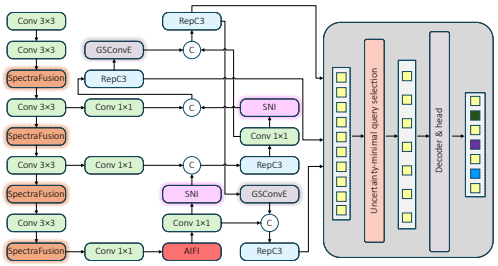
OEE | 光电工程 Opto-Electronic Engineering

Article

2026 年 5 月, 第 53 卷, 第 5 期

SRL-DETR：多场景轻量化行人检测算法



张　蕊[1,2]，李怡欣[1,2]，高　张[3*]，张　上[1,2]，朱　帅[1,2]

[1] 三峡大学计算机与信息学院，湖北 宜昌 443002；

[2] 三峡大学湖北省建筑质量检测装备工程技术研究中心，湖北 宜昌 443002；

[3] 三峡大学信息技术中心，湖北 宜昌 443002

摘要：针对小尺度和复杂场景下行人检测存在的精度不足、漏检和误检问题，以及推理效率和模型轻量化需求，提出一种改进 RT-DETR 的轻量化行人检测算法—SRL-DETR。首先，提出空间频率特征融合模块，以增强主干网络在空间与频率上的特征表达能力，提高对小目标的感知能力。其次，设计特征缩放精细增强模块，优化多尺度特征融合过程，从而提升模型对密集目标和复杂场景的检测性能。最后，引入基于通道重要性评估的自适应剪枝，在保留关键特征的前提下降低冗余计算。实验结果表明，所提算法在 CityPersons 数据集上 mAP50 提升至 71.6%，相较于基准模型提升 2.0%，参数量和计算量分别下降 50.5% 和 42.1%；并进一步在 TinyPerson 与 WiderPerson 数据集上进行泛化性能验证，mAP50 分别提升 3.7% 和 1.0%，展现出较强的跨场景适应性。综上所述，所提算法在显著提升行人检测精度的同时，有效保持了模型轻量化，适用于多种复杂场景。

关键词：行人检测；RT-DETR；小目标检测；多尺度；轻量化

The English title and abstract appear at the end of the paper.

DOI: 10.12086/oee.2026.250363 | CSTR: 32245.14.oee.2026.250363 | 中图分类号：TP391.41 | 文献标识码：A

张蕊，李怡欣，高张，等. SRL-DETR：多场景轻量化行人检测算法 [J]. 光电工程，2026，**53**(5): 250363

Zhang R, Li Y X, Gao Z, et al. SRL-DETR: lightweight pedestrian detection algorithm for multiple scenarios[J]. *Opto-Electron Eng*, 2026, **53**(5): 250363

1　引　言

随着智能监控、自动驾驶与智慧城市的发展，行人检测作为目标检测领域的重要研究方向，受到广泛关注。然而，复杂环境下的行人检测仍面临诸多挑战，例如在远距离或大范围监控场景中，行人目标往往尺寸较小、特征模糊；在人群密集区域，则易出现尺度差异大、遮挡严重等问题。如何在保持检测精度的同时，实现模型的轻量化与高效推理，成为当前行人检测研究的关键问题。

随着深度学习技术的迅速发展，行人检测算法取得了显著进展，逐步突破了传统方法在特征表达与检测精度方面的局限。当前，目标检测方法主要分为两阶段 (two-stage) 与一阶段 (one-stage) 两类。两阶段算法的代表包括 R-CNN[1]、Fast R-CNN[2]、Faster R-CNN[3] 和 Cascade R-CNN[4] 等，它们通常先通过区域提议网络生

收稿日期：2025-11-25；修回日期：2026-01-28；录用日期：2026-02-02；网络出版日期：2026-05-25

基金项目：湖北省国家级大学生创新创业训练计划 (S202311075047)；湖北省数字经济试点示范建设专项 (2312-420625-04-02-996363)；2024 年度三峡大学人才科研启动资金 (2024RCKJ074)；2024 年三峡大学教学研究重点项目 (J2024018)

*通信作者：高张，gaozhang@ctgu.edu.cn

版权所有©2026 中国科学院光电技术研究所

成候选框，再进行特征提取与目标分类。由于多阶段处理和区域裁剪，推理效率受到限制，难以满足实时应用的要求。相比之下，一阶段算法如 YOLO[5] (you only look once) 系列和 SSD[6] (single shot multibox detector) 等则跳过候选框生成阶段，直接在特征图上进行密集预测。这类方法结构更为紧凑，推理速度快，更适合嵌入式或边缘设备部署。

近年来，随着 Transformer 架构在计算机视觉中的广泛应用，端到端目标检测方法逐渐成为研究热点。代表性工作如 DETR[7] (detection transformer) 首次引入纯 Transformer 结构实现端到端目标检测，摆脱了传统检测框架中对候选框生成与后处理模块的依赖，显著简化了模型结构。然而，DETR 收敛速度较慢、对小目标不敏感，限制了其实际应用。为此，后续研究如 Deformable DETR 通过引入可变形注意力机制提升模型对不同尺度目标的适应性，显著改善了检测精度与收敛效率。在此基础上，RT-DETR[8] (real-time detection transformer) 进一步面向实时场景进行结构精简与解码器并行优化，有效缩短了推理延迟，同时在 COCO 等主流数据集上取得了良好的检测性能，展现出极强的实用性与推广价值。

为提升小目标与复杂场景下的检测稳定性，近年来的研究围绕模型结构设计与特征提取方法开展了多种改进，以增强目标检测算法的鲁棒性。陈海秀等[9] 提出 MLD-YOLO 算法，在 C2f 模块中引入自适应点移动与卷积门控机制，并构建 SLC-P2 特征金字塔和轻量检测头，有效提升密集场景下小目标行人检测的精度与鲁棒性。黄德意等[10] 提出了基于 YOLOv8 的多尺度特征增强算法 MFE-YOLO，设计 EGM 特征提取模块与 MLHS-FPN 融合结构，有效解决复杂场景下行人多尺度分布、密集遮挡及噪声干扰导致的漏检和误检问题。杨迪等[11] 提出 LDD-YOLO 算法，通过引入 RELAN、SPPF-LSKA 与 FFDM 等模块，在降低模型复杂度的同时有效提升了密集行人检测的精度。除 YOLO 系列外，RT-DETR 模型也受到了广泛关注，已有很多基于该模型的优化方法被提出。樊嵘等[12] 设计了一种基于注意力机制的改进方法，引入可变形注意力编码器以增强对多尺度特征的聚合能力，同时采用融合 CBAM 的 EfficientNet 主干网络提升特征提取与提纯效率，并通过损失函数优化提升模型收敛精度，但由于结构相对复杂，训练与部署效率仍有进一步提升空间。宋天泽等[13] 设计了 Pe-DETR 算法，通过引入 DWSE-FFN 模块强化局部上下文感知，同时引入 Swin Transformer-L 主干以实现多尺度特征的充分融合。实验结果表明该方法在精度方面有所提升，但整体结构较复杂，对计算资源要求较高，在轻量化部署中尚有优化空间。李青云等[14] 提出了 RSH-RTDETR 算法，结合轻量级 Regocn 主干结构、Haar 小波下采样模块及多尺度融合策略，显著提升了对遮挡和小目标的检测能力，并在降低参数量的同时保持较高精度。不过，该算法在极端密集的小尺度行人场景或更复杂的实际动态环境中，检测稳定性仍有进一步优化的空间。王栋等[15] 提出改进模型 MAF-RTDETR，通过引入多分支加权融合模块 MAFF 增强特征金字塔中不同层级特征的有效表征，结合 VSS 模块强化局部特征提取，并采用可学习位置编码提升空间适应性，有效提升了城市行人和车辆检测精度，优化了小目标与密集目标检测性能，但该模型在复杂场景下的泛化能力仍有提升空间。

尽管现有基于 RT-DETR 的行人检测研究通过多尺度特征增强、注意力机制引入或主干网络优化，在一定程度上提升了小目标与密集场景下的检测性能，但多数方法仍主要集中于局部结构或单一模块的改进，往往伴随模型复杂度和计算开销增加，使其在极端密集、小尺度行人场景下的检测稳定性与轻量化部署之间难以兼顾。在小目标和密集环境下，行人的姿态多样、外观差异、遮挡以及复杂背景容易导致特征信息丢失，从而增加漏检与误检风险；同时，不同光照、天气条件及个体差异进一步加大了检测难度。针对上述不足，本文提出一种基于改进 RT-DETR 的多场景轻量化行人检测算法 SRL-DETR，以提升复杂环境下小目标与密集行人的检测性能，同时在保持精度的前提下有效降低模型复杂度。本文主要改进如下：

1) 提出空间频率特征融合模块 (spectral-feature fusion network, SpectraFusion) 以优化主干网络结构，本文以 FasterNet Block 为基础，将其中的 PConv 替换为具备空间与频率选择能力的 SFSConv 算子 (space-frequency selection convolution, SFSConv)，并融合 CSP (cross-stage partial) 结构思想，实现特征流动与融合的优化。该设计有效增强了主干网络的空间建模能力与特征表达能力，尤其在处理小尺度行人及复杂背景时表现出更强的鲁棒性。

2) 设计特征缩放精细增强模块 (refined enhancement for feature scaling, REFS) 以提升特征融合效率。针对原始 RT-DETR 中多尺度特征融合阶段上下文建模能

力不足的问题，该模块引入 SNI (soft nearest-neighbor interpolation) 和 GSConvE (grouped shuffling convolution enhancement) 分别优化上采样与下采样过程，使多尺度特征融合更加高效、准确，从而增强模型对密集目标的感知能力。

3) 引入基于层自适应幅值的剪枝策略 (layer-wise adaptive magnitude-based pruning, LAMP) 以实现模型的轻量化优化。针对检测模型在实际部署中参数冗余与计算开销较高的问题，本文采用 LAMP 剪枝方法在全局范围内根据各层权重重要性自适应分配剪枝比例，从而在保持检测性能的同时有效降低参数量与计算量，提升模型的推理效率与部署可行性。

2 SRL-DETR 算法

RT-DETR 是一种基于 Transformer 的端到端目标检测模型，结合 CNN 特征提取与全局建模优势，在保持检测精度的同时实现较高推理效率，相比传统 DETR 更适合实时检测和轻量化部署。然而，在复杂场景下，RT-DETR 对小尺度行人和局部细节特征的建模能力有限，颈部多尺度信息融合也不充分。这些局限在密集区域或远距离目标检测中容易导致特征弱化和漏检，同时其整体计算量仍限制了进一步轻量化部署。

为了在复杂环境下兼顾行人检测精度与模型轻量化，本文提出 SRL-DETR 算法，名称来源于本文提出的三个关键模块 SpectraFusion、REFS 和 LAMP 的首字母，概括了算法的整体改进思路。该算法基于 RT-DETR-R18 进行改进，网络结构如图 1 所示。首先，提出空间频率特征融合模块，在主干网络中引入 SFSConv 以增强空间与频率建模能力，并结合 CSP 结构实现特征分流与融合。接着，设计特征缩放精细增强模块，结合 SNI 与 GSConvE 优化颈部特征融合结构，增强多尺度信息感知。最后，引入基于层自适应幅值的 LAMP 剪枝策略，根据各层权重的重要性自适应分配剪枝比例，在保证检测精度的前提下有效降低模型复杂度与计算开销，实现高效的轻量化检测模型。

2.1 空间频率特征融合模块 (SpectraFusion)

传统 ResNet-18 主干虽然引入残差学习机制，但在提取复杂空间信息和处理小尺度目标方面存在不足，且多尺度特征融合能力有限，易导致细节丢失，影响检测精度。为此，本文提出空间频率特征融合模块，将 FasterNet Block[16] 中的 PConv 替换为具备空间频率建模能力的 SFSConv[17]，并融合 CSP 结构，在主干网络中引入空间域与频率域信息的协同建模机制，详细结构如图 2 所示。该设计不仅提升了多尺度特征的感知与表达能力，同时减少了冗余计算，增强了网络在复杂背景和小目标场景下的鲁棒性，有效弥补了 ResNet-18 在结构上的不足。

空间频率特征融合模块主要包括三个阶段：特征划分、空间与频率信息的混合处理以及最终的特征重构融合。

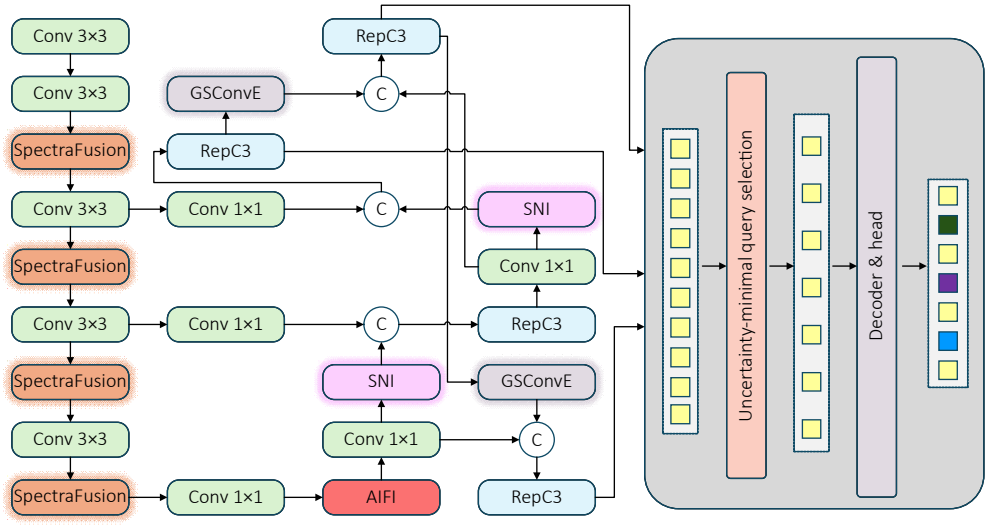


图 1 SRL-DETR 网络结构图

Fig. 1 SRL-DETR network structure

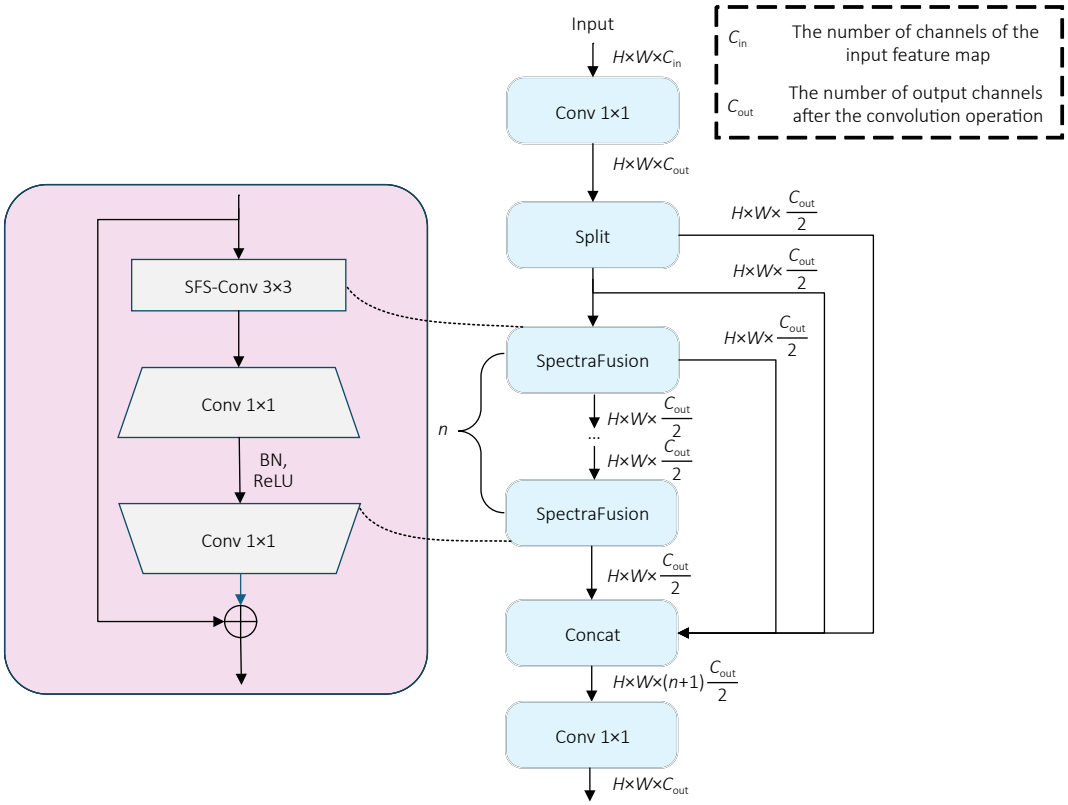


图 2 空间频率特征融合模块图

Fig. 2 Spectral-feature fusion network module

首先将输入特征在通道维度划分为两部分，实现信息的并行建模与跨层信息保留。一部分直接通过捷径路径进行残差传递，以保持原始特征的完整性；另一部分进入主支路，进行深度特征变换和空间频率特征增强。其具体计算过程如下所示：

$$\begin{cases} \boldsymbol{X}_1 + \boldsymbol{X}_2 = Split(X) \\ \boldsymbol{X}_1 \in \mathbb{R}^{H\times W\times C_1} \\ \boldsymbol{X}_2 \in \mathbb{R}^{H\times W\times C_2} \\ C_1 + C_2 = C \end{cases}, \tag{1}$$

式中：$\boldsymbol{X}_1$进入空间频率特征融合模块进行深度建模；$\boldsymbol{X}_2$保留作为跨阶段残差路径；$\boldsymbol{X}$表示模块的输入特征图；H和W分别表示特征图高度和宽度，C为特征图通道数；C_1和C_2分别表示特征划分后两部分特征图的通道数.

SFSConv 空间频率混合部分是模块的核心所在，利用空间感知单元 (spatial perception unit, SPU) 与频率感知单元 (frequency perception unit, FPU) 分别提取空间上下文关系与频率域纹理特征，并通过通道选择单元 (channel selection unit, CSU) 完成多分支特征融合。具体而言，SFSConv 由 SPU 与 FPU 组成，分别提取空间细化特征$\boldsymbol{Y}^{\mathrm{s}}$与频率细化特征$\boldsymbol{Y}^{\mathrm{f}}$。随后，通过 CSU 对两个分支输出进行融合，动态保留最具判别力的特征响应，最终得到融合输出特征 Y，SFSConv 的结构如图 3 所示。

在 SFSConv 中，输入特征$\boldsymbol{X}_1$再次划分为空间支路与频率支路，比例为α，$\boldsymbol{X}^{\mathrm{s}} \in \mathbb{R}^{(1-\alpha)C_1\times H\times W}$，$\boldsymbol{X}^{\mathrm{f}} \in \mathbb{R}^{\alpha C_1\times H\times W}$，以便各自专注于空间域与频率域的特征提取与建模。接下来将分别介绍两个支路的具体工作过程。

空间感知单元旨在通过动态感受野实现上下文建模。输入$\boldsymbol{X}^{\mathrm{s}}$被分为 g 个通道组，其中$g \in \{1,2,\cdots,n\}$，每组采用不同大小的卷积核构建多尺度特征感知，并通过逐组残差连接扩大感受野，最后将所有尺度特征拼接后通过 1×1 卷积整合，得到空间特征Y^{s}：

$$\boldsymbol{Y}_g^{\mathrm{s}} = \begin{cases} \boldsymbol{X}_g^{\mathrm{s}} * K_g, & g = 1 \\ \left(\boldsymbol{X}_g^{\mathrm{s}} + \boldsymbol{Y}_{g-1}^{\mathrm{s}}\right) * K_g, & 1 < g \leqslant n \end{cases}, \tag{2}$$

式中：对于每个X_g^{s}，存在一个大小为k_g的卷积核，总记为K_g。

频率感知单元使用分数阶 Gabor 变换 (fractional Gabor transform, FrGT) 引导的卷积核提取不同尺度与方向的频率纹理响应，实现对高频细节、边缘及纹理的多角度建模。对于输入的特征$\boldsymbol{X}^{\mathrm{f}}$，其卷积过程可表示为

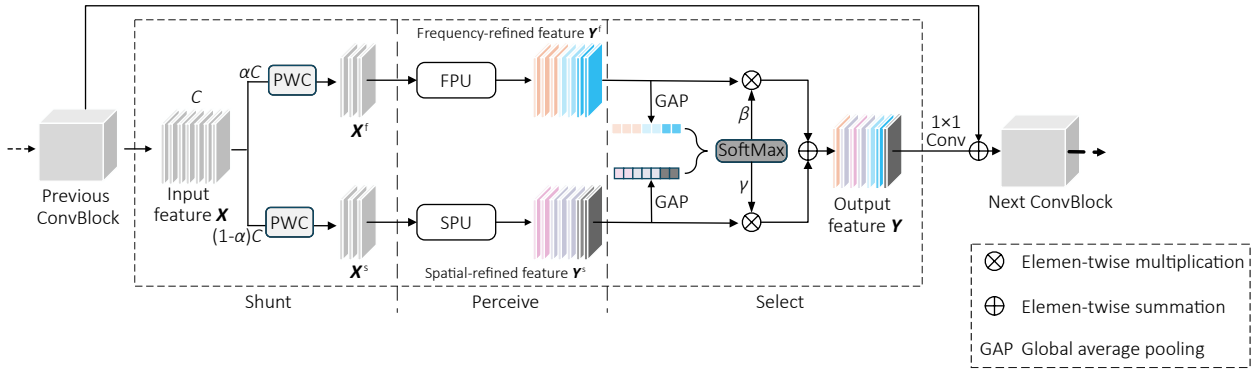


图 3 SFSConv 结构图

Fig. 3 Space-frequency selection convolution structure

$$K_{i,u}^{v} = K_{i,o} * G(u,v)\ , \tag{3}$$

$$Y^{\mathrm{f}v} = X^{\mathrm{f}v} * K_{i,C/V}^{v}\ , \tag{4}$$

式中：$v \in \{0,1,\cdots,V-1\}$；$K_{i,o}$是一个已学习的、具有 i 个输入通道和 o 个输出通道的 $k\times k$ 卷积核；$G(u,v)$代表一组具有不同方向和尺度的 FrGT 滤波器；$\boldsymbol{K}_{i,C/V}^{v}$为经分数阶 Gabor 变换作用后的卷积核，能够在不同方向 v 上提取对应的频率特征。最后，将所有方向的特征拼接形成频率增强特征$\boldsymbol{Y}^{\mathrm{f}} = [\boldsymbol{Y}^{\mathrm{f}_0}, \boldsymbol{Y}^{\mathrm{f}_1}, ..., \boldsymbol{Y}^{\mathrm{f}_{V-1}}]$。

接着，在通道选择单元对$\boldsymbol{Y}^{\mathrm{s}}$和$Y^{\mathrm{f}}$分别做通道平均池化，得到统计特征$S^{\mathrm{s}}$与$S^{\mathrm{f}}$，再用 Softmax 获得融合权重：

$$\gamma = \frac{\mathrm{e}^{S^{\mathrm{s}}}}{\mathrm{e}^{S^{\mathrm{s}}} + \mathrm{e}^{S^{\mathrm{f}}}},$$

$$\beta = \frac{\mathrm{e}^{S^{\mathrm{f}}}}{\mathrm{e}^{S^{\mathrm{s}}} + \mathrm{e}^{S^{\mathrm{f}}}}\ . \tag{5}$$

最终输出$\boldsymbol{Y}_1 = \gamma \boldsymbol{Y}^{\mathrm{s}} + \beta Y^{\mathrm{f}}$。

前两个阶段完成后，进入最终的特征重构与融合阶段。将深度建模输出$\boldsymbol{Y}_1$与绕过路径$\boldsymbol{X}_2$在通道维度拼接，实现原始信息与增强语义的融合，该过程既整合了局部增强与全局保留特征，又通过 1×1 卷积统一通道维度，如下所示：

$$\boldsymbol{Y} = Conv_{1\times1}(Concat(Y_1, \boldsymbol{X}_2))\ . \tag{6}$$

通过提出的空间频率特征融合模块，有效克服了传统 ResNet-18 在空间信息建模和多尺度特征融合方面的局限，从而提升了网络对小目标与复杂背景的感知能力，实现了更优的检测精度与计算效率平衡。

2.2 特征缩放精细增强模块 (REFS)

在原始 RT-DETR 结构中，其颈部网络主要依赖于最近邻上采样与常规卷积实现多尺度特征融合。这种设计虽然计算效率较高，但存在明显问题：1) 最近邻插值无法捕获连续空间中的上下文变化，导致高层语义在上采样过程中产生失真。2) 常规下采样结构在多尺度融合中缺乏对纹理细节和上下文多样感受野的建模能力，进而影响小目标或边缘目标的检测性能。

因此，本文提出特征缩放精细增强模块来重构原始的颈部网络，以解决上述问题，其结构如图 4 所示。该模块主要通过引入软最近邻插值 SNI[18] 和增强卷积 GSConvE 来改进特征重建与多尺度建模能力。具体来说，其两大核心为：SNI 用于上采样阶段的对齐软化与 GSConvE 用于下采样阶段的多感受野选择性增强。

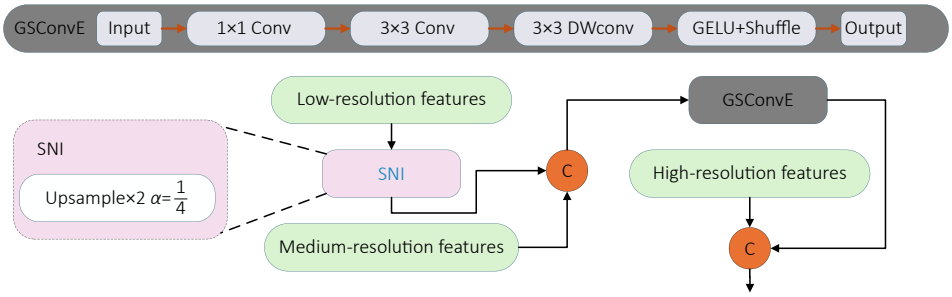


图 4 特征缩放精细增强模块图

Fig. 4 Refined enhancement for feature scaling module

SNI 的主要作用是改进上采样过程，通过软插值策略在空间域中实现语义一致且尺度自适应的特征映射，从而避免传统最近邻插值带来的信息丢失和错位问题。对于输入的低分辨率特征图 $\boldsymbol{X}$，该模块在保持最近邻插值计算效率的同时，引入与放大倍数相关的调节因子，对上采样特征的幅值进行约束，使语义特征在不同尺度间保持合理平衡。其计算过程如下：

$$\boldsymbol{X}_{\text{up}} = \alpha \cdot Upsample_{\text{nearst}}(X),\ \alpha = \frac{1}{s^2}, \tag{7}$$

式中：s 是上采样倍数；α是插值权重，用于控制高层语义特征在放大过程中的幅值衰减。

通过这种方式，SNI 能在上采样阶段对高层语义特征进行适度约束，使其在与低层特征融合时实现语义信息的一致性，增强融合特征的稳定性和可靠性。

在下采样阶段，为进一步增强颈部网络的特征表达能力，本文引入改进型轻量卷积结构 GSConvE，其能够在单个卷积块中实现多感受野的建模。与传统卷积层相比，GSConvE 在结构上融合了常规卷积、深度可分离卷积和通道重排 (shuffle) 技术，以在保持轻量化计算的同时实现更加丰富的纹理特征和多尺度的特征学习。

首先，输入特征 $\boldsymbol{X}$ 通过 1×1 卷积进行通道压缩与初步特征提取，在保持主要语义信息的同时降低通道维度，减少后续卷积计算量，计算过程如下：

$$\boldsymbol{X}_1 = Conv_{1\times1}(\boldsymbol{X})\ . \tag{8}$$

接着，压缩后的特征$\boldsymbol{X}_1$被输入至增强路径，以实现多感受野特征建模。在该路径中，特征依次经过两个卷积操作：一个普通的 3×3 卷积和一个深度可分离卷积。通过采用通道独立卷积的设计，在降低计算复杂度的同时增强了对局部结构细节的建模能力，其计算过程为

$$\boldsymbol{X}_2 = DWConv_{3\times3}(\boldsymbol{X}_1)\ . \tag{9}$$

最后，将两部分输出进行拼接，并通过通道重排操作增强通道间的信息交互：

$$\boldsymbol{Y} = Concat(\boldsymbol{X}_1, \boldsymbol{X}_2)\ , \tag{10}$$

$$\boldsymbol{Y}_{\text{shuffled}} = ReshapePermute(\boldsymbol{Y})\ . \tag{11}$$

通过这些步骤，GSConvE 能够在下采样过程中有效缓解特征表达单一与感受野不足的问题，为颈部网络提供更丰富且稳健的多尺度特征。

综上所述，特征缩放精细增强模块有效解决了原始颈部网络在上下采样过程中的信息损失与感受野不足问题，实现了更精细的特征重建与更充分的多尺度信息融合。

2.3 模型剪枝 (LAMP)

在实际部署中，行人检测模型往往受到计算资源和响应速度的限制。为提升推理效率，本文引入一种基于层自适应的幅值剪枝方法——LAMP[19] (layer-adaptive magnitude-based pruning)，以进一步降低网络冗余并提升推理效率。传统的全局幅值剪枝方法通常依据权重的绝对值大小统一设定阈值进行裁剪，但由于各层权值分布差异较大，固定阈值容易造成部分层被过度剪枝，而其它层保留冗余参数，从而影响整体性能。LAMP 的核心思想是在幅值剪枝的基础上，通过对层内权重幅度进行归一化计算，自适应地为不同层分配合适的剪枝比例，从而在全局范围内实现更平衡的稀疏性分布。LAMP 的权重评分公式如下：

$$score(u;\boldsymbol{W}) = \frac{(\boldsymbol{W}[u])^2}{\sum\limits_{v\geq u}(\boldsymbol{W}[v])^2}\ , \tag{12}$$

式中：$\boldsymbol{W}$ 表示某一层展平并按权值幅度升序排列后的权重向量；u 为当前权重索引。分子表示该权重的幅值大小，分母为该层中剩余权重幅值的平方和。该公式反映了单个连接在层内总体权重分布中的相对重要性。通过计算每一层的得分并在全模型范围内排序后，剪除得分较低的权重即可实现自适应剪枝。

剪枝操作会使网络输出发生一定变化，而这种变化主要取决于被删除权重的大小。较大的权重在特征提取中起关键作用，若被剪除，可能导致性能明显下降，而幅值较小的权重对结果的影响较弱，剪除后对整体性能的影响有限。基于此原理，LAMP 通过归一化处理各层的权重幅值，综合衡量不同层内各连接的重要性，使模型能够在全局剪枝时自动平衡不同层的稀疏度，从而在减少计算量的同时保持检测性能稳定。LAMP 的剪枝方式如图 5 所示。

综上，LAMP 以减小剪枝后模型输出误差为目标，通过层内能量归一化评估权重的重要性，使模型在全局剪枝时能够自动确定各层的稀疏比例。该方法不需要额外的超参数调节，计算复杂度与传统幅值剪枝相当，但能更合理地控制各层的剪枝强度，从而在有效降低参数量与计算量的同时保持模型性能稳定。

3 实验设计与分析

3.1 实验环境及参数配置

本实验所用的操作系统为 Windows 11，CPU 为

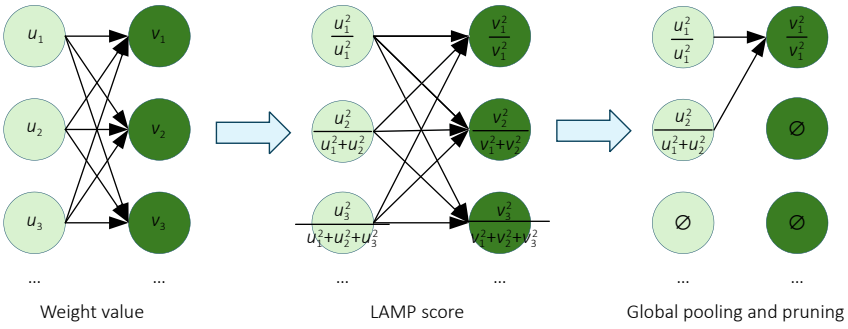


图 5 LAMP 剪枝流程图

Fig. 5 Pruning procedure diagram of LAMP

AMD Ryzen 55600 (3.50 GHz)，GPU 为 NVIDIA RTX 4060 Ti，显存大小为 8 GB，编程语言为 Python 3.10，深度学习框架采用 PyTorch 2.0.0，CUDA 版本为 11.8。实验过程中，输入图像统一调整为 640 pixel×640 pixel，训练轮次设定为 200，批次大小为 4，使用 AdamW 优化器对模型参数进行自适应调整，初始学习率设定为 1×10^{-4}。

3.2 实验数据集

CityPersons[20] 数据集是 Cityscapes 数据集的一个子集，包含超过 5000 张带有精确标注的高分辨率图像。数据集中的行人样本涵盖了各种姿势、尺度和遮挡情况，且拍摄场景包括复杂的城市街道、不同天气和光照条件。其中包含 2975 张训练集，500 张和 1575 张图片用于验证集和测试集，图像标签标注为“pedestrian”。

TinyPerson[21] 数据集是专为极小目标检测任务构建的数据集，其行人目标尺寸小、遮挡严重、分布密集，具有极高挑战性。该数据集总共包含 1610 张图片，其中 794 张图片作为训练集，816 张图片作为测试集，且具有两个类别: “sea people”和“earth people”。

WiderPerson[22] 是一个覆盖广泛的真实场景行人检测数据集，包含 13382 张图像，其中训练集有 8000 张，验证集有 1000 张，测试集有 4382 张，涵盖行人、骑车人、人群等 5 个类别，本文将所有类别统一归类为“person”类别。

3.3 实验评估指标

为全面评估所提出行人检测算法的性能，本文选取精确率 (precision)、召回率 (recall)、平均精度均值 (mAP)、参数量 (params)、计算量 (FLOPs) 以及帧率 (FPS) 作为主要性能评估指标。其中，Precision、Recall 和mAP的具体计算方式如下所示：

$$Precision = \frac{TP}{TP+FP}, \tag{13}$$

$$Recall = \frac{TP}{TP+FN}, \tag{14}$$

$$mAP = \frac{1}{N}\sum_{i=0}^{N} AP_i, \tag{15}$$

式中：TP为预测结果与真实标注匹配正确的检测数量；FP为预测结果与真实标注不匹配的错误检测数量；FN为未被检测出的真实标注数量；mAP是目标检测任务中常用的综合性能评价指标，用于衡量模型在所有类别上的检测准确性和稳定性。

3.4 剪枝结果分析

为减少模型的计算量和参数量，并提升推理速度，本文采用 LAMP 剪枝方法优化改进后的模型。通过设置不同的加速比 Speed-up (Speed-up=MACs (base)/MACs (pruned)) 来控制剪枝强度，从而比较不同压缩程度下模型的参数规模、计算复杂度与检测精度的变化。不同剪枝程度对应的实验结果如表 1 所示。

当 Speed-up 为 1.25 时，模型参数量降至 9.8 M，计算量降至 33 GFLOPs，推理速度得到提高，同时 mAP50 提升 0.4%，表明该剪枝比例在加快推理速度的同时，精度有所提升。随着 Speed-up 增至 1.5，模型参数量与计算量进一步降低，但 mAP50 从 71.2% 降至 70.9%，精确率与召回率均出现下降，说明模型的检测能力已接近稳定精度的临界点。当 Speed-up 进一步增至 1.75、2.0 与 2.25 时，模型的 mAP50 从 70.1% 持续降至 59.1%，召回率由 59.2% 下降至 49.8%，表明过度剪枝显著削弱了模型表征能力，精度损失已超出实际应用可接受范围。综合考虑压缩率与检测精度，本研究选择 Speed-up 为 1.25 作为最优方案，在保证模型轻量化的同时，实现了检测精度的最优表现。图 6 展示了

表 1 不同 Speed-up 剪枝实验结果

Table 1 Pruning experiment results corresponding to different acceleration ratios

Speed-up	Precision/%	Recall/%	mAP50/%	mAP50-95/%	Params/M	FLOPs/G	FPS
1.0	81.9	61.3	71.2	45.2	12.8	43.4	77.0
1.25	**83.8**	**61.5**	**71.6**	**46.0**	**9.8**	**33.0**	**85.3**
1.5	81.6	60.8	70.9	45.2	8.8	26.7	91.3
1.75	81.5	59.2	70.1	44.0	8.2	23.5	96.7
2.0	80.4	57.4	67.8	41.5	7.8	20.9	102.4
2.25	74.7	49.8	59.1	34.2	7.2	18.2	95.8

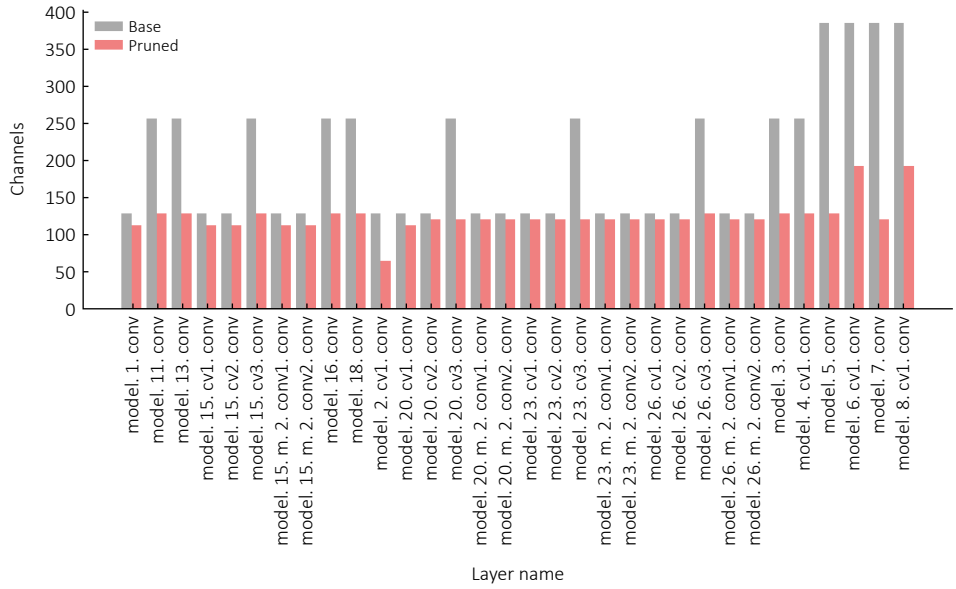


图 6 剪枝前后通道对比图

Fig. 6 Comparison of channels before and after pruning

Speed-up 为 1.25 时剪枝前后各关键层通道数的对比，横坐标为通道数发生变化的网络层名称，纵坐标表示对应层的输出通道数。

3.5 消融实验

为深入探究各改进模块对模型整体性能的影响，本文在保持训练环境与超参数一致的条件下，以 RT-DETR-R18 作为基准模型，逐步引入 SpectraFusion 模块、REFS 模块以及 LAMP 剪枝方法 (Speed-up=1.25)。通过不同模块组合的实验对比，可进一步分析各模块在提升检测精度与模型效率方面的作用，实验结果如表 2 所示。

组 A 采用基准模型，用于后续优化结果的对比分析。

组 B 加入 SpectraFusion 模块后，精确率、召回率和 mAP50 分别提升 1.0%、1.5% 和 1.1%，同时参数量与计算量分别下降至 13.1 M 与 44.0 GFLOPs。说明 SpectraFusion 模块通过在主干网络中融合空间域与频率域特征，有效增强了特征表达能力并减少了冗余计算，从而在精度和效率上均取得提升。

组 C 引入 REFS 模块，mAP50 略升至 69.9%，参数量与计算量分别下降 0.3 M、0.6 GFLOPs。该模块通过优化特征的多尺度建模和表达能力，使模型能够更有效地捕捉关键特征信息。表明 REFS 模块能够进一步增强模型的判别能力和鲁棒性。

组 D 同时引入 SpectraFusion 模块和 REFS 模块后，精确率、召回率、mAP50 分别提高 2.5%、1.7%、1.6%，参数量和计算量降至 12.8 M 和 43.4 GFLOPs。两者的协同效应表明，SpectraFusion 模块提供丰富的空间和频率特征，REFS 模块优化特征融合和多尺度建模，使网络能够更全面地捕捉关键信息，从而提升检测精度。同时，

表 2 消融实验结果

Table 2 Ablation experiment results

Group	SpectraFusion	REFS	LAMP	*P*/%	*R*/%	mAP50/%	Params/M	FLOPs/G	FPS
A				79.4	59.6	69.6	19.8	57.0	88.7
B	√			80.4	61.1	70.7	13.1	44.0	77.8
C		√		81.7	59.1	69.9	19.5	56.4	87.5
D	√	√		81.9	61.3	71.2	12.8	43.4	77.0
E	√		√	82.3	61.0	71.1	9.9	33.0	86.6
F		√	√	82.4	59.6	70.6	10.6	44.7	**95.8**
G	√	√	√	**83.8**	**61.5**	**71.6**	**9.8**	**33.0**	85.3

通过特征整合与通道优化，两模块在增强表达能力的同时避免了冗余计算，使模型参数量和计算量略有下降。

组 E 与组 F 分别在单独引入各自模块的基础上，进一步采用 LAMP 剪枝。精确率和 mAP50 在组 E 中分别提高 2.9% 和 1.5%，在组 F 中分别提高 3.0% 和 1.0%。参数量分别降至 9.9 M 和 10.6 M，计算量分别降至 33.0 GFLOPs 和 44.7 GFLOPs。此外，组 F 的 FPS 提升至 95.8 帧。这表明，LAMP 剪枝通过稀疏化网络权重，去除冗余计算和无关特征通道，从而在提升精度的同时有效降低了计算开销。

组 G 加入 SpectraFusion 模块、REFS 模块并结合 LAMP 剪枝后，精确率达到 83.8%，mAP50 提升至 71.6%，模型参数量下降 10 M，计算量下降 24 GFLOPs。SpectraFusion 模块和 REFS 模块提升了特征融合与提取能力，使得模型能够更精准地识别目标。并且 LAMP 剪枝在减少计算量和参数量的同时，保持了精度的提升。尽管剪枝可能在一定程度上导致细节损失，但由于特征提取模块的优化，模型能够在保持推理效率的同时，获得更高的精度。因此，模块的结合不仅优化了计算效率，还增强了模型的检测性能，达到精度和效率的最佳平衡。

3.6 对比实验

为验证本文算法的检测性能，将其与 Faster-RCNN、SSD、YOLO 系列算法、最新目标检测算法 DEIM 和 RT-DETR 系列模型及其他行人检测方法进行对比。在统一的实验环境下，于 CityPersons 数据集进行对比实验，结果如表 3 所示。

由于传统检测方法表现受限，Faster-RCNN 的 mAP50 仅为 44.3%，且参数量和计算量分别达到 137 M 和 207.1 GFLOPs，模型规模庞大、推理成本高，难

表 3 对比实验结果

Table 3 Comparative experimental results

Model	Precision/%	Recall/%	mAP50/%	mAP50-95/%	Params/M	FLOPs/G	FPS
Faster-RCNN	37.0	52.5	44.3	19.5	137.0	207.1	20.0
SSD	**91.0**	13.7	31.4	11.3	26.2	43.7	45.0
YOLOv9m	83.4	54.3	63.9	40.7	16.7	60.5	153.1
YOLOv10m	82.6	55.3	66.2	42.0	15.3	58.9	**177.1**
YOLOv11m	80.2	56.6	64.9	41.7	20.3	67.6	147.0
YOLOv12m	79.9	56.0	64.4	41.2	19.5	59.5	120.6
Ref.1[23]	80.5	58.4	68.5	43.5	**6.5**	–	–
Ref.2[24]	76.9	56.8	64.8	39.4	11.8	34.0	92.0
Ref.3[25]	82.7	59.2	68.9	43.9	10.4	–	–
Ref.4[26]	81.3	60.5	71.4	43.2	16.0	45.0	62.0
DEIM[27]	79.8	59.5	70.1	42.8	51.3	104.6	102.0
RT-DETR-R18	79.4	59.6	69.6	43.4	19.8	57.0	88.7
RT-DETR-R34	83.8	59.7	70.9	45.3	31.3	88.8	72.8
SRL-DETR	83.8	**61.5**	**71.6**	**46.0**	9.8	**33.0**	85.3

以满足实际部署需求；SSD 虽精确率达 91.0%，但召回率只有 13.7%，漏检问题突出，难以适应复杂场景。YOLOv9m、YOLOv10m、YOLOv11m 与 YOLOv12m 的 mAP50 分别为 63.9%、66.2%、64.9% 和 64.4%，精确率最高为 83.4%，召回率最高为 56.6%。相比 YOLO 系列，本文算法在精度和召回率上均实现提升，显示出更强的密集小目标检测能力。DEIM 虽在 mAP50 上略高于 YOLO 系列，但其参数量和计算量较大；文献 [1]、文献 [2] 和文献 [3] 方法的检测精度也相对较低；文献 [4] 虽在精度上有所提升，但计算复杂度仍然较高。相比之下，本文改进算法 mAP50 达到 71.6%，较基准算法提升 2.0%，mAP50-95、精确率和召回率分别提升至 46.0%、83.8% 和 61.5%，显著改善了漏检和检测不稳定的问题。同时，模型参数量减少至 9.8 M，计算量下降至 33.0 GFLOPs，推理速度仍保持在实时水平，表明本文算法在保持轻量化设计的同时，实现了检测性能与计算效率的兼顾。

为直观比较各模型在多指标下的综合性能，绘制模型性能雷达图，如图 7 所示。该图包含精确率、召回率、mAP50、mAP50-95、Params、FLOPs 及 FPS 等关键指标，能够全面反映模型在精度与效率维度的整体表现。从图中可以看出，Faster-RCNN 参数量与计算量较高，整体模型较为庞大，导致精度指标覆盖范围收缩；SSD 虽在精确率上表现较好，但召回率与 mAP 指标不足，存在明显漏检倾向。YOLOv10m 在速度指标上具有优势，但精度较低。相较之下，SRL-DETR 在精确率、召回率及 mAP50 上均呈现较大的覆盖半径，且保持较为适宜的参数规模和 FPS，体现出较好的精度、效率平衡能力，综合性能优于对比模型。

3.7 可视化分析

行人检测在不同场景下面临光照、背景与目标尺度变化等多重挑战。为分析模型在真实环境中的表现，本文选取三类典型城市场景，对比 RT-DETR-R18 与 SRL-DETR 的检测结果及对应热力图分布，如图 8 所示。

在小目标行人场景中，RT-DETR-R18 出现误检。图中黄色标注区域可见，该模型将路旁树干识别为行人，而 SRL-DETR 未产生此类误检，能够正确排除与行人形态相近的背景结构。同时，SRL-DETR 的注意力分布也更集中于行人主体区域。在白天场景中，SRL-DETR 的

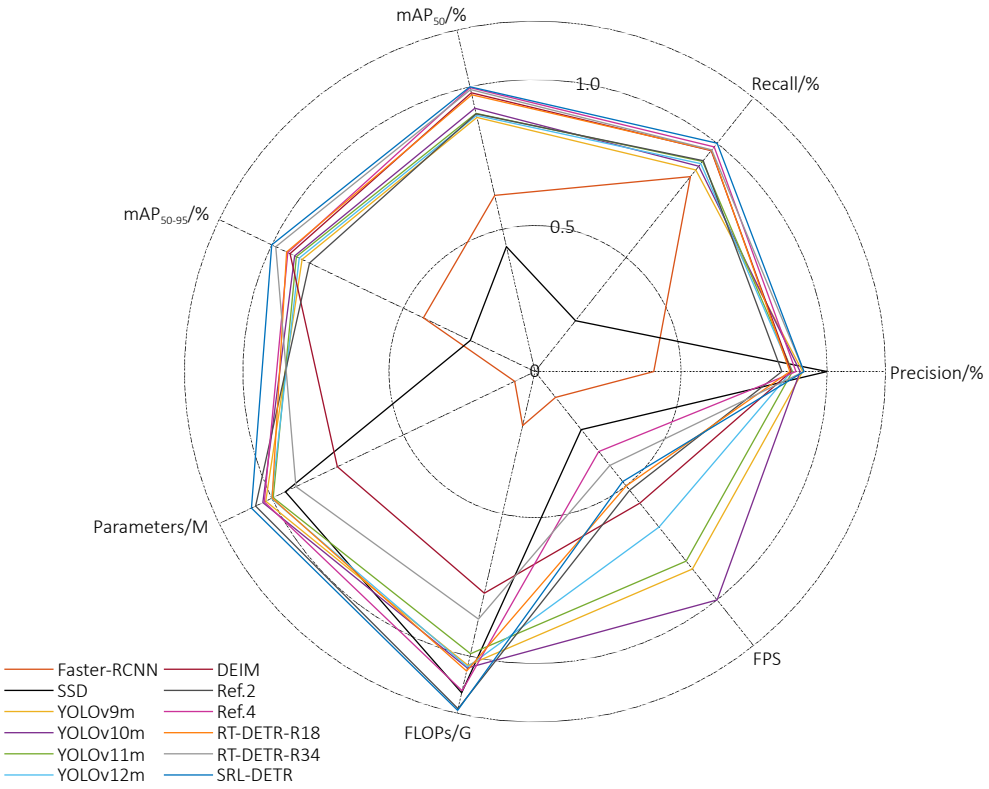


图 7 各模型性能指标雷达图

Fig. 7 Radar chart of performance indicators for various models

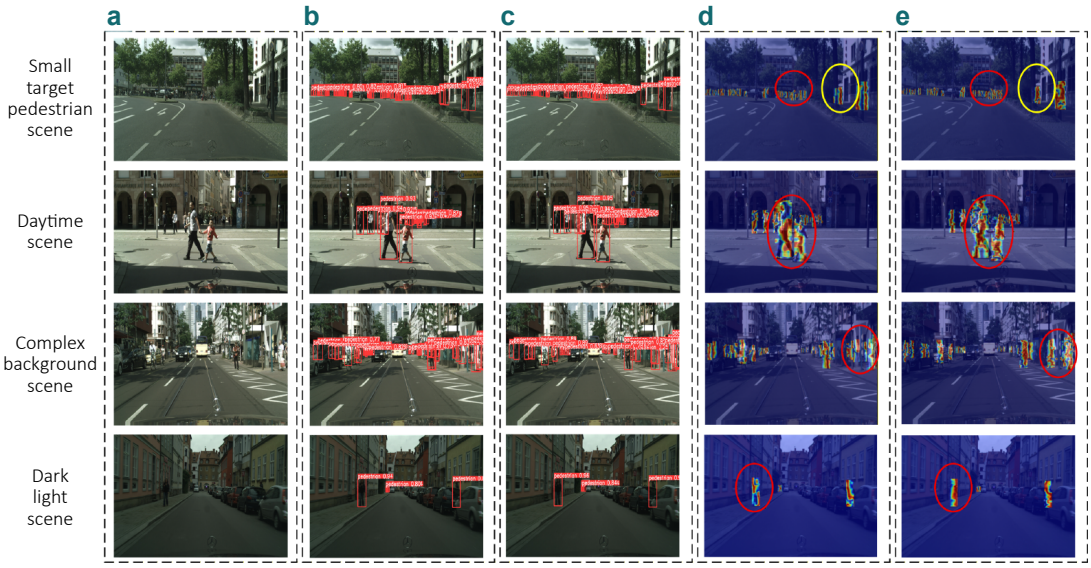


图 8 CityPersons 数据集检测结果可视化。(a) 原图；(b) RT-DETR-R18 检测图；(c) SRL-DETR 检测图；(d) RT-DETR-R18 热力图；(e) SRL-DETR 热力图

Fig. 8 Visualization of detection results on the CityPersons dataset. (a) Original image; (b) RT-DETR-R18 detection image; (c) SRL-DETR detection image; (d) RT-DETR-R18 thermal map; (e) SRL-DETR thermal map

检测框与行人轮廓贴合度更高，定位更精准；其热力图主要覆盖行人本身，而不会延展至多余背景区域。相比之下，RT-DETR-R18 的热力图虽能完整覆盖行人，但热力图较为分散。在背景复杂场景中，RT-DETR-R18 的检测结果出现误检与冗余框，而 SRL-DETR 的检测结果保持单一且稳定。在暗光场景中，SRL-DETR 仍能可靠检测行人，定位偏差较小，热力图对行人区域的响应更集中，显示出较好的低照度适应性。

可视化结果显示，改进后的模型在多类场景中均保持稳定检测表现，能够准确定位行人并有效抑制背景干扰。在小目标、复杂背景和低照度条件下，其检测结果仍具有较好的完整性与一致性，体现出良好的鲁棒性。

3.8 泛化性实验

3.8.1 TinyPerson 数据集

为进一步验证本文算法在小目标密集行人检测任务中的泛化能力，本文在 TinyPerson 数据集上与不同算法开展对比实验，综合比较各个方面的表现，结果如表 4 所示。本文算法 mAP50 与 mAP50-95 分别提升至 24.8%、9.9%，较基准算法提高了 3.7%、1.9%，对比 YOLO 系列模型提升幅度更为显著。LMFI-YOLO 虽在参数量上具备一定优势，但 mAP50 略低。并且本文算法的精确率和召回率分别为 45.8% 和 28.8%，均高于其他对比模型，表明其在提升检测准确性的同时，有效降低了漏检率。与此同时，本文算法在轻量化方面也展现出明显优势，其参数量和计算量仅为 9.9 M 和 33.2 GFLOPs。SRL-DETR 在 TinyPerson 数据集上表现出显著提升，得益于增强的特征表达能力和优化的多尺度特征融合，使模型能够更有效捕捉微小行人信息，同时保持轻量化和高稳定性。综上所述，所提方法在精度、召回率及模型效率之间取得了更优的平衡，展现出在小尺寸行人检测任务中的显著优势。图 9 展示了具有代表性算法的精度曲线，以直观呈现各方法在小目标检测任务中的性能差异。由图可见，本文算法在检测精度上相对于其他方法取得了明显提升。同时结合检测效果图进行可视化分析，以多维度评估本文算法在小目标行人检测任务中的性能和泛化能力，检测结果如图 10 所示。在场景一和场景二中，RT-DETR-R18 均出现漏检现象，同时在场景一将沙滩躺椅的椅脚误识别为人，在场景二则对同一人产生了重复检测。相比之下，SRL-DETR 在这两个场景中均实现了完整且准确的目标识别。在场景三中，基准算法将浪花误识别为行人，而 SRL-DETR 有效抑制了此类误检。综上，改进后的算法在不同场景下均表现出更强的检测鲁棒性与泛化能力，能够有效减少漏检与误检问题。

3.8.2 WiderPerson 数据集

为评估所提算法在复杂场景下的行人检测泛化

表 4 TinyPerson 数据集对比实验表

Table 4 Comparison experiment of TinyPerson dataset

Model	Precision/%	Recall/%	mAP50/%	mAP50-95/%	Parameters/M	FLOPs/G
YOLOv8m	38.4	20.4	18.8	7.4	23.2	67.4
YOLOv9m	38.3	19.2	18.0	7.0	16.7	60.5
YOLOv10m	37.2	19.5	17.9	6.9	15.3	58.9
YOLOv11m	39.9	20.9	19.9	7.8	20.0	67.7
LMFI-YOLO[28]	38.3	24.6	21.7	7.7	**1.8**	–
LPAE-YOLOv8[29]	36.3	22.1	18.7	6.5	1.97	**10.5**
PPD-YOLO[30]	39.9	26.6	22.7	8.3	–	–
STD-DETR[31]	37.2	22.1	18.5	6.1	12.2	37.1
TCF-DETR[32]	–	–	24.2	7.7	19.8	61.9
RT-DETR-R18	41.1	25.2	21.1	8.0	19.8	56.9
RT-DETR-R34	42.8	26.4	22.5	8.9	31.1	88.8
SRL-DETR	**45.8**	**28.8**	**24.8**	**9.9**	9.9	33.2

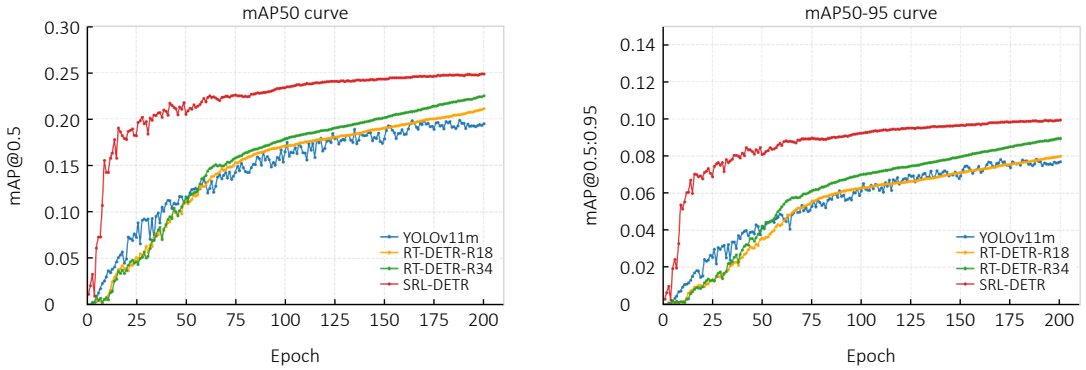


图 9 TinyPerson 数据集上 mAP 值比较

Fig. 9 Comparison of mAP values on the TinyPerson dataset

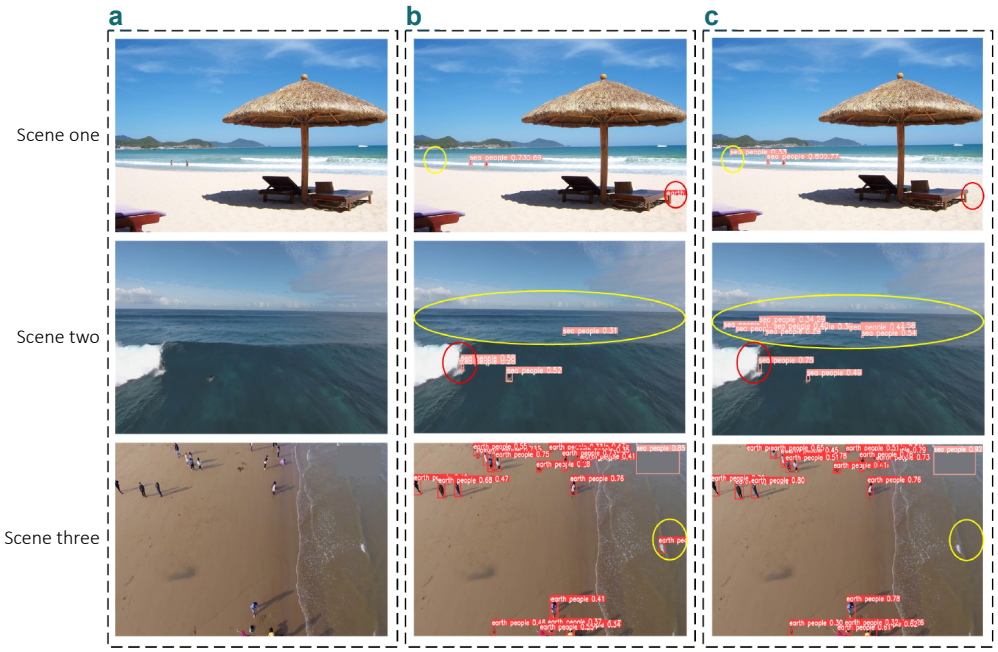


图 10 TinyPerson 数据集检测结果可视化。 (a) 原图; (b) RT-DETR-R18; (c) SRL-DETR

Fig. 10 Visualization of detection results on the TinyPerson dataset. (a) Original image; (b) RT-DETR-R18; (c) SRL-DETR

能力，本文选取 WiderPerson 数据集进行实验，并与 YOLOv8m、YOLOv10m、YOLOv11m 及其他方法进行对比，实验结果如表 5 所示。结果表明，本文算法准确率达到 88.6%，召回率为 83.8%，较基准模型分别提升 1.7% 和 1.1%；mAP50 提升至 91.5%，优于 YOLO 系列、LDD-YOLO-m 和 RT-DETR-MSS 等方法，体现出在复杂场景中更强的鲁棒性与适应性。这主要归因于改进后的模型在复杂背景、遮挡和多尺度行人场景下能更稳定地捕捉行人信息，从而提升整体检测性能。图 11 为基准模型和本文算法的检测效果对比图。从三个场景的可视化结果可以看出，SRL-DETR 在不同复杂度场景中均展现出更稳定的检测能力。在场景一与场景二中，SRL-DETR 表现出更高的检测准确率。尤其是在场景三中，RT-DETR-R18 出现了明显漏检，未能识别远处小目标，而 SRL-DETR 能够成功检测到该行人。可见，本文算法在复杂背景和不同尺寸目标的场景中表现更稳定可靠。

4 结　论

本文针对复杂场景下的密集及小目标行人检测任务，提出了一系列优化方法。首先，提出空间频率特征融合模块，以增强主干网络在空间与频率维度上的特征表达与建模能力。其次，设计特征缩放精细增强模块，以优化多尺度特征融合过程中的上下文建模和信息传递。最后，引入 LAMP 剪枝对通道重要性进行评估与筛选，从而在保留关键特征的同时，使模型更为轻量。以上改进

表 5　WiderPerson 数据集对比实验表

Table 5　Comparison experiment of WiderPerson dataset

Model	Precision/%	Recall/%	mAP50/%	mAP50-95/%	Parameters/M	FLOPs/G
YOLOv8m	86.4	83.6	90.8	**66.8**	23.2	67.4
YOLOv10m	84.2	83.1	90.2	65.7	15.3	58.9
YOLOv11m	86.2	83.5	90.7	66.2	20.0	67.6
LDD-YOLO-m	87.0	82.9	90.9	65.7	13.2	41.4
RT-DETR-MSS[33]	87.6	83.6	91.1	64.2	18.4	–
RT-DETR-R18	86.9	82.7	90.5	63.2	19.8	56.9
SRL-DETR	**88.6**	**83.8**	**91.5**	64.4	**9.9**	**33.7**



图 11　WiderPerson 数据集检测结果可视化。(a) 原图; (b) RT-DETR-R18; (c) SRL-DETR

Fig. 11　Visualization of detection results on the WiderPerson dataset. (a) Original image; (b) RT-DETR-R18; (c) SRL-DETR

有效增强了模型对密集和小尺度行人的感知能力，同时降低了计算冗余。

在 CityPersons、TinyPerson 和 WiderPerson 三个数据集上的实验结果表明，所提方法在不同数据分布和目标密度条件下均取得了稳定的精度提升。进一步结合剪枝策略，在保持检测性能的同时显著降低了模型的计算量与参数量，实现了模型的轻量化与高效推理。上述结果验证了该方法在复杂场景、密集行人及小目标行人检测任务中的有效性和良好的泛化性能。

未来工作中，将进一步探索可见光与红外光融合的行人检测方法，以充分利用多模态信息提升模型在低光照、夜间及复杂环境下的检测鲁棒性。通过结合可见光的纹理信息和红外光的热特征，有望进一步增强模型对密集和小尺度目标的感知能力，同时保持高效推理与轻量化设计，为实际应用提供更加稳健的检测方案。

作者贡献

张蕊：指导论文选题、论文修改；李怡欣：研究资料的收集、分析，论文写作；高张：实验数据集查询、采集；张上：资料调研；朱帅：数据整理。

利益冲突

所有作者声明无利益冲突

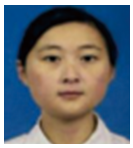
参考文献

1. Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//*Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014: 580–587. https://doi.org/10.1109/CVPR.2014.81.
2. Girshick R. Fast R-CNN[C]//*Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2015: 1440–1448. https://doi.org/10.1109/ICCV.2015.169.
3. Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE Trans Pattern Anal Mach Intell*, 2017, **39**(6): 1137−1149.
4. Cai Z W, Vasconcelos N. Cascade R-CNN: delving into high quality object detection[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 6154–6162. https://doi.org/10.1109/CVPR.2018.00644.
5. Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 779–788. https://doi.org/10.1109/CVPR.2016.91.
6. Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[C]//*Proceedings of the 14th European Conference on Computer Vision*, 2016: 21–37. https://doi.org/10.1007/978-3-319-46448-0_2.
7. Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers[C]//*16th European Conference on Computer Vision*, 2020: 213–229. https://doi.org/10.1007/978-3-030-58452-8_13.
8. Zhao Y A, Lv W Y, Xu S L, et al. DETRs beat YOLOs on real-time object detection[C]//*Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024: 16965–16974. https://doi.org/10.1109/CVPR52733.2024.01605.
9. Chen H X, Chen Z A, Ning X, et al. Improved YOLOv10n dense pedestrian small target detection method[J]. *J Chin Comput Syst*, 2026, **47**(4): 999−1007.
陈海 , 陈子昂, 宁馨, 王思源, 陆亚舒, 卢海涛. 改进 YOLOv10n 的密集行人小目标检测方法[J]. 小型微型计算机系统, 2026, **47**(4): 999−1007.
10. Huang D Y, Huang D Q, Huang H F, et al. MFE-YOLO: multi-scale feature enhancement algorithm for pedestrian detection in complex scenes[J]. *Comput Eng Appl*, 2026, **62**(1): 124−139.
黄德意, 黄德启, 黄海峰, 等. MFE-YOLO: 复杂场景下多尺度特征增强行人检测算法[J]. 计算机工程与应用, 2026, **62**(1): 124−139.
11. Yang D, Zhang X L, Wang P, et al. LDD-YOLO: improved lightweight dense pedestrian detection algorithm based on YOLOv8[J]. *J Front Comput Sci Technol*, 2026, **20**(1): 251−265.
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.
12. Fan R, Ma X L. Improved DETR for crowded pedestrian detection[J]. *Comput Eng Appl*, 2023, **59**(19): 159−165.
樊嵘, 马小陆. 面向拥挤行人检测的改进 DETR 算法[J]. 计算机工程与应用, 2023, **59**(19): 159−165.
13. Song T Z, Cao C J, He J Q, et al. Research on dense pedestrian detection algorithm based on improved DETR[J]. *Comput Eng*, 2024.
宋天泽, 曹从军, 何佳琪, 等. 基于改进 DETR 的密集行人检测算法研究[J]. 计算机工程, 2024.
14. Li Q Y, Wei J. Improvement of the dense pedestrian detection algorithm of RT-DETR[J]. *Electron Meas Technol*, 2025, **48**(21): 148−156.
李青云, 魏佳. 改进 RT-DETR 的密集行人检测算法[J]. 电子测量技术, 2025, **48**(21): 148−156.
15. Wang D, Jia L. Improved pedestrian and vehicle detection algorithm based on RT-DETR[J]. *Comput Eng Des*, 2025, **46**(12): 3620−3627.
王栋, 贾磊. 基于 RT-DETR 改进的行人车辆检测算法[J]. 计算机工程与设计, 2025, **46**(12): 3620−3627.
16. Chen J R, Kao S H, He H, et al. Run, don't walk: chasing higher FLOPS for faster neural networks[C]//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023: 12021–12031. https://doi.org/10.1109/CVPR52729.2023.01157.
17. Li K, Wang D, Hu Z Y, et al. Unleashing channel potential: space-frequency selection convolution for SAR object detection[C]//*IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024: 17323–17332. https://doi.org/10.1109/CVPR52733.2024.01640.
18. Li H L. Rethinking features-fused-pyramid-neck for object detection[C]//*18th European Conference on Computer Vision (ECCV)*, 2024: 74–90. https://doi.org/10.1007/978-3-031-72855-6_5.
19. Lee J, Park S, Mo S, et al. Layer-adaptive sparsity for the magnitude-based pruning[C]//*9th International Conference on Learning Representations*, 2021.
20. Zhang S S, Benenson R, Schiele B. CityPersons: a diverse dataset for pedestrian detection[C]//*IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017: 4457–4465. https://doi.org/10.1109/CVPR.2017.474.
21. Yu X H, Gong Y Q, Jiang N, et al. Scale match for tiny person

detection[C]//*2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2020: 1246–1254. https://doi.org/10.1109/WACV45572.2020.9093394.

22. Zhang S F, Xie Y L, Wan J, et al. WiderPerson: a diverse dataset for dense pedestrian detection in the wild[J]. *IEEE Trans Multimedia*, 2020, **22**(2): 380−393.
23. Tang J W, Lai H C, Wang T G. Improved YOLOv8 pedestrian detection algorithm for long-distance situations[J]. *Comput Eng*, 2025, **51**(4): 303−313.
汤静雯, 赖惠成, 王同官. 远距离情形下的改进 YOLOv8 行人检测算法[J]. 计算机工程, 2025, **51**(4): 303−313.
24. Li Y Y, Hu J, Wen Y, et al. Rethinking vision transformers for MobileNet size and speed[C]//*IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023: 16843–16854. https://doi.org/10.1109/ICCV51070.2023.01549.
25. Ma Y X, Liu B L, Liu J R. Research on pedestrian detection technology based on YOLO11s algorithm[C]//*8th International Conference on Computer Network, Electronic and Automation*, 2025: 1–7. https://doi.org/10.1109/ICCNEA66167.2025.11211939.
26. Wu T Y, Li X, Dong Q X. An improved transformer-based model for urban pedestrian detection[J]. *Int J Comput Intell Syst*, 2025, **18**(1): 68.
27. Huang S H, Lu Z C, Cun X D, et al. DEIM: DETR with improved matching for fast convergence[C]//*2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025: 15162–15171. https://doi.org/10.1109/CVPR52734.2025.01412.
28. Yuan T T, Lai H C, Tang J W, et al. LMFI-YOLO: lightweight pedestrian detection algorithm in complex scenes[J]. *Comput Eng Appl*, 2025, **61**(15): 111−123.
袁婷婷, 赖惠成, 汤静雯, 等. LMFI-YOLO: 复杂场景下的轻量化行人检测算法[J]. 计算机工程与应用, 2025, **61**(15): 111−123.
29. Zhao Y, Chen G P. LPAE-YOLOv8: lightweight aerial small object detection via LSE-Head and adaptive attention[J]. *Sci Rep*, 2025, **15**(1): 44780.
30. Li J, Tang R Q, Yi J B, et al. PPD-YOLO: a lightweight model for small target detection in extreme light and dense scenes for UAVs[J]. *Signal Image Video Process*, 2025, **19**(18): 1445.
31. Yin Z Y, Yang B, Chen J L, et al. Lightweight small object detection algorithm based on STD-DETR[J]. *Laser Optoelectron Prog*, 2025, **62**(8): 0815002.
尹泽宇, 杨波, 陈金令, 等. 基于 STD-DETR 的轻量化小目标检测算法[J]. 激光与光电子学进展, 2025, **62**(8): 0815002.
32. Lei H, Wu Z, Shang L, et al. TCF-DETR: multi-scale token-channel fusion transformer for enhanced small object detection[J]. *Multimed Syst*, 2025, **31**(4): 292.
33. Song Y L, Qian P, Zhang K X, et al. An improved Multi-Scale Fusion and Small Object Enhancement method for efficient pedestrian detection in dense scenes[J]. *Multimed Syst*, 2025, **31**(3): 151.

作者简介



张蕊 (1980-)，女，博士，副教授，研究方向为人工智能、深度学习、图像处理。

E-mail：5421791@qq.com



【通信作者】高张 (1980-)，男，硕士，高级工程师，研究方向为大数据分析处理、人工智能。

E-mail：gaozhang@ctgu.edu.cn

SRL-DETR: lightweight pedestrian detection algorithm for multiple scenarios

Zhang Rui[1,2], Li Yixin[1,2], Gao Zhang[3*], Zhang Shang[1,2], Zhu Shuai[1,2]

[1] College of Computer and Information Technology, China Three Gorges University, Yichang, Hubei 443002, China;
[2] Hubei Province Engineering Technology Research Center for Construction Quality Testing Equipment, China Three Gorges University, Yichang, Hubei 443002, China;
[3] Information Technology Center of China Three Gorges University, Yichang, Hubei 443002, China

Abstract

Objective Pedestrian detection faces challenges from small-scale targets, dense object distributions, and complex backgrounds, which frequently result in insufficient detection reliability, missed detections, and false positives. Under such conditions, variations in pedestrian scale and scene complexity place higher demands on feature perception and multi-scale modeling. Existing detection models still exhibit limited robustness under small-scale and complex scene conditions, particularly when computational efficiency and model complexity are constrained, making it difficult to achieve a balance between accuracy and efficiency. These limitations remain critical obstacles to reliable pedestrian detection in practical scenarios. The objective of this study is to improve detection reliability and robustness under small-scale and complex scene conditions while reducing computational cost and model size, thereby enhancing inference efficiency and facilitating deployment in real-world pedestrian detection applications.

Methods The proposed SRL-DETR is a lightweight pedestrian detection model developed based on the RT-DETR architecture. Firstly, a spatial-frequency feature fusion module was introduced for the backbone, integrating the SFSConv operator with a CSP structure. This design enhances the network's perception of small-scale pedestrians by capturing both spatial and frequency features and facilitating more effective feature flow, enabling the model to extract richer and more discriminative information from low-resolution targets. Secondly, a refined feature scaling enhancement module was introduced to optimize multi-scale feature fusion in the neck network. SNI was introduced during upsampling to align high-level semantic features across scales, reducing information misalignment and preserving contextual consistency. GSConvE was introduced during downsampling to capture multi-scale textures and receptive fields, improving detection of dense pedestrians in complex scenes. These enhancements allow more accurate reconstruction and fusion of multi-scale features, strengthening the model's robustness in challenging detection scenarios. Finally, LAMP was introduced to evaluate channel importance and remove less critical channels. This strategy effectively reduces model size and computation while preserving essential feature information, improving inference efficiency and making the model more suitable for deployment in resource-constrained environments. The proposed SRL-DETR model was evaluated on the CityPersons, TinyPerson, and WiderPerson datasets to validate both detection accuracy and generalization under varying pedestrian densities and scale distributions. The combination of enhanced small-target perception, refined multi-scale feature fusion, and adaptive pruning contributes to improved detection precision, robustness, and efficiency across different datasets.

Results and Discussions The proposed SRL-DETR was evaluated on multiple pedestrian detection datasets, achieving high detection accuracy while simultaneously reducing model size and computational cost. On the CityPersons dataset, SRL-DETR achieved an mAP50 of 71.6%, representing a 2.0% improvement over the baseline model, while reducing parameters and computation by 50.5% and 42.1%, respectively. These results demonstrate that the proposed model maintains strong detection performance even under substantially reduced computational complexity. Comparison experiments further show that SRL-DETR outperformed YOLO and other compared methods across multiple evaluation metrics, indicating more balanced overall detection performance. In particular, SRL-DETR exhibited more stable detection behavior, suggesting improved capability in distinguishing pedestrians from complex backgrounds and reduced performance degradation caused by scale variation and scene clutter. On the TinyPerson dataset, SRL-DETR improved mAP50 by 3.7% compared with the baseline model. This improvement is mainly attributed to enhanced sensitivity to small-scale pedestrians and more effective multi-scale feature interaction, which alleviated missed detections and false positives in dense pedestrian scenarios. The results indicate that SRL-DETR is more effective in handling large scale variations and dense pedestrian distributions, which are commonly encountered challenges in pedestrian detection tasks.Similarly, on the WiderPerson dataset, SRL-DETR achieved a 1.0% improvement in

Foundation item: Hubei Provincial National College Student Innovation and Entrepreneurship Training Program of China (S202311075047), Hubei Provincial Digital Economy Pilot Demonstration Construction Special Project (2312-420625-04-02-996363), the 2024 Talent Research Startup Fund of China Three Gorges University (2024RCKJ074), and 2024 Annual Key Teaching Research Project of China Three Gorges University (J2024018)
*Correspondence: Gao Z, E-mail: gaozhang@ctgu.edu.cn

mAP50 over the baseline, indicating consistent robustness across varying pedestrian densities and complex scene configurations. The stable performance across datasets with different data distributions further confirms the generalization ability of the proposed model under diverse detection conditions.Furthermore, the introduction of adaptive channel pruning enabled a substantial reduction in parameters and computation while preserving detection accuracy, highlighting the effectiveness of the lightweight optimization strategy. Overall, the coordinated improvements in small-scale target perception, multi-scale feature interaction, and model compression jointly contributed to improved detection accuracy, stronger robustness in dense and complex scenes, and enhanced efficiency, providing a practical and effective solution for pedestrian detection in real-world environments.

Conclusions The proposed SRL-DETR model effectively balances detection accuracy, robustness, and computational efficiency for pedestrian detection under small-scale and complex scene conditions. Experimental results on the CityPersons, TinyPerson, and WiderPerson datasets demonstrate that SRL-DETR achieves consistent improvements in both accuracy and efficiency compared with the baseline and other compared methods, maintaining stable performance across varying pedestrian densities and scale distributions. The coordinated improvements enhance the model's sensitivity to small-scale pedestrians, improve the handling of dense pedestrian distributions in complex scenes, and reduce model size and computation without compromising detection reliability. As a result, SRL-DETR maintains robust detection performance while satisfying efficiency requirements under practical constraints. In summary, SRL-DETR provides an effective pedestrian detection model that achieves a well-balanced trade-off between performance and computational efficiency, making it suitable for pedestrian detection tasks in small-scale and complex scene conditions.

Keywords: pedestrian detection; RT-DETR; small object detection; multi-scale; lightweight

Zhang R, Li Y X, Gao Z, et al. SRL-DETR: lightweight pedestrian detection algorithm for multiple scenarios[J]. *Opto-Electron Eng*, 2026, **53**(5): 250363; DOI: 10.12086/oee.2026.250363

Opto-Electronic Journals Group
www.oejournal.org

扫描二维码，获取 PDF 全文