

综述与专论

DOI: 10.11949/0438-1157.

过程分析技术中的图像数据增强方法及其应用

范云龙, 周光正, 张京康, 桑雅妮, 袁玥, 王学重

(北京石油化工学院新材料与化工学院, 恩泽生物质精细化工北京市重点实验室, 北京 102617)

摘要: 随着原位显微成像与工业视觉系统的不断发展, 图像分析技术已成为生产过程智能化调控的重要手段。深度学习在图像分析任务中得到广泛应用, 但受高质量数据采集成本高、人工标注工作量大等现实问题限制。数据增强技术通过计算机算法扩大数据集规模、丰富样本特征多样性, 从而有效提高模型性能。本文对过程分析技术 (process analytical technology, PAT) 领域常用的各类图像数据增强方法做了系统介绍, 包含传统的非深度学习方法和基于深度学习的先进方法, 同时重点梳理了它们在结晶、气液多相流、细胞培养等典型应用场景中的研究进展, 强调了对物理一致性的相关保障。文章还展望了数据增强在提高在线检测智能化水平、加快工业应用方面的发展方向。

关键词: 成像; 结晶; 多相流; 过程分析技术; 深度学习; 数据增强

中图分类号: TQ 014

文献标志码: A

文章编号: 0438-1157 (XXXX) XX-0001-19

Image data augmentation methods and their applications in process analytical technology

FAN Yunlong, ZHOU Guangcheng, ZHANG Jingkan, SANG Yani, YUAN Yue, WANG Xuezhong

(Beijing Key Laboratory of Enze Biomass Fine Chemicals, College of New Materials and Chemical Engineering, Beijing Institute of Petrochemical Technology, Beijing 102617, China)

Abstract: With the advance of in situ microscopic imaging and industrial vision systems, image analysis technology has become quite important for the intelligent regulation and control of production processes. Deep learning is increasingly applied to various image analysis tasks, but its implementation is often hindered by practical issues, such as the high cost of acquiring high-quality data and the large manual labor of data annotation. Data augmentation techniques expand the dataset size and enrich the diversity of sample features via computer algorithms, and thus effectively enhance model performance. This paper gives a systematic review of different augmentation strategies for image data in the field of process analytical technology (PAT), including traditional methods and advanced deep learning methods. Moreover, recent progresses in the applications of those approaches in some typical scenarios are introduced in details, such as crystallization, gas-liquid multiphase flow, and cell culture, while the strategies for maintaining physical consistency are particularly stressed. Finally, some future directions of data augmentation are prospected for enhancing the intelligence level of online monitoring and accelerating its industrial applications.

收稿日期: 2026-02-06 修回日期: 2026-04-30

通信作者: 周光正(1981—), 男, 博士, 副研究员, zhouguangzheng@bipt.edu.cn; 王学重(1963—), 男, 博士, 教授, wangxuezhong@bipt.edu.cn

第一作者: 范云龙(2000—), 男, 硕士研究生, 2023520281@bipt.edu.cn

基金项目: 北京市自然科学基金(123033); 国家自然科学基金重点项目(61633006); 北京市教育委员会项目(22019821001)

引用本文: 范云龙, 周光正, 张京康, 桑雅妮, 袁玥, 王学重. 过程分析技术中的图像数据增强方法及其应用[J]. 化工学报, XXXX, XX(XX): 1-19

Citation: FAN Yunlong, ZHOU Guangcheng, ZHANG Jingkan, SANG Yani, YUAN Yue, WANG Xuezhong. Image data augmentation methods and their applications in process analytical technology[J]. CIESC Journal, XXXX, XX(XX): 1-19

Keywords: imaging; crystallization; multiphase flow; process analytical technology; deep learning; data augmentation

文章类型:综述与专论 DOI: 10.11949/0438-1157.20260182

引 言

过程检测技术在化工、制药、食品以及生物制造等行业的生产流程中得到广泛应用,以提高产品质量并充分保障生产安全^[1,2]。对原材料特性、生产中间状态以及最终产品指标的实时监测是流程优化、质量把控和风险防控的基础。尤其在制药行业,关键质量指标的检测直接与药物疗效、患者安全密切相关,因此一直受到监管机构与行业的高度重视^[3,4]。然而,传统检测方式大多依靠离线手段,如实验室化学分析、采样后的显微观察等。在实际生产过程中,工艺状态可能在短时间内明显变化,过长的离线检测周期往往无法实时获得重要的动态信息,将造成生产过程的调控滞后。频繁采样不仅增加人力成本,还可能给无菌和高风险生产环境带来潜在干扰。随着工业生产向连续化、智能化、高质量的方向发展,传统检测方法的相关不足变得更加突出。

过程分析技术(process analytical technology, PAT)是基于质量监控与管理理念的技术,它强调实时在线采集生产过程中的各类信息^[5]。相较于传统的离线检测方法,PAT致力于“在过程中保证质量”,即将质量控制环节提前到生产过程中。PAT最早是在制药行业中被提出^[6],快速推广到生物技术与生物制药、食品饮料、精细化工与石化、新能源与材料等不同领域^[7,8]。PAT利用近红外与中远红外光谱、拉曼光谱、紫外/可见光谱、原位成像、超声等技术对生产中的关键工艺参数以及产品质量特性做持续监测,以保障工艺建模、优化控制以及质量源于设计(quality by design, QbD)的落地实施^[9-11]。原位成像技术依靠光学原理在工业环境里直接获取观测目标,直观地展示对象特征^[12]。为实时预测和分析生产过程的关键指标,高精度的图像分析模型是关键因素^[13,14]。原位采集的图像通常比较复杂,普遍存在图像质量低、目标模糊、对比度不足等问题,而传统的图像处理方法虽操作简单,但分析准确度普遍不高。人工智能在近些年快速发展,深度学习(deep learning, DL)技术可依托特殊的深度神经网络

结构自动提取图像的高层抽象特征,大幅提高了对复杂图像的识别能力^[15,16]。深度学习模型一般非常依赖大规模、多样化且标注规范的训练数据集。然而,因为数据获取周期长、标注成本高、人工标注的主观偏差等原因,高质量数据不足、数据分布不均等情况在实际应用中很常见,它是先进检测算法在PAT场景中发挥作用的主要问题之一^[17]。

图像数据增强技术(data augmentation, DA)在真实过程数据的基础上,通过各种计算机手段进一步扩增数据规模与覆盖面,服务于模型开发、算法验证及过程优化等^[18,19]。除生成新图像外,一些增强技术可直接基于原图像的标注信息而同时提供新图像的精确标注,从而大量节约数据获取与标注成本。因此,将数据增强技术引入深度学习建模过程是提升过程分析技术中图像分析性能的一种可行途径。本质上,PAT应用中数据增强技术的核心价值并不限于完全逼真的视觉效果,而是通过扩展数据分布与增加有效训练样本以提高下游计算机视觉任务(如分割、分类、参数测量等)模型的精度与泛化能力。本文后续内容将在介绍图像分析结果评价指标的基础上,重点阐述各种类型的非深度学习传统增强方法与基于深度学习的生成方法,包括其基本原理以及在过程分析技术中的代表性应用成果。最后,展望图像数据增强技术为推动不同领域PAT应用的精准化与智能化的未来研究方向。

1 图像分析结果评价

图像数据增强方法虽为缓解数据稀缺与分布不均问题提供了有效途径,但其应用效果的科学评价仍是当前研究中的主要问题之一。数据增强技术应用于PAT的最终目标是提升过程参数测量的精度。本文聚焦于图像数据的增强方法,而非光谱等其他数据类型,因此不涉及光谱模型相关的物理参数评价,如组分浓度等。图像增强技术的效果评估不仅是在视觉层面分析新图像的质量(如逼真度),还需在实际物理层面评价其对预测结果准确性与稳定性的影响^[20],特别是尺寸、形状、数密度这类参数,如晶体粒径与形状分布、气泡尺寸与数密度分布、多相流型特征参数等。因此,需要评估数据增强模型在预测过程参数时是否存在系统偏差,

同时也要确认测量误差能否满足工业在线监测的精度要求。

物理一致性评估一般采用多个指标全面判断图像分析结果是否可靠,比如目标特征、统计属性这些指标就可以反映真实的物理特性,进而给模型优化和方法对比给出系统的定量参考。表1列出了图像分析物理精度评价的一些代表性指标。其中,一类指标直接对生成图像和真实图像在物理属性或统计分布上的一致性做量化,常用指标有尺寸平均绝对误差^[21]、面积相对比值^[22]、尺寸/形状分布的Kullback-Leibler (KL) 散度^[23]、形貌与结构的二维Minkowski 泛函^[24]等。另一类指标则主要根据预测结果和真实信息之间的匹配程度以评价模型在对

应计算机视觉任务上的性能表现。典型指标包括平均精确率(average precision, AP)及其对各类别的均值(mAP)、平均召回率(average recall, AR)及其对各类别的均值(mAR)、F1 分数、准确率(accuracy)、受试者工作特征曲线下面积(area under curve, AUC)等。这类指标依靠预设好阈值的特定参数从不同维度反映模型的识别精确性和检测完整度,如分割或目标检测任务中用于描述空间重叠度的交并比(intersection over union, IoU)、分类任务的预测概率都属于这类。此外,分割任务在像素级识别层面,常用的指标还有像素精确率(pixel precision)、像素召回率(pixel recall)等,这类指标直接依靠图像像素点统计模型的相关性能。

表1 图像分析的物理精度评价指标

Table 1 Physical accuracy evaluation metrics for image analysis

指标	定义	功能
尺寸平均绝对误差	所有目标预测尺寸相对其真实值的绝对误差的平均值,值小优	衡量目标几何尺寸的预测精度
面积相对比值	所有目标预测面积总和相对其真实值的比值,值接近1.0 优	基于像素点评估两者尺寸与形状的相似性
尺寸/形状分布的KL散度	利用信息熵计算的尺寸/形状分布差异,值小优	衡量生成图像在目标尺寸/形态维度的全局统计分布匹配程度
形貌与结构的二Minkowski 泛函	通过面积、周长以及二维 Euler 特征数对几何形态、拓扑结构特征的系统化表征,绝对值小优	完整描述二维空间中研究对象的几何属性与拓扑特性
平均精确率及其均值	不同阈值下特定类别被正确识别的比例,均值是各类别结果的算术平均值,值大优	分别评估模型对单类与多类目标的查准性能,反映目标识别的准确性
平均召回率及其均值	不同阈值下特定类别被正确检出的比例,均值为不同类别结果的算术平均值,值大优	分别评估模型对单类与多类目标的查全能力,体现检测结果完整程度
F1 分数	精确率和召回率的调和平均值,值大优	结合精确率与召回率对模型的整体性能评估
准确率	预测正确的样本数在总样本数的占比,值大优	综合体现模型对各类别的预测效果
AUC	受试者工作特征曲线下方的面积,值大优	反映模型在二分类任务中的整体判别能力
像素精确率	某一类别中被正确识别的像素占该类别所有预测像素的比例,值大优	评估像素级分割任务的识别性能

2 非深度学习数据增强方法

基于图像的PAT技术通过视觉传感模块采集数据,常用的采集设备主要包括在线显微成像装置、工业摄像设备等。图像分析技术已被应用于多种PAT场景中,比如颗粒粒度与形貌特征分析、气泡及液滴动态行为监测、细胞生长阶段判定以及复杂体系混合与分散效果评估等^[25, 26]。与传统的标量过程参数不同,图像信息具有多维度结构和复杂时空属性,能够提供生产流程中状态变化的全面信息。为克服深度学习模型对大规模标注数据的依

赖,传统非深度学习数据增强技术是常用手段,它们基本包括三类,即几何变换、图像扰动以及其他高级方法(表2)。这些方法无需额外实验与标注成本即可提升样本空间的覆盖度,缓解模型的过拟合问题,为精准的过程监控与质量预测提供数据保障。下面分别介绍这三类方法的特征与一些具体应用。

2.1 几何变换与图像扰动

几何变换主要利用旋转、翻转、缩放及平移等方式调整原始图像的空间相位特征,模拟相机视角、距离、物体方向等的动态变化,从而反映图像的

表2 非深度学习数据增强方法汇总

Table 2 Summary of non-deep learning data augmentation methods

类型	方法	具体操作	功能
几何变换	旋转	旋转一定角度	通过改变图像中像素的空间坐标,模拟拍摄角度、距离及物体朝向的变化,提升空间不变性
	翻转	基于水平或竖直轴翻转	
	缩放	按比例缩小或放大	
	平移	沿水平、垂直方向移动	
	裁剪	裁剪出图像的子区域	
	透视变换 ^[27]	使用高维变换矩阵模拟视角变化等	
图像扰动	添加噪声	添加高斯、椒盐等类型噪声	直接修改图像的像素数值,模拟真实世界中拍摄环境、成像质量、光照、噪音等影响,提升光学鲁棒性
	灰度化	三通道颜色信息转换为单通道灰度值	
	亮度/对比度调整	调亮、调暗或改变明暗对比	
	色彩抖动	轻微随机调整颜色值	
	模糊/锐化	使用卷积核对图像进行滤波	
	随机擦除 ^[28]	随机区域的像素值替换为随机噪声	
其他方法	Mixup ^[29] /CutMix ^[30]	混合两个样本及其对应标签	通过一些高级操作生成特殊样本,进一步提升复杂场景鲁棒性,如极端条件等
	Mosaic ^[31]	随机组合四张训练图像	
	随机风格化 ^[32]	随机添加多样化的图像风格特征	
	FGSM ^[33] /PGD攻击 ^[34]	以对抗攻击方式添加扰动	
	Copy-Paste ^[35]	裁剪目标粘贴至新背景	
	物理渲染技术 ^[36]	通过三维结构和光线追踪模拟物体外观	

空间不变性。图像扰动则是在像素层面加入高斯噪声、改变亮度与对比度参数、实施色彩抖动(color jittering)操作等。它直接修改图像的像素值,以此模拟现实场景下拍摄环境差异、成像质量波动、光照条件变化以及噪音干扰等多重因素的影响。这两类数据增强方法能够在不改变目标主体物理语义与全局结构的基础上,重构图像的空间排布规律与像素强度分布特征,扩大训练数据集的样本覆盖度,有效提高深度学习模型应对成像条件变化的泛化能力与抗干扰鲁棒性。下文将介绍几何变换与图像扰动在PAT领域的应用,重点探讨其在实例分割、语义分割及其他图像分析任务中的研究进展。

2.1.1 实例分割 实例分割(instance segmentation)能够同时完成图像中多个目标的类别判断及其像素级的空间坐标定位。从检测分割联合实现的角度而言,算法架构可以分为单阶段和双阶段两大类。双阶段算法遵循“先检测后分割”的运行逻辑,对于复杂图像分割任务可能达到更优的精度。Mask R-CNN^[37]是一种典型的双阶段算法,它在过程

分析技术的各种场景中应用广泛。Liu等^[38]搭建了一套依托计算机视觉技术的高通量筛选体系,可以高效判断不同添加剂种类与浓度对琥珀酸晶体的尺寸参数、形状特征以及聚集规律的影响。为扩充琥珀酸晶体图像样本库,研究使用了多种数据增强方法,具体包括调整亮度参数、垂直翻转、水平翻转、垂直与水平组合翻转(图1)以及原始图像重采样处理。最终数据集的样本规模从60张增加到395张,经过人工目视检查,Mask R-CNN模型的识别准确率超过90%,训练阶段的平均精确率最高达到73.9%。Liu等^[39]建立了包含晶体、液滴、气泡与团聚体等四种目标类型的开放化学微粒图像数据库(open-CMD),采用与前述基本相同的数据增强策略将数据规模扩充至2,500张(超过50,000个标记实体)。研究表明,数据增强策略能显著提高Mask R-CNN模型的准确性,平均精确率提升达3.8%;但由于数据集增大,其达到最优性能所需训练时间较无数据增强情况变长。在高斯模糊、均值滤波模糊、亮度调节、椒盐噪声、高斯噪声、图像缩放等多

种数据增强方法的基础上, Kim等^[40]搭建了一套可端到端运行的蛋白质结晶动力学评估深度学习框架。研究以溶菌酶晶体的特征尺寸参数的图像分析结果、运动模型预测值的均方误差作为优化目标进行参数调优, 利用贝叶斯优化方法求解得到晶体生长动力学相关参数。模型输出的晶体尺寸随时

间的演化规律和实验观测结果整体保持一致。该研究有效缓解了样本量不足、图像噪声这两类问题对晶体尺寸检测和动力学参数估算的干扰, 在提高蛋白质工业结晶生产效率、优化产品品质方面有较高应用价值。

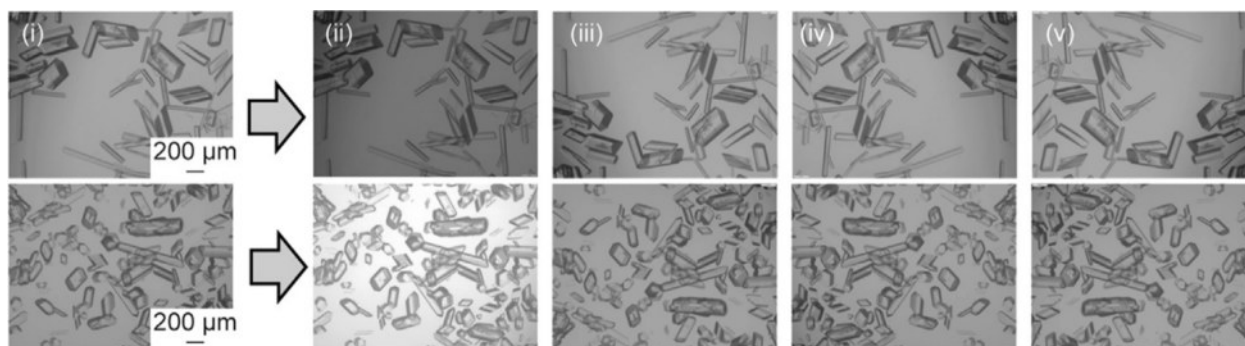


图1 琥珀酸晶体的数据增强:(i) 原始图像;(ii) 亮度调整;(iii) 垂直翻转;

Fig.1 Data augmentation of succinic acid crystals: (i) raw image; (ii) brightness adjustment;

注:(iv) 水平翻转;(v) 垂直与水平组合翻转^[38]

(iii) vertical flip; (iv) horizontal flip; (v) simultaneous vertical and horizontal flip^[38]

对于液液分散系统, Zhang等^[21]基于Mask R-CNN提出Segfit方法, 实现了微流控通道中微液滴尺寸分布的精确与快速测量。原始训练集为不同实验条件下的20张显微图像(1,957个液滴), 普遍存在液滴重叠、边界模糊及卫星液滴等复杂情况。为缓解小样本条件下模型泛化能力不足的问题, 该研究引入翻转、仿射变换、模糊变换等数据增强方法, 而且采用了测试时增强技术(test-time augmentation, TTA)输出最终预测结果, 即综合输入图像多个变换版本的识别结果。经图像增强优化后, 小液滴检测精度显著改善, 液滴直径平均绝对误差由1.01 μm 降至0.75 μm (相当于平均液滴直径的2.14%), mAP提升3.8%, mAR提升2.5%。

动物细胞培养是生物制药的一项重要基础技术, 在线实时检测对全面优化培养过程尤为关键。对于常见的中国仓鼠卵巢细胞培养, 为解决其原位图像中存在的细胞与背景之间区分度不高、边界不清晰、聚集行为很严重等问题对图像分析造成的挑战, Wang等^[25]利用了水平翻转、180度旋转以及垂直翻转的数据增强方法。训练集的图像数量从46张增加到184张, 对应的细胞总数也从45,760个扩充至183,040个。同时测试了多尺度边缘检测方法^[41]、一种集成边缘检测、强度阈值、高级分水岭的方法^[42]等传统图像处理技术, 结果反映深度学习可以更加准确地拆分成团细胞群体中的所有单个细胞, 其性

能指标值也更加突出, 精确率、召回率和F1分数分别为0.901、0.827和0.862。

相对液液系统, 气液分散系统中的气泡形状变化更为丰富, 涉及复杂的气泡破碎、聚并等行为。Qu等^[22]将Mask R-CNN的主干网络由ResNet-101替换为VMamba^[43], 并结合数据增强技术实现了高重叠密集气泡的精确分割。VMamba将选择性状态空间模型(state space models, SSMs)^[44]与视觉任务特性相结合, 引入适配二维空间结构的建模方式, 使状态空间模型能够更好地捕获图像中的全局上下文与空间依赖关系, 从而为高效视觉表示学习提供了一条兼具性能与可扩展性的技术路径。其数据增强方法包括随机水平翻转、随机裁剪以及多尺度缩放。在空隙率为0.10–0.20的条件下, 仅基于25张图像增强后的模型在IoU为0.5–0.95时的平均精确率为0.476–0.738; 对100张图像执行增强, 平均精确率进一步提升至0.525–0.861。然而, 本研究的测试集仅包含的30张真实图像可能无法覆盖工业两相流的各种流动特征, 导致影响统计结果的鲁棒性。

与二阶段算法不同, 单阶段算法将检测与分割任务整合于统一的前向网络中, 通常在推理效率方面具有优势, 更适用于对实时性要求较高的工业在线监测场景。Manee等^[45]以改进的RetinaNet算法^[46]为基础构建了单阶段分割模型, 用于NaCl–乙醇–水

反溶剂结晶体系的在线监测,包括对结晶过程中晶体颗粒的在线识别、分割与尺寸统计。在多种不同温度与反溶剂流量下采集显微图像数据,从中随机抽取100张用于训练模型,进一步采用随机裁剪、旋转、平移、颜色通道操作等数据增强方法。模型分割精度在大多数工况下与Mask R-CNN相当,但对于晶体较小且存在团聚现象的结晶初始阶段,其性能略低于两阶段的Mask R-CNN。针对高密度结晶浆料中存在大量晶体间团聚的图像分析挑战,Li等^[47]以牛磺酸杆状晶体为研究对象,通过水平/垂直翻转以及90°旋转等数据增强策略提升模型泛化能力。研究发现,较为先进的单阶段YOLOv8算法的mAP值与平均交并比皆稍优于Mask R-CNN算法。

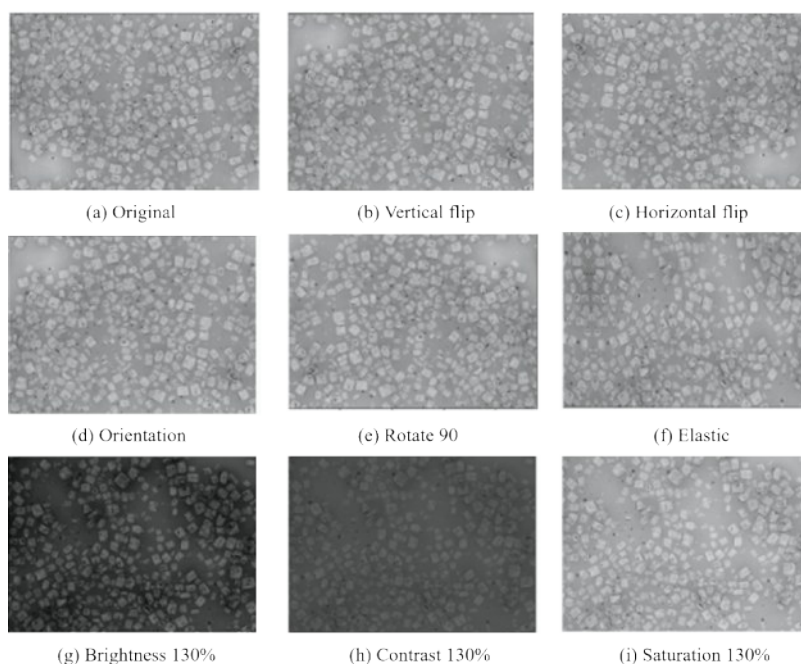
2.1.2 语义分割 语义分割(semantic segmentation)旨在为图像中的每个像素分配类别标签,从而实现对不同类别物体分布区域的精准划分,但并不区分同类别的不同个体。U-Net是一种经典的语义分割算法,其在实际应用过程中常结合数据增强技术以提升模型的泛化能力^[48]。Osiecka-Drewniak等^[49]通过显微成像考察液晶化合物9BA4冷却过程中的相变行为,不同相具有各自独特的图像纹理特征。由于获取的图像数量有限,通过亮度调整、翻转、缩放、旋转扩充数据规模。基于U-Net模型的图像识别结果,通过二值化处理定量表征七种降温速率下冷却过程的结晶程度。依靠非侵入式在线成像技术,Huo等^[50]基于改进的U-Net,研究冷却结晶过程中针状的 β 型L-谷氨酸晶体(LGA)尺寸分布的变化规律。数据增强手段包括亮度与对比度调整、噪声处理、平移、旋转、翻转等,从而减轻光照问题的影响并丰富了样本多样性。经过在线分析得到的晶体长度、宽度分布直方图和离线电子显微镜测量结果高度吻合,借助概率密度函数完成了双向生长速率的评估计算。然而,相对于实例分割技术,语义分割并不能区分图像中的晶体个体与聚团,导致生长速度的相关统计结果可能存在误差。

为保证糖类蒸发结晶的产品质量且提升经济效益,有必要在结晶过程中不断检测其产物性质以确定合适的补料时机。针对木糖结晶过程原位图像中的固体分数很高、晶体尺寸分布范围广、存在大量聚团等分析挑战,Zhu等^[51]使用了将Swin Transformer^[52]作为特征提取骨干网络的UperNet语义分割框架。相对卷积神经网络感受野有限的局部性,Transformer^[53]的自注意力机制更能反映不同

元素间的全局依赖关系。采用水平翻转、180°旋转和垂直翻转的数据增强方法将训练集中的图像数量由69张扩增至276张,晶体数达到105,852个。进一步结合迁移学习技术,模型精确率、召回率与F1分数分别为0.903、0.887和0.894,明显优于基于卷积神经网络的Mask R-CNN以及一种较先进的传统图像处理方法。因此,数据增强与Swin Transformer网络的结合对糖类蒸发结晶的在线检测具有显著优势,服务于其生产工艺的自动化与智能化发展。

2.1.3 其他分析任务 计算机视觉的目标检测(object detection)任务旨在准确预测图像中特定目标的类别及其具体位置(边界框坐标),而无需给出每个像素点的类别属性。Su等^[54]研发了一套集成微流控技术的高通量结晶条件筛选系统,实现了在微反应器芯片上生成尺寸均一的水凝胶液滴。晶体检测采用二阶段的Faster R-CNN算法;原始数据集为498张吡哆美辛晶体图像,包括片状、棒状、线状以及胶状相四种典型形貌。其数据增强采用基础的旋转、翻转、缩放等几何变换。相应模型对胶状相与线状的AUC值为1.00与0.99,棒状与片状的亦分别达到0.90和0.89,验证了数据增强对提升不同晶体形貌辨识与区分能力的有效性。在牛磺酸冷却结晶的原位成像研究中,Wu等^[55]采用了定向目标检测算法S²A-Net同时检测晶体与晶体簇。利用步长为5°的旋转变换及亮度扰动等手段对29张原始图像进行扩充,显著提升了数据集的多样性。相应的检测模型在不同晶种载荷(低、中、高浓度)下均保持了较高的精确率和召回率,展现出良好的工业应用鲁棒性。

图像分类(image classification)任务旨在根据图像的全局特征对其分配预定义的类别标签,实现图像内容的整体理解与判别。Zhang等^[56]将融合多尺度感知与残差学习的InceptionResNetV2算法应用于蔗糖结晶过程的在线监控,通过原位图像分类实现了对结晶过程中不同阶段的自动辨识与区分(如溶液澄清、成核发生、晶体生长等)。若图2所示,该研究引入了八种数据增强手段以提高模型的泛化性能,涵盖了水平/垂直翻转、全向随机旋转(-180°至180°)、90°旋转、弹性形变以及针对亮度、对比度和饱和度的随机调节(范围均为70%–130%)。分类测试结果显示,模型的准确率、精确率及召回率分别达到0.901、0.906和0.900。

图2 蔗糖结晶图像的数据增强^[56]Fig.2 Data augmentation of cane sugar crystallization image^[56]

除图像分类外,基于全域图像的内容识别与理解还可实现对系统理化性质的定量表征,体现了“图像信息学”在跨学科研究中的深度应用。Marcato等^[57]提出了一种基于场发射扫描电子显微镜图像直接预测冻干产品传质系数的端到端深度学习框架。研究使用9种不同配方的冻干样品,其显微镜图像分辨率为2,048×1,536像素。为缓解数据稀缺问题,研究采用水平翻转、垂直翻转和180°旋转操作扩充将数据量,各增强图像均保留原配方样本的传质系数标签。结果表明,除一个表现异常的样品外,数据增强有效降低了其他8个配方传质系数的预测误差,平均降低程度达45.5%,证明了基于图像分析实现传质性质预测的可行性。

几何变换与图像扰动是基础且应用最广泛的传统数据增强方式。其优势在于维持目标主体结构的基础上,能够快速扩充样本的视角、尺度、色彩与纹理多样性,从而有效提升模型对拍摄相对位置、光照、传感器、噪音等的泛化能力。此外,它实现简单、计算高效,易于控制,且解释性强。然而,这类数据增强只是原样本的简单加工,对多样性、数据分布覆盖的提升程度较为有限;参数依赖人工经验选择,如过度的噪音或模糊可能掩盖关键的细微特征,不当的参数甚至可能对模型学习产生负面影响。

2.2 其他增强方法

除常见的几何变换与图像扰动外,非深度学习图像增强方法还包括一些更复杂的技术,它们通过生成一些特殊样本以进一步提升模型对复杂场景的鲁棒性,如样本混合、图像拼接、随机风格化、生成极端样本、区域重组、物理渲染等技术。Mixup^[29]和CutMix^[30]算法通过一定比例混合两个样本及其对应标签达到数据增强目的。Mixup的主要思路是在特征空间中对两个随机样本及其标签使用线性插值,构建出介于二者之间的新样本,以此生成大量虚拟样本。CutMix则是在图像空间中完成更直观的局部混合,从一张图像内裁剪出一块矩形区域,再把这块区域粘贴到另一张图像的对应位置,同时按照区域面积比例混合两张图像的标签。与上述仅混合两张样本的操作不同,Mosaic算法^[31]通过混合四张训练图像,拼接成单张完整图像,不仅丰富了目标的上下文语义,还缩小了原始目标的视觉占比,有效解决了模型对小目标感知不足的问题。

随机风格化策略(style randomization)^[32]是一种基于风格迁移的数据增强方法,利用训练过程随机更改图像的低层视觉属性,比如纹理、颜色、对比度等,同时保留图像原有的几何结构与语义信息不发生变化,以此扩大数据分布范围,提高模型对不同视觉环境的鲁棒性。FGSM^[33]与PGD^[34]分别通过单步生成扰动和多步迭代投影约束扰动生成极端对

抗样本。FGSM主要思路是利用模型损失函数相对图像的梯度信息,在梯度方向添加有限幅度的扰动。PGD用小步长完成多次迭代更新,每步扰动都被重新投影到预设的范数约束区域。然而,如上对抗攻击数据增强将显著增加模型训练时间,加入这些特殊样本虽能提升复杂场景下的决策安全性,但可能会导致模型在普通样本上的精度下降。对于一般的计算机视觉任务,常规的数据增强通常已经足够,反而能够学到更加稳健的全局结构特征。

Copy-Paste^[58]的核心操作是以裁剪方式提取图像中的真实目标,再嵌入到全新的背景场景。针对低固体浓度的氧化铝颗粒过滤过程的在线显微图像,Daus等^[35]利用带标记控制的分水岭方法提取约500个颗粒和100个气泡(由搅拌器带入或系统排气不完全导致),将其作为“前景对象库”。利用Copy-Paste技术合成600张训练图像,覆盖对应不同离散相浓度场景的三个目标面积分数(0.01、0.05、0.10)。每一张合成图像都自动附带像素级标注,大幅降低了图像标注成本以及人工操作的可能偏差。完全用合成数据训练获得的实例分割模型甚至对于高固体浓度也可精准检测颗粒,颗粒尺寸统计结果和激光衍射数据一致性很高。然而,该技术要求精确的分割掩码,质量不高的标注会给新样本带来噪声,反而干扰训练;另外,前景实例在新背景上的粘贴具有随机性,可能导致环境、光照、尺寸等方面的不匹配。

物理渲染技术主要利用三维材质体系与光线追踪运算。首先,基于几何类参数完成目标结构建模,包括尺寸规格、形态比例、曲面曲率、边缘轮廓特征等。再结合物理光学原理计算环境光场作用于物体时产生的各类光学效应,包括反射、折射、透射、吸收等,最终输出具有高度视觉真实感的合成图像。为优化蛋白质结晶过程的自动化检测模型,Bischof等^[36]合成了大规模高质量图像,训练集的322,558张图像包含16,801,555个晶体。对随机初始化训练且跳过首个最大池化层的S²A-Net算法的测试结果表明,小、中、大三种尺寸目标的平均精确率分别为41.87%、65.90%、77.80%,效果明显优于基于ImageNet预训练的微调模型。

相较于几何变换与图像扰动方法在像素空间的简单几何或代数调整的局限性,如上这些方法主要通过多类操作组合、多图像融合等高级方式生成一些特殊样本,从而进一步提升模型应对各类复杂

场景甚至极端场景的泛化性能。然而,此类方法基于原有真实样本重构而非生成全新样本,对数据丰富性的扩展存在局限性;合成样本的物理语义也较弱。因此,其效果与应用场景强相关。

3 生成对抗网络

由于实现相对简便且具有较强的可解释性,传统数据增强方法在PAT任务中被广泛应用。然而,此类方法对复杂非线性特征及多变工况的刻画能力相对受限,难以全面支撑一些复杂场景下深度学习模型对高保真、多样性训练数据的需求。相对而言,基于深度学习的数据增强技术依托其强大的自动特征学习能力,能够在真实样本数量有限的条件下生成结构复杂、逼真度较高的合成图像,从而为模型训练有效引入更丰富的数据结构特征与分布变化。生成对抗网络(generative adversarial networks, GANs)^[59, 60]作为该领域的代表性框架,通过博弈机制实现对原始数据分布的精准拟合,其基本架构如图3所示。生成器(generator, G)以随机噪声作为输入,生成伪造样本;判别器(discriminator, D)通过真实样本与生成样本的比较不断提升判别能力。生成器根据判别器的反馈持续优化生成结果,尽可能生成以假乱真的样本,最终在二者的博弈中而达到动态平衡。

3.1 GAN算法与质量评估指标

传统GAN^[61]的算法结构比较简单,主要功能是直接学习真实数据的整体分布特征,不需要依靠工艺相关信息或者标注信息。因为缺少对工况参数和质量状态的明确调控机制,GAN模型无法做到面向特定工艺条件或者异常工况生成定向样本,主要被用于条件生成模型的参照基准。早期GAN算法的生成效果欠佳,还存在训练不稳定、模式崩溃等弊端。训练不稳定问题的主要原因是生成器和判别器的动态博弈失去平衡:训练之初,判别器的收敛速度通常更快,会让生成器的学习效率大幅降低;如果生成器更新速度太快,又会让判别器无法给出可靠的分类判断结果。模式崩溃指生成器没有完整学到真实数据的全部特征,最终输出的样本存在多样性不足的问题。

为解决传统GAN存在的如上问题,改进方向主要从网络结构、损失函数、训练机制三个层面调整,在提高模型生成图像质量的同时保证训练过程稳

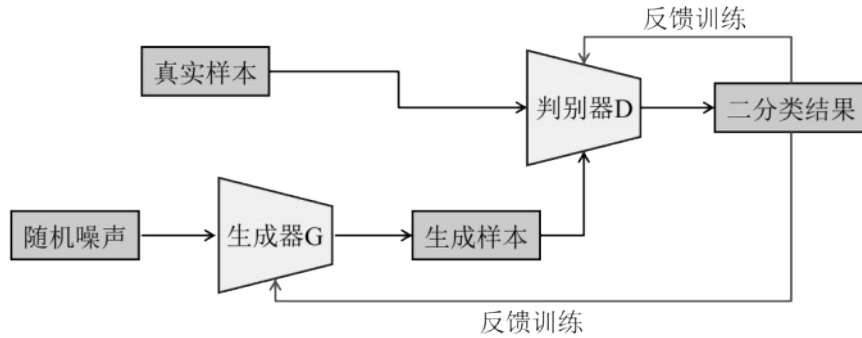


图3 生成对抗网络基本架构

Fig.3 Basic architecture of generative adversarial networks

定^[62]。DCGAN^[63]用深度卷积网络替换了原始的全连接层结构,明显提高了模型对局部纹理特征和空间结构信息的表征能力。U-Net GAN^[64]加入带跳跃连接的编解码结构,既能同时保留图像全局信息,又可整合多尺度上下文语义特征,提高生成结果的画面连贯性。基于Transformer结构的TransGAN^[65]在生成器中逐步提高特征分辨率,判别器则使用多尺度结构以同时抓取语义信息和底层纹理;StransGAN^[66]进一步缩小了与传统CNN判别器之间的性能差距。通过加入循环一致性约束,CycleGAN^[67]具有无配对样本场景下的跨域图像转换功能。MSG-GAN^[68]引入多尺度梯度监管机制,针对不同分辨率层级同步完成生成器参数优化,能够生成高分辨率的图像,也让训练过程收敛更稳定。为了强化生成过程的可控性,cGAN^[69]把类别标签或者工艺参数作为额外条件输入模型,让模型能在给定约束下生成符合目标分布的过程图像;InfoGAN^[70]是一种基于信息论扩展的GAN,能够在无监督的情况下实现学习可解耦表征的任务,并且通过最大化部分隐变量与观测值之间的互信息、优化互信息的下界来实现高效训练。

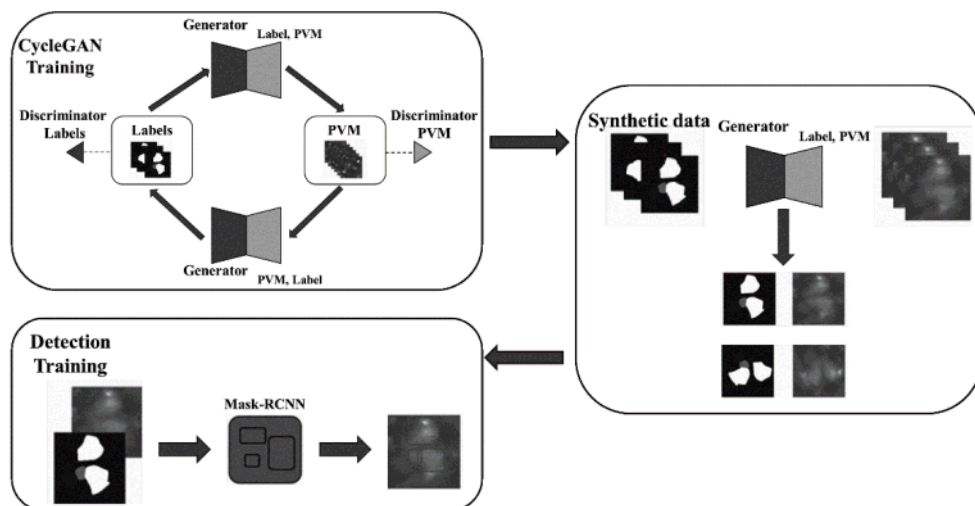
计算机视觉领域的一些通用的图像生成质量评估指标^[71,72]分别从不同角度评价合成图像的视觉保真度与分布一致性。峰值信噪比(peak signal-to-noise ratio, PSNR)与均方误差(mean squared error, MSE)都属于像素维度的评估参数:前者计算生成样本和实际图像的像素匹配程度,值大优;后者则统计两类图像对应像素点差值的平方均值以判定差异,值小优。结构相似性指数(structural similarity index measure, SSIM)从亮度、对比度、结构三个维度做综合判定,值大优。学习感知图像块相似度(learned perceptual image patch similarity, LPIPS)通过计算特征向量的加权距离判断相似度,值小优。

Fréchet Inception 距离(fréchet inception distance, FID)属于分布级评价指标,核心是计算生成图像和实际图像在特征空间内的分布偏差,能够同时覆盖生成结果真实感和丰富度两个评价维度,值小优。

3.2 代表性应用

随着对抗生成网络的不断改进,其在基于图像的过程分析技术中的应用也在增加中。基于颗粒视觉与测量(particle vision and measurement, PVM)的多种不同规格探头采集到的图像,Vagenknecht等^[73]进一步借助CycleGAN^[67]生成训练图像和标签对。图4给出了颗粒图像生成与检测模型训练的流程。首先,构建满足物理约束的图像标签集:根据工业结晶过程中颗粒粒径尺寸分布的先验信息限定生成标签的多边形尺寸范围,并通过设置多级像素强度的掩码以反映真实PVM图像中各目标的不同可见度情形(清晰程度)。基于400张合成图像训练CycleGAN模型以建立随机多边形标签与非配对PVM图像之间的映射函数,并通过CycleGAN的循环一致性损失约束保证生成图像在结构、纹理等方面与真实PVM图像的物理匹配性。然后,通过向“标签→PVM图像”生成器输入样本多边形标签,自动生成图像-标签对的数据集用于训练检测颗粒的Mask R-CNN模型。模型对中等尺寸颗粒(40~110微米)的mAP达到0.444,优于整体水平(0.275)。在低浓度工况下对两种不同晶体形貌化合物的中值粒径预测结果与激光衍射分析具有良好的一致性。尽管如此,模型在高固相浓度环境下的鲁棒性仍有待优化。

填料塔是吸收、精馏等气液传质作业的常用设备,液泛是它运行过程中代表性的流体力学故障,不仅会打乱塔器正常运行节奏、限制生产负荷上限,严重时还可能造成设备停运,甚至引发安全生产事故。为解决填料塔液泛实时在线监测的需求,

图4 基于CycleGAN的颗粒图像合成与识别模型训练流程^[73]Fig.4 CycleGAN-based workflow for particle image synthesis and recognition model training^[73]

Liu等^[74]搭建了融合数据增强模块与图像分类算法的全局上下文建模(global context modeling, GCM)框架,基于变分自编码器(variational autoencoder, VAE)与WGAN^[75]结合而设计了图像生成模型VA-WGAN。该研究采集实际生产中塔内发生液泛时的工况图像作为基础训练样本,在损失函数中同时加入VAE的KL散度约束项、WGAN的Wasserstein距离项以及梯度惩罚项,通过缩小生成分布与真实分布在隐空间的偏差,完成对真实数据分布的拟合建

模,还能从统计层面保证生成样本和实际液泛图像的特征分布一致。针对不同近液泛宽度,加入生成数据的模型即使在小样本训练场景依然可以保持较高的液泛状态分类精度,验证了生成图像能够较好还原真实工况的特征。然而,本研究的模型主要基于实验室规模的填料塔,其在高噪声水平工业环境中的泛化能力尚需全面评估。另外,液泛过程演化具有明显的时间序列特征,基于图像序列的建模有望进一步提升预测精度。

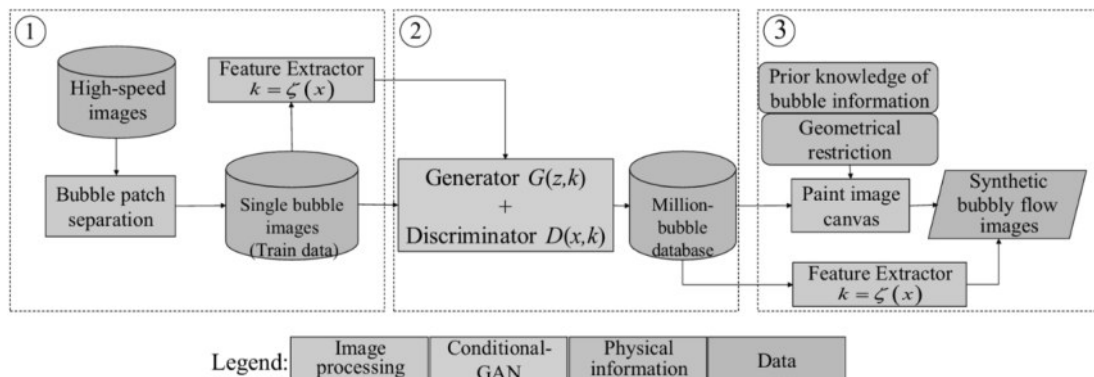


图5 BubGAN算法流程:(1)训练数据提取;(2)大规模气泡库生成;

Fig.5 Workflow of the BubGAN algorithm: (1) training data extraction; (2) large-scale bubble library generation;

注:(3)基于物理信息的图像合成^[76](3) physics-informed image synthesis^[76]

对于气液两相流过程的在线监测,传统生成对抗网络输出的图像不带有明确的像素层级标注内容,比如气泡边界信息,无法直接给图像分割类有监督学习任务提供有效支撑。受自身结构设计逻辑与算力负载上限限制,输出结果和工业级高速摄像设备采集的高清图像相比,空间分辨能力与局部

细节还原程度仍存在明显差距。针对真实气泡样本不足、人工标注成本过高的问题,Fu等^[76]搭建了气泡生成对抗网络(bubble generative adversarial network, BubGAN)。该算法根据“分而治之”思路生成高清晰度的带标注图像,图5给出了整体流程。第一步,使用分水岭、自适应阈值等常规图像处理

方法从高速相机采集的原始图像中批量提取独立气泡样本。第二步,基于上述真实单气泡数据集,以旋转角度、纵横比、边缘比、圆度四项几何参数的统计分布规律为约束条件,训练深度卷积生成对抗网络(DCGAN)^[63],最终建立覆盖百万级样本量的单气泡图像数据集。合成单气泡的形态受到实验数据统计规律的先验约束,从而避免不符合物理特征的结构。第三步,基于设定的气泡尺寸分布、数密

度分布等信息从气泡数据库采样,并将其融合成气泡流图像。图6给出了气泡实验图像与BubGAN合成图像的对比,两者的相似度很高。然而,BubGAN模型的训练数据只是从低到中等空隙率的气泡流图像中提取,而高空隙率气泡流中的气泡聚并和破碎更为频繁、形状更不规则。另外,BubGAN尚不能反映由气泡重叠导致的外观变化。

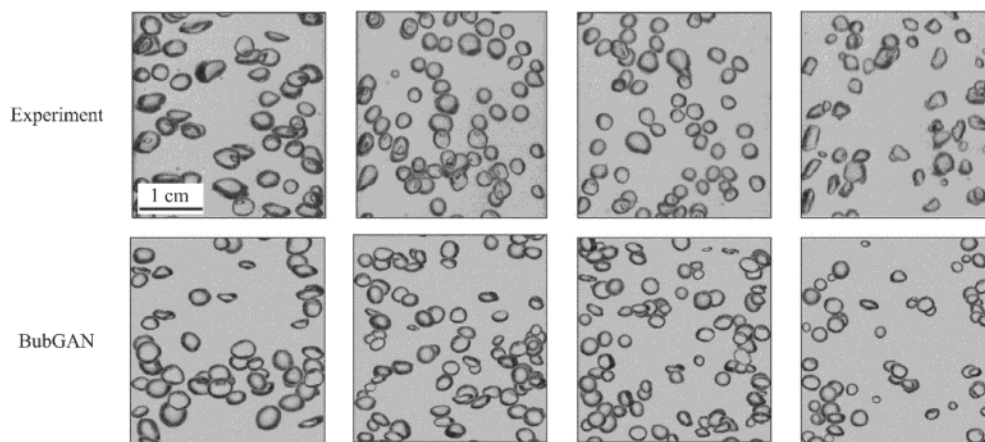


图6 不同流动条件下的气泡流实验图像与BubGAN合成图像^[76]

Fig.6 Experimental and BubGAN synthesized images of bubbly flow under different flow conditions^[76]

凭借在生成带标注气泡流图像方面的强大功能,BubGAN算法已被应用于多类气液两相流在线监测任务。Haas等^[77]基于Faster R-CNN架构提出了BubCNN算法,实现了气泡目标检测与几何特征参数(如长短轴、倾角等)的耦合测量。研究证实,引入BubGAN合成图像进行数据增强,能够在气泡高度聚集、尺寸分布范围较宽等复杂工况下显著提升模型的稳健性。进一步构建了迁移学习框架,在预训练模型的基础上辅以少量真实气泡流图像进行微调,便可实现模型向新场景的高效迁移,表明合成样本在降低人工标注成本方面具有重要价值。Kim等^[78]则将Mask R-CNN应用于气泡羽流、核电站棒束通道鼓泡流以及池沸腾气泡等多种场景的实例分割任务。该研究同样借助BubGAN扩充训练集,有效改善了模型在复杂两相流环境下的泛化性能。

气液液三相流的相界面演化过程具有很强的非线性特征,其流型结构复杂多变,对运行参数有高度依赖性。受实际作业场景的运行特性限制,部分流型能采集到的图像样本普遍偏少,给三相流流型的准确识别还有工况的动态监测带来技术难题。针对90°弯头水平管内的气-油-水三相流动过程的

十四种流型,Yaqub等^[79]利用WGAN-GP模型^[23]扩充241张真实图像到840张,有效改善了数据集类别不均衡的问题。与传统GAN架构相比,WGAN-GP选择Wasserstein距离来衡量真实样本分布和生成样本分布的差异,还加入梯度惩罚机制强制满足Lipschitz连续性约束,可以高精度还原出气液液三相分布的微观细节。不同流型的原始图像和对应生成结果的对比见图7,所有流型生成图像的像素准确率与精确率都达到100%,平均结构相似度为0.915。这些图像质量评价指标间接证明生成图像在流型空间结构、相态分布规律等核心物理特征上有着较高的还原度。使用预训练的基于Transformer架构的ViT模型在扩充后的数据集上完成迁移学习,四项分类精度指标都展现出优秀的流型判别效果。该研究提出的WGAN-GP与ViT结合的混合建模方案为小样本、多类别、分布不均衡条件下的三相流图像识别给出了实用的解决思路。

在细胞显微成像与分析领域,不同细胞的形态特征通常存在明显差异,边缘辨识度低,而且标注过程高度依赖操作人员的专业经验。以pix2pix算法为基础^[80],Baniukiewicz等^[81]提出了携带荧光标记蛋白的细胞图像的合成方法。它可以较好地保留

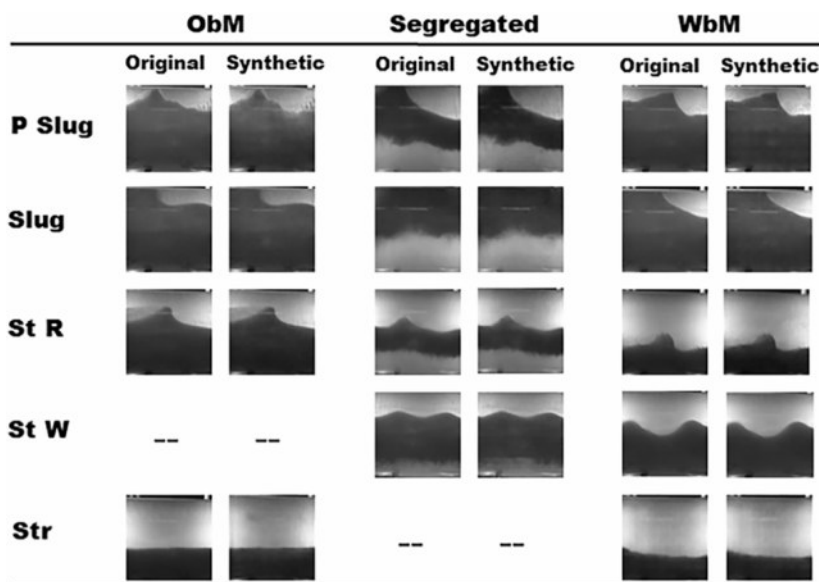
图7 管道气液液三相流不同流型的实验图像与WGAN-GP合成图像^[79]

Fig.7 Experimental and WGAN-GP synthesized images of different flow regimes

注:in pipe gas-liquid-liquid three-phase flow^[79]

细胞的尺寸特征(如细胞膜厚度与细胞大小基本无关),克服了剪切、缩放之类的传统仿射变换方法可能影响细胞相关几何统计的问题。定量表征结构相似性的 S_3 指标^[82]表明,合成图像的细节保真度要优于仿射变换基于双线性/三次插值函数的结果。该研究还设计了算力消耗较高的三维生成网络的一种替代方案,即利用二维生成对抗网络输出的高分辨率2D切片而合成伪3D细胞图像。Zargari等^[83]使用优化后的CycleGAN算法^[67]扩大显微细胞图像的数据规模,通过在损失函数中加入衍生自VGG架构的感知损失项^[84]等措施,让生成图像的质量(对应FID值明显下降)和样本多样性都得到有效增强。用增强后的数据集训练细胞分割模型,当IoU阈值设为0.5时,不同细胞类别与模型配置下的F1得分提高了0.02–0.10,高密度细胞图像的F1得分更是从0.77提高到0.86,证明所提出的数据增强方案确实可以有效增强模型在复杂应用场景下的泛化能力。需要注意,VGG损失增加了额外的算法复杂性与计算需求,对计算资源提出了更高要求。

工业场景中的成像条件往往存在很多限制,在线采集到的检测图像通常带有噪声干扰强、特征细节模糊等明显的质量问题,直接影响后续图像解析的精度。生成对抗网络也被应用于去噪复原、分辨率提高等基础视觉处理环节,更好地还原图像承载的物理信息。基于盲超分辨率生成对抗网络(BSRGAN)^[85]与注意力增强的超分辨率生成对抗网

络(A-ESRGAN)^[86],Neves等^[87]对气液两相流原始图像做超分辨率重建与去噪预处理。A-ESRGAN的整体表现最好,对应PSNR达到29.24 dB、LPIPS为0.062,性能明显超过传统的双三次插值方法。生成对抗网络可以有效拟合低质量图像和高质量样本之间的映射关系,经过增强处理的图像在边缘辨识度和细节保留程度上都有明显提高,有助于提高气泡检测和形态参数分析的结果可靠性。针对多孔介质多相流成像场景,Mathew等^[24]的研究主要解决多色X射线显微CT高速成像过程中噪声带来的结构信息丢失问题。利用CycleGAN^[67]获取噪声分布和真实结构特征的差异规律,精准还原多相流的孔隙结构、相分布特征。去噪处理后,图像SSIM指标从原始数据的0.377提高到0.663;针对非湿润相的二维Minkowski泛函(包含面积、周长、二维欧拉特征数三个参数),该算法的处理效果明显好于非局部均值滤波与自适应维纳滤波方案。尽管三种方法相应的欧拉特征数误差已经降到0,但经过GAN优化之后,另外两种方法的周长误差从原来的-9%、-10%缩小到-4%,面积误差也从-13%、-11%减少到-6%。本次研究只选用本廷海默砂岩作为试验样本,还没有验证该模型在其他岩石样本上的泛化能力。

工业表面缺陷检测领域经常存在正负样本分布极度不均的问题,数据增强技术被用于生成还原度较高的缺陷仿真样本。针对钢板表面缺陷图像

分类的场景,Jain等^[88]对比了平移、缩放、旋转、翻转等传统数据增强方法和DCGAN生成新样本对分类模型效果的影响。深度学习方法相比传统方法在模型准确率、召回率、特异性方面分别提高5.5%、5.05%与1.1%。Mohammed等^[89]设计了基于条件图像到图像翻译的cGAN架构,补充织物缺陷数据。该方法的约束条件不是单一的缺陷类别标签,而是由缺陷掩码和无缺陷织物图像组合得到的双输入结构。缺陷掩码以分割图的形式准确定义了需要生成缺陷的大小、形态、位置和类别,无缺陷织物图像则提供原始的纹理基底;生成器采用U-Net结构把两类输入沿通道维度拼接后作为条件信号输入,基于掩码引导在指定区域生成符合要求的缺陷。依靠这种空间对齐的结构化条件生成机制,训练集覆盖的缺陷类型和纹理丰富度都得到了有效拓展,用合成数据训练出来的模型在各项性能指标上都有明显提高。该研究仅通过掩码传递缺陷信息,如果进一步完善模型的条件输入形式,比如引入多维度参数来表示缺陷属性,就能对缺陷生成过程做更精准的控制,进而搭建精度更高、多样性更强的数据集为织物缺陷检测与分类模型的训练提供支持。

生成对抗网络凭借独特的双向对抗训练机制学习真实数据的分布特征,生成异于原始数据但真实感很强的全新样本,克服了传统数据增强仅能微调原样本的不足。然而,GAN的训练不稳定、模式崩溃等问题并未彻底解决,部分生成图像可能存在细节失真、训练难以收敛等情况。另外,GAN模型

缺乏可解释性,生成结果的可控性较弱。将GAN生成的图像用于过程分析技术,关键是保证数据符合物理一致性,因此往往有必要将一些物理规律约束融入相应场景的生成算法。建立基于物理约束的GAN算法通常具有较大的难度,这方面的工作仍在发展中。

4 扩散模型等其他类型

除生成对抗网络外,近年来还出现了多种其他类型的深度学习图像增强策略。其中一类方法基于精心设计的图像掩码,通过特定的深度神经网络结构生成具有较高逼真度的新图像^[23];另一类为扩散模型(diffusion models, DMs)^[90],作为生成式人工智能领域的前沿范式,在PAT的复杂工况下高质量图像合成方面展现出巨大的应用潜力。扩散模型的基本思想是通过一个正向扩散过程逐步将数据“破坏”成噪声,再通过一个反向扩散过程恢复原始数据,它具有卓越的复杂概率分布拟合能力。近些年,扩散模型在生成图像质量、速度、计算资源需求等方面不断优化^[91]。其代表性算法包括去噪扩散概率模型(denoising diffusion probabilistic models, DDPMs)^[92]、去噪扩散隐式模型(denoising diffusion implicit models, DDIMs)^[93]、潜空间扩散模型(latent diffusion models, LDMs)^[94]等。

Li等^[23]以L-丙氨酸结晶过程的原位显微图像为对象,研究图像分割相应的数据增强方法。传统

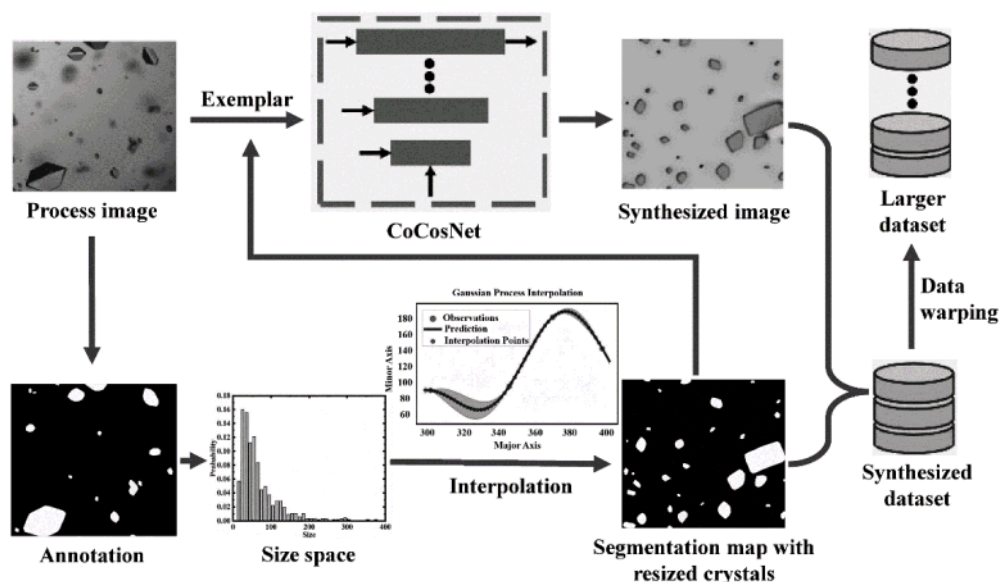


图8 基于CoCosNet与插值法的L-丙氨酸晶体图像合成流程^[23]

Fig.8 Flowchart of L-alanine crystal image synthesis based on CoCosNet and interpolation^[23]

数据增强方法本质上只是在原始数据集中体现更多信息而不会生成具有新潜在信息的样本,从而导致增强图像中晶体尺寸丰富性提升有限。因此,该研究提出采用CoCosNet算法^[95]和插值法生成虚拟图像,其具体流程如图8所示。该算法由两个子网络组成:跨域对应网络将源自不同领域的输入转换到一个中间特征域以建立可靠的密集对应关系;翻译网络通过整合经过扭曲但在语义上与掩码对齐的范例图像的细节特征,逐步合成输出图像。鉴于CoCosNet无法仅使用少量训练数据合成高保真度图像,该研究采用了迁移学习技术。同时,重点对晶体分布相对稀疏的较大、中等尺寸分别应用高斯过程插值与中值插值,以补充相应的晶体样本。仅基于25%真实图像并结合合成数据集的分割模型,其mAP(97.9%)可媲美基于全量人工标注真实数据集的模型(97.6%)。KL散度统计显示前者的晶体尺寸与形状分布更接近真实状态,晶体长轴偏差由后者的0.0336降至前者的0.0320,短轴偏差由0.0081降至0.0079,长宽比偏差由0.0784降至0.0766。

除上述特殊网络算法模型外,随着人工智能技术不断更新迭代,扩散模型已经逐渐成为复杂工业场景中数据增强的重要方法。针对材料显微图像的晶体分析需求,Saleh等^[96]构建了由概念驱动的扩散数据生成框架,主要思路是把晶体尺寸参数、几何形态、空间取向这些带有物理意义的结构特征作为隐性约束融入生成过程。最终生成的3000张合成图像既增加了数据多样性,又拥有可靠的物理合理性,也满足统计层面的一致性要求。该研究比较合成图像与真实透射电镜(TEM)图像在晶粒形貌、灰度分布两个维度的视觉匹配程度,证明了生成结果的高保真度。使用增强后数据集训练的分割模型在晶界提取任务中的平均准确率能达到97.23%,性能明显优于结合生成对抗网络的RCF算法的93.00%。经扩散模型生成的晶体图像能大幅拓宽训练数据的覆盖范围,可以在小样本场景下有效提高晶体分析模型的识别精度与泛化能力。

油井溢油监测场景关注溢油区域在时间维度上的演化过程。Lyu等^[97]提出的DiffusionLSTM算法框架集成了扩散模型与长短期记忆网络(long short-term memory, LSTM)^[98]的优势,实现了对复杂溢油扩散过程在空间与时间上的联合刻画。通过基于流体力学原理的OpenOil模型模拟溢油漂移扩散过程,并将模拟结果作为合成图像的物理约束,保障

生成符合图像统计特征与物理规律的溢油场景。扩散模型负责高质量地学习单帧图像的空间分布特征,而LSTM擅长建立图像序列之间的时间依赖关系模型,从而实现对未来时刻图像序列的生成与预测。DiffusionLSTM在生成图像序列的空间一致性、时间连续性以及对溢油形态演化趋势的刻画能力方面均优于传统时序生成方法。该框架为将扩散模型引入动态过程监测和PAT相关时序图像分析提供了有价值的范例。

在材料结构健康监测研究领域,Cao等^[99]针对纤维增强复合材料的结构状态划分和损伤辨识需求,提出了一种结合条件DDPM的高性能数据扩增方案。先把材料刚度衰减规律映射到Weibull分布,利用该分布的反函数推导得到结构健康评价指标,将它作为生成过程的约束边界,再搭配迁移学习训练模式完成不同损伤场景下样本的定向生成。试验阶段使用最大均值差异(maximum mean discrepancy, MMD)与FID指标量化对比生成样本和实测样本的特征分布契合度。与GAN、WGAN-GP、VAE以及原始DDPM相比,该方法得到了最低的MMD值,FID表现只比VAE稍低。在健康指标预测任务中,该方法的均方根误差为0.0750,比传统数据增强方法降低了25.5%,性能也明显优于上述几类生成模型。三个疲劳阶段的误差分布特征可以清晰呈现复合材料疲劳损伤的内在物理演化规律,进一步证实了该数据增强方法的实用价值。目前生成样本仅在统计层面满足合理性要求,后续仍需结合导波传播原理等物理模型以及领域专家经验,对生成样本的物理一致性做深入校验。Huang等^[100]的研究则把从碳纤维增强聚合物复合板采集到的Lamb波信号转换成二维灰度图,提出了一种改进型DDPM扩散模型,再将数据扩增结果用于DenseNet算法^[101]以实现缺陷类别辨识。结果表明,该DDPM方法得到的FID值为13.74,分类模型的准确率达到88.10%,性能明显优于DCGAN(对应指标为33.56、86.17%)与VAE(对应指标为41.65、71.25%)。该方法目前还无法识别尺寸很小的微损伤(可能小于1毫米),后续还需验证它在其他材料体系和其他损伤类型场景下的有效性。

对于高保真图像生成任务,扩散模型的性能通常能超越各类生成对抗网络模型,得到细节丰富、结构合理、真实感强的合成数据。相对GANs,由于其基于概率建模与变分下界优化,训练过程更稳

定,可以稳健地学习完整数据分布;能够兼顾生成质量与样本多样性;易于与其他模型融合(如VAE、GANs等),形成性能更强的混合生成框架。然而,它对硬件资源要求高、训练耗时长而导致成本高昂,且生成推理效率低而难以满足实时性要求。另

外,一些场景下的生成结果仍可能存在细节瑕疵,将其应用于过程分析技术也要解决物理一致性问题。

5 结语与展望

表3 三类图像增强方法在PAT应用中的比较

Table 3 Comparison of three image augmentation methods in PAT applications

方法	优势	劣势	发展方向
传统增强	提升空间不变性、光照/传感器鲁棒性,多样性有限,数据分布覆盖不足,参数依赖	根据场景与任务的方法组合自动筛选与参数自适应	实现简单,计算成本低,易于控制,解经验,过度的噪音或模糊可能掩盖关键的细能力,通过物理模型反映扰动,与深度学习数据增强
生成对抗网络	解释性强	微特征	方法融合
扩散模型	学习数据真实分布,生成高多样性、高质量的新样本	可能存在训练不稳定、模式崩溃等问题,对参数敏感,可能生成看似合理但物理错误的结构	提升训练稳定性,提高细节保真度,基于任务的条件化与可控生成,引入物理一致性约束
神经网络	训练稳定,多样性更丰富,高细节保真	算力成本高,对计算机硬件要求高,推理速度慢,可能生成看似合理但物理错误的结构	提升推理效率,基于任务的条件化与可控生成,引入物理一致性约束,少样本增强

大量的高质量标注数据是基于深度学习的人工智能技术的基础,数据增强对于缓解工业场景中高质量标注样本稀缺、样本分布不均衡、极端工况数据不足等问题具有重要作用^[27]。除了以几何变换、图像扰动等类型为代表的传统非深度学习方

法,生成式人工智能技术(生成对抗网络、扩散模型等)在近些年快速发展,其合成图像能够显著地增加数据的多样性与分布范围。就数据分布建模而言,不同方法在机制上存在本质差异。传统非深度学习方法的

不同数据生成能力较为有限,只能通过指定操作生成一些特定状态的样本。GAN属于隐式分布建模方法,通过学习从随机变量到数据空间的映射关系而合成样本;其核心在于能够从目标分布中采样,但并不显式给出数据分布的概率密度函数。扩散模型则属于显式分布建模方法,通过构建可计算的概率模型对样本的生成概率进行评估与优化。表3系统总结了这三类方法在过程分析技术应用中的各自优势、劣势及未来发展方向。

不同生成式算法在PAT应用中的模型训练数据量需求并不存在统一标准,而是取决于算法架构、任务复杂度、图像特征以及是否采用迁移学习等多重因素。所需要的图像数量本质上还与图像中的可用目标数密切相关。例如,在结晶、气泡流等场景中,单幅图像通常包含大量目标,因此在图像数量较有限的情况下仍可能满足模型训练需求。在引入物理先验信息或条件约束(如分割掩码、结构参数等)的条件生成框架下,一些模型甚至能够在几十至数百张图像规模上实现有效训练并保证

生成多样性。本文围绕过程分析技术中图像数据增强方法的研究进展与应用状况进行了较为全面的总结与评述。该领域尚待解决的挑战与未来发展方向主要包括以下方面。

(1)动态过程的序列生成。工业过程通常具有显著的时间演化特征,如晶体的成核生长、气泡的动力学行为等。相对单帧图像生成,进一步结合扩散模型、LSTM、Transformer^[102]等时序生成网络实现对动态图像序列的生成与预测,从而更全面地获取过程信息,相应建立的分析模型也将更加准确。另外,实时监测与预测模型相结合将为工业过程的闭环控制提供数据支撑。

(2)极少样本与特定工况样本生成^[103]。在工业生产场景中,异常状态的样本采集往往难度很高,样本分布不平衡的问题会大幅降低下游图像分析模型对高价值特殊工况的泛化性能。研发将工艺参数、实际运行条件等作为限制条件的生成类算法,在小样本量(甚至没有对应样本)的条件下以高度可控的方式合成稀缺工况数据。

(3)物理一致性与可解释性^[20]。在图像增强算法中需考虑相应的物理机理模型、领域统计经验或是实际生产工艺的约束规则,保证生成的仿真图像在结构特征、目标尺度分布规律以及具体任务相关的参数统计表现等方面和真实场景中样本特征高度吻合,从而为生产机理研究、工艺优化设计提供高可信数据支撑。

(4)图像与多维传感数据的深度融合^[104]。将图像和光谱、温度、压力等多类监测数据配合使用,

发展多模态过程检测模型的数据增强方法,更有效地揭示不同类型监测参数中隐藏的重要信息,从而

提高模型对系统复杂工况、异常事件的全局感知能力。

(5)工程化应用与可扩展性。图像数据增强技术还需考虑过程分析系统的扩展潜力,匹配不同生产场景、不同工况下的差异化使用需求。搭建高模块化、支持功能灵活调用的图像数据生成支撑平台,结合实际生产场景灵活调整增强方案,促进在线智能监测系统的大范围应用落地。

参考文献

- [1] Ge Z Q, Song Z H, Gao F R. Review of recent research on data-based process monitoring[J]. *Industrial & Engineering Chemistry Research*, 2013, **52**(10): 3543–3562.
- [2] Rao S L, Wang J T. A comprehensive fault detection and diagnosis method for chemical processes[J]. *Chemical Engineering Science*, 2024, **300**: 120565.
- [3] 赵绍磊, 王耀国, 张腾, 等. 制药结晶中的先进过程控制[J]. *化工学报*, 2020, **71**(2): 459–474.
Zhao S L, Wang Y G, Zhang T, et al. Advanced process control of pharmaceutical crystallization[J]. *CIESC Journal*, 2020, **71**(2): 459–474.
- [4] 张妍, 程景才, 杨超, 等. 药物多晶型的过程控制和工程技术进展[J]. *中国医药工业杂志*, 2018, **49**(5): 537–546.
Zhang Y, Cheng J C, Yang C, et al. Progress in the process control of pharmaceutical polymorphism and engineering[J]. *Chinese Journal of Pharmaceuticals*, 2018, **49**(5): 537–546.
- [5] Simon L L, Pataki H, Marosi G, et al. Assessment of recent process analytical technology (PAT) trends: a multiauthor review[J]. *Organic Process Research & Development*, 2015, **19**(1): 3–62.
- [6] Food and Drug Administration. Guidance for industry: PAT—a framework for innovative pharmaceutical development, manufacturing and quality assurance[R]. Maryland: FDA, 2004.
- [7] Rathore A S, Kapoor G. Application of process analytical technology for downstream purification of biotherapeutics[J]. *Journal of Chemical Technology & Biotechnology*, 2015, **90**(2): 228–236.
- [8] van den Berg F, Lyndgaard C B, Sørensen K M, et al. Process analytical technology in the food industry[J]. *Trends in Food Science & Technology*, 2013, **31**(1): 27–35.
- [9] Grangeia H B, Silva C, Simões S P, et al. Quality by design in pharmaceutical manufacturing: a systematic review of current status, challenges and future perspectives[J]. *European Journal of Pharmaceutics and Biopharmaceutics*, 2020, **147**: 19–37.
- [10] Fonteyne M, Vercruysse J, De Leersnyder F, et al. Process Analytical Technology for continuous manufacturing of solid-dosage forms[J]. *TrAC Trends in Analytical Chemistry*, 2015, **67**: 159–166.
- [11] Zhang F K, Du K, Guo L Y, et al. Progress, problems, and potential of technology for measuring solution concentration in crystallization processes[J]. *Measurement*, 2022, **187**: 110328.
- [12] 周光正, 钟子翰, 黄彦群, 等. 基于原位成像与图像分析技术的结晶过程智能监测[J]. *化工学报*, 2025, **76**(9): 4351–4368.
Zhou G Z, Zhong Z H, Huang Y Q, et al. Intelligent monitoring of crystallization processes based on in situ imaging and image analysis[J]. *CIESC Journal*, 2025, **76**(9): 4351–4368.
- [13] Zong S L, Zhou G Z, Li M, et al. Deep learning-based on-line image analysis for continuous industrial crystallization processes[J]. *Particuology*, 2023, **74**: 173–183.
- [14] Wang L Y, Zhu Y L, Gan C Y. Predictive control of particlesize distribution of crystallization process using deep learning based image analysis[J]. *AIChE Journal*, 2022, **68**(11): e17817.
- [15] Lu M J, Rao S L, Yue H, et al. Recent advances in the application of machine learning to crystal behavior and crystallization process control[J]. *Crystal Growth & Design*, 2024, **24**(12): 5374–5396.
- [16] Archana R, Eliahim Jeevaraj P S. Deep learning models for digital image processing: a review[J]. *Artificial Intelligence Review*, 2024, **57**(1): 11.
- [17] Peres R S, Jia X D, Lee J, et al. Industrial artificial intelligence in industry 4.0 – systematic review, challenges and outlook[J]. *IEEE Access*, 2020, **8**: 220121–220139.
- [18] Shorten C, Khoshgoftaar T M. A survey on image data augmentation for deep learning[J]. *Journal of Big Data*, 2019, **6**(1): 60.
- [19] Wang Z T, Wang P F, Liu K P, et al. A comprehensive survey on data augmentation[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2026, **38**(1): 47–66.
- [20] Banerjee C, Nguyen K, Fookes C, et al. Physics-informed computer vision: a review and perspectives[J]. *ACM Computing Surveys*, 2024, **57**(1): 17.
- [21] Zhang S Y, Liang X, Huang X Y, et al. Precise and fast microdroplet size distribution measurement using deep learning[J]. *Chemical Engineering Science*, 2022, **247**: 116926.
- [22] Qu Z M, Wang Q. A VMamba-based deep learning framework for bubble detection in dense flows[J]. *Physics of Fluids*, 2025, **37**(7): 073387.
- [23] Li M Y, Yao T, Liu J, et al. Deep learning-based in situ micrograph synthesis and augmentation for crystallization process image analysis[J]. *Mathematics*, 2024, **12**(22): 3448.
- [24] Mathew E S, Jackson S J, Wildenschild D, et al. Deep learning assisted denoising of fast polychromatic X-ray micro-CT imaging of multiphase flow in porous media[J]. *Computers & Geosciences*, 2025, **204**: 105990.
- [25] Wang X L, Zhou G Z, Liang L P, et al. Deep learning-based image analysis for in situ microscopic imaging of cell culture process[J]. *Engineering Applications of Artificial Intelligence*, 2024, **129**: 107621.
- [26] Liu J, Kuang W J, Liu J Q, et al. In-situ multi-phase flow imaging for particle dynamic tracking and characterization: Advances and applications[J]. *Chemical Engineering Journal*, 2022, **438**: 135554.
- [27] Mumuni A, Mumuni F. Data augmentation: a comprehensive survey of modern approaches[J]. *Array*, 2022, **16**: 100258.
- [28] Zhong Z, Zheng L, Kang G L, et al. Random erasing data augmentation[C]//*Proceedings of the AAAI Conference on Artificial Intelligence*. Palo Alto, California, USA: AAAI, 2020, **34**(7): 13001–13008.

- [29] Zhang H, Cisse M, Dauphin Y N, et al. Mixup: beyond empirical risk minimization[EB/OL]. 2017: arXiv: 1710.09412. <https://arxiv.org/abs/1710.09412>.
- [30] Yun S, Han D, Chun S, et al. CutMix: regularization strategy to train strong classifiers with localizable features[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South): IEEE, 2019: 6021–6031.
- [31] Bochkovskiy A, Wang C Y, Liao H M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. 2020: arXiv: 2004.10934. <https://arxiv.org/abs/2004.10934>.
- [32] Jackson P T, Atapour-Abarghouei A, Bonner S, et al. Style augmentation: data augmentation *via* style randomization[EB/OL]. 2018: arXiv: 1809.05375. <https://arxiv.org/abs/1809.05375>.
- [33] Goodfellow I J, Shlens J, Szegedy C. Explaining and harnessing adversarial examples[EB/OL]. 2014: arXiv: 1412.6572. <https://arxiv.org/abs/1412.6572>.
- [34] Madry A, Makelov A, Schmidt L, et al. Towards deep learning models resistant to adversarial attacks[EB/OL]. 2017: arXiv: 1706.06083. <https://arxiv.org/abs/1706.06083>.
- [35] Daus S, Buchwald T, Peuker U A. In-line image analysis of particulate processes with deep learning: optimizing training data generation *via* copy-paste augmentation[J]. Powder Technology, 2024, **443**: 119884.
- [36] Bischoff D, Walla B, Weuster-Botz D. Machine learning-based protein crystal detection for monitoring of crystallization processes enabled with large-scale synthetic data sets of photorealistic images[J]. Analytical and Bioanalytical Chemistry, 2022, **414**(21): 6379–6391.
- [37] He K M, Gkioxari G, Dollár P, et al. Mask R-CNN[C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017: 2980–2988.
- [38] Liu J, Yao T, Li M Y, et al. Computer vision-assisted high-throughput screening of crystallization additives for crystal size, shape, and agglomeration regulation[J]. Engineering, 2025, **54**: 308–319.
- [39] Liu J, Zhang Q Y, Chen M Y, et al. A verified open-access AI-based chemical microparticle image database for *in-situ* particle visualization and quantification in multi-phase flow[J]. Chemical Engineering Journal, 2023, **451**: 138940.
- [40] Kim H B, Cho Y H, Hong M S. End-to-end system for estimating crystallization kinetics using a deep learning-based approach[J]. Crystal Growth & Design, 2026, **26**(2): 861–873.
- [41] Calderon De Anda J, Wang X Z, Roberts K J. Multi-scale segmentation image analysis for the in-process monitoring of particle shape with batch crystallisers[J]. Chemical Engineering Science, 2005, **60**(4): 1053–1065.
- [42] Vancleef A, Maes D, Van Gerven T, et al. Flow-through microscopy and image analysis for crystallization processes[J]. Chemical Engineering Science, 2022, **248**: 117067.
- [43] Liu Y, Tian Y J, Zhao Y Z, et al. VMamba: visual state space model[C]//Advances in Neural Information Processing Systems 37 (NeurIPS 2024). Vancouver, BC, Canada: 2024: 103031–103063.
- [44] Gu A, Dao T. Mamba: linear-time sequence modeling with selective state spaces[EB/OL]. 2023: arXiv: 2312.00752. <https://arxiv.org/abs/2312.00752>.
- [45] Manee V, Zhu W, Romagnoli J A. A deep learning image-based sensor for real-time crystal size distribution characterization[J]. Industrial & Engineering Chemistry Research, 2019, **58**(51): 23175–23186.
- [46] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]// Proceedings of the IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017: 2980–2988.
- [47] Li M Y, Liu J, Yao T, et al. Deep-learning based *in-situ* micrograph analysis of high-density crystallization slurry using image and data enhancement strategy[J]. Powder Technology, 2024, **437**: 119582.
- [48] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Cham: Springer, 2015: 234–241.
- [49] Osiecka-Drewniak N, Galewski Z, Piwowarczyk M, et al. Deep learning analysis of crystallization using polarized light microscopy and U-Net segmentation[J]. The Journal of Physical Chemistry B, 2025, **129**(40): 10521–10527.
- [50] Huo Y, Li X, Tu B B. Image measurement of crystal size growth during cooling crystallization using high-speed imaging and a U-Net network[J]. Crystals, 2022, **12**(12): 1690.
- [51] Zhu Q H, Zhou G Z, Hou G H, et al. On-line image analysis for evaporative crystallization of xylose[J]. Powder Technology, 2025, **452**: 120446.
- [52] Liu Z, Lin Y T, Cao Y, et al. Swin Transformer: hierarchical vision Transformer using shifted windows[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, QC, Canada: IEEE, 2022: 9992–10002.
- [53] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [C]//Advances in Neural Information Processing Systems 30 (NIPS 2017). Long Beach, CA, USA: Curran, 2017: 5998–6008.
- [54] Su Z N, He J X, Zhou P P, et al. A high-throughput system combining microfluidic hydrogel droplets with deep learning for screening the antisolvent-crystallization conditions of active pharmaceutical ingredients[J]. Lab on a Chip, 2020, **20**(11): 1907–1916.
- [55] Wu Y Y, Gao Z G, Rohani S. Deep learning-based oriented object detection for in situ image monitoring and analysis: a process analytical technology (PAT) application for taurine crystallization [J]. Chemical Engineering Research and Design, 2021, **170**: 444–455.
- [56] Zhang J L, Meng Y M, Wu J F, et al. Monitoring sugar crystallization with deep neural networks[J]. Journal of Food Engineering, 2020, **280**: 109965.
- [57] Marcato A, Boccardo G, Pisano R. Enhancing mass transfer coefficient prediction from field emission scanning electron microscope images through convolutional neural networks and data augmentation techniques[J]. Processes, 2025, **13**(2): 365.
- [58] Ghiasi G, Cui Y, Srinivas A, et al. Simple copy-paste is a strong data augmentation method for instance segmentation[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA: IEEE, 2021: 2917–2927.
- [59] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, **63** (11): 139–144.
- [60] Gui J, Sun Z N, Wen Y G, et al. A review on generative

- adversarial networks: algorithms, theory, and applications[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, **35** (4): 3313–3332.
- [61] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal, QC, Canada: MIT Press, 2014: 2672–2680.
- [62] 董永生, 范世朝, 张宇, 等. 生成对抗网络的发展与挑战[J]. 信号处理, 2023, **39**(1): 154–175.
- Dong Y S, Fan S C, Zhang Y, et al. Development and challenge of generative adversarial network[J]. Journal of Signal Processing, 2023, **39**(1): 154–175.
- [63] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks[EB/OL]. 2015: arXiv: 1511.06434. <https://arxiv.org/abs/1511.06434>.
- [64] Schonfeld E, Schiele B, Khoreva A. A U-Net based discriminator for generative adversarial networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020: 8204–8213.
- [65] Jiang Y, Chang S, Wang Z. TransGAN: two pure transformers can make one strong GAN, and that can scale up[C]// Advances in Neural Information Processing Systems 34 (NeurIPS 2021). Curran, 2021: 14745–14758.
- [66] Xu R, Xu X Y, Chen K, et al. The nuts and bolts of adopting Transformer in GANs[EB/OL]. 2021: arXiv: 2110.13107. <https://arxiv.org/abs/2110.13107>.
- [67] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017: 2242–2251.
- [68] Karnewar A, Wang O. MSG-GAN: multi-scale gradients for generative adversarial networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020: 7796–7805.
- [69] Mirza M, Osindero S. Conditional generative adversarial nets[EB/OL]. 2014: arXiv: 1411.1784. <https://arxiv.org/abs/1411.1784>.
- [70] Chen X, Duan Y, Houthoofd R, et al. InfoGAN: interpretable representation learning by information maximizing generative adversarial nets[EB/OL]. 2016: arXiv: 1606.03657. <https://arxiv.org/abs/1606.03657>.
- [71] Liu Q, Yang S L, Li Z J, et al. Image generation evaluation: a comprehensive survey of human and automatic evaluations[J]. Frontiers of Information Technology & Electronic Engineering, 2025, **26**(7): 1027–1065.
- [72] Baraheem S S, Le T-N, Nguyen T V. Image synthesis: a review of methods, datasets, evaluation metrics, and future outlook[J]. Artificial Intelligence Review, 2023, **56**(10): 10813–10865.
- [73] Vagenknecht M, Soukup J, Chen A T, et al. A deep learning solution for particle size analysis in low resolution inline microscopy images based on generative adversarial network[J]. Powder Technology, 2023, **426**: 118641.
- [74] Liu Y, Jiang Y X, Gao Z L, et al. Generative convolutional monitoring method for online flooding recognition in packed towers [J]. Journal of the Taiwan Institute of Chemical Engineers, 2024, **165**: 105719.
- [75] Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial networks[C]//Proceedings of the 34th International Conference on Machine Learning. Sydney, NSW, Australia: PMLR, 2017: 214–223.
- [76] Fu Y C, Liu Y. BubGAN: bubble generative adversarial networks for synthesizing realistic bubbly flow images[J]. Chemical Engineering Science, 2019, **204**: 35–47.
- [77] Haas T, Schubert C, Eickhoff M, et al. BubCNN: bubble detection using Faster RCNN and shape regression network[J]. Chemical Engineering Science, 2020, **216**: 115467.
- [78] Kim Y, Park H. Deep learning-based automated and universal bubble detection and mask extraction in complex two-phase flows [J]. Scientific Reports, 2021, **11**: 8940.
- [79] Yaqub M W, Chen X Z. Vision transformers for three-phase flow classifications with data augmentation through generative adversarial networks[J]. AIChE Journal, 2025, **71**(10): e70002.
- [80] Isola P, Zhu J Y, Zhou T H, et al. Image-to-image translation with conditional adversarial networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017: 5967–5976.
- [81] Baniukiewicz P, Lutton J E, Collier S, et al. Generative adversarial networks for augmenting training data of microscopic cell images [J]. Frontiers in Computer Science, 2019, **1**: 10.
- [82] Vu C T, Phan T D, Chandler D M. S_3 : a spectral and spatial measure of local perceived sharpness in natural images[J]. IEEE Transactions on Image Processing, 2012, **21**(3): 934–945.
- [83] Zargari A, Topacio B R, Mashhadi N, et al. Enhanced cell segmentation with limited training datasets using cycle generative adversarial networks[J]. iScience, 2024, **27**(5): 109740.
- [84] Johnson J, Alahi A, Li F F. Perceptual losses for real-time style transfer and super-resolution[C]//Computer Vision – ECCV 2016. Cham: Springer, 2016: 694–711.
- [85] Zhang K, Liang J Y, Van Gool L, et al. Designing a practical degradation model for deep blind image super-resolution[C]// 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, QC, Canada: IEEE, 2022: 4771–4780.
- [86] Wei Z H, Huang Y D, Chen Y A, et al. A-ESRGAN: training real-world blind super-resolution with attention U-Net discriminators [C]//PRICAI 2023: Trends in Artificial Intelligence. Singapore: Springer, 2024: 16–27.
- [87] Neves M C, Filgueiras J, Kokkinoginis Z, et al. Enhancing experimental image quality in two-phase bubbly systems with super-resolution using generative adversarial networks[J]. International Journal of Multiphase Flow, 2024, **180**: 104952.
- [88] Jain S, Seth G, Paruthi A, et al. Synthetic data augmentation for surface defect detection and classification using deep learning[J]. Journal of Intelligent Manufacturing, 2022, **33**(4): 1007–1020.
- [89] Mohammed S S, Clarke H G. Conditional image-to-image translation generative adversarial network (cGAN) for fabric defect data augmentation[J]. Neural Computing and Applications, 2024, **36**: 20231–20244.
- [90] Yang L, Zhang Z L, Song Y, et al. Diffusion models: a comprehensive survey of methods and applications[J]. ACM Computing Surveys, 2023, **56**(4): 105.
- [91] Croitoru F A, Hondru V, Ionescu R T, et al. Diffusion models in vision: a survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, **45**(9): 10850–10869.
- [92] Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models

- [C]//Advances in Neural Information Processing Systems 33 (NeurIPS 2020. Curran, 2020: 6840–6851.
- [93] Song J M, Meng C L, Ermon S. Denoising diffusion implicit models [EB/OL]. 2020: arXiv: 2010.02502. <https://arxiv.org/abs/2010.02502>.
- [94] Rombach R, Blattmann A, Lorenz D, et al. High-resolution image synthesis with latent diffusion models[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE, 2022: 10674–10685.
- [95] Zhang P, Zhang B, Chen D, et al. Cross-domain correspondence learning for exemplar-based image translation[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020: 5142–5152.
- [96] Saleh A S, Croes K, Ceric H, et al. Novel concept-oriented synthetic data approach for training generative AI-Driven crystal grain analysis using diffusion model[J]. Computational Materials Science, 2025, **251**: 113723.
- [97] Lyu X R, Han H B, Ren P, et al. DiffusionLSTM: a framework for image sequence generation and its application to oil spill monitoring and prediction[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, **62**: 5632313.
- [98] Graves A, Mohamed A R, Hinton G. Speech recognition with deep recurrent neural networks[C]//2013 IEEE International Conference on Acoustics, Speech and Signal Processing. Vancouver, BC, Canada: IEEE, 2013: 6645–6649.
- [99] Cao L C, Gao F, Zhang T X, et al. A conditional diffusion model-based data augmentation method for structural health monitoring of composites[J]. Mechanical Systems and Signal Processing, 2025, **238**: 113155.
- [100] Huang T R, Gao Y, Li Z L, et al. A hybrid deep learning framework based on diffusion model and deep residual neural network for defect detection in composite plates[J]. Applied Sciences, 2023, **13**(10): 5843.
- [101] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017: 2261–2269.
- [102] Dubey S R, Singh S K. Transformer-based generative adversarial networks in computer vision: a comprehensive survey[J]. IEEE Transactions on Artificial Intelligence, 2024, **5**(10): 4851–4867.
- [103] Hu T, Zhang J N, Yi R, et al. AnomalyDiffusion: few-shot anomaly image generation with diffusion model[C]// Proceedings of the AAAI Conference on Artificial Intelligence. Vancouver, BC, Canada: AAAI, 2024, **38**(8): 8526–8534.
- [104] Baltrušaitis T, Ahuja C, Morency L P. Multimodal machine learning: a survey and taxonomy[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, **41**(2): 423–443.