

DOI: 10.13715/j.issn.2096-644X.20250614.0001

引用格式:孙晓纪,王文龙,李明鸿,等.基于分阶递进增强的低分辨率图像目标检测方法[J].湘潭大学学报(自然科学版),2026,48(2):10-20.

Citation: SUN Xiaoji, WANG Wenlong, LI Minghong, et al. Object detection on low-resolution images with hierarchical progressive enhancement[J]. Journal of Xiangtan University (Natural Science Edition), 2026, 48(2): 10-20.

基于分阶递进增强的低分辨率图像目标检测方法*

孙晓纪^{1,2}, 王文龙^{1,2}, 李明鸿², 王 焘², 赵于前², 桂卫华²

(1. 镍钴共生资源开发与综合利用全国重点实验室, 甘肃 金昌 737104;

2. 中南大学 自动化学院, 湖南 长沙 410083)

摘要:【目的】目标检测是计算机视觉领域的核心任务之一,在智能监控、自动驾驶等实际场景中具有重要应用价值.然而,在真实采集过程中,受成像设备性能限制、传输干扰等因素影响,图像往往会出现低分辨率(LR)退化,并伴随噪声和模糊等问题.现有多数基于卷积神经网络(CNN)的目标检测算法主要针对高分辨率(HR)图像设计,在复杂退化的LR图像上,由于HR与LR特征域差异显著,检测性能明显下降.针对上述问题,本文提出一种面向目标检测任务的超轻量级预增强网络(HPEN).【方法】为降低复杂退化带来的图像重构难度,HPEN采用分阶递进增强策略.第一阶段设计小波复原模块(WRM)恢复图像关键纹理信息,实现图像复原.WRM首先利用离散小波变换将输入图像分解为低-低频、低-高频、高-低频和高-高频4个频率分量,并在每个分量支路中引入残差单元进行特征提取.为实现低频结构信息与高频细节信息的有效融合,在相邻频率支路之间构建跨层特征融合单元,通过仿射变换方式将外部指导特征融合到当前特征中,以增强特征表达能力.第二阶段设计分辨率增强模块(REM)对复原图像进行细节重建并提升图像分辨率.REM首先利用 3×3 卷积提取浅层特征,并通过残差单元进行特征细化;随后采用双线性插值将特征缩放至固定尺寸 128×128 ,再通过像素重排层将特征上采样至 512×512 .同时,通过跳跃连接结合双线性插值方式引入残差图像,以促进模块内部信息流动.将HPEN与任意检测器级联,可构建适用于LR图像的目标检测方法(HPELOD).【结果】为验证所提方法的有效性,在降质后的PASCAL VOC2007 trainval(VOC-L)和COCO val 2017(COCO-L)数据集上,与基准检测方法CenterNet、基于预处理的方法以及联合增强与检测的方法进行对比实验.实验结果表明,在保持检测速度的前提下,HPELOD仅引入少量网络参数和计算开销,即可显著提升LR图像的目标检测性能.例如,在VOC-L和COCO-L数据集上,HPELOD相较于自监督学习分辨率自适应编码方法(AERIS)分别提升1.2和1.1的mAP.此外,在主观视觉效果方面,所提方法在较强噪声和模糊的情况下仍能够检测出更多目标.【结论】该文提出的超轻量级预增强网络HPEN及其分阶递进增强策略能够有效缓解退化LR图像的重构与增强难题,所构建的HPELOD方法为低分辨率图像目标检测提供了一种高效且轻量化的解决方案.该方法在检测精度与计算效率之间取得良好平衡,具有较高实际应用价值,可为后续低分辨率图像目标检测研究提供参考.

关键词:卷积神经网络;图像增强;目标检测;深度学习;特征融合

中图分类号: TP391.41

文献标志码: A

文章编号: 2096-644X(2026)02-0010-11

* 收稿日期: 2025-06-14

基金项目: 国家自然科学基金(U23B2063, 62573439, 62473381); 湖南省科技创新重点研发计划项目(2023GK2021, 2024JK2028)

通信作者: 赵于前(1973-), 男, 湖南长沙人, 博士, 教授. E-mail: zyuq@csu.edu.cn

Object detection on low-resolution images with hierarchical progressive enhancement

SUN Xiaoji^{1,2}, WANG Wenlong^{1,2}, LI Minghong², WANG Tao², ZHAO Yuqian², GUI Weihua²

(1. National Key Laboratory of Ni&Co Associated Minerals Resources Development and Comprehensive Utilization, Jinchang 737104, China; 2. School of Automation, Central South University, Changsha 410083, China)

Abstract: **【Objective】** Object detection is one of the core tasks in computer vision and plays an important role in practical applications such as intelligent surveillance and autonomous driving. However, in real-world acquisition scenarios, images often suffer from low-resolution (LR) degradation due to limitations of imaging devices and transmission interference, and are frequently accompanied by noise and blur. Most existing convolutional neural network (CNN)-based object detection methods are designed for high-resolution (HR) images. When applied to LR images with complex degradation, the significant feature-domain discrepancy between HR and LR images leads to notable performance degradation. To address this issue, this paper proposes an ultra-lightweight pre-enhancement network named HPEN for object detection tasks. **【Method】** To reduce the difficulty of image reconstruction caused by complex degradations, HPEN adopts a progressive enhancement strategy. In the first stage, a wavelet restoration module (WRM) is designed to recover critical texture information and achieve image restoration. Specifically, WRM employs discrete wavelet transform to decompose the input image into four frequency components: low-low, low-high, high-low, and high-high. A residual unit is introduced in each frequency branch to extract features. To effectively integrate low-frequency structural information and high-frequency detail information, a cross-layer feature fusion unit (CLFU) is constructed between adjacent branches. CLFU incorporates external guidance features into the current features through an affine transformation mechanism, thereby enhancing feature representation capability. In the second stage, a resolution enhancement module (REM) is proposed to reconstruct image details and improve image resolution. REM first extracts shallow features using a 3×3 convolution and further refines them through residual units. The refined features are then resized to a fixed size of 128×128 via bilinear interpolation, followed by upsampling to 512×512 using a pixel shuffle layer. Meanwhile, residual images are introduced through skip connections combined with bilinear interpolation to facilitate information flow within the module. By cascading HPEN with any detector, an LR image object detection framework termed HPELOD is constructed. **【Result】** To evaluate the effectiveness of the proposed method, experiments were conducted on degraded versions of the PASCAL VOC2007 *trainval* (VOC-L) and COCO val 2017 (COCO-L) datasets. Comparisons were made with the baseline detector CenterNet, preprocessing-based methods, and joint enhancement-detection approaches. Experimental results show that, while maintaining detection speed, HPELOD introduces only a small number of additional parameters and computational overhead, yet significantly improves object detection performance on LR images. For example, on the VOC-L and COCO-L datasets, HPELOD achieves improvements of 1.2 and 1.1 mAP, respectively, compared with the second-best method AERIS. In terms of qualitative results, the proposed method can detect more objects even when images suffer from severe noise and blur. **【Conclusion】** The proposed ultra-lightweight pre-enhancement network HPEN and its progressive enhancement strategy effectively alleviate the reconstruction and enhancement challenges of degraded LR images. The resulting HPELOD framework provides an efficient and lightweight solution for object detection in low-resolution images. The proposed method achieves a good balance between detection accuracy and computational efficiency, demonstrating strong practical value and offering useful insights for future research on LR image object detection.

Key words: convolution neural network; image enhancement; object detection; deep learning; feature fusion

0 引言

随着深度学习技术的快速发展,卷积神经网络(CNN)被广泛应用于各种视觉任务如图

像超分辨率 (SR) 重建^[1-5] 和目标检测^[6-8] 等. 近年来, 基于 CNN 的目标检测算法越来越流行, 特别是单阶段检测算法^[9-10], 因其能快速进行目标识别和定位而广受关注. 由于基于 CNN 的目标检测算法大多是面向高分辨率 (HR) 图像的, 若在边缘、纹理等重要细节信息丢失的低分辨率 (LR) 图像上直接进行检测, 则模型的检测精度容易因为 HR 和 LR 图像之间的特征域差距过大而出现明显下降, 从而难以达到实际要求. 值得注意的是, 虽然已有部分研究关注 LR 图像目标检测或伪装检测等特定场景, 但现有方法在应对多种复杂退化因素 (如噪声、模糊) 的低分辨率图像时, 其检测精度、模型效率或泛化性仍存在明显挑战.

为解决以上问题, 一种简单方法是将目标检测模型在 LR 图像上重新训练. 由于 LR 图像丢失大量细节信息, 这种方式容易使得目标检测模型难以从中学习到准确判别特征, 从而限制检测准确率的提升. 考虑到这个问题, 一些学者尝试将 SR 模型与目标检测器级联以改善 LR 图像目标检测性能, 这类方法称为基于预处理 (PP) 的方法. 由于现有的 SR 模型^[11-14] 大多通过像素损失函数如 L1 和 L2 等来训练, 其重建的图像虽然符合人类视觉层面上的需求, 但其中有益于目标检测任务的特征可能因此遭到破坏. 为更好地提升 LR 图像目标检测准确率, 研究人员开发了一种联合图像增强与目标检测 (JED) 的方法^[17-20]. 如 Wang 等^[17] 提出了一个面向 LR 图像目标检测的特征驱动超分辨率方法 (FDSR) 方法. 与 PP 方法不同, 该方法中的 SR 模型是通过目标检测损失和像素损失联合训练, 从而使生成的 SR 图像更好地被检测器感知. 然而, FDSR 采用大尺度 SR 模型作为整个框架的预增强部分, 导致模型推理时间较长, 难以满足实际需求. 此外, 目标检测器本身计算量大, 若再采用大尺度 SR 网络进行预增强, 会存在以下问题: 由于预增强网络不仅采用像素损失还需用到检测损失来引导其训练, 因此, 若 SR 网络过深, 会增加训练过程中检测损失梯度回传的难度; 大尺度 SR 网络不仅会导致整体框架复杂度和参数明显增加, 还会极大延长模型推理时间, 不利于在实际应用中部署. 为解决这个问题, Zhang 等^[19] 开发的 SuperYOLO 和 Cui 等^[20] 设计的自监督学习分辨率自适应编码方法 (AERIS) 都选择采用多任务学习方式来训练检测器. 不同的是, SuperYOLO 是将 SR 支路部署到检测器骨干网络处, 而 AERIS 则是在检测器颈部额外构建一个 SR 解码器. 虽然这两种方法都得益于多任务学习策略使得其在推理阶段可以丢弃 SR 组件而避免引入额外参数和计算量, 但这类方法同样容易遭遇与重训练方法一样的问题: 直接对 LR 图像检测容易提取到不准确特征, 限制模型检测准确率的提高.

另一方面, 考虑到 LR 图像可能存在多种降质因素如噪声、模糊和下采样等, 若直接对其进行 SR, 会存在较大的图像重建挑战. 因此, 在设计预增强网络时, 采用一种分阶递进增强策略: 总体图像增强任务由图像复原和分辨率增强两个阶段依次实现. 前者旨在修复图像受损信息 (如噪声、模糊), 后者则进一步提升复原图像细节信息和分辨率. 值得注意的是, 在增强网络中应先执行图像复原, 而不是分辨率增强. 原因在于若先执行后者, 容易放大 LR 图像中有害信息如噪声和模糊, 且图像分辨率增强后再进行图像复原会引入大量计算负担.

为能以较低参数和计算负担有效提升 LR 图像目标检测精度, 本文提出了一种用于 LR 图像的目标检测方法 (HPELOD), 其主要贡献在于以下 3 点.

1) 提出一个超轻量级 SR 网络超轻量级预增强网络 (HPEN) 来处理具有不同退化类型和任意比例因子的 LR 图像, 以在检测器前实现图像层增强. 为降低复杂降质带来的图像重建挑战, HPEN 采用分阶递进增强策略, 依次解决图像复原和分辨率增强子问题.

2) 提出一个小波复原模块 (WRM), 结合离散小波变换 (DWT) 和频率特征渐进融合策略, 有效实现图像复原. 在 WRM 中, 构建一种跨层特征融合单元 (CLFU), 其通过仿射变

换整合低层指导信息到高层频率分量中, 以提升特征的表达能力.

3) 通过将 HPEN 与目标检测器进行级联, 可以得到一种新的 HPELOD. 相较于其他方法, HPELOD 能在模型性能、参数和复杂度之间实现更好的平衡.

1 本文方法

本节将详细介绍所提出的低分辨率图像目标检测方法 HPELOD, 包括其总体框架、分阶递进增强网络结构以及损失函数设计.

1.1 HPELOD 总体框架

本文提出一种新的 LR 图像目标检测框架 HPELOD, 它由 HPEN 和一个目标检测器构成, 如图 1 所示. 一般来说, 基于视觉的任务如目标检测都依赖于输入图像的结构和纹理信息, 以便提取出准确特征来实现最终定位和分类功能. 然而, 对于丢失重要细节和分辨率的 LR 图像, 若直接使用在高质量图像上训练得到的检测器对其直接检测, 则会因为检测特征域差距过大, 最终检测性能下降. 为避免这个问题, 设计 HPEN 用于在检测器前对 LR 图像进行预增强, 其具体结构将会在 1.2 节详细说明.

1.2 分阶递进增强网络

现有基于 CNN 的 SR 方法大多通过像素损失来引导网络构建从 LR 到 HR 图像之间的非线性映射来实现 SR 任务. 与这类 SR 模型不同, HPEN 的目的是提升目标检测精度, 而不是改善图像的可视化质量. HPEN 的具体流程如图 1 所示. 首先, 给定一个 LR 图像为 $I_{LR} \in \mathbf{R}^{512/s \times 512/s \times 3}$, 其中 s 为采样率, 将 I_{LR} 输入 WRM 中进行图像复原, 得到中间复原图像 $I_{LR} \in \mathbf{R}^{512/s \times 512/s \times 3}$. 随后, 将 I_{LR} 输入分辨率增强模块 (REM) 中进行分辨率增强, 得到最终 SR 图像 $I_{SR} \in \mathbf{R}^{512 \times 512 \times 3}$. 通过采用这种分阶递进增强策略, 有效降低了图像增强挑战, 从而改善模型性能.

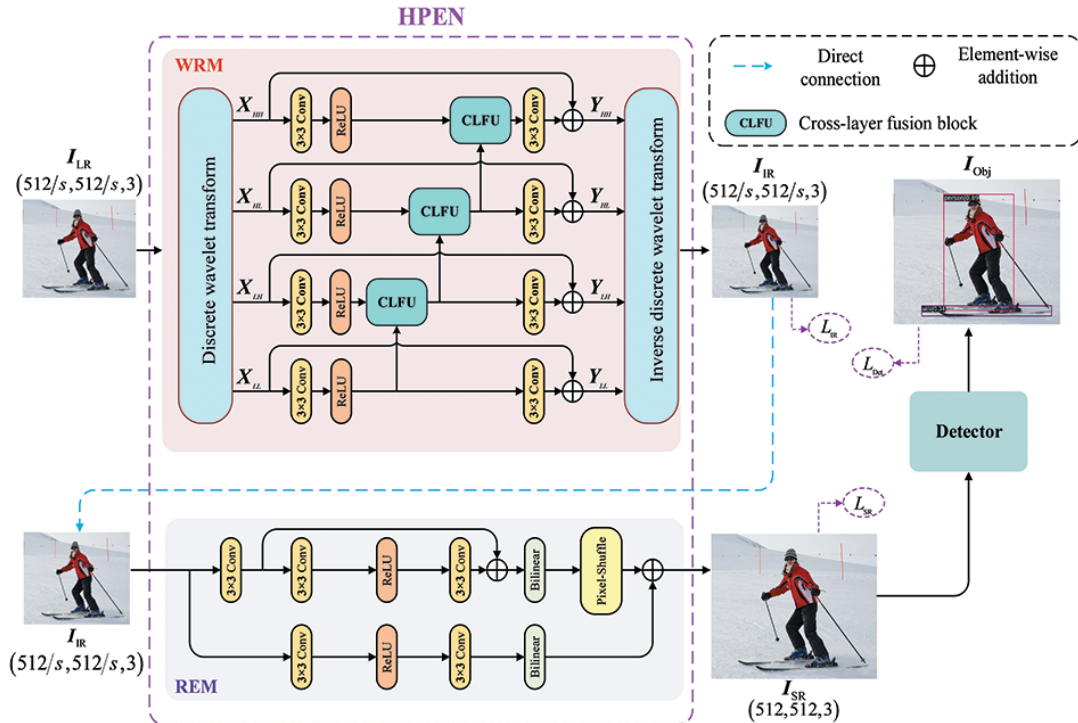


图 1 HPELOD 总体框架

Fig. 1 The overall framework of HPELOD

1.2.1 小波复原模块

为降低预增强网络参数和计算量,同时利用频率相关的信息强化图像中对检测友好的结构化特征,提出 WRM 作为 HPEN 的图像复原部分. WRM 主要在频率域中对 LR 图像进行增强,其具体结构如图 1 所示. WRM 通过使用 2 维 (2D) DWT 将输入图像分解成 4 个频率分量,包括低-低频 (LL) 分量、低-高频 (LH) 分量、高-低频 (HL) 分量和高-高频 (HH) 分量. 其中, LL 为低频分量,其保留了输入图像的内容信息,可以认为是输入图像的一个近似表示;而 LH、HL、HH 属于高频分量,分别保留了输入图像垂直、水平和对角线方向的高频信息. LR 图像分解过程可由以下公式表示:

$$[X_{LH}, X_{HL}, X_{HH}, X_{LL}] = M(I_{LR}), \quad (1)$$

式中: $M(\cdot)$ 为 DWT 函数; X_{LL} 为低频分量 LL, 而 X_{LH} 、 X_{HL} 、 X_{HH} 分别为高频分量 LH、HL、HH. 这些频率分量的维度都为 $512/2s \times 512/2s \times 3$.

在得到 4 个频率分量后,采用残差单元分别对每个分量进行频率特征提取. 其中,残差单元主要由两个 3×3 卷积层和一个线性激活层 ReLU 构成. 由于每个频率分量之间存在很强的空间相关性. 因此,本文开发了一个 CLFU,旨在利用这相关性来改善每个高频分量的特征表达能力,其具体结构如图 2 所示. 为降低参数和计算复杂度,CLFU 选择用 3×3 深度卷积而不是标准 3×3 卷积去实现特征提取和学习仿射变换参数. 给定一个特征图 $H \in \mathbf{R}^{h \times w \times c}$, 其中, h 和 w 表示特征图高和宽, c 表示特征图的通道数. 如图 2 所示, H 输入第一个 3×3 深度卷积后得到待变换特征图 $P \in \mathbf{R}^{h \times w \times c}$. 与此同时,外部指导特征图 $G \in \mathbf{R}^{h \times w \times c}$ 输入到两个 3×3 深度卷积后分别得到缩放参数 $\gamma \in \mathbf{R}^{h \times w \times c}$ 和偏移参数 $\beta \in \mathbf{R}^{h \times w \times c}$. 随后,对 P 执行仿射变换后得到变换特征 $\hat{P} \in \mathbf{R}^{h \times w \times c}$. 这个过程可由如下公式表示:

$$\hat{P} = \gamma \odot P + \beta, \quad (2)$$

式中: \odot 和 $+$ 分别表示像素乘和加操作. 经过仿射变换后, \hat{P} 输入最后一个 3×3 深度卷积做进一步细化,然后与经过跳跃连接传输过来的 H 进行像素相加,产生最终融合特征 $\hat{H} \in \mathbf{R}^{h \times w \times c}$.

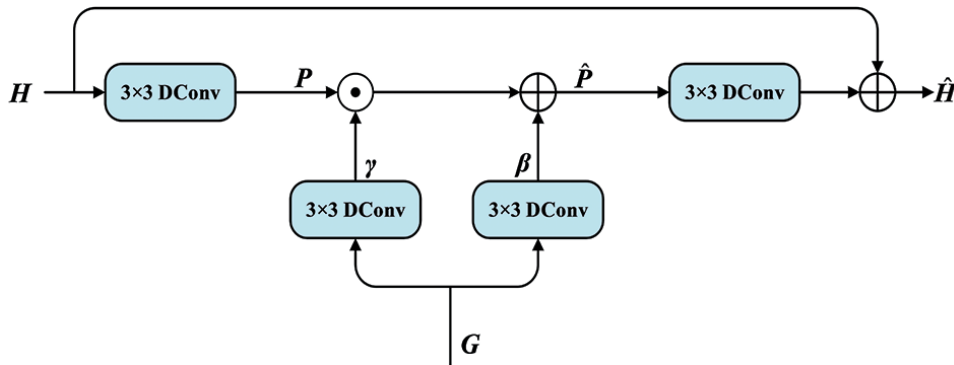


图 2 CLFU 的结构

Fig. 2 The structure of CLFU

在 WRM 中,将第一个 CLFU 设置在 LH 分量所在支路中,其以 LL 分量作为指导特征. 第二个和第三个 CLFU 分别设置在 HL 和 HH 分量支路中,并以第一个和第二个 CLFU 的输出作为指导特征. 通过这种渐进融合方式,最终的 HH 分量可以融合之前所有频率分量信息,最大程度重建出完整图像纹理.

经过上述特征提取和细化后, 最终复原图像可以由 4 个增强后的频率分量通过逆离散小波变换 (IDWT) 重构得到. 这个过程表示如下:

$$\mathbf{I}_{\text{IR}} = R(Y_{\text{LH}}, Y_{\text{HL}}, Y_{\text{HH}}, Y_{\text{LL}}), \quad (3)$$

式中: Y_{LH} 、 Y_{HL} 、 Y_{HH} 、 Y_{LL} 分别为增强后的 LH、HL、HH、LL 分量; $R(\cdot)$ 为 IDWT 函数. 因为 Haar 小波的计算高效性, 故将其用于 DWT 和 IDWT 中的小波基.

1.2.2 分辨率增强模块

分辨率增强模块 (REM) 具体结构展示在图 1 中. 如图 1 所示, REM 先采用一个 3×3 卷积层从复原图像 \mathbf{I}_{IR} 中提取基础浅层特征, 然后采用一个残差单元进行特征细化. 与 WRM 旨在从 LR 图像中恢复干净图像不同, REM 的目标是增强图像细节信息和分辨率. 但由于在 HPELOD 训练过程中, 尺度因子 s 并不是固定的. 因此, 传统 SR 方法所常用固定采样因子的上采样器如反卷积层并不适用于 REM. 为此, REM 先采用双线性插值 (Bilinear) 将细化后特征的空间维度缩放到固定尺寸 128×128 , 紧接着采用一个像素洗牌层将缩放后的特征上采样到尺寸为 512×512 的残差图像. 此外, 为促进 REM 内信息流动, 额外将 \mathbf{I}_{IR} 通过一条整体特征提取支路后与残差图像像素相加, 形成最终增强图像 \mathbf{I}_{SR} . 与普通跳跃连接支路仅采用直连操作不同, 整体特征提取支路采用两个 3×3 卷积和一个 ReLU 去提取图像总体结构特征, 然后使用 Bilinear 将特征的空间尺寸直接缩放到 512×512 以便与残差图像进行像素相加.

1.3 损失函数

为使 HPEN 学习到目标检测所关注的结构化特征, 采用一种联合 HPEN 和检测器的端到端方式来训练 HPELOD. 为保证 HPEN 中 WRM 和 REM 能分别实现各自的任务, 采用一种深度监督策略: WRM 和 REM 分别由图像复原损失函数 L_{IR} 和分辨率增强损失函数 L_{RE} 进行监督. 此外, 引入针对目标检测任务的检测损失函数 L_{Det} 来指导 HPELOD 模型训练. 总体损失函数 L_{T} 由如下公式表示:

$$L_{\text{T}} = L_{\text{IR}} + L_{\text{RE}} + L_{\text{Det}}, \quad (4)$$

假定第 i 个 HR 图像和对应的 LR 图像分别为 \mathbf{I}_{HR}^i 和 \mathbf{I}_{LR}^i . 则 L_{IR} 和 L_{RE} 分别由如下式子表示:

$$L_{\text{IR}} = \frac{1}{N} \sum_{i=1}^N \| F_{\text{WRM}}(\mathbf{I}_{\text{LR}}^i) - F_{\text{bic}}(\mathbf{I}_{\text{HR}}^i) \|_1, \quad (5)$$

$$L_{\text{RE}} = \frac{1}{N} \sum_{i=1}^N \| F_{\text{REM}}(\mathbf{I}_{\text{IR}}^i) - \mathbf{I}_{\text{HR}}^i \|_1, \quad (6)$$

式中: N 为训练批次大小; $F_{\text{WRM}}(\cdot)$ 和 $F_{\text{REM}}(\cdot)$ 分别为 WRM 和 REM 函数; $\| \cdot \|_1$ 为 L_1 范数; \mathbf{I}_{IR}^i 为 WRM 复原的第 i 个图像; $F_{\text{bic}}(\cdot)$ 定义为双三次插值.

2 实验结果

为验证所提方法的有效性, 本节将在多个标准数据集上对 HPELOD 进行客观指标对比、消融实验、泛化性验证以及复杂度分析.

2.1 实验设置

采用数据集 PASCAL VOC 和 COCO 来评估本文方法性能. 在这些数据集中, 所有图像被缩放到 512×512 大小后作为 HR 图像. 对于本文训练和测试所用的 LR 图像, 采用 AERIS 提供的降质模型来产生. 在 PASCAL VOC 数据集中, VOC2007 trainval 和 VOC2012 trainval 用于训练, VOC2007 test 用于测试. 在 COCO 数据集中, COCO train2017 和 COCO val2017 分别作为训练集和测试集. 为更好区分 HR 图像和降质图像测试集, 分别定义降质版本的

VOC2007 test 和 COCO val2017 为 VOC-L 和 COCO-L.

本文采用 MMDetection 来搭建 HPELOD 框架. 与 AERIS 一致, 选择 CenterNet^[9] 作为 HPELOD 的检测器. 在训练阶段, 采用随机裁剪和随机翻转进行数据增强. 总训练 epoch 为 80, 采用 SGD 作为优化器; 学习率为 0.005, 并在 epoch 为 40 和 70 时降低为原来十分之一.

本文采用交并比 (IoU) 为 0.50 时所得到的平均精度均值 (记为 mAP_{50}) 来衡量模型在 VOC-L 上的检测性能. 对于 COCO-L, 除了采用 $mAP_{50:95}$ 外, 还采用在 IoU 分别为 0.5 和 0.75 情况下的 AP 值 (分别记为 AP_{50} 和 AP_{75}) 来衡量模型性能.

2.2 不同方法的性能比较

所比较方法分为 3 类: 基准 (Baseline) 方法 CenterNet (H)^[9] 和 CenterNet (L)^[9] 分别表示在 HR 和 LR 图像数据集上训练得到的 CenterNet 模型; PP 方法表示在检测器 CenterNet (H) 前采用 SR 模型 RFDN^[11]、DSSR^[12]、RLFN^[13]、MDRN^[14]、SPAN^[15] 和 SRConvNet^[16] 来对 LR 图像进行预增强; JED 方法包括 SuperYOLO^[19]、AERIS^[20] 和 HPELOD. 表 1 展示了不同方法在 VOC-L 和 COCO-L 上的检测结果. 从表 1 可知, 在所有比较方法中, CenterNet (H) 准确率最低, 表明检测器在 HR 和 LR 图像上分别提取的特征之间存在较大域差距. 虽然 CenterNet (L) 的性能优于 CenterNet (H), 但 AP 值提升幅度并不够明显. 这可能是由于 CenterNet (L) 从 LR 图像上提取的特征不够准确, 影响后续检测器的判断. 相比 CenterNet (H), PP 方法通过对 LR 图像进行预增强, 提升了图像质量, 一定程度上改进了模型检测准确率. 即使与最新的 PP 方法 SPAN 和 SRConvNet 相比, 本文方法 HPELOD 在检测精度上仍能保持显著优势. 总的来说, 在所有比较方法中, HPELOD 在所有测试集上都获得了最优结果. 这主要归功于 HPELOD 中的 HPEN 通过渐进地执行图像复原和分辨率增强阶段, 更好地恢复 LR 图像中检测器所关注的结构化信息.

表 1 不同方法在 VOC-L 和 COCO-L 上的性能比较

Tab. 1 Performance comparison of different methods on VOC-L and COCO-L

方法		VOC-L	COCO-L
		$mAP_{50} / \%$	$mAP_{50:95} / \%$
Baseline	CenterNet (H) ^[9]	39.5	14.7
	CenterNet (L) ^[9]	53.8	15.9
PP	RFDN ^[11]	43.8	15.9
	DSSR ^[12]	40.9	15.7
	RLFN ^[13]	46.5	17.4
	MDRN ^[14]	46.4	15.9
	SPAN ^[15]	44.3	15.7
	SRConvNet ^[16]	45.7	17.0
	SuperYOLO ^[19]	55.4	17.8
JED	AERIS ^[20]	57.5	18.3
	HPELOD	58.7	19.4

图 3 和图 4 展示了所提方法 HPELOD 与其他方法的主观检测结果对比. 其中, CenterNet (H) 在 HR 图像上的检测结果可以认为是性能上限. 值得注意的是, 其他方法都是在 LR 图像上进行检测. 从图 3 中可以发现, 相比其他方法, 即使在 LR 图像存在大量噪声和模糊情况下, HPELOD 仍能检测出更多人和马. 从图 4 中也可以观察到, 直接以 LR 图像作为检测器

输入而未经过预增强的方法如 AERIS 和 SuperYOLO 更容易出现误检问题. 例如, AERIS 将小车误识别为卡车, 而 SuperYOLO 将狗识别成长颈鹿. 这主要是因为这类方法直接从 LR 图像中提取到不准确判别特征, 误导了检测器. 以上结果表明 HPELOD 所采用的 HPEN 通过分阶递进增强策略, 能有效重构面向检测任务的 SR 图像, 进而提升模型检测精度.

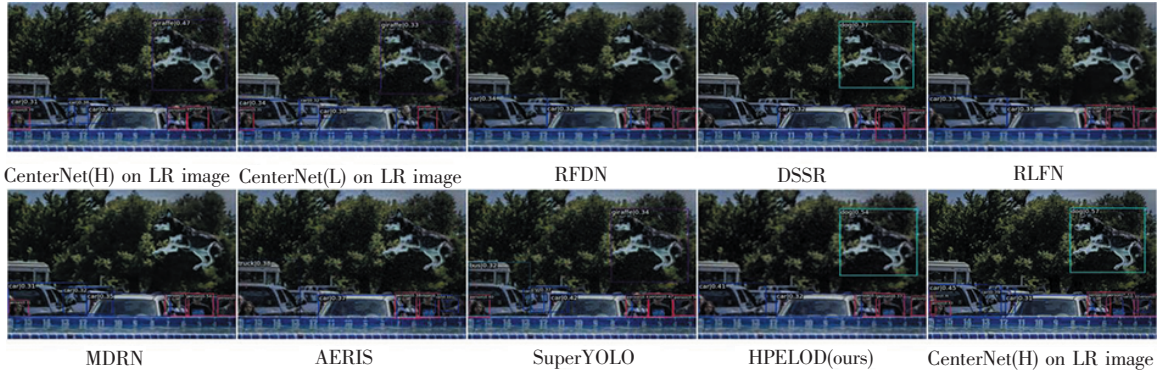


图 3 不同方法在 COCO-L 数据集 000000067213 图像上获得的检测结果

Fig. 3 Detection results obtained by different methods on image 000000067213 of the COCO-L dataset

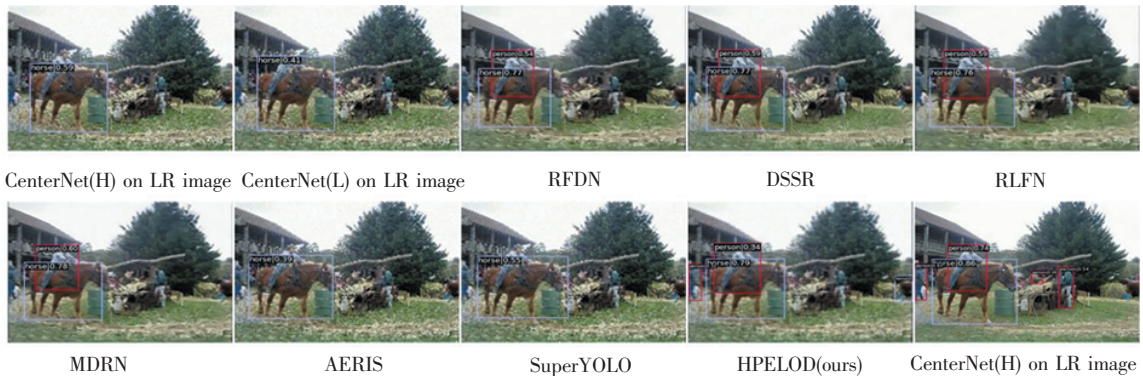


图 4 不同方法在 VOC-L 数据集 000166 图像上获得的检测结果

Fig. 4 Detection results obtained by different methods on image 000166 from the VOC-L dataset

2.3 消融实验

为深入研究 HPEN 中各个模块对模型性能的影响, 使用 VOC-L 进行消融实验, 其结果展示在表 2 中. 其中, Base 是基准模型网络, 其等同于表 1 中的 CenterNet (L); DWT 在表 2 中是指完整 WRM 去除 CLFU 后产生的模块. 从表 2 中知, 在仅执行图像复原阶段时, 通过应用 DWT 到 Base 中, 在 VOC-L 上得到 2% mAP 提升. 结果表明, DWT 通过将 LR 图像分解成不同频率分量并分别进行特征细化来改善图像细节和纹理, 能使后续检测器提取更准确的特征. 在此基础上, 引入 CLFU 整合不同频率分量特征, 模型性能得到进一步提升, 说明 CLFU 和 DWT 的能力是互补的. 在使用 WRM 的基础上, 采用 REM 执行分辨率增强阶段, 能在 VOC-L 上得到 2.4% 的 mAP 值提升, 表明 REM 能通过提升复原图像分辨率和细节信息来改善模型检测性能. 同时, 也说明了对 LR 图像采用分阶递进增强策略, 能有效降低图像重建难度.

表 2 HPEN 不同模块的消融实验

Tab. 2 Ablation experiments on different modules of HPEN

方法	WRM		REM	mAP ₅₀ /%
	DWT	CLFU		
Base	×	×	×	53.8
	√	×	×	55.8
HPELOD	√	√	×	56.3
	√	√	√	58.7

2.4 泛化性实验

为验证 HPELOD 的泛化能力, 将知名 YOLOv7-t^[10] 代替所有方法原有目标检测器 CenterNet 进行比较实验, 所有调整后模型都重新训练. 表 3 展示了不同方法在 COCO-L 上的检测结果. 由表 3 可知, 相比其他方法, 更换检测器后, HPELOD 仍能实现最优 AP 值. 表明了 HPELOD 具有良好的泛化性, 对于其他检测器仍然适用.

2.5 网络规模和计算复杂度分析

表 4 展示了不同方法的参数和帧率 (FPS) 值, 其中 CenterNet 作为检测器. FPS 值是通过使用单张 GTX 1080Ti GPU 在 VOC-L 上测试得到的. 由表 4 可知, 即使与最新的 PP 方法 SPAN 和 SRConvNet 相比, HPELOD 所需增加的参数 (仅 0.03 M) 明显更少. 原因在于 HPELOD 中 HPEN 采用超轻量级的结构设计和结合分阶递进策略. 此外, HPEN 中的 WRM 采用 DWT 将输入 LR 图像转换到频率域进行特征提取时, 每个频率分量的分辨率仅为输入的一半, 使引入的计算负担大大降低. 由表 4 还可知, 虽然 SuperYOLO 通过采用多任务学习策略而没有引入额外参数与计算, 但 HPELOD 同样能以轻微增加的参数和运行时间实现明显更好的性能 (详见表 1 和表 3). 总的来说, 相比其他方法, HPELOD 能在模型性能、参数数量与推理时间之间实现更好平衡.

表 3 不同方法在 COCO-L 上的性能比较

Tab. 3 Performance comparison of different methods on COCO-L

方法	COCO-L			
	mAP _{50:95} /%	mAP ₅₀ /%	mAP ₇₅ /%	
Base	YOLOv7-t (H) ^[10]	17.4	28.2	17.7
	YOLOv7-t (L) ^[10]	20.6	33.8	21.2
PP	RFDN ^[11]	13.4	21.6	13.9
	DSSR ^[12]	13.6	21.4	14.4
	RLFN ^[13]	15.5	24.7	16.2
	MDRN ^[14]	12.9	20.6	13.3
	SuperYOLO ^[19]	20.9	34.2	21.5
JED	AERIS ^[20]	21.2	34.7	21.9
	HPELOD	23.0	36.6	24.0

表 4 不同方法参数、FPS 和 mAP 值比较

Tab. 4 Comparison of different method parameters, FPS, and mAP values

	方法	参数/M	FPS	mAP ₅₀ /%
Baseline	CenterNet (H) ^[9]	14.43	167	39.5
	CenterNet (L) ^[9]	14.43	167	53.8
PP	RFDN ^[11]	15.07	32	43.8
	DSSR ^[12]	22.20	2.59	40.9
	RLFN ^[13]	14.95	41	46.5
	MDRN ^[14]	14.75	15	46.4
	SPAN ^[15]	16.27	41.7	44.3
	SRConvNet ^[16]	15.43	10.9	45.7
	SuperYOLO ^[19]	14.43	167	55.4
JED	AERIS ^[20]	14.59	143	57.5
	HPELOD	14.46	143	58.7

3 结语

针对现有大多目标检测模型处理多种降质条件下的 LR 图像时检测精度明显下降问题, 开发了一种超轻量 SR 网络 HPEN, 以在检测器前实现图像层增强. 由于复合降质条件下直接增强 LR 图像难度大, 因此 HPEN 将图像增强分为图像复原、分辨率增强两阶段并依次执行. 在第一阶段, WRM 通过利用 DWT 将图像转换为频率域并辅以 CLFU, 使其能以较低计算负担结合不同频率的特征信息. 在第二阶段, REM 负责对不同分辨率的复原图像进行统一增强, 以改善图像细节信息. 通过级联 HPEN 和任意一个检测器, 可以得到最终的 HPELOD. 实验结果表明, 相比其他算法, 在所有测试集中, HPELOD 的检测精度最高, 且能在检测速度和检测性能之间达到最好的平衡.

参 考 文 献

- [1] DONG C, LOY C C, HE K, et al. Image super-resolution using deep convolutional networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(2): 295-307.
- [2] SHI W, CABALLERO J, HUSZAR F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2016:1874-1883.
- [3] CHEN R, ZHANG H, LIU J X. Multi-attention augmented network for single image super-resolution[J]. Pattern Recognition, 2022, 122: 108349.
- [4] ZHAO L L, GAO J Y, DENG D H, et al. SSIR: Spatial shuffle multi-head self-attention for single image super-resolution[J]. Pattern Recognition. 2024, 148: 110195.
- [5] 邱春林, 王冰莹, 胡凯. 一种结合 Mamba 和 YOLOv8 的结肠镜图像息肉检测算法[J]. 湘潭大学学报(自然科学版), 2025, 47(3): 54-64.
- [6] 马仪, 周仿荣, 文刚, 等. 基于融合注意力机制的电力设施遥感识别方法[J]. 测绘通报, 2024(增刊 2): 182-186.

- [7] 徐博文, 童孟军. 基于改进的 Faster R-CNN 的小麦麦穗检测识别[J]. 湘潭大学学报(自然科学版), 2022, 44(4): 48-59.
- [8] 王志涛, 张瑞菊, 王坚, 等. 基于 YOLOv5 改进的交通场景目标检测算法 HCAM-YOLO[J]. 测绘通报, 2024(11): 61-67.
- [9] ZHOU X, WANG D, KRÄHENBÜHL P. Objects as points[EB/OL]. (2019-04-16)[2025-3-14]. <https://arxiv.org/abs/1904.07850>.
- [10] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2023: 7464-7475.
- [11] LIU J, TANG J, WU G. Residual feature distillation network for lightweight image super-resolution[C]//European Conference on Computer Vision Workshops, Glasgow, UK, 2020: 41-55.
- [12] LI F, WU Y, BAI H, et al. Learning detail-structure alternative optimization for blind super-resolution [J]. IEEE Transactions on Multimedia, 2022, 25: 2825-2838.
- [13] KONG F, LI M, LIU S, et al. Residual local feature network for efficient super-resolution[C]//IEEE Conference on Computer Vision and Pattern Recognition. 2022: 766-776.
- [14] MAO Y, ZHANG N, WANG Q, et al. Multi-level dispersion residual network for efficient image super-resolution[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2023: 1660-1669.
- [15] WAN C, YU H, LI Z, et al. Swift parameter-free attention network for efficient super-resolution[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2024: 6246-6256.
- [16] LI F, CONG R M, WU J J, et al. SRConvNet: A transformer-style convnet for lightweight image super-resolution[J]. International Journal of Computer Vision, 2024, 133(1): 173-189.
- [17] WANG B, LU T, ZHANG Y. Feature-driven super-resolution for object detection[C]//Control, Robotics and Cybernetics, 2020: 211-215.
- [18] HARIS M, SHAKHNAROVICH G, UKITA N. Task-driven super resolution: Object detection in low-resolution images[C]//International Conference on Neural Information Processing, 2021: 387-395.
- [19] ZHANG J, LEI J, XIE W, et al. SuperYOLO: Super resolution assisted object detection in multimodal remote sensing imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-15.
- [20] CUI Z, ZHANG Y, GU L, et al. Exploring resolution and degradation clues as self-supervised signal for low quality object detection[C]//European Conference on Computer Vision, 2022: 473-49.

(责任编辑:夏金玉)