

基于计算机视觉的混凝土高坝裂缝智能检测方法研究

叶飞¹, 王赞¹, 汪诚¹, 钟洪星¹, 施炎², 包腾飞², 李扬涛²

(1. 三峡金沙江云川水电开发有限公司禄劝乌东德电厂, 云南 昆明 651512; 2. 长江三峡勘测研究院有限公司 (武汉), 湖北 武汉 430074)

摘要:【目的】针对混凝土高坝无人机航拍图像中裂缝目标细长、尺度变化大、背景干扰强, 以及人工巡检效率低、主观性强等问题, 构建一种基于计算机视觉的混凝土高坝裂缝智能检测方法, 为大坝安全巡检与病害识别提供技术支撑。【方法】以无人机航拍获取的混凝土高坝裂缝图像为研究对象, 在 YOLOv5 目标检测模型基础上进行结构改进。采用 Swin Transformer 替换原有跨阶段局部模块, 以增强模型的全局上下文建模能力; 增设 4 倍下采样检测头, 以提升细微裂缝和多尺度目标检测能力; 嵌入卷积块注意力模块, 以强化关键裂缝特征并抑制复杂背景干扰。【结果】改进 YOLOv5 模型训练过程收敛稳定, 在混凝土高坝裂缝检测任务中取得了较优性能, 平均精度 mAP_{0.5}、mAP_{0.5:0.95} 和帧率分别达到 93.29%、74.39% 和 34.3。与经典目标检测模型相比, 所提模型在检测精度和实时性方面均表现更优, 能够有效减少细微裂缝漏检和复杂背景误检现象。【结论】提出的改进 YOLOv5 模型能够提升无人机航拍场景下混凝土高坝裂缝的特征表征能力、多尺度检测能力和背景抗干扰能力, 兼具较好的检测精度、实时性和工程应用潜力, 可为混凝土高坝智能巡检与安全评估提供参考。

关键词: 混凝土坝; 深度学习; 机器视觉; 缺陷检测

中图分类号: TV33; TP391.41 **文献标志码:** A

大坝在保障能源安全与水资源合理调配方面发挥不可替代的基础性作用, 是支撑国民经济稳健发展的重要工程设施, 在防洪、灌溉和水力发电等领域具有重要意义^[1]。我国目前约有 9.8 万座大坝正在服役, 其中 200m 以上大坝中混凝土坝约占 56%^[2]。混凝土坝在长期运行过程中不可避免地会产生累积性损伤, 在荷载等因素作用下, 局部损伤可能逐步演化为整体性破坏, 存在较大安全隐患, 溃坝事故将造成难以估量的损失^[3-5]。因此, 构建科学完善的大坝安全诊断体系, 对保障混凝土坝长期安全运行具有重要意义。

混凝土高坝表观缺陷类型较为复杂, 主要包括裂缝、剥蚀、表层脱落、析出物及渗痕等。相较于其他类型的表观病害, 裂缝不仅发生频率较高、分布范围较广, 而且与坝体应力重分布、材料性能劣化以及局部损伤演化过程密切相关, 因此成为高坝安全巡检中最受关注的关键缺陷类型之一。尤其在无人机航拍条件下, 裂缝通常表现出细长形态、低对比度、小目标及多尺度并存等特征, 其识别难度明显高于一般表观缺陷。基于此, 本文将研究重点聚焦于混凝土高坝裂缝识别问题, 并围绕细微裂缝的检测需求, 在模型结构设计、数据处理及实验分析等方面开展系统研究。目前混凝土高坝表观缺陷检测仍以人工定期巡检为主, 该方式依赖人员经验, 主观性强、效率低、成本

高, 且巡检频率有限, 难以及时发现关键缺陷^[6,7]。随着无人机等无人智能设备的发展, 利用搭载高清摄像头的无人系统对混凝土高坝进行缺陷图像采集逐渐成为大坝安全巡检的重要手段, 该方法具有安全性高、覆盖范围广等优势^[8]。然而, 受混凝土高坝结构复杂、尺度跨度大的影响, 航拍图像数据量庞大, 人工分析难以满足工程应用需求。

近年来, 人工智能技术的发展为结构缺陷自动识别提供了新途径。现有缺陷图像检测方法主要包括基于传统机器学习和基于深度学习的目标检测方法。传统机器学习方法通过提取图像颜色、纹理等人工特征并结合分类器实现缺陷检测, 代表性算法包括支持向量机和随机森林。杨扬等^[9]基于颜色与纹理特征并采用支持向量机构建了裂缝识别模型; Shi Yong 等^[10]利用随机森林算法提高了裂缝分布不均条件下的检测稳定性; 靳华中等^[11]结合梯度直方图与分水岭算法实现了网状裂缝的多类别检测; 王睿等^[12]基于裂缝几何特征, 结合径向基函数与支持向量机构建了裂缝自动判别模型。该类方法依赖人工特征提取, 在复杂背景条件下的适应性和鲁棒性仍有限。

深度学习方法能够自动学习缺陷特征, 在检测精度和效率方面具有明显优势。孟诗乔等^[13]基于改进卷积神经网络构建了裂缝识别与分割模型; 杨杰文等

收稿日期/Received: 2026-01-27

修回日期/Revised: 2026-07-07

基金项目: 三峡金沙江云川水电开发有限公司禄劝乌东德电厂项目 (Z522402021); 国家自然科学基金 (52508277)。

第一作者: 叶飞 (1995—), 男, 工程师, 硕士, 主要从事水工建筑物维护、大坝安全监测方面的研究。E-mail: ye_fei@ctg.com.cn。

通信作者: 包腾飞 (1974—), 男, 正高级工程师, 博士, 主要从事大坝安全监控方面的研究。E-mail: baotf@hhu.edu.cn。

[14]将 U-Net 与 Haar-like 特征相结合,实现了复杂背景下的裂缝检测;李松等[15]提出了一种基于改进 YOLOv8 的轻量化裂缝检测模型;吴刚等[16]融合 K-means 聚类与 YOLOv4 改进算法,实现了隧道衬砌损伤检测;王燕华等[17]通过调整多尺度先验框设置,提高了复杂噪声背景下的裂缝识别精度。尽管如此,针对混凝土高坝航拍图像背景复杂、小尺度裂缝特征显著等特点,现有模型仍易产生漏检和误检现象,且卷积网络受限于局部感受野,难以充分建模全局上下文信息。与一般混凝土坝相比,混凝土高坝在巡检过程中面临更显著的高差条件、更大的观测距离以及更复杂的光照和视角变化。受此影响,航拍图像中的裂缝往往呈现尺度更小、背景更复杂、纹理干扰更强等特征,给自动检测带来更大挑战。因此,面向混凝土高坝场景开展裂缝智能检测方法研究会面临更高的技术难度与更复杂的工程挑战。

针对上述问题,本文提出一种基于深度学习的无人机图像裂缝识别方法。以 YOLOv5 目标检测模型为基础,引入 Swin Transformer 网络以增强全局特征建模能力,并增加小目标预测层以提升细微裂缝的检测性能;同时在特征融合阶段引入卷积块注意力模块(Convolutional Block Attention Module, CBAM),提高模型在复杂背景和多尺度条件下的预测能力。

1 YOLOv5 目标检测模型

YOLOv5 是 Glenn Jocher 等人于 2020 年发布的一种单阶段目标检测模型,该模型在 YOLOv4 的基础上进行了进一步的优化,对输入端、主干网络、特征融合网络和检测头部结构均做了一定的改进,具有特征提取能力强、检测速度快等优势。混凝土高坝裂缝缺陷航拍图像数据集具有数量多、特征细微等特点,故将 YOLOv5 作为主干网络引入到混凝土高坝航拍图像缺陷的快速检测中。

1.1 输入端

输入端表示对输入图像的预处理,包括将输入图像缩放到网络输入尺寸、归一化等步骤。在 YOLOv5 网络训练阶段中,提出了使用 Mosaic 数据增强方法、自适应锚框计算和自适应图片缩放方法。该方法指的是将随机的 4 张照片通过随机裁剪、随机缩放、随机排布的方式拼接成为 1 张新的图片,该方法可以有效丰富数据集数量,为增强网络的鲁棒性和提高小目标的检测性能提供了帮助。此外,由于对 4 张图片进行了新的拼接,使得模型网络可以直接计算 4 张图片的数据,大大减少了显存的占用。

此外 YOLOv5 模型将自适应锚框计算嵌入到了模型之中,其原理是网络在初始锚框的设定基础上输出预测框,在与数据集中真实锚框比对后进行反向更

新,迭代网络参数。优势在于减少了误检和漏检的情况,有助于提高模型泛化能力。自适应图片缩放方法是对原始图像中填充最少的像素将其缩放为标准尺寸,减少冗余,提高推理速度。

1.2 主干网络

主干网络是用以提取输入图像特征的结构,在 YOLOv5 中,使用了 Focus 结构及跨阶段局部结构(Cross Stage Partial, CSP)。

1) Focus 结构。Focus 结构是输入图像进入主干网络时,对图像进行切片操作的结构。该结构将高分辨率图像切分为多个低分辨率图像,具体操作为在 1 张图片中每隔一个像素提取一个值形成 4 张照片,采样结果在通道维度进行融合,即将通道扩充为原来的 4 倍。以 YOLOv5 为例,将空间尺寸为 608×608 像素、通道数为 3 的原始图像输入网络,经过切片操作后形成空间尺寸为 304×304 像素、通道数为 12 个的特征图该方法在没有信息丢失的情况下实现了二倍下采样特征图。

2) CSP 结构。YOLOv5 中使用了 CSP 结构,用以优化网络中梯度重复计算导致的计算量过大问题,其设计了两种 CSP 结构,分别是应用于主干网络的 CSP1_X 模块和应用于特征融合网络中的 CSP2_X 模块。CSP1_X 模块的原理是将原输入的特征图按照通道拆分为两个部分,其一通过基本卷积单元模块和残差网络构建残差组件并进行卷积操作,其二直接进行卷积操作,随后将两个部分进行合并形成新的特征图。

1.3 特征融合网络

输入图像在主干网络中进行特征提取时,图像会产生局部信息丢失的问题,利用特征融合结构可以从主干网络中提取不同尺度的特征图并进行融合,丰富图像的特征信息。

YOLOv5 的特征融合结构采用了特征金字塔的思想,其是一种多尺度目标检测技术,通过添加不同尺度的特征层来实现。在 YOLOv5 网络中,采用的是结合特征金字塔(Feature Pyramid Network, FPN)和路径聚合网络(Path Aggregation Network, PAN)的特征融合结构,FPN 结构自上而下传递高层信息,丰富了特征信息,而 PAN 结构自下而上将浅层信息传递到高层,补充了定位信息。此外,如上文所述,YOLOv5 设计了两种 CSP 结构,其中 CSP2_X 在特征融合结构中用卷积层替代残差结构,将输入的特征图分为两个部分分别处理,能保留更多图像特征信息。

1.4 检测头部结构

目标检测头部结构是对特征融合结构输出的特征图进行目标预测和检测框回归计算的部分,主要包括卷积层、池化层和全连接层等。在 YOLOv5 网络

中,检测头部结构主要包括了锚框、卷积层、预测层和非极大值抑制结构。

锚框是一组预设的边界框,通过提前定义尺寸的锚框在特征图上生成候选框;卷积层通过调整核大小与通道数来应对目标检测任务,处理特征图并提取目标特征;预测层由卷积层和输出层构成,每个预测层对应一组边界框和目标类别;非极大值抑制用于筛选概率最高的预测框,在计算中,会生成大量的预测框,而其中绝大部分均是无效的,YOLOv5 采用了非极大值抑制方法,通过分类器对产生的预测框进行排序,利用交并比指标进行筛选迭代,最终保留概率最大的预测框实现目标的精确检测。下面介绍 YOLOv5 中使用的损失函数和非极大值抑制方法。

1) 损失函数。在目标检测中,需要对预测边框与标定边框进行比对与评判。传统的边框回归损失往往只关注对应参数之间的距离,当预测框与标注框不相交时,交并比 (Intersection over Union, IoU) 为 0,难以为模型优化提供有效梯度。为克服这一局限,YOLOv5 在边框回归中引入了广义交并比损失 (Generalized Intersection over Union, GIoU),使模型在预测框与真实框无重叠的情况下仍能够获得合理的回归约束。

GIoU 的原理是在损失函数中添加了标定框 A 和预测框 B 所组成的闭环框 C,其惩罚项是闭环框减去两个框的并集后的面积在闭环框中的比例越小越好,计算方法如式 (1) 和式 (2) 所示。

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

$$GIoU = IoU - \frac{|C / (A \cup B)|}{|C|} \quad (2)$$

式中, A 表示标注框, B 表示预测框, C 表示同时包围 A 和 B 的最小闭包矩形; $|A \cap B|$ 表示预测框与标注框的交集面积, $|A \cup B|$ 表示二者并集面积, $|C - (A \cup B)|$

表示闭包区域中除去并集后的面积。GIoU 越大,说明预测框与真实框的重合程度越高。

2) 非极大值抑制。非极大值抑制 (Non-Maximum Suppression, NMS) 的核心思想是搜索局部最大值,抑制非极大值。在目标检测任务中,目标位置一般会生成大量的候选框,此时就需要找到最佳的预测框并消除多余的候选框。NMS 算法根据置信度得分对预测候选框进行排序,将得分最高的候选框添加到输出列表,然后将其与其余的候选框比对,若重叠面积大于设定阈值,则删除该候选框,重复该操作直至候选框列表清空,从而实现检测任务。

2 基于改进 YOLOv5 的混凝土高坝裂缝检测模型

混凝土高坝的无人机航拍图像因视角与距离多变,同时存在大尺度与细微裂缝,且背景复杂,干扰严重。YOLOv5 目标检测模型在物体检测的一般应用场景下具有良好的识别效果,但在精细小目标、细微特征的检测中存在一定的不足。在利用卷积神经网络进行图像特征提取时,主干网络通过卷积层进行特征提取操作,卷积核将遍历图像的所有像素点。小尺度目标的像素占比远低于大尺度目标在图像中的像素占比,这导致卷积层更容易学习到大尺度目标的特征而对小目标造成漏检和误检,严重降低目标检测精度。

针对 YOLOv5 的上述问题,为提高模型小尺度目标检测性能,更加准确地对混凝土高坝航空图像进行裂缝识别,对 YOLOv5 目标检测模型进行如下改进: ①引入 Swin Transformer 网络,替换原模型特征融合结构中的 CSP 模块,使模型充分利用检测目标与周围环境的关联信息,提高微小裂缝特征提取能力; ②在模型检测头部结构中增加了一个下采样检测头,提高模型对各尺寸裂缝目标检测的能力; ③在特征融合结构中引入 CBAM 机制,提高模型在复杂背景下的抗干扰能力。改进后整体模型结构如图 1 所示。

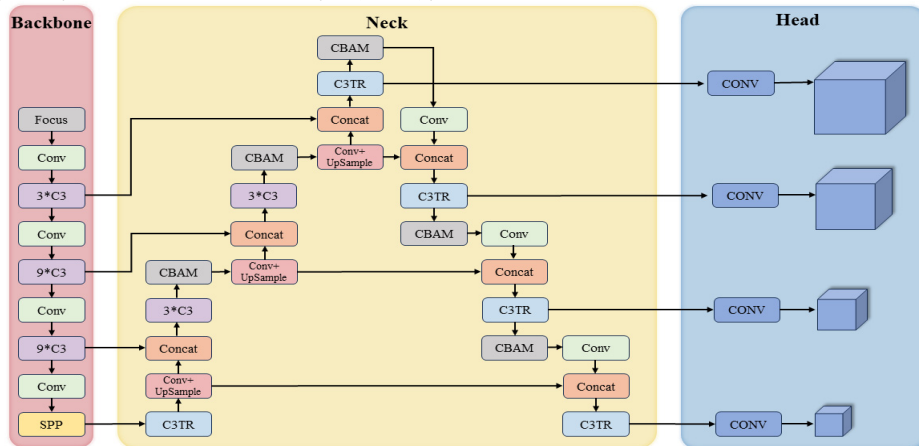


图 1 改进 YOLOv5 目标检测模型整体结构

Fig.1 Architectural modifications to the YOLOv5 object detection model

2.1 特征融合网络的改进

1) Swin Transformer 网络结构。在 YOLOv5 网络模型中，大量使用了 CSP 模块，而该模块由于卷积层结构较为简单，导致其对裂缝小目标的检测能力较低。其次，由于其卷积核大小为 3×3 ，所以在检测长条形裂缝时，会出现滑动窗口无法覆盖的问题，使得裂缝检测出现不连续和漏检的情况。此外，该模块参数量较少，使得其需要更多的训练样本以保证模型的精度和泛化能力。上述缺陷严重降低了 YOLOv5 模型在混凝土高坝航拍图像缺陷识别任务中的性能。因此引入了 Swin Transformer 网络替换该结构，有效地解决了这一问题。

Swin Transformer 的核心是将特征图分割为固定大小的窗口，在每个窗口内进行局部窗口内注意力计算，然后对窗口进行移位并再次计算局部注意力，以此实现窗口间的信息交流，提高全局特征获取能力，其结构如图 2 所示。Swin Transformer 模型将输入的 $H \times W \times 3$ 的红绿蓝三通道（Red Green

Blue，RGB）原始图像通过补丁分区层（Patch Partition）进行分块操作，每相邻的 4×4 个像素成为一个补丁，然后将所有补丁在 RGB 通道上进行叠加，每个补丁的特征维度为 $\frac{H}{4} \times \frac{W}{4} \times 48$ ，也即将

图像展开为维度为 $\frac{H}{4} \times \frac{W}{4} \times C$ 的张量。在阶段一中，将划分好的补丁块输入线性嵌入层（Linear Embedding），将图像的特征维度映射到维度 C ，随后将图像传入 Swin Transformer Block 模块进行特征提取并输出特征图。为了生成层级化结构，特征图在继续输入补丁合并层（Patch Merging）时将缩小图像分辨率，增大感受野，以获取全局特征信息。阶段二中，补丁合并层将输入的 2×2 的补丁块进行合并，通过一个卷积层后，输出图像维度变为 $\frac{H}{8} \times \frac{W}{8} \times 4C$ 。阶段三和阶段四将重复补丁合并和特

征转换操作，输出尺寸为 $\frac{H}{16} \times \frac{W}{16} \times 4C$ 和 $\frac{H}{32} \times \frac{W}{32} \times 8C$ 。

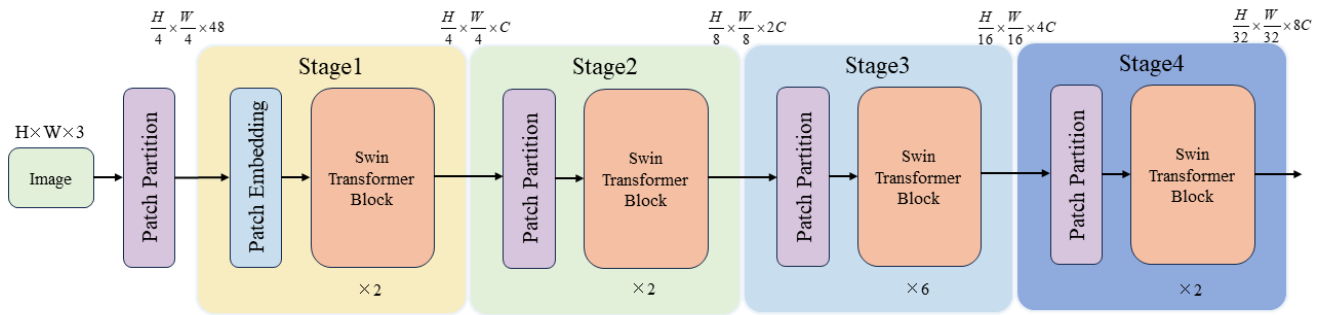


图 2 Swin Transformer 网络结构
Fig.2 Network architecture of Swin Transformer

2) CBAM 机制。在无人机对混凝土高坝进行图像数据采集时，时常会拍摄到坝体附近的环境背景，这些复杂的干扰信息将严重影响裂缝缺陷的检测精度，造成背景信息的误判。同时，坝体表面的水渍等干扰也会影响检测模型的判断。基于此，本节将卷积注意力模块（Convolutional Block Attention Module，CBAM）引入到了模型特征融合网络中，提高了模型的抗干扰能力。

CBAM 是一种应用于前馈神经网络的注意力机制。其流程和原理是给定一个中间特征图，卷积注意力模块将沿着通道和空间两个独立维度依次判断注意力图，然后将注意力图与输入特征图进行自适应特征优化。通道注意力着重关注最有用的输入图像，空间注意力则关注于最具信息量的部分，是对通道注意力的加强与补充。

2.2 检测头部的改进

YOLOv5 网络模型中包含了 3 个不同尺度的特征图检测头，每个尺度设定了 3 种尺寸的锚框，分别对应检测大、中、小型目标，模型在数据集训练中通过

自适应锚框计算得到 9 种最佳锚框值，其最小感知范围为 8×8 的区域，而在混凝土高坝航拍缺陷图像中，大多图像处于远距离拍摄，其中的裂缝像素占比较低，当低于 8 像素时，极易出现漏检现象。基于此，为改善模型对于细微裂缝的检测性能，本节在模型中增加了一个 4 倍下采样检测头，使得模型能在 4 倍、8 倍、16 倍和 32 倍进行下采样操作，生成 12 种多尺度锚框，提高模型的小目标检测能力。

3 实际工程验证

3.1 数据预处理

样本采集是构建高精度识别模型的基础，本研究系统地采集某大型水电站混凝土高坝坝体、廊道、消力池及边坡支护等关键部位的高清图像，并从中筛选和标注裂缝缺陷样本用于模型训练与测试。

采用以大疆无人机航拍为主、人工近景补充为辅的图像采集方式进行立体化数据获取。其中，无人机主要负责大范围、高空域、多角度图像采集，并通过多角度高清成像对坝体及相关结构实施精细化拍摄；

人工近景拍摄主要用于补充局部细部区域和遮挡区域影像,从而提高样本完整性与代表性。具体采集过程中,设置了多组航线系统,围绕某大型水电站的坝体主体、廊道、消力池及边坡支护等关键部位进行系统性图像采集。坝体典型缺陷的图像采集过程如图3所示。

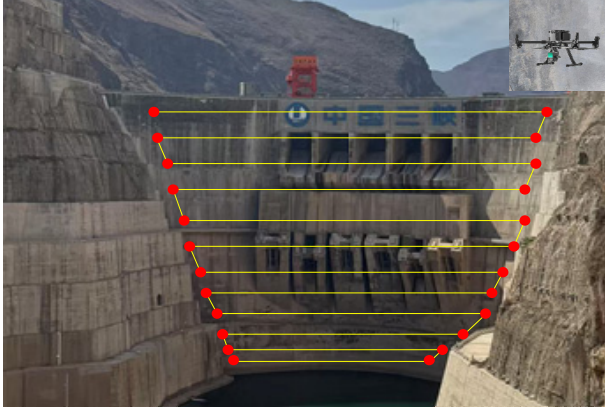


图3 坝体缺陷无人机采集

Fig.3 UAV-based imaging of dam body defects.

而在进行深度学习模型训练前,需要对收集到的图像数据集进行标注用于模型的特征学习,包括缺陷位置信息与缺陷类别信息。采用 labelme 软件进行标注,图像标注完成后生成 JavaScript 对象表示法 (JavaScript Object Notation, JSON) 文件,根据基准模型 YOLOv5 的要求,将标注数据集转换为通用物体语境 (Common Objects in Context, COCO) 数据集格式,按照 6:2:2 的比例划分训练集、验证集和测试集。

3.2 实验环境与参数设置

将进行的混凝土结构缺陷检测实验采用的实验环境与参数设置如表1所示。

表1 模型训练参数表

Tab.1 Model training hyperparameters

参数名称	参数值	参数说明
Image_size	608×608×3	训练图像分辨率
Epochs	300	迭代运行次数
Batch_size	4	训练批次
Optimizer	Adam	优化器
Lr	0.001	初始学习率
Momentum	0.937	学习率动量
Weight_decay	0.0005	权重衰减系数

3.3 评价指标

混凝土缺陷目标检测任务的主要目标有2个:①检测图像中裂缝的位置;②检测图像中裂缝的大小。在混凝土结构裂缝目标检测任务中,引入损失函数 (Loss)、精确率 (Precision)、召回率 (Recall)、平均精度均值 (Mean Average Precision, mAP) 和帧率 (Frames Per Second, FPS) 这5个指标来定量的描

述所提模型对裂缝目标检测的效果。

1) 损失函数。损失函数主要用于描述神经网络预测信息与期望信息之间的差距,预测值越接近于标签期望值,则损失函数值越小,其在训练过程中的趋势代表了模型训练的收敛情况。在本研究所提的改进 YOLOv5 模型中,损失函数为 $Loss = Loss_{rect} + Loss_{obj}$, 式中, $Loss_{rect}$ 表示定位损失函数, $Loss_{obj}$ 表示置信度损失函数,损失函数越小,表示模型效果越好。

2) 精确率与召回率。在目标检测任务中,真正例 (True Positives, TP) 表示期望值为正样本且被划分为正样本的样本数量,真反例 (True Negatives, TN) 表示期望值为负样本且被划分为负样本的样本数量,假正例 (False Positives, FP) 表示期望值为负样本但被划分为正样本的样本数量,假反例 (False Negatives, FN) 表示期望值为正样本但被划分为负样本的样本数量。

精确率又称查准率,表示预测出的正样本数量占所有预测为正样本数量的比例,用以反映预测结果中,正样本正确预测的概率;召回率表示预测出的正样本数量占实际正样本数量的比例,反映模型漏检情况,具体公式如式3和式4所示。

$$P = \frac{TP}{TP + FP} \quad (3)$$

$$R = \frac{TP}{TP + FN} \quad (4)$$

式中, TP 表示真正例数量, FP 表示假正例数量, FN 表示假反例数量, TN 表示真反例数量。

3) 平均精度均值。精确率与召回率是一组相互制约的评价指标,具有单点值局限性,故引入平均精度均值来均衡两个指标的评价结果,以充分评估模型的性能。以召回率为横坐标,精确率为纵坐标,可以得到精确率-召回率曲线 (Precision-Recall curve, PR 曲线),曲线与坐标轴围成的面积表示模型的平均精确度 (Average Precision, AP), mAP 表示数据集中所有类别的 AP 的均值,由于本次任务是二分类任务,故 mAP 与 AP 相等。计算公式如式5所示,阈值设置为 0.5,即交并比大于 0.5 的预测框属于有效,表示为 mAP_0.5, mAP_0.5:0.95 交并比从 0.5 一直取到 0.95,每隔 0.05 计算一次 mAP 值,最后取均值。

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (5)$$

4) 帧率。帧率 (Frames Per Second, FPS) 用以描述模型推理和检测的速度,即目标网络每秒可以处理图像的数量。

3.4 结果分析

根据前文标注的数据集和设定的参数对改进 YOLOv5 模型进行训练, 训练过程中输出损失函数、精确率等可视化训练结果。训练完成后, 利用结果文件中的最优权重文件对测试集进行推理, 根据所提的 5 项指标定量分析模型的检测效果。

为进一步评估模型对细微裂缝的检测能力, 本文在总体测试集评价之外, 额外统计了细微裂缝样本子集上的 Precision、Recall 和 mAP 指标。结果表明, 相比基线模型, 改进模型在细微裂缝子集上的检测性能提升更为明显, 说明所提出的 Swin Transformer、

下采样检测头与 CBAM 组合能够有效增强小目标裂缝的特征表达和背景抑制能力。如图 4 所示为改进 YOLOv5 模型在本次裂缝缺陷检测训练任务中的可视化结果, 可以看出, 两项损失函数均在训练轮次内快速下降, 最终基本稳定在 0.01 附近, 表明模型训练收敛情况良好; 模型召回率和精确率在第 50 个训练轮次前快速上升, 最终稳定在 0.91 和 0.94 附近, 表示模型在类别判断和检测框回归计算任务中表现良好。根据 4 项训练可视化结果可以看出, 模型在训练过程中收敛良好, 预测效果精确, 模型在裂缝检测任务中表现良好。

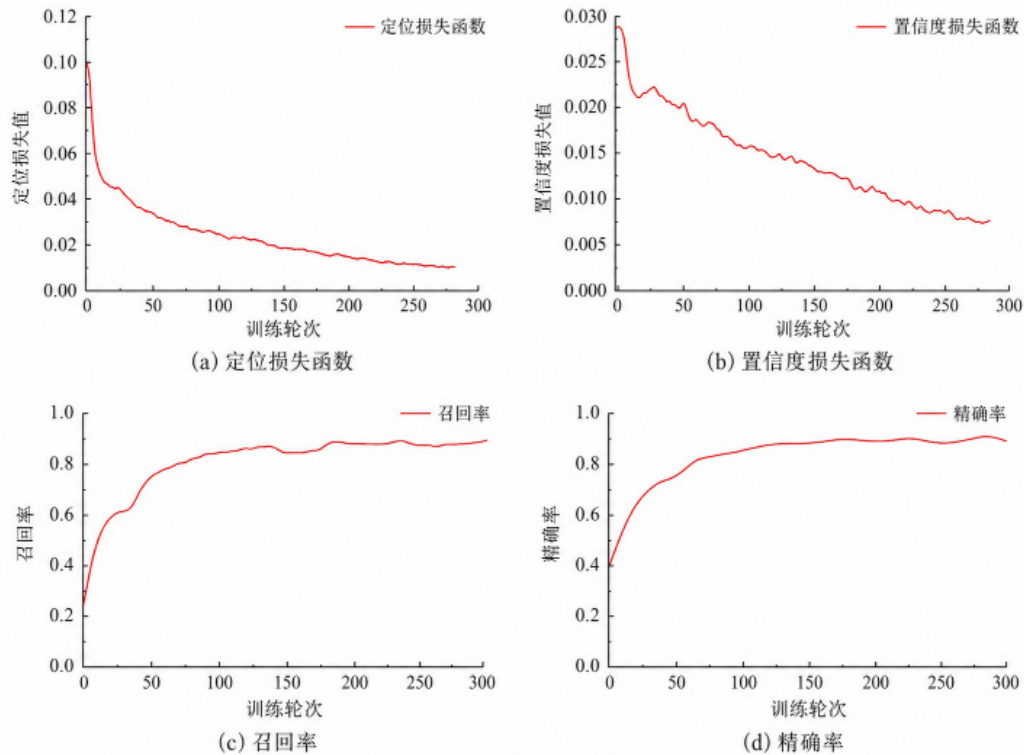


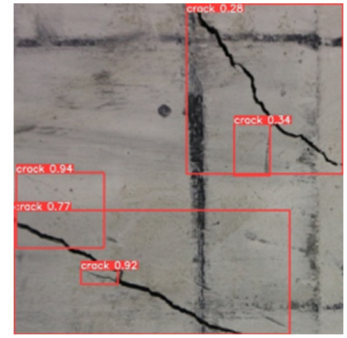
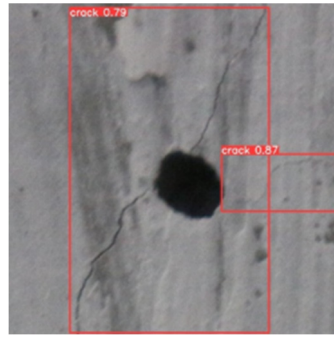
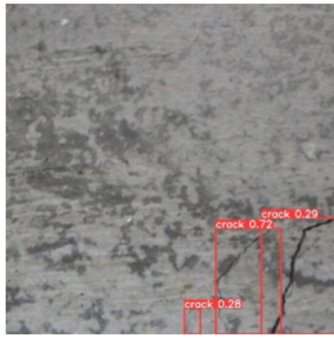
图 4 改进 YOLOv5 网络模型训练可视化结果

Fig.4 Visual results from the improved YOLOv5 architecture

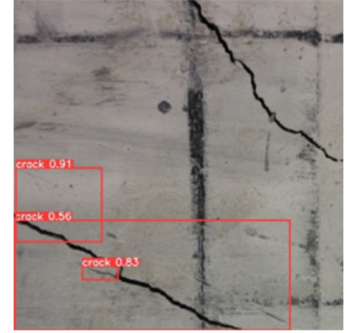
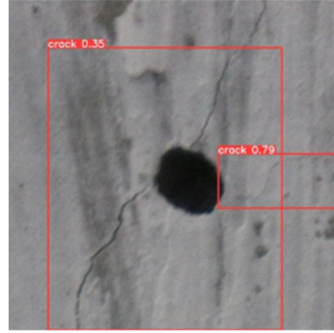
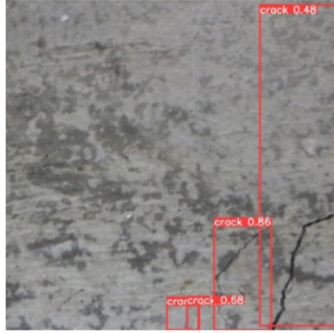
为确保对比实验结论的公平性与稳健性, 本次研究中的所有模型均遵循统一的实验协议。具体而言, 参与对比的基准模型 (RetinaNet, YOLOX, CornerNet) 与本文方法均采用相同的输入分辨率 (608×608)、训练周期 (300 epoch) 及 ImageNet-1k 预训练权重初始化。数据增强策略严格统一, 包括 Mosaic、随机仿射变换、水平翻转与色相-饱和度-明度 (Hue Saturation Value, HSV) 色彩空间扰动, 不引入任何模型特有的额外增强。所有模型的超参数 (如学习率、权重衰减) 均通过相同的贝叶斯优化流程在设定空间内搜索确定, 并采用一致的早停准则。为评估结果的统计稳健性, 每组实验均在仅随机种子不同的条件下独立重复 5 次。后续所述评估指标均为 5 次实验的均值±标准差, 并以平均结果作为最终比较依据, 以降低随机性偏差, 确保结论可靠。

为了进一步验证本研究所提改进 YOLOv5 模型的有效性与优越性, 引入了 RetinaNet、CornerNet 以

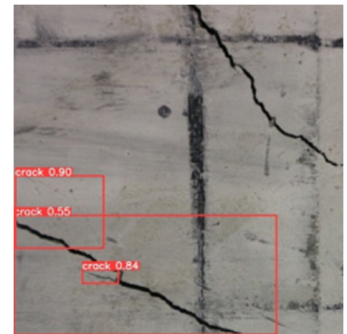
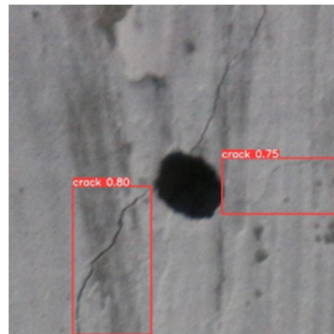
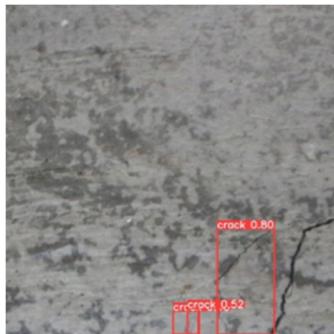
及 YOLOX 共 3 种经典检测算法进行对比实验。图 5 对比了 4 种模型在数据集上的检测效果。从图中可以看出, 该数据集包含不同尺寸的裂缝缺陷、背景存在一定干扰等特点, 基本符合无人机航空摄影图像集所具备的特点。相较而言, RetinaNet 模型在对细微裂缝的检测中存在漏检, 且检测框定位出现偏差, 不能准确地框选出裂缝的完整位置, 预测置信度较低; CornerNet 模型和 YOLOX 模型在细微裂缝的检测上具有一定的优势, 但同样存在漏检、检测框定位不精确的问题; 综合性能较优的是改进 YOLOv5 检测模型, 该模型能检测到完整的多尺度裂缝目标, 有效缓解了漏检现象, 且检测框定位准确, 能精确地描述裂缝的位置, 置信度较高, 表明本研究所提改进 YOLOv5 模型具有较强的特征提取能力与检测精度, 在针对无人机航拍图像的混凝土裂缝检测中具有较好的效果。



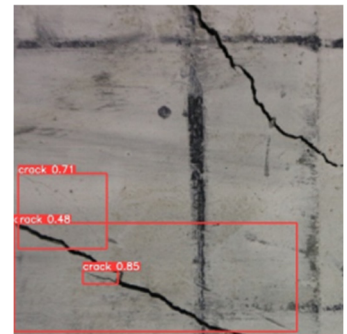
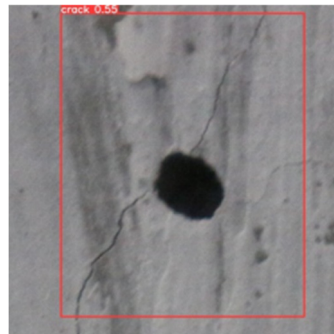
(a) 改进 YOLOv5 模型检测结果



(b) YOLOX 模型检测结果



(c) RetinaNet 模型检测结果



(d) CornerNet 模型检测结果

图 5 混凝土裂缝检测结果

Fig.5 Concrete crack detection results.

为了进一步验证所提改进 YOLOv5 模型的有效性与优越性, 这里引入了 RetinaNet^[18]、CornerNet^[19]以及 YOLOX^[20]共 3 种经典检测算法进行对比实验。表 2 为不同网络之间精确度对比情况, 可以看出, 本研究所提改进 YOLOv5 网络模型的检测精度在该数据集下优于各类经典算法, 相较于 CornerNet 模型,

所提模型的 $mAP@0.5$ 和 $mAP@0.5:0.95$ 分别提高 10.16 和 17.67 个百分点; 相较于 YOLOX, 分别提高 3.23 和 1.80 个百分点, 表明模型在检测精度上具有一定的优势。表 2 还可以看出, 本研究所提改进 YOLOv5 模型平均检测为 34.3FPS, 高于所比较的 3 种模型, 表明该模型具有实时检测能力。

表 2 不同模型检测性能对比

Tab.2 Comparison of detection performance among different models

模型	mAP@0.5/%	mAP@0.5:0.95/%	推理速度/(frame·s ⁻¹)
改进 YOLOv5	93.29±0.41	74.39±0.38	34.3±0.5
RetinaNet	87.16±0.67	63.32±0.59	29.8±0.4
CornerNet	83.13±0.92	56.72±0.84	27.3±0.6
YOLOX	90.06±0.35	72.59±0.46	32.5±0.4

综合上述结论与分析,本研究所提改进策略行之有效,有效提升了模型的特征提取能力与检测精度,在针对细微裂缝目标的检测任务中具有一定的优势,有效解决了传统 YOLOv5 模型面临复杂背景干扰、目标尺度微小等情况下出现的误检和漏检现象,具备良好的泛化能力。

4 结论

本研究面向混凝土高坝裂缝目标细长、尺度变化大及易受复杂背景干扰等特点,开展了基于 YOLOv5 的目标检测模型改进研究。在充分分析 YOLOv5 网络结构与特征表达机制的基础上,围绕航拍场景下裂缝多尺度表征不足的问题,对模型的特征融合路径与检测头结构进行针对性优化,以增强关键特征提取能力并提升检测精度。最终通过消融实验与对比实验,对各改进模块的贡献及整体性能提升进行了系统验证,结果表明所建模型在裂缝检测任务上具有良好的有效性与综合优势。本研究的主要工作如下:

1) 针对无人机航拍图像中裂缝目标尺度变化显著、细长且易受复杂背景干扰的特点,本研究对 YOLOv5 网络进行结构性增强。具体而言,引入 Swin Transformer 以替换特征融合结构中的 CSP 模块,从而更充分地建模目标与周围环境的上下文关联,提升微小裂缝的特征表征能力。同时,在检测头部分增设额外的下采样尺度,以强化多尺度检测与对不同尺寸裂缝目标的适应性。进一步地,在特征融合过程中嵌入 CBAM 注意力机制,实现对关键裂缝区域的自适应强化与背景噪声抑制,提高复杂场景下的鲁棒性与抗干扰能力。

2) 实验结果表明,改进 YOLOv5 模型在训练过程中具有较好的收敛性和稳定性,并在混凝土高坝裂缝检测任务中表现出更优的综合性能。其 mAP_{0.5}、mAP_{0.5:0.95} 和 FPS 分别达到 93.29%、74.39% 和 34.3。相较于 RetinaNet、YOLOX 和 CornerNet 模型,所提模型在检测精度和实时性方面均有所提升,说明该方法能够有效增强复杂背景下细微裂缝及多尺度裂缝目标的识别能力。

3) 根据消融实验量化结果可以看出,本研究所提改进策略有效地提升了 YOLOv5 模型针对细微

裂缝的特征提取能力与检测精度,在针对无人机航拍图像的混凝土裂缝检测中具有较好的效果。同时,根据各经典模型精度对比情况,验证了本研究提出的改进 YOLOv5 模型有效性与优越性。

参考文献

- [1] 吴中如,陈波.大坝变形监控模型发展回眸[J].现代测绘,2016,39(5):1-3.[Wu Zhongru,Chen Bo.Review on the development of dam deformation monitoring models[J].Modern Surveying and Mapping,2016,39(5):1-3.]
- [2] Jia,Jinsheng.A technical review of hydro-project development in China[J].Engineering,2016,2(3):302-312.
- [3] 吴中如,徐波,顾冲时,等.大坝服役状态的综合评判方法[J].中国科学:技术科学,2012,42(11):1243-1254.[Wu Zhongru,Xu Bo,Gu Chongshi,et al.Comprehensive evaluation method for dam service state[J].Science China Technological Sciences,2012,42(11):1243-1254.]
- [4] 顾冲时,李占超,徐波.基于动力学结构突变的混凝土坝裂缝转异诊断方法研究[J].中国科学:技术科学,2011,41(7):1000-1009.[Gu Chongshi,Li Zhanchao,Xu Bo.Diagnosis method of concrete dam crack transition based on dynamic structural mutation[J].Science China Technological Sciences,2011,41(7):1000-1009.]
- [5] 顾冲时,李波,虞鸿,等.碾压混凝土坝力学参数的反分析[J].中国科学:技术科学,2010,40(6):651-656.[Gu Chongshi,Li Bo,Yu Hong,et al.Back analysis of mechanical parameters of roller compacted concrete dams[J].Science China Technological Sciences,2010,40(6):651-656.]
- [6] Dung C V.Autonomous concrete crack detection using deep fully convolutional neural network[J].Automation in Construction,2019,99:52-58.
- [7] Fang Fen,Li Liyuan,Gu Ying,et al.A novel hybrid approach for crack detection[J].Pattern Recognition,2020,107:107474.
- [8] 鲍跃全,李惠.人工智能时代的土木工程[J].土木工程学报,2019,52(5):1-11.[Bao Yuequan,Li Hui.Civil engineering in the era of artificial intelligence[J].China Civil Engineering Journal,2019,52(5):1-11.]
- [9] 杨扬,王连发,张宇峰,等.基于多特征融合的混凝土结构表面病害图像分类算法[J].长安大学学报(自然科学版),2021,41(3):64-74.[Yang Yang,Wang Lianfa,Zhang Yufeng,et al.Image classification algorithm for surface defects of concrete structures based on multi-feature fusion[J].Journal of Chang'an University (Natural Science Edition),2021,41(3):64-74.]
- [10] Shi Yong,Cui Limeng,Qi Zhiqian,et al.Automatic road crack detection using random structured forests[J].IEEE Transactions on Intelligent Transportation Systems,2016,17(12):3434-3445.
- [11] 靳华中,万方,叶志伟.结合路面深度影像梯度方向直方图

- 和分水岭算法的裂缝检测[J].华中师范大学学报(自然科学版),2017,51(5):715-722.[Jin Huazhong,Wan Fang,Ye Zhiwei. Crack detection combining histogram of oriented gradients of pavement depth images and watershed algorithm[J].Journal of Central China Normal University (Natural Sciences),2017, 51(5):715-722.]
- [12] 王睿,漆泰岳.基于机器视觉检测的裂缝特征研究[J].土木工程学报,2016,49(7):123-128.[Wang Rui,Qi Taiyue. Study on crack characteristics based on machine vision detection[J].China Civil Engineering Journal,2016,49(7): 123-128.]
- [13] 孟诗乔,张啸天,乔魁阳,等.基于深度学习的网格优化裂缝检测模型研究[J].建筑结构学报,2020,41(增刊2):404-410.[Meng Shiqiao,Zhang Xiaotian,Qiao Suyang,et al.Research on grid-optimized crack detection model based on deep learning[J].Journal of Building Structures,2020,41(S2):404-410.]
- [14] 杨杰文,章光,陈西江,等.基于深度学习的较复杂背景下桥梁裂缝检测[J].铁道科学与工程学报,2020,17(11):2722-2728.[Yang Jiewen,Zhang Guang, Chen Xijiang,et al.Bridge crack detection under complex background based on deep learning[J].Journal of Railway Science and Engineering,2020,17(11):2722-2728.]
- [15] 李松,史涛,井方科.改进 YOLOv8 的道路损伤检测算法[J].计算机工程与应用,2023,59(23):165-174.[Li Song, Shi Tao,Jing Fangke.Road damage detection algorithm based on improved YOLOv8[J].Computer Engineering and Applications,2023,59(23):165-174.]
- [16] 吴刚,罗炜,王小龙,等.基于深度学习的盾构隧道衬砌表面病害检测模型研究[J].现代隧道技术,2023,60(4):67-75.[Wu Gang,Luo Wei,Wang Xiaolong,et al.Research on surface defect detection model of shield tunnel lining based on deep learning[J].Modern Tunnelling Technology,2023,60(4): 67-75.]
- [17] 王燕华,何俊泽,张明洲,等.基于 SSD 及剪枝神经网络的复杂环境下混凝土裂缝识别[J].Journal of Southeast University (English Edition),2023,39(4):393-399.[Wang Yanhua,He Junze,Zhang Mingzhou,et al.Concrete crack identification in complex environments based on SSD and pruned neural networks[J].Journal of Southeast University (English Edition),2023,39(4):393-399.]
- [18] Lin T Y,Goyal P,Girshick R,et al.Focal loss for dense object detection[C]//Proceedings of the IEEE International Conference on Computer Vision.Piscataway:IEEE,2017: 2980-2988.
- [19] Law H,Jia Deng.Cornet:Detecting objects as paired keypoints[C]//Proceedings of the European Conference on Computer Vision.Cham:Springer,2018:734-750.
- [20] Ge Zheng,Liu Songtao,Wang Feng,et al.YOLOX:exceeding YOLO series in 2021[PP/OL].V.2.(2021-08-06)[2026-08-05].<https://doi.org/10.48550/arXiv.2107.08430>.

Research on Intelligent Crack Detection Method for Concrete High Dams Based on Computer Vision

YE Fei¹, WANG Yun¹, WANG Cheng¹, ZHONG Hongxing¹, SHI Yan², BAO Tengfei², LI Yangtao²

(1. Luquan Wudongde Power Plant of Three Gorges Jinsha River Yunchuan Hydropower Development Co., Ltd., Kunming 651512, China; 2. Three Gorges Geotechnical Consultants Co., Ltd., Wuhan 430074, China)

Abstract: 【Objective】 Concrete dams are critical structures in hydraulic engineering systems and play essential roles in flood control, irrigation, and power generation. To address the challenges of slender crack morphology, large scale variation, complex background interference, and the low efficiency and subjectivity of manual inspection in UAV images of concrete high dams, this study proposes an intelligent crack detection method based on computer vision, aiming to provide technical support for dam safety inspection and defect identification. 【Methods】 UAV-acquired crack images of concrete high dams were used as the research object, and structural improvements were made to the YOLOv5 object detection model. Specifically, the Swin Transformer was introduced to replace the original CSP module to enhance global contextual modeling capability; a 4× downsampling detection head was added to improve the detection of fine cracks and multi-scale targets; and the CBAM attention mechanism was embedded to strengthen key crack features while suppressing complex background interference.

【Results】 Experimental results show that the improved YOLOv5 model achieved stable training convergence and superior performance in concrete high dam crack detection, with mAP@0.5, mAP@0.5:0.95, and FPS reaching 93.29%, 74.39%, and 34.3, respectively. Compared with classical detection models, the proposed model performed better in detection accuracy and real-time capability, effectively reducing missed detections of fine cracks and false detections caused by complex backgrounds. 【Conclusion】 The improved YOLOv5 model can enhance feature representation, multi-scale detection, and background interference suppression for concrete high dam cracks in UAV-based inspection scenarios. The proposed method demonstrates favorable detection accuracy, real-time performance, and engineering application potential, and can provide a reference for intelligent inspection and safety assessment of concrete high dams.

Key words: concrete dam; deep learning; machine vision; crack detection