



汕 頭 大 學

SHANTOU UNIVERSITY

硕 士 学 位 论 文

题 目 基于深度学习的电厂仪表识别

英文题目 Instrument Identification in Power Plant Based on
Deep Learning

姓 名 安康 学 号 151709039

所在学院 工学院 导师姓名 崔岩

专 业 电子与通信工程

入学日期 2017.9 答辩日期

学位论文原创性声明

本论文是我个人在导师指导下进行的工作研究及取得的研究成果。论文中除了特别加以标注和致谢的地方外，不包含其他人或其它机构已经发表或撰写过的研究成果。对本文的研究做出贡献的个人和集体，均已在论文中以明确方式标明。本人完全意识到本声明的法律责任由本人承担。

作者签名：_____ 日期：_____年____月____日

学位论文使用授权声明

本人授权汕头大学保存本学位论文的电子和纸质文档，允许论文被查阅和借阅；学校可将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或其它复制手段保存和汇编论文；学校可以向国家有关部门或机构送交论文并授权其保存、借阅或上网公布本学位论文的全部或部分内容。对于保密的论文，按照保密的有关规定和程序处理。

作者签名：_____ 导师签名：_____

日期：_____年____月____日 日期：_____年____月____日

摘要

在发电厂中，各种仪表显示的信息反映了不同设备的工作状态，因此，需要工作人员定时的执行巡检任务，以保证设备正常运行。然而，在实际场景中仍然需要大量的人工进行各类仪表所监测数据的采集和录入。这种数据采集的方式费时费力，不但需要工作人员不间断的去记录数据，而且由于工作人员的主观差异造成一定的数据误差。同时，往往大部分仪表所存在装置的工作环境十分恶劣，导致人工无法通过人工操作方式方便地对其进行数据信息传输和状态监控。因此，如何解决这些问题，并对各类仪表进行高效的、精确的自动数据录入变得尤为重要。

在发电厂中最常用的仪表包括字符仪表、指针式仪表、油压表。近年来，数据采集的设备和方法日益增多，计算机视觉技术亦渐趋成熟，本文选择图像处理和深度学习的相关技术解决了发电厂的字符仪表、指针式仪表、油压表的自动识别问题，具体包括仪表定位、仪表分类、仪表示数识别。其中，仪表盘定位采用了尝试 Faster RCNN、SSD、YOLO 三种算法，经过实验对比，以平均准确率均值(mAP)作为评价指标，Faster RNN 算法达到了 0.96 以上，故选其作为目标检测算法；仪表类别判断尝试 LeNet5、AlexNet、VGG、InceptionV4 四种分类算法，经过实验对比，以准确率和召回率为主要评价指标，InceptionV4 算法的两项指标都达到了 98%以上，故选其作为类别判断的算法。仪表示数识别分别借助两种途径：字符仪表示数识别通过使用 CRNN 算法进行，经试验对比，该方法的准确性为 90%；而指针式仪表和油压表示数识别则打破当前主流的方法。由于传统图像处理方法具有一定的局限性，因此，本文提出了一种基于卷积神经网络的指针式仪表和油压表示数识别方法，该方法主要思想是使用卷积神经网络做回归任务，以仪表图片和其示数作为输入进行训练，使用均方误差函数作为损失函数，经过实验证明，该方法准确性达到了 99%。基于卷积神经网络的指针式仪表和油压表示数识别方法降低了传统图像处理对仪表图片质量的依赖性，增加了示数识别的准确性且具有一定的泛化性。最后，经实验证明，本文选择和提出的方法切实有效，实现了在发电厂复杂环境中仪表的定位、类别判断、示数读取的自动化。

关键词：指针式仪表；深度学习；图像处理；目标检测；图像分类

Abstract

In terms of power plants, the information displayed by various instruments reflects the working status of different equipments. Therefore, it is necessary for staff to regularly perform inspection tasks to ensure the normal operation of equipments. However, a lot of manual labor is still needed to collect and input the data monitored by all kinds of instruments in the actual application scenarios. This collection method is time-consuming and laborious, which not only requires the staff to record data continuously, but also causes certain data errors due to subjective differences of the personnel and efficiency of data collection. Moreover, the working environment of most instruments is very tough and challenging, which makes it very difficult to perform data transmission and status monitoring by manual operation. Therefore, to find a way to solve these problems and carry out efficient and accurate automatic data collection for various instruments becomes particularly important.

Generally speaking, the most commonly used instruments in power plants include character instruments, pointer instruments and oil pressure gauges. In recent years, data collection equipments and methods have increased day by day, and computer vision technology has gradually matured. This article chose to use image processing and deep learning related technologies to solve the automatic identification problems of the above-mentioned instruments in power plants, including instrument positioning, instrument classification and instrument representation number recognition.

Firstly, the three algorithms - Faster RCNN, SSD and YOLO – have been tested for the instrument panel positioning. After experimental comparison, taking the average accuracy precision(mAP) as the evaluation index, the Faster RNN algorithm has reached more than 0.96, so it is selected as the target detection algorithm. Secondly, the four classification algorithms - LeNet5, AlexNet, VGG and InceptionV4 have been tested for instrument category judgment. After experimental comparison, The accuracy and recall rate are the main evaluation indicators. The two indicators of the InceptionV4 algorithm have reached more than 98%, so they are selected as the algorithm for category judgment. Finally, the instrument number recognition involves two methods: the one is CRNN algorithm for character instrument number recognition, after experimental comparison, the accuracy of this method is 90% ; as for the pointer instrument and the oil pressure number

recognition, this article thinks out of the mainstream method based on traditional image processing, and proposes a sign recognition method based on convolutional neural network, the main idea of this method is to use convolutional neural network for regression tasks. The instrument picture and its indication are used as input for training, and the mean square error function is used as the loss function, and experiments show that the accuracy of the method reaches 99%. The method which reduces the dependence of traditional image processing methods on the image quality of the meter, increases the accuracy of sign recognition and has a universality for potential widespread application scenarios.

Experiments have proven that the methods selected and proposed in this thesis are effective which realizes the automation of instrument positioning, category judgment and indicator reading under the complex environment of power plants.

Keywords: pointer instrument, deep learning, image processing, object detection, image classification

目 录

摘要.....	I
Abstract.....	II
目录.....	IV
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
第 2 章 深度学习模型理论.....	8
2.1 感知器.....	8
2.2 全连接神经网络.....	9
2.3 卷积神经网络.....	10
2.4 本章小结.....	12
第 3 章 电厂仪表定位.....	13
3.1 Faster R-CNN 基本原理简介.....	13
3.2 YOLO 基本原理简介.....	17
3.3 SSD 基本原理简介.....	18
3.4 仪表检测算法实验分析.....	19
3.5 本章小结.....	21
第 4 章 电厂仪表类别识别.....	23
4.1 LeNet-5 基本原理简介.....	23
4.2 AlexNet 基本原理简介.....	24
4.3 VGG-16 基本原理简介.....	25
4.4 GoogLeNet 基本原理简介.....	27
4.5 仪表类别识别实验分析.....	28
4.6 本章小结.....	30
第 5 章 电厂仪表示数识别.....	31
5.1 字符仪表示数识别.....	31
5.2 基于传统方法的压力表示数识别.....	34
5.3 基于卷积神经网络的压力表示数识别.....	36

5.4 本章小结.....	38
第6章 工作总结和未来展望.....	39
6.1 工作总结.....	39
6.2 未来展望.....	39
参考文献.....	40
致谢.....	44
攻读硕士学位期间主要的工作成果.....	45

第 1 章 绪论

1.1 研究背景及意义

随着社会的进步和发展以及工业 4.0 的到来，全国工业行业发生了翻天覆地的变化，作为国家命脉的电力行业更是如此。由于社会对电力需求需求的不断增长，发电厂和电网的规模也越来越大，对电力系统的安全生产和稳定输出的要求也越来越高，因此，建立一个能既保证高质量电力输出又能保持安全和稳定生产的电力系统极其重要^[1]。

对发电厂来说，对其系统内各种设备的运行状态保持长期且定时的巡检是必不可少的任务，是安全和稳定生产的前提。其中，电厂中的各类仪表是反应各种设备运行状态的最直观的工具，对仪表进行巡检是整个巡检任务中最重要的一环。发电厂巡检任务通常是由电厂员工在一天 24 小时内交替、定时的执行，这种靠人工执行巡检^[2]的方式虽然在一定程度上能通过对仪表的巡检，以保障各种设备在某些时间段内处于正常运行状态，但是这种方法并不能做到实时监控。为了达到实时的效果，部分电厂使用摄像头对仪表进行监控^[3]，其本质还是依靠人工实时的在操作盘面前进行监测。依靠人力定期去现场巡检的这种方法，常常会收到现场复杂环境的影响，导致巡检效率极低甚至会有安全隐患；依靠人力在操作盘面前进行检测这种方式虽然能避免危险，但是由于人工的主观性和疲劳性不可避免，这种方式的检测效率依然极低。

随着人工智能^[4]和机器人产业的发展，巡检机器人^[5]的出现可以代替或辅助人工解决这一问题。巡检机器人搭载一套对仪表巡检的智能系统，其最大的特点是不再需要人工长期定时的操作，其系统通过巡检机器人的摄像头采集仪表图像，使用图像处理和人工智能技术可以对各种仪表的图片进行自动的、多功能的识别。这种方法既避免了传统巡检方式效率低、耗费人工的缺点，又完成了实时监控的需求，保证了发电系统安全和稳定的生产。图 1-1 为电厂巡检机器人在执行仪表巡检任务。



图 1-1 电厂巡检机器人

1.2 国内外研究现状

早在 1950 年到八十年代中期，国内外对电力设备状态的监测的方法主要是事后维护^[6]、定期维修^[7]、状态监测^[8]。事后维护，顾名思义，是等设备发生了故障之后才去找到导致故障的原因并进行检修，这种方式不但不能保证整个电力系统稳定和正常的运行，还会缩短设备的寿命，进而增加了电厂的安全隐患；定期维修是指在电力系统运行到一定时间后，进行停电操作来排查有故障的设备，这种方法并不能保证维修间隔内设备不发生故障；状态检测这种方法相对来说比较有效，它主要是结合计算机和传感器对所需监测的设备进行远程监测。

随着机器视觉和图像处理^[9]等技术的发展，将这些技术应用到电力设备的自动检测已经成为了新的趋势。人工智能的火热发展使得很多相关技术人员将其应用到了电厂设备监测之上，尽管目前为止还没有效果稳定的产品应用到实际场景中，但是使用深度学习^[10]和图像处理技术应用到电力系统设备监测已经证明了其有效性和可行性。

现阶段的发电厂中用来反映设备状态的仪表主要有指针式仪表、字符仪表和油压表，油压表相较于前两种仪表使用较少，识别方法简单，故主要介绍字符仪表和指针式仪表为主。目前应用机器视觉技术进行指针仪表读数识别的主要分为两个步骤：仪表定位并获取仪表盘和仪表指针的识别及示数读取。仪表盘定位属于目标检测范畴，现阶段已有很多成熟的算法可以使用。常用的算法有 SIFT^[11]、SURF^[12]、Haar^[13]特征+Adaboost^[14]+Cascade^[15]、HOG^[16]特征（方向梯度直方图）+SVM^[17]（支持向量机）等。SIFT 算法对旋转、尺度缩放、亮度变化会保持不变性，对视角变化、仿射变换、噪声也有一定程度的稳定性，还具有独特性、多样性、高速性等优点。但是，SIFT 算法因为要不断地进行下采样和差值等操作，所以导致其实时性不高。同时，对边缘平滑的目标无法准确提取特征，例如圆。SURF 算法虽然是 SIFT 的改进版，但是依然需要人工选取特征，达不到自动化的要求。Haar 特征+Adaboost+Cascade 算法具有较高的检测速度，但是其选用的特征其实是有些弱的，总体效果并不好。HOG 特征（方向梯度直方图）+SVM（支持向量机）算法虽然对目标的检测效果有显著的提升，但是所需时间过长，并不能满足实时的要求。

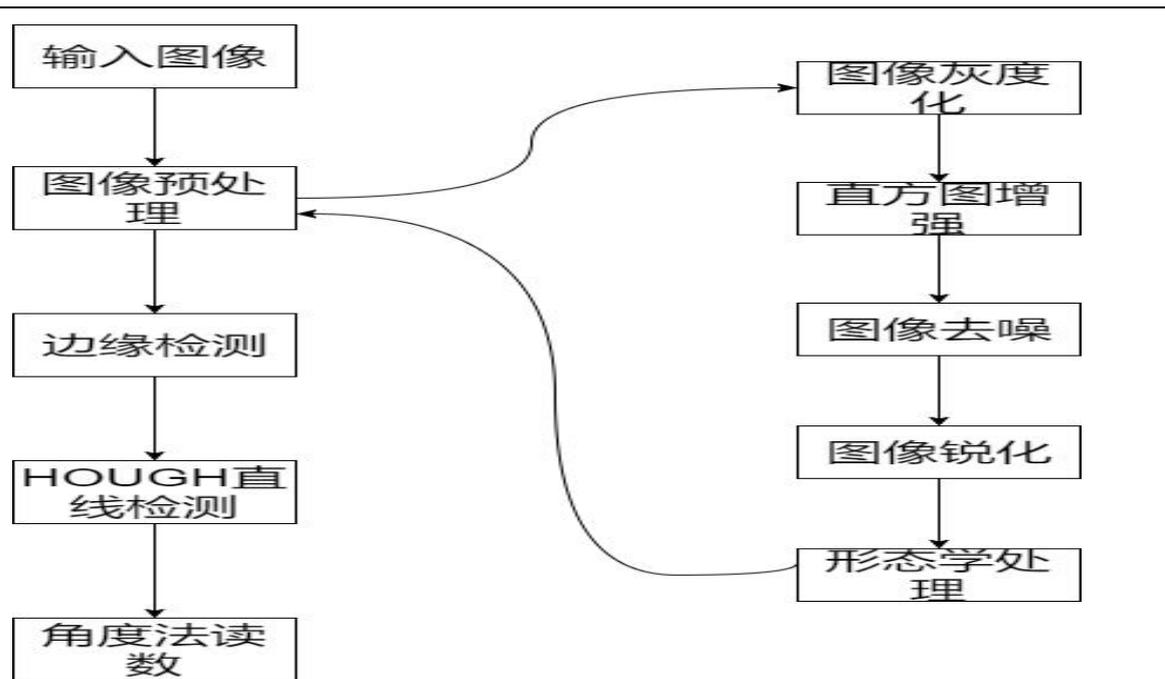


图 1-2 指针式仪表示数识别流程图

在指针式仪表读数识别的任务中，仪表指针的识别及示数读取这部分才是关键。这一部分目前采用的方法大致相同，可以分为五个步骤，如图 1-2，分别为输入图像、图像预处理、边缘检测^[18]、HOUGH 直线检测^[19]、角度法读数。其中，图像预处理又分为 RGB 图像的灰度化^[20]、直方图均衡^[21]、图像去噪^[22]、图像锐化、形态学^[23]处理。根据提取出的指针所在的直线与零度角的角度差值、仪表量程的比例关系来计算指针式仪表的读数。具体计算方法如下所示：

$$N = \frac{\theta - \theta_0}{\theta_{\max} - \theta_0} (N_{\max} - N_0) \quad (1-1)$$

公式 (1-1) 中， N 表示由角度法所求的仪表示数， θ 表示仪表指针所在直线与水平线之间的角度， θ_0 表示零刻度线与水平线之间的角度， θ_{\max} 表示最大量程所在直线与水平线之间的角度， N_{\max} 和 N_0 为仪表的最大、最小量程。

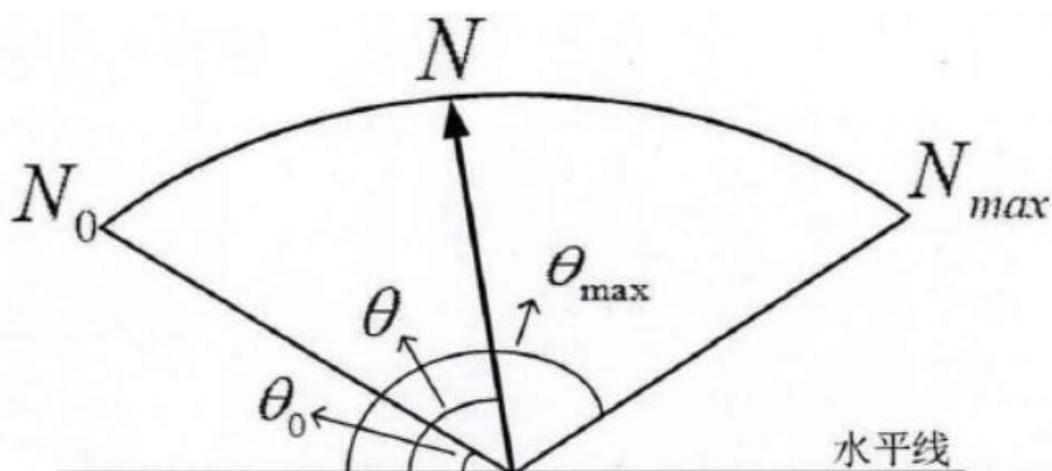


图 1-3 角度法原理示意图

使用角度法读取仪表示数的优点在于，此算法实现相对简单，且不需要获得每个刻度的具体位置，只需要指针所在直线和水平线之间的角度、最大和最小量程即可，但是该算法的缺点就是依赖图片质量，一旦图片被噪声影响严重，其算法精度也随之发生变化。

对于字符仪表识别来说，除去图片预处理主要分为两个步骤：字符分割、字符识别。字符分割是用图像处理和深度学习的技术将图片中的字符一一分割出来，传统图像处理分割的方法主要有聚类法、区域生长法、垂直投影法、固定间隔切分法等。聚类法的思想是计算像素和聚类中心的距离，对比像素和所有类的距离，根据最小距离将像素判定为某一类，从而达到了分割整个图像的目的，但是这种方法依赖于提前设置的超参数，因此并不能保证是最佳的方法且达不到自动化。区域生长法的基本原理是将图片中具有相似性质的像素进行集合进而得到区域，具体为随机选择一个没有归类的区域中的像素作为该区域的生长点，让生长点对周围的跟其有相似性的像素进行合并。垂直投影法通过字符投影的直方图的峰值和峰谷判断出字符分割的位置，从而实现单字分割^[24]。边缘检测法是利用图像连通域的原理对数字进行定位和切分。尽管这些字符分割方法有一定的效果，但还是避免不了传统图像处理方法容易被噪声和光照等因素影响的缺点。

随着深度学习的发展，图像分割的方法也有了很大的改进。基于深度学习的图像分割方法的核心思想为：采用卷积神经网络对图片中的像素进行特征提取，对每一个像素点进行分类从而实现图像分割的目的。具体为：使用 labelme 等数据标注工具，提前标注感兴趣区域，将这些数据送入到卷积神经网络分割网络中进行学习和训练最终得到分割结果。当前主流的深度学习分割模型有 FCN^[25]、DeepLab^[26]、PSPNet^[27]、Deepface^[28]、Onepose^[29]、U-Net^[30]、Mask R-CNN^[31]等。虽然基于深度学习的图像分

割方法效果远远强于传统图像处理方法,但是训练分割模型需要花费很多时间进行图像标注和训练,因此,如何减少花费的时间是一个很重要的问题。

字符分割之后,就是字符识别部分。关于字符识别问题,切割后的字符识别的目前主流的方法有传统图像处理和机器学习两种,基于传统图像处理的方法常用模板匹配法、基于机器学习的方法用分类方法如 SVM、逻辑回归、KNN^[32]等,基于深度学习中的 LeNet5 算法等。尽管这些方法在字符识别领域效果较好,但还是依赖于字符分割的结果,一旦字符分割不清晰或错误,就会导致后面的识别部分失准。

1.2.1 国内研究现状

郭子海^[33]等人是我国将数字图像处理技术应用到指针式仪表读数自动识别中的早期研究者之一,提出了利用数字图像处理技术进行指针式仪表读数的整个流程。其主要思想是通过相机采集到指针式仪表图片,并将其预处理为二值化图像,再用霍夫变换法确定指针位置并计算其与零刻度的夹角,根据夹角计算出读数。

石睿^[34]等人提出了一种基于支持向量机(SVM)算法的指针式仪表自动读数方法,包括仪表图像预处理、SVM 识别刻度并读数两个部分。预处理部分:基于二值化方法,加入了标准差分析;分割得到刻度和指针后,采用极坐标变换方法对刻度区域进行极坐标变换,这样可以得到刻度和指针的相对位置。识别与读数部分:得到刻度与指针相对位置后,通过 SVM 算法识别数字部分,解决了自动化程度较低的问题。

邢浩强^[35]等人提出了一种基于深度学习卷积神经网络的指针式仪表自动检测和识别方法,包括目标定位、图像消畸变和读数识别 3 个部分。该方法有点在于通过深度学习目标检测算法 SSD 可快速的定位仪表,获取仪表在图像中具体坐标后,通过裁剪以及图像消畸变方法最终得到高质量的仪表图像,解决了多角度采集图像引起的图像变形的问题。获得仪表目标高质量图像后,通过预处理、表盘和指针检测、指针检测与刻度对应等步骤识别仪表最终读数。该算法具有良好的迁移性、稳定性和鲁棒性。

周杨浩^[36]等人针对自然场景中指针式仪表的自动读数问题,提出了一种基于机器学习和图像处理算法的方案。使用全卷积网络(Full Convolutional Networks, FCN)网络对输入图像进行语义分割,完成对仪表位置的检测,并提取出仪表所在的部分。利用直方图均衡化、中值滤波和双边滤波减小光照和阴影对指针识别的干扰,并利用仿射变换矫正拍摄时的倾斜,再结合改进的霍夫变换识别仪表中指针的位置,从而计算指针角度获取读数。总体来说,此研究方法有良好的鲁棒性、抗噪性和准确性。

万吉林^[37]等人提出一种基于深度学习目标定位和语义分割的方法。首先使用 Faster R-CNN 算法定位仪表图像中表盘区域,然后使用改进的语义分割 U-Net 算法精准提取了指针和刻度。此方法创新点在于,根据指针式仪表特点,重新构造了 U-Net

的 Dice 损失函数，有效代替了传统霍夫变换提取指针算法。

1.2.2 国外研究现状

Carlos Behaine^[38]等人提出了一种新型的算法进行指针式仪表识别。该算法的新型之处在于使用 Active Shape Model 模型 (ASM) 作为一种灵活的模型来表示指针式仪表及指针在其整个范围内的可变形形状。ASM 是一种自适应形状匹配技术，已用于表示图像中的可变形对象特征。ASM 方法基于其标记点来表示可变形对象的形状，这些标记点随对象变形而改变其位置，可以使用这些界标点自动识别模拟显示图像中的感兴趣区域 (ROI)，然后根据识别出的仪表形状读取仪表。

Ricardo Ocampo-Vega^[39]等人针对用智能手机采集到的指针式仪表图片，设计了一种鲁棒性较强的自动读数算法。该算法包括：首先用 SIFT (Scale-Invariant Feature Transform) 算法进行仪表区域的定位；其次，使用 OSTU (大津算法) 进行阈值分割；然后，运用形态学处理的相关技术进行指针提取；最后，计算指针与水平线之间的角度，根据相应的线性数学关系得到仪表的读数。

Yifan Ma^[40]等人提出了一种完整的基于机器视觉的指针式仪表自动读取算法，该算法可以应用于复杂照明环境下的检查机器人。首先，利用圆度法识别仪表中心，具体为采用八个链码形状来表示像素图像的形状，克服了某些连接点的圆度的反常计算域；其次，提出了一种基于对称度的二值化阈值确定方法，可以有效地确定指针区域从指针阴影分段；然后，该作者还提出了一种改进的 RANSAC 算法来识别指针线，解决了图像周围的图像干扰指针；最后，根据指针线和零刻度所在直线之间的角度计算仪表示数。实验结果证明该算法快速准确，可以有效地去除指针阴影并具有良好的抗干扰性。

1.3.1 论文的主要研究内容

如图 1-3，发电厂中最常用的仪表有字符仪表、指针式仪表、油压表，随着近年来数据采集的设备和方法越来越多，计算机视觉技术也日趋成熟，本文使用图像处理和深度学习的相关技术解决了发电厂的字符仪表、指针式仪表、油压表的自动识别问题，包括仪表定位、仪表分类、仪表示数识别。其中，仪表盘定位采用了 Faster RCNN、SSD、YOLO 三种算法，经过实验对比，选择 Faster RNN 算法进行目标检测；仪表类别判断采用了 LeNet5、AlexNet、VGG、InceptionV4 四种分类算法，经过实验对比，选择了 InceptionV4 算法进行类别判断；在仪表示数识别部分，使用 CRNN 算法进行字符仪表示数识别，指针式仪表和油压表示数识别当前主流的方法还是基于传统图像处理方法，由于传统图像处理方法具有一定的局限性，因此，本文提出了一种基于卷积神经网络的示数识别方法，该方法降低了传统图像处理对仪表图片质量的依赖性，增加了示数识别的准确性且具有一定的泛化性。



图 1-3 仪表示例



图 1-4 算法流程图

本项目基于 Python 语言,所用的编译器是 Anaconda,所用开源框架为 Tensorflow,使用 Tensorflow API 进行算法实现, TensorFlow 具有多种语言可用的 API,可用于构建和执行 TensorFlow 图。 Python API 目前是最完整和最容易使用的,但是其他语言 API 可能更易于集成到项目中,并且可能在图形执行方面提供一些性能优势。

在实际环境中,仪表图像的采集并不能时刻保证角度正常、图像清晰。因此,本文所使用的仪表图片均在实地采集,以保证所提方法的实用性。经实验证明,本文使用和提出的方法十分有效,各部分准确率都达到了 90%及以上,实现了在复杂环境中指针式仪表的定位、类别判断、示数读取的自动化。

1.3.2 论文的组织结构

本文主要研究内容为发电厂指针式仪表、字符仪表、油压表的识别技术,本文各个章节的内容如下:

第 1 章:绪论。简述发电厂各种仪表识别的背景意义和主流方法、国内外研究现状,以及概述文章研究内容。

第 2 章:深度学习模型理论。介绍深度学习的基本理论与发展,卷积神经网络的基本组成结构,经典的卷积神经网络以及常用的深度学习实现平台。

第 3 章:电厂仪表定位。介绍常用目标检测算法 Faster R-CNN、SSD、YOLO 的基本原理以及相应的实验结果。

第 4 章:电厂仪表分类。介绍常用图像分类算法 LeNet5、AlexNet、VGG-Net、InceptionV1~V4 的基本原理以及相应的实验结果。

第 5 章:电厂仪表识别。介绍卷积神经网络算法做回归的方式进行指针式仪表和油压表识别的基本原理、字符仪表识别以及实验结果。

第 6 章:工作总结和未来展望。该部分主要是对本文研究工作和实验的总结,同时也对未来工作和研究的展望。

第 2 章 深度学习模型理论

深度学习是一种模仿人类大脑运作的计算机技术,大脑运作的基本原理是通过眼睛获得外界的图像之后,由大脑中的第一阶神经元来分析该图像中的一些简单信息,例如判断眼前的物体是什么颜色的,然后,由第二阶段的神经元来进行更高一级的图像分析,得到一些更进一步的信息,例如判断眼前的物体是什么形状。人类的大脑有大量的神经元,由低到高层层递进,每一层的神经元分析图像得到的信息由简到繁,越往上得到的信息得到的信息越具体和细致,直到达到最上层将这些神经元得到的信息组合在一起,形成相应的图像,从而让人类准确的区分物体。深度学习机理也是如此,它通过事前给定的一些互相对应的数据作为初入和输出,通过类似人类神经元的数学分析,经过大量的计算后逐渐得到这些数据中潜藏的规律,最终可以通过这些规律来预测新的数据。

2.1 感知器

深度学习的本质就是人工神经网络,说起人工神经网络就离不开感知器。感知器^[41]是由费兰克·罗森布拉特(Frank Rosenblatt)基于MPC神经元模型提出。感知器是一个输入一个信号,通过数学操作可以判断这个信号属于有且仅有两个类中的哪一类的一种算法。可以预测数字向量所表示的输入是否属于特定的类(其将种类假定为标记+1和-1),分类器试图通过线性分离器来划分这两个类。感知器是最简单的神经网络,且感知器一般是指单层感知器,它是指只含有输入层和输出层的神经网络。

一个感知器由如下部分组成(图2-1): 1.输入层,该层包含多个输入分量,每一个输入分量 P_i ($i=1,2,\dots,n$)代表不同的数据,并且都会与一个权重分量 W_j ($j=1,2,\dots,n$)对应相乘,将得到的结果求总和作为阈值函数的输入。偏差 b 可以看做是特殊权重的分量,往往由常数充当,其含义是表示激活函数被激活难易程度; 2.输出层。输出层包含一个求和操作和函数 f 映射,其中 y 为感知器的输出(多输入,单输出), a 为输入加权求和的结果, f 为一种对 a 的激活函数,对感知器来说, f 一般为符号函数,即输出结果只有+1和-1(每次输入,对应一个单值得输出;多次输入,得到输出向量),输出层可以表示为:

$$a = \sum_{i=1}^{i=n} p_i \cdot w_j + b \quad y = f(a) \quad (2-1)$$

感知器的学习规则是梯度下降法,即计算实际输出和目标输出之间的误差,利用

梯度下降法修正权重和偏置，当达到一定的误差要求或者满足最大迭代次数时，训练结束，得到感知器模型。虽然感知器可以达到分类效果，但是它只能够解决一些简单的线性问题，一旦遇到一些略微复杂的非线性问题就无法解决。

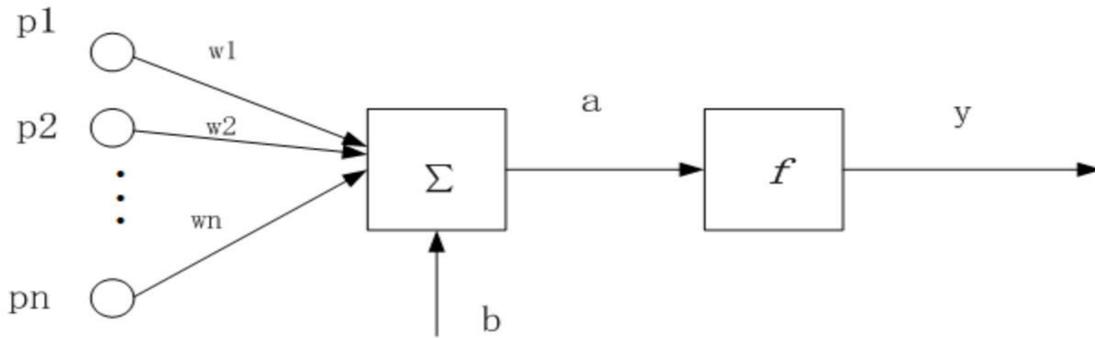


图 2-1 感知器

2.2 全连接神经网络

神经网络和感知器本质是一样的，神经网络是多层感知器，除了输入层和输出层中间还有隐藏层。我们说感知器的时候，其常常用符号函数作为激活函数进行一些线性的计算；神经网络也叫作特殊感知器，神经网络使用 sigmoid 函数或 tanh 函数处理非线性问题，代替符号函数作为阶跃函数，并且神经网络的层数要远远多于感知器。

神经网络其实就是按照一定规则连接起来的多个神经元，图 2-2 展示了一个全连接(fully connected,FC)神经网络。由图可知神经网络的主要组成部分为输入层、中间层、输出层。数据通过输入层进行网络，经过一系列数学计算后通过输出层得到相应的计算结果。中间层也叫隐藏层，因为其存在并不能够被外界所看到。相同层中的所有神经元都是独立存在的，在整个网络中，除去输入层之外，每一层的神经元都和其上一层的神经元相连并会有相应的连接权重，上一层的输出为下一层的输入，这也就是被称为全连接的原因。

从输入层到输出层之间的计算工程叫做前向传播，相邻两层连接的神经元根据式 2-2 从输入层开始，经过中间层，最后到达输出层，层层递进，依次计算每个神经元的值。

$$\bar{y} = f_{network}(\bar{x}) \quad (2-2)$$

图 2-2 中，输入层有三个输入神经元，依次编号为 1、2、3；隐藏层有四个神经元，依次编号为 4、5、6、7；输出层有两个神经元，依次编号为 8、9。其中，每个神经

元都与上一层的所有神经元有链接。以隐藏层神经元 4 为例，它与输入层的三个神经元都有连接，权重依次为 w_{41} 、 w_{42} 、 w_{43} ，根据式 2-3 可以计算神经元 4 的输入值 a_4 ，依次类推最终可以得到输出值。然后，计算实际输出和目标输出之间的误差，利用反向传播法修正权重和偏置，当达到一定的误差要求或者满足最大迭代次数时，训练结束，得到神经网络模型。

$$a_4 = \text{sigmoid}\left(\sum w \cdot x\right) \quad (2-3)$$

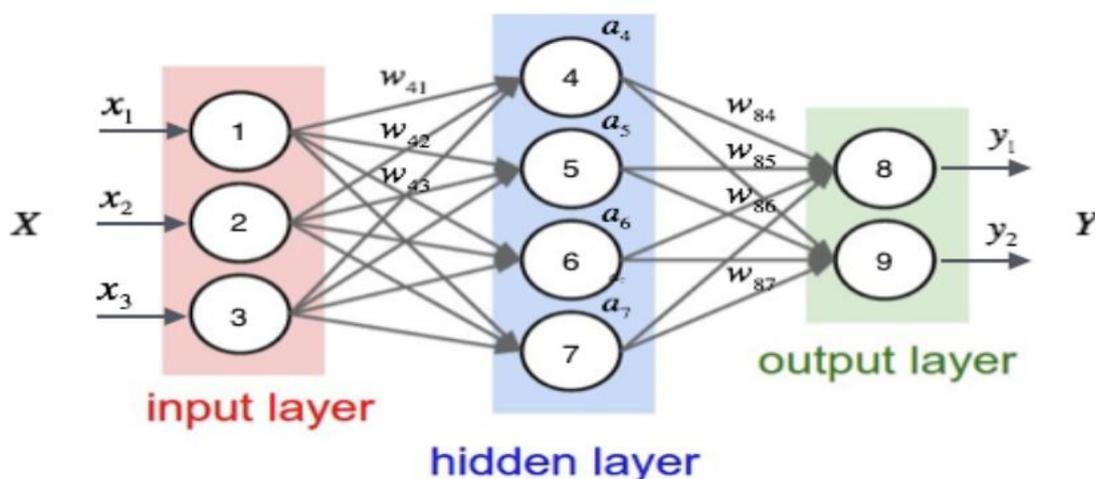


图 2-2 全连接神经网络

2.3 卷积神经网络

全连接神经网络成功解决了非线性问题，尤其在图像识别领域也取得了很好的成绩，例如经典的手写数字识别，可以达到 98% 的准确率。但是，随着全连接神经网络广泛的应用，其缺点也开始暴露了出来，主要的缺点为参数量过多，如果输入一张 1000×1000 大小的图片，输入层就有 100 万个神经元，假设隐藏层有 100 个神经元，按照前向传播计算方式，这一层的参数量就达到了 1 亿！这还仅仅是一层，如果有多层隐藏层的话，那么参数量就会无法估量。

卷积神经网络(CNN, Convolutional Neural Network)的提出成功解决了这个问题，并成为了当下深度学习主流的算法。其主要思路是：神经元由部分连接代替了全连接，除去输入层，中间层和输出层的所有神经元都只会和所在层的上一层中的神经元部分相连；权重共享，每一组连接共同使用一个权重，代替了之前每个连接就有一个权重；通过下采样的方式进一步减少参数量。

卷积神经网络主要有三部分构成：卷积层、池化层和非线性函数。其中，卷积层

的计算方式如图 2-3，输入一张图片，使其一个不同尺寸的卷积核进行卷积操作，原理为：

$$S(i, j) = \sum_m \sum_n I(m, n)K(i - m, j - n) \quad (2-3)$$

具体为，从图片左上角开始，从左到右，从上到下的选出与卷积核一样大小的区域，此区域中的像素与卷积核里的权重，一一对应相乘后相加，操作结束后卷积核可以根据不同的步长与其他图片区域进行卷积，在此过程中，图片中的所有区域都与同一个卷积核进行卷积操作，这就是权重共享的思想。

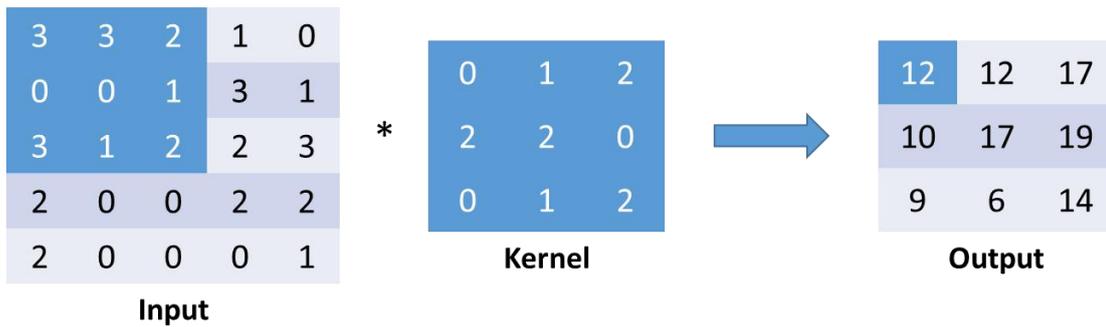


图 2-3 卷积操作

卷积操作之后紧接着就是下采样操作，下采样层也可以称为池化层(pooling)。池化操作和卷积类似，目前最常用的就是最大池化，如图 2-4，其主要是选择不同大小的窗口，使得卷积之后的矩阵按照窗口的大小选出最大的数作为输出矩阵的相应元素的值，从而进一步减少了参数量。

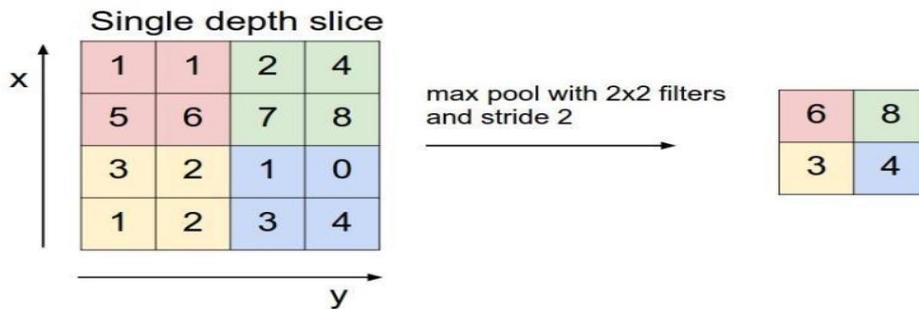


图 2-4 池化操作

卷积神经网络相较全连接神经网络，除了卷积和池化操作之外，最大的不同就是非线性函数的选择。全连接神经网络通常使用 sigmoid 函数作为激活函数，但是 sigmoid 函数在趋向无穷的地方时，函数值变化很小，会产生梯度消失的现象。因此，卷积神经网络选择了 relu 函数作为常用激活函数，其优点是可以使网络训练的更快，增加了网络的非线性和稀疏性，防止梯度消失。图 2-5 为 sigmoid 函数和 relu 函数的函数图像。

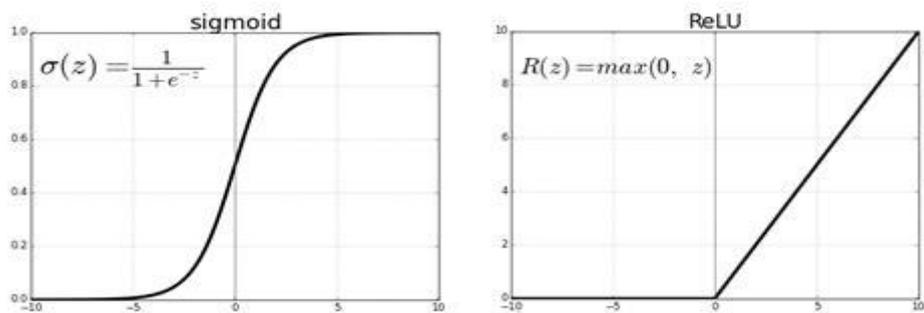


图 2-5 sigmoid 函数和 relu 函数

2.4 本章小结

本章简要介绍了深度学习的感知器、全连接神经网络、卷积神经网络的原理和发展情况。本文所用方法全部基于深度学习卷积神经网络，故了解其原理为重中之重。

第3章 电厂仪表定位

电厂仪表定位是电厂仪表识别流程中第一个环节，该环节是后面所有环节的基础。由于发电厂环境复杂，各类仪表分布不均，光照不均衡，这些因素对仪表定位造成了很大的阻碍。仪表定位属于图像处理中目标检测的范畴，传统目标检测的方法主要有模板匹配、SIFT等，这些方法虽然有一定的效果，但是需要预先手动选取特征，自适应性较弱，并且容易受到噪声影响。目前，随着深度学习技术的大爆炸，目标检测算法也有了很大的发展，主要是用卷积神经网络代替传统图像处理方法的手工提取特征来自动提取特征。基于CNN的目标检测算法主要分为两个分支：**two stage**和**one stage**。两个分支的主要区别为，**two stage**算法需要先生成候选框，然后再将所有的候选区域送入分类器进行分类，例如Faster R-CNN算法；而**one stage**算法不需要，通过一个卷积神经网络就可以得到最终的候选框，例如YOLO算法。

3.1 Faster R-CNN 基本原理简介

说起Faster R-CNN就不得不从R-CNN说起。2014年，Ross Girshick等人提出了一种针对高准确度的目标检测与语义分割的多特征层级方法(R-CNN)，其本质就是一个用来做目标检测和语义分割的神经网络。R-CNN^[42]算法的提出具有很多比较重要的意义：1.在Pascal VOC 2012的数据集上能够将目标检测的验证指标mAP提升到53.3%，这相对于之前最好的结果提升了整整30%；2.R-CNN证明了可以将神经网络应用在自底向上的提取候选区域，这样就可以进行目标分类和目标定位；3.R-CNN的出现也带来一个观点，当缺乏大量的标注数据时，比较可行的方法是进行神经网络的迁移学习，采用在其他大型数据训练过后的神经网络，然后在小规模的特定的数据集中进行fine-tune微调。

R-CNN主要流程为，如图3-1，输入一张图片，使用选择性搜索(Selective Search)^[43]从图像中生成2000个左右的区域候选框(region proposal)，对于每一个区域候选框使用CNN进行特征提取，并组合特征。使用经典机器学习分类算法SVM对选择性搜索算法生成的候选区域进行分类，得到目标类别，最后使用一个回归算法对候选区域的坐标进行修正进而得到最终的边界框。其中，在特征提取阶段，R-CNN抽取了一个4096维的特征向量，采用的是AlexNet，但是通过Selective Search产生的区域候选框大小是不一的，为了与AlexNet兼容，R-CNN强制区域候选框大小变为227*227的尺寸。

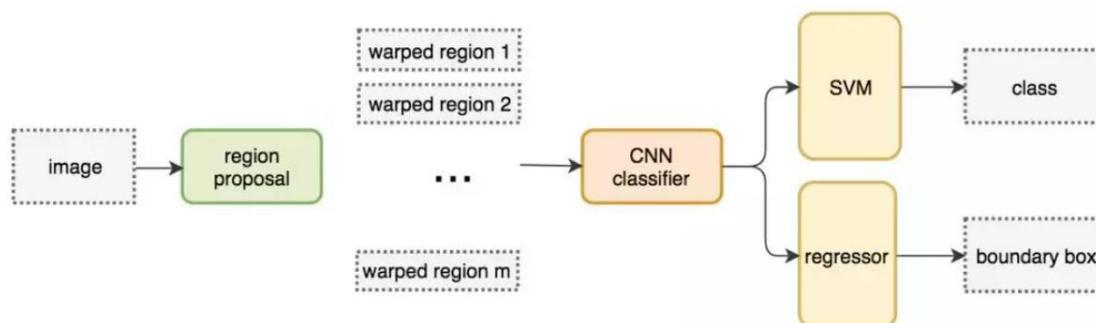


图 3-1 R-CNN 网络结构

在测试阶段，R-CNN 在每张图片上抽取近 2000 个候选区域。然后将每个候选区域进行尺寸的修整变换，送进神经网络以读取特征，然后用 SVM 进行类别的识别，并产生分数。候选区域有 2000 个，所以很多候选区域会产生重叠。针对每个类，通过计算 IoU 指标，采取非极大性抑制，以最高分的区域为基础，剔除掉那些重叠位置的区域。其中，IoU 为交并比，是预测框和真实框之间的重叠区域与并集区域的比值。

尽管 R-CNN 有诸多优势，但是也存在一些缺陷，例如训练分为多个阶段，步骤繁琐；训练耗时，占用磁盘空间大；测试速度慢，每个区域候选框需要运行整个前向 CNN 计算；SVM 和回归是事后操作，在此过程中 CNN 特征没有被学习和更新。

为了解决 R-CNN 的缺点，Ross Girshick 等人又提出了 Fast R-CNN，顾名思义，它是加速版的 R-CNN。如图 3-2，Fast R-CNN^[44] 相较 R-CNN 改进的地方主要是使用 CNN 先对整个图像进行特征的提取，然后使用 Selective Search 算法在这些特征图上生成候选框。具体为，Fast R-CNN 使用 VGG16 网络中 $H=W=7$ 的参数，即将一个区域候选框分割成 $H*W$ 大小的网络，然后将这个区域候选框映射到最后一个卷积层输出的特征图，将特征图变为固定大小的尺寸，然后送入到全连接层进行目标分类和候选框坐标回归修正。将回归器放入网络一起训练，每一个类别对应一个回归器，同时使用 softmax 代替原来的 SVM 作为分类器，从而整体上减少了计算量和处理时间，达到了加速的效果。

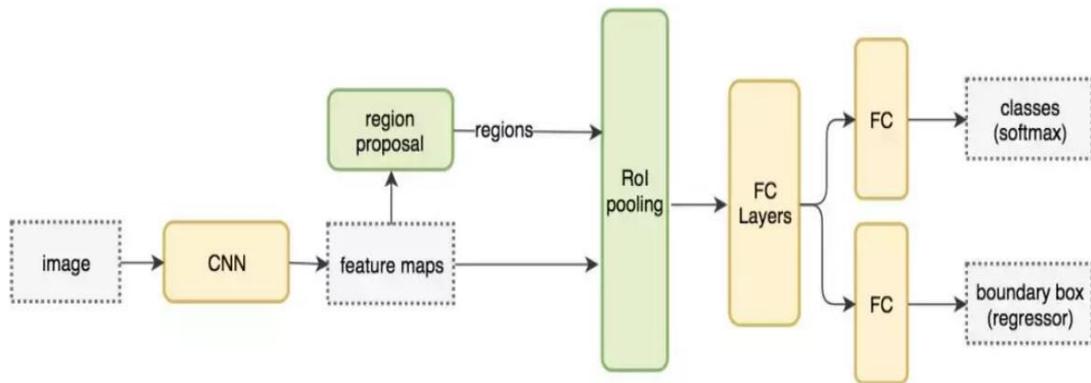


图 3-2 Fast R-CNN 网络结构

虽然 Fast R-CNN 相较于 R-CNN 在速度上有了很大的改进，但是在提取候选区域这一操作上，Fast R-CNN 还是使用的 Selective Search 方法，该方法在 CPU 上的运行非常耗时。为了进一步提升速度，Shaoqing Ren, Kaiming He 等人在 2015 年提出了 Faster R-CNN 算法。如图 3-3，Faster R-CNN^[45]整体结构可以看做主要有两部分组成，一部分是 RPN 网络(Region Proposal Network,RPN)，另一部分还是 Fast R-CNN 网络框架。Faster R-CNN 相较于 R-CNN 和 Fast R-CNN 最大的创新点就是使用 RPN 网络生成区域候选框，RPN 在生成区域候选框时效率更高，并且以每张图像 10 毫秒的速度运行。

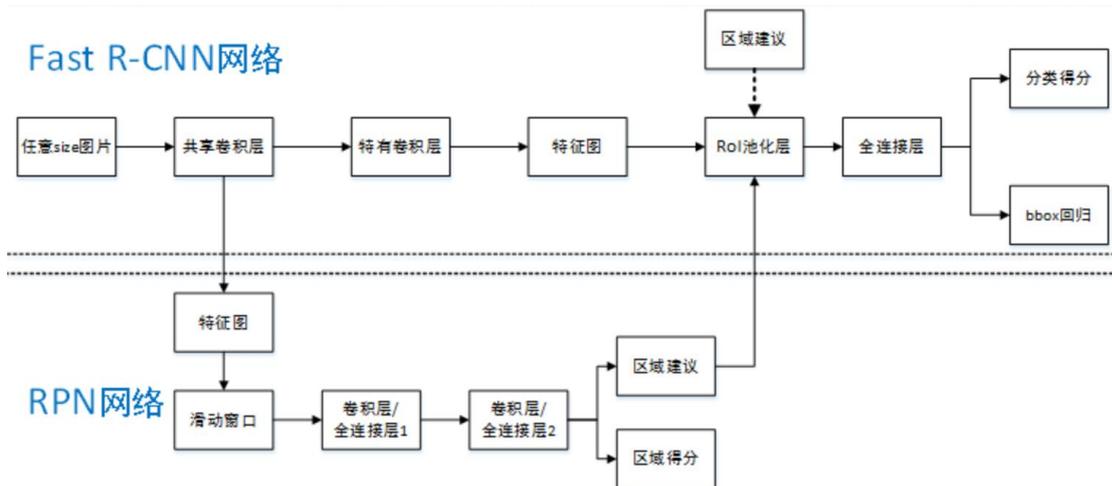


图 3-3 Faster R-CNN 网络结构

RPN 的主要工作机理是，将 Faster R-CNN 的公共特征图作为输入，使用 3×3 大小的卷积核在这些特征图上进行卷积操作，最后得到一个通道数为 256 的特征图，将得到的特征图送入到两个全连接层，一个得到前景或背景的 2 个分数，另一个得到 4 个坐标，这 4 个坐标是基于原图的偏移量。

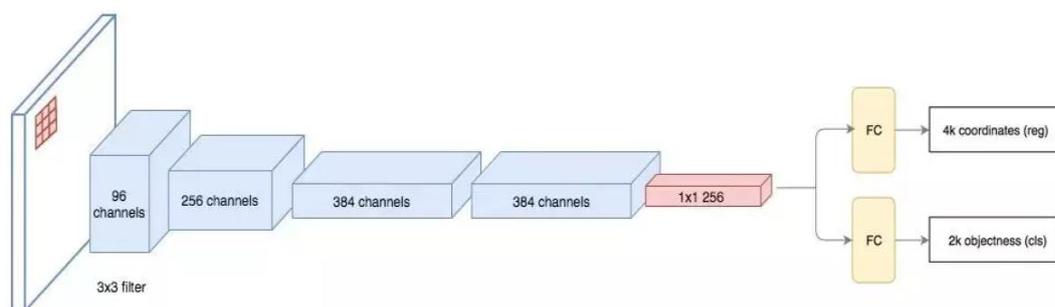


图 3-4 RPN 网络结构

对于特征图中的每一个位置，RPN 会做 k 次预测。因此，RPN 将输出 $4 \times k$ 个坐标和每个位置上 $2 \times k$ 个得分。图 3-5 展示了 8×8 的特征图，且有一个 3×3 的卷积核执行运算，它最后输出 $8 \times 8 \times 3$ 个 ROI（其中 $k=3$ ）。图 3-5（右）展示了单个位置的 3 个候选区域。

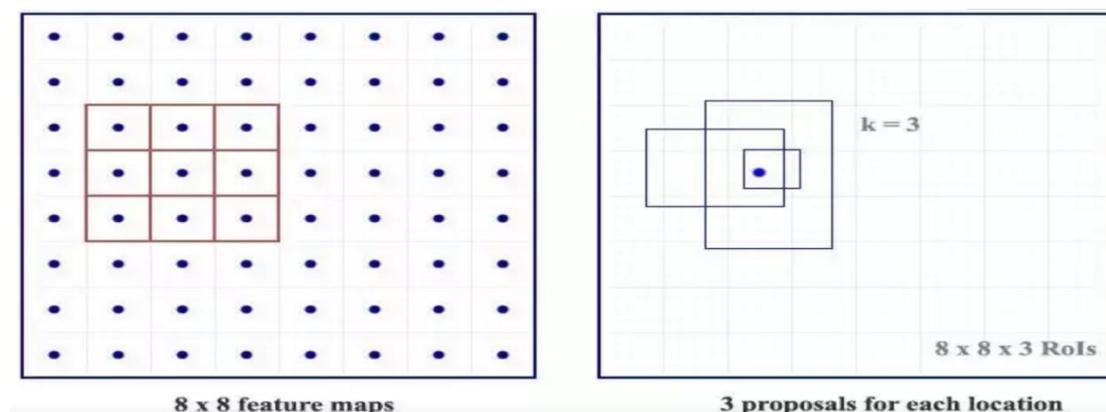


图 3-5 8×8 的特征图

Faster R-CNN 会预测一些与左上角名为锚点的参考框相关的偏移量，要对每个位置进行 k 个预测，得到以每个位置为中心的 k 个锚点。这些锚点是精心挑选的，因此它们是多樣的，且覆盖具有不同比例和宽高比的现实目标。Faster R-CNN 部署 9 个锚点框，如图 3-6，3 个不同宽高比的 3 个不同大小的锚点框。每一个位置使用 9 个锚点，每个位置会生成 2×9 个目标分数和 4×9 个坐标，然后送入分类器和回归器得到相应的目标和候选框。

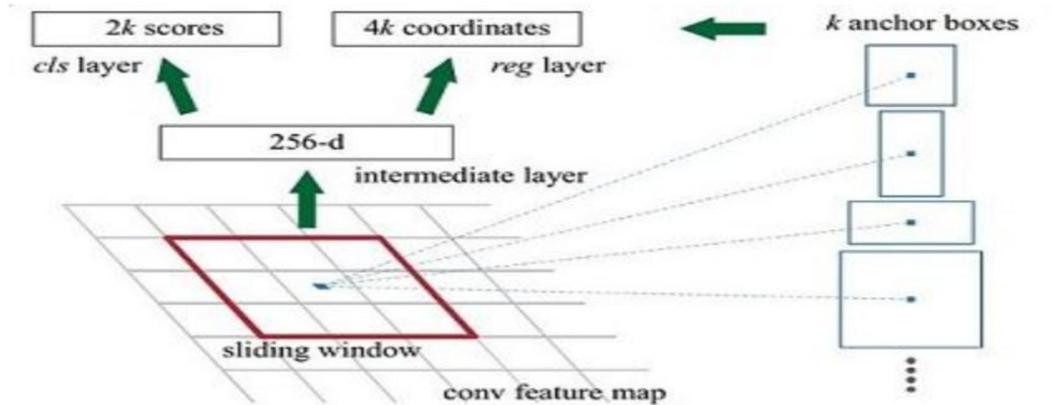


图 3-6 锚点生成流程图

3.2 YOLO 基本原理简介

Faster R-CNN 是典型的 two stages 算法，其使用 RPN 网络先在图像中生成区域候选框，然后在这些候选框上训练分类器。在分类完成之后，之后的处理用来精细化处理边界框，消除重复检测，并基于场景中的其他对象重新定位框。因为每一个步骤都需要单独训练，这导致操作流程非常复杂并且速度还是较慢。

基于此，Joseph Redmon, Santosh Divvala, Ross Girshick 等人提出了一种 one stage 算法，YOLO (You Only Look Once)。如图 3-7，整体来看，YOLO^[46] 算法的主要思路是不需要预先生成区域候选框，而是通过把整张图片分割成多个固定大小的格子，对这些格子用一个 CNN 网络直接将目标识别和候选框修正等问题一次性解决。其过程主要分三步：1. 将图片的尺寸变成固定的大小，多用 448*488 大小；2. 使用 CNN 网络对图片进行特征提取，进而得到目标类别和框坐标等信息；3. 采用非极大抑制机制得到最后的结果。YOLO 是典型的 one stage 端到端的算法，其速度更快。

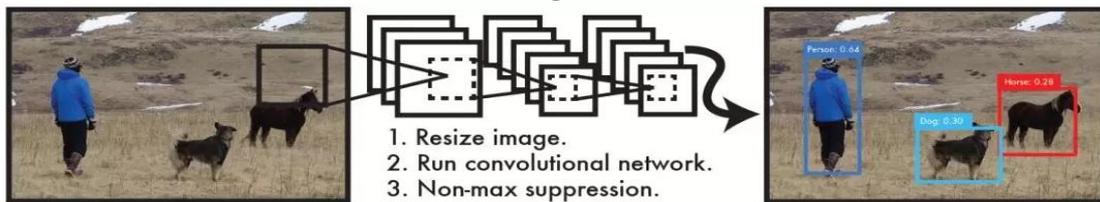


图 3-7 YOLO 网络流程图

如图 3-8，YOLO 采用 CNN 来提取特征，然后用全连接层来得到预测值，整个网络包括 24 个卷积层和 2 个全连接层，对于卷积层主要使用 1*1 卷积来减少通道数，然后紧跟 3*3 卷积。具体流程为：先将输入的图片以固定的尺寸划分成多个小方格，每一个小方格对落在其内部的内容进行检测，通常固定的尺寸为 448*448*3，经过一系列卷积操作之后，得到的输出维度为：

$$S * S * (B * 5 + C) \tag{3-1}$$

其中， S 代表图片总共分割的方格数， B 代表小方格存在的多少个目标数， C 代表图片里总共有多少个类别。每一个小方格会产生 B 个区域候选框，以小方格的中心作为区域候选框中心，最大范围是图片的宽高来搜寻物体的候选框位置；产生的每个区域候选框都会有一个与其对应的得分，该得分是指这个小方格内是否有检测的物体以及其定位的准确性 $P(object) * IoU_{pred}^{truth}$ ，每个小方格都会对应整个图片中所有类别数 C 个概率值，只要计算出最大概率值并且找出最大概率显示的具体类别 $P(Class | object)$ ，就可以并认该小方格内有物体或者是物体的一部分。

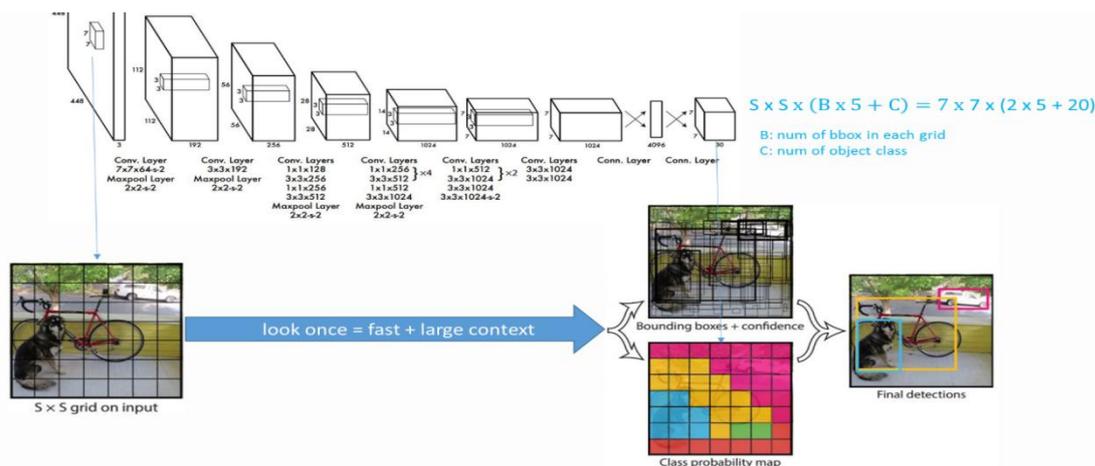


图 3-8 YOLO 网络结构图

3.3 SSD 基本原理简介

YOLO 虽然比较简洁并且速度快，但其检测精度并不如 two stage 算法，尤其是检测密集问题时效果十分不稳定。为了同时保证算法的精度和速度，Wei Liu 等人提出了一种新的 one stage 算法，SSD(Single Shot MultiBox Detector)^[47]。

SSD 的主要核心思想为，使用应用于要素图的小卷积滤波器来预测固定的默认边界框的类别得分和框偏移；从不同尺度的特征图中快速检测不同尺度的准确度，并通过纵横比明确区分预测；即使在低分辨率输入图像，这些设计特征也可实现简单的端到端训练和高精度，从而进一步提高速度与精度之间的权衡。

如图 3-9，SSD 使用经典卷积神经网络 VGG16 为基础，在此基础上额外增加了卷积层作为核心网络进行特征提取。算法步骤为，输入一张 300*300 大小的图片，将其输入到预训练好的分类网络中来获得不同大小的特征映射，将 VGG16 网络中的第六全连接层和第七全连接层两层全连接层转化为卷积层。以图中的第六、7 卷积层为例，去掉所有的 Dropout 层和第 8 全连接层，将第五个池化层的窗口大小和步长变换到 3*3 大小、步长为 1；抽取 Conv4_3、Conv7、Conv8_2、Conv9_2、Conv10_2、Conv11_2 层的特征图，然后分别在这些特征图上面的每一个点构造 6 个不同尺度大小的边界框，然后分别进行检测和分类，生成多个边界框；将不同特征图活动的边界框结合起

来, 经过非极大抑制方法来抑制一部分重叠或者不正确的边界框, 生成最终的边界框集合(即检测结果)。

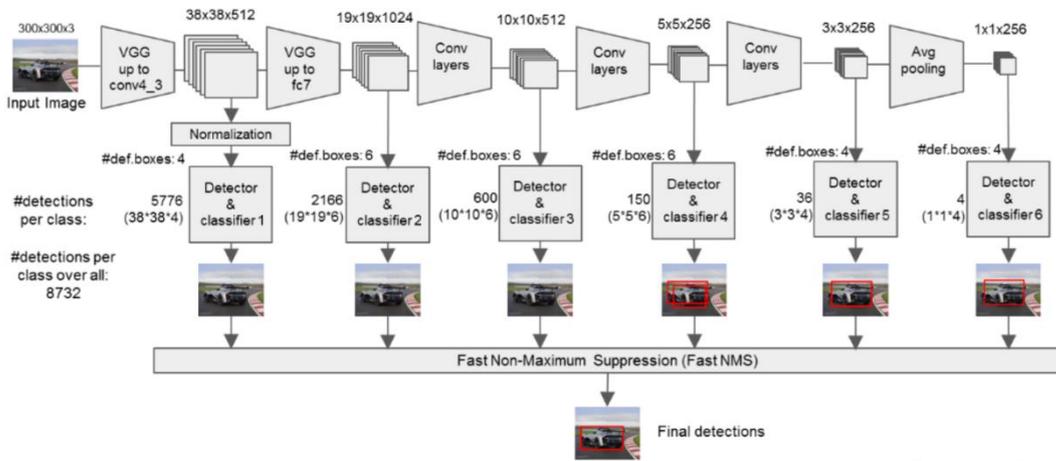


图 3-9 SSD 网络结构图

3.4 仪表检测算法实验分析

在基于深度学习仪表检测算法实验中, 实验以 GPU 集群 Linux 服务器作为实验硬件平台, 其主要计算资源为 4 块 12GB 显存, 计算能力为 3.6 的 NVIDIA-GTX-Titan-XP 型号的 GPU, 基于 CUDA 和 cuDNN 的底层计算框架。使用 TensorFlow 框架中的目标检测 API 进行 Faster R-CNN、YOLO、SSD 算法的实现。

由于在现场采集的指针式仪表不具有多样性, 其中, 所谓的多样性是指电厂整个运行状态趋于稳定, 指针式仪表所示数值几乎不变, 对我们后续的指针式仪表识别部分影响很大。所以, 我们在实验室自己搭建了一个实验平台, 使用气泵对两种指针式仪表进行多示数、多角度、多光照下的图像采集, 如图 3-10, 两种指针式仪表的量程分别为 0~1 和 0~0.4。





图 3-10 实验室采集指针式仪表图片示例

本文所用数据集包含四类仪表，共有 3480 张图片，每张图片都是以类别+序号命名。仪表类别分别为，两类不同量程的指针式仪表，量成为 0~1 和 0~0.4，两类指针式仪表各有 1240 张图片；字符仪表和油压表共有 1000 张图片,图 3-11 为字符仪表和油压表示例图片。



图 3-11 字符仪表和油压表示例

我们使用 `labelimg` 软件对每张图片做标记框(ground truth),得到相应图片的 `xml` 文件, `xml` 文件里记录了对应图片的名字、尺寸、仪表框坐标等信息。将这些仪表图片和对应的 `xml` 文件按照 8:2 和 9:1 两种不同的比例分为训练集和测试集,用 `creat_pascal_tf_record.py` 脚本将训练集和测试集转换成对应的 `tf.record` 格式的文件:`train.record` 和 `test.record`,同时,将 `xml` 文件转换成 `txt` 格式的文件,分别为 `trainImagePath.txt` 和 `vallImagePath.txt`,该文件为 YOLO 算法做准备。然后,建立一个 `labellist.pbtxt` 文件,改文件包含了本次目标检测的目标序号和名字信息,将 `laebellist.pbtxt`、`train.record`、`test.record` 三个文件放入 `Data` 文件夹下。至此,数据准备工作结束。

接下来下载所用目标检测算法的模型文件,本文所用的 `Faster R-CNN` 和 `SSD` 两种算法的模型文件为 `faster_rcnn_inceptionV2_coco.config`、`ssd_inceptionV2_coco.config`。将 `config` 文件中的数据信息、模型训练地址等信息根据我们自己的数据做出相应的修改,并且不使用模型微调功能,重新训练我们自己的模

型。然后，使用 TensorFlow 目标检测 API 中的 train.py 文件进行模型训练，训练迭代次数为 50000 次，将训练好的模型放入到 model 文件中。模型训练结束后，使用 API 中的 eval.py 文件对模型进行评估。

使用 YOLO 算法进行训练的流程为，下载 YOLO 模型文件，在“Image”文件夹下存放所有的图片样本，包括训练数据和验证数据，建立“XML”文件夹作为“Image”的同级文件夹，该文件夹存放所有图片的 xml 文件。然后，修改 YOLO 模型的相应 config 文件，方法与上述类似；下载预训练文件 cfg/darknet19_448.conv.23,放在 darknet 文件夹下，在 darknet 文件下运行训练指令，训练迭代次数为 50000 次，模型训练好之后运行相应的评估指令进行模型评估。

在本文仪表检测实验中，使用平均准确率均值(mAP)作为仪表检测模型准确性的评价指标。表 3-1 为三种算法最后的测试结果，综合来看，Faster R-CNN 的效果是最好的，故在仪表检测这部分我们最终选择了 Faster R-CNN 算法做为检测算法。图 3-12 为 Faster R-CNN 算法的测试结果图示例。

表 3-1 三种定位算法实验结果

	Train Set	Faster R-CNN	SSD	YOLO
mAP	8: 2	0.96	0.90	0.84
mAP	9: 1	0.99	0.93	0.89

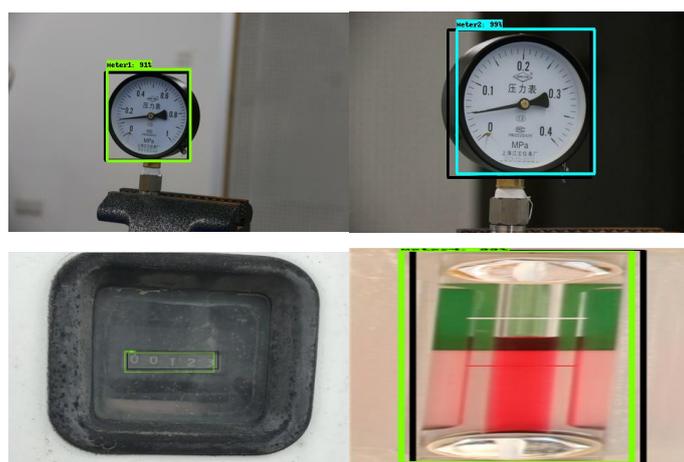


图 3-12 仪表定位结果

3.5 本章小结

本章主要介绍了仪表检测所用的方法和实验结果。简要介绍了当前主流的目标检测算法 Faster R-CNN、YOLO、SSD 的基本原理，并介绍了相关实验的具体流程，训

练了对应算法的模型，使用 mAP 作为评估指标。经过实验对比，Faster R-CNN 算法在仪表检测中效果最好，在训练集和测试集的不同比例下都达到了 95% 以上的准确率，故在仪表检测这一部分中选择了 Faster R-CNN 作为检测算法，为后续的仪表类别识别、仪表示数识别起到了关键作用。

第 4 章 电厂仪表类别识别

仪表种类识别是仪表识别整体算法中的第二步，本文使用深度学习卷积神经网络的 4 个经典算法，对 Faster R-CNN 算法对仪表定位之后，根据位置框的四个坐标进行裁剪得到的仪表图片进行种类识别。

4.1 LeNet-5 基本原理简介

LeNet-5^[48]是 Yann LeCun 在 1998 年设计的用于手写数字识别的卷积神经网络，当年美国大多数银行就是用它来识别支票上面的手写数字的，他是早期卷积神经网络中最有代表性的试验系统之一。

如图 4-1，尽管 LeNet-5 整体的网络结构比较简单，但是它在有限的结构中却涵盖了深度学习的基本模块：卷积层，池化层，全连接层。可以说该网络是其他深度学习模型的基础。LeNet-5 除了输入层之外，共有 7 层，其中 C1、C3、C5 是卷积层，S2、S4 是池化层，F6 是全连接层。

具体步骤为，C1 层对图像进行第一次卷积运算，使用 6 个大小为 5×5 的卷积核，得到 6 个大小为 28×28 的特征图(feature maps)；S2 层是对 C1 层得到的特征图使用最大池化进行下采样，池化的尺寸为 2×2 ，得到 6 个 28×28 的特征图变成了 14×14 的特征图，减少数据处理量的同时保留了有用信息；C3 层是一个卷积层，卷积和 C1 层相同，不同的是 C3 层的每个节点与 S2 层中的多个特征图图相连，表示本层的特征图是上一层提取的特征图的不同组合，C3 层使用 10 个 5×5 的卷积核，对个 S2 层得到的特征图进行特殊运算得到 16 个 10×10 的特征图；S4 是池化层，窗口大小依旧是 2×2 ，C3 层得到的特征图经过池化后得到 16 个 5×5 的特征图；C5 层使用 120 个 5×5 的卷积核对 S4 层得到的特征图进行卷积操作，由于卷积核和特征图的尺寸一样，所以卷积后形成的特征图的大小为 1×1 ；F6 层使用 84 个神经元与 C5 层得到的特征图进行全连接，得到的结果与输出层的 10 个神经元进行全连接，10 个神经元代表手写数字 0-9 这 10 个类别，最终计算出相应的结果。

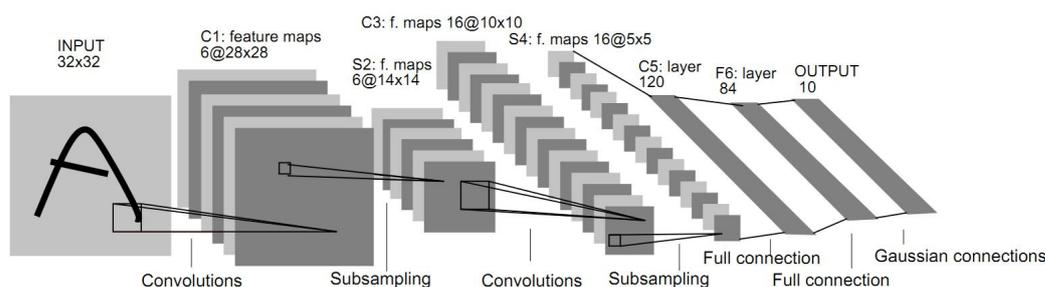


图 4-1 LeNet5 网络结果

LeNet-5 是最采用 CNN 的网络，该网络证明了仅仅需要些许的预处理就可以得到原图片很多的信息和识别视觉上的规律。然而，由于当时缺乏大规模训练数据，计算机的计算能力也跟不上，LeNet-5 对于复杂问题的处理结果并不理想。

4.2 AlexNet 基本原理简介

由于受到计算机性能的影响，虽然 LeNet-5 在图像分类中取得了较好的成绩，但是并没有引起很多的关注。直到 2012 年，Alex 等人提出的 AlexNet 网络在 ImageNet 大赛上以远超第二名的成绩夺冠，卷积神经网络乃至深度学习重新引起了广泛的关注。

AlexNet^[49]的网络结构在整体上与 LeNet-5 相似，二者皆为输入图像后进行卷积、池化、然后最后进行全连接。但是两个网络在很多细节上截然不同，可以说 AlexNet 要比 LeNet-5 更为复杂。AlexNet 共有 60 million 个参数和 65000 个神经元，最终的输出层输出 1000 个类别。除了有更深的网络深度之外，AlexNet 使用了层叠的卷积层，增加了 Dropout 层和数据增强的手段来抑制过拟合，使用 Relu 函数作为激活函数，避免了 sigmoid 函数造成的梯度消失的问题，使用两块 GPU 同时计算，大大提高了运算的效率。

如图 4-2，AlexNet 网络一共含有 8 层网络，包括五层卷积层和 3 层全连接层。具体步骤为：输入图片尺寸为 $224 \times 224 \times 3$ ，经过卷积层 C1 时，由 96 个 $11 \times 11 \times 3$ 的卷积核，以步长为 4，不填充边缘的方法对图片进行卷积操作，两块 GPU 分别计算 48 个卷积核，得到的特征图维度为 $55 \times 55 \times 96$ ，然后使用 3×3 的窗口，步长为 2 对得到的特征图进行池化操作，输出为 $27 \times 27 \times 96$ 的特征图，然后使用 $k=2, n=5, \alpha=10^{-4}, \beta=0.75$ 进行局部归一化，输出的特征图仍为 $27 \times 27 \times 96$ ，将输出分为两组，每组的大小为 $27 \times 27 \times 48$ ；经过卷积层 C2 时，使用两组，每组 128 个 $5 \times 5 \times 48$ 的卷积核，边缘填充为 2，步长为 1 对 C1 层输出的特征图进行卷积操作，每组得到新的特征图为 $27 \times 27 \times 128$ ，然后对其进行，以 3×3 的窗口，步长为 2 的池化操作，输出为 $13 \times 13 \times 256$ ，随即使用和 C1 层相同的局部归一化操作，并将输出分为两组，每组的大小为 $13 \times 13 \times 128$ ；

经过卷积层 C3 时，该层使用 384 个 $3 \times 3 \times 256$ 的卷积核对上层输出的特征图进行以边缘填充为 1，步长为 1 的卷积操作，得到的特征图大小变为 $13 \times 13 \times 384$ ；经过卷积层 C4 时，将上层输出的卷积图分为两组，每组大小为 $13 \times 13 \times 192$ ，然后使用 2 组，每组以 128 个 $3 \times 3 \times 192$ 的卷积核，边缘填充为 1，步长为 1 对其进行卷积操作，随即进行以 3×3 的窗口，步长为 2 的池化操作，得到的特征图为 $6 \times 6 \times 256$ ；经过全连接层 FC6 时，使用和上层输出的特征图一样大小的卷积核，共 4096 个，与特征图一一对应相乘，得到的特征图为 $1 \times 1 \times 4096$ ，即 4096 个神经元，然后采用 Dropout 手段，随机的断开某些神经元的连接从而抑制过拟合；经过全连接层 FC7 时，该层和 FC6 基本相同，使用 4096 个神经元进行全连接，同样采用 Drop 手段，得到 4096 个输出，并与输出层的 1000 个神经元相连，最终得到 1000 个输出。

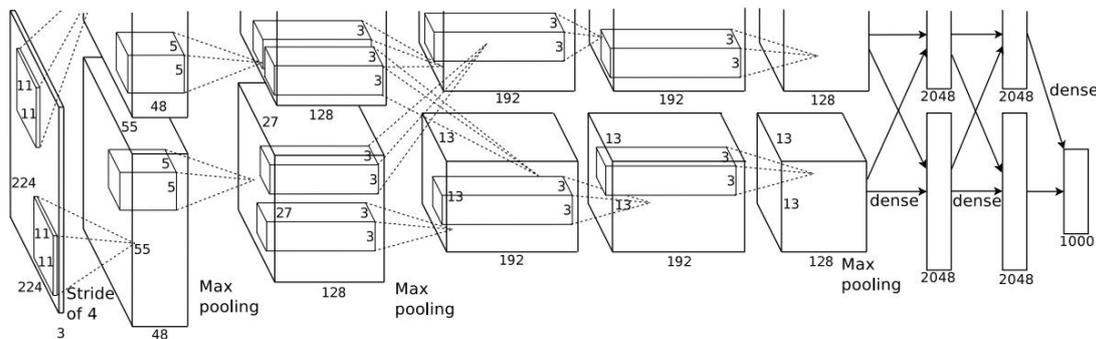


图 4-2 AlexNet 网络结构图

4.3 VGG-16 基本原理简介

VGGNet 是在从 AlexNet 网络基础上发展而来。VGGNet^[50]和 AlexNet 的不同之处除了深度要比 AlexNet 深之外，VGG 使用了大量的 3×3 尺寸的卷积核替换了 AlexNet 的卷积核， 3×3 卷积核是能够感受到上下、左右、重点的最小的感受野尺寸。VGGNet 根据层数分为了 VGG-16 和 VGG-19 两种结构的网络，其效果非常好，在 2014 年的 ILSVRC 比赛中，VGG 在 Top-5 中取得了 92.3% 的正确率。

VGG 之前的卷积神经网络在提取特征时候往往采用尺寸比较大的卷积核，例如 AlexNet 中的 11×11 大小的卷积核，虽然卷积核的大小代表提取的特征和信息也就越多，但是同时带了一个问题，那就是参数量的增加。针对这个问题，VGG 网络最大的创新点是使用多个 3×3 大小的卷积核进行组合，用以代替大尺寸的卷积核。举个例子，用 3 个 3×3 的卷积核来卷积得到像素对应的感受野大小与一个 7×7 卷积核得到的感受野大小是一样的。但是，参数量却是差了近一倍($3 \times 3 \times 3 = 27$ ， $7 \times 7 = 49$)。

如图 4-3，VGG 网络根据卷积核大小和卷积层数目的不同，可分为 A，A-LRN，B，C，D，E 这六种不同层数及结构(ConvNet Configuration),其中 D、E 也就是人们

常说的 VGG-16 和 VGG-19。图中每一列是其相应的具体配置，例如，图中 D 列就是 VGG-16 所采用的结构，VGG-16 共有 13 个卷积层(Convolutional Layer)，分别用 conv3-XXX 表示，3 个全连接层(Full connected Layer)，分别用 FC-XXX 表示，5 个池化层(Pool Layer),分别用 maxpool 表示，其中卷积层和全连接层具有权重系数，因此也被成为权重层，总数目为 13+3=16，这就是 VGG-16 的来源。(池化层不涉及权重，因此不属于权重层，不被计数)

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv1-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

图 4-3 VGG 网络结构图

VGG-16 的突出特点就是简单，具体为：卷积层均采用相同的卷积核参数，卷积层均表示为 conv3-XXX，其中 conv3 说明该卷积层采用的卷积核尺寸是 3，3*3 是很小的卷积核尺寸，结合其他参数(步长为 1，边缘填充为 same)，这样就能够使得每一个卷积层与前一层保持相同的宽和高，XXX 代表卷积层的通道数，也是卷积核的数量；池化层均采用相同的池化层参数，均为 2*2 窗口；模型是由若干卷积层额池化层堆叠(Stack)的方式构成，比较容易形成较深的网络结构。总之，VGG 的优点可以概括为：Small filters, Deeper networks。

在 VGG-16 的整个结构中，卷积层和池化层可以划分为不同的块(Block)，从 input 开始到 maxpool 一共有 5 个 Block，从前到后依次编号为 Block1~Block5。每一个块内包含若干卷积层和一个池化层。例如，Block4 包含 3 个卷积层，conv3-512，1 个池化层，maxpool。并且同一块内，卷积层的通道数是相同的，例如，Block2 中包含 2 个卷积层，每个卷积层用 conv3-128 表示，即卷积核为 3*3*3，通道数都是 128；Block3 中包含 3 个卷积层，每个卷积层用 conv3-256 表示，即卷积核为 3*3*256，通

道数都是 256。VGG-16 的输入图像是 $224 \times 224 \times 3$ ，卷积核数量依次为 $64 \rightarrow 128 \rightarrow 256 \rightarrow 512$ ，图像高宽减半，依次为 $224 \rightarrow 112 \rightarrow 56 \rightarrow 28 \rightarrow 28 \rightarrow 14 \rightarrow 7$ 。

尽管 VGG 的结构简单，使用了小尺寸卷积核叠加代替大尺寸卷积核，但是其参数量依然多的惊人。这些参数包括卷积核权重和全连接层权重，这导致训练时间会很长，调参难度变大，需要的存储容量大，不利于部署。

4.4 GoogLeNet 基本原理简介

GoogLeNet(Inception V1^[51])最早是出现在 ILSVRC 2014 的比赛中,Inception V1 以 6.67%的错误率夺得了冠军。Inception V1 的特点就是控制了神经网络的计算量和参数量,该网络只有 500 万的参数量,而经典网络 AlexNet 的参数量为 6000 万。Inception V1 一共有 22 层,比 VGG Net 更深。Inception 之所以选择了降低参数量,是因为,参数量的增加就意味着计算量和计算时间的增加,进而会加大消耗硬件的资源。Inception V1 效果好的主要原因为:第一,使用了一维卷积代替了全连接层,在达到降低维度的同时也觉少了参数量量、消除了全连接层引起的过拟合问题;第二,Inception V1 最大的创新点就是设计了 Inception Module,其作用是提高参数的利用率,Inception Module 其实就是大网络中的一个小网络,通过反复堆叠 Inception Module 来形成一个大网络。

图 4-4 是一个 Inception Module,它包含了 4 个分支,第一个分支是一个 1×1 的卷积,其作用是跨通道组织信息和特征变换,提高网络的表达能力,同时还可以对输出通道进行升维和降维;第二个分支先使用了一个 1×1 的卷积,然后再连接了一个 3×3 的卷积,相当于进行了两次特征变换;第三个分支的结构与第二个分支类似,只是将 3×3 的卷积换成了更大的 5×5 的卷积;最后一个分支,是

先进行 3×3 的较大池化,然后再使用 1×1 的卷积。最后,通过聚合操作在输出通道上将四个分支的结果进行合并。

在 Inception Module 中, 1×1 卷积的比例较高, 3×3 和 5×5 的卷积比例稍低。在整个网络中,堆叠了多个 Inception Module,目的是捕捉到更高阶的抽象特征。Inception V1 一共含有 22 层,除了最后一层的输出,其中间节点的分类效果也很好。所以,在 Inception Net 中,还使用了辅助分类节点(auxiliary classifiers),将中间某一层的输出用作分类,并按一个较小的权重(0.3)加到最终的分类结果中,相当于做了模型融合,同时还给网络增加了方向传播的梯度信号。

Inception Net 包括 V1~V4,Inception V2^[52]借鉴了 VGG Net 的思想,使用两个 3×3 的卷积来代替 5×5 的卷积。Batch Normalization(BN)层的主要作用是参数的正则化,提升了网络训练的速度和泛化能力,BN 层的本质其实是一个归一化网络层,可以替代 LPN 层。Inception V3^[53]相较于 InceptionV2 有两方面的改进,第一是引入了将大

尺寸卷积核分解成小卷积核，如将 7×7 的卷积拆成 1×7 和 7×1 的卷积，除了可以减少参数之外，还可以降低模型的过拟合，同时还增加了一层非线性扩展模型表达能力，这种非对称卷积结构拆分，其结果比对称拆分为几个相同的小卷积核效果更好，可以处理更多、更丰富的空间特征；第二，Inception V3 优化了 Inception Module 的结构，包含了 35×35 、 17×17 和 8×8 三种结构。Inception V4^[54] 相对于 Inception V3 主要结合了微软的 ResNet。

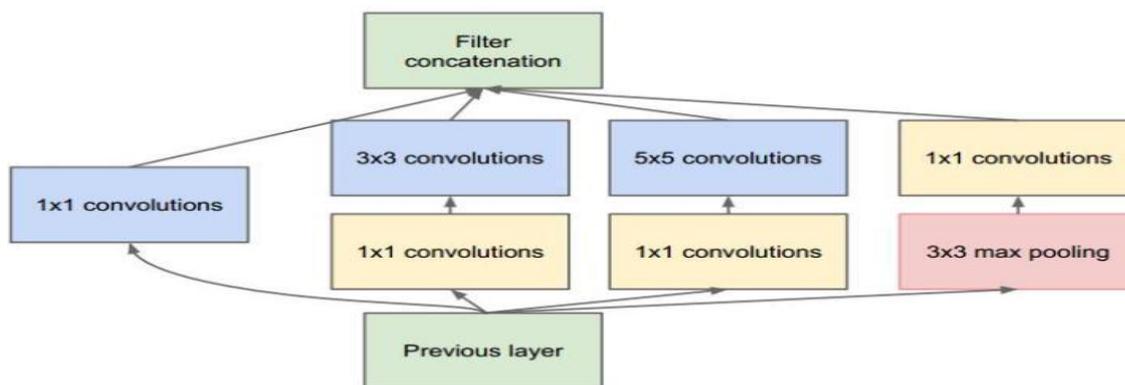


图 4-4 Inception Module 结构

4.5 仪表类别识别实验分析

仪表类别识别是仪表识别整体算法流程的第二步，也是相当关键的一步，如图 4-5 所示，仪表类别识别算法流程为：输入一张仪表定位之后剪裁的仪表图像，送入到训练好的卷积神经分类网络，输出为具体的仪表类别。图 4-4 是使用 Faster R-CNN 定位之后，根据定位框位置剪裁之后得到的仪表图像。

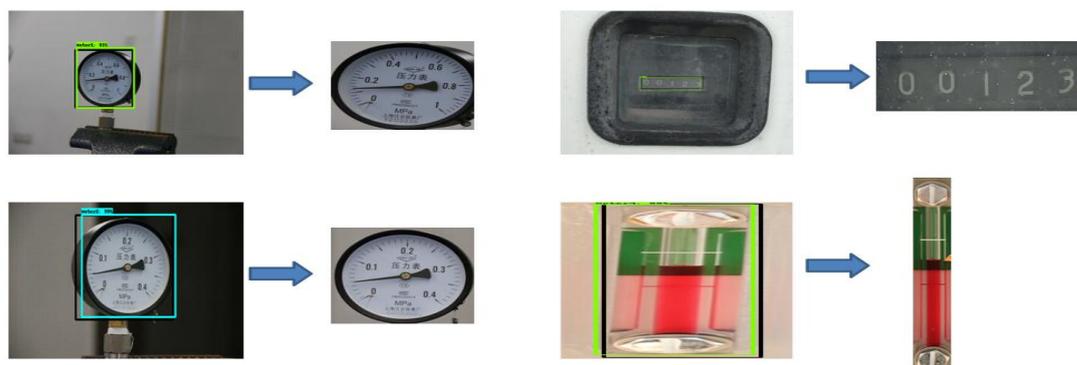


图 4-4 仪表图像剪裁



图 4-5 仪表分类流程图

仪表类别识别实验是基于 Tensorflow、Keras 框架，搭建了四种经典卷积神经网络，分别为 LeNet5、AlexNet、VGG-16 Net、InceptionV4。将四类仪表，3840 张图像以 6:4、7:3、8:2 三种比例分为训练集和验证集进行训练，训练迭代为 100 个 epoch(训练迭代数)。用 TP、FP、FN、ACC、RECALL、F1 六个指标作为模型分类效果好坏的指标，其中，TP 为真正例、FP 为假正例、FN 为假反例、ACC 为准确率、RECALL 为召回率、F1 为 ACC 和 RECALL 的调和指标。

$$ACC = \frac{TP}{TP + FP} \quad (4-1)$$

$$RECALL = \frac{TP}{TP + FN} \quad (4-2)$$

表 4-1 为四种经典分类网络，在三种比例的训练集和验证集下的准确率，可以看出，根据网络复杂度、参数量、准确率等因素，最终 InceptionV4 算法效果最好。图 4-2 是 InceptionV4 训练的准确率图以及测试的 ROC 图。

表 4-1 仪表分类算法结果

	6:4	7:3	8: 2
LeNet5	0.79	0.82	0.89
AlexNet	0.84	0.90	0.92
VGG16	0.89	0.92	0.95
InceptionV4	0.90	0.98	0.99

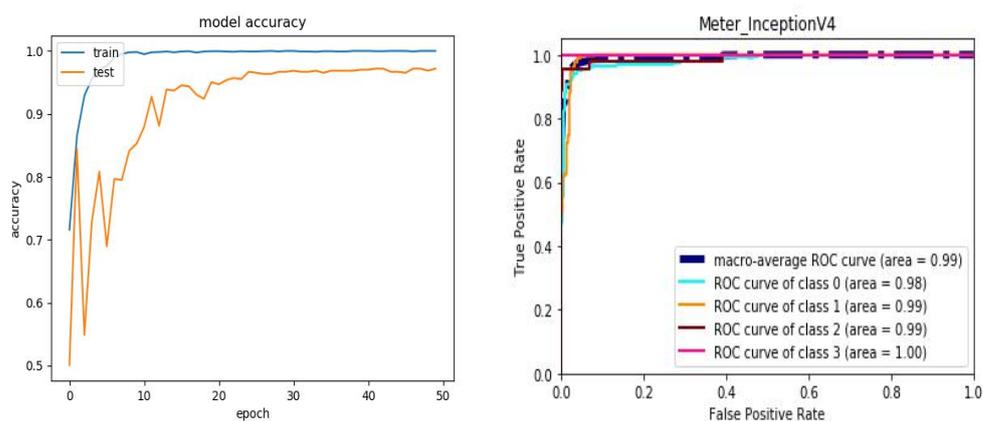


图 4-6 InceptionV4 算法实验结果

4.6 本章小结

本章主要介绍了经典卷积神经网络分类算法的基本原理以及仪表类别识别的算法流程。具体为，将仪表图片输入到训练好的卷积神经网络，输出为仪表的类别。将数据按照 6:4、7:3、8:2 三种不同比例，分别使用 LeNet5、AlexNet、VGG-16、InceptionV4 四种算法进行训练，经过对比和评估，在仪表类别识别部分选择了 InceptionV4 作为分类算法，为后面的仪表示数识别打下了良好的基础。

第 5 章 电厂仪表示数识别

本章开始将介绍仪表识别的最后也是最重要的一部分，仪表示数识别。根据仪表类别识别得到的结果之后，根据不同的类别，分为字符仪表示数识别、两种不同量程的指针式仪表示数识别、油压表示数识别。字符仪表识别部分使用了经典算法和当前主流算法两种算法，经过效果对比，使用 CRNN 算法作为字符仪表示数识别的算法；使用传统图像处理方法对两种不同量程的指针式仪表和油压表进行了示数识别，但是传统图像处理方法具有一定的局限性，因此准确率并不稳定。由于指针式仪表示数识别和油压表示数识别都可以看做是一种求面积比例的问题，故本文提出了一种新型的基于卷积神经网络的示数识别方法，可以对两种仪表同时进行示数识别，经试验证明，该方法准确率高，并且具有一定的泛化性。

5.1 字符仪表示数识别

目前主流的字符识别流程主要为：输入字符图片、二值化操作、滤波降噪、形态学操作、字符分割、字符送入卷积神经网络识别、输出预测的字符结果。其中，卷积神经网络识别所用的网络最经典的是手写数字识别网络--LeNet5 网络，该网络是将大量的 0-9 的手写图片送入网络进行学习，输入为 10 个字符类别，在模型评估的时候准确率高达 95%以上。图 5-1 为使用字符分割的方法来进行字符仪表识别，具体为：将定位之后的字符仪表进行剪裁、将剪裁后的图片进行黑白二值化、使用中值滤波进行降噪、使用开操作进一步消除斑点噪声、根据像素进行字符分割、将分割后的图片送入训练好的 LeNet5 网络进行识别。

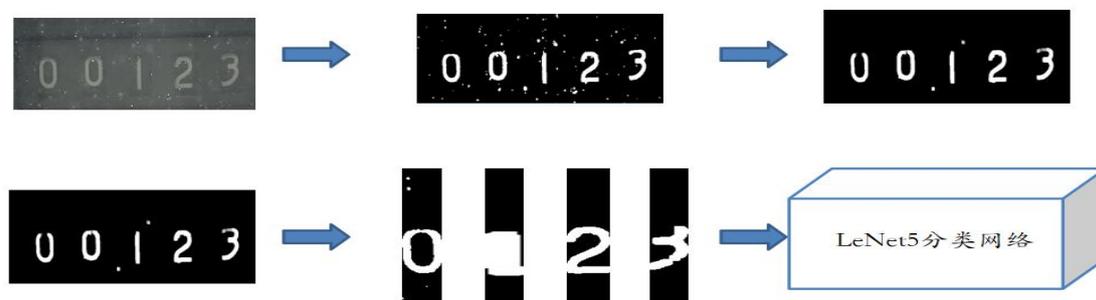


图 5-1 字符仪表示数识别流程图

虽然这种方法具有较高的准确率，但是该方法特别依赖图片预处理和字符分割部

分，一旦字符分割不准确，那么后续的卷积神经网络识别部分就会受到较大的影响，进而准确率降低。

为了解决这一问题，本文使用了 CRNN 网络进行字符仪表识别。CRNN^[55]全称为 Convolutional Recurrent Netural Network，是一种端到端的卷积循环神经网络结构，主要用于端到端的对不定长的文本序列进行识别，不需要图片预处理以及对单个字符进行切割，而是将文本识别转化为时序依赖的序列学习问题，就是基于图像的序列问题。

图 5-2 为 CRNN 结构图，由图可知，整个 CRNN 网络包含三部分，从下到上依次为：CNN(卷积层)，使用深度 CNN 对输入图像提取特征，得到特征图；RNN(循环层)，使用双向的 RNN 对特序列进行预测，对序列中的每个特征向量进行学习，并输出预测标签(真实值)分布；CTC(转录层)，把从循环层获取的一系列标签分布转换成最终的标签序列。其中，CNN 部分采用的 VGG 网络的结构，如图 5-3，为了能将 CNN 提取的特征作为输入，将第三个和第四个最大池化层的核尺度从 2*2 变成了 1*2，为了加速网络的训练，在第五和第六个卷积层后面加上了 BN 层；RNN 部分是对 CNN 输出的特征序列 X，每一个 X 都对应一个输出 Y。为了防止训练时梯度的消失，采用了 LSTM 神经单元作为 RNN 单元。对于序列的预测，序列的前向信息和后向信息都有助于序列的检测，所以采用了双向 RNN 网络；CTC 转录层在测试时，转录分为两种，一种是带字典的，一种是不带字典的。带字典的就是在测试的时候，测试集是有字典的，测试的输出结果计算出所有字典的概率，取最大的即为最终的预测字符串。不带字典的，是指测试集没有给出测试集包含哪些字符串，预测时就选取输出概率最大的作为最终的预测字符串。

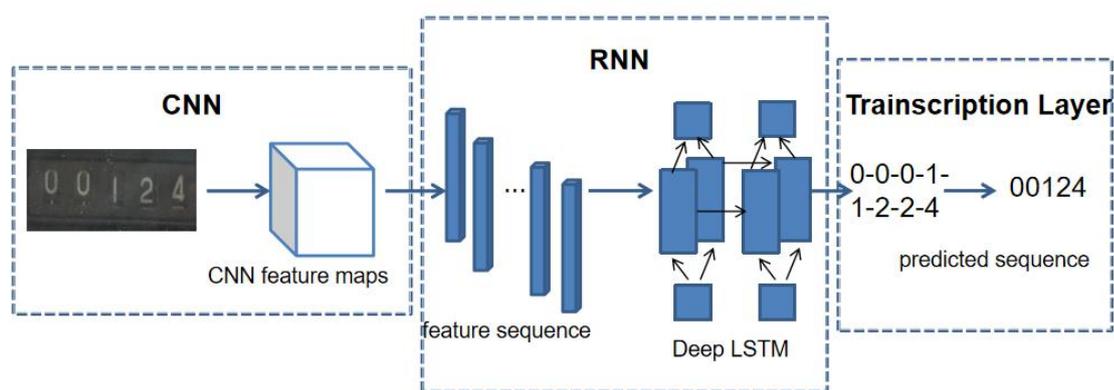


图 5-2 CRNN 网络结构图

Type	Configurations
Transcription	-
Bidirectional-LSTM	#hidden units:256
Bidirectional-LSTM	#hidden units:256
Map-to-Sequence	-
Convolution	#maps:512, k:2 × 2, s:1, p:0
MaxPooling	Window:1 × 2, s:2
BatchNormalization	-
Convolution	#maps:512, k:3 × 3, s:1, p:1
BatchNormalization	-
Convolution	#maps:512, k:3 × 3, s:1, p:1
MaxPooling	Window:1 × 2, s:2
Convolution	#maps:256, k:3 × 3, s:1, p:1
Convolution	#maps:256, k:3 × 3, s:1, p:1
MaxPooling	Window:2 × 2, s:2
Convolution	#maps:128, k:3 × 3, s:1, p:1
MaxPooling	Window:2 × 2, s:2
Convolution	#maps:64, k:3 × 3, s:1, p:1
Input	$W \times 32$ gray-scale image

图 5-3 CRNN 网络结构图

我们随机选取 10 张字符仪表剪裁后的图片，分别使用字符分割识别法和 CRNN 算法进行识别，然后进行效果对比。如图 5-4，9 张字符仪表示数为：00155、00094、00096、00124、00106、00139、00138、00150、00107、00093。由表 5-1 可知，字符分割法在图片质量较差的情况下，平均准确率只有 50%，而 CRNN 在图片质量较差的情况下有 90%的准确率，故在字符仪表识别部分中使用 CRNN 作为识别算法。



图 5-4 字符仪表示例图片

表 5-1 字符分割法和 CRNN 实验结果对比

真实值	字符分割法	CRNN
00155	00055	00155

00094	00004	00094
00096	00096	00096
00124	00124	00124
00106	00100	00106
00139	00130	00139
00138	10138	00138
00150	00150	00150
00107	00107	00007
00093	00096	00093
平均准确率	50%	90%

5.2 基于传统方法的压力表示数识别

压力表示数识别部分分为指针式仪表示数识别、油压表识别。首先来看一下油压表识别，如图 5-5，通过定位、剪裁、分类步骤之后的油压表图片，接下来就是识别示数，也就是识别油位高度部分。油压表比较特殊，其装置简单，管道偏细窄，且不像温度计这一类型的仪表有刻度显示，所以识别起来还是有一定的难度。为此，我们采用了传统图像处理方法进行油位高度识别，主要思想是求出油位高度占整体油管的面积比例，就可以得到具体的油位示数，具体为：输入图像、阈值分割、中值滤波、腐蚀、膨胀、计算黑色像素占整体像素的比例，进而可以得到油压表的示数。该方法虽然有效，但是当图片质量较差、光照不均衡时准确性就会大大降低。

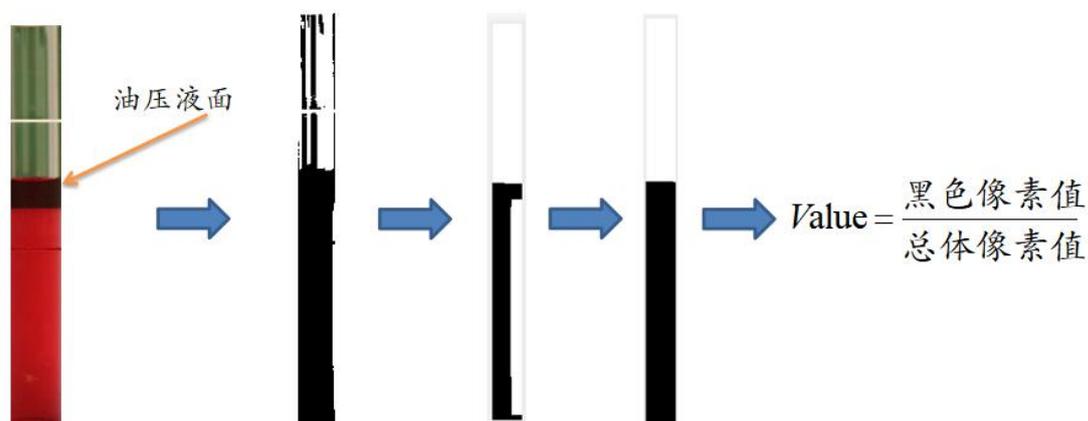


图 5-5 油压表示数识别流程图

然后，指针式仪表示数识别部分同样是在定位、剪裁、分类之后对仪表进行识别。

目前指针式仪表示数识别的方法有很多，最新的方法是基于 Faster R-CNN 和 U-Net 的识别方法。如图 5-6，其具体步骤为，使用 Faster R-CNN 对仪表指针进行定位，然后用 U-Net 提取指针，最后通过角度法计算最终的读数。角度法是基于指针和最大、最小刻度线、水平线之间的夹角以及最大量程之间的线性关系计算出最终的示数。虽然此方法在提取指针时使用 U-Net 代替了传统的霍夫圆和直线的检测，但是在最终读数的时候还是使用的角度法。当仪表图片收到光照不均衡、消畸变不完全、指针位置提取不准确时，角度法的准确性就会受到很大的影响。

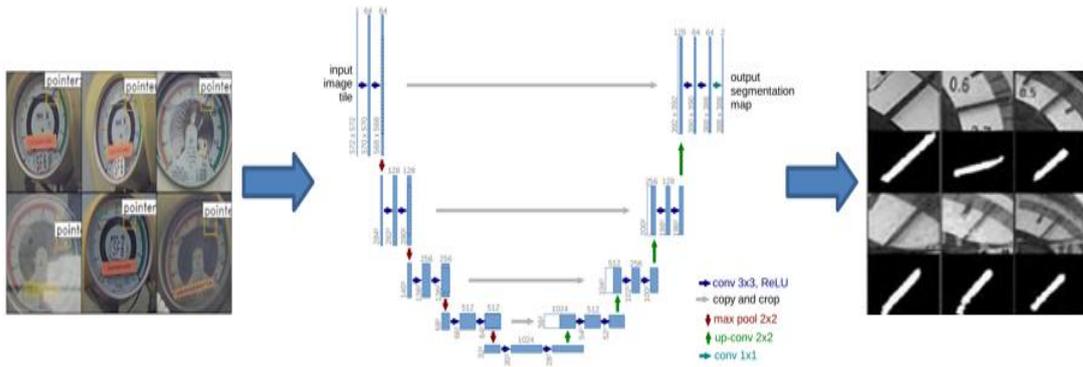


图 5-5 基于 Faster R-CNN 和 U-Net 的指针式仪表示数识别

图 5-6 为本文数据集中的三张指针式仪表图片，为了更加贴近在复杂环境下的仪表图片，我们在原有图片基础上加入了光照不均衡和障碍物遮挡两种效果，然后使用传统图像处理方法对其分别进行示数识别，具体为：输入仪表图像、仪表图像灰度化、仪表图像中值滤波降噪、仪表图像边缘检测、仪表图像霍夫变换圆检测、仪表图像直线检测提取指针、角度法计算仪表示数。图 5-7 为三张仪表图片的预处理结果，绿色圆圈为获取仪表盘的结果，红色直线为提取仪表指针的结果，可以得知在图片发生畸变时，霍夫变换提取表盘的过程会受到很大影响，进而影响之后的识别结果；当仪表受到光照和遮挡影响时，提取指针就会变得十分困难，从而示数识别部分也难以准确。



图 5-6 原图、加光照、加遮挡的指针式仪表图片



图 5-7 使用传统图像处理方对三种不同情况下的指针式仪表进行示数识别

5.3 基于卷积神经网络的压力表示数识别

目前指针式仪表和油压表的示数识别主要还是使用传统图像处理方法进行识别，由于传统图像处理方法具有一定的局限性，对图片的质量、拍摄角度、光照等有较高的要求，导致准确性不稳定。所以，本文提出了一种基于卷积神经网络的压力表示数识别方法，该方法的核心思想是使用卷积神经网络做回归任务，主要为：输入指针式仪表或油压表图片以及对应的示数，将示数作为 label(油压表的示数是人工标定的)，送入卷积神经网络进行学习，把卷积神经网络的最后一层多个神经元输出层改为一个神经元的输出层，输出为卷积神经网络预测的示数值，使用均方误差函数代替交叉熵函数作为损失函数。

如图 5-8，输入图片为指针式仪表，其数值为 0.12，或油压表，其数值为 0.6，将数值作为 label，送入一个含有 4 层卷积层、3 层池化层、2 层全连接层、1 个输出层的神经网络进行学习，输出为对应仪表图片的预测数值。

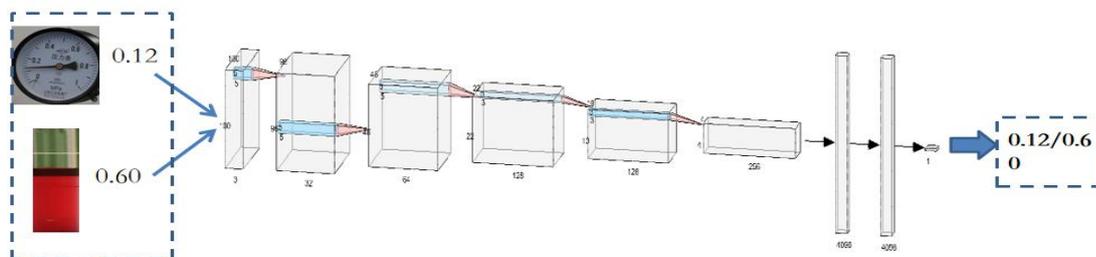


图 5-8 基于卷积神经网络的指针式仪表、油压表示数识别流程图

基于卷积神经网络的压力表示数识别方法具体为，将定位、剪裁、分类之后的指

针式仪表图像和油压表图像尺寸缩放到 100*100*3，送入网络进行学习。C1 为第一个卷积层，用 32 个 5*5 大小的卷积核，以步长为 1 对图片进行卷积操作，得到的特征图大小为 96*96*32，然后使用 2*2 大小的窗口，对得到的特征图进行最大池化操作，特征图变为 48*48*32；C2 为第二个卷积层，用 64 个 5*5 大小的卷积核核，以步长为 1 对 C1 输出的特征图进行卷积操作，得到的特征图大小为 44*44*64，然后使用 2*2 大小的窗口，对得到的特征图进行最大池化操作，特征图变为 22*22*64；C3 为第三个卷积层，用 128 个 3*3 大小的卷积核，以步长为 1 对 C2 输出的特征图进行卷积操作，得到的特征图大小为 20*20*128，然后使用 2*2 大小的窗口，对得到的特征图进行最大池化操作，特征图变为 10*10*128；C4 为第四个卷积层，用 256 个 3*3 大小的卷积核，以步长为 1，对 C3 输出的特征图进行卷积操作，得到的特征图大小为 4*4*256；F5、F6 为全连接层，分别有 4096 个神经元，输出层为一个神经元，为数值的预测值。损失函数为均方误差函数：

$$Loss = \frac{1}{2} \sum_{i=1}^{i=n} (y - f(x))^2 \quad (5-1)$$

其中，y 为仪表图片的真实数值，f(x)为卷积神经网络学习后的预测值。

如图 5-9，为了更好的贴近现实情景，我们在原有的数据集中对每张指针式仪表和油压表进行了数据增广，对每张仪表图片都进行了增加光照和遮挡效果，三类仪表总计 11420 张图片。5-10 是增广后的仪表图片在上述卷积神经网络在训练 100 个 epoch(训练迭代次数)后的损失图，可以看出在 100 个 epoch 后网络的损失接近为 0。

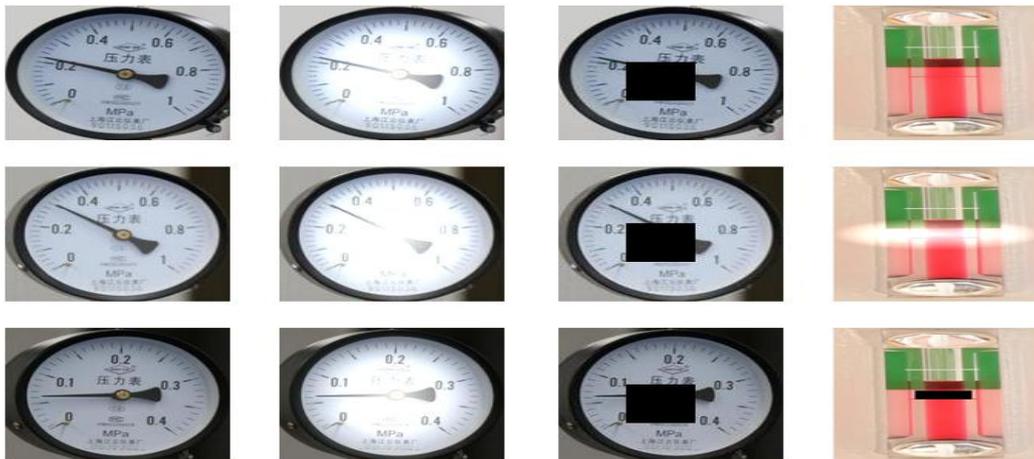


图 5-9 数据增广图片

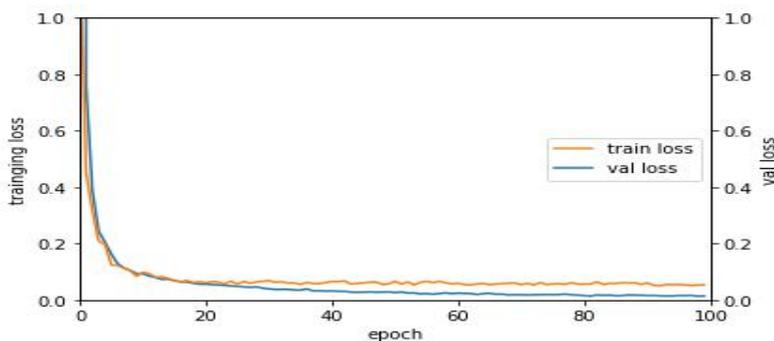


图 5-9 指针式仪表和油压表所用卷积神经网络训练 100 个 epoch 后的损失图

在卷积神经网络模型训练好之后，随机挑选指针式仪表和油压表各 6 张图片进行测试，并与两种仪表的传统识别方法进行了对比，指针式仪表的传统识别方法为角度法，油压表为传统图像处理方法。如表 5-2，经过实验对比，本文提出的基于卷积神经网络的压力表示数识别方法(CNN 回归法)误差在 1%之内，相较于传统方法，CNN 回归法有较好的准确性。

表 5-2 CNN 回归法和角度法、传统图像处理方法的实验对比

指针仪表	角度法	CNN 回归法	油压表	传统方法	CNN 回归法
0.12	0.101245	0.110462	0.55	0.511462	0.543016
0.14	0.111573	0.145531	0.60	0.567824	0.596735
0.16	0.201136	0.159672	0.50	0.482064	0.496587
0.40	0.362314	0.398516	0.65	0.632973	0.649876
0.35	0.331596	0.350365	0.70	0.680421	0.692683
0.52	0.497631	0.519416	0.45	0.427415	0.449637
误差	2%-4%左右	1%之内	误差	2%-4%左右	1%之内

5.4 本章小结

本章主要介绍了字符仪表、指针式仪表、油压表的识别。字符仪表示数识别使用 CRNN 算法进行识别，该方法相较于传统方法有较高的准确性。在指针式仪表和油压表的识别方面，提出了一种基于卷积神经网络的示数识别方法，该方法的主要思路是使用 CNN 网络做回归的方式，将仪表图片、仪表示数作为 label 送入网络训练，通过网络得到示数的预测。经试验证明，该方法有较好的准确性和泛化性，同时减少了传统方法对图片质量的依赖和预处理操作。

第 6 章 工作总结和未来展望

6.1 工作总结

本文的主要工作是针对发电厂的指针式仪表、字符仪表、油压表，使用深度学习的方法对其进行示数识别。主要包括仪表定位、仪表图片剪裁、仪表分类、仪表示数识别四部分。

仪表定位部分使用了主流深度学习目标检测算法 Faster R-CNN、YOLO、SSD 三种算法，经过实验对比，选择了效果最好的 Faster R-CNN 作为仪表定位的算法。

仪表分类部分是对仪表定位之后的仪表图片进行剪裁，送入训练好的卷积神经网络进行类别识别。分类网络使用了目前主流的四种算法，分别是 LeNet5、AlexNet、VGG、InceptionV4，经过实验对比，选择了效果最好的 InceptionV4 作为分类算法。

仪表示数识别部分分为两个部分：字符仪表示数识别和压力表识别，压力表包括指针式仪表和油压表。字符仪表示数识别采用了 CRNN 算法，经过实验对比，该方法比传统方法鲁棒性更好；压力表识别部分，由于传统图像处理方法对指针式仪表和油压表进行读数识别有一定的局限性，因此提出了一种基于卷积神经网络的示数识别方法，该方法的主要思路是使用 CNN 网络做回归的方式，将仪表图片、仪表示数作为 label(标签)同时送入网络训练，通过网络得到示数的预测值。经试验证明，该方法有较好的准确性和泛化性，同时减少了传统方法对图片质量的依赖和预处理操作。经实验证明，本文使用和提出的方法十分有效，各部分准确率都达到了 90%及以上，实现了在复杂环境中指针式仪表的定位、类别判断、示数读取的自动化。

6.2 未来展望

本文提出的一种基于卷积神经网络的新型示数识别方法虽然准确度和创新性较好，但是还是需要依赖仪表定位和仪表分类两个步骤为前提。为了进一步减少算法的复杂度，未来的研究方向是提出一种端到端的网络，即输入一张仪表图片，送入网络，输出为仪表的位置、类别、示数识别等信息。

参考文献

- [1] 孙华东, 汤涌, 马世英. 电力系统稳定的定义与分类述评[J]. 电网技术, 2006(17):35-39.
- [2] 耿宁. 电力设备状态检修实施策略研究[J]. 工程技术(文摘版), 2016:00076-00076.
- [3] Levergood T M , Stewart L C , Morris S J , et al. Internet server access control and monitoring systems[J]. 1998.
- [4] Nils, J, Nilsson. Artificial intelligence: A modern approach: Stuart Russell and Peter Norvig, (Prentice Hall, Englewood Cliffs, NJ, 1995); xxviii + 932 pages[J]. Artificial Intelligence, 1996.
- [5] 房桦, 蒋涛, 李红玉, et al. 一种适用于智能变电站巡检机器人的双针仪表读数的识别算法[J]. 山东电力技术, 2013(03):12-16+72.
- [6] Rihs T A , Michel C M , Thut G . A bias for posterior α -band power suppression versus enhancement during shifting versus maintenance of spatial attention[J]. Neuroimage, 2009, 44(1):190-199.
- [7] Endrenyi, J. The Present Status of Maintenance Strategies and the Impact of Maintenance on Reliability[J]. Power Engineering Review IEEE, 2001, 21(12):68-68.
- [8] Kuntz, Paul, A. A Reliability Centered Optimal Visual Inspection Model for Distributed Feeders.[J]. IEEE Transactions on Power Delivery, 2001.
- [9] M.D. Abràmoff, P.J. Magalhães, Ram S J . Image Processing with Image J[J]. Biophotonics International, 2004, 11(7):36-41.
- [10] Lecun Y , Bengio Y , Hinton G . Deep learning[J]. nature, 2015, 521(7553):436.
- [11] Lowe D G . Object Recognition from Local Scale-Invariant Features[C]// Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on. IEEE, 1999.
- [12] Bay, Herbert, Ess, Andreas, Tuytelaars, Tinne. Speeded-Up Robust Features[J]. computer vision & image understanding, 2008, 110(3):404-417.
- [13] Takeshi Mita, Toshimitsu Kaneko, Osamu Hori. Joint Haar-like Features for Face Detection[J]. Proc.intl.conf.on Computer Vision, 2005, 2:1619-1626 Vol. 2.
- [14] Zhu, Ji, Arbor, Ann, Hastie, Trevor. Multi-class AdaBoost[J]. statistics & its interface, 2006, 2(3):349-360.
- [15] The Nitrogen Cascade. BioOne Online Journals - BioScience - 53(4):341 - Full

Text[J]. bioone.

[16] O'Rourke S M , Herskowitz I , O'Shea E K . Yeast go the whole HOG for the hyperosmotic response[J]. Trends in Genetics, 2002, 18(8):405-412.

[17] Cherkassky V , Ma Y . Practical selection of svm parameters and noise estimation for svm regression[J]. Neural Networks, 2004, 17(1):113-126.

[18] 段瑞玲, 李庆祥, 李玉和. 图像边缘检测方法研究综述[C]// 全国光电技术学术交流会. 2005.

[19] 滕今朝, 邱杰. Fast and Precise Detection of Straight Line with Hough Transform% 利用 Hough 变换实现直线的快速精确检测 [J]. 中国图象图形学报, 2008, 013(002):234-237.

[20] Vincent L . Morphological grayscale reconstruction in image analysis: applications and efficient algorithms[J]. IEEE Transactions on Image Processing, 1993, 2(2):P.176-201.

[21] Wang Y , Chen Q . Image enhancement based on equal area dualistic sub-image histogram equalization method[J]. IEEE Transactions on Consumer Electronics, 1999, 45(1):P.68-75.

[22] Buades A , Coll B , Morel J M . A non-local algorithm for image denoising[C]// Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. IEEE, 2005.

[23] Wells, T P N . Handbook of Image and Video Processing[J]. Physiological Measurement, 2001, 22(1):263-264.

[24] 贺嘉琪. 基于深度学习的指针式仪表示数自动识别的研究与应用[D]. 2019.

[25] Long J , Shelhamer E , Darrell T . Fully Convolutional Networks for Semantic Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(4):640-651.

[26] Chen L C, Papandreou G, Kokkinos I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs[J]. IEEE transactions on pattern analysis and machine intelligence, 2018, 40(4): 834-848.

[27] Li J , Chen Y , Xiao S , et al. Estimation of Affective Level in the Wild with Multiple Memory Networks[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, 2017.

[28] Taigman Y, Yang M, Ranzato M A, et al. Deepface: Closing the gap to human-level performance in face verification[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 1701-1708.

[29] Cao Z, Hidalgo G, Simon T, et al. OpenPose: realtime multi-person 2D pose

- estimation using Part Affinity Fields[J]. arXiv preprint arXiv:1812.08008, 2018.
- [30] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234-241.
- [31] He K, Gkioxari G, Dollár P, et al. Mask r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2961-2969.
- [32] 李航.统计学习方法[M].2012.
- [33] 郭子海, 郭健, 谭志强. 用图象识别概念判读指针式仪表读数的研究[J]. 仪器仪表学报, 1996.
- [34] 石睿, 谢将剑, 赵暄, et al. 基于 SVM 的精密指针式仪表自动读数方法 %Auto-reading Method for Precision Pointer Meter Based on Intersect Cortical Model[J]. 现代制造技术与装备, 2016, 000(011):22-24.
- [35] 邢浩强, 杜志岐, 苏波. Detection and recognition method for pointer-type meter in transformer substation%变电站指针式仪表检测与识别方法[J]. 仪器仪表学报, 2017, 038(011):2813-2821.
- [36] 周杨浩, 刘一帆, 李琛. An automatic reading method for pointer meter%一种自动读取指针式仪表读数的方法[J]. 山东大学学报(工学版), 2019, 049(004):1-7.
- [37] 基于 Faster R-CNN 和 U-Net 的变电站指针式仪表读数自动识别方法[J]. 万吉林, 王慧芳. 电网技术. 2020(01)
- [38] Behaine C , Scharcanski J . A new approach for automatic visual monitoring of analog meter displays[J]. Conference Record IEEE Instrumentation & Measurement Technology Conference, 2012.
- [39] Ocampo-Vega R , Sanchez-Ante G , Falcon-Morales L E , et al. Automatic Construction of Radial-Basis Function Networks Through an Adaptive Partition Algorithm[C]// 8th Mexican Conference on Pattern Recognition. Springer, Cham, 2016.
- [40] Ma Y , Jiang Q . A robust and high-precision automatic reading algorithm of pointer meters based on machine vision[J]. Measurement Science & Technology, 2019, 30(1).
- [41] Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain[J]. Psychological review, 1958, 65(6): 386.
- [42] Girshick R , Donahue J , Darrell T , et al. Rich feature hierarchies for accurate object detection and semantic segmentation[J]. 2013.
- [43] Van d S K E A , Uijlings J R R , Gevers T , et al. [IEEE 2011 IEEE International Conference on Computer Vision (ICCV) - Barcelona, Spain (2011.11.6-2011.11.13)] 2011 International Conference on Computer Vision - Segmentation as selective search for object

recognition[J]. 2011:1879-1886.

[44] Yu, Licheng, Park, Eunbyung, Alexander C Berg, 等 . Proceedings - 2015 IEEE International Conference on Computer Vision, ICCV 2015[J]. 2016.

[45] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[C]//Advances in neural information processing systems. 2015: 91-99.

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

[47] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//European conference on computer vision. Springer, Cham, 2016: 21-37.

[48] Lecun Y , Bottou L . Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11):P.2278-2324.

[49] Krizhevsky A , Sutskever I , Hinton G . ImageNet Classification with Deep Convolutional Neural Networks[C]// NIPS. Curran Associates Inc. 2012.

[50] Simonyan K , Zisserman A . Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014.

[51] 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.

[52] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[J]. arXiv preprint arXiv:1502.03167, 2015.

[53] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2818-2826.

[54] Szegedy C , Ioffe S , Vanhoucke V , et al. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning[J]. 2016.

[55] Shi B , Bai X , Yao C . An End-to-End Trainable Neural Network for Image-Based Sequence Recognition and Its Application to Scene Text Recognition[J]. IEEE transactions on pattern analysis & machine intelligence, 2017, 39(11):2298-2304.

致谢

时光荏苒，岁月如梭，又是一年凤凰花开的季节。不知不觉已经在美丽的汕头大学学习和生活了三年，我的硕士求学之路也即将画上句号。回首过去的种种，我感慨万千，因此我想在这里向所有给予过我帮助的人表达衷心的感谢。

首先，我要感谢我的导师崔岩教授和指导老师范衡教授，感谢两位教授对我的知遇和教导之恩。尤其是我的指导老师范衡教授，他知识渊博，为人和蔼，待徒如子。在学习中，范教授为我们尽可能的提供一个好的实验平台，给我们的科研工作提供了很大的便利。不仅如此，每当我们遇到问题时，范教授总会不厌其烦的为我们讲解，使我们在专业领域中快速成长；在生活中，范教授以身作则，他以严谨的科研态度、无私的奉献精神对我们言传身教，让我们懂得了很多做人做事的道理。

其次，我要感谢我实验室的师兄弟们，无论是科研还是生活他们都给予了我很大的帮助。尤其是邱本章、卞新超、姜涛、袁宇彤四位同学，他们扎实的专业知识、积极向上的态度、乐于助人的品质是我一生学习的榜样；同时，我要感谢我的舍友戴金祥和汪士明同学，感谢他们这三年对我的陪伴，使远在他乡的我感受到了家乡的温暖；我也要感谢我家乡的几位挚友：王子豪、王晨昕、刘志欣、杨洁、刘沫。在我迷茫、无助的时候，他们都会竭尽所能的给予我安慰和鼓励，使我重整旗鼓，怀揣着一颗温暖的心继续上路。

然后，我要感谢我亲爱的父母，感谢你们多年来对我的关心和爱护，正因为你们的付出和支持，我才能走到今天。

最后，我想感谢自己。一路走来，磕磕绊绊，经历了痛苦和磨难，才懂的感恩和付出。希望在以后的日子里，依旧坚持不抛弃自己、不放弃希望，心中谨记，但行好事，莫问前程！

安康
2020年4月

攻读硕士学位期间主要的工作成果

一、论文

[1] Automated Steel Bar Counting and Center Localization with Convolutional Neural Networks. IEEE Transactions on Industrial Informatics, Under Review.
<https://arxiv.org/abs/1906.00891>

二、专利

[1] 一种基于卷积神经网络的变电站指针式仪表的识别方法 申请号：CN201911143610.6 公开号：CN110929723A

[2] 一种手眼触屏机械臂 申请号：CN201911075012.X 公开号：CN110900599A

[3] 一种基于卷积神经网络的电厂电表字符定位和识别方法 申请号：CN201910316734.3 公开号：CN110263790A

[4] 一种基于深度卷积神经网络的钢筋端面识别方法 申请号：CN201811618063.8 公开号：CN109815950A