

中图分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2025)11-3451-14

论文引用格式: Zhang K, Sheng X, Xiao Y J, Yang J Y, Chen M J and Ren Z H. 2025. Stable diffusion model for few-shot generation of meter defect images. Journal of Image and Graphics, 30(11):3451-3464(张珂, 盛鑫, 肖扬杰, 杨济远, 陈美娟, 任泽华. 2025. 面向少样本表计缺陷图像生成的稳定扩散模型. 中国图象图形学报, 30(11):3451-3464)[DOI:10.11834/jig.240777]

面向少样本表计缺陷图像生成的稳定扩散模型

张珂^{1,2,3*}, 盛鑫¹, 肖扬杰¹, 杨济远¹, 陈美娟¹, 任泽华¹

1. 华北电力大学电子与通信工程系, 保定 071003; 2. 华北电力大学河北省电力物联网技术重点实验室, 保定 071003;
3. 电力物联智慧化技术河北省工程研究中心, 保定 071003

摘要: 目的 变电站缺陷检测对电力系统的安全稳定运行具有重要意义, 表计作为关键设备, 其缺陷检测对保障电力调度和运行监控的准确性尤为重要。目前表计缺陷相关数据极其稀缺, 成为实现高效检测的主要挑战。然而, 现有缺陷生成方法多依赖于对已有缺陷样本的简单变换或直接生成, 难以在少样本条件下生成多样化、高质量的缺陷图像, 限制了实际应用效果。为此, 本文提出一种面向少样本表计缺陷图像生成的稳定扩散模型, 旨在生成高质量、形式丰富的缺陷图像, 以满足实际场景需求。**方法** 首先, 针对现有模型生成图像与实际表计图像差距较大的问题, 通过微调方法将独特标识符与表计图像绑定, 实现表计知识嵌入, 提升生成图像与实际变电站表计的相似度; 其次, 为克服传统模型在生成多样性和缺陷样式控制上的不足, 设计裂纹特征建模方法, 通过逐像素融合线稿图与裂纹掩码, 并结合约束图限制建模区域, 生成具备几何约束的控制图像, 精准表达缺陷特征; 最后, 基于生成的控制图像, 利用超网络机制动态调整生成过程, 严格控制缺陷形状、位置及样式, 确保生成结果符合实际需求。**结果** 在真实巡检图像构建的表计数据集上进行实验, 结果表明所提方法在轮廓细节、缺陷样式和表计多样性方面表现更优, 弗雷歇特启动距离(Fréchet inception distance, FID)和启动分数(inception score, IS)指标分别达到 76.72 和 2.45。在下游检测任务中, 加入生成数据后, 检测精度提升 26.9%, mAP50(mean average precision)增加 19.1%, 充分验证了生成数据对提升检测性能的有效性。**结论** 所提方法在少样本条件下有效解决了现有生成模型在图像多样性不足和生成质量不稳定等问题, 显著提升了生成样本的实用价值。生成的高质量缺陷图像为电力巡检系统的高效、可靠运行提供了有力的数据支持, 展现了广泛的工业应用潜力。

关键词: 稳定扩散模型; 变电站表计; 缺陷图像生成; 裂纹特征建模; 缺陷检测

Stable diffusion model for few-shot generation of meter defect images

Zhang Ke^{1,2,3*}, Sheng Xin¹, Xiao Yangjie¹, Yang Jiyuan¹, Chen Meijuan¹, Ren Zehua¹

1. Department of Electronic and Communication Engineering, North China Electric Power University, Baoding 071003, China;

2. Hebei Key Laboratory of Power Internet of Things Technology, North China Electric Power University, Baoding 071003, China;

3. Hebei Engineering Research Center of Intelligent Technology for Power Internet of Things, Baoding 071003, China

Abstract: **Objective** Meters, as critical components of substations, are essential for maintaining power grid stability. However, prolonged exposure to harsh environments such as extreme weather and temperature fluctuations makes them

收稿日期: 2024-12-30; 修回日期: 2025-03-04; 预印本日期: 2025-03-11

* 通信作者: 张珂 zhangkeit@ncepu.edu.cn

基金项目: 国家自然科学基金项目(62076093, 61871182, 62206095); 中央高校基本科研业务费专项资金资助(2022MS078, 2023JG002, 2023JC006); 河北省自然科学基金项目(F2024502017)

Supported by: National Natural Science Foundation of China (62076093, 61871182, 62206095); Fundamental Research Funds for the Central Universities (2022MS078, 2023JG002, 2023JC006); Natural Science Foundation of Hebei Province, China (F2024502017)

prone to defects such as cracks and deformations. These issues can disrupt operations and threaten grid reliability. Early detection is vital to prevent cascading failures. While recent advancements in computer vision have improved defect identification, training high-performance models relies on large, accurately labeled datasets, which are costly and limited by the scarcity of real-world defect data for specialized equipment including meters. Generative data augmentation methods offer an effective and promising solution to this issue. Techniques such as generative adversarial networks and denoising diffusion probabilistic models have proven their capability to generate visually compelling images by training on large-scale datasets. These methods are widely used to supplement existing datasets, enhance data diversity, and improve model training efficiency. However, when applied to small-sample datasets of substation meter defects, these approaches face great challenges, such as contour distortion, insufficient texture details, and excessive similarity to original images. These issues degrade the quality of generated images, fail to capture the subtle characteristics of meter defects, and limit their usefulness in tasks such as defect detection and segmentation. To overcome these limitations, this study proposes a novel defect generation method based on the Stable Diffusion model, which is specifically designed for small-sample scenarios. By leveraging its capability to balance high-quality and diverse image generation, this approach addresses the weaknesses of existing methods through improving the fidelity and variability of generated defect images. The proposed method ensures better alignment with real-world applications and enhances the applicability of synthetic images in downstream detection and analysis tasks. Ultimately, it contributes to improved defect detection performance and increased reliability in industrial applications.

Method This study proposes a novel small-sample defect generation method for substation meters based on a diffusion model. This method addresses the limitations of existing approaches in capturing structural and defect-specific features while achieving high-quality image generation to substantially enhance downstream applications. First, the pre-trained Stable Diffusion model was fine-tuned using a meter knowledge embedding strategy. This process effectively integrated the structural characteristics of substation meters and defect features into the model weights, which enhanced the capability of the model to comprehend meter patterns and improved the accuracy of key feature representation and reconstruction. Second, a crack feature modeling module was developed. This module utilized a structured preprocessing approach to process normal meter images and seamlessly integrated them with existing defect masks, which generated control images with precise geometric and spatial constraints. The module effectively delineated the spatial distribution of defects, which ensured accuracy and consistency in defect localization, as well as provided reliable conditional guidance for subsequent generation processes. Finally, an innovative hypernetwork-based conditional generation mechanism was introduced. While maintaining the diversity of generated images, this mechanism achieved precise manipulation of defect shapes, positions, and other characteristics. By dynamically adjusting model weights and refining input conditions, the hypernetwork effectively ensured local constraints and global coherence during the generation process. It enabled precise control over defect generation while moderately reducing strict constraints on other details, which provided the model creative flexibility to balance high-quality and diverse image generation.

Result Comprehensive experimental validation was conducted on the constructed substation meter dataset, which demonstrated that the proposed method can generate high-quality images with diverse and precise defect characteristics. The resulting images closely aligned with real-world application scenarios. The introduction of synthetic data significantly improved the performance of the model in downstream defect detection tasks. Notably, when 40% synthetic data were added to the training set, model precision increased by 26.9%, and mAP50 improved by 19.1%, which further verified the effectiveness of the proposed method in enhancing detection accuracy and robustness. Moreover, comparative experiments with advanced mainstream methods highlighted the superiority of the proposed approach. Fréchet inception distance (FID) and inception score (IS) were used as evaluation metrics to measure the similarity between generated and real images and the diversity of generated images, respectively. A lower FID score indicates higher generation quality, which reflects a smaller gap between the distributions of generated and real images. Meanwhile, a higher IS score demonstrates better clarity and diversity of the generated images. Experimental results show that the proposed method achieved the best performance in FID and IS metrics, with scores of 76.72 and 2.45, respectively, which significantly surpassed other mainstream methods.

Conclusion This study proposes a small-sample generation method based on the Stable Diffusion model, which focuses on the generation of substation meter defect images. Experimental results demonstrate that the proposed method effectively addresses issues associated with existing

generation models, such as poor quality and high redundancy of images generated from small-sample specialized datasets.

By producing high-quality defect images, the method significantly enhances the accuracy and robustness of downstream defect detection tasks. Thus, it provides a solid and reliable technical foundation for the stable operation of power systems.

Key words: Stable Diffusion; substation meter; defect image generation; crack feature modeling; defect detection

0 引言

变电站是电网的重要组成部分,其设备的安全稳定运行直接影响电网整体的可靠性(谢林枫等, 2019)。表计作为关键设备之一,是电力调度与运行监控的基础工具。然而,由于长期暴露于复杂的外部环境中,表计容易受到日晒、狂风、暴雨和暴雪等极端天气的影响,从而产生裂纹、碎裂等物理缺陷(Siddiqui和Park, 2020)。这些缺陷不仅可能导致表计功能失效,无法正常监测或记录数据,还可能引发设备损坏甚至系统性隐患,严重威胁电网的安全性和稳定性。

随着计算机视觉技术的发展,电力视觉在设备检测与故障识别领域取得了显著成果(赵振兵等, 2021; 戚银城等, 2021)。裂纹缺陷作为工业设备中常见且重要的故障类型,其检测与建模在工业场景中具有广泛应用价值。然而,视觉模型的高效训练通常依赖大规模且精确标注的数据集,而此类数据集的构建成本高昂、周期漫长,同时受制于数据稀缺性,成为进一步推广应用的瓶颈问题。

为缓解数据稀缺问题,数据增强技术成为一种有效手段。传统的数据增强方法通过几何变换(如翻转、旋转、裁剪、缩放)或添加噪声等操作扩展数据集,从而提高模型的鲁棒性(朱晓慧等, 2021; Royle等, 2007)。这些方法虽然适用于多数图像类型和深度学习模型训练,但通常忽略了图像的具体语义内容,在处理复杂场景或细粒度任务时表现有限,难以有效提升数据分布的多样性和复杂性。

近年来,生成对抗网络(generative adversarial network, GAN)和去噪扩散概率模型(denoising diffusion probabilistic model, DDPM)逐渐成为数据增强领域的核心技术(Goodfellow等, 2014; Ho等, 2020)。GAN通过对抗训练生成高质量图像,在样本量有限时显著提升样本多样性。Arjovsky和Bottou(2017)分析了JS(Jensen-Shannon)散度在GAN训练中指导不足的原因,Gulrajani等人(2017)提出WGAN(Wasserstein GAN),通过优化散度计算提升了训练稳定

性和收敛速度。Karras等人(2019, 2020)提出StyleGAN及其改进版本StyleGAN2,分别通过风格控制机制和正则化技术实现对图像细节的精细控制和一致性的提升。此外,Kang等人(2023)通过结合多尺度训练进一步提高GAN在复杂场景中的适用性。

扩散模型凭借细节控制和生成质量的显著优势,逐渐超越GAN。Ho等人(2020)基于概率推断框架提出扩散模型,避免了GAN中的模式崩溃问题。Song等人(2022)提出去噪扩散隐式模型(denoising diffusion implicit model, DDIM),通过减少迭代步骤显著提升推理效率,同时保持优异的生成效果。Dhariwal和Nichol(2021)优化U-Net结构,增强扩散模型对噪声处理的适应性,使其在生成任务中首次超越GAN。Rombach等人(2022)提出潜在扩散模型(latent diffusion model, LDM),通过潜在空间的加噪和去噪操作降低计算成本,并引入跨注意力机制(cross-attention)增强多模态生成能力。Kim和Kim(2024)结合潜在扩散模型与隐式表示,提出一种高效生成任意分辨率图像的新架构,进一步提升了生成的通用性和灵活性。

尽管现有的数据增强方法在多数场景下能够生成高质量的逼真图像,但在实际变电站复杂场景中仍然面临诸多挑战,主要包括以下两点:1)数据集规模差异显著。现有生成方法大多基于大规模数据集,如MS-COCO(Microsoft common object in context)(Lin等, 2014)和CC12M(conceptual 12M)(Sharma等, 2018),这些公共数据集的图像数量少则数万,多则上亿。相比之下,变电站表计缺陷数据由于收集难度大且样本稀少,现有生成方法在少样本数据集上的表现往往难以达到预期,生成结果难以满足实际应用需求。2)生成过程缺乏有效控制。现有生成方法(Croituru等, 2023; Yang等, 2023)大多依赖去噪过程中的概率机制,生成结果具有随机性,难以实现定量控制,满足特定需求。而具备强控制能力的编辑方法(Couairon等, 2022; Wang等, 2023)则过于依赖原始图像,仅能在图像上进行强制性修改,限制了生成图像的多样性。这要求生成方法需要在保证多样性的同

时提供适度控制,以生成符合实际需求的缺陷图像。

针对上述问题,本文提出一种面向少样本表计缺陷图像生成的稳定扩散模型,包括以下3个步骤:1)表计知识嵌入。通过微调方法将独特标识符与表计绑定,实现缺陷相关特征注入预训练的文生图模型中,增强模型对变电站表计结构特征的适配能力,得到特征知识嵌入的模型权重。2)裂纹特征建模。创新性地提出裂纹特征建模方法,以完整图像为输入,提取线稿图与交互式分割生成的约束图,将裂纹掩码与线稿图结合,生成包含几何约束的控制图像,

实现对裂纹形状和位置的精准控制。3)条件控制生成。在主干网络基础上,结合超网络机制添加额外分支,通过标签驱动,将裂纹特征建模图像的特征逐层引入扩散过程,有效引导生成,确保生成图像在形态和位置上的一致性多样性。图1展示了本文方法流程。生成模型完成训练后,可用于生成缺陷图像,进一步优化检测模型。实验结果表明,所提方法能够显著提升缺陷图像的生成质量和多样性,同时在工业场景的表计缺陷检测中表现出更高的可靠性和精确性。

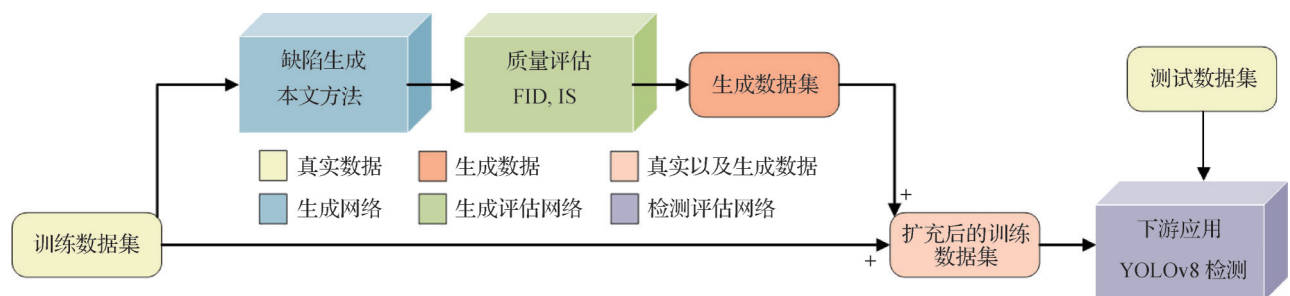


图1 本文方法流程

Fig. 1 Pipeline of the proposed approach

1 方法

针对现有生成模型在少样本表计缺陷数据集中的应用问题,本文提出一种专业生成模型。首先,通过表计知识嵌入,将新数据集的知识注入初始文生图模型;随后,提出裂纹特征建模方法,生成包含几何约束的控制图像;最后,在知识嵌入和控制图像的基础上,利用超网络机制与标签驱动约束引导生成,构建面向少样本表计缺陷的生成模型。总体架构如图2所示。

1.1 表计知识嵌入

训练条件生成大模型需要大规模标注数据集和高昂计算资源,而变电站表计数据集数量稀少、特征复杂,与自然图像数据分布差异明显。因此,研究如何利用现有大模型的生成能力,推动其在下游任务中的应用具有重要意义(Li等,2024;Hu等,2021)。图3展示了Stable Diffusion模型对“变电站表计”概念的“理解”与实际图像的差异,直接利用预训练模型生成表计缺陷图像存在局限性。

为实现表计知识的有效嵌入,本文采用了DreamBooth方法(Ruiz等,2023),通过对Stable Dif-

fusion进行微调,将独特标识符[表计]赋予新含义,作为输入提示注入训练过程,使现有模型适配变电站表计特征知识,从而增强模型生成表计缺陷图像的能力。在此过程中,模型可能会遗忘从大规模数据中学到的先验知识,从而削弱其通用生成能力。为此,本文使用了包含主体损失与先验保持损失的优化目标,用以在新概念学习与先验知识保持之间实现平衡。

1)主体损失。主体损失项用于优化模型在特定条件下生成包含新特征的图像,通过最小化模型预测与目标分布的差距,引导模型学习特定的表计缺陷特征。其表达式为

$$L_{\text{subj}} = \mathbb{E}_{x,c,\epsilon,t} [\omega_t \|\hat{x}_\theta(\alpha_t x + \sigma_t \epsilon, c) - x\|^2] \quad (1)$$

式中, x 表示真实图像, c 是文本条件, \hat{x}_θ 表示模型的预测图像, α_t 和 σ_t 是扩散噪声调度参数, ϵ 是随机噪声。优化主体损失,模型能够在独特标识符条件下生成具备特定缺陷特征的高质量图像。

2)先验保持损失。先验保持损失项旨在约束模型的生成结果,使其不偏离原有数据分布,从而避免因少样本过拟合而导致模型泛化能力的下降。通过对比当前生成结果与原始分布生成的样本,最大程度保留原有先验知识。损失函数表达为

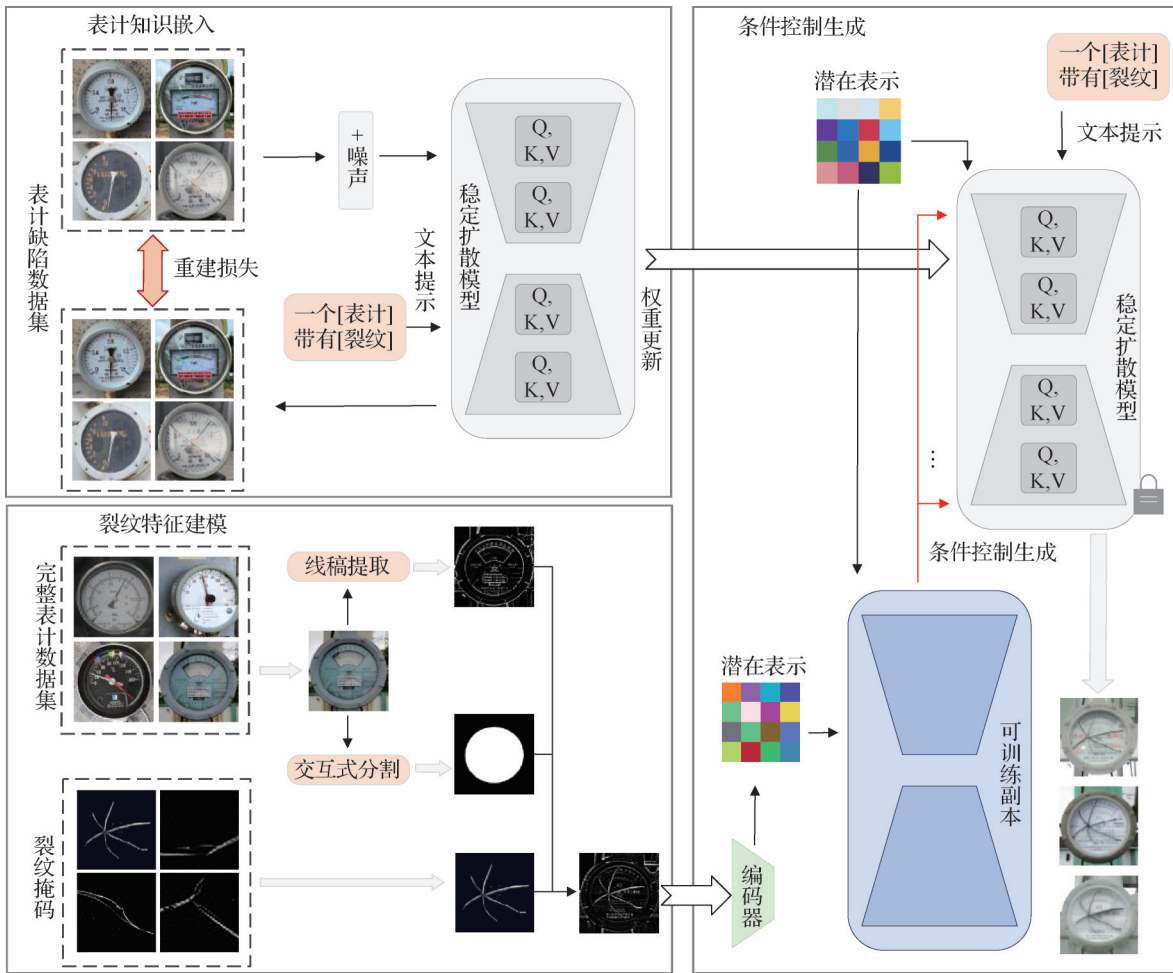


图2 本文方法整体流程

Fig. 2 The overall process of the proposed method



(a) 模型理解 (b) 真实概念

图3 模型理解以及真实概念

Fig. 3 Model understanding and real-world concepts

((a) model understanding; (b) real-world concepts)

$$L_{pr} = \mathbb{E}_{\mathbf{x}_{pr}, c_{pr}, \epsilon'} \left[\omega_r \|\hat{\mathbf{x}}_{\theta}(\alpha_r \mathbf{x}_{pr} + \sigma_r \epsilon', c_{pr}) - \mathbf{x}_{pr}\|^2 \right] \quad (2)$$

式中, \mathbf{x}_{pr} 表示模型在先前生成的图像, c_{pr} 是相应的文本条件, ϵ' 是另一随机噪声项。通过引入大规模未标注数据的生成样本, 有效避免模型遗忘通用生成能力。

损失函数整合实现两种目标的动态平衡, 以加

权组合形式表示为

$$L = L_{sbj} + \lambda L_{pr} \quad (3)$$

式中, $\lambda \in [0, 1]$ 是权重参数, 用于调整主体损失和先验保持损失权重, 根据任务灵活优化生成能力。在对损失函数进行相应的优化之后, 使用独特标识符 [表计] 将变电站表计这一特征知识注入到模型中, 使其能够生成具有特定目标特征的图像, 且保留了基础模型的泛化能力和对先验知识的理解。

1.2 裂纹特征建模

裂纹是工业环境中常见的缺陷, 形式多样, 特征复杂, 已成为工业检测与图像生成的重要研究对象 (Niu 等, 2022a, b)。相比于常见裂纹, 变电站表计裂纹表现出更大的多样性和更丰富的视觉效果, 增加了特征知识嵌入的难度。同时, 表计裂纹样本数量稀少且数据分布不平衡, 限制了模型在生成裂纹缺陷图像时对多样性和表现形式的覆盖能力。

在本文提出的裂纹特征建模方法中, 无缺陷原

图 $I_o \in \mathbf{R}^{h \times w \times 3}$ 经过边缘检测、图像平滑和风格增强等预处理, 提取得到线稿图像 $I_l \in \mathbf{R}^{h \times w \times 1}$ 。随后, 线稿图像与裂纹掩码 $I_d \in \mathbf{R}^{h \times w \times 1}$ 逐像素融合, 生成控制图像 $I'_c \in \mathbf{R}^{h \times w \times 1}$ 。融合过程表示为

$$I'_c(x, y) = \max(I_d(x, y), I_l(x, y)) \quad (4)$$

式中, $I'_c(x, y)$ 、 $I_d(x, y)$ 和 $I_l(x, y)$ 分别为控制图像、裂纹图像和线稿图像在像素点 (x, y) 的灰度值, 范围限制在 $[0, 255]$ 。在建模过程中, 裂纹图像的高灰度值突出裂纹特征, 线稿图像体现表计的整体结构, 通过融合完成裂纹特征建模。

然而, 由于裂纹分布的特殊性, 直接将线稿图像与裂纹图像结合可能导致裂纹特征出现在表盘之外的异常现象, 无法针对表盘区域进行精确建模。因此, 仅通过简单的像素融合难以有效实现裂纹特征的精准建模。

为解决上述问题, 可借助分割一切 (segment anything, SAM) 模型的强大功能实现精准建模 (Kirillov 等, 2023)。SAM 是一种通用的交互式分割模型, 具备对任意目标区域进行精确分割的能力。其核心由图像编码器、提示编码器和掩码解码器组成, 通过对图像的全局特征进行提取, 并结合用户输入的提示, 快速生成对应的分割掩码。凭借出色的分割能力, SAM 已广泛应用于图像处理、图像编辑等领域, 为多场景下的精确建模提供了高效且灵活的解决方案。

本文基于 SAM 提出一种裂纹特征建模方法, 通过限制裂纹的建模范围, 提升其真实性。对于输入的原始图像 I_o 和用户提供的提示条件 c , SAM 首先通过图像编码器将 I_o 映射至嵌入空间, 生成特征表示, 具体为

$$f = E(I_o) \quad (5)$$

同时, 通过提示编码器对提示条件 c 进行编码, 生成提示特征, 具体为

$$g = E'(c) \quad (6)$$

式中, f 为嵌入特征, g 为提示特征。 E 和 E' 分别表示图像编码器和提示编码器。随后, 特征 f 和 g 在掩码解码器中融合, 生成表盘约束图 I_r , 具体为

$$I_r = D(f, g), I_r \in \mathbf{R}^{h \times w \times 1} \quad (7)$$

式中, D 为掩码解码器, 在获取表盘约束后, 将其约束于线稿图像和掩码图像逐像素融合过程, 生成最终控制图像, 融合过程可更新为

$$I_c(x, y) = \begin{cases} I'_c(x, y), & I_r(x, y) = 255 \\ I_l(x, y), & I_r(x, y) = 0 \end{cases} \quad (8)$$

式中, $I_c(x, y)$ 和 $I_r(x, y)$ 分别表示最终控制图像和约束图像在像素点 (x, y) 的灰度值。 $I_r(x, y)$ 的灰度值仅取 0 和 255 两种, 前者表示允许结合的区域, 后者表示保留线稿图像原始灰度值。通过引入约束图, 能够在裂纹特征建模过程中施加更精确的区域控制, 有助于建模合理性更高的控制图像。

本文所提出的裂纹特征建模方法具体过程如图 4 所示, 充分发挥了 SAM 在精确区域分割中的强大能力, 结合裂纹掩码、线稿图和约束图三者的特性, 以协同方式实现对裂纹特征的精确建模。裂纹掩码定义了裂纹的基本形态, 确保裂纹形态的清晰表达; 线稿图完整保留了表计的结构特征, 避免了对表计轮廓的破坏; 而约束图进一步限定了裂纹的分布范围, 有效避免了裂纹特征跨区域不合理分布的问题, 从而在多方约束下得到表达能力更为均衡的控制图像。

1.3 条件控制生成

在完成裂纹特征建模控制图像的生成后, 下一步是将其引入图像生成过程, 以实现整体图像外观的全局约束, 同时精确定位和控制裂纹的位置与形态。然而, 从粗略的线稿图像生成风格多样且与之对应的表计缺陷图像, 依然是一项极具挑战性的任务。在 GAN 生成领域, 已有研究通过引入超网络提升生成质量和精度 (Alaluf 等, 2022; Dinh 等, 2022), 并取得了显著成效。然而, 这些方法主要依赖于 StyleGAN 架构, 难以推广至其他生成模型。

为解决这一问题, 本文参考了 Zhang 等人 (2023) 的研究方法, 引入超网络对微调后的 Stable Diffusion 模型进行控制, 使其能够接受额外的条件输入。超网络是一种小型神经网络, 能够根据输入条件动态调整更大型网络的权重。如图 5 所示, 通过在保持 Stable Diffusion 主模型原始权重不变的前提下, 加入可训练的超网络层, 为生成过程引入了灵活且精确的条件约束。

具体而言, 首先将 Stable Diffusion 主模型的权重冻结, 以保留其原有的图像先验知识。在此基础上, 构建一个可训练的模型副本作为超网络, 通过共享部分特征表示, 动态调整主模型的生成过程。同时, 输入的条件控制图像经编码器处理后, 生成对应的

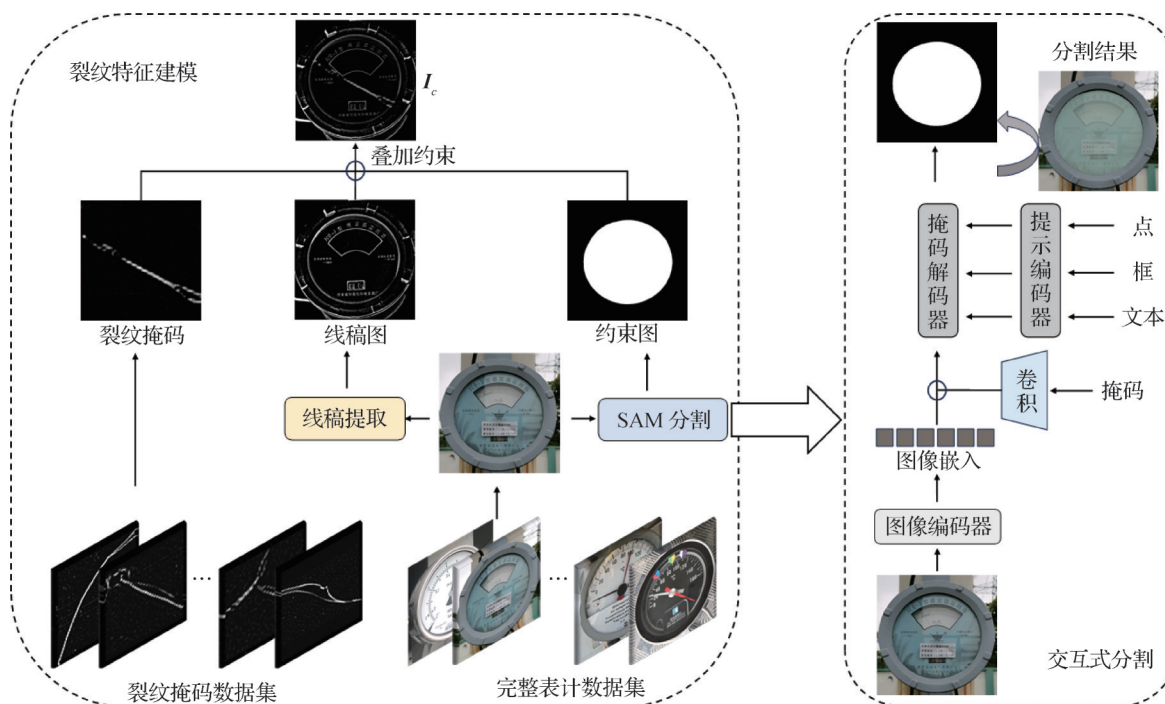


图4 裂纹特征建模

Fig. 4 Crack feature modeling

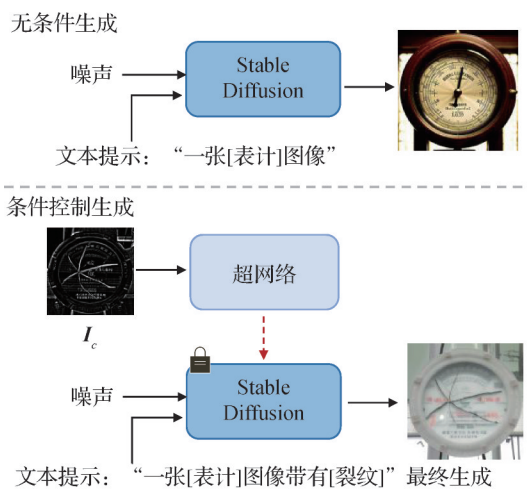


图5 基于超网络的条件控制生成

Fig. 5 Conditional generation based on hypernetworks

潜在表示(latent representation)。这一潜在表示与超网络生成的动态权重相结合,有效调节生成图像的整体轮廓与裂纹细节。同时,基础生成模型会生成初步图像,两者协同作用,共同完成最终图像的生成,从而确保生成结果严格符合设定条件,达到精确控制的效果。

最后,本文以约束图像作为输入条件,用于展现整体图像的粗略轮廓以及裂纹的位置和形态。在此阶段,设计了特定的损失函数以优化生成过程,定

义为

$$L = \mathbb{E}_{x,c,c',\epsilon,\delta} [\|\hat{x}_\theta(\alpha_t x + \sigma_t \epsilon, c, c') - x\|^2] \quad (9)$$

式中, c 为文本提示条件, c' 为控制图像的条件控制信息, $\alpha_t x + \sigma_t \epsilon$ 为当前扩散步的图像表示。在扩散生成过程中,控制信号多次注入到模型的各特征层中,既影响生成初始阶段,又在中后期提供结构约束,从而确保生成结果符合设定条件。

2 实验结果及分析

2.1 数据集

本文研究面向少样本表计缺陷图像生成的稳定扩散模型,由于当前尚无相关研究,首先需要构建专门的表计缺陷数据集。为此,构建了变电站表计数据集 SMD(substation meter dataset),涵盖多种不同类型的表计图像,包括带裂纹缺陷的样本和无缺陷的正常样本。部分示例图像如图3(b)所示。

最终数据集包含629幅图像,其中100幅为带裂纹缺陷的样本。所有图像均进行了精确的类别标注,用于后续缺陷检测实验中评估生成效果。数据集按照8:2的比例划分为训练集和验证集,以确保数据分布的合理性。

在生成部分的表计知识嵌入阶段,本文仅使用

带裂纹缺陷图像对 Stable Diffusion 的预训练权重进行微调,未引入无缺陷图像,以避免数据泄露风险。在裂纹特征建模阶段,则基于无缺陷图像及相应的裂纹掩码生成控制图像。裂纹掩码的来源包括现有工业数据集中的裂纹缺陷图像,以及由行业专家手绘得到的裂纹图像。其中,专家手绘的裂纹掩码为裂纹建模提供了可靠补充,尤其在可能出现 SAM 分割失败的情况下,可有效弥补表盘区域分割的不足,确保裂纹特征的完整性和精准性,从而降低对分割依赖的风险,进一步提高生成模型的鲁棒性和泛化性能。

此外,为满足模型输入要求并提升生成质量,本文对所有图像进行了裁剪,每幅图像仅保留单个表计,并统一调整分辨率为 512×512 像素。

2.2 评价指标

采用精度(precision, P)、召回率(recall, R)和平均精度均值(mean average precision, mAP)作为图像检测的主要评价指标。计算方法为

$$P = \frac{TP}{TP + FP} \quad (10)$$

$$R = \frac{TP}{TP + FN} \quad (11)$$

式中, TP (true positive)、 FP (false positive) 和 FN (false negative) 分别表示真阳性、假阳性和假阴性样本数量。

mAP 通过计算不同交并比(intersection over union, IoU) 阈值下的 P-R 曲线面积衡量检测模型的综合性能,其中,mAP50 指 IoU 阈值为 50% 时的平均精度。mAP50-95 表示在 IoU 阈值从 50% 到 95% 范围内,以 5% 为步长,计算各阈值下平均精度的均值。

此外,为评估生成图像的质量和多样性,本文引入弗雷歇特启动距离(Fréchet inception distance, FID)和启动分数(inception score, IS)作为评价指标。FID 计算生成图像与真实图像在特征空间的分布距离,值越小表示生成质量越高。IS 通过类别分布衡量图像清晰度与多样性,值越大表示生成效果越优。

2.3 生成效果评估

本文基于所构建的 SMD 数据集对提出的方法进行了全面实验评估,并与现有相关工作进行了对比分析。

如图 6 所示,本文方法能够生成高度逼真的表计图像,使 Stable Diffusion 在超出原始训练分布的情况下,能够生成高质量样本。生成图像中的缺陷

外观严格遵循输入裂纹掩码的形状和位置,同时整体轮廓受到固定约束。此外,生成图像的整体颜色可能与初始图像存在差异。这一现象并非源于传统的图像编辑操作,而是通过裂纹特征建模模块生成的控制图,结合条件控制生成过程,实现对图像内容的完整生成,并确保其符合预设约束。

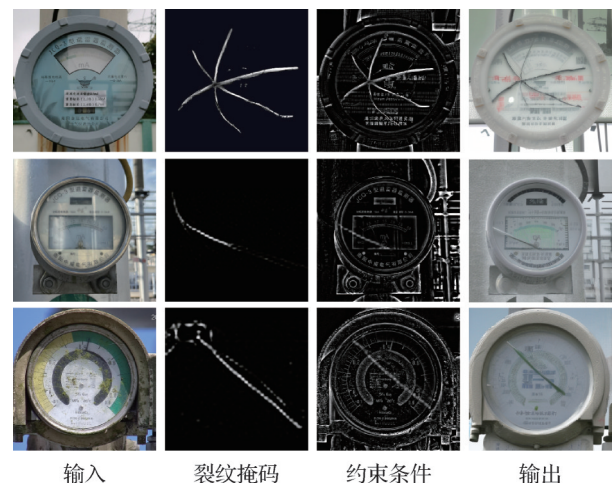


图6 条件控制生成过程

Fig. 6 Conditional generation process

此外,本文探讨了合成数据在实际变电站应用场景中的可行性,尤其是其在训练模型执行缺陷检测任务中的作用。为验证这一点,本文采用 YOLOv8n (you only look once) 模型,在多种数据配置下进行了实验,包括仅使用真实数据,以及在真实数据基础上添加不同比例的生成数据。通过在独立的真实数据集上对模型性能进行评估,量化了合成数据对缺陷检测指标的影响,具体实验结果如表 1 所示。实验结果表明,引入合成数据显著提升了模型的精度和 mAP50 等指标,尤其是在生成数据比例接近真实数据一半时,对检测性能的提升最为显著。当额外添加 40% 生成数据时,精度提高 26.9%, mAP50 增加了 19.1%,表明生成数据在增强模型检测能力方面具有重要价值。在召回率指标上,模型的提升较为有限。这可能是由于生成数据虽然增强了正样本的多样性,但在细粒度缺陷特征的覆盖上仍存在不足,导致部分真实数据中的少数缺陷样本未能被有效检测。此外,当生成数据比例进一步增加至 80% 时,尽管召回率保持稳定,但精度略有下降。本文认为,这可能是由于生成数据在特征表示上仍与真实数据存在一定差距,无法完全模拟真实数据的分布特性。

表1 不同比例生成数据对检测效果的影响

Table 1 Impact of different proportions of generated data on detection performance

数据集	精度	召回率	mAP50	mAP50-95
仅包含真实数据	0.698	0.56	0.587	0.464
真实数据 + 20% 生成数据	0.784	0.58	0.654	0.483
真实数据 + 40% 生成数据	0.967	0.52	0.778	0.561
真实数据 + 60% 生成数据	0.867	0.52	0.761	0.545
真实数据 + 80% 生成数据	0.877	0.56	0.642	0.491
真实数据 + 100% 生成数据	0.776	0.64	0.652	0.515

注:加粗字体表示各列最优结果。

P-R曲线以召回率为横坐标、精度为纵坐标,用于衡量模型的检测性能。图7展示了检测模型在基础数据集和扩充40%数据条件下的P-R曲线。扩充数据后的曲线整体高于基础数据曲线,表明通过数据扩充,本文方法显著提升了检测性能。

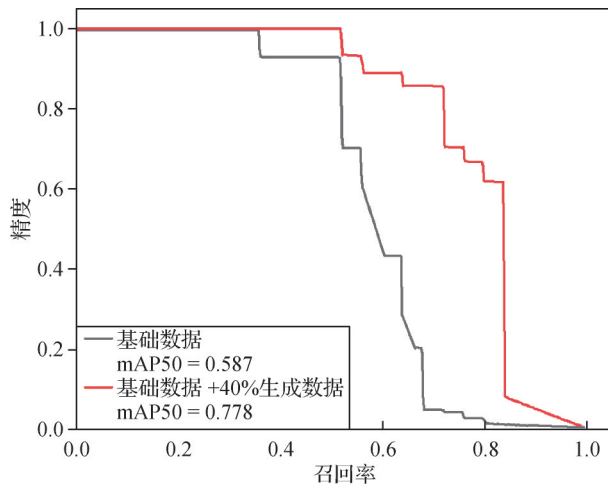


图7 不同数据比例的P-R曲线对比

Fig. 7 Comparison of P-R curves for different data proportions

为了进一步对比分析本文方法与现有少样本生成方法的性能,本文选择了5种主流生成模型进行实验,分别为FastGAN(Liu等,2021)、ProjectedGAN(Sauer等,2021)、DFMGAN(defect feature manipulation GAN)(Duan等,2023)、IDDPM(improved denoising diffusion probabilistic models)(Nichol和 Dhariwal, 2021)以及RDDM(residual denoising diffusion models)(Liu等,2024)。其中,FastGAN和ProjectedGAN是基于生成对抗网络的经典少样本生成模型,已在先前研究中展示了良好的生成效果;IDDPM和RDDM则为改进的扩散模型,通过优化扩

散过程,在少样本条件下生成更清晰且多样性更高的图像;DFMGAN采用两阶段训练策略,首先在大量无缺陷图像上进行预训练,随后结合缺陷图像及其缺陷部分对应掩码进行微调,以生成缺陷图像。

为确保对比的公平性,本文在相同的SMD数据集和一致的实验环境下,对各模型进行了少样本生成训练和测试,相关结果如图8和表2所示。

如图8所示,FastGAN生成的图像存在较为严重的失真问题,整体模糊且缺陷拖影现象显著,语义表达缺乏合理性。ProjectedGAN在一定程度上改善了模糊问题,但生成结果仍存在失真。例如,生成的表计表面形状偏离了正常的圆形结构,且生成图像与初始缺陷图像过于相似,这是生成对抗网络模型中常见的问题之一,这表明其生成质量在多样性和独立性上仍然有限。DFMGAN方法由Duan等人(2023)提出,采用了两阶段训练策略:第1阶段在大量正常表计图像上进行预训练;第2阶段通过引入缺陷图像及其对应掩码进行微调。然而,在表计缺陷生成实验中,DFMGAN未能生成预期的缺陷特征,而是出现了随机分布的小黑点。这一现象可能源于表计结构的复杂性,导致模型在少量数据上难以复刻相应的特征分布。

与生成对抗网络相比,基于扩散模型的生成方法在视觉效果上展现出一定优势。IDDPM在减少计算量的同时,能够在有限计算资源下生成图像,但生成结果与原始图像的相似度较高,部分图像还存在亮度偏差和模糊问题。RDDM通过引入ResNet结构提升了生成能力,尽管部分生成图像的质量较高,但仍容易出现噪点,影响整体效果。此外,其生成结果与原始缺陷图像的相似性问题依然突出,限制了模型在多样化生成上的表现。



图8 与现有方法生成效果对比图

Fig. 8 Comparison of generation results with existing methods

表2 与现有方法生成效果对比

Table 2 Comparison of generation results with existing methods

模型	FID ↓	IS ↑
FastGAN	127.90	1.98
ProjectedGAN	99.83	1.77
DFMGAN	94.32	0.96
IDDPM	131.69	1.86
RDDM	219.83	1.88
本文方法 + SD1.5	76.72	2.45
本文方法 + SDXL	70.71	2.68
本文方法 + FLUX	75.28	2.74

注:加粗字体表示各列最优结果。“↑”和“↓”分别表示值越大越好和值越小越好。

与传统生成方法相比,本文方法在生成图像的真实度、裂纹多样性以及与原图的差异性方面展现了显著的优势。在基础的SD1.5模型上应用本文方法,不仅可以生成裂纹清晰、形态多样的表盘图像,还能通过结合完整图像与相应的裂纹掩码,灵活生成多样化的缺陷表计图像,充分满足实际需求。

进一步地,在性能更强的SDXL(stable diffusion extreme large)模型上应用本文方法,生成图像的清

晰度和背景真实度得到了显著提升。SDXL生成的裂纹特征更加细腻、背景纹理更加丰富且贴近真实场景,进一步验证了本文方法在高分辨率生成任务中的优越性。

而在FLUX(first large-unified X-to-image model)模型上的实验结果表明,FLUX生成的图像在清晰度和多样性方面表现尤为突出,特别是在背景细节和裂纹特征的丰富性上,明显优于SD1.5。然而,在条件控制生成任务中,FLUX模型未能严格遵循输入的控制信息,生成图像存在一定程度的自由发挥。这可能是由于FLUX模型本身结构过于复杂,导致其在条件信息处理上的精确性有所欠缺。

通过量化指标可以更直观地比较各方法的性能。如表2所示,本文方法在FID和IS指标上均表现优异,明显优于其他方法,展现了其在生成图像类别清晰度和多样性上的优势。此外,随着基础模型版本的升级,例如SDXL和FLUX,生成图像的FID进一步降低,IS指标显著提升,这表明更强大的基础模型能够在一定程度上增强本文方法的生成效果。然而,本文仍主要基于SD1.5模型进行实验分析,以验证方法在少样本条件下的通用性和实际应用价值。

综上,传统生成方法和现有少样本生成方法在表计缺陷生成任务中仍存在不足。相比之下,本文方法通过将正常表计图像与缺陷特征相结合,创新

性地实现了从正常表计生成缺陷图像的能力,不仅有效提高生成图像的多样性,还准确描绘了缺陷的形态特征,生成了多样化且具备显著缺陷特征的高质量表计图像,为后续检测任务提供了更为丰富和可靠的数据支持。

2.4 消融实验

为验证本文方法在少样本表计数据集上的生成效果,在基础的SD1.5模型上设计了一系列消融实验,以分析各模块对最终生成质量的具体贡献。如图9所示,不同实验阶段生成的图像在质量上表现

出明显差异。基础模型虽然在生成的多样性和形式多样化方面具有优势,但在特征知识掌握上仍显不足。生成的图像全局结构与真实表计图像存在较大差距,难以准确呈现表计的关键特征和裂纹缺陷的具体形态。通过引入 DreamBooth 方法进行表计知识嵌入后,模型在调整预训练权重的基础上,显著改善了生成图像的样式,使其更贴近真实变电站表计特征。然而,此阶段生成的图像在缺陷特征的表现上仍存在一定不足,特别是在裂纹形状的多样性和细节刻画方面,尚未达到理想水平。

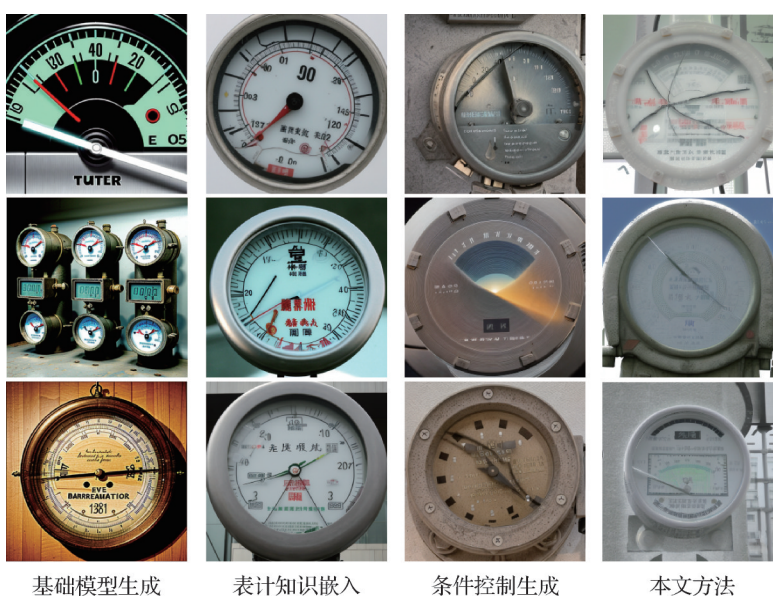


图9 消融实验结果样例

Fig. 9 Examples of ablation experiment results

仅使用条件控制生成模块进行生成时,虽然在形态上能够严格符合输入条件的几何约束,但生成图像在配色上存在一定偏差,且部分图像未能准确结合表计概念与输入约束,导致生成结果偏离预期,呈现出其他风格的内容。条件控制模块在几何约束上表现出色,但是在语义理解上仍然存在不足。

结合条件控制生成模块后,生成图像的整体质量进一步提升。通过引入裂纹特征建模模块,生成图像的轮廓严格遵循输入控制条件中的几何结构,裂纹缺陷区域也得到了精准复现。最终生成的图像不仅在整体结构一致性上表现优异,还在缺陷细节刻画和形态多样性方面取得了显著改进。

为定量评估生成图像的质量与多样性,本文采用 FID 和 IS 两项指标对实验结果进行分析。通过将生成图像与 SMD 数据集的真实分布进行对比,本文

方法的 FID 值显著降低至 76.72, IS 指标提升至 2.45,具体结果如表3所示。从表3可以看出,表计知识嵌入模块显著提升了生成图像的质量,与基线模型相比, FID 值降低 20.57%, IS 值提升 18.75%。在此基础上,结合条件控制生成模块后, FID 值进一步下降 33.47%, IS 值实现翻倍增长。这表明,表计知识嵌入与条件控制生成模块的结合在提升生成图像质量与多样性方面具有重要作用。

具体来看,表计知识嵌入通过引入表计的特定结构特征,有效改善了生成图像的全局一致性;条件控制生成模块则借助裂纹特征建模方法提供了精确的几何约束,使模型能够准确控制裂纹的形状和位置。两者的模块化设计相辅相成,不仅强化了模型的生成能力,还为生成高质量且多样化的缺陷表计图像提供了有力支持。

表3 消融实验的FID和IS指标数据

Table 3 FID and IS metrics of ablation experiments

条件控制生成	表计知识嵌入	FID ↓	IS ↑
×	×	115.31	1.12
×	√	91.59	1.33
√	×	89.45	2.39
√	√	76.72	2.45

注:加粗字体表示各列最优结果。“√”和“×”分别表示使用和未使用对应模块。“↑”和“↓”分别表示值越大越好和值越小越好。

3 结论

针对变电站表计缺陷图像生成任务的特点,本文提出一种面向少样本表计缺陷图像生成的稳定扩散模型。主要贡献包括:1)针对现有模型缺乏变电站表计相关知识的问题,设计了表计知识嵌入模块,将表计结构特征有效融入预训练模型,显著增强了模型在表计领域的适配能力。2)针对裂纹缺陷的分布特性,提出裂纹特征建模模块。通过结合线稿图、裂纹掩码及约束图生成控制图像,精准表达裂纹的形态与位置,为生成任务提供了可靠的几何约束。3)引入超网络机制和标签驱动生成方式,动态调整生成过程中的权重分布,在提升生成图像多样性的同时,严格控制缺陷形态与分布条件,确保生成结果满足设定需求。

实验结果表明,本文方法在生成图像质量和多样性上优于对比的生成方法,并显著提升了下游检测任务的实用效果。同时,本文方法在生成裂纹缺陷图像方面的能力,具备扩展至其他工业设备和复杂环境中的潜在应用价值,有望为更广泛的工业检测任务提供新方案。然而,当前方法主要针对裂纹缺陷展开研究,尚未涉及其他类型的缺陷生成问题。未来研究将进一步扩展至更多样化的缺陷类型,并探索如何提升生成图像的特征表现,为工业场景下的表计缺陷检测与应用提供更强有力的支持。

参考文献(References)

Alaluf Y, Tov O, Mokady R, Gal R and Bermano A. 2022. HyperStyle: StyleGAN inversion with hypernetworks for real image editing//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and

Pattern Recognition. New Orleans, USA: IEEE: 18490-18500 [DOI: 10.1109/CVPR52688.2022.01796]

Arjovsky M and Bottou L. 2017. Towards principled methods for training generative adversarial networks [EB/OL]. [2024-12-30].

<https://arxiv.org/pdf/1701.04862.pdf>

Couairon G, Verbeek J, Schwenk H and Cord M. 2022. DiffEdit: diffusion-based semantic image editing with mask guidance [EB/OL]. [2024-12-30]. <https://arxiv.org/pdf/2210.11427.pdf>

Croitoru F A, Hondru V, Ionescu R T and Shah M. 2023. Diffusion models in vision: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(9): 10850-10869 [DOI: 10.1109/TPAMI.2023.3261988]

Dhariwal P and Nichol A. 2021. Diffusion models beat GANs on image synthesis [EB/OL]. [2024-12-30].

<https://arxiv.org/pdf/2105.05233.pdf>

Dinh T M, Tran A T, Nguyen R and Hua B S. 2022. HyperInverter: improving StyleGAN inversion via hypernetwork//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 11379-11388 [DOI: 10.1109/CVPR52688.2022.01110]

Duan Y X, Hong Y, Niu L and Zhang L Q. 2023. Few-shot defect image generation via defect-aware feature manipulation//Proceedings of the 37th AAAI Conference on Artificial Intelligence. Washington, USA: AAAI: 571-578 [DOI: 10.1609/aaai.v37i1.25132]

Goodfellow I J, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y. 2014. Generative adversarial networks [EB/OL]. [2024-12-30].

<https://arxiv.org/pdf/1406.2661.pdf>

Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V and Courville A. 2017. Improved training of Wasserstein GANs [EB/OL]. [2024-12-30].

<https://arxiv.org/pdf/1704.00028.pdf>

Ho J, Jain A and Abbeel P. 2020. Denoising diffusion probabilistic models [EB/OL]. [2024-12-30]. <https://arxiv.org/pdf/2006.11239.pdf>

Hu E J, Shen Y L, Wallis P, Allen-Zhu Z, Li Y Z, Wang S A, Wang L and Chen W Z. 2021. LoRA: low-rank adaptation of large language models [EB/OL]. [2024-12-30].

<https://arxiv.org/pdf/2106.09685.pdf>

Kang M, Zhu J Y, Zhang R, Park J, Shechtman E, Paris S and Park T. 2023. Scaling up GANs for text-to-image synthesis//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 10124-10134 [DOI: 10.1109/CVPR52729.2023.00976]

Karras T, Laine S and Aila T. 2019. A style-based generator architecture for generative adversarial networks//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 4396-4405 [DOI: 10.1109/CVPR.2019.00453]

Karras T, Laine S, Aittala M, Hellsten J, Lehtinen J and Aila T. 2020. Analyzing and improving the image quality of StyleGAN//Proceed-

- ings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 8110-8119 [DOI: 10.1109/CVPR42600.2020.00813]
- Kim J and Kim T K. 2024. Arbitrary-scale image generation and upsampling using latent diffusion model and implicit neural decoder//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 9202-9211 [DOI: 10.1109/CVPR52733.2024.00879]
- Kirillov A, Mintun E, Ravi N, Mao H Z, Rolland C, Gustafson L, Xiao T T, Whitehead S, Berg A C, Lo W Y, Dollár P and Girshick R. 2023. Segment anything//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 3992-4003 [DOI: 10.1109/ICCV51070.2023.00371]
- Li C Y, Gan Z, Yang Z Y, Yang J W, Li L J, Wang L J and Gao J F. 2024. Multimodal foundation models: from specialists to general-purpose assistants. *Foundations and Trends® in Computer Graphics and Vision*, 16(1/2): 1-214 [DOI: 10.1561/0600000110]
- Lin T Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P and Zitnick C L. 2014. Microsoft COCO: common objects in context//Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland: Springer: 740-755 [DOI: 10.1007/978-3-319-10602-1_48]
- Liu B C, Zhu Y Z, Song K P and Elgammal A. 2021. Towards faster and stabilized GAN training for high-fidelity few-shot image synthesis [EB/OL]. [2024-12-30]. <https://arxiv.org/pdf/2101.04775.pdf>
- Liu J W, Wang Q, Fan H J, Wang Y N, Tang Y D and Qu L Q. 2024. Residual denoising diffusion models//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 2773-2783 [DOI: 10.1109/CVPR52733.2024.00268]
- Nichol A and Dhariwal P. 2021. Improved denoising diffusion probabilistic models [EB/OL]. [2024-12-30]. <https://arxiv.org/pdf/2102.09672.pdf>
- Niu S L, Li B, Wang X G and Peng Y R. 2022a. Region- and strength-controllable GAN for defect generation and segmentation in industrial images. *IEEE Transactions on Industrial Informatics*, 18(7): 4531-4541 [DOI: 10.1109/TII.2021.3127188]
- Niu T Z, Li B, Li W F, Qiu Y H and Niu S L. 2022b. Positive-sample-based surface defect detection using memory-augmented adversarial autoencoders. *IEEE/ASME Transactions on Mechatronics*, 27(1): 46-57 [DOI: 10.1109/TMECH.2021.3058147]
- Qi Y C, Wu X L, Zhao Z B, Shi B Q and Nie L Q. 2021. Bolt defect detection for aerial transmission lines using Faster R-CNN with an embedded dual attention mechanism. *Journal of Image and Graphics*, 26(11): 2594-2604 (戚银城, 武学良, 赵振兵, 史博强, 聂礼强. 2021. 嵌入双注意力机制的 Faster R-CNN 航拍输电线路螺栓缺陷检测. *中国图象图形学报*, 26(11): 2594-2604) [DOI: 10.11834/jig.200793]
- Rombach R, Blattmann A, Lorenz D, Esser P and Ommer B. 2022. High-resolution image synthesis with latent diffusion models//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 10674-10685 [DOI: 10.1109/CVPR52688.2022.01042]
- Royle J A, Dorazio R M and Link W A. 2007. Analysis of multinomial models with unknown index using data augmentation. *Journal of Computational and Graphical Statistics*, 16(1): 67-85 [DOI: 10.1198/106186007X181425]
- Ruiz N, Li Y Z, Jampani V, Pritch Y, Rubinstein M and Aberman K. 2023. DreamBooth: fine tuning text-to-image diffusion models for subject-driven generation//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 22500-22510 [DOI: 10.1109/CVPR52729.2023.02155]
- Sauer A, Chitta K, Müller J and Geiger A. 2021. Projected GANs converge faster [EB/OL]. [2024-12-30]. <https://arxiv.org/pdf/2111.01007.pdf>
- Sharma P, Ding N, Goodman S and Soricic R. 2018. Conceptual captions: a cleaned, hypernymed, image alt-text dataset for automatic image captioning//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Melbourne, Australia: Association for Computational Linguistics: 2556-2565 [DOI: 10.18653/v1/P18-1238]
- Siddiqui Z A and Park U. 2020. A drone based transmission line components inspection system with deep learning technique. *Energies*, 13(13): #3348 [DOI: 10.3390/en13133348]
- Song J M, Meng C L and Ermon S. 2022. Denoising diffusion implicit models [EB/OL]. [2024-12-30]. <https://arxiv.org/pdf/2010.02502.pdf>
- Wang Q, Zhang B, Birsak M and Wonka P. 2023. InstructEdit: improving automatic masks for diffusion-based image editing with user instructions [EB/OL]. [2024-12-30]. <https://arxiv.org/pdf/2305.18047.pdf>
- Xie L F, Jiang C, Sun Q Q, Wang H W, Song Q W and Guan G F. 2019. The global map's creating and positioning of substation inspection robot based on adaptive Monte Carlo particle filter algorithm. *Electric Power Engineering Technology*, 38(5): 16-23 (谢林枫, 蒋超, 孙秋芹, 王昊炜, 宋庆武, 官国飞. 2019. 基于 AMC 算法的变电站巡检机器人地图创建与定位. *电力工程技术*, 38(5): 16-23) [DOI: 10.12158/j.2096-3203.2019.05.003]
- Yang L, Zhang Z L, Song Y, Hong S D, Xu R S, Zhao Y, Zhang W T, Cui B and Yang M H. 2023. Diffusion models: a comprehensive survey of methods and applications. *ACM Computing Surveys*, 56(4): #105 [DOI: 10.1145/3626235]
- Zhang L M, Rao A Y and Agrawala M. 2023. Adding conditional control to text-to-image diffusion models//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 3813-3824 [DOI: 10.1109/ICCV51070.2023.00355]
- Zhao Z B, Jiang Z G, Li Y X, Qi Y C, Zhai Y J, Zhao W Q and Zhang

K. 2021. Overview of visual defect detection of transmission line components. *Journal of Image and Graphics*, 26(11): 2545-2560 (赵振兵, 蒋志钢, 李延旭, 戚银城, 翟永杰, 赵文清, 张珂. 2021. 输电线路部件视觉缺陷检测综述. *中国图象图形学报*, 26(11): 2545-2560) [DOI: 10.11834/jig.200689]

Zhu X H, Qian L P and Fu W. 2021. Overview of research on image data enhancement technology. *Software Guide*, 20(5): 230-236 (朱晓慧, 钱丽萍, 傅伟. 2021. 图像数据增强技术研究综述. *软件导刊*, 20(5): 230-236) [DOI: 10.11907/rjdk.202372]

作者简介

张珂,男,教授,博士生导师,主要研究方向为计算机视觉和电力人工智能技术。E-mail:zhangkeit@ncepu.edu.cn

盛鑫,男,硕士研究生,主要研究方向为电气设备的图像生成。E-mail:220232215019@ncepu.edu.cn

肖扬杰,男,博士研究生,主要研究方向为电力视觉研究。E-mail:jay_xyj@ncepu.edu.cn

杨济远,男,硕士研究生,主要研究方向为电气设备多模态检测。E-mail:y7217yjy@163.com

陈美娟,女,硕士研究生,主要研究方向为电气设备的图像生成与编辑。E-mail:220232215004@ncepu.edu.cn

任泽华,男,硕士研究生,主要研究方向为电气设备的图像生成。E-mail:220242215106@ncepu.edu.cn