

# 面向工业过程的图像生成及其应用研究综述

汤健<sup>1,2,3</sup> 郭海涛<sup>1,2,3</sup> 夏恒<sup>1,2,3</sup> 王鼎<sup>1,2,3</sup> 乔俊飞<sup>1,2,3</sup>

**摘要** 在面向工业过程的计算机视觉研究中,智能感知模型能否实际应用取决于其对复杂工业环境的适应能力。由于可利用的工业图像数据集存在分布不均、多样性不足和干扰严重等问题,如何生成符合多工况分布的期望训练集是提高感知模型性能的关键。为解决上述问题,以城市固废焚烧(Municipal solid wastes incineration, MSWI)过程为背景,综述目前面向工业过程的图像生成及其应用研究,为进行面向工业图像的感知建模提供支撑。首先,梳理面向工业过程的图像生成定义和流程以及其应用需求;随后,分析在工业领域中具有潜在应用价值的图像生成算法;接着,从工业过程图像生成、生成图像评估和应用等视角进行现状综述;然后,对下一步研究方向进行讨论与分析;最后,对全文进行总结并指出未来挑战。

**关键词** 工业过程,视觉感知,图像生成,生成图像评估与应用,城市固废焚烧

**引用格式** 汤健,郭海涛,夏恒,王鼎,乔俊飞.面向工业过程的图像生成及其应用研究综述.自动化学报,2024,50(2):211-240

**DOI** 10.16383/j.aas.c230126

## Image Generation and Its Application Research for Industrial Process: A Survey

TANG Jian<sup>1,2,3</sup> GUO Hai-Tao<sup>1,2,3</sup> XIA Heng<sup>1,2,3</sup> WANG Ding<sup>1,2,3</sup> QIAO Jun-Fei<sup>1,2,3</sup>

**Abstract** In computer vision research for industrial process, the practical implementation of intelligent perception models is contingent upon their capacity for adapting to complex environments. As a result of issues such as non-uniform distribution, inadequate diversity, and significant interference within available image datasets, generating a training set that meets the multi-condition distribution is pivotal to enhance model performance. In order to address these issues, with the municipal solid wastes incineration (MSWI) process as background, this article focuses on current research on image generation and its application for industrial process, providing support for perceptual modeling for industrial images. Firstly, the definition and process of image generation for industrial process are summarized, as well as their application requirements in industrial process. Subsequently, the image generation algorithms with potential application value in the industrial domain are analyzed. Then, an overview is provided from the perspectives of industrial process image generation, generated image evaluation and application. Next, the future research direction is discussed and analyzed. Finally, we summarize the article and provide future challenges.

**Key words** Industrial process, visual perception, image generation, generated image evaluation and application, municipal solid wastes incineration (MSWI)

**Citation** Tang Jian, Guo Hai-Tao, Xia Heng, Wang Ding, Qiao Jun-Fei. Image Generation and Its Application Research for Industrial Process: A Survey. *Acta Automatica Sinica*, 2024, 50(2): 211-240

工业物联网、大数据、人工智能、云计算等新一代信息技术的发展,使得工业过程能够在传统的控制与决策基础上融入视觉感知信息<sup>[1]</sup>。目前,计算机

视觉模型已能够依据工业图像建立运行工况识别模型、产品质量检测模型和难测参数量化模型<sup>[2-4]</sup>,这些模型对复杂工业环境适应能力的强弱通常是决定其能否实际应用的关键<sup>[5]</sup>。

基于深度学习(Deep learning, DL)的视觉感知模型已在诸多领域得到广泛应用<sup>[6-9]</sup>,其具有以下优势:1)能够自动学习特征;2)能够获得具有完备性和非冗余性、强于人工获取方式的特征;3)能够学习复杂问题的非线性可分“分界面”;4)具有通用的问题解决思路和技术框架。复杂工业过程中的图像存在可解释性差、干扰性强、标记成本高等问题,这导致大量数据难以有效使用<sup>[10]</sup>,使得视觉感知模型在应用中存在识别精度低、鲁棒性差等现状<sup>[11]</sup>。以城市固废焚烧(Municipal solid wastes incineration, MSWI)过程<sup>[12]</sup>为例,存在的问题包括<sup>[13]</sup>:

收稿日期 2023-03-10 录用日期 2023-10-12

Manuscript received March 10, 2023; accepted October 12, 2023

科技创新 2030——“新一代人工智能”重大项目(2021ZD0112302)资助

Supported by National Key Research and Development Program of China (2021ZD0112302)

本文责任编辑 张军平

Recommended by Associate Editor ZHANG Jun-Ping

1. 北京工业大学信息学部 北京 100124 2. 北京工业大学智慧环保北京实验室 北京 100124 3. 北京工业大学智能感知与自主控制教育部工程研究中心 北京 100124

1. Faculty of Information Technology, Beijing University of Technology, Beijing 100124 2. Beijing Laboratory of Smart Environmental Protection, Beijing University of Technology, Beijing 100124 3. Engineering Research Center of Intelligent Perception and Autonomous Control, Ministry of Education, Beijing University of Technology, Beijing 100124

1) 燃烧过程中固有的飞灰、高温等因素使得火焰图像清晰度差; 2) 在炉排前端和后端进行燃烧的极端异常火焰图像稀缺; 3) 物料组分的不可控性和控制参数的波动性导致火焰图像的可解释性差; 4) 火焰图像难以标记. 因此, 该领域对视觉信息的处理依然依靠运行专家, 存在难以避免的主观性和随意性<sup>[14]</sup>. 可见, 因存在异常图像稀缺、图像对比度低和噪声干扰大等问题, 常用视觉模型难以适用于具有强污染、多噪声和图像类别不完备等特性的工业过程. 显然, 实际训练集的分布不符合期望全集分布已成为制约计算机视觉应用和发展的主要因素之一.

如何获取符合期望分布的训练图像集仍是一个开放性的难题. 图像生成<sup>[15]</sup>是解决该难题的方法之一. 目前, 已有的相关研究包括: 文献 [16–17] 阐述玻尔兹曼机研究进展, 包括亥姆霍兹机、深度玻尔兹曼机 (Deep Boltzmann machines, DBM) 和深度置信网络 (Deep belief network, DBN) 等; 文献 [18] 梳理传统自编码器 (Auto-encoder, AE) 模型及其衍生变体模型的研究现状、分析其存在的问题与挑战和展望未来的发展趋势; 文献 [19–21] 概述生成对抗网络 (Generative adversarial networks, GAN) 的基本思想、梳理相关理论与应用研究; 文献 [22] 根据似然函数处理方法对深度生成模型进行分类, 包括基于受限玻尔兹曼机 (Restricted Boltzmann machines, RBM)、变分 AE (Variational AE, VAE) 的近似方法<sup>[23]</sup>、能够避免求极大似然过程的诸如 GAN 的隐式方法、对似然函数进行适当变形的流模型和自回归模型; 文献 [24] 介绍基于去噪扩散概率模型 (Denoising diffusion probabilistic models, DDPMs)<sup>[25–26]</sup>、噪声条件分数网络 (Noise conditioned score networks, NCSNs)<sup>[27]</sup> 和随机微分方程 (Stochastic differential equations, SDEs)<sup>[28]</sup> 3 种通用扩散模型框架, 并讨论与其他深度生成模型的关系. 但是, 这些文献综述主要聚焦于图像生成在计算机领域的应用, 其核心问题是如何更好地拟合训练集的概率密度分布. 因工业过程具有强污染、多噪声和不确定等特性而使得图像生成更加复杂, 其核心在于: 如何结合过程机理, 借助小样本集“创造”出期望的图像集. 因此, 有必要结合工业过程的实际特性, 针对性地对工业图像生成及其应用研究进行综述.

本文面向实际需求, 对工业过程图像生成、生成图像评估与应用进行综述, 主要贡献包括: 1) 梳理面向工业过程的图像生成技术和工业领域潜在图像生成技术; 2) 结合图像生成领域的研究成果, 面向实际工业过程需求, 依据流程将现有算法从工业图像生成、生成图像评估和应用 3 个方面进行综述; 3) 提出面向工业过程图像生成及其应用的未来研究方向与挑战.

## 1 面向工业过程的图像生成技术

### 1.1 图像生成的定义与分类

图像生成的目标函数如下

$$G_* = \arg \min_G \text{Div}(p_G, p_{\text{data}}) \quad (1)$$

式中,  $G_*$  表示最优生成模型,  $p_G$  和  $p_{\text{data}}$  表示生成数据和真实数据的概率分布,  $\text{Div}(\cdot)$  表示散度.

由式 (1) 可知, 图像生成的定义为: 寻找生成模型参数, 使生成的数据与真实的数据概率分布的散度最小. 本文给出如图 1 所示的深度生成模型分类框架.

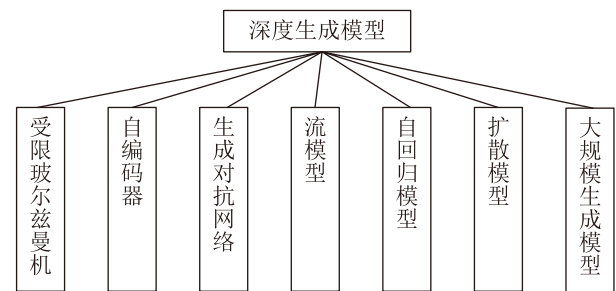


图 1 深度生成模型分类

Fig. 1 Classification of deep generation model

由图 1 可知, 深度生成模型包括: 1) 受限玻尔兹曼机<sup>[29]</sup> 及其为基础模块的深度置信网络<sup>[30]</sup>、深度玻尔兹曼机<sup>[31]</sup> 等模型; 2) 自编码器及其改进模型; 3) 生成对抗网络<sup>[32]</sup> 以及改进模型; 4) 以非线性独立分布估计 (Non-linear independent components estimation, NICE) 为基础的常规流 (Normalizing flow) 模型<sup>[33]</sup> 及其改进模型; 5) 包括神经自回归密度估计 (Neural autoregressive distribution estimation, NADE)<sup>[34]</sup>、像素循环神经网络 (Pixel recurrent neural network, PixelRNN)<sup>[35]</sup>、掩码 AE 分布估计 (Masked AE for distribution estimation, MADE)<sup>[36]</sup> 以及 WaveNet<sup>[37]</sup> 等在内的自回归模型; 6) 扩散模型及其改进模型; 7) 以 ChatGPT 和 GPT-4 为代表的大规模生成模型.

在上述模型中, 用于图像生成的 GAN、AE、流模型和扩散模型的论文出版情况如图 2 ~ 5 所示.

### 1.2 面向工业过程的图像生成定义

面向工业过程的图像生成任务可表示为

$$G_*^{\text{ind}} = \arg \min_G \text{Div}(p_G^{\text{ind}}, p_{\text{data}}^{\text{ind}}) \quad (2)$$

式中,  $G_*^{\text{ind}}$  表示最优的生成模型,  $p_G^{\text{ind}}$  表示工业过程生成数据的概率分布,  $p_{\text{data}}^{\text{ind}}$  表示真实数据的概率

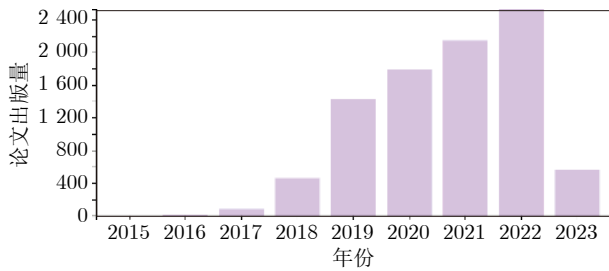


图 2 GAN 模型论文出版情况

Fig.2 Publication status of GAN model

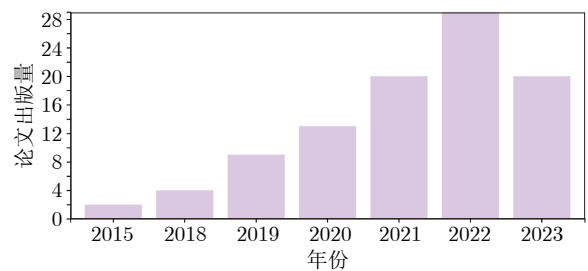


图 5 扩散模型论文出版情况

Fig.5 Publication status of diffusion model

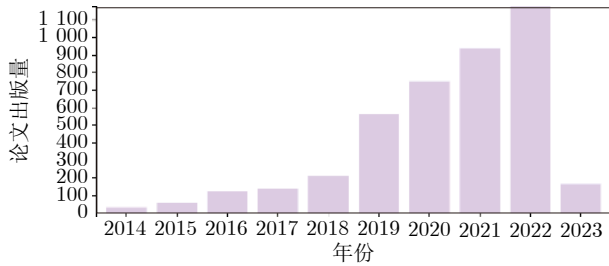


图 3 AE 模型论文出版情况

Fig.3 Publication status of AE model

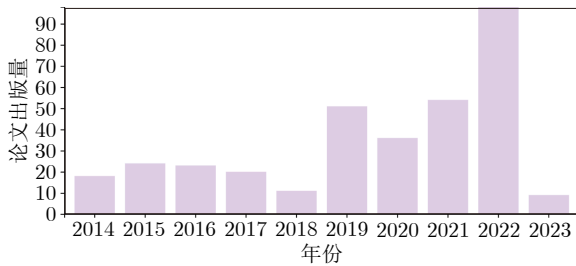


图 4 流模型论文出版情况

Fig.4 Publication status of flow-based model

分布。

由式 (2) 可知, 当训练集足够大时, 从训练集中抽样获取的  $p_{data}^{ind}$  能够近似表征真实数据的分布, 其难点在于如何定义  $Div(\cdot)$  能够使模型得到有效的训练. 面向工业过程, 由于样本的稀缺, 除了存在上述问题外, 如何表征  $p_{data}^{ind}$  也是难以解决的问题, 需要针对性地采用不同的策略予以解决.

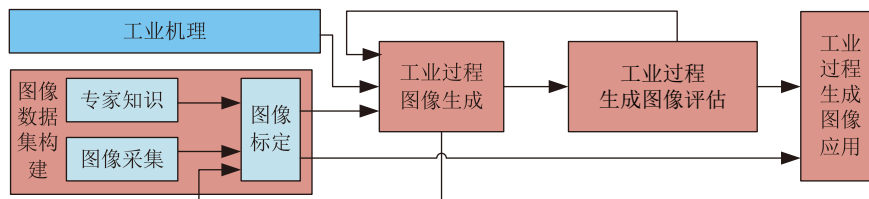


图 6 面向工业过程的图像生成及应用流程

Fig.6 Image generation and application process for industrial process

### 1.3 面向工业过程的图像生成及应用流程

面向复杂工业过程的图像生成及其应用流程如图 6 所示。

由图 6 可知, 具体流程为: 首先, 构建图像数据集, 主要包括图像采集和基于专家知识的图像标定; 然后, 针对真实图像数据存在的问题, 结合工业机理构建图像生成模型; 接着, 定性或定量评估生成图像的质量和多样性并选择合格图像, 若再次进行图像生成仍不满足要求, 重新通过图像采集和图像标定构建图像数据集; 最后, 构建基于生成和真实图像的数据集以进行生成图像应用。

本文重点关注工业过程图像的生成模型构建、生成图像评估和应用等方面。

### 1.4 面向工业过程的生成图像评估框架

虽然生成模型, 特别是 GAN, 得到了广泛的关注, 但如何对所生成的图像进行评估和选择仍然是待解决的开放性问题. 面向生成图像的评估模型架构<sup>[38]</sup>如图 7 所示。

传统生成图像的评估框架为: 先将真实图像集和生成图像集进行特征提取, 再对所提取的特征向量进行度量计算. 该评估框架涉及多种不同的特征提取网络和度量准则. 以真实图像集  $X_r$ 、生成图像集  $X_g$  和特征提取器为输入, 以度量准则 FID (Fréchet inception distance) 值为输出的评估过程为: 首先, 加载特征提取器提取两个图像集的特征矩阵  $z_r$  与  $z_g$ ; 然后, 计算特征矩阵的多元正态分布均值  $\mu_r$

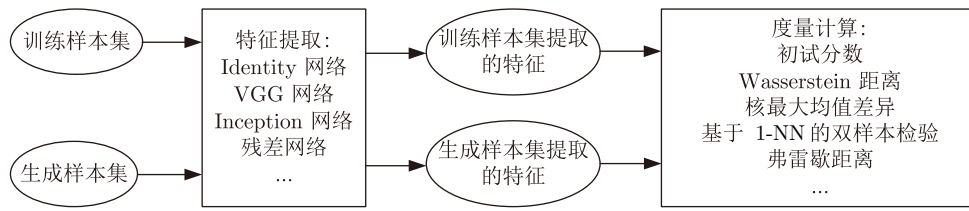


图 7 对生成图像的评估架构

Fig. 7 Evaluation architecture for generated images

与  $\mu_g$  以及协方差矩阵  $Cov_r$  与  $Cov_g$ ; 接着, 计算矩阵的迹  $\text{Tr}(\cdot)$ ; 最后, 根据式 (3) 计算 FID 值

$$\text{FID} = \|\mu_r - \mu_g\|^2 + \text{Tr}(Cov_r + Cov_g - 2(Cov_r \times Cov_g)^{\frac{1}{2}}) \quad (3)$$

这类算法旨在度量生成数据集与真实数据集之间的相似度。

### 1.5 面向工业过程的图像生成应用需求

目前, 卷积神经网络 (Convolutional neural network, CNN) 是视觉领域的主要研究方向. 例如, 文献 [39] 结合 CNN 和图像分割定位变压器内部热缺陷的故障; 文献 [40] 改进 CNN 以预测设备部件的坐标、方向角和类别类型; 文献 [41] 在 CNN 中引入局部聚集描述符向量以增加特征表示的鲁棒性和增强识别模型的精度. 但是, 以 CNN 为代表的监督网络模型的准确率常取决于训练样本标签的质量与规模.

工业过程的图像采集设备长期处于强干扰环境中, 这导致图像的获取和标定存在困难<sup>[14]</sup>; 此外, 数据的不均衡分布也是工业过程中的常见问题<sup>[42-43]</sup>. 诸多研究表明, 在数据分布不平衡的情况下, 数据增强处理有助于提高模型性能<sup>[44-46]</sup>. 传统数据增强是通过几何变换 (如平移、缩放和旋转) 和通道变换合成图像<sup>[47]</sup>, 其局限性在于无法真实地创建新的样本, 所生成的样本仍然受限于原始数据的范围和特征. 图像生成技术采用 GANs、VAEs 或 DDPMs 等生成模型, 能够更加逼真地生成新样本, 结合工业领域特有的机理知识, 理论上能够创造出更丰富、更贴近真实的数据样本, 能够扩展数据的多样性和覆盖范围, 进而能够在训练过程中更好地捕捉数据分布的细微特征以提升模型的泛化能力.

针对复杂工业过程中的图像存在可解释性差、干扰性强、标记成本高等问题, 其图像生成方法可从以下角度进行分析研究: 1) 样本分布不均问题, 正常和异常数据分布存在严重偏差<sup>[48]</sup>; 2) 样本多样性问题, 极端异常的图像缺失; 3) 样本可解释性问题, 特定图像在不同程度上与工业机理相关.

## 2 工业领域潜在图像生成相关技术

GAN、VAE、流模型、PixelRNN 和扩散模型等算法及其变体在工业领域的图像生成中均具有潜在应用前景.

### 2.1 GAN

GAN 由生成器  $G$  和判别器  $D$  组成, 前者通过随机噪声  $z$  生成图像, 后者判断输入图像为真的概率. 具体而言,  $G$  与  $D$  是相互竞争的判别过程和欺骗过程, 前者为  $D$  判别图像真假时的参数更新过程, 后者为  $G$  企图欺骗  $D$  时通过  $D$  的损失更新  $G$  的过程. GAN 的目标函数如下

$$\min_G \max_D V(D, G) = \mathbb{E}_{\mathbf{x} \sim p_r(\mathbf{x})} (\log_2(D(\mathbf{x}))) + \mathbb{E}_{z \sim p_z(z)} (\log_2(1 - D(G(z)))) \quad (4)$$

式中,  $V(D, G)$  表示真实数据与生成数据的差异程度, 下标  $r$  表示真实数据,  $p_r$  和  $p_z$  表示真实数据的概率分布和  $z$  服从的高斯分布.

为便于 GAN 的数学描述, 采用下标  $g$  表示生成数据的概率分布. 假设真实数据和生成数据的概率分布  $p_r$  和  $p_g$  为定值,  $D$  可拟合任意函数.

首先, 考虑任何给定  $G$ , 求解最佳  $D$ , 即  $D_*$ . 训练  $D$  时, 固定  $G$  的参数, 在  $\max_D V(D, G)$  的过程中,  $D_*$  表示如下

$$D_*(\mathbf{x}) = \frac{p_r(\mathbf{x})}{p_r(\mathbf{x}) + p_g(\mathbf{x})} \quad (5)$$

然后, 假设每轮  $D$  均是最优的且  $G$  可拟合任意函数. 此时, 固定判别网络参数, 更新生成网络参数,  $\min_G V(D, G)$  的结果如下

$$\min_G V(D, G) = \min_G (2D_{\text{JS}}(p_r || p_g) - 2 \ln 2) \quad (6)$$

式中,  $D_{\text{JS}}$  为 JS (Jensen-Shannon) 散度. 由式 (6) 可得  $p_g = p_r$  为最优解, 即生成器能够拟合真实数据的概率分布.

最后,  $G$  根据其学习到的概率分布生成符合真实数据概率分布的新样本.

GAN 利用判别器的特性避开了求解似然函数的复杂过程, 这使得 GAN 非常灵活且适用性强, 其代表模型包括 DCGAN、WGAN 和 BigGAN 等. GAN 的变体模型如表 1 所示.

## 2.2 VAE

VAE 由编码器和解码器组成, 其核心理念是学习潜在的空间. 理论上, 空间中的每个点均对应着数据的一个潜在表示, 进而可在空间中进行数据的插值、生成和探索. VAE 的简要过程为: 首先, 编码器将输入数据映射到潜在空间中的概率分布, 即将输入数据转换为潜在变量的均值  $\mu$  和标准差  $\sigma$  以

表征该分布; 然后, 采用重参数化技巧对该分布进行可微分采样以获得潜在变量; 最后, 解码器接收在潜在空间采样的潜在变量并将其映射回原始数据空间, 进而生成一个与原始输入数据相似的样本.

VAE 采用变分下界作为其优化目标函数, 如下

$$L_{VAE} = L_{recon} + L_{reg} \quad (7)$$

式中,  $L_{recon}$  表示重构损失, 其常采用均方差 (Mean squared error, MSE) 或交叉熵 (Cross-entropy) 作为度量准则以衡量生成样本与原始输入数据间的差异; 设  $\mathbf{x}$  为输入数据,  $\hat{\mathbf{x}}$  为重构数据, 则重构损失可表示为  $MSE(\mathbf{x}, \hat{\mathbf{x}})$  或  $Cross-entropy(\mathbf{x}, \hat{\mathbf{x}})$ .

表 1 基于 GAN 的变体  
Table 1 Variants based on GAN

序号	模型名称	主要贡献	文献与年份
1	cGAN	将标签信息作为附加信息输入到生成器中, 并将其与生成样本一同输入到判别器中, 进而增强生成样本与标签之间的关联性.	[49], 2014
2	DCGAN	采用卷积网络作为生成器和判别器以及采用无监督的训练方法.	[50], 2015
3	LAPGAN	基于拉普拉斯金字塔结构逐层增加样本分辨率, 上层高分辨率图像的生成以下层低分辨率图像作为条件.	[51], 2015
4	VAE-GAN	结合 VAE 和 GAN 的混合模型. VAE 用于学习输入数据的潜在空间表示, GAN 中的判别器用于学习两个概率分布之间的相似度.	[52], 2016
5	BiGAN	采用两个生成器和两个判别器学习训练数据的潜在表示和生成新数据. 其中, 一个生成器将随机噪声映射到数据空间, 另一个生成器将数据映射到潜在空间. 相应地, 两个判别器分别评估从潜在空间到数据空间和从数据空间到潜在空间的一对样本.	[53], 2016
6	CoGAN	提出新的联合训练方法和共享权重策略, 能够同时学习多个领域之数据空间和从数据空间到潜在空间并且能够生成跨域图像. 1) 联合训练: 用于同时训练多个 GAN, 每个 GAN 对应一个领域. 通过联合训练, CoGAN 可以学习多个领域之间的相关性, 并且可生成跨域图像. 2) 共享权重: CoGAN 的生成器和判别器之间共享权重, 可共同学习多领域间的相关性提高模型的泛化能力. 此外, CoGAN 的生成器和判别器也可共享一部分权重, 进而减少模型参数的数量.	[54], 2016
7	Info-GAN	引入了信息理论的概念, 使得 GAN 能够更有效地学习到有意义的表示和结构化的表示. 1) 引入信息瓶颈: 从随机噪声向量中提取有意义的信息, 并将其与潜在变量结合生成图像, 使得 GAN 可对生成图像中的信息进行控制, 例如生成特定的数字或对象. 2) 信息瓶颈的优化: 通过最大化互信息 (Mutual information, MI) 优化信息瓶颈. 具体地, 通过在训练过程中最大化生成数据和信息瓶颈的 MI, 即目标函数是最大化生成数据与潜在编码间 MI 的同时最小化生成器输出与噪声间的 MI, 从而实现生成图像中信息的控制.	[55], 2016
8	f-GAN	证明了任意散度都适用于 GAN 框架.	[56], 2016
9	Improved-GAN	采用多种方法对 GAN 的稳定性和生成效果进行进一步加强. 1) 同步批量标准化 (Synchronized batch normalization, SyncBN): 将生成器和判别器的训练过程同步, 从而减少训练过程中的不稳定性, 提高训练效率和生成图像的质量. 2) 动量梯度下降 (Momentum gradient descent, MGD): 采用 MGD 代替传统的随机梯度下降 (Stochastic gradient descent, SGD) 训练生成器和判别器, 加速训练过程, 减少震荡和不稳定性. 3) One-sided label smoothing: 将真实样本的标签从 1 降低到 0.9, 减少判别器对真实样本的过度自信, 提高训练过程的稳定性和鲁棒性. 4) Spectral normalization: 限制判别器中权重矩阵的最大奇异值, 减少模式崩溃的风险, 提高生成图像的多样性和质量.	[57], 2016
10	WGAN-GP	将判别器的梯度作为正则项加入到判别器的损失函数中.	[58], 2017
11	ACGAN	生成图像类别的控制. 1) 实现生成图像类别控制: 在 cGAN 的基础上增加判别器以预测生成图像的类别, 在生成图像的同时学习图像的分类能力. 2) 改善生成效果: 通过引入类别信息提高 GAN 的生成效果和多样性, 生成器和判别器相互协作使得生成的图像具有高质量以及不同的类别和特征. 3) 推广应用: 不仅生成图像, 还可应用于其他数据类型, 例如声音和文本数据.	[59], 2017

表 1 基于 GAN 的变体 (续表)  
Table 1 Variants based on GAN (continued table)

序号	模型名称	主要贡献	文献与年份
12	StackGAN	<p>基于多阶段生成更高分辨率和更逼真的图像。</p> <p>1) 多阶段生成: 第一阶段生成低分辨率图像, 第二阶段将低分辨率图像转化为高分辨率图像, 可使得生成器更容易学习到复杂的图像结构和细节信息。</p> <p>2) 条件 GAN 结构: 将类别信息嵌入到生成器和判别器中, 能够生成指定类别的图像。同时引入文本信息作为条件, 可根据文本描述生成图像。</p> <p>3) 特征金字塔 (Feature pyramid): 采用特征金字塔结构, 可同时学习不同分辨率和不同层次的图像特征, 从而生成更加逼真和细致的图像。</p> <p>4) 应用广泛: 可应用于不同的数据集和任务, 例如自然图像生成、文本到图像生成等。</p>	[60], 2017
13	BigGAN	<p>实现了高分辨率图像的生成和模型的可解释性。</p> <p>1) 大规模模型: 最大的生成对抗网络模型之一, 具有高度可扩展性和并行性, 可生成高分辨率的真实感图像。采用分层架构、条件归一化、投影判别器和分布式训练等技术, 可在不增加训练时间的情况下生成更高质量的图像。</p> <p>2) 可解释性: 提供类向量插值的图像生成方式, 可在类别之间进行平滑过渡, 生成更具艺术价值的图像。可通过对类别和噪声向量的操纵控制生成图像的特定属性, 如颜色、纹理和形状等, 增强模型的可解释性和应用性。</p> <p>3) 领域拓展: 可用于各种数据类型的应用场景, 如图像生成、自然语言生成和音频合成等。</p>	[61], 2018
14	SAGAN	<p>引入自注意力机制使得生成图像具有全局一致性和结构性。</p> <p>1) 自注意力机制: 在生成器和判别器中引入自注意力模块, 可在不同空间位置上学习到图像的相互依赖关系, 使生成的图像更具有全局一致性和结构性。自注意力机制可看作是对局部区域的特征加权融合, 得到更具有语义信息的全局特征表示。</p> <p>2) 混合正则化: 批归一化和实例归一化相结合, 从而有效地解决生成器和判别器中的内部协变量偏移问题, 提高模型的鲁棒性和稳定性, 减少训练时间和计算成本。</p> <p>3) 高分辨率图像生成: 可生成更具有真实感和艺术价值的高分辨率图像, 用于许多实际场景, 例如人脸生成、自然场景生成、图像修复和视频生成等。</p>	[62], 2019
15	LSGAN	采用最小二乘损失函数, 可将图像的分布尽可能接近决策边界。	[63], 2020
16	DivCo	<p>引入对比学习, 增强条件图像生成的多样性。</p> <p>1) 引入对比损失: 采用对比学习以增加生成图像多样性。</p> <p>2) 潜在增强对比损失: 以对比的方式区分生成样本的潜在表征, 进而使得模式崩溃问题得到缓解。</p>	[64], 2021
17	Semanticspatial aware GAN	<p>语义空间感知 GAN, 通过联合训练语义分割网络和生成网络, 生成更真实、更多样且与输入条件更一致的图像。</p> <p>1) 引入语义信息: 通过使用语义分割网络提取输入条件的语义信息, 使生成器更好地理解输入条件。</p> <p>2) 对齐特征图: 通过将语义分割网络的特征图与生成器的特征图进行对齐, 使生成器更好地利用语义信息。</p> <p>3) 空间感知损失: 引入空间感知损失以保持生成图像与输入条件的空间一致性, 从而进一步提高生成器的性能。</p>	[65], 2022
18	RCF-GAN	提出互相对偶的特征方程 GAN (Reciprocal GAN through characteristic functions, RCF-GAN), 进而能够学习到有意义的嵌入空间, 能够避免在数据域中使用均方差所产生的平滑伪影, 进而捕捉图形数据之间固有的关系。兼具自编码器和 GAN 的优点, 即能够双向生成清晰图像。	[66], 2023
19	WGAN	从理论上分析 GAN 训练不稳定的原因, 通过采用 Wasserstein 距离等方法提高训练稳定性。	[67], 2017
20	SNGAN	<p>基于谱归一化 (Spectral normalization, SN) 的 GAN, 提高了生成器和判别器的稳定性和性能。</p> <p>1) SN: 采用 SN 技术对判别器的权重矩阵进行处理, 控制判别器范数大小, 提高稳定性和泛化能力, 减少训练过程中梯度爆炸和消失的问题, 使得训练更加稳定和快速。</p> <p>2) 训练技巧: 采用批次训练、增强数据和生成器先行等训练技巧, 使得生成器和判别器可以更好地学习数据分布和特征信息。生成器先行技巧可使得生成器更容易学习到真实数据的特征, 从而生成更高质量的图像。</p>	[68], 2018
21	PGGAN	<p>实现高分辨率图像的生成和训练过程的稳定性。</p> <p>1) 逐步生长: 从低分辨率图像开始训练, 逐渐增加分辨率, 增加模型的深度和复杂度, 可生成高分辨率的真实感图像和提高模型的稳定性与可训练性。</p> <p>2) 非线性映射: 采用非线性映射技术将噪声向量转换为高维的潜在空间向量, 增强模型的表达和生成能力, 可学习到更复杂的图像特征和结构, 生成更具艺术价值的图像。</p> <p>3) 归一化技术: 采用像素归一化技术, 平衡不同分辨率图像的亮度和对比度, 减少训练过程中的内部协变量偏移问题, 提高模型的可训练性和生成能力, 减少训练时间和计算成本。</p>	[69], 2018

在式 (7) 中,  $L_{\text{reg}}$  表示正则化项, 通常是通过最小化 KL (Kullback-Leibler) 散度  $D_{\text{KL}}$  约束潜在变量分布与预定义先验分布间的相似性, 进而使得潜

在空间具有平滑性。设  $q(\mathbf{z}|\mathbf{x})$  是给定输入数据  $\mathbf{x}$  时潜在变量  $\mathbf{z}$  的后验分布,  $p(\mathbf{z})$  是预定义的先验分布, 则  $L_{\text{reg}}$  表示如下

$$L_{\text{reg}} = D_{\text{KL}}(q(\mathbf{z}|\mathbf{x})||p(\mathbf{z})) = -\frac{1}{2} \sum_{i=1}^N (1 + \ln(\sigma_i^2) - \mu_i^2 - \sigma_i^2) \quad (8)$$

式中,  $\sigma_i$  和  $\mu_i$  分别表示编码器输出的第  $i$  个潜在变量的均值和标准差。

VAE 的训练过程旨在学习适当的潜在表示和解码器, 进而使得样本能够在潜在空间中平滑地插值并进行生成, 其代表模型包括重要性加权自编码

器和辅助深度生成模型等. VAE 的变体类型如表 2 所示.

### 2.3 流模型

流模型通常由多个能够进行可逆变换的耦合层组成, 其核心理念是通过逐层的可逆变换将简单的先验分布映射为复杂的数据分布, 进而实现对数据分布的建模. 耦合层由变换层 (Transform layer) 和标量变换层 (Scalar transform layer) 组成, 前者

表 2 基于 VAE 的变体  
Table 2 Variants based on VAE

序号	方法名称	主要贡献	文献与年份
1	CVAE	<p>将类别信息引入潜在空间的表示中, 能够生成特定类别的数据.</p> <ol style="list-style-type: none"> <li>1) 引入类别信息: 将条件信息作为额外输入, 并与潜在变量合并后作为一个新的潜在表示.</li> <li>2) 采用重参数技巧: 允许模型进行梯度反向传播, 可有效训练深度生成模型, 即将潜在在变量转换为确定性变量和随机噪声的乘积.</li> <li>3) 最大化后验概率: 采用最大后验估计训练模型, 最大化给定输入和标签的后验概率, 可有效控制生成的数据符合特定的类别信息.</li> </ol> <p>采用对抗训练的方式实现无监督学习和数据生成, 同时将潜在在变量编码成固定的噪声分布, 使得数据具有可解释性.</p> <ol style="list-style-type: none"> <li>1) 无监督学习: 在无监督情况下学习数据表示, 可在不需要标签的情况下生成新数据, 对于诸多现实问题非常有效.</li> <li>2) 对抗训练: 采用对抗训练学习数据的表示和生成, 能够学习到数据分布的本质特征, 对于训练的稳定性有帮助.</li> <li>3) 潜在在变量编码: 将潜在在变量编码为固定的噪声分布, 而不是传统的正态分布, 能够更好地控制生成器的输出, 进而使得生成的数据更加多样化.</li> <li>4) 可解释性: 所学习到的数据表示具有可解释性, 其由一个编码器生成, 可通过调整潜在在变量控制生成器生成的数据, 这种可解释性对图像修复、图像生成、数据增强等应用非常有价值.</li> </ol>	[70], 2015
2	AAE	<p>提高变分自编码器的似然下界 (Evidence lower bound, ELBO) 的上界, 能够更准确地估计后验分布.</p> <ol style="list-style-type: none"> <li>1) 提高 ELBO 的上界: 提出新的 ELBO 上界, 通过对潜在在变量的重要性权重进行平均, 提高原始 ELBO 的上界, 使得对数似然估计更加准确.</li> <li>2) 更准确的后验分布估计: 通过引入具有多个重要性的采样样本, 准确地估计后验分布, 提高模型的生成能力.</li> <li>3) 自适应重要性采样: 选择重要性权重可更准确地估计 ELBO 上界和后验分布, 提高模型的生成能力.</li> <li>4) 基于平均的方法: 计算重要性权重, 可更准确地估计后验分布, 提高模型生成能力.</li> </ol>	[71], 2015
3	IWAE	<p>可从输入图像中逆推出图像中的物体形状、位姿、材质等信息, 可用于图像编辑和重构.</p> <ol style="list-style-type: none"> <li>1) 逆图形生成: 使得模型可理解图像中的物体结构和属性.</li> <li>2) 图像编辑: 可采用逆推出的物体形状、位姿、材质等信息对输入图像进行编辑, 修改图像中的物体属性.</li> <li>3) 高效的网络结构: 采用高效的卷积神经网络结构, 可在较短的时间内学习到图像中的物体信息, 快速地生成新的图像.</li> <li>4) 数据集构建: 可学习到多种物体的形状、位姿、材质等信息, 可在不同的视角下观察物体, 提高模型生成能力.</li> </ol>	[72], 2015
4	DC-IGN	<p>可更好地学习层次化的特征表示, 在生成样本的同时学习特征的表达方式.</p> <ol style="list-style-type: none"> <li>1) 层次化的结构: 可学习更加丰富和复杂的特征表示, 可生成更加准确和多样化的样本.</li> <li>2) 递归推断: 可在学习低层次特征表示的同时学习高层次特征表示, 使得模型可学习更加完整和准确的特征表示.</li> <li>3) 共享权重: 不同层次之间的特征表示可共享信息, 减少模型参数数量, 提高模型泛化能力.</li> </ol>	[74], 2016
5	LVAE	<p>新的模型结构和优化方法, 可利用有标注数据 and 无标注数据进行训练, 在半监督学习中性能优秀.</p> <ol style="list-style-type: none"> <li>1) 无监督和有监督的 VAE 结构: 无监督的 VAE 用于无标注数据的特征提取, 有监督的 VAE 用于有标注数据的特征提取.</li> <li>2) 优化方法: 采用 SGD 和重参数化技术优化无监督和有监督的 VAE, 采用一个判别器优化整个模型.</li> <li>3) 应用于半监督学习: 性能优于其他经典的半监督学习算法, 例如半监督支持向量机 (Semi-supervised SVM) 和 DBN 等.</li> </ol>	[75], 2017
6	SSVAE	<p>通过最大化信息瓶颈提高 VAE 的信息提取能力.</p> <ol style="list-style-type: none"> <li>1) 信息瓶颈目标函数: 通过最大化信息瓶颈提高 VAE 的信息提取能力, 该目标函数的重构误差项用于保证模型能够还原原始数据, 互信息项用于最大化编码器和解码器之间的互信息.</li> <li>2) 优化方法: 采用 SGD 和重参数化技术优化模型, 采用额外的网络估计互信息项.</li> </ol>	[76], 2017
7	infoVAE		

表 2 基于 VAE 的变体 (续表)

Table 2 Variants based on VAE (continued table)

序号	方法名称	主要贡献	文献与年份
8	MSVAE	新的多尺度结构可同时处理不同尺度的信息, 并且能够生成高质量的图像。 1) 多尺度结构: 包括一个全局编码器和多个局部编码器, 每个局部编码器对应一个尺度, 该结构可生成高质量图像。 2) 融合策略: 将不同尺度的信息融合以生成最终的图像, 首先采用全局编码器生成一个潜在向量, 然后采用多个局部解码器将潜在向量解码为不同尺度的图像, 最后通过融合得到最终的图像。	[77], 2017
9	RVAE	采用递归变分自编码器 (Recurrent variational autoencoders, RVAE) 学习非线性生成模型, 能够在存在异常值的情况下对数据进行建模。	[78], 2018
10	CIVAE	结合自编码器 s 和变分自编码器的思想, 并在此基础上引入自省机制, 用于生成高质量的、多样化的、可控的图像。 1) 自省机制: 帮助模型自监督和修正, 计算每个样本的重构误差和 KL 散度, 将这些信息用于调整模型的参数, 使模型在生成图像时更稳定和准确。 2) 条件生成: 可接受额外的条件信息, 如标签或文本描述, 用于控制生成图像的特征和风格。 3) 多样性: 通过引入随机噪声扰动隐含变量, 从而生成具有多种不同特征的图像。	[79], 2020
11	DALL-E	基于 Transformer 的生成模型, 通过将文本描述编码为低维向量, 采用解码器将该向量转化为图像: 采用大规模的无监督数据集, 通过最大化 ELB (Evidence lower bound) 的策略优化模型参数, 实现生成与训练数据中的不同图像。	[80], 2021
12	DALL-E2	采用 CLIP 潜变量进行分层文本条件图像生成。 1) 分层图像生成: 通过引入分层的生成过程, 将图像生成任务分解为多个子任务, 从而提高生成图像的多样性和控制能力。 2) 采用 CLIP 模型的潜变量: 将 CLIP 模型生成的图像嵌入以作为潜变量, 能够捕捉到图像中的语义和风格信息, 并且可通过文本描述控制生成的图像。 3) 文本条件生成: 将文本描述作为输入, 生成与描述相匹配的图像, 并可通过操纵文本描述进而生成不同层次的图像。	[81], 2022
13	VPORN	1) 提出新的潜在特征增强和分布正则化框架, 用于 FSL (Few-shot learning), 包括先验关系网络 (Prior relationship network, PRN) 和基于 VAE 的后验关系网络 (Variational posterior prediction and regularization network, VPORN)。通过 PRN 和 VPORN, 从少量样本中学习到更多的关键类内特征和类间特征。 2) 基于正则化分布估计降低标记数量不足的新颖样本的方差, 更关注关键和独特的特征, 进而避免不可控的转移。	[82], 2023

对部分的潜在变量进行非线性变换, 后者对另外部分的潜在变量进行可逆的线性变换, 进而在保持模型可逆性的同时引入复杂的非线性。由于变换是可逆的, 流模型可实现从数据空间到潜在空间的反向映射, 因此其具有计算潜在变量和生成样本的能力。

流模型的训练目标是最大似然估计, 即最大化训练数据在流模型下的概率。由于流模型结构的可逆性, 可通过变量变换法计算样本的概率密度函数, 其目标函数可表示为如下式所示的负对数似然 (Negative log-likelihood, NLL) 损失

$$L_{\text{Flow}} = -\frac{1}{N} \sum_{i=1}^N \ln p(\mathbf{x}_i) \quad (9)$$

式中,  $L_{\text{Flow}}$  表示负对数似然损失,  $\mathbf{x}_i$  表示第  $i$  个样本,  $N$  表示输入样本的总数。根据变量变换法,  $p(\mathbf{x}_i)$  可表示为  $p(\mathbf{z}_i) \times \left| \det \frac{d\mathbf{z}_i}{d\mathbf{x}_i} \right|$ ,  $\mathbf{z}_i$  表示  $\mathbf{x}_i$  的潜在变量,  $\frac{d\mathbf{z}_i}{d\mathbf{x}_i}$  表示潜在变量与数据样本间的雅可比矩阵。式 (9) 可进一步展开, 如下式所示

$$L_{\text{Flow}} = -\frac{1}{N} \sum_{i=1}^N \ln p(\mathbf{z}_i) - \frac{1}{N} \sum_{i=1}^N \ln \left| \det \frac{d\mathbf{z}_i}{d\mathbf{x}_i} \right| \quad (10)$$

式中, 第一项表示  $\mathbf{z}_i$  在模型先验分布下的负对数似然, 通常假设  $\mathbf{z}_i$  服从多维高斯分布; 第二项为雅可比行列式的负对数, 用于考虑变换从潜在空间到数据空间的缩放。

在实际流模型的训练中, 可采用 SGD 等优化算法最小化负对数似然损失。

综上, 流模型将真实数据分布通过转换函数映射到给定的简单分布, 是一种采用可逆函数构造似然函数、直接优化模型参数和利用可逆结构的特性生成图像的精确模型, 其代表模型是常规流模型、变分流模型和可逆残差网络。流模型类型如表 3 所示。

## 2.4 PixelRNN

PixelRNN 采用循环神经网络 (Recurrent neural network, RNN) 建模像素间的条件概率分布生成图像, 能够实现逐像素的生成, 其关键是捕获了像素之间的横向和纵向依赖关系, 后者能够通过循



表 3 流模型  
Table 3 Flow-based model

序号	名称	主要贡献	文献与年份
1	NICE	作为首个流模型, 提出其基本框架并提出 3 个重要的模型结构层, 即加线耦合层、维数混合层和维数压缩层.	[33], 2014
2	VINF	在 VAE 推断过程引入流模型结构的归一化流变分推断.	[83], 2015
3	Real NVP	在耦合层中引入卷积层, 可更好地处理图像问题, 设计多尺度结构以降低模型的计算量和存储空间.	[84], 2016
4	IAF	将自回归结构的流模型应用在 VAE 变分推断中的模型.	[85], 2016
5	MAF	IAF 的衍生模型, 将 Real NVP 中的掩码卷积层引入到 IAF 中, 能够更好地处理图像样本, 然后提出了条件掩码自回归流 CMAF, 将 MAF 应用到监督模型中.	[86], 2017
6	GLOW	采用可逆变换将简单分布 (如高斯分布) 映射到目标分布, 从而实现高质量的样本生成和概率密度估计. 1) 可逆性变换的设计和实现: 为实现高效的样本生成和概率密度估计, 采用特殊的可逆性变换, 即耦合层 (Coupling layer), 将输入数据的一部分作为输出, 另一部分通过函数变换后与输出部分进行结合, 实现了输入与输出之间的可逆性映射, 能够快速计算样本的概率密度. 2) 计算图的优化和并行计算: 为加速计算和提高可扩展性, 采用计算图的优化技术, 如存储复用、内存分配和子图合并等. 此外, 采用并行计算技术以加速模型的训练和推理, 可高效地处理大规模数据集并实现快速的训练和推理. 3) 在图像生成和数据压缩等领域的成功应用.	[87], 2018
7	i-ResNet	以残差网络为基础的生成模型, 利用约束使残差块可逆, 用近似方法计算残差块的雅可比行列式. 无偏差的图像风格转移方法.	[88], 2019
8	ArtFlow	1) 无偏差的图像风格转移: 采用可逆神经流进行图像风格转移, 避免了传统方法中出现的偏差问题. 2) 支持多种风格: ArtFlow 可以同时支持多种风格的转移, 从而使其更加灵活.	[89], 2021

神经网络或其他序列模型建模. 在生成某个像素值时, PixelRNN 考虑了该像素的左边和上边的像素值以及已经生成的像素值.

PixelRNN 通常可采用交叉熵最小化负对数似然损失, 其目标函数如下式所示

$$L_{\text{PixelRNN}} = -\ln P(\mathbf{q}) = -\ln \prod_{i=1}^N P(q_i | q_1, q_2, \dots, q_{i-1}) \quad (11)$$

式中,  $q_i$  表示第  $i$  个像素值,  $N$  表示像素的总数,  $P(q_i | q_1, q_2, \dots, q_{i-1})$  表示在给定生成的像素值情况下预测当前像素值的条件概率.

由于每个像素的生成均依赖于已生成的像素, 所以 PixelRNN 的生成速度慢. PixelCNN 通过像素生成过程的并行化能够同时生成一个像素位置的所有通道值, 进而提高了生成速度.

综上, PixelRNN 是将图片的像素作为循环神

经网络的输入, 在本质上是自回归神经网络在图片处理上的应用, 其代表模型是 Row LSTM 和 Diagonal BiLSTM 等. PixelRNN 技术及其变体类型如表 4 所示.

## 2.5 扩散模型

扩散模型由逆流模型和生成模型组成, 训练过程涉及逆扩散和正扩散阶段. 其中, 前者通过逆流模型将真实数据样本逐步转化为噪声样本, 进而使转化后的噪声样本尽可能接近噪声分布; 后者通过生成模型将噪声样本逐步转化为真实数据样本, 进而使转化后的噪声样本逐步逼近真实数据分布. 在生成过程中, 从初始噪声样本开始, 通过多次迭代正向扩散将噪声样本逐步转化为逼真的数据样本, 其每个迭代步骤均涉及逆流模型和生成模型的操作.

在本质上, 扩散模型的训练过程是通过最大似然估计获取生成数据的概率分布, 如下式所示

表 4 PixelRNN 模型  
Table 4 PixelRNN model

序号	名称	主要贡献	文献与年份
1	PixelRNN	将图像的每个像素点当作时间序列信号, 采用深度自回归神经网络预测图片的像素值, 提出 3 种不同结构的深度生成模型: PixelCNN、Row LSTM 和 Diagonal BiLSTM.	[35], 2016
2	PixelSNAIL	采用 PixelCNN 模型生成逐像素递归的图像, 采用了可学习的自注意力机制捕捉像素之间的长距离依赖关系.	[90], 2017
3	Auto-regressive GANs	GAN 中, 生成器采用自回归模型生成连续的视频帧.	[91], 2022

$$L_{\text{Diffusion}} = -\ln p(\mathbf{X}) = -\ln \prod_{t=1}^T p(\mathbf{x}_t | \mathbf{x}_{t-1}, \dots, \mathbf{x}_1) \quad (12)$$

式中,  $\mathbf{x}_t$  表示在时间步  $t$  时的数据样本,  $T$  表示总时间步数,  $p(\mathbf{x}_t | \mathbf{x}_{t-1}, \dots, \mathbf{x}_1)$  表示在给定时间步的数据样本下预测当前时间步数据样本的条件概率分布. 通常, 扩散模型的目标函数会被分解为每个时间步的预测误差, 进而保证可采用诸如均方差或交叉熵等标准的损失函数.

综上, 扩散模型通过控制噪声信号的逐步变化生成数据样本, 并且支持逆向过程和条件生成, 能够产生高质量的样本且具有较高的灵活性和可解释性. 代表性的扩散模型如表 5 所示.

## 2.6 Visual ChatGPT 大规模模型

将 ChatGPT 和多个 SOTA 视觉基础模型连接, 能够在对话系统中理解和生成图片的 Visual ChatGPT 大规模模型, 其详细描述如表 5 所示.

## 3 工业过程图像生成及其评估与应用研究现状

结合本文第 1.3 节给出的流程, 本节将从面向

工业过程的图像生成、生成图像评估和生成图像应用共 3 个方面进行研究现状的综述 (如图 8 所示), 并展开叙述每个方向的子类.

### 3.1 工业过程图像生成研究现状

本文从复杂工业过程图像生成存在问题的视角出发, 将工业过程图像生成从面向样本分布不均、面向多样性不足和面向噪声干扰大等 3 个方向进行综述.

#### 3.1.1 面向样本分布不均的图像生成现状

针对工业过程样本分布不均的问题, 可采用生成器拟合小样本分布以扩充样本数据的策略, 主要包括 VAE、GAN 和混合 VAE 与 GAN 模型等策略.

1) VAE 模型: 通过将原始数据映射至低维表示空间进而有效地捕捉数据特征. 由于在低维空间中表征的数据分布可能会更加的均匀, VAE 有助于减少样本分布不均所导致的问题, 从而有效地提高生成器的性能. 相关的研究包括: 文献 [96] 采用卷积编码器 (Convolution encoder, CE) 进行数据增强和文献 [97] 采用卷积自编码器 (Convolution autoencoder, CAE) 进行数据扩充.

2) GAN 模型: 通过改进 GAN 模型的策略控制样本数量的均衡, 进而实现更好的数据生成效果.

表 5 扩散模型和 Visual ChatGPT 大规模模型

Table 5 Diffusion model and Visual ChatGPT large-scale model

序号	名称	主要贡献	文献与年份
1	NICS	基于梯度估计的生成建模, 其通过得分匹配估计数据分布的梯度, 利用 Langevin 动力学生成样本. 1) 梯度估计: 当数据处于低维流形上时, 梯度未定义且难以估计, 采用不同水平的高斯噪声扰动数据并通过联合估计获得相应的得分, 即构建针对所有噪声水平扰动数据分布的梯度向量场. 2) 样本生成: 采用基于退火 Langevin 动力学方法, 在采样过程逐渐接近数据流形时, 利用与逐渐减小的噪声水平相对应的梯度.	[27], 2019
2	DDPM	引入噪声扩散模型作为生成模型, 通过迭代梯度和高斯噪声生成样本, 采用退火 Langevin 动力学逐渐逼近数据流形, 有效地处理低维流形上的数据生成与原始数据相似的样本.	[26], 2020
3	ILVR	针对 DDMP 生成过程的随机性问题, 提出基于迭代潜变量细化 (Iterative latent variable refinement, ILVR) 条件的方法, 在控制图像生成的同时生成高质量图像.	[92], 2021
4	LDM	该潜在空间扩散模型 (Latent diffusion models, LDMs) 能够在有限的计算资源上进行扩散模型的训练, 同时保持其质量和在潜空间中训练扩散模型. 通过引入交叉注意力层, 将扩散模型转变为能够处理通用条件输入 (如文本或边缘框) 的强大且灵活的生成器; 通过在潜空间中进行训练能够达到细节保留与复杂度降低的平衡点, 进而实现计算需求降低和高分辨率图像合成.	[93], 2021
5	ADM-G	采用 UNet 结构的扩散模型, 其通过增加模型的深度和宽度以使得模型的大小保持相对恒定; 增加了注意力机制的 Heads, 采用 $32 \times 32$ 、 $16 \times 16$ 和 $8 \times 8$ 的分辨率进行注意力计算; 采用 BigGAN 残差模块进行上采样和下采样; 通过大量消融实验, 能够在 LSUN 和 ImageNet $64 \times 64$ 的图像生成效果上达到 SOTA, 打破 GANs“垄断”.	[94], 2022
6	Visual ChatGPT	基于视觉基础模型, 通过对话、绘画和编辑等方式进行交互. 1) 视觉基础模型: 能够处理图像输入, 通过图像与文本的联合建模进而能够更好地理解和生成与图像相关的文本. 2) 多模态对话生成: 通过将视觉输入与文本对话相结合实现多模态对话的生成, 使得模型能够根据图像内容生成相关的回复. 3) 绘画和编辑交互: 模型支持用户通过绘画和编辑改变生成结果, 提供了更加直观和灵活的控制方式.	[95], 2023

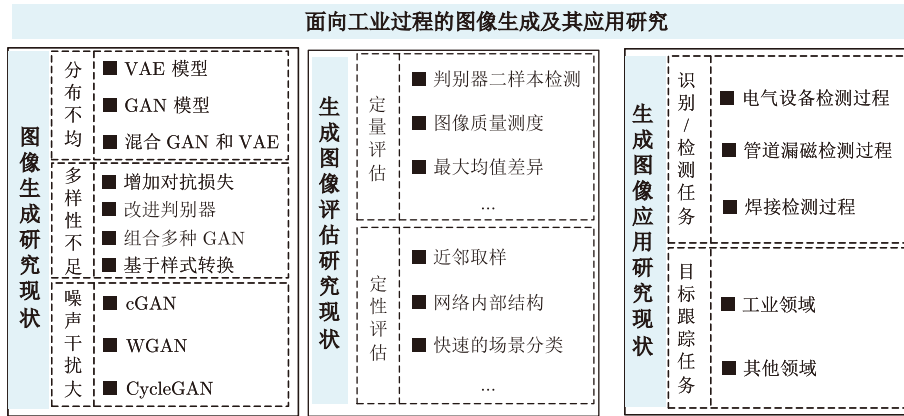


图 8 面向工业过程的图像生成及其应用研究现状结构图

Fig. 8 Structure diagram of the current research status of image generation and its application on industrial process

相关研究如下文所示。

文献 [98] 提出了自适应平衡生成网络 (Adaptive balance GAN, AdaBalGAN), 其创新点包括: 改进条件 GAN (Condition GAN, cGAN) 以生成高保真度的模拟晶圆图并对缺陷类别进行分类, 设计自适应生成控制器后根据分类准确性平衡每种缺陷类型的样本数量. 该对抗过程分为两个阶段: 训练生成器和判别器用来生成指定类别的高保真晶圆图; 训练分类器用来准确识别真实或合成晶圆图的缺陷模式, 其结构图如图 9 所示。

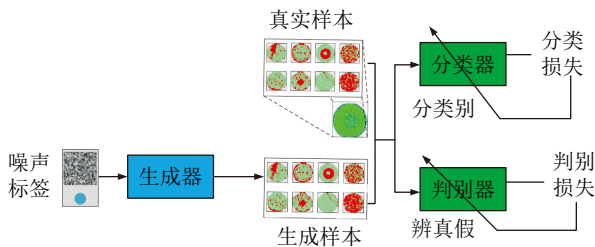


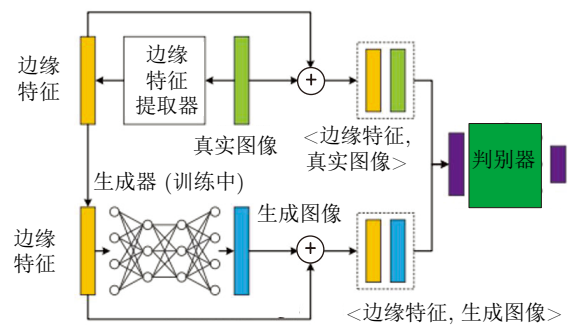
图 9 AdaBalGAN 结构图<sup>[98]</sup>

Fig. 9 Structure of AdaBalGAN<sup>[98]</sup>

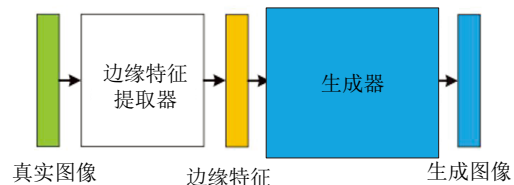
文献 [99] 开发了面向边缘的 GAN (Edge-oriented GAN, EOGAN) 以创建逼真的红外图像, 核心是将所提取的边缘特征作为先验知识指导红外图像生成, 其训练和测试过程如图 10 所示。

文献 [100] 提出了旨在增强漏磁信息有效性的改进 cGAN, 在轴向-径向-轴向空间处理融合后的漏磁信息, 利用生成器的损失函数构建多传感器信息进而增强漏磁信息, 提高模型在缺陷生成方面的效果。

文献 [101] 提出将生成样本与实际样本按比例混合后替代生成样本的方式增强模型生成能力的改进 GAN (Improving GAN, IGAN), 能够在一定程度上避免模态崩溃, 其结构如图 11 所示, 其中随机



(a) 训练过程  
(a) Training process



(b) 测试过程  
(b) Testing process

图 10 EOGAN 训练和测试过程<sup>[99]</sup>

Fig. 10 Training and testing process of EOGAN<sup>[99]</sup>

噪声  $z$  服从高斯分布,  $\alpha$  是取值区间在 (0, 1) 内的混合比例系数。

3) 混合 VAE 与 GAN 模型: 通过结合 VAE 的潜在变量建模和 GAN 的数据生成优点, 期望实现更好的数据分布捕捉和生成效果。

面向自动缺陷检测过程, 文献 [102] 提出基于 GAN 的缺陷样本生成框架, 其贡献体现在: 1) 提出区域训练策略, 即在图像局部区域建立损失函数以适用于局部图像到图像的转换任务; 2) 设计编码器-解码器的图像生成器, 能够结合不同尺度的图像特征生成局部缺陷, 同时保持无缺陷区域基本不变; 3) 引入小波对生成图像进行细化, 恢复高频信息,

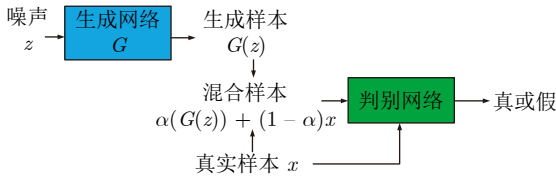


图 11 IGAN 结构<sup>[100]</sup>

Fig. 11 Structure of IGAN<sup>[100]</sup>

避免图像模糊。

综上,上述方法均是通过生成稀缺区域的图像以达到弥补样本分布不平衡的目的,其难点在于如何设计合适的模型结构、网络结构和学习算法等。

### 3.1.2 面向样本多样性不足的图像生成现状

多采用图像到图像的转换,即通过源域特征丰富目标域特征的方式解决此处的样本多样性不足的问题,其核心是基于 GAN 的循环一致性网络。为进一步增加样本的多样性,针对不同问题采取以下方法。

1) 增加对抗损失: 引入能够增加多样性的对抗损失。文献 [103] 提出用于图像生成的表面检测 GAN (Surface defect-GAN, SDGAN), 通过引入 D2 对抗损失<sup>[104]</sup> 增加多样性, 通过采用循环一致损失生成模型学习少量的缺陷样本和大量的无缺陷样本以获得更全面的特征, 其结构如图 12 所示。

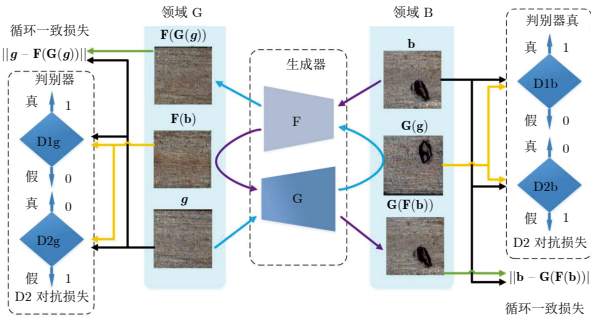


图 12 SDGAN 结构<sup>[103]</sup>

Fig. 12 Structure of SDGAN<sup>[103]</sup>

2) 改进判别器: 增加对细节特征的学习。文献 [105] 针对基于 GAN 的循环一致性结构难以学习到更为丰富的图像特征问题, 提出引入胶囊网络作为判别器以学习细节特征的 DuCaGAN, 其结构如图 13 所示。

3) 组合多种 GAN: 针对不同的工业需求特点采用不同的 GAN 网络。针对自动缺陷检测系统的图像分割任务, 文献 [106] 提出采用 DCGAN 和 CycleGAN 生成缺陷图像、采用 PatchMatch 和周期性空间 GAN (Periodic space GAN, PSGAN) 生成无缺陷合成图像的策略, 如图 14 所示。

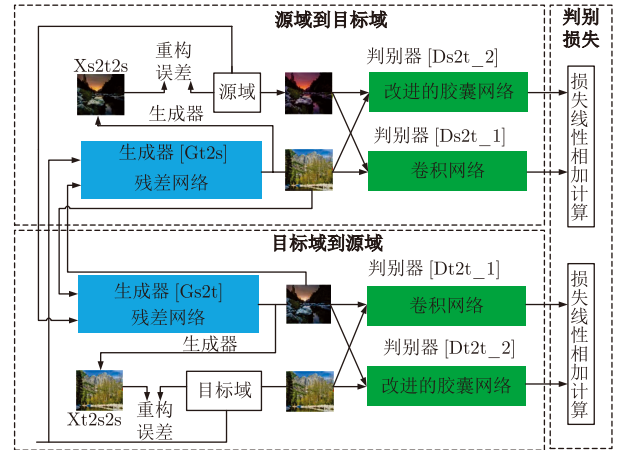


图 13 DuCaGAN 结构<sup>[105]</sup>

Fig. 13 Structure of DuCaGAN<sup>[105]</sup>

4) 基于样式转换: 通过不同部件或背景转换实现主体特征保留和复杂场景处理。

面向焊接过程中的对接板形变预测, 文献 [107] 提出基于 cGAN 产生焊接参数和形变数据以获得新样式。面向夜间环境中的目标检测任务, 针对背景模糊和光线暗淡导致检测困难的问题, 文献 [108] 提出将不易检测的夜间图像转换成易检测的白天图像的策略, 其首先以夜间图像为输入, 基于 GAN 生成与白天环境相似的虚拟目标场景, 再通过深度卷积特征融合和多尺度 ROI (Region of interest) 池化构建基于 Faster R-CNN (Region-convolution neural networks) 目标检测系统。

虽然上述方法能够增加样本的多样性, 但面向工业领域而言, 如何生成符合期望的样本集仍是有待解决的开放性问题, 原因在于特定的工业图像特征通常是与特定的工业机理相关联的。显然, 为更好地生成工业图像样本, 需要更为深入地了解相关机理并将其融入到图像生成过程中。这需要与特定行业的领域专家进行密切合作, 以便更好地生成符合实际需求的样本集。由上述研究现状可知, 此类算法的研究关键点之一是式 (1) 中的  $p_{data}$  难以表征真实数据的概率分布  $p_{real}$ , 即如何表征  $p_{real}$  为新增的难点。

### 3.1.3 面向样本噪声干扰大的图像生成现状

噪声干扰大的样本采用图像采样和转换方法进行处理, 即将强噪声转化为弱噪声或无噪声, 主要包括基于 cGAN、WGAN (Wasserstein GAN) 和 CycleGAN 的方法。

1) 基于 cGAN 的去噪: 模型通过学习从带有噪声的图像到干净图像的映射, 实现图像的噪声去除, 同时利用额外的条件信息, 有助于模型更好地理解

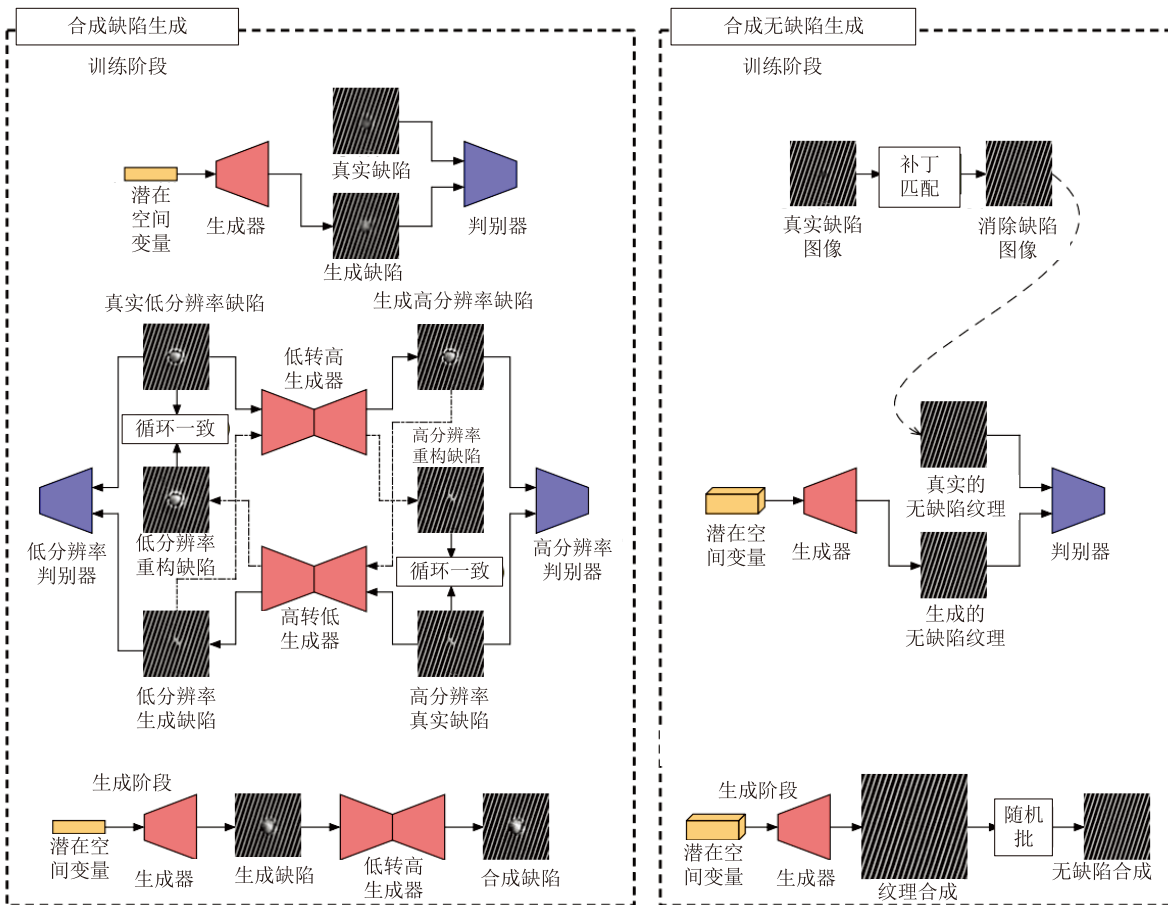


图 14 周期性纹理图像的数据增强过程<sup>[106]</sup>

Fig.14 Data enhancement process for periodic texture image<sup>[106]</sup>

噪声的性质, 从而能够更准确地去除噪声.

文献 [109] 采用由生成器和判别器组成的基于 cGAN 的图像到图像转换模型 pix2pix<sup>[110]</sup> 去除光反射噪声. 其中, 原始噪声图像作为输入, 手动去噪图像作为判别器和生成器的训练目标, 去噪示意如图 15 所示.

文献 [111] 提出了基于 cGAN 的印刷电路板 (Printed circuit boards, PCBs) 图像去噪策略, 其生成器与判别器之间通过对抗性训练的迭代优化提高生成图像的质量.

2) 基于 WGAN 的去噪: 本质上是采用 Wasserstein 距离量化生成图像与真实图像之间的分布差异, 以减少训练中的模式崩溃和梯度消失问题, 从而能够更准确地衡量生成图像的质量, 促使生成器产生更真实干净的图像.

医疗领域, 为了解决细胞图像的模糊性问题, 文献 [112] 提出基于 WGAN 的图像去噪训练框架, 包括生成子网络、基于 MSE 的学习和对抗学习 3 个模块. 其中, 生成子网络模块用于学习噪声图像

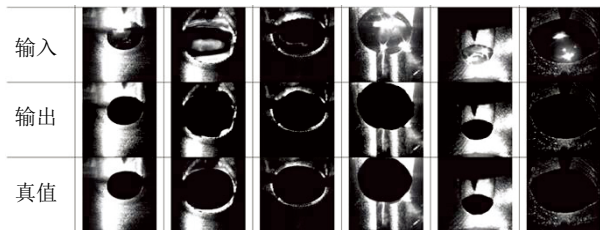


图 15 焊接点去噪示意图<sup>[111]</sup>

Fig.15 Schematic diagram of welding point denoising<sup>[111]</sup>

与去噪图像之间的映射关系, 基于 MSE 的学习模型用于指导生成子网络快速学习映射关系, 对抗学习子模块用于帮助生成子网络学习真实干净图像的分布空间.

3) 基于 CycleGAN 的去噪: 针对具有噪声的原始图像领域和去除噪声后的图像领域, 目标是通过 CycleGAN 的生成器和判别器将噪声图像映射至干净图像领域, 进而实现噪声的去除.

面向无人机航空摄影任务, 文献 [113] 针对因天气原因导致的照片噪声问题, 提出了 SlimR-

GBD (Slim reCNN-GAN) 去噪系统; 针对沙漠地震数据中噪声强且与有效信号的频带存在严重重叠的问题, 文献 [114] 引入 CycleGAN 进行去噪; 针对单幅图像去雾, 文献 [115] 提出基于 CycleGAN 的端到端注意力网络, 其生成器设计除包括注意力模块、编码器-解码器结构和密集块外, 引入了诸如暗通道、颜色衰减和最大对比度等多个先验以获取注意力图, 同时提出颜色损失补偿机制以避免颜色失真.

针对医疗领域, 面向低剂量计算机断层扫描 (Low-dose computed tomography, LDCT) 图像存在的高噪声问题, 文献 [116] 提出了基于 CycleGAN 的图像域去噪; 面向原始光学相干层析成像 (Optical coherence tomography, OCT) 图像的质量问题, 文献 [117] 提出了能够实现视网膜 OCT 图像的端到端散斑抑制和对比度增强的 cGAN 框架, 以解决散斑噪声会模糊视网膜结构、影响视觉质量以及降低后续图像分析任务性能等问题; 进一步, 文献 [118] 提出基于风格转换和 cGAN 的 OCT 图像斑点噪声抑制模型, 其包括: a) 采用 CycleGAN 学习两个 OCT 图像数据集间的样式转移以获得真值数据集; b) 基于 PatchGAN 机制采用小型 cGAN 模型抑制 OCT 图像中的斑点噪声. 可见, 与处理样本噪声相关的技术已在医疗等领域得到广泛应用.

上述研究将强噪声转化为弱噪声或零噪声以提升图像质量和进行信息还原. 在工业实际应用中, 在选择适当的方法时还需要考虑噪声的类型、数据分布和任务需求.

综上所述, 针对上述 3 类工业过程图像生成问题, 所采用的解决策略在本质上是不同的, 对应的数学视角分析如表 6 所示.

面向样本分布不均的问题, 其生成模型和下游任务模型可表示为  $f_{G1}(\mathbf{X}_{\text{training}})$  和  $f_{DT1}(f_{G1}(\mathbf{X}_{\text{training}}), \mathbf{X}_{\text{training}})$ , 对应的不采用  $f_{G1}(\cdot)$  的下游任务直接模型可表示为  $f_{DT}^{\text{nobalance}}(\mathbf{X}_{\text{training}})$ , 此时  $f_{G1}(\cdot)$  的目的是使  $f_{DT1}(\cdot)$  模型不会因为样本的分布不均而产生过拟合现象. 假设  $f_{DT}^{\text{nobalance}}(\cdot)$  能够拟合任意函数, 那么对于任意的  $f_{G1}(\cdot)$  和  $f_{DT1}(\cdot)$ , 则一定存在  $f_{DT}^{\text{nobalance}}(\cdot)$ ,

使得下式成立

$$f_{DT1}(f_{G1}(\mathbf{X}_{\text{training}}), \mathbf{X}_{\text{training}}) = f_{DT}^{\text{nobalance}}(\mathbf{X}_{\text{training}}) \quad (13)$$

由式 (13) 可得, 受限于样本分布不均, 下游任务模型性能的提升必然是有限的. 当  $\mathbf{X}_{\text{training}}$  不服从  $\mathbf{X}_{\text{Real}}$  分布时, 在借助领域知识的情况下, 必然存在  $f_{G1}(\mathbf{X}_{\text{training}})$  也无法服从  $\mathbf{X}_{\text{Real}}$  分布的现象. 此时, 虽然  $f_{DT1}(f_{G1}(\mathbf{X}_{\text{training}}), \mathbf{X}_{\text{training}})$  在  $\mathbf{X}_{\text{training}}$  上的测试准确率可能会得到提升, 但在面对  $\mathbf{X}_{\text{Real}}$  时, 其准确率会较低.

面向样本多样性不足的问题, 其生成模型和下游任务模型可表示为  $f_{G2}(\mathbf{X}_{\text{training}}, \mathbf{X}_{\text{no\_label}})$  和  $f_{DT2}(f_{G2}(\mathbf{X}_{\text{training}}, \mathbf{X}_{\text{no\_label}}), \mathbf{X}_{\text{training}})$ , 对应的不采用  $f_{G2}(\cdot)$  下游任务直接模型可表示为  $f_{DT}^{\text{shortage}}(\mathbf{X}_{\text{training}})$ , 此时  $f_{G2}(\cdot)$  的目的是要获取符合全局分布的数据集以使得下游模型的鲁棒性更强. 当  $\mathbf{X}_{\text{training}}$  的分布不服从  $\mathbf{X}_{\text{Real}}$  分布、而  $\mathbf{X}_{\text{no\_label}}$  的分布却能够很接近  $\mathbf{X}_{\text{Real}}$  分布时, 下游模型  $f_{DT2}(\cdot)$  的性能相对于  $f_{DT}^{\text{shortage}}(\cdot)$  是能够得到显著提升的.

面向样本噪声干扰大问题, 其生成模型和下游任务模型可表示为  $f_{G3}(\mathbf{X}_{\text{training}})$  和  $f_{DT3}(f_{G3}(\mathbf{X}_{\text{training}}))$ , 对应的不采用  $f_{G3}(\cdot)$  的下游任务直接模型可表示为  $f_{DT}^{\text{noise}}(\mathbf{X}_{\text{training}})$ , 此时  $f_{G3}(\cdot)$  的目的是使下游模型能够更好“理解”工业图像. 当  $f_{DT}^{\text{noise}}(\cdot)$  是具有自适应去噪能力的端到端模型时, 则对于任意的  $f_{G3}(\cdot)$  和  $f_{DT3}(\cdot)$ , 一定会存在  $f_{DT}^{\text{noise}}(\cdot)$  使得下式成立

$$f_{DT3}(f_{G3}(\mathbf{X}_{\text{training}})) = f_{DT}^{\text{noise}}(\mathbf{X}_{\text{training}}) \quad (14)$$

由式 (14) 可得, 受限于样本中的噪声干扰, 下游任务模型性能的提升有限.

综上所述, 针对样本分布不均和噪声干扰大问题, 已有生成模型难以有效表征  $p_{\text{real}}$ ; 针对多样性不足问题而言, 已有生成模型通过对  $\mathbf{X}_{\text{no\_label}}$  数据的学习能够尝试获得足以表征  $p_{\text{real}}$  的图像集. 由于涉及数据采集与标定时场景缺失和知识缺乏等因素, 此类研究在工业领域现场中的应用较少. 因此, 如何更有效地生成具有较强表征性的工业图像集, 仍需结合工业机理和领域知识进行深入研究.

表 6 工业过程图像生成问题的本质

Table 6 The essence of image generation problems in industrial process

问题描述	生成模型的表达式	下游任务采用模型的表达式
样本分布不均	$f_{G1}(\mathbf{X}_{\text{training}})$	$f_{DT1}(f_{G1}(\mathbf{X}_{\text{training}}), \mathbf{X}_{\text{training}})$
样本多样性不足	$f_{G2}(\mathbf{X}_{\text{training}}, \mathbf{X}_{\text{no\_label}})$	$f_{DT2}(f_{G2}(\mathbf{X}_{\text{training}}, \mathbf{X}_{\text{no\_label}}), \mathbf{X}_{\text{training}})$
样本噪声干扰大	$f_{G3}(\mathbf{X}_{\text{training}})$	$f_{DT3}(f_{G3}(\mathbf{X}_{\text{training}}))$

### 3.2 工业过程生成图像评估研究现状

通常, 生成图像的评估指标应该满足以下标准: 1) 能够检测生成图像的生成质量; 2) 能够检测生成图像集的多样性; 3) 能够检测生成图像的可理解性或可控性; 4) 评价指标具备有界性; 5) 不可承受过高计算复杂度; 6) 能够检测生成图像的语义不变性; 7) 能够检测图像细微的形变和瑕疵.

因缺乏明确的似然概率度量<sup>[119]</sup>, 早期研究采用具有主观性的视觉方式对 GANs 所生成的图像进行评价. 之后, 文献 [68] 提出 GAN-train 策略, 通过对比基于 cifar100 数据集和基于合成图像训练的 CNN 判别器性能进行生成图像评估. 进一步, 文献 [120] 提出了同时进行 GAN-train 和 GAN-test 的评估策略, 具体为: GAN-train 采用 GAN 生成的图像训练判别器并测量其在真实测试图像上的性能, 进而评价 GAN 图像的多样性和真实感; GAN-test 采用真实图像训练判别器并基于 GAN 生成的图像进行评估, 进而衡量 GAN 图像的真实感. 此外, 初始得分 (Inception score, IS)<sup>[121]</sup> 和 FID<sup>[122-123]</sup> 也常被用作评估生成图像质量的指标.

本文将文献 [124-125] 所综述的定量和定性的生成图像评价指标进行整理, 如表 7 所示.

在面向样本多样性不足的研究中, 期望的生成样本需要服从全局分布, 而不是仅拟合训练样本的概率分布. 以 MSWI 过程为例, 燃烧线极端异常的火焰图像不存在于通常所获取的训练样本集中, 因此, 通常的评估方法不适用于这种情况. 如何结合特定行业图像固有关键特征对应的机理知识对其进行综合评估和筛选, 是工业生成图像评估算法研究的难点之一.

### 3.3 工业过程生成图像应用研究现状

当前针对工业过程图像生成的研究主要聚焦于故障识别和工况监测等任务, 基于工业图像进行反馈控制和关键参数量化等方面的研究较为缺乏. 以 MSWI 过程为例, 构建完备的火焰图像模板库可用于量化燃烧线, 从而支撑基于图像的实时燃烧控制. 但是在数据分析中发现, 燃烧线异常的火焰图像稀缺<sup>[14]</sup>, 极端异常的火焰图像缺失. 因此, 如何基于真实图像和生成图像构建完备模板库是实现上述目标的关键<sup>[13]</sup>. 图像生成技术在基于模板匹配的目标跟踪领域具有广泛的应用潜力. 具体来说, 模板匹配是指采用预定义的图像模板与实际场景中的图像进行匹配, 以达到目标检测和识别的目的. 在进行燃烧线量化时, 首先采用图像生成技术获取系列的火焰图像, 然后构建完备模板库, 最后采用模板匹配

技术实现图像量化. 这种方法可代替传统的基于经验的燃烧线控制方法, 进而提升智能化水平.

综上, 本节从面向识别任务和面向以目标跟踪任务为代表的参数量化两个方面进行研究现状的分析.

#### 3.3.1 面向识别/检测任务的生成图像应用现状

在实际生产线上, 因缺陷图像不足和标注成本过高等原因, 很难获得具有足够多样性的缺陷样本数据. 文献 [98, 102, 105] 均基于 GAN 提出面向特定工业行业的图像生成策略, 构建缺陷识别模型. 文献 [103] 提出基于 SDGAN 提升缺陷识别准确率, 其首先采用循环损失和 D2 对抗损失训练 SDGAN, 然后采用 FID 评估缺陷图片质量, 最后构建缺陷识别 CNN 模型, 结构如图 16 所示.

文献 [96] 提出基于 CE 增强和深度可分卷积构建缺陷分类模型, 采用 mobilenet V1 和 V2 训练两个模型, 表明轻量化的深度卷积可减少模型参数和计算量. 文献 [97] 提出基于 CAE 数据增强的缺陷检测模型, 采用 Xception 进行晶圆缺陷检测和分类.

面向自动缺陷检测系统中的缺陷图像分割任务, 文献 [106] 提出采用 DCGAN 和 CycleGAN 生成缺陷图像、采用 PatchMatch 和 PSGAN 生成无缺陷合成图像的策略, 其在数据增强基础上提出基于 CycleGAN 的周期性纹理缺陷分割框架, 表明该框架优于现有的弱监督分割方法, 结构如图 17 所示.

以 GAN 为代表的图像生成算法已应用于工业缺陷检测领域, 因其具有更佳的数据增强、特征提取、样本平衡能力而备受关注. 然而, 由于训练过程不稳定、生成数据存在噪声等问题, 其性能和稳定性仍需进一步研究. 因此, 在采用 GAN 进行工业缺陷检测时, 需要进行充分的实验和评估, 以验证其应用效果并对其优化方法进行探讨, 从而提高工业应用中的实用性.

下面综述在电气设备识别、管道漏磁检测和焊接接头检测等过程中的应用现状.

##### 3.3.1.1 电气设备识别过程

该过程存在的问题包括工业过程温度分布过于集中、设备形状尺寸不确定以及识别背景复杂等. 文献 [99] 提出基于 EOGAN 生成数据的电气设备识别框架, 首先采用预处理和分割技术提取边缘信息, 然后在训练阶段将图像数据与边缘特征进行匹配以生成红外图像, 最后在基于边缘特征的图像生成阶段构建识别模型, 其策略如图 18 所示.

上述研究存在的局限性在于: 1) 采用的数据集规模相对较小, 所提方法的适用性和鲁棒性有待采用更为广泛和多样化的数据集进行验证; 2) 未考虑

表 7 生成图像评估指标  
Table 7 Evaluation index for generated image

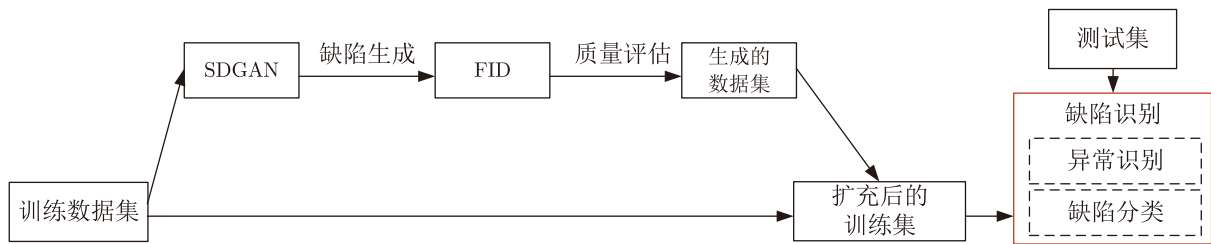
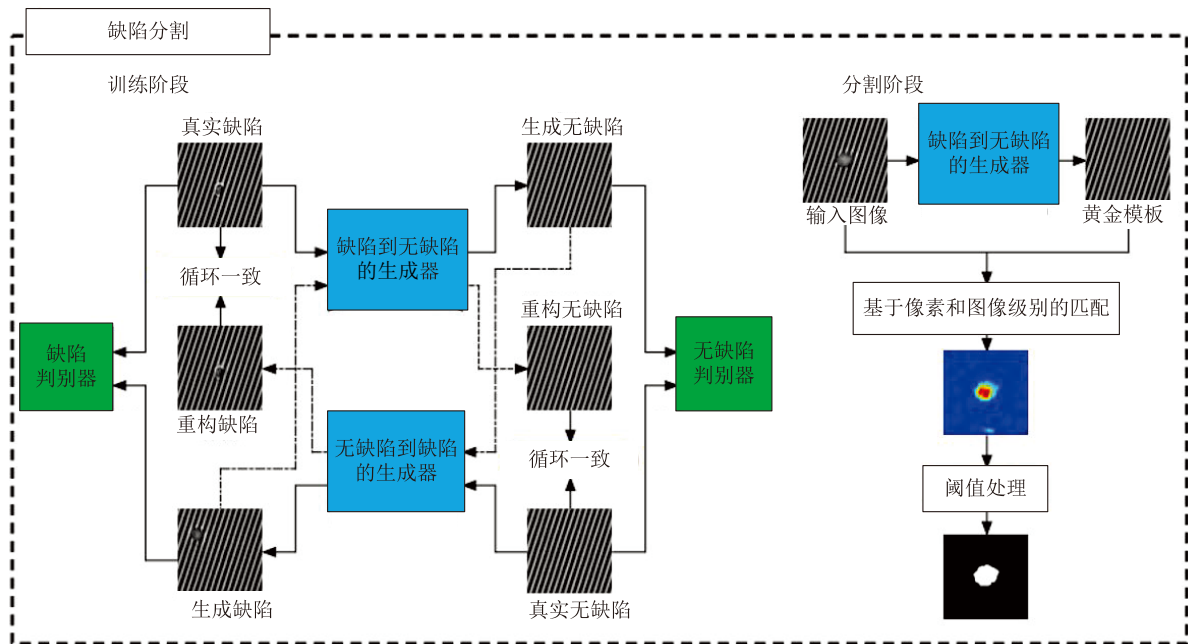
方法	描述	文献与年份	
1) 平均对数似然	采用从生成数据中估计的密度 (例如, 采用 KDE 或 Parzen 窗口估计) 解释真实/测试数据的对数可能性, 即	[32], 2014	
	$L = \frac{1}{N} \sum_{i=1}^N \log_2 p_{\text{model}}(\mathbf{x}_i)$	[126], 2016	
2) 分类性能	评估无监督表示质量的间接技术 (例如, 特征提取、FCN 得分), 可参见 GAN 质量指数 (GQI) <sup>[127]</sup> .	[50], 2016 [110], 2017	
3) 初试分数 (Inception score, IS)	生成数据的条件与边缘标签分布之间的 KLD, 即 $\exp(\mathbb{E}_{\mathbf{X}}(\text{KL}(p(\mathbf{y} \mathbf{x})  p(\mathbf{y}))))$ .	[121], 2016	
4) 判别器二样本检测	采用二分类判别器识别两个样本是否源自同一分布.	[128], 2005 [129], 2004	
5) 图像质量测度	采用 SSIM、PSNR 和清晰度差异等指标.	[130], 2015 [131], 2017	
6) 最大均值差异	从每个分布独立采样, 测量概率分布之间的差异, 如下 $M_k(p_r, p_g) = \mathbb{E}_{\mathbf{X}, \mathbf{X}' \sim p_r}(k(\mathbf{x}, \mathbf{x}')) - 2\mathbb{E}_{\mathbf{X} \sim p_r, \mathbf{Y} \sim p_g}(k(\mathbf{x}, \mathbf{y})) + \mathbb{E}_{\mathbf{Y}, \mathbf{Y}' \sim p_g}(k(\mathbf{y}, \mathbf{y}'))$	[132], 2012	
7) 模式分数 (Mode score, MS)	与 IS 类似, 同时考虑标签在真实数据上的先验分布, 即 $\exp(\mathbb{E}_{\mathbf{X}}(\text{KL}(p(\mathbf{y} \mathbf{x})  p(\mathbf{y}^{\text{train}})))) - \text{KL}(p(\mathbf{y})  p(\mathbf{y}^{\text{train}}))$	[133], 2016	
定量 评估	8) 图像检索性能	测量图像间的最近邻距离分布: 方式 1. 设 $\mathbf{d}_{i,j}^k$ 是由方法 $k$ 生成的第 $j$ 个图像与测试图像 $i$ 的最近邻距离, $\mathbf{d}_{i,j}^k = \{d_{1,j}^k, \dots, d_{n,j}^k\}$ 是单张图像到所有测试图像的最近邻距离的集合, 采用 Wilcoxon 符号秩检验假设: 两个生成器之间最接近的两个距离分布之间的差值的中值为零. 若该假设成立, 则两个生成器一样好, 否则, 其结果可用于评估哪种方法在统计上更好. 方式 2. 设 $\mathbf{d}_j^t$ 是第 $j$ 个训练图像到数据集的距离, 考虑到训练集和测试集源自同一个数据集, $\mathbf{d}_j^t$ 可被认为是生成器达到的最优分布; 计算平均最近邻距离的相对增量以度量生成样本与理想样本间的差异, 如下式所示 $\hat{d}_j^k = \frac{\bar{d}_j^k - \bar{d}_j^t}{\bar{d}_j^t}, \bar{d}_j^k = \frac{1}{N} \sum_{i=1}^N d_{i,j}^k, \bar{d}_j^t = \frac{1}{N} \sum_{i=1}^N d_{i,j}^t$	[134], 2016
	9) 生成对抗度量 (Generative adversarial metric, GAM)	通过交换判别器和生成器比较两个 GAN.	[135], 2016
	10) 覆盖率度量	生成数据覆盖真实数据的概率质量 $C := p_{\text{data}}(d_{p_{\text{model}}} > t)$ , 其中 $t$ 使 $p_{\text{model}}(d_{p_{\text{model}}} > t) = 0.95$ .	[136], 2017
	11) 改进的初始分数 (Modified inception score, m-IS)	侧重于从特定类别中采样图像的多样性 $\exp(\mathbb{E}_{\mathbf{X}_i}(\mathbb{E}_{\mathbf{X}_j}(\text{KL}(p(\mathbf{y} \mathbf{x}_i)  p(\mathbf{y} \mathbf{x}_j))))))$	[137], 2017
	12) 激活最大化分数 (Activation maximization score, AM Score)	考虑训练标签与预测标签之间的 KLD 分布以及预测值的熵 $\text{KL}(p(\mathbf{y}^{\text{train}})  p(\mathbf{y})) + \mathbb{E}_{\mathbf{X}}(H(\mathbf{y} \mathbf{x}))$	[138], 2017
	13) 弗雷歇距离 (Fréchet inception distance, FID)	多元高斯数据的特征空间数据的 Wasserstein-2 距离 $\text{FID} = \ \mu_r - \mu_g\ ^2 + \text{Tr}(\text{Cov}_r + \text{Cov}_g - 2(\text{Cov}_r \times \text{Cov}_g)^{\frac{1}{2}})$	[122], 2017 [123], 2017
	14) 沃特斯汀评判 (The Wasserstein critic)	训练神经网络, 评判真实样本为高值和低值 $\hat{W}(\mathbf{x}_{\text{test}}, \mathbf{x}_g) = \frac{1}{N} \sum_{i=1}^N \hat{f}(\mathbf{x}_{\text{test}}[i]) - \frac{1}{N} \sum_{i=1}^N \hat{f}(\mathbf{x}_g[i])$	[67], 2017
	15) 生日悖论测试 (Birthday paradox test)	通过计算重复数 (接近重复数) 度量离散 (连续) 分布的大小.	[139], 2017
	16) 对抗准确率和散度	计算两个判别器的分类精度, 一个训练于真实数据, 一个训练于生成数据, 在验证集上评估 $p_g(\mathbf{y} \mathbf{x})$ 和 $p_r(\mathbf{y} \mathbf{x})$ .	[140], 2017
	17) 重构误差 (Reconstruction error)	通过优化 $\min \ G(\mathbf{z}) - \mathbf{x}^{\text{test}}\ ^2$ 测量测试图像与最近生成图像间的重构误差.	[141], 2017
	18) 低层次的图像统计 (Low-level image statistics)	从平均功率谱、随机滤波器响应分布、对比度分布等方面评估生成图像的低层统计特征与自然场景的相似程度.	[142], 2017 [143], 2017
	19) 精确度、召回率和 F1 分数	用于量化 GANs 中的过拟合程度.	[123], 2017
	20) 边界失真	采用分类方法测量生成样本的多样性和协变量漂移.	[144], 2018
	21) 统计显著性差异 (Number of statistically-different bins, NDB)	假设存在两组源自同一分布的样本集, 那么落入给定区间的样本数量在采样误差的范围内应该相等.	[145], 2018



表 7 生成图像评估指标 (续表)

Table 7 Evaluation index for generated image (continued table)

方法	描述	文献与年份
22) 比赛胜率和技能等级	设计比赛: 玩家要么是一个试图区分真实数据和虚假数据的判别器, 要么是一个试图欺骗判别器将虚假数据当作真实数据接受的生成器.	[146], 2018
23) 归一化相对判别分数 (Normalized relative discriminative score, NRDS)	若生成样本与真实样本更接近, 则需要更多的训练批次才能将它们与真实样本进行区分, 从而对 $n$ 个 GAN 进行比较.	[147], 2018
24) 几何分数	在真实数据与生成数据之间比较底层数据流形的几何属性.	[148], 2018
25) 切片 Wasserstein 距离 (Sliced Wasserstein distance, SWD)	基于切片 Wasserstein 距离的生成图像评估指标. 可测量生成图像与真实图像之间的分布差异, 同时避免了 FID 的缺点.	[69], 2018
26) 类别感知 FID (Class-aware FID, CAFD)	采用高斯混合模型更好地拟合特征分布; 此外包含类别信息, 计算每个 $K$ 类的 FID 并对结果求平均, 得到 CAFD $\text{CAFD}(p_r, p_g) = \frac{1}{K} \sum_{i=1}^K \ \mu_i^r - \mu_i^g\  + \text{Tr}(Cov_i^r + Cov_i^g - 2(Cov_i^r \times Cov_i^g)^{\frac{1}{2}})$	[149], 2018 [150], 2021
定量评估		
27) 无偏的 FID 和 IS	采用外推方法来获得分数的无偏估计, 称为 $\overline{\text{FID}}_\infty$ 和 $\overline{\text{IS}}_\infty$ , 用无限个样本计算代替有限样本分数.	[151], 2020
28) 快速 FID (Fast FID)	加快 FID 计算的方法: 真实的样本在训练期间不会改变, 其 Inception 编码仅计算一次.	[152], 2020
29) 空间 FID (Spatial FID, SFID)	FID 变体, 采用空间特征取代标准的池化特征. 采用标准 pool3 初始特征和中间混合 6/conv 特征映射的前 7 个通道来计算 FID.  FID 的扩展, 除度量分布相似度, 还考虑模型对生成图像的记忆程度 $\text{MiFID}(S_g, S_t) = m_\tau(S_g, S_t) \times s(S_g, S_t)$ 其中, $S_g$ 是生成数据集, $S_t$ 是真实数据集, $m_\tau$ 是基于阈值的记忆惩罚, $s(S_g, S_t)$ 是距离度量 $s(S_g, S_t) = \frac{1}{ S_g } \sum_{x_g \in S_g} \min_{x_t \in S_t} (1 - \frac{ \langle x_g, x_t \rangle }{\ x_g\  \ x_t\ })$ $m_\tau(S_g, S_t) = \begin{cases} \frac{1}{s(S_g, S_t)}, & s(S_g, S_t) < \tau \\ 1, & \text{otherwise} \end{cases}$	[153], 2021 [154], 2021
30) 记忆感知 FID (Memorization-informed FID, MiFID)		
31) 核密度估计 (Kernel density estimation, KDE)	用于估计概率密度函数形状的非参数估计方法, 通过采用核函数在数据点周围形成局部的概率密度贡献相加得到全局的概率密度函数.	[155], 2022
1) 快速的场景分类	参与者被要求在很短的展示时间内 (例如 100 ms) 区分生成图像和真实图像.	[32], 2014
2) 近邻取样	为了检测过拟合, 生成的样本将与训练集中最近的“邻居”同时展示.	[156], 2001 [157], 2014
3) 网络内部结构	用于探索和说明模型的内部表征和动力学 (如空间连续性) 以及可视化学习的特征.	[5], 2016 [158], 2016 [159], 2016 [160], 2017 [161], 2017 [162], 2017
4) 偏好判断	参与者被要求根据生成图像的保真度对模型进行排名.	[163], 2017 [164], 2018
定性评估		
5) 模式掉落和崩溃	对于已知模式的数据集 (例如高斯混合模型或带标签的数据集), 通过测量生成数据与模式中心的距离进行度量.	[165], 2017 [166], 2017
6) 人眼感知评估 (Human eye perceptual evaluation, HYPE)	基于人眼视觉感知特性的图像质量评估方法. 在 HYPE 中, 评价者根据主观感受评估图像的质量, 评价指标通常包括多个评价者的评分, 得出图像的颜色、对比度、清晰度等, 统计质量评价结果.	[167], 2019 [168], 2019
7) 神经得分 (Neuroscore)	采用语义分割网络比较生成图像和真实图像中分割对象的分布情况, 揭示 GAN 忽略某些对象类别的统计差异, 比较真实图像与 GAN 生成图像间的差异.	[169], 2019
8) 通用的假与真检测器 (A universal fake versus real detector)	创建一个“通用”的检测器, 用于区分真实图像和生成图像, 如 CNN 或 GAN. 在训练阶段, 该模型采用真实图像和生成图像作为输入, 以学习如何区分两类图像. 在测试阶段, 该模型对新的图像进行真假检测.	[170], 2019 [171], 2020 [172], 2020 [173], 2021
9) 细节对比	通过细节对比检验不同 GAN 模型的生成效果.	[174], 2022 [175], 2023

图 16 基于 SDGAN 数据增强的检测过程<sup>[103]</sup>Fig.16 Detection process based on data enhancement of SDGAN<sup>[103]</sup>图 17 周期性纹理图像的缺陷分割过程<sup>[106]</sup>Fig.17 Defect segmentation process of periodic texture image<sup>[106]</sup>

在实际应用中可能存在的不确定性因素,如设备老化、环境变化等因素可能会导致生成的合成数据与真实数据存在差异;3)数据需要进行后续处理才能用于设备识别,在实际应用中需要考虑更快速和实时的识别方法。

### 3.3.1.2 管道漏磁检测过程

存在的问题是检测环境不稳定和设备异常会导致信息不完整.文献[100]提出基于多传感器融合的方法,用于处理信息不完整情况下的缺陷表征问题.基于3种不同类型的传感器信号(磁场强度、相位和高斯噪声)并采用改进的cGAN生成缺陷信号样本,以弥补传感器数据缺失的情况.研究表明,该方法可显著提高缺陷表征的准确性和鲁棒性,降低传感器数据缺失造成的影响,其策略如图19所示。

值得注意的是,该研究仅基于模拟的信号数据集进行,并未在实际漏磁检测场景中进行验证.此

外,该方法可能会受到缺失数据位置和数量等因素的影响,因此其应用范围具有局限性。

### 3.3.1.3 焊接头检测过程

该过程存在问题是:1)不合格焊接点的样本稀缺;2)焊点具有纹理弱、对比度弱和存在腐蚀等复杂特性;3)噪声干扰大。

文献[101]通过集成GAN和AE构建了具有图像数据生成、特征提取和模式识别功能的网络结构,首先采用改进GAN扩展不合格点焊接头的图像数据集,然后结合专家经验通过AE选择图像的特征向量,最后利用隐马尔科夫模型判断点焊质量.该方法能够解决标准GAN中生成的图像样本分布不均和多样性不足的问题。

文献[176]采用CycleGAN实现工件与焊缝之间的样式转换,进而实现数据集的扩展,其应用过程如图20所示。

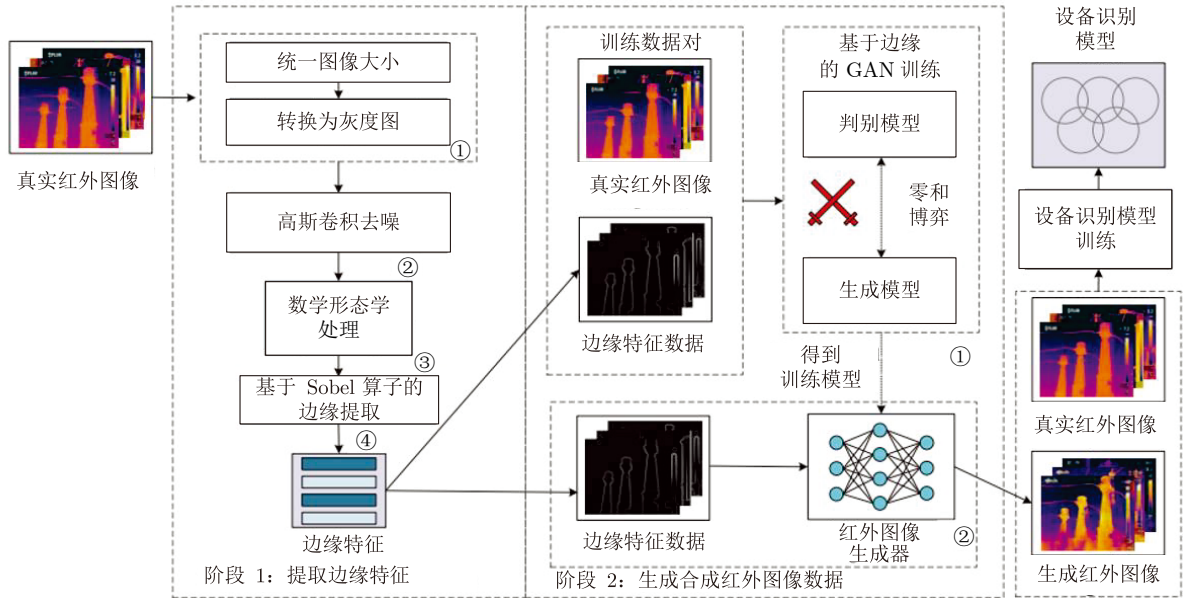


图 18 电气设备识别方法的结构<sup>[99]</sup>

Fig. 18 Architecture of identification method for electrical equipment<sup>[99]</sup>

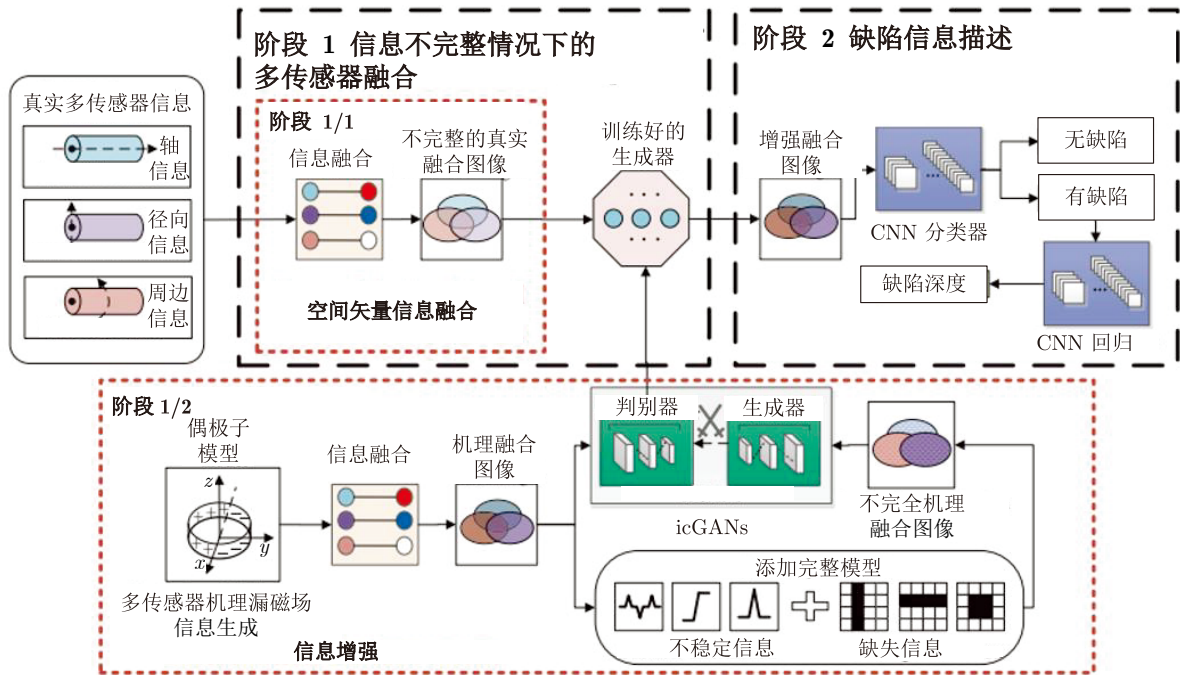


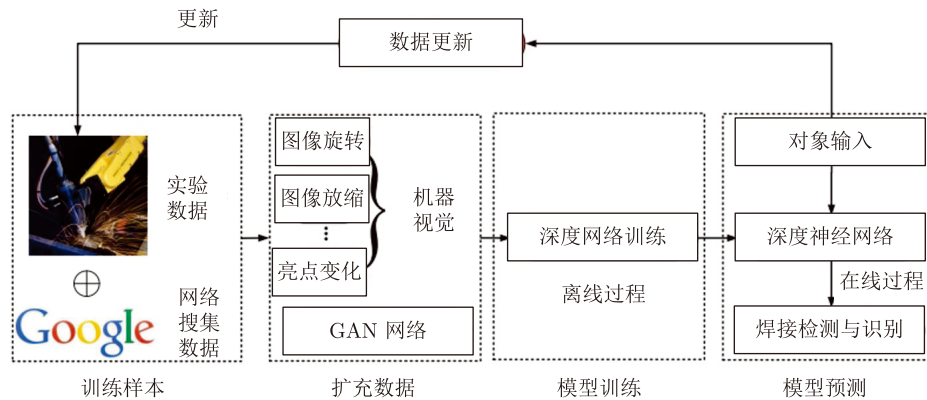
图 19 信息不完全下漏磁缺陷表征的多传感器融合框架<sup>[100]</sup>

Fig. 19 Multi-sensor fusion framework for magnetic flux leakage defect characterization under incomplete information<sup>[100]</sup>

首先, 针对训练样本不足的问题, 结合图像处理和 GAN 生成高质量的训练样本; 然后, 建立训练样本的更新机制以保证深度神经网络模型能够覆盖所有样本; 最后, 利用深度神经网络实现焊接头的检测与识别. 实验表明, 该方法能够快速高效地完

成焊接头的检测与识别.

文献 [109] 应用 pix2pix<sup>[110]</sup> 消除焊接图像中的噪声以实现图像增强, 提出基于多源传感图像进行熔透检测的新框架, 其能够同步分析多种类型的光学传感图像, 包括图像预处理、图像选择和焊缝熔

图 20 焊接头检测与识别框架<sup>[176]</sup>Fig. 20 Welded joint inspection and identification framework<sup>[176]</sup>

透分类 3 个采用深度学习进行增强的阶段。

综上, 基于 GAN 的图像增强在焊接头检测中应用广泛. 该技术通过学习原始图像中的信息生成高质量图像, 能够提高焊接头缺陷、几何形状、尺寸和质量检测的准确性与可靠性. 通过增强图像的细节、清晰度和纹理等特征, 可显著提高焊接头检测的效果.

### 3.3.2 面向目标跟踪任务的生成图像应用现状

#### 3.3.2.1 工业领域

近年来, 孪生网络 (Siamese network) 的突出精度和速度使其在目标跟踪研究中广受关注, 尤其是在视频监控和自动驾驶等领域<sup>[177-178]</sup>, 但在复杂工业过程中的应用仍需要进一步深入.

以对 MSWI 过程的燃烧火焰进行跟踪为例, 目前还存在如下开放性问题: 1) 生成虚假火焰图像以增强数据集的数量和质量, 提高火焰目标跟踪算法的鲁棒性和泛化能力; 2) 生成虚拟火焰图像用于训练火焰边界检测模型, 以便能够更准确地检测火焰的位置和形态; 3) 火焰的形态和颜色等信息会不断发生变化, 而成像设备故障、物质遮挡等因素可能会导致火焰图像中出现缺失区域, 这需要采用图像生成技术进行填充, 从而更好地进行燃烧火焰的跟踪; 4) 火焰的形态和颜色等信息能够反映炉内的燃烧状态, 生成虚拟火焰图像可用于训练模型预测火焰状态的变化趋势, 从而更好地控制和优化炉内燃烧过程.

因此, 面向目标跟踪的图像生成技术在工业领域具有广泛的应用前景, 如在 MSWI 过程中, 其可用于增强数据集的数量和质量、检测火焰边界、填充缺失数据和预测火焰状态的变化趋势等.

#### 3.3.2.2 其他领域

SINT (Siamese instance search tracker) 算法<sup>[179]</sup>

是基于孪生网络的跟踪算法的代表之一, 其采用孪生网络训练识别与初始目标外观相匹配的候选图像位置, 进而将目标跟踪问题转化为匹配问题. 相比之下, SiamFC 算法<sup>[180]</sup>在初始阶段训练全卷积网络用于解决相似性学习问题, 在推理期间对学习到的匹配函数进行在线评估, 能够以超实时速度运行. SiamRPN 算法<sup>[181]</sup>将 RPN (Region proposal network) 网络应用于目标跟踪任务, 用锚框替代多尺度卷积过程, 进一步提升了跟踪速度和精度.

尽管上述跟踪算法表现出色, 但均采用固定模板进行跟踪, 在目标出现旋转、形变和运动模糊等外观变化时容易出现模板匹配错误, 进而导致目标跟踪失败. 为此, CFNet 算法<sup>[182]</sup>在 SiamFC 算法框架中引入可微相关滤波器, 采用岭回归微调初始模板, 从而可在跟踪过程中更新模板. 另外, 文献 [183] 提出了动态 Siamese 网络, 通过快速转换学习模型, 能够有效地从历史帧中在线学习目标外观变化. 此外, 文献 [184] 在目标跟踪过程中采用 GAN 生成所需要的模板.

以上算法在一定程度上提高了跟踪性能, 但在目标外观发生变化时, 会存在跟踪失败的情况. 由于图像生成算法能够在目标跟踪中生成跟踪目标的模板、学习目标的潜在表示、进行数据的增强更新以适应变化<sup>[184]</sup>, 因此应用前景广阔.

## 4 讨论与分析

### 4.1 方法比较

面向工业过程的图像生成、生成图像评估和生成图像应用共 3 个视角的相关研究统计结果详见表 8.

由表 8 可知:

1) 工业过程图像生成. 从问题角度分析, 其主

要能够解决样本分布不均、多样性不足和噪声干扰大 3 类问题. 具体而言, 从解决问题的视角出发, 针对不同的类别应采用不同的策略: 面向样本分布不均, 主要采用通过生成样本拟合小样本分布的方式, 主要度量指标为样本稀疏度和样本相似度, 该策略能够有效提高下游识别模型的泛化性能; 面向样本多样性不足, 主要采用图像到图像的转换方式, 借助源域特征扩充目标域特征, 相较于原始目标域图

像集, 目标域特征更加丰富, 能够在一定程度上解决目标域图像缺失问题, 该策略能够有效提高下游识别模型的鲁棒性; 面向样本噪声干扰大, 主要采用图像到图像的转换方式, 借助清晰图像集的概率分布修正噪声干扰大图像集的概率分布, 进而去除噪声, 该策略能够有效提高图像的可解释性.

2) 工业过程生成图像评估. 生成图像评估模型的架构涵盖了输入图像与生成图像之间的比较和分

表 8 面向工业过程的图像生成及其应用与评估统计表  
Table 8 Statistical table of image generation, application and evaluation for industrial process

子类别	方法	年份	优劣	文献	
样本分布不均	生成器为编码器-解码器的 GAN, 引入小波对生成图像细化	2019	无缺陷区域的基本不变, 同时避免图像模糊.	[102]	
	AdaBalGAN	2019	根据分类准确性平衡每种缺陷类型的样本数量.	[98]	
	图像处理和 GAN	2019	焊接头的检测与识别.	[176]	
	卷积编码器 (CE)	2020	基于编码的图像生成, 具有可解释性, 但多样性不足.	[96]	
	卷积自编码器 (CAE)	2020	基于编码的图像生成, 具有可解释性, 但多样性不足.	[97]	
	EOGAN	2020	边缘特征作为先验知识指导红外图像生成.	[99]	
	生成器是 U-Net 的改进 GAN	2021	细化生成器的损失函数构建多传感器信息, 从而更好地增强漏磁信息.	[100]	
工业过程图像生成研究	混合生成样本与实际样本训练 GAN	2022	一定程度上避免模态崩溃.	[101]	
	引入工业过程数据的 GAN 的图像生成	2023	创新的应用, 采用 GAN 来生成高炉料槽的热图; 需克服数据依赖性、模型复杂性和实际应用验证等方面的挑战.	[185]	
	SDGAN	2020	引入 D2 对抗损失和循环一致损失, 学习更全面的特征.	[103]	
	采用 DCGAN 和 CycleGAN 生成缺陷图像, 采用 PatchMatch 和 PSGAN 生成无缺陷合成图像	2020	同时生成缺陷和无缺陷图像.	[106]	
	DuCaGAN	2020	在 CycleGAN 基础上加入胶囊网络的视觉不变性和旋转, 能够学习细节特征.	[105]	
	缺陷区域和强度可控 GAN	2021	生成缺陷的区域和强度具有可控性.	[186]	
	具有注意力机制的循环一致性 GAN (AttenCGAN)	2022	循环一致对抗网络和注意力机制解决小样本表面缺陷分类问题; 缺乏详细的实验结果和定量指标, 存在适用性的局限性以及可解释性的问题.	[187]	
样本噪声干扰大	pix2pix	2019	消除焊接图像中的噪声.	[109]	
	cGAN	2019		[111]	
	PLSGAN	2021	能够增强和降噪红外热成像图像中的缺陷.	[188]	
工业过程生成图像应用研究	面向识别任务	基于 GAN 增强的缺陷识别	2019	有效提高识别模型的性能.	[102]
			2019		[98]
			2020		[105]
		基于 CAE 数据增强的缺陷检测模型	2020	轻量化的深度卷积可减少模型参数和计算量, 提高缺陷检测性能.	[96]
			2020		[97]
		基于 CycleGAN 的周期性纹理缺陷分割框架	2017	优于现有的弱监督分割框架.	[104]
		基于红外图像的电气设备识别	2020	能够根据边缘特征生成图像; 采用的数据集规模相对较小, 适用性和鲁棒性有待采用更为广泛和多样化的数据集验证.	[99]
	管道漏磁检测	2021	集成 GAN 和 AE, 解决标准 GAN 中生成的图像样本分布不均和多样性不足的问题.	[100]	

表 8 面向工业过程的图像生成及其应用与评估统计表 (续表)

Table 8 Statistical table of image generation, application and evaluation for industrial process (continued table)

子类别	方法	年份	优劣	文献	
工业过程生成 图像应用研究	面向目标 跟踪任务	焊接检测			
		具有多维信息感知和时域运动补偿的孪生网络, 进行卫星视频中的目标跟踪	2019	结合 GAN 和图像处理, 建立训练样本的更新机制, 以保证模型能够覆盖所有样本.	[176]
		基于孪生多尺度网络的露天矿变化检测	2019	采用 pix2pix 技术去噪, 包括图像预处理、图像选择和焊缝熔透三个阶段, 提高检测效果.	[109]
			2016	引入多维信息感知和时域运动补偿技术, 对于解决卫星视频中目标跟踪问题具有潜力; 缺乏实验结果的详细分析及对使用的数据集的说明.	[179]
		2018	Siamese 网络和多尺度信息的引入能够提高变化检测的准确性; 露天矿应用具有创新性.	[181]	
工业过程生成 图像评估现状	定量	2005 ~ 2023	能够给出客观定量的评价指标, 具有通用性; 在面向样本多样性不足的研究中, 期望的生成样本需要服从全局分布, 而不是仅拟合训练样本的概率分布, 此类样本无法被有效评估.	[32] [110] [122] [123] [126-155]	
	定性	2001 ~ 2023	近邻取样网络内部结构、快速的场景分类、偏好判断、模式掉落和崩溃、HYPE、神经得分、通用的假与真检测器和细节对比 (详见表 7)	能够观测出是否生成期望特征; 无法定量描述或给出客观指标.	[32] [156-175]

析过程, 用于确定两者之间的相似性和差异, 其流程通常包括预处理、特征提取、相似性度量和评估指标计算等步骤. 定量指标虽然能够提供客观的数值结果, 但在极端异常工业过程图像的评估方面可能存在一定困难, 无法准确反映生成图像的真实质量和特殊特征. 定性指标主要是指基于人工方式的主观评估, 包括专家评审、用户调查和视觉感知实验等方法, 尽管该指标主要依赖于经验和主观判断, 但在评估特殊生成图像时却具有一定优势, 能够更全面、准确地评估图像的视觉质量、真实性和感知一致性. 在未来研究中, 需要改进和发展更为准确可靠的定量指标, 需要在评估中融入更多的工业特征和领域知识. 此外, 还可以探索基于深度学习的生成图像质量评估算法, 以提高评估的自动化程度和准确性.

3) 工业过程生成图像应用. 目前, 对于工业过程生成图像应用的研究主要侧重于故障识别和工况监测等任务, 面向工业生成图像进行反馈控制和关键参数量化等方面的相关研究相对较少. 本文将生成图像应用划分为面向识别/检测任务和面向目标跟踪任务两类. 针对识别/检测任务, 主要集中在电气设备识别、漏磁检测和焊接检测等方面, 研究者采用不同的图像生成方式获取具有特定目标的图像, 以便进行算法的训练和性能的评估. 针对目标跟踪任务, 目前的研究主要涉及其他非工业过程领域, 能够生成具有已知运动轨迹和特征的目标, 进而提供训练数据或进行算法测试. 未来研究中, 工

业生成图像还可以用于过程难测参数的检测和控制, 模拟在不同参数条件下的系统性能和响应, 进而拓宽其在工业领域中的潜在应用价值.

#### 4.2 讨论与分析

结合以上分析, 本文总结了面向工业过程的图像生成及其应用的未来研究方向, 如下所示.

1) 结合工业背景知识增加生成图像的多样性. 目前的生成研究主要采取拟合已有真实小样本数据和使用图像到图像转换的策略, 通过将大量无标签样本融入生成过程以增加样本多样性<sup>[103, 105-107]</sup>. 然而, 在工业环境中, 如何生成同时具有全局分布特征和强解释性的样本仍然是一个挑战. 笔者认为, 解决方案之一是引入工业机理知识和领域专家知识约束, 进而生成符合工业过程实际物理规律的工业图像, 以提高生成结果的可靠性和可控性. 此外, 也可以考虑将不同类型的工业过程数据 (如传感器数据、文本数据等) 与工业图像相结合, 实现基于多模态的工业过程图像生成.

2) 结合扩散模型、Prompt learning 和大型生成模型的图像生成. 笔者认为, 在面向工业过程图像生成及其应用的未来研究中, 结合扩散模型、Prompt learning 和大型生成模型将具有里程碑式的重要意义. 扩散模型能够通过控制噪声信号的逐步变化生成数据样本, 支持逆向过程和条件生成, 可产生与 GAN 方法相媲美的高质量样本. 通过改进和扩展扩散模型应对不同工业领域的特定挑战, 显然能够提高视觉感知模型的准确性和泛化能力.

基于 Prompt learning 技术, 理论上能够通过设计适当的提示信息引导生成模型输出符合工业过程特征的图像; 如何设计提示信息, 如何在生成过程中引入领域专家知识与物理约束信息, 是提高生成结果可控性和准确性的关键之一. 大型生成模型能够学习大规模数据集的知识, 通过文本描述能够生成高质量、可解释的图像, 但受限于缺乏可训练样本和领域知识等因素, 如何生成工业领域性强的图像还有待解决. 因此, 结合扩散模型、Prompt learning 和大型生成模型的图像生成技术将能够为工业过程图像生成的未来研究提供新的契机.

3) 对图像去噪与图像生成进行协同优化的图像生成. 受传感器噪声、光照变化或物体振动等多种因素的影响, 工业过程图像常常包含噪声和伪影<sup>[109, 111]</sup>, 过度拟合噪声数据会导致模型记忆随机噪声的模式, 进而影响模型的泛化能力<sup>[189]</sup>. 因此, 需要研究如何设计能够同时学习图像去噪和图像生成任务的一体化模型, 通过共享特征表示和参数, 使得去噪和生成过程能够相互促进, 进而提高生成结果的质量和准确性. 此外, 针对具体的工业过程, 其噪声结构往往具有独特性, 例如在 MSWI 过程中, 噪声主要源自飞灰、杂质以及高温对摄像头的影响. 笔者认为, 未来研究需要根据噪声结构中所存在的先验信息选择或改进更为适合的算法, 这将有助于提高工业图像去噪效果.

4) 基于生成图像的工业模板库构建与更新及在下游应用. 通常的工业图像模板库因存在异常样本分布不均和极端异常样本缺失等问题, 导致难测参数和控制模型的泛化性能和鲁棒性能受到限制<sup>[190]</sup>. 从模板库的构建和更新视角而言, 简单的线性更新策略显然难以满足工业现场的所有可能情况; 此外, 该更新策略在空间维度上是恒定的, 无法实现局部更新. 因此, 有必要采用基于生成图像的策略构建完备的工业图像模板库以解决上述问题. 从难测参数检测的角度分析, 基于图像生成的模板库能够提供完备的图像特征, 能够涵盖各种可能的异常情况, 从而提高难测参数的检测能力. 从基于图像进行控制<sup>[191]</sup>的角度分析, 基于图像生成的模板库的数据具有完备性, 能够与控制策略建立较为完整的映射, 进而支撑实现更为准确和精细的工业过程控制. 此外, 生成模型还可以根据不同的控制策略生成相应的图像样本, 用于评估控制系统的性能. 因此, 通过构建和更新完备的工业图像模板库并将其应用于难测参数检测和控制优化, 在未来有望提升工业过程控制系统的性能.

5) 基于生成图像与过程参数工业模板库的产品质量优化研究. 本文认为, 通过利用模板库中的图像样本和相关参数信息, 进行基于生成图像与过程参数工业模板库的产品质量优化将是未来的重要研究领域. 首先, 需要收集、整理和标注大量的样本数据, 即构建的模板库应包含不同工业产品的图像样本以及与其相关的质量参数, 以支持后续的产品质量优化研究. 然后, 利用生成模型技术扩充样本库中的样本数量, 即通过生成图像模拟不同参数和条件下的产品. 最后, 结合图像处理和机器学习方法, 通过分析生成的图像样本和与之相关的参数信息, 建立预测模型以优化产品质量. 显然, 上述研究将会推动工业领域质量管理和控制技术的发展, 实现对产品质量的优化和改进, 提高工业生产的效率和竞争力.

6) 基于生成图像跟踪特征目标的工业过程控制和优化. 本文认为, 基于生成图像跟踪特征目标在助力工业过程的控制和优化方面具有较大的应用潜力, 具体表现在: 可用于工业过程目标的感知和识别, 目标位置、状态和运动轨迹的实时监测; 基于跟踪信息实现自适应的控制和优化策略, 进行更高效、精确和可靠的工业生产过程跟踪控制; 通过比较生成图像与预期目标图像的差异进行异常检测和故障诊断, 及时检测和识别工业过程中的异常行为或故障情况, 支撑容错控制算法的设计与实现; 通过实时监测目标的位置和状态对资源进行优化分配和利用, 进而提高生产效率和资源利用率; 通过自动化的生成图像分析和目标跟踪, 促进工业过程的智能化和无人化, 减少对人力资源的依赖, 提高对生产过程的可控性和安全性. 未来的研究方向包括提高生成图像跟踪的精度和稳定性, 探索更为有效的目标识别和跟踪算法以及将其与工业过程的控制和优化相结合, 提高工业过程的智慧化水平, 实现更为高效、可持续和安全的工业生产.

## 5 总结与展望

本文首先概述了图像生成研究现状, 阐释了工业过程图像生成的定义、流程、评估和应用需求; 然后, 简要分析了在工业领域具有潜在应用价值的图像生成算法; 接着, 依据图像生成流程, 从图像生成、生成图像评估和应用 3 个视角进行详细综述; 最后, 讨论了这些算法的技术特点和研究难点. 笔者认为, 为获得具有全局分布特性的工业过程生成图像样本集, 未来的研究主要面临着以下挑战:

1) 大规模模型的融入: 随着深度学习的快速发

展,大规模模型在图像生成任务中已经展现出巨大潜力.考虑到将大规模模型应用于工业过程中需要解决计算资源消耗、模型复杂度和训练效率等问题,未来的挑战应致力于如何通过高效地融入大规模模型以提升特定行业的生成图像的质量和效率.

2) 多模态场景的生成:工业过程涉及多种场景和特征,单一的生成模型难以覆盖所有工况下的图像生成需求.研究人员可探索设计具有多模态特性的生成模型,通过将每个模态专注于特定的工业场景或特征等方式提高生成图像的逼真度和多样性,包括考虑但不限于在光照、材质、形状等方面的变化.

3) 基于生成图像的关键参数检测和工业过程控制:在工业过程中,生成的图像中不仅包含视觉信息,其还蕴含关键过程参数和控制信息.研究如何准确提取生成图像中的关键过程参数并将其应用于工业过程的控制和优化是一个重要的开放性问题.未来的研究可探索基于生成图像的关键过程参数检测和工业过程优化控制策略,进而实现工业过程的智慧运行.

综上所述,面向工业过程的图像生成及其应用研究面临着大规模模型的融入、多模态场景的生成和基于生成图像的关键参数检测和工业过程控制等挑战.解决这些挑战将为工业领域提供更高质量、多样化和可控的图像生成技术,并推动工业过程的创新和进步.

## References

- Pan Y J, Yang C J, An R Q, Sun Y X. Fault detection with improved principal component pursuit method. *Chemometrics and Intelligent Laboratory Systems*, 2016, **157**: 111–119
- Wu Gao-Chang, Liu Qiang, Chai Tian-You, Qin S. Joe. Abnormal condition diagnosis through deep learning of image sequences for fused magnesium furnaces. *Acta Automatica Sinica*, 2019, **45**(8): 1475–1485  
(吴高昌, 刘强, 柴天佑, 秦泗钊. 基于时序图像深度学习的电熔镁炉异常工况诊断. *自动化学报*, 2019, **45**(8): 1475–1485)
- Yi Ning-Chun, Gui Wei-Hua, Liang Xiao-Jun, Zhang Chao-Bo, Tang Feng-Run, Yang Chun-Hua. Keypoints identification and position monitoring of the blower pipe for volatilization kiln. *Control Theory & Applications*, DOI: 10.7641/CTA.2023.20541 (易俊纯, 桂卫华, 梁晓俊, 张超波, 唐峰润, 阳春华. 挥发窑鼓风管的关键点识别及其摆放位置监测. *控制理论与应用*, DOI: 10.7641/CTA.2023.20541)
- Li Pei, Yang Chun-Hua, He Jian-Jun, Gui Wei-Hua. Smelting condition identification and prediction for submerged arc furnace based on shadow-trend-comparison. *Acta Automatica Sinica*, 2021, **47**(6): 1343–1354  
(李沛, 阳春华, 贺建军, 桂卫华. 基于影子趋势对比的矿热炉炉况在线辨识及趋势预测. *自动化学报*, 2021, **47**(6): 1343–1354)
- Yang J C, Wang C G, Jiang B, Song H B, Meng Q G. Visual perception enabled industry intelligence: State of the art, challenges and prospects. *IEEE Transactions on Industrial Informatics*, 2021, **17**(3): 2204–2219
- Roy S K, Krishna G, Dubey S R, Chaudhuri B B. HybridSN: Exploring 3-D-2-D CNN feature hierarchy for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 2020, **17**(2): 277–281
- Ahammad S H, Rajesh V, Rahman M Z U, Lay-Ekuakille A. A hybrid CNN-based segmentation and boosting classifier for real time sensor spinal cord injury data. *IEEE Sensors Journal*, 2020, **20**(17): 10092–10101
- Sun Y N, Xue B, Zhang M J, Yen G G, Lv J C. Automatically designing CNN architectures using the genetic algorithm for image classification. *IEEE Transactions on Cybernetics*, 2020, **50**(9): 3840–3854
- Zhou P, Gao B H, Wang S, Chai T Y. Identification of abnormal conditions for fused magnesium melting process based on deep learning and multi-source information fusion. *IEEE Transactions on Industrial Electronics*, 2022, **69**(3): 3017–3026
- Gui J, Sun Z N, Wen Y G, Tao D C, Ye J P. A review on generative adversarial networks: Algorithms, theory, and applications. *IEEE Transactions on Knowledge and Data Engineering*, 2023, **35**(4): 3313–3332
- Qiao Jun-Fei, Duan Hao-Shan, Tang Jian, Meng Xi. Recognition of MSWI combustion conditions based on feature extraction of flame images. *Control Engineering of China*, 2022, **29**(7): 1153–1161  
(乔俊飞, 段涛杉, 汤健, 蒙西. 基于火焰图像特征提取的MSWI燃烧工况识别. *控制工程*, 2022, **29**(7): 1153–1161)
- Tang Jian, Xia Heng, Yu Wen, Qiao Jun-Fei. Research status and prospects of intelligent optimization control for municipal solid waste incineration process. *Acta Automatica Sinica*, 2023, **49**(10): 2019–2059  
(汤健, 夏恒, 余文, 乔俊飞. 城市固废焚烧过程智能优化控制研究现状与展望. *自动化学报*, 2023, **49**(10): 2019–2059)
- Guo Hai-Tao, Tang Jian, Xia Heng, Qiao Jun-Fei. Combustion line extreme abnormal flame image adversarial generation for municipal solid waste incineration processes. *Proceedings of the CSEE*, 2022  
(郭海涛, 汤健, 夏恒, 乔俊飞. 城市固废焚烧过程燃烧线极端异常火焰图像对抗生成. *中国电机工程学报*, 2022)
- Guo Hai-Tao, Tang Jian, Ding Hai-Xu, Qiao Jun-Fei. Combustion states recognition method of MSWI process based on mixed data enhancement. *Acta Automatica Sinica*, DOI: 10.16383/j.aas.c210843  
(郭海涛, 汤健, 丁海旭, 乔俊飞. 基于混合数据增强的MSWI过程燃烧状态识别. *自动化学报*, DOI: 10.16383/j.aas.c210843)
- Rosenfeld A. Picture processing by computer. *Computer of Monographs and Textbooks*, 1969, **1**(3): 147–176
- Salakhutdinov R. Learning deep generative models. *Annual Review of Statistics & Its Application*, 2015, **2**(1): 361–385
- Liu Jian-Wei, Liu Yuan, Luo Xiong-Lin. Research process of Boltzmann machine. *Journal of Computer Research and Development*, 2014, **51**(1): 1–16  
(刘建伟, 刘媛, 罗雄麟. 玻尔兹曼机研究进展. *计算机研究与发展*, 2014, **51**(1): 1–16)
- Zhai Zheng-Li, Liang Zhen-Ming, Zhou Wei, Sun Xia. Research overview of variational auto-encoders models. *Computer Engineering and Applications*, 2019, **55**(3): 1–9  
(翟正立, 梁振明, 周炜, 孙霞. 变分自编码器模型综述. *计算机工程与应用*, 2019, **55**(3): 1–9)
- Wang Kun-Feng, Gou Chao, Duan Yan-Jie, Lin Yi-Lun, Zheng Xin-Hu, Wang Fei-Yue. Generative adversarial networks: The state of the art and beyond. *Acta Automatica Sinica*, 2017, **43**(3): 321–332  
(王坤峰, 苟超, 段艳杰, 林懿伦, 郑心湖, 王飞跃. 生成式对抗网



- 络 GAN 的研究进展与展望. *自动化学报*, 2017, **43**(3): 321–332
- 20 Lin Yi-Lun, Dai Xing-Yuan, Li Li, Wang Xiao, Wang Fei-Yue. The new frontier of AI research: Generative adversarial networks. *Acta Automatica Sinica*, 2018, **44**(5): 775–792 (林懿伦, 戴星原, 李力, 王晓, 王飞跃. 人工智能研究的新前线: 生成式对抗网络. *自动化学报*, 2018, **44**(5): 775–792)
- 21 Wang Kun-Feng, Zuo Wang-Meng, Tan Ying, Qin Tao, Li Li, Wang Fei-Yue. Generative adversarial networks: From generating data to creating intelligence. *Acta Automatica Sinica*, 2018, **44**(5): 769–774 (王坤峰, 左旺孟, 谭莹, 秦涛, 李力, 王飞跃. 生成式对抗网络: 从生成数据到创造智能. *自动化学报*, 2018, **44**(5): 769–774)
- 22 Hu Ming-Fei, Zuo Xin, Liu Jian-Wei. Survey on deep generative model. *Acta Automatica Sinica*, 2022, **48**(1): 40–74 (胡铭菲, 左信, 刘建伟. 深度生成模型综述. *自动化学报*, 2022, **48**(1): 40–74)
- 23 Kingma D P, Welling M. Auto-encoding variational bayes. arXiv: 1312.6114, 2022.
- 24 Croitoru F, Hondru V, Ionescu R T, Shah M. Diffusion models in vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, **45**(9): 10850–10869
- 25 Sohl-Dickstein J, Weiss E A, Maheswaranathan N, Ganguli S. Deep unsupervised learning using nonequilibrium thermodynamics. arXiv: 1503.03585, 2015.
- 26 Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. arXiv: 2006.11239, 2020.
- 27 Song Y, Ermon S. Generative modeling by estimating gradients of the data distribution. arXiv: 1907.05600, 2020.
- 28 Song Y, Sohl-Dickstein J, Kingma D P, Kumar A, Ermon S, Poole B. Score-based generative modeling through stochastic differential equations. arXiv: 2011.13456, 2021.
- 29 Smolensky P. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Cambridge: MIT Press, 1987. 194–281
- 30 Hinton G E, Osindero S, Teh Y. A fast learning algorithm for deep belief nets. *Neural Computation*, 2006, **18**(7): 1527–1554
- 31 Salakhutdinov R, Hinton G E. Deep Boltzmann machines. In: *Proceedings of the International Conference on Artificial Intelligence and Statistics*. Shanghai, China: PMLR, 2009. 448–455
- 32 Goodfellow I J, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. In: *Proceedings of the 27th International Conference on Neural Information Processing Systems*. Montreal, Canada: MIT Press, 2014. 2672–2680
- 33 Dinh L, Krueger D, Bengio Y. NICE: Non-linear independent components estimation. arXiv: 1410.8516, 2015.
- 34 Larochelle H, Murray I. The neural autoregressive distribution estimator. In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. Fort Lauderdale, USA: PMLR, 2011. 29–37
- 35 Oord A V D, Kalchbrenner N, Kavukcuoglu K. Pixel recurrent neural networks. arXiv: 1601.06759, 2016.
- 36 Germain M, Gregor K, Murray I, Larochelle H. MADE: Masked autoencoder for distribution estimation. In: *Proceedings of the 32nd International Conference on International Conference on Machine Learning*. Lille, France: JMLR, 2015. 881–889
- 37 Socher R, Pennington J, Huang E H, Ng A Y, Manning C D. Semi-supervised recursive autoencoders for predicting sentiment distributions. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Edinburgh, UK: Association for Computational Linguistics, 2011. 151–161
- 38 Xu Q T, Huang G, Yuan Y, Guo C, Sun Y, Wu F, et al. An empirical study on evaluation metrics of generative adversarial networks. arXiv: 1806.07755, 2018.
- 39 Duan J J, He Y G, Du B L, Ghandour R M R, Wu W J, Zhang H. Intelligent localization of transformer internal degradations combining deep convolutional neural networks and image segmentation. *IEEE Access*, 2019, **7**: 62705–62720
- 40 Gong X J, Yao Q, Wang M L, Lin Y. A deep learning approach for oriented electrical equipment detection in thermal images. *IEEE Access*, 2018, **6**: 41590–41597
- 41 Zhao Z B, Fan X Q, Xu G Z, Zhang L, Qi Y C, Zhang K. Aggregating deep convolutional feature maps for insulator detection in infrared images. *IEEE Access*, 2017, **5**: 21831–21839
- 42 Jia F, Lei Y G, Lu N, Xing S B. Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization. *Mechanical Systems and Signal Processing*, 2018, **110**: 349–367
- 43 Gadermayr M, Li K, Müller M, Truhn D, Krämer N, Merhof D, et al. Domain-specific data augmentation for segmenting MR images of fatty infiltrated human thighs with neural networks. *Journal of Magnetic Resonance Imaging*, 2019, **49**(6): 1676–1683
- 44 Zhu X Y, Liu Y F, Qin Z C, Li J H. Data augmentation in emotion classification using generative adversarial networks. arXiv: 1711.00648, 2017.
- 45 He K M, Zhang X Y, Ren S Q, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, USA: IEEE, 2016. 770–778
- 46 Gao X, Deng F, Yue X H. Data augmentation in fault diagnosis based on the Wasserstein generative adversarial network with gradient penalty. *Neurocomputing*, 2020, **396**: 487–494
- 47 Wong S C, Gatt A, Stamatescu V, McDonnell M D. Understanding data augmentation for classification: When to warp? In: *Proceedings of the International Conference on Digital Image Computing: Techniques and Applications (DICTA)*. Gold Coast, Australia: IEEE, 2016. 1–6
- 48 Cuello K C, Chacón M, Mery D, Acuña G. Neural network method for failure detection with skewed class distribution. *Insight*, 2004, **46**(7): 399–402
- 49 Mirza M, Osindero S. Conditional generative adversarial nets. arXiv: 1411.1784, 2014.
- 50 Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv: 1511.06434, 2016.
- 51 Denton E, Chintala S, Szlam A, Fergus R. Deep generative image models using a Laplacian pyramid of adversarial networks. arXiv: 1506.05751, 2015.
- 52 Larsen A B L, Sonderby S K, Larochelle H, Winther O. Autoencoding beyond pixels using a learned similarity metric. In: *Proceedings of the 33rd International Conference on International Conference on Machine Learning*. New York, USA: JMLR, 2016. 1558–1566
- 53 Donahue J, Krahenbühl P, Darrell T. Adversarial feature learning. arXiv: 1605.09782, 2017.
- 54 Liu M Y, Tuzel O. Coupled generative adversarial networks. In: *Proceedings of the 30th International Conference on Neural Information Processing Systems*. New York, USA: Curran Associates Inc., 2016. 469–477
- 55 Chen X, Duan Y, Houthoofd R, Schulman J, Sutskever I, Abbeel P. InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets. In: *Proceedings of the 30th International Conference on Neural Informa-*

- tion Processing Systems. Barcelona, Spain: Curran Associates Inc., 2016. 2180–2188
- 56 Nowozin S, Cseke B, Tomioka R. *f*-GAN: Training generative neural samplers using variational divergence minimization. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, Spain: Curran Associates Inc., 2016. 271–279
- 57 Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A, Chen X. Improved techniques for training GANs. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, Spain: Curran Associates Inc., 2016. 2234–2232
- 58 Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville A C. Improved training of wasserstein GANs. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc., 2017. 5769–5779
- 59 Odena A, Olah C, Shlens J. Conditional image synthesis with auxiliary classifier GANs. In: Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia: JMLR, 2017. 2642–2651
- 60 Zhang H, Xu T, Li H S, Zhang S T, Wang X G, Huang X L, et al. StackGAN: Text to photo-realistic image synthesis with stacked generative adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017. 5908–5916
- 61 Brock A, Donahue J, Simonyan K. Large scale GAN training for high fidelity natural image synthesis. arXiv: 1809.11096, 2019.
- 62 Zhang H, Goodfellow I, Metaxas D, Odena A. Self-attention generative adversarial networks. arXiv: 1805.08318, 2019.
- 63 Qi G J. Loss-sensitive generative adversarial networks on Lipschitz densities. *International Journal of Computer Vision*, 2020, **128**(5): 1118–1140
- 64 Liu R, Ge Y X, Choi C L, Wang X G, Li H S. DivCo: Diverse conditional image synthesis via contrastive generative adversarial network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE, 2021. 16372–16381
- 65 Liao W T, Hu K, Yang M Y, Rosenhahn B. Text to image generation with semantic-spatial aware GAN. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, USA: IEEE, 2022. 18166–18175
- 66 Li S X, Yu Z Y, Xiang M, Mandic D. Reciprocal GAN through characteristic functions (RCF-GAN). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, **45**(2): 2246–2263
- 67 Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial networks. In: Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia: JMLR, 2017. 214–223
- 68 Miyato T, Kataoka T, Koyama M, Yoshida Y. Spectral normalization for generative adversarial networks. arXiv: 1802.05957, 2018.
- 69 Demir U, Unal G. Patch-based image inpainting with generative adversarial networks. arXiv: 1803.07422, 2018.
- 70 Sohn K, Yan X C, Lee H. Learning structured output representation using deep conditional generative models. In: Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal, Canada: MIT Press, 2015. 3483–3491
- 71 Makhzani A, Shlens J, Jaitly N, Goodfellow I, Frey B. Adversarial autoencoders. arXiv: 1511.05644, 2016.
- 72 Burda Y, Grosse R, Salakhutdinov R. Importance weighted autoencoders. arXiv: 1509.00519, 2016.
- 73 Kulkarni T D, Whitney W F, Kohli P, Tenenbaum J B. Deep convolutional inverse graphics network. In: Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal, Canada: MIT Press, 2015. 2539–2547
- 74 Sonderby C K, Raiko T, Maaloe L, Sonderby S K, Winther O. Ladder variational autoencoders. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, Spain: Curran Associates Inc., 2016. 3745–3753
- 75 Abbasnejad M E, Dick A, Hengel A V D. Infinite variational autoencoder for semi-supervised learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017. 781–790
- 76 Zhao S J, Song J M, Ermon S. InfoVAE: Information maximizing variational autoencoders. arXiv: 1706.02262, 2018.
- 77 Cai L, Gao H Y, Ji S W. Multi-stage variational auto-encoders for coarse-to-fine image generation. arXiv: 1705.07202, 2017.
- 78 Wang Y, Dai B, Hua G, Aston J, Wipf D. Recurrent variational autoencoders for learning nonlinear generative models in the presence of outliers. *IEEE Journal of Selected Topics in Signal Processing*, 2018, **12**(6): 1615–1627
- 79 Zheng K, Cheng Y F, Kang X J, Yao H, Tian T. Conditional introspective variational autoencoder for image synthesis. *IEEE Access*, 2020, **8**: 153905–153913
- 80 Ramesh A, Pavlov M, Goh G, Gray S, Voss C, Radford A, et al. Zero-shot text-to-image generation. arXiv: 2102.12092, 2021.
- 81 Ramesh A, Dhariwal P, Nichol A, Chu C, Chen M. Hierarchical text-conditional image generation with CLIP latents. arXiv: 2204.06125, 2022.
- 82 Xu J, Liu B, Xiao Y S. A variational inference method for few-shot learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023, **33**(1): 269–282
- 83 Rezende D J, Mohamed S. Variational inference with normalizing flows. In: Proceedings of the 32nd International Conference on International Conference on Machine Learning. Lille, France: JMLR, 2015. 1530–1538
- 84 Dinh L, Sohl-Dickstein J, Bengio S. Density estimation using real NVP. arXiv: 1605.08803, 2016.
- 85 Kingma D P, Salimans T, Jozefowicz R, Chen X, Sutskever I, Welling M. Improving variational inference with inverse autoregressive flow. arXiv: 1606.04934, 2016.
- 86 Papamakarios G, Pavlakou T, Murray I. Masked autoregressive flow for density estimation. arXiv: 1705.07057, 2018.
- 87 Kingma D P, Dhariwal P. Glow: Generative flow with invertible  $1 \times 1$  convolutions. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal, Canada: Curran Associates Inc., 2018. 10236–10245
- 88 Behrmann J, Grathwohl W, Chen R T Q, Duvenaud D, Jacobsen J. Invertible residual networks. arXiv: 1811.00995, 2019.
- 89 An J, Huang S Y, Song Y B, Dou D J, Liu W, Luo J B. ArtFlow: Unbiased image style transfer via reversible neural flows. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE, 2021. 862–871
- 90 Chen X, Mishra N, Rohaninejad M, Abbeel P. PixelSNAIL: An improved autoregressive generative model. arXiv: 1712.09763, 2017.
- 91 Huang J H, Chia Y K, Yu S, Yee K, Küster D, Krumhuber E

- G, et al. Single image video prediction with auto-regressive GANs. *Sensors (Basel)*, 2022, **22**(9): Article No. 3533
- 92 Choi J H, Kim S, Jeong Y, Gwon Y, Yoon S. ILVR: Conditioning method for denoising diffusion probabilistic models. arXiv: 2108.02938, 2021.
- 93 Dhariwal P, Nichol A. Diffusion models beat GANs on image synthesis. arXiv: 2105.05233, 2021.
- 94 Rombach R, Blattmann A, Lorenz D, Esser P, Ommer B. High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, USA: IEEE, 2022. 10674–10685
- 95 Wu C F, Yin S M, Qi W Z, Wang X D, Tang Z C, Duan N. Visual ChatGPT: Talking, drawing and editing with visual foundation models. arXiv: 2303.04671, 2023.
- 96 Tsai T, Lee Y. A light-weight neural network for wafer map classification based on data augmentation. *IEEE Transactions on Semiconductor Manufacturing*, 2020, **33**(4): 663–672
- 97 Cha J, Oh S, Kim D, Jeong J. A defect detection model for imbalanced wafer image data using CAE and xception. In: Proceedings of the International Conference on Intelligent Data Science Technologies and Applications (IDSTA). Valencia, Spain: IEEE, 2020. 28–33
- 98 Wang J L, Yang Z L, Zhang J, Zhang Q H, Chien W K. AdaBalGAN: An improved generative adversarial network with imbalanced learning for wafer defective pattern recognition. *IEEE Transactions on Semiconductor Manufacturing*, 2019, **32**(3): 310–319
- 99 Niu Z W, Reformat M Z, Tang W H, Zhao B N. Electrical equipment identification method with synthetic data using edge-oriented generative adversarial network. *IEEE Access*, 2020, **8**: 136487–136497
- 100 Fu M R, Liu J H, Zhang H G, Lu S X. Multisensor fusion for magnetic flux leakage defect characterization under information incompleteness. *IEEE Transactions on Industrial Electronics*, 2021, **68**(5): 4382–4392
- 101 Wang B. A study on spot welding quality judgment based on improved generative adversarial network and auto-encoder. *Mechanical Systems and Signal Processing*, 2022, **165**: Article No. 108318
- 102 Liu L Z, Cao D H, Wu Y B, Wei T R. Defective samples simulation through adversarial training for automatic surface inspection. *Neurocomputing*, 2019, **360**: 230–245
- 103 Niu S L, Li B, Wang X G, Lin H. Defect image sample generation with GAN for improving defect recognition. *IEEE Transactions on Automation Science and Engineering*, 2020, **17**(3): 1611–1622
- 104 Nguyen T D, Le T, Vu H, Phung D. Dual discriminator generative adversarial nets. arXiv: 1709.03831, 2017.
- 105 Shao G F, Huang M, Gao F Q, Liu T D, Li L D. DuCaGAN: Unified dual capsule generative adversarial network for unsupervised image-to-image translation. *IEEE Access*, 2020, **8**: 154691–154707
- 106 Kim M, Jo H, Ra M, Kim W. Weakly-supervised defect segmentation on periodic textures using CycleGAN. *IEEE Access*, 2020, **8**: 176202–176216
- 107 Yi J A, Chen Z, Li D Y, Li J, Liu J. Conditional generative adversarial network for welding deformation field prediction of butt-welded plates. *Journal of Constructional Steel Research*, 2023, **201**: Article No. 107755
- 108 Wang K, Liu M Z. Object recognition at night scene based on DCGAN and faster R-CNN. *IEEE Access*, 2020, **8**: 193168–193182
- 109 Feng Y H, Chen Z Y, Wang D L, Chen J, Feng Z L. Deep-Welding: A deep learning enhanced approach to GTAW using multisource sensing images. *IEEE Transactions on Industrial Informatics*, 2020, **16**(1): 465–474
- 110 Isola P, Zhu J Y, Zhou T H, Efros A A. Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017. 5967–5976
- 111 Lin H, Menendez P. Image denoising of printed circuit boards using conditional generative adversarial network. In: Proceedings of the IEEE 10th International Conference on Mechanical and Intelligent Manufacturing Technologies (ICMIMT). Cape Town, South Africa: IEEE, 2019. 98–103
- 112 Chen S K, Shi D M, Sadiq M, Cheng X C. Image denoising with generative adversarial networks and its application to cell image enhancement. *IEEE Access*, 2020, **8**: 82819–82831
- 113 Wu C X, Ju B B, Wu Y, Xiong N X. SlimRGBD: A geographic information photography noise reduction system for aerial remote sensing. *IEEE Access*, 2020, **8**: 15144–15158
- 114 Li Y, Wang H Z, Dong X T. The denoising of desert seismic data based on Cycle-GAN with unpaired data training. *IEEE Geoscience and Remote Sensing Letters*, 2021, **18**(11): 2016–2020
- 115 Jiang X, Lu L, Zhu M, Hao Z C, Gao W. Haze relevant feature attention network for single image dehazing. *IEEE Access*, 2021, **9**: 106476–106488
- 116 Li Z H, Huang J Z, Yu L F, Chi Y J, Jin M W. Low-fuse CT image denoising using cycle-consistent adversarial networks. In: Proceedings of the IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC). Manchester, UK: IEEE, 2019. 1–3
- 117 Ma Y H, Chen X J, Zhu W F, Cheng X N, Xiang D H, Shi F. Speckle noise reduction in optical coherence tomography images based on edge-sensitive cGAN. *Biomedical Optics Express*, 2018, **9**(11): 5129–5146
- 118 Zhou Y, Yu K, Wang M, Ma Y H, Peng Y Y, Chen Z Y, et al. Speckle noise reduction for OCT images based on image style transfer and conditional GAN. *IEEE Journal of Biomedical and Health Informatics*, 2022, **26**(1): 139–150
- 119 Theis L, Oord A V D, Bethge M. A note on the evaluation of generative models. arXiv: 1511.01844, 2016.
- 120 Shmelkov K, Schmid C, Alahari K. How good is my GAN? In: Proceedings of the European Conference on Computer Vision (ECCV). Berlin, Heidelberg: Springer-Verlag, 2018. 218–234
- 121 Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A, Chen X. Improved techniques for training GANs. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, Spain: Curran Associates Inc., 2016. 2234–2242
- 122 Heusel M, Ramsauer H, Unterthiner T, Nessler B, Hochreiter S. GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc., 2017. 6629–6640
- 123 Lucic M, Kurach K, Michalski M, Bousquet O, Gelly S. Are GANs created equal? A large-scale study. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal, Canada: Curran Associates Inc., 2018. 698–707
- 124 Borji A. Pros and cons of GAN evaluation measures. *Computer Vision and Image Understanding*, 2019, **179**: 41–65
- 125 Borji A. Pros and cons of GAN evaluation measures: New developments. *Computer Vision and Image Understanding*, 2022,

- 215: Article No. 103329
- 126 Oord A V D, Kalchbrenner N, Vinyals O, Espeholt L, Graves A, Kavukcuoglu K. Conditional image generation with Pixel-CNN decoders. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, Spain: Curran Associates Inc., 2016. 4797–4805
- 127 Ye Y C, Wang L J, Wu Y, Chen Y P, Tian Y L, Liu Z C, et al. GAN quality index (GQI) by GAN-induced classifier. In: Proceedings of the 6th International Conference on Learning Representations. Vancouver, Canada: ICLR, 2018.
- 128 Lehmann E L, Romano J P, Casella G. *Testing Statistical Hypotheses*. New York: Springer, 2005.
- 129 Wang Z, Bovik A C, Sheikh H R, Simoncelli E P. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 2004, **13**(4): 600–612
- 130 Snell J, Ridgeway K, Liao R J, Roads B D, Mozer M C, Zemel R S. Learning to generate images with perceptual similarity metrics. arXiv: 1511.06409, 2017.
- 131 Juefei-Xu F, Boddeti V N, Savvides M. Gang of GANs: Generative adversarial networks with maximum margin ranking. arXiv: 1704.04865, 2017.
- 132 Gretton A, Borgwardt K M, Rasch M J, Scholkopf B, Smola A. A kernel two-sample test. *The Journal of Machine Learning Research*, 2012, **13**(1): 723–773
- 133 Che T, Li Y R, Jacob A P, Bengio Y, Li W J. Mode regularized generative adversarial networks. arXiv: 1612.02136, 2016.
- 134 Wang Y X, Zhang L C, Weijer J V D. Ensembles of generative adversarial networks. arXiv: 1612.00991, 2016.
- 135 Im D J, Kim C D, Jiang H, Memisevic R. Generating images with recurrent adversarial networks. arXiv: 1602.05110, 2016.
- 136 Tolstikhin I, Gelly S, Bousquet O, Simon-Gabriel C, Scholkopf B. AdaGAN: Boosting generative models. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc., 2017. 5430–5439
- 137 Gurumurthy S, Sarvadevabhatla R K, Babu R V. DeLiGAN: Generative adversarial networks for diverse and limited data. arXiv: 1706.02071, 2017.
- 138 Zhou Z M, Cai H, Rong S, Song Y X, Ren K, Wang J. Activation maximization generative adversarial nets. arXiv: 1703.02000, 2018.
- 139 Arora S, Zhang Y. Do GANs actually learn the distribution? An empirical study. arXiv: 1706.08224, 2017.
- 140 Yang J W, Kannan A, Batra D, Parikh D. LR-GAN: Layered recursive generative adversarial networks for image generation. arXiv: 1703.01560, 2017.
- 141 Xiang S T, Li H. On the effects of batch and weight normalization in generative adversarial networks. arXiv: 1704.03971, 2017.
- 142 Zeng Y, Lu H C, Borji A. Statistics of deep generated images. arXiv: 1708.02688, 2019.
- 143 Karras T, Aila T, Laine S, Lehtinen J. Progressive growing of GANs for improved quality, stability, and variation. arXiv preprint arXiv: 1710.10196, 2018.
- 144 Santurkar S, Schmidt L, Madry A. A classification-based study of covariate shift in GAN distributions. arXiv: 1711.00970, 2018.
- 145 Richardson E, Weiss Y. On GANs and GMMs. arXiv: 1805.12462, 2018.
- 146 Olsson C, Bhupatiraju S, Brown T, Odena A, Goodfellow I. Skill rating for generative models. arXiv: 1808.04888, 2018.
- 147 Zhang Z F, Song Y, Qi H R. Decoupled learning for conditional adversarial networks. arXiv: 1801.06790, 2018.
- 148 Khrulkov V, Oseledets I. Geometry score: A method for comparing generative adversarial networks. arXiv: 1802.02664, 2018.
- 149 Liu S H, Wei Y, Lu J W, Zhou J. An improved evaluation framework for generative adversarial networks. arXiv: 1803.07474, 2018.
- 150 Solovetchik M, Diskin T, Morin E, Wiesel A. Conditional Fréchet inception distance. arXiv: 2103.11521, 2022.
- 151 Chong M J, Forsyth D. Effectively unbiased FID and inception score and where to find them. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE, 2020. 6069–6078
- 152 Mathiasen A, Hvilshøj F. Fast Fréchet inception distance. arXiv: 2009.14075, 2020.
- 153 Nash C, Menick J, Dieleman S, Battaglia P W. Generating images with sparse representations. arXiv: 2103.03841, 2021.
- 154 Bai C, Lin H, Raffel C, Kan W C. On training sample memorization: Lessons from benchmarking generative modeling with a large-scale competition. In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. New York, USA: ACM, 2021. 2534–2542
- 155 Zhou X K, Hu Y Y, Wu J Y, Liang W, Ma J H, Jin Q. Distribution bias aware collaborative generative adversarial network for imbalanced deep learning in industrial IoT. *IEEE Transactions on Industrial Informatics*, 2023, **19**(1): 570–580
- 156 Neal R M. Annealed importance sampling. *Statistics and Computing*, 2001, **11**(2): 125–139
- 157 Zeiler M D, Fergus R. Visualizing and understanding convolutional networks. In: Proceedings of the European Conference on Computer Vision. Zurich, Switzerland: Springer, 2014. 818–833
- 158 Higgins I, Matthey L, Pal A, Burgess C P, Glorot X, Botvinick M, et al. beta-VAE: Learning basic visual concepts with a constrained variational framework. In: Proceedings of the 5th International Conference on Learning Representations. Toulon, France: ICLR, 2016.
- 159 Mathieu M, Zhao J B, Sprechmann P, Ramesh A, LeCun Y. Disentangling factors of variation in deep representation using adversarial training. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, Spain: Curran Associates Inc., 2016. 5047–5055
- 160 Bau D, Zhou B L, Khosla A, Oliva A, Torralba A. Network dissection: Quantifying interpretability of deep visual representations. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017. 3319–3327
- 161 Huang X, Li Y X, Poursaeed O, Hopcroft J, Belongie S. Stacked generative adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017. 1866–1875
- 162 Xian C, Zhang D, Dai C, Wang C C L. Fast generation of high-fidelity RGB-D images by deep learning with adaptive convolution. *IEEE Transactions on Automation Science and Engineering*, 2021, **18**(3): 1328–1340
- 163 Yi Z L, Zhang H, Tan P, Gong M L. DualGAN: Unsupervised dual learning for image-to-image translation. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017. 2868–2876
- 164 Xiao C W, Li B, Zhu J Y, He W, Liu M Y, Song D. Generating adversarial examples with adversarial networks. arXiv: 1801.02610, 2019.
- 165 Srivastava A, Valkov L, Russell C, Gutmann M U, Sutton C. VEEGAN: Reducing mode collapse in GANs using implicit variational learning. In: Proceedings of the 31st International

- Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc., 2017. 3310–3320
- 166 Lin Z N, Khetan A, Fanti G, Oh S. PacGAN: The power of two samples in generative adversarial networks. arXiv: 1712.04086, 2017.
- 167 Zhou S, Gordon M L, Krishna R, Narcomey A, Li F F, Bernstein M S. HYPE: A benchmark for human eye perceptual evaluation of generative models. arXiv: 1904.01121, 2019.
- 168 Kolchinski Y A, Zhou S, Zhao S J, Gordon M, Ermon S. Approximating human judgment of generated image quality. arXiv: 1912.12121, 2019.
- 169 Bau D, Zhu J Y, Wulff J, Peebles W, Zhou B L, Strobel H, et al. Seeing what a GAN cannot generate. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, South Korea: IEEE, 2019. 4501–4510
- 170 Yu N, Davis L, Fritz M. Attributing fake images to GANs: Learning and analyzing GAN fingerprints. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, South Korea: IEEE, 2019. 7555–7565
- 171 Wang S Y, Wang O, Zhang R, Owens A, Efros A A. CNN-generated images are surprisingly easy to spot... for now. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE, 2020. 8692–8701
- 172 Chai L, Bau D, Lim S, Isola P. What makes fake images detectable? Understanding properties that generalize. In: Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer, 2020. 103–120
- 173 Gagnaniello D, Cozzolino D, Marra F, Poggi G, Verdoliva L. Are GAN generated images easy to detect? A critical analysis of the state-of-the-art. In: Proceedings of the IEEE International Conference on Multimedia and Expo (ICME). Shenzhen, China: IEEE, 2021. 1–6
- 174 Liu J T, Hou W M, Luo X, Su J, Hou Y L, Wang Z Z. SI-SA GAN: A generative adversarial network combined with spatial information and self-attention for removing thin cloud in optical remote sensing images. *IEEE Access*, 2022, **10**: 114318–114330
- 175 Liu Z M, Fang Y, Liu Y, Zheng V W. Neighbor-anchoring adversarial graph neural networks. *IEEE Transactions on Knowledge and Data Engineering*, 2023, **35**(1): 784–795
- 176 Yang L, Liu Y H, Peng J Z. An automatic detection and identification method of welded joints based on deep neural network. *IEEE Access*, 2019, **7**: 164952–164961
- 177 Nie Y D, Bian C J, Li L G. Object tracking in satellite videos based on siamese network with multidimensional information-aware and temporal motion compensation. *IEEE Geoscience and Remote Sensing Letters*, 2022, **19**: Article No. 6517005
- 178 Li J, Xing J H, Du S H, Du S H, Zhang C Y, Li W. Change detection of open-pit mine based on siamese multiscale network. *IEEE Geoscience and Remote Sensing Letters*, 2022, **20**: Article No. 2500105
- 179 Tao R, Gavves E, Smeulders A W M. Siamese instance search for tracking. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE, 2016. 1420–1429
- 180 Bertinetto L, Valmadre J, Henriques J F, Vedaldi A, Torr P H S. Fully-convolutional siamese networks for object tracking. arXiv: 1606.09549, 2021.
- 181 Li B, Yan J J, Wu W, Zhu Z, Hu X L. High performance visual tracking with siamese region proposal network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 8971–8980
- 182 Valmadre J, Bertinetto L, Henriques J, Vedaldi A, Torr P H S. End-to-end representation learning for correlation filter based tracking. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017. 5000–5008
- 183 Guo Q, Feng W, Zhou C, Huang R, Wan L, Wang S. Learning dynamic siamese network for visual object tracking. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017. 1781–1789
- 184 Guo J, Xu T F, Jiang S W, Shen Z Y. Generating reliable online adaptive templates for visual tracking. In: Proceedings of the 25th IEEE International Conference on Image Processing (ICIP). Athens, Greece: IEEE, 2018. 226–230
- 185 Cheng X M, Cheng S S. Thermal image generation for blast furnace chute based on generative adversarial network. *Signal, Image and Video Processing*, 2023, **17**: 2595–2606
- 186 Niu S L, Li B, Wang X G, Peng Y. Region-and strength-controllable GAN for defect generation and segmentation in industrial images. *IEEE Transactions on Industrial Informatics*, 2021, **18**(7): 4531–4541
- 187 Wen L, Wang Y, Li X Y. A new cycle-consistent adversarial networks with attention mechanism for surface defect classification with small samples. *Statistics and Computing*, 2022, **18**(12): 8988–8998
- 188 Fang Q, Ibarra-Castanedo C, Duan Y X, Erazo-Aux J, Garrido I, Maldague X. Defect enhancement and image noise reduction analysis using partial least square-generative adversarial networks (PLS-GANs) in thermographic nondestructive evaluation. *Journal of Nondestructive Evaluation*, 2021, **40**: Article No. 92
- 189 Huang T, Li S J, Jia X, Lu H C, Liu J Z. Neighbor2Neighbor: Self-supervised denoising from single noisy images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE, 2021. 14776–14785
- 190 Tang Jian, Cui Can-Lin, Xia Heng, Qiao Jun-Fei. A survey of virtual sample generation for complex industrial processes. *Acta Automatica Sinica*, DOI: [10.16383/j.aas.c221006](https://doi.org/10.16383/j.aas.c221006) (汤健, 崔璨麟, 夏恒, 乔俊飞. 面向复杂工业过程的虚拟样本生成综述. *自动化学报*, DOI: [10.16383/j.aas.c221006](https://doi.org/10.16383/j.aas.c221006))
- 191 Wang Yao-Nan, Jiang Yi-Ming, Jiang Jiao, Zhang Hui, Tan Hao-Ran, Peng Wei-Xing, et al. Key technologies of robot perception and control and its intelligent manufacturing applications. *Acta Automatica Sinica*, 2023, **49**(3): 494–513 (王耀南, 江一鸣, 姜娇, 张辉, 谭浩然, 彭伟星, 等. 机器人感知与控制关键技术及其智能制造应用. *自动化学报*, 2023, **49**(3): 494–513)



汤健 北京工业大学信息学部教授。主要研究方向为小样本数据建模, 城市固废处理过程智能控制。本文通信作者。

E-mail: [freeflytang@bjut.edu.cn](mailto:freeflytang@bjut.edu.cn)

(TANG Jian Professor at the Faculty of Information Technology, Beijing University of Technology. His research interest covers small sample data modeling and intelligent control of municipal solid waste treatment process. Corresponding author of this paper.)



**郭海涛** 北京工业大学信息学部硕士研究生. 主要研究方向为面向城市固废焚烧过程的图像处理研究.

E-mail: guoht@emails.edu.cn

**(GUO Hai-Tao** Master student at the Faculty of Information Technology, Beijing University of Technology.

His main research interest is image processing research for municipal solid waste incineration process.)



**夏 恒** 北京工业大学信息学部博士研究生. 主要研究方向为面向决策树的深/宽度学习结构设计与优化, 城市固废焚烧过程二噁英排放预测.

E-mail: xiaheng@emails.bjut.edu.cn

**(XIA Heng** Ph.D. candidate at the Faculty of Information Technology,

Beijing University of Technology. His research interest covers structure design and optimization of deep/broad learning based on decision trees and dioxin emission prediction of municipal solid waste incineration process.)



**王 鼎** 北京工业大学信息学部教授. 2009 年获得东北大学硕士学位, 2012 年获得中国科学院自动化研究所博士学位. 主要研究方向为强化学习, 智能控制.

E-mail: dingwang@bjut.edu.cn

**(WANG Ding** Professor at the

Faculty of Information Technology, Beijing University of Technology. He received his master degree from Northeastern University in 2009 and received his Ph.D. degree from Institute of Automation, Chinese Academy of Sciences in 2012. His research interest covers reinforcement learning and intelligent control.)



**乔俊飞** 北京工业大学信息学部教授. 主要研究方向为污水处理过程智能控制, 神经网络结构设计与优化.

E-mail: junfeiq@bjut.edu.cn

**(QIAO Jun-Fei** Professor at the Faculty of Information Technology, Beijing University of Technology.

His research interest covers intelligent control of waste water treatment process, structure design and optimization of neural networks.)