

分类号：

密级：

U D C :

学号： 411014519027

南 昌 大 学 专 业 学 位 硕 士 研 究 生
学 位 论 文

基于深度学习的瓷砖瑕疵检测识别
的研究和实现

Research and Implementation of Ceramic Tile Defect Detection
and Recognition Based on Deep Learning

洪诚

培养单位（院、系）：数学与计算机学院 计算机科学与技术系

指导教师姓名、职称：陈萌 副教授

指导教师姓名、职称：郭晓峰 副教授

专业学位种类：工程硕士

专业领域名称：计算机技术

论文答辩日期：2022 年 5 月 22 日

答辩委员会主席：_____邓承志_____

评阅人：_____盲审_____

_____盲审_____

2022 年 5 月 22 日

一、学位论文独创性声明

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

学位论文作者签名（手写）：洪诚

签字日期：2022 年 6 月 9 日

二、学位论文版权使用授权书

本学位论文作者完全了解南昌大学有关保留、使用学位论文的规定，同意学校有权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，允许论文被查阅和借阅。本人授权南昌大学可以将学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编本学位论文。同时授权北京万方数据股份有限公司和中国学术期刊（光盘版）电子杂志社将本学位论文收录到《中国学位论文全文数据库》和《中国优秀博硕士学位论文全文数据库》中全文发表，并通过网络向社会公众提供信息服务，同意按“章程”规定享受相关权益。

学位论文作者签名（手写）：洪诚

导师签名（手写）：陈东

签字日期：2022 年 6 月 9 日

签字日期：2022 年 6 月 9 日

论文题目	基于深度学习的瓷砖瑕疵检测识别的研究和实现				
姓 名	洪 诚	学号	411014519027	论文级别	博士 <input type="checkbox"/> 硕士 <input checked="" type="checkbox"/>
院/系/所	数学与计算机学院		专业	计算机技术	
E_mail					
备注：					

☒公开 ☐保密（向校学位办申请获批准为“保密”，____年__月后公开）

摘 要

缺陷检测是工业上非常重要的一个应用,与此同时,卷积神经网络在特征提取和定位上取得巨大成就。因此深度学习在目标检测领域取得了长足的发展,使得瓷砖表面缺陷检测成为一个有潜力的研究方向。但是瓷砖瑕疵检测依然受到检测目标中小目标较多,形状多变且不规则,特征不明显等因素的困扰,企业在生产制造的过程中依然无法避免的生产出具有各类瑕疵的瓷砖。如何解决以上种种问题,提高小目标瑕疵检测的精度是亟需解决的难题。本文主要是从对原始数据集进行增强和增强网络的特征提取能力两方面解决上述难题。

考虑到瓷砖瑕疵数据集中图像尺寸大,但是目标瑕疵较小的情况,主要从对原始数据集进行数据增强和增强网络的特征提取能力两方面解决。首先通过图像裁剪的方式进行数据增强,分别从机器学习和深度学习两个方面进行实验验证,结果表明数据增强的有效性。其次选择 Faster RCNN 为目标检测网络,使用分散注意力网络进行替换原始网络结构,加强网络特征提取能力,为进一步提取多尺度图像特征,引入了循环特征金字塔(RFPN)模块进行融合高低层特征信息,由于目标瑕疵的形状具有多变性和不规则性,引入可形变卷积(DCN)进行目标检测。最后经过实验验证,分别统计 Faster RCNN 替换特征为分散注意力网络(网络 1),替换后的 Faster RCNN 融合 RFPN 特征金字塔(网络 2)和基于可形变卷积 Faster RCNN 融入 RFPN(网络 3)这三种网络模型的平均精度和准确率两个综合指标,实验表明经过优化后的网络,网络 2 的 mAP 比网络 1 的 mAP 高 1.61%,说明引入 RFPN 模块提取到更多的图像特征,网络 3 的 mAP 比网络 2 的 mAP 值高 1.23%,说明在加入 DCN 模块之后提高了网络的检测精度。

考虑到现有技术传统工业生产制造受一定限制,人工检测产品速度慢且主观性会影响检测结果的准确性。本文基于提出的改进模型设计并实现了瓷砖瑕疵检测系统,包括登录注册,图像检测和后台管理三大功能,能够对导入系统的瓷砖瑕疵数据集进行识别和记录,促进企业进行智能化改造。

关键词: 瓷砖瑕疵; 图像分类; 数据增强; 深度学习

ABSTRACT

In industry, defect detection is a very important application. At the same time, deep learning has made great achievements in feature extraction and location. Therefore, convolutional neural network has made great progress in the field of target detection, which makes ceramic tile surface defect detection a potential research direction. However, ceramic tile defect detection is still plagued by many small and medium-sized targets, changeable and irregular shapes, unclear characteristics and other factors. Enterprises still can not avoid producing ceramic tiles with various defects in the process of production and manufacturing. An urgent problem to be tackled is how to solve the foregoing problems and increase the accuracy of small target defect detection. This thesis mainly solves the above problems from two aspects: enhancing the original data set and enhancing the feature extraction ability of the network.

Considering that the image size in the tile defect data set is large, but the target defect is small, it is mainly solved from the two aspects of data enhancement of the original data set and enhanced feature extraction ability of the network. Firstly, the data enhancement is carried out by means of image cropping, and the experimental verification is conducted out using both machine learning and deep learning techniques, respectively, and the results show the effectiveness of data enhancement. Secondly, Faster RCNN is selected as the target detection network, the original network structure is replaced by the distracting network, and the network feature extraction ability is strengthened, and in order to further extract multi-scale image features, the recursive feature pyramid network(RFPN) module is introduced to fuse the high and low layer feature information, and due to the variable and irregular shape of the target defect, the deformation convolution network(DCN) is introduced for target detection. Finally, after experimental verification, the mean average accuracy and accuracy of the three network models based on the replacement characteristics of Faster RCNN are scattered network (network 1), the replaced Faster RCNN fusion RFCNN feature pyramid

(network 2) and the average accuracy and accuracy of the three network models based on the variable convolution Faster RCNN into the RFPN (network 3), and the experiment shows that after the optimized network, the mAP of network 2 is 1.61% higher than the mAP of network 1. This thesis shows that the introduction of the RFPN module extracts more image features, and the mAP value of network 3 is 1.23% higher than that of network 2, indicating that the detection accuracy of the network is improved after the introduction of DCN.

Considering that the prior art is limited in traditional industrial manufacturing, the slow and subjective nature of manual testing products will affect the accuracy of the test results. Based on the proposed improved model, this thesis designs and implements the tile defect detection system, including login registration, image detection and background management, which can identify and record the ceramic tile defect data collection of the imported system and promote the intelligent transformation of enterprises.

Key Words: ceramic tile defect; image classification; data enhancement; deep learning

目 录

第 1 章 绪论	1
1.1 研究背景及意义	1
1.2 国内外研究现状	2
1.2.1 机器学习方法的瓷砖瑕疵检测识别研究现状	3
1.2.2 深度学习方法的瓷砖瑕疵检测识别研究现状	4
1.2.3 深度学习与工业相结合的现状	4
1.3 本文的研究内容	5
1.4 本文的组织结构	6
第 2 章 相关理论及技术介绍	7
2.1 机器学习方法	7
2.1.1 图像特征提取	7
2.1.2 图像分类器	8
2.2 深度学习方法	10
2.2.1 卷积神经网络	10
2.2.2 ResNet.....	13
2.2.3 VGG.....	14
2.2.4 R-CNN.....	14
2.2.5 Faster RCNN	15
2.3 本章小结	17
第 3 章 数据准备与评价标准	18
3.1 瓷砖瑕疵图像数据集的构建	18
3.2 瓷砖瑕疵图像的数据增强	20
3.2.1 增强方法	21
3.2.2 数据增强	21
3.3 类别平衡重采样	23

3.4 数据集划分	24
3.5 模型评价指标	24
3.6 本章小结	25
第 4 章 基于改进 Faster RCNN 的瓷砖瑕疵检测	26
4.1 引言	26
4.2 网络结构	26
4.2.1 主干网络	27
4.2.2 RPN 层	27
4.2.3 ROI Pooling 层	28
4.3 改进的 Faster RCNN 模型	29
4.3.1 基于分散注意力网络(ResNeSt)的改进	29
4.3.2 融合 RFPN 模块的改进	32
4.3.3 利用可形变卷积(DCN)的 Faster RCNN 与 RFP 融合	34
4.4 实验结果与分析	36
4.4.1 实验平台	36
4.4.2 深度学习网络参数	37
4.4.3 深度学习网络训练过程	38
4.4.4 基于机器学习的实验结果分析	39
4.4.5 基于深度学习的实验结果分析	40
4.5 本章小结	43
第 5 章 瓷砖瑕疵检测系统设计与实现	44
5.1 业务流程梳理	44
5.2 瓷砖瑕疵检测系统设计	45
5.2.1 需求分析	45
5.2.2 可行性分析	46
5.2.3 系统开发运行环境	47
5.2.4 瓷砖瑕疵系统架构设计	48
5.2.5 系统功能模块设计	48

5.2.6 数据库设计	49
5.3 瓷砖检测原型系统实现	51
5.3.1 登录注册模块功能实现	52
5.3.2 图像检测模块	53
5.3.3 管理模块功能实现	54
5.4 本章小结	56
第 6 章 总结和展望	57
6.1 论文工作总结	57
6.2 未来工作展望	58
致 谢	59
参考文献	60

第1章 绪论

1.1 研究背景与选题依据

随着我国城镇化比率的提高,我国建筑陶瓷产业持续迅猛发展,成绩斐然。20 世纪末期,中国建筑陶瓷产品产量超越发达国家的瓷砖产量,连续 20 多年居世界之首。根据中国建筑卫生陶瓷协会数据显示,2021 年由于受到全球疫情的影响,年全国建筑陶瓷产量为 81.74 亿平方米^[1],同比下降 4.61%,但在我国江西、四川等地的产量都有不同程度的增长,新建的陶瓷生产线超过 200 条,陶瓷的出口量为 6.01 亿平方米,出口额达到 40.99 亿美元,同比下降 0.70%。近几年我国建筑陶瓷的发展趋势如图 1.1 所示,可以看出产量虽有所下降,但是我国仍是当之无愧的建筑陶瓷出口大国。

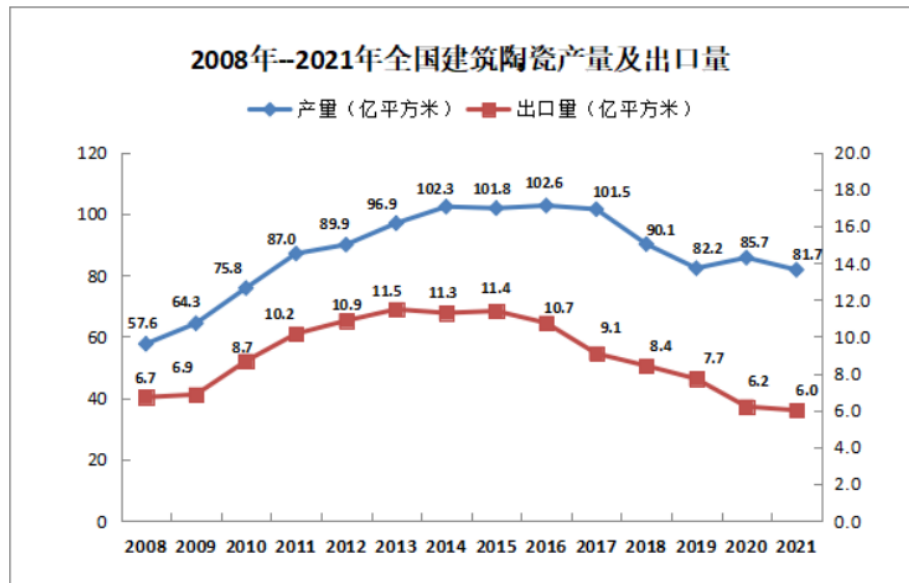


图 1.1 建筑陶瓷发展折线图

2020 年我国规模以上建筑陶瓷企业 1160 家,总产量 82.25 亿平方米^[2],且一直趋于上升的趋势。然而在瓷砖的生产过程中,从原材料的采集到最后制作成为成品一共需要多个工艺步骤,其中花纹面砖的工艺相对于其他瓷砖更是繁琐复杂。因此,制作过程中的一点疏忽就容易导致最后的瓷砖产生污点、划痕等瑕

疵,例如砖坯厚度相差太大和平整性较差等因素,或者其他工作上的疏忽都可能会导致瓷砖产生瑕疵,虽然生产环节已经基本实现智能化,然而瓷砖表面瑕疵检测环节依然过度依靠人工。通常来说,一条生产线大概需要配置 4~6 名工作人员,长时间在高强度下寻找瓷砖的表面瑕疵,视网膜容易受到刺激,容易出现视疲劳,这会导致检测的质量容易受到检测人员主观判断的影响,造成质检质量参差不齐,评价标准缺乏稳定性和成本居高不下等问题。

瓷砖表面瑕疵检测是瓷砖行业生产和质量管理中重要且关键的环节,也是限制企业智能化的技术难题,出于对上述原因的考虑,提出一种高效的识别检测瓷砖瑕疵的方法,对瓷砖企业发展而言具有十分重要的现实意义。

伴随着技术爆炸,人工智能和计算机视觉渐渐进入了人们的视野,计算机视觉技术在工业产业相结合中也取得了较好的成果。如果能够通过计算机视觉等技术手段来辅助瓷砖表面瑕疵质检的工艺流程,革新瓷砖质量检测流程,解决人工进行瑕疵检测的痛点和难点,不仅能够高效稳定的完成质检任务,降低检错和检漏的概率,而且还能提供客观且一致的检测识别结果,对瓷砖的瑕疵数据进行大数据分析和统计,这有助于瓷砖企业生产智能化进程更加顺利,然而由于瓷砖的瑕疵种类繁多,纹理组织复杂,观察识别技术困难较大,瓷砖表面瑕疵检测依然是困扰建筑瓷砖行业生产效率提高的技术难题。因此,通过机器学习的智能算法对瑕疵数据进行学习,探索出瓷砖瑕疵精确智能检测的优秀算法,辅助瓷砖生产过程中瑕疵检测识过程,减轻工人的负担,解决生产过程中的痛点和难点是很有必要的。

1.2 国内外研究现状

伴随着高速发展的人工智能和计算机视觉技术,计算机视觉与工业生产结合已经十分紧密。就瓷砖瑕疵检测而言,瓷砖表面瑕疵检测是瓷砖生产的重要一步,已经被科学家广泛研究并且取得了丰硕的成果。通过阅读相关论文,在瓷砖检测方向中主要有两个角度进行研究:一类是从机器学习角度出发的检测识别的方法,另一类是从深度学习角度出发的检测识别方法。

1.2.1 机器学习领域的瓷砖瑕疵检测识别研究现状

在瓷砖瑕疵检测模型中,核心部分是如何进行瓷砖图像的特征提取,并且在完成提取之后进行图像的分类。在传统机器学习方法中,B Zorić等^[3]学者提出了一种新的傅里叶频谱环空特征提取的方法,它是基于瓷砖表面图像的傅里叶频谱,并且在瓷砖行业的实际瓷砖中得到验证,特征提取方法优于其他方法。B Mariyadi 等^[4]学者提出基于灰度共生矩阵的特征提取,生成的特征作为人工神经网络(Artificial Neural Network, ANN)输入,构建了一个基于人工神经网络的瓷砖表面缺陷二维检测模型。Sioma 等人^[5]提出了使用激光三角法(LTM)构造瓷砖表面的三维图像,并且提出了用于图像滤波和缺陷识别的瓷砖三维表面图像分析算法。Eva Barreira 等^[6]提出了使用主动热像采集替换被动热像采集方法,并且在瓷砖的热图像中扩展了缺陷区域的分割和热像处理的自动化,最后在 facades 建筑瓷砖中进行了测试并且取得了良好的结果。权小霞等^[7]利用经过数据增强的图像和标准图像进行差分得到差值图像,通过在差值图像上进行特征提取来提取缺陷区域的轮廓,对缺陷细节的处理使用方法加权信息熵来修正,进而让缺陷的细节更加的突出,该算法在缺陷图像细节的保留和检测便利这两方面都有很好的突破。苏彩虹等^[8]受到纹理特征在空间分布上的特点启发,研究图像空间中两个像素之间的灰度关系,结合颜色统计特征构造了一个判断矢量,总结出一种用于墙地砖的缺陷检测的算法。Khaled 等^[9]为了提高检测瓷砖中斑点和裂纹缺陷所需要的时间,提出了将瓷砖图像划分为非重叠分区,识别缺陷分区然后仅对缺陷分区执行缺陷分类的并行算法,通过仿真实验表明并行算法的计算速度加快最终实现了 45 倍的提速。Luan 等^[10]提出了应用灰度共生矩阵,基于分割的分形纹理分析,离散小波变换,局部二值模式,以及基于分割的分形纹理分析和离散小波变换相结合的新方法,并且通过遗传算法对参数进行优化。Drlje 等^[11]提出基于离散小波变换(DWT)模拟人类视觉系统行为,提取故障检测的重要特征,和径向基概率神经网络(PNN)来根据缺陷的存在对瓦片进行分类方法。周向等^[12]首先使用中值滤波器对图像进行降噪处理之后,再提取图像轮廓,然后采用图像掩膜技术,抓取图像感兴趣的区域进行 MRI 图像的分割,最后采用斜外接的几何变换信息与标准位置的几何偏差作为反馈信息进行矫正,提出了图像缺陷检测领域中一种全新的定位分割方法。张怀亮等人^[13]在考虑到瓷砖纹理属性对瑕疵干扰的问题,在图像分割校正后,采用显著检测的方法对瓷砖表面进行

第一次检测后，对含有缺陷区域的图像字块进行二次检测，最终剔除缺陷。

1.2.2 深度学习领域的瓷砖瑕疵检测识别研究现状

近年来，深度学习模型在机器学习领域的迅速崛起，许多学者为瓷砖瑕疵检测与深度学习相融合提供了新思路。腾博文^[14]在 YOLOv3 目标检测算法的基础上结合卫浴陶瓷表面缺陷特征表现，对多尺度特征图先验框以及激活函数进行改进优化，并且在测试集上达到 94.90% 的 mAP。张涛川等^[15]通过图像增强的方式扩充样本数量，图像处理中使用双边滤波器，计算图像的梯度使用 Sobel 算子，提出了一种采用最大值融合策略的双流卷积神经网络模型，实现了 fc6 层的决策特征融合。戴玉昭^[16]通过残差神经网络 ResNet，谷歌深度神经网络 InceptionV3 以及密集卷积网络 DenseNet 在 ImageNet 数据集的预训练权重进行网络初始化，提取不同瓷砖瑕疵的底层特征，最后将三者训练得到的特征进行融合后，使用一个全连接层来提取顶层的特征，实现高层的分类。GuangWan 等^[17]在 YOLOv5s 目标检测方法中在骨干网中深化网络层，增加关注(CBAM)模块和小尺度检测层，在颈部网络中增强网络特征融合，用深度可分卷积替代原有的卷积，有效的解决了特性信息不足的问题。赵鹏^[18]利用通过聚类算法 K-means 解决实际裂纹的尺寸，融合路径聚合特征金字塔来获取高分辨率，提出了一种光伏电池缺陷检测的方法。孔兰芳^[19]在针对小目标的检测任务时，提出了一种采用 K-means 聚类方法的自适应的目标区域形态约束的方法，模型里同时采用反卷积的方法综合上下文的信息，进而提高小目标检测的精确率。

综上所述，它们都从不同的角度在瓷砖瑕疵检测这个方向中取得了一定的成果，但是同时也存在诸如瑕疵检测识别不稳定，精确度较低，开发成本较高的问题。本文从数据集的特点出发，选择了经典的两阶段目标检测算法，在数据增强的基础之上进行改进和验证，在相对较低的成本实现较高的识别准确率。

1.2.3 深度学习与工业相结合的现状

基于深度学习的表面异常检测已经成为一个热门的研究领域，对视觉检测的应用领域有着非常直接的影响。Weimer 等人^[20]应用了范围从 5 层到 11 层的网络来评估六种类型的错误，结果表明，深度网络算法的性能优于任何经典算法，并且能够在几个像素点的精度范围内定位误差，但是这种方法定位时的效率较

低。Masci 等人^[21]使用五层卷积神经网络对钢缺陷类型进行图像分类,证明深度学习的方法优于将手工工程特征与支持向量机(SVM)结合起来的经典机器视觉方法,但是他们的工作仅限于没有批量标准化的浅网络。Cha 等人^[22]在不同条件下进行实验,利用 CNN 网络结构来检测混凝土裂缝,得到的结果都比传统的 Canny 和 Sobel 边缘检测方法要好。Tabernik 等人^[23]提出由 11 个卷积和 3 个最大池化层的网络深度学习架构,来检测和分割表面的异常。Lin 等人^[24]提出了一种基于视觉注意机制和特征映射深度学习的鲁棒检测方法,采用帧间深度卷积神经网络(DCNN)策略消除误检,并且通过实验结果表明,铸件缺陷检测的误检率和漏检率均在 4% 以下,缺陷检测的准确率在 95% 以上,说明方法具有鲁棒性。刘金松^[25]在金属缺陷的缺陷检测中从优化特征提取网络,解决数据样本存在的问题等方面解决金属数据集中包含小面积缺陷较多的问题,提高了 Faster RCNN 的检测精度。张振宇^[26]在进行布匹瑕疵检测的时候在网络模型中引入了注意力机制和多尺度特征差异,网络提高了对未知花色和已知花色的检测 mAP。吕小倩^[27]利用对抗生成网络对水下图像进行处理,提出了基于 Faster RCNN 的水下前景分割的弱监督目标检测方法。李丹^[28]在卫星数据集上通过去雾,翻转等数据增强操作,训练一个良好的 Faster RCNN 目标检测模型。

1.3 本文的研究内容

本文将阿里云天池提供的瓷砖瑕疵数据集作为实验数据集,以提高瑕疵检测准确率为研究目标,提出了新的瑕疵检测识别方案。主要研究内容如下:

(1) 首先调研了瓷砖企业表面瑕疵检测识别该课题的选题背景与意义,掌握国内外研究现状,从传统方式和深度学习领域了解目前所取得的研究成果,深入分析以深度学习为基础的检测算法,探索与工业结合的发展前景。

(2) 以阿里云天池大赛提供的瓷砖产业园图像作为实验数据集,在输入模型训练前,需先对数据集完成数据预处理操作,归纳总结其特点并进行数据增强,主要采取了数据裁剪和类别平衡重采样等操作,保证数据的丰富性。

(3) 由于原始的 FasterRCNN 网络结构简单,泛化能力低,根据数据集本身特点,需对其进行改进。为提高特征提取能力,引入深度残差网络变体 ResNeSt,加深网络结构,并结合特征金字塔模块进行特征融合,获取图像多尺度特征。针对数据集信息的多样化,考虑使用可形变卷积代替普通卷积,提高检测精度。

(4) 为验证本文提出算法的有效性,解决人工质检费时、效率低等问题,本文设计并实现了瓷砖检测识别系统,便于用户和管理人员更好的开展工作,梳理瑕疵检测业务,同时验证了本文提出模型的有效性。

1.4 本文的组织结构

本文研究内容大致划分成六章,详细内容安排如下:

第一章:引言。首先详细说明了我国建筑瓷砖企业在表面瑕疵检测识别工艺流程中所面临的处境,道出了本文的选题意义,并且从国内和国外两个角度,描述了机器学习和深度学习在瑕疵检测识别上所积累的成果,简要概述了深度学习与工业相结合的现状。

第二章:相关理论及技术介绍。首先介绍了传统的机器学习方法在图像分类中的一般方法和步骤,其次介绍了深度学习相关基础知识,重点介绍机器学习中的 SVM 分类器,以及深度学习中的 ResNet 等网络的架构和特点,最后介绍了第四章实验用到的卷积神经网络模型 Faster RCNN。

第三章:数据准备与评价标准。本章主要对实验数据集合的来源进行基本的介绍,分析数据集的原始标签和特点,针对其中突出的目标较小和样本不平衡问题,进行数据裁剪和类别平衡重采样。

第四章:基于改进 Faster RCNN 的瓷砖瑕疵检测实验。本章主要包括基于 Faster RCNN 的改进和实验验证两个部分。首先详细分析了 Faster RCNN 网络的组成部分,同时创造性的加入 RFPN 和 DCN 等模块,替换了主干的特征提取网络为 ResNeSt,经过消融对比实验,改进后模型的实验结果表明,准确率和精度较优化之前的网络模型均得到了提升。

第五章:瓷砖瑕疵检测识别系统的设计与实现。本章首先对系统的业务进行基本的梳理,然后按照软件工程中系统设计理论对系统进行设计开发,最后进行主要功能的实现,并进行成果展示。

第六章:总结与展望。本章节对本文所做的工作进行总结性的梳理,并且分析本文所做的工作不足,为下一步的工作提供切入角度。

第2章 相关理论与技术

2.1 机器学习方法介绍

一般是首先对图像进行预先处理，瑕疵定位，进而进行特征提取，最后选择合适的分类器基于特征进行瑕疵分类，它们都属于模式识别的范畴。一般来说，其中核心步骤是提取特征和瑕疵的分类。图像预处理和特征工程共同决定了机器学习的最终识别效果，最后选择合适的算法不断训练进一步优化检测效果。

传统的机器学习方法主要包括选取理论基础、定义手工提取特征(Hand-crafted)方法以及选择分类器，具有较好的可解释性，但是该学习方法过度依赖于特征，手工设计的特征存在较大的局限性且难于设计等问题。因此，传统机器学习的模型性能的差异主要是由特征提取和分类器这两个因素所导致。机器学习图像瑕疵分类流程如图 2.1 所示。

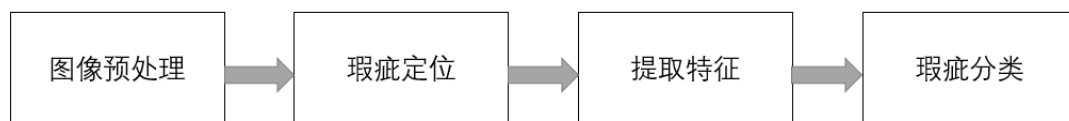


图 2.1 机器学习瑕疵分类流程

2.1.1 图像特征提取

特征是一事务异于其他事务的本质特点，或者是由这些特点所组成的集合。从图像出发，图像特征可以是图像中的任何东西，任何对应用有用的东西，例如建筑物中的边缘轮廓就是可以消除透视形变的特征，除了边缘特征之外还有点特征等。图像的特征有人为特征和自然特征这两大类，其中，自然特征是指不需要人们通过后期的挖掘就直接呈现出的特征，其中主要包括图像的颜色、轮廓、边缘和大小等；人为特征则是需要采取一定的技术手段才能获取到作为图像的特征，其中比较典型的有灰度直方图，矩阵征和熵。需要注意的是，特征信息是不容易受随机因素波动的。因此，图像特征提取是指将图像中所包含的丰富信息，通过一定的方法挖掘出来作为图像的特征信息。

通过原始数据进行模型的训练不仅仅耗时而且耗力，同时也没有达到预期的效果，因此，特征提取成为模型训练的关键一步。从传统机器学习的角度出发来讨论图像特征分类，图像特征通常有空间关系、纹理形状、颜色和统计等。下面简要描述提取这几类特征的方法。

(1) 空间特征主要包括点线面特征。其中，点特征提取的方法主要有 Forstner 算子等；线特征提取的方法有梯度算子、Roberts 算子以及 Prewitt 算子。

(2) 纹理是一种非常普遍的视觉特征，它的定义是物体表面的许多相同局部模型重复的结果，常用的特征提取方法主要有模型法和信号处理方法等。

(3) 颜色特征的常见方法主要有颜色集和颜色聚合等。

传统的机器学习方法在缺陷识别领域虽得到一定的效果，但是大多数时候苦于对数据集的要求比较高，在复杂的场景特征提取的效果不是很好，不利于高效的应用。

不难发现，传统的目标检测方法在已经确定了物体可能位置的时候，需要通过以上的方法来进行特征的提取，主要依靠人工设计提取器进行，一般有方向梯度直方图、局部二值模式^{[29][30]}和尺度不变特征变换法。与此同时，深度学习一般是通过网络自己去学习到图像的特征，再用于分类和回归的预测，每个网络都会选择合适的特征提取网络进行特征提取。现在热门的特征提取网络主要是 ResNet 网络和它的相关变体。

2.1.2 图像分类器 SVM

机器学习中用于分析的工具一般有贝叶斯分类器、随机森林以及支持向量机(SVM)等。本节简要介绍 SVM 分类算法。

随着统计学的兴起，支持向量机于 1995 年被 Corites 和 Vapnik 等人^[31]提出。SVM 的主要特点是使用核函数计算映射后的向量内积，降低了算法的时间复杂度，能够将非线性问题转换成为线性问题，具备良好的通用性，鲁棒性和有效性。作为一种经典的二分类模型^[32]，其目的是找到最佳的分类超平面，划分出两类特征数据，并使用间隔最大化分类策略，寻找凸二次规划问题中的最优解。最终模型训练结束时，仅保留与支持向量相关的数据，其他数据则被丢弃。因此，SVM 能够胜任目标分类任务，也是机器学习中常用的分类模型。结构如图 2.2 所示。

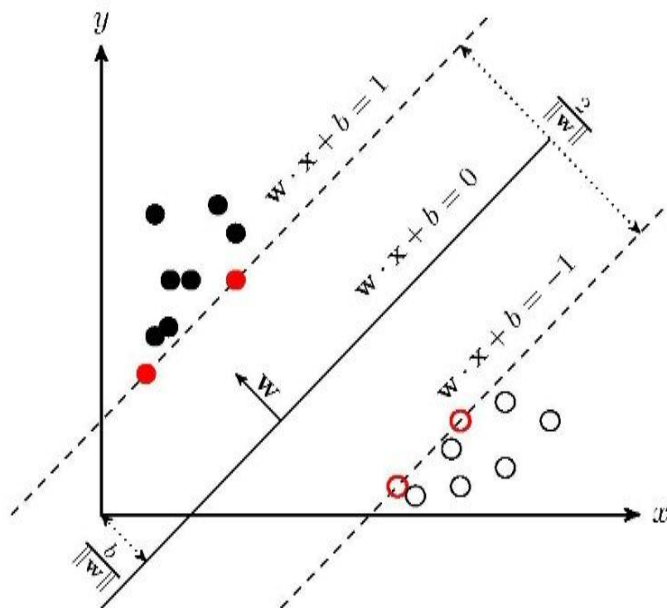


图 2.2 SVM

正如图 2.2 所示，上图是一个线性可分的 SVM，决策边界为实线，间隔边界为虚线，间隔是两个虚线之间的距离，黑白点代表的是两种样本，其中支持向量是红点。

公式 2.1 中的 w 代表 n 维法向量的超平面， b 是一个常数代表的是超平面的截距。SVM 一般是用来解决线性可分的问题，线性可分指的是存在任意的 w 和 b ，都能满足训练集合 $D=\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ ，满足 $y_i=+1$ 时， $w^T x + b \geq 0$ ， $y_i=-1$ 时， $w^T x + b \leq 0$ ，将两式综合表达公式如 2.1 所示。

$$y_i(w^T x + b) \geq 0 \quad (2.1)$$

由于当 $a \in R^+$ 时， $w^T x + b = 0$ 和 $aw^T x + ab = 0$ 为同一个平面，所以最终在所有支持向量 x_0 上，公式如 2.2 所示。

$$|wx_0 + b| = 1 \quad (2.2)$$

并且向量 x_0 到超平面的 $w^T x + b = 0$ 的距离，公式如 2.3 所示。

$$d = \frac{|w^T x_0 + b|}{\|w\|} \quad (2.3)$$

那么支持向量机到平面的距离，公式如 2.4 所示。

$$d = \frac{1}{\|N\|} \quad (2.4)$$

最大化 d 的问题可以最小化为 $\|w\|^2 / 2$ 问题，但是与此同时样本到超平面的距离大于等于 d ，则 $|w^T x_i + b| \geq 1$ ，其中 i 的取值范围为1到 n 。因此，SVM就可以转化为凸优化的问题。

2.2 深度学习方法介绍

机器学习规模较小，并且机器学习过分依赖挖掘数据的先验知识和人工设计特征，这种方式过程复杂并且对设计者的技术要求较高^[33]，但是深度学习(Deep Learning, DL)对数据的先验知识较弱，而是通过层次表达学习到数据的规律性，因此得到广泛应用。深度学习最初通过图像识别技术展现在人们的眼前，解决了机器学习所不能解决的问题，它利用设定好的多层网络结构，通过模拟人脑，能够自动的学习和提取数据的特征，在图像识别领域取得了瞩目的成果。下面介绍卷积神经网络的组成结构以及其他的相关卷积神经网络。

2.2.1 卷积神经网络

卷积神经网络(Convolutional Neural Networks, CNN)来自于 Hubel 和 Wiesel 对哺乳动物大脑中视觉系统的研究^[34]，在了解到哺乳动物视觉皮层是有一定层次的深度结构后，人们便试图通过构建一组数学公式来表达这种抽象的过程，更进一步的发现视觉系统中，通过采用划分清晰的层次结构来进行信息的获取。主要表现为浅层中一般是用来提取物体边缘的特征信息，深层的网络则可以提取到图像的抽象信息，更能反映出图像的高级语义信息。深度学习就是通过参考这种层次堆叠的体系进行构建网络。

卷积神经网络就是一种多层网络结构的深度学习模型，它通常由卷积层，池化层，非线性映射层等构成，在输入端和输出端之间建立一种非线性的映射来无限逼近视觉系统机制。与传统机器学习浅层网络感知域较小相比较，它具有较大的感知域，能够学习到更加抽象的特征。与传统的方法相比，CNN 巧妙的利用权值共享的方式大大的简化了网络结构，使得模型在多个应用场景的适应能力较好^[35]。它具有池化操作，权值共享，局部连接，以及多层结构的四个特点^[36]，

并且网络结构的深度使得它具有很强的表达能力和学习能力^[37]，从而能够处理更为复杂问题。

卷积神经网络在图像处理领域的产出成果较为丰富，当下卷积神经网络在瑕疵检测领域主要作为主干网络来进行特征提取，它能够帮助瑕疵检测网络更快的分辨出目标。当前应用的比较多的特征提取网络主要有 VggNet, ResNet 等，也有新的网络在不断的推出，卷积神经网络如图 2.3 所示。

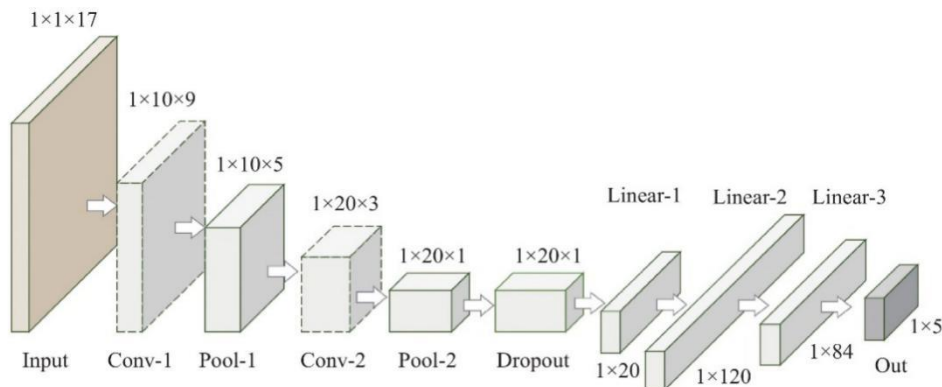


图 2.3 卷积神经网络

(1) 卷积层

卷积层是用来进行图像的特征提取，它是识别图像核心的部分，对一张图像的感知首先是通过局部感知，再通过深层的网络结构对已经获取到的信息进行处理。低层和高层的卷积分别承担不同的卷积提取任务，低层提取图像边缘、线条等细节信息，高层提取图像空间、几何等语义信息。卷积核表现为一个权值矩阵，如图 2.4 所示。图像经过处理之后得到 5×5 的矩阵，接下来我们用 3×3 卷积核（滤波器）对图片进过滤，就是求卷积的过程。

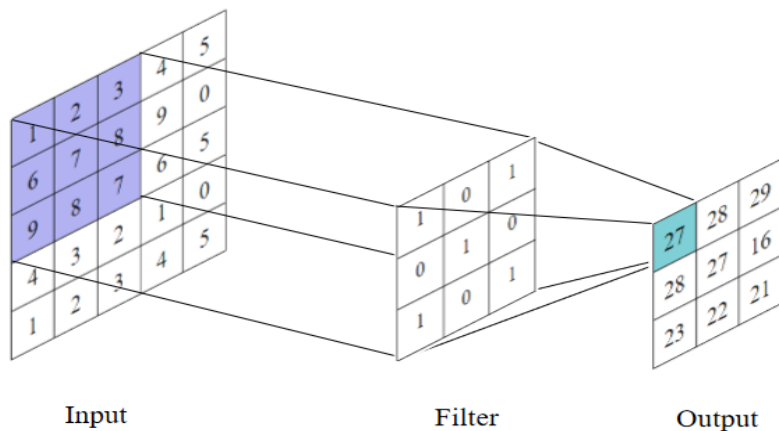


图 2.4 卷积运算

(2) 激活层

激活层主要是用来保证网络的非线性，根本在于线性函数在经过计算之后依然是线性，卷积层的输出结果通过激活层进行非线性的转换，主要作用是将神经元输出映射输出端，使得目标检测模型的特征表达更加具体生动。在神经网络开始的阶段使用 Sigmoid 作为激活函数，但是由于后面出现的网络，网络的层数在不断的加深，Sigmoid 函数在层数不断加深的情况下会出现梯度消失，后来出现的 Tanh 函数在 Sigmoid 函数的基础之上加快了模型收敛的速度，为了避免出现输入较大的时候会出现网络更新的速度较慢的问题，经过多次迭代之后目前常用的激活函数是 ReLU。如图 2.5 所示。

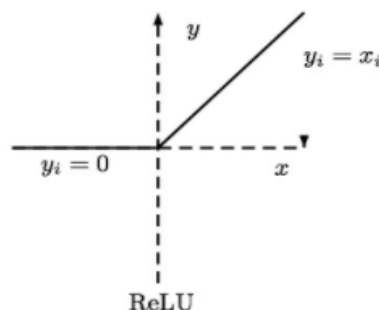


图 2.5 修正线性单元

(3) 池化层

池化通常也称为下采样，主要用于防止过拟合，突出重要特征，减少网络的计算量，增加非线性的表达能力，增加模型的泛化能力和提高模型的容错性。主要有最大池化如图 2.6 和平均池化如图 2.7 两种方法，通过池化使得模型中的神经元个数骤减，输入空间具有平移不变性也更具有鲁棒性^[38]，让 CNN 训练起来更加高效。

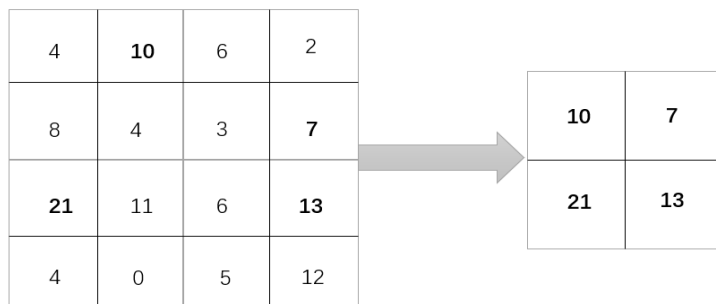


图 2.6 最大池化

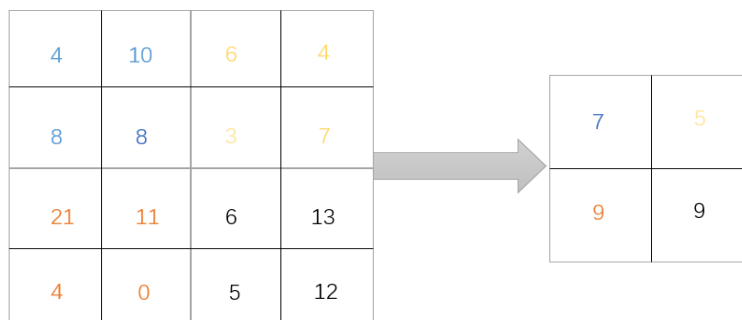


图 2.7 平均池化

(4) 全连接层

全连接层就是将最后一层卷积得到的矩阵展开为一维向量，并为分类器提供输入，它的主要作用就是将卷积层和池化层计算得到的特征空间映射样本，标记空间，达到弱化特征位置对分类结果干扰的目的，提高整个网络的健壮性。

2.2.2 ResNet

残差网络是微软亚洲研究院何恺明等人^[39]在 2015 年提出。在 ImageNet 图像分类性能上超过人类水平，它主要解决了随着网络的层数越来越多的时候出现的网络退化现象。作者何恺明通过一种恒等映射的方式如图 2.8 所示来同等的构造深层次的网络结构。

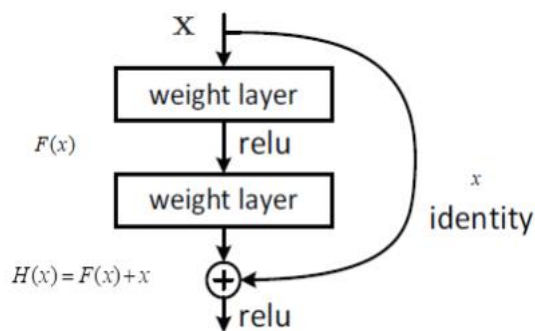


图 2.8 恒等映射

ResNet 之所以能够解决网络退化的原因，首先是恒等映射能够将顶层的信号传到底层，也能将深层的信息传回到浅层，其次是学习到类似于长短记忆神经网络(LSTM)的优点，残差网络能够做到对重要的模块进行输出，不重要的进行抹平的操作。最后是传统的线性神经网络很难拟合线性的网络，但是残差网络中

的 Skip Connection 能够让模型自行选择是否需要更新，同时弥补了高度非线性造成的不可逆的信息损失。ResNet 还做到了并行的多潜在路径的集成，也就是去掉中间的几个残差块并不会影响性能。

2.2.3 VGG

VGG 是由牛津大学提出具体模型^[40]，详情如图 2.9 所示。VGG 具有两个经典的网络结构 VGG16 和 VGG19，它们最大的区别是在网络层数的不同，常作为迁移学习的骨干网络，将 LetNet^[41]和 AlexNet 奠定的经典串行卷积神经网络结构的深度和性能发挥到极致。

最深的 VGG 网络保留了原生的串行网络结构，它的整体结构采用了同样大小的卷积核尺寸(3×3)和最大池化尺寸(2×2)，这样构造的网络结构较为简洁。通过 2 层的 3×3 的卷积替换了 5×5 的卷积，使得模型的非线性次数变多，模型的学习能力和表示能力变强。同时模型像素层面的长宽度越来越小，语义通道越来越长，进行迁移学习，将像素信息成功转换为语义信息。除此之外还通过将其中一层全连接层转换为卷积层，以此来适应不同的图像尺寸。网络的缺点是耗费更多的计算机资源，并且使用了更多的参数，导致与其他网络结构相比更占内存。

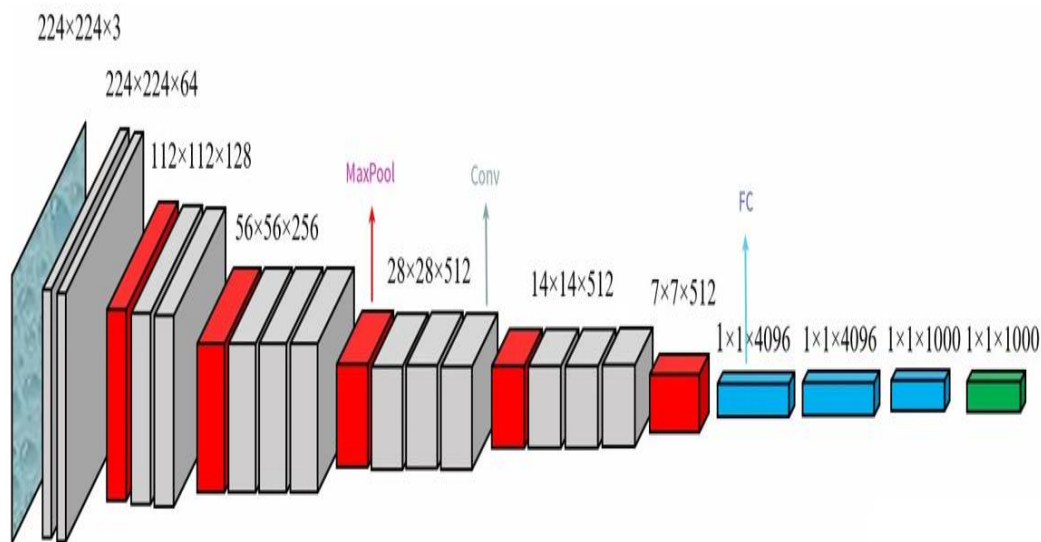


图 2.9 VGG 结构图

2.2.4 R-CNN

R-CNN(Regions with CNN features, R-CNN)是 Two-stage 目标检测算法，提

出了两阶段目标检测算法的大致流程，为后面的 Faster RCNN 算法的创新起到奠基作用^[42]，大致流程如图 2.10 所示。首次将深度学习和卷积神经网络相结合用于目标检测，最终在 mAP 指标上得到了大幅度提升，但是也存在速度慢，结构松散等缺点。后续的 Fast RCNN、Faster RCNN 目标检测算法都是在 R-CNN 的基础之上进行的改进。

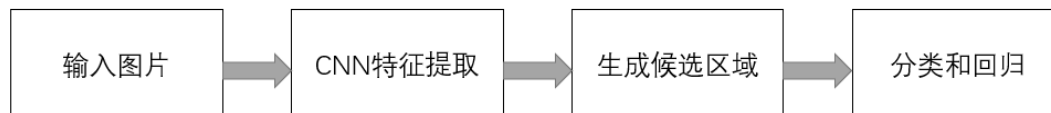


图 2.10 Two-stage 基本流程

R-CNN 算法训练过程如图 2.11 所示，大致分为以下几步：

(1) 对每张输入的图像使用区域性搜索的算法。就是使用一种过分割的方法，将一张完整的图片分割成比较小的区域，然后计算相邻区域之间颜色、纹理和尺度的相似性，不断的进行迭代合并区域直到整个图片变成一个区域。最后，生成多个有层次结构的提议区域，为每个提议区域生成物体类别和真实边界框。

(2) 选取一个预先训练好的卷积神经网络，对每个提议区域将其尺寸大小变成卷积神经网络需要的输入尺寸之后，进行前向计算到第五层，抽取这一层的特征。

(3) 使用支持向量机对特征进行分类输出。

(4) 针对特定的类别训练一个线性回归模型来预测真实边界框。

(5) 进行非极大值的抑制，计算交并比剔除重复多余的框。

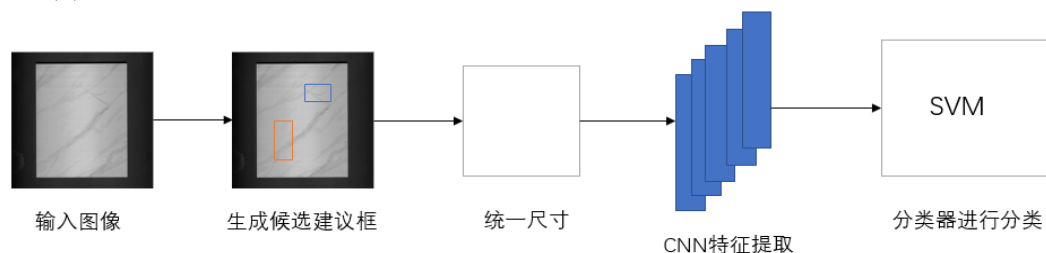


图 2.11 RCNN 网络处理流程

2.2.5 Faster RCNN

Faster RCNN 是 Kaiming 等^[43]提出的一种基于区域的快速卷积算法目标检测网络方法，Faster RCNN 采用了与 Fast RCNN^[44]相同的设计详细流程图。让卷

积网络建议框和目标检测网络共享卷积网络，这一创造性的改变使得提议区域的数目得到骤减，同时提议框的质量得到了本质的提升。**Faster RCNN** 算法的基本流程框图如图 2.12 所示。

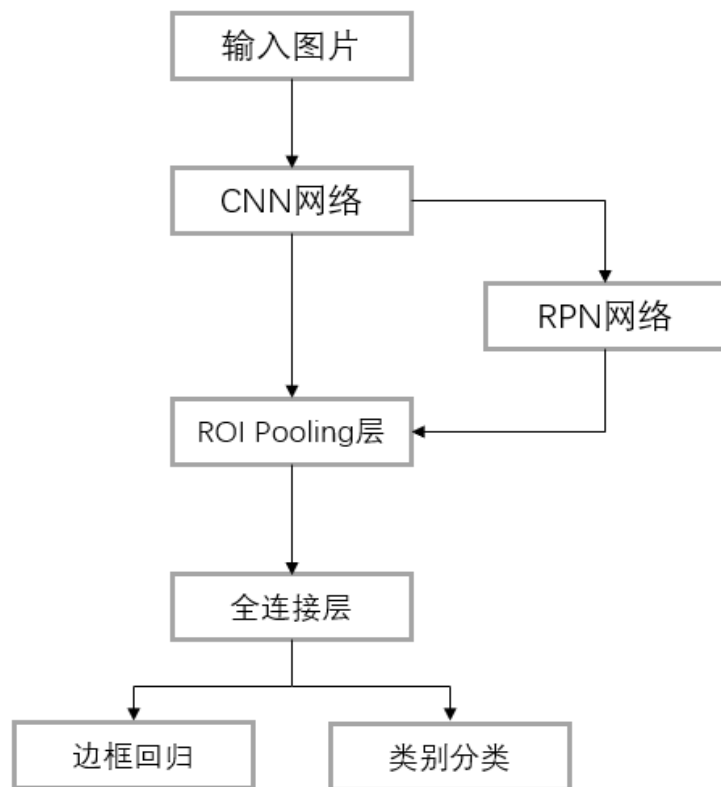


图 2.12 Faster RCNN 算法框架图

Faster RCNN 将特征抽取、建议框提取、边界框回归和分类回归层融合在一个网络中。与其他的目标检测算法相比，**Faster RCNN** 采用 **RPN** 的网络结构生成检测框，直接通过预测出候选框，然后产生推荐区域，同时很大比例的候选框预测工作在 **GPU** 中完成，加快了网络训练的速度^[45]。**RPN** 不仅能参与进行共享卷积层，还能通过随机梯度下降和反向传播进行端到端的训练，使得模型的综合性能有了较大的提高，将提议和回归进行分离开来进行操作，所以 **Faster RCNN** 是典型的二阶段目标检测算法，在小目标检测上发挥着巨大的作用。

2.3 本章小结

本章节主要介绍了后面章节进行对比实验在深度学习和机器学习领域的相关前置知识，传统机器学习主要介绍图像特征提取和图像分类的基本方法，深度学习主要以卷积神经网络为切入角度进行介绍，除此之外，还介绍了 Faster RCNN 的核心模块，例如像 ResNet，VGG，R-CNN 等经典网络。

第3章 数据准备与评价标准

3.1 瓷砖瑕疵图像数据集的构建

本文使用的瓷砖瑕疵数据集是由阿里云提供的广东佛山瓷砖产业园图像数据集，在生产线上经过专业的人员架设设备进行实地采集而来，数据集中涵盖到瓷砖生产线频繁出现困扰企业发展的常见瑕疵，图像的编码均为 JPG 格式。在进行训练的时候需要将数据集格式全部转换成为 COCO 格式，主要包括粉背景、边异常和白色点瑕疵等七种瑕疵的分布，如图 3.1 所示。

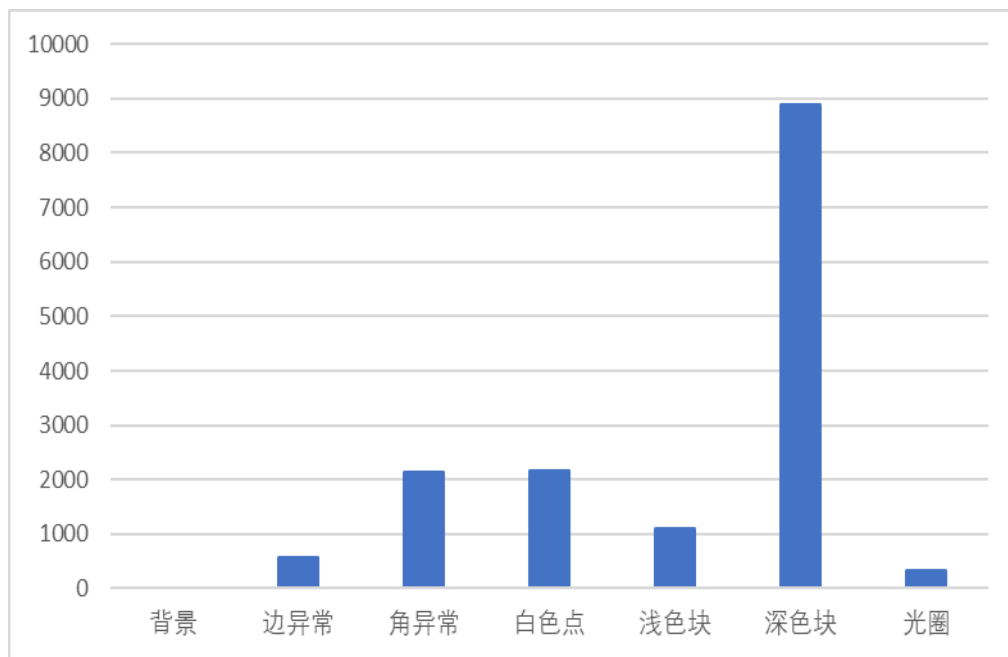


图 3.1 瓷砖瑕疵数据训练集缺陷样本 BBox 数量

图 3.1 的纵坐标为瑕疵目标包围框(BBox)，从图上很明显的可以看出 BBox 数量不均衡深色块有 8000 多，而光圈只有 300 多，边异常有 500 多。

由于某些缺陷只有在一定的角度和特定的光照情况下才能拍摄到，每块砖拍摄了三张图像，分别用相机的编号作为后缀进行命名。三张照片分别从不同的角度拍摄了黑白图和彩色图。具体照片效果展示如图 3.2 所示。

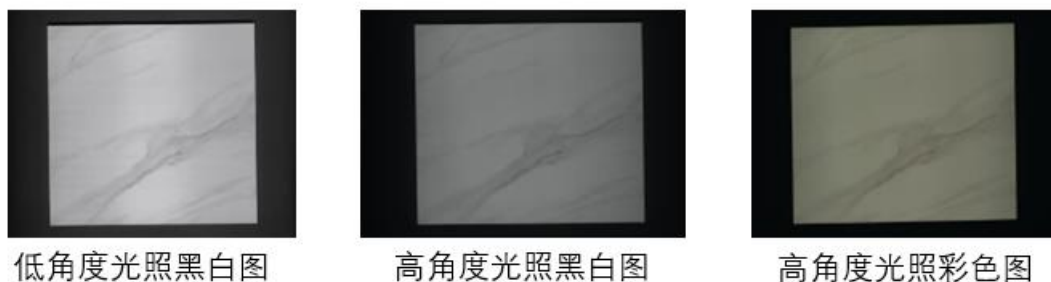


图 3.2 数据集采样照片

对数据集挖掘出以下显著的特点：

(1) 数据集中训练集的图像总共有 5388 张，测试集中的图像有 1762 张。训练集通过 JSON 的数据格式进行给出，主要包括姓名、高度、宽度和种类标瑕疵框(BBox)等字段。图像的平均高为 5562.21，图像的平均宽度为 7474.73。

(2) 每张图像中包含的瑕疵呈现出长尾分布的状态。与此同时，由于大部分包围框非常小，所以占比也比较小，包围框比也呈现一个长尾的分布，如图 3.3 所示。主要包括物体的大小分布不均匀、小目标占比比较多和长宽比例和瑕疵尺度变化大等特点，例如划伤的最大框为 2000×1800 ，最小框为 7×7 ，白色点的最小框为 5×2 。每张图的目标包围框(BBox)的数量大部分分布在 1 到 5 个，少量的图分布在 10 个以上。瑕疵类别分布不均匀，例如深色点瑕疵和边异常的比例为 27: 1。

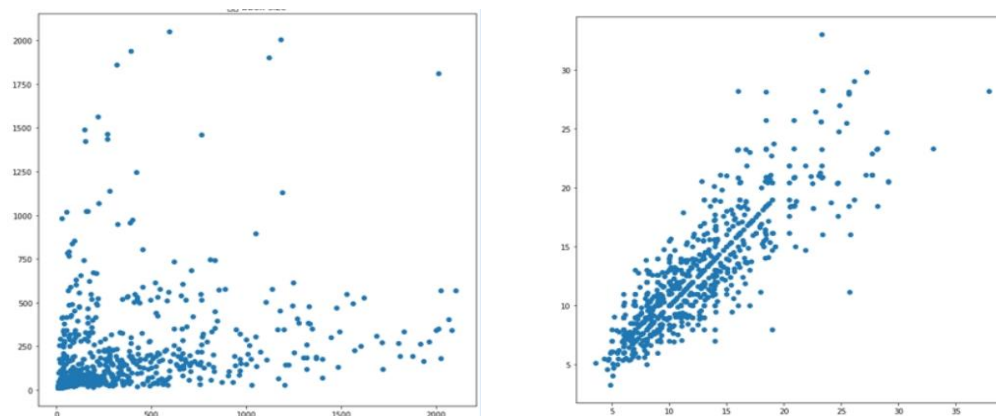


图 3.3 瑕疵点分布图

(3) 图像的分辨率较大且分布在 6000×8192 不等，但是每个检测的目标都很小。图像目标包围框分布如图 3.4 所示，图像中最大的目标包围框(BBox)是 404813.20，最小的目标包围框(BBox)是 8.26，不难发现这是瓷砖瑕疵检测所面

临的一个难点。

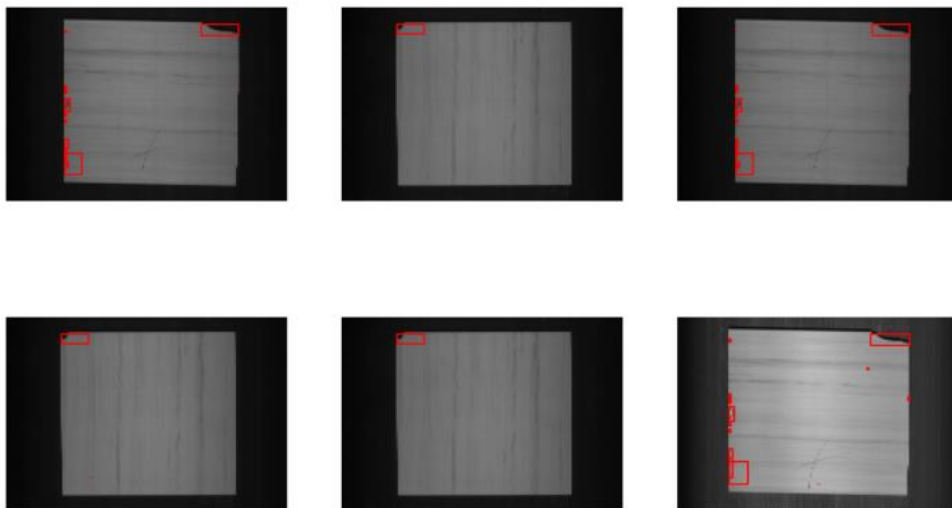


图 3.4 瑕疵目标包围框样例

3.2 瓷砖瑕疵图像的数据增强

数据信息的挖掘是深度学习的前导性工作，一般通过数据的预处理来挖掘出对深度学习可能有用的数据特点。数据样本多样性的好坏决定了训练模型的泛化能力好坏。深度学习为了增强模型的泛化能力的主要方向就是通过增加数据样本的多样性和数据量来强化模型的泛化能力，常见的方式是人为的添加图像噪声进行干扰。例如人为添加瑕疵，或者是对目标检测对象进行增强，弱化周围干扰因素，对图像进行降噪，去雾等。

目前常见的数据增强^[46]的方法有图像翻转和图像裁剪等。由于本章节的数据集存在瓷砖瑕疵类别不均匀和图像尺寸过大瑕疵太小等问题，计划对图像集的图像做一系列的改变，进而提高模型的准确率和泛化能力。下面对数据增强的方法从传统领域深度领域进行对比最后结合数据集的特点选择比较合适的数据增强的方法。

3.2.1 增强方法对比

(1) 传统领域常见数据增强的方法

① 颜色增强(Color Jittering): 主要通过改变图像亮度、饱和度和对比度的方式来增加数据量。

② 随机裁剪: 以目标为中心进行裁剪, 常用的方法有中心裁剪、随机裁剪和随机长宽比裁剪。

③ 旋转、翻转: 旋转是指以图像的某点为中心进行转动, 角度一般取 90 或者 180, 否则会出现边缘缺失或者超出问题, 翻转不同于旋转 180 度, 而是和人在镜子中映射类似, 比较常用的是概率水平翻转和概率垂直翻转。

④ 添加噪声: 高频率特征学习可能会导致网络的过拟合, 为了避免这种情况的发生, 可以随机加入噪声数据来抵抗这种因素。使用比较广泛的有 Gaussian 噪声^[46]和椒盐噪声^[48]。

⑤ 尺度变换: 通过不同的尺度因子对图像进行一定的尺度缩放, 通常使用的尺度因子有 0.8、0.9 和 1.1 等。

⑥ 平移: 平移指的是对图像沿着 x 轴或者是 y 轴进行移动, 在平移的过程中需要对平移的背景进行假设。

(2) 使用神经网络进行数据增强的方法

① 神经风格变化: 会对损失函数进行一些操作, 操控 CNN 的表达从而改变图像的原始风格。

② 对抗训练: 对抗训练是提升图片鲁棒性的一种方法, 例如 FGSM(fast gradient sign method)算法^[49], 通过添加人眼不可见的扰动, 将加入扰动后的图像作为样本加入数据集中来提高鲁棒性。

③ 基于 GAN 的数据增强的方法^[50]: 常用的是 DCGAN 一类的 GAN, 它训练比较稳定, 非常适合图像分类任务。

3.2.2 数据增强

传统领域数据增强的方法与神经网络方法相比, 工作量较小, 适合比较常规数据集的增强, 神经网络方法工作量较大, 相对传统方法效果比较明显, 但是网络训练时间加长。在进行数据分析的时候, 发现图像的分辨率较大但是瑕疵检测

点较小,并且背景浓厚和图像中标注缺失等特点,本文打算采用传统领域数据增强的数据裁剪方法对数据集进行增强。

(1) 瑕疵特征对于整个场景图来说占比不大,如果通过将原始瓷砖的特征利用 ROI-Pooling 层提取,然后进行卷积运算会导致小目标丢失,训练效果不是很理想。因此本文打算对数据集中的所有图片进行裁剪的操作操作如下:将原图像尺寸为 6000×8192 的图像通过离线裁剪的方式切成 640×640 的大小,为了防止图像裁剪的时候目标检测对象被裁剪,受到卫星图像目标检测的启发^[50],取图像的重叠比例为 20%,从原图的左上角开始滑动,将切出来图像的左上角标记为 x 坐标和 y 坐标,切出来的图像 y 坐标依次是 0, 512, 1024 等。需要注意的细节是,在切图的最后一步要考虑到原始图像的高度限制,需要以图像的高度反向计算最后一次的 y 坐标,这是整个切图环节的难点。切图可视化如图 3.5 所示。

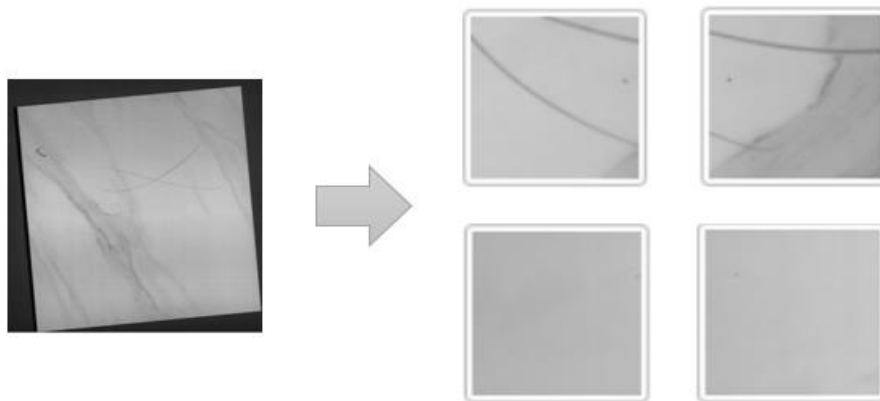


图 3.5 切图可视化图

(2) 根据切图的整体思路,切图完成之后的目标标签 x_{\min} , y_{\min} , x_{\max} , y_{\max} , 要分配给每个切出来的小图(640×640)剩下的图片当作背景,然后过滤掉。

(3) 在结合模型考虑切图的时候有两种策略可选,一种是在线切图使用,也就是使用训练的时候进行切图,可以使用的一些成熟的框架,比如 MMDet 中封装好的库函数来进行裁剪,但原图太大,在线切图的时候会消耗太多的 CPU 时间,增加训练时长。与之对比的是离线切图,通过离线切好图之后再输入到模型中。切完图像可以通过可视化证明标签的变化是正确的如图 3.6 所示。

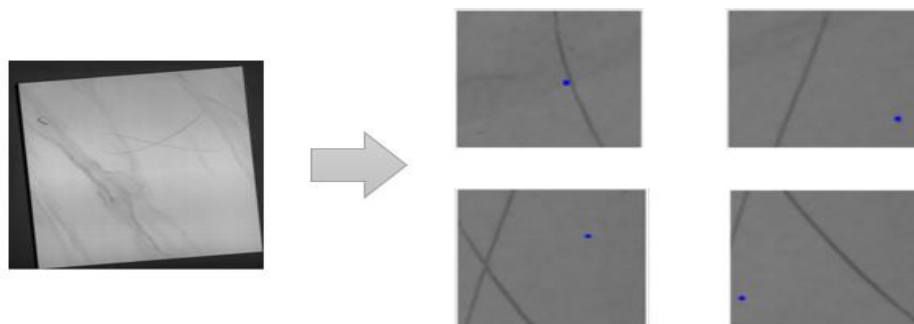


图 3.6 标签不丢失图

3.3 类别平衡重采样

通过分析数据，发现不同类别间样本数不平衡问题十分的严重，不平衡的问题会导致模型在数据样本丰富类别上过拟合，样本数量比较少时提取特征不是很明显，很难捕捉到强有效的特征信息。本小节使用重采样的方法解决这个问题。

瓷砖瑕疵数据集的重采样的主要通过增加稀有样本数的上采样和减少大类样本数的下采样这两种途径解决类别不平衡的问题。本文中对训练样本按照类别进行重采样，具体步骤分三步如下：

- (1) 计算各类别出现的频率发 $f(c)$ 。
- (2) 计算每个类别的重采样系数 $r(c) = \max(1, \sqrt{0.1/f(c)})$ 。
- (3) 对训练集的每张图片，计算重新采样的次数 $r(I) = \max r(c)$ 。

数据经过重新采样后的前后对比图如图 3.7 所示。

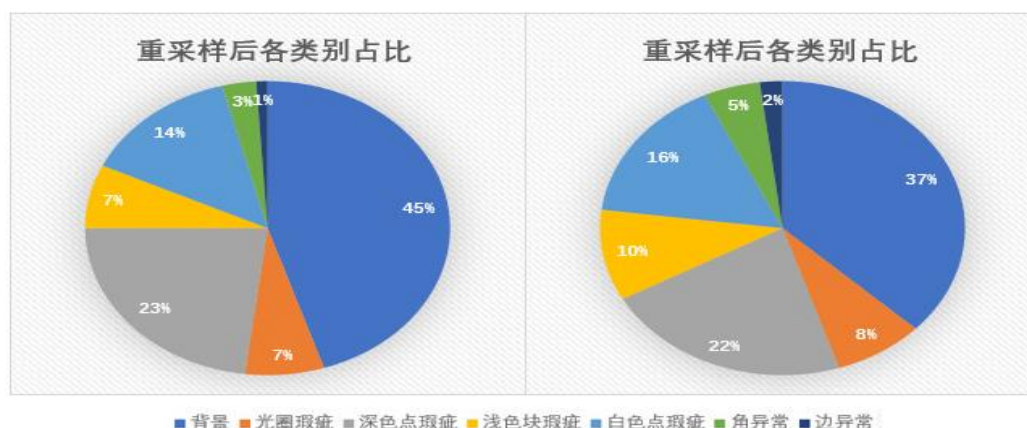


图 3.7 类别平衡重采样

3.4 数据集划分

经过上述的数据平衡重采样，解决了数据集合之间的不平衡性，同时经过数据增强的操作之后，数据图像的总数从原来的 5388 张扩展为 38340 张，按照 8:1:1 的比例划分之后训练集总共有 30672 张，验证集有 3834 张，测试集有 3834 张图片。

3.5 评价指标

在深度学习中，有很多评价指标从不同的角度来评价模型的性能，不同的指标的应用场景不同，选择合理的评价指标非常重要。在模型经过训练完成后，本文采用以下几种常见的目标检测评价指标来度量模型。

(1) 准确率(Accuracy, ACC)

准确率是模型预测正确数量所占总量的比例，反应了模型对样本总体的判断能力。其计算公式如 3.1 所示。

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \quad (3.1)$$

其中：

TP(True Positive): 预测是真，真实值也是真。

TN(True Negative): 预测是假，真实值也是假。

FP(False Positive): 预测是真，真实值是假。

FN(False Negative): 预测是假，但真实值是真。

(2) 查全率(Recall, R)

查全率又名召回率，它具体含义是查询出来为正样例占查询总数的比例。其计算公式如 3.2 所示。

$$R = \frac{TP}{TP+FN} \quad (3.2)$$

(3) 查准率(Precision, P)

查准率，它和查全率是一对矛盾的度量，两者成反比关系，查准率越高查全率越低，具体表达含义是在正确样本中有多少个样本被预测出来了。用 P 表示，

其计算公式如 3.3 所示。

$$P = \frac{TP}{TP+FP} \quad (3.3)$$

(4) 平均精确率(Average Precision, AP)

平均精确率是对不同召回率点上的准确率进行平均, 在 PR 曲线图上表现为 PR 曲线下方的面积, AP 的值越大, 表示模型的性能越好。

(5) 均值平均精确率(mean Average Precision, mAP)

均值平均精度是所有类别的平均精度求和除以所有的类别, 是衡量模型在所有类别上的好坏, 是目标检测中一个非常常见的指标, 评估一个目标检测模型通常就是查看这个评价指标。mAP 的值分布在 0 和 1 之间, 越大越好。其计算公式如 3.4 所示。

$$mAP = \frac{1}{|Q_R|} \sum AP(q) \quad (3.4)$$

(6) 时间指标

时间指标是指模型预测一张图像所耗费的时间, 时间指标需要结合其他的指标一起来评价模型的好坏, 通常情况下检测时间越短, 说明模型的性能越好。

3.6 本章小结

本章对训练时的数据集进行构建并分析了数据集的特点, 对数据增强的方法进行了总结, 从数据集的特点出发决定采用裁剪的方法对数据集进行增强操作, 让其能够突出检测目标满足模型训练的要求, 为后面实验训练做好前导工作, 然后对数据进行了类别平衡重采样, 最后简要介绍了模型所需要的评价指标, 以及测试集和训练集的划分。

第4章 基于改进 Faster RCNN 的瓷砖瑕疵检测模型

4.1 引言

上一章通过对数据集的分析,根据数据集的特点,采用裁剪的策略对数据集裁剪来进行数据增强,同时对数据类别不均衡的问题进行了类别平衡的重采样,使得数据的特征更加的明显,从而更加有利于提升模型训练的效果。从工业上对目标检测精度角度要求出发,本文将采用 Faster RCNN 的算法对瓷砖进行瑕疵检测的任务。

本章主要是对 Faster RCNN 网络模型在瓷砖瑕疵缺陷上进行原理上的分析。深入剖析它的三大组成部分,对相关模块进行改进优化。到目前为止 Faster RCNN 依然是一种非常有效的两阶段目标检测算法,同一阶段的算法相比更加的复杂且速度较慢,但是精度更高。

4.2 网络结构

Faster RCNN 整个网络结构大致可以划分四个部分,如图 4.1 所示。主要包括主干网络、RPN 网络、ROI Pooling 模块和分类回归。

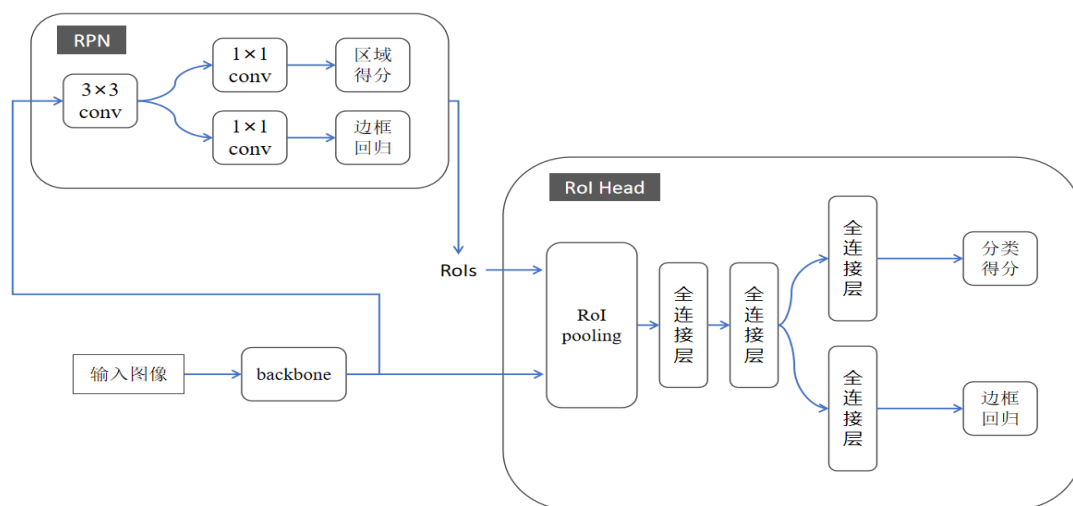


图 4.1 Faster RCNN 网络模块构成图

4.2.1 主干特征提取网络

主干特征提取网络主要负责图像的特征提取，它是目标检测的核心，通常情况下，采用在公共数据集上进行训练好的网络充当主干网络，来进行网络的特征提取。主干网络对目标检测网络的性能起着决定性的作用，Faster RCNN 原始论文中使用 VGG 作为特征提取网络，由于 ResNet 网络的性能优势，目前深度学习的大部分的主干网络的采用的是 ResNet。

ResNet50 是由 Conv Block 和 Identity Block 这两个基础块构建而成，如图 4.2 所示。其中 Conv Block 它通过输输入和输出的维度不一样来改变网络的维度，反之 Identity Block 利用输入维度相同的特点可以加深网络的深度，Faster RCNN 在利用 ResNet 进行特征提取的时候，它只进行了四次的长宽压缩，将第五次进行压缩的内容输出给 ROI Pooling 模块进行使用。

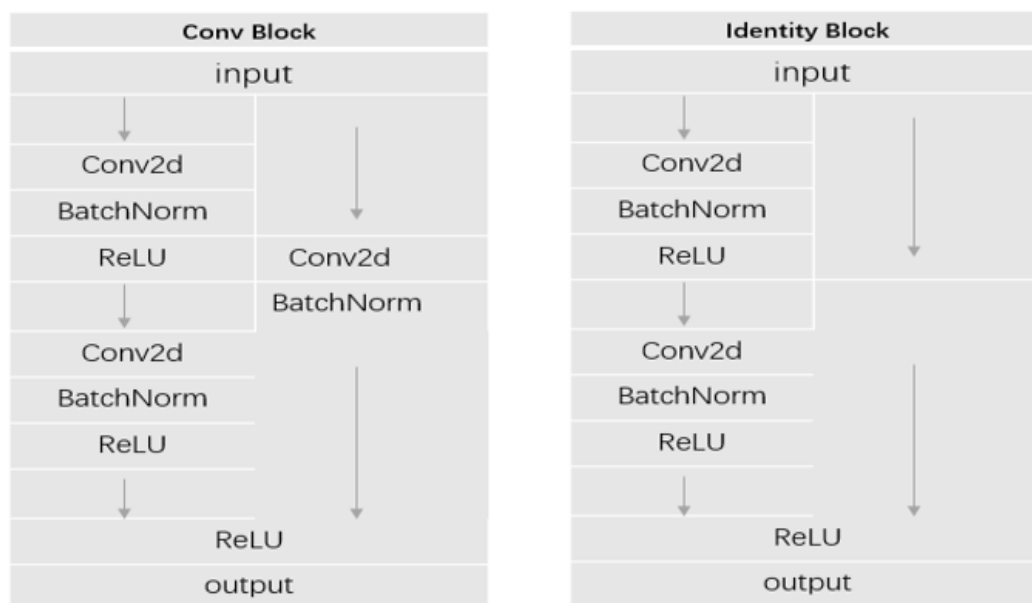


图 4.2 Conv Block & Identity Block

4.2.2 RPN 层

RPN 区域候选网络是 Faster RCNN 的富有创造性的部分，它成功的取代了选择性搜索算法做到在检测精度和检测检测速度的两全。它为 Faster RCNN 加快检测速度做出了核心的贡献。与 ROI 共享卷积，使用全卷积的计算方式减少

计算量，提高了检测速度不少，RPN 结构图如图 4.3 所示。

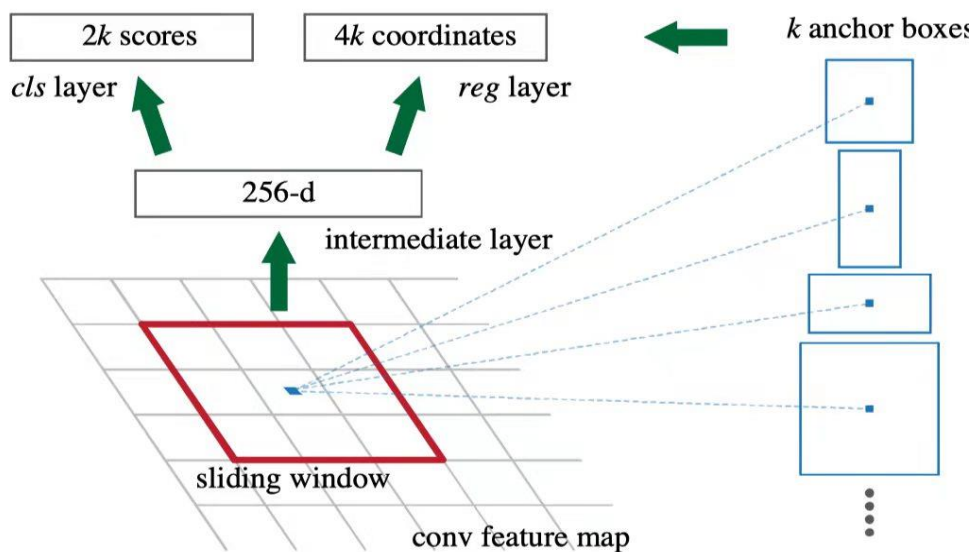


图 4.3 区域建议网络

正如图上所表达，RPN 网络的基本原理是在生成建议框的过程中，在主干网络输出的特征图上使用 3×3 的窗口(Sliding window)对输出的特征图进行遍历，使用三种不同的尺度搭配三种高宽比生成 9 框，然后每个滑窗都被映射为低维的向量，最后将两个低维向量输出到分类(cls)和回归(Reg)中进行分类和回归识别。

4.2.3 ROI 池化层

由于 RPN 网络提出的建议框的大小尺寸是不确定的，而神经网络输入的图片固定，因此，Faster RCNN 中为了保证将所有的候选框转换成固定的尺寸，提出了 ROI 池化层。它的作用是为了解决图像尺寸不统一的问题而出现，允许端到端的训练和测试目标检测系统，实现训练和测试的显著加速并提高准确率，详细的细节步骤如下：

- (1) 以输入的图片作为基准，将感兴趣的区域映射到 feature maps 上。
- (2) 划分映射之后的区域为相同的大小区域。
- (3) 取每个划分而来的区域的最大像素，进行最大池化操作。

经过上述的操作之后就实现了从不一样的大小方框中获得相同尺寸的对应

的 feature map。综上所述，Faster RCNN 的网络模型各个模块的流程如图 4.4 所示。

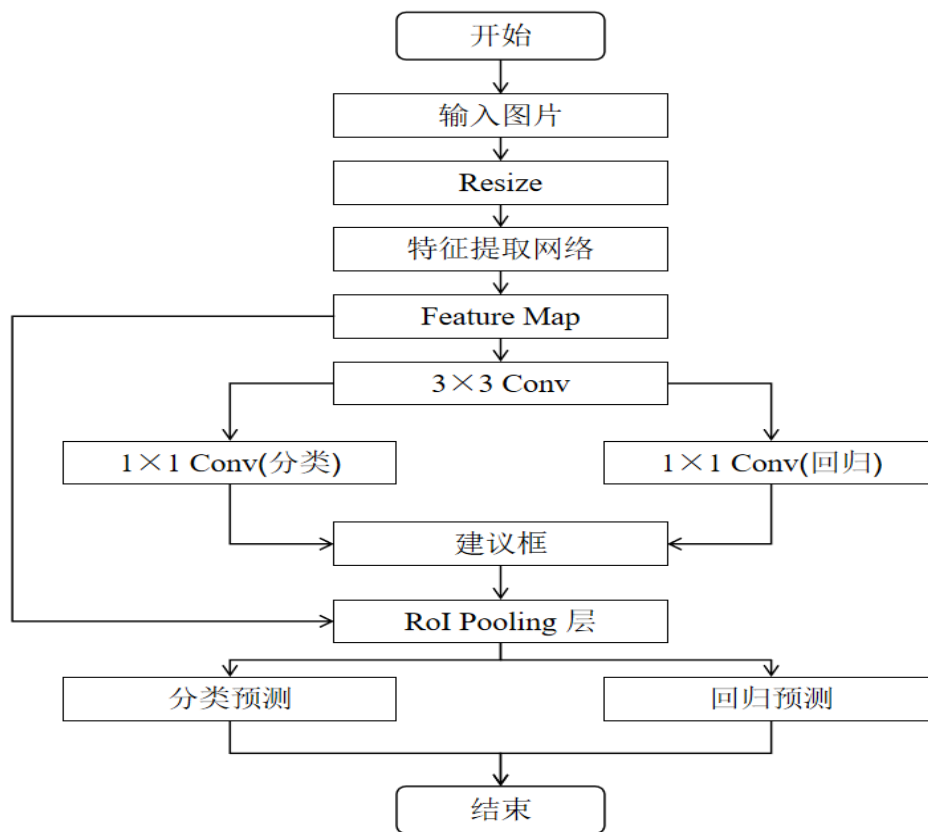


图 4.4 Faster-RCNN 执行流程图

4.3 改进的 Faster RCNN 模型

4.3.1 基于分散注意力网络(ResNeSt)的改进

通常情况下，卷积神经网络主干网络代表网络提取特征的能力，网络的性能越好，检测的准确度就会得到明显的提升，因此，选择良好的特征提取网络尤为重要。2020 年，HangZhang 等人提出了一种性能较好的深度残差网络变体 ResNeSt^[53]。ResNeSt 融合了 ResNeXt^[54]、Inception block^[55]和 SK-Net^[56]的思想，是 ResNet 的最强改进版本，同 ResNet 相比较，在参数的数量并没有显著增加的情况下，提高了模型的准确率，它弥补了 ResNet 的感受野大小固定单一，并且

无法融合不同尺寸的特征，加强了跨通道之间的相互作用。

ResNeSt 在基于 ResNet 的基础之上采用多路径机制，保证在网络的层数足够深的前提下，能够提取到不同尺度的特征信息，结合组卷机和多分支的架构的优势让网络学到更丰富更明显的特征，与此同时让网络更加注重对局部信息的提取，排除其他不必要的噪声，更有利于网络实现精准的图像识别任务。ResNeSt 引入注意力机制建立了通道之间的重要程度，将通道维度的注意力扩展到特征图组进行表达，最终形成模块化。ResNeSt 模块结构如图 4.5 所示。

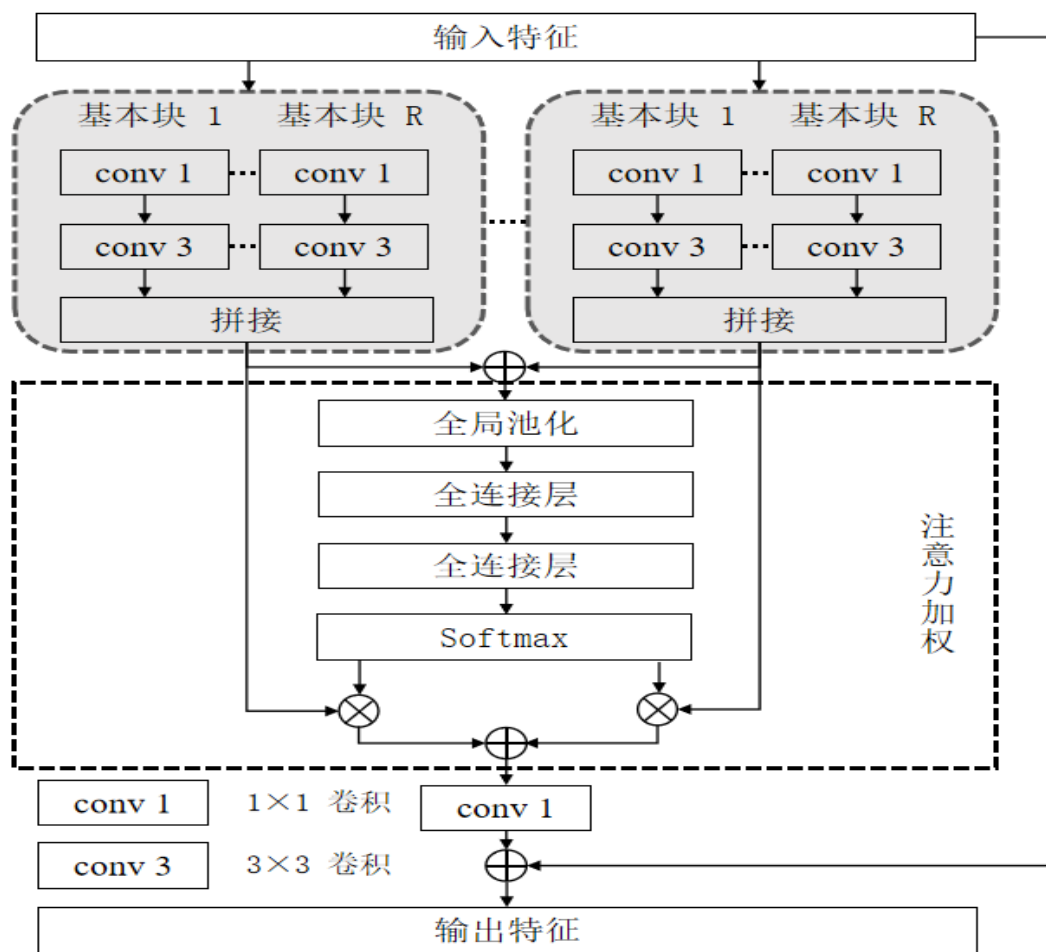


图 4.5 ResNeSt 模块

其中分散注意力模块(split attention)可以跨越特征图组实现信息的交互。分离注意力模块是计算单元，主要由特征图组合分离注意力操作组成，计算公式如 4.1 所示。

$$s_c^k = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W U_c^k(i, j) \quad (4.1)$$

式 4.1 中: $H \times W$ 为空间维度的尺寸, S_c^k 是全局纹理表达的 S^k 的第 c 个通道的每一组权重, U_c^k 为 U^k 的第 c 个通道的特征向量卷积的结果。

基数组通过通道维度软注意力实现加权融合, 然后加权产生每个通道的特征图, 公式如 4.2 所示。

$$V_C^K = \sum_{i=1}^R a_i^k(c) U_{R(k-1)+r} \quad (4.2)$$

公式中 $a_i^k(c)$ 的公式表达如 4.3 所示。

$$a_i^k(c) = \begin{cases} \frac{\exp(g_i^c(s^k))}{\sum_{j=0}^R \exp(g_j^c(s^k))} & \text{if } R > 1 \\ \frac{1}{1 + \exp(-g_i^c(s^k))} & \text{if } R = 1 \end{cases} \quad (4.3)$$

卷积神经网络的经典残差网络模型在计算机视觉中大放异彩, 它可以通过增加网络的深度加强特征提取的能力, 但是, 如果图像的局部存在复杂的噪声, 则不能充分的提取图像的特征, 对特征提取提出了新的挑战, 对网络的检测和识别起到了干扰的作用。残差网络改进版 ResNeSt 的提出, 在一定程度上解决了这个问题, 提高了目标检测的性能。

表 4.1 展示了 ResNeSt 在 MS-COCO 数据集上的实验对比结果。

表 4.1 不同主干网络的性能比较 mAP/%

网络模型	主干特征提取网络	mAP
Faster RCNN	ResNet50	39.25
	ResNet101	41.37
	ResNeSt50	42.33
	ResNeSt101	44.72
Cascade RCNN	ResNet50	42.52
	ResNet101	44.03
	ResNeSt50	45.41
	ResNeSt101	47.50

根据上表的结果, 为了提高目标检测的精度, 本文选择 ResNeSt 作为主干网络对 Faster-RCNN 进行改进。其主要的模块 Split-Attention 通过将上层各组的输

出进行融合之后再行相加,最后通过共享卷积层,按照通道进行分组,对分组卷积做 R-softmax,实现跨通道的注意力机制,解决了 ResNet 作为特征提出网络有限感受野存在的问题。

通过第三章的瓷砖瑕疵数据集进行对比实验的结果如表 4.2 所示。

表 4.2 替换特征提取网络之后的实验结果

网络模型	mAP	ACC
Faster RCNN(ResNet50)	59.71	83.57
Faster RCNN(ResNeSt50)	64.22	79.10

4.3.2 融合 RFPN 模块的改进

考虑到瑕疵的检测图像中小目标过多,为了提升小目标检测的性能可以通过引入特征金字塔进行加强目标特征提取的性能,在没有引入 FPN 之前的网络例如 Faster RCNN 等网络都是利用最后一层的特征层进行网络的预测,而 SDD 则是绕开上采样的过程进行多尺度特征融合来预测。但是它们都有值得改进的地方,没有充分利用到底层的特征信息,所以通过引入特征金字塔网络进行处理。Faster RCNN 是利用深层网络进行目标检测的算法,对小目标的检测会出现性能下降的情况,本质原因是小目标物体本身的像素就比较少,随着降采样的累积,它们的图像更容易丢失。在引入特征金字塔之后可以通过特征金字塔的每一层进行独立预测。

特征金字塔(FPN)是特征尺度的金字塔操作,与图像金字塔不同,它是通过将自顶向下和自底向上的特征图进行融合,来实现特征金字塔的操作,FPN 提供的是一个特征融合机制,在没有引入太多参数的情况下,实现了在增加极少计算代价的情况下,完成目标检测的能力的提升。它有以下三个显著特点。

(1) 自底向上路径

自底向上也就是使用步长为 2 的降采样,换句话说就是进行特征提取,在提取的过程中会有 conv1, conv2, conv3, conv4, conv5 的 5 个 Stage。它们分别是通过对原图进行{2, 4, 8, 16, 32}的下采样,随着网络层次增加而增加,它所包含的语义信息就越丰富。不会考虑 conv1 的相关信息,主要是因为它的语义信息太低,占用内存太大。

(2) 自顶向下路径

把高层的特征图使用最邻近插值法进行上采样，这样能够最大程度的保留语义信息来和自底向上的路径进行融合，融合使用的是横向连接，就是将相同大小的特征层进行融合，通过按照元素进行相加，将上采样的映射与自底向上的映射进行合并生成最终的分辨率结果图。并且每一个层级进行独立的预测。

(3) 横向连接

横向连接主要是指采用 1×1 的卷积进行连接。

经过上述分析可得 FPN 结构如图 4.6 所示。

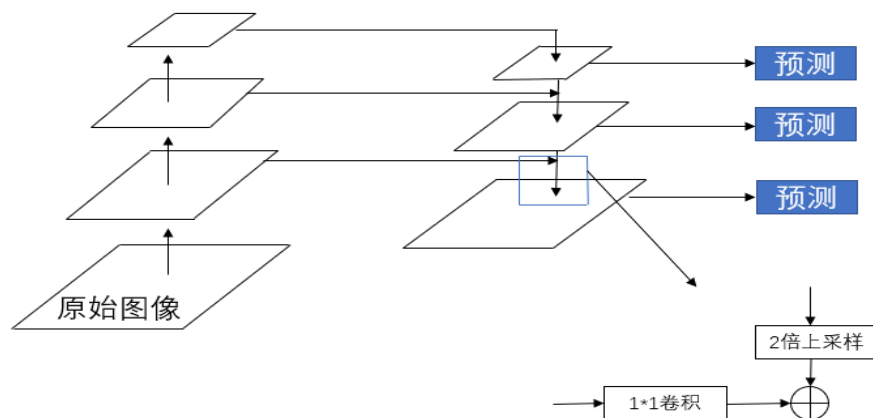


图 4.6 FPN 网络模块

为了获得更好的特征，在 FPN 的基础之上，引入了 DetectoRS 算法当中的 RFP 结构，将 FPN 的输出特征图再次连接特征提取网络进行融合，从而可以实现查看网络两次或多次网络的结构，让网络能够接收原始的特征图和反馈信息作为输入。改进之后的 RFPN 如图 4.7 所示。

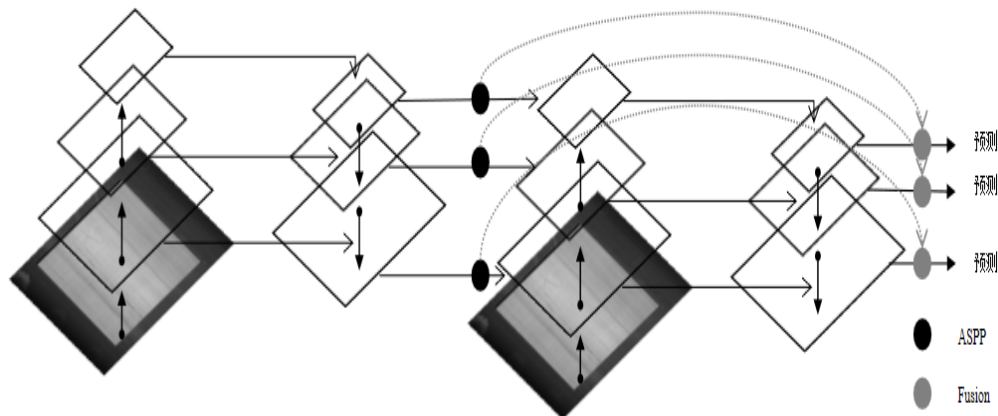


图 4.7 RFPN 网络结构图

通过采用具有反馈结构的递归金字塔网络结构，对小目标的瑕疵特征进行深度的融合，在一定程度上解决了小目标检测过程中 FPN 特征信息融合不充分的问题。另一方面，在分类和回归预测梯度信息的回传过程中，让目标检测的信息能够更加快速的调用主干网络的参数，进而提高目标检测的性能。所以本文采用递归金字塔来替换传统的金字塔来进行不同尺度特征的融合，进而达到增强瓷砖瑕疵检测的目的。Faster RCNN 融合 RFPN 如图 4.8 所示。

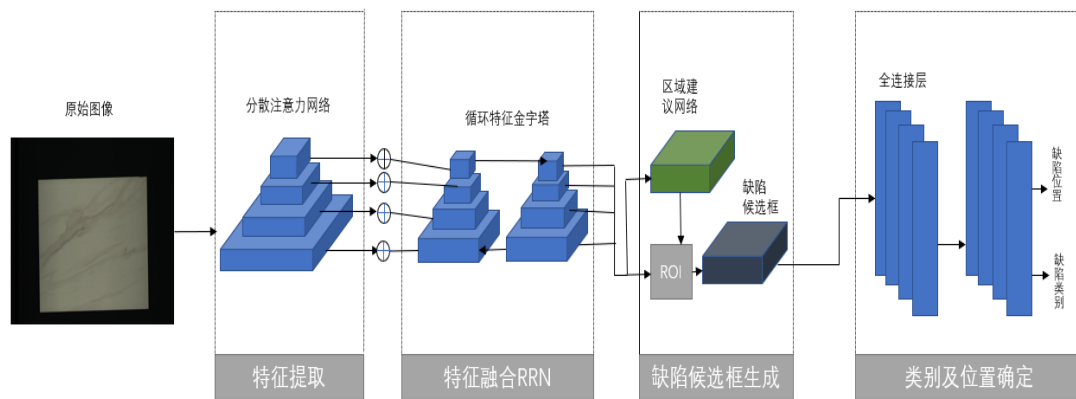


图 4.8 Faster RCNN 融合 RFPN 图

4.3.3 可形变卷积(DCN)的 Faster RCNN 与 RFP 融合

目标检测的对象通常在图像中呈现出不同的形状，甚至不同的角度。瓷砖的瑕疵也是复杂多样的，神经网络通常依靠数据的多样性来对抗各种各样的形状，网络的本身并没有采取相应的措施来解决这种问题，可形变卷积的提出大大改善了这一情况。标准卷积通常采用固定卷积核的大小，不能很好的适应对未知的变化。这种从规格点中进行采样，使得卷积没有能力去适应图像中可能出现的不同非刚体的形变，最终让检测目标被遮挡，然而在可形变卷积中使用可学习的偏移量进行移动对未知的改变进行适应。卷积就可以作用在这些可以移动的点上，进而使得卷积变得具有空间变形的能力。因此，可变卷积的引入改善了检测对象中存在形状困扰的状况。

可形变卷积它通过在原始的卷积层中增加一条新的卷积分支来进行学习卷积核的位置偏移，使得它具有了不规则形状感受野的特点。可形变卷积作用效果可视化如图 4.9 所示。



图 4.9 可形变卷积作用效果图

原始卷积层输出公式如 4.4 所示。

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n) \quad (4.4)$$

其中 p_0 是卷积核心中心点像素的坐标，式子中的 R 是卷积核覆盖的图像区域，集合 $R=\{(-1, -1), (-1, 0), \dots, (0, 1), (1, 1)\}$ 表示的是标准 3×3 卷积核中每个位置向不同方向的偏移， p_n 表示图像中其他像素点的坐标。

形变卷积的输出公式表达如 4.5 所示。

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n + \Delta p_n) \quad (4.5)$$

$w(p_n)$ 为卷积核的位置权重， $x(p_0 + p_n)$ 为输入特征图与卷积核对应的区域， Δp_n 是位置学习偏移，同时也就是形变卷积比标准卷积多出来的部分，因此可形变卷积具有更强的特征表达能力。

ResNeSt50 总共有 5 个阶段，每个阶段都基于前面介绍的两个残差块组成，在网络中引入可以形变卷积就是将残差块中的 3×3 卷积替换成为可以形变的卷积，经过替换之后 ResNeSt50 能够感应到瑕疵的长度且动态的调节感受野的大小，大大的提高网络检测不同形状瑕疵的鲁棒性。引入 DCN 的特征提取网络结果图如图 4.10 所示。

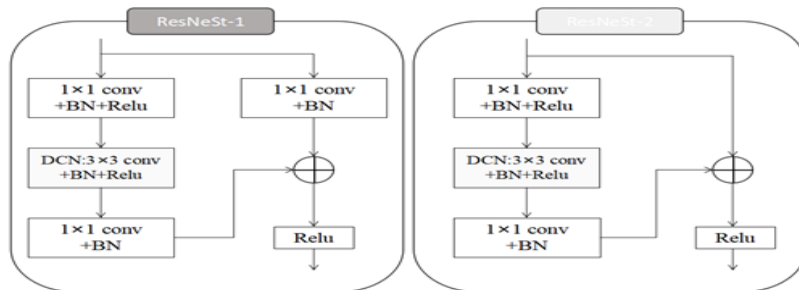


图 4.10 引入 DCN 改进之后 ResNeSt 模块

4.4 实验结果和结果分析

4.4.1 实验所需环境

表 4.3 列举了几种常见的深度学习框架进行对比。这些框架在计算机视觉都取得了不错的成果，都突出了各自的特点。

表 4.3 深度学习框架

框架	API	平台	年份
Torch	C++, Lua	Linux, macOS, iOS, Android	2002
Caffee	Python, C++	Windows, Linux, macOS	2013
TensorFlow	Python, C++	Windows, Linux, macOS, Android	2015
Pytorch	Python, C++	Windows, Linux, macOS	2017

Pytorch 是一个以 Python 优先的深度学习框架，具有支持 GPU，轻松扩展等特点，使得它在众多深度学习框架中脱颖而出。它不仅仅能够实现强大的 GPU 加速，还能动态的支持神经网络，其次它通过线性的设计思路，让他简便并且易于使用，通过快捷信息快速的定位到错误，解决了在不透明引擎上的时间浪费问题。与 Tensorflow 等高度工业化代码相比较，Pytorch 的代码更加简洁直观，它也更方便理解。所以本文实验框架选择 Pytorch 作为深度学习框架。

本文实验环境配置版本号如表 4.4 所示。

表 4.4 实验环境及配置表

实验环境	环境配置
操作系统	Windows10
CPU	Intel(R)Core(TM)i7-9700LCPU@3.60GHz
GPU	TeslaV100
内存容量	32G
深度学习框架	Pytorch
编程语言	python
CUDA 工具版本	10.1
CUDDN 版本号	7.6
计算机图像视觉库	OpenCV

4.4.2 深度学习网络参数

在深度神经网络训练的过程中，超参数的设置在一定程度上决定实验结果的好坏。通过观察检测指标损失率和准确率折线图，分析判断模型在训练过程中处于一个什么样的运行状态，并且以此来进行指导模型的训练，达到更加科学有效的训练模型。超参数的设定在一定程度上决定了实验最后结果的好坏，经过实验反复验证，本实验中的网络模型选定的参数如表 4.5 所示。从表中可以看到，本实验的学习率为 0.0025，学习率决定着目标函数能不能收敛或者是目标函数何时收敛。模型的训练回合数(Epoch)为 120，训练数据的批尺寸是 Batch-size，它代表每次送入网络模型中的数量规模，它的选定主要和随机梯度噪音和收敛速度有关。Dropout 代表随机让网络中的一些节点失效。

表 4.5 网络模型超参数设置

超参数	值
学习率(Learning rate)	0.0026
训练回合数(Epoch)	200
批量大小(Batch-size)	24
Dropout(概率)	0.2

交叉熵代价函数的作用是衡量人工神经网络的预测值与实际值，能够正向有效的促进人工神经网络模型的训练，能够让预测更加真实，所以本文选取交叉熵代价函数(Cross Entropy cost function)。其公式的定义如 4.6 所示。

$$CE(\theta) = -\sum_{i=1}^n y_i * \log(\hat{y}_i) \quad (4.6)$$

其中 n 表示总的样本数量， y_i 为期望输出， \hat{y} 表示预测概率分布输出为 θ_i ，则 \hat{y} 的计算公式如 4.7。

$$\hat{y}_i = \text{soft max}(\theta_i) = \frac{\exp(\theta_i)}{\sum_j \exp(\theta_j)} \quad (4.7)$$

深度神经网络中存在庞大规模的训练参数，所以合适的优化器能够合理的减少网络训练的参数。本文选择 Adam 优化器，它吸取了 AdaGrad 和 RMSProp 两种算法的优点，参数的更新不受梯度的伸缩变换的影响，能够非常有效的自动调整学习率，解释性很好，更新的步长在一定的范围之内，它在深度学习很多场景都是工作性能比较优秀的优化器。适合于深度学习各种常见的场景。表 4.6 简

述了 Adam 优化算法的具体运算过程。

表 4.6 Adam 优化算法过程

算法: Adam 算法

Require: 步长 ε (建议默认为: 0.001)

Require: 矩估计的指数衰减速率, ρ_1 和 ρ_2 在区间[0,1)内部(默认分别为 0.9 和 0.999)

Require: 用于数值稳定的小常数 δ (默认建议为: 10^{-8})

Require: 初始参数 θ

初始化一阶和二阶矩变量 $s = 0, r = 0$

初始化时间步 $t = 0$

While 没有达到停止准则 do

从训练集中采包含 m 个样本 $\{x^{(1)}, \dots, x^{(m)}\}$ 的小批量, 对应目标为 $y^{(i)}$

计算梯度: $g \leftarrow -\frac{1}{m} \nabla_{\theta} \sum_i L(f(x^{(i)}; \theta), y^{(i)})$

$t \leftarrow t + 1$

更新有偏一阶矩估计: $s \leftarrow \rho_1 s + (1 - \rho_1)g$

更新有偏二阶矩估计: $r \leftarrow \rho_2 r + (1 - \rho_2)g \otimes g$

修正一阶矩的偏差: $\hat{s} \leftarrow \frac{s}{1 - \rho_1^t}$

修正二阶矩的偏差: $\hat{r} \leftarrow \frac{r}{1 - \rho_2^t}$

计算更新: $\Delta\theta = -\varepsilon \frac{\hat{s}}{\sqrt{\hat{r}} + \delta}$

应用更新: $\theta \leftarrow \theta + \Delta\theta$

End while

4.4.3 网络训练过程

本文选取改进后的三种网络对比实验, 使用第三章经过数据增强的数据集, 在进行网络训练之前, 训练的时候需要将数据集整个训练过程分为两个阶段。

(1) 前向传播

输入图像就是将每一层的输出结果导入到下一层中进行输入, 然后通过激活函数得到本层的结果, 最后进行识别得到分类结果, 计算结果和预期值的误差。

(2) 反向传播

如果上一步的误差值较大,计算统计误差,将误差进行返回并且使用优化算法更新网络中的权值。重复上述步骤进行反复训练,直到损失函数收敛才停止训练。前向传播和反向传播如图 4.11 所示。

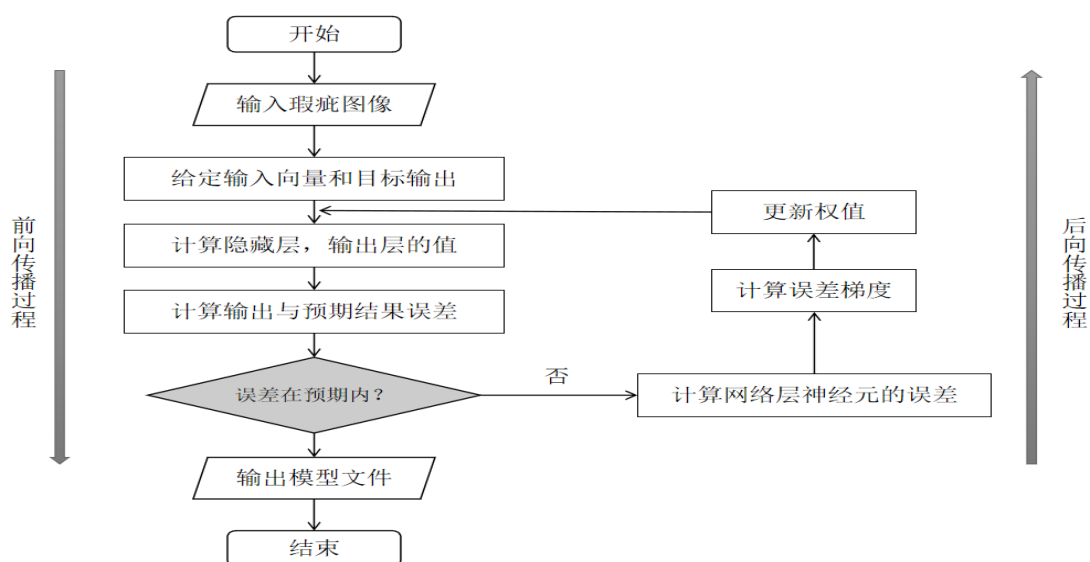


图 4.11 前向传播和反向传播

4.4.4 机器学习的实验结果分析

本文将经过数据增强之后的数据在传统机器学习的方法上进行实验验证。实验结果在准确率指标上的表现如表 4.7 所示。

表 4.7 使用传统机器学习方法进行瓷砖瑕疵分类的准确率 ACC/%

分类器	缺陷类别	原始数据集	增强后的数据集
SVM	边异常	61.23	61.35
	角异常	56.72	57.32
	浅色块	54.68	56.34
	白色块	51.73	53.76
	深色块	55.17	60.78
	光圈	49.32	51.54

表 4.7 中记录了瓷砖瑕疵的六个类别在传统分类器 SVM 上的准确率结果,

分别对比了进行数据增强之后的准确率和没有进行数据增强的准确率。从表中很容易得出,准确率最高准确率是 61.35%,光圈的准确率最低只有 51.54%,出现这种结果是由于像边异常或者是角异常的缺陷,在经过裁剪之后更容易得到加强。而光圈这种瑕疵检测的准确率较低,是由于瓷砖的纹理背景对特征信息进行了干扰。在原始的瓷砖瑕疵数据集上 SVM 的总体准确率是 54.9%,在经过数据增强后的瓷砖瑕疵数据集上 SVM 的总体准确率是 56.8%。

综上所述,进行数据的增强使得数据集整体的准确率得到了提升,尽管如此但是机器学习的准确还是不能满足实际应用的要求,还需要在深度学习模型进行数据增强的验证。

4.4.5 深度学习的实验结果分析

在进行深度学习实验时,同样采用 AP 的值作为实验的评价指标。基于本章节前面的改进,本实验在 Faster RCNN 验证数据增强的检测效果。

表 4.8 Faster RCNN 数据增强前后的 AP/% 值

缺陷类别	原始数据集	增强后的数据集
边异常	62.23	63.67
角异常	64.33	64.97
浅色块	60.15	63.34
白色块	59.34	62.23
深色块	62.13	63.12
光圈	61.85	61.93

从上述表 4.8 中可以发现,在 Faster RCNN 上原始数据的 mAP 是 61.67%,在进行数据增强之后的 mAP 是 63.21%,增强后数据集的 mAP 比原始的数据集高 1.54%,说明在经过数据裁剪的数据增强之后能够有效的提高检测率。其中浅色块的 AP 值增长的幅度最大为 3.19%,光圈的增长值最小 0.08%,其余的都有微幅的上涨,从整体的范围看,经过数据增强之后检测的效果均得到了提升。在原始数据集上 AP 最高的是角异常的缺陷,在增强之后数据集上 AP 最高的也是角异常,说明角度异常的缺陷特征在数据集中比较明显,并且数据增强的效果没有影响到数据类别的稳定性。

以上结果再次证明数据增强是比较有效果的。综上所述经过数据的增强,实

验的指标得到了实实在在的提升。

表 4.9 三种改进网络各缺陷目标 AP/% 值

缺陷类别	网络 1	网络 2	网络 3
边异常	63.89	64.37	65.56
角异常	65.02	65.73	67.78
浅色块	63.56	65.21	66.36
白色块	64.21	66.34	66.39
深色块	63.78	66.45	67.75
光圈	61.12	63.18	64.79

表 4.9 在 Faster RCNN 的基础之上替换特征提取网络为 ResNeSt 作为网络 1, 替换了特征提取网络为 ResNeSt 的 Faster RCNN 融入特征金子塔作为网络 2, 基于可形变卷积 DCN 的 Faster RCNN 融入特征金子塔作为网络 3。从表中可以看出, 网络 3 的 AP 值高于网络 2 的 AP 值, 同时网络 3 的 AP 值也高于网络 1 的 AP 值。其中比较最突出的是角异常, 网络 3 比网络 2 增长了 2.02%, 网络 3 比网络 1 高出 2.76%。说明基于可行变卷积(DCN)的 Faster RCNN 与 FPN 进行特征融合之后网络的检测精度更高。

表 4.10 三种网络的综合指标/%

指标	网络 1	网络 2	网络 3
MAP	63.60	65.21	66.44
ACC	84.49	85.63	86.34

表 4.10 统计了三种网络模型的平均精度和准确率两个网络模型的综合参考指标, 在网络在经过一步一步的优化之后, 网络 3 的 mAP 比网络 2 的 mAP 值高 1.23%, 说明在引入 DCN 之后网络的检测精度有所提高, 网络 2 的 mAP 比网络 1 的 mAP 高 1.61%, 说明引入 RFPN 之后检测精度也得到了提高。除此之外网络的准确率也在一步一步的优化过程中的到了一定深度的提升。

综上所述, 实验结果表明基于可形变卷积(DCN)的 Faster RCNN 与 RFPN 融合的瓷砖瑕疵检测网络精度更高, 更优秀。

下图 4.12 为基于可形变卷积的 Faster RCNN 与 RFPN 融合的表面缺陷检测效果, 其中边异常和角度异常的置信度能够达到 0.91 之高, 但是像浅色块和深色块的置信度只有 0.6 左右, 说明网络有待进一步的提升。

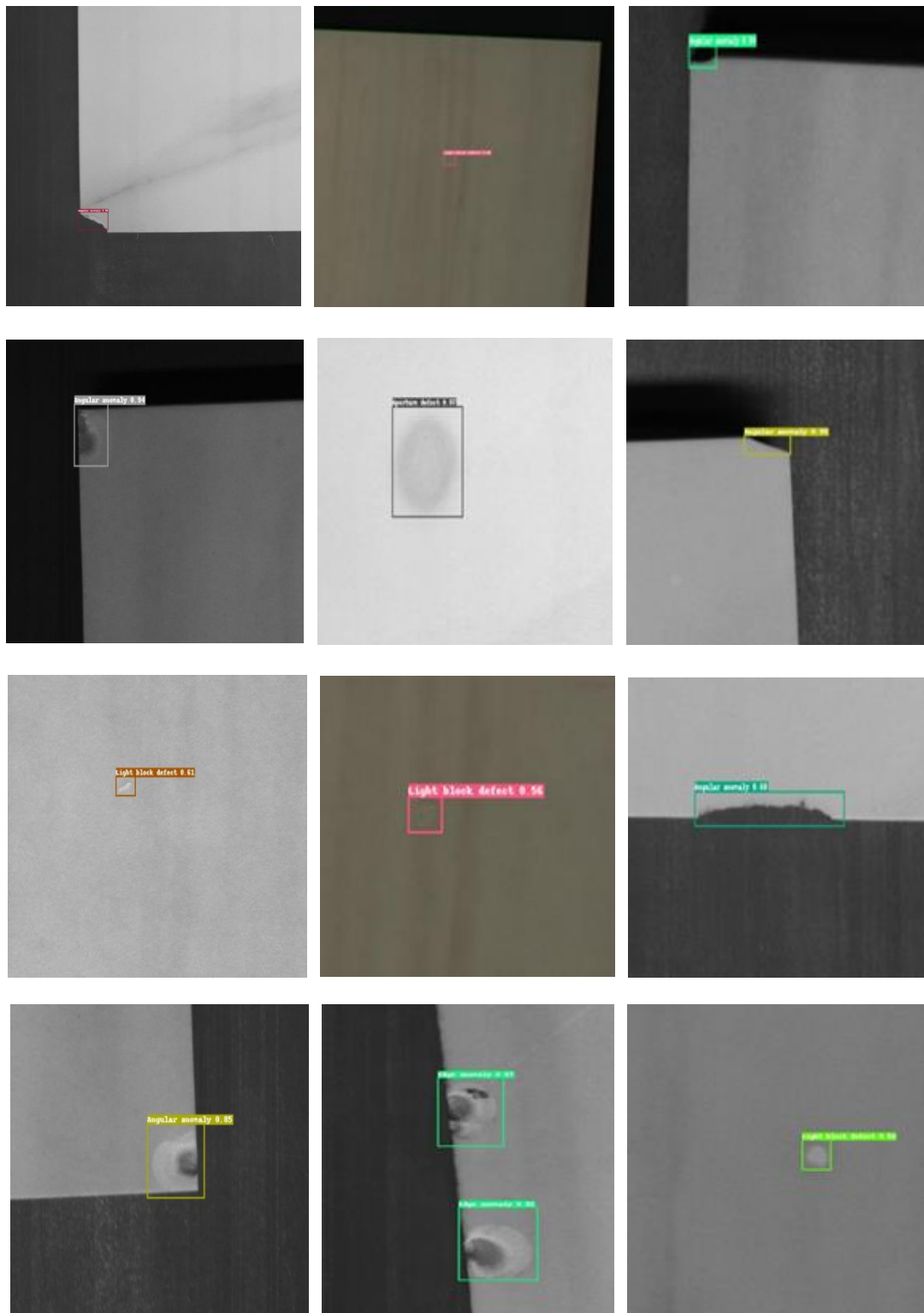


图 4.12 Faster RCNN+DCN+RFPN 融合的网络效果图

4.5 本章小结

综合上所述,本章节首先详细的 Faster RCNN 网络的详细情况,对网络的主干网络、RPN 层、ROI Pooling 层和分类层进行了深入的剖析。然后从瓷砖瑕疵数据集的实际特点出发,首先使用 ResNeSt 替换掉初始的卷积层,进一步加强网络模型进行特征信息提取的能力,然后通过引入 RFPN 特征金字塔的网络结构,加强特征信息的融合,实现可以查看特征信息两次。最后考虑到瑕疵的形状不规则且千变万化,通过引入形状卷积(DCN)来解决这种形状改变所带来的问题,实验验证部分,在传统方法和深度学习进行对比的方式验证了数据增强在精确度上的提升,除此之外还验证了 Faster RCNN 网络替换特征提取网络, Faster RCNN 替换特征提取网络后融合 RFPN, 替换特征提取网络 Faster RCNN 融合 RFPN 后添加 DCN 三种网络在综合指标上的实验效果。实验结果也表明 Faster RCNN 融合 RFPN 添加 DCN 的网络模型实验效果最好。

第5章 瓷砖瑕疵检测系统设计与实现

本章针对瓷砖企业依靠人工进行瓷砖瑕疵存在质检效率低下，质量参差不齐并且成本居高不下等问题，对瓷砖瑕疵检测业务进行梳理，并且以瓷砖瑕疵检测为重点，设计并实现符合瓷砖瑕疵检测工作所需要的瓷砖瑕疵检测系统。

5.1 业务逻辑梳理

瓷砖瑕疵识别分类业务参与者主要有生产一线的技术人员和企业管理人员，其中技术人员是用户的角色，生产管理人员是管理员角色。用户主要是通过获取到的高质量图像输入到检测系统中，对图像进行识别和分类，系统会记录有瑕疵的瓷砖编号和类型，并做好数据统计，之后可以对分类后的瑕疵进行定制化的训练模型。管理员可以通过系统对工程进度进行评估，导出报表，及时的反馈给现场制作的工人，还可以进行规划下一轮的生产任务。系统的业务逻辑如图 5.1 所示。

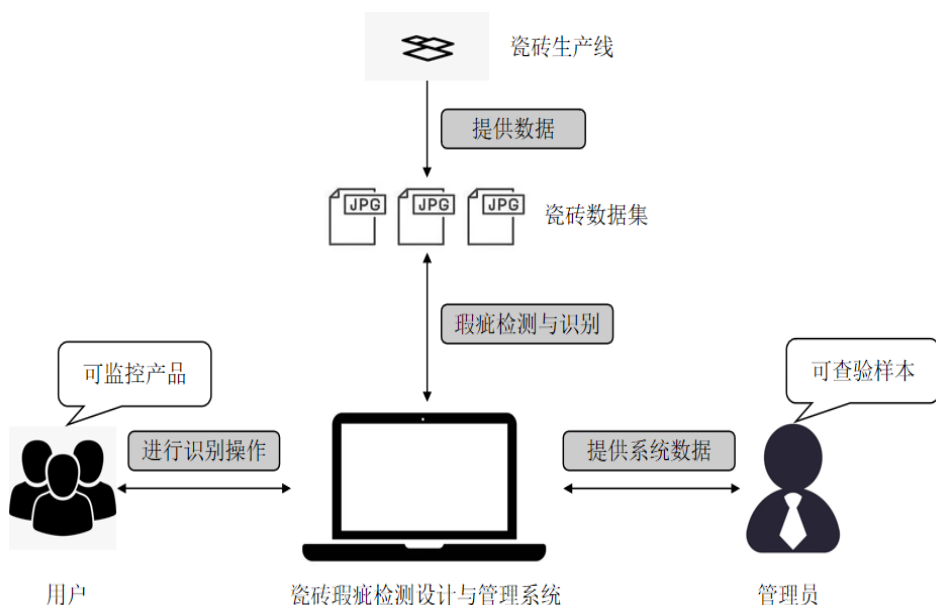


图 5.1 系统的业务逻辑流程图

5.2 瓷砖瑕疵检测系统设计

5.2.1 系统需求分析

(1) 总体需求分析

瓷砖是一种非常常见的建筑装饰材料，它主要用于建筑物室内外地面和墙面的表面等，主要用途为建筑装修装饰，所以瓷砖行业与房地产行业，建筑装修装饰行业紧密相关，城镇化率的不断提升更是瓷砖行业发展的核心动能。瓷砖行业的变化发展是城镇化，工业化的重要体现，瓷砖的瑕疵检测更是在瓷砖的生产过程中重要的工作内容。

现行的瓷砖瑕疵检测作业，依然是通过在一个生产线上安排多名的工人进行瓷砖的瑕疵检测，在这类工作中通过延长工作时间加快瑕疵检测的总进度，实际的工作模式存在工程量大、工作周期长和质检技术人员容易由于视觉疲劳等问题，以上原因都有可能企业生产效率过低。

现实的业务需求急切需要寻找一种更加高效的技术手段和相应的网络平台来进行瓷砖的瑕疵检测。本系统将从完善企业智能化改造，提高工作人员的工作效率的角度出发，以瓷砖的瑕疵检测为核心内容，根据瓷砖生产的流程和计算机视觉技术相结合设计并且实现，对检测人员、检测记录和良品率等数据，实现数据管理和维护。本系统的构建将为现场质检员工提供良好的科学依据和快速的检测结果。最后值得一提的是，本系统是建立在可以获得高质量图像的基础之上，所以如何获得高质量的图像不属于本文的研究内容。

(2) 功能性需求

需求分析是软件计划过程中前期的重要工作，同时也是软件开发过程中的基础工作之一。这个阶段所做的主要工作是对用户提出来的“要求”进行归纳总结，转换成为软件开发的专业文档，确定需要完成那些工作。对系统不同的用户角色分析出具体的需求，进行分模块的功能设计，本系统主要设计的有管理员和普通用户这两个角色。

本系统是瓷砖瑕疵检测系统，检测模型是经典检测网络 Faster RCNN,系统的核心功能是瓷砖瑕疵的检测和识别。用户通过上传图像到图像检测识别模块，系统从服务器响应到前端提交的任务之后，调用 Faster RCNN 模型对其进行检测识别，将识别分类之后的图像进行编号保存在服务器，然后将识别检测的结果

响应到前端页面进行展示。从用户角度而言,用户使用本系统的时候,系统首先需要提供注册的功能,让用户在每次使用系统的时候必须先进行登录验证,确保一人一号。其次是图像的上传和识别功能让用户能够实时的观察到瑕疵检测的结果。最后提供历史记录查询的功能,每个普通用户可以在系统上搜索到自己检测的历史内容。从管理员角度而言,他们拥有系统管理的分配权限,是最高权限等级,可以增删查改普通的用户信息,查看普通员工的相关信息,设置密码,赋予普通角色相关权限等功能。除此之外,比较重要的功能还能重新添加优化之后的模型。

(3) 非功能性需求

非功能性需求是指不具体的描述软件应该如何做的,更多的是从维护软件稳态的角度进行考虑。非功能性需求主要是指能够满足系统的正常使用和操作优化之后,还需要在安全性、稳定性和可扩展性等方面的需求。系统的安全性是指对系统进行保护,做到每个用户之间相互隔离,同时还要防止外部黑客的攻击。稳定性是指为了应对未来的风险和未来的风险做对抗,从而来提高系统运行的占比时间。可扩展性是指系统适应变化的能力,随着用户量和业务量的提升也能保证系统的合理运行。

5.2.2 可行性分析

本小节对用户提出的理想化的需求进行分析,主要从经济和技术两层面来论证瓷砖瑕疵检测系统的可行性。

(1) 技术层面

当前比较流行的软件架构主要有 C/S 架构和 B/S 架构,其中 C/S 架构由于适用面窄,通常用于局域网中,而且又因为需要安装才能使用,导致用户群体比较固定,维护程序比较复杂,版本迭代比较繁琐,成本较高。所以客户端的程序市场占有率不高。

与此同时,B/S 架构可以随时通过输入网址进行查询,没有安装客户端烦恼,业务扩展比 C/S 简单。迭代方便,维护方便,开发容易,共享性强,成本低。数据可以通过云服务等应用进行存储,这样可以大大减轻本地存储的负担,同时也不用担心数据的丢失。综上所述,本系统采用优点众多的 B/S 的架构进行开发。

Java 语言简洁快速的特点使得它成为深度学习中比较热门的语言,并且有

一套比较成熟的开发框架。所以本文使用成熟稳定的 Java 开发语言，选择 Spring Boot 作为本系统的开发框架。前端开发使用具有快速响应特点的 Vue 前端框架和 Themeleaf 前端模板进行实现，不仅能保证页面响应的及时，还能构建简单优雅的界面。

因此，使用以上热门技术进行瓷砖瑕疵系统实现在技术上是可行的。

(2) 经济层面

从设备成本的角度出发，需要一台笔记本电脑用于软件开发，和一台服务器用于数据的存储和功能模块的测试。从员工技术培训的角度，员工只要会使用电脑和手机等通讯工具，就能快速的上手系统。所以，本系统的开发成本很低，投入使用以后，企业可以快速的部署起来并且使用它辅助工人进行瑕疵识别，大大的提高工人的工作效率，降低企业因为检错而付出的返工成本。

因此，根据以上的论证本系统在经济层面是可行的。

(3) 结论

综上所述，使用以上技术进行系统开发比较容易，经济成本较低，使用的技术生态成熟，系统一经部署，就能够帮助建筑瓷砖企业高效的进行瓷砖瑕疵检测和分类，所以本系统的开发是可行的。

5.2.3 系统开发所需环境

系统在 Window10 环境下进行开发，使用 Java 作为编程语言，后端开发工具是 IntelliJ IDEA，前端开发工具是 Visual Studio2020。数据库采用性能优越的关系型数据库 MySQL，它不仅仅开源而且安全性高，容错性高，出问题容易找到解决方案。图像处理库使用的是 OpenCV 3.4 版本。具体的开发所需环境如图 5.1 所示。

表 5.1 系统开发环境

工具类型	版本与名称
操作系统	Windows 10
编程语言	Java
开发工具	Visual Studio 2020
数据库	MySQL 5.7
图像处理库	OpenCV 3.4

5.2.4 瓷砖瑕疵系统架构设计

本系统采用 MVC 的三层架构进行设计，它凭借可定制、代码清晰和潜在框架结构的特点，成为市场上最为流行的软件设计架构，系统可以划分为应用层，服务层，持久层。如图 5.2 所示，应用层主要负责与用户进行交互，服务层要实现业务员逻辑，持久层作用主要是数据的备份和数据的持久化。



图 5.2 软件三层架构

5.2.5 功能模块设计

根据前面的业务梳理和对瓷砖瑕疵检测的需求进行分析，三个核心的模块分别是：登录模块，检测模块和管理模块。

(1) 登录模块

系统的登录用户分为管理员和用户。表现在应用层就是通过一人一账号的方式在前台页面输入账号密码，确保不同角色的权限相互隔离，首次登录的用户通过短信验证的方式进行账号的注册，注册成功则保存用户信息到数据库，否则跳转到注册页面修改相关信息。服务层进行密码的校验，验证通过之后跳转到操作界面，否则在登录页面进行错误信息的提示。管理员进行登录的时候可以对质检人员管理，进行工作排班，生成相关报表和绩效评估。登录样式采用 Bootstrap 进行设计。

(2) 检测模块

质量检测人员登录账号之后，首先上传获取到的瓷砖原图到系统，然后调用封装了第四章中改进后的 Faster RCNN 模块，进行瓷砖瑕疵检测，经过系统的处理之后，将检测结果一起返回给前端进行对比，方便质量检测员进行瓷砖的瑕疵检测，系统会将瑕疵图进行标记。详细的检测流程如图 5.3 所示。

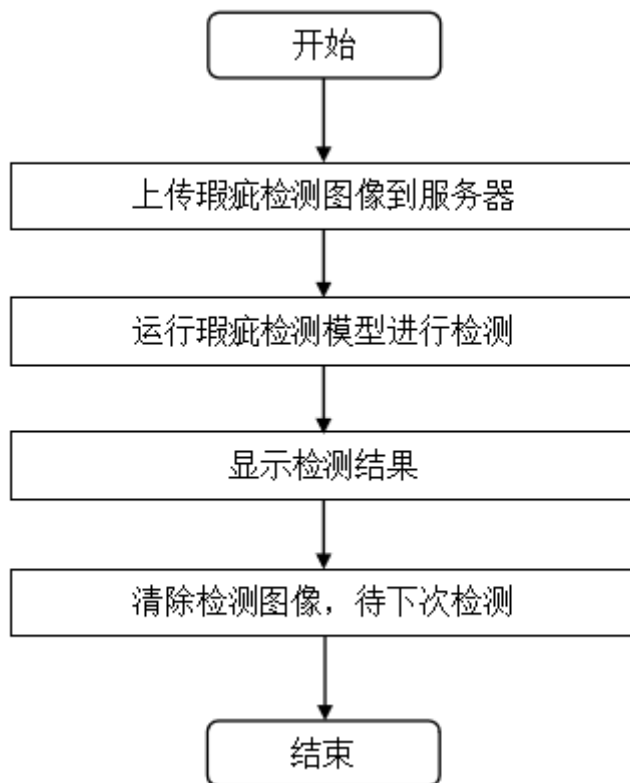


图 5.3 检测流程图

(3) 管理模块

系统的管理模块主要分为人员的管理和图像生成的数据报表管理，对人员的管理主要分为对人员的基本资料的增加、删除、修改和查询，对检测后的数据可以通过系统生成柱状图和折线图，方便以后进行数据分析。

5.2.6 数据库关系设计

数据库作为业务支撑，主要围绕系统的会员管理，模型管理，检测模型三大模块进行设计，其主要逻辑有，用户和检测业务构成一对多的关系，一个用户进

图像的识别, 查看历史记录, 和替换检测模块, 而权限管理部分则都是一对一的关系。

数据库的物理设计是采用性能良好且开源的 MySQL 作为数据的存储。为了进一步改善用户的使用体验, 缓存选用数据类型丰富的 Redis 作为本系统的缓存, 解决系统的稳定性。Redis 还可以充当分布式锁的角, 解决行分布式部署中的分布不一致问题。具体的数据库表之间关系如图 5.4 所示。

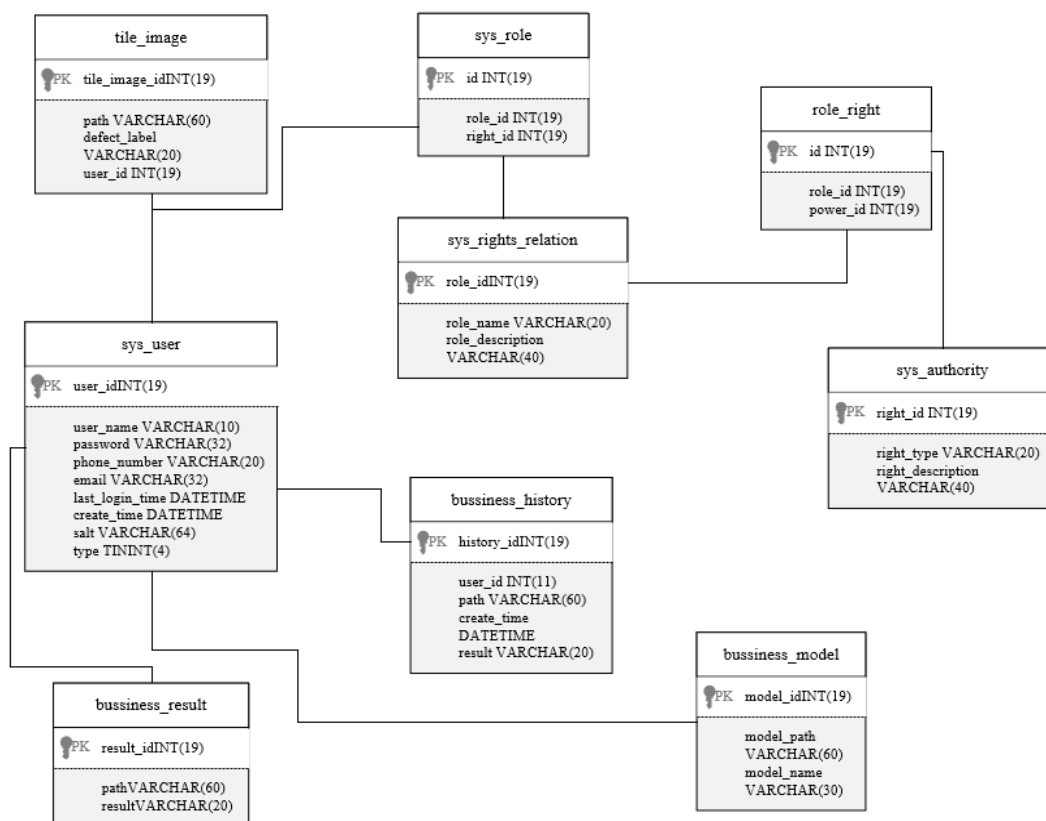


图 5.4 数据库关系图

经过上述数据库关系的分析, 一共有 9 张表, 其中 5 张作为系统基础支撑表, 分别为用户表 (SYS_USER), 角色表 (SYS_ROLE), 权限表 (SYS_AUTHORITY), 用户角色关系表 (SYS_ROLE_RELATION), 角色权限关系表 (SYS_RIGHTS_RELATION) 这五张表是用来保存系统权限录入的相关信息。其余的 4 张表记录的是瑕疵检测业务所需要的表分别是瓷砖图像信息表 (BUSSINESS_IMAGE), 记录表 (BUSSINESS_HISTORY), 模型更新表 (BUSINESS_MODEL) 和结果表 (BUSSINESS_RESULT)。其中图像信息表记录的

是上传图像的原始信息，记录表主要是检测记录的相关信息，模型表记录的是模型的迭代版本和更新内容，结果表记录瑕疵检测的结果。主要业务表的详细信息如表 5.2-5.5 所示。

表 5.2 瓷砖图像表

字段名称	类型	主键	字段说明
tile_image_id	INT(19)	是	主键，唯一
image_path	VARCHAR(40)	否	路径存储
defect_category	VARCHAR(30)	否	瑕疵类别
sys_user_id	INT(19)	否	用户 id 号

表 5.3 瓷砖瑕疵检测识别结果表

字段名称	类型	主键	字段说明
result_id	INT(19)	是	主键，唯一
image_path	VARCHAR(60)	否	路径存储
image_result	VARCHAR(20)	否	识别结果

表 5.4 模型表

字段名称	类型	主键	字段说明
detect_model_id	INT(19)	是	主键，唯一
detect_model_path	VARCHAR(60)	否	检测模型路径
detect_model_name	VARCHAR(20)	否	检测模型名称

表 5.5 检测记录表

字段名称	类型	主键	字段说明
record_id	INT(19)	是	主键，唯一
sys_user_id	INT(12)	否	用户 id 号
image_path	VARCHAR(50)	否	路径存储
record_create_time	DATETIME	否	创建时间记录
image_result	VARCHAR(30)	否	识别结果

5.3 瓷砖检测原型系统实现

5.3.1 登录注册模块功能实现

通过在浏览器中输入网址进入系统登录页面，操作员可以根据不同的身份选择登录，登录模块获取到用户名和密码，后台调用数据库中加盐密码在服务层中进行密码的比对，比对成果进入检测系统，反之比对失败的执行跳转，返回到登陆界面，页面效果图如图 5.5 所示。



图 5.5 登陆界面展示

系统在首次使用的时候，如果没有注册账号会跳转到注册页面进行注册，注册的时候需要收集到姓名、性别、身份证号码、联系方式和证件照等相关信息。注册登陆细节效果如图 5.6 所示。

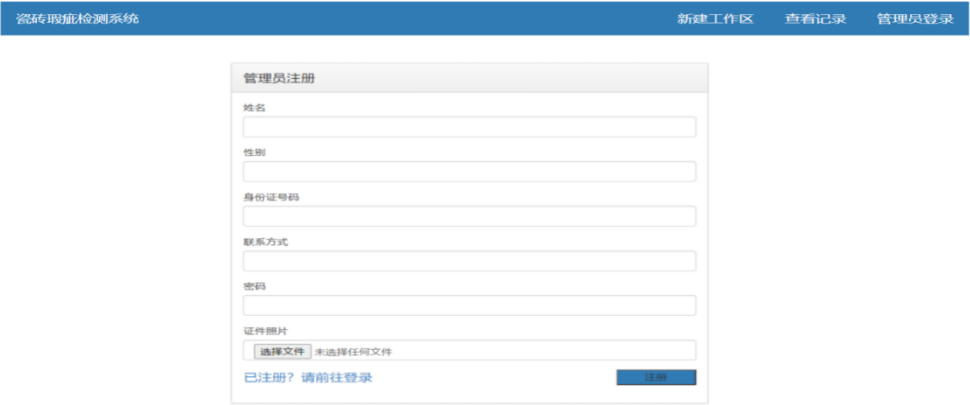


图 5.6 注册页面展示

5.3.2 图像检测模块

操作员进入系统之后，在页面的右上角会显示操作员，和查看记录的选项，界面的左边就是需要选择需要检测的文件夹，也就是图像的集合，然后调用系统的检测模型进行检测，同时也提供了重新识别的按钮方便再次进行结果的验证。如图 5.7 所示。



图 5.7 缺角检测结果

选择后需要识别的文件夹和图片之后，点击识别，系统使用训练好的优化后的模型进行检测，并且能够通过文字的方式将结果返回给前台页面。如图 5.8-5.10 分别展示了不同种类的瑕疵导入系统后进行识别反馈回来的结果。

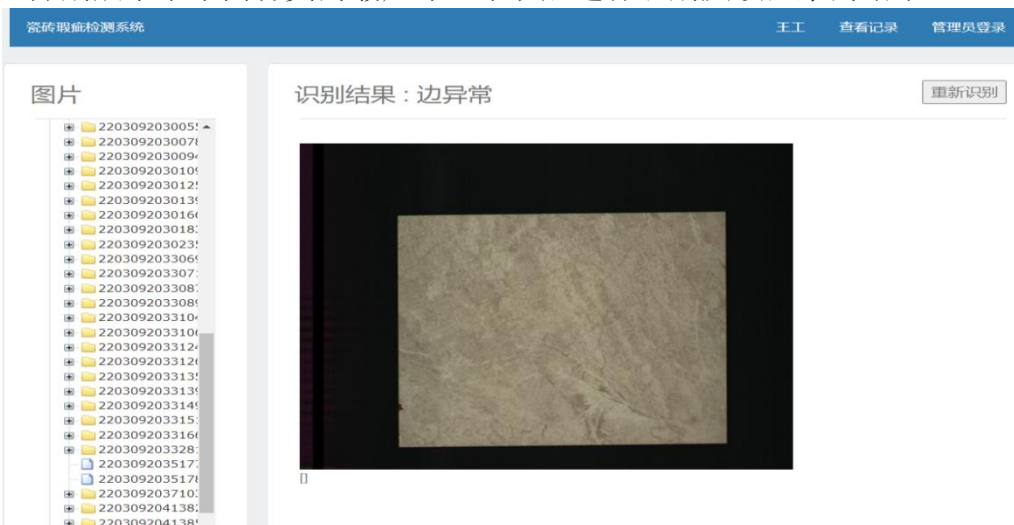


图 5.8 边异常的识别结果



图 5.9 深色点瑕疵识别结果



图 5.10 划伤检测结果图

5.3.3 管理模块功能实现

(1) 用户管理

管理模块主要实现功能是对员工的增加、删除和修改等功能，通过表格显示每个员工的详细信息，同时还提供以表格的形式，进行导出用户的相关信息，具体的操作界面如图 5.11 所示。

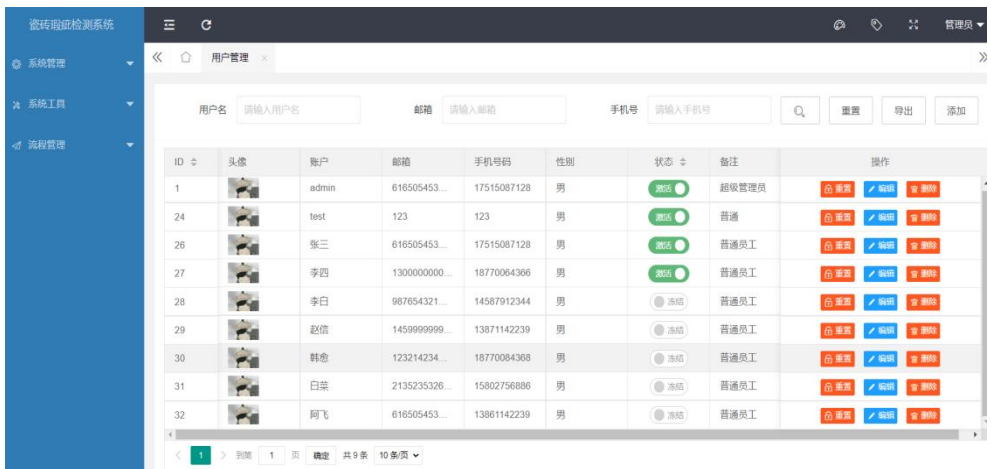


图 5.11 用户管理图

(2) 管理模型

管理员工登录系统后，可以通过添加的方式进行模型的更新，每个模型还可以关联一个键用来标记该模型的生产状态，通过模型的搜索框进行搜索模型，来查看模型的详细信息，详细信息通过表格的形式进行展示，主要包括模型的运行次数，名称等相关信息。模型管理页面实现如图 5.12 所示。



图 5.12 模型管理页面

(3) 角色管理

角色管理模块主要是以经典的 RPAC 权限设计思想为指导，通过上文中设计的五张数据库表进行实现。本系统中主要分为管理员和普通员工两种角色，其界面如图 5.13 所示。



图 5.13 角色管理页面

5.4 本章小结

本章节使用 Java 语言进行开发，后端采用 Springboot 和 Opencv。针对改进的 Faster RCNN 的瓷砖瑕疵检测系统进行了实现，主要实现了登录注册、图像检测和后台管理三大模块，本系统能够对导入系统的瓷砖瑕疵数据集进行识别和记录，管理员还能够对操作人员进行记录和管理以及进行角色权限的分配。

第6章 总结和展望

6.1 工作总结

本文针对建筑瓷砖企业目前存在的通过人工进行瓷砖瑕疵检测的低效率，质检质量参差不齐且成本居高不下等问题，设计了一个基于改进 Faster RCNN 的瓷砖瑕疵检测系统。它可以用来帮助企业提高瑕疵检测的质量和效率，本文的具体工作如下：

(1) 研究背景和意义。介绍了瓷砖瑕疵检测的研究背景和意义，从国内和国外两个角度介绍了瓷砖瑕疵检测的研究现状，介绍了选题的背景和意义，列举了一些深度学习与工业相结合的成果。

(2) 目标检测相关理论知识的介绍。主要从深度学习和传统机器学习两个角度进行了相关理论知识的介绍，深度学习的内容主要有卷积神经网络的主要组成部分和经典卷积神经网络模型的介绍，传统机器学习则主要有整个学习的过程和典型的特征提取的方法和分类器。

(3) 瓷砖瑕疵数据的增强。通过对瑕疵数据进行数据分析，发现数据集合存在检测目标太小和样本数据分布不均衡这两个明显的特点，针对检测目标太小的问题，利用卫星图像处理的方式进行了在线的数据裁剪。针对类别不平衡的问题，对样本数据进行类别平衡重采样的方式进行解决。

(4) 瑕疵检测模型实验的改进。使用了 SVM 传统机器学习方法与 Faster RCNN 卷积神经网络，进行瓷砖瑕疵检测识别的对比实验，然后通过对比实验结果，发现残差神经网络的效果最好。然后再替换特征提取模块为 ResNeSt 网络，同时针对瑕疵的目标大小不一致的情况，通过 FPN 进行不同尺度的特征进行融合，加入了可形变卷积(DCN)解决形状不规则的问题，经过实验结果的对比，改进后的模型在准确率上得到了提升。

(5) 瓷砖瑕疵检测识别系统的设计和开发。本章以 Faster RCNN 为作为系统的识别模块，开发了 B/S 架构的瓷砖瑕疵检测系统，最终参考软件开发的流程对系统进行了业务流程梳理、需求分析和可行性分析，并且对系统的模块进行了详细的设计与实现，并进行检测成果的展示。

6.2 工作展望

本文的研究在形状各异的小目标检测方面取得了一定的成效，但是依然存在很多需要改进的地方。主要有以下方面：

(1) 数据增强部分。可以通过 GAN(比如 CycleGAN)来进行数据增强，通过 CycleGAN 的方式生成高质量的数据集，进行模型的训练，今后的学习中，若有合适的实验平台，将会进一步进行实验验证。

(2) 模型改进部分。本文使用的是经典的 Faster RCNN 目标检测模型，通过网络模型的改进之后，在准确率和 mAP 等指标上得到了提升。通过网络模型架构剖析，未来可以引入 IOU 级联的方式来进一步提升模型的准确率。在速度上也尝试计算机比较新的网络，例如 scaled-yolov4 进一步提升检测的速度。

(3) 系统友好性优化部分。本文的瑕疵系统前端界面比较朴实，可以进一步进行用户友好化的美化设计，后台也可以添加数据分析的模块挖掘，有利于生产的信息。

致 谢

行笔至此，三年的求学生涯已接近尾声，这期间的学习和收获所给予我的成长与进步，让我对这一切充满了感激与不舍。在这里，我想对有缘的大家真心的说一句谢谢！

首先感谢陈萌导师！您是一位认真且负责的老师，三年间和老师之间有关学习、生活的交流与沟通，都是我学生生涯中非常珍贵的财富。特别是论文写作期间，老师对我提出的很多有益建议，感谢您的专业指导和及时帮助，让我顺利结业最后一堂课。同时，谢谢您生活中的对我们的照顾，祝您事事顺心，事业有成，万事如意！

其次感谢一直在背后支持的家人们！感谢父亲和母亲积极进取、脚踏实地的生活态度，这也是我一直以来为人处事的原则，感谢你们对我的支持与培养，祝你们身体健康，永远幸福！兄弟如手足，感谢我的哥哥一直以来对我的支持与鼓励，愿你步步高升，天天开心！

最后感谢给与我无私帮助的朋友！感谢我的室友，有趣的寝室生活让我在繁忙的学习中很好的缓解了压力，谢谢你们在学习和生活中对我的帮助，毕业将至，虽以后身处各地，但友谊地久天长！感谢实验室的小伙伴们，我想日后定会想念大家一起奋斗的日子，很开心认识有意思的你们，祝大家前程似锦，越来越好！

三年的时光转瞬即逝，未来将一直在路上，无论未来发生什么，我都将披荆斩棘，勇往直前，为美好的明天努力奋斗！当然，谢谢一直以来努力且不放弃的自己，而努力对我的意义，是当下一个机会到来的时候，我想我可以。

洪诚

2022 年 4 月

参考文献

- [1] 段淳林, 高辉, 杜鹃鸿. 中国建筑陶瓷品牌消费者研究[J]. 2012, (17): 53-64.
- [2] 谈一兵, 伍川生. 我国建筑陶瓷行业新产品发展现状[J]. 中国建材科技, 2020, 29(3):2.
- [3] Zorić B, Matić T, Hocenski Ž. Classification of biscuit tiles for defect detection using Fourier transform features[J]. ISA transactions, 2021.
- [4] Mariyadi B, Fitriyani N, Sahroni T R. 2D Detection Model of Defect on the Surface of Ceramic Tile by an Artificial Neural Network[C]. Journal of Physics: Conference Series. 2021.
- [5] Sioma A. Automated control of surface defects on ceramic tiles using 3D image analysis[J]. Materials, 2020, 13(5): 1250.
- [6] Garrido I, Barreira E, Almeida R M S F, et al. Introduction of active thermography and automatic defect segmentation in the thermographic inspection of specimens of ceramic tiling for building façades[J]. Infrared Physics & Technology, 2022, 121: 104012.
- [7] 权小霞, 李军华, 汪宇玲. 基于局部方差加权信息熵的瓷砖表面缺陷检测[J]. 中国陶瓷, 2019, 55(10): 46-55.
- [8] 苏彩红, 朱学峰, 刘笛. 基于共生矩阵和颜色特征的墙地砖缺陷自动检测的研究[J]. 计算机工程与应用, 2004, 40(11): 229-231.
- [9] Ragab K, Alsharay N. Developing parallel cracks and spots ceramic defect detection and classification algorithm using CUDA[C]. 2017 IEEE 13th International Symposium on Autonomous Decentralized System (ISADS). 2017: 255-261
- [10] Casagrande L, Macar: ni L A B, Bitencourt D, et al. A new feature extraction process based on SFTA and DWT to enhance classification of ceramic tiles quality[J]. Machine Vision and Applications, 2020, 31(7): 1-15.
- [11] Rimac-Drlje Snjezana, Zagar Drago, Rupcic Slavko. Adaptive Image Processing Technique for Quality Control in Ceramic Tile Production[J]. 2010, 52(2): 205-215.
- [12] 周向, 何伟, 何涛等. 一种基于机器视觉的瓷砖定位分割方法[J]. 中国陶瓷, 2016, 52(7): 43-48.
- [13] Zhang H, Peng L, Yu S, et al. Detection of Surface Defects in Ceramic Tiles With Complex Texture[J]. IEEE Access, 2021, 9: 92788-92797.
- [14] 滕博文. 基于深度学习的卫浴陶瓷缺陷检测[D]. 北京:北京机械工业自动化研究所学位论文. 2021.
- [15] 张涛川. 基于深度学习的瓷砖在线快速无损检测系统研发[J]. 2020-09-25.
- [16] 戴玉昭. 基于深度学习的瓷砖釉面瑕疵分类系统研究与实现[D]. 2021-7.
- [17] Guang Wan, Hongbo Fang. Ceramic tile surface defect detection based on deep learning[J]. 2021-12-31.

- [18] 赵鹏. 基于卷积神经网络的光伏电池红外图像缺陷检测[D]. 河北工业大学, 2020.
DOI:10.27105/d.cnki.ghbgu.2020.000899.
- [19] 孔兰芳. 基于卷积神经网络的输电线路小目标检测算法研究[D]. 华中科技大学, 2019.
DOI:10.27157/d.cnki.ghzku.2019.001093.
- [20] Weimer D, Scholz-Reiter B, Shpitalni M. Design of deep convolutional neural network architectures for automated feature extraction in industrial inspection[J]. CIRP Annals, 2016, 65(1): 417-420.
- [21] Masci J, Meier U, Cireşan D, et al. Steel defect classification with max-pooling convolutional neural networks[C]. The 2012 international joint conference on neural networks (IJCNN). IEEE, 2012: 1-6.
- [22] Cha Y J, Choi W, Büyüköztürk O. Deep learning - based crack damage detection using convolutional neural networks[J]. Computer - Aided Civil and Infrastructure Engineering, 2017, 32(5): 361-378.
- [23] Tabernik D, Šela S, Skvarč J, et al. Segmentation-based deep-learning approach for surface-defect detection[J]. Journal of Intelligent Manufacturing, 2020, 31(3): 759-776.
- [24] Lin J, Yao Y, Ma L, et al. Detection of a casting defect tracked by deep convolution neural network[J]. The International Journal of Advanced Manufacturing Technology, 2018, 97(1): 573-581.
- [25] 刘金松. 基于优化 Faster Rcnm 的金属缺陷检测 [D]. 重庆大学, 2020.
DOI:10.27670/d.cnki.gcqdu.2020.002867.
- [26] 张振宇. 基于 Faster R-CNN 的布匹瑕疵检测算法研究 [D]. 温州大学, 2020.
DOI:10.27781/d.cnki.gwzdx.2020.000313.
- [27] 吕晓倩. 基于 Faster R-CNN 的水下目标检测方法研究与实现[D]. 哈尔滨工业大学, 2019.
DOI:10.27061/d.cnki.ghgdu.2019.006214.
- [28] 李丹. 基于 Faster R-CNN 的遥感图像目标检测 [D]. 河北地质大学, 2019.
DOI:10.27752/d.cnki.gsjzj.2019.000057.
- [29] Ojala T, Pietikäinen M, Harwood D. A comparative study of texture measures with classification based on featured distributions[J]. Pattern recognition, 1996, 29(1): 51-59.
- [30] Ahonen T, Hadid A, Pietikäinen M. Face recognition with local binary patterns[C]. European conference on computer vision. Springer, Berlin, Heidelberg, 2004: 469-481..
- [31] Corites C, Vapnik V N. Support-Vector Networks [J]. Machine Learning 1995, 20(3):273-297.
- [32] 李航. 统计学习方法[M]. Qing hua da xue chu ban she, 2012.
- [33] 殷琪林, 王金伟. 深度学习在图像处理领域中的应用概述[J]. 高教学刊, 2018, (9):72-74.
- [34] Hubel D H, Wiesel T N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex [J]. The Journal of physiology, 1962, 160(1): 106-154.
- [35] Huang J T, Li J, Gong Y. An analysis of convolutional neural networks for speech recognition[C]. 2015 IEEE International Conference on Acoustics, Speech and Signal

- Processing (ICASSP). IEEE, 2015: 4989-4993..
- [36] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. nature, 2015, 521(7553): 436-444.
- [37] Wang Xiao-Gang. Deep learning in image recognition Communications of the CCF,2015,11(80):15-23.
- [38] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [39] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
- [40] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.
- [41] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [42] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2014:580-587.
- [43] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 2015, 28: 91-99.
- [44] Girshick R. Fast r-cnn[C]. Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.
- [45] 孙朝云, 裴莉莉, 李伟等. 基于改进 Faster R-CNN 的路面灌封裂缝检测方法[J]. 华南理工大学学报(自然科学版), 2020, 48(2): 84-93.
- [46] Perez L, Wang J. The Effectiveness of Data Augmentation in Image Classification using Deep Learning[J]. arXiv e-prints, 2017: arXiv: 1712.04621.
- [47] Walt S, Schnberger J L, Nunez-Iglesias J, et al. scikit-image: Image processing in Python[J]. PeerJ, 2014, 2: e453.
- [48] Chan R H, Ho C W, Nikolova M. Salt-and-pepper noise removal by median-type noise detectors and detail-preserving regularization[J]. IEEE Transactions on image processing, 2005, 14(10): 1479-1485.
- [49] Goodfellow I J, Shlens J, Szegedy C. Explaining and harnessing adversarial examples[J]. arXiv preprint arXiv:1412.6572, 2014.
- [50] Deng W, Zheng L, Ye Q, et al. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 994-1003.
- [51] Van Etten A. You only look twice: Rapid multi-scale object detection in satellite imagery[J]. arXiv preprint arXiv:1805.09512, 2018.
- [52] Ye Z F, Wen Y M, Lu B L. A survey of imbalanced pattern classification problems[J]. CAAI Transactions on Intelligent Systems, 2009, 4(2): 148-156.
- [53] Zhang H, Wu C, Zhang Z, et al. Resnest: Split-attention networks[J]. arXiv preprint arXiv:2004.08955, 2020.

- [54] Xie S, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1492-1500.
- [55] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.
- [56] Qiao S, Chen L C, Yuille A. Detectors: Detecting objects with recursive feature pyramid and switchable atrous convolution[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 10213-10224.