

分类号: TP39

单位代码: 10335

密 级: 公开

学 号: 22060443

浙江大学

硕士专业学位论文



中文论文题目: 基于语义分割的钢材缺陷检测算法研究

英文论文题目: Research on Steel Defect Detection Algorithm
based on Semantic Segmentation

申请人姓名: 张训飞

指导教师: 胡浩基 副教授

合作导师: 尹航 高级工程师

专业学位类别: 工程硕士

专业学位领域: 电子信息

所在学院: 工程师学院

论文提交日期 2023 年 3 月

基于语义分割的钢材缺陷检测算法研究



论文作者签名: 张训飞

指导教师签名: 胡浩基

论文评阅人 1: 隐名评审

评阅人 2: 隐名评审

评阅人 3: 隐名评审

答辩委员会主席: 马龙华\教授\浙大宁波理工学院

委员 1: 沈会良\教授\浙江大学信息与电子工程学院

委员 2: 刘之涛\教授\浙江大学控制科学与工程学院

委员 3: 王朝\硕导\浙江大学宁波科创中心

委员 4: 朱想先\正高\宁波普瑞均胜汽车电子科技有限公司

答辩日期: 2023 年 3 月 4 日

浙江大学研究生学位论文独创性声明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得浙江大学或其他教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

学位论文作者签名：张训

签字日期：2023 年 3 月 4 日

学位论文版权使用授权书

本学位论文作者完全了解浙江大学有权保留并向国家有关部门或机构送交本论文的复印件和磁盘，允许论文被查阅和借阅。本人授权浙江大学可以将学位论文的全部或部分内容编入有关数据库进行检索和传播，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。

（保密的学位论文在解密后适用本授权书）

学位论文作者签名：张训

导师签名：胡浩基

签字日期：2023 年 3 月 4 日

签字日期：2023 年 3 月 4 日

致谢

西子湖畔，求是学府，
海纳江河，灿若星辰。
青衿之志，履践致远，
满怀赤忱，朝乾夕惕。
流离琐尾，矢志不移，
道不相谋，砥砺前行。
幸遇恩师，承蒙教诲，
益者三友，情谊长存。
仰取俯拾，秣载而归，
行而不辍，未来可期。

感谢胡浩基老师！

感谢同行的朋友们！

感恩一直陪伴的父母！

摘要

面向工业领域的异常检测在现实工程中必不可少。随着工业的高速发展,针对工业生产的异常检测的必要性逐步体现。对于真实工业场景下的钢材表面出现的针孔、破洞、压印、折印等缺陷,需要及时地、准确地检测出来以降低风险,避免后续的一系列工业事故和经济损失。相比于高人力成本消耗、低效率、低精度的人工异常检测,基于视觉的机器智能化异常检测方向逐步体现出其价值意义所在。随着深度学习在图像分类、目标检测、语义分割等相关领域的快速发展,研究人员开始着力部署这些高精度的算法去解决工业实际的问题,其中就包括本文主要涉及的工业缺陷异常检测领域。

然而,在真实工业场景之中存在着诸多难题:一方面,算法模型对于数据有很大的依赖,需要人工定义缺陷,并且消耗大量的人力去完成真实工业场景的数据标注,这样不仅效率低,而且人工标注也势必会带来在定义和标注上的误差;另一方面,真实工业场景数据中异常缺陷总是存在于极少数的样本之中,异常缺陷一般都比较细小,难以识别,且正负样本数据不均衡,所以也会带来一系列的问题。

针对上述的问题,本文旨在设计一个可以充分利用正常样本数据,对数据依赖更少,精度更高的基于语义分割的工业钢材异常缺陷检测算法。本文提出了一种基于“异常重建+对比判别”的半监督对比学习的异常缺陷检测网络架构。其通过对于正常样本进行异常模拟,得到模拟异常样本,然后利用异常重建网络实现对这些模拟异常样本的“去异常化”的复原重建,将复原重建样本与正常样本进行损失计算,从而训练出一个由异常样本生成复原重建样本的异常重建网络。对于异常样本与其复原重建样本,通过后续的对比如判别网络,形成两个样本图像的对比如学习,检测异常缺陷位置,最终得到异常检测结果。得益于无标签的正常样本的充分利用,半监督对比学习的异常缺陷检测网络模型架构实现了更高的精度,同时对于数据依赖更低。我们在真实工业场景下的钢材表面缺陷数据集中进行实验,以实现钢材表面的折印和边裂两种缺陷的检测,以及对机床区域的定位。通过对比基础语义分割算法,展示了我们提出的网络架构的优越性。

其次,针对上述提出的半监督对比学习的异常缺陷检测网络架构,为了提高异常重建网络和对比如判别网络的性能,我们设计了一种掩膜特征学习模块 **MFLM**,其包含了掩膜空洞卷积和 **Transformer** 注意力机制,建立了一个高效的异常重建网络优化模块,更好地优化了重建网络的异常复原的能力;同时设计了 **CALM** 的对比如注意力学习模块,

提高对比判别网络在空间和通道上的对比学习能力。最终在真实工业场景下的钢材表面缺陷数据集中进行实验，对比了基础网络架构，实现了更优的检测效果和性能。

关键词：半监督学习，对比学习，语义分割，异常检测，注意力机制

Abstract

Anomaly Detection for industrial field is essential in engineering. With the rapid development of industry, the necessity of anomaly detection for industrial production is gradually embodied. Defects such as pinholes, holes, imprints and scratches on steel surfaces in real industrial scenarios need to be detected rapidly and accurately to reduce risks and avoid a series of industrial accidents and economic losses. Compared with manual anomaly detection with high labor cost, low efficiency and low accuracy, anomaly detection based with vision intelligence gradually shows its value. With the rapid development of deep learning in image classification, object detection, semantic segmentation and other related fields, researchers have begun to deploy these high-precision algorithms to solve industrial practical problems, including the industrial defect anomaly detection field.

However, there are so many problems in real industrial scenarios: on the one hand, the algorithm model has a great dependence on data and requires manual definition of defects. It consumes a lot of manpower to complete the data annotation, which is not only inefficient, but also will absolutely make errors in the definition and annotation; On the other hand, there are always very few anomaly defects in real industrial scenarios. The anomaly defects are generally small, difficult to identify, and the positive and negative sample data are unbalanced, which will lead to a series of problems.

To solve the above problems, we aims to design an anomaly detection method for industrial defects based on semantic segmentation, which can make the most of normal sample data, rely less on data and has higher accuracy. we propose an anomaly detection network architecture based on semi-supervised contrast learning of "anomaly reconstruction + contrastive discrimination". It gets simulated anomaly samples by anomaly information simulation of normal samples, and then uses the reconstruction network to realize the simulated anomaly samples recovery and reconstruction. It calculates the loss between the restored and reconstructed samples and the normal samples, so as to train a reconstruction network that generates restored and reconstructed samples from anomaly samples. For the anomaly samples and their recovery and reconstruction samples, through the contrastive discrimination network, the comparative

learning of two sample images is formed to detect the location of anomaly defects, and finally the anomaly detection result is obtained. Thanks to the full use of unlabeled normal samples, the semi-supervised comparative learning network architecture model for anomaly detection achieves higher accuracy and lower dependence on data. We carried out experiments on the steel surface defect dataset in real industrial scenarios to detect the scratch and the flaw on the steel surface and locate the lathe area. By comparing the basic semantic segmentation algorithms, the advantages of our proposed network architecture are demonstrated.

Secondly, in order to improve the performance of the anomaly reconstruction network and the contrastive discrimination network, we designed a Masked Feature Learning Module (MFLM), which includes the collaborative application of masked dilated convolution and Transformer attention mechanism, to establish an efficient anomaly reconstruction network optimization module, which better optimizes the ability of anomaly recovery of the reconstruction network; At the same time, the contrastive attention learning module of CALM is designed to improve the contrastive learning ability of the contrastive discrimination network in space and channel. Finally, the experiment was carried out in the steel surface defect dataset in real industrial scenarios, and achieve better detection effect and performance.

Keywords: Semi-supervised Learning, Comparative Learning, Semantic Segmentation, Anomaly Detection, Image Reconstruction, Attention Mechanism

目录

致谢	I
摘要	II
Abstract	IV
目录	VI
图目录	VIII
表目录	IX
1 绪论	1
1.1 研究背景和研究意义	1
1.2 国内外研究现状	2
1.2.1 语义分割算法	2
1.2.2 工业缺陷检测算法	5
1.3 钢材表面缺陷检测项目背景	7
1.4 本文的主要研究内容和创新点	8
1.5 本文的结构安排	9
2 基于语义分割的工业缺陷检测背景知识	10
2.1 基于语义分割的工业缺陷检测简介	10
2.2 工业缺陷检测的问题与挑战	11
2.2.1 工业缺陷检测数据难点	11
2.2.2 工业缺陷检测任务挑战	12
2.3 基于数据情况划分的工业缺陷检测任务	12
2.3.1 缺陷模式已知	13
2.3.2 缺陷模式未知	14
2.3.3 少量缺陷标注	20
2.4 本章小结	21
3 半监督对比学习的网络架构设计	23
3.1 问题阐述	23
3.2 本章方法概述	23
3.3 算法设计	24
3.3.1 异常模拟模块	25
3.3.2 异常重建网络 and 对比判别网络模块	26
3.3.3 损失函数设计	28
3.4 实验分析	32
3.4.1 数据集介绍	32
3.4.2 实验细节	34
3.4.3 对比实验	35
3.4.4 可视化结果分析	39
3.5 本章小结	41
4 面向重建和判别网络的优化模块设计	43

4.1 问题阐述	43
4.2 本章方法概述	43
4.3 算法设计	43
4.3.1 掩膜特征学习模块	43
4.3.2 对比注意力学习模块	48
4.3.3 优化的模型结构	52
4.4 实验分析	53
4.4.1 实验细节	53
4.4.2 对比实验	54
4.5 本章总结	56
5 总结与展望	57
参考文献	58
附录	65
A 作者简历	65

图目录

图 2.1 基于语义分割的工业缺陷检测任务	10
图 2.2 有监督语义分割网络的基本结构	13
图 2.3 自编码器类的结构	15
图 2.4 生成对抗类网络的结构	17
图 2.5 Memory 类的结构	18
图 3.1 半监督对比学习网络架构	24
图 3.2 异常模拟模块	25
图 3.3 两阶段半监督学习训练模式	27
图 3.4 损失函数对比	29
图 3.5 钢材表面缺陷数据集	34
图 3.6 不同数据量的 MIoU 变化趋势	38
图 3.7 不同损失函数的重建效果对比	39
图 3.8 半监督对比学习架构与基础网络的检测效果对比	41
图 4.1 掩膜空洞卷积	44
图 4.2 掩膜空洞卷积的特征提取流程	45
图 4.3 Transformer 的 Tokens 提取流程	46
图 4.4 Transformer 的多头注意力机制模块	47
图 4.5 CBAM 通道注意力模块	49
图 4.6 CBAM 空间注意力模块	49
图 4.7 CBAM 流程	50
图 4.8 CA 流程	51
图 4.9 CALM 流程	51
图 4.10 优化后的重建网络和判别网络	53

表目录

表 3.1 数据集信息 32

表 3.2 与基础网络的缺陷检测性能对比 36

表 3.3 与基础网络的机床区域检测性能对比 36

表 3.4 不同数据量对模型性能的影响分析 37

表 4.1 重建网络优化模块的性能对比 54

表 4.2 重建网络优化模块的计算量与参数量对比 55

表 4.3 判别网络优化模块的性能对比 55

表 4.4 判别网络优化模块的计算量与参数量对比 56

1 绪论

1.1 研究背景和研究意义

近年来,人工智能技术发展速度迅猛,在各种领域都取得了傲人的成绩,比如图像分类^[1,2]、人脸识别^[3,4]、语义分割^[5,6]、目标检测^[7,8]等等领域,并逐渐向更多领域深耕发展。与图像分类这种的基于图像的分类任务不同,语义分割是基于像素级的精细分类任务,任务难度要求更高,不仅需要基于图像层面的识别能力,也需要对于目标的精确定位的识别能力。比如在当下火热的自动驾驶领域^[9,10]的任务中,实验车辆通过前端摄像头获取的道路视野的实时画面,通过语义分割算法能够准确地感知到视野环境下的诸多道路环境要素,包括道路边界、斑马线、行人、不同功能类型的车辆等等,从而实现高精度的智能驾驶辅助系统。再比如在工业视觉领域^[11,12]中,需要对金属表面实现异常缺陷的检测,语义分割算法的目标就是找出异常缺陷的区域位置,并且获得异常的准确轮廓。这不同于图像分类检测结果的有异常和无异常的判断,也不同于目标检测结果的统一矩形框粗略选定异常范围,语义分割的结果具有更多的定位信息和使用价值。

在工业缺陷检测领域,语义分割算法的重要价值毋庸置疑。工业缺陷检测是指对于工业界上各种产品,包括大到航天飞行器,小到微纳级别的电子元器件,都需要在生产加工流程前后进行一定的工业检测保障措施,所以,工业缺陷检测是对于工业产品质量保障、生产安全稳定的重要环节和重要技术之一。相较于以往高成本和低效率的人工筛查,工业界一直急需一种能够实现大规模应用的工业缺陷检测技术。所以近年来,各种各样的工业缺陷检测技术层出不穷,伴随着工业界的不断发展而有所突破,获得了工业界一定程度上的认可。其中,基于视觉的工业缺陷检测技术成了工业缺陷检测中尤为重要的技术领域之一,毕竟代替人眼执行操作和决策具有良好的可靠性和普适性。所以,基于视觉的高精度和高效率的工业缺陷检测技术备受学术界和工业界的关注,广泛应用于无人化工业质检、生产过程控制、异常问题在线反馈等等场景。语义分割技术能更好地满足了工业缺陷检测的实现目标,已经成为了智能生产和制造领域不可或缺的基础研究和关键应用技术之一。

然而,基于语义分割的工业缺陷检测技术面临着诸多难点,这与工业界的现实环境有着密不可分的关系。比如缺陷样本资源不足,缺陷类型的定义不清晰、不全面,缺陷的可视性较差,缺陷的形状多变等等。这就导致了一系列的实际场景的技术难题,致使

工业缺陷检测发展道路受阻。具体问题可概括为：1. 真实工业场景下的大批量数据标注困难，人为标注存在误差；2. 数据标注量有限，算法模型对于数据的依赖较大；3. 真实工业场景下的存在大量的正常样本，而异常样本数量过少，正负样本不均衡。

因此，真实工业场景中，对于有限标注的样本信息的充分利用，减少网络对数据依赖十分重要。尤其是针对大量的无标签的正常样本，对这类样本需要有更好的价值利用。同时，不同于图像分类、目标检测，使用语义分割算法实现的高精度要求的工业缺陷检测，需要更高效的网络设计。

考虑到钢材表面缺陷数据集中，钢材是在机床上高速滚动的，生成的图像相对模糊，且由于工业场景的光线暗淡，颜色辨别度低，导致缺陷难以清晰体现，同时，钢材表面缺陷的形状和大小多种多样，使得单纯的有监督和无监督学习的方案对于此类工业检测难以使用，无法实现更高精度的算法要求。所以，对于这类真实工业数据的特点，半监督学习的使用具有重要价值。

本文以上述问题为出发点，致力于针对工业缺陷检测领域的数据样本问题，设计更高精度、更低数据依赖的半监督语义分割算法网络，进而在工业缺陷检测领域有更好的应用价值。

1.2 国内外研究现状

近年来，大量的语义分割算法和异常检测算法不断涌现，并随着工业缺陷应用领域做出了针对性的突破，下面我们将分别介绍语义分割算法以及工业异常缺陷检测算法的发展之路。

1.2.1 语义分割算法

语义分割算法大致可以划分为两种类型：基于传统方法的语义分割算法和基于深度学习方法的语义分割算法。接下来，我们将分别展开介绍。

(1) 基于传统方法的语义分割算法

基于边缘检测分割的方法，主要是充分利用目标物体的天然纹理、颜色以及外形轮廓等基本特征属性，通过传统图像处理中图像增强算法的边缘检测算子进行物体边缘的勾勒。例如常见的 Sobel^[13]、HOG^[14]、Laplace、Kirsch^[15]等等算子，通过一阶、二阶或者其他设计的滤波器，针对图像像素进行空间锐化滤波操作，得到像素突变的边缘位置。

在这个基础之上, **Malik**^[16]等人利用基于像素周围纹理的控制算子, 获得相邻像素属于同一区域的可能性度量, 从而将图像的像素进行简单的区域划分, 形成了简单的语义分割结果; 除此之外, 基于像素值统计属性的算法^[17,18]通过设置像素的阈值划分像素, 利用概率图模型实现语义分割, 对于边缘明显的图像分割有一定效果, 但是鲁棒性较差, 易受光线环境影响。

基于像素相似性分割的算法, 算法默认对于同一物体的像素之间具有高度的相似性, 即为整体性, 所以可以通过区域范围的像素堆进行粗略地划分, 再使用细粒度算法对像素区域进行拆分或合并。如 **Fulkerson**^[19]等的基于 **superpixel** 的分割方法, 采用先粗略后细致策略, 以及 **Farabet**^[20]等人的层次化策略, 使用多尺度卷积, 先大致定位后细致分类, 对识别稀有的物体有所帮助。但这类方法计算的复杂度较高, 分割结果也容易出现一些零碎区域, 影响分割结果。

基于图构建的分割方法, 即视每个像素为图结构的一个个结点, 将图像视为一张结点和对应边构成的图来进行分割学习。如 **GraphCut**^[21], 根据图随机游走的遍历算法对边权重进行添加, 对于边权重过低的边进行断开处理, 如此实现图的规范化切割, 对于不同分组的图就代表不同的分割区域; **CRF**^[22]用于构建概率模型以分割和标记序列数据, 使用无向图学习像素之间的上下文信息。此类基于图的方法的计算消耗巨大, 很难应用在实际的场景之中。

基于聚类的分割方法, 对图像进行分布、频谱域的信息转换以实现聚类划分。高斯混合模型分割算法^[23]使用基于高斯混合模型与 **K** 均值来实现图像的聚类初始分割, 初始分割结果为 **EM**^[24]算法的初始值, 加速算法迭代得到最优解; 其他的聚类方法包括了 **K-MEANS** 聚类^[25]、谱聚类^[26-28]也是同样的聚类分割思路。然而这一类方法的计算量过大, 方法也过于理论, 难以适配实际场景, 算法性能有限。

(2) 基于深度学习方法的语义分割算法

近年来, 深度学习促使了语义分割领域走向了一个新的高度, 深度学习技术不断在各个研究领域得到有力的证明。语义分割可以理解为是对像素的分类任务, 所以起初的深度学习语义分割算法是基于原有的图像分类算法之上, 进行像素分类目标的改进, 如 **Long**^[6]等人在 **VGG**^[1]和 **GoogleNet**^[29]等分类网络的基础上设计添加了一种全卷积层, 通过简单的特征图上采样的方式, 还原输入图像的大小, 虽然这会导致图像的细节信息丢失, 但也取得了一定的成效, 基于深度学习的图像分割算法的大门也因此逐步开启。之

后的 U-Net^[5]为了优化这个信息损失的问题,在降采样和上采样之间建立了跳级连接来实现特征的融合,保留图像的一部分细节信息;为了在特征融合上进一步优化,U-Net++^[30]做出了更为复杂的特征融合方式,将多个层级的图像信息进行深度融合,这样的结构也有利于后续的剪枝优化操作。还有后续的方法^[31-34]都在语义分割网络上做出改进,比如尽可能保留全局信息和细节信息,获取更多的图像像素之间的上下文信息,强化同类像素一致性等等。具体分类如下:

基于多尺度特征融合模型。此类模型的优化目的是尽可能学习到不同尺度的图像特征,加强特征融合能力。常见的方法比如空洞卷积,借助空洞卷积可以实现感受野的扩大,对于图像卷积的信息获取范围扩大,从而实现保持较低的计算量的同时,提高对较大物体的特征信息学习。具体示例如 Deeplab^[35]就是充分利用了不同空洞率的空洞卷积,外加不同尺度的特征融合,实现对不同尺度的物体都有着很好的检测效果。PSPNet^[36]提出了金字塔池化模块(Pyramid Pooling Module, PPM),通过并联四个不同大小的全局池化层,以获取不同尺度的图像信息,聚合多尺度的图像特征,融合了不同尺度和不同子区域之间的信息完成预测。此外,DenseASPP^[37]在 Deeplab^[35]提出的 ASPP^[35]的基础之上将生成的特征密集连接化,能够生成覆盖范围很大的特征,包含了大范围的多尺度信息。EncNet^[38]则是提出了 Context Encoding Module,对特征图进行编码,捕获场景的上下文语义信息并选择性地强调与类别相关的特征信息。SAENet^[39]使用了对抗性的 dropout-activation 块,引入了一种灵活的自监督变换机制,利用全局池化函数来压缩特征图,从而获得了不同通道上的全局特征,最后用这些全局特征优化各个像素的特征,实现全局到细节的特征融合。

基于注意力机制的模型,近几年注意力机制在自然语言处理领域大获成功,于是研究人员寄希望于将其应用于图像领域。不同于图像的卷积操作,图像上的注意力机制^[40,41]方法不局限于常用卷积的规则型卷积核,所获得的感受野并非局部的,而是直接对图像分块进行编码,对图像块进行上下文的自注意力机制的融合,对像素与像素之间的相关性进行稠密计算。具体示例如 DANet^[42]设计了双重注意力网络来自适应地集成局部特征和全局的依赖,分别为空间维度的注意力模块和通道维度的注意力模块,局部特征拥有丰富的上下文依赖关系,通过评估像素之间的特征距离,融合相似的特征通道来优化特征图。CCNet^[43]只分析水平和垂直方向上的上下文信息,提高了注意力机制的运算效率,减少了计算消耗。Li^[44]等人提出了 EMANet,将注意力机制与传统的高斯混合模型

相融合,构造期望最大化注意力网络,计算像素与聚类得到的像素簇之间的相关性,从而强化分割效果。**ACFNet**^[45]从分类的角度提取全局上下文信息,计算像素和粗分割的结果中的像素间注意力关系,构建一个从粗到细的细分网络,提升了分割性能。

基于图学习的模型。近几年图神经网络逐渐进入了研究人员的视野,其在知识图谱和推荐系统上拥有不错的效果,于是研究人员尝试将图神经网络中的图学习技术引用到图像领域,基于图学习^[46,47]的语义分割工作也就逐步出现了,具体示例如 **DGMP**^[48]使用动态图消息传递网络,对图像结点的邻域进行动态采样,动态消息回馈并预测结点,该方法不同于卷积和注意力机制,边连接是可存在可不存在的,所以会屏蔽一些不必要的像素邻域信息干扰。**GCU**^[49]使用类似图的聚类算法,基于视觉特征进行图推理,将像素结点聚类成多个像素簇,学习像素间的相似性信息。

1.2.2 工业缺陷检测算法

基于视觉的工业异常缺陷检测主要针对微小的缺陷进行检测,对于可能严重危害产品功能、质量乃至生产、运输等相关问题的解决有着尤为重要的意义。大多数情况下,对于工业缺陷的异常检测算法集中于针对数据缺陷模式已知的情况,使用常见的视觉算法对大量已标注的数据做出检测。

工业缺陷检测任务主要包括分类和定位两种功能,主流的工业缺陷检测算法包括了传统的工业缺陷检测以及基于深度学习的工业缺陷检测:

(1) 传统的工业缺陷检测

传统的算法主要是根据已定义缺陷的颜色、尺寸、位置、形状等基本特征,利用传统的图像处理方法对其进行检测筛选,亦或者结合一些简单的传统机器学习方法来对缺陷进行决策判别。对于绝大部分缺陷来说,其在像素区域上表现往往是突变的,所以边缘检测算子对于缺陷检测也是具有一定效果的,常用的边缘检测算子包括了 **Prewitt**^[50]、**Sobel**^[13]和 **Canny**^[51]以及优化算子等等,不过这一类的方法不能针对所有的异常缺陷,尤其是在像素特征不够明显,或者是背景噪声的影响因素较大时,该类方法的问题就会凸显。从频谱的角度分析,这种突变的缺陷特征往往是高频的特征信号。可以使用傅里叶变换^[52]、**Gabor**变换^[53]、小波变换^[54]等周期性方法进行频域转换,然后使用滤波器加以过滤。具体示例如 **Chetverikov**^[55]等基于纺织品的规则和流状纹理的特征检测结构缺陷,**Zhen**^[56]等采用 **Gabor** 小波变换算子,应用自适应算子选择方案来提取频域特征,然后基

于支持向量机(Support Vector Machine, SVM)方法进行分类任务,从而实现在纹理表面上的缺陷检测。对于大面积的缺陷,由于会影响图像的统计特性,可以使用 Peng^[57]等的灰度变化差异性、Murino^[58]等的灰度直方图、颜色特性等统计方法进行表征,除此之外还可以使用决策树、随机森林^[59]等传统机器学习方法实现像素分类。从上述可以看出,传统的缺陷检测对于缺陷的检测有很大的局限性,上述方法大都针对突变型缺陷有一定成果,由于缺陷存在未知性和多样性,传统方法操作复杂,很难应用到更高要求的实际场景之中。

(2) 基于深度学习的工业缺陷检测

当数据体量足够多,缺陷标注足够充足的情况之下,基于有监督的深度学习方法就能够很好地学习出这些标注的缺陷样本模式。相较于传统方法,深度学习模型具有很好的学习和泛化能力。基于深度学习的工业缺陷检测方法起初是面向图像分类的一些方法^[60-63],这些方法在各种工业生产领域均展现出较高的性能。具体示例如 Cha^[64]等针对多变的照明和阴影的现实情况利用卷积神经网络结合滑动窗口技术,实现了粗糙的混凝土裂缝的定位。Xu^[65]等提出的 SDD-CNN,针对正负样本不均衡的问题,通过集成方法对小数据进行预处理,应用标签扩张来解决类分布不平衡的问题,首先利用标签膨胀解决样本不均衡问题,然后使用半监督数据扩充方法,以更有效和更可控的方式来扩展数据集,从而优化了数据集,使其充分且均衡。

在实际应用中,我们知道区分有缺陷和无缺陷的图片分类是有意义的,但常常我们不仅需要判断图像是否存在缺陷,还需要获得缺陷的具体位置信息,也就是定位功能。对此,研究人员开始对一些通用的分类检测算法框架进行调整,将工业缺陷检测方向转移到了目标检测领域,使得检测结果可以显示目标框的区域选择,以满足工业缺陷检测的定位需求。Cha^[66]基于 Faster R-CNN^[7]检测钢筋和混凝土表面的缺陷,提供对多种类型缺陷的准实时型检测。类似的方法^[67]面向工业 CT 图像实现了精细缺陷分割。Tao^[68]等针对航空图像中绝缘体的缺陷进行检测,应对实际场景中的缺陷图像的稀缺性,使用仿射变换、高斯模糊、亮度变换等数据增强方式,网络选择级联 Faster R-CNN,从而实现绝缘体中的定位和缺陷检测。He^[69]等使用多级特征融合网络将多个分层的不同尺度特征融合,这些特征包含了很多缺陷的位置细节信息,最终有效实现了钢板缺陷检测和定位。Li^[70]等基于 YOLO^[8]检测算法进行特征融合改进,提升对细小缺陷的检测能力,为带钢表面缺陷检测提供了端到端的解决方案;Zhang^[71]等使用 YOLOv3^[72]检测,实现

一个更快、更简单的单级检测器，并且使用数据增强扩充了训练数据集，增强模型泛化性，同时引入 Focal Loss^[73]提升训练能力。Chen^[74]等使用 DenseNet^[75]作为主干网络，得到密集型的 YOLOv3 网络。

语义分割作为像素级检测任务，定位更为精准，展现了缺陷的形状信息，在工业缺陷检测中更具有价值。主流的语义分割网络架构主要包括了两种：1. FCN^[6]系列 (Fully Convolutional Networks)用卷积成替代了传统卷积神经网络中的全连接层，实现了一种端到端的简单语义分割算法；2. U-Net^[5]系列设计了编码器-解码器 (Encoder-Decoder) 结构，利用编码器实现基础的图像多尺度解析，使用图像上采样复原解析度，并融合同尺度的特征，实现了高精度要求，并在医学图像分割领域有所成就。在工业缺陷检测领域，研究人员常常会基于这两种语义分割框架，辅以多尺度特征融合、深度特征融合、网络轻量化等改良方案，设计具有针对性的工业缺陷检测模型。具体示例如 Qiu^[76]等提出了一种三阶段执行的全卷积表面缺陷检测方法，即为先分割，再分类，最后精修，分割阶段使用 FCN 对缺陷区域进行像素级别的分割，检测阶段对分割结果进行更精细的修正，去除缺陷误判项，修边阶段对检测结果的边缘进行微调，使其预测结果更加精确。Tabernik^[77]等设计一个在大分辨率图像中检测小的表面缺陷的算法，在较高的层中增加一个下采样层并使用较大的卷积核尺寸，增加感受野的大小，增强网络捕捉小细节的能力。Huang^[78]等人在 U-Net 架构的基础上进行优化，设计一种显著性检测模型检测磁瓦的表面缺陷，对于缺陷图片进行多种显著性方式的处理，强化分割效果，同时使用边界框突出显示预测的缺陷，提高缺陷检测精度，达到了显著性检测的目标，然而，面对不同显著程度的缺陷，在仅考虑空间域的情况下仍然存在一定的局限性。针对频域的方法^[79,80]选择在图像频域上也进行分析，使用小波变换在频域上提取有价值的信息，然后与图像的空间域上的信息相结合，扩充了信息的价值，提升了模型在多尺度特征上的细小缺陷的异常检测能力。

1.3 钢材表面缺陷检测项目背景

基于深度学习的钢材表面缺陷检测研究，源于与上海某科技公司合作的基于机器学习的钢材缺陷检测系统项目。数据由某真实工厂提供，包含正常样本和具有缺陷标注的异常样本。

本项目的目标是设计基于深度学习的带钢上表面缺陷检测系统。该系统通过前端摄

像机实时捕捉在线带钢运行实时高清图像,通过视觉算法识别表面折印、边部边裂等肉眼可清晰分辨缺陷,用于带钢生产过程中的缺陷检测。基于深度学习的语义分割算法,目标是开发一套完整的带钢表面缺陷检测系统,实现的算法功能包括表面折印检测、边裂检测与带钢机床区域定位三个功能模块,并针对实际工业数据和精度的要求,对模型进行优化及算法性能提升。

项目针对当前的工业中金属钢材表面的缺陷问题提供有效的解决方案。考虑到钢材生产加工非常恶劣的环境,工业生产中每一道工序都有可能会对金属钢材表面质量造成一定程度的影响,导致生产产品质量无法保障的问题以及生产过程中的严重工业安全问题,所以只要金属钢材表面一旦出现了折印、边裂等具有重要影响因素的缺陷,如果不及时进行检测并发现就会直接给企业带来重大的损失和商业影响。

同时,对于工业界的实际数据,存在背景环境复杂、数据获取困难、数据缺陷标注不明确、缺陷特征不明显等等相关问题,致使工业缺陷检测的研究难度加大,亟待具有更好实际场景适应性的工业缺陷检测方案,来为无人质检、智能巡检、生产控制和异常溯源等场景提供便利,为广大的生产制造群体提供高效的解决方案。

因此,为了应对上述实际工业生产场景的问题,为了响应国家制造业改革,加速推进信息化与工业化的深度融合,推进生产的智能化的过程。本项目的面向钢材表面缺陷检测研究有非常重要的意义和价值。

1.4 本文的主要研究内容和创新点

整体上看,本文以基于语义分割的钢材缺陷检测算法作为研究目标,设计了一个半监督对比学习的工业缺陷检测的语义分割算法模型架构,充分利用真实工业场景中已取得的正常样本,对比学习异常缺陷的位置信息,得到更高性能、更低数据依赖的语义分割模型。

从局部来看,本文设计了一个半监督对比学习的语义分割网络架构和面向重建和判别网络的优化方案来实现上述目标:

本文的第一个工作提出了一个基于异常重建网络与对比判别网络组合的半监督对比学习的语义分割网络模型架构。该模型采用两阶段的训练方式进行,第一阶段充分利用正常样本,使用柏林噪声构建类似真实场景的缺陷异常,这些缺陷将是随机位置、随机纹理、随机形状的,对正常样本进行异常模拟。然后通过训练学习模拟异常样本复原为

正常样本的重建网络，得到最终的异常重建网络。第二阶段输入异常样本，得到重建网络输出的复原重建样本，合并两个样本形成对比，输入判别网络，最终得到分割检测的结果。实验部分采用真实场景的钢材表面缺陷数据集，对钢材表面的边裂、折印缺陷以及机床区域进行检测和定位，在该数据集上实验对比现有的语义分割算法，证明所提出的算法架构的优越性，并且在数据量减少的情况下，依旧保持较好的稳定性。

本文的第二个工作提出了面向重建和判别网络的优化方案。针对异常重建网络，设计掩膜特征学习模块，其结合掩膜空洞卷积和 Transformer 的注意力机制，充分利用了掩膜空洞卷积的更大感受野优势，加强了异常样本的复原能力。针对对比判别网络，设计对比注意力学习模块，促使样本对比学习通道和空间上的注意力信息，优化分割效果。实验部分通过对比未优化的基础网络架构，证明了优化方案的效果。

1.5 本文的结构安排

第一章，绪论。主要介绍了语义分割和工业缺陷检测的研究背景及意义，然后分析了国内外的研究现状，重点介绍了基于深度学习的语义分割算法和工业缺陷检测算法，以及项目的背景。最终总结了本文的主要研究内容和创新点。

第二章，基于语义分割的工业缺陷检测背景知识。首先分析了工业缺陷检测现存的问题与挑战，并针对性地分类阐述不同真实工业场景下的语义分割算法，包括了三种数据模式场景下的语义分割方案，最后是本章的内容总结。

第三章，半监督对比学习的网络架构设计。首先分析现阶段语义分割模型在工业缺陷检测领域的问题，并针对性地阐述了本章所提出的模型设计思路。随后介绍了设计的网络架构和损失函数的选择。紧接着介绍并展示了本章使用的网络在真实钢材表面缺陷数据集上的实验细节和相应的对比实验，最后是本章的内容总结。

第四章，面向重建和判别网络的优化模块设计。首先分析了当前网络架构中可优化之处，随后介绍了本章方法的设计思路。之后详细介绍了所设计的两种优化模块 MFLM 和 CALM。紧接着介绍了实验细节并利用对比实验展示了本章的优化模块方案在钢材表面缺陷数据集中针对网络的优化效果，最后是本章的内容总结。

第五章，总结与展望。总结本文的研究内容，归纳本文的创新点并展望未来的研究方向。

其中，第三、四章是本文的重点章节。

2 基于语义分割的工业缺陷检测背景知识

作为后续章节的背景知识，本章将详细介绍基于语义分割的工业缺陷检测的研究发展内容、问题与挑战以及细分任务领域。

2.1 基于语义分割的工业缺陷检测简介

工业缺陷检测作为工业视觉领域中最重要的高新技术研究领域之一，在伴随着人工智能、深度学习的快速发展之下，很好地实现了不同场景的实际算法落地。得益于卷积神经网络的强大特征提取能力，深度学习在工业检测领域的效果相比于传统算法有着极大的优势，其中基于语义分割的工业缺陷检测算法更是由于其满足高精定位需求和强化的泛化能力，逐渐占据了深度学习在工业缺陷检测的主流位置。起初为了适配工业缺陷检测场景，在语义分割算法的基础之上，研究人员加入一些辅助技术来提高算法的实用性，比如使用数据增强或者数据合成的方法解决数据不足的问题，为模型提供多类型的训练集以强化训练；使用模型压缩和算法加速的技术解决对于存储、算力有限或者对算法推理的实时性有一定要求的情况。这些辅助方法在一定程度上改良了语义分割算法，解决了一部分工业缺陷检测的问题，但是仍存在一定的实际限制，这就需要对语义分割算法进行整体的改进，设计面向真实工业场景需求的工业缺陷检测算法。

基于语义分割的工业缺陷检测算法根据实际工业的不同场景中数据标注情况可以大致分为 3 种不同的类别进行分析，分别是缺陷模式已知、缺陷模式未知与少量缺陷标注 3 种。

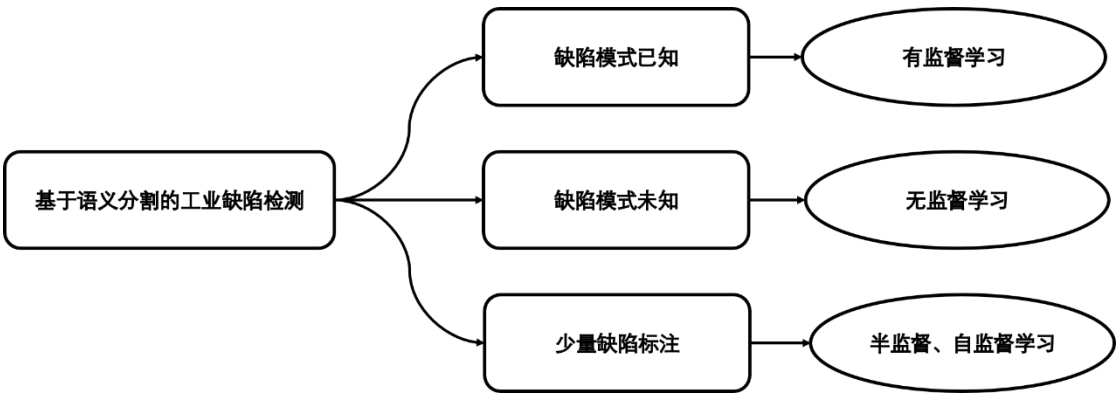


图 2.1 基于语义分割的工业缺陷检测任务

2.2 工业缺陷检测的问题与挑战

无论是图像分类、目标检测还是语义分割任务，在工业缺陷检测的实际场景应用时都会遇到一些现实因素的问题，比如工业缺陷数据样本和部署环境的多样性、复杂性等，因此既有困难，又有挑战，需要根据具体任务需求做具体分析。本节将分析工业缺陷检测的真实数据难点以及工业缺陷检测任务所面临的一系列挑战。

2.2.1 工业缺陷检测数据难点

基于语义分割算法的工业缺陷检测在大量数据标注的情况下已经获得了一定的效果，也能基本满足了各种任务场景的基础目标需求。然而面向工业实际场景，诸多存在的问题仍然未能得到很好的解决。比如真实工业场景之下，是否能够对大量的缺陷样本提供标注？大量的正常样本导致的正负样本不均衡问题是否可以很好地解决？不同工业场景下的缺陷具有多样性，是否能在定义缺陷时就能很好的描述？现有模型对数据的高度依赖与真实场景的有限数据体量的矛盾如何解决？这些问题所指向的就是对于工业缺陷检测数据的实际难点的反馈，具体难点细分如下：

(1) 数据样本不足。工业场景下所能提供的数据数量是相对有限的，一方面由于工业大规模生产过程中对次品率有很高的要求，缺陷的出现必然是小概率的事件，所以所能采集到的正负样本会出现明显的不均衡现象。另一方面，加入人工标注的缺陷样本更是难以获取。由于缺陷需要精确的标注，人力成本是非常高昂的，对于稀少的缺陷样本需要花费时间查找并精确标注。

(2) 缺陷定义未知且主观。对于缺陷的定义，需要花费大量的时间讨论，同时还需要整理大量的可能性样本进行比对，所以缺陷的定义一直都是很模糊的，人为的定义毕竟过于主观，很难能把握好准确的缺陷描述，甚至可能会导致数据标注的错误理解。更何况人为归纳局部缺陷是无法覆盖全局类型的缺陷的，这会导致一些算法模型对与新的异常缺陷无法检测。

(3) 工业缺陷可视性差。一方面，缺陷的存在必然是细小的，一般是难以察觉的，对于图像来说，其涉及的区域往往都是极小一部分；另一方面，由于工业环境的影响，视觉的背景会存在一系列的噪点、亮度、对比度的问题，缺陷的特征不够明显，这会影响对于缺陷的判断，也影响标注的精度和模型的训练。

(4) 工业场景复杂多样。一些常见的工业场景数据集主要针对工业产品搭配简单统

一的产品背景，然而在真实工业场景中，工业环境是极其复杂的，产品本身具有复杂的结构与纹理，其背景也是复杂的，比如不同的光线、不同的视角、不同的阴影、不同的背景物体干扰等等。

2.2.2 工业缺陷检测任务挑战

(1) 高精度需求。工业缺陷检测算法不仅对于缺陷的识别精度有所要求，对于语义分割的工业缺陷检测算法来说，还需要对于缺陷的定位精度有所要求，对于缺陷的形状能准确勾勒，这就需要有高精度、大批量、高可靠性的数据标注来保证模型训练。同时也需要对于工业场景中的所有缺陷，不论大小、类型、形状，都是应检尽检，这对于算法的性能要求是极高的。

(2) 低开销成本。随着所使用的图像分辨率越高，模型的计算量和参数量都会相应地爆炸式增长，这对模型的效率有极大的影响，所以对于真实工业场景，需要模型在算法的效率上有所保证，在精度提升的同时，模型的规模得到一定的限制，即为模型轻量化。对于数据来源的成本也需要得到很好的控制，希望能在更少的数据上实现较高的模型性能，降低数据的依赖。

2.3 基于数据情况划分的工业缺陷检测任务

本节主要从 3 个角度讨论。已知缺陷标注时，主要是基于有监督深度学习方法，需要充足而精确的样本标注前提。未知缺陷标注时，主要是基于无监督深度学习方法，通过构造比较的对象，生成伪标签，根据图像和特征的相似度进行伪标签的训练学习。不同于已知缺陷标注和未知缺陷标注这两种极端的理想情况，真实的工业场景往往具有大量的正常样本和少量的有标注的异常缺陷样本，这种情况称之为部分缺陷标注。这种情况下，常常需要考虑结合已知缺陷标注和未知缺陷标注两种情况的方法综合考虑，也就是将有监督和无监督的方法进行结合，所以通常会考虑采用半监督、自监督乃至弱监督等等一系列的方法，对正常样本进行辅助学习，然后对异常样本进行有监督的标注学习，获取精准缺陷定位验证，对于正负样本都有了很好的用武之地。这一类方法在工业缺陷检测的实际场景中得到了广泛的应用。

2.3.1 缺陷模式已知

在第 1 章中我们已经重点介绍了有监督的深度学习工业缺陷检测算法，有监督的语义分割模型的基本结构主要分为 FCN 和 U-Net 两种，其中 FCN 生成的结果和输入的分辨率是不同的，而 U-Net 通过上采样或者反卷积的方式，实现了输出分辨率的复原。有监督方法在缺陷模式已知的情况下广泛使用，在数据均衡和标注有所保证的情况下效果较好。

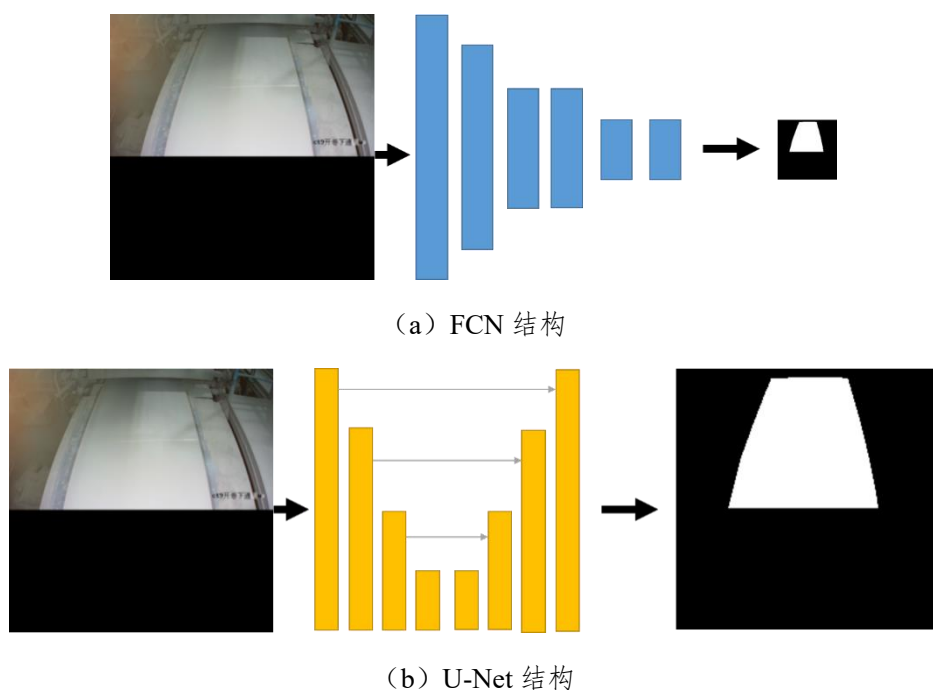


图 2.2 有监督语义分割网络的基本结构

但由于有监督方法所需要的数据环境要求非常之高，实际工业落地的难度很大。实际情况中，工业难以提供充足且均衡的数据，或者不能提供很好的数据标注样式，即使利用图像的数据增强技术来扩充数据多样性，均衡数据比例，也无法有效地解决这类数据问题。于是，研究人员尝试利用数据合成技术来生成异常缺陷，可以理解为是模拟数据标注，然后利用有监督的训练方式完成网络优化。具体示例如 Valente^[81]等人合成训练打印缺陷数据，通过图像处理和计算机图形学技术来模拟打印缺陷和打印扫描效果，应用于语义分割网络的训练。Zhang^[82]等人使用生成对抗网络在具有不同纹理和外观的各种图像背景中生成真实的缺陷来扩充训练数据集。除此之外，还有一些方法^[83,84]利用剪贴的方式实现正常样本的异常模拟，完成自监督学习。

有监督的方法对于已知条件充裕,数据均衡,数据标注充足的情况,有很好的应用价值。当然,这也需要对数据标注有较高的要求,不可避免带来了高昂的标注成本。对于真实的工业缺陷检测场景,虽然在理论上此类方法在数据和标注保障的情况下具有优良的性能,然而现实是真实场景的缺陷过于复杂,人肉眼都不一定能快速判断,有监督学习的方法表现也会有一定的精度问题,仍有很大的发展空间,对于目前的工业缺陷检测场景仍不能很好地适配,存在不足。

2.3.2 缺陷模式未知

在真实的工业场景中,异常的出现是极少数的情况,所以有缺陷的异常样本是难以获取的。对于工业不同任务场景,缺陷存在未知性、多样性和不规则性,而无监督的方法,不使用样本标签,完全不依赖人工的判断,实现一定程度上的分割效果。异常缺陷被理解为是在正常样本的基础上有所差异的模式,通过这种差异程度的判断,实现无监督的异常检测算法在工业缺陷检测上的应用。对于无监督学习的语义分割,同样会对缺陷区域进行精准位置定位,实现像素区域上的分割效果。

传统的无监督方法如模板匹配法,通过轮廓^[84]、特征点^[85]、极值区域^[86]等特征信息,对异常缺陷样本和正常样本进行简单差值对比,实现缺陷的定位。这种方法对于图像有一定的特殊要求,需要图像之间尽可能保持对齐一致,对于主体目标位置随机、背景多样化的情况不适用,而且对于模板的匹配,需要消耗大量的时间进行逐一的计算,模板数据分布选择需要足够充分,具有典型的代表性,并且数据体量也不能太大,否则模型的计算量和参数量将会巨大。基于稀疏编码的方法是使用线性组合的基向量来表征多样的正常样本数据信息,通过学习正常样本的特征,构建具有普遍正常特征描述的最佳稀疏矩阵,具体在管道^[87]、织物纹理表面^[88]等多样的工业缺陷检测场景中有所应用。其他的方法还包括,基于图像特征统计^[89,90]、低秩分解^[91]的方法。这些传统的方法能实现基本的缺陷检测,但随着场景的复杂化和缺陷的特征削弱,传统方法的局限性也将愈加明显。

基于深度学习的无监督方法,同样针对正常样本进行模式学习。借鉴和传统思想相同的模板匹配的基本思维,对于正常样本训练学习其特征模式,也就是说,在训练阶段,仅针对正常样本进行学习,这一类的方法因为其缺陷没有定义,通过正常样本的学习,获得具有普遍正常样本特征的描述,匹配异常样本时能学习到一些未知模式的缺陷,这

也是相较于有监督学习的优势所在。得益于深度学习、语义分割技术的成熟，相比于传统方法，基于深度学习的无监督方法也已显示出了优异的特征表达能力，因而成为无监督工业缺陷检测领域的主流方法。同样是模板匹配的思路，在深度学习方法领域，根据差异比较的维度的不同，主要分为基于图像重建的方法与基于图像复原的方法两种。

(1) 基于图像重建的方法

这种方法是基于图像的像素维度的对比，核心的思路就是通过重建方法，对于输入的样本构建其正常样本的图像，然后进行两个图像的差异的对比，从而学习捕获缺陷的位置信息。下面会对此类的一些常见方法做介绍：

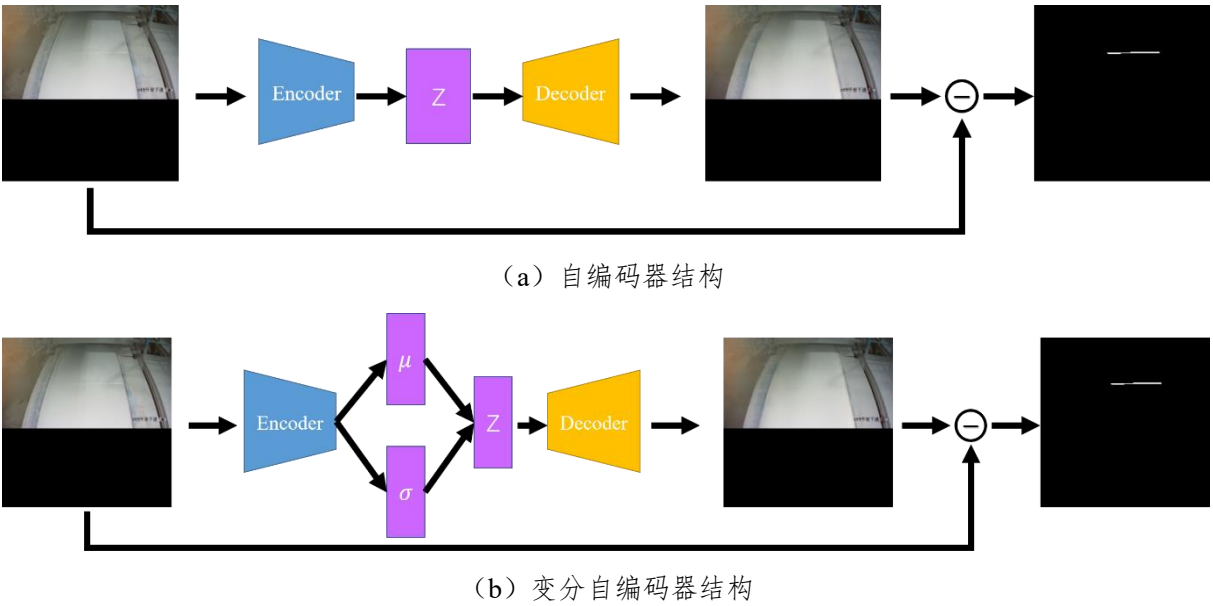
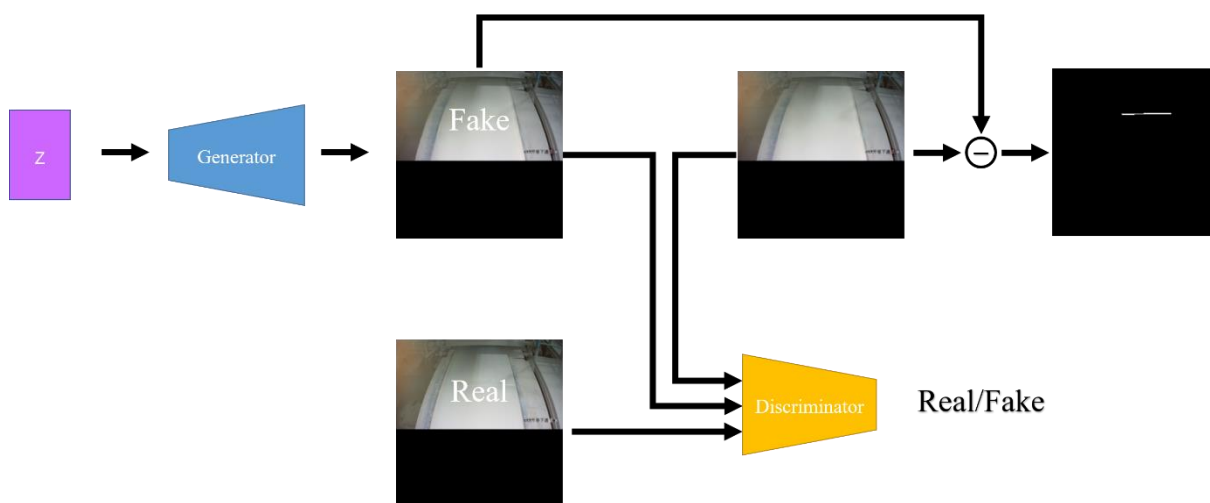


图 2.3 自编码器类的结构

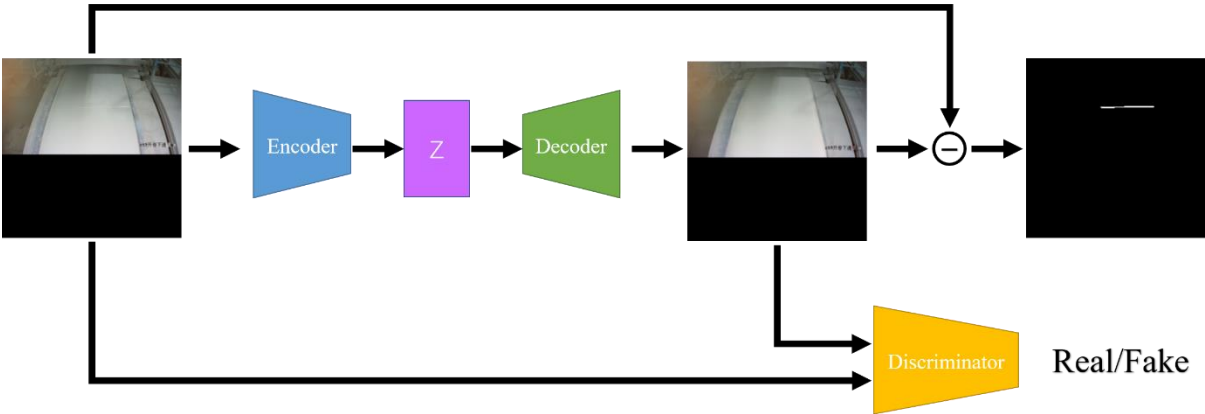
基于自编码器（Auto-Encoder，AE）的方法，如图 2.3(a)所示，结构采用编码器-解码器的基础结构，编码器用于对图像进行编码，得到隐空间变量 z ，之后利用解码器对 z 进行重建处理，生成工业图像的复原样本，将复原样本和输入的样本图像进行对比，计算误差作为损失，从而训练基于自编码器的重建网络，最终训练好的重建网络通过对比输入和输出的图像，实现精确的缺陷定位。对于这种方法，核心是希望重建复原的图像更加清晰且图像质量有所保障，所以可以在多尺度的信息上进行获取并对比上下文信息，从而对图像特征有更好的捕获和重建。具体示例如 MS-FCAE^[92]和 DFR^[93]就是针对多尺度特征进行优化，为图像重建提供更好地设计，使得重建图像的质量有所提高。MS-

FCAE 基于少量无缺陷的纹理样本, 利用不同尺度的多个子网络来重建多个纹理背景图像, 通过从输入图像中单独去除这些纹理背景来获得残差图像, 最后将它们融合到一个缺陷图像中, 子网络通过特征聚类来提高编码特征图的判别能力, 实现了高效的检测精度。DFR 则是设计了一种多尺度区域特征生成器, 可以在预训练的深度卷积网络中为输入图像的每个子区域生成多个空间上下文信息的感知表示, 对各个子区域包含的局部特征进行编码, 结合了空间上下文信息, 使编码信息具有鉴别能力, 对工业异常缺陷检测非常有益。除此之外, Chung^[94]等人在 AE 的基础上, 针对过度检测的问题, 引入了风格蒸馏网络, 模拟了 AE 失真问题, 通过风格转换, 实现图像的对齐和重建效果。

基于变分自编码器 (Variational Auto-Encoder, VAE) 的方法, 不同于自编码器的方法, 其通过编码器先将图像映射到先验分布之中, 再对分布进行随机采样从而得到隐空间变量 z , 后续的操作与自编码器的方法保持一致, 目的都是重建图像, 由于 VAE 引入了先验分布, 使得 VAE 可以构建更为可控的隐空间信息, 用于更好地学习正常样本图像的共性信息, 由于是随机采样, 重建的结果也难以得到清晰的效果, 于是也诞生一些优化方案。具体示例如 Dehaene^[95]等人通过对从 VAE 的损失函数使用梯度下降, 使用正则化项来增强, 根据重建误差迭代更新, 完善先验分布, 避免了高频信息的丢失, 实现重建更高质量的图像。Zhou^[96]等人结合了 VAE 和高斯混合模型 (GMM), 得到一种混合半监督的织物缺陷检测方法, 利用 VAE 提取特征和重建图像, 然后利用 GMM 进行密度估计, 从而可以更准确地构建缺陷区域边界。



(a) GAN 生成对抗网络结构



(b) AE 与 GAN 结合的结构

图 2.4 生成对抗类网络的结构

基于生成对抗网络（GAN）的方法，可以充分利用 GAN 的分布学习能力，生成高质量的重建图像。GAN 包括了生成器（Generator）和判别器（Discriminator）两个组成部分，利用生成器和判别器形成对抗训练。具体来说，对于获得的图像的隐空间变量 z ，生成器的目的是重建生成一张尽可能与输入图像高度相似的图像，判别器用于判别输入图像的真假。从而形成对抗学习，最终使得重建生成的图像足够逼真，说明重建的图像清晰且真实，最终利用训练好的生成器实现重建网络的功能，并在异常样本出现时，与之形成对比，更好地观察到异常缺陷的位置。具体示例如：AnoGAN^[97]首次提出用 GAN 来实现异常检测，利用正常样本训练生成器，然后固定生成器参数，在测试时对于输入图像匹配最佳的生成样本，如果可以匹配就是正常样本，否则异常，这个方法的优化训练过程需要反复迭代，影响模型的推理速度。所以，优化之后的 f-AnoGAN^[98]在速度上有所提升，它训练分两个阶段，第一阶段就是利用 DCGAN^[99]进行训练，得到一个可以生成正常图像的生成器，保留图像的核心特征，第二阶段使用异常样本图像，通过生成器生成它所对应的正常图像。这符合流形分布定律，即异常图像和所生成的正常图像集中在低维流形附近，它们对应的隐空间向量是近似的。

除了上述的方法，将 AE 或 VAE 与 GAN 相结合的方法也层出不穷，具体而言，使用 AE 或者 VAE 的编码器-解码器结构，解码器视为 GAN 的生成器，也就是充分利用了 AE 和 VAE 的编码器能力，获取图像隐空间变量，并使用 GAN 的生成器的对抗学习能力，获得图像重建的出色效果。具体示例如 Baur^[100]等人使用 VAE 和 GAN 重建图像，得到正常样本的图像，通过将输入图像与重建图像进行比较来检测并描绘脑部 MR 图像中的异常。GANomaly^[101]通过结合 AE 和 GAN 进行联合学习，在高维图像特征空间也

保持一致性,从而减小重建误差,优化缺陷检测效果。**Skip-GANomaly**^[102]在 **GANomaly** 的基础之上进行优化,学习了 **U-Net** 的跨层连接方式,使用具有跳跃连接的 AE 的编码器-解码器的结构,捕获高维图像空间中正态数据分布的多尺度分布,利用 GAN 的对抗性训练方案在高维图像空间和低维潜在向量空间编码两者中获得优越的重建效果。**Bergmann**^[103]等人出于对输入输出的结构对齐的考量,加入了结构相似性的损失函数,有效提升了模型在工业检测领域的性能。**Zhou**^[104]等人利用图像纹理和结构之间的关系来设计用于异常检测的深度神经网络,对于视网膜图像的结构信息进行提取,然后结合编码器的编码特征,输入解码器,以实现重建高质量图像。

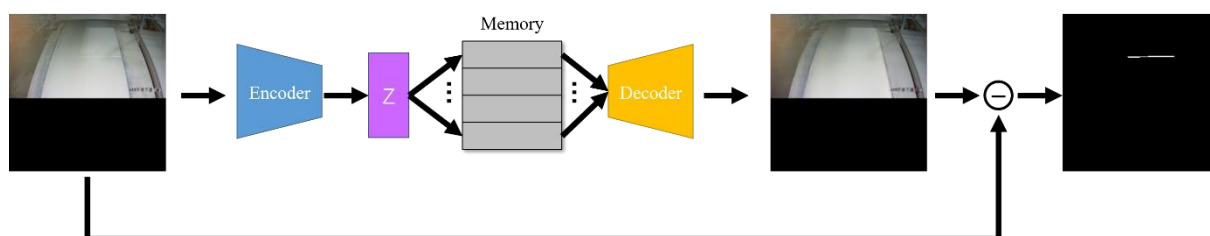


图 2.5 Memory 类的结构

基于 **Memory** 的方法,主要是希望利用一块额外空间对图像的正常样本特征信息进行存储,执行测试时直接对比 **Memory** 库里的正常样本特征信息,计算出差异信息和异常定位从而实现分割效果。具体示例如 **Gong**^[105]等人提出的 **MemoryAE**,也是最早提出这种思想的方法,**Memory** 中存储的是正常图像经过编码器计算得到的隐空间向量,代表了正常图像的普遍特征,并与输入的图像进行根据相似度计算。这种方法不同于直接把图像存入 **Memory**,增加对比和模型推理的时间消耗,使用隐空间向量可以更加灵活地实现重要特征的对比。**Yang**^[106]等人基于 **Memory** 思想,加入了多尺度特征融合机制来提升重建的效果。**Park**^[107]等人将正常特征分为多个聚类中心,提高了 **Memory** 中正常样本数据信息的多样性,优化了重建效果。**TrustMAE**^[108]使用信任区间来防止只包含正常样本信息的 **Memory** 被异常数据污染,强化了 **Memory** 的稳定性,对于工业检测的真实数据情况有了很好的适应性,缺陷检测性能也远远优于 **MemoryAE**,**Memseg**^[109]使用 **VGG** 实现了基础的编码器功能,并且在 **Memory** 中保留了三种尺度的信息作为对比参考,在匹配后融合了多尺度信息和空间注意力机制,实现缺陷的准确定位,但是考虑到真实工业图像能难满足对齐的前提,并且 **Memory** 仍然需要使用大量的额外空间和对比

时间消耗，最终的工业实际场景中效果并不理想。

总言之，基于图像重建的方法，主要针对正常样本进行特征提取，对比输入图像，获取异常差异和缺陷定位，取得一定的优异分割效果。但是也凸显出两个重要的问题：首先，对于输入图像和输出图像，图像重建的方法基本依赖于对齐的前提，然而在真实工业场景下的图像具有多样性，很难满足对齐，这就会导致像素上的偏差，导致不必要的误判。其次，基于正常样本的训练，使用编码器提取特征也存在泛化到缺陷特征的情况，如此就失去了异常检测的优势，重建图像的质量难以保证。同时由于模型的输入与输出基本接近，此类方法可能会导致一定程度上的恒等映射的问题。

（2）基于图像复原的方法

这类方法的目的是实现缺陷噪声的去除，核心思想是在正常图像的基础上加入缺陷特征，应用网络复原成原来的正常图像。复原的网络使用 AE 或者 U-Net 即可实现。训练好的模型具有良好的图像复原能力，可以和输入图像进行对比，计算重建误差，实现分割。由于模型的输入与输出的不同，此类方法能一定程度上避免恒等映射的问题。

这类方法的难点在于如何利用正常样本构造异常缺陷样本图像。往往大多数方法都是利用图像的差异性和掩膜异常覆盖等方式生成的。具体示例如 Sabokrou^[110]等人使用高斯噪声模拟异常，与原图叠加。Fei^[111]等人通过对原始图像中删除选定属性的方式来模拟异常样本，比如图像的颜色、方向等等属性。Salehi^[112]等人选择直接对图像进行分块拆分，对图像块进行打乱重组，形成了噪声图像，用于训练。

以上的方法与真实的工业缺陷或者生活中正式的异常有所区别，所以图像复原的训练价值有限，对于真实缺陷的模拟，更常用的方法是直接对于原图叠加异常的掩膜，用于训练复原去除掩膜。具体示例如 Haselmann^[113]等人对正常样本进行随机地使用掩膜块覆盖，这一类方法的缺点很明显，首先是掩膜块的形状是规则的，这与真实场景的缺陷是完全不一样的，不能很好地模拟异常，其次，掩膜块的颜色与样本本身无关，对于重建网络的复原能力的学习简单化，重建复原的效果一般。Li^[114]等人随机选择图像的超像素作为掩膜，但是这种方法会导致计算量的加大，效率低下。Yang^[115]等人借鉴隐空间聚类的思想，人工构造缺陷生成缺陷图像，之后在隐空间上进行特征聚类，对异常特征进行重建复原去除，实现了最终的分割。Mei^[116]等人考虑真实工业场景下的缺陷尺度差异较大，所以在设计掩膜时，随机化不同尺度和不同样式的掩膜，利用图像金字塔，将图

像缩放到多个尺度并加入噪声,最后融合多个尺度信息用于分割。**Zavrtanik**^[117]等人随机将图像划分出多个子图像块,对这些子图像块进行上下文信息复原,并使用基于梯度相似度的损失函数计算差异,提高分割精度。**Yan**^[118]等人选择在图像空间上的两个方向设计多尺度的条纹掩膜,应用于图像复原。

基于图像复原的方法,主要是充分利用异常构建的知识,设计更合理更真实的缺陷,然后利用重建网络实现特征复原,这种模型学习需要很好的上下文信息提取能力,以及更多图像自身信息的挖掘。对于异常缺陷的设计,需要考虑真实场景的不同尺寸,不同颜色、有无规则性等等去设计异常缺陷的掩膜,也尽可能与原图像保持一定的相关性,符合原图像的真实缺陷分布,同时需要具备缺陷的多样性,避免特征的过于一致导致模型训练的过拟合。

2.3.3 少量缺陷标注

在现实的工业场景情况下,往往是介于无监督和有监督方法之间的设置的,即为不仅仅需要使用相对充分的正常样本,还需要对一定数量的异常缺陷样本进行标注和分析。可以理解为使用正常样本完成无监督的初步训练,再使用缺陷样本对检测进行细化,提升检测效果。由于纯无监督和有监督学习的方法不能全面地利用样本信息和标注信息,所以不能成为最佳的解决方案。针对样本过少、数据正负样本不均衡、标注精度有限等问题,为了充分利用数据和标注的信息,研究人员开始将重心转移到半监督学习的设置之上,来设计更高效、更适配真实工业缺陷检测场景的算法。

在实际的工业缺陷检测场景中,存在大量的正常样本数据,毕竟对于工业生产,成品率需要一定的保障。而对于缺陷数据的获取是很难的,一方面由于缺陷样本的数量极其稀少,另一方面缺陷数据的标注工作需要消耗大量的人力成本。所以对于大量的正常样本和少量的标注缺陷样本,使用半监督学习方法以减少对样本标注的需求和依赖,利用了少量的标注缺陷数据充分地学习了缺陷的精确定位,相较于无监督学习方法的分割效果有明显提升。半监督学习的工业缺陷检测算法具体示例如:**He**^[119]等人利用半监督学习方法来对钢材的表面缺陷进行分类。针对无标注样本,选择使用卷积自编码器(Convolutional AutoEncoder, CAE)提取细节特征,最终用于特征提取器,用于输入半监督生成对抗网络 SGAN^[120],提高了方法的泛化能力,利用生成器和有标注样本的判别器进行训练,得到分类结果,具体包括了正常类别和具体缺陷的几种类别。**Gao**^[121]等人

为了避免模型对于有标注数据的过拟合情况,使得无标签数据变得毫无意义,于是选择利用无标注数据构造伪标签,通过构造模型来对无标签数据生成伪标签,达到协调无标签数据参与度的目的。**Yu**^[122]等人使用了两个不同网络来实现半监督学习的工业缺陷分类任务,分别是生成对抗网络和残差网络,通过 GAN 生成大量的无标签样本,对数据集进行扩充,使用不同学习策略的多分类器进行算法训练,对于判别器和分类器训练分类结果一致的样本,进行伪标签生成,加强分类器的训练,反之则削弱分类器的训练。**Chu**^[123]等人考虑到异常的发生概率很低,数据分布严重偏向于大量的正常样本数据,提出使用一种用于异常检测和语义分割的半监督学习算法,使用异常分类器作为输入,通过自编码器处理样本数据的损失轮廓,对于训练期间的重建损失值构建损失分布,然后引入了一种基于强化学习的元算法,放大异常和正常样本数据之间的损失分布差异,在具有多样的纹理和缺陷目标的数据集中能更好地表征缺陷异常的分布模式。最终可以根据损失函数的曲线变化趋势实现对于像素是否属于缺陷的分割判断。**CAM-UNet**^[124]方法利用足量的正常样本和少量的缺陷异常标注的样本进行训练,第一阶段使用 VGG-16 作为 backbone 的 U-Net 网络作为编码器对正常和异常样本进行分类和粗定位,第二阶段训练两个解码器对特征图和先验知识的激活图进行融合,用两个解码器实现两轮融合,同时对两个解码器进行约束,使两个解码器相互促进,从而细化分割效果。**Kimura**^[125]等人则是在其半监督的基础之上,使用额外的判别器,实现了对于少量有标注的缺陷样本的利用,从而网络能够更好地训练更佳的注意力信息图。

总言之,在工业缺陷检测之中,当训练的数据集存在无标签的缺陷样本,半监督的学习方法就可以很好地解决这一类的问题,不过真实场景下的缺陷特征多样,对于缺陷样本的训练,会束缚网络模型,局限于已知的缺陷特征。所以,对于半监督语义分割任务,更多地是使用无标签的正常样本,利用无监督学习方法,对于正常样本的模式进行学习,然后与少量的有标签的异常缺陷样本进行对比学习,从而优化缺陷定位细节。这样既借鉴了无监督学习的多样缺陷挖掘的特点,也利用了有监督学习方法的强学习能力,对于工业实际场景的缺陷检测具有很好的适应性,检测效果出众。

2.4 本章小结

本章首先介绍了基于语义分割的工业缺陷检测研究现状,包括了基于语义分割的工业缺陷检测的研究发展内容、问题与挑战以及细分任务领域,并详细介绍了基于语义分

割的工业缺陷检测算法在三种数据模式下的细分：依据实际工业场景中数据情况的不同，被分为缺陷模式已知、缺陷模式未知与少量缺陷标注 3 种设置。上述内容为本文后续章节的网络架构设计和算法优化做了铺垫。

3 半监督对比学习的网络架构设计

3.1 问题阐述

半监督学习是一种将有监督学习和无监督学习相结合的学习方法。半监督学习选择使用大量未标注的数据进行学习分析,同时也会使用有标注的数据进行学习优化。这种情况在真实工业场景之中最为常见,因为在实际工业缺陷检测之中,数据集普遍存在大量的无标签的情况,大量的样本难以得到有效的标注。一方面原因是样本获取对于工业界是容易的,但对于样本的标注工作却需要消耗大量的人力成本,这样的工作不仅繁琐而且效率低下;另一方面,由于数据标注是人工完成的,势必会出现标注的一系列问题,比如数据标注存在少标、漏标、标注范围不全、标注类型不正确等问题,这些对于后续的模型学习都会带来不必要的训练影响和精度损失。当然,在数据定义上,缺陷因为过小或者可视性过差等问题,也不可避免地会导致一些视觉上的判断误差,缺陷的定义一旦出了问题,就会导致大量的不同数据标注结果的产生,可能这些标注都是正确的,但也都存在一定的问题,这都是数据定义不统一、不规范导致的一系列问题,在真实工业场景中,这种定义问题只能通过不断地沟通来做一定程度上的弥补,但是难以从根本上解决问题,所以上述的一系列数据问题势必会影响到模型检测的效果。

半监督的对比学习架构,可以结合有标签和无标签的数据一起使用,对于无标签数据来说,没有了人工标注的误差和定义问题,直接使用这些数据,可以保留其真实的信息价值,避免了人工标注的误差,再结合有标签数据作为补充学习,学习人工定义的精确的缺陷类型,从而健壮了整个网络体系。同时针对正常样本和异常样本进行特征对比,形成特征比较,学习类内差距,可以更好地定位缺陷位置。所以,半监督对比学习的网络架构在性能得到保障的情况之下,对数据的依赖有所下降,这也得益于无标签数据的充分利用,模型的性能也有所提升。

3.2 本章方法概述

对上述问题,本章提出了一种半监督对比学习的网络架构,用于学习异常样本复原重建为正常样本的异常重建网络,再形成异常样本和复原重建样本的对比,通过对比判别网络学习缺陷的精确定位结果。从而得到一种“异常重建+对比判别”的半监督对比学习的网络架构(ARCD, Anomaly Reconstruction and Contrastive Discrimination)。

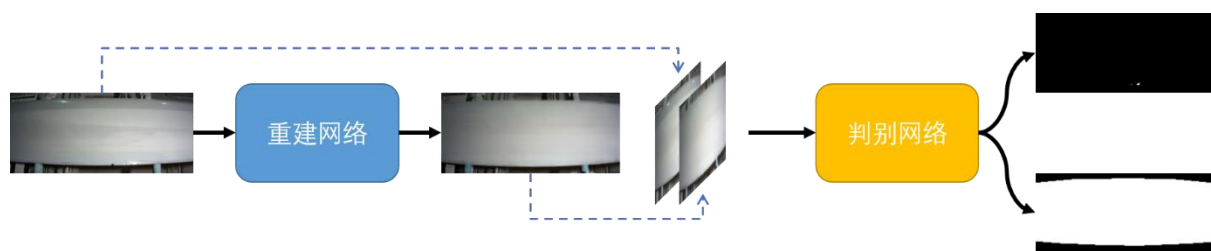


图 3.1 半监督对比学习网络架构

整体上，如图 3.1 所示，对于输入的样本，使用重建网络对其进行样本复原重建，实现“去异常化”，从而得到了该样本的“正常化”复原重建样本，即在缺陷处复原为无异常的状态，然后通过输入样本和生成的“正常化”复原重建样本合并输入到判别网络之中，使用判别网络对比学习两个样本之间的差异，通过语义分割算法输出最终分割结果。考虑到机床区域位置的检测结果和缺陷检测结果会出现覆盖的情况，所以在最终的分割结果上进行区分，分别为针对缺陷的检测和针对机床区域位置的检测两种结果。

针对真实工业场景，钢材表面缺陷数据集中拥有大量的无标签正常样本数据和有标签异常缺陷样本数据。综合考虑下，我们设计了一种模拟异常方法来充分利用正常样本进行重建网络学习，匹配网络架构，并且设计一种两阶段的半监督学习的训练模式，从而实现一个完整的半监督对比学习的网络模型架构 ARCD。

具体而言，本章的主要工作和贡献如下：

(1) 本章针对工业检测场景，提出了一种“异常重建+对比判别”的网络模型架构 ARCD，有效优化了工业缺陷检测性能。

(2) 本章针对工业检测场景，提出了一种工业缺陷异常模拟方案，它能针对真实工业钢材表面缺陷数据集进行数据的预处理、异常模拟、异常融合等操作。

(3) 本章针对工业检测场景，提出了一种两阶段的半监督学习的训练模式，它能先后学习无标签的正常样本和有标签的异常样本的信息。

3.3 算法设计

本节将详细介绍半监督对比学习网络架构中的各个模块的算法细节，并且针对损失函数进行介绍。

3.3.1 异常模拟模块

模拟异常方法：对于正常样本，因为这些正常样本是没有标注的，所以需要使用无监督学习的方式进行学习。需要通过模拟异常，生成模拟工业真实场景的异常柏林噪声，与正常样本进行结合，然后用于异常重建网络。

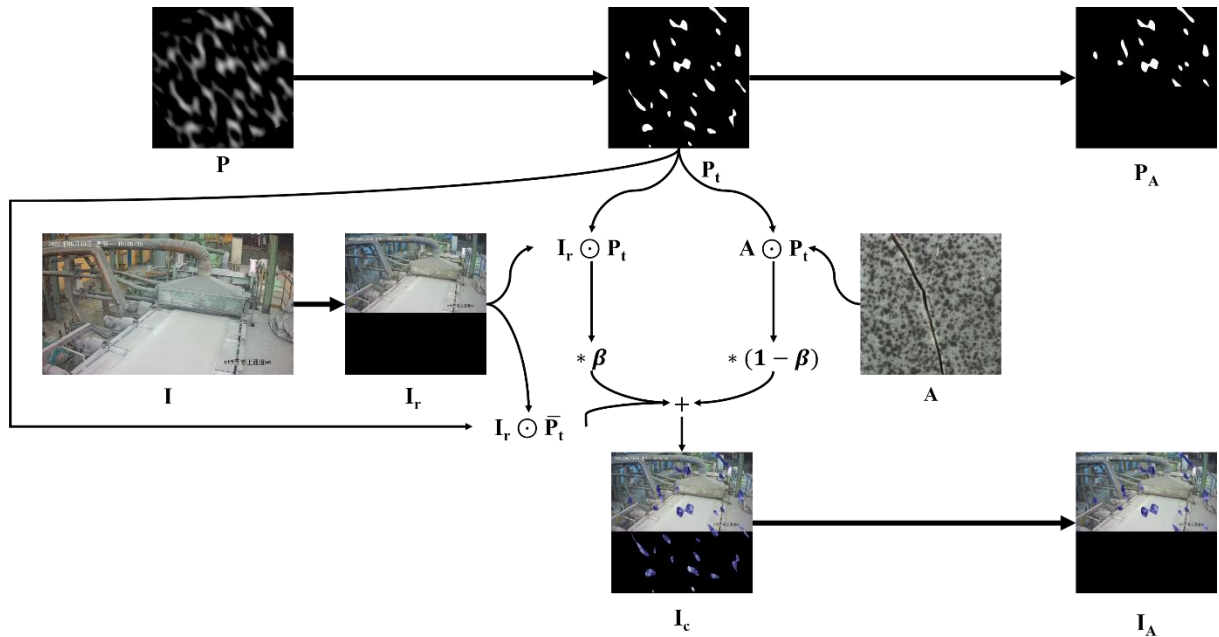


图 3.2 异常模拟模块

具体步骤如下：

首先，随机生成噪声图像 P ，噪声图像由 Perlin 噪声^[126]生成，Perlin 噪声可以很好地模拟各种随机缺陷，可以模拟不同形状和大小的异常缺陷。由于 Perlin 噪声一开始是并非二值化的图像，所以不能直接用来作为缺陷异常图像使用，所以需要通过随机均匀采样的阈值进行二值化操作从而得到了如图中所示的 P_t 图像。

对于异常的模拟，我们不仅需要用 P_t 二值化图像来实现模拟缺陷的位置随机化，同时我们也需要对缺陷的纹理信息进行随机化，这里我们引入了一个随机的纹理图案库，对于不同的纹理随机地选取，作为缺陷的纹理，如图中的 A 即为所选取的随机的异常缺陷纹理图案，基于这个纹理，需要辅以随机化的数据增强^[127]来加强模拟缺陷的多样性，包括了旋转、仿射变换、图像亮度、图像锐利度、图像均衡值、图像对比度与饱和度等等的随机增强，从中随机选取三种使用。增强后的 A 与异常二值化图像 P_t 相覆盖，得到了随机纹理和增强的模拟异常缺陷部分。

对于原图像 I ，由于数据集中的图像是统一的 1440×2560 的分辨率大小，考虑到大部分的缺陷过小，所以如果对图像进行强制的分辨率压缩到 256×256 ，以匹配网络输入的分辨率，会导致细小缺陷的信息有很大的损失。所以考虑到这一问题。我们选择对于原图先进行 2560×2560 的方形图像叠加，然后再整体进行 256×256 分辨率的压缩。对于细小缺陷的形状信息做了一定程度的保留，对后续的检测有一定帮助，最终得到了如图中所示的压缩图像 I_r ， I_r 与异常二值化图像 P_t 相覆盖，得到了包含原图像信息的模拟异常部分，同时 I_r 与异常二值化图像 P_t 的反向取值图像相覆盖，得到了无模拟异常的图像区域部分。

最后，混合三个部分，得到模拟的异常图像 I_c ：

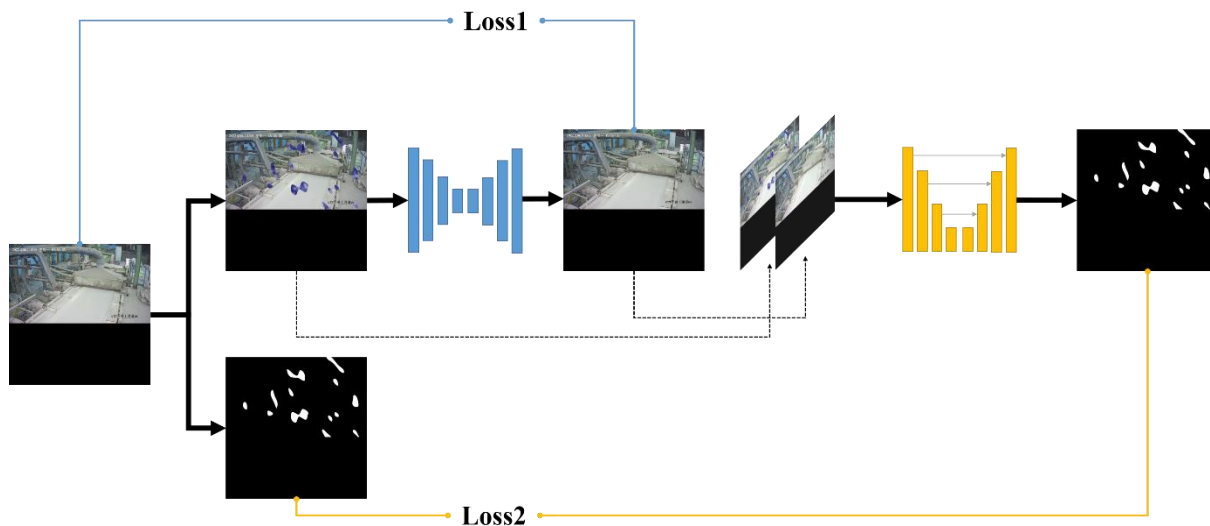
$$I_c = I_r \odot \bar{P}_t + \beta (I_r \odot P_t) + (1 - \beta)(A \odot P_t) \quad (3-1)$$

其中， β 表示比例系数，是原图像和模拟异常缺陷之间的比例融合，这样既设计了一种随机的噪声异常，又结合了原图像本身的信息，使得最终的模拟缺陷异常的信息更接近原图像的像素分布特性。 \odot 表示图像矩阵之间的逐像素的点乘操作， \bar{P}_t 表示异常二值化图像 P_t 的反向取值图像。

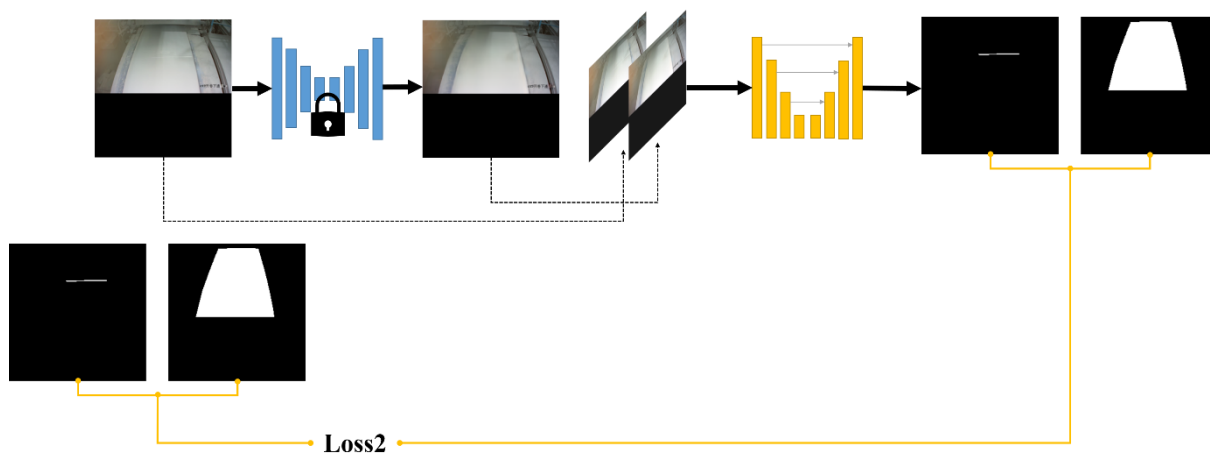
由于考虑到原图的实际范围，对于模拟的异常图像 I_c 和模拟异常 P_t 进行区域截取得到了最终的结果： I_A 和 P_A 。

3.3.2 异常重建网络和对比如判别网络模块

异常重建网络和对比如判别网络模块分两个部分组成：异常重建网络（图 3.3 中蓝色网络部分）使用编码器-解码器的结构，即为基本的特征降采样和上采样的卷积网络实现，将输入的异常样本图像的局部模式转换为更接近样本分布的模式，该网络的训练输入样本使用模拟异常样本。对比如判别网络（图 3.3 中黄色网络部分）使用生成的复原重建样本和原输入样本作为网络输入，输出两个异常缺陷和一个机床区域定位的检测结果。从而构建一个“异常重建+对比如判别”的半监督对比学习架构 ARCD。针对这一网络，我们设计了两阶段的半监督学习训练模式，即为先训练正常样本和重建网络，后训练异常样本和判别网络，如此可以充分利用以往不可使用的正常样本，发挥其信息价值，对异常样本的重建起到了关键性的作用。



(a) 正常样本的异常重建网络的训练阶段



(b) 异常样本的对比判别网络的训练阶段

图 3.3 两阶段半监督学习训练模式

两阶段的半监督学习训练模式的两个阶段具体如下：

第一阶段，仅输入正常样本的模拟异常样本进行学习，对于异常重建网络进行学习，生成“去异常化”样本输出，然后与实际正常样本进行损失（图 3.3(a)中 $Loss1$ ）计算，训练异常重建网络，得到一个稳定的异常重建网络。需要注意的是，对于正常样本，没有任何标注，所以对于后续的对比判别网络可以选择性进行训练，判别网络的训练损失（图 3.3(b)中 $Loss2$ ）优化主要在第二阶段体现。另外，正常样本训练阶段生成的模拟异常样本仅是模拟缺陷异常，不能模拟机床区域定位的分割，所以不会涉及机床区域定位的训练学习。

第二阶段, 仅输入异常样本进入已经预训练好的异常重建网络之中, 生成“去异常化”样本, 然后和输入的异常样本进行通道合并, 一起输入对比判别网络, 学习输出缺陷检测结果和机床区域定位结果。该阶段对于在第一阶段已经训练好的异常重建网络将不进行任何训练调整, 直接用于对异常样本的复原工作, 为对比学习提供了良好的条件。

3.3.3 损失函数设计

(1) 重建网络的损失函数

L1 损失函数表示最小绝对误差, 也被称为最小绝对值误差 (LAE), 即为预测值和真实值之间差值的绝对值差异。具体公式如下:

$$L1Loss = \sum_{i=1}^N |F(x_i) - y_i| \quad (3-2)$$

其中, y_i 代表第 i 个样本 x_i 的真实值, $F(x_i)$ 代表样本 x_i 的模型预测值, $L1Loss$ 表示了所有样本的绝对值差异之和。

L2 损失函数^[128,129]是广泛用于基于重建的异常检测方法的损失函数, 也被称为最小平方误差 (LSE), 表示预测值和真实值之间差值的平方, 它假设了相邻像素的独立性, 不考虑像素之间的相关性, 使用损失进行硬性地计算差异。具体公式如下:

$$L2Loss = \sum_{i=1}^N (F(x_i) - y_i)^2 \quad (3-3)$$

其中, y_i 代表第 i 个样本 x_i 的真实值, $F(x_i)$ 代表样本 x_i 的模型预测值, $L2Loss$ 表示了所有样本的差异平方之和。

对比 L1 和 L2, 二者都是硬性地计算模型预测值和样本真实值的差异的指标, 都是用于最小化误差。虽然 L1 损失函数相比于 L2 损失函数的鲁棒性更好。这是因为 L2 范数将误差进行了平方化, 对于误差大于 1 的情况会放大很多, 敏感度更高, 所以在模型误差计算上一般都是 L2 范数更大一些。然而往往很大可能这种较大误差的样本就是一个异常值, 模型就会尽力去调整以拟合这种异常值, 那么就会导致训练的模型的出现偏差。而对于图像重建任务, L2 反而会更加有效, 因为, 图像重建任务中通常预测值和真实值之间的差异不会太大, 因此需要用 L2 损失来放大强调差异的存在, 进而加快模型的训练优化。而且 L1 存在一定的特殊问题, L1 计算的梯度在极值点会发生突变, 即使是很小的差异也会计算出很大的梯度, 需要学习率的调整, 不利于模型的学习, 相比之下 L2 自身就会对梯度进行一定的缩放, 对于学习率的调整依赖性不强。

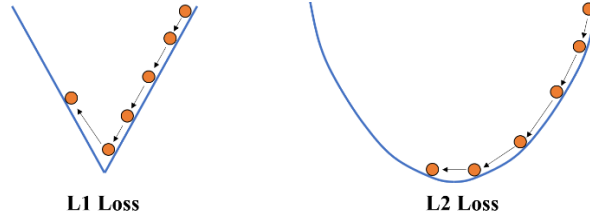


图 3.4 损失函数对比

SSIM^[130]结构相似性的一种衡量两幅图像相似度的指标之一。人们对于像素的绝对亮度和颜色是相对不敏感的，而对于边缘和纹理的位置信息是非常敏感的，SSIM 就是通过主要关注这种边缘和纹理的相似性信息，来捕获一些人类感知的信息。

SSIM 将来两张图像的相似性分为三个维度进行比较，分别是亮度相似性、对比度相似性以及结构相似性，具体如下：

(a) 亮度 (luminance) 相似性：

$$\mu_x = \frac{1}{N} \sum_{i=1}^N x_i \quad (3-4)$$

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad (3-5)$$

其中 $l(x, y)$ 即为在图像 x 和图像 y 之间的亮度相似性， C_1 常数是为了去掉分母为零的情况，可以看出亮度相似性的取值范围在 $[0, 1]$ 之间。其中， x_i 代表单张图像每个像素点， N 为单张图像像素点总数， μ_x 表示单张图像的像素均值。

(b) 对比度 (contrast) 相似性：

$$\sigma_x = \left(\frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)^2 \right)^{\frac{1}{2}} \quad (3-6)$$

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (3-7)$$

其中 $c(x, y)$ 即为在图像 x 和图像 y 之间的对比度相似性， C_2 常数是为了去掉分母为零的情况，可以看出对比度相似性的取值范围在 $[0, 1]$ 之间。其中， x_i 代表单张图像每个像素点， N 为单张图像像素点总数， μ_x 表示单张图像的像素均值， σ_x 表示单张图像的像素值的标准差，也就是图像明暗的变化剧烈程度。

(c) 结构 (structure) 相似度

$$\begin{aligned} s(x, y) &= \left(\frac{1}{\sqrt{N-1}} \frac{x - \mu_x}{\sigma_x} \right) \cdot \left(\frac{1}{\sqrt{N-1}} \frac{y - \mu_y}{\sigma_y} \right) \\ &= \frac{1}{\sigma_x \sigma_y} \left(\frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y) \right) \end{aligned} \quad (3-8)$$

考虑到其中的协方差，以及防止分母为 0，设置 C_3 常数：

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3} \quad (3-9)$$

令 $C_3 = C_2/2$ ，简化得到最终的 SSIM 公式如下：

$$\begin{aligned} SSIM(x, y) &= f(l(x, y), c(x, y), s(x, y)) \\ &= l(x, y) \cdot c(x, y) \cdot s(x, y) \\ &= \frac{(2\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \end{aligned} \quad (3-10)$$

由此可以看出 SSIM 在三种维度提供了相似性度量，在图像重建领域有着重要的使用价值。

在图像超分辨率重建网络的双通道残差网络^[131]中指出，对于为了在图像重建任务中应对图像高频细节的不足问题，为了使重建网络的复原能力达到人眼相似的主观视觉感受能力，一般选择结合 L1 或者 L2 这一类硬性损失函数，再配合 SSIM 这种结构相似度作为损失函数，如此能够更好地保留图像的视觉效果并且表现出一定的优越性。所以，我们最终在重建网络上使用 L2 和 SSIM 的组合，通过实验比较，考虑到 $L2Loss$ 和 $SSIM(x, y)$ 之间实际量级差距，我们最终确定损失函数为 $L2Loss + \frac{1}{100}(1 - SSIM(x, y))$ 。

(2) 判别网络的损失函数

考虑到真实工业缺陷检测的图像特点，异常缺陷往往是占据极小的像素部分的，对于这一类较小范围的异常，希望模型能够学习拟合，然而相较于大面积的背景和正常像素区域，模型的损失计算很难因为较小的异常损失而发生改变，所以针对判别网络，需要建立一个更合适的损失函数。

Focal 损失^[73]的使用就是针对上述的问题一种很好的解决方案，它主要就是为了解决检测中正负样本数量极不平衡的问题。

Focal 损失是基于二分类任务交叉熵损失 **CE** (Cross Entropy Loss) 的。**Focal** 损失具有动态的缩放能力, 通过动态缩放因子, 可以动态调整样本训练的权重, 将训练的重心转移聚焦到一些难例样本之上。

CE 是一种非常常见的分类损失函数, 相比于利用均方误差损失 + **Sigmoid** 实现的分类损失, 收敛速度更快, 能衡量细微的差异。**CE** 的公式如下:

$$CE(p, y) = \begin{cases} -\log(p) & \text{if } y = 1 \\ -\log(1 - p) & \text{otherwise} \end{cases} \quad (3-11)$$

其中, 对于二分类任务, 真实值 y 常设置为 0 和 1, 分别代表背景和前景目标。预测结果 p 的取值范围为 0~1, 表示预测像素是前景目标的概率。

基于 **CE** 优化后的 **BCE** (Balanced Cross Entropy) 已经开始针对类不平衡的问题下手了, 其引入了权重因子 $\varepsilon \in [0, 1]$, 对于正样本来说, 权重因子就是 ε , 而对于负样本来说, 权重因子则是 $1 - \varepsilon$ 。具体公式如下:

$$BCE(p, y) = \begin{cases} -\varepsilon \log(p) & \text{if } y = 1 \\ -(1 - \varepsilon) \log(1 - p) & \text{otherwise} \end{cases} \quad (3-12)$$

其中, 对于二分类任务, 真实值 y 一般设置为 0 和 1, 分别代表背景和前景目标。预测结果 p 的取值范围为 0~1, 表示预测像素是前景目标的概率, 这里的 ε 的具体公式表达如下:

$$\frac{\varepsilon}{1 - \varepsilon} = \frac{n}{m} \quad (3-13)$$

其中, m 和 n 分别代表 y 为 1 和 0 时, 也就是正样本和负样本的数量, 如此可以更好地根据正负样本的数量做出权重调整。

相比于 **BCE** 的权重思想, **Focal** 损失则是在 **CE** 基础之上增加了类别权重 α 和样本难度权重调节因子 $(1 - p_t)^\gamma$ 来优化。在 **BCE** 的基础上增加一个调节因子和一个可聚焦参数 $\gamma \geq 0$, 可以更好地聚焦于稀有缺陷样本的学习。

$$FL(p, y) = \begin{cases} -(1 - p)^\gamma \log(p) & \text{if } y = 1 \\ -p^\gamma \log(1 - p) & \text{if } y = 0 \end{cases} \quad (3-14)$$

其中, 设置 $p_t = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{if } y = 0 \end{cases}$, 则公式简化为:

$$FL(p_t) = -(1 - p_t)^\gamma \log(p_t) \quad (3-15)$$

其中 p_t 反映了与真实值的接近程度, 可通过 $\gamma \geq 0$ 实现调节, 常取值为 2。 p_t 越大则说明距离真实值越接近, 分类效果就越好, 即对分类的置信度越高, 说明分类越容易, 对于分类效果好的样本, 权重调节因子 $(1 - p_t)^\gamma$ 趋近于 0, 而对于分类效果一般的样本, 权重调节因子 $(1 - p_t)^\gamma$ 趋近于 1, 即对分类的置信度较低, 说明分类有难度, 整体上来看,

Focal 损失对于分类效果不准确的样本损失不做改变，而对分类准确的样本的损失进行缩小，如此凸显对于难例的侧重学习，加强了对稀少缺陷像素区域的学习能力。

Focal 损失与 **BCE** 的目的都是希望解决样本不平衡带来的模型训练困难问题，后者从样本的正负分布不均衡角度对损失函数添加权重因子，而前者则是从样本分类的难易程度出发，使得损失聚焦于难分样本。

结合 $p_t = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{if } y = 0 \end{cases}$ 和 $\varepsilon_t = \begin{cases} \varepsilon & \text{if } y = 1 \\ 1 - \varepsilon & \text{if } y = 0 \end{cases}$ ，可以得到对于正负样本和难易样本不均衡的问题的双重解决，也就得到了最终的 **Focal** 损失：

$$FL(\varepsilon_t, p_t) = -\varepsilon_t(1 - p_t)^{\gamma} \log(p_t)$$

(3 - 16)

对于工业真实场景的缺陷检测，**Focal** 损失能够很好地应对正负样本不均衡和工业缺陷过于细小的问题，所以最终在判别网络中使用 **Focal** 损失作为损失函数。

3.4 实验分析

3.4.1 数据集介绍

本文采用的数据集是源于真实工业场景下的数据，具体场景为钢材表面的缺陷检测选取的四种不同视角拍摄的机床区域范围图片，图中钢材随机床传动，会出现边裂和折印等常见缺陷，这些缺陷可能会导致生产环境的破坏，生产成品的质量得不到保证，为此使用语义分割算法实现工业缺陷检测尤为必要。

表 3.1 数据集信息

	Train		Test	
	Good	Anomaly	Good	Anomaly
Total	4287		175	
	2238	2049	26	149
D01	603	522	5	50
D02	605	642	10	45
D03	506	19	5	5
D04	524	866	6	49

数据集包含了训练集图片 4287 张，测试集图片 175 张。其中，训练集中正常样本有

2238 张，异常样本有 2049 张，对于每一类样本都有四种不同的视角图像（D01-D04），其中正常样本没有标注；在测试集中，针对四个不同视角，分别随机选取了一定数量的图片作为每个视角图像的测试数据，最终获得了 175 张测试集图像，具体包含了 26 张正常样本以及 149 张异常样本用于测试。

训练集的异常样本和测试集的所有样本，都配有两种数据标注，分别代表对于缺陷检测的数据标注，包括了折印和边裂的缺陷标注；对于机床定位的标注，主要用对机床范围的像素级标注。



a) D01 视角数据集样张



b) D02 视角数据集样张



c) D03 视角数据集样张



d) D04 视角数据集样张

图 3.5 钢材表面缺陷数据集

由上图 3.5 所示，我们展示了四个视角的正常样本(左)和异常样本(右)的样张，异常样本中包括了边裂和折印两种类型的缺陷，如图 3.5(a)(c)中是边裂缺陷，图 3.5(b)(d)是折印缺陷。

3.4.2 实验细节

A. 评测指标

本章采用的实验指标为平均像素精度 MPA (Mean Pixel Accuracy)，平均交并比 MIoU (Mean Intersection over Union)，以及两类缺陷的 AUC (Area Under Curve) 数值。具体而言：

对于 k 类检测目标的任务，一共需要预测 $k + 1$ 个类（包含一个背景类）， p_{ij} 表示实际属于类 i 而模型预测为类 j 的像素数量。所以， p_{ii} 表示实际模型正确判断的数量，而 p_{ij} 和 p_{ji} 分别可以被理解为假反例和假正例。

平均像素精度 MPA 是对于像素精度 PA (Pixel Accuracy) 的一种提升，其计算每个类的被正确分类的像素数目的比例，实际为真却被判断为假的像素数目将会影响该精度。最后将所有类的结果取平均：

$$MPA = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij}} \quad (3-17)$$

平均交并比 MIoU 作为语义分割的一种标准型的度量，其主要是计算两个像素集合的像素之间的交集和并集的比例，也就是除了交集之外的集合区域越少越好，即为对于每个类，预测为真和实际为真的两类的像素都视为并集的组成部分，交集则仅仅是

预测且实际为真的像素数目，在每个类上计算 IoU 之后取平均值：

$$MIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji}} \quad (3-18)$$

AUC 即为 AUROC (Area Under the Receiver Operating Characteristic curve)，即为 ROC 曲线下面积，AUC 必在 0~1 之间，AUROC 值越大的分类器，正确率越高，表示分类器可以较好的分类正负样本。

B. 基于不同的判别网络的对比

本章采用四种基础的语义分割网络 U-Net、SegNet、PSPNet、DeepLabV3+ 作为判别网络进行对比，并分别将其与我们提出的半监督对比学习架构 ARCD 进行结合，与原四种基础网络进行对比，并采用不同的数据量进行对比，证明架构的低数据依赖性、有效性和优越性。

C. 训练和推理流程

本章的实验选择 (256, 256) 固定分辨率大小的图像作为统一输入，采用不同的数据采样率实现不同数据量进行训练，采样率包括了 100%、80%、50%、30% 等，对于采样的 seed 进行统一设置，保证数据随机选取的一致性。学习率设定为 0.0001，优化器使用 Adam。实验采用 3090 单卡进行训练。

3.4.3 对比实验

本小节，我们做了两组对比实验来验证所提出的半监督对比学习网络架构 ARCD 的有效性。

A. 与基础网络的性能对比

本章采用四种基础的语义分割网络 U-Net、SegNet、PSPNet、DeepLabV3+ 作为对比的判别网络，分别与所设计的 ARCD 半监督的对比学习语义分割网络进行结合，与 U-Net、SegNet、PSPNet、DeepLabV3+ 原基础网络进行对比，使用的是全量的钢材表面缺陷数据集，即数据采样率为 100%。

表 3.2 与基础网络的缺陷检测性能对比

Model	MPA (%)	MIoU (%)	AUC(%)	
			Scratch 折印	flaw 边裂
U-Net	77.56	73.76	96.69	95.84
ARCD (U-Net)	80.05	75.24	98.82	97.81
SegNet	75.34	72.80	96.44	94.94
ARCD (SegNet)	78.01	74.95	97.63	95.31
PSPNet	77.76	73.81	96.98	96.35
ARCD (PSPNet)	80.13	75.56	98.78	97.90
DeepLabV3+	78.53	75.03	97.55	97.34
ARCD (DeepLabV3+)	82.19	76.88	98.95	98.03

表 3.3 与基础网络的机床区域检测性能对比

Model	MPA (%)	MIoU (%)	AUC (%)
U-Net	98.79	98.58	99.00
ARCD (U-Net)	99.13	99.06	99.22
SegNet	98.35	98.11	98.59
ARCD (SegNet)	98.50	98.43	98.88
PSPNet	98.81	98.58	99.11
ARCD (PSPNet)	99.05	99.01	99.12
DeepLabV3+	99.11	98.97	99.33
ARCD (DeepLabV3+)	99.32	99.05	99.41

由上表可知，本章所提出的 ARCD 半监督对比学习架构实现了一致的优越性，对比了几种基础的语义分割模型，在实际的钢材表面缺陷数据集中，有明显的性能提升，在 MPA 指标上分别超越了 U-Net 2.49%，SegNet 2.67%，PSPNet 2.37%，DeepLabV3+ 2.66%，在 MIoU 指标上也分别高出四个基础网络 1.48%，1.15%，1.75%，1.45%，在 AUC 上也高出四个基础网络。

对于上述被广泛应用的常见语义分割模型，所提出的半监督对比学习架构都能在原网络基础之上增强网络对正常样本的充分利用，并在重建网络中学习正常样本的模式信息，更好地对比学习出缺陷的精确分割效果，保证了足够的灵活性和可迁移性。

对于机床区域检测，因为区域面积较大，整体具有较强的规则性，检测难度不大，重建网络可以提供图像的结构信息作为辅助，提升微弱效果，分别在 MPA 指标上超越了 U-Net 0.34%，SegNet 0.15%，PSPNet 0.24%，DeepLabV3+ 0.21%。分别在 MIoU 指标上超越了 U-Net 0.48%，SegNet 0.32%，PSPNet 0.43%，DeepLabV3+ 0.08%。

B.不同数据量对模型性能的影响分析

为了更好地对比我们所提出的半监督对比学习网络架构 ARCD 的优势，考虑真实工业场景之下，数据体量的有限性，尤其是异常样本的数据量有限，数据标注成本也相对较高，所以我们希望在使用更少的数据量和更少数据标注的情况下，依旧能够保持模型的性能。

为此，我们对比了不同数据采样率下的数据量缩减对于模型的影响，这里我们同样对比了四种网络模型 U-Net、SegNet、PSPNet、DeepLabV3+，并且通过设置相应的对比判别网络进行合理的比较。

表 3.4 不同数据量对模型性能的影响分析

Model	采样率	MPA (%)	MIoU (%)	AUC (%)	
				Scratch 折印	flaw 边裂
U-Net	80%	75.42	70.30	96.69	95.84
	50%	72.22	67.77	95.81	94.04
	30%	69.08	66.67	95.68	93.68
ARCD (U-Net)	80%	79.55	74.33	98.54	97.68
	50%	78.04	73.20	98.32	97.55
	30%	77.84	72.80	97.97	97.03
SegNet	80%	75.04	69.38	96.02	94.84
	50%	71.74	67.11	95.51	92.37
	30%	68.87	65.89	94.81	91.07

ARCD (SegNet)	80%	77.57	73.23	97.83	95.63
	50%	76.98	72.35	97.50	95.60
	30%	75.49	70.50	97.32	94.94
PSPNet	80%	75.89	71.72	96.77	96.02
	50%	73.61	68.10	96.06	95.27
	30%	70.98	66.79	95.74	94.10
ARCD (PSPNet)	80%	79.65	75.06	98.69	97.80
	50%	78.88	73.69	98.44	97.65
	30%	78.21	73.11	98.05	97.13
DeepLabV3+	80%	76.01	73.12	97.10	97.01
	50%	74.78	70.86	96.66	96.44
	30%	71.35	67.63	96.19	95.39
ARCD (DeepLabV3+)	80%	81.54	76.00	98.73	97.90
	50%	81.09	75.83	98.46	97.77
	30%	80.67	75.22	98.19	97.27

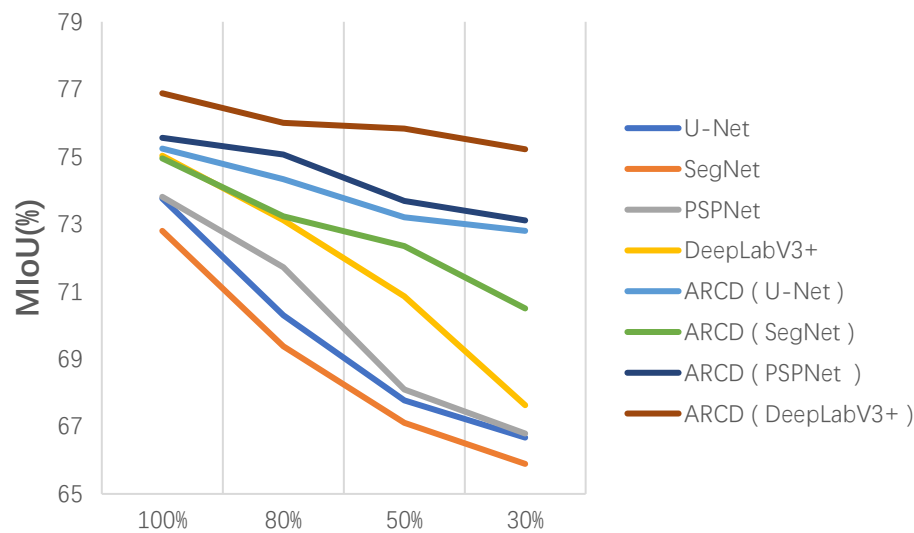


图 3.6 不同数据量的 MIoU 变化趋势

由表 3.4 和图 3.6 可知，在不同数据采样率下，即不同的异常数据使用量下，我们提出的半监督对比学习网络架构 ARCD 依旧保持着出色的检测效果，且随着采样率的降

低,模型表现出更好的稳定性,说明半监督对比学习网络架构 ARCD 对于数据依赖的程度相比与基础语义分割网络更低,网络对数据的需求量也更少,具有很强的工业实用性。

3.4.4 可视化结果分析

A.不同损失函数的重建效果

对于重建网络的分析,我们希望将模拟异常图像重建为与原正常样本图像尽可能相似的重建图像,我们通过对比仅使用 L2 损失函数和 L2 与 SSIM 的组合损失函数的实际重建效果,从而更好地判断差异:

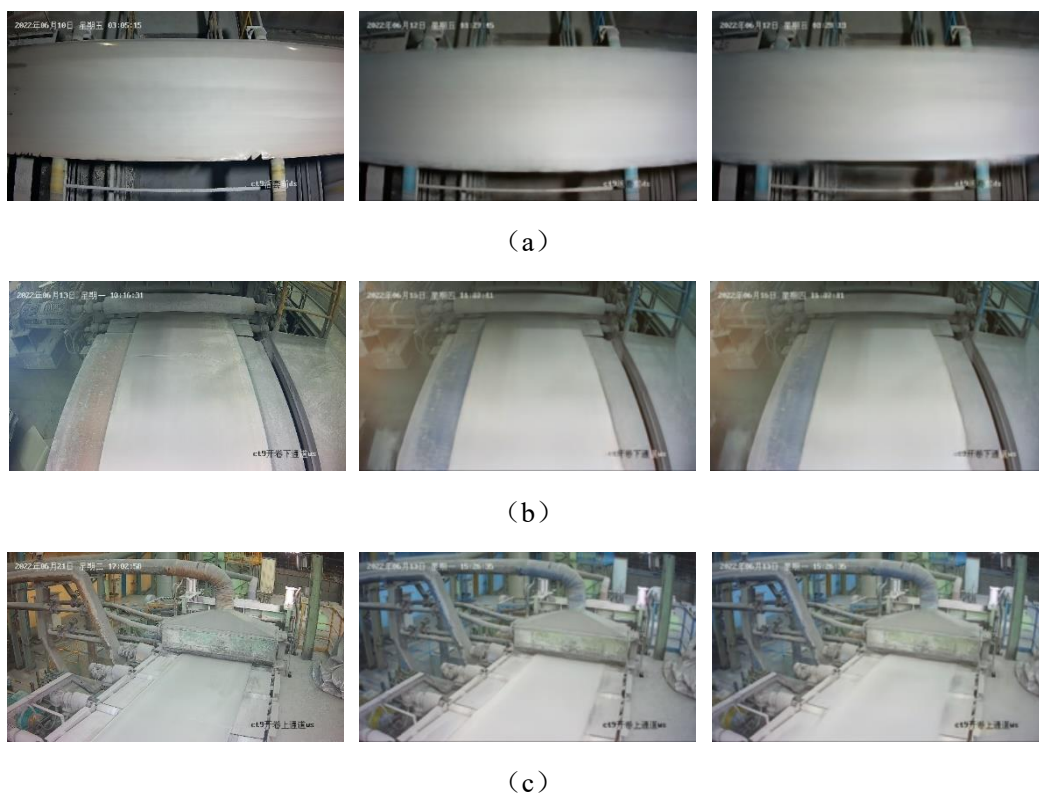


图 3.7 不同损失函数的重建效果对比

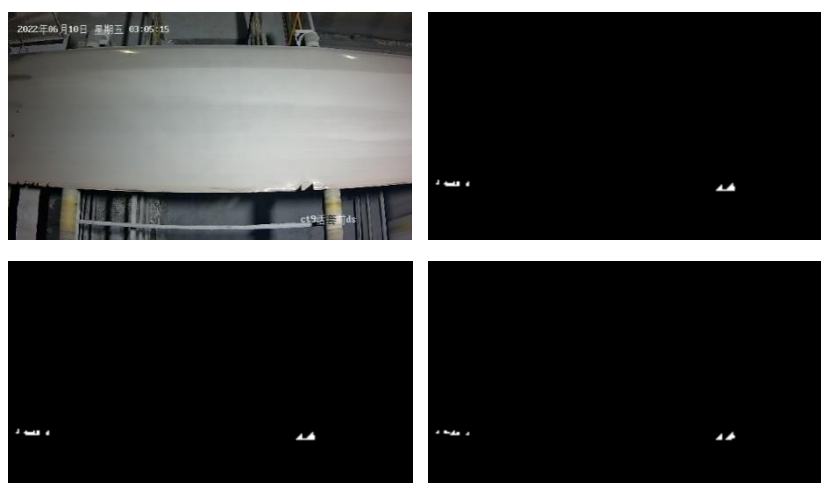
如图 3.7 所示,每组的三个图分别代表原异常缺陷样本(左)、基于 L2 和 SSIM 的组合损失函数的重建效果(中)、基于仅 L2 损失函数的重建效果(右)。可以明显看出,(a)中仅 L2 的效果边缘存在很多模糊的重建区域,(b)中仅 L2 的效果图中存在了边缘扭曲的情况,(c)中仅 L2 的效果图整体的模糊程度过高。相比之下 L2 和 SSIM 的组合损失函数效果表现出明显的优势,这也得益于 SSIM 的结构相似性的度量效果,对于

重建的优化至关重要。

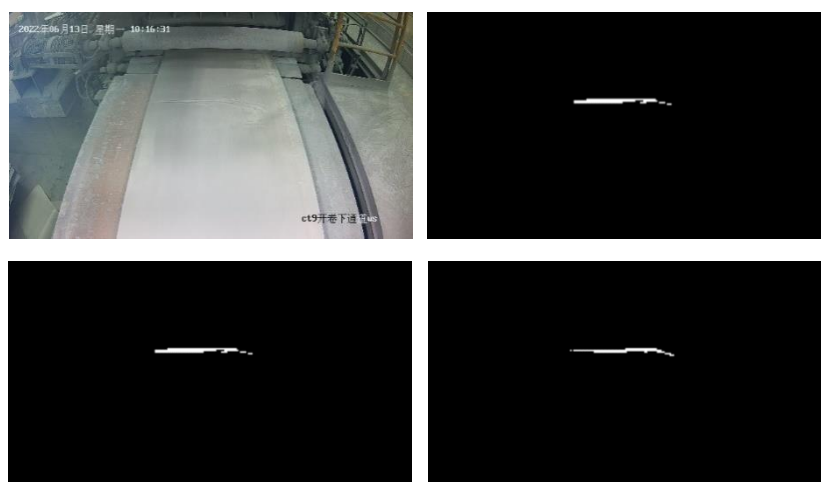
B. 半监督对比学习架构与基础网络的检测效果对比

实验展示对比基于 U-Net 的半监督学习的语义分割网络架构 ARCD 和 U-Net 的基础语义分割网络的最终对比效果展示。

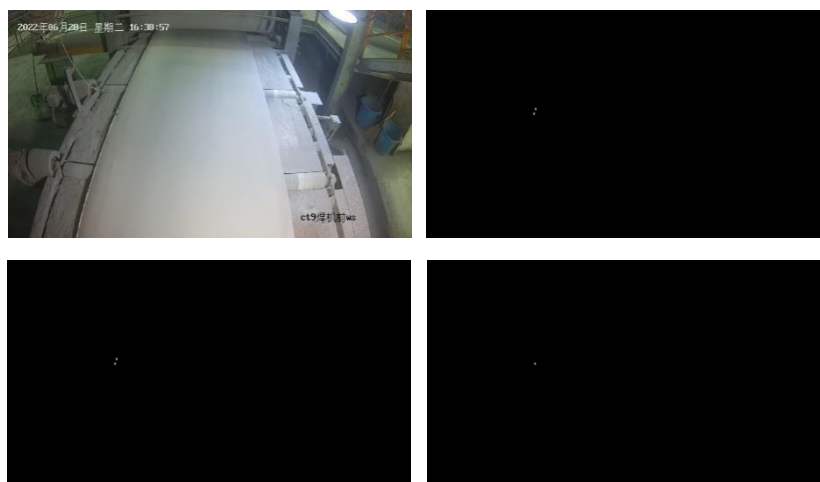
如图 3.8 所示，每组图包含 4 个图，分别代表原图像（左上）、原标签（右上）、基于 U-Net 的半监督对比学习架构的检测效果（左下）、基础 U-Net 网络的检测效果（右下）。



(a)



(b)



(c)



(d)

图 3.8 半监督对比学习架构与基础网络的检测效果对比

可以看出基础的网路在对于机器细小的缺陷，如图 3.8(c)中的两个细小的边裂，U-Net 基础网络只检测出了一个，而使用半监督对比学习架构 ARCD 的方法效果有所提升。整体上半监督对比学习架构 ARCD 的效果更好，证明了架构的有效性。

3.5 本章小结

本章提出了一个“异常重建+对比判别”的半监督对比学习的网络架构 ARCD，其能够更为充分地利用正常样本数据，减少工业场景的数据标注工作量和误差影响。具体而言，设计一种两阶段的半监督学习训练模式：使用了异常模拟模块对于正常样本生成类似真实工业场景的伪异常样本，用于重建网络训练；利用异常样本，通过预训练的

重建网络构建复原重建样本，进行对比，更准确地学习其缺陷的准确定位。使用的数据集是真实工业场景的钢材表面缺陷数据集，源于真实的生产加工产品线。通过对比实验证明了相较于基础的语义分割网络，性能上具有优越性，且在更少的数据量上，性能更稳定，对数据的依赖程度更低。

4 面向重建和判别网络的优化模块设计

4.1 问题阐述

如第3章所介绍的,我们使用半监督的对比学习架构能够更好地利用正常样本数据的价值,并且减少了对数据的依赖,提高了算法的性能。然而,我们仅仅使用的是较为简单的编码器-解码器网络作为异常重建网络,用基础的U-Net网络作为对比判别网络。对于异常重建网络来说,需要重视异常图像复原的网络性能,所以需要对于异常区域的像素有更好的预测能力;对于对比判别网络来说,基础的异常样本和复原重建样本的通道简单叠加的方式不能很好地表征缺陷的精细定位。所以,网络需要对于自身的特征信息进行注意力自主提取学习,加深特征融合能力,提升检测性能。

4.2 本章方法概述

为了解决上述问题,本章讨论对于异常重建网络,设计掩膜特征学习模块(Masked Feature Learning Module, MFLM),利用掩膜空洞卷积来扩大重建网络的感受野,实现对像素周围的信息聚合能力,从而更好地预测当前像素的复原结果。同时,使用Transformer替代基础的全连接整合操作,更好地实现了特征的聚合。对于对比判别网络,设计对比注意力学习模块(Contrastive Attention Learning Module, CALM),使用自注意力机制来获取通道和空间上的自注意力,从而对于异常样本和复原重建样本之间的信息有了更好的注意力判断,优化了缺陷检测性能。

具体而言,本章的主要工作和贡献而下:

(1) 本章针对异常重建网络,设计掩膜特征学习模块,扩大网络感受野,获取像素周围的信息,加强信息重建和特征融合的能力;

(2) 本章针对对比判别网络,设计对比注意力学习模块,使用自注意力机制,在空间和通道维度,获取注意力信息,优化缺陷检测性能。

4.3 算法设计

4.3.1 掩膜特征学习模块

对于异常重建网络,我们起初使用的是基础的编码器-解码器结构,使用的卷积也是最基础的 3×3 卷积,这种结构在感受野、特征融合能力方面仍有很大的提升空间,也缺

乏从局部特征获取全局特征的能力，对于重建性能有一定的影响。

于是我们设计了一种结合掩膜空洞卷积 (Masked Dilated Convolution) 和 Transformer 注意力机制的掩膜特征学习模块 MFLM，对于重建网络扩大感受野，加强对重建信息的复原能力，优化重建效果。对于掩膜特征学习模块，接下来将分别具体介绍其中的掩膜空洞卷积和 Transformer 注意力机制部分：

掩膜空洞卷积的核心思想是基于 SSPCAB^[132] (Self-Supervised Predictive Convolutional Attentive Block) 的方法。该方法使用空洞卷积和通道注意力机制组成了一种自监督学习的自预测模块，对于所需要预测的重建像素进行掩膜处理，使用掩膜空洞卷积对其周围的上下文信息进行融合学习，与此同时也可以选择通过空洞率的超参数和卷积核尺寸来控制上下文信息的覆盖范围，从而获取不同大小的感受野范围，实现从局部模式到全局模式的自监督学习。

具体而言，该模块能够学习重建掩膜覆盖的信息，在我们的重建网络中是通过像素周围的信息的自监督学习当前像素本身的信息，同时这些信息也为后续网络层提供了有用的特征。

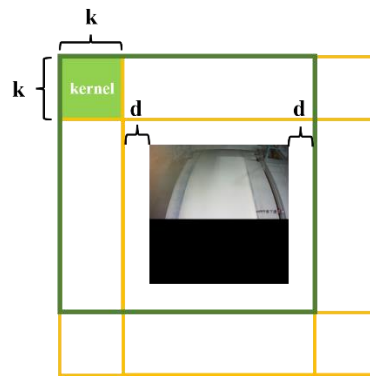


图 4.1 掩膜空洞卷积

如上图所示，掩膜空洞卷积需要确定两个参数，一个是空洞率 d ，一个是使用的卷积核大小 k 。从而确定使用的卷积核尺寸是 $k \times k \times c$ 的，其中 c 是卷积核的通道数。

对于一个 $h \times w \times c$ 的图像，首先需要对其进行 pad 扩展操作，对于图像四周扩展 $k + d$ 的距离长度，从而得到 $(h + 2k + 2d) \times (w + 2k + 2d) \times c$ 的扩展图像。

对于扩展图像，将其分割为 4 个部分，分别为左上角（如图 4.1 深绿色框的区域所示）、左下角、右上角、右下角的四个子图像块，每个子图像块的都是 $(h + k + 2d) \times (w + k + 2d) \times c$ 的尺寸大小。接下来分别对着四个子图像块 $P_i \in \mathbb{R}^{(h+k+2d) \times (w+k+2d) \times c}$,

$\forall i \in \{1, 2, 3, 4\}$ 进行卷积、合并、激活，相当于对于原图像的每个像素的周围四个视角进行自监督学习，预测该像素的实际值，从而实现了自监督学习。具体操作见下图 4.2。

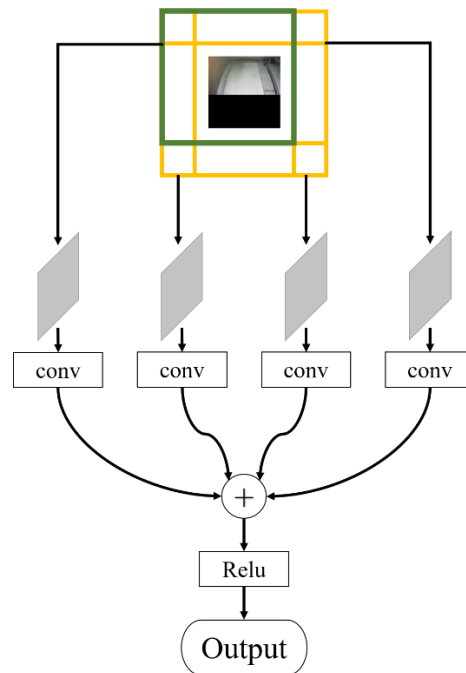


图 4.2 掩膜空洞卷积的特征提取流程

如上图 4.2 所示，最终我们得到了对于一个图像的掩膜空洞卷积提取的特征，实现了周围特征信息的有效聚合，这个特征图可以很好的预测当前像素的信息。在 SSPCAB^[132] 中选择使用 SE Layer 实现后续的特征的融合，在通道上实现自注意力机制的学习，具体为通过对上述特征图进行 **flatten** 扁平化处理，然后在通道维度使用两次全连接层和激活函数直接实现特征信息的自注意力机制学习，最后将通道自注意力信息恢复为原特征图的尺寸，并于原特征图结合，得到最终的特征融合结果。这种方法在一定程度上对于优化了特征的信息，很好地融合了特征，然而直接的扁平化处理对于图像的结构信息有一定的影响，对于通道数过大的特征图，直接执行全连接层计算也会影响整体的训练效率。

自从 Vaswani 等人^[40]引入了自注意力机制的思想并获得成功，计算机视觉领域引发了大量的基于自注意力机制的架构研究，包括 ViT^[133] 等的视觉领域 Transformer 研究。这些模型效果很快就在计算机视觉领域打破了记录，其优异的性能水平得到了研究人员的认可。

为了更好地利用由掩膜空洞卷积得到的不同区域掩膜学习的特征图的激活映射关系，

基于 SSMCTB^[134] (self-supervised masked convolutional transformer block) 的方法思想, 将 SSPCAB 中的 SE Layer 模块替换为基于 Transformer 的注意力模块。使用新的基于 Transformer 的注意力模块能够发挥其更高的建模性能, 捕获更复杂的通道相互关系, 因其学习将注意力权重分配给对应于每个掩膜空洞卷积得到的重建信息, 减少了重建网络的重建误差。

具体操作如下:

对于掩膜空洞卷积得到的特征图输出 $Z \in \mathbb{R}^{h \times w \times c}$, 其表示 c 个掩膜空洞卷积得到的 $h \times w$ 大小的特征图。首先我们将使用空间平均池化来对特征图进行压缩, 得到池化后的特征图 $\hat{Z} \in \mathbb{R}^{h' \times w' \times c}$, 其中 $h' \leq h$ 且 $w' \leq w$ 。接下来对 \hat{Z} 根据通道进行 reshape 操作, 得到扁平化处理的特征 $A \in \mathbb{R}^{c \times n}$, 其中 $n = h' \times w'$, A 表示了不同通道的扁平化特征。之后, 通过线性层对 A 提取出 Tokens $T \in \mathbb{R}^{c \times d_t}$, 其中 d_t 表示每个通道的 Token 的维度。Tokens 将会加入位置编码, 得到最终的 Tokens $T^* \in \mathbb{R}^{c \times d_t}$ 。

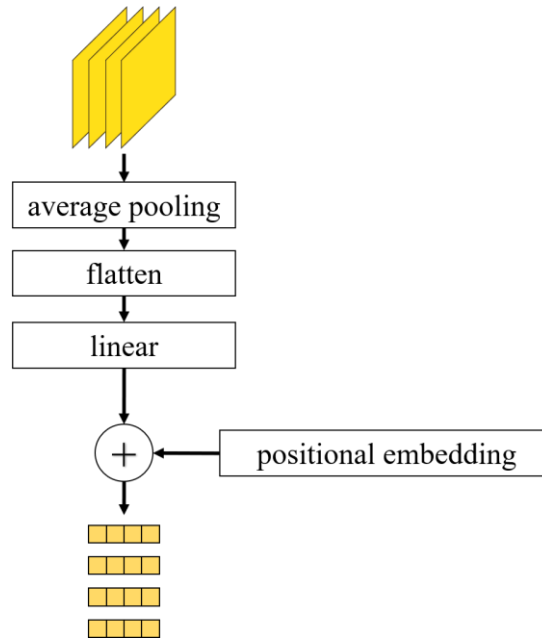


图 4.3 Transformer 的 Tokens 提取流程

对于注意力机制部分, 具体公式如下:

$$P = f(\text{norm}(R)) + R \quad (4-1)$$

$$R = g(\text{norm}(P)) + P \quad (4-2)$$

如上述公式所示,注意力机制模块需要重复上述公式(4-1)(4-2)的步骤实现注意力提取。其中,起始的 R 即为 T^* 的输入, f 表示多头注意力机制模块(multi-head attention layer), g 表示多层感知机(Multi-Layer Perceptron, MLP)。对于输入的 Tokens 经过了 $norm$ 的归一化处理,输入多头注意力机制后与自身相加,然后再次经过归一化处理和多层感知机,并再次与自身相加,如此重复多次。

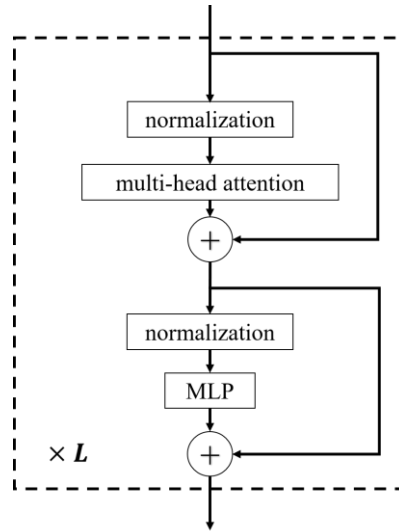


图 4.4 Transformer 的多头注意力机制模块

Transformer 的多头注意力机制模块的目的是获取 c 个通道上的 Tokens 信息的交互学习,从而捕获上下文的信息,并且在通道维度完成多头注意力机制的训练。在多头注意力机制中,对于每个头 $j \in \{1, 2, \dots, H\}$ 都利用 Tokens 生成三种可学习的权重矩阵,分别代表了 Q (Query) 查询量, K (Key) 被查询量, V (Value) 被查询的信息,具体矩阵分别是:

$$W^{Q_j} \in \mathbb{R}^{d_t \times d_q}$$

$$W^{K_j} \in \mathbb{R}^{d_t \times d_k}$$

$$W^{V_j} \in \mathbb{R}^{d_t \times d_v}$$

$$Q_j = \text{norm}(R) \cdot W^{Q_j} \quad (4-3)$$

$$K_j = \text{norm}(R) \cdot W^{K_j} \quad (4-4)$$

$$V_j = \text{norm}(R) \cdot W^{V_j} \quad (4-5)$$

其中, d_q, d_k, d_v 分别代表三种矩阵的信息维度, 且 $d_q = d_k$ 。这三个矩阵分别与 Tokens 进行线性计算得到代表当前头 j 的不同通道的 Q_j, K_j, V_j 三种信息。接下来就是在通道之间, 对着三种信息进行融合, 得到每个通道的注意力结果:

$$Y_j = \text{softmax}\left(\frac{Q_j \cdot K_j'}{\sqrt{d_q}}\right) \cdot V_j \quad (4-6)$$

$$Y = \sum_{j=1}^H Y_j \quad (4-7)$$

$$P = Y + R \quad (4-8)$$

其中, K_j' 是 K_j 的转置, 对于多头注意力机制, 最后将多头的信息进行相加得到最终的注意力信息 Y , 并最后与自身的原信息相加得到最终结果。

综上所述, 我们设计了 MFLM 的掩膜特征学习模块, 具体为使用掩膜空洞卷积的自监督学习模块实现更大感受野的周围信息获取, 得到像素的预测特征信息, 使用基于 Transformer 的注意力机制实现对这些特征信息的高效利用, 对于通道上的自注意力信息进行提取, 最终与原特征信息相加, 得到了掩膜像素的最佳预测, 对于工业缺陷检测中的正常样本的模拟异常区域能够很好地重建复原, 对于重建网络的性能有所提升。

4.3.2 对比注意力学习模块

对于对比判别网络, 一般使用简单的语义分割的编码器-解码器结构可以有基本的效果, 或者使用 U-Net 系列代表的层级间信息共享结构, 实现初步的成效。

然而, 对于所提出的半监督的对比学习网络架构来说, 对比判别网络的输入是两张具有细小异常缺陷差异的样本, 对于这种输入, 其本身就具有一定的信息价值有待提取, 需要加入自注意力机制, 让两张样本之间可以充分地对比学习, 精确地感知异常缺陷的位置定位。

所以, 本节考虑针对对比判别网络, 设计对比注意力学习模块 CALM, 在空间和通道层面对于判别网络引入自注意力机制, 对于 CNN 系列的网络中的特征图的信息在空间和通道维度充分挖掘, 从而优化对比判别网络, 提升工业缺陷检测的分割效果精度。

对于卷积网络的自注意力模块, 常用两种方法, 分别是基于 CBAM^[135] (Convolutional Block Attention Module) 的双维度自注意力方法和基于 CA^[136] (Coordinate Attention) 的

轻量级协调自注意力方法，接下来本节将具体介绍这两种方法，并且基于这两种方法的思想，介绍所设计的对比注意力学习模块 CALM：

（1）基于 CBAM 的自注意力方法

CBAM 是一种简易的卷积网络的自注意力方案，通过对特征图的通道和空间两个维度分别进行注意力的提取，以获得注意力图，最后与原特征图相乘合并，得到自适应的特征细化。

CBAM 采用先通道注意力模块后空间注意力模块的方案：

针对通道注意力模块，采用压缩特征图空间的方法，不同于普遍采用的仅平均池化的方法来获取注意力，CBAM 利用平均池化和最大池化两种方法来有效地计算空间的信息统计，收集了不同通道的注意力特征，提高了网络表示能力。

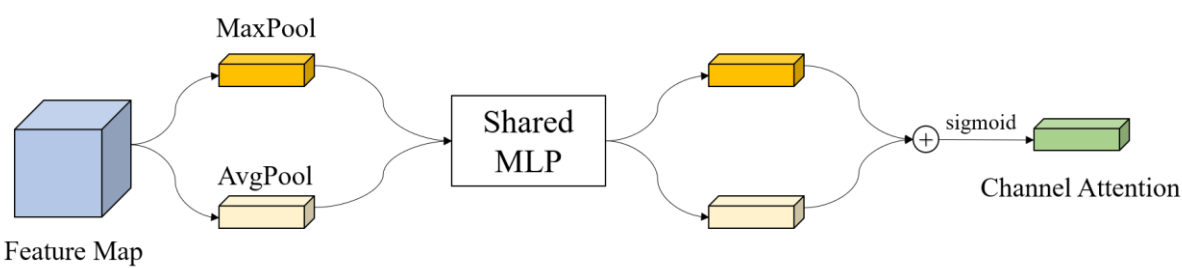


图 4.5 CBAM 通道注意力模块

针对空间注意力模块，同样采用平均池化和最大池化两种方法，针对通道维度进行压缩，有效计算通道上的信息统计，最终得到了不同空间的注意力特征。

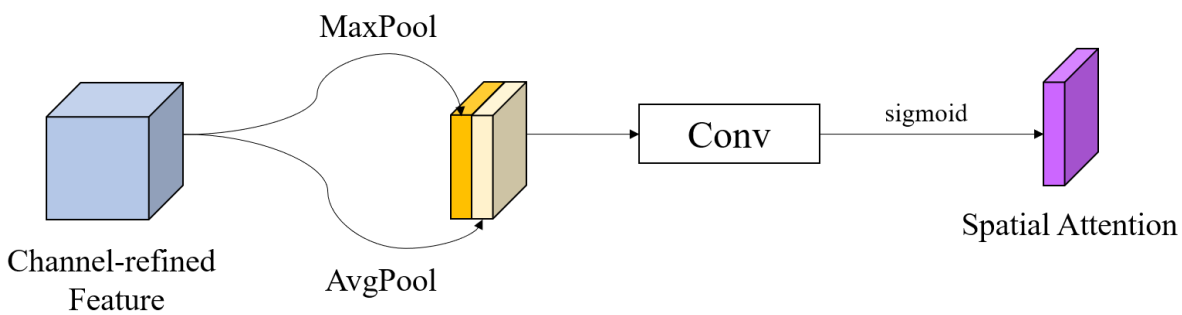


图 4.6 CBAM 空间注意力模块

最后，结合两种注意力模块，对于特征图采用先通道后空间的注意力提取方式，并

且在每次提取注意力后，与原特征图进行相乘回馈。从而得到了最终的注意力机制模块，对于检测性能有所助力。

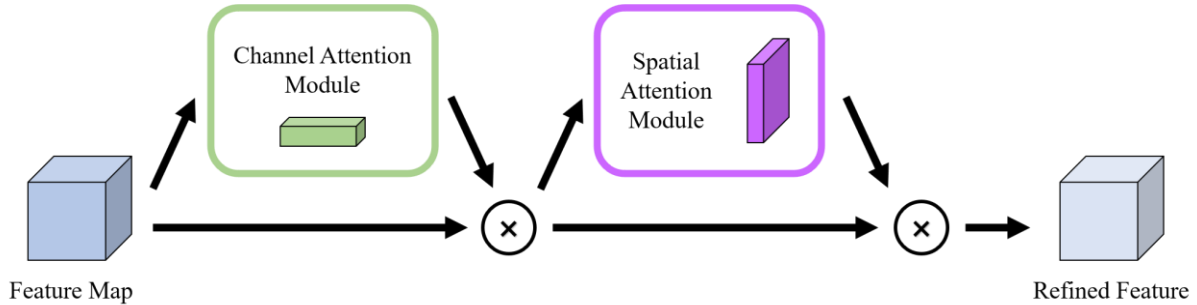


图 4.7 CBAM 流程

(2) 基于 CA 的自注意力方法

基于 CA 的自注意力方法与 CBAM 不同，它在空间的两个方向上的聚合特征，如此可以在一个空间方向上获取较远的依赖关系，同时沿着另一个方向可以保留精确的位置信息，这个对于缺陷异常检测领域有很好的可用价值。最终生成的特征图在方向和位置两个方面的特征信息都会敏感化，增强了对于输入特征图的关注对象的表示。

具体如下图所示，对于特征图 $S \in \mathbb{R}^{h \times w \times c}$ ，通过平均池化对空间上两个方向进行特征提取，分别得到维度为 $h \times 1 \times c$ 和 $1 \times w \times c$ 的两个方向上的注意力信息，级联这两个方向的注意力信息得到维度 $1 \times (h + w) \times c$ 的空间注意力信息，再使用卷积、归一化和激活对特征进行提取，得到维度为 $1 \times (h + w) \times 1$ ，通过 split 操作拆分回两个方向的注意力信息，维度分别为 $h \times 1 \times 1$ 和 $1 \times w \times 1$ ，然后分别通过卷积和激活恢复原通道数，得到最终的空间上双方向的注意力信息，维度分别为 $h \times 1 \times c$ 和 $1 \times w \times c$ 。通过与原特征图进行扩张乘法，将两个方向的注意力信息加入特征图，得到最终结果。

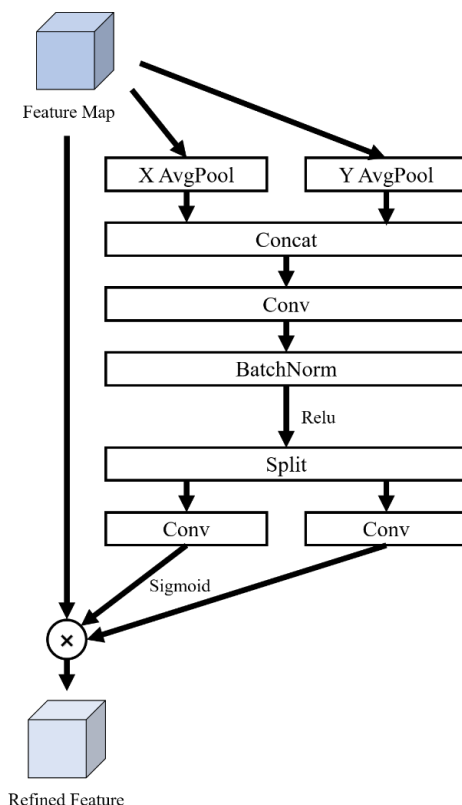


图 4.8 CA 流程

可以看出, CA 方法针对空间层面进行两个方向上的信息捕捉, 利用两个方向信息, 获取精确的位置信息和方向信息, 增强了对全局的空间信息的感知能力, 对于工业缺陷检测的缺陷定位性能有所帮助。

(3) 对比注意力学习模块

针对上述的两种 CNN 自注意力信息提取的方法思想, 我们针对对比判别网络, 设计了对比注意力学习模块 CALM, 对于对比图像的通道和空间上的注意力信息进行提取, 同时针对空间上的信息进行两个维度的细分优化, 最终优化了分割效果。

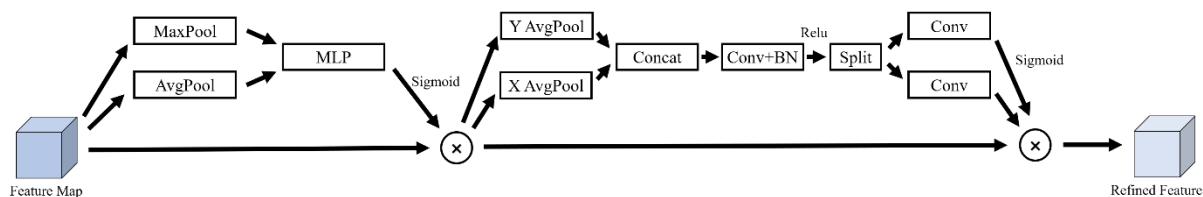
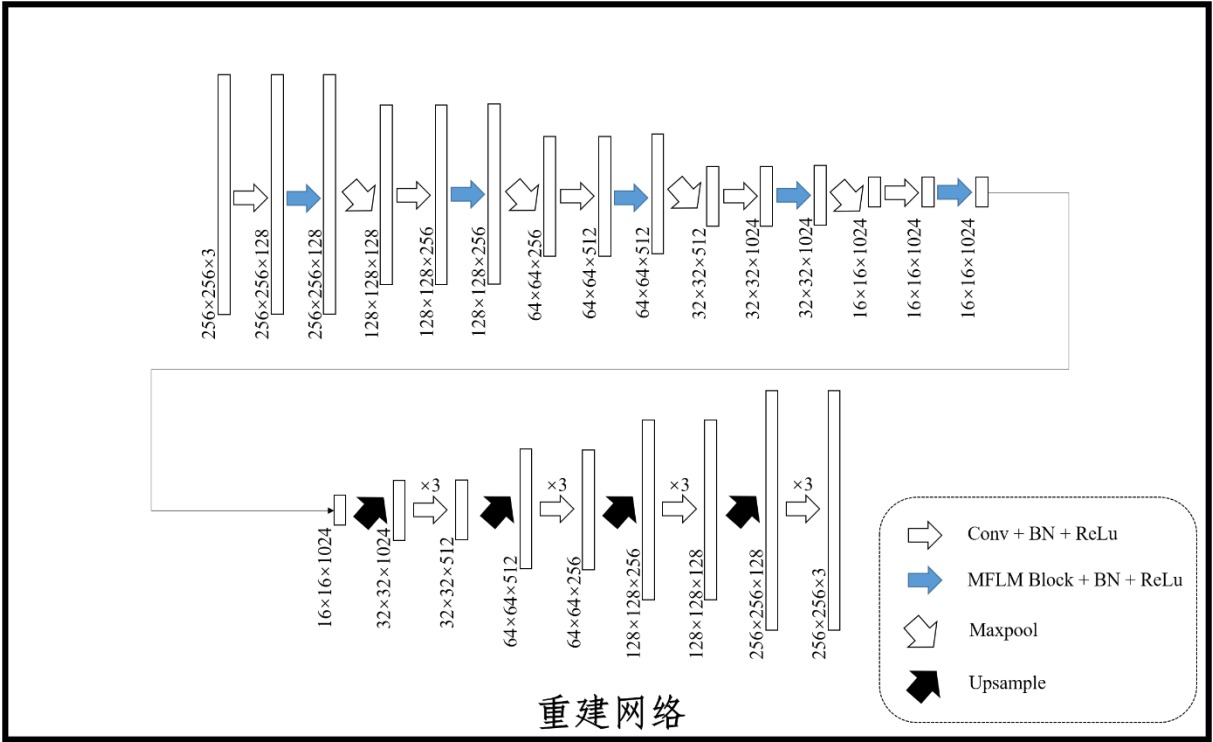


图 4.9 CALM 流程

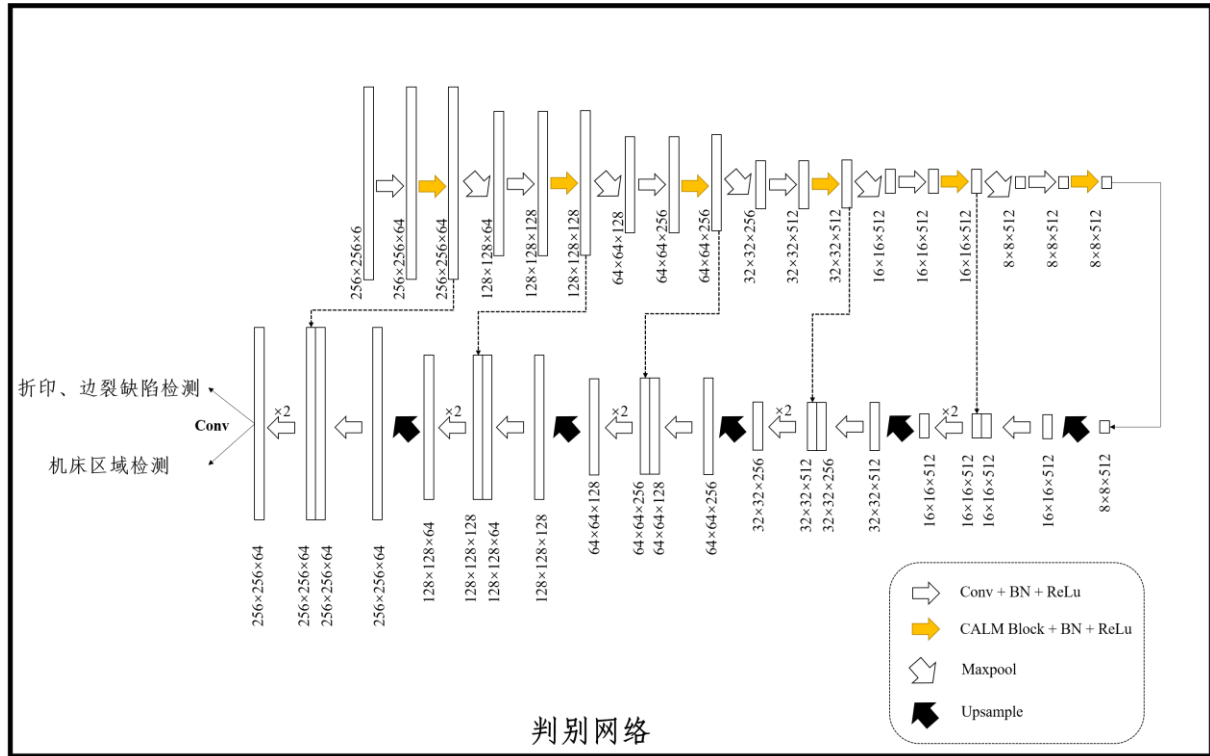
基于对比注意力学习模块 CALM，可以在通道层面进行注意力信息提取，即通过最大池化和平均池化获取通道注意力信息，形成信息对比，然后在空间层面获取定位和方向的注意力信息，通过对 X 轴和 Y 轴的信息分别进行平均池化提取，加强模型的分割性能。

4.3.3 优化的模型结构

最终，结合使用掩膜特征学习模块 MFLM 优化的异常重建网络，以及使用对比注意力学习模块 CALM 优化的对比判别网络，得到了最终的优化后的模型结构，如下图 4.10 所示：



(a) 优化后的重建网络



(b) 优化后的判别网络

图 4.10 优化后的重建网络和判别网络

4.4 实验分析

4.4.1 实验细节

A. 评测指标

本章采用的实验指标为平均像素精度 MPA, 平均交并比 MIoU, 以及两类缺陷的 AUC 数值, 同时也参考 FLOPs (Floating Point Operations) 浮点运算数, 可以理解为计算量, 以及 Params 参数量指标, 用来衡量算法模型的复杂度。

对于重建网络的重建效果, 我们设定了重建效果指标 RA (Reconstruction Accuracy):

$$RA = e^{-10000(L2Loss + \frac{1}{100}(1-SSIM(x,y)))} \quad (4-9)$$

其中, $L2Loss$ 源于公式(3-3), $SSIM(x,y)$ 源于公式(3-10), 考虑到实际 $L2Loss$ 和 $SSIM(x,y)$ 量级都比较小, 我们选择设置系数 10000 来使得 RA 随着损失值的增大而快速下降, 以凸显重建效果的差距。

B. 优化模块对比

通过对比不使用优化模块设计的半监督对比学习架构和使用优化模块设计的半监督对比学习架构，验证所提出的优化模块的有效性。掩膜特征学习模块 MFLM 中，掩膜空洞卷积使用的空洞率为 1，卷积核选择 3×3 大小尺寸，Transformer 使用的参数是 5 头，特征向量统一维度 64 维，考虑减少模型大小，只构建 1 层的 Transformer 训练。

4.4.2 对比实验

本小节，我们做了两组对比实验来验证优化方案的有效性。

A. 重建网络的优化模块对比

本节将会对比使用原有设计重建网络与仅使用掩膜空洞卷积的优化模块、结合掩膜空洞卷积和 Transformer 的 MFLM 掩膜特征学习模块，完成对比实验，证明 MFLM 优化模块的优越性。

表 4.1 重建网络优化模块的性能对比

Model	RA
Base	0.5536
Base + Masked Dilated Convolution	0.6031
Base + MFLM	0.7075

通过表 4.1 可以看出，原有的 Base 基础重建网络在 RA 的重建效果指标上已经拥有一定成效了，而相比之下，仅使用掩膜空洞卷积的优化模块之后，RA 重建效果标准提升了 8.94%，对于重建效果的提升显著。而所提出的 MFLM 优化方案成功推动重建网络效果再度提升，相比于仅使用掩膜空洞卷积的方案，RA 重建效果标准提升了 17.31%，效果显著，相比于原有的重建网络的效果提升幅度更为显著，RA 重建效果标准提升了 27.80%。证明了优化方案的有效性。

表 4.2 重建网络优化模块的计算量与参数量对比

Model	FLOPs (G)	Params (M)
Base	160.38	69.05
Base + Masked Dilated Convolution	283.59	135.58
Base + MFLM	137.85	57.26

由表 4.2 可知，仅使用掩膜空洞卷积的方案改进虽然在性能上得以优化，但是在计算量和参数量上也增加了不少。相比之下，所提出的效果更优的 MFLM 方案，计算量反而减少了 14.05%，参数量减少了 17.07%。在工业场景中实用性更强。

B.判别网络的优化模块对比

本节将会实验对比无自注意力模块的原基础判别网络和加入 CA、CBAM 以及所提出的 CALM 自注意力方案之后的效果。重建网络采用 Base 基础重建网络，该网络是通过 L2 与 SSIM 的组合损失函数训练的，用于参考。判别网络选择基于 U-Net 网络进行优化。

表 4.3 判别网络优化模块的性能对比

Model	MPA (%)	MIoU (%)	AUC (%)	
			Scratch 折印	flaw 边裂
Base	80.05	75.24	98.82	97.81
Base + CBAM	80.65	75.52	98.91	97.84
Base + CA	80.77	75.60	98.96	97.85
Base + CALM	81.13	75.80	98.99	97.95

从表 4.3 可以看出，CALM、CBAM 和 CA 的加入都会对判别网络的效果带来一定的提升，CBAM 相较于 Base 基础判别网络 MPA 提升了 0.6%，MIoU 提升了 0.28%，CA 相较于 Base 基础判别网络 MPA 提升了 0.72%，MIoU 提升了 0.36%，而我们所设计的 CALM 优化方案，相较于 Base 基础判别网络 MPA 提升了 1.08%，MIoU 提升了 0.56%，相比 CA、CBAM 效果更优，证明了 CALM 优化方案的有效性。

表 4.4 判别网络优化模块的计算量与参数量对比

Model	FLOPs (G)	Params (M)
Base	383.90	28.38
Base + CBAM	279.88	20.63
Base + CA	279.91	20.61
Base + CALM	279.80	20.63

从表 4.4 可以看出, CALM、CBAM 和 CA 在计算量和参数量上都相比于原判别网络更少, CBAM 和 CA 相较于原判别网络, 在计算量减少了约 27.09%, 在参数量上 CBAM 减少了 23.31%, CA 减少了 23.38%。而我们所设计的 CALM 的计算量相比 CBAM 和 CA 方案更低, 参数量上和 CA、CBAM 接近, 模型的工业实用性更强。

4.5 本章总结

本章讨论设计了 MFLM 和 CALM 两种优化方案, MFLM 使用掩膜空洞卷积和 Transformer 注意力模块优化了异常重建网络, 扩大了重建网络的感受野, 促进了注意力特征融合的能力, 优化重建性能。CALM 使用自注意力机制模块优化了对比判别网络, 提升了通道和空间上的自注意力信息挖掘能力, 对异常样本和复原重建样本提取注意力信息, 优化了缺陷检测性能。

5 总结与展望

近些年来,人工智能技术的蓬勃发展确实引人注目。人工智能在现实生活的应用落地需要更多地结合实际场景需求做出调整。语义分割作为计算机视觉领域的像素级高精度要求的任务,在实际场景的应用非常广泛,对于真实场景的应用,如智能网联汽车、医学图像分割、工业缺陷检测等等,可以做的还有很多,需要考量的也很多。

本文致力于设计一种更贴合工业实际场景的基于语义分割的工业缺陷检测算法。

(1) 对于工业上大量正常样本数据的情况,我们充分地利用工业上常见易得的正常样本数据,有针对性地设计了自监督的学习模式,基于图像重建复原的网络思路,用异常模拟的方法构建正常样本的伪缺陷,实现正常样本的自监督学习训练,得到了异常样本复原重建的网络。对于异常样本的输入,得到复原重建样本,与异常样本本身一起加入判别网络进行对比学习,加强了异常缺陷位置的定位能力,优化分割效果。如此构建了最终的完整的半监督对比学习的语义分割网络架构,充分利用了正常样本。由于重建网络学习到了正常样本的模式,加强分割性能的同时,也减少了对数据整体体量的依赖,符合工业实际需求,解决了工业的实际问题;

(2) 对于重建网络和判别网络进行优化。针对异常重建网络,因为需要准确预测模拟异常样本的真实像素值,设计了 **MFLM** 的掩膜特征学习模块,使用掩膜空洞卷积和 **Transformer** 注意力机制进行优化,加强了感受野范围,优化了重建性能。针对对比判别网络,需要对复原重建样本和异常样本之间的差异信息进行提取对比学习,设计了 **CALM** 的对比注意力学习模块,更好地在通道和空间维度上进行自注意力信息获取,增强了复原重建样本和异常样本的特征表示,更精确地感知缺陷的定位,优化了分割效果。

尽管本文提出了优化后的半监督对比学习的工业缺陷检测方案,但仍有许多值得思考的地方。本文潜在的未来工作如下:(1) 是否可以直接合并异常重建网络 and 对比判别网络,由原来的两阶段训练转变了端到端的直接学习模式。(2) 是否可以更多地使用 **Transformer** 替代原有卷积操作,将模型进行再次优化,并提高模型效率。(3) 是否可以考虑引入弱监督的思想,进一步优化模型对真实工业数据的敏感性,更贴合真实的工业数据模式。

参考文献

- [1] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014.
- [2] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet Classification with Deep Convolutional Neural Network[C]// Advances in Neural Information Processing Systems: vol. 25. 2012:1097-1105.
- [3] CHENG D, GONG Y, ZHOU S, et al. Person Re-Identification by Multi-Channel Parts-based CNN with Improved Triplet Loss Function[C]//The IEEE Conference on Computer Vision and Pattern Recognition. 2016:1335-1344.
- [4] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: A Unified Embedding for Face Recognition and Clustering[C]//The IEEE Conference on Computer Vision and Pattern Recognition. 2015:815-823.
- [5] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional Networks for Biomedical Image Segmentation[J]. Springer International Publishing, 2015.
- [6] LONG J, SHEHMER E, DARRELL T. Fully Convolutional Networks for Semantic Segmentation[C]//The IEEE Conference on Computer Vision and Pattern Recognition. 2015:3431-3440.
- [7] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6):1137-1149.
- [8] REDMON J, DIVVALA S, GIRSHICK R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]//The IEEE Conference on Computer Vision and Pattern Recognition. 2016.
- [9] 高凯珺,孙韶媛,姚广顺等.基于深度学习的无人车夜视图像语义分割[J].应用光学, 2017, 38(3):8.
- [10] 王中宇,倪显扬,尚振东.利用卷积神经网络的自动驾驶场景语义分割[J].光学精密工程, 2019,27(11): 10.
- [11] 蒋成,张骥,梁晓辉等.机器视觉在互联网工业技术领域的重要应用[J].现代传输,2021(02):42-47.
- [12] 杨俊叶,刘佳,王丽.计算机视觉技术在工业领域中的应用[J].科技创新导报,2020,17(01):108-109.
- [13] SOBEL I. Camera Models and Machine Perception[J]. Dissertation Stanford University, 1970.
- [14] DALAL N, TRIGGS B. Histograms of Oriented Gradients for Human Detection[C]//The IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2005.
- [15] Pratt W K. Quantitative Design and Evaluation of Enhancement/ Thresholding EdgeDetectors in Procession[J]. IEEE transaction on image procession, 1979.67(5)
- [16] MALIK J, BELONGIE S, LEUNG T, et al. Contour and Texture Analysis for Image Segmentation[J]. International Journal of Computer Vision, 2001, 43(1): 7-27.
- [17] 冯希龙. 基于 RGB-D 图像的室内场景语义分割方法[D].南京邮电大学,2015.
- [18] 张晓雪. 基于概率图模型的图像语义分割技术研究[D].厦门大学,2014.
- [19] FULKERSON B, VEDALDI A, SOATTO S. Class Segmentation and Object Localization with Superpixel Neighborhoods[C]//The IEEE International Conference on Computer Vision. 2009: 670-677.
- [20] FARABET C, COUPRIE C, NAJMAN L, et al. Learning Hierarchical Features for Scene

- Labeling[J].IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 35(8): 1915-1929.
- [21] SHIJ, MALIK J. Normalized Cuts and Image Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8): 888-905.
- [22] LAFFERTY J, MCCALLUM A, PEREIRA F C. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data[J], 2001.
- [23] 包广斌,杨旭鹏,康宏.基于改进的高斯混合模型牙齿图像分割研究[J].兰州理工大学学报,2020,46(03):100-104.
- [24] DOC B, BATZOGLOU S. What Is the Expectation Maximization Algorithm?[J]. Nature Biotechnology, 2008, 26(8): 897-899.
- [25] 陈科尹,吴崇友,关卓怀等.基于统计直方图 k-means 聚类的水稻冠层图像分割[J].江苏农业学报,2021,37(06):1425-1435.
- [26] 魏宏超,王永丽,丁晓云等.基于线性谱聚类超像素分割和联合稀疏表示的高光谱图像分类算法[J].山东科技大学学报(自然科学版),2022,41(04):85-97.
- [27] WEISS Y. Segmentation Using Eigenvectors: A Unifying View[C]//The IEEE International Conference on Computer Vision: vol. 2.1999: 975-982.
- [28] STELLA X Y, SHIJ. Multiclass Spectral Clustering[C]//The IEEE International Conference on Computer Vision: vol. 2.2003:313-313.
- [29] SZEGEDY C, LIU W, JIA Y, et al. Going Deeper with Convolutions[J]. IEEE Computer Society,2014.
- [30] Zhou Z, Siddiquee M, Tajbakhsh N, et al. UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation[J]. IEEE Transactions on Medical Imaging, 2020, 39(6):1856-1867.
- [31] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation[J] IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.
- [32] LIN G, MILAN A, SHEN C, et al. RefineNet: Multi-Path Refinement Networks for High-Resolution Semantic Segmentation[C]//The IEEE Conference on Computer Vision and Pattern Recognition.2017: 1925-1934.
- [33] WU T, TANG S, ZHANG R, et al. CGNet: A Light-Weight Context Guided Network for Semantic Segmentation[J]. IEEE Transactions on Image Processing, 2020, 30: 1169-1179.
- [34] ZHAO H, QI X, SHEN X, et al. ICNet for Real-Time Semantic Segmentation on High-Resolution Images[C]//The European Conference on Computer Vision. 2018: 405-420.
- [35] Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4):834-848.
- [36] ZHAO H, SHIJ, QI X, et al. Pyramid Scene Parsing Network[C]//The IEEE Conference on Computer Vision and Pattern Recognition. 2017: 2881-2890.
- [37] YANG M, YU K, CHI Z, et al. DenseASPP for Semantic Segmentation in Street Scenes[C]//The IEEE Conference on Computer Vision and Pattern Recognition. 2018.
- [38] ZHANG H, DANA K, SHI J, et al. Context Encoding for Semantic Segmentation[C]//The IEEE Conference on Computer Vision and Pattern Recognition. 2018: 7151-7160.
- [39] ZHONG Z, LIN Z Q, BIDART R, et al. Squeeze-and-Attention Networks for Semantic Segmentation[C]//The IEEE Conference on Computer Vision and Pattern Recognition. 2020: 13065-13074.

- [40] VASWANI A, SHAZEER N, PARMAR N, et al. Attention Is All You Need[C]//Advances in Neural Information Processing Systems. 2017: 5998-6008.
- [41] WANG X, GIRSHICK R, GUPTA A, et al. Non-Local Neural Networks[C]//The IEEE Conference on Computer Vision and Pattern Recognition. 2018: 7794-7803.
- [42] FUJ, LIUJ, TIAN H, et al. Dual Attention Network for Scene Segmentation[C]//The IEEE Conference on Computer Vision and Pattern Recognition. 2019: 3146-3154.
- [43] HUANG Z, WANG X, HUANG L, et al. CCnet: Criss-Cross Attention for Semantic Segmentation[C]//The IEEE International Conference on Computer Vision. 2019: 603-612.
- [44] LI X, ZHONG Z, WU J, et al. Expectation-Maximization Attention Networks for Semantic Segmentation[C]//The IEEE International Conference on Computer Vision. 2019: 9167-9176.
- [45] ZHANG F, CHEN Y, LI Z, et al. ACFnet: Attentional Class Feature Network for Semantic Segmentation[C]//The IEEE International Conference on Computer Vision. 2019: 6798-6807.
- [46] LIANG X, HU Z, ZHANG H, et al. Symbolic Graph Reasoning Meets Convolutions[J]. Advances in Neural Information Processing Systems, 2018, 31.
- [47] CHEN Y, ROHRBACH M, YAN Z, et al. Graph-based Global Reasoning Networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 433-442.
- [48] ZHANG L, XU D, ARNAB A, et al. Dynamic Graph Message Passing Networks[C]//The IEEE Conference on Computer Vision and Pattern Recognition. 2020: 3726-3735.
- [49] LI Y, GUPTA A. Beyond Grids: Learning Graph Representations for Visual Recognition[J]. Advances in Neural Information Processing Systems, 2018, 31: 9225-9235.
- [50] 任小强, 王佩, 胡波, 等. Prewitt 图像边缘检测算法的优化设计与实现[J]. 中国集成电路, 2016(7):5.
- [51] 李牧, 闫继红, 李戈, 等. 自适应 Canny 算子边缘检测技术[J]. 哈尔滨工程大学学报, 2007, 28(9):6.
- [52] Zuo B, Wang F. Surface cutting defect detection of magnet using Fourier image reconstruction[J]. Computer Engineering and Applications, 2016.
- [53] 魏道昀, 祝美龙, 陶亮. 基于 DCT 的实值离散 Gabor 变换[J]. 电子与信息学报, 2007, 29(11):4.
- [54] 钮心忻, 杨义先. 基于小波变换的数字水印隐藏与检测算法[J]. 计算机学报, 2000, 23(1):7.
- [55] Chetverikov D, Hanbury A. Finding Defects in Texture Using Regularity and Local Orientation[J]. Pattern Recognition, 2002, 35(10):2165-2180.
- [56] Zhen H, Parker J M. Texture Defect Detection Using Support Vector Machines with Adaptive Gabor Wavelet Features[C]// Application of Computer Vision, 2005. WACV/MOTIONS '05 Volume 1. Seventh IEEE Workshops on. IEEE, 2005.
- [57] Peng X, Chen Y, Yu W, et al. An online defects inspection method for float glass fabrication based on machine vision[J]. International Journal of Advanced Manufacturing Technology, 2008, 39(11-12):1180-1189.
- [58] Murino V, Bicego M, Rossi I A. Statistical classification of raw textile defects[C]// International Conference on Pattern Recognition. IEEE, 2004.
- [59] Yong S, Cui L, Qi Z, et al. Automatic Road Crack Detection Using Random Structured Forests[J]. IEEE Transactions on Intelligent Transportation Systems, 2016, 17(12):3434-3445.
- [60] Deng Y S, Luo A C, Dai M J. Building an Automatic Defect Verification System Using Deep Neural Network for PCB Defect Classification[C]// 2018 4th International Conference on Frontiers of Signal Processing (ICFSP). 2018.

- [61] Liang Q , Zhu W , Sun W , et al. In-line Inspection Solution for Codes on Complex Backgrounds for the Plastic Container Industry[J]. Measurement, 2019, 148:106965.
- [62] Deitsch S , Christlein V , Berger S , et al. Automatic classification of defective photovoltaic module cells in electroluminescence images[J]. Solar Energy, 2019, 185(JUN.):455-468.
- [63] Zhang Z , Wen G , Chen S . Weld image deep learning-based on-line defects detection using convolutional neural networks for Al alloy in robotic arc welding[J]. Journal of Manufacturing Processes, 2019, 45.
- [64] Cha Y J , Choi W , Oral Büyüköztürk. Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks[J]. Computer-Aided Civil and Infrastructure Engineering, 2017.
- [65] Xu X , Zheng H , Guo Z , et al. SDD-CNN: Small Data-Driven Convolution Neural Networks for Subtle Roller Defect Inspection[J]. Applied Sciences, 2019, 9(7).
- [66] Cha Y J , Choi W , Suh G , et al. Autonomous Structural Visual Inspection Using Region-Based Deep Learning for Detecting Multiple Damage Types[J]. Computer-Aided Civil and Infrastructure Engineering, 2018, 33(9):731-747.
- [67] 吴晓元, 常海涛, 苟军年. Faster R-CNN 定位后的工业 CT 图像缺陷分割算法研究[J]. 电子技术应用, 2019, 45(1):5.
- [68] Tao X , Zhang D , Wang Z , et al. Detection of Power Line Insulator Defects Using Aerial Images Analyzed With Convolutional Neural Networks[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2020, 50(4):1486-1498.
- [69] He Y , Song K , Meng Q , et al. An End-to-end Steel Surface Defect Detection Approach via Fusing Multiple Hierarchical Features[J]. IEEE Transactions on Instrumentation and Measurement, 2019, PP(99):1-1.
- [70] Li J , Su Z , Geng J , et al. Real-time Detection of Steel Strip Surface Defects Based on Improved YOLO Detection Network - ScienceDirect[J]. IFAC-PapersOnLine, 2018, 51(21):76-81.
- [71] Zhang C , Chih-chen Chang, Jamshidi M . Concrete bridge surface damage detection using a single-stage detector[J]. Computer-Aided Civil and Infrastructure Engineering, 2020, 35(4):389-409.
- [72] Redmon J , Farhadi A . YOLOv3: An Incremental Improvement[J]. arXiv e-prints, 2018.
- [73] Lin T Y , Goyal P , Girshick R , et al. Focal Loss for Dense Object Detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, PP(99):2999-3007.
- [74] Chen S H , Tsai C C . SMD LED chips defect detection using a YOLOv3-dense model[J]. Advanced Engineering Informatics, 2021, 47(1):101255.
- [75] Huang G , Liu Z , Laurens V , et al. Densely Connected Convolutional Networks[J]. IEEE Computer Society, 2016.
- [76] Qiu L T , Wu X , Yu Z Y . A High-Efficiency Fully Convolutional Networks for Pixel-Wise Surface Defect Detection[J]. IEEE Access, 2019:1-1.
- [77] Tabernik D , Ela S , Skvar J , et al. Segmentation-based deep-learning approach for surface-defect detection[J]. Journal of Intelligent Manufacturing, 2020, 31.
- [78] Huang Y , Qiu C , Yuan K . Surface Defect Saliency of Magnetic Tile[J]. Visual Computer, 2018.
- [79] 李月娥, 刘青芳. 小波变换在图像分割中的应用研究(英文)[J]. 山西大学学报:自然科学版, 2009, 32(4):566-571.
- [80] Xie Y , Zhu F , Fu Y . Main-Secondary Network for Defect Segmentation of Textured Surface Images[C]// 2020 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE,

- 2020.
- [81] Valente A C , Wada C , Neves D , et al. Print Defect Mapping with Semantic Segmentation: IEEE, 10.1109/WACV45572.2020.9093470[P]. 2020.
- [82] Zhang G , Cui K , Hung T Y , et al. Defect-GAN: High-Fidelity Defect Synthesis for Automated Defect Inspection[J]. 2021.
- [83] Devries T , Taylor G W . Improved Regularization of Convolutional Neural Networks with Cut-out[J]. 2017.
- [84] Wang Z , Wu F , Hu Z . MSLD: A robust descriptor for line matching[J]. Pattern Recognition, 2009, 42(5):941-953.
- [85] Harris C G , Stephens M J . A combined corner and edge detector[C]// Alvey vision conference. 1988.
- [86] D Nistér, H Stewénus. Linear Time Maximally Stable Extremal Regions[C]// Computer Vision - ECCV 2008, 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part II. Springer, Berlin, Heidelberg, 2008.
- [87] Movafeghi A , Mohammadzadeh N , Yahaghi E , et al. Defect Detection of Industrial Radiography Images of Ammonia Pipes by a Sparse Coding Model[J]. Journal of Nondestructive Evaluation, 2018, 37(1):3.
- [88] Le T , Wong W K , Kwong C K . Fabric defect detection for apparel industry : a nonlocal sparse representation approach[J]. IEEE Access, 2017, PP(99):1-1.
- [89] Yuan H , Cuyt A , Lee W S , et al. Decomposing Textures using Exponential Analysis[C]// International Conference on Acoustics, Speech, and Signal Processing. IEEE, 2021.
- [90] Hocenski, Željko; Vasilic, Hocenski S , et al. Improved Canny Edge Detector in Ceramic Tiles Defect Detection[C]// Conference of the IEEE Industrial Electronics Society. IEEE, 2006.
- [91] Li C , Liu C , Gao G , et al. Robust low-rank decomposition of multi-channel feature matrices for fabric defect detection[J]. Multimedia Tools and Applications, 2018, 78.
- [92] Yang H , Chen Y , Song K , et al. Multiscale Feature-Clustering-Based Fully Convolutional Autoencoder for Fast Accurate Visual Inspection of Texture Surface Defects[J]. Annals of the American Thoracic Society, 2019(3):16.
- [93] Yang J , Shi Y , Qi Z . DFR: Deep Feature Reconstruction for Unsupervised Anomaly Segmentation:, 10.48550/arXiv.2012.07122[P]. 2020.
- [94] Chung H , Park J , Keum J , et al. Unsupervised Anomaly Detection Using Style Distillation[J]. IEEE Access, 2020, 8:221494-221502.
- [95] Dehaene D , Frigo O , Combrexelle S , et al. Iterative energy-based projection on a normal data manifold for anomaly localization[J]. 2020.
- [96] Zhou Q , Mei J , Zhang Q , et al. Semi-supervised fabric defect detection based on image reconstruction and density estimation:[J]. Textile Research Journal, 2021, 91(9-10):962-972.
- [97] Schlegl T , Seebeck P , Waldstein S M , et al. Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery[C]// International Conference on Information Processing in Medical Imaging. Springer, Cham, 2017.
- [98] Schlegl T , Seebeck P , Waldstein S M , et al. f-AnoGAN: Fast Unsupervised Anomaly Detection with Generative Adversarial Networks[J]. Medical Image Analysis, 2019, 54.
- [99] Radford A , Metz L , Chintala S . Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks[J]. Computer science, 2015.
- [100] Baur C , Wiestler B , Albarqouni S , et al. Deep Autoencoding Models for Unsupervised

- Anomaly Segmentation in Brain MR Images[J]. 2018.
- [101] Akcay S , Atapour-Abarghouei A , Breckon T P . GANomaly: Semi-supervised Anomaly Detection via Adversarial Training[J]. 2019.
- [102] Akay S , Atapour-Abarghouei A , Breckon T P . Skip-GANomaly: Skip Connected and Adversarially Trained Encoder-Decoder Anomaly Detection[J]. IEEE, 2019.
- [103] Bergmann P , Lwe S , Fauser M , et al. Improving Unsupervised Defect Segmentation by Applying Structural Similarity to Autoencoders[C]// 14th International Conference on Computer Vision Theory and Applications. 2019.
- [104] Zhou K , Xiao Y , Yang J , et al. Encoding Structure-Texture Relation with P-Net for Anomaly Detection in Retinal Images[C]// 2020.
- [105] Gong D , Liu L , Le V , et al. Memorizing Normality to Detect Anomaly: Memory-Augmented Deep Autoencoder for Unsupervised Anomaly Detection[C]// 2019 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2020.
- [106] Yang Y , Xiang S , Zhang R . Improving unsupervised anomaly localization by applying multi-scale memories to autoencoders:, 10.48550/arXiv.2012.11113[P]. 2020.
- [107] Park H , Noh J , Ham B . Learning Memory-guided Normality for Anomaly Detection[J]. 2020.
- [108] Tan D S , Chen Y C , Chen P C , et al. TrustMAE: A Noise-Resilient Defect Classification Framework using Memory-Augmented Auto-Encoders with Trust Regions[J]. 2020.
- [109] Yang M , Wu P , Liu J , et al. MemSeg: A semi-supervised method for image surface defect detection using differences and commonalities[J]. 2022.
- [110] Sabokrou M , Khalooei M , Fathy M , et al. Adversarially Learned One-Class Classifier for Novelty Detection[J]. IEEE, 2018.
- [111] Fei Y , Huang C , Cao J , et al. Attribute Restoration Framework for Anomaly Detection[J]. IEEE Transactions on Multimedia, 2020, PP(99):1-1.
- [112] Salehi M , Eftekhari A , Sadjadi N , et al. Puzzle-AE: Novelty Detection in Images through Solving Puzzles:, 10.48550/arXiv.2008.12959[P]. 2020.
- [113] Haselmann M , Gruber D P , Tabatabai P . Anomaly Detection Using Deep Learning Based Image Completion[J]. IEEE, 2018.
- [114] Li Z , Li N , Jiang K , et al. Superpixel Masking and Inpainting for Self-Supervised Anomaly Detection.[C]// British Machine Vision Conference. 2020.
- [115] Yang H , Zhou Q , Song K , et al. An Anomaly-Feature-Editing-Based Adversarial Network for Texture Defect Visual Inspection[J]. IEEE Transactions on Industrial Informatics, 2020, PP(99):1-1.
- [116] Mei S , Yang H , Yin Z . An Unsupervised-Learning-Based Approach for Automated Defect Inspection on Textured Surfaces[J]. IEEE Transactions on Instrumentation and Measurement, 2018:1266-1277.
- [117] Vz A , Mk A , Ds A . Reconstruction by inpainting for visual anomaly detection[J]. Pattern Recognition, 2020.
- [118] Yan X , Zhang H , Xu X , et al. Learning Semantic Context from Normal Samples for Unsupervised Anomaly Detection[C]// National Conference on Artificial Intelligence. 2021.
- [119] He D , Xu K , Zhou P , et al. Surface defect classification of steels with a new semi-supervised learning method[J]. Optics and Lasers in Engineering, 2019, 117(JUN.):40-48.
- [120] Odena A . Semi-Supervised Learning with Generative Adversarial Networks: arXiv,

- 10.48550/arXiv.1606.01583[P]. 2016.
- [121] Gao Y , Gao L , Li X , et al. A semi-supervised convolutional neural network-based method for steel surface defect recognition[J]. Robotics and Computer Integrated Manufacturing, 2020, 61(Feb.):101825.1-101825.8.
- [122] Yu He, Song K , Dong H , et al. Semi-supervised defect classification of steel surface based on multi-training and generative adversarial network[J]. Optics and Lasers in Engineering, 2019.
- [123] Chu W H , Kitani K M . Neural Batch Sampling with Reinforcement Learning for Semi-supervised Anomaly Detection[J]. Springer, Cham, 2020.
- [124] Lin D , Li Y , Prasad S , et al. CAM-UNET: Class Activation MAP Guided UNET with Feedback Refinement for Defect Segmentation[C]// 2020 IEEE International Conference on Image Processing (ICIP). IEEE, 2020.
- [125] Kimura D , Chaudhury S , Narita M , et al. Adversarial Discriminative Attention for Robust Anomaly Detection[C]// Workshop on Applications of Computer Vision. IEEE, 2020.
- [126] Ken Perlin. An image synthesizer. ACM Siggraph Computer Graphics, 19(3):287–296, 1985
- [127] Cubuk E D , Zoph B , Shlens J , et al. RandAugment: Practical automated data augmentation with a reduced search space[J]. 2019.
- [128] 李银辉, 吕晓琪, 于荷峰. 基于 L1 和 L2 混合范式的序列图像超分辨率重建[J]. 计算机应用, 2015, 35(3):4.
- [129] 马敏, 郭琪, 闫超奇,等. 基于 l2, p-范数的 ECT 图像重建算法[J]. 计量学报, 2017, 38(5):5.
- [130] Zhou W , Bovik A C , Sheikh H R , et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE Trans Image Process, 2004, 13(4).
- [131] 左龙, 张鹏, 荆树旭,等. 用于图像超分辨率重建的双通道残差网络[J]. 西安交通大学学报, 2022, 56(1):158-164.
- [132] Ristea N C , Madan N , Ionescu R T , et al. Self-Supervised Predictive Convolutional Attentive Block for Anomaly Detection[J]. 2021.
- [133] A. Dosovitskiy, L. Beyer, A. Kolesnikov, et al., “An image is worth 16x16 words: Transformers for image recognition at scale,” in Proceedings of ICLR, 2021.
- [134] Neelu M, Nicolae-Catalin R, Radu T I , et al. Self-Supervised Masked Convolutional Transformer Block for Anomaly Detection[J]. arXiv.2209.12148
- [135] Woo S , Park J , Lee J Y , et al. CBAM: Convolutional Block Attention Module[C]// European Conference on Computer Vision. Springer, Cham, 2018.
- [136] Hou Q , Zhou D , Feng J . Coordinate Attention for Efficient Mobile Network Design[J]. 2021.

附录

A 作者简历

发明专利：

- 学生第二作者（导师一作），已受理，一种基于半监督对比学习的钢材表面异常缺陷检测方法，202211687353.4

在校荣誉：

- 担任党支部书记，负责管理和组织党支部相关事宜。
- 在校荣获三好研究生、优秀研究生、优秀研究生干部、优秀团员等荣誉称号。
- 在读期间，参加 5 项科研项目和 2 场实习。