分类号：TP391　　学校代码：10363
密　级：公开　　学　　号：2220110123

安徽工程大学
Anhui Polytechnic University

硕士学位论文

题目　基于深度学习的交通标志牌文本检测方法研究

论文作者　柳 浩
校内导师　潘道远 副教授
校外导师　李公文 高级工程师
专业学位类别　机 械
研究方向（领域）　无人驾驶

论文提交日期：2025 年 5 月 20 日

分类号：TP391　　单位代码：10363
密　级：公开　　学号：2220110123

基于深度学习的交通标志牌文本检测方法研究

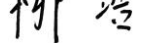
Research on Traffic Sign Text Detection Method Based on Deep Learning

学 生 姓 名 ：柳 浩
校 内 导 师 ：潘道远(副教授)
校 外 导 师 ：李公文(高级工程师)
专　　业 ：机 械
研究方向(领域)：无人驾驶

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　　05 月 20 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：柳浩　　指导教师签名：潘道远

日期：2025年05 月 20日　　日期：2025 年05月 20 日

基于深度学习的交通标志牌文本检测方法研究

摘 要

随着无人驾驶技术的不断发展，各种辅助技术相继问世，其中交通标志牌文本检测技术是无人驾驶的重要组成部分。无人驾驶汽车可以根据交通标志牌文本信息和地图信息，调整车辆的速度、行驶方向和其他相关参数，以确保行驶的安全性和合规性。本文将基于无人驾驶平台，以深度学习理论为基础，重点探索复杂环境下交通标志牌文本检测方法。

针对复杂环境下交通标志牌文本检测容易出现漏检和错检问题，提出了基于双向特征金字塔的交通标志牌文本检测模型（BPI-DBNet）。BPI-DBNet 采用综合通道空间注意力模块对骨干网络进行优化，减少多尺度特征提取过程中关键信息的损失。利用 Haar 小波下采样模块、自校准空间金字塔模块和高效多尺度注意力模块优化颈部网络双向特征金字塔，以提升特征表达能力和文本上下文关系。为了验证 BPI-DBNet 的适用性和鲁棒性，在数据集 CTST-1600、ICDAR2015 和 MLT-2017 上与其它文本检测模型进行对比实验。实验结果表明，BPI-DBNet 在三个数据集上的 F 值分别为 94.04%、85.03%和 74.29%，均优于其它文本检测模型。

针对无人驾驶汽车车载平台计算资源有限的问题，提出了基于旋转框的交通标志牌文本检测模型（IYOLOV8-OBB）。由于交通标志

牌文本设计较为工整，IYOLOV8-OBB 首先利用旋转框定位由拍摄角度引起的倾斜文本，然后适配计算资源需求较低的基于回归的文本检测方法。为了验证 IYOLOV8-OBB 的检测性能，选取较为权威的目标检测模型进行对比实验。目标检测模型用于交通标志牌文本检测时，需将检测头替换为旋转框检测头。实验结果表明：IYOLOV8-OBB 的 P 值和 F 值分别为 94.50%和 93.95%，均优于其它文本检测模型；IYOLOV8-OBB 与 BPI-DBNet 相比，检测性能略有下降，但 Params 值减小近 90 倍。

为了进一步满足无人驾驶汽车低计算资源的要求，首先利用 LAMP 剪枝算法对 IYOLOV8-OBB 进行剪枝处理，缩减模型大小，然后采用 CWD 知识蒸馏算法进行蒸馏，提升模型检测性能，最后搭建无人驾驶汽车交通标志牌文本检测系统进行检测实验。实验结果表明，实验车在城市道路、校园场景中均能准确检测并定位交通标志牌文本，充分体现了检测系统在实际道路环境中的鲁棒性和实用性。

关键词：无人驾驶，交通标志牌，文本检测，深度学习，模型轻量化

RESEARCH ON TRAFFIC SIGN TEXT DETECTION METHOD BASED ON DEEP LEARNING

ABSTRACT

With the continuous development of autonomous driving technology, various auxiliary technologies have emerged, among which traffic sign text detection plays a crucial role in autonomous driving systems. Autonomous vehicles can adjust their speed, direction, and other relevant parameters based on traffic sign text information and map data to ensure safe and compliant operation. This paper, based on an autonomous driving platform and grounded in deep learning theory, focuses on exploring methods for traffic sign text detection in complex environments.

To address the issues of missed and false detection in traffic sign text detection under complex environments, a traffic sign text detection model based on a bidirectional feature pyramid (BPI-DBNet) is proposed. BPI-DBNet optimizes the backbone network using a comprehensive channel-spatial attention module, reducing the loss of key information during multi-scale feature extraction. It employs a Haar wavelet down sampling module, a self-calibrating spatial pyramid module, and an efficient multi-scale attention module to optimize the neck network’s bidirectional feature

pyramid, enhancing feature representation and capturing text-context relationships. To validate the applicability and robustness of BPI-DBNet, comparative experiments were conducted on the CTST-1600, ICDAR2015, and MLT-2017 datasets with other text detection models. The experimental results demonstrate that BPI-DBNet achieves F-scores of 94.04%, 85.03%, and 74.29% on the three datasets, out-performing other text detection models.

To address the issue of limited computational resources on autonomous vehicle onboard platforms, a traffic sign text detection model based on rotation boxes (IYOLOV8-OBB) is proposed. Since traffic sign text is generally designed with a structured and orderly layout, IYOLOV8-OBB first uses rotation boxes to localize skewed text caused by varying camera angles, and then adapts a regression-based text detection approach that requires lower computational resources. To evaluate the detection performance of IYOLOV8-OBB, comparative experiments were conducted with several authoritative object detection models. In these experiments, the detection head of the models was replaced with a rotation box detection head for traffic sign text detection. The experimental results show that IYOLOV8-OBB achieves precision (P) and F-scores of 94.50% and 93.95%, respectively, outperforming other text detection models. Compared to BPI-DBNet, IYOLOV8-OBB exhibits a slight decline in

detection performance but reduces the model's parameter size by nearly 90 times.

To further meet the low computational resource requirements of autonomous vehicles, the LAMP pruning algorithm is first applied to prune the IYOLOV8-OBB model, reducing its size. Then, the CWD knowledge distillation algorithm is employed to distill the model, enhancing its detection performance. Finally, a traffic sign text detection system for autonomous vehicles is developed and used for experimental testing. The experimental results demonstrate that the test vehicle can accurately detect and localize traffic sign text in both urban roads and campus environments, fully showcasing the robustness and practicality of the detection system in real-world driving scenarios

KEY WORDS: Autonomous Driving, Traffic Signs, Text Detection, Deep Learning, Model Lightweighting

目录

插图清单

插表清单

第1章 绪论

1.1 研究背景和意义

城市化在全球范围内迅速推进，城市人口密度不断增加。随着城市扩展和人口流动，交通流量显著增加，给城市交通管理带来巨大挑战[1]。随着城市化进程的加快，构建高效智能的交通管理体系显得尤为重要。在大都市中，交通拥堵和交通事故频发，严重影响居民的日常生活与工作效率。联合国预测，到 2050 年全球城市人口将占总人口的 68%，这一趋势进一步加剧了对智能交通系统的需求。在此背景下，交通标志牌文本检测技术因其在实时交通指挥与信息获取中的关键作用，成为提升交通管理水平的重要支撑手段。因此，推动交通标志牌文本检测技术的发展，对于建设智慧城市和保障交通安全具有重要意义。

近年来，随着大众对自动驾驶技术需求的增加，该技术成为各大车企发展的重要方向[2]。自动驾驶车辆需要依赖各种传感器和算法，实时检测和识别道路上交通标志牌、交通信号灯以及车牌等信息，以确保行车安全和效率[3-6]。

深度学习和计算机视觉技术的不断发展为文本检测任务提供了坚实的技术基础。卷积神经网络（CNN）和循环神经网络（RNN）等模型的引入显著提升了检测的准确性与实时性，尤其是在复杂场景下表现出较强的鲁棒性。例如，近年来兴起的基于深度学习的光学字符识别（OCR）技术，能够高效地定位和识别自然场景中的文本信息，为交通标志牌文本检测等实际应用提供了关键支持。借助深度学习模型的强大特征提取能力，文本检测在多样化场景中的应用潜力不断拓展。

交通标志牌文本检测作为自然场景文本检测的一种，自然场景的文本检测方法能应用于交通标志牌文本检测场景中，但需针对交通标志牌文本的特性进行改进，以提高交通标志牌文本检测性能和实际场景的检测效果。交通标志牌文本检测的目的是确保交通标志牌文本能被准确的识别，如图 1-1 所示，交通标志牌文本出现在复杂的场景中（如店牌和交通标志牌文本在同一场景中）。交通标志牌文本检测模型要准确地检测出交通文本，而非其它文本，从而减少无人驾驶汽车

对其它文本的识别，避免浪费算力。若误检过多，会造成车辆误解交通环境，发生误判情况，导致发生交通事故。

现阶段文本识别模型相对成熟，识别准确率较高，只需对交通标志牌文本区域进行定位，把定位区域输入到文本识别模型中，就可对交通文本进行识别，因此如何准确的检测到交通标志牌文本区域就显得尤为重要。针对交通标志牌文本检测的研究较为有限，因此开展交通标志牌文本检测研究具有重要意义。

图 1-1 复杂环境下的交通标志牌文本

1.2 国内外研究现状

1.2.1 文本检测研究现状

文本检测方法最早出现在 20 世纪 90 年代。早期的方法通过基于滑动窗口和基于组件拼接的方式获取文本候选区域，利用人工设计的特征对这些候选区域进行验证，以此确定文本的位置。但人工设计的特征有较大局限性，难以有效检测复杂自然场景中的文本。

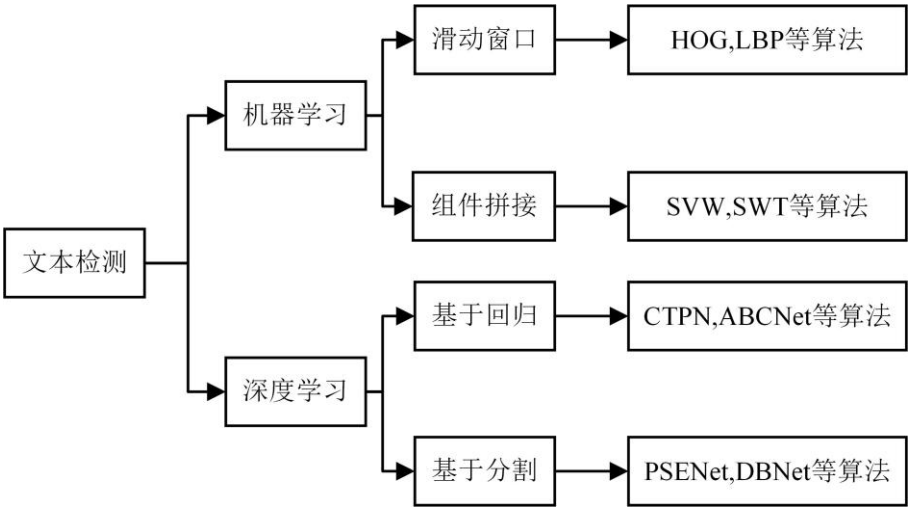


图 1-2 文本检测方法分类

近年来，深度学习技术的快速发展为文本检测带来了突破性的进展。研究者广泛采用卷积神经网络（CNN）对大规模文本图像数据进行训练，实现了对图像中潜在文本特征的自动提取。与传统人工设计特征的方法相比，深度学习模型能够更深入地挖掘图像中的语义信息，从而显著提升文本检测的准确性和鲁棒性。深度学习方法已成为当前文本检测研究的主流方向，并在各类实际应用中展现出广泛的适应性和良好的性能表现。由图 1-2 可知，目前的深度学习文本检测模型主要分为两类：基于回归的文本检测方法和基于分割的文本检测方法。

(1) 机器学习的文本检测方法

自然场景文本检测的研究起步较早，早在 20 世纪 90 年代初便已引起学术界关注。在初期阶段，研究人员尝试引入传统的机器学习方法，以克服手工设计特征在表达能力和适应性方面的局限。这一技术路线在当时取得了一定成效，并逐渐发展成为文本检测领域的主流方法。尽管受到时代技术水平的限制，但传统机器学习为自然场景文本检测奠定了重要基础，并为后续深度学习方法的兴起提供了实践经验与研究思路。基于传统机器学习的文本检测方法主要分成两类：一种是基于滑动窗口的方法，另一种是基于组件拼接的方法。

基于滑动窗口的方法采用自顶向下的扫描策略，对整幅图像逐步扫描，将每个窗口覆盖的区域视为潜在的文本候选区域。为了适应不同大小和长度的文本行，通常引入多尺度滑动窗口机制，从而提高对各类文本区域的检测能力。对于每个候选区域，提取一系列手工设计的特征，如梯度边缘特征、局部二值模式（LBP）、方向梯度直方图（HOG）等，并将其输入预先训练好的文本分类器中进行判别。文本分类器通常采用传统的机器学习方法，例如自适应增强算法[7]（Adaboost）、决策树[8]（Decision Tree）或随机蕨类[9]（Random Ferns），对滑动窗口内的图像区域进行分类，从而判断其属于文本区域还是背景区域。Wang 等人[10]通过多尺度的滑动窗口分类来检测图像中潜在的字符位置，选择随机蕨类作为分类器，该分类器适合多类别的检测，在训练和测试时具有较高的效率。通过该方法，系统能够识别图像中的字符，并进一步使用图结构（Pictorial Structures）框架将这些字符组合，形成单词的检测。

基于滑动窗口的文本检测方法在实际应用中存在一定的局限性，该方法通过

在图像上滑动窗口生成多个候选区域进行文本检测，计算成本较高，尤其是在处理图像中尺寸变化较大的文本时，效率较低。此外，由于该方法依赖固定大小的窗口，导致其在检测严重扭曲、弯曲或变形的文本时表现不佳，难以适应复杂自然场景中的多样性和变形性。因此，尽管滑动窗口方法在早期的文本检测中有一定应用，但其效率和适应性限制了其在复杂场景中的表现，促使了其他更高效、鲁棒性更强的方法的研究。

基于组件拼接的方法受传统机器学习分类器（如支持向量机[11]（SVM）和随机森林[12]）的启发，通过分析图像中具有相似视觉属性的区域，将文本组件进行分类，并将其逐步拼接为可识别的单词或文本行。Epshtein 等人[13]提出了一种基于笔画宽度的图像文本检测方法（SWT），其核心思想是通过估算图像中每个像素的笔画宽度来检测文本区域。Epshtein 用 Canny 边缘检测算法提取图像边缘信息，该算法沿边缘像素的梯度方向查找与之匹配的另一边缘像素，并将这两点之间的距离作为路径上像素的笔画宽度值。依据笔画宽度值的相似性对像素进行分组，从而构建出潜在的文本区域。SWT 能适应不同语言、尺寸及方向的文本检测任务，但在光照剧烈变化或背景复杂的图像中，检测效果会受到一定影响。Neumann 等人[14]采用最大稳定极值区域（MSER）方法进行文本检测，在输入图像中提取 MSER 区域作为候选文本区域，后借助分类器进行初步筛选，结合文本行的几何特性进一步去除非文本区域，从而实现对文本区域的准确定位。

基于组件拼接方法不受单一方向限制，能从多角度识别字符，为处理倾斜文本提供更广阔的视角，成为深度学习时代组件拼接技术的基础。但这类方法的准确性高度依赖组件分类的准确性，在歧义或环境条件较差的情况下，容易产生误分类和错误的字符拼接，导致文本检测失败。

传统的场景文本检测方法在基础结构理解和特征提取方面积累了丰富的经验。例如，基于组件拼接的方法在图像分割和特征理解方面，为深度学习的文本检测方法提供了启发。此外，传统方法在计算效率以及小数据集的处理方面具有独特优势，这些特点可以被深度学习模型借鉴并融合。因此，尽管深度学习在许多方面优于传统方法，但传统方法的核心思想和技术仍然具有重要的参考价值。

(2) 深度学习的文本检测方法

随着深度学习的发展，其改变了处理图像的基本思路，同时为文本检测领域提供了坚实的技术支撑。目前，文本检测主要分为两种方法：一种是基于回归的文本检测方法，另一种是基于分割的文本检测方法。

基于回归的文本检测方法因其结构简洁、效率较高，在文本检测研究早期阶段得到了广泛关注和应用。该方法受到传统目标检测技术的启发，通过建模文本的几何形状，将文本视为一种特殊目标，直接回归预测其边界框，有效应对图像中复杂的文本布局。相较于其他方法，该方法简化了检测流程，降低了对计算资源的依赖，具有较强的实用性。因此，基于回归的文本检测方法成为深度学习初期在文本检测任务中应用的重要技术路径。

在水平文本检测方面，Tian 等人[15]通过结合深度卷积神经网络和循环神经网络，在自然图像的卷积特征图上精确定位文本行。利用垂直锚定的回归机制、细粒度文本提议序列以及双向长短期记忆网络（Bi-LSTM）来捕捉图像中的上下文信息，从而提高文本检测的准确性和效率。Liao 等人[16]通过网络的多个层级直接预测文本的存在性和预设默认框的偏移量，实现对文本边界框的快速且精确回归。该方法采用独特的文本框层和不规则的 1×5 卷积核，以适应文本多变的长宽比，并通过多尺度输入提升对极端尺寸和长宽比文本的检测准确度。

在任意角度文本检测方面，Liao 等人[17]通过端到端的训练预测自然场景中任意方向的文本边界框，用四边形或旋转矩形适应文本的不同方向和形状。该网络采用特制的“长”卷积核和改进的锚框（Anchor Boxes），有效地覆盖文本区域并提升回归精度。Shi 等人[18]通过将文本分解为可局部检测的段（Segment）和连接段的链接（Link）。Shi 采用全卷积神经网络在多个尺度上进行密集检测段和链接，通过组合由链接相连的段，生成文本检测框。He 等人[19]通过引入文本特征对齐模块（TFAM）和位置感知的非最大抑制（PA-NMS），提高文本检测性能。TFAM 动态调整特征接收域以适应初步检测结果，PA-NMS 利用位置信息优化重叠检测框的合并过程。该方法采用实例化的 IoU 损失函数，以平衡不同规模文本实例的训练，确保在多尺寸文本上的一致性和准确性。在处理极端纵横比和多尺度变化的文本实例时，He 提出的方法展现出卓越的性能，实现文本边界框高效且精确的回归。

在弯曲文本检测方面，Liu 等人[20]通过回归多边形顶点，实现对自然场景中曲线文本的检测。Liu 用横纵偏移连接（TLOC）技术和循环神经网络（RNN）学习文本位置偏移，从而实现端到端的训练，提升检测的平滑性和准确性。Liu 还提出两种后处理方法：非多边形抑制（NPS）和多边形非最大抑制（PNMS）进一步提升检测精度。Liu 等人[21]通过参数化的贝塞尔曲线适应文本中任意形状。Liu 将文本检测问题转化为具有八个控制点的边界框回归问题，其中每个控制点的坐标通过最小二乘法从标注的多边形文本区域中计算得到。

基于回归的文本检测方法在处理水平方向排列的文本时具有良好的检测效果。然而，当面临形状复杂或背景杂乱的文本实例时，该方法的性能会下降。例如，在背景噪声较大或光照条件强烈的场景中，检测准确性明显降低；在文本排列紧密、字体尺寸过小或过大的情况下，也难以实现精确定位。因此，该方法在复杂场景中的适应性和鲁棒性仍存在一定局限，亟需进一步优化与改进。

基于分割的文本检测方法在复杂形状文本的定位中展现出较高的灵活性和精度。该方法借鉴图像分割技术，通过像素级的掩膜间接标注文本区域或边界，通常采用 U 形网络结构生成分割结果，并结合特定的后处理技术，对检测到的文本区域进行精细分析，从而实现高精度的边界框定位。依托于像素级建模能力，该方法尤其适用于处理形状不规则或排列松散的文本，成为提升检测准确性的有效手段。Wang 等人[22]通过多尺度预测，结合渐进式尺度扩展算法，使检测区域能够从最小尺度的核心区域逐步扩展，直至覆盖文本实例。Wang 提出的算法在区分紧密相邻文本实例方面表现出色，同时能有效处理复杂背景以及不同颜色、字体和尺度的文本。Liu 等人[23]通过自底向上的路径增强和自适应特征的池化技术，优化神经网络中的信息流动，提升实例分割的准确性，其利用低层特征中的精确定位信号，缩短低层到顶层的特征传递路径。Liu 还通过引入额外的分支，捕获每个提议不同的视图，增强掩码预测的能力。Liao[24]通过引入可微分二值化模块，并结合概率图与自适应阈值图协同优化，实现对任意形状文本的精确检测。同时，Liao 还简化了后处理流程，提升了检测速度和准确性。有效解决了传统分割方法中二值化后处理不能优化的难题。

基于分割的文本检测方法在复杂文本环境中展现了显著的优势。与基于回归

的方法相比，基于分割的检测方法能够更精细地提取文本区域的轮廓，特别是在处理任意形态的文本和多语言文本时，表现尤为出色。这种方法通过像素级的精确定位，提升了文本检测的完整性和准确性。此外，得益于其对文本形态变化的高度适应性，基于分割的检测策略在复杂多变的自然场景中展现出更强的鲁棒性和泛化能力。因此，基于分割的文本检测方法已成为提升检测精度和处理复杂场景的关键技术之一。

1.2.2 交通标志牌文本检测难点和方法

无人驾驶技术发展迅速，作为其中的重要组成部分，交通标志牌文本检测取得了一定的发展。交通标志牌文本检测是自然场景文本检测中的一个特殊分支，不仅面临着通用自然场景下的诸多技术难题，还有特定于交通场景的独特挑战。

(a) (b)

图 1-3 交通标志牌文本语言多样性

在中国，交通标志牌文本涉及多种语言，这为文本检测带来了挑战。标识中通常包括中文、英文以及少数民族语言，每种语言都有其独特的字符集。英文由 52 个字母（大小写）组成，中文则包含多达 6763 个常用汉字，这些汉字已在 1980 年通过国标 GB2312-80 进行编码定义。除此之外，中国还有超过 80 种少数民族语言和文字。这种语言的多样性使得交通标志牌文本检测任务变得更加复杂，需要处理不同字符集和语言的识别问题。图 1-3(a)中，交通牌包含有中文、拼音、英语以及阿拉伯数字。图 1-3(b)表示在少数民族地区交通标志牌会同时出现少数民族语言和中文的情况。通过对比可知，不同语言或同一语言中不同文本类型的交通标志牌文本，在视觉特征上存在显著差异。

(a) (b)

图 1-4 交通标志牌文本排列方向多样性

交通标志牌上的文本排列方向具有多样性，主要包括横向和竖向两种形式。这种排列方向的变化使得文本检测更加复杂，尤其是在不同拍摄角度下，文本定位的准确性受到影响。如图 1-4(a)和图 1-4(b)所示，车载摄像头通常无法正对交通标志牌拍摄，因此，实际拍摄的文本可能呈现不同的角度。这使得文本的排列方向在实际检测中可能出现多种变化，从而增加了检测的难度。

(a) (b)

图 1-5 交通标志牌文本行大小和比例差异、字符稀疏度变化

交通标志牌上的文本行大小、比例差异以及字符稀疏度的变化，显著增加了文本检测的复杂性。不同的文本行可能有不同的字体大小、行间距和字符间距，且在不同标志牌上，字符的排列可能较为稀疏或紧凑，这对检测系统提出了更高的精度要求。图 1-5(a)中，左侧交通牌的文本行比右侧小，且左侧三个文本行的长度不同，同时“创业路”文本稀疏；而图 1-5(b)中，每行单词间距和文本的长宽比都不相同，“春风路”文本稀疏。因此，检测文本行时须考虑文字的尺度、文本行的长宽比以及文本的稀疏度，这进一步增

加了文本检测的复杂性。

2019 年何璇等人[25]提出了一种基于深度学习的交通标志牌文本检测模型，该模型由预处理模块、全卷积网络（FCN）和简化的后处理模块组成。其中，预处理模块用于应对复杂场景，网络中引入尺度转换层（Scale-transfer Layer）以提升计算效率，加快网络推理速度。此外，何璇构建并公开 CTST-1600 数据集，为交通标志牌文本检测领域提供了重要的公共数据资源。2020 年彭西明[26]对基于对比度受限的直方图均衡化的图像增强方法进行优化，并构建了一个字符型交通标志数据集。此外，彭西明提出了两种检测算法：一是基于深度学习的图像语义分割检测算法，用于精准定位字符型交通标志；二是改进的 YOLOv3[27]文本检测算法，专门用于字符型交通标志区域的文本检测。这些方法显著提升了字符型交通标志的检测准确率和效率。2022 年胡高丽等人[28]通过基于 PSENet 文本检测模型进行交通标志牌文本检测，并引入特征增强模块（FEM）和空洞卷积，以增强模型的感受野和多尺度特征融合能力，提高交通标志牌文本检测精度。2023 年李金鹏[29]在文本检测领域对 DBNet[24]算法进行了改进。针对 DBNet 在特征提取过程中可能导致信息丢失的问题，李金鹏用可变型卷积[30]替换了骨干网络中的 3×3 标准卷积，并在特征提取的下采样阶段引入了全局注意力机制。此外，为解决数据不足的问题，该研究构建了交通标志牌文本数据集，并最终开发了交通标志文本检测与识别系统。2024 年朱彦斌等人[31]提出了一种基于多粒度特征增强网络（MFENet）的交通标志牌文本检测方法，该方法通过设计多粒度文本特征增强模块和联合损失函数，提高了车载摄像头对交通标志牌文本的检测性能。在交通标志牌文本数据集上，MFENet 取得了良好的检测精度，并在自然场景文本数据集上展现出良好的泛化能力和鲁棒性。

在自然场景文本检测竞赛中，模型不仅需准确定位文本区域，还要实现词组级别的完整文本序列检测。这对模型的文本感知能力、语义理解能力及字符关联机制提出了更高要求。相比通用自然场景中的文本检测任务，交通标志牌文本检测更具挑战性，因交通标志牌文本语言种类多样性、文本排列方向多样性以及文本行大小和比例差异、字符稀疏度变化等因素。这些复杂性使得传统文本检测方法在交通标志牌场景中难以取得理想效果。

尽管已有模型引入多尺度感知结构、注意力机制和特征增强模块等方法，有效提升了对复杂文本形态的处理能力，但面对语言种类多样、排列方向多变及字符稀疏等问题，交通标志牌文本仍易出现漏检和误检。自动驾驶对检测系统的实时性要求更高，而深度学习模型计算资源消耗较大，尤其在多尺度特征提取与多语言识别场景中更为明显。为此，后续研究需在确保检测精度的前提下，持续优化模型结构，降低计算复杂度，提升检测速度，以提高交通标志牌文本检测模型的实用性与可靠性。

1.3 主要研究内容

将交通标志牌中的文本型交通标志作为研究对象。利用深度学习知识来优化现有的文本检测方法。考虑到 1.2.2 节中交通标志牌文本检测的难点，旨在解决交通标志牌文本检测中的漏检和错检问题，以此来提高交通标志牌文本检测性能。其次针对模型较大，计算资源消耗较大的问题，需采用轻量化文本检测方法和模型轻量化技术，来实现交通标志牌文本检测模型的高检测性能和轻量化。最后，将高检测性能和轻量化的交通标志牌文本检测模型嵌入到智能网联测试装调实验车进行实时检测。研究内容关系图，如图 1-6 所示。

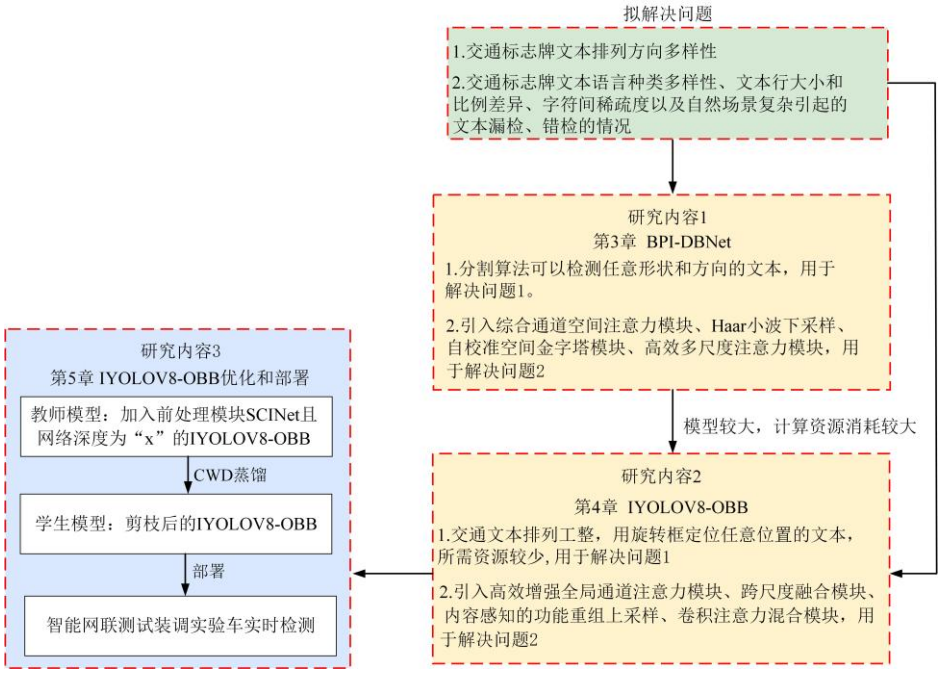


图 1-6 研究内容关系

（1）针对交通标志牌文本检测中漏检和错检的问题，提出 BPI-DBNet 的交通标志牌文本检测模型。该模型在 DBNet 文本检测模型的基础上进行优化。在骨干网络中引入综合通道空间注意力模块（ICSAM）。减少特征提取过程中的信息的损失，保留更多交通标志牌文本定位和分类信息，颈部网络使用双向特征金字塔网络模块（BiFPN），并引入 Haar 小波下采样模块（HWD）、自校准空间金字塔模块（SC-SPPF）以及高效多尺度注意力模块（EMA）对 BiFPN 进行优化。HWD 减少下采样过程中特征丢失和削弱。SC-SPPF 增强高层次特征对文本几何大小的表达能力和上下文关系。EMA 善于捕捉特征图的局部细节和全局上下文，增强特征丰富度和区分度。

（2）针对 BPI-DBNet 模型较大，计算资源消耗较大的问题。提出基于旋转框的交通标志牌文本检测模型 IYOLOV8-OBB。由于交通标志牌文本设计较为工整，IYOLOV8-OBB 首先利用旋转框定位由拍摄角度引起的倾斜文本，然后适配计算资源需求较低的基于回归的文本检测方法。为了减少漏检和错检的现象。在骨干网络中引入高效全局通道注意力模块（EEGCA），该模块增强了全局视角信息和不同尺度大小的特征。引入跨尺度融合模块（CCFM）、引入内容感知功能重组上采样模块（CARAFE）以及卷积注意力混合模块（CAMixing）对颈部网络进行优化，CCFM 增强了对尺度变化的适应性。CARAFE 增大感受野，聚合上下文信息，增强对重要内容的感知。CAMixing 捕捉全局上下文信息和局部特征。

（3）针对 IYOLOv8-OBB 模型，引入 LAMP 方法对其进行剪枝优化。LAMP 通过识别并剔除对模型性能贡献较小的冗余参数，即离群点，有效降低了模型的复杂性，并生成轻量化的学生模型。尽管剪枝操作在一定程度上会导致检测性能下降，但随后引入的 CWD 知识蒸馏策略，通过教师模型对学生模型进行特征层级的指导学习，在不增加模型体积的前提下，有效弥补了剪枝带来的性能损失。最后就将蒸馏后的模型部署到无人驾驶汽车交通标志牌文本检测系统进行检测实验，验证优化后的 IYOLOv8-OBB 模型再在实际场景中的应用效果。

1.4 论文组织结构

本文组织结构分为：

第一章：绪论。阐述了交通标志牌文本检测的相关背景及其研究意义。描述文本检测和交通标志牌文本研究的现状，并进一步分析交通标志牌文本检测面临的问题，明确主要研究内容，简要介绍论文的组织结构。

第二章：文本检测基本理论。对交通标志牌文本检测涉及的深度学习理论和知识进行概述，包括交通标志牌文本检测模型中主要的三种结构：骨干网络、颈部网络和检测头以及模型轻量化中剪枝和知识蒸馏。此外，还介绍了实验评价标准和实验数据集。

第三章：基于双向特征金字塔的交通标志牌文本检测。介绍交通标志牌文本检测模型 BPI-DBNet 的基本原理、网络结构以及改进方案中四个模块的特点和优点。通过实验数据和模型检测可视化验证模型的有效性，并提出模型缺点，为下一章研究做铺垫。

第四章：基于旋转框的交通标志牌文本检测。为解决第三章模型过大，消耗计算资源较多的问题，提出基于旋转框的交通标志牌文本检测方法（IYOLOV8-OBBO）。介绍该方法的基本原理、网络结构以及改进方案中四个模块的特点和优点，通过实验数据和模型检测可视化验证模型的有效性。

第五章：交通标志牌文本检测系统搭建。IYOLOV8-OBB 的优化，介绍模型优化中剪枝与知识蒸馏所需的算法，并阐述知识蒸馏中教师模型与学生模型生成步骤。最后搭建无人驾驶汽车交通标志牌文本检测系统进行检测实验。

第六章：总结与展望。总结本文工作，提出研究的不足。

第2章 基于深度学习的文本检测基本理论

为提升交通标志牌文本检测的精度与效率，理解深度学习模型的基本结构和轻量化策略至关重要。简要介绍文本检测模型的三个核心组成部分：骨干网络、颈部网络和检测头，并阐述模型轻量化的方法，包括模型剪枝与知识蒸馏。介绍评价检测性能的常用指标以及所采用的数据集，为后续模型设计与实验验证提供理论基础和参考依据。

2.1 深度学习

2.1.1 骨干网络

(1) 残差网络

网络深度是影响目标检测模型分类与识别性能的关键因素。理论上，较深的网络能够提取更复杂、更抽象的特征，从而提升模型表现。然而，随着网络深度的增加，梯度消失或梯度爆炸等问题逐渐显现，严重影响模型的训练效果。在反向传播过程中，梯度需层层传递，每经过一层都会与该层权重的导数相乘。当导数较小时，梯度将快速衰减，导致深层网络的权重更新缓慢；反之，当导数较大时，梯度会迅速放大，引发梯度爆炸，使网络训练过程变得不稳定。为了解决这个问题，残差网络（ResNet[32]）引入了残差块结构，如图 2-1 所示。该结构通过跨层连接引入恒等映射，有效缓解了深层网络中的梯度传递困难，显著提升了训练的稳定性与模型的表现能力。

ResNet 用跨层连接来构建残差模块，使梯度信息更容易传递到浅层网络。残差网络提供了残差学习的概念。残差学习是网络中每个残差块的输出和输入的差值。假设输入 X，通过恒等映射得到输出 Y，网络为恒等映射 $F(X)$。若网络中存在一些非线性的变换（卷积，池化等），则 $F(X)\neq X$，称为残差块。

若学习的特征是 $H(X)$，则残差学习的表达式为 $F(X)=H(X)-X$。为了更好的学习原始特征，特征表达式记作 $F(X)+X$，若残差为 0，说明堆叠层执行是恒等映射，这有助于提升网络性能。在实际的情况中，残差通常不为 0，因此堆叠层能

从输入的特征中学习到新信息，从而进一步提高网络性能。

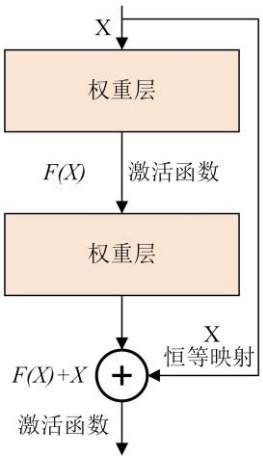


图 2-1 残差块网络结构

(2) MobileNet

随着深度神经网络结构日益复杂，模型层数和参数量的增加虽然提升了检测精度，但也带来了训练时间增长和资源消耗加剧等问题，尤其在计算资源受限的场景中更为突出。因此，如何在保持模型精度的同时有效减少参数量和模型大小，已成为当前研究的关键方向。为应对这一挑战，MobileNet[33]提出了一种轻量化网络架构，通过深度可分离卷积等结构优化手段，显著降低模型复杂度和计算成本，同时尽可能保持较高的准确率，使其更适用于边缘设备和嵌入式系统等资源受限环境。

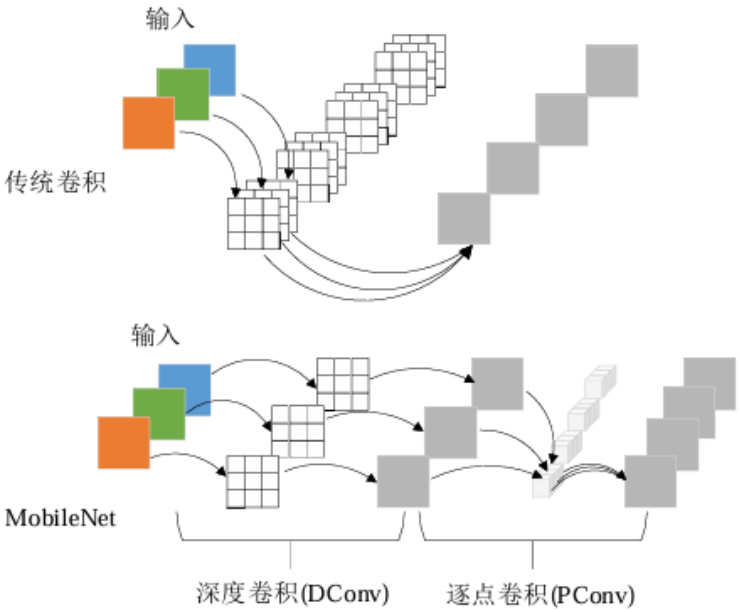


图 2-2 传统卷积与深度可分离卷积网络结构

由图 2-2 可知，MobileNet 使用深度可分离卷积代替传统卷积，在检测精度

变化较小的情况下，降低网络的计算量。传统卷积通过学习卷积核参数，对输入数据进行空间滤波和特征混合，生成新的特征图。而深度可分离卷积则优化了传统卷积，将空间滤波和特征混合分为两个独立的步骤：深度卷积（DConv）对每个输入通道分别应用独立的卷积核进行空间滤波。逐点卷积（PConv）使用 1x1 卷积核对 DConv 的输出进行通道间的线性组合，生成最终的特征图。该方法减少了模型的计算量和参数量，同时在一定程度上保持模型的准确性。

2.1.2 颈部网络

特征金字塔网络（FPN[34]）的核心理念是构建多层次特征金字塔，通过自顶向下的路径和横向连接，在不同层级之间融合特征，从而整合低层的细节信息和高层的语义信息，提升对不同尺寸目标的感知能力。FPN 生成的特征图在各个尺度上均富含语义信息，为精确的多尺度目标检测提供支持。

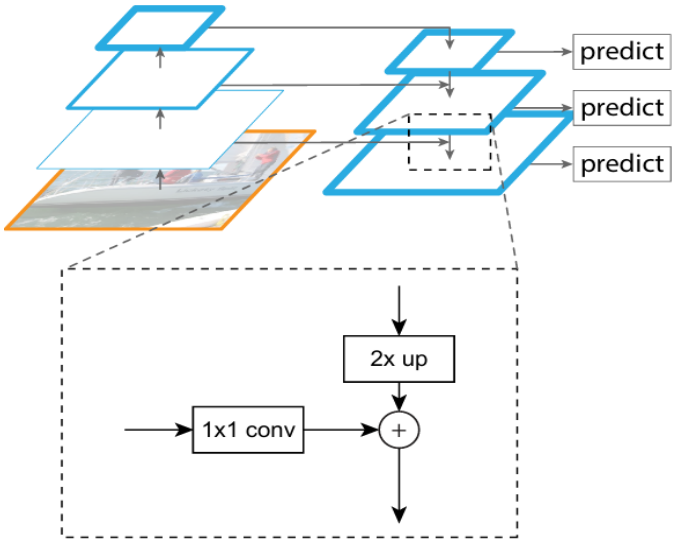


图 2-3 FPN 网络结构

由图 2-3 可知，FPN 的架构由自底向上路径、横向连接、自顶向下路径、特征融合以及多尺度特征输出组成。自底向上路径通过多层卷积逐步提取和细化输入图像的特征，生成不同分辨率的特征图。横向连接则在各层级上，将自底向上路径中的特征图与上采样后的特征图合并，实现不同层级间的信息共享和整合。自顶向下路径中的上采样，将高层语义信息与底层的高分辨率特征相结合，以增强特征图的语义和空间信息。特征融合过程中，先通过 1×1 卷积调整通道数，再将上采样后的特征图与自底向上的特征图相加，以进一步增强特征表征能力。融

合后的特征图经 3×3 卷积处理，生成一组空间尺寸相同但语义深度不同的多尺度特征图，从而支持精确的多尺度目标检测。

2.1.3 检测头

(1) 可微分二值化检测头

可微分二值化检测头[24]（DB）网络结构如图 2-4 所示。DB 检测头流程（红色箭头），输入的图像通过分割网络生成概率图和阈值图。该阈值图由网络学习得到，可为每个像素点提供最优的二值化阈值。利用概率图和阈值图，通过可微分的近似二值化函数生成近似二值图。由于该函数可微分，因此允许在训练期间通过网络的反向传播进行优化。相比之下，传统的文本检测流程（蓝色箭头）则有所不同。输入图像首先被送入图像分割网络，该网络负责生成概率图，该概率图表示图像中每个像素属于文本区域的可能性。该概率图通过固定的阈值转换成二值图，由于该过程不可微分，因此无法在网络训练过程中进行优化。最终，通过启发式技术（如像素聚类），将二值图中的像素点分组成文本实例，生成文本检测结果。

DB 检测头的设计使文本检测能在训练期间通过反向传播进行端到端的优化，从而提高检测的准确性和效率。

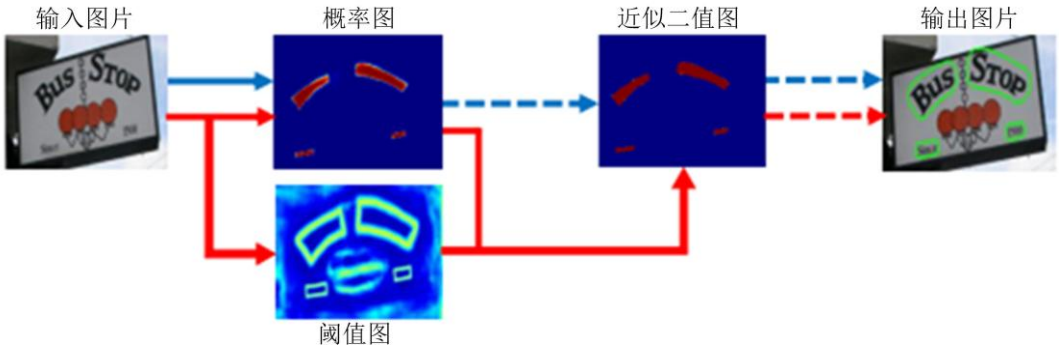


图 2-4 可微分二值化检测头流程图

(2) 解耦检测头

YOLO 系列自 YOLOv2[35]起采用耦合检测头设计，该设计采用 1×1 卷积层，同时预测类别分数、边界框回归参数以及物体存在概率，从而简化模型结构。但 YOLOX[36]的研究者发现，该耦合方式可能会导致分类和回归任务之间产生冲突，从而影响模型的精确度和训练的收敛速度。

在此基础上，YOLOv8[37]在目标检测模型中引入了一种新的检测头结构，即解耦检测头（Decoupled Head），如图 2-5 所示，该结构将原本耦合的分类与边界框的回归任务分开，形成两个独立的预测分支。YOLOv8 通过解耦的方式优化了检测头，使分类和回归任务能更专注于特定目标，从而提升整体检测性能。

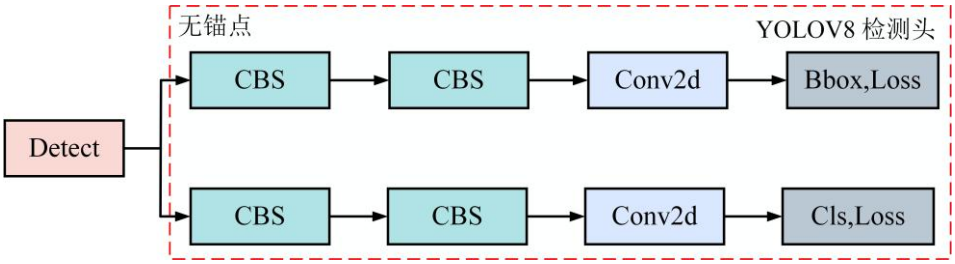


图 2-5 解耦检测头网络结构

YOLOv8 通过采用无锚点（Anchor Free）技术，解决了传统基于锚点（Anchor based）目标检测方法中手动设计的锚框尺寸受限、锚框生成计算量大以及正负样本不均衡等关键问题。该改进不仅提升了检测任务的效率和灵活性，还降低了超参数调优的难度。

2.2 模型轻量化

深度神经网络的模型训练与推理是两个关键阶段。其中，训练阶段是模型学习数据特征并调整参数的过程，而推理阶段则是利用学习到的参数进行预测。尽管深度模型凭借庞大的参数量和复杂结构具备出色的表征能力，但过多的参数可能会降低推理速度并增加部署难度。因此，模型剪枝等压缩技术应运而生，通过减少冗余参数，不仅缩小了模型大小，使其便于部署，加快推理速度，降低了计算资源需求和能耗。

2.2.1 模型剪枝

在深度学习领域，神经网络模型有海量的参数，其中部分权重较小的参数对模型预测结果影响有限。模型剪枝技术通过识别并移除这些影响较小的参数，从而实现模型精简。这一过程涉及两个关键步骤：确定需要剪除模型的哪些部分，选择合适的标准进行剪枝。

根据剪枝方式的不同，模型剪枝可分为非结构化剪枝和结构化剪枝。非结构

化剪枝主要针对神经网络中的单个神经元或权重进行优化。例如，经典的Dropout[38]算法，该算法通过随机关闭某些神经元来模拟剪枝效果，而Dropconnect[39]算法则通过随机将一些权重设为 0 来减少模型复杂度。这类方法不改变网络的拓扑结构，而是在权重层面上进行优化，从而实现模型的简化和加速。

结构化剪枝是一种有组织的剪枝策略，通过系统性地移除整个通道或滤波器，以减少模型的参数和计算量。与非结构化剪枝不同，结构化剪枝并非随机地删除单个权重，而是根据权重对模型性能的贡献度，有选择性地删除一组权重。Network Slimming[40]算法是结构化剪枝的代表性算法，其主要流程如图 2-6 所示。

图 2-6 Network Slimming 算法流程图

在剪枝标准上，通常考虑以下三个主要因素：权重的大小、梯度幅度以及局部剪枝的策略。权重较小的通道对模型的贡献不大，会被优先考虑剪除。梯度幅度较小的权重变化不大，其对模型的贡献有限，也会被剪除。而局部剪枝则是一种平衡策略，通过在模型的不同层或组内进行适度剪枝，避免全局剪枝可能带来的过度简化和模型性能下降。

2.2.2 知识蒸馏

模型剪枝技术能够有效地减少模型尺寸和冗余参数，降低部署成本，并使模型能适配嵌入式等资源受限的设备。这一过程可能会移除部分携带有用特征信息的通道或滤波器。即使对剪枝后模型进行再训练，其检测或识别的精度仍可能受到影响，导致性能下降。因此，为了弥补剪枝带来的信息损失，提高剪枝后模型的性能，知识蒸馏成为了一种有效的解决方案。

由图 2-7 可知，知识蒸馏可概括为教师模型辅助学生模型，以加速训练并提升性能。教师模型凭借其更为复杂且成熟的架构，将自身所蕴含的知识高效传递给结构相对简单、更小的学生模型（如剪枝后模型），从而提升学生模型的学习能力。凭借教师模型的精准指导，学生模型能够在较短的时间内，达到较高的性能表现。

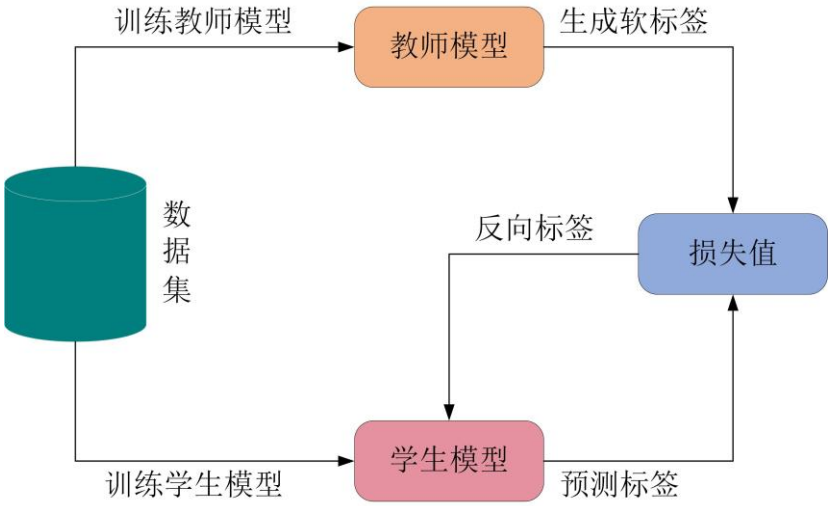


图 2-7 知识蒸馏过程

2.3 检测评价标准

为了评估文本检测方法的性能，采用常用的文本检测指标作为评估标准。文本检测的评价标准包括：准确率（P）、召回率（R）和调和平均数（F）。评价指标定义为

$$P = \frac{TP}{TP + FP} \tag{2-1}$$

$$R = \frac{TP}{TP + FN} \tag{2-2}$$

$$F = \frac{2 \times P \times R}{P + R} \tag{2-3}$$

式中：TP 是正类预测为正类的数目，FP 是负类预测为正类的数目，FN 是正类预测为负类的数目。

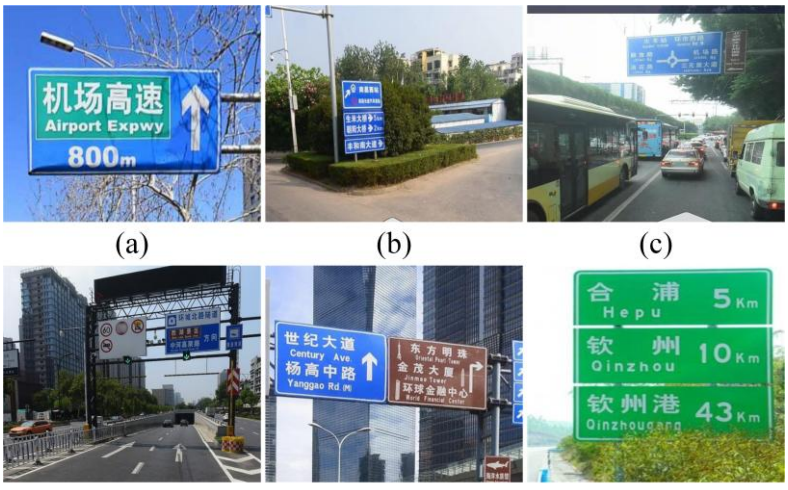
式(2-3)中，F 值是准确率和召回率的调和平均值。提高 F 值通常意味着模型在减少漏检和错检之间取得了较好的平衡，模型检测性能得到提升。

为了更好的测试交通标志牌文本检测模型的性能，使用模型大小（Params）和每秒检测帧数（FPS）作为评价指标。

2.4 数据集介绍

CTST-1600 是中国交通标志牌文本数据集，由何璇[25]提出。该数据集包含 1600 张图片，其中训练集有 1320 张，测试集有 280 张，是交通标志牌文本检测

模型主要的研究对象。



(a) (b) (c)

(d) (e) (f)

图 2-8 CTST-1600 数据示例

该数据集有 1600 张图片，由图 2-8 可知，该数据集涵盖了多种拍摄角度与方向，并呈现出阴暗、曝光等不同程度的光照条件。此外，交通标志牌文本标志形状各异，自然背景复杂，这些因素都增加交通标志牌文本检测的难度。该数据集涵盖了众多交通标志牌文本场景，如高速公路，城市街景，景区等。

ICDAR 2015 数据集是 2015 年场景文本检测竞赛所使用的数据集。该数据集为纯英文，且配有相应的注释。数据集内的所有图像分辨率均设定为 1280×720。该数据集包含 1500 张图片，其中训练集有 1000 张，测试集有 500 张。

(a) (b) (c)

(d) (e) (f)

图 2-9 ICDAR2015 数据集示例

该数据集包含来自多种场景和环境的图像，如书籍、网络图片、街景、场馆、商店以及广告牌等。由图 2-9 可知，图像种文本形式多样性，包含水平、垂直、弯曲以及曲线文本。数据集中每个文本行均标注。同时还提供了字符级别的标注

信息，包括字符的类别和位置信息。

MLT-2017 数据集是一个多语言文本数据集，涵盖了来自 6 种不同文字的 9 种语言。该数包含 9000 张图片，其中训练集有 7200 张，测试集有 1800 张。

(a) (b) (c)

(d) (e) (f)

图 2-10 MLT-2017 数据集示例

由图 2-10 可知，MLT-2017 数据集包含来自多种场景和环境的图像，如街道标志、广告牌、商店招牌、文档以及书籍等不同环境的文本图像。文本排布形式多样，包括水平、垂直、斜向和曲线排列。数据集中每个文本行均标注，同时还提供了字符级别的标注信息，包括字符的类别和位置信息。由于涉及多种语言、字体及复杂背景，该数据集对模型的泛化能力提出了较高的挑战。

2.5 本章小结

本章介绍了深度学习文本检测中三个主要部分：骨干网络，颈部网络以及检测头的典例算法。介绍了模型轻量化的方法：剪枝和知识蒸馏的基本原理。阐述了实验评价的标准和所使用的数据集。为后续提出的交通标志牌文本检测模型奠定理论基础。

第3章 基于双向特征金字塔的交通标志牌文本检测

3.1 引言

文本检测作为目标检测的一种特殊形式，与传统的实体目标检测存在显著差异。普通目标检测对象如植物、动物或飞机等具有明确的结构和整体性，即便候选框存在一定偏差，模型仍可通过局部特征进行有效识别。而文本由多个独立字符组成，字符之间存在空隙，缺乏连续性和封闭边缘，整体结构较为松散。这使得文本检测对定位精度要求更高，稍有偏差便可能导致字符漏检或错检。此外，文本信息具有高度语义严谨性，任何一个字符的误检或漏检都可能引起语义上的重大偏差，进而影响后续的识别与理解任务。因此，文本检测系统必须具备更高的精度与鲁棒性，以确保信息提取的完整性与可靠性。

针对交通标志牌文本排列多样性问题，采用基于分割的 DBNet[24]算法进行检测，该算法可以检测任意方向的文本。同时，因交通标志牌文本语言多样性、自然场景复杂以及交通标志牌文本行大小和比例差异、字符稀疏度的变化等问题，DBNet 模型在检测过程中易出现漏检和错检现象，影响整体检测精度。因此，需在 DBNet 模型基础上引入针对性改进策略，以增强模型对复杂场景中文本特征的感知能力，从而有效提升交通标志牌文本检测的鲁棒性与准确性。

3.2 交通标志牌文本检测模型

基于 DBNet 文本检测模型，提出 BPI-DBNet 的交通标志牌文本检测模型。其网络结构如图 3-1 所示，该模型由三个部分组成：骨干网络、颈部网络和检测头。主要对骨干网络和颈部网络进行改进，引入 ICSAM 模块来优化骨干网络 ResNet18（Improved ResNet18，IResNet18），以进行多尺度特征的提取，从而获得分辨率为 1/4、1/8、1/16 和 1/32 的特征图 $C2$、$C3$、$C4$ 和 $C5$。颈部网络用改进的 BiFPN（Improved BiFPN，IBiFPN）进行特征融合，引入 HWD 下采样模块，获得分辨率为 1/4、1/8、1/16 和 1/32 的特征图 $F2$、$F3$、$F4$ 和 $F5$。$F5$ 特征图通过 SC-SPPF 模块后，与剩下的三个特征图融合，融合后的特征图通过 EMA 模

块，获得分辨率为 1/4 特征图 F。在检测头部分，通过特征图 F 来预测概率图和阈值图，再通过可微分二值化获得近似二值图，以此获得检测结果。

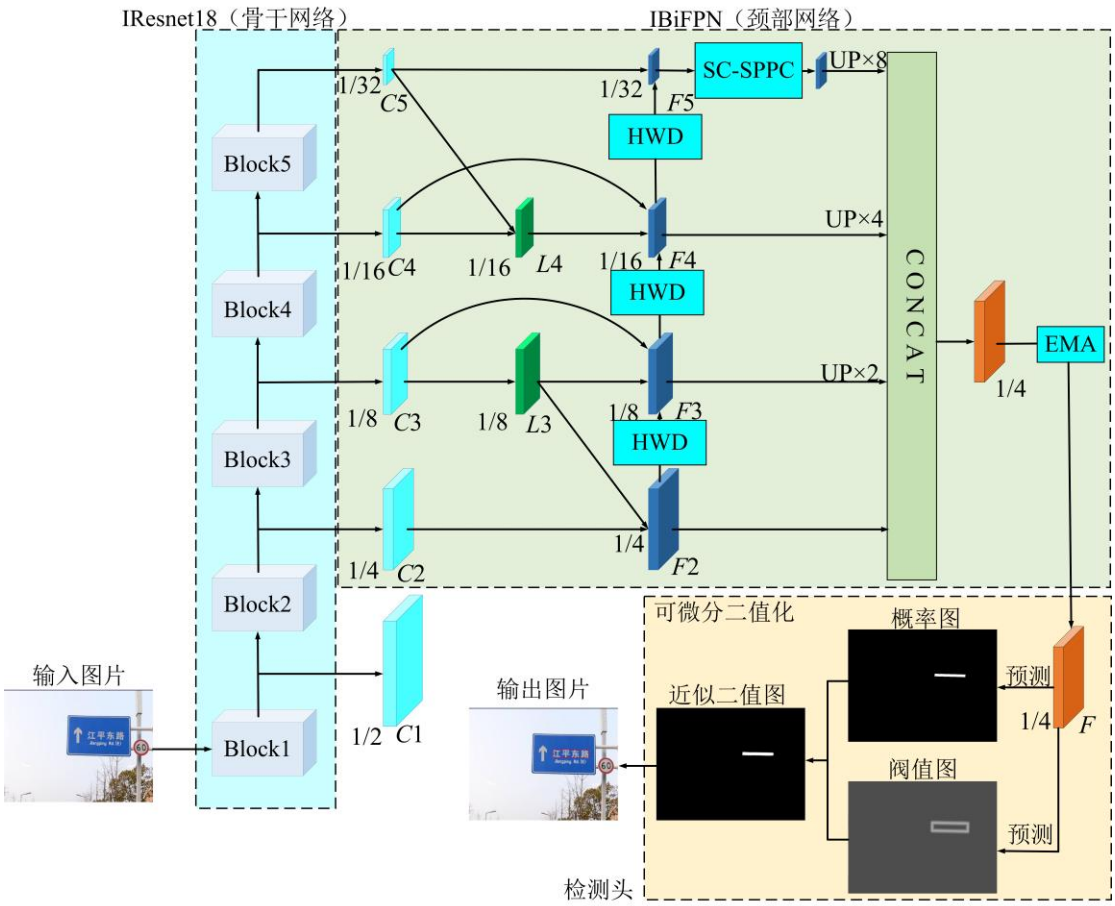


图 3-1 BPI-DBNet 交通标志牌文本检测模型

3.2.1 综合通道空间注意力模块

由图 3-2 可知，综合通道空间注意力模块（ICSAM）包括两个模块：通道注意力模块和空间注意力模块。在通道注意力模块中，先通过全局平均池化对特征图进行处理，以压缩其空间维度，再利用具有降维作用的全连接层来提取不同通道之间的关系。将降维后的特征通过多个大小相同的并行全连接层进行处理，并聚合这些层的输出，进而学习到全局的特征。后采用无压缩比例的全连接层，将汇聚后的特征恢复到原始特征图的形状。恢复后的特征与原始特征图相加，形成残差连接，有效促进信息的流通和梯度的传播。

在空间注意力模块中，先利用两个卷积层对输入特征图进行处理，从而学习特征图的权重。权重经过 Sigmoid 函数，将其值限制在 0 到 1 之间，进而形成注

意力权重。把原始的输入特征图与这些注意力权重进行逐元素相乘，以此来增强或抑制不同区域的特征响应，从而提升模型对于重要特征的关注度。ICSAM 模块中，r 的取值均为 4，旨在减少计算复杂度，提高模型的计算效率。

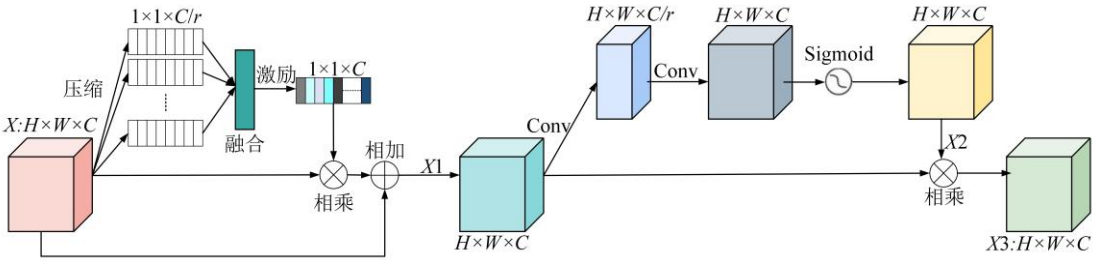


图 3-2 ICSAM 网络结构

输入特征图 $X{:}H\times W\times C$，则输出的特征图 X_1 为

$$X_1 = X + X \otimes Ex(\sum_{i=1}^{r} Sq_i(X)) \tag{3-1}$$

式中：$Sq_i(X)$ 表示通过不同分支的全连接层来实现的压缩操作，$Ex(X)$ 是激励操作，$\sum_{i=1}^{r} Sq_i(X)$ 表示分支的输出被融合，其中 r=4，$\otimes$ 为矩阵相乘。

得到的特征图 X_2 为

$$X_2 = \sigma(\mathcal{F}(\mathcal{F}(X_1)) = \sigma(K*(K*X_1)) \tag{3-2}$$

式中：σ 代表 sigmoid 函数，$\mathcal{F}(X)$表示卷积核函数，*为卷积操作，K 是滤波器，大小为 7×7。

最后输出特征图 X_3 为

$$X_3 = X_1 \otimes X_2 \tag{3-3}$$

由图 3-3 可知，ICSAM 被插入到 ResNet18 残差块中的具体位置。

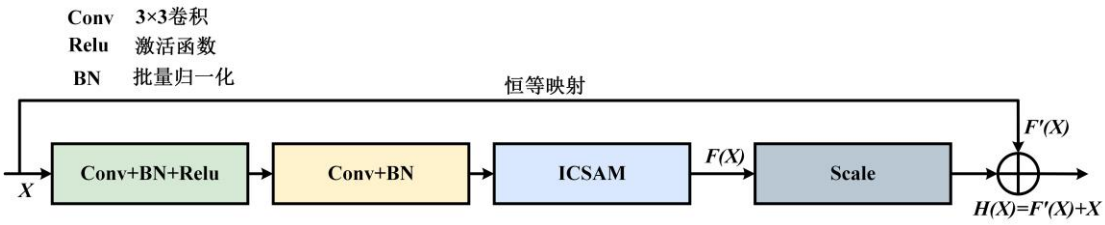
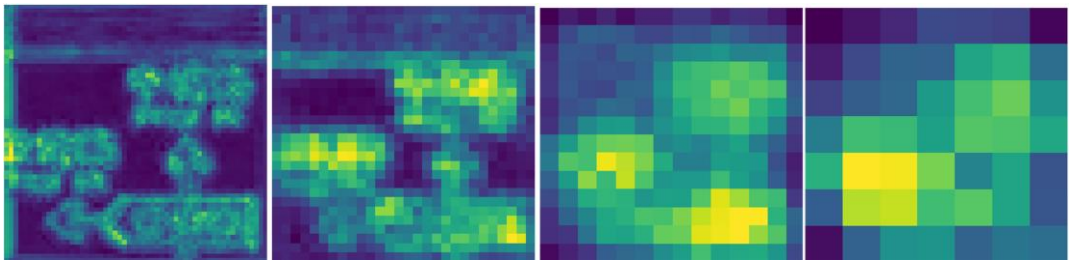


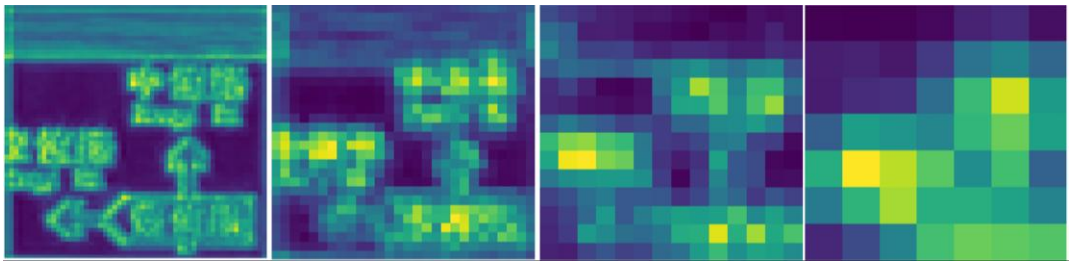
图 3-3 残差块中 ICSAM 位置

图 3-4 为 ResNet18 和 IResNet18 两种骨干网络提取特征图的可视化，其中图 3-4(a)为 ResNet18 提取结果，图 3-4(b)为 IResNet18 提取结果。第一列和第四

列对应图 3-1 中 *C*2、*C*3、*C*4 和 *C*5 四个特征图。



(a) ResNet18 提取结果



(b) IResNet18 提取结果

图 3-4 ResNet18 和 IResNet18 特征提取可视化对比

由图 3-4 可知，IResNet18 提取的特征图更加清晰，表示其能更好地保留图像中的关键细节和特征，减少多尺度处理过程中重要信息的丢失。

3.2.2 双向特征金字塔

双向特征金字塔网络（BiFPN[41]）通过双向跨尺度连接和加权特征融合来聚合不同分辨率的特征图。图 3-5 是 BiFPN 模块具体的网络结构。

双向跨尺度连接方面，该方法旨在优化网络结构，因此移除缺乏特征融合能力且对整体特征网络贡献有限的单条输入边节点。同时，对于处于同一层级的原始输入和输出节点，增加了直接连接，这种连接在不增加计算负担的情况下进一步增强了特征融合能力。此外，BiFPN 摒弃了单一的自顶向下或自底向上路径，而是采用多次重复的双向路径结构，将每个包含自顶向下和自底向上过程的完整循环视为一层特征网络，以此促进更深层次的特征整合。

加权特征融合方面，该方法在融合不同分辨率的特征图时，不再简单相加，而是为每个输入特征分配可学习权重，使网络能够自适应地强调对当前任务更为重要的特征，同时抑制无用的信息。通过这种方式，BiFPN 能够更高效地利用不同层级的特征图，优化特征融合过程，从而提升目标检测的性能。

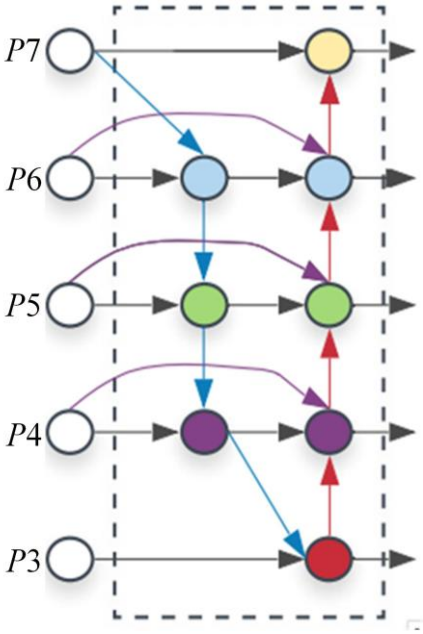


图 3-5 BiFPN 网络结构

BPI-DBNet 交通标志牌文本检测方法采用 BiFPN 的特征融合机制。相比 DBNet 使用的 FPN，BiFPN 具备更强的双向特征融合能力，使高层语义信息与低层细节信息能够双向流动，从而提升特征表达能力。进一步引入 HWD 模块、SC-SPPF 模块以及 ECA 模块对 BiFPN 进行优化，以提升其特征表达能力和文本上下文关系。

(1) Haar 小波下采样

Haar 小波下采样模块（HWD）主要由无损特征编码模块与特征学习模块两部分组成。在无损特征编码阶段，通过引入 Haar 小波变换（HWT）对输入特征图进行处理，在降低特征图空间分辨率的同时，将部分空间信息转化编码至通道维度中，实现空间到通道的信息重构。这种方式能够有效减少信息丢失，从而保持原始特征图中的关键信息。

特征学习模块利用标准卷积层（Conv）、批量归一化（BN）及 ReLU 激活函数对编码后的特征进行进一步处理，既实现了通道数的适配调整，也强化了与分割任务相关的显著特征表达。HWD 不仅显著降低了模型在多尺度特征提取过程中的计算成本，还在保留图像结构信息的同时提升了网络对文本区域的感知能力。

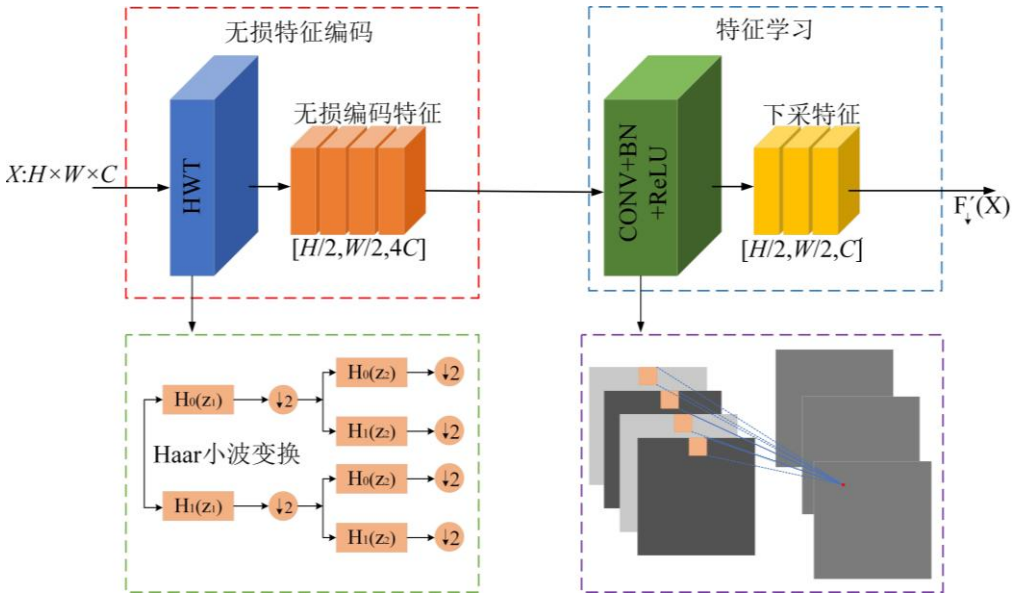


图 3-6 HWD 网络结构

HWD 模块的核心是 Haar 小波变换，一级一维 Haar 小波变换函数和尺度函数定义为

$$\begin{cases} \phi_1(X)=\dfrac{1}{\sqrt{2}}\phi_{1,0}(X)+\dfrac{1}{\sqrt{2}}\phi_{1,1}(X) \\ \psi_1(X)=\dfrac{1}{\sqrt{2}}\phi_{1,0}(X)-\dfrac{1}{\sqrt{2}}\phi_{1,1}(X). \end{cases} \tag{3-4}$$

$\phi_{j,k}(X)$ 的定义为

$$\phi_{j,k}(X)=\sqrt{2^j}\phi(2^jX-k),k=0,1,\cdots,2^j-1 \tag{3-5}$$

式中：参数 j 和 k 分别表示为 Haar 函数的阶数（或图像处理域中的尺度）和顺序（或 2D 图像的方向）。

$\phi_{0,0}(X)$ 定义为

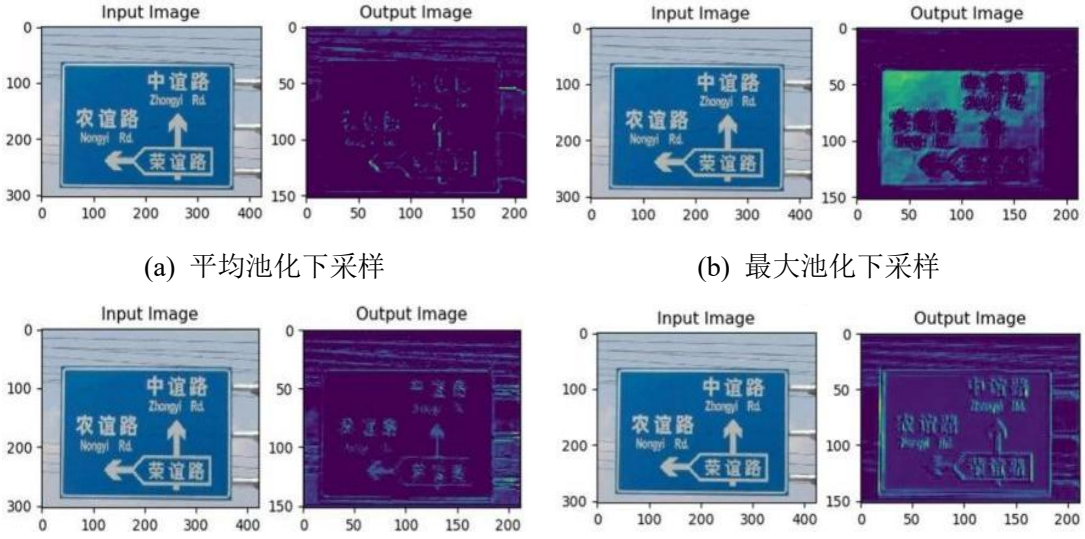
$$\phi_{0,0}(X)=\phi_0(X)=\begin{cases} 0, & X<0 \\ 1, & 0\leqslant X<1 \\ 0, & X\geqslant 1 \end{cases} \tag{3-6}$$

因此，1 阶段 Haar 变换可以使用 0 阶段 Haar 基函数来表示，转换公式为

$$\begin{cases} \phi_1(x)=\phi_0(2x)+\phi_0(2x-1) \\ \psi_1(x)=\phi_0(2x)-\phi_0(2x-1) \end{cases} \tag{3-7}$$

这意味着长度为 L 的信号可以分为长度为 $L/2$ 的两个部分，两个部分可以解

释为高通和低通解调滤波器，当 Haar 小波变换应用在二维信号时，产生的四个分量，每一个分量是原始信号的一半空间分辨率。



(a) 平均池化下采样　　(b) 最大池化下采样

(c) 步幅卷积下采样　　(d) HWD 下采样

图 3-7 平均池化，最大池化，步幅卷积和 HWD 的下采样特征图

图 3-7 中表示平均池化，最大池化，步幅卷积和 HWD 四个下采样后所产生的特征图。最大池化、平均池化和步幅卷积是常用于下采样操作的方法，用来聚合局部特征，扩大感受野，减少计算开销，但其在分割算法中会导致重要的空间信息丢失。由图 3-7 可知，HWD 相对于最大池化、平均池化和步幅卷积下采样，在信息的丢失方面有所减少。

(2) 自校准空间金字塔模块

由图3-8可知，自校准空间金字塔模块（SC-SPPF）由空间金字塔模块（SPPF）和自校准模块（SC）组成。在 SPPF 模块中，先输入特征图 $X = R^{C\times H\times W}$，经过卷积层处理，再使用三个尺寸相同的 5×5 池化核池化，并通过跳跃连接将池化结果与原始特征图相加。

池化操作有助于捕捉文本在不同尺度上的高层次特征信息。将提取的特征与跳连接分支进行拼接，融合局部与全局高层次特征，增强模块对文本几何尺寸的表达能力，从而更好处理不同大小和形状的文本。

SC 模块采用不同尺度的卷积操作，结合非线性融合方式，有效整合多尺度的空间上下文信息，从而增强了各空间位置对文本上下文关系的建模能力。

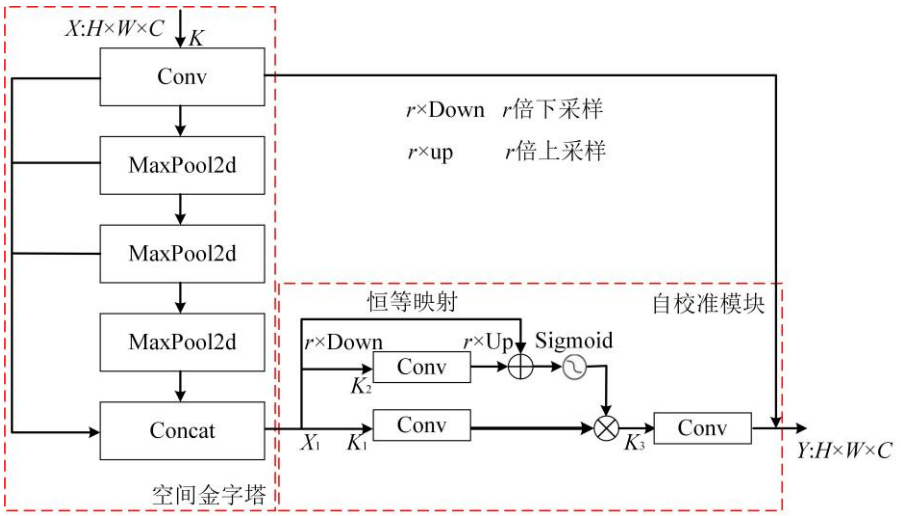


图 3-8 SC-SPPF 网络结构

将输入尺寸为 $X{:}C\times H\times W$ 的原始图像，通过空间金字塔得到特征图为

$$X_1=\mathrm{Concat}(Y_1(X),Y_2(Y_1(X)),Y_2(Y_2(Y_1(X))),Y_2(Y_2(Y_2(Y_1(X)))),1) \tag{3-8}$$

式中：$Y_1(X)=\mathcal{F}(X)=K*X$，其中 K 为滤波器，大小为 1×1，$Y_2(X)=\mathrm{MaxPool}_5(X)$，其中 5 表示池化核大小为 5×5。

进入自校准模块，利用滤波器大小 4×4 和步长为 4 的平均池化进行下采样，具体公式如式(3-9)所示。

$$T=\mathrm{AvgPool}_4(X_1) \tag{3-9}$$

式中：4 是池化过程下下采样频率和步长，T 是 X_1 进行下采样后的特征。

对于 T 进行卷积和上采样操作。

$$X_1'=\mathrm{Up}(\mathcal{F}_2(T))=\mathrm{Up}(T*K_2) \tag{3-10}$$

式中：UP 代表上采样操作，是一个双线性插值运算符，将小比例空间尺寸映射到原始空间尺寸。

校准的卷积操作公式为

$$Y'=\mathcal{F}_1(X_1)\otimes\sigma(X_1+X_1') \tag{3-11}$$

$$Y=Y_1\left(\mathcal{F}_3(Y')\right) \tag{3-12}$$

式中：$\mathcal{F}_1(X)=X_1*K_1$，$\mathcal{F}_3(Y')=Y'*K_3$，K_i 为滤波器，大小都为 3×3。

为提升文本检测中特征表达的准确性与鲁棒性，提出了 SC-SPPF 模块。在

原有 SPPF 模块的基础上，虽然其通过多尺度池化增强了高层特征对文本几何尺寸的表达能力，但仍存在忽略不同空间位置间上下文关系的问题。为此，SC 模块引入了上下文建模机制，有效弥补了该不足。整合两者优势后，SC-SPPF 模块不仅进一步强化了几何信息的表达能力，还实现了空间位置间上下文关系的有效建模，从而显著提升了整体特征表示的准确性与鲁棒性。

(3) 高效多尺度注意力模块

高效多尺度注意力模块（EMA）旨在提升特征图在通道与空间维度上的上下文建模能力。其结构如图 3-9 所示，输入特征图 $X = R^{C\times H\times W}$，将特征图按通道分成 G 组，每组捕捉不同语义信息，通道分组表示为 X=[X_0, X_1, …, $X_{G\text{-}1}$]，$X_i = R^{C//G}$。通过三个并行子网络，得到分组特征图。在 1×1 分支中，EMA 沿着 X、Y 方向对通道编码，利用两个一维全局平均池化捕捉通道间依赖。该分支的两个通道特征图，经乘法操作合并，再通过全局平均池化对上层输入编码，得到第一个输出 $R_1^{1\times C//G}\times R_3^{C//G\times H\times W}$。经二维全局平均池化和矩阵点积运算，合并并行处理结果，形成第一个空间特征图。二维全局平均池化公式为

$$Z_c = \frac{1}{H\times W}\sum_{j}^{H}\sum_{i}^{W} X_c(i,j) \tag{3-13}$$

式中：H 和 W 分别为输入特征图的高度和宽度，$X_c(i,j)$ 为特征图第∂通道上的像素值。

在 3×3 分支中，通过 3×3 卷积核提取多尺度特征，采用与 1×1 分支相似的编码方法，经过系列调整后得到第二个输出 $R_3^{1\times C//G}\times R_1^{C//G\times H\times W}$，进而得到第二个空间特征图。将第一和第二空间特征图结合，通过 Sigmoid 激活函数处理，有效捕捉像素间的配对关系，增强特征图的表达能力与鲁棒性。整体来看，EMA 通过多分支协同机制与跨空间注意力机制，有效建模多尺度上下文信息，显著增强了模型在复杂场景下的感知能力。

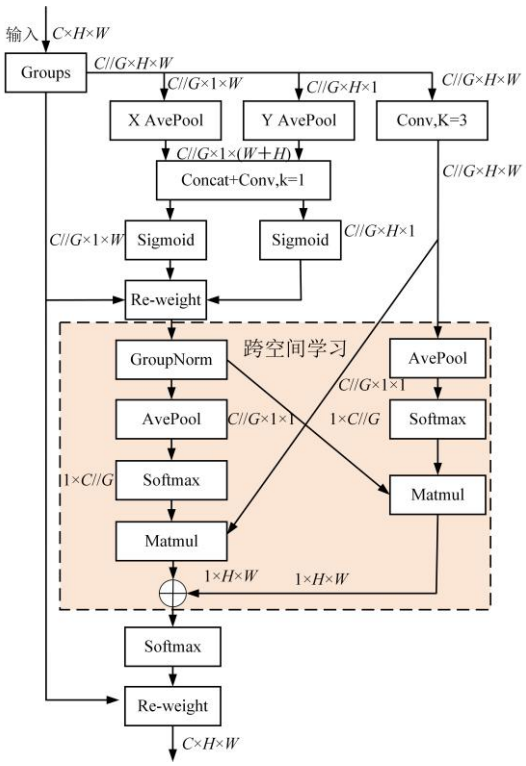


图 3-9 EMA 注意力网络结构

3.2.3 后处理模块

(1) 可微分二值化

基于分割的文本检测中，采用标准二值化函数对概率图进行处理，获取二值图，标准二值化函数为

$$B_{i,j}=\begin{cases}1 & (P_{i,j}\geqslant t)\\ 0 & (\text{otherwise})\end{cases} \tag{3-14}$$

式中：$P_{i,j}$ 是概率图中像素点(i, j)的值，$B_{i,j}$ 是阀值图中像素点(i, j)的值，t 为预设阈值。文字区域为 1，非文字区域为 0。

由式(3-14)可知，标准二值化不可微分，其不能在网络中进行优化学习。为了解决二值化不可微分的问题，使用近似阶跃函数来代替标准二值化。近似阶跃函数公式为

$$\hat{B}_{i,j}=\frac{1}{1+e^{-k(P_{i,j}-T_{i,j})}} \tag{3-15}$$

式中：$\hat{B}_{i,j}$ 是近似二值图中像素点(i, j)的值，$T_{i,j}$ 是网络学习阈值图像素点(i, j)的值，k 是放大因子，k=50。

(2) 损失函数

损失函数（Loss）值由三个部分组成，整体公式为

$$L = L_s + \alpha \times L_b + \beta \times L_t \tag{3-16}$$

式中：L_s 是概率图损失，L_b 是二值图损失，L_t 是阈值图损失，α 和 β 分别取 1 和 10。

概率图损失和二值图损失为

$$L_s = L_b = \sum_{i \in S_i} y_i \log x_i + (1 - y_i) \log(1 - x_i) \tag{3-17}$$

式中：S_i 是采样数据集，其正样本和负样本比值为 1:3。

阈值图损失为

$$L_t = \sum_{i \in R_d} |y_i^* - x_i^*| \tag{3-18}$$

式中：R_d 是原始文本框和扩张后文本框之间的区域，y_i^* 是阈值图标签，x_i^* 是预测的阀值图。

3.3 实验结果和分析

3.3.1 实验参数设置

实验环境如表 3-1 所示。网络训练过程中采用 Adam 优化器，初始学习率为 0.001，模型在数据集 CTST-1600 上训练周期是 400 epoch，而在 ICDAR2015 和 MLT-2017 的训练周期，参考 DBNet[24]的设置是 1200 epoch，批次大小都为 16。

表 3-1 实验环境参数

配置	参数
操作系统	Ubuntu20.04
处理器	Xeon(R)
内存	25G
显存	V100 32G 显存
框架	PyTorch 2.0.0
Cuda	11.8
编程软件	Pycharm

为提升模型对不同场景和文本形态的鲁棒性，在训练阶段对原始图像进行数据增强处理，将数据集中的图片在(-10°, 10°)角度范围内随机旋转，以增强模型对文本倾斜角度变化的适应能力，同时执行随机裁剪与水平翻转以丰富图像的空间分布特征，防止模型对特定位置或排列方式产生依赖。为保证模型输入维度一致并兼顾计算效率，将图像调整为 640×640 像素大小。

3.3.2 消融实验

为了验证 ICSAM，HWD，SC-SPPF 和 EMA 的有效性，在中国交通标志牌文本检测数据集（CTST-1600）上进行大量的消融实验，结果见表 3-2。BPI-DBNet 主要在特征提取的骨干网络和特征融合的颈部网络进行改进。实验中，ResNet18 为骨干网络，BiFPN 为颈部网络，以此作为可微分二值化方法的复现结果和实验基准线。其中，IResNet18 表示经过 ICSAM 优化的 ResNet18，IBiFPN 表示经过 HWD、SC-SPPF 和 EMA 优化的 BiFPN。ResNet+BiFPN 表示以 ResNet 为骨干网络进行特征提取，BiFPN 为颈部网络进行特征融合。以此类推 IResNet+BiFPN 等也是如此。

表 3-2 消融实验结果

骨干网络	HWD	SC-SPPF	EMA	P(%)	R(%)	F(%)
ResNet18	O	O	O	93.80	90.48	92.11
IResNet18	O	O	O	93.58	91.15	92.34
IResNet18	P	O	O	93.48	92.06	92.77
IResNet18	O	P	O	93.04	92.99	93.01
IResNet18	O	O	P	93.56	92.14	92.84
IResNet18	P	P	O	94.89	91.81	93.32
IResNet18	P	O	P	93.30	92.75	93.02
IResNet18	O	P	P	93.88	93.43	93.65
IResNet18	P	P	P	94.59	93.50	94.04

(1) IResNet+BiFPN

表 3-2 中，“O”表示 BiFPN 未加入相应模块，“P”则相反。由表 3-2 可知，用 IResNet 作为骨干网络后，R 值和 F 值分别提高了 0.67%和 0.23%。原因在于，IResNet 进行特征提取时，减少了骨干网络在多尺度特征提取过程中的信

息损失，因特征图分辨率高，从而获得更具有判别力的特征。图 3-10 中第一行是模型检测的结果，第二行表示相应的二值图。由图 3-10 可知，ResNet18 作为骨干网络进行特征提取时，会造成交通标志牌文本的漏检，使用 IResNet 作为骨干网络后，减少了因交通标志牌文本行大小和比例差异引起的漏检。

(a) ResNet+BiFPN (b) IResNet+BiFPN (c) ResNet+BiFPN (d) IResNet+BiFPN

图 3-10 ResNet18+BiFPN 和 IResNet18+BiFPN 检测效果对比

(2) IResNet+BiFPN(HWD)

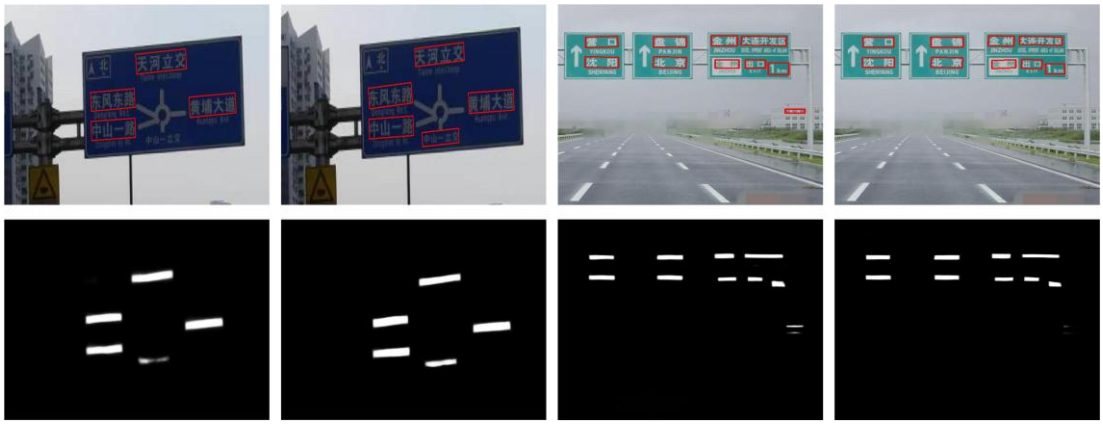
由表 3-2 可知，在保持骨干网络 IResNet 不变的前提下，用 HWD 优化 BiFPN 后的模型。在 R 值和 F 值上分别提高了 0.91%和 0.43%。原因在于，HWD 替换 BIFPN 下采样时，会把特征图分解为不同频率的成分，以此来保留特征图的关键信息和更多的空间信息。由图 3-11 可知，该优化不仅提高了不同背景下的交通标志牌文本检测精度，同时减少了因交通标志牌文本行大小和比例差异、复杂自然场景导致的文本漏检和错检。

(a) IResNet+BiFPN (b) HWD 优化 (c) IResNet+BiFPN (d) HWD 优化

图 3-11 IResNet18+BiFPN 和 IResNet18+BiFPN(HWD)检测效果对比

(3) IResNet+BiFPN(SC-SPPF)

由表 3-2 可知，在保持骨干网络 IResNet 不变的前提下，用 SC-SPPF 优化 BiFPN 后的模型，在 R 值和 F 值分别提高了 1.84%和 0.67%。原因在于，SC-SPPF 增强高层次特征对文本的几何大小的表达能力和上下文关系，增强文本中各元素的关联性和交互性。由图 3-12 可知，该优化不仅减少因交通标志牌文本行大小和比例差异引起的漏检，还减少了自然环境中雾天引起的错检。



(a) IResNet+BiFPN　(b) SC-SPPF 优化　(c) IResNet+BiFPN　(d) SC-SPPF 优化

图 3-12 IResNet18+BiFPN 和 IResNet+BiFPN(SC-SPPF)检测效果对比

(4) IResNet+BiFPN(EMA)

由表 3-2 可知，在保持骨干网络 IResNet 不变的前提下，用 EMA 优化 BiFPN 后的模型，在 R 值和 F 值分别提高了 0.99%和 0.5%。原因在于，EMA 善于捕捉局部细节与全局上下文，从而增强特征丰富度和区分度，由图 3-13 可知，该优化减少了因交通标志牌文本字符稀疏而导致的交叉错检，以及因语言种类不同而引起的错检。

(a) IResNet+BiFPN　(b) EMA 优化　(c) IResNet+BiFPN　(d) EMA 优化

图 3-13 IResNet18+BiFPN 和 IResNet18+IBiFPN(EMA)检测效果对比

(5) 组合实验

此外，还随机组合了 HWD，SC-SPPF 和 EMA，并测试每种组合的检测性能（如表 3-2 的第七行至第十行所示）。实验结果表明，所提模块有效的提高交通标志牌文本检测性能。

IResNet、HWD、SC-SPPF 以及 EMA 优化后的交通标志牌文本检测模型 BPI-DBNet 与基线模型相比，在 CTST-1600 数据集上的 P 值、R 值和 F 值分别提高了 0.79%、3.02%和 1.93%。

由表 3-3 可知，在 CTST-1600 数据集上，BPI-DBNet 相较于 DBNet，P 值、R 值和 F 值分别提高了 2.88%、1.21%和 2.04%。同时，表 3-2 第一行数据和 DBNet 比较可知，BiFPN 替换 FPN，R 值和 F 值分别提高了 2.09%和 0.11%。实验结果表明，基于双向特征金字塔的交通标志牌文本检测方法优于 DBNet，在交通标志牌文本检测任务中具有更好的适用性和性能优势。

3.3.3 对比实验分析

为了验证 BPI-DBNet 的适用性和鲁棒性，在数据集 CTST-1600、ICDAR2015 以及 MLT-2017 上与多种文本检测模型进行比较。由表 3-3 可知，在 CTST-1600 数据集上，BPI-DBNet 在 P 值、R 值和 F 值上达到 94.59%、93.50%和 94.04%。相比于其它的交通标志牌文本检测模型，BPI-DBNets 在检测性能指标 F 上分别提高了 18.04%、35.04%、13.04%、10.04%、2.04%、2.95%、2.25%、0.34%。从 F 值来看，BPI-DBNet 是表 3-3 模型中检测性能优异的交通标志牌文本检测方法。

表 3-3 在 CTST-1600 数据集上各方法比较

模型	P(%)	R(%)	F(%)
EAST[42]	83.00	72.00	76.00
CTPN[15]	71.00	62.00	59.00
CTST[43]	86.00	79.00	81.00
HTSTL[25]	88.90	79.40	84.00
DBNet(ResNet18)[24]	91.71	92.29	92.00
DBNet++(ResNet50)[44]	93.24	89.95	91.09
CC-DBNet[45]	93.70	89.95	91.79
MFENet[31]	93.60	93.80	93.70
BPI-DBNet	94.59	93.50	94.04

由表 3-4 可知，在 ICDAR2015 数据集上，BPI-DBNet 在 P 值、R 值和 F 值上达到 92.08%、78.99%和 85.03%，与 DBNet 相比，分别提高了 5.28%、0.59%和 2.73%。从 F 值来看，BPI-DBNet 是表 3-4 模型中检测性能优异的文本检测方法。

表 3-4 在 ICDAR2015 数据集上各方法比较

模型	P(%)	R(%)	F(%)
CTPN[15]	74.20	51.60	60.90
MCN[46]	72.00	80.00	76.00
SSTD[47]	80.20	73.90	76.90
WordSup[48]	79.30	77.00	78.20
Textboxes++[17]	87.20	76.70	81.70
DB(ResNet18)[24]	86.80	78.40	82.30
PANet[23]	84.00	81.90	82.90
Textsnake[49]	84.90	80.40	82.60
DBNet++(ResNet18)[44]	90.10	77.20	83.10
SAE[50]	85.10	84.50	84.80
文献[51]	88.40	79.50	81.90
BPI-DBNet	92.08	78.99	85.03

由表 3-5 可知，在 MLT-2017 数据集上，BPI-DBNet 在 P 值、R 值和 F 值上达到 85.88%，65.45%和 74.29%，与 DBNet 相比，分别提高了 3.98%、1.65%和 2.59%。从 F 值来看，BPI-DBNet 是表 3-5 模型中检测性能优异的文本检测方法。

表 3-5 在 MLT-2017 数据集上各方法比较

模型	P(%)	R(%)	F(%)
SARI_FDU_RRPN_V1[52]	71.20	55.50	62.40
Sensetime_OCR[52]	56.90	69.40	62.60
SCUT_DLVlab1[52]	80.30	54.50	65.00
Corner[53]	83.80	55.60	66.80
LOMO[54]	78.80	60.60	68.50
e2e_ctc01_multi_scale[52]	79.80	61.20	69.30
SPCNet[55]	73.40	66.90	70.00
PSENet[22]	73.80	68.20	70.90
DBNet(ResNet18)[24]	81.90	63.80	71.70
BPI-DBNet	85.88	65.45	74.29

由表 3-3 至表 3-5 可知，BPI-DBNet 不仅在中国交通标志牌文本数据集（CTST-1600）上检测性能表现卓越，同时在 ICDAR2015，MLT-2017 公共数据集上也表现出优异的检测性能，体现了 BPI-DBNet 模型较强的适用性和良好的鲁棒性。

由第二章可知，轻量化的网络可以减少模型的大小，利于车辆的部署，表 3-6 中利用轻量化网络对 DBNet 和 BPI-DBNet 进行优化。

表 3-6 轻量化网络效果对比

模型	P(%)	R(%)	F(%)	Params(M)	FPS
DBNet(ResNet18+FPN)	91.71	92.29	92.00	187.00	37.57
DBNet(MobilenetV3[56]+FPN)	92.89	83.08	87.71	28.60	30.12
DBNet(Shufflenetv2[57]+FPN)	92.43	85.61	88.89	19.20	32.14
BPI-DBNet(IResNet18+IBiFPN)	94.59	93.50	94.04	457.00	21.44
BPI-DBNet(MobilenetV3[56]+IBiFPN)	95.77	84.29	89.75	68.64	17.18
BPI-DBNet(Shufflenetv2[57]+IBiFPN)	95.31	86.82	90.93	46.08	18.34

由表 3-6 可知，轻量化网络 MobilenetV3，Shufflenetv2 优化的 DBNet 和 BPI-DBNet，在减小模型大小的同时，检测性能却显著减少，这与交通标志牌文本检测对高检测性能的要求不符。表 3-6 间接证明基于分割方法需要更多的计算资源。BPI-DBNet 在交通标志牌文本检测上有较高的检测性能，但模型的大小 Params 为 457M，模型过大不利于车辆部署。

3.4 本章小结

本章提出了一种基于双向特征金字塔的交通标志牌文本检测方法 BPI-DBNet。该方法在 DBNet 模型基础上进行优化，在骨干网络 ResNet18 中引入 ICSAM，同时在颈部网络 BiFPN 中引入 HWD、SC-SPPF 和 EMA 模块。优化后的 BPI-DBNet 模型在 CTST-1600、ICDAR-2015 以及 MLT-2017 三个公开数据集上进行了实验验证，F 值分别为 94.04%、85.03%和 74.29%，整体检测性能优于现有主流文本检测模型。然而，BPI-DBNet 模型结构较为复杂，计算资源消耗较大，难以直接部署于资源受限的无人驾驶车载平台，亟需进一步开展模型轻量化设计以满足实际应用需求。

第4章 基于旋转框的交通标志牌文本检测

4.1 引言

在无人驾驶汽车车载平台计算资源有限的背景下，虽 BPI-DBNet 在交通标志牌文本检测任务中表现出较高的检测精度，但其模型结构复杂、参数量较大，对计算资源的依赖较强，限制了其在资源受限场景中的实际部署效果。针对交通标志牌文本的特性以及实际应用场景的需求，进一步优化检测方法显得尤为重要。

交通标志牌具有显著的规范性，其中文字排列工整、字体统一，并且通常采用高对比度的背景色进行突出显示，这些特征相较于城市中店铺广告中多样化、弯曲或变形的文本样式，更利于自动检测与识别。这种规则化的文本形态在一定程度上降低了检测难度。然而，在实际拍摄过程中，交通标志牌可能因拍摄角度、视角偏移等因素导致文本发生倾斜，由图 4-1(a)可知，传统水平框难以精确框选文本区域。相比之下，由图 4-1(b)可知，旋转框能够更准确地拟合文本的倾斜角度，从而提高定位精度。旋转框属于基于回归的文本检测方法，该类方法计算开销相对较低，尤其适合部署在计算资源有限的车载系统中。

(a) 水平框文本定位

(b) 旋转框文本定位

图 4-1 水平框和旋转框文本选定对比

YOLOV8-OBB 是为检测倾斜目标而设计的方法。但由于交通标志牌文本语言多样性、自然场景复杂以及交通标志牌文本行大小和比例差异、字符稀疏度变化，导致交通标志牌文本漏检和错检，使其检测性能较低。因此，需要在 YOLOV8-OBB 的基础上改进，以提高交通标志牌文本检测性能。

4.2 交通标志牌文本检测模型

YOLOV8[37]是 Ultralytics 公司在 YOLOV5[58]基础上优化和改进的模型，主要由骨干网络、颈部网络和检测头三个部分组成。YOLOV8-OBB 是 YOLOV8 的拓展版本，专为旋转框体检测场景设计，在 YOLOV8 的基础上对检测头进行修改，解耦头中新增了角度预测输出头，并将原来的 CIoU 损失函数替换为 ProbIOU[59]损失函数，以实现旋转框的预测。

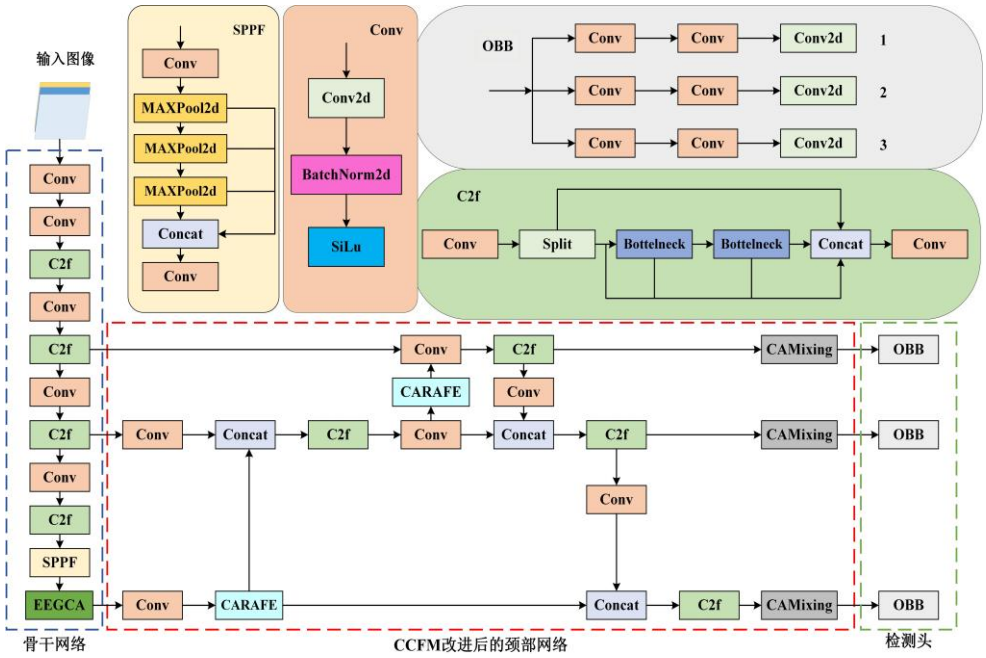


图 4-2 IYOLOV8-OBB 交通标志牌文本检测模型

基于 YOLOV8-OBB 文本检测模型，提出 IYOLOV8-OBB（Improved YOLOV8-OBB）的交通标志牌文本检测模型。由图 4-2 可知，引入 EEGCA 注意力机制对主干网络进行优化，引入 CCFM 模块来改进 YOLOV8-OBB 的颈部网络。将上采样模块替换为 CARAFE，在三个检测头前加入 CAMixing 模块。为了模型的轻量化，将模型的尺寸和复杂度设定为“n”。

4.2.1 高效全局通道注意力模块

高效全局通道注意力模块（EEGCA）是在 ECA[60]注意力机制基础上进行改进，ECA 通过对不同通道的特征进行加权，使网络能够更有效地捕捉长距离依赖关系。然而，ECA 在保留全局视角信息和表达不同尺度特征方面存在一定不足。为克服这一弱点，将对输入层分别进行平均池化（AvgPool）和最大池化（MaxPool），再相加，从而更全面地整合全局信息并增强特征表达能力。

由图 4-3 可知，先对输入特征图进行平均池化和最大池化，再相加。将特征图从[*H*, *W*, *C*]的矩阵转换为[1, 1, *C*]的向量，根据特征图的通道数 *C* 计算自适应的一维卷积核大小，再利用一维卷积核进行一维卷积，以获得特征图中每个通道的权重。将归一化后的权重与原始输入特征图逐通道相乘，生成加权后的特征图。优化后的 EEGCA 注意力模块不仅增强了对全局视角信息及不同尺度大小特征的捕捉能力，还具有出色的跨通道信息捕捉能力。

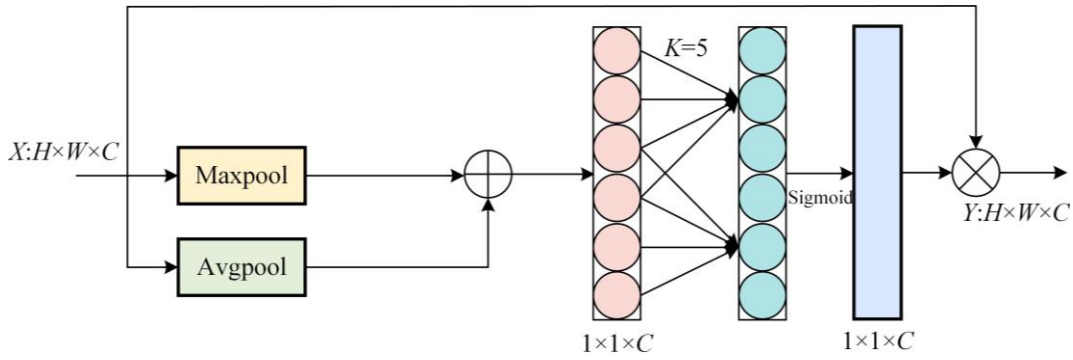


图 4-3 EEGCA 网络结构

4.2.2 多尺度特征融合

为了提升多尺度特征融合能力、扩大感受野，并整合全局与局部特征。引入 CCFM、CARAFE 以及 CAMixing 对 YOLOV8-OBB 中用来多尺度特征融合的颈部网络进行优化，优化位置如图 4-2 可知。

(1) 跨尺度融合模块

跨尺度融合模块（CCFM）由 RT-DETR[61]提出的，用于优化特征融合路径。该模块在融合路径中插入多个由卷积层组成的融合块，融合块的作用是将相邻的两个尺度特征融合成一个新的特征。

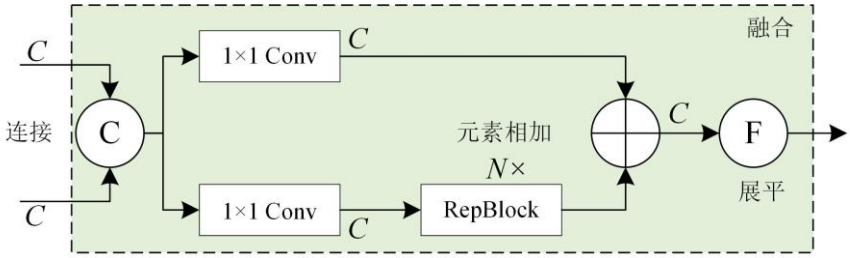


图 4-4 CCFF 融合模块

图 4-4 中，C 表示输入通道。融合模块包含两个 1×1 卷积以调整通道数，并包含 N 个由 RepConv[62]组成的 RepBlock 进行特征融合。通过元素相加的方式融合两路输出。将跨尺度融合模块（CCFM）结合到 YOLOV8-OBB 网络结构中，整合不同尺度的特征，增强模型对尺度变化的适应性以及对小尺度目标的检测能力。同时，该模块能有效融合 YOLOV8-OBB 的细节特征和上下文信息，从而提升整体性能。

(2) 内容感知功能重组

内容感知功能重组模块（CARAFE）通过引入更大的感受野，有效聚合上下文和特征信息，并结合输入特征图的语义信息，增强了对重要内容的感知。该模块计算量较低，保持了模型的轻量级特性，同时通过保留图像细节信息，提高了模型的文本检测性能。

由图 4-5 可知，CARAFE 上采样主要分为上采样预测模块和特征重组模块，在上采样预测模块中，将输入的 $H\times W\times C$ 特征图的通道数压缩到 C_m，得到 $C_m\times H\times W$ 特征图，以减少后续计算量。通过大小为 $k_{up}\times k_{up}$ 的卷积核对内容编码，生成重组核，得到尺寸为 $\sigma^2\times k_{up}^2$ 的特征图。如式(4-1)所示。

$$C_m=\sigma^2\times k_{up}^2 \tag{4-1}$$

式中：C_m 表示降维后的特征层通道数，σ 为上采样的倍数（σ=2），k_{up} 为预测上采样核的大小。将通道在空间维度上展开，重新排列组合，得到尺寸为 $\sigma_H\times\sigma_W\times k_{up}^2$ 的上采样核。进行 Softmax 归一化，使其权重之和为 1。

特征重组模块中，利用上采样核进行特征重组，提取目标特征。对于输出特征图的每一个目标位置，以目标为中心取一个大小为 $k_{up}\times k_{up}$ 的区域，与该点位置预测得到的上采样核做点积，将其映射给输入特征图，得到大小为 $\sigma_H\times\sigma_W\times C$ 的特征图。

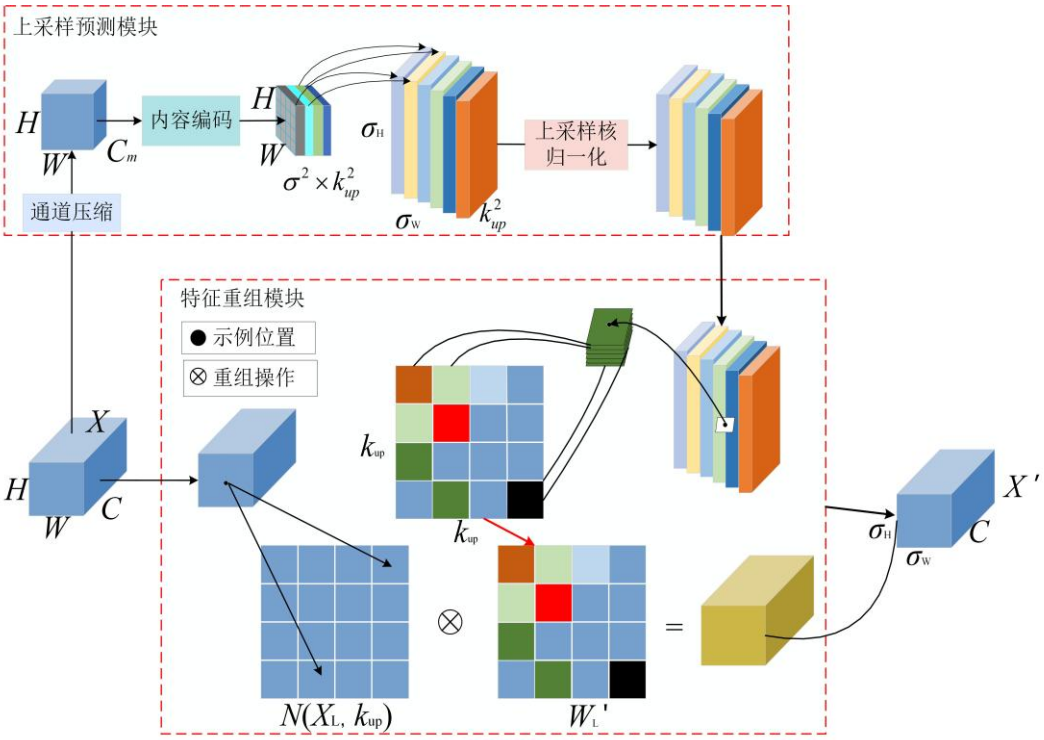


图 4-5 CARAFE 网络结构

CARAFE 上采样结构能根据不同特征的分布情况，生成不同的上采样核，从而关注特征在全局特征图中的分布。将 CARAFE 上采样应用于 YOLOv8-OBB 模型中，替代原本的最近邻插值上采样模块。该优化在不增加额外参数和计算量的情况下，提高重要特征的检测能力。

(3) 卷积注意力混合模块

由图 4-6 可知，卷积注意力混合模块（CAMixing）包含两个主要部分：卷积注意力融合模块（CAFM）和多尺度前馈网络（MSFN）。

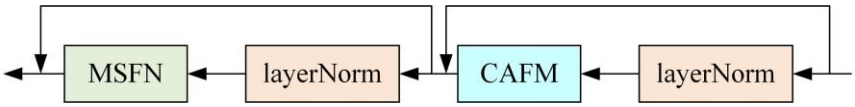


图 4-6 CAMixing 网络结构

在卷积注意力融合模块中卷积运算因局部有限的感知领域，难以有效模拟全局特征。而 Transformer[63]在注意力机制的支撑下，在提取全局特征和捕获远程依赖方面表现出色。卷积和注意力机制相辅相成，能对全局和局部特征进行建模。

由图 4-7 可知，卷积注意力融合模块（CAFM）由局部和全局分支组成，分别用于提取局部和全局特征。其中，局部分支利用卷积和通道洗牌变换提取局部特征，而全局分支通过注意力机制用于建模远程特征依赖。

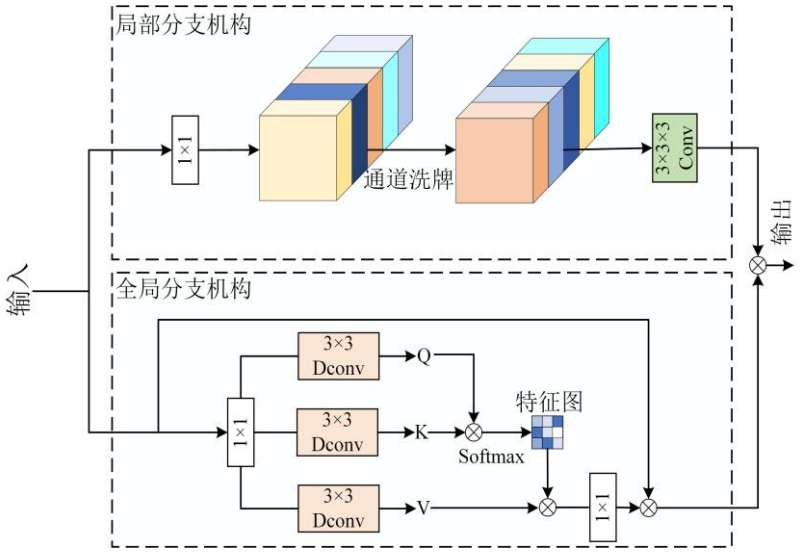


图 4-7 CAFM 网络结构

在局部分支中，使用 1×1 卷积调整通道维度后，执行通道洗牌，进一步混合通道信息。通道洗牌过程将输入张量沿通道维度划分为若干组，每组内部采用深度可分卷积进行特征提取。并通过通道重组来增强信息交互。将所有组的输出张量沿通道维度拼接，形成新的输出张量。利用 3×3×3 卷积进一步提取特征。局部分支表述为

$$F_{\text{conv}} = \mathcal{F}_{3\times3\times3}(\text{CS}(\mathcal{F}_{1\times1}(Y))) \tag{4-2}$$

式中：F_{conv} 是局部分支的输出，$\mathcal{F}_{1\times1}(X)$ 表示 1×1 卷积，$\mathcal{F}_{3\times3\times3}(X)$ 表示 3×3×3 卷积，CS 表示通道洗牌操作，Y 是输入特征。

在全局分支中，通过 1×1 卷积和 3×3 深度卷积生成的查询 Q，关键字 K 和值 V，得到形状为 $\hat{H}\times\hat{W}\times\hat{C}$ 的三个张量。Q 被重塑为 $\hat{Q}\in R^{\hat{H}\hat{W}\times\hat{C}}$，$K$ 被重塑为 $\hat{K}\in R^{\hat{C}\times\hat{H}\hat{W}}$。通过 $\hat{Q}$ 和 $\hat{K}$ 交互来计算特征图 $\text{A}\in R^{\hat{C}\times\hat{C}}$，为减少计算负担，而不是计算大小为 $R^{\hat{H}\hat{W}\times\hat{H}\hat{W}}$ 巨大的注意力图。全局分支的输出 F_{att} 定义为

$$F_{\text{att}} = \mathcal{F}_{1\times1}\,\text{Attention}\left(\hat{Q},\hat{K},\hat{V}\right)+Y \tag{4-3}$$

$$\text{Attention}\left(\hat{Q},\hat{K},\hat{V}\right)=\hat{V}\,\text{Softmax}\left(\hat{K}\hat{Q}/\alpha\right) \tag{4-4}$$

式中：α为学习的比例参数，应用在 Softmax 函数之前，控制 $\hat{Q}$ 和 $\hat{K}$ 的矩阵乘法大小。

CAFM 模块输出为

$$F_{\text{out}} = F_{\text{att}} + F_{\text{conv}} \tag{4-5}$$

Vit[64]的原始前馈网络（FFN）由两个线性层组成，用于单尺度特征聚合。FFN仅能聚合单一尺度的特征，所包含的信息有限，为了增强非线性特征转换，在每个 CAMixing 模块中，将 CAFM 的输出输入到多尺度前馈网络（MSFN），以实现多尺度特征聚合，提升非线性信息转换能力。

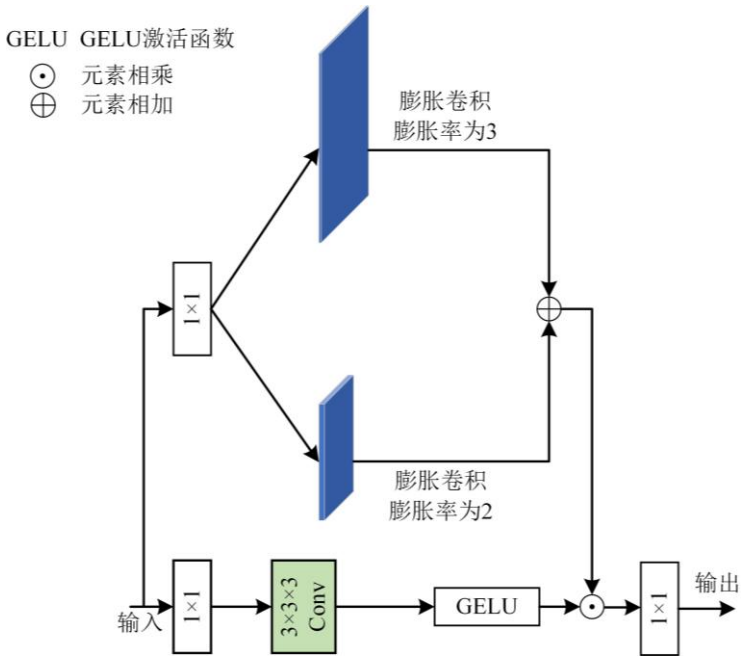


图 4-8 MSFN 网络结构

MSFN 网络结构图如 4-8 所示。使用两个 1×1 卷积扩展特征通道，扩展比 γ=2。输入特征在两个并行路径中处理，并引入门控机制，通过两个路径的特征逐元素乘积，以增强非线性变化能力。在上方的路径中，采用多尺度扩张卷积进行多尺度特征提取。使用两个 3×3 扩张卷积，扩张率为 2 和 3。在下方的路径中，使用深度卷积进行特征提取。

输入张量 X，MSFN 的公式为

$$g(X) = \phi(\mathcal{F}_{3\times3\times3}\mathcal{F}_{1\times1}(X)) \odot (\mathcal{F}^2_{3\times3}\mathcal{F}_{1\times1}(X) + \mathcal{F}^3_{3\times3}\mathcal{F}_{1\times1}(X)) \tag{4-6}$$

$$X_{\text{out}} = \mathcal{F}_{1\times1}\,\text{Gating}(X) \tag{4-7}$$

式中：$\odot$表示逐元素乘法，ϕ为 GELU 激活函数，$\mathcal{F}_{1\times1}(X)$ 表示 1×1 卷积，$\mathcal{F}^2_{3\times3}$ 表示膨胀率为 2 的 3×3 膨胀卷积。$\mathcal{F}^3_{3\times3}$ 表示膨胀率为 3 的 3×3 膨胀卷积。与 CAFM 相比，MSFN 具有独特的作用，专注于上下文信息丰富的功能。两者相结合，能捕捉全局上下文信息和局部特征。

4.2.3 损失函数

与水平框目标检测一样，旋转框体的损失函数同样包含三个部分：$Loss_{\text{Box}}$、$Loss_{\text{CLS}}$ 以及 $Loss_{\text{DFL}}$，其分别表示回归框的损失、类别损失以及角度预测损失。与水平框的损失函数不同，$Loss_{\text{Box}}$ 的定义有所变化。总体损失函数为

$$Loss = \alpha Loss_{\text{Box}} + \beta Loss_{\text{CLS}} + \gamma Loss_{\text{DFL}} \tag{4-8}$$

水平矩形框的损失函数 $Loss_{\text{Box}}$ 为 CIOU，旋转矩形框是 ProbIOU[59]，下面将会具体的介绍 ProbIOU 函数。

ProbIOU 函数将真实框和预测转化为高斯分布 $\mathcal{G}(\mu,\Sigma)$，如式(4-9)、式(4-10)所示。

$$\Sigma = \begin{bmatrix} a & c \\ c & b \end{bmatrix} = R_\theta \begin{bmatrix} a' & 0 \\ 0 & b' \end{bmatrix} R_\theta^T = \begin{bmatrix} a'cos^2\theta + b'sin^2\theta & \frac{1}{2}(a'-b')sin2\theta \\ \frac{1}{2}(a'-b')sin2\theta & a'sin^2\theta + b'cos^2\theta \end{bmatrix} \tag{4-9}$$

$$\mu = (x_0, y_0)^T \tag{4-10}$$

式中：R_θ表示二维旋转矩阵，该矩阵和旋转角度 θ 有关，a'，b'两个系数和旋转矩形框的宽和高有关，(x_0, y_0)为旋转矩形框的中心坐标。

ProbIOU 通过设计回归的(a', b', θ)的参数网络，使得旋转预测框和旋转的角度和真实矩形框更加接近。

通过计算巴士系数来获取两个高斯分布的相似度，两个二维概率密度函数之间的巴士系数可以被定义为

$$B_C(p,q) = \int_{R^2} \sqrt{p(x)q(x)}dx \tag{4-11}$$

式中：$p \sim \mathcal{G}(\mu_1, \Sigma_1)$，$q \sim \mathcal{G}(\mu_2, \Sigma_2)$，$p(x)$，$q(x)$分别表示真实框和预测框高斯分布的概率密度函数。

引入 Hellinger 距离来拉近两个高斯分布，使得真实框和预测框更多的重叠在一起，Hellinger 距离计算公式、ProbIOU 计算公式以及损失函数如式(4-12)、式(4-13)及式(4-14)所示。

$$H_{\mathrm{D}}(p,q)=\sqrt{1-B_c(p,q)} \tag{4-12}$$

$$ProbIoU=1-H_D(p,q) \tag{4-13}$$

$$L_{ProbIoU}(p,q)=1-ProbIoU(p,q) \tag{4-14}$$

4.3 实验结果和分析

4.3.1 实验参数设置

实验环境如表 4-1 所示。网络训练采用 SGD 优化器，初始学习率设为 0.01，训练周期为 400 epoch。为提高实验效率，若训练过程中模型性能不再提升，则提前终止训练。训练批次大小为 16。

为提升模型在复杂场景中的泛化能力，在训练过程中引入了在线数据增强策略，以增强训练样本的多样性，IYOLOV8-OBB 模型每次迭代时，从原始数据集中随机选取一个样本，并对其进行一系列随机变换以生成增强样本，然后用于当前批次的训练。在线数据增强方法包括马赛克增强、混合增强、随机扰动和颜色扰动四种方式。

表 4-1 实验环境参数

配置	参数
操作系统	Ubuntu20.04
处理器	Xeon(R)
内存	12G
显卡	RTX 3080Ti 显卡
框架	PyTorch 1.11.0
Cuda	11.3
编程软件	Pycharm

4.3.2 消融实验

为了验证 EEGCA、CCFM、CARAFE 以及 CAMixing 的有效性，在数据集 CTST-1600 上进行了大量消融实验，其结果如表 4-2 可知，以 YOLOV8-OBB 检测模型为实验的基线进行消融实验。在骨干网络中引入 EEGCA 进行优化，在颈

部网络中引入 CCFM、CARAFE 和 CAMixing 进行优化。通过实验数据分析与模型检测结果，评估这些模块对模型性能的提升。

表 4-2 消融实验结果

骨干网络	CCFM	CARAFE	CAMixing	P(%)	R(%)	F(%)	Params(M)	FPS
O	O	O	O	92.40	92.20	92.29	6.5	76.48
EEGCA	O	O	O	92.70	92.80	92.74	6.4	73.67
EEGCA	P	O	O	93.10	93.20	93.15	4.3	68.04
EEGCA	O	P	O	93.20	93.10	93.15	6.4	68.73
EEGCA	O	O	P	93.10	93.00	93.05	9.1	70.55
EEGCA	P	P	O	93.70	93.30	93.50	4.6	75.90
EEGCA	P	O	P	93.40	93.10	93.25	4.8	69.30
EEGCA	O	P	P	94.70	92.80	93.74	9.4	66.67
EEGCA	P	P	P	94.50	93.40	93.95	5.1	74.82

(1) EEGCA 优化的骨干网络

表 4-2 中，“O”表示未加入相应模块，“P”则相反，由表 4-2 可知，用 EEGCA 优化后的骨干网络，在 P 值、R 值和 F 值上分别提高了 0.3%、0.6%和 0.45%。原因在于，EEGCA 优化后的骨干网络增强了全局视角信息和不同尺度大小的特征，同时其良好的跨通道信息捕捉能力提高了交通标志牌文本的检测效果。由图 4-9(a)和图 4-9(b)可知，EEGCA 优化后的模型，减少了因交通标志牌文本行大小和比例差异引起的漏检。

(a) YOLOV8-OBB 模型检测结果

(b) EEGCA 优化后模型检测

图 4-9 EEGCA 优化后的检测效果对比

(2) EEGCA+CCFM 优化的检测模型

由表 4-2 可知，用 EEGCA 优化后的骨干网络，再使用 CCFM 对其颈部网络

进行优化，P 值、R 值和 F 值分别提高了 0.4%、0.4%和 0.41%。原因在于，CCFM 能整合不同尺度特征，增强模型对尺度变化的适应性。由图 4-10(a)和图 4-10(b)可知，EEGCA 和 CCFM 优化后的模型，减少了因语言多样性，交通标志牌文本行大小和比例差异引起的漏检。

(a) EEGCA 优化后模型检测

(b) EEGCA+CCFM 优化后模型检测结果

图 4-10 EEGCA 与 EEGCA+CCFM 检测效果对比

(3) EEGCA+CARAFE 优化的检测模型

由表 4-2 可知，用 EEGCA 优化后的骨干网络，再使用 CARAFE 对其颈部网络进行优化，P 值、R 值和 F 值分别提高了 0.5%、0.3%和 0.41%。原因在于，CARAFE 引入了更大的感受野，能有效地聚合了上下文信息，增强对重要内容的感知。由图 4-11(a)和图 4-11(b)可知，EEGCA 和 CARAFE 优化后的模型，减少了因语言多样性，文本行大小和比例差异引起的漏检和错检。

(a) EEGCA 优化后模型检测

(b) EEGCA+CARAFE 优化后模型检测

图 4-11 EEGCA 与 EEGCA+CARAFE 检测效果对比

(4) EEGCA+CAMixing 优化的检测模型

由表 4-2 可知，用 EEGCA 优化后的骨干网络，再使用 CAMixing 对其颈部网络进行优化，P 值、R 值和 F 值分别提高了 0.4%、0.2%和 0.31%。原因在于，CAMixing 捕捉了全局上下文信息和局部特征。由图 4-12(a)和图 4-12(b)可知，EEGCA 和 CAMixing 优化后的模型，减少因交通标志牌文本行的大小和比例差异，字符间稀疏度引起的漏检，同时减少了复杂背景下交通标志牌文本的错检。

(a) EEGCA 优化后模型检测

(b) EEGCA+CAMixing 优化后模型检测

图 4-12 EEGCA 和 EEGCA+CAMixing 检测效果对比

(5) 组合实验

此外，随机组合 CCFM、CARAFE 和 CAMixing，并测试每种组合的检测性能（由表 4-2 的第七行至第十行所示）。实验结果表明提出的方法是有效的。

使用 EEGCA、CCFM、CARAFE 以及 CAMixing 对 YOLOV8-OBB 模型进行改进，与基线相比，模型在 CTST-1600 数据集上的 P 值、R 值和 F 值分别提高了 2.1%、1.2%和 1.56%，模型大小 Params 减少了 1.4M。实验数据表明 EEGCA、CCFM、CARAFE 以及 CAMixing 模块不仅提高了交通标志牌文本检测性能，还使模型更轻量化。IYOLOV8-OBB 检测性能上略差于第三章的 BPI-DBNet，但模型大小减少近 90 倍。

4.3.3 对比实验分析

为了体现 IYOLOV8-OBB 交通标志牌文本检测模型在基于回归的目标检测方法中检测性能优异，表 4-3 中选取了较为权威的目标检测模型，将其检测头替换成旋转框检测头（OBB），用于交通标志牌文本检测。

表 4-3 回归模型对交通标志牌文本检测结果

模型	P(%)	R(%)	F(%)	Params(M)	FPS
YOLOV3[27]-OBB	92.00	93.10	92.55	210.20	65.66
YOLOV5[58]-OBB	92.00	92.60	92.30	5.6	97.48
YOLOV6[65]-OBB	92.10	93.5	92.79	8.9	95.10
YOLOV8[37]-OBB	92.40	92.20	92.29	6.5	76.48
Re-detr[61]-OBB	93.60	92.60	93.10	60.90	71.46
YOLOV9[66]-OBB	92.9	93.0	92.95	4.9	91.71
YOLOV10[67]-OBB	92.5	93.4	92.95	5.8	113.17
IYOLOV8-OBB	94.50	93.40	93.95	5.1	74.82

由表 4-3 可知，IYOLOV8-OBB 的 P 值和 F 值分别达到 94.50%和 93.95%。在检测性能上优于其它模型且模型大小 Params 值相对较小。

表 4-4 BPI-DBNet 和 IYOLOV8-OBB 实验结果对比

模型	P(%)	R(%)	F(%)	Params(M)	FPS
BPI-DBNet	94.59	93.50	94.04	457	21.44
IYOLOV8-OBB	94.50	93.40	93.95	5.1	74.82

由表 4-4 可知，IYOLOV8-OBB 的模型大小与 BPI-DBNet 相比，减少了近 90 倍，但检测性能略差于 BPI-DBNet。需对 IYOLOV8-OBB 进一步优化，使其检测性能提高的同时，模型更加轻量化。

4.4 本章小结

本章采用旋转框来定位文本区域。考虑到基于回归的 YOLOV8-OBB 方法计算资源占用较少，但在交通标志牌文本检测任务中的性能仍存在不足，本章提出改进模型 IYOLOV8-OBB。该模型在 YOLOV8-OBB 基础上，引入 EEGCA、CCFM、CARAFE 以及 CAMixing，四个模块来对 YOLOV8-OBB 进行优化。实验结果表明，IYOLOV8-OBB 在 CTST-1600 数据集上的 P 值和 F 值分别为 94.50%和 93.95%，均优于其它文本检测模型；IYOLOV8-OBB 与 BPI-DBNet 相比，Params 值减小近 90 倍。显示出更强的轻量化优势，但在检测精度上略逊一筹。后续工作将进一步优化 IYOLOV8-OBB 模型，在保证检测精度的同时降低模型复杂度，以实现更高效的部署性能。

第5章 交通标志牌文本检测系统构建

5.1 引言

在智能交通系统迅速发展的背景下，交通标志牌文本的高效检测对于保障自动驾驶系统的准确感知与决策具有重要意义。然而，受限于无人驾驶车辆车载平台的计算资源，如何在确保检测精度的前提下，降低模型复杂度与计算开销，成为实际部署中的关键问题。为此，针对 IYOLOv8-OBB 模型在交通标志牌文本检测中的应用，提出了一种基于模型剪枝与知识蒸馏的轻量化优化方法。首先引入 LAMP[68]剪枝算法，削减模型冗余参数，构建高效的轻量化网络结构，随后，结合 CWD[69]知识蒸馏策略，以 SCINet[70]增强的 IYOLOv8-OBB 作为教师模型，其模型尺寸与复杂度设定为“x”，通过中间特征的迁移引导学生模型学习，进一步提升剪枝后模型的检测性能。最终，在实验车平台上完成模型部署与实车验证，验证其在实际道路场景中的实时检测能力与工程可行性。

5.2 交通标志牌文本检测模型优化

5.2.1 LAMP 剪枝

LAMP 剪枝是模型压缩技术，通过识别并移除对模型性能影响较小的参数（即离群点），以降低模型复杂性。该方法能加快模型的推理速度，使其更适用于资源受限的环境。此外，LAMP 剪枝还通过减少离群数据来降低模型的过拟合风险。

LAMP 算法的核心在于其 LAMP 评分机制，该机制突破了传统神经网络剪枝中难以确定的层级稀疏度问题。通过评估模型的临界失真度量，实现自动确定权重剪枝阈值，从而在保持模型性能的同时优化模型的稀疏度。LAMP 评分方法继承了幅度剪枝（MP）的优点，避免复杂的超参数调整。在深度前馈神经网络中，LAMP 评分机制对不同层权重张量($W[1] \ldots W[d]$)进行评估，其中，卷积层的权重张量为四维矩阵，而全连接层的张量则为二维矩阵。为了确保评分标准的统

一性，无论权重张量原始维度如何，LAMP 评分均要求将其展开成一维向量。LAMP 评分按照特定的索引映射，对权重值进行升序排列，确保当索引 n 小于 m 时，$W[n]$的值小于等于 $W[m]$的值。这种有序排列方式为权重项的比较提供了一种有序的框架。其中，$W[n]$和 $W[m]$分别表示展开后张量在索引 n 和 m 处的权重值，而权重张量 W 的第 n 个指标 LAMP 评分定义为

$$\text{score}(n;W)=\frac{((W[n])^2)}{\sum_{m\geq n}(W[m])^2} \tag{5-1}$$

由图 5-1 可知，LAMP 分数是优化连接重要性的方法，其通过计算每一层内部连接的重要性，识别并修剪权重较小的连接，同时考虑不同索引映射的影响，即使权重相同，连接的 LAMP 分数因结构不同而有所变化。

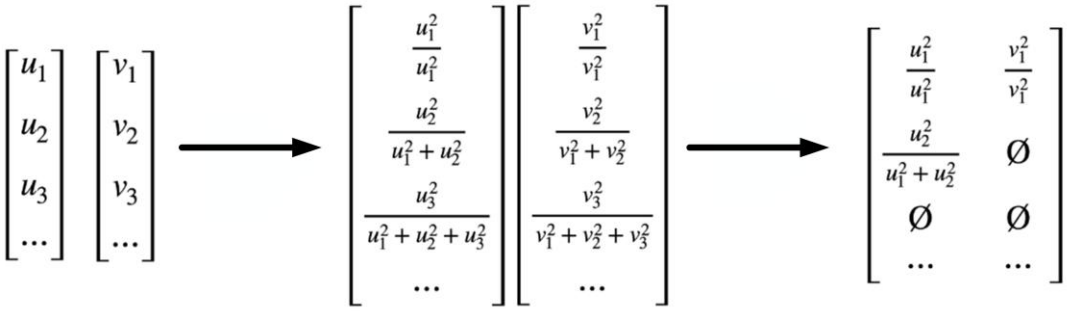


图 5-1 LAMP 剪枝算法流程图

在完成 LAMP 分数计算后，系统将进行全局搜索，依次剪除 LAMP 分数最小的连接，直到满足预设的稀疏性约束。这过程类似于自适应执行分层稀疏度选择的 MP 算法，通过全局得分判断并最小化失真，从而在模型的稀疏度和性能之间实现平衡。全局得分判断及最小化失真公式为

$$(W[n])^2>(W[m])^2\Rightarrow score(n;W)>score(m;W) \tag{5-2}$$

$$\min_{\sum_{i=1}^{d}||M^{(i)}||_0\leq k||x||_0\leq 1} sup\,||f(x;W^{(1:d)})-f(x;W^{(1:d)})||_2 \tag{5-3}$$

式中：k 代表模型级稀疏性约束，$W^{(i)}=M^{(i)}\odot W^{(i)}$ 表示第 i 层权重矩阵的剪枝。

5.2.2 知识蒸馏

CWD 知识蒸馏方法用于指导 LAMP 剪枝后的 IYOLOV8-OBB 模型（学生模型）。

教师模型是在 IYOLOV8-OBB 交通标志牌文本检测模型的基础上加入自校正照明网络（SCINet），其用于低光照图像增强。第四章使用的 IYOLOV8-OBB 尺寸和复杂度是“n”，教师模型改为“x”。“n”和“x”的深度（d）和宽度比率（w）参数分别为 0.33 和 0.25，1 和 1.25。“n”具有更小的模型和更快的推理速度，但检测性能相对较低，而“x”则相反，具有更高的检测性能，但模型的大小较大，推理速度较慢。由于教师模型对检测性能要求较高，且对模型的大小和推理速度无严格限制。因此，选择复杂度“x”作为教师模型，以指导学生模型的训练。

SCINet 基于 Retinex[71]理论，低光照环境下的观测图像 η 可以通过清晰图像 z 与光照 ε 的点乘来表示。在照明网络中，优化的核心在于对光照的准确估计。一旦获得精确的光照估计，便可利用这一关系恢复出清晰的图像。该方法受到现有文献中关于分阶段优化光照策略的启发[72]，设计出分阶段的光照优化网络。该网络在优化过程中，每个阶段包含由残差和光照估计网络组成的基本单元，其具体形式如公式(5-4)所示。

$$F(\varepsilon_t)=\begin{cases} u_t = H_\theta(\varepsilon_t), \varepsilon_0 = \eta \\ \varepsilon_{t+1} = \varepsilon_t + u_t, \text{other} \end{cases} \tag{5-4}$$

式中：η 代表低照度图像，u_t 和 ε_t 分别代表在第 t 阶段的残差和光照，θ 是权重，H_θ 是光照估计网络。H_θ 的结构和参数在所有阶段都是共享的，且不随阶段变化。

此外，为确保网络在各个阶段的输出接近目标值，且能保持一致，该方法引入了校准模块。该模块通过分析不同阶段之间的相互关系，确保在训练过程中，输出都能够稳定地收敛到一致的状态。该模块如式(5-5)所示。

$$G(\varepsilon_t):\begin{cases} z_t = \eta \odot \varepsilon_t \\ \delta_t = K_\theta(z_t) \\ v_t = \eta + \delta_t \end{cases} \tag{5-5}$$

式中：z 表示希望获得的清晰图像，δ 是自校正映射，K_θ 是参数化的运算符，v_t 是经过校准后，用于下一阶段处理的输入，符号 $\odot$ 表示按元素进行除法操作。

通过自校正映射和参数化操作符对输入进行校准，并采用逐元素除法，以生成下一阶段的输入，从而逐步逼近期望的清晰图像。

在原始光照学习过程中，每个后续阶段的输入是前一阶段的结果。随着每个

阶段的进展，光照优化过程的基本单元按公式(5-6)重新定义为

$$F(\varepsilon_t) \rightarrow F(G(\varepsilon_t)) \tag{5-6}$$

原本需多阶段进行的级联测试，现可简化为单阶段测试，从而降低推理过程中的计算成本。此外，由于网络的参数数量保持较小且恒定，其空间复杂度为常数级，无论输入规模如何增长，空间复杂度都不会随之增加。由图 5-2 可知，自校准照明网络（SCI）的结构主要分为两个部分：自校正模块，辅助降低级联模式的计算负担。照明估计模块，负责进行光照估计。

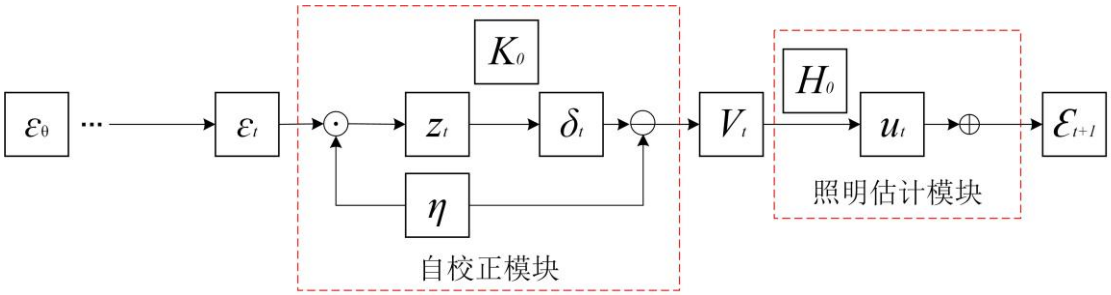


图 5-2 自校正照明网络（SCINet）结构

图 5-3 是 SCINet 算法处理低光照图片的效果图。

(a) 未经过处理的低光照图像

(b) SCINet 算法处理低光照图片

图 5-3 SCINet 算法处理低光照图像情况

图 5-3(a)为未经过处理的低光照图像，图 5-3(b)为处理过的低光照图像，由图 5-3 可知，经过 SCINet 处理后的低光照图像清晰度有所改善。

通道级知识蒸馏（CWD）利用教师网络的中间特征图来指导和辅助学生网络的训练，从而提升剪枝后模型在交通标志牌文本检测准确率，具体流程如图 5-4 所示。

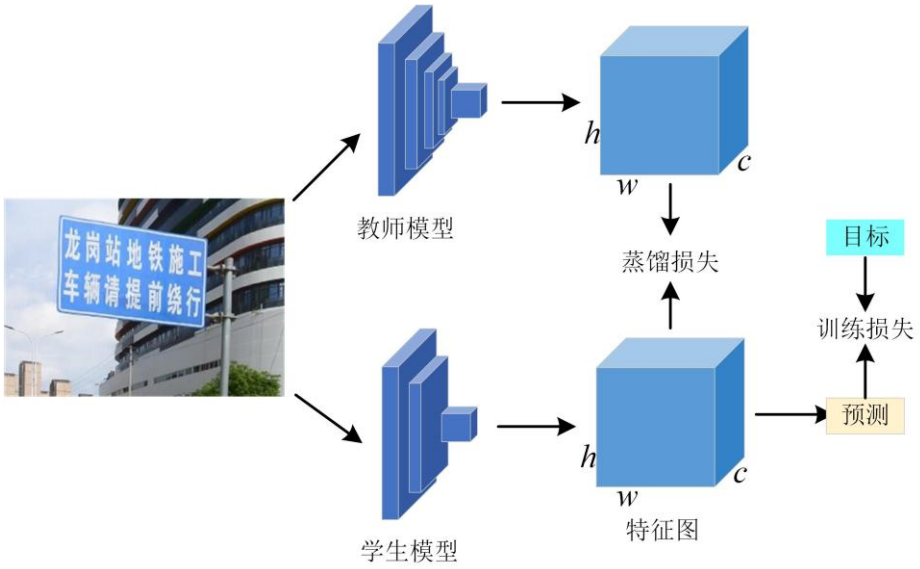


图 5-4 CWD 网络结构

CWD 通过将教师网络特定层的每个通道进行特征激活值归一化，形成空间位置的概率分布。对学生网络相应层级的每个通道执行相同的归一化处理。为了优化学生网络的学习过程，CWD 通过最小化这两个概率分布之间的 KL 散度，使学生模型在交通标志牌文本检测任务中更关注具有显著特征激活值的区域，从而提高定位准确性。

教师网络和学生网络分别用 t 和 s 表示，其特征激活值分别为 y^t 和 y^s，由式(5-7)可知，在计算蒸馏损失时，蒸馏温度 T 是一个关键参数，影响教师网络的概率分布形状。当 T 增大时，教师网络的概率分布变得平缓，这表明模型在蒸馏过程中会考虑通道C 内更广泛的空间区域，而不是仅仅关注激活值较高的局部区域。

$$L_{distill}=\frac{T^2}{C}\sum_{C=1}^{C}\sum_{i=1}^{W\cdot H}\phi(y_{C,i}^{T})\cdot\ln\left[\frac{\phi(y_{C,i}^{t})}{\phi(y_{C,i}^{s})}\right] \tag{5-7}$$

$$\phi(y_C)=\frac{\exp\left(\frac{y_{C,i}}{T}\right)}{\sum_{i=1}^{W\cdot H}\exp\left(\frac{y_{C,i}}{T}\right)} \tag{5-8}$$

式中：C 是通道索引，i 是通道 C 中的空间位置。

根据式(5-8)，将教师网络和学生网络每个通道特征激活值转换为概率分布，有助于消除两者在幅度尺度上的差异。针对通道数量不一致的情况，使用 1×1 卷积层来调整教师网络的特征通道数，以确保与学生网络一致。

5.3 实验结果与分析

5.3.1 实验参数设置

实验环境参数从表 5-1 可知。剪枝中训练轮次设置为 300，训练批次设置为 16，优化器为 SGD，使用全局剪枝，剪枝前的模型计算量除以剪枝后的模型计算量设置为 2。在蒸馏中训练的轮次设置为 300，训练批次设置为 16，优化器为 SGD。

表 5-1 实验环境参数

配置	参数
操作系统	Ubuntu20.04
处理器	Xeon(R)
内存	12G
显卡	RTX 3080Ti 显卡
框架	PyTorch 1.11.0
Cuda	11.3
编程软件	Pycharm

5.3.2 对比实验分析

六种剪枝算法对 IYOLOV8-OBB 的交通标志牌文本检测模型进行剪枝，其检测效果由表 5-2 可知。

表 5-2 各剪枝算法实验结果

剪枝算法	P(%)	R(%)	F(%)	Params(M)	FPS
L1[73]	91.90	91.60	91.75	3.5	72.82
Slim[74]	90.50	86.90	88.66	2.9	69.73
Group_hessian[75]	91.80	92.80	92.29	3.0	71.58
Group_norm[76]	90.70	89.00	89.84	3.1	66.93
Group_slim[77]	91.70	88.50	90.07	2.9	67.28
LAMP	92.20	92.50	92.35	2.2	69.47

由表 5-2 可知，LAMP 算法相对于其它剪枝算法，F 值达到 92.35%，模型大小为 2.2M，其在检测性能减少较小的情况下，模型大小明显下降。

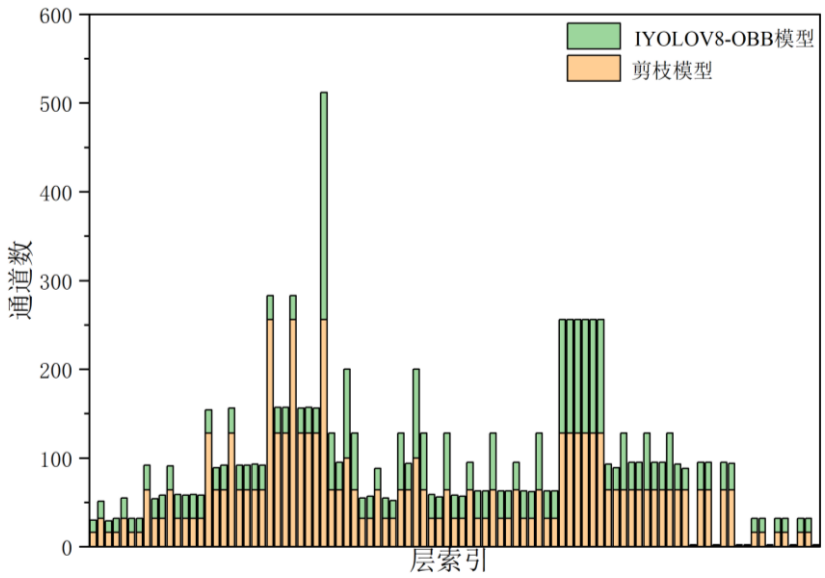


图 5-5 IYOLOV8-OBB 剪枝前后通道数对比

图 5-5 是 IYOLOV8-OBB 模型和剪枝模型在不同层索引下的通道数对比。绿色柱状图表示原始 IYOLOV8-OBB 模型的通道数，橙色柱状图表示经过 LAMP 剪枝后的通道数。LAMP 通过移除权重较小的通道来减少模型的参数量。由图 5-5 可知，多数层的通道数在剪枝后有所减少，部分层的通道数更是大幅下降，这表明 LAMP 剪枝能够有效去除冗余通道，从而降低了模型参数量和计算量，同时保持了部分关键层的通道数，以维持模型的核心特征提取能力。

表 5-3 教师模型和蒸馏后模型的实验结果

模型	P(%)	R(%)	F(%)	Params(M)	FPS
教师模型	93.7	96.20	94.93	105.5	43.96
蒸馏后模型	93.5	96.10	94.78	2.2	90.23

由表 5-2 和表 5-3 可知，剪枝后的模型在教师模型的指导下，其检测性能得到了提升，同时模型大小没有改变，适合部署在实验车上。相比于 BPI-DBNet 和 IYOLOV8-OBB，其具备了两者的优点，不仅模型小，益于部署，同时检测性能优异，满足交通标志牌文本检测高检测性能的要求。

5.4 无人驾驶汽车交通标志牌文本检测系统搭建

在无人驾驶汽车的行驶过程中，目标检测系统作为关键的感知模块，承担着对环境中关键目标的快速识别与定位任务。其中，交通标志牌文本检测作为目标检测的一项重要子任务，主要负责精确定位交通标志牌上的文字区域，为后续的

文本识别和语义理解提供基础支持。

图 5-6 智能网联测试装调实验车

为了验证剪枝与知识蒸馏优化后 IYOLOV8-OBB 模型的实际应用效果，本文将该模型嵌入至图 5-6 所示的智能网联测试装调实验车中，并在真实道路场景下进行部署测试。通过对检测性能、响应速度和资源消耗等指标的评估，全面考察所提出模型在复杂环境中运行的稳定性与实用性，为其在智能驾驶感知系统中的落地应用提供支撑依据。

5.4.1 感知层硬件配置

无人驾驶汽车的感知系统类似于智能神经系统，通过集成的传感器网络实时获取车辆位置信息、识别周围障碍物，并判断可行驶区域，从而确保车辆在遵守交通规则的前提下安全高效地行驶。图 5-7 展示实验车辆的传感器部分。

图 5-7 实验车上的传感器

在环境感知教学实训平台中，前向毫米波雷达主要用于探测车辆前方 100 米范围内的障碍物信息，包括相对距离和相对速度。激光雷达则负责监测车辆周围的障碍物，与导航融合以提高定位精度和方向感知。单目相机可以通过拍摄周围环境图片，送入到工控机上进行目标的检测和识别。无线路由器在平台中为导航模块提供 4G 网络支持，并兼具路由功能。360 全景摄像头通过安装在实验车前后及外后视镜上的摄像头采集影像，在启用 360 全景辅助系统时，实验车的工控机接收并处理这些影像，生成俯视视角，实现全景环视功能。系统同步采集车辆四周的影像，并利用图像处理技术和智能算法进行处理，最终生成一幅全景俯视图，在屏幕上直观显示车辆的位置及其周边环境。超声波雷达通过发射超声波信号，并通过接收从目标反射回来的回波来计算目标的距离、速度和方向。

表 5-4 智能网联测试装调实验车工控机硬件配置

硬件清单	硬件信息
CPU	Inter Core i7
硬盘	≥ 256GB
RAM	≥ 16GB
键盘	K620U
鼠标	N1200
显卡	GTX1660 6G
CAN 卡	/
Pcie 视频采集卡	/

感知系统要求工控机具有强大的数据处理能力，以确保能够迅速准确地处理大量实时传感器数据，保障无人驾驶汽车的稳定和安全运行，表 5-4 是工控机的配置，该配置保证了工控机在处理各类传感器数据时，依然可以完成交通标志牌文本检测任务。

5.4.2 检测系统检测流程

目标检测系统作为无人驾驶汽车感知系统的核心组件，通过分析单目相机捕获的实时视频流，可以检测出交通标志牌文本的位置，并将该位置的文本进行识别，以此来获取道路上的位置和环境信息，智能的控制无人驾驶汽车运动和选择最佳行驶路径。检测系统的处理过程如图 5-8 所示，主要分为以下四步：

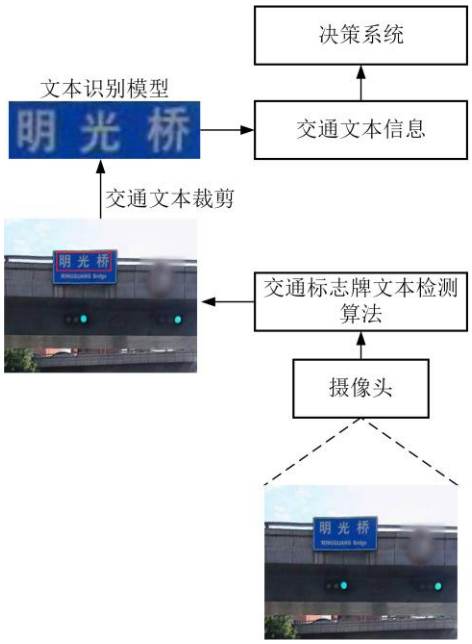


图 5-8 交通标志牌文本检测模型流程图

（1）摄像头初始化阶段，该阶段需要调整其相对位置参数以校准并减少目标检测误差，同时设置为采样模式以准备捕获实时图像数据。

（2）图像预处理阶段，该阶段应用滤波和归一化技术来改善图像质量，从而提高交通标志牌文本检测准确性。该过程能有效降低采集和数据传输引起的噪声，并减少不同光照条件下对图像分析结果的影响，确保目标检测系统在各种环境下都能稳定运行并输出准确的检测结果。

（3）模型检测阶段，将步骤（2）预处理后的图像输入到交通标志牌文本检测模型中，以此来检测出交通标志牌文本的位置。

（4）结果输出阶段，将步骤（3）中交通标志牌文本检测中文本的位置裁剪下来，并用文本识别的模型进行识别，来获取交通信息，把交通信息输入到决策系统，为接下来的控制决策提供信息依据。

5.4.3交通标志牌文本检测模型应用

如图5-9所示，交通标志牌文本检测模型部署到智能网联测试装调实验车上，并在实际道路环境中进行验证。图 5-9(a)展示了模型部署后的车辆外观，可见车身顶部或尾部安装了用于感知和采集道路信息的摄像头和传感器。图 5-9(b)则显示了车辆在真实路况下进行实验的场景。实验车行驶过程中对交通标志牌文本进

行实时检测，实验车能够及时定位到标志牌上的文字信息，为后续的驾驶决策或提示功能提供支持。这种基于实际道路测试的方式，有助于检验模型在光照变化、视角偏移以及复杂背景等多种不确定因素下的稳定性和实用价值。

(a) 交通模型部署到实验车　　(b) 部署后的实验车在路上实测

图 5-9 交通模型部署实测

图 5-10 展示了模型部署到实验车辆后，在实际道路环境中对交通标志牌文本的检测效果。由图 5-10 中可以看出，实验车在城市道路和校园场景中均能准确检测并定位交通标志牌文本，无论标志牌是悬挂在横梁上还是竖立于路旁，模型在不同光照条件和复杂背景下均能保持稳定的检测性能，对拍摄角度变化导致的文本倾斜也具有良好适应性，能够准确框选出文字区域，检测过程可在车辆行驶中实时完成，充分体现了模型的鲁棒性与实用性。

此外，图 5-10 中部分场景还具有一定的挑战性，但模型仍表现出良好的适应能力。例如，第 1 行第 2 列图像中，标志牌处于树荫遮挡区域，光照条件较弱，模型依然准确识别文本，体现了其在弱光环境下的鲁棒性；第 1 行第 3 列中，标志牌被树叶部分遮挡，模型依然正确检测关键交通信息，显示出良好的抗遮挡能力；第 2 行第 1 列中，标志牌文字呈纵向排列且结构复杂，模型依旧准确检测，说明其对文本排列方向具有较强的适应性；第 2 行第 2 列图像中，标志牌悬挂距离远、分辨率较低，模型仍能完成有效识别，体现其在小尺度目标检测方面的能力；第 3 行第 3 列图像中，标志牌上方存在强烈阳光直射导致光照过曝，模型依旧能定位文本区域，表现出较强的抗曝光能力。

图 5-10 实验车检测交通标志牌文本效果

5.5 本章小结

本章围绕交通标志牌文本检测系统的构建，主要开展了三个方面的工作：模型优化、实验验证及系统部署。针对自动驾驶场景下对高精度与低算力的需求，提出了一种融合 LAMP 剪枝与 CWD 知识蒸馏的轻量化优化方案。该方法通过剪枝有效减少冗余参数与计算开销，提升模型运行效率；同时结合知识蒸馏策略，引导学生模型学习教师模型的特征提取能力，提升检测精度与鲁棒性。优化后模型在 CTST-1600 数据集上 F 值达到 94.78%，参数量仅为 2.2M，较 BPI-DBNet 与原始 IYOLOv8-OBB 在性能与复杂度上均有明显优势。最终模型成功部署于实验车辆，验证了其在实际道路场景中的实时检测能力与工程可行性。

第6章 总结与展望

6.1 总结

准确识别交通标志牌上的文本对于自动驾驶系统理解交通规则、规范行驶以及减少交通事故具有重要意义。作为交通文本识别的关键步骤，交通标志牌文本检测直接影响自动驾驶车辆获取道路环境信息的准确性，进而关系到行车安全。然而，当前交通文本检测方法面临漏检和错检较高的问题，同时模型复杂度高导致计算资源消耗较大。针对这些问题，本文主要工作内容包括以下三个方面：

（1）针对复杂环境下交通标志牌文本出现漏检和错检的问题，提出了名为BPI-DBNet 的交通标志牌文本检测方法。BPI-DBNet 在 DBNet 的基础上，采用双向特征金字塔（BiFPN）作为颈部网络，并引入综合通道空间注意力模块（ICSAM）、Haar 小波下采样模块（HWD）、自校准空间金字塔模块（SC-SPPF）以及高效多尺度注意力模块（EMA）。ICSAM 减少特征提取中的信息损失，保留更多交通标志牌文本的定位和分类信息，有效的减少因交通标志牌文本行大小和比例差异、自然场景复杂引起的漏检和错检。HWD 减少下采样过程中特征丢失和削弱，SC-SPPF 增强高层次特征对文本几何大小的表达能力和上下文关系，两模块有效的减少因文本行大小和比例差异引起的漏检。EMA 善于捕捉局部细节与全局上下文，增强特征丰富度与区分度，有效的减少因字符稀疏度变化、语言多样性引起的错检。实验结果表明，BPI-DBNet 在 ICDAR 2015、MLT-2017 数据集上的 F 值上分别为 85.03%、74.29%，相比 DBNet 提高了 2.7%、3%。在中国交通标志牌文本数据集 CTST-1600 上 F 值达到 94.04%，展现出了优异的检测性能，但模型大小 Params 值为 457M，所需计算资源较多。

（2）针对无人驾驶汽车车载平台计算资源有限的问题，提出基于旋转框的交通标志牌文本检测模型（IYOLOV8-OBB）。因交通文本工整，IYOLOV8-OBB 首先利用旋转框定位由拍摄角度引起的倾斜文本，然后适配计算资源需求较低的基于回归的文本检测方法。IYOLOV8-OBB 在 YOLOV8-OBB 的基础上，引入高效全局通道注意力模块（EEGCA）、跨尺度融合模块（CCFM）、内容感知功能

重组模块（CARAFE）以及卷积注意力混合模块（CAMixing）。EEGCA 增强了全局视角信息和不同尺度大小的特征，有效减少因交通标志牌文本行大小和比例差异、自然场景复杂引起的漏检和错检。CCFM 增强了对尺度变化的适应性，CARAFE 增大感受野，聚合上下文信息，增强对重要内容的感知，两模块有效的减少因语言种类不同、文本行大小和比例差异引起的错检。CAMixing 捕捉全局上下文信息和局部特征，有效的减少因交通标志牌文本行的大小和比例差异、字符间稀疏度引起的漏检。实验结果表明，IYOLOV8-OBB 在 CTST-1600 上 F 值达到 93.95%，Params 值为 5.1M，与 YOLOV8-OBB 相比，F 值提高了 1.66%，Params 值减少了 1.4M。IYOLOV8-OBB 和 BPI-DBNet 相比，Params 值减少近 90 倍，检测性能略有下降，需进一步优化。

（3）为了进一步满足无人驾驶汽车低计算资源的要求，首先利用 LAMP 剪枝算法对 IYOLOV8-OBB 进行剪枝处理，缩减模型大小，但检测性能下降，然后采用 CWD 知识蒸馏算法进行蒸馏，该算法通过教师模型对剪枝后模型进行指导，在不增加模型大小的前提下，实现对模型检测性能的提升。教师模型是在 IYOLOV8-OBB 网络结构前加入了前处理模块 SCINet，提高低光照情况下的图片质量，其网络深度为“x”。实验数据表明，教师模型在 CTST-1600 数据集上 F 值达到 94.93%，检测性能较高，适用于教师模型。优化后模型在同一数据集上的 F 值达到了 94.78%，Params 值为 2.2M，相比 BPI-DBNet 和 IYOLOV8-OBB，F 值提高了 0.74%、0.83%，Params 值下降了 454.8M、2.9M。优化后模型具备高检测性能和轻量化。最后搭建无人驾驶汽车交通标志牌文本检测系统进行检测实验。实验结果表明，实验车在城市道路、校园场景中均能准确检测并定位交通标志牌文本，充分体现了检测系统在实际道路环境中的鲁棒性和实用性。

主要创新点可归纳为：

（1）针对复杂环境下交通标志牌文本漏检、错检问题，提出了基于双向特征金字塔的检测模型 BPI-DBNet。通过引入综合通道空间注意力、Haar 小波下采样、自校准空间金字塔和高效多尺度注意力机制，显著提升了模型的特征表达能力和检测精度。

（2）为解决模型复杂度高、难以部署的问题，设计了基于旋转框的轻量化

检测模型 IYOLOV8-OBB，融合多种注意力机制与结构优化模块。在检测性能略微下降的情况下，模型大小较 BPI-DBNet 减少近 90 倍，显著提升了车载平台的实时检测能力。

6.2 展望

通过研究，尽管在交通标志牌文本检测方法上取得了一些进展，但还存在值得探索和改进的地方，主要包括一下几个方面：

（1）交通标志牌文本数据集的扩充，数据集的数量决定着目标检测模型的训练深度和泛化能力。目前，公开的交通标志牌文本数据集相对较少，CTST-1600 数据集在数据量和场景复杂性方面相较于其他数据集仍存在一定不足。

（2）优化后的 IYOLOV8-OBB 交通标志牌文本检测算法，兼具了高精度和轻量化，但其对非工整文本的检测效果低，鲁棒性不强。后续将会研究鲁棒性较强同时兼具高检测性能和轻量化的交通标志牌文本检测方法。

参考文献

[1] 袁晓玲, 黄晓洲, 刘壤. 中国城市人口高质量发展水平测度及其驱动因素研究 [J]. 西安交通大学学报(社会科学版), 2025, 45 (02): 26-40.

[2] 魏宇晨, 林潭晨, 胡晓辉. 转型实验室视角下中国城市无人驾驶汽车发展特征与潜力分析[J]. 现代城市研究, 2024, (12): 116-123.

[3] 刘登峰, 郭文静, 陈世海. 基于内容引导注意力的车道线检测网络[J]. 浙江大学学报(工学版), 2025, 59 (03): 451-459.

[4] 牛文杰, 伊力哈木·亚尔买买提. 面向无人驾驶场景下的道路多目标检测算法[J]. 计算机应用与软件, 2024, 41 (08): 282-288.

[5] 张蕊, 高诗博, 赵霞, 等. 基于改进YOLOv5s的无人驾驶夜间车辆目标检测算法[J]. 电子测量技术, 2023, 46 (17): 87-93.

[6] 徐进, 陈钦, 陈正委, 等. 适应无人驾驶汽车的道路设施设计综述[J]. 西南交通大学学报, 2023, 58 (06): 1366-1377.

[7] Schapire R E, and Singer Y. Improved boosting algorithms using confidence-rated predictions[J]. Machine learning, 1999, 37(3): 297–336.

[8] Lee J, Lee P, Lee S, Yuille A L, and Koch C. Adaboost for text detection in natural scene[C]. in Proc. Int. Conf. on Document Anal. and Recognition, 2011: 429–434.

[9] Bosch A, Zisserman A, and Munoz X. Image classification using random forests and ferns[C]. in Proc. IEEE Int. Conf. on Comp. vision, 2007: 1–8.

[10] Wang K, Babenko B, Belongie S. End-to-end scene text recognition[C] //2011 International conference on computer vision. IEEE, 2011: 1457-1464.

[11] Neumann L, and Matas J. A method for text localization and recognition in real-world images[C]. in Proc. Asian Conf. on Comp. Vision. Springer, 2010: 770–783.

[12] Yao C, Bai X, Liu W, Ma Y, and Tu Z. Detecting texts of arbitrary orientations in natural images[C]. in Proc. IEEE Conf. on Comp. Vision and Pattern Recognit, 2012: 1083–1090.

[13] Epshtein B, Ofek E, and Wexler Y. Detecting text in natural scenes with stroke

width transform[C]. in Proc. IEEE Conf. on Comp. Vision and Pattern Recognit, 2010: 2963–2970.

[14] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M. Imagenet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211–252.

[15] Tian Z, Huang W, He T, et al. Detecting text in natural image with connectionist text proposal network[C] //Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VIII 14. Springer International Publishing, 2016: 56-72.

[16] Liao M, Shi B, Bai X, et al. Textboxes: A fast text detector with a single deep neural network[C] //Proceedings of the AAAI conference on artificial intelligence. 2017, 31(1).

[17] Liao M, Shi B, Bai X. Textboxes++: A single-shot oriented scene text detector[J]. IEEE transactions on image processing, 2018, 27(8): 3676-3690.

[18] Shi B, Bai X, Belongie S. Detecting oriented text in natural images by linking segments[C] //Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2550-2558.

[19] He M, Liao M, Yang Z, et al. MOST: A multi-oriented scene text detector with localization refinement[C] //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 8813-8822.

[20] Yuliang L, Lianwen J, Shuaitao Z, et al. Detecting curve text in the wild: New dataset and new solution[J]. arXiv preprint arXiv:1712.02170, 2017.

[21] Liu Y, Chen H, Shen C, et al. Abcnet: Real-time scene text spotting with adaptive bezier-curve network[C] //proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 9809-9818.

[22] Wang W, Xie E, Li X, et al. Shape robust text detection with progressive scale expansion network[C] //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 9336-9345.

[23] 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.

[24] 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.

[25] He X, Wang R, Li X, et al. HTSTL: Head-and-tail search network with scale-transfer layer for traffic sign text detection[J]. IEEE Access, 2019, 7: 118333-118342.

[26] 彭西明. 自然场景下的交通标志文本检测方法研究[D]. 武汉: 武汉理工大学, 2020.

[27] ul Haq Q M, Ruan S J, Haq M A, et al. An incremental learning of yolov3 without catastrophic forgetting for smart city applications[J]. IEEE Consumer Electronics Magazine, 2021, 11(5): 56-63.

[28] 胡高丽, 文成玉. 自然场景下交通标识文本检测与识别算法研究[J]. 成都信息工程大学学报, 2022, 37 (02): 171-176.

[29] 李金鹏. 基于深度学习的交通标志文本检测及识别的研究与实现[D]. 哈尔滨: 黑龙江大学, 2023.

[30] Dai J, Qi H, Xiong Y, et al. Deformable convolutional networks[C] //Proceedings of the IEEE international conference on computer vision. 2017: 764-773.

[31] 朱彦斌, 王润民, 陈华, 等. 基于多粒度特征增强网络的交通文本检测方[J]. 计算机工程, 2024(3).

[32] 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.

[33] Howard A G. Mobilenets: Efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint arXiv:1704.04861, 2017.

[34] 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.

[35] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C] //Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7263-7271.

[36] Ge Z. Yolox: Exceeding yolo series in 2021[J]. arXiv preprint arXiv:2107.08430, 2021.

[37] Li Y, Fan Q, Huang H, et al. A modified YOLOv8 detection network for UAV aerial image recognition[J]. Drones, 2023, 7(5): 304.

[38] Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. The journal of machine learning research, 2014, 15(1): 1929-1958.

[39] Wan L, Zeiler M, Zhang S, et al. Regularization of neural networks using dropconnect[C] //International conference on machine learning. PMLR, 2013: 1058-1066.

[40] Liu Z, Li J, Shen Z, et al. Learning efficient convolutional networks through network slimming[C] //Proceedings of the IEEE international conference on computer vision. 2017: 2736-2744.

[41] Tan M, Pang R, Le Q V. Efficientdet: Scalable and efficient object detection[C] //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 10781-10790.

[42] 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.

[43] Yao P, Cuan Y. Chinese Traffic Sign Text Detection Method Based on Improved EAST[C] //2021 6th International Symposium on Computer and Information Processing Technology (ISCIPT). IEEE, 2021: 713-716

[44] 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.

[45] 姜文恒. 基于深度学习的中文交通标志文本检测方法研究与应用[D]. 济南: 济南大学, 2024.

[46] Liu Z, Lin G, Yang S, et al. Learning markov clustering networks for scene text detection[J]. arXiv preprint arXiv:1805.08365, 2018.

[47] He P, Huang W, He T, et al. Single shot text detector with regional attention[C] //Proceedings of the IEEE international conference on computer vision. 2017: 3047-3055

[48] Hu H, Zhang C, Luo Y, et al. Wordsup: Exploiting word annotations for character based text detection[C] //Proceedings of the IEEE international conference on computer vision. 2017: 4940-4949.

[49] Deng D, Liu H, Li X, et al. Pixellink: Detecting scene text via instance segmentation[C] //Proceedings of the AAAI conference on artificial intelligence. 2018, 32(1).

[50] Tian Z, Shu M, Lyu P, et al. Learning shape-aware embedding for scene text detection[C] //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 4234-4243.

[51] 吕伶, 李华, 王武. 基于增强特征提取网络与语义特征融合的多方向文本检测 [J]. 图学学报, 2024, 45 (01): 56-64

[52] Lyu P, Liao M, Yao C, et al. Mask textspotter: An end-to-end trainable neural network for spotting text with arbitrary shapes[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 67-83.

[53] Lyu P, Yao C, Wu W, et al. Multi-oriented scene text detection via corner localization and region segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7553-7563.

[54] Zhang C, Liang B, Huang Z, et al. Look more than once: An accurate detector for text of arbitrary shapes[C] //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 10552-10561.

[55] Xie E, Zang Y, Shao S, et al. Scene text detection with supervised pyramid context network[C] //Proceedings of the AAAI conference on artificial intelligence. 2019, 33(01): 9038-9045.

[56] Howard A, Sandler M, Chu G, et al. Searching for mobilenetv3[C] //Proceedings of the IEEE/CVF international conference on computer vision. 2019: 1314-1324.

[57] Ma N, Zhang X, Zheng H T, et al. Shufflenet v2: Practical guidelines for efficient cnn architecture design[C] //Proceedings of the European conference on computer vision (ECCV). 2018: 116-131.

[58] Chen S, Liao Y, Lin F, et al. An improved lightweight YOLOv5 algorithm for detecting strawberry diseases[J]. IEEE Access, 2023, 11: 54080-54092.

[59] Llerena J M, Zeni L F, Kristen L N, et al. Gaussian bounding boxes and probabilistic intersection-over-union for object detection[J]. arXiv preprint arXiv:2106.06072, 2021.

[60] 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:

[61] Zhao Y, Lv W, Xu S, et al. Detrs beat yolos on real-time object detection[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 16965-16974.

[62] 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.

[63] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.

[64] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.

[65] Li C, Li L, Jiang H, et al. YOLOv6: A single-stage object detection framework for industrial applications[J]. arXiv preprint arXiv:2209.02976, 2022.

[66] Wang C Y, Yeh I H, Liao H Y M. Yolov9: Learning what you want to learn using

programmable gradient information[J]. arXiv preprint arXiv:2402.13616, 2024.

[67] Wang A, Chen H, Liu L, et al. Yolov10: Real-time end-to-end object detection[J]. arXiv preprint arXiv:2405.14458, 2024.

[68] Lee J, Park S, Mo S, et al. Layer-adaptive sparsity for the magnitude-based pruning[J]. arXiv preprint arXiv:2010.07611, 2020.

[69] Shu C, Liu Y, Gao J, et al. Channel-wise knowledge distillation for dense prediction[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 5311-5320.

[70] Ma L, Ma T, Liu R, et al. Toward fast, flexible, and robust low-light image enhancement[C] //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 5637-5646.

[71] Wei C, Wang W, Yang W, et al. Deep retinex decomposition for low-light enhancement[J]. arXiv preprint arXiv:1808.04560, 2018.

[72] Ma L, Ma T, Liu R, et al. Toward fast, flexible, and robust low-light image enhancement[C] //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 5637-5646.

[73] Li H, Kadav A, Durdanovic I, et al. Pruning filters for efficient convnets[J]. arXiv preprint arXiv:1608.08710, 2016.

[74] Liu Z, Li J, Shen Z, et al. Learning efficient convolutional networks through network slimming[C] //Proceedings of the IEEE international conference on computer vision. 2017: 2736-2744.

[75] LeCun Y, Denker J, Solla S. Optimal brain damage[J]. Advances in neural information processing systems, 1989, 2.

[76] Fang G, Ma X, Song M, et al. Depgraph: Towards any structural pruning[C] //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 16091-16101.

[77] Friedman J, Hastie T, Tibshirani R. A note on the group lasso and a sparse group lasso[J]. arXiv preprint arXiv:1001.0736, 2010.

攻读学位期间发表的学术成果

专利：

[1] 潘道远, **柳浩**, 等. 一种夹持式梅花花蕾采收设备及使用方法[P]. 中国专利: CN115152424B, 2023. (发明授权)

[2] 潘道远, **柳浩**, 等. 一种复杂天气情况下的交通标志牌文本检测方法[P]. 中国专利: CN117333839A, 2024. (发明实审)

[3] 潘道远, **柳浩**, 等. 一种轻量化的交通文本检测模型及方法[P]. 中国专利: CN118587679A, 2024. (发明实审)

[4] 潘道远, **柳浩**, 等. 一种基于旋转框体的文本检测方法[P]. 中国专利: CN118314563A, 2024. (发明实审)

[5] 潘道远, **柳浩**, 等. 一种基于改进的DBNet的文本检测方法及装置[P]. 中国专利: CN118262368A, 2024. (发明实审)

[6] 潘道远, 吴宜涛, **柳浩**, 等. 一种撞击式梅花花蕾用的采摘设备及梅花花蕾的采摘方法[P]. 中国专利: CN115152425A, 2023. (发明授权)

[7] 潘道远, **柳浩**, 等. 一种线束支撑机构及具有该线束支撑机构的支撑装置和缠绕系统[P]. 中国专利: CN221420324U, 2024. (实用新型)

科研竞赛：

[1] 排名第二，“中国光谷·华为杯”第十九届中国研究生数学建模竞赛三等奖.

[2] 排名第一，“华为杯”第二十届中国研究生数学建模竞赛三等奖.

[3] 第六届全国高校大学生外语水平能力大赛（英语赛道）特等奖.

[4] 2023年第十二届芜湖市大学生专利创新创业大赛二等奖.

致 谢

在攻读学位、撰写论文的岁月里，我深切感受到身边每一位给予我关心、支持与帮助的人所带来的温暖与力量。正是因为有了诸位的陪伴与鼓励，我才能在科研之路上披荆斩棘、砥砺前行，顺利完成这项工作。在此，我怀着无比感激之情，向所有给予我帮助的人致以最诚挚的谢意。

首先，我要衷心感谢我的导师潘道远副教授。在整个研究与写作过程中，潘老师始终以诲人不倦的精神和严谨求实的态度给予我悉心指导与无私支持。从研究方向的确立到具体问题的攻克，从论文框架的构思到细节内容的打磨，他都倾囊相授、悉心点拨，使我在困顿迷茫中重拾信心、坚定前行。论文的顺利完成，离不开潘老师的倾力相助与殷切教诲。

同时，我诚挚感谢本研究中所参考与引用的众多文献作者。正是他们精益求精、孜孜以求的治学精神和宝贵的学术成果，为我的研究奠定了坚实的理论基础，提供了重要的思路启发。他们的严谨态度与深邃思想，不仅拓宽了我的学术视野，也为本论文的深入展开提供了有力支撑。

我还要感谢在研究期间一路陪伴、鼎力相助的同学与朋友们。感谢室友们三年来的风雨相伴、肝胆相照，我们共同走过一个又一个寒来暑往的日夜，是你们让我的求学之路充满温情与力量。感谢课题组的师弟师妹们在实验与比赛中的密切配合、通力合作，让我在科研的征途上不再孤单。

最深沉的感激，献给我的父母。感谢你们始终如一的理解与支持，是你们的含辛茹苦、无私奉献给予我披星戴月、奋发向前的动力。每当我踟蹰不前，是你们给予我鼓励与信念；每当我意气风发，是你们静静守护在我背后。你们是我永远的依靠，是我人生路上最坚强的后盾。

三年的研究生生涯即将落下帷幕，这段征程虽已走到终点，但其中所凝结的知识、情谊与成长，将成为我人生中弥足珍贵的精神财富。我将带着这份感恩与铭记，继续砥砺奋进、勇往直前，追寻属于自己的星辰大海。

再次向所有关心、帮助与支持过我的人致以最诚挚的谢意，衷心祝愿大家前程似锦、万事胜意！

尚德教学 唯实惟新