

基于级联乘性融合的夜间多尺度行人检测网络

蔡华健^{1,2,3}, 李俊^{1,2,3}, 李琦铭^{1,2,3}

(1. 福建师范大学 计算机与网络空间安全学院, 福州 350117;

2. 中国科学院 福建物质结构研究所, 福州 350108;

3. 中国科学院 海西研究院泉州装备制造研究中心, 泉州 362200)

摘要: 针对夜间低光照环境下行人检测面临的多尺度特征捕获困难和夜间背景噪声干扰等问题, 提出了一种双流乘性融合网络 DMFNet; 该网络结合 CMF 模块和 DCA 模块, 有效解决了夜间场景中行人尺度差异与夜间噪声的关键问题; CMF 模块采用自顶向下的乘性特征融合机制, 通过元素级乘法操作实现了跨尺度信息的高效整合, 有效抑制了背景噪声干扰; DCA 模块采用一维卷积和全连接层构建双路径注意力机制, 通过可学习的混合因子自适应调整特征通道权重, 实现了对行人区域特征的精准增强; 实验结果表明, DMFNet 在 NightOwls 数据集上的性能超越现有主流方法, Reasonable 子集漏检率为 6.85%, 检测精度更高且鲁棒性更强, 能够有效检测夜间复杂场景中的行人目标。

关键词: 夜间行人检测; 乘性连接; 特征融合; 通道注意力; 多尺度; 低光照检测

Nighttime Multi-scale Pedestrian Detection Network Based on Cascaded Multiplicative Fusion

CAI Huajian^{1,2,3}, LI Jun^{1,2,3}, LI Qiming^{1,2,3}

(1. College of Computer and Cyber Security, Fujian Normal University, Fuzhou 350117, China;

2. Fujian Institute of Research on the Structure of Matter, CAS, Fuzhou 350108, China;

3. Quanzhou Institute of Equipment Manufacturing, Haixi Institute, CAS, Quanzhou 362200, China)

Abstract: To address the challenges of multi-scale feature capture and severe background noise interference in nighttime low-light pedestrian detection, this paper proposes a dual-flow multiplicative fusion network (DMFNet). This network integrates a cascaded multiplicative fusion (CMF) module and a dual-flow channel attention (DCA) module to effectively solve the issues of pedestrian scale variation and nighttime noise. Specifically, the CMF module adopts a top-down multiplicative mechanism, which integrates cross-scale information through element-wise multiplication, thereby suppressing background noise. Meanwhile, the DCA module constructs a dual-path attention mechanism by using one-dimensional convolution and fully connected layers. By employing a learnable mixing factor to adaptively adjust feature channel weights, the DCA module accurately enhances the feature in pedestrian regions. Experimental results on the NightOwls dataset demonstrate that the DMFNet outperforms existing mainstream methods, achieving a missed detection rate of 6.85% on the Reasonable subset. The proposed method exhibits superior detection accuracy and robustness, proving its effectiveness in detecting pedestrians within complex nighttime scenes.

Keywords: nighttime pedestrian detection; multiplicative connection; feature fusion; channel attention; multi-scale; low-light detection

收稿日期:2025-12-12; 修回日期:2026-01-29。

基金项目:福建省自然科学基金项目(2025J01248);泉州市科技计划项目(2024QZC001R);福建省科技计划项目(2025T3027, 2025T3023)。

作者简介:蔡华健(2001-),男,硕士。

通讯作者:李琦铭(1987-),男,博士,副研究员。

引用格式:蔡华健,李俊,李琦铭. 基于级联乘性融合的夜间多尺度行人检测网络[J]. 计算机测量与控制, 2026, 34(7): 268-275, 309.

0 引言

行人检测是目标检测中一个重要的研究方向^[1], 在智能交通与安防监控^[2]中应用广泛。该技术旨在通过检测算法对行人进行识别, 并输出相应的位置信息^[3], 为行人重识别、行人跟踪等更高层次的视觉任务提供基础支持。

近年来, 深度学习推动了基于卷积神经网络 (CNN, convolutional neural network) 的行人检测方法在日间场景取得显著进展^[4-6]。但在夜间低光照条件下, 由光照不足与背景噪声导致的多尺度表征困难严重影响了检测性能^[7], 现有方法仍难满足实际需求。

针对这些挑战, 现有研究主要分为两类: 多模态方法利用红外图像补充信息, 如 MBNet^[8]通过差分模态感知融合模块实现可见光与红外特征的互补; 文献 [9] 引入模态间和模态内注意力机制实现不同模态特征的自适应融合; DCMNet^[10]则结合局部和非局部互补信息实现多模态特征互补, 但这类方法存在模态对齐困难与硬件成本高等限制。单模态方法则基于可见光图像, 文献 [11] 构建了 NightOwls 夜间数据集; TFAN^[12]利用时序信息来应对检测中的遮挡问题, 但依赖时序数据且影响实时性; LSFN^[13]提出超像素金字塔池化融合网络, 但其性能依赖于预训练分类模型 HRNet^[14]; FATNet^[15]设计特征注意与变换模块增强夜间特征, 但忽略了低对比度问题, 且采用两阶段训练流程复杂。FB-CA^[16]通过前后景对比注意力机制改善低对比度问题, 但依赖手动设计的先验信息, 在复杂背景下泛化能力有限。此外, 现有的通用多尺度融合 (如 FPN^[17]、PANet^[18]) 常因加法或拼接操作放大夜间噪声, 并且常规注意力机制 (如 CBAM^[19]、CA^[20]) 缺乏对夜间光照的自适应调节。因此, 本文提出了一种双流乘性融合网络

(DMFNet, dual-flow multiplicative fusion network), 通过级联乘性融合模块 (CMF, cascaded multiplicative fusion) 和双流通道注意力模块 (DCA, dual-flow channel attention) 有效解决夜间多尺度行人检测中的关键技术难题。具体而言, CMF 模块通过自顶向下的乘性特征融合, 利用深层语义特征作为门控权重对浅层细节特征进行引导, 通过元素级乘法实现跨尺度信息整合, 减轻夜间背景噪声干扰; 并设计 DCA 模块采用一维卷积与全连接两条路径构建双路径通道注意力, 分别建模通道间的局部依赖与全局依赖以应对夜间全局光照感知和局部噪声抑制的双重挑战, 通过可学习的混合因子自适应融合局部与全局依赖, 增强夜间多尺度特征的判别能力。

实验结果表明, DMFNet 在公开数据集上的性能显著优于现有主流算法, 且在日夜间场景下均展现出良好的鲁棒性与泛化能力, 为自动驾驶及智能监控提供了有效的技术支撑。

1 双流乘性融合网络设计

针对夜间行人尺度差异与光照差异, 本文提出了一种基于双流乘性融合的夜间行人检测网络, 通过创新的特征融合机制和通道注意力设计, 实现对夜间多尺度行人目标的精准检测。网络结构如图 1 所示, DMFNet 的整体架构由特征提取骨干网络、双流乘性融合模块和检测头 3 个核心部分构成。

在特征提取阶段, DMFNet 采用高分辨率网络 HRNet-W32 作为骨干网络, 该网络通过并行多分辨率子网络设计, 在整个特征提取过程中始终保持高分辨率表征, 避免了传统网络中因多次下采样导致的空间细节丢失问题。对于输入的夜间图像 $I \in R^{H \times W \times 3}$, HRNet 输出 4 个不同尺度的特征图 $\{F_1, F_2, F_3, F_4\}$, 它们

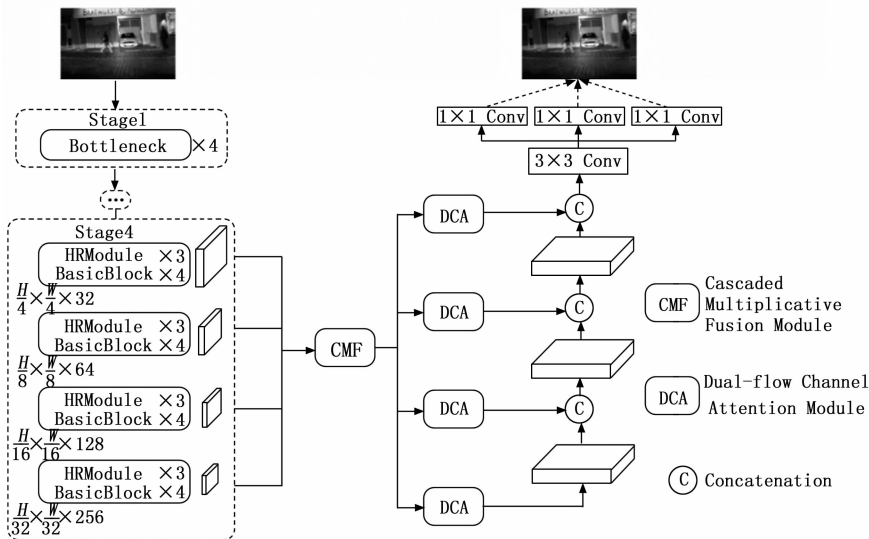


图 1 基于双流乘性融合的夜间行人检测网络

的空间分辨率依次为原图的 1/4、1/8、1/16、1/32，通道数分别为 32、64、128、256。这种多尺度特征金字塔结构呈现出明显的层次化特性，其中浅层特征 F_1 保留了丰富的空间细节和边缘信息，适合捕获小尺度行人，而深层特征 F_4 则蕴含高层语义和全局上下文，有利于识别大尺度行人并抑制夜间背景干扰。

然而，传统的特征融合方法如加法或拼接操作在整合多尺度信息时，往往会同时放大有用信号和背景噪声，且缺乏对通道重要性的自适应调节能力。针对这些问题，本文设计了双流乘性融合模块，该模块由级联乘性融合模块和双流融合通道注意力模块协同工作，通过自顶向下的传递机制与乘性融合策略实现深层语义对浅层细节的引导作用，并完成空间与通道双维度的特征优化。在检测阶段，DMFNet 采用基于中心点的无锚框检测头，通过 3 个并行分支分别预测行人中心点热图、目标尺度和中心点偏移量，通过后处理过程，筛除冗余的边界框，最终生成精准的检测边界框。

1.1 级联乘性融合模块

在夜间行人检测任务中，模型面临着目标尺度差异显著与环境噪声干扰严重的双重挑战。首先，夜间行人的尺度跨度极大，而卷积神经网络固有的下采样操作导致深层特征丢失了小目标的边界细节，难以满足精准定位的需求。其次，浅层特征虽保留了高分辨率的空间信息，但在低光照条件下往往混杂着大量背景噪声，导致单一层级的特征难以兼顾语义识别与细节捕获。

传统解决方案通常采用高分辨率输入策略，即使用更高分辨率的图像以获取更大尺寸的特征图，从而保留更丰富的行人细节。然而，这种方法存在明显缺陷：一方面，它不可避免地导致计算量和显存占用激增，严重限制了算法在实际应用中的部署效率；另一方面，简单提高分辨率并不能从根本上解决夜间低对比度、高噪声环境下的特征提取难题，因为高分辨率输入可能同时放大背景噪声和无用信号，反而干扰模型对真正有效行人特征的识别。

针对这些问题，本文提出了级联乘性融合模块，该模块通过精心设计的特征交互机制实现深浅层特征的高效互补。CMF 模块采用自顶向下的乘法连接策略，将深层特征中蕴含的语义信息（高抽象、低分辨率）精准注入到浅层特征（高分辨率、低语义）中，形成一种由深至浅的信息流动路径。利用深层语义特征作为引导信号，指导浅层特征关注真正的行人区域，同时抑制背景噪声的干扰。通过这种方式，CMF 模块不仅能够有效保留和增强关键的行人区域特征，还能在维持较低计算成本的前提下，显著提升夜间行人检测的精度和鲁棒性。

在深度学习的特征融合机制中，乘性连接相较于传

统的加法或拼接方式具有显著优势。为了直观阐述其原理，我们将包含丰富细节但混杂夜间噪声的浅层特征 F_{low} 建模为有效目标信号 S 与背景噪声 N 的叠加（即 $F_{\text{low}} = S + N$ ）。传统的加法融合（ $Y_{\text{add}} = F_{\text{high}} + F_{\text{low}}$ ）在整合特征时，会将背景噪声 N 无差别地保留并传递至下一层，同时放大有用信号与背景噪声，导致特征的信噪比未能得到有效改善。而拼接融合虽然能保留更多信息，但会成倍增加通道维度，显著增加模型计算量和参数量，且同样缺乏对背景噪声的抑制能力。相比之下，乘法操作确保只有在两个层次上都表现显著的特征才会被保留和强化，实现了语义引导下的空间精准定位。本文采用的乘性融合（ $Y_{\text{mul}} = F_{\text{high}} \odot F_{\text{low}}$ ）利用深层语义特征 F_{high} 作为权重项，实现了显式的“门控”效应：在前向传播阶段，对于非行人的背景区域，深层特征的响应值趋近于 0（ $F_{\text{high}} \rightarrow 0$ ），使得乘法操作能够过滤掉该区域的噪声（ $F_{\text{high}} \odot N \approx 0$ ）；而在行人区域，深层特征的高响应值则确保了有效信号 S 的保留与增强。在反向传播阶段，乘性操作进一步调节了梯度的流动，使得背景噪声区域的梯度权重受 F_{high} 调制而显著降低，从而避免网络在训练过程中去拟合无效的背景干扰。

同时 CMF 模块的自顶向下特征融合策略能够让具有丰富语义信息的高层特征有效指导低层特征的提取和优化。由于夜间场景中行人的边缘和纹理细节往往不明显，传统自底向上的特征提取难以在浅层形成有效的行人表征，容易被背景噪声干扰影响。而通过自顶向下的语义引导，网络可以在深层语义特征的指导下，更准确地地在浅层特征中定位和识别行人区域，实现高分辨率空间信息和丰富语义信息的有效结合，使网络能够同时应对夜间场景中的远小行人和近大行人检测问题。因此本文设计的 CMF 模块采用了这种自顶向下的特征融合策略，并结合乘性连接机制实现了多尺度特征的高效融合，为夜间行人检测提供了强大而精准的特征表达能力。

CMF 模块具体结构如图 2 所示，CMF 模块接收来自骨干网络 HRNet 的 4 个不同尺度特征图 $\{F_1, F_2, F_3, F_4\}$ ，其中 F_1 为浅层特征，具有最高的空间分辨率和最少的通道数， F_4 为深层特征，分辨率最低但包含最丰富的语义信息，通道数最多。CMF 模块采用自顶向下的特征融合机制，从最深层特征 F_4 开始，向最浅层 F_1 逐级上采样并进行乘性融合，形成一种金字塔式的信息流动路径。该模块使用深层特征作为语义权重，通过逐元素乘法有效抑制浅层噪声，获得更全面的特征表示，其核心操作可形式化表示为公式（1）：

$$F_{i-1}^{\text{CMF}} = F_{\text{up}}(F_i^{\text{CMF}}) \odot F_{\text{conv}}(F_{i-1}) \quad (1)$$

其中： F_i^{CMF} 表示第 i 层经过 CMF 模块处理后的特征图， F_{i-1} 表示第 $i-1$ 层的原始输入特征图， $F_{\text{up}}(\cdot)$

表示上采样和 1×1 卷积操作, $\odot$ 表示逐元素乘法操作, $F_{\text{conv}}(\cdot)$ 表示一系列卷积、批量标准化和激活函数等操作, 用于对原始特征进行预处理和通道维度对齐。

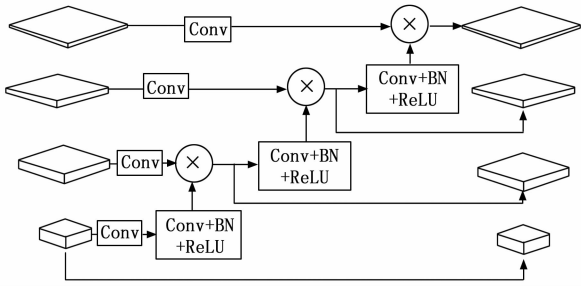


图 2 级联乘性融合模块结构图

融合过程从最深层开始, 首先保留最高层特征 $F_4^{\text{CMF}} = F_4$, 以此捕获夜间场景下的全局语义信息和上下文关系。随后对其进行上采样, 并与经过卷积处理的 F_3 进行逐元素乘法, 形成增强特征 F_3^{COM} 。通过这一过程逐级向浅层传递, 实现了深层语义对浅层细节的精准引导, 最终得到一组显著增强的夜间多尺度特征。这些特征在保留了各自尺度优势的同时, 获得了语义引导增强与背景噪声过滤, 为后续的行人检测任务提供了更具判别性和鲁棒性的特征表征。

1.2 双流通道注意力模块

经过 CMF 模块处理后, 多尺度特征在空间维度上实现了语义引导下的噪声抑制与行人区域增强。然而, 仅依靠空间维度的特征融合仍不足以应对夜间场景的复杂挑战。在夜间环境中, 不同特征通道对行人检测的贡献存在极大差异, 且这种通道重要性分布随场景光照条件动态变化, 固定的通道权重难以适应夜间场景的多样性。传统通道注意力机制如 CA、CBAM 等虽然在日间场景中表现出色, 但其单一路径设计难以同时捕获局部细节与全局语义, 无法有效地区分重要的通道特征与冗余的噪声信息, 导致夜间背景与行人目标的区分度不足; 且缺乏针对夜间场景的自适应调节机制, 无法有效应对光照不均带来的特征退化问题, 因此在面对夜间低对比度环境时表现出明显局限性。

针对这一问题, 本文提出了双流通道注意力模块。夜间场景往往同时存在大范围的非均匀照明和局部强噪声区域, 因此既需要从全局尺度感知整体光照分布与显著目标位置, 又需要在局部尺度精细抑制噪声、增强与行人相关的细节响应。基于这一考虑, DCA 模块分别构建两条互补的注意力路径: 一条在全局范围内建模通道响应的整体分布, 用于突出对行人语义更敏感的特征通道; 另一条在局部范围内刻画细粒度的通道依赖关系, 用于强化行人轮廓与纹理特征。通过可学习的自适应融合机制动态调整两条路径的权重, 使网络能够根据输入图像的光照条件和纹理特性自适应地增强与行人相

关的特征通道, 同时抑制噪声等无关通道的影响, 从而在复杂夜间环境中获得更加稳健的表示。

如图 3 所示, DCA 模块主要由双流通道注意力单元和自适应特征融合单元组成。在双流通道注意力单元中, 模块首先应用全局平均池化对输入特征进行空间维度压缩, 将 $H \times W \times C$ 的特征图压缩为 $1 \times 1 \times C$ 的通道描述符, 获取每个通道的全局响应, 从而将注意力建模从三维空间简化到一维通道域。这种全局聚合策略不仅能够显著降低后续注意力计算的复杂度, 还能够捕获每个通道在整个图像空间上的整体响应模式以评估不同通道的重要性。此外, 针对夜间挑战, 全局平均池化能够有效感知整幅图像的光照条件和噪声水平等全局上下文信息, 同时避免局部强光源、高亮区域等局部噪声的干扰, 确保通道注意力机制基于整体场景特性而非局部异常进行学习。随后, 模块对池化后的特征进行双路径并行处理, 分别采用一维卷积和全连接层处理, 生成两条互补的信息流。

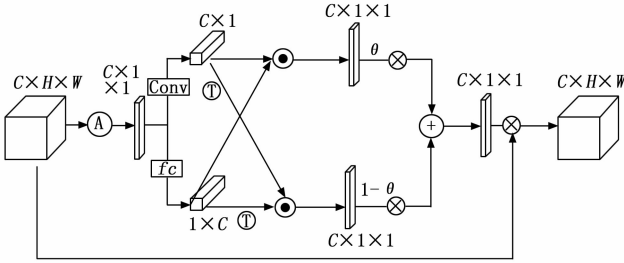


图 3 双流通道注意力模块结构图

在一维卷积路径中, 为了实现局部跨通道交互的自适应调节, 本文摒弃了固定卷积核大小的设计, 采用与通道维度相关的自适应核大小策略。具体而言, 我们将一维卷积的步长设置为 1, 填充设置为 $k/2$, 以确保输出特征保持原有的通道数。卷积核大小 k 通过公式动态计算, 具体操作如公式 (2) 所示:

$$k = \left\lceil \frac{\log_2(C) + 1}{2} \right\rceil_{\text{odd}} \tag{2}$$

其中: C 为输入特征的通道数, $\lceil \cdot \rceil_{\text{odd}}$ 表示取最近的奇数。不同大小的卷积核可以在不同特征层级上实现最优的局部信息捕获。

在全连接路径中, 为了避免夜间特征在降维压缩过程中丢失, 本文采用了全通道映射结构。具体而言, 该路径由单层线性映射层构成, 采用核大小为 1×1 的卷积层实现, 其可学习参数总量为 $C^e + C$ (其中 C^e 为权重参数, C 为偏置参数)。以此构建全局通道依赖关系, 学习长程通道间的全局关联, 捕获更大范围的上下文信息。

双流通道注意力单元通过矩阵乘法操作分别计算两个路径的特征交互, 并通过 sigmoid 激活函数生成互补的注意力权重, 具体操作如公式 (3) 所示:

$$out_1 = \sigma(x_1 \cdot x_2), out_2 = \sigma(x_2^T \cdot x_1^T) \quad (3)$$

其中： x_1 与 x_2 分别表示通过一维卷积和全连接层处理后的两条路径特征， out_1 捕获前向通道关系， out_2 捕获反向通道关系， σ 表示 sigmoid 激活函数。这种双向交互机制实现了局部细节与全局语义的互补增强。

然而，不同场景和不同尺度特征对局部和全局注意力的依赖程度并不一致。在夜间高噪声场景下，全局通道依赖可能更有助于抑制噪声干扰；而在细节丰富的区域，局部通道关系可能对行人的精细结构更重要。为了实现场景自适应的通道权重分配，DCA 模块引入了自适应特征融合单元，该单元通过引入可学习的混合因子 θ ，实现了对前向与反向通道注意力的动态加权融合，为不同场景下的特征增强提供了灵活的调控机制，具体操作可形式化表示为公式 (4)：

$$out = out_1 \cdot \sigma(\theta) + out_2 \cdot (1 - \sigma(\theta)) \quad (4)$$

其中： θ 通过训练自动学习最优路径组合。

融合后的注意力向量 out 通过广播操作与输入特征进行逐元素乘法，自适应特征融合单元通过参数 θ 自适应融合前向通道与反向通道的互补依赖关系。通过这种可学习的自适应机制，网络能够根据输入图像的具体特性动态调整通道重要性，同时关注局部显著特征（如行人轮廓、关键点等）和全局语义信息（如姿态、上下文等），从而学习出最优的特征融合策略，有效提升模型对夜间行人的特征感知能力。

1.3 行人检测头模块

检测头模块采用三支设计，将融合增强后的特征图 $F_{out} \in R^{H/4 \times W/4 \times 512}$ 转换为最终检测结果。该模块包括一个 3×3 卷积层和 3 个并行的 1×1 卷积层。首先， 3×3 卷积层将特征通道数从 512 降至 256，在优化特征表达的同时控制计算复杂度；随后，3 个并行的 1×1 卷积层分别生成中心点预测、尺度预测和偏移预测。在训练过程中，需要为检测头的每个输出分支构建相应的监督标签。基于数据集提供的边界框标注信息，可自动生成中心点、尺度和偏移的真值标签。

中心点预测分支采用基于高斯核的软标签策略处理定位不确定性，对于输入图像中的每个行人目标，中心热力图的真值标签定义为：

$$M_{ij} = \max_{k=1 \dots N} e^{-((i-x_k)^2/2\sigma_x^2 + (j-y_k)^2/2\sigma_y^2)} \quad (5)$$

其中： (i, j) 表示中心热力图中的空间坐标位置， (x_k, y_k) 表示第 k 个行人的中心坐标， N 为图像中的行人目标数量，最大值操作 $\max_{k=1 \dots N}$ 确保多目标重叠场景下标签的一致性。

尺度预测分支基于行人目标的高度信息构建监督标签。考虑到行人检测通常采用固定宽高比的边界框表示，本文采用 0.41 的统一宽高比，使得网络仅需预测行人高度即可完全确定边界框尺寸。尺度真值 S_{ij} 定

义为：

$$S_{ij} = \log(h_k) \quad (6)$$

其中： h_k 表示第 k 个行人目标的实际高度。引入对数变换有效压缩了不同尺度目标的数值范围，防止大尺度目标对小尺度目标学习过程的干扰。

偏移预测分支用于精细化边界框的中心位置，其真值标签 Q_{ij} 定义为：

$$O_{ij} = \left(\frac{x_n}{r} - \left\lfloor \frac{x_n}{r} \right\rfloor, \frac{y_n}{r} - \left\lfloor \frac{y_n}{r} \right\rfloor \right) \quad (7)$$

其中： (x_n, y_n) 表示原始图像中第 n 个行人的中心坐标， r 表示特征图相对于输入图像的下采样率。该分支通过像素级的位置修正，进一步提升了边界框的定位精度。

最后，通过整合 3 个预测分支的结果，结合非极大值抑制等后处理操作，自动生成最终的行人边界框。这种基于中心点的检测策略在夜间复杂场景下表现出色，能够有效处理行人目标的尺度变化和密集分布问题。

1.4 网络损失函数及推理

实验中，本方法采用多任务联合优化策略来约束检测网络的端到端训练。对于中心热力图预测，为缓解严重的正负样本不平衡问题，采用 Focal Loss^[21] 进行优化，通过动态权重因子降低易分负样本的权重，使网络将学习重心转移至困难样本；对于尺度和偏移预测，采用 Smooth L1 损失^[22] 进行回归优化，相比 L2 损失对异常值具有更好的鲁棒性。DMFNet 总体训练损失定义为：

$$L_{total} = \lambda_1 L_{center} + \lambda_2 L_{scale} + \lambda_3 L_{offset} \quad (8)$$

其中： L_{center} 、 L_{total} 和 L_{offset} 分别表示中心点预测损失、尺度预测损失和偏移预测损失，权重系数 λ_1 、 λ_2 和 λ_3 用于平衡各个任务对总损失的不同贡献。

在推理阶段，输入图像经过骨干网络和检测头的前向传播过程后，首先对中心热力图应用 0.01 的置信度阈值过滤低置信度位置，提取高置信度中心点及其对应的尺度预测。随后，结合中心位置和尺度信息构建初始检测边界框，并利用偏移图进行像素级位置优化。最后，将预测结果映射回原始图像尺寸，并采用阈值为 0.4 的非极大值抑制（NMS, non maximum suppression）消除重复检测，输出最终的行人检测结果。

2 实验结果与分析

2.1 数据集及评价指标

为全面验证所提出方法在不同光照条件下的性能，本文在两个具有挑战性的行人检测数据集上进行了实验：NightOwls^[11] 和 CityPersons^[23]。NightOwls 数据集专门为单模态夜间行人检测设计，包含 279k 个完全标注的帧，共 42 273 个行人。该数据集涵盖了不同季节、黎明和夜间时段以及多样化的天气条件，包含大量动态

物体、变化的场景布局和异构背景环境, 为评估夜间检测性能提供了理想的基础; CityPersons 是一个城市行人检测数据集, 包含 5 000 张高分辨率图像: 2 975 张用于训练, 500 张用于验证, 1 575 张用于测试。虽然主要是日间场景, 但也包含一些夜间和黄昏场景, 适合评估模型在不同光照条件下的泛化能力。通过在这两个互补数据集上的评估, 可以全面验证所提方法在不同场景和光照条件下的鲁棒性。

漏检率 (MR, Miss Rate) 定义为未检测到的行人数量占真值行人总数的比例, 可用公式 (9) 表示:

$$MR = \frac{FN}{(TP + FN)} = 1 - Recal \tag{9}$$

其中: FN 表示假阴性数量, TP 表示真阳性数量。本文采用对数平均漏检率 $(MR)^{-2}$ 作为评价指标, 该指标计算每图像假阳性 (FPPI) 在 $[10^{-2}, 10^0]$ 范围内漏检率的对数平均值。 MR^{-2} 值越低表示检测性能越好。

2.2 训练细节

本文基于 PyTorch 框架实现 DMFNet, 采用在 ImageNet^[24] 上预训练的 HRNet-W32 作为骨干网络, 在配备四块 NVIDIA PCIe A100 GPU 的高性能服务器上进行训练。在实验过程中, 采用了多种数据增强策略, 包括亮度调整、水平翻转和随机缩放。优化器采用 Adam^[25], 初始学习率设为 2×10^{-4} , 采用阶梯衰减策略每 100 次迭代将学习率减半。在第 50 个 epoch 时, 学习率进一步降至初始值的 10%。为确保训练稳定性, 在前 3 个 epoch 实施线性预热策略。对于 NightOwls 和 CityPersons 数据集, 训练图像的输入尺寸设置为 $640 \times 1\,024$, 测试图像的输入尺寸分别为 $640 \times 1\,024$ 和 $1\,024 \times 2\,048$ 。模型共训练 150 个 epoch, 每个 epoch 包含 2 000 次迭代, 每个 GPU 的批量大小为 2, 总批量大小为 8。

2.3 消融实验

为评估所提模块对夜间行人检测性能的贡献, 本文在 NightOwls 数据集上进行了消融实验。实验中, 基线模型仅采用 HRNet 骨干网络, 随后逐步引入级联乘性融合模块和双流通道注意力模块, 与其他注意力机制对比, 并验证各模块的有效性。表 1 展示了不同配置下的检测性能对比。从表中可以看出, 基线模型在 Reasonable 子集上的漏检率为 11.82%; 单独引入 CMF 模块后, Reasonable 子集漏检率降至 7.92%; 单独引入 DCA 模块后, Reasonable 子集漏检率为 8.27%; 当同时引入 CMF 和 DCA 两个模块时, 漏检率进一步降至 6.85%, 验证了两个模块的独立有效性。此外, 为对比不同注意力机制的性能, 我们在 CMF 基础上分别引入 CBAM 和 CA 模块。实验结果表明, CMF+DCA 配置达到最优的 6.85%。

表 1 不同核心模块及注意力机制消融实验

核心模块					MR^{-2}
HRNet	CMF	DCA	CBAM	CA	Reasonable
✓					11.82
✓	✓				7.92
✓		✓			8.27
✓	✓		✓		8.45
✓	✓			✓	7.73
✓	✓	✓			6.85

为验证乘性融合相比传统融合方式的优势, 本文对比了加法融合、拼接融合和乘性融合 3 种特征融合策略。所有实验均保持相同的网络结构和训练配置, 仅改变 CMF 模块中的融合操作方式。表 2 展示了不同融合策略的性能对比。实验结果表明, 本文提出的乘性融合策略在 Reasonable 和 All 子集的漏检率分别达到了 6.85% 和 15.61%, 取得了最优性能。

表 2 不同融合策略的消融实验

融合策略	Reasonable	All
加法融合	7.76	16.89
拼接融合	7.42	16.36
乘性融合	6.85	15.61

在 DCA 模块中, 可学习参 θ 数决定了局部路径与全局路径在特征融合时的初始权重分配。根据公式 (4), 融合权重由 Sigmoid 函数 $\sigma\theta$ 生成。为了探究 θ 的初始值对模型收敛及最终性能的影响, 本文在 NightOwls 数据集的 Reasonable 和 All 子集上进行了参数敏感性分析实验。我们设置了不同的 θ 初始值进行对比, 包括 0 (等权重初始化)、-0.8、-0.5、-1.0 以及正值 0.5。实验结果如表 3 所示, 在 θ 为 -0.8 时, 模型取得最好的性能。

表 3 DCA 模块混合因子的敏感性分析

$\theta_{\text{初始值}}$	Reasonable	All
0	7.22	16.35
-0.5	6.94	15.82
-0.8	6.85	15.61
-1.0	7.08	16.18
0.5	7.35	16.75

2.4 客观对比试验

为全面评估所提出的 DMFNet 在夜间行人检测任务上的有效性, 本文在 NightOwls 和 CityPersons 两个具有代表性的数据集上与现有先进方法进行了对比实验。所有方法均采用对数平均漏检率 (MR^{-2}) 作为评价指标, 数值越低表示性能越好。

表 4 展示了 DMFNet 在 NightOwls 数据集上的性

能对比。从表中可以看出, DMFNet 在所有测试子集上均实现了显著性能提升。在 Reasonable 子集上, DMFNet 取得了 6.85% 的漏检率; 在小尺度行人检测方面, 漏检率降至 14.72%。实验结果表明, 本文方法在处理夜间行人检测的多尺度等关键挑战方面具有显著优势。

表 4 不同算法在 NightOwls 数据集上的性能对比

方法	MR ⁻²			
	Reaso-nable	Small	Occ	All
ALFNet ^[26]	21.01	33.09	61.06	32.18
TFAN ^[12]	16.50	—	—	—
EGCL ^[27]	15.93	26.25	50.36	28.73
FATNet ^[15]	14.02	22.43	43.32	23.28
FBCNet ^[16]	13.64	22.70	46.91	25.61
Ours	6.85	14.72	42.21	15.61

为验证 DMFNet 的泛化能力, 本文在 CityPersons 上进行了对比实验, 结果如表 5 所示。本文方法在不同划分的子集下均取得了最佳结果, 在 Reasonable 子集上, DMFNet 取得了 9.6% 的漏检率。实验结果表明, DMFNet 在日间场景中依然保持竞争力, 验证了所提方法的泛化能力。

表 5 不同算法在 CityPersons 数据集上的性能对比

方法	MR ⁻²			
	Reaso-nable	Partial	Bare	Heavy
DAFC+ ^[28]	11.8	—	—	44.8
DA ^[29]	11.5	—	—	51.1
EGCL ^[27]	10.9	11.6	7.4	46.4
CSP ^[30]	10.4	9.3	6.8	48.1
OSS-OCL ^[31]	10.2	—	—45.9	
SMPD ^[32]	9.9	9.0	6.5	45.6
VA-CSP ^[33]	9.6	9.1	6.6	48.1
Ours	9.6	6.94	7.8	44.3

2.5 可视化分析

为进一步验证 DMFNet 在行人特征表达方面的改进效果, 本文对基线模型与 DMFNet 在夜间场景下的特征热图进行了可视化对比分析。

如图 4 所示, 左图为基线方法输出的行人响应热图, 右图为 DMFNet 的响应结果。可以观察到, 基线方法对行人的响应较弱, 热区分布较为分散, 并且将摩托车误检为行人; 本文方法在夜间场景下, 对行人的响应更加集中和稳定, 从而消除了误检。



图 4 DMFNet 与基线模型的特征热图对比

在此基础上, 为直观展示 DMFNet 在检测结果层面的优势, 本文进一步对模型在 NightOwls 和 CityPersons 验证集上进行了可视化分析。图 5 展示了基线方法与 DMFNet 在不同夜间场景下的检测效果对比, 从图 5 可以观察到, 基线方法在复杂的夜间场景中存在以下典型问题。在低光照条件下 (图 5 (a)), 由于行人目标与背景对比度极低, 基线方法对远处的小尺度行人出现明显漏检, 而本文方法成功检测出所有尺度的行人目标。在夜间小尺度行人的部分遮挡场景中 (图 5 (b)), 基线方法对部分遮挡的小尺度行人检测失败, 而本文方法对部分遮挡行人实现了准确检测。此外, 在 Citypersons 验证集中, 在图 5 (c) 的复杂背景场景下, 基线方法容易将路灯、树木等非行人目标误识别为行人, 产生较多误检, 而本文方法抑制了背景噪声干扰, 未出现误检情况。综合结果表明, 本文方法在夜间低光照以及多尺度行人等多种挑战场景下均展现出稳定可靠的检测性能, 验证了所提方法的有效性和鲁棒性。

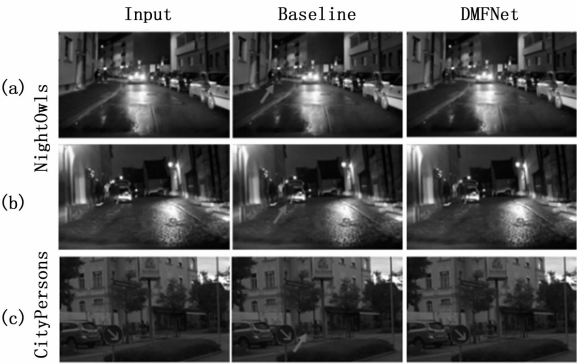


图 5 不同数据集中行人检测的可视化结果对比

3 结束语

本文针对夜间低光照环境下行人检测面临的多尺度特征捕获困难和背景噪声干扰等问题, 提出了一种基于级联乘性融合与双流通道注意力的夜间行人检测网络 DMFNet。该方法采用 HRNet 作为骨干网络提取多尺度特征, 设计了级联乘性融合模块实现跨尺度特征的高效整合并抑制背景噪声干扰, 提出了双流融合通道注意力模块增强行人多尺度特征的判别能力。在 NightOwls 和 CityPersons 等公开基准数据集上的实验结果表明, DMFNet 在多项评价指标上均达到了先进水平, 消融实验进一步验证了不同核心模块的有效性。本文方法为夜间行人检测技术提供了新的解决方案, 在自动驾驶、智能交通监控、夜间安防巡检等实际场景中具有广阔的应用前景。然而, 本文方法仍存在一些需要进一步完善的技术问题。当前, 本文方法的计算复杂度相对较高, 在嵌入式设备上的实时性能需要进一步优化。此外, 本文方法主要针对交通场景下的夜间行人检测, 未来工作将

进一步优化网络模型,探索该方法在更广泛应用场景中的泛化能力和适应性。

参考文献:

- [1] 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(7):2094-2111.
- [2] 赵才荣,齐 鼎,窦曙光,等.智能视频监控关键技术:行人再识别研究综述[J].中国科学:信息科学,2021,51(12):1979-2015.
- [3] 包晓安,朱晓芳,张 娜,等.基于PHOG特征的行人检测算法研究[J].计算机测量与控制,2018,26(8):158-162.
- [4] 王晨宇,郑利民,李 延.基于YOLOv8的遮挡行人检测算法改进研究[J].现代车用动力,2025(2):1-5.
- [5] KHAN A H, MUNIR M, VAN ELST L, et al. F2dnet: Fast focal detection network for pedestrian detection [C] //2022 26th International Conference on Pattern Recognition, IEEE, 2022: 4658-4664.
- [6] LIU M, JIANG J, ZHU C, et al. Vlpd: Context-aware pedestrian detection via vision-language semantic self-supervision [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 6662-6671.
- [7] 李宝兵,符长友.基于改进YOLOv8的低光照行人检测算法[J].西华大学学报(自然科学版),2025,44(3):29-36.
- [8] ZHOU K, CHEN L, CAO X. Improving multispectral pedestrian detection by addressing modality imbalance problems [C] //European Conference on Computer Vision, Cham: Springer International Publishing, 2020: 787-803.
- [9] ZHANG H, FROMONT E, LEFÈVRE S, et al. Guided attentive feature fusion for multispectral pedestrian detection [C] //Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021: 72-80.
- [10] XIE J, ANWER R M, CHOLAKKAL H, et al. Learning a dynamic cross-modal network for multispectral pedestrian detection [C] //Proceedings of the 30th ACM International Conference on Multimedia, 2022: 4043-4052.
- [11] NEUMANN L, KARG M, ZHANG S, et al. Night-owls: A pedestrians at night dataset [C] //Asian Conference on Computer Vision, Cham: Springer International Publishing, 2018: 691-705.
- [12] WU J, ZHOU C, YANG M, et al. Temporal-context enhanced detection of heavily occluded pedestrians [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 13430-13439.
- [13] KHAN A H, NAWAZ M S, DENGEL A. Localized semantic feature mixers for efficient pedestrian detection in autonomous driving [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 5476-5485.
- [14] CHENG B, XIAO B, WANG J, et al. Higherhrnet: Scale-aware representation learning for bottom-up human pose estimation [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 5386-5395.
- [15] LI G, ZHANG S, YANG J. Nighttime pedestrian detection based on feature attention and transformation [C] //2020 25th International Conference on Pattern Recognition, IEEE, 2021: 9180-9187.
- [16] YAO H, ZHANG Y, JIAN H, et al. Nighttime pedestrian detection based on Fore-Background contrast learning [J]. Knowledge-Based Systems, 2023, 275: 110719.
- [17] 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.
- [18] 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.
- [19] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module [C] //Proceedings of the European Conference on Computer Vision, 2018: 3-19.
- [20] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13713-13722.
- [21] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C] //Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980-2988.
- [22] REN S, HE K, GIRSHICK R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks [J]. Advances in Neural Information Processing Systems, 2015, 28.
- [23] ZHANG S, BENENSON R, SCHIELE B. Citypersons: A diverse dataset for pedestrian detection [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 3213-3221.
- [24] DENG J, DONG W, SOCHER R, et al. Imagenet: A large-scale hierarchical image database [C] //2009 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2009: 248-255.

(下转第309页)