

DOI:10.19322/j.cnki.issn.1006-4710.2023.04.013

<https://xuebao.xaut.edu.cn>

引文格式:晏润冰,李明辉,李煜博,张梦圆,李金彪,张奇. 基于改进 TResNet 的下水道缺陷识别算法[J]. 西安理工大学学报, 2023, 39(4):567-575.

YAN Runbing, LI Minghui, LI Yubo, ZHANG Mengyuan, LI Jinbiao, ZHANG Qi. An improved TResNet algorithm for sewer defect identification [J]. Journal of Xi'an University of Technology, 2023, 39(4):567-575.

基于改进 TResNet 的下水道缺陷识别算法

晏润冰, 李明辉, 李煜博, 张梦圆, 李金彪, 张 奇

(陕西科技大学 机电工程学院, 陕西 西安 710021)

摘要: 同一张下水道图像可能同时反映多种类型的下水道缺陷, 所以应对下水道图像进行多标签缺陷预测。为了进一步解决城市下水道复杂场景中缺陷的多标签自动分类问题, 本文提出了一种两阶段的下水道缺陷类型识别方法。第一阶段, 使用二分类卷积神经网络(CNN)模型检测下水道图像是否存在缺陷; 第二阶段, 使用改进 TResNet 模型对下水道图像具体缺陷类型进行多标签分类, 该模型在 TResNet 基础上引入了捕获空间相关类别信息的特定类别残差注意力(CSRA)模块, 从而提升了网络的多标签识别能力。考虑到不同类型缺陷造成的经济影响程度不同, 使用加权缺陷类别经济影响指数的 F_{2-CIW} 指标以及标准 F_1 分数指标来评价模型的性能。实验结果表明, 在下水道数据集 Sewer-ML Dataset 上, 本文方法的 $F_{1-Normal}$ 和 F_{2-CIW} 指标取得了 90.41% 和 53.57% 的测试得分, 相比其它方法具有更好的缺陷识别性能。

关键词: 下水道缺陷识别; 多标签分类; TResNet; 特定类别的残差注意力模块

中图分类号: TP391.4

文献标志码: A

文章编号: 1006-4710(2023)04-0567-09

An improved TResNet algorithm for sewer defect identification

YAN Runbing, LI Minghui, LI Yubo, ZHANG Mengyuan, LI Jinbiao, ZHANG Qi

(College of Mechanical and Electrical Engineering, Shaanxi University of Science & Technology, Xi'an 710021, China)

Abstract: The same sewer image may reflect multiple types of sewer defects at the same time, so the sewer image should be of multi-label defect prediction. In order to further solve the problem of multi-label automatic classification of defects in the complex scene of urban sewers, a two-stage sewer defect type identification method is proposed in this paper. In the first stage, the binary classification convolutional neural network(CNN) model is used to detect whether there are defects in sewer images. In the second stage, the improved TResNet model is used to classify the specific defect types of sewer images with multiple labels. Based on TResNet, the model introduces the class-specific residual attention(CSRA) module to capture spatially-relevant category information, thus improving the multi-label recognition capability of the network. Considering the different degrees of economic impact caused by different defects, F_{2-CIW} index which weights economic impact index of defect class and standard F_1 score were used to evaluate the performance of the model. The experimental results show that our model obtained 90.41% and 53.57% scores on $F_{1-Normal}$ and F_{2-CIW} indexes of Sewer-ML Dataset, which has a better defect identification performance than other methods.

收稿日期: 2023-07-12; 网络首发日期: 2023-08-22

网络首发地址: <https://link.cnki.net/urlid/61.1294.N.20230821.1551.004>

基金项目: 陕西省科技厅重点研发计划资助项目(2023-YBGY-0697); 陕西省咸阳市重点研发计划资助项目(L2022ZDYF-GY-008)

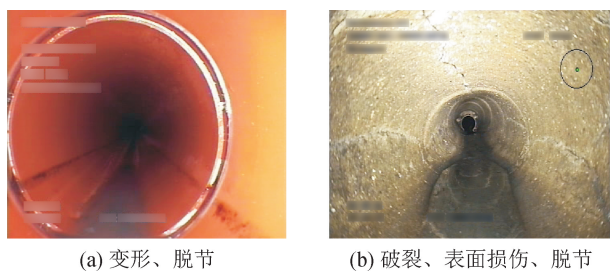
第一作者: 晏润冰, 男, 硕士生, 研究方向为机器视觉、数字图像处理、嵌入式系统与应用。E-mail: 504955844@qq.com

通信作者: 李明辉, 男, 博士, 教授, 研究方向为自动控制工程、嵌入式系统与应用。E-mail: liminghui@sust.edu.cn

Key words: sewer defect identification; multi-label classification; TResNet; CSRA

城市下水道被用来排放污水和雨水,是一座城市流动的“血管”。我国城市下水道众多,且铺设年限久远,因下水道老化造成的泄露、坍塌等事故,对环境、社会经济及人身安全造成了严重影响,故定期对城市下水道进行检修是非常必要的。目前,对城市下水道的检测主要采用 CCTV(closed circuit television)的方法,即使用专用的管道机器人在下水道内部采集图像信息,再将采集的图像交给专业人员,判别下水道存在的缺陷类型。人工判定大量的下水道图像是一个费时费力的事情,效率低下,且存在主观性等问题。近年来,深度学习方法在各类数字图像处理领域都得到了突破发展,这让城市下水道缺陷的自动化识别成为了可能,已经有部分学者开始将深度学习方法用于下水道缺陷识别领域的研究。

城市下水道内部场景极为复杂,不同类型的下水道之间结构又不尽相同(图 1)。相比过去采用手工提取特征的传统数字图像处理方法,自动学习提取特征的深度学习方法在图像识别领域取得了更好的效果,已有一些学者开始使用深度学习方法对下水道缺陷图像进行识别。Hassan 等^[1]使用经典的 AlexNet^[2]网络对下水道图像进行缺陷识别,同样地, Li 等^[3]使用改进的 ResNet-18^[4]网络对下水道图像做了端到端的缺陷识别;不同于 Hassan 和 Li 等端到端的检测方法,陈克凡^[5]将下水道图像缺陷识别分为两个阶段,在第一个阶段使用轻量化的 SqueezeNet^[6]网络来确定图像是否存在缺陷,第二个阶段则使用 Inception V3^[7]网络来识别具体的缺陷类型;Xie 等^[8]同样使用了两阶段的方法来识别下水道图像。



(a) 变形、脱节

(b) 破裂、表面损伤、脱节

图 1 下水道缺陷图片

Fig.1 Sewer defect image

在以往的下水道缺陷识别研究中,研究方向主要集中在识别出下水道的一种缺陷类别,而在实际场景中,一张下水道图像往往能同时反映出多个下水道缺陷类别,故下水道缺陷识别为多标签分类问题。Kumar 等^[9]利用二进制分类器的集合,使用以一对一方式训练的小型双层 CNN 模型对 4 种类型

的缺陷进行分类;曾保誌等^[10]提出了改进标签图的图卷积网络(GCN)模型,并对 7 类下水道缺陷进行了多标签分类;Haurum 等^[11]将多标签识别算法领域性能较好的网络模型对下水道缺陷图像的识别效果进行了对比,最后选取测试性能最好的 TResNet^[12]网络对下水道缺陷图像进行多标签分类识别。

在过去的下水道多标签缺陷识别研究中,存在缺陷类型识别不全面、识别准确率低等问题。为了进一步解决城市下水道复杂场景下的缺陷多标签分类问题,本文提出了一种两阶段识别方法:第一个阶段,设计小型 CNN 模型识别下水道图像是否存在缺陷;第二个阶段,在 TResNet 模型中引入特定类别的残差注意力模块(class-specific residual attention,CSRA)^[13]来识别下水道图像存在的缺陷类型。

1 下水道缺陷识别方法

在实际的下水道图像数据集中,由于正常的下水道数量远多于存在缺陷的下水道数量,故正常类别与各种缺陷类别的图像数量往往是极不平衡的。为了改善数据的均衡性,提升缺陷识别的性能,首先对下水道缺陷图像进行图像预处理,再将下水道缺陷识别分为正异常识别和缺陷类别识别两个阶段。其中,正异常识别阶段将所有缺陷的类别合并为异常类,与正常类形成二分类;缺陷类型识别阶段则对前一阶段的异常子类进行缺陷的具体分类。在缺陷识别的第一个阶段,使用一个二元 CNN 模型进行二分类;在第二个阶段,使用改进 TResNet 网络对缺陷图像进行缺陷类型多标签分类。下水道缺陷识别的整体方法如图 2 所示。

2 图像预处理

由于下水道原始图像的分辨率各不相同,因此在处理之前,首先将它们调整为 224×224 像素。为了避免图像数据的过拟合和增加数据集的大小,以 50% 的概率水平翻转,将亮度、对比度、饱和度和色调抖动原始值的 $\pm 10\%$,并使用均值和标准差对数据进行归一化处理。在图像推理过程中,图像也被调整为 224×224 像素并归一化。

3 正异常识别模型

近年来,一些大规模的训练数据集陆续出现,为了尽可能多地学习特征以达到更高的精度,学者们提出了更深更广的卷积神经网络架构。然而,直接

叠加更多的层并不能保证更好的分类结果^[4],相反,深层网络可能会导致退化问题,甚至出现比浅层网络更差的表现,因为深层网络很难训练,且深层网络对硬件有更高的要求。由于此阶段是面向所有使用的下水道数据集,考虑到工程的实际需求,在正异常识别阶段,应设计一个深度相对较浅的 CNN 网络来解决下水道图像二分类问题,从而筛选正常下水道图像。

二元分类网络的整体架构如图 3 所示,其网络的卷积层数及参数的选择都通过反复实验来确定。该网络主要包括三个卷积层和三个全连接层。每个卷积层后面都接上一个最大池化层,这有助于防止模型过拟合,而且还能减少需要学习的网络参数,降低计算成本。

由图 3 可知,第一次卷积操作采用 64 个 $3 \times 11 \times 11$ 的卷积核,步长为 1 像素。第二次和第三次卷积采用 3×3 的卷积核,步长调整为 2 像素。其中每次卷积操作都使用了零填充处理,以保持特征尺寸的完整性。这样做是为了防止下水道图像边缘信息的丢失,因为在下水道图像的边缘也可能出现缺陷,如图 4 所示。

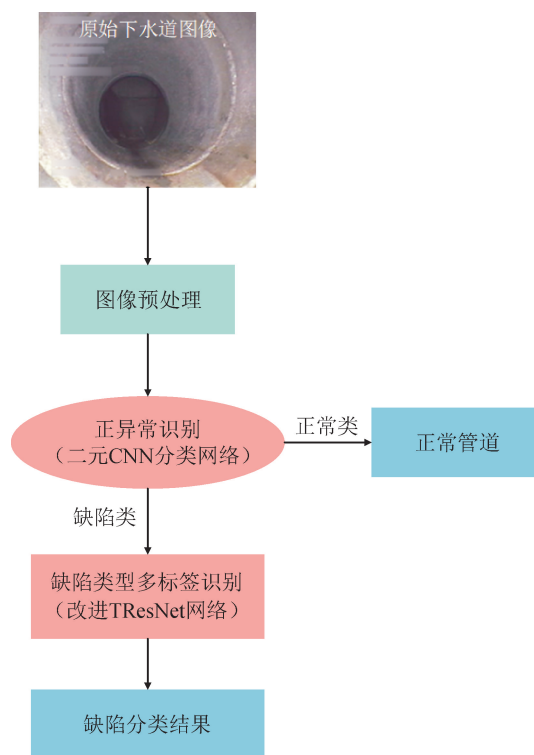


图 2 下水道缺陷识别方法
Fig. 2 Sewer defect identification method

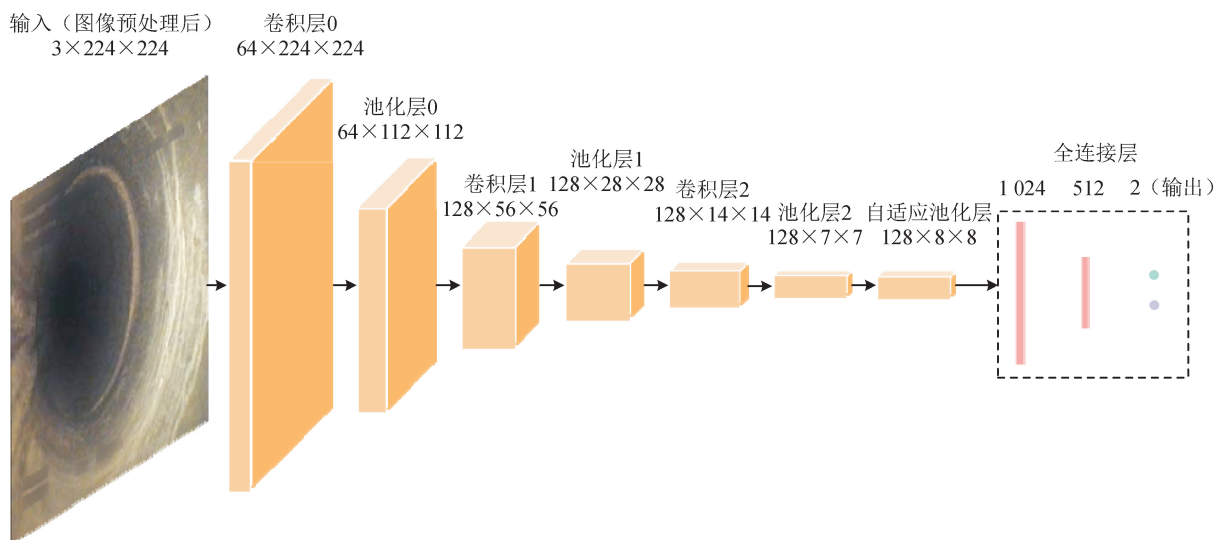


图 3 二元分类网络
Fig. 3 Binary classification network

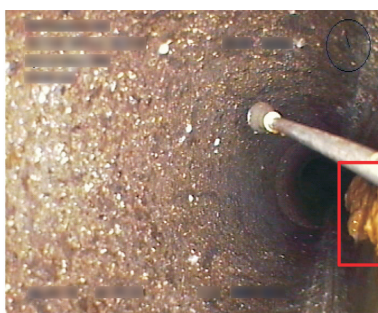


图 4 下水道图像边缘缺陷
Fig. 4 Sewer image edge defect

在整个网络中,最大池化层均采用 2×2 窗口,池化步长为 2 像素,每次池化操作前使用 ReLU 函数归一化。

网络在接入全连接层之前,将特征图自适应池化到 8×8 的大小。

网络末端的全连接层同样使用 ReLU 函数归一化神经元参数,使用 dropout 策略防止网络过拟合,最终输出两个神经元,分别表示“正常”类别和“缺陷”类别的概率。

4 缺陷类型识别模型

TResNet 网络具有较高的实际 GPU 训练和推理效率^[11],且在图像多标签分类任务领域性能表现良好。为了进一步提高城市下水道复杂场景下的缺陷识别准确性,在 TResNet 网络末端引入专为提升多标签分类性能的 CSRA 模块,对下水道缺陷图像进行多标签分类识别。

4.1 TResNet 网络结构

近年来开发的许多深度学习模型在 ImageNet 等数据集上达到了比 ResNet50^[4]更高的精度,但在测量实际 GPU 训练和推理吞吐量时,ResNet50 通常比最近的竞争对手^[14-16]快得多,提供了更好的吞吐量-精度权衡。TResNet 通过在 ResNet50 构架基础上进行专门的改进和优化,在与 ResNet50 相似的 GPU 吞吐量下达到了更好的性能。它包含三个变体,即 TResNet-M、TResNet-L 和 TResNet-XL,它们仅在深度和通道数量上有所不同,本文使用的模型为 TResNet-L。其中,主要的修改包括 SpaceToDepth 转换层的替换、Inplace-ABN 层的替换、SE 层的添加以及一系列代码和部署优化。

ResNet50 网络一开始采用大小为 7×7 、步长为 2 像素的卷积核及一个池化层来完成图像的特征提取以及 4 倍下采样,TResNet 网络使用 SpaceToDepth 转换层^[17]替代了这个模块来降低特征图尺寸。使用 SpaceToDepth 转换层的目的是降低由激进的降图像尺寸过程产生的“信息丢失”,以实现更高的精度。SpaceToDepth 转换层与 ResNet50 第一层的替换关系如图 5 所示。

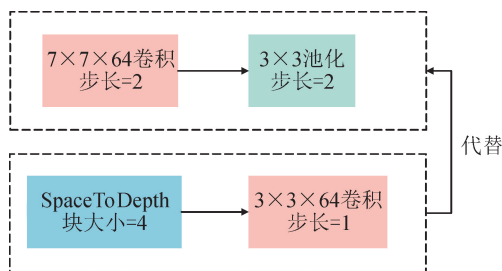


图 5 SpaceToDepth 转换层替换关系

Fig. 5 SpaceToDepth conversion layer replacement relation

TResNet 使用 Anti-Alias (AA) Downsampling^[18]取代降尺寸模块,具体关系如图 6 所示。

TResNet 采用 Inplace-ABN 层^[19]取代了所有的 BatchNorm+ReLU 层,大大减少了训练深层网络所需要的内存,从而可以显著增加训练时的批量大小。在激活函数上,使用 Leaky-ReLU 函数取代普通 ReLU 函数。

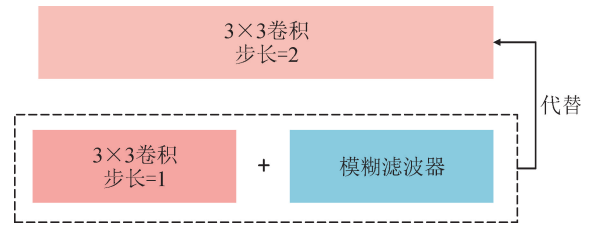


图 6 AA Downsampling 替换关系

Fig. 6 AA Downsampling replacement relation

ResNet 网络家族包含 BasicBlock 层和 Bottleneck 层两种基本的构建块。Bottleneck 层由两个 1×1 的卷积块和一个 3×3 的卷积块组成,相比 BasicBlock 层通常具有更好的精度,而 BasicBlock 层由两个 3×3 的卷积块组成,相比 Bottleneck 层具有更大的感受野,且具有较小的 GPU 使用率。不同于 ResNet50 网络在构建块上全部使用 Bottleneck 层,TResNet 网络在前两个阶段使用 BasicBlock 层来接受更大的感受野,在后两个阶段则继续使用 Bottleneck 层提取特征。使用这样的混合设计是为了让模型具有更好的速度-精度平衡。

TResNet 在两种基本块中添加了 Squeeze-and-Excitation (SE) 块^[20],用来提高模型精度。不同于标准的 SE 块设计,本文在 BasicBlock 层中将折减系数设置为 4,在 Bottleneck 层中将折减系数设置为 8。TResNet 网络的两种基本块设计如图 7 所示。

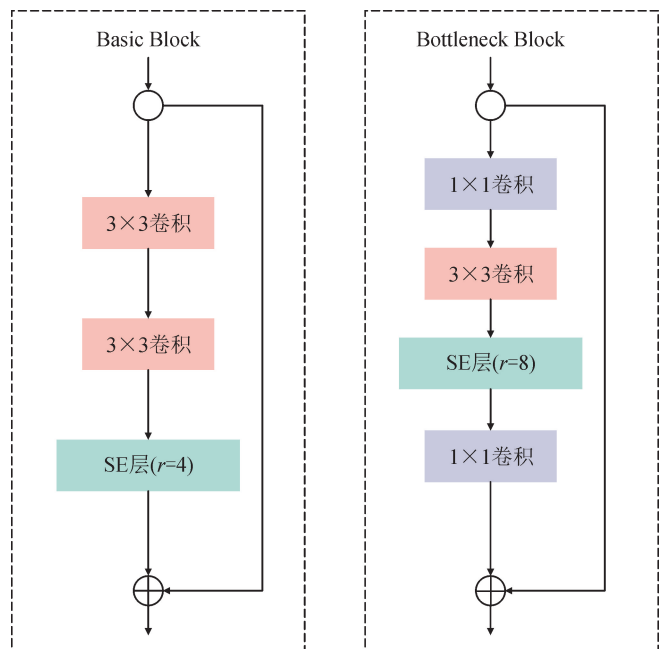


图 7 TResNet 的 BasicBlock 层和 Bottleneck 层设计

Fig. 7 BasicBlock layer and Bottleneck layer design of TResNet

对于预处理后的下水道缺陷图像 I ,将其发送给特征提取器 ϕ (TResNet-L),得到特征张量 $\mathbf{p} \in \mathbb{R}^{d \times h \times w}$,其中 d, h, w 分别为特征张量的维度、高度和宽度:

$$\mathbf{p} = \phi(I; \theta) \quad (1)$$

式中: θ 为 TResNet-L 网络主干参数。

输入下水道图像分辨率为 $3 \times 224 \times 224$,经过特征提取器 TResNet-L 后,输出特征张量的形状为 $2432 \times 7 \times 7$ 。

4.2 特定类别的残差注意力模块(CSRA)

CSRA 通过一个空间注意力得分机制来生成每个类别的特定类别特征,将其与类别无关的平均池化特征相结合,能有效捕获不同类别的缺陷所占据的不同空间区域,从而提升下水道多标签分类的性能。

将下水道图像输入特征提取器 TResNet-L 后,输出为 $2432 \times 7 \times 7$ 的特征张量 \mathbf{p} ,可将其分解为 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{49}$ ($\mathbf{x}_k \in \mathbb{R}^{2432}$)。则第 i 类 ($i \in \mathbb{R}^{17}$) 在特征张量 \mathbf{p} 第 j 个位置 ($j \in \mathbb{R}^{49}$) 的 class-specific 的注意力得分^[13]为:

$$s_j^i = \frac{\exp(\mathbf{T}\mathbf{x}_j^T m_i)}{\sum_{k=1}^{49} \exp(\mathbf{T}\mathbf{x}_k^T m_i)} \quad (2)$$

式中: T 为可手动调整的超参数; $m_i \in \mathbb{R}^{2432}$ 是第 i 类的分类器。将 s_j^i 看作第 i 类出现在位置 j 的概率。

得到第 i 类在第 j 个位置上出现的概率之后,把这个概率对原特征进行加权,并同维度全局求和,

得到 class-specific 的特征向量^[13] \mathbf{a}^i :

$$\mathbf{a}^i = \sum_{k=1}^{49} s_k^i \mathbf{x}_k \quad (3)$$

式中: s_k^i 为第 i 类的注意力得分; \mathbf{x}_k 为特征图在每个位置上分解的向量。

特征图的全局特征 \mathbf{g} 可通过对所有位置的特征求平均^[21] 获得:

$$\mathbf{g} = \frac{1}{49} \sum_{k=1}^{49} \mathbf{x}_k \quad (4)$$

由于特征向量 \mathbf{g} 得到了广泛应用,并取得了良好的效果,故将其作为主要的特征向量,将 \mathbf{a}^i 作为残差分量。给 \mathbf{a}^i 一个权重系统 λ (本文取 $\lambda = 0.4$),然后与 \mathbf{g} 相加,得到第 i 个类别的残差注意力特征^[13] \mathbf{f}^i :

$$\mathbf{f}^i = \mathbf{g} + \lambda \mathbf{a}^i \quad (5)$$

最后,将所有残差注意力特征 \mathbf{f}^i 与每个类别的分类器 m_i 相乘,得到最终的得分向量^[13]:

$$\hat{\mathbf{y}} \triangleq (y^1, y^2, \dots, y^C) = (m_1 \mathbf{f}^1, m_2 \mathbf{f}^2, \dots, m_C \mathbf{f}^C) \quad (6)$$

式中: C 表示类的数量。

以上为 CSRA 的单头注意力机制,CSRA 可进一步拓展至多头注意力机制,让多个不同的 CSRA 模块相加结合,从而产生更好的分类结果。其模块原理如图 8 所示。

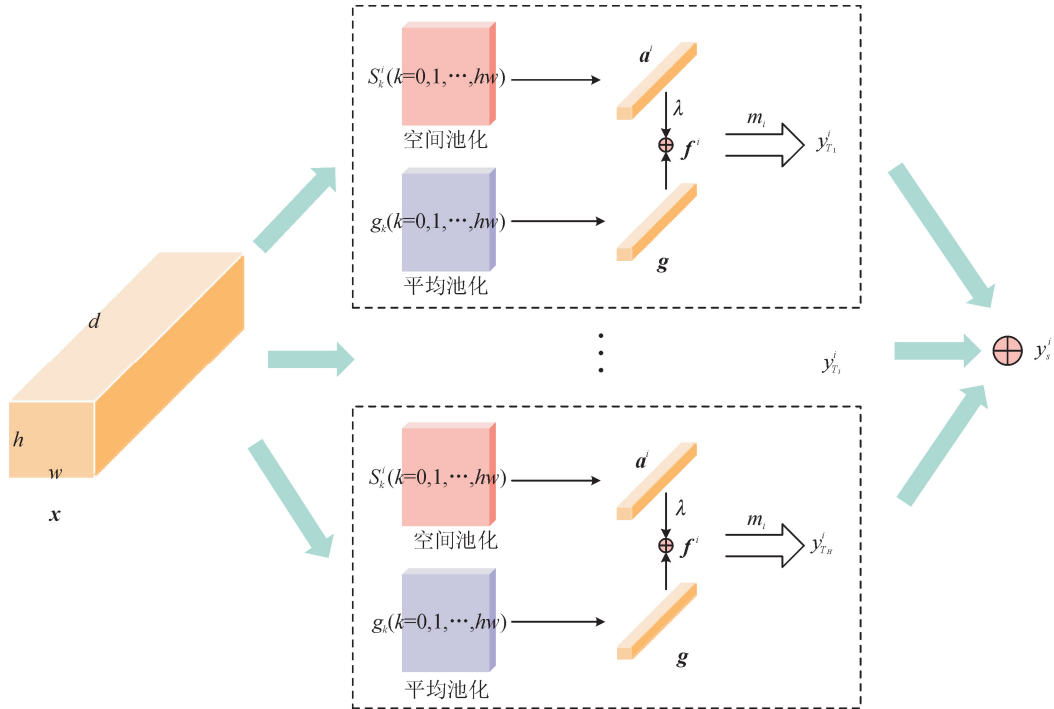


图 8 多头 CSRA 模块

Fig. 8 Multi-head CSRA module

多头注意力的最终得分向量^[13] \hat{y}_s 为:

$$\hat{y}_s = \sum_{h=1}^H \hat{y}_{T_i} \quad (7)$$

式中: T_i 为第 i 个注意力头的超参数。拓展为多头注意力后,可以根据头数 H 确定超参数 T_i ,为了实现更好的缺陷多标签分类效果,本文采用多头 $H=4$,根据文献[13],每个注意力头的超参数 T 选择 $T_{1:4} = 1, 2, 4, \infty$ 。其中 $T_4 = \infty$ 的效果相当于一个全局最大池化层。

缺陷类型识别阶段的对象为下水道缺陷图像。此阶段首先使用 TResNet-L 特征提取骨架来提取预处理后的图像特征, TResNet-L 特征提取骨架包含五个阶段:第一个阶段使用 SpaceToDepth 转换层来降低图像尺寸;第二个阶段和第三个阶段都使用 BasicBlock 层来提取特征,分别重复 4 次和 5 次;第四个阶段使用 Bottleneck 层,重复 18 次,进一步提取特征;第五个阶段使用原始 Bottleneck 层提取特征,重复 3 次。在提取缺陷图像特征后,使用 4 头 CSRA 注意力模块对特征图进行类相关空间位置信息捕捉,并输出类别预测(logits),如图 9 所示。

4.3 损失函数

本文采用二值交叉熵损失^[22] 作为损失函数,该损失函数常用于图像多标签分类领域,如式(8)所示:

$$L = \frac{1}{C} \sum_c - [w_c y_c \log(p_c) + (1 - y_c) \log(1 - p_c)] \quad (8)$$

式中: C 是缺陷类别的数量; y_c 表示当前图像是否有缺陷类 c ; p_c 表示模型预测类 c 为正的机率; w_c 表示类 c 在当前图像中的权重。

由于下水道图像数据集中各类缺陷的数量是不平衡的,故在损失函数里对不同类别进行加权,让出现次数较少的缺陷获得一个相对更高的权重,出现次数较多的缺陷则权重相对更低,如式(9)所示:

$$w_c = \frac{N - N_c}{N_c} \quad (9)$$

式中: N 是训练集的总图像数量; N_c 为训练集里包含缺陷类的图像数量。

5 实验结果与分析

为了评估改进后的模型在下水道缺陷多标签分类任务中的性能,利用下水道数据集 Sewer-ML Dataset^[11] 进行实验,并通过对比实验进行模型效果的评估。本节首先介绍数据集、实验环境、参数设置和评价指标,然后与目前最具代表性的下水道缺陷

分类算法进行对比。

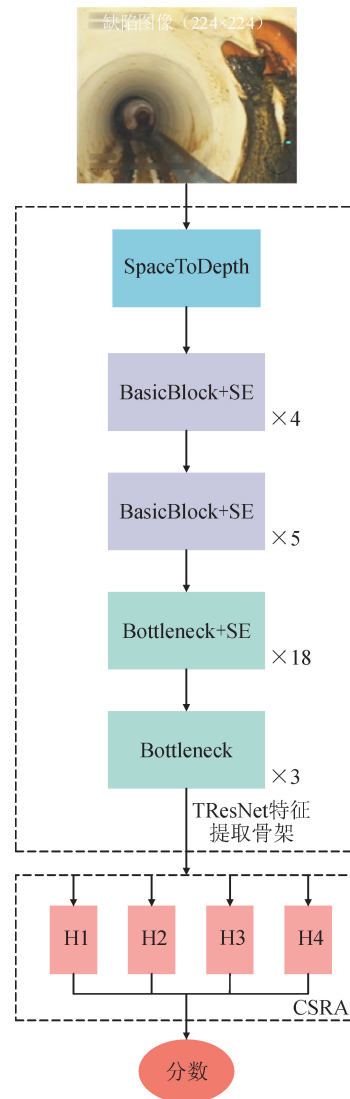


图9 缺陷类型多标签分类网络
Fig. 9 Multi-label classification network for defect type

5.1 数据集介绍

本文选用数据集中 131 631 张下水道图像对视觉算法进行了实验,其中无缺陷的正常图像 65 751 张,缺陷样本 65 880 张,数据集划分如表 1 所示。缺陷类型包括破裂、表面损伤、脱节、变形、树根等 17 类下水道缺陷^[23]。为了客观反映本文算法对实际下水道的缺陷多标签分类性能,本实验并未手动平衡各类缺陷的类别数量,而是按照下水道场景中缺陷的真实分布进行分析。在本文缺陷类别中,表面损伤、脱节和施工变更相关缺陷为出现次数最多的缺陷,分别占比数据集 24. 68%、41. 19% 和 18. 73%,侧向修复切口、过渡剖面相关缺陷为下水道最为罕见的缺陷,分别占比 0. 44% 和 0. 08%。考虑到不同缺陷会对经济造成不同的影响,故应给每个缺陷类别都赋予一个相应的经济影响得分,其分

数由专业人员判定^[24]。在模型推理时,令每个类别的得分除以经济影响最大类的得分,使其归一化到区间 $[0,1]$,并将归一化分数作为类别权重(class-importance weight, CIW)。缺陷类型及其对应的 CIW 权重如表 2 所示。

表 1 数据集分布
Tab.1 Data set distribution

类型	训练集	测试集	验证集	合计
正常图像	51 909	6 881	6 961	65 751
缺陷图像	55 463	5 325	5 092	65 880
合计	107 372	12 206	12 053	131 631

表 2 缺陷类型及其对应 CIW 权重
Tab.2 Defect type and its corresponding CIW weight

代表符号	缺陷类型	CIW
RB	破裂	1.000 0
OB	表面损伤	0.551 8
PF	生产错误	0.289 6
DE	变形	0.162 2
FS	脱节	0.641 9
IS	接口材料脱落	0.184 7
RO	树根	0.355 9
IN	渗透	0.313 1
AF	沉积	0.081 1
BE	结垢	0.227 5
FO	障碍物	0.247 7
GR	支管暗接	0.090 1
PH	凿孔连接	0.416 7
PB	钻孔连接	0.416 7
OS	侧向修复切口	0.900 9
OP	过渡剖面相关缺陷	0.382 9
OK	施工变更相关缺陷	0.382 9

5.2 实验设置

本实验的主要硬件环境:CPU 为 Intel(R) Core (TM) i5-12400F,内存 16 GB;GPU 为 NVIDIA GeForce RTX 3060,显存 12 GB。本实验在 Windows 10 操作系统上进行,采用的深度学习框架为 PyTorch。使用带动量的随机梯度下降(SGD)训练,最大迭代次数设置为 100 次,批量大小设为 64。采用的学习

率为 0.1,动量为 0.9,权重衰减为 0.000 1,并在 30 次、60 次和 80 次时给学习率乘以 0.1。

5.3 评价指标

F_β 分数^[25]综合考虑了模型的精度和召回率,常被用于评价模型的性能,其计算方法为:

$$F_\beta = (1 + \beta^2) \frac{Prc \cdot Rcll}{\beta^2 Prc + Rcll} \quad (10)$$

式中: Prc 为模型的精度; $Rcll$ 为模型的召回率; β 为召回率的权重。

在实际的下水道图像缺陷检查中,找出含有缺陷的图像往往比找出图像内的缺陷类型更重要,即模型的召回率比精度更重要。如果两者不能同时满足,那么应赋给召回率更高的权重,故本文设置 $\beta=2$ 。考虑到每个缺陷类型的经济影响程度,将表 2 中的 CIW 权重整合到 F_2 评价指标中,得到 F_{2-CIW} 指标^[11]:

$$F_{2-CIW} = \frac{\sum_{c=1}^c F_{2-c} \cdot CIW_c}{\sum_{c=1}^c CIW_c} \quad (11)$$

式中: CIW_c 和 F_{2-c} 分别为 c 类的 CIW 和 F_2 分数; c 为缺陷类别序数。

由于正常类别没有 CIW 权重,故同时使用 F_1 -Score 来评价正常下水道,以下称为 $F_{1-Normal}$ 。

5.4 对比实验

ResNet-101 解决了深度学习容易出现的梯度消失和模型退化问题,在图像分类任务研究中取得了非常好的效果,是图像分类通用领域的基准算法。为了更好地体现算法的有效性,在相同的实验条件下,将本文算法与具有代表性的原 TResNet-L 算法及 ResNet-101 算法进行对比,使用 $F_{1-Normal}$ 及 F_{2-CIW} 两个指标对 3 种算法进行客观评价,如表 3 所示。

表 3 对比实验
Tab.3 Contrast experiment

算法	验证/%		测试/%	
	$F_{1-Normal}$	F_{2-CIW}	$F_{1-Normal}$	F_{2-CIW}
ResNet-101	90.78	50.46	90.31	50.03
TResNet-L	90.81	51.38	90.35	51.06
本文算法	90.88	53.93	90.41	53.57

三种方法都使用 0.5 的阈值对各个缺陷类别的预测分数进行二值化,且三种方法在第一阶段都使用前文所述的二元分类网络来识别是否存在缺陷。若二元分类网络识别到正常图像,则使用二元分类网络的结果;否则以第二阶段的预测分数作为每个

缺陷类别的分数。在训练时,三种算法经历 80~90 次 epoch 后会得到最低的验证损失,此后网络开始过拟合。在预测时,使用训练轮次里验证损失最小的模型权重来获得结果。

综上所述,本文所提方法在 TResNet 模型基础上引入了多头 CSRA 注意力模块来关注不同缺陷类别在不同位置上的信息,相比 TResNet-L、ResNet-101,具有更好的缺陷多标签识别能力。

5.5 每类缺陷的性能

为了直观地理解模型在不同缺陷类型上的性能表现及其经济影响程度,本文将对每类缺陷的 F_{2-CIW} 分数及其 CIW 权重进行比较,如图 10 所示。

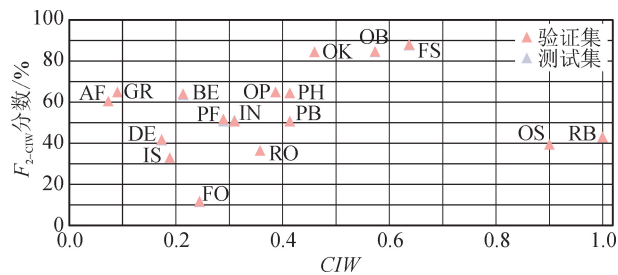


图 10 各类缺陷性能

Fig. 10 All kinds of defect type performance

本文模型在经济影响程度较高的 FS(脱节)、OB(表面损伤)、OK(施工变更相关缺陷)缺陷类上表现出较高的性能,这三类缺陷是下水道最常出现的缺陷。对于功能型缺陷 FO(障碍物)、IS(接口材料脱落)、RO(树根),模型表现出相对较低的识别性能,可能是由于这几类缺陷图像特征相差较大,比如障碍物可能是断裂的墙体、易拉罐堆或穿过主下水道的另一根下水管道等,这几类缺陷的经济影响程度相对较低。RB(破裂)、OS(侧向修复切口)类缺陷具有最高的经济影响程度,但是模型在这两类缺陷上表现出的性能却并不高,如果想要进一步提高 F_{2-CIW} 分数,提升这两类缺陷的分类性能非常重要。

5.6 消融实验

本实验采用两阶段的缺陷识别方法。在第一阶段,把所有下水道缺陷合成一类,使用二元分类器进行有无缺陷识别;在第二阶段,对缺陷类型进行具体分类。为了验证两阶段方法的有效性,将本文的两阶段分类方法与仅使用改进 TResNet 模型的单阶段方法在测试集上进行比较,使用 F_{2-CIW} 和 $F_{1-Normal}$ 指标来评价性能,消融实验结果如表 4 所示。

由表 4 可知,使用两阶段方法第一阶段的二元分类器过滤正常下水道,相比仅使用多标签分类器的单阶段方法,在 F_{2-CIW} 和 $F_{1-Normal}$ 指标上都有了提高,特别是在考虑正常下水道分类的 $F_{1-Normal}$ 指标上

提升了 10.79%,说明本文采用的两阶段方法对识别下水道图像类别是非常有效的。

表 4 消融实验对比

Tab. 4 Comparison of ablation experiments

方法	$F_{1-Normal} / \%$	$F_{2-CIW} / \%$
单阶段方法	79.62	52.87
两阶段方法	90.41	53.57

6 结 语

城市下水道缺陷对社会安全、稳定以及经济发展有着非常重要的影响。本文提出了一种两阶段的水下道多标签分类算法,在第一阶段,设计二元分类网络识别图像有无缺陷;在第二阶段,改进 TResNet 算法,对下水道缺陷类型进行多标签识别,在 TResNet 网络基础上引入特定类别的残差注意力模块 CSRA,通过一个与空间相关的注意力机制提升每类缺陷的分类性能。与原 TResNet-L、ResNet-101相比,本文算法在下水道数据集 Sewer-ML Dataset 上取得了更好的多标签识别效果。

参考文献:

- [1] 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.
- [2] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. Advances in Neural Information Processing Systems, 2012, 25: 1097-1105.
- [3] 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.
- [4] 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 Computer Society, 2016: 770-778.
- [5] 陈克凡. 基于视觉的地下管道缺陷检测方法研究[D]. 哈尔滨:哈尔滨工业大学,2019.
CHEN Kefan. Research on detecting method of sewer defects based on vision [D]. Harbin: Harbin Institute of Technology, 2019.
- [6] LANDOLA F N, HAN S, MOSKEWICZ M W, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size[J]. arXiv preprint arXiv, 2016, 1602.07360.

- [7] 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 Computer Society, 2016: 2818-2826.
- [8] XIE Q, LI D, XU J, et al. Automatic detection and classification of sewer defects via hierarchical deep learning[J]. IEEE Transactions on Automation Science and Engineering, 2019, 16(4): 1836-1847.
- [9] 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.
- [10] 曾保誌, 罗建桥, 熊鹰, 等. 基于改进标签图的图卷积网络管道缺陷检测模型[J]. 激光与光电子学进展, 2022, 59(18): 54-62.
- ZENG Baozhi, LUO Jianqiao, XIONG Ying, et al. Graph convolutional network detection model for pipeline defects based on improved label graph[J]. Laser & Optoelectronics Progress, 2022, 59(18): 54-62.
- [11] HAURUM J B, MOESLUND T B. Sewer-ML: a multi-label sewer defect classification dataset and benchmark[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE Computer Society, 2021: 13456-13467.
- [12] RIDNIK T, LAWEN H, NOY A, et al. TResNet: high performance GPU-dedicated architecture [C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa: IEEE Computer Society, 2021: 1400-1409.
- [13] ZHU K, WU J. Residual attention: a simple but effective method for multi-label recognition[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal: IEEE Computer Society, 2021: 184-193.
- [14] TAN M, LE Q. EfficientNet: rethinking model scaling for convolutional neural networks [C]//International Conference on Machine Learning. Long Beach: ICML, 2019: 6105-6114.
- [15] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: transformers for image recognition at scale[J]. arXiv preprint arXiv: 2021, 2010.11929.
- [16] LIU Z, LIN Y, CAO Y, et al. Swin transformer: hierarchical vision transformer using shifted windows [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal: IEEE Computer Society, 2021: 10012-10022.
- [17] SANDLER M, BACCASH J, ZHMOGINOV A, et al. Non-discriminative data or weak model? On the relative importance of data and model resolution [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. Seoul: IEEE Computer Society, 2019: 1036-1044.
- [18] ZHANG R. Making convolutional networks shift-invariant again [J]. arXiv preprint arXiv, 2019, 1904.11486.
- [19] BULO S R, PORZI L, KONTSCHIEDER P. In-place activated batchnorm for memory-optimized training of DNNs [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE Computer Society, 2018: 5639-5647.
- [20] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE Computer Society, 2018: 7132-7141.
- [21] LIN M, CHEN Q, YAN S. Network in network [J]. arXiv preprint arXiv, 2013, 1312.4400.
- [22] BROWNLEE J. Probability for machine learning: discover how to harness uncertainty with Python [M]. Austin: Machine Learning Mastery, 2019.
- [23] A/S B, BRUNHØJVEJ. Fotomanualen: TV-inspektion af afløbsledninger [J]. Dansk Vand og Spildevandsforening (DANVA), 2010, 6: 8-56.
- [24] PEDERSEN M, LADEN B. Fotomanualen: beregning af fysisk indeks ved TV-inspektion [J]. Dansk Vand og Spildevandsforening (DANVA), 2005, 3: 9-16.
- [25] VAN R C J. Information retrieval [M]. London: Butterworth-Heinemann, 1979.

(责任编辑 周 蓓)