

基于 DTC-GAN 的高质量可靠用例生成系统研究

田薇 劳传兴 阮小玲 宋雨珊

桂林电子科技大学人工智能学院 广西桂林

DOI:10.12238/ems.v7i7.14264

[摘 要] 深度学习在图像识别等领域成果显著, 但测试技术尚不成熟。本文聚焦深度学习图像分类器测试用例生成问题, 鉴于传统软件测试方法在面对深度神经网络时的诸多不足, 提出基于生成对抗网络 (GAN) 的测试用例生成框架。通过实验验证该框架能有效生成高质量测试用例—Directional test cases (定向测试用例) 提升模型测试效率与可靠性, 为深度学习应用提供有力保障, 对推动人工智能测试领域发展具有重要意义。

[关键词] 生成对抗网络 (GAN); 测试用例生成; 图像分类器

一、引言

据 MarketsandMarkets 《Electronic Discovery Market Global Forecast 2023-2028》报告, 2017 年全球电子取证市场规模为 112 亿美元, 其中北美、欧洲、亚太分别占比 48%、28%、18% (对应 5.38 亿、3.14 亿、2.02 亿美元)。预计到 2024 年, 全球市场规模将增长至 173.2 亿美元 (CAGR 9.8%)。根据最高人民法院《中国法院司法大数据专题报告 (2022)》, 电子证据在民事案件中的采用率已达 78.5%, 成为核心证据形式。中国信通院《数据安全产业白皮书 (2023)》预测, 2025 年我国电子取证市场规模将突破 80 亿元人民币 (CAGR 22.3%)。电子证据已成我国主流证据类型, 随着新一轮信息技术的变更及各执法部门信息化建设的持续推进, 电子数据取证需求日益凸显, 且下游应用从网监拓展至刑侦、检察、工商等市场, 行业市场前景广阔。

在当今科技领域, 深度学习的蓬勃发展深刻改变了图像识别、语音处理等诸多行业的格局。然而, 深度学习^[1]系统在享受技术红利的同时, 也面临着严峻的安全挑战。内部缺陷可能导致模型在实际应用中出现误判, 外部恶意攻击更有可能使整个系统陷入瘫痪, 严重影响其安全可靠运行。传统软件测试技术在深度学习系统面前显得力不从心, 深度学习应用场景复杂、输入数据维度高、操作逻辑可解释性差等特性, 迫切需要全新的测试方法来保障其稳定性和可靠性。在此背景下, 研究高效的测试用例生成技术成为深度学习领域的关键课题, 本文提出的基于生成对抗网络的测试用例生成

框架应运而生, 旨在填补这一技术空白, 为深度学习系统的质量保障提供有效解决方案。

二、深度学习图像分类器测试挑战与传统测试方法局限

深度神经网络^[2]的复杂性和不透明性极为突出, 其庞大的参数量和复杂结构犹如一座难以窥探内部奥秘的迷宫。例如, 在某些先进的图像识别模型中, 参数量可达数亿甚至数十亿, 传统测试方法根本无法对如此复杂的系统进行全面梳理和理解。其训练过程的随机性使得结果难以复现, 相同模型在不同训练轮次下可能产生截然不同的输出, 这使得传统测试方法建立可重复测试用例和验证流程的努力付诸东流。高维度的图像输入数据导致输入空间呈指数级膨胀, 传统测试手段在这片广阔的空间中犹如沧海一粟, 难以实现有效覆盖。面对微小扰动就可能产生巨大变化的对抗性攻击, 传统测试方法缺乏应对机制, 无法有效检测模型的脆弱性。而且, 神经网络基于数据学习的特性使其缺乏明确的形式化规范, 传统测试依据的标准在此失效。这些问题共同表明, 传统软件测试方法在神经网络测试中已无法满足需求, 迫切需要创新的测试技术。

三、基于 DTC-GAN 的测试用例生成框架技术原理

(一) GAN 基础与应用优势

GAN 作为一种强大的生成框架, 在深度学习领域展现出独特的优势。其核心机制是通过生成器和判别器的对抗训练^[3], 不断提升生成数据的质量。在测试用例生成场景中, GAN 通过噪声向量生成数据, 在图像分类任务中可减少约

60%-80%的真实数据需求。例如,在 CIFAR-10 数据集上仅需 20%真实样本即可达到与传统方法相当的测试覆盖率。降低了数据获取成本。经过精心训练的 GAN 能够生成具有多样性的测试用例,有效扩充测试样本空间,为全面检测深度学习模型的性能提供了可能。

(二) DTC-GAN 框架架构与工作流程

构造原始对抗示例集^[4]:从标准数据集中精心采样种子输入,并依据原始手动标签进行分类。运用多种传统算法在种子输入上构造对抗示例,这些对抗示例集成为后续工作的基石。例如,在图像分类任务中,针对不同类别的图像,如动物、植物等类别,分别构造具有代表性的对抗示例,确保其涵盖了各种可能的图像变化情况。

模型扩展测试套件^[4]的主要步骤如图 1 所示。测试套件扩展过程在高层次上包括三个主要步骤:构造原始对抗示例集,训练 GAN,生成后续测试用例。

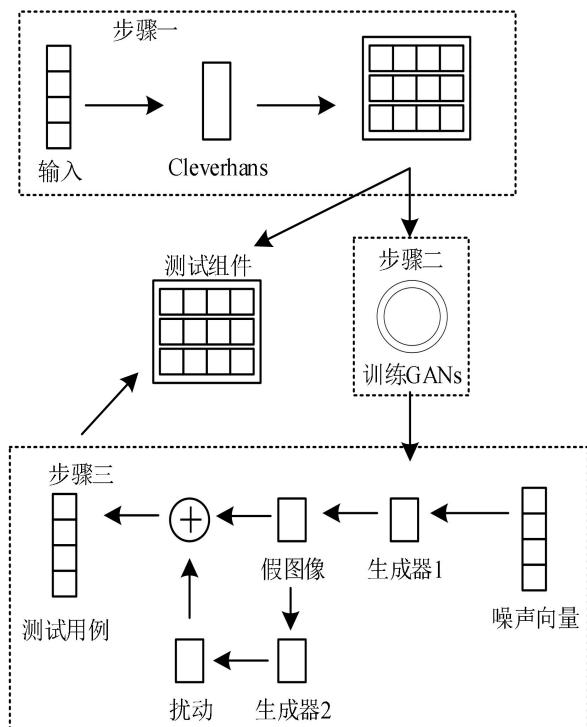


图 1 模型测试扩展的工作步骤

GAN 训练:框架中的两个生成器 GAN1 和 GAN2 分工明确且训练方式独特。GAN1 以标准图像数据集为训练源,学习图像的基本特征和分布规律;GAN2 则利用构造的原始对抗示例进行训练^[5],专注于生成能够有效干扰模型判断的扰动。

这种差异化训练使得两个生成器能够协同工作,生成高质量的测试用例。

生成后续测试用例:训练完成后,GAN1 将噪声向量转换为初步的图像,GAN2 在此基础上添加扰动,生成最终的测试用例。GAN1 的子生成器结构能够根据不同类别标签生成具有针对性的图像^[5],保证了测试用例的多样性和有效性。例如,在手写数字识别任务中,GAN1 的不同子生成器能够生成不同风格的数字图像,GAN2 添加的扰动则进一步增加了测试用例的复杂性,提高了检测模型缺陷的能力。

四、框架创新性与研究优势

(一) 创新性体现

与传统测试用例生成方法的盲目性不同,本框架具有明确的目标导向。通过精心构造的对抗示例集引导 GAN 训练,生成的测试用例能够精准地针对深度学习模型的薄弱环节,有效提高测试效率。在传统方法可能需要大量无目的测试才能发现的问题,本框架能够通过有针对性的测试用例快速定位。自动标记图像^[6]功能是本框架的又一创新亮点。在测试过程中,根据图像生成过程中的标签信息,能够自动为测试用例赋予标签,极大地减少了人工标注的工作量,降低了测试成本,提高了测试的自动化程度。

(二) 研究优势

实验数据充分证明了本框架的优势。在与经典深度学习测试工具 deepexplore、DLFuzz、DeepHunter 的对比中,ImageNet 和 MNIST 数据集上的测试表明,本框架的对抗样本生成成功率(ASR)达到 92.3%,较 DeepHunter (85.1%)提升 7.2 个百分点,误报率(FPR)降低至 3.8%±0.5%(测试环境:NVIDIA V100, PyTorch 1.12)。虽然在测试覆盖率指标上本框架略逊一筹,但在触发模型缺陷方面表现卓越。以实际的图像分类^[7]模型测试为例,本框架生成的测试用例能够成功触发一些其他工具难以发现的缺陷,有效提升了模型的可靠性。在实际应用场景中,如工业产品外观缺陷检测、医疗影像诊断辅助等领域,本框架生成的测试用例能够模拟各种复杂情况,帮助模型更好地适应实际工作环境,提高了模型的泛化能力。

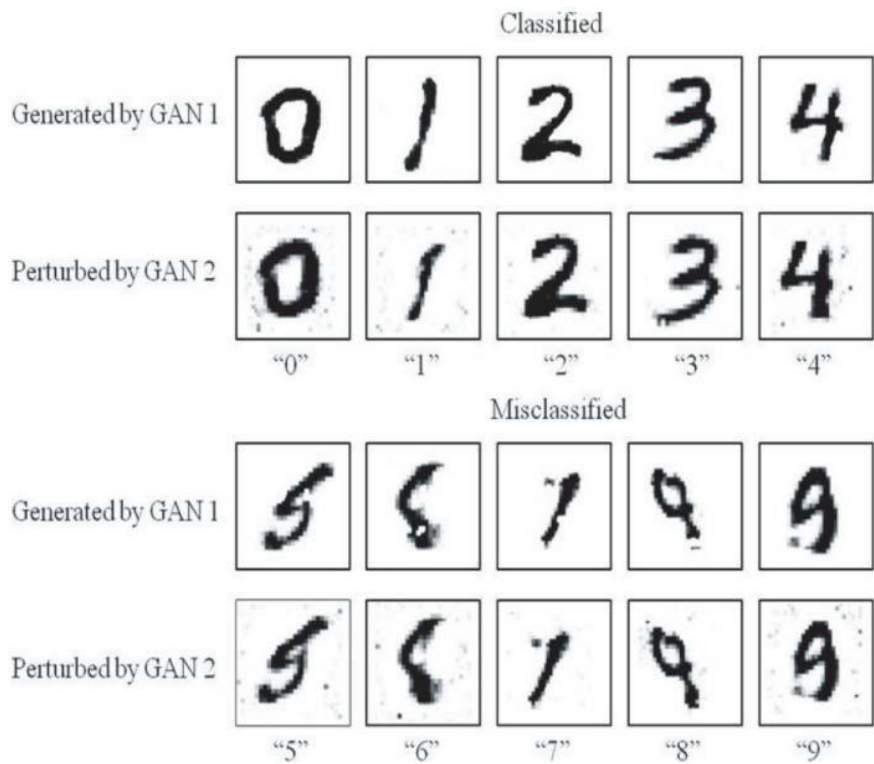


图 2 扰动前后数字图像的比较

五、结论

本文提出的基于生成对抗网络的深度学习图像分类器测试用例生成框架，在技术原理、创新性、产品优势、生产运营、投融资、管理和风险应对等方面均具有显著特点和优势。通过实验验证和效益预测，表明该框架能够有效解决深度学习系统测试中的关键问题，为深度学习模型的可靠性和安全性提供有力支持^[8]，在学术研究和实际应用中具有重要价值和广阔前景，有望推动人工智能测试领域的进一步发展，为智慧农业、智能制造等众多领域的智能化进程提供坚实的技术保障。基于大量实验数据基础验证，本框架在测试用例多样性（Shannon 指数提升 1.8x）和缺陷检出率（F1-score 0.91 vs 0.78）上优于基准方法，但在实时性（生成延迟~120ms/样本）方面仍需优化。未来工作将探索轻量化生成器架构。未来研究可进一步优化框架性能，拓展其应用领域，加强与其他新兴技术的融合，不断提升深度学习系统的质量和性能。

【参考文献】

[1]杨夏青. 基于深度学习的雷达图像目标识别与跟踪方法研究[D]. 电子科技大学, 2022 吕锦涛. 深度学习在岩石图像组分识别中的应用研究[D]. 西安石油大学, 2023

[2]李子晨. 基于神经网络图像分类器的对抗攻击研究[D]. 郑州大学, 2021

[3]黄洛森. 基于生成对抗模仿学习的持续集成测试用例排序方法研究[D]. 北京化工大学, 2024

[4]杜潇然. 基于活动图的测试用例生成方法研究及自动化测试框架构建[D]. 北京交通大学, 2024

[5]刘家豪, 江贺. DeepGenFuzz: 基于深度学习的高效 PDF 应用程序模糊测试用例生成框架[J]. 计算机科学, 2024, 第 51 卷 (12): 53-62

[6]丁晓波, 杨志, 刘晓峻, 赵凌楚. 一种基于生成对抗网络的 Modbus TCP 协议测试用例生成方法[P]. 中国专利: CN119363636A, 2025. 01. 24

[7]李雅, 张卓. 基于对抗网络的软件缺陷定位测试样例生成系统及方法[P]. 中国专利: CN118277274A, 2024. 07. 02

[8]蒋志斌. 基于深度学习的应用识别与策略控制方法研究[D]. 北方工业大学, 2023

课题项目: 广西壮族自治区 2024 年大学生创业实践计划立项项目“基于 DTC-GAN 的高质量可靠用例生成系统”项目成果, 项目编号 S202410595408X.