一种轻量级地下管道缺陷多标签分类模型MG-MobileViT

文源,李波*,戴路,吴佳妮,罗理
(中南民族大学 计算机学院,湖北 武汉 430074)

摘 要 城市地下管道缺陷检测和分类是保障城市地下管道设施安全运行的关键环节,也是典型的多标签分类任务.地下管道缺陷具有复杂背景、缺陷尺度和形态差异大、缺陷种类不均衡、标签间语义关联复杂等特点,对其准确实时多标签分类具有较大挑战性.针对多标签分类任务中标签关联性建模不足等问题,提出了一种基于MobileViTv3架构的改进多标签分类模型MG-MobileViT.该模型在MobileViTv3主干网络的基础上,在分类器前引入并串联集成了多阶门控聚合模块(MoGA)和图注意力网络(GAT),以提升模型对复杂缺陷特征的捕捉能力;同时采用ASL损失函数以解决地下管道缺陷数据中固有的长尾和样本不平衡问题.实验使用Sewer-ML公开数据集,并采用地下管道领域中F2-CIW等指标作为核心评价指标.结果表明:所提出的模型在F2-CIW指标上达到54.15%,相较于主流模型Swin Transformer(46.75%)实现了显著的性能提升;mF1达到78.27%;改进后模型参数量较小,满足实时性能要求.实验充分验证了MG-MobileViT模型的有效性.

关键词 MG-MobileViT模型;多标签分类;城市地下管道缺陷;图注意力网络
中图分类号 TP391 **文献标志码** A **文章编号** 1672-4321(2026)04-0559-10
doi:10.20056/j.cnki.ZNMDZK.20260417

A lightweight multi-label classification model for underground pipeline defects using MG-MobileViT

Wen Yuan, Li Bo*, Dai Lu, Wu Jiani, Luo Li
(College of Computer Science, South-Central Minzu University, Wuhan 430074, Hubei China)

Abstract Defect detection and classification of urban underground pipelines is a key part of ensuring the safe operation of urban underground pipeline facilities, and it is also a typical multi-label classification task. Underground pipeline defects are characterized by complex backgrounds, significant difference in defect scale and morphology, uneven defect categories, and complex semantic relationships between labels, making accurate and real-time multi-label classification highly challenging. To address issues such as insufficient modeling of label correlations in multi-label classification tasks, an improved multi-label classification model, MG-MobileViT, based on the MobileViTv3 architecture, is proposed. Before the classifier, the model introduces and sequentially integrates a multi-level gated aggregation module and a graph attention network on the backbone of MobileViTv3 to enhance the model's ability to capture complex defect features; meanwhile, it uses the ASL loss function to address the inherent long-tail and sample imbalance issues in underground pipeline defect data. The publicly available Sewer-ML dataset is used in experiments. Core evaluation metrics in the underground pipeline domain, such as F2-CIW, were employed. The results show that the proposed model achieved 54.15% on the F2-CIW metric, significantly outperforming the mainstream Swin Transformer model (46.75%); the mF1 reached 78.27%; the improved model has a small number of parameters, meeting real-time performance requirements. Experiments fully validate the effectiveness of the MG-MobileViT model.

收稿日期 2026-01-02
* **通信作者** 李波(1975-),男,教授,研究方向:模式识别和机器学习,E-mail:libo@mail.scuec.edu.cn
基金项目 国家自然科学基金资助项目(61976226)

Keywords MG-MobileViT model; multi-label classification; urban underground pipeline defects; graph attention network

城市地下管道作为一项重要基础设施,承担城市污水的收集与输送职能,它在生产、生活环境保障中承担着重要职责[1].随着城市化的发展,地下管道负载越来越大,容易出现管道裂缝、变形、腐蚀、长树根、有沉积物等现象,进而造成城市内涝等问题.地下管道缺陷检测的相关技术有激光检测技术、声纳检测技术和闭路电视技术(Closed Circuit Television, CCTV)等[2].由于具有成本低、部署灵活等优势,CCTV检测技术已经成为地下管道检测任务中应用最广泛的技术[3].该技术由机器人进入到管道内部拍摄图像,再由专业人员对获取到的图像数据进行判读,对管道内的缺陷进行定位、分类和等级评定[4].

人工判读费时且易出错,因此自动化检测技术成为此领域的研究热点[5].Kumar等[6]提出了一种基于深度学习的方法,利用Faster R-CNN网络从闭路电视视频中自动检测并分类地下管道缺陷,显著提高了检测效率和准确性.Yin等[7]利用闭路电视视频作为输入数据,并采用YOLOv3网络作为目标检测器,以实现实时自动化的缺陷检测.Cha等[8]通过视觉图像自动检测混凝土表面的裂缝,利用CNN自动学习图像特征,避免了传统图像处理技术(IPTs)中复杂的特征提取步骤.Mondal等[9]通过融合深度信息显著提升了损伤分割算法的性能,并提出了一种新的体积损伤量化方法.田淙文等[10]提出FCC-Deeplabv3+模型,提升了模型在缺陷分割中的鲁棒性.

本文所讨论的多标签分类(Multi-Label Classification)[11]是另一种重要的视觉识别任务,其核心特征在于单个样本可以同时关联多个语义标签.多标签分类是图像级别的.在地下管道缺陷分类任务中,一张CCTV图像通常包含多种缺陷类型.多标签分类能够将缺陷描述得更精细化与丰富化,可以赋能精准的养护决策与维修优先级排序.多标签分类可以直接将原始的视觉检测结果,转化为富含语义信息和工程价值的“决策数据”,为管网的精准养护、科学管理和风险防控提供重要的技术支撑.然而地下管道多标签分类面临一些挑战:类别严重不均衡、标签间语义关联复杂、背景干扰强、缺陷特征微弱,等等.Chen等[12]首次用图卷积神经网络建模标签关系并生成相关依赖分类器.Zuo等[13]通过结合掩码注意力特征增强和标签关联学习来提升检测的性能.Zhao等[14]提出了一种嵌入打包方法,结合标签特定特征生成和模型归纳,解决多标签分类中的缺失标签问题.Hu[15]等提出一种图卷积网络的双分支结构,分别建模图像内部的结构关系和标签间的语义依赖.An等[16]提出一种基于Swin Transformer的方法,用于处理GPR(Ground Penetrating Rader,探地雷达)数据并实现缺陷的分类和定位.

城市地下管道样本背景复杂,缺陷种类多且数量极不均衡.针对此类任务,模型需在较低的参数量的约束下,具备强大的空间特征提取能力以识别微弱缺陷.尽管近年来涌现了FasterNet[17](通过部分卷积PConv追求极致推理速度)和MobileOne[18](利用结构重参数化实现超低延迟)等更先进的轻量级纯卷积模型,但在地下管道巡检这类视觉退化严重的任务中,纯卷积架构表现出明显的鲁棒性不足.针对内壁反光、淤泥遮蔽等造成的非均匀背景干扰,纯卷积算子受限于局部感受野,难以捕获跨区域的语义依赖,易导致特征判别失效.本文选择轻量级混合架构MobileViTv3[19]作为基准模型,其CNN-ViT混合设计能够协同提取局部细节与全局表征,有效应对管道图像背景复杂且缺陷特征微弱的挑战,是当前工业任务中兼顾性能与部署效率的理想平衡点.本文主要贡献包括:

(1)针对轻量级MobileViTv3骨干网络难以有效聚合多尺度、复杂空间上下文信息,导致特征判别力不足的问题,引入多阶门控聚合模块(Multi-order Gated Aggregation, MoGA)[20]通过动态门控机制,增强模型对复杂空间上下文的理解能力;

(2)针对地下管道缺陷特征的非局部空间依赖和标签之间的复杂关联难以显式建模的问题,引入图注意力机制(Graph Attention Network, GAT)[21]捕捉像素间的非局部依赖和标签间的复杂关联关系;

(3)针对地下管道高危缺陷样本稀少、训练过程中负样本淹没的问题,使用不对称损失函数(Asymmetric Loss, ASL)[22],提升对高危缺陷分类的召回率,以有效解决样本不平衡问题.

1 MobileViTv3模型原理

MobileViTv3[19]是一种专为移动端设计的轻量级混合架构,结合了CNN的局部特征提取能力与Vision Transformer的全局建模优势.模型以3×3卷

积起始，进行初步下采样；浅层（Layer 1–2）采用 MobileNetV2 倒残差块快速提取特征，深层（Layer 3–5）引入 MobileViTv3 块：先通过局部卷积捕获细节，再将特征重塑为序列送入轻量级 Transformer 编码器以建立全局依赖，最后通过逐点卷积将全局信息映射回空间域并与局部特征融合，形成“局部→全局→局部”的信息流．主干输出经 1×1 卷积扩展通道，再通过全局平均池化和全连接层，生成最终的类别预测．

尽管 MobileViTv3 在初步辨识管道背景方面具有优势，但其轻量化架构在多标签缺陷分类中仍存在局限：深层特征的语义聚合能力不足，难以捕捉“接口位移”与“渗漏”等缺陷间的空间协同规律；同时，全局平均池化前的特征表示易受环境噪声的干扰，限制了模型对长尾分布下细微高危缺陷的判别精度．为此，本文针对性地在分类器头部分引入了增强模块．

2　本文的改进模型 MG-MobileViT

为了提升 MobileViTv3 模型在城市地下管道多标签缺陷分类任务中的性能，本文引入 MoGA[20] 模块和 GAT[21] 模块对分类器进行增强．模型总体框架如图 1 所示，红色背景模块为本文改进之处．本文将改进重点置于分类头而非主干结构，主要基于计算效率与决策深度的权衡．一方面，在低分辨率的高层特征图上构建图网络，能显著降低 GAT 模块的计算复杂度和内存开销，规避浅层特征下的计算冗余；另一方面，在特征提取末端嵌入 MoGA 的空间重校准与 GAT 的非局部建模，可使模型在语义层面能更有效地捕捉缺陷间的共生规律，实现从基础特征提取到复杂关联决策的逻辑增强．主干网络输出的原始特征图 F 首先进入 MoGA 模块进行空间自适应重校准，生成精炼特征 F_{moga}；随后，该特征被转化为图节点序列 X_{flat} 并输入 GAT 模块，利用邻接矩阵 Adj 建模缺陷间的非局部空间关联．通过这种递进式的特征增强，模型最终经由 ASL 损失函数引导，实现对长尾分布下高危缺陷的精准识别．关于各模块的数学定义与实现细节，将在 2.1 至 2.3 节详细展开．

2.1　多阶门控聚合模块

为增强模型对复杂空间里上下文的理解能力，本文在 MobileViTv3 主干网络的分类器中，于池化层

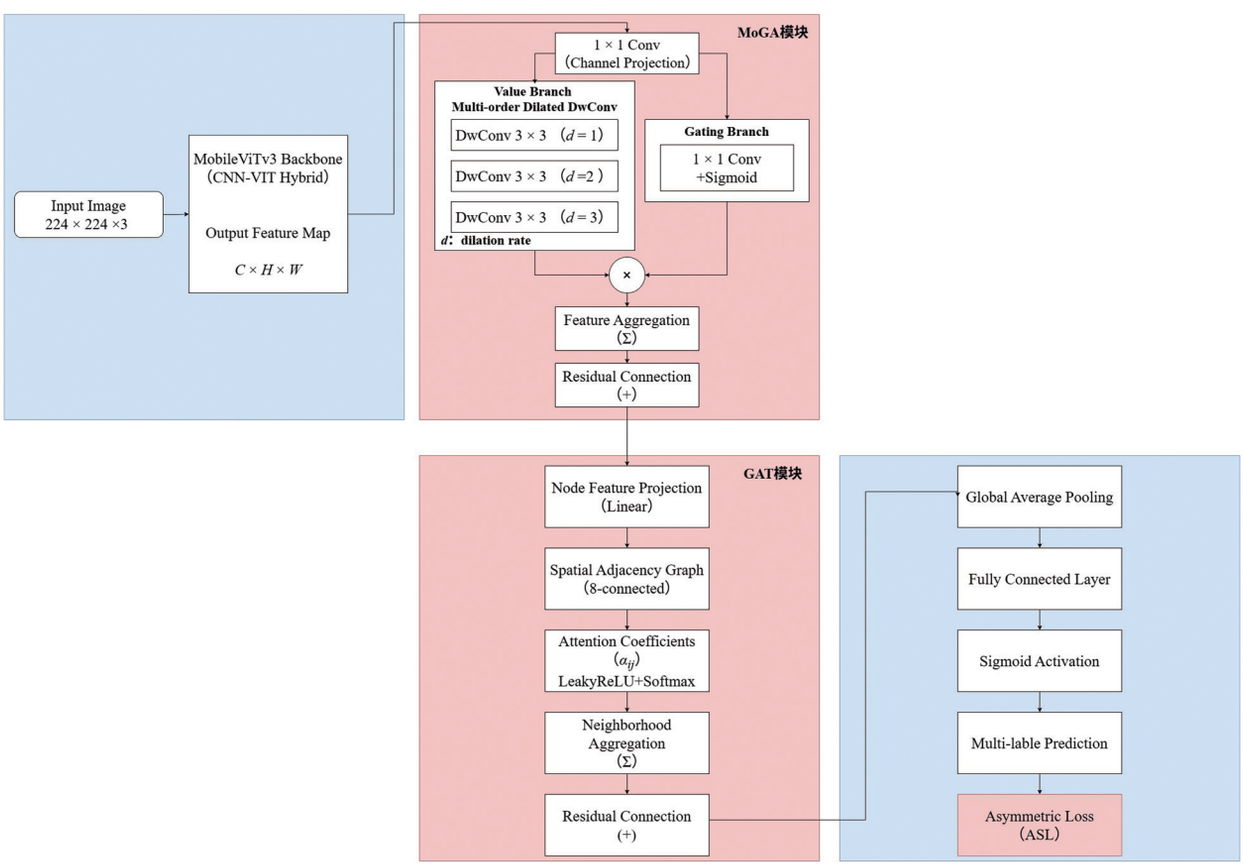


图 1　本文模型总体框架图

Fig. 1　Overall framework diagram of this model

之前,集成了一个多阶门控聚合模块MoGA,其核心设计旨在通过并行的多尺度特征提取和动态门控机制,有效捕获和融合不同层次的上下文结构信息,其中上下文理解能力主要体现在模型对不同空间尺度下特征依赖关系的综合感知,多阶空洞深度可分离卷积显示扩大有效感受野,而门控分支生成的动态权重则用于自适应调节各尺度特征的重要性,从而在管道复杂场景中强化关键信息、抑制冗余响应.

MoGA模块的输入首先经过一个1×1卷积进行初步的通道投影,随后特征被送入两个并行分支:值分支和门控分支.值分支利用多阶空洞深度可分离卷积将特征通道分组,并为每组应用不同膨胀率的深度卷积,从而同时捕捉精细的局部细节和广阔的全局上下文;与此同时,门控分支通过一个1×1卷积生成动态的门控系数,将值分支的输出进行逐元素相乘,以自适应地调整和加权各个通道的重要性,实现对多尺度特征的精准控制.

具体而言,膨胀率决定了卷积核的有效感受野:膨胀率1、2、3分别对应3×3 、7×7、15×15的感受野,使模型在单次前向传播中同时获得局部、中尺度和全局三个层次的特征表示.对于稀有高危缺陷(如OS仅35个训练样本),门控机制通过自适应权重分配,自动选择该缺陷在特征最显著尺度下的表示,避免在不相关尺度上浪费稀缺的梯度信号,从而在有限样本条件下实现有效学习.

此外,模块内部的特征分解步骤还利用全局平均池化来获取特征的全局统计信息,并结合可学习的缩放参数进行微调,进一步增强了对特征的判别力.最终,经过门控聚合的特征与模块的原始输入通过残差连接相加,共同形成了最终增强后的输出 $F_{\text{moga}} \in \mathbb{R}^{B\times N\times C}$.

2.2 图注意力网络

图注意力网络是一种基于图的神经网络,核心思想是通过注意力机制来聚合邻居节点的特征.在多标签分类中,不同标签之间可能存在复杂的关联,本文所指的标签间关联性主要体现在不同缺陷类型在空间分布和语义层面上的共现与依赖关系,例如管道的裂缝(RB)极易伴随渗水(IN/IS)的发生 ,而大量的沉积物(AF)可能导致障碍物(FO)的形成,理解并显式建模这种关联性对于准确识别复杂缺陷至关重要.本文引入的GAT模块旨在显式地建模特征图中像素或区域间的非局部、非规则依赖关系,其中这种特征层面的依赖可视为多标签缺陷在表征空间中的关联映射,并通过图注意力机制在特征传播过程中自适应融合相关节点信息,从而使模型在预测阶段能够综合考虑不同缺陷标签之间的依赖关系,更好地把握多标签之间的关联结构.它将MoGA输出的特征图 F_{moga} 视为一个图结构,其中每个空间位置是一个节点,并利用图注意力机制计算节点间的相互重要性,从而捕捉长距离的空间关联,这对于理解复杂缺陷形态如裂缝走向、孔洞分布等至关重要.GAT首先将输入特征图 F_{moga} 重塑为节点特征序列 X_{flat},过程如公式(1):

$$F_{\text{moga}} \in \mathrm{R}^{B\times C\times H\times W} \rightarrow X_{\text{flat}} \in \mathrm{R}^{B\times N\times C}, \tag{1}$$

其中,B 为批次大小,C 为通道数,H 和 W 为特征图的高度和宽度,N 为空间维度展平后的节点总数,$N = H \times W$.$\mathbb{R}$ 表示实数域,表明特征图中的所有元素均为实数.

然后,根据预设的图类型构建邻接矩阵 $\mathrm{Adj} \in \mathrm{R}^{N\times N}$ 以定义节点连接先验,本文使用的是8-连通空间邻接图,通过多头图注意力机制处理特征:对每个注意力头,节点特征 X_{flat} 经线性变换后,计算节点对的注意力分数 e_{ij},并通过Adj掩码和Softmax归一化得到注意力权重 α_{ij},用于聚合邻居信息.

从运算机制看,注意力权重 α_{ij} 的计算使得模型能够学习特征节点间的软连接关系.当训练集中FS与IN频繁共现时,反向传播会使模型参数倾向于增大这两类特征节点间的注意力分数,从而在推理阶段,即使某样本中IN的原始特征较弱(因样本不足导致特征提取不充分),通过加权聚合操作,强FS特征也能传播至IN节点,提升IN的预测置信度.这种基于注意力权重的特征传播机制,本质上是在特征空间中编码了缺陷间的共现统计规律,使模型能够利用频繁类别的强特征来辅助判别稀有类别,从而在长尾分布下提升召回率.

多头输出经拼接和可选的线性变换后,得到增强的节点特征.最后,这些特征被重塑回原图维度 (B, C, H, W),并与输入特征图 F_{moga} 通过残差连接相加,输出增强后的特征图.

2.3 损失函数

标准的二元交叉熵损失(BCE)对正负样本的误分类施加了对称的惩罚.然而,在地下管道多标签分类中,数据呈现稀疏性与非对称性:一方面,一张图片通常只有少数几个标签为1(正样本),绝大多数标签为0(负样本);另一方面,背景噪声产生的简单负样本占据主导.这种天然的不平衡会导致模型

被大量的负样本“淹没”，难以捕捉到稀有但高危的缺陷特征.ASL损失函数通过引入不对称的聚焦参数，降低简单负样本的梯度贡献，同时宽容对待正样本的漏检，从而促使模型优先学习识别稀有但高危的缺陷.ASL损失函数的计算公式如下：

$$\mathrm{ASL}=\begin{cases}L_{+}= & (1-p)^{\gamma_{+}}\log(p)\\ L_{-}= & (p_m)^{\gamma_{-}}\log(1-p_m)\end{cases},\tag{2}$$

其中，p代表模型对当前样本是正类的预测概率，p_m是针对负样本的修正概率，γ用于控制正负样本损失的权重.

3　实验与结果分析

3.1　实验数据

3.1.1　数据来源

本文的数据来源于CVPR会议的Sewer-ML[23]公开数据集.考虑到原始数据集规模非常庞大、训练成本极高，且本文研究重点在于模型架构的研究及改进，本文从原数据集中选取了一个具有代表性的子集（共30380张图像）作为实验数据来源.样本筛选主要依据以下原则：(1)覆盖全部17个有效缺陷类别，各类别均有样本；(2)兼顾单一缺陷及多缺陷共存的不同场景，保证样本多样性；(3)保持各缺陷类别原有的长尾分布特征.将数据集按照7∶2∶1划分为训练集、验证集和测试集，该比例在训练充分性与评估可靠性之间达到了最优平衡，在保证训练样本充分的同时，为模型调参和泛化性能评估提供了稳定的数据支撑，划分结果如表1所示.该数据集标记有18种不同地下管道缺陷类别，根据检验标准，每个类别都会得到一个代表该类别经济影响的分数，该分数越大，造成的经济影响越大，这个类别应更被检测者重点关注.这个分数由地下管道检测领域的专业人员确定.研究人员通过将所有分数除以最大分数归一化为区间[0,1]，将新值表示为类别重要权重（Class-importance Weight，CIW）.其中VA类与检测时的水位相关联，与本文目标无关，因此后续处理时忽略VA标签.地下管道缺陷类别数量及重要性权重[23]如表2所示，由表2可知缺陷分布呈现出典型的长尾分布，部分缺陷数量少，会使得模型在训练时对稀有类别特征提取不足，影响模型对高危缺陷识别的鲁棒性，后续将对数据进行预处理.划分后各子集所涵盖的缺陷数量见表3，其中ND是正常（即无缺陷）样本.

表1　数据集分布

Tab. 1　Dataset distribution

类型	训练集	验证集	测试集	总样本
正常	9901	2819	1418	14138
缺陷	11377	3264	1601	16242
总样本	21278	6083	3019	30380

表2　缺陷类别数量及其对应的CIW权重

Tab. 2　The number of defect categories and their corresponding CIW weights

缺陷代码	缺陷描述	CIW	缺陷数量	缺陷代码	缺陷描述	CIW	缺陷数量
VA	水位，水在管道深处	0.0310	-	AF	顽固沉积	0.0811	2945
IN	存在渗透物	0.3131	1176	BE	附着沉积	0.2275	2411
OB	表面损坏，被腐蚀等	0.5518	8447	FO	存在障碍物	0.2477	214
PF	管道生产的问题	0.2869	268	GR	存在支管	0.0901	1776
DE	管道变形	0.1622	276	PH	凿刻连接	0.4167	344
FS	接口位移	0.6419	11628	PB	孔洞连接	0.4167	103
IS	管道渗透物	0.1847	318	OS	横面的复原切口	0.9009	35
RO	存在树根	0.3559	606	OP	过渡连接缺陷	0.3829	58
RB	管道有裂缝、断裂和坍塌	1.000	1353	OK	与建筑风格相关的管道连接	0.4396	4093

3.1.2　数据预处理

为提升模型的泛化能力和对地下管道缺陷识别的鲁棒性，本文对训练数据实施了一套全面且多层次的数据预处理策略.首先，所有图像均被统一缩放至256×256像素，随后进行随机裁剪至224×224像素，以模拟不同尺度的观测视角.同时，为模拟真实场景中的遮挡和噪声，本文以0.3的概率进行随机灰度化，以0.2的概率进行随机擦除.所有图像均使用本数据集计算得出RGB三个颜色通道统计值进行归一化.通道均值（[0.523,0.472,0.367]）和标准差（[0.260,0.243,0.217]）分别用于中心化和统一尺度，以确保模型的输入特征分布稳定，并加速收敛.此外，针对横面的复原切口（OS）样本是极度稀有的高危缺陷类别，本文设计了增强策略：对包含这些标签的样本，采用更高的参数值（如更大的旋转角度（±45°）和更高的擦除概率（20%）），以人

工扩充其数据的多样性,缓解长尾分布带来的负面影响.

表3 各数据子集中缺陷数量

Tab. 3 Number of defects in each subset

缺陷代码	训练集	验证集	测试集	缺陷代码	训练集	验证集	测试集
RB	968	255	130	BE	1725	462	224
OB	5989	1687	862	FO	145	42	27
PF	190	54	24	GR	1262	341	173
DE	196	56	24	PH	253	58	33
FS	8117	2345	1165	PB	72	20	11
IS	221	72	25	OS	23	7	5
RO	450	110	46	OP	43	12	5
IN	819	232	125	OK	2885	835	373
AF	2035	604	306	ND	9901	2819	1418

3.1.3 实验环境及参数设置

实验硬件环境为:CPU为14 vCPU Intel(R) Xeon(R) Gold 6330 CPU@2.00GHz, GPU为RTX 3090(24GB),使用的深度学习框架是Pytorch2.5.1,CUDA版本是12.4.模型输入图片大小是224×224.模型的超参数设置为:Epoch为100,BatchSize为16,使用AdamW优化器,初始学习率为0.0001,权重衰减系数为0.0001,学习率用余弦退火策略.

在实验阶段,本文对公式(2)的γ值进行了对比测试,结果表明,针对数据集的标签特性,设置γ=4能在不损失整体准确率(mF1)的前提下,最大程度地提升模型对长尾分布中高危缺陷类的敏感度,因此本文实验γ的值据此设定.

3.2 评价指标

为科学地、全面地评估模型在地下管道多标签缺陷识别任务上的性能,本文采用一套多层次的评价体系,包括微观F1分数(mF1)、正常管道F1分数(F1-normal)、平均精度(mAP)以及加权F2分数(F2-CIW)[23].mF1是评估多标签分类模型整体性能的常用指标,能宏观地、均衡地评估模型的分类能力,计算公式如下:

$$\mathrm{mF1}=\frac{2\cdot P\cdot R}{P+R}, \tag{3}$$

其中,P和R分别指微观精确率(Micro Precision)和微观召回率(Micro Recall),计算公式如下:

$$P=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}, \tag{4}$$

$$R=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}, \tag{5}$$

其中,TP代表模型正确预测为正例的样本数量,FP代表模型错误预测为正例的样本数量,FN代表模型错误预测为负例的样本数量.

F1-normal是计算模型在"无缺陷"(ND)类别上的F1分数,计算公式如下:

$$\mathrm{F1-normal}=2\cdot\frac{P_{\mathrm{ND}}\cdot R_{\mathrm{ND}}}{P_{\mathrm{ND}}+R_{\mathrm{ND}}}, \tag{6}$$

其中,P_{ND}和R_{ND}分别是"无缺陷"类别的精确率和召回率.

AP是某类缺陷在P-R(Precision-Recall)曲线下的面积,评价其在对应召回率下的精确率.mAP是对每个缺陷类别的AP取算术平均值.AP和mAP的计算公式如下:

$$\mathrm{AP}=\int_0^1 P(R)\mathrm{d}R, \tag{7}$$

$$\mathrm{mAP}=\sum_{i=1}^{C}\frac{\mathrm{AP}_i}{C}. \tag{8}$$

F2-CIW是由国际权威CVPR2021会议论文[23]针对地下管道缺陷检测任务提出的评价指标,该指标通过将统计学度量与工程风险评估相结合,精准反应模型识别高危缺陷的能力.根据论文[23],F2-CIW为地下管道缺陷多标签分类领域的核心评价指标,其计算公式如下:

$$\mathrm{F2-CIW}=\frac{\sum_{c=1}^{C}\mathrm{F2}_c\cdot\mathrm{CIW}_c}{\sum_{c=1}^{C}\mathrm{CIW}_c}, \tag{9}$$

其中,C是缺陷类别的总数(本文中C=17),$\mathrm{F2}_c$是第c个缺陷类别的F2分数,CIW_c是对应类别的权重.

除上述分类性能指标外,本文还采用FPS(Frames Per Second,每秒帧数)评估模型的推理速度.在地下管道实际检测中,CCTV检测机器人需要实时处理视频流,FPS越高意味着模型实时性越好,检测效率越高.本文实验中,FPS在GPU环境(RTX 3090)下测试,输入图像尺寸为224×224像素.

虽然宏观F1分数(Macro-$F1$)常用作对不平衡数据的补充评估(它赋予每个缺陷类别相等的权重),但在地下管道缺陷检测中,不同缺陷的经济影响和工程风险差异巨大.由于CVPR2021会议论文[23]提出的F2-CIW指标已通过CIW权重显式地整合了各缺陷类别的重要性,使其成为更符合工程实际的加权宏观评价指标,因此,本文省略了标准的Macro-F1,而将F2-CIW作为衡量模型在识别高危缺陷方面性能的核心评价指标.

3.3 实验结果分析

3.3.1 不同多标签分类算法对比实验

为了体现本文所用MG-MobileViT在多标签分

类效果上的提升，实验从多个角度对模型的性能进行分析，将本文模型与目前常用的多标签分类模型进行了对比，实验结果如表4所示.

表4　不同分类算法对比实验结果

Tab. 4　Comparison experimental results of different classification algorithms

Model	F2-CIW/%	F1-normal/%	mF1/%	mAP/%	Params	FPS
MobileViTv3[19]	19.55	82.44	60.18	45.06	5.82M	487.41
ResNet-50[24]	20.96	81.25	68.97	54.95	23.54M	470.13
EfficieNet-B0[25]	34.15	89.56	75.55	64.80	4.03M	491.73
Swin Transformer[26]	46.75	91.46	76.38	70.53	86.76M	310.26
ConvNeXt[27]	45.95	90.55	79.69	71.54	87.58M	354.78
本文模型	54.15	89.66	78.27	68.52	7.03M	399.78

由表4可以看出，本文模型在多标签分类领域最核心评价指标 *F*2-CIW 上有着最好的检测效果，达到了 54.15%，此外 mF1 也达到了 78.27%. 与 Swin Transformer 和 ConvNeXt 相比，本文模型的 F1-normal 和 mAP 均有差距，但仍领先于 ResNet 和 EfficieNet 两个模型，主要是因为 Swin Transformer 和 ConvNeXt 有巨大的参数量和计算复杂度，有更强的通用特征提取能力，尤其在大型数据集上更具优势. 本文模型在保持轻量化（参数量 7.03M）的基础上，通过 MoGA 和 GAT 模块实现了对缺陷局部细节和复杂标签关联性的有效建模，所得到的结果仍保持了较高的竞争力. 尤其在 F2-CIW 核心评价指标上，领先于 Swin Transformer 和 ConvNeXt. 本文模型充分体现了在资源受限场景下，通过对高危缺陷特征的精准增强，实现了对识别能力优化的高效性和实用性. 从多标签分类效果和模型轻量化等方面综合考量，本文模型达到了最优平衡.

3.3.2　各类缺陷的 F2-CIW 分数及其 CIW 权重比较

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 F2-CIW 分数及其 CIW 权重进行比较，如图2所示.

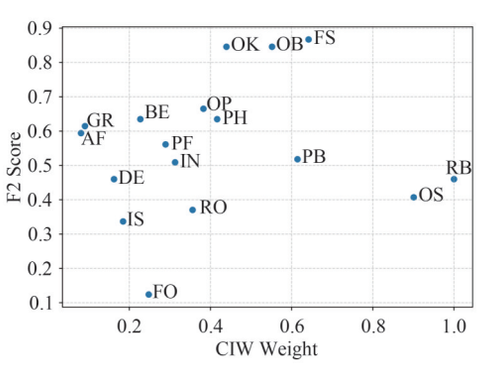


图2　缺陷的 CIW 权重与 F2-CIW 分数

Fig. 2　CIW weight and F2-CIW score of defect

本文模型在对经济影响程度较高的 FS（接口位移）、OB（表面损坏、腐蚀等）以及 OK（与建筑风格相关的管道连接）等缺陷类别上表现出较好的识别性能，这三类缺陷在实际地下管道检测中不仅出现频率较高，而且往往直接影响管道的结构稳定性和运行安全性，因此在 CIW 加权评价体系中被赋予了相对较高的权重，模型在这些类别上的性能优势能够在 F2-CIW 指标上得到充分体现，说明本文提出的方法在关注高经济影响程度的缺陷方面具有良好的适应性和工程实用价值. 相比之下，对于功能型缺陷 FO（存在障碍物）、IS（管道渗透物）和 RO（存在树根），模型的识别性能整体偏低，这一现象一方面源于该类缺陷在形态结构和视觉表现上的高度多样性，例如障碍物类型复杂、树根生长形态跨度较大，导致显著的类内差异；另一方面也与其在 CIW 权重体系中对应的经济影响程度相对较低有关，其识别性能的波动对整体 F2-CIW 评价结果的影响相对有限. 对于 RB（管道有裂缝、断裂和坍塌）和 OS（横面的复原切口）这两类缺陷，作为工程安全评估中尤为关键的高风险缺陷类型，其 CIW 权重显著高于其他类别，模型当前对其识别性能处于中等水平，这表明，尽管引入的 ASL 损失函数和多尺度特征建模策略在一定程度上缓解了稀有高危缺陷的识别难度，但在复杂场景下仍存在进一步优化的空间，考虑到 F2-CIW 指标对高权重类别具有更高的敏感性，若在后续工作中进一步提升 RB 和 OS 两类缺陷的召回性能，将对整体 F2-CIW 分数的提升产生更为显著的促进作用.

3.3.3　消融实验分析

为了分析各改进模块对本文模型性能的影响，实验依次添加每一种优化方法，对各方法进行消融实验，并分析优化方法对模型检测效果的影响，结果如表5所示.

由表5可知，原始 MobileViTv3 模型在处理复杂特征时，其特征提取能力明显不足. 增加 MoGA 模块

表5 消融实验对比

Tab. 5 Comparison of ablation experiments

Model	F2-CIW/%	F1-Normal/%	mF1/%	mAP/%	Params
MobileViTv3	19.55	82.44	60.18	45.06	5.82M
+MoGA	50.06	90.21	74.15	50.44	6.33M
+GAT	52.07	90.10	71.05	49.29	7.03M
+ASL	54.15	89.66	78.27	68.52	7.03M

后,各项指标都有显著提升,尤其是F2-CIW提升了30.51%.引入GAT后,3项指标都有轻微下降,主要是因为GAT将计算资源和注意力转移到更复杂的非局部空间关联上,将F2-CIW提升了2.01%.ASL为了追求稀有标签的高召回率,鼓励模型将决策阈值放宽,使得原本很容易被识别的无缺陷类(ND)中一些位于决策边上的图片被错误地预测为存在缺陷.以F1-normal的微小降低为代价,提升了其他3项指标,这对于地下管道缺陷检测任务是更有价值的.

3.3.4 多标签分类结果及缺陷关联性分析

为直观展示MG-MobileViT模型在实际多标签分类任务中的性能,本文选取了若干具有代表性的测试样本,其预测标签结果如图3所示.与单分类任务中使用Softmax激活函数不同,多标签分类任务需要允许单个样本同时属于多个类别.因此,类别存在的预测概率P_k采用Sigmoid激活函数进行计算,对每一类缺陷分别进行独立概率建模,从而实现对复杂场景下多种缺陷共存情况的有效识别,公式如下:

$$P_k = \frac{1}{1 + e^{-O_k}}, \tag{10}$$

其中,P_k是第k个缺陷存在的预测概率,O_k是模型对第k个类别的原始输出分数(Logit).该函数将O_k映射到[0,1]的范围内,每个P_k彼此独立.

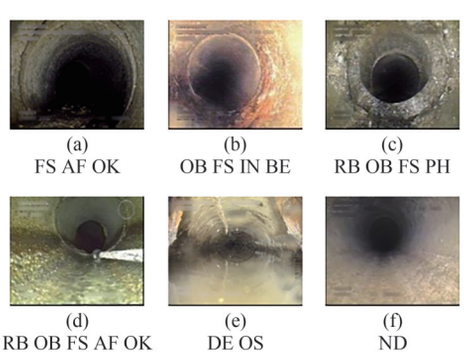


图3 多标签分类结果示例

Fig. 3 Example of multi-label classification results

图3中预测Label为图片存在预测缺陷标签.预测概率P_k的阈值为0.5,大于0.5则判定该缺陷存在.结果表明,本文所提出的MG-MobileViT模型成功实现了对复杂场景下多标签缺陷的高效识别,并充分体现了GAT模块对标签间复杂关联的有效建模能力.具体而言,在图3(a)所示的场景中,模型同时预测出FS、AF和OK三类缺陷,其对应的预测概率分别为0.7421、0.7079和0.7147,均高于设定阈值,表明模型能够在复杂背景下稳定识别接口位移、顽固沉积以及与建筑风格相关的管道连接等空间上高度耦合的共存缺陷,体现了模型对多标签共现特征的良好判别能力.在图3(b)中,模型预测出OB、FS、IN和BE四类缺陷,其预测概率分别为0.7547、0.7862、0.7455和0.7139,反映了模型不仅能够准确识别接口位移和表面损坏等显著缺陷,还能够进一步捕捉由此引发的渗透现象及其对沉积物附着形成的影响,说明模型具备刻画缺陷间关联演化关系的能力.对于图3(c),模型预测出RB、OB、FS和PH,其对应预测概率分别为0.7770、0.7624、0.7597和0.7521,成功反映了由局部结构破坏引发的应力集中过程,并进一步演化为裂缝或坍塌等高危缺陷,表明模型能够对复杂结构性失效链进行有效建模.在图3(d)中,模型同时识别出RB、OB、FS、AF和OK五类缺陷,其预测概率分别为0.7331、0.6850、0.7081、0.6619和0.7662,说明即使在局部结构性缺陷与大范围表面缺陷同时存在的复杂场景下,模型仍能够保持稳定的多标签判别性能,体现了其对多尺度特征的综合建模能力.在图3(e)所示场景中,模型对DE和OS的预测概率分别为0.7122和0.7635,能够准确识别由管道变形引发的接口结构异常及其后续失效形式,而在图3(f)中,模型仅预测出ND(无缺陷)这一标签,其预测概率达到0.8508,说明模型在无缺陷场景下同样具有较高的判别置信度和良好的误检抑制能力.综合上述结果可以看出,借助GAT模块在特征层面对不同缺陷之间关联关系的显式建模,模型能够在预测阶段综合考虑多种缺陷之间的协同与依赖信息,从而在保证伴随缺陷全面检测的同时,实现对高危缺陷(如RB)的稳定识别和可靠判别.

4 总结

本文针对城市地下管道缺陷检测中效率低、标签关联建模不足及样本不平衡等问题,将MoGA模块和GAT模块串联集成到MobileViTv3主干网络,并通过ASL损失函数优化了模型对稀有高危缺陷

的识别能力.这一改进使得模型在Sewer-ML数据集上的核心指标F2-CIW达到54.15%,显著超越了SwinTransformer等现有最优算法(46.75%).这不仅证实了MoGA和GAT模块在增强复杂上下文理解和捕捉标签关联方面的有效性,更为地下管道自动化检测提供了高效且高性能的解决方案.尽管本模型在常见高经济影响缺陷上表现良好,但未来工作将集中于进一步提升对功能型缺陷和极高重要性稀有缺陷的识别能力,以实现更全面、更可靠的城市管道安全评估.

参 考 文 献

[1] Wang J, Liu G H, Wang J, et al. Current status, existent problems, and coping strategy of urban drainage pipeline network in China[J]. Environmental Science and Pollution Research, 2021, 28(32): 43035-43049.

[2] Sadeghikhah 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.

[3] 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.

[4] Chandure P, Waghmare C, Ansari K. Review on guidelines of sewer inspection[J]. Chemik, 2020, 2: 160-167.

[5] Moradi S, Zayed T, Golkhoo F. Review on computer aided sewer pipeline defect detection and condition assessment[J]. Infrastructures, 2019, 4(1): 10.

[6] 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: 04019047.

[7] Yin X, Chen Y, Bouferguene A, et al. A deep learning-based framework for an automated defect detection system for sewer pipes[J]. Automation in Construction, 2020, 109: 102967.

[8] 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.

[9] Ghosh Mondal T, Jahanshahi M R, Wu Z Y. Deep learning-based RGB-D fusion for multimodal condition assessment of civil infrastructure[J]. Journal of Computing in Civil Engineering, 2023, 37(4): 04023017.

[10] 田淙文, 李波, 蓝雯飞, 等. 基于FCC-Deeplabv3+的城市地下管道缺陷语义分割方法[J]. 中南民族大学学报(自然科学版), 2025, 44(1): 107-117.

[11] Tarekegn A N, Giacobini M, Michalak K. A review of methods for imbalanced multi-label classification [J]. Pattern Recognition, 2021, 118: 107965.

[12] Chen Z M, Wei X S, Wang P, et al. Multi-label image recognition with graph convolutional networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA. IEEE, 2019: 5172-5181.

[13] Zuo X, Sheng Y, Shen J, et al. Multilabel sewer pipe defect recognition with mask attention feature enhancement and label correlation learning[J]. Journal of Computing in Civil Engineering, 2025, 39: 04024050.

[14] Zhao D, Tan Y, Sun D, et al. Multi-label learning of missing labels using label-specific features: An embedded packaging method[J]. Applied Intelligence, 2024, 54(1): 791-814.

[15] Hu J, Mou X, Li B, et al. Compound facial expression recognition based on graph convolutional multi-label learning [C]//2024 4th International Conference on Electronic Information Engineering and Computer Science (EIECS). Yanji, China. IEEE, 2024: 433-437.

[16] An J, Yang L, Hao Z, et al. Investigation on road underground defect classification and localization based on ground penetrating radar and Swin transformer [J]. International Journal for Simulation and Multidisciplinary Design Optimization, 2024, 15: 7.

[17] Chen J, Kao S H, He H, et al. Run, don't walk: Chasing higher FLOPS for faster neural networks [C]// 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, BC, Canada. IEEE, 2023: 12021-12031.

[18] Vasu P K A, Gabriel J, Zhu J, et al. MobileOne: An improved one millisecond mobile backbone [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, BC, Canada. IEEE, 2023: 7907-7917.

[19] Wadekar S N, Chaurasia A. MobileViTv3: Mobile-

friendly vision transformer with simple and effective fusion of local, global and input features [EB/OL]. (2022-09-21) [2026-01-02]. https://arxiv. org/abs/2209.15159.

[20] Li S, Wang Z, Liu Z, et al. Moganet: Multi-order gated aggregation network [EB/OL]. (2022-11-04) [2026-01-02]. https://arxiv.org/abs/2211.03295.

[21] Veličković P, Cucurull G, Casanova A, et al. Graph attention networks [EB/OL]. (2017-10-30) [2026-01-02]. https://arxiv.org/abs/1710.10903.

[22] Ridnik T, Ben-Baruch E, Zamir N, et al. Asymmetric loss for multi-label classification [C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, QC, Canada. IEEE, 2021: 82-91.

[23] 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, TN, USA. IEEE, 2021: 13451-13462.

[24] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA. IEEE, 2016: 770-778.

[25] Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks [C]//International Conference on Machine Learning. Long Beach: PMLR, 2019: 6105-6114.

[26] Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows [C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, QC, Canada. IEEE, 2021: 9992-10002.

[27] Liu Z, Mao H, Wu C Y, et al. A ConvNet for the 2020s [C]// 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA. IEEE, 2022: 11966-11976.

(责编 曹东,校对 雷建云)