

基于视觉注意力机制的下水管病害识别方法

朱家松^{1,2*}, 马天柱^{1,3***}, 杨昊坤^{1,3}, 方旭^{2,4}, 李庆^{1*}

¹城市智慧交通与安全运维研究院, 广东 深圳 518000;

²深圳市人工智能与机器人研究院, 广东 深圳 518000;

³深圳大学土木与交通工程学院, 广东 深圳 518000;

⁴深圳大学电子与信息工程学院, 广东 深圳 518000

摘要 城市排水系统是城市公用设施的重要组成部分,排水管道的定期检查和维修对于地下管网的安全运行十分重要。漂流式胶囊机器人具有作业便捷、高效、低成本等特点,可满足地下管网大范围普查需求。然而,高效的作业方式带来了剧增的待处理数据,同时漂流式作业采集的视频数据包含振动、光照等多项干扰因素,传统数据处理方法无法适用,亟需开发新的智能病害识别方法。基于此,提出了一种基于改进残差注意力的病害识别方法。该方法以视频片段为输入,利用卷积神经网络对每一帧进行特征提取,然后将不同层沿特定维度融合后进行分类识别。实验结果表明,所提方法的识别准确率可达 89.6%,优于未改进的残差网络,可有效提高漂流式胶囊机器人的识别精度和效率。

关键词 机器视觉; 注意力机制; 下水管道; 病害识别; 图像处理

中图分类号 TP274+.5

文献标志码 A

DOI: 10.3788/LOP202259.1815001

Detection Method of Downpipe Diseases Based on Visual Attention Mechanism

Zhu Jiasong^{1,2*}, Ma Tianzhu^{1,3***}, Yang Haokun^{1,3}, Fang Xu^{2,4}, Li Qing^{1*}

¹Institute of Urban Smart Transportation & Safety Maintenance, Shenzhen 518000, Guangdong, China;

²Shenzhen Institute of Artificial Intelligence and Robotics for Society, Shenzhen 518000, Guangdong, China;

³College of Civil and Transportation Engineering, Shenzhen University, Shenzhen 518000, Guangdong, China;

⁴College of Electronics and Information Engineering, Shenzhen University, Shenzhen 518000, Guangdong, China

Abstract Urban drainage system is a crucial part of urban public facilities; thus, the regular inspection and maintenance of drainage pipe is essential for safely operating underground pipe network. The drifting capsule robot developed by the project team is characterized by its convenient operation, high efficiency, and inexpensiveness, which meet the requirements of large-scale survey of underground pipe network. However, the high-efficiency operation mode results in a huge amount of data that needs to be processed. Simultaneously, video data collected by drifting operation contains several unwanted features, such as vibrations and illumination, thus traditional data processing methods are unsuitable. Therefore, there is an urgent need to develop new intelligent disease recognition methods. This study presents a disease identification method based on an improved residual attention network. This method considered video clips as input, used convolutional neural networks to extract the features of each frame, and then fused different layers along specific dimensions for classification and recognition. Experimental results show that the improved method can achieve an accuracy of 89.6%, better than unimproved residual network, and effectively improve the recognition accuracy and efficiency of the drifting capsule robot.

Key words machine vision; attention mechanism; sewer pipe; defect recognition; image processing

收稿日期: 2021-06-11; 修回日期: 2021-07-04; 录用日期: 2021-07-20

基金项目: 国家自然科学基金(41871329)、深圳市水务科技信息中心科技项目(0832-SFCX19SZC020)

通信作者: *qingli@szu.edu.cn; **zjsong@szu.edu.cn; ***1910473004@email.szu.edu.cn

1 引言

城市排水系统是处理和排出城市污水和雨水的工程设施系统,是城市公用设施的重要组成部分,在维护环境健康、城市安全方面发挥着重要作用^[1-3]。地下排水管道的开裂、腐蚀、淤塞等因素不仅会影响城市的雨水和日常污水排放,在极端情况下还会出现积水渗漏、环境污染、地面下陷等问题。因此,定期检查排水管道对于避免下水道管线恶化至关重要。

在管道病害检测领域,目前主流的方法是闭路电视(CCTV)^[4-5]。此法检测时需将管内积水抽干,再使用装有摄像头的爬行器对下水管道的内表面进行录像。检查员检查录制的视频,然后手动识别各类病害。由于清空积水费时费力、影响居民生活,且手动检测依赖于检查人员的主观评估,容易出错^[6]。受肠胃检测胶囊的启发,漂流式管道检测方法被提出。此法很好地改进了 CCTV 检测方法在此方面的不足,它是一种流体驱动的排水管道检测方法,设备在污水中漂流式行进,检测时无须将水抽干,遇到阻碍时亦可及时回收,作业效率达 CCTV 检测方法的数倍,更加适用于大规模管网检测的需求。

然而,漂流式检测方法采集的数据有着独特的特点。一方面,因其操作简便、作业快捷的设备特性,检测机器人可以快速将管道内部信息采集,在短时间内容易累积大量数据。因此,数据量大且集中是它的重要特点。另一方面,机器人在漂流中进行拍摄,光照变化、污水的流速、自身的运动都易使设备因受力不均而产生拍摄模糊、图像变形、残缺等问题,这是此数据的第 2 个显著特征。这些特征给下水管网病害的识别带来了新的挑战,其数据处理也成为了新的研究热点。

许多研究人员将深度学习应用于城市管网基础设施病害检测与健康状态评估^[7-8]。Hassan 等^[9]提出了基于卷积神经网络(CNN)的排水管网病害分类方法,该方法可识别出 6 种病害类型,且取得了高达 96% 的检测精度。其他研究者针对排水管自动化检测任务,在深度卷积神经网络中引入了层次分类的方法,来监督不同级别的学习过程,提高了深度卷积神经网络在不平衡数据集下的分类性能^[10-17]。Tian 等^[18]通过更快的基于区域的卷积神经网络模型(Faster R-CNN)^[19]开发了一种基于深度学习的检测方法,其 mean average precision(mAP)为 83%。

以上研究都是基于 CCTV 的视频数据进行的,针对漂流式管道机器人采集的数据的研究较少。漂流式检测方法虽然有成本低、作业便捷等优势,但数据量大,质量差且受到抖动噪声干扰导致传统的基于 CCTV 数据的处理算法并不适用,亟须一种适合其数据特点的方法来解决这个难题。

综上所述,基于数据采集方式不稳定而造成的图像之间时间和空间相关性被切断及因图像模糊造成的

识别准确性降低这些特点,本文提出了一种高效的基于视觉注意力机制的下水管病害自动识别方法,并将其应用于胶囊管道视频数据来提高分类准确率,同时在此过程中探究了图像融合方法在注意力残差神经网络中的效果。

2 残差注意力网络

2.1 注意力机制

注意力机制在自然语言处理、统计学习及图像识别等多个领域都表现出良好的效果,并成为深度学习中一个非常重要且有效的技巧^[20]。该机制源于人类的感知机制,人类会对复杂环境下的重要信息进行周边或者相关信息的整合来达到记住关键信息的目的。该机制的原理也是如此,它可以帮助模型给输入的各个部分赋予不同的权重,达到即使在模糊状态下也可提取出重要信息的效果,使模型作出更加准确的判断,同时不会对模型的计算和存储带来更大的消耗。这种方法既简单,又可以赋予模型更强的辨别能力。

检测机器人在漂流过程中,管内光线变化、污水的流速及设备抖动造成的拍摄模糊、图像变形会使图像丢失部分信息。如果覆盖一个病害的多张图像信息丢失过多就会致使图像的深层特征不能被融合,识别的准确率将显著降低。这会对于后续的图像检测与病害识别造成较大的困扰。而注意力机制的引入,可将深层信息进行整合,使其信息学习更加丰富。传统的检测方法以单帧图片作为输入,对于模糊的胶囊数据来说,图像之间的信息丢失严重,效果极差^[20]。而本实验组将胶囊数据以图像序列的形式输入模型,并采用注意力机制网络来提取出多幅模糊图像的共性特征与深层信息,以此来改善振动、光照影响导致的图像模糊,为胶囊数据的处理带来了新的研究方法。

2.2 残差注意力网络

本实验组引入残差注意力神经网络用于病害识别,该网络是通过堆叠多个可生成注意力感知功能的模块构建而成的。随着模块的不断深入,来自不同模块的注意力感知功能会自适应地发生变化。网络中每个模块都由两部分组成,一部分是主干分支,主要负责处理图像特征;另一部分是掩膜分支,负责生成注意力信息(即权重)。该网络同时使用激活前的 Residual Unit(可解决因深度增加而导致性能下降的问题)、ResNeXt(包含的每个分支的拓扑结构相同,可实现分组卷积)和 Inception(包含多个卷积核,可提取图像不同尺度的信息进行融合)作为残差注意力网络基本单元构建注意力模块。

将主干分支的输入设为 x , 输出为 $T_{i,c}(x)$, 而掩膜输出设为 $M_{i,c}(x)$ 。掩膜分支内部是一种 bottom-up 和 top-down 注意力相结合的结构,用来学习输出与主干大小相同的掩膜,可将其视为 $T_{i,c}(x)$ 的软权重。故注

注意力模块的输出 $H_{i,c}(x)$ 的表达式为

$$H_{i,c}(x) = M_{i,c}(x) * T_{i,c}(x), \quad (1)$$

式中: i 表示空间位置; c 是表示通道的下标; $*$ 代表卷积。

在注意力模块中, 注意力掩膜因其特殊结构既可作为向前反馈的特征选择器选择特征, 又可以作为向后传播的梯度更新器来更新梯度, 其梯度可表示为

$$\frac{\partial M(x, \theta) M(x, \theta)}{\partial \phi} = M(x, \theta) \frac{\partial T(x, \theta)}{\partial \phi}, \quad (2)$$

式中: θ 是掩膜分支的参数; ϕ 是主干分支的参数。

为了改善单一堆叠的注意力模块导致性能下降的问题, 应用了改善的注意力模块, 则

$$H_{i,c}(x) = [1 + M_{i,c}(x)] * F_{i,c}(x), \quad (3)$$

式中: $M_{i,c}(x) \in (0, 1]$, 如果该值接近于 0, 则 $H_{i,c}(x)$ 近似于 $F_{i,c}(x)$ 。

堆叠式注意力模块通过其增量性对注意力残差学习特征进行备份, 因注意力残差学习具有保持原始特征的良好性能, 通过该方法, 整个残差注意网络随着深度的增加, 学习性能可以得到持续改善。所提残差注意力模型如图 1 所示。

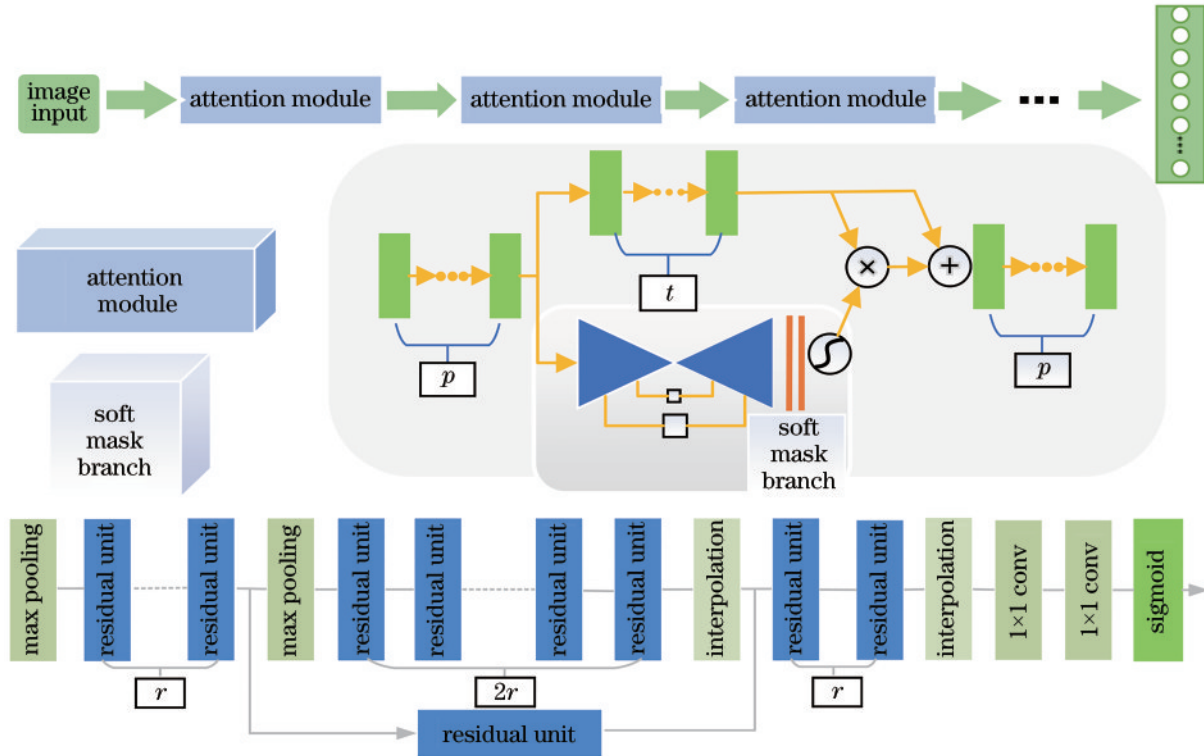


图 1 残差注意力模型网络图

Fig. 1 Residual attention model network diagram

图 1 中, p, r, t 为残差注意力模型中的 3 个超参数。超参数 p 表示在划分为主干分支和掩膜分支之前的预处理剩余单元数。 r 表示掩膜分支中相邻池化层之间的剩余单元数。 t 表示主干分支中的剩余单元数。实验中, 设 $p = 1, t = 2, r = 1$ 。

由于管道内部视频背景相似, 仅在病害处存在区别, 网络使用不含其他限制的混合注意力 f_i 对每个图像的通道和空间位置采用 sigmoid 函数作为激活函数, 来获取与其相关的通道及位置信息。

$$f_i(x_i, c) = \frac{1}{1 + \exp(-x_{i,c})}. \quad (4)$$

实验中选取交叉熵作为模型的损失函数, 即

$$L = \frac{-1}{N} \sum_{i=1}^N \sum_{c=1}^M y_{i,c} \log p_{i,c}, \quad (5)$$

式中: N 表示样本总数, 即图像序列的数量, 在训练集

中为 5600; M 表示类别总数, 在本实验中为 4; $y_{i,c}$ 表示样本 i 的标签, 如果与类别 c 相同则为 1, 否则为 0; $p_{i,c}$ 表示模型预测观察到的样本; i 为属于类别 c 的概率。

3 实验设计与分析

3.1 实验设计

3.1.1 数据集

所用数据来源于 2018 年安徽省合肥市的某住宅小区内总长度约 7 km 的污水及雨水下水管道。该小区地下水管的管道主要由聚氯乙烯材料制成, 其直径从 300~600 mm 不等, 其中直径 300 mm 和 400 mm 的占 86%。胶囊机器人在漂流前进的过程中录制管内信息, 并保存为 mp4 格式的 1920×1080 视频。

采集的视频中大部分信息位于鱼镜头图像的中心且包含着很多不相关的背景, 所以实验中把采集的

图像在中心切割成 1040×980 的尺寸,以消除不相关的背景。原始视频文件为每秒 24 帧,抽取每帧图像进行标记并制作成数据集,筛选后的数据集包含 1100 个下水道图像序列,每个序列由 8 个连续帧组成,共计 8800 个图像。

在这些序列中,有 825 个包含病害的(裂纹、变形、腐蚀各 275 个)和 275 个无病害的。由于采集数据的鱼眼畸变,图像的实际有效部分以圆圈的形式呈现。标记病害的原则是在一个序列的 8 幅图像中,如果其中至少有一幅图像存在某种病害,则该序列被贴上该病害的标签。每个序列分配 1 个标签,代表序列中最严重的病害。当一个序列中的所有图像都没有病害时,该序列被标记为无病害。

为了提高泛化能力,实验前在数据集上执行图像增强。将原始的数据旋转 90° 、 180° 和 270° ,然后对图像进行镜像变换,最终获得 8800 个视频。同时将数据在输入网络之前缩放为 224×224 。数据集被随机分

成训练集、验证集和测试集,分别具有 5600、1600 和 1600 个图像序列。

3.1.2 实验设备

本实验在 Ubuntu 14.04 的操作系统上进行,该操作系统使用 Intel® Xeon E5-2680v2 处理器、4 个 NVIDIA RTX2080Ti 11 GB GPU 和 128 GB DDR4RAM。

3.1.3 实验方法

融合不同尺度的特征是提高分割性能的一个重要手段。众所周知在图像特征中,低层特征分辨率较高,也包含更多位置、细节信息,但是由于经过的卷积更少,语义性更低,噪声更多^[10]。而高层特征具有更强的语义信息,但同时面临着分辨率低,对细节的感知能力较差等特点。

本实验组在浅层融合(shallow fusion)、中层融合(middle fusion)、底层融合(bottom fusion)中依次选择在第 1 层、第 3 层及最后 1 层分别进行融合,如图 2 所示。

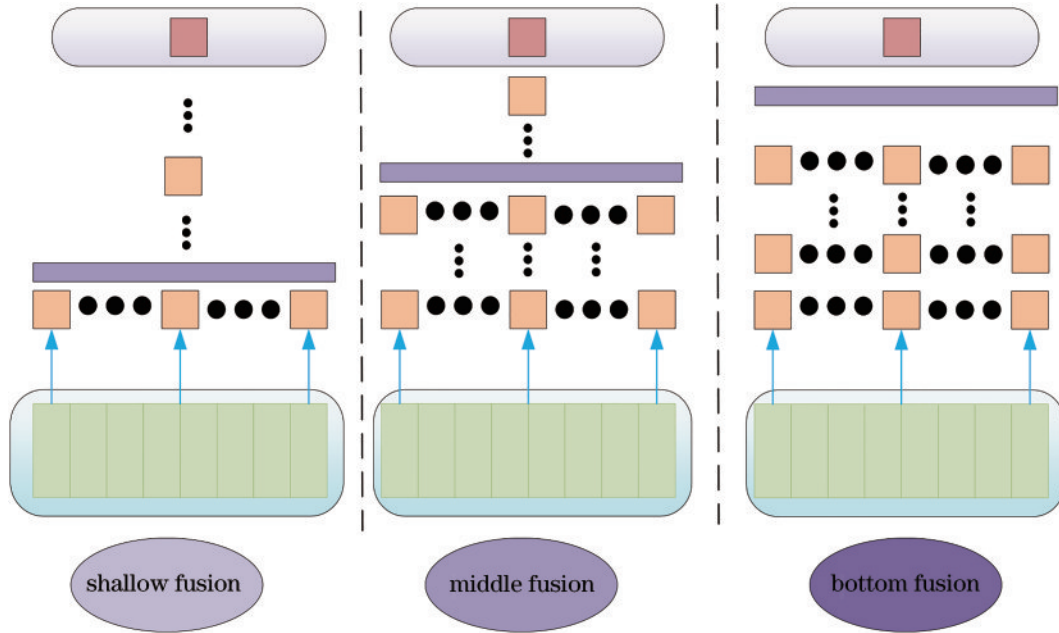


图 2 融合策略示意图

Fig. 2 Schematic diagram of fusion strategy

本实验以 8 frame/s 的速度将原始数据输入网络,并采用平均融合策略分别进行浅层融合、中层融合及深层融合,为进一步探究注意力机制的有效性,还进行了将胶囊数据输入同等深度及条件下,结构相似的不加注意力机制的 ResNet50 中的对照实验。

3.2 评价指标

为了评估分类结果,使用常用的准确度(accuracy)、正确率(precision)、召回率(recall)和 F1 作为评价指标,其表达式分别为

$$R_{\text{accuracy}} = \frac{N_{\text{TP}} + N_{\text{TN}}}{N_{\text{TP}} + N_{\text{FN}} + N_{\text{FP}} + N_{\text{TN}}}, \quad (6)$$

$$R_{\text{precision}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}}}, \quad (7)$$

$$R_{\text{recall}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}}, \quad (8)$$

$$S_{\text{F1}} = 2 \times \frac{R_{\text{precision}} \times R_{\text{recall}}}{R_{\text{precision}} + R_{\text{recall}}}, \quad (9)$$

式中: N_{TP} 为被模型预测为正类的正样本数量; N_{TN} 为被模型预测为负类的负样本; N_{FP} 为被模型预测为正类的负样本; N_{FN} 为被模型预测为负类的正样本。

3.3 实验分析

3.3.1 注意力残差网络适用于胶囊检测机器人数据
表 1、表 2、表 3 分别展示了引入注意力残差神经网络

络后不同融合策略下的表现结果。从表中数据可以看出,该网络在识别胶囊数据中的破损(Broken)、腐蚀(Corrosion)、变形(Deformation)这几种病害的效果较好,其各项指标均可达 70% 以上,对于无病害的情况也能准确识别。这表明所提方法能够正确识别三种类型的病害和正常图像。

表 1 浅层融合策略下不同病害的结果
Table 1 Results of different defects under shallow fusion strategy

Style	Precision	Recall	F1	Number of samples
Broken	0.82	0.78	0.80	400
Corrosion	0.89	0.90	0.90	400
Deformation	0.88	0.86	0.87	400
Normal	0.92	0.97	0.95	400

表 2 中层融合策略下不同病害的结果
Table 2 Results of different defects under middle fusion strategy

Style	Precision	Recall	F1	Number of samples
Broken	0.82	0.73	0.77	400
Corrosion	0.88	0.94	0.91	400
Deformation	0.84	0.80	0.82	400
Normal	0.89	0.97	0.93	400

表 3 底层融合策略下不同病害的结果
Table 3 Results of different defects under bottom fusion strategy

Style	Precision	Recall	F1	Number of samples
Broken	0.86	0.86	0.86	400
Corrosion	0.88	0.91	0.89	400
Deformation	0.87	0.83	0.85	400
Normal	0.98	0.99	0.98	400

此外,表 1~3 中的数据表明,无论是哪种融合策略,腐蚀(Corrosion)这一病害的各项指标明显高于其他两类病害,这表明此网络对于腐蚀的识别度较好,而对变形(Deformation)和破损(Broken)识别度相对较低,性能略有下降。特别是对于破裂,在中层融合中的召回率仅为 0.73,究其原因可能是两者共存,当变形足够严重时,可能会造成破裂。同样,一些破裂病害中也不可避免地伴随着变形。

3.3.2 底层融合策略

表 4 为不同融合策略的结果。在底层融合实验中,得到的整体精度约为 0.896、平均召回率可达 0.896、F1 值达到 0.895 的结果,三项指标都明显优于浅层融合和 中层融合,这表明高层特征融合适用于此数据,该策略可提取更多、更强的语义信息,分辨率低,对细节的感知

表 4 不同融合策略的结果

Table 4 Results of different fusion strategies

Neural network	fusion strategy	F1	accuracy	recall
Residual	Shallow fusion	0.860	0.859	0.859
Attention	Middle fusion	0.878	0.879	0.878
Network	Bottom fusion	0.895	0.896	0.896

能力较差等不足在此不能明显影响实验结果。

3.3.3 对于胶囊数据,残差注意力神经网络的适用性优于 Resnet50

底层融合后效果表现更好,表 5 为底层融合策略下不同神经网络的表现结果。从表 5 可以看出,在同等条件下,所提网络与 Resnet50 网络相比,各项指标也明显要高,验证了残差注意力网络对于胶囊数据的适用性且改善后的效果确实更好。

表 5 底层融合策略下不同神经网络的结果

Table 5 Results of different neural networks based on bottom fusion strategy

Neural network	F1	Accuracy	Recall
Residual Attention Network	0.895	0.896	0.896
Resnet50	0.874	0.876	0.875

4 结 论

探究了注意力神经网络在胶囊数据上的适用性及如何提高检测性能,并通过特征融合的方法,对神经网络内部层次融合策略进行了深入的探讨。本研究为管道病害检测提供了更为高效准确的检测方法,该方法可对漂流式胶囊管道机器人采集的数据进行处理,并能够较为准确地检测出污水管道的四种状态,即破裂、腐蚀、变形及正常。除此之外,所提方法对于漂流式检测机器采集的数据具有一定的针对性,可高效快捷地处理并完成下水管道病害检测,具有一定的实用性和工程价值。

实验结果表明,所提智能检测分析方法效果较为理想,但受限于学习样本量不足仍有一定限制。用于实验处理的数据仅来自于某一小区且管道的材质较为相似,多为聚氯乙烯材料,此类材料所包含的病害和其他材料的管道所含病害存在差异,且更大的样本量可能会包含更多的信息,对病害识别的准确性有一定的提升,故对于方法的普适性需要进一步的探究和对比实验。随着胶囊检测新技术的应用推广,本研究将获取更多实测及标定数据以扩大训练样本数,以求得包含更多的病害信息的数据,有望进一步提升所提方法的效率与准确性。

致谢 本文由深圳市人工智能与机器人研究院资助完成,在此表示诚挚的感谢。

参 考 文 献

- [1] 王鸣霄, 范娟娟, 周磊, 等. 基于深度学习的排水管道缺陷自动检测与分类[J]. 给水排水, 2020, 56(12): 106-111. Wang M X, Fan J J, Zhou L, et al. Automatic detection and classification of sewer defects via deep learning revoke[J]. Water & Wastewater Engineering, 2020, 56(12): 106-111.
- [2] Yang M D, Su T C. Automated diagnosis of sewer pipe defects based on machine learning approaches[J]. Expert Systems With Applications, 2008, 35(3): 1327-1337.
- [3] Cha Y J, Choi W, Büyüköztürk O. Deep learning-based crack damage detection using convolutional neural networks[J]. Computer-Aided Civil and Infrastructure Engineering, 2017, 32(5): 361-378.
- [4] Chen F C, Jahanshahi M R. NB-CNN: deep learning-based crack detection using convolutional neural network and naïve Bayes data fusion[J]. IEEE Transactions on Industrial Electronics, 2018, 65(5): 4392-4400.
- [5] Xue Y D, Li Y C. A fast detection method via region-based fully convolutional neural networks for shield tunnel lining defects[J]. Computer-Aided Civil and Infrastructure Engineering, 2018, 33(8): 638-654.
- [6] 单倩文, 郑新波, 何小海, 等. 基于改进多尺度特征图的目标快速检测与识别算法[J]. 激光与光电子学进展, 2019, 56(2): 021002. Shan Q W, Zheng X B, He X H, et al. Fast object detection and recognition algorithm based on improved multi-scale feature maps[J]. Laser & Optoelectronics Progress, 2019, 56(2): 021002.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] Li D S, Cong A R, 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.
- [11] Cheng J C P, Wang M Z. Automated detection of sewer pipe defects in closed-circuit television images using deep learning techniques[J]. Automation in Construction, 2018, 95: 155-171.
- [12] 刘鑫, 陈思溢, 陈小龙, 等. 基于深度学习的深层次多尺度特征融合目标检测算法[J]. 激光与光电子学进展, 2021, 58(12): 1210029. Liu X, Chen S Y, Chen X L, et al. Deep multi-scale feature fusion target detection algorithm based on deep learning[J]. Laser & Optoelectronics Progress, 2021, 58(12): 1210029.
- [13] Lu J S, Xiong C M, Parikh D, et al. Knowing when to look: adaptive attention via a visual sentinel for image captioning[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 3242-3250.
- [14] Meijer D, Scholten L, Clemens F, et al. A defect classification methodology for sewer image sets with convolutional neural networks[J]. Automation in Construction, 2019, 104: 281-298.
- [15] Haurum J B, Moeslund T B. A survey on image-based automation of CCTV and SSET sewer inspections[J]. Automation in Construction, 2020, 111: 103061.
- [16] 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.
- [17] Zuo X, Dai B, Shan Y W, et al. Classifying cracks at sub-class level in closed circuit television sewer inspection videos[J]. Automation in Construction, 2020, 118: 103289.
- [18] Tian Z, Shen C H, Chen H, et al. FCOS: fully convolutional one-stage object detection[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 9626-9635.
- [19] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [20] 郭列, 张团善, 孙威振, 等. 融合空间注意力机制的图像语义描述算法[J]. 激光与光电子学进展, 2021, 58(12): 1210030. Guo L, Zhang T S, Sun W Z, et al. Image semantic description algorithm with integrated spatial attention mechanism[J]. Laser & Optoelectronics Progress, 2021, 58(12): 1210030.