工业缺陷产品的分类及检测

目录

[1 绪论 3](#_Toc12625827)

[1.1 研究背景及意义 3](#_Toc12625828)

[1.2 国内外研究现状 4](#_Toc12625829)

[2 算法模型 4](#_Toc12625830)

[2.1 神经网络的选择 4](#_Toc12625831)

[2.2 ResNet50结构描述 6](#_Toc12625832)

[2.3 训练过程描述 7](#_Toc12625833)

[2.3.1 环境配置 7](#_Toc12625834)

[2.3.2 数据预处理 7](#_Toc12625835)

[2.3.3 设置训练模型参数 7](#_Toc12625836)

[3 结论 8](#_Toc12625837)

[3.1 baseline分类结果 8](#_Toc12625838)

[3.2 avgpool分类结果 9](#_Toc12625839)

[3.3 实验结果分析 9](#_Toc12625840)

[4 参考文献 10](#_Toc12625841)

1 绪论

1.1 研究背景及意义

传统的工业产品缺陷检测依靠工人手工操作完成，工人的劳动强度大，检测经验也不同，而且长时间的检测也非常影响人的精神状态以及危害人的身心健康，导致检测效率不高，结果一致性差，且不可避免的存在错检、漏检。同时，二极管的图像检测需要极高的技术要求，这就需要投入大量的培养投入，但是仅仅依靠工人去学习标注检测就耗费了大量的人力物力，而且最终的效果因为种种原因也并不理想。

机器视觉检测相对于传统的人工缺陷检测工序简单，检测速度快，效率高。特别是深度学习的崛起让视觉缺陷检测的优势更加明显。而卷积神经网络在图像分类、目标检测等领域取得了巨大的突破。其强大的特征学习和分类能力是我们作为选择和应用的关键。卷积神经网络在图像识别的问题上，准确率可以高达百分之九十九以上，模型仅需训练一次，就能高效的应用，而且预测速度快，准确率高，并且在不断的发展之中。

1.2 国内外研究现状

通过调研以及查阅文献和资料，国内大多数学者或论文在进行图片分类时主要运用了Faster R-CNN分类网络结构、ResNet101特征提取网络、ResNet50、ZFNet、VGGNet等神经网络，通过对不同工业产品数据集的分类效果进行分析，例如：金属轴的表面缺陷、轮胎缺陷、电弧熔积表面缺陷等，得出结论，虽然ResNet在速度上可能略微逊色于其他网络结构，但是准确率却要远远高于其他网络结构，也因此作为我们主要的研究方向和使用结构。

通过对比国内外各大工业缺陷检测公司所使用的基本都是Faster R-CNN以及ResNet网络结构对各类特征进行识别和分类，而且在表面缺陷的分类及检测上在微观上均达到了微米级以上，但同时也发现更多的公司都只注重了表面缺陷的检测和研究，而针对工业产品内部的结构的研究较少，因此，在x射线下成型的工业产品结构缺陷的检测具有很大的应用空间。

2 算法模型

2.1 神经网络的选择

常用的图像分类的经典CNN网络模型种类繁多，例如LeNet、AlexNet、GoogleNet、VGGNet、ResNet等多种模型。

LeNet模型：LeNet是最早的卷积神经网络之一，涉及到60k参数。该模型最早应用于手写数字的识别中，取得了不错的效果，但由于受当时计算效率的低下，该模型深度浅、参数少且结构单一，不适用于复杂的图像分类任务。

AlexNet模型：该网络共涉及约60Ｍ参数。AlexNet有着和LeNet相似的网络结构，但网络层数更深，有更多的参数。相较于LeNet，该模型使用了ReLU激活函数，其梯度下降速度更快，因而训练模型所需的迭代次数大大降低。同时该模型使用了随机失活（dropout）操作，在一定程度上避免了因训练产生的过拟合，训练模型的计算量也大大降低。该模型相较于LeNet模型其深度仅仅增加了3层，其对图像的特征描述及提取能力仍然十分有限。

GoogleNet模型：该网络共涉及5M参数。该模型最大的特点在于引入了Inception模块，GoogleNet相较于之前的网络模型其深度大大增加，达到了史无前例的22层，由于其参数量仅为AlexNet的1/12，模型的计算量大大减小，但对图像分类的精度又上升到了一个新的台阶。虽然GoogleNet模型层次达到了22层，但想更进一步加深层次却是异常困难，原因在于随着模型层次的加深，梯度弥散问题愈发严重，使得网络难以训练。

VGGNet模型：VGGNet是从AlexNet模型发展而来，主要修改了如下两方面：（a）使用几个带有小滤波器的卷积层代替一个大滤波器的卷积层，即卷积层使用的卷积核较小，但增加了模型的深度；（b）采用多尺度训练策略，具体来说，首先将原始图像等比例缩放，保证短边大于224，再在经过处理的图像上随机选取224\*224窗口，因为物体尺度变化多样，这种训练策略可以更好地识别物体。但该网络使用的参数过多，训练速度缓慢。

ResNet模型：该模型旨在解决“退化”问题，即当模型的层次加深后错误率却提高了。其原因在于：当模型变复杂时，随机梯度下降（SGD）的优化变得更加困难，导致了模型达不到好的学习效果。提出了Residual结构，即增加一个恒等映射，将原始所需要学的函数H(x)转换成F(x)+x，假设F(x)的优化会比H(x)简单的多，则这两种表达的效果相同，但是优化的难度却并不相同。该模型的出现，使得网络模型深度得到了很大的提升。ResNet通过额外的分支（gap-fc-fc-sign）来得到每个通道的［0,1］权重，自适应地校正原各通道激活值响应，以提升有用通道响应并抑制对当前任务用处不大的通道响应。该模型不仅在一定程度上减少了计算量，防止了模型训练的过拟合，同时更有利于对图像特征的描述。

因此在进行了上述模型的比较和调研之后，ResNet模型有效减少了计算量，加深了网络深度，在大部分场景下都可以取得较好的结果，故选取ResNet50。

2.2 ResNet50结构描述

Residual Networks是MSRA何凯明团队在2015年提出的，在多项比赛中获得第一名的成绩，并在其相关论文Deep Residual Learning for Image Recognition中提出了Residual block的结构：

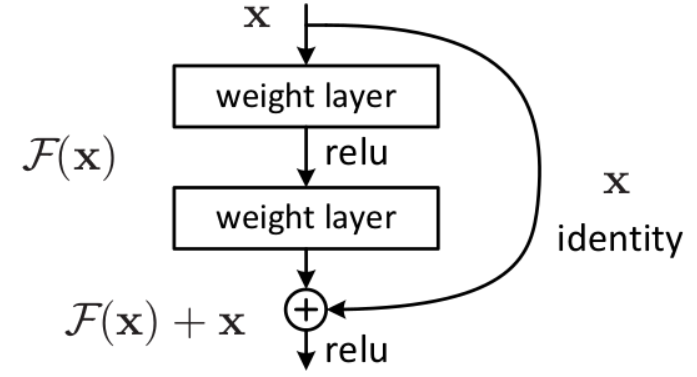


图2.2.1

即增加了一个identity mapping（恒等映射），将原来需要学习的函数H(x)转换成F(x)+x。论文中提出，这二者的优化难度并不相同，相比较于H(x)，F(x)+x的优化要简单，这一想法源自图像处理中的**残差向量编码**，通过一个reformulation，将一个问题分解成多个尺度之间的残差问题，可以更好的获得优化效果。

残差块 Residual block通过shortcut connection实现，通过shortcut将block的输入输出进行一个简单的叠加，这个操作不会给网络增加额外的参数和计算量，却可以大大增加模型的训练速度，提高训练效果，当模型的深度增加时，该结构能够很好的解决退化问题。相比较于学习原始特征H(x)=F(x)+x，我们学习其残差F(x)=H(x)−x要容易。如果我们学习的残差F(x)=0，此时堆积层仅仅作了恒等映射，至少保证网络性能不会下降，实际上残差不会为0, 这会使得堆积层在输入特征的基础上学习到新的特征，从而获得更好的性能。

对于更深的网络，论文中提出了新的Residual block，考虑到计算成本，对深层的网络，将两个3×3的卷积层替换为1×1+3×3+1×1的卷积层，新结构中的3×3卷积首先在一个降维1×1卷积层下减少了计算量，然后在另一个1×1的卷积层下做了还原，既保持了精度又减少了计算量（降维程度为输入输出的1/4），其结构如下：

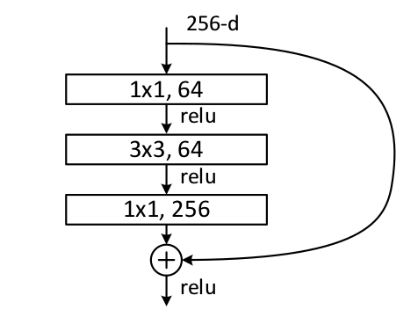


图2.2.2

2.3 训练过程描述

2.3.1 环境配置

本项目使用pytorch实现，环境要求python>=3.5,pytorch>=1.0,numpy,PIL,cv2。

2.3.2 数据预处理

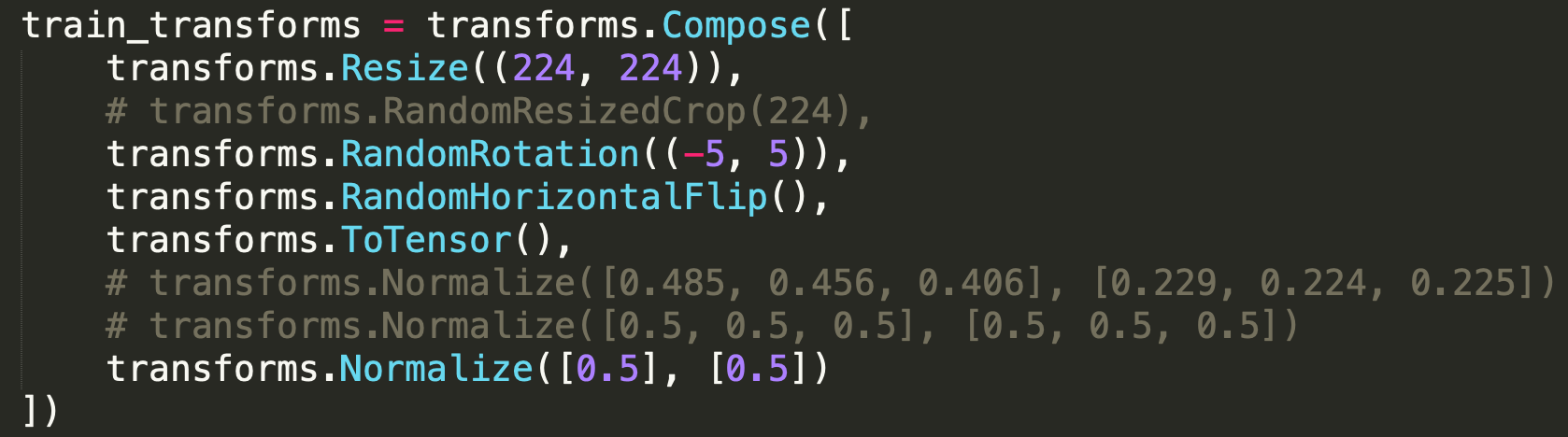


图2.3.1

1. 将数据集中的图像resize为224\*224。
2. 图像在-5到+5随机旋转。
3. 图像进行随机水平翻转。
4. 对图像进行归一化处理，均值设为0.5，标准差设为0.5，使图像在[-1,1]内标准化

2.3.3 设置训练模型参数

本项目采用resnet50网络架构，类别数为9（需要将图片分为BUFF毛边、LF、pin、划痕、金线、气泡、缺陷、未成型、异物9大类），batch\_size为64，学习速率η为0.001，优化参数为0.9，学习率预热参数设为100，迭代次数为50。

$ python train.py

...

06/23 07:11:20 Train Epoch: [44][205/1172] Batch Time 0.453(0.453) Loss 0.359(0.290) Lr 0.000594

06/23 07:11:22 Train Epoch: [44][210/1172] Batch Time 0.450(0.453) Loss 0.259(0.290) Lr 0.000594

06/23 07:11:24 Train Epoch: [44][215/1172] Batch Time 0.448(0.453) Loss 0.348(0.290) Lr 0.000594

...

3 结论

3.1 baseline分类结果

在使用基本ResNet50网络模型的情况下，原始的残差块downsample层shortcut采用了stride=2的1x1卷积分类，准确率达到0.8687，其中pin的分类准确率较低，只达到0.5100。

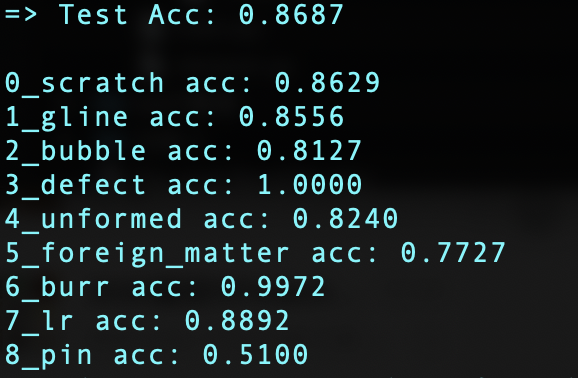


图3.1.1

3.2 avgpool分类结果

通过对网络结构进行分析，发现原始的残差块downsample层shortcut采用的是stride=2的1x1卷积, 丢失了部分信息, 这里将shortcut修改为2x2, stride=2的AvgPool加stride=1的1x1卷积，分类准确率提升到0.8751。

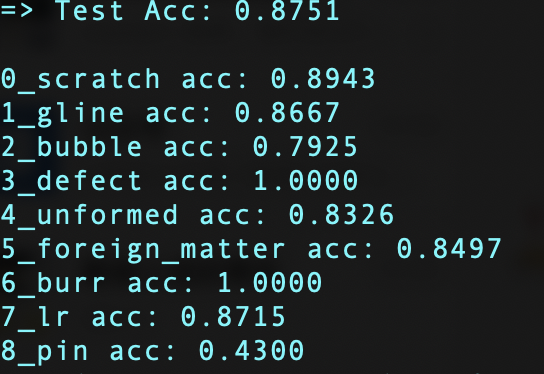


图3.2.1

3.3 实验结果分析

在对二极管进行的检测分类中，各个缺陷的差别较小，且类别pin的数据量较小，影响分类的准确率。但总体准确率已经可以达到0.8751，可以在一定程度上替代原本的人眼产品质检，从而大幅提升工业产品合格率和降低人力成本。

4 参考文献

1. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in CVPR, 2016.
2. 周飞燕，金林鹏，董军.卷积神经网络研究综述[J].计算机学报，2017，40(6):1230-1247.
3. 陈强，朱立新.结合Canny算子的图像二值化[J].计算机辅助设计与图形学学报，2005，17(6):1302-1306.
4. Krizhevsky A, Sutskever II, Hinton G. Imagenet classification with deep convolutional neural networks[J]. Proceedings of the Advances in Neural Information Processing Systems. Lake Tahoe, USA, 2012: 97-110.