

DOI:10.19344/j.cnki.issn1671-5276.2025.06.031

基于 DPAM 特征提取的 GAN 合成水管泄漏图像检测分析

范海蛟¹, 罗政星², 李峰³

(1. 宁夏宁东水务有限责任公司, 宁夏回族自治区 银川 750409;

2. 宁夏易通联创科技有限公司, 宁夏回族自治区 银川 750002;

3. 宁夏大学 信息工程学院, 宁夏回族自治区 银川 750021)

摘 要:针对生成对抗网络(GAN)合成的逼真图像存在难以识别的问题,提出一种基于 DPAM 特征提取的 GAN 合成水管泄漏图像检测方法,融合集成分类器来实现准确且高效的 GAN 合成图像识别功能。研究结果表明:小样本训练下展现出卓越的检测性能,具备较高稳定性;在单类样本数量充足时能够有效区分自然图像与 GAN 生成图像;能够有效辨识出 StyleGAN 生成的伪图像,对于其他 GAN 类图像展现出了极高识别精确度,识别准确度维持在不低于 99.88% 的水平。由此证明融合多通道特征的策略是行之有效的。该研究有助于提高水管泄漏的识别能力,也可以拓展到其它图像识别领域。

关键词:水管泄漏;图像检测;图像取证;色彩通道;特征融合;生成对抗网络;集成分类器

中图分类号:TP391.4 **文献标志码:**A **文章编号:**1671-5276(2025)06-0169-04

GAN Synthetic Water Pipe Leakage Image Detection Analysis Based on DPAM Feature Extraction

FAN Haijiao¹, LUO Zhengxing², LI Feng³

(1. Ningxia Ningdong Water Co., Ltd., Yinchuan 750409, China;

2. Ningxia E-Link Technology Co., Ltd., Yinchuan 750002, China;

3. School of Information Engineering, Ningxia University, Yinchuan 750021, China)

Abstract:For the difficulty in recognizing realistic images synthesized by generative adversative network (GAN), a water pipe leakage image detection method based on DPAM feature extraction is proposed, which fuses integrated classifier to realize accurate and efficient GAN image recognition. The results show that the proposed method, with small sample training, has excellent detection performance and high stability. With sufficient number of single samples, the natural image can be effectively distinguished from GAN generated image. And the method can effectively identify the pseudo-images generated by StyleGAN, and show high recognition accuracy for other GAN images, with recognition accuracy higher than 99.88%, which proves the effectiveness of the strategy of fusing multi-channel feature. The research is helpful to improve the recognition ability of water pipe leakage, and can be extended to other image recognition fields as well.

Keywords:water pipe leakage; image detection; image forensics; color channel; feature fusion; generate adversarial network; integrated classifier

0 引言

随着当前网络信息技术的快速进步,许多新型数字处理技术逐步涌现,该技术在输水管道检测装备方面得到了迅速推广,极大推动了数字图像分析方法在人们日常生活中及工业生产过程的应用拓展,由此成为网络信息传递的关键手段,同时也成为人们信息交流的核心技术^[1-2]。根据相关文献报道可知,这类技术需要基于繁杂数学理论,对技术能力有较高要求,通常需要专业技术人

员参与,其普及率并不高^[3-5]。现阶段,大部分生成对抗网络(GAN)图像检测技术都是建立在深度神经网络(DNN)上,这些方法往往需依赖大量训练样本^[6]。

尽管当前以 DNN 进行特征提取的技术不断涌现,但在某些特定问题处理方面,例如区分真实图像和 GAN 图像数据时,传统基于手工特征提取方法并不一定逊色于新 DNN 方法^[7-8]。DNN 模型尽管可实现较高识别精度,但因其存在结构复杂、缺乏解释性,以及对计算资源占用率高的问

基金项目:宁夏自然科学创新基金项目(2023SQ20012)

第一作者简介:范海蛟(1969—),男,宁夏回族自治区吴忠人,高级工程师,本科,研究方向为水利水电工程, fanhaijiao123132@163.com。

题,在应用过程中可能难以获得广泛应用。因此,结合传统机器学习算法的低维度特征提取过程,设计一种“轻量级”机器学习模型显得尤为重要^[9-10]。这种方法可以达到快速计算的效果,并且模型结构简单,易于理解和部署,同时还能实现较高识别精度。特别是在训练样本数量有限或只有单一数据集的情况下,使用手工特征提取方法被认为可以有效区分真实图像与 GAN 生成图像^[11-14]的一种手段。本研究设计了一种基于差分像素邻接数组(DPAM)特征提取的 GAN 合成水管泄漏图像检测方法。

1 检测框架设计

1.1 色彩通道转换

自然景物图片数据与 GAN 生成的图像在成源方面存在差异,由此导致它们在色彩通道方面具有不同表现^[15]。针对这种差异,可将这些色彩差异转化为辨识度更高的特征向量,并以此区分生成器和鉴别器两种并行思路。GAN 生成图像的流程如图 1 所示。对于(红绿蓝)三原色彩体系,将色相、亮度以及饱和度进行整合后,这种方式未能对色彩与亮度元素进行细致区分^[16]。在多个色彩通道上对自然图片和 GAN 生成数据开展深入分析,可以筛选出最合适的通道,以进行有效的特征提取。

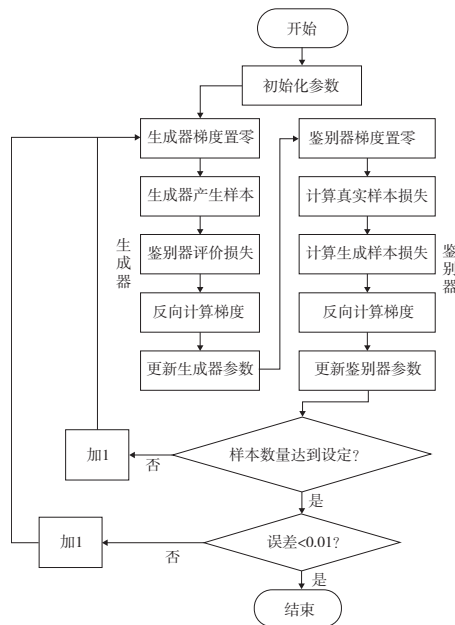


图1 本文方法流程

本研究分别对红绿蓝三原色彩空间制定了评估准则,由此判断哪些色彩通道在辨识 GAN 生成

图像时更高效。针对每幅图像各通道,计算相邻像素相关性。假设图像尺寸 $a \times b$,由此设定下述评估准则 L 为

$$L = \frac{\sum_{m=1}^{a-1} \sum_{n=1}^b (F_{m+1,n} - \bar{F})(F_{m,n} - \bar{F})}{\sqrt{\sum_{m=1}^{a-1} \sum_{n=1}^b (F_{m+1,n} - \bar{F})^2 \sum_{m=1}^{a-1} \sum_{n=1}^b (F_{m,n} - \bar{F})^2}} \quad (1)$$

式中: a 为图像高度, b 为宽度, a, b 都设定在 512; F 表示像素数组; $\bar{F}$ 为 F 的均值。根据以上条件可以推断, L 增大说明邻近像素间关联度获得了提升。本研究仅分析了竖直向像素关联性,但相似结论也适用于其他方向的分析需求。

1.2 差分像素邻接数组特征提取

随着像素个数增加,像素直方图中包含的数据量也保持指数级别上升的趋势,这使得像素层面搜索差异变得更复杂^[17]。考虑到自然图像像素值通常表现出强烈空间相关性,导致相邻像素非常接近,因此相邻像素 $(F_{m,n}, F_{m,n+1})$ 同时出现概率 $P(F_{m,n}, F_{m,n+1})$ 呈现出中心对称的分布形态。基于以上结果,像素间差值 $F_{m,n} - F_{m,n+1}$ 可以看作是不受像素 $F_{m,n}$ 影响。通过 DPAM 来深刻揭示自然图像与 GAN 生成图像之间的区别,并依此构建了以下算法步骤。

先对 8 个方向的差值参数开展计算:

$$T_{m,n}^{\rightarrow} = F_{m,n} - F_{m,n+1} \quad (2)$$

式中:满足 $m \in \{i | 1 \leq i \leq a\}, n \in \{i | 1 \leq i \leq b-1\}$,上标 $\rightarrow$ 为计算方向; T 为差值数组。对其他方向进行计算的过程与水平方向相近,因此本文只设置水平方向算法。

利用 2 阶马尔可夫链建立相邻像素差分数组转移概率模型,获得 2 阶 DPAM 特征^[18]:

$$K_{x,y}^{\rightarrow} = P(T_{m,n+1}^{\rightarrow} = x | T_{m,n}^{\rightarrow} = y) \quad (3)$$

式中: $x, y \in \{i | -L \leq i \leq L\}$; $P(T_{m,n+1}^{\rightarrow} = x | T_{m,n}^{\rightarrow} = y)$ 是在 $T_{m,n}^{\rightarrow} = y$ 条件下, $T_{m,n+1}^{\rightarrow} = x$ 的概率。当 $P(T_{m,n+1}^{\rightarrow} = x) = 0$,可以得到 $K_{x,y}^{\rightarrow} = P(T_{m,n+1}^{\rightarrow} = x | T_{m,n}^{\rightarrow} = y) = 0$ 。

建立水平 2 阶特征算式如下:

$$K_{x,y,z}^{\rightarrow} = P(T_{m,n+2}^{\rightarrow} = x | T_{m,n+1}^{\rightarrow} = y, T_{m,n}^{\rightarrow} = z) \quad (4)$$

满足

$x, y, z \in \{i | -L \leq i \leq L\}$; $P(T_{m,n+2}^{\rightarrow} = x | T_{m,n+1}^{\rightarrow} = y, T_{m,n}^{\rightarrow} = z)$ 为 $T_{m,n+1}^{\rightarrow} = y$ 与 $T_{m,n}^{\rightarrow} = z$ 时, $T_{m,n+2}^{\rightarrow} = x$ 的概率。当 $P(T_{m,n+1}^{\rightarrow} = y, T_{m,n}^{\rightarrow} = z) = 0$ 时, $K_{x,y,z}^{\rightarrow} = P(T_{m,n+2}^{\rightarrow} = x | T_{m,n+1}^{\rightarrow} = y, T_{m,n}^{\rightarrow} = z) = 0$ 。

虽然可以通过 GAN 技术输出逼真图像,但两类图像是以不同方式形成,特别是在色彩波段和图片质感方面展现出独特性能。通过自然图像对称的特点,DPAM 特性缩减了特征向量维度,同时借助 2 阶马尔可夫链来分析图片中邻近像素间变异情况,进而对该变异参数进行量化。该技术可以有效辨识出与模型不相符的差值。

2 实验方案

本实验设置了一个专为模型测试建立的大

型图像数据库,该数据库包含了部分真实自然环境下的图像参数(将部分图像展示在图 2 中)。在初始阶段,本文从 CelebA-HQ 数据集中采集了 100 张尺寸为 1 024×1 024 的天然水管渗漏图像,同时选择了 100 张来自 CelebA-HQ 数据库的自然图像组成训练样本,将其导入 GAN 算法中进行训练。相关的实验数据设置如表 1 所示。在评估模型效果方面,以准确率 ACC 组成评价标准。



图 2 正样本示例图(来自 CelebA-HQ 的真实自然图像)

表 1 实验数据配置

数据项目	配置内容
色彩通道	R, G, B
图像尺寸/px	1 024×1 024
图像格式	JPEG
样本对数量	10 000

表 2 不同小样本数的准确率 单位:%

训练样本	100	50	10	5	2
本文	98.88± 0.02	98.88± 0.03	99.97± 0.03	99.94± 0.06	99.37± 0.34
SRM	99.99± 0.02	98.88± 0.02	99.97± 0.04	99.93± 0.07	99.56± 2.41
ccJRM	99.87± 0.03	99.97± 0.04	99.77± 0.17	99.43± 0.72	91.48± 9.65

3 实验结果分析

3.1 小样本训练及检测结果

本节对降低训练样本规模条件下本研究方法的验证结果进行了测试。因为神经网络对训练数据量需求较高,本次只选择 SRM、ccJRM 特征整合至分类器中再进行比较分析。表 2 详细给出了在不同样本数量的情况下图像样本验证结果。即便在训练数据极为有限下,本方法依然展现出卓越检测性能。在本研究检测架构中,本研究特征与 SRM 特征的性能几乎持平,在样本数量仅为 2 时,其准确率达到 99% 以上,且标准差较小,而 ccJRM 准确率在同样样本数量下降低至 91.48%。特别是本研究仅包含 2 对训练样本时, σ 值低至 0.34%,由此反映出本方法具备较高稳定性。

3.2 单类样本训练及检测结果

在当前条件下若采用二元分类器,则分类过程倾向于数量较多的样本类别。本文旨在研究单一类别样本的训练,并对分类器性能进行检测对比。在现实应用条件下,获取真实图像数据通常比获得由 GAN 生成的图像更简单,并且 GAN 生成的图像来源难以追溯。鉴于以上结果,本研究利用真实图像特征进行 SVM 训练,采用网格搜索法寻找最佳参数。根据表 3 分类结果,本研究提出的方法可以在样本仅有 5 000 个的情况下依旧展现出优异性能,其准确率能够达到 97.25%。该结果表明,在单类样本数量充足时,本文提出的特征提取方法能够有效区分自然图像与 GAN 生成图像。

表 3 单类训练样本的准确率 单位: %

训练个数	5 000	500	50	10	5
本文	97.25	96.36	93.28	73.16	59.08
SRM	50.00	50.00	50.00	50.00	50.00
ccJRM	50.00	50.00	50.00	50.00	50.00

3.3 多种 GAN 模型检测结果

本文对检测器实施深入效能评估与对比,各项数据如表 4 所示。根据实验数据可以发现,本研究构建的检测算法不仅能够有效辨识出 GAN 生成的伪图像,对于其他 GAN 类图像也展现出了极高识别精确度,即便是在单一通道特征情况下,识别准确度也能维持在不低于 99.88% 的水平。特别是在处理 GAN 模型时,如果使用 FFHQ 数据集作为基准,能够实现不低于 99.51% 的识别准确率,而采用 LSUN 数据集时,准确率同样高达 99.97%。综合来看,三通道检测性能明显优于独立通道,由此证明融合多通道特征的策略是行之有效的。此外,本方法同样适用于检测 GAN 生成的漏水图像以及其它各类图像,结果显示其在识别 GAN 合成图像方面具备广泛适用性。

表 4 多种 GAN 模型生成图像识别的准确率 单位: %

色彩通道	R	G	B	R+G+B
GAN(CelebA-HQ)	99.88	99.99	99.88	99.99
GAN2(FFHQ)	99.76	99.51	99.69	99.97
GAN2(LSUN)	99.85	99.81	99.64	99.97
PGGAN(CelebA-HQ)	99.94	99.94	99.97	99.97

4 结语

本文开展基于 DPAM 特征提取的 GAN 合成水管泄漏图像检测分析,取得如下有益结果。

1) 小样本训练下本方法依然展现出卓越检测性能,样本数量仅为 2 时准确率依然达到 99% 以上,标准差低至 0.34%,具备较高稳定性。

2) 本方法在样本仅有 5 000 个下依旧展现出优异性能,准确率达到 97.25% 以上,在单类样本数量充足时本方法能够有效区分自然图像与 GAN 生成图像。

3) 本方法能够有效辨识出 GAN 生成的伪图像,对于其他 GAN 类图像也展现出了极高识别精确度,识别准确度维持在不低于 99.88% 水平,证明融合多通道特征的策略是行之有效的。

参考文献:

[1] 郭新蕾,郑飞飞,马朝猛,等. 输水管道泄漏的探地雷达检测原理和现象机理分析[J]. 水利学报,2021,

52(7):781-792.

- [2] 彭道刚,潘俊臻,王丹豪,等. 基于改进 YOLOv5 的电厂管道油液泄漏检测[J]. 电子测量与仪器学报,2022,36(12):200-209.
- [3] 张万远,卞红雨,周天,等. 基于多波束测深声呐的海底气体泄漏实时报警系统[J]. 声学技术,2022,41(6):813-820.
- [4] 王艺博,张珂,孔英会,等. 人脸年龄合成的生成对抗网络方法综述[J]. 中国图象图形学报,2023,28(10):3004-3024.
- [5] 但志平,方帅领,孙航,等. 基于双判别器异构 CycleGAN 框架下多阶通道注意力校准的室外图像去雾[J]. 电子学报,2023,51(9):2558-2571.
- [6] 刘雅蕾,和红杰,陈帆,等. 基于水印神经网络的可溯源 DNN 模型保护方法[J]. 应用科学学报,2023,41(2):183-196.
- [7] 熊若鑫,宋元斌,王宇轩. 基于 DNN 的作业姿势评估方法及应用[J]. 中国安全科学学报,2018,28(5):105-110.
- [8] 陈凯健,李二强,周漾. 基于相对坐标控制的非均匀纹理合成方法[J]. 计算机辅助设计与图形学学报,2023,35(2):284-292.
- [9] FRIDRICH J, KODOVSKY J. Rich models for steganalysis of digital images[J]. IEEE Transactions on Information Forensics and Security, 2012, 7(3):868-882.
- [10] KODOVSKY J, FRIDRICH J. Steganalysis of JPEG images using rich models[C]//Media Watermarking, Security, and Forensics 2012. Burlingame, California, USA. SPIE,2012:81-93.
- [11] 张润,邱志斌,童志鹏,等. 基于红外图像目标检测与偏斜矫正的变压器套管发热缺陷检测方法[J]. 南方电网技术,2024,18(9):59-68.
- [12] 李新良,蒋庆,党胜茂,等. 基于线阵相机的微通道扁管表面缺陷检测系统[J]. 现代制造工程,2022(3):114-121.
- [13] 刘昱,陈永祺,王瀚,等. 基于通道协同与色彩校正的 RAW 数据到 RGB 图像重建方法[J]. 天津大学学报(自然科学与工程技术版),2025,58(4):395-405.
- [14] 乔通,陈彧星,谢世闯,等. 多色彩通道特征融合的 GAN 合成图像检测方法[J]. 电子学报,2024,52(3):924-936.
- [15] 栗凤永,魏璐,曾玮姝. 融合神经网络变换和通道置乱的彩色图像加密[J]. 计算机工程与设计,2023,44(7):2118-2124.
- [16] 周景春,卫晓靖,史金余. 基于蓝绿通道自适应色彩补偿的水下图像增强[J]. 电子与信息学报,2022,44(8):2932-2939.
- [17] 牛士铭,薛茹. 基于耦合的 Henon 和 Sine 映射及彩色图像加密[J]. 科学技术与工程,2024,24(16):6809-6820.
- [18] 马啸,姚刚,张峰,等. 基于像素差分神经网络的断层识别方法[J]. 地球物理学报,2023,66(4):1649-1663.

收稿日期:2024-08-12