文章编号:1673-0291(2026)02-0153-12 DOI:10.11860/j.issn.1673-0291.20250053

基于两阶段视觉检测的动车螺栓缺陷识别算法

吕 涂,周 航,张一凡,陈业泓,王 慧
(北京交通大学 电子信息工程学院,北京 100044)

摘 要:针对动车组运行状态图像检测中转向架精密部件缺陷识别需求,提出一种结合模板匹配与改进YOLOv5的转向架螺栓松动与丢失的识别算法.该算法首先进行转向架提取,通过双线性插值缩小图像尺寸以显著减少计算量,并采用归一化相关系数模板匹配在缩小图像上保持高精度的目标定位;然后进行缺陷识别,使用改进YOLOv5模型.针对缺陷识别任务,模型改进主要包括三方面:一是针对模型参数量大、计算开销高的问题,引入跨尺度特征融合模块(CNN-based Cross-scale Feature-fusion Module,CCFM)替代路径聚合网络(Path Aggregation Network,PANet)精简参数量和模型大小;二是针对数据集中微小目标的检测需求,增设小目标检测层强化微小缺陷捕捉能力;三是为解决特征感知范围不足的问题,在浅层网络嵌入大型选择性卷积核(Large Selective Kernel,LSK)扩展特征感知范围,提高检测精度.最后,对算法的转向架提取效率与缺陷识别性能进行实验验证.实验结果表明:在转向架提取阶段,缩小图像后的模板匹配平均处理时间由929.58 ms/张降至74.87 ms/张,缩短约92%;在缺陷识别阶段,相较于基准模型在IoU阈值0.5和0.5~0.95下的平均精度均值(mean Average Precision,mAP)为89.1%和43.1%,改进模型在参数量减少约23%的情况下分别提升至91.5%和46.2%,提高了2.4%和3.1%.该方法有效提升了动车组精密部件的检测效率和准确性,为动车组的安全运维提供了技术参考.

关键词:缺陷识别;视觉检测;转向架螺栓;YOLOv5;模板匹配
中图分类号:TP391.4;U266 **文献标志码**:A

Bolt defect recognition algorithm for EMU based on two-stage visual detection

LYU Tu, ZHOU Hang, ZHANG Yifan, CHEN Yehong, WANG Hui
(School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China)

Abstract: To address the need for defect identification in precision components of electric multiple unit (EMU) bogies during operational image inspection, this paper proposes a recognition algorithm for detecting loose and missing bogie bolts by combining template matching with an improved YOLOv5 model. The proposed algorithm first performs bogie extraction. During bogie extraction, bilinear inter-

收稿日期:2025-04-15;**修回日期**:2025-12-08
基金项目:国家重点研发计划(K24B05200020);中国国家铁路集团有限公司科技研究开发计划重点课题(M23D00101);太原市"揭榜挂帅"项目(W25M200071);北京交通大学自然科学横向项目(W21L00390)
Foundation items: National Key R&D Plan (K24B05200020); Science and Technology Research and Development Key Program of China Railway Corporation (M23D00101); Taiyuan "Leading the Charge with Open Competition" Project (W25M200071); Project of Beijing Jiaotong University (W21L00390)
第一作者:吕涂(2002—),男,安徽界首人,硕士生.研究方向为基于计算机视觉的缺陷检测.**email**:23125038@bjtu.edu.cn.
通信作者:周航(1974—),男,四川成都人,副教授,博士.**email**:hangzhou@bjtu.edu.cn.

引用格式:吕涂,周航,张一凡,等.基于两阶段视觉检测的动车螺栓缺陷识别算法[J].北京交通大学学报,2026,50(2):153-164.
LYU Tu, ZHOU Hang, ZHANG Yifan, et al. Bolt defect recognition algorithm for EMU based on two-stage visual detection[J]. Journal of Beijing Jiaotong University, 2026, 50(2): 153-164. (in Chinese)

polation is utilized to downscale the images, significantly reducing computational overhead. Simultaneously, normalized cross-correlation template matching is employed to maintain high-precision target localization on the downscaled images. Then, the proposed algorithm performs defect recognition, for which an improved YOLOv5 model is applied. For the defect recognition task, the model improvements mainly include three aspects: first, to address the large parameter size and high computational cost of the baseline model, a CNN-based Cross-Scale Feature-Fusion Module (CCFM) replaces the Path Aggregation Network (PANet), effectively reducing both the parameter count and overall model size; second, to meet the requirements for detecting tiny targets within the dataset, a small-object detection layer is added to enhance the capture of minute defects; third, to overcome the limited feature perception range, Large Selective Kernel (LSK) modules are embedded into the shallow network layers to expand the receptive field and improve detection accuracy. Finally, the bogie extraction efficiency and defect recognition performance of the algorithm are experimentally verified. Experimental results demonstrate that, in the bogie extraction stage, the average processing time of template matching on downscaled images is reduced by approximately 92%, from 929.58 ms to 74.87 ms per image. In the defect recognition stage, compared with the baseline model's mean Average Precision (mAP) values of 89.1% at an IoU threshold of 0.5 (mAP@0.5) and 43.1% over IoU thresholds from 0.5 to 0.95 (mAP@0.5:0.95), respectively, the improved model achieves 91.5% and 46.2% while simultaneously reducing the parameter count by approximately 23%, yielding improvements of 2.4% and 3.1%. The proposed method effectively enhances both the efficiency and accuracy of defect detection for precision components in EMUs, providing a valuable technical reference for safe operation and maintenance.

Keywords: defect recognition; visual inspection; bogie bolts; YOLOv5; template matching

动车组运行状态图像检测技术主要用于动车组侧部与底部关键运行部件的图像采集与状态监测，相关检测数据可依托集中化平台进行综合分析与应用管理[1].该类技术对于保障动车组高速运行的安全性和稳定性具有重要意义.随着检测系统在铁路中的持续应用，动车组运行安全保障需求对关键部件缺陷识别的自动化与智能化水平提出了更高标准.尤其是在转向架螺栓等微小部件的检测任务中，仍存在一定技术难点：一是螺栓等微小部件的缺陷特征不明显，识别难度较大；二是异常样本数量有限，模型训练数据不足；三是实际应用中需要在保证检测精度的同时具备较高的图像处理效率，以适应多场景运行环境下的实时检测需求[2].

为解决动车组部件异常检测中的目标区域定位与缺陷识别问题，现有研究主要分为传统机器视觉方法和深度学习方法两类.在基于传统机器视觉的动车故障检测方面，文献[3]改进了尺度不变特征变换(Scale Invariant Feature Transform, SIFT)算法，用于检测动车底部图像的异常，通过自动提取和匹配特征点，利用交叉相关修正图像定位异常区域，标记并显示异常位置.但在复杂背景和微小缺陷检测中鲁棒性不足，容易受到光照和噪声干扰.文献[4]提出了一种自适应融合局部SIFT特征匹配和全局模板匹配的图像故障识别算法，用于动车组故障检测，通过图像对齐、局部比对粗略定位故障区域，再利用模板匹配精准定位故障，以提高故障检测准确性和降低误报率，但其计算开销较大，实时性较差.文献[5]采用了一种基于空间尺度标准化的图像对齐模型和改进的背景差分法，用于动车组底部异常检测，通过解决列车速度波动导致的图像失真问题，提高了异常区域的定位与识别效率和准确性，但在缺陷检测的适应性方面仍作用有限.文献[6]提出了一种基于拓扑交叉数和均值漂移聚类的弱纹理图像配准算法，用于车底故障区域的定位，能有效提高故障检测的准确率，但在多目标和小目标检测中的泛化能力不足.总体来看，传统视觉方法在定位精度方面具有一定优势，但在复杂背景和微小缺陷检测中的鲁棒性和适应性仍显不足.

随着深度学习的快速发展，越来越多的研究开始应用深度学习模型进行轨道交通装备的缺陷检测.文献[7]介绍了一种基于卷积神经网络(Convolutional Neural Networks, CNN)的两阶段自动故障识别系统.该系统首先利用CNN模型对侧架销和轴端螺栓的目标区域进行定位，再构建另一CNN模型对检测到的目标区域进行多故障判定，实现典型故障识别.该方法实现了目标定位与故障识别的结合，

但多模型串联使其结构较复杂,推理效率仍有待提升.文献[9]结合改进的SSD算法与DeepLabv3 plus语义分割模型,实现了高铁接触网顶紧螺栓的高精度检测,但模型参数量较大,限制了其在实时场景中的部署.文献[10]提出了基于注意力反向知识蒸馏的车轮踏面异常检测模型,通过U-Net分割和注意力机制提升了检测性能,但检测效果依赖前置分割精度.文献[11]提出了基于新型对抗自编码器的铁路货车车辆异物检测算法,通过无监督学习方式利用无标注正常图像训练模型,并引入注意力机制和异常评分机制,在缺陷检测中减少了标注依赖,但在实时检测需求下仍存在性能瓶颈.综上,深度学习方法在微小目标识别和检测精度方面取得了显著成效,但模型复杂度高、处理效率低,限制了其在实际场景中的应用效果.

基于上述分析,针对传统视觉方法在复杂场景下鲁棒性不足、深度学习方法在检测效率和模型复杂度上的局限,提出一种结合模板匹配与改进YOLOv5的转向架螺栓松动与丢失检测方法.该方法首先采用模板匹配实现转向架区域的快速、准确提取,降低图像处理规模;然后在改进的YOLOv5模型中引入跨尺度特征融合模块(CNN-based Cross-scale Feature-fusion Module, CCFM)和小目标检测层,以提高微小部件检测精度与实时性能,同时嵌入大型选择性卷积核(Large Selective Kernel, LSK)增强模型的特征感知能力,提升复杂背景下的检测鲁棒性.

1　动车组转向架螺栓缺陷检测方法

本文所使用的图像来源于动车段模拟故障图库中的原始底图,目标检测区域为动车组底部两侧的转向架区域.原始图像分辨率为26 100×832,若直接使用原图进行目标检测,由于分辨率较高且背景信息复杂,容易导致计算量增大并影响检测效率[12].因此,本文将动车组转向架螺栓缺陷识别问题划分为两个阶段:先对转向架区域进行图像提取,然后对螺栓进行缺陷检测.整体处理流程如图1所示.

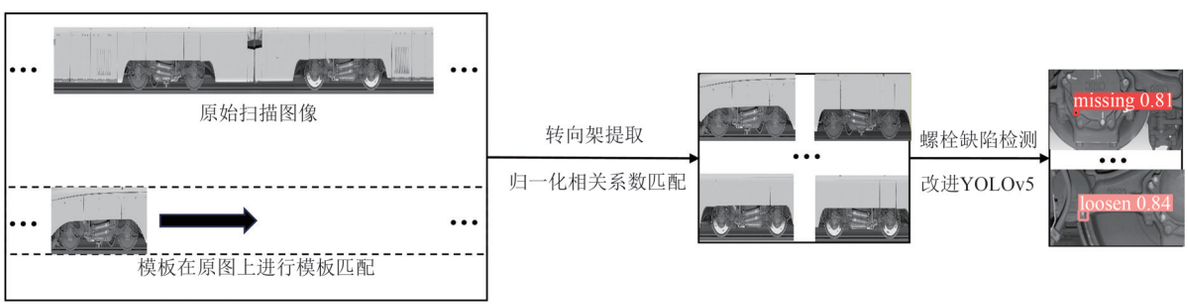


图1　检测方法流程图

Fig.1　Flow diagram of the detection method

1.1　基于模板匹配的转向架提取

为了提升检测效率,针对动车关键部件的识别过程,先精确定位部件所在区域,然后再进行检测,这种方法相比将整张图像裁剪后直接送入检测网络的策略,提高了计算资源的利用率及检测速度.尤其在侧视扫描图像的场景下,转向架包含多个关键部件,其状态对动车组运行具有重要影响,是一个需要重点关注的检测部位.

在图像中定位一个已知目标的最直接方法是在像素级上搜索需要识别的目标,其中模板匹配是一种计算量较小的目标识别算法.在模板匹配中,首先选定一个小的图像作为目标图像,然后在一个大的输入图像中滑动这个小的目标图像,从而寻找与其最相似的部分.基于像素灰度的模板匹配方法中,绝对误差和算法与误差平方和算法稳定性容易受图像亮度、对比度等因素的影响.归一化互相关系数算法(Normalized Cross Correlation,NCC)通过计算两个图像对应像素之间的互相关系数来度量相似度,具有一定的抗噪声及抗亮度变化干扰能力,能够提高模板匹配的精度[13],因此,该算法在实际系统工作过程中对环境变化具有更好的鲁棒性.NCC的计算式为

$$\mathrm{NCC}(x,y)=\frac{\sum_{j=1}^{N}\sum_{i=1}^{M}I(x+i,y+j)\cdot T(i,j)}{\sqrt{\sum_{j=1}^{N}\sum_{i=1}^{M}I(x+i,y+j)^2}\cdot\sqrt{\sum_{j=1}^{N}\sum_{i=1}^{M}T(i,j)^2}} \tag{1}$$

式中:$I(x+i,y+j)$表示原图像中位于$(x+i,y+j)$位置的像素值,其中(x,y)为原图像中当前计算的参考位置,(i,j)为模板图像像素相对位置偏移;$T(i,j)$为模板图像中坐标(i,j)的像素值;M、N为模板图像的行、列像素数;$\mathrm{NCC}(x,y)$为计算出的归一化互相关系数.

在匹配过程中，选择与原图等高的转向架图像作为模板，因此计算出的归一化互相关系数为一维数组，含4个转向架的原图匹配结果可视化如图2所示.

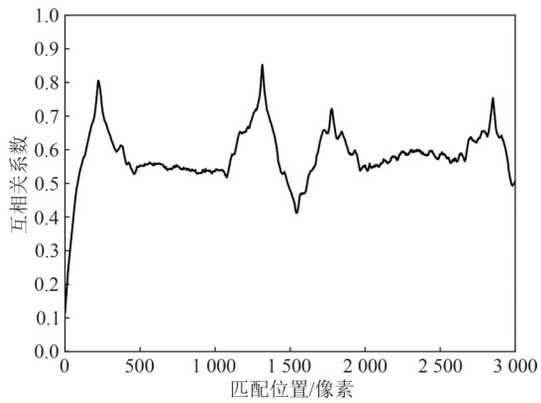


图2 模板匹配的互相关系数分布

Fig.2 Distribution of correlation coefficients for template matching

图2中：纵轴代表匹配结果，即归一化互相关系数，该系数越接近于1的区域，意味着原图中区域与模板图像的匹配程度越高；横轴为匹配的位置，在本文中，模板图像沿原图水平方向从左至右遍历.原图的尺寸约为26 100×832像素，而模板图像的尺寸为2 048×832像素.为提升匹配效率并保证鲁棒性，两幅图像均等比例缩小至原图像的八分之一进行处理，因此，横轴的覆盖范围计算为：(原图宽度－模板宽度)/缩小比例≈(26 100－2 048)/8≈3 000.采取等比例缩小两幅图像的策略再进行匹配，显著优化了计算资源利用并极大缩减了匹配所需时间，使之与后续缺陷识别算法相契合.为验证缩放策略的有效性，选取100张原始侧视图像进行测试，分别在未缩放与等比例缩小至1/8分辨率两种情况下执行NCC模板匹配，并统计完成一次完整匹配所需的平均时间.结果表明，未缩放图像的平均匹配耗时为929.58 ms，而缩放至1/8的图像仅需74.87 ms，处理效率提升约92%，且未出现任何漏检或误检，100张图像的匹配准确率均为100%.

转向架区域的定位过程主要包括以下步骤：首先，在所得的相关系数的数组中识别最大值，当该值超越预设的阈值时，将其周围相当于模板宽度的区域标记并记录，同时将此区域内的相关系数置为零，以避免重复检测.然后，此过程循环往复，直至搜寻并标记所有转向架区域为止.最后，依据保留的区域坐标，从原始图像中裁剪出所有转向架区域图像，并将其输入缺陷检测网络进行进一步分析.

1.2 基于改进YOLOv5的螺栓缺陷检测

在动车组部件中，螺栓属于精密部件，其在缺陷检测数据集图像中所占像素比例微小，属于典型的小目标检测问题.鉴于此，本文对YOLOv5模型进行针对性优化，改进后的网络结构如图3所示.该网络以CBS、C3与SPPF等模块为基础构建主干特征提取网络.其中，CBS模块由卷积(Conv)、批归一化(BN)及SiLU激活函数构成，负责基础的特征提取与非线性变换；C3模块借鉴CSPNet结构，通过拆分与融合特征图以优化梯度流并丰富特征表达能力，C3_n表示该模块内部残差组件的重复堆叠次数，用于控制不同深度特征层的非线性变换强度；SPPF模块则通过串行与并行的最大池化操作来融合多尺度上下文信息.本文对YOLOv5的改进包括以下3方面：将特征融合部分路径聚合网络(Path Aggregation Network，PANet)更换为CCFM；增加一个检测小目标的检测头部；在主干网络与CCFM网络浅层连接处和CCFM网络内部嵌入LSK大核注意力模块.

1.2.1 CCFM替换颈部网络

CCFM是RT-DETR[14]中提出的一种模块，它是一种类似于PANet[15]结构的特征融合网络，两者都采用了自底向上的路径增强，更有效地利用底层特征中的定位信息，从而获得较高的检测性能与定位精度.本文引入CCFM作为模型的Neck部分来替换原YOLOv5模型的PANet，在检测精度小幅下降的情况下，达到显著的轻量化效果.CCFM设计的核心是其中的融合块即图3中的Concat操作与RepC3模块的组合，具体结构如图4所示.图4中，N表示RepBlock的重复堆叠次数，用于控制RepC3模块中重参数化卷积块的数量.通过调整N的取值，可以改变模块的特征提取深度，从而在特征表达能力与计算复杂度之间进行平衡.

RepC3的设计旨在增强特征表示能力，同时保持计算效率.该模块通过整合多个卷积操作，以一种高效而灵活的方式丰富了网络的表示能力.RepC3模块通过多分支卷积结构增强特征表示能力，并利用重参数化思想在推理阶段降低计算开销.模块内部通过1×1卷积进行通道转换，并由多个RepBlock进行特征提取，最后通过残差相加完成特征融合，在保持较低计算复杂度的同时提升网络表达能力.

1.2.2 引入小目标检测层

在构建的螺栓缺陷数据集中，检测的目标即螺栓的丢失与松动问题均属于小目标.随着神经网络层的加深，尽管网络能够捕获更丰富的语义信息，但

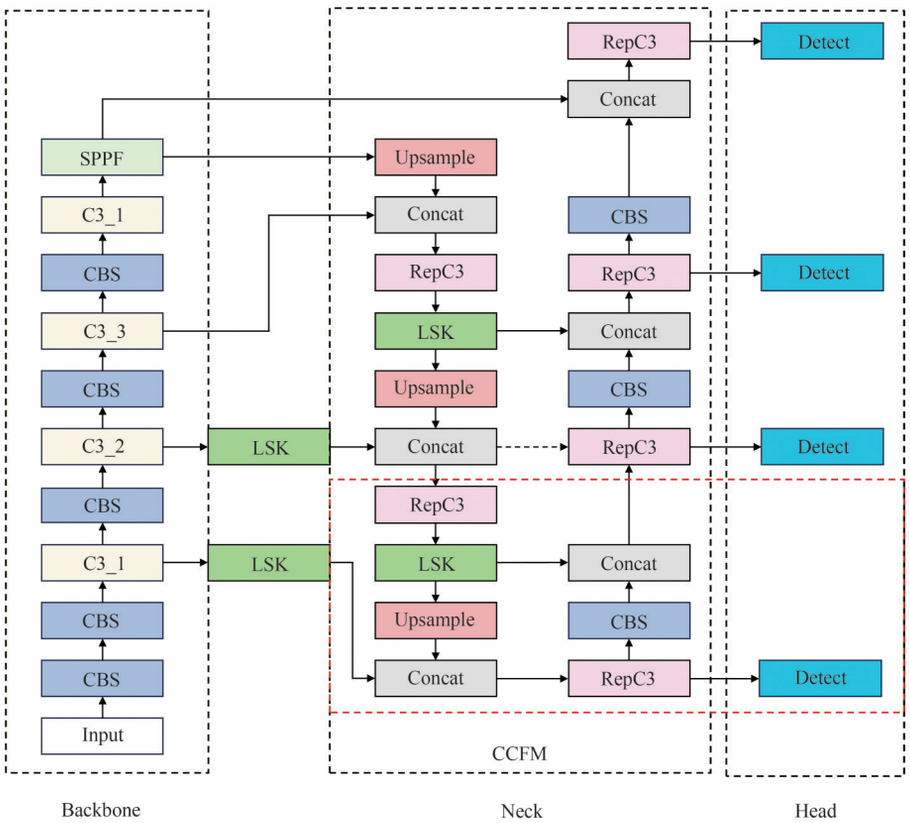


图 3　改进后的 YOLOv5 网络结构

Fig.3　Improved YOLOv5 network structure

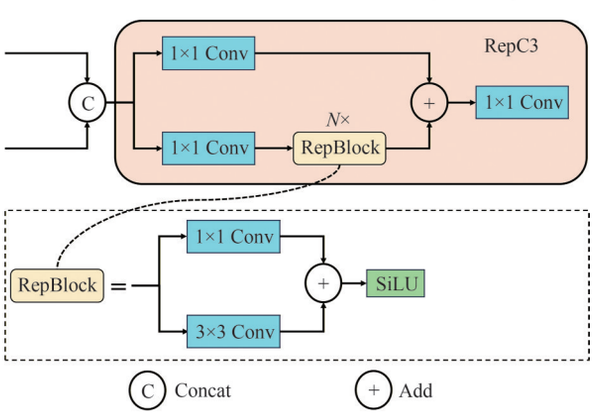


图 4　CCFM 中的融合块

Fig.4　Fusion block in CCFM

位置精确度却有所下降,尤其是对于小目标,其关键特征在下采样过程中容易丢失[16].鉴于此,本文在引入的CCFM结构上增加小目标检测层,如图5所示.该层采用下采样率为4倍的浅层特征图作为头部网络的补充输入,与改造的CCFM原先集成的8倍、16倍及32倍下采样特征进行融合.通过综合不同尺度的特征信息,增强模型对小目标细节的捕捉能力,从而提升整体的检测精度.

1.2.3　添加LSK模块

LSKNet(Large Selective Kernel Network)[17]是一种为遥感目标检测任务设计的深度学习网络,它强调了在图像中精准捕获目标所必需的广泛且动态的场景上下文信息.LSKNet的核心是LSK模块,LSK模块通过分解大核卷积来构建具有大感受野的卷积层.这是通过一系列深度卷积实现的,其中每个卷积具有逐渐增大的核心尺寸和扩张率,从而以较小的参数和计算量快速扩展感受野.在通过分解得到多个具有不同感受野的特征之后,LSK模块使用空间选择机制来动态选择不同目标最相关的空间上下文区域.这包括将经过不同卷积核处理的特征进行拼接,然后通过平均池化和最大池化提取空间关系,接着通过卷积层生成空间注意力图,最后通过sigmoid激活函数获得每个分解大核心的个别空间选择掩码.

LSK模块的空间选择过程主要包括以下6个步骤.

步骤1:将输入$\boldsymbol{X}$从具有不同感受野范围的不同内核获得的特征拼接起来,得到$\tilde{\boldsymbol{U}}$为

$$\tilde{\boldsymbol{U}}=[\tilde{\boldsymbol{U}}_1;\ \tilde{\boldsymbol{U}}_2] \tag{2}$$

式中:$\tilde{\boldsymbol{U}}_1$为输入$\boldsymbol{X}$经过卷积核为5的深度可分离卷积操作并通过1×1卷积层得到的特征;$\tilde{\boldsymbol{U}}_2$为输入$\boldsymbol{X}$先后经过卷积核大小为5和7的深度可分离卷积操

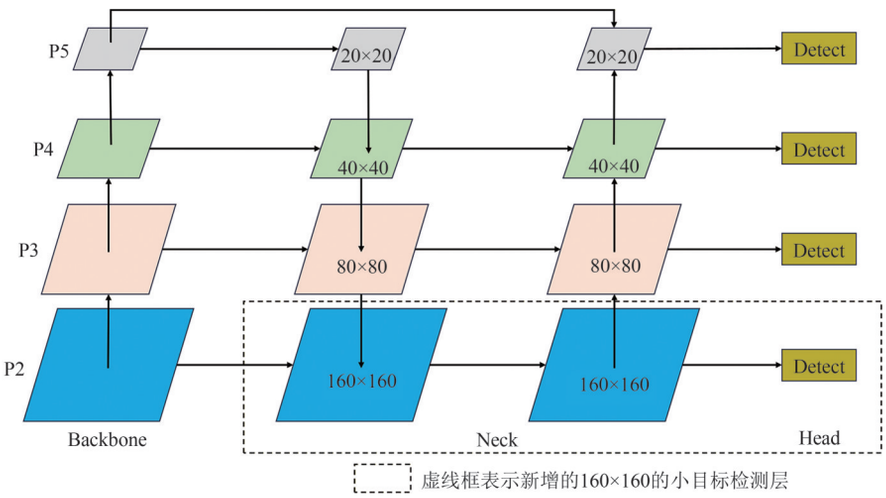


图 5 增加小目标检测层

Fig.5 Adding a small target detection layer

作并通过1×1卷积层得到的特征.

步骤2:通过应用基于通道的平均池化P_{avg}和最大池化$P_{\max}$来有效提取空间关系的计算式为

$$\begin{cases}\text{SA}_{\text{avg}}=P_{\text{avg}}(\tilde{U})\\ \text{SA}_{\max}=P_{\max}(\tilde{U})\end{cases}\tag{3}$$

式中:SA_{avg}和$\text{SA}_{\max}$分别是经过平均池化和最大池化后的空间特征描述符.

步骤3:连接空间池化特征并使用卷积层$\mathcal{F}^{2\to M}$将池化特征(具有2个通道)转换为M个空间注意力图,以实现不同空间描述符之间的信息交互,空间注意力图集合$\widehat{\text{SA}}$计算式为

$$\widehat{\text{SA}}=\mathcal{F}^{2\to M}([\text{SA}_{\text{avg}};\ \text{SA}_{\max}])\tag{4}$$

步骤4:使用sigmoid激活函数为每个分解的内核获取单独的空间选择掩码$\widetilde{\text{SA}_i}$,其计算式为

$$\widetilde{\text{SA}_i}=\sigma(\widehat{\text{SA}_i})\tag{5}$$

式中:$\sigma(\cdot)$为sigmoid激活函数;$\widehat{\text{SA}_i}$为每个空间注意力图.

步骤5:将分解后的大核序列中的特征与相应的空间选择性掩码进行加权,并通过卷积层$\mathcal{F}(\cdot)$进行融合,得到注意力特征$\boldsymbol{S}$,其计算式为

$$\boldsymbol{S}=\mathcal{F}\left(\sum_{i=1}^{N}(\widetilde{\text{SA}_i}\cdot\widetilde{\boldsymbol{U}_i})\right)\tag{6}$$

步骤6:LSK模块的最终输出$\boldsymbol{Y}$是输入特征$\boldsymbol{X}$和$\boldsymbol{S}$之间的逐元素乘积,其计算式为

$$\boldsymbol{Y}=\boldsymbol{X}\cdot\boldsymbol{S}\tag{7}$$

LSK模块的概念图如图6所示.

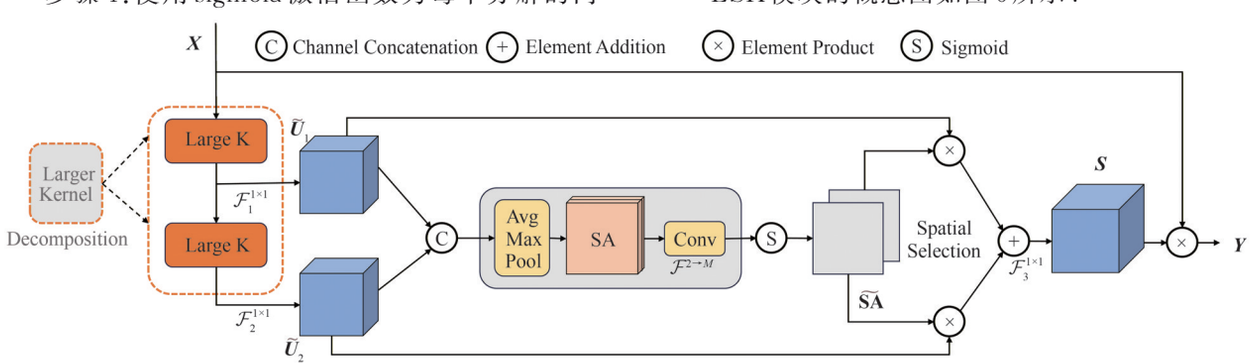


图 6 LSK模块结构图

Fig.6 LSK module structure diagram

LSK模块通过有效分解大核卷积,构建具有大感受野的特征表示,鉴于螺栓缺陷往往需依托于丰富的上下文信息才能从复杂环境中被准确识别,LSK的这一特性有效提升了检测精度.进一步地,LSK的空间选择机制可以增强网络对缺陷所在区域的关注,通过空间注意力图来加强模型对缺陷区域的特征提取.

2 实验及结果分析

2.1 数据集与实验环境

实验所使用的原始图像数据来源于动车段模拟故障图库中的原始底图.本文研究对象为侧部扫描

图像中转向架区域内的螺栓缺陷,旨在实现对螺栓丢失和松动状态的准确检测.统计分析表明,螺栓缺陷在原始图像中的平均像素面积约为15×15像素,最大不超过26×26像素,远低于通用小目标检测阈值32×32像素,因此属于典型的小目标检测任务.由于真实故障图像数量有限且获取难度较大,结合相关技术人员提供的故障特征信息,对部分缺陷样本进行了人工模拟构建.具体方法为在原始图像中随机复制缺陷区域并粘贴至对应位置,以生成更多缺陷样本[18],该方法在扩充缺陷样本数量的同时,也起到了增强检测模型训练效果的作用.随后按照6:2:2的比例将数据集划分为训练集、验证集和测试集.为在有限背景条件下缓解成像扰动带来的性能波动,对整个数据集进一步采用数据增强策略,包括随机添加噪声、调节亮度以及模拟不同天气条件(如雨、雪、雾)等方式,使模型对光照变化与成像噪声具备一定的适应能力.最终训练集为780张,验证集和测试集均为260张.数据增强示例如图7所示.

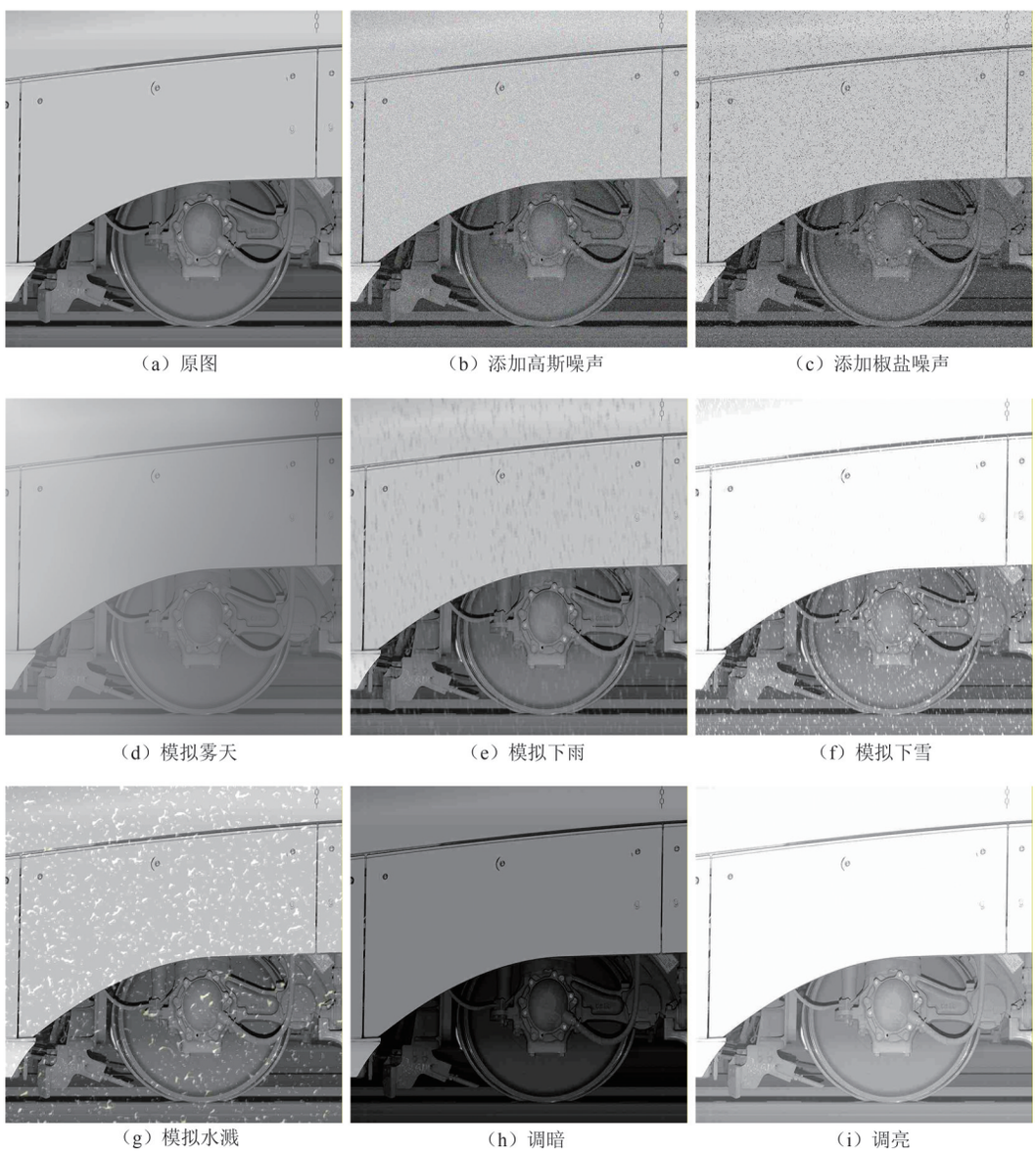


图7　数据增强示例

Fig.7　Data enhancement example

实验在以下软硬件环境及超参数配置下进行:硬件环境采用NVIDIA GeForce RTX 3090显卡,显存容量24 GB;软件环境基于Python 3.9,深度学习框架为PyTorch 1.9.0,并结合CUDA 11.1进行加速.训练过程中,Batch size设为32,采用Adam优化器,初始学习率为0.001,动量参数为0.937.

2.2　评价指标

本文采用检测精度与计算效率两类指标系统评估模型性能.精度指标包括精确率(Precision)、召回率(Recall)及平均精度均值(mean Average Preci-

sion, mAP).效率指标涵盖模型大小、参数量及推理速度.

精确率是指预测为正样本且预测正确的结果占预测为正样本的比例,召回率是指在实际为正样本的情况下,被正确预测为正样本的比例,精确率与召回率的计算式为

$$\text{Precision}=\frac{\text{TP}}{\text{TP}+\text{FP}} \tag{8}$$

$$\text{Recall}=\frac{\text{TP}}{\text{TP}+\text{FN}} \tag{9}$$

式中:TP为正确预测的正样本数量;FP为错误预测为正样本的负样本的数量;FN为错误预测为负样本的正样本的数量.

为综合评价模型在不同召回率下的精度表现,引入平均精度(Average Precision, AP),其定义为精确率-召回率(P-R)曲线下的面积.对于多类别任务,采用所有类别AP的平均值,即平均精度均值(mAP),AP与mAP的计算式为

$$\text{AP}=\int_0^1 P(R)\mathrm{d}(R) \tag{10}$$

$$\text{mAP}=\frac{\sum_{i=1}^{C}\text{AP}}{C} \tag{11}$$

式中:C为类别总数,本文涉及螺栓丢失与松动两个类别,故C=2;mAP为IoU阈值从0.5~0.95范围内计算的mAP平均值,还使用IoU阈值分别为0.5和0.75时的mAP_{50}和mAP_{75}来更好地评估模型在不同定位精度要求下的表现.

除了检测精度外,本文使用模型大小(Model Size)、参数量(Params)和推理速度(FPS)作为效率评价指标.FPS是指模型每秒能处理的图像数量,其计算式为

$$\text{FPS}=\frac{1}{\bar{t}} \tag{12}$$

式中,$\bar{t}$为单张图像的平均推理时间,s.

2.3 LSK模块位置优化实验与有效性分析

为探索LSK模块在网络中的有效集成方式,并为其在后续模型中的应用提供一个相对更优的配置,首先对其放置位置进行消融研究.具体设计了4种实验方案.

方案1:Backbone.将LSK模块添加到Backbone中的C3模块之后;

方案2:Head.将LSK模块添加到Neck的所有输出层(4个RepC3)中;

方案3:Neck-d.将LSK模块添加到Neck部分的前两个RepC3后,并添加到Backbone深层与Neck连接的部分;

方案4:Neck-s.将LSK模块添加到Neck部分的前两个RepC3后,同时添加到Backbone浅层与Neck连接的部分.

4种方案的实验结果如表1所示.由表1可知:Neck-s方案在Precision、Recall、mAP、mAP_{50}四项指标中均取得最优结果;将LSK模块添加到Backbone和Neck部分的浅层连接处并结合Neck浅层特征融合部分可以产生最佳的检测性能;因此最终选择Neck-s方案作为模型的配置,此配置将被用于后续所有包含LSK模块的实验中.

表1 LSK模块在不同放置位置的实验结果

Tab.1 Experimental results of LSK modules placed in different locations %

位置	Precision	Recall	mAP	mAP_{50}
Backbone	89.5	94.2	45.8	90.9
Head	90.3	92.8	44.8	89.8
Neck-d	89.9	93.7	44.3	90.1
Neck-s	**91.0**	**94.6**	**46.2**	**91.5**

最终确定的LSK模块放置位置如图8所示.

为了直观展示LSK模块对模型注意力分布的影响,本文采用梯度加权类激活映射(Gradient-weighted Class Activation Mapping, Grad-CAM)生成网络关注区域的热力图.热力图能够可视化模型在输入图像中关注的关键位置,从而帮助分析LSK模块是否有效增强了模型对缺陷区域的感知能力.网络增加LSK模块前后的热力图如图9所示.由图9可知:增加LSK模块前,热力图中模型对螺栓缺陷区域的关注较为分散.相比之下,增加LSK模块后,热力图显示模型对缺陷所在区域的关注明显增强,特征提取更加集中且准确,体现了LSK模块通过动态选择相关空间上下文区域有效提升模型检测能力的优势.

2.4 消融实验

为了评估所提出的改进算法的检测性能,进行消融实验,以评估每个改进对其性能的影响.评价指标包括mAP、mAP_{50}、mAP_{75}、Params、FPS,结果如表2所示.

由表2可知:①引入CCFM后,相较于基准模型(参数量7.02M,推理速度126.0FPS),参数量显著减少至4.96M,推理速度提升至142.1FPS,尽管检测精度指标略有下降,但模型计算效率明显提高.②进一步集成小目标检测层(P2)后,相较于仅使用CCFM的模型,增加P2层后,R值由92.2%提升至93.6%,mAP由42.3%提升至45.4%,表明P2层能够显著提升召回率和整体检测精度.③在CCFM和P2基础上

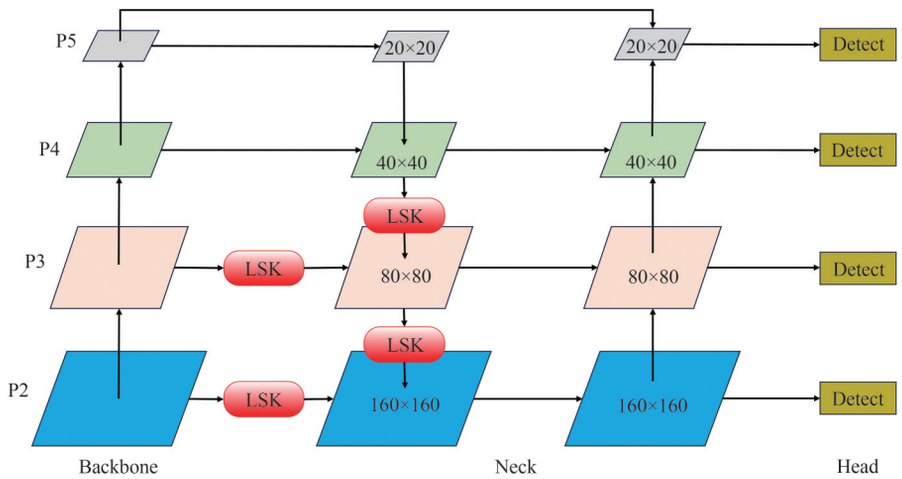
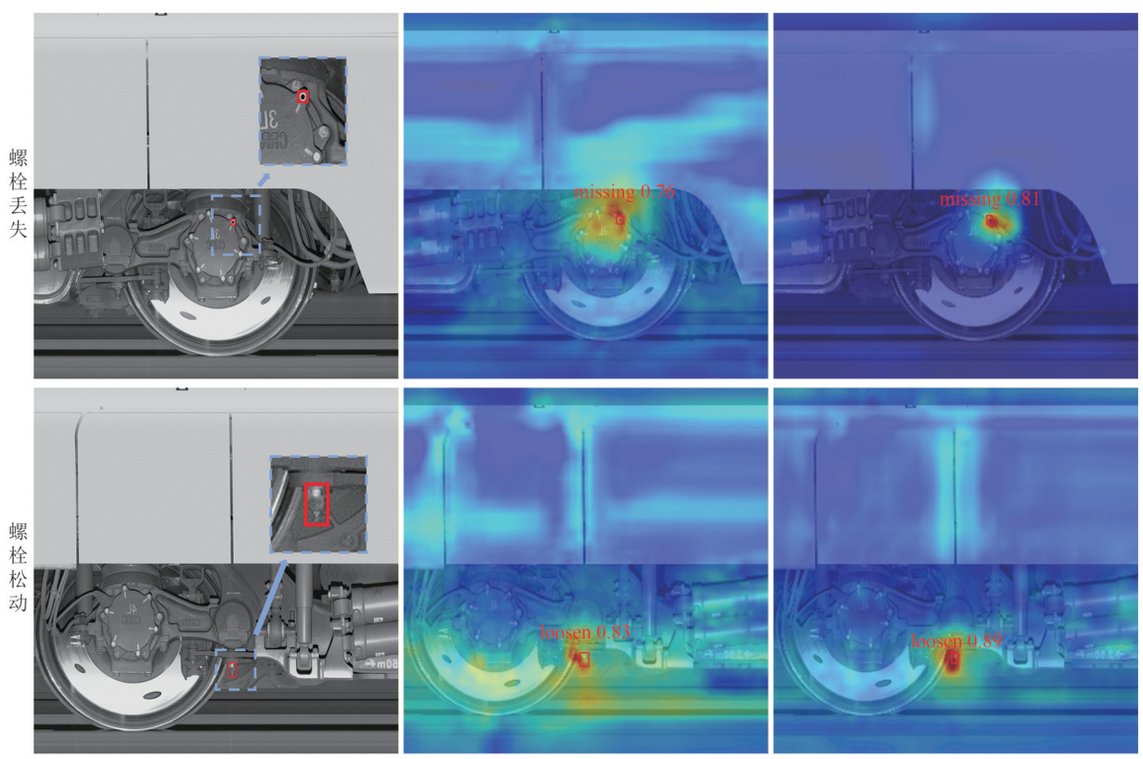


图 8　最终确定的LSK模块放置位置

Fig.8　Final placement of the LSK modules



（a）原图　　（b）添加LSK模块前　　（c）添加LSK模块后

图 9　LSK模块加入前后缺陷区域的热力图对比

Fig.9　Comparison of heatmaps in defective areas before and after adding LSK modules

表 2　消融实验结果

Tab.2　Results of ablation experiments

CCFM	P2	LSK	Precision/%	Recall/%	mAP/%	mAP_{50}/%	mAP_{75}/%	Params/M	FPS
—	—	—	90.2	92.8	43.1	89.1	33.8	7.02	126.0
√	—	—	89.9	92.2	42.3	88.3	31.7	**4.96**	**142.1**
—	√	—	90.5	94.1	46.0	90.3	38.7	7.16	111.8
—	—	√	90.4	92.7	43.5	89.3	34.0	7.09	123.6
√	√	—	89.9	93.6	45.4	89.7	36.8	5.30	99.8
√	—	√	90.1	92.4	42.9	88.9	32.1	5.00	133.1
—	√	√	90.3	93.7	45.6	90.4	38.2	7.31	79.5
√	√	√	**91.0**	**94.6**	**46.2**	**91.5**	**40.3**	5.35	85.8

进一步引入LSK模块后，尽管推理速度由99.8FPS降至85.8FPS，但检测精度显著提升，Precision值由89.9%提升至91.0%，Recall值由93.6%提升至94.6%，mAP由45.4%提升至46.2%，表明LSK模块能够有效增强模型对缺陷区域的特征提取能力.最终集成模型在减少参数量的情况下，mAP_{50}和mAP较基准模型YOLOv5分别提高了2.4%和3.1%，在精度和实时性之间取得了良好平衡.

2.5 对比实验

为验证由上述消融实验得出的最终模型的综合性能，本文在第一阶段模板匹配提取转向架区域的基础上，进一步进行了数据增强处理，构建了统一的检测数据集.所有对比模型均在该数据集上进行训练与测试，最终得到了综合的检测结果数据.本文将改进后的模型与多种主流先进目标检测模型进行对比，结果如表3所示.

表3 不同网络模型结果对比

Tab.3 Comparison of results from different network models

Module	mAP/%	mAP_{50}/%	Model Size/MB	Params/M	FPS
YOLOv5	43.1	89.1	13.6	7.02	126.0
YOLOv7[19]	36.4	86.4	71.3	36.49	44.2
YOLOv8	33.1	70.5	21.4	11.13	103.3
YOLOv8-p2	41.5	85.9	20.6	10.63	80.4
RT-DETR[14]	41.3	86.5	63.1	31.99	37.1
YOLOv10[20]	27.6	62.6	15.7	8.04	**145.2**
YOLO11	34.2	70.4	18.2	9.41	84.3
本文模型	**46.2**	**91.5**	**10.8**	**5.35**	85.8

由表3可知，本文提出的改进算法在mAP和mAP_{50}上均取得最高精度，且模型大小和参数量最小，推理速度满足实时检测要求；YOLOv10虽然FPS最高，但检测精度不理想；YOLOv7和RT-DETR尽管检测精度较好，但参数量过大且FPS偏低；YOLOv8和YOLO11的检测精度不高，但YOLOv8通过添加小目标检测层P2后，检测精度显著提升.虽然在其他模型COCO等自然场景数据集上表现优异，但在本文的小目标螺栓检测任务中，实验结果显示这些模型在检测精度与实时性之间未能取得较优的平衡，可能与其网络结构更复杂、对大规模数据依赖更强有关.相比之下，YOLOv5具有轻量化、适合小目标和少类别检测的优势，因此被选为基准模型，并在其基础上进行改进，使模型在精度与实时性之间取得最佳平衡.

为了更直观地展示所提改进模型与其他模型在不同召回率水平下的精度表现，进一步绘制了精度-召回率(P-R)曲线如图10所示.

由图10可知：改进模型的P-R曲线位于最上方，特别是在高召回率区域，这表明在追求最大化检测召回率的同时，改进模型依然能维持较高的检测精度，证明了所做改进对于提高综合检测能力的有效性；相比之下，其他模型如YOLOv7和RT-DETR虽在某些点上接近改进模型的表现，但在整体曲线上未能持续保持同等水平的精度与召回率平衡.

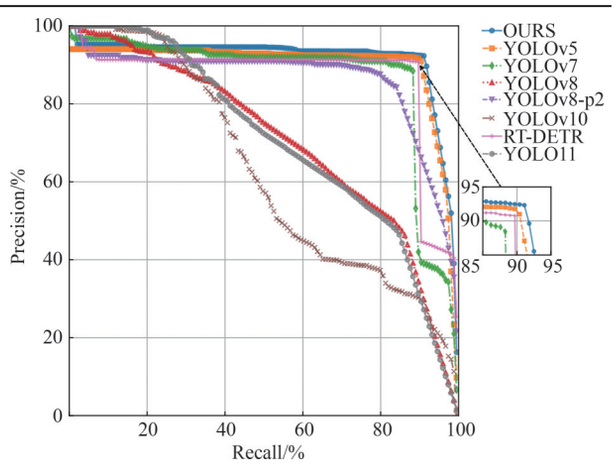


图10 不同模型的PR曲线对比

Fig.10 Comparison of PR curves for different models

2.6 检测效果对比分析

不同模型在同一缺陷位置的检测效果对比如图11所示.

由图11可知：YOLOv5、YOLOv7、RT-DETR以及本文提出的改进模型均能够准确识别出缺陷；然而，YOLOv8、YOLOv8-p2、YOLOv10以及YOLO11在检测螺栓缺失方面存在漏检问题，且YOLOv10还错误地将正常螺栓识别为有缺陷的螺栓；相比之下，本文提出的改进模型在测试集上具有较好的性能，其漏检和误检的情况相对较少.

为精确评估改进模型对两类缺陷的检测能力，其在“螺栓丢失”与“螺栓松动”两个类别上的详细性能指标如表4所示.

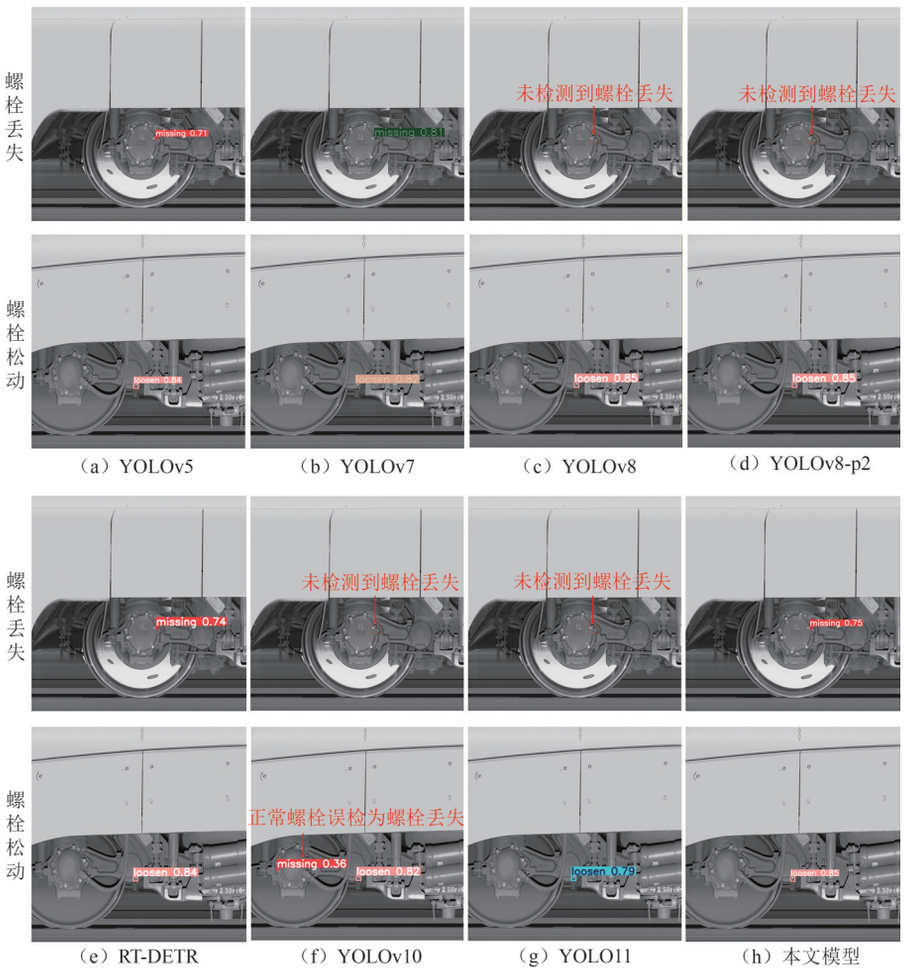


图 11　不同模型的检测效果对比

Fig.11　Comparison of detection effect for different models

表 4　改进模型对不同缺陷类型的检测性能

Tab.4　Detection performance of the improved model for different defect types　　%

缺陷类别	Precision	Recall	mAP	mAP_{50}
螺栓丢失	98.9	91.6	41.9	96.4
螺栓松动	83.1	97.6	50.5	86.6

由表 4 可知:对于螺栓丢失这类特征明显的缺陷,模型表现出极高的检测精确率;对于螺栓松动这一类别,模型则表现出极高的召回率,但其精确率相对较低,这反映了此类缺陷因特征变化大而带来的检测挑战;“螺栓丢失”的mAP_{50}更高而“螺栓松动”的mAP更高,这反映了两者在检测挑战上的差异:前者易于检测但边界框定位精细化难度大,后者虽更难初始召回但一旦检测到则定位更准.该结果证明,最终改进模型能够针对两类不同特性的缺陷,自适应地调整其性能表现,具备均衡且鲁棒的检测能力.

3　结论

1）针对转向架螺栓这种精密部件的缺陷检测问题,提出了一种基于模板匹配和深度学习的检测方法.在转向架提取方面,采用了归一化相关系数的模板匹配方法,并结合双线性插值技术对图像进行预处理,显著提升了匹配速度并减少计算开销.在缺陷检测方面,改进了 YOLOv5 目标检测模型,通过引入 CCFM 网络、增加小目标检测层以及嵌入 LSK 大核注意力机制,有效提升了模型对小目标的检测性能.

2）实验结果表明:改进后的模型在减少参数量的情况下,mAP_{50}和 mAP 分别达到了 91.5% 和 46.2%,较基准模型 YOLOv5 分别提高了 2.4% 和 3.1%,且优于其他主流目标检测模型,漏检和误检均有所降低,验证了所提方法的有效性.

在后续的工作中,将加入更多视角的采集图像和更多类别的异常,以扩大算法的应用场景并提高模型的泛化能力.

参考文献(References):

[1] 杨凯.动车组运行故障图像监测系统(TEDS)自动分类

监测作业技术研究[J]. 铁道运输与经济, 2023, 45(2): 76－82, 146.
YANG Kai. TEDS automatic classified monitoring technology research[J]. Railway Transport and Economy, 2023, 45(2): 76－82, 146. (in Chinese)

[2] ZHANG H D, YUAN X, LI D Y, et al. An effective framework using identification and image reconstruction algorithm for train component defect detection[J]. Applied Intelligence, 2022, 52(9): 10116－10134.

[3] GAO X R, WU Y Y, YANG K, et al. Vehicle bottom anomaly detection algorithm based on SIFT[J]. Optik, 2015, 126(23): 3562－3566.

[4] 刘祖胜, 方凯, 刘硕研. 用于动车组故障检测的图像识别算法[J]. 铁路计算机应用, 2015, 24(12): 1－4.
LIU Zusheng, FANG Kai, LIU Shuoyan. Image recognition algorithm for EMU trouble detection[J]. Railway Computer Application, 2015, 24(12): 1－4. (in Chinese)

[5] 耿庆华, 刘伟铭, 刘瑞康. 基于空间尺度标准化的动车组底部异常检测[J]. 铁道学报, 2022, 44(5): 67－75.
GENG Qinghua, LIU Weiming, LIU Ruikang. Anomaly detection of bottom of EMU based on space-scale standardization[J]. Journal of the China Railway Society, 2022, 44(5): 67－75. (in Chinese)

[6] 黄粤豫, 周航, 陈业泓, 等. 借助弱纹理匹配的TEDS车底故障区域定位算法[J]. 智能系统学报, 2024, 19(3): 670－678.
HUANG Yueyu, ZHOU Hang, CHEN Yehong, et al. TEDS underbody fault location algorithm in virtue of weak texture matching[J]. CAAI Transactions on Intelligent Systems, 2024, 19(3): 670－678. (in Chinese)

[7] SUN J H, XIAO Z W, XIE Y X. Automatic multi-fault recognition in TFDS based on convolutional neural network[J]. Neurocomputing, 2017, 222: 127－136.

[8] 周雯, 史天运, 李平, 等. 基于卷积神经网络的动车组行车安全图像缺陷检测与分割[J]. 铁道学报, 2019, 41(10): 76－83.
ZHOU Wen, SHI Tianyun, LI Ping, et al. Defects detection and segmentation of operation safety image of EMU based on convolutional neural network[J]. Journal of the China Railway Society, 2019, 41(10): 76－83. (in Chinese)

[9] 罗隆福, 叶威, 王健. 基于深度学习的高铁接触网顶紧螺栓的缺陷检测[J]. 铁道科学与工程学报, 2021, 18(3): 605－614.
LUO Longfu, YE Wei, WANG Jian. Defect detection of the puller bolt in high-speed railway catenary based on deep learning[J]. Journal of Railway Science and Engineering, 2021, 18(3): 605－614. (in Chinese)

[10] 秦荣荣, 高晓蓉, 罗林, 等. 基于注意力反向知识蒸馏的车轮踏面异常检测[J]. 激光与光电子学进展, 2023, 60(24): 2415002.
QIN Rongrong, GAO Xiaorong, LUO Lin, et al. Wheel tread anomaly detection based on attentional reverse knowledge distillation[J]. Laser & Optoelectronics Progress, 2023, 60(24): 2415002. (in Chinese)

[11] 丁凤霞. 基于新型对抗自编码器的铁路货车车辆异物检测算法[J]. 北京交通大学学报, 2023, 47(5): 25－33.
DING Fengxia. Foreign object detection algorithm for railway freight cars based on a novel adversarial autoencoder[J]. Journal of Beijing Jiaotong University, 2023, 47(5): 25－33. (in Chinese)

[12] ZENG Y, ZHANG P P, LIN Z, et al. Towards high-resolution salient object detection[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, 2019: 7233－7242.

[13] BRIECHLE K, HANEBECK U D. Template matching using fast normalized cross correlation[C]// SPIE Optical Pattern Recognition XII. Orlando, 2001: 95－102.

[14] 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, 2024: 16965－16974.

[15] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 8759－8768.

[16] DUAN Y L, WU X, DENG H Y, et al. Content-adaptive non-local convolution for remote sensing pansharpening[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, 2024: 27738－27747.

[17] LI Y X, HOU Q B, ZHENG Z H, et al. Large selective kernel network for remote sensing object detection[C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris, 2024: 16748－16759.

[18] KISANTAL M. Augmentation for Small Object Detection [EB/OL]. (2019-02-19) [2025-12-08]. https://arxiv.org/abs/1902.07296.

[19] WANG C Y, BOCHKOVSKIY A, LIAO H M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, 2023: 7464－7475.

[20] WANG A, CHEN H, LIU L, et al. Yolov10: Real-time end-to-end object detection[J]. Advances in neural information processing systems, 2024, 37: 107984－108011.