

引用格式: 李旭东, 杨瞻远, 周雪. 基于细粒度分类的一体化地下排水管道缺陷检测算法研究 [J]. 电子科技大学学报, 2025, 54(5): 676-689.
LI X D, YANG Z Y, ZHOU X. Research on integrated sewer pipe defect detection algorithm based on fine-grained classification[J]. Journal of University of Electronic Science and Technology of China, 2025, 54(5): 676-689.

基于细粒度分类的一体化地下排水管道缺陷检测算法研究



李旭东^{1,3}, 杨瞻远², 周雪^{2*}

(1. 电子科技大学 自动化工程学院, 成都 611731; 2. 电子科技大学 (深圳) 高等研究院, 深圳 518110;
3. 中国核动力研究设计院 先进核能技术全国重点实验室, 成都 610213)

摘要: 随着智慧城市的不断建设, 缺陷检测在地下排水管道中扮演的角色愈发重要。然而, 当前的通用目标检测方法主要面向差异较大目标的识别场景, 不能很好地解决地下排水管道缺陷检测场景下存在的缺陷类别间易混淆、缺陷等级差异小的问题。基于此, 该文首先探究通用检测方法在管道缺陷检测任务中存在局限性的原因, 从增强模型的细粒度分类性能入手, 提出了多尺度细粒度增强方法下的一体化联合学习算法, 旨在同时提高模型的缺陷分类和缺陷分级性能。在两个自建数据集 Sewer-Complete 和 Sewer-Part 上进行大量实验, 验证了该方法的有效性和泛化性, 与多个现有检测方法进行对比实验和可视化分析验证了该方法的先进性。

关键词: 细粒度分类; 特征交互; 多任务学习; 地下排水管道缺陷检测

中图分类号: TP391

文献标志码: A

DOI: 10.12178/1001-0548.2024184

Research on integrated sewer pipe defect detection algorithm based on fine-grained classification

LI Xudong^{1,3}, YANG Zhanyuan², and ZHOU Xue^{2*}

(1. School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China;

2. Shenzhen Institute for Advanced Study, University of Electronic Science and Technology of China, Shenzhen 518110, China;

3. State Key Laboratory of Advanced Nuclear Energy Technology, Nuclear Power Institute of China, Chengdu 610213, China)

Abstract: With the continuous development of smart cities, defect detection in sewer pipes has become increasingly crucial. However, current generic object detection methods are primarily tailored to recognition scenarios featuring substantial differences among objects, failing to effectively address the challenges present in defect detection within sewer pipes scenario, characterized by easily confusable defect categories and minimal disparities in defect levels. In light of this, this paper first explores the limitations of generic detection methods in the task of sewer pipes defect detection. Beginning with enhancing the fine-grained classification performance of models, a unified joint learning algorithm under the method of multiscale fine-grained enhancement is proposed, aiming to simultaneously enhance the defect categories classification and defect level classification capabilities of the model. Extensive experiments conducted on two self-constructed datasets, Sewer-Complete and Sewer-Part, validate the effectiveness and generalization ability of the proposed method. Comparative experiments and visual analysis against multiple existing detection methods further validate the superiority of this approach.

Key words: fine-grained classification; feature interaction; multi-task learning; tracking sewer pipe defect detection

地下排水管道作为城市建设中的基础设施之一, 在长期运行后可能会出现由于管道淤塞、变形等

问题引发的各类安全事故。因而, 需要对地下排水管道进行定期检测以及时规避事故的发生。目前, 管

收稿日期: 2024-07-18

基金项目: 国家自然科学基金 (62372082); 中央高校基本科研业务费 (ZYGX2024Z017); 深圳市自然科学基金 (JCYJ20240813114206010)

作者简介: 李旭东, 主要从事计算机视觉及软件方面的研究。

*通信作者 E-mail: zhouxue@uestc.edu.cn

道检测主要采用闭路电视(closed-circuit television, CCTV)检测系统和潜望镜检测方法,通过上述方法采集到实景视频或图像,然后由技术人员进行管道缺陷的判读评估,由于该类方法严重依赖专家经验,判读过程耗时耗力,不同专家判读出的缺陷结果也可能不同。根据《城镇排水管道检测与评估技术规程》(CJJ181-2012)^[1]的定义,管道缺陷共有16个类别,每个类别对应2~4个等级,也即缺陷的严重程度。定义中将管道缺陷分为功能性缺陷和结构性缺陷。功能性缺陷主要判别管道的过水截面是否发生变化,即可能影响管道通畅性能的6种缺陷,包括沉积、浮渣、结垢、树根、残墙坝根和障碍物;结构性缺陷主要判别管体结构本身是否受损,即可能影响其使用寿命、强度和刚度的10种缺陷,包括破裂、腐蚀、变形、脱节、支管暗接、起伏、错口、异物插入、接口材料脱落和渗漏。

按照评估规程,每一类缺陷根据其严重程度可划分为不同的“缺陷等级”。等级越高,代表缺陷程度越严重。如沉积有4个等级,1级代表沉积物厚度为管道直径长度的20%以下,4级代表沉积物厚度为管道直径长度的50%以上。

随着智慧城市的不断建设,管道缺陷检测逐渐向智能自动化识别的方向发展。近年来,以卷积神经网络为代表的深度学习方法在目标检测任务中获得了成功,其中不乏出现YOLOX^[2]这类优秀的检测算法,可应用于各类场景当中。地下排水管道缺陷检测方法作为通用目标检测领域的拓展应用,同样受到工业领域的青睐,是一个重要的研究方向。现有做法是直接将通用检测方法迁移到缺陷检测场景中,并利用管道缺陷数据集进行模型微调。本文地下排水管道缺陷检测任务主要包括缺陷有无检测、缺陷类别检测和缺陷等级检测3类。然而,由于部分管道缺陷类别间容易混淆且每个缺陷类别的等级间差别较小,这使得地下排水管道缺陷检测任务具有一定的挑战性。因此,如何提高检测算法对易混淆缺陷和缺陷等级的分辨能力,是解决地下排水管道场景下缺陷检测任务难题的关键。文献[3]研究表明引入细粒度图像分类算法是一个较好的解决方案。

细粒度图像分类又称细粒度分类(fine-grained classification),是一种基于细化图像特征的分类方法,其目的是对属于同一个大类别的图像进行更加细致的子类划分。由于部分类别易混淆且子类间差异细微,要想顺利地对一些极为相似的缺陷类别或等级进行细粒度识别,最重要的是在图像中定

位到能够区分这些目标的易辨别区域块,并能够对这些区域块的特征进行较好地表示。已有许多缺陷检测研究表明将细粒度分类的方法引入缺陷检测算法可以取得更好的缺陷识别效果。从不同的技术角度出发,细粒度分类大致分为3个方法:基于定位识别、基于网络集成和基于高阶特征编码的方法^[4]。基于定位识别的细粒度分类方法^[5-8]与人类区分物体的过程相似,其将细粒度图像识别分为两个部分,分别是区别性区域定位和区域中的细粒度特征学习。在区别性区域定位时,通常会以强监督或弱监督的方式利用深度卷积特征响应;而在细粒度特征学习时,则从定位到的各个区域中分别抽取特征,并将各特征组合到一起最后进行分类。基于网络集成的细粒度分类方法^[9-11]旨在将数据集划分为多个相似子集或使用多个神经网络来提高分类性能。虽然其准确率得到了较大提升,但是模型复杂度却远高于基于定位识别的方法,因此在实际应用场景中基于定位识别的方法仍然是主流。基于高阶特征编码的细粒度分类方法^[12-14]主要使用双线性池化的方法来完成更强特征的提取,后续方法大多从降秩的角度出发来简化该任务。目前基于定位识别和基于网络集成的细粒度分类方法在钢轨识别^[15]、绝缘子自爆识别^[16]、金属表面缺陷识别^[17]等缺陷检测场景中均有应用,并且均具有较好的推理效果。然而,由于管道缺陷数据集的标注难度和缺陷尺寸特点不同于其他场景下的数据集,将细粒度分类方法直接应用于地下排水管道场景下的检测任务不能获得预期效果。本文基于上述问题分析了基线算法YOLOX^[2]在地下排水管道场景下存在的问题,从提升模型对易混淆缺陷类别和缺陷等级的分类能力入手,探究了检测算法不能使地下排水管道缺陷检测任务达到最佳性能的原因。

1) 数据集标注难度大、尺寸分布与通用数据集存在差别。本文主要研究对象是管道潜望镜获取的真实地下排水管道图像及视频数据。由于本研究缺乏适用的公共数据集,因此参照《城镇排水管道检测与评估技术规程》(CJJ181-2012)^[1]所定标准,由非专业判读人员进行管道的缺陷类别和缺陷等级标注并制作了本文自建数据集。由于部分缺陷类别间易混淆、缺陷等级间差异小,因此整体数据集的标注难度较大,在标注过程中不免会出现错标的情况,造成训练时标签带有不同程度的噪声。YOLOX模型在训练期间采用SimOTA标签分配策略选择参与损失计算的正样本点,如果在匹配过程中计算

成本矩阵^[18]所用标签为带有噪声的真实 (ground truth, GT) 标签, 模型将会学习到越来越多的错误样本信息, 这使得网络向错误的方向收敛。此外, 由于检测算法采用 PAFPN 网络进行特征提取工作, 由特征金字塔网络 (feature pyramid network, FPN)^[19] 和路径聚合网络 (path aggregation network, PAN)^[20] 构成, 两者的集成极大地提升了模型对中、小尺寸目标的检测性能。由于本文数据集存在较多的大尺寸管道缺陷, 因此采用通用 PAFPN 的检测效果较差。原因在于 PAFPN 网络中浅层的特征图能够产生更多的正样本锚框, 而比较关注大尺寸缺陷的 20×20 尺度深层特征图则只能产生小数量的锚框。

2) 难以分辨易混淆缺陷类别。虽然 YOLOX 算法融合了多尺度信息, 使模型具有初步的细粒度信息分辨能力, 但对于地下排水管道场景中特征相似的缺陷, 其细粒度信息的提取能力还有待提升。

3) 缺陷分类和缺陷分级任务间存在竞争学习问题。在一阶段检测算法中, 需要同时完成回归、置信度和分类任务, 如果在分类任务输出后直接串联或并联分级任务输出头, 即使算法采取解耦输出的形式, 存在不同粒度特征的分类和分级任务也不可避免会出现相互竞争的情况。以上情况使得检测模型只能给予一种粒度关注, 而忽略另一种粒度特征的学习, 这将导致算法的识别性能下降。部分方法^[21]采用检测器和分类器级联的方式实现了先检测目标大类再细分小类的流程。然而, 这类方法却没有考虑到检测网络传递给分类网络的错误定位, 结果将会加剧整体算法的识别错误率。因此, 采用两阶段级联的方法并不能显著提高算法的细粒度分类性能。

基于上述分析, 本文对目标检测算法在管道缺陷检测任务中存在的局限性展开研究, 从增强模型的细粒度特征提取性能入手, 逐步构建了一种具备多粒度特征的多任务联合学习地下排水管道缺陷检测算法。一方面从提升模型对数据集的适应性和细粒度分类性能入手, 旨在提高缺陷分类性能; 另一方面从多任务协同学习的角度出发, 缓解了粗、细粒度的竞争学习问题, 同时提高了缺陷分类和缺陷分级性能。具体而言, 本文设计了一种多任务联合学习的一体化地下排水管道缺陷检测算法, 将其分为两个层面。从提升缺陷分类性能的层面上, 本文通过多尺度增强方法和细粒度增强方法分别缓解了

模型对数据集适应性不足和细粒度分类能力不足的问题; 从同时提升缺陷分类和缺陷分级性能的层面上, 本文针对任务间相互关联的关系提出了网络间特征交互模块, 根据基于定位识别的细粒度分类方法范式, 通过共享特征图的方式巧妙地利用了检测网络输出位置较准确的特点, 使检测网络协助分类网络较细粒度任务的学习, 这缓解了地下排水管道缺陷检测场景下粗粒度任务和细粒度任务间的竞争学习问题。与其他现有检测方法相比, 通过实验证明了本文所提方法细粒度分类能力更强、检测能力更优。

1 相关工作

本节针对地下排水管道场景下的缺陷检测方法进行概述, 将其分为基于分类、基于检测和基于分割的地下排水管道缺陷检测方法。然后, 对细粒度图像分类方法在缺陷检测领域中应用的研究现状进行综述性总结。

1.1 地下排水管道缺陷检测方法

现有通用场景下的算法适用于识别图像中特征差异显著的目标下, 在该情况下通常能得到较好的识别性能。从不同的任务角度出发, 基于深度学习的地下排水管道缺陷检测算法可分为 3 类, 即基于分类、基于检测和基于分割的地下排水管道缺陷检测方法^[22]。

基于分类的地下排水管道缺陷检测方法, 主要使用常见的分类网络如 AlexNet^[23]、ResNet^[24]、GooGleNet^[25] 等完成整张图像的缺陷特征提取并进行分类识别任务。目前常见的缺陷分类算法大致分为两类: 二元分类和多类别分类。文献 [26] 通过训练 AlexNet 用于管道缺陷分类, 模型精确度最高可达到 96%, 但是仅研究了 6 种管道缺陷类别。文献 [27] 通过将分层分类方法引入 ResNet-18 网络中, 监督不同级别任务的学习。低级检测任务用来识别图像是否带有缺陷, 高级分类任务对缺陷图像中的各缺陷类别进行分类。文献 [28] 介绍了一个包含多个二分类器的分类系统, 这些二分类器通过逐层递进的方法最终将所有类别单独分开, 但该研究仅涉及 3 类缺陷。基于分类的地下排水管道缺陷检测方法准确率虽高, 但是仅能识别出图像中最明显的一种缺陷, 而且不能准确指出缺陷的具体位置。因此, 为增强地下排水管道缺陷检测算法的适用能力, 更多研究将目光聚焦在了基于检测的方法上。

基于检测的地下排水管道缺陷检测方法,旨在使用检测框架完成缺陷位置的检测和缺陷分类任务。文献[29]通过微调YOLOv3^[30]算法,使其可端到端自动检测出管道缺陷的数量、位置和类别,对于部分类别的分类性能可达到96%,但因某些缺陷类别间易混淆且等级不易划分,使得该检测器仍具有一定的局限性。文献[31]基于SSD^[32]、YOLO^[33]和Faster R-CNN^[34]这3种检测框架开发了一种用于检测树根和沉积的缺陷检测分类和定位系统,该系统对于结构性缺陷和等级分类效果较差。文献[35]研究比较了YOLO和Faster R-CNN检测算法的性能,在5种缺陷类别下进行实验,表明Faster R-CNN的预测精确率更高。两种模型检测脱节和沉积这两种缺陷的性能最好,而对破裂和树根的检测性能相对较差。该结果有助于更好地了解经典方法的优缺点,为自动化地下排水管道缺陷检测的实际应用提供了指导。上述基于粗粒度的地下排水管道缺陷检测算法虽然能够较为准确定位并分类大部分管道缺陷类别,但在检测16个缺陷类别以及各类别对应等级的场景中却远达不到上述的高准确率。

基于分割的地下排水管道缺陷检测方法,旨在使用常见的分割框架如SegNet^[36]、U-Net^[37]、Mask RCNN^[38]等完成缺陷几何分割和分类任务。基于分割的方法在目前的水下管道缺陷检测中较少使用,该方法对人工标注要求高且实时性较差,因此在实际应用中也较少使用基于分割的方法。

1.2 基于细粒度分类的缺陷检测方法

基于细粒度分类的缺陷检测算法旨在对图像中存在细微差异的缺陷类别进行检测和分类,能够较好地处理缺陷类别之间视觉表征相似的识别问题。基于细粒度分类的算法思路是:首先找出具有明显特征的局部区域并利用CNN提取这些区域的特征,然后级联这些特征并训练分类器实现分类或检测功能。然而,由于先前大部分地下排水管道缺陷检测算法仅研究部分缺陷类别,没有涉及很多难以分辨的混淆缺陷类别,同样也没有涉及缺陷的等级,因此多数研究并不涉及细粒度图像分类算法。为更广泛地探究细粒度算法的应用,本文将细粒度分类方法的应用范围扩展到了所有缺陷检测领域。

文献[14]提出一种由粗到细的模型来识别轨缺陷,定义了3种尺度级别,包括图像级别、区域级别和像素级别,使用3种粒度共存的模型来达到检测各类缺陷的目的。文献[16]提出基于细粒度分类的绝缘子自爆缺陷检测模型,以绝缘子自爆

缺陷检测作为研究任务,将研究问题转换为对航拍图片进行细粒度分类。首先构建了一个用于自爆缺陷检测的多任务学习模型,其次基于航拍图片的标签树构建了层次分类模型,将其与多任务学习模型结合,其中层次分类模型使用B-CNN网络^[11]更加细致地区分了自爆绝缘子和正常绝缘子。文献[17]提出一种基于细粒度的表面缺陷检测方法,该方法由特征提取模块、注意力引导数据增强模块和BAP特征融合模块组成。特征提取模块采用InceptionV3网络^[39]作为骨干获取图像特征层和注意力图。将通过数据增强模块的数据和原始数据一起输入特征融合模块用于学习。特征融合模块采用双线性注意力池化融合数据特征和注意力图,使网络聚焦于关键部位从而得到最终结果。该方法在6种主流表面缺陷检测数据集上均表现出良好性能。文献[40]基于Faster R-CNN中的RPN(region proposal network)网络引入一种新的两阶段检测网络SRPN,通过利用多层全局特征,使网络具有检测所有缺陷区域并进行细粒度缺陷分类的能力。具体来说,SRPN(strengthened region proposal network)通过融合来自骨干网络的多尺度特征图生成代表性区域提案,该步骤不仅可提取用于缺陷分类的建议区域,还可将建议区域特征和相应图像的全局上下文特征结合,用于增强细粒度缺陷分类网络的特征表示,但该研究在地下排水管道场景下仅涉及5种缺陷类别。文献[41]提出适用于航天应用的细粒度密封环表面缺陷检测算法,其包含3种膨胀卷积组成的特征提取网络和特征融合树模块。首先,量化并生成参与卷积运算的原始像素热力图,根据热力图提出了使用不同膨胀率卷积组合的特征提取优化方法,在此基础上应用带有不同膨胀卷积的特征提取网络;其次,在多尺度分类和位置回归预测头之前应用特征融合树模块,保证相同尺度下不同感受野的响应特征可以被重用和压缩。该方法使检测模型在航天密封环缺陷数据集O-rings-3000中对航空航天O型圈的5种表面缺陷测试的准确率达到最高值。上述在其他缺陷检测领域中成功应用细粒度分类方法的研究给地下排水管道缺陷检测领域引入细粒度因素提供了坚实的依据。

2 多任务联合学习的一体化地下排水管道缺陷检测算法

本节介绍地下排水管道缺陷检测算法的整体框

架, 根据方法的不同, 依次介绍了细粒度增强和特征交互模块在本场景下的设计细节和应用原理。

2.1 整体框架

本文设计的多任务联合学习的一体化地下排水管道缺陷检测算法基于 YOLOX^[2] 检测框架构建, 能够同时提升模型的缺陷分类和缺陷分级性能, 如图 1 所示。整体算法框架由检测网络和分类网络组

成, 其中检测网络完成回归任务、缺陷有无识别任务和粗粒度缺陷分类及分级识别任务, 分类网络完成细粒度缺陷分类及分级识别任务。对于一张给定的图像, 检测网络的输入为该图像经过预处理之后的 3 通道张量, 分类网络的输入则来源于该图像经过真实标注 (ground truth, GT) 框裁剪后得到的目标图像组。

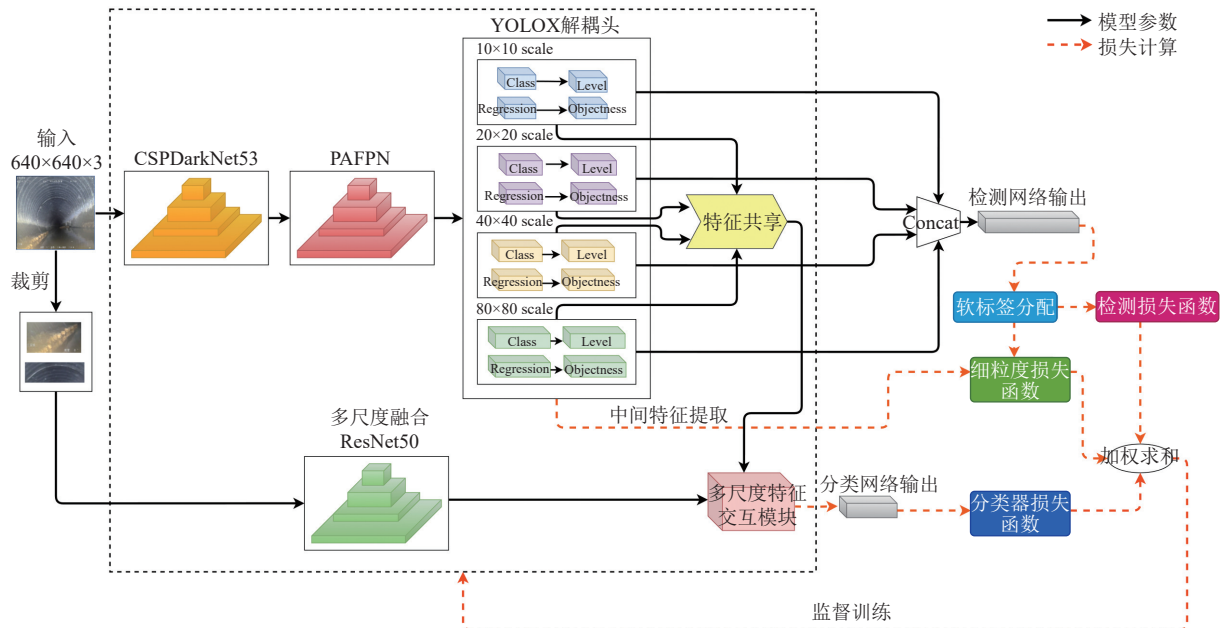


图 1 本文方法整体框架

在检测网络中, 经过预处理的图像首先经过骨干提取网络 CSPDarkNet53 和扩增尺度后的颈部网络 (path aggregation feature pyramid network, PAFPN) 用以提取并融合得到多尺度特征, 在 YOLOX 解耦头输出前, 提出网络间共享多尺度回归和分类特征图。在损失计算过程中, 根据细粒度图像分类中 MC-Loss 方法^[42], 利用在检测网络输出前的卷积特征图上施加损失的机制, 本文提出了多尺度细粒度损失监督, 通过该损失使检测网络能够感知不同尺度特征图的缺陷目标, 从而初步提升了模型的细粒度分类性能。在标签分配方面, 采用软标签分配的方式, 减小了模型训练过程中的噪声并提高了模型鲁棒性。

为解决地下排水管道缺陷检测下粗、细粒度任务间的相互竞争问题, 本文在检测网络的基础上引入了新的分类网络, 并将之同样改良为具有多尺度特性的多尺度融合 ResNet50 网络。

在分类网络中, 经过标注框裁剪后的目标图像组首先经过多尺度融合 ResNet50 网络提取缺陷类别和等级特征, 然后经过多尺度特征交互模块融合

共享特征图和分类网络提取的特征。其中, 多尺度特征交互模块参照基于定位识别的细粒度分类方法范式设计, 在模型训练的中后期, 检测网络具备更加成熟的定位和分类能力, 此时特征交互模块可以辅助分类网络完成粒度更细的缺陷分类和缺陷分级任务, 进一步提升了模型的细粒度分类性能。在损失计算过程中, 采用交叉熵损失计算分类和分级损失。算法的总体损失是检测部分损失、细粒度损失和分类部分损失的加权求和。

2.2 多尺度细粒度增强方法

检测算法对数据集是敏感的, 由于本文数据集存在标注错误率较高、缺陷类别间易混淆的问题。在网络前向传播阶段, YOLOX 算法中颈部网络 PAFPN 缺乏对大尺寸目标的关注, 同时整体算法缺乏细粒度特征的分辨、提取能力。因此本文通过扩增特征层和添加细粒度模块的方式缓解上述问题给模型训练过程带来的困难。在参数反向传播阶段, 采用软标签形式计算成本矩阵分配正样本锚点, 相比原先直接采用硬标签分配的形式, 该方法允许模型输入数据存在一定的噪声。与此同时, 利用细粒度模

块对齐不同缺陷类别的特征通道, 使得每个特征图都具备足够且不同的细粒度特征判别能力。多尺度细粒度增强方法主要由多尺度注意力增强、软标签分配策略和多尺度细粒度损失设计 3 个部分组成。

2.2.1 多尺度注意力增强

YOLOX 中颈部网络 PAFPN 虽融合了深层语义信息和浅层定位信息, 但更多关注的是 COCO 数据集中小尺寸和中等尺寸的目标。因此在 3 个特征层 $P3^{C \times 80 \times 80}$ 、 $P4^{C \times 40 \times 40}$ 和 $P5^{C \times 20 \times 20}$ 的基础上融入一层更深的小尺度特征层 $P6^{C \times 10 \times 10}$ 以增加对大目标的关注。考虑到加深 PAFPN 网络会使信息传递路径加长, 导致信息逐渐流失, 因此在每一层采样操作之后均加入空间-通道注意力机制 (convolutional block attention module, CBAM)^[43]。总而言之, $P6$ 特征层使网络更聚焦于大尺寸缺陷的语义信息, 其来源于经过卷积、下采样、注意力机制处理后的骨干网络深层特征和经过下采样、注意力机制处理后的颈部网络深层特征的融合特征层, 旨在保持 PAFPN 结构的稳定性。

2.2.2 软标签分配策略

在目标检测中, 损失函数的计算一半来源于于标签分配得到的正样本信息。考虑到在算法动态标签分配的过程中, 成本矩阵的计算非常依赖人工标

注, 因此减少算法对标注数据的依赖程度是使模型具备较强泛化性的必要措施。具体而言, 软标签的设计遵循两个原则。第一, 预测框和 GT 框越接近, 则该 GT 对应的类别标签越可信; 第二, 若 GT 类别和预测类别相差较大, 就应给予成本函数更大的权重从而加大惩罚。因此, 类别软标签的计算如式 (1)~式 (3) 所示:

$$S_i^{\text{cls}} = G_i^{\text{cls}} \times \text{IOU}_{ij} \quad (1)$$

$$\mu_i^{\text{cls}} = S_i^{\text{cls}} - P_j^{\text{cls}} \quad (2)$$

$$c_{ij}^{\text{cls}} = - \sum_{i=1}^n \sum_{j=1}^m [S_i^{\text{cls}} \times \log(P_j^{\text{cls}})] \times |\mu_i^{\text{cls}}|^2 \quad (3)$$

式中, G^{cls} 和 S^{cls} 分别表示类别硬、软标签; S^{cls} 表示类别预测; IOU_{ij} 表示 GT 框和预测框间的 IoU 得分; μ^{cls} 作用类似于 Focal Loss^[44] 中的缩放因子。

2.2.3 多尺度细粒度损失

由于检测模型在学习时仍然只关注差异较大的类别, 而忽略了差异较小类别中具有辨别性的部位和特征。为使网络在输出前就开始利用具有高语义信息的特征图, 从而让模型能关注到不同的部件级特征。本文通过扩散 MC-Loss^[42] 尺度的损失计算方式使模型更关注缺陷的不同部位特征, 损失函数细节如图 2 所示。

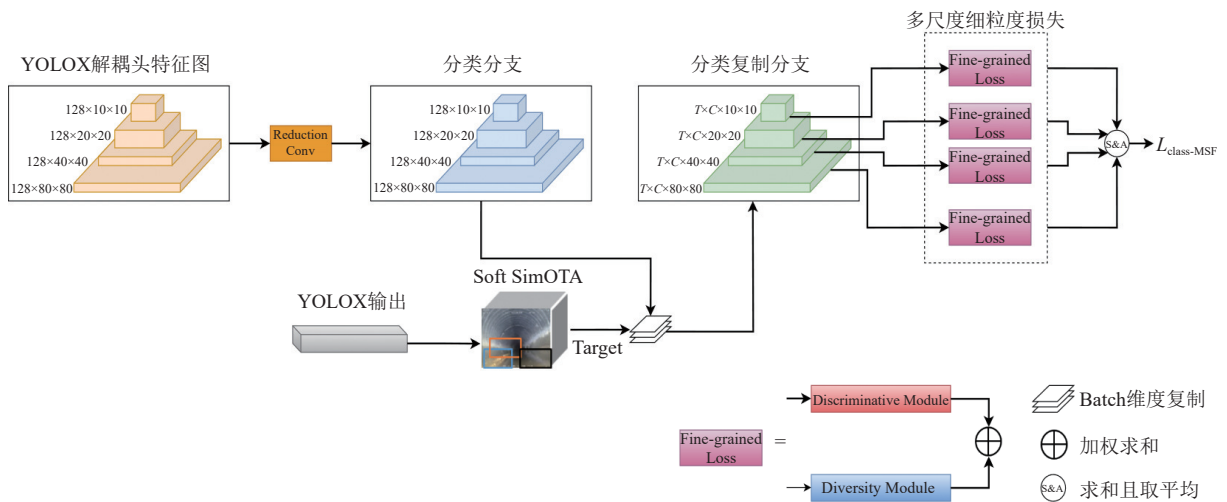


图 2 多尺度细粒度损失实施细节

图中仅展示分类分支特征图参与损失计算的细节。算法分类输出头中间层提取出的特征层为 $F^{N \times 128 \times W \times H}$, 其中 $W \times H$ 代表 4 个特征层的尺度, 之后再将特征图经过卷积从 128 通道处理为人为可设定的 C 通道, 得到特征图 $F^{N \times C \times W \times H}$ 。 C 的计算如式 (4) 所示:

$$C = c \xi_1 \quad (4)$$

式中, c 表示缺陷总类别数 16; ξ_1 表示每一缺陷类别需要分配的特征通道数。实施流程为输出分支输出后首先经过软标签分配处理, 假设每张图分配 T 个正样本预测, 将通道处理后的特征图复制 T 次并扩展到 Batch Size 维度, 特征图可分组为

$F_i \in \mathbb{R}^{T \times \xi_1 \times W \times H}$, $i = 0, 1, \dots, c-1$, 因此每一个和正样本匹配完成的 GT target 的有效相关特征图为 $F_i = \{F_{i \times \xi_1 + 1}, F_{i \times \xi_1 + 1}, \dots, F_{i \times \xi_1 + \xi_1}\}$ 。其后, 分配好的特征图进入 MC-Loss 中的判别性模块和多样性模块计算出总细粒度损失, 最后利用式 (5) 将 4 个尺度的损失做加权求和并取平均, 得到细粒度缺陷分类的平均损失:

$$L_{\text{class-MSF}} = \frac{1}{T} \left[\frac{1}{4} \sum_{s=1}^4 \eta_1 L_{\text{dis}}^{\text{cls}} + \eta_2 L_{\text{div}}^{\text{cls}} \right] \quad (5)$$

式中, η_1 和 η_2 为两个模块的预设超参数; s 表示不同特征层索引; $L_{\text{dis}}^{\text{cls}}$ 和 $L_{\text{div}}^{\text{cls}}$ 分别表示经过判别性模块和多样性模块后的损失:

$$L_{\text{dis}}^{\text{cls}}(F) = \sum_{r=0}^{T-1} \text{Loss}_{\text{SCE}} \left(t_r^{\text{cls}}, \frac{[e^{g(F_0)}, \dots, e^{g(F_{c-1})}]^T}{\sum_{i=0}^{c-1} e^{g(F_i)}} \right) \quad (6)$$

$$L_{\text{div}}^{\text{cls}}(F_i) = \frac{1}{c} \sum_{r=0}^{T-1} \sum_{i=0}^{c-1} \sum_{k=1}^{W_s H_s} \max_{j=1,2,\dots,\xi_1} \left[\frac{e^{F_{i,j,\bar{k}}}}{\sum_{\bar{k}} e^{F_{i,j,\bar{k}}}} \right] \quad (7)$$

式 (6) 的计算是为了使不同类别对应指定数量的通道, 使任意一个通道都可将该类别和其他类别

区分开。式 (7) 的计算是为了使不同类别的不同特征通道关注目标的不同区域, 通过多样化每个类别的特征通道从而减少冗余信息。

本节主要针对基线算法 YOLOX 在地下排水管道数据集中存在的问题进行了改进, 上述改进增强了检测网络的细粒度分类能力, 但是却无法解决缺陷分类与缺陷分级间的竞争学习问题。

2.3 特征交互增强方法

基于定位识别的细粒度分类方法^[45-47]大都让模型关注目标的大致范围, 之后再关注目标的判别性区域, 这种处理方式能够极大地提高分类模型的细粒度分类性能。为解决单个模型的粒度任务竞争学习问题, 在前文的基础上添加了额外的分类网络 ResNet-50, 将其改良为多尺度融合的形式, 与 PAFPN 网络保持对齐。此外, 通过设计网络间特征交互模块建立检测网络和分类网络的联系, 使得检测网络的成熟定位信息帮助分类网络进行更细粒度化的特征学习, 进一步加强了算法的细粒度分类性能。

2.3.1 多尺度特征交互模块

因本文基线为检测网络, 自身具备定位识别功能, 要在不增加分类网络复杂度的基础上提升细粒度能力, 需要融合检测网络提取的回归和分类特征。图 3 显示了检测和分类网络间的多元线性回归特征交互模块设计细节。

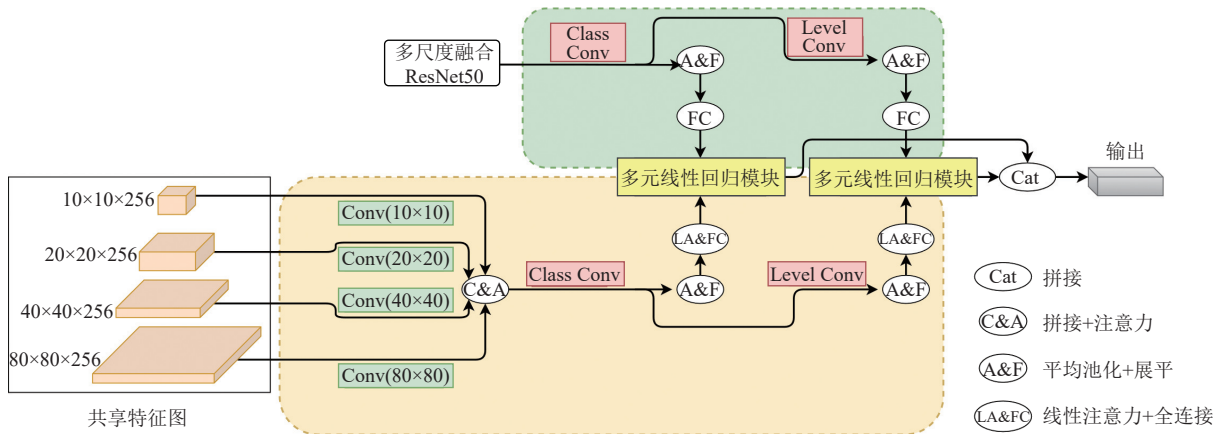


图 3 多尺度特征交互模块结构

具体而言, 检测网络输出前共有 4 层不同尺度的特征图 $S_i \in \mathbb{R}^{W_s \times H_s}$, $i = 0, 1, 2, 3$, 为统一特征维度以便后续特征融合, 利用不同数量的 3×3 卷积将特征图尺度均处理为 10×10 大小, 其后拼接为共享融合特征图 $F^{10 \times 10 \times 2048}$, 该特征图融合了检测网络获取到的多尺度定位和较粗粒度分类信息, 在多任务

联合学习模型中充当强化分类模型的作用。

由于共享特征图信息比较混杂, 为得到与细粒度分类更加相关的特征, 考虑采用类似 SE 注意力机制的压缩和激励操作完成线性注意力模块功能。

在分别得到来自共享融合特征图和分类网络的线性向量之后, 如何有效融合两个向量是决定分类

网络是否能较好地利用检测网络提供的定位、分类特征的关键一环。较好的解决方案是使融合参数成为可学习参数, 即采用多元线性回归的方式, 其相较于直接对两个向量进行加权求和的方法具备着更强的模型拟合能力。线性回归拟合过程如式 (8) 所示:

$$P_i^f = \sigma \left[\sum_{n=1}^2 l(P_i^n) \times P_i^n \right] \quad (8)$$

式中, P_i^f 是类别 i 的最终融合置信度; $l(\cdot)$ 和 σ 分别表示线性变换操作和 Sigmoid 激活函数, 线性回归拟合的目标损失函数则采用标准二元交叉熵损失。

模型间特征融合流程如下: 分类网络的输出来源于其自身和检测输出前的共享特征分支, 两个分支分别经过式 (9) 和式 (10) 的操作得到两个等长向量; 此后再经过线性回归拟合模块得到最终的细粒度缺陷分类和缺陷分级预测结果:

$$F_{cls} = l \left[\frac{1}{W_c H_c} \sum_{k=1}^{W_c H_c} \text{Conv}_i(F_i^{W_c \times H_c \times C_c}) \right] \quad (9)$$

$$F_{fuse} = l \left\{ \text{la} \left[\frac{1}{W_f H_f} \sum_{k=1}^{W_f H_f} \text{Conv}_i(F_i^{W_f \times H_f \times C_f}) \right] \right\} \quad (10)$$

式中, i 表示分类或等级分支索引; W 、 H 、 C 分别是输入特征图的宽、高、通道数; f 和 c 分别表示特征图来自于共享特征图和分类网络输出特征图; $\text{la}(\cdot)$ 代指线性注意力操作。多尺度特征交互模块利用检测网络成熟的回归定位信息, 帮助分类网络获取大致的缺陷位置信息, 使其可更加专注于细粒度分类特征的学习, 详细分析结果及数据见实验部分。

2.3.2 总损失设计

检测算法的损失 L_{det} 由分类、分级、回归和置信度损失加权求和得出, 本文在此基础上加入了细粒度损失 L_{fg} , 因此检测网络部分的损失 $L_{classifier}$ 如式 (11) 所示。细粒度损失 L_{fg} 由细粒度缺陷分类损失和细粒度缺陷分级损失组成, 如式 (12) 所示。分类网络仅进行缺陷分类和缺陷分级任务, 其损失构成如式 (13) 所示:

$$L_{det} = L_{cls} + L_{lv} + L_{obj} + \lambda L_{reg} + L_{fg} \quad (11)$$

$$L_{fg} = \lambda_1 L_{CM} + \lambda_2 L_{LM} \quad (12)$$

$$L_{classifier} = L_{class} + L_{level} \quad (13)$$

式中, λ 、 λ_1 和 λ_2 均为平衡损失的权重系数; 将细

粒度分类损失和分级损失分别简写为 L_{CM} 和 L_{LM} 。总体损失为两个网络损失的简单线性求和:

$$L_{total} = L_{det} + L_{classifier} \quad (14)$$

3 实验

本节介绍了文中所用自建数据集、评价指标和有关实验设置, 通过消融实验验证了多尺度细粒度增强方法和特征交互模块的有效性, 最后与基于现有检测方法搭建的基线算法进行对比并进行可视化分析, 证明本文所提方法能够有效提升地下排水管道缺陷检测算法的细粒度分类性能。

3.1 实验设置

本文数据集来源于横向项目中的管道图像数据, 根据数据的来源时间和标注方式, 将全部数据划分为 Sewer-Complete 和 Sewer-Part 两个数据集。其中 Sewer-Complete 将图中的全部缺陷悉数标出, 而 Sewer-Part 仅标注图中最明显的 1~3 个缺陷。两个数据集均包含全部 16 种缺陷类别图片、标签及正常管道图片, 均有 2 400 张图像。

根据任务类型, 地下排水管道缺陷检测场景可分为缺陷有无、缺陷类别和缺陷等级的检测。对于缺陷有无任务, 采用常见的 Accuracy、Recall、Precision 和 F1-score 指标。对于其他两个任务, 本文在 mAP (mean average precision) 指标基础上进行了部分改动, 即在判定正样本的条件上加入了 IoU 和中心点欧式距离的计算, 当满足两个条件时预测框才可被判定为正样本框。对于缺陷类别检测任务而言, TP、FP、FN 的判定则与类别有关, 给定一张图像, 若 GT 框对应的类别与检测框对应的类别相同, 则该类别的 TP 样本数加一, FP、FN 同理。针对 16 类缺陷计算得到平均精确率 AP、平均召回率 AR 和平均 F1 分数 AF1 之后, 再按照标注时各类别的区域数量计算权重占比, 则类均精度 C-mAP、类均召回率 C-mAR 和类均分数 C-mAF1 的定义为:

$$C-mX = \sum_{i=1}^c X_i \times \text{weight}_i \quad (15)$$

式中, X 表示 AP、AR 和 AF1 中的一个符号; i 表示 c 类, 即 16 类缺陷的索引; weight 表示参与计算的各缺陷类别的区域数量权重占比。对于缺陷等级检测任务而言, 其判定标准更加严苛, 样本需满足类别相同且等级也相同的条件才可被判定为 TP。

等级均精度 L-mAP、等级均召回率 L-mAR 和等级均 F1 分数 L-mAF1 的计算与缺陷类别检测指

标相似，其定义为：

$$L-mX = \sum_{i=1}^{c \times 4} X_i \times \text{weight}_i \quad (16)$$

式中，假设每个缺陷类别均有 4 个缺陷等级， i 表示等级索引， weight^* 代表参与计算的各缺陷等级的区域数量权重占比。本文实验结果中符号“↑”表示该项指标越高性能越好，反之则越差。

本文涉及的模型训练和推理基于 PyTorch 深度学习框架搭建，编程语言为 Python3.8，CUDA 版本为 11.3，操作系统版本为 Ubuntu-20.04。实验所用硬件设备为一块型号为 GeForce RTX 4090 的 GPU 和型号为 Intel (R) Core (TM) i7-14700KF 的 CPU。实验过程将两个数据集均采取 8:2 的比例划分为训练集和测试集。

3.2 训练与推理设置

训练过程中数据增强策略采用加灰条、缩放、长宽扭曲、左右翻转、色域变换等方法。分类网络将裁剪后的图像尺寸处理为 224×224 ，检测网络则是将图像尺寸处理为 640×640 。使用 Adam 优化器进行梯度计算，批次大小 (Batch Size) 设置为 32，迭代总次数设为 180，其他参数如学习率衰减、初始学习率等均采用官方 YOLOX 默认参数。检测模型采用 YOLOX-S，训练过程开启混合精度 FP16 模式，实验中采用随机种子初始化参数。采用冻结训练策略，其中前 120 次迭代中，冻结联合学习模型的分类网络和特征交互网络，后 60 次迭代中冻结检测网络的骨干和颈部部分，而解冻分类网络和特征交互网络。训练过程中式 (4) 中的类别通道权重 ξ_1 设为 3，等级通道权重 ξ_2 设为 2，式 (5) 中的权重系数 η_1 和 η_2 分别为 5 和 20，式 (11) 的权重系数 λ 设为 5，式 (12) 的权重系数 λ_1 和 λ_2 分别为 1 和 2。

推理过程中采用两阶段形式推理得到最终的预测结果。第一个推理阶段由检测网络完成，该阶段得到解码输出矩阵后首先过滤置信度小于 0.3 的预测框，然后用阈值为 0.1 的 NMS 后处理方法再次过滤得到检测结果。第二个推理阶段则是遍历并裁剪每一个预测框，然后将裁剪后的图片和上一阶段中的共享信息作为输入传入模型分类网络得到分类结果。最后，按照置信分数高低判断是否采纳分类网络的类别和等级结果。

3.3 消融实验

本文主要针对所提多尺度细粒度增强方法和特征交互模块进行消融实验。值得说明的是，在实验

过程中，多尺度细粒度增强方法是在一体化框架提出前使用的。为验证数据集标注方式对模型推理结果产生的影响，选择合并两个数据集的测试集作为本文的公共验证测试集，实验中从公共验证测试集随机选择一半作为验证集，另一半作为测试集，重复 5 次上述过程。实验结果如表 1 和表 2 所示，表中最优指标用黑体高亮字体显示，两个数据集最优指标分开统计，后续实验结果也采用该形式。表中第 1、5 行分别代表在两个数据集上仅采用基线算法的模型性能，在此基础上引入多尺度细粒度增强方法（即多尺度）之后，缺陷有无指标几乎无增长，但缺陷分类指标 C-mAF1 分别提升了 3.6% 和 2.4%，缺陷分级指标 L-mAF1 分别提升了 1.5% 和 0.4%，说明该方法可以有效提升模型对混淆缺陷类别组的辨别能力，但在提升模型分辨缺陷等级的性能上却比较有限。在将检测网络和分类网络联合训练（即一体化）之后，大部分指标均有下降趋势，说明直接一体化的模式会使检测网络和分类网络学习时产生竞争，原本性能尚可的检测网络会因此受到影响。而在引入多尺度特征交互模块（即交互）之后，模型在 Sewer-Complete 数据集上的缺陷有无指标 Accuracy、Recall 和 F1 分数在基线基础上分别提升了 2.0%、2.9% 和 1.4%，在 C-mAF1 和 L-mAF1 上分别提升了 5.4% 和 4.1%，这说明特征交互模块可以同时缓解不同粒度任务间与不同网络间的竞争学习问题，能够在前一个方法的基础上带来更大的缺陷分级性能提升，两个方法的协同作用使得模型的细粒度分类性能取得了极大提升。

由于本文主要对缺陷分类和缺陷分级任务进行针对性设计提升，因此后续消融实验表格中不再过多涉及缺陷的有无检测指标。

多尺度细粒度增强方法中涉及的重要超参数为 ξ_1 和 ξ_2 ，两者分别表示每个缺陷类别和每个缺陷等级应分配到的特征通道数。本小节仅探索影响细粒度缺陷分类的超参数 ξ_1 对实验结果产生的影响，表 3 显示了在仅有多尺度细粒度增强方法下， ξ_1 取值为 2、3、4、5 的实验结果比较。可以发现，当 $\xi_1 < 4$ 时，多尺度细粒度损失会过度约束少数特征通道，迫使其承载过量信息。这样的方式不仅不会提高每个特征通道的判别性，反而会使一部分判别信息丢失，从而影响细粒度分类性能；当 $\xi_1 > 4$ 时，判别性信息会保留得更多，C-mAR 指标达到了最高值，但考虑到持续增大 ξ_1 带来的模型增强效果愈发减小，而训练负担却愈发增大。因此，本文

最终将 ξ_1 设置为 4, ξ_2 消融实验同 ξ_1 , 本文不再过多赘述, 直接将 ξ_2 设置为 2。

表 1 本文改进方法消融实验 (缺陷有无指标)

数据集	多尺度	一体化	交互	A↑	R↑	P↑	F1↑
Sewer-Complete	—	—	—	86.0	87.0	95.8	91.2
	√	—	—	86.2(+0.2)	85.8(-1.2)	97.3(+1.5)	91.3(-0.1)
	√	√	—	87.2(+1.2)	88.8(+1.8)	95.4(-0.4)	92.0(+0.8)
	√	√	√	88.0(+2.0)	89.9(+2.9)	95.4(-0.4)	92.6(+1.4)
Sewer-Part	—	—	—	81.8	79.3	98.5	87.4
	√	—	—	81.0(-0.8)	79.3(+0.0)	97.3(-1.2)	87.4(+0.0)
	√	√	—	82.0(+0.2)	80.3(+1.0)	97.7(-0.8)	88.1(+0.7)
	√	√	√	85.0(+3.2)	85.1(+5.8)	96.5(-2.0)	90.4(+3.0)

表 2 本文改进方法消融实验 (缺陷分类和缺陷分级指标)

数据集	多尺度	一体化	交互	C-mAP↑	C-mAR↑	C-mAF1↑	L-mAP↑	L-mAR↑	L-mAF1↑
Sewer-Complete	—	—	—	49.4	52.1	49.9	36.1	35.9	34.5
	√	—	—	54.4(+5.0)	55.6(+3.5)	53.5(+3.6)	39.0(+2.9)	37.1(+1.2)	36.0(+1.5)
	√	√	—	51.2(+1.8)	54.5(+2.4)	50.7(+0.8)	36.6(+0.5)	35.5(-0.4)	33.1(-1.4)
	√	√	√	58.6(+9.2)	54.2(+2.1)	55.3(+5.4)	42.5(+6.4)	39.4(+3.5)	38.6(+4.1)
Sewer-Part	—	—	—	58.8	47.5	51.3	36.5	31.2	32.4
	√	—	—	62.7(+3.9)	48.7(+1.2)	53.7(+2.4)	36.8(+0.3)	31.8(+0.6)	32.8(+0.4)
	√	√	—	59.7(+0.9)	45.9(-1.6)	50.9(-0.4)	40.9(+4.4)	28.5(-2.7)	32.2(-0.2)
	√	√	√	58.1(-0.7)	54.1(+6.6)	54.9(+3.6)	42.3(+5.8)	31.6(+0.4)	34.7(+2.3)

表 3 多尺度细粒度增强方法超参数消融实验

ξ_1	C-mAP↑	C-mAR↑	C-mAF1↑
2	54.0	51.8	52.1
3	54.1	52.3	52.8
4	54.4	53.8	53.3
5	53.9	53.9	53.1

由于特征交互网络的输入分别来源于共享特征图和分类输出向量, 不同的组合结构会产生不同的效果。本节共讨论 3 种组合结构: 线性融合特征交互模块、辅助特征交互模块以及本文采用的多元线性回归特征交互模块。图 4a 为线性融合形式的特征交互模块, 来源于两个分支的特征向量经过拼接操作后, 采用线性注意力机制使模型关注有效向

量, 最终通过全连接层输出结果。这种融合形式的缺点是拼接信息量过大且无策略, 加大了训练和推理的负担。图 4b 为辅助形式的特征交互模块, 来源于共享特征图的分支仅在训练阶段参与损失函数的计算和参数更新, 推理阶段则不考虑该分支的计算, 而是直接将分类向量分支的结果作为算法分类网络的输出。这种融合形式的缺点在于推理中只能利用分类网络提供的特征信息, 两个网络只是隐式交互的关系。表 4 展示了 3 种特征交互模块在 Sewer-Complete 数据集上的实验结果。可以看出, 本文方法采用不同形式的特征交互模块均较基线算法有所提升, 但采用线性回归融合方式的模块对缺陷分类和缺陷分级的性能提升最明显, 因此本文最终采用多元线性回归融合形式构建特征交互模块。

表 4 特征交互模块设计消融实验

方法	C-mAP↑	C-mAR↑	C-mAF1↑	L-mAP↑	L-mAR↑	L-mAF1↑
无交互	51.2	54.5	50.7	36.6	35.5	33.1
线性融合	56.2(+5.0)	53.6(-0.9)	53.6(+2.9)	40.2(+3.6)	36.3(+0.8)	36.7(+3.6)
辅助交互	53.1(+1.9)	53.8(-0.7)	51.5(+0.8)	38.2(+1.6)	36.2(+0.7)	35(+1.9)
线性回归	58.6(+7.4)	54.2(-0.3)	55.3(+4.6)	42.5(+5.9)	39.4(+3.9)	38.6(+5.5)

3.4 与现有检测方法的对比实验

将本文方法与现有检测方法 YOLOv5、YOLOv8、YOLOX 和 EfficientDet 在两个自建数

据集上进行对比。由于地下排水管道缺陷检测领域具备一定的商业模式, 涉及的方法很少开源, 因此本文方法均由官方代码改写而成。训练中的

相关参数设置均与官方实现一致，实验结果如表 5 所示。

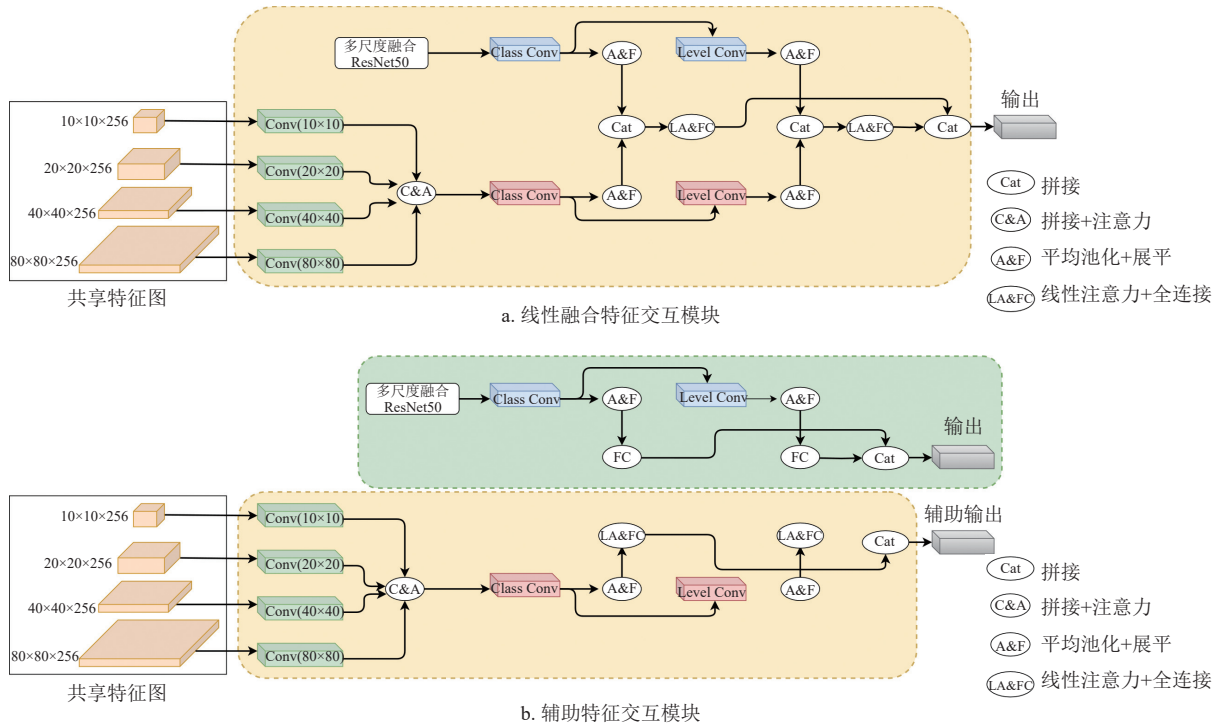


图 4 特征交互模块消融实验结构示意图

表 5 与现有检测方法对比实验

数据集	方法	C-mAP↑	C-mAR↑	C-mAF1↑	L-mAP↑	L-mAR↑	L-mAF1↑
Sewer-Complete	YOLOv5	38.3	38.8	36.9	25.3	25.5	22.6
	YOLOv8	43.8	49.3	44.7	30.1	33.6	29.4
	YOLOX ^[2]	49.4	52.1	49.9	36.1	35.9	34.5
	EfficientDet ^[48]	37.1	45.4	39.7	18.6	25	20.5
	本文方法	58.6	54.2	55.3	42.5	39.4	38.6
Sewer-Part	YOLOv5	45.2	36.3	39.1	28.6	21.6	23.1
	YOLOv8	44.7	50.7	46.8	30.1	33.6	29.4
	YOLOX ^[2]	58.8	47.5	51.3	36.5	31.2	32.4
	EfficientDet ^[48]	42.6	46.5	43.7	27.8	29.2	26.4
	本文方法	58.1	54.1	54.9	42.3	31.6	34.7

将现有检测方法与本文方法进行比较，发现本文方法分别在两个数据集上训练出的模型性能均较优，在 Sewer-Complete 数据集上训练的模型其 C-mAF1 和 L-mAF1 指标分别较基线 YOLOX 算法提升了 5.4 和 4.1 个百分点，在 Sewer-Part 数据集上训练的模型其 C-mAF1 和 L-mAF1 指标分别较基线算法提升了 3.6 和 2.3 个百分点，本文方法性能远优于其他检测方法，进一步证明本文方法能够逐步提升模型对于缺陷类别和缺陷等级的细粒度判别能力，在地下排水管道缺陷场景中具有一定的先进性。

训练过程中优化器策略采用 cosine 衰减策略，迭代次数为 180，其中前 150 次迭代冻结分类网络

和特征交互网络，后 30 次迭代冻结检测网络。训练过程的迭代曲线如图 5 所示。

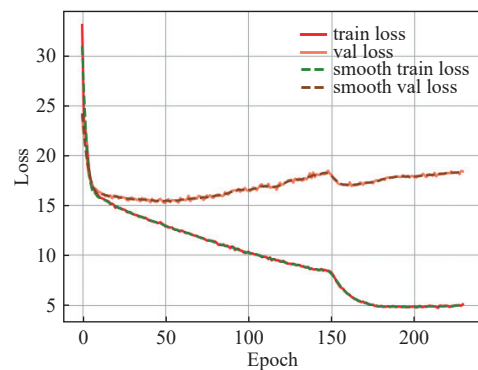


图 5 本文方法训练过程迭代曲线图

3.5 可视化分析

在本文验证集部分数据上进行本文算法与基线 YOLOX 算法的检测结果和热力图输出对比, 可视化效果如图 6 所示。图中第 1 行为带有标签的输入图像, 各列表示不同的管道场景, 图像的输出热力图代表此处是某个类别的响应值, 在热力图中响应值越高的地方越趋近于红色, 反之则趋近于蓝色。由图 6 场景 1 可以发现, 不论基线算法还是本文方法, 都有足够的力量检测得到图像中的全部缺陷。比较两个方法的热力图结果发现, 本文方法在巩固 YOLOX 性能的同时加强了对缺陷的聚焦度。由图 6 场景 2 的结果可发现, 原先算法不关注的缺陷“变形”, 在采用本文方法后, 可正常关注到相关区域, 并能以较高置信分数输出结果。值得说明的是, 图 6 场景 3 的标注是非常容易出错的, 图中 GT 框类别为“沉积”, 但有时也会被错标为“浮渣”类。基线算法辨认不出缺陷区域的真实类别, 因此输出区域有两个重叠度极高的框, 类别分别是“沉积”和“浮渣”。本文方法则是以较高的置信度输出“沉积”类并过滤了“浮渣”类, 但仍会输出一个错误类别“障碍物”, 不过通过热力图可看出算法已有抑制错误框“障碍物”附近关注度的趋势。总之, 本文提出的方法能够较大程度地提升算法的性能, 整体算法具备了一定的细粒度特征辨别能力。

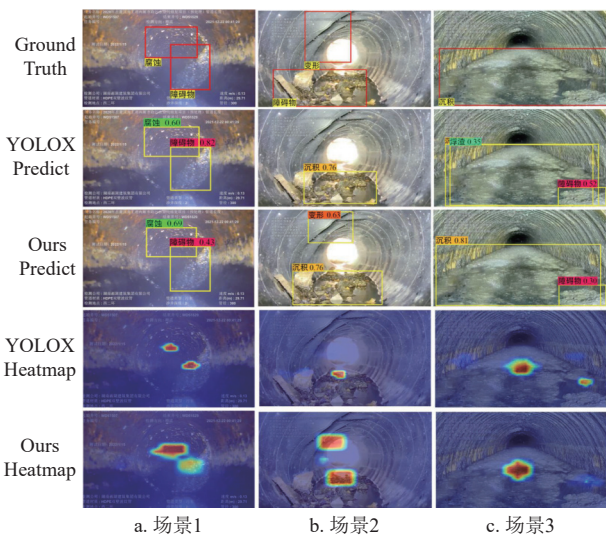


图 6 本文方法与基线方法检测结果和输出热力图对比

图 7 为本文方法与其他检测方法的检测结果可视化对比结果。在图 7 场景 4 中, GT 为“插入 3 级”, YOLOv5 和 YOLOv8 算法均出现漏检的问题, 而 YOLOX 出现误检的问题, EfficientDet 检测是正确的, 但置信度较低, 本文算法则是以较高

置信度在正确识别缺陷类别的基础上再次正确识别到该类别对应的等级, 这说明本文算法的细粒度分类能力已能够较好地完成地下排水管道缺陷检测中的分类和分级任务。图 7 场景 5 中的图像来源于 Sewer-Part 不完整标注数据集, 该图像还应有“脱落 2 级”和“错口 1 级”两个另外的缺陷标注框, 除本文方法之外的其他方法均未检测到全部缺陷区域, 本文方法则是检测到全部的缺陷区域, 并且正确识别出了这些区域对应的缺陷类别和等级。图 7 场景 6 下图像还应有“结垢 1 级”缺陷标注框, 需要说明的是这张图像左边的标注框可以是渗漏 2 级或渗漏 3 级, 右边的标注框应是渗漏 3 级, 其他方法均出现漏检和误检的问题, 但本文方法可以达到精准无误的检测。后两个场景案例证明了本文算法具有较强的综合能力和泛化性, 能够在缺少标注的情况下较为正确地检测出缺陷的位置、类别和等级。

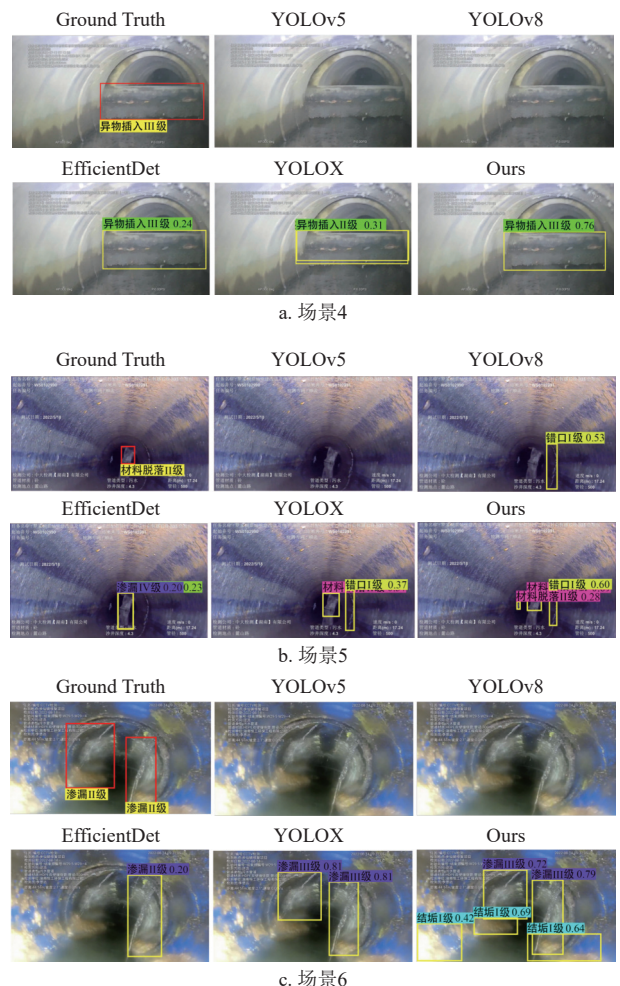


图 7 本文方法与其他方法的检测结果可视化对比

4 结束语

本文针对地下排水管道缺陷检测任务提出了一

种基于细粒度分类的一体化联合学习方法, 本方法根据自建数据集的特点和当前基线算法模型存在的问题提出一系列细粒度改进方案, 有效提升了算法的细粒度分类性能。具体而言, 本文通过利用多个尺度的特征图信息, 重新设计了细粒度损失函数, 以引导模型聚焦于缺陷区域中更具区分性的局部特征, 从而提高其对易混淆缺陷类别的辨别能力。进一步提出一体化检测网络和分类网络的框架以初步建模多粒度检测模型, 然后设计了多尺度特征交互网络, 通过共享特征图的方式巧妙利用了检测网络输出位置较准确的特点, 使检测网络协助粒度更细的分类网络进行细粒度缺陷分类和分级任务学习, 缓解了缺陷检测任务中不同粒度任务间的竞争学习问题。本文在两个自建数据集、多个先进检测方法上进行了大量实验和可视化分析, 证明本方法能够有效提升算法的细粒度分类性能, 与其他现有检测方法相比具备更强的综合能力和泛化性。本文提出的算法可实际应用于地下排水管道缺陷检测场景中, 能够更准确、智能快速地判定管道内的各类缺陷, 极大地减少了缺陷判读人员的工作量。并且及时发现问题可以优化维护策略, 降低维修成本, 延长管道使用寿命。因此本文工作对地下排水管道的维护和管理有着重要的应用价值。

虽然本文所提方法在一定程度上增强了地下排水管道缺陷检测模型的细粒度特征提取能力, 但仍有改进和其他应用的空间。首先, 当前地下排水管道缺陷检测算法仅使用检测器中原有的骨干网络来提取特征, 若采用类似 CLIP 等大模型用于特征提取可能会进一步提升模型的泛化能力; 其次是算法的速度问题, 由于加入了特征交互模块增加了算法的复杂度, 后续可以从轻量化模型的角度入手提升部署的速度; 最后, 本文算法也可应用于其他具有细粒度类别的目标检测场景, 如相似花果种类、相似鸟类的检测等。

参考文献

- [1] 王和平, 安关峰, 谢广永. 《城镇排水管道检测与评估技术规程》(cjj181-2012) 解读[J]. *给水排水*, 2014, 40(2): 124-127.
WANG H P, AN G F, XIE G Y. Technical regulations for inspection and evaluation of urban drainage pipelines(cjj181-2012) interpretation[J]. *Water Supply and Drainage*, 2014, 40(2): 124-127.
- [2] GE Z, LIU S, WANG F, et al. Yolox: Exceeding yolo series in 2021[EB/OL]. (2021-08-06)[2024-06-18]. <https://arxiv.org/abs/2107.08430>.
- [3] 邓泉, 叶蔚, 谢睿, 等. 基于深度学习的源代码缺陷检测研究综述[J]. *软件学报*, 2023, 34(2): 625-654.
DENG X, YE W, XIE R, et al. Survey of source code bug detection based on deep learning[J]. *Journal of Software*, 2023, 34(2): 625-654.
- [4] WEI X S, SONG Y Z, MAC A O, et al. Fine-grained image analysis with deep learning: A survey[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 44(12): 8927-8948.
- [5] ZHANG N, DONAHUE J, GIRSHICK R, et al. Part-based R-CNNs for fine-grained category detection[C]//Computer Vision—ECCV 2014 13th European Conference. Zurich: Springer International Publishing, 2014: 834-849.
- [6] HUANG S, XU Z, TAO D, et al. Part-stacked CNN for fine-grained visual categorization[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 1173-1182.
- [7] XIAO T, XU Y, YANG K, et al. The application of two-level attention models in deep convolutional neural network for fine-grained image classification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015: 842-850.
- [8] ZHENG H, FU J, MEI T, et al. Learning multi-attention convolutional neural network for fine-grained image recognition[C]//Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 5209-5217.
- [9] GE Z Y, MCCOOL C, SANDERSON C, et al. Subset feature learning for fine-grained category classification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015: 46-52.
- [10] GE Z Y, BEWLEY A, MCCOOL C, et al. Fine-grained classification via mixture of deep convolutional neural networks[C]//2016 IEEE Winter Conference on Applications of Computer Vision (WACV). Las Vegas: IEEE, 2016: 1-6.
- [11] WANG D, SHEN Z, SHAO J, et al. Multiple granularity descriptors for fine-grained categorization[C]//Proceedings of the IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 2399-2406.
- [12] LIN T Y, ROYCHOWDHURY A, MAJI S. Bilinear CNN models for fine-grained visual recognition[C]//Proceedings of the IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 1449-1457.
- [13] KONG S, FOWLKES C. Low-rank bilinear pooling for fine-grained classification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 365-374.
- [14] YU C, ZHAO X, ZHENG Q, et al. Hierarchical bilinear pooling for fine-grained visual recognition[C]//Proceedings of the European Conference on Computer Vision (ECCV). Munich: Springer, 2018: 574-589.
- [15] YU H, LI Q, TAN Y, et al. A coarse-to-fine model for rail surface defect detection[J]. *IEEE Transactions on Instrumentation and Measurement*, 2018, 68(3): 656-666.
- [16] 徐建军, 黄立达, 闫丽梅, 等. 基于层次多任务深度学习的绝缘子自爆缺陷检测[J]. *电工技术学报*, 2021, 36(7): 1407-1415.
XU J J, HUANG L D, YAN L M, et al. Insulator self-explosion defect detection based on hierarchical multi-task deep learning[J]. *Transactions of China Electrotechnical Society*, 2021, 36(7): 1407-1415.
- [17] 文龙, 李新宇. 深度学习[M]. 北京: 清华大学出版社, 2022, 58-65.
WEN L, LI X Y. Deep learning[M]. Beijing: Tsinghua University Press, 2022, 58-65.
- [18] GE Z, LIU S, LI Z, et al. Ota: Optimal transport assignment for object detection[C]//Proceedings of the

- IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 303-312.
- [19] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 2117-2125.
- [20] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 8759-8768.
- [21] WU Y, LI Z, CHEN Y, et al. Real-time traffic sign detection and classification towards real traffic scene[J]. *Multimedia Tools and Applications*, 2020, 79(25): 18201-18219.
- [22] LI Y, WANG H, DANG L M, et al. Vision-based defect inspection and condition assessment for sewer pipes: A comprehensive survey[J]. *Sensors*, 2022, 22(7): 2722.
- [23] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2012, 60(6): 84-90.
- [24] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770-778.
- [25] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015: 1-9.
- [26] HASSAN S I, DANG L M, MEHMOOD I, et al. Underground sewer pipe condition assessment based on convolutional neural networks[J]. *Automation in Construction*, 2019, 106: 102849.
- [27] LI D, CONG A, GUO S. Sewer damage detection from imbalanced CCTV inspection data using deep convolutional neural networks with hierarchical classification[J]. *Automation in Construction*, 2019, 101: 199-208.
- [28] KUMAR S S, ABRAHAM D M, JAHANSHAHI M R, et al. Automated defect classification in sewer closed circuit television inspections using deep convolutional neural networks[J]. *Automation in Construction*, 2018, 91: 273-283.
- [29] YIN X, CHEN Y, ZHANG Q, et al. A neural network-based application for automated defect detection for sewer pipes[C]//Proceedings of the 2019 Canadian Society for Civil Engineering Annual Conference. Montreal: CSCE, 2019: 1-8.
- [30] REDMON J, FARHADI A. Yolov3: An incremental improvement[EB/OL]. (2018-04-08)[2024-06-18]. <https://arxiv.org/abs/1804.02767>.
- [31] KUMAR S S, WANG M, ABRAHAM D M, et al. Deep learning-based automated detection of sewer defects in CCTV videos[J]. *Journal of Computing in Civil Engineering*, 2020, 34(1): 04019047.
- [32] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector[C]//Computer Vision—ECCV 2016 14th European Conference. Amsterdam: Springer International Publishing, 2016: 21-37.
- [33] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 779-788.
- [34] REN S, HE K, GIRSHICK R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. *Advances in Neural Information Processing Systems*, 2015, 28: 91.
- [35] ZHOU Q, SITU Z, TENG S, et al. Comparison of classic object-detection techniques for automated sewer defect detection[J]. *Journal of Hydroinformatics*, 2022, 24(2): 406-419.
- [36] BADRINARAYANAN V, KENDALL A, CIPOLLA R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(12): 2481-2495.
- [37] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-assisted Intervention—MICCAI 2015. Munich: Springer International Publishing, 2015: 234-241.
- [38] HE K, GKIOXARI G, DOLLÁR P, et al. Mask r-cnn[C]//Proceedings of the IEEE International Conference on Computer Vision. Honolulu: IEEE, 2017: 2961-2969.
- [39] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 2818-2826.
- [40] LI D, XIE Q, YU Z, et al. Sewer pipe defect detection via deep learning with local and global feature fusion[J]. *Automation in Construction*, 2021, 129: 103823.
- [41] LIU H, HE B, HE Y, et al. Lightweight detection algorithm for fine-grained surface defects of aerospace seal rings[J]. *Journal of Intelligent & Fuzzy Systems*, 2021, 40(6): 11423-11440.
- [42] CHANG D, DING Y, XIE J, et al. The devil is in the channels: Mutual-channel loss for fine-grained image classification[J]. *IEEE Transactions on Image Processing*, 2020, 29: 4683-4695.
- [43] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV). Munich: Springer, 2018: 3-19.
- [44] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 2980-2988.
- [45] FU J, ZHENG H, MEI T. Look closer to see better: Recurrent attention convolutional neural network for fine-grained image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 4438-4446.
- [46] SONG J, YANG R. Learning granularity-aware convolutional neural network for fine-grained visual classification[EB/OL]. (2021-03-04) [2024-06-18]. <https://arxiv.org/abs/2103.02788>.
- [47] HANSELMANN H, NEY H. Elope: Fine-grained visual classification with efficient localization, pooling and embedding[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. [S.l.]: IEEE, 2020: 1247-1256.
- [48] TAN M, PANG R, LE Q V. Efficientdet: Scalable and efficient object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 10781-10790.