

文章编号: 1007-2780(2026)05-0717-17

FW-YOLO: 一种基于频率引导和小波增强的 密集行人检测算法

黄子骏, 凌 萍*

(江苏师范大学 人工智能与计算机学院, 徐州 江苏 221116)

摘要: 在密集目标检测场景中, 传统方法存在高频边缘信息丢失、上采样细节模糊及尺度适应性不足等问题。为此, 本文提出一种基于频率引导的目标检测算法 FW-YOLO。该算法设计频率引导全维动态卷积模块(FGODC), 利用 Haar 小波变换分解特征, 并结合高频响应引导卷积核选择, 以增强下采样阶段的边缘信息保留能力; 提出频率残差引导上采样模块(FRUpsample), 将高频特征注入上采样路径, 实现边缘细节恢复; 构建频率引导共享检测头(FGSHHead), 结合跨尺度权重共享与 FiLM 机制, 提升不同尺度目标的检测能力。实验结果表明, 在 CrowdHuman 数据集上, 该方法的精确率、召回率、mAP50 和 mAP50:95 分别提升 1.1%、3.8%、1.8% 和 2.0%, 验证了频率引导策略在密集目标检测中的有效性, 为密集目标检测任务提供了一种新的解决思路。

关键词: 目标检测; 频率引导; 小波变换; 动态卷积; 上采样; 密集行人检测

中图分类号: TP391.4 文献标识码: A doi: 10.37188/CJLCD.2026-0044 CSTR: 32172.14.CJLCD.2026-0044

FW-YOLO: a dense pedestrian detection algorithm based on frequency-guided and wavelet enhancement

HUANG Zijun, LING Ping*

(College of Artificial Intelligence and Computer Science, Jiangsu Normal University,
Xuzhou 221116, China)

Abstract: In dense object detection scenarios, conventional detection methods often suffer from the loss of high-frequency edge information, blurred details during upsampling, and insufficient scale adaptability. To address these issues, this paper proposes a frequency-guided object detection algorithm, termed FW-YOLO. Specifically, a Frequency-Guided Omni-Dimensional Dynamic Convolution (FGODC) module is designed to decompose features by means of the Haar wavelet transform, while high-frequency responses are further exploited to guide convolution kernel selection, thereby enhancing the preservation of edge information during downsampling. In addition, a Frequency-Residual Guided Upsampling (FRUpsample) module is introduced to inject high-frequency features into the upsampling path, enabling the recovery of fine edge details. Furthermore, a Frequency-Guided Shared Head (FGSHHead) is constructed by combining cross-scale weight sharing with the FiLM mechanism, so as to improve the detection capability for objects at

收稿日期: 2026-03-18; 修订日期: 2026-04-10.

基金项目: 国家自然科学基金面上项目 (No. 61872168)

Supported by General Program of National Natural Science Foundation of China (No. 61872168)

*通信联系人, E-mail: 6020000012@jsnu.edu.cn

different scales. Experimental results on the CrowdHuman dataset show that the proposed method improves Precision, Recall, mAP50, and mAP50:95 by 1.1%, 3.8%, 1.8%, and 2.0%, respectively. These results verify the effectiveness of the frequency-guided strategy in dense object detection and provide a new perspective for dense object detection tasks.

Key words: object detection; frequency guidance; wavelet transform; dynamic convolution; upsampling; dense pedestrian detection

1 引 言

由于存在严重的相互遮挡、尺度变化剧烈以及复杂光照条件,密集人群场景下的行人检测面临极高的漏检与误检风险,严重制约了其在智慧城市流量监控、自动驾驶环境感知及公共安全预警等领域的实际应用。

从应用需求看,密集行人检测不仅是通用目标检测中的重要研究内容,也是成像与显示系统实现智能感知与信息呈现的关键技术基础。在车载显示与辅助驾驶场景中,前端检测结果直接影响行人目标的高亮标注、风险提示与驾驶决策信息呈现;在智慧安防监控与公共安全可视化场景中,检测结果又是多目标跟踪、异常事件预警和显示终端信息调度的重要前提。由于上述应用普遍伴随遮挡严重、光照变化明显、成像质量波动及目标尺度差异大等问题,如何从图像成像特性出发提升密集行人的边缘细节保留能力与检测稳定性,已成为图像处理、计算机视觉与智能显示应用交叉领域值得深入研究的重要课题。

近年来,YOLO系列算法凭借其在精度与速度之间的卓越平衡,已成为工业界首选的检测框架^[1]。从YOLOv8^[2]通过C2f模块与解耦头设计提升计算效率,到YOLO11^[3]引入C3k2模块与C2PSA注意力机制进一步增强特征提取的鲁棒性,再到最新的YOLO26^[4]引入无NMS端到端架构,提升边缘设备的推理效率,该系列在通用目标检测任务中展现了持续的性能突破。

为解决密集拥挤场景下行人特征模糊与严重遮挡的难题,研究者在YOLO基线模型之上进行了大量针对性改进。针对密集人群中特征提取能力不足的问题,DED-YOLOv8^[5]算法在骨干网络中引入DCNv2可变形卷积,并结合EMA^[6]高效多尺度注意力机制与DyHead动态检测头,显著增强了模型对多尺度遮挡行人的感知能力。为

提升复杂背景下的特征融合效果,YOLO-RAD^[7]模型创新性地感受野注意力融入卷积模块,并设计了自适应空间特征融合策略,在CrowdHuman^[8]数据集上有效降低了漏检率。FA-YOLO^[9]算法提出一种特征增强模块以整合全局与局部信息,并结合自适应稀疏自注意力机制,进一步强化了网络在低光照与高密度场景下对关键特征的捕捉能力。针对红外与夜间场景下的微小目标检测,MDCFvit-YOLO^[10]通过引入视觉Transformer模块与Focaler-ShapeIoU损失函数,优化了极端环境下的边界框回归精度。针对密集行人检测中因拥挤遮挡而导致的漏检问题,曹洁等^[11]在YOLOv7框架基础上提出基于优化权重的密集行人检测算法,通过优化权重分配与特征融合策略,提升了拥挤场景下的行人检出能力。针对遮挡场景下仅依赖整体外观信息易造成漏检的问题,马晞茗等^[12]提出基于头部与整体信息联合的行人检测算法,通过融合头部关键部位与整体表征,提高了遮挡行人的检测鲁棒性。

然而,现有方法普遍忽视了一个根本的结构瓶颈:卷积神经网络中广泛使用的步长卷积(Strided Convolution)或池化操作,在执行下采样时会依据采样定理而丢弃高频分量(High-Frequency Components)。在密集行人检测中,这些高频分量对应着区分相邻行人的边缘纹理与轮廓细节,一旦在骨干网络(Backbone)阶段丢失,后续颈部网络(Neck)的上采样过程仅能通过插值恢复低频平滑特征,导致深层特征图中相邻行人发生“特征混叠”,边界模糊不清。此外,现有的检测头通常为每个检测尺度配置独立的卷积参数,不仅导致了明显的参数冗余,更缺乏跨尺度的自适应调节机制,使得单一固定的卷积核难以灵活应对密集人群中目标尺度剧烈变化与极端尺寸的检测需求。

近期研究发现,在目标密集且遮挡严重的场

景中,仅依赖空间域特征提取容易导致边缘纹理弱化与判别信息混叠。针对这一问题,部分工作将二维离散小波变换^[13]引入检测框架,以分离低频结构与高频细节并尽可能保留纹理信息。同时,动态卷积^[14]可根据局部内容自适应调整聚合权重^[15];空间注意力^[16]与基于FiLM的线性调制^[17]则有助于跨层特征重标定,降低多尺度分支的参数冗余。基于这些研究,我们认为密集行人检测中的主要问题并不只在于空间信息不足,而在于关键信息在传递过程中被逐步稀释,因此将重点放在“先保留细节、再按层次选择性融合”的信息流重构上。

针对上述问题,本文提出一种基于频率感知与闭环信息流的密集行人检测算法——FW-YOLO (Frequency-Guided and Wavelet Enhancement YOLO)。不同于以往仅在空间域修补的方法,本研究创新性地构建了覆盖“下采样-上采样-检测头”全链路的频率域闭环设计:

(1)提出频率引导全维动态卷积(FGODC),在下采样阶段引入小波分解获取高频细节,并结

合动态卷积进行自适应聚合,从而增强密集目标边缘与纹理表征能力。

(2)设计频率残差引导下采样(FRUpsample),以高频残差注入替代常规插值恢复,并通过空间自适应门控调节融合比例,在保留底层细节的同时提升特征金字塔的轮廓重建能力。

(3)构建频率引导共享检测头(FGSHHead);将层级权重共享与FiLM特征调制结合,用于统一跨尺度特征映射与参数生成,缓解多尺度分支冗余并提升密集场景下的尺度适应性与检测稳定性。

2 FW-YOLO 模型设计

2.1 整体架构设计

本文在YOLOv11的基础上构建FW-YOLO,通过引入频率域闭环设计,在降低检测头参数冗余的同时,增强了密集遮挡场景下对目标边缘与纹理的感知能力。如图1所示,首先在Backbone中引入FGODC模块以替换传统下采样层,利用Haar小波分解与频率引导动态路由机制,在特征

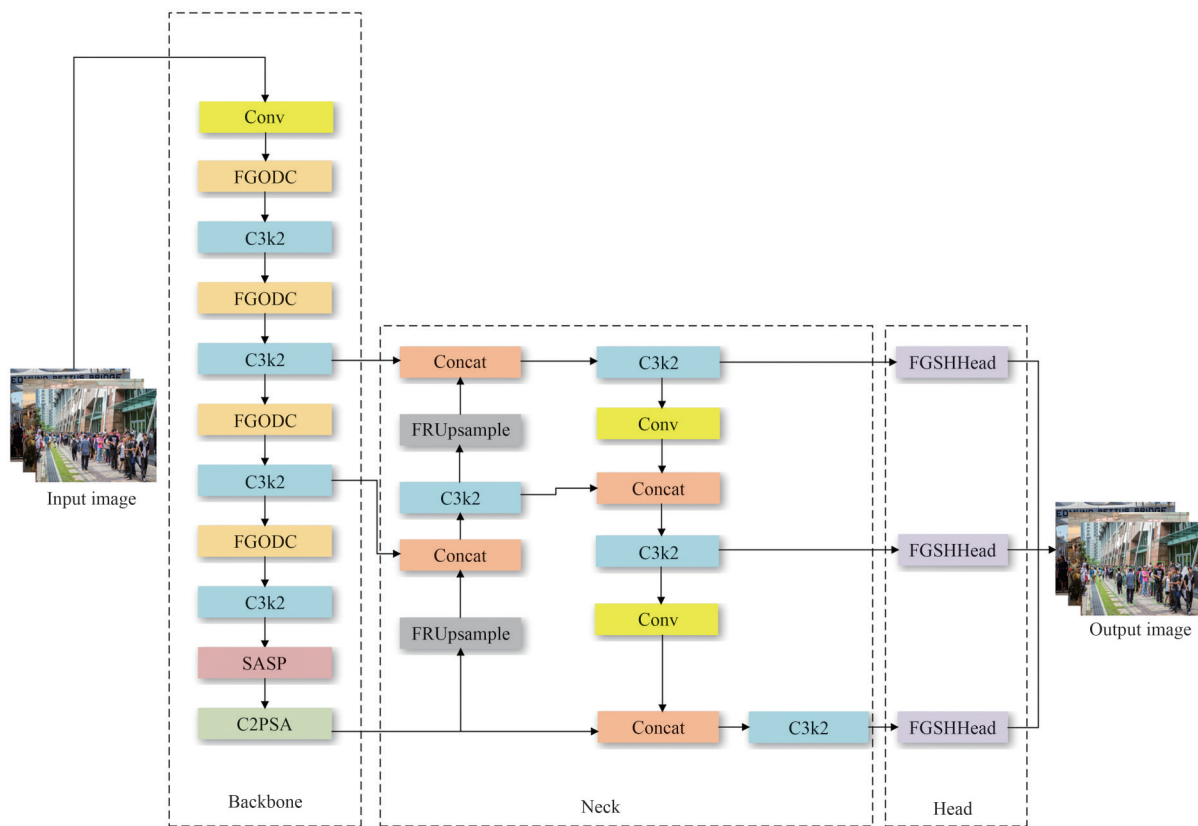


图1 FW-YOLO 模型架构图

Fig. 1 FW-YOLO model architecture

降维过程中显式保留高频边缘细节,解决标准卷积导致的细微特征丢失问题;同时在 Neck 中设计 FRUpsample 模块,通过空间自适应注意力门控机制将缓存的高频特征注入上采样过程,有效补偿插值造成的边缘模糊并强化特征金字塔的轮廓重构能力;最后,构建 FGSHHead 检测头用于取代标准检测头,利用跨尺度权重共享与 FiLM 频率调制技术,实现尺度自适应特征变换,为检测器在复杂密集环境中提供更具判别力的特征表示。

2.2 频率引导全维动态卷积

在深度卷积神经网络中,下采样操作是构建

多尺度特征表示的关键步骤。传统的步长卷积通过设置步长 (Stride) 参数来实现空间分辨率的降低,然而该操作在下采样过程中不可避免地造成高频信息的丢失。根据奈奎斯特采样定理^[18],当采样频率低于信号最高频率的 2 倍时,将产生频谱混叠现象,导致图像边缘、纹理等高频细节信息的不可逆损失。

针对上述问题,本文提出频率引导全维动态卷积 FGODC (Frequency-Guided Omni-Dimensional Dynamic Convolution) 模块。如图 2 所示,该模块通过引入 Haar 小波变换对输入特征进行频率域

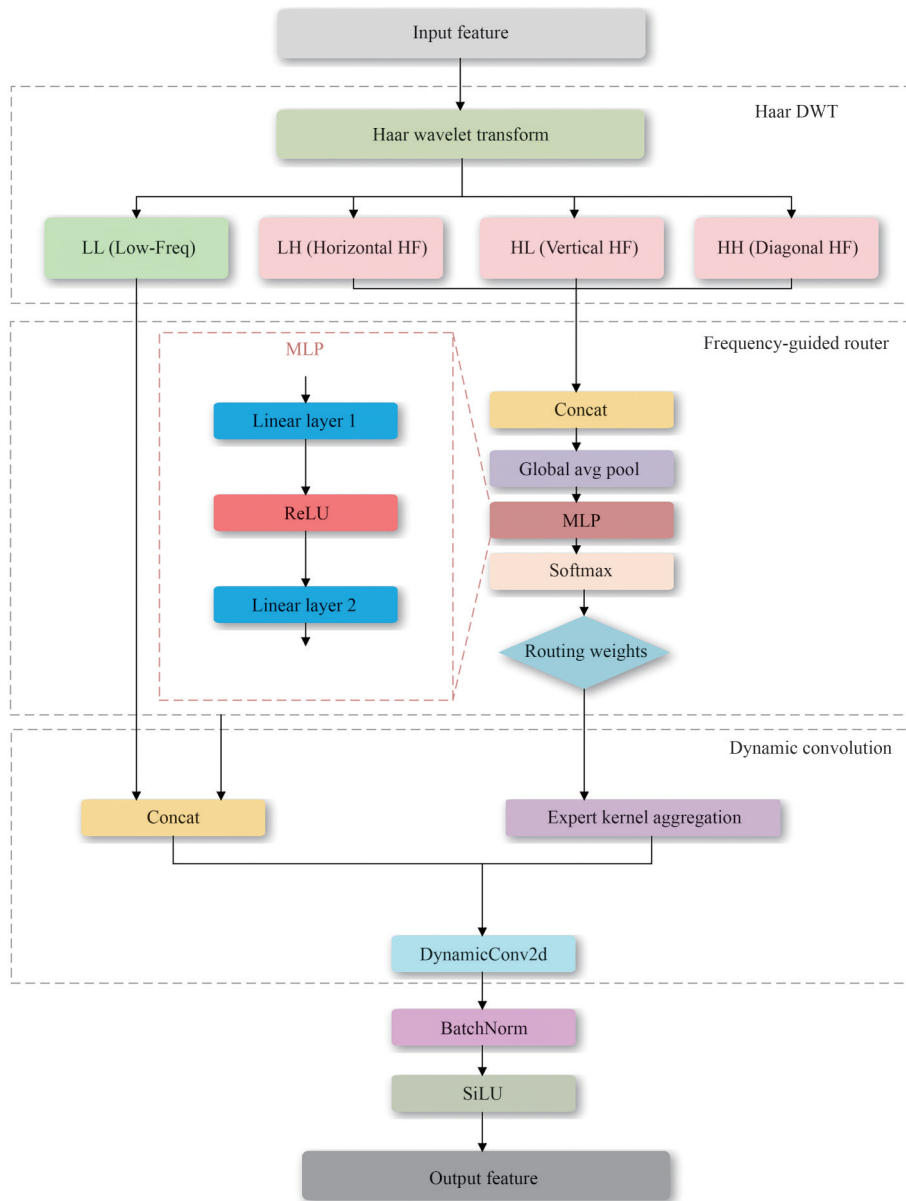


图 2 FGODC 结构图

Fig. 2 FGODC structure diagram

分解,显式分离低频近似分量与高频细节分量,并利用高频响应强度指导动态卷积核的自适应选择,从而在保留边缘信息的同时实现高效的特征下采样。

Haar小波变换作为最简单的正交小波变换,具有计算高效、实现简洁的优点,被广泛应用于图像处理与信号分析领域。对于输入特征图 $X \in \mathbb{R}^{C \times H \times W}$, Haar小波变换将其分解为4个频率子带:低频近似分量(LL)与3个高频细节分量(LH、HL、HH)。具体而言,对于特征图中任意 2×2 像素块 $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$, Haar小波变换定义如公式(1)所示:

$$\begin{cases} \text{LL} = \frac{1}{4}(a + b + c + d) \\ \text{LH} = \frac{1}{4}(a - b + c - d) \\ \text{HL} = \frac{1}{4}(a + b - c - d) \\ \text{HH} = \frac{1}{4}(a - b - c + d) \end{cases}, \quad (1)$$

其中:LL表示低频近似分量,反映图像的整体亮度与平滑区域信息;LH表示水平方向高频分量,用于检测垂直边缘;HL表示垂直方向高频分量,用于检测水平边缘;HH表示对角方向高频分量,用于检测对角边缘与纹理细节。

经过Haar小波变换后,各子带的空间分辨率降为原始特征图的1/2,该变换在实现下采样的同时,将高频边缘信息显式保留于独立的子带中,为后续的频率引导机制提供了基础。

同时为充分利用高频分量所蕴含的边缘与纹理信息,该模块中设计了频率引导路由器(Frequency-Guided Router),根据局部区域的高频响应强度自适应地选择最优的卷积核组合。

设3个高频子带经拼接后得到 $H_f = [\text{LH}; \text{HL}; \text{HH}]^{3C \times \frac{H}{2} \times \frac{W}{2}}$, 频率引导路由器首先通过全局平均池化^[19]提取全局频率统计信息 z :

$$z = \text{GAP}(H_f) = \frac{1}{\frac{H}{2} \times \frac{W}{2}} \sum_{i,j} H_f(:, i, j) \in \mathbb{R}^{3C}, \quad (2)$$

其中: H_f 表示三者通道维度拼接得到的高频特征, C 为通道数, H 、 W 为空间尺寸。随后,通过多层感知机(MLP)^[20]将频率统计向量映射为 K 个

专家卷积核的注意力权重,如公式(3)所示:

$$\alpha = \text{Softmax}(W_2 \cdot \sigma(W_1 \cdot z)) \in \mathbb{R}^K, \quad (3)$$

其中: α 为注意力权重, W_1 与 W_2 为可学习参数, $\sigma(\cdot)$ 为ReLU激活函数, K 为专家卷积核数量。

当获取专家卷积核的注意力权重 α 后,动态卷积模块对 K 个专家卷积核进行加权聚合,获得聚合后的卷积 W_{agg} ,如公式(4)所示:

$$W_{\text{agg}} = \sum_{k=1}^K \alpha_k \cdot W_k, \quad (4)$$

其中, W_k 表示第 K 个专家卷积核的参数。聚合后的卷积核 W_{agg} 可以根据输入特征的特征特性进行动态调整,使得边缘复杂区域能够自动选择更适合捕捉细节的卷积核组合。

最终,FGODC的输出特征计算如下:

$$Y = \text{SiLU}(\text{BN}(W_{\text{agg}} * X_{\text{cat}})), \quad (5)$$

其中: X_{cat} 为4个频率子带的拼接,*表示卷积操作,BN(\cdot)表示批归一化,SiLU(\cdot)表示SiLU激活函数。

本模块相较于传统的步长卷积,通过Haar小波变换将高频边缘信息分离至独立子带,避免了传统下采样中高频信息的直接丢弃;使用频率引导路由机制根据局部区域的边缘复杂度动态调整卷积核权重,使网络能够针对不同内容区域采用差异化的特征提取策略;采用分组卷积技巧实现批量动态卷积的高效计算,避免了逐样本卷积带来的计算开销。

2.3 频率残差引导上采样

在目标检测网络的特征金字塔结构(如FPN、PANet)^[21-22]中,上采样操作是实现多尺度特征融合的关键步骤。然而,传统的上采样方法存在固有缺陷:高频信息完全丢失,最近邻插值只是简单地复制像素值,这是一个纯低通滤波操作,无法恢复下采样时丢失的边缘、纹理等高频细节;由于直接复制像素,放大后的特征图会出现明显的“马赛克”效果,边缘呈现阶梯状锯齿;相邻像素之间没有平滑过渡,导致小目标边界模糊不清,相邻目标容易粘连,细粒度特征难以区分等现象^[23]。

基于上述分析,本文提出频率残差引导上采样FRUpsample(Frequency-Residual Upsample)模块,核心思想是:在上采样过程中显式补偿被丢失的高频信息,实现“先分解、后重建”的频率感知上采样。

FRUpsample 模块由 3 个核心组件构成:

(1) Haar 小波高频提取器: 从下采样路径提取高频残差;

(2) 空间自适应注意力门控: 根据局部高频响应强度动态调整融合权重;

(3) 特征融合与压缩网络: 将上采样特征与高频残差有效融合。

Haar 小波是最简单的正交小波基, 具有紧支撑特性, 非常适合实时图像处理。对于二维信号, Haar 小波变换将输入分解为 4 个子带, 其中 3 个高频子带通过可学习权重进行自适应融合:

$$H = \sum_{i \in \{LH, HL, HH\}} \varphi_i \cdot H_i, \quad (6)$$

其中可学习权重 φ_i 通过 Softmax 归一化获得:

$$\varphi_i = \frac{\exp(w_i)}{\sum_j \exp(w_j)}, \quad (7)$$

其中: $w_i \in \mathbb{R}$ 为可学习参数, 这种设计可以使网络自动学习不同方向高频信息的重要性。

同时, 对于不同空间位置对高频信息所需程度不同, 本文设计了空间自适应注意力机制, 根据局部高频响应强度动态调整融合权重。对于给定融合后的高频特征 $H \in \mathbb{R}^{C \times H \times W}$, 空间注意力图计算如式(8)所示:

$$A =$$

$$\sigma(\text{Conv}_{1 \times 1}(\text{Conv}_{3 \times 3}(\text{ReLU}(\text{Conv}_{1 \times 1}(H))))), \quad (8)$$

其中: $\sigma(\cdot)$ 表示 Sigmoid 函数, Conv 表示卷积块。如公式(9)所示, 门控后的高频特征通过逐元素乘法获得:

$$H_{\text{gated}} = A \odot H. \quad (9)$$

如公式(10)所示, 深层特征 X 先通过双线性插值上采样, 再与门控后的高频特征拼接, 两层卷积网络执行通道压缩和特征精炼:

$$\begin{cases} X_{\text{cat}} = \text{Concat}(\text{Upsample}_{2 \times}(X), \\ \text{Upsample}_{2 \times}(H_{\text{gated}})) \\ F_{\text{out}} = \text{Conv}_{3 \times 3}(\text{SiLU}(\text{Conv}_{1 \times 1}(X_{\text{cat}}))) \end{cases} \quad (10)$$

首先, 将特征图 X 与门控特征 H_{gated} 分别进行 2 倍上采样, 并在通道维度进行拼接, 得到融合特征 X_{cat} ; 随后, 通过卷积与激活函数进一步提取目标特征 F_{out} 。

所提出的 FRUpsample 模块架构如图 3 所示。浅层特征首先通过 Haar 小波变换分解为 4 个子带。3 个高频子带(LH、HL、HH)通过可学习权重融合, 并由空间自适应注意力门控调制。门控后的高频特征与上采样的深层特征拼接, 实现特征的最终融合。

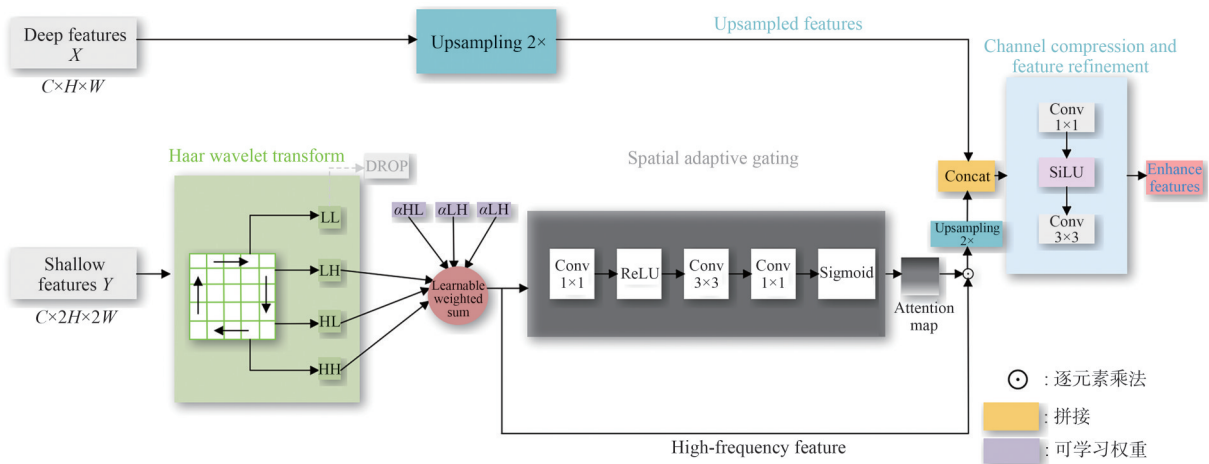


图 3 FRUpsample 结构图

Fig. 3 FRUpsample structure diagram

2.4 频率引导共享检测头

检测头是目标检测网络的最终输出模块, 负责将多尺度特征图转换为目标的类别预测与边界框回归结果。现代目标检测器通常采用解耦

检测头(Decoupled Head)架构^[24], 将分类任务与回归任务分离至独立的分支进行处理。然而, 现有检测头设计存在以下问题: 现阶段的分类任务侧重于语义判别, 需要关注目标的整体特征; 而

回归任务侧重于边界定位,需要关注目标的边缘细节。传统共享检测头难以同时满足两种任务的差异化特征需求。同时,不同尺度的特征图在特征分布上存在差异,直接共享检测头参数可能导致尺度适应性不足^[25]。

针对上述问题,本文提出频率引导共享检测头 FGSHHead(Frequency-Guided Shared Head)。如图 4 所示,该模块通过引入频率感知的任务路由机制,在共享参数的框架下实现分类与回归任务的差异化特征处理,同时借鉴自适应归一化思

想^[26]设计了一种尺度自适应归一化策略(Scale-Adaptive Normalization, SAN)。该策略针对不同尺度学习独立的归一化参数,同时共享卷积核参数,以提升共享检测头对多尺度特征的适应能力。

该模块的核心创新在于利用输入特征的特征特性指导分类与回归任务的特征处理策略。其设计基于以下观察:高频分量主要包含边缘与纹理信息,对边界框回归任务更为重要;低频分量主要包含整体结构与语义信息,对分类任务更为关键。

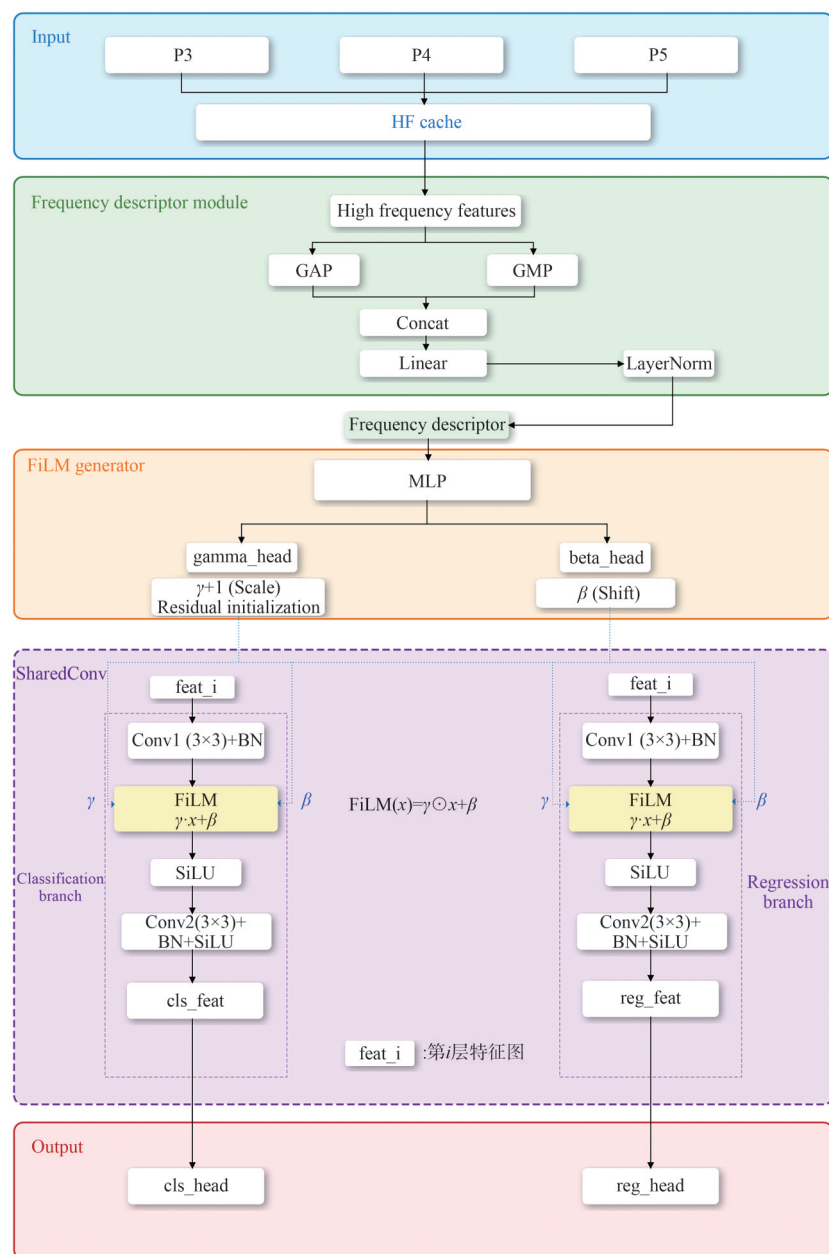


图 4 FGSHHead 结构图

Fig. 4 FGSHHead structure diagram

对于输入特征 $F \in R^{C \times H \times W}$, 首先通过 HaarWavelet 变换提取频率分量:

$$LL, LH, HL, HH = \text{HaarWavelet}(F). \quad (11)$$

基于小波变换后提取的频率分量, 使用公式 (12) 计算高频能量比 E_{HF} , 作为频率特征描述子:

$$E_{\text{HF}} = \frac{|LH|_F^2 + |HL|_F^2 + |HH|_F^2}{|LL|_F^2 + |LH|_F^2 + |HL|_F^2 + |HH|_F^2}, \quad (12)$$

其中, $|\cdot|_F$ 表示 Frobenius 范数。高频能量比 E_{HF} 反映了输入特征中边缘信息的丰富程度。

基于频率特征描述子, FGSHHead 设计了双分支动态特征调制机制, 分别为分类分支与回归分支生成差异化的特征表示。分类任务需要强调低频语义信息, 因此采用低频增强策略生成分类分支特征 F_{cls} :

$$F_{\text{cls}} = F \odot (1 + \lambda_{\text{cls}} \cdot (1 - E_{\text{HF}}) \cdot M_{\text{cls}}), \quad (13)$$

其中: $M_{\text{cls}} \in R^{C \times 1 \times 1}$ 为可学习的通道注意力向量, λ_{cls} 为分类分支的调制强度系数, \odot 表示逐元素乘法。

回归任务需要强调高频边缘信息, 因此采用高频增强策略生成回归分支特征 F_{reg} :

$$F_{\text{reg}} = F \odot (1 + \lambda_{\text{reg}} \cdot E_{\text{HF}} \cdot M_{\text{reg}}), \quad (14)$$

其中: $M_{\text{reg}} \in R^{C \times 1 \times 1}$ 为回归分支的通道注意力向量, λ_{reg} 为回归分支的调制强度系数。

通过上述调制机制, 分类分支与回归分支在共享基础特征的同时, 能够根据各自任务的需求对特征进行差异化增强。

为缓解多尺度特征分布不一致的问题, FGSHHead 引入了尺度自适应归一化策略。该策略为每个尺度学习独立的归一化参数, 同时共享卷积核参数。

设第 s 个尺度的输入特征为 F_s , 自适应归一化定义如下:

$$\text{SAN}_s(F_s) = \gamma_s \cdot \frac{F_s - \mu_s}{\sqrt{\sigma_s^2 + \epsilon}} + \beta_s, \quad (15)$$

其中: μ_s 与 σ_s^2 分别为第 s 个尺度特征的均值与方差, γ_s 与 β_s 为尺度相关的可学习仿射参数, ϵ 是用于数值稳定的极小常数。

通过尺度自适应归一化, 不同尺度的特征在进入共享卷积层之前被归一化至相似分布空间, 从而提升共享参数对多尺度特征的适应能力。

FGSHHead 采用参数共享的解耦检测头架

构, 包含共享的特征提取模块与独立的任务预测模块。对于共享特征提取模块, 即对调制后的特征进行深度特征提取, 采用堆叠的深度可分离卷积 (Depthwise Separable Convolution)^[27] 实现高效计算:

$$F_{\text{shared}} = \text{DWConv}_N \circ \dots \circ \text{DWConv}_i(F_{\text{in}}), \quad (16)$$

其中: F_{shared} 表示共享特征提取模块的输出特征, DWConv_i 表示第 i 个深度可分离卷积层, N 为堆叠层数。

任务预测模块分为分类预测模块和回归预测模块。其中分类预测模块可将分类分支特征映射为类别概率分布 P_{cls} :

$$P_{\text{cls}} = \text{Sigmoid}(\text{Conv}_{1 \times 1}^{\text{cls}}(F_{\text{shared}}^{\text{cls}})) \in \mathbb{R}^{N_{\text{cls}} \times H \times W}, \quad (17)$$

其中, N_{cls} 为类别数量。

回归预测模块可将回归分支特征映射为边界框参数 P_{reg} :

$$P_{\text{reg}} = \text{Conv}_{1 \times 1}^{\text{reg}}(F_{\text{shared}}^{\text{reg}}) \in \mathbb{R}^{4 \times H \times W}. \quad (18)$$

输出的 4 个通道分别对应边界框的中心偏移 (dx, dy) 与尺寸缩放 (dw, dh)。

综上所述, FGSHHead 的设计不仅优化了模型的部署效率, 更关键的是利用频率引导的双分支调制与自适应归一化技术, 成功实现了任务感知的差异化特征处理及多尺度特征的分布对齐。加之全架构的可微分特性支持端到端优化, 使得该模块在保持结构紧凑的同时, 显著增强了对不同尺度目标的检测能力与分类回归的综合性能。

2.5 模块协同与整体架构

通过上述 3 个核心模块的有机协同, FW-YOLO 框架构建了围绕频率域信息显式建模与传递的闭环架构。具体而言, 特征提取阶段利用 FGDC 通过 Haar 小波变换实现了高频边缘信息的有效保留与内容自适应提取; 随后, 上采样阶段借助 FRUpsample 将存储的高频分量注入, 完成了边缘信息的跨尺度传递与增强; 最终, 检测阶段利用 FGSHHead 依据频率特性引导分类与回归任务的差异化处理。该整体架构不仅遵循了频率域显式建模与任务驱动利用的核心设计原则, 克服了传统卷积网络隐式丢失高频信息的缺陷, 还有效兼顾了计算效率与检测性能的平衡, 为密集行人检测任务提供了强有力的特征表达基础。

3 实验结果及分析

3.1 数据集介绍

为全面评估本文所提算法的检测性能与泛化能力,实验选取 CrowdHuman^[8]、WiderPerson^[28]和 BDD100K^[29]3个公开数据集进行模型训练与测试验证。通过在不同来源、不同特性的数据集上进行实验对比,能够充分检验算法在多样化场景下的适应性与稳定性。

CrowdHuman 是针对高密度人群场景构建的行人检测数据集,收录约 24 000 张高清图像,其中训练集 15 000 张、验证集 4 370 张、测试集 5 000 张,累计标注行人实例超过 47 万个。该数据集的显著特征在于单幅图像中行人数量众多、个体间相互遮挡现象普遍,同时涵盖街道、商场、车站等多类真实场景,标注信息详尽准确。CrowdHuman 为评估模型在目标高度密集、遮挡关系复杂条件下的检测表现提供了理想的测试平台。

WiderPerson 是一个面向拥挤场景的大规模行人检测基准数据集,总计包含 13 382 张图像,标注行人实例约 40 万个。数据集按照 8 000 张、1 000 张、4 382 张的比例分别构建训练集、验证集和测试集。标注类别包括行人、骑车人、部分可见目标、人体模型及密集人群等,本实验将上述类别统一视为行人进行检测。WiderPerson 数据集具有场景覆盖广泛、目标遮挡程度高、标注实例数量大等特点,能够有效评估检测模型应对复杂拥挤环境的能力。

BDD100K 源自多个城市的实际行车记录视频,是一个大规模自动驾驶场景数据集。该数据集覆盖了晴朗、多云、雨天、雪天、雾天等多种气象状况,包含高速公路、城区道路、居民区、停车场、隧道等典型交通场景,拍摄时间从清晨延续至深夜,充分反映了真实驾驶环境中光照条件与天气因素的多变性。本文针对行人检测任务对原始数据进行筛选处理,仅提取行人类别标注,并按照 8:1:1 的比例重新划分为训练集、验证集与测试集。BDD100K 场景类型丰富、环境条件多变,适合用于检验模型在不同气象与光照条件下的鲁棒性表现。

图 5 为 3 个公开数据集中部分实例图,从



图 5 数据集示例图

Fig. 5 Sample dataset image

上至下分别为 CrowdHuman、WiderPerson 和 BDD100K。

3.2 实验环境设置

本文所有实验均在统一的软硬件平台上进行,以确保实验结果的可比性和可复现性。具体实验环境配置如表 1 所示。

表 1 实验环境配置

Tab. 1 Experimental environment configuration

参数	配置
系统环境	Windows 11
CPU	AMD 9700X
GPU	RTX 5070ti
深度学习框架	Pytorch 2.10.0
CUDA 环境	CUDA 12.8
Python 环境	Python 3.12.4
Ultralytics	Ultralytics 8.3.238

为保证实验的公平性,本文对基线模型和改进模型采用相同的训练策略和超参数设置。具体训练参数配置如表 2 所示。

训练轮数设置为 300 轮,确保训练结果收敛; batch-size 设置为 16,输入图片尺寸统一为 640 × 640;为了避免消融实验与泛化实验所带来的误差,将初始学习率统一设置为 0.01,并使用 SGD 优化器对网络参数加以优化。

表2 实验参数配置

Tab. 2 Experimental parameter configuration

参数	配置
Epochs	300
Batch-size	16
Image-size	640×640
Optimizer	SGD
Lr0	0.01
Lrf	0.1
Workers	4
Warmup-epochs	3

3.3 评价指标

为全面评估模型的检测性能,本文采用目标检测领域广泛使用的评价指标体系,包括精确率(Precision, P)、召回率(Recall, R)、平均精度均值(mean Average Precision, mAP)。

精确率表示模型预测为正样本的结果中,真正为正样本的比例,反映了模型预测结果的准确程度,计算公式为:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (19)$$

其中:TP(True Positive)表示模型正确检测到的目标数量,当预测边界框与真实边界框的交并比(IoU)超过预设阈值(通常为0.5),且预测类别与真实类别一致时,该检测结果被判定为真正例。FP(False Positive)表示模型错误检测的目标数量,即“误检”。产生误检的情况主要包括两类:一是将背景区域错误地检测为行人;二是预测框与真实框的IoU低于阈值,导致定位不准确而被判定为错误检测。在密集场景中,由于目标间相互遮挡,误检问题尤为突出。精确率的取值范围为 $[0, 1]$,值越大表示模型的误检率越低。当精确率为1时,意味着模型的所有预测结果都是正确的,不存在任何误检。在实际应用中,高精确率意味着系统发出的检测警报具有较高的可信度,能够减少因误报带来的资源浪费和用户干扰。

召回率表示所有真正样本中被正确检测出的比例,反映了模型对目标的检出能力,计算公式为:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (20)$$

其中,FN(False Negative)表示模型未能检测到的真实目标数量,即“漏检”。当真实存在的行人未被任何预测框正确匹配时,该目标被计为假负例。在拥挤场景下,由于严重的遮挡和目标重叠,漏检是影响检测性能的主要因素之一。其余符号含义等同于公式(19)中的含义。

平均精度均值(mean Average Precision, mAP)是在多类别目标检测任务中评估模型整体性能的综合指标,它将各类别的AP值进行平均,得到一个统一的性能度量。mAP定义为所有类别AP值的算术平均,其数学表达式为:

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^N \text{AP}_i, \quad (21)$$

其中: N 为检测任务中的类别总数, AP_i 为第 i 个类别的平均精度。mAP的取值范围为 $[0, 1]$,值越大表示模型的整体检测性能越好。在行人检测任务中,由于通常只关注单一类别(行人),此时mAP等于该类别的AP值。

根据IoU阈值的不同设定,mAP可以细分为多种变体,本文主要采用mAP50以及mAP50:95两种指标。

mAP50是指在IoU阈值为0.5时计算的mAP值。这是目标检测领域最常用的评价指标。mAP50对边界框定位精度的要求相对宽松,主要评估模型是否能够正确检测到目标。计算公式为:

$$\text{mAP50} = \frac{1}{N} \sum_{i=1}^N \text{AP}_i|_{\text{IoU}=0.5}. \quad (22)$$

mAP50:95^[30]是在IoU阈值0.5~0.95(步长0.05)的10个阈值下分别计算mAP,然后取平均值。这对模型的定位精度提出了更高要求。计算公式为:

$$\text{mAP50:95} = \frac{1}{10} \sum_{t \in \{0.5, 0.55, \dots, 0.95\}} \text{mAP}_t, \quad (23)$$

其中, mAP_t 为在阈值 t 下计算得到的mAP。mAP50:95相比mAP50更加严格,能够更好地区分不同模型在边界框定位精度上的差异。一个在mAP50:95上表现优异的模型,不仅能够准确检测目标,还能够提供精确的边界框定位。

3.4 实验结果分析

本节从消融实验、对比实验、可视化分析和泛化性验证4个方面对所提出的改进方法进行全方位评估。

3.4.1 消融实验

为验证本文所提各个改进模块的有效性,本节设计了一系列消融实验。以YOLOv11n为基线模型,逐步添加或两两结合各改进模块,分析每个模块对检测性能的贡献。实验结果如表3所示,分别考察了FGODC(A)、FRUpsample(B)、FGSHHead(C)在不同组合下对模型评估指标的影响。

表3 消融实验结果

Tab. 3 Ablation experiment results

A	B	C	P/%	R/%	mAP50/%	mAP50:95/%
		Baseline	83.4	64.8	75.5	47.7
✓			84.4	65.4	76.0	48.4
	✓		84.4	64.5	75.5	47.9
		✓	83.4	64.1	75.4	47.6
✓	✓		84.5	65.8	76.4	49.0
		✓	83.7	64.4	75.4	48.3
✓		✓	84.3	66.5	76.6	49.0
✓	✓	✓	84.5	68.6	77.3	49.7

由表3可知,加入FGODC模块替换骨干网络中的传统步长卷积,模型的精确率、召回率、mAP50、mAP50:95分别提升了1.0%、0.6%、0.5%、0.7%。该模块通过Haar小波变换将输入特征显式分解为低频近似分量与高频细节分量,并利用高频响应强度指导动态卷积核的自适应选择,有效缓解了传统下采样过程中边缘信息丢失的问题。在密集行人检测任务中,显式保留高频边缘信息对区分相邻遮挡目标具有重要作用。

加入FRUpsample模块替换颈部网络中的传统上采样,模型的精确率、mAP50:95分别提升了1.0%、0.2%。传统上采样方法本质上是低通滤波器,在放大特征图的同时不可避免地丢失高频细节信息。FRUpsample模块通过Haar小波变换从下采样路径提取高频残差,并设计空间自适应注意力门控机制,根据局部高频响应强度动态调整融合权重,使边缘区域获得强高频补偿而平滑区域抑制噪声注入。该设计有效解决了跨尺度特征融合过程中边缘模糊的问题,提升了模型对小目标和遮挡目标边界的定位精度。

加入FGSHHead模块替换原始检测头,模型的4个指标并没有获得提升。这是由于其与前两个模块达成互补,需要使用前两个模块提供的低频

与高频分量进行计算来获得高频能量比作为频率特征描述子,并针对分类分支采用低频增强策略、针对回归分支采用高频增强策略。

通过两两组合消融可以得出,在骨干网络中引入FGODC并在颈部网络中引入FRUpsample后,模型的精确率、召回率、mAP50、mAP50:95分别提升了1.1%、1.0%、0.9%、1.3%。这一结果表明,FGODC在下采样阶段提取并保留的高频边缘信息能够被FRUpsample在上采样阶段有效利用,两个模块在频率域形成了信息的闭环传递。具体而言,FGODC通过Haar小波变换分离的高频分量为FRUpsample提供了精确的边缘残差信息,而FRUpsample的空间自适应门控机制能够根据局部特征的高频响应强度智能地融合这些残差,从而在特征金字塔的各层级间实现边缘信息的无损传递。

在上述基础上进一步引入FGSHHead模块,构成完整的改进模型。实验结果表明,完整模型的mAP50达到77.3%,mAP50:95达到49.7%,相比基线模型分别提升了1.8%和2.0%;精确率和召回率分别达到84.5%和68.6%,相比基线模型分别提升了1.1%和3.8%。

上述实验结果充分验证了本文所提3个改进模块的有效性与协同性。FGODC在特征提取阶段显式保留高频边缘信息,FRUpsample在上采样阶段实现边缘信息的跨尺度传递与增强,FGSHHead在目标检测阶段根据任务需求差异化利用频率信息。3个模块围绕频率域信息的提取、传递与利用形成完整的技术闭环,协同解决了密集行人检测中边缘信息丢失与多尺度目标检测的关键问题。

3.4.2 模型对比实验

为了进一步验证本文算法在密集行人检测中的有效性,本文将提出的FW-YOLO模型与YOLO系列进行比较,包括YOLOv8^[2]、YOLOv10^[31]、YOLOv12^[32]、YOLOv26^[4]等模型以及近年来的轻量化模型GS-YOLO^[33]、DAMO-YOLO^[34]、LDD-YOLO^[35]和YOLO-CROWD^[36]在相同环境使用CrowdHuman数据集进行实验。实验结果如表4所示,其中‘-n’表示Nano模型,‘-m’表示Middle模型,‘-s’表示Small模型,粗体数字表示最优值,下划线数字表示次优值。

表 4 CrowdHuman 数据集上不同模型的对比结果

Tab. 4 Comparison of different models on the CrowdHuman dataset

Network architecture	P/%	R/%	mAP50/%	mAP50:95/%	Params/M	GFLOPs	FPS/
YOLOv8-n	84.3	65.1	75.6	48.1	3.2	8.9	431
YOLOv10-n	83.4	66.1	76.3	<u>48.4</u>	2.8	8.7	<u>304</u>
YOLOv11-n(Baseline)	83.4	64.8	75.5	47.7	2.6	6.6	293
YOLOv12-n	84.3	64.5	75.3	47.8	2.6	6.7	160
YOLOv26-n	82.1	64.9	74.8	47.7	<u>2.4</u>	5.3	133
YOLO-CROWD-s	<u>84.4</u>	70.3	76.9	44.5	18.4	20.6	99
DAMO-YOLO-m	76.7	61.1	73.3	41.1	28.2	61.8	233
LDD-YOLO-n	81.1	65.5	77.4	46.6	4.7	1.8	—
GS-YOLO-n	78.2	62.6	73.7	38.6	1.7	<u>3.8</u>	100
FW-YOLO(Ours)	84.5	<u>68.6</u>	<u>77.3</u>	49.7	12.3	40.2	155

由表4可以看出,在与YOLO系列模型的对比中,FW-YOLO的检测精度整体优于YOLOv8-n、YOLOv10-n、YOLOv11-n、YOLOv12-n等模型,尤其在精确率和mAP50:95指标上具有明显提升。与最新的YOLOv26-n模型相比,FW-YOLO在精确率、召回率、mAP50和mAP50:95上分别提升0.2%、3.7%、2.5%和2.0%,进一步印证了其在结构优化方面的有效性。

与其他改进算法相比,FW-YOLO的精确率和召回率高于其他改进算法,说明模型在密集场景下对目标边界与冗余背景的分隔能力更强。与mAP指标最高的LDD-YOLO-n模型以及召回率指标最高的YOLO-CROWD-s模型相比,虽然FW-YOLO的mAP50指标与召回率均有所降低,但是mAP50:95指标却高出LDD-YOLO-n模型3.1%,高出YOLO-CROWD-s模型5.2%。这证明了FW-YOLO模型在处理密集遮挡人群的高效性和准确性,在实现高精度检测的同时兼顾“不漏检”、“不误检”。

为全面评估改进对模型复杂度与推理效率的影响,对模型参数量(Params)、计算量(GFLOPs)与推理速度(FPS)进行补充。与基线YOLOv11-n相比,FW-YOLO的Params由2.6M增加至12.3M,GFLOPs由6.6增加至40.2,说明频率引导与小波增强模块带来额外计算开销。由于预处理、后处理(如图像解码/resize、NMS等)与具体工程实现、库版本及算子优化强相关,且并非结构改进的核心贡献点,为保证不同模型对比的公平性与可复现性,FPS仅统计模型前向传播

(Forward)耗时,用以反映网络结构本身带来的推理开销变化。在输入尺寸 640×640 、batch=1、FP16关闭的统一设置下,FW-YOLO可达到155FPS(约6.45ms/帧),表明该方法在引入一定计算代价的同时仍具有较高的推理效率。因此,FW-YOLO体现了以适度计算代价换取密集遮挡场景下精确率、召回率与mAP提升的精度和效率的权衡,适用于对检测精度与定位质量要求更高的应用场景。

总体而言,FW-YOLO在高频边缘信息表征与多尺度特征融合方面表现出较好的有效性,能够更充分地挖掘密集场景下目标的边界与细节信息,从而提升目标检测的准确性。相较于原模型与其他模型,该模型在复杂场景中的目标识别与定位能力均有所增强,体现出该方法在密集目标检测任务中的优越性。

3.4.3 可视化结果分析

为评估各模型在密集场景下的检测效果,本文选取CrowdHuman数据集中具有代表性的3张测试图像进行可视化对比,所选样本涵盖了远距离、密集、遮挡等典型场景。图6展示了基线模型与本文方法在相同输入下的检测结果对比。

由图6可以观察到,在密集遮挡场景中,基线模型存在明显的漏检问题,从图中第二行左图可以看出,红圈标注区域并没有检测到任何目标。但FW-YOLO模型的检测结果表明,该模型能够准确检测出密集遮挡区域的微小目标。

进一步观察图6中的第1行与第3行的对比图片可以发现,基线模型对正常的行人目标检测



图6 检测结果可视化对比

Fig. 6 Visualization comparison of detection results

置信度为80%左右,而FW-YOLO模型的置信度平均在85%~90%左右,验证了FW-YOLO在检测精度方面的有效性。

上述整体结果验证了FW-YOLO模型在密

集遮挡场景下,特别是在密集遮挡区域的小目标检测方面,展现出更强的目标检测能力。进一步说明该模型具备更优的检测性能和实际应用价值。

3.4.4 模型泛化性验证

泛化性是衡量模型实用价值的重要指标,良好的泛化能力意味着模型能够适应不同场景和数据分布的变化。为验证本文方法的泛化能力,本文选取 WiderPerson 和 BDD100K 两个数据集进行跨数据集泛化性验证。这两个数据集与 CrowdHuman 在场景分布、图像风格和标注方式上存在一定差异,能够有效评估模型的泛化能力。

WiderPerson 是一个大规模行人检测数据集,包含 13 382 张图像,约 40 万个行人标注。该数据集涵盖了多种场景,包括运动场、街道、购物中心等,行人密度和遮挡程度各异。

BDD100K 是一个大规模自动驾驶数据集,包含 10 万张驾驶场景图像。本文选取其中的行人检测子集进行实验,该数据集的图像主要来自车载摄像头,与 CrowdHuman 的互联网图像在视角和场景上存在明显差异。

泛化性对比实验结果如表5所示。

表5 泛化性对比实验结果

Tab. 5 Generalization comparison experiment results

Dataset	Model	$P/\%$	$R/\%$	mAP50/%	mAP50:95/%
WiderPerson	YOLOv11-n	81.1	64.1	75.3	48.3
	FW-YOLO	82.8	64.9	76.4	49.1
BDD100K	YOLOv11-n	73.3	57.0	64.9	32.7
	FW-YOLO	73.8	57.1	65.1	32.8

由表5可知,在 Widerperson 数据集上的泛化实验中,FW-YOLO 在精确率、召回率、mAP50、mAP50:95 四个方面分别提升了 1.7%、0.8%、1.1%、0.8%,有效降低了在密集场景下的误检率与漏检率,验证了其在密集行人检测场景下的鲁棒性和准确性。

在 BDD100K 数据集检测中,FW-YOLO 相比基线模型效果提升有限,这是由于 BDD100K 数据集中包含晴天、阴天、雨、雪、雾等恶劣天气场景,且覆盖从白天到夜间多个时间段。由此得出,FW-YOLO 虽然在密集行人检测方面提升显著,但在复杂环境和恶劣天气条件下的适应能力

并不强。

3.4.5 恶劣场景下的频率域可视化分析

为进一步分析模型在恶劣天气场景下的表现,我们从频率域角度对雨夜眩光样本进行了定性可视化对比。如图7所示,基线模型与 FW-YOLO 均能够在主要目标区域产生较强响应,但 FW-YOLO 在 Backbone 与 Neck 两个阶段的高频能量分布更强、更广。除真实行人边界外,雨丝、湿滑路面的高反射区域以及灯光眩光边缘也被明显激活,说明这些由天气退化和成像干扰引起的非语义结构,可能在 Haar 小波分解中被部分映射为高频边缘信息,并在后续频率增强过程中被进一步强化。

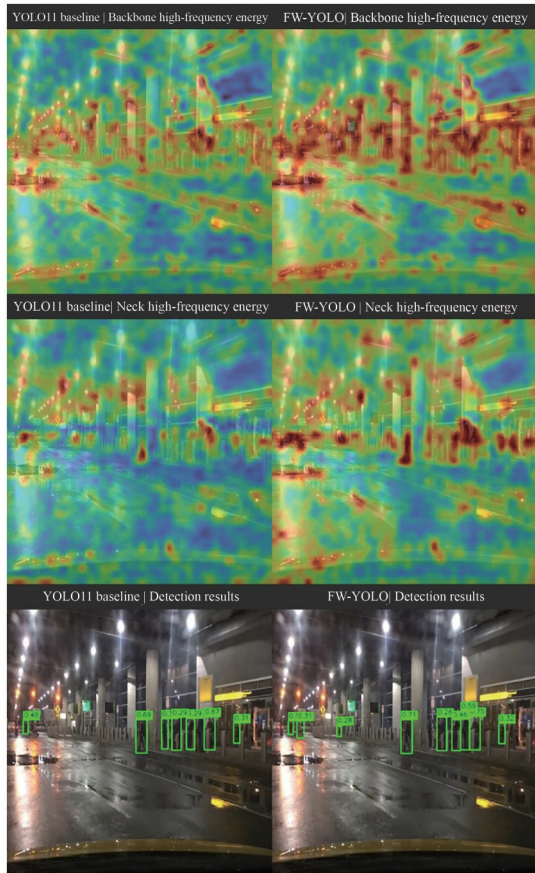


图7 基线模型与FW-YOLO在雨夜眩光场景下的频率域可视化对比

Fig.7 Frequency-domain visual comparison between Baseline and FW-YOLO in rainy-night glare scenes

结合第3行的检测结果可以看出,FW-YOLO仍能保持对主要行人的有效检测,但其高频响应并未完全集中于目标轮廓,而是部分扩散到背景干扰区域。这一现象表明,频率增强机制在复杂雨夜场景中存在一定的过度增强风险,可能限制模型在BDD100K恶劣天气条件下的鲁棒性。

在雪天场景中,类似现象同样可以观察到。如图8所示,相较于基线模型,FW-YOLO在Backbone与Neck两个阶段均呈现出更强的高频能量响应,且这种响应不仅分布于行人边界附近,也覆盖了积雪纹理、路面颗粒扰动以及建筑边缘等背景区域。由于雪粒和积雪纹理本身会引入大量细碎的局部高频成分,这些非语义结构在Haar小波分解后可能被部分映射为边缘性高频信息,并在后续频率增强过程中得到进一步放大。

从第3行的检测结果可以看出,FW-YOLO对主要目标仍保持较好的检测能力,但其中间频

率响应表现出更明显的背景扩散趋势。这说明在雪天场景下,频率增强机制同样存在将天气退化纹理与真实目标边界共同强化的倾向,从而可能削弱模型对复杂背景噪声的判别能力。该现象进一步支持了本文对于恶劣天气条件下鲁棒性受限原因的分析。

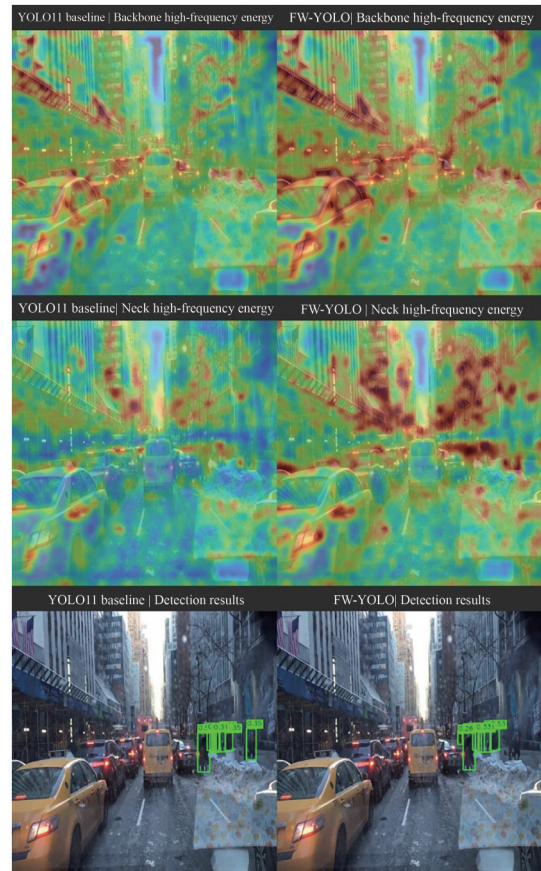


图8 基线模型与FW-YOLO在雪天场景下的频率域可视化对比

Fig.8 Frequency-domain visual comparison between Baseline and FW-YOLO in snowy scenes

通过对雨夜和雪天典型样本中基线模型与FW-YOLO在Backbone、Neck以及最终检测结果上的对比可以观察到,所提出的频率增强机制在强化目标轮廓信息的同时,也会对部分由天气退化引起的高频扰动产生额外响应。这一现象为FW-YOLO在BDD100K恶劣天气场景下性能提升受限提供了一种合理解释。后期工作可以围绕增强恶劣天气下的密集行人检测任务进行改进与提升。

4 结 论

针对密集行人检测任务中普遍存在的高频边缘信息丢失、上采样特征模糊及多尺度适应性差等关键问题,本文提出了一种基于频率引导和小波增强的目标检测算法——FW-YOLO。该算法创新性地构建了覆盖“下采样-上采样-检测头”全链路的频率域闭环设计,通过引入FGODC、FRUpsample以及FGSHHead三个核心模块,实现了高频边缘信息的显式保留、精确恢复与任务自适应利用。

实验结果充分验证了FW-YOLO在密集遮挡场景下的卓越性能。在CrowdHuman数据集上的测试表明,改进后模型的精确率、召回率、mAP50和mAP50:95分别达到了84.5%、68.6%、77.3%和49.7%。与基线模型相比,上述指标分别提升了1.1%、3.8%、1.8%和2.0%。特别是在衡量高精度定位能力的mAP50:95指标上,

FW-YOLO优于最新的YOLOv26-n模型2.0%,且召回率的大幅提升证明了该方法能有效缓解密集人群中的漏检现象。消融实验进一步证实,FGODC与FRUpsample的协同作用在频率域形成了有效的信息闭环,显著增强了特征金字塔对轮廓细节的重构能力。

此外,泛化性实验表明,FW-YOLO在WiderPerson数据集上仍保持了较好的检测性能,mAP50达到76.4%,说明该方法在复杂拥挤场景下具有一定的泛化能力。然而,在包含雨、雪、雾等复杂气象条件的BDD100K数据集上,模型性能提升相对有限,表明当前方法在应对恶劣天气导致的图像退化、边缘弱化和细节干扰方面仍存在不足。后续研究将进一步结合复杂环境下的鲁棒特征建模与退化场景增强策略,以提升模型在真实开放场景中的适应能力,并拓展其在车载显示辅助感知、智能监控显示及公共安全可视化等成像与显示相关应用中的实际价值。

参 考 文 献:

- [1] TERVEN J, CORDOVA-ESPARZA D M, ROMERO-GONZALEZ J A. A comprehensive review of YOLO architectures in computer vision: from YOLOv1 to YOLOv8 and YOLO-NAS [J]. *Machine Learning and Knowledge Extraction*, 2023, 5(4): 1680-1716.
- [2] JOCHER G, CHAURASIA A, QIU J, *et al.* Ultralytics YOLO: ultralytics YOLOv8 [EB/OL]. 2023[2026-04-16]. <https://github.com/ultralytics/ultralytics>.
- [3] KHANAM R, HUSSAIN M. YOLOv11: an overview of the key architectural enhancements [J/OL]. *arXiv*, 2024: 2410.17725.
- [4] SAPKOTA R, CHEPPALLY R H, SHARDA A, *et al.* YOLO26: key architectural enhancements and performance benchmarking for real-time object detection [J/OL]. *arXiv*, 2025: 2509.25164.
- [5] LI X L, MA L S. DED-YOLOv8: dense pedestrian detection algorithm based on YOLOv8 [C]. 2024 4th International Symposium on Computer Technology and Information Science (ISCTIS 2024). Xi'an: IEEE, 2024: 545-548.
- [6] OUYANG D L, HE S, ZHANG G Z, *et al.* Efficient multi-scale attention module with cross-spatial learning [C]// *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Rhodes Island: IEEE, 2023: 1-5.
- [7] FANG Y C, PANG H L. An improved pedestrian detection model based on YOLOv8 for dense scenes [J]. *Symmetry*, 2024, 16(6): 716.
- [8] SHAO S, ZHAO Z J, LI B X, *et al.* CrowdHuman: a benchmark for detecting human in a crowd [J/OL]. *arXiv*, 2018: 1805.00123.
- [9] SUI H, HAN H Y, CUI Y Z, *et al.* FA-YOLO: A pedestrian detection algorithm with feature enhancement and adaptive sparse self-attention [J]. *Electronics*, 2025, 14(9): 1713.
- [10] ZHANG H Y, ZHANG Q H, GONG Y F, *et al.* MDCFVit-YOLO: a model for nighttime infrared small target vehicle and pedestrian detection [J]. *PLoS One*, 2025, 20(6): e0324700.

- [11] 曹洁, 牛瑜, 梁浩鹏. 基于优化权重的 YOLOv7 密集行人检测算法[J]. 液晶与显示, 2025, 40(3): 505-515.
CAO J, NIU Y, LIANG H P. Dense pedestrian detection algorithm based on YOLOv7 with optimized weights [J]. *Chinese Journal of Liquid Crystals and Displays*, 2025, 40(3): 505-515. (in Chinese)
- [12] 马晞茗, 李宁, 吴迪, 等. 基于头部与整体信息联合的行人检测算法[J]. 光学精密工程, 2025, 33(14): 2278-2290.
MA X M, LI N, WU D, et al. Pedestrian detection algorithm based on joint head and overall information [J]. *Optics and Precision Engineering*, 2025, 33(14): 2278-2290. (in Chinese)
- [13] MALLAT S G. A theory for multiresolution signal decomposition: the wavelet representation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1989, 11(7): 674-693.
- [14] LI C, ZHOU A J, YAO A B. Omni-dimensional dynamic convolution [C]//*Proceedings of the 10th International Conference on Learning Representations (ICLR)*. Online: OpenReview.net, 2022.
- [15] WANG J Q, CHEN K, XU R, et al. CARAFE: content-aware ReAssembly of FEatures [C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul: IEEE, 2019: 3007-3016.
- [16] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C]//*Proceedings of the 15th European Conference on Computer Vision (ECCV)*. Munich: Springer, 2018: 3-19.
- [17] PEREZ E, STRUB F, DE VRIES H, et al. FiLM: visual reasoning with a general conditioning layer [C]//*Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*. New Orleans: AAAI Press, 2018, doi: 10.1609/aaai.v32i1.11671.
- [18] SHANNON C E. Communication in the presence of noise [J]. *Proceedings of the IRE*, 1949, 37(1): 10-21.
- [19] LIN M, CHEN Q, YAN S C. Network in network [J/OL]. *arXiv*, 2013: 1312.4400.
- [20] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagating errors [J]. *Nature*, 1986, 323(6088): 533-536.
- [21] 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 (CVPR)*. Honolulu: IEEE, 2017: 2117-2125.
- [22] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Salt Lake City: IEEE, 2018: 8759-8768.
- [23] LIU W Z, LU H, FU H T, et al. Learning to upsample by learning to sample [C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. Paris: IEEE, 2023: 6027-6037.
- [24] GE Z, LIU S T, WANG F, et al. YOLOX: exceeding YOLO series in 2021 [J/OL]. *arXiv*, 2021: 2107.08430.
- [25] DAI X Y, CHEN Y P, XIAO B, et al. Dynamic head: unifying object detection heads with attentions [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Nashville: IEEE, 2021: 7373-7382.
- [26] HUANG X, BELONGIE S. Arbitrary style transfer in real-time with adaptive instance normalization [C]//*Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. Venice: IEEE, 2017: 1501-1510.
- [27] HOWARD A G, ZHU M L, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications [J/OL]. *arXiv*, 2017: 1704.04861.
- [28] ZHANG S F, XIE Y L, WAN J, et al. WiderPerson: a diverse dataset for dense pedestrian detection in the wild [J]. *IEEE Transactions on Multimedia*, 2020, 22(2): 380-393.
- [29] YU F, CHEN H F, WANG X, et al. BDD100K: a diverse driving dataset for heterogeneous multitask learning [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle: IEEE, 2020: 2636-2645.
- [30] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context [C]//*Proceedings of the 13th European Conference on Computer Vision (ECCV)*. Zurich: Springer, 2014: 740-755.
- [31] WANG A, CHEN H, LIU L H, et al. YOLOv10: real-time end-to-end object detection [J/OL]. *arXiv*, 2024: 2405.14458.
- [32] TIAN Y J, YE Q X, DOERMANN D. YOLOv12: attention-centric real-time object detectors [J/OL]. *arXiv*, 2025: 2502.12524.
- [33] LI H L, LI J, WEI H B, et al. Slim-neck by GSConv: a lightweight-design for real-time detector architectures [J].

Journal of Real-Time Image Processing, 2024, 21(3): 62.

- [34] XU X Z, JIANG Y Q, CHEN W H, *et al.* DAMO-YOLO: a report on real-time object detection design [J/OL]. *arXiv*, 2022: 2211.15444.
- [35] 杨迪,张喜龙,王鹏. LDD-YOLO:改进YOLOv8的轻量级密集行人检测算法[J]. *计算机科学与探索*, 2026, 20(1):251-265.
YANG D, ZHANG X L, WANG P. LDD-YOLO: a lightweight and dense pedestrian detection algorithm improving YOLOv8 [J]. *Computer Science and Exploration*, 2026, 20(1): 251-265. (in Chinese)
- [36] MERABET M Z. YOLO-CROWD [EB/OL]. [2026-04-16]. <https://github.com/zaki1003/YOLO-CROWD>.

作者简介:



黄子骏,男,硕士研究生,2025年于江苏师范大学科文学院获得学士学位,主要从事计算机视觉、目标检测方面的研究。
E-mail:hzj026977bb@163.com



凌萍,女,博士,副教授,2010年于吉林大学获得博士学位,主要从事人工智能、数据分析、目标检测方面的研究。E-mail:6020000012@jssnu.edu.cn