

# 基于XS-BiSeNetV2的城市地下管道缺陷语义分割模型

田淙文, 蓝雯飞\*, 李波, 潘禹欣, 姚为

(中南民族大学 计算机科学学院, 武汉 430074)

**摘要** 针对城市地下管道缺陷分割任务对实时分割的需求, 基于BiSeNetV2模型提出了一种改进的分割模型XS-BiSeNetV2. 针对城市地下管道缺陷形态多样、空间特征复杂的问题, 设计了互相关联的双分支交叉特征提取策略; 针对传统跨步卷积存在的空间信息损耗, 使用Haar小波变换下采样模块替换细节分支的跨步卷积, 保留更多空间信息; 针对轻量级模型存在的上下文特征不匹配和空间特征偏移的问题, 使用上下文和空间特征校准模块提高模型的分割精度. 通过实验验证了提出的缺陷分割模型的实时性和有效性, 对比原始BiSeNetV2模型, mIoU提升了2.65%, mPA也提升了1.47%, 且拥有每秒49帧的处理速度, 具有良好的实时性. 相比其他基于深度学习的实时语义分割模型, 也具有一定优势.

**关键词** XS-BiSeNetV2模型; 缺陷分割; 城市地下管道; Haar小波下采样

**中图分类号** TP391 **文献标志码** A **文章编号** 1672-4321(2025)04-0536-10

**doi**: 10.20056/j.cnki.ZNMDZK.20250413

## Sematic segmentation method for urban underground pipeline defects based on XS-BiSeNetV2

TIAN Congwen, LAN Wenfei\*, LI Bo, PAN Yuxin, YAO Wei

(College of Computer Science, South-Central Minzu University, Wuhan 430074, China)

**Abstract** In order to meet the demand of real-time segmentation for urban underground pipeline defects segmentation task, this paper proposes an improved segmentation model XS-BiSeNetV2 based on the BiSeNetV2 model. For the problems of urban underground pipeline defects with diverse morphology and complex spatial features, an interrelated two-branch crossover feature extraction strategy is designed. For the loss of spatial information in traditional step-by-step convolution, the Haar wavelet transform downsampling module is used to replace the step-by-step convolution of detail branches to retain more spatial information. For the problems of mismatch of context features and offset of spatial features in lightweight models, the context and spatial feature calibration module is used to improve the segmentation accuracy of the model. The real-time and effectiveness of the proposed defect segmentation model is verified through experiments. Compared with the original BiSeNetV2 model, the mIoU is improved by 2.65%, and the mPA is also improved by 1.47%, and it possesses a processing speed of 49 frames per second, which provides good real-time performance. It also has some advantages compared to other real-time semantic segmentation models based on deep learning.

**Keywords** XS-BiSeNetV2 model; defect segmentation; urban underground pipeline; Haar wavelet downsampling

城市地下管道是关系到居民的正常生活、城市的公共卫生以及城市的发展建设的重要基础设施之一, 随着使用年限的增长, 会出现各类缺陷<sup>[1]</sup>. 其

内部情况评估是通过采用相关的设备和技术检查管道内部来实现, 目前应用最广泛的技术是管道闭路电视 (Closed-Circuit Television, CCTV) 检测技

收稿日期 2024-07-07

\* 通信作者 蓝雯飞(1966-), 女, 教授, 数据库技术和深度学习, E-mail: 446487309@qq.com

基金项目 国家自然科学基金资助项目(61976226)

术<sup>[2]</sup>.由于CCTV检测具有成本低、部署灵活等优势,基于CCTV的图像自动检测技术成为城市地下管道缺陷检测任务中亟待重点发展的技术<sup>[3]</sup>.

基于图像的地下管道缺陷分割不仅可以对多种缺陷进行检测和定位,还可以获取缺陷的形态边缘信息,便于缺陷评级工作.传统的分割方法主要基于数学形态学,SU等提出了一种基于边缘检测的形态学分割方法,并验证该方法比数学形态学方法中的开顶帽方法有更好的分割性能<sup>[4]</sup>.随着深度学习技术的发展,WANG等<sup>[5]</sup>首次将深度学习模型应用于城市地下管道缺陷分割任务中.HAURUM等<sup>[6]</sup>为城市地下管道缺陷分类工作提供了一个名为Sewer-ML的大型公开数据集.ZHOU等<sup>[7]</sup>将DeepLabv3+模型应用于城市排水管道缺陷语义分割,对比了不同骨干网络下模型的表现.PAN等<sup>[8]</sup>基于UNet基础模型提出了PipeUNet模型.这些学者的研究为地下管道缺陷语义分割领域做出了大量贡献.

本文基于BiSeNetV2<sup>[9]</sup>模型,提出了XS-BiSeNetV2分割模型(XS中的X表示改进的双分支交叉的结构,S为Sewer污水管道的简写),并首次应用到城市

地下管道缺陷分割,主要贡献包括:(1)设计了双分支交互指导的特征提取策略;(2)使用Haar小波变换下采样<sup>[10]</sup>替换了传统跨步卷积下采样,减少空间信息损耗;(3)在模型预测前加入上下文和空间特征校准模块<sup>[11]</sup>,减少模型的上下文信息失配和边缘预测偏移.此外,针对目前城市地下管道缺陷语义分割领域尚未有公开数据集的情况,本文在基于Sewer-ML<sup>[6]</sup>城市地下管道缺陷图像数据集构建的缺陷分割任务数据集上进行实验.

## 1 BiSeNetV2 模型原理

BiSeNetV2模型是2021年由YU等<sup>[9]</sup>提出的轻量级语义分割模型,他们认为对于语义分割任务来说,低级的空间细节信息和高级的语义信息都至关重要,为了提高模型的速度和精确度而不牺牲这些信息,他们采用了分别处理低级空间信息和高级语义信息的双分支的架构设计了BiSeNetV2模型.如图1所示.

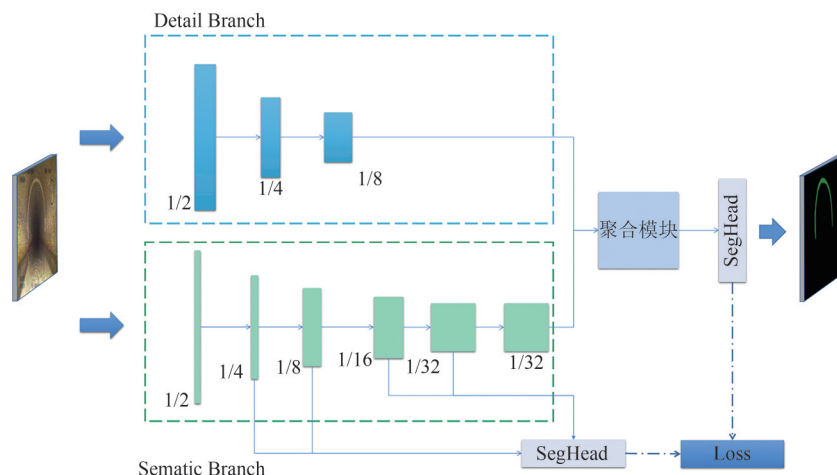


图1 BiSeNetV2结构图

Fig. 1 Structure of BiSeNetV2

模型包含两个分支,一个是具有较多通道数量、但整体层次更浅的细节分支,用于提取空间细节特征如边缘信息、形态信息等;一个是具有较少通道数量、但整体层次更深的语义分支,用于提取任务相关的高级语义特征.细节分支和语义分支并行,分别提取低级空间信息和高级语义信息,由于这两个分支是独立的,因此需要一个聚合模块来将这两个分支提取的信息融合,对于融合的结果会通过分割头得到最终的预测输出.此外他们还引入了额外的辅助深度监督模块,对语义分支的第二到第

四阶段的输出,会分别通过一个分割头得到预测输出,将这些预测输出都与实际标签计算损失,并将损失汇总到优化模型.

## 2 本文提出的XS-BiSeNetV2模型

本文基于BiSeNetV2模型,提出了一种用于城市地下管道缺陷分割的改进模型XS-BiSeNetV2模型,其结构如图2所示,红色框内为改进模块.对比BiSeNetV2模型,本文改进了模型独立的双分支结

构,对原始模型细节分支与语义分支的对应尺度特征图进行X型交叉融合,保证模型的细节分支能得到一定的语义分支信息,语义分支也能得到一定的细节分支信息,使细节分支和语义分支相辅相成地学习信息.考虑到传统的跨步卷积可能带来的空间细节信息丢失,模型采用HWD模块替换细节分支

中用于下采样的卷积模块,保证细节分支能够保留更多有用的空间细节信息.为了减少像素上下文不匹配和空间错位的问题,模型引入了上下文特征校准模块和空间特征校准模块,在模型预测之前分别对特征图中的上下文信息和空间细节信息进行校准,提高模型的分割能力.

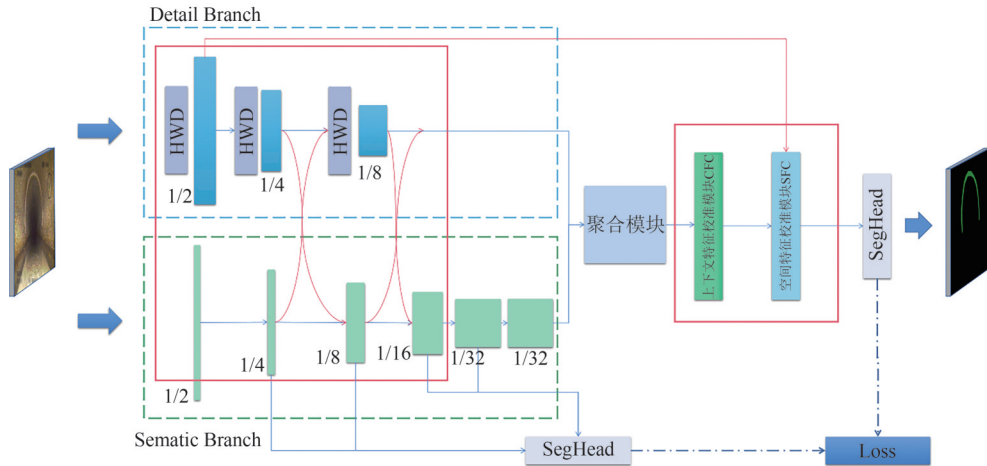


图2 XS-BiSeNetV2结构图  
Fig. 2 Structure of XS-BiSeNetV2

2.1 交叉特征提取模块

BiSeNetV2模型使用了双分支的架构,空间细节分支和语义分支相互独立,在语义分割任务中,空间细节信息和语义信息被认为在模型的预测中拥有同等重要的意义<sup>[12]</sup>.由于城市地下管道缺陷各类别边缘形态特征复杂,类别语义差异较大.在复杂的任务上为提高速度使用语义和空间相互独立的分支可能会使两种信息的提取都存在偏差,导致精度的下降,考虑到两种信息的同等重要性,本文放弃了原始模型的双分支独立信息提取,采取了双分支交叉特征提取,使细节分支影响语义分支特征提取的同时,保证语义分支指导细节分支进行特征提取,互相影响以提高两种信息的融合,在不耗费太多计算量的情况下,加强模型对两种信息的获取,提升模型的分割能力.

交叉特征提取模块结构如图3所示,对模型细节分支和语义分支第二、第三阶段的输出进行交叉融合.由于这两个阶段的特征图通道数不同,使用 $1 \times 1$ 卷积调整细节分支特征图的通道数再通过BN层和ReLU激活函数,采用相加的操作与语义分支的特征图融合,之后将特征图输入到语义分支的第三阶段.对语义分支的特征图同样使用 $1 \times 1$ 卷积调整通道数,然后通过BN层和ReLU激活函数,并与此时第二阶段的细节分支的特征图相加融合,输入

到细节分支的第三阶段.同样,在两个分支的第三阶段使用相同的融合方式,在这种相互影响下使空间细节特征与语义特征进行融合,提高模型的分割能力.

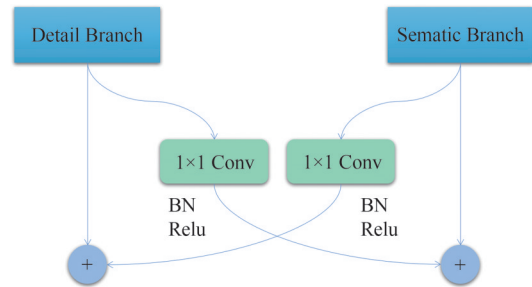


图3 交叉特征提取模块结构  
Fig. 3 Structure of cross feature extraction module

2.2 Haar小波变换下采样模块

深度学习任务中常用的下采样方法包括平均池化、最大池化和跨步卷积操作<sup>[13]</sup>.上述3种传统的下采样方法往往会导致语义分割模型在一定程度上丢失重要的空间信息,不利于语义分割模型的预测.针对这个问题XU等提出了Haar小波变换下采样(HWD, Haar Wavelet Downsampling)的方法.Haar小波变换主要被应用于图像压缩、去噪等方面,优势是能够简单快速无损地处理图像<sup>[14]</sup>.传统的池化和跨步卷积进行下采样时都不可避免地导致空间信息的损失,而HWD模块能显著减少特征图在下

采样过程中的空间信息损耗,提高模型的分割性能.城市地下管道缺陷类别形态各异,边缘特征复杂,更多空间信息有益于缺陷分割精度的提升,本文将原始模型中细节分支的跨步卷积下采样方式改为 Haar 小波变换下采样,HWD 模块结构如图 4 所示,主要包括两个模块:无损特征编码模块和特征表示学习模块.无损特征编码模块负责变换特征并降低空间分辨率,利用 Haar 小波变换可以将一张大小为  $W \times H$  灰度图转换为 4 张大小为  $\frac{W}{2} \times \frac{H}{2}$  的灰度图,为一张原图的近似图 A 和三张原图各方向上的细节信息,包括原图的水平分量  $H$ 、原图的垂直分量  $V$ 、原图的对角线分量  $D$ ,可以被认为是将图像的部分信息从空间整理到了通道上.特征表示学习模块用于从上一个模块的特征图中学习信息和过滤冗余信息调整通道数,该模块主要由一个  $1 \times 1$  的卷积层、一个 BN 层和一个激活函数组成.

2.3 上下文和空间特征校准模块

轻量级的实时语义分割模型常常会以降低网络深度、裁剪通道数或设计高效的解码器方式来提高模型的推理速度<sup>[15]</sup>,而这些操作往往意味着舍弃一定量的特征信息,导致模型精度的下降,模型的预测输出发生类别不准确或预测目标区域错位.金字塔池化模块<sup>[16]</sup>、空洞空间金字塔池化<sup>[17]</sup>以及密集连接的空洞空间金字塔池化模块<sup>[18]</sup>等方法融合了多尺度信息,并聚合上下文信息.但这些方法对所有的像素都引入了相似的上下文信息聚合操作,没有

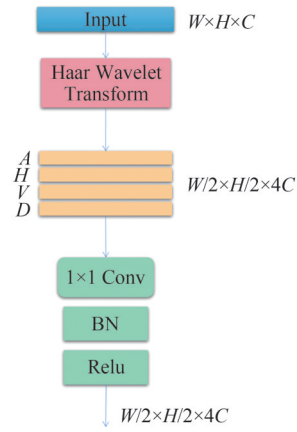


图 4 HWD 模块结构图

Fig. 4 Structure of Haar wavelet downsampling module

考虑到不同像素对不同上下文信息的差异.LI 等<sup>[11]</sup>提出了上下文和空间校准的方法来应对上述问题.

考虑到将实时语义分割模型应用于城市地下管道缺陷分割任务所带来的精度不高、类别错判、边缘预测错位的问题,本文在 BiSeNetV2 的模型尾部引入了上下文特征校准模块和空间特征重校准模块,输入的特征图会依次通过这两个模块,且会将模型细节分支的第一阶段的特征图输出引入空间特征校准模块补充模型浅层提取的空间信息.上下文特征校准模块利用级联金字塔池化模块提取上下文信息,然后根据相似度对每一个像素选取合适的语义上下文信息,空间特征重校准模块按通道维度将特征图分割,采用特征重采样的方式重构特征,弥补下采样造成的空间细节损失.上下文特征校准模块如图 5 所示.

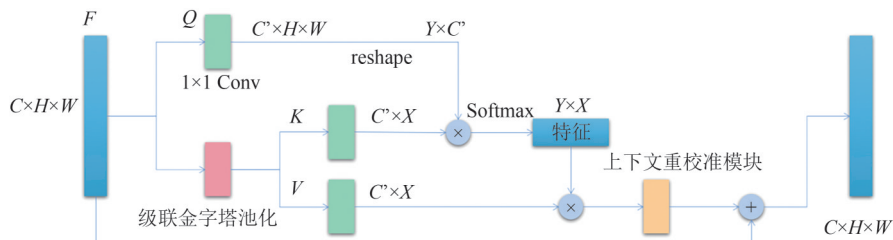


图 5 上下文特征校准模块

Fig. 5 Context feature calibration module

上下文特征重校准模块会选取语义上更接近的区域来聚合上下文,对输入特征图  $F \in \mathbb{R}^{C \times H \times W}$ ,通过如图 6 所示的级联金字塔池化模块,使用不同的自适应的平均池化调整到宽高为  $1 \times 1, 2 \times 2, 3 \times 3, 6 \times 6$ ,然后分别对其展平并拼接,生成多尺度上下文  $M \in \mathbb{R}^{C \times X}$ .

将特征图  $F$  使用  $1 \times 1$  的卷积进行特征降维生成特征图  $Q \in \mathbb{R}^{C' \times H \times W}$ ,并将  $Q$  形状调整为  $\mathbb{R}^{Y \times C'}$ ,此时  $C'$  远小于  $C$ ,且  $Y = H \times W$ .

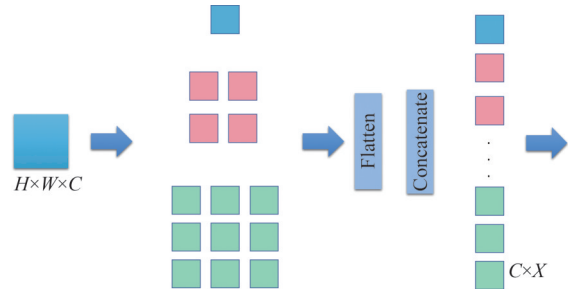


图 6 级联金字塔池化模块

Fig. 6 Cascade pyramid pooling module

对多尺度上下文  $M$  分别使用两个  $1 \times 1$  的卷积生成两种上下文表示  $K \in \mathbb{R}^{C \times X}$  和  $V \in \mathbb{R}^{C \times X}$ . 对特征图  $Q$  和上下文表示  $K$  进行矩阵乘法运算并计算 softmax 生成像素和上下文的仿射  $A \in \mathbb{R}^{Y \times X}$ , 可以表示为:

$$A_{i,j} = \frac{\exp(Q_i \cdot K_j)}{\sum_{j=1}^n \exp(Q_i \cdot K_j)}, \quad (1)$$

其中  $A_{i,j}$  表示特征图  $Q$  第  $i$  个像素到上下文  $K$  的第  $j$  个元素的仿射. 然后对  $A$  和  $V$  计算矩阵乘法得到语义校准的上下文  $S \in \mathbb{R}^{C \times X}$  并将其形状调整为  $\mathbb{R}^{C \times H \times W}$ , 最后将  $S$  通过上下文重校准模块并与原始特征图  $F$  进行相加, 得到上下文重校准后的特征图输出. 上下文重校准模块如图 7 所示.

空间特征重校准模块用于减少在模型向前推理的过程中多次下采样造成的空间细节损失, 会使

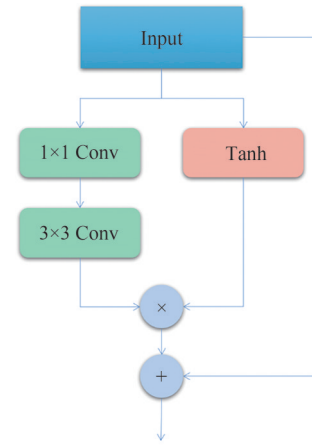


图 7 上下文重校准模块

Fig. 7 Context re-calibration module

用到当前的低分辨率特征图  $L \in \mathbb{R}^{C_i \times H_i \times W_i}$  和一个高分辨率的浅层特征图  $H \in \mathbb{R}^{C_h \times H_h \times W_h}$ , 对这两个特征图进行如图 8 所示的操作.

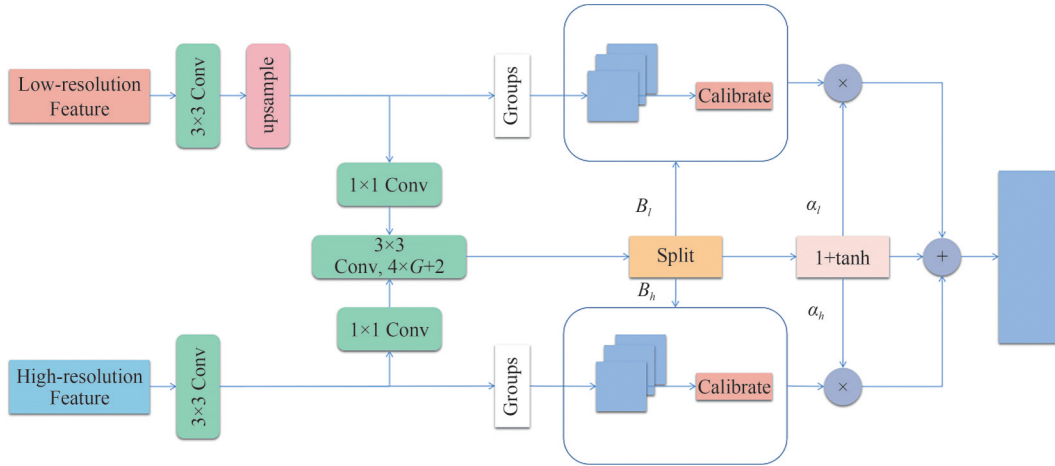


图 8 空间特征重校准模块

Fig. 8 Spatial feature re-calibration module

首先同时使用  $3 \times 3$  的卷积层调整通道数, 且对低分辨率特征图进行上采样, 调整到和高分辨率特征图相同的尺寸, 此时两张特征图的大小均调整为  $\mathbb{R}^{C \times H \times W}$ , 然后对两张特征图拼接并使用卷积块进行卷积操作, 得到特征图  $B \in \mathbb{R}^{(4 \times G + 2) \times H \times W}$ , 并拆分成两张对齐偏移量特征图  $B_l \in \mathbb{R}^{(2 \times G) \times H \times W}$ 、 $B_h \in \mathbb{R}^{(2 \times G) \times H \times W}$  和两张门限特征图  $\alpha_l \in \mathbb{R}^{1 \times H \times W}$ 、 $\alpha_h \in \mathbb{R}^{1 \times H \times W}$ , 之后分别将低分辨率特征图  $L$  和高分辨率特征图  $H$  分成  $G$  组, 对每组特征图分别使用上述偏移量进行横向和纵向上的校准得到校准后的特征图  $L'$  和  $H'$ . 校准操作以对一张单通道的图像为例, 如公式 (2) 所示:

$$O_{h,w} = \sum_{h'} \sum_{w'} F_{h',w'} \cdot \max(0, 1 - |h + \Delta_{h,w}^1 - h'|) \cdot \max(0, 1 - |w + \Delta_{h,w}^2 - w'|), \quad (2)$$

其中  $F$  代表校准前的原始特征图,  $\Delta \in \mathbb{R}^{2 \times H \times W}$  表示偏移量特征图, 表示在点  $(h + \Delta_{h,w}^1, w + \Delta_{h,w}^2)$  采样得到输出  $O_{h,w}$ . 由于低分辨率特征图包含丰富的语义信息, 高分辨率特征图包含丰富的空间细节信息, 仅仅校准不能消除这两者之间的差异, 因此在校准之后还要通过门控策略减小差异. 最终的输出如公式 (3) 所示:

$$D = \alpha_l \otimes L' + \alpha_h \otimes H', \quad (3)$$

其中  $\alpha_l$  和  $\alpha_h$  为使用卷积块生成的门限特征图.

### 3 实验与结果分析

#### 3.1 实验数据

##### 3.1.1 实验数据来源

目前城市地下管道缺陷分割领域尚无具有分

割标注的公开数据集,本文的实验数据来源于2021年CVPR会议论文发布的Sewer-ML数据集<sup>[6]</sup>.该数据集主要应用于地下管道多标签缺陷分类任务,没有缺陷分割任务的标注,不能直接用于分割实验,本文选取了在实际占比中较多、较为典型的缺陷类型进行缺陷语义分割数据的标注,共标注了900张图像.缺陷类型包括管道错口(DJ, Displaced Joint)、

腐蚀(CR, Corrosion)、沉积(DP, Deposit)、树根(RO, Root)、支管暗接(BP, Branch Pipe)、管道结垢(SG, Surface Gathering),加上背景类别(BG, Background)共7种类别,为了更易于区分,对不同类别使用不同颜色进行展示,如图9所示,第一行为原始缺陷样本,第二行为缺陷标签.

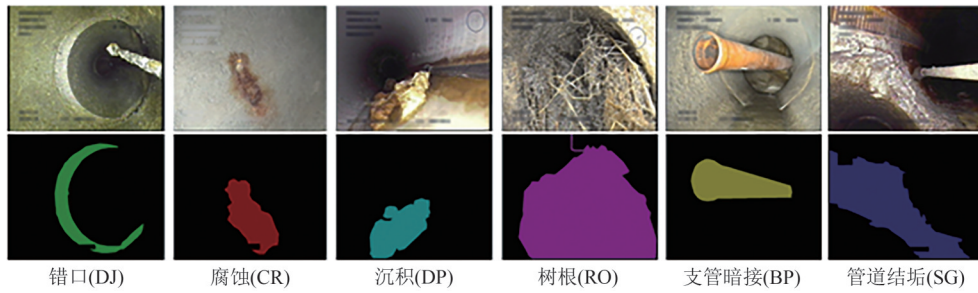


图9 地下管道缺陷样本类型

Fig. 9 Underground pipeline defects samples

### 3.1.2 数据增强

为了验证模型的泛化能力,本文取原始数据的20%作为测试集,对剩余数据采取随机色彩抖动、

随机角度旋转、垂直翻转的方式进行数据增强,如图10所示,增强后的数据集达到4320张.对增强后的数据集按照9:1划分为训练集和验证集.

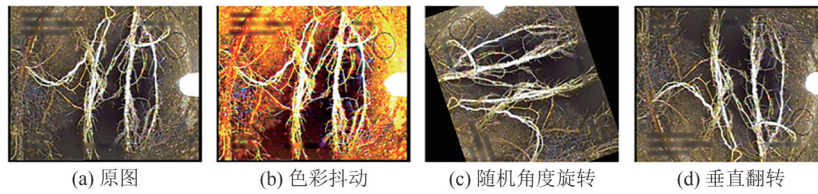


图10 数据增强

Fig. 10 Data augmentation

### 3.1.3 实验环境及参数设置

本文实验硬件环境为:操作系统为64位的Windows11, CPU为Intel i5-11400,内存16G, GPU为NVIDIA GeForce RTX 3070,显卡内存为8GB;软件环境为:Pycharm平台下Python 3.8,深度学习框架为Pytorch 1.10.1和CUDA 11.4.最大学习率设置为 $5 \times 10^{-4}$ ,最小学习率为 $3 \times 10^{-6}$ ,使用Adam优化器,损失函数使用Focal Loss和Dice Loss,学习率优化策略采用带有热身机制的余弦下降策略.迭代次数设置为120次,前50次迭代冻结骨干网络权重,只训练其他部分的参数,冻结阶段批处理大小设置为8,后70次迭代对模型整体进行训练,批处理大小设置为4.

### 3.2 评价指标

语义分割任务中,常用的评价指标有平均交并比(mIoU)、平均像素精确度(mPA)、频率权重交并比(FWIoU).

IoU用于衡量模型的预测区域和真实标签区域

的重合程度,是两个区域的交集和并集的比值,mIoU则是数据集中每一个类别的IoU的平均值,其计算公式如下:

$$mIoU = \frac{1}{k} \sum_{i=1}^k \frac{TP}{FN + FP + TN}, \quad (4)$$

其中 $k$ 为类别数量,TP表示预测为正样本且真实标签也为正样本的像素,FN是预测为负样本但真实标签是正样本的像素,FP是预测为正样本但真实标签是负样本的像素,TN是预测为负样本且真实标签也为负样本的像素.

PA用于计算预测正确的像素数占总像素数的比例,mPA就是对每一个类别取平均,其计算公式如下:

$$mPA = \frac{1}{k} \sum_{i=1}^k \frac{TP + TN}{TP + FN + FP + TN}. \quad (5)$$

频率权重交并比(FWIoU)是根据每一类别的频率和该类别的分割IoU进行加权计算,其常常应用于类别不均衡的情况,在这种情况下直接对所有类

别求 mIoU 存在一定的不合理性,因为模型针对样本较少的类别无法得到充分的训练,在这个类别上的精度不可避免地偏低,因此引入了频率权重交并比作为衡量指标,计算公式如下:

$$FWIoU = \frac{TP + TN}{TP + FP + FN + TN} \times \frac{TP}{TP + FP + FN} \quad (6)$$

Accuracy 用于计算预测正确的样本占全部样本的比例,计算公式如下:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (7)$$

表 1 不同语义分割模型对比实验

Tab. 1 Comparative experiments of different semantic segmentation models

Method	Backbone	mIoU/%	mPA/%	FWIoU/%	Accuracy/%	Params	Fps
UNet <sup>[20]</sup>	VGG-16	64.65	73.19	84.45	91.25	24.9M	21.92
PSPNet <sup>[16]</sup>	ResNet-50	67.59	75.97	85.54	92.02	46.7M	48.92
DANet <sup>[20]</sup>	ResNet-50	65.04	77.03	84.22	91.03	47.4M	48.34
Deeplabv3+ <sup>[20]</sup>	MobileNetV2	68.29	76.8	85.60	92.01	<b>5.81M</b>	63.09
DDRNet <sup>[23]</sup>	—	67.01	79.26	85.18	91.69	32.65M	47.42
CSFCN <sup>[11]</sup>	—	66.37	79.68	84.63	91.23	13.4M	<b>85.49</b>
BiSeNetV2 <sup>[9]</sup>	—	68.52	78.25	85.69	92.03	13.5M	62.13
<b>XS-BiSeNetV2</b>	—	<b>71.17</b>	<b>79.72</b>	<b>86.30</b>	<b>92.43</b>	13.74M	49.08

由表 1 可知,本文提出的改进模型与原始的 BiSeNetV2 模型相比,在参数量上略微增加,但在 mIoU 上提升了 2.65%,mPA 提升了 1.47%,速度下降了约 13 帧.特征提取网络为 MobileNetV2<sup>[19]</sup> 的 Deeplabv3+ 模型以最小的参数量 5.81 M 做到了 68.29% 的 mIoU 和 63 帧的速度.CSFCN 模型拥有最高的速度,达到了 85.49 帧,但在分割精度上较低,只有 66.37% 的 mIoU.对比常用的轻量级模型,提出的 XS-BiSeNetV2 在 mIoU 和 mPA 上为最高,达到了 71.17% 和 79.72%,且在分割速度上能达到 49 帧/s.

图 11 展示了不同语义分割模型在城市地下管道缺陷图像上的分割效果.由图 11 可知,改进的 XS-BiSeNetV2 模型对比几种常用的语义分割模型在上述缺陷中拥有更好的表现.对比第 1 行样本可以发

### 3.3 实验结果分析

#### 3.3.1 不同语义分割模型对比实验

为验证 XS-BiSeNetV2 模型针对城市地下管道缺陷的分割能力,将该模型与常用的轻量级语义分割模型进行了对比实验,实验结果如表 1 所示,表中的列分别表示语义分割模型名称、该模型选用的主干特征提取网络、模型在数据集上的 mIoU、模型在数据集上的 mPA、模型的 FWIoU、模型的 Accuracy 和参数量以及模型的分割速度.

现 U-Net 和 PSPNet 以及 DANet 对错口类别分割效果不佳,DANet 和 CSFCN 模型错将一部分树根缺陷分割成了结构缺陷.对比第 2 行样本可以发现 U-Net、Deeplabv3+、DDRNet、CSFCN、BiSeNetV2 都错误地将一部分管道结构缺陷分割成了管道沉积缺陷,只有 DANet 和本文提出的模型将缺陷区域正确分割.对比第 3 行样本可以发现几乎所有的模型都存在误测,PSPNet 未分割出腐蚀区域,DANet、CSFCN 和 BiSeNetV2 模型错将部分腐蚀区域分割成管道结垢区域,提出的 XS-BiSeNetV2 模型在样本的左半区域的分割最接近真实标签,但右半区错误识别出了腐蚀缺陷,Deeplabv3+ 模型的分割效果最接近真实标签.对比第 4 行样本可以发现,几乎所有模型对支管暗接缺陷都有较好的分割效果,DANet 对样本左上

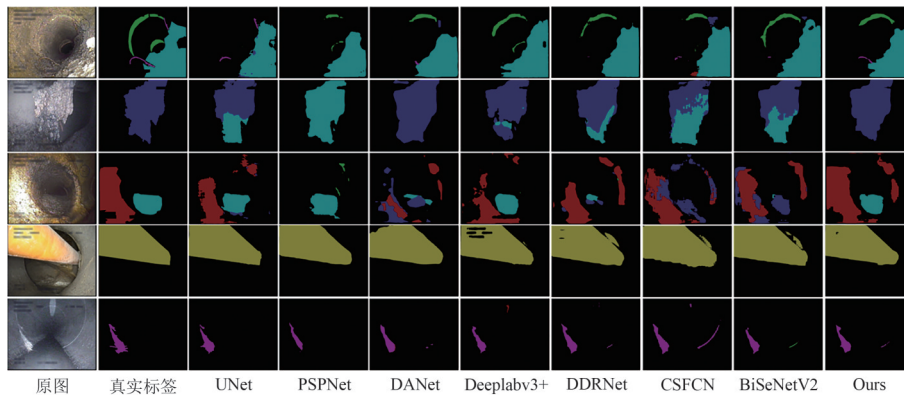


图 11 对比实验分割结果可视化

Fig. 11 Visualization of segmentation results in contrast experiments

角区域存在错误分割,DDRNet、CSFCN、BiSeNetV2 模型都存在割裂的区域,Deeplabv3+模型正确地将马赛克区域分割.对比第 5 行样本可以发现 Deeplabv3+错误地识别出了腐蚀缺陷,CSFCN 错误地识别出了树根缺陷,BiSeNetV2 模型错误地识别出了错口缺陷.多种模型评价指标的结果和在与训练集互斥的测试集的可视化结果表明:对比其他常用模型,本文提出的模型拥有较好的泛化能力.

### 3.3.2 消融实验

为了验证添加的各个模块的有效性,本文在原

BiSeNetV2 模型中依次添加改进的几种模块,以 BiSeNetV2 模型为基线模型,对改进的交叉特征提取模块、Haar 小波变换下采样模块和上下文和空间校准模块进行了消融实验,结果如表 2 所示,最左侧一列表示模型添加的模块,其中 2X 代表交叉特征提取模块、HWD 代表 Haar 小波变换下采样模块、CS 代表上下文和空间校准模块.中间 7 列为腐蚀(CR)、管道错口(DJ)、支管暗接(BP)、管道结垢(SG)、树根(RO)、沉积(DP)、背景(BG)的分割 IoU,右侧 3 列为模型评价指标.

表 2 各模块消融实验

Tab. 2 Ablation experiments between modules

Method	IoU/%							mIoU/%	mPA/%	FPS
	CR	DJ	BP	SG	RO	DP	BG			
Baseline	58.49	65.03	78.66	52.3	68.69	64.75	91.69	68.52	78.25	<b>62.13</b>
+2X	64.16	63.32	79.07	55.54	74.03	62.21	<b>91.86</b>	70.03	79.3	61.68
+HWD	59.66	<b>65.97</b>	76.48	55.41	73.61	<b>68.49</b>	91.71	70.19	79.76	60.61
+CS	65.05	62.24	77.41	53.61	72.45	63.37	91.84	69.42	78.89	53.96
+2X&HWD	61.78	65.83	<b>80.64</b>	55.82	73.65	64.57	91.79	70.60	<b>79.86</b>	57.06
+2X&HWD&CS	<b>65.17</b>	62.87	80.56	<b>58.0</b>	<b>76.23</b>	63.72	91.63	<b>71.17</b>	79.72	49.08

由表 2 可知,使用原始的 BiSeNetV2 模型的 mIoU 为 68.52%,在 BiSeNetV2 中添加改进的模块,模型的分割 mIoU 都会提升,只加入交叉特征提取模块后 mIoU 提升了 1.51%,只使用 Haar 小波变换下采样模块替换细节分支的跨步卷积下采样后 mIoU 提升了 1.67%,只加入上下文和空间特征校准模块 mIoU 提升了 0.9%,但分割速度降低较多,同时使用交叉特征提取模块和 Haar 小波变换下采样模块后模型的 mIoU 提升了 2.08%.在同时进行两种改进,并加入上下文和空间特征校准模块后,mIoU 提升了

2.65%,mPA 也提升了 1.47%.模型引入的 3 个模块在提升分割精度的同时,会增加模型的参数,降低分割速度,改进的模型仍能以每秒 49 帧的速度处理图片,证明能够较好地支持实时视频分割.

图 12 展示了不同改进后的模型和原始 Deeplabv3+模型在城市地下管道缺陷图像上的分割效果.由图 12 可知,对比原始方法,加入 3 种改进的方法在分割精度和准确度上有优于原始模型的表现.对比第 1 行图像可以发现,加入交叉特征提取模块和 Haar 小波变换下采样模块,或者同时使用上述两个

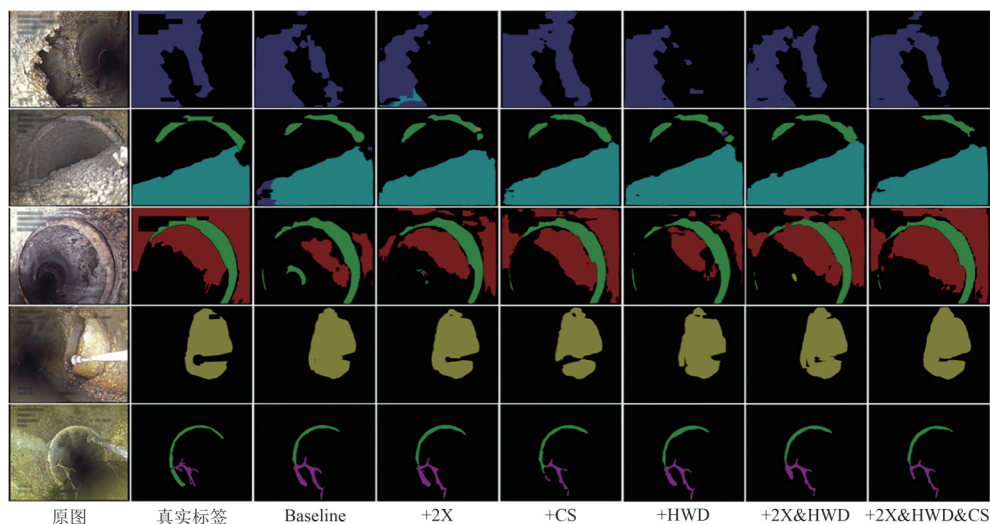


图 12 消融实验分割结果可视化

Fig. 12 Visualization of segmentation results in ablation experiments

模块的模型在对管道结垢缺陷的分割效果上都低于原始模型,同时使用3种改进的模型分割效果最接近于真实标签,拥有最高的IoU.对比第2行图像可以发现,原始模型错误地将部分沉积区域分割成了管道结垢区域,增加交叉特征提取模块和Haar小波变换下采样模块都提升了空间细节的预测效果,但仍存在类别错判,增加上下文和空间特征校准模块消除了错判的类别.对比第3行样本可以发现,同时使用3种改进的模型对错口和腐蚀样本拥有最好的分割效果.对比第4行样本可以发现,使用交叉特征提取模块的模型对支管暗接缺陷类别的分割效果优于使用Haar小波变换下采样模块和同时使用上述两种模块的模型,同时使用3种模块的模型分割效果最优.对比第5行样本可以发现,几种模型拥有类似的分割效果,都存在错误地将错口缺陷识别成树根缺陷的问题,同时使用3种模块的模型效果最好.

## 4 总结

本文结合了交叉特征提取策略、Haar小波变换下采样模块、上下文和空间特征校准模块及BiSeNetV2模型提出了基于XS-BiSeNetV2的城市地下管道缺陷语义分割方法,将模型的双分支策略改为交叉特征提取策略,替换传统的跨步卷积下采样为Haar小波变换下采样,在模型解码器部分融入多尺度信息,在模型预测时引入上下文和空间特征校准模块.对于目前没有城市地下管道缺陷语义分割数据集的情况,基于Sewer-ML数据集标注了用于该任务的数据集.在该数据集上的实验结果表明:XS-BiSeNetV2模型具有良好的语义分割能力.在后续的相关工作中,将从以下方面进行改进:(1)城市地下管道缺陷分割图像存在大量的噪声和对比度,可以引入更多的数据增强方法;(2)模型提升分割精度时,参数量也变大,后续在速度上仍有提升空间.

### 参 考 文 献

- [1] HUANG R, TAO Z, LIN Y, et al. Current situation of drainage pipe network in China and its detection technology: A brief review [J]. Polish Journal of Environmental Studies, 2024, 33(1): 19-29.
- [2] 王俊岭, 邓玉莲, 李英, 等. 排水管道检测与缺陷识别技术综述[J]. 科学技术与工程, 2020, 20(33): 13520-13528.
- [3] SADEGHKHAH A, AHMED E, KREBS P. Towards a decentralized solution for sewer leakage detection - a review [J]. Water Science and Technology, 2022, 86(5): 1034-1054.
- [4] 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.
- [5] WANG M, CHENG J C P. Semantic segmentation of sewer pipe defects using deep dilated convolutional neural network [C]//Proceedings of the International Symposium on Automation and Robotics in Construction. Banff: IAARC, 2019: 586-594.
- [6] HAURUM J B, MOESLUND T B. Sewer-ML: A multi-label sewer defect classification dataset and benchmark [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville: IEEE, 2021: 13451-13462.
- [7] ZHOU Q, SITU Z, TENG S, et al. Automatic sewer defect detection and severity quantification based on pixel-level semantic segmentation [J]. Tunnelling and Underground Space Technology, 2022, 123: 104403.
- [8] PAN G, ZHENG Y, GUO S, et al. Automatic sewer pipe defect semantic segmentation based on improved U-Net [J]. Automation in Construction, 2020, 119: 103383.
- [9] YU C, GAO C, WANG J, et al. BiSeNet V2: Bilateral network with guided aggregation for real-time semantic segmentation [J]. International Journal of Computer Vision, 2021, 129(11): 3051-3068.
- [10] XU G, LIAO W, ZHANG X, et al. Haar wavelet downsampling: A simple but effective downsampling module for semantic segmentation [J]. Pattern Recognition, 2023, 143: 109819.
- [11] LI K, GENG Q, WAN M, et al. Context and spatial feature calibration for real-time semantic segmentation [J]. IEEE Transactions on Image Processing, 2023, 32: 5465-5477.
- [12] HUANG Y, SHI P, HE H, et al. Senet: Spatial information enhancement for semantic segmentation neural networks [J]. The Visual Computer, 2024, 40(5): 3427-3440.
- [13] ZHOU D X. Theory of deep convolutional neural networks: Downsampling [J]. Neural Networks, 2020, 124: 319-327.
- [14] STANKOVIĆ R S, FALKOWSKI B J. The Haar wavelet transform: Its status and achievements [J]. Computers & Electrical Engineering, 2003, 29(1): 25-44.
- [15] SIAM M, GAMAL M, ABDEL-RAZEK M, et al. RTSeg: Real-time semantic segmentation comparative study [C]//

- 2018 25th IEEE International Conference on Image Processing. Athens: IEEE, 2018: 1603-1607.
- [16] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 6230-6239.
- [17] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.
- [18] YANG M, YU K, ZHANG C, et al. DenseASPP for semantic segmentation in street scenes [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 3684-3692.
- [19] SANDLER M, HOWARD A, ZHU M, et al. MobileNetV2: Inverted residuals and linear bottlenecks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 4510-4520.
- [20] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation [C]// Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015. Cham: Springer International Publishing, 2015: 234-241.
- [21] FU J, LIU J, TIAN H, et al. Dual attention network for scene segmentation [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 3141-3149.
- [22] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [C]// Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 833-851.
- [23] PAN H, HONG Y, SUN W, et al. Deep dual-resolution networks for real-time and accurate semantic segmentation of traffic scenes [J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 24(3): 3448-3460.

(责编 曹东,校对 雷建云)