分类号：TP391　　学校代码：10363
密　级：公开　　学　　号：2210320102

安徽工程大学
Anhui Polytechnic University

硕士学位论文

题目　基于深度学习的雨天环境下行人检测算法研究

论文作者　崔福荣
指导教师　韩超（教授）
学科（专业）　控制科学与工程
研究方向　模式识别与人工智能

论文提交日期：2025 年 5 月 15 日

分类号：TP391　　单位代码：10363

密类级：公开　　学　　号：2210320102

题　　目 基于深度学习的雨天环境下行人检测算法研究

英文题目 Research on Deep Learning-Based Pedestrian Detection Algorithms in Rainy Environments

论文作者：崔福荣

导　　师：韩超教授

专　　业：控制科学与工程

研究方向：模式识别与人工智能

论文答辩日期：2025 年 5 月 12 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 14 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。

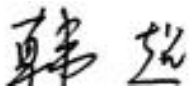
本学位论文属于

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 14 日

日期： 2025 年 5 月 14 日

基于深度学习的雨天环境下行人检测算法研究

摘要

近年来，随着城市化进程的加速，人们在智慧交通、自动驾驶等场景中对于目标检测的需求日益增大。基于深度学习的目标检测技术是社会智能化发展的重要技术之一，其核心原理是利用卷积神经网络学习图像层次化特征，再结合锚框机制与回归算法实现目标的定位与分类。行人检测作为目标检测的重要分支，它从多种传感器上获取行人的特征信息，并依赖这些信息判断行人的空间位置和预判行人的未来轨迹，以此来保障行人的人身安全。但是，目前的行人检测算法在精度上还不能够满足复杂气候以及密集遮挡场景下的需求。基于此，本文采用端到端的目标检测器算法作为行人检测的基础方法，设计了一个“雨天+密集人群”场景下行人检测的网络结构，并通过实验验证其有效性和可行性。本文具体工作如下：

（1） 针对雨天环境下行人特征提取困难导致检测精确度不高的问题，本文提出了一种结合动态直方图注意力机制与多特征聚合的去雨算法。该算法通过引入多特征聚合模块、全局空间感知模块和动态直方图注意力模块来有效提升去雨效果。具体而言，多特征聚合模块采用多尺度卷积技术，精确提取雨线的细微结构及复杂背景信息；全局空间感知模块通过深度可分离卷积和空间注意力机制，增强了特征之间的空间关系感知能力；而动态直方图注意力模块利用动态直方图注意力机制，有效去除雨线噪声并恢复图像的细节。实验结果表明，该算法能够在去除雨线噪声的同时，显著保留图像细节并提升图像质量。结合目标检测模型的实验进一步验证了该算法在恶劣天气下对行人检测任务的有效性，尤其在雨线遮挡和低能见度场景下，检测精度得到了显著提升，充分证明了该算法在雨天环境下的实用性和可靠性。

（2）针对密集行人检测精度较低的问题，本文改进了基于深度学习的行人检测算法。首先，针对拥挤行人经常出现的小目标众多、像素较低的问题，提出

在骨干网络中添加一个更高分辨率的特征尺度层来提升模型对小局部特征的敏感度，从而增强网络对小目标行人的感知能力，提升网络对行人检测的准确率。然后，针对因遮挡导致的不规则目标行人的特征提取能力下降的问题，引入多分支模块，提升网络对遮挡行人的多尺度特征提取能力。接着，针对复杂背景噪声干扰等问题引入高效通道注意力机制，通过动态调整特征图的权重分布，使网络在行人检测时更能突出感兴趣的区域，减少受背景噪声的影响，以提升行人检测时的精度。最后，融入一种对行人尺度变化低敏感的损失函数，能更准确地评估小目标之间的相似性，克服了小目标行人检测中因尺度变化导致的漏检误检。

（3）将优化后的图像去雨算法与优化后的行人检测模型结合，并进行对比实验来验证其有效性。改进后的模型在公开数据集 Crowdhuman 和自建数据集 Rain-Crowdhuman 数据集上进行实验，验证了其有效性和可行性。

关键词：YOLOv10，行人检测，恶劣天气，密集人群，Transformer

RESEARCH ON DEEP LEARNING-BASED PEDESTRIAN DETECTION ALGORITHMS IN RAINY ENVIRONMENTS

ABSTRACT

In recent years, with the accelerated urbanization process, demands for object detection in smart transportation, autonomous driving, and other scenarios have grown exponentially. Deep learning-based object detection technology has emerged as a pivotal component in societal intelligent development, leveraging convolutional neural networks (CNNs) to learn hierarchical image features and combining anchor box mechanisms with regression algorithms for accurate object localization and classification. As a critical subfield of object detection, pedestrian detection extracts feature information from diverse sensors and utilizes this data to determine spatial positions and predict future trajectories of pedestrians, thereby ensuring their safety. However, existing pedestrian detection algorithms still fall short in meeting accuracy requirements under complex weather conditions and in densely occluded scenarios. To address these limitations, this paper employs an end-to-end object detector as the foundational approach and proposes a specialized network architecture for pedestrian detection in "rainy weather + crowdhuman" scenarios, with its effectiveness and feasibility validated through extensive experiments. The main contributions of this work include:

(1) To address the challenge of degraded pedestrian feature extraction in rainy conditions leading to reduced detection accuracy, this paper proposes a novel rain removal algorithm that integrates dynamic histogram attention mechanisms with multi-feature aggregation. The algorithm significantly enhances deraining performance through three key components: the Multi-Scale Feature Aggregation

module, Global Spatial Perception module, and Dynamic Histogram Attention module. Specifically, the MSFA module employs multi-scale convolution techniques to precisely extract fine rain streak structures and complex background information. The GSP module combines depthwise separable convolution with spatial attention mechanisms to strengthen spatial relationship awareness among features. The DHAT module utilizes dynamic histogram attention to effectively eliminate rain noise while restoring image details. Experimental results demonstrate that our algorithm achieves superior rain streak removal while remarkably preserving image details and improving overall image quality. When integrated with object detection models, the proposed solution shows particular effectiveness in adverse weather conditions, with significant accuracy improvements observed in scenarios involving rain occlusion and low visibility. These findings thoroughly validate the algorithm's practical utility and reliability in rainy environments.

(2) To address the low detection accuracy in dense pedestrian scenarios, this paper proposes improves the deep learning-based pedestrian detection algorithm. Firstly, to address the issues of numerous small targets and low pixel resolution commonly found in crowded pedestrian scenarios, this paper proposes adding a higher-resolution feature scale layer to the backbone network. This modification enhances the model's sensitivity to small local features, thereby improving the network's perception of small-scale pedestrian targets and boosting the accuracy of pedestrian detection. Secondly, to overcome the degraded feature extraction performance caused by occlusions that lead to irregular pedestrian shapes, we introduce the Diverse Branch Block (DBB) module to strengthen the network's multi-scale feature extraction capacity for occluded pedestrians. Thirdly, to mitigate interference from complex background noise, we incorporate the Efficient Channel Attention (ECA) mechanism that dynamically adjusts feature map weight distribution, enabling the network to better focus on regions of interest while reducing background noise impact. Finally, we implement a scale-insensitive Normalized Wasserstein

Distance (NWD) loss function that provides more accurate similarity measurement for small targets, effectively addressing both missed detections and false alarms caused by scale variations in small-target pedestrian detection.

(3) The optimized image deraining algorithm was combined with the optimized pedestrian detection model, and comparative experiments were conducted to verify its effectiveness. The improved model was tested on the public dataset CrowdHuman and the self-built dataset Rain-CrowdHuman, validating its effectiveness and feasibility.

KEY WORDS: YOLOv10, pedestrian detection, adverse weather, crowded pedestrians, Transformer

目 录

第 1 章 绪论

1.1 课题研究背景及意义

伴随着全球城市化进程的迅猛推进与机动车保有量的持续攀升，现代交通系统正承受着前所未有的负荷。截止至 2025 年根据国际能源署（IEA）的最新预测，全球汽车保有量已突破 20 亿辆大关，尤其在发展中国家，机动车年均增长率维持在 5%至 7%的高位。与此同时，城市人口密度的不断提升进一步加剧了交通流量的复杂性，例如在中国，超过 60%的城市居民集中于人口稠密的城区。与这一趋势相伴的是交通安全问题的日益严峻。世界卫生组织（WHO）2023 年的统计报告指出，全球每年约有 135 万人因交通事故丧生，其中行人伤亡比例占 22%，而在低收入国家和地区，这一比例甚至接近 40%。复杂环境因素进一步放大了安全隐患，美国国家公路交通安全管理局（NHTSA）的数据揭示，夜间和雨天的事故发生率分别较白天和晴朗天气高出 50%和 30%，主要归因于能见度的显著下降以及行人识别难度的陡增[1]。

(a) 雾天　(b) 雨天

(c) 夜间　(d) 拥挤人群

图 1-1 复杂环境下车辆行人图像

智能交通系统[2]旨在通过技术创新优化交通流量、缓解拥堵并提升安全性。而自动驾驶汽车则依赖于对周围环境的精准感知，以确保在复杂城市场景中的安全导航。然而，复杂环境下的行人检测面临诸多技术瓶颈，例如：遮挡现象、光照条件的变化以及恶劣天气的干扰，这类问题均显著削弱了传统检测方法的可靠性[3]，如图 1-1 所示。

解决恶劣天气条件下的行人目标检测问题是提升自动驾驶安全性与交通管理效能的核心挑战。在雨雪、雾霾或沙尘等恶劣环境中，传统视觉传感器易受噪声干扰与激光雷达性能衰减，导致漏检或误检率激增，直接威胁自动驾驶系统的决策可靠性。突破该技术瓶颈不仅可挽救数万生命，更能推动全天候自动驾驶的商业化进程。随着智慧城市和新基建发展，该技术对应急救援、物流运输等场景具有战略价值。当前研究聚焦多模态传感器融合、物理模型驱动的数据增强、以及抗干扰深度学习算法，这些技术的突破将带动车载芯片、边缘计算等产业链升级，为构建零事故交通生态提供关键技术支撑。

从理论意义上看，本文聚焦于基于深度学习的雨天环境下行人检测算法，旨在突破现有目标检测技术在复杂场景下的局限性。在雨天环境下进行行人检测面临多维度挑战，这对特征提取和模型泛化能力提出了更高要求。通过优化深度学习模型结构或引入多模态融合策略，可以更深入地理解神经网络在雨天场景中的特征表达机制，从而为目标检测算法的鲁棒性与适应性提供新的理论依据。此外，探索图像去雨技术与行人检测的协同作用，将进一步拓展深度学习在多任务协同优化中的应用边界。从实践意义上看，本研究的成果在智能交通系统与自动驾驶领域具有广泛的应用前景。随着城市交通流量的激增和自动驾驶技术的快速发展，行人检测的精度与实时性直接关乎生命安全与交通效率。

1.2 国内外研究现状

1.2.1 行人检测研究现状

行人检测作为计算机视觉领域[4]的重要组成部分，经历了从传统手工特征方法到深度学习的深刻转型。随着自动驾驶和智慧交通等领域对行人检测需求的快速增长，国内外学者针对雨天环境下的行人检测问题展开了系统研究，逐步形成了多层次的技术突破。

（1）基于传统方法的行人检测

在深度学习普及之前，国内外行人检测研究主要依赖传统图像处理技术，这类传统检测方式主要衍生出两大技术分支：基于梯度通道特征的方法[5]与基于可变形部件模型的检测体系[6]，其中方向梯度直方图（Histogram of Oriented

Gradients，HOG）[7]结合支持向量机（Support Vector Machine，SVM）[8]成为当时的主流方法。DPM[9]最初由 P.Felzenszwalb 提出，随后 R.Girshick 对其进行了多项改进。作为 HOG 检测器的扩展，DPM 不仅继承了其在行人检测方面的优势，还突破了对单一独立目标的局限，进一步发展为一种基于弱监督学习的检测方法。罗森林等人[10]提出一种融合加权 KNN 与自适应牛顿法的鲁棒 Boosting 框架，通过动态权重调节机制：在强化误分类样本学习权重的同时，对高噪声先验概率样本施加抑制约束，从而平衡数据偏差与噪声干扰，实现分类器抗噪能力的系统性优化。Dalal 等[11]开发了基于 HOG 特征与 SVM 分类器的行人检测框架，通过方向梯度直方图捕捉轮廓特征，显著提升检测精度。早期的行人检测技术主要采用传统算法框架，无需依赖预训练模型即可实现目标识别。其典型处理流程包含三个核心环节：首先通过候选区域生成算法定位图像中的潜在目标位置；继而运用人工设计的特征描述子（如 HOG 等）进行手工特征提取；随后采用机器学习分类器（如 SVM 等）对特征向量进行判别分析并筛选出最优检测结果。

（2）基于深度学习的行人检测

随着时代的发展，人们对于行人检测的实时性与准确性需求日趋增长，但传统机器学习的目标检测方法在实时性和准确性之间难以平衡，这一局限性促使深度学习算法迅速崛起。通过卷积神经网络的自动特征学习和端到端的优化，深度学习在检测精度和速度上实现了显著突破。以 Fast R-CNN[12]等二阶段模型则通过区域建议网络进一步提升了检测精度，而以 YOLO（You Only Look Once）[13]为代表的一阶段检测器在保持较高精度的同时达到实时处理，推动了目标检测技术的实用化发展。

二阶段行人检测算法以其高精度和鲁棒性在复杂静态场景中表现出色。在发展初期，Girshick 等人提出了 R-CNN（Region-based Convolutional Neural Network）[14]，通过选择性搜索算法提取约 2000 个候选目标区域，使用 CNN 网络对每个候选区域进行特征编码，为二阶段检测器的发展奠定了基础。然而，R-CNN 的训练和检测效率较低。为此，Girshick 等人提出了 Fast R-CNN[15]，引入 ROI Pooling 池化层实现特征共享，取消了 SVM 分类器，大幅提升了训练和检测效率。尽管如此，Fast R-CNN 的区域生成效率仍有待提升。对于区域生成效率问题，Ren 等人提出了 Faster R-CNN[16]，通过区域建议网络（Region Proposal Network，RPN）

替代选择性搜索，实现端到端区域生成，显著提升了检测速度与精度。Zhang 等人提出了 Double Anchor R-CNN[17]，通过同时生成并匹配全身与头部检测框，在密集场景下提升了检测性能，相比于 Faster R-CNN，大大提升了区域矩阵计算效率。与此同时，Tian 等人提出了 DeepParts[18]，通过将人体分割为多个独立区域进行检测，并整合各区域的检测结果，有效提升了遮挡情况下的检测性能。然而，独立区域检测针对小目标检测存在漏检缺检的问题。为此，Song 等人提出了 RT-DETR[19]，提出多尺度融合模块与小目标增强结构，将底层细粒度特征集成到高层特征中，提高了小目标检测能力。但是，多尺度融合又带来了参数量剧增的问题。Liu 等人提出了改进的轻量级多尺度行人检测算法[20]，通过引入轻量级级联融合网络和注意力模块，改进了多尺度特征语义和位置信息的表征。

一阶段行人检测算法以其高效性和实时性成为近年来的研究热点。2016 年，Joseph Redmon 等人提出的 YOLOv1[21]将目标检测视为单次回归问题。与传统的二阶段检测方法不同，YOLOv1[22]通过基于全局图像上下文的建模，直接预测目标类别和边界框，实现了端到端的检测。然而，YOLOv1 在处理小目标和高度重叠目标时仍存在精度不足的问题。为了解决精度不足的问题，Redmon 等人提出了 YOLOv2[23]，通过引入 Anchor 机制、批量归一化、高分辨率分类器以及多尺度训练，显著提升了定位精度和召回率。之后，Redmon 等人提出了 YOLOv3[24]，采用 Darknet-53[25]残差网络增强特征提取能力，结合 FPN（特征金字塔网络）融合深浅层特征，提升了对小目标的检测能力。然而，YOLOv3 在处理密集场景中的遮挡问题时效率较低。在提升效率方面，Li 等人提出了 YOLO-CAN[26]，通过引入注意力机制、CIoU 损失函数、Soft-NMS 和深度可分离卷积等策略，提高了小目标和遮挡物体的检测精度与检测效率。Wang 等人提出了 TRC-YOLO[27]，优化了 YOLOv4-Tiny 的卷积核，并在网络残差模块中加入了额外的卷积层，成功构建了 CSPResNet 结构，提升了网络的训练效率。由于是串行卷积结构，ResNet 网络对于细节特征提取有一定局限性。针对 YOLOv8[28]在处理小尺度行人检测时存在的特征提取不足问题，Ang 等人提出了改进方案[29]。他们将可变形卷积应用于骨干网络，有效增强了模型对细节特征的提取能力，从而显著提升了小尺度行人的检测性能。在此基础上，Li 等人提出了 GR-YOLO[30]，同样对模块骨干网络进行优化，增强了特征提取能力，提升了密集场景检测的准确率。Deng 等人从

多尺度出发，提出了交叉融合模块[31]，通过将空间信息融合成多尺度特征，并对每个尺度的特征进行多次融合，以获得更多被遮挡的小尺度行人特征信息，改善了网络遮挡导致的特征丢失问题。

在行人检测发展的这些年来，YOLO 系列算法[32-36]经过了一次次快速的更新迭代。截止 2024 年，YOLO 系列算法已更新至 YOLOv10[37]，尽管其核心原理没变，但每个版本都在行人检测上有所提升，通过不同的网络改进，基于 YOLO 系列的行人检测算法在速度、精度和部署效率上达到了更好的平衡效果，更适合实时性要求高、资源受限的应用场景。

1.2.2 图像去雨技术研究现状

在自动驾驶、视频监控等依赖计算机视觉的领域中，雨天拍摄的图像常因雨水干扰产生严重降质。具体而言，雨水不仅会遮挡目标物体的纹理细节，还会形成半透明噪声层，导致图像对比度下降、边缘模糊等问题，严重干扰计算机视觉系统的正常工作。为此，图像去雨技术成为提升视觉系统可靠性的关键研究方向。当前主流方法主要分为两类：基于传统图像去雨的方法通过图像分解、滤波等方法去除雨纹；而深度学习方法则利用深度神经网络从海量数据中自动学习雨纹特征与背景的复杂映射关系，逐步实现更精准的雨痕分离。接下来在本章小节中详细分析基于传统方法与基础深度学习的研究现状，并分析其优势与劣势。

（1）基于传统方法的图像去雨算法研究

传统的去雨方法通常依赖于图像处理技术，通过分析雨天图像中的雨纹与背景的差异来进行去除。Kang[38]等人提出了一个基于单图像的去雨框架，该方法不直接采用传统的图像分解技术，而是首先使用双边滤波器将图像分解为低频段和高频频段。然后通过字典学习和稀疏编码将高频部分分解为“雨分量”和“非雨分量”通过对不同分量进行处理来恢复图像。但是高低频信息处理在面对复杂情况下容易出现损失图像细节的情况，针对此问题 Santhaseelan 等人[39]提出了一种基于相位一致性特征的雨水检测方法，通过分析视频帧间的特征动态变化，避免了直接分离图像信息，实现了雨天视频的雨水去除。而 Ding 等人[40]则从另一个角度出发，提出了一种基于 L0 梯度最小化的引导平滑滤波方法，通过类似于机器学习反向传播的方法，降低梯度，从而去除雨图中的雨条纹。Zhu 等人[41]利

用分层先验信息来区分雨区。尽管在某些场景中表现良好，但这些基于先验的方法依赖于特定假设，可能无法适用于雨纹复杂多样的真实场景，因此其泛化能力有限。针对此问题，Li 等人[42]利用高斯混合模型（GMM）作为先验，分别对雨纹和背景进行建模，实现去雨任务，并且提升了模型的泛化能力。

传统去雨方法在一定程度上能够处理简单的雨天场景，但在复杂雨天环境中表现不佳。由于雨纹与背景可能高度相似，去雨效果往往不理想；同时，雨纹形态、密度和方向的多样性进一步增加了处理难度。为了克服传统去雨方法在复杂雨天场景中的局限性，近年来，基于深度学习的图像去雨算法逐渐成为研究的热点。这些方法通过更先进的网络架构和训练策略，能够有效处理雨纹与背景高度相似、雨纹形态多样等问题，提高去雨效果。基于深度学习的方法不仅能处理简单的雨天场景，还能够应对复杂的环境，具有更强的泛化能力。

（2）基于深度学习的图像去雨算法研究

近年来，基于深度学习的图像去雨方法取得了显著进展，主要分为三大类：基于卷积神经网络的方法、基于生成对抗网络的方法以及基于 Transformer 的方法。这些方法各自通过不同的网络架构和优化策略，致力于解决图像去雨任务中的关键问题，为复杂环境下的视觉任务提供了重要技术支持。

基于卷积神经网络的图像去雨方法通过多层卷积操作有效提取局部特征，实现对雨纹的高效去除和图像细节的恢复。Li 等人[43]设计了一种两阶段神经网络，结合创新的条纹感知分解方法，自适应地将图像分离为高频成分和低频成分。然而，分离的方法容易丢失复杂图片的细节信息，同时增加网络参数量。针对这一问题，Ren 等人[44]提出了渐进式 ResNet（PRN）解决方案。该方法摒弃了传统的分离策略，转而采用重复展开浅层 ResNet[45]的创新架构。通过递归计算机制，PRN 在显著减少网络参数量的同时，保持了优异的学习性能。Qin 等人[46]提出了一个新颖的特征注意（FA）[47]模块，通过结合通道注意和像素注意机制，灵活地处理不同类型的特征和像素，显著提升了图像处理的能力。同样的，Hu 等人[48]提出了一种基于深度引导注意力机制的模型，通过提取深度注意力特征并生成残差图实现雨水去除。为进一步验证算法性能，他们还构建了新的数据集 RainCityscapes，为后续研究提供了重要数据支持。Kim 等人[49]则从另一个角度

出发，构建了一个多失真数据增强数据集，并提出了一种混合专家网络，利用多任务学习来恢复多重失真图像。

基于 Transformer 的去雨方法通常利用自注意力机制捕捉长距离依赖，通过编码器-解码器架构学习输入图像的雨滴模式，从而生成去雨后的清晰图像。Chen 等人[50]提出 DRSformer，其针对传统 Transformer 在注意力机制中易受无关信息干扰的问题，提出 top-k 稀疏注意力机制。该机制通过筛选并保留最相关的自注意力值，优化了特征聚合效果，值得注意的是模型在计算效率方面仍存在不足。Wang 等人[51]则在网络中引入一种新的局部增强窗口 Transformer 块，通过基于非重叠窗口的自注意机制，减少了特征图的计算复杂度。相比于 DRSformer，在提升了计算效率的同时增强了局部上下文信息的获取能力。Chen 等人[52]开发了一种端到端多尺度 Transformer，结合各种尺度的潜在有用特征，促进高质量的图像重建。该方法利用基于像素坐标的尺度内隐式神经表征与退化输入结合，增强了模型在复杂场景下的鲁棒性。Xiao 等人[53]设计了互补的基于窗口的 Transformer 和空间 Transformer，增强了局部性，并捕获了远程依赖关系。Gao 等人[54]则从结构出发，提出了混合层次结构，通过逐步恢复退化图像中的上下文信息和空间细节。

基于生成对抗网络的去雨方法是通过训练生成器生成无雨图像，判别器则用于判断图像是否真实，从而优化生成器去除雨滴。Wei 等人[55]将半监督迁移学习引入单图像去雨任务。具体而言，基于对残差分布的理解他们设计了一种参数化模型，并利用似然项准确刻画了干净图像与含雨图像之间的差异关系，显著提升了模型的泛化能力。在此基础上，Jin 等人[56]提出了一种无监督去雨生成对抗网络（UD-GAN），通过引入自监督约束，从未配对的雨天图像和干净图像中提取内在先验信息，实现去雨任务。Zhu 等人[57]提出了一种基于 GAN 的无监督去雨网络 RR-GAN，利用未配对图像生成无雨图像。该网络采用多尺度注意力生成器和深度监督判别器，性能接近监督方法。然而，其仅通过一致性损失约束雨天与无雨图像，导致约束较弱且不适定。

相比较于基于传统方法的去雨算法，基于深度学习方法有较好的去雨效果。但是深度学习方法需要依赖大量的训练数据和计算资源，使得其训练时间长，无法很好地适应不同的雨水类型和环境条件。

1.3 行人检测研究难点

尽管以卷积神经网络为代表的深度学习技术已经在目标检测领域取得了突破性进展，但行人检测在雨雾等复杂环境下的性能仍然存在显著不足。这些复杂环境带来的挑战主要体现在目标特征模糊、背景干扰增强以及光照条件变化等方面，导致现有检测方法的性能往往大幅下降。针对这些问题，本节将系统性地分析复杂环境下行人检测面临的主要技术难点，并探讨相应的解决思路。

恶劣天气对图像质量的破坏严重制约了行人检测性能，主要体现在两个方面：其一，雨滴在成像过程中引入的斑驳纹理会干扰局部图像结构；其二，视野能见度的急剧下降导致图像出现低对比度和色彩失真等问题。在如此严重的信息损失情况下，即使是最先进的检测器也难以准确提取行人的判别性特征，从而导致检测精度显著下降。此外，雨天行人样本的稀缺性进一步加剧了这一挑战：当前主流的行人检测数据集（如 CityPersons、CrowdHuman）主要包含晴天场景，而真实雨天行人样本的获取和标注成本高昂，这使得构建大规模雨天数据集面临现实困难。这种数据匮乏导致模型在少量雨天样本上训练时容易欠拟合，而直接使用晴天数据训练的模型在雨天场景中又存在明显的分布差异问题。因此，如何在数据有限的条件下有效克服雨天干扰、提升行人检测的环境适应性，已成为一个亟待解决的关键技术难题。

密集行人检测面临的主要技术难点集中在遮挡问题、多尺度变化以及复杂背景噪声等方面。在遮挡问题上，行人密集场景中普遍存在部分遮挡和完全重叠现象，这使得传统基于锚框的检测方法难以获取完整的视觉特征，模型必须依赖上下文推理或全局信息来补偿被遮挡区域的特征缺失。在多尺度变化方面，行人在图像中的尺度差异显著，导致特征提取困难且容易被背景噪声混淆，从而增加漏检风险。此外，复杂背景噪声的干扰进一步降低了图像质量，表现为行人与背景的对比度下降、边缘信息模糊，甚至引发误检问题。这些挑战共同要求行人检测算法不仅要具备强大的特征提取能力，还需要显著提升其鲁棒性，从而满足实际应用中对模型泛化性、实时性和抗干扰能力的严苛要求。

最后，实时性与计算效率也是影响行人检测实际部署的重要因素。在自动驾驶、辅助驾驶等场景下，行人检测不仅要做到精准，还需满足实时性的苛刻要求。然而，雨天和密集行人会大大增加图像的信息量，这就需要更大的计算量来完成

检测工作。在这种背景下，如何在有限硬件资源条件下实现检测精度与速度的最佳平衡，确保系统能够同时满足高性能和实时性的双重需求，已成为当前行人检测技术亟待解决的核心问题之一。

综上所述，雨天环境下的行人检测仍面临诸多亟待解决的技术挑战。具体而言，恶劣天气条件导致的图像质量退化、密集行人场景中的目标遮挡与尺度变化，以及检测精度与速度的平衡问题，都对现有检测方法提出了严峻考验。为应对这些挑战，需要在特征提取、模型架构设计以及计算效率优化等多个层面进行理论创新和技术突破，从而全面提升行人检测系统在复杂环境下的鲁棒性和实用性。

1.4 本文主要研究内容和结构安排

本文研究的主要内容是雨天环境下的行人检测方法，主要研究的场景是恶劣天气（雨天）和拥挤人群场景。雨天天气会导致图像对比度降低、噪声增强，而拥挤场景中行人密集遮挡、肢体重叠易引发误检或漏检。因此，针对现实场景中遇到的这些挑战，本文聚焦于图像去雨和行人检测技术的改进和优化，目的是为了提升去雨算法和行人检测算法在雨天+密集行人场景下的性能。本文的章节结构与研究内容安排如下：

第 1 章：绪论。本章节首先介绍了行人检测的研究背景与意义，然后分别阐述了行人检测算法和图像去雨算法的国内外研究现状与发展趋势，接着分析了雨天场景下行人检测的研究难点，最后简介了本文的整体架构。

第 2 章：图像去雨与行人检测算法相关理论与技术基础。本章首先详细介绍了卷积神经网络的结构与理论知识，然后对所用的 YOLOv10 基础目标检测网络框架做了介绍，最后对当前所使用的去雨算法网络架构进行总结概括。

第 3 章：基于动态直方图注意力机制与多特征聚合的图像去雨算法。本章首先简介了优化后的去雨网络模型的网络架构，然后分别对改进模块的作用以及原理进行了详细介绍。此外，对雨天行人数据集进行构建与解释，并通过在自建数据集上进行消融实验、对比实验验证了该模型的有效性与可行性。

第 4 章：改进的密集行人检测算法。本章节首先简介了改进后的网络整体架构，然后针对不同模块在不同方向上提出优化方案，并对改进的模块及其在原始模型的位置进行详细的说明，接着，用本章的行人检测算法集成第三章图像去雨

算法得到最终的集成算法。最后，通过消融以及与其他算法的对比实验来验证算法的有效性与可行性。

第 5 章：总结与展望。对本文的工作与研究成果进行总结，分析当前工作尚存在哪些问题与不足，并对后续的工作与未来研究方向进行可行性规划。

第 2 章 图像去雨与行人检测算法相关理论基础

在本章中，将会介绍雨天+密集行人这样的复杂环境下行人目标检测所涉及到的相关理论知识，以便能够更好地理解本文后续的研究内容。首先，介绍了卷积神经网络的结构与原理。然后，细致讲解了端对端的 YOLOv10 行人检测网络模型，包括其结构与损失函数。紧接着又对基于 U-net 和 Transformer 的去雨网络的结构进行分析。最后，构建本文所需要雨天+密集行人数据集，包括基础数据集的来源以及图像中雨纹噪声的引入方法等，方便后续实验的进行。

2.1 卷积神经网络

卷积神经网络（Convolutional Neural Network, CNN）[58]作为一种深度学习模型，在图像处理、视频分析及自然语言处理等领域展现出卓越的性能。其核心设计灵感源于人类视觉感知机制，通过卷积操作提取输入数据的局部特征。具体而言，CNN 的典型架构由卷积层、池化层和全连接层构成：卷积层利用卷积核（或滤波器）在局部感受野内提取特征；池化层则通过下采样操作降低数据维度，同时保留显著特征；全连接层最终将提取的特征映射至输出层，用于分类或回归任务。CNN 的显著优势在于其局部连接与权重共享机制，这不仅能够有效捕捉数据的空间层次结构信息，还大幅减少了模型参数量，从而提升了计算效率与泛化能力。

2.1.1 卷积层

卷积是卷积神经网络中的核心操作，其作用是通过卷积核从输入数据中提取局部特征。卷积核是一个尺寸较小的权重矩阵，它在输入数据上滑动并与数据的局部区域进行逐元素相乘并求和，从而生成一个新的特征图。在卷积操作中，步长决定了卷积核在输入数据上滑动的步幅：当步长为 1 时，卷积核每次移动一个像素；当步长大于 1 时，卷积核根据步长的大小进行跳跃性地移动，这会导致输出特征图的尺寸减小。填充则是在输入数据的边缘添加额外的像素，其目的是确保卷积核能够覆盖到输入数据的边缘区域，或者使输出特征图的尺寸与输入数据

保持一致。当图像输入特征图的尺寸为$H \times W$，卷积核大小为$C \times C$，步长为I，零填充为O，则输入特征图经过卷积运算得到输出特征图大小的计算公式如（2-1）和（2-2）所示：

$$H' = \left\lfloor \frac{H - C + 20}{I} \right\rfloor + 1 \tag{2-1}$$

$$W' = \left\lfloor \frac{W - C + 20}{I} \right\rfloor + 1 \tag{2-2}$$

式中，$\lfloor \bullet \rfloor$表示向下取整。另外，我们假设输入图像为 5×5，卷积核的大小为 3×3，如图 2-1 所展示卷积运算过程。

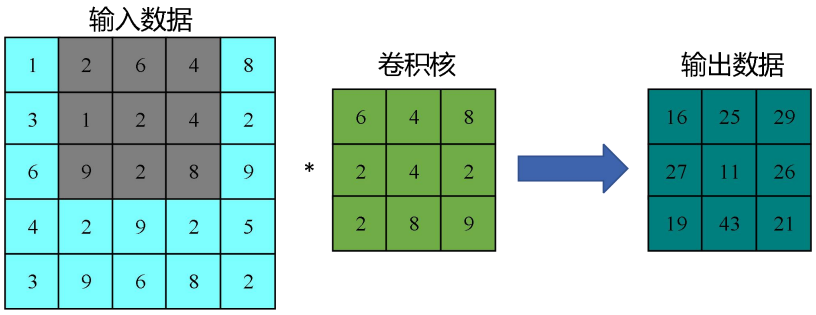


图 2-1 卷积运算过程

2.1.2 激活函数

激活函数作用于每一层神经元的输出，其核心作用是为网络引入非线性能力，从而使其能够学习和表达复杂的模式与关系。如果没有激活函数，无论神经网络有多少层，都只能执行线性变换，无法拟合复杂的数据分布，这将极大地限制模型的表达能力。常见的激活函数包括 Sigmoid[59]、Tanh[60]、ReLU[61]和 Leaky ReLU[62]，如图 2-2 所示。它们各自具有不同的特性，适用于不同的场景和需求。

（1）Sigmoid 激活函数

Sigmoid 函数是一种常用的激活函数，它将输入的实数值映射到 0 到 1 之间，这种特性使其特别适合用于二分类问题，因为其输出可以被直接解释为概率值，便于表示二元决策。同时，Sigmoid 函数通过引入非线性，使神经网络能够学习复杂的非线性关系，其平滑的特性也使其在梯度计算和反向传播中表现良好。然而，Sigmoid 函数存在一个显著缺点：当输入值过大或过小时，其梯度会趋近于零，导致训练过程中学习效率降低甚至停滞，这种现象被称为梯度消失问题。

Sigmoid 激活函数计算公式如式（2-3）所示：

$$\vartheta(x)=\frac{1}{1+e^{-x}} \tag{2-3}$$

（2）ReLu 激活函数

ReLU（Rectified Linear Unit，修正线性单元）工作原理为：当输入值大于零时，它直接返回输入值；当输入值小于或等于零时，它输出零。这种高效的规则使得 ReLU 激活函数在处理大规模数据时具有高效的计算性能。其最大优势在于有效缓解了深层神经网络中的梯度消失问题，而传统激活函数在反向传播时易产生梯度接近零的现象，导致权重更新缓慢。然而，ReLU 也存在神经元"死亡"问题，即当神经元输入持续为负时，其输出始终为零，导致权重无法更新，影响模型性能。

Relu 激活函数计算公式如式（2-4）所示：

$$\mathrm{Re}\,lu(x)=\max(0,x) \tag{2-4}$$

（3）Leaky Relu 激活函数

Leaky Relu（Leaky Rectified Linear Unit）是 Relu 激活函数的一种变种，旨在解决 Relu 的“死神经元”问题。而 Leaky Relu 通过在负输入区间引入一个小的负斜率，使得输入值小于零时输出一个小的负值，而不是零。这种调整避免了神经元完全失活，从而提高了网络的学习能力，并且有助于加速深度神经网络的训练过程。

Leaky Relu 激活函数计算公式如式（2-5）所示：

$$LeakReLU=\begin{cases}x, & if x>0\\ ax, & if x\le 0\end{cases} \tag{2-5}$$

（4）Tanh 激活函数

Tanh 函数作为激活函数具有输出对称分布于零值附近的特性，这一特性相较于 Sigmoid 函数更有利于模型收敛。具体而言，Tanh 函数将输入值映射到[-1, 1]区间，这种对称性特征在处理零均值分布的数据时表现出显著优势，可有效避免数据分布偏差对模型训练的影响。然而与 Sigmoid 函数类似，Tanh 函数在输入值趋近于正负无穷时其梯度会急剧减小，导致梯度消失问题从而影响深层神经网络的训练效率。因此虽然 Tanh 函数在隐藏层中具有一定优势，但其仍存在梯度消失的固有缺陷。

Tanh 激活函数计算公式如式（2-6）所示：

$$\mathrm{Tanh}(x)=\frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \tag{2-6}$$

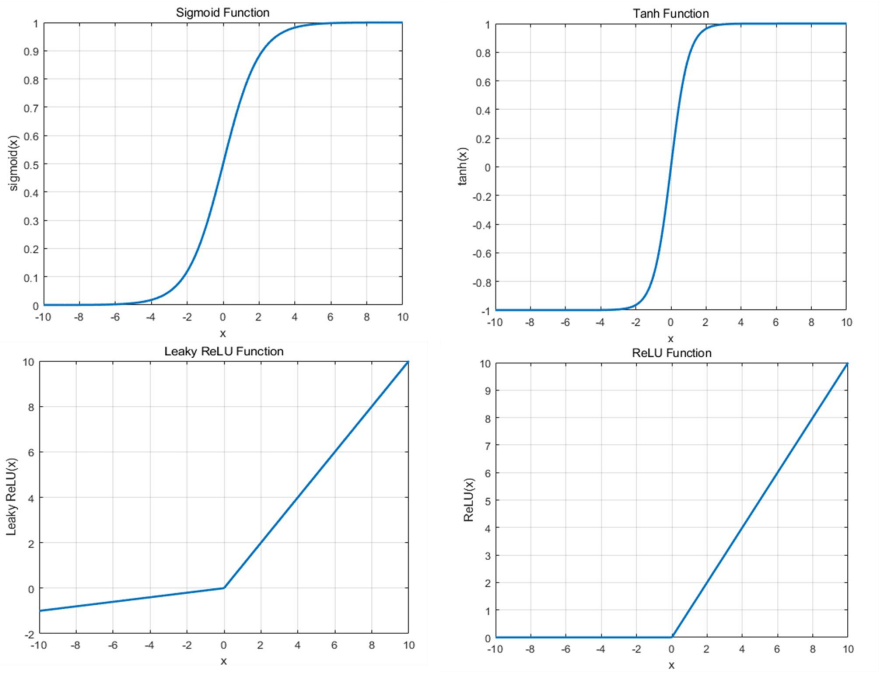


图 2-2 四种不同的激活函数

2.1.3 池化层

池化层（Pooling Layer）[63]通过下采样操作降低特征图的空间维度，这一过程在减少计算量和参数量的同时保留了重要特征信息。其独特的特性使网络对输入数据的平移、缩放和旋转等变换具有较强的适应能力，从而显著提升模型的泛化性能并有效防止过拟合。池化操作常见的方式有最大池化和平均池化：

（1）最大池化通过提取局部区域的最大值作为输出，能够有效保留特征图中的显著信息。在图像处理中，这一操作能够捕捉边缘和纹理等关键特征，从而增强网络的特征提取能力，最大池化运算过程如图 2-3 所示。

（2）平均池化则是对池化窗口内的所有数值取平均值作为输出，与最大池化不同，平均池化倾向于保留输入区域的整体特征，而不是突出局部最显著的特征。它常用于需要平滑数据的任务中，能有效地减少噪声和细节，保留数据的整体信息，平均池化运算过程如图 2-4 所示。

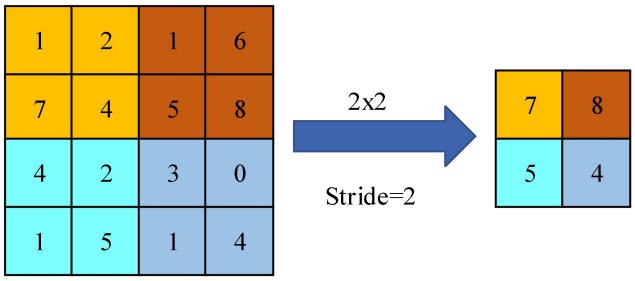


图 2-3 最大池化运算过程

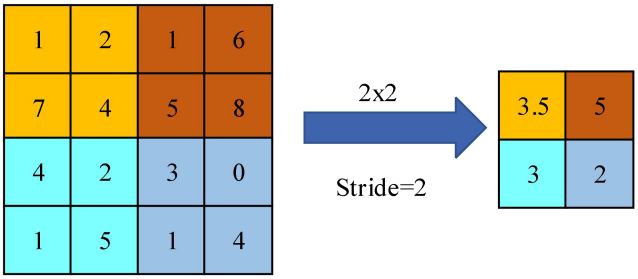


图 2-4 平均池化运算过程

2.1.4 全连接层

全连接层（Fully Connected Layer）[64]作为神经网络的典型结构，通常应用在卷积神经网络的末端。其主要功能是通过密集连接整合前层提取的特征，将多维特征映射到输出空间。在分类任务中，全连接层将展平后的特征向量转换为最终的预测输出。尽管其密集连接结构导致参数量较大，但全连接层能够有效学习复杂的非线性关系，在特征综合与决策输出中发挥关键作用，全连接层结构如图 2-5 所示。

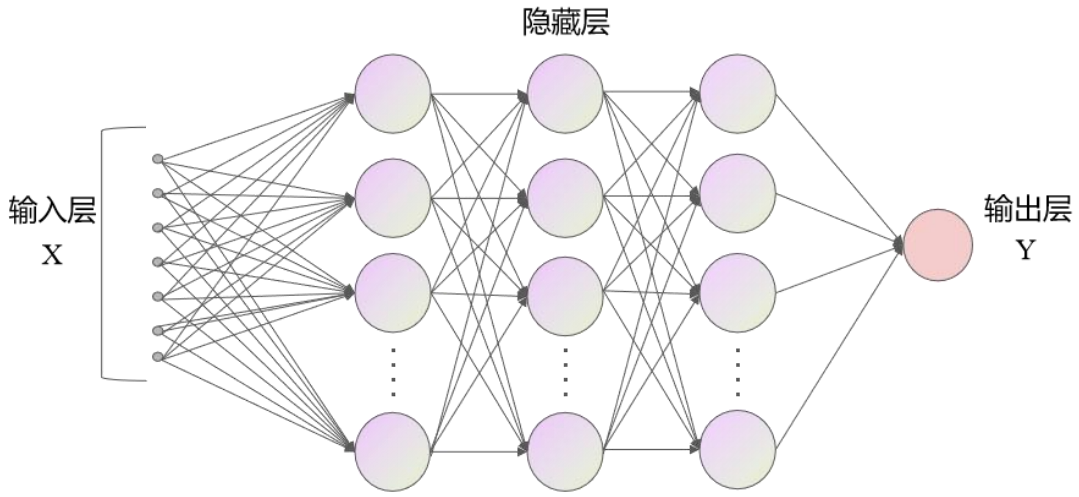


图 2-5 全连接层

2.2 YOLOv10 模型理论基础

在目标检测领域，YOLO 算法无疑是具有里程碑意义的。它开创性地提出了一种全新的检测思路，即通过构建一个单一的卷积神经网络，直接对图像进行处理，从而一次性预测出目标所属的类别以及其边界框的坐标信息。这种端到端的检测方式，不仅简化了目标检测的流程，还极大地提升了检测的效率，为后续的目标检测技术发展开辟了新的道路。

YOLOv10 的理论基础则是建立在无 NMS 训练策略上，致力于解决传统 YOLO 模型对非极大值抑制（Non-Maximum Suppression，NMS）后处理的依赖问题，通过一致的双重分配结合一对一和一对多标签分配，消除了模型在推理过程中的非最大抑制（NMS）的使用。通过一致性匹配度量协调两个分支的监督信号，结合分类分数和 IoU 评估预测与真实值的匹配，提高模型推理的准确度。

2.2.1 YOLOv10 的网络架构

YOLOV10 延续了 YOLOv8 的网络结构来进行搭建，对 YOLOv8 中的一些模块在细节上进行了改进。同样的，YOLOv10 的网络结构由骨干网络（Backbone）、颈部网络(Neck)和检测头(Head)三部分组成，如图 2-6 所示为 YOLOv10 的网络结构示意图。

YOLOv10 的骨干网络基于增强 CSPNet（Cross Stage Partial Network）[65]，通过分割特征图减少冗余，提升信息流动。它采用紧凑倒置块（Compact Inverted Bottleneck，CIB）[66]进行空间和通道混合，并嵌入高效层聚合网络（Extended-ELAN，ELAN）。此外，YOLOv10 使用大核卷积（7×7）扩大感受野，增强全局信息捕捉能力，提升小目标识别，并通过结构重参数化优化设计，不增加推理计算负担。

YOLOv10 的颈部网络总体上采用了 PANet（Path Aggregation Network）[67]结构，连接着 Backbone 和 Head，起到多尺度特征融合的作用，其中 YOLOv10 的颈部网络对原有的结构进行了轻微的修改，使用 SCDown 代替原来的下采样模块，在更深的网络当中，使用 C2fCIB 来代替了 C2f 模块，以此来提高模型的特征融合能力。

YOLOv10 检测头采用双检测头架构，训练时结合一对一和一对多分配策略，推理时仅用一对一分配，避免 NMS。一对多分配为模型提供丰富监督信号，加速回归收敛；一对一分配避免冗余，从而提升推理速度。

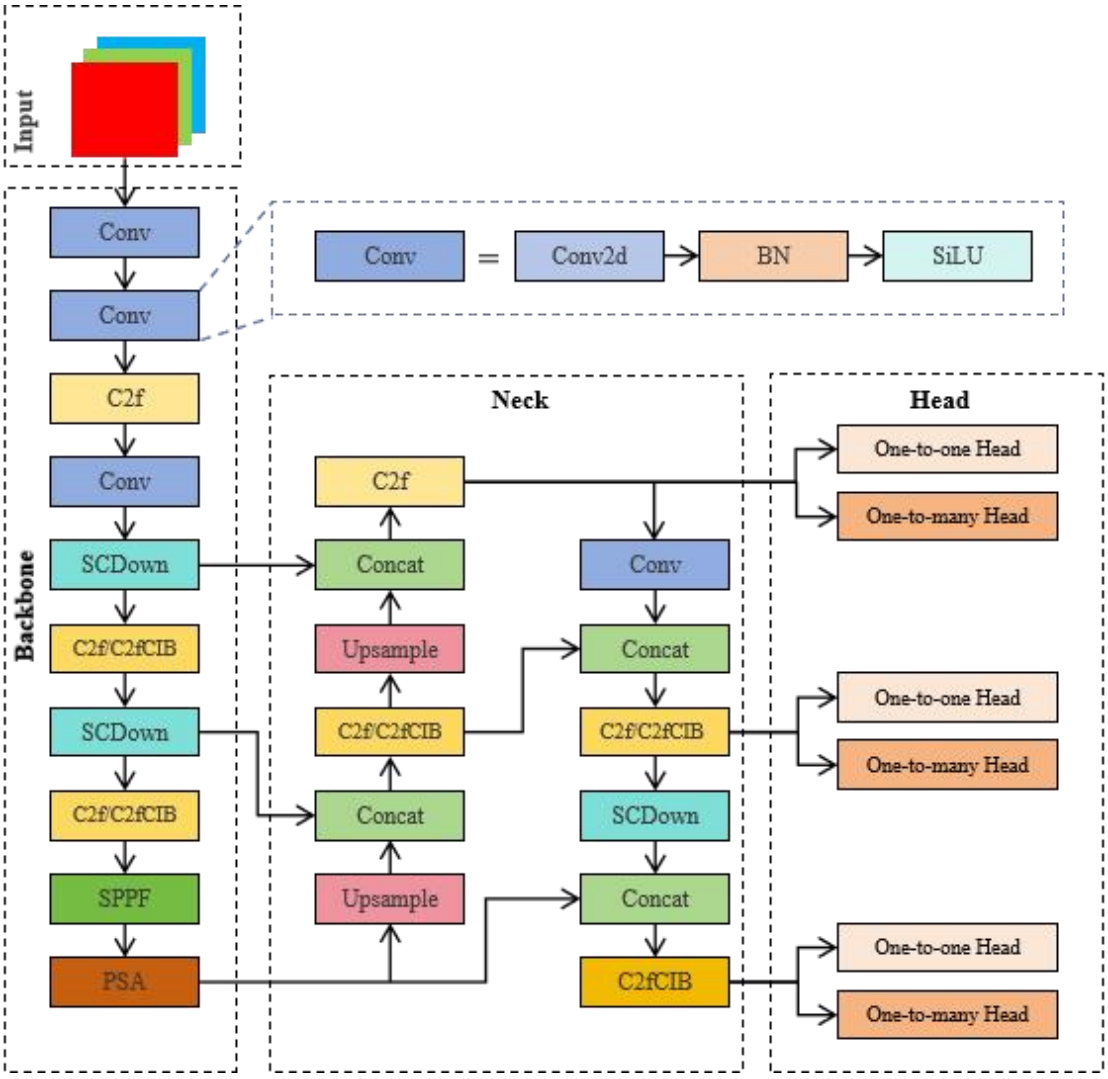


图 2-6 YOLOv10 网络结构示意图

在 YOLOv10 较大尺寸的模型网络中，C2fCIB 层就是使用紧凑的倒置块（Compact Inverted Block, CIB）替换了原本 C2f 模块中的 Bottleneck，CIB 模块将原本 Bottleneck 中的标准卷积用深度卷积加逐点卷积进行替换，提高了计算效率。

YOLOv10 中的下采样模块由原来单个的 CBL（Convolution BatchNorm LeakyReLU）模块替换为了 SCDown（Spatial-channel decoupled downsampling）模块。SCDown 在空间和通道上进行解耦。先通过 1×1 的逐点卷积来调节通道数，再通过 3×3 的深度卷积做空间下采样，在降低计算开销的同时最大限度保留特征

图中的信息。

SPPF 结构是具有残差结构的连续三次最大池化，卷积核统一为 5×5，最后将池化前和每次池化后的结果拼接，结合了每一层的输出，保证了多尺度融合的同时，降低了计算量的同时进一步增大了感受野。

YOLOv10 在 SPPF 层后面增加了 PSA（Partial Self-Attention）模块，是一种高效的自注意力模块。PSA 模块的设计核心在于它将卷积得到的特征图分为两部分进行处理。一部分特征被输入到多头自注意力模块（MHSA）和前馈网络（FFN）构成的子模块中，增强对全区上下文信息的理解。另一部分特征则保留了局部的空间信息。然后将这两部分特征进行连接在通过卷积操作进行融合，提高的全局建模的能力。YOLOv10 中各模块结构详细示意图如图 2-7 所示。

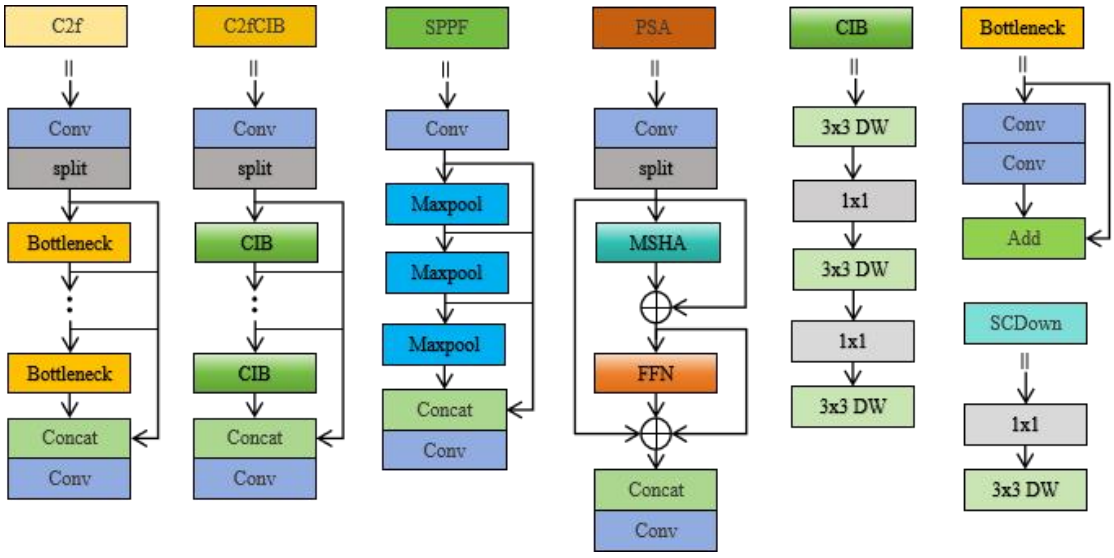


图 2-7 各模块结构示意图

2.2.2 YOLOv10 的损失函数

在 YOLOv10 中，损失函数对于算法性能的提升起着至关重要的作用，其损失函数通常是由多种损失函数加权求和得到的，包含分类和两种坐标损失函数。这三种损失函数协同工作，共同对模型在目标检测和定位方面进行优化，引导模型更好地学习目标检测任务。

分类损失主要用于优化模型对目标类别的区分和预测能力。具体而言，它通过计算模型对各类别的预测值与真实值之间的差异，促使模型提升对不同目标类别的区分能力。在常规的单标签分类任务中，模型输出的各个类别是互斥的，模

型最后将概率最高的类别作为最终的预测结果输出。分类损失的计算公式如式（2-7）所示：

$$CELoss(y,\hat{y}) = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \cdot \log(p_i) + (1-y_i)\cdot \log(1-p_i)\right] \tag{2-7}$$

式中，y_i表示真实标签，p_i为模型预测的概率。为了解决多标签分类问题，YOLOv10 对每个类别的预测结果采用了二元交叉熵损失（BCE Loss）进行训练。首先将数据带入 Sigmod，再借助二元交叉熵损失进行下一步计算，这能确保保证每个类别的预测概率处于 0 到 1 之间，并且可以有效处理多标签分类的情形。BCE Loss 计算公式如式（2-8）和（2-9）所示：

$$y_i = Sig\,\mathrm{mod}(x_i) = \frac{1}{1+e^{-x_i}} \tag{2-8}$$

$$BCELoss(y,\hat{y}) = -\sum_{i=1}^{N}\left[y_i \cdot \log(\hat{y}_i) + (1-y_i)\cdot \log(1-\hat{y}_i)\right] \tag{2-9}$$

坐标损失又称为定位损失，主要对模型预测的目标位置、大小进行优化，在 YOLOv10 中使用了 CIoU Loss 和 DFL。CIoU Loss 计算了正样本的定位损失，DFL 则计算了 anchor point 的中心点到左上角和右下角的偏移量。

交并比（IoU）是目标检测任务中的一个重要概念，即为预测框与真实框分别取交集和并集后这两者的比值，以此可以引导目标检测模型预测框向真实框靠近，最初的坐标损失如式（2-10）所示：

$$IoULoss = 1 - IoU(B,B_{gt}) = 1 - \frac{\left|B \cap B_{gt}\right|}{\left|B \cup B_{gt}\right|} \tag{2-10}$$

式中，IoU为交并比，B为样本。但是，当真实框和预测框完全不重叠时，它们的 IoULoss 恒为 1，无法准确反应预测框与真实框之间的距离，这将使得模型训练进入瓶颈；与此同时在出现预测框被真实框包含时，无论预测框在真实框中位置如何，IoULoss 的结果都将保持不变。

而 YOLOv10 使用的 CIoULoss 则比 IoULoss 更全面地考虑了以上两个问题，增加了预测框与真实框中心点间的欧式距离ρ、预测框与真实框的最小闭包区域的对角线距离c以及纵横比影响因子αv，使得模型更全面地学习目标框的特征，有助于提高模型在复杂场景中的性能。CIoULoss 如式（2-11）所示：

$$CIoULoss(B,B_{gt}) = 1 - IoU(B,B_{gt}) + \frac{\rho^2(b,b^{gt})}{c^2} + \alpha v \tag{2-11}$$

式中，α 是权衡长宽比造成的损失和 IoU 部分造成的损失的平衡因子，v 是用来衡量长宽比一致性的参数。

2.3 图像去雨算法网络架构

2.3.1 U-Net 网络架构

U-Net 是由德国弗赖堡大学的 Olaf Ronneberger 等人于 2015 年提出的一种深度学习网络架构，其独特的 U 形结构由编码器、瓶颈层和解码器三部分构成，并通过跳跃连接实现特征信息的有效传递。编码器部分通过卷积、ReLU 激活和最大池化操作逐步提取图像特征并降低空间分辨率。随着层数增加，特征图深度逐步增加，从而捕捉到更复杂的高层次特征。瓶颈层位于编码器和解码器之间，负责进一步提取图像的全局上下文信息，为分割任务提供深层的语义支持。解码器部分通过转置卷积逐步恢复图像的空间分辨率，并结合跳跃连接将编码器的低层特征与解码器的高层特征拼接，有效保留图像细节。最后，输出层通过 1x1 卷积将特征映射到分类结果。U-Net 通过这种结构在降低空间分辨率的同时提取高级特征，并在恢复分辨率的过程中结合低层次细节，成为深度学习领域的重要基础架构。U-Net 的网络结构图如图 2-8 所示[68]。

在图像去雨任务中，U-Net 凭借其独特的编码器-解码器架构和跳跃连接机制，成为解决雨水干扰问题的有效方案。其核心优势在于能够同时处理全局特征和局部细节：编码器部分通过卷积和池化操作逐步提取图像的高级特征，捕捉雨水干扰的全局模式；解码器部分则通过反卷积操作逐步恢复图像的空间分辨率。跳跃连接机制在这一过程中起到关键作用，它将编码器提取的低级细节信息直接传递到解码器，使网络在去除雨水的同时能够保留图像的精细结构和纹理信息。这种设计使 U-Net 在处理雨水造成的模糊和噪声时表现出色，能够生成更自然、更清晰的去雨图像。此外，U-Net 的对称结构使其在训练过程中能够高效收敛，而多尺度特征融合能力进一步提升了其去雨性能。因此，U-Net 不仅能够有效应对复杂的雨水干扰模式，还能保持背景图像的完整性，成为图像去雨任务中兼具高效

性和可靠性的解决方案。

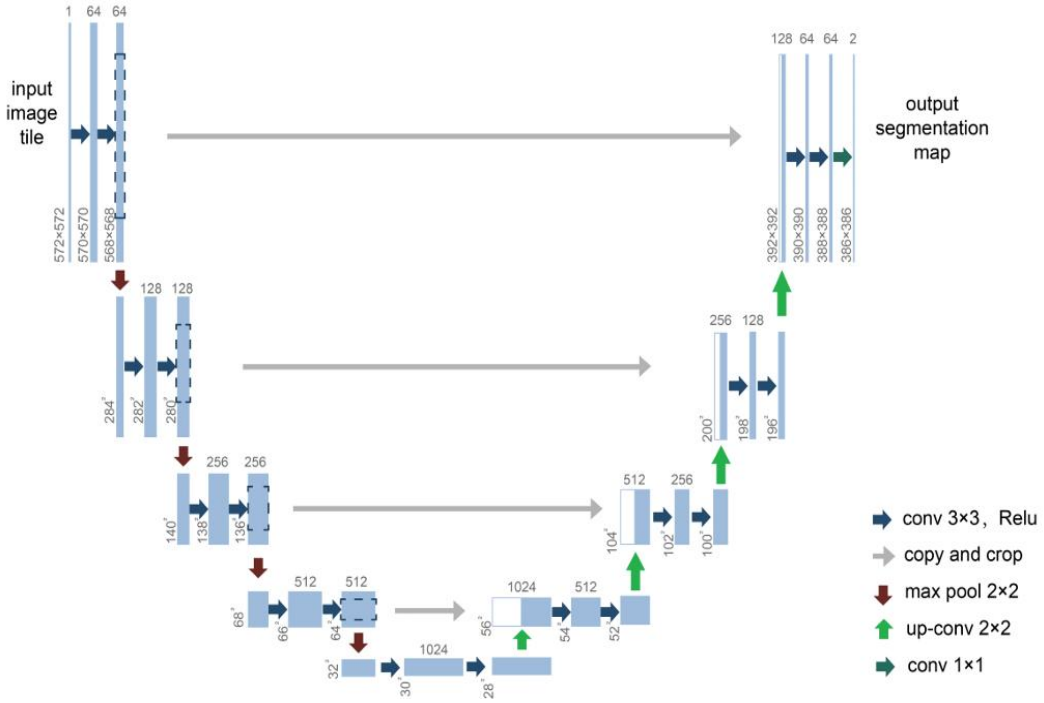


图 2-8 U-Net 网络架构

2.3.2 Transformer 网络架构

Transformer 由 Vaswani[69]等人于 2017 年提出，是一种基于自注意力机制的深度学习架构，其在自然语言处理（Natural Language Processing，NLP）领域取得了显著突破，它的结构如图 2-9 所示，其核心设计摒弃了传统的循环神经网络（RNN）和长短时记忆网络（Long Short-Term Memory，LSTM）序列处理方式，采用并行计算处理整个输入序列，显著提升了训练效率和性能。Transformer 由编码器和解码器两部分组成：编码器通过多层堆叠的自注意力机制和前馈神经网络将输入序列转化为高维特征表示；解码器则在此基础上增加了编码器-解码器注意力机制，用于生成输出序列。为解决并行计算中的顺序信息丢失问题，Transformer 引入了位置编码。这种设计使其在处理长序列时表现出色，有效避免了传统序列模型的长距离依赖问题。

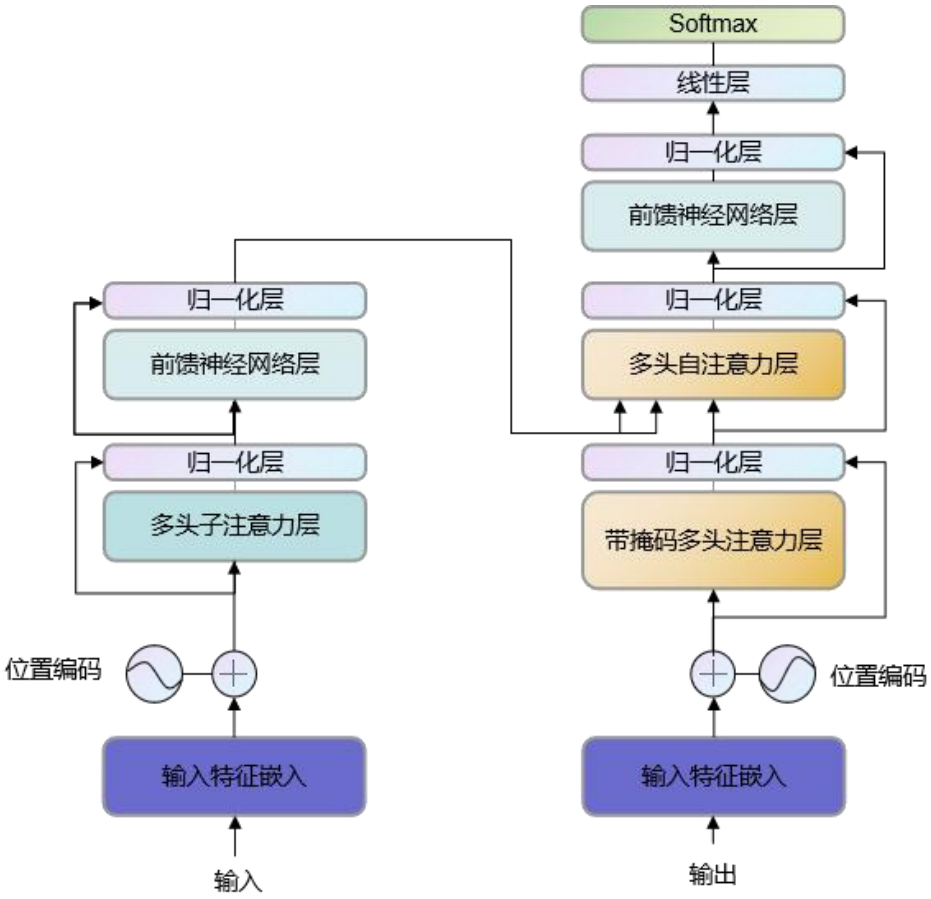


图 2-9 Transformer 结构

自注意力机制（Self-Attention）是一种通过计算输入序列中各个元素之间的关系来动态生成每个元素表示的机制，它的网络结构如图 2-10 所示。具体来说，输入的每个元素会被转换为三个向量：查询（Query）、键（Key）和值（Value）。查询向量用于衡量当前元素与其他元素的相关性，键向量表示其他元素的特征，而值向量则包含元素的具体信息。通过计算查询向量与所有键向量之间的相似度，得到一组注意力权重，这些权重反映了当前元素与序列中其他元素的关联程度。接着，利用这些权重对值向量进行加权求和，生成当前元素新的表示。这种机制使得模型能够同时关注序列中的多个位置，捕捉长距离依赖关系，无论这些位置在序列中的距离有多远。与 RNN 和 LSTM 相比，自注意力机制具有更高的并行计算能力，能够更高效地处理长序列，并避免了信息丢失或梯度消失的问题。自注意力机制的计算公式如式（2-12）所示：

$$ATT(Q,K,V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2-12}$$

式中，Q,K,V 分别代表查询向量、键向量与值向量， *Softmax* 代表 Softmax

激活函数，d_k为键向量的维度。

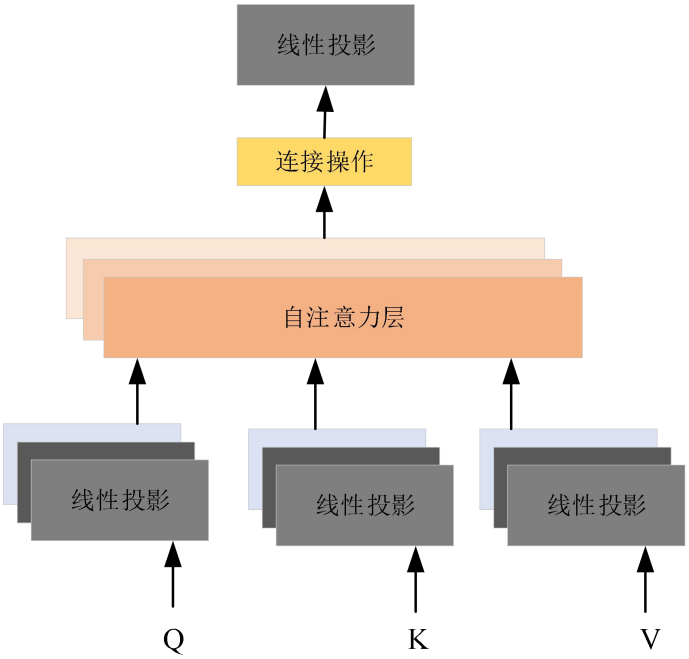


图 2-10 自注意力层

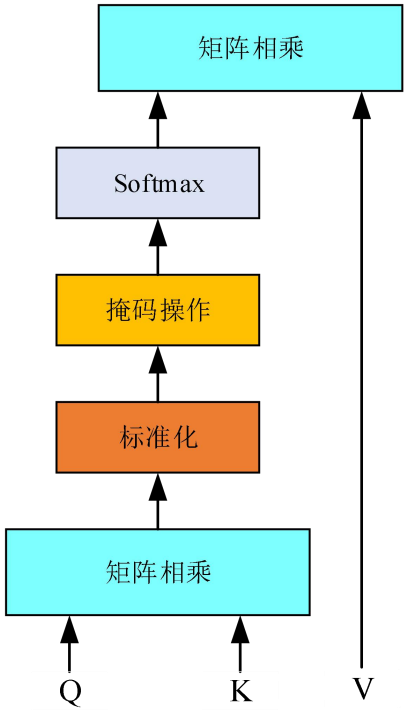


图 2-11 多头自注意力

多头注意力机制（Multi-Head Attention）是 Transformer 中的关键组件，它通过并行计算多个自注意力头来增强模型的表示能力，其结构如图 2-11 所示。在多头注意力中，输入的查询、键和值分别通过不同的线性变换生成多个子空间中的查询、键和值，每个注意力头独立地计算注意力输出，捕捉序列中不同部分

的特征。然后，所有头的输出会被连接起来，并通过一个线性变换得到最终结果。通过这种方式，多头注意力机制能够从多个角度关注序列中的不同依赖关系，提升模型在处理复杂信息时的灵活性和表达能力。多头注意力机制计算公式如式（2-13）和（2-14）所示：

$$MH(Q,K,V) = Conact(H_1, H_2, \cdots, H_n)W^o \tag{2-13}$$

$$H_j = ATT(Q_j, K_j, V_j) \tag{2-14}$$

式中，$MH(Q,K,V)$ 为多头注意力机制输出的结果，$Conact$ 代表拼接操作，W^o 为权重矩阵。

2.4 本章小结

本章系统介绍了深度卷积神经网络，对其卷积层、池化层、激活层和全连接层等核心模块的结构及原理做了详细介绍，以便后续实验工作的开展。然后对所要用到的 YOLOv10 网络进行了细致分析，主要介绍了其网络结构与损失函数。最后对基于 U-net 和 Transformer 的去雨算法网络的结构进行细致的分析。

第 3 章 基于动态直方图注意力机制与多特征聚合的图像去雨算法

在智能驾驶系统中，行人检测是一项至关重要的任务。然而，在雨天等恶劣天气条件下，图像质量往往因雨滴、雨线等噪声干扰而显著下降，这直接影响了目标检测算法的准确性。为了提升雨天环境下的行人检测精度，本章提出了一种基于动态直方图注意力机制和多特征聚合的图像去雨算法。该算法通过结合了多尺度特征聚合模块（Multispectral Filter Array，MSFA）、动态直方图注意力模块（Dynamic Histogram Adjustment Technique，DHAT）以及全局空间感知模块（Graph Signal Processing，GSP），不仅能够有效去除图像中的雨线噪声，还能较好地保留图像的细节信息，从而为后续行人检测任务提供更高质量的输入图像。

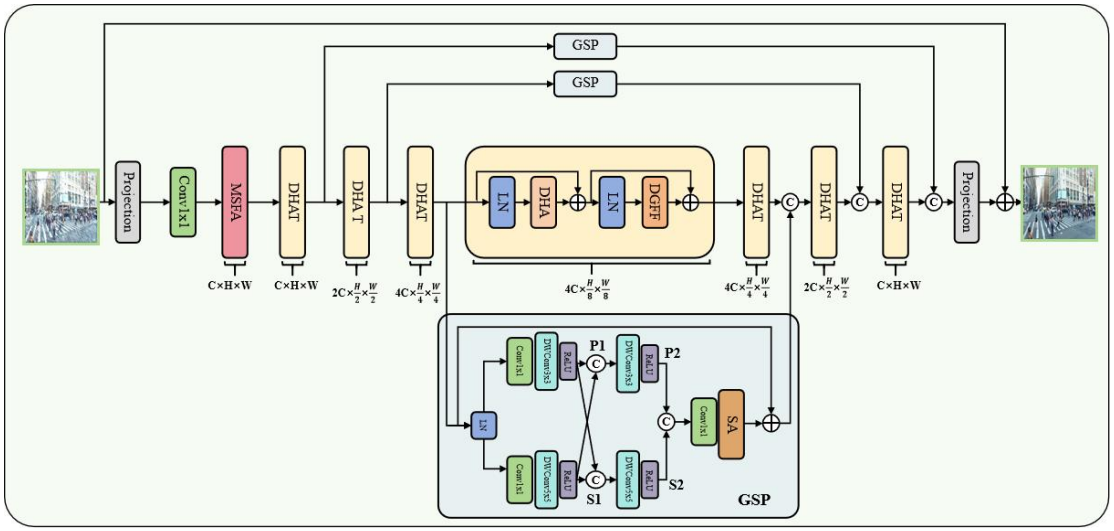


图 3-1 基于动态直方图注意力机制和多特征聚合的图像去雨网络结构

本章提出的去雨算法基于 U-Net 的编码器-解码器结构，并在此基础上引入了一系列模块以提升去雨效果，改进的算法结构如图 3-1 所示。在编码器部分，加入了多特征聚合模块（MSFA），通过多尺度卷积从不同尺度提取图像特征，从而更好地捕捉雨线的细微结构和复杂背景信息。此外，在编码器的跳跃连接中嵌入了全局空间感知模块（GSP），通过深度可分离卷积和空间注意力机制，增强特征之间的空间关系感知能力，突出雨线区域的特征并抑制背景噪声。在编码器和解码器采样部分，引入了动态直方图注意力模块（DHAT），通过动态直方

图注意力机制捕捉特征之间的复杂关联并进行特征变换，有效去除雨线噪声的同时恢复图像的细节信息。随后，通过与现有去雨算法的对比实验，本章验证了所提算法在图像质量改善方面的显著优势。实验结果表明，该去雨算法有效去除雨线噪声，同时保留图像细节和视觉可见性。去雨算法结合目标检测模型后，检测算法在恶劣天气下的行人检测任务中显著提升了检测精度，验证了算法在恶劣天气条件下的实用性和可靠性。下面小节将详细介绍各个模块与内容。

3.1 多特征聚合模块（MSFA）

在雨天场景中，雨滴常常与背景产生较强的对比，这种对比可能在不同的雨滴形态、大小、分布以及光照、物体轮廓等因素的影响下有所变化。传统的单一特征提取方法很难适应这些变化，因为不同类型的雨滴可能具有不同的纹理特征，而背景中的复杂元素（如建筑物、道路等）也会影响雨滴与背景的分离。因此，单一尺度的特征提取无法全面捕捉这些信息，容易导致去雨效果不理想。

为了解决这一问题，本文在网络开头部分引入一个多特征聚合模块（MSFA）[70]提取不同尺度的特征信息，其结构如图 3-2 所示。该模块具有三个分支，其中两个特征图 F_1 与 F_2 是含有不同卷积核尺度的特征信息，能够从多个尺度上捕捉到不同雨滴大小、形态和分布特征的细微差异。而另一个为输入特征图 $F \in R^{C*H*W}$，确保模型能够同时考虑全局和局部特征，避免在处理细节时丢失重要信息。通过不同尺度下的注意力权重加权不同区域的特征，从而增强模型面对不同去雨图像内容时的鲁棒性和泛化能力。

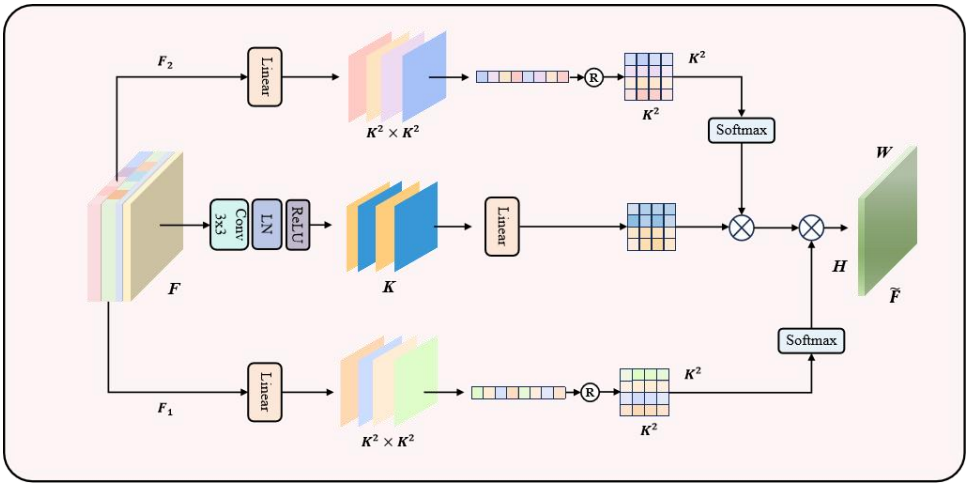


图 3-2 多特征聚合模块结构

该网络中 F_1 与分别是不同尺度的分支，通过较大的卷积核来增加感受野以此捕捉更广泛的空间特征。其中，在每个空间位置 $F(m,n)$，MSFA 通过空间特征来计算 (m,n) 为中心的 $K\times K$ 的注意力权重。具体来说，对于给定输入特征 $F\in R^{C*H*W}$ 依次通过卷积层，归一化层，激活层进行融合。其次，通过线性层权重将 $W_v\in R^{C*c}$ 映射到值向量 $V\in R^{H*W*C}$。值向量在每个相邻窗口进行展开，聚合每个位置的邻域信息。$V_{m,n}\in F^{C*K^2}$ 表示以 (m,n) 为中心局部窗口的所有值，中心局部窗口计算公式如式（3-1）所示。

$$V_{m,n}=\left[V_{m+o-\left|\frac{K}{2}\right|,n+g-\left|\frac{K}{2}\right|}\right],\quad 0\le o,g\le K \tag{3-1}$$

其次，将两个不同尺度的分支通过不同的线性层生成权重图，生成权重图计算公式如式（3-2）所示：

$$Att_{F1}=W_{F1}*F_1,\quad Att_{F2}=W_{F2}*F_2 \tag{3-2}$$

俩分支特征图的权重形状为 $W_{F_1},W_{F_2}\in R^{C*K^4}$，注意力权重为 $Att_{F_1},Att_{F_2}\in R^{H*W*K^2}$。在位置 (m,n) 处两者的权重重塑为 $\tilde{A}tt\in\mathbb{R}^{K^2\times K^2}$，使其经过 $Soft\max$ 激活函数。最后，对展开的 $V_{m,n}$ 向量进行加权，对加权值进行融合得到最终特征融合图。生成最终特征融合图计算公式如式（3-3）和（3-4）所示：

$$\tilde{V}_{m,n}=\mathrm{Softmax}(\tilde{A}\mathrm{tt}F_{1m,n})\otimes\left(\mathrm{Softmax}(\tilde{A}\mathrm{tt}F_{2m,n})\otimes V_{m,n}\right) \tag{3-3}$$

$$\tilde{F}_{m,n}=\sum_{0\le p,q<d}\tilde{V}_{m+p-\left|\frac{K}{2}\right|,n+q-\left|\frac{K}{2}\right|} \tag{3-4}$$

式中，$\otimes$ 表示矩阵乘法。通过这种方法在去雨任务中能够精细地识别和恢复被雨滴覆盖的区域，去除雨水对图像的干扰，同时保留图像的细节和结构。

3.2 动态直方图注意力模块（DHAT）

传统 Transformer 能够有效捕捉长距离的依赖关系，但在处理细小雨滴与复杂背景之间的微小差异时，其局部细节处理能力较弱，这导致它难以精确区分雨滴与背景。此外，Transformer 在全局特征建模时未能充分考虑图像的空间局部特征，导致它在捕捉雨滴的局部形态和方向信息时存在不足，进而影响去雨效果。最后，面对复杂的噪声（如雨滴和光斑等）的鲁棒性较差，难以有效区分噪声与

真实图像内容，可能导致去雨效果不精细，甚至产生伪影或丢失重要细节。

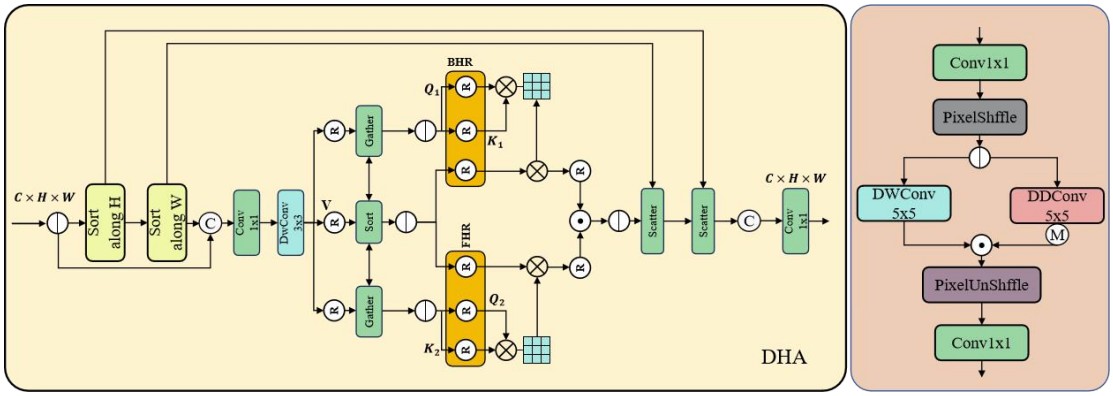


图 3-3 动态直方图注意力和双尺度门控前馈网络结构

为了使得模型能够更精确地定位和去除雨滴，并保留背景细节，引入动态直方图注意力模块（DHAT）[71]。该模块主要分为两部分：动态直方图注意力（Dynamic Hyperparameter Adaptation，DHA）和双尺度门控前馈网络（Directed Graph Feature Fusion，DGFF）。动态直方图注意力（DHA）通过动态调整图像中特征的注意力范围，使模型能够根据图像的不同区域自动适配特征提取策略。具体来说，DHA 机制通过捕捉雨滴区域的特征分布，调整特征的权重，使得模型能更精确地识别并去除雨滴噪声，同时避免背景细节的损失。它通过直方图自适应调整，确保在不同的雨天条件下，网络能高效地处理变化多端的雨滴形态。而双尺度门控前馈网络（DGFF）则通过在两个不同尺度下进行特征提取与融合，有效增强了网络对雨滴和背景复杂性的处理能力。同样的，DGFF 模块能够根据不同的图像区域及雨滴的大小和形态调整尺度，确保每个区域的特征能得到恰当的加权，从而提高去雨效果。这两个部分的结构如图 3-3 所示。

动态直方图自注意力机模块是一种结合了动态范围卷积和双路径直方图自注意力机制的特征提取模块。动态范围卷积通过对输入特征的空间分布进行重新排序，扩展了有效感受野。在输入特征 $I \in R^{C\times H\times W}$ 沿通道维度进行拆分为两个分支：对于第一个分支 I_1，特征会在水平 H 和垂直 V 方向上进行排序，以重新组织空间分布；第二个分支 I_2 则保持不变。重新排序后的特征会被拼接并通过可分离卷积进行处理，从而使模块能够同时捕捉局部和长距离的依赖关系。I_1 和 I_2 的处理计算公式如式（3-5），（3-6）和（3-7）所示:

$$I_1, I_2 = Split(I) \tag{3-5}$$

$$I_1 = Sort_v(Sort_h(I_1)) \tag{3-6}$$

$$I = DWConv_{3\times3}(Conv_{1\times1}(Conact(I_1, I_2)) \tag{3-7}$$

式中，$DWConv_{3\times3}$ 代表 3×3 的深度卷积，$Conv_{1\times1}$ 代表 1×1 的卷积，*Conact* 代表拼接操作，*Split* 代表沿通道维度进行拆分特征（其中，H 与 V 代表水平与垂直排序操作）。在动态范围卷积模块输出后，可以将特征分为值 V 向量、两对查询 Q 与键 K 向量，然后将 Q,K,V 注意力向量传递到两个分支中。首先对 V 进行序列排序，根据索引排列 Q，K 注意力向量对。对 Q，K，V 注意力向量的排序计算公式如式（3-8）和（3-9）所示：

$$V, i = Sort(R_{C\times H\times W}^{C\times HW}(V)) \tag{3-8}$$

$$Q_n, K_n = Split(Gather(R_{C\times H\times W}^{C\times HW}(I_{QK}, n), i)), n \in [1, 2] \tag{3-9}$$

式中，$R_{C\times H\times W}^{C\times HW}$ 代表将特征 $R^{C\times H\times W}$ 重塑为 $R^{C\times HW}$，i 进行排序的索引值。*Gather* 表示基于给定索引从张量中检索元素的操作。双路径直方图自注意力机制通过逐箱直方图重塑（Bin-wise Histogram Reshaping，BHR）和逐频率直方图重塑方法（Frequency-wise Histogram Reshaping，FHR）对两个 Q-K 注意力向量分别进行处理，从而提高模型的表达能力。给定箱（bin）数 B 时，将特征 $C\times HW$ 重塑为 $C\times B\times HW/B$。BHR 是直接分配等于 B 的箱数，并且每个 bin 包含 HW/B 个元素，使得每个 bin 包含大量动态分布的像素，同时能够提取大尺度的信息。FHR 则是将每个 bin 的频率设置为 B，并且 bin 为 HW/B，每个 bin 只包含少量相邻的像素，能够使得模型提取更细粒度的信息。计算注意力参数的公式如式（3-10），（3-11）和（3-12）所示：

$$Att_{BHR} = soft\max(\frac{R_B(Q_1)R_B(K_1)^T}{\sqrt{k}})R_B(V) \tag{3-10}$$

$$Att_{FHR} = soft\max(\frac{R_F(Q_2)R_F(K_2)^T}{\sqrt{k}})R_F(V) \tag{3-11}$$

$$Att = Att_{BHR} \odot Att_{FHR} \tag{3-12}$$

式中，k 代表注意力头的数量，Att_{BHR} 和 Att_{FHR} 代表不同重塑方向上的注意力参数。

双尺度门控前馈网络：该网络首先通过1×1的卷积将输入特征图通道扩展，提供更多的特征维度。然后使用3×3和5×5的深度可分离卷积，分别提取不同尺度的特征。在门控机制后，第二个分支的输出在经过激活之后充当另一个分支的门控映射。双尺度门控前馈网络完整计算公式如式（3-13），（3-14）和（3-15）所示：

$$I_{l,1}, I_{l,2} = Split(Shuffle(Conv_{1*1}(I_l))) \tag{3-13}$$

$$I_{l,1} = DWConv_{5\times5}(I_{l,1}), I_{l,2} = Conv_{3\times3}^{d,dilated}(I_{l,2}) \tag{3-14}$$

$$I_{l+1} = Conv_{1\times1}(UnShuffle(Mish(I_{l,2}) \odot I_{l,1})) \tag{3-15}$$

式中，$Conv_{3\times3}^{d,dilated}$ 代表3×3的扩展深度卷积，*Shuffle* 与 *UnShuffle* 表示像素重洗和恢复像素顺序，*Mish* 代表 *Mish* 激活函数。最终，完整的直方图 Transformer 块定义计算公式如式（3-16）和（3-17）所示：

$$I_j = I_{j-1} + DHA(LN(I_{j-1})) \tag{3-16}$$

$$I_j = I_j + GFF(LN(I_j)) \tag{3-17}$$

式中，LN 代表层归一化，I_j 代表第 j 级的特征。

3.3 全局空间感知模块（GSP)

在去雨的 U-net 网络中，传统跳跃连接直接将浅层特征与深层特征进行简单相加或拼接，这可能导致特征冗余或引入噪声信息。同时，直接的信息传递没有对浅层特征进行筛选或优化，导致有用信息（如背景细节）和噪声信息（如雨线）被同等对待。这种缺乏选择性的特征融合方式，使得模型难以有效区分雨线与背景，从而影响去雨精度。

为了解决这一问题，本章提出了全局空间感知模块（GSP）[72]，它的结构如图 3-4 所示，GSP 模块通过深度可分离卷积和空间注意力机制，对浅层特征进行优化和筛选，有效去除雨线噪声，同时保留背景细节。这种有选择的特征融合方式，避免了传统跳跃连接中特征冗余和噪声干扰的问题。其次，通过关系矩阵，GSP 模块能够提取并融合不同尺度的特征，适应雨线形态的多样性。通过多尺度

特征提取和全局信息融合，模型能够更精确地捕捉雨线的分布和形态。提升信息传递效率。

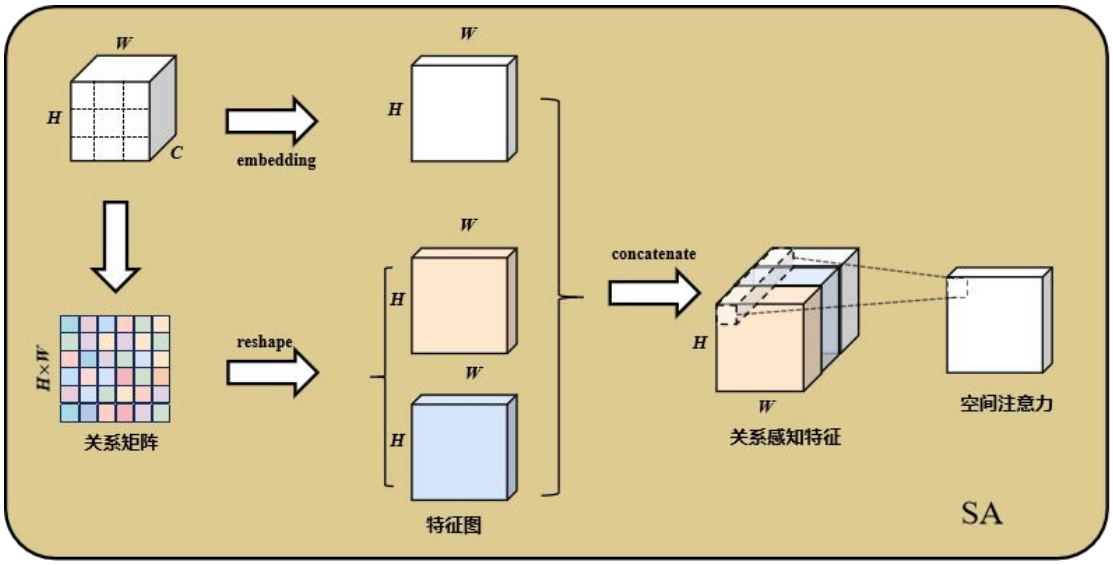


图 3-4 空间注意力结构

全局空间感知模块具体算法如下：

首先，对输入特征图 $X_{l-1} \in \mathbb{R}^{H\times W\times C}$ 进行1×1卷积操作，将通道数扩展为 $r\times C$，扩展通道的计算公式如式（3-18）所示：

$$\hat{X}_l = Conv_{1\times1}(LN(X_{l-1})) \tag{3-18}$$

式中，$Conv_{1\times1}$ 表示1×1卷积操作，$LN(\cdot)$ 表示层归一化，拓展通道为后续的多尺度卷积操作提供更丰富的特征表示。之后，在扩展后的特征图 $\hat{X}_l$ 上，引入两条并行的卷积分支，分别使用3×3和5×5的深度可分离卷积进行特征提取。经过特征提取的两条分支计算公式如式（3-19）和（3-20）所示：

$$X_{p1}^{l} = \sigma(DWConv_{3\times3}(\hat{X}_l)) \tag{3-19}$$

$$X_{s1}^{l} = \sigma(DWConv_{5\times5}(\hat{X}_l)) \tag{3-20}$$

式中，$DWConv_{3\times3}$ 和 $DWConv_{5\times5}$ 分别表示3×3和5×5的深度可分离卷积，σ 表示 *ReLU* 激活函数，通过不同尺度的深度卷积，能同时处理小尺度的雨线结构与大范围的雨线分布和全局特征。然后将两条分支的输出特征图 X_{p1}^{l} 和 X_{s1}^{l} 进行通道拼接，并通过1×1卷积将通道数还原为 C。拼接操作的计算公式如式（3-21）所示：

$$x = Conv_{1\times1}(\left[X_{p2}^{l}, X_{s2}^{l} \right]) + X_{l-1} \tag{3-21}$$

式中，x 表示融合之后的特征图，$[\cdot]$表示通道拼接操作。在融合了不同尺度的特征信息之后，再进行空间注意力操作。通过构建成对关系矩阵和注意力机制，捕捉特征之间的全局关系，从而进一步提升去雨任务的性能。

首先，空间注意力需要计算特征图$x \in \mathbb{R}^{H\times W\times C}$中每个空间位置与其他所有位置的成对关系来捕捉全局结构信息。成对关系的具体计算公式如式（3-22）所示：

$$r_{i,j} = f_s(x_i, x_j) = \theta_s(x_i)^T \phi_s(x_j) \tag{3-22}$$

式中，x_i 和 x_j 分别表示位置i和j的特征向量，θ_s 和 ϕ_s 是两个嵌入函数，通常通过1×1卷积实现，$r_{i,j}$ 不同位置的关系值。之后，将所有位置的成对关系值组合为成对关系矩阵 $R_s \in \mathbb{R}^{N\times N}\left(N = H\times W\right)$，并对每个位置 i 构建其关系特征 r_i ，构建关系特征的具体计算公式如式（3-23）所示：

$$r_i = \left[R_s(i,:), R_s(:,j) \right] \tag{3-23}$$

式中，$R_s(i,:)$ 是成对关系矩阵的第 i 行，表示位置 i 与其他所有位置的成对关系。$R_s(:,j)$ 为成对关系矩阵第 j 列，表示其他所有位置与位置i的成对关系。然后，将位置 i 的原始特征 x_i 与其关系特征 r 结合，构建关系感知特征 y_i ，构建关系特征的具体计算公式如式（3-24）所示：

$$y_i = \left[x_i, r_i \right] \tag{3-24}$$

式中，y_i 表示融合信息，这一步将局部特征和全局关系信息结合在一起，为后续的注意力学习提供更丰富的输入。最后，通过一个浅层卷积模型计算每个位置的注意力值 a_i，并与原始特征 x_i 相乘得到增强后的特征，最后的增强特征信息计算公式如式（3-25）和（3-26）所示。

$$a_i = Sigmoid(W_2 \,\mathrm{Re}\, LU(W_1 y_i)) \tag{3-25}$$

$$\hat{x}_i = a_i \cdot x_i \tag{3-26}$$

式中，$\hat{x}_i$ 为增强后的特征图。

3.4 损失函数

在去雨任务中，损失函数的设计对生成高质量的去雨图像至关重要。本章采用 L1 损失与 L2 损失相结合的方式作为去雨任务的损失函数。在雨天场景中，雨滴往往表现为局部的噪声点或条纹，L1 损失通过计算去雨图像与目标图像之间的绝对误差，能够更精确地捕捉这些局部特征，避免因过度平滑而丢失细节。并且，L1 损失能够有效抑制雨滴稀疏噪声，同时保留背景的清晰度。而 L2 损失则通过对小像素误差进行强烈的惩罚，精细地优化图像的整体误差，确保去雨后的图像在颜色和亮度等全局特征上尽可能接近真实的干净图像。此外，L2 损失在处理均匀分布的噪声时表现优异，能够有效平衡图像的整体质量，确保去雨后的图像在视觉效果上更加自然。通过合理地权衡这两种损失，我们能够在去雨效果和图像质量之间取得良好的平衡。具体的损失公式如式（3-27），（3-28）和（3-29）所示：

给定输入的有雨图像 I_{rain} 和网络生成的去雨图像 I_{derain}，我们定义损失函数如式（3-27）所示：

$$\mathcal{L} = \lambda_1 \mathcal{L}_1 + \lambda_2 \mathcal{L}_2 \tag{3-27}$$

式中，λ_1 和 λ_2 分别为 $L1$ 损失和 $L2$ 损失的权重系数，用于控制两种损失在总损失中的贡献。$L1$ 和 $L2$ 损失函数定义如式（3-28）所示：

$$\mathcal{L}_{L1} = \frac{1}{N}\sum_{i=1}^{N}\left|I_{\text{derain}}^{(i)} - I_{\text{gt}}^{(i)}\right| \tag{3-28}$$

式中中 N 图像中的像素总数，$I_{derain}^{(i)}$ 和 $I_{gt}^{(i)}$ 分别表示去雨图像和目标图像的第 i 个像素值，$L1$ 损失用于计算去雨图像与目标无雨图像 I_{gt} 之间的绝对误差，L_2 损失的计算公式如式（3-29）所示：

$$\mathcal{L}_{L2} = \frac{1}{N}\sum_{i=1}^{N}\left(I_{\text{derain}}^{(i)} - I_{\text{gt}}^{(i)}\right)^2 \tag{3-29}$$

L_2 损失用于计算去雨图像与目标无雨图像之间的平方误差。

3.5 检测评价指标

3.5.1 图像质量评价指标

在图像去雨任务中，为了客观评估生成图像的质量，通常需要采用多种图像质量评价指标。这些指标能够从不同角度量化去雨图像与目标无雨图像之间的相似性。本章采用结构相似性指数（Structural Similarity Index，SSIM）和峰值信噪比（Peak Signal-to-Noise Ratio，PSNR）两种广泛使用的图像质量评价指标，作为去雨后图像质量的客观评价标准。

（1）SSIM 是一种衡量两幅图像在亮度、对比度和结构方面相似性的指标。SSIM 的值范围在[-1, 1]之间，值越接近 1，表示两幅图像越相似。SSIM 的计算流程如式（3-30），（3-31），（3-32），（3-33）和（3-34）所示：

其中，亮度对比函数如式（3-30）所示：

$$l(i,j)=\frac{2\alpha_i\alpha_j+b_1}{\alpha_i^2+\alpha_j^2+b_1} \tag{3-30}$$

式中，i 表示第一张图像窗口中的数据，j 表示第二张图像窗口中的数据，α_i,α_j 表示 i 和 j 的均值，b_1 是常数，为了避免分母为 0。

对比度对比函数如式（3-31）所示：

$$c(i,j)=\frac{2\beta_i\beta_j+b_2}{\beta_i^{\,2}+\beta_j^{\,2}+b_2} \tag{3-31}$$

式中，β_i 和 β_j 表示 i 和 j 的方差，b_2 也是常数，为了避免分母为 0。

结构对比函数如式（3-32）所示：

$$s(i,j)=\frac{\beta_{ij}+b_3}{\beta_i\beta_j+b_3} \tag{3-32}$$

式中，β_{ij} 表示 i 和 j 之间的协方差。

最后，SSIM 的计算公式如式（3-33）和（3-34）所示：

$$SSIM(i,j)=[l(i,j)^a\cdot c(i,j)^c\cdot s(i,j)^d] \tag{3-33}$$

式中，a,c,d 都大于 0，令其都为 1，则得到常用的简化 SSIM 公式：

$$SSIM(i,j)=\frac{(2\beta_i\beta_j+b_1)(2\beta_{ij}+b_2)}{(\beta_i^2+\beta_j^2+b_1)(\sigma_x^2+\sigma_y^2+b_2)} \tag{3-34}$$

（2）PSNR 是一种基于均方误差（MSE）的图像质量评价指标，通常用于衡量去雨图像与目标图像之间的像素级差异。PSNR 的值越高，表示图像质量越好。PSNR 的计算公式如式（3-35）所示：

$$PSNR=20\times\log_{10}(\frac{P_{\max}}{MSE}) \tag{3-35}$$

式中，$P_{\max}$ 是图像像素的最大值，MSE 为均方误差。

3.5.2 目标检测评价指标

混淆矩阵是目标检测任务中用于评估模型性能的重要工具，通过将模型的预测结果与真实标签进行对比，生成一个矩阵。矩阵的每一行表示真实类别，每一列表示预测类别。混淆矩阵包含四个关键指标：真正例（True Positive, TP）（模型正确预测为正例的数量）、假正例（False Positive, FP）（模型错误预测为正例的数量）、真负例（True Negative, TN）（模型正确预测为负例的数量）和假负例（False Negative, FN）（模型错误预测为负例的数量）。通过混淆矩阵，可以直观地分析模型在不同类别上的表现，尤其是模型在正例和负例上的分类能力。

在全面衡量模型的检测能力的部分，通常采用四个核心指标：精确率（Precision，P）、召回率（Recall，R）、平均准确率（Average Precision，AP）和平均精度均值（mean Average Precision，mAP）。这些指标分别从不同角度反映了模型在检测正例、避免误检和漏检以及多类别检测任务中的综合表现。下面详细介绍各个指标：

精确率: 精确率是模型在预测为正例的样本中，真实为正例的比例，反映了模型在检测过程中避免误检的能力。精确率的计算公式如式（3-36）所示：

$$P=\frac{TP}{TP+FP} \tag{3-36}$$

式中，TP 表示模型正确检测到的正例数量，FP 表示模型错误检测为负例的数量。精确率越高，说明模型的误检率越低，检测结果的可信度越高。

召回率：召回率是模型在所有真实正例中成功检测到的比例，反映了模型在

检测过程中避免漏检的能力。召回率计算公式如式（3-37）所示：

$$R = \frac{TP}{TP + FN} \tag{3-37}$$

式中，FN 表示模型未能检测到的正例数量。召回率越高，说明模型的漏检率越低，检测结果的覆盖率越高。

平均准确率：平均准确率是对不同置信度阈值下精确率和召回率的综合评估，反映了模型在多阈值下的整体性能。平均准确率计算公式如式（3-38）所示：

$$AP = \sum_{i=1}^{M} P(i) \times \Delta R(i) \tag{3-38}$$

式中，M 是测试样本的数量，$P(i)$ 和 $R(i)$ 分别表示在第 i 个阈值下的精确率和召回率。AP 高，说明模型在不同置信度阈值下的检测性能越稳定。

平均精度均值：平均精度均值是所有类别的 AP 的平均值，用于评估模型在多类别目标检测任务中的整体性能。平均精度均值的计算公式如式（3-39）所示：

$$mAP = \frac{1}{M} \sum_{i=1}^{C} AP_i \tag{3-39}$$

式中，M 为检测类别的数量，AP_i 表示第 i 个类别的值。mAP 越高，说明模型在多类别检测任务中的整体性能越好。

在目标检测任务中，评估模型的性能不仅需要关注检测精度，还需综合考虑检测效率和训练过程中的优化效果。为了量化模型的检测速度，通常采用每秒传输帧数（Frames Per Second, FPS）作为评价标准。FPS 表示模型在单位时间内能够处理的图像帧数，其值越高，表明算法的实时性越好，计算效率越高。此外，模型大小（单位 MB）、参数量和浮点运算次数（Floating Point Operations，FLOPs）也是衡量模型性能的重要指标，它们直接反映了模型存储成本与计算的复杂度。

3.6 雨天行人数据集构建

3.6.1 雨滴的物理特性

雨滴在下落过程中受到多种物理因素的影响，包括表面张力、静水压力、空气动力压力以及外部干扰等。这些因素共同作用导致雨滴发生形变，并在图像中

形成具有不同方向和亮度的雨纹。这些雨纹不仅会遮挡图像中的物体和背景，还会造成图像模糊，严重影响视觉信息的完整性。雨滴下落过程中的物理特性主要由它的大小决定，空中的雨滴形状表达为公式（3-40）所示：

$$r(\theta) = a(1 + \sum_{n=1}^{0} c_n \cos(n\theta)) \tag{2-15)(2-16}$$

式中，a 为球形雨滴半径，c_n 是雨滴的形状系数,θ 为雨滴倾斜的角度。

当雨滴自高空下落时，受空气阻力作用最终将趋于匀速运动状态，该速度值与雨滴直径密切相关，被定义为终端速度。研究人员基于大量实验数据的分析，揭示了雨滴终端速度v_0与其直径d之间的数量关系为公式（3-41）和（3-42）所示：

$$v_0 = -0.2 + 5d - 0.9d^2 + 0.1d^3 \tag{2-17}$$

$$v = v_0(\frac{\rho_0}{\rho})^{0.4} \tag{2-18}$$

式中，ρ_0 是雨滴开始降落位置的空气密度,ρ 是雨滴所在位置的空气密度。

3.6.2 数据集构建

针对现有数据集中缺乏代表性雨天拥挤行人场景的问题，本研究提出了一种基于雨线特征的数据增强方法。具体而言，利用 Python 算法模拟雨线的物理特征，并将其叠加到 Crowdhuman[73]数据集的原始图像中，构建了一个能够真实反映雨天拥挤行人场景的增强数据集。该数据集通过引入雨纹干扰，更好地模拟了实际雨天环境中行人检测所面临的挑战，为相关研究提供了可靠的数据支持。

Crowdhuman 数据集主要采集于背景复杂多变的密集行人场景，其中存在大量因拥挤导致被遮挡的行人目标，同时行人目标的背景情况也十分复杂。这个数据集一共拥有 24370 张左右的拥挤行人照片，平均每张照片上有 23 个行人，一共有多达 480000 个行人实例。其中 15000 张用于训练，5000 张用于测试，4370 张用于验证。

在 Python 环境中，可通过引入随机噪声模型对图像进行处理，以仿真雨滴运动轨迹实现无雨图像的雨天场景渲染。具体操作是通过控制不同密度的噪声分布来模拟雨滴直径的差异，其基础噪声样本图像如图 3-5 所示。

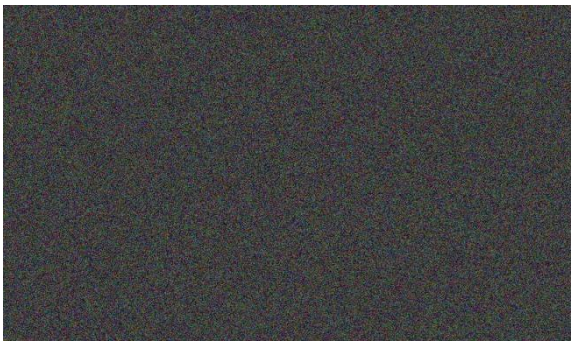
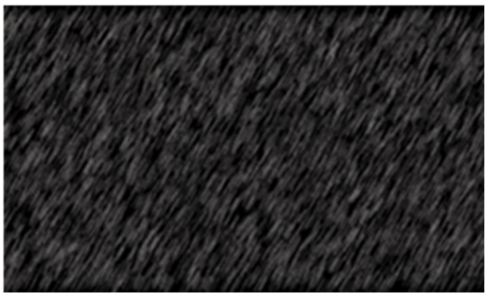
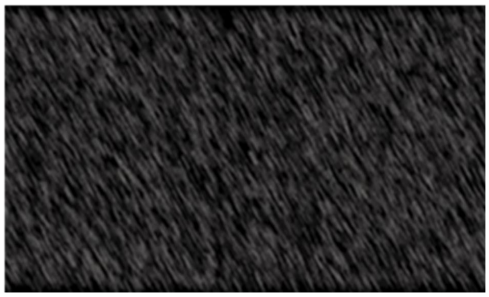


图 3-5 基础噪声样本图像

随后，需对基础噪声形态进行变换，通过调整噪声单元的长度参数并施加几何旋转操作，可有效模拟不同直径雨滴在风力作用下的运动轨迹特征。经过参数调节的雨纹形态示意图如图 3-6 所示。



(a) 雨纹图a



(b) 雨纹图b

图 3-6 雨纹噪声效果图

最后将生成的雨纹噪声与原始图像进行叠加，构建雨天行人场景仿真图像，与原数据集对比效果如图 3-7 所示。本文共合成 24370 张左右雨天行人数据集，命名为 Rain-Crowdhuman。

图 3-7 雨纹添加前后效果对比图

3.7 仿真实验与结果分析

3.7.1 实验环境

实验平台配备 NVIDIA RTX 3090 GPU, Intel Core i5-13600KF CPU, 3.5 GHz, 16 GB 内存，Windows10 操作系统。实验在深度学习框架 Pytorch 上进行。具体配置如表 3-1 所示。

表 3-1 实验配置表

实验配置	型号
CPU	16 vCPU Intel(R) Xeon(R) Gold 6430
GPU	NVIDIA RTX 3090
操作系统	Windows10
深度学习架构	Pytorch 2.2.2
运行内存	16G
运行库	CUDA 12.1
软件	Python 3.9.19

3.7.2 实验结果与分析

本文在整个自建数据集 Rain-CrowdHuman 上将我们的方法与 MPRNet[74]和 TMAM[75]进行综合定量对比分析，结果如表 3-2 所示。从表 3-2 可以看出本文模型平均的 PSNR 与 SSIM 均为最高值，表明图像去雨效果最好，证明了本文方法的可行性。

表 3-2 在 Rain-CrowdHuman 数据集的定量评价

算法	No Rain	MPRNet	TMAM	Our
SSIM	1.0	0.906	0.894	0.916
PSNR	Inf	30.083	29.765	31.244

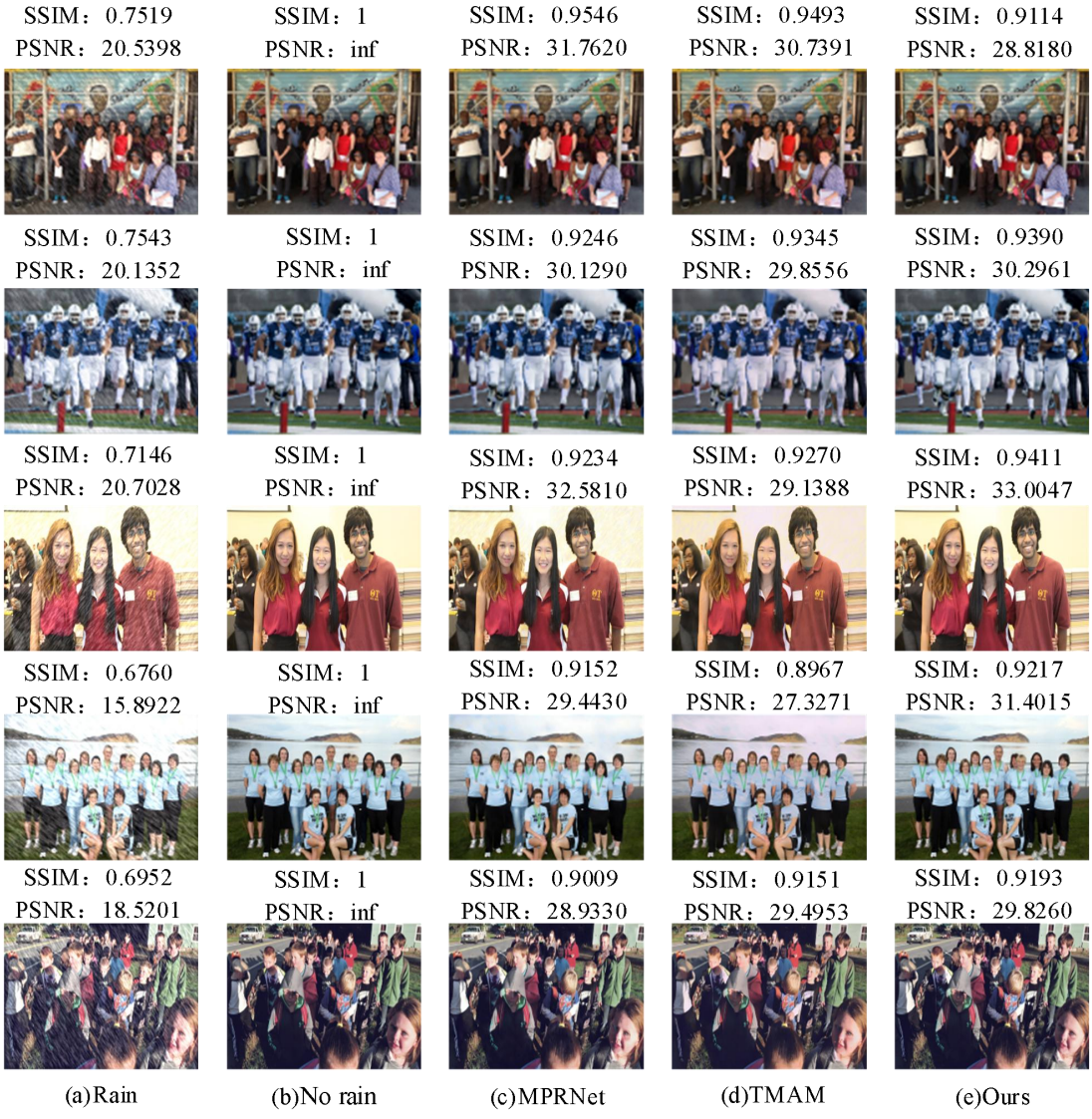


图 3-8 在 Rain-CrowdHuman 数据集上的去雨效果对比图

本文算法的主观实验效果对比如图 3-8 所示。 TMAM 方法在第三行与第四行图像的背景引入不必要的色调，其第一行图像中具有最高的 SSIM 值。MPRNet 与 TMAM 在第一行图像均高于本文算法，且 MPRNet 在第一行图像具有最高的 PSNR 与 SSIM 值。本章模型除第一行外，在剩下四幅图像评价指标方面均排名第一，得到了较好的去雨效果。

为了进一步评估算法性能与去雨效果，本章随机选取了两张合成数据集中的图片进行局部细节的验证，如图 3-9 所示。通过对比实验，结果表明本章提出的模型在去除图像中的雨滴方面表现出色。相较于其他现有的去雨方法，本章模型不仅能够精确去除雨水的干扰，还能有效恢复被雨水遮挡的细节，显著提升了图像的视觉质量。这一优势对于后续的行人检测至关重要，因为更清晰、更真实的图像为行人检测提供了更加可靠和高质量的输入数据，减少了因雨水干扰所带来的误检测或漏检问题。

图 3-9 在 Rain-CrowdHuman 数据集进行细节对比实验

在本小节中，我们通过消融实验验证了模型中各个模块的有效性，选取 SSIM 和 PSNR 作为评估指标。实验中，我们分别移除了 MSFA、GSP 和 DHAT 模块，并对比了不同模块缺失情况下的模型性能。实验结果如表 3-3 和图 3-10 所示，完整模型在去雨任务中表现最佳，PSNR 达到 31.267，SSIM 为 0.920，显著优于其他模块缺失的组合。这一结果充分证明了 MSFA、GSP 和 DHAT 三者结合在细节恢复和全局一致性优化方面的有效性，能够显著提升去雨效果。

表 3-3 消融实验的定量分析

Methods	MSFA	GSP	DHAT	Test-Our	
				PSNR	SSIM
-wo/MSFA	✗	✓	✓	30.348	0.911
-wo/GSP	✓	✗	✓	30.564	0.912
-wo/DHAT	✓	✓	✗	29.750	0.896
Full model	✓	✓	✓	31.267	0.920

进一步分析发现，不同模块的缺失对去雨效果的影响存在显著差异。具体而言，当缺少 MSFA 模块时，模型性能略有下降，PSNR 为 30.348，SSIM 为 0.911；而缺少 GSP 模块时，性能反而有所提升，PSNR 达到 30.564，SSIM 为 0.912。这一结果表明，多特征聚合模块（MSFA）对细节恢复具有重要作用。相比之下，缺少 DHAT 模块时，模型性能的下降最为明显，PSNR 仅为 29.750，SSIM 为 0.896，这充分证明了多头注意力机制在图像去雨任务中起着关键作用。综合来看，MSFA、GSP 和 DHAT 三个模块的结合充分发挥了各自的优势，形成了有效的互补机制，从而显著提升了整体去雨效果。

图 3-10 消融实验主观图

在目标检测实验部分，本文在自建数据集 Rain-CrowdHuman 数据集上，分别采用 SSD、Fast-Rcnn、YOLOv3、YOLOv5s、YOLOv8s 和 YOLOv10s 算法模型进行对比实验。本次实验批量大小（batch_size）设置为 8，且一共训练 200 个 epoch，初始学习率大小 0.01，对比实验结果如表 3-4 所示。

由表 3-4 可知，在相同数据集和实验环境下，YOLOv8s 和 YOLOv10s 在检测精确度和检测速度上整体要优于其他算法。其中 YOLOv10s 算法在精确度 P 和 mAP_0.5 分别为 81.0%和 77.2%，相比于 YOLOv3 分别提升了 0.2%和 0.3%，相比于 YOLOv5s 分别提升了 0.6%和 0.5%，相比于 SSD 和 Fast-Rcnn 提升更多，在几个算法模型中也仅仅略低于 YOLOv8s，但 YOLOv10s 的 FPS 较 YOLOv8 提升了约 15%，拥有更快的检测速度。综上所述，与其他三个模型相比，YOLOv10s 取得了最快的检测速度和较高的检测精度。由于 YOLOv10s 在模型精确度上损失很小，且行人检测对检测速度要求颇高，所以本文采用 YOLOv10s 作为基线模型。

为了进一步验证改进的去雨算法对行人检测的有效性，在同一数据集 Rain-CrowdHuman 上进行实验，这里采用基础的 YOLOv10s 算法和添加去雨算法后的 YOLOv10s-dr 算法进行对比实验，对比结果如表 3-5 所示。

表 3-4 不同算法模型实验结果对比

算法模型	P/%	R/%	mAP_0.5/%	FPS
SSD	74.7	44.6	70.7	60
Fast-Rcnn	76.9	46.8	72.9	8
YOLOv3	80.8	49.7	76.9	45
YOLOv5s	80.4	47.5	76.7	95
YOLOv8s	**81.3**	49.9	**77.4**	105
YOLOv10s	81.0	**50.8**	77.2	**120**

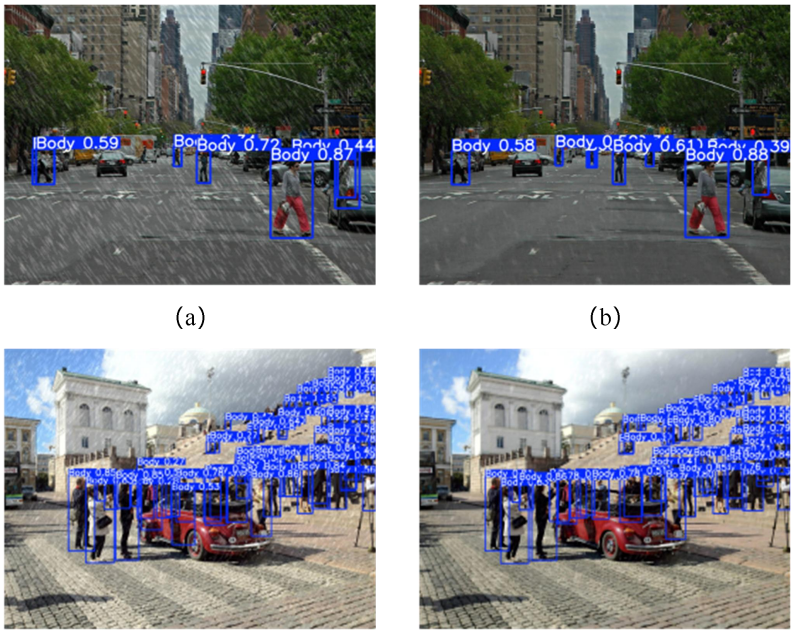
表 3-5 模型实验结果对比

算法模型	P/%	R/%	mAP_0.5/%
YOLOv10s	81.0	50.8	77.2
YOLOv10s-dr	83.5	51.9	79.1

从表 3-5 可知，添加去雨算法后的 YOLOv10s-dr 相较于原始的 YOLOv10s 算法，在 P、R 和 mAP_0.5 上分别提高了 2.5%、1.1%和 1.9%，充分证明了本章提出的基于动态直方图注意力机制与多特征聚合的图像去雨算法有助于提升行人检测算法的检测性性能。

与此同时，为了进一步验证检测性能的有效性，进行去雨前后行人检测可视

化效果对比，结果如图 3-11 所示。对比图（a）与图（b）可以看出，未经过去雨网络的 YOLOv10s 算法在雨线遮挡的小目标行人时，会存在漏检率高的情况，会将图片远处左侧白色车头的行人与右侧骑自行车的行人漏检，而经过去雨网络的 YOLOv10s 算法能检测出这两位漏检的行人。同样，对比图（c）与图（d），可以看出在行人过于密集的场景中，经过去雨的图像在小目标行人检测的能力上同样能做到更好。具体体现在，图像经过去雨后，能够检测出去雨前漏检的位于图像上方中央的远处台阶上的小目标行人若干。综上所述，证明了本章去雨算法的有效性和可行性。



(a) (b)

(c) (d)

图 3-11 去雨前后行人检测对比图

3.8 本章小结

为了解决雨天环境下行人检测精度下降的问题，本章提出了一种基于动态直方图注意力机制与多特征聚合的图像去雨算法。该算法通过学习有雨图像和对应的无雨图像之间的特征映射关系，有效捕捉雨线的分布模式与背景的空间关系，从而实现雨线去除并恢复图像的细节。为验证算法有效性，本章自建 Rain-CrowdHuman 数据集并进行了对比和消融实验，结果表明该算法优于现有方法，并验证了本章模型的合理性。进一步地，将本文去雨模型与 YOLOv10s

目标检测模型结合，在雨天场景下的行人检测任务中进行了验证。实验结果显示，经过去雨预处理后，目标检测模型在雨线遮挡区域场景下的检测精度有显著提升。

第 4 章 基于改进 YOLOv10 的密集行人检测算法

本文针对拥挤场景下行人检测准确率低、漏误检率高等问题，在 YOLOv10s 架构基础上提出了改进的 MDEN-YOLOv10 网络，其结构如图 4-1 所示。该网络首先通过新增 160×160 特征尺度并采用多尺度特征融合机制，在原有 80×80、40×40、20×20 这三个尺度基础上扩展特征提取范围，有效缓解了因遮挡导致的特征信息丢失问题；其次引入多样分支块（DBB）模块，通过集成不同尺寸的卷积核和池化操作构建多分支结构，结合结构重参数化技术，在训练阶段保持高特征表达能力的同时，在推理阶段简化为单个卷积层，从而显著提升了特征提取能力和计算效率；接着引入的 ECA 通道注意力机制凭借轻量级设计，易于集成到现有 CNN 架构中，通过动态重加权通道重要性，在降低参数负载的同时，显著提升了对关键特征的关注度，有效抑制了背景干扰，从而整体提升了模型性能。最后结合交并比（IoU）和归一化瓦瑟斯坦距离（NWD）损失函数，利用 NWD 对目标尺度变化不敏感的特性，优化了小目标检测性能，同时保留 IoU 对大目标的准确评估能力。

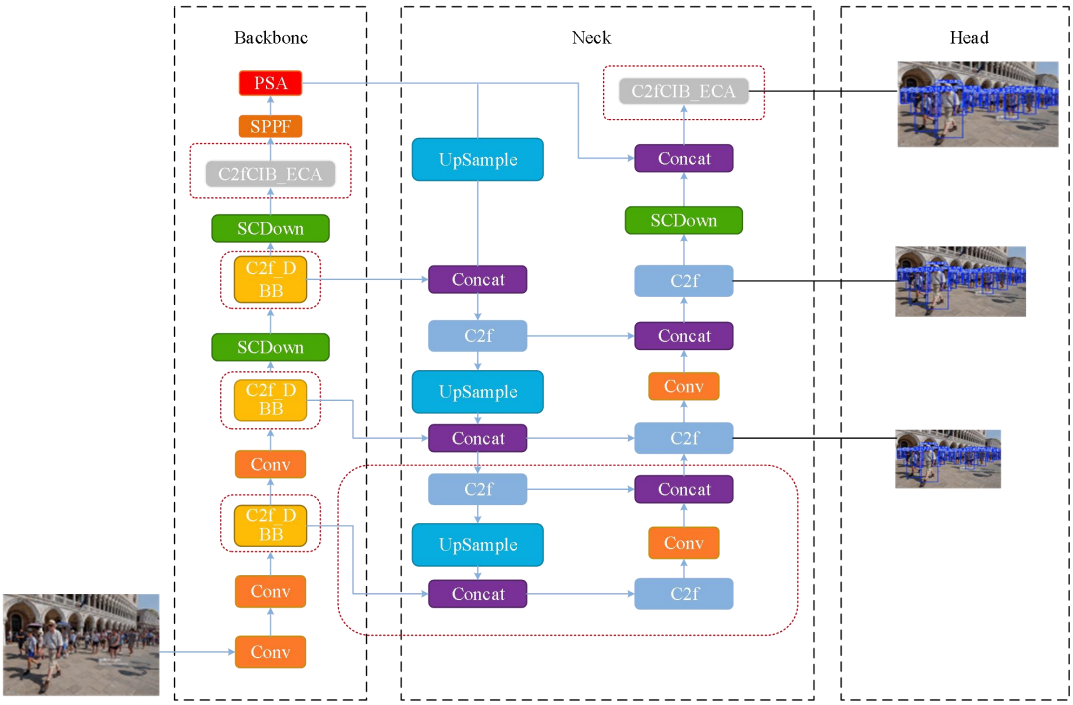


图 4-1 MDEN-YOLOv10 网络结构示意图

4.1 多尺度特征融合模块

在 YOLOv10 算法中，小目标检测面临三个核心挑战：其一，尺寸小于 32×32 像素的行人目标因分辨率不足，在检测过程中易出现漏检和误检现象，这在包含密集小目标和遮挡行人的 CrowdHuman 数据集中表现尤为突出；其二，随着网络层级的加深，特征图的分辨率逐级降低，虽然深层特征图具有更大的感受野和语义表征能力，但小目标的细节信息在多次下采样过程中逐渐衰减；其三，传统卷积架构对遮挡场景中的细粒度局部特征提取能力不足，难以有效整合被遮挡目标的残缺特征，导致漏检率显著升高。这三个方面的缺陷共同制约了复杂场景下的行人检测性能。

为了克服这一挑战，本文提出的网络延续了 YOLOv10 骨干网络中提取和融合关键特征的方法，即从 80×80、40×40 和 20×20 大小的特征图中获取信息。为了更好地捕捉被遮挡人员的暴露局部细节，本文在现有三个尺度的基础上，增加了一个 160×160 的特征尺度，如图 4-2 所示。该方法通过引入更大尺寸的特征图，能够有效提取更多的细粒度信息，尤其是在复杂环境下的遮挡场景中，进一步提升了模型对细节的感知能力。具体而言，该方法从骨干网络的原始 160×160 特征图中提取特征，并将其与现有的三个尺度进行融合，从而提供更加全面的特征描述，这种设计确保了在小目标和被遮挡目标的情况下，模型仍能保持较高的检测精度。多尺度融合模块具体结构如图 4-2 所示。

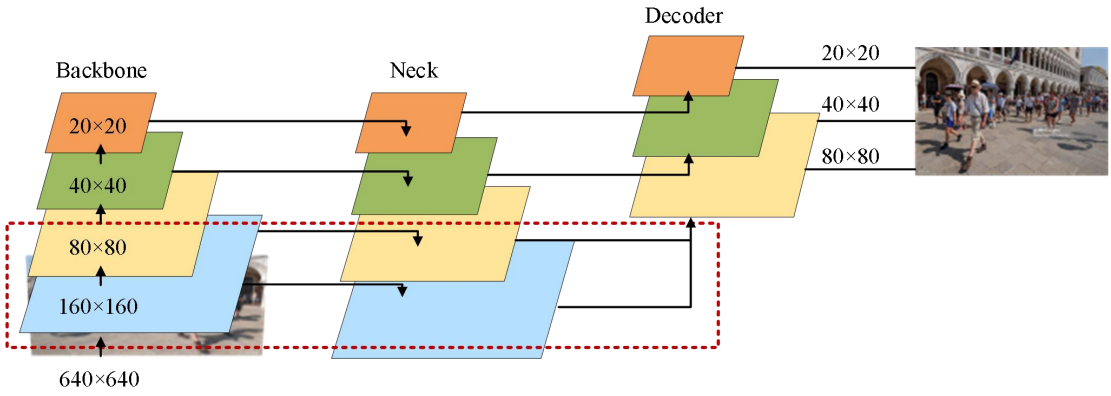


图 4-2 改进后的多尺度特征融合

首先，将较小尺寸的特征图（如 F_{80}，F_{40} 和 F_{20}）过上采样操作扩展到 160×160 分辨率，扩展过程具体计算公式如式（4-1）所示。

$$F_{80}^{up} = Upsample(F_{80}), F_{40}^{up} = Upsample(F_{40}), F_{20}^{up} = Upsample(F_{20}) \tag{4-1}$$

式中，$Upsample(\cdot)$ 表示上采样。接着，将上采样后的特征图和 F_{160} 进行融合，生成融合后的特征图F_{fused}具体表达式如式（4-2）所示：

$$F_{fused} = Concat(F_{160}, F_{80}^{up}, F_{40}^{up}, F_{20}^{up}) \tag{4-2}$$

式中，$Concat(\cdot)$ 表示言通道维度拼接操作。在融合后的特征图 F_{fused} 上，新增的160×160层级检测头负责检测小目标。通过预测每个特征图(i, j)的边界框偏移量进行预测。预测过程的具体计算公式如式（4-3）所示：

$$\hat{B}_{i,j} = (\Delta x, \Delta y, \Delta w, \Delta h) \tag{4-3}$$

式中，Δx 和Δy 表示中心点偏移，Δw和Δh 表示宽度和高度的缩放。对于每个边界框，预测目标类别概率 p_c 计算公式如式（4-4）所示：

$$p_c = Soft\max(W_c \cdot F_{fused} b_c) \tag{4-4}$$

式中，W_c和b_c 分别为分类器的权重和偏置。

4.2 C2f_DBB 模块

在训练过程中，DBB 模块[76]通过多分支结构的设计显著提升了特征表达能力，DBB 模块结构如图 4-3 所示。该模块集成了不同尺寸的卷积操作、平均池化以及分支求和等多种计算路径，这些并行分支结构能够从多个维度捕捉图像特征，从而有效丰富了特征空间，增强了模型对复杂特征的表达能力。DBB 的独特之处在于，尽管它在训练时采用复杂的多分支结构，但在推理时可以通过六种不同的变换方法将这些分支等效转换为单个卷积层。这些变换使得 DBB 能够保留与原始网络相同的推理结构，在不增加推理成本的情况下显著提升模型性能。此外，DBB 模块通过灵活的变换方式，使得模型的通用性更强，能够更好地适应各种网络架构和不同类型的任务。在不同规模的数据集上，DBB 都能有效提升特征提取的能力，并且避免了传统多分支结构中因参数增加带来的性能瓶颈。DBB 在不同阶段具体算法流程如下：

在训练阶段，DBB 模块通过多分支结构来丰富特征空间。输入特征图为$F \in \mathbb{R}^{H\times W\times C}$，DBB 模块的输出特征公式如式（4-5）所示：

$$F_{out}=\sum_{j}B_j((AvgPool(Conv_{k\times k}(\sum_{i=1}^{N}B_i(F)))))\tag{4-5}$$

式中：$k\times k$ 表示卷积核尺寸。B_i 表示第 i 个分支的操作，N 为分支数量，通过对不同分支的卷积池化操作，来提升网络的特征提取性能。

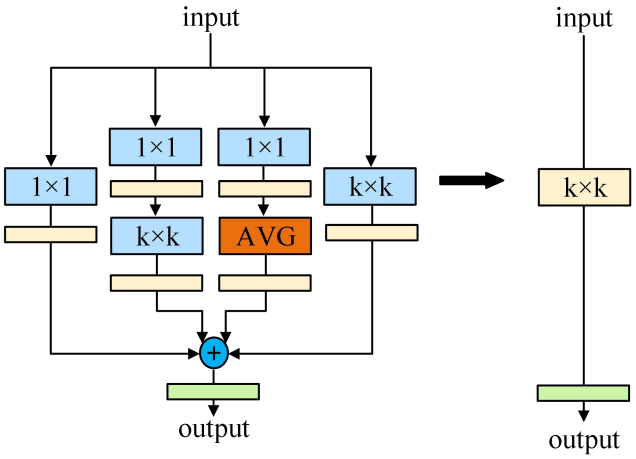


图 4-3 多样分支块（DBB）模块在训练和推理阶段的情况

在推理阶段，DBB 等效转换方法具体计算公式如式（4-6），（4-7），（4-8），（4-9），（4-10）和（4-11）所示：

卷积层与批归一化层合并：

$$Conv_{seq}(F)=Conv_1(Conv_2(F))\tag{4-6}$$

多个卷积层输出求和：

$$Conv_{sum}(F)=\sum_{i}Conv_i(F)\tag{4-7}$$

卷积序列合并：

$$Conv_{seq}(F)=Conv_1(Conv_2(F))\tag{4-8}$$

深度连接卷积层合并：

$$Conv_{depth}(F)=Conv_1(F)+Conv_2(F)\tag{4-9}$$

平均池化转换：

$$Conv_{avg}(F)=Conv_{avg\_kernel}(F)\tag{4-10}$$

多尺度卷积转换：

$$Conv_{multi}(F)=\sum Conv_{k\times k}(F)\tag{4-11}$$

通过上述变换，DBB 模块在推理阶段简化为单个卷积层，从而降低计算复杂度。

为了进一步增强网络骨干的特征提取能力，本文引入了一种改进的 C2f_DBB 模块，其融入了结构重参数化的概念。具体而言，C2f_DBB 模块通过将 DBB 模块引入到瓶颈结构中，对 C2f 模块进行了改进，用一个 DBB 模块取代了瓶颈结构中的3×3 卷积，从而形成了 Bottleneck_DBB 模块。该模块通过结合1×1 卷积、$K \times K$ 卷积和平均池化层构建了一个多分支结构。在训练阶段，每个分支有效地整合了来自不同尺度和感受野的特征信息，使模型能够更全面地捕捉到更精细的图像细节。训练完成后，这些多分支结构被等效转换为单个3×3 卷积，在不影响检测性能的前提下简化了网络，同时显著加快了推理速度。通过这种设计，C2f_DBB 模块不仅在训练阶段增强了网络的学习能力，而且在推理阶段有效地降低了计算复杂度，在实时处理中保持了较高的效率，它的结构如图 4-4 所示。

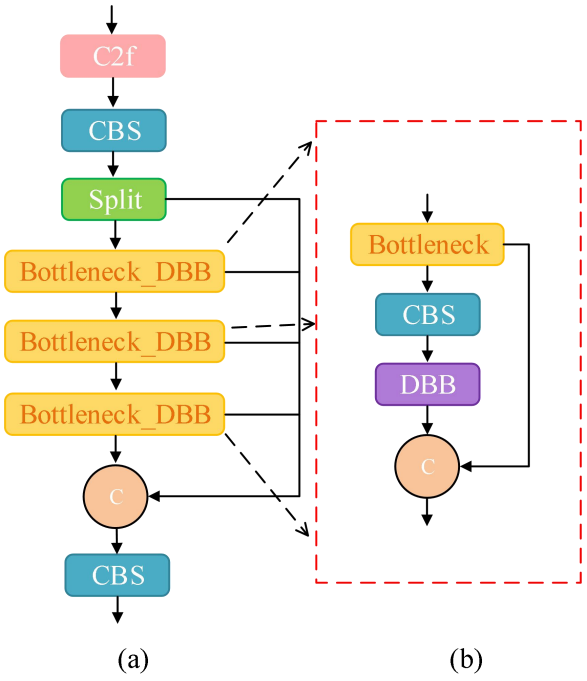


图 4-4 C2f_DBB 模块。(a) C2f_DBB，(b) Bottleneck _DBB 模块

4.3 改进的 C2fCIB 注意力模块

在复杂的公共场景中，目标识别会受到建筑物、车辆遮挡以及无规则背景等因素的影响。这些干扰因素会给机器的视觉系统引入噪声，降低识别效率。而注

意力机制利用自适应特征加权策略，能够动态地从输入图像中筛选出关键信息，从而突出重要特征并抑制来自无关区域的响应。在此背景下，本文将高效通道注意力（ECA）机制[77]引入到 YOLOv10 中，其中 ECA 结构如图 4-5（a）所示。ECA 模块是在 SE 模块的基础上构建的，它去除了全连接层，转而采用 1×1 卷积来执行局部跨通道交互。这种方法减轻了降维带来的负面影响，同时显著减小了模型的参数规模和计算复杂度。该模块对特征进行重构和重新分配，有效地增加了小目标特征图的权重和比例，这弥补了卷积模块在通道处理方面的局限性，增强了网络的特征提取能力。此外，ECA 模块通过自适应调整通道权重，在多目标重叠或复杂背景场景中有效区分目标与背景。该模块通过抑制背景干扰和动态分配特征图权重，显著提升了检测的准确性和鲁棒性，使模型在复杂场景中表现出更强的泛化能力。ECA 模块的具体算法流程如下：

对于输入特征图 $F \in \mathbb{R}^{H\times W\times C}$ 进行全局平均池化，生成通道描述符 $z \in \mathbb{R}^{1\times 1\times C}$：

$$z_c = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}F_c(i,j) \tag{4-12}$$

式中，z_c 表示第 c 个通道的全局平均池化值。接着通过1×1卷积对通道描述符 z 进行局部跨通道交互，生成通道权重 s，权重 s 计算公式如式（4-13）所示，$s \in \mathbb{R}^{1\times 1\times C}$：

$$s = \sigma(Conv_{1\times 1}(z)) \tag{4-13}$$

式中，$Conv_{1\times 1}$ 表示1×1的卷积操作，σ 表示表示 Sigmoid 激活函数。之后，将生成的通道权重 s 与原始特征图 F 相乘，得到重校准后的特征图 $\hat{F}$ 计算公式如式（4-14）所示，其中 $\hat{F} \in \mathbb{R}^{H\times W\times C}$：

$$\hat{F} = s_c \cdot F_c \tag{4-14}$$

式中，s_c 表示第 c 个通道的权重，F_c 表示重校准后的第 c 个通道特征图。

YOLOv10 引入了一种名为紧凑倒置块（CIB）的新型神经网络结构。该模块通过利用成本效益高的深度可分离卷积进行空间维度混合，以及使用低成本的逐点卷积进行通道维度混合，实现了高效的特征混合。其可以在保持有效信息传递的同时，显著降低了计算复杂度。如图 4-5（b）所示，C2fCIB 模块用高效的 CIB 构建块取代了 C2f 模块中的所有瓶颈结构。通过这种设计，YOLOv10 在保

持较高检测精度的同时，显著降低了计算开销，实现了性能与效率的最佳平衡。这种改进不仅提升了模型的实时处理能力，还使其能够适应更广泛的实际应用场景，为复杂环境下的目标检测提供了更高效的解决方案。

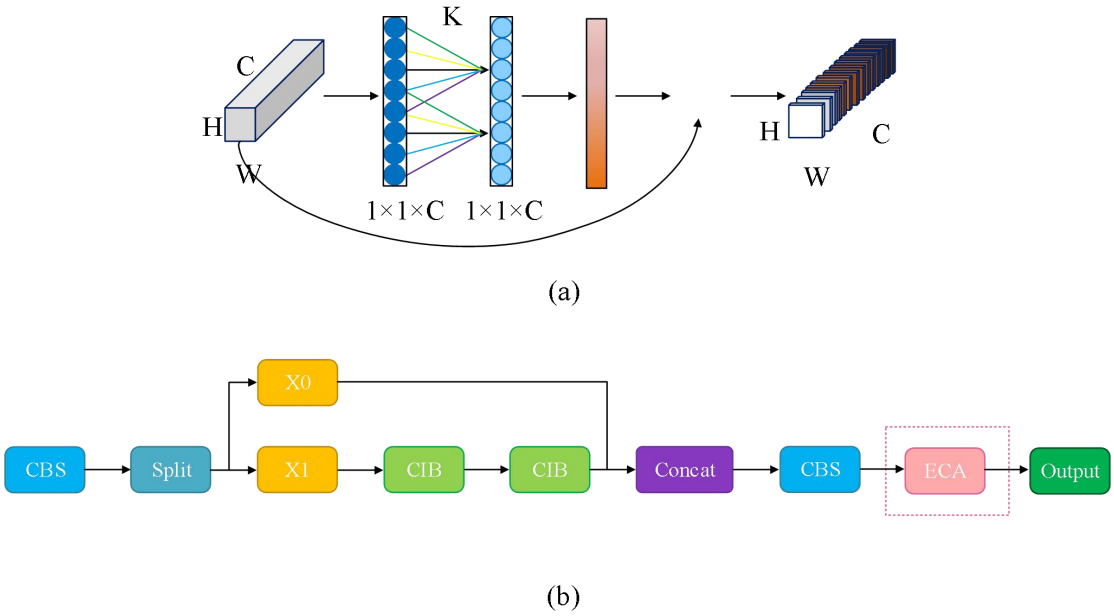


图 4-5 ECA 模块和 C2fCIB_ECA 模块的结构。(a) ECA 的结构，(b) C2fCIB_ECA 模块

4.4 归一化高斯瓦瑟斯坦距离（NWD）损失函数

YOLOv10 中使用的交并比（IoU）[78]指标在处理不同尺度的目标时表现出显著的敏感性差异。对于小至 6×6 像素的小目标而言，即使是微小的位置偏差也可能导致 IoU 值大幅下降，这会对标签分配的准确性产生负面影响，进而可能导致误检或漏检。相比之下，对于较大且更具代表性的目标（例如 36×36 像素的目标），相同的位置偏差只会使 IoU 发生相对较小的变化。因此，在实际应用中，传统 IoU 指标在处理小目标时的局限性尤为明显，这凸显了需要一个更强大的损失函数来提高小目标检测性能的必要性。

为应对这一挑战，提出了一种基于分布相似性的归一化高斯瓦瑟斯坦距离（NWD）[79]方法，该方法通过计算检测到的目标之间的分布相似性来评估它们之间的重叠程度。NWD 特别适用于小目标检测，因为它对目标尺度变化的敏感度较低，能够更准确地评估小目标之间的相似性。

NWD 损失函数的核心思想是将边界框（Bounding Box）建模为二维高斯分

布，通过计算两个高斯分布之间的瓦瑟斯坦距离来衡量它们的相似性。具体来说，NWD 方法采用了一个归一化过程，即将目标边界框的中心坐标和宽高作为二维高斯分布的均值和方差，并通过最优传输理论（Optimal Transport, OT）[80]计算这两个高斯分布之间的相似性。通过这种方式，NWD 能够减轻目标尺度变化对检测结果的影响，确保在不同大小的目标上都能保持较为一致的检测性能。

（1）将边界框建模为二维高斯分布

在归一化瓦瑟斯坦距离（NWD）方法中，检测到的目标的边界框被视为二维高斯分布。其中边界框中心具有边界框的位置和尺度（即宽度和高度）被建模为高斯分布的均值和方差，即中心像素具有最高权重，随着中心到边界重要性逐渐减小。具体来说，定义边界框 $B=(kx,ky,w,h)$，其内接椭圆方程与其高斯分布的概率密度函数如式（4-15），（4-16）和（4-17）所示：

$$\frac{(x-\varpi_x^2)}{o_x^2}+\frac{(y-\varpi_y^2)}{o_y^2}=1 \tag{4-15}$$

$$l(i,j)=\frac{2\alpha_i\alpha_j+b_1}{\alpha_i^2+\alpha_j^2+b_1} \tag{4-16}$$

$$f(x\mid\rho,\Sigma)=\frac{\exp\left(-\frac{1}{2}(x-\rho)^T\Sigma^{-1}(x-\rho)\right)}{(2\pi)^{\frac{d}{2}}\mid\Sigma\mid^{\frac{1}{2}}} \tag{4-17}$$

式中，(ϖ_x,ϖ_y)是椭圆的中心坐标。o_x,o_y为x与y轴的半轴长度，x、ρ和$\sum$表示坐标向量(x,y)、均值向量和协方差矩阵。当$(x-\rho)^T\sum^{-1}(x-\rho)=1$时，其为椭圆的 2D 高斯分布的密度轮廓。然后可以将水平边界框建模为 2D 高斯分布。水平边界框建模过程具体计算公式如式（4-18）所示：

$$\rho=\begin{bmatrix}kx\\ky\end{bmatrix},\sum=\begin{bmatrix}\frac{w^2}{4}&0\\0&\frac{h^2}{4}\end{bmatrix} \tag{4-18}$$

其中，ρ为高斯密度矩阵，矩阵描述了高斯分别的形状和方向。

此外，边界框 A 和 B 之间的相似性可以通过它们对应的高斯分布之间的距离来衡量，从而将几何上的重叠问题转化为分布相似性的计算问题。

（2）计算归一化高斯瓦瑟斯坦距离

在将边界框建模为高斯分布之后，归一化瓦瑟斯坦距离（NWD）会计算两个高斯分布之间的瓦瑟斯坦距离，以此来衡量它们的相似性。这个距离会被归一化处理，并通过一个指数函数转换为一个相似性度量指标，NWD 的具体公式如式（4-19）和（4-20）所示：

$$NWD(N_a,N_b)=exp(-\frac{\sqrt{W_2^2(N_a,N_b)}}{c}) \tag{4-19}$$

$$W_2^2(N_a,N_b)=\left\|\left(\left[cx_a,cy_a,\frac{w_a}{2},\frac{h_a}{2}\right]^T,\left[cx_b,cy_b,\frac{w_b}{2},\frac{h_b}{2}\right]^T\right)\right\|_2^2 \tag{4-20}$$

式中，$W_2^2(N_a,N_b)$表示两个高斯分布之间的瓦瑟斯坦距离，C是一个与数据集相关的常数。这一度量指标能够有效地处理目标大小的变化，使其适用于小目标的检测。

基于这一原理，本文引入了归一化瓦瑟斯坦距离（NWD）损失函数，以应对传统交并比（IoU）损失函数在小目标检测方面的局限性。具体来说，我们采用了一种加权混合损失函数，它将 NWD 损失和 IoU 损失相结合，二者的权重比为 0.5。通过上述做法，模型能够利用 NWD 在小目标检测上的优势，同时保持 IoU 在检测中大型目标时的出色性能。这种方法使网络能够更有效地学习不同目标尺度的特征信息，从而在复杂的现实场景中得到一个更强大、更可靠的检测器。

4.5 实验结果与分析

为了进一步验证 MDEN-YOLOv10 模型在行人检测方面的性能，我们使用基线模型 YOLOv10 和改进后的模型在公开数据集 Crowdhuman 上进行了实验。图 4-6 展示了两个模型在整个训练周期内的精确率、召回率、平均精度值（mAP_0.5）以及平均精度值（mAP_0.5:0.95）曲线。该图清晰地表明，MDEN-YOLOv10 模型的性能始终优于基线模型。改进后模型的最终精确率（P）和召回率（R）值分别为 85.5%和 68.4%，平均精度值（mAP_0.5）和平均精度值（mAP_0.5:0.95）分别为 83.5%和 47.3%，与基线模型 YOLOv10s 相比，mAP_0.5 和 mAP_0.5:0.95 分别提高了 4.4%、和 2.0%。

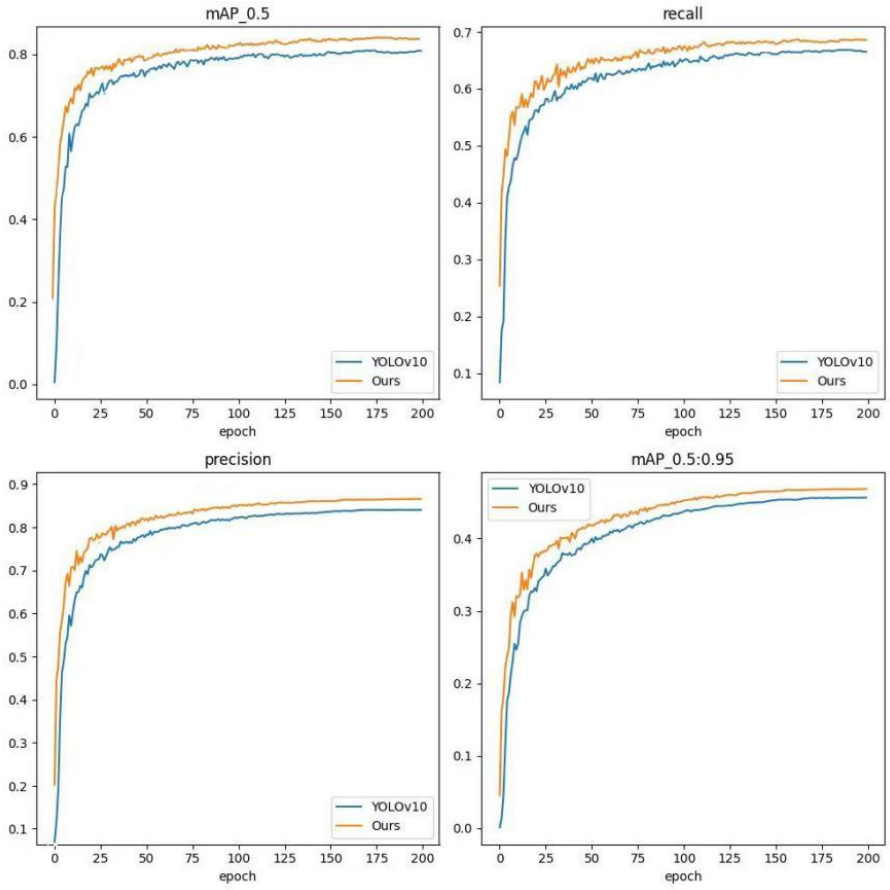


图 4-6 模型改进前后结果对比

模型通常部署在资源受限的边缘设备上。因此，在保持检测精度的同时，考虑模型的轻量化至关重要，这样才能确保在边缘计算平台上的检测速度，实现实时检测。我们从参数数量和浮点运算次数（FLOPs）方面对 MDEN-YOLOv10 模型和 YOLOv10 模型进行了比较。改进后的模型参数数量增加了 13.0%，浮点运算次数增加了 18.4%。总而言之，MDEN-YOLOv10 在保持原模型高精度的同时，提高了其对小目标的检测效率。尽管模型规模略有增大，但它仍然适用于实时行人检测的任务。

4.5.1 消融实验

在相同的实验条件下，使用相同的人群数据集进行了一项消融研究，以验证所提出的四项改进措施的有效性，并对检测性能进行对比分析。以 YOLOv10 作为基线模型，依次引入了四种改进方法，实验结果见表 4-1。方法 A 代表引入多尺度特征融合，方法 B 在 C2f 模块中融入了用于瓶颈融合的 DBB 模块，方法 C 将 ECA 注意力机制集成到 C2fCIB 模块中，方法 D 引入了归一化瓦瑟斯坦距离

（NWD）损失函数作为改进后的损失函数。

表 4-1 消融实验结果

算法	P/%	R/%	mAP_0.5%	模型大小/MB	FLOPs/G
YOLOv10	80.1	65.0	79.1	8.03	24.5
YOLOv10+A	81.8	65.7	81.0	8.81	27.1
YOLOv10+B	81.9	66.1	80.9	8.03	25.3
YOLOv10+C	81.2	65.8	80.0	8.20	25.0
YOLOv10+D	80.9	65.3	80.2	8.03	24.5
YOLOv10+A+B	84.5	67.5	82.7	8.84	28.0
YOLOv10+B+C	83.6	67.7	82.5	8.21	25.2
YOLOv10+C+D	82.9	66.6	82.1	8.20	25.0
YOLOv10+B+C+D	84.9	67.8	82.9	8.21	25.2
YOLOv10+A+B+D	85.0	68.1	82.8	8.84	28.0
YOLOv10+A+B+C	85.2	67.9	83.1	8.97	28.6
YOLOv10+A+B+C+D	85.5	68.4	83.5	9.10	29.0

如表 4-1 所示，添加 160×160 的特征尺度使模型能够从更大的特征图中捕获更多的局部信息，显著提高了对被遮挡目标的识别精度，将被遮挡目标的平均精度值（mAP_0.5）从 79.1%提升到了 81.0%。当同时引入 C2f_DBB 模块和多尺度特征提取时，DBB 模块进一步增强了网络的特征提取能力，提高了其捕获和融合细微生长状态变化目标中细粒度特征的能力，使得检测目标行人的平均精度值（mAP_0.5）提升到 82.7%，而参数数量仅略有增加。此外，融入轻量级且结构简单的 ECA 注意力机制，进一步提高了模型对相关特征的关注程度，将平均精度值（mAP_0.5）提升到 83.1%。最后，通过将归一化瓦瑟斯坦距离（NWD）损失与原始的交并比（IoU）损失相结合，充分发挥了两者的优势，引导模型更好地识别不同大小的目标，将检测目标行人的平均精度值（mAP_0.5）提升到 83.5%。当与其他改进措施相结合时，它进一步提升了模型的性能，从而有助于对不同大小的目标进行稳定检测。

当将这四个模块全部集成到网络中时，该模型实现了最佳的检测精度和最快的检测速度，与其他模型相比展现出了更优越的性能。我们为基线模型设计了一种新的结构，其中包括一个经过优化的特征提取网络、一种新颖的特征融合策略以及一种协同注意力机制，从而提升了该模型检测多尺度目标的能力。

4.5.2 对比实验

为进一步分析改进后的 MDEN-YOLOv10 算法性能，现选取当下主流的目标检测算法与之进行对比，相关对比结果如表 4-2 所示。

表 4-2 各检测模型对比实验结果

模型	mAP_0.5/%	mAP_0.5:0.95/%	参数量 /million	模型大小 /MB	FPS	FLOPs/G
SSD	72.6	37.6	26.8	45	62.3	30.54
Fast-Rcnn	74.8	39.2	137.3	71.3	8.1	90.7
RT-DETR[81]	73.9	39.8	28.5	20	68.0	108.0
YOLOv5n	76.8	41.5	2.0	3.9	114.9	18
YOLOv7-tiny	75.4	42.2	6.1	11.7	118.5	18.6
YOLOv8n	80.2	45.1	3.5	6.0	115.1	28.6
YOLOv9t	78.4	43.7	3.1	5.8	118.6	25.2
YOLOv10s	79.1	45.3	2.3	8.03	120.8	24.5
Ours	83.5	47.3	2.6	9.1	122.7	29.0

由表 4-2 可知：基于同一数据集，与传统目标检测算法 SSD、Fast-Rcnn 和 RT-DETR 相比，本文改进的 MDEN-YOLOv10 在检测精度（mAP_0.5/ mAP_0.5:0.95）、模型参数量和推理速度方面均展现出显著优势。本文的算法在 mAP_0.5 指标相较于 YOLOv7-tiny 模型提升 8.1 个百分点，mAP_0.5:0.95 综合指标增长 5.1%，同时实现了参数规模的显著缩减与检测速度的优化。相较于 YOLOv5n，改进算法在 mAP_0.5 上提高 6.7%，mAP_0.5:0.95 提升 5.8%；对比 YOLOv8n，mAP_0.5 增长 3.3%，mAP_0.5:0.95 提升 2.2%，同时保持了更快的推理速度；针对 YOLOv9t，mAP_0.5 提高 5.1%，mAP_0.5:0.95 增长 3.6%，在性能提升的同时实现了更优的效率平衡。相较于原 YOLOv10s，改进后的 MDEN-YOLOv10 算法在参数量、模型大小和检测速度均未受太大影响的情况下 mAP_0.5 上提升了 4.4%，在 mAP_0.5:0.95 上提升了 2.0%。综合分析可以得出，本文提出的算法优于以上行人检测算法，拥有更好的检测效果。

4.5.3 检测效果可视化及其分析

本实验选取了 CrowdHuman 数据集中三种代表性场景来展示 YOLOv10 与改进的 MDEN-YOLOv10 在人群密集，遮挡严重、背景复杂情况下的检测效果对比。

（a）Original　　（b）YOLOv10　　（c）Ours

图 4-7 人群密集场景检测效果对比

（a）Original　　（b）YOLOv10　　（c）Ours

图 4-8 严重遮挡场景检测效果对比

（a）Original　　（b）YOLOv10　　（c）Ours

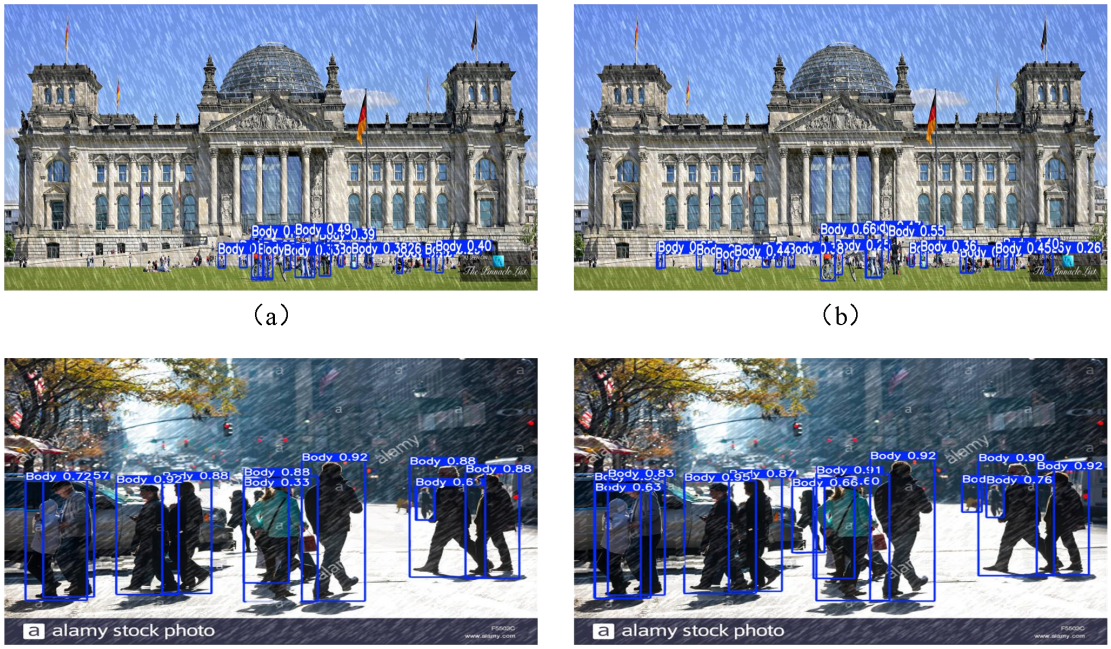
图 4-9 复杂背景场景检测效果对比

在图 4-7（a）人群密集和图 4-8（a）严重遮挡的场景中，对于遮挡程度较低的大尺度行人目标，这两种算法都展现出了良好的检测效果，检测框的置信度也比较高。然而，原算法在面对遮挡严重的小尺度行人目标时，检测效果不尽如人意，相比之下，改进后的算法能够更为精准地识别这类小尺度行人目标。在图 4-7（b）和图 4-7（c）中可以看出，在检测密集行人方面，两种算法都能做到准确检测，然而原算法会存在误检的情况，原算法误检了图片右侧大树树根的右侧垃圾桶与远处汽车车头为行人，与之不同的是，改进后的算法有效地对这类问题进行了改善,避免了误检。对比图 4-8（b）和图 4-8（c）可以看出，在面对远处被栏杆遮挡的小尺度行人检测问题上，原算法存在漏检的情况，而改进的算法则

是能够检测出来这些被遮挡的小目标行人。

在图 4-9（a）复杂背景的场景中，相比于原算法，改进后的算法降低了误检漏检率。对比图 4-9（b）和图 4-9（c）可以看出，在车辆众多且光影繁杂的夜间街道，原算法漏检了一位背着黑色双肩包的灰衣男子和一位骑着自行车的绿衣男子，改进后的算法展现出了更出色的效果，成功检测出了这两位男子。。

同样的，本文也基于自建数据集 Rain-CrowdHuman 进行了行人检测的实验，检测结果如图 4-10 所示。



（a）　（b）

（c）　（d）

图 4-10　雨天行人检测对比图

左图为 YOLOv10 算法检测结果，右图为我们改进的算法的检测结果，我们选取了两张复杂行人的图片进行检测。对比图 4-10（a）和图 4-10（b）可以看出，在雨天环境下，改进后的算法嫌贵与原 YOLOv10 算法在对小目标行人的检测能力上有所提高，虽然还是有对微小行人的漏检情况存在，但原算法左侧漏检的几个小目标行人，改进后的算法检测到了。而对比图 4-10（c）和图 4-10（d）可以看出，图片左边银灰色车辆头部位置的远处的一位行人以及图片中央偏右远处的一位穿黄色裙子的行人。改进后的算法展现出了更出色的效果，成功检测出了这两位行人。

综上所述，在人群密集、遮挡情况严重以及背景复杂的场景中，原 YOLOv10 算法存在较高的漏检误检率。而经过改进后的 MDEN-YOLOv10 算法能够改善这

种情况，能有效降低漏检误检率，且检测框置信度也相对更高。

4.6 雨天集成算法结构

本文将改进的去雨技术与优化的 MDEN-YOLOv10 目标检测模型进行融合，并在专门针对雨天场景设计的 Rain-CrowdHuman 数据集上进行了系统的对比分析。实验结果表明，这种技术融合不仅有效改善了雨天环境下的图像质量，还显著提升了行人检测的准确率和鲁棒性，为复杂天气条件下的行人检测提供了一种切实可行的技术方案。

4.6.1 检测效果可视化及其分析

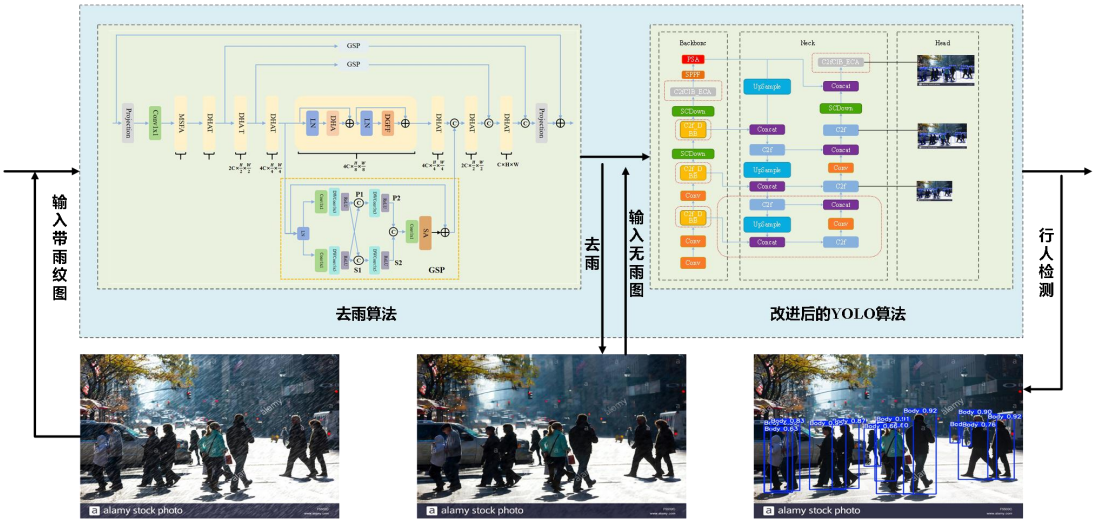


图 4-11 雨天行人检测集成算法结构图

雨天环境下的行人检测集成算法如图 4-11 所示。本文提出的集成算法检测网络采用分阶段处理流程：首先，将自建数据集 Rain-CrowdHuman 中的原始图像输入到第三章设计的去雨网络中进行信号增强处理，通过去雨操作有效恢复图像细节，输出高质量清晰图像；随后，将这些经过预处理后的清晰图像输入到第四章提出的基于 YOLOv10 的改进行人检测算法中，利用优化后的检测网络准确识别和定位行人目标。

4.6.2 对比实验思路

该部分使用自建数据集 Rain-Crowdhuman 作为测试数据集，采用原

YOLOv10s 行人检测算法、原数据增强去雨算法结合原 YOLOv10s 行人检测算法与本文提出的基于改进的数据增强去雨算法结合改进的 MDEN-YOLOv10 行人检测算法进行对比实验。该实验的对比框架展示图如图 4-12 所示。

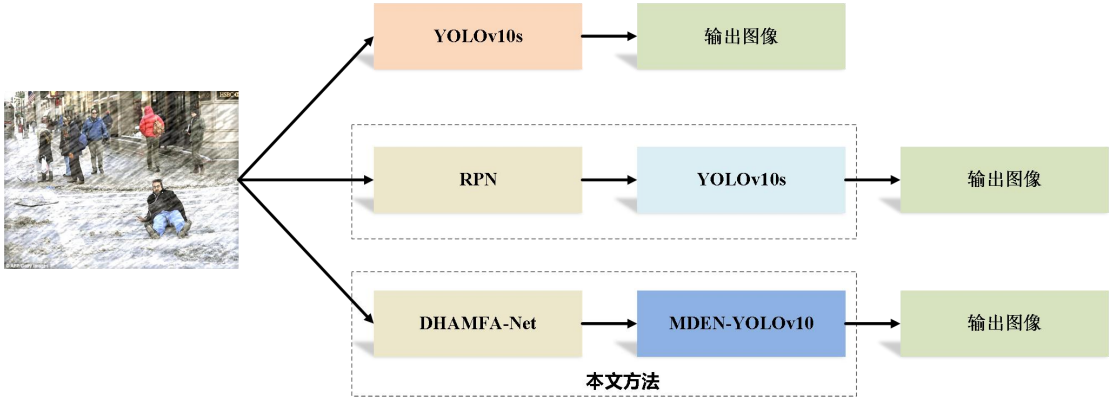


图 4-12 对比实验框架

4.6.3 实验结果与分析

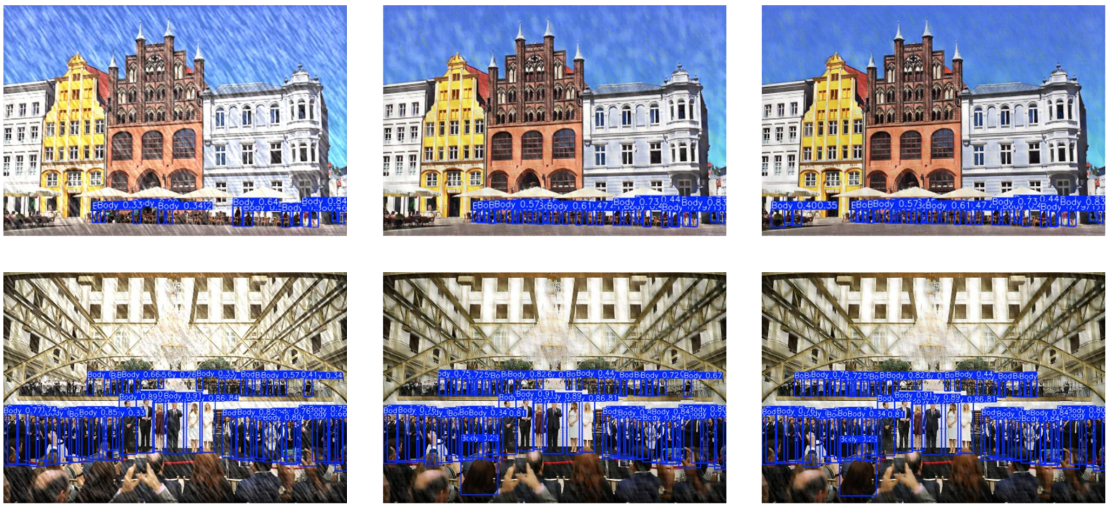
以上一节的对比实验框架在自建数据集 Rain-Crowdhuman 上进行 3 组对比实验。实验结果见表 4-3。

表 4-3 检测精度对比

算法	P/%	R/%	mAP_0.5/%	mAP_0.5:0.95/%
YOLOv10s	77.1	61.2	76.8	43.1
YOLOv10s+RPN	79.0	63.7	78.5	44.9
Ours	83.7	65.7	82.4	46.8

由表 4-3 可以分析出，相比于单独的 YOLOv10、YOLOv10+RPN 去雨算法。本文的行人检测算法在 mAP_0.5 上分别提升了 5.6%和 3.9%，在 mAP_0.5:0.95 上分别提升了 3.7%和 1.9%。证明了本文方法的有效性和可行性。

同时，对以上 3 组实验进行可视化分析，其检测结果对比图如图 4-13 所示。实验结果表明，不管是置信框精度的提升还是检测出漏检的小目标行人，都表明本文的行人检测算法明显优于其他 2 组算法。



YOLOv10　　YOLOv10+去雨　　MDEN-YOLOv10+本文去雨

图 4-13 检测效果对比图

4.7 本章小结

本章聚焦密集场景下行人检测难题，以 YOLOv10s 为基础提出改进的 MDEN-YOLOv10 算法，本章从多尺度特征融合、模块优化、注意力机制引入以及损失函数改进等方面对网络进行改进，并在公开数据集 CrowdHuman 上进行验证，改进算法的 mAP_0.5 和 mAP_0.5:0.95 分别达到 83.5%和 47.3%，相比原算法提升 4.4%和 2.0%，尽管模型规模略有增大，但它仍然适用于实时行人检测的任务。消融实验验证了各项改进措施的有效性，对比实验显示该算法在多指标上优于其他主流检测算法，在密集场景下检测效果更优。同时，在自建数据集 Rain-CrowdHuman 上的验证行人检测效果，也同样取得了更高的检测精度。最后，集成图像去雨+密集行人检测算法，把经过去雨处理的图像作为改进的 MDEN-YOLOv10 算法的输入，再进行行人检测，相较于未经过去雨处理的图像、未经过改进的去雨算法+未经过改进的 YOLOv10s 算法，本章的集成算法体现出更优的检测效果。

第 5 章 总结与展望

5.1 本文总结

行人检测作为自动驾驶与环境感知的关键技术，其可靠性直接影响智能驾驶系统在雨天场景下的决策安全性。尽管深度学习的出现推动该领域取得显著进展，但雨线干扰引发的图像质量降低，导致行人检测模型准确率降低。同时在行人密集场景下，行人之间会存在严重遮挡等问题，这同样会影响行人目标的检测精确度。本文主要针对恶劣天气（雨天）及拥挤行人场景下的行人检测方法进行研究，利用深度学习技术提升来提升行人目标的检测性能，同时针对雨天环境进行图像去雨处理，并进行多组实验来验证分析其可靠性。本文的主要工作如下：

（1）提出一种基于动态直方图注意力机制与多特征聚合的图像去雨方法。该方法主要通过以下三个模块达到去雨的效果：多特征聚合模块（MSFA）通过多尺度卷积提取图像特征，捕捉雨线的细微结构和复杂背景；全局空间感知模块（GSP）利用深度可分离卷积和空间注意力机制，增强特征间的空间关系感知，突出雨线区域并抑制噪声；动态直方图注意力模块（DHAT）通过动态直方图注意力机制捕捉特征关联并进行变换，有效去除雨线噪声并恢复细节。通过对比实验表明，该算法显著提升了去雨效果和图像质量。

（2）提出一种改进的 MDEN-YOLOv10 行人检测网络。第一步，在骨干网络中添加一个 160×160 的特征尺度层来增强多尺度特征融合网络，提高模型对小局部特征的敏感度，特别是对被遮挡行人的检测能力。第二步，高效通道注意力机制（ECA）集成到骨干网络的 C2fCIB 卷积模块中，以提高网络对输入图像中感兴趣区域的关注度。第三步，引入了多样分支块（DBB）模块来提升网络对多尺度行人的特征提取能力。第四步，提出了归一化瓦瑟斯坦距离（NWD）损失函数，以有效减少对密集排列和高度重叠的目标行人的漏检情况。最后进行消融实验与对比实验，证明了该优化方案的有效性和可行性。

5.2 下一步工作展望

本文主要针对复杂天气（雨天）与拥挤情况下的行人检测，设计了一种图像去雨算法，并对基于 YOLOv10 的行人检测算法进行了改进与优化，实验体现了其具有一定的效果，但任然存在一些不足之处，可以在以下方面进行改进：

（1）本文只是研究了复杂天气下的一种雨天天气，在以后的研究中可以加入其他复杂天气，例如雾天，雪天，甚至沙尘暴这种极端天气。

（2）本文检测目标单一，仅为行人目标，在后续的研究中可以加入其他交通主体，例如车辆，指示牌，红绿灯等，并同时进行检测。

（3）本文实验用的数据集为合成的雨天环境数据集，虽然已经尽可能接近真实雨天的环境，但还是会与真实雨天环境存在偏差。在后续的实验中可以加入真实雨天的环境来训练，并验证算法的有效性与可行性。

（4）本文设计的行人检测网络侧重于精确度的提高，下一步可以考虑推进一下轻量化的研究，能够使其在边缘设备上部署并工作，使其有拥有实际应用的价值。

参考文献

[1] 王士林, 董猛. 国内城市道路建设发展综述[J]. 城市道桥与防洪, 2024, (11): 1-3+16.

[2] 牛苗苗, 张宏芳, 王莹. 高影响天气下的公路交通气象研究进展[J]. 陕西气象, 2025, (01): 34-41.

[3] 齐晨, 孙广林. 易受恶劣天气影响主要通道交通安全管理对策研究[J]. 道路交通管理, 2023, (11): 22-25.

[4] Guo M H, Xu T X, Liu J J, et al. Attention mechanisms in computer vision: A survey[J]. Computational visual media, 2022, 8(3): 331-368.

[5] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). IEEE, 2005, 1: 886-893.

[6] 罗艳, 张重阳, 田永鸿, 等. 深度学习行人检测方法综述[J]. 中国图象图形学报, 2022, 27(07): 2094-2111.

[7] Wang X, Han T X, Yan S. An HOG-LBP human detector with partial occlusion handling[C]//2009 IEEE 12th international conference on computer vision. IEEE, 2009: 32-39.

[8] Huang S, Cai N, Pacheco P P, et al. Applications of support vector machine (SVM) learning in cancer genomics[J]. Cancer genomics & proteomics, 2018, 15(1): 41-51.

[9] Felzenszwalb P F, Girshick R B, McAllester D, et al. Object detection with discriminatively trained part-based models[J]. IEEE transactions on pattern analysis and machine intelligence, 2009, 32(9): 1627-1645.

[10] 罗森林,赵惟肖,潘丽敏.结合加权 KNN 和自适应牛顿法的稳健 Boosting 方法[J].北京理工大学学报,2021,41(01):112-120.

[11] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). Ieee, 2005, 1: 886-893.

[12] Fang Q, Jiang A, Liu M, et al. Faster R-CNN model for target recognition and diagnosis of scapular fractures[J]. Journal of Bone Oncology, 2025, 51: 100664.

[13] Kang S, Hu Z, Liu L, et al. Object Detection YOLO Algorithms and Their Industrial Applications: Overview and Comparative Analysis[J]. Electronics, 2025, 14(6): 1104.

[14] Peng X, Schmid C. Multi-region two-stream R-CNN for action detection[C] //European conference on computer vision. Cham: Springer International Publishing, 2016: 744-759.

[15] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.

[16] 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.

[17] Zhang K, Xiong F, Sun P, et al. Double anchor R-CNN for human detection in a crowd[J]. arXiv preprint arXiv:1909.09998, 2019.

[18] Tian Y, Luo P, Wang X, et al. Deep learning strong parts for pedestrian detection[C] //Proceedings of the IEEE international conference on computer vision. 2015: 1904-1912.

[19] Song Y, Qian P, Zhang K, et al. An improved Multi-Scale Fusion and Small Object Enhancement method for efficient pedestrian detection in dense scenes[J]. Multimedia Systems, 2025, 31(2): 1-16.

[20] Liu Q, Ye H, Wang S, et al. YOLOv8-CB: dense pedestrian detection algorithm based on in-vehicle camera[J]. Electronics, 2024, 13(1): 236.

[21] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.

[22] Mao M, Hong M. YOLO Object Detection for Real-Time Fabric Defect Inspection in the Textile Industry: A Review of YOLOv1 to YOLOv11[J]. Sensors (Basel, Switzerland), 2025, 25(7): 2270.

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

[24] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.

[25] Kumar G S, Maheshwari G U, Selvan C, et al. Deep learning approach using modified DarkNet-53 for renal cell carcinoma grading[J]. International Journal of Bioinformatics Research and Applications, 2025, 21(1): 1-25.

[26] Li Y, Li S, Du H, et al. YOLO-ACN: Focusing on small target and occluded object detection[J]. IEEE access, 2020, 8: 227288-227303.

[27] Wang G, Ding H, Yang Z, et al. TRC‐YOLO: A real‐time detection method for lightweight targets based on mobile devices[J]. IET Computer Vision, 2022, 16(2): 126-142.

[28] Zhong J, Qian H, Wang H, et al. Improved real-time object detection method based on YOLOv8: a refined approach[J]. Journal of Real-Time Image Processing, 2025, 22(1): 1-13.

[29] Ang G, Xing Z L, Chen X X, et al. A dense pedestrian detection algorithm with improved YOLOv8[J]. Journal of graphics, 2023, 44(5): 890-898.

[30] Li N, Bai X, Shen X, et al. Dense pedestrian detection based on GR-YOLO[J]. Sensors, 2024, 24(14): 4747.

[31] Deng S, Li J. Efficient dense pedestrian detection based on transformer[C]//2024 IEEE 7th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC). IEEE, 2024, 7: 992-997.

[32] Hussain M. YOLO-v1 to YOLO-v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection[J]. Machines, 2023, 11(7): 677.

[33] Li Q P, Zhang J W, Zhao Y, et al. A YOLOX object detection algorithm based on bidirectional cross-scale path aggregation[J]. Neural Processing Letters, 2024, 56(1): 35.

[34] Sirisha U, Praveen S P, Srinivasu P N, et al. Statistical analysis of design aspects of various YOLO-based deep learning models for object detection[J]. International Journal of Computational Intelligence Systems, 2023, 16(1): 126.

[35] Su P, Han H, Liu M, et al. MOD-YOLO: Rethinking the YOLO architecture at the level of feature information and applying it to crack detection[J]. Expert Systems with Applications, 2024, 237: 121346.

[36] Xue Z, Xu R, Bai D, et al. YOLO-tea: A tea disease detection model improved by YOLOv5[J]. Forests, 2023, 14(2): 415.

[37] Wang A, Chen H, Liu L, et al. Yolov10: Real-time end-to-end object detection[J]. Advances in Neural Information Processing Systems, 2024, 37: 107984-108011.

[38] Kang L W, Lin C W, Fu Y H. Automatic single-image-based rain streaks removal via image decomposition[J]. IEEE transactions on image processing, 2011, 21(4): 1742-1755.

[39] Santhaseelan V, Asari V K. Utilizing local phase information to remove rain from video[J]. International Journal of Computer Vision, 2015, 112: 71-89.

[40] Ding X, Chen L, Zheng X, et al. Single image rain and snow removal via guided L0 smoothing filter[J]. Multimedia Tools and Applications, 2016, 75: 2697-2712.

[41] Zhu L, Fu C W, Lischinski D, et al. Joint bi-layer optimization for single-image rain streak removal[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2526-2534.

[42] Li Y, Tan R T, Guo X, et al. Rain streak removal using layer priors[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2736-2744.

[43] Li R, Cheong L F, Tan R T. Heavy rain image restoration: Integrating physics model and conditional adversarial learning[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 1633-1642.

[44] Ren D, Zuo W, Hu Q, et al. Progressive image deraining networks: A better and simpler baseline[C]// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 3937-3946.

[45] Wu W, Huo L, Yang G, et al. Research into the Application of ResNet in Soil: A Review[J]. Agriculture, 2025, 15(6): 661.

[46] Qin X, Wang Z, Bai Y, et al. FFA-Net: Feature fusion attention network for single image dehazing[C]// Proceedings of the AAAI conference on artificial intelligence. 2020, 34(07): 11908-11915.

[47] Liu X, Zhou J. Short-term wind power forecasting based on multivariate/multi-step LSTM with temporal feature attention mechanism[J]. Applied Soft Computing, 2024, 150: 111050.

[48] Hu X, Fu C W, Zhu L, et al. Depth-attentional features for single-image rain removal[C]// Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition. 2019: 8022-8031.

[49] Kim S, Ahn N, Sohn K A. Restoring spatially-hetero-geneous distortions using mixture of experts network[C]// Asian Conference on Computer Vision. Cham: Springer,2021: 185-201.

[50] Chen X, Li H, Li M, et al. Learning a sparse transformer network for effective image deraining[C]// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 5896-5905.

[51] Wang Z, Cun X, Bao J, et al. Uformer: A general u-shaped transformer for image restoration[C]// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 17683-17693.

[52] Chen X, Pan J, Dong J. Bidirectional multi-scale implicit neural representations for image deraining[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 25627-25636.

[53] Xiao J, Fu X, Liu A, et al. Image de-raining transformer[J]. IEEE transactions on pattern analysis and machine intelligence, 2022, 45(11): 12978-12995.

[54] Gao H, Zhang Y, Yang J, et al. Mixed hierarchy network for image restoration[J]. Pattern Recognition, 2025, 161: 111313.

[55] Wei W, Meng D, Zhao Q, et al. Semi-supervised transfer learning for image rain removal[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 3877-3886.

[56] Jin X, Chen Z, Lin J, et al. Unsupervised single image deraining with self-supervised constraints[C]//2019 IEEE International Conference on Image Processing (ICIP). IEEE, 2019: 2761-2765.

[57] Zhu H, Peng X, Zhou J T, et al. Singe image rain removal with unpaired information: A differentiable programming perspective[C]//Proceedings of the AAAI conference on artificial intelligence. 2019, 33(01): 9332-9339.

[58] Zhao X, Wang L, Zhang Y, et al. A review of convolutional neural networks in computer vision[J]. Artificial Intelligence Review, 2024, 57(4): 99.

[59] Ramapuram J, Danieli F, Dhekane E, et al. Theory, analysis, and best practices for sigmoid self-attention[J]. arXiv preprint arXiv:2409.04431, 2024.

[60] De Ryck T, Lanthaler S, Mishra S. On the approximation of functions by tanh neural networks[J]. Neural Networks, 2021, 143: 732-750.

[61] Mirzadeh I, Alizadeh K, Mehta S, et al. Relu strikes back: Exploiting activation sparsity in large language models[J]. arXiv preprint arXiv:2310.04564, 2023.

[62] Qiu Q, Zhu T, Gong H, et al. Relu-kan: New kolmogorov-arnold networks that only need matrix addition, dot multiplication, and relu[J]. arXiv preprint arXiv:2406.02075, 2024.

[63] Jian J, Xia W, Zhang R, et al. Multiple instance convolutional neural network with modality-based attention and contextual multi-instance learning pooling layer for effective differentiation between borderline and malignant epithelial ovarian tumors[J]. Artificial intelligence in medicine, 2021, 121: 102194.

[64] Bhavana N, Kodabagi M M, Kumar B M, et al. POT-YOLO: Real-Time Road Potholes Detection using Edge Segmentation based Yolo V8 Network[J]. IEEE Sensors Journal, 2024，24(15)：24802 – 24809.

[65] Qiu X, Chen Y, Cai W, et al. LD-YOLOv10: A lightweight target detection algorithm for drone scenarios based on YOLOv10[J]. Electronics, 2024, 13(16): 3269.

[66] Wang H, Zhang Y, Zhu C. YOLO-LFD: A Lightweight and Fast Model for Forest Fire Detection[J]. Computers, Materials & Continua, 2025, 82(2).

[67] Yu H, Li X, Feng Y, et al. Multiple attentional path aggregation network for marine object detection[J]. Applied intelligence, 2023, 53(2): 2434-2451.

[68] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. Springer international publishing, 2015: 234-241.

[69] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [C]//Proc of the 31st Annual Conference on Neural Information Processing Systems. Cambridge: MIT Press: 2017: 6000-6010.

[70] Chhapariya K, Ientilucci E, Buddhiraju M K, et al. Target Detection and Characterization of Multi-Platform Remote Sensing Data[J]. Remote Sensing, 2024, 16(24): 4729-4729.

[71] Zhang D, Zhou H, Zhou T, et al. Using 2D U-Net convolutional neural networks for automatic acetabular and proximal femur segmentation of hip MRI images and morphological quantification: a preliminary study in DDH[J]. Bio Medical Engineering OnLine,2024,23(1):98-98.

[72] Wang X, Girshick R, Gupta A, et al. Non-local neural networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. IEEE, 2018: 7794-7803.

[73] Shao S, Zhao Z, Li B, et al. Crowdhuman: A benchmark for detecting human in a crowd[J]. arXiv preprint arXiv:1805.00123, 2018.

[74] Mehri A, Ardakani P B, Sappa A D. MPRNet: Multi-path residual network for lightweight image super resolution[C]. Waikoloa: IEEE Winter Conference on Applications of Computer Vision, 2021: 2703-2712.

[75] Kwon H J, Lee S H. Raindrop-removal image translation using target-mask network with attention module[J]. Mathematics, 2023, 11(15): 3318.

[76] Ding X, Zhang X, Han J, et al. Diverse Branch Block: Building a Convolution as an Inception-like Unit[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA: IEEE, 2021: 10881-10890.

[77] Wang Q, Wu B, Zhu P, et al. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020: 11531-11539.

[78] Zhang Y F, Ren W, Zhang Z, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. Neurocomputing, 2022, 506: 146-157.

[79] Wang J, Xu C, Yang W, et al. Detecting tiny objects in aerial images： A normalized Wasserstein distance and a new benchmark[J]. ISPRS Journal of Photogrammetry and Remote Sensing，2022，190：79-93.

[80] Courty N, Flamary R, Tuia D, et al. Optimal transport for domain adaptation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(9): 1853-1865.

[81] Peng H, Chen S. FedsNet: the real-time network for pedestrian detection based on RT-DETR [J]. Journal of Real-Time Image Processing, 2024, 21 (4): 142-142.

攻读硕士期间取得的学术成果

[1] 崔福荣, 韩超. 基于YOLOv10的拥挤人群行人检测算法研究[J]. 安徽工程大学学报, 2026, 41(1).（已录用）

致谢

在研究生生涯的三年里，本人收获良多。

首先，我要向我的导师韩超教授表达诚挚的感谢与敬意。从选题构思到最终定稿，导师都给予了我不间断的关怀和指导。导师更是以身作则，以其严谨的治学精神、深厚的专业知识和高尚的人格魅力深深感染着我，激励着我在学术的道路上不断前行。

同时，我也要向在我论文写作过程中给予我支持和帮助的老师、同学以及朋友们表达深深的感谢。我的论文不断地完善与提高，离不开他们一次次的建议与意见。同时，这样一段一起互相讨论、互相支持、互相鼓励的科研生涯，也会成为我心底一抹永生难忘的奋斗时光。

此外，我还要感谢我的家人，他们的理解与支持就是我最坚实的后盾。他们在生活上的默默付出和精神上的鼓励，使我能心无旁骛地投入到学术研究中。亲情的包容与信任，是我不断前行的动力。

最后，我要感谢我的母校安徽工程大学提供的优质学术平台。先进的实验设备、丰富的文献资源以及跨学科的交流机会，都为我课题的研究创造了良好条件，让我能够在这里更好地完成我的三年硕士学业。在未来的日子里，我将谨记母校“尚德敏学 唯实惟新”的校训，继续努力，不断学习，为以后的学术事业奉献自己的一份力量。

再次感谢所有帮助与支持过我的人，我将永远铭记，并带着这份感恩的心，再次出发。

尚德教学　唯实惟新