

**计算机科学技术学院**

**本科生毕业论文**

**基于OCR的身份证要素提取设计与实现**

**班 级： 云计算数据中心运营**

**学 号： 20161104492**

**姓 名： 高尚**

**指导教师： 李艳玲**

**2020 年 3 月 1 日**

摘要 本文针对OCR系统中的身份证扫描图片设计出一套要素信息提取系统，系统基于CNN深度卷积神经网络搭建，利用CTC技术解决中文字符不定长识别问题；扫描证件图像首先经过图片预处理模块得到干净且无噪声的待识别图片，然后将图片送入文字识别模型经过计算得出特征向量，最后由SoftMax函数转为分类概率得到识别结果；此系统针对特定应用场景，

经实验测试，具有较高的识别率。

关键字 OCR、深度卷积神经网络、图像识别

目录

[第一章 绪论 3](#_Toc443)

[1.1 研究背景及意义 4](#_Toc9371)

[1.3 研究目标和内容 4](#_Toc16422)

[1.4 论文结构安排 4](#_Toc14152)

[第二章 相关理论与技术 5](#_Toc24436)

[2.1 图像处理技术 6](#_Toc12969)

[2.1.1 图像分割技术 6](#_Toc14478)

[2.1.2 图像旋转矫正技术 6](#_Toc16447)

[2.1.3 图像匹配技术 6](#_Toc21821)

[2.1.4 图像形态学技术 6](#_Toc27012)

[2.2 深度学习技术与开发框架 7](#_Toc27800)

[2.2.1 深度学习技术 7](#_Toc13971)

[2.2.2 深度学习开发框架 7](#_Toc26155)

[2.3 CNN网络结构 7](#_Toc22844)

[2.4 相关数据集 8](#_Toc15152)

[2.5 本章小结 8](#_Toc25038)

[第三章 图片预处理 8](#_Toc20144)

[3.1 图像滤波和二值化 9](#_Toc21318)

[3.1.1 图像滤波操作 9](#_Toc23429)

[3.1.2 图像二值化 9](#_Toc22964)

[3.2 定位身份证顶点坐标 9](#_Toc20372)

[3.3 去除图片水印 9](#_Toc15644)

[3.4 矫正身份证方向 10](#_Toc26044)

[3.4.1 矫正身份证倒置问题 10](#_Toc15395)

[3.4.2 矫正身份证正反面问题 10](#_Toc7239)

[3.5 本章小结 10](#_Toc5438)

[第四章 文字处理和识别 11](#_Toc21477)

[4.1 获取身份证中文本行信息 11](#_Toc31294)

[4.2 建立字符集 11](#_Toc886)

[4.3 模型识别文本行信息 11](#_Toc18806)

[4.4 纠正模型识别结果 11](#_Toc3886)

[4.5 本章小结 11](#_Toc203)

[第五章 识别结果展示 12](#_Toc1500)

[第六章 总结与展望 13](#_Toc16993)

[6.1 方案总计 13](#_Toc23894)

[6.2 方案展望 13](#_Toc16344)

致谢

参考文献

全文共\_\_\_\_\_\_\_\_\_页\_\_\_\_\_\_\_\_\_\_\_\_字

1. 绪论
   1. 研究背景及意义

文字是人们日常交流使用最多的形式之一，近些年互联网和智能手机的快速发展，使得互联网数据量开始爆发式增长，大量的文字变成了语音和图像，面对形式多样的语音数据和图像数据，计算机固有的处理方式逐渐变得低效，如何让计算机

可以看懂图像甚至理解图像一直是众多学者的研究重点。

Optical Character Recognition(OCR，光学字符识别)是对含有文字内容的图像进行文字提取并数字化的一个过程．传统OCR识别系统主要应用于从扫描文档中提取出文本，它的处理流程主要包括:图像预处理、图像二值化、单字符切割、链接识别结果等。

中文具有类别多、结构复杂的特点，传统OCR识别系统的单字符切割模块难以准确识别汉字，例如，＂从＂误分为＂人　人＂；得益于深度学习在图像领域的快速发力，Convolutional Neural Networks(CNN，深度卷积神经网络)配合使用Connectionist Temporal Classification(CTC)技术可以实现端到端的免字符切割不定长文本识别。

综上所述，本文研究的基于OCR的身份证要素提取设计与实现，结合了CNN卷积神经网络提取图像特征和CTC不定长字符识别的有关技术，具有较高的研究意义。

* 1. 研究目标和内容

本文主要的研究目标是对身份证中各要素信息进行文本识别。传统OCR识别系统对复杂汉字结构识别率低，针对传统技术方案的缺陷，本文利用深度学习技术设计一套高效且识别率高的图文识别系统。

主要研究内容包括：对图片进行消除噪声和去除印章水印等图像预处理方法；使用数据集训练基于AlexNet网络构建的模型；利用公开的全国省市区级行政编码纠正识别结果。

* 1. 论文章节安排

本文的章节安排如下：

1. 绪论。首先介绍基于OCR的身份证要素提取设计与实现的研究背景，然后简单介绍了传统OCR识别系统的处理流程和技术弊端，提出了本文的技术方案和创新点，最后阐述了本文的研究目标和内容。
2. 相关理论与技术。主要介绍了图像预处理技术、深度学习开发框架、深度卷积神经网络的基础理论和相关技术。
3. 图片预处理流程。主要介绍了图片预处理过程中用到的相关方法，并说明图片预处理的重要性。
4. 文本处理和识别．主要介绍切分文本行信息、建立字符集和模型识别的过程，并说明了校正识别结果的方法。
5. 识别结果展示。主要介绍了整个识别系统的设计结构，并构建可视化界面向用户展示识别结果。
6. 总结与展望。回顾了本文的主要工作以及难点，并且对识别系统提出一些改进建议。
7. 相关理论与技术

概括：

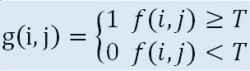
* 1. 图像处理技术

图文识别的成功率一定程度上依赖于图片处理结果的好坏，因此，剔除待处理图片中的噪声数据，凸显有效识别区域是图像预处理的主要目的，好的处理结果可以有效提高识别准确率。目前，常用的图像处理技术包括:图像分割技术、图像旋转矫正技术、图像匹配技术、图像形态学技术。

* + 1. 图像分割技术

图像分割技术的目的在于提取有效识别区域，主要包括:阈值分割、区域分割、边缘分割．本文采用阈值分割．

在阈值分割中，灰度阈值分割是应用次数最广的一类方法，它的处理流程是输入待处理图片，将图片与阈值一一对比，筛选出物体元素与背景元素；可以看出，灰度阈值分割的关键是确定合适的阈值；阈值分为全局阈值和自适应阈值，要分情况合理的使用不同类型的阈值。



* + 1. 图像旋转矫正技术

图像旋转矫正技术的目的在于纠正人为拍摄或处理图片时的误差，经过纠正后的图片更利于模型提取特征，可提高识别结果的准确性，旋转矫正技术的关键在于检测待处理图像中的倾斜方向和角度，主要技术包括:基于投影的办法、基于霍夫变换的办法；本文采用基于投影的办法。

* + 1. 图像匹配技术

图像匹配技术的目的在于寻找待处理图像中的特定元素，进而确定图像中的水印位置、字符位置．主要技术包括:相似性测度匹配方法、模板匹配方法、尺度不变特征变换，本文采用模板匹配方法。

模板匹配方法的处理流程是：在待处理图像中进行平滑移动得到匹配目标，此方法要求匹配目标不可旋转，否则无法找到匹配目标；本文数据集中待处理的印章水印和logo位置相对固定，不会影响匹配结果。

* + 1. 图像形态学技术

图像形态学技术的目的在于剔除待处理图像中的噪声数据，主要技术包括:二值腐蚀和膨胀操作、二值开闭操作；开操作可以使图像轮廓变得光滑，消除图像中的小黑点，并且不会对图像的形状和面积造成明显改变；而闭运算则是相反的过程，会填充图像中的小黑点。



* 1. 深度学习技术与开发框架

概括:

* + 1. 深度学习技术

深度学习是机器学习的分支，由多层感知器构筑起来的深度学习网络可以组合低层特征形成抽象的高层特征，进而识别出数据的分布式特征，实现既定的任务要求。深度学习的原理在于模拟人脑进行分析学习解释数据，它在2012的ImageNet比赛中大放光彩，一举将图像分类的正确率提高了约10%，此后，深度学习被广泛应用在图像和视觉领域，大大超越了传统机器学习所取得成绩。

* + 1. 深度学习开发框架

TensorFlow是由Google主导开发的深度学习框架，是目前主流的深度学习框架之一，它拥有良好的延展性和支持分布式系统，在多GPU和多机上拥有极强的灵活性，随着Google不断更新，现在使用TensorFlow框架的应用可快速部署在各类服务器、PC终端和网页并支持GPU和TPU高性能数据运算。

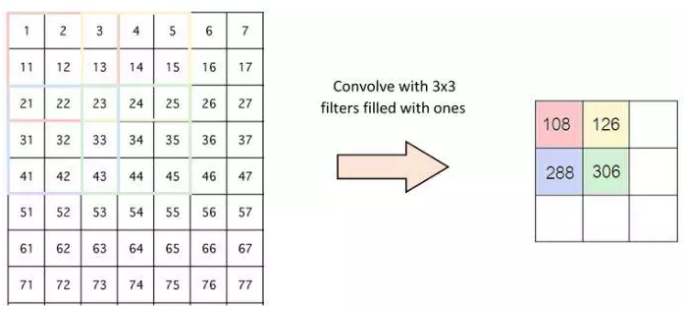
Torch是由Facebook主导开发的深度学习框架，在Python平台上名为PyTorch，它拥有GPU加速的张量计算和包含自动求导系统的深度神经网络，同时它还有入门简单和快速高效的特性受到广大研究人员的喜爱，本文采用PyTorch深度学习框架。

Caffe是由伯克利大学主导开发的深度学习框架，它具有完全开源的特点，同时提供了用于训练、测试等完整工具包帮助开发者快速入门。

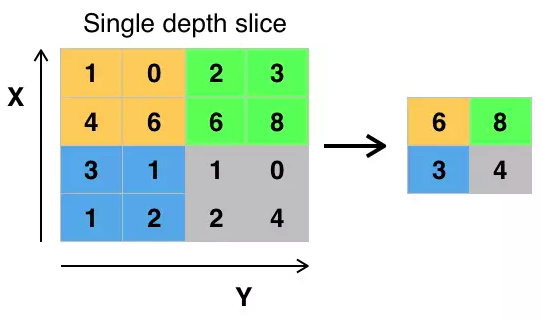
* 1. CNN网络结构

卷积神经网络(Convolutional Neural Network,简称CNN)，是一种前馈型神经网络结构，主要结构包括卷积层、池化层。

卷积层用于提取输入图像的特征，设置不同的卷积核可以提取不同的特征图，通过多种特征图准确了解输入图像的特征，它的基本计算流程是对输入图像的局部矩阵和卷积核矩阵各个位置上的元素相乘，然后相加所有结果得到卷积层结果。



池化层又称采样层，用于特征图降维，减少卷积层的参数数量和压缩数据，提高模型的容错性，主要分为最大池化和平均池化。



* 1. 相关数据集

赛方数据提供10000张图片进行训练，按9:1比例分为训练集和测试集，数据有模糊和重叠等影响

* 1. 本章小结

本章介绍了目前主流的图像预处理技术和深度学习框架，说明了卷积神经网络系统中各层网络的作用，分析了赛方提供的数据集，同时介绍中文合成数据集的情况．

1. 图片预处理

3.1 图像滤波和二值化

3.1.1 图像滤波操作

图像滤波是图片预处理中常用的方法，它需要在不丢失图像细节特征的条件下对目标图像的噪声进行抑制；选择一个可以有效消除噪声的滤波器尤为重要，在选择滤波器时需要考虑两点：能有效地去除背景和背景中的噪声；同时，能很好地保护目标图像的形状、大小和独特的几何结构。滤波操作的合理与否会影响到后续图文识别的有效性和可靠性。通过对数据集的分析，图片有不同程度的模糊现象，直接输入模型会难以提取身份证轮廓信息，因此需预先采用锐化内核对图片进行滤波操作，消除图片中的模糊现象，使后续身份证的轮廓信息更易于提取。

3.1.2 图像二值化

图像二值化是指将灰度图中的像素值设置为0或者255，使整个图片呈现出黑白效果，便于凸显出目标区域的轮廓，同时经过二值化后的图像，其数据量减小，图像变得简单，利于后续的图像处理。

本文二值化方法使用openCV开源库中的adaptive Threshold函数，它是一种局部性的阈值，可以动态自适应地调整属于自己像素点的阈值，相比于全局阈值，会有更好的处理结果；它通过人为设定一个区域，在此区域内比较各像素点与区域里面像素点的平均值的大小关系确定此像素点是属于黑或者白。

3.2 定位身份证顶点坐标

定位身份证顶点的前提是检测身份证的轮廓信息，轮廓检测也是图像预处理中常用的技术，本文轮廓检测方法采用openCV开源库中的findContours函数，因为身份证轮廓为矩形且只需要保存矩形的四顶点坐标，在参数上选择cv2.CHAIN\_APPROX\_SIMPLE压缩水平、垂直、对角线方向的元素，只保留该方向的终点坐标。

对检测到的轮廓进行面积筛选，轮廓面积大小占OCR扫描件图片的0.05~0.5即为身份证轮廓信息，否则为无效信息；接着计算轮廓信息的最小外接矩形得到身份证的中心和顶点坐标．

3.3 去除图片水印

图片中的印章水印是此次图文识别任务的一个难点，水印的位置不固定且有一定几率会遮挡到文字部分，影响后续的识别效果；本文采用的方案是先利用模板匹配定位到水印位置，再对水印区域的灰度值进行线性变换以弱化水印。

3.4 矫正身份证方向

3.4.1 矫正身份证倒置问题

在身份证的左上角标有＂仅限BDCI比赛使用＂的logo，这可以作为判断身份证方向的标志，与定位水印位置原理相同，先提取出logo的模板图像，分别将模板图像与身份证的左上角和右下角进行卷积运算，如果右下角的卷积值大于左上角，说明身份证方向倒置，则进行图像翻转操作。

3.4.2 矫正身份证正反面问题

身份证的正反面信息各不相同，为了准确识别出身份证的要素信息，确定身份证的正反面尤为重要，这关系到后续的文本行信息切分；这里仍需要用到模板匹配的方法，不同的是这次模板图像采用拉普拉斯方法计算得到，因为拉普拉斯方法计算出的图像保留了丰富的形状信息，这可以精确识别到身份证正面或背面的要素信息。

3.5 本章小结

本章介绍了在图片预处理流程中具体用到的方法，以及选择某种方法的原因和实现后的效果，本章所有方法的目的在于得到干净且无噪声的数据，这不仅有利于模型的特征提取，同时可以提高结果的准确性。

1. 文字处理和识别

4.1 获取身份证中文本行信息

在确保身份证正反面都被准确识别的情况下，获取身份证中各要素的位置就变得较为容易，首先计算得到正反面模板图像中各要素的坐标和宽高信息，将正反面图像的尺寸缩放至模板图像尺寸的445\*280，在正反面模板匹配过程中，会得到两个值，一个是最高响应值代表两者的匹配程度，另一个是最高响应值的坐标代表目标图像相对于模板的位置偏差，利用位置偏差可以修正各要素的坐标信息，实现精准切分文本行。

4.2 建立字符集

在汉字中，常用汉字为3000个，为了使识别结果覆盖更为全面，首先获取国标一二级字库作为基础字符集，遍历训练集结果将基础字符集中不存在的汉字增加至字符集中，最终字符集个数为6853，覆盖了常用汉字和部分生僻字。

4.3 模型识别文本行信息

CNN网络结构具有卷积层、池化层，基于AlexNet的网络模型，本文添加了dropout层解决过拟合问题，添加Bottleneck层对特征图进行归一化，提高训练速度，加快模型收敛过程。

模型识别时，需加载训练好的模型权重数据，依次将切分后的文本行图像输入模型中进行计算得出文字识别结果。

4.4 纠正模型识别结果

获取公开的全国省市区级行政代码编号用于校正签发机关的识别结果

4.5 本章小结

本章介绍了文字识别部分的过程，说明了网络结构增加的层以及增加的原因和作用．

1. 识别结果展示
2. 总结与展望

6.1 方案总计

本套技术方案采用CNN+CTC的技术路线，提供了一种端到端的识别方案，旨在解决OCR图像中的身份证信息识别，竞赛方提供的数据集为生成数据，与真实样本存在差异；此方案针对竞赛数据设计实现，故依赖许多强假设，当应用场景发生变化时，需要对方案进行相应的调整。

6.2 方案展望

在图像预处理的去水印模块，可利用数据生成器制造出有水印和无水印的两种身份证照片，将图片输入条件生成对抗(CGAN)网络中进行训练，理论上可以实现更好的去水印效果，提高识别准确率；因个人电脑性能有限和研究时间较短，本文并没有应用此网络．

致谢

大学四年的学习时光是快乐的，也是短暂的，在这四年中，我不仅收获了丰富的专业知识，也得到了班里同学们和老师们的热心帮助，在此，我向他们表示衷心的感谢。

首先，真诚地感谢我的论文指导老师李艳玲教授，感谢她在我做毕业设计期间给我的耐心指导，指引我在深度学习方向一步步成长。

最后，我要感谢我的家人在我做毕业设计期间对我生活上的帮助和精神上的鼓励，使我以积极的心态面对作品中的种种困难。

参考文献

内容

1. 戈嘉宇,刘为嵩.基于深度学习的身份证识别系统的设计与实现[J].电子世界,2020(02):109.

[2]冯海. 基于深度学习的中文OCR算法与系统实现[D].中国科学院大学(中国科学院深圳先进技术研究院),2019.

[3]尹佳利. 自然场景文字检测与识别研究[D].石家庄铁道大学,2019.