帮助！

并且，我也要感谢 MVIC 其他帮助过的老师和同学们，让我能在读研期间如此顺利，离不开你们的帮助，也让我在 MVIC 实验室里收获了很好的让很多人都羡慕的实验室氛围与帮助，谢谢你们在我读研期间的帮助!

最后，在学术之外，要谢谢乐乐，其实我是一个特别容易钻牛角尖的人，但是我遇到了你，从遇到了你开始，我发现我的生活变得有了除学术之外的东西，你总是可以在我钻牛角尖不知道该怎么办的时候来安慰我，陪我度过那段日子，你总是不信我说的，你就像一个阳光照进了我的生活，但是工科做实验搞学术就是这样容易出不来效果然后难受，再加上我自己的性格，但是你的出现，能让我很快的走出难过的时候，每一次在操场上散步，去公园溜达，都让我的生活多姿多彩，而且你也教会了我如何爱自己如何对待自己的生活和感情，虽然前路还有一段，但是希望我们可以一直走下去！

在此，我再次向所有帮助过我的人表示衷心的感谢。未来的日子里，我将继续努力，对得起老师的教诲，对得起自己读研三年来帮助过我的人，为了美好的明天。

最后的最后，希望未来能去读博！