

文章编号: 1003-1421(2026)09-0199-10

引用格式: 白玉龙. 基于RT-DETR的轻量化铁路侵限异物检测模型研究[J]. 铁路运输与经济, 2026, 48(9): 199-208.

BAI Yulong. Lightweight Railway Intrusion Obstacle Detection Model Based on RT-DETR[J]. Railway Transport and Economy, 2026, 48(9): 199-208.

基于RT-DETR的轻量化铁路侵限异物检测模型研究

白玉龙

(国能新朔铁路有限责任公司 调度指挥中心, 内蒙古 呼和浩特 010000)

摘要: 针对铁路侵限异物检测模型参数量大、计算复杂而难以嵌入资源受限的铁路边侧设备, 导致实时预警能力不足的问题, 提出一种基于RT-DETR的轻量化模型。该模型从特征提取、特征融合和计算压缩3个层面进行优化: 在特征提取阶段引入空间状态建模机制, 通过线性复杂度与硬件感知设计增强关键特征的捕捉能力; 在特征融合阶段采用分组混洗卷积替换标准卷积, 降低参数量与计算负担; 在计算压缩阶段引入哈尔小波下采样模块, 对特征图进行高效压缩, 显著减少解码器的计算开销。实验结果表明, 所提模型在仅占用25.995 MB参数和71.411 G FLOPs计算量的条件下, 在多项复杂度指标上均优于现有先进模型, 同时保持高精度检测能力, 满足了铁路侵限检测高精度、低延迟的核心业务需求。

关键词: 铁路侵限; 异物检测; RT-DETR; 空间状态模型; 哈尔小波下采样

中图分类号: U282+.1; TP391.4 文献标识码: A DOI: 10.16668/j.cnki.issn.1003-1421.20251010001

Lightweight Railway Intrusion Obstacle Detection Model Based on RT-DETR

BAI Yulong

(Scheduling Center, China Energy Xinshuo Railway Co., Ltd., Hohhot 010000, Inner Mongolia, China)

Abstract: The railway intrusion obstacle detection models, due to their large parameter size and high computational complexity, are difficult to deploy on resource-constrained railway trackside equipment, resulting in insufficient real-time warning capability. In response to this problem, a lightweight model based on RT-DETR was proposed. The optimization of this model was carried out from three aspects: feature extraction, feature fusion, and computational compression. In the feature extraction stage, a state-space modeling mechanism was introduced to enhance the ability to capture key features through linear complexity and hardware-aware design. In the feature fusion stage, grouped shuffle convolution was adopted to replace standard convolution, reducing parameter count and computational burden. In the computational compression stage, a Haar wavelet downsampling module was incorporated to efficiently compress feature maps, significantly lowering the computational overhead of the decoder. Experimental results demonstrate that the proposed model, with only 25.995 MB parameters and 71.411 G FLOPs, outperforms the existing state-of-the-art models on multiple complexity metrics while maintaining high detection accuracy, thus satisfying the core business requirements of high-precision and low-latency railway intrusion detection.

Keywords: Railway Intrusion; Obstacle Detection; RT-DETR; State-Space Model; Haar Wavelet Downsampling

收稿日期: 2025-10-10

0 引言

铁路线路、接触网为列车运行提供稳定可靠的线路基础和电力传输通道^[1]。侵入铁路限界的异物可能引发事故,严重威胁铁路运行安全^[2]。针对铁路侵限异物的检测,已有研究从多个角度展开,包括传统图像处理的方法和深度学习的目标检测方法。在传统图像处理方法中,多利用异物的形态学特征进行识别,其核心思路是通过前景与背景的分离判断异物是否存在。这类方法高度依赖算法参数设定与特征提取策略,对复杂背景适应性差^[3]。随着深度学习技术的发展,基于神经网络的目标检测方法为铁路异物识别提供了新思路。其中,以卷积神经网络(Convolutional Neural Network, CNN)为基础的模型在异物检测中得到广泛应用。Guo等^[4]基于Faster R-CNN构建了针对传输线异物的检测模型,但存在检测速度较慢的问题;Song等^[5]通过改进YOLOv4,引入K-means聚类与距离交并比非极大值抑制(Distance Intersection over Union-Non Maximum Suppression, DIoU-NMS)算法,提升了高压线路异物的检测效果,但检测精度仍有待提高,YOLO系列模型已更新到YOLOv8,模型的推理速度与精度获得极大提高。部分研究者尝试通过任务对齐检测头、动态标签分配机制或解耦样本策略等方法优化检测性能,提升模型的鲁棒性与空间一致性^[6-7]。然而,基于CNN的模型普遍存在手工设计复杂、后处理流程繁琐等问题,限制了其扩展性与训练效率。为克服上述局限,研究者开始探索基于Transformer架构的检测模型。Carion等^[8]提出一种基于Transformer的端到端目标检测模型(Detection Transformer, DETR),可有效简化检测流程,提升模型的可扩展性,其变体DAB-DETR^[9]将解码器的查询显式设计为可学习的锚框,并在每层动态更新,让查询有了明确的物理意义;Zhang^[10]提出密集且独特的查询生成策略,根据图像内容动态生成查询并强化查询间的区分度,减少重复预测、降低计算冗余。实时目标检测模型(Real-Time Detection Transformer, RT-DETR)^[11]通过轻量化编码器、稀疏注意力机制与没有非极大值抑制(Non-Maximum Suppression, NMS)的设计,在实时检测的同时兼顾

了高精度与部署便利性,成为当前铁路异物检测中具有潜力的模型之一。尽管如此,现有基于Transformer的模型仍存在计算复杂度高、对硬件要求较高等问题,在复杂铁路场景下的轻量化与实时性仍有优化空间。

综上所述,既有研究在铁路侵限异物检测方面已取得一定成果,但仍存在以下不足:传统方法依赖人工特征,泛化能力有限;基于CNN模型在检测速度与精度之间难以兼顾;基于Transformer的模型计算开销大,难以适应资源受限的部署环境。为此,在RT-DETR基础上,从3个方面进行优化:首先,采用轻量级空间状态模型(Object Detection Mamba, ODMamba)^[12]作为骨干网络,兼顾计算效率与长程依赖建模能力;其次,使用分组混洗卷积(Group Shuffle Convolution, GSConv)^[13]替代标准卷积(Standard Convolution, SC),降低模型参数量与计算负担;最后,在编码器与解码器之间引入哈尔小波下采样(Haar Wavelet Downsampling, HWD)^[14]模块,对特征图进行高保真压缩,显著降低了解码器的计算复杂度,使得模型在保障检测精度的同时,实现了对铁路侵限异物的高效识别,有效满足铁路现场对高精度、低延迟侵限检测与预警的核心业务需求。

1 RT-DETR 模型结构

RT-DETR是兼顾精度与实时性的端到端目标检测模型,沿用了模型DETR的核心思想,即利用自注意力机制(Self Attention, SA)对图像中的目标进行全局建模并直接回归边界框,而非依赖传统的候选框生成与后处理过程。其网络结构由主干网络(Backbone)、高效混合编码器(Efficient Hybrid Encoder)和带辅助预测头的Transformer解码器3部分级联构成。主干网络输出多尺度特征图后,高效混合编码器通过尺度内特征交互(AIFI)与跨尺度特征融合(CCFM)2个解耦模块,先利用自注意力增强各尺度内部语义,再聚合多尺度信息,从而在保证全局上下文建模能力的同时大幅降低计算复杂度;随后,解码器采用IoU感知的查询选择机制,从编码特征中筛选兼具高分类置信度与高IoU潜力的初始目标查询,使解码器多层迭代优化更具针对性,并通过

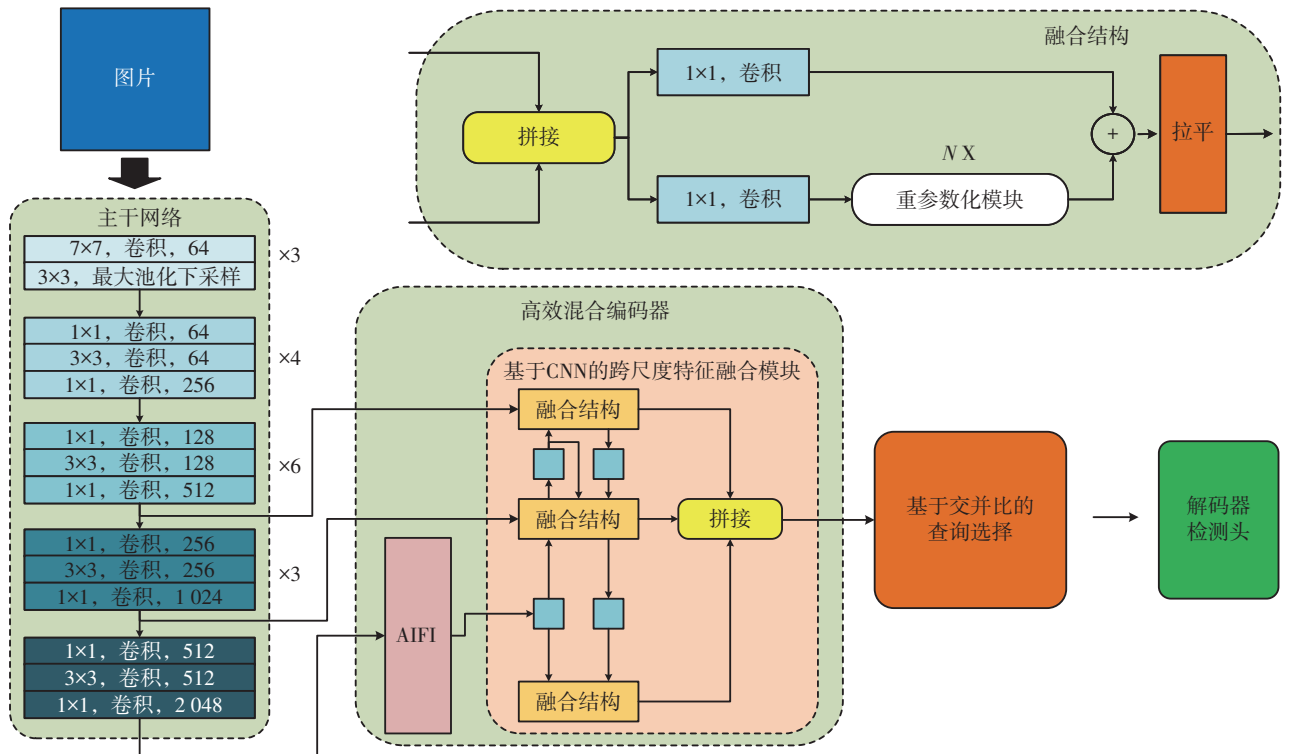
每层辅助预测头施加监督以加速收敛。此外,该模型支持灵活调节解码器层数以适应不同算力需求,且全程无需非极大值抑制(NMS)后处理,从而在实时性要求下,实现了检测精度与推理速度的平衡。相比传统DETR及其改进模型,RT-DETR在常见物体数据集(Common Objects in Context, COCO)^[15]上实现了更快的推理速度和更高的检测效率。因此,将RT-DETR作为改进的基线模型,RT-DETR模型结构如图1所示。实验中采用上海人工智能实验室主导的OpenMMLab目标检测工具箱中的MMDetection3.0^[16]目标检测框架中以残差网络^[17]为主干网络的RT-DETR。

2 改进的RT-DETR

2.1 轻量级空间状态模型

传统的基于CNN的特征提取网络因其局部感受野的局限性,难以捕获全局上下文信息,而基于Transformer的特征提取网络虽然具备建模全局依赖的能力,但其自注意力机制在高分辨率特征图上,存在计算成本高、复杂度为二次方增长的问题,从

而限制了其在实时目标检测中的应用^[18]。为克服上述不足,引入轻量级空间状态模型ODMamba作为RT-DETR的主干网络,以提升特征提取的效率和表达能力,ODMamba结构如图2所示。ODMamba结合状态空间模型(State Space Model, SSM)^[19]和深度学习优化设计,具备线性复杂度的计算效率,并通过模块化结构实现了全局与局部特征的深度融合。ODMamba包括初始模块(Simple Stem), ODSSBlock与视觉线索融合模块(Vision Clue Merge, VCM),通过几个模块的堆叠实现对输入图片的4个阶段特征提取。初始模块是网络的入口,它采用2个步长为2、卷积核大小为3的卷积层进行堆叠,通过2次采样初步提取并压缩输入图像的特征,为后续网络提供合适的输入尺度。ODSSBlock是ODMamba的核心特征提取模块,该模块首先通过卷积、批量归一化及激活层对输入特征进行通道变换与非线性增强,随后将特征分流为2条并行的计算路径:在主路径中,二维选择扫描模块(Selective-Scan-2D, SS2D)机制从4个对称方向对二维特征图展开并执行状态空间序列建模,以线性复杂度捕获全局长程



注:“1×1, 卷积, 512”表示该卷积核的宽度为1×1, 输出通道数为512, 其他同; AIFI表示基于注意力的尺度内特征交互。

图1 RT-DETR模型结构
Fig.1 RT-DETR model structure

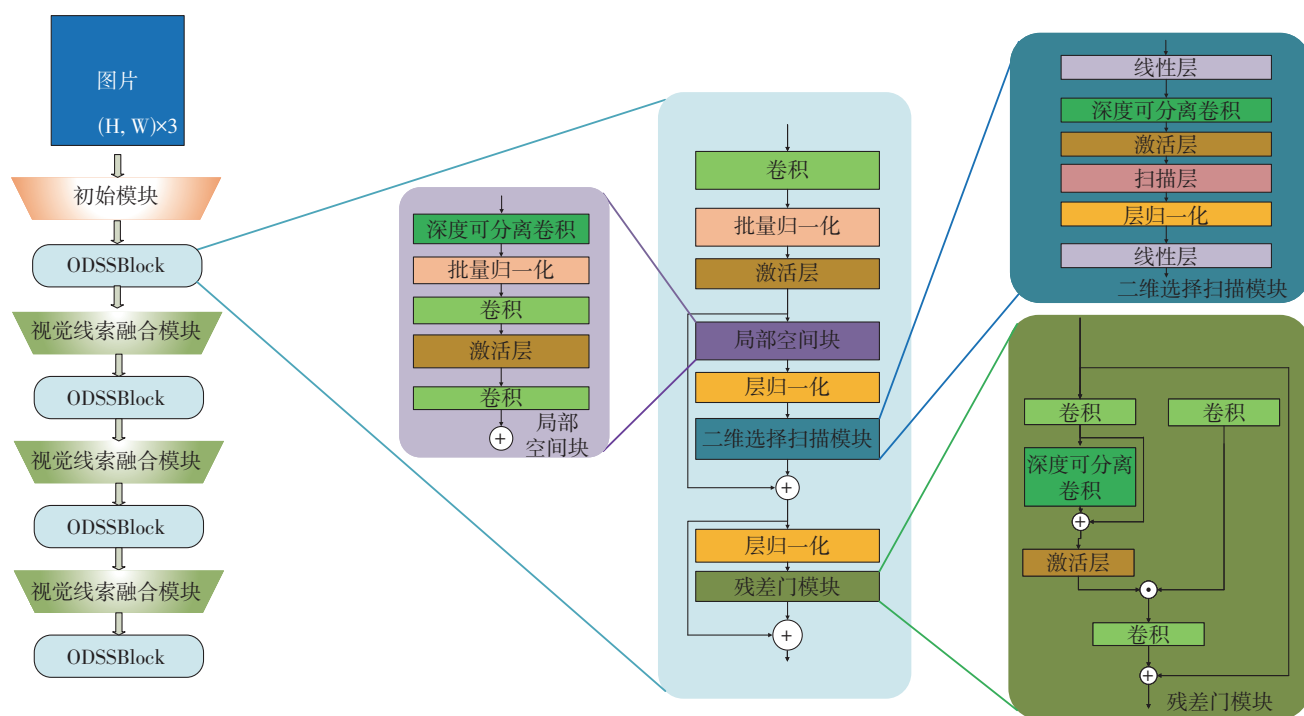


图2 ODMamba结构
Fig.2 ODMamba structure

依赖关系；在辅助路径中，局部空间模块(Local Spatial Block, LSB)利用深度可分离卷积(Depthwise Separable Convolutional, DSC)^[20]强化局部纹理特征提取，而残差门模块(Residual Gated Block, RGB)则通过门控机制对通道间的响应关系进行精细化建模，以弥补状态空间模型在细粒度空间定位上的不足。2条路径的输出经元素级相加后再与输入残差连接相叠加，使得ODSSBlock在保持实时推理效率的同时，显著增强了对多尺度目标尤其是小目标的特征判别能力。VCM是各特征提取阶段之间的过渡下采样层，起到特征的提炼与信息压缩。

2.2 GSConv改进编码器

RT-DETR的编码器结构在整体架构中扮演着连接主干特征提取网络编码器的桥梁作用。该部分的设计直接关系到特征的多尺度融合效果与编码器的输入质量，是RT-DETR实现高效与高精度并重的重要组件。RT-DETR的特征融合结构主要借鉴了特征金字塔(Feature Pyramid Network, FPN)^[21]和路径聚合特征金字塔网络(Path Aggregation FPN, PAFPN)^[22]的设计思想，能够将来自不同层级的特征图进行融合，使得最终输入到Transformer的特征具有更强的语义表达能力与空间分辨率。RT-DETR

从主干网络中提取多层次特征图，这些特征图具有不同的空间分辨率和语义丰富度。通过横向连接与上采样策略，将高层语义信息向下传播，与底层空间信息进行融合，以实现对小、中、大目标的全局建模能力。但特征融合过程中的卷积操作增加了RT-DETR的计算量，因此需要进行轻量化设计。目前的轻量化设计普遍采用DSC^[23]替代SC^[24]来减小参数量和计算成本，DSC与SC的核心差异在于空间特征提取与通道特征融合的解耦。SC在一个卷积核内同时执行所有输入通道的加权求和与空间邻域聚合，其参数量为 $C_{in} \times C_{out} \times k \times k$ ，其中 C_{in} 为输入通道数， C_{out} 为输出通道数， k 为卷积核大小，计算复杂度随通道数呈2次增长。而DSC将此过程分解为2个独立步骤：首先进行深度卷积(Depthwise Convolution)，每个输入通道单独使用1个 $k \times k$ 的卷积核进行空间滤波，该阶段仅聚合空间信息而不跨通道混合，参数量为 $C_{in} \times k \times k$ ；随后进行逐点卷积(Pointwise Convolution)，利用 1×1 卷积核在通道维度上进行线性重组，以生成指定数量的输出通道。这种解耦策略将总参数量降低至 $C_{in} \times k \times k + C_{in} \times C_{out}$ ，约为SC的 $1/C_{out} + 1/k^2$ (例如当 $k=3$ ， $C_{out}=256$ 时，可缩减至原来的1/8至1/9)，显著减少了模型参数与浮点

运算量。然而, 由于深度卷积对每个通道独立操作, 其通道间信息交互仅在逐点卷积阶段发生, 导致特征表达的联合优化能力弱于SC。SC结构如图3所示, DSC结构如图4所示。

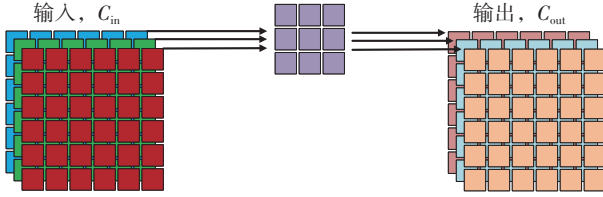


图3 SC结构

Fig.3 Standard convolution structure

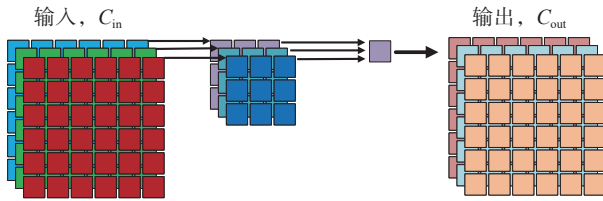


图4 DSC结构

Fig.4 Depth-wise separable convolutional structure

为了提高卷积网络的效率, Qin等^[25]提出的移动网络是通过 1×1 卷积融合通道信息, Chen等^[26]提出的混洗网络和Hou等^[27]提出的幻影网络, 则是利用通道混洗或廉价操作优化, 但这些方式要么计算量高, 要么性能仍未达到SC的水平。GSConv结合了SC和DSC的优点, 通过通道混合策略, GSConv的具体操作流程可概括为“分流-并行-拼接-混洗”4个有序步骤。首先, 对于输入特征图, GSConv在通道维度上将其划分为2个分支: 一个分支采用SC进行处理, 负责充分融合各输入通道间的空间与语义关联; 另一个分支则依次经过 1×1 点卷积和 3×3 DSC, 以极低的计算成本提取空间特征并调整通道数。完成2条路径的并行计算后, 将它们的输出在通道维度上进行拼接, 恢复至目标输出通道数。然而, 简单的拼接会导致SC生成的信息与DSC生成的信息在通道上呈区块化分布, 彼此缺乏交互, 为此GSConv引入了通道混洗操作, 通过将拼接后的通道均匀分组并重新排列, 使SC产生的丰富语义信息渗透至DSC的轻量通道中, 实现跨分支信息的充分交融。GSConv结构如图5所示。因此, 将RT-DETR编码器结构中的SC替换为GSConv, 从而优化模型的计算效率和特征表达能力, 降低RT-DETR的浮

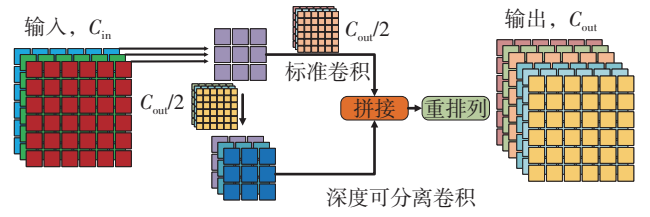


图5 GSConv结构

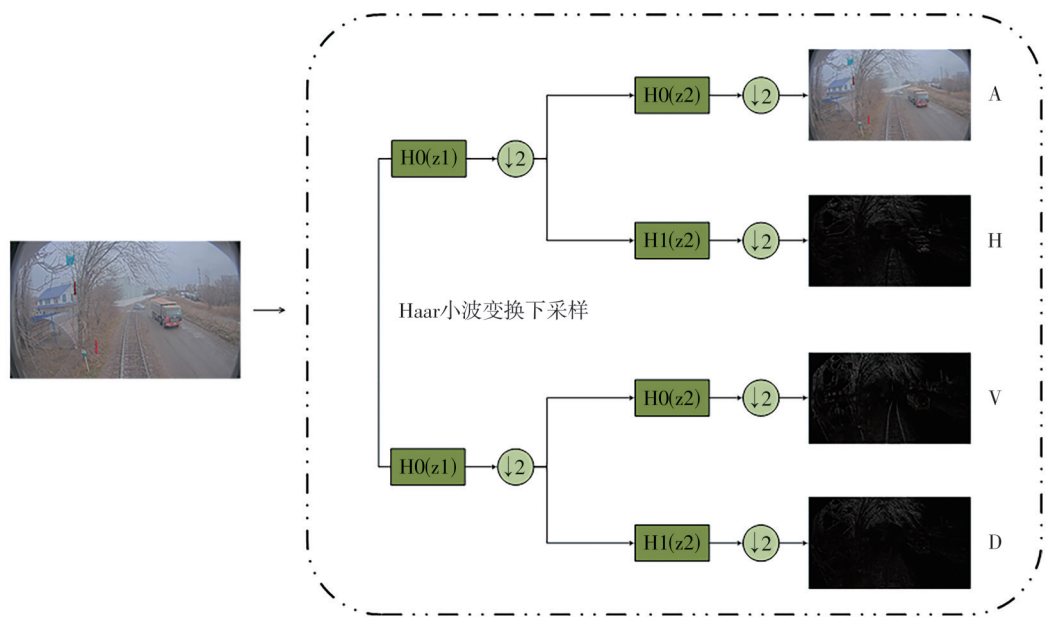
Fig.5 GSConv structure

点运算量。

2.3 特征图高保真压缩

RT-DETR中的编码器与解码器主要由Transformer层堆叠而成, Transformer层中核心的自注意力模块的计算复杂度为 $O(n^2)$ 。为了降低原始Transformer中自注意力机制的平方计算复杂度, 研究者提出了多种高效改进策略。Zhang等^[28]采用滑动窗口与全局稀疏注意力机制, 将全局交互限定于局部邻域及预设全局节点, 实现复杂度随序列长度的线性增长, 但其局限性在于感受野固定, 可能导致长距离依赖建模能力下降; Jin等^[29]利用低秩投影对键值矩阵进行降维处理, 通过低秩近似重构注意力分布, 使自注意力计算在保持全局建模能力的同时复杂度降至线性阶, 然而其依赖的核函数与低秩假设在复杂分布下容易引入累积误差。总体而言, 上述方法虽在理论上将复杂度降至接近 $O(n)$ 或 $O(n \log n)$, 但其各自引入的近似误差、信息损失或计算开销问题尚未得到根本性解决。HWD模块是一种基于HWD的下采样方法, 通过降低特征图的空间分辨率同时尽可能保留信息, 有效地将部分空间信息编码到通道维度, 从而实现信息的无损压缩, 基于HWD的图像压缩方法如图6所示。当对图像应用HWD时, 会产生4个分量, 每个分量的空间分辨率都是原始信号的一半。图6中HWD生成4个分量: 近似(低频)分量(A)以及水平(H)、垂直(V)和对角(D)方向的细节分量(高频)。每个部件的大小为 $H/2 \times W/2$, 每个分量的分辨率降低到 $H/2 \times W/2$, 而特征映射的通道数增加了4倍。

与最大池化下采样方法相比, HWD采样模块能够保留更多细节信息, 最大池化下采样与HWD采样后的结果对比如图7所示。针对图像中的树枝, HWD采样得到的图像细节显著优于最大池化下采样。因此, 本研究编码器与解码器之间加入HWD



注：H0和H1分别表示低通和高通分解滤波器，分别用于从图像中提取近似信息和高频信息；↓2表示应用于近似构件和详图构件的向下采样。

图6 基于HWD的图像压缩方法
Fig.6 HWD-based image compression method

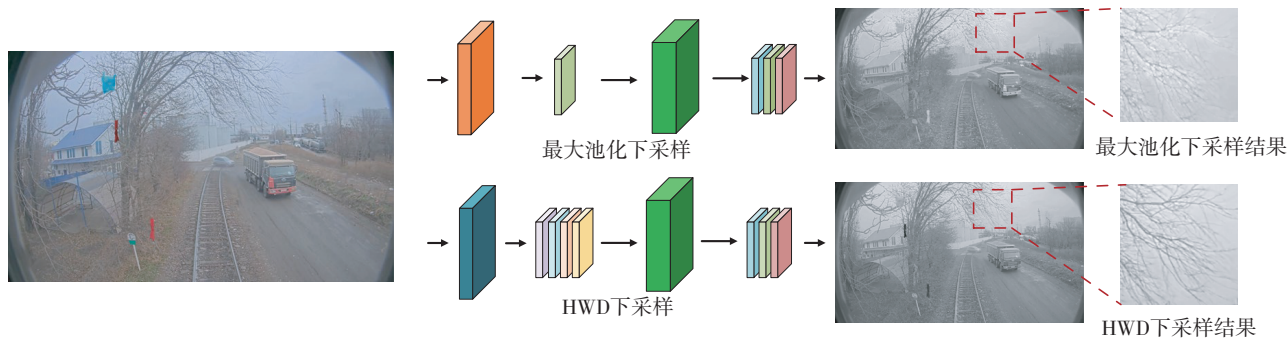


图7 最大池化下采样与HWD下采样后的结果对比
Fig.7 Comparison of results after max-pooling and HWD downsampling

采样模块，对特征图进行高保真压缩，从而在保持检测精度的前提下，显著降低模型的计算复杂度。

2.4 改进的RT-DETR模型

经过特征提取、特征融合和计算压缩3个层面进行优化，完成了对RT-DETR的改进工作。改进的RT-DETR模型结构如图8所示，主要从3方面对基线模型进行改进：首先，为了克服CNN难以建模长程时空依赖与自注意力计算开销过高的问题，使用ODMamba作为骨干网络，同时参数量降低至基线模型的约50%。其次，利用GSConv替代编码器中特征融合模块的SC，进一步降低模型的计算量与参数量。最后，为了降低解码器中自注意力的计算复杂度，通过HWD对编码器输出的特征图压

缩，模型的整体计算开销降低。

3 实验与分析

3.1 实验数据集

实验硬件软件配置如表1所示。实验中采用公开的铁路接触网异物检测数据集RailFOD23^[30]，包含14 615张高分辨率图像和40 541个标注对象，覆盖塑料袋、飘动物体、鸟巢和气球4类常见异物。各类标签数量及其分布如表2所示，鸟巢与飘动物体数量最多，气球最少，类别不平衡现象存在。该数据集由图像增强合成、AIGC生成、Photoshop合成3种来源构成，确保多样性与真实性。所有图片按照8:2的比例划分训练集与测试集，RailFOD23

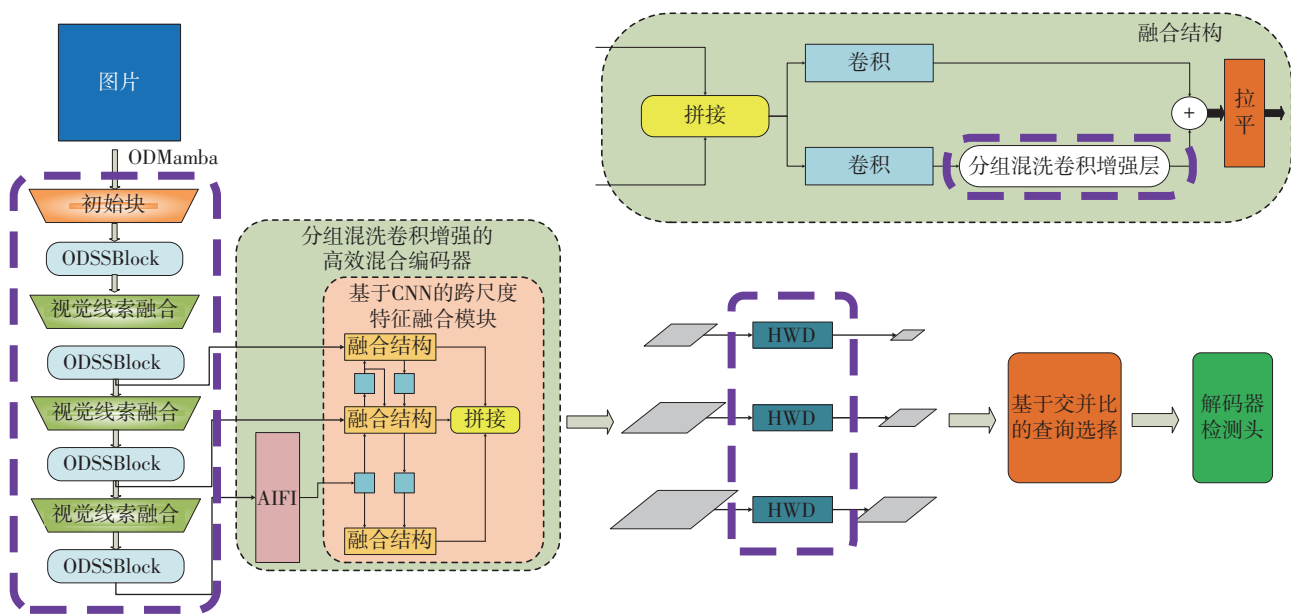


图8 改进的RT-DETR模型结构
Fig.8 Improved RT-DETR model structure

图像数据样本如图9所示。

表1 实验硬件软件配置

Tab.1 Experimental hardware and software configuration

参数	配置
系统	Ubuntu22.04 系统
底层框架	PyTorch2.1.2
处理器	16 vCPU Intel(R) Xeon(R) Gold 6430 2.10GHz
加速显卡	NVIDIA RTX 4090 GPU-24GB 单卡
CUDA 版本	11.8
模型构建框架	MMDetection3.0 ^[30]

表2 各类标签数量及其分布

Tab.2 Quantity and distribution of various labels

标签	标签数量	特点描述
塑料袋	4 561	易被风吹起, 缠绕电线
飘动物体	14 614	如风筝、横幅、布料等
鸟巢	15 795	筑巢于输电塔或电线上
气球	5 571	漂浮性强, 易造成接触网短路

3.2 评价标准

通过多个指标全面评估模型性能, 如平均精度均值(mean Average Precision, mAP)来衡量识别精度, 参数量(Params Size)评估模型的空间复杂度, 浮点运算次数(Floating Point Operations, FLOPs)评估模型的计算复杂度, 每秒帧数(Frames Per Second, FPS)评估模型的实际推理速度, 模型显存占用(Memory)评估模型运行中的实际显存占用。mAP结

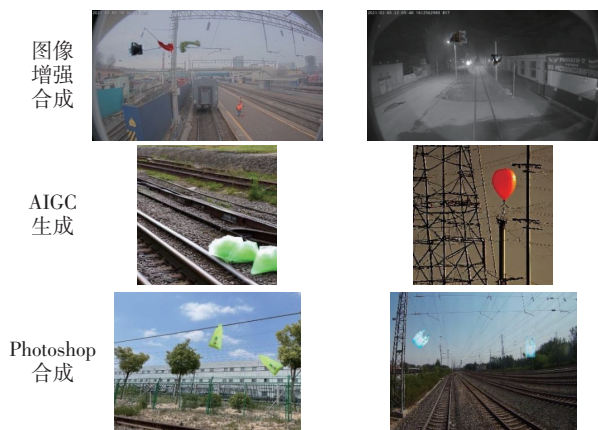


图9 RailFOD23图像数据样本

Fig.9 RailFOD23 image data samples

合了精确率(Precision)、召回率(Recall)和IoU的概念, 能够全面反映模型的准确性, 根据IoU 阈值不同分为 $mAP_{0.5}$, $mAP_{0.75}$ 和 $mAP_{0.5\sim0.95}$, 分别对应不同的精度要求。 $mAP_{0.5}$ 和 $mAP_{0.75}$ 对应IoU 阈值为0.5和0.75时的mAP。 $mAP_{0.5\sim0.95}$ 覆盖了0.5~0.95的IoU 阈值范围。AP是精度-召回率曲线(PRC)下的面积, 表示模型在召回率从0到1变化时的平均精度。 mAP 计算方法如下所示。

$$mAP = \frac{\sum_{k=1}^{N_{image}} P(k) \Delta r(k)}{m} \quad (1)$$

式中: N_{image} 为输入模型的图片数量; $P(k)$ 为模型识

别 k 个图像时的精度值； $\Delta r(k)$ 为图片数量从 $k-1$ 增加到 k 时召回率的变化值； m 为数据集类别个数。

3.3 消融实验

为了验证 RT-DETR 模型改进的有效性，对采用的改进方法进行消融实验，模型训练参数配置如表3所示。

表3 模型训练参数配置
Tab.3 Model training parameter configuration

参数	配置
训练批大小	16
训练轮数	108
基础学习率	0.000 1
优化器	Adamw
预热	前2 000次迭代
动量	0.000 1
权重衰减系数	0.000 1

基线模型训练损失变化情况如图10所示，在108个训练轮数内已收敛，因此设置训练轮数为108。消融实验结果如表4所示。将改进后模型命名为Mamba RT-DRTR。首先，将RT-DETR原网络的骨干网络改进为ODMamba，模型的数量降低了约38%，模型的FLOPs值降低了约43%，模型Memory值下降了约41%；其次，将编码器中特征融合模块中的SC替换为GSConv，模型的数量增加了约1%，模型的FLOPs值下降了约10%；最后，在编码器与解码器加入HWD模块对特征图压缩后送入编码器，模型的数量增加了约2%，模型的FLOPs值降低了约4%。当所有改进全部应用于RT-DETR时， $mAP_{0.5-0.95}$ 降低了0.4%， $mAP_{0.75}$ 与 $mAP_{0.5}$ 都有所提升，模型的数量降低了约40%，模型的FLOPs值降低了约58%，模型Memory值下降了约36%。在检测精度有所提升的基础上，模型参数量、FLOPs值、Memory值大幅下降，由此证明对模型所提出

的各项改进是有效的，每项改进对于提高模型的检测性能都有重要作用。

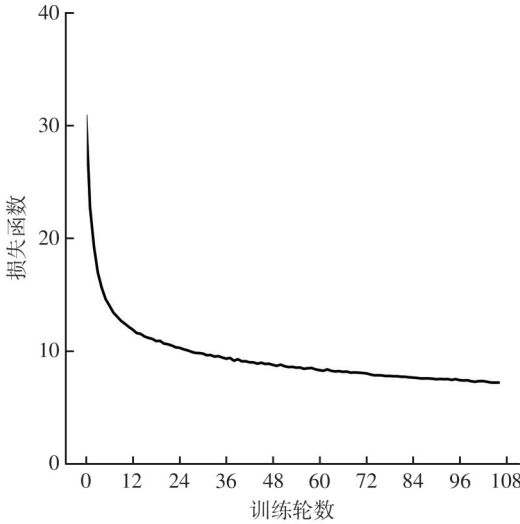


图10 基线模型训练损失变化情况
Fig.10 Training loss variation of baseline model

3.4 对比实验

为了验证 Mamba RT-DRTR 模型的性能，将其与Faster R-CNN，YOLOv4，TOOD，DOOD，DETR，Deformable DETR，DINO，DDQ等先进的实例分割模型进行对比。实验中所有模型均采用迁移学习的方法，特征提取网络预加载ImageNet1k数据集上的预训练参数。Faster R-CNN，TOOD，DOOD采用MMDetection3.0中的9X训练策略，YOLOv4采用默认训练参数，DETR，DAB DETR，DINO，DDQ将epoch均设置108，其他训练参数采用默认配置，模型对比实验结果如表5所示。

在检测精度上，Mamba RT-DRTR模型 $mAP_{0.5-0.95}$ 状态上比最优模型DDQ低0.8%，在 $mAP_{0.75}$ 状态上低0.3%，在 $mAP_{0.5}$ 状态上达到最优性能，为94.2%；在模型参数量方面，Mamba RT-DRTR模型的参数量约为DDQ的50%，计算量FLOPs约为DDQ的25%，

表4 消融实验结果
Tab.4 Results of ablation experiments

模型	$mAP_{0.5-0.95}/\%$	$mAP_{0.75}/\%$	$mAP_{0.5}/\%$	Params Size/MB	FLOPs/GFLOPs	FPS/(f/s)	Memory/MB
基线模型	84.8	87.8	93.6	42.769	168.000	46.4	174
ODMamba 骨干网络	84.9	87.5	93.7	26.17	95.799	47.0	102
GSConv	85.1	88.0	94.0	43.383	152.000	34.9	174
HWD	84.1	87.7	93.9	43.558	160.000	42.8	177
Mamba RT-DETR	84.4	88.1	94.2	25.388	71.411	47.1	111

表5 模型对比实验结果
Tab.5 Comparative experimental results of models

模型	mAP _{0.5-0.95} /%	mAP _{0.7} /%	mAP _{0.5} /%	Params Size/MB	FLOPs/GFLOPs	FPS/(f/s)	Memory/MB
基线模型 ^[11]	84.8	87.8	93.6	42.769	168.000	46.4	174
Faster R-CNN ^[4]	78.7	83.2	91.4	41.364	211.000	53.0	169
YOLOv4 ^[5]	83.8	87.1	93.2	44.105	154.000	83.7	168
YOLOv8 ^[5]	84.9	87.8	94.1	43.633	211.000	61.5	160
TOOD ^[6]	82.7	85.6	92.1	32.025	202.000	41.0	133
DDOD ^[7]	82.5	85.3	92.3	32.203	182.000	52.3	134
DETR ^[8]	76.8	83.0	91.0	41.556	99.803	58.6	170
DAB DETR ^[9]	79.6	85.0	93.5	43.702	105.000	39.1	178
DINO	84.3	87.1	93.8	47.546	284.000	23.5	193
DDQ ^[10]	85.2	88.4	94.1	48.317	284.000	18.8	227
Mamba RT-DETR	84.4	88.1	94.2	25.388	71.411	47.1	111

检测速度FPS达到实时水平，显存占用Memory达到最优。Mamba RT-DRTR模型在参数量和计算复杂度远低于DDQ的情况下，仍取得了与其接近的检测精度。在铁路侵限异物检测任务中，Mamba RT-DRTR模型在精度损失极小的前提下，参数量、计算复杂度与实际运行显存占用均优于多个主流模型，展现出良好的综合性能。

4 结束语

针对铁路侵限异物检测在复杂场景下存在的模型复杂度与检测精度难以平衡、边侧设备部署困难等现实挑战，构建的轻量化检测模型在显著降低参数量与计算开销的同时，有效保持了高精度检测能力，验证了状态空间建模与高效特征压缩在铁路安全检测领域的应用潜力。当前工作主要聚焦于可见光图像下的单一模态检测，未来将重点开展3方面研究：其一，结合自监督预训练策略以增强模型对稀有小目标异物的表征学习能力；其二，推进模型量化与结构化剪枝的联合优化，构建面向嵌入式平台的轻量化部署方案，支撑无人机巡检与车载实时预警等多场景应用；其三，探索可见光与红外、激光雷达等多源传感信息的深度融合机制，提升模型在夜间、雨雾等恶劣气象条件下的感知鲁棒性，最终形成覆盖全天候、全场景的铁路侵限智能检测体系。

参考文献：

[1] FENG Z C, YANG J, CHEN Z C, et al. LRseg: An EFFICIENT RAILWAY REGION EXTRACTION METHOD

BASED on Lightweight Encoder and Self-Correcting Decoder[J]. Expert Systems with Applications, 2024, 238: 122386.

[2] 杨文, 胡昊, 李凌志, 等. 基于机器视觉的铁路限界入侵检测方法[J]. 铁道科学与工程学报, 2025(3): 1328-1343.

[3] 栗文韬, 徐成伟, 张万鹏, 等. 基于激光雷达+视频的城际铁路周界入侵监测报警技术研究[J]. 铁道运输与经济, 2023, 45(7): 134-142.

[4] GUO S Q, BAI Q L, ZHOU X X. Foreign Object Detection of Transmission Lines Based on Faster R-CNN[C]//Information Science and Applications. Singapore: Springer, 2020: 269-275.

[5] SONG Y H, ZHOU Z Z, LI Q, et al. Intrusion Detection of Foreign Objects in High-Voltage Lines Based on YOLOv4[C]//2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP). Xi'an: IEEE, 2021: 1295-1300.

[6] FENG C J, ZHONG Y J, GAO Y, et al. TOOD: Task-Aligned One-Stage Object Detection[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2022: 3490-3499.

[7] CHEN Z H, YANG C, LI Q F, et al. Disentangle Your Dense Object Detector[C]//Proceedings of the 29th ACM International Conference on Multimedia. New York: ACM, 2021: 4939-4948.

[8] CARION N, MASSA F, SYNNAEVE G, et al. End-to-End Object Detection with Transformers[C]//Computer Vision-ECCV 2020. Cham: Springer, 2020: 213-229.

[9] YANG J C, ZHANG H, ZHOU Y T, et al. Improved DAB-DETR Model for Irregular Traffic Obstacles Detection in Vision

- Based Driving Environment Perception Scenario[J]. Applied Intelligence, 2025, 55(7): 541.
- [10] ZHANG S L, WANG X J, WANG J Q, et al. Dense Distinct Query for End-to-end Object Detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver: IEEE, 2023: 7329–7338.
- [11] ZHAO Y A, LV W Y, XU S L, et al. DETRs Beat YOLOs on Real-Time Object Detection[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2024: 16965–16974.
- [12] YUAN C, LI S C, WANG K, et al. Mamba-YOLO-ML: A State-Space Model-Based Approach for Mulberry Leaf Disease Detection[J]. Plants, 2025, 14(13): 2084.
- [13] 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.
- [14] XU G P, LIAO W T, ZHANG X, et al. Haar Wavelet Downsampling: A Simple but Effective Downsampling Module for Semantic Segmentation[J]. Pattern Recognition, 2023, 143: 109819.
- [15] ZOU Z X, CHEN K Y, SHI Z W, et al. Object Detection in 20 Years: A Survey[J]. Proceedings of the IEEE, 2023, 111(3): 257–276.
- [16] YANG J Q, YANG X, LIAO Y Z, et al. Multispectral Sample Augmentation and Illumination Guidance for RGB-T Object Detection by mm Detection Framework[C]//4th International Conference on Laser, Optics, and Optoelectronic Technology (LOPET 2024). Chongqing: SPIE, 2024: 69.
- [17] HE K M, ZHANG X Y, REN S Q, et al. Deep Residual Learning for Image Recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016: 770–778.
- [18] HAN K, WANG Y H, CHEN H T, et al. A Survey on Vision Transformer[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(1): 87–110.
- [19] GU A, DAO T. Mamba: Linear-Time Sequence Modeling with Selective State Spaces[J]. 2024, 47(1): 88–99.
- [20] LU G Z, ZHANG W Z, WANG Z. Optimizing Depthwise Separable Convolution Operations on GPUs[J]. IEEE Transactions on Parallel and Distributed Systems, 2022, 33(1): 70–87.
- [21] GONG Y Q, YU X H, DING Y, et al. Effective Fusion Factor in FPN for Tiny Object Detection[C]//2021 IEEE Winter Conference on Applications of Computer Vision (WACV). Waikoloa: IEEE, 2021: 1159–1167.
- [22] PAN B, XIANG J, ZHANG N, et al. A Fine-Grained Attributes Recognition Model for Clothing Based on Improved the CSPDarknet and PAFPN Network[J]. Signal, Image and Video Processing, 2025, 19(3): 230.
- [23] LIU F C, XU H, QI M, et al. Depth-Wise Separable Convolution Attention Module for Garbage Image Classification[J]. Sustainability, 2022, 14(5): 3099.
- [24] KHAKI S, PHAM H, KHALILZADEH Z, et al. High-Throughput Image-Based Plant Stand Count Estimation Using Convolutional Neural Networks[J]. PLoS One, 2022, 17(7): e0268762.
- [25] QIN D F, LEICHER C, DELAKIS M, et al. MobileNetV4: Universal Models for the Mobile Ecosystem[C]//Computer Vision–ECCV 2024. Cham: Springer, 2025: 78–96.
- [26] CHEN Z C, YANG J, CHEN L F, et al. Garbage Classification System Based on Improved ShuffleNet V2[J]. Resources, Conservation and Recycling, 2022, 178: 106090.
- [27] HOU H T, GUO M Z, WANG W, et al. Improved Lightweight Head Detection Based on GhostNet-SSD[J]. Neural Processing Letters, 2024, 56(2): 126.
- [28] ZHANG P C, DAI X Y, YANG J W, et al. Multi-Scale Vision Longformer: A New Vision Transformer for High-Resolution Image Encoding[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2022: 2978–2988.
- [29] JIN Y L, WU Y F, GAO Y, et al. LinFormer: A Linear-Based Lightweight Transformer Architecture for Time-Aware MIMO Channel Prediction[J]. IEEE Transactions on Wireless Communications, 2025, 24(9): 7177–7190.
- [30] CHEN Z C, YANG J, FENG Z C, et al. RailFOD23: A Dataset for Foreign Object Detection on Railroad Transmission Lines[J]. Scientific Data, 2024, 11: 72.

(责任编辑 刘宁馨)