中国科学: 技术科学 2026 年 第 56 卷 第 7 期: 1724 ~ 1740
SCIENTIA SINICA Technologica techcn.scichina.com
《中国科学》杂志社 SCIENCE CHINA PRESS
论 文 复杂恶劣条件下水电开发基础科学问题专刊

多模态感知复杂场景下振捣质量识别方法研究

任炳昱[1*], 张承轩[1], 杨志华[1], 关涛[1], 何伟[2]

1. 天津大学水利工程智能建设与运维全国重点实验室, 天津 300072
2. 水电水利规划设计总院有限公司, 北京 100011
* E-mail: renby@tju.edu.cn

收稿日期: 2025-12-16; 接受日期: 2026-03-23; 网络版发表日期: 2026-07-10
国家自然科学基金项目(批准号: 52222907, 52379131, 52350420)和中国电力建设股份有限公司科技项目(编号: DJ-ZDXM-2020-50)资助

摘要 振捣质量对混凝土强度和耐久性至关重要, 现场振捣质量识别采用人工旁站监督方式, 存在主观性强、劳动强度大的缺点; 基于计算机视觉方法虽有效提高了识别客观性和自动化水平, 但存在数据利用深度不足和不能满足现场全天候施工需求的缺陷, 模型在夜晚、雾天等复杂场景下的识别能力显著下降. 针对上述问题, 提出一种感知并融合混凝土表面图像数据和振捣施工数据(振捣时长、振捣深度)、混凝土材料信息(坍落度、含气量等)参数数据的多模态识别方法. 首先, 在振捣台车上集成姿态传感器、激光传感器、GNSS、工业摄像头等设备, 有效监控并得到振捣多模态数据集, 使用图像增强技术模拟复杂场景下的混凝土图像效果; 其次, 提出MetaData-EfficientNet-B4深度学习模型, 采用EfficientNet-B4网络提取图像模态特征, 在提出的MetaData模块中实现图像模态特征和参数模态特征的深度融合; 最后, 基于B/S架构开发简单易操作的振捣质量识别系统, 用于辅助现场施工和仓面验收. 工程应用表明, 提出的多模态识别模型相比传统方法在白天、夜晚和雾天的识别准确率由95.8%, 90.7%和91.7%提高到96.8%, 96.2%和95.8%, 多模态融合有效提高了模型在复杂场景下的性能, MetaData模块融合方法相比其他融合方法在TOPSIS评价方法中更具优势.

关键词 混凝土大坝, 计算机视觉, 多模态融合, MetaData-EfficientNet-B4, 振捣质量识别

1 引言

振捣是保障混凝土硬化质量与力学性能, 进而确保大坝工程结构安全的关键施工工艺之一. 目前针对振捣质量的研究重点在振捣密实机理和施工质量评价方法上. 理论研究常假设混凝土为Bingham流体[1], 采用理论推导、离散元法等分析其流动特性和能量传播过程等; 温家馨等人[2]根据能力传递将振捣过程分为两部分, 第一过程为水泥浆对空隙填充, 第二过程为混凝土流动, 气泡上浮逸出; Zhao等人[3]指出振捣过程本质是通过振捣设备产生的冲击波使水泥砂浆发生液化作用, 引发混凝土颗粒之间的相对运动, 从而促使内

引用格式: 任炳昱, 张承轩, 杨志华, 等. 多模态感知复杂场景下振捣质量识别方法研究. 中国科学: 技术科学, 2026, 56: 1724–1740
Ren B Y, Zhang C X, Yang Z H, et al. Research on vibration quality recognition method based on multi-modal sensing in complex scenes (in Chinese). Sci Sin Tech, 2026, 56: 1724–1740, doi: 10.1360/SST-2025-0415

© 2026 《中国科学》杂志社

部气泡向上迁移并逸出, 减小颗粒间的空隙, 提高混凝土的密实度与耐久性. 振捣施工过程中认为施工混凝土表面无明显塌陷, 汽泡不再上浮即振捣合格. 现实中影响施工质量的因素是多方面的, 包括振捣插入深度、振捣时长等施工控制因素, 以及混凝土坍落度和含气量等材料特性[4].

目前施工中常采用事中监理旁站监督和事后抽芯检测抗压强度方式保证振捣施工质量, 该方法高度依赖现场人员施工经验, 且检测方式具有明显滞后性, 难以补救施工损失. 物联网和计算机技术的发展有效带动了行业进步[5~7], Gong等人[8]采用UWB技术定位实现手持式振捣棒定位并可视化振捣范围; Tian和Bian[9]提出通过卫星定位技术, 开发定量检测振捣轨迹和振捣时长的监控系统, 评估混凝土浇筑过程中的缺陷; 钟登华等人[10]使用随机森林算法实时预测混凝土抗压强度, 有效提高振捣判断实时性; 刘亚结[11]提出基于立体视觉的混凝土振捣质量监测方法, 用算法追踪振捣棒位置及振捣插入深度; Lee和Skibniewski[12]提出一种振捣预警系统, 当传感器收集数据超出阈值, 可以对现场工人发出预警信息; Liu等人[13]应用卷积神经网络和图像分析技术来精确识别振捣时长, 其结果可以快速评估振捣质量; Wang等人[14]将混凝土表面图像输入至ResNet网络, 实现振捣质量实时识别, 有效应用于现场施工. 然而, 仅混凝土图像特征不能全面反映振捣施工质量, 振捣数据利用深度不足. 模型表现高度依赖输入模型的图像质量, 在夜晚、雾天等低照度场景下模型性能下降, 不能满足现场全天候施工需求[15].

多模态融合是指将来自多个模态的数据(如图像、文本、音频、视频等)进行综合处理, 以获得比单一模态更丰富的语义表达和更好的任务性能, 通过结合不同模态的信息, 可以更全面地理解数据, 其被广泛应用于地貌演变研究[16]、情感识别[17]和医疗诊断[18]等领域的多个场景. 常用的多模态融合方法有四种[19], 分别是特征级融合、模型级融合、决策级融合和混合级融合. 张君等人[20]综合声学和运动学模态数据识别碾压机机械活动识别过程, 实现堆石坝施工仿真参数的实时建模; 姜智颖等人[21]拼接红外和可见光数据获得广角合成图, 但这种线性融合方式结构简单, 不能充分发掘不同模态数据内含的信息, 对模型准确率提高不大. 随着深度学习技术发展, 有学者尝试将其用于多模态融合领域, Li等人[22]提出MetaNet模块用于疾病识别, 该方法使用一维卷积来提取参数数据特征, 得到的结果与CNN训练图像数据的Feature Map相乘, 大大提高了模型预测准确率; Yuan等人[23]提出多模态融合模型DMF, 结合图像、近红外光谱与介电数据实现砂砾含水率高精度测量, 在复杂干扰下模型仍保持优异性能; Pacheco和Krohling[24]受LSTM思想启发, 设计MetaBlock模块用于融合图像和元数据信息, 也取得较好结果; 陈辉等人[25]提出一种以生成式对抗网络为基本框架, 结合多尺度增强网络实现对水下退化图像的增强; Dan等人[26]提出融合称重与视频信息的桥梁移动载荷识别方法, 可实现桥面纵横向载荷的实时全桥识别; Shi等人[27]融合了红外图像和可见光图像, 并提出了一种基于动态稀疏注意力机制的深度学习网络用于路面裂缝检测, 为像素级路面裂缝检测提供了技术支持; Zhang等人[28]收集图像和点云数据, 先使用特征级融合两种模态信息, 后使用坐标注意力机制融合方法挖掘更深层次信息, 提高了3D目标检测准确性. 振捣过程涉及丰富的多模态数据, 包括混凝土表面图像模态、振捣插入深度、振捣时长、混凝土材料信息等模态. 有效感知并融合这些数据用于质量识别, 可以有效提高模型识别准确率, 且在低照度场景中, 不同模态数据互补可以保证模型性能稳定.

综上, 本文针对混凝土大坝振捣施工过程中存在人工劳动强度大, 计算机模型适用范围有限等难点, 提出一种基于多模态感知的复杂场景下振捣质量识别方法; 针对振捣领域缺乏公开多模态数据集的问题, 采用物联网技术监控振捣施工过程, 并采用图像增强方式获得多场景图像数据, 匹配构建训练数据集; 针对多模态数据如何高效融合问题, 提出MetaData-EfficientNet-B4网络, 已有研究多聚焦于感官类数据融合, 其多模态数据均为不同传感器对对象不同角度的物理感知信号, 而本文融合的是视觉图像与结构化施工及材料参数这类过程性知识类数据, 更模拟现场施工人员判断方式, 其挑战在于建立低纬度特征数据到高纬度视觉特征数据的联系, 本文提出的MetaData模块正是针对上述难点设计, 该模块采用多分支结构和注意力机制, 提高模态数据间交互表达能力. 工程应用表明, 相比于不融合, 线性融合、深度学习Meta-Block等融合方法, 本文提出的MetaData融合方法识别精度更高, 尤其在夜晚和雾天低照度场景下能够大幅提升模型性能.

2 研究路线

本文提出的振捣施工质量识别方法包括数据获取与处理、振捣质量识别模型和工程应用三部分. 研究框架如图1所示, 具体实现步骤如下.

(1) 在振捣台车上集成GNSS、姿态传感器、激光测距传感器、工业摄像头等设备, 在不同时段监控得到振捣多模态数据, 对图像模态采用基于灰度变化暗化处理和基于标准光学模型雾化处理模拟夜晚和雾天场景下摄像头得到的图像效果, 匹配构建用于模型训练的数据集.

(2) 提出MetaData模块用于多模态数据融合, 该模块采用多分支结构、不同分支采用不同大小卷积核提取图像数据特征, 引入注意力机制增强与参数模态数据交互, 引入残差机制, 增强模型训练稳定性和表达能力.

(3) 以东庄水利工程现场为例进行分析, 将MetaData-EfficientNet-B4模型与传统模型和其他融合方法进行对比分析, 从识别精度、鲁棒性和性能评价角度证明本研究提出的识别方法优越性.

3 多模态感知复杂场景下振捣质量识别模型

3.1 多模态数据实时采集方法

在振捣台车上集成物联网和多源传感器, 采用工业摄像机拍摄振捣施工区域, 另外采用人工补充拍摄盲区作业, 采用固定帧率截取即获得原始混凝土表面图像. 混凝土材料信息选择坍落度、含气量、加水量和减水剂四项指标. 其中, 坍落度[29]反映了混凝土工作性, 直接影响成型后的硬化性能和耐久性; 含气量[30]影响混凝土强度, 当混凝土承载压力时, 大气泡的存在可能导致应力集中, 造成混凝土开裂, 威胁大坝体安全性; 加水量及减水剂[31]也会影响混凝土流动性. 根据现场规定, 每一方混凝土在拌合楼出机口由专业人员测量记录并上传后方数据库, 当所有指标符合标准后才可用于仓面浇筑. 现场调研发现, 同一天

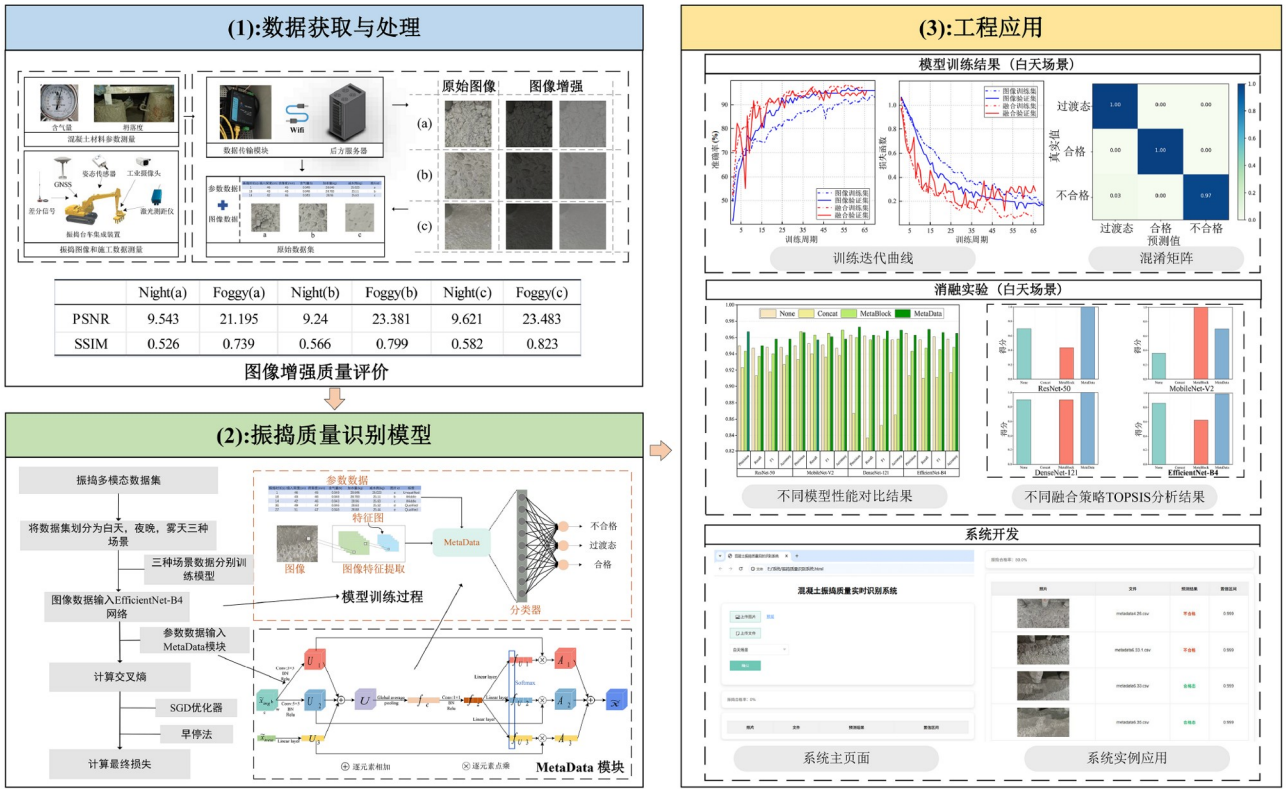


图 1 (网络版彩图)研究框架

Figure 1 (Color online) Research framework.

生产的混凝土材料信息有高度相似性, 因此, 选取系统中最近一方混凝土材料信息可作为此时刻施工区域混凝土材料信息的代表.

振捣过程可分为四部分: 移动、插入、振捣和拔出. 激光测距传感器安装在振捣台架处, 测量台架处距离混凝土表面距离, 施工按照国标要求秉持快插慢拔原则, 一次振捣过程在数据上表现为折线图, 如图2所示, 根据式(1)线性变化可得到振捣插入深度, 根据式(2)可得到振捣时长:

$$L = I - d, \tag{1}$$

其中, L为振捣插入深度, cm; I为传感器安装位置距离振捣棒末端距离, cm; d为传感器测量距离混凝土表面距离, cm.

$$T = T_4 - T_2, \tag{2}$$

其中, T为振捣时长, s; T_4为拔出开始时刻; T_2为插入结束时刻.

所有数据经过系统解析计算实时传入后方数据库, 按照相同时刻将图像模态和参数模态数据匹配, 得到原始振捣多模态数据集, 过程如图3所示.

3.2　多模态数据预处理

大规模图像数据是计算机视觉技术成功应用的前提, 原始振捣图像数据相似性高, 据此训练的模型鲁棒性较差, 为了提高模型鲁棒性并模拟复杂场景施工环境, 我们采用图像增强技术对得到的图像数据进行处理, 包括基于灰度变化的暗化处理模拟夜晚施工场景和基于标准光学模型的雾化处理模拟雾天施工场景.

3.2.1　基于灰度变化暗化处理

尽管夜晚的现场施工有灯光照明, 但仍存在模板、两岸高仓或振捣台车自身遮挡的低照度盲区(本研究统一将其称为夜晚场景), 这些区域的振捣图像数量少, 我们采用算法合成的方式获取.

调节图像亮度的方法种类繁多[32], 典型的方法是基于直方图均衡化和基于灰度变化, 直方图均衡化方法能提高图像对比度, 但易出现色差问题, 并且其变换函数较难控制, 容易造成图像失真. 考虑到单次振捣工作范围小且图像像素细节变化小, 区域亮度大体一致的特征, 基于灰度变化的方法即可满足夜间图像需求, 该方法通过遍历图像元素生成不同程度暗化的图像, 效果如图4所示.

3.2.2　基于标准光学模型雾化处理

随着大坝施工逐渐迈向高寒高海拔地区, 山上温度随着海拔升高而下降, 山谷、盆地等地形较为封闭, 不易形成对流, 山地雾气不易消散. 不可避免会遇到雾天施工场景, 这类数据数量少且收集困难, 本文采用算法合成方式得到雾天图像数据.

计算机加雾算法在工程上已有一定应用[33], 本文采用基于标准光学模型合成方法生成雾天图像, 该方法基于Koschmieder等人[34]建立的大气散射模型, 该模型认为雾天成像是由物体反射光透射部分和大气光散

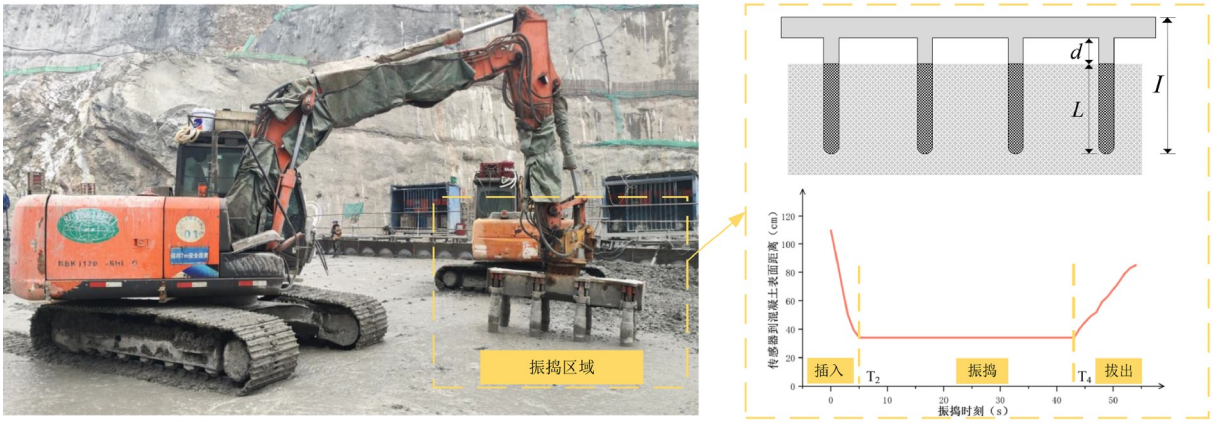


图 2　(网络版彩图)振捣施工参数获取

Figure 2　(Color online) Acquisition of vibration parameters.

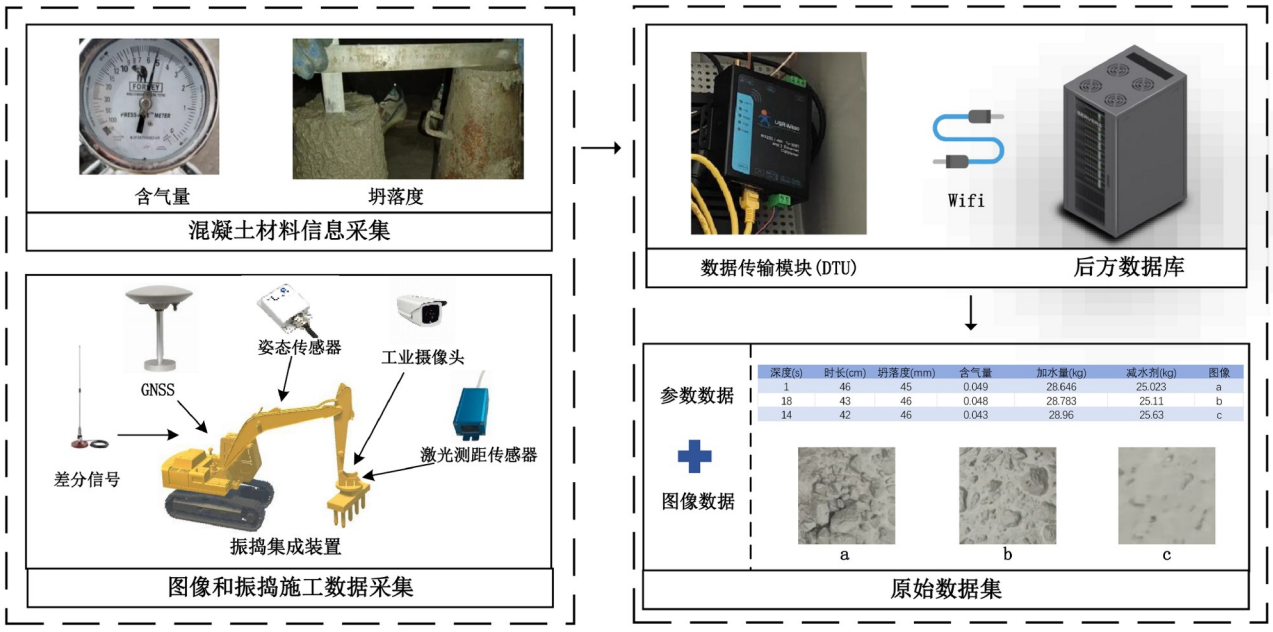


图 3 (Color online)多模态数据收集过程
Figure 3 (Color online) Multi-modal data collection process.



图 4 (网络版彩图)暗化处理, e为调节暗化程度参数
Figure 4 (Color online) Image darkening, e is the gray level processing parameter.

射后进入系统的部分之和, 其数学表达式为

$$I(x) = J(x)t(x) + A(1 - t(x)), \tag{3}$$

其中, $I(x)$为观测到的有雾图像; $J(x)$为对应的无雾图像; A为大气光强度; $t(x)$为截止透射率, 当大气密度均匀时, $t(x)$可表示为

$$t(x) = \mathrm{e}^{-\beta d(x)}, \tag{4}$$

其中, β为大气散射系数, 用来控制雾的厚度, β越大, 雾就越厚; $d(x)$代表像素点处场景深度.

通过此算法合成的有雾图像效果如图5所示.

3.3 多模态融合振捣质量识别深度学习模型

目前计算机视觉技术在振捣质量识别研究中存在两方面局限性, 其一是训练数据仅采集于白天的理想场景, 据此训练的模型难以适用在其他低照度场景下. 其二是仅利用单一图像模态数据, 振捣数据利用深度不足, 模型性能有限. 为解决上述问题, 本研究提出多模态融合振捣质量识别深度学习模型, 该模型由两部分组成: (1) 主干网络采用EfficientNet-B4模型提取图



图 5　(网络版彩图)雾化处理, β为调节雾化程度参数
Figure 5　(Color online) Foggy images, β is the atomization level processing parameter.

像模态特征; (2) 融合方法采用创新提出的MetaData模块, 试验证明该模块可以有效融合多模态信息提高模型识别准确率. 模型融合流程如图6所示.

3.3.1　主干网络EfficientNet-B4

EfficientNet是由Tan和Le[35]提出的一种深度学习模型. 其系列网络在ImageNet top-1上取得了本年度最高的准确率(84.3%). 根据以往经验, 增加网络深度可以获得更丰富、更复杂的特征. 此外模型更便于迁移到其他视觉任务中. 但是过深的网络会出现梯度消失或梯度爆炸的问题. 反而降低网络性能; 增加网络宽度可以获得更高的细粒度特征, 使模型更便于训练. 然而对于宽度过大而深度过小的网络, 在任务中很难学习到更丰富特征; 提高图像分辨率可能获得更精细的特征, 但大分辨率图像无疑增加了模型的计算量.

EfficientNet-B4网络采用神经架构搜索(NAS)技术获取到最合适的网络宽深比, 有效提高了模型在图像识别领域的上限, 其主要结构MBConv如图7所示.

3.3.2　多模态融合MetaData模块

受SK-Net网络[36]启发, 我们提出MetaData模块用于图像模态和参数模态数据融合任务, 该模块采用多分支结构, 有效解决异构信息维度不匹配问题, 其中图像模态使用3×3和5×5两种尺度卷积核提取特征, 3×3小卷积核擅长捕捉局部细节, 如振捣图像细节纹理等特征, 5×5大卷积核擅长捕捉图像轮廓等全局结构特征. 参数模态经过线性连接层提取特征, 通过注意力机制进行权重分配, 动态调整融合时的信息优先级, 得到最终用于预测的特征$\tilde{x}$. MetaData模块结构如图8所示.

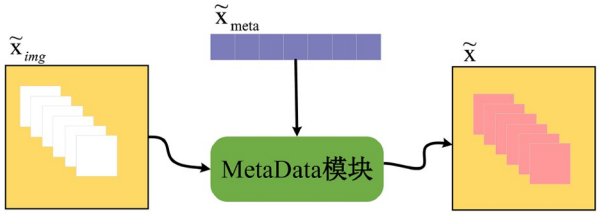


图 6　(网络版彩图)模型融合流程图
Figure 6　(Color online) Model fusion process.

我们定义特征提取器为ψ, 图像模态经过EfficientNet-B4网络提取得到特征$\tilde{x}_{img}$, 参数数据经过人工筛选去掉错误、重复的信息, 得到参数特征$\tilde{x}_{meta}$, 如式(5)和(6)所示:

$$\tilde{x}_{img} = \psi_{img}(x_{img}), \tag{5}$$

$$\tilde{x}_{meta} = \psi_{meta}(x_{meta}). \tag{6}$$

$\tilde{x}_{img}$经过两种不同尺度卷积核(3×3和5×5)进行卷积得到U_1和U_2特征图, 如式(7)所示, 经过BN归一化层, ReLU函数激活, $\tilde{x}_{meta}$经过单层神经网络得到U_3特征图, BN归一化层, ReLU函数激活:

$$\begin{cases} U_1 = \psi_{3\times3}(\tilde{x}_{img}), \\ U_2 = \psi_{5\times5}(\tilde{x}_{img}), \\ U_3 = \psi_{\text{Linear-layer}}(\tilde{x}_{meta}). \end{cases} \tag{7}$$

相加得到融合两种模态及不同感受野的信息U, 如式(8)所示:

$$U = U_1 + U_2 + U_3. \tag{8}$$

对U经过平均池化层得到关于通道信息的一维向

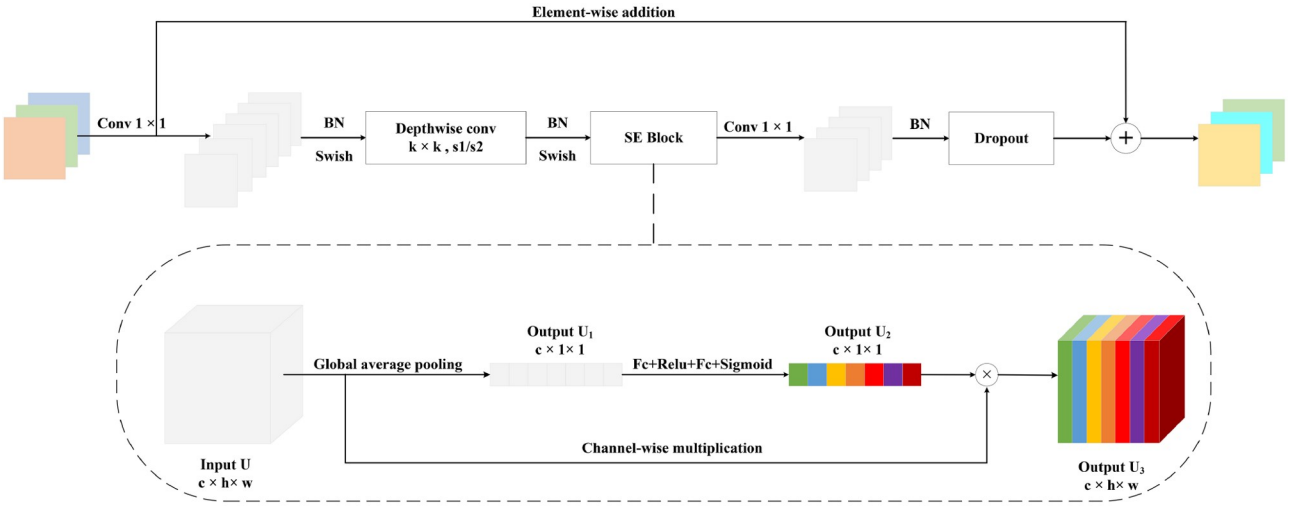


图 7 (网络版彩图)MBConv结构图

Figure 7 (Color online) MBConv structural schematic.

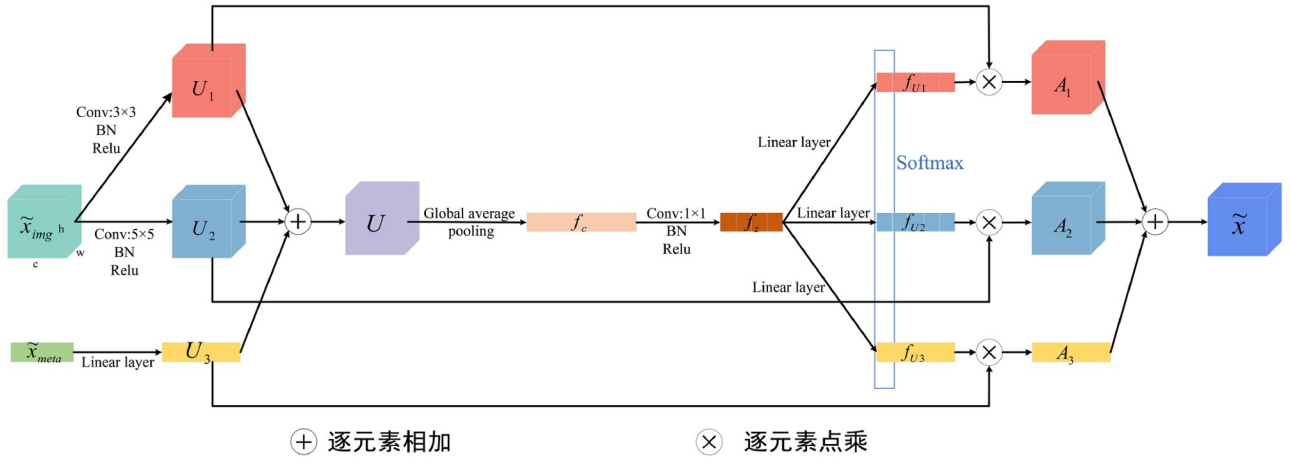


图 8 (网络版彩图)MetaData结构

Figure 8 (Color online) MetaData architecture.

量f_c, 如式(9)所示, 反映了各通道上信息重要程度:

$$f_c = \psi_{\text{pool}}(U). \tag{9}$$

如式(10)所示, f_c经过线性变化得到f_z, 对f_z经过三个线性变化得到三个特征图, 一系列操作完成对通道信息的提取, 对结果进行Softmax操作重新分配权重, 如式(11)所示, 代表原通道的重要程度, 与U_1, U_2, U_3相乘得到各分支重分配权重后的结果A_1, A_2, A_3, 如式(12)所示, 相加得到MetaData模块处理后的特征$\tilde{x}$, 输入致分类层得到综合了两种模态互补信息的振捣分类结果, 如式(13)所示:

$$f_z = \psi_{\text{linear}}(f_c), \tag{10}$$

$$\begin{cases} f_{U1} = \text{Softmax}\left(\psi_{\text{linear}}(f_z)\right), \\ f_{U2} = \text{Softmax}\left(\psi_{\text{linear}}(f_z)\right), \\ f_{U3} = \text{Softmax}\left(\psi_{\text{linear}}(f_z)\right), \end{cases} \tag{11}$$

$$\begin{cases} A_1 = f_{U1} \times U_1, \\ A_2 = f_{U2} \times U_2, \\ A_3 = f_{U3} \times U_3, \end{cases} \tag{12}$$

$$\tilde{x} = A_1 + A_2 + A_3. \tag{13}$$

我们在公开的PAD-UFES-20数据集上训练和评估了我们的模型，该数据集包含6种皮肤癌类型(分别为ACK, BCC, MEL, NEV, SCC, SEK)的医学图像以及相应的患者元数据(如年龄、身高和病史). 实验绘制ROC曲线，并选取ACC、BACC、AUC作为模型性能评价指标. ROC曲线能够直观体现模型单类二分类区分能力；ACC为整体识别准确率；BACC是平衡准确率，适配类别数量失衡的数据集；AUC代表ROC曲线下的面积, 其数值越大, 模型对对应类别的识别区分效果越好. 为了验证所提模块的有效性, 我们将基线EfficientNet-B模型(仅图像)与包含患者信息的MetaData-EfficientNet-B4模型进行了比较. 如表1所示. 鉴于数据集中存在严重的类不平衡, 与基线模型相比, MetaData-EfficientNet-B4模式的BACC提高了近11个百分点，证明了所提出的MetaData模块在多模态融合任务中的有效性. 图9所示为ROC曲线. 与仅图像模型相比, 融合模型保持了整体性能(Micro AUC=0.93，Macro AUC=0.91)，同时显著提高了特定类别(如MEL，NEV和SCC)的分类能力. 值得注意的是，MEL的AUC达到了1.00的满分，证明了多模态融合在提高模型性能方面的有效性.

4 工程应用

本研究以我国东庄水利枢纽工程为研究平台. 将匹配构建后的多模态数据分为三类：不合格、过渡态和合格. 得到1180份样本数据, 包括353份不合格样本, 334份过渡态样本和493份合格样本. 不合格样品的特点是粗骨料堆积严重，骨料之间有较大间隙，没有下沉现象；过渡态样本表现为骨料初步下沉，混凝土表面不规则；合格样本表现为骨料下沉明显，气泡基本逸出, 混凝土表面平整.

表 1　PAD-UFES-20数据集上模型性能

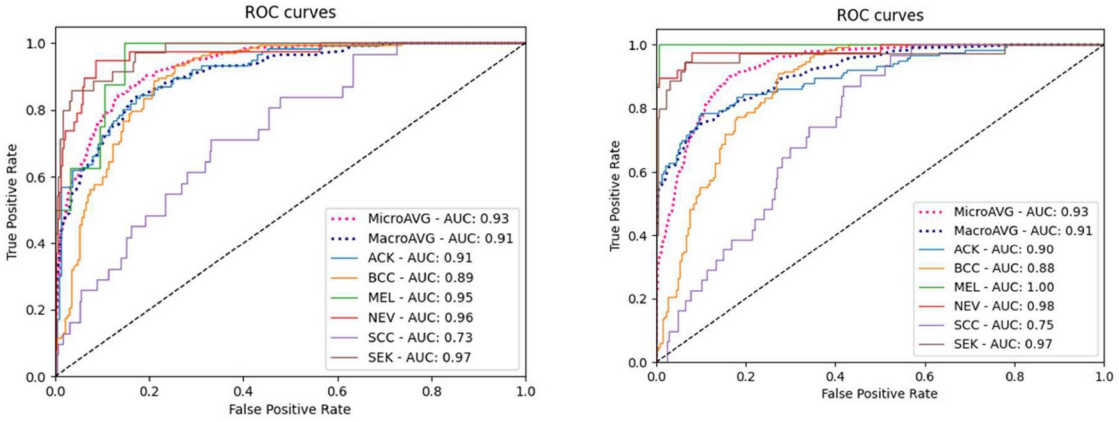
Table 1　Performance of the models for the PAD-UFES-20 dataset

模型	ACC	BACC	AUC
EfficientNet-B4	0.664	0.613	0.905
MetaData-EfficientNet-B4	0.761	0.721	0.933

4.1 图像增强处理

对上述数据集进行图像增强处理，为避免图像过分失真, 暗化图像取参数e=0.5, 雾化参数取β=0.08, 随机选择不同标签样本，如图10所示，采用峰值信噪比PSNR和结构相似性SSIM两项指标评价, PSNR值越高表示失真越小，SSIM值越接近1表示结构保存越完好. 结果如表2所示.

以一组不合格标签为例, 分析结果发现, 雾天场景下PSNR值为21.195, 夜晚场景下PSNR值为9.543; 雾天场景下SSIM值为0.739, 夜晚场景下SSIM值为0.26. 雾天场景两项指标值相比都大于夜晚场景, PSNR结果表



(a)EfficientNet-B4 Model　　(b) MetaData-EfficientNet-B4

图 9　(网络版彩图)不同模型ROC曲线

Figure 9　(Color online) ROC curves for different models.

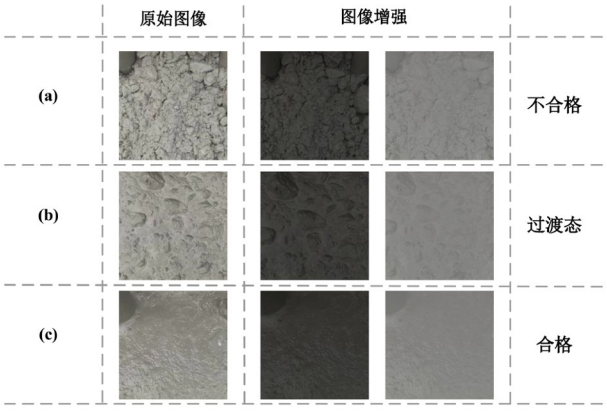


图 10 随机抽取混凝土样本
Figure 10 Randomly selected concrete samples.

明雾天图像数据包含的噪声较小, 信号强度强, 图像质量相对较好. SSIM指标结果表明, 雾天图像相比夜晚与参考图像结构更相似, 视觉特征更明显. 综合来看, 夜晚和雾天两种场景相比白天场景得到的图像质量都更差, 其中夜晚场景对于模型学习更具有挑战性.

4.2 模型训练策略

本实验先按5:1划分训练集和验证集, 再对训练集采用五折交叉验证训练模型, 将三种场景分开输入, 训练专注于一种场景的模型, 保证后续系统性能最高效. 模型提前在ImageNet数据集上预训练, 使用从torchvision API得到的权重, 采用SGD优化器, 同时, 为避免过拟合, 采用早停法策略, 参数选择15, 即模型在15个训练周期内损失函数不再减小, 则模型停止训练. Batchsize设置为24, 最大Epoch设置为150, 初始学习率设置为0.001, 采用学习率调整优化器加速模型收敛, 稳定周期设置为10, 衰减因子设置为0.1, 即当模型在10个训练周期内损失函数不再减小, 学习率降低为0.0001.

4.3 模型训练结果

仅使用图像模态数据, MetaData-EfficientNet-B4模型在白天、夜间和雾天场景测试集上平均准确率分别达到95.8%, 90.7%和91.7%. 而融合多模态数据后, 模型在白天、夜间和雾天条件下的平均准确率分别达到了96.8%, 96.2%和95.8%. 结果表明, 多模态融合能显著提高模型的识别准确性, 尤其是大幅提升模型在复杂场景下的表现, 所提出的MetaData模块在多模态数据融合任务中表现稳定, 模型利用互补信息有效提高了识别精度. 模型训练的迭代曲线如图11所示.

对模型混淆矩阵的结果进行分析. 如图12所示, 模型在区分“合格”与“不合格”状态方面表现出良好的辨别能力, 未出现将“合格”误判为“不合格”或反向误判的情况. 分类错误主要集中在复杂场景下的“过渡态”类别, 统计结果发现, 本文融合方法相比传统不融合方法, 夜晚场景的过渡态误判率降低了8个百分点, 雾天场景的误判率降低了4个百分点然而需指出的是, 过渡态本身就具有一定模糊性, 其表观图像特征与合格态较为相似, 尤其在气泡基本逸出, 混凝土密实性已明显提高的情况下, 从外观上较难进行准确区分, 少量误判为过渡态的样本对大坝结构安全性影响较小, 属于可接受范围.

4.4 消融实验

为验证多模态融合在综合利用图像与参数信息判断混凝土振捣质量方面的有效性, 平行选择ResNet-50[37]、 MobileNet-V2[38]、DenseNet-121[39]作为主干网络训练模型, 分析结果. 同时, 设计不融合参数信息, 线性的Conca融合方法, 非线性的MetaBlock融合方法做平行试验, 训练模型得到的结果如表3~5所示.

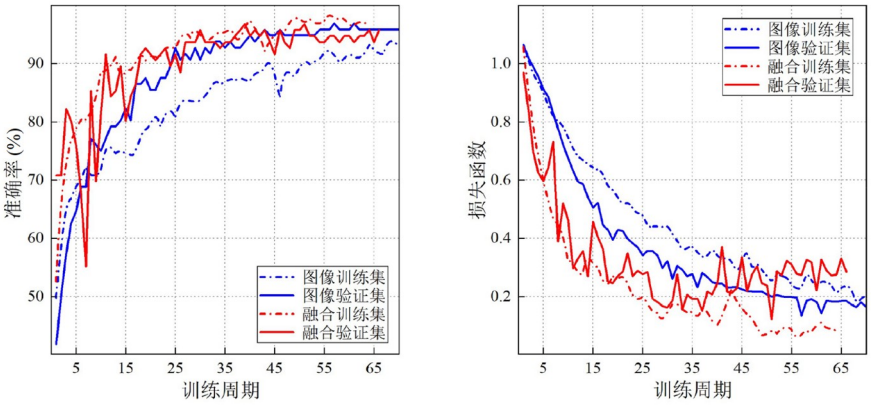
(1) 一致性分析

对上述试验结果进行分析, 可以发现在未融合多模态数据的训练策略下, 各模型在白天场景下的识别

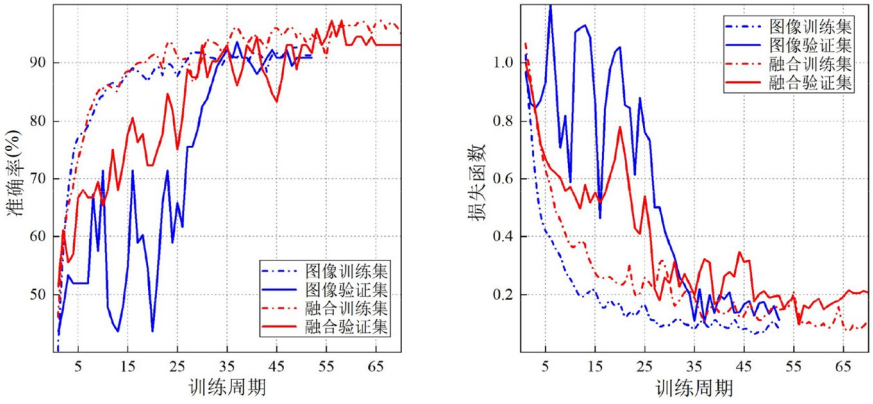
表 2 图像质量评价[a)]
Table 2 Image quality assessment

指标	Night (a)	Foggy (a)	Night (b)	Foggy (b)	Night (c)	Foggy (c)
PSNR	9.543	21.195	9.240	23.381	9.621	23.483
SSIM	0.26	0.739	0.566	0.799	0.582	0.823

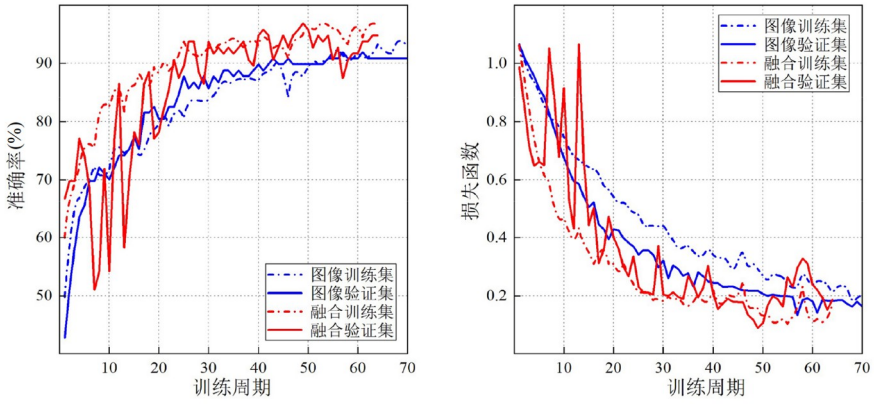
a) a, b, c对应图10中的a, b, c.



(a)白天场景下模型训练迭代曲线



(b)夜晚场景下模型训练迭代曲线



(c)雾天场景下模型训练迭代曲线

图 11　(网络版彩图)复杂场景下模型迭代曲线

Figure 11　(Color online) Model iteration curve in complex scenes.

准确率整体普遍高于夜晚与雾天场景, 平均高出4~5个百分点，这一现象表明仅依赖单一图像模态训练的模型在低照度或图像质量下降的复杂场景下识别性能不足，所训练模型难以适应全天候的振捣质量识别精度需求，证明了本研究在复杂场景下提升识别能力的现实意义与必要性．此外，采用本文提出的MetaData模

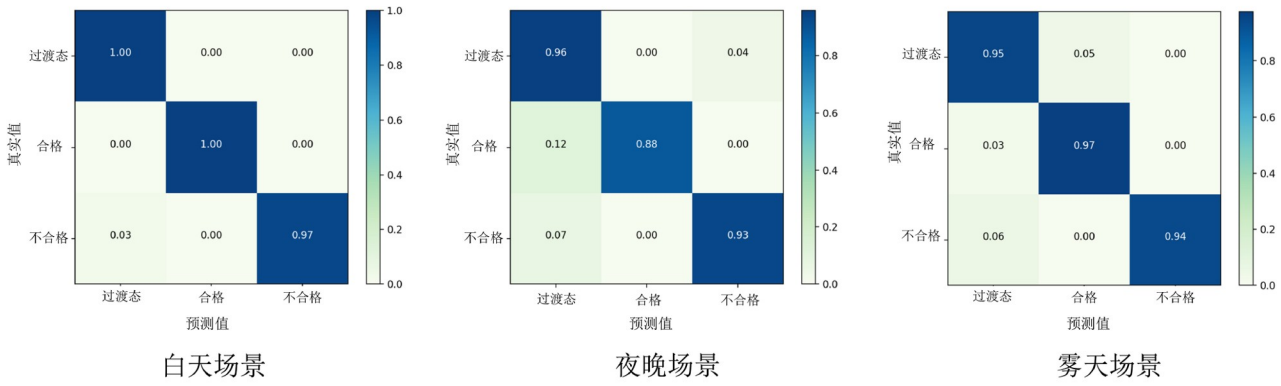


图 12 (网络版彩图)混淆矩阵

Figure 12 (Color online) Confusion matrix.

表 3 白天场景下试验结果比较

Table 3 Comparison results of daytime working conditions

模型	融合方法	Precision	Recall	F1	Accuracy
ResNet-50	None	0.950	0.947	0.948	0.948
	Concat	0.923	0.913	0.918	0.927
	MetaBlock	0.943	0.937	0.940	0.938
	MetaData	0.967	0.950	0.958	0.958
MobileNet-V2	None	0.950	0.953	0.951	0.947
	Concat	0.933	0.940	0.936	0.938
	MetaBlock	0.967	0.963	0.965	0.969
	MetaData	0.966	0.957	0.961	0.958
DenseNet-121	None	0.963	0.962	0.962	0.957
	Concat	0.867	0.837	0.852	0.865
	MetaBlock	0.960	0.957	0.958	0.958
	MetaData	0.973	0.963	0.968	0.969
EfficientNet-B4	None	0.965	0.957	0.961	0.958
	Concat	0.913	0.910	0.911	0.917
	MetaBlock	0.943	0.947	0.945	0.948
(试验方法)	MetaData	0.963	0.970	0.966	0.965

块融合多模态数据后，模型在各类场景下的识别准确率均有显著提升，尤其在夜晚与雾天环境下提升更为明显，充分验证了多模态融合策略在增强模型鲁棒性与泛化能力方面的有效性.

(2) 差异性分析

分析表3~5四项评价指标的结果可知，采用Concat拼接方法的模型性能波动较大，尤其在夜晚和雾天等复杂场景下，F1值常低于0.7，准确率常停留在80%左右. 其主要原因是该方法仅对多模态特征进行简单堆叠，未显式建模模态之间的关联性. 相比之下，MetaBlock方法借鉴LSTM架构的思想，通过非线性方式对图像特征与参数特征进行映射与重构，在白天场景下，MetaBlock模块取得了与MetaData模块相近甚至更优的性能, 例如, 主干网络为MobileNet-V2时, MetaBlock的F1值达0.965，略高于MetaData模块的0.961，准确率也达到0.969, 超过MetaData的0.958. 同时, 在夜晚与雾天等低质量图像场景中, MetaBlock的融合效果虽未超过MetaData模块, 但同样表现出较明显的性能提高, 表

表 4　夜晚场景下试验结果比较

Table 4　Comparison results of night working conditions

模型	融合方法	Precision	Recall	F1	Accuracy
ResNet-50	None	0.903	0.907	0.905	0.904
	Concat	0.823	0.840	0.831	0.833
	MetaBlock	0.907	0.910	0.908	0.889
	MetaData	0.940	0.953	0.946	0.944
MobileNet-V2	None	0.927	0.920	0.923	0.916
	Concat	0.846	0.847	0.846	0.847
	MetaBlock	0.943	0.943	0.943	0.931
	MetaData	0.953	0.957	0.955	0.942
DenseNet-121	None	0.917	0.903	0.910	0.903
	Concat	0.877	0.887	0.882	0.861
	MetaBlock	0.913	0.920	0.916	0.903
	MetaData	0.963	0.967	0.965	0.958
EfficientNet-B4	None	0.906	0.910	0.908	0.907
	Concat	0.703	0.650	0.675	0.639
	MetaBlock	0.913	0.917	0.915	0.903
(试验方法)	MetaData	0.963	0.970	0.966	0.962

表 5　雾天场景下试验结果比较

Table 5　Comparison results of foggy working conditions

模型	融合方法	Precision	Recall	F1	Accuracy
ResNet-50	None	0.900	0.921	0.910	0.906
	Concat	0.753	0.730	0.741	0.729
	MetaBlock	0.923	0.917	0.920	0.927
	MetaData	0.967	0.940	0.953	0.958
MobileNet-V2	None	0.890	0.891	0.890	0.896
	Concat	0.810	0.817	0.813	0.802
	MetaBlock	0.927	0.926	0.926	0.928
	MetaData	0.967	0.964	0.965	0.969
DenseNet-121	None	0.907	0.906	0.906	0.918
	Concat	0.733	0.723	0.728	0.740
	MetaBlock	0.933	0.947	0.940	0.938
	MetaData	0.961	0.956	0.958	0.960
EfficientNet-B4	None	0.905	0.900	0.902	0.917
	Concat	0.780	0.773	0.776	0.802
	MetaBlock	0.943	0.947	0.945	0.949
(试验方法)	MetaData	0.960	0.951	0.955	0.958

明其在一定程度上提升了模型对复杂场景的适应性.

本文采用TOPSIS方法[40]，基于表3~5中四个评价指标的结果，对不同融合方法进行综合评分. 结果显示，本文提出的融合策略在大多数场景下表现最优，

以EfficientNet-B4为主干网络, 结合MetaData融合模块的训练策略, 在白天, 夜晚及雾天等场景下均获得了最高的综合得分, 其融合效果稳定, 鲁棒性强. 而MetaBlock模块在与MobileNet-V2网络搭配时表现良好, 尤其在白天场景下取得了较高得分, 显示出一定的适配性与应用潜力. 相比之下, Concat方法在所有场景中的表现均处于劣势, 得分显著偏低. 这表明简单的特征拼接方式未能有效捕捉模态间的关联信息, 甚至干扰了模型对数据特征的正确学习, 影响最终识别性能. 结果如图13所示, 进一步验证了融合方式选择对模型性能的显著影响, 也从侧面证明了MetaData模块在多模态数据融合任务中的有效性和优越性.

4.5 系统开发

为了便于将研究成果应用于现场施工, 我们基于B/S架构开发了简单易操作的振捣质量识别系统. 系统支持不同的施工场景选择, 背后关联预训练过表现最好的模型. 用户可上传当前施工振捣图像和数据库中对应的施工参数和混凝土材料参数, 系统快速返回振捣质量识别结果和可信度. 此外, 随着振捣施工进行, 系统自动计算该轮次振捣合格率, 用于辅助仓面质量验收. 对白天场景数据进行测试, 结果如图14所示.

5 结论

针对传统振捣质量监测方式存在人为主观性强, 劳动强度大的缺点, 计算机视觉技术数据利用深度不足, 模型适用施工场景有限的缺陷. 本文开展基于多模态感知的复杂场景下振捣质量识别研究, 主要结论如下.

(1) 基于物联网(IoT)、全球导航卫星系统(GNSS)和无线传输技术, 开发了混凝土质量监测系统, 采集振捣施工参数、混凝土参数和振捣表面图像数据, 利用

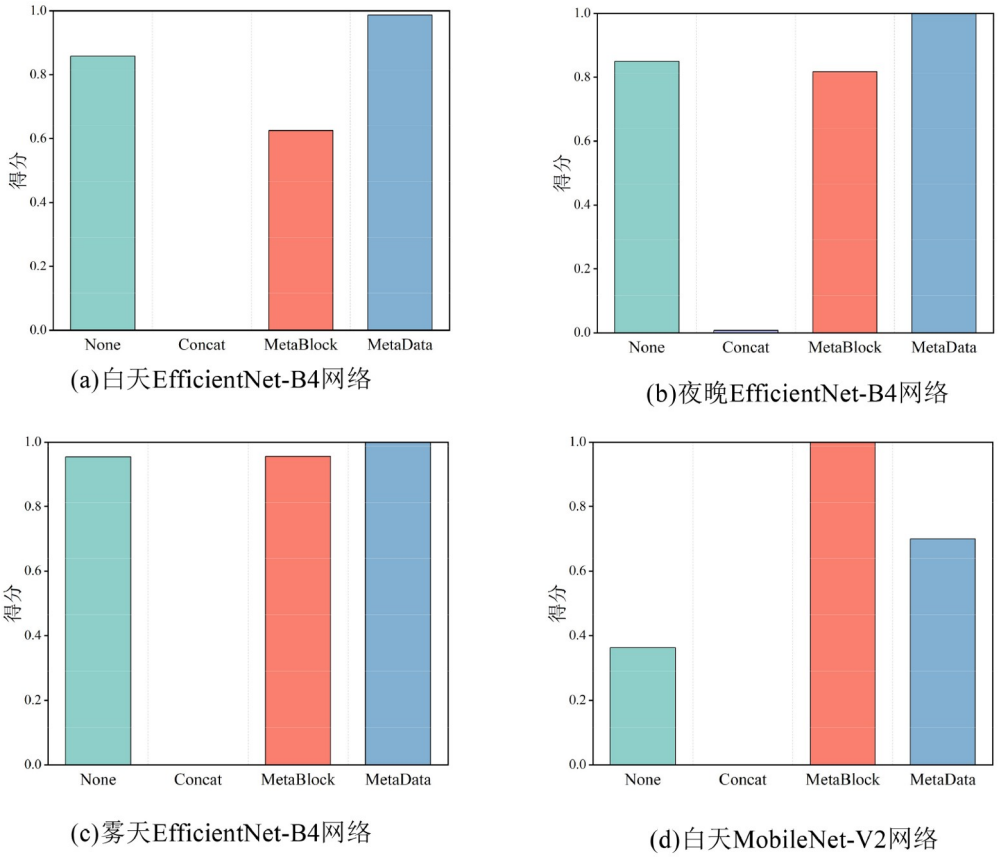


图 13 (网络版彩图)TOPSIS方法分析

Figure 13 (Color online) Experimental results of the TOPSIS analysis method.

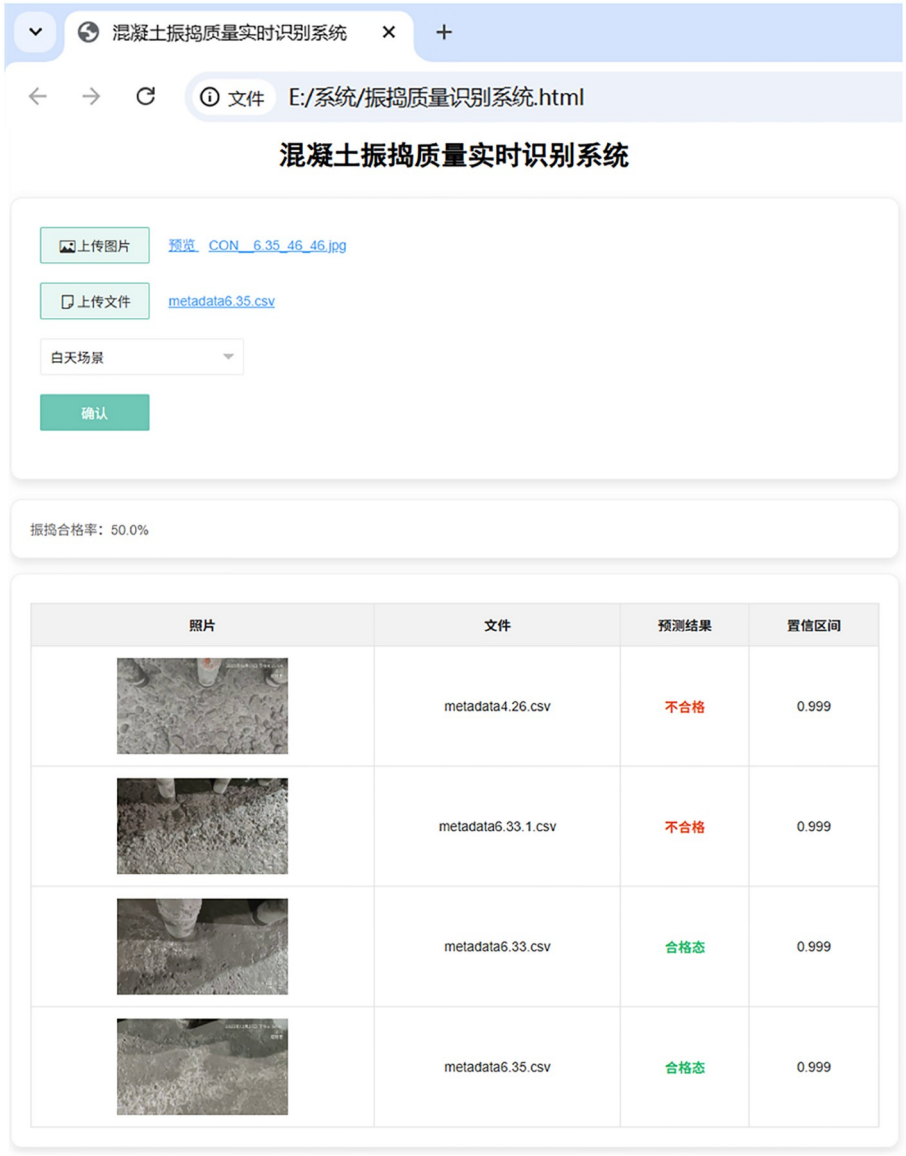


图 14　(网络版彩图)白天场景系统应用

Figure 14　(Color online) Daytime scene system application.

图像增强技术构建多场景下的振捣多模态数据集.

(2) 提出了一个深度学习模型MetaData-Efficient-Net-B4, 用于整合多模态信息. 该模型使用Efficient-Net-B4主干网络, 结合创新提出的MetaData模块, 有效完成多模态数据融合任务, 大幅提高模型在图像识别领域的识别准确率和稳定性.

(3) 将所提模型应用于我国东庄水利枢纽工程. 模型精度满足现场全天候施工需求, 在复杂场景下保持较高准确率, MetaData融合策略与其他融合方法比较取得明显优势, 此外, 还基于B/S架构开发了简单易操作的振捣质量识别系统, 方便现场人员使用, 支持仓面验收工作.

参考文献

1　Liu Y. Measurement, modeling and application of rheology of freshly mixed concrete (in Chinese). Dissertation for the Doctoral Degree. Beijing:

China General Research Institute of Building Materials Science, 2020 [刘豫. 新拌混凝土流变的测量, 模型及其应用. 博士学位论文. 北京: 中国建筑材料科学研究总院, 2020]
2 Wen J X, Huang F L, Wang Z, et al. Research status and development trend of concrete vibration technology (in Chinese). Bull Chin Ceram Soc, 2021, 40: 3326–3336 [温家馨, 黄法礼, 王振, 等. 混凝土振捣技术研究现状与发展趋势. 硅酸盐通报, 2021, 40: 3326–3336]
3 Zhao X, Huang Y, Dong W, et al. A review of compaction mechanisms, influencing factors, and advanced methods in concrete vibration technology. J Build Eng, 2024, 93: 109847
4 Bian C, Tian Z H, Liu Z M, et al. Experimental study on the characterization of vibratory compactness evaluation indexes for fresh concrete (in Chinese). J Hydroelectr Eng, 2020, 39: 67–80 [边策, 田正宏, 刘志明, 等. 新拌混凝土振动密实评价指标表征性试验研究. 水力发电学报, 2020, 39: 67–80]
5 Xu R L, Tian D, Shao B, et al. Automatic identification of construction safety hazards in hydropower projects under dual-attention mechanism (in Chinese). J Hydroelectr Eng, 2025, 44: 119–128 [许仁乐, 田丹, 邵波, 等. 双重注意力机制下的水电工程施工安全隐患自动辨识方法. 水力发电学报, 2025, 44: 119–128
6 Shen Y, Tian D, Liu H, et al. Intelligent construction of knowledge graph for concrete dam construction management (in Chinese). J Hydroelectr Eng, 2022, 41: 118–128 [沈扬, 田丹, 刘昊, 等. 面向混凝土坝施工管理的知识图谱智能构建. 水力发电学报, 2022, 41: 118–128]
7 Zheng R T, Chen Z T, Guan X H, et al. Deep learning-driven foreign object monitoring and localization technology for river inspection by unmanned aerial vehicles (in Chinese). J Hydroelectr Eng, 2025, 44: 87–98 [郑荣添, 陈泽同, 管孝汉, 等. 深度学习驱动的无人机河道巡检异物监测及定位技术研究. 水力发电学报, 2025, 44: 87–98]
8 Gong J, Yu Y, Krishnamoorthy R, et al. Real-time tracking of concrete vibration effort for intelligent concrete consolidation. Automat Constr, 2015, 54: 12–24
9 Tian Z, Bian C. Visual monitoring method on fresh concrete vibration. KSCE J Civ Eng, 2014, 18: 398–408
10 Zhong D H, Shen Z Y, Wang J J, et al. Research on dynamic evaluation of concrete dam vibration construction quality based on real-time monitoring (in Chinese). J Hydraul Eng, 2018, 49: 775–786 [钟登华, 沈子洋, 王佳俊, 等. 基于实时监控的混凝土坝振捣施工质量动态评价研究. 水利学报, 2018, 49: 775–786]
11 Liu Y J. Development of concrete vibration quality monitoring system based on stereo vision (in Chinese). Dissertation for the Master's Degree. Harbin: Harbin Institute of Technology, 2018 [刘亚洁. 基于立体视觉的混凝土振捣质量监测系统的开发. 硕士学位论文. 哈尔滨: 哈尔滨工业大学, 2018]
12 Lee S, Skibniewski M J. Automated monitoring and warning solution for concrete placement and vibration workmanship quality issues. AI Civ Eng, 2022, 1: 4
13 Liu H, Liu C, Fu J, et al. Intelligent analysis of hydraulic concrete vibration time based on convolutional neural network. Adv Civil Eng, 2023, 2023: 1–11
14 Wang D, Ren B, Cui B, et al. Real-time monitoring for vibration quality of fresh concrete using convolutional neural networks and IoT technology. Automat Constr, 2021, 123: 103510
15 Wang D. Research on intelligent vibration of concrete in high arch dam driven by air and space sensing data (in Chinese). Dissertation for the Doctoral Degree. Tianjin: Tianjin University, 2022 [王栋. 空天地感知数据驱动下高拱坝混凝土智能振捣研究. 博士学位论文. 天津: 天津大学, 2022]
16 Zhang Z H, Sun J, Han K, et al. Study on geomorphologic evolution of braided river flow with multi-source data fusion (in Chinese). J Hydroelectr Eng, 2025, 44: 1–11 [张子浩, 孙健, 韩凯, 等. 多源数据融合的辫状河流地貌演变研究. 水力发电学报, 2025, 44: 1–11]
17 Pan B, Hirota K, Jia Z, et al. A review of multimodal emotion recognition from datasets, preprocessing, features, and fusion methods. Neurocomputing, 2023, 561: 126866
18 Wang Z, Luo S, Chen J, et al. Multi-modality deep learning model reaches high prediction accuracy in the diagnosis of ovarian cancer. iScience, 2024, 27: 109403
19 Zhang H C, Li L X, Liu D J. A review of multimodal data fusion research (in Chinese). J Front Comput Sci Technol, 2024, 18: 2501–2520 [张虎成, 李雷孝, 刘东江. 多模态数据融合研究综述. 计算机科学与探索, 2024, 18: 2501–2520]
20 Zhang J, Wang J G, Yu J, et al. A deep learning model for parameter integration in high rockfill dam construction simulation driven by multimodal perception (in Chinese). J Hydraul Eng, 2022, 53: 1049–1063, 1072 [张君, 王金国, 余佳, 等. 多模态感知驱动下高堆石坝施工仿真参数集成深度学习模型. 水利学报, 2022, 53: 1049–1063, 1072]

21 Jiang Z Y, Zhang Z X, Liu J Y, et al. Wide-angle image generation by multimodal data feature fusion (in Chinese). J Image Graph, 2025, 30: 173–187 [姜智颖, 张曾翕, 刘晋源, 等. 多模态数据特征融合的广角图像生成. 中国图象图形学报, 2025, 30: 173–187]

22 Li W, Zhuang J, Wang R, et al. Fusing metadata and dermoscopy images for skin disease diagnosis. In: Proceedings of the 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). Iowa: IEEE, 2020. 1996–2000

23 Yuan Q, Wang J, Wu B, et al. Deep multimodal fusion model for moisture content measurement of sand gravel using images, NIR spectra, and dielectric data. Measurement, 2024, 227: 114270

24 Pacheco A G C, Krohling R A. An attention-based mechanism to combine images and MetaData in deep learning models applied to skin cancer classification. IEEE J Biomed Health Inform, 2021, 25: 3554–3563

25 Chen H, Wang S, Xu J C, et al. Underwater image enhancement based on multi-scale feature fusion generative adversarial network (in Chinese). Comput Eng Appl, 2023, 59: 231–241 [陈辉, 王硕, 许家昌, 等. 基于多尺度特征融合生成对抗网络的水下图像增强. 计算机工程与应用, 2023, 59: 231–241]

26 Dan D, Ge L, Yan X. Identification of moving loads based on the information fusion of weigh-in-motion system and multiple camera machine vision. Measurement, 2019, 144: 155–166

27 Shi M, Li H, Yao Q, et al. Vision based nighttime pavement cracks pixel level detection by integrating infrared visible fusion and deep learning. Constr Build Mater, 2024, 442: 137662

28 Zhang X, He L, Chen J, et al. Multiattention mechanism 3D object detection algorithm based on RGB and LiDAR fusion for intelligent driving. Sensors, 2023, 23: 8732

29 Zhao K, Zhao L, Hou J, et al. Effect of vibratory mixing on the slump, compressive strength, and density of concrete with the different mix proportions. J Mater Res Tech, 2021, 15: 4208–4219

30 Li P D, Zhao Y, Wu Y F, et al. Effect of defects in adhesive layer on the interfacial bond behaviors of externally bonded CFRP-to-concrete joints. Eng Struct, 2023, 278: 115495

31 Li Q, Yang K, Yang C. An alternative admixture to reduce sorptivity of alkali-activated slag cement by optimising pore structure and introducing hydrophobic film. Cement Concrete Compos, 2019, 95: 183–192

32 Li M, Liu J, Yang W, et al. Structure-revealing low-light image enhancement via robust retinex model. IEEE Trans Image Process, 2018, 27: 2828–2841

33 Ancuti C O, Ancuti C, Timofte R. NH-HAZE: An image dehazing benchmark with non-homogeneous hazy and haze-free images. In: Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle: IEEE, 2020. 1798–1805

34 Koschmieder H. Theorie der horizontalen Sichtweite. Beiträge Phys Atmos, 1924, 12: 33–53

35 Tan M, Le Q V. EfficientNet: Rethinking model scaling for convolutional neural networks. arXiv: 1905.11946

36 Li X, Wang W, Hu X, et al. Selective kernel networks. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019. 510–519

37 He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016. 770–778

38 Sandler M , Howard A, Zhu M, et al. MobileNetV2: Inverted residuals and linear bottlenecks. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018

39 Huang G, Liu Z, Laurens V D M, et al. Densely connected convolutional networks. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2016

40 Yu D, Pan T. Tracing knowledge diffusion of TOPSIS: A historical perspective from citation network. Expert Syst Appl, 2021, 168: 114238

Research on vibration quality recognition method based on multi-modal sensing in complex scenes

REN BingYu[1], ZHANG ChengXuan[1], YANG ZhiHua[1], GUAN Tao[1] & HE Wei[2]

[1] *State Key Laboratory of Hydraulic Engineering Intelligent Construction and Operation, Tianjin University, Tianjin 300072, China*
[2] *China Renewable Energy Engineering Institute, Beijing 100011, China*

Vibration quality is crucial for the strength and durability of concrete. On-site vibration quality identification typically relies on manual supervision, which suffers from strong subjectivity and high labor intensity. While computer vision-based methods have improved objectivity and automation, they exhibit limitations such as insufficient data utilization depth and an inability to meet all-weather construction demands, with significantly degraded performance in complex scenarios (e.g., nighttime, foggy conditions). To address these issues, this study proposes a multi-modal recognition method that integrates concrete surface image data, vibration construction parameters (duration, depth), and material properties (slump, air content, etc.). First, attitude sensors, laser sensors, GNSS, and industrial cameras are integrated onto the vibration trolley to monitor and acquire a multi-modal dataset, while image enhancement techniques simulate concrete images under complex conditions. Second, the MetaData-EfficientNet-B4 deep learning model is proposed, where EfficientNet-B4 extracts image features, and the novel MetaData module enables deep fusion of image and parameter features. Finally, a user-friendly B/S-based vibration quality recognition system is developed to assist on-site construction and placement section acceptance. Engineering applications demonstrate that the proposed multi-modal model improves recognition accuracy (compared to traditional methods) from 95.8%, 90.7%, and 91.7% (daytime, nighttime, foggy conditions) to 96.8%, 96.2%, and 95.8%, respectively, enhancing performance in complex scenarios. The MetaData fusion approach outperforms other fusion methods in TOPSIS evaluation.

concrete dam, computer vision, multi-modal fusion, MetaData-EfficientNet-B4, vibration quality recognition

doi: 10.1360/SST-2025-0415