

分类号: TP391

密 级: 公开

学校代码: 10363

学 号: 2220320152



安徽工程大学

Anhui Polytechnic University

硕士学位论文

题目 基于自监督学习的工业图像
异常检测与定位技术研究

论文作者 范伟平

校内导师 程凡永（副教授）

校外导师 刘海广

专业学位类别 电子信息

研究方向（领域） 智能信息处理与模式识别

论文提交日期: 2025 年 5 月 15 日

分类号: TP391.4

密 级: 公开

单位代码: 10363

学 号: 2220320152

题 目 基于自监督学习的工业图像异常
检测与定位技术研究

题 目 Research on Self-supervised Learning-based
Anomaly Detection and Location Technology in
Industrial Image

姓 名: 范伟平

专 业(领域): 电子信息

指 导 教 师: 程凡永（副教授）

校 外 导 师: 刘海广

研 究 方 向: 智能信息处理与模式识别

答 辩 日 期: 2025 年 5 月 12 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：范伟平

日期：2025 年 5 月 15 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：范伟平

指导教师签名：程永

日期：2025 年 5 月 15 日

日期：2025 年 5 月 15 日

基于自监督学习的工业图像异常检测与定位技术研究

摘 要

随着现代工业制造智能化水平的不断提升,基于深度学习的工业图像异常检测技术已成为工业产品质量检测的一种高效解决方案。深度学习方法主要分为有监督学习和自监督学习两类,有监督方法需要大规模标注样本作为支撑,且训练数据需具备精细的像素级标注,耗费大量人力和物力,难以满足大规模的质检需求。相比之下,自监督方法仅需正常样本进行训练即可检测异常区域,不仅能够解决有监督方法在面对复杂或微小缺陷时准确率不足的问题,还具备更强的泛化能力,具有训练速度快、检测效率高和广泛适用性等优势。因此,开展基于自监督学习的工业图像异常检测研究具有重要的现实意义,具体内容如下:

(1)针对工业图像表面异常检测中缺陷样本稀缺、漏检率高等问题,提出了一种基于自监督结构化重建的工业图像异常检测与定位模型,该模型仅需正常样本进行训练,即可实现异常区域的检测与定位。首先,模型基于自编码器重建输入图像,由于其泛化能力较强,异常区域也可能被较好地重建,导致输入图像与重建图像之间的重构误差减小,从而影响检测精度。然后,提出了一种多互补掩码融合方法,通过多角度遮挡,降低异常区域的重建可能性,同时保持正常区域的高质量重建,从而提高异常检测的准确性。最后,通过计算输入图像与融合重建图像之间的多尺度结构相似性损失,实现异常检测与定位。

实验结果表明,在图像级和像素级的 AUROC 评价指标下,该方法在 MVTec AD 数据集上的异常检测率和定位率分别达到 94.9%和 95.6%,相较于对比的最优方法,分别提升了 1.6%和 1.7%,具有更有效的检测性能。

(2)针对工业图像异常检测中缺陷样本不足、检测效率和精度不平衡等问题,设计了一种自监督深度记忆特征嵌入的工业图像异常检测与定位模型,该模型采用正常样本和模拟异常样本训练即可准确检测和定位异常区域。首先,通过模拟异常策略构造负样本,解决了异常样本稀缺的问题。然后,引入深度记忆特征嵌入模块,捕捉输入样本与记忆模块中正常样本之间的特征差异信息,并将其与输入样本特征拼接,从而增强模型对异常区域的推断能力。最后,设计了一种多尺度特征融合模块,加强了局部细节与全局上下文信息的有效结合,进一步提高了模型对多尺度目标的检测精度和鲁棒性。实验结果表明,在 AUROC 指标下,该方法在 MVTec AD 和 BTAD 数据集上的异常检测准确率比最高的对比方法分别提高了 0.7%和 4.9%;异常定位准确率分别提高了 3.2%和 2.7%,更适合当前工业图像异常检测任务的需求。

(3)基于以上研究,为了实现更快捷、高效的工业图像异常检测任务,设计了一种基于自监督学习的工业图像异常检测系统。首先,该系统设有用户登录窗口,以确保系统的安全性和数据的保密性。然后,系统构建了多个功能模块,包括模型训练与测试、工业图像异常检测以及检测结果等控件,方便用户直观了解异常检测的完整流程及操作

步骤。最后，实验结果表明，该系统能够快速、准确地识别工业产品中的缺陷，验证了其有效性和实用性。

综上所述，本文围绕基于自监督学习的工业图像异常检测方法展开研究，实现了对复杂纹理和工业零件工业图像的异常检测与定位。此外，该方法不仅可以判断工业图像是否存在缺陷，还可精确定位缺陷区域和尺寸，为工业产品质量评估提供了技术支持。

关键词：自监督学习，工业图像，异常检测，结构化重建，记忆特征嵌入

RESEARCH ON SELF-SUPERVISED LEARNING-BASED ANOMALY DETECTION AND LOCALIZATION IN INDUSTRIAL IMAGES

ABSTRACT

With the continuous advancement of intelligent manufacturing in modern industry, deep learning-based industrial image anomaly detection has emerged as an efficient solution for industrial product quality inspection. Deep learning methods can be broadly categorized into supervised and self-supervised learning. Supervised methods require large-scale annotated datasets with precise pixel-level labels, which demand significant human and material resources, making it challenging to meet the needs of large-scale quality inspection. In contrast, self-supervised methods only require normal samples for training to detect anomalous regions. These methods not only address the limitations of supervised approaches in achieving high accuracy for complex or subtle defects but also exhibit stronger generalization capabilities, faster training speeds, higher detection efficiency, and broader applicability. Therefore, conducting research on industrial defect detection based on self-supervised learning holds significant practical importance. The specific research directions are as follows:

(1) To address the challenges of scarce defect samples and high miss rates in industrial image surface anomaly detection, this paper proposes a self-supervised structured reconstruction-based model for anomaly detection and localization in industrial images. The model requires only normal samples for training and can achieve accurate detection and localization of anomalous regions. First, the model reconstructs the input image using an autoencoder. However, due to the strong generalization capability of the autoencoder, anomalous regions may also be well-reconstructed, leading to reduced reconstruction errors between the input and reconstructed images in anomalous areas, thereby affecting detection accuracy. To mitigate this issue, this paper introduces a multi-complementary mask fusion method, which reduces the likelihood of reconstructing anomalous regions through multi-angle occlusion while maintaining high-quality reconstruction of normal regions, thus improving the accuracy of anomaly detection. Finally, multi-scale structural similarity loss between the input image and the fused reconstructed image is computed to achieve anomaly detection and localization. Experimental results demonstrate that, under image-level and pixel-level AUC evaluation metrics, the proposed method achieves anomaly detection and localization rates of 94.9% and 95.6%, respectively, on the MVTec AD dataset. Compared to the best-performing baseline method, these results represent improvements of 1.6% and 1.7%, showcasing effective detection

(2) To address the challenges of insufficient defect samples and the imbalance between detection efficiency and accuracy in industrial image anomaly detection, this paper proposes a self-supervised deep memory feature embedding model for anomaly detection and localization in industrial images. The model can accurately detect and localize anomalous regions by training on normal samples and simulated anomaly samples. First, a simulated anomaly strategy is employed to construct negative samples, effectively addressing the scarcity of anomaly samples. Then, a deep memory feature embedding module is introduced to capture feature differences between input samples and normal samples stored in the memory module. These differences are concatenated with the input sample features, thereby enhancing the model's ability to infer anomalous regions. Finally, a novel multi-scale feature fusion module is designed to strengthen the effective integration of local details and global contextual information, further improving the model's detection accuracy and robustness for multi-scale targets. Experimental results demonstrate that, under the AUC metric, the proposed method achieves a 0.7% and 4.9% improvement in anomaly detection accuracy on the MVTec AD and BTAD datasets, respectively, compared to the best baseline methods. Additionally, it achieves a 3.2% and 2.7% improvement in anomaly localization accuracy on these datasets, making it more suitable for the requirements of current industrial image anomaly detection tasks.

(3) Based on the aforementioned research, a self-supervised learning-based industrial image anomaly detection system has been designed to achieve faster and more efficient industrial image anomaly detection tasks. First, the system incorporates a user login interface to ensure security and data confidentiality. Then, the system constructs multiple functional modules, including model training and testing, industrial image anomaly detection, and detection result controls, enabling users to intuitively understand the complete workflow and operational steps of anomaly detection. Finally, experimental results demonstrate that the system can rapidly and accurately identify defects in industrial products, validating its effectiveness and practicality.

In summary, this paper focuses on self-supervised learning-based methods for industrial image anomaly detection, achieving effective anomaly detection and localization for complex textures and industrial component images. Furthermore, the proposed method not only identifies the presence of defects in industrial images but also precisely localizes defect regions and measures their dimensions, providing robust technical support for industrial product quality assessment.

Fan Weiping(Electronic Information)

Supervised by Cheng Fanyong

KEYWORDS: self-supervised learning, industrial images, anomaly detection, structured reconstruction, memory feature embedding

目录

摘 要.....	I
ABSTRACT.....	IV
第 1 章 绪论.....	1
1.1 研究背景与意义.....	1
1.2 国内外研究现状.....	2
1.2.1 基于传统技术的图像异常检测方法.....	3
1.2.2 基于深度学习的图像异常检测方法.....	4
1.3 主要内容.....	7
1.4 章节安排.....	9
第 2 章 相关基础理论与技术.....	10
2.1 引言.....	10
2.2 深度学习基础理论.....	10
2.2.1 卷积神经网络简介.....	10
2.2.2 经典卷积模块介绍.....	13
2.2.3 深度学习框架.....	15
2.3 自监督异常检测算法概述.....	16
2.3.1 基于图像重建的异常检测.....	17
2.3.2 基于特征嵌入的异常检测.....	18
2.3.3 评价指标.....	20
2.4 本章小结.....	21
第 3 章 基于自监督结构化重建的工业图像异常检测与定位.....	22
3.1 结构化重建异常检测模型.....	23
3.1.1 掩码图像设计.....	23
3.1.2 结构化重建模型.....	25
3.1.3 损失函数设计.....	26
3.2 数据集和评价指标介绍.....	28
3.2.1 数据集介绍.....	28

3.2.2 评价指标介绍.....	29
3.3 实验结果与分析.....	30
3.3.1 实验平台和参数设置.....	30
3.3.2 消融实验分析.....	30
3.3.3 异常检测与定位分析.....	32
3.4 本章小结.....	34
第 4 章 基于自监督深度记忆特征嵌入的工业图像异常检测与定位.....	36
4.1 深度记忆特征嵌入异常检测模型.....	37
4.1.1 异常模拟策略.....	37
4.1.2 深度记忆特征嵌入模块.....	39
4.1.3 多尺度特征融合模块.....	40
4.1.4 损失函数.....	41
4.1.5 异常检测得分.....	41
4.2 数据集和评价指标介绍.....	42
4.2.1 数据集介绍.....	42
4.2.2 评价指标介绍.....	43
4.3 实验细节及结果分析.....	44
4.3.1 实验平台和参数设置.....	44
4.3.2 消融实验分析.....	44
4.3.3 异常检测分析.....	45
4.3.4 异常定位分析.....	46
4.4 本章小结.....	49
第 5 章 工业图像异常检测系统的设计与实现.....	50
5.1 异常检测系统框架设计.....	50
5.1.1 系统框架设计.....	50
5.1.2 运行环境设置.....	51
5.2 异常检测系统的 UI 界面设计.....	52
5.2.1 用户登录界面设计.....	52
5.2.2 系统功能选择页面设计.....	52

5.2.3 模型训练与测试页面设计.....	53
5.2.4 工业图像异常检测页面设计.....	54
5.3 工业图像异常检测系统的应用与实现.....	54
5.3.1 结构化重建模型异常检测系统实现.....	54
5.3.2 特征嵌入模型异常检测系统实现.....	55
5.4 本章小结.....	56
第 6 章 全文总结与展望.....	57
6.1 全文工作总结.....	57
6.2 后续研究展望.....	58
参考文献.....	60
攻读硕士学位期间所发表的论文及其它研究成果	67
致谢	68

第 1 章 绪论

1.1 研究背景与意义

制造业是我国经济的重要支柱，承担着国民经济的主体地位，是国家发展的基础和强盛的基石^[1]。然而，随着制造总量的迅速增长，我国制造业也面临着诸多挑战，比如产品质量不高与生产智能化程度不足等问题，这些因素限制了其进一步发展。在这种背景下，对工业缺陷的研究，尤其是针对常见工业产品表面缺陷的检测技术，面临着更大的需求和更高的标准。在现代化的大规模工业生产过程中，由于生产技术、设备精度以及环境条件等多重因素的影响，产品表面往往会出现各类缺陷，如污渍、划痕、裂纹等异常现象。随着市场对产品质量要求的不断提升，如何快速、准确地识别这些异常产品已成为工业生产中的关键问题。与传统的异常检测任务不同，工业图像异常检测不仅需要对图像进行整体异常判断，还需要精确定位异常的具体区域，以便生产人员能够及时采取针对性的修复措施或调整生产参数。因此，近年来，工业图像异常检测与定位技术的研究受到了学术界和工业界的广泛关注。

在工业制造领域，产品表面缺陷的检测与定位是确保产品质量的关键环节。然而，传统的人工视觉检测方法存在诸多局限性。例如，人工检测需要在流水线上对每件产品进行逐一检查，耗时较长，且随着工作时间的延长，工人的注意力会逐渐下降，导致漏检率和误检率明显上升。此外在精密电子元器件的生产中，微小的焊点缺陷或细微的裂纹往往无法通过人工被发现，而这些缺陷却可能对产品的性能和可靠性产生重大影响，不能满足大规模的生产需求。

随着现代化工业制造智能化水平的发展，基于计算机视觉的工业产品表面缺陷检测已成为一项重要的技术研究，并取得了显著的成效。这项技术主要采用摄像传感与智能分析技术，通过多模态数据采集与深度学习算法，实现对目标场景异常状况的自动识别与判定。系统包含数据采集、特征提取、模式识别和决策输出等核心模块，能有效识别偏离正常预期的异常数据，既降低了成本也提升了效率与性能，被广泛应用于纹理瑕疵检测、印刷电路板缺陷检测以及工业金属表面异常检测等多个工业场景。

尽管如此,基于机器视觉的图像异常检测技术仍面临诸多挑战。例如监督学习算法在模型训练过程中需要依赖大量有标签的数据,而在实际工业生产中,获取这类数据尤其是像素级数据的成本极高,导致此类模型的应用范围受到限制^[2],在视觉检测系统中,针对较大且规则的缺陷识别准确率较为理想,然而面对微观尺度瑕疵时,检测效果则明显下降。由于工业产品在制造过程中产生的缺陷往往呈现微小特征,这对检测算法的灵敏度提出了更高要求。随着现代制造业对产品精密度要求的提升,微小缺陷检测的重要性日益凸显。同时,考虑到生产线对检测效率的严格要求,理想的检测算法不仅需要保持较高的识别准确率,还应具备快速响应的特性。然而,现有基于监督学习的检测框架普遍存在计算复杂度高、模型参数量大和成本等问题,难以满足轻量化部署的需求,这在一定程度上限制了其在工业现场的应用。鉴于以上问题,研究者们开始探索基于自监督学习的异常检测方法,这类方法在训练过程中仅需使用正常样本即可构建检测模型,其核心原理在于通过对比学习、掩码图像建模等自监督策略,构建正常样本的特征空间分布,进而通过重构误差或特征偏离度识别异常区域。这种技术革新不仅将标注成本降低 90%以上,更重要的是解决了工业场景中"异常形态未知、缺陷类型动态变化"的核心难点,为工业图像异常检测提供了一种高效且经济的解决方案,推动工业视觉检测进入了新阶段^[3]。

综上所述,本文选择采用自监督学习方法对工业图像异常检测算法进行深入研究。随着主流监督学习模型的日益成熟,模型精度、速度和成本平衡的问题逐渐成为研究者面临的主要挑战,而自监督学习仅需正常样本进行训练即可快速检测异常区域,展现出巨大的发展潜力。尽管自监督学习在多个研究领域已取得初步成果,但其应用深度和广度仍有待进一步探索,尤其是在工业图像异常检测等复杂任务中,其研究价值和应用前景尤为显著。因此,本文旨在通过自监督学习方法挖掘工业图像中的潜在特征,为异常检测提供一种更加高效和灵活的解决方案。

1.2 国内外研究现状

图像异常检测技术作为计算机视觉领域的关键研究方向,已经发展出多种解决方案,主要可划分为传统技术方法和深度学习方法两大类^[4]。首先,将系统地

分析传统检测方法的演进过程，分析其核心技术突破与重要进展。随后，重点探讨深度学习方法的理论发展脉络和实际应用中的表现。基于对现有研究的深入分析，本文将进一步分析无监督图像异常检测方法存在的局限性，包括特征提取的方法、数据集的不足以及在复杂场景下的适应性问题。

1.2.1 基于传统技术的图像异常检测方法

在计算机技术未得到大力发展之前，一些传统方法在图像异常检测领域起了很大的作用。基于机器学习的图像异常检测主要有基于统计模型方法^[5]、基于频域分析方法^[6]、一类分类方法^[7]。基于统计模型方法主要运用统计学原理对图像数据进行建模，可以分析像素或区域的统计特性来进行异常检测，其中常见的方法包括高斯混合模型和马尔可夫随机场。例如，针对高光谱图像像素信息的分布特征建模问题，李等人^[8]采用单高斯分布函数对像素特征进行建模分析，但该方法在数据分布形态偏离高斯假设时，检测性能会出现明显下降。针对以上数据分布的复杂性等问题，后续研究引入了高斯混合模型（GMM），通过多个高斯分布的线性组合来构建概率密度函数，从而更准确地描述数据的真实分布特征。然而，基于统计模型方法仍面临两个主要挑战：首先，模型训练需要依赖大规模标注样本，这在实际应用中往往难以获取；其次，在背景复杂的场景中，模型构建过程存在较大困难，且对超出假设范围的异常样本容易出现识别错误^[9]。基于频域分析的异常检测方法主要通过将图像从空间域转换到频域来实现异常识别。这类方法通常采用傅里叶变换或小波变换等技术，提取图像的频率成分。在频域中，通过分析频率分布及其特征，能够有效检测出与正常模式显著差异的异常区域。

例如，Aiger 等人^[10]提出了一种创新的缺陷检测方法，该方法基于纯相位傅里叶变换并结合自适应阈值技术，实现了图像异常区域的自动识别。相较于传统的背景频谱消除技术，该方案具有计算复杂度低、实现简单的优势，但在处理周期性分布的大面积缺陷时，其检测效果可能不够理想。值得注意的是，随着工业检测场景的多样化，频域分析方法的应用范围受到明显限制，主要适用于具有规则纹理特征的图像检测任务。作为一种经典的单分类器，支持向量机（Support Vector Machine, SVM）在机器学习与模式识别领域具有重要地位。该算法通过在高维特征空间中构造最优分割超平面，并最大化间隔来实现样本的有效分类。特别是

在处理非线性可分问题时，SVM 通过核函数映射展现出独特的优势。基于这一特性，SVM 被广泛应用于异常检测任务中，通过寻找正常样本与异常样本之间的最优边界来实现目标识别。例如，郭慧等^[11]提出了一种基于 SVM 的钢板缺陷检测方法。该方法首先对原始图像进行滤波、降维和二值化等预处理操作，随后提取多维特征并利用 SVM 进行训练，最终实现缺陷区域的识别。然而，这种依赖人工设计特征的方法存在一定局限性：一方面，在干扰较多的工业环境中，其适应性较差；另一方面，针对图像等高维数据，手工特征设计的计算复杂度较高，导致检测效率较低。尽管基于传统技术的异常检测方案在某些特定应用场景中能够取得一定的检测效果，但这类方法往往依赖手工特征设计，并需要构建复杂的先验知识体系，导致其在实际应用中存在适应性差、计算效率低等局限性。近年来，随着深度学习技术的突破性进展，特别是卷积神经网络在特征学习方面的优势，基于深度学习的异常检测方法逐渐成为该领域的研究热点，为图像异常检测提供了新的解决思路^[12]。

1.2.2 基于深度学习的图像异常检测方法

近年来，深度学习技术的快速发展推动了图像异常检测领域的显著进步。现有方法主要可分为两类：有监督异常检测方法和无监督异常检测方法^[13]。以下将分别对这两类方法的研究进展进行详细介绍。

(1)有监督异常检测方法

有监督异常检测方法是利用标注数据进行异常检测。其核心思想是通过训练正常样本和异常样本构建分类模型，从而实现对正常数据和异常数据的区分^[14]。例如，Xue 等人^[15]设计了一种改进 Faster R-CNN 的盾构隧道衬砌表面缺陷检测方法，该方法采用改进型 Inception 全卷积网络作为特征提取器，通过优化锚框设计策略，新增两种尺度比例，并创新性地使用位置敏感型 ROI 池化层替代标准池化操作。实验结果显示，在保证检测效率的前提下，该模型对各类缺陷的识别准确率达到 95% 以上，较传统方法有显著提升。Ding 等人^[16]设计了一种用于 PCB 表面缺陷检测的网络，该方法利用 K 均值聚类优化锚框尺寸，并将多尺度金字塔网络引入 Faster R-CNN 中，增强了底层结构信息的融合能力，从而提升了微小缺陷的检测效果。此外，针对小数据集和样本不平衡问题，训练阶段采

用了在线困难样本挖掘技术。该方法在 PCB 缺陷数据集上实现了 98.90%的 mAP^[17]。Wang 等人^[18]针对小缺陷难以检测的问题，在 Faster R-CNN 的基础上，通过多层特征网络提取焊缝缺陷的多尺度特征，并改进了 RPN 网络，增加了多种滑动窗口，优化了模型锚框的纵横比，使得识别准确率达到了 98%以上。以上有监督方法虽然能够准确检测缺陷，但是需要大量训练样本和标记样本，而且训练时间较长，存在精度与效率不平衡的问题，难以满足只有少量无标记工业图像异常检测的需求^[19]。

(2) 自监督图像异常检测

在实际工业场景中，异常检测面临的核心挑战在于异常样本的稀缺性和多样性。尤其是在工业流水线等环境中，异常事件的发生频率极低，且异常形态复杂多变，导致难以收集足够的标注数据来训练有监督分类模型。相比之下，自监督方法仅需利用易于获取的正常样本进行模型训练，即可有效检测异常区域，从而避免了有监督学习中对大量标注数据的依赖以及标注过程的繁琐性。此外，自监督方法在训练效率和泛化性能方面展现出显著优势，能够更好地适应复杂多变的工业场景。目前，自监督工业图像缺陷检测主要分为两类方法：一类是基于图像重建的方法，通过重建误差识别异常；另一类是基于深度特征嵌入的方法，利用特征空间的分布差异检测异常。这两种方法各有特点，以下将详细介绍。

在图像异常检测领域，基于图像重构的方法主要采用自编码器（AutoEncoder, AE）或生成对抗网络（Generative Adversarial Network, GAN）作为基础架构。这类方法的基本原理是利用模型对正常样本具有较好的重构能力，而对异常样本则会产生较大的重构偏差，通过计算输入图像与重建结果之间的差异来实现异常识别。卷积自编码器因其能够有效提取图像的空间特征，已成为该领域的主流模型之一。在实际应用中，通常采用均方误差（MSE）作为损失函数，量化原始输入与重建输出之间的像素级差异，从而判断样本是否异常^[20]。例如 Haselman 等人^[21]提出了一种深度卷积自编码器实现工业零件表面缺陷检测，该方法损失函数采用了均方误差，只适用于一些均匀纹理、亮度差异较小、噪声较小的图像，图片重构效果不好，导致输入图片与重构图片之间的重构误差区分不明显，影响检测效果。然而，传统自编码器由于受到瓶颈层结构的限制，在特征压缩过程中容易出现信息损失问题，这导致重建图像的边缘轮廓细节区域出现模糊现象。考

考虑到局部特征之间的空间关联性，研究者们开始探索新的特征提取机制，以期更精确地建模图像局部区域间的依赖关系，这种改进对于提升异常检测的定位精度具有重要意义。例如，Bergmann 等人^[22]将结构相似性损失 SSIM 作为自编码器的损失函数，改善了重构图像的质量，提高了异常检测与定位准确率。He 等人^[23]对输入图片进行随机掩码，不仅加速了模型训练，而且提高了模型的泛化性能。但单一掩码会在一定程度上丢失一些重要细节特征和边缘信息，从而影响重构图像的质量。为了更多保留图像边缘和局部细节，Vitjan 等人^[24]提出了一种基于 UNet 自编码器的工业图像缺陷检测方法。该方法提出了多掩码融合的方法降低重建异常区域可能性，同时保持正常区域能较好地重建，提高重构误差识别异常的准确性，在 MVTec AD 数据集上的准确率达 90.7%。

基于特征空间的方法是一种通过分析正常样本与缺陷样本在特征空间中的分布差异来实现缺陷检测的技术^[25]。该方法首先提取图像的特征表示，然后在特征空间中计算正常样本与待测样本之间的差异，这种差异通常被称为异常分数。当异常分数超过预设阈值时，即可判定为缺陷。这种方法的核心在于构建能够有效区分正常与异常特征的特征空间，并通过距离度量方法来量化特征差异。其优势在于能够捕捉到图像的高维特征信息。例如，Aqsa 等人^[26]提出了一种基于自监督学习的图像数据生成策略，该方法通过构建自编码器网络，采用随机裁剪机制将输入图像分割为若干局部区域，并以剩余区域作为网络输入来预测被裁剪部分的内容。这种方法通过挖掘图像自身的语义信息来构建训练样本，不仅降低了对外部数据的依赖，还显著提升了模型的泛化能力。实验结果表明，经过该方法训练的模型在图像修复任务中表现出优异的性能，为图像复原领域提供了新的技术思路。在深度特征建模方面，Defard 等人^[27]提出了一种基于预训练卷积神经网络的特征提取框架，该方法首先利用深度卷积网络提取输入图像的多层次特征表示，随后通过构建多元高斯分布模型来刻画正常样本的概率分布特性。这种概率建模方法不仅能够有效捕捉特征间的统计相关性，还为异常检测提供了可靠的决策依据。在上下文信息提取领域，Xu 等人^[28]提出了高分辨率上下文信息提取网络（HRCNet），该网络架构采用双分支注意力机制，通过并行处理局部和全局特征来捕获丰富的上下文信息。为了更好地融合上下文语义信息，研究者设计了一种轻量级的特征增强金字塔结构，实现了多尺度特征的深度融合，有效解决了传

统方法中特征丢失的问题。此外，为了提升边缘区域的检测精度，该方法还引入了边界感知模块，通过自适应权重调整机制来强化边界特征的学习。

尽管自监督方法在检测效率、精度和成本上优于有监督方法，但仍面临一些挑战。例如，基于图像重建的方法利用重构误差判别异常时，异常区域可能被有效重构，导致正负样本重构误差的差异不明显，影响检测精度。其次，基于特征嵌入的方法在推理阶段计算成本高，且对异常检测任务的适应性不足。因此，本课题将致力于研究引入更复杂的损失函数和多尺度特征融合模块来提升模型性能，同时探索更高效的训练策略和模型架构，以实现更精准、快速、低成本的工业图像异常检测。

1.3 主要内容

以上内容探讨了工业图像异常检测的发展历程，讲述了从传统人工的异常检测到基于机器视觉的异常检测的创新发展。经过进一步分析各类方法的优缺点，发现自监督异常检测方法无论是在检测效率、精度和成本均优于其他方法，已成为主流的工业图像异常检测方法之一。本文主要研究内容如图 1-1 所示，它主要分为两类方法：基于图像重建的方法和基于深度特征嵌入的方法，虽然这些方法在工业图像中表现出较好的性能，但仍然存在一定的局限性。具体分析如下：

(1) 利用重构误差判别异常失效

基于图像重建方法虽然具有较强的泛化性能，但这种特性可能导致异常区域被错误地重构为正常模式，使得重建图像与原始输入之间的差异显著降低。这种现象直接导致正常样本和异常样本的重建误差分布重叠，从而影响了模型的异常检测能力。因此，需要着重研究如何优化正常样本的特征表示方法，以扩大正常与异常样本在重建误差空间中的可分性。同时，需要设计更具判别力的异常检测框架，以提高网络对正负样本的识别能力，从而获得更优异的检测性能。

针对以上问题，设计了一种基于自监督结构化重建的工业图像异常检测与定位模型。首先，利用正常样本进行训练，并引入多尺度结构相似度损失 MS-SSIM 和多尺度梯度幅度相似性损失 MSGMS，以提升输入图像与重建图像在结构和纹理上的相似性。随后，提出了一种多互补掩码重构方法，通过对测试图像的异常像素进行多角度遮挡，降低模型重建异常区域的可能性，减少单次掩码重构结果

的随机性，从而提高异常检测与定位的准确性。

(2)模型精度、速度和成本的平衡问题

基于特征嵌入的方法在推理阶段需要进行复杂的特征匹配操作，增加了计算成本。此外，由于模型训练未使用针对异常的数据集，而是直接采用预训练参数进行特征提取和异常检测，其对异常检测任务的适应性仍有不足。因此，如何在更精准、更快、更低成本的条件下实现图像异常检测，已成为当前学术研究和工业领域中面临的一个重大挑战。

针对上述问题，设计了一种基于自监督学习的深度记忆特征嵌入模型。首先，通过模拟异常策略生成异常样本，有效解决了工业图像中异常样本稀缺的问题。然后，利用正常样本和模拟异常样本进行训练，并引入深度记忆特征嵌入模块，通过对比输入样本与记忆模块中存储的正常样本，提取两者之间的高级特征差异信息，从而加速模型训练过程。最后，为了进一步提升模型的检测能力，设计了多尺度特征融合模块，充分融合不同层次特征的上下文语义信息，增强模型对复杂场景的适应能力。

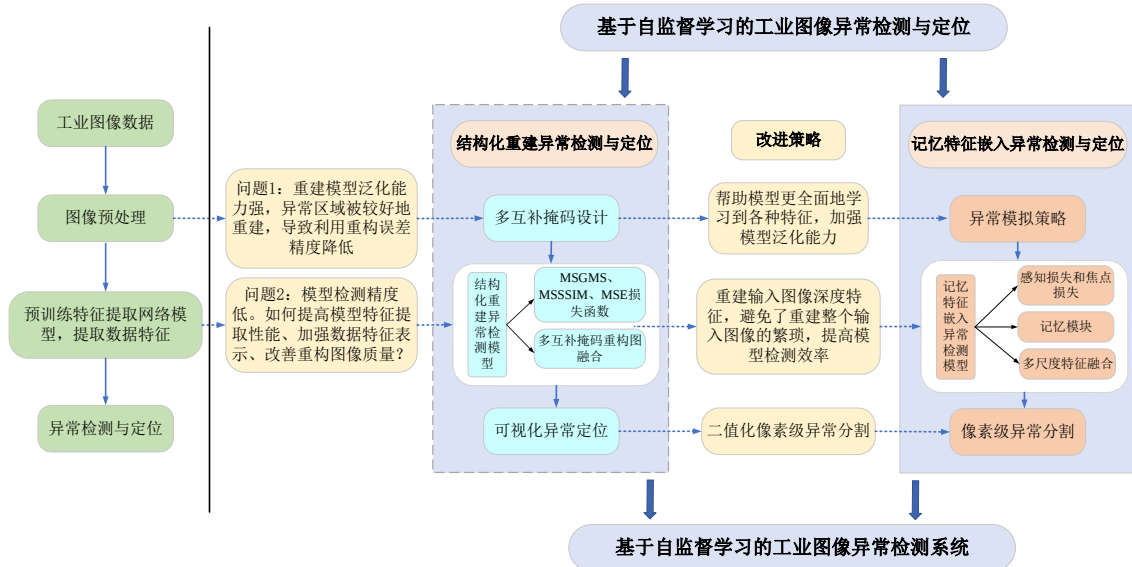


图 1-1 研究内容框图

(3)基于以上研究，为了实现更快捷、高效的工业图像异常检测任务，基于两种改进的自监督学习异常检测模型构建了工业图像异常检测系统软件实现。用户可以通过该系统的功能菜单或选项进行操作，其中包括模型训练与测试、工业图像异常检测以及检测结果等控件，从而实现命令、文件调用、程序启动以及其他一些常规操作，实现输入图像物体表面异常检测。

1.4 章节安排

第 1 章：绪论。阐述了异常检测算法的研究背景与意义以及它的历史发展，确立了本文的研究论点和研究方向，并且说明了现存的研究难点与不同的研究路径，并介绍了本文的研究重点是自监督学习机制的实现与优化，最后介绍了本课题的主要研究内容。

第 2 章：阐述了相关理论基础知识和自监督学习异常检测的方法。本章首先介绍了深度学习基础理论，其中包括卷积神经网络简介、经典卷积神经网络模块和深度学习框架基本介绍。然后介绍了自监督学习异常检测的两种比较主流的算法，分别是基于图像重建的异常检测和基于特征嵌入的异常检测方法，最后对评价指标进行了简要介绍

第 3 章：基于自监督结构化重建的工业图像异常检测与定位。首先，利用正常样本进行训练。然后，采用多互补掩码重构方法对测试图像的异常像素进行多角度遮挡，并引入多尺度结构相似度损失和多尺度梯度幅度相似性损失作为模型的损失函数。最后，通过计算输入图像与融合重建图像之间的多尺度结构相似性损失，实现异常区域的检测与定位。

第 4 章：基于自监督深度记忆特征嵌入的工业图像异常检测与定位。首先，通过模拟异常策略生成异常样本，并利用正常样本和模拟异常样本进行训练。然后，引入深度记忆特征嵌入模块，通过对比输入样本与记忆模块中存储的正常样本，提取两者之间的高级特征差异信息。最后，本文设计了多尺度特征融合模块，充分融合不同层次特征的上下文语义信息。

第 5 章：工业图像异常检测系统的软件实现。物体表面异常检测系统的设计与实现包括多个步骤。首先，我们使用 PyQt5 来设计系统的可视化界面，利用该界面完成各个控件功能的具体配置。然后，系统选择图像模式进行信息采集和读取，从而实现对图像的异常检测。

第 6 章：总结与展望。本章将对本文的研究工作进行总结，分析在研究过程中出现的问题与存在的不足，并对本文工作的未来发展趋势给予展望。

第 2 章 相关基础理论与技术

2.1 引言

近年来，深度学习技术在计算机视觉等多个领域取得了显著进展，推动了人工智能的快速发展。为了更好地奠定本文所提算法的理论基础，本章将重点探讨基于深度学习的图像异常检测技术。首先，从基础的卷积网络入手，介绍深度学习的核心概念和基本原理。然后，深入探讨更复杂的卷积神经网络模块及其架构设计。最后，详细分析自监督学习在异常检测中的常用方法，为后续算法的验证提供理论支持和参考依据。

2.2 深度学习基础理论

2.2.1 卷积神经网络简介

卷积神经网络（Convolutional Neural Networks, CNN）是深度学习领域的重要架构，主要采用独特的卷积运算机制构建深度前馈网络，在视觉任务中展现出卓越性能^[29]。该网络通过数学卷积操作取代传统的全连接计算，能够高效处理高维像素数据，特别适用于计算机视觉相关应用。如图 2-1 所示，典型的 CNN 架构由多个功能模块组成，包括数据输入层、卷积层、池化层和全连接层等。这些基础组件可根据任务需求灵活组合，构建出适应不同场景的网络模型。其中，卷积核的权重共享机制是 CNN 的核心特征，通过在特征图上滑动执行局部特征提取，有效捕获输入数据的空间层级结构信息。这种设计不仅降低了参数数量，还增强了模型对平移、缩放等变换的鲁棒性，使其成为图像处理领域的核心工具。

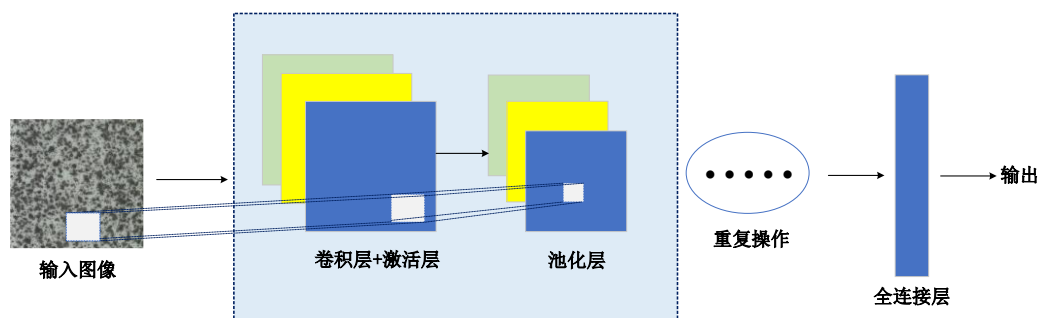


图 2-1 卷积神经网络架构设计

(1)卷积层运算

卷积层作为卷积神经网络的核心步骤，通过卷积核以滑动窗口的方式高效提取图像特征^[30]。卷积核的参数通过前向传播和反向传播算法进行优化学习，由于同一卷积核应用于整张输入图像，卷积层具备参数共享的特性，显著减少了模型参数量。卷积核在图像上逐像素滑动，对图像的不同区域和特征进行响应，通过计算卷积核与图像局部区域的点积，生成一系列特征图。特征图的尺寸可通过公式计算得出，具体取决于输入图像大小、卷积核尺寸、步长以及填充方式。具体公式如下：

$$F' = \frac{(F - K + 2P)}{S} + 1 \quad (2-1)$$

式中 S 为步长， P 为填充值， F 为输入图像的尺寸， K 为卷积核的尺寸大小

卷积操作如图 2-2 所示，以滑动步长为 1、尺寸为 2×2 的卷积核在尺寸为 3×3 的输入图像上滑动，输入图像中像素与卷积核对应位置参数相乘再相加，得到尺寸为 2×2 的输出特征图。

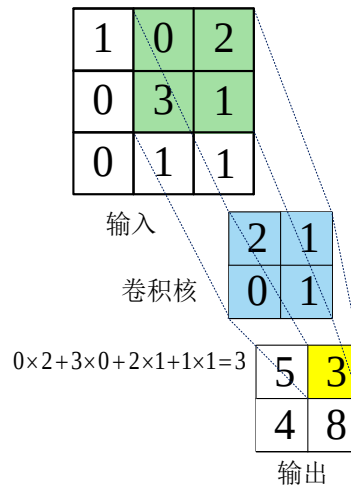


图 2-2 卷积计算图

(2)池化层

池化层是卷积神经网络中的重要组成部分，主要通过下采样机制降低特征图的维度，有效减少了计算资源的消耗^[31]。这种降维策略不仅能够显著降低模型的计算复杂度，还可以通过控制特征图的尺寸来缓解过拟合问题。此外，池化操作还赋予了网络对输入数据的几何变换不变性，包括局部平移和旋转等，从而增强了模型的泛化能力和鲁棒性。在实现方式上，池化操作主要在空间维度执行，常用的方法包括最大值池化和均值池化等。如图 2-3 所示，最大值池化通过选取感

受野内的极值来保留最显著的特征响应,而均值池化则通过计算局部区域的平均值来获得平滑的特征表示。这些不同的池化策略各具特点,为神经网络架构设计提供了多样化的选择方案。

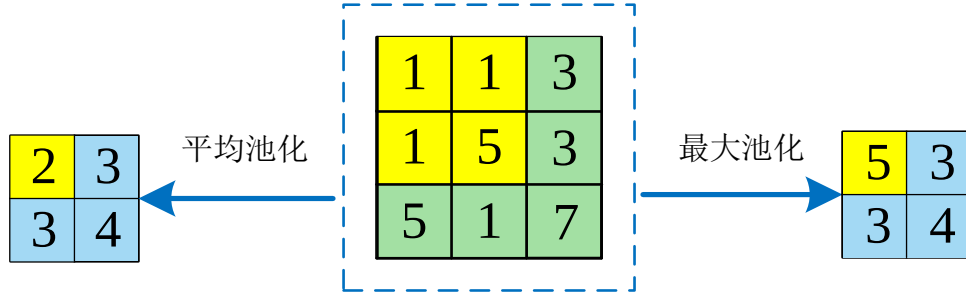


图 2-3 二种类型池化

(3) 激活函数

激活函数层,也称为非线性变换层,该层的主要作用是通过引入非线性映射关系,增强网络的表达能力,使模型能够学习复杂的特征表示^[32]。激活函数层由特定的激活函数构成,通常需要满足以下几个条件:首先,它们必须是连续且可导的非线性函数;然后,其计算过程应尽可能高效,以避免引入过多的计算过程;最后,其输出值域应被限制在一个合理的范围内,以确保模型的稳定性。在卷积神经网络中,常用的激活函数包括 Sigmoid、Tanh 和 ReLU 等。Sigmoid 函数将输入值映射到(0,1)区间,适用于二分类任务;Tanh 函数则将输入值映射到(-1,1)区间,具有零中心化的特性,有助于加速模型收敛;而 ReLU (Rectified Linear Unit) 因其计算简单、效率高且能有效缓解梯度消失问题,成为深度学习中最广泛使用的激活函数之一。其中 Sigmoid、Tanh 和 ReLU 函数如下式所示:

$$\text{Sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (2-2)$$

$$\text{Tanh}(x) = 2\text{Sigmoid}(2x) - 1 = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2-3)$$

$$\text{ReLU}(x) = \begin{cases} \max(0, x) & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (2-4)$$

目前 ReLU 已成为了最常用的激活函数,因为相比于其他激活函数,ReLU 计算更简单高效。ReLU 在 $x > 0$ 时,导数值恒为 1,这在一定程度上缓解梯度消失问题。

(4)全连接层

全连接层通常被部署在卷积神经网络的末端，主要用于实现分类功能^[33]。在这一层中，每个神经元都与前一层的所有神经元完全连接，导致其参数量极为庞大，占据了模型总参数的主要部分。然而，随着深度学习技术的不断演进，全连接层的局限性逐渐暴露，尤其是在计算效率和灵活性方面表现不足。为了解决这些问题，现代模型越来越多地采用全卷积网络结构，摒弃了传统的全连接层设计。全卷积网络通过卷积操作直接生成空间化的输出，不仅大幅减少了参数量，还保留了重要的空间信息，使其更适用于像素级预测任务，如图像分割和目标检测。这种结构设计在提升模型性能的同时，显著降低了计算复杂度，已成为当前计算机视觉领域的主流趋势，为高效处理复杂视觉任务提供了新的解决方案。

2.2.2 经典卷积模块介绍

(1)注意力机制

在卷积神经网络框架中，注意力机制^[34]作为一种高效的资源分配策略，能够有效应对计算资源受限场景下的信息过载挑战。该机制的核心原理是通过动态权重分配，从海量输入数据中提取与当前任务最相关的特征表示，同时抑制无关或冗余信息的干扰。这种选择性关注机制不仅能够显著降低计算复杂度，还可以通过增强关键特征的表达来提高模型的判别能力。通过自适应地调整特征权重分布，注意力机制实现了对重要信息的精准捕捉，从而提升了网络的处理效率和预测精度。

传统的注意力机制通常关注通道注意力，通过全局信息对通道特征响应进行重新校准。相较之下，坐标注意力（Coordinate Attention, CA）从空间维度进行创新，引入了对特征图中横向和纵向坐标信息的编码。这一机制的目标在于通过提升空间位置的显著性表示，从而增强模型的整体性能^[35]。CA 不仅能够有效捕捉特征图中的长距离依赖关系，还能够帮助模型识别图像中远距离相隔但仍然相关的对象和区域，这一特性在异常检测和目标检测等计算机视觉任务中显得尤为重要。此外，相较于传统的其他注意力机制，CA 在增加的计算和参数量极小的基础上，依然实现了显著的性能提升。这种设计合理性使得 CA 在实际应用场景中表现出了更强的实用性和灵活性，有助于处理复杂的视觉信息。

如图 2-4 所示，坐标注意力机制网络结构，在接收到输入后，先获取输入特征图的宽度 W 和高度 H 。接着，采用两个一维全局平均池化操作，将特征图在垂直和水平方向上分别分解为两个独立的特征图。其中， $XAPool$ 用于在宽度方向进行平均池化，而 $YAPool$ 则用于在高度方向进行平均池化。在获得平均池化后，将这两个特征图在空间维度上进行拼接，以形成一个综合特征图。其次，使用卷积操作对拼接后的特征图进行通道压缩。这一过程之后，进行归一化处理，并施加激活函数，以进一步增强特征表示。然后，将完整的特征向量重新分为两个方向的向量，针对每个方向进行卷积操作。此时，调整这两个方向特征向量的通道数，以便它们能够匹配。最后，通过 Sigmoid 函数对调整后的特征向量进行处理，并将其与原始输入信息在两个方向上进行加权融合，最终生成所需的输出特征图。这种方法不仅提高了特征表达的效率，还充分利用了空间信息，有助于后续的任务执行。

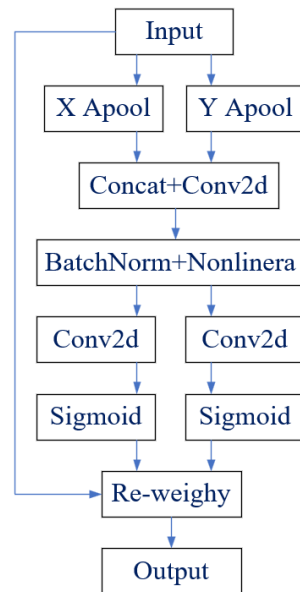


图 2-4 CA 网络结构

(2) UNet 网络介绍

UNet 网络的核心思想是通过编码器-解码器架构实现图像的精确分割^[36]。网络的左侧为编码器路径，通过一系列卷积和池化操作逐步提取高层次特征，同时降低特征图的分辨率；右侧为解码器路径，通过上采样和卷积操作逐步恢复分辨率，并融合来自编码器的低层次特征。UNet 的创新性在于引入了跳跃连接（Skip Connection），将编码器路径中的特征图与解码器路径的对应层直接连接，从而有

效保留了更多的空间细节信息，显著提升了分割精度。UNet 网络不仅有高效的特征提取与融合机制，而且更巧妙结合了多层次信息，为图像分割任务提供了一种高效且可靠的解决方案。

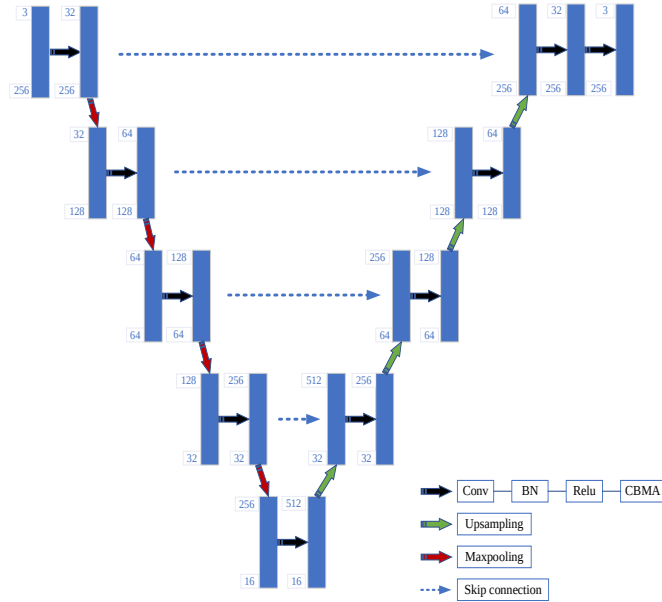


图 2-5 UNet 模型框架图

如图 2-5 所示，编码器部分由卷积组合操作和 Maxpooling 组成，其中蓝色箭头代表卷积组合操作，主要由卷积层、归一化层、Relu 激活函数层组成。首先，将输入图片经过卷积操作组合得到的特征图进行 Maxpooling 操作。然后，下采样得到的特征图再经过三个卷积组合操作共得到 5 组不同层级的特征图，并且每次下采样操作特征图的通道数均增加 1 倍。

解码器部分是用于将编码器提取的特征图恢复到原始图像尺寸的结构。解码器由多个上采样和卷积操作组成，通常包括反卷积或上采样层，旨在逐步扩大特征图的空间维度，以实现输入图像的高分辨率输出。在每一层上采样之后，UNet 应用跳跃连接，将相应层的编码器特征图与解码器的特征图进行拼接。这种设计允许解码器在逐步恢复图像细节时，利用从编码器中获取的高分辨率特征，从而有效地保留了图像的空间信息和上下文信息。

2.2.3 深度学习框架

深度学习框架是为深度神经网络模型的构建、训练及优化提供支持的软件工具库，涵盖了模型设计、数据集加载、数据预处理以及参数优化等核心功能。目

前，国内外涌现了众多深度学习框架，各具特色，但在学术界和工业界中，TensorFlow 和 PyTorch 仍是应用最为广泛的两大主流框架。以下对这两个框架进行简要分析：

TensorFlow 由 Google 团队开发并于 2015 年开源发布，现已成为应用最广泛的框架之一^[37]。该框架采用 C++ 作为底层实现，同时提供 Python、C++ 和 Java 等多语言 API 接口。其核心特点是使用数据流图来描述神经网络的计算过程，支持多设备并行计算和分布式训练，具有优异的可扩展性和灵活性，能够满足不同类型神经网络模型的构建需求。尽管 TensorFlow 在学术界和工业界都获得了广泛应用，但近年来面临 PyTorch 的激烈竞争。为此，TensorFlow 在 2.0 版本中进行了重大更新，引入了即时执行模式以支持动态计算图，从而提升其市场竞争力。

另一主流框架 PyTorch 由 Facebook 人工智能研究院开发，其前身是基于 Lua 的 Torch 框架，于 2017 年正式发布^[38]。该框架以 Python 作为主要编程语言，对 GPU 加速计算提供了原生支持。与 TensorFlow 的静态计算图不同，PyTorch 采用动态计算图机制，使得模型构建和调试过程更加直观灵活。由于其 API 设计符合 Python 编程习惯，且提供了丰富的工具库，PyTorch 在科研领域迅速普及，已成为众多前沿研究项目的首选开发平台。特别地，PyTorch 的自动微分机制和模块化设计大大简化了复杂模型的实现过程，为深度学习研究提供了强有力的支持。

总体而言，TensorFlow 和 PyTorch 各有优势：TensorFlow 在工业界的部署和规模化应用中表现突出，而 PyTorch 则在学术研究和快速原型开发中更受欢迎。两者的持续竞争与创新推动了深度学习技术的快速发展，为研究者和开发者提供了多样化的选择空间。

2.3 自监督异常检测算法概述

在工业智能化进程中，图像异常检测技术已获得广泛应用并取得显著成效。然而，实际应用场景中存在诸多挑战：首先，异常事件的发生概率极低，在样本集中占比往往不足 5%；其次，异常模式具有高度不确定性，难以预先定义和收集；此外，工业环境中的复杂背景和噪声干扰进一步增加了检测难度。为应对这些挑战，研究者们将目光转向了自监督学习方法。近年来，随着深度学习的突破性进展，基于自监督图像异常检测方法逐渐成为该领域的主流研究方向。这类方

法主要可分为两个技术路线：第一种是基于图像重建的异常检测；第二种是基于特征嵌入的异常检测。这些方法通过挖掘数据自身的监督信号，有效缓解了异常样本稀缺的问题，为工业检测提供了新的解决方案。

基于图像重建的方法通过自编码器学习正常数据的分布，并利用生成对抗网络 AnoGAN^[39]或自动编码器 AE^[40]生成输入图像的重建结果。随后，通过计算输入图像与重建图像之间的重构误差，实现异常检测与定位。这种方法虽然直观，但在复杂场景中可能面临重建精度不足的问题，限制了其在实际应用中的效果。

近年来，基于特征嵌入的方法因其优异的性能受到广泛关注。该方法利用预训练网络提取图像的高级特征，通过计算测试样本与正常样本特征之间的距离来判定异常，并定位异常区域。相比重建方法，特征嵌入方法避免了重建整个图像的繁琐过程，降低了计算成本，同时在工业应用中展现出更高的效率和更大的发展潜力，成为当前研究的热点方向。

2.3.1 基于图像重建的异常检测

目前，基于重构思想的检测方法是当前图像异常检测领域的主流技术路线之一，其核心在于利用卷积神经网络构建图像重建模型。该方法使用正常样本训练网络，使其学习正常数据的分布特征，从而具备重建正常模式的能力。在推理阶段，由于异常样本偏离了学习到的正常分布，重建过程会产生显著误差。通过分析输入图像与重建结果的差异，可以实现异常区域的精确定位。这类方法的主要优势在于其直观的检测机制，能够通过像素级差异分析实现精确的缺陷分割。然而，卷积神经网络强大的泛化能力可能导致异常区域也被较好地重建，导致正常样本和异常样本之间的重建误差减小，从而影响了模型的异常检测能力。这种现象在复杂背景或细微缺陷场景下尤为明显，因此需要设计更精细的特征提取机制来提高模型的判别能力。此外，常用的重建损失函数包括均方误差（MSE）、感知损失^[41]和结构相似性 SSIM^[42]等，如何选择合适的损失函数来评估重建质量也是一个挑战。

自动编码器（AE）是基于图像重构的经典卷积神经网络模型，如图 2-6 所示，它由编码器 $f(x)$ 和解码器 $g(z)$ 两部分组成。编码器 $f(x)$ 通过一系列卷积操作提取输入图像 x 的潜在特征，将输入数据压缩为低维表示 z ，以学习正常数据的

高维分布。随后，解码器 $g(z)$ 通过反卷积操作将潜在特征 z 重建为近似原始数据的输出 $\tilde{x}$ 。该过程的目标是最小化重建误差，即输入图像与重建输出之间的差异。通常使用均方误差（MSE）来衡量输入图像与重建图像之间的差异，并基于此进行异常检测。公式如下：

$$z = f(x) \quad (2-5)$$

$$\tilde{x} = g(z) = g(f(x)) \quad (2-6)$$

$$L_{MSE} = \|x - \tilde{x}\|_2^2 \quad (2-7)$$

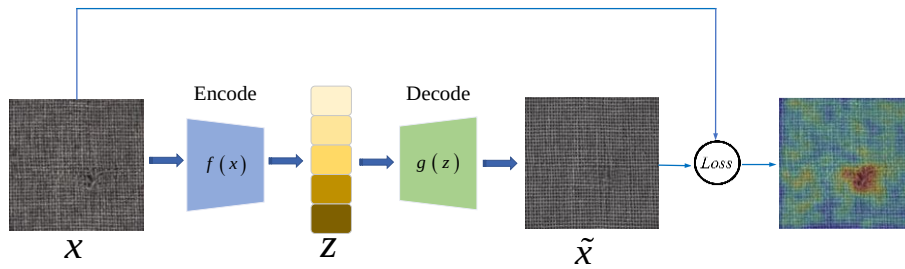


图 2-6 自编码器图像异常检测流程图

2.3.2 基于特征嵌入的异常检测

在图像异常检测领域，基于特征嵌入的方法是一种重要的主流技术，其核心思想在于构建一个能够有效区分正常与异常特征的特征空间，并利用距离度量方法来量化特征之间的差异。该方法首先通过预训练模型或特征提取网络获取图像的高级特征表示，随后在特征空间中计算待测样本与正常样本之间的差异，这种差异通常被称为异常分数。当异常分数超过预设阈值时，即可判定样本为异常。与基于重构的方法不同，特征嵌入技术专注于高维特征空间中的异常检测，而非直接在原始像素空间中进行操作。由于其具备高计算效率、无需复杂的重建过程，并能够捕捉更丰富的语义信息。这种技术不仅简化了异常检测的流程，还提升了检测的精度和鲁棒性，为实际应用提供了高效且可靠的解决方案。

如图 2-7 所示，展示了一种基于特征嵌入的异常检测模型示意图，该方法的核心创新点在于引入了知识蒸馏机制^[43]，通过预训练的教师网络提取多层次特征表示，并以此监督学生网络的训练过程。具体而言，教师网络生成的多尺度特征图被用作监督信号，指导学生网络学习更鲁棒的特征表示。在推理阶段，通过

计算多层特征空间中的差异度，实现对异常区域的精准定位与判别。

在训练阶段，首先使用正常样本数据集预训练一个深度卷积神经网络作为教师网络，并创建一个结构相同但参数随机初始化的学生网络。然后，从教师网络和学生网络中分别选取若干特征层，将其输出特征图作为真实标签和预测标签，并计算两者之间的损失。最后，通过反向传播算法，迭代优化学生网络的参数，使其能够更好地模仿教师网络的特征表示。以上损失函数主要采用两种损失函数，第一种采用欧式距离对特征进行约束，公式如下：

$$L_{val} = \sum_{i=1}^{N_{CP}} \frac{1}{N_i} \sum_{j=1}^{N_i} (a_T^{CP_i}(j) - a_S^{CP_i}(j))^2 \quad (2-8)$$

其中， CP_i 表示网络的第 i 个特征层， $a_T^{CP_i}(j)$ 表示学生网络在第 i 层的激活值， $a_S^{CP_i}(j)$ 表示教师网络在第 i 层的激活值， N_i 表示 CP_i 层中神经元的数量，而 N_{CP} 示被使用的特征层的总数。

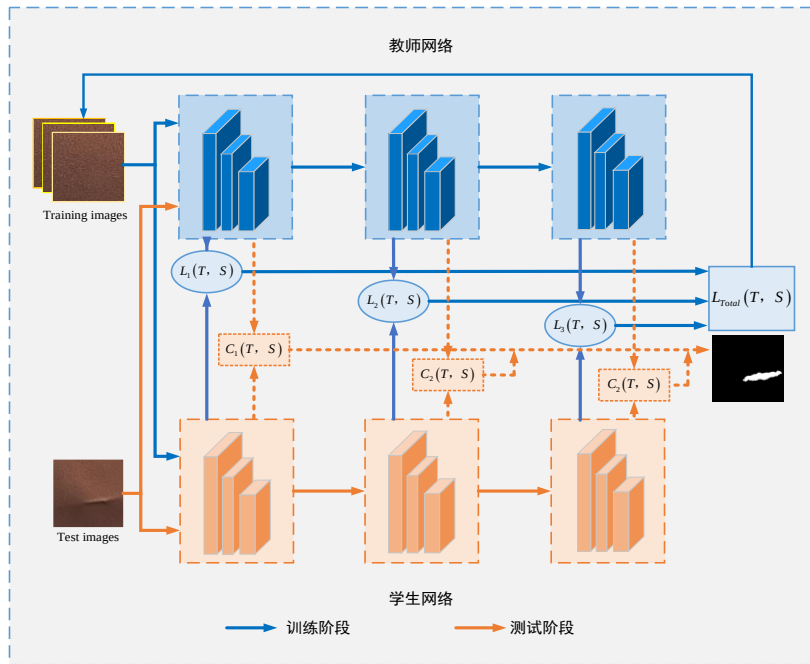


图 2-7 基于特征嵌入的异常检测模型示意图

在测试阶段，将测试图像分别输入已训练的学生网络和教师网络，提取两者的特征图。通过比较这些特征图之间的差异，计算异常分数。如果异常分数超过预设阈值，则判定图像中存在异常区域。这种方法避免了复杂的图像重建过程，直接在高维特征空间中进行异常检测，具有较高的计算效率和检测精度。

2.3.3 评价指标

在工业产品表面缺陷检测与分割任务的性能评估中，传统方法通常需要人为设定判别阈值，将网络输出的异常分数与该阈值进行比较以确定缺陷区域。然而，这种阈值选择方式容易引入主观偏差，影响评估结果的客观性。为了消除阈值设定的影响，研究者普遍采用受试者工作特征曲线（ROC）及其对应的曲线下面积（AUROC）作为评价指标^[44]。这些指标通过综合考虑不同阈值下的真阳性率和假阳性率，这种方法能够有效消除阈值选择的偏差，能够更全面地反映模型的检测性能。因此，AUROC 值越大表明模型的判别能力越强，为模型性能评估提供更加可靠和全面的依据。

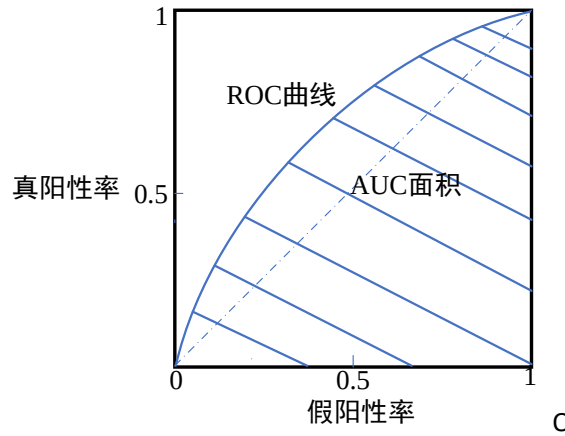


图 2-8 ROC 曲线示例

接受者操作特性（ROC）曲线是二分类任务的重要评估工具，能够全面反映分类器在不同判别阈值下的性能表现，特别适用于本文研究的缺陷检测场景。如图 2-8 所示，展示了一个典型的 ROC 曲线示例，其中深蓝色曲线表示分类器的性能轨迹，浅蓝色阴影区域对应曲线下面积（AUC），红色标记点则指示最优阈值位置。纵轴表示真阳性率（TPR），反映模型识别正样本的能力，计算方式为正确分类的正样本数与实际正样本总数的比值。横轴则代表假阳性率（FPR），表示模型将负样本误判为正样本的概率，其值为错误分类的负样本数与实际负样本总数的比值。计算公式如下：

$$TPR = \frac{TP}{TP + FN} \quad (2-9)$$

$$FPR = \frac{FP}{FP + TN} \quad (2-10)$$

式中：FP 表示被模型错误判断为正类的负样本数，TP 表示被模型正确判断为正类的正样本数，FN 表示被模型错误判断为负类的正样本数，TN 表示被模型正确判断为负类的负样本数。

2.4 本章小结

本章系统性地阐述了图像异常检测方法的理论基础和算法模型。首先，对卷积神经网络中的关键组件进行了深入分析，包括卷积层、池化层、激活函数和全连接层等模块的工作原理及其相互作用机制。然后，详细阐述了经典卷积神经网络的核心模块，包括注意力机制和 UNet 网络架构。这些模块通过不同的特征提取和融合策略，能够有效提升网络的表达能力。进一步分析了这些模块在图像异常检测任务中的应用价值。最后，重点探讨了自监督学习在工业图像异常检测中的应用，详细分析了基于图像重建和特征嵌入的两类主流算法框架。这些理论知识的梳理为后续章节的研究工作奠定了坚实的理论基础，同时也为新型检测算法的设计提供了重要的参考依据。

第3章 基于自监督结构化重建的工业图像异常检测与定位

在现代工业生产中，随着生产线自动化程度的不断提升，为确保产品质量的一致性和可靠性，对产品进行异常检测并实施分类处理已成为关键环节^[45]。与传统的语义级图像异常检测不同，当前任务不仅要求模型具备图像级的异常判断能力，还需要实现像素级的精确定位，这对模型的性能提出了更高要求。值得注意的是，在实际应用场景中，异常样本的获取难度较大，且像素级标注需要耗费大量人力物力，导致传统的监督学习方法难以适用。相比之下，无缺陷的正常样本相对容易收集，因此现有的工业图像异常检测与定位方法大多采用自监督学习方法。这种学习方式通过挖掘正常样本的内在特征分布，能够有效解决异常样本稀缺的问题。

目前，基于重构的思想是图像异常检测的主流解决方案之一，其通常利用自编码器重构模型。这类方法的核心思想是通过正常数据训练网络，使其能够重建正常样本。在测试阶段，由于异常样本与正常样本存在差异，使用训练好的模型重建测试图像时会产生较大的重建误差，通过比较原始图像与重建图像，可以识别并定位异常区域。这类方法的优势在于思路直观，能够通过逐像素对比实现像素级缺陷分割，快速定位异常区域。然而，重建模型在提高重建能力的同时，泛化能力也会增强，导致异常区域也可能被重构，难以通过重建误差进行准确的异常检测与定位。此外，重建模型的损失函数通常只适用于均匀纹理和结构差异较小的图像，无法很好地反映重构图像的内部结构和纹理细节，从而导致缺陷漏检率较高。

针对上述问题，提出了一种基于自监督结构化重建的工业图像异常检测与定位模型。首先，利用正常样本进行训练，通过引入多尺度结构相似度损失和多尺度梯度幅度相似性损失，增强输入图像与重建图像在结构和纹理上的相似性。然后，采用多互补掩码重构方法对测试图像的异常像素进行多角度遮挡，从而降低模型重建异常区域的可能性，并减少单个掩码重构结果的随机性，提升异常检测与定位的精度。最后，通过计算输入图像与融合重建图像之间的多尺度结构相似性损失，实现异常区域的检测与定位。

3.1 结构化重建异常检测模型

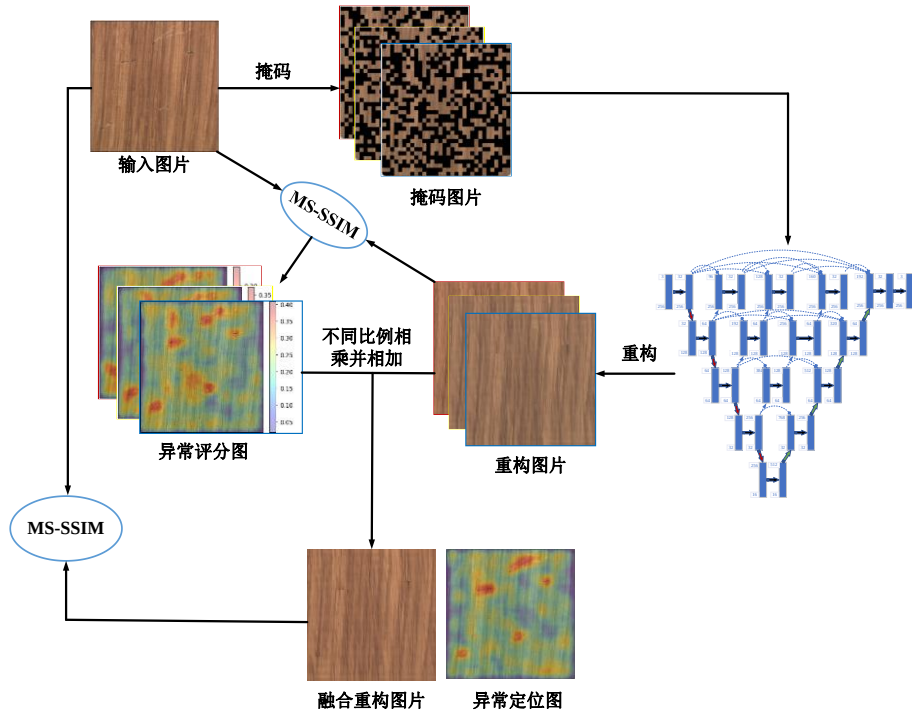


图 3-1 异常检测与定位流程图

如图 3-1 所示，设计了工业图像异常检测与定位流程图。首先，将测试图像进行 n 次互补掩码，以获得 n 个互补掩码图像，其中掩码大小为 $f \times f$ 个像素。然后，使用预训练模型对掩码图像进行结构化重建，根据输入图像和重建图像之间的重建误差，得到每幅重构图像的异常评分图，计算每个异常评分图中 $f \times f$ 像素区域里所有像素的均值，其中最大异常得分是缺陷的可能性最大，并把它作为重构图的异常得分。最后，计算每幅重构图异常得分所占全部重构图异常得分的权重，把每幅重构图乘以各自对应的权重并相加得到融合重构图片，再根据输入图像和融合重构图像之间的多尺度结构相似性误差实现异常定位。

3.1.1 掩码图像设计

首先，把输入图像 I 分割为 $N = \frac{H}{f} \times \frac{W}{f}$ 个网格得到分割后的图像 I' ，每个网格包含 $f \times f$ 像素，其中 W 和 H 是输入图像的高度和宽度。然后，输入图像中的网格随机分为 n 个互补集合 $y_i (i = 1, 2 \dots n)$ ，其中每个集合 y_i 中有 $\frac{N}{n}$ 个小方框，并把每个集合 y_i 复原成与输入图像高度和宽度相同的二进制掩码图像 M_{y_i} ，获得 n 个二进

制掩码图像 M_{y_i} ，掩码图像 M_{y_i} 把之前选中的集合 y_i 小方框置 0，其余小方框置 1。

最后，输入图像分别乘以二进制掩码图像 M_{y_i} 得到 n 副掩码图像 I_{mi} ：

$$I_{mi} = I' \odot M_{y_i} \quad (3-1)$$

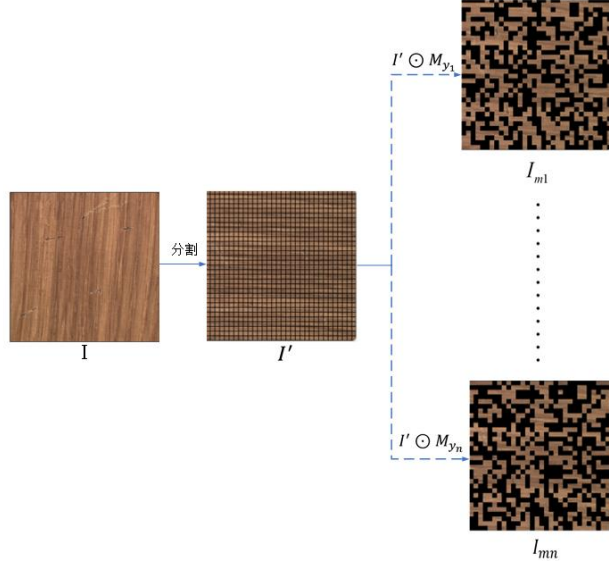


图 3-2 掩码图像设计流程图

如图 3-2 所示，设置测试图像高度 W 和宽度 H 为 256，并分成 32×32 个网格，每个网格包含 $f \times f = 8 \times 8$ 个像素，输入图像中的网格随机分为 n 个互补集合，然后输入图像分别乘以由 n 个互补集合生成的二进制掩码图像 M_{y_i} ，最终得到 n 副互补掩码图像。

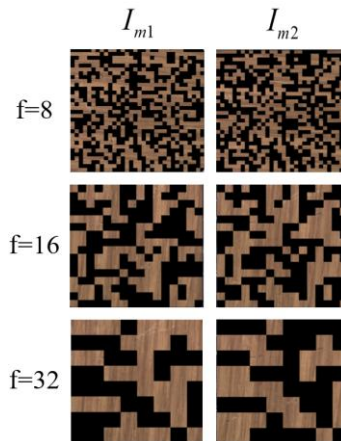


图 3-3 $n = 2$ 时互补集合掩码图像示例图

如图 3-3 所示，掩码像素大小为 $f \in (8, 16, 32)$ 的掩码图像，其中 2 张掩码图像互补。

3.1.2 结构化重建模型

本文方法基于 UNet++ 模型实现掩码图像的重建,相较于 UNet 模型,UNet++ 具有更强的语义表达能力和泛化性能^[46]。UNet++ 通过在不同尺度的特征图中引入密集跳跃连接,将编码器输出的特征图与解码器各阶段的特征图紧密连接,显著减少了重要细节的丢失^[47]。此外,UNet++ 还引入了多尺度特征融合模块,通过融合不同层级的特征,将局部信息与全局信息有效结合,从而更好地捕捉图像的全局上下文信息^[48]。

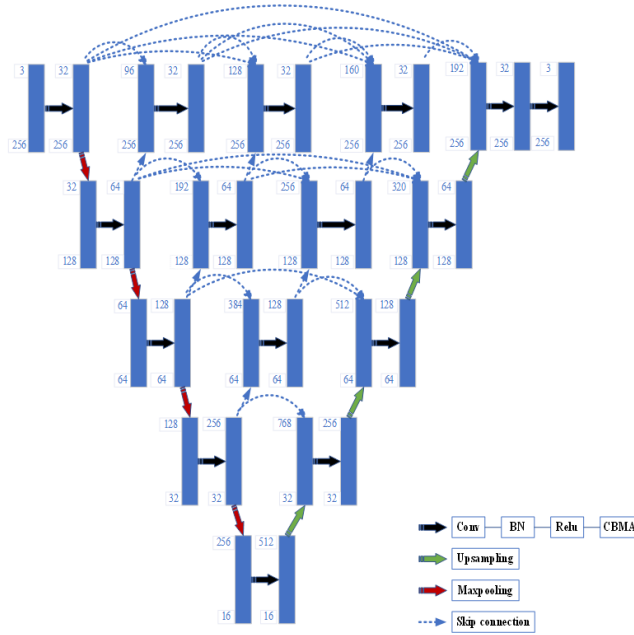


图 3-4 改进 UNet 重建模型示意图

如图 3-4 所示,编码器部分由卷积组合操作、注意力模块和最大池化 (Maxpooling) 组成,其中黑色箭头表示卷积组合操作,包括卷积层、归一化层、ReLU 激活函数层以及注意力模块^[49]。首先,输入图像经过卷积操作生成特征图,随后进行 Maxpooling 下采样。然后,下采样后的特征图通过三次卷积组合操作,生成五组不同层级的特征图。最后,将上采样的特征图与前一层的特征图融合,构建同一层级但深度不同的特征表示。在解码器部分,最后一层的特征图经过上采样,并与编码器对应层级的特征图融合。最终,通过逐层卷积解码上采样特征图,生成重建图像。这种方法充分利用了多尺度特征融合的优势,提升了重建精度和细节保留能力^[50]。

3.1.3 损失函数设计

均方误差（MSE）在处理均匀纹理和亮度差异较小的图像时表现良好，但对于研究的非均匀纹理和复杂表面图像，其局限性显而易见。为了弥补这一不足，引入了多尺度结构相似性损失（MSSSIM），该损失函数从亮度、对比度和结构三个维度综合评估图像全局像素之间的相关性，能够更好地捕捉图像的全局特征。此外，为了进一步增强对图像边缘和纹理细节的关注，本文还结合了多尺度梯度幅度相似性损失（MSGMS），该损失函数通过计算图像梯度的多尺度差异，有效保留了图像的局部细节信息。通过引入 MSSSIM 和 MSGMS，本文方法在复杂场景下实现了更精确的图像重建与异常检测，提升了模型的鲁棒性和适应性。

(1)SSIM 损失函数

1、亮度对比度函数：

$$l_{(x,y)} = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \quad (3-2)$$

式中： μ_x 、 μ_y 分别表示输入图像和重构图像的所有像素值的平均值， c_1 为常数。

2、对比相似度函数：

$$c_{(x,y)} = \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \quad (3-3)$$

式中： σ_x 、 σ_y 分别表示输入图像和重构图像的所有像素值标准差的平均值

3、结构相似度函数：

$$s_{(x,y)} = \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3} \quad (3-4)$$

式中： σ_{xy} 表示输入图像和重构图像的所有像素值协方差的平均值。

根据 SSIM 定义得可知：

$$SSIM(x, y) = l_{(x,y)} \cdot c_{(x,y)} \cdot s_{(x,y)} \quad (3-5)$$

则 SSIM 损失可以表示为：

$$L_{SSIM}(x, y) = \frac{1}{N} \sum_{m=1}^H \sum_{n=1}^W 1 - SSIM(x, y)_{(m,n)} \quad (3-6)$$

本文采用 $MSSSIM$ 损失可以表示为 $SSIM$ 图在 3 个尺度上的均值：

$$L_S(x, y) = \frac{1}{3} \sum_{k=1}^3 \frac{1}{N_k} \sum_{m=1}^{H_k} \sum_{n=1}^{W_k} 1 - SSIM(x_k, y_k)_{(m,n)} \quad (3-7)$$

式中： N_k 是尺度 K 下的像素个数， H_k 和 W_k 是图像的高度和宽度， x_k 和 y_k 分别是在尺度 K 下的输入图像和重构图像， $SSIM(x_k, y_k)_{(m,n)}$ 是 x_k 和 y_k 以像素 (m, n) 为中心的两个补丁之间的 $SSIM$ 结构相似度图的值。

(2)MSGMS 损失^[24]

$MSGMS$ 损失定义为 GMS 图在 3 个尺度上的均值：

$$L_G(x, y) = \frac{1}{3} \sum_{k=1}^3 \frac{1}{N_k} \sum_{m=1}^{H_k} \sum_{n=1}^{W_k} 1 - GMS(x_k, y_k)_{(m,n)} \quad (3-8)$$

式中： N_k 是尺度 K 下的像素个数， H_k 和 W_k 是图像的高度和宽度， x_k 和 y_k 分别是在尺度 K 下的输入图像和重构图像， $GMS(x_k, y_k)_{(m,n)}$ 是 x_k 和 y_k 像素 (m, n) 之间的 GMS 梯度幅度相似度图的值。

(3)总损失

总损失 L 由 $MSSSIM$ 损失、 $MSGMS$ 损失以及 L_2 损失相加构成：

$$L_{\text{总}} = W_1 L_S + W_2 L_G + L_2 \quad (3-9)$$

式中： W_1 和 W_2 表示权重参数。

(4)异常检测与定位得分计算

首先，输入图像和重建图像下采样到不同尺度 K ，利用公式(7)计算输入图像和重建图像不同尺度 K 下的 $SSIM$ 图。然后，将不同尺度 K 下的 $SSIM(x_k, y_k)$ 上采样到原始分辨率，并计算不同尺度 K 下以像素 (m, n) 为中心 $N \times N$ 滑动窗口的 $SSIM$ 图平均值，得到多尺度结构相似度 $MSSSIM$ 图。最后，采用大小为 s 均值滤波器对 $MSSSIM$ 图平滑处理，并用单位像素矩阵减去 $MSSSIM$ 图，得到异常评分图 $G(x, y)$ ，其中 $G(x, y)$ 里的像素值 $m, n \in (0, 1)$ ，异常评分图中像素值越接近 1 表示异常的可能性越大，公式如下：

$$G(x, y) = 1_{H \times W} - (MSSSIM * f_{s \times s}) \quad (3-10)$$

式中： $1_{H \times W}$ 是像素值为 1，且长为 H 、宽为 W 的像素矩阵， $f_{s \times s}$ 是卷积核为 $s \times s$

的均值滤波器。

使用掩码区域为 $f \times f$ 个像素对输入图片 n 次互补掩码并重构，然后计算异常图中每个 $f \times f$ 像素区域里所有像素的均值即：

$$G_v(x, y)_n = \frac{\sum G(x, y)_{fn}}{n} \quad (3-11)$$

式中： $G(x, y)_{fn}$ 是由公式 10 中掩码区域为 $f \times f$ 像素产生的第 n 个异常图。

综上所述，第 n 个重构图的图像级异常得分通过取 $G_v(x, y)_n$ 的最大值来计算即：

$$S_n = \max(G_v(x, y)_n) \quad (3-12)$$

(5) 重构图融合算法

由于单次掩码重构图像可能存在误差或仅从单一角度重建图像信息，为了综合评估多次掩码重构结果并降低单个掩码重构图像的随机性，本文设计了一种重构图像融合算法。首先，通过公式(12)计算每幅重构图像异常得分占总异常得分的权重，随后将每幅重构图像与其对应的权重相乘并累加，从而得到融合后的最终重构图像。这种方法不仅有效减少了单次掩码重构的随机误差，还通过加权融合提升了异常检测的稳定性和准确性，为复杂场景下的图像分析提供了更可靠的解决方案。公式如下：

$$I_r = \sum_{i=1}^n \left(I_{ri} \odot \frac{S_i}{\sum_{i=1}^n S_i} \right) \quad (3-13)$$

式中： I_{ri} 表示第 i 个重构图像， S_i 表示第 i 个重构图像的图像级异常得分。

3.2 数据集和评价指标介绍

3.2.1 数据集介绍

MVTec AD 是一个模拟真实工业场景的图像数据集，涵盖了 15 种不同的图像类别，其中包括 5 种复杂纹理数据集和 10 种实物表面数据集，每个类别均包含多种类型的缺陷样本^[51]。如图 3-5 所示，本文重点研究复杂纹理和实物表面的

工业图像，复杂纹理部分选用了木材和毛毯数据集，分别包含 240 张和 270 张正常训练样本，以及 75 张和 110 张异常测试样本。实物表面部分则选用了瓶子和螺母数据集，分别包含 200 张和 210 张正常训练样本，以及 80 张和 110 张异常测试样本。在训练和测试过程中，所有输入图像均被裁剪为 256×256 像素，以确保数据的一致性和模型的高效处理。通过对这些多样化的数据集进行深入研究，本文旨在提升复杂工业场景下的异常检测精度与鲁棒性。

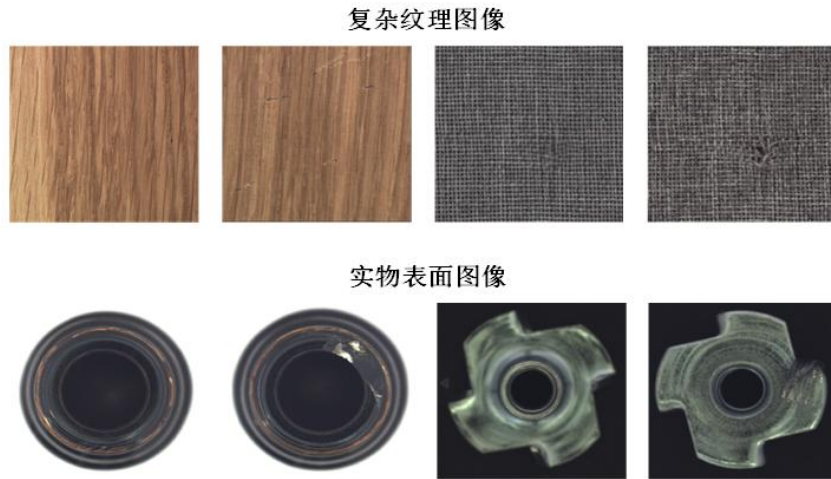


图 3-5 MVTec AD 数据集图像示例图

3.2.2 评价指标介绍

文采用 AUC 值作为异常检测模型的性能评价指标。AUC 值是 ROC 曲线与横坐标轴围成的面积，其中 ROC 曲线以假阳性率（FPR）为横坐标、真阳性率（TPR）为纵坐标绘制而成。通过设置不同的阈值，计算各阈值下的 FPR 和 TPR 值，进而绘制 ROC 曲线并计算 AUC 值。AUC 值越接近 1，表明模型的异常检测性能越优异^[52]。

具体而言，首先计算重构图像中所有像素的异常得分，根据的阈值设置，绘制所有像素异常得分的 ROC 曲线，并计算曲线与横坐标轴围成的面积，作为像素级 AUC 得分，用于评估异常定位的精度。然后，将每幅图像中最大的异常得分作为该图像的全局异常得分，绘制每一个图像异常得分的 ROC 曲线并计算 AUC 值，作为异常检测的整体性能评价指标。这种方法不仅能够全面评估模型在像素级和图像级的检测能力，同时也增强了模型在复杂场景中的适用性与鲁棒性。

3.3 实验结果与分析

3.3.1 实验平台和参数设置

本文实验环境基于 Intel CPU i7-12700 处理器，配备 64GB 内存，并采用 NVIDIA RTX 3090 GPU 进行加速计算，CUDA 版本为 11.6，操作系统为 Windows 11。深度学习框架基于 Python 3.9.15 和 PyTorch 1.12.1 构建。本文采用结构化重建模型进行异常检测与定位，如图 4 所示。在训练阶段，模型训练次数设置为 120 轮，批量处理大小（Batch Size）为 8，优化器选用 Adam 函数，损失函数权重 w_1 和 w_2 均设为 1，初始学习率为 0.001。此外，为了提升模型的鲁棒性，训练过程中引入了学习率衰减策略，以动态调整优化过程。最后，模型采用大小为 21 的均值滤波器对 MS-SSIM 图进行平滑处理，以进一步优化检测结果。这一配置确保了实验的高效性和结果的可靠性。

3.3.2 消融实验分析

(1) 损失函数消融实验

为了验证不同损失函数对模型性能的影响，设计了几组消融实验，如表 3-1 所示，模型训练时，方法 1 和方法 2 分别用均方误差 L_2 、多尺度梯度幅度相似性函数 L_S 和多尺度结构相似性函数 L_G 作为损失函数时，平均异常检测率分别为 86.1% 和 92.8%。方法 3 叠加以上三种损失函数作为损失函数时，平均异常检测率为 94.9%，比以上两种方法分别提高 8.8% 和 2.1%。综上所述，本文设计的损失函数提高了异常检测的准确率，模型性能有了明显改善。

表 3-1 损失函数消融实验（单位：%）

方法\模块	L_2	$L_G + L_S$	木材	毛毯	瓶子	螺母	平均
1	√	-	88.5	81.2	91.4	83.2	86.1
2	-	√	93.3	90.4	97.6	89.7	92.8
3	√	√	93.3	93.8	98.7	93.7	94.9

如图 3-6 所示，当使用正常瓶子作为训练样本时，仅采用均方误差 L_2 作为损

失函数，模型生成的重构图像较为模糊，细节丢失明显。然而，当引入多尺度梯度幅度相似性函数（MSGMS）和多尺度结构相似性函数（MSSSIM）作为损失函数时，重构图像的清晰度明显提升，内部纹理结构得到明显改善。当将上述三个损失函数叠加使用时，模型生成的重构图像在纹理细节上最为清晰，且与输入图像的相似度最高。实验结果表明，本文设计的损失函数能够有效优化重构图像的质量，提升了模型的重构性能，为复杂纹理和细节的保留提供了可靠的支持。

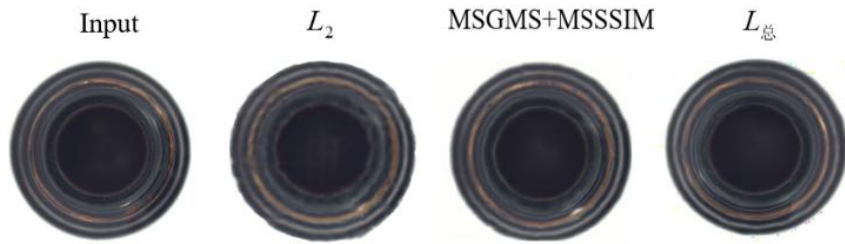


图 3-6 不同损失函数的模型重构性能对比图

(2)掩码次数对比实验

为了探究多次互补掩码融合对测试图片异常检测的影响，对测试图像进行 $n \in (1, 2, 3, 4)$ 次掩码。如表 3-2 所示，对 4 种测试图像进行 $n=1$ 次掩码，并直接对单个掩码重构，平均异常检测率为 80.9%。然后再对 4 种测试图像进行 $n \in (2, 3, 4)$ 次掩码重构并融合，此时模型异常检测准确率分别为 89.5%、93.3%、94.4%，均高于单次掩码测试图像时模型异常检测准确率。实验结果表明，本文提出的掩码方法和融合方法可以提高模型异常检测准确率，进一步提高了模型异常检测的性能。

表 3-2 掩码次数对比实验（单位：%）

种类\ n	1	2	3	4
木材	80.1	87.4	93.1	92.8
毛毯	78.7	84.7	89.1	93.3
瓶子	88.3	97.4	99.7	98.2
螺母	76.5	88.4	91.2	93.2
总平均	80.9	89.5	93.3	94.4

如图 3-7 所示，4 种测试图片在 $n \in (1, 2, 3, 4)$ 次掩码下模型生成的异常

评分图，红色越深，得分越高，代表该位置属于缺陷的置信度越高。单次掩码生成的异常图虽然能够定位缺陷区域，但缺陷周围的正常区域也会出现深浅不一的红色响应，导致正常区域与异常区域之间的缺陷置信度得分差距减小。这表明单次掩码重构图像可能存在误差或仅从单一角度重建图像信息，从而降低了缺陷定位的准确性。相比之下，通过多次掩码重构并融合生成的异常图中，缺陷区域的得分高于周围正常区域，两者之间的置信度得分差距明显增大。这种多掩码融合策略有效减少了单个掩码重构结果的随机性，提升了缺陷定位的精确度。实验结果表明，多次掩码重构方法能够更可靠地识别异常区域，为工业图像异常检测提供了更精确的解决方案。

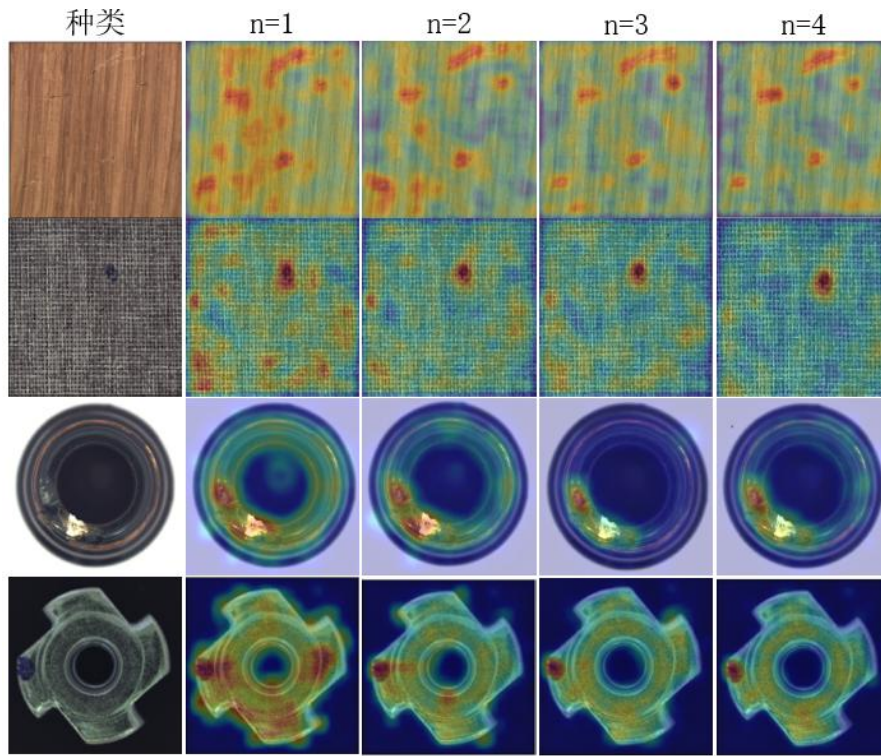


图 3-7 不同互补掩码的缺陷可视化对比图

3.3.3 异常检测与定位分析

(1) 异常检测结果分析

模型在 MVTEC AD 数据集上的异常检测结果如表 3-3 所示，异常检测结果以 AUC 评分为评估指标。为了更好地验证结构化重建模型的性能，将该方法与 MOCCA^[53]、DCAE^[54]、MSSCL^[55]和 RIAD^[24]方法在 MVTEC AD 数据集上进行异常检测对比。

表 3-3 异常检测方法实验结果比较（单位：%）

种类	MOCCA	DCAE	MSCL	RIAD	Ours
木材	87.6	89.2	82.6	92.8	93.1
毛毯	96.6	86.4	92.6	92.4	93.3
瓶子	95.4	92.3	88.4	99.2	99.7
螺母	86.2	78.1	89.4	88.6	93.2
Avg-T	92.1	87.8	87.6	92.6	93.2
Avg-O	90.8	85.2	88.9	93.9	96.5
总平均	91.5	86.5	88.3	93.3	94.9

如表 3-3 所示，本文方法的平均 AUC 值为 94.9%，对比的 4 种算法中最高的 RIAD 高了 1.6%，其中在纹理数据集 Avg-T（木材、毛毯）中提高了 0.6%，在实物数据集 Avg-O（瓶子、螺母）中提高了 2.6%。根据以上算法实验结果比较，本文算法优于以上 4 种算法。

(2)异常定位结果分析

模型在 MVTec AD 数据集的异常定位结果如表 3-4 所示，异常定位结果以 AUC 评分为评估指标，为了更好地验证结构化重模型的性能，将该方法与 CAVGA^[56]、AE-SSIM^[22]、UTAD^[57]和 RIAD^[24]方法在 MVTec AD 数据集上进行异常定位对比。

表 3-4 异常定位方法实验结果比较（单位：%）

种类	CAVGA	AE_{SSIM}	UTAD	RIAD	Ours
木材	88.6	74.4	83.2	87.2	90.2
毛毯	94.2	88.2	97.1	96.2	98.6
瓶子	91.4	93.4	98.2	98.2	98.4
螺母	85.8	89.7	85.6	94.1	95.2
Avg-T	91.4	81.3	90.2	91.7	94.4
Avg-O	88.6	91.6	91.9	96.2	96.8
总平均	90.0	86.5	91.1	93.9	95.6

如表 3-4 所示，本文方法的异常定位 AUC 值达到 95.6%，对比的 4 种算

法中最高的 RIAD 高了 1.7%，其中在纹理数据集 Avg-T（木材、毛毯）中提高了 2.7%，在实物数据集 Avg-O（瓶子、螺母）中提高了 0.6%。根据以上算法实验结果比较，本文算法优于以上 4 种算法。

异常定位可视化图如图 3-8 所示，其中颜色表示缺陷置信度得分，红色越深，得分越高，代表该位置属于缺陷的置信度越高。

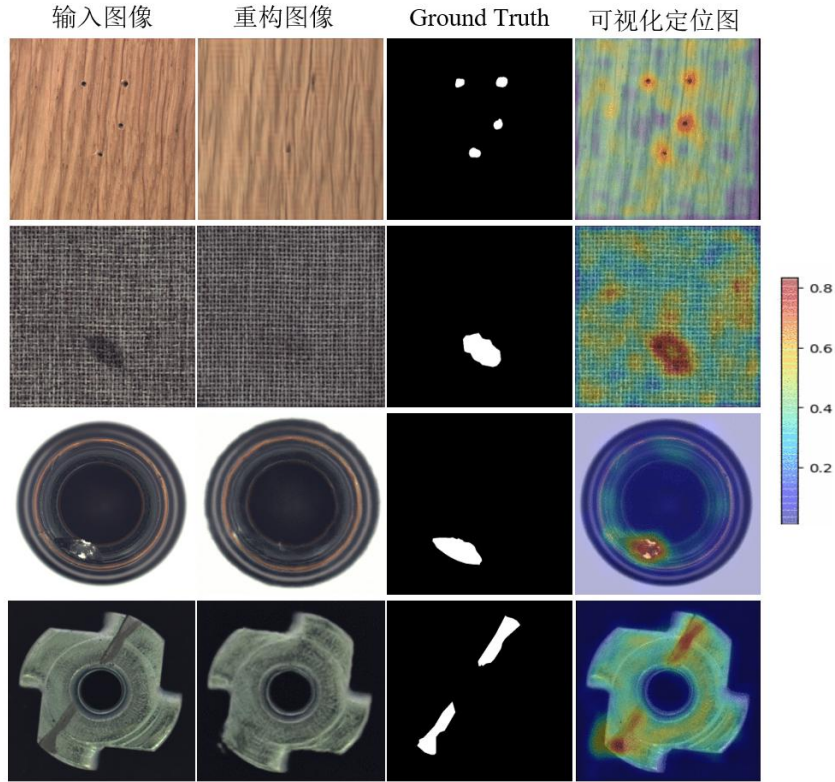


图 3-8 异常定位可视化图

3.4 本章小结

本文针对工业图像缺陷检测中训练样本不足和漏检率高等问题，提出了一种基于自监督结构化重建的工业图像异常检测与定位模型。首先，在模型训练阶段，通过引入多尺度结构相似度损失（MS-SSIM）和多尺度梯度幅度相似性损失（MSGMS），提升了模型的重建性能，使其能够更好地捕捉图像的细节和结构信息。然后，本文提出了一种多互补掩码融合方法，通过多角度遮挡和融合策略，减少了单次掩码重构的随机误差，有效提高了异常检测和定位的准确率。最后，实验结果表明，该模型在 MVTec AD 数据集上的平均异常检测率达到 94.9%，比

对比算法中性能最优的模型提升了 1.6%；同时，平均异常定位率达到 95.6%，对比算法中性能最优的模型提高了 1.7%，验证了该模型在工业图像异常检测与定位任务中的可行性和有效性。随着工业技术的快速发展，未来需要进一步优化和压缩结构化重建模型，以提升其在更复杂工业场景下的检测精度和运行效率。

第4章 基于自监督深度记忆特征嵌入的工业图像异常检测与定位

在前一章中,本文介绍了一种自监督结构化重建的工业图像异常检测与定位模型前,该方法在缺陷识别和定位精度方面表现出色。然而,由于采用了多层次特征提取、多尺度特征融合以及多级特征重建等复杂模块,导致模型参数量大幅增加,计算复杂度显著提升,进而影响了检测效率。鉴于工业检测环境对实时处理能力的严苛要求,同时受限于部署设备的计算能力和功耗约束^[58],如何在维持较高检测准确率的前提下有效降低计算成本,这已成为当前研究中迫切需要解决的核心技术难题。因此,本章提出了一种深度记忆特征嵌入模型,通过引入轻量化设计思想和特征压缩机制,在保持较高检测准确率的前提下提升了计算效率。与基于图像重建的方法相比,该方法具有以下优势:

(1)特征嵌入方法直接利用预训练网络提取高级特征,避免了复杂的图像重建过程,降低了计算开销。

(2)图像重建方法需要生成完整的重建图像,而特征嵌入方法通过提取图像的高级特征并直接比较特征之间的差异来实现异常检测,从而避免了重建过程中可能引入的误差。

(3)图像重建方法通常依赖复杂的网络结构生成高质量重建图像,而特征嵌入方法采用轻量化的特征提取网络,降低了模型复杂度,进一步提升了检测速度和效率。

然而,基于特征嵌入的方法虽然减少了训练时间,但在推理阶段需要进行复杂的特征匹配操作,增加了计算成本。此外,由于模型训练未使用针对异常的数据集,而是直接采用预训练参数进行特征提取和异常检测,其对异常检测任务的适应性仍有不足^[59]。

针对上述问题,本文设计了一种基于自监督学习的深度记忆特征嵌入模型。首先,通过模拟异常策略生成异常样本,有效解决了工业图像中异常样本稀缺的问题。然后,利用正常样本和模拟异常样本进行训练,并引入深度记忆特征嵌入模块,通过对比输入样本与记忆模块中存储的正常样本,提取两者之间的高级特

征差异信息，从而加速模型训练过程。最后，为了进一步提升模型的检测能力，本文设计了多尺度特征融合模块，充分融合不同层次特征的上下文语义信息，增强模型对复杂场景的适应能力。

4.1 深度记忆特征嵌入异常检测模型

如图 4-1 所示，本文设计了深度记忆特征嵌入的异常检测模型。模型训练时引入异常模拟图像和正常图像，实现自监督学习，使模型获得正常图像和异常图像之间的更具鲁棒性的决策边界。首先，输入图像经过编码器提取三层特征，然后，计算输入图像的特征信息与所有记忆信息 M_i 之间的 L_2 距离和结构相似度 $SSIM$ 距离，获得输入图像深度特征与所有记忆信息 M_i 之间的距离信息，寻找最小距离信息和输入图像特征进行同尺度拼接。最后，将拼接信息经过多尺度特征融合模块进行特征融合，融合特征通过跳连接流向解码器逐层解码，得到输入图像异常区域的预测图像^[60]，并采用感知损失和焦点损失作为损失函数，使其预测值更加接近其真实标签。

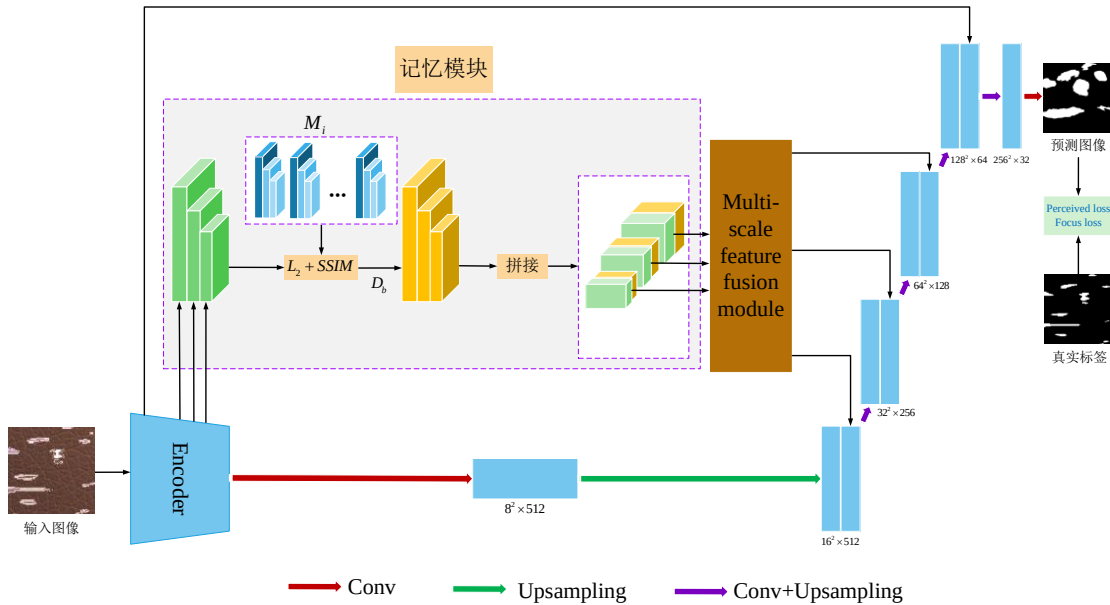


图 4-1 深度记忆特征嵌入的异常检测模型流程图

4.1.1 异常模拟策略

在工业场景中，异常以各种形式出现，在进行数据收集时不可能覆盖所有的

异常，为了帮助模型更全面地学习到各种特征，而不仅仅依赖于真实数据中的少量异常，设计了一种更有效的策略来模拟异常样本，该方法通过引入多样化的异常情况，丰富训练数据，解决数据不平衡问题，加速模型收敛，并增强特征学习的有效性^[61]。

异常模拟设计如下：

(1)纹理异常模拟设计：

如图 4-2 所示，生成二维柏林噪声，然后通过阈值 T 二值化得到掩码。将二值掩码图像和纹理图像进行逐元积，得到纹理感兴趣区域(ROI)图像，并引入透明因子提高掩码图像和纹理图像的融合效果^[62]，使模拟异常设计更接近真实异常即：

$$M = \rho \times (M_p \odot I_t) \quad (4-1)$$

式中 M_p 、 I_t 表示二值化掩码和纹理图像， ρ 表示透明度因子。

然后再将 ROI 掩码图像叠加到原始图像得到模拟纹理异常图像即：

$$I_a = M + I \quad (4-2)$$

式中 M 和 I 表示纹理感兴趣区域(ROI)图像和输入图像。

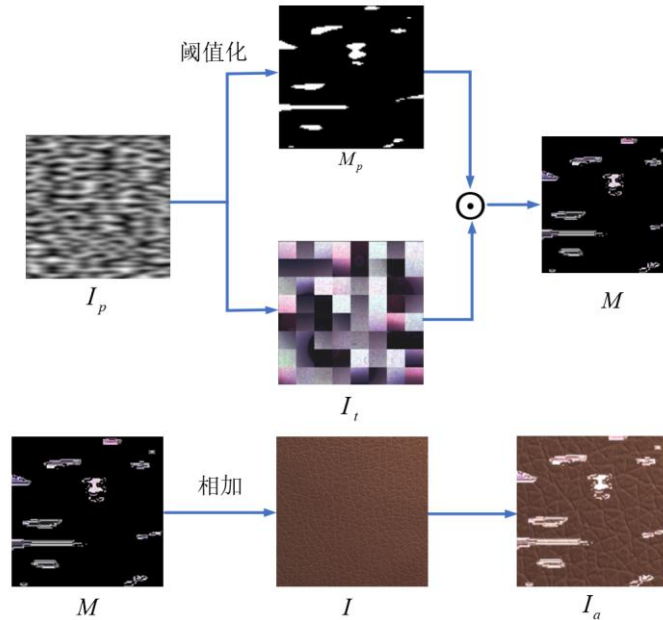


图 4-2 纹理异常模拟技术流程图

(2)结构异常模拟设计：

如图 4-3 所示，CutPaste 数据增强是从正常样本中切割出一个小的矩形区域，

并将其粘贴在附近相邻的位置（黄色虚线框），构造模拟结构异常图像^[63]。

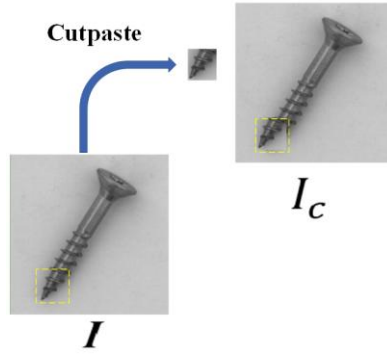


图 4-3 结构异常模拟技术流程图

4.1.2 深度记忆特征嵌入模块

模型引入了深度记忆特征嵌入模块，获得输入样本和记忆模块中正常样本之间特征的差异信息，提高模型对异常区域的有效猜测。

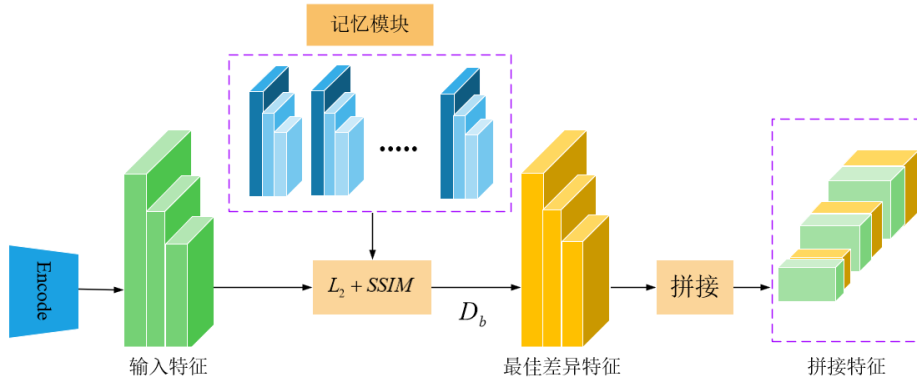


图 4-4 记忆模块

如图 4-4 所示，首先，在模型初始化阶段，从正常样本集中随机选取图像作为记忆模块，从 DenseNet-121 网络中提取 $N \times 64 \times 64 \times 64$ 、 $N \times 128 \times 32 \times 32$ 和 $N \times 256 \times 16 \times 16$ 三个尺度的特征表示，这些多分辨率特征被固定并构成记忆模块 M_i 。然后，在训练过程中，编码器对输入图像进行特征提取，获得 $64 \times 64 \times 64$ 、 $128 \times 32 \times 32$ 和 $256 \times 16 \times 16$ 三个层次的特征表示，共同形成输入特征 I 。其次，通过计算输入特征 I 与所有记忆模块 M_i 之间的 L_2 距离和结构相似性度量 (SSIM)，可以获得输入图像与正常样本之间的差异信息即：

$$D = \bigcup_{i=1}^N \left\{ \left(\frac{1}{K} \sum_{m=1}^H \sum_{n=1}^W 1 - SSIM(M_i, I)_{(m,n)} \right) + \|M_i - I\|_2 \right\} \quad (4-3)$$

式中： N 为记忆样本个数， H 和 W 是图像的高度和宽度， K 为图像中像素总和， $SSIM(M_i, I)_{(m,n)}$ 是 M_i 和 I 以像素 (m, n) 为中心的两个补丁之间的 $SSIM$ 结构相似度的值。取最小值为最佳差异信息 D_b 即：

$$D_b = \arg \min_{D_i \in D} \sum_{d \in D_i} d \quad (4-4)$$

其中 $i \in [1, N]$ ，最佳差异信息 D_b 是输入图像与最相似的记忆样本之间的差异，一个位置差值越大，则该位置对应的输入图像区域异常的概率越高。

经过上述操作，最佳差异信息 D_b 与之前提取的三个维度特征相拼接，得到最后的拼接特征。

4.1.3 多尺度特征融合模块

传统的多尺度特征融合模块虽然在提取多尺度特征方面具有一定效果，但仍存在一些局限性。首先，传统方法通常采用简单的特征拼接方式，导致融合过程中容易出现信息丢失或冗余问题。其次，传统方法忽略不同尺度特征之间的差异性，未能有效平衡局部细节与全局上下文信息。

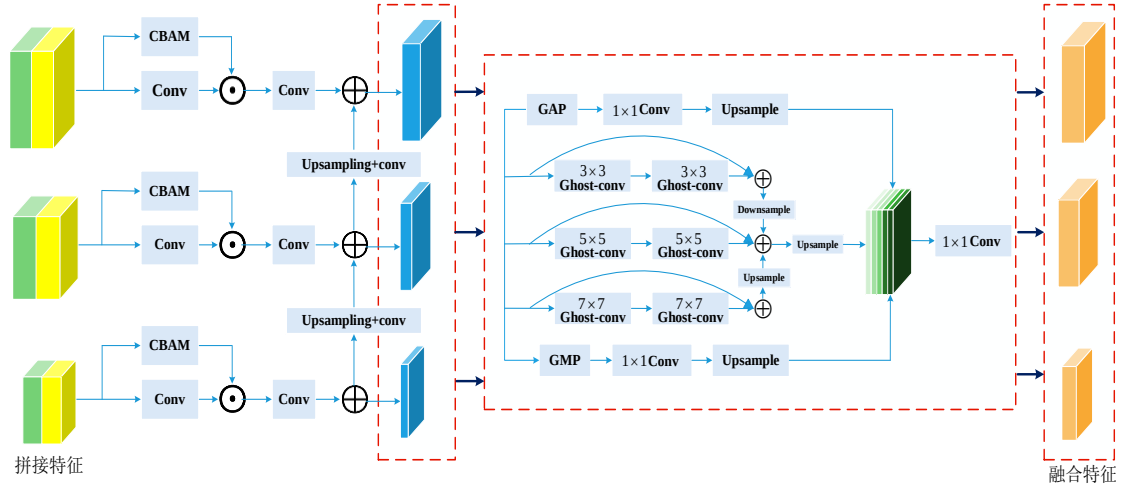


图 4-5 多尺度特征融合模块

针对上述问题，提出了一种多尺度融合模块。如图 4-5 所示，该模块分为两个阶段的融合过程：第一阶段是记忆模块中拼接特征的初步融合，第二阶段是对融合特征进行多卷积和池化操作的深度融合。首先，在不同尺度的拼接特征中引入坐标注意力机制（CA），用于同时捕获通道和空间信息，并通过分解和编码增强模型对目标位置边缘细节的精确感知能力。其次，通过多次卷积操作，将低分

分辨率拼接特征逐一上采样并与高分辨率拼接特征进行融合^[64]。然后，这些融合特征经过全局均值池化（GAP）和全局最大池化（GMP）处理。其中，GAP 能够反映整体特征分布，适合捕捉全局上下文信息；而 GMP 则能够提取特征图中最显著的区域，强调局部细节信息。通过同时使用 GMP 和 GAP，模块能够有效结合局部显著特征和全局上下文信息，进一步增强特征表达能力并提升模型的鲁棒性。最后，利用1×1卷积调整通道数，生成聚合多尺度语义信息的特征图。

4.1.4 损失函数

为了保证模型的预测值在条件约束下接近真实标签，本文采用感知损失和焦点损失训练模型，感知损失可以缓解了图像中正常区域与异常区域面积不平衡的问题，使模型更加关注难样本的分割，提高异常分割的准确率^[16]。焦点损失^[65]则可以捕捉图像的高级特征，能够很好地保留图像的纹理和局部的细节，并且可以生成更加真实、自然和高质量的图像。

(1)感知损失函数：

$$L_p = \sum_{i=1}^N \frac{1}{N_i} \left\| \phi_i(y) - \phi_i(\tilde{y}) \right\|_1 \quad (4-5)$$

式中： $\Phi_i(y)$ ， $\Phi_i(\tilde{y})$ 分别表示真实标签第*i*层特征图和预测图像的第*i*层特征图。

(2)焦点损失函数：

$$L_f = \begin{cases} -\alpha(1-\tilde{y})^\gamma \log \tilde{y}, y=1 \\ -(1-\alpha)\tilde{y}^\gamma \log(1-\tilde{y}), y=0 \end{cases} \quad (4-6)$$

式中： y 表示真实值， $\tilde{y}$ 表示该类别的预测值。

(3)总损失：

$$L_{all} = w_1 L_p + w_2 L_f \quad (4-7)$$

式中： w_1 和 w_2 表示权重参数。

4.1.5 异常检测得分

在图像异常检测与定位的过程中，每个测试图像都会产生一个像素级的异常得分图。该异常得分图每个像素值对应模型预测此像素属于异常的概率。针对图

像级的异常检测，将异常得分图中最异常的前 20 个像素点分数平均值作为图像级异常分数。然后，计算每个阈值下的 TPR、FPR、精确率和召回率，最终的异常判别阈值 T 取决于使精确率和召回率达到平衡状态的最大值。而在检测阶段，如果测试图像的得分大于异常判别阈值 T ，则被认为是异常样本。为了定位图像中的异常区域，可以通过选取合适的分割阈值 T_1 将异常得分图进行二值化展示其定位效果。

4.2 数据集和评价指标介绍

4.2.1 数据集介绍

为了验证本文所提算法在工业图像异常检测与定位任务中的性能表现，本研究选取了 MVTec AD 和 BTAD 两个具有代表性的工业数据集进行实验验证。其中 MVTec AD 数据集样本采集自真实工业环境，具有较高的应用价值^[51]。如图 4-6 所示，该数据集包含多种工业场景：复杂纹理类别选取瓷砖和皮革为代表，工业零件类别选取螺钉和螺母为代表。该数据集的训练集仅包含正常样本，而测试集则同时包含正常和多种缺陷类型的异常样本。此外，测试集不仅提供了图像级分类标签，还包含像素级精度的异常区域标注，这使得 MVTec AD 成为评估异常检测与定位性能的理想基准。

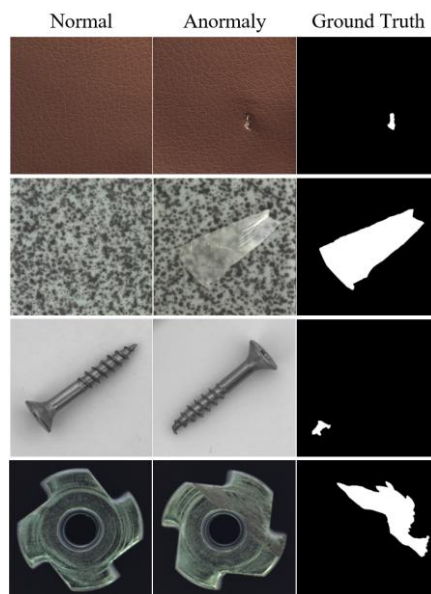


图 4-6 MVTec AD 数据集示例图

BTAD 数据集^[66]包含三类工业场景图像，样本总量达到 2830 张。此外，该数据集的设计充分考虑了工业检测的实际需求，具有较高的应用价值。与 MVTec AD 类似，该数据集的训练集仅包含正常样本，而测试集则同时包含正常和异常样本，并提供了图像级分类标签和像素级标注信息。如图 4-7 所示，展示了 BTAD 数据集的样本示例：第一列为正常样本，第二列和第三列分别展示了异常样本及其对应的像素级标注。

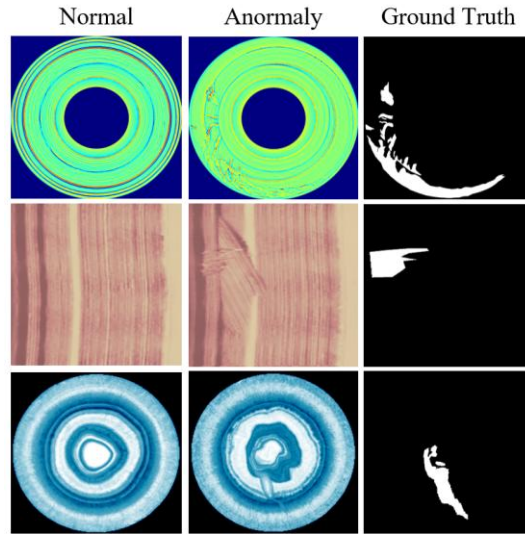


图 4-7 BTAD 数据集示例图

4.2.2 评价指标介绍

本研究采用像素级和图像级的 AUC 值作为模型性能的评价指标。在像素级评估中，通过计算输入图像中所有像素点的异常得分，绘制不同阈值下的 ROC 曲线，并计算曲线下面积 AUC 来评估异常定位的精度。对于图像级评估，则以每幅图像中异常得分最高的 20 个像素点的均值作为图像级异常得分，进而绘制 ROC 曲线并计算 AUC 值，用于衡量模型的整体检测性能。

在实验过程中，首先需要计算混淆矩阵中的假阳性率（FPR）和真阳性率（TPR）。此外，这些指标能够全面反映模型在不同阈值下的性能表现，为算法优化提供重要依据。然后，通过调整判别阈值，并以 FPR 为横轴，TPR 为纵轴构建 ROC 曲线。曲线下面积 AUC 越大，表明模型的查准率越高且误检率越低，检测性能越优异。在具体计算过程中，模型分别以图像和像素为单位计算 TPR 和 FPR，根据设定的异常阈值绘制 ROC 曲线，并最终计算得到对应的曲线下面积。

4.3 实验细节及结果分析

4.3.1 实验平台和参数设置

为了验证本文模型的性能，并搭建相应的实验平台，实验环境基于 Inter CPU i7-12700 处理器，内存为 64GB，显卡采用 NVIDIA RTX3090 GPU，CUDA 版本为 11.6，操作系统为 Windows11，深度学习框架基于 Python3.9.15、Pytorch1.12.1。本文采用深度记忆特征嵌入模型进行异常检测与定位，如图 1 所示，对于解码器部分，上采样层包含双线性插值层、由卷积层、批处理归一化和 ReLU 激活函数。只有最后一个卷积层包含一个 2 通道卷积层。在训练阶段，设置模型训练次数为 600 轮，输入图像的大小设置为 256×256 ，批量处理 Batch 为 8，其中包含 4 个正常样本和 4 个模拟异常样本，深度学习优化器选择 Adam 函数，差异权重 α_1 、 α_2 为 1，损失函数权重 W_1 、 W_2 为 1，初始学习率为 0.05。

4.3.2 消融实验分析

(1) 损失函数和多尺度特征融合模块消融实验

为了验证不同损失函数和多尺度特征融合模块对模型性能的影响，本文在 MVTec AD 数据集上进行了几组消融实验，如表 4-1 所示，模型训练过程中，方法 1 和方法 2 分别用感知损失 L_p 、感知损失 L_p 叠加焦点损失 L_f 作为损失函数时，平均异常定位率分别为 87.4% 和 93.6%。方法 3 在以上基础上引入了多尺度特征融合模块，平均异常检定位为 96.7%，比以上两种只使用损失函数的方法分别提高 9.3%、3.1%。综上所述，本文设计的损失函数和多尺度特征融合模块提高了异常定位的准确率，模型性能有了明显改善。

表 4-1 不同模块对实验（单位：%）

方法	L_p	L_f	$MSFF$	皮革	瓷砖	螺钉	螺母	平均
1	√	-	-	89.5	89.3	86.4	84.2	87.4
2	√	√	-	93.3	94.7	94.6	91.8	93.6
3	√	√	√	96.8	97.8	95.4	96.7	96.7

如图 4-8 所示，以螺母为例，使用感知损失 L_p 训练过程中，由于异常模拟

样本中异常区域占图像的比例很小，可能造成像素级上的训练数据不平衡，导致分割图像中异常区域边缘模糊。当使用感知损失 L_p 叠加焦点损失 L_f 训练模型，使模型更加关注训练样本中的异常区域，分割图像中异常区域边缘更加精确，进一步提高了异常定位精确度。当模型在之前损失函数训练基础上引入多尺度特征融合模块，增强不同尺度信息的特征表示，提供更丰富的上下文信息，此时分割图像中异常区域边缘最清晰，异常定位精确度最高。

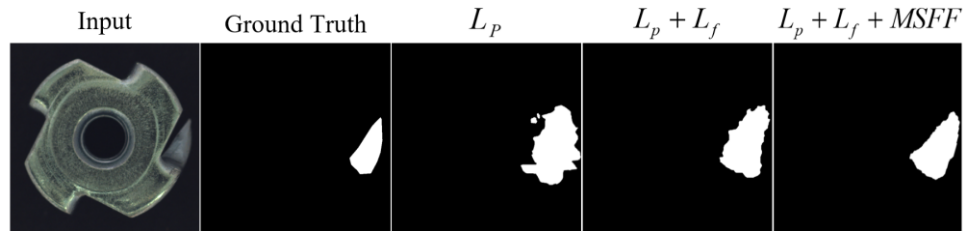


图 4-8 不同损失函数的异常分割对比图

4.3.3 异常检测分析

(1) MVTEC AD 数据集结果与分析

为了更好地验证深度记忆特征嵌入模型的异常检测性能，将该方法与 RIAD^[24]、CusPaste^[63]、SVDD^[67]和 InTra^[68]方法在 MVTEC AD 数据集上进行异常检测对比。如表 4-2 所示，列出了每个方法在各个类别上的图像级 AUROC，2 个纹理类和 2 个工业零件类的平均图像级 AUROC，以及整体的图像级 AUROC。

表 4-2 异常检测方法实验结果比较（单位：%）

类/方法	RIAD	CusPaste	SVDD	InTra	Ours
皮革	93.1	96.4	96.2	96.4	97.3
瓷砖	96.3	93.1	92.9	97.1	98.8
螺钉	83.3	91.3	92.1	95.7	95.4
螺母	89.2	94.5	93.8	96.4	96.9
Avg-T	94.7	94.8	94.5	96.7	98.1
Avg-O	86.2	92.9	92.9	96.1	96.2
总平均	90.5	93.8	93.7	96.4	97.1

如表 4-2 所示，本文方法的平均 AUC 值为 97.1%，与其他方法相比，无论

是在纹理类还是工业零件类上都达到了最佳的图像级 AUROC。对比其他 4 种算法中最高的 InTra 高了 0.7%，其中在纹理类 Avg-T（皮革、瓷砖）和工业零件类 Avg-O（螺钉、螺母）中分别提高了 1.4%和 0.1%。实验结果表明，本文算法异常检测性能优于以上 4 种算法，能够更好地满足工业图像异常检测的需求。

(2)BTAD 数据集结果与分析

为了进一步验证模型在不同工业场景中的异常检测能力，将该方法与 CusPaste^[63]、SVDD^[67]和 VT-ATL^[69]方法在 BTAD 数据集上进行异常检测对比。如表 4-3 所示，列出了每个方法在各个产品上的图像级 AUROC 以及整体的图像级 AUROC。

表 4-3 异常检测方法实验结果比较（单位：%）

类/方法	VT-ATL	CusPaste	SVDD	Ours
产品 1	93.8	93.4	96.2	97.3
产品 2	91.1	91.3	94.5	96.8
产品 3	80.8	92.3	85.8	97.7
总平均	88.6	92.4	92.2	97.3

如表 4-3 所示，本文方法的平均 AUC 值为 97.3%，较其他三种算法中表现最好的 CusPaste 提升了 4.9%。实验结果表明，所提出的算法具有更强的鲁棒性与泛化能力，在不同工业场景下均表现出稳定的检测性能，能够更好地满足工业图像异常检测的需求。

4.3.4 异常定位分析

(1)MVTec AD 数据集结果与分析

为了更好地验证深度记忆特征嵌入模型的异常定位性能，将该方法与 RIAD^[24]、SVDD^[67]和 InTra^[68]方法在 MVTec AD 数据集上进行异常定位对比。如表 4-4 所示，详细列出了各方法在不同类别上的像素级 AUROC 值，包括 2 个纹理类和 2 个工业零件类的平均像素级 AUROC，以及整体像素级 AUROC。此外，这些数据充分反映了算法在不同场景下的性能表现。

表 4-4 异常定位方法实验结果比较（单位：%）

类/方法	RIAD	SVDD	InTra	Ours
皮革	93.2	92.5	91.4	96.8
瓷砖	95.1	91.4	95.2	97.8
螺钉	86.7	90.3	94.9	95.4
螺母	87.6	91.3	92.8	96.7
Avg-T	94.2	91.9	93.3	97.3
Avg-O	87.2	90.8	93.8	96.1
总平均	90.7	91.3	93.5	96.7

如表 4-4 所示，本文方法异常定位平均 AUC 值为 96.7%，与其他方法相比，无论是在纹理类还是物体类上都达到了最佳的像素级 AUROC。比其他 3 种算法中最高的 InTra 高了 3.2%，其中在纹理 Avg-T（瓷砖、皮革）和工业零件类 Avg-O（螺钉、螺母）中分别提高了 4.0%和 2.3%。实验结果表明，本文算法异常检测性能优于以上 3 种算法。

为更直观地展示所提出方法的异常定位性能，本研究在 MVTec AD 数据集的部分类别上对比了不同方法的异常分割结果。

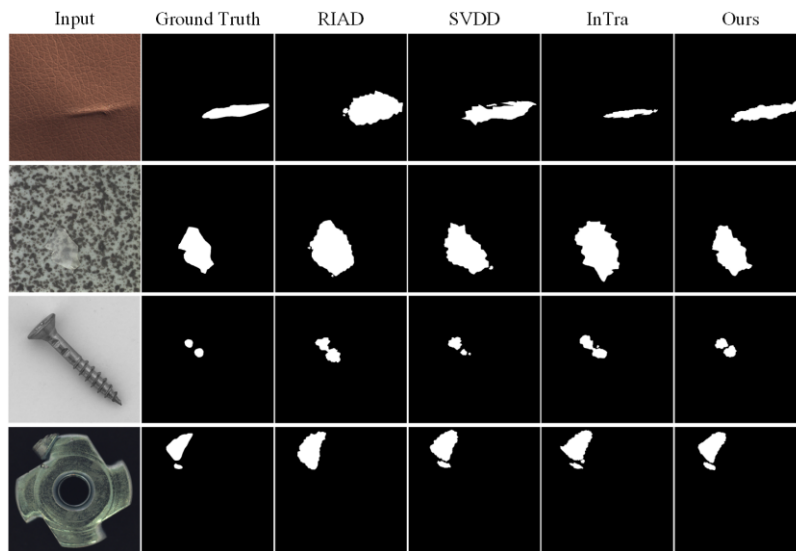


图 4-9 各种方法在 MVTec AD 数据集上的异常分割对比图

如图 4-9 所示，展示了皮革、瓷砖、螺钉和螺母四个类别在不同方法下的异

常分割效果。每行依次展示了测试图像、真实标注的异常区域,以及 RIAD、SVDD、InTra 和本文方法的异常分割结果。实验结果表明,本文方法在皮革、瓷砖和螺母三个类别上的异常定位精度明显优于对比方法,能够更准确地识别异常区域的边界和范围。但是,在螺钉类别上,虽然各方法的性能差距较小,但本文方法在细节处理方面仍展现出一定优势。这些对比实验充分验证了所提出方法在异常检测任务中的有效性和鲁棒性,为工业检测应用提供了可靠的技术支持。

(2)BTAD 数据集结果与分析

为了验证模型在不同工业场景中的异常定位性能,将该方法与 VT-ATL^[69]、CusPaste^[63]和 SVDD^[67]方法在 BTAD 数据集上进行异常定位对比。如表 4-5 所示列出了每个方法在各个产品上的像素级 AUROC 以及整体的像素级 AUROC。

表 4-5 异常定位方法实验结果比较 (单位: %)

类/方法	VT-ATL	CusPaste	SVDD	Ours
产品 1	89.8	93.5	94.9	98.8
产品 2	90.2	91.4	92.3	94.4
产品 3	79.8	90.3	91.5	93.4
总平均	86.6	91.7	92.9	95.6

如表 4-5 所示,本文方法的平均 AUC 值为 95.6%,对比其他 4 种算法中最高的 SVDD 高了 2.7%。实验结果表明,本文方法具有更强的鲁棒性与泛化能力,能够更好地满足工业图像异常定位质检需求。

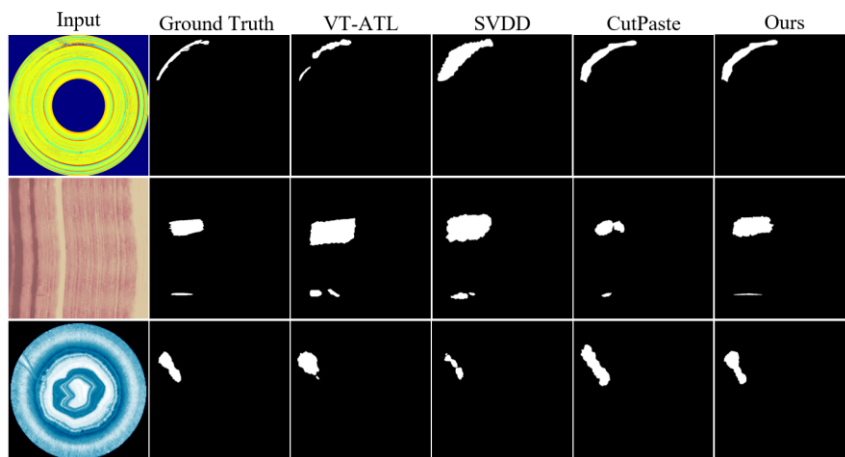


图 4-10 各种方法在 BTAD 数据集上的异常分割对比图

为了直观展示所提出方法的异常定位性能,本研究在 BTAD 数据集上对比

了不同方法的异常分割结果。如图 4-10 所示，展示了产品 01、02 和 03 三个类别在不同方法下的异常分割效果。每行依次展示了测试图像、真实标签，以及 VT-ATL、SVDD、CusPaste 和本文方法的异常分割结果。实验结果表明，本文方法在异常定位精度上优于对比的方法，能够更准确地识别异常区域的边界和范围。这些对比实验充分验证了所提出方法在异常检测任务中的有效性和鲁棒性^[69]。

4.4 本章小结

针对工业图像中缺陷训练样本少和检测效率低等问题，提出一种自监督深度记忆特征嵌入的工业图像异常检测与定位模型。首先，正常样本经过异常模拟策略构造异常样本，解决异常样本不充分问题。然后，提出深度记忆特征模块和多尺度特征融合模块，并引入感知损失和焦点损失进行训练，提高了异常检测与定位的准确性。最后，实验结果表明，本文方法在 MVTec AD 数据集和 BTAD 数据集上平均异常检测率分别为 97.1%和 97.3%，相较于对比的最好方法，分别提高了 0.7%和 4.9%；平均异常定位率分别为 96.7%和 95.6%，相较于对比的最好方法，分别提高了 3.2%和 2.7%，进一步提高工业图像异常检测与定位的准确性。随着计算机视觉的飞速发展，需要进一步将深度记忆特征嵌入模型进行优化压缩，实现快速精确的工业图像异常检测与定位。

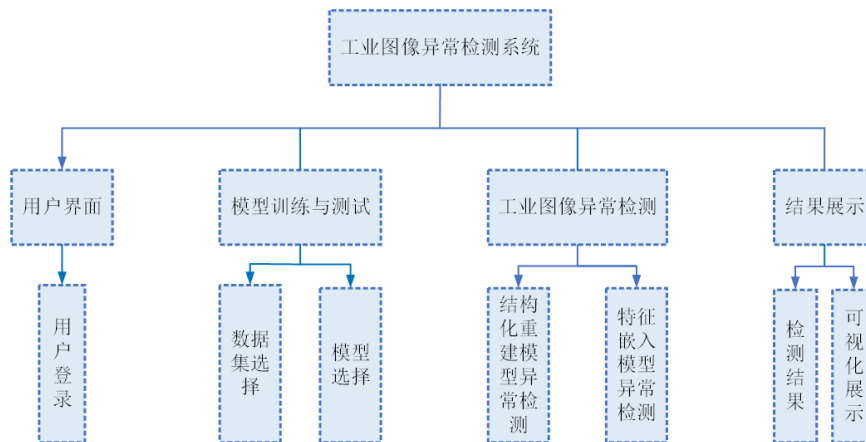
第 5 章 工业图像异常检测系统的设计与实现

基于深度学习的异常检测技术已成为当前图像分析领域的重要研究方向,该方法首先利用图像传感器采集目标物体表面信息,随后对获取的数字图像进行预处理和分析,以提取关键特征信息。尽管我国科技水平不断提升,但许多制造企业仍依赖人工进行产品表面缺陷检测,这种方法存在检测效率低、误检漏检率高以及人力成本昂贵等问题。在检测微小缺陷时,人工检测的准确性和可靠性难以保证。基于前文的理论分析,本章设计并实现了一个基于自监督学习的工业图像异常检测系统。该系统提供了便捷的用户界面,用户可通过输入设备与系统交互,执行图像输入、模型训练与测试、异常检测按键等操作,从而实现高效准确的表面缺陷检测。

5.1 异常检测系统框架设计

5.1.1 系统框架设计

本系统以 PyCharm 为开发工具,通过 Qt Designer 系统实现可视化界面的设计,系统的整体框架如图 5-1 所示。为了使图像异常检测系统的功能更为针对性、通用性更强,系统共分为三个模块,分别为用户界面、模型训练与测、工业图像异常检测和结果展示模块。



5-1 工业图像异常检测系统框架流程图

用户界面模块的核心功能之一是用户身份认证。为确保系统安全性和数据隐

私，只有通过身份验证的授权用户才能访问系统功能，这种设计有效保障了系统的合法性和数据安全性。

系统功能模块主要包含三个核心组件：模型训练与测试单元、工业图像异常检测单元以及结果可视化单元。各模块具体性能如下：

其中，模型训练与测试提供了数据集选择和模型配置功能，用户可以根据实际需求灵活选择数据集和模型架构，快速启动训练和测试流程。这种模块化设计不仅增强了系统的灵活性和易用性，还支持用户根据具体应用场景优化模型配置，显著提升了操作效率和检测精度，为工业图像异常检测提供了可靠的技术支持。

工业图像异常检测模块主要包括结构化重建模型异常检测和特征嵌入模型异常检测。用户可以根据需求，从平台中已上传的数据集中选择目标数据集，并灵活选择不同的异常检测模型进行检测与分析，为用户提供高效、便捷的操作体验，满足多样化的工业检测需求。

结果展示模块涵盖了检测结果的详细呈现与可视化展示功能。该模块能够将每次检测的结果清晰地展示出来，包括异常区域的热力图、分割图及统计信息，方便用户直观观察和分析检测细节，为用户提供全面、便捷的数据管理体验。

5.1.2 运行环境设置

针对不同应用场景和功能需求，系统所需的开发环境和运行环境存在显著差异。为确保工业图像异常检测系统的稳定运行和高效执行，必须配置相应的硬件和软件环境。根据实际应用需求，如表 5-1 所示，详细列出了该系统支持的操作系统版本及其最低运行环境配置要求。

表 5-1 系统开发环境配置表

设备	资源名称	配置
硬件	CPU	Inter CPU i7-12700
	GPU	NVIDV RTX3090
	内存	64GB
软件	操作系统	Windows11
	编程语言	C++、pythonn
	开发环境	Pyharm 2022

5.2 异常检测系统的 UI 界面设计

系统是基于自监督学习的工业图像异常检测方法为基础进行设计与实现的。以下将具体介绍系统各模块的设计过程。主界面中有多个功能，其中包括系统登录、功能选择、检测结果等功能。工业图像异常检测系统的界面具有图像显示区域，各操作按钮可进行相关的操作。

5.2.1 用户登录界面设计

为确保系统安全性和数据隐私保护，本系统采用了用户认证机制。如图 5-2 所示，用户登录界面提供了注册和登录功能。新用户需点击"立即注册"按钮，填写用户名和密码完成账户创建，而已注册用户则直接在登录页面输入凭证信息。该系统采用了加密存储技术，有效保障了用户信息安全。在输入完成后，用户需点击"登录"按钮进行身份验证。当输入的用户名和密码与数据库中的加密记录匹配时，系统将授权用户访问平台功能。

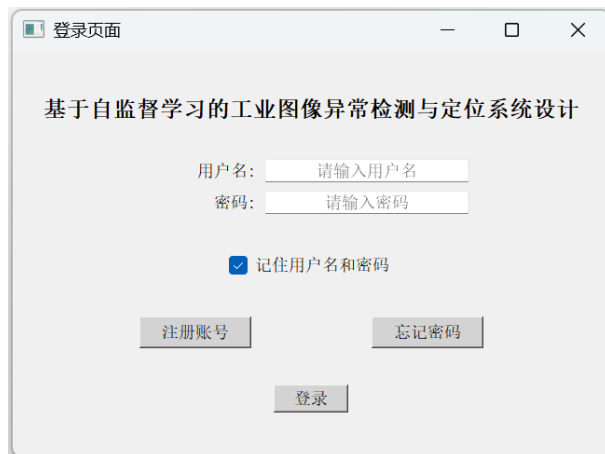


图 5-2 系统登录界面

5.2.2 系统功能选择页面设计

用户登陆成功后跳转至工业图像异常检测系统功能选择主页面，如图 5-3 所示，主页面包含模型训练与测试功能、工业图像异常检测功能和检测结果功能，为用户提供便捷、灵活的操作体验，确保系统响应迅速、功能易用，全面提升用户的工作效率与使用满意度。

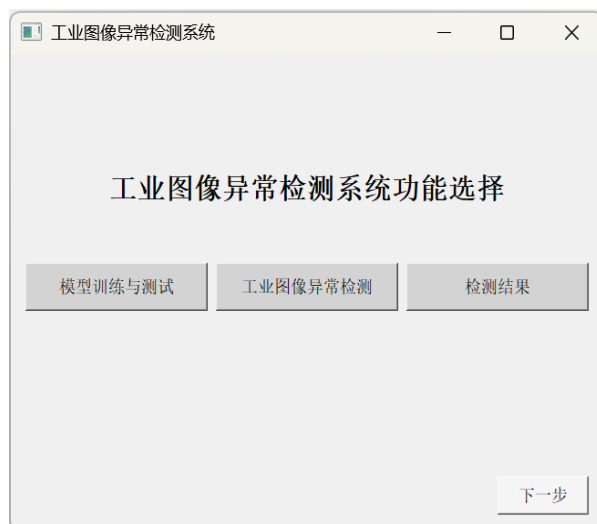
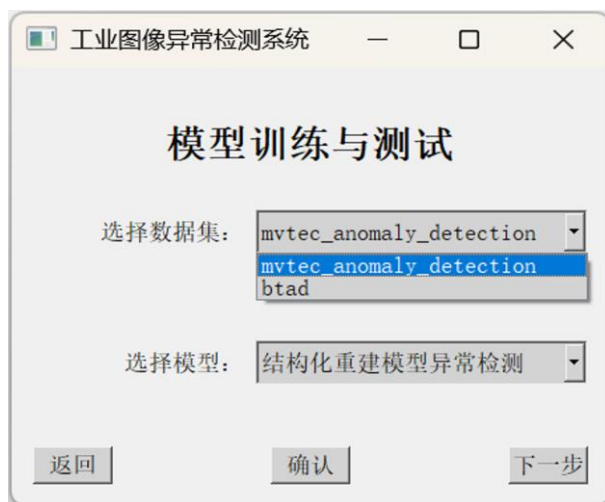


图 5-3 系统功能选择页面

5.2.3 模型训练与测试页面设计

模型训练与测试界面如图 5-4 所示，用户首先点击数据集选择下拉框，从已上传的数据集中选择目标数据集。据集选择下拉框支持用户查看并切换不同的数据集列表，方便用户根据任务需求灵活选择目标数据。然后，点击模型选择下拉框，从可用模型列表中选择适合的模型进行检测。模型选择下拉框则提供了多种预训练模型选项，用户可根据具体场景选择最优模型，确保检测任务的准确性和效率。



5-4 异常检测可视化定位展示

首先，点击选择数据集下拉框，之后会弹出各种数据集名称，并且选择合适的数据集。然后，再次点击选择模型下拉框，之后会弹出各种模型名称，并且选择合适的模型。最后，点击确认，完成模型训练与测试。

5.2.4 工业图像异常检测页面设计

在完成模型训练与测试后,用户可通过系统功能界面进入工业图像异常检测模块。如图 5-5 所示,该模块界面设计简洁直观,主要包含七个功能按键:待检测图像选择、结构化重建模型检测、特征嵌入模型检测、结果显示、返回主界面、保存检测结果以及下一步操作。这些功能按键经过精心设计,能够充分满足用户多样化的检测需求,为工业图像异常检测提供了便捷的操作体验

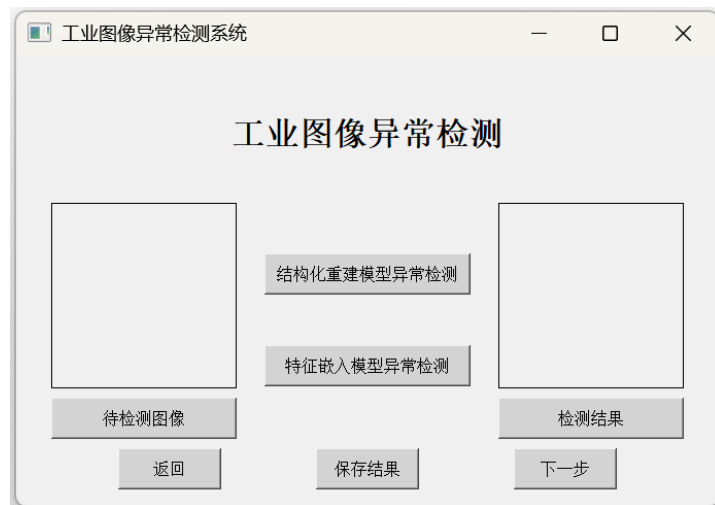


图 5-5 工业图像异常检测页面

5.3 工业图像异常检测系统的应用与实现

5.3.1 结构化重建模型异常检测系统实现

如图 5-6 所示,从功能选择界面首先选择 MVTEC AD 数据集中的一张测试图像,然后选择结构化重建模型异常检测,该系统就会对输入图像进行检测。

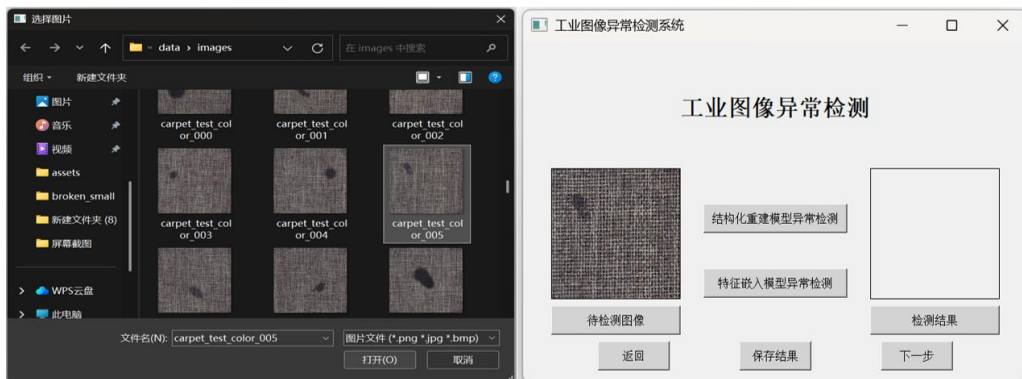


图 5-6 选择 MVTEC AD 待检测图片

待检测图像选取成功后，点击检测结果。如图 5-7 所示，当检测到图像存在异常时，就会发出提示“异常图片”，提醒用户该产品存在问题，并生成可视化定位图，便于快速定位缺陷的准确位置。



图 5-7 检测结果

5.3.2 特征嵌入模型异常检测系统实现

如图 5-8 所示，从功能选择界面首先选择 BTAD 数据集中的一张测试图像，然后选择结构化重建模型异常检测，该系统就会对输入图像进行检测。

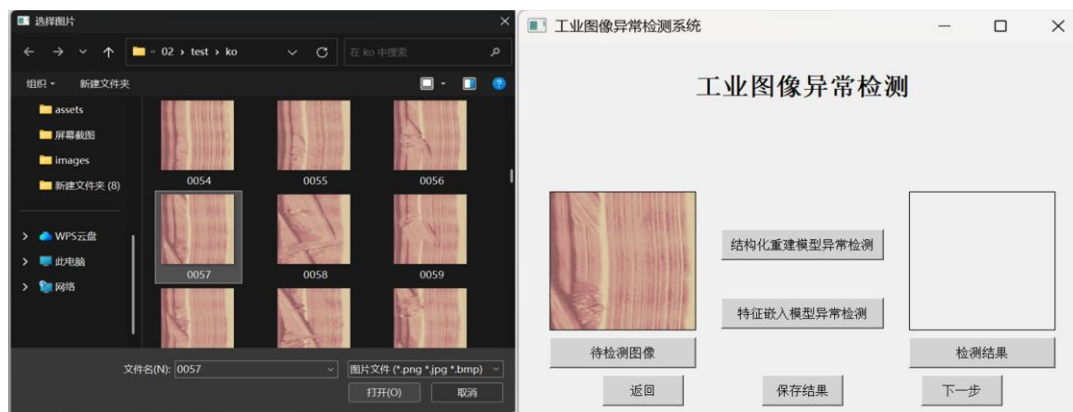


图 5-8 选择 BTAD 待检测图片

当检测图像选取完成后，用户可点击查看检测结果。此外，系统采用了智能提示机制，能够实时反馈检测状态。如图 5-9 所示，若检测到图像存在异常，系统将发出“异常图片”提示，提醒用户该产品存在质量问题，同时生成可视化定位图，帮助用户快速准确地定位缺陷位置。



图 5-9 检测结果

5.4 本章小结

为了实现更高效、便捷且精准的异常检测与定位，本章设计并构建了一套工业图像异常检测系统，并详细介绍了其整体框架设计。系统从用户界面、模型训练与测试、工业图像异常检测以及结果展示等多个模块进行了全面设计，使用户能够清晰地了解异常检测的完整流程。实验结果表明，本章设计的工业图像异常检测系统具有实用性和有效性。

第 6 章 全文总结与展望

6.1 全文工作总结

在现实工业生产中，异常样本数据稀缺且采集困难，同时训练过程需要精细的像素级标注数据集，耗费大量人力和物力，难以满足大规模质检需求。因此，本文针对工业图像异常检测与定位中的关键问题展开了一系列研究，提出了两种自监督学习异常检测算法模型，并通过在两个真实工业缺陷检测数据集上的实验验证了其有效性。主要内容和贡献总结如下：

论文的主要研究成果包括以下三个方面：

(1) 针对工业图像缺陷训练样本不足和漏检率高等问题，提出了一种基于自监督结构化重建的工业图像异常检测与定位模型。首先，在模型训练中引入多尺度结构相似度损失（MS-SSIM）和多尺度梯度幅度相似性损失（MSGMS），明显提升了模型的重建性能。其次，提出了一种多互补掩码融合方法，进一步提高了异常检测的准确率。实验结果表明，该模型在 MVTec AD 数据集上的平均异常检测率达到 94.9%，对比算法中性能最优的模型提升了 1.6%；同时，平均异常定位率达到 95.6%，对比算法中表现最好的模型提高了 1.7%，充分验证了该模型的有效性和实用性。随着工业技术的快速发展，未来需要进一步优化和压缩结构化重建模型，以提升其在更复杂工业场景下的检测精度和运行效率。

(2) 针对工业图像异常检测中缺陷样本不足、检测效率和精度不平衡等问题，提出了一种基于自监督深度记忆特征嵌入的工业图像异常检测与定位模型。首先，通过异常模拟策略生成异常样本，解决了异常样本不足的问题。然后，设计了深度记忆特征模块和多尺度特征融合模块，并引入感知损失和焦点损失进行训练，加快了模型训练速度，同时提高了异常检测与定位的精度。最后，实验结果表明，本文方法在 MVTec AD 数据集和 BTAD 数据集上的平均异常检测率分别达到 97.1% 和 97.3%，对比算法中性能最优的模型分别提升了 0.7% 和 4.9%；平均异常定位率分别达到 96.7% 和 95.6%，对比算法中表现最好的模型分别提高了 3.2% 和 2.7%。随着计算机视觉技术的快速发展，未来需要进一步优化和压缩深度记忆特征嵌入模型，以实现更快速、更精确的工业图像异常检测与定位。

(3)利用 Qt 等开发工具,设计并实现了一种自监督工业图像异常检测与定位系统。该系统包括登录界面、功能选择模块和检测结果显示模块等,能够根据不同的异常检测模型对输入图像进行分析,并输出和保存可视化热力图及异常分割图。

综上所述,本文的研究成果为工业图像异常检测与定位提供了理论支持和技术手段,具有重要的理论意义。

6.2 后续研究展望

本研究围绕自监督图像异常检测这一核心问题展开,在深入分析现有研究成果的基础上,提出了两种创新性的检测模型。尽管所提出的算法在一定程度上缓解了该领域的关键挑战,但由于研究周期和实验条件的限制,仍存在若干需要改进的方面。未来的研究工作可以从以下几个方向进行深入探索和优化:

(1)异常区域边缘检测精度不足

本文所提两种异常检测模型虽然在异常边缘检测方面取得了明显改进,但相较于目标分割任务的精度仍存在差距。未来可探索两种优化方向:一是引入目标分割中的有监督学习方法,利用伪标签指导网络学习异常特征;二是扩展卷积感受野,例如采用大尺寸卷积核、增加网络深度、构建多尺度特征金字塔,或利用 Transformer 的大感受野特性,以进一步提升边缘检测性能。

(2)跨领域自适应性与泛化性

本文提出的两种工业图像异常检测算法已在工业数据集上得到验证。然而,自监督异常检测对数据质量要求较高,而当前可用的公开数据集数量有限。因此,未来的研究工作可以扩展到其他领域进行实验验证,评估算法在不同场景下的适应能力和泛化性能,从而拓宽其应用范围。

(3)模型轻量化设计

本研究提出的两种异常检测算法均采用预训练特征提取器,其表征能力在一定程度上制约了检测性能的上限。虽然大规模模型通常具有更好的表现,但模型复杂度的增加影响其在终端设备的部署效率。因此,未来可探索模型剪枝、量化压缩以及知识蒸馏等轻量化技术,设计更高效的算法架构,在维持检测精度的同时降低计算资源消耗。

综上所述，尽管本文的研究取得了一定的成果，但在边缘检测精度、跨领域适应性和模型轻量化等方面仍有改进空间。未来的工作可以围绕这些方向展开，以进一步提升自监督图像异常检测技术的实用性和普适性。

参考文献

- [1] 苑玮琦, 薛丹. 基于机器视觉的隧道衬砌裂缝检测算法综述[J]. 仪器仪表学报, 2017, 38(12):3100-3111.
- [2] 代心茹. 基于知识蒸馏的小型深度神经网络优化研究[D]. 华东师范大学, 2024.
- [3] 万千. 基于深度特征表示的工业图像异常检测方法[D]. 华中科技大学, 2023.
- [4] 于瀚文, 吴一全. 基于机器视觉的锂电池缺陷检测研究进展[J]. 仪器仪表学报, 2024, 45(09):1-23.
- [5] Liu J, Xie G, Wang J, et al. Deep Industrial Image Anomaly Detection: A Survey[J]. Machine Intelligence Research, 2024, 21(1):104-135.
- [6] Shvetsova N, Bakker B, Fedulova I, et al. Anomaly detection in medical imaging with deep perceptual autoencoders[J]. IEEE Access, 2021, 9:118571-118583.
- [7] Tao X, Gong X, Zhang X, et al. Deep learning for unsupervised anomaly localization in industrial images: A survey[J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71:1-21.
- [8] 李茗欣, 黄远程, 竞霞等. 融合视觉注意机制的高光谱RX异常检测算法[J]. 红外技术, 2023, 45(04):402-409.
- [9] 黄洪海. 基于多模态掩码的工业图像异常检测[D]. 哈尔滨理工大学, 2024.
- [10] Aiger D, Talbot H. The phase only transform for unsupervised surface defect detection[C]//2010 IEEE computer society conference on computer vision and pattern recognition. IEEE, 2010: 295-302.
- [11] 郭慧, 徐威, 刘亚菲. 基于支持向量机的钢板表面缺陷检测[J]. 东华大学学报(自然科学版), 2018, 44(04):635-639.
- [12] 吴浪. 基于特征重构与知识蒸馏的无监督图像异常检测算法研究[D]. 电子科技大学, 2024.
- [13] 韩成生. 基于深度学习的光伏电池板缺陷异常检测研究[D]. 山东大学, 2023.
- [14] Qin J, Gu C, Yu J, et al. Multilevel saliency-guided self-supervised learnin

- g for image anomaly detection[J].Signal, Image and Video Processing, 2024, 18(8-9):6339-6351.
- [15] Xue Y D, Li Y C. A fast detection method via region-based fully convolutional neural networks for shield tunnel lining defects. Computer-Aided Civil and Infrastructure Engineering, 2018, 33(8):638–654
- [16] Ding R W, Dai L H, Li G P, Liu H. TDD-net: A tiny defect detection network for printed circuit boards. CAAI Transactions on Intelligence Technology, 2019, 4(2): 110–116.
- [17] 陶显, 侯伟, 徐德. 基于深度学习的表面缺陷检测方法综述[J]. 自动化学报, 2021, 47(05):1017-1034.
- [18] Wang H, Yu X M. Research on image recognition algorithm of weld defect based on deep learning[J]. Advances in Computer, Signals and Systems, 2022, 6(1):83-89.
- [19] Tu Y, Zhang B, Liu L, et al. Self-supervised feature adaptation for 3d industrial anomaly detection[C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2024:75-91.
- [20] Sabokrou M , Fathy M , Hoseini M . Video anomaly detection and localisation based on the sparsity and reconstruction error of auto-encoder[J]. Electronics Letters, 2016, 52(13):1122-1124.
- [21] Haselmann M, Gruber D P, Tabatabal P. Anomaly detection using deep learning based image completion[C]. 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE, 2018:1237-1242.
- [22] Bergmann P, Lowe S, Fauser M, et al. Improving unsupervised defect segmentation by applying structural similarity to autoencoders[J]. arXiv preprint arXiv:1807.02011, 2018.
- [23] He K, Chen X, Xie S, et al. Masked autoencoders are scalable vision learners[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022:16000-16009.
- [24] Vitjan Z, Matej K, Danijel S. Reconstruction by inpainting for visual anomaly detection[J]. Pattern Recognition, 2021, 112:107706.

- [25] Wang X, Wang Y, Pan Z, et al. Unsupervised anomaly detection and localization via bidirectional knowledge distillation[J]. Neural Computing and Applications, 2024, 36(29):18499-18514.
- [26] Aqsa R, Bushra Z, Amina R, et al. Fabric Defect Detection Using Computer Vision Techniques: A Comprehensive Review[J]. Mathematical Problems in Engineering, 2020, 2020(1): 8189403.
- [27] Defard T, Setkov A, Loesch A, et al. Padim: a patch distribution modeling framework for anomaly detection and localization[C]//International Conference on Pattern Recognition. Cham: Springer International Publishing, 2021:475-489.
- [28] Xu Z, Zhang W, Zhang T, et al. HRCNet: high-resolution context extraction network for semantic segmentation of remote sensing images[J]. Remote Sensing, 2020, 13(1):71-93.
- [29] Meha D , Manan S. An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and Convolutional neural network (CNN)[J]. Clinical Health, 2021, 41-11.
- [30] 司念文, 张文林, 屈丹,等. 卷积神经网络表征可视化研究综述[J]. 自动化学报, 2022, 48(08):1890-1920.
- [31] Kangkang W, Linglong Y, Lei Z, et al. Optimized active layer morphology toward efficient and polymer batch insensitive organic solar cells[J]. Nature Communications, 2020, 11(1):2855.
- [32] Cirocchi R, Fearnhead N, Vettoretto N, et al. The role of emergency laparoscopic colectomy for complicated sigmoid diverticulitis: A systematic review and meta-analysis[J]. The Surgeon, 2019, 17(6):360-369.
- [33] Ghose P, Ghose A, Sadhukhan D, et al. Improved polyp detection from colonoscopy images using finetuned YOLO-v5[J]. Multimedia Tools and Applications, 2024, 83(14):42929-42954.
- [34] 韩冉, 曾广淼, 王荣杰. 基于残差块网络的图像去雨算法[J]. 仪器仪表学报, 2021, 41(08):175-182.

- [35] 张重生, 陈杰, 李岐龙,等. 深度对比学习综述[J]. 自动化学报, 2023, 49(01): 15-39.
- [36] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Bio medical Image Segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer International Publishing, 2015.
- [37] 曹希戡, 陈鑫, 魏同权. 基于RISC-V的深度可分离卷积神经网络加速器[J]. 计算机学报, 2024, 47(11):2536-2551.
- [38] Yin S, Li H, Laghari A, et al. An Anomaly Detection Model Based on Deep Auto-Encoder and Capsule Graph Convolution via Sparrow Search Algorithm in 6G Internet of Everything[J]. IEEE Internet of Things Journal, 2024, 11(18):29402-29411.
- [39] Liu S, Ni H, Li C, et al. Defect GAN: Synthetic Data Generation for EMU Defects Detection With Limited Data[J]. IEEE Sensors Journal, 2024(11):24.
- [40] Kang X, Li Y, Zhang Y, et al. Anomaly detection in concrete dam using memory-augmented autoencoder and generative adversarial network (MemAE-GAN)[J]. Automation in Construction, 2024, 168:105794.
- [41] An T, Mao B, Xue B, et al. Patch loss: A generic multi-scale perceptual loss for single image super-resolution[J]. Pattern Recognition, 2023, 139:109510.
- [42] 李博文, 杨续来, 葛肖尽,等. 锂离子电池极片表面缺陷检测方法研究进展[J]. 仪器仪表学报, 1-22.
- [43] 王纪康, 赵旭俊. 基于知识蒸馏的图像异常检测方法[J]. 计算机技术与发展, 2024, 34(05):149-156.
- [44] Wang Q, Wu B, Zhu P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020:11534-11542.
- [45] 吕承侃, 沈飞, 张正涛,等. 图像异常检测研究现状综述[J]. 自动化学报, 2022, 48(06):1402-1428.

- [46] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9,2015, Proceedings, Part III 18. Springer International Publishing, 2015:234-241.
- [47] Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. Deep Learn Med Image Anal Multimodal Learn Clin Decis Support (2018). 2018, Sep;11045:3-11.
- [48] 史雨馨, 朱继杰, 凌志刚. 基于特征增强YOLOv4的无人机检测算法研究[J]. 电子测量与仪器学报, 2022, 36(07):16-23.
- [49] 卢印举, 马芳, 戴曙光, 等. 融合多尺度特征的马尔可夫随机场路面裂缝分割算法[J]. 计算机辅助设计与图形学学报, 2022, 34(05):711-721.
- [50] 马敏, 高晓波. 自适应加权的多特征融合ECT图像重建算法[J]. 电子测量技术, 2022, 45(21):130-135.
- [51] Bergmann P, Batzner K, Fauser M, et al. The MVTec anomaly detection dataset: a comprehensive real-world dataset for unsupervised anomaly detection[J]. International Journal of Computer Vision, 2021, 129(4):1038-1059.
- [52] 孙隽姝, 王树徽, 杨晨雪, 等. 附加特征图增强的图卷积神经网络[J]. 计算机学报, 2023, 46(09):1900-1918.
- [53] Massoli F V, Falchi F, Kantarci A, et al. MOCCA: Multilayer one-class classification for anomaly detection[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 33(6): 2313-2323.
- [54] Yoa S, Lee S, Kim C, et al. Self-supervised learning for anomaly detection with dynamic local augmentation[J]. IEEE Access, 2021, 9:147201-147211.
- [55] Reiss T, Hoshen Y. Mean-shifted contrastive loss for anomaly detection[C]. Proceedings of the AAAI Conference on Artificial Intelligence. 2023, 37(2): 2155-2162.
- [56] Venkataramanan S, Peng K C, Singh R V, et al. Attention guided anomaly localization in images[C]. European Conference on Computer Vision. Cham

- m: Springer International Publishing, 2020:485-503.
- [57] Song H, Li C, Fu Y, et al. A two-stage unsupervised approach for surface anomaly detection in wire and arc additive manufacturing[J]. Computers in Industry, 2023, 151:103994.
- [58] Kwon D, Kim H, Kim J, et al. A survey of deep learning-based network anomaly detection[J]. Cluster Computing, 2019, 22:949-961.
- [59] Gao H Y, Zhang M C, Ge Q B. Defect sample image generation method based on point set matching[J]. Journal of Intelligent Systems, 2023, 18(05):1030-1038.
- [60] Chen S, Cheng Z, Zhang L, et al. Snipe Det: Attention-guided pyramidal prediction kernels for generic object detection[J]. Pattern Recognition Letters, 2021, 152:302-310.
- [61] Zhang R, Lai P, Wan X, et al. Lesion-aware dynamic kernel for polyp segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, 2022:99-109.
- [62] Singh, Anand S, Desai, et al. Automated surface defect detection framework using machine vision and convolutional neural networks[J]. Journal of Intelligent Manufacturing, 2022, 34(4):1-17.
- [63] Yang L, Yang W, Wang Z. CutPaste-ROI: An Industrial Anomaly Data Detection Method based on Self-supervised Learning[J]. Journal of Imaging Science and Technology, 2024, 68(2):1-12.
- [64] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017:4700-4708.
- [65] Li X, Wang W, Hu X, et al. Generalized focal loss v2: Learning reliable localization quality estimation for dense object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021:11632-11641.
- [66] Mishra P, Verk R, Fornasier D, et al. Vt-adl: A vision transformer network

- for image anomaly detection and localization[C]. 2021 IEEE 30th International Symposium on Industrial Electronics (ISIE), 2021, 01-06.
- [67] Yi J, Yoon S. Patch SVDD: Patch-Level SVDD for Anomaly Detection and Segmentation[C]//Asian Conference on Computer Vision. 2020: 375-390.
- [68] Pirnay J, Chai K. Inpainting transformer for anomaly detection[C]//International Conference on Image Analysis and Processing. Cham: Springer International Publishing, 2022:394-406.
- [69] Mishra P, Verk R, Fornasier D, et al. VT-ADL: A vision transformer network for image anomaly detection and localization[C]//2021 IEEE 30th International Symposium on Industrial Electronics (ISIE). IEEE, 2021:01-06.

攻读硕士学位期间所发表的论文及其它研究成果

[1]范伟平,程凡永,张明艳.自监督结构化重建的工业图像异常检测与定位[J].哈尔滨商业大学学报(自然科学版),2025,41(02):152-160.

致谢

时光飞逝，转眼间研究生生涯已接近尾声。回望这段求学之路，有艰辛，有收获，更有许多需要感谢的人。在此，我谨以最朴实的文字，表达我内心最真挚的谢意。

首先，我要衷心感谢我的导师程凡永教授。感谢您在学术上的悉心指导，从论文选题到实验设计，从数据分析到最终成稿，每一步都离不开您的耐心点拨。您严谨的治学态度和务实的科研精神深深影响了我，让我明白做学问不仅需要热情，更需要踏实和坚持。同时，也要感谢课题组其他老师的帮助，你们的建议让我的研究更加完善。

感谢实验室的师兄师弟师妹们。感谢同门在科研和生活中的互相支持，那些一起熬夜赶数据、互相修改论文的日子，将成为我最珍贵的回忆。此外，也要感谢我的室友和朋友们，你们的陪伴让这段求学时光充满温暖。

家人的支持是我前进的最大动力。感谢我的父母，你们始终尊重我的选择，默默承担生活的压力，让我能够安心求学。你们的理解与鼓励，是我面对困难时最坚实的后盾。

最后，感谢所有在研究生阶段给予我帮助的人，无论是课堂上老师的传授，还是生活中陌生人的善意，都让我受益匪浅。这段经历不仅让我收获了知识，更让我学会了坚持与感恩。

未来的路还很长，我将带着这段时光的馈赠，继续踏实前行。

尚德敏学 唯实惟新

