

基于矢量量化变分自编码器的混凝土表观裂缝检测算法

刘超, 吴纪曙

(同济大学 土木工程学院, 上海 200092)

摘要: 提出了一种基于第 2 代矢量量化变分自编码器(VQ-VAE-2)的自监督混凝土表观裂缝检测算法,可以在缺少裂缝样本的条件下实现高效检测。以重建误差为检测指标,利用无裂缝图片训练 VQ-VAE-2,使其在重建裂缝图片时产生更大的重建误差;在计算重建误差时将原图和重建图片均分割成若干图块,取对应图块间重建误差最大值作为图片的重建误差,以增大 2 类图片的重建误差差异。结果表明,该算法的精确率为 0.954,召回率为 0.959,准确率为 0.956, F1 分数为 0.957。在无裂缝样本作为训练集的情况下,该算法能较好地完成混凝土表观裂缝检测任务。

关键词: 桥梁工程;混凝土表观裂缝检测;深度学习;变分自编码器;异常检测

中图分类号: U445.7⁺1

文献标志码: A

Detection Algorithm of Concrete Structural Apparent Cracks Based on VQ-VAE-2

LIU Chao, WU Jishu

(College of Civil Engineering, Tongji University, Shanghai 200092, China)

Abstract: In this paper, we propose a self-supervised algorithm based on VQ-VAE-2 for the automated detection of concrete structural apparent cracks. The algorithm demonstrates the capability to effectively detect cracks without available crack samples. By taking the reconstruction error as detection index, VQ-VAE-2 is trained on crack-free images. When applied to images with cracks, VQ-VAE-2 produces higher reconstruction errors. The original and reconstructed images are partitioned into blocks for calculating the reconstruction error. The maximum value of the reconstruction error between corresponding blocks is taken as the reconstruction error of the image. This approach increases the difference in

reconstruction error between the two types of images. The results show that the algorithm achieves a precision of 0.954, a recall of 0.959, an accuracy of 0.956, and an F1 score of 0.957. These results indicate that the algorithm can effectively detect concrete structural apparent cracks even without crack samples in the training set.

Keywords: bridge engineering; concrete structural apparent crack detection; deep learning; variational autoencoders; anomaly detection

混凝土是桥梁结构的主要材料。开裂是混凝土结构的主要病害之一,若不及时修复,则会导致桥梁结构耐久性能和承载能力退化。因此,裂缝检测是混凝土结构后期管养中的一项重要任务。然而,传统的裂缝检测方法严重依赖人工,既不高效率也不准确。为了提高混凝土裂缝检测效率,许多基于计算机视觉技术的检测手段被提出。目前基于计算机视觉技术的检测手段主要采用深度学习神经网络,可分为有监督模型和无监督模型。有监督模型,如 AlexNet、GoogLeNet、ResNet、YOLO 等^[1-5],在混凝土裂缝检测上表现良好。然而,有监督模型都依赖裂缝训练集,为了保证裂缝检测的准确性,必须采集大量的裂缝图片以供模型训练。实际工程中,标注的裂缝图片样本一般较少,训练的有监督模型容易发生拟合。

为了克服裂缝检测任务中训练集数据量不足困难,有学者通过训练神经网络生成裂缝图片进行数据增广。侯越等^[6]和 Xu 等^[7]分别采用生成对抗网络(GAN)和卷积自编码器网络(CAE)生成裂缝图片,然后将生成的裂缝图片与真实的裂缝图片组成有监督模型的训练集。上述生成模型仍然需要一定

收稿日期: 2022-11-28

第一作者: 刘超,副教授,博士生导师、工学博士,主要研究方向为智能监测技术、精细化设计理论等。

E-mail: lctj@tongji.edu.cn



论文
拓展
介绍

数量的裂缝图片作为训练数据,并存在生成的裂缝图片清晰度不高、多样性较差的问题。另一种可行的思路是将裂缝视为异常,利用无监督算法检测混凝土裂缝。其中,基于重建误差的检测方法应用广泛。在训练阶段仅输入正常图片并使得输出图像尽可能接近输入图像,能较好地重建正常图片,而对于异常图片则重建能力较差。2类图片的重建误差存在明显差异,根据其是否超过阈值可以达到检测异常图片的目的^[8]。

基于重建误差的检测方法常用的深度学习模型为自编码器。An等^[9]介绍了基于自编码器的异常检测原理,并提出以变分自编码器(VAE)的重建概率取代重建误差作为异常检测指标。Hasan等^[10]提出了一种基于全卷积自编码器的异常检测方法。Zhou等^[11]提出了Robust Deep Autoencoders,提高了自编码器的降噪能力,并将其应用于异常检测任务。虽然上述自编码器在异常检测公开数据集上取得了良好效果,但是在实际工程应用中仍具有泛化能力太强和重建图片模糊的缺陷。Gong等^[12]在自编码器中加入记忆模块,组成记忆自编码器(MemAE)以增强对正常数据特征的学习,从而增大正常与异常的重建误差差距。虽然MemAE改善了自编码器泛化能力过强的问题,但是其同样难以重建出清晰的图片,不利于实际工程的应用。

第2代矢量量化变分自编码器(VQ-VAE-2)采用层级结构,融合多尺度感受野特征,具有强大的图片重建能力,能重建出高清晰度图片^[13]。将混凝土表面裂缝检测转化为异常检测任务,基于VQ-VAE-2提出了一种无监督混凝土表面裂缝检测算法。首先采用正常混凝土图片训练用于图片重建的VQ-VAE-2,然后利用VQ-VAE-2重建待检测的图片,将原图与重建图片均分割成若干图块,计算各相应图块间的重建误差,取图块间的重建误差最大值作为图片的重建误差,最后根据图片重建误差进行裂缝检测。该算法旨在解决裂缝检测任务中裂缝样本不足的问题,能在没有裂缝样本作为训练数据的情况下准确地检测出包含裂缝的图片。

1 VQ-VAE-2简介

VQ-VAE-2是由DeepMind团队于2019年提出的一种变分自编码器。本文所提出的算法是源于基于重建误差的异常检测方法,这类方法利用正常样本训练深度学习模型来重建正常样本,从而使模型

在重建正常样本时产生较小的重建误差,而重建异常样本时产生较大的重建误差。VQ-VAE-2在本文算法中是用于重建输入样本的深度学习模型。

1.1 离散的潜变量

变分自编码器是在2013年被提出的一种自监督深度学习模型,由编码器和解码器两部分组成。在训练过程中,编码器将输入的训练图片降维并匹配潜变量 z 的概率分布,然后解码器从概率分布中抽样出潜变量 z 并用其重建输入的图片。本质上,变分自编码器是在推理样本的概率分布,再从中抽取样本。概率分布 $p(x)$ 计算式为

$$p(x) = \frac{p(x|z)p(z)}{p(z|x)} \quad (1)$$

式中: x 为图片样本; $p(x|z)$ 为似然分布; $p(z|x)$ 为后验分布; $p(z)$ 为先验分布。

根据式(1),为了求解图片样本的概率分布 $p(x)$,需得知似然分布 $p(x|z)$ 、后验分布 $p(z|x)$ 和先验分布 $p(z)$ 。在传统的变分自编码器中,潜变量的概率分布一般假定为标准正态分布,即 $z \sim N(0, I)$ 。后验分布 $p(z|x)$ 和似然分布 $p(x|z)$ 分别由编码器和解码器进行拟合。其中,后验分布 $p(z|x)$ 的拟合采用变分推理。变分推理是一种采用简单的分布来近似推理复杂的未知分布的方法。在VAE中,采用高斯分布来近似推理后验分布 $p(z|x)$ 。

VQ-VAE-2假设潜变量为离散分布,采用矢量量化实现离散的潜变量分布。一方面,减小了计算难度,增加了模型的推理速度;另一方面,离散分布的隐变量可以逼近连续分布,满足解码器重建输入图片的需求。此外,在真实世界中许多事物本身也是离散的^[14]。如图1所示,VQ-VAE-2采用矢量量化(VQ)的方式将潜变量的分布离散化,即模型在训练过程中训练一个编码表。编码器的输出向量 w 采用最邻近匹配的思路,按下式映射到编码表中的向量 w_k 。

$$Q(w) = w_k, k = \arg \min_j \|w - w_j\| \quad (2)$$

式中: w 为编码器的输出; w_k 、 w_j 均为编码表中的向量。

为了使编码表中的向量 w_k 尽可能表征输入图片的特征,应让其逼近编码器的输出向量 w ,因此在训练过程中需添加损失函数对编码表的参数更新进行约束。模型的损失函数如下:

$$L(x, y) = \|x - y\|_2^2 + \|\text{sg}(w) - w_k\|_2^2 + \beta \|\text{sg}(w_k) - w\|_2^2 \quad (3)$$

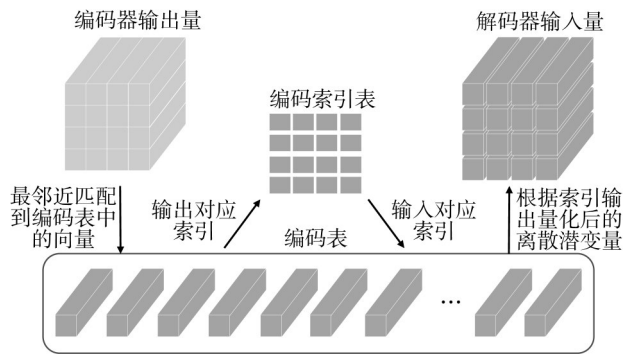


图 1 矢量量化过程

Fig.1 Process of vector quantization

式中: y 为解码器的输出; sg 为梯度截止计算符,此部分的偏导数为零; β 为权重系数。

模型的损失函数分为 3 项,第 1 项代表输入图像与重建图像的差异,后 2 项则计算编码器的输出向量 w 与映射的编码表中向量 w_k 间的差距。

1.2 层级结构

难以重建出清晰的图片是大多数自编码器的缺陷,限制了其在异常检测任务中的精度。为了重建出清晰的图片,VQ-VAE-2 采用层级结构,通过融合不同尺寸感受野特征获得更清晰的输出。如图 2 所示,该 VQ-VAE-2 具有 2 层编码器,其中上层具有更大的感受野。训练过程中,上层先将编码器的输出量化,并将量化的潜变量解码后与下层编码器的输出结合起来计算下层的潜变量。然后,将上下层潜变量共同作为解码器的输入并输出重建的图片,计算原图与重建图片的重建误差。采用这种方法实现了局部信息与全局信息分别编码,上层编码器负责编码全局信息,下层编码器在顶部潜变量的基础上编码局部特征,从而使模型能学习更多的纹理细节特征。

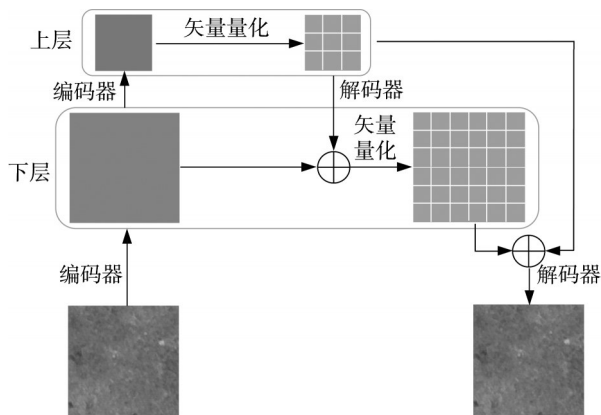


图 2 VQ-VAE-2 结构

Fig.2 Architecture of VQ-VAE-2

2 基于 VQ-VAE-2 的裂缝检测方法

2.1 检测方法

重建误差是表征深度学习模型重建图片与输入图片间差异程度的指标。在训练用于重建图片的深度学习模型时,常以重建误差作为损失函数,通过迭代更新模型的参数使重建误差最小化,从而达到输出图片尽可能接近输入图片的优化目标。模型训练时参数是根据预测值与训练集误差进行更新迭代,因此模型的重建能力很大程度上依赖训练集。当训练集仅包含正常混凝土图片时,训练得到的 VQ-VAE-2 只能学习到正常混凝土图片的特征,也就只能较好地重建正常混凝土图片。当输入图片中包含裂缝时,重建输入图片则更为困难。这导致重建的裂缝图片与原图具有更大差异,即具有更大的重建误差。如图 3 所示,以输入与输出图片的均方误差作为重建误差,训练的模型重建某张无裂缝图片时产生的重建误差为 0.000 08,远小于其重建某张裂缝图片时产生的重建误差 0.002 09。因此,选择合适的重建误差阈值可以比较准确地将图片进行分类。

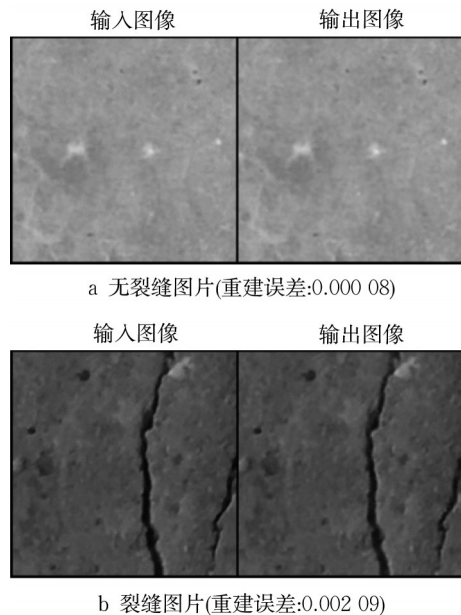


图 3 混凝土图片的重建误差

Fig.3 Reconstruction error of concrete images

以重建误差为判断指标,本文提出了一种基于 VQ-VAE-2 的无监督裂缝检测方法,如图 4 所示。首先,以仅包含正常混凝土图片的训练集训练 VQ-VAE-2;其次,将待检测图片输入 VQ-VAE-2 中进行重建;然后,将原图与重建图片均分割成 $k \times k$ 个图

块(k 为超参数,本文中 k 取值为4),并计算重建图片的图块与原图中相应的图块之间的重建误差,取所有图块重建误差的最大值作为此图片的重建误差代

表值;最后,对比重建误差及其阈值,若重建误差大于阈值则判断图片中包含裂缝,反之则为正常图片。

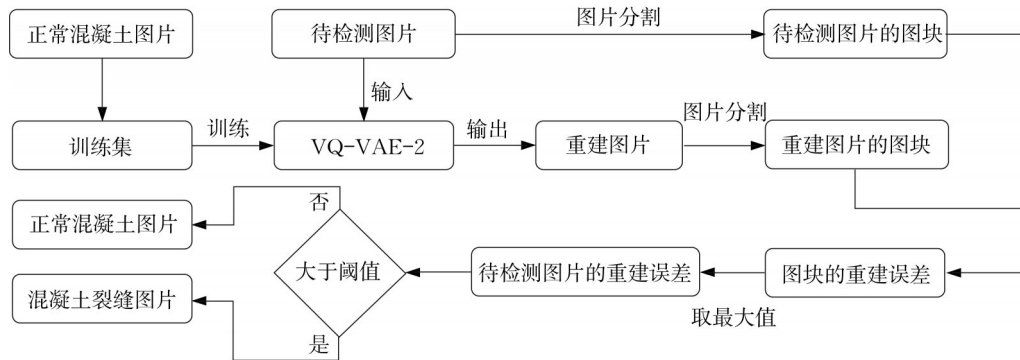


图4 基于VQ-VAE-2的混凝土表面裂缝检测算法

Fig.4 Detection algorithm of concrete structural apparent cracks based on VQ-VAE-2

2.2 重建误差计算方法

采用均方误差 e_{MSE} 作为重建误差,计算式为

$$e_{MSE} = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (I(i,j) - K(i,j))^2 \quad (4)$$

式中: M 为图片 I 和 K 的纵向像素尺寸; N 为图片 I 和 K 的横向像素尺寸。

在异常检测中,均方误差可以评估重建图片与原图之间像素级别上的平均差异程度,其值越低则2张图片相似程度越高。由于计算均方误差需要取均值,因此重建图片与原图之间的差异可能会被掩盖。图5展示了某张裂缝图片不同区域重建误差的差异,在裂缝区域,重建图片与原图存在较大差异,其他区域两者则差异不明显。然而,大多数情况下,裂缝对于整张图片而言往往只占小部分,因此在计算均值之后,某些图片虽然存在裂缝,但重建误差并不明显,从而导致算法误判。在计算重建误差时,先将重建图片与原图分别分割成同样数量的图块,然后计算对应图块之间的均方误差,最后取这些图块均方误差中的最大值作为图片的重建误差。这种重建误差的计算方法能有效增大无裂缝图片与有裂缝图片的重建误差差距,提高检测任务的准确性。

3 算法验证

3.1 数据集

为了评估所提出算法的有效性,采用公开数据集 CCIC (concrete crack images for classification)^[15] 进行算法验证。CCIC 数据集共由 40 000 张像素 227×227 的混凝土图片组成,其中正常混凝土图片

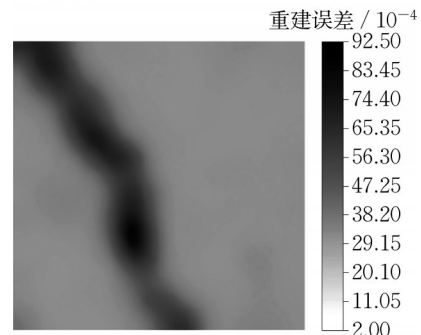


图5 某裂缝图片的异常分数图

Fig.5 Anomaly score map of a crack image

与裂缝图片各 20 000 张。选取 20 000 张正常混凝土图片及 5 000 张裂缝图片,划分成训练集和验证集。其中,训练集为 15 000 张正常混凝土图片,验证集为 5 000 张正常混凝土图片、5 000 张裂缝图片。

3.2 VQ-VAE-2 训练

在 CPU Intel (R) Core (TM) i7-10700K、GPU NVIDIA GEFORCE RTX3070ti 平台上,以 Python3.8 语言训练基于 PYTORCH 1.11.0 框架的 VQ-VAE-2。训练阶段学习率为 0.0003,输入图片像素为 224×224 ,批尺寸为 32,采用 Adam 优化器,共训练 560 代。训练过程中的损失曲线如图 6 所示。前期损失快速下降,经过 560 代训练后,模型的损失函数值已经收敛,且数值很小。

图 7 展示了训练 560 代后 VQ-VAE-2 重建训练集图片示例,重建效果良好,输入图片与输出图片基本难以进行区分。

3.3 评估指标

采用验证集评估模型的混凝土表面裂缝检测能

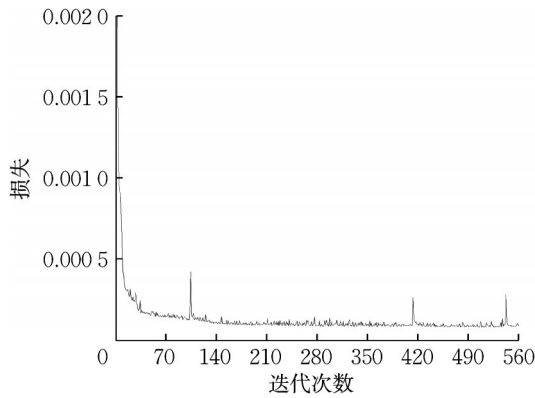


图6 损失曲线
Fig.6 Loss curve

力。评估指标采用精确率 P 、召回率 R 、准确率 A 、F1分数(F_1)和曲线包围面积 S_{AUC} 。精确率为正确预测裂缝图片数占全部预测裂缝图片数的比例。召回率为正确预测裂缝图片数占全部实际裂缝图片数的比例。准确率是正确分类的图片数与总图片数之比。 F_1 为精确率和召回率的调和平均。为了计算上述指标,首先将检测结果统计在混淆矩阵中,如表1所示。混淆矩阵是异常检测精度评价的重要工具。

表1 混凝土表观裂缝检测的混淆矩阵

Tab.1 Confusion matrix for concrete structural apparent crack detection

真实值	预测值	
	裂缝	无裂缝
裂缝	真正类 n_{TP}	假负类 n_{FN}
无裂缝	假正类 n_{FP}	真负类 n_{TN}

P 、 R 、 A 、 F_1 、 S_{AUC} 的计算式为:

$$P = \frac{n_{TP}}{n_{TP} + n_{FP}} \quad (5)$$

$$R = \frac{n_{TP}}{n_{TP} + n_{FN}} \quad (6)$$

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

$$F_1 = 2 \frac{PR}{P + R} \quad (8)$$

受试者工作特征曲线(ROC曲线)是以误识别率 F 为横轴、召回率为纵轴绘制的曲线,反映模型在不同分类阈值时误识别率和召回率的关系。误识别率计算式为

$$F = \frac{n_{FP}}{n_{TN} + n_{FP}} \quad (9)$$

S_{AUC} 为ROC曲线与横坐标轴围成的面积,其值介于0和1之间,值越大表征模型分类效果越好。

3.4 性能评估

选用MemAE^[12]与CAE^[16]作为对照。将验证集的图片输入VQ-VAE-2、MemAE、CAE中进行图片重建并计算重建误差。根据重建误差预测图片中是否包含裂缝,并将预测结果与真实标签进行对比,混淆矩阵结果如图8所示。根据混淆矩阵计算精确率 P 、召回率 R 、准确率 A 、F1分数(F_1),3种方法结果如图9a所示。结果表明,在没有裂缝混凝土图片作为训练数据的情况下,本文提出的检测算法可以比较准确地检测出包含裂缝的图片,并在精确率 P 、召回率 R 、准确率 A 、F1分数(F_1)上优于MemAE和CAE。3种方法的ROC曲线如图9b所示。根据曲线与横坐标围成面积可知,VQ-VAE-2在 S_{AUC} 指标上优于MemAE与CAE。

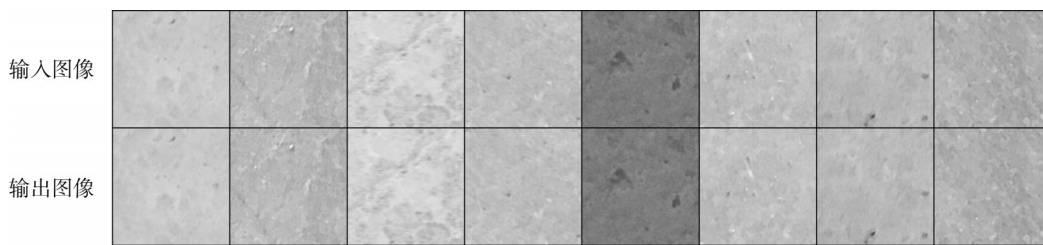


图7 重建训练集图片示例

Fig.7 Example of reconstructed training set images

为了验证将图块级均方误差的最大值作为图像重建误差的有效性,进行了消融实验。分别采用直接计算图像的均方误差作为重建误差以及将图块级均方误差最大值作为重建误差2种计算方式。消融实验结果如表2所示。结果表明,采用后一种计算方法更有效,在各项指标上均更优于前者。

3.5 可视化检验

为了实现图片中裂缝的位置检测并以可视化的方式展示检测结果,可将图片划分成更小的图块,并采用本文算法计算各图块均方误差,以类热力图的形式可视化裂缝。热力图中高数值的区域即为开裂区域,表现出较深的颜色。分割的图块数量越多,热

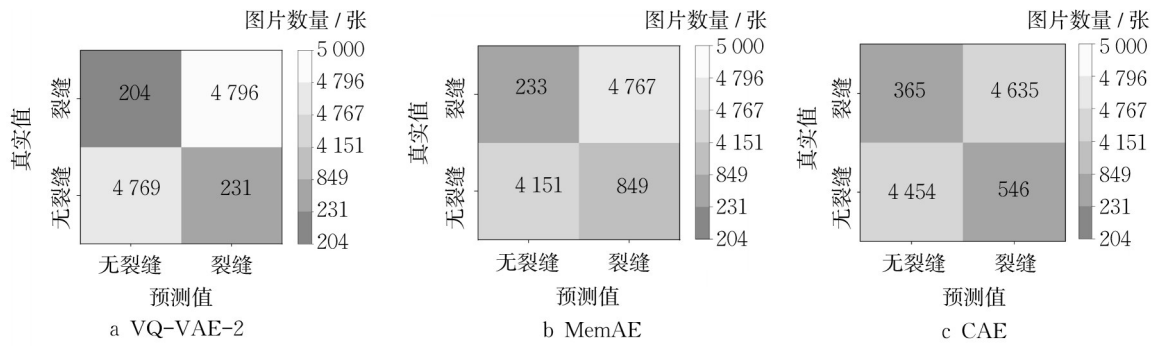


图 8 实验结果的混淆矩阵

Fig.8 Confusion matrices of test results

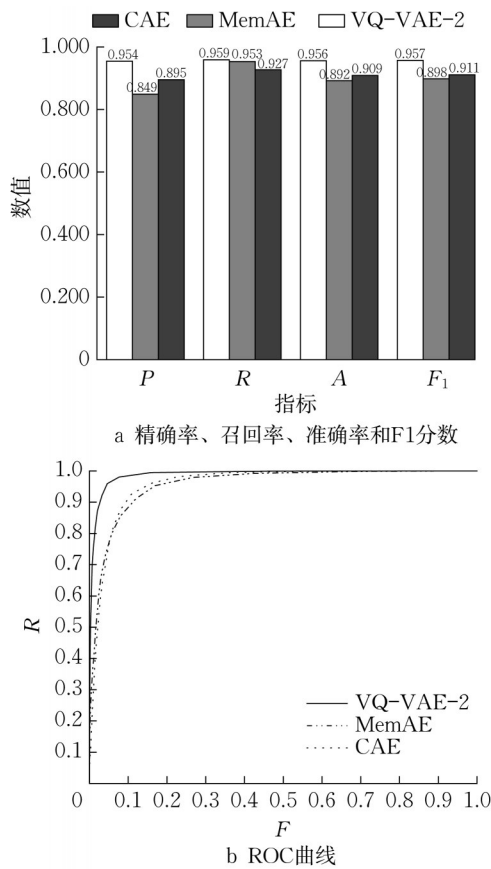


图 9 VQ-VAE-2、CAE 和 MemAE 的精确率、召回率、准确率、F1 分数和 ROC 曲线

Fig.9 Precision, recall, accuracy, F1 score and ROC curve of VQ-VAE-2, CAE and MemAE

表 2 不同重建误差计算方法性能

Tab.2 Performance of different reconstruction error calculation methods

损失函数计算方法	P	R	A	F ₁
图像级均方误差	0.898	0.956	0.924	0.926
图块级均方误差最大值	0.954	0.959	0.956	0.957

力图中裂缝位置越准确,但会导致计算量的显著增加。因此,将图片分割成 256 个图块,计算图块均方

误差(16×16 的矩阵),然后上采样至 224×224 后进行颜色映射,并随机选取了 4 张图片的裂缝可视化,如图 10 所示。结果表明,颜色较深的区域基本上与图像中裂缝所处的位置重叠,能较好地检测到裂缝在图片中的位置。

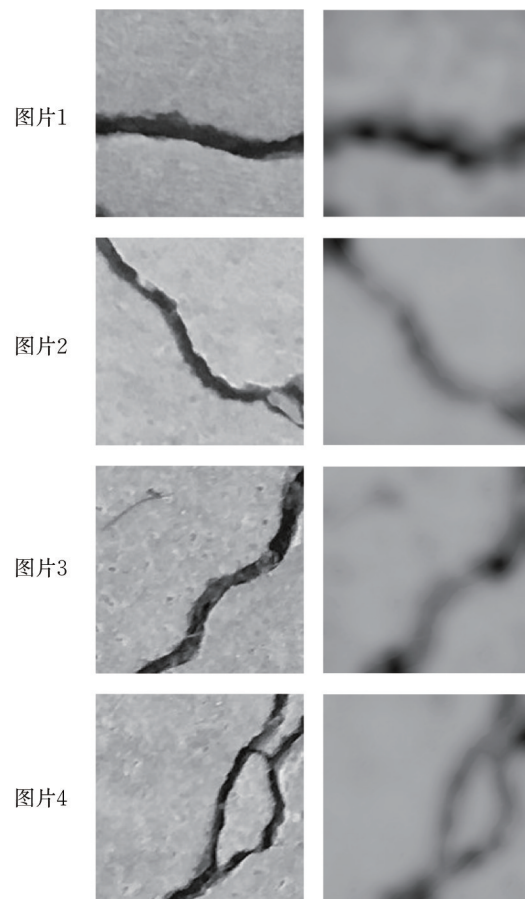


图 10 裂缝可视化分析结果

Fig.10 Visualization analysis results of cracks

4 结语

本文提出了一种无监督混凝土表面裂缝的视觉

检测方法,采用以无裂缝混凝土图片训练的VQ-VAE-2重建检测图片,并计算原图与检测图片间的重建误差,根据重建误差的大小进行裂缝检测。为了增大2类图片重建误差的差异,将重建图片与原图分割成若干图块,取相应图块间均方误差的最大值作为检测图片的重建误差。在验证实验中,本文算法的F1分数为0.957,高于MemAE和CAE,在其他评估指标上也表现最优。结果表明,在没有裂缝混凝土图片作为训练数据的情况下,该方法在混凝土表面裂缝检测任务中性能良好。

作者贡献声明:

刘 超:算法提出,论文撰写指导。

吴纪曙:具体算法编写,论文撰写。

参考文献:

- [1] YOU Y, ZHANG Z, HSIEH C, *et al.* ImageNet training in minutes[C]//Proceedings of the 47th International Conference on Parallel Processing. New York: Association for Computing Machinery, 2018: 1-10.
- [2] SALINI R, XU B, PAPLAUSKAS P. Pavement distress detection with PICUCHA methodology for area-scan cameras and dark images [J]. *Civil Engineering Journal*, 2017, 26(1): 34.
- [3] 梁雪慧,程云泽,张瑞杰,等.基于卷积神经网络的桥梁裂缝识别和测量方法[J].*计算机应用*, 2020, 40(4): 1056.
LIANG Xuehui, CHENG Yunze, ZHANG Ruijie, *et al.* Bridge crack classification and measurement method based on deep convolutional neural network [J]. *Journal of Computer Applications*, 2020, 40(4): 1056.
- [4] ZHUANG X, WANG D, PENG B, *et al.* Life-cycle civil engineering: innovation, theory and practice [M]. London: CRC Press, 2021.
- [5] 毛莺池,唐江红,王静,等.基于Faster R-CNN的多任务增强裂缝图像检测方法[J].*智能系统学报*, 2021, 16(2): 286.
MAO Yingchi, TANG Jianghong, WANG Jing, *et al.* Multi-task enhanced dam crack image detection based on Faster R-CNN[J]. *CAAI Transactions on Intelligent Systems*, 2021, 16(2): 286.
- [6] 侯越,陈逸涵,顾兴宇,等.基于卷积自编码的沥青路面目标与裂缝智能识别[J].*中国公路学报*, 2020, 33(10): 288.
HOU Yue, CHEN Yihan, GU Xingyu, *et al.* Automatic identification of pavement objects and cracks using the convolutional auto-encoder [J]. *China Journal of Highway and Transport*, 2020, 33(10): 288.
- [7] XU B, LIU C. Pavement crack detection algorithm based on generative adversarial network and convolutional neural network under small samples[J]. *Measurement*, 2022, 196: 111219.
- [8] 黄超,胡志军,徐勇,等.基于视觉的车辆异常行为检测综述[J].*模式识别与人工智能*, 2020, 33(3):234.
HUANG Chao, HU Zhijun, XU Yong, *et al.* Vision-based abnormal vehicle behavior detection: a survey [J]. *Pattern Recognition and Artificial Intelligence*, 2020, 33(3):234.
- [9] AN J, CHO S. Variational autoencoder based anomaly detection using reconstruction probability [J]. *Special Lecture on IE*, 2015, 2(1):1.
- [10] HASAN M, CHOI J, NENUMANN J, *et al.* Learning temporal regularity in video sequences [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 733-742.
- [11] ZHOU C, PAFFENROTH R C. Anomaly detection with robust deep autoencoders [C]//Proceedings of the 23rd ACM SIGKDD, International Conference on Knowledge Discovery and Data Mining. New York: Association for Computing Machinery, 2017: 665-674.
- [12] GONG D, LIU L, LE V, *et al.* Memorizing normality to detect anomaly: memory-augmented deep autoencoder for unsupervised anomaly detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019: 1705-1714.
- [13] RAZAVI A, OORD A, VINVALS O. Generating diverse high-fidelity images with VQ-VAE-2[C]//Advances in Neural Information Processing Systems. New York: Curran Associates, 2019:14837-14847.
- [14] OORD A, VINVALS O, KAVUKCUOGLU K. Neural discrete representation learning [C]//Advances in Neural Information Processing Systems. New York: Curran Associates, 2017: 6309-6318.
- [15] ÖZGENEL Ç, SORGUÇ A. Performance comparison of pretrained convolutional neural networks on crack detection in buildings[C]//Proceedings of the International Symposium on Automation and Robotics in Construction. Berlin: IAARC Publications, 2018: 1-8.
- [16] BERGMANN P, LWE S, FAUSER M, *et al.* Improving unsupervised defect segmentation by applying structural similarity to autoencoders [C]//Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications. Setúbal: Science and Technology Publications, 2019: 1-8.