

分类号: TP391

密 级: 公开

学校代码: 10363

学 号: 2210330124



安徽工程大学

Anhui Polytechnic University

硕士学位论文

题 目 面向工业小样本的
图像增强技术研究

论 文 作 者 陶秀文

校 内 导 师 江明（教授）

校 外 导 师 王冬

专业学位类别 电子信息

研究方向（领域） 人工智能与系统集成

论文提交日期: 2024 年 5 月 31 日

分类号：TP391

单位代码：10363

密类级：公开

学 号：2210330124

题 目 面向工业小样本的图像增强技术研究

题 目 RESEARCH ON IMAGE ENHANCEMENT
TECHNIQUES FOR INDUSTRIAL
SMALL SAMPLES

论 文 作 者：陶秀文

校 内 导 师：江明（教授）

校 外 导 师：王冬

专业学位类别：电子信息

研究方向（领域）：人工智能与系统集成

论文答辩日期：2024 年 5 月 19 日

安徽工程大学学位论文原创性声明

本人郑重声明：我恪守学术道德，崇尚严谨学风。所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已明确注明和引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品及成果的内容。论文为本人亲自撰写，我对所写的内容负责，并完全意识到本声明的法律后果由本人承担。

学位论文作者签名：陈庆云

日期：2024年5月31日

安徽工程大学学位论文版权使用授权书

学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅或借阅。本人授权安徽工程大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

保密 ☐，在 ____ 年解密后适用本版权书。

本学位论文属于

不保密 ☒。

学位论文作者签名：陈庆文

指导教师签名：江峰

日期：2024年5月31日

日期：2024年5月31日

面向工业小样本的图像增强技术研究

摘要

随着“中国制造 2025”战略的不断深化，工业制造正逐渐向着数字化、智能化的方向迈进。计算机视觉技术不断发展，为工业自动化提供了更加精准、高效的解决方案，推动了工业制造领域的智能化升级。在工业生产过程中，准确的图像分析和处理对于质量控制、产品检测以及设备维护至关重要。然而，传统的图像处理技术由于受限于数据量和算法的局限性，往往难以满足工业生产中图像处理精度和效率的需求。随着人工智能领域的迅速发展，基于深度学习的图像增强技术在工业领域里逐渐显现出巨大潜力和优势。但由于工业生产中存在数据集采集困难、样本数量不足以及样本缺乏多样性等制约高效自动化生产的问题，因此，本文针对工业小样本的图像增强技术进行了深入分析和研究。具体研究如下：

首先，针对工业样本数据集采集困难、样本数量不足和缺乏多样性的问题，提出两种虚拟样本现实化的方法。一种通过传统的图像增强方法(如裁剪、旋转和拼接)增加样本数量，再利用 CycleGAN 网络将虚拟样本转换成与真实样本风格相一致的现实样本；另一种是使用结合随机算法的虚拟引擎建模方法生成虚拟样本，以增加样本的多样性，再利用 CycleGAN 将虚拟样本转换成与真实样本风格相一致的现实样本。最后，针对两种方法得到的数据集，分别采用 SSIM 和 PSNR 两种图像评价指标对其进行质量评估，对生成样本的真实性和质量进行把控。实验结果表明，该方法可以有效降低工业样本数据集采集的难度、扩充样本数量且增加样本的多样性。

其次，为进一步提升 CycleGAN 网络性能，本文提出一种优化

算法框架和生成器的新网络框架。首先在原算法中融入 CBAM 和 FPN 突出细节特征；其次提出 C-U-Net 生成器网络，使用 U-Net 网络作为生成器网络，并将其部分传统卷积层替换为特征提取能力更强的 Convnext 卷积结构，提升图像生成质量；最后将改进后的 CycleGAN 网络应用于工业小样本领域。改进后的 CycleGAN 网络生成的虚拟样本分别使用 SSIM 和 PSNR 两种图像评价指标对其进行质量评估，验证生成样本的真实性和质量。

最后，为了验证所提出方法对工业小样本目标检测网络性能提升的有效性，本文将现实样本分别与改进方法得到的扩充样本数据集相融合，在 YOLOv5 和 YOLOv7 两个目标检测网络上进行实验，以验证添加扩充数据集后对目标检测网络的性能提升效果和实际应用中的效果。实验结果表明，本文提出的方法可以有效改善工业小样本问题。

关键词：工业小样本，图像增强，CycleGAN，虚拟样本，目标检测

RESEARCH ON IMAGE ENHANCEMENT TECHNIQUES FOR INDUSTRIAL SMALL SAMPLES

ABSTRACT

As the "Made in China 2025" strategy deepens, industrial manufacturing is progressively moving towards digitization and intelligence. Computer vision technology has been able to develop, providing more precise and efficient solutions for industrial automation, thereby driving the intelligent upgrade of the industrial manufacturing sector. Accurate image analysis and processing are crucial for quality control, product inspection, and equipment maintenance in industrial production processes.

However, traditional image processing techniques often struggle to meet the demands for image processing accuracy and efficiency in industrial production due to limitations in data volume and algorithmic constraints. With the rapid development in the field of artificial intelligence, image enhancement techniques based on deep learning are gradually demonstrating immense potential and advantages in the industrial domain. However, challenges such as difficulty in collecting datasets, insufficient sample quantities, and lack of sample diversity hinder efficient automation production in industrial settings. Therefore, this paper conducts an in-depth analysis and research on image enhancement techniques for industrial small samples. The specific research is as follows:

Firstly, addressing the challenges of difficult industrial sample dataset collection, insufficient sample quantities, and lack of diversity,

two methods for virtual sample realization are proposed. One approach involves augmenting sample quantities using traditional image enhancement techniques (such as cropping, rotation, and stitching), followed by utilizing CycleGAN to transform virtual samples into realistic samples consistent with the style of real samples. The other method entails utilizing a virtual engine modeling approach combined with stochastic algorithms to generate virtual samples, thus enhancing sample diversity, followed by employing CycleGAN to transform virtual samples into realistic samples consistent with the style of real samples. Finally, quality assessment of the datasets obtained through the two methods was conducted using two image evaluation metrics, SSIM and PSNR, respectively. This evaluation aimed to control the authenticity and quality of generated samples. Experimental results indicate that the proposed method effectively reduces the difficulty of collecting industrial sample datasets, expands sample quantity, and enhances sample diversity.

Furthermore, to further enhance the performance of the CycleGAN network, this paper proposes a new network framework for optimization algorithms and generators. Firstly, CBAM and FPN are integrated into the original algorithm to highlight detailed features. Secondly, a C-U-Net generator network is introduced, utilizing the U-Net network as the generator network and replacing some traditional convolutional layers with Convnext convolutional structures possessing stronger feature extraction capabilities to enhance the quality of image generation. Finally, the improved CycleGAN network is applied to the field of industrial small samples. Virtual samples generated by the improved CycleGAN network are assessed for their quality using both SSIM and PSNR image evaluation metrics, validating the authenticity and quality of the generated samples.

Finally, to validate the effectiveness of the proposed method in

enhancing the performance of industrial small sample object detection networks, this paper integrates real samples with augmented sample datasets obtained through the improved method. Experiments are conducted on two object detection networks, YOLOv5 and YOLOv7, to verify the performance improvement of the object detection networks and their practical applicability after adding augmented datasets. The experimental results demonstrate that the proposed method can effectively address the issue of industrial small sample problems.

Tao Xiuwen (Electronic Information)

Supervised by Professor Jiang Ming

KEY WORDS: industrial small sample, image enhancement, CycleGAN, virtual sample, target detection

目录

摘要	I
ABSTRACT	III
第 1 章 绪论	1
1.1 研究背景和意义	1
1.2 国内外研究现状	2
1.2.1 工业小样本数据集研究现状	3
1.2.2 虚拟样本生成技术研究现状	4
1.2.3 深度学习研究现状	6
1.3 论文主要工作	7
1.3.1 论文研究内容	7
1.3.2 论文技术路线	8
1.3.3 论文章节安排	9
第 2 章 工业小样本图像增强相关知识理论	11
2.1 传统图像增强方法	11
2.2 基于虚拟引擎的图像增强方法	12
2.3 深度神经网络生成算法	14
2.3.1 生成对抗网络	15
2.3.2 变分自编码器	16
2.3.3 自回归模型	18
2.3.4 Transformer 模型	19
2.4 实验评价指标	22
2.4.1 图像质量评价指标	22
2.4.2 目标检测网络评价指标	23
2.5 本章小结	23
第 3 章 基于特征增强的工业样本图像生成方法	25
3.1 传统图像增强引起的问题	25
3.2 改善传统工业样本图像增强方法	27

3.2.1 数据预处理	27
3.2.2 基于旋转、裁剪、拼接的工业小样本图像生成方法	27
3.2.3 融入 CycleGAN 的工业小样本图像生成	29
3.2.4 改善传统工业样本图像增强方法的图像质量评价实验	30
3.3 基于虚拟引擎的工业小样本图像增强方法	30
3.3.1 基于 blender 的工业小样本对象模型建立	31
3.3.2 结合随机算法的工业小样本生成方法	32
3.3.3 融入 CycleGAN 的域转换	37
3.3.4 图像质量评价实验	38
3.4 本章小结	39
第 4 章 改进 CycleGAN 算法	40
4.1 CycleGAN 图像生成算法	40
4.1.1 CycleGAN 模型整体架构	40
4.1.2 CycleGAN 工业小样本图像生成实验	41
4.2 CycleGAN 模型改进	42
4.2.1 融入 CBAM 和 FPN 突出细节特征	42
4.2.2 改进 U-Net 生成器提升细节效果	43
4.2.3 PatchGAN 增强判别能力	48
4.2.4 改进 CycleGAN 的工业样本图像生成实验	48
4.2.5 改进 CycleGAN 的工业样本图像质量评价结果	49
4.3 本章小结	49
第 5 章 实验验证与分析	50
5.1 基于传统图像增强与 CycleGAN 结合的图像增强方法	50
5.1.1 改善传统图像增强方法的目标检测网络验证实验	50
5.1.2 改善传统图像增强方法的实验分析	51
5.2 基于虚拟引擎与 CycleGAN 结合的图像增强方法	53
5.2.1 虚拟引擎与 CycleGAN 结合方法的目标检测网络验证实验	53
5.2.2 虚拟引擎与 CycleGAN 结合方法的实验分析	54
5.3 基于改进 CycleGAN 的图像增强方法	55

5.3.1 改进 CycleGAN 图像增强方法的目标检测网络验证实验	55
5.3.2 改进 CycleGAN 图像增强方法的实验分析	58
5.4 本章小结	58
第 6 章 总结与展望	60
6.1 总结	60
6.2 展望	60
参考文献	62
攻读学位期间发表的学位论文及参加的科研项目	69

第 1 章 绪论

1.1 研究背景和意义

图像增强技术是图像处理领域中的一项基本技术，旨在针对特定的应用需求对图像进行改进。通过采用各种特定方法，如旋转、对比度增强、裁剪等。图像增强技术能够显著改善图像的质量，并使得处理后的图像更适合人眼观察或机器识别^[1]。这些改进能够使图像更易于分析，从而提高图像的实用性和可用性，为各种应用场景提供更加有价值的图像数据。在工业领域，图像增强技术对于提高产品质量检测、故障诊断、自动化监控等环节的效率和准确性至关重要。当前阶段，制造业正经历着从单一的大规模批量生产模式向多元化、个性化的定制生产模式转型，这一发展趋势对提升自动化生产的灵活性与适应性提出了更为严苛的标准和需求。并且许多工业应用场景下的图像往往受限于环境因素如光照、噪声等，导致图像质量不佳，影响后续的图像分析和处理^[2]。与此同时，工业环境下的图像数据往往存在样本量少、标注困难等问题^[3]，传统的依赖大规模标注数据来训练和优化图像处理技术的方法，在此种情境下并不容易直接实施和奏效。在工业应用中，图像增强技术不仅能提升视觉检测的准确性，还能显著增强系统应对环境变化的能力，是图像处理领域持续研究的重要课题。

目前，以深度学习为代表的目标检测算法，如：YOLO、SSD 等单阶段目标检测算法已经体现出较高的检测率和稳定的检测效果，但是，此类检测算法在一些特定的小目标检测应用中，如专业定制和单件小批量灵活生产的工业对象、不同场景下工业对象杂乱放置的情况等，很难获取到足量且质量高的数据集。而基于深度学习的图像处理算法训练的模型的好坏与数据集的质量、规模息息相关，但由于客观原因存在，出现检测率低、误检率高等问题。面向工业小样本的图像增强技术的研究有助于推动工业自动化和智能化的发展，这类研究对于理解和提升小样本学习算法的泛化能力、适应性及效率具有重要意义。对于工业领域特有的问题和挑战，该研究方向的进展能够提供有效的技术解决方案^[4]。图 1-1 所示为工业机器人目标检测应用场景。

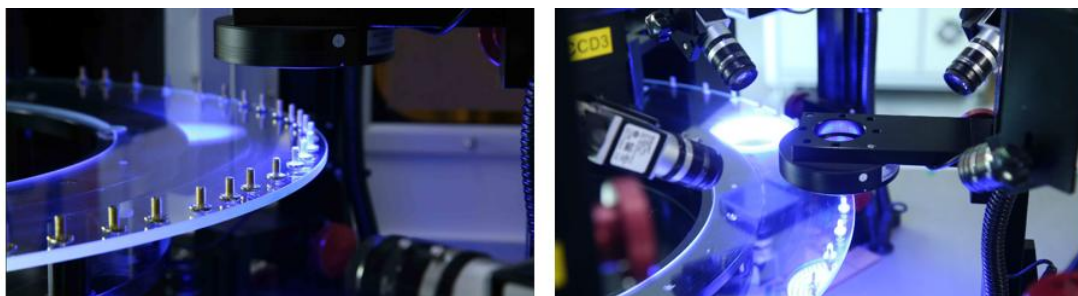


图 1-1 工业机器人目标检测应用场景

综上问题分析，针对工业领域中样本获取较难、耗费时间较长、占用成本较大等原因导致检测率低等问题，面向工业小样本图像增强技术的研究不仅对于推进工业领域的技术创新具有重要意义，也对于深入理解小样本学习机制、提升图像处理技术的实用性和效率具有深远的影响^[5]。因此，面向工业小样本的图像增强技术研究很有必要。常见图像增强技术对比如下图 1-2 所示。

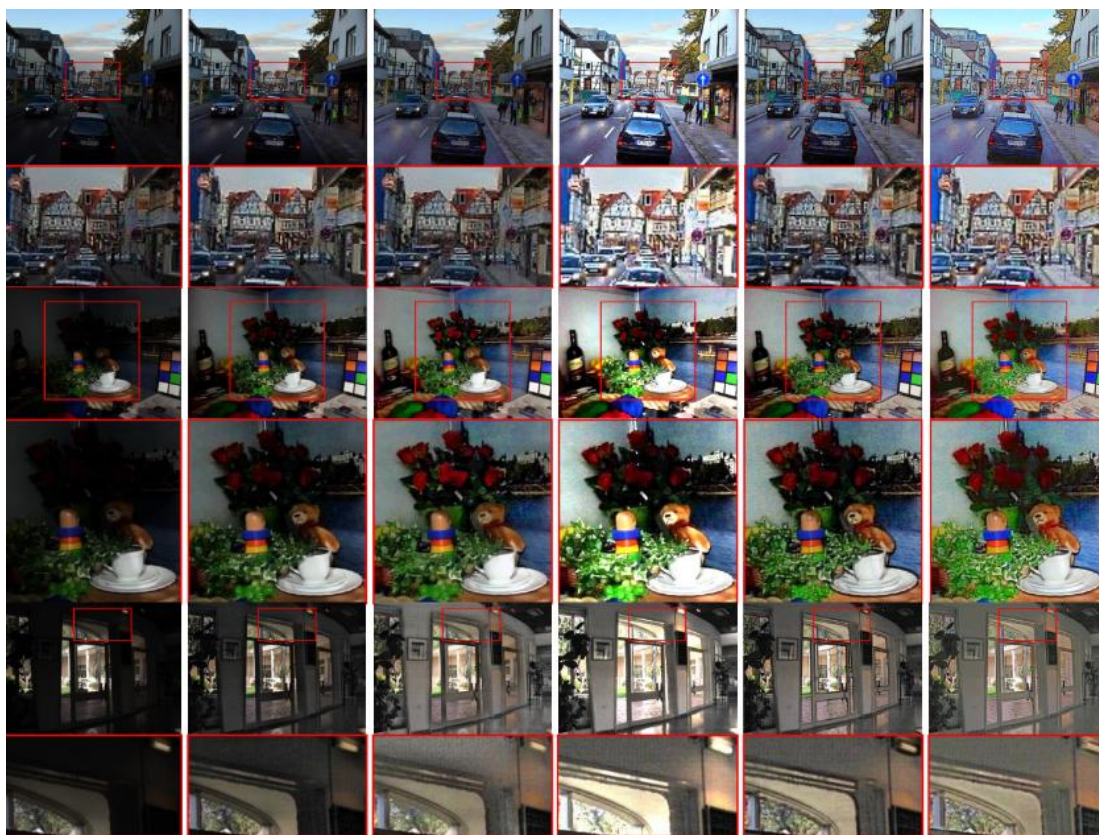


图 1-2 常见图像增强技术对比图

1.2 国内外研究现状

样本量不充足的问题，一直是各领域研究中的一大难题。面对这一挑战，不同领域的专家学者们积极探索，取得了诸多重要的成果。针对此类问题，研

究者们在机器学习、数据扩充技术两个方面，不断探索新的解决方案^[6]。

机器学习方面，研究者们利用原始数据构建灰色模型支持向量机(Support Vector Machine, SVM)等预测模型^[7-9]，这种方法通过挖掘数据间的潜在关联，提升了模型在小样本条件下的预测性能。同时，有学者基于特征构建了 box 图预测模型^[10-12]。这些模型在样本量有限的情况下，能够更准确地捕捉数据的分布和变化规律。

数据扩充方面，以蒙特卡洛方法为代表的虚拟样本生成技术成为了主要的研究焦点。这些方法通过生成新的虚拟样本，有效增加了训练数据的数量，提高了模型的泛化能力。近年来，迁移学习也成为了解决标签数据集不足问题的重要手段。这种方法不仅能够解决标签数据不足的问题，还能够实现知识的跨领域迁移，为扩充可用数据集提供了新的途径。

1.2.1 工业小样本数据集研究现状

工业小样本数据集的研究现状主要集中在如何有效利用有限的样本进行准确的模式识别和决策，这对于工件检测尤为关键。在国际上，许多研究团队和公司正在探索深度学习、迁移学习、元学习等技术来解决小样本学习问题。例如，通过元学习方法快速适应新任务，或者迁移在大型数据集上学到的知识到小样本工业场景^[13]。

在工业工件检测领域，数据集通常包含各种工件的图像，可能包括不同的尺寸、形状、材料和缺陷类型。虽然国际上存在一些公开的数据集，但这些数据集的多样性、质量和规模往往无法满足实际应用需求。此外，工件图像的复杂性和多变性，如不同尺寸、形状、材料和缺陷类型的影响，也为模型的泛化能力和鲁棒性提出了更高要求。

在国内，虽然工业 4.0 和智能制造的推进为工件检测技术的需求提供了动力，但如何在有限的标注数据下实现高效准确的学习，以及如何结合机器视觉等技术有效检测各种复杂工件表面缺陷，仍是亟待解决的问题。此外，中国在人工智能领域的快速发展虽然为研究提供了技术和数据支持，但如何将这些研究成果转化为实际生产力，以及如何确保技术的可靠性和安全性，也是重要的研究方向。

总体来说，工业小样本数据集的研究尚存在不足之处，包括数据集的质量和多样性问题，模型的泛化能力和鲁棒性挑战，以及实际应用中的转化效率和技术可靠性等。这些问题的存在，为广大研究者提供了广阔的研究空间，值得投入更多的精力和资源去探索 and 解决。

1.2.2 虚拟样本生成技术研究现状

文献[14]首次提出虚拟样本概念，其的核心思想是在不掌握样本概率分布函数的具体细节的情况下，通过充分利用已有的样本数据生成合理的、具有代表性的新样本。为确保新生成的样本能够保持原有样本的特性和结构，从而更有效地应用于后续的机器学习或模式识别任务中，因此虚拟样本遵循特定的数学表达式，即式(1-1)：

$$(x, y) \mapsto (Tx, y) \quad (1-1)$$

上述表达式中， (x, y) 代表原始样本，即用于生成新数据的基础数据点； T 表示利用原始样本生成新的、具有合理性的样本数据的虚拟样本生成技术； (Tx, y) 则代表通过这些技术生成具有与原始样本相似特性的虚拟样本。在其他领域中，虚拟样本的概念和应用方式可能有所不同，因此，虚拟样本在特定情境下被重新定义为式(1-2)，以适应不同的数据结构和处理需求。

$$(x, y) \mapsto (Tx, y_T f(x)) \quad (1-2)$$

在给出的表达式中， (x, y) 同样代表原始样本，为生成虚拟样本的基础数据。当面对分类任务时，代表离散类别标签的 y 用于区分不同的数据类别。而在回归任务中， y 则代表描述数据的某种连续变化趋势或结果的连续输出值。由式可知，需要对输入变换和输出变换进行深入分析和研究，以确保生成的虚拟样本既能够保持原始样本的某些特性，又能够在新的输入下产生合理的输出^[15]。

虚拟样本生成(Virtual Sample Generation, VSG)方法^[16]于1992年被首次提出，通过对样本数据集数量进行有效扩充，以此提升模型的性能，并将其运用于模式识别领域。文献[17]深入探讨了真实与虚拟样本在空间分布上的关联，强调VSG方法通过“填充”数据空间中的信息间隙，有效扩充样本，增强数据多样性，为模型训练和性能提升奠定坚实基础。

文献[18]阐述了 VSG 技术最初在图像识别领域得到了广泛应用，研究者们成功地生成了有效的虚拟图像样本，提升了模型的泛化能力。文献[19]进一步证明了结合先验知识的 VSG 方法与正则化策略具有等效性，即通过将先验知识整合为正则化矩阵^[20]，可以有效地约束和优化模型的训练过程。基于先验知识的 VSG 在图像识别领域成效显著，但工业过程先验知识常受限。为克服此挑战，研究者日益关注利用小样本数据提取知识生成虚拟样本的研究^[21]。这种方法旨在通过分析和学习有限样本的内在规律和特征，生成更多具有代表性的虚拟样本，从而丰富模型的训练数据，提高模型的性能。

进入 21 世纪后，VSG 技术引起了学术界广泛关注和研究^[22]。文献[23]探索了在随机神经网络模型中注入噪声的方法生成非线性虚拟样本的方法，这种方法对于解决样本不足问题导致的建模精度不足问题具有显著效果。此外，文献[24]还证明了在样本中注入噪声的 VSG 方法与神经网络结构设计的正则化方案具有相似性，其中正则化系数与噪声标准差密切相关，这一发现为 VSG 方法提供了更深入的理论支持。然而，文献[25]指出基于噪声和扰动的 VSG 方法在改进模型泛化性能方面具有一定的随机性和局限性。虽然这种方法能够在一定程度上提升模型的性能，但其效果往往受到噪声和扰动参数的影响，且可能不适用于所有类型的模型和数据集。

在解决小样本问题的过程中，研究者们不断探索创新方法。除了通过 Bagging^[26]，Foosting^[27]或 Bootstrap^[28] 等重采样技术来增加训练样本数量外，VSG 方法也提供了另一种有效途径。这些方法确实能够扩充样本数量，然而，它们不可避免地会改变原始样本的概率分布，进而可能导致建模过程中出现偏差。

为了克服小样本分布不平衡的局限性，文献[29]、文献[30]和文献[31]提出整体趋势扩散技术。该技术运用三角隶属度函数估算数据分布，扩展小样本空间。在扩展域内，整体趋势扩散技术生成多样虚拟样本，丰富模型训练信息^[32]。进一步，文献[33]对整体趋势扩散方法进行了改进，引入了差分进化算法。这一改进有效地约束了虚拟样本的生成范围，使得生成的虚拟样本更加贴近原始数据的分布，从而更准确地估计样本的趋势。差分进化算法的引入不仅提高了虚拟样本的质量，还增强了 VSG 方法的稳定性和可靠性。

文献[34]提出了一种创新的 VSG 方法,该方法基于区间核密度估计,通过分析小样本的分布情况来估计数据的总体分布,进而生成虚拟样本。为了应对小样本的不同分布特性,文献[35]、文献[36]和文献[37]进一步提出并探讨了基于双参数威布尔分布估计和多模态分布估计等多种 VSG 方法的优势和缺点。

为了克服上述文献中所提方法的局限性,文献[38]提出了合成少数过采样(synthetic minority oversampling, SMOTE)技术。该技术通过在小样本之间进行线性插值,从而提高了模型的泛化能力。文献[39]和文献[40]等人提出了基于神经网络隐含层插值的 VSG 方法,该方法能够生成非线性的虚拟样本,进一步丰富了虚拟样本的多样性。文献[41]提出了基于遗传算法(Genetic Algorithm, GA)和粒子群优化算法(Particle Swarm Optimization, PSO)优化生成虚拟样本的方法,通过优化算法来搜索最佳的虚拟样本生成策略,从而提高了虚拟样本的质量和有效性。

综上所述,目前各个领域都发展出了各自有效的样本生成方法,这些方法为处理小样本问题提供了新的思路,有助于提升模型的泛化能力和性能。其中,VSG 方法在各类生成任务中展现出显著的优势。然而,在工业小样本目标检测领域,尽管已有一些方法被提出,但这些方法无法排除一些低质量且对模型精度提升作用不大的虚拟样本。针对上述导致模型精度提升受到限制的问题,本文将分别从虚拟输入和虚拟输出两个维度进行研究^[42-44]。

1.2.3 深度学习研究现状

近年来,随着深度学习技术的迅猛进步,越来越多的研究者开始将其应用于目标检测任务中。卷积神经网络(Convolutional Neural Network, CNN)作为深度学习在计算机视觉领域的先驱,率先引领了使用深度学习技术进行视觉处理任务的潮流。文献[45]在 CNN 的基础上进行了优化,提出了创新的模型 AlexNet¹。尽管 AlexNet 解决了 CNN 中的许多问题,但单纯加深网络深度并非持续提升模型精度的有效途径。文献[46]首次提出两阶段目标检测算法 R-CNN,通过滑动窗口和特征提取,结合 SVM 分类,极大提高了检测精度。后续算法如 Fast-RCNN、Mask-RCNN 使用 RPN 网络筛选候选区域,减少冗余,但仍存在检测时间较慢的问题。

文献[47]提出单阶段目标检测框架 YOLO (You Only Look Once)，直接回归预测类别和位置，检测时间快但准确度略低。后续 YOLOV2、YOLOV3 等模型提高了单阶段检测精度。文献[12]提出的 SSD 算法(Single Shot MultiBox Detector)，则利用多尺度特征图进行同步回归预测，有效提升了小型物体的检测精度。这些算法的发展，为实时目标检测提供了有力支持。文献[48]提出了 DenseNet，利用逐层残差结构减少梯度消失，提高特征利用率，但不同尺度特征组合效果有限。

文献[49]和文献[50]提出了 RetinaNet 以改进版，通过优化训练损失，有效解决样本类别不平衡问题，提升检测性能。文献[51]提出了 DCN 算法，通过增强卷积网络几何变换能力，解决目标检测中的几何变换限制问题。文献[52]提出了压缩和激励网络(Squeeze-and Excitation Networks, SENet)，对特征提取网络输出进行权重计算，强化重要特征，预测物体类别和位置。文献[53]提出了 Cascade R-CNN，基于 Faster R-CNN 框架，使用多分支回归，探究不同交并比对性能的影响。文献[54]提出了无 anchor 的特征选择模块特征选择性无锚模块 (Feature Selective Anchor-Free Module, FSAF)，该算法优化单阶段目标检测，自适应选择特征并回归参数，无需人工设定锚框。

综上所述，这些算法在结构设计和性能上各具特色，且应用多样化。然而，它们均依赖于大量数据进行训练，当应用于新场景时，需重新训练网络，这涉及大量人工标注工作，成本高昂。此外，对于图像中物体尺寸较小的情况，这些算法的泛化性能不佳，凸显了深度学习在目标检测中广泛应用所面临的挑战。

1.3 论文主要工作

1.3.1 论文研究内容

本课题面向工业小样本的图像增强技术研究，旨在有以下三个方面研究内容：第一，改善传统图像增强的解决方案；第二，探索虚拟现实技术增加数据集的方法；第三，设计虚拟样本现实化方案。最后在深度学习目标检测网络上进行验证。本文研究内容框图如下图 1-3 所示。

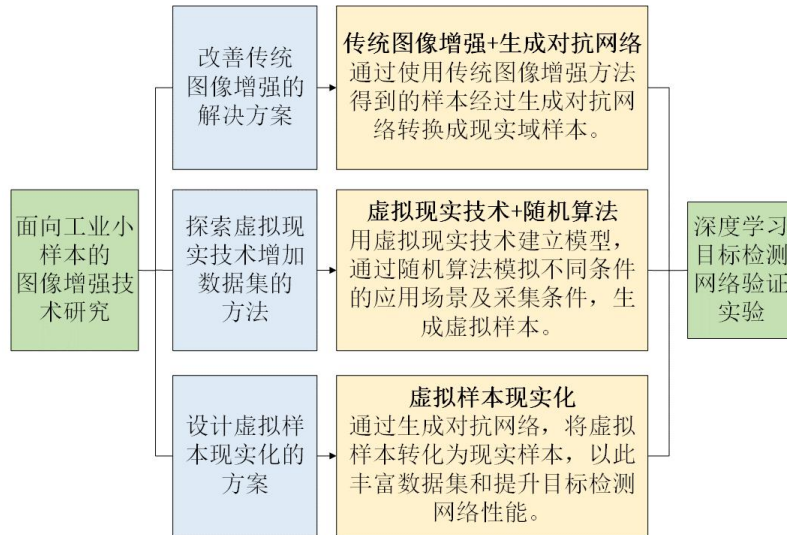


图 1-3 本文研究内容

1.3.2 论文技术路线

本文主要研究工业小样本的图像增强技术，旨在通过一系列技术手段扩充有限的小样本数据集，提升数据多样性和质量。本研究技术路线如图 1-4 所示。

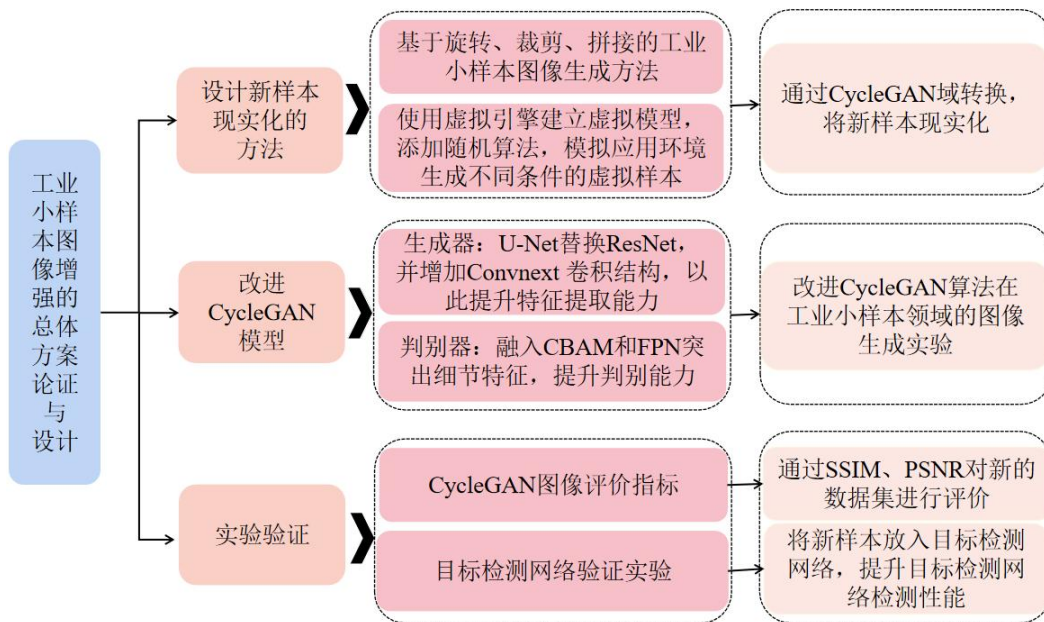


图 1-4 本文技术路线

首先通过旋转、裁剪、拼接等基础图像变换方式扩充原始样本，同时利用虚拟引擎建立虚拟样本模型，结合随机算法模拟不同应用环境生成具有不同条件的虚拟样本。在此基础上，进一步引入 CycleGAN 进行域转换，将虚拟样本转化为更贴近原始样本的分布。通过改进 CycleGAN 模型，实现更精确、更稳

定的转换效果，使得生成的虚拟样本在外观和特征上更加接近真实样本。

在完成上述一系列的图像增强与转换步骤后，采用图像评价指标 SSIM 和 PSNR 对生成的样本进行了细致的评估，筛选出一批高质量的样本。最后，将这些经过转换的虚拟样本与原始样本一起用于目标检测网络的训练与验证，从而证明本文提出的图像增强技术的有效性。

1.3.3 论文章节安排

本论文针对工业小样本存在的相关问题展开了研究，各章节安排如下：

第 1 章：绪论。首先概述了图像处理的研究现状与发展趋势，随后深入分析了工业小样本数据集、虚拟样本生成技术以及深度学习的研究现状。最后，明确了各个章节的研究内容与重点。

第 2 章：工业小样本图像增强相关知识理论。简要分析了工业小样本图像增强的相关理论知识。首先引出了传统图像增强方法，然后探讨了基于虚拟引擎的图像增强方法，最后研究了深度神经网络生成算法。这些方法为解决工业小样本图像增强问题提供了理论基础与方法路径。

第 3 章：基于特征增强的工业样本图像生成方法。首先深入提出了对传统工业图像增强方法的改进，随后探讨了基于虚拟引擎的工业小样本图像增强方法。这些方法旨在提高工业样本图像的质量和样本丰富度，为实践应用提供了有效的技术路径。

第 4 章：针对 CycleGAN 算法改进。从总体结构、生成器和判别器三个方面对 CycleGAN 进行改进，以此提升 CycleGAN 网络性能。最后利用改进后的 CycleGAN 算法进行工业小样本图像生成实验。

第 5 章：实验验证与分析。针对本文提出的基于特征增强的工业样本生成方法，融合改进后的 CycleGAN 工业小样本生成方法进行消融实验。同时，利用 YOLOv5 和 YOLOv7 模型进行对照实验，观察本文提出的图像增强方法对工业小样本目标检测性能提升的有效性，以此验证本文提出的图像增强方法的实用性，为实际应用提供了有力支持。

第 6 章：总结与展望。对本文提出的工业小样本图像增强方法进行概括总结，针对本文尚且存在的不足之处提出了一些改进方法。对未来发展方向进行

展望，未来设计一种基于工业小样本图像增系统，该系统可以有效对多场景、多样本进行集成，并加强对多样本工业场景的适应性，且结合迁移学习等技术提高性能。

第2章 工业小样本图像增强相关知识理论

工业小样本图像增强是连接传统制造业与现代智能技术的关键，它在提高生产效率、保证产品质量以及降低运营成本方面发挥着重要作用。随着技术的进步，这一领域将继续发展和创新，为工业自动化和智能化提供更多支持。图像增强在工业应用中扮演着至关重要的角色，尤其是在处理小样本数据时，其重要性更是不言而喻。本章将从三个不同的角度来审视图象增强技术：首先是传统图像增强方法，这些基础技术虽然简单，但在特定情况下仍然非常有效；其次是基于虚拟引擎的图像增强方法，这类方法利用先进的计算机图形技术来生成或改善图像；最后探讨基于深度学习的图像增强方法，这是近年来在图像处理领域的一大突破。

2.1 传统图像增强方法

传统图像增强技术通过一系列基于原始图像的操作来改进图像质量，如直方图均衡化、空间域滤波、频率域滤波、图像锐化与边缘增强、对比度调整以及伪彩色增强等多种方法。针对各异的应用场景，选取相宜的增强技术显得尤为关键。本部分将进一步对比剖析数种常见传统图像增强方法的优点与局限性，旨在为特定应用条件下选择最佳增强策略提供实际参考。

直方图均衡化^[55-57]作为一种重要的对比度增强技术，具有显著的优势与潜在局限性。其主要优点体现在：通过有效地重新分配像素灰度级，该方法能够显著提升图像的整体对比度，尤其对初始对比度较低的图像，效果尤为明显。这种技术无需依赖于复杂的参数调整，即可自动优化图像的灰度分布，实现视觉效果的显著改善，因此在实际应用中具有较高的通用性和便捷性。然而，直方图均衡化亦存在一定的不足之处。在增强图像全局对比度的过程中，由于过分追求灰度的均匀分布，可能会导致图像在亮部和暗部区域的细节信息被淹没，特别是对于那些本身就拥有较高对比度的图像，过于激进的增强处理可能会引发局部过曝或欠曝现象，从而造成图像某些重要细节的损失，影响后续的图像分析和理解精度。

空间域滤波作为图像处理的核心技术，在提升图像质量方面具有直观性和灵活性的双重优势。其通过直接操作像素值，实现平滑滤波以降噪、锐化滤波以增强细节。然而，此方法在优化图像的同时，也存在明显局限：平滑过程中可能导致边缘模糊；而在锐化操作中可能放大图像噪声，损害整体视觉效果。

频率域滤波技术在图像处理中扮演着关键角色。其优点体现在对噪声的高效抑制，特别是在对抗周期性噪声方面表现卓越；相较于空间域滤波，它更能精确地保留图像中的微小细节和结构信息。然而，该方法也存在局限性：计算过程较为复杂，需进行傅里叶变换与逆变换步骤，增加了运算；此外，对于非专业用户而言，使用频率域滤波技术要求具备一定的频域分析基础，降低了操作友好度。

图像锐化及边缘增强技术在提升视觉效果方面具有显著优势，能够有效增强图像中边缘和纹理部分的清晰度，赋予图像更加鲜明、立体的观感。但同时，此类方法的不足在于，在高对比度区域易引入额外噪声，并且若处理强度控制不当，可能引发图像过度锐化现象，导致局部失真或不自然。

对比度拉伸作为一种直观而便捷的图像处理手段，尤其适用于改善暗调或低对比度图像的可视性，通过调整灰度动态范围以提升整体画面层次感。然而，这种方法受限于其动态范围调整机制，可能无法妥善优化所有类型的图像质量，而且在极端亮度条件下，有可能损失亮部或暗部的关键细节信息。

伪彩色增强技术则通过将灰度图像映射为彩色表示来增强特征识别，尤其在诸如卫星遥感图像分析和医学影像等领域展现出了独特价值。尽管如此，这一技术亦伴随着一些潜在问题，如可能导致颜色失真，使得色彩表达与实际场景有所偏差；并且，不同的颜色映射方案选择会直接影响最终的增强效果，因此结果可能存在主观性和多样性。

2.2 基于虚拟引擎的图像增强方法

虚拟现实技术核心在于利用计算机技术，创造出逼真的虚拟环境，不仅包括视觉上的沉浸感，还包括听觉、触觉等多种感官的体验。通过模拟现实世界的场景和交互方式，虚拟现实技术为用户提供了与真实世界不同的体验，广泛应用于游戏、教育、医疗等领域，极大地丰富了人们的生活和工作体验。

Blender 是一款功能强大的开源软件，集成了多种 2D 和 3D 创作功能。Blender 使用 C/C++ 和 OpenGL(Open Graphics Library) 编写，这使得它具有高度的性能和灵活性^[58]。作为一款跨平台软件，Blender 在 Windows、MacOS 和 Linux 等操作系统上都有良好的兼容性。它不仅支持模型建模、动画制作和渲染，还具备视频编辑、特效合成等丰富功能。Blender 界面如图 2-1 所示，是以“数据——可视化——编辑”循环为基础。其在工业小样本图像增强领域的应用主要体现在其 3D 建模、渲染和动画功能，这些功能可以用于生成高质量的合成图像。这对于小样本数据集来说尤其重要，因为在许多实际情况下，获取大量的真实训练数据既昂贵又不切实际^[59]。

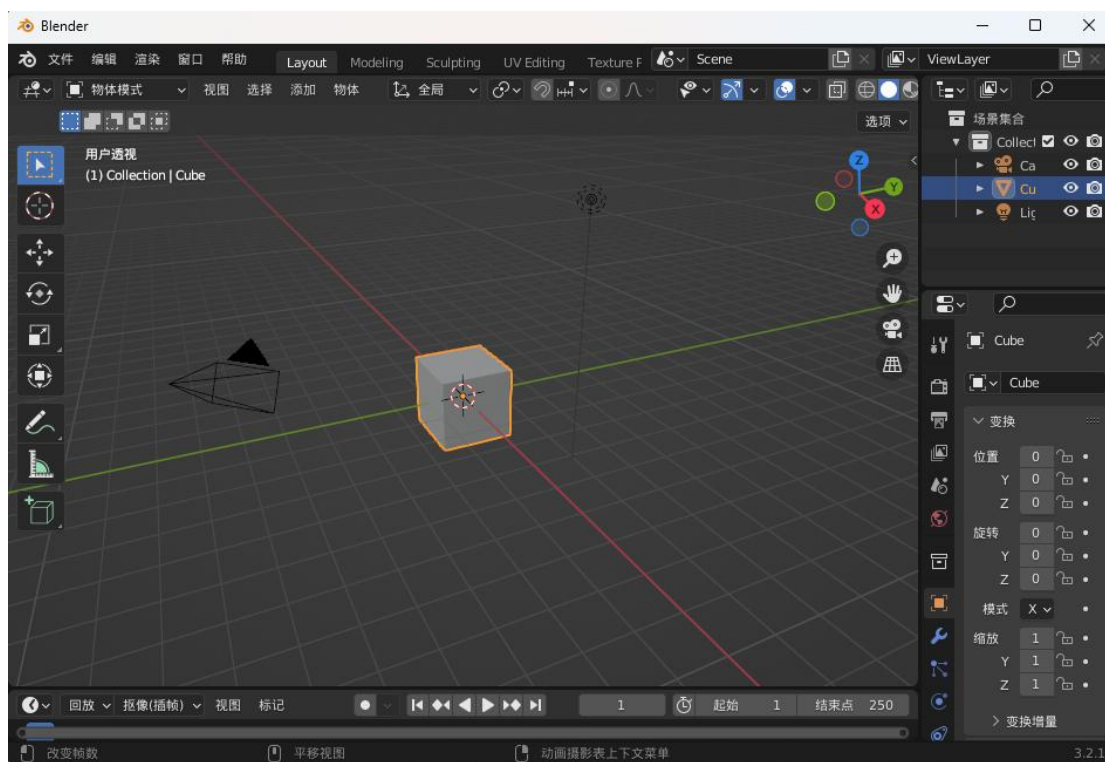


图 2-1 Blender 界面图

Blender 在工业小样本领域的一些应用如下：**合成图像生成：**在小样本数据集中，Blender 可以用来生成合成图像，这些图像可用于训练机器学习模型。通过在 3D 环境中创建场景和对象，然后从不同的角度和设置下渲染图像，可以生成多样化的数据集。

数据增强：Blender 可以用于数据增强，这是通过改变已有图像(如缩放、旋转、改变光照条件)来增加数据集多样性的过程。在小样本情况下，通过这种方式增加数据多样性对于提高机器学习模型的泛化能力是非常有帮助的。

场景模拟和变化：Blender 可以创建复杂的 3D 场景，模拟真实世界的条件，如不同的光照、天气情况和环境设置。这在需要对工业图像进行增强时尤其有用，因为可以准确控制和变化场景参数。

注释和标记：为了训练机器学习模型，通常需要对图像进行注释和标记。使用 Blender，可以在图像生成时就包含这些信息，例如通过在 3D 模型中直接嵌入标记。

生成难以获取的图像：在某些工业应用中，获取特定类型的图像可能极其困难或危险(例如在极端温度或压力条件下的设备图像)。Blender 可以用来模拟这些条件，并生成相应的图像用于训练。

通过这些方法，Blender 为工业小样本图像增强提供了强大的工具，可以显著提高数据集的质量和多样性，从而改善机器学习模型的性能。

关于 Blender 在工业小样本应用方面的使用并不广为人知。Blender 主要被认为是一个强大的 3D 建模、动画和渲染工具，并且在电影制作、游戏开发、艺术创作和教育领域中使用最为广泛。然而，随着机器学习和人工智能的迅速发展，Blender 开始被用于一些新的、更专业的领域，如工业设计、建筑可视化、以及作为数据增强工具在小样本学习中的应用^[60]。

随着深度学习技术的兴起，Blender 开始被用于生成合成数据，这对于训练机器学习模型非常有价值，特别是在获取真实数据成本高或不可行的情况下。Blender 能够创建详细的 3D 环境和对象，模拟真实世界条件和生成增强数据，对于小样本学习尤其重要。2010 年，随着计算能力的提高和深度学习的发展，Blender 开始被用于创建用于机器学习训练的数据集。近年来，随着技术的进一步发展，越来越多的研究小组和公司开始探索使用 Blender 生成合成数据来支持机器学习项目，尤其是在自动驾驶汽车、机器人视觉和工业自动化领域。

2.3 深度神经网络生成算法

经过大量实验研究与深入分析发现，深度学习技术在图像生成领域展现出了显著的优势。通过输入大量已知图像至多层级神经网络架构中，系统能够有效捕获并学习这些图像所蕴含的所有复杂多元特征。这些特征经逐层深度提取和整合后，构造和生成新的高质量图像。这一过程彰显了深度学习在图像特征

理解与创造性合成上的强大能力，有效证实了其在高保真、多样化的图像生成任务中的核心作用。常用于图像生成的方法有：生成对抗网络、简化后的变分自编码器、自回归模型和 Transformer 模型。

2.3.1 生成对抗网络

生成对抗网络(Generative Adversarial Nets, GAN)模型由 Goodfellow 等人于 2014 年提出^[61]。根据零和博弈理论，GAN 将图像生成识别问题视作两者间的博弈。训练中，生成器 G 学习真实数据分布以生成近似真实的假样本，而判别器 D 则通过评分判断输入数据的真伪。G 努力使生成数据难以区分，而 D 则力求准确识别。当 D 以 1/2 概率分辨真假数据时，GAN 达到纳什平衡。整个训练过程仅通过反向传播来优化 G 和 D。

生成对抗网络总体目标函数 $V(D, G)$ 为式(2-1)：

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (2-1)$$

其中 x 为真实样本数据， z 表示输入随机噪声， $p_{data}(x)$ 为真实样本分布， p_z 为随机噪声变量分布， $D(x)$ 表示当判别器输入为真实数据时的输出值， $D(G(z))$ 表示判别器输入为假样本时其判别器的输出值。 $E_{x \sim p_{data}(x)}$ 表示判别器输入为真实样本时的输出均值，同理 $E_{z \sim p_z(z)}$ 为判别器输入为假样本时其判别器的输出均值。生成对抗网络训练过程中需分别对生成器和判别器进行固定即分步训练，如式(2-2)所示：

$$V(D, G)_{\max} = \nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^m [\log D(x^{(i)}) + \log(1 - D(G(z^i)))] \quad (2-2)$$

判别器参数 θ_d 定义了其与真实数据空间的映射关系，表示为 $D(x; \theta_d)$ 。当输入真实数据 x 时，判别器努力给出最高判别结果 $D(x)$ 以验证数据的真实性。而对于生成器生成的假数据 $G(z)$ ，判别器则力求给出最低判别结果 $D(G(z))$ ，以证明其虚假性。在训练过程中，判别器通过递增式随机梯度参数 θ_d 进行更新。同时，生成器 G 也在学习随机数据 z 的概率分布，努力生成与真实样本相似分布的虚假数据，使判别器难以区分数据的真假。

生成器是一种深度学习模型，旨在学习将高斯分布映射到目标域概率分布 $p(x)$ 的转换关系，从而生成具有目标域图像特征的图像。在训练生成器网络时，需要将生成器和判别器作为一个整体考虑，并固定判别器的参数，从而生成器能够逐渐生成更真实、更自然的图像。图 2-2 所示为 GAN 模型结构示意图。

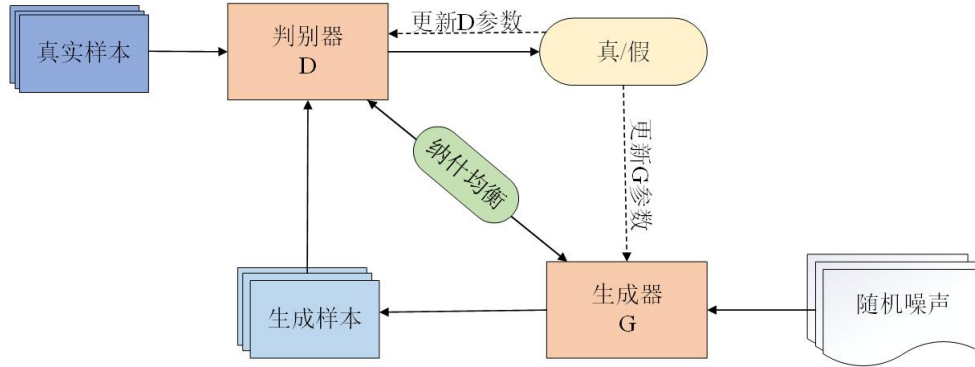


图 2-2 生成对抗网络模型结构示意图

自 2014 年 GAN 被首次提出以来，其在无监督和弱监督领域均取得了迅猛的发展。原始 GAN 采用全连接层构建生成器和判别器，导致生成能力受限、训练难度高且易陷入局部最优。该模型直接从高斯噪声生成样本，难以精确控制图像内容。随后出现了 CGAN、CycleGAN 和 StyleGAN 等基于 GAN 架构的弱监督模型变体，广泛应用于图像数据增强、生成及图文转换领域。如图 2-3 所示为 GAN 的发展脉络图^[62]。

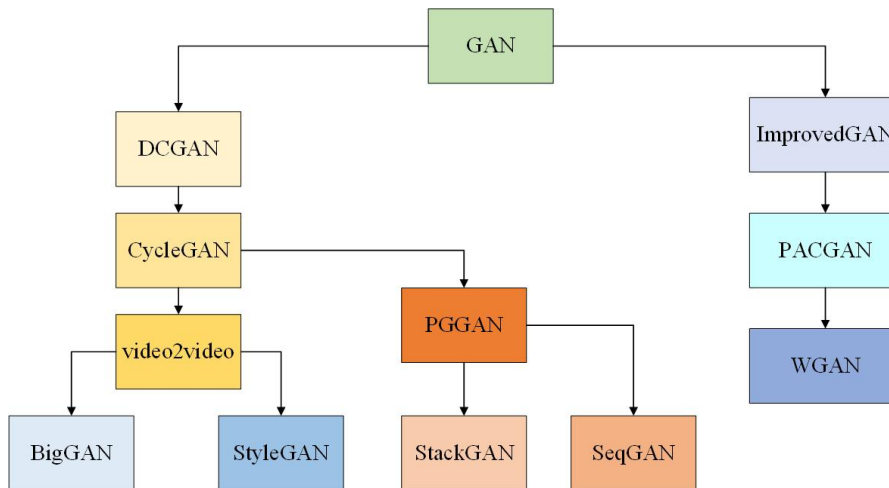


图 2-3 GAN 发展脉络图

2.3.2 变分自编码器

变分自编码器(Variational Autoencoder, VAE)^[63]是一种生成模型，通过变分

贝叶斯推断学习数据分布，并采样生成新样本。鉴于数据分布复杂，VAE 在潜在空间约束隐变量接近高斯分布。VAE 在自编码器基础上发展，差异在于自编码器学习压缩特征并重建数据，而 VAE 则融入变分处理构建数据概率分布参数，实现新样本生成^[64]。图 2-4 展示为变分自编码器模型结构图。

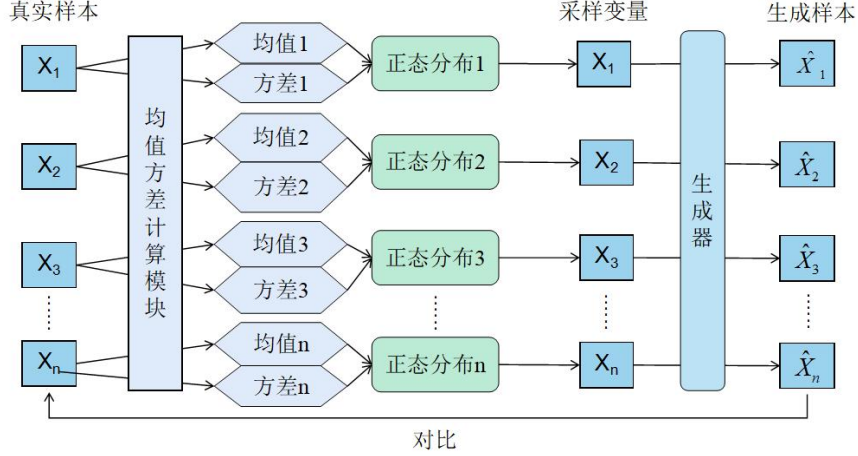


图 2-4 变分自编码器模型结构图

假设训练数据集为 $X = \{x_i\}_{i=1}^N$ 的 N 个样本，其中每个样本均为相互独立、遵循连续或离散分布的变量样本 x ， $X' = \{x'_i\}_{i=1}^N$ 为解码器生成的数据集合。变分自编码器的结构与自编码器结构类似，但引入了变分推断的概念，变分自编码器的编码器端输入样本数据 x ，输出隐变量 z ，权值和偏置项表示为 ϕ ；解码器端的输入是隐变量 z ，输出是数据概率分布的参数，权值和偏置表示为 θ 。根据贝叶斯定理和全概率分布可得式(2-3)和式(2-4)：

$$p_{\theta}(z|x) = \frac{p_{\theta}(x|z)p_{\theta}(z)}{p_{\theta}(x)} \quad (2-3)$$

$$p_{\theta}(x) = \int p_{\theta}(x|z)p_{\theta}(z)dz \quad (2-4)$$

上述公式中的后验分布 $p_{\theta}(z|x)$ 不能通过计算直接得到，因此变分自编码器引入了一个可以通过计算得到的先验分布 $q_{\phi}(z|x)$ 去近似后验分布 $p_{\theta}(z|x)$ 。

为了使 $q_{\phi}(z|x)$ 和 $p_{\theta}(z|x)$ 尽可能相等，其计算公式如式(2-5)所示：

$$\begin{aligned} \text{KL}(q_{\phi}(z|x) \| p_{\theta}(z|x)) &= E_{q_{\phi}(z|x)} \log \frac{q_{\phi}(z|x)}{p_{\theta}(z|x)} \\ &= E_{q_{\phi}(z|x)} \log \frac{q_{\phi}(z|x)}{p_{\theta}(z,x)} + \log p_{\theta}(x) \end{aligned} \quad (2-5)$$

由上式可得式(2-6):

$$\log p_{\theta}(x) = \text{KL}(q_{\phi}(z|x) \| p_{\theta}(z|x)) + L(\theta, \phi; X) \quad (2-6)$$

式(2-6)中的 $L(\theta, \phi; X)$ 成为变分下界(Evidence Lower Bound, ELBO), 可以写成式(2-7):

$$\begin{aligned} \log p_{\theta}(x) &\geq L(\theta, \phi; X) = -E_{q_{\phi}(z|x)} \log \frac{q_{\phi}(z|x)}{p_{\theta}(z, x)} \\ &= E_{q_{\phi}(z|x)} \log p_{\theta}(z, x) - E_{q_{\phi}(z|x)} \log q_{\phi}(z|x) \\ &= -\text{KL}(q_{\phi}(z|x) \| p_{\theta}(z)) + E_{q_{\phi}(z|x)} \log p_{\theta}(x|z) \end{aligned} \quad (2-7)$$

因此, 变分自编码器的优化目标为式(2-8):

$$L_{VAE}(\theta, \phi; X) = -\text{KL}(q_{\phi}(z|x) \| p_{\theta}(z)) + E_{q_{\phi}(z|x)} \log p_{\theta}(x|z) \quad (2-8)$$

由于模型无法进行反向传播, 因此变分自编码器引入了重参数化技巧。假设 $q_{\phi}(z|x)$ 是均值为 μ , 标准差为 σ 的高斯分布, 隐变量 z 的采样可由下式重参数化方式得到式(2-9):

$$z = \mu + \sigma \odot \varepsilon \quad (2-9)$$

其中 ε 为服从 $\varepsilon \sim N(0,1)$ 的辅助噪声变量。通过重参数化技巧, 可以使 $p_{\theta}(z) \sim N(0,1)$ 由于模型每次只采样一个数据样本, 因此式(2-4)的变分下界公式可以简化为式(2-10):

$$L_{VAE}(\theta, \phi; X) = \frac{1}{2} \sum (1 + \log \sigma^2 - \mu^2 - \sigma^2) + \log p_{\theta}(x|z) \quad (2-10)$$

引入辅助噪声 ε , 所有参数都可以使用随机梯度下降(Stochastic Gradient Descent, SGD)直接进行优化^[65]。

2.3.3 自回归模型

自回归模型(Autoregressive Model, AR)^[66]是时间序列分析中的一种模型, 它预测未来值基于过去的观测值。在图像生成中, 这种方法被用来顺序生成图像的每个像素值, 每个像素的值依赖于之前生成的像素值。自回归模型在图像生成中的应用是基于一种顺序决策过程, 旨在一步步地生成图像的每个像素或区块。这种方法源自于时间序列分析中的自回归模型, 其核心思想是预测未来值基于过去的观测值。

对于一幅图像 I ，可以将其视为像素序列 $p_1, p_2, \dots, p_N$ ，其中 N 是像素的总数。在自回归模型中，每个像素 p_i 的分布被建模为给定之前所有像素 $p_1, \dots, p_{i-1}$ 的条件概率： $P(p_i | p_1, \dots, p_{i-1})$ 。整幅图像的联合概率分布可以表示为所有像素条件概率的乘积式(2-11)：

$$P(I) = \prod_{i=1}^N P(p_i | p_1, \dots, p_{i-1}) \quad (2-11)$$

训练该模型通常涉及最大化训练数据上的对数似然，即式(2-12)：

$$\log P(I) = \sum_{i=1}^N \log P(p_i | p_1, \dots, p_{i-1}) \quad (2-12)$$

通过这种方式，模型学习到如何有效地表示和生成像素值，从而能够生成新的图像。在生成过程中，一旦模型被训练好，可以逐像素生成图像。这个过程从第一个像素开始，逐步生成每一个后续的像素，每一步都基于之前所有已生成的像素。这种方法允许模型在生成每个新像素时考虑到图像中的局部结构和全局上下文，从而生成高质量和高度真实感的图像。

自回归模型在图像生成领域的应用不仅限于静态图像，随着技术的进步和模型的改进它们也被用于动态场景的视频生成、条件图像合成以及其他多种复杂任务。尽管自回归模型在图像生成中取得了显著成就，但它们也面临着一些挑战和限制，例如生成高分辨率图像的速度较慢，以及在捕捉图像中的全局一致性方面的困难。为了解决这些问题，研究者提出了多种改进的自回归模型，如 PixelRNN 和 PixelCNN，通过改进的结构和训练策略来提高性能和生成质量。此外，与其他类型的生成模型相结合的方法也在探索中，旨在充分利用各自的优势，进一步推动图像生成技术的发展。

2.3.4 Transformer 模型

Google 提出的 Transformer 模型^[67]，提高了训练效率和模型性能。该模型不仅在自然语言处理任务中表现出色，还成功应用于视觉任务，展现了其广泛应用潜力。Transformer 的整体结构遵循“编码器——解码器”架构，如图 2-5 所示。编码器模块主要由多头注意力机制(Multi-Head Self-attention Mechanism MHSM)和全连接前馈层(Fully Connected Feed-Forward Network)组成，每个子层后都配备了层归一化和残差连接，以增强模型的稳定性和性能。

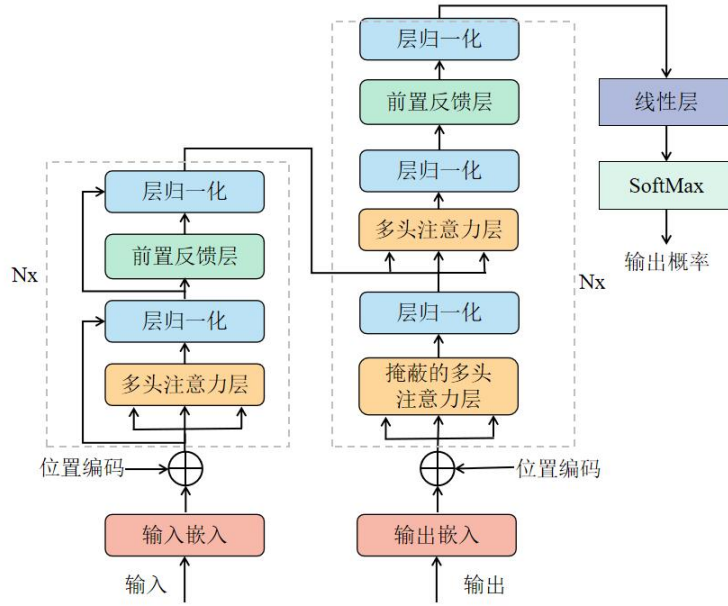


图 2-5 Transformer 网络结构示意图

解码器块包含三个子层：一为掩蔽多头注意力层，防信息泄露；二为整合编码器与上层解码器输出的多头注意力层，促信息交互融合；三为全连接前馈层，辅以层归一化与残差连接^[68]。点积注意力机制的对比如图 2-6 所示。

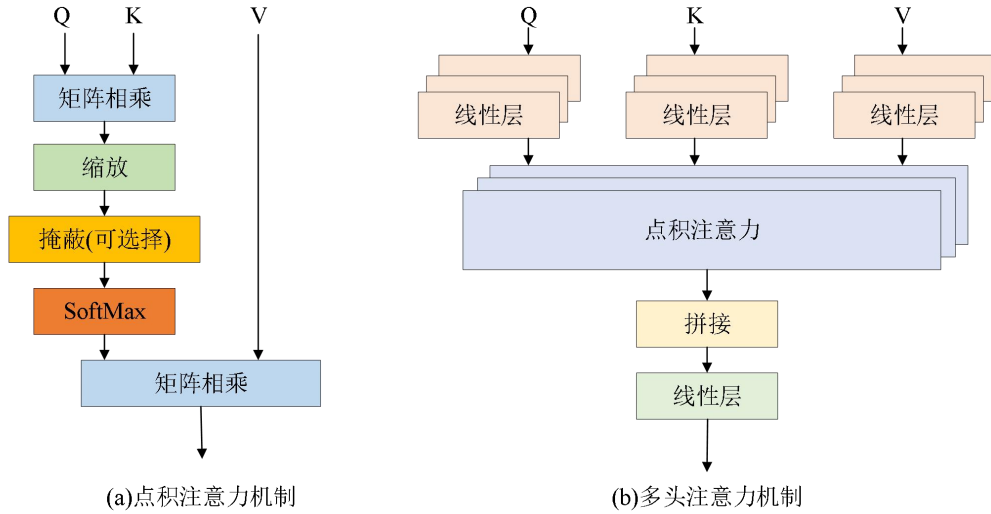


图 2-6 点积注意力机制与多头注意力机制示意图

模型中点积注意力机制可以由式(2-13)表示：

$$Attention(Q, K, V) = \text{soft max} \left(\frac{QK^T}{\sqrt{d_k}} \right) V \quad (2-13)$$

而多头注意力机制在 Transformer 模型中就是通过多个“头”或独立的注意力组件，并行地对输入进行 Q、K、V 的线性变换，并将结果拼接，最终输出更丰富的信息。多头注意力的公式表示如式(2-14)与式(2-15)所示。

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_h)W^O \quad (2-14)$$

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (2-15)$$

其中 d 为模型的维度，并且 $d_k = d / h$ ， W^O 为最后的线性投影层的权重矩阵。全连接前馈层的作用是为模型提供非线性变化，由两个线性层和一个激活函数组成，可以由式(2-16)表示：

$$FFN(x) = \max(0, xW_1 + b_1)W_2 + b_2 \quad (2-16)$$

由于序列顺序和位置信息对实验也很重要，因此 Transformer 提出了利用正弦余弦函数对序列进行编码，并与输入嵌入语句向量进行求和，如式(2-17)和式(2-18)所示。

$$PE_{(pos, 2i)} = \sin\left(\frac{pos}{10000^{2i/d_{model}}}\right) \quad (2-17)$$

$$PE_{(pos, 2i+1)} = \cos\left(\frac{pos}{10000^{2i/d_{model}}}\right) \quad (2-18)$$

Transformer 模型已经成功应用于计算机视觉任务中，除了三大视觉任务之外，在识别任务、图像增强、图像生成等任务上 Transformer 模型都表现优秀。

图像生成是一项重要的计算机视觉任务，旨在基于现有数据生成新的图像。最初基于 Transformer 的图像生成方法主要采用基于像素的生成方法，其中典型的代表包括 Image-Transformer 和 iGPT。这两种方法将图像生成视为自回归序列生成问题，iGPT 采用了无监督的训练方式，其准确率可达到 99.0%。然而，这些方法存在着计算成本高和无法处理高分辨率图像的缺点。

目前，基于 Transformer 的图像生成方法大多与 GAN 网络相结合，其中以 VQGAN、TransGAN 等为代表。图 2-7 为视觉 Transformer 发展历程图。

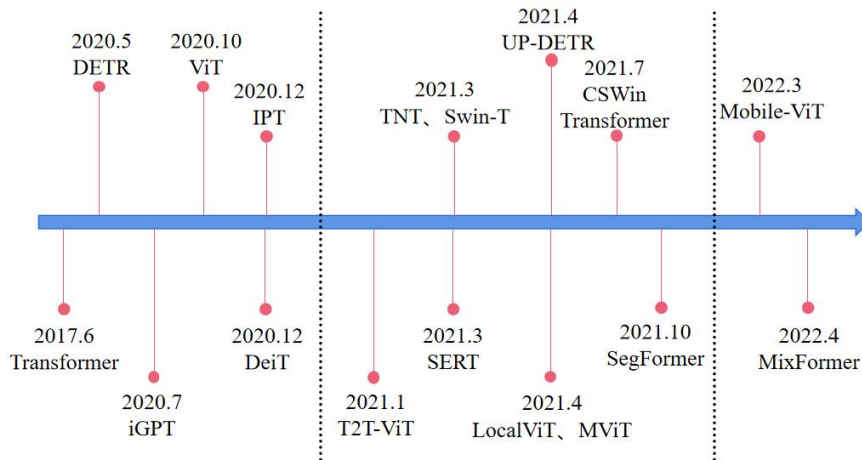


图 2-7 视觉 Transformer 发展历程

TransGAN 采用了纯 Transformer 架构，生成器首先将生成的图像切分为图像块，然后输入到判别器中，判别器利用编码器来区分真实图像和生成图像。与此不同的是，VQGAN 利用感知丰富的图像成分来表示图像，然后利用 Transformer 对这些图像成分的长距离关系进行建模。这使得 VQGAN 更适用于处理高分辨率的图像，并能够更好地捕捉图像中的细节和复杂性。

各类深度神经网络生成算法广泛应用于图像生成、自然语言处理等多个场景时，彰显出其独一无二的效能特点。这些算法能够依据训练集中的大量视觉信息，自动生成高分辨率、细节丰富且逼真的图像，甚至能够创造原本不存在于训练数据中的新奇视觉景象。而在自然语言处理方面，深度生成模型不仅能够理解和模仿人类语言的各种复杂结构和语境，还能生成连贯、逻辑严谨且富含创意的文本内容。随着深度学习理论研究的不断深入与实践应用的持续拓展，此类技术将有力满足不同行业领域对智能化创造性成果日益攀升的需求。

2.4 实验评价指标

为了系统地评估其生成结果的质量并筛选出具有较高真实度与细节保留程度的图像样本，本研究采用两个广泛应用的图像质量评价指标——结构相似性指数 (Structural Similarity Index, SSIM)^[69] 以及峰值信噪比 (Peak Signal-to-Noise Ratio, PSNR)^[70]，对生成的图像样本进行了严谨而细致的分析。

2.4.1 图像质量评价指标

图像质量评价中，常使用 SSIM 和 PSNR 两个指标。SSIM 衡量图像结构相似性，值域为(0,1)，越高表示生成的工件图像与真实图像越相似。PSNR 评估图像失真和噪声水平，值越大表示图像质量越高。

其结构相似性计算公式如式(2-19)所示：

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\mu_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (2-19)$$

其中 x 和 y 表示输入的两幅图像， μ_x 是 x 的平均值， μ_y 是 y 的平均值， σ_x^2 是 x 的方差， σ_y^2 是 y 的方差， μ_{xy} 表示 x 和 y 的协方差。 c_1 和 c_2 为常数。

PSNR(峰值信噪比)是一种广泛使用的图像质量评估指标，用于衡量图像的失真程度和噪声水平。式中 MSE 为用于量化修复后的图像 i 与原始图像 k 之间像素级别差异程度的均方误差，其公式如式(2-20)和式(2-21)所示：

$$MSE = \frac{1}{M \cdot N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [I(i, j) - K(i, j)]^2 \quad (2-20)$$

$$PSNR = 10 \log \left(\frac{MAX_I^2}{MSE} \right) \quad (2-21)$$

2.4.2 目标检测网络评价指标

为了验证本文方法的可用性，本文使用平均精度均值(mean average precision, mAP)作为目标工件检测模型的性能评价指标。根据样本的真实类别和模型检测类别的组合，将样本划分为模型正确检测目标工件数量(true positive, TP)，模型误检目标工件数量(false positive, FP)，模型漏检目标工件数量(false negative, FN)3 种类型。

精确率 P 表示正确检测的目标数量在全部检测结果中的比率， P 值越大表示检测精确度越高，误检越低。计算公式为式(2-22)：

$$Precision = \frac{TP}{TP + FP} = \frac{TP}{\text{all detctions}} \quad (2-22)$$

召回率 R 表示正确检测的目标数量占有所有目标真实数量的比率， R 值越高表示漏检测越低。计算公式如式(2-23)：

$$Recall = \frac{TP}{TP + FN} = \frac{TP}{\text{all ground truths}} \quad (2-23)$$

mAP 是各类 AP 值的平均， mAP 值越高，目标检测模型各个类别的平均检测效果越好。计算公式如式(2-24)和式(2-25)：

$$AP = \int_0^1 PdR \quad (2-24)$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (2-25)$$

2.5 本章小结

本章阐述了工业小样本图像增强的各种方法和理论。从传统的图像处理技术到最新的虚拟引擎和深度学习方法，每种技术都有其独特的优势和应用场景。

传统方法在简单性和直观性方面具有优势，而基于虚拟引擎的图像增强方法在生成高质量和逼真图像方面显示了其力量。深度学习方法，凭借其强大的数据处理能力和学习能力，正在逐渐成为图像增强领域的主导技术。总的来说，这些技术为处理和增强工业小样本图像提供了广泛的工具和可能性，对于推动工业自动化和智能化具有重要意义。

第3章 基于特征增强的工业样本图像生成方法

在前一章节详尽阐述了图像生成技术的基础理论与应用背景下，本章节将进一步聚焦于一种创新的工业场景下样本图像生成策略，该策略强调了特征增强技术的核心作用及其在实际应用中的实现原理。通过系统性地对原始工件图像进行精细化预处理步骤，并引入虚拟引擎建模方法等内容组合机制，切实提升所生成工业样本图像的整体质量和多样性，旨在丰富样本集，优化其对后续学习任务的适应性和效能。

3.1 传统图像增强引起的问题

在工业检测中，深度学习网络的泛化能力依赖于大量样本数据训练。小样本数据集直接训练易导致过拟合，影响实际应用性能。因此，数据扩增技术至关重要。主要包括空间几何变换如随机旋转、裁剪和拼接，以增加样本形态多样性；像素层面调整如亮度、色度、对比度调整及噪声添加，模拟真实场景变化；以及利用 GAN 在图像级别创新生成逼真图像，丰富训练数据集。这些措施共同提升模型性能，解决小样本训练问题。

通过分析现有方法的特性，可以发现基于空间几何变换和像素级操作的数据扩增技术虽能在一定程度上通过诸如旋转、裁剪、拼接等手段模拟多样化的视角和局部细节变化，同时利用对比度、亮度调整以及高斯模糊等像素处理增强模型对不同视觉条件的适应性，但此类方法本质上是对已有样本的再加工和变形。对于提升模型在处理未见过或少见过的样本类型上的检测精度成效依然有限，特别是在面对实际工业环境中的极端或者边缘案例时，仅依靠这些基本变换无法凭空创造出新的样本类别，也就难以弥补稀有样本的不足。

在实践中，数据集内部各类样本往往呈现不均匀分布，部分类型的样本频繁出现，而其他罕见类型的样本数量则极其有限。这种显著的分布不均衡性导致可用作模型训练与性能评估的有效样本基数较小，限制了模型对工业现场可能遭遇的各种复杂、多变情况的全面覆盖能力。图 3-1 为经过对比度调整、亮度校正、高斯模糊及多种方向旋转等典型预处理步骤后的样本变化效果。



图 3-1 传统方法效果图

3.2 改善传统工业样本图像增强方法

鉴于传统工业样本图像增强方法在实践过程中暴露的一些局限性，本节内容将阐述一种融合现有图像增强技术和 CycleGAN 域迁移算法的创新方案。通过整合这两种策略，旨在强化并凸显工业样本图像的关键特征，从而有效克服传统方法可能带来的不足。具体地，结合 CycleGAN 卓越的域迁移特性，在保持传统图像增强技术优点的同时，进一步提高工业样本图像的保真度和实用性，从而为后续深度学习模型训练以及工业自动化视觉识别系统的性能优化提供有力支撑，构建更为坚实的技术基础。

3.2.1 数据预处理

在处理工业小样本图像增强问题时，首先需要确保每个工件都在均匀光照下进行高分辨率的图像采集。这样做可以减少阴影和光斑对图像的影响，并保证工件细节的清晰度。同时，对于因拍摄错误或损坏而无法代表正常工件状态的图像，应进行筛选并去除，以清洗数据集。此外，从可用的图像中精心选择具有代表性和多样性的样本也至关重要。这些样本应覆盖工件可能出现的各种状态和变化，以确保模型训练的有效性和鲁棒性。通过这些细致的图像采集和数据清洗步骤，可以为后续的图像处理和增强任务打下坚实的基础。

3.2.2 基于旋转、裁剪、拼接的工业小样本图像生成方法

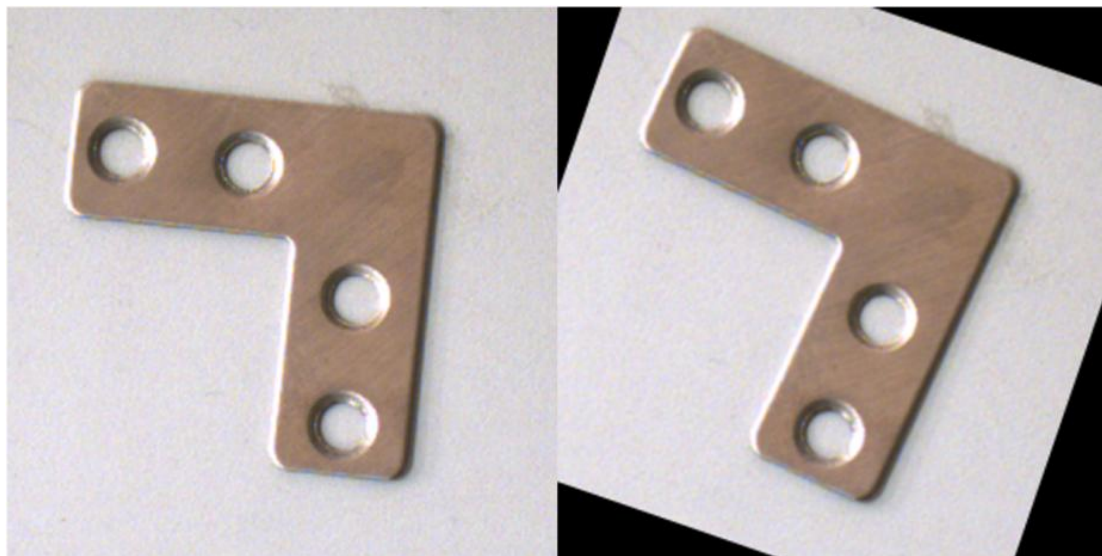
传统图像增强方法有很多，本文使用图像旋转、图像裁剪和图像拼接三种方法对原始样本进行图像增强。

图像旋转是将图像绕着某个点或某个轴进行旋转操作，使图像的方向或角度发生变化。旋转操作通过矩阵运算实现。具体地，假设要将图像旋转角度为 θ ，绕图像中心点旋转，则需要先将旋转中心点移动到原点，然后进行旋转，最后再将旋转中心点移回原来的位置。旋转操作可以使用式(3-1)和式(3-2)实现：

$$x' = x \cos \theta - y \sin \theta \quad (3-1)$$

$$y' = x \sin \theta + y \cos \theta \quad (3-2)$$

其中， x 和 y 是原始图像中的像素坐标， x' 和 y' 是旋转后图像中的像素坐标， θ 是旋转角度。如图 3-2 为工件样本旋转对比图。

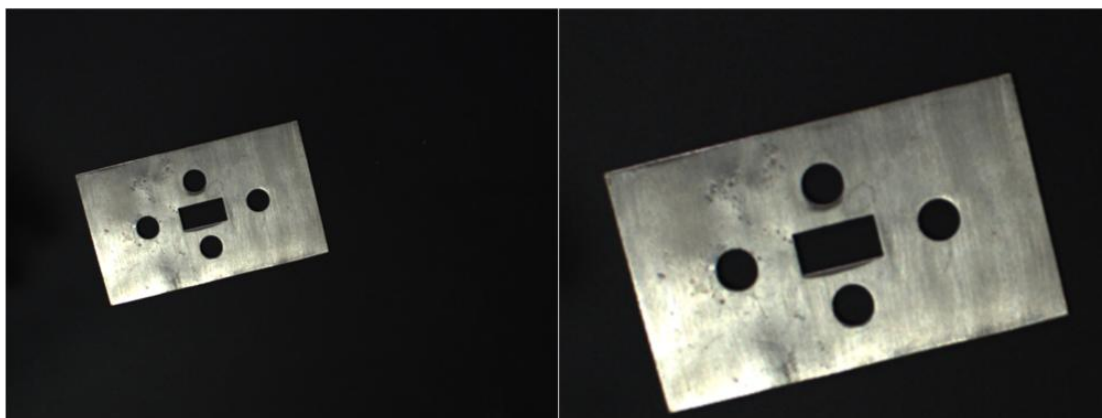


(a) 旋转前

(b) 旋转后

图 3-2 工件样本旋转对比图

图像裁剪是将图像的一部分截取下来，形成一个新的图像。裁剪操作可以通过修改图像的像素坐标范围来实现。具体地，假设要将图像的左上角坐标为 (x_1, y_1) ，右下角坐标为 (x_2, y_2) 的区域截取下来，则可以将图像的像素坐标限制在这个范围内，超出这个范围的像素将被忽略。截取后的图像的大小将由裁剪区域的大小决定。如图 3-3 为工件样本裁剪对比图。



(a) 裁剪前

(b) 裁剪后

图 3-3 工件样本裁剪对比图

图像拼接是将多张图像进行配准和拼接，通过图像处理技术将小图与大图进行融合。在这种方法中，需要进行图像配准、特征提取、特征匹配、图像变换等多个步骤。通过这些步骤，可以将小图拼接到大图的背景上，并实现无缝连接。如图 3-4 为工件样本拼接对比图。

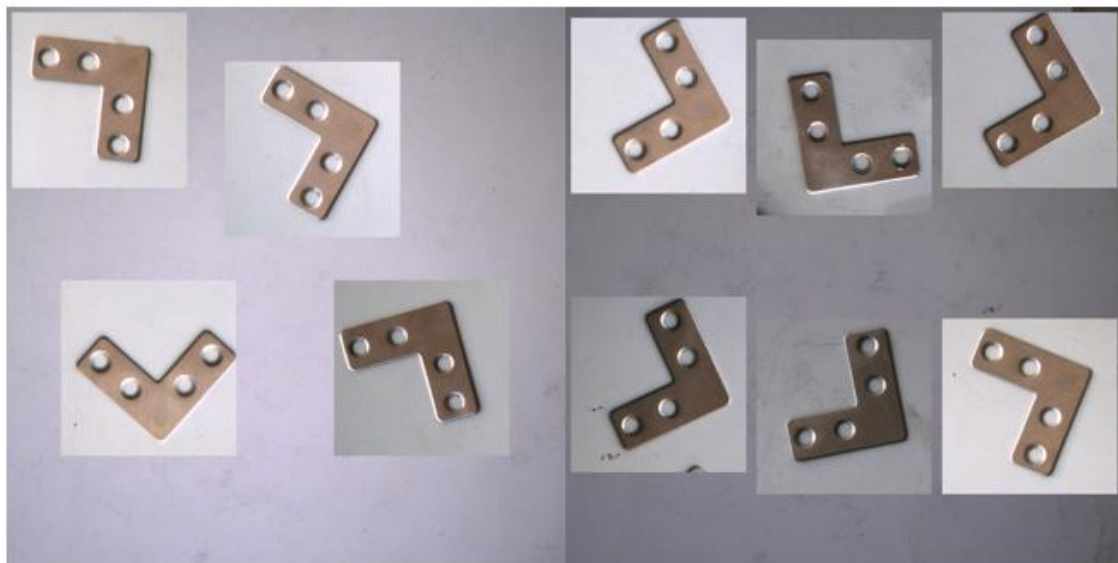


图 3-4 工件样本拼接对比图

3.2.3 融入 CycleGAN 的工业小样本图像生成

通过对原始工业图像执行一系列的传统增强操作后，扩充了初始的小样本数据集，增加了其内在的多样性和表征能力。然而，仅依靠传统增强手段尚不足以完全弥补工业图像数据量匮乏的问题，为此，针对前期预处理后扩增的图像数据集，本部分融入了 CycleGAN 框架以实现域的有效转换。在 CycleGAN 模型训练过程中，两个生成网络分别负责从增强图像到真实图像，以及反向从真实图像到增强图像的转换任务。同时，每一对生成器均有对应的判别器用于鉴别输入图像的真实性，通过循环一致性损失促使生成的图像在两次转换后尽可能接近原图，确保图像在风格迁移的过程中保留其本质特征。

如图 3-5 所示，经过 CycleGAN 深度学习架构的精心转换，处理后的图像不仅细致地保持了原始图像的关键结构信息，还注入了真实工业环境中多元且精细的纹理细节以及特有的背景特性，实现了对工业实物视觉表现的高度还原与仿真，从而在实质上缩小了由增强图像向真实应用场景过渡的差距。一系列严谨的实验验证证明，将传统的图像增强手法与 CycleGAN 的图像域转换技术相互融合，对于优化小样本工业图像的学习性能展现出显著的效果提升。通过实证研究及可视化结果的呈现，深入了解并确证这种创新策略如何在有限数据集的基础上，有力地促进了模型的理解深度、学习速度和泛化能力，为解决工业检测中的小样本学习难题提供了新的解决方案。

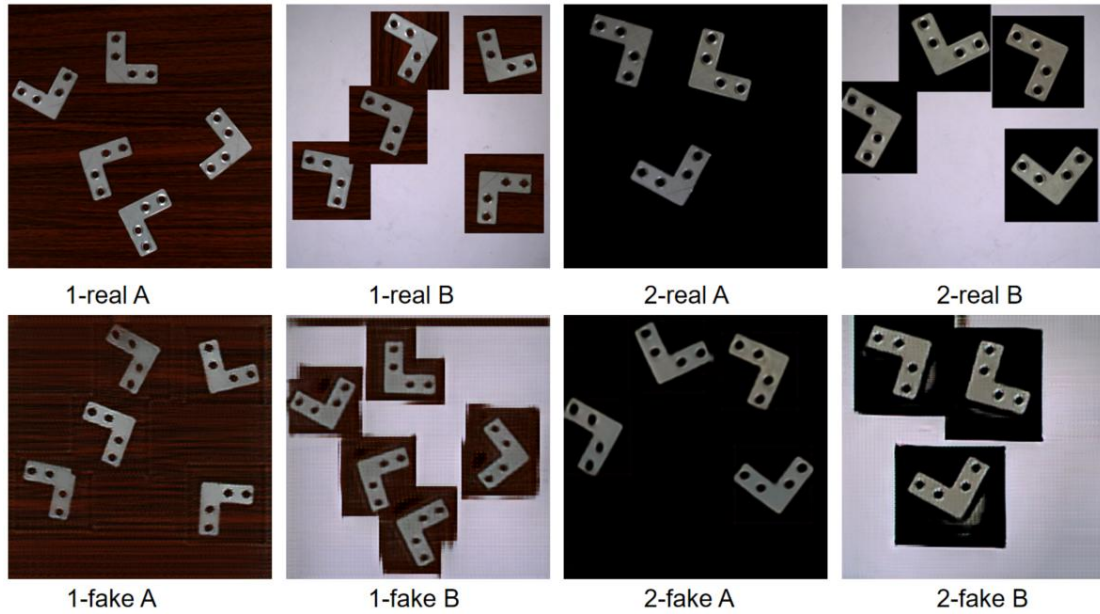


图 3-5 拼接图与原图经过 CycleGAN 域转换后图像对比

3.2.4 改善传统工业样本图像增强方法的图像质量评价实验

本节通过传统图像增强与 CycleGAN 结合的图像增强方法生成了一系列的数据集，采用结构相似性指数(SSIM)和峰值信噪比(PSNR)两种图像质量评价标准，对转化后图像的质量进行了量化评估。其中，SSIM 值越高则图像在结构和纹理上的相似度更高，PSNR 值越高则图像在信号恢复和降噪性能上越优越。实验结果如下表 3-1 所示，证实了该方法在提升工业图像质量和适用性方面的有效性和先进性。

表 3-1 实验客观评价指标

	PSNR/dB	SSIM
工件 1	32.767	0.677
工件 2	33.464	0.695
工件 3	31.325	0.667
工件 4	31.114	0.673

3.3 基于虚拟引擎的工业小样本图像增强方法

本部分采用虚拟现实技术生成虚拟样本，在真实拍摄的数据集中引入虚拟样本以此增加样本的数量，从而在小样本条件下增加目标检测网络的准确率。整体流程如下：首先在有限的条件下使用相机尽可能地采集真实样本；再使用虚拟引擎建立目标工件样本，模拟工件在各种环境下被视觉系统采集的状态；

进而使用随机算法，生成不同种类、不同材质、不同磨损程度、不同位姿、不同数量的目标工件，并生成包含目标工件的虚拟图片；最后使用 CycleGAN 进行真实域与虚拟域之间的转换。整体方案流程图如图 3-6 所示。

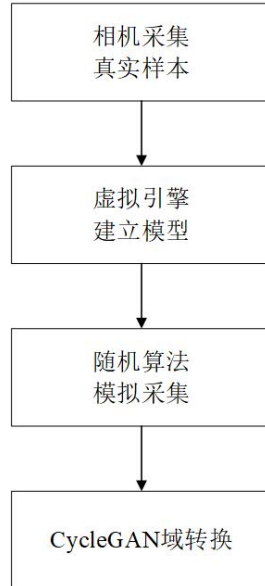


图 3-6 整体方案流程图

3.3.1 基于 blender 的工业小样本对象模型建立

本节旨在研究探索一种基于 Blender 的方法来构建工业小样本对象的三维模型，重点关注数量有限但重要性高的工业工件。鉴于在工业领域中常常只能获得少量的工件样本，这一研究的目标是开发一种能够有效应对样本稀缺情况的建模方法。通过利用 Blender 这一强大的开源三维建模软件，本研究提出了一个系统化的建模流程，以解决小样本环境下的具体挑战。

首先，研究确定了建模过程中需要面对的主要挑战，包括有限样本数量、工件的复杂几何形状，以及对高精度模型的需求。接着，详细探讨了如何利用 Blender 的高级功能，包括但不限于其建模工具、材质和纹理设置、以及灯光和渲染技术，来创建精确且细致的工业对象三维模型。

在数据准备阶段，本研究描述了如何收集工件的实物样本，进行必要的清洗和预处理，以及记录关键的几何和外观特征。随后，研究进入了创建基本几何形状的阶段，详细说明了如何使用 Blender 中的工具，如 Bezier 曲线和 NURBS 曲面等，精确地模拟工件的轮廓和主要特征。在模型细节增强阶段，研

究在 **Blender** 中添加更复杂的细节和纹理映射，以提高模型的真实感和视觉效果。此外，利用 **Blender** 的光照和材质设置进一步增强模型的视觉呈现。

建模过程的最后一步是模型的验证和迭代改进。通过与实物样本进行比较来验证模型的准确性，并根据反馈对模型进行迭代，以确保其最大程度地反映了实际工件的特征和细节。通过一系列的迭代和优化，模型的质量得到了显著提升，更好地满足了实际应用中的需求。如图 3-7 所示，为真实样本与虚拟引擎生成的虚拟样本对比，(a)组为真实样本，(b)组为虚拟样本。



图 3-7 真实样本与虚拟引擎生成的虚拟样本对比图

3.3.2 结合随机算法的工业小样本生成方法

Blender 是一个开源的三维建模和渲染软件，深受艺术家、设计师以及科学研究和工业领域专家的喜爱。其与 **Python** 脚本的无缝整合能力为用户提供了极大的灵活性和扩展性。**Blender** 与 **Python** 脚本结合使用，可以显著提升工作效率和创造力，具有多方面的优势。在工业领域，特别是当面对小样本工件的三维建模时，传统方法往往受限于样本数量的不足和工件复杂性的增加。**Blender** 通过与 **Python** 脚本的强大整合能力，能够在不同的光照条件、视角、纹理、材质以及背景组合下动态生成多变的工业图像。这一过程确保了生成样本的多样性和现实感，从而克服了传统小样本集在特征表达上的局限性。

首先，利用 **Python** 脚本在 **Blender** 中自动化繁琐的任务可以极大地提高效率。对于小样本工件，可以编写 **Python** 脚本来自动导入工件的测量数据，根据

这些数据生成基本的几何形状。由于工件数量有限，每一件精确建模都至关重要，Python 脚本能够确保每一步骤都按照精确的参数执行，减少人为错误，提高整体建模的精度和一致性。Python 的另一个优势在于其能够轻松集成各种算法，包括随机算法。在小样本情况下，通过引入随机化算法，可以模拟和生成工件可能的变异和偏差，这对于理解工件在实际生产中的表现至关重要。例如，可以使用 Python 编写的随机化脚本来调整模型的尺寸、形状、灯光、场景等属性，生成一系列可能的工件变体。这不仅增加了模型库的多样性，也为后续的质量控制和故障检测提供了宝贵数据。

此外，利用 Python 的强大计算能力，可以处理和分析大量数据，对于理解和优化工件模型特别有帮助。比如，可以编写脚本来分析工件的应力分布、热特性或其他重要参数，这些分析对于优化设计和提高工件的性能至关重要。最后，Blender 和 Python 的结合提供了一个高度可定制和扩展的平台。用户可以根据具体的需求编写脚本，开发定制的建模工具和流程。这种灵活性意味着无论工件的复杂性如何，Blender 都能提供有效的解决方案。

因此，Blender 结合 Python 脚本在工业小样本工件模型建立中的应用不仅展现了其高度的灵活性和效率，还提供了强大的功能，能够满足工业领域对精确和高质量三维模型的需求。通过自动化流程、引入随机算法、处理复杂数据和提供可定制的解决方案，这一组合成为了在面对小样本工件建模挑战时的一个强有力的工具。

随机数生成器是一种用于生成随机数序列的工具或算法。这些生成器通常依赖于一些确定性的数学公式，该公式使用一个称为“种子”的初始值，并根据一定的计算规则产生随机数。线性同余法(Linear Congruential Generator, LCG)是一种简单且常见的随机数生成器。它的数学公式如式(3-3)所示：

$$X(n+1) = (a * X(n) + c) \% m \quad (3-3)$$

式中，模 m ， $m > 0$ ；系数 a ， $0 < a < m$ ；增量 c ， $0 \leq c < m$ ；原始值(种子)。其中参数 c 、 m 和 a 直接影响了伪随机数产生的质量。LCG 能够生成一个具有足够随机特性的长周期序列，以满足大多数应用场合的需求。然而，尽管 LCG 因其简单高效而受到青睐。一般高 LCG 的 m 是 2 的指数次幂，这样取模操作截断最右的 32 或者 64 位即可。

场景选择：定义一系列可能的应用场景，如纯黑色背景环境、木质材质表

面以及接驳平台等具体情境。采用随机数生成算法对预先设定的场景集合进行抽样选取，从而减少人为因素影响，确保所研究的方法能够在不同背景及条件下稳健有效。如图 3-8 所示为随机生成不同场景下的同种工件图。



图 3-8 随机生成不同场景下的同种工件图

数量选择：明确规定一个用于决定生成物体数量的预设区间，该区间确立了生成物体数量的最小值与最大值界限。采用随机数生成技术从给定的数值范围内精确抽取一个整数值，以此来动态决定实际生成物体的具体数目。从而确保每次实验的独立性，并探究所研究方法在不同物体数量规模下的性能表现。如图 3-9 所示为随机生成不同数量的同种工件图。

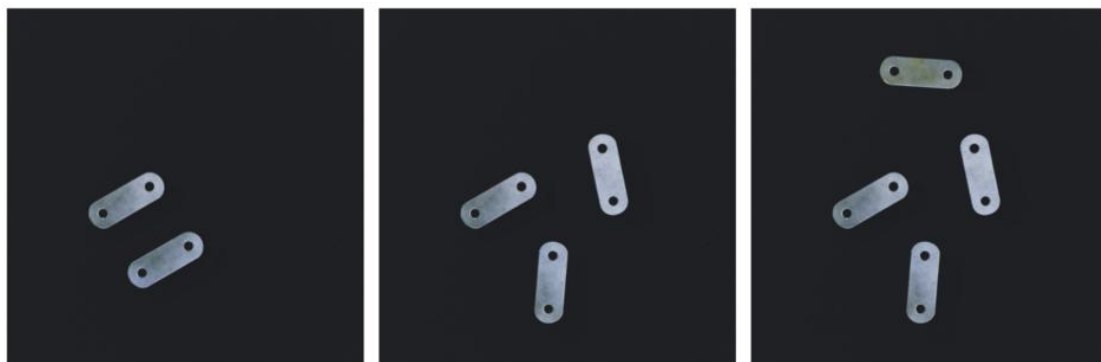


图 3-9 随机生成不同数量的同种工件图

种类选择：创建一个包含不同物体种类的列表，使用随机数生成器从列表中随机选择一个物体种类。设计一个详尽的工件种类集合，该集合囊括了一系列预先建模的不同形态工件，例如圆形工件、长方形工件以及具有复杂表面特征的凹凸形工件等。通过随机数生成器从该预设种类列表中随机选出一种具体工件类型，以此确保工件种类选取的随机性和多样性，且全面检验各类工件在不同情境下的适应性及性能差异，提升整体实验结果的有效性和可靠性。

如图 3-10 所示为随机生成同种场景下的不同种类工件对比图。



图 3-10 随机生成同种场景下的不同种类工件图

材质选择：针对每个已确定的物体种类，系统地设立一系列潜在的材质配置方案，包括但不限于木材质感、金属质地以及玻璃材质等多元选项。当某一类工件被随机选定后，将运用随机数生成技术，从与该种类相对应的材质子集里精确且随机地抽取一种材质属性，确保了每次选择的材质都具有随机性和不可预测性，从而大大增加了数据集的多样性和丰富性。

为了进一步提升生成图像的真实感和变异性，本方法还模拟并随机生成了不同程度的锈迹、磨损或其他环境影响因素。这些环境因素能够更准确地反映不同材质在实际应用中的老化状态和视觉效果，使得生成的图像更加贴近真实世界的场景。通过综合运用多种材质选项和环境影响因素，能够生成出既多样化又高度逼真的图像数据。

如图 3-11 所示为随机生成不同材质的同种工件图。

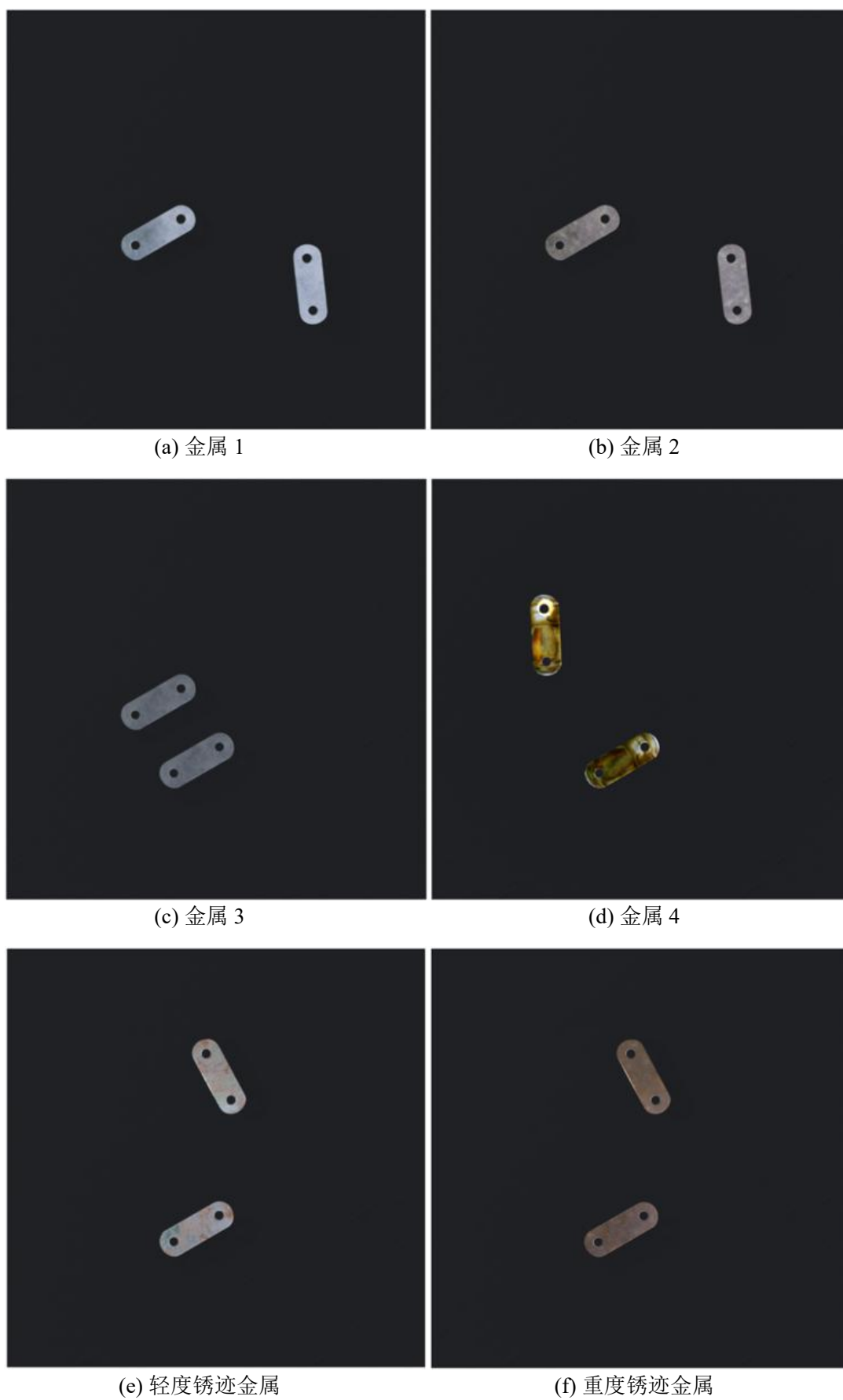


图 3-11 随机生成不同材质的同种工件图

位姿生成：在三维空间布局模拟阶段，针对各异的场景构建及各类工件特性差异，首先为每一个具体场景与工件类型精确定义其适用的位姿分布区间及特定约束条件。这些范围和条件充分考虑了物理可行性、功能需求等因素，从而反映现实世界中物体质心的位置布局 and 方向姿态的合理变化。随后，在这一严谨界定的框架内，运用随机数生成算法来创设符合实际应用场景的多样化位姿组合。每一种生成的位姿都严格遵循预设的界限与规则，确保无论是在位置坐标还是旋转角度上，都能准确无误地落在指定的有效区域之内，并满足所有施加的位姿限制条件。

通过随机化手段实现对复杂场景中众多工件多样而自然布局的高效模拟，进而增强整体设计或实验的真实性与可行性。如图 3-12 为随机生成在位姿范围内的随机位姿。



图 3-12 随机生成在位姿范围内的随机位姿

3.3.3 融入 CycleGAN 的域转换

利用 Blender 构建工业部件的三维模型，并通过设计和应用随机算法来增加图像的多样性和真实性，可以生成一个具有一定规模且覆盖多种可能场景变化的工业图像合成数据集。然而，为了进一步提升这些合成图像的真实感，使之更接近实际应用场景，本节将研究如何借助 CycleGAN 技术实现从虚拟图像域到真实图像域的有效转换。

针对本研究中的工业样本生成任务，配置 CycleGAN 模型，使其接受由 Blender 生成的虚拟图像作为源域，并以真实的工业环境图像作为目标域。在模型训练阶段，两个相互对抗的生成网络分别致力于将虚拟图像转化为真实风格的图像，以及将真实图像转换回虚拟风格。同时，两个判别网络负责鉴别输入

图像，以此推动生成器不断优化其转换能力。此外，CycleGAN 特有的循环一致性损失确保了图像在源域和目标域之间往返转换后的连贯性。

通过集成 CycleGAN 的训练流程，将前期通过 Blender 和随机算法生成的工业虚拟图像映射至真实图像空间内，使得原本的虚拟数据具备了更加真实和自然的视觉表现。这一方法极大地拓宽了可用于训练和测试工业视觉系统的样本范围，对于提高模型在复杂工业环境下的识别准确率和鲁棒性具有显著意义。

最终，经过 CycleGAN 转换后的图像样本与原始真实样本在视觉效果上达到了高度一致，实验证明了此种方法在增强工业小样本图像集方面的有效性及可行性。这一研究成果为解决工业视觉领域中普遍存在的小样本问题开辟了新的路径，实验效果如图 3-13 所示。

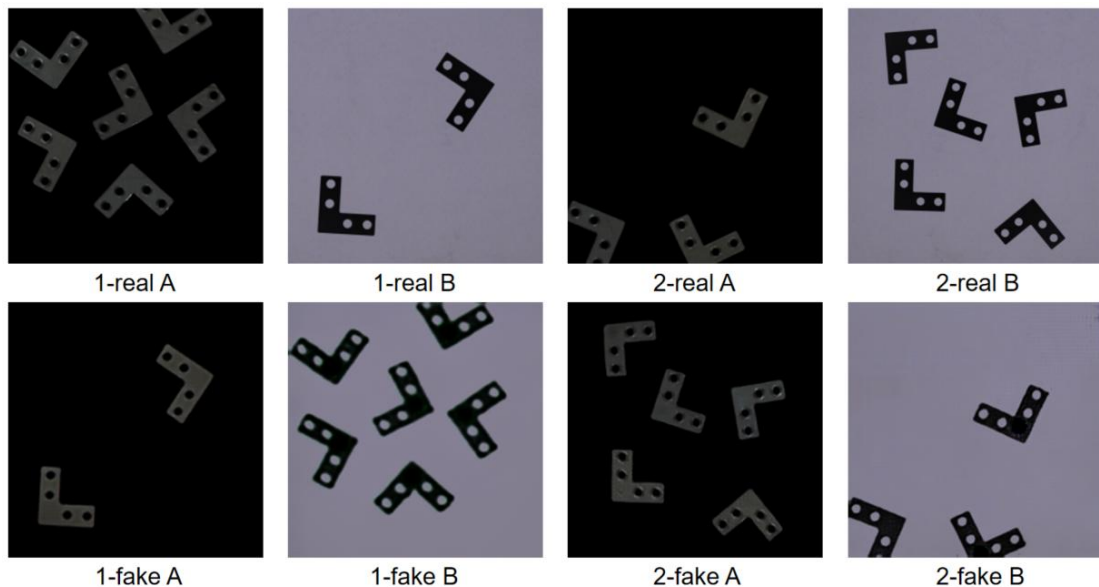


图 3-13 基于虚拟引擎建模生成的虚拟样本与真实样本域转换效果

3.3.4 图像质量评价实验

本节阐述了将虚拟引擎技术与 CycleGAN 模型相结合的图像增强实践，系统性地创建了一个扩展的图像数据集集合。为进一步量化评估这些经过增强处理后的图像品质，本文采纳了行业标准的图像质量评测指标，即结构相似性指数(SSIM)与峰值信噪比(PSNR)。按照这两个度量体系的原则，数据集中图像的 SSIM 值与 PSNR 值越高，则意味着该图像的视觉保真度与清晰度越佳，从而证实了采用这种方法所生成图像具备优良的质量特性。实验结果如下表 3-2 所示。

实验结果表明，在应用虚拟引擎技术与 CycleGAN 模型对初始图像数据集进行深度学习增强之后，生成的新图像数据在各项评测指标上均有显著提升。证明了经过增强的图像在结构和纹理细节方面与真实场景保持了更高的相似性，新生成图像具有更低的噪声水平和更高的清晰度。这一系列量化结果有力证明了结合虚拟引擎与 CycleGAN 的方法在图像增强任务上的有效性和可行性，对于提高图像质量和扩大训练样本规模具有积极的应用价值。

表 3-2 实验客观评价指标

	PSNR/dB	SSIM
工件 1	34.141	0.739
工件 2	32.766	0.718
工件 3	33.282	0.720
工件 4	33.710	0.724

3.4 本章小结

本章详细阐述了将虚拟图像转换为现实图像的方法。首先，通过两种途径获取虚拟样本：一是利用传统图像增强方法生成虚拟样本，二是利用虚拟引擎建立对象模型，添加随机算法生成虚拟样本。随后，借助 CycleGAN 网络，对现实工件样本和两种虚拟样本进行域转换，使虚拟样本现实化。这一过程旨在增强样本特征，丰富数据集，以提升工业小样本图像的质量和多样性，为进一步的实验与应用提供可靠基础。

第 4 章 改进 CycleGAN 算法

在工业自动化的进程中，图像处理和目标检测对于提升生产效率和保证质量至关重要。然而，特定工件的小样本问题限制了传统深度学习方法的效果，因为这些方法通常需要大量标记数据以实现可靠性能。生成对抗网络(GAN)特别是 CycleGAN，在图像增强方面显示出了潜力。为进一步提升对工件的检测精度，在此基础上进行优化改进，本章节研究提出了改进的 CycleGAN 算法。该算法更专注于增强生成图像的质量和相关性，以有效支持小样本下的目标检测。改进措施包括引入注意力机制以精确捕捉工件细节，优化生成器以增强对小样本的适应性，以及实施先进的数据增强技术以提升泛化能力并在自建数据集中进行生成效果验证。通过改进模型生成工件样本的结果分析比对，总结出优势与不足，为工业图像处理领域提供改进的思路方向。

4.1 CycleGAN 图像生成算法

CycleGAN 是一种生成对抗网络(GAN)，特别设计用于图像到图像的转换任务，即无需成对数据也能将图像从一个域转换到另一个域。例如，它可以将夏天的风景转换为冬天的样子，或将马的图像转换为斑马的样子。CycleGAN 的核心在于它的循环一致性损失，这使得原始图像经过两次域转换(A 到 B，再从 B 回到 A)后能够重建回原始图像。图 4-1 是将马的图像转换为斑马图像。



图 4-1 CycleGAN 模型图像生成案例

4.1.1 CycleGAN 模型整体架构

CycleGAN 模型最早由 Zhu 等^[71]提出，旨在解决非配对图像翻译问题。该

模型引入了独特的循环一致性损失函数作为关键约束，确保转换过程的连贯性和准确性。它通过在源域与生成的假源域之间构建损失函数，确保了即便在缺乏成对数据的情况下，生成器也能学习到有效的风格转换映射。这种损失函数的设计使得 CycleGAN 能够实现不同风格间的相互转化，同时保持图像内容的一致性和风格的转换效果。

在 CycleGAN 诞生之前，pix2pix 算法曾是图像风格转化的主流方法。然而，pix2pix 的一个显著局限在于它要求训练数据必须是成对的。这种成对的数据获取难度极大，特别是在处理不同风格、不同场景的图像时。为了克服这一限制，CycleGAN 引入了双向循环生成的结构，这种设计使得 CycleGAN 能够利用非成对数据集进行训练，从而显著拓宽了其应用场景。

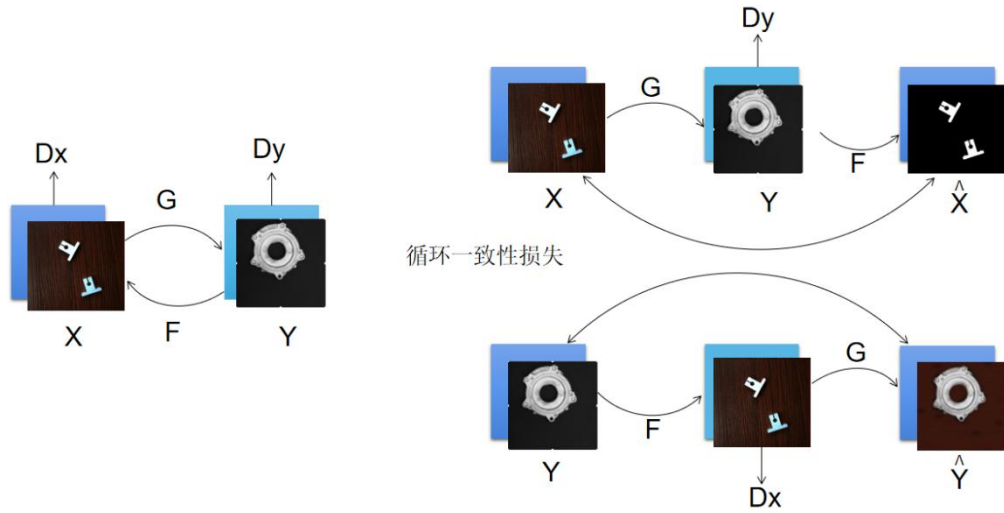


图 4-2 CycleGAN 算法框架图

图 4-2 展示了 CycleGAN 的算法框架。CycleGAN 的算法框架展示了 CycleGAN 的基本结构和训练流程。具体来说，X 域数据被映射到 Y 域，构建前向循环一致性损失；同时，Y 域数据被映射回 X 域，构建反向循环一致性损失。通过这些损失函数的优化，生成器 G 和判别器 F 能够逐步实现非配对图像间的风格迁移。

4.1.2 CycleGAN 工业小样本图像生成实验

本节实验旨在分析研究 CycleGAN 模型在工业小样本图像生成中的性能，两种风格工件样本数据集无需成对出现。

实验过程中，首先对数据进行预处理。所有工业小样本图像被统一调整为三通道，并设定了统一的大小。其中，2000 个工件图像被用作训练集，而 1000 个工件图像则作为测试集。为了更高效地训练网络模型，本次实验所选的 Batch Size 大小为 1，网络模型的学习率在前 100 个 epoch 中为 2×10^{-4} ，在之后的 epoch 中，学习率则线性递减到 0。此外，损失函数 $L_{\text{cyc}}(G, F)$ 的超参数 A 被设置为 10。为加快模型数据的加载速度，设定实验数据加载器的 num workers 数量为 4。整个模型最终训练了 500 个 epoch。在经过第 300 个 epoch 后，CycleGAN 模型在工件图像生成能力上的提升几乎停滞，这表明模型在此时已经达到了相对稳定的性能水平。

4.2 CycleGAN 模型改进

4.2.1 融入 CBAM 和 FPN 突出细节特征

为了提升网络对有效信息的捕捉能力，并增强对小型工件标志的检测精度，本文特别引入了轻量级卷积块注意力机制(Convolutional Block Attention Module, CBAM)，以优化网络模型，提高检测性能。CBAM 由两个子模块组成：通道注意模块(Channel Attention Module, CAM)和空间注意模块(Spatial Attention, SAM)，这两个模块协同工作，以实现特征的深度关注，如图 4-3 所示。

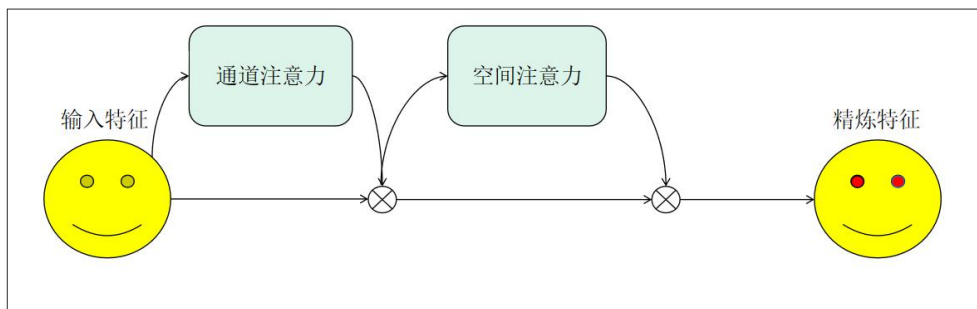


图 4-3 CBAM 示意图

CAM 首先通过一维卷积 $M_C \in R^{C \times 1 \times 1}$ 对特征进行处理，然后将处理后的特征与原始输入特征残差 $F \in R^{C \times H \times W}$ 相乘，实现特征的初步筛选与强化。随后，CAM 的输出结果将作为 SAM 的输入，SAM 进一步通过二维卷积 $M_S \in R^{1 \times H \times W}$

提取空间维度上的特征信息，并与残差相乘，完成第二次的特征细化提取。通过这种双重特征细化的方式，CBAM 能够有效地突出关键特征，进而提升模型的性能。其整体过程描述为式(4-1)、式(4-2)所示。

$$F' = M_C(F) \otimes F \quad (4-1)$$

$$F'' = M_C(F') \otimes F' \quad (4-2)$$

CenterNet 模型采用了一种多层次特征提炼策略，首先通过在其主干网络结构基础上施加连续三次反卷积运算过程，从而从高层特征映射中生成分辨率为 128×128 且包含 64 个特征通道的中间表征。此特征层被进一步输入至独特的检测子模块中，用于执行关键点定位预测任务。尽管深层特征在网络架构中承载了丰富的语义含义，但随着网络层次的逐渐加深，特别是对于细粒度目标如小型工件标志的局部纹理及微细结构信息，在传播过程中会遭受一定的损失。

鉴于此，在本研究中借鉴了 FPN 的网络构建，创造性地整合了 ResNet50 架构中的跨层特性。具体措施是对 ResNet50 内部的 ID Block 层特征进行了适应性通道调整，并将其作为 FPN 多尺度特征金字塔的基础组成部分。接着，通过上采样流程，将这些深层次特征逐级放大并与较低层的高分辨率特征实现无缝拼接和融合。

经过三层递进式的上采样与特征聚合机制，可以有效地弥合深层语义特征与浅层空间细节之间的差距，进而显著提升对小尺寸标志物的检测精度及减少遗漏检测的情况发生^[72]。

4.2.2 改进 U-Net 生成器提升细节效果

CycleGAN 模型的生成器网络基于 ResNet 残差结构构建，ResNet 残差结构的引入，不仅有效缓解了深度神经网络中常见的梯度消失问题，还显著加速了训练过程的收敛速度。通过允许信息在不同层级之间直接流动，残差结构使得模型能够更好地捕捉和保留关键特征信息，这种结构为模型带来了显著的性能提升。每个残差块包含两个卷积层和一个跳跃连接，使得输入数据能够直接跳过中间的一些计算步骤，直接加至输出，这种设计极大地简化了学习任务并加速了梯度传播，使得模型在训练过程中更加高效。通过卷积操作匹配维度，确保张量相加无误。这种结构有效缓解了梯度消失问题，加速了训练收敛，提升

了模型性能。单个残差块结构如图 4-4 所示。

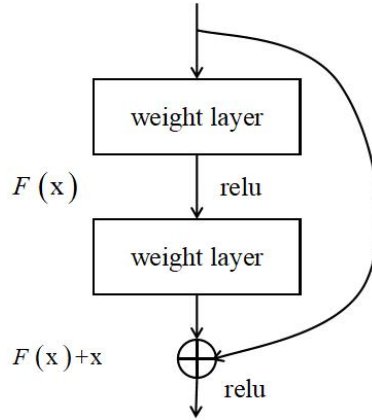


图 4-4 单个残差块结构图

在图 4-4 中， x 代表输入， $F(x)$ 代表输出，其中 $F(x)$ 的如式(4-3)：

$$F(x) = W_2 \sigma(W_1 x) \quad (4-3)$$

式中， W_1 和 W_2 代表权重参数， σ 表示 relu 激活函数。通过和跳跃连接传递的进行叠加操作后，最终的输出变为 $\sigma(F(x) + x)$ 。

在 CycleGAN 模型的原始设计中，生成器网络的核心组件采用了 ResNet 残差结构。该结构虽在图像风格迁移任务中表现较好，但仍存在一些如背景颜色失真和纹理细节难以处理的局限性：相比之下，U-Net 结构在图像生成任务中展现出了其独特的优势。长跳跃连接结构有助于保留更多的图像有效信息，这种设计使得 U-Net 在处理局部信息和重建细节信息方面表现出色，从而能够生成更加逼真和清晰的图像。此外，U-Net 采用反卷积层进行上采样，进一步增强了其在图像生成任务中的性能。因此，在处理具有复杂纹理的图像迁移任务时，U-Net 结构可能是一个更为合适的选择。ResNet 网络原理图与 U-Net 网络原理图对比如图 4-5 所示。

原始 U-Net 结构在特征提取方面存在局限性，其 10 层深度可能不足以充分捕获图像的高阶特征，因此在 CycleGAN 中直接应用可能效果不佳。为了改进这一点，本文提出了 C-U-Net 生成器结构。该结构通过加深网络深度并融合 Convnext 卷积结构，显著增强了特征提取能力，从而优化了图像风格迁移的效果。同时，C-U-Net 还注重平衡计算效率和性能，确保在实际应用中能够取得良好的综合表现。

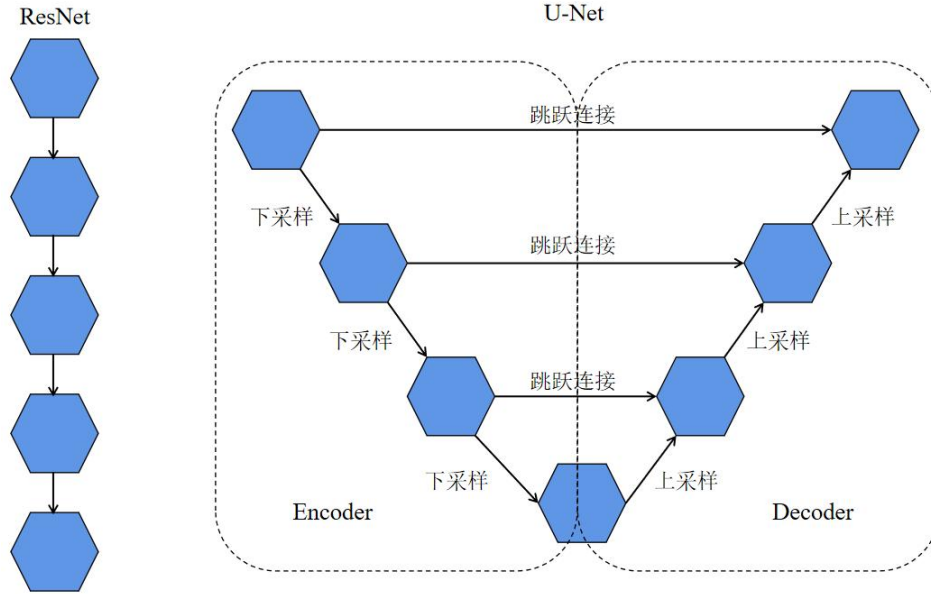


图 4-5 ResNet 网络与 U-Net 网络原理图对比

在 C-U-Net 中，部分传统卷积层升级为 Convnext 卷积结构，从而显著提升了特征提取效能。Convnext 不仅继承了 Swin-Transformer 的精髓，还对 ResNet 的残差结构进行了创新优化，实现了性能飞跃。这使得 C-U-Net 在图像风格迁移等任务中表现出色，提升了模型整体性能。

相较于 ResNet 的残差结构，Swin-Transformer 的基本构件块具有以下特点：首先，它将原始输入数据的通道转换为一种创新的“编码——解码”模式，通过通道扩展和缩减，实现对输入信息的深度抽取与高效重建。其次，在数据预处理阶段，Swin-Transformer 前置了具有更大感受野的处理层，以便在早期阶段捕捉更广阔的空间上下文信息。最后，为了优化模型结构和提升性能，Swin-Transformer 采取了精简策略，减少了不必要的激活函数和归一化层的数量。这些特点共同增强了 C-U-Net 的特征提取能力，使其在图像风格迁移等任务中取得优异表现。

此外，特别引入了多层感知机层（Multi-Layer Perceptron, MLP）作为滤波器机制的一部分，这种改进使得模型在特征表达和抽象能力上得到了强化。MLP 层是神经网络中常用的隐藏层组件，由多个相互连接的感知机组成。感知机作为基本单元，由神经元构成，能接收并处理多个输入信号，输出处理结果。在 MLP 层中，感知机通过线性操作连接形成线性层。为优化性能，Convnext 引入 GeLU 激活函数和层归一化技术，处理线性层数据，提升网络性能与稳定性。

如图 4-6 所示为 Convnext 模块图。

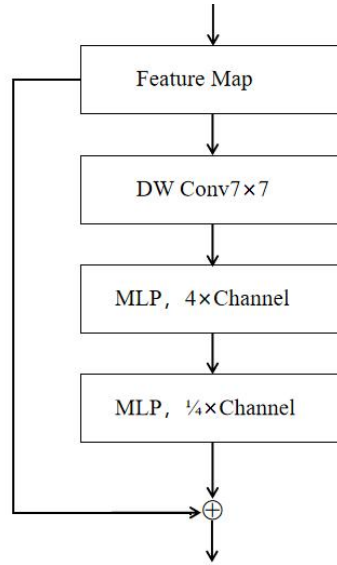


图 4-6 Convnext 模块图

Convnext 进行了结构化调整：其核心在于采用 7×7 卷积核替代传统的标准卷积核。为适配 MLP 层，需对特征图进行转置操作。MLP 层通过调整通道数，实现类似通道注意力的机制，提升信息提取与整合效率。这种调整使 Convnext 成为融合卷积与 MLP 的基础模块，显著增强网络在特征提取和学习方面的能力。鉴于 CycleGAN 模型参数多、计算复杂，全面采用 Convnext 可能导致计算时间过长。因此，在构建 CycleGAN 的 U-Net 结构时，保留用于下采样和上采样的部分原始卷积结构，而 U-Net 中的其他卷积层则采用 Convnext 结构，以利用其强大的特征提取能力，同时避免计算负担，提高训练效率。

图 4-7 为原始 U-Net 结构生成器，图 4-8 为 C-U-Net 生成器网络结构图，可以直观观察到改进带来的变化和潜在优势。通过这种混合型设计，既能充分利用 Convnext 的优势提升计算效率，又能维持 CycleGAN 在无配对样本图像转换任务上的优秀表现。

U-Net 结构凭借独特的编解码器架构和跳跃连接，有效地提取并融合不同层次的图像信息，助力生成器网络学习源域和目标域图像间的复杂映射关系。Convnext 卷积结构，融合 Swin-Transformer 设计思想，既保留卷积操作的局部特征提取优势，又引入自注意力机制捕捉全局上下文信息，从而在特征提取上展现更强能力，提取更丰富、具代表性的特征。

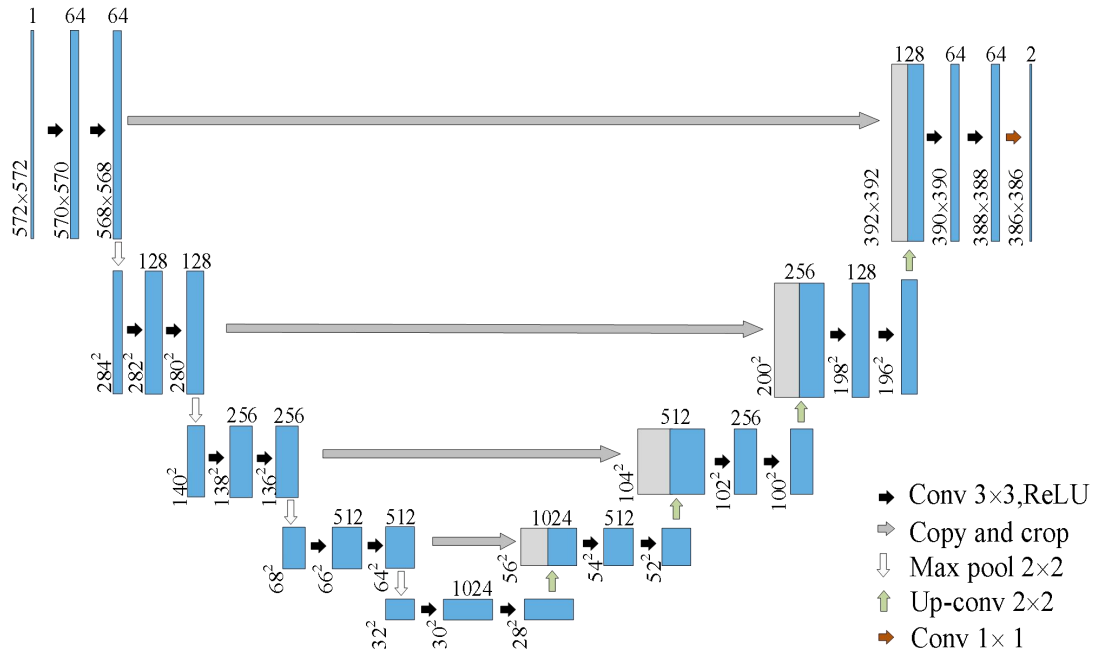


图 4-7 U-Net 结构图

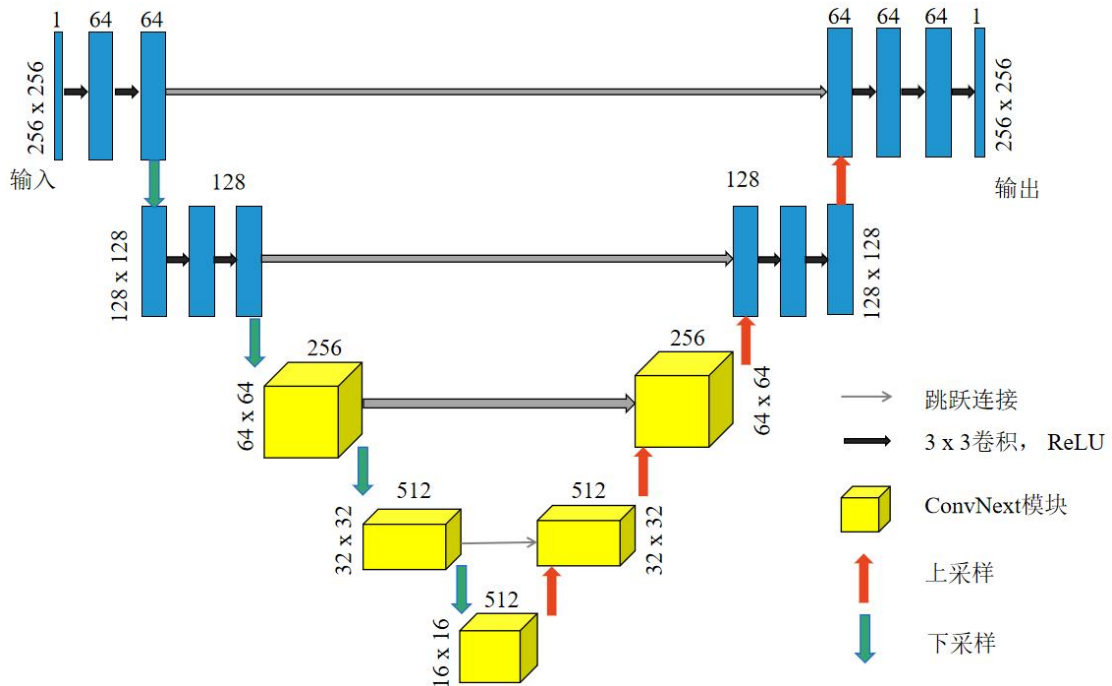


图 4-8 C-U-Net 生成器网络结构图

将 C-U-Net 结构应用于 CycleGAN 模型的生成器中，可以显著提高模型的图像风格迁移能力。这种组合结构能够充分发挥两者在特征提取和映射学习方面的优势，使得生成器能够更准确地捕捉到源图像和目标图像之间的风格差异，并生成出质量更高、更自然的风格迁移图像。

4.2.3 PatchGAN 增强判别能力

判别器模型采用 PatchGAN 结构，整体架构借鉴了全卷积神经网络的设计理念，整个模型的核心是运用连续的卷积操作，以逐步深化地抽取图像的内在特征信息。当特征提取完毕后，进一步将所得的特征图展开为一维特征向量以完成特征输出。

相较于原始 GAN 网络仅输出一个反映输入样本真实程度的实数值，即样本被判定为真实样本的概率，PatchGAN 则选择将输入空间映射到一个二维矩阵，并基于此计算该矩阵与相应标签矩阵之间的均方误差。这样调整的优势在于：常规 GAN 的单一数值输出方式可能过于绝对，无法细致地评价整幅图像的质量。而 PatchGAN 通过比较二维矩阵中每个像素点的差异并取均值得出评估结果，这种方式使训练过程更加侧重于图像局部细节的学习和判别，从而提升了模型的精细化识别能力。判别器网络结构如下图 4-9 所示：

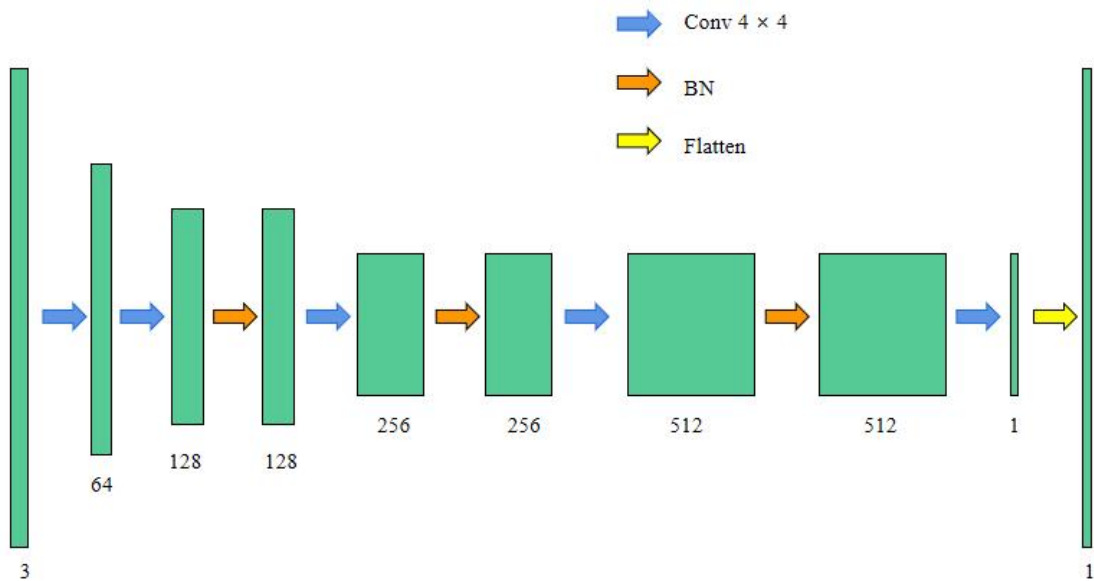


图 4-9 CycleGAN 判别器网络结构

4.2.4 改进 CycleGAN 的工业样本图像生成实验

本节实验在 Windows 10 操作系统环境下进行的，具体的实验环境配置和参数详情如表 4-1 所示。

表 4-1 实验环境

Parameter	Value
Pytorch	1.10.2
CUDA	11.3
GPU	RTX 3060, 12GB
CPU	Intel i7-10700
Image Size	1024 × 1024
Python	3.8

4.2.5 改进 CycleGAN 的工业样本图像质量评价结果

为了进一步说明本文改进的 CycleGAN 模型针对工业小样本的图像增强的优越性，在实验过程中给出了各种改进模型的 PSNR 和 SSIM 两个图像质量评价指标进行对比评价。

由下表 4-2 可知，本文提出方法在 PSNR 和 SSIM 值均为最大。相较于原始 CycleGAN 模型在 PSNR 和 SSIM 两项指标上分别提升了 4.382dB 和 0.102。这说明本文算法改进后图像与真实图像更加接近。

表 4-2 各算法图像评价指标对比

算法	PSNR/dB	SSIM
原始 CycleGAN	33.732	0.695
CycleGAN+CBAM	35.234	0.715
CycleGAN+ C-U-Net	35.742	0.732
CycleGAN+CBAM+PatchGAN	36.183	0.728
My-Method	38.114	0.797

4.3 本章小结

本章主要针对 CycleGAN 进行一些改进。该部分基于 CycleGAN 图像生成框架，结合 CBAM 注意力机制，致力于增强特征信息提取能力。在此基础上，通过对生成器进行改进，提高生成样本质量，精细化生成样本细节。同时，通过改进判别器来提高网络的判别性能。这一综合性的改进框架旨在优化工业小样本的生成过程，从而更好地满足实际应用的需求，为工业图像处理领域的发展提供新的思路和方法。

第 5 章 实验验证与分析

在第 3 章中，采用两种方式进行虚拟图像现实化，以扩充数据集。首先，对原始图像采用传统图像增强方法生成虚拟样本，再利用 CycleGAN 将这些虚拟样本从虚拟域转换为源域，实现虚拟图像现实化。其次，通过虚拟引擎建模生成虚拟样本，再利用 CycleGAN 将其与原始图片进行域转换，实现虚拟图像现实化。在第 4 章中，针对 CycleGAN 进行改进，并应用于工业小样本。本章将利用目标检测网络验证扩充后的数据集对目标检测的提升效果。

具体实验对比流程图如图 5-1 所示。

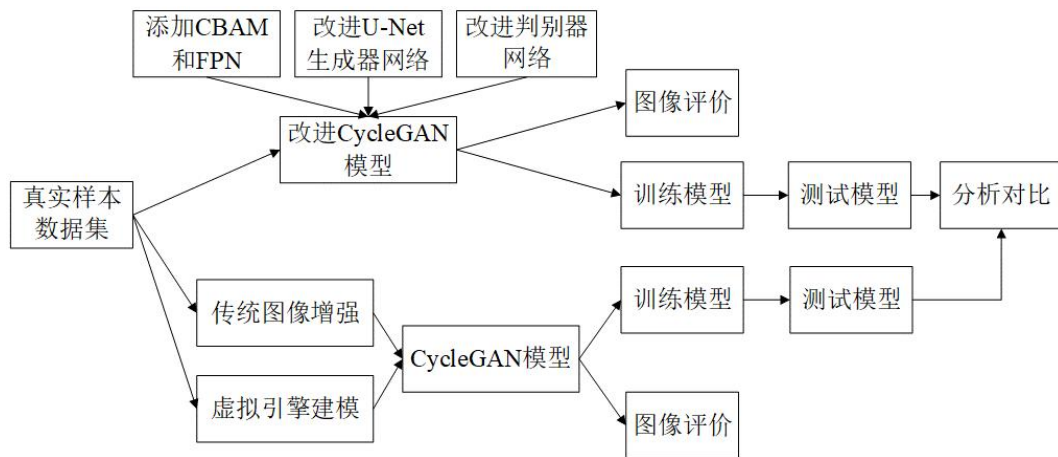


图 5-1 实验对比流程图

5.1 基于传统图像增强与 CycleGAN 结合的图像增强方法

5.1.1 改善传统图像增强方法的目标检测网络验证实验

本文实验是在 Windows10 系统下进行的，采用 Pytorch1.10.2 深度学习框架，CUDA 版本为 11.3。硬件配置为 NVIDIA GeForce RTX 3060,12GB 显存，Intel(R) Core(TM)i7-10700 CPU@2.90GHz，16GB RAM。训练过程中参数设置如下：输入大小为 640×640；采用 SGD 优化器，初始学习率为 0.01，权重衰减系数为 0.005。

在本节研究中，实验采用了一个涵盖了四种不同工件类别的数据集，其具体类别如图 5-2 所示。该数据集融合了真实的原始样本与应用本文所阐述的图

像增强技术生成的虚拟样本，共同构成了实验所使用的混合数据集。

为了确保模型训练、验证及测试阶段的有效性和公正性，该数据集遵循 6:2:2 的比例划分原则，分别对应形成了训练集、验证集以及测试集。对于训练集部分，采取了文中所介绍的方法来创建虚拟样本数据子集，以此增强目标检测网络性能。实验共制备了 2400 张各类型工件的增强图像，利用 labelImg 标准图像标注工具对所有工件图像进行详细的标注处理，并将这些标注信息转化成 XML 格式的标注文件，为深度学习模型训练提供完整且精准的标注支撑。

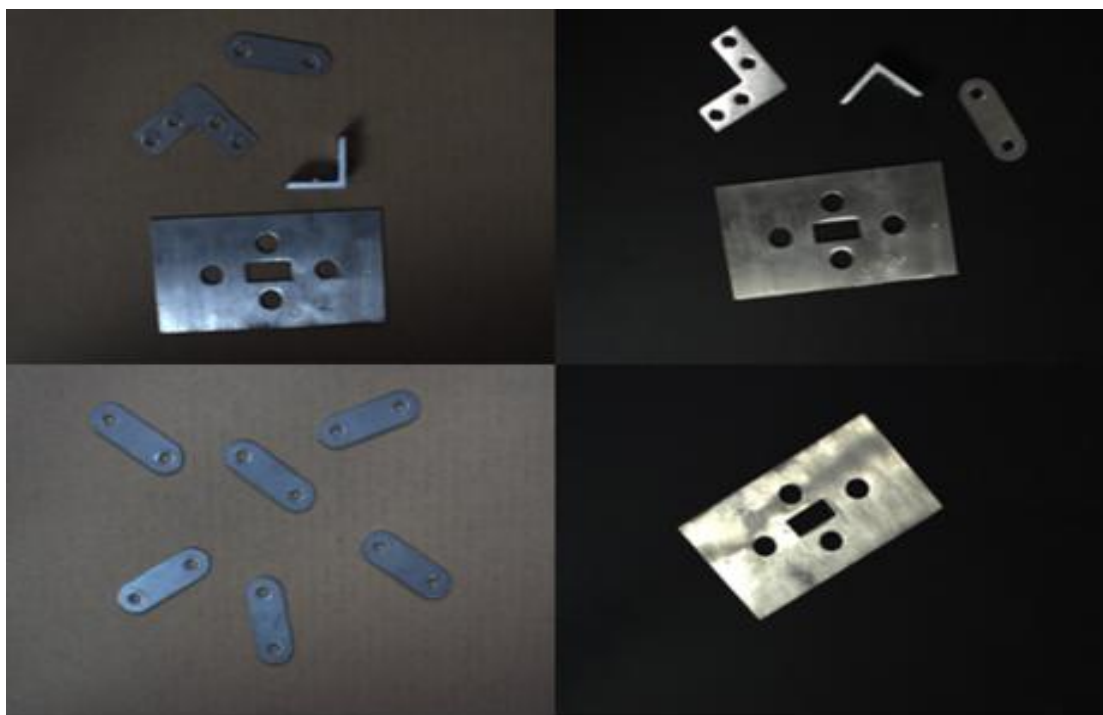
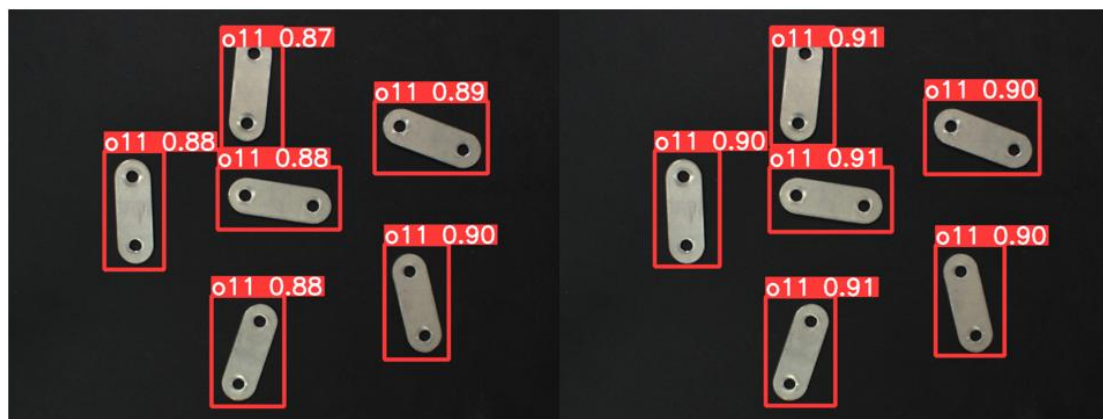


图 5-2 部分工件数据集展示

5.1.2 改善传统图像增强方法的实验分析

在实际的工件检测流水线上，工作台的工件包括相同类型和不同类型的工件，种类繁多且复杂。因此，为验证本文所提出的虚拟样本扩充方法对工件目标检测的有效性，实验阶段使用 YOLOv5 和 YOLOv7 这两个主流的深度学习目标检测网络在原有的少量工件样本数据集和增加虚拟样本的总工件数据集上，随机选取同种类别工件样本和不同种类别工件样本进行混合检测，如图 5-3 所示为同种类别工件样本使用 YOLOv5 进行检测的对比图，如图 5-4 所示为不同种类别工件样本使用 YOLOv7 进行检测的对比图。



(a) 增强前检测概率

(b) 增强后检测概率

图 5-3 同类工件检测结果对比图



(a) 增强前检测概率

(b) 增强后检测概率

图 5-4 多类工件检测结果对比图

为了能够综合评估使用本文所提到的图像增强方法得到的虚拟样本对工件目标检测的有效性，由以上消融实验，分别得到 yolov5 和 yolov7 两种目标检测网络在 4 类工件的均值平均精度 mAP(Mean Average Precision)。mAP 为模型的测评指标，mAP 的值越高表示检测精度越高。最终各模型检测结果如表 5-1 和图 5-5 所示。

表 5-1 各模型的检测结果

	mAP@.5:.95			
	YOLOv5		YOLOv7	
	原小样本训练数据集	新样本训练数据集	原小样本训练数据集	新样本训练数据集
工件 1	0.759	0.782	0.779	0.804
工件 2	0.761	0.783	0.771	0.803
工件 3	0.760	0.782	0.780	0.812
工件 4	0.760	0.784	0.783	0.819

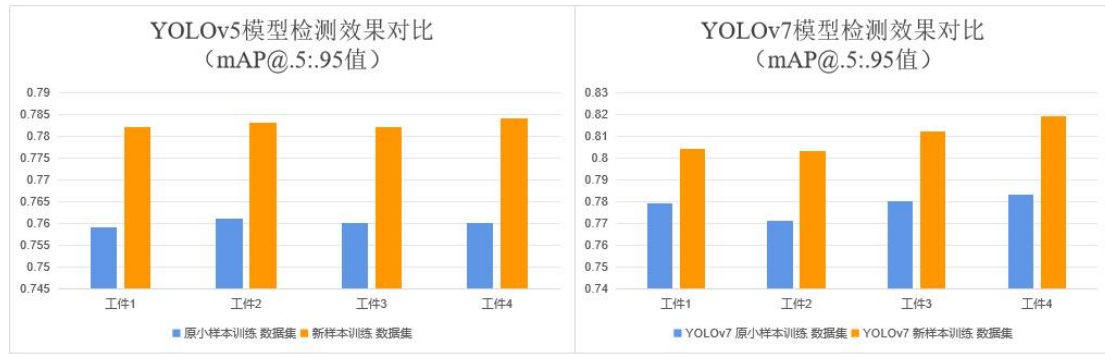


图 5-5 各模型的检测结果示意图

由表 5-1 和图 5-5 可以看出，目标检测网络经过增加虚拟样本的新样本训练数据集训练之后，mAP 均有所提高。其中，yolov5 平均提升了 2.28%，yolov7 平均提升了 3.12%。显然，通过对比实验可知，工件小样本问题情况下，使用了本文图像增强的方法在主流的深度学习目标检测网络上效果更佳。

综上所述，本节提出的方法可以有效地增加工件样本数量和质量，进而提升深度学习目标检测网络的性能，从而改善工件小样本目标检测的问题。该方法节省大量人工成本，为日后工业目标检测领域提供了新思路。

5.2 基于虚拟引擎与 CycleGAN 结合的图像增强方法

5.2.1 虚拟引擎与 CycleGAN 结合方法的目标检测网络验证实验

由于本部分生成的虚拟样本是经过虚拟引擎和随机算法完成的，在深入研究其如何影响数据模型性能之前，必须首先验证这些虚拟样本的有效性，从而确保它们在目标检测中的表现与实际样本保持一致。

本实验环境如下表 5-2 所示。

表 5-2 实验环境

Parameter	Value
Blender	2.9
Image Size	1280 × 1024
Pytorch	1.10.2
Python	3.8
CUDA	11.3
GPU	RTX 3060, 12GB
CPU	Intel i7-10700

本研究在单阶段深度学习目标检测模型 YOLO 上进行实验，以 YOLOv5 和 YOLOv7 为验证网络。将 6 种工件的实际样本按照 7:3 的比例划分为训练样本

和测试样本，为了进一步验证虚拟样本的引入对目标检测模型性能的影响，试验过程中依次向训练样本中增加 200 组、400 组和 600 组虚拟样本。在依次放入虚拟样本后，需重新训练模型并观察其检测模型的检测准确度变化，以评估虚拟样本对模型性能的影响。实验结果如下图 5-6 所示。



图 5-6 不同工件引入不同数目虚拟样本后模型的 mAP@.5:.95 结果对比

5.2.2 虚拟引擎与 CycleGAN 结合方法的实验分析

由上述实验结果可以看出，在引入的虚拟样本数量为 200 张、400 张和 600 张后，相较于未加入虚拟样本时 YOLOv5 模型对应的 mAP@.5:.95 值分别平均提升为：2.06%、3.13%和 3.98%，总体平均提升 3.06%；YOLOv7 模型对应的 mAP@.5:.95 值分别平均提升为：1.28%、1.93%和 3.47%，总体平均提升和

2.23%。通过以上实验分析可以看出，随着训练样本中虚拟样本数量的增加，目标检测模型的 $mAP@.5:.95$ 值不断升高，证实了虚拟样本的有效性。

5.3 基于改进 CycleGAN 的图像增强方法

5.3.1 改进 CycleGAN 图像增强方法的目标检测网络验证实验

本研究在单阶段深度学习目标检测模型 YOLO 上进行实验，以 YOLOv5 和 YOLOv7 作为验证网络。和上一小节实验方法一致，在引入不同数目的经过改进 CycleGAN 模型生成的工件虚拟样本后，对目标检测网络在 $mAP@.5:.95$ 上的结果进行了比较，具体对比结果如下图 5-7、图 5-8、图 5-9 和图 5-10 所示。



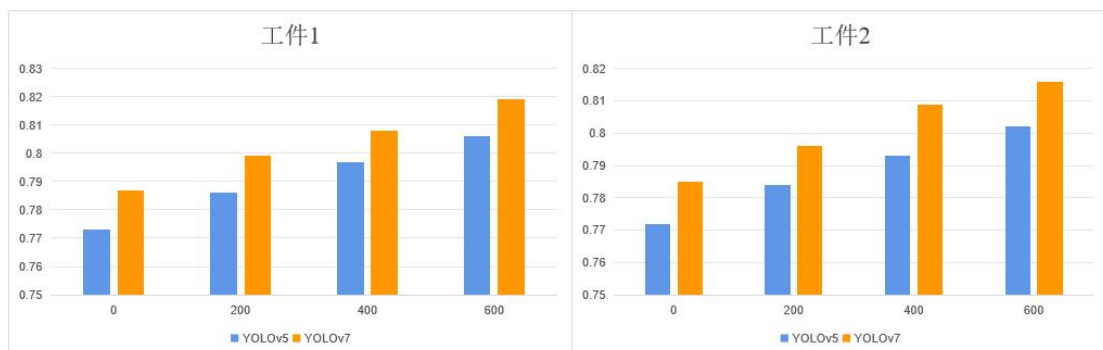
图 5-7 CycleGAN+CBAM 生成的虚拟样本在目标检测网络中的效果

如下图 5-8 所示为使用 C-U-Net 生成器的改进 CycleGAN 网络生成的虚拟样本在目标检测网络中的效果对比图。



图 5-8 CycleGAN+C-U-Net 生成的虚拟样本在目标检测网络中的效果

图 5-9 为集成 CBAM 模块和 PatchGAN 判别器的改进版 CycleGAN 网络生成的虚拟样本在目标检测任务中的显著效果。这一对比有力验证了所提出的改进措施对于提高目标检测网络性能的有效性。



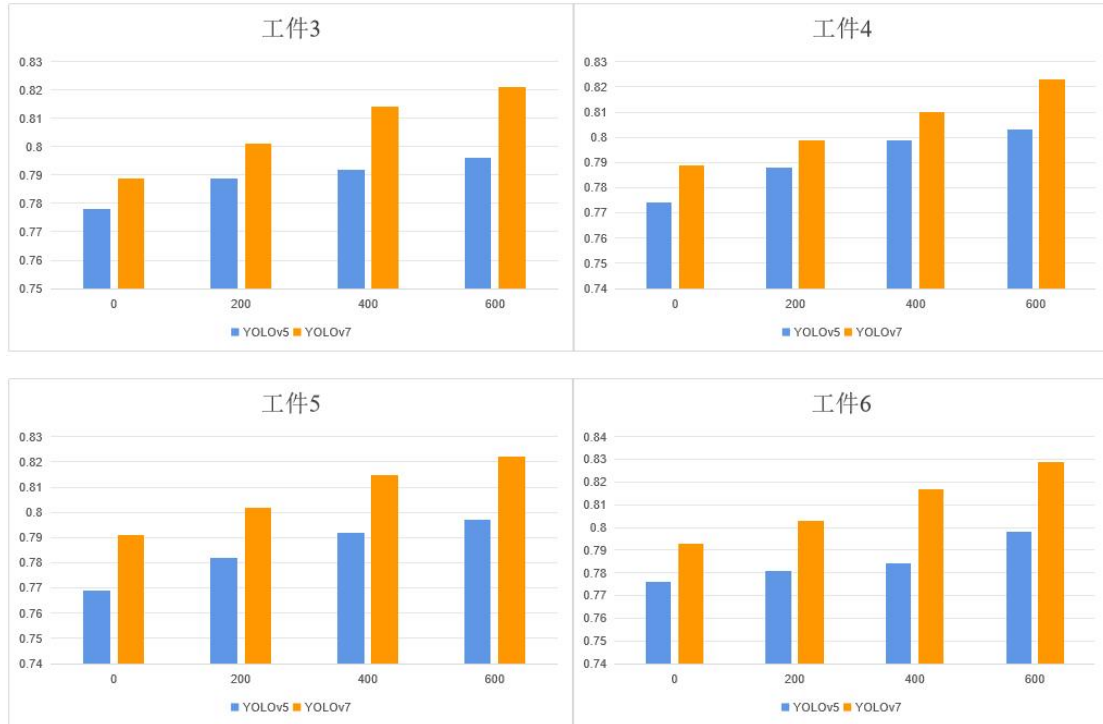
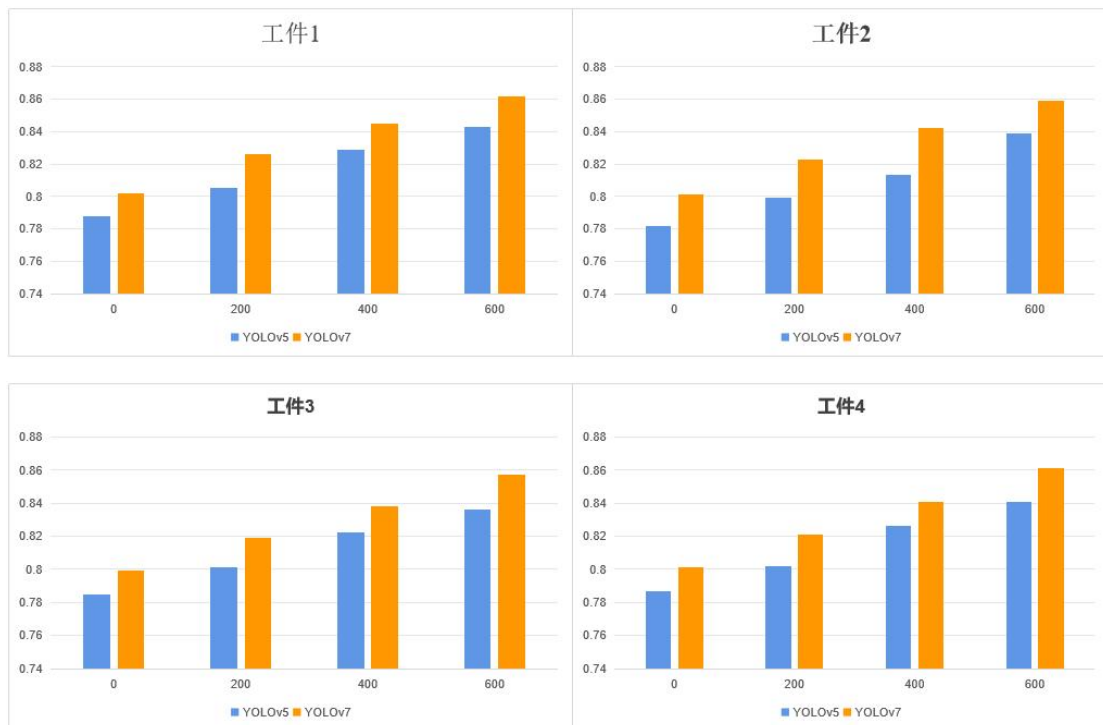


图 5-9 CycleGAN+CBAM+PatchGAN 生成的虚拟样本在目标检测网络中的效果

如图 5-10 所示，本文所提出的改进 CycleGAN 模型在生成应用于目标检测任务的虚拟样本方面的优秀性能。通过整合 CBAM 注意力模块、C-U-Net 生成器和 PatchGAN 局部判别网络，该改进模型能够有效地提升合成图像的质量及其在目标检测算法中的适用性，从而证实了该改进方案对于提升目标检测网络训练效果的有效性。



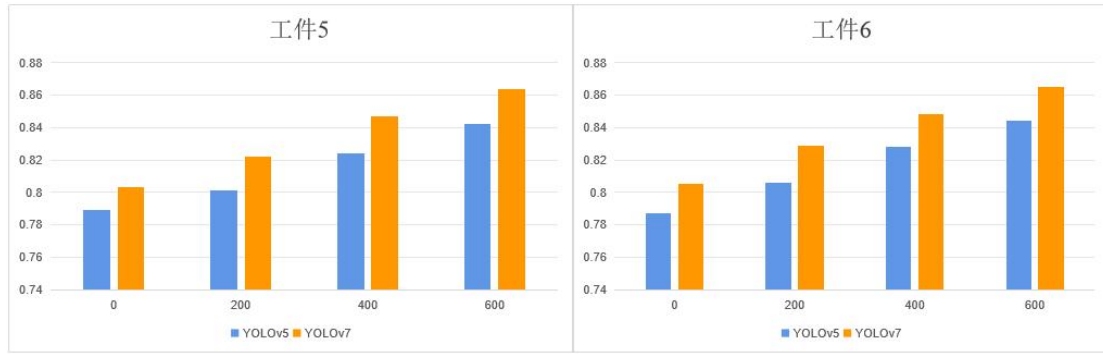


图 5-10 My-Method 生成的虚拟样本在目标检测网络中的效果

5.3.2 改进 CycleGAN 图像增强方法的实验分析

由上述实验结果可以看出，在引入的虚拟样本数量为 200 张、400 张和 600 张后，四种方法相较于未加入虚拟样本时 YOLOv5 和 YOLOv7 模型对应的 $mAP@.5:.95$ 值分别平均提升的结果如下表 5-3 所示。

表 5-3 四种方法对 YOLO 模型对应的 $mAP@.5:.95$ 值的平均提升情况

	CycleGAN+ CBAM		CycleGAN+ C-U-Net		CycleGAN+CBAM+ PatchGAN		My-Method	
	YOLO v5	YOLO v7	YOLO v5	YOLO v7	YOLO v5	YOLO v7	YOLO v5	YOLO v7
200	1.18%	1.10%	1.28%	1.18%	1.13%	1.10%	1.60%	2.15%
400	2.83%	2.22%	2.07%	1.72%	1.92%	2.32%	3.73%	4.17%
600	4.00%	3.18%	3.33%	3.20%	2.67%	3.27%	5.45%	5.95%

通过以上实验分析可以看出，随着训练样本中虚拟样本数量的增加，目标检测模型的 $mAP@.5:.95$ 值不断升高。本文提出的方法在提升效果上表现最为显著，相较于未加入虚拟样本时 YOLOv5 模型对应的 $mAP@.5:.95$ 值分别平均提升为：1.60%、3.73%和 5.45%，总体平均提升 3.59%；YOLOv7 模型对应的 $mAP@.5:.95$ 值分别平均提升为：2.15%、4.17%和 5.95%，总体平均提升 4.09%。证实了所提出的方法的有效性。

5.4 本章小结

本章对本文提出的图像增强方法进行了实验验证。通过使用目标检测网络，验证了本文提出的方法在增加虚拟样本的情况下对目标检测网络性能的提升效

果。实验结果显示，本文方法在提高目标检测网络性能方面具有一定的有效性。这些结果表明，本文提出的方法不仅在图像质量方面表现出色，而且对于增强目标检测网络性能也具有潜力和实用性。随着虚拟样本技术的不断发展和完善，有望进一步提升目标检测模型的性能，从而更好地应对实际场景中的复杂情况，为计算机视觉领域的发展做出更大的贡献。

第 6 章 总结与展望

6.1 总结

本文研究面向工业小样本的图像增强技术，首先分析了工业小样本图像生成的发展现状，发现目前有的工件在人工收集的情况下存在很大的局限性，脚本生成的数据不具有代表性，且目前工件定制化比较多、应用场景复杂，工件图像的收集越来越难，对工业小样本的检测和识别研究造成很大的影响，体现了本文研究工业小样本图像生成形成近真实、近完备的工件数据库的必要性。主要工作有：

(1)基于样本多样性的工业样本图像生成方法。分别从传统图像增强和 CycleGAN 相结合的方法与虚拟引擎制作工件建模和 CycleGAN 相结合的方法，丰富样本的多样性；

(2)为了提升网络性能，使得生成样本效果更好，从而改进 CycleGAN 网络，该方法对进行工业样本图像生成具有一定的先进性。首先两个训练集不需要成对，意味着可以大大减少人力分类和标注，其次改进生成器和判别器可以循环抑制生成假图像，相较于原始 CycleGAN 还能减少模式崩溃问题。

(3)针对上述两种方式生成的样本数据集，在目标检测网络上进行不同生成方法的工件样本数据集分组验证，实验证明相较于原始小样本数据集，提出的方法效果更好。

此研究解决了真实场景下的工业小样本图像数量少、获取困难、标注耗费人力物力等问题。根据虚拟引擎的生成方法可以提前设定工件的种类和各项参数，可以模仿真实场景和类别不平衡进行调配，生成更有利于工业小样本目标检测的训练样本。这样对于不同实验不同效果都可以得到相应的数据集和训练网络，体现了此研究的先进性。

6.2 展望

本文针对工业小样本图像增强提出了新的工件生成方法，用传统图像增强

方法、虚拟引擎建模和深度学习方法对于工件样本数据集进行了扩充。但受限用于训练的工件样本数据量多，网络模型较大，分析起来较为困难，耗费大量时间和存储空间等约束条件，为此提出如下展望：设计一种基于工业小样本的图像增强系统。

基于工业小样本的图像增强系统可以有效对多场景、多样本进行集成，并加强对多样本工业场景的适应性，且结合迁移学习等技术提高系统性能。系统中的工件数据库，可以根据不同生成方法分为各个典型的子数据集，这些子数据集可以提供不同需求来训练模型和检测识别实验，处理各种不同情况下的问题子数据集包括真实的工件图片、虚拟生成的工件图片、以及对抗样本，还可以在每个子数据集下分子类进行分类研究，也可以用整个数据库进行全面测试。这样不仅可以为工件研究提供一个标准来进行评测，也可以加固工件识别算法，实现更高精度的工件识别，使工业小样本数据集更加丰富，从而使得工业小样本目标检测更加精准。

参考文献

- [1] 王浩, 张叶, 沈宏海, 等. 图像增强算法综述[J]. 中国光学, 2017, 10(4): 438-448.
- [2] 梅菠萍. 小规模数据集工件检测方法研究与实现[D]. 四川: 西南科技大学, 2023.
- [3] 朱晓慧, 钱丽萍, 傅伟. 图像数据增强技术研究综述[J]. 软件导刊, 2021, 20(5): 230-236.
- [4] 徐惠灵, 尚政国, 董胜波, 等. 面向深度神经网络应用的小样本学习技术研究[J]. 南京航空航天大学学报, 2022, 54(S01): 80-86.
- [5] 赵凯琳, 靳小龙, 王元卓. 小样本学习研究综述[J]. 软件学报, 2021, 32(2): 349-369.
- [6] 陈良臣, 傅德印. 面向小样本数据的机器学习方法研究综述[J]. 计算机工程, 2022, 48(11): 1-13.
- [7] Deng W, Zheng L, Ye Q, et al. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, Salt Lake City, UT, USA, 18-23 June, 2018: 994-1003.
- [8] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//Proceedings of the IEEE international conference on computer vision. IEEE, Venice, Italy, 22-29 October, 2017: 2223-2232.
- [9] Xia P, Niu H, Li Z, et al. Enhancing backdoor attacks with multi-level mmd regularization[J]. IEEE Transactions on Dependable and Secure Computing, 2022, 20(2): 1675-1686.
- [10] 安广琳, 李宗刚, 杜亚江, 等. 基于深度学习的多工件抓取点定位方法[J]. 激光与光电子学进展, 2023, 60(12): 311-321.
- [11] Li D, Zhang X, Jiang T, et al. Application of improved wavelet threshold denoising algorithm in work piece inspection[C]//2013 6th International Congress on Image and Signal Processing (CISP). IEEE, Hangzhou, China, 2013: 256-260.

- [12]Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector[C]//Computer Vision–ECCV 2016: 14th European Conference, Amsterdam. Springer International Publishing, Netherlands, 2016: 21-37.
- [13]李小丽. 基于深度学习的小样本目标检测算法研究[D]. 广东: 广州大学, 2022.
- [14]Poggio T, Hurlbert A. Observations on Cortical Mechanisms for Object Recognition and s Learning[J]. Massachusetts Institute of Technology, 1994, 6(1): 153-182.
- [15]张晓晗. 基于机器学习的工业过程数据驱动建模及数据扩充方法研究[D]. 北京: 北京化工大学, 2023.
- [16]Chen Z S, Hou K R, Zhu M Y, et al. A virtual sample generation approach based on a modified conditional GAN and centroidal Voronoi tessellation sampling to cope with small sample size problems: Application to soft sensing for chemical process[J]. Applied Soft Computing, 2021, 101: 1-12.
- [17]Liu X, Sun Q, Meng Y, et al. Hyperspectral image classification based on parameter-optimized 3D-CNNs combined with transfer learning and virtual samples[J]. Remote Sensing, 2018, 10(9): 1-16.
- [18]Li L, Peng Y, Qiu G, et al. A survey of virtual sample generation technology for face recognition[J]. Artificial Intelligence Review, 2018, 50: 1-20.
- [19]Belkin M, Niyogi P. Laplacian eigenmaps for dimensionality reduction and data representation[J]. Neural computation, 2003, 15(6): 1373-1396.
- [20]Zhu Q X, Zhang X H, He Y L. Novel virtual sample generation based on locally linear embedding for optimizing the small sample problem: Case of soft sensor applications[J]. Industrial & Engineering Chemistry Research, 2020, 59(40): 17977-17986.
- [21]刘德平. 基于空间插值的虚拟样本生成技术应用研究[D]. 北京: 北京化工大学, 2023.
- [22]Pan S, Tian S, Wang X, et al. Comparing different spatial interpolation methods to predict the distribution of fishes: A case study of *Coilia nasus* in the Changjiang River Estuary[J]. Acta Oceanologica Sinica, 2021, 40: 119-132.

- [23]He Y L, Wang P J, Zhang M Q, et al. A novel and effective nonlinear interpolation virtual sample generation method for enhancing energy prediction and analysis on small data problem: A case study of Ethylene industry[J]. Energy, 2018, 147: 418-427.
- [24]Sutojo T, Rustad S, Akrom M, et al. A machine learning approach for corrosion small datasets[J]. npj Materials Degradation, 2023, 7(1): 1-10.
- [25]徐雯, 汤健, 夏恒, 等. 基于 Bagging 半监督深度森林回归的二噁英排放浓度软测量[J]. 仪器仪表学报, 2022, 43(6): 251-259.
- [26]Błaszczyński, J, Stefanowski J. Neighbourhood sampling in bagging for imbalanced data[J]. Neurocomputing, 2015, 150: 529-542.
- [27]Zhu, Q X, Gong H F, Xu Y, et al. A bootstrap based virtual sample generation method for improving the accuracy of modeling complex chemical processes using small datasets[C]//2017 6th Data Driven Control and Learning Systems (DDCLS). IEEE, Chongqing, China, 2017: 84-88.
- [28]Chernozhukov V, Chetverikov D, Kato K, et al. High-dimensional data bootstrap[J]. Annual Review of Statistics and Its Application, 2023, 10: 427-449.
- [29]Lin L S, Li D C, Chen H Y, et al. An attribute extending method to improve learning performance for small datasets[J]. Neurocomputing, 2018, 286: 75-87.
- [30]Li D C, Lin L S. A new approach to assess product lifetime performance for small data sets[J]. European Journal of Operational Research, 2013, 230(2): 290-298.
- [31]Li D C, Lin Y S. Using virtual sample generation to build up management knowledge in the early manufacturing stages[J]. European Journal of Operational Research, 2006, 175(1): 413-434.
- [32]张晓晗. 基于机器学习的工业过程数据驱动建模及数据扩充方法研究[D]. 北京: 北京化工大学, 2023.
- [33]Koa C H, Baeb J S, Yeumc J H, et al. Application of photoacoustic spectroscopy to spatially resolved, non-destructive measurements of organics distributed in zeolite membranes[J]. Journal of Ceramic Processing Research, 2009, 10: S9-S13.

- [34]Lin L S, Li D C, Yu W H, et al. Generating Multi-modality Virtual Samples with Soft DBSCAN for Small Data Set Learning[C]//2015 3rd International Conference on Applied Computing and Information Technology/2nd International Conference on Computational Science and Intelligence. IEEE, Okayama, Japan, 2015: 363-368.
- [35]Cai W, Ma B, Zhang L, et al. A pointer meter recognition method based on virtual sample generation technology[J]. Measurement, 2020, 163: 1-11.
- [36]He G, Yuan X, Zhang W, et al. An effective method for face recognition by creating virtual training samples based on pixel processing[C]//2018 10th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC). IEEE, Hangzhou, China, 2018: 177-180.
- [37]Luo J, Tjahjadi T. Multi-set canonical correlation analysis for 3D abnormal gait behaviour recognition based on virtual sample generation[J]. IEEE Access, 2020, 8: 32485-32501.
- [38]Tang J, Xia H, Aljerf L, et al. Prediction of dioxin emission from municipal solid waste incineration based on expansion, interpolation, and selection for small samples[J]. Journal of Environmental Chemical Engineering, 2022, 10(5): 1-13.
- [39]Chen Z S, Zhu Q X, Xu Y, et al. Integrating virtual sample generation with input-training neural network for solving small sample size problems: application to purified terephthalic acid solvent system[J]. Soft Computing, 2021, 25: 6489-6504.
- [40]Zhu J, Ge Z, Song Z, et al. Review and big data perspectives on robust data mining approaches for industrial process modeling with outliers and missing data[J]. Annual Reviews in Control, 2018, 46: 107-133.
- [41]Chen Z S, Zhu B, He Y L, et al. A PSO based virtual sample generation method for small sample sets: Applications to regression datasets[J]. Engineering Applications of Artificial Intelligence, 2017, 59: 236-243.
- [42]Li D C, Wen I H. A genetic algorithm-based virtual sample generation technique to improve small data set learning[J]. Neurocomputing, 2014, 143: 222-230.

- [43] Qiao J F, Guo Z H, Tang J. Virtual sample generation method based on improved megatrend diffusion and hidden layer interpolation with its application[J]. Journal of chemical industry, 2020, 71(12): 331-345.
- [44] He Y L, Wang P J, Zhang M Q, et al. A novel and effective nonlinear interpolation virtual sample generation method for enhancing energy prediction and analysis on small data problem: A case study of Ethylene industry[J]. Energy, 2018, 147: 418-427.
- [45] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [46] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, Columbus, OH, USA, 2014: 580-587.
- [47] 丁文龙, 费树珉. 基于改进 YOLOv3 的安全帽检测方法研究[J]. 电子测试, 2022, 11: 84-86.
- [48] 杨其睿. 基于改进的 DenseNet 深度网络火灾图像识别算法[J]. 计算机应用与软件, 2019, 36(2): 258-263.
- [49] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, Honolulu, HI, USA, 2017: 2117-2125.
- [50] 宋欢欢, 惠飞, 景首才, 等. 改进的 RetinaNet 模型的车辆目标检测[J]. 计算机工程与应用, 2019, 55(13): 225-230.
- [51] Li X, Jiang H, Liu S, et al. A unified framework incorporating predictive generative denoising autoencoder and deep Coral network for rolling bearing fault diagnosis with unbalanced data[J]. Measurement, 2021, 178: 1-12.
- [52] Li X, Wang W, Hu X, et al. Selective kernel networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. IEEE, Long Beach, CA, USA, 15-20 June, 2019: 510-519.

- [53]Cai Z, Vasconcelos N. Cascade r-cnn: Delving into high quality object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, Salt Lake City, UT, USA, 2018: 6154-6162.
- [54]Zhu C, He Y, Savvides M. Feature selective anchor-free module for single-shot object detection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. IEEE, Long Beach, CA, USA, 2019: 840-849.
- [55]Shankar T, Eappen G, Suresh V, et al. Contrast enhancement using quantile separation and bi-histogram equalization[C]//2019 Innovations in Power and Advanced Computing Technologies (i-PACT). IEEE, New Delhi, 2019, 1: 1-4.
- [56]Chen S D, Ramli A R. Minimum mean brightness error bi-histogram equalization in contrast enhancement[J]. IEEE transactions on Consumer Electronics, 2003, 49(4): 1310-1319.
- [57]Wang Y, Pan Z. Image contrast enhancement using adjacent-blocks-based modification for local histogram equalization[J]. Infrared Physics & Technology, 2017, 86: 59-65.
- [58]刘文韬. Blender 3D 渲染机制研究[D]. 南京: 南京邮电大学, 2016.
- [59]Villar O. Learning Blender[M]. Addison-Wesley Professional, 2021.
- [60]胡冰. 基于 Blender 的虚拟摄影机插件开发[J]. 现代电影技术, 2022(09): 46-52.
- [61]Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.
- [62]Zhu Q X, Hou K R, Chen Z S, et al. Novel virtual sample generation using conditional GAN for developing soft sensor with small data[J]. Engineering Applications of Artificial Intelligence, 2021, 106: 1-9.
- [63]崔琳琳, 沈冰冰, 葛志强. 基于混合变分自编码器回归模型的软测量建模方法[J]. 自动化学报, 2022, 48(2): 398-407.
- [64]Zhang L, Zhang H, Cai G. The multiclass fault diagnosis of wind turbine bearing based on multisource signal fusion and deep learning generative model[J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 1-12.

- [65] Wang X, Liu H. Data supplement for a soft sensor using a new generative model based on a variational autoencoder and Wasserstein GAN[J]. Journal of Process Control, 2020, 85: 91-99.
- [66] 郝飞, 史金飞, 张志胜, 等. 应用通用自回归模型实现图像的自适应滤波[J]. 光学精密工程, 2014, 22(1): 186-192.
- [67] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30: 1-16.
- [68] Wu P, Jiang L, Hua Z, et al. Multi-focus image fusion: Transformer and shallow feature attention matters[J]. Displays: Technology and Applications, 2023, 76: 1-5.
- [69] Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE transactions on image processing, 2004, 13(4): 600-612.
- [70] Hore A, Ziou D. Image quality metrics: PSNR vs. SSIM[C]//2010 20th international conference on pattern recognition. IEEE, Istanbul, Turkey, 2010: 2366-2369.
- [71] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//Proceedings of the IEEE international conference on computer vision. IEEE, Venice, Italy, 2017: 2223-2232.
- [72] 易星, 潘昊, 赵怀慈, 等. 基于改进循环生成对抗网络实现红外图像生成[J]. 电子测量技术, 2023, 46(18): 171-178.

攻读学位期间发表的学位论文及参加的科研项目

一、发表的学术论文

- [1] 以第一作者在第 42 届中国控制会议（CCC 2023）发表会议论文一篇（EI 收录）；
- [2] 以第一作者在第 2023 年中国自动化大会（CAC 2023）发表会议论文一篇（已发表）。

二、实审的专利

- [1] 以第二发明人（导师第一）申请与课题相关的一件发明专利，进入实审阶段。

致谢

花开花落万物道，聚散离别终有时，二十余载求学之路渐进尾声。何其有幸成为一名研究生，感谢党和国家的政策扶持，让一名满怀理想抱负的乡村女孩子有机会接受平等教育、有了改变命运的机会。回首这三年，感谢老师们的悉心教诲，师兄师弟师妹们的无私帮助，感谢家人无限支持。

首先，感谢我的导师江明教授，有幸在这三年里跟随导师学习，从督促我学习进度，到给我许多宝贵意见，都是我的人生道路及自身修养的巨大财富。感谢您的悉心指导，师恩当永记于心。

感谢万国扬老师对我研究课题专业内容的耐心指导，让我对图像增强有了更好的认知，给我的大论文和小论文都提出了宝贵意见。这段时间，给您添了很多麻烦，谢谢老师。感谢柏受军副教授，从入学到现在，关心着我们的研究生生活，督促着我们的课题进展，时时刻刻监督我们的学习状态，也为我的论文格式提出很多修改意见。感谢您的细心帮助，在这里由衷的说一声感谢!

感谢徐印赞、卢志远、吴云飞、刘富春、葛凌枫、赵朝朝、殷宗琨、贾凯、高宝林师兄们，张健、黄志远、刘鹏、周阳、陈金城同门，汪倩倩和何琴师妹对我学习和生活上的帮助，让我的研究生生活更加丰富多彩!同时也要感谢我的家人对我所做的一切的支持。

感谢好友小吴同学坚定我读研的决心，给予我各种支持，感谢你在我快要枯萎时接住了那个糟糕的我。感谢好友小翁同学，谢谢你让我认识自己，谢谢你让我明白我也是朵值得被爱的玫瑰。

最后，深深感谢我敬爱的祖父陶善华先生。亲人之教，得以成才。在周围环境充满着重男轻女的世俗喧嚣，您力排众议，鼎力支持我继续读书，成为家族中首位女研究生。感谢您坚定支持，给予我最大的温暖和爱的港湾，为我提供最深厚的情感寄托与精神支柱。您对我的偏爱是我成长路上最珍贵的力量，愿您在天堂能够看到孙女始终在努力进步，不断完善自我。谢谢您，永怀感激，永远怀念!

何其有幸，生于华夏!何其有幸，有你们在!

尚德藏学 唯实唯新

