分类号：TP391　　学校代码：10363
密　级：公开　　学　　号：2210110101

安徽工程大学
Anhui Polytechnic University

硕士学位论文

题目　基于机器视觉的金属表面缺陷检测算法研究

论文作者　赵亚玲
指导教师　王海(教授)
学科（专业）　机械工程
研究方向　机器视觉

论文提交日期：2024 年 5 月 31 日

分类号：TP391　　学校代码：10363
密 级：公开　　学　号：2210110101

基于机器视觉的金属表面缺陷检测算法研究

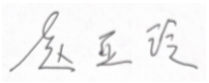
RESEARCH ON DEFECT DETECTION ALGORITHM OF METAL SURFACE BASED ON MACHINE VISION

学生姓名：　　赵亚玲
指导教师：　　王海（教授）
专　　业：　　机械工程
研究方向：　　机器视觉

论文答辩日期：2024 年 5 月 20 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： 

日期：2024 年 5 月 20 日

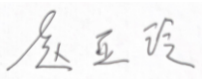
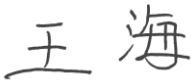
安徽工程大学学位论文版权使用授权书

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：  　　指导教师签名： 

日期：2024 年 5 月 20 日　　日期：2024 年 5 月 20 日

基于机器视觉的金属表面缺陷检测算法研究

摘 要

在当今的工业生产中，金属表面缺陷的自动检测和分类是提高产品质量和生产效率的关键。由于金属表面缺陷的多样性和微妙性，以及在实际应用中遇到的复杂背景和光照条件，传统的检测技术常常无法达到既高精度又高效率的要求。因此，利用先进的图像处理技术和机器学习方法成为了研究和工业界的重要课题，用以进行金属表面缺陷的自动检测和分类。

基于此，本研究致力于通过先进的图像处理与机器学习技术，提高金属表面缺陷检测的性能与准确率。研究主要围绕以下四个方面展开：数据集构建与优化、金属表面缺陷分割技术改进、金属表面缺陷检测的高效解决方案开发，以及半监督学习网络的应用创新。主要研究内容如下：

(1)金属表面缺陷检测数据集的构建和优化研究

本研究通过应用对数变换、中值滤波、直方图均衡化等图像预处理技术，成功构建了针对金属表面缺陷检测的分割与分类数据集。这些数据集覆盖多种缺陷类型，并通过精确的手动标注，为自动化缺陷识别与分类提供了丰富资源。

(2)金属表面缺陷分割技术的改进研究

通过引入改进型 DeepLabV3+模型(Improved-DeepLabV3+)，集成了挤压激励块(Squeeze and Excitation Networks, SENets)、三级注意力单元(tri-level attention unit, TAUs)和密集空洞空间金字塔池化(Dense Atrous

Spatial Pyramid Pooling, DASPP)等先进技术，显著提升了金属表面缺陷分割的效率与精确性。

(3)金属表面缺陷检测的高效解决方案研究

针对检测精度与效率的平衡难题，本研究提出了信息增强 YOLOv5 网络模型(Information Enhancement YOLOv5 Network, IE-YOLOv5)。该模型采用轻量级卷积 GSConv、联邦学习算法(FedFusion)优化的特征提取结构(FedFusion Slim Neck, FF-Slim-Neck)、及无参数注意力机制(Parameter-free Spatial Attention mechanism, PSA)，实现了金属表面缺陷的快速精准检测，显著提升了处理速度与效率。

(4)半监督学习网络的创新研究

为了解决实际应用中缺陷人工标定成本太大的问题，提出半监督检测网络。通过引入伪标签分配器(Pseudo Label Assigner, PLA)及基于伪标签的潜在特征提取方法，本研究在半监督学习网络中实施了一种高效的学生-教师方法。利用大量未标记数据，并结合半监督框架与均值教师(Mean Teacher, MT)模型，显著提升了模型的数据利用率与识别性能。

关键词：金属表面，缺陷检测，IE-YOLOv5，Improved-DeepLabV3+，PLA

RESEARCH ON DEFECT DETECTION ALGORITHM OF METAL SURFACE BASED ON MACHINE VISION

ABSTRACT

In contemporary industrial production, the automatic detection and classification of metal surface defects play a pivotal role in enhancing product quality and productivity. The intricacies and subtleties of metal surface defects, coupled with challenging background and lighting conditions in practical applications, often render traditional detection methods inadequate in meeting the demands for high accuracy and efficiency. Consequently, research and industry have turned their attention towards leveraging advanced image processing techniques and machine learning methods to address this critical issue.

This research is dedicated to elevating the performance and precision of metal surface defect detection through the integration of advanced image processing and machine learning techniques. The study focuses on four key areas: dataset construction and optimization, enhancement of metal surface defect segmentation techniques, development of efficient solutions for defect detection, and the innovative application of semi-supervised learning networks. The primary research contributions are outlined as follows:

(1) Dataset Construction and Optimization

A comprehensive investigation into the establishment and optimization of

a dataset for detecting defects on metal surfaces is conducted. Utilizing image preprocessing techniques, such as logarithmic transformation, median filtering, and histogram equalization, segmentation and classification datasets are constructed. These datasets cover diverse defect types and serve as a valuable resource for automated defect recognition and classification through meticulous manual labeling.

(2) Improvement of Metal Surface Defect Segmentation Technique

This study introduces the Improved-DeepLabV3+ model, incorporating advanced techniques like tri-level attention units (TAUs), Squeeze Excitation Blocks (SENets), and Dense Avoidance Space Pyramid Pooling (DASPP). This enhancement significantly improves the efficiency and accuracy of defect segmentation on metal surfaces.

(3) Efficient Solutions for Detecting Defects on Metal Surfaces

Addressing the challenge of balancing detection accuracy and efficiency, this study proposes the IE-YOLOv5 model. The model integrates lightweight convolutional GSConv, the feature extraction structure FF-Slim-Neck optimized by FedFusion, a federated learning algorithm, and the parameter-free attention mechanism PSA. This approach achieves fast and accurate defect detection, substantially improving processing speed and efficiency.

(4) Innovative Study of Semi-Supervised Learning Networks

To mitigate the high cost of manual calibration in practical applications, this study introduces a semi-supervised detection network. By incorporating a pseudo-labelling distributor and a pseudo-labelling-based potential feature extraction method, the research implements an efficient student-teacher approach within a semi-supervised learning network. This approach leverages large amounts of unlabeled data, enhancing model data utilization and recognition performance through the integration of a based semi-supervised framework with the Mean Teacher (MT) model.

KEY WORDS: metal surfaces, defect detection, IE-YOLOv5, Improved-DeepLabV3+, PLA

目录

插图清单

插表清单

第 1 章 绪论

1.1 研究背景及意义

在现代工业生产领域，金属作为一种基础且关键的原料[1]，承担着不可替代的作用，其应用范围横跨多个行业，包括但不限于汽车生产、航空航天以及机械加工等。金属材料的质量和性能直接影响着机械零部件产品的质量和功能。然而，在金属制品生产过程中，金属表面经常遭遇多种缺陷形式，包括裂纹、凹痕和气泡等，这类缺陷不只损害了产品的外表审美，也可能引起金属制品的性能降低及使用年限减少。

传统的人工检测方法虽然能够辨识部分金属表面缺陷，但存在着效率低、准确性不高、易受主观因素影响等诸多问题[2]。因此，迫切需要一种高效、准确的自动化检测方法来满足日益复杂的生产需求。基于机器视觉的金属表面缺陷检测算法因其在高速、高精度图像处理方面的优势而备受关注[3]。

首先，采用机器视觉的检测系统通过高分辨率和高速度的图像捕获及处理，显著提升了金属表面检测的灵敏度和准确率。基于机器视觉的检测系统利用图像传感器获取金属表面的图像数据，经过适当的图像处理和识别方法，实现对表面缺陷的自动化检测，并可同时进行定量分析和评估。与传统的手工检测技术相比，采用机器视觉的自动化检测系统在效率和准确度方面都展现出显著优势，能够迅速识别并纠正金属表面的缺陷问题，进而提升金属制品的质量与生产效能。

其次，这些系统具备了强大的数据处理和分析能力，能够实时监测生产过程中的异常情况，并提供及时的反馈和处理措施。通过对大量数据的分析和比对，系统可以识别出不同类型的表面缺陷，并对其进行有效分类和处理。这助于生产企业能够及时识别并解决生产过程中潜在的问题，来增强生产线的稳定性与可靠性。

此外，采用机器视觉的自动化金属表面检测系统能显著增加生产效率，降低人力成本，为企业节省大量的时间和人力资源。通过引入智能化的检测系统，企业可以更加高效地进行生产，提高产品的生产率和质量，进一步提升企业的竞争力和市场份额。

总的来说，研究基于机器视觉的金属表面缺陷检测算法，具有显著的实践意义和学术价值。它不仅能够确保机械零部件的质量和稳定性，提高产品的竞争力，还能够推动工业生产的智能化和自动化发展，为工业制造的现代化转型做出重要贡献。在未来，随着机器视觉技术的不断发展和完善，在工业生产领域，基于机器视觉的金属表面缺陷检测算法的作用将日益增强，为实现智能制造和工业 4.0 的目标提供有力支持。

1.2 国内外研究发展现状

1.2.1 目标识别算法

目标识别技术[4-7]，作为计算机视觉领域的一个核心研究方向，旨在让计算机通过分析图像或视频来识别出特定目标。这项技术在多个领域中发挥着至关重要的作用，包括但不限于安防监控[8]、自动驾驶[9, 10]、工业自动化[11, 12]以及医疗图像分析[13]。随着图像处理、机器学习和深度学习等学科的交叉融合，尤其是深度学习技术的飞速进步，目标识别的准确性和处理速度都得到了显著提高。

当前深度学习在目标检测领域的识别算法即检测器可分为三种主要类型：两阶段目标检测器、单阶段目标检测器以及锚点自由的检测器。

(1) 两阶段目标检测器

在双阶段目标检测技术领域，Ross 等人[14]提出的 R-CNN 研究标志着深度学习在目标检测上的首次应用。R-CNN 通过区域提案算法识别图像内可能的目标候选区，然后对这些区域利用卷积神经网络(CNN)提取特征，并借助支持向量机(SVM)完成目标分类。紧接着，为了克服 R-CNN 在速度和效率上的限制，Ross 等人[15]进一步提出了 Fast R-CNN，这一改良版本通过对整幅图像仅执行一次 CNN 运算，并将得到的特征图映射至各个提案区域，显著加快了训练和测试流程。基于 Faster R-CNN 的进展，He[14]引入了 Mask R-CNN，新增了一个用于生成目标分割掩码的分支，实现了高效目标检测和精准像素级目标分割的同步实现，极大地扩展了目标检测技术的应用领域。此外，Cai 等人[4]的 Cascade R-CNN 通过构建一个多级阶段的检测框架，有效提高了检测过程中对小目标及难以识别目标的准确率，通过逐级细化检测框来不断提升检测质量。Guo 等人[17]通过语义增强模块(SEM)和语义注入模块(SIM)

来改善用于对象检测的特征金字塔网络(FPN)，以提高不同尺度特征融合的效果，减少语义差距，并提升检测性能。

(2) 单阶段目标检测器

鉴于双阶段目标检测算法虽提供较高的检测准确率，但往往以减慢检测速度为代价，J Redmon 等人[18]引入了具有里程碑意义的 YOLO (You Only Look Once)模型。YOLO 模型创新地将目标检测转变为单步回归问题，实现从图像像素直接映射到目标边界框坐标及其类别概率的转换，代表了基于回归方法的实时目标检测算法的兴起，有效地解决了传统两阶段目标检测算法检测速度慢的问题。继 YOLO 之后，Liu 等人[19]推出的 SSD 算法通过引入多尺度特征图和默认框机制，实现了在多分辨率特征图上直接预测目标位置和类别，既提升了检测的精度，又保证了处理速度。进一步，Lin 等人[5]开发的 RetinaNet 通过采用焦点损失(Focal Loss)机制，有效应对了在检测众多背景类别时出现的类别不平衡问题，显著增强了网络在识别小尺寸目标上的能力。最后，Tan[20]提出的 EfficientDet 网络，通过一种基于复合系数的可伸缩策略，实现了网络深度、宽度和分辨率的均衡优化，不仅保持了高效率，还大幅提升了检测精度。

(3) 锚点自由检测器

锚点自由(Anchor-free)检测方法代表了目标检测技术的一项新进展，区别于传统的基于锚点的策略，这一方法直接预测图像中目标的关键点或中心点以实现目标的精确定位和识别。通过减少对预设参数的依赖，该方法简化了模型的构建和训练流程，为目标检测提供了一种更为直观和高效的途径。

在这一新兴领域中，Law 等人[21]开发的 CornerNet 标志性地采用了一种创新策略，通过直接检测物体角点的成对关键点，实现了对物体的识别和定位，摆脱了对锚点的依赖。紧随其后，Zhou 等人[22]提出的 CenterNet 进一步简化了目标检测流程，将其转化为关键点检测和回归任务，有效提升了检测效率和精度。Yang 等人[23]的 RepPoints 方法采用点集来表示物体，而非传统的锚框，通过自适应调整点集以精确刻画物体形状，提高了检测准确性。Li 等人[24]提出了一个适用于航空图像检测的基于 FCOS 的简单锚点自由旋转目标检测器，该模型侧重于训练阶段的标签分配策略，并采用椭圆中心采样方法定义旋转边界框(OBB)。Wang 等人[25]则提出了基于高斯中

心度的锚点自由定向对象检测方法 AOGC，该方法以 FCOS 为基线，增加了针对定向对象的检测分支。

1.2.2 工业缺陷检测算法

传统的缺陷诊断主要依赖于人工目视检测[26]，这一方法虽广泛应用，但其检测精度和效率通常较低。随技术的持续发展，机器视觉和深度学习基础上的检测技术已经变成了缺陷检测领域内的两个主要方法。机器视觉检测技术[27]采用了机械设备来模拟人眼进行测量和判断，与传统的人工视觉检测相比，它显著降低了人力和时间成本，因此被广泛采用于多个领域。例如，Suvdaa 等人[28]提出了一种利用尺度不变特征变换(SIFT)与支持向量机(SVM)相结合的表面缺陷识别技术，该技术通过 SIFT 来识别缺陷并从目标中提取特征，随后使用 SVM 对这些特征进行有效分类。进一步，Xiao-Cong[29]开发了一种新型的缺陷检测模型，该模型融合了支持向量机和量子粒子群优化(SVM-QPSO)技术，以更准确地识别表面缺陷。此外，为了提取更加鲁棒的特征，Gyimah 等人[30]采用了一种结合小波阈值滤波和完全局部二值模式(CLBP)的非局部(NL)技术，之后将提取的特征用于表面缺陷的分类检测。这些研究均表明，通过先进的特征提取方法和机器学习技术的结合，可以显著提高缺陷检测的精度和效率。

然而，大多数缺陷的分布具有不规则性且纹理特征不显著，这使得仅依赖传统特征提取算法难以准确地识别和提取这些缺陷特征，从而导致基于这些算法的缺陷检测在应用上遇到诸多挑战，特别是在检测精度和效率方面的表现较差。因此，提升对缺陷图像特征提取的能力显得尤为关键。针对这一挑战，深度学习技术逐渐成为解决方案之一。不同于传统方法，基于深度学习的缺陷检测技术能够直接将缺陷图像作为输入，自动学习并提取图像中的关键特征，极大地简化了特征提取过程缺陷检测[31]。相较于传统技术，深度学习不仅大幅提升了检测的准确度，还实现了更快的识别速度和更好的适应性，有效地克服了传统方法在处理复杂缺陷时的局限性。

在实际生产环境中，基于深度学习的方法已经成为检测部件缺陷的重要工具。早期采用的检测算法主要包括 SSD[32]和 YOLO 系列网络[18, 33-35]。以 SSD 算法为基础，Lv 等人[36]开发了一种高效而准确的图像缺陷检测模型，通过对骨干网络 ResNet 的改进增强了检测能力。Hatab 等研究者[37]对 YOLOv3 算法进行了调整，改

变了其超参数设置，如大小和输入图像尺寸，以适应缺陷检测任务。Li 等人[38]通过将注意力机制模块整合到 YOLOv4 算法的骨干网络中，并对网络结构进行调整，以增强缺陷识别的精度。虽然与两阶段检测算法相比，这些方法在速度上取得了显著的提高，但在提升检测精度方面的成果并不突出。

为了同时保持快速检测的优势并提高检测精度，YOLOv5 算法被提出并受到了广泛研究。Wang 等人[39]在 YOLOv5 框架中集成了卷积注意模块(CBAM)，有效地通过权重分配来抑制复杂背景下的干扰，强化了对目标特征的识别。Le 等研究者[40]通过引入坐标关注机制和利用 BiFPN 结构来融合多尺度特征，有效降低了小目标的漏检和误检。Wang 等人[41]提出了一种基于 YOLOv5 的带钢表面缺陷检测算法，强调了多尺度特征融合模块在提高检测准确性方面的作用，并且通过实验验证了模型在 NEU-CLS 数据集上的优异泛化性能。Chen 等人[42]开发的基于 YOLOv5 的模型采用了新型聚类方法生成适用于 PCB 缺陷检测的锚框，并引入 Swin Transformer 作为特征提取网络，显著提高了检测的准确性和效率。此外，Li 等人[43]提出了一种新颖的两阶段目标缺陷检测方法，该方法通过两个不同的模型分别完成定位和分类任务，进一步提升了检测的准确率。Liu 等[44]的研究则在 YOLOv5 基础上提出了一种增强模型，通过结合混合膨胀卷积和新设计的残差结构，增强了模型提取浅层特征和位置信息的能力。最近，基于 YOLOv5 版本成功经验的 YOLOv8 被提出[45]，引入了新的骨干网络、全新的无锚点(Anchor-Free)检测头，以及一个新颖的损失函数。这些改进使得该模型在各种硬件平台上都能高效运行，从 CPU 到 GPU，都表现出色。

在工业缺陷检测领域，针对少样本和无监督学习的方法逐渐受到重视。Guo 等人[46]研究了一种基于生成对抗网络(GAN)的无监督少样本缺陷检测方法 ISU-GAN，旨在克服小样本环境中的缺陷检测难题。Karmakar 等人[47]探索了卷积神经网络(CNN)和多层感知器(MLP)在少样本环境下的应用，他们通过调整网络结构，有效提高了缺陷检测的效率。Min 等人[48]开发的 FS-RSDD 模型，一个针对铁路表面缺陷检测的简易而有效方案，采用少量样本训练，并利用原型学习技术，实现了精确的缺陷检测和定位。Pang 等人[49]采用了基于少样本学习的 MAML 框架，在样本量有限的条件下，对工业表面缺陷进行了深入研究，体现了该方法在概念学习中的有效性。

现有的缺陷检测算法主要面临以下不足之处或未解决的问题：首先，对于具有不规则分布和纹理特征不显著的缺陷，传统的特征提取方法和机器视觉技术往往难以准确识别和分类。尽管深度学习技术在这方面取得了进展，提升了自动化程度和处理复杂缺陷的能力，但在高精度和高效率方面仍有待改进。其次，现有算法在面对小目标缺陷时容易出现漏检和误检，尽管一些改进策略如引入注意力机制和多尺度特征融合已经被尝试，但这些方法的效果在实际应用中仍存在限制。最后，尽管 YOLOv5 和 YOLOv8 等较新的算法框架对提高检测速度和适应性方面做出了贡献，但如何在快速检测和高精确度之间找到最佳平衡，特别是在不同的硬件平台上，依然是现有缺陷检测方法需要解决的关键问题。针对以上问题，提出了基于机器视觉的金属表面缺陷检测的研究。

1.3 主要研究内容和创新点

本研究全面探讨了基于深度学习的金属表面缺陷检测算法，并成功构建了一套高效、适用于工业场景的缺陷检测模型架构。针对实际工业检测需求，本研究有效利用了真实工业场景中的金属表面缺陷数据和公共数据集，进而开发出了性能优异的分割模型、目标检测模型以及半监督检测模型。这些模型不仅提升了缺陷检测的准确性和效率，也体现了深度学习技术在工业质量控制领域的强大潜力。

本研究的贡献主要体现在以下几个关键方面：

数据库构建：成功创建了全面扩展的数据库，包括分割数据集和目标检测数据集，为开发高效缺陷识别算法提供了强大的数据基础。

分割网络的优化：通过对 DeepLabV3+分割网络的改进，实现了金属表面缺陷图像的高精度分割，得到了高质量的缺陷区域掩膜。

目标检测算法的改进：采用改进的信息增强型 YOLOv5 (IE-YOLOv5) 算法，显著提升了缺陷检测的速度与准确性，为实时缺陷检测提供了有效的技术手段。

创新的半监督检测网络：提出了一种新型的半监督检测网络设计，有效提高了模型在未标记数据上的学习效率和识别能力。

1.4 本文结构安排

第 1 章：首先介绍了金属检测的可研究性以及实用性，同时介绍了国内外对于目标检测的研究方法，其次介绍了国内外对于缺陷检测的现状，以及项目背景，详细介绍了深度学习在工业检测分析上的成果。最终总结了本文的主要研究内容和创新点。

第 2 章：金属表面缺陷检测数据集构建。首先对数据的来源进行了详细说明，并对数据集制作流程进行说明，并进行了一系列数据预处理步骤，以确保数据的质量和可用性。这些预处理步骤包括对数变换、图像滤波和直方图均衡化等操作。并在此基础上制作分割和分类模型训练所需的数据集，最后是本章内容总结。

第 3 章：金属表面缺陷语义分割技术分析。首先介绍了金属表面缺陷分割所使用的模型 DeepLabV3+原理，并介绍了 DASPP 模块和、三级注意单元(TAU)和挤压激励网络(SENets)，并改进出更适合金属表面缺陷分割的网络 Improved-DeepLabV3+。然后通过对比实验得出了不同语义分割网络模型对金属表面缺陷提取的影响，体现了 Improved- DeepLabV3+对缺陷检测的优越性，最后是本章内容总结。

第 4 章：金属表面缺陷目标检测技术分析。首先介绍了 YOLOv5 目标检测网络的原理，提出了 IE-YOLOv5 网络，随后介绍新的细颈特征融合模块的 FF-Slim-Neck 结构，之后并详解介绍了对比注意力学习模块——无参数空间注意机制 PSA ，介绍了该模型的损失函数和评价标准，通过训练该网络得到可用模型，实现了各类金属表面缺陷的高精度识别。并进行对比各个经典目标检测网络，进行消融实验和进行实验结果分析，最后是本章内容总结。

第 5 章：监督对比金属表面缺陷检测技术分析。首先，详细介绍了所设计的三个优化模块均值教师模式、CAE 模块和对比判断模块。随后介绍了损失函数和检测性能评估的选择。紧接着介绍了本章使用的网络在半监督数据集上的实验细节和消融实验对比，最后是本章内容总结。

第 6 章：综述了研究工作和取得的成果，同时指出了存在的不足之处以及改进的空间。对于金属表面缺陷检测技术未来的发展方向也进行了展望。

第 2 章 金属表面缺陷检测数据集构建

2.1 数据来源及其流程

由于钢材具备广泛的用途、持久的耐用性和经济效益，它在多个行业中得到了广泛应用，并且促进了可持续发展的实现。这在电力和建筑行业尤为明显，其中钢材的质量对于保障最终产品性能发挥了关键作用。然而，钢材表面可能出现的各种缺陷，如凹坑、划痕和裂纹，不仅削弱了材料的性能，还可能引发安全隐患和经济损失。因此，在毛坯阶段和加工过程中对金属零部件表面缺陷的检测和分类变得尤为重要，这有助于及时进行质量控制，提成品率和加工质量。

在工业缺陷检测领域，目前已有多个可用于缺陷检测的数据集。其中金属表面缺陷数据集已东北大学表面缺陷数据集(NEU-DET)最为典型。该数据集收集了热轧带钢 6 种典型的表面缺陷，轧内垢、斑块、裂纹、点蚀面、夹杂物和划痕。数据集共计 1800 张灰度图像，每种缺陷 300 个样本，每张图像的分辨率为 200×200 像素[50]。基于本课题目标应用对象为芜湖某机械零部件厂商提供的用于避雷器的金属电阻片，为了提供数据集的泛化性，拟结合 NEU-DET 和自采样图片，构建表面缺陷混合数据集。自建金属表面缺陷数据集和东北大学表面缺陷数据集(NEU-DET)共享一些金属材料的基本特征，并且可能在某些应用中存在相关性。通过融合不同类型金属材料的数据集，可以增加数据的多样性和丰富度，提高模型的泛化能力，减少过拟合的风险，同时也能够使训练出的模型具有更广泛的适用性。然而，在进行融合时需要注意数据质量、特征选择和领域知识的引入，以确保融合后的数据集对所需任务具有良好的效果。

电阻片表面缺陷图片数据的采集主要通过如图 2-1 所示数据采集平台实现。这一平台包含了光源、镜头、相机、图像采集模块以及图像处理模块等关键组成部分，用以实现图像的捕获、处理和分析，以便进行自动化的缺陷检测和识别。具体而言，光源对于确保捕获到的图像质量至关重要，能够保证对象被适当照明；镜头的作用是将光聚焦并传输到相机的感光元件上；相机则将光学图像转换为电子信号，随后由图像采集模块转换为数字信号，供计算机处理；图像处理模块最终对这些数字图

像进行分析，完成特征提取、模式识别等步骤。这些元件的整合构成了完整的机器视觉数据采集平台，对于基于机器视觉的金属表面缺陷检测至关重要。

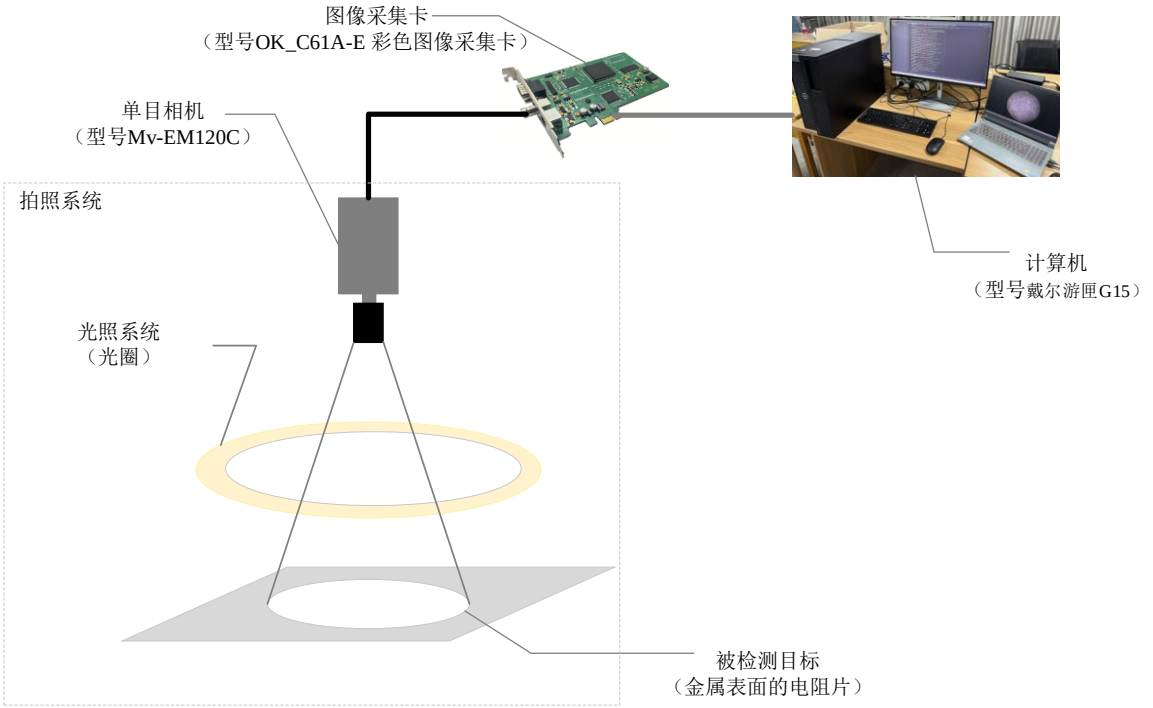


图2-1 数据采集平台原理图

具体的采集环境如图 2-2 所示，本研究采用的是光圈的照明方式，采集平台的背景为黑色，所采用的单目相机的型号为 Mv-EM120C，关键参数如表 2-1 所述，像素尺寸为 3.75μm*3.75μm，外形尺寸为 29*29*48.9，靶面尺寸为 1/3 英寸。它能够以最高 40fps 的速率采集图像，提供大约 130 万像素(1280 宽*960 高)的分辨率，配备了 4mm 的有效焦距镜头。这款相机的能耗低(1.2W)，具有紧凑的设计，与数据采集模块相结合时，支持 I/O 接口，以及 Windows 系统的驱动程序。

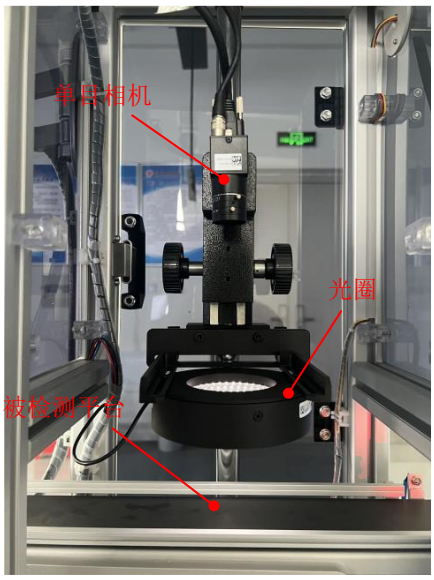
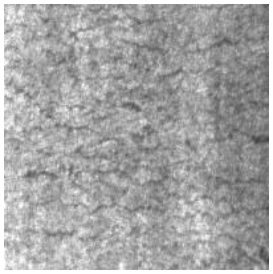


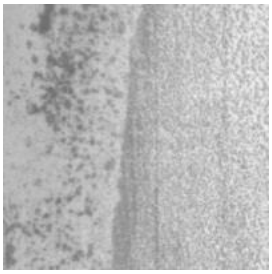
图 2-2 数据采集环境

表 2-1 相机主要参数

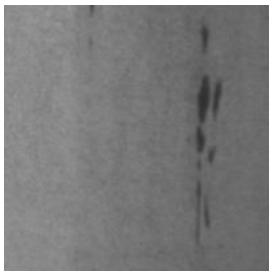
参数	数值
靶面尺寸	1/3 inch CMOS
像素尺寸	3.75μm*3.75μm
帧率	40fps
分辨率	1280H*960V
有效焦距	4mm
电源	12V

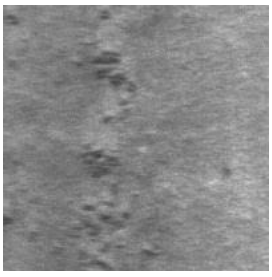
本研究所设计的分割和识别算法所采用的数据集为自建的混合数据集，支持多分类任务，专注于金属表面缺陷识别。该数据集包括金属表面图像及其对应的缺陷类型标签，所构建的混合数据集覆盖了广泛的缺陷类型和不同金属表面条件，保证了模型能够适应于多样化的实际应用场景。数据集部分图像来源于 NEU-DET 数据库，如图 2-3 所示，包括六种缺陷类型轧内垢(RS)、斑块(Pa)、裂纹(Cr)、点蚀面(PS)、夹杂物(In)、划痕(Sc)。


(a) 裂纹(Cr)


(b) 夹杂物(In)


(c) 斑块(Pa)


(e) 点蚀面(PS)


(f) 轧内垢(RS)

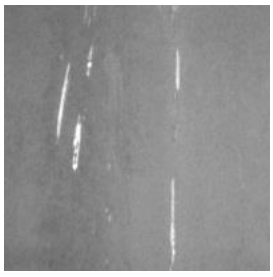
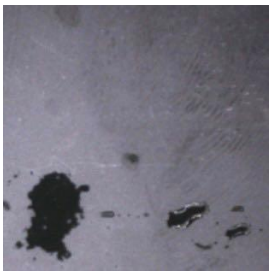
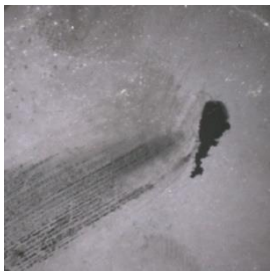

(g) 划痕(Sc)

图 2-3 NEU-DET 缺陷图像

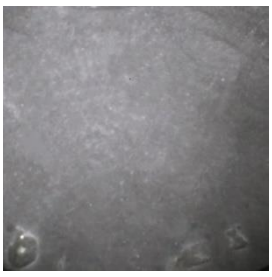
此外，图 2-4 展示了使用高精度数据采集平台从缺陷电阻片获取的金属表面图像。


(a) 裂纹(Cr)


(b) 斑块(Pa)


(c) 斑块(Pa)和夹杂物(In)


(e) 点蚀面(PS)


(f) 轧内垢(RS)


(g) 划痕(Sc)

图 2-4 自采集数据图像

通过对多个数据集的综合性分析，可以得出，混合数据集作为一个综合性的缺陷检测数据库，其全面性表现在丰富的金属表面缺陷图像收录，极为适宜本研究课题的训练数据需求。鉴于本课题的研究方法旨在结合分割和分类网络来处理缺陷检

测问题。因此，数据集的适当预处理和对标签系统的精细重构显得尤为重要。对数据集中的图片数据的预处理及标签的重新定义成为必要的步骤。具体而言，标签的重定义将遵循图 2-5 所示的实验流程，确保数据处理的准确性与实验结果的可靠性。

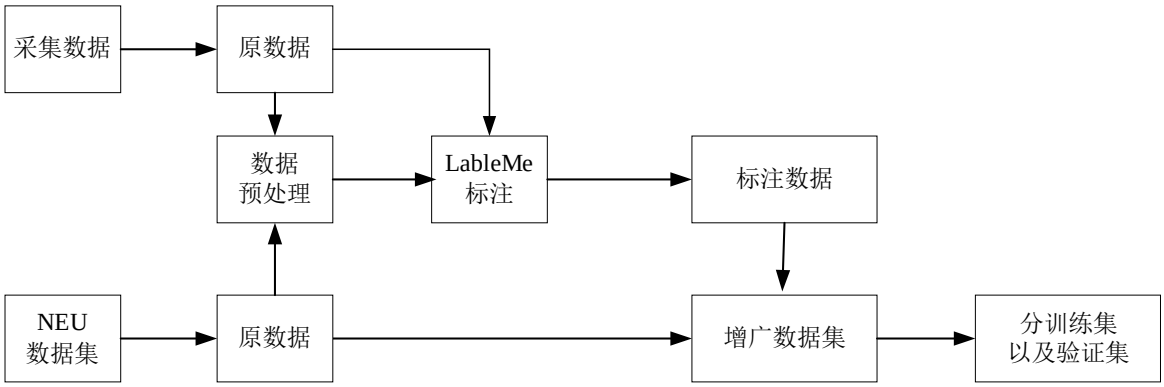


图 2-5 标签制定流程图

2.2 数据预处理

2.2.1 对数变换

在数字图像处理领域，对图像进行增强是提高图像质量和可视化效果的关键步骤之一。对数变换[51]作为一种常见的图像增强方法，在调整图像对比度方面发挥着重要的作用。

对数变换是通过对图像的像素值进行对数运算来调整图像的对比度。对数变换的数学表达式如(2-1)所示：

$$g(x,y)=c\cdot log(1+f(x,y)) \tag{2-1}$$

其中，$f(x,y)$表示输入图像的像素值，$g(x,y)$表示变换后的像素值，c 是调节对比度的常数。

其函数图像如图 2-6 所示，由于 0 的对数是未定义的，因此该函数是在 0.01 到 2 的范围内绘制的。从对数函数的特征曲线图可以看出，对数函数随着 x 的增大而逐渐增大。

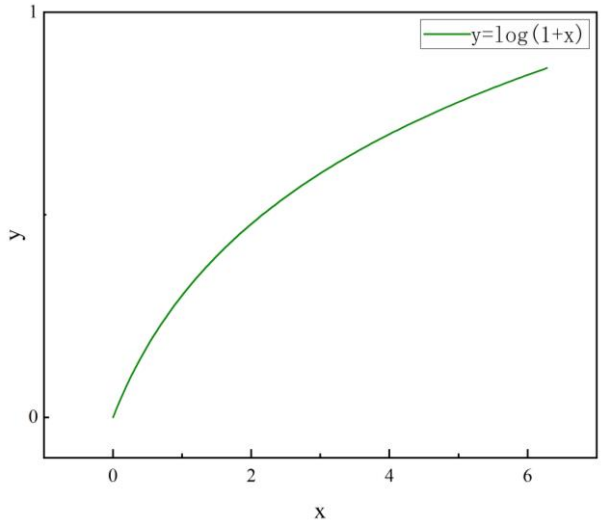
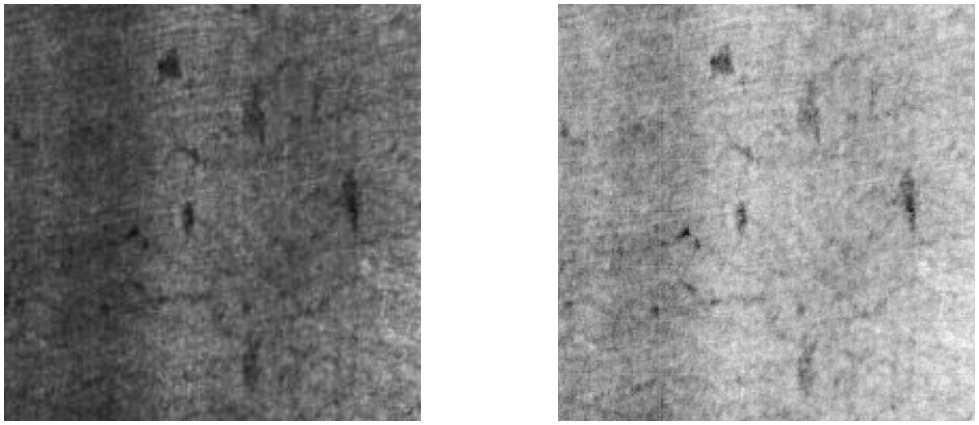


图 2-6 对数变换函数曲线图

对数变换通过对每个像素的灰度值进行对数运算，有效扩展了图像的动态范围。图 2-6 通过直观的视觉对照，展示了对比变换的效果。这组图像被分为两部分：图 2-7(a)展示原始图像的基本特征和细节，提供了对比变换前状态的基准；图 2-7(b)则展示了经过对比度调整后的图像，更清晰地突出了缺陷。原图中的缺陷可能由于对比度、亮度或色彩分布的限制而不够明显或难以识别。相比之下，图 2-7(b)中的图像通过调整对比度，增强了图像中原本不明显的特征，如裂纹、瑕疵或其他类型的缺陷，使之变得更加明显。这种变换主要涉及调整亮度和对比度参数，以便在图像中更好地区分不同区域。



(a) 原始图像　　(b) 对数变换后的图像

图 2-7 对数变换后的缺陷对比图片

定量分析：评价对数变换处理效果的指标包括平均像素值和标准偏差，这两者是在图像处理领域中用于衡量像素分布均匀性的关键统计参数。其计算公式如下：

平均像素值 M：

$$M = \frac{1}{N}\sum_{i=1}^{N} p_i \tag{2-2}$$

其中 N 为像素数，并且 p_i 表示第 i 个像素的值。

标准偏差 SD 衡量像素值：

$$SD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(p_i - M\right)^2} \tag{2-3}$$

其中 p_i 是第 i 个像素的值，M 是平均像素值，N 是图像中像素的总数。

图 2-7 为对数变换后的缺陷对比图片，对数变换前后图像的平均像素值和标准偏差具体数值列在表 2-2 中。在对数变换之前，图 2-7(a)的原始图像的平均像素值为 89.64，标准偏差为 23.54。对数变换后，图 2-7(b)的平均像素值显著提高至 214.11，标准偏差降低至 12.82。这一变化表明对数变换显著增加了图像的平均亮度，并改善了暗区域的细节可见性，同时缩小了像素值的分布范围，从而在低亮度区域增强了对比度。

表 2-2 对数变换评价指标

量化指标	原始图像	对数变换后的图像
平均像素值(M)	89.64	23.54
标准偏差(SD)	214.11	12.82

2.2.2 图像滤波

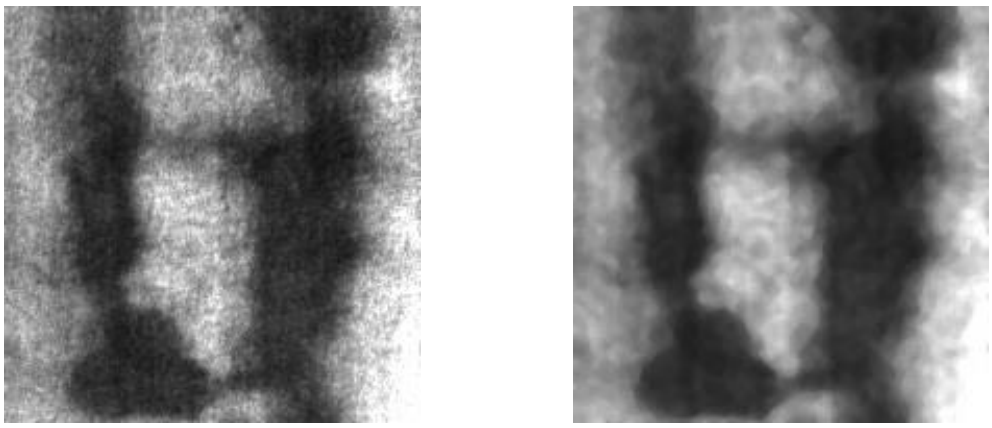
图像滤波[52]是数字图像处理中一项关键技术，其目的在于去除图像中的噪声、平滑细节、增强图像特征等。本节将深入研究图像滤波的原理、常见滤波器以及在不同应用场景中的效果。

在图像处理中，中值滤波的核心理念是将每个像素的灰度值替换为其周围邻域的中间值。与其他线性滤波技术相比，中值滤波在处理椒盐噪声等异常值时表现更为出色。其数学表达式为：

$$I_{new}(x,y) = median(I_{old}(x-1,y-1),\ldots,I_{old}(x+1,y+1)) \tag{2-4}$$

其中 $I_{new}(x,y)$ 是新的像素值，$I_{old}(x,y)$ 是原始图像的像素值，中值函数对邻域内像素值进行排序并取中值。

对缺陷图像进行中值滤波后的图像如 2-8 所示，图 2-8(a)为原图像，图 2-8(b)为中值滤波后的图像。



(a) 原图　　(b) 中值滤波后的图片

图 2-8 中值滤波的缺陷对比图片

其函数图像如图 2-9 所示，该图说明了中值滤波对噪声信号的影响。原始信号(绿色)是添加了随机噪声的正弦波。中值滤波信号(红色)是通过应用核大小为 5 的中值滤波器得到的。该滤波器能有效减少噪音，使信号更加平滑，同时保留了正弦波的整体形状。

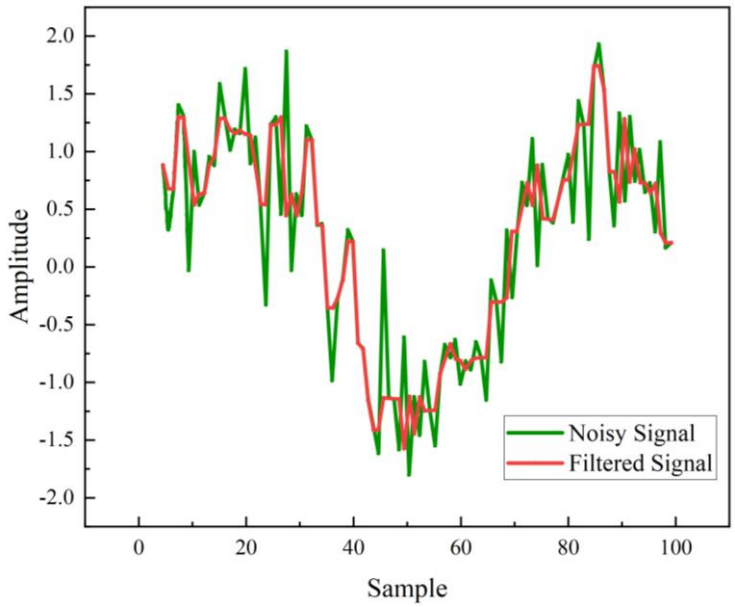


图 2-9 中值滤波函数图

定量分析：采用结构相似性指数(SSIM)和峰值信噪比(PSNR)作为评价指标，对经过中值滤波处理前后的图像进行了定量分析。

结构相似性指数(SSIM)是一项用于比较两张图像相似性的指标，其考虑了亮度、对比度和结构三个方面的变化。对于两个窗口 x 和 y(窗口是图像的一部分)，SSIM 定义为：

$$SSIM(x,y)=\frac{\left(2\mu_x\mu_y+c_1\right)\left(2\sigma_{xy}+c_2\right)}{\left(\mu_x^2+\mu_y^2+c_1\right)\left(\sigma_x^2+\sigma_y^2+c_2\right)} \tag{2-5}$$

其中 μ_x 是 x 的平均值，μ_y 是 y 的平均值，σ_x^2 是 x 的方差，σ_y^2 是 y 的方差，σ_{xy} 是 x 和 y 的方差，c_1 和 c_2 是为了避免分母为零而加入的小常数，通常 $c_1=(k_1L)^2$，$c_2=(k_2L)^2$ 其中 L 是像素值的动态范围，k_1 和 k_2 是常数，通常取值 0.01 和 0.03。

PSNR 是最常用于测量图像重构质量的指标之一，特别是在图像压缩领域。PSNR 被定义为：

$$PSNR=10\cdot\lg\left(\frac{MAX_I^2}{MSE}\right) \tag{2-6}$$

在这个过程中，*MAX* 代表着图像可能的最大像素值，比如对于 8 位图像来说，它就是 255。而 *MSE* 则表示了原始图像与估计图像之间的均方误差，计算方法为：

$$MSE=\frac{1}{mn}\sum_{i-1}^{m}\sum_{j=1}^{n}(I(i,j)-K(i,j))^2 \tag{2-7}$$

其中 I 是原始图像，K 是估计图像，m 和 n 是图像的尺寸。

通过应用结构相似性指数(SSIM)和峰值信噪比(PSNR)，对原始图像(图 2-6(a))和中值滤波处理后的图像(图 2-6(b))进行了比较，具体数值见表 2-3。这两幅图像的 SSIM 值大约为 0.656，指示在结构信息方面存在明显差异。同时，PSNR 值达到 28.09 分贝(dB)，表明虽然滤波处理后的图像在视觉质量上与原始图像有所不同，但这种差异程度通常视为在可接受的范围内。

表 2-3 中值滤波评价指标

量化指标	数值
SSIM	0.656
PSNR	28.09 dB

2.2.3 直方图均衡化

直方图均值化[53]是图像处理中一种常用的增强技术，旨在调整图像的对比度和亮度，使其分布均匀。通过对图像的像素值进行调整，采用直方图均值化技术可以改善图像的视觉效果，进而突出细节，使得图像更易于进行分析和理解。

直方图均值化的核心原理是通过调整图像的灰度级分布，使其接近均匀分布。具体而言，对于输入图像的灰度直方图，直方图均值化将调整图像的灰度级，以使输出图像的直方图更加平均。这一过程可以通过(2-8)数学公式表示：

$$G(x,y)=T[F(x,y)] \tag{2-8}$$

其中，$G(x,y)$是均值化后的图像，$F(x,y)$是原始图像，T 而表示灰度变换函数。通过对灰度变换函数的设计，可以实现对图像的调整，使其灰度级更加均匀。

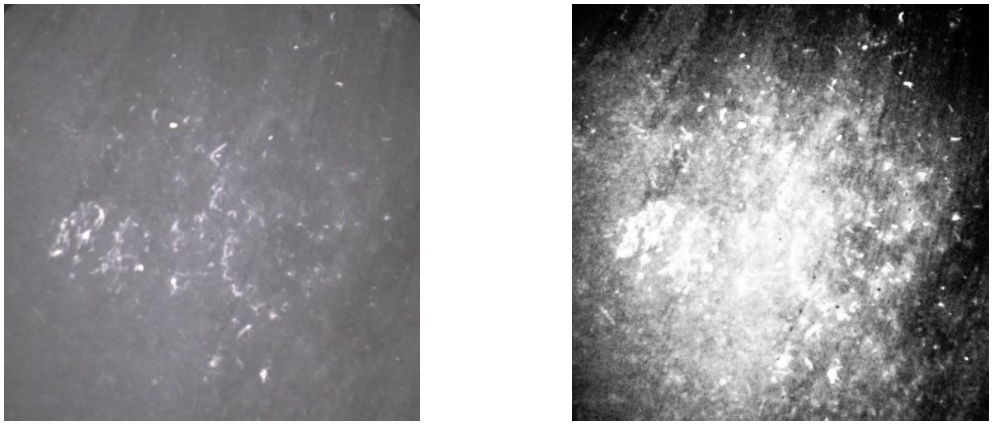
直方图均衡化的过程通常涵盖以下步骤：首先，计算原始图像的灰度直方图，这涉及统计每个灰度级别的像素数量以获得图像的灰度分布。接着，基于该灰度直方图，计算累积分布函数(CDF)，为灰度变换做准备。其公式为：

$$CDF(v)=\frac{1}{MN}\sum_{i=0}^{v}h(i) \tag{2-9}$$

然后，设计灰度变换函数，该函数基于 CDF，目的是将原始图像的灰度级映射至一个更均匀的分布。其公式为：

$$g(v)=round((L-1)\cdot CDF(v)) \tag{2-10}$$

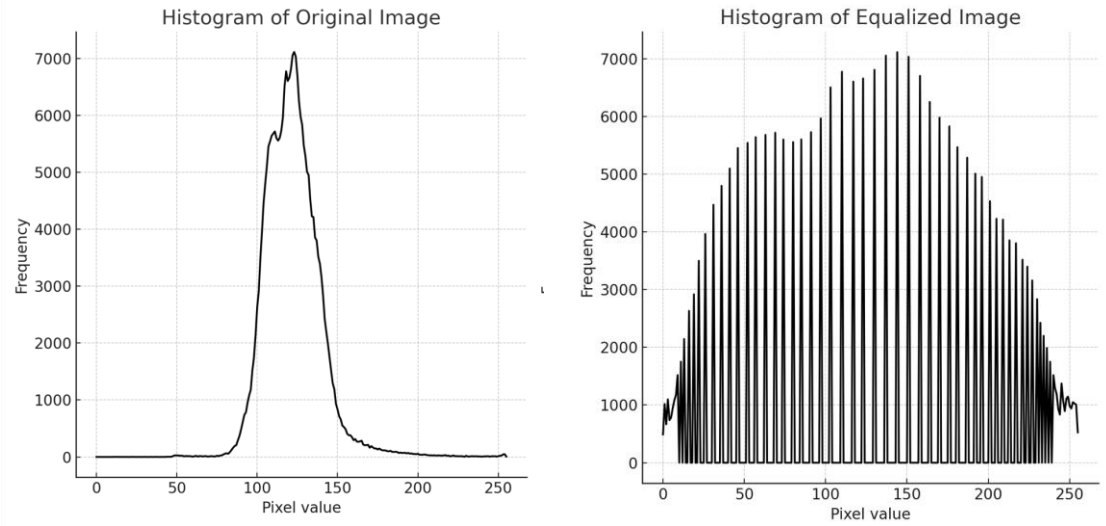
最后，应用这一灰度变换，生成均衡化的图像，以增强其对比度。图 2-10(a)中展示了因照明条件不佳或相机设置错误等原因而对比度较低的原始图像，这导致图像在暗部或亮部区域，细节识别变得困难。相反，在图 2-10(b)中，可以看到经过直方图均衡化处理，使得原有的灰度分布被扩散，覆盖整个亮度范围，改善了图像对比度和细节的可见性。



(a) 原图　　(b) 直方图均衡化后的图像

图 2-10 直方图均衡化前后对比图像

如图 2-11 所示，图 2-11(a)显示的是原始图像的直方图，可以看到像素值集中在某些范围内。图 2-11(b)显示的是均衡化后图像的直方图。均衡化过程重新分配了像素强度，使其更均匀地覆盖整个范围，这通常会增加图像的整体对比度，尤其是当图像的可用数据由接近的对比度值表示时。这种效果可提高图像中细节的可见度。



(a) 原图直方图　　(b) 直方图均衡化后的直方图

图 2-11 直方图对比图

定量分析：对直方图均衡化的评价指标为亮度分布均匀性：直方图均衡化能够使得图像的亮度分布更加均匀，避免了图像中出现过亮或过暗的区域。标准差是描述图像亮度分布均匀性的常用指标之一。标准差越小，表示图像的亮度分布越均匀。标准差的计算公式如(2-11)所示：

$$\sigma=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2} \tag{2-11}$$

其中，N 表示像素总数，x_i 表示每个像素的灰度值，$\bar{x}$ 表示灰度值的平均值。

亮度分布的均匀性通过原始图像(图 2-10(a))与直方图均衡化处理后的图像(图 2-10(b))的标准偏差进行了比较，其数值详细列于表 2-4。

表 2-4 直方图均衡化评价指标

量化指标	原图	直方图均衡化后的图像
σ	809.78	814.59

由表 2-3 可知，直方图均衡化前，图 2-9(a)的标准偏差大约是 809.78，而处理后的图 2-9(b)的标准偏差为约 814.59。这表明通过均衡化处理，像素强度在直方图上的分布略有增加，指示出分布的均匀性得到了提高。直方图均衡化的目标是增强图像对比度，实现这一点通过使像素强度分布更均匀。比较处理前后的图像，通过标准偏差的变化可直观地评估对比度和图像特征可见度的改善，进一步验证均衡化处理对图像质量的积极影响。

2.3 缺陷分割数据集建立

对于分割网络模型，需要特别注重图像的局部细节和缺陷边缘的清晰度，因此在数据筛选时倾向于选择那些边缘更为清晰、缺陷类型更为明显的图像。缺陷检测与分割是图像处理领域中重要的任务，尤其在工业生产和医学影像等应用中具有广泛的应用。为了有效地处理和分析缺陷图像数据，本小节探讨了使用 LabelMe 工具进行缺陷检测与分割数据标注的方法，以及这种方法在提高数据标注质量和训练深度学习模型方面的潜在优势。

2.3.1 缺陷分割

在图像处理领域，缺陷分割[54]任务扮演着至关重要的角色，尤其在工业质量检测与医学影像分析等关键领域中，其应用范围广泛。该任务的核心目标在于通过高精度的技术手段，从复杂的图像数据中准确提取并定位缺陷区域，这不仅对识别图像中的异常部分至关重要，而且为深入分析与决策制定提供了坚实的基础。

为了深入探索缺陷分割技术的有效性及其应用潜力，本研究通过使用 LabelMe 工具，精心标注了一个包含 3000 张图像的数据集。这些图像覆盖了多种类型的缺陷，包括轧内垢(RS)、斑块(Pa)、裂纹(Cr)、点蚀面(PS)、夹杂物(In)、划痕(Sc)，以及图像的背景(Bg)。

2.3.2 缺陷数据标注

经过对数据库组合的细致分析并结合现实情况考量，观察到在金属表面缺陷提取过程中存在大量的噪声和其他干扰因素(见图 2-12)。这些干扰因素显著影响了缺陷分割和提取的效率，尤其是在金属表面的关键区域。为了改善这一问题，对图像进行了详细的标注，以便在后续工作中更有效地进行缺陷检测和特征提取。

图 2-12 干扰因素

LabelMe 是一个开源在线图像标注工具，设计理念是提供一个用户友好的界面以便用户能够轻松地进行图像中的对象、区域或边界框的标注。这款工具支持多标签和多对象标注，并允许添加文字注释，使得图像的细粒度标注变得既简单又灵活。用户可以直观地在图像上标示出缺陷区域，并为每个缺陷指定类型，例如划痕、凹陷或锈蚀等。此方法不仅提升了标注的精确度，也由于其直观的界面，显著降低了标注过程的时间成本和复杂度。

采用 LabelMe 手动标注增强了数据集的精确度，并为模型训练奠定了高质量的基础。精确的数据标注使模型能更高效地学习并识别不同类型的金属缺陷，这对提升检测网络的准确性和可靠性极为关键。此外，一个全面和多层次的标注及验证流程保证了数据集中每个缺陷标记的精确度和可靠性。

在缺陷检测和分割的数据标注环节中，LabelMe 工具支持图像的上传及其标注工作：缺陷图像可以被上传到 LabelMe 平台，并且通过绘制工具对缺陷区域进行精细的标注，这包括缺陷的轮廓和边缘信息。进一步，用户还能为这些标注的缺陷区域增加标签和文字说明，以区分不同的缺陷类型并提供附加信息。完成的图像数据能够以标准格式导出，比如 JSON 格式，以便于进行后续的数据处理和为模型训练做准备。图 2-13 展现了 LabelMe 的一个标注界面实例。

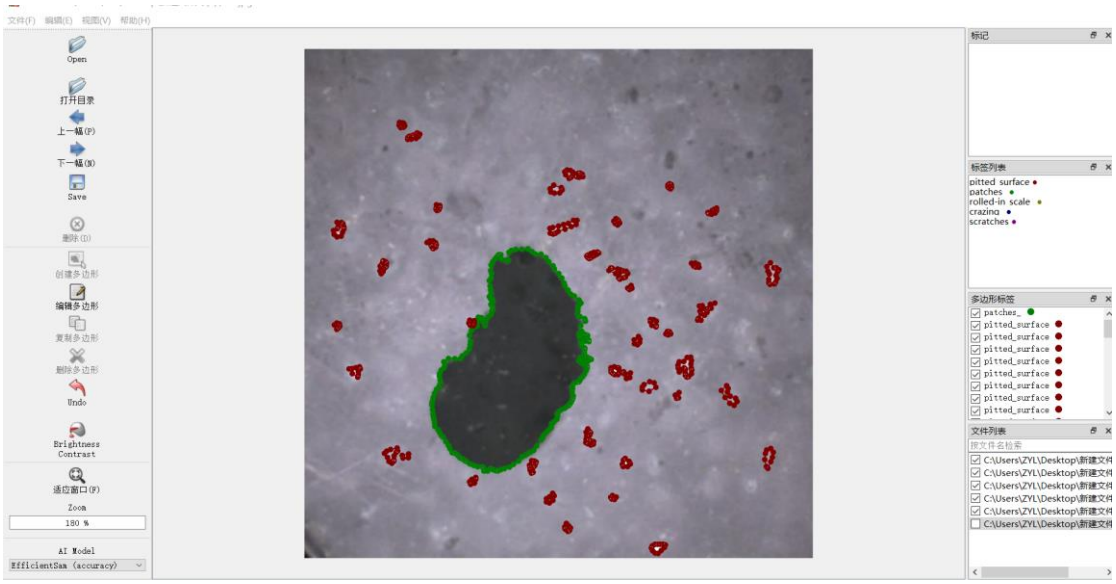
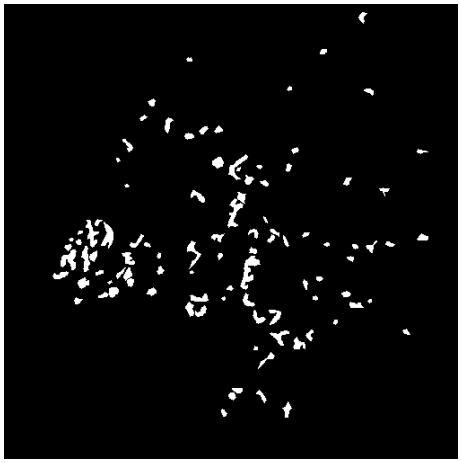


图 2-13 LabelMe 缺陷标注界面

为了确保标注数据具有极高的准确性，采用了复合验证策略。最初的审查工作交由具有丰富经验的审核人员执行，目的是初步判断标注的准确度。随后，这些标注信息将被送至另一名审核人员手中进行交叉检验，以此保障识别及分类的缺陷达到最高的准确标准。该流程设计旨在增强标注数据的精确度与信赖度。标注过程结束后，所有数据将被储存为 JSON 格式，该格式详细记录了包括图像尺寸(高、宽及层次)、标注类型以及特定的标记点位置和大小等关键信息。示例文件的内容如图 2-14 所示。此类系统化的标注手段不仅增进了数据的有序性，同时也便于后续的处理与分析工作。

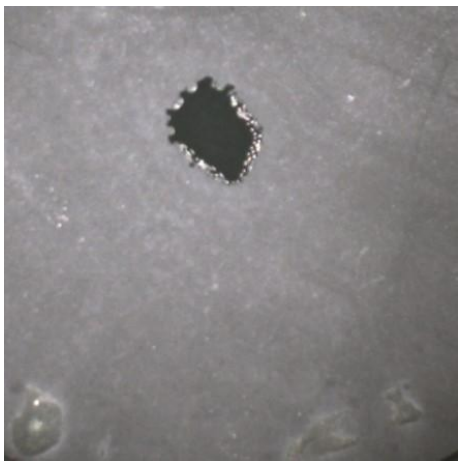
(a) 缺陷标签 json 文件



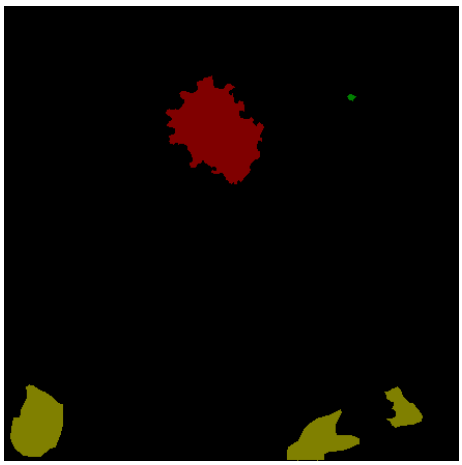
(b) 缺陷标签 mask 文件

图 2-14 缺陷标注数据标签

虽然 JSON 文件中包含了网络训练所需的关键信息，但它们不适合直接应用于训练过程。因此，需要通过 Python 脚本先行提取这些信息。在处理过程中，关键信息(如类别、中心点以及锚框的尺寸)将被保存在 TXT 文件中，用于缺陷检测的标签提取。与此同时，为了缺陷标注，这些关键信息也将被转换并保存为 PNG 格式的图像。此步骤为网络训练的进行提供了必要的数据准备。具体的操作流程和结果如图 2-15 所示。因此采用这种方法有助于维持数据格式的统一和操作的简便性，来提高训练的效率和精度。



(a) 工业相机原图像



(b) 缺陷标签图像

图 2-15 缺陷分割数据集

选取 3000 张经过 2.2 章节预处理后的金属表面图像，随后每张图像均通过 LabelMe 标注工具进行了详尽的标记工作，以确保数据的高精度和高一致性。图像中的缺陷被细致地标注为七个不同的类别，包括轧内垢(RS)、斑块(Pa)、裂纹(Cr)、

点蚀面(PS)、夹杂物(In)、划痕(Sc)以及图像的背景(Bg)，旨在覆盖金属表面缺陷的主要类型。

2.4 缺陷分类数据集建立

对于分类网络模型，考虑到其对整体图像特征的敏感性，选取了包含不同尺寸、形状和对比度缺陷的图像。在数据预处理阶段，对图像进行了多种处理，以提高训练效率和模型准确性。这些预处理步骤包括 2.2 章节所采用的对数变换、图像滤波和直方图均衡化后的图像以符合模型的输入要求；对数变换以增强缺陷特征；以及还需要数据增强技术，如旋转、翻转和噪声添加，这有助于提升数据集的多样性及模型的泛化性能。

2.4.1 缺陷分类

本研究旨在通过构建一个分类网络数据集，加强机器学习算法在此领域的应用效率和准确性。经过精密的预处理和标注工作，将金属表面缺陷图片详细分类为六个主要类别：轧内垢(RS)、斑块(Pa)、裂纹(Cr)、点蚀面(PS)、夹杂物(In)和划痕(Sc)。该数据集的类别及其对应的标签名详细列在了表 2-5 中，每一类别的定义都基于缺陷的形态特征和产生机理，为分类网络提供了明确的学习目标。

表 2-5 对应类别及其标签名

类别	标签类别	类别	标签类别
0	轧内垢(RS)	1	斑块(Pa)
2	裂纹(Cr)	3	点蚀面(PS)
4	夹杂物(In)	5	划痕(Sc)

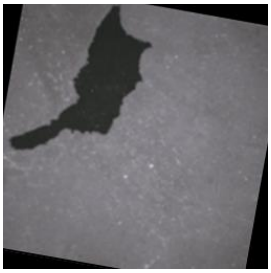
此分类体系的建立，不仅有助于提高分类网络在实际检测中的性能，更是对工业自动化视觉检测系统的一项重要贡献。通过对数据集的深入学习，分类网络能够更有效地识别和区分不同缺陷，这为之后的产品质量评估和故障诊断提供了可信的决策支持。

2.4.2 数据增强

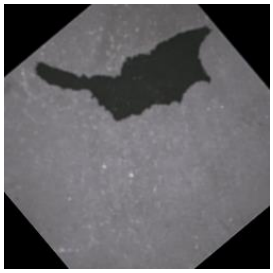
数据扩充通过对原始数据应用一系列变换，旨在创建新的训练样本，以此提升训练集的数量和多样性[55]。此方法不仅增强了模型的泛化性能，也有助于减少过拟

合的可能性，同时提升了模型对输入数据的稳健性。鉴于原数据集中样本的显著差异性，例如，斑块(Pa)和夹杂物(In)的数量不足 500 张，而划痕(Sc)的数量超过 1200 张，采用数据增强可以防止模型过拟合，并优化网络模型的性能及其检测准确率。本节介绍了在原有数据集基础上实施的数据增强技术，包括随机旋转、翻转、添加图像噪声以及调整图像的亮度和锐度。数据增强的示例见图 2-16。

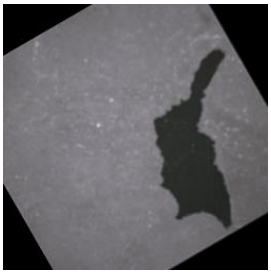
(a) 原图



(b) 旋转 20°



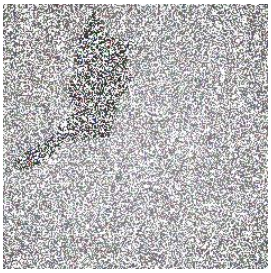
(c) 旋转 45°



(d) 旋转 120°

(e) 垂直翻转

(f) 水平翻转



(g) 添加噪音

(h) 亮度变换

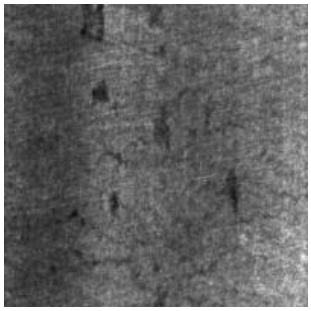
(i) 锐度变换

图 2-16 数据增强图

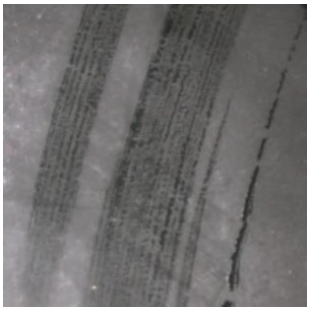
数据扩充有助于模型学习更广泛、更复杂的数据分布，提高模型在未见过数据上的性能。数据扩充通过增加多样性，有助于缓解模型对训练数据的过拟合问题，并提升了模型的泛化能力。在数据有限的情况下，数据扩充可通过有效利用现有数据，使模型更好地学习任务特征，提高整体性能。

在本研究中，采集并处理了 6000 张金属表面缺陷图像，并将其严格分类为六个独立的缺陷类别，轧内垢(RS)、斑块(Pa)、裂纹(Cr)、点蚀面(PS)、夹杂物(In)和划痕(Sc)，如图 2-17 所示，构建了一个多类别金属表面缺陷分类数据集。这为缺陷识

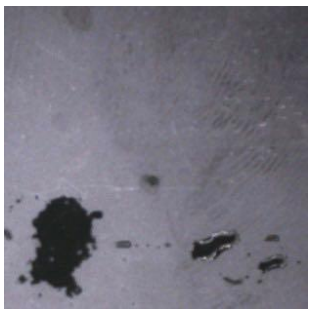
别提供了明确的分类标准和参照。进一步地，为了确保模型训练的有效性及评估的准确性，按照 6:2:2 的比例，将整个数据集划分为训练集、验证集和测试集三个子集。



(a) 裂纹(Cr)



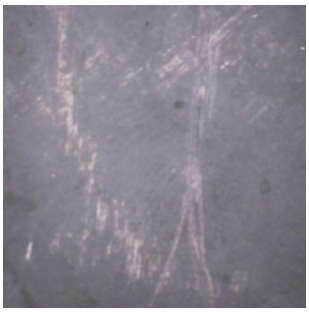
(b) 夹杂物(In)



(c) 斑块(Pa)

(d) 点蚀面(PS)

(e) 轧内垢(RS)



(f) 划痕(Sc)

图 2-17 分类数据集

此外，本文所述的金属表面缺陷分类数据集的详细组成和样例图像如图 2-18 所示。

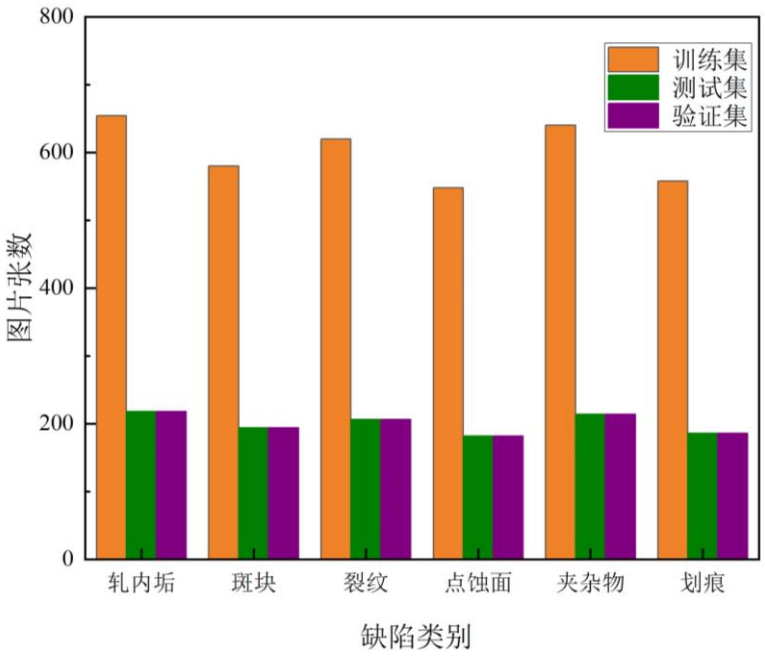


图 2-18 金属表面分类数据集数据分布情况

2.5 本章小结

本章的主要工作集中在建立和优化两种金属表面缺陷检测数据集：金属表面缺陷分割数据集和金属表面缺陷分类数据集。每个数据集的构建都是为了应对特定的检测和分类挑战，来提高整个缺陷检测系统的综合性能和准确性。

在构建金属表面缺陷分割和分类数据集时，采取了一系列的图像预处理措施，这些步骤的选择和设计旨在确保数据集的一致性和标准化，从而为后续的实验研究提供准确和可重复的结果。具体的预处理措施包括对数变换、中值滤波和直方图均衡化等。对数变换作为一个有效的对比度增强技术，能够使图像中的低对比度区域增强，进而突出那些在原始图像中可能不太明显的细微缺陷。这一变换对于改善图像的可视性尤为关键，因为它直接影响到后续分类和分割算法的识别能力。中值滤波是图像去噪的经典方法，它通过替换像素点的值为其邻域像素值的中位数来消除随机噪声，这样操作不仅可以有效消除噪声，同时在最大程度上保留图像的边缘和纹理信息，对于保持图像细节至关重要。另外，还进行了直方图均衡化处理，这是一种调整图像强度分布的方法，使得图像有一个更均匀的亮度分布。这种处理能够提升图像整体的对比度，使得缺陷区域与背景之间的差异更加明显，进一步增强分类和分割算法的性能。

对于金属表面缺陷分割数据集的构建，特别关注于缺陷的种类和特征。通过对现有数据集的细致分析和分类，将轧内垢(RS)、斑块(Pa)、裂纹(Cr)、点蚀面(PS)、夹杂物(In)、划痕(Sc)和背景(Bg)七个类别。每个类别都经过精确的标注，使用了 LabelMe 工具进行手动标注，确保了标注的准确性和一致性。最终，这个数据集包含了约 3000 张精细标注的影像，为金属表面缺陷的自动识别和分类提供了丰富的资源。

对于金属表面缺陷分类数据集的构建，对数据集进行数据增强进行扩充，扩充到 6000 张，再对数据集进行分类，分成轧内垢(RS)、斑块(Pa)、裂纹(Cr)、点蚀面(PS)、夹杂物(In)、划痕(Sc)六个类别。

综上所述，本章的工作不仅在于数据集的量化构建，更体现在对数据质量的精细控制和处理方法的创新。通过这些精心设计的数据集，为金属表面缺陷的自动检

测和分类提供了坚实的数据基础，这对于提高工业生产的自动化水平和产品质量具有重要意义。

第 3 章 金属表面缺陷语义分割技术分析

3.1 本章方法概述

针对 DeepLabV3+模型在缺陷边缘分割的不精准性、小尺寸目标的漏分割现象以及网络参数量庞大等挑战。因此，本研究的目标是提出一种有效的深度学习网络，利用多种注意力技术来减少编码器和解码器之间的语义信息差距，实现金属表面缺陷图片的精确分割。

3.2 基于 DeepLabV3+缺陷分割网络

DeepLabV3+[56]分割网络能够自动识别和分割图像中的缺陷，减少了人工干预的需求，提高了检测效率。这对于大规模生产中的质检任务尤为重要。分割网络能够提供高精度的缺陷分割，使得检测结果更为准确。与传统的基于规则或阈值的方法相比，深度学习分割模型可以更好地适应复杂的场景和不规则形状的缺陷。DeepLabV3+分割网络能够识别和分割多种类型、多样性的缺陷，包括形状、大小、颜色等方面的变化。这使其适用于处理各种实际场景中可能出现的缺陷。

DeepLabV3+的工作流程如 3-1 所示，待测图像首先被输入到网络中，接着通过卷积、池化等步骤进行特征提取，并通过注意力机制进行增强。然后，对编码器和解码器提取的特征进行融合，调整特征层的通道数，通过上采样操作匹配输入输出图像的大小，最终得到预测结果。

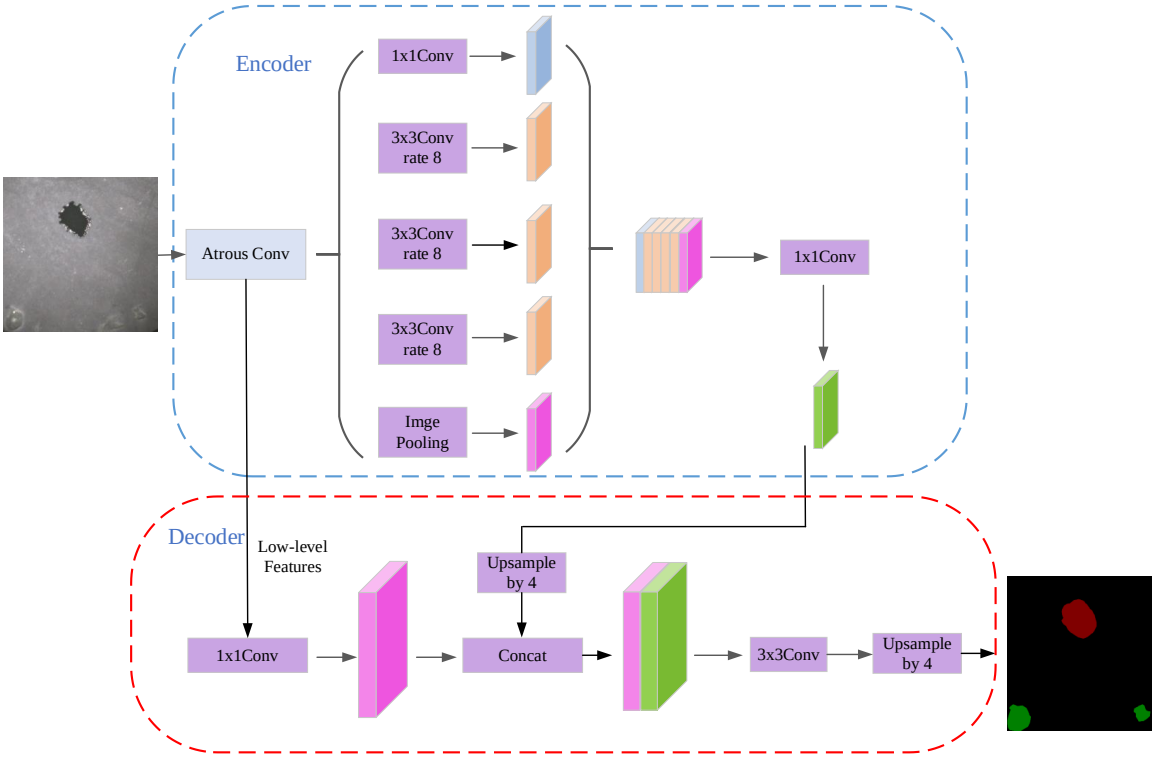


图 3-1 DeepLabV3+结构

在图像输入后，DeepLabV3+模型首先通过深度卷积网络(DCNN)[57]从图像中提取特征，区分为高级和低级语义特征。低级特征直接送入解码器，而高级特征进入编码器。在编码器阶段，引入 Atrous Spatial Pyramid Pooling(ASPP)模块[58]以捕获金属表面缺陷的特征，并需求大的接收野。但是，扩张率的增加导致像素采样变稀疏，进而可能导致细节信息丢失。因此，随着扩张卷积的效率降低，ASPP 模块的性能也受影响。在解码器阶段，模型采用 4 倍上采样，这对于精确边缘分割尤其是金属表面缺陷而言可能不利。此外，仅将解码器的输出与网络底层特征融合，可能导致信息丢失，进而影响分割精度。

3.2.1 DeepLabV3+分割模型的改进

为了解决以上问题，所提出的 Improved- DeepLabV3+模型将 DeepLabV3+架构与三级注意单元(TAUs)和挤压激励网络(SENets)相结合，提高了分割效率和精度。更具体地说，将 TAU 与该架构的 DASPP 部分进行了紧密的集成。在每个卷积层之后，引入了 TAU，以有效地提取更加细致和具有代表性的特征。这种结

合使得模型能够更好地捕获图像中的重要信息和细微差异，从而提高了模型在复杂场景下的性能。

此外，引入卷积注意模块增强有效特征信息，抑制无关信息响应，提高特征提取和表示能力。最后，在解码器后嵌入残差细化模块，映射从上层传输的重要信息，优化缺陷边界，提高分割精度。

同时，在网络的解码器部分引入了挤压激励网络(SENets)。SENets 通过对特征图的通道进行加权，有针对性地增强重要特征的响应，提高了模型对关键信息的关注程度。该机制对模型的特征提取过程进行了进一步的优化，使得模型能够更加有效地适应多样化的图像场景，从而增强了模型的泛化能力和鲁棒性。

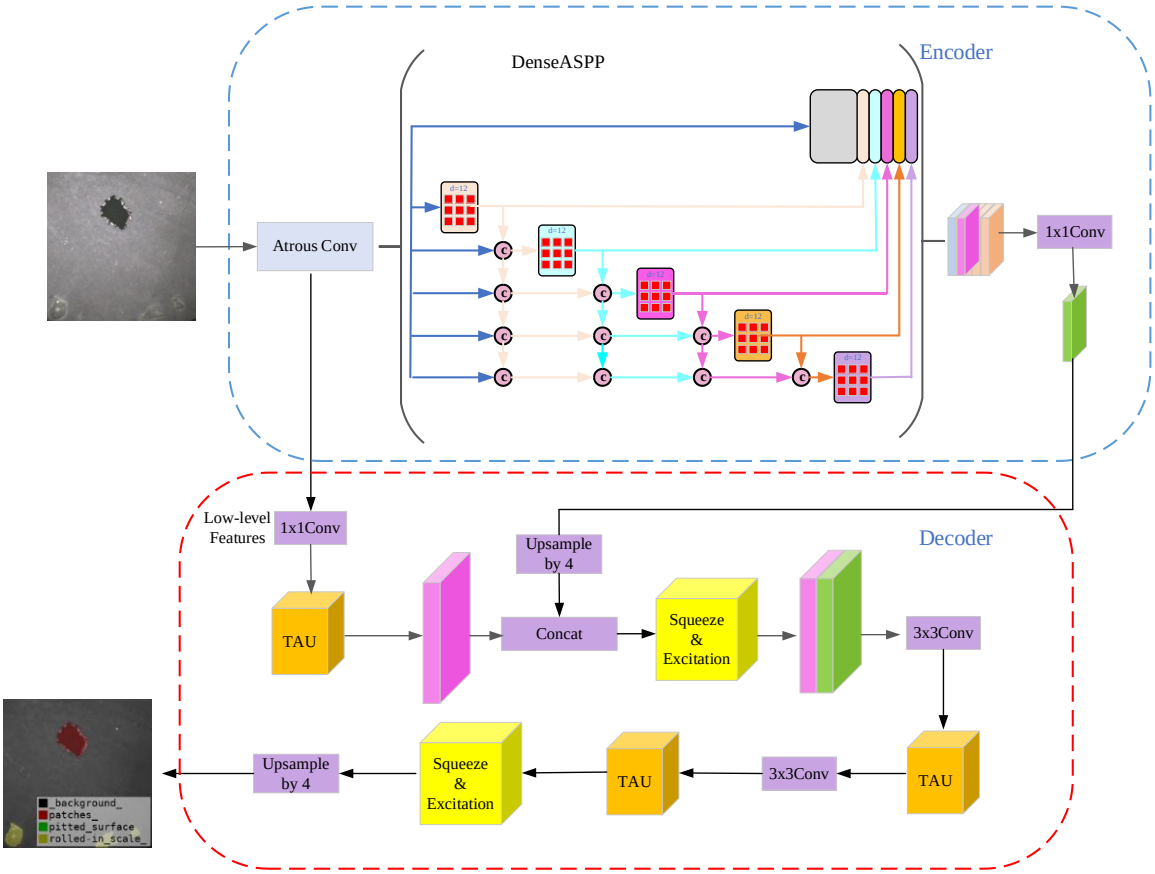


图 3-2 Improved DeepLabV3+网络结构

综合起来，改进后的网络架构在保持原有 DeepLabV3+优势的基础上，通过融合 TAU、DASPP、SENets 等先进技术，实现了更加精细的特征提取和更高效的信息利用，显著提升了模型在图像分割任务中的性能和效果。该网络的详细架构如图 3-2 所示。

3.2.2 DASPP 模块

DASPP[59]代表“密集的空间金字塔池”。在 DASPP 模块的结构中，亚属性卷积被组合成级联融合操作。膨胀率逐层增加，膨胀率较低的层放置在较低的部位，膨胀率较高的层放置在较高的部位。后续层利用其特性共享信息，与前一层共享信息。这种密集的连接允许更密集的像素利用率。每个属性层将输入与前一层的输出连接起来作为其输入，最终生成由多尺度属性卷积生成的特征图。

与传统的 ASPP 相比，DASPP 利用密集连接在不同膨胀速率的层之间建立互连。每个集合都可以看作是一个不同尺度的卷积核，代表不同的接受域。这一改变引入了特征金字塔的密集化和感受野的扩展，允许更好地识别和整合目标缺陷的语义特征。

DASPP 模块的结构如图 3-3 所示，其原理基于对输入特征图进行空洞卷积处理，接着应用具有 12 的扩张率的多个空洞卷积层。这些层通过调整其视野大小捕获不同尺度的信息。扩张率设置决定了卷积核点之间的距离，从而有效增加了接收域，同时不降低空间分辨率或增加参数数量。接下来，各空洞卷积层的输出特征图会被合并，这样整合了从各层得到的多尺度上下文信息。合并后，通常会施加 3×3 卷积层进一步细化特征，帮助整合多尺度特征成为更一致的特征图。为简化深层网络的训练，还可能引入残差连接，这些连接在反向传播时直接传递梯度，有助于维持信息穿越网络深度的质量。最终输出的特征图可被模型 Improved-DeepLabV3+中的 decoder 利用来进行像素级别分类。

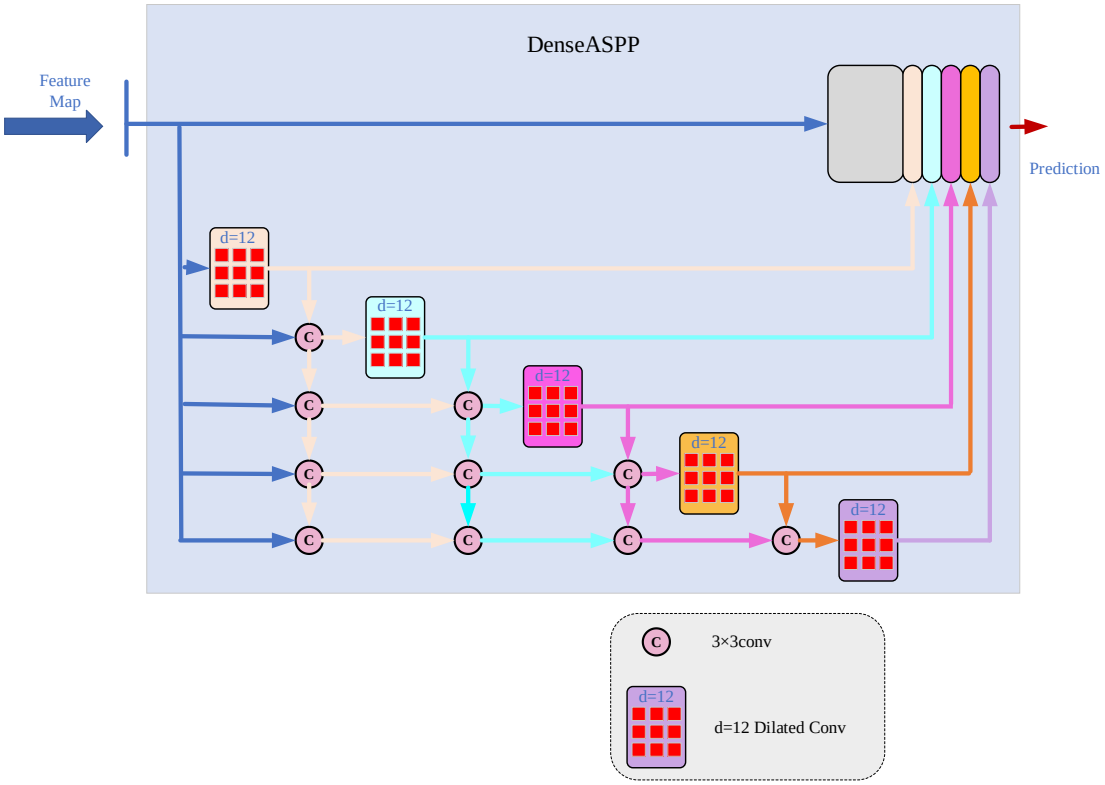


图 3-3 DASPP 模块示意图

3.2.3 三级注意单元

Tanvir Mahmud [60]等人提出的三层注意单元(TAU)是深度学习领域的一个创新。该单元巧妙地整合了三种不同层次的注意力机制，通道注意力(CA)[61]、空间注意力(SA)[62]和像素注意力(PA)[63]，来更加精准地捕获和概括数据的相关上下文信息。这一多层次结构能够在不同的抽象水平上理解图像内容，从而实现更为丰富和动态的特征表示。

具体来说，通道注意力(CA)机制通过分析整个特征图，能够识别并强调包含更多信息的通道，从而为模型提供一种全局的信息视角。这类似于在一组特征中找到最有用的信息，并确保网络能够重点处理这些信息。接着，空间注意力(SA)机制进一步优化这些特征表示，通过关注图像中的局部空间区域，尤其是那些包含关键目标或感兴趣区域的部分，用来提升模型对位置和上下文的感知能力。最后，像素注意力(PA)机制在最细粒度上运作，它逐个像素地评估特征相关性，确保网络能够捕捉到最细微的细节变化，这对于诸如图像分割、目标检测等任务至关重要。

TAU 单元的这个三层架构使其在整个 Improved DeepLabV3+解码器网络中发挥了核心作用。通过频繁地在网络中应用 TAU 单元，模型能够不断地重新校准和提升特征表达的相关性，这样的机制显著提高了网络对复杂图像内容的理解能力。这种自监督的注意力技术，通过不断自我优化来强化特征表达，代表了一种在深度学习中处理高级视觉任务的前沿方法。因此，TAU 单元模块在整个 Improved DeepLabV3+解码器网络中被频繁使用，以重新校准特征并提高特征相关性。

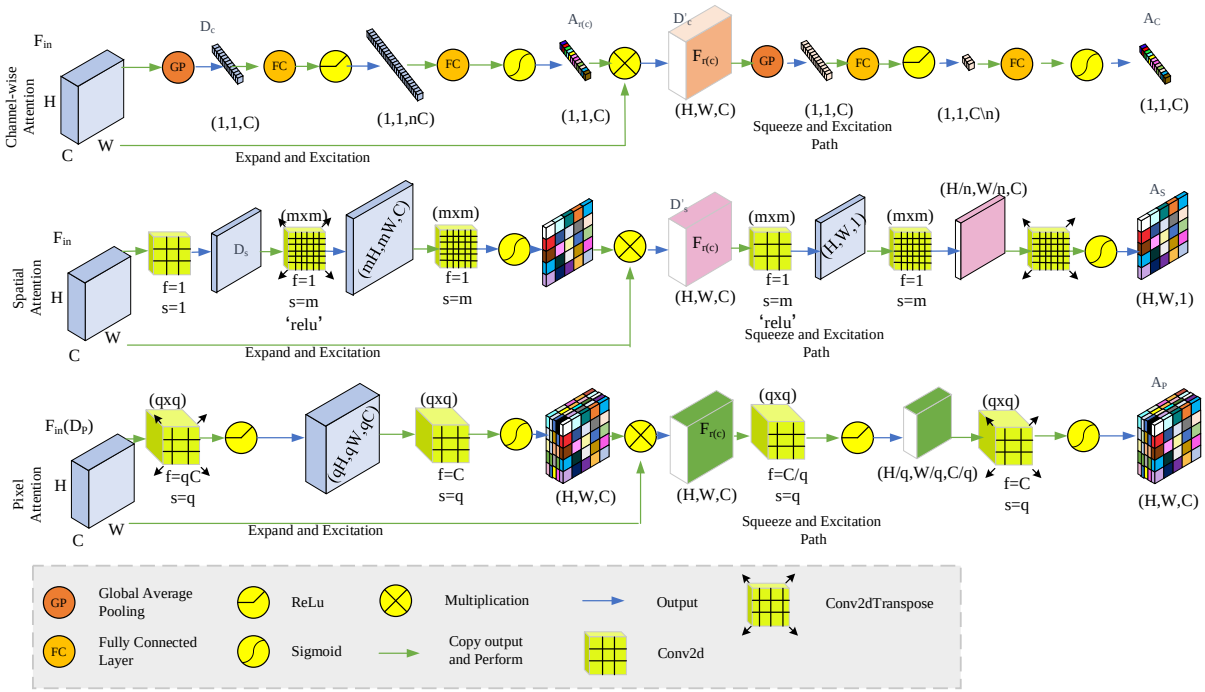


图 3-4 三层注意单元结构

TAU 架构如图 3-4 所示，分为两个阶段:特征重新校准阶段和特征泛化阶段。在每个阶段中，提取预期泛化水平的统计描述，稍后进行处理生成相应的注意图。

设 $F_{\text{in}} \in \mathbb{R}^{H\times W\times C}$ 为输入特征图，其中(H,W,C)分别表示特征图的高度、宽度和通道。其中，通道描述 $D_c \in \mathbb{R}^{1\times 1\times C}$ 是通过对特定通道的像素进行全局平均生成的，空间描述 $D_s \in \mathbb{R}^{H\times W\times C}$ 是通过卷积滤波生成的，输入特征映射 F_{in} 表示像素描述 $D_p \in \mathbb{R}^{H\times W\times C}$ 本身。

然后，将描述子向量 D 投影到高维空间进行特征再标定，然后对原维空间进行恢复，生成再标定注意图 A_r，利用该注意图获得再标定特征图 F_r。该过程有助于在

随后的特征泛化阶段重新分配特征空间，通过锐化有效的代表性特征来更好地泛化特征。它可以表示为：

$$\begin{aligned} F_r &= F_{in} \otimes A_r = F_{in} \otimes \sigma\left(W_R\left(W_E(D)\right)\right) \\ &= F_{in} \otimes \sigma\left(W_R\left(W_E\left(W_D\left(F_{in}\right)\right)\right)\right) \end{aligned} \tag{3-1}$$

其中⊗表示需要进行维度广播运算的元素智能乘法，W_D 表示统计描述子提取器，W_E 表示维度展开滤波，W_R 表示维度恢复滤波，$\sigma(\cdot)$表示 Sigmoid 激活。对于 CA 机制，W_E 和 W_R 由全连接层实现，而对于空间和 PA，使用卷积滤波器。

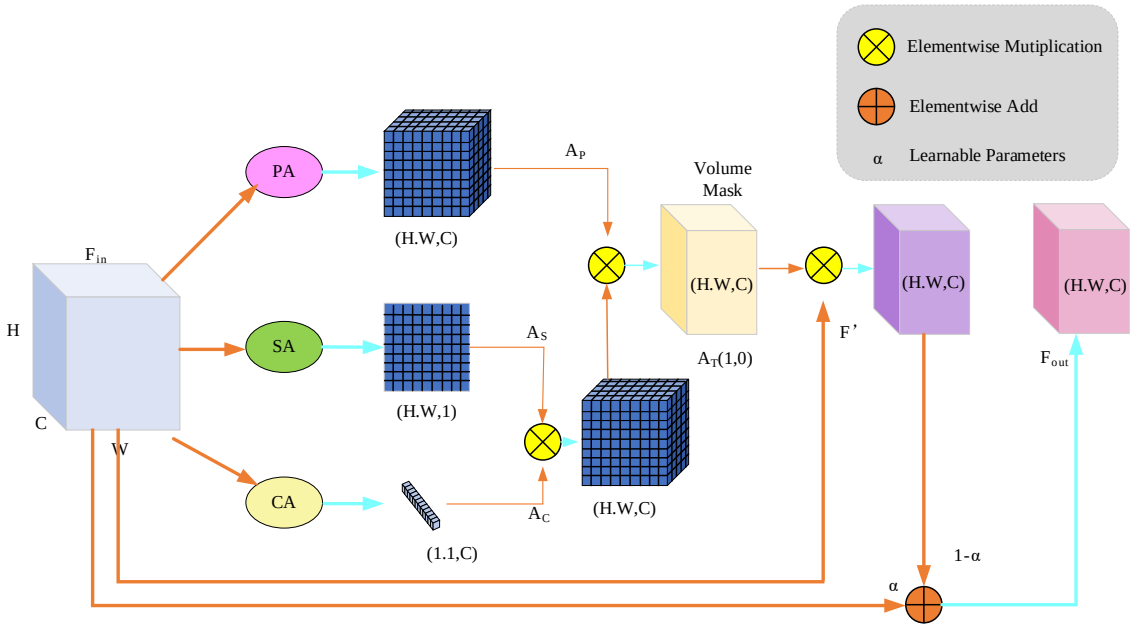


图 3-5 TAU 机制

随后，通过对重新校准的特征空间 F_r 进行挤压和激励操作，进行特征泛化操作，生成有效的注意图 A。在此阶段，将提取的特征描述符 D'投影到较低维空间中，提取最有效的表征特征，然后重建回原维。这种顺序的降维和重建操作提供了一个强调的机会，在减少冗余特征的同时得到广义特征。

因此，生成的注意图 A 为减少冗余特征的影响提供了机会，为有效特征提供了更多的关注，可以表示为：

$$A = \sigma\left(W_{R'}\left(W_S\left(D'\right)\right)\right) = \sigma\left(W_{R'}\left(W_S\left(W_{D'}\left(F_r\right)\right)\right)\right) \tag{3-2}$$

其中，W_E、W_R 分别表示相应的压缩滤波和恢复滤波，W_D 表示统计描述符提取器。因此，生成了三个层次的注意图，即 CA 映射 $A_c \in \mathbb{R}^{1\times1\times C}$，SA 映射

$A_S \in \mathbb{R}^{H\times W\times 1}$，PA 映射 $A_P \in \mathbb{R}^{H\times W\times C}$ 。如图 3-5 所示的 TAU 生成了有效的体积三重注意力掩码 A_T，整合了所有三个 maps，可以用(3-3)表示:

$$A_T = A_P \otimes \left(A_S \otimes A_C\right) \tag{3-3}$$

然后利用这个累积的注意掩码 A_T 将输入特征图 F_{in} 变换为 F'，增强兴趣区域，最后将输入特征图和变换后的特征图加权相加，得到输出特征图 F_{out}，总结如下:

$$F^{'} = F_{in} \otimes A_T \tag{3-4}$$

$$F_{\text{out}} = T\left(F_{\text{in}}\right) = \alpha F_{in} + (1-\alpha)F^{'} \tag{3-5}$$

其中 $T(\cdot)$ 表示提出的三层注意机制，α 是一个可学习的参数，与其他参数一起通过反向传播算法进行优化。

3.2.4 挤压激励网络

SENet 的核心思想是通过引入“Squeeze”和“Excitation”操作，以自动调节每个通道上特征映射的权重[64]。通过执行全局平均池化(Global Average Pooling)操作，SENet 实现了对每个通道特征映射的压缩，即计算每个通道特征图的平均值，从而获取每个通道的全局统计信息。

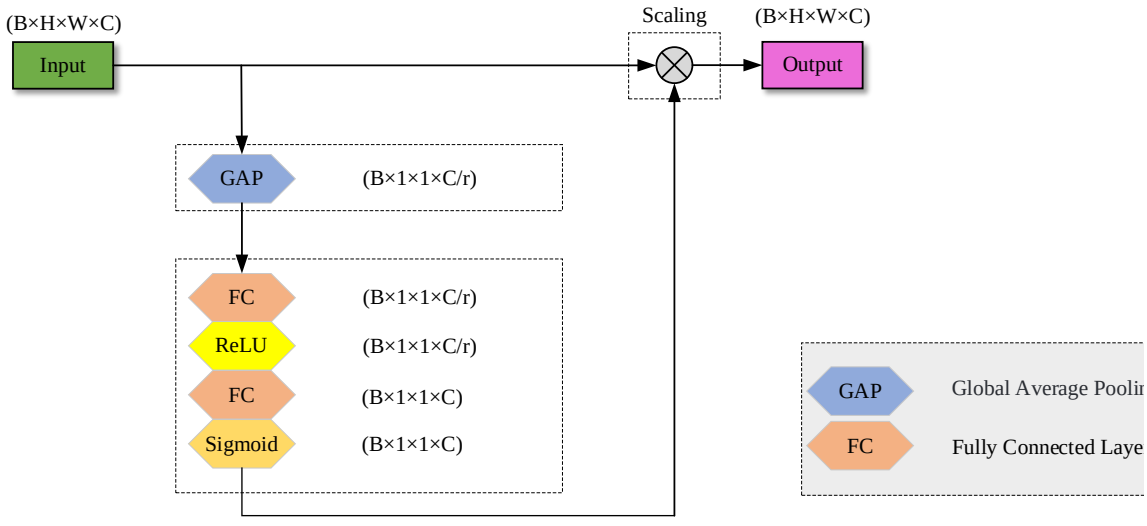


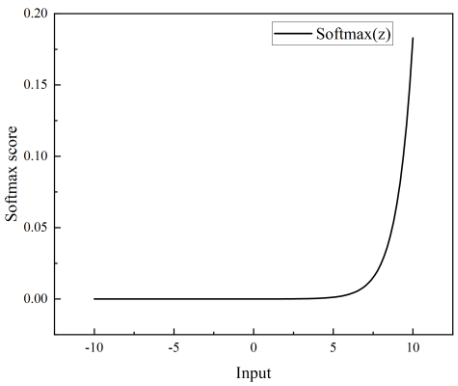
图 3-6 SENets 架构

SENet 使用一个小型全连接神经网络来学习通道权重，如图 3-6 所示，SENet 网络包括一个隐藏层和一个 Sigmoid 激活函数。该隐藏层接收 Squeeze 操作的输出，并

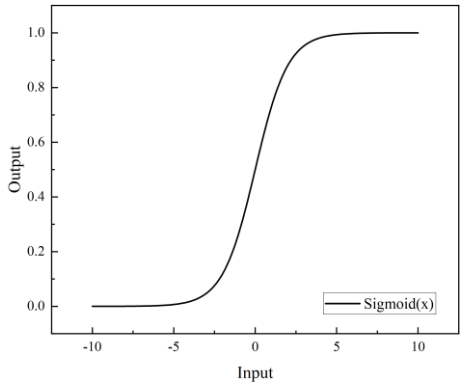
产生用于调整通道权重的激活权重。这些激活权重表示每个通道对于特定任务的重要性。最后，SENet 将 Excitation 操作的输出与原始特征映射相乘，再按通道重新加权特征映射。这个尺度操作有助于强化重要通道的信息，减少不重要通道的影响。挤压激励网络(SENets)为卷积神经网络(CNN)引入了一种构件，旨在增强各通道间的依赖关系，而计算成本几乎无增加。在汇总 CNN 所有输出通道的信息前，它通过先识别每个输入通道上的空间特征来实现。传统上，网络在形成输出特征图时对各通道权重一视同仁。SENets 通过融入一个内容识别机制，为每个通道实施自适应权重分配，致力于转变这一处理过程。在最简单的形式中，这可能需要为每个通道提供一个单一参数和一个线性标量，代表每个通道的相关程度。通过将特征映射浓缩为单一数值，它们可以首先获得对每个通道的广泛理解。结果，产生了一个大小为 n 的向量，其中 n 是卷积通道的数量。然后将其输入一个双层神经网络，该网络会产生一个相同大小的输出向量。现在，这 n 个值可以作为权重应用到原始特征图中，根据每个通道的重要性来确定其大小。

3.2.5 激活函数

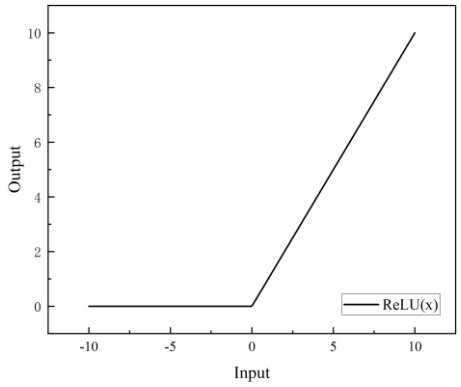
在本章节中，采用了 Softmax 激活函数[65]，与图 3-7(b)、3-7(c)、3-7(d)中使用的 Sigmoid[66]、ReLU[67]和 Tanh 激活函数[69]不同。Softmax 函数的核心作用是将输入向量转化成概率分布，其中每个元素的输出代表了该元素属于某一特定类别的可能性。在处理多类别分类问题时，Softmax 在 Improved-DeepLabV3+模型的输出层应用，以提供各类别的预测概率。



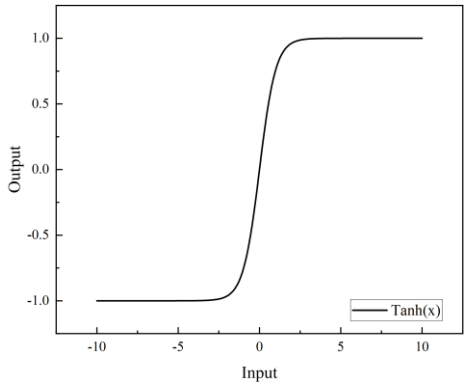
(a) Softmax 函数



(b) Sigmoid 函数



(c) ReLU 函数



(d) Tanh 函数

图 3-7 激活函数

这个函数的作用是将一个实数向量转换成一个概率分布。通常在网络的最终层使用 Softmax 激活函数，目的是产生每个像素点对应各类别的概率分布。Softmax 函数的标准公式为：

$$\text{softmax}(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \tag{3-6}$$

其中，z 是一个包含 K 个类别的分数或逻辑回归的向量，z_i 是向量 z 中第 i 个元素的值。首先。对向量 z 中每个元素应用指数函数，得到 e^{z_i}；接着，计算这些指数值的总和，即 $\sum_{j=1}^{K} e^{z_j}$；最终每个指数值会被其总和所除。这样对于向量 z 中每个元素 z_i，得到一个介于 0 到 1 之间值，这些值的总和为 1。

鉴于 Softmax 函数的输出之和等于 1，故此输出可被理解为一个概率分布。这对于分类问题尤其有用，因为它允许模型估计输入属于每个可能类别的概率；由于使用了指数函数，Softmax 会放大输入值之间的差异。即使是非常小的输入值差异，经过 Softmax 处理后，也可能导致输出概率的显著不同；在实际应用中，直接计算指数可能会导致数值不稳定。因此，通常会使用一些技巧来稳定计算，例如在计算之前从每个分数中减去最大分数。

Softmax 应用于最后的卷积层的输出。模型针对每个像素位置输出一个与待识别类别数量相等的向量长度。这些向量的每个元素都会独立地通过 Softmax 函数进行处理，使之被转换成概率分布。这意味着对于每个像素位置，将得到一个表示该像素属于每个类别的概率的向量。

3.2.6 损失函数

DeepLabV3+是一种专为语义图像分割设计的深度学习模型，其主要采用交叉熵损失函数(Cross-Entropy Loss)[69]来评估性能。而 Improved-DeepLabV3+在此基础上，引入了 Dice Loss[70]与交叉熵损失的结合，以提升图像分割效果。这种方法使得模型能够同时兼顾全局的准确性(通过交叉熵损失)和处理类别不平衡或划分难度高的区域(通过 Dice Loss)。

交叉熵损失函数用于评估模型预测的概率分布与实际标签的概率分布之间的不同。对于多类分类问题，交叉熵损失特别有效，因为它能够直接针对每个类别的预测概率进行优化。交叉熵损失可以表示为:

$$L = -\sum_{i=1}^{N} y_i \log(p_i) \tag{3-7}$$

其中 N 是类别，y_i 是真实标签的独热编码，p_i 是模型预测的概率。

Dice Loss 是基于 Dice 系数的，这是一个评估两个样本相似度的度量，非常适合于解决类别不均衡的问题。Dice Loss 的计算公式为：

$$DiceLoss = 1 - \frac{2 \times |Y \cap P|}{|Y| + |P|} \tag{3-8}$$

其中 $|Y \cap P|$ 表示真实标签 Y 和预测标签 P 之间的交集的大小，而 $|Y|$ 和 $|P|$ 分别是真实标签和预测标签中的元素数量。Y 是真实标签，P 是预测结果。该公式可以被理解为两次预测和真实值之间重叠部分的大小除以它们各自的大小之和。它的值范围从 0(没有重叠)到 1(完全重叠)。

组合交叉熵损失和 Dice Loss 通常是通过简单地加权两者来完成的：

$$\text{Total}Loss = \alpha \times CrossEntropyLoss + \beta \times FocalLoss \tag{3-9}$$

其中，α 和 β 是用来平衡两个损失的权重。

通过将交叉熵损失与 Dice Loss 或 Focal Loss 结合使用，可以使模型在不同方面(如整体准确性、类别不平衡处理、关注难分类样本等)取得更好的平衡。

3.3 实验结果与分析

3.3.1 实验平台

所有实验均在统一平台上进行，如表 3-1 所示：

表 3-1 实验平台

实验设备	型号
设备	戴尔游匣 G15
平台	Unbuntu 20.04
CPU	Intel(R) Core(TM) i7-10875H
GPU	GeForce RTX 2060
System memory	16GiB
SSD	Samsung SSD 970 EVO Plus 1TB

在本研究中，模型训练的相关参数被设定如下：batchsize 为 8，采取不冻结网络的策略，并设定初始学习率为 0.01。作为学习率调整策略，引入了余弦退火(Cos)算法[71]，它在训练过程中周期性地调整学习率，可以有效避免训练结果陷入局部最优并增加寻找全局最优解的可能性，从而促进网络的良好收敛并防止过拟合。动量设定为 0.9，以适应学习速度的变化。损失函数选用的是 TotalL*oss*，旨在解决前景与背景不平衡的问题。超参数的具体设置详见表 3-2。

表 3-2 超参数表

Super Parameter	Values
weight in the loss function	0.7
batch size	8
warmup_bias_lr	0.1
epochs	300
initial learning rate	0.01
final learning rate	0.2
momentum	0.937
weight_decay	0.0005
image-size	256*256

3.3.2 评价标准

为验证所提方法的有效性，采用了 DICE 系数、像素精度、模型大小、平均交并比以及平均像素精度进行了评估，在使用这些评价标准时，需要引入混淆矩阵，

假定一定有 k+1 类(包括 k 个目标类和 1 个背景类)，p_{ij} 表示本属于 i 类却预测为 j 类的像素点总数，具体地，p_{ii} 表示 True Postives，p_{ij} 表示 False Positives，p_{ji} 表示 False Negatives，则有：

像素精度表示的是被正确标记的像素数量占总像素数量的比例，具体可参照公式 3-10 进行计算：

$$PA=\frac{\sum_{i=0}^{k}p_{ii}}{\sum_{i=0}^{k}\sum_{j=0}^{k}p_{ij}} \tag{3-10}$$

类别平均像素准确度是通过首先计算每个类别中正确分类的像素点与该类别像素点总数的比例，接着求这些比例的平均值来获得的一个指标，详见公式 3-11：

$$MPA=\frac{1}{(k+1)}\sum_{(i=0)}^{k}\frac{p_{ii}}{\sum_{j=0}^{k}p_{ij}} \tag{3-11}$$

类别平均交并比是指每个类别被正确分类像素的比例，如公式 3-12 所示：

$$MIoU=\frac{1}{(k+1)}\sum_{(i=0)}^{1}\frac{p_{ii}}{\sum_{j=0}^{k}p_{ij}+\sum_{j=0}^{k}p_{ji}-p_{ii}} \tag{3-12}$$

在上式中，当 i=m，p_{mm} 表示 True Positives，$\sum_{j=0}^{k}p_{mj}$ 中表示属于 m 类却预测为其他像素点数的计数中包含了 p_{mm}，$\sum_{j=0}^{k}p_{ji}$ 中表示属于其他类别却预测为第 m 类的像素点数，也包含了 p_{mm}，计算两次所以要减去一个 p_{ii}。

3.3.3 对比实验与结果分析

本研究在相同数据集上比较了 Improved-DeepLabV3+与常用分割模型如 BASNet[72]、UNet[73]、TransFuse[74]和 DeepLabV3+，以验证改进模型的卓越分割性能。使用 IoU(交并比)、PA(像素准确率)、mIoU(平均交并比)、mPA(平均像素准确率)和 F1_Score 作为评价指标，从定性和定量角度分析了不同分割模型在金属表面图像分割的效果。

(1) 定量分析

如表 3-3 所示，Improved-DeepLabV3+在处理多类缺陷分割时，特别是在裂纹(Cr)和点蚀面(PS)方面，显示了其在 IoU 和 PA 指标上相较于其他模型的明显优势。同时，对于整个图像的语义分割，Improved-DeepLabV3+在平均 Intersection over

Union (mIoU)、平均 Pixel Accuracy (mPA)方面具有明显的优势。这些指标显著地提升，验证了本文提出的改进策略的有效性。

表 3-3 不同分割网络实验结果对比

模型	Background		Cr		In		Pa		PS		RS		Sc		mixed	
	IoU	PA	IoU	PA	IoU	PA	IoU	PA	IoU	PA	IoU	PA	IoU	PA	IoU	PA
BASNet	0.99	0.99	0.52	0.61.	0.73	0.78	0.91	0.87	0.64	0.72	0.73	0.78	0.75	0.80	0.48	0.54
UNet	0.99	0.99	0.74	0.79	0.81	0.78	0.91	0.89	0.79	0.83	0.82	0.85	0.79	0.81	0.71	0.75
TransFuse	0.99	0.99	0.72	0.81	0.83	0.76	0.88	0.87	0.81	0.81	0.81	0.85	0.74	079	0.69	0.74
DeepLabV3+	0.99	0.99	0.73	0.78	0.81	0.75	0.91	0.90	0.81	0.83	0.82	0.85	0.78	0.80	0.70	0.73
Improved-DeepLabV3+	0.99	0.99	0.84	0.94	0.84	0.85	0.95	0.91	0.85	0.87	0.84	0.86	0.85	0.88	0.81	0.83

表 3-4 的数据表明了改进版 DeepLabV3+(Improved-DeepLabV3+)在金属表面缺陷分割任务上的表现。相比于原始 DeepLabV3+，TransFuse，UNet 和 BASNet，Improved-DeepLabV3+的模型参数减少至分别为 74.8%，68.23%，61.6%和 62.8%，这意味着它能够在使用更少计算资源的同时，实现更优的性能。Improved-DeepLabV3+相较于 DeepLabV3+原版模型，Dice 系数提升了约 0.52%，像素准确度(PA)提升了约 0.32%，平均像素准确度(mPA)提升了约 1.23%，平均交并比(mIoU)提升了约 2.10%，参数数量减少了约 25.15%。这些提升百分比表明，Improved-DeepLabV3+在 Dice 系数和 mIoU 上有轻微但重要的提升，而在参数效率上则有显著的优化。减少了超过四分之一的参数数量，对于提高计算效率和模型部署的灵活性尤为重要。

表 3-4 不同分割网络实验结果对比

项目	BASNet	UNet	TransFuse	DeepLabV3+	Improved
Dice	0.9658	0.9759	0.9715	0.9806	0.9857
pa	0.9262	0.946	0.945	0.9585	0.9616
mpa	0.9118	0.9366	0.9215	0.9426	0.9542
miou	0.891	0.905	0.9047	0.915	0.9342
Params	97M	98.9M	89.4M	81.5M	61M

在金属表面图像分割数据集上，对 BASNet、UNet、TransFuse、DeepLabV3+以及 Improved-DeepLabV3+模型进行了训练和测试。训练过程如图 3-8 所示，其中横

坐标表示训练的轮次，纵坐标则显示了损失值。与其他分割网络相比，Improved-DeepLabV3+在训练仅 100 轮后就达到了约 0.1 的损失值，并保持了该水平的稳定性。相对而言，DeepLabV3+在同样轮数的训练后达到了 0.13 的损失值，BASNet 需要 200 轮训练以降至 0.18 的损失值，UNet 在 100 轮训练后达到 0.1 的损失值，而 TransFuse 在 150 轮训练后损失值为 0.20。这表明 Improved-DeepLabV3+模型的性能表现较为出色。

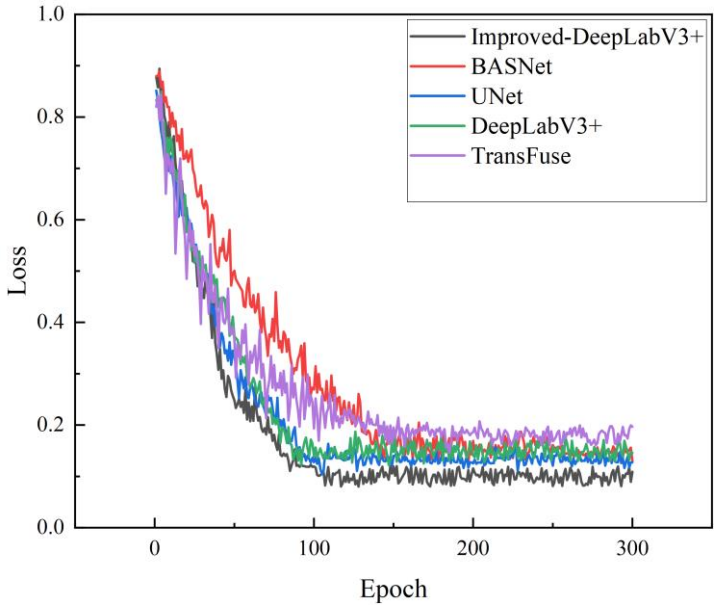


图 3-8 不同分割模型的训练和验证过程

图 3-9 展示了 Improved-DeepLabV3+模型与其他分割网络相比的性能优势。相较于原始 DeepLabV3+模型，在验证 Dice 得分上提高了 0.52%；与 UNet 模型相比，提高了 1.0%；与 TransFuse 模型相比，提高了 1.46%；与 BASNet 模型相比，提高了 1.63%。

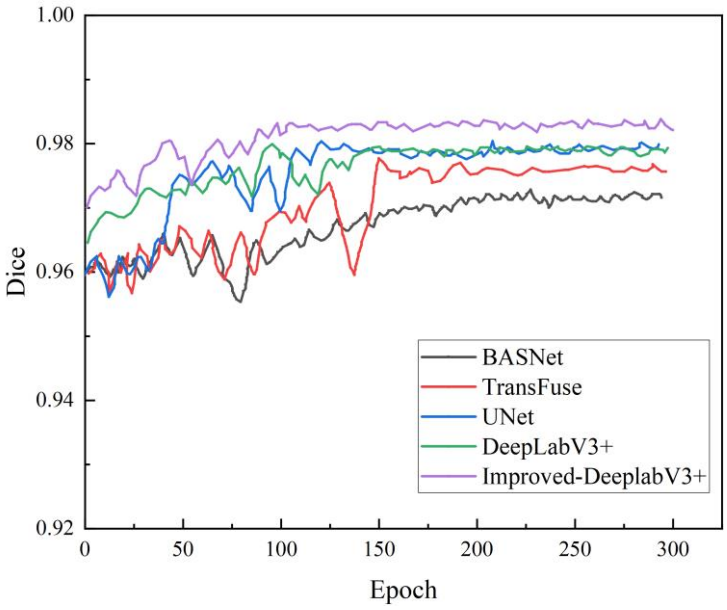


图 3-9 不同分割模型的验证 Dice 分数

(2) 定性分析

图 3-10 通过视觉对比直观展示了 DeepLabV3+、BASNet、UNet、TransFuse 以及 Improved-DeepLabV3+在七种不同类型的金属表面缺陷分割任务上的性能。结果清楚地显示，Improved-DeepLabV3+在金属表面缺陷分割方面显著优于原始版本，这种性能提升主要体现在更精确的缺陷边界识别和在复杂背景下的识别准确性。

图 3-10 中的对比揭示，在处理复杂分割场景时，UNet 和 DeepLabV3+模型表现出了过度分割的倾向。这一现象可能根源于模型在区分特征能力上的局限，导致它们将邻近的缺陷错误地识别为单一实体。另一方面，TransFuse 模型在特定场合下出现了明显的分割不足，未能充分识别所有缺陷，特别是体积较小或对比度不明显的缺陷，揭示了其在捕捉精细特征上的局限性。

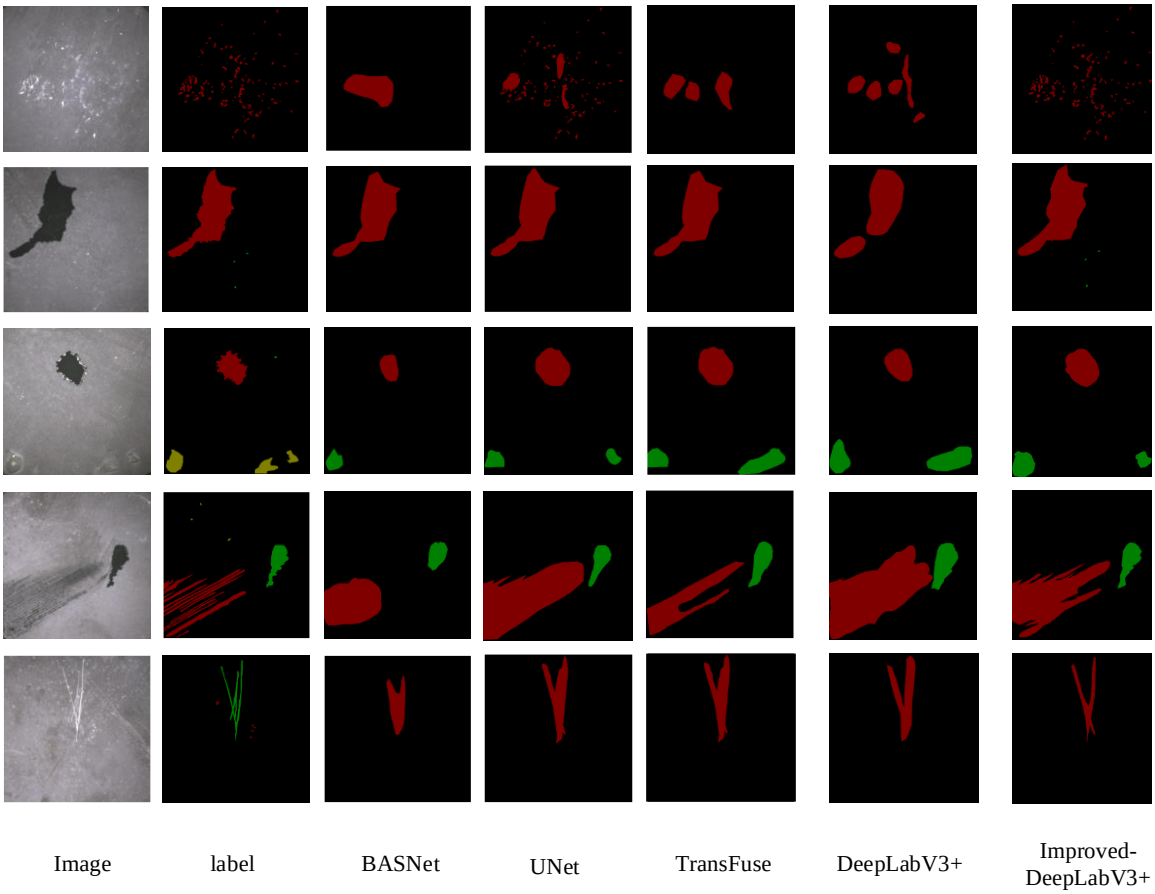


图 3-10 不同分割网络的测试结果

在评估各个模型的绩效时，Improved-DeepLabV3+模型不仅在图像分割准确率方面超越了其它模型，而且还证明了在不显著增加计算负担的情况下，通过细致的网络架构调整与训练策略的优化，能够有效提升模型性能。这一发现强调了即便在计算资源有限的条件下，依然可以通过智能化的设计实现高效的缺陷分割，这对于推动工业自动化及质量控制系统的进步具有显著意义。改进模型在处理具有高度变异性和复杂性的图像方面所表现出的鲁棒性，为金属表面缺陷的自动检测与分类提供了一条有效途径，在实际工业应用中推动了计算机视觉技术的研究和应用前沿。

3.4 本章小结

本章节对金属表面缺陷分割问题进行了深入探讨，并详细分析了传统语义分割模型在处理此类图像时遇到的性能瓶颈。通过引入改进型模型——Improved-DeepLabV3+，研究取得了显著进步。该模型融合了 DeepLabV3+的结构、三级注意力单元(TAUs)和挤压激励块(SENets)，大幅提升了分割效率和准确性。密集空洞空

间金字塔池化(DASPP)的应用代替了原始的 ASPP 模块，有效解决了稀疏采样带来的信息丢失问题。模型通过引入 TAUs 和 SENets，增强了对关键特征的捕获能力。

本研究的贡献主要包括：

1、在解码器中加入 TAUs 和 SE 块，该结构创新地模拟了通道间的相互作用，自适应地调整通道的特征响应，从而增强了模型在分割任务上的效率和精确度。

2、用DASPP替换传统的ASPP，通过在不同扩张率的卷积层间建立密集连接，形成了一个更为细致和多层次的特征表示。这一改进加强了模型的特征提取能力，为识别和整合金属表面的缺陷提供了更大的接受野。

3、通过在解码器末端融入残差细化模块，本研究进一步提升了对图像边缘信息的捕捉能力，优化了边界细节的处理，并明显增强了模型在金属表面图像分割方面的精度。模型训练过程中引入了具有多种状态特征的缺陷图像，以增强模型的鲁棒性。综合比较和评估证明，Improved-DeepLabV3+在高精度金属表面缺陷分割方面展现了其显著优势。

第 4 章 金属表面缺陷目标检测技术分析

4.1 本章方法概述

本章介绍了一种改进的 IE-YOLOv5 模型，旨在提高工业环境下金属零件缺陷检测的效率和准确性。通过在 YOLOv5 的 Neck 模块中采用轻量级卷积方法 GSConv 替代传统卷积，并结合 GSConv 与联邦学习特征融合技术，设计了一种新型的细颈结构 FF-Slim-Neck，以此改进 YOLOv5 的 Neck 层。此外，引入了无参数的 PSA 注意力模块，旨在优化网络结构调整的复杂度，同时提升对关键信息特征的提取能力。通过对主干网络的全面优化与调整，显著降低了模型参数数量，进而提升了网络的轻量化水平。

4.2 基于 YOLO 的缺陷识别算法

YOLOv5 的网络结构由四个主要部分组成：输入层、主干网络(Backbone)、颈部(Neck)和检测头(Head)[75]，如图 4-1 所示。图像在进入网络之前首先进行预处理，尺寸调整为 640×640×3 以符合模型输入标准。随后，图像通过 CSPDarkNet53 构成的深度卷积神经网络，实现特征提取。该网络利用多尺度卷积核捕捉信息，并通过跳跃连接增强了特征传递的效率；再将所得到的特征通过特征金字塔网络(FPN)和路径聚合网络(PAN)进行融合，以集成不同层级的语义和细节信息，从而提升对不同大小物体的检测能力。然后，利用一系列预定义的多种比例和尺寸的 Anchor Boxes，来实现图像中不同位置预测物体的位置和类别的特征提取。YOLOv5 的检测头在三个不同分辨率的特征图上工作，以便检测各种尺寸的物体。在模型的每个特征图层内，针对每个格子预测多个边界框及其对应的类别概率。边界框的参数包括框的中心点坐标(x，y)、宽度高度(w，h)、以及对象存在的置信度评分。此外，模型还会为每个边界框预测其属于各个类别的可能性。为解决同一个物体被多个边界框识别的情况，YOLOv5 应用了非极大值抑制(NMS)技巧，通过置信度来筛选边界框，去除那些重合且置信度较低的边界框。最后，模型将输出每个识别物体的边界框坐标、类别预测及置信度评分

在图 4-1 中 CBL 表示基本卷积单元；CSP_x 注释特征提取模块；Concat 表示连接模块;Up-sample 为上采样模块；SPP 是高效空间池化的空间金字塔池化层;输出的

特征映射的维度分别为 20×20×255、40×40×255 和 80×80×255，分别表示长度、宽度和深度。这些维度对应于网络中不同尺度的待检测目标。

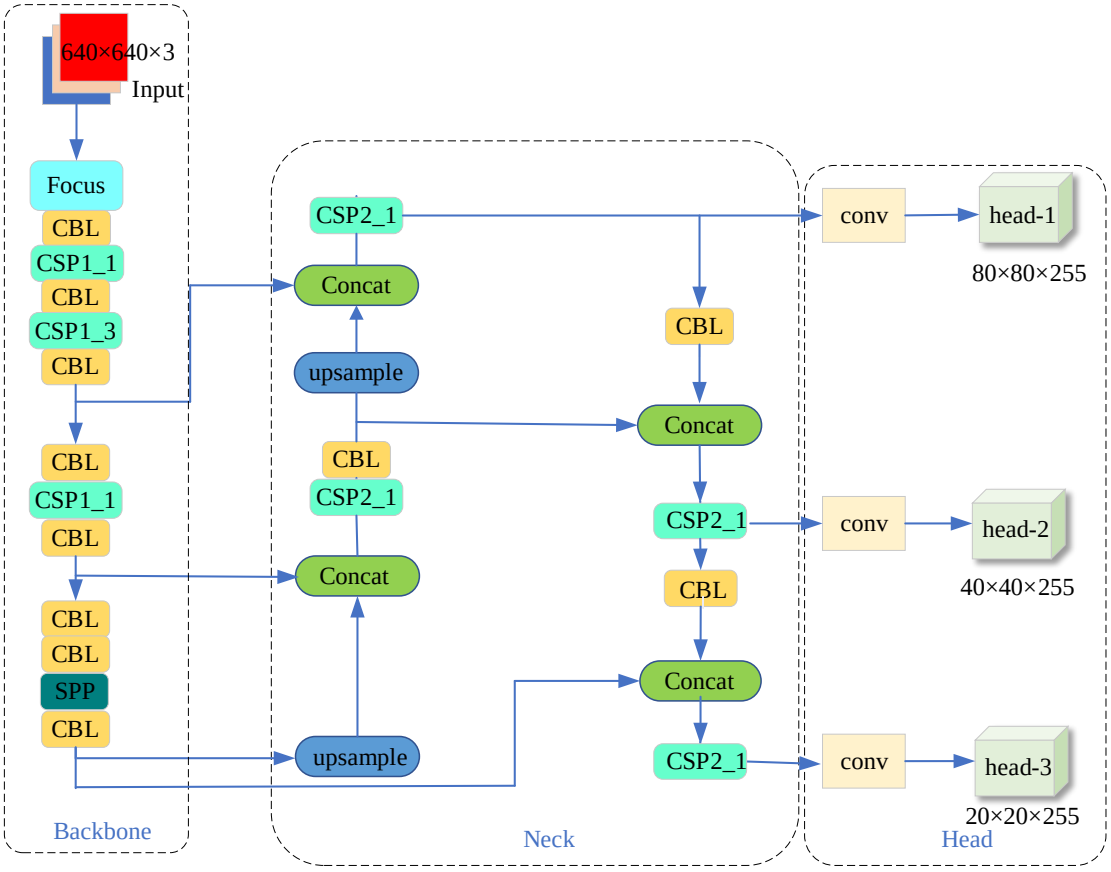


图 4-1 YOLOv5 结构图

4.2.1 YOLOv5 目标检测网络的改进

因为在原始的 YOLOv5 中，输入图像在 Neck 模块中经过卷积和上采样处理，使得空间信息逐渐传递到通道中。当压缩特征图的空间维度和扩展通道时，会造成部分语义信息的丢失。所以可以在 Neck 模块中引入 GSConv 结构来解决这个问题。

于此同时，在 GSConv 中引入 FF-Slim-Neck 来平衡检测精度、检测速度和计算成本之间的权衡，这有助于在保持整体效率的同时解决检测小物体的挑战。

为了进一步提高检测算法的性能，在图像特征提取、特征聚合与融合、网络轻量化设计等方面进行了各种改进。颈部结构中原有的 FPN+PAN 特征融合网络被 FF-Slim-Neck 和 GSConv 组合特征融合网络所取代，以便增强关键信息提取能力，提高了特征融合效率。

此外，Head层引入了PSA关注机制，利用三维权重的概念对语句进行评估，减少了推理时间，提高了网络的检测速度。

基于YOLOv5的改进算法框架IE-YOLOv5如图4-2所示。

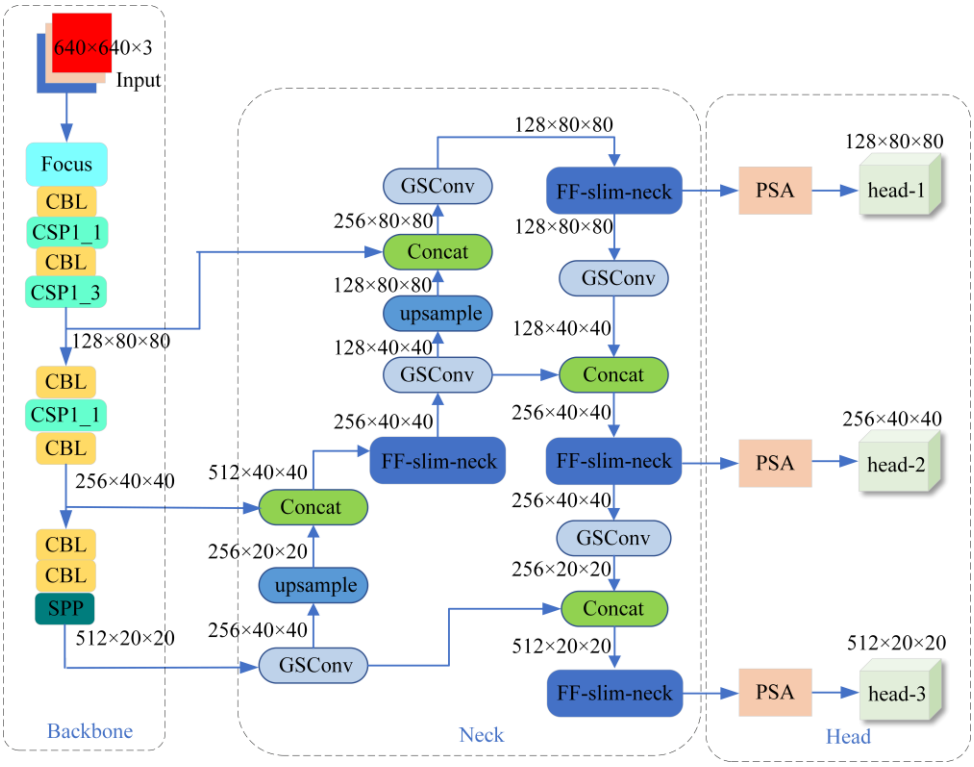


图4-2 IE-YOLOv5网络结构

4.2.2 特征融合模块

在原始的YOLOv5中，输入图像在Neck模块中经过卷积和上采样处理，使得空间信息逐渐传递到通道中。当压缩特征图的空间维度和扩展通道时，会造成部分语义信息的丢失。可以在Neck模块中引入GSConv结构来解决这个问题。

GSConv可以有效地保留每个信道的隐藏连接[76]。给定大小为$C_1\times H\times W$的输入特征图，GSConv模块应用两种类型的卷积:标准卷积(SC)和深度可分离卷积(DSC)[77]。SC使用大小为$C_2/2\times C_1\times H_1\times W_1$的卷积核对每组输入特征图进行卷积，得到大小为$C_2/2\times H\times W$的特征图。然后将DSC用于特征图分割和通道融合，生成通道仍然在$C_2/2\times H\times W$的输出特征图。使用Concat操作将SC和DSC的特征图融合在一起，得到大小为$C_2\times H\times W$的特征图。最后，使用Shuffle模块进行特征重组，生成大小为$C_2\times H\times W$的输出特征映射，如图4-3所示。GSConv模块允许SC生成的信息

渗透到 DSC 生成的信息的各个部分。处理后的图像特征包含较少的冗余和重复信息，减少了压缩的需要，减少了参数和浮点运算的次数，从而提高了特征提取的效率。

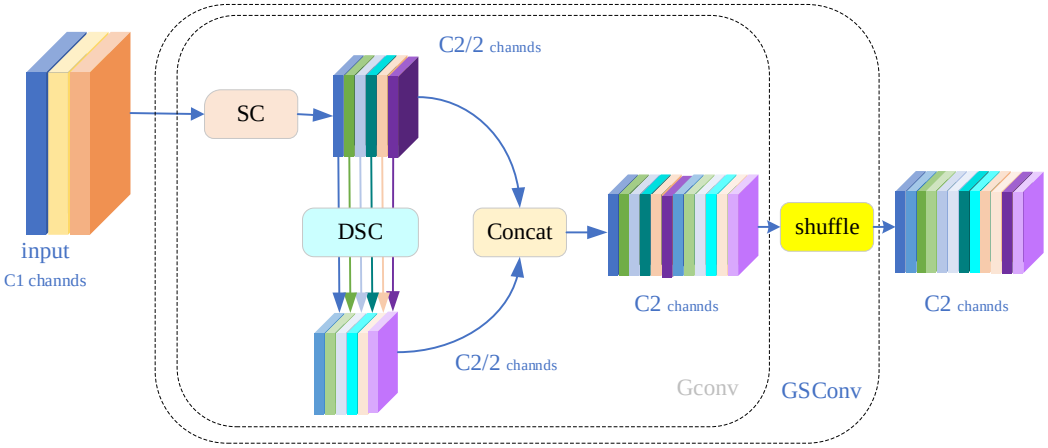


图 4-3 GSConv 结构

引入 GSConv 后，网络深度增加，检测时间明显增加，如图 4-4 所示。更深的网络可以引入更多的数据流阻力，显著增加检测时间。为了解决这个问题，所提出的 FedFusion Slim Nck (FF-Slim-Neck)结构的设计灵感来自于 DensNet[78]、VoVNet[79]和 CSPNet[80]学习能力的广义方法。

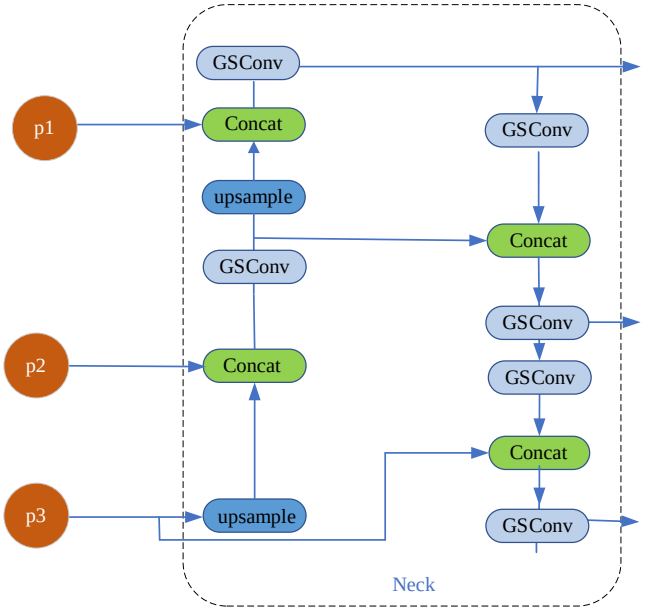


图 4-4 neck 模块单独使用 GSConv

结合 GSConv 的 FF- Slim-Neck 结构可以平衡网络深度和效率之间的权衡。这种设计允许网络提取重要的特征，同时将计算成本和检测时间保持在合理的水平。

FF-Slim-Neck 结构将 GSConv 的优点与这些广义 CNN 学习方法的原理相结合，提高了检测金属零件缺陷的性能。

在 IE-YOLOv5、FCConv 中引入了一种基于融合的联邦学习算法(FedFusion)[81]。图 4-5 显示了 FCConv 的卷积过程。通过局部特征提取器 E_l 和全局特征提取器将输入图像 X 变换到两个特征空间，分别得到特征映射 $E_l(x)$，$E_g(x) \in \mathbb{R}^{C\times H\times W}$。随后，使用 1×1 卷积将 E_l 和 E_g 嵌入到融合特征空间中。

$$F_{conv}\left(E_l(x), E_g(x)\right) = W_{conv}\left(E_l(x) \| E_g(x)\right) \tag{4-1}$$

其中 $W_{conv} \in \mathbb{R}^{2C\times C}$ 表示学习到的权矩阵，||表示沿通道轴连接特征映射的操作。

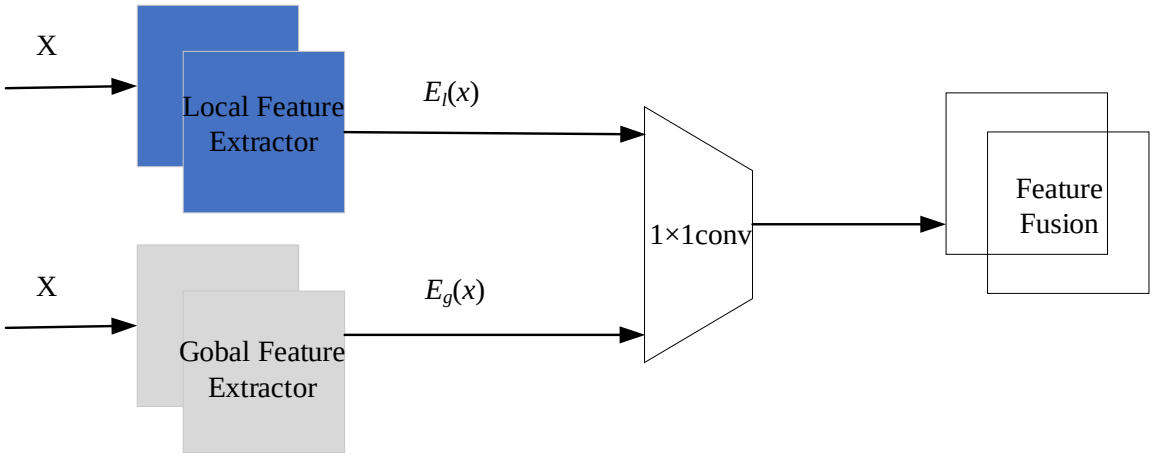


图 4-5 FCConv

颈部模块采用 GSConv 和 FedFusion 设计 FF- Slim-Neck 结构。采用一步聚合的方法设计跨阶段部分网络 FC 模块，降低了检测过程的计算复杂度和推理时间。如图 4-6 所示，FF-Slim-Neck 结构在特征提取阶段促进了局部和全局模型的特征融合。输入特征图为 C_1×H×W。它通过 Conv 模块生成 $C_1/2$×H×W 的输出特征映射。随后，GSConv 模块生成一个输出特征图 $C_2/4$×H×W。特征融合通过 FCConv 模块进行，产生输出特征图 $C_2/2$×H×W。最后，将 Conv 和 FCConv 模块的输出连接起来，得到 C_2×H×W 的输出特征图。

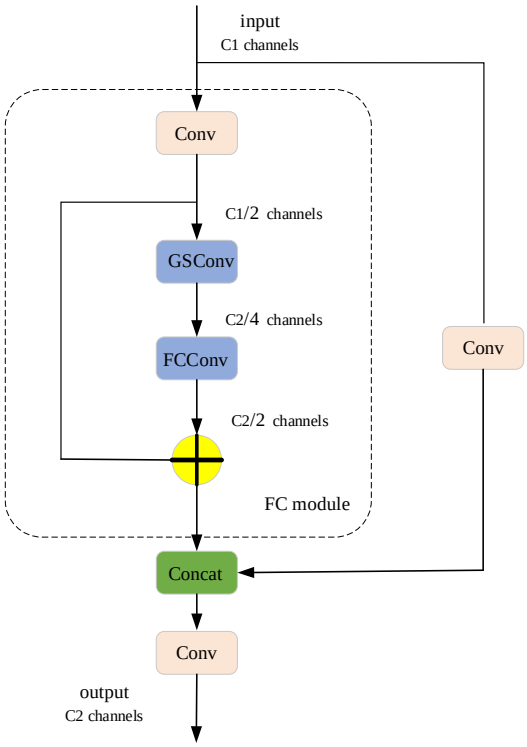


图 4-6 FF-Slim-Neck 结构

4.2.3 注意力学习模块

被称为“挤压-激励模块”(SE)的注意力机制[43]通过全局最大池化(Global Max Pooling, GMP)捕获上下文信息，以确定不同通道的重要性。然而，SE 注意机制对通道或空间位置上的所有神经元都是平等的，这限制了它在计算真正的 3D 权重时的有效性。此外，SE 中的权重计算需要大量的计算，增加了网络的计算成本，阻碍了其从神经元中提取更多信息的能力。

为了克服这些挑战，引入了无参数空间注意机制(PSA)，如图 4-7 所示，其中，W 和 H 代表特征图的宽度与高度，而 C 代表通道数。输入特征图的尺寸为 W×H×C。图中不同的颜色用以标识特征图上不同的通道、空间位置或点，每个点对应一个独立的标量值。PSA 区分不同的位置，并单独处理所有通道，为每个神经元分配唯一的权重。因此，PSA 提供了 3D 权重，超越了传统 1D 和 2D 注意机制的能力。此外，PSA 将现有的注意力模块整合到每个块中，增强了领导层的输出。

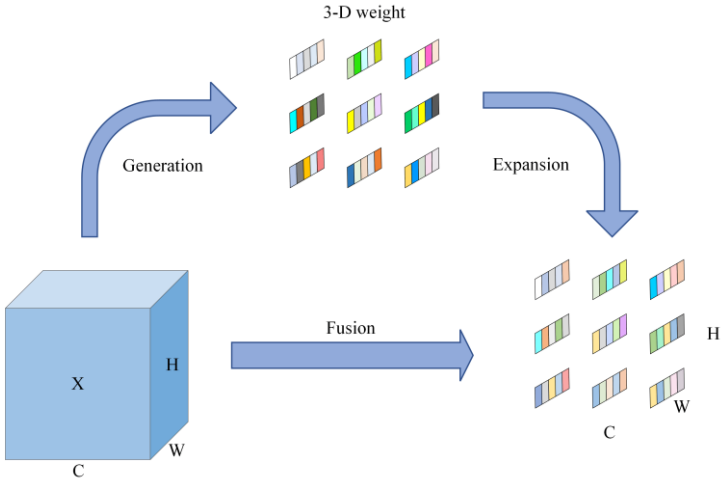


图 4-7 PSA 注意力机制原理图

为了在三维权值中同时包含空间维度和通道维度，并有效地区分目标神经元，需要测量一个将目标神经元与其他神经元区分开来的值，从而使网络能够学习到更多的判别特征。因此，每个神经元的能量函数定义如下:

$$e_t(w_t,b_t,y,x_i)=\frac{1}{N-1}\sum_{i=1}^{N-1}\left(-1-\left(w_t x_i+b_t\right)\right)^2+\left(1-\left(w_t t+b_t\right)\right)^2+\lambda w_t^2 \tag{4-2}$$

其中，w_t和b_t分别表示目标神经元转换的权重和偏置，t指目标内的神经元，而x_i代表其他神经元，i是空间索引。每个通道拥有$N=H\times W$个能量函数。基于上述公式，可以得到以下的解析解：

$$w_t=-\frac{2\left(t-\mu_t\right)}{\left(t-\mu_t\right)^2+2\sigma_t^2+2\lambda}$$
$$b_t=-\frac{1}{2}\left(t+\mu_t\right)w_t \tag{4-3}$$

因此，最小能量e_t^*可以通过如下公式得到：

$$e_t^*=\frac{4\left(\hat{\sigma}+\lambda\right)}{\left(t-\hat{\mu}\right)^2+2\hat{\sigma}^2+2\lambda} \tag{4-4}$$

在此通道上，除目标神经元 t 外，所有其他神经元的计算结果分别对应平均值$\hat{u}$和$\hat{\sigma}^2$方差。根据公式(4-4)，可以观察到能量值越低，神经元与其周围神经元的差异越显著，从而其重要性越高。因此，可以得到神经元的重要性。

对特征进行增强处理：

$$\tilde{X} = sigmoid\left(\frac{1}{E}\right) \otimes X \tag{4-5}$$

在增强处理的步骤中，E 操作将 e_t^* 空间和通道维度的所有值进行分组，并通过 Sigmoid 函数限制了 E 的输出。除了计算通道均值 $\hat{u}$ 和方差 $\hat{\sigma}^2$ 之外，该模块内的所有计算过程均为元素级操作，这一点与其他注意力机制相比，极大地提升了网络的灵活性。

4.2.4 损失函数

CioU(Complete Intersection over Union)[82] 损失对于加速预测框回归过程显示了一定程度的有效性，但仍然面临着一个明显的问题：当预测框与真实框(GT 框)的宽高比例呈线性关系时，CioU 损失中的相对比例惩罚项将不再有效。当 w 和 h 中的一个增加时，另一个必须减小，它们不能同步增减，为了解决这个问题，EIoU(Enhanced Intersection over Union)[83]将损失函数分成了三个部分，L_{EIoU} 表示预测框和真实框之间的重叠损失；L_{dis} 用于衡量预测框和真实框中心之间的距离损失；L_{asp} 则关注预测框与真实框在宽度和高度上的差异；此外，还引入了一种针对宽度 w 和高度 h 预测值直接进行惩罚的损失函数：

$$L_{EIoU} = L_{IoU} + L_{dis} + L_{asp} = 1 - IoU + \frac{\rho^2\left(b, b^{gt}\right)}{c_w^{\ 2} + c_h^{\ 2}} + \frac{\rho^2\left(w, w^{gt}\right)}{c_w^2} + \frac{\rho^2\left(h, h^{gt}\right)}{c_h^2} \tag{4-6}$$

其中 IoU 为交并比，c_w^2 和 c_h^2 分别是预测框和 GT 框最小外接矩形的宽和高，$\rho^2\left(b, b^{gt}\right)$ 为预测框和目标框的欧式距离，$\rho^2\left(w, w^{gt}\right)$ 表示预测框和目标框宽度的差值，$\rho^2\left(h, h^{gt}\right)$ 表示预测框和目标框高度的差值。

EIoU 损失的前两部分沿用了 CIoU 损失的计算方法，说明它们在计算上的相似性。另一方面，宽高损失部分则通过直接优化预测框与真实框之间的宽度和高度差异来进行改进，从而加快了模型的收敛速度。

4.3 实验结果与分析

4.3.1 实验平台

实验数据集采用的 2-3 章节创建的缺陷分类数据集，存在六种不同的缺陷类型，它们分别是轧内垢(RS)、斑块(Pa)、裂纹(Cr)、点蚀面(PS)、夹杂物(In)以及划痕(Sc)，共计含有 6000 张金属表面缺陷图像，每类图像包含 1000 张图片。所有实验均在统一平台上进行，如表 4-1 所示：

表 4-1 实验平台

实验设备	型号
设备	戴尔游匣 G15
平台	Pycharm 2020
CPU	Intel(R) Core(TM)i7-11800H
GPU	GeForce RTX 3060
System memory	16GiB
SSD	Samsung SSD 970 EVO Plus 1TB

在本研究中，提出了一种策略以解决训练过程可能遇到的震荡或梯度爆炸问题。实验旨在尽可能提升训练效率，因此，最初采用了一个较小的学习率来训练模型。这个阶段通常被认为是稳定模型训练的关键，因为较小的学习率有助于减少训练过程中的不稳定性。一旦模型达到了一定的稳定性，并且梯度的分布也相对均匀，会逐步恢复学习率至正常水平。这样做的目的是确保训练过程中不会出现梯度爆炸的情况，同时获得更高的训练速度。具体的超参数配置可见表 4-2。

表 4-2 超参数配置

超参数	数值
warmup_epochs	10
warmup_momentum	0.8
warmup_bias_lr	0.1
epochs	150
initial learning rate	0.1
final learning rate	0.2
momentum	0.937
weight_decay	0.0005
image-size	608×608

4.3.2 评价标准

IE-YOLOv5 的评估指标涵盖了精度(precision)、召回率(recall)、平均精度(AP)、平均精度均值(mAP)以及每秒帧数(FPS)。FPS 衡量的是模型每秒能处理的图像数，这一指标直接反映了检测的速度[84]。精确度衡量了在模型判定为阳性的样本里，真正阳性样本的占比。而召回率评估了在全部真阳性样本中，模型成功识别的阳性样本的比例。精度和召回率分别由式(4-7)和式(4-8)表示:

$$\mathrm{Pr}ecision = \frac{TP}{TP+FP} - \frac{YP}{prediction} \tag{4-7}$$

$$\mathrm{Re}call = \frac{TP}{TP+FN} = \frac{TP}{groundtruths} \tag{4-8}$$

其中，真阳性(TP)反映了模型正确判定的阳性样本数量，假阳性(FP)表示模型错误地将样本判定为阳性的数量，假阴性(FN)则指模型将阳性样本误判为阴性的数量。通过计算检测框与真实框之间的交并比(IoU)，依据特定的 IoU 阈值，可以据此确定 TP、FP 和 FN 的数量。

平均精度(AP)是指精度-召回率(P-R)曲线下方与坐标轴形成的面积大小。通过计算所有类别的 AP，可以获得平均精度(mAP)。AP 由式(4-9)定义，其中 K 代表所有类别的数量，P 代表精度，R 代表召回率。mAP 表示不同 AP 的平均值，mAP 的计算公式如(4-10)所示:

$$AP = \int_0^1 P(R)dR \tag{4-9}$$

$$mAP = \frac{\sum_{i=1}^{K} AP_i}{K} \tag{4-10}$$

4.3.3 消融实验

为了验证 IE-YOLOv5 网络算法在 NEU-DET 缺陷检测数据集中的检测效果，采用消融实验来验证每个改进模块的优化效果。YOLOv5-A 使用轻量级卷积方法 GSconv 替换标准卷积，YOLOv5-B 使用是在 GSConv 的基础进行特征融合，加入 FF-Slim-Neck 结构，YOLOv5-C 中，在主干网的末端引入了无参数 PSA 注意力机制。IE-YOLOv5 则同时融合了以上三个改进模块。

YOLOv5-A、YOLOv5-B、YOLOv5-C 各模型参数如表 4-3 所示，其中 Inference Time 是指模型推断单幅图像所需的平均时间，Weight 表示模型的体积。

表 4-3 各模型参数比较

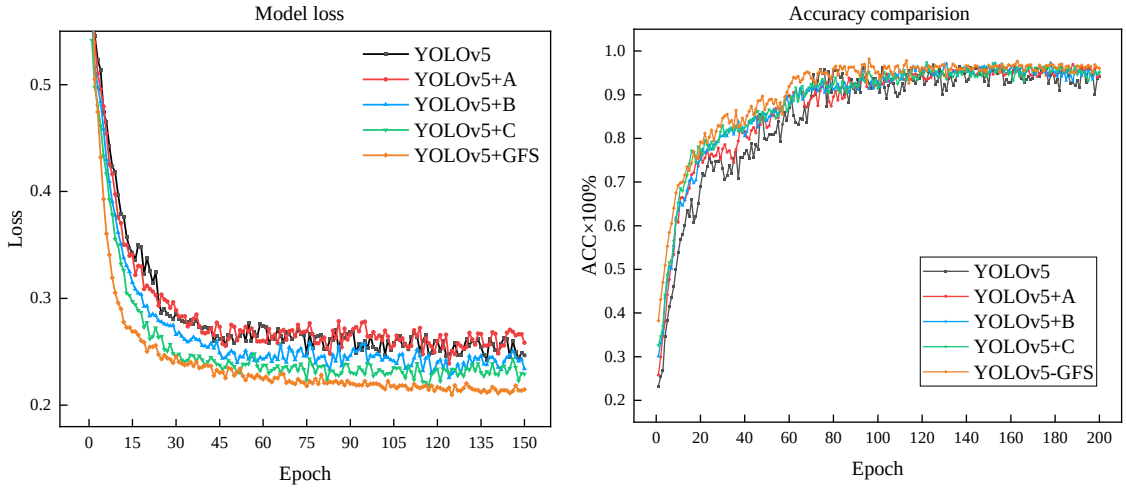
Model	GSconv	FF-Slim-Neck	PSA	Weight	Inference Time
YOLOv5	×	×	×	23.2MB	2.8ms
YOLOv5-A	√	×	×	23.2MB	2.7ms
YOLOv5-B	×	√	×	27.6MB	2.5ms
YOLOv5-C	×	×	√	28.2MB	2.7ms
IE-YOLOv5	√	√	√	28.2MB	2.5ms

消融实验结果表明如表 4-4 所示，与原始的 YOLOv5 模型相比，模型 YOLOv5-A 的推理时间缩短了 3.6%，mAP 提高了 0.8%，FPS 提高了 1.6%，这表明 GSConv 替换标准卷积 Conv 减少了推理时间，也提高了模型的检测精度和速度。对于模型 YOLOv5-B, mAP 比原来的 YOLOv5 模型增加了 1.3%，从而验证了引入 FF-Slim-Neck 的有效性，并且几乎没有影响模型的大小和推理时间。这表明 FF-Slim-Neck 结构提高了对缺陷的泛化能力和检测精度。模型 YOLOv5-C 使用了 SPF 注意力机制，体积增加了 21.6%，检测速度降低了 3.6%。而 mAP 则增加了 1%，说明该所提出的 SPF 注意力机制的有效性和效率。说明改进的特征融合网络对算法的提升有效。

表 4-4 模型评价指标比较

Model	AP						mAP	FPS
	craze	inclusion	patch	pit	roll	scratch		
YOLOv5	93.9%	93.8%	94.7%	94.3%	93.4%	95.7%	94.3%	48.65
YOLOv5-A	95.2%	96.1%	95.2%	95.5%	94.2%	96.5%	95.5%	49.45
YOLOv5-B	95.7%	96.5%	95.6%	96.3%	95.1%	97.1%	96.1%	48.69
YOLOv5-C	94.6%	96.3%	95.3%	96.3%	95.6%	97.9%	96.0%	46.29
IE-YOLOv5	95.7%	97.2%	96.3%	96.5%	96.4%	98.1%	96.7%	46.17

如图 4-8(a)所示，回归损失曲线的对比图显示，改进的模型在达到拟合状态时比原始的 YOLOv5 及其他消融模型(即 YOLOv5-A、YOLOv5-B、YOLOv5-C)更为迅速。在训练的初始阶段，原始的 YOLOv5 算法显示出较大的不稳定性，而改进后的 IE-YOLOv5 算法的训练损失曲线则显得更加平稳。训练完成时，与 YOLOv5 相比，IE-YOLOv5 的训练损失降低了 27%。这一结果表明，改进的方法有效减少了损失，同时优化了过拟合的问题。



(a) Loss 对比图　　(b) Accuracy 对比图

图 4-8 消融实验对比图

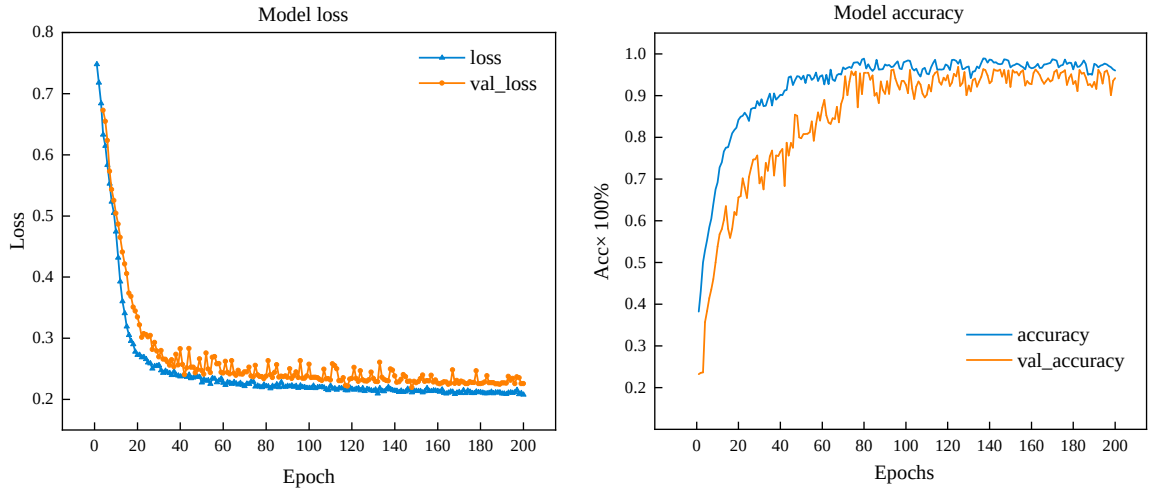
图 4-8(b)展示了 YOLOv5、YOLOv5-A、YOLOv5-B、YOLOv5-C 以及 IE-YOLOv5 各自的平均精度均值(mAP)曲线的比较。YOLOv5 的 mAP 在 120 次迭代后趋于稳定，最终大约达到了 94%。YOLOv5-A 的 mAP 在 110 次迭代之后稳定，最终值大约为 95%。YOLOv5-B 和 YOLOv5-C 的 mAP 在 100 次迭代后稳定，它们的最

终 mAP 均约为 96%。IE-YOLOv5 网络精度相较于 YOLOv5、YOLOv5-A、YOLOv5-B 和 YOLOv5-C 分别提高了 2.4%、1.2%、0.6%、0.7%。

4.3.4 对比实验与结果分析

依据本研究提出的算法构建的网络模型，图 4-9(a)展示了 IE-YOLOv5 网络在训练及验证阶段的损失曲线。曲线在训练初期显示出显著的下降趋势，在进行了大约 100 轮迭代之后，无论是训练集还是验证集上的损失曲线都显示出了稳定的趋势，并最终稳定在 0.2 左右。这表明所设置的网络超参数是合理的，能够确保网络训练和验证过程的有效收敛。

图 4-9(b)显示了 IE-YOLOv5 网络的平均精度均值(mAP)曲线，反映了训练集与验证集的表现。在大约 80 次迭代之后，两者的 mAP 曲线趋向于稳定，并且最终的平均精度均值超过了 96%。这一结果表明，本研究构建的金属表面缺陷检测模型结构合理，具有稳定的性能，并且没有出现过拟合或欠拟合的问题。



(a) 训练和验证 Loss 曲线　　(b) 平均精度均值曲线

图 4-9 IE-YOLOv5 网络的训练和验证

为了进一步验证本文算法的优越性，将其与 Faster R-CNN、SSD、YOLOv3、YOLOv4 目标检测算法进行对比实验。

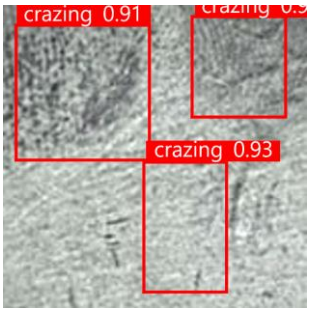
表 4-5 的检测结果表明，本文提出的算法在金属表面缺陷检测领域取得了优异的表现，实现了最佳的综合性能。具体而言，此算法在检测 In、Ps、Sc 三类金属表面缺陷时，其平均精度均值(mAP)达到了 97%，与 SSD 算法相比，处理密集缺陷的

漏检率有了显著的降低，且检测精度增加了 5.3%。在算法的执行速度上，由于采用了高效的细颈结构，减少了计算资源的消耗，从而实现了 46.17fps 的高检测帧率，与 Faster R-CNN 和 SSD 网络相比，分别提高了 28.5fps 和 17.28fps。这一进展不仅在精度方面超越了现有技术，同时在速度上也实现了显著的提高。

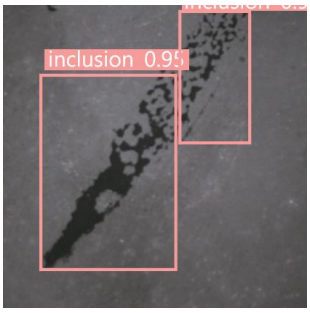
表 4-5 不同算法检测结果

Model	AP						mAP	FPS
	Cr	In	Pa	Ps	RS	Sc		
Faster R-CNN	92.1%	93.5%	90.4%	91.2%	93.6%	92.5%	92.2%	17.67
SSD	90.5%	92.1%	90.2%	93.5%	91.4%	90.7%	91.4%	28.89
YOLOV3	92.5%	93.9%	92.2%	93.8%	95.1%	94.8%	93.7%	34.18
YOLOV4	93.8%	94.7%	93.1%	96.3%	96.2%	97.1%	95.2%	40.19
IE-YOLOv5	95.7%	97.2%	96.3%	96.5%	96.4%	98.1%	96.7%	46.17

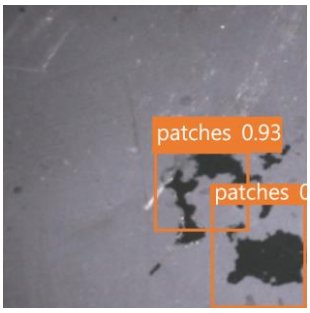
本文提出的改进 IE-YOLOv5 算法，通过与各个经典目标检测网络进行对比实验，验证了该网络的轻量化和准确度高，并通过消融实验进行定量和定性的分析，验证了 GSconv|、FF-Slim-Neck 和 PSA 模块的有效性，IE-YOLOv5 训练结果应用于自建分类数据集，检测效果在图 4-10 所示。从图 4-10 可以明显看出，本文提出的改进版算法相较于原始网络模型，展现了更加优越的性能。



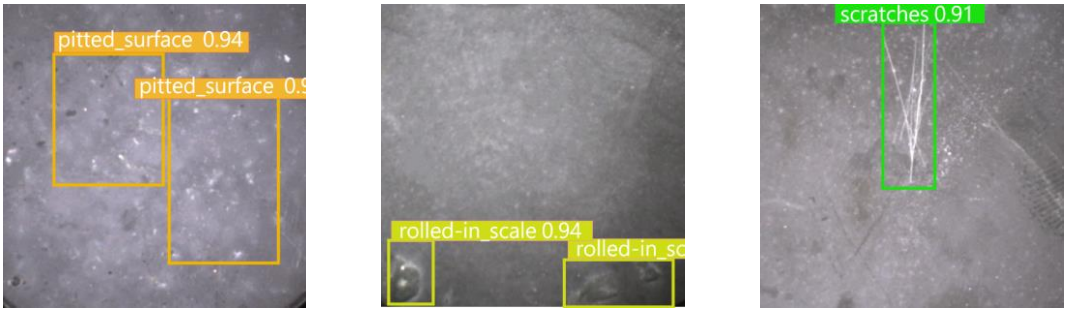
(a) 裂纹(Cr)



(b) 夹杂物(In)



(c) 斑块(Pa)



(d) 点蚀面(PS)　　(e) 轧内垢(RS)　　(f) 划痕(Sc)

图 4-10 部分检测实例

4.4 本章小结

在本章节中，本研究针对现有金属表面缺陷检测中精度和效率难以平衡的挑战，提出了基于改进型 YOLOv5 网络的解决方案，命名为 IE-YOLOv5。这一模型采用轻量级卷积 GSConv 替换了传统的标准卷积，构建了高效的特征提取 Neck 模块，以提升特征提取的效率。通过引入联邦学习算法 FedFusion，本研究开发了特征融合细颈结构 FF-Slim-Neck，这一结构优化了原有的 Neck 模块，并减少了计算成本。此外，整合了无参数注意力机制 PSA，使得 IE-YOLOv5 模型不仅能够快速精准地检测金属表面缺陷，还显著减少了模型参数，提高了网络的轻量化程度。

本章的主要贡献如下：

1、引入轻量级卷积 GSConv 替代 YOLOv5 中的标准卷积 Conv，实现了高效的特征提取并保留了每个通道的隐藏链接。

2、采用联邦学习算法 FedFusion 引入特征融合模块，在特征提取阶段实现局部模型与全局模型的特征聚合，与 GSConv 相结合，形成 FF-Slim-Neck 结构，有效减少计算开销。

3、提出无参数注意力机制 PSA，减轻网络参数负担并提高网络表达能力，同时优化损失函数为 EIoU，提升模型的预测准确率和收敛速度。

综上，IE-YOLOv5 模型通过这些创新和改进，不仅增强了金属表面缺陷检测的性能，还显著提高了处理速度和效率，有力地支持了工业级应用的需求。

第 5 章 半监督对比金属表面检测技术分析

5.1 本章方法概述

本章节专注于解决实际应用中由于标注资源缺乏所带来的挑战，提出了一种半监督检测网络来准确识别缺陷区域。首先介绍了基于高效密集输入处理能力的密集检测器上加入伪标签分配器的高效教师方法。该方法旨在高效利用可靠和不确定的伪标签。接着，通过卷积自动编码器(CAE)从数据集中提取缺陷数据分布信息，这对于增强金属表面缺陷的检测是至关重要的。进一步，提出了一个基于伪标签分配器的潜在特征提取框架，强调了该策略在提升识别精度和效率方面的重要性。

5.2 半监督检测网络

半监督网络是一类机器学习模型，其目标是通过同时利用大量未标记数据和少量标记数据进行训练和学习。这种网络通过结合无标签数据的隐式知识和有标签数据的显式信息来改善学习过程和模型性能。半监督学习适用于标签获取成本高昂或数据标注困难的情况，能有效提高数据利用率，增强模型的泛化能力。半监督学习的常见方法包括自训练、一致性正则化、伪标签使用，以及学生-教师模型等策略。这些方法有助于克服有限标注数据的挑战，实现更精确的预测。

在半监督学习的学生-教师模型流程中，如图 5-1 所示，教师模型最初通过有标签数据进行训练。完成训练之后，该模型被用来对无标签数据进行预测，产生伪标签。随后，学生模型利用这些伪标签与有标签数据共同进行训练，目的是学习教师模型的知识并进一步提升性能。学生模型训练后的版本可以成为新的教师模型，进行迭代训练，以此不断精炼伪标签的质量和模型的性能。

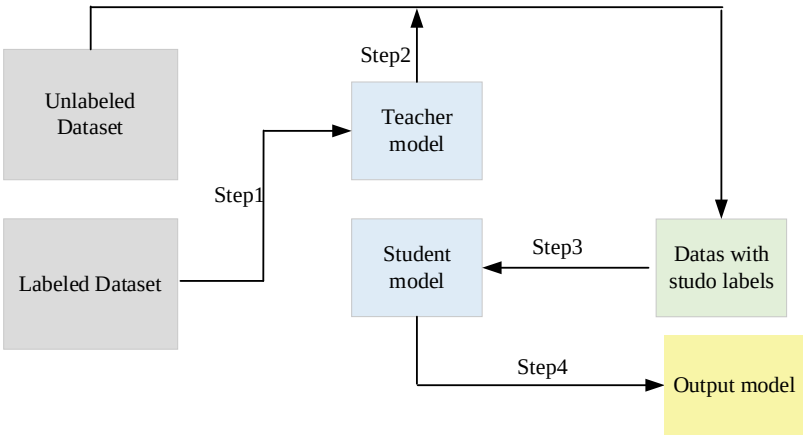


图 5-1 半监督学习的学生-教师模型流程

5.3 算法设计

本研究提出了一种新颖的半监督潜在特征提取技术，用于识别缺陷区域。这一方法基于伪标签分配器(Pseudo Label Assigner)，通过使用师生网络架构处理未标记数据，从而增强分类模型的性能。该技术的核心在于利用未标记数据的潜在信息，通过生成伪标签来训练模型，使得模型能够在有限的标记数据情况下也能实现更高的准确性和泛化能力，整个架构如图 5-2 所示。本研究采取了一系列关键措施：首先，结合了半监督学习框架和均值教师法，通过生成伪交叉训练图像来缓解图像级别上的域差异，应用半监督学习的更新损失来补偿跨域的不一致性，并引入了一种新的一致性损失来进一步校正跨域的偏差，旨在使学生模型能够识别未标注目标域中的实例级特征。此外，算法还通过场景风格的变换，生成不同域间的伪图像，减少图像层面上的差距。然后，发展了一个初始的教师模型，并在有限的已标注数据集上进行训练，确保其具有指导初学者模型的能力。教师模型建立之后，将对未标注数据进行筛选，并利用卷积自编码器(CAE)中的中间层来捕捉图像特征，此过程依赖教师模型的输出为未标注图像赋予伪标签。之后，选取置信度最高的前 K 张伪标签图像，与已有的标注数据结合，形成一个新的训练集。在学生模型的训练阶段，它通过模拟教师模型的判断来学习。训练完成后，对学生模型进行细致调整，以进一步提升其性能，使之更加契合实际应用需求。

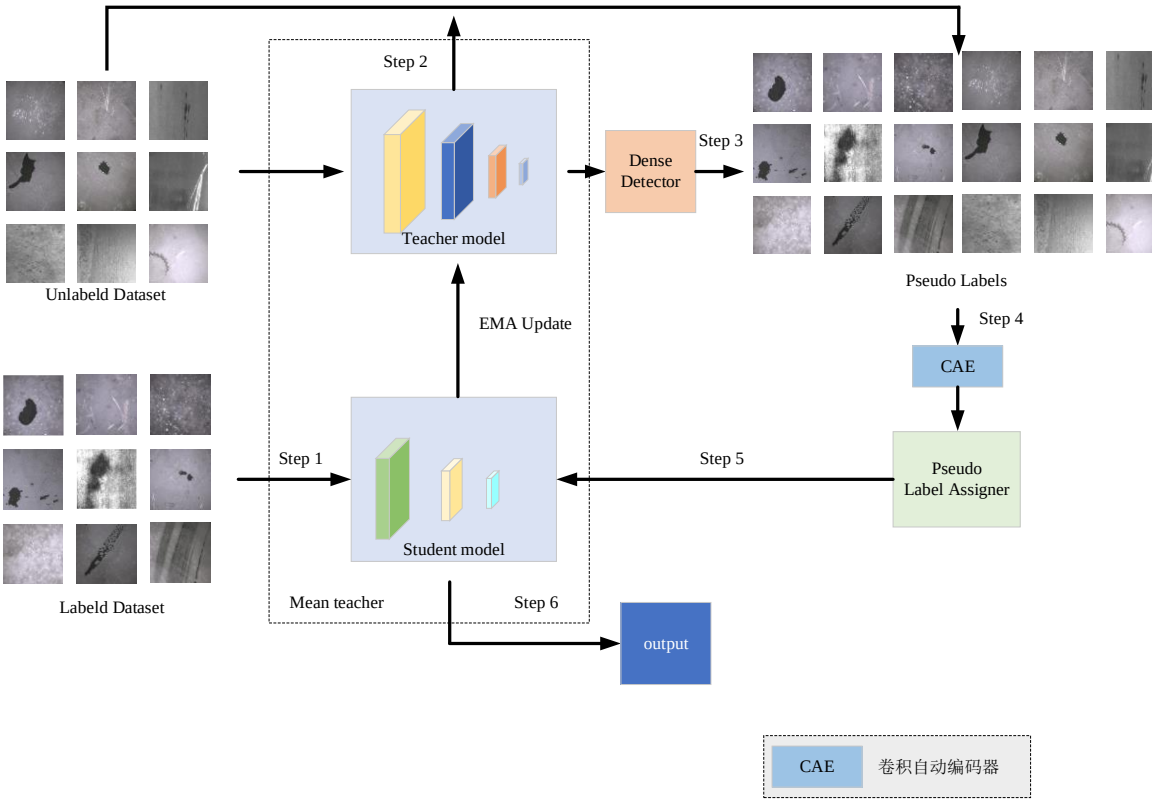


图 5-2 半监督网络结构

5.3.1 均值教师模式

Mean Teacher (MT)模型，最早在图像分类的半监督学习场景中被引入，结合了传统知识蒸馏的框架和两个结构相同的模型（即学生模型和教师模型）[85]。在进行域适应任务时，采用梯度下降法对源域的有标签数据进行训练。在 MT 模型的架构下，教师模型权重的更新依赖于学生模型权重的指数移动平均(EMA)。具体而言，设学生模型的权重参数为 p_s，而教师模型的权重参数为 p_t，在每个训练批次步骤更新 p_t 如下:

$$p_t = \gamma p_t + (1-\gamma) p_s \tag{5-1}$$

其中 γ 是指数衰减。它的合理值接近 1.0，通常在多个 9 的范围内:0.99、0.999 等。

MT 模型把未标记的目标域样本 Dt 作为教师模型的唯一输入，并在这些未标记的样本上 I^t 对学生模型进行部分训练。在知识蒸馏的过程中，学生模型通过选取教师模型预测中的高概率边界框作为伪标签，旨在减少目标域的方差，从而提升模型

的鲁棒性。假设从相同的图像中为教师模型增加了 $\hat{I}^t$ 的目标输入，为学生模型增加了 $\bar{I}^t$ 的目标输入，两个模型之间的预测不一致可以使用定义如下的蒸馏损失进行惩罚:

$$\zeta_{dis}(\hat{I}^t,\bar{I}^t)=\zeta_{\det}(\bar{I}^t,g_B\left[f_B(\bar{I}^t)\right],g_C\left[f_C(\bar{I}^t)\right]) \tag{5-2}$$

其中 $f_B(\cdot)$ 和 $f_C(\cdot)$ 是教师模型对边界框坐标和对象的预测分支在 $\hat{I}^t$ 上获得最高类别分数的班级。 $g_B(\cdot)$ 和 $g_C(\cdot)$ 是对应的过滤器。具体而言，在训练过程中每一步都将 MT 模型设置为评估模式，并使用非最大抑制(NMS)方法[86]对预测边界框按照具有 IoU 阈值 τ_{box} 的目标置信度进行排序并过滤。随后，挑选出那些类别得分超过阈值 τ_{cls} 的边界框。

在均值教师(Mean Teacher)模型的框架下，教师模型和学生模型通过相互促进来不断提升检测的精确度。提高的检测精度意味着教师模型能够生成更加准确且稳定的伪标签，这一点对于显著增强算法的性能发挥了至关重要的作用。此外，教师模型也被视作是不同时间点上学生模型的一个累积体，这与观察到的现象一致，即教师模型在准确度上通常会超越学生模型。

对于 MT 模型中的教师模型和学生模型，都采用了残差网络(ResNet-50)[87]作为骨干架构，实验中使用的网络结构详情如表 5-1 所示。ResNet 模型分成五个阶段进行，其中第 0 阶段的结构比较简单，主要作为输入数据的预处理环节。紧随其后的四个阶段则是由瓶颈结构构成的，结构相对相似。阶段 1 包含三个瓶颈，其余三个阶段分别包含四个、六个和三个瓶颈。每个阶段都包含不同的信息，并受到卷积和感受野的限制；早期阶段主要处理局部信息，而后期阶段则逐渐汇总局部信息以获得全局信息。

表 5-1 Resnet50 的详细信息

	Layer Name	Output Size	50-layer	Repeats
ResNet50	Conv1	112×112	7×7,64	-
	Conv2	56×56	1×1,64 3×3,64 1×1,256	3
	Conv3	28×28	1×1,128 3×3,128 1×1,512	3
	Conv4	14×14	1×1,256 3×3,256 1×1,1024	6
	Conv5	7×7	1×1,512 3×3,512 1×1,2048	3
	-	1×1	Average pool	-

5.3.2 CAE 模块

卷积自动编码器(CAE)是一种特殊的自动编码器，主要用于图像处理应用[75]。如图 5-3 所示。卷积自编码器(CAE)的架构主要由两部分构成：编码器部分和解码器部分。首先，在编码器部分，输入图像首先经过卷积层处理，这些卷积层利用滤波器来提取图像的特征，并且每次卷积操作之后都会应用激活函数，例如 ReLU，以引入非线性特性。随后，池化层减少特征的空间维度，这有助于提取更高级的特征并减少计算需求。经过一系列卷积和池化操作后，图像被压缩成一个瓶颈层，这个层包含了图像的压缩特征表示。随后，在解码器环节，通过上采样层将特征图的维度增大，接近于原始图像的尺度。然后，利用卷积层对上采样后的特征进行深入处理，以便重建原始图像中的细节信息。最终，输出层负责生成重构的图像，其主要目标是让这个重构后的图像与原始输入图像尽量相似。

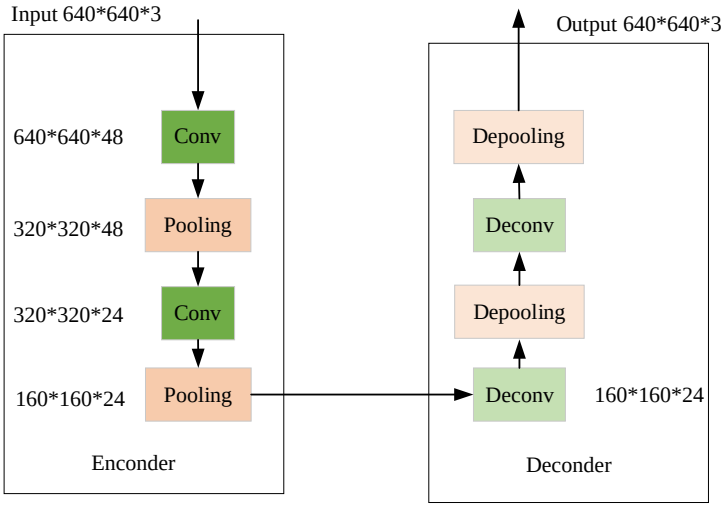


图 5-3 卷积自编码器(CAE)的结构

卷积层对输入应用多个滤波器。这些层可捕捉空间层次以及边缘、纹理等特征。每个卷积层的运算可以表示为：

$$h^{(l)} = f(W^{(l)} * h^{(l-1)} + b^{(l)}) \tag{5-3}$$

其中 $h^{(l-1)}$ 是第 *l-1* 层的输出，或者是第一层的输入图像； $W^{(l)}$ 和 $b^{(l)}$ 是第 *l* 层的权重和偏置；*表示卷积运算； f 是非线性激活函数

紧随卷积层的是池化操作，它降低了输入的空间维度。对于最大池化，其可以表示为：

$$p_{i,j}^{(l)} = \max(K \subseteq h^{(l)}) \tag{5-4}$$

其中 K 是输出 $h^{(l)}$ 中池化窗口的子集， $p_{i,j}^{(l)}$ 是池化后的输出。

解码器通常使用转置卷积(有时称为反卷积)来将数据扩展回原始输入大小。转置卷积操作可以表示为：

$$h^{(l)} = f((W^{(l)})^T * h^{(l-1)} + b^{(l)}) \tag{5-5}$$

其中 $(W^{(l)})^T$ 是在编码器中使用的第 *l* 层权重的转置.

训练中的损失函数是均方误差(MSE)，对每个像素定义为：

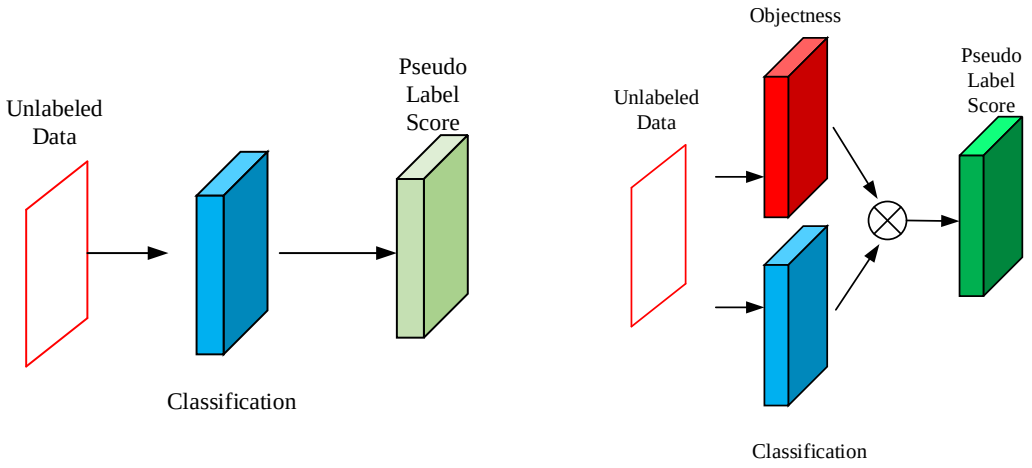
$$L = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2 \tag{5-6}$$

其中 N 是像素数量， x_i 是像素 i 的实际值， $\hat{x}_i$ 是像素 i 的重建值。

卷积自动编码器善于从图像中学习空间层次特征，这对于识别缺陷至关重要，缺陷通常表现为视觉数据中的异常模式或异常现象。通过学习“正常”图像的压缩表示法，可以对网络进行训练，使其能够识别出可能表示缺陷的规范偏差。卷积自动编码器还可以通过训练来去除图像中的噪声，这对于图像可能受到各种干扰或存在质量问题的工业环境非常有益。能对图像进行去噪处理的模型可以更好地关注真正的缺陷，而不是成像过程中的伪影。通过学习重建“正常”图像，卷积自动编码器可以通过测量重建误差来检测异常。较大的重构误差通常指示异常情况，因此，经过训练的卷积自编码器中的编码器部分可以被用作学生模型的特征提取器。

5.3.3 对比判断模块

为了克服无监督数据缺标签的挑战，本研究采用了伪标签法来训练学生网络，以此利用无监督数据。伪标签法在目标域中为学生模型实例提供了特征，同时比较了不同伪标签选择策略的影响。改进的密集检测器(Dense Detector)基于 ResNet-50-FPN 主干的 RetinaNet 进行改进[88]，如图 5-4(a)所示。该检测器对特征金字塔网络(FPN)[89]的输出层数进行了调整，将其从 5 层减少至 3 层，并取消了不同检测头之间的权重共享。同时，在训练和推理过程中，输入分辨率从 1333 降低到了 640。Dense Detector 主要包含三个输出分支：边界框偏移量、分类得分以及对象性得分。具体而言，Dense Detector 通过计算预测框与真实框(GT 框)之间的完全交并比(Complete Intersection over Union，CIoU)来确定目标得分[90]。对象性得分反映了预测框的位置质量，成为提升检测性能的重要信息源。在半监督检测中，伪标签代表未标记数据的预测框，其对象性得分衡量了伪标签的定位准确性。因此，相比仅具有一个分类分支的 RetinaNet(如图 5-4(b)所示)，引入对象性得分能有效表示半监督伪标签的定位质量。



(a) RetinaNet 结构　　(b) Dense Detector 结构

图 5-4 RetinaNet 和 Dense Detector 的比较

在半监督网络中，Pseudo Label Filter 方法[91, 92]通常被采用来改善伪标签的质量。然而，若阈值设定过低，则会产生不准确的伪标签；反之，设定过高则可能排除了一些实际上可靠的伪标签，这样的设定可能会导致次优的标签分配，从而不利于网络的训练效果。为了克服这一问题，本研究引入了一种新的伪标签分配方法(Pseudo Label Assigner, PLA)，该方法能够基于高低阈值系统将伪标签分为可靠和不确定两大类。在 PLA 框架下，通过对不确定的伪标签分配软标签作为 L_{obj} 的目标的策略，旨在提升半监督检测网络中伪标签的整体质量。进一步地，PLA 通过非最大抑制(Non-Maximum Suppression, NMS)[93]精确地处理伪标签，将它们划分为可靠和不确定两种类型。具体来说，阈值 τ_1 和 τ_2 用于区分这两种伪标签；位于 τ_1 和 τ_2 之间的伪标签被视为不确定[94]，而忽略这些标签可以直接提高 Dense Detector 的性能。除了解决伪标签过滤问题，PLA 还特别强调了如何有效利用不确定伪标签的无监督损失，以进一步增强检测模型的性能，如图 5-5 所示。

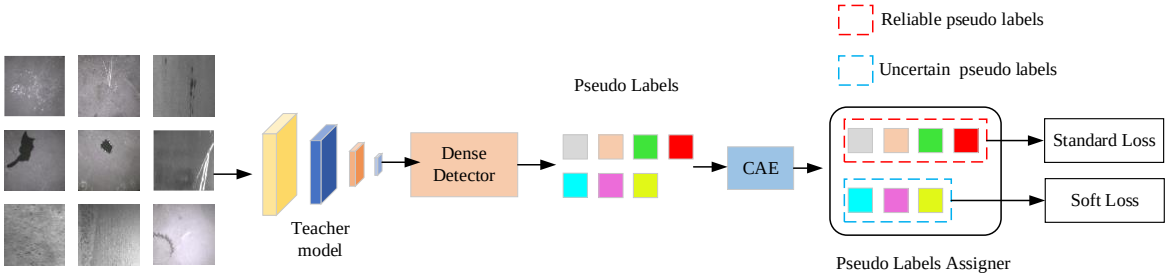


图 5-5 Pseudo Label Assigner 工作结构

5.3.4 损失函数

定义该半监督网络结构中 Dense Detector 的损失为一对单标记图像和单未标记图像:

$$L = L_S + \lambda L_U \tag{5-7}$$

其中 L_S 表示在标记图像上计算的损失函数，L_U 表示在未标记图像上计算的损失函数，λ 用于平衡监督损失和半监督损失，本研究将其设置为 3.0。L_S 为(5-8)中的标准损失函数:

$$L_s = \sum_{h,w}(CE(X^{cls}_{(h,w)}, Y^{cls}_{(h,w)}) + CIoU(X^{reg}_{(h,w)}, Y^{reg}_{(h,w)}) + CE(X^{obj}_{(h,w)}, Y^{obj}_{(h,w)})) \tag{5-8}$$

其中 CE 为交叉熵损失函数， $X_{(h,w)}$为学生模型的输出， $Y_{(h,w)}$为 Dense Detector 的标签分配器生成的采样结果

L_u 的定义如下:

$$L_u = L_u^{cls} + L_u^{reg} + L_u^{obj} \tag{5-9}$$

$$L_u^{\text{cls}} = \sum_{h,w}(\mathbb{R}_{\{\text{p}(h,w)>=r_2\}} CE(X^{cls}_{(h,w)}, \hat{Y}^{\text{cls}}_{(h,w)})) \tag{5-10}$$

$$L_u^{reg} = \sum_{h,w}(\mathbb{R}_{\{\text{p}(h,w)>=r_2 or \hat{obj}(h,w)>0.99\}} CIoU(X^{reg}_{(h,w)}, \hat{Y}^{reg}_{(h,w)})) \tag{5-11}$$

$$\begin{aligned} L_{\text{u}}^{cls} = \sum_{h,w}(&\mathbb{R}_{\{\text{p}(h,w)<=r_1\}} CE(X^{obj}_{(h,w)}, 0) + \mathbb{R}_{\{\text{p}(h,w)>=r_2\}} CE(X^{obj}_{(h,w)}, \hat{Y}^{\text{obj}}_{(h,w)}) \\ &+ \mathbb{R}_{\{r_1<\text{p}(h,w)<r_2\}} CE(X^{obj}_{(h,w)}, \hat{obj}^{\text{obj}}_{(h,w)})) \end{aligned} \tag{5-12}$$

其中 $\hat{Y}^{\text{cls}}_{(h,w)}$，$\hat{Y}^{reg}_{(h,w)}$，$\hat{Y}^{reg}_{(h,w)}$ 为位置(h,w)PLA 采样结果的分类得分、回归得分、客观性得分。$\hat{obj}(h,w)$是在(h,w)位置的伪标签，$p(h,w)$是在(h,w)位置的伪标号，$\mathbb{R}_{\{\}}$为指标函数，满足条件$\{\cdot\}$输出 1，不满足则输出 0。

PLA 与其他伪标签选择策略[93, 94]的区别在于设计了一个软损失来单独处理不确定的伪标签。PLA 区分两种不确定伪标签：高分类分数和高客观分数。对于第一种类型，只计算对象损失项 L^{obj}_u。交叉熵 $\hat{Y}^{\text{obj}}_{(h,w)}$ 被替换为软标签 $\hat{obj}_{(h,w)}$，这表明这些伪标签既不被分类为背景样本，也不被分类为阳性样本。对于第二种类型，PLA 计算对象得分大于 0.99 时的回归损失项 L^{reg}_u 这些伪标签有很好的回归结果，但是没有足够的分类分数来确定它们的标签类别。PLA 的目标是使用 L^{reg}_u 将更多不确定的伪标签转换为真阳性。

5.4 实验分析与结果

5.4.1 实验平台

所有的实验都是在一个统一的平台上完成的，具体细节如表 5-2 所示，使用 Python3.8 和 PyTorch v1.9.1 框架，网络优化采用了 Adam 优化器[97]，初始学习率(LR)为 0.002、动量参数 β_1=0.9，β_2=0.99，并且将训练 epoch 设置为 100。

表 5-2 实验平台

实验设备	型号
设备	戴尔游匣 G15
平台	Pycharm 2020
CPU	Intel(R) Core(TM)i7-11800H
GPU	GeForce RTX 3060
System memory	16GiB
SSD	Samsung SSD 970 EVO Plus 1TB

5.4.2 评价标准

本章节使用了五个主要的评估指标来衡量实验结果，包括召回率、F1 分数、精确度、准确度以及平均精确率的均值。这些指标是评估识别准确率的标准，尤其适用于类别不均衡的数据集，可以全面评价模型的性能。具体而言，精确度衡量的是在模型标识为正例的样本中，实际为正例的比例，而召回率则评价了在所有实际为正例的样本中，被模型正确标识的比例。这两个指标共同揭示了模型在识别正类样本方面的能力。F1 分数，作为精确度和召回率的调和平均，提供了一个评估模型整体表现的指标，实现了对查准率和查全率之间影响的平衡。准确度直接表示模型正确预测的样本占总样本的比例，是评估模型表现的最直观指标。mAP(平均精确率均值)则用于多类别识别任务中，衡量模型在各类别上的平均表现。通过这些综合性指标，能够详细了解模型在处理复杂数据场景下的准确性和效能。

1、精确度(Precision)计算公式为:

$$Precision = \frac{TP}{TP + FP} \tag{5-13}$$

2、召回率(Recall)的计算公式为：

$$Recall = \frac{TP}{TP + FN} \tag{5-14}$$

3、F1 分数(F1-sore)则是精确度和召回率的调和平均，计算公式为：

$$F1-\text{score} = \frac{2Precision \cdot \text{Re}call}{Precision + \text{Re}call} \tag{5-15}$$

4、精度(Accuracy)的计算公式为：

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + TN + FP} \tag{5-16}$$

其中 TP 是真正例的数量，FP 是假正例的数量,其中 FN 是假反例的数量。

5、mAP(mean Average Precision)的计算公式为：

$$\text{mAP} = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{5-17}$$

5.4.3 消融实验

为了验证改进后的半监督网络算法在半监督缺陷检测数据集中的检测效果，在 ResNet 模型中比较了各个模块的性能。两个模型都是在相同的设置下从头开始训练的。采用消融实验来验证每个改进模块的优化效果。各模型参数如表 5-3 所示，五种模型在测试集上的比较结果如表 5-4 所示。

表 5-3 各模型参数

Model	MT	CAE	PLA
Original model	×	×	×
Model-A	√	×	×
Model-B	×	√	×
Model-C	×	×	√
Improved model	√	√	√

表 5-4 定量比较

Model	P	R	F1	A
Original model	0.621	0.662	0.614	0.695
Model-A	0.641	0.689	0.693	0.709
Model-B	0.693	0.696	0.662	0.725
Model-C	0.6868	0.721	0.658	0.706
Improved model	0.717	0.724	0.705	0.752

从表 5-4 中可以看出，通过在模型中加入均值教师模型(Mean Teacher, MT)，F1 得分得到了明显的提升，提升了约 12.87%。均值教师模型通过融合多个模型的预测结果，有效地提高了模型对不同数据集的泛化能力，这在提高 F1 得分方面尤为重要。同时，引入卷积自动编码器(Convolutional Autoencoder, CAE)模块后，模型的整体分类性能获得了显著提高，准确率(Precision)提高了约 11.59%，召回率(Recall)提高了约 5.14%，F1 得分(F1 Score)提高了约 7.82%，准确度(Accuracy)提高了约 4.32%。CAE 模块通过有效地提取和学习数据的深层次特征，增强了模型在处理复杂数据时的分类能力。此外，加入伪标签识别器之后，召回率(Recall)提高了约 8.91%。这表明伪标签识别器在区分真实标签和伪标签方面表现出色，有效地减少了误分类的情况，特别是在数据标签质量不高或存在噪声的情况下更为明显。在改进模型(Improved model)相比于原始模型(Original model)方面，观察到准确率(Precision)提高了约 15.46%，召回率(Recall)提高了约 9.37%，F1 得分(F1 Score)提高了约 14.8%，

准确度(Accuracy)提高了约 8.2%。这些数据明确显示了改进模型在多个关键性能指标上的全面提升。

5.4.4 实验结果分析

(1)定量分析

在半监督数据集下，消融实验 Original model、Model-A、Model-B、Model-C 和 Improved model 模型的 mAP 比较如图 5-6 所示。Original model 初始性能增长迅速，但在 300 个 epoch 左右达到峰值后迅速下降，显示出显著的过拟合现象。从图 5-3 可以看出 Model-A 均值教师模型通过融合多个模型预测的方式，该模型在整合各个子模型的信息方面表现出色，进一步增强了模型对数据的泛化能力。在实验中，MT 模型的引入有效地减少了过拟合的风险，由于 CAE(卷积自编码器)在特征提取上的高效性，Model-B 能够从数据中学习到更深入且更具鲁棒性的特征表征，性能增长在开始时略显缓慢，后来加快，但在约 230 个 epoch 后出现性能下降的迹象。Model-C 通过伪标签识别器的引入使得识别器有效降低了误分类率，初始性能增长迅速，但在 230 个 epoch 左右达到峰值后迅速下降，显示出轻微的过拟合现象。Improved model 这个模型由于结合 Model-A、B、C 的优势，表现出稳定的性能提升，并且在大约 200 个 epoch 后接近饱和，达到最高的 mAP 值。

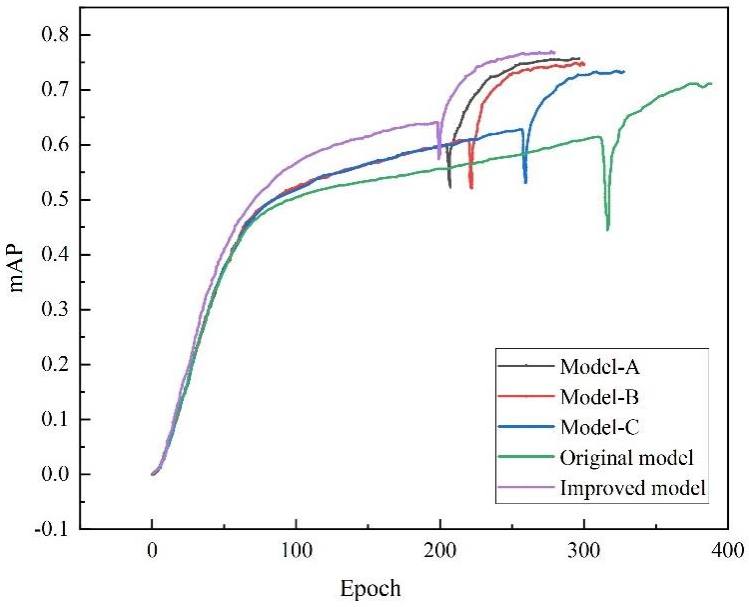


图 5-6 mAP 对比图

如图 5-7 所示，在最初的 100 个 epoch 中，所有模型的 loss 都显著下降，这反映了在学习的早期阶段，模型的学习效率较高。Model-A 在大约 150 个 epoch 后，loss 值稳定在 0.35 左右，这可能表明其采用的均值教师模式通过新型一致性损失成功地减少了跨域的客观性偏差。相比之下，Original model 的 loss 降至约 0.4，并显示出较大的波动，这暗示了过拟合的发生。Improved model 表现出最佳性能，其 loss 在 200 个 epoch 左右降至最低，约为 0.25，并保持稳定，这表明所采取的改进措施是有效的，并且该模型可能具有更强的泛化能力。Model-B 和 Model-C 的 loss 在降至约 0.3 之后出现显著波动，这表明仅使用 PLA 和 EA 模块可能导致学习过程中的不稳定性。由于 PLA 模块提供了更精确的伪标签分配，它比 Original model 的 loss 低约 0.12，尽管这可能增加了学习的 epoch 数。而 EA 模块在特征提取方面的效率导致其 loss 比 Original model 低约 0.15，从长远来看，Model-B 和 Model-C 的 loss 在 0.3 左右波动，这一水平比其他模型略高，表明它们可能受到过拟合影响。

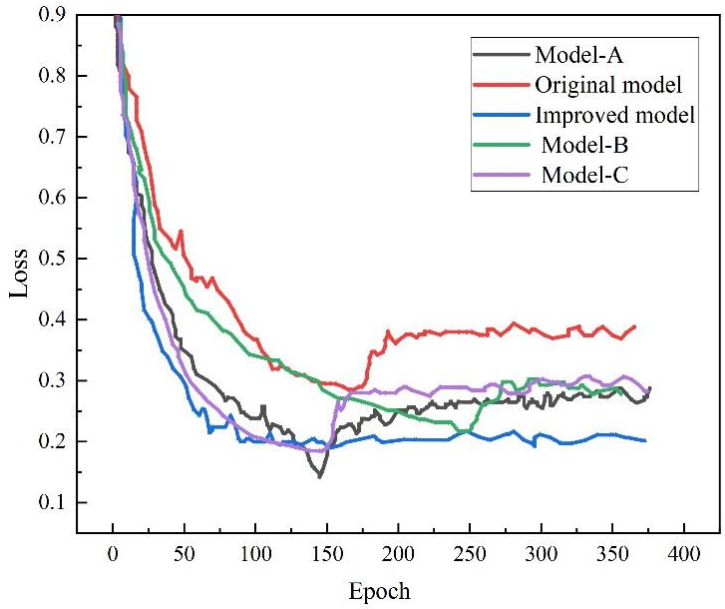


图 5-7 loss 对比图

(2)定性分析

对均值教师模型(MT)的作用的探讨：均值教师模型通过整合多个模型的预测，显著提高了信息融合的质量，因此模型对不同数据集的泛化能力得到了增强。实验结果显示，MT 模型的融入对抗了过拟合的倾向，尤其在处理那些复杂和不断变化的数据集时，这一优势尤为明显。

卷积自动编码器(CAE)模块的效益：CAE 在高效特征提取方面的卓越性能，使其能够从数据中学习更深层次且鲁棒的特征表示。在分类任务中，尤其是那些涉及到复杂金属表面缺陷的任务，CAE 的特征提取能力提供了精准的特征描述，显著提升了分类精度。

伪标签识别器的作用：引入伪标签识别器后，模型在降低误分类率方面取得了显著效果，特别是在区分难以识别的样本方面更是如此。这一创新在提升模型整体性能方面发挥了关键作用，这直接提高了模型对正类样本的识别精度，尤其在类别不平衡的数据集中，这种改进表现得尤为明显。该识别器的能力在自动生成高置信度伪标签方面表现优异，这不仅提升了数据的利用率，此外，这还显著提升了模型对于复杂数据集的泛化能力，对于数据稀疏的领域尤为重要。

整体模型性能的综合提升：与原始模型相比，改进后的模型在准确率、召回率、F1 得分和总体精度方面都实现了显著提高。这些进步不仅证明了结合 MT 模型、CAE 模块和伪标签识别器所带来的优化效果覆盖了模型的多个关键方面，而且准确率和召回率的提高直观地反映了模型对实际情况预测能力的增强，F1 得分和总体精度的提升则全面地展现了模型性能的整体提升。

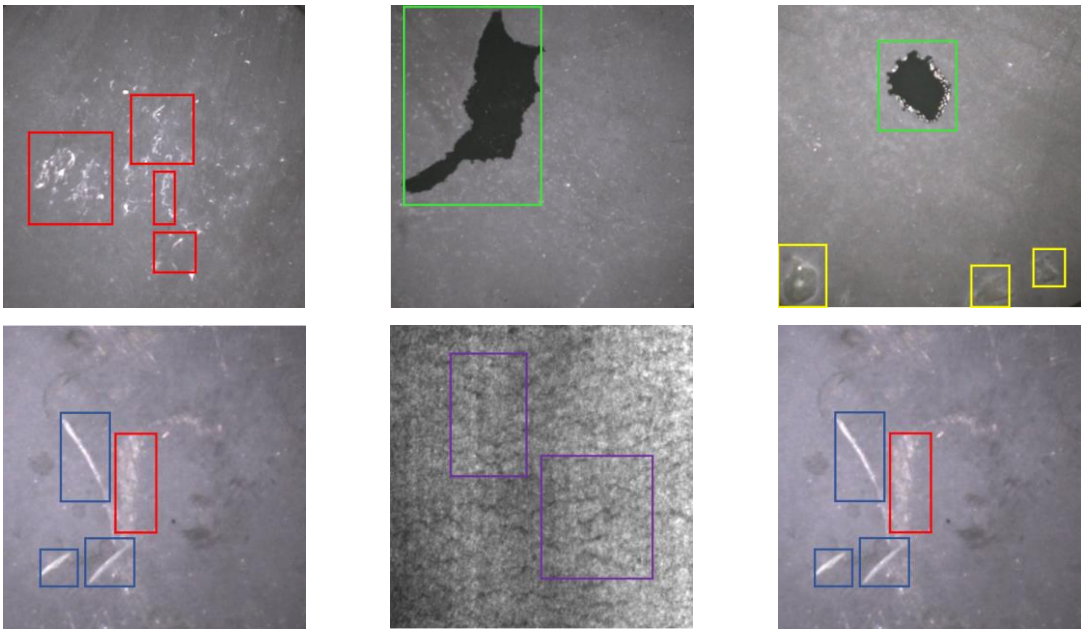


图 5-8 部分图片半监督网络检测结果

通过半监督网络检测验证结果如图5-8所示，红色框表示点蚀面(PS)，绿色框表示斑块(Pa)，黄色框表示轧内垢(RS)，紫色框表示裂纹(Cr)，点蚀面(PS)、、裂纹(Cr)、、橙色框表示夹杂物(In)、蓝色框表示划痕(Sc)。

实验结果表明，这种基于均值教师模式和卷积自动编码器的半监督学习方法，构建教师模型、标记、训练学生模型和微调的方法，成功地提高了模型在处理无标记数据时的性能。并且该网络在金属表面缺陷检测任务中表现出色。提出的方法在灵敏度值上远远高于标准方法。这一结果表明在金属表面缺陷分类这一特定应用场景中，改进后的模型能够以更高的准确性识别出微小的缺陷。这对于确保金属材料的质量控制和安全性检验至关重要。高灵敏度值不仅提升了模型在实际工业应用中的可靠性，还能够减轻人工检验的负担，提高自动化检测的效率。总体来说，实验结果显示了通过引入新技术和方法所带来的多方面性能提升，这不仅在理论上验证了所提方法的有效性，也为金属表面缺陷分类等实际应用提供了有价值的参考。

5.5 本章小结

在本章节中，提出了一种高效的教师方法，通过建立在密集检测器的高效密集输入处理之上，引入了伪标签分配器，以有效地利用可靠和不确定的伪标签，基于对半监督网络中伪标签分配的分析。同时，本章节也提出了一种基于伪标签分配器的潜在特征提取方法来检测金属表面的缺陷。

本章贡献为：

1、结合半监督学习框架和均值教师模型，通过生成伪标签训练图像以减轻图像级别的域差异，引入了更新的半监督损失来补偿跨域差异，以及一种新的一致性损失用于更进一步矫正跨域的偏差。通过利用师生网络在半监督学习场景下，有效地利用了大量的未标注数据，显著提升了数据的使用效率和模型的识别能力。

2、提出了一种探索金属表面图中缺陷分布的新方法，利用卷积自动编码器(CAE) 从数据集中获取缺陷数据分布信息。

3、提出了一种创新性的伪标签分配器，设计用于半监督学习网络中。这个分配器的核心目的是有效利用未标记数据，增强模型在有限监督信息条件下的学习能力。伪标签分配器通过为这些未标记数据自动生成高可信度的伪标签，架起了监督学习与未标记数据之间的桥梁。这种方法不仅提高了数据的利用效率，还显著增强

了模型在真实世界复杂数据集上的泛化能力。采用这种方法后，模型在经过少量有标签数据的训练之后，能够更有效地理解和处理大量未标记数据，这一点对于那些标注数据稀缺的领域来说尤其关键。

4、通过在数据集上执行的消融研究，从多种视角证实了经过改良的师生模型在实用性和有效性方面的显著优势。

第 6 章 总结与展望

6.1 研究成果总结

本文针对缺陷图像的分割技术方法、目标检测算法以及半监督检测算法开展了研究，提供了相关的技术实现模型与算法，为构建先进且成熟的环境感知系统提供了强有力的技术基础。主要完成工作如下：

(1)金属表面缺陷检测数据集的建立和优化：通过精心设计的预处理措施(如对数变换、中值滤波和直方图均衡化)和精确的手动标注，建立了两种专用于金属表面缺陷检测的数据集——分割数据集和分类数据集。分割数据集聚焦于缺陷的种类和特征，包含约 3000 张精细标注的影像，而分类数据集则经过数据增强扩充到 6000 张影像，涵盖六个缺陷类别。这两个数据集的建立为自动识别和分类金属表面缺陷提供了丰富的资源，是提高检测系统综合性能和准确性的关键。

(2)金属表面缺陷分割技术的进步：通过改进型 DeepLabV3+模型(Improved-DeepLabV3+)，结合三级注意力单元(TAUs)、挤压激励块(SENets)和密集空洞空间金字塔池化(DASPP)，显著提升了金属表面缺陷分割的效率和精确性。这些技术的集成不仅增强了模型的特征提取能力，还优化了边界细节的处理，在高精度金属表面缺陷分割方面展现出显著优势。

(3)金属表面缺陷检测的高效解决方案：提出的基于改进型 YOLOv5 网络(IE-YOLOv5)的方法，通过引入轻量级卷积 GSConv、特征融合的细颈结构 FF-Slim-Neck 和无参数注意力机制 PSA，实现了金属表面缺陷图像的快速、精准检测。IE-YOLOv5 模型的创新改进显著提高了检测性能、处理速度和效率，满足了工业级应用的需求。

(4)半监督学习网络的创新应用：通过引入伪标签分配器和基于伪标签的潜在特征提取方法，结合知识蒸馏框架和均值教师模型，本研究提出了一种高效的半监督方法。这一方法不仅有效利用了大量未标记数据，提高了数据利用率和识别性能，还增强了模型在有限监督信息条件下的学习能力，显著提升了模型在真实世界复杂数据集上的泛化能力。该模型有效地利用了未标记数据来提高学习效率，减少了标注成本，并增强了模型的实用性和有效性。

综上所述，本研究不仅在数据集的量化构建方面做出了贡献，更在数据质量控制、处理方法创新以及模型设计与优化方面取得了显著进展。通过这些成果，本研究为金属表面缺陷的自动检测和分类提供了坚实的数据基础和技术支持，这对于提升工业生产的自动化程度和产品的质量具有重大的意义。

6.2 未来展望

本研究在金属表面缺陷检测领域取得了显著成果，特别是在数据集构建、图像预处理技术、缺陷分割与识别模型的开发与优化等方面。未来工作可从以下几个方向进行扩展和深化：

数据集的扩展与多样性：由于本研究只针对同材料的金属缺陷进行研究，未来工作可着重于进一步扩展数据集的规模和多样性，例如引入更多种类的金属材料、不同制造过程的缺陷类型，以及不同环境条件下的影像数据。这将有助于模型适应更广泛的应用场景，并提升其泛化能力。

更先进的预处理技术：探索和应用新的图像预处理技术，如深度学习驱动的图像增强方法，可能会进一步提高模型对细微缺陷的识别率。

模型的实时性和集成性：由于本研究只是对以采集的数据对模型进行训练，并没有考虑实际生产中的实际情形，为了满足实际工业应用的需求，未来研究应致力于提高模型的实时性和集成性，以便将这些模型直接部署在生产线上，实现即时的缺陷检测。

半监督学习和自监督学习的深入：鉴于标注数据集的成本很高，未来研究可以继续探索半监督学习和自监督学习技术，以减少对大量标注数据的依赖，同时提高模型利用未标记数据的能力。

参考文献

[1] 赵军生.金属材料在现代工业产品设计中的创新应用研究[J].世界有色金属,2019,(06):243+245.

[2] 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.

[3] 朱云,凌志刚,张雨强.机器视觉技术研究进展及展望[J].图学学报,2020,41(06):871-890.

[4] Cai Z, Vasconcelos N. Cascade r-cnn: Delving into high quality object detection[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 6154-6162.

[5] 郭庆梅,刘宁波,王中训,等.基于深度学习的目标检测算法综述[J].探测与控制学报,2023,45(06):10-20+26.

[6] 石瑶,陈美玲.基于深度学习算法的三维激光雷达主动成像目标检测[J].激光杂志,2023,44(12):70-74.

[7] 吕向东,彭超亮,陈治国,等.基于 RSSD 的遥感图像目标检测算法[J].现代电子技术,2024,47(07):49-53.

[8] 钱烺, 罗小娟, 宋璐璐, 等. 基于物联网的智能家居安防监控系统设计[J]. 物联网技术, 2021, 11(03): 28-30.

[9] 陈商盈, 倪受东, 童林. 改进 YOLOX 的自动驾驶场景目标检测算法[J]. 计算机工程与应用: 1-11.

[10] 任柯燕, 谷美颖, 袁正谦, 等. 自动驾驶 3D 目标检测研究综述[J]. 控制与决策, 2023, 38(04): 865-89.

[11] 诸竹君, 袁逸铭, 焦嘉嘉. 工业自动化与制造业创新行为[J]. 中国工业经济, 2022, (07): 84-102.

[12] 吕张成, 张建业, 陈哲钥, 等. 基于深度学习的工业零件识别与抓取实时检测算法[J]. 机床与液压, 2023, 51(24): 33-8.

[13] 孔令军, 王茜雯, 包云超, 等. 基于深度学习的医疗图像分割综述[J]. 无线电通信技术, 2021, 47(02): 121-30.

[14] Girshick R, Donahue J, Darrell T, et al. R-CNN for object detection[C]. IEEE Conference. 2014.

[15] Girshick R. Fast r-cnn[C]. Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.

[16] He K, Gkioxari G, Dollár P, et al. Mask r-cnn[C]. Proceedings of the IEEE international conference on computer vision. 2017: 2961-2969.

[17] Guo D, Wu Z, Feng J, et al. Multi-scale semantic enhancement network for object detection[J]. Scientific Reports, 2023, 13(1): 7178.

[18] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.

[19] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]. proceedings of the Computer Vision–ECCV 2016: 21-37.

[20] Tan M, Pang R, Le Q V. Efficientdet: Scalable and efficient object detection[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 10781-10790.

[21] Law H, Deng J. Cornernet: Detecting objects as paired keypoints[C]. Proceedings of the European conference on computer vision (ECCV). 2018: 734-750.

[22] Zhou X, Wang D, Krähenbühl P. Objects as points[J]. arXiv preprint arXiv:190407850, 2019.

[23] Yang Z, Liu S, Hu H, et al. Reppoints: Point set representation for object detection[C]. Proceedings of the IEEE/CVF international conference on computer vision. 2019: 9657-9666.

[24] Li Z, Hou B, Wu Z, et al. FCOSR: A simple anchor-free rotated detector for aerial object detection[J]. Remote Sensing, 2023, 15(23): 5499.

[25] Wang Z, Bao C, Cao J, et al. AOGC: Anchor-Free Oriented Object Detection Based on Gaussian Centerness[J]. Remote Sensing, 2023, 15(19): 4690.

[26] 罗东亮, 蔡雨萱, 杨子豪, 等. 工业缺陷检测深度学习方法综述[J]. 中国科学:信息科学, 2022, 52(06): 1002-39.

[27] 费佳杰, 李宏胜, 任飞, 等. 基于深度学习的钢表面缺陷检测方法综述[J]. 现代信息科技, 2023, 7(19): 107-12.

[28] Suvdaa B, Ahn J, Ko J. Steel surface defects detection and classification using SIFT and voting strategy[J]. International Journal of Software Engineering and Its Applications, 2012, 6(2): 161-6.

[29] Xiao-Cong L. A hybrid SVM-QPSO model based ceramic tube surface defect detection algorithm[C]. 2014 Fifth International Conference on Intelligent Systems Design and Engineering Applications. IEEE, 2014: 28-31.

[30] Gyimah N K, Girma A, Mahmoud M N, et al. A robust completed local binary pattern (rclbp) for surface defect detection[C]. 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, 2021: 1927-1934.

[31] Luo Q, Fang X, Su J, et al. Automated visual defect classification for flat steel surface: a survey[J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(12): 9329-49.

[32] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]. Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.

[33] Bochkovskiy A, Wang C-Y, Liao H-Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:200410934, 2020.

[34] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7263-7271.

[35] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:180402767, 2018.

[36] Lv L, Yao Z, Wang E, et al. Efficient and Accurate Damage Detector for Wind Turbine Blade Images[J]. IEEE Access, 2022, 10: 123378-86.

[37] Hatab M, Malekmohamadi H, Amira A. Surface defect detection using YOLO network[C]. Intelligent Systems and Applications: Proceedings of the 2020 Intelligent Systems Conference (IntelliSys) Volume 1. Springer International Publishing, 2021: 505-515.

[38] Li M, Wang H, Wan Z. Surface defect detection of steel strips based on improved YOLOv4[J]. Computers and Electrical Engineering, 2022, 102: 108208.

[39] Wang L, Cao Y, Wang S, et al. Investigation into recognition algorithm of helmet violation based on YOLOv5-CBAM-DCN[J]. IEEE Access, 2022, 10: 60622-32.

[40] Le H F, Zhang L J, Liu Y X. Surface Defect Detection of Industrial Parts Based on YOLOv5[J]. IEEE Access, 2022, 10: 130784-94.

[41] Wang H, Yang X, Zhou B, et al. Strip surface defect detection algorithm based on YOLOV5[J]. Materials, 2023, 16(7): 2811.

[42] Chen W, Huang Z, Mu Q, et al. PCB Defect Detection Method Based on Transformer-YOLO[J]. IEEE Access, 2022, 10: 129480-9.

[43] Li Z, Tian X, Liu X, et al. A two-stage industrial defect detection framework based on improved-yolov5 and optimized-inception-resnetv2 models[J]. Applied Sciences, 2022, 12(2): 834.

[44] Liu Z, Gao Y, Du Q, et al. YOLO-extract: improved YOLOv5 for aircraft object detection in remote sensing images[J]. IEEE Access, 2023, 11: 1742-51.

[45] Lou H, Duan X, Guo J, et al. DC-YOLOv8: small-size object detection algorithm based on camera sensor[J]. Electronics, 2023, 12(10): 2323.

[46] Guo Y, Zhong L, Qiu Y, et al. Using ISU-GAN for unsupervised small sample defect detection[J]. Scientific Reports, 2022, 12(1): 11604.

[47] Karmakar S, Banerjee A, Gidde P, et al. Convolutional Ensembling based Few-Shot Defect Detection Technique[C]. Proceedings of the Thirteenth Indian Conference on Computer Vision, Graphics and Image Processing. 2022: 1-7.

[48] Min Y, Wang Z, Liu Y, et al. FS-RSDD: few-shot rail surface defect detection with prototype learning[J]. Sensors, 2023, 23(18): 7894.

[49] Pang S, Zhao W, Wang S, et al. Permute-MAML: exploring industrial surface defect detection algorithms for few-shot learning[J]. Complex & Intelligent Systems, 2024, 10(1): 1473-82.

[50] Bao Y, Song K, Liu J, et al. Triplet-graph reasoning network for few-shot metal generic surface defect segmentation[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1-11.

[51] Benoit K. Linear regression models with logarithmic transformations[J]. London School of Economics, London, 2011, 22(1): 23-36.

[52] Chandel R, Gupta G. Image filtering algorithms and techniques: A review[J]. International Journal of Advanced Research in Computer Science and Software Engineering, 2013, 3(10).

[53] Lin S C-F, Wong C Y, Rahman M A, et al. Image enhancement using the averaging histogram equalization (AVHEQ) approach for contrast improvement and brightness preservation[J]. Computers & Electrical Engineering, 2015, 46: 356-70.

[54] Huang Y, Jing J, Wang Z. Fabric defect segmentation method based on deep learning[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1-15.

[55] Su H, Zhu X, Gong S. Deep learning logo detection with data expansion by synthesising context[C]. 2017 IEEE winter conference on applications of computer vision (WACV). IEEE, 2017: 530-539.

[56] Azad R, Asadi-Aghbolaghi M, Fathy M, et al. Attention deeplabv3+: Multi-level context attention mechanism for skin lesion segmentation[C]. European conference on computer vision. Cham: Springer International Publishing, 2020: 251-266.

[57] Mallat S. Understanding deep convolutional networks[J]. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2016, 374(2065): 20150203.

[58] He H, Yang D, Wang S, et al. Road extraction by using atrous spatial pyramid pooling integrated encoder-decoder network and structural similarity loss[J]. Remote Sensing, 2019, 11(9): 1015.

[59] Dong G, Yan Y, Shen C, et al. Real-time high-performance semantic image segmentation of urban street scenes[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 22(6): 3258-74.

[60] Mahmud T, Alam M J, Chowdhury S, et al. CovTANet: a hybrid tri-level attention-based network for lesion segmentation, diagnosis, and severity prediction of COVID-19 chest CT scans[J]. IEEE Transactions on Industrial Informatics, 2020, 17(9): 6489-98.

[61] Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 13713-13722.

[62] Guo C, Szemenyei M, Yi Y, et al. Sa-unet: Spatial attention u-net for retinal vessel segmentation[C]. 2020 25th international conference on pattern recognition (ICPR). IEEE, 2021: 1236-1242.

[63] Zhao H, Kong X, He J, et al. Efficient image super-resolution using pixel attention[C]. Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16. Springer International Publishing, 2020: 56-72.

[64] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.

[65] Kouretas I, Paliouras V. Simplified hardware implementation of the softmax activation function[C]. 2019 8th international conference on modern circuits and systems technologies (MOCAST). IEEE, 2019: 1-4.

[66] Pratiwi H, Windarto A P, Susliansyah S, et al. Sigmoid activation function in selecting the best model of artificial neural networks[C]. Journal of Physics: Conference Series. IOP Publishing, 2020, 1471(1): 012010.

[67] Eckle K, Schmidt-Hieber J. A comparison of deep networks with ReLU activation function and linear spline-type methods[J]. Neural Networks, 2019, 110: 232-42.

[68] Dubey S R, Singh S K, Chaudhuri B B. Activation functions in deep learning: A comprehensive survey and benchmark[J]. Neurocomputing, 2022, 503: 92-108.

[69] Zhang Z, Sabuncu M. Generalized cross entropy loss for training deep neural networks with noisy labels[J]. Advances in neural information processing systems, 2018, 31.

[70] Zhao R, Qian B, Zhang X, et al. Rethinking dice loss for medical image segmentation[C]. 2020 IEEE International Conference on Data Mining (ICDM). IEEE, 2020: 851-860.

[71] Loshchilov I, Hutter F. Sgdr: Stochastic gradient descent with warm restarts[J]. arXiv preprint arXiv:160803983, 2016.

[72] Qin X, Zhang Z, Huang C, et al. Basnet: Boundary-aware salient object detection[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 7479-7489.

[73] Huang H, Lin L, Tong R, et al. Unet 3+: A full-scale connected unet for medical image segmentation[C]. ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, 2020: 1055-1059.

[74] Zhang Y, Liu H, Hu Q. Transfuse: Fusing transformers and cnns for medical image segmentation[C]. Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24. Springer International Publishing, 2021: 14-24.

[75] Wu W, Liu H, Li L, et al. Application of local fully Convolutional Neural Network combined with YOLO v5 algorithm in small target detection of remote sensing image[J]. PloS one, 2021, 16(10): e0259283.

[76] Li H, Li J, Wei H, et al. Slim-neck by GSConv: A better design paradigm of detector architectures for autonomous vehicles[J]. arXiv preprint arXiv:220602424, 2022.

[77] Masci J, Meier U, Cireşan D, et al. Stacked convolutional auto-encoders for hierarchical feature extraction[C]. Artificial Neural Networks and Machine Learning–ICANN 2011: 21st International Conference on Artificial Neural Networks, Espoo, Finland, June 14-17, 2011, Proceedings, Part I 21. Springer Berlin Heidelberg, 2011: 52-59.

[78] Singh D, Kumar V, Kaur M. Densely connected convolutional networks-based COVID-19 screening model[J]. Applied Intelligence, 2021, 51: 3044-51.

[79] Lee Y, Hwang J, Lee S, et al. An energy and GPU-computation efficient backbone network for real-time object detection[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2019: 0-0.

[80] Wang C Y, Liao H Y M, Wu Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2020: 390-391.

[81] Yao X, Huang T, Wu C, et al. Towards faster and better federated learning: A feature fusion approach[C]. 2019 IEEE International Conference on Image Processing (ICIP). IEEE, 2019: 175-179.

[82] Wang X, Song J. ICIoU: Improved loss based on complete intersection over union for bounding box regression[J]. IEEE Access, 2021, 9: 105686-95.

[83] Hou F, Lei W, Li S, et al. Improved Mask R-CNN with distance guided intersection over union for GPR signature detection and segmentation[J]. Automation in Construction, 2021, 121: 103414.

[84] Wang P, Liang J, Wang L V. Single-shot ultrafast imaging attaining 70 trillion frames per second[J]. Nature communications, 2020, 11(1): 2091.

[85] Tarvainen A, Valpola H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results[J]. Advances in neural information processing systems, 2017, 30.

[86] Gong M, Wang D, Zhao X, et al. A review of non-maximum suppression algorithms for deep learning target detection[C]. Seventh Symposium on Novel Photoelectronic Detection Technology and Applications. SPIE, 2021, 11763: 821-828.

[87] Koonce B, Koonce B. ResNet 50[J]. Convolutional neural networks with swift for tensorflow: image recognition and dataset categorization, 2021: 63-72.

[88] He Y, Zhu C, Wang J, et al. Bounding box regression with uncertainty for accurate object detection[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 2888-2897.

[89] Zhu L, Lee F, Cai J, et al. An improved feature pyramid network for object detection[J]. Neurocomputing, 2022, 483: 127-39.

[90] Zhou H, Ge Z, Liu S, et al. Dense teacher: Dense pseudo-labels for semi-supervised object detection[C]. European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 35-50.

[91] Liu Y-C, Ma C-Y, He Z, et al. Unbiased teacher for semi-supervised object detection[J]. arXiv preprint arXiv:210209480, 2021.

[92] Xu M, Zhang Z, Hu H, et al. End-to-end semi-supervised object detection with soft teacher[C]. Proceedings of the IEEE/CVF international conference on computer vision. 2021: 3060-3069.

[93] Hosang J, Benenson R,Schiele B. Learningnon-maximum suppression[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4507-4515.

[94] Chen B, Chen W, Yang S, et al. Label matching semi-supervised object detection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 14381-14390.

[95] Chen B, Li P, Chen X, et al. Dense learning based semi-supervised object detection[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 4815-4824.

[96] Liu Y C, Ma C Y, Kira Z. Unbiased teacher v2: Semi-supervised object detection for anchor-free and anchor-based detectors[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 9819-9828.

[97] Zhang Z. Improved adam optimizer for deep neural networks[C]. 2018 IEEE/ACM 26th international symposium on quality of service (IWQoS). Ieee, 2018: 1-2.

攻读硕士学位期间发表学术论文及参研项目

（一）参与的科研项目

[1] 面向复杂零部件的打磨机器人的智能控制及其应用.鸠江区产业协同创新专项基金项目，课题编号: 2022cyxtb4.

（二）发表的学术论文

[1] An Enhanced YOLOv5-Based Algorithm for Metal Surface Defect Detection[J]. Applied Sciences, 2023, 13(20): 11473. (JCR Q4，对应第四章，作者一作)

[2] Weld Image Recognition based on Deep Learning[C]. 2022 9th International Conference on Dependable Systems and Their Applications (DSA). IEEE, 2022: 241-248. (EI，对应第四章，作者一作)

[3] Research on Key Technologies of Human Upper Limb Bone Diagnosis based on Convolutional Neural Network[C]. 2022 9th International Conference on Dependable Systems and Their Applications (DSA). IEEE, 2022: 909-917.(EI，对应第二章，作者二作)

（三）申请及已获得的发明专利

[1] 一种锥形声子束型能量采集器[P].安徽省:CN114421810A,2022-04-29. (作者六作)

致谢

在此，我谨向在学术论文撰写过程中给予我无限支持和宝贵指导的所有人表示最深切的感谢。

首先，我必须表达我对我的导师王海教授的衷心感激。在研究生三年期间，他的专业知识、细致入微的指导和不懈努力为我的研究工作奠定了坚实的基础。王海老师具有严谨的科研态度和一丝不苟的工作精神，值得我学习，他的言传身教将使我终生受益。

感谢实验室杨春来老师对我的帮助和支持，也感谢桑元俊老师，对我两年助教工作提供支持，祝各位老师身体健康、工作顺利！

感谢我的同门张子豪、李帅。感谢你们为我带来的欢乐、灵感、慰藉、动力、安全感，也愿我们仨顶峰相见。同时，感谢室友江彤、卢娜、陈清清在生活上的照顾与帮助，也愿我们友谊长存。

我还要感谢张庆同学，感谢在我论文遇到困难的时候鼓励我，为我提供了宝贵的意见和新的视角。

对于我的父母，我欠他们一份深重的谢意！在我二十多年追学的道路上，爸爸妈妈给我树立了很好的榜样作用，同时他们提供了无条件和唯一的爱、鼓励和支持，在我人生遇到重大挫折的时候，成为我坚不可摧的后盾。感谢爸爸妈妈在我有时候闹脾气的时候，可以无条件的包容我理解我原谅我，感谢他们无私的奉献和坚强的信念，谢谢父母是我的底牌，作为你们女儿真的很幸福！

总之，这篇论文的完成离不开你们每一个人的贡献。我的感激之情，难以用言语完全表达，只愿我们共同努力的成果能为学术界带来微小却有意义的贡献。

尚德教学 唯实惟新