

2019 年（第六届）全国大学生统计建模大赛

千里之堤，溃于蚁穴——大型建筑设施裂纹
图像的自动识别研究

参 赛 单 位：长沙理工大学

参赛者姓名：罗均、李桢、朱璐

目录

一、 前言.....	1
(一)、研究背景.....	1
(二)、国内外研究现状.....	1
二、 图片预处理.....	4
三、 卷积神经网络简介.....	5
(一)、卷积的定义.....	5
(二)、卷积神经网络.....	7
3.2.1 卷积神经网络的特征.....	7
3.2.2.稀疏交互与参数共享.....	9
3.2.3 池化.....	9
四、 卷积神经网络的构架.....	11
(一)、Vgg16 模型.....	11
五、 模型的建立.....	12
(一)、预备知识.....	12
5.1.1 迁移学习.....	12
5.1.2 SVM 模型.....	12
5.1.3 ROI(Region of interest)池化.....	14
5.1.4 边框回归.....	14
5.1.5 NMS(非极大值抑制).....	17
(二)、R-CNN 模型.....	18
(三)、Fast R-CNN 模型.....	19
(四)、Faster R-CNN.....	20
5.4.1 RPN 网络.....	20
六、 优化算法.....	22
(一)、Minibatch 梯度下降算法.....	22
(二)、动量梯度下降.....	23
(三)、Adagrad 优化算法.....	24
(四)、Rmsprop 算法.....	24
(五)、Adam 优化算法.....	25
七、 建模及结果分析.....	26

(一)、模型构建.....	26
(二)、训练流程.....	27
(三)、调参过程.....	27
(四)、结果分析.....	27
7.4.1 模型一结果分析.....	27
7.4.2 模型二结果分析.....	29
7.4.3 模型三结果分析.....	30
7.4.4 模型结果对比.....	31
7.4.5 裂纹智能检测系统的建立.....	33
八、模型的改进.....	34
参考文献.....	35
附录.....	37
致谢.....	37

表格和插图清单

图 1	大津分割 (OSTU) 图像二值化.....	2
图 2	裂纹智能检测示意图.....	3
图 3	原图与灰度图.....	4
图 4	原图与直方图均衡.....	4
图 5	原图与同态滤波增强的图片.....	5
图 6	二维离散的卷积计算示意图.....	7
图 7	卷积的局部感受野.....	7
图 8	裂纹滤波器及卷积训练.....	8
图 9	池化下采样.....	8
图 10	卷积网络与全连接网络的对比.....	9
图 11	卷积层的构造.....	10
图 12	池化示意图.....	10
图 13	VGG 网络架构.....	11
图 14	ReLU 函数.....	12
图 15	RoI 池化示意图.....	14
图 16	最小外接矩形与候选区域.....	15
图 17	IOU 示意图.....	17
图 18	非极大值抑制示意.....	17
图 19	传统物体检测与区域卷积神经网络对比.....	18
图 20	R-CNN 模块划分.....	19
图 21	Fast R-CNN 结构图.....	20

图 22	Fast R-CNN 的流程.....	21
图 23	动量梯度下降形象示意图.....	23
图 24	Adagrid 优化算法的适用函数.....	24
图 25	复杂的等高线图.....	25
图 26	模型一训练的总损失.....	28
图 27	模型一未经图片预处理的 Faster R-CNN 测试结果.....	28
图 28	模型二训练的总损失.....	29
图 29	经过同态滤波处理图片的 Faster R-CNN 的测试图片.....	29
图 30	模型三的总损失.....	30
图 31	改进优化策略 Adam 算法的 Faster R-CNN 测试图片.....	31
图 32	三种模型对比.....	32
图 33	模型一，模型二，模型三测试图片对比.....	33
图 34	智能裂纹检测系统.....	34

千里之堤，溃于蚁穴——大型建筑设施裂纹图像的自动识别研究

摘要

国民经济的平稳发展离不开大型建筑设施的保驾护航，因此对于建筑物的安全检测成为了学术界和工业界的焦点之一。众所周知，裂纹是引起安全隐患的重要原因，裂纹检测已是当今的一大研究热点，也是一大难点。本文通过改进传统的裂纹检测算法，采用更快的区域卷积神经网络（Faster R-CNN）识别大型建筑设施的裂纹，并且精确给出裂纹的具体位置及其概率。

传统的裂纹识别算法需要手动提取特征进行分类，而且不能很好的区别划痕、水痕、刻痕等伪裂纹，复杂不高效，其分类效果也不理想。更快的区域卷积神经网络不仅可以自动提取图片的特征，一定程度上区分了伪裂纹，还提高了分类精度且可以定位裂纹在图片中的位置和概率。本文通过对比了三种模型：图片不做预处理 Faster R-CNN、做同态滤波预处理的 Faster R-CNN，和使用 Adam 优化的 Faster R-CNN，发现做同态滤波等预处理的 Faster R-CNN 效果最好，损失值最小为 0.3，分类精度最大可达 0.892。最后，本文还从模型的鲁棒性，模型的功能，模型的应用范围三个方面论证了同态滤波预处理的 Faster R-CNN 优于传统方法。

基于最优的模型，建立了一套自动识别裂纹的系统，其流程为拍摄图片→预处理→Faster R-CNN 处理→得出裂纹位置及概率，具有很强的实用性，极大程度上解放了生产力。需要注意的是，由于实际裂纹图片的分布会随着时间发生变化，因此需要经常加入新的训练图片微调。最后，本文建立的系统对于细小的裂纹无法识别，在识别对安全性能要求非常高的建筑，比如金属建筑，还需要改进。

关键词：Faster R-CNN；支持向量机；边框回归；自动识别系统；裂纹检测

A small leak will sink a great ship-- large building facilities crack image automatic identification research

Abstract:

The steady development of national economy is inseparable from the escort of large building facilities, so the safety inspection of buildings has become one of the focal points of academia and industry. As is known to all, crack is an important cause of safety hazards. Crack detection has become a hot research topic nowadays, especially the crack detection based on complex background images, which is also a difficult problem. In this paper, by improving the traditional crack detection algorithm, Faster regional convolutional neural network (Faster R-CNN) is adopted to identify large building facility cracks, and the exact location and probability of cracks are given.

Compared with the traditional crack recognition algorithm, which requires manual feature extraction for classification, it is complex and inefficient, and it cannot distinguish the pseudo-cracks such as scratches, water marks and nick marks well, so its classification effect is not ideal. The faster regional convolutional network can not only automatically extract the features of the image, but also distinguish the pseudo-cracks to some extent, improving the classification accuracy and locate the location and probability of cracks in the image. In this paper, three models are compared: Faster R-CNN without preprocessing, Faster R-CNN of homomorphic filtering and other preprocessing, and Faster R-CNN of the use of Adam optimized. Then, it is concluded that the best effect is achieved by Faster R-CNN of homomorphic filtering and other preprocessing, with the minimum loss value of 0.3 and the maximum classification accuracy of 0.892. Finally, it is proved that the Faster R-CNN preprocessing with homomorphic filter is superior to the

traditional method from the three aspects of robustness, function and application scope of the model.

Based on the best model, a set of automatic crack identification system is established. The process : taking pictures -- > pretreatment -- >Faster r-cnn processing -- >obtain the crack location and probability , which has a strong practicability and liberates productivity to a great extent.

It should be noted that as the distribution of actual crack images will change over time, new training images should be added to fine-tune frequently. Finally, the system established in this paper cannot recognize fine cracks, and it needs to be improved in identifying buildings with high safety performance requirements, such as metal buildings.

Keywords : Faster R-CNN; SVM; Bounding box regression ; Automatic identification system ; Crack detection

一、前言

（一）研究背景

中国建筑业的稳步增长，对国民经济发展，奉献出了强有力的支撑。与此同时，在中国建筑业蓬勃发展的今天，为确保社会主义建设稳中求进，就需要对建筑物的安全性能做评估。

在评估建筑物的安全性能时，有很多方面需要考虑，如建筑物的裂纹、抗震性能、耐久性能和承载能力等都是很重要的因素。由于裂纹是引起安全隐患的重要原因，本文专门探讨大型建筑物表面的裂纹检测。

本文研究的另一重点就是区分伪裂纹与裂纹。建筑上的划痕、水痕、刻痕也属于研究范围内，这些痕迹与裂纹十分相像，但这些对建筑物的安全性能不构成威胁，所以被称为伪裂纹。根据观察可得：裂纹的长度一般大于伪裂纹结构；裂纹所在片段比例一般小于伪裂纹片段比例；裂纹走向和分布无规律，伪裂纹一般比较规律，其形状近似于曲线。

（二）国内外研究现状

传统的建筑物裂纹检测是人工攀登建筑物，肉眼检测。此种做法不安全，不高效率也不准确全面。后来利用机器人摄像，高清摄像机拍摄建筑表面，再基于图像检测裂纹。

如今，基于裂纹的图像研究已有了显著的成果，但是主要是针对混凝土建筑墙面、沥青材质的路面、桥梁建筑地面等，其研究的背景相对简单，而且以上建筑物的裂纹在图片中较为明显，通常采用传统图像识别就可以达到要求。

基于图像的传统裂纹识别方法包含两个步骤，第一个步骤是图片预处理，以降低图片噪声并增强图片的裂纹特征；第二个步骤是根据之前预处理的图片人工进行裂纹特征提取，构建一个二分类器如决策树，支持向量机，逻辑回归来判断是否有裂纹。

针对图片预处理这一步骤：Fujita et al 在 2006 年通过图像拉伸技术对图片进行预处理，实现降噪性能[1]。Jabakhanji 在 2007 年运用直方图均衡技术，提升图片的清晰度，达到图片整体对比度的效果，这项技术在裂纹图像处理中十

分高效[2]。除此之外，Jabakhanji D et al 在 2007 年[2]，Chen H O 在 2010[3] 使用灰度级腐蚀预处理图像，差异化背景和前景，取得良好的效果。但是这些处理的方法在复杂背景和包含伪裂纹的图片中效果不显著，前景和背景区分不明显。比如日本学者大津于 1979 年提出 OSTU，利用前景与背景图像类间方差最大，确定阈值二值化图像。利用该算法可以识别背景结构简单的图像，略去了很多重要的裂纹细节。如图 1，左图背景结构简单，无伪裂纹，二值化效果良好，右图背景结构复杂，存在伪裂纹，二值化去掉了重要的细节，效果差。

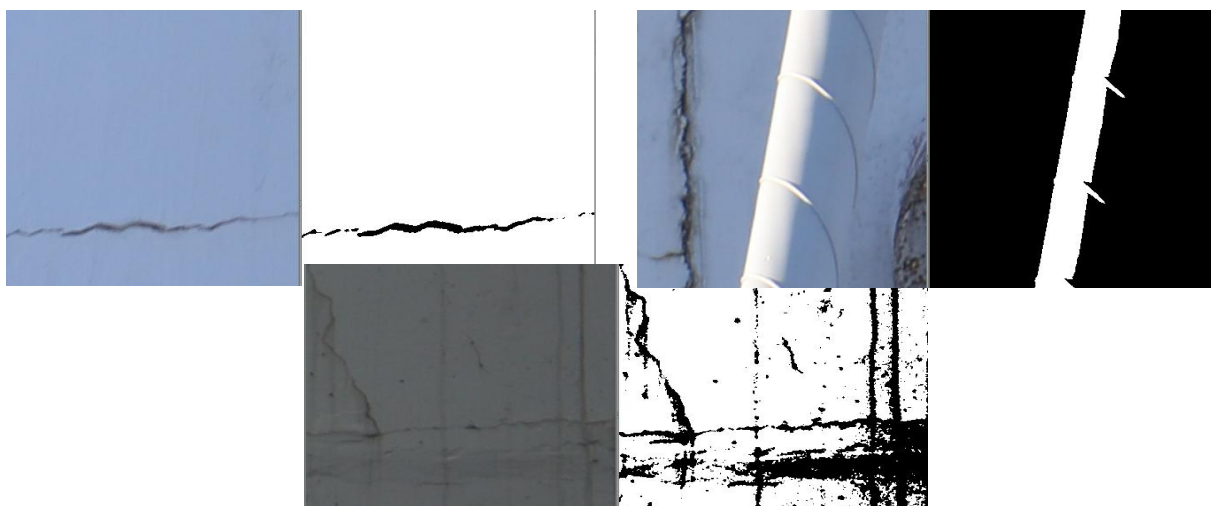


图 1 大津分割（OSTU）图像二值化

传统裂缝识别的第二个步骤是人工提取图片的特征：Zhenhua Guo 在论文中提出了一种基于局部二值模式(LBP)的裂纹检测算法[4]，[5]提出了一种基于树结构的裂纹自动检测方法——裂纹树，[6]也引入了一套完整的用于检测和表征路面裂缝病害的图像处理算法。一方面，在基于图片的裂纹检测中，要尽可能区分裂纹和伪裂纹就需要选择合适的特征。因此针对以上的特征一般采用下面的特征来区分图片是否有裂纹：区域面积，周长，灰度均值，最大连通区域宽高比，图片二阶矩，偏心率等等。需要指出的是，这种手工提取特征的方法在区分复杂背景结构和区分伪裂纹上不够有辨别力。这是因为：准确选择区分图片是否存在裂纹的特征存在困难，常常受到伪裂纹特征的干扰，特征存在漏掉不全的可能性大；特征需要人工选择，花费的人力物力较大；在特征提取时信息不可必然的遗漏，和人工选择特征存在的误差也使得分类效果并不十分显著。于是，对于裂纹图像的研究有了新的难度。

另一方面，随着科学技术的发展，计算机运算速度的极大提高，深度学习技术在近十年内得到飞速发展，在许多医学成像和计算机视觉任务中表现出了许多令人印象深刻的性能，明显地展示了深度神经网络学习的深度特征的有效性[7-12]，这些特征可能会取代传统手工制作的特征[13]。深度学习在裂纹检测领域也卓有成效，Lei Zhang 等人提出了一种用于路面裂纹检测的深度卷积神经网络，通过与传统的手工提取特征相比，其提出的深度学习框架所提取的深度特征具有更好的裂纹检测性能[14]。常海涛等将 Faster R-CNN 用于工业 CT 图像缺陷检测，可以有效识别出裂纹并准确缺陷的标志标记[15]。

本文采用 3 种 Faster R-CNN 模型检测大型建筑的裂纹，摆脱了传统的手动选取特征技术，对图片进行自动的学习数据特征，进行分类，并且能够定位出裂纹在图片中的位置。模型包括图片不做预处理 Faster R-CNN，做同态滤波预处理的 Faster R-CNN，使用 Adam 优化策略的 Faster R-CNN。最终基于分类精度选择出同态滤波预处理的 Faster R-CNN 模型。

在与现有传统方法对比时，本文以卷积神经网络提取特征然后再分类的这一现有方法（简称卷积神经网络分类方法）为例，分别从模型的鲁棒性，模型的功能，模型的应用范围三个方面探讨 Faster R-CNN 裂纹识别与现有裂纹识别之间优缺点，发现同态滤波预处理的 Faster R-CNN 综合效果更好。

基于上述最优模型，建立裂纹智能检测系统，整个流程如图 2：拍摄图片→预处理→输入模型（更快的区域卷积神经网络）→裂纹识别并定位。便能够对裂纹图像进行可靠的识别，区分出裂纹与伪裂纹。高效方便快捷，省时省力。

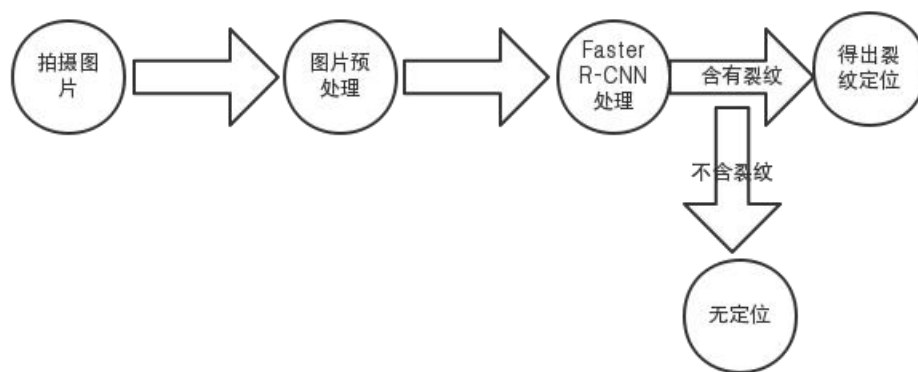


图 2 裂纹智能检测示意图

二、图片预处理

本文使用湖南拓达结构监测技术有限公司提供的图片数据。由于原图片的裂纹识别能力较低，伪裂纹的干扰性及光照强度不一，所以需要图片进行一定程度的预处理，以便更好的识别出裂纹。首先人工标注图片中的裂纹，再生成 xml 文件，xml 本身是一种格式规范，是一种包含了数据，以及数据说明的文本格式规范。其次，运用图片灰度化技术，对图片进行灰度化操作，即将彩色图片的三通道 (r, g, b) 加权变成单通道 gray 的灰度图片：

$$\text{gray} = 0.229r + 0.587g + 0.114b \quad (2.1)$$

这样可以减小图片的数据量，减少计算量，同时也是后续对图片进行滤波的处理的基础，见图 3；

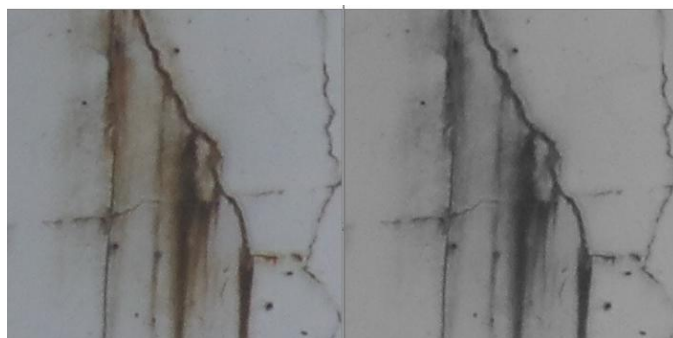


图 3 原图与灰度图

再利用直方图均衡化技术，此技术能够增强图片的对比度，使得图片看起来更加清晰，它将一副图片的直方图分布变成近似均匀分布，来增强图像的对比度。其处理后的效果非常明显，见图 4；

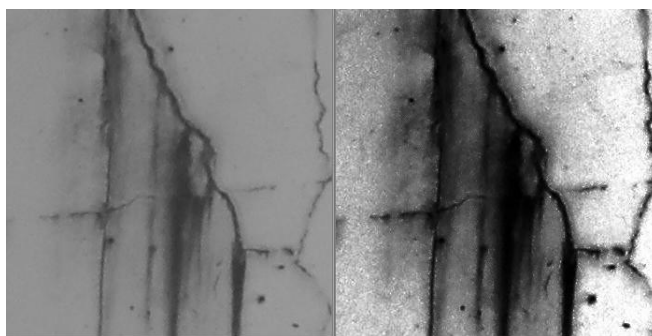


图 4 原图与直方图均衡

最后进行同态滤波增强，同态系统是一种非常特殊的非线性系统，它遵从广义的叠加原理，然后利用此原理进行滤波操作，同态滤波是把频率过滤和灰度变换结合起来的一种图像处理方法，它需要依靠图片的照度模型，以此作为频域处理的基础，使用压缩亮度范围、增强图片对比度来改善图片的质量。其技术的基本原理是将像元灰度值，看成照度和反射率两个组份的产物。由于图片照度的相对变化很小，即可以将之看作为图片的低频成份，而反射率则可看作图片的高频成份。通过分别处理图片的照度和反射率对像元灰度值的影响，达到揭示阴影区细节特征的目的；

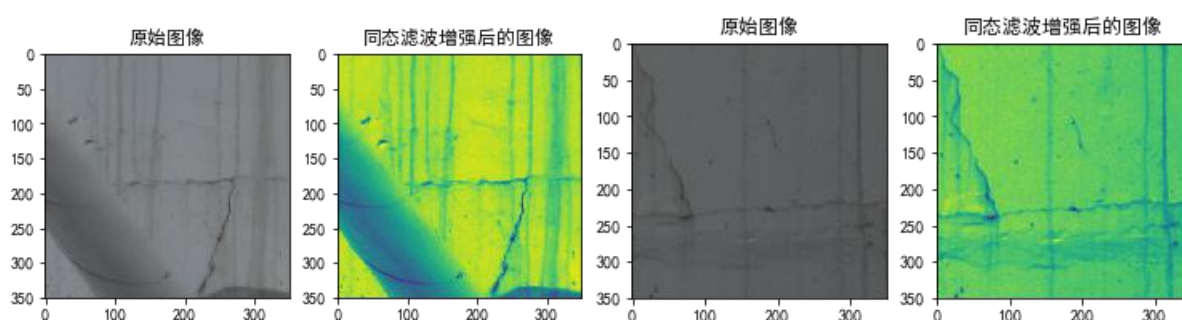


图 5 原图与同态滤波增强的图片

三、卷积神经网络简介

用传统的 bp 神经网络架构，直接是将图片的像素矩阵展开成一列或者一行直接作为输入向量，这样做的缺点是计算量太大，而且将许多背景信息也考虑进去了，给神经网络的分类或者定位任务增添了许多噪声，因此做出来的效果不理想。因此我们引入了卷积神经网络。卷积神经网络因其权值共享的特点，降低了网络模型的复杂度，减少了权值的数量，避免了传统识别算法中复杂的特征提取和数据重建过程，同时能够有效挖掘出数据的局部特征用于分类，目前广泛应用于图像识别和语言处理等领域。

(一) 卷积的定义

卷积是对两个实变函数的一种数学操作。卷积的操作通常可以用星号表示：

$$s(t) = (x * w)(t) \quad (3.1)$$

其中 w 为有效的概率密度函数, 不然其输出不会是一个加权平均。除此之外, 当参数小于 0 时, w 的取值不能不为 0, 不然它预测的是未来, 而正常情况下, 这不是我们可以预测出来的。一般情况下, 卷积可以被定义在满足上面积分表达式的任意一个函数上, 也可以用作加权平均之外的目的, 具体视实际情况而定。

卷积的第一个参数 x 是卷积运算的输入, 第二个参数, 即函数 w 被称为核函数, 输出被称为特征映射。

在假设 x 和 w 被定义在整数时刻 t , 定义离散卷积如下:

$$s(t) = (x * w)(t) = \sum_{a=-\infty}^{\infty} x(a)w(t-a) \quad (3.2)$$

例如, 如果把一张二维的图像 I 作为卷积的输入, 那么其核函数 K 通常也是一个二维的, 表达式如下:

$$s(i, j) = (I * K)(i, j) = \sum_{a=-\infty}^{\infty} I(m, n)K(i-m, j-n) \quad (3.3)$$

易得, 卷积具有交换性

$$s(i, j) = (K * I)(i, j) = \sum_{a=-\infty}^{\infty} I(i-m, j-n)K(m, n) \quad (3.4)$$

如下图, 二维离散的卷积可以看做是矩阵的矩阵对应元素相乘再求和。

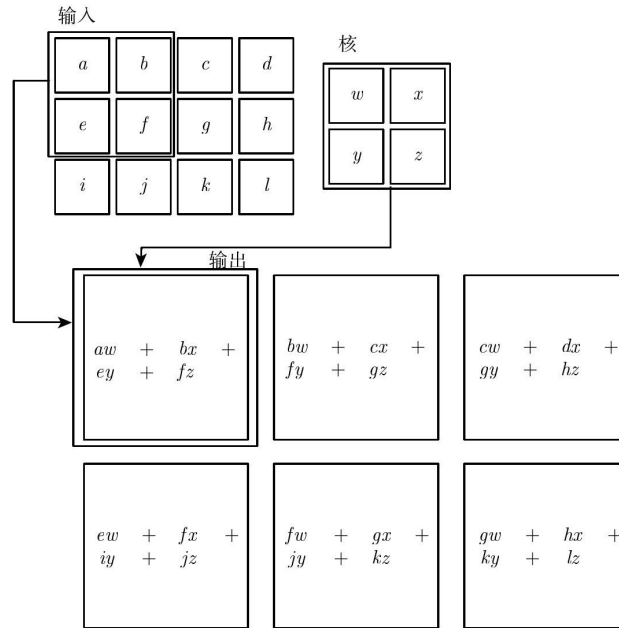


图 6 二维离散的卷积计算示意图

(二) 卷积神经网络

卷积神经网络是一种多层的监督学习神经网络，它的结构主要包含六个部分：输入层、卷积层、池化层、激活函数层、全连接层和分类层，其中提取特征的核心模块是卷积层与池化层。

3.2.1 卷积神经网络的特征

卷积神经网络有三个特征：1，局部感受野；2，共享权重和偏置；3，池化。

卷积神经网络具有局部感受野,如图 7，不同全连接网络考虑单个像素，而卷积考虑的是红色框中的所有像素。由于局部像素之间联系紧密，距离较远的像素联系较弱神经元没有必要对全图进行感知，先通过感知局部，然后再在更深层将这些局部信息综合，层层提取信息，综合起来也能达到获取全部信息的效果。通常物体的目标特征在图片的不同区域，卷积神经网络可以通过卷积的滑动，在不同的位置提取相同的特征，大大减少了参数的数量，避免了神经网络的过拟合。



图 7 卷积的局部感受野

卷积神经网络具有共享权重和偏置，通过卷积识别一定范围内的特征，即视野更宽。如下图 8，设置一个裂纹滤波器，即识别出裂纹。裂纹的区域通常比图片要小得多，卷积神经网络可以共享权重和偏置，大大减少参数的数量。

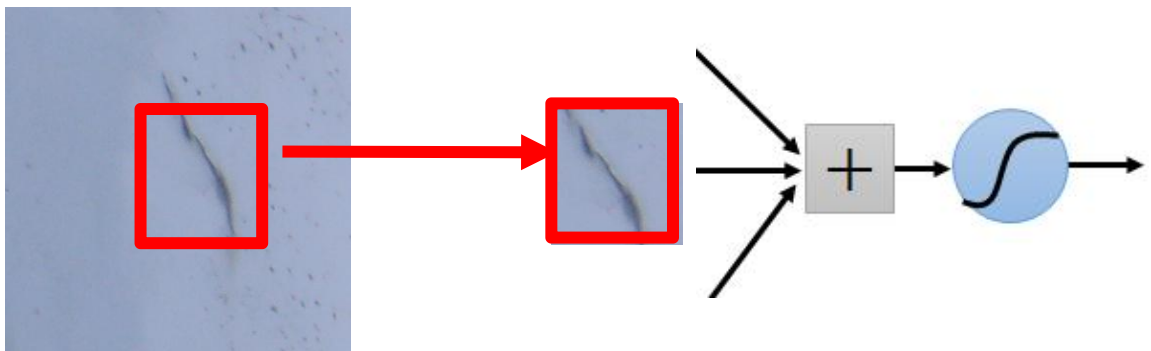


图 8 裂纹滤波器及卷积训练

池化是一种下采样的方法，即用单个像素值代替一个小区域内的像素值，将一张分辨率较高的图片转化为分辨率较低的图片，只保留图中重要的特征信息，但是却不改变物体的结构和空间关系。这种做法的优点是减少输入参数，在减少计算量的同时也防止了过拟合。如图 9，对裂纹图像进行下采样，虽然分辨率下降了，即数据量减少，但是并未丢失裂纹的重要特征。

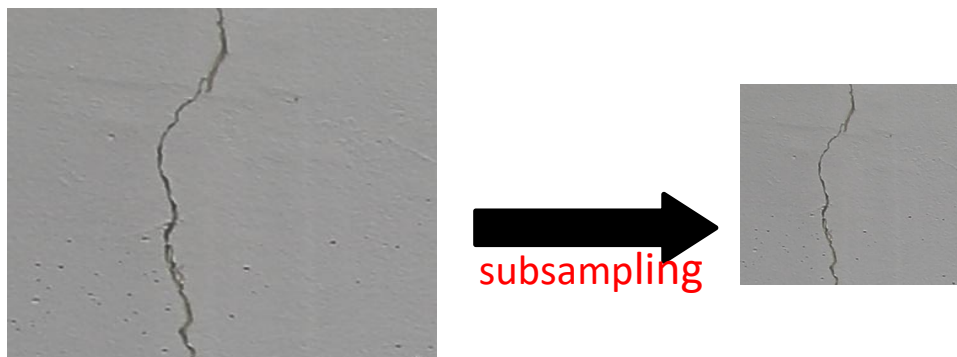


图 9 池化下采样

3.2.2.稀疏交互与参数共享

传统的神经网络可以看成是矩阵的乘法和一系列非线性函数激活的组合。首先建立相应的损失函数，通过输入矩阵和网络的结构进行梯度下降，迭代循环至收敛，从而求得参数矩阵。参数矩阵的每一个元素都描述了一个输入神经元的交互。如图 10，换句话说，传统的神经网络是全连接网络，每一个输入单元都和输出单元产生交互。反观卷积神经网络则是稀疏交互的，原因是卷积的核函数大小远小于输入矩阵的大小。例如，在图片处理中，卷积神经网络的输入图片经常是成百上千个像素点，如果使用全连接网络，那么储存的模型参数非常大，提高了存储模型的硬件要求，而且统计效率不大。在卷积神经网络中，设置只有几十或者几百个参数的卷积核，就可以检测到图片的边缘特征，这些特征然可以在高层的卷积层进一步提取。

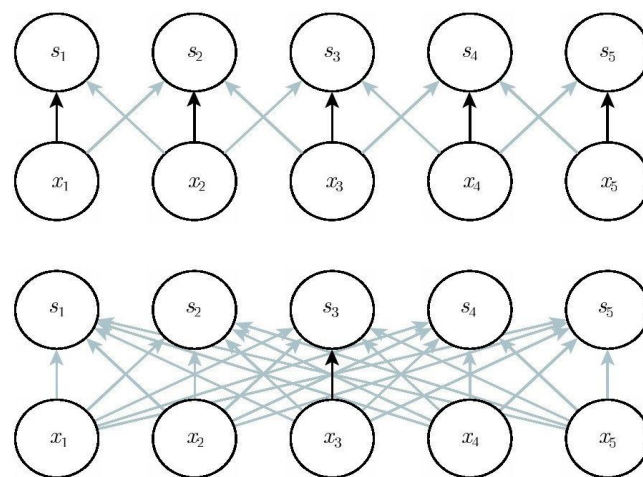


图 10 卷积网络与全连接网络的对比

参数共享，即在一个模型中，各参数的值在多个函数中被重复使用。由于传统神经网络是全连接结构，在计算神经元的输出时，权重矩阵的每一个参数只用了一次。

3.2.3 池化

卷积神经网络的一个典型层包含三级。如图 11，第一级叫卷积层，并行地计算卷积，然后使用线性激活函数对其激活。第二级叫非线性激活层，对前一步的结果做非线性激活，常用的激活函数包括 Relu, sigmoid, tanh 等。这一级也被称为探测级。在第三级中，使用池化函数对上一级的输出进行下采样，进一步调整这一层的结果。

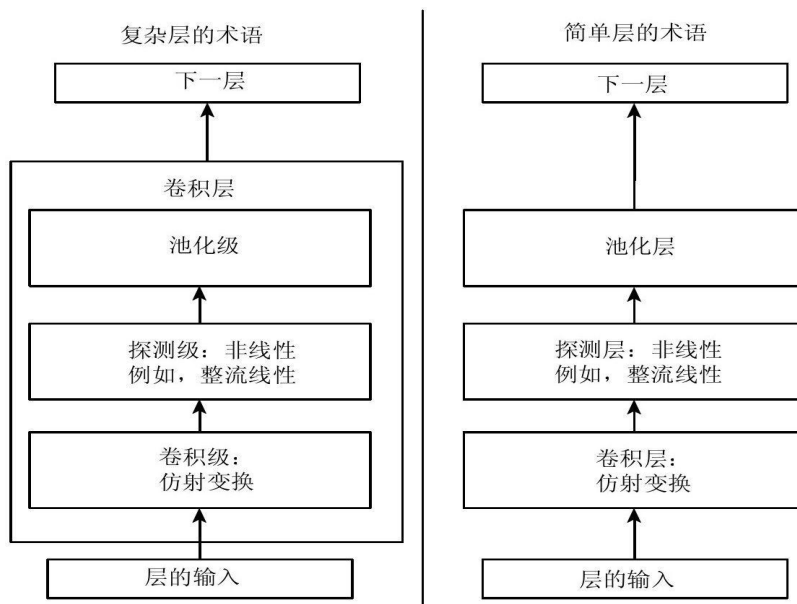


图 11 卷积层的构造

池化，即是用样本某一个位置相邻输出的总的统计特征代替网络在该位置的输出。简单地讲是用一个统计特征代替一定区域的特征。池化主要包括最大值池化和均值池化。例如，平均池化是对相邻矩形区域求平均，用平均值代替区域特征值；最大池化则是用相邻矩形区域内的最大值代替这一区域特征值。常见的池化还有，L2 范数，基于中心的加权平均池化。

事实上，池化具有近似平移不变形，即对输入做轻微的平移，池化后的结果不会和之前的结果发生太大的变化。池化的这种性质十分有用，特别是当我们只关心某个特征出现与否而不关心其出现的具体位置时。在进行裂纹识别时，我们只关心是否出现裂纹，却不用关心裂纹出现的位置。

除此之外，池化还具有下采样的功能，如图 12 所示，这是一个带有下采样功能的最大值池化。其中池化的范围是三个神经元，池化之间的步长是两个神经元。这样做的好处是保存了一定区域内神经元的主要特征，而且又把输出神经元减少了一半，极大地降低了计算和统计的负担。

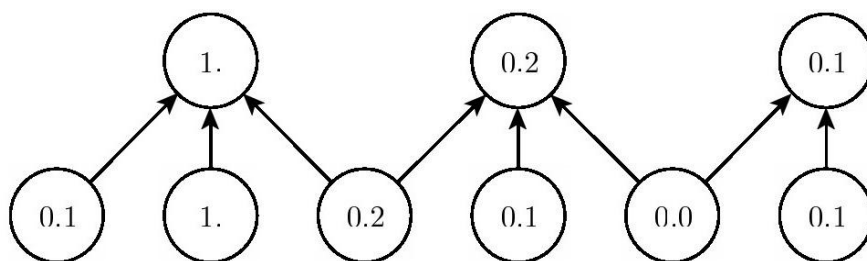


图 12 池化示意图

四、卷积神经网络的构架

(一) Vgg16 模型

2014 年, Karen Simonyan 在研究中表明, 在现有的技术配置上加深卷积神经网络的层数, 即提升至 16 到 19 个权重层能够有效提升网络的性能, 提取图像中的特征用于精准分类。该论文提出的模型即 vgg16, 网络结构包含 16 个权重层, 其中 13 个为卷积层 (包含 RELU 激活函数层), 3 个为全连接层, 卷积层全部使用大量 3X3 大小的卷积核, 如图 13:

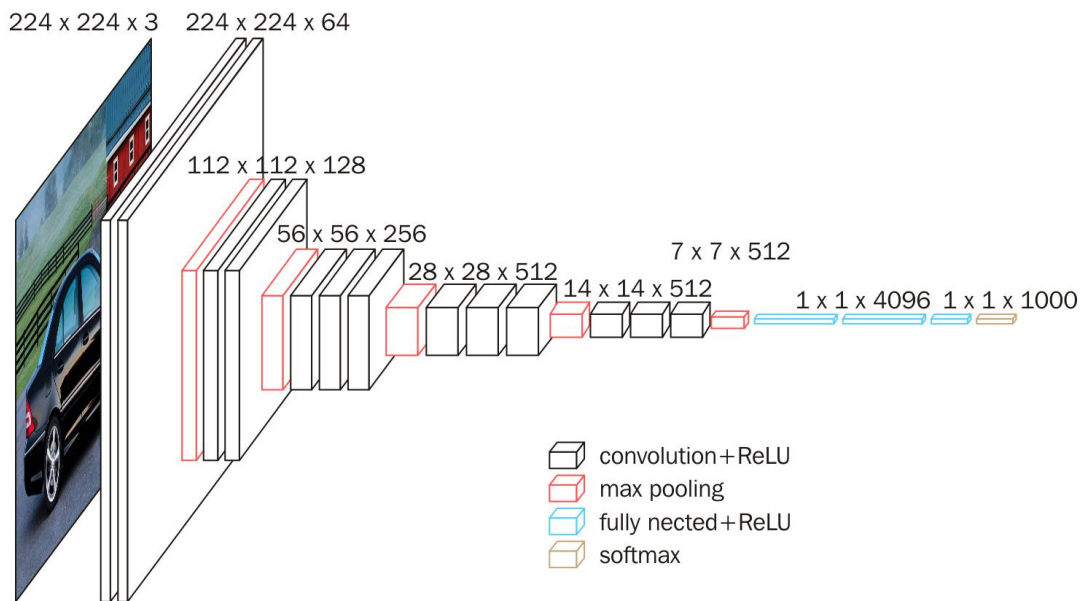


图 13 VGG 网络架构

在神经网络中, 激活函数的作用是能够给神经网络加入一些非线性因素, 解决线性模型表达能力不够的问题, 避免神经网络中的梯度消失问题, 使得神经网络可以更好地解决较为复杂的问题。

在标准神经网络中最常使用的非线性函数是 S 形函数, 例如 $y = \tanh(x)$, 但是 Nair 和 Hinton (2010) [16] 的研究表明饱和和非线性会减慢计算速度, 因此, 他们在研究中使用 ReLU($y = \max(0, x)$) 作为非线性激活函数。如图 14, ReLU 的梯度始终为 0 和 1, 这个特性有助于加快计算速度, 并且提高计算的精度。

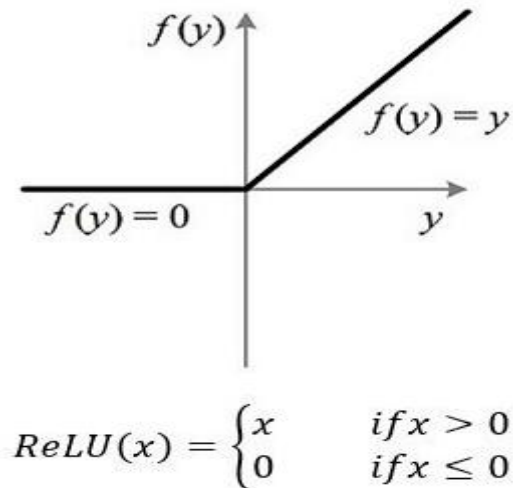


图 14 ReLU 函数

五、模型的建立

(一) 预备知识

5.1.1 迁移学习

在训练深度卷积网络时，通常需要大量的数据。当数据量较小时，建立的模型通常会出现严重的过拟合，不能达到很好的效果。这个时候常常采用迁移学习的方法，通俗地讲，迁移学习就是把训练好的某一个网络的某些层的节点的权重迁移到一个全新的网络，而不是从头开始，为某个特定的任务重新训练一个网络。

常用的方法是用 AlexNet、VGG 等深度网络结构在 ImageNet 等训练集上进行网络预训练，然后再对这些网络最后面的权重层进行重新微调训练。一般来说，在大型数据集上预训练得出的模型，越靠近底层，提取出来的特征越基础，如边框特征、颜色特征，这些特征在小数据集上依然适用。在越靠近顶部的地方提取出的特征越具体，比如说如果训练猫狗分类的数据集，最后几层网络得出来的特征就是猫狗的眼睛，耳朵，鼻子这些特定数据集的特征。因此，在定位裂纹的特征时，往往采用先在 ImageNet 数据集上预训练，然后锁定卷积网络的前几层，对最后几层进行重新训练。这样的好处一是避免了过拟合，二是可以利用别人的模型参数，节约大量的时间成本。

5.1.2 SVM 模型

SVM (Support Vector Machine) 即支持向量机，是由 Vapnik 等人在研究统

计学理论的基础上，针对线性分类器，提出的一种设计最佳准则，可以看成一种二分类模型，SVM 总的来说可分为两大类：线性 SVM 和非线性 SVM，该模型的主要思想是找到空间中的一个超平面，该超平面能够将所有的数据样本划开，且能够使得这些样本集合中的所有数据，到此超平面的距离最短。

该模型的基本原理可以从线性可分思想开始，接着扩展到线性不可分的情况。然后再扩展到非线性函数中，即找到一个非线性映射 m ，把数据样本空间映射到一个高维的特征空间中（甚至无穷维的特征空间中），该空间可以为我们的熟悉的 Hilbert 空间，这样便可以在原来的数据样本空间里，把非线性可分问题，转化为在高维特征空间中的线性可分问题。这种操作可简单概括为升维和线性化。

升维，即把样本空间映射到高维空间，在一般的问题中，这种做法会增加计算的复杂性，使得计算量大大超过实际的应用，因而这种升维的方法没有受到广大群众的欢迎，但是如果只在作为分类、回归等问题中，有些样本集在低维的样本空间中无法进行线性处理，但是在高维特征空间中，可以通过线性超平面进行线性分类或者回归。

基于上述升维的复杂性，SVM 模型十分巧妙的解决了该难题：它利用核函数的展开定理后，不需要知道非线性映射的显式表达式，并且由于 SVM 模型，是在高维特征空间中创建线性学习机，所以与原线性模型比较，几乎不会增加计算的复杂性，并且在某种程度上来说，会避免维数过高带来的计算问题，以上的实现都要归功于核函数的展开定理，和核函数的计算理论。

选择不同的核函数，可以生成不同的 SVM，常用的核函数有以下 4 种：

$$\text{线性核函数: } K(x, y) = x^T * y; \quad (5.1)$$

$$\text{径向基核函数: } K(x, y) = \exp(-|x - y|^2 / d^2); \quad (5.2)$$

$$\text{多项式核函数: } K(x, y) = (\alpha x^T * y + c)^d; \quad (5.3)$$

$$\text{Sigmoid 核函数: } K(x, y) = \tanh(\alpha x^T y + c); \quad (5.4)$$

注：基于核函数的选取，任何半正定的函数都可以作为核函数，不同的核函数，对应的 SVM 模型各异。

SVM 模型问题可以简要的表示成为凸优化模型问题，即最优化的应用，因此我们可以利用优化领域中已知的有效算法，来计算出目标函数的全局最小值。而对于局部最小值来说，一般是基于规则的分类器和人工神经网络这种分类方法，采用基于贪心学习的策略来搜索假设空间。SVM 控制模型的能力是基于最大化决策边界的边缘。但是，用户还是需要提供某些参数，如上面说到的核函数，或者松弛变量等。

5.1.3 ROI(Region of interest)池化

R-CNN 采用的是先通过选择性搜索获得 2000 个候选区域，然后利用卷积神经网络提取候选区域的特征，这样做十分耗内存和时间，因为一张图片相当于用了 2000 次左右的卷积神经网络候选区域。在改进的 fast R-CNN 和 faster R-CNN 中，对于每一张图片，先用卷积神经网络对其提取特征，然后再在特征图上提取候选区域，如此可以减少大量内存空间和时间成本。

ROI 池化就是对提取的候选区域进行处理的一种方法。首先在原图上利用选择性搜索的方法提取相关候选区域，然后利用卷积不改变图片中物体的相对位置的这一特性，将在原图上提取的候选区域一一对应到特征图上。最后再将特征图上的候选区域进行 ROI 池化。注意对于任何尺寸的特征图都可以将其变成固定长度的特征向量，最后输入到全连接层。

如图 15，特征图上的候选区域（绿色边框内的区域）即为 ROI 区域。首先将 ROI 区域的卷积特征分成 $H \times W$ 网格（注：在 VGG 结构中使用的是 7×7 的网格），每个格子里面的特征进行 Max pooling。最后将向量展开成一列即可得到 $H \times W$ 维的列向量。

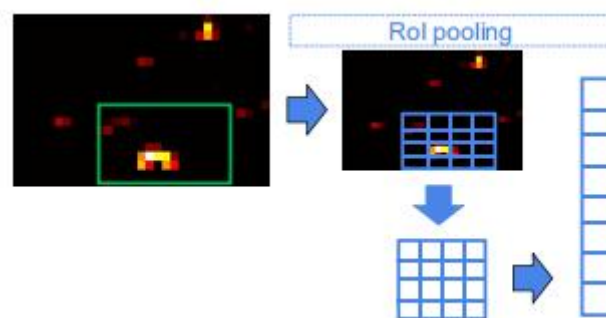


图 15 ROI 池化示意图

5.1.4 边框回归

如图 16，绿色框是飞机的真实框 Ground Truth，又称最小外接矩形。红色框为 Region proposal（候选区域）。虽然候选区域包含的绝大部分的目标，但是仍然遗漏了一部分，定位不精确。因此需要对候选区域的边框进行缩放和平移以修正，使得红色边框和绿色边框尽可能重合。

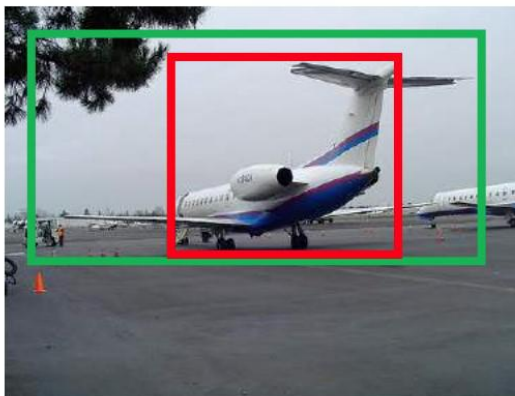


图 16 最小外接矩形与候选区域

无论是 Ground Truth 边框还是 Bounding box 边框，它们都是用四维向量 (x, y, w, h) 表示。其中， (x, y) 表示边框的中心位置， (w, h) 代表其宽和高。

边框回归，顾名思义就是将红色边框的位置大小变换使得尽可能和绿色框重合。公式如下表示：

$$f(p_x, p_y, p_w, p_h) \approx (G_x, G_y, G_w, G_h) \quad (5.5)$$

其中 (p_x, p_y, p_w, p_h) 表示 Bounding box1 边框， (G_x, G_y, G_w, G_h) 为 Ground Truth 边框。

首先，对 (x, y) 进行位置修正，即对边框的中心位置进行平移 $(\Delta x, \Delta y)$ ：

$$\Delta x = P_w d_x(P) \quad (5.6)$$

$$\Delta y = P_h d_y(P) \quad (5.7)$$

那么修正后得到的中心位置为：

$$\hat{G}_x = P_x + P_w d_x(P) \quad (5.8)$$

$$\hat{G}_y = P_y + P_h d_y(P) \quad (5.9)$$

然后对边框 (w, h) 进行缩放，缩放尺度为 (S_w, S_h) ，公式为：

$$S_w = \exp(d_w(P)) \quad (5.10)$$

$$S_h = \exp(d_h(P)) \quad (5.11)$$

则修正过后的边框的宽和高为：

$$\hat{G}_w = P_w \exp(d_w(P)) \quad (5.12)$$

$$\hat{G}_h = P_h \exp(d_h(P)) \quad (5.13)$$

综上，修正后的边框 $(\hat{G}_x, \hat{G}_y, \hat{G}_w, \hat{G}_h)$ 要近似等于 Ground Truth 边框 (G_x, G_y, G_w, G_h) 。

在 RCNN 模型中，损失函数为采用平方损失函数并采用 L_2 正则化：

$$W_* = \arg \min_{\hat{w}_*} \sum_i^N (t_*^i - \hat{w}_*^T \phi_5(P^i))^2 + \lambda \| \hat{w}_* \|^2 \quad (5.14)$$

其中 $\phi_5(P^i)$ 表示卷积神经网络第五层 CONV5 的特征向量，上标 i 表示第 i 个边框对 $\{(P^i, G^i)\}_{i=1, \dots, N}$ ，而 t_* 表示如下：

$$t_x = (G_x - P_x) / P_w \quad (5.15)$$

$$t_y = (G_y - P_y) / P_h \quad (5.16)$$

$$t_w = \log(G_w / P_w) \quad (5.17)$$

$$t_h = \log(G_h / P_h) \quad (5.18)$$

经过最小化损失函数，得到 W_* ，并使得 $d_*(P) = W_*^T \phi_5(P)$ ，最后得到估计：

$$\hat{G}_x = P_x + P_w d_x(P) \quad (5.19)$$

$$\hat{G}_y = P_y + P_h d_y(P) \quad (5.20)$$

$$\hat{G}_w = P_w \exp(d_w(P)) \quad (5.21)$$

$$\hat{G}_h = P_h \exp(d_h(P)) \quad (5.22)$$

5.1.5 NMS(非极大值抑制)

IOU 是目标检测中一个非常重要的概念，是在图片数据集中检测相应物体准确度的一个测量标准。如图 17, IOU 表示两个框重合区域（交集）与总区域（并集）的比值。IOU 越接近 1 表示两个框越接近。

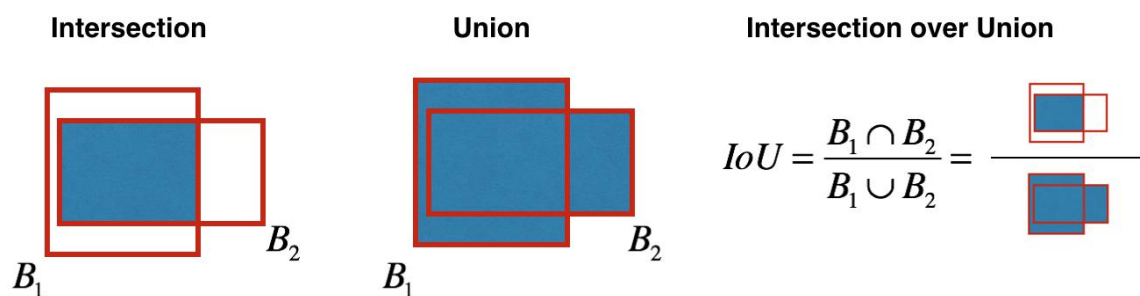


图 17 IOU 示意图

非极大值抑制 (Non-Maximum Suppression), 从字面便可以理解, 抑制不是极大值的元素, 即可理解成优化中的局部最大搜索。其局部表示的是某个邻域, 该邻域有两个可变参数, 邻域的维数和邻域的大小。非最大值抑制采用的是边缘细化技术, 一般应用于“减薄”边缘, 它可以抑制除了局部最大值以外的所有梯度值 (应用梯度计算)。在裂纹检测中, 最后的输出是裂纹的坐标, 根据坐标可以画出框框定物体。由于一个裂纹可以生成众多的框, 所以需要去掉一部分多余的和概率小的框。非极大值抑制的功能就是去掉多余的边框。简而言之就是两步, 第一步是去掉 IOU 小的边框, 第二步是去掉与最大边框与其余边框重合多的边框。如图 18, 在对汽车进行物体检测时, 需要判断图片中是否有坐标并且给出框框住汽车。对于 Faster R-CNN 的输出结果, 框住汽车的边框有多个, 利用非极大值抑制可以去掉多余的框, 最后剩下 IOU 的最大的框。



图 18 非极大值抑制示意

(二) R-CNN 模型

区域卷积神经网络 R-CNN，其模型由三个部分部分组成。第一部分是选择性搜索算法提取图片的物体区域，这一步会搜索图片的不同位置，生成大量的不同尺寸的候选区域。第二部分是改进传统手工提取特征，利用卷积神经网络自动进行特征提取，第三部分在传统分类器的基础上（此处使用支持向量机）添加边框回归，如图 19 所示。注意到根据相似度产生的候选区域的尺寸是不同的，所以在对其用卷积神经网络进行分类之前必须要重新修改它的尺寸，使其适合卷积网络的输入尺寸。



图 19 传统物体检测与区域卷积神经网络对比

R-CNN 具体如图 20，第一部分是选择性搜索获取候选区域，每一张图片根据图片相似度生成的候选区域大概有 2000 个左右。模块二是将候选区域的图片进行预处理，使得候选区域尺寸符合卷积神经网络的特定尺寸。卷积神经网络之所以需要固定输入的大小是因为其卷积层和池化层的输出尺寸由输入尺寸决

定的。

第二部分即模块二是提取特征的卷积神经网络 AlexNet，本文采用的是 VGG 网络架构，其作用是取代传统提取图片特征的方法。具体来说，首先在 ImageNet 数据集上对该卷积神经网络进行预训练得到模型 M ，然后在模块一预处理的图片上对预训练的模型 M 进行微调得到模型 M' ，最后得到两个输出。

第三部分由模块三和模块四组成。模块三是以模块二中模型 M' 的第七层即全连接层为输入，利用支持向量机分类。模块四是利用模型 M' 的第五层特征图为输入，进行边框回归，对预测的边框做出相应的调整。

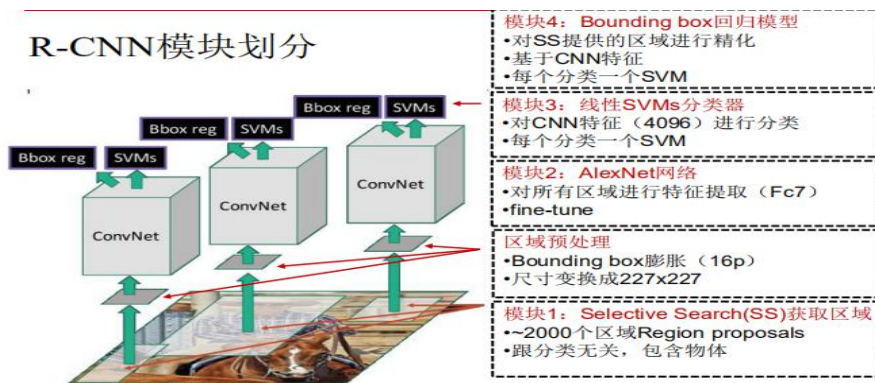


图 20 R-CNN 模块划分

R-CNN 相比传统的方法精确度得到提升，但是它的局限性质也十分明显：首先，每一张图片大约提取 2000 个左右的候选框，每个候选框要经过 $N+1$ (N 是分类类别数目) 次的支持向量机分类，所以需要进行大量的计算和存储大量特征；其次，需要经过多阶段训练。选择性搜索提取候选框特征，特征提取，支持向量机分类和边框回归是 3 个独立的步骤，所以训练过程十分复杂；

综上所述，RCNN 网络模型训练与测试所需时间长，不能进行实时检测，因此并不是最好的选择。

(三) Fast R-CNN 模型

基于 R-CNN 改进的 FastR-CNN，比起 R-CNN，能够有效提升训练和测试的速度，缩短时长，同时能够提升模型的训练精度，如图 21。

这一切都归功于三个改进：1、R-CNN 是先提大概 2000 个取候选区域然后利用神经网络提取特征，所以每张图片需要训练 2000 次左右，大量的时间花在提取特征上。Fast R-CNN 则是先通过卷积神经网络提取图片特征，然后再特征图上进行选择性搜索生成候选区域，节约了大量提取特征的时间。2、引入了 RoI

池化。3、引入了多任务损失函数，从而实现端对端的单阶段训练。

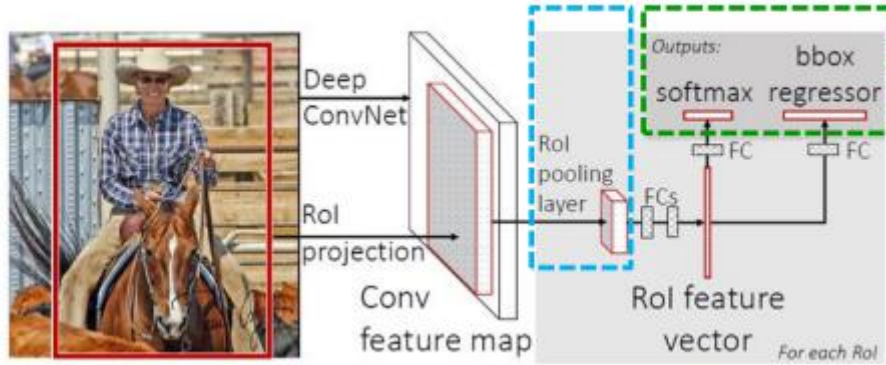


图 21 Fast R-CNN 结构图

其中，多任务算是函数的提出使得复杂的多阶段训练变成了端到端训练，节约了时间，整个网络也更趋于一个整体。多任务损失函数：

$$L(p, u, t^u, v) = L_{cls}(p, u) + \lambda[u \geq 1]L_{loc}(t^u, v) \quad (5.23)$$

其中对于分类部分： $L_{cls}(p, u)$ 表示分类损失且 $L_{cls}(p, u) = -\log pu$,

$p = (p_0, \dots, p_k)$ 表示每个 RoI 的概率分布， u 表示 Ground truth 的类别

对于边框回归部分：

$$L_{loc}(t^u, v) = \sum_{i \in \{x, y, w, h\}} smooth_{L_1}(t_i^u - v_i), \quad smooth_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases}, \quad (5.24)$$

(四) Faster R-CNN

5.4.1 RPN 网络

Faster R-CNN 相较于 Fast R-CNN 速度更快，精度更高。这主要归功于 Fast R-CNN 中选择性搜索 (Selective Search) 模块被 RPN 网络取代。RPN 网络能够解决选择性搜索速度慢，内存高，候选区域量多质差的瓶颈。RPN 网络通过进一步共享卷积层的计算，并引进 Attention 注意机制，引导 Fast R-CNN 关注区域。RPN 网络提取出来的候选区域大概是 300 个左右，而且拥有更高的查全率和查准率。

RPN 网络的损失函数：

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(P_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (5.25)$$

其中分类部分： L_{cls} 表示两类，一类是包含物体的另一类是不包含物体的。

回归部分：

$$L_{reg}(t_i, t_i^*) = R(t_i - t_i^*) \quad (5.26)$$

损失函数如Fast R-CNN 同样使用 smooth L_1 , x, x_a, x^* 分别表示对应的预测框, 锚框和 ground truth 框, 同理对于 y, w, h 也是如此。具体：

$$t_x = (x - x_a) / w_a, \quad t_y = (y - y_a) / h_a \quad (5.27)$$

$$t_w = \log(w / w_a), \quad t_h = \log(h / h_a) \quad (5.28)$$

$$t_x^* = (x^* - x_a) / w_a, \quad t_y^* = (y^* - y_a) / h_a \quad (5.29)$$

$$t_w^* = \log(w^* / w_a), \quad t_h^* = \log(h^* / h_a) \quad (5.30)$$

P_i^* 则表示指示函数, 包含物体时表示 1, 否则为 0。

Faster R-CNN 是 Fast R-CNN 的改进, 它主要是用 RPN 代替后者的选择性搜索。由前面 FPN 可知, 选择性搜索的生成的候选区域不高效精准, 数量大以致于内存耗费多, 整个训练测试过程时间长。RPN 生成的 候选区域则不同, 数量少且精准有效, 化繁为简。其余部分和 Fast R-CNN 相同, 训练过程如图 22。

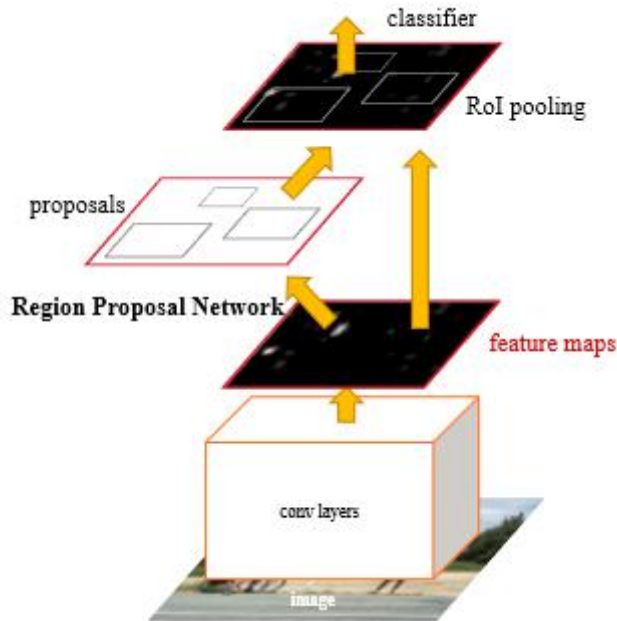


图 22 Fast R-CNN 的流程

六、优化算法

(一) Minibatch 梯度下降算法

机器学习尤其是深度学习，它是一个十分依赖经验的过程，这其中伴随着大量复杂的迭代过程，在一个项目中，需要训练诸多的模型，才能从中选出最适合，精度最高的模型，所以寻找合适的优化算法能够帮助我们快速的训练模型，大大提高团队的运行效率。

一般来说，使用 $\text{batch}=1$ 时，是随机梯度下降 (SGD)，即神经网络每次只训练一个样本就使用梯度下降算法，这样做的优点是训练速度快，但是 SGD 产生的噪音比较多，并不能保证每次的迭代都是朝着损失函数下降的方向，由于包含着随机性，所以 SGD 有时可以跳出局部最优，可能寻找更好地局部极小值。

$$\theta = \theta - \eta \nabla_{\theta} J(\theta; x^{(i)}; y^{(i)}) \quad (6.1)$$

其中， θ 表示参数向量， η 表示学习率， $x^{(i)}$ ， $y^{(i)}$ 中 i 表示样本编号， $\nabla_{\theta} J(\theta; x^{(i)}, y^{(i)})$ 表示损失函数关于参数向量的导数。

当 $\text{batch}=\text{N}$ ，即每次利用所有的样本进行梯度下降，这样做的好处是综合考虑了所有的样本信息，这样更全面。但是也正是如此，对整个数据集计算梯度，每次迭代的计算量非常大。当遇到大量的数据集时就非常棘手，通常内存溢出，而且不能实时增加新数据。互联网大数据时代，海量数据实时更新，如果训练出来的神经网络无法适应这种实时性，那么这种网络的也会由于输入样本数据分布的改变而下降，做不到动态更新也就不能长时间的应用该模型。

$$\theta = \theta - \eta \nabla_{\theta} J(\theta) \quad (6.2)$$

其中， θ 表示参数向量， η 表示学习率， $\nabla_{\theta} J(\theta)$ 表示损失函数关于参数向量的导数。

当 $\text{batch}=\text{n}$ (50-256)，即每次利用一定批量的样本进行梯度下降，这样可以降低参数更新的方差，收敛会更加稳定，两外一方面。在这种矩阵操作可以用 GPU 进行加速，减少训练的时间。但是 mini-batch 不能保证很好的收敛性，学习率如果选择太小，收敛速度就会慢，如果选择太大，损失函数就会在极小值附近震荡甚至是偏离。并且由于 minibatch 的学习率是固定的，梯度大的区域和梯度小的区域我们希望能分别对待，对梯度大的区域需要小的学习率，反之，对于梯度小的区域，需要给其大一点的学习率。否则，损失函数的收敛会非常慢。而

这一点 minibatch 不能做到，有待改进。

$$\theta = \theta - \eta \nabla_{\theta} J(\theta; x^{(i:i+n)}; y^{(i:i+n)}) \quad (6.3)$$

其中， θ 表示参数向量， η 表示学习率， $x^{(i:i+n)}$ ， $y^{(i:i+n)}$ 中 i 表示样本编号， n 表示 batch 数， $J(\theta; x^{(i:i+n)}; y^{(i:i+n)})$ 表示损失函数关于参数向量的导数。

(二) 动量梯度下降

SGD 在 ravines 的情况下极易被困住，ravines 指等高线图中，一个方向的梯度比另一个方向的梯度大，从等高线图上来看，即一个方向的曲线密集，一个方向的曲线比较稀疏。这种情况下，SGD 会发生震荡而迟迟不能接近极小值。并且在梯度平缓时，梯度更新慢，在梯度为零或鞍点时、在局部极小值时处停止更新，这样往往不能到达最小值处。由此，提出了动量梯度下降：

$$v_t = \gamma v_{t-1} + \eta \nabla_{\theta} J(\theta) \quad (6.4)$$

$$\theta = \theta - v_t \quad (6.5)$$

其中， v_t 表示第 t 次的动量， θ 表示参数向量， η 表示学习率， $\nabla_{\theta} J(\theta)$ 表示损失函数关于参数向量的导数。

Momentum 通过加入 γv_{t-1} ，加速 SGD 的更新，而且抑制震动。其物理含义是当一个小球从山上滚下来时（模拟梯度下降），会产生一个动量，在遇到梯度平缓的地方时，由于运动的惯性，会加快梯度的更新。当遇到鞍点或局部极小值时，会继续更新梯度，有机会寻找更加的局部极小值。但是缺点就是，学习率没有根据梯度的平缓与否发生变化，所以更新的效率不高。

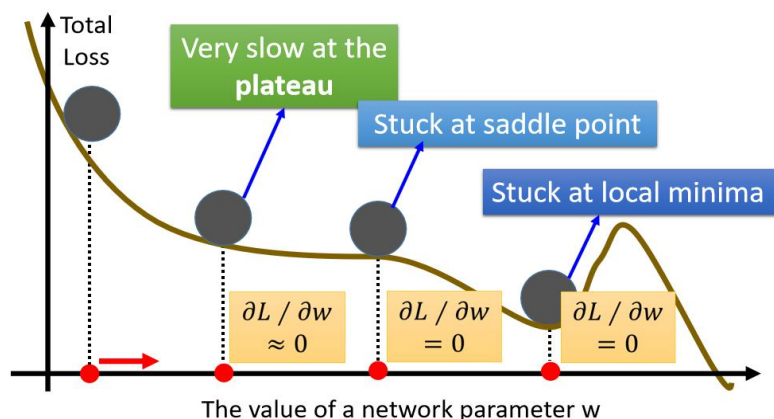


图 23 动量梯度下降形象示意图

(三) Adagrid 优化算法

为了对低频的参数做较大的更新，对高频的参数做小的更新，提出该种下降算法。所以在对稀疏的数据时它的表现良好，极大提高 SGD 的稳健型。如图 24，为了提升更新的效率，在 x 轴方向上的梯度变化平缓，需要一个大的学习率在 y 方向上的梯度比较陡峭，所以需要一个小学习率。

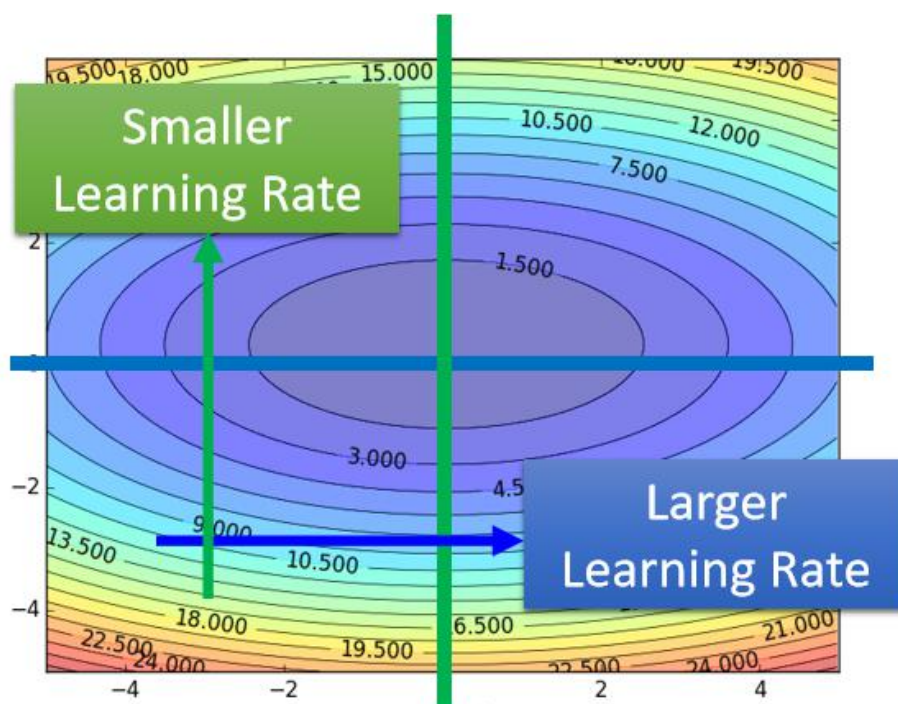


图 24 Adagrid 优化算法的适用函数

$$W^{t+1} \leftarrow W^t - \eta g^t / \sqrt{\sum_{i=0}^t (g^i)^2} \quad (6.6)$$

其中 w 表示参数向量， η 表示学习率， $g^{(i)}$ 表示第 i 次的梯度。

(四) Rmsprop 算法

Adagrad 只考虑了某个方向上的学习率，但是对于更一般的损失函数，等高线在某一个方向上的趋势更加复杂多变，考虑到图 25 的等高线，沿着 x 轴方向梯度先陡峭在平缓在陡峭，这种情况下 Adagrad 不再使用，由此提出了 Rmsprop，它是一种自适应学习率的方法。

$$s_t = \gamma s_{t-1} + (1-\gamma)g_t^2 \quad (6.7)$$

$$\theta_t = \theta_{t-1} - \eta g_t / \sqrt{s_t + \varepsilon} \quad (6.8)$$

其中 γ 表示衰减率， g_t^2 表示第 t 次梯度向量的元素对应相乘， ε 是一个极小的常数。

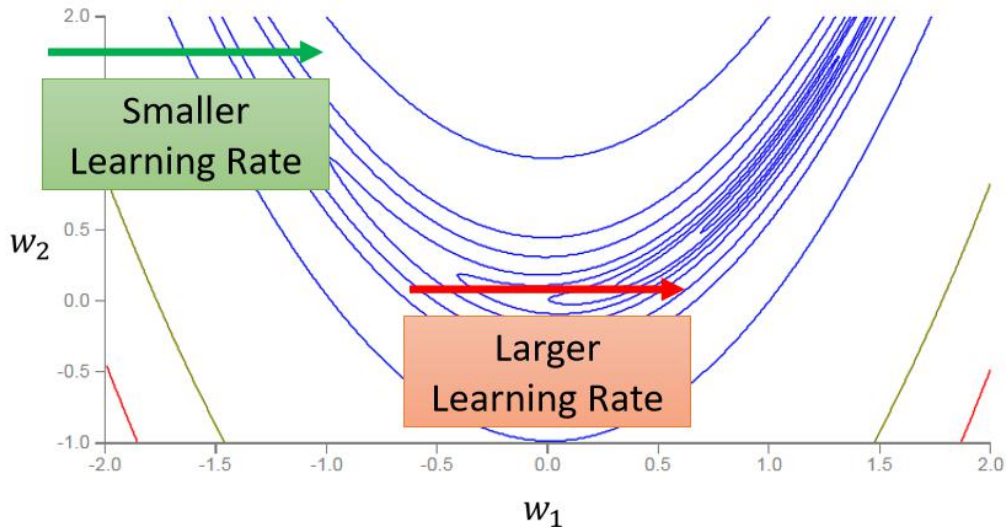


图 25 复杂的等高线图

(五) Adam 优化算法

动量梯度下降只考虑了添加动量，RMSprop 只改变了学习率，为了结合二者的优点，既考虑了学习率的衰减，又加入了动量，使得逃离局部最优成为可能，于是提出了 Adam 算法。更新方程如下：

$$g_t = \nabla_{\theta} f(\theta_{t-1}) \quad (6.9)$$

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t \quad (6.10)$$

$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2 \quad (6.11)$$

$$\hat{m}_t = m_t / (1 - \beta_1^t) \quad (6.12)$$

$$\hat{v}_t = v_t / (1 - \beta_2^t) \quad (6.13)$$

$$\theta_t = \theta_{t-1} - \alpha \cdot \hat{m}_t / (\sqrt{\hat{v}_t} + \varepsilon) \quad (6.14)$$

其中 g_t 表示迭代第 t 次时的梯度， $\beta_1, \beta_2 \in [0,1)$ 表示指数衰减率， α 表示学习率， g_t^2 表示第 t 次梯度向量的元素对应相乘， ε 是一个极小的常数。

七、建模及结果分析

(一) 模型构建

综合比较 R-CNN，Fast R-CNN，Faster R-CNN 模型的优劣以及优化算法。本文采取三种模型：

模型一，用图片未经预处理的 Faster R-CNN 模型。由于 VGG 网络架构的输出是固定尺寸的，因此需要图片的输入也是固定尺寸的。首先将裂纹图片尺寸进行重新调整至 $224 \times 224 \times 3$ ，分别表示图像的高，宽，颜色通道数。然后将图片输入到 VGG 卷积神经网络，最小化损失函数，采用随机梯度下降策略（SGD），经过多次迭代得到图像的特征图。接着用 RPN 网络对特征图进行候选区域的提取，生成 300 个候选框。紧接着结合特征图和候选框的位置，对候选框进行感兴趣区域池化（RoI 池化），虽然候选区域大小不一，但是经过 RoI 池化后可展开得到相同尺寸的特征向量，经过全连接层。一方面得到支持向量机的分类结果，一方面得到预测边框的初始坐标，再用边框回归调整，用非极大值抑制（NMS）去掉重复的多余的框。最终结合 SVM 分类和边框回归可以得到裂纹的具体位置和概率。

模型二，用同态滤波预处理的 Faster R-CNN 模型。首先对图片进行灰度化操作，然后直方图均衡，接着同态滤波处理，最后扩充图片通道，调整图片尺寸得到预处理后的图片。之后具体建模过程如模型一所述。

模型三，改进优化策略 Adam 算法的 Faster R-CNN。模型三结合了模型一和模型二，首先对图片进行模型二中的相同图片预处理，后续建模过程大体如模型一，只是在对损失函数进行梯度下降的时候采用了 Adan 优化策略。

（二）训练流程

训练一个 Faster R-CNN 模型，需要分别训练 RPN 和 Faster R-CNN，有四个步骤：

第一步训练 RPN，首先在 ImageNet 数据集上预训练模型参数，得到卷积层的初始化参数。

第二步是训练 Fast R-CNN 网络，首先在 ImageNet 上预训练模型参数，得到卷积层的初始化参数，然后把有第一步中 RPN 生成的候选区域输入其中。

第三步是调优 RPN 网络，第四步是调优 Fast R-CNN 网络。

（三）调参过程

训练深度卷积网络是一个十分精细的过程。训练过程中要查看训练精度和测试精度。当训练精度和测试精度都不高时，模型处于欠拟合的状态，由于采用 VGG 网络固定架构，通过加大迭代次数，提高精度。当模型的训练精度高，测试精度不高时，模型处于过拟合状态，通过提前停止的策略，在训练精度和测试精度都高的位置结束训练。

（四）结果分析

7.4.1 模型一结果分析

模型一是未经过图片预处理的 Faster R-CNN，如图 26，在经过迭代 10000 次迭代之后，最终模型的总损失值降到 0.87，模型的 AP 值为 0.856，具体损失值见附件。最终测试部分图片如图 27，所有图片见附件。可以观察到，对于比较明显的大的裂纹，可以检测到，而对于较小的裂纹，则不容易检测到。识别出部分的一伪裂纹。总体来看模型效果良好，小的裂纹不容易检测得到，关系也不大，英文对于工程建筑中墙体上的特别细小裂纹检测不到影响，只要较大的裂纹可精确识别即可。

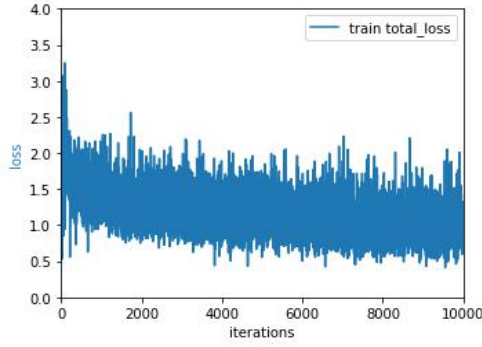


图 26 模型一训练的总损失

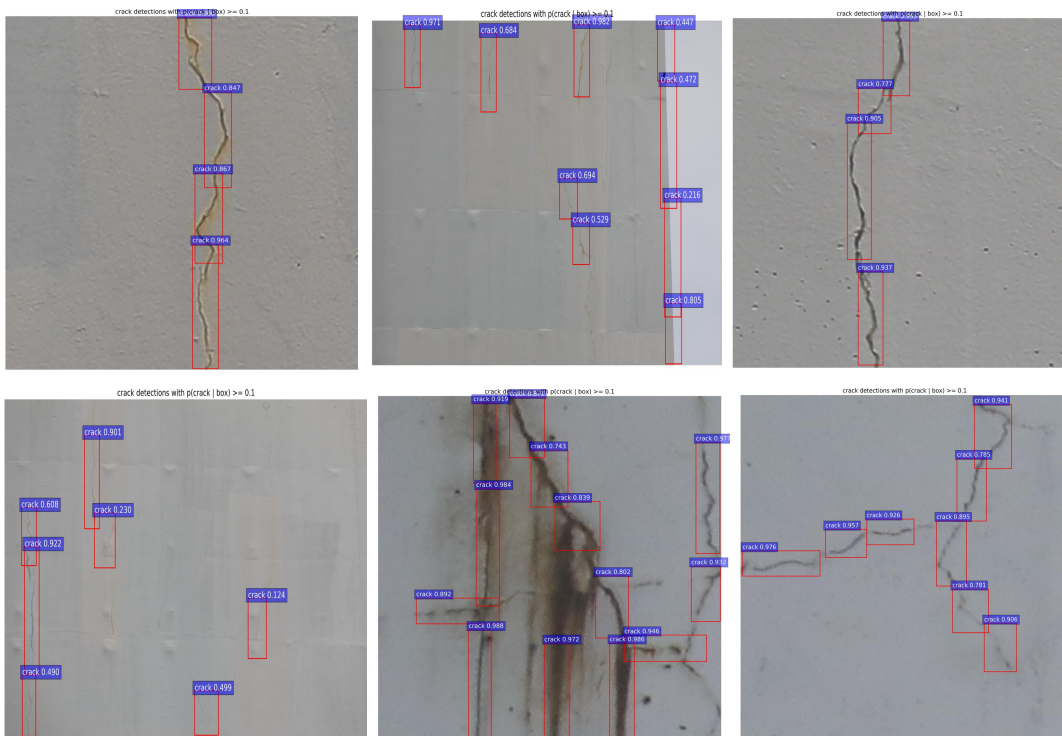


图 27 模型一未经图片预处理的 Faster R-CNN 测试结果

具体来看，如图 27，简称第一行第一列的图片为(1,1)，后同。(1,1)和(1,3)中的裂纹较大，较明显，所以被轻而易举的识别出来，给出的位置精确，概率高。对于(1,2)来说，图中的较小的裂纹也识别出来了，但是由于背景明暗突变处发生了梯度变化，这里被错误当成了裂纹，形成了误判。(2,1)识别了较小的裂纹，(2,3)识别了横向裂纹和颜色较淡的裂纹，模型给出的位置精确概率高。而对于(2,2)来说，图中伪裂纹较多，有两处伪裂纹被错误当成了裂纹。综上所述，模型一也比

较容易识别出较小的裂纹，但是该模型容易受到背景的干扰，错把背景变化和伪裂纹都当成裂纹。

7.4.2 模型二结果分析

模型二是经过同态滤波等预处理图片的Faster R-CNN,如图 28,在经过 10000 次迭代后，模型二总损失值降到 0.3，训练精度 AP 提升至 0.892，所有损失值见附件。最终的测试图片如图 29，完整图片见附件，可以观察到经过同态滤波处理的测试图片明显变亮，裂纹部分更加明显，可以明显从背景区分出来。对于伪裂纹，区分的效果比模型一更好。

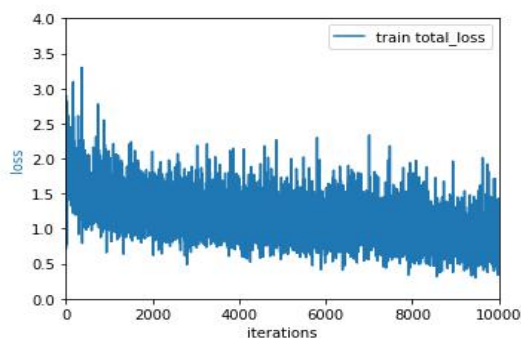


图 28 模型二训练的总损失

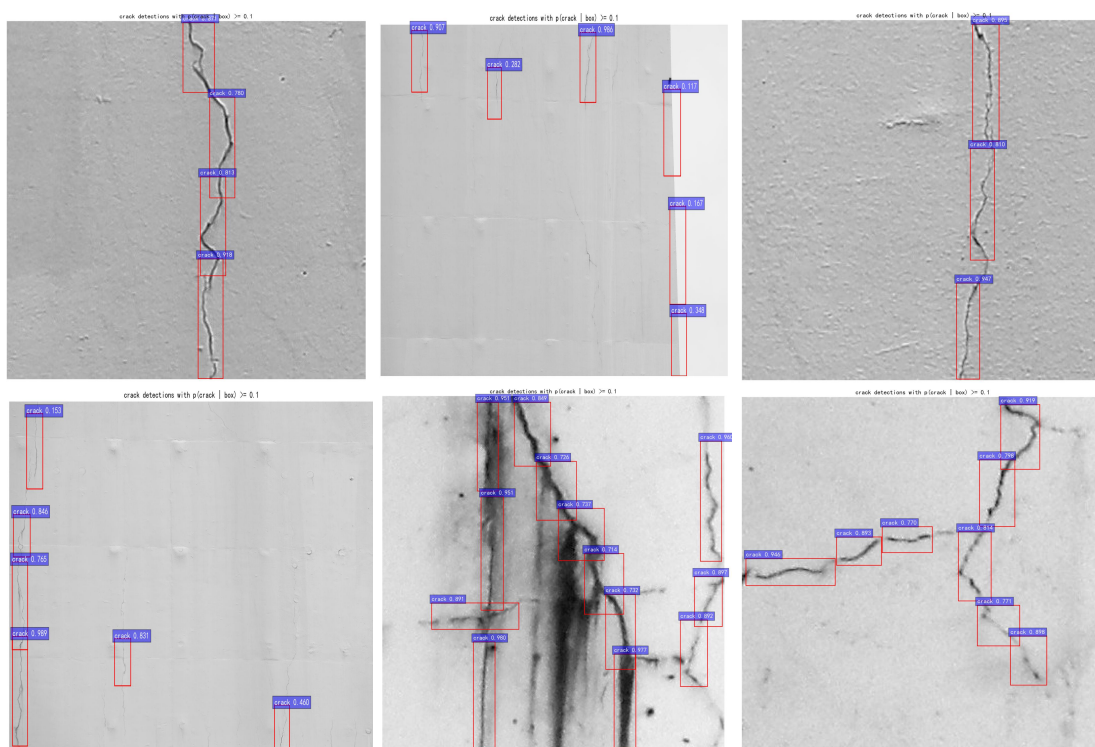


图 29 经过同态滤波处理图片的Faster R-CNN 的测试图片

具体来看，如图 29，图片明显比模型一明亮清晰，但是细小的裂纹更淡了。(1,1) 由于背景简单，裂纹较为明显，其中的裂纹可以被全部检测出来，给出的裂纹边框清晰，裂纹概率高，(1,3) 中的裂纹检测也是如此。(1,2) 中的有的裂纹比较细小，无法检测出来，而且在深色背景与浅色背景处的梯度变化使得误将此作为裂纹检验出来。对于 (2,2)，画面中裂纹较多而且包含了大量的伪裂纹，其中就大部分的裂纹都被检测出来，伪裂纹只有一处被当成了裂纹，并且还是在上半部分还框住了裂纹的情况下。图中的黑点较多，但是没被当成裂纹。而对于 (2,3)，大部分横向的和斜向的裂纹都被检测出来了。综上所述，模型二的检测效果好，不容易将伪裂纹当成裂纹，错误识别的情况少，漏掉了一些非常小的裂纹。裂纹给出的概率相对模型一小，认证观察这些确实是裂纹，因此只需要调整非极大值抑制的阈值就不会把这部分概率小的裂纹漏掉。

7.4.3 模型三结果分析

模型三是使用改进优化策略 Adam 算法的 Faster R-CNN，如图 30，在经过迭代 10000 次迭代之后，最终模型的总损失值降到 0.382，模型的 AP 值为 0.886，具体损失值见附件。最终测试部分图片如图 31，所有图片见附件。可以观察到，对于比较明显的大的裂纹，可以检测到，而对于较小的裂纹，则仍然不容易检测到。注意到可识别出部分的伪裂纹，较模型二差。

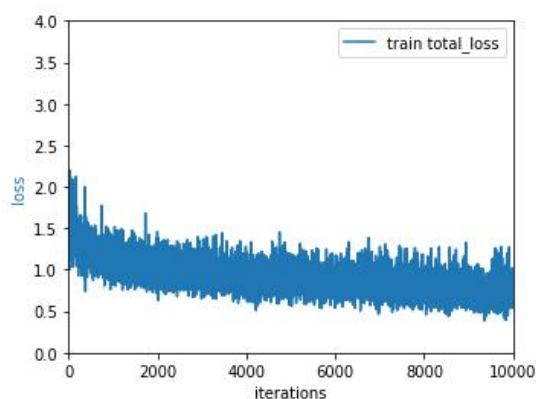


图 30 模型三的总损失

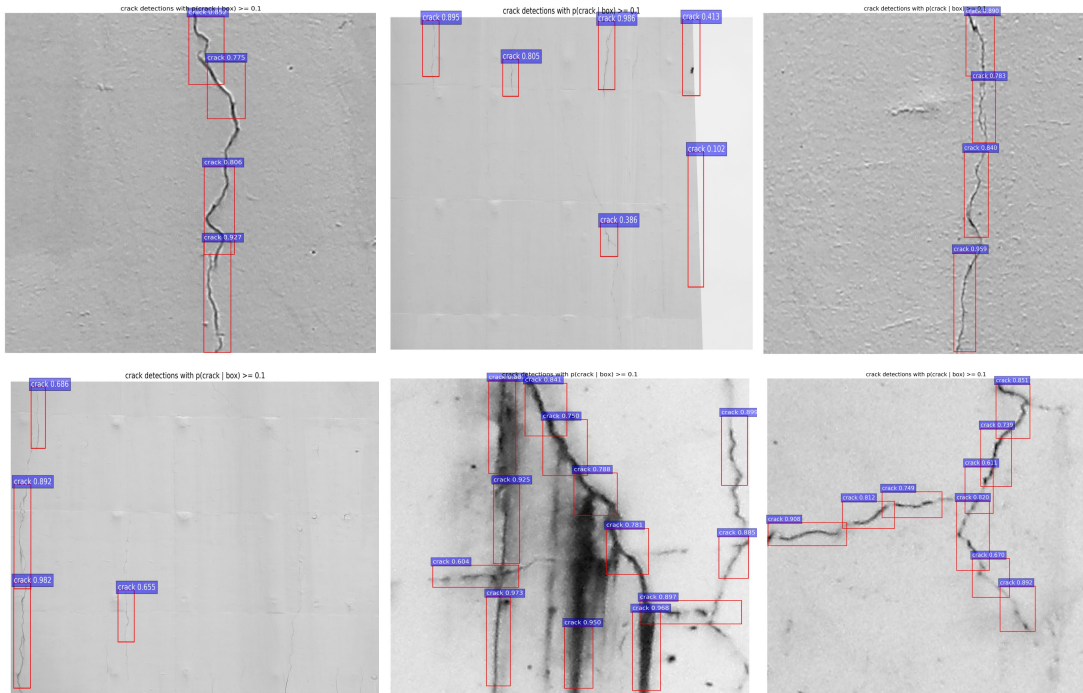


图 31 改进优化策略 Adam 算法的 Faster R-CNN 测试图片

具体来看，如图 31，图片明显比模型一明亮清晰，但是细小的裂纹更淡了。总体效果看起来和模型二给出的测试结果相同。(1,1)由于背景简单，裂纹较为明显，其中的裂纹可以被全部检测出来，给出的裂纹边框清晰，裂纹概率高，(1,3)中的裂纹检测也是如此。(1,2)中的有的裂纹比较细小，无法检测出来，而且在深色背景与浅色背景处的梯度变化使得误将此作为裂纹检验出来。对于(2,2)，画面中裂纹较多而且包含了大量的伪裂纹，有两处伪裂纹都被当成了裂纹。而对于(2,3)，大部分横向的和斜向的裂纹都被检测出来了。综上所述，模型二的检测效果好，不容易将伪裂纹当成裂纹，错误识别的情况少，漏掉了一些非常小的裂纹。裂纹给出的概率相对模型一小，认证观察这些确实是裂纹，因此只需要调整非极大值抑制的阈值就不会把这部分概率小的裂纹漏掉。

7.4.4 模型结果对比

综上所述，如图 32，模型二的损失值最小为 0.3，模型三次之，以模型一的损失值最大。究其原因是模型一的图片未经过预处理，图片中的噪声使然。至

于模型精度的比较，模型的精度提升不大，但是仍以模型二精度最高，为 0.892，这是因为通过同态滤波预处理，图片更加清晰，背景和前景区分更好。

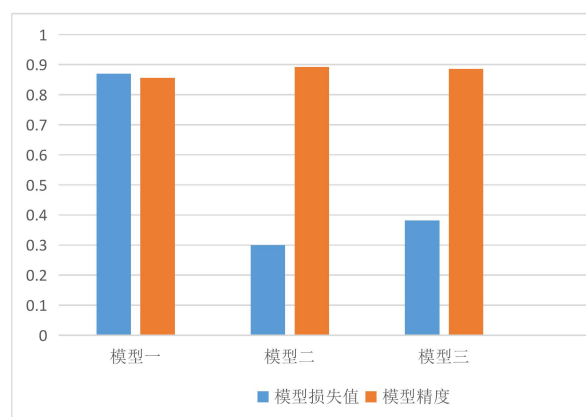


图 32 三种模型对比

如图 33，进一步通过对比测试图片，在伪裂纹识别方面，发现模型二的检测(1,2)，识别出的伪裂纹更少。在对细小裂纹的敏感程度方面，模型一的测试图片(2,1)较好，但模型二没有差太多。考虑到过于细小的裂纹对大型建筑物的墙面的影响不太，完全可以通过定期检查排除其风险。综上所述，相对来说，模型二的效果最好。比较三者，最终选择模型二。

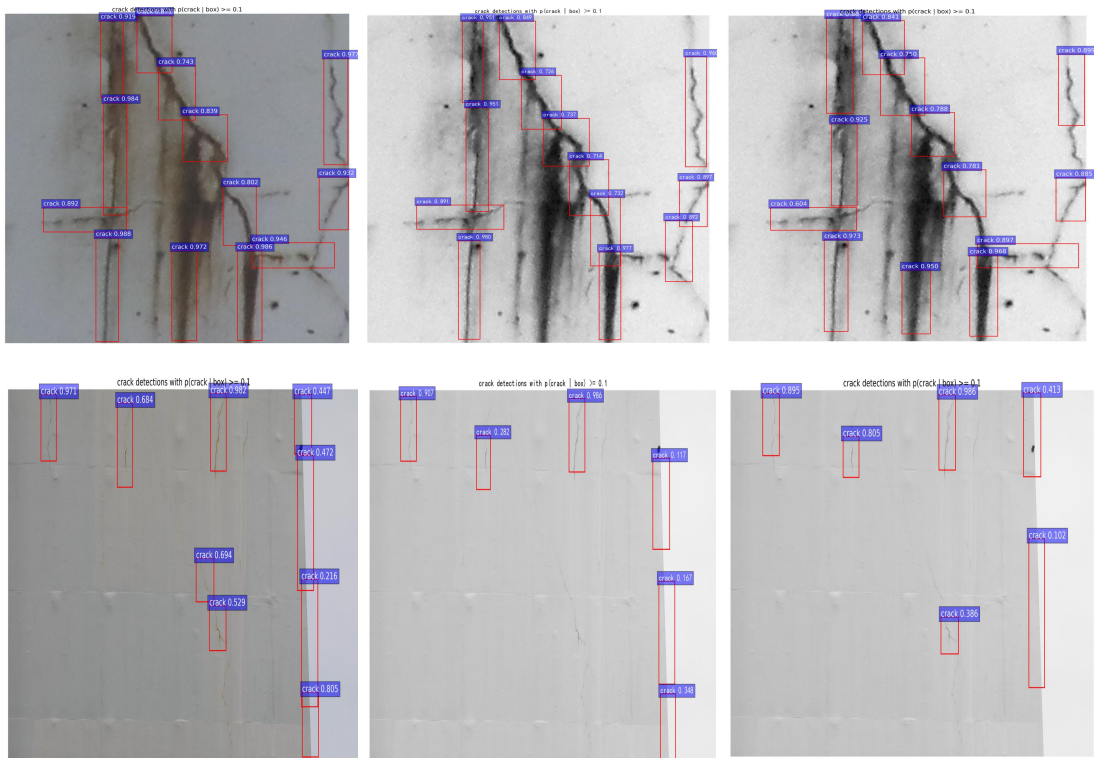


图 33 模型一，模型二，模型三测试图片对比

7.4.5 与现有方法对比

本文以卷积神经网络提取特征然后再分类的这一现有方法（简称卷积神经网络分类方法）为例，从模型的鲁棒性，模型的功能，模型的应用范围三个方面探讨 Faster R-CNN 裂纹识别效果与现有裂纹识别之间优缺点。

卷积神经网络分类方法分为如下步骤：首先，用原始图片制作数据集，将其分成背景图和裂纹图。然后采用灰度化预处理，再经过图像卷积，使用 Adam 优化下降策略，最后得到裂纹的二分类精度。

其中，数据集的背景图制作对结果会产生非常大的影响，当背景图为干净，无伪裂纹的墙体表面，识别精度高，平均为 92%，当数据集背景图包含伪裂纹时，识别精度低，平均为 40%，由此可见模型的鲁棒性低。

在功能方面，相比 Faster R-CNN，卷积神经网络分类方法功能单一，无法做到对裂纹的定位，仅仅能做到分类。

除此之外，由于传统方法对背景图的要求严格。然而现实环境中墙面伪裂纹的存在很常见，所以模型很容易将伪裂纹误判成裂纹，识别精度低，应用面较狭窄。

综上所述，对比上述模型的鲁棒性，模型的功能，模型的应用范围，得出使用 Faster R-CNN 的效果更好。

7.4.6 裂纹智能检测系统的建立

至此，经过比较，选择经过同态滤波等预处理图片的 Faster R-CNN 作为模型，建立智能裂纹检测系统。该系统只需要拍摄图片，然后经直方图均衡，同态滤波预处理，最后经过 Faster R-CNN 处理，即可检测出是否有裂纹以及裂纹的精准位置。另外，考虑到建筑物的实际分布会随着时间而改变，该系统中 Faster R-CNN 可以不用重头开始训练，只需要定期加入新的图片又可精准检测出裂纹。综上，其具体流程如图 34：

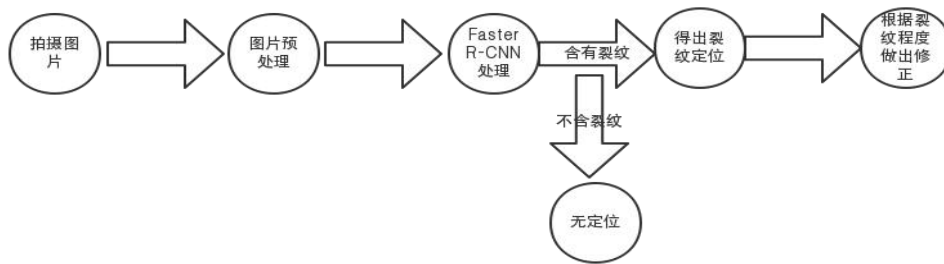


图 34 裂纹智能检测系统

八、模型的改进

综上所述，本文通过对比了三种 Faster R-CNN 模型的精度和效果，选择出模型三，即选择经过同态滤波处理图片的 Faster R-CNN 模型，其具有低损失值，高精度的特点。随后建立一套智能裂纹检测系统，流程为拍摄图片，在经过图片预处理，Faster R-CNN 处理后，随后即可得到裂纹的精确位置。但是该系统还存在如下缺点：

其一是该系统并未实现完全的自动化，需要多次选择图片预处理的方法，和神经网络的调参。

其二是随着时间的改变，裂纹的分布可能发生变化，即用作训练的图片分布可能和测试的图片分布不同，那么对裂纹的检测精度就会下降。这就需要经常将裂纹的图片输入系统微调。

其三是最终选择的更快的区域卷积神经网络虽然速度较之前的模型更快，但是训练的时间仍然很长，如果没有 GPU 加速运算，训练时间会长达数天。

其四是对于微小的裂纹识别度不高，如果是那种安全性能要求更高的建筑，或者不是本文研究的墙体表面，而是金属建筑表面的裂纹，那么就需要更进一步改善模型对小物体的识别。可以考虑使用 YOLO V3 模型。

参考文献

引文文献

- [1] Fujita Y , Mitani Y , Hamamoto Y. A method for crack detection on a concrete structure[C]. Pattern Recognition, 2006, 3:901-904.
- [2] Jabakhanji D , Shams M. A CMOS imager with on-pixel gray-scale erosion[C]. Symposium on Circuits & Systems, 2007:61-64
- [3] Ooi C H , Kong N S P , Ibrahim H. Bi-histogram equalization with a plateau limit for digital image enhancement[J]. IEEE Transactions on Consumer Electronics, 2010, 55(4):2072-2080
- [4] Hu Y, Zhao C. A local binary pattern based methods for pavement crack detection[J]. Journal of Pattern Recognition Research, 2010, 5(1): 140-147
- [5] Zou Q , Cao Y , Li Q , et al. CrackTree: Automatic crack detection from pavement images[J]. Pattern Recognition Letters, 2012, 33(3):227-238.
- [6] Oliveira H, Correia P L. Crackit-an image processing toolbox for crack detection and characterization[C]. Proceedings of IEEE International Conference on Image Processing, 2014:798-802.
- [7] Roth H R , Lu L , Liu J , et al. Improving Computer-aided Detection using Convolutional Neural Networks and Random View Aggregation[J]. IEEE Transactions on Medical Imaging, 2015, 35(5):1170-1181.
- [8] Dan C Ciresan, Giusti A , Gambardella L M , et al. Deep neural networks segment neuronal membranes in electron microscopy images[J]. Advances in Neural Information Processing Systems, 2012:2843-2851.

- [9] Dan C Ciresan, Giusti A , Gambardella L M , et al. Mitosis detection in breast cancer histology images with deep neural networks[J]. in Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2013:411-418.
- [10] Krizhevsky A , Sutskever I , Hinton G . Hinton. Imagenet classification with deep convolutional neural networks[J]. Advances in Neural Information Processing Systems, 2012:1097 - 1105.
- [11] Zhang Y , Sohn K , Villegas R , et al. Improving object detection with deep convolutional networks via Bayesian optimization and structured prediction[J]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2015:249 - 258.
- [12] Kivinen J , Williams C , Heess N . Visual boundary prediction: A deep neural prediction network and quality dissection[J]. Proceedings of International Conference on Artificial Intelligence and Statistics, 2014: 512 - 521.
- [13] Lecun Y , Bengio Y , Hinton G . Deep learning[J]. Nature, 2015, 521(7553):436-444.
- [14] Zhang L , Yang F , Zhang D , et al. Road crack detection using deep convolutional neural network[C]. Conference:IEEE International Conference on Image Processing (ICIP 2016)
- [15]常海涛, 苟军年, 李晓梅. Faster R-CNN 在工业 CT 图像缺陷检测中的应用[J]. 中国图象图形学报, 2018, 267(7):129-139.
- [16] Hinton G E . Rectified linear units improve restricted Boltzmann machines[C]. Proceedings of the 27th International Conference on Machine Learning (ICML-10), 2010.
- [17]Abdel-Qader I , Pashaie-Rad S , Abudayyeh O , et al. PCA-Based algorithm for unsupervised bridge crack detection[J]. Advances in Engineering Software, 2006, 37(12):771-778.
- [18]Sondes T, Hassense S, Zouhair M, et al. RGB image de-noising using new low-pass filter with Variable Gaussian core real time optimized by neural networks[C].

[19]Werbos P J .Werbos P J . Applications of advances in nonlinear sensitivity analysis[M]. System Modeling and Optimization. Springer Berlin Heidelberg, 1970.

附录

由于论文中用到的所有程序和三个模型的测试结果太大，见附件。
由于本文用到的训练图片数据太大，部分见附件。

致谢

本次大赛中，梁小林副教授对该论文的指导尽心竭力，对本论文提出了很多有建设性的意见。在实际应用中，谢永钦教授也提供了许多宝贵的经验。同时非常感谢湖南拓达结构监测技术有限公司提供的图片数据！在此，对他们致以最诚挚的感谢！