



## 通道混洗与跨尺度增强的轻量级铁路全景分割

陈永 周方春 周建宇

### Lightweight railway panoramic segmentation based on channel shuffle and cross-scale enhancement

CHEN Yong, ZHOU Fangchun, ZHOU Jianyu

引用本文:

陈永, 周方春, 周建宇. 通道混洗与跨尺度增强的轻量级铁路全景分割[J]. *北科大: 工程科学学报*, 2025, 47(8): 1722–1730. doi: 10.13374/j.issn2095–9389.2024.10.23.001

CHEN Yong, ZHOU Fangchun, ZHOU Jianyu. Lightweight railway panoramic segmentation based on channel shuffle and cross-scale enhancement[J]. *Chinese Journal of Engineering*, 2025, 47(8): 1722–1730. doi: 10.13374/j.issn2095–9389.2024.10.23.001

在线阅读 View online: <https://doi.org/10.13374/j.issn2095–9389.2024.10.23.001>

## 您可能感兴趣的其他文章

### Articles you may be interested in

#### 一种基于轻量级神经网络的高铁轮对轴承故障诊断方法

Fault diagnosis of high-speed train wheelset bearing based on a lightweight neural network

工程科学学报. 2021, 43(11): 1482 <https://doi.org/10.13374/j.issn2095–9389.2020.12.09.001>

#### 结合多尺度分割和随机森林的变质矿物提取

Extraction of metamorphic minerals by multiscale segmentation combined with random forest

工程科学学报. 2022, 44(2): 170 <https://doi.org/10.13374/j.issn2095–9389.2020.09.08.004>

#### 基于SIMP安定的高速列车组合式座椅底架轻量化设计与分析

Lightweight design and analysis of a combined seat bracket for a high-speed train based on SIMPshakedown

工程科学学报. 2024, 46(6): 1151 <https://doi.org/10.13374/j.issn2095–9389.2023.05.05.002>

#### 一种轻量型人体行为识别学习模型

Lightweight human activity recognition learning model

工程科学学报. 2022, 44(6): 1072 <https://doi.org/10.13374/j.issn2095–9389.2021.03.18.001>

#### DS-TransFusion: 基于改进Swin Transformer的视网膜血管自动分割

DS-TransFusion: Automatic retinal vessel segmentation based on an improved Swin Transformer

工程科学学报. 2024, 46(10): 1889 <https://doi.org/10.13374/j.issn2095–9389.2023.06.27.004>

#### 时速350 km高速列车用铜基闸片材料的摩擦性能

Friction performance of Cu-based brake pad for high-speed railway trains

工程科学学报. 2023, 45(3): 389 <https://doi.org/10.13374/j.issn2095–9389.2021.10.20.003>

# 通道混洗与跨尺度增强的轻量级铁路全景分割

陈永<sup>1,2)</sup>✉, 周方春<sup>1)</sup>, 周建宇<sup>1)</sup>

1) 兰州交通大学电子与信息工程学院, 兰州 730070 2) 甘肃省人工智能与图形图像处理工程研究中心, 兰州 730070  
✉通信作者, E-mail: [edukeylab@126.com](mailto:edukeylab@126.com)

**摘要** 针对高速铁路场景下图像全景分割时存在全景分割精确度低, 难以实现轻量级实时分割等问题, 提出了一种通道混洗与跨尺度增强的轻量级铁路全景分割方法. 首先, 基于 FasterNet 网络, 提出了轻量化 CS\_FasterNet 特征提取网络, 通过部分卷积和通道混洗优化了特征信息的聚合, 实现对铁路场景下全景分割轻量化特征提取. 其次, 设计了多尺度特征交互增强模块, 利用特征交互和跨特征融合, 全面地捕捉局部的细节和全局信息, 提高图像特征提取的质量. 最后, 改进预测融合模块对语义结果与实例结果进行融合, 提升网络对图像分割的准确性, 得到更加精细的全景分割输出结果. 实验结果表明: 所提轻量级模型在模型每秒处理帧率和计算量等评价指标均优于对比方法, 相较于 UPSNet 方法, 本文方法的每秒约处理 11.5 帧, 全景分割质量提升了约 9.9%, 能够实现对不同铁路场景下图像全景分割的准确性和实时性.

**关键词** 全景分割; 高速铁路; 轻量级; 通道混洗; 跨尺度特征增强

**分类号** TP391.4

## Lightweight railway panoramic segmentation based on channel shuffle and cross-scale enhancement

CHEN Yong<sup>1,2)</sup>✉, ZHOU Fangchun<sup>1)</sup>, ZHOU Jianyu<sup>1)</sup>

1) School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China  
2) Gansu Artificial Intelligence and Graphics and Image Processing Engineering Research Center, Lanzhou 730070, China  
✉Corresponding author, E-mail: [edukeylab@126.com](mailto:edukeylab@126.com)

**ABSTRACT** A lightweight railway panoramic segmentation method based on channel mixing and cross-scale enhancement was proposed to address the challenges of low accuracy and difficulty of achieving lightweight real-time panoramic segmentation in high-speed railway scene images. The model consists of three main components: a lightweight CS\_FasterNet feature extraction network, multi-scale feature interaction enhancement module, and prediction fusion output module. First, based on the FasterNet network, a lightweight CS\_FasterNet feature extraction network was proposed. FasterNet reduces redundant calculations through partial convolution to enhance processing speed while preserving high detection accuracy. However, the original design applies filters to only a portion of the input channels, potentially limiting feature extraction for the remaining channels. This limitation was addressed by optimizing the aggregation of local and global feature information through partial convolution and channel shuffling, combined with feature recombination techniques to reduce computational complexity and improve feature extraction in lightweight railway scenes. Second, based on the completion of the lightweight CS\_FasterNet feature extraction, a multiscale feature interaction enhancement module was designed to improve the network's image segmentation ability and enhance the representation of features. This module consists of an attention-based intrascale feature interaction module and a cross-scale feature fusion module. The attention-based intrascale feature interaction module applies a multihead attention mechanism to extract pixel-level semantic features from high-level

收稿日期: 2024-10-23

基金项目: 国家自然科学基金资助项目(62462043, 61963023); 兰州交通大学重点研发项目资助(ZDYF2304)

image features, expanding the receptive field and capturing fine-grained information. The cross-scale feature fusion module adopts both bottom-up and top-down fusion paths to integrate feature maps of different scale outputs from the backbone network, improving scale feature utilization and enabling comprehensive extraction of local details and global information. Finally, the prediction fusion module was refined to integrate the semantic and instance results. In the panoramic segmentation task, the Soft NMS method was used to improve the accuracy of pixel classification. Soft NMS reduces confidence scores for detection boxes based on the intersection-to-union ratio and uses a Gaussian weighted score to identify the true detection box, leading to improved segmentation accuracy and more refined panoramic segmentation outputs. Experimental results indicate that the proposed lightweight model excels in frame rate per second and computational complexity. A higher frame rate indicates faster segmentation speed, and lower computational complexity favors lightweight segmentation. In this model, evaluation metrics, such as the processing frame rate per second and computational complexity, outperformed the comparison methods. Compared with the UPSNet method, this method increased the processing frame rate by approximately 11.5 frames per second and improved the quality of the panoramic segmentation by approximately 9.9%. This method achieves accurate, real-time panoramic segmentation across various railway scenarios.

**KEY WORDS** panoramic segmentation; high-speed railway; lightweight; channel shuffle; cross-scale feature enhancement

随着高速铁路智能化发展, 列车安全运行备受关注<sup>[1]</sup>. 由于列车运行时速度不断地提高和交通路网的不断扩张, 行人、车辆等异物侵入铁路限界, 极易造成安全事故, 危害极大<sup>[2]</sup>. 实现对铁路全景感知对于我国智能高铁的发展至关重要.

全景分割 (Panoptic segmentation, PS)<sup>[3]</sup> 是通过将图像中区域进行语义类别的划分, 为每个事物赋予其类别标签和实例标识的技术. 全景分割对于智能交通、无人驾驶、医学等领域<sup>[4]</sup> 及高速铁路<sup>[5]</sup> 具有重要的意义. 通过全景分割能将场景分割成不同的语义区域, 增强对于交通异物的感知能力, 提升车辆运行的安全性和智能性<sup>[6]</sup>, 为智能驾驶提供了更加可靠的决策基础.

随着深度学习的推广和发展, 图像全景分割逐渐成为热点研究问题. Kirillov 等<sup>[7]</sup> 提出了全景分割概念, 将实例分割网络<sup>[8]</sup> 和语义分割网络金字塔<sup>[9]</sup> 结合得到全景分割结果, 为后续全景分割的研究奠定了基准. Geus 等<sup>[10]</sup> 提出了统一的全景分割端到端的网络框架 (Joint semantic and instance segmentation network, JSIS-Net), 通过引入公共共享特征提取网络来降低网络的计算量, 但该方法存在特征提取网络结构单一导致特征提取不充分和分割精度低的问题. Xiong 等<sup>[11]</sup> 提出全景融合机制 (Unified panoptic segmentation network, UPSNet) 的全景分割方法, 通过全卷积网络<sup>[12]</sup> 和实例分割结合来提高全景分割性能, 但忽略了对全景场景中无定形区的关注, 易出现对无定形目标分割结果不够准确的问题. Chen 等<sup>[13]</sup> 提出了双路空间金字塔和残差网络结合的 Panoptic-DeepLab 方法, 但该方法网络结构复杂度较高, 存在实时性差的问题. Chen 等<sup>[14]</sup> 提出了一种信息聚合的网络 (Bidirectional

aggregation network, BANet), 利用语义分割和实例分割间特征交互来提升全景分割效果, 但在全景融合时会出现冲突和重复的问题, 影响全景分割精度. Liu 等<sup>[15]</sup> 提出采用空间排序模块来提高分割精度, 但该方法基于 Mask R-CNN 仍采用 RoIAlign 的方法, 在像素级运算中仍然存在精度损失的问题. Fu 等<sup>[16]</sup> 提出一种多尺度注意力引导的全景分割网络 MAPSNet, 利用引导卷积核生成注意力模块, 提升物体分割准确率, 但该模型存在复杂度高, 实时性较差的问题.

综上所述, 针对现有高速铁路场景下图像全景分割时仍存在全景分割精确度低, 难以实现轻量级实时分割等问题, 提出了一种基于通道混洗与跨尺度增强的轻量级铁路全景分割方法. 主要工作包括: (1) 基于 FasterNet 特征提取网络, 提出了轻量化的 CS\_FasterNet 特征提取网络, 通过部分卷积和通道混洗优化了特征信息的聚合, 实现了对铁路全景分割特征提取的轻量化. (2) 提出了多尺度特征交互增强模块, 通过特征交互和跨特征融合, 增强对特征的代表能力, 提高全景分割的特征提取质量. (3) 改进预测融合模块对语义分割结果与实例分割结果进行融合, 提升网络对图像分割的准确性, 实现对铁路场景下图像的实时全景分割.

## 1 所提方法

### 1.1 网络整体架构

本文提出了一种基于通道混洗与跨尺度增强的轻量级铁路全景分割方法, 该模型主要由: ①轻量化 CS\_FasterNet 特征提取网络、②多尺度特征交互增强模块和③预测融合输出模块三部分构成. 其整体网络结构如图 1 所示. 模型工作时, 首先采

用提出的轻量 CS\_FasterNet 特征提取网络, 通过部分卷积和通道混洗优化了特征信息的聚合, 实现轻量化特征提取. 然后, 设计多尺度特征交互增强模块, 捕获更丰富的上下文信息, 进一步提高图像

特征提取质量. 接着, 通过实例分割分支模块和语义分割分支模块获取图像的实例信息和语义信息, 最后, 将两分支预测结果进行预测融合, 从而完成对铁路场景下图像的轻量化全景分割.

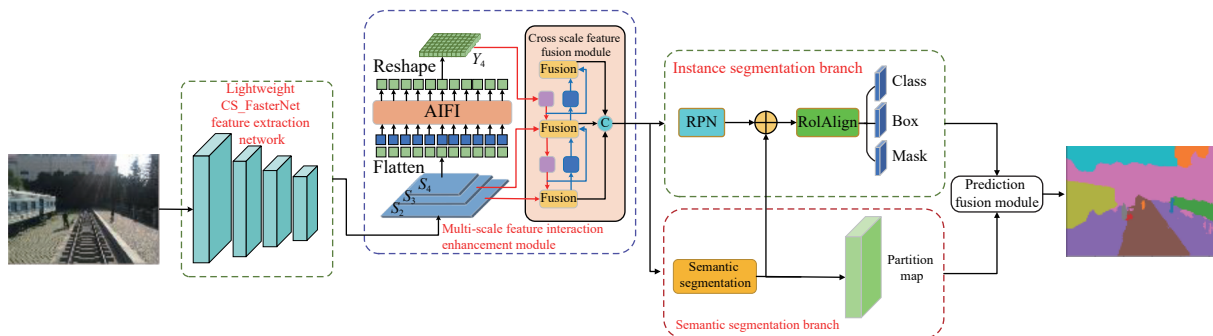


图 1 网络整体结构

Fig.1 Overall network structure

### 1.2 轻量 CS\_FasterNet 特征提取网络

在图 1 中, 针对现有铁路场景下全景分割网络中, 存在网络模型计算量较大, 实时性差的问题, 本文在原始 FasterNet<sup>[17]</sup> 网络的基础上, 提出了一种轻量 CS\_FasterNet 特征提取网络. 原始 FasterNet 是一种通过部分卷积 (Partial convolution, PConv) 来减少冗余计算的特征提取网络, 旨在提高计算速度并保留较高的检测准确性. 但部分卷积仅对输入部分通道应用滤波器, 易造成其余通道特征提取不充分的问题<sup>[18]</sup>. 鉴于以上 FasterNet 通道操作的缺点, 本文提出了一种轻量 CS\_FasterNet 特征提取网络, 通过部分卷积和通道混洗 (Channel

shuffle) 优化局部和全局特征信息的聚合, 使用特征重组技术, 减小模型的计算量, 从而达到轻量化铁路场景下的特征提取的目的.

本文提出的轻量化 CS\_FasterNet 网络, 如图 2 所示, 主要由四个 CS\_FasterNet 模块组成, 每个模块中  $C_i$  表示每个阶段的通道数. 四个 CS\_FasterNet 模块对全景分割铁路图像进行不同尺度特征提取, 每个层级前面均有一个卷积为  $4 \times 4$ 、步幅为 4 的嵌入层和卷积为  $2 \times 2$ 、步幅为 2 的合并层, 用于空间的下采样和扩展通道数. 最后使用全局平均池化层、卷积层和全连接层进行特征转换和分类.

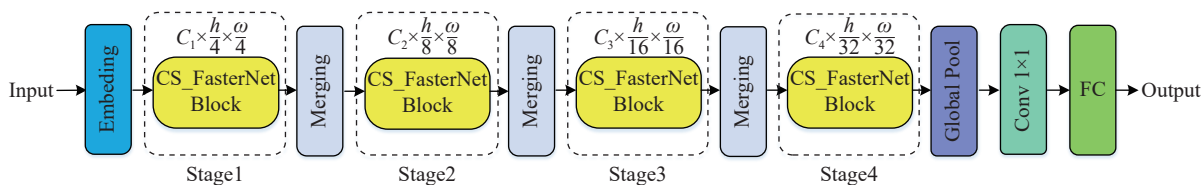


图 2 轻量 CS\_FasterNet 特征提取网络结构图

Fig.2 Structure of the lightweight CS\_FasterNet feature extraction network

图 2 中, 每个 CS\_FasterNet 模块由部分卷积和通道混洗相结合的思想进行设计, 结构图如图 3 所示. 首先, 通过分组卷积 (Group convolution, GConv) 对不同特征图进行分组, 接着使用的部分卷积 PConv 计算特征图来减少计算中的冗余信息, 同时加入通道混洗机制来增加不同通道间的特征交互, 来克服铁路全景分割特征提取不充分的问题<sup>[19]</sup>. 最后, 再次加入分组卷积减少资源消耗过大的问题, 可进一步提升模型效率, 并与全局平均池化层

进行相加操作.

图 3 中部分卷积利用特征映射中的冗余性, 避免了对所有输入通道进行全面卷积计算<sup>[17]</sup>, 其浮

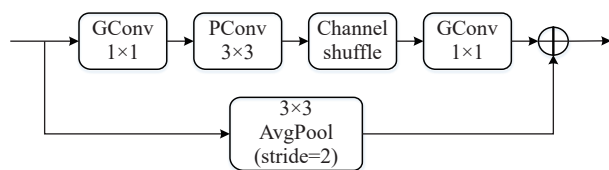


图 3 CS\_FasterNet Block 结构图

Fig.3 Structure of the CS\_FasterNet Block

点运算次数如式 (1) 所示.

$$FLOPs = h \times w \times k^2 \times c_p^2 \quad (1)$$

式中,  $c$  表示通道数;  $h$  为特征图高,  $w$  为特征图宽,  $k$  表示卷积核的大小. 对于典型的缩减因子  $r$ , 设定只对其中  $1/4$  的通道进行部分卷积, 即  $r = c_p/c = 1/4$  时, PConv 的 FLOPs 只有常规卷积的  $1/16$ .

在特征提取骨干网络中加入通道混洗<sup>[20]</sup>, 可以增强通道间的特征交互, 提升网络模型性能, 将不同组中进行统一的通道交换, 降低计算的复杂度. 通道混洗还可以改变输入特征图不同通道之间的依赖关系<sup>[21]</sup>, 提高特征的提取能力.

### 1.3 多尺度特征交互增强模块

在完成轻量化 CS\_FasterNet 特征提取的基础上, 为了提升网络的图像分割能力, 增强对特征的表达, 进一步设计了铁路全景分割多尺度特征交互增强模块, 该模块由①基于注意力的尺度内特征交互模块和②跨尺度特征融合模块两部分组成.

基于注意力的尺度内特征交互模块, 结构如图 4 所示, 该模块应用多头注意力机制<sup>[22]</sup> 从高层级语义特征中进一步提取图像的像素级语义特征信息, 扩大感受野范围<sup>[23]</sup>, 有助于捕获更细粒度的信息. 对于主干网络的最后输出二维特征  $S_4$ , 将其展平为列向量, 然后对其进行多头注意力操作, 完成后重组为二维特征  $Y_4$ , 该过程如式 (2) 和式 (3) 所示:

$$Q = K = V = Flatten(S_4) \quad (2)$$

$$Y_4 = Reshape(Attention(Q, K, V)) \quad (3)$$

式中,  $Q$  为查询向量,  $K$  为键向量,  $V$  为值向量, Flatten 为展平操作, 通过该操作可以将输入特征图拉为列向量; Reshape 为重组操作, 通过该操作将列

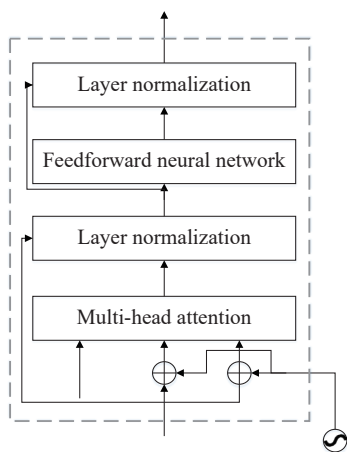


图 4 基于注意力的尺度内特征交互模块

Fig.4 Attention-based feature interaction module within scale

向量还原为二维特征图.

跨尺度特征融合模块采用自底向上和自顶向下两条路径的特征融合, 通过将主干网络输出的不同尺度  $S_2$ 、 $S_3$  和  $Y_4$  进行融合来提高尺度特征的利用率, 如式 (4) 所示:

$$Output = CCFM(\{S_2, S_3, Y_4\}) \quad (4)$$

式中, Output 为输出的新特征; CCFM 为跨尺度特征融合模块操作, 具体融合操作如下图 5 所示. 首先将两个输出特征作为输入进行拼接, 然后分为两路经过  $1 \times 1$  卷积和可重参数化模块, 最后通过逐元素加法进行融合重组后, 产生新的融合后特征.

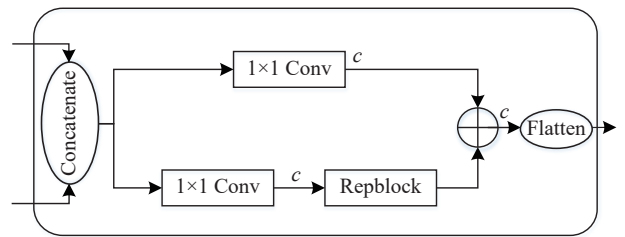


图 5 跨尺度特征融合模块内部结构

Fig.5 Internal structure of the cross-scale feature fusion module

为了直观验证所提多尺度特征交互增强模块的有效性, 下面进行可视化比较, 热力图结果如图 6 所示. 图 6(b) 为改进前特征信息图, 可以看出其无法对目标进行准确聚焦, 极易受到背景杂波的影响, 如图 6(b) 中箭头所指的钢轨中间行人未能有效聚焦. 图 6(c) 为使用本文多尺度特征交互增强模块改进后所获取的特征图, 从结果可以看出能够有效抑制背景杂波的干扰, 能够对目标进行更加准确的关注, 获取了高质量特征提取信息.

### 1.4 预测融合模块

最后, 在完成多尺度特征交互增强后, 进行预测输出分割结果, 其模块结构如图 7 所示.

在全景分割结果中, 图像经过实例分割后会出现同一个像素点可能被多个实例或者类别同时覆盖结果的情况, 影响分割的精度. 为了克服上述缺点, 本文在全景分割任务中, 采用 Soft-NMS<sup>[24]</sup> 方法来提高像素类别划分的准确性, Soft-NMS 通过减小交并比大于检测框的置信度来完成检测框的选取, 将获得的交并比数值取高斯指数后与原得分相乘, 最后通过加权的方式选出真正的检测框, 计算公式如下.

$$S_i = S_i \exp \left[ -\frac{IoU(B, b_i)^2}{\delta} \right] \quad (5)$$

式中, IoU 为交并比;  $B$  为置信度最高的检测框;  $b_i$  为第  $i$  个检测框;  $S_i$  为第  $i$  个检测框置信度. IoU 的

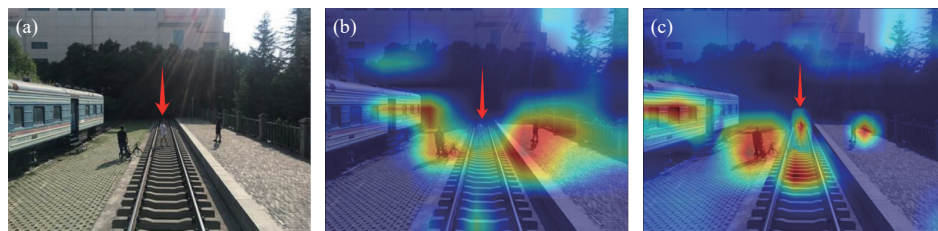


图 6 热力图可视化实验。(a) 铁路场景图像; (b) 改进前; (c) 本文改进后

Fig.6 Heatmap visualization experiment: (a) railway scene image; (b) before improvement; (c) after improvement

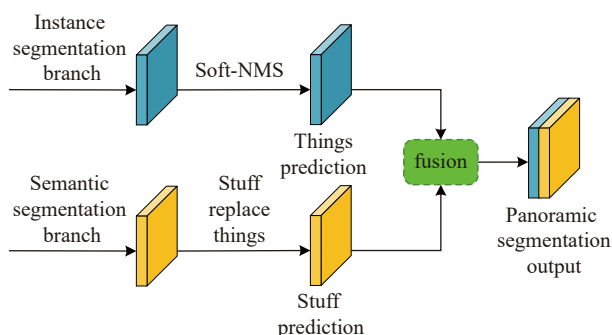


图 7 全景预测融合模块

Fig.7 Panoramic-prediction fusion module

值小于 1, IoU 的数值越大, 对其取指数运算后所得数值越小,  $\delta$  为超参数, 用于控制衰减的速度. 再与  $S_i$  相乘后使得到的检测框分数越低, 经过不断地循环, 最终筛选出所需要的目标框, 输出全景分割结果.

### 1.5 模型损失函数

本文损失函数包括语义分割损失、实例中心点损失和实例中心偏移损失 3 个部分, 公式如下:

$$L = \lambda_{\text{semantic}} L_{\text{semantic}} + \lambda_{\text{center}} L_{\text{center}} + \lambda_{\text{offset}} L_{\text{offset}} \quad (6)$$

语义分割损失采用加权交叉熵损失函数<sup>[25]</sup>, 公式如下所示

$$L_{\text{semantic}} = -\frac{1}{K} \sum_{i=1}^N \mathbb{C}[P_{i,y_i} < t_K] \cdot \ln P_{i,y_i} \quad (7)$$

式中,  $K = 0.15 \times N$ ,  $N$  为图像中总像素点个数,  $K$  的值采用向上取整的方式保证值为整数, 通过向上取整保留足够数量的像素点用于语义分割操作, 避免因像素点过少导致重要信息丢失的问题.  $y_i$  为像素  $i$  的真实标签,  $P_{i,y_i}$  为像素  $i$  属于类  $y_i$  的可能性. 对于  $\mathbb{C}[x]$ , 若  $x$  为真, 则  $\mathbb{C}[x]=1$ , 否则为 0,  $t_K$  为超参数.

实例中心点损失计算如下:

$$L_{\text{center}} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (8)$$

式中,  $N$  是样本数量,  $y_i$  是第  $i$  个样本真值,  $\hat{y}_i$  是第  $i$  个样本预测值.

实例中心偏移损失使用 SmoothL1 Loss 损失函数<sup>[26]</sup>, 公式如下所示

$$L_{\text{offset}}(x) = \begin{cases} \frac{1}{2}x^2, & |x| \leq 1 \\ |x| - \frac{1}{2}, & \text{otherwise} \end{cases} \quad (9)$$

式中,  $x$  是预测值与真实值之间的差值.

## 2 实验结果分析

### 2.1 实验数据与环境配置

数据集采用 Microsoft COCO<sup>[27]</sup> 数据集和铁路场景下现场拍摄的铁路相关图像进行实验, 通过数据增强扩充后, 并通过 LabelImg 工具标注, 得到包含 20000 张的全景铁路分割数据集, 其中 70% 为训练集, 30% 为测试集. 软硬件配置环境为: PyTorch, Windows 11, Intel(R) Core i7-10700, NVIDIA GeForce GTX 3060. 为了验证所提方法的有效性, 选择与文献<sup>[11]</sup>的 UPSNet 方法、文献<sup>[13]</sup>的 Panoptic-DeepLab 方法、文献<sup>[16]</sup>的 MAPSNet 方法和文献<sup>[28]</sup>的 MP-Former 方法进行对比分析.

### 2.2 评价指标

定量性能采用全景质量 PQ、平均精度 AP 以及平均交并比 mIoU 进行衡量. 其中 PQ 定义如下

$$PQ = \frac{\sum_{(p,q) \in TP} \text{IoU}(p,q)}{|TP|} \times \frac{|TP|}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|} \quad (10)$$

式中,  $\text{IoU}(p,q)$  表示两者之间的交集,  $p,q$  分别为预测值和真值; TP 为正样本数; FN 为漏报正样本数; FP 为误报负样本数.

AP 为实例的平均精度, 计算如下:

$$AP = \int_0^1 p(r) dr \quad (11)$$

式中,  $p$  表示准确率,  $r$  表示召回率,  $p(r)$  表示准确率-召回率曲线.

mIoU 表示分割结果与真值的重合程度, 计算如下:

$$\text{mIoU} = \frac{1}{n} \sum_{i=0}^n \frac{TP}{FN + FP + TP} \quad (12)$$

式中,  $n$  为类别总数.

### 2.3 铁路整体场景全景分割实验

为了验证所提轻量化方法的整体图像分割效果, 首先进行铁路全景分割实验, 结果如图8所示. 图8(b)为文献[11]中UPSNet方法分割结果, 可以看出该方法存在明显的漏分割问题, 其原因为该分割时易忽略对全景中未定型区域的关注, 容易出现漏分割的问题, 如第3幅和第4幅中对于钢轨区域均未能有效实现分割. 图8(c)为文献[13]提出的Panoptic-DeepLab方法, 该方法通过残差网络特征来提高全景分割精度, 可以看出该方法较文献[11]UPSNet方法有所改进, 但从分割结果可以看出, 该方法在图像分割时缺少精细度控制策略, 边界易出现混淆的问题, 分割结果易出现误差. 图8(d)为文献[16]提出的MAPSNet方法, 该方法利用引

导卷积核生成注意力模块来提高分割精度, 从分割结果可以看出, 该方法相对于前两种方法有很大提升, 但是仍然存在一定的分割误差, 如第4幅左下钢轨出现了锯齿效应. 图8(e)为文献[28]提出的MP-Former方法结果, 从分割结果可以看出, 该方法分割精度有所提高, 这是因为该方法通过掩码引导模块能够更好地捕捉图像中目标的形状和位置信息, 部分缓解了全景分割中不准确预测的问题, 但仍出现误分割的现象, 如第1幅对人的不同物体分割时出现颜色错乱的问题. 图8(f)为所提方法分割结果, 可以看出采用本文提出的通道混洗及多尺度特征交互增强模块后, 对于分割精度有了较大的视觉提升, 对于铁路整体场景下不同类别分割更加精准.

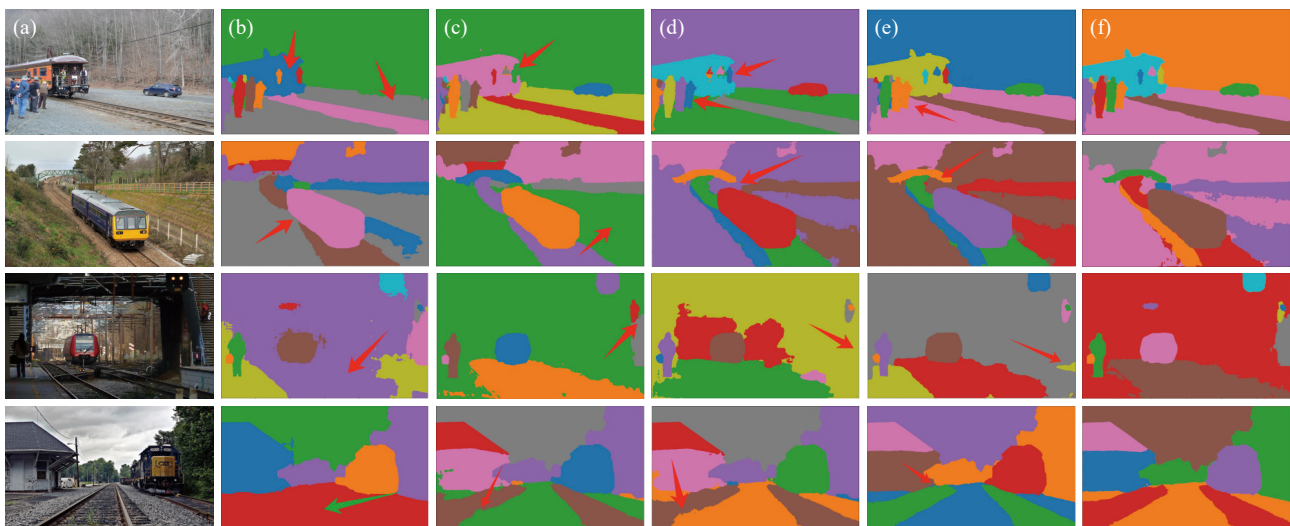


图8 铁路场景全景分割. (a) 铁路场景图像; (b) UPSNet; (c) Panoptic-DeepLab; (d) MAPSNet; (e) MP-Former; (f) 所提方法

Fig.8 Panoramic segmentation of railway scenes: (a) railway scene image; (b) UPSNet; (c) Panoptic-DeepLab; (d) MAPSNet; (e) MP-Former; (f) proposed method

### 2.4 异物侵入铁路场景下全景分割实验

下面进一步对异物侵入铁路场景进行全景分割实验, 结果如图9所示. 图9(b)为文献[11]中UPSNet方法的分割结果, 可以看出该方法在全景分割过程中, 对异物出现了漏分割的问题, 如第2幅铁路沿线左侧, 侵入限界区的行人出现了大量的漏分割现象. 图9(c)为文献[13]提出的Panoptic-DeepLab方法的分割结果, 可以看出该方法同样出现了漏分割的问题, 如第1幅钢轨中间行人出现了漏识别分割. 图9(d)为文献[16]提出的MAPSNet方法的分割结果, 可以看出该方法分割时容易出现不同物体分割结果颜色错乱的情况, 如第2幅左侧山体与钢轨路基等区域出现误分割. 图9(e)为文献[28]提出的MP-Former方法的分割结果,

可以看出该方法分割结果虽有所改善, 但仍出现误分割问题, 如第4幅对铁轨的错误分割. 图9(f)为所提方法的分割结果, 可以看出所提方法不仅能够获得目标更多的特征信息, 而且分割时细节信息和相邻问题的边界问题处理得更好, 全景分割质量优于对比算法.

为了验证所提算法在铁路场景下对于动态异物全景分割的有效性及其可视化展示, 以快速侵入铁路的犬科动物在铁轨处不同位置奔跑为例, 进行全景分割实验, 结果如图10所示. 其中, 图10(a)为原始图像, 图10(b)为文献[11]中UPSNet算法分割结果, 可以看出该方法整体分割效果不佳, 对于铁轨和地面的分割出现了严重偏差. 文献[13]Panoptic-DeepLab算法结果图10(c)和文献[16]MAPSNet算

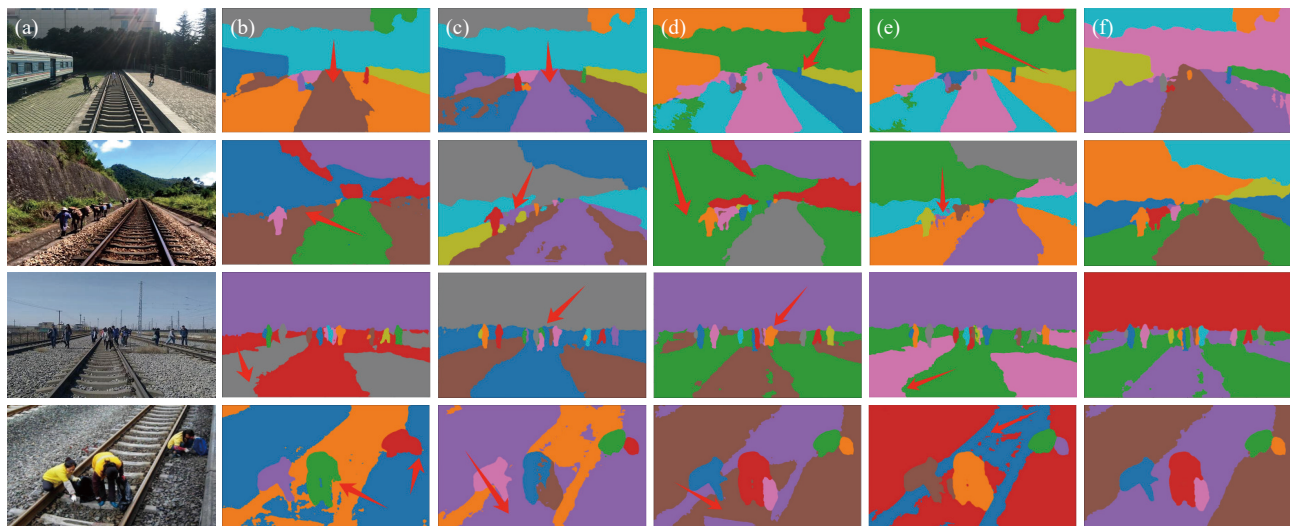


图 9 异物侵入铁路场景. (a) 铁路场景图像; (b) UPSNet; (c) Panoptic-DeepLab; (d) MAPSNet; (e) MP-Former; (f) 所提方法

Fig.9 Experimental results of foreign object intrusion in railway scenes: (a) railway scene image; (b) UPSNet; (c) Panoptic-DeepLab; (d) MAPSNet; (e) MP-Former; (f) proposed method

法结果图 10(d), 相较于 UPSNet 方法对于铁轨和地面分割有所改进, 但铁轨中仍有部分区域未准确分割, 钢轨区域分割后出现了锯齿效应. 图 10(e) 为文献 [28] 中 MP-Former 算法, 与前三种算法相

比边缘轮廓分割精度有所提高, 但也出现了误分割的现象. 图 10(f) 为所提方法结果, 可以看出所提方法能够快速准确分割出犬科动物、铁轨限界等区域, 分割效果更优.

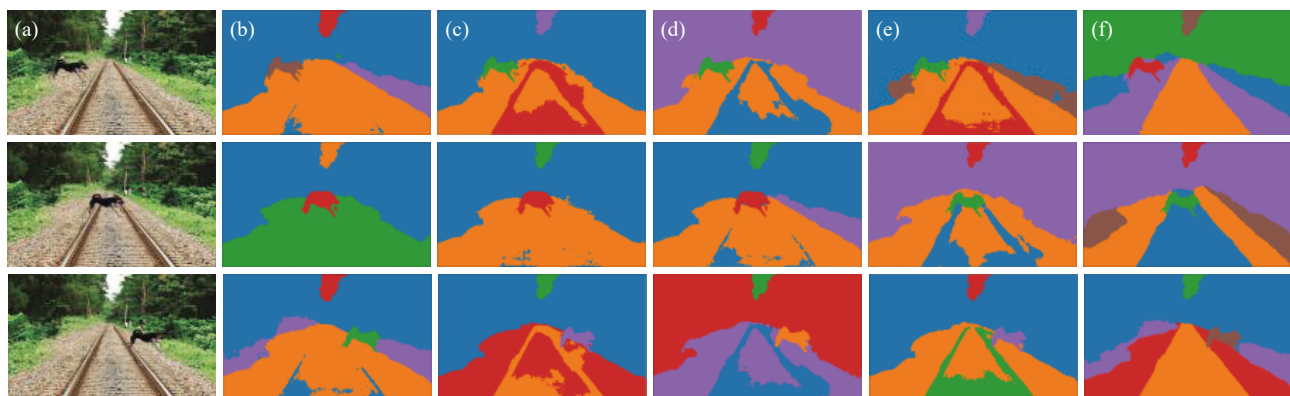


图 10 动态异物全景分割. (a) 铁路场景图像; (b) UPSNet; (c) Panoptic-DeepLab; (d) MAPSNet; (e) MP-Former; (f) 所提方法

Fig.10 Dynamic foreign object panoramic segmentation: (a) railway scene image; (b) UPSNet; (c) Panoptic-DeepLab; (d) MAPSNet; (e) MP-Former; (f) proposed method

### 2.5 模型分割精度指标分析

为了客观衡量各种分割精度方法的综合检测性能, 使用 PQ、AP 和 mIoU 进行定量衡量. 以上评价指标数值越高, 代表全景分割效果越好. 其实验数据如表 1 所示.

对表 1 中数据进行分析可以看出, UPSNet 方法采用语义分割全卷积网络和实例分割 Mask R-CNN 的结果进行融合, 模型对图像中特征的提取能力有限, 特征信息提取不充分, 导致其 PQ 仅有 45.3%, AP 的值为 30.7%, mIoU 为 72.6%. Panoptic-DeepLab 采用双路空间金字塔池化的方式用于提

表 1 不同场景下分割方法性能对比

Table 1 Performance comparison of segmentation methods across different scenarios

Experimental methods	Evaluating indicator		
	PQ/%	AP/%	mIoU/%
UPSNet	45.3	30.7	72.6
Panoptic-DeepLab	51.5	33.2	74.8
MAPSNet	53.6	33.8	76.4
MP-Former	54.7	35.1	77.9
Proposed method	55.2	35.9	78.3

升特征信息的提取能力,进而增强分割精度,该方法指标略高于 UPSNet 方法. MAPSNet 方法通过聚合多尺度空间特征信息和关注通道特征信息来引导卷积核权重提高分割准确率, MAPSNet 方法指标均优于前两种. MP-Former 方法通过掩码引导的方法,较前三种方法指标均有所提升. 本文所提方法定量评价指标均优于对比方法,其原因为所提方法设计轻量 CS\_FasterNet 特征提取网络和多尺度特征交互增强模块,在增强特征提取能力的同时,能够获取更全面细节性的全局信息,此外通过预测融合模块,提升网络对图像分割的准确性.

## 2.6 模型轻量化指标分析

在完成不同模型分割精度比较后,下面进行轻量化性能评价,轻量化的性能采用每秒处理帧数 FPS,模型参数量 (Model parameters, MP) 和计算量 (Giga floating-point operations per second, GFLOPs) 进行定量实验对比, FPS 值越大表示分割速度越快, MP 和 GFLOPs 越小代表越有利于轻量化分割. 不同方法比较结果如表 2 所示.

表 2 不同分割方法性能对比

Table 2 Performance comparison of different methods

Experimental methods	Evaluating indicator		
	FPS	MP	GFLOPs
UPSNet	23.9	8.64	95.52
Panoptic-DeepLab	12.7	36.87	176.93
MAPSNet	10.3	40.1	192.7
MP-Former	11.5	37.2	158.4
Proposed method	35.4	12.59	52.76

对表 2 中数据进行分析, Panoptic-DeepLab 模型采用残差网络提取了分割精度,但导致 GFLOPs 值较高,不利于轻量化分割. MAPSNet 方法主要是利用引导卷积核生成注意力模块,在模型参数量 MP 及计算量 GFLOPs 是对比方法中值最高的,其 FPS 最低,仅有 10.3,说明其分割的实时性最低. MP-Former 模型采用了 Transformer 架构和多尺度特征融合等技术,模型的结构相对复杂,从表 2 可以看出, MP-Former 的参数较高,不利于实时性分割. UPSNet 虽然模型参数和 GFLOPs 相对于 Panoptic-DeepLab 方法和 MAPSNet 方法较小,但从表 1 可知,其分割精度最低. 所提方法实现了轻量化的特征增强和实时轻量分割,同时每秒可以处理 35.4 帧,实时性更好,其计算量也是最小,轻量化程度最好.

## 2.7 消融实验

最后进行消融实验. 以 UPSNet 为基础网络进

行实验,分别逐步添加提出的轻量 CS\_FasterNet 特征提取网络称为 M1、多尺度特征交互增强模块叫做 M2 和预测融合模块为 M3. 将全景质量 PQ 作为指标验证各改进模块对模型性能的作用,消融实验结果如下表 3 所示.

表 3 所提模型消融实验比较

Table 3 Comparison of ablation experimental indexes for the proposed method

Reference	M1	M2	M3	PQ/%
√				45.3
√	√			47.8
√	√	√		52.1
√	√	√	√	55.2

从表 3 中可以看出,采用 CS\_FasterNet 特征提取网络对全景分割性能有所提升,增强结构对多尺度信息的提取多样性,全景质量精度由 45.3% 提升到 47.8%,提升了 2.5%. 在此基础上加入多尺度特征交互增强模块,增强信息间的交互和特征的表达能力,将模型的全景质量提升到 52.1%,提升了 4.3%. 最后在预测过程中通过改进预测融合模块进行优化,所提方法的全景质量达到 55.2%.

## 3 结论

(1) 针对高速铁路场景下图像全景分割时存在全景分割精确度低,难以实现轻量级实时分割等问题,提出了一种通道混洗与跨尺度增强的轻量级铁路全景分割方法.

(2) 提出了轻量 CS\_FasterNet 特征提取网络,实现对铁路场景下图像不同尺度特征的轻量化特征提取;并设计了多尺度特征交互增强模块,提高图像特征分割质量.

(3) 实验结果表明,所提轻量化模型在保持较高分割准确率的同时轻量化模型的每秒处理帧数与计算量均优于比较方法,客观评价优于对比方法.

## 参 考 文 献

- [1] Liu H Y, He S B, Chen J M. Data-driven adaptive adjustment strategy for strong wind alarm in high-speed railway. *Acta Autom Sin*, 2019, 45(12): 2242  
(刘昊侯, 贺诗波, 陈积明. 数据驱动的高速铁路强风报警自适应解除策略. *自动化学报*, 2019, 45(12): 2242)
- [2] Chen Y, Wang Z, Zhou F C. Railway foreign objects tracking detection based on spatial location and feature generalization enhancement. *J Beijing Univ Aeronaut Astronaut*, 2025, 51(1): 9  
(陈永, 王镇, 周方春. 空间定位与特征泛化增强的铁路异物跟

- 踪检测. 北京航空航天大学学报, 2025, 51(1): 9)
- [3] Li Z Q, Wang W H, Xie E Z, et al. Panoptic SegFormer: Delving deeper into panoptic segmentation with transformers // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans, 2022: 1270
- [4] Yang B C, Wang J Y, Jin H B. DS-TransFusion: Automatic retinal vessel segmentation based on an improved Swin Transformer. *Chin J Eng*, 2024, 46(10): 1889  
(杨本臣, 王建宇, 金海波. DS-TransFusion: 基于改进 Swin Transformer 的视网膜血管自动分割. 工程科学学报, 2024, 46(10): 1889)
- [5] Liu W Z, Chen H, Ren J X, et al. Research on knowledge extraction from street scene images based on hybrid intelligence. *Acta Geod Cartogr Sin*, 2024, 53(9): 1817  
(刘万增, 陈杭, 任加新, 等. 基于混合智能的街景影像知识提取方法. 测绘学报, 2024, 53(9): 1817)
- [6] Ning B, Mo Z S, Li K C. Application and development of intelligent technologies for high-speed railway signaling system. *J China Railw Soc*, 2019, 41(3): 1  
(宁滨, 莫志松, 李开成. 高速铁路信号系统智能技术应用及发展. 铁道学报, 2019, 41(3): 1)
- [7] Kirillov A, He K M, Girshick R, et al. Panoptic segmentation // 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, 2019: 9396
- [8] Sapkota R, Ahmed D, Karkee M. Comparing YOLOv8 and Mask R-CNN for instance segmentation in complex orchard environments. *Artif Intell Agric*, 2024, 13: 84
- [9] Zhao H S, Shi J P, Qi X J, et al. Pyramid scene parsing network // 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, 2017: 6230
- [10] De Geus D, Meletis P, Dubbelman G. Panoptic segmentation with a joint semantic and instance segmentation network [J/OL]. *arXiv preprint* (2018–9–6) [2024–10–23]. <https://arxiv.org/abs/1809.02110>
- [11] Xiong Y, Liao R, Zhao H, et al. UPSNet: a unified panoptic segmentation network // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Long Beach, 2019: 8818
- [12] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. IEEE, 2017: 640
- [13] Cheng B W, Collins M D, Zhu Y K, et al. Panoptic-DeepLab: A simple, strong, and fast baseline for bottom-up panoptic segmentation // 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, 2020: 12472
- [14] Chen Y F, Lin G C, Li S Y, et al. BANet: Bidirectional aggregation network with occlusion handling for panoptic segmentation // 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, 2020: 3792
- [15] Liu H Y, Peng C, Yu C Q, et al. An end-to-end network for panoptic segmentation // 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, 2019: 6165
- [16] Fu D, Qu S J, Fu Y. Multiscale attention-guided panoptic segmentation network. *Comput Eng Appl*, 2023, 59(22): 223  
(付都, 瞿绍军, 付亚. 多尺度注意力引导的全景分割网络. 计算机工程与应用, 2023, 59(22): 223)
- [17] Chen J R, Kao S H, He H, et al. Run, don't walk: Chasing higher FLOPS for faster neural networks // 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Vancouver, 2023: 12021
- [18] Ma X L, Guo F M, Niu W, et al. PCONV: The missing but desirable sparsity in DNN weight pruning for real-time execution on mobile devices [J/OL]. *arXiv preprint* (2019–9–6) [2024–10–23]. <https://arxiv.org/abs/1909.05073>
- [19] Chen Y, Wang Z, Zhang J J. Lightweight detection of railway object intrusion based on spectral pooling and shuffled-convolutional block attention module enhancement. *J Southwest Jiaotong Univ*, 2024, 59(6): 1294  
(陈永, 王镇, 张娇娇. 频谱池化与混洗注意力增强的铁路异物轻量检测. 西南交通大学学报, 2024, 59(6): 1294)
- [20] Zhang X Y, Zhou X Y, Lin M X, et al. ShuffleNet: An extremely efficient convolutional neural network for mobile devices // 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, 2018: 6848
- [21] Deng K W, Ge C Y. Research on lightweight of improved YOLOv5 infrared traffic detection network. *Comput Eng Appl*, 2023, 59(12): 184  
(邓楷文, 葛晨阳. 改进 YOLOv5 的轻量化红外交通目标检测. 计算机工程与应用, 2023, 59(12): 184)
- [22] Zhao Y A, Lv W Y, Xu S L, et al. DETRs beat YOLOs on real-time object detection [J/OL]. *arXiv preprint* (2023–4–17) [2024–10–23]. <https://arxiv.org/abs/2304.08069>
- [23] Yu C H, Chen X W. Railway rutting defects detection based on improved RT-DETR. *J Real Time Image Process*, 2024, 21(4): 146
- [24] Bodla N, Singh B, Chellappa R, et al. Soft-NMS—improving object detection with one line of code [J/OL]. *arXiv preprint* (2017–4–14) [2024–10–23]. <https://arxiv.org/abs/1704.04503>
- [25] Zhong P X, Wang D, Miao C Y. An affect-rich neural conversational model with biased attention and weighted cross-entropy loss. *Proc AAAI Conf Artif Intell*, 2019, 33(1): 7492
- [26] Ren S Q, He K M, Girshick R, et al. Object detection networks on convolutional feature maps. *IEEE Trans Pattern Anal Mach Intell*, 2017, 39(7): 1476
- [27] Lin T T, Maire M, Belongie S, et al. Microsoft COCO: Common objects in context // *European Conference on Computer Vision*. Zurich, 2014: 740
- [28] Zhang H, Li F, Xu H Z, et al. MP-former: mask-piloted transformer for image segmentation // 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Vancouver, 2023: 18074