

基于局部 - 全局注意和聚焦卷积的工业管道缺陷检测方法

崔文韬¹
CUI Wentao

摘要

工业管道的日常运行中因为外界扰动、腐蚀、受力不均匀等,导致管体的缺陷产生,为工业管道的安全运行带来隐患。然而,目前的管道缺陷检测方法以井下电视配合人工检查为主,费时、主观性强且成本高昂,因此需要一种自动化的缺陷检测方法。文章通过引入了一种两阶段的检测方法,并以聚焦卷积和局部 - 全局注意的方法提升网络对缺陷特征的提取和学习能力。通过重卷积的方法建立各个卷积核之间的联系,使每个卷积核都能关注到其他通道的学习的表征,增强网络的特征提取能力。通过局部 - 全局注意方法结合上下文信息,同时不缺少对细节信息的捕捉,增强网络对特征的捕捉能力,以此增强对管道缺陷的检测性能。通过大量的实验证明,所提出网络的 mAP 达到了 89.9%,在管道缺陷检测领域有良好的效果,改进了管道缺陷检测领域的自动化。

关键词

深度学习; 目标检测; 管道缺陷; 注意力机制; 重聚焦卷积

doi: 10.3969/j.issn.1672-9528.2025.05.019

0 引言

工业管道是现代工业体系中至关重要的一环。现代工业管道规模越来越大,而大部分的管道使用时间已达到 30 ~ 100 年,产生各种缺陷带来困扰。传统的基于井下电视 (CCTV)^[1]的方法直观地显示管道内壁状况,并依靠专业人员观看采集的视频来进行判定,随着工业管道系统的规模和复杂性不断增加,人工方法已很难满足需求。工业管道系统的运行和维护所需的巨大资金促使研究一种高效率、自动化的管道缺陷检测方法成为必要。

计算机视觉技术可以自动判读视频中出现的物体,辅助人工识别缺陷,节省时间和成本。其中一些利用算子的方法^[2-4],这些方法实现简单,但依赖于简单的阈值保持技术,难以应对变化的井下环境。还有一些结构学的方法,Su 等人^[5]提出基于形态学理论数据处理缺陷的理想检测模型。Khalifa 等人^[6]利用结构学结合滤波方法识别管道缺陷。但形态学的方法受限于所使用的结构元素 (SE) 参数。虽然计算机视觉技术一定程度上解决了管道缺陷检测的自动化,但这些方法缺少鲁棒性,对复杂环境检测效果不尽如人意。

机器学习技术在管道缺陷的检测领域的应用,进一步提升检测的准确性以及速度。Laurentys 等人^[7]采用模糊逻辑、遗传算法、统计算法完成管道检测。Myrans 等人^[8]用随机森林算法识别管道缺陷。然而这些方法需要人工对特定的缺陷

设计特征,只适用于特定的缺陷,限制了这些方法的适用性。

深度学习技术拥有更强的特征提取和学习能力,不再依赖于人工设计特征。拥有更高的准确率和实时性。Li 等人^[9]采用增强区域网络 SRPN 检查管道缺陷,利用多层特征融合技术进行缺陷分类。Wang 等人^[10]将 Faster RCNN 应用于管道检测领域,但这种方法并非端到端的训练模式,需要大量的耗时。Yin 等人^[11]提出以 YOLOv3 为基础的检测方法 VIASP,可以解决包括缺陷检测、视频解释、文本识别在内的任务。Chen 等人^[12]用 YOLOv5 技术实现管道检测,并用生成式网络 GAN 解决数据问题。Dang 等人^[13]利用 Transformer 技术进行管道检测。Ma 等人^[14]利用卷积神经网络融合 StyleGAN-SDM 进行管道缺陷检测,并利用数据增强原理解决数据采集困难的问题。

由于管道内部环境复杂性高、缺陷特征少、边缘纹理不清晰等,使精确检测管道缺陷存在挑战。本文提出一种名为局部 - 全局重聚焦网络 (local-global refocusing network, LGR-Net),促进深度学习在管道缺陷检测领域的应用。

本文主要贡献如下:

(1) 利用重聚焦卷积在各个卷积核之间建立联系,增强网络对管道缺陷特征提取能力。

(2) 引入了局部 - 全局注意力机制。先用局部注意力对局部信息进行关联,后使用全局注意进一步捕捉局部关联信息之间的长距离依赖关系,提高模型的性能。

(3) 构建了一个管道缺陷数据集,对所提出的方法进行

1. 武汉轻工大学 湖北武汉 430023

行验证,证实了所提出方法的有效性。

1 方法

1.1 LGR-Net 的结构

LGR-Net 的结构如图 1 所示,由输入、主干(包含卷积网络和局部-全局注意力机制(local-global attention mechanism, LGAM)、颈部(包含 FPN 和 RPN 网络)、头部(包含检测头和 NMS)以及输出 5 部分组成。输入为 $640 \text{ px} \times 640 \text{ px}$ 。主干提取目标特征。FPN 将主干提取的特征进行融合。RPN 将生成感兴趣区域的特征图传递给检测头部,生成检测结果。主干网络中所有 Bottleneck 以及颈部 FPN 网络中使用重卷积层(RefConv)^[15]替换普通卷积层。主干网络中增加 LGAM,使 LGR-Net 更全面理解输入数据,同时捕捉长距离依赖关系和细节信息。

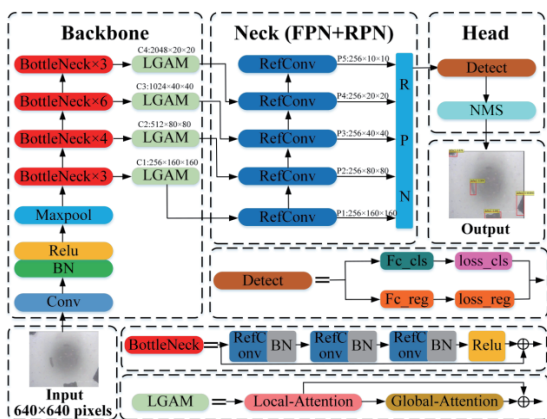


图 1 LGR-Net 网络结构图

主干网络中 Bottleneck 会生成 4 个不同大小的特征图 $C_1 \sim C_4$, 分别为 $256 \text{ px} \times 160 \text{ px} \times 160 \text{ px}$ 、 $512 \text{ px} \times 80 \text{ px} \times 80 \text{ px}$ 、 $1024 \text{ px} \times 40 \text{ px} \times 40 \text{ px}$ 、 $2048 \text{ px} \times 20 \text{ px} \times 20 \text{ px}$, 这些特征图由 LGAM 处理后,被用于特征融合过程。每个 Bottleneck 都包含三组 RefConv 和批归一化 BN。FPN 通过接收 $C_1 \sim C_4$ 融合生成 $P_1 \sim P_4$, 此外 $P_2 \sim P_3$ 通过与上一层进行下采样,保证每一层中都融合了其他尺度的信息,完成特征融合。检测头通过全连接层预测类别以及生成边界框,通过 NMS 过滤部分边界框生成最后的输出。

1.2 重聚焦卷积模块

LGR-Net 使用 RefConv 代替普通的卷积层,引入一个额外的可学习参数矩阵,通过重新参数化输入的特征图增强卷积核的特征提取能力。因此,RefConv 的卷积核可以聚焦于更广泛的区域,获得更大范围内的信息。

RefConv 的流程如图 2 所示,包括两个关键的机制:重参数化和重聚焦变换。在重参数化中,RefConv 会变换卷积层的权重,增加一个可学习的重聚焦权重 W_r ,调整并增强卷积核提取不同输入特征的能力,提升模型性能。这一过程中,

RefConv 引入的 W_r 具有可学习的能力,可以不断调整优化自身的权重,并且 W_r 还会与基本权重 W_b 进行重聚焦变换,从而增强和优化 W_b 对输入数据的适应能力。

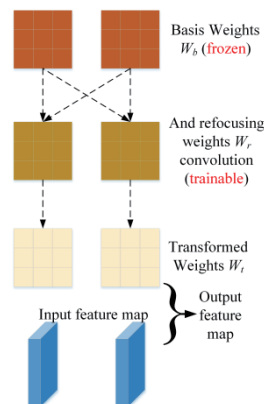


图 2 RefConv 流程图

重聚焦变换过程中,基本权重 W_b 与重聚焦权重 W_r 卷积,生成对输入特征的适应能力更强的变换权重 W_t 。这一过程中,因为卷积了基本权重 W_b 与重聚焦权重 W_r 的信息,整合了不同卷积核的跨通道信息,使新生成的变换权重 W_t ,不仅依赖于其原本输入通道中的信息,还整合了其他通道信息。因此,变换权重 W_t 多维度提取分析输入特征。提升了模型提取目标特征的能力。变换权重 W_t 的计算过程公式为:

$$W_t = W_b * W_r + W_b \quad (1)$$

1.3 局部-全局注意力机制

局部注意机制以局部信息为对象,仅依靠局部注意,会使模型对图像信息的捕捉能力减弱。全局注意虽然将对象放大到整个图像,但计算量会增加,影响模型实时性。LGR-Net 将局部注意和全局注意相结合,使用局部-全局注意机制,以较低的计算复杂度同时增强局部信息关注和全局关联。

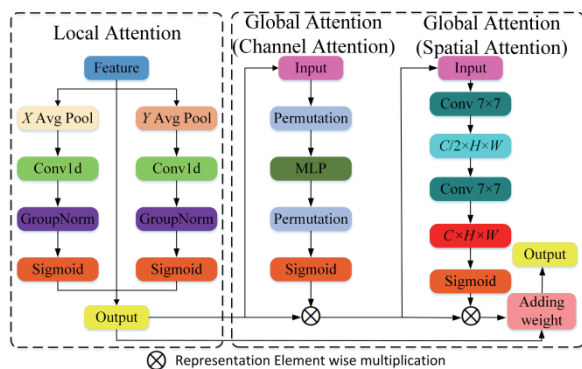


图 3 LGAM 结构图

LGAM 的结构如图 3 所示,分为局部注意和全局注意,全局注意包括通道注意和空间注意。局部注意中,沿着 X 轴的水平方向和 Y 轴的垂直方向进行局部注意计算,然后通过一个 1D 卷积层进行卷积,并执行组归一化。最后通过激活函数,将两个方向的输出合并,得到最后的输出。

在水平和垂直方向上的输出计算过程用公式分别表示为:

$$z_c^h(h) = \frac{1}{H} \sum_{0 \leq i < H} x_c(h, i) \quad (2)$$

$$z_c^w(w) = \frac{1}{W} \sum_{0 \leq j < W} x_c(j, w) \quad (3)$$

式中: $z_c^h(h)$ 和 $z_c^w(w)$ 分别表示在水平方向和垂直方向的第 h 个通道 c 和第 w 个通道 c 的平均池化输出; H 和 W 表示特征图的高度和宽度; i 和 j 表示卷积块此时的高度和宽度; $x_c(h, i)$ 和 $x_c(j, w)$ 表示卷积块的输出。

在得到 $z_c^h(h)$ 和 $z_c^w(w)$ 之后, 会经过 1D 卷积层、组归一化和激活函数处理。这一部分计算过程用公式分别表示为:

$$y^h = \sigma(G_n(F_n(z_c^h))) \quad (4)$$

$$y^w = \sigma(G_n(F_n(z_c^w))) \quad (5)$$

式中: y^h 和 y^w 分别表示水平和垂直方向上的输出; σ 表示激活函数; G_n 表示组归一化; F_n 表示 1D 卷积层; 卷积层的卷积核大小被设置为 7。

由此可以得到局部注意机制的最后输出:

$$Y = x_c \times y^h \times y^w \quad (6)$$

式中: Y 表示局部注意机制的输出; x_c 表示最初输入特征图 c 通道上的值。

全局注意对局部注意的输出进一步处理, 得到对局部信息的全局关联。全局注意中, 首先进行通道注意, 将特征图从 $C \times W \times H$ 排列为 $W \times H \times C$, 之后再排列回 $C \times W \times H$, 在两个重新排列的特征图之间通过多层感知器 MLP 连接, 捕捉不同通道之间的对应关系, 然后通过激活函数得到最后的输出和局部注意的输出进行逐元素乘法。这一过程计算公式为:

$$Y_2 = M_c(Y) \otimes Y \quad (7)$$

式中: Y_2 表示通道注意的输出; M_c 表示通道注意; $\otimes$ 表示逐元素乘法。

通道注意之后进行空间注意, 通道注意的输出会先后经过两个 7×7 的卷积层, 在两个卷积层之间特征图的通道数被减半, 因此通过两个卷积层之后, 可以捕捉到不同像素点之间的关联关系, 然后通过激活函数, 与通道注意的输出进行逐元素乘法得到最后输出。计算过程为:

$$Y_3 = M_s(Y_2) \otimes Y_2 \quad (8)$$

式中: Y_3 表示空间注意的输出; M_s 表示空间注意。

LGAM 最后将局部注意和全局注意的输出进行加权操作, 得到的输出既保留局部注意得到的局部细节信息, 也具有全局注意在局部注意输出的关键信息之间的全局关联关系。因为全局注意的计算对象并非整个原始输入, 而是局部注意得到局部细节信息, 避免了冗余计算, 保证了模型的实时性。这一过程计算公式为:

$$Y_4 = \alpha Y + (1 - \alpha) Y_3 \quad (9)$$

式中: Y_4 表示 LGAM 的输出; α 表示局部注意和全局注意之间的权重比例。

1.4 损失函数

LGR-Net 将分类损失和回归损失分别计算, 根据平衡系数将二者合并计算最后的损失。其中分类损失的计算过程为:

$$L_{cls}(y, y^{\wedge}) = -\frac{1}{C} \sum_{c=1}^C y_c \log(y_c^{\wedge}) \quad (10)$$

式中: $L_{cls}(y, y^{\wedge})$ 表示分类损失, y 表示真实类别, $y^{\wedge}$ 表示模型的预测类别; C 表示类别数目; 如果预测框的标签为正, 即 y_c 为 1, 反之 y_c 为 0; $y_c^{\wedge}$ 表示模型对类别 c 的预测概率。

而回归损失采用平滑的 L1 损失进行计算, 其计算公式为:

$$L_{reg}(y, y^{\wedge}) = \begin{cases} 0.5(y - y^{\wedge})^2 & \text{if } |y - y^{\wedge}| < 1 \\ |y - y^{\wedge}| - 0.5 & \text{otherwise} \end{cases} \quad (11)$$

式中: $L_{reg}(y, y^{\wedge})$ 表示回归损失; y 表示真实框的坐标; $y^{\wedge}$ 表示预测框的坐标。

总的损失 L 计算公式为:

$$L = L_{cls} + \gamma L_{reg} \quad (12)$$

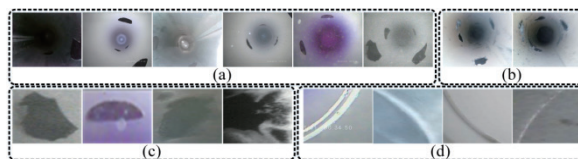
式中: γ 表示分类损失相对于回归损失的平衡系数, 用于平衡两者在总损失中的影响。

2 结果和讨论

这项工作中, 所有实验均在 Windows11 操作系统, 32 GB 内存以及 NVIDIA RTX 4070TI 显卡的计算机上进行。学习率设置为 0.01, 权重衰减率设置为 0.000 1, batch_size 设置为 4。NMS 阈值设置为 0.5, 训练 200 个 epochs。使用的深度学习框架为 PyTorch1.13.1, 编程语言为 Python3.7。

2.1 数据集构建

本文制作了包括真实管道和模拟管道的数据集。真实管道数据在石油和污水管道中采集。模拟管道数据为了减少与真实数据的差异, 设计了多种环境, 分为昏暗、明亮、强光、清水、液态浑浊以及固态浑浊 6 种环境, 如图 4 所示, 共采集 17 000 张图片。



注: (a) 5 种不同环境下的典型图片; (b) 真实工业管道环境下的典型图片; (c) defect 类缺陷典型图片; (d) hoop 类缺陷典型图片

图 4 不同环境下采集的典型图片及缺陷类型示例

表1中,分为两个类别 defect 和 hoop, 含有 3 367 个 defect 和 431 个 hoop。所有数据按照 9:1 的比例分为训练集和测试集。训练集包含 3 011 个 defect 和 369 个 hoop, 验证集包含 356 个 defect 和 62 个 hoop。用 DarkLabel 开源标注工具进行人工标注。

表1 管道数据集目标数目

数据集	训练集	测试集	总计
defect	3 011	356	3 367
hoop	369	62	431
总计	3 380	418	3 798

2.2 网络验证实验

2.2.1 重聚焦卷积与普通卷积的对比实验

为验证 Refconv 和 Conv 在 LGR-Net 中的性能表现。设计了对比实验,实验的结果如表2所示。使用 Refconv 时在 mAP 和 AP 指标上均比 Conv 有较大提升。这表明 Refconv 比 Conv 拥有更强的特征提取能力。

表2 Refconv 和 Conv 效果对比实验

方法	Conv	Refconv	mAP/% ↑	AP/% ↑
LGR-Net	√		87.8	90.6
		√	89.9	91.3

2.2.2 局部-全局注意力机制与其他注意力机制的对比实验

为验证 LGAM 相比单独使用局部注意力或者全局注意力的表现,设计了对比实验。实验的结果如表3所示。LGAM 对性能的提升大于其他方法,这表明先通过局部注意捕捉细节信息,再通过全局注意建立全局细节信息的关联是更为有效的方法。

表3 LGAM 与其他局部或者全局注意的对比实验结果

方法	mAP/% ↑	AP/% ↑	AR/% ↑
CBAM[16]	87.9	90.4	87.8
SE[17]	87.3	90.3	87.1
ELA[18]	88.6	90.8	88.1
GAM[19]	88.5	90.9	88.5
LGAM (Ours)	89.9	91.3	89.0

2.3 网络消融实验

为验证 LGR-Net 中所设计的模块对性能的贡献,对各个模块进行消融实验。结果如表4所示。LGAM 对 mAP 的改进最大达到 2.6%, Refconv 对 mAP 的改进为 1.9%。这表明所设计的模块对模型的性能都有提升。当两个模块都被添加时,对 mAP 的改进最大达到了 3.7%,这表明两个模块可以很好地相互匹配,对模型性能的提升会更加明显。在实时性方面, LGAM 的影响最大为 8.7, 而 Refconv 几乎没有影响, LGR-Net 的实时性为 46.1。

表4 LGR-Net 中各模块的消融实验结果

方法	Base	Refconv	LGAM	mAP/%	FPS/(帧·s ⁻¹)
LGR-Net	√			85.2	58.5
	√	√		87.1	55.2
	√		√	87.8	49.0
	√	√	√	89.9	46.1

2.4 LGR-Net 对各种环境下检测性能展示

为了验证 LGR-Net 在各种环境内检测的准确性,在多种复杂环境内进行实验。结果如图5所示,(a)~(g)分别为昏暗、明亮、强光、清水、液体杂质浑浊、固体杂质浑浊以及真实管道环境检测结果。在各种环境中, LGR-Net 对 defect 和 hoop 缺陷都精确识别检测,这说明 LGR-Net 对环境的改变具有适应性,不会因环境改变引起检测性能下降。

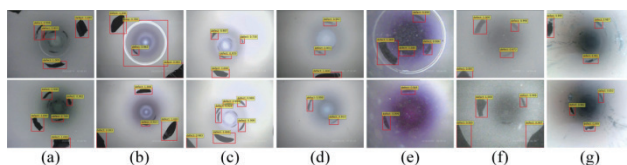


图5 LGR-Net 在各种环境下的检测结果展示

2.5 网络对比实验

为验证 LGR-Net 的先进性,将其与其他检测网络进行对比实验。实验结果如表5所示。在准确性方面 LGR-Net 的 mAP 最高达到 89.9%。在实时性方面 LGR-Net 的 FPS 性能为 46.1 帧/s,仅次于 YOLOv5 和 YOLOv7。说明 LGR-Net 可在保持效率的前提下保证对管道缺陷检测的准确性。

表5 LGR-Net 与其他检测网络的性能对比实验结果

Method	mAP/% ↑	AP/% ↑	AR/% ↑	FPS/(帧·s ⁻¹) ↑
YOLOv5	78.8	77.6	81.1	56.1
YOLOv6	79.3	78.1	81.8	50.1
YOLOv7	78.5	77.7	80.3	45.3
YOLOv8	85.6	84.4	86.9	48.6
FaterRCNN	80.7	79.9	84.3	30.3
LGR-Net	89.9	91.3	89.0	46.1

图6展示了 LGR-Net 与其他检测网络的对比实验结果,可以看出 LGR-Net 避免了其他方法的检测错误,实现了精准地识别和检测。所提出的方法克服了管道缺陷检测的困难,提升了检测的准确性。

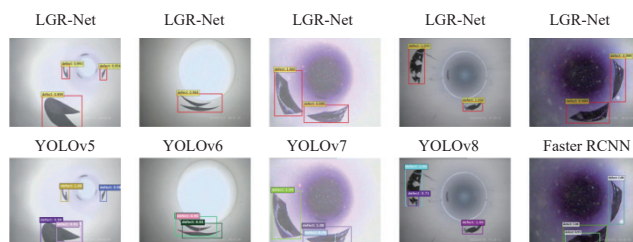


图6 LGR-Net 与其他检测网络对比实验典型结果

3 结论

本文针对工业管道缺陷的检测问题, 提出一个管道缺陷检测网络 LGR-Net, 并构建一个管道缺陷数据集。在 LGR-Net 中, 首先对管道缺陷难以提取特征的问题, 以 RefConv 代替普通的 Conv, 增强模型对管道缺陷特征的提取能力。其次, 提出 LGAM 保留细节信息并实现全局信息关联, 并且不增加模型计算复杂性。最后, 在构建的数据集上进行大量实验。结果表明, LGR-Net 的 mAP 为 89.9%, FPS 为 46.1 帧/s, 优于现在流行的检测网络。这项工作有望进一步推动深度学习在管道缺陷检测方面的应用, 使管道检查自动化进程更加完善。

参考文献:

- [1] HAWARI A, ALAMIN M, ALKADOUR F, et al. Automated defect detection tool for closed circuit television (cctv) inspected sewer pipelines[J]. Automation in construction, 2018, 89: 99-109.
- [2] GANEGEDARA H, ALAHAKOON D, MASHFORD J, et al. Self organising map based region of interest labelling for automated defect identification in large sewer pipe image collections[C//OL]//The 2012 International Joint Conference on Neural Networks (IJCNN). Piscataway: IEEE, 2012[2024-09-11]. <https://ieeexplore.ieee.org/document/6252482>. DOI: 10.1109/IJCNN.2012.6252482.
- [3] IYER S, SINHA S K. A robust approach for automatic detection and segmentation of cracks in underground pipeline images[J]. Image and vision computing, 2005, 23(10): 921-933.
- [4] IYER S, SINHA S K. Segmentation of pipe images for crack detection in buried sewers[J]. Computer - aided civil and infrastructure engineering, 2006, 21(6): 395-410.
- [5] SU T C, YANG M D, WU T C, et al. Morphological segmentation based on edge detection for sewer pipe defects on CCTV images[J]. Expert systems with applications, 2011, 38(10): 13094-13114.
- [6] KHALIFA I, ABOUTABL A E, BARAKAT G S A. A new image-based model for predicting cracks in sewer pipes[J]. International journal of advanced computer science and applications, 2013, 4(12): 65-71.
- [7] LAURENTYS C A, BOMFIM C H M, MENEZES B R, et al. Design of a pipeline leakage detection using expert system: A novel approach[J]. Applied soft computing, 2011, 11(1): 1057-1066.
- [8] MYRANS J, EVERSON R, KAPELAN Z. Automated detection of fault types in CCTV sewer surveys[J]. Automation in construction, 2018, 95: 64-71.
- [9] LI D W, XIE Q, YU Z H, et al. Sewer pipe defect detection via deep learning with local and global feature fusion[J]. Automation in construction, 2021, 129: 103823.
- [10] WANG M Z, KUMAR S S, CHENG J C P. Automated sewer pipe defect tracking in CCTV videos based on defect detection and metric learning[J]. Automation in construction, 2021, 121: 103438.
- [11] YIN X F, MA T X, BOUFERGUENE A, et al. Automation for sewer pipe assessment: CCTV video interpretation algorithm and sewer pipe video assessment (SPVA) system development[J]. Automation in construction, 2021, 125: 103622.
- [12] CHEN K, LI H T, LI C S, et al. An automatic defect detection system for petrochemical pipeline based on cycle-gan and yolo v5[J]. Sensors, 2022, 22(20): 7907.
- [13] DANG L M, WANG H X, LI Y F, et al. DefectTR: end-to-end defect detection for sewage networks using a transformer[J]. Construction and building materials, 2022, 325: 126584.
- [14] MA D, LIU J H, FANG H Y, et al. A multi-defect detection system for sewer pipelines based on Style GAN-SDM and fusion CNN[J]. Construction and building materials, 2021, 312: 125385.
- [15] CAI Z C, DING X H, SHEN Q, et al. Refconv: re-parameterized refocusing convolution for powerful convnets[J]. IEEE transactions on neural networks and learning systems, 2025: 1-15.
- [16] WOO S, PARK J C, LEE J Y, et al. CBAM: convolutional block attention module[C]//Computer Vision-ECCV 2018. Berlin: Springer, 2018: 3-19.
- [17] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 7132-7141.
- [18] XU W, WAN Y. ELA: efficient local attention for deep convolutional neural networks[EB/OL]. (2021-03-02)[2024-06-11]. <https://doi.org/10.48550/arXiv.2403.01123>.
- [19] LIU Y C, SHAO Z R, HOFFMANN N. Global attention mechanism: retain information to enhance channel-spatial interactions[EB/OL]. (2021-12-10)[2024-05-25]. <https://doi.org/10.48550/arXiv.2112.05561>.

【作者简介】

崔文韬 (1998—), 男, 河南安阳人, 硕士研究生, 研究方向: 深度学习、图像处理与工程技术。

(收稿日期: 2025-01-18 修回日期: 2025-05-19)