中国科学院大学

University of Chinese Academy of Sciences

硕士学位论文

基于文本识别的手写汉字识别平台的设计与实现

作者姓名：董春生

指导教师：于金刚 研究员 中科院沈阳计算技术研究所

刘锦 副教授 中南大学

学位类别：工程硕士

学科专业：计算机技术

培养单位：中国科学院沈阳计算技术研究所

2022 年 6 月

Design and implementation of handwritten Chinese character recognition platform based on text recognition

A thesis submitted to

University of Chinese Academy of Sciences

in partial fulfillment of the requirement

for the degree of

Master of Engineering

in computer technology

By

Dong Chunsheng

Supervisor **Researcher Yu Jingang**

Associate Professor Liu Jin

Shenyang Institute of Computing Technology

Chinese Academy of Sciences

June 2022

摘 要

近年来，随着计算机技术的发展，手机等电子设备的普及使得人机间的交互越来越频繁。在人机交互方向中针对汉字的识别被广泛用于文档转录和文档数字化中。汉字识别在处理文件、邮件分类、商务和其他社会活动中具有巨大的潜力。目前，汉字识别已应用于许多领域，例如：网上阅卷系统、OCR 系统、医院病例的识别等。但是目前主流的汉字识别算法都是把文本检测和文本识别分开进行实现，这样做的效率较低，并且对于手写体汉字的识别算法较少，因此本论文针对于手写体汉字设计了集文本检测与识别于一体的手写体汉字识别平台。

本文以北京邮电大学发布的 HCL2000 数据集和高中生的语文作文图片作为数据源，设计出一个针对于手写体汉字文本图片的检测模块以及识别模块于一体的端到端的模型，将检测和识别统一到一个模型中去，不仅可以减少网络模型参数还可以减少网络模型进行推理的时间。网络模型主要有特征提取、检测与识别模块，其中检测模块和识别模块共用一个卷积神经网络，达到参数共享、减少计算量的目的。该模型通过一次前向传播就可以将文本信息进行定位并识别，与传统的文本识别模型来比较，在进行训练和推理的时候卷积特征的权重不用分开计算，从而节省了前向传播的时间。同时针对手写体汉字引入数据增强的方法，增加模型的泛化能力，并在文本识别部分融入双向长短期记忆网络，根据文本的前后文信息，有效地提高了对手写体汉字识别的准确率。实验结果表明，相比于其他文本识别算法，本文提出的基于文本识别的手写体汉字识别算法相比于其他的文本识别的算法在准确率和速度上均有所提升。

平台通过算法将输入的手写体文本图像识别出来字符，把字符通过前端页面展示出来并保存在后台数据库。同时可以批量上传图片，进行批量识别。本文通过对用户的需求进行了分析归纳，据此对平台的整体架构和数据库进行了总体设计，并完成了登录模块、上传模块、算法模型、显示模块、数据平台等模块的研发，并且经过严格的功能测试，本系统功能满足应用需求，性能测试达到了预期。

关键词：文本识别，卷积神经网络，数据增强，双向长短期记忆网络

Abstract

In recent years, with the development of computer technology and the popularization of electronic devices such as mobile phones, the interaction between humans and machines has become more and more frequent. The recognition of Chinese characters in the direction of human-computer interaction is widely used in document transcription and document digitization. Chinese character recognition is widely used in document transcription and document digitization. Chinese character recognition has great potential in document processing, mail sorting, business and other social activities. At present, Chinese character recognition has been applied in many fields, such as: on-line marking system, OCR system, hospital disease list recognition and so on. However, the current mainstream Chinese character recognition algorithm is to separate text detection and text recognition, which is low efficiency, and there are few recognition algorithms for handwritten Chinese characters. Therefore, this paper designs a handwritten Chinese character recognition platform integrating text detection and recognition.

Based on the Beijing university of posts and telecommunications publishing HCL2000 image data sets and high school students of Chinese composition as a data source, and design a for handwritten Chinese character text image detection and recognition module in an end-to-end model, the detection and identification of unified into one model, can save network model parameters and to shorten the calculation time. The main feature extraction, detection and recognition of the network model, in which the detection module and recognition module share a convolutional neural network, to achieve the purpose of parameter sharing and reduce the amount of calculation. This model can locate and identify the text information through a single forward propagation. Compared with the traditional text recognition model, the weights of the convolution features need not be calculated separately during training and reasoning, thus saving the time of forward propagation. At the same time, the method of data enhancement is introduced for handwritten Chinese characters to

increase the generalization ability of the model, and the bidirectional long and short-term memory network is integrated into the text recognition part, which effectively improves the accuracy of handwritten Chinese characters recognition according to the information of the text. Experimental results show that compared with other text recognition algorithms, the handwritten Chinese character recognition algorithm proposed in this paper has improved accuracy and speed compared with other text recognition algorithms.

The platform identifies characters from the input handwritten text image through algorithm, and displays the characters through the front page and saves them in the background database. At the same time, you can upload pictures in batches for batch identification. This article through to the user's requirements is analyzed and summarized, on the basis of the overall architecture of platform and the database has carried on the overall design, and completed the login module, the upload module, algorithm model and display module, data platform and module of research and development, and the function of the strict test, to achieve the expected performance test.

Key words: Text recognition, Convolutional neural network, Data Augmentation, Bidirectional long and short-term memory network,

目 录

图表目录

第 1 章 绪论

1.1 研究背景及意义

当今，随着电子产品在人们生活中的普及以及科学技术的快速发展，电脑、手机等新的科技设备走进家家户户，成为人们生活的必需品之一。由于人们对计算机功能需求的提高，越来越多的信息需要计算机处理。手机等电子设备的普及使得人和机器间的交互也越来越频繁，并且在很多领域都有涉及，比如通信、学习、娱乐以及生活等方面。但是对于人机交互领域，针对于汉字的识别方向是人机交互的重要技术，如果计算机可以快速且准确的识别出来汉字那么在人机交互领域就会有所创新，但是目前为止计算机针对于汉字的识别，但这远远不能满足人们的要求。智能交互的使用提高了人们对汉字识别的要求。

文字的产生对于人类几千年文明发展有着里程碑的意义。文字的出现使得人类能够记录生活中的信息以及传播信息，并且文字是无法替代的。它是人类生活中日常信息的载体，还可以克服人类大脑的有限的记忆力的短板，使得信息能够一直保存下去。人类可以通过书写、记录各种信息，并且在全国各地之间互相传播、交流。随着科学技术的蓬勃发展，我们可以在信息处理中使用电子文本信息，这样做的话可以提高效率。如今的电子设备大多都可以识别出纸质的文字，把图片上面的文字给识别出来字符版本，但是对于手写体汉字的识别有很多难点，原因在于手写体汉字的书写不一致以及有些字符比较接近，因此对于手写体汉字的识别并不是很理想，而且对于手写体汉字的识别仍然是人工智能领域的一大难题。

手写汉字识别技术根据是否在线书写可以分为两类，第一类是基于在线手写汉字识别，第二类是基于离线手写汉字的识别。目前，针对于手写体汉字的识别技术正处于逐渐稳步上升的过程，并逐步向实用化阶段迈进。自从 21 世纪 60 年代以来，国内外的大量学者针对手写体汉字的识别做了大量开创性的工作。针对于汉字的识别分为两大类，第一类是针对于手写体的汉字识别，第二类是针对于印刷体汉字的识别。尽管手写汉字识别的基本思想与印刷汉字识别是一致的，但从技术上讲，对于手写体汉字来说，它的识别难度要比印刷体汉字的识别难度

要大很多。由于手写汉字的字体差异很大，因人而异，各种成熟的印刷汉字识别方法无法使用。目前针对于数字的识别算法相对比较成熟，因为数字的类别很少并且区分性强，但是针对于汉字识别的算法并不是很完善，特别是对于手写汉字来说，识别准确率并不高。

1.2 研究发展现状与存在问题

1.2.1 传统手写体汉字识别技术

手写体字符识别已经经过了许多年的发展与技术突破，已经形成了一套很成熟的技术体系。识别流程为：输入汉字文本图像、对手写体汉字文本图像进行图像的预处理（降噪、平滑、矫正等）、对预处理后的文本图像进行特征提取、对特征进行压缩以及转换、根据特征训练分类器、分类器进行分类。如下图 1.1 所示：

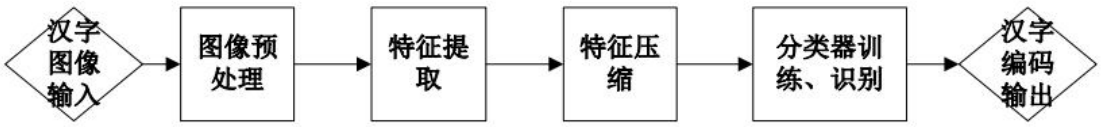

图 1.1 传统手写体汉字识别流程

Figure 1.1 Traditional handwritten Chinese character recognition process

1. 图像预处理：由于书写、环境、拍照等因素，拍摄出来的汉字文本图像经过设备输入到识别设备中，图像中会有各种各样的噪声，例如：光照因素造成的拍摄图片过白、图片上面有各种各样的黑点、书写不规范造成的平滑断笔、拍摄角度不同造成的图片倾斜等。并且输入图片一般是 RGB 图片，即彩色图片，但是针对于图片上面的汉字识别，我们仅仅关注的是灰度图，彩色图片虽然富含丰富的色彩信息，但同时也会伴随着计算量的提升，在不考虑色彩信息的情况下，需要先把彩色图片进行图像处理得到单通道的灰度图，从而减少计算量，提高识别速度。将需要处理的图像进行规范化是图像预处理最重要的一步，也是第一步，一般都是先将图像进行归一话处理之后在进行后面的操作。归一化的目的是：消除因为个人书写字迹差别大而带来的字体不均衡分布、避免线性归一化带来的不

良影响。

2. 特征提取：每一类物体都有自己独有的特征，根据特征这一特点可以把一类对象区别于另一类对象，计算机可以通过不同的特征对物体进行分类。对于图像来说，每一幅图都有属于其本身的特征，如：边缘特征、纹理特征、直方图特征等。对于文本图片的特征提取来说，主要作用就是提取该图片中所有的文本信息，这些信息可以区分不同类别的文字，从而达到对不同字符进行分类的效果。

3. 特征压缩：特征压缩的主要目的就是把第 2 步所提取到的文本特征进行降维，即减少特征的维度。因为特征包含的向量维度是非常大的，不利于分类器对这些高维特征进行训练，因此要把高维特征信息降维到低维度，之后在进行训练。

4. 分类器的训练：分类器的作用就是对于用户输入的文本图像进行识别，之后把汉字特征分类为不同的字符。分类器想要十分准确的对输入特征进行识别分类，就要经过大量的训练，也就是从大量的样本中进行学习，把所有的训练样本所对应的特征空间中获得每个特征的概率分布。分类器是整个汉字文本图像识别的核心，常见的分类器算法有 SVM（支持向量机）、HMM（隐马尔科夫模型）、Logistic（逻辑回归）、Naive Bayes（朴素贝叶斯）等。

传统的汉字识别是基于单字的识别分类，因此需要先对输入的汉字文本图像进行逐个字符的提取。提取出单个字符之后在进行识别分类，不能做到直接对文本图像进行字符识别，这就造成了计算资源的浪费以及识别速度过慢。

1.2.2 文本识别技术研究现状

自从 2006 年开始，深度学习的研究在学术界引起了巨大的轰动，它是含有多层隐藏层、多感知机的一种网络结构，能够更加深层、抽象的描述物体的特征以及属性。Krizhevsky 等（2012）于年发表的 AlexNet 网络对 ImageNet 图像分类任务上取得突出贡献，使用基于深度学习的特征提取无论是在效率上面还是在效果上都要好于传统的方法。在图像的识别方向上，卷积神经网络相对于其他的神经网络有着很好的性能，其中卷积神经网络中的卷积层对于图像的提取特征有着巨大的优点，卷积层在提取特征的过程中不会丢失空间信息、也不会造成过多的参数，因次卷积神经网络已经在图像的识别、分类和检测任务中有了一些列的

突破，并被广泛应用。

Zhi Tian,Weilin Huang 等（2016）年提出的 CTPN 网络模型，该网络模型的创新点在于作者把双向长短期记忆网络（BLSTM）引入到卷积神经网络模型中，该论文的亮点结合了 CNN（卷积神经网络）与 BLSTM，这样做的目的是为了增加文本的上下文特征信息。能够快速有效的检测出文本信息，属于文本检测在深度学习领域中的开篇之作。该网络的特点在网络上进行了改进，引入双向 LSTM 可以结合文本的上下文信息去对文字进行识别，避免只识别一个字而带来的准确率的提升。该网络还采用了一组不同高宽比的 Anchors，用来对文字位置的定位。Xinyu Zhou, Cong Yao 等（2017）年提出的 EAST 网络模型，该模型是一种端到端的快速有效的文本检测方法，消除了候选框的区域聚合、文本分词、后处理等内容。可以对文本行进行直接预测，即可以检测单词级别的文本，也可以检测文本行级别的，也就是不用先对文本图片中的每一个汉字进行单独分割了，而是可以快速的对整行文本进行检测。该模型训练的时候采用了加权损失函数，即分别采用交叉熵损失，和 Iou 损失，最终把二者的损失值加以不同的权重。针对 EAST 网络模型来说，对于长文本的检测是比较困难的，也就是说直接把整个文本图片作为输入，来对它上面的字符进行检测识别是较为困难的，针对这一难点，Baoguang Shi, Xiang Bai 等（2017）年提出了 SegLink 网络模型，该网络模型是在基于目标检测进行改进，使得改进后的网络模型对于文本图片的检测更加适用。该网络模型的改进是将每个字符分割成更容易检测的并且带有方向性的小字符块，之后在将各个小字符块拼接成字符。虽然该网络模型是基于英文单词来设计的，但是他提出的网络模型的设计思路以及训练时候的小技巧是非常值得借鉴的。Baoguang Shi, Xiang Bai 等人发表了 TextBoxes 文本检测网络，该网络模型主要是关于自然场景中（街道、广告牌、路牌等）的文本检测，该网络模型修改了锚框长和宽的比例，修改成了长条形状。并且修改了分类器中的卷积核的大小为 1*5，而在 SSD 网络结构中卷积核的大小为 3*3，这样做有助于文本的检测，因为对于文本来说属于长条形状的文本居多。关于自然场景中（街道、场景文字、路牌等）的文本图片检测也是目前研究的热点，因此也诞生了许多优秀的文本检测模型。

文本识别现在主要分为两大分支，一种是基于 CNN、RNN、CTC Loss 的网络模型，另一种是基于 CNN、Seq2Seq 和 Attention 的网络模型。Weilin Huang 等（2016）年发表的 CRNN 网络，作者认为文本识别和普通的物体识别是不一样的，文本中的信息上下文是有联系的，这里作者采用的办法是引入了时间序列模型，分别用卷积网络和循环网络对空间和时间两个维度提取特征，最后进行预测。Zbigniew Wojna 等人（2017）年提出 Attention-based 网络模型主要分为两个部分，分别为矫正模型以及识别模型。矫正模型主要是处理弯曲、透视等形状不友好的文本，识别模型采用了注意力机制。商汤科技和深圳先进院乔宇老师等合作发表的 FOST 网络模型是端到端的文本识别模型，在模型前向传播的过程中检测了文本的候选框的同时还对文本进行了识别，这样做一方面是为了节约时间，另一方面模型会比两阶段模型学习到更多的特征信息。

1.2.3 手写汉字识别难点

目前针对于文本检测以及文本识别的方法和技术已经非常成熟了，但是大多数算法模型都是只适用于印刷体或者是英文字符的识别，并不适合用于手写体汉字的识别。因为汉字的数量是非常大的，若采用基于汉字字体的方法来对汉字进行识别，那么就要对这些汉字一一制定规则，工作量十分巨大。如果采用传统的识别方法，第一步需要先把整篇文本图像中的所有汉字进行单独的分割出来之后再单个进行识别，显而易见这是十分费时的策略，并且十分浪费计算机资源，只适用于对单字的识别，并不适用于识别整篇文本。

相比之下，采用基于深度学习的方法对手写体汉字的识别技术更加适用。利用卷积神经网络的特征提取能力可以提取更加丰富的特征信息，有助于手写体的识别。针对于手写体汉字识别的难点在于以下方面：

1. 汉字的数量不同于英文单词，数量十分庞大，常用的汉字就达到上千个之多，因此对于汉字的分类就相当于进行一个规模巨大的分类问题，可想而知构造出能够分类如此多的类别的算法模型是十分困难的。

2. 汉字中有许多相似的字，如: 土和士、日和曰、天和夭等。这些图形字符的相似性很难控制。为了区分这些汉字，需要提取汉字的细节特征，手写汉字的变量因素很多。在预处理中，需要消除这些噪声干扰，但很难保持汉字的细节。

3. 汉字结构复杂。虽然基本笔画可分为横、纵、撇、钉、点、折，但汉字的结构并不像字母那样以排列组合的形式表示词义。它是通过穿插和组合基本笔画的二维空间而形成的，因此结构特征是汉字识别中的关键特征。汉字的合字包括单字和合字。

4. 手写字形样式因人而异。由于书写状态和不同的书写风格，也会呈现各种风格的字形。特别是在行书中，草书会出现笔迹粘连，对字符分割和识别有很大影响。

1.3 研究目标和研究内容

手写汉字识别平台的主要任务是将强大的深度学习技术与传统的文本识别技术相结合，开发实用的手写汉字文本识别平台，它能对用户上传的文本图片进行识别，并且可以输出文本图片包含的文本信息，再把识别出来的文本信息反馈给用户。具体可划分为以下内容：

1. 识别模型的设计：这是本论文的核心内容，功能是对手写体汉字的文本图像进行识别，采用深度学习的技术，搭建网络模型。设计出一个针对于手写体汉字文本图片的检测模块以及识别模块于一体的模型结构，将检测和识别统一到一个模型中去，可以大量节省网络模型参数并缩短计算时间。网络模型主要特征提取、检测与识别，其中检测模块和识别模块共用一个卷积神经网络，达到参数共享、减少计算量的目的。该模型通过一次前向传播就可以将文本信息进行定位并识别，与传统的文本识别模型来比较，在进行训练和推理的时候卷积特征的权重不用分开计算，从而节省了前向传播的时间。

2. 模型的创新点在于：

1）集检测与识别于一体的端到端的文本识别模型。

2）网络模型基于目前比较常用的模型进行改良，准确度有所提高。

3. 平台的搭建：平台的识别步骤为：用户上传文本图片->对文本图片进行存储->图片预处理->识别图片->对识别的结果进行存储并反馈给用户。

1)数据存储：主要作用是把用户上传的文本图片识别出来的文本信息进行存储，方便用户查询历史识别记录。

2)数据预处理：对用户上传的图片进行预处理，先把图片归一化。由于书写、环境、拍照等因素，拍摄出来的汉字文本图像经过设备输入到识别设备中， 图像中会有各种各样的噪声，例如：光照因素造成的拍摄图片过白、图片上面有各种各样的黑点、书写不规范造成的平滑断笔、拍摄角度不同造成的图片倾斜等。并且输入图片一般是 RGB 图片，即彩色图片，但是针对于图片上面的汉字识别，我们仅仅关注的是灰度图，彩色图片虽然富含丰富的色彩信息， 但同时也会伴随着计算量的提示，在不考虑色彩信息的情况下， 需要先把彩色图片进行图像处理得到单通道的灰度图，从而减少计算量，提高识别速度。

1.4 研究方法

本论文采用深度学习的方法对手写体汉字进行识别，并且基于中小学生中文作文搭建手写汉字识别平台。首先通过一定手段进行数据采集，可以从网上寻找一些开源的数据集，或者自己制作一些数据集，并对其进行标注。再通过对手写字符特点和对设计不同的网络模型，通过消融实验来查看哪些变量可以提高识别精度，后续拟采用的研究方法有：

1. 文献研究法：

通过相关文献调研可实施的方法。查阅现有的相关文献资料，如期刊，研究报告，书籍等，最为主要的两个调研方向是对于文本数据图像预处理方式以及文本检测的研究现状发展趋势和文本识别的构建与发展情况。

2. 对比分析法：

采用消融研究的方法，即控制变量法。改变网络模型中的一个或多个参数，从而从新训练网络模型，看哪个参数对预测的结果的权重更大。删除一些模块，或用随机的模块替换一些训练有素的功能而不会降低性能。消除研究过程中的噪音，进行消融研究。

3. 实验研究法：

对于模型的准确性优调以及不断的改进模型架构，针对于不同人写的汉字来进行识别，并针对结果来调整模型，使网络模型针对不同的样本都能达到很好的识别效果，增加网络的泛化能力。

1.5 本文结构组织

本文主要由六个章节组成，包括关于手写体汉字的背景调研、研究现状以及关键技术、深度学习的基本技术、基于手写体汉字文本识别的算法研究与试验、手写体汉字文本识别平台的设计与实现以及平台的各项功能的测试都做了较为详细的论述，以下从各个章节进行简易概述：

第一章为绪论部分，从手写汉字识别的背景以及技术现状来分析，从传统汉字识别到基于深度学习的手写汉字文本识别发展以及各个时间段的技术。之后是关于论文的研究目的以及各个章节的介绍。

第二章是关于深度学习的基本知识，以及深度学习在图像识别上面的应用，主要为前期的准备工作，为第三章网络模型的搭建提供技术支持。本章主要讲解了 CNN（卷积神经网络）、网络模型的搭建、网络模型的训练等。

第三章是关于手写体汉字识别的网络模型的搭建，本章主要描述了数据集的制作、关于数据图片怎么进行预处理、采取数据增强的方法、手写汉字识别算法的原理以及训练和测试效果的对比。主要从数据集中提取特征，然后利用神经网络模型对文本图片的特征信息进行学习。训练完成的网络模型就可以直接对新的文本图片进行预测，也是整篇论文的核心部分。

第四章的内容是对于平台的设计与搭建，该平台是基于中小学生的手写中文作文识别进行设计的。在手写汉字识别算法完成的基础下，着重对于平台的各种功能需求的分析与设计。从该平台的需求分析开始，研究平台都需要哪些功能，并且还要注重平台的安全以及稳定性。分析完需求之后，又详细的介绍了该平台的各种模块的设计，例如用户登录、上传作文图片、对作文图片进行汉字识别、查询历史识别记录等模块。

第五章是关于平台的各项功能的测试，测试目的有三点，一是为了检测平台是否按照功能需求进行设计，是否有遗漏的功能没有设计。二是为了测试平台的安全性以及稳定性。三是为了测试该平台是否能够给到用户一个良好的体验。针对这三个目的，结合了大量的用户案例来对该平台进行测试，测试结果表明，基于文本识别的手写汉字识别平台各项功能都完整并且安全稳定性能良好。

第六章是总结与展望，本章主要概述了本论文从初期的准备工作到中间的算

法实现工作以及最后的平台设计工作，对整体的流程进行了概述，以及手写体汉字识别平台的不足之处和后续可扩展的功能做了进一步的探讨。

第 2 章　相关理论与技术研究

2.1　深度学习的理论研究

2.1.1　深度学习的基本概念

深度学习是机器学习领域下的一个分支，它是一种尝试使用包含复杂结构或多个非线性变换的多个处理层来高层抽象数据的算法。近年来，有监督的深度学习方法 (用反馈算法训练 CNN、LSTM 等) 取得了前所未有的成功，而半监督或无监督的方法 (如 DBM、DBN、stacked autoencoder) 在深度学习的兴起中起到了重要作用，但他们仍处于研究阶段，并取得了良好的进展。未来，无监督学习将是深度学习的重要研究方向，因为大多数人类和动物学习都是无监督的。我们可以通过观察发现世界的结构，而不是事先被告知所有物体的名称。深度学习是通过分层学习实现多个抽象表示的图。输入层可以直接从图中观察到。第一层可以学习相对简单的特征 (例如边缘)。第二层网络可以在第一层网络中学习到的特征的基础上学习稍微复杂的特征，例如角度或曲率。第三层可以学习对象的复杂特征，从而识别输入对象。

卷积神经网络 (CNN) 不仅包括深度卷积网络 (如 AlexNet) 来执行目标识别任务 (2012 Imagenet champion)，还包括许多用于处理任务 (如目标检测、语义分割和超分辨率) 的优秀模型。他们以不同的方式将卷积过程应用于不同的任务，并在这些任务上产生非常好的结果。基本上，与常规的全连接网络相比，卷积具有许多出色的性能。例如，它只与前一层的神经元产生部分连接。相同的卷积内核可以在输入张量上重用。也就是说，卷积层可以在输入特征图的每一层上面都来重复检测输入图的局部特征。这是卷积网络的一个非常好的属性，它大大减少了两层之间的参数数量。卷积层加上池化层可以达到平移不变性的优点，即无论怎么平移特征图都不会影响最终的结果。

2.1.2　全连接层和卷积层的对比

在卷积神经网络出现之前，对于图像的特征提取主要是使用全连接层，但是全连接层有两个缺点：

（1）全连接层中的参数量是非常大的，导致成本很高，效率很低。

随着卷积神经网络的引入，首先解决问题的方法是 “简化复杂问题”，将大量的参数减少为少量的参数，然后进行处理。最重要的是: 在大多数情况下，降低维度不会影响结果。例如，如果 1000*1000 分辨率图片降低到 200*200 分辨率，不会影响肉眼识别图片是猫还是狗，机器也是如此。因此，卷积神经网络在较大范围内减小参数时，不会导致图像识别的准确性下降。

（2）全连接层在提取图像特征的时候会打乱图像原有的空间结构，导致图像识别的准确率不高，如下图 2.1 所示：

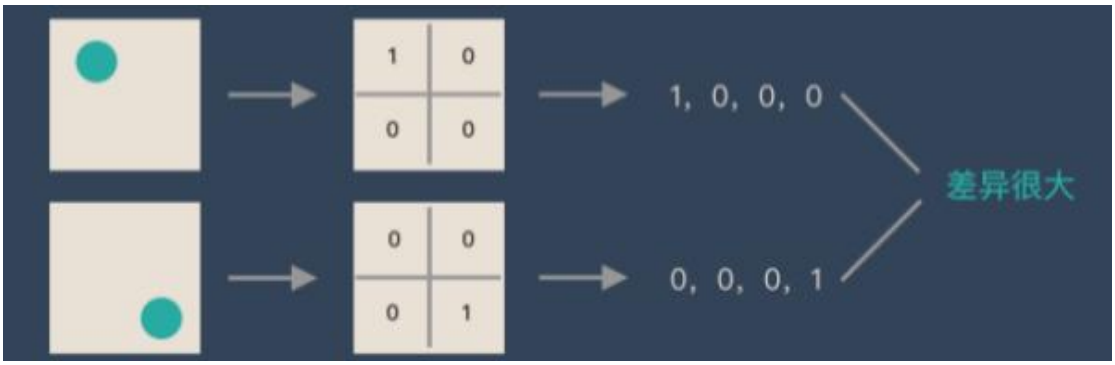

图 2.1　不同图像的空间结构

Figure 2.1　Spatial Structure of Different Images

如果一个圆是数字 1，没有圆的话是数字 0，那么圆的不同位置会产生完全不同的数据表达式。但是如果从视觉上看，图像的内容 (本质) 并没有改变，只是位置发生了变化。因此，当我们移动图像中的对象时，以传统方式获得的参数会有很大差异，不符合图像处理的要求。卷积神经网络已经解决了这个问题。他以视觉般的方式保留了图像的特征。如果一个图像被翻转、旋转或改变时，它也可以有效地识别相似的图像，这一点也是卷积神经网络的三个特性之一的平移不变性。

2.2　卷积神经网络及基本组件的理论研究

卷积神经网络与常规的神经网络的最大不同之处在于神经元的规格不一样。对于全连接神经网络来说网络之间的每一层都和上一层的神经元节点全部链接，

而卷积神经网络之和上一层的神经元节点是局部连接的。常规的神经网络如下图 2.2 所示：

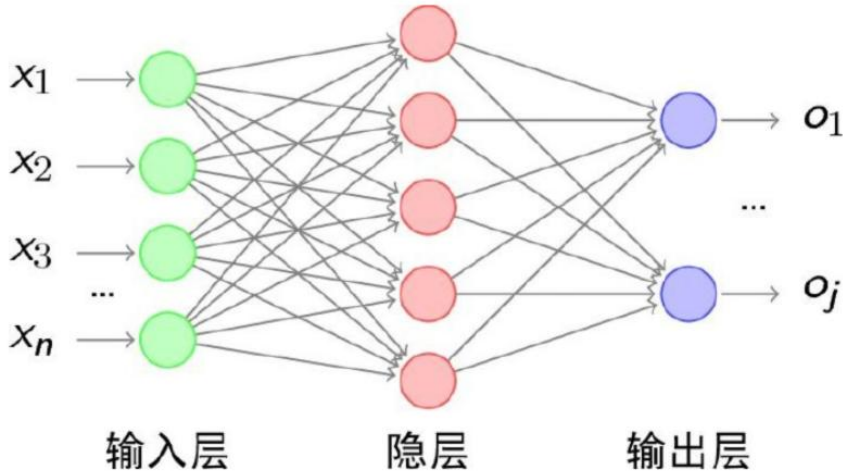

图 2.2　全连接层神经网络

Figure 2.2　Fully Connected Layer Neural Network

该神经网络为三层神经网络，其中包含了输入层（输入层就是该网络模型的输入），隐层（其中隐层可以是单层也可以是多层，里面包含着大量的参数，只要目的就是提取输入层输入的特征，其中隐层和上一层的神经元以及下一层的神经元节点全部链接），输出层（输出层是网络模型的最后一层，经过输出层之后就是对输入的直接输出了，目的就是把输入的东西进行分类）。可以看出常规神经网络中输入层和隐藏层直接的所有神经元节点都是链接在一起的，因此也叫全连接层，这样的做的好处是一般放在网络的最后，可以综合所有信息做分类，但是最大的缺点就是参数量十分庞大，效率不高。对于卷积神经网络来说，中间层的参数相相比于常规的神经网络的参数就少的多了。卷积神经网络如下图 2.3 所示：

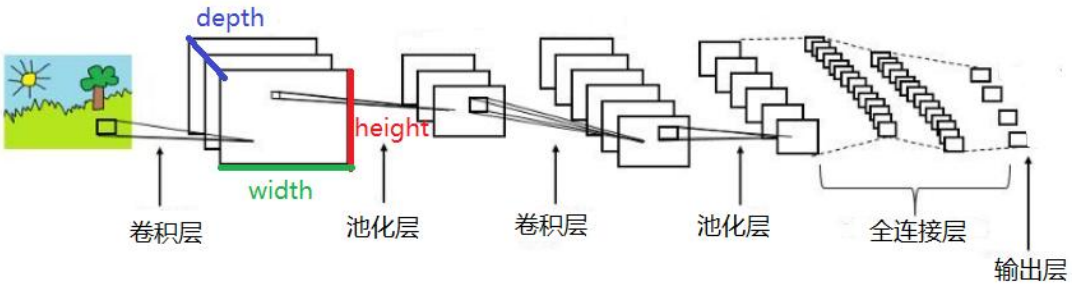

图 2.3　卷积神经网络示意图

Figure 2.3　Schematic Diagram of Convolutional Neural Network

卷积神经网络是由输入层（输入层就是该网络模型的输入），卷积层（主要作用是提取图像的特征，是由卷积核在特征图上面做卷积），池化层（主要作用是进一步提取特征，并且降低分辨率从而减少参数量）和输出层（输出层是网络模型的最后一层，经过输出层之后就是对输入的直接输出了，目的就是把输入的东西进行分类）组成。上图里的 depth、height、width 是特征图的通道数以及高和宽。卷积层相比于全连接层的优点有：1、减少参数来。2、保留图像的空间结构，下面会具体讲解。

2.2.1　卷积层

卷积层的作用是提取图像的特征，卷积层使用时经常会堆叠多层，从底层往上看，底层卷积负责提取不同位置区域的信息，越往上层的卷积相当于是对底层卷积得到的信息进行组合并提取更深层次的信息。只要底层和上层卷积配合得当，是可以在上层提取到全局特征的。

卷积的过程实际上就是用一个带有数值的卷积核，在特征图上上下左右移动，每次移动的过程中都会进行一次运算。当窗口在特征图上进行移动的时候，会计算窗口的权重与特征图的像素的卷积，得到的结果就为新的特征图。卷积的公式如下所示：

$$(f * g)(n) \qquad \text{...（2.1）}$$

卷积运算如下图 2.4 所示：

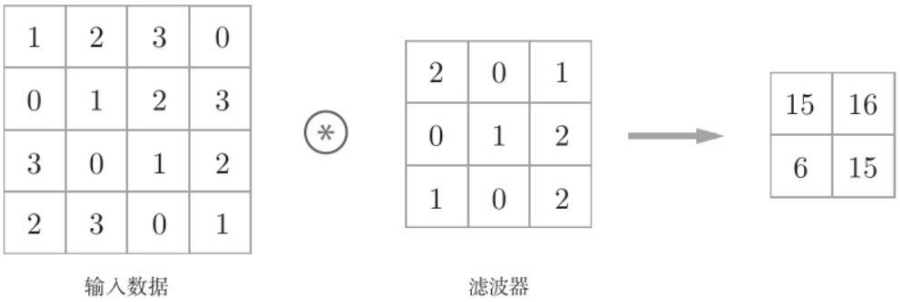

图 2.4　卷积操作（1）

Figure 2.4　Convolution Operation（1）

对于输入数据，卷积操作以一定的间隔滑动过滤器的窗口并应用它。如下图所示，将每个位置的滤波器元素与输入的相应元素相乘再求和。最后，将求和得到的值赋值给新的特征图上面，注意赋值的位置是要根据运算的位置一致。卷积运算的输出可以通过在所有位置执行此过程来获得。如下图 2.5 所示：

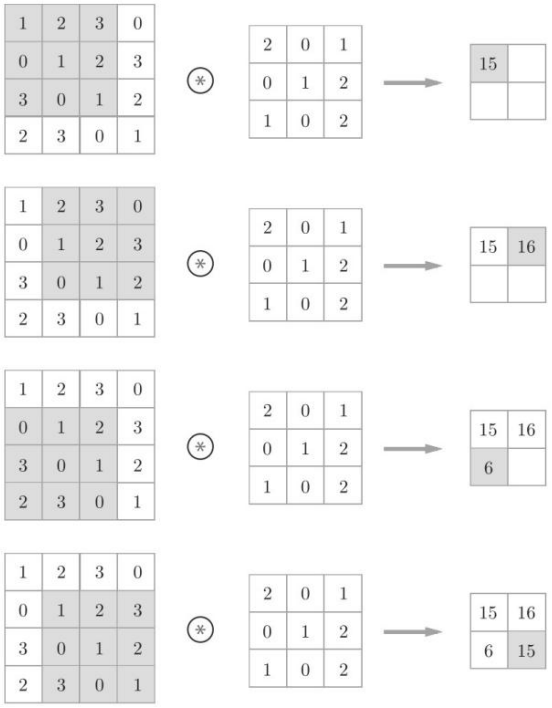

图 2.5　卷积操作（2）

Figure 2.5　Convolution Operation（2）

在全连接层的神经网络中，隐藏层中除了权重参数外，还存在偏置。卷积层中同样也存在偏置，卷积层中滤波器的参数就是对应的权重，除此之外还存在有偏置。带有偏置的卷积运算的处理如下图 2.6 所示：

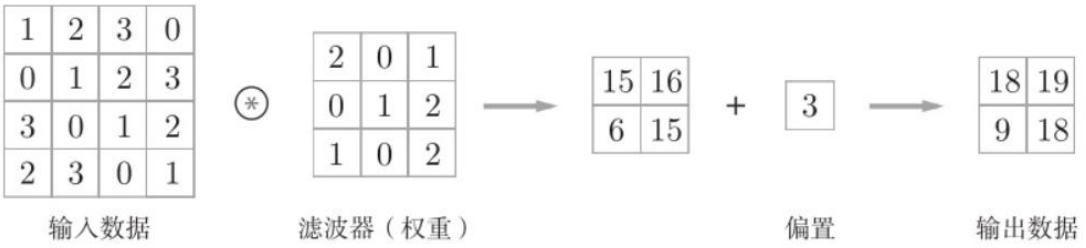

图 2.6　带偏置的卷积操作

Figure 2.6　Convolution Operation with Bias

当滤波器在特征图上面进行卷积操作的过程中可能并不是一次只移动一格而是有间隔的移动，该间隔称为步幅，上面的的例子中步幅都是 1，下图 2.7 就是步幅为 2 时候的运算过程。

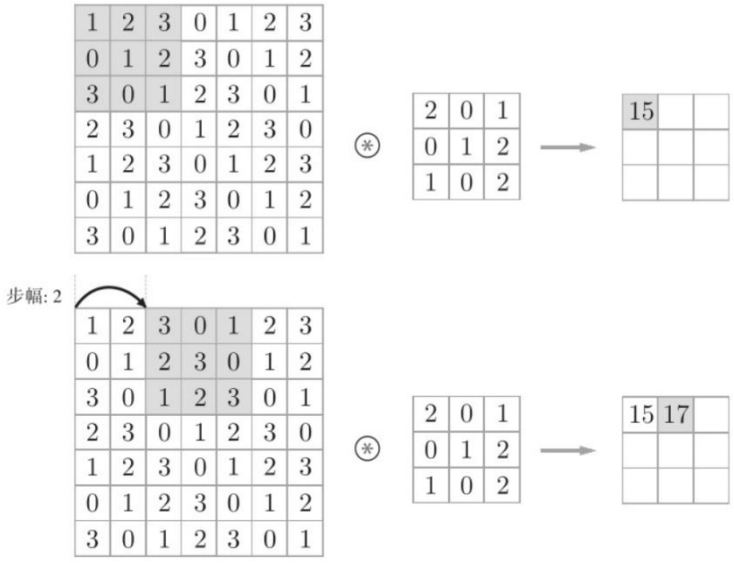

图 2.7　带步长的卷积操作

Figure 2.7　Convolution Operation with Step Size

当采用带有步长的卷积运算的时候，如果步长增加的情况下，输出的特征图的长和宽就会变小。假设输入图片的维度是(H, W)，滤波器大小为(FH, FW)，输出大小为(OH, OW)，填充大小为 p，步长大小 s。此时，输出特征图的大小可通过式（2.2 和 2.3）进行计算。

$$OH = \frac{H+2P-FH}{S} + 1 \quad \text{...（2.2）}$$

$$Ow = \frac{w+2P-FW}{S} + 1 \quad \text{...（2.3）}$$

之前的卷积运算的例子都是以 2 维形状为对象的。但是，图像是 3 维数据，除了高、宽方向之外，还需要处理通道方向。将数据和滤波器结合长方体的方块来考虑，3 维数据的卷积运算会很容易理解。方块是如图 2.8 所示的 3 维长方体。把 3 维数据表示为多维数组时，书写顺序为（channel, height, width），其中 channel 代表通道数（维度），height 代表高，width 代表宽。

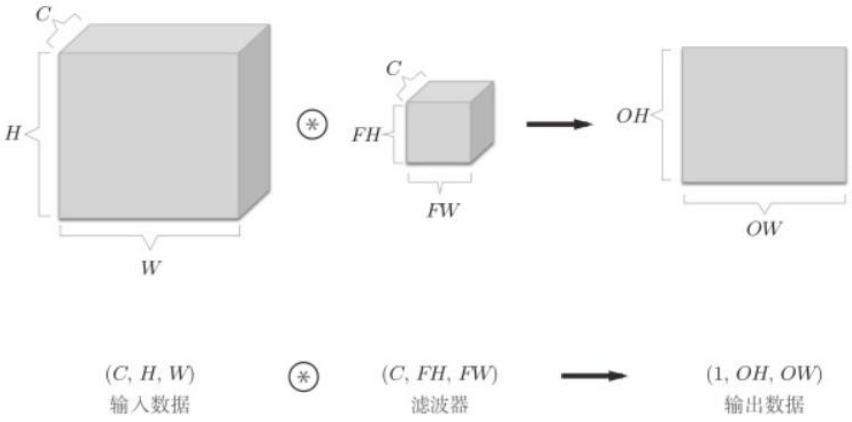

图 2.8　带通道的卷积运算

Figure 2.8　Convolution Operation With Channel

卷积层的特点有三个：局部链接、参数共享、平移不变性。

（1）局部连接：局部链接指的是卷积层之间的节点的链接方式只有一些节点连接，而不是全部连接。在计算机视觉中，像素之间的相关性和图像的某个区域中的像素之间的距离也是相关的。距离较近的像素之间的相关性较强，如果距

离较长，相关性较弱。可见，局部相关理论同样适用于计算机视觉的图像处理领域。局部连接的优点在于可以减少一定的参数量从而并加快了训练速度，也在一定程度上降低了过度拟合的可能性。

（2）参数共享：一个 3*3 大小的卷积核在对特征图进行特征提取的过程中，在通道为 1 的情况下，在每个位置进行特征提取的时候都是共享一个卷积核，假设有 k 个通道，则参数总量为 3*3*k，因为不同通道的参数是不能共享的。在不使用参数共享的情况下，卷积核在输入特征图上移动的过程中在每一位置的权重是不一样的，则卷积核的参数量就会与输入特征提的大小保持一致，假设有 k 个通道，则参数量就是 width*height*k，显然对于大分辨率的输入图像来说是不可取的。通过参数共享的方法，如果只用了局部链接，那个参数量会大大减少，从而提高计算效率。

（3）平移不变性：卷积层的主要目的是为了提取图像的特征，当图像因为卷积操作而发生变化的过程中，并不会发生因为目标改变位置而影响特征提取的情况。比如人脸被移动到了图像左下角，卷积核直到移动到左下角的位置才会检测到它的特征。比如最大池化，它返回感受野中的极大值，如果图片被移动或者改变位置了，但是该区域的最大值是不会改变的，这就有点平移不变的意思了。所以这卷积操作和池化操作共同提供了一些平移不变性，即使图像被平移，卷积也能够提取到该图像本来就具有的特征，并不会被改变。

2.2.2 池化层

在卷积神经网络中，卷积层过后一般都会紧接着进入池化层，池化层的作用是按比例缩小特征图的分辨率，从而减少最后进入全连接层的参数，因此加入池化层可以加快网络模型的预测速度，还可以起到一定的防止过拟合的作用。池化层不仅可以达到降采样的作用，还可以进一步的提取图像的特征。池化层可以分为两种形式：分别为平均池化和最大池化，其中最大池化最为常见。最大池化和平均池化的区别：一般来说，平均池化能减小第一种误差（邻域大小受限造成的估计值方差增大），更多的保留图像的背景信息，最大池化能减小第二种误差（卷积层参数误差造成估计均值的偏移），可以把更多的纹理信息以及边缘信息给保留下来。

最大池化将输入特征图划分为若干个矩形区域，矩形区域的个数是根据池化层中的参数 k 来决定的，然后寻找每个区域中的像素最大值，把每个区域的最大值按照每个区域的排列顺序进行重组之后进行输出。如下图 2.9 所示：

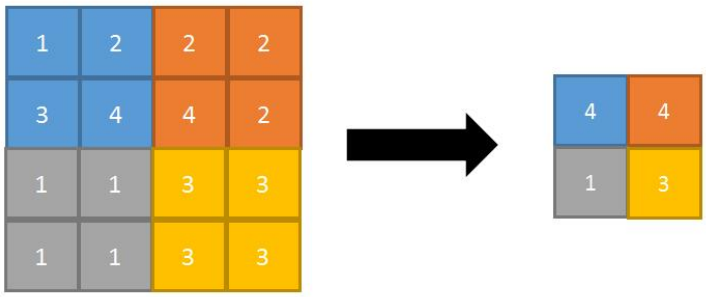

图 2.9　池化操作

Figure 2.9　Pooling Operation

平均池化的过程和最大池化类似，不同点在于最大池化中取得是每个区域的最大值，而平均池化取得是每个区域的平均值。平均池化如下图 2.10 所示：

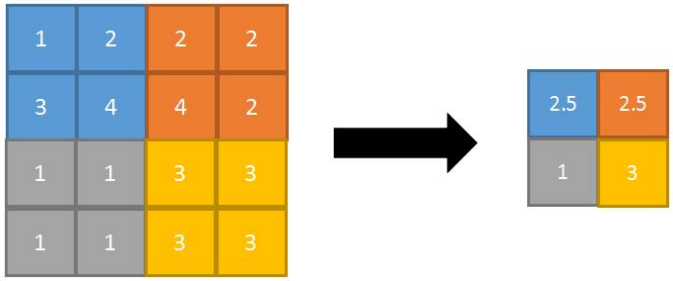

图 2.10　平均池化

Figure 2.10　Average Operation

对于池化的选取：最大池化有利于提取特征图的纹理特征，平均池化有利于提取特征图的背景结构。对于采用最大池化还是采用平均池化需要根据具体的情况而定，也可以二者都采用，之后根据实验结果相互比较效果再决定采用哪一种方式。本文所采取的池化方式为最大池化。

2.2.3 激活函数

在隐藏层中，当某一层的神经元经过运算得到的输出会先经过一个非线性函数在传入下一层，该层会接收到输入特征之后和该层神经元的参数进行运算然后最为下一次神经元的输入。在下一层神经元接收上一层神经元的输入值的时候，输入值会先经过一个函数，这个函数就是激活函数。激活函数的作用就是增加网络模型的表达能力，激活函数分为线性激活函数。以及非线性激活函数，我们采用的是非线性激活函数，可以增加网络模型的非线性因素从而增加表达能力，常见的激活函数如下所示：

1. sigmoid 函数

如图 2.11 所示，Sigmoid 函数是一个曲线，在原点位置的值为 0.5，在 x 轴负半轴位置的值接近于 0，在 x 轴正半轴位置的值接近于 1。Sigmoid 函数有一个特性就是在 x 轴的两端的导数趋向于 0，在远点位置的导数的值是最大的。由于该函数在原点两端呈现不同结果的特点，常常被用在二分类任务中，输出的结果若大于 0.5 就为正类，小于 0.5 就为负类，因此经常被用在二分类任务中。

$$\mathrm{Sigmoid}(z) = \frac{1}{1 + e^{-z}} \quad \text{...（2.4）}$$

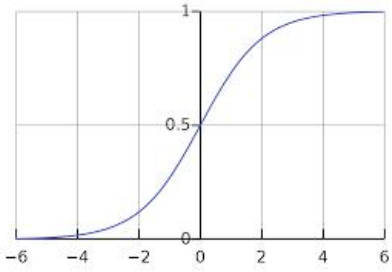

图 2.11 Sigmoid 函数

Figure 2.11 Sigmoid Function

2. Softmax 函数

Softmax 函数不同于 Sigmoid 函数用于二分类的任务，它通过将 0 到正无穷的预测结果归一化到 0 与 1 之间，这使它可以用于多分类的任务。分子函数的值域范围是零到正无穷，这保证了概率值始终为正，并且确保所有预测结果的概率之和等于 1，将转化后的结果除以所有转化后结果之和，这样就将在负无穷到正

无穷的预测结果转换成了 0 到 1 之间的概率。函数表达式为：

$$\mathrm{Softmax}(z_i) = \frac{e^{z_i}}{\sum_{c=1}^{C} e^{z_c}} \quad \text{...（2.5）}$$

3. ReLU 函数

ReLU 激活函数在原点处的值为 0，在 x 轴负半轴的值都为 0，在 x 轴正半轴的值等于输入值，并且在 x 轴正半轴位置的导数都为 1，ReLU 激活函数的图像如下图 2.12 所示。

$$g'(z) = \begin{cases} 1 & \text{if } z > 0 \\ \text{undefined if } z = 0 \\ 0 \text{ if } z < 0 \end{cases} \quad \text{...（2.6）}$$

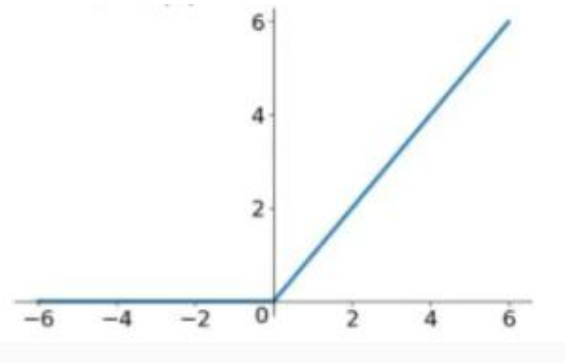

图 2.12　ReLU 函数

Figure 2.12　ReLUFunction

ReLU 激活函数和 Sigmoid 激活函数来说，可以解决 Sigmoid 的梯度消失和梯度爆炸的问题。因为 ReLU 激活函数在 x 轴正半轴的位置时候恒等于 1，但是 Sigmoid 的导数变化较大，若导数较大的话，在反向传播过程中就会造成梯度爆炸，因此 ReLU 激活函数可以解决这一问题，并且较为常用。

激活函数的选择：通常来说，在选择激活函数的时候，同一个网络模型中不能把多个激活函数串起来一起使用。现在在很多网络模型中都避免使用 signoid 激活函数，因为它可能导致梯度消失和梯度爆炸的情况。目前比较常用的激活函数为 ReLU 激活函数，但是在使用它的时候需要注意学习率的设置。但是 ReLU 激活函数也有一定的缺点会导致很多神经元的死亡，采用 Relu 激活函数的变体

Leaky ReLU 或者 PReLU 可以解决此问题。

4. Leaky ReLU 函数

Leaky ReLU 函数和 ReLU 函数的区别在于，ReLU 激活函数是将 x 轴横坐标的左侧的值全部设置为 0，而 Leaky ReLU 激活函数是将 x 轴横坐标的左侧的所有负值都赋予一个非零的斜率，这样就可以保留了 x 轴横坐标左侧的一些值，使得信息不会丢失，从而解决 ReLU 激活函数导致很多神经元死亡的缺点。Leaky ReLU 的函数如下所示：

$$\text{LeakyReLU} = \begin{cases} Z & Z > 0 \\ \alpha Z & Z <= 0,\ \alpha = 0.1 \end{cases} \quad \text{...（2.7）}$$

2.3 卷积神经网络的训练

2.3.1 损失函数

损失函数（Loss function）经常用来评估网络模型的输出值。与输入样本的真实值之间的差别。损失函数是一个非负实值函数，用 L(y, f (x))来表示，f(x)为网络模型的输出值，y 为样本的真实标签。损失函数的值越小，代表网络模型的输出值和真实值越接近，则网络模型的效果越好。

1. 0-1 损失函数（Zero-one Loss）

0-1 损失函数最为简单，若预测值和真实值不相等为 1，若相等为 0，函数表达式如 2.6 所示：

$$L(y, f(x)) = \begin{cases} 1, y != f(x) \\ 0,\ y = f(x) \end{cases} \quad \text{...（2.8）}$$

特点：0-1 损失函数与网络模型判断错误的个数直接对应，是一个非凸函数，不太适用。

2. 均方差损失函数（Mean Square Error）

均方差损失函数是计算网络模型的输出值和输入的真实标签的欧式距离，即输出值和真实值越接近，两者的均方差越小。常用在线性回归中。函数表达式如 2.8 所示：

$$\mathrm{L}(y, f(x)) = \frac{1}{2m}\sum_{i=1}^{m}(f(x)-y)^2 \quad \text{...（2.9）}$$

3. 对数损失函数（Logarithmic Loss Function）

对数损失函数，即对数似然损失，是在概率估计上定义的，经常被用在逻辑回归和神经网络模型中。函数表达式如 2.9 所示：

$$L(y, P(y|x)) = -\log P(y|x) \quad \text{...（2.10）}$$

特点：对数损失函数能非常好的表征概率分布，在很多场景尤其是多分类，如果需要知道结果属于每个类别的置信度，那它非常适合。

4. 交叉熵损失函数（CrossEntropy Loss）

交叉熵损失函数经常被用于分类问题中，交叉熵是香农信息论中的一个重要概念，经常被用作衡量两个概率的差别。在信息论中，交叉熵是表示两个概率分布 p,q，其中 p 表示真实分布，q 表示非真实分布，在相同的一组事件中，其中，用非真实分布 q 来表示某个事件发生所需要的平均比特数。公式如 2.10 所示：

$$\mathrm{L(y, f(x))} = -[f(x)\log y + (1-f(x)\log(1-y)] \quad \text{...（2.10）}$$

交叉熵在神经网络中极为重要，经常被用为神经网络中的损失函数，本论文所采用的损失函数用的也是交叉熵。

2.3.2 反向传播

神经网络之所以有着很强大的提取特征能力，原因在于隐藏层的神经元之间连接的权重，这些权重在构建神经网络的时候一般都会先随机化一组数值，随机化的方式有很多，例如正态分布、Kaiming 初始化等。权重随机化之后，输入特征进入神经网络中会先进行前向传播，即从输入层到隐藏层最终到输出层的一系列运算，最终得到输出。但是如果仅仅使用随机化的权重来对输入值提取特征的话，提取到的特征并不是想要的，也就是神经网络的输出值和输入特征的真实值之间的差别十分大，并不理想。如果想要输出值十分接近输入特征的真实值的话

就要对这些权重进行训练，只有训练出一组很好的参数的情况下，神经网络的输出值才会和输入特征的真实值十分接近。权重的训练就要涉及到反向传播，也就是 BP 算法，反向传播的原理就是根据网络模型输出的值和真实值之间的误差来不断的调整权重，误差从网络模型的最后一层不断的先上一层进行传播，传播的过程中会更新每一层的权重。假设现在训练一个图片的分类网络模型，输入一张图片后逐层的向前计算后，最终会输出一个概率，这个概率就是这张图片属于哪一类的概率，由于刚开始每个神经网络的参数都是随机赋予的，因此输出的概率会和真实情况差别特别大，这时可以根据网络输出值与真实值之间的差距，从最后一层开始逐层向前调整神经网络的参数，若误差值为负，我们就提示权重，若误差值为正，我们就减少权重，调整的程度受一定的比率即学习率的约束，学习率主要用于控制参数调整的幅度，在一次次输入数据和反向调整的过程中，网络就能逐渐给出不错的输出，直到输出十分接近真实结果为止。反向传播的过程如下图 2.13 所示：

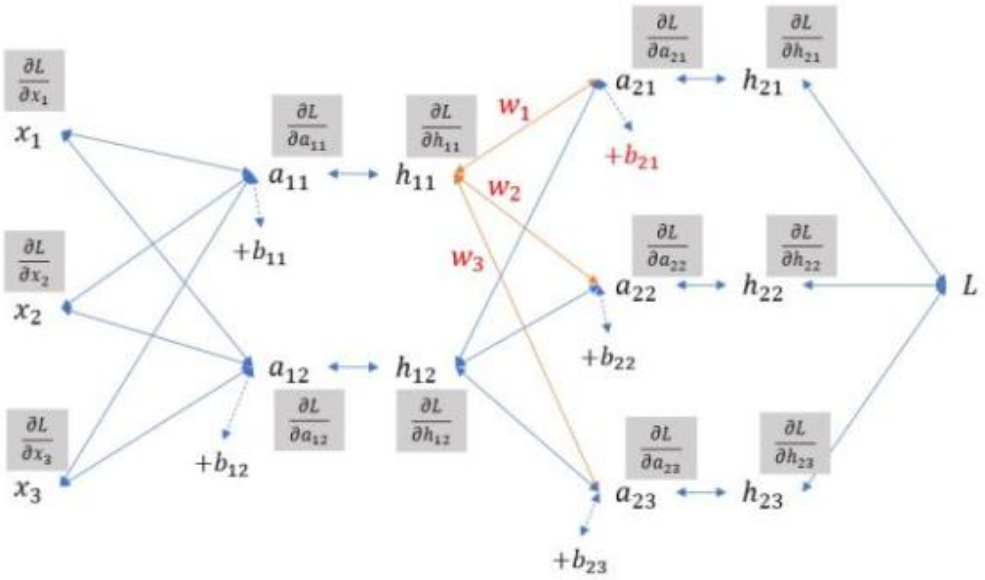

图 2.13　反向传播

Figure 2.13　Backpropagation

如上图所示，x 为输入特征，a 为输出值，h 是 a 经过非线性激活函数的输出值，L 为输出。反向传播开始的时候会先计算 L 和真实值之间的误差，之后将误差传播给 h， ∂L/∂a 是 L 对 a 求偏导，之后把求得的偏导向后传播给输入值 x，这是对于隐藏层是单层的情况下进行的计算，如果隐藏层是多层的情况下同样适用，只是进行重复求偏导在向后传播的循环。

2.4 批标准化

批标准化和普通的数据归一化十分类似，主要作用就是将分散的数据特征给统一到一定范围的一种做法，经常放在神经网络的隐藏层之间，起到优化神经网络的作用。

如果输入到神经网络中的数据具有统一规范性的话能够让机器学习更容易学习到数据之间的特征。在神经网络中数据的分布对网络模型的训练会产生很大的影响，比如说某个神经元的值是 1，与他相连接的权重的初始值为 0.1，这样后一层神经元计算的结果为 0.1，与他相连接的权重的初始值为 20，那么后一层神经元计算结果就是 2，假若使用的非线性激活函数是 tanh 函数的话，那么 0.1 经过激活函数后就会输出 0.1，0.2 进入激活函数后的输出就会变成 0.96，0.2 的输出会在 tanh 的饱和区，无论神经元的值怎么扩大，都会输出在 tanh 的饱和区，换句话说也就是神经网络在初始阶段已经不对那些比较大的 x 的特征范围敏感了，这样会造成神经网络不能充分的学习输入的特征。这时候就需要批标准化对数据进行处理了，批标准化是用在每一个隐藏层之间，大多用在激活函数之前，主要作用就是把该层的输入特征进行标准化，把值过大的特征给予较小的权重，把值过小的特征给予较大的权重，把该层的输入特征固定到一个范围内避免过大或过小，让输入特征经过激活函数非饱和区，也就是激活函数的导数变化较大的部分。

2.5 本章小结

本章主要是对本论文所使用的深度学习方法的一些研究，从深度学习的概念到网络模型的搭建以及训练所用到的方法都有涉及。第一节主要讲的是深度学习

的基本概念以及全连接神经网络和卷积神经网络的对比，经过对比得出，卷积神经网络更适合于图像识别，因此本文所设计的网络模型的搭建也是基于卷积神经网络进行设计的。第二节主要讲的是有关神经网络模型的基本搭建组件，包含了卷积层、池化层、激活函数等。第三节讲的是针对于搭建好的网络模型的训练，包含了损失函数、反向传播两大模块，其中反向传播的目的就是为了训练神经网络模型，使网络模型的输出尽量接近数据的真实标签。最后一节介绍了批标准化，批标准化是一种优化方法，能够使得网络模型能够加速收敛并且能够防止过拟合，该方法作为一种优化方法经常被使用在神经网络中。

第 3 章　基于文本识别的手写汉字识别算法设计

3.1　手写汉字数据集

3.1.1　数据集来源

本实验采用的数据集是由北京邮电大学分布的 HCL2000 数据集，此数据集一共有 1000 多人书写，近 4000 张图片，都是手写的简体中文。HCL2000 经常被用作手写体识别数据集，除此之外，还选取了一些高中生的作文图片作为数据集的来源。

3.1.2　手写汉字特点

手写汉字有着如下特点：

1. 书写风格不规范：每个人都有不同的书写风格，如下图 3.1 所示。比如字体粘连，两个字之间距离过近，书写不规范导致的不在同一水平线上面而是波浪型状的，或者有的人写错字之后就随便涂抹。

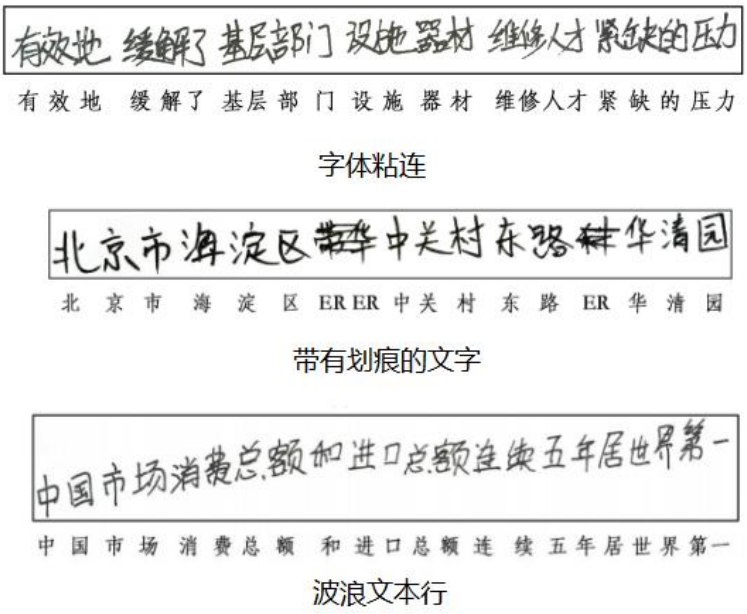

图 3.1　书写不规范的文本

Figure 3.1　Write irregular text

以上这些情况都会导致识别效果不好，因为对于神经网络来训练这些图片其实就是为了从这些图片中提取到特征，但是不规范的书写风格没有固定的特征。

2. 汉字结构复杂：汉字是由笔画组成的，不像英语一样由 26 个英文字母组成，针对于笔画较多的汉字就有 30 多划，并且每个汉字平均的笔划约为 11 划，因此有的汉字结构是非常复杂的，例如“赢”、“繁”，“攥”等，由于笔划较多的原因，携带的特征就比较多，对于汉字的识别不好做特征处理。

3. 汉字有很多形近字：汉字中有着特别多的形近字，如图 3.2 所示。有些字只是一点之差，但是意义完全不同，例如“人、入”、“天、大、夭”、“已、己、巳”等一些相似的汉字。对于这些汉字来说，有着很大的可能会被算法误判为它的形近字，因为算法是基于手写汉字图片的特征来进行识别的，如果两个字十分接近的话，那么这两个字的特征就会十分相似，不好区分两者的特征。

针对于以上手写体汉字的特点，可以采用数据增强的方法来增加网络模型的识别准确率。

3.1.3 数据增强

手写体汉字有着书写风格多样性、存在很多形似的汉字和字符种类庞大的特点，因此相比印刷体来说手写体汉字更难识别。为了提高针对于手写体汉字识别的准确率，除了改进网络模型的结构外，良好的性能同样离不开大规模数据集的支撑。一般来说，准确率较高、泛化能力较强的神经网络模型训练时候需要的数据集是庞大的，并且多样性的。小规模的数据集可能会对网络模型造成过拟合的风险。

面对数据集不足导致的训练效果差的问题，可以想办法从获取更多的数据或者采用数据增强的方法来增加更多的数据集这两方面入手。对于手写体汉字的识别来说，每个人的书写字迹都是不一样的，因此除了可以多收集一些不同人书写的汉字图片外，也可以利用数据增强去增加一些多变的数据集。本文采取数据增强的方法，针对手写体汉字进行翻转、平移、旋转等针对位置的处理来创造更多的样本，来提高模型的鲁棒性。做法如下：

1. 对原始图片按照一定的角度旋转，防止出现拍照不正的图片识别。对文本框进行上下和左右移动，并按照比例来对图片进行增大或缩小。

2. 把文本进行不同程度的扭曲，来达到适应不同人的字迹。对文本进行不同程度的拉伸，增加泛化能力。并对文本做投射变化，来进行不同位置的识别。

3. 小范围内修改每张图片的像素值，达到颜色扰动的效果，来模拟相机拍摄图片时候由于光照等原因造成的颜色干扰。

3.2 文本行区域的检测

文本检测的目的是检测出文本图片里面的所有文本行，但是对于文本检测有一个难点就是关于文本行的大小，文本行的大小一般都是长方形并且大小不是固定的而是变化的。要是采用常规的目标检测的话会面临着怎么生成一些适用的候选框。关于这个问题，本论文提出一种方案就是可以设置一个小的并且固定高宽的矩形框，用这个矩形框去检测每一个文本行的字符，最后把矩形框检测出来的所有字符拼接在一起，最终得到文本行。

在网络模型的设计上面，借鉴了 CTPN 网络模型，网络模型结构如下图 3.2 所示：

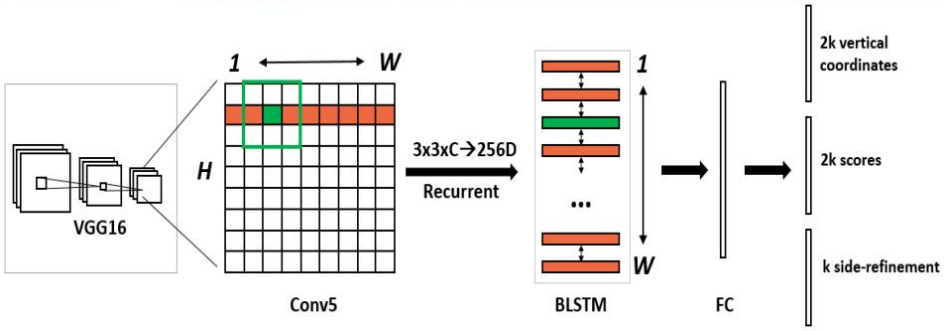

图 3.2 文本行区域检测网络

Figure 3.2 Text Line Area Detection Network

首先使用 VGG16 网络模型作为 backbone，对输入图片进行特征提取，得到的特征图作为 feature map，大小是 W×H×C。在 backbone 中一共进行了 5 次下采样，分辨率缩小了 5 倍。之后采用 3×3 大小的窗口在该特征图上面移动，每次移动过后都能得到一个维度为 3×3×C 大小的特征向量，使用这些特征向量来对锚框的偏移量做预测。经过全连接层的信息整合后最终输出有两个，分别为：

2k 个坐标向量，这个 2 分别为锚框的高和宽，以及 2k 个分数，因为预测了 k 个候选框，所以有 2k 个分数，文本信息和非文本信息各有一个分数。

3.3 基于 CRNN 网络模型的文本识别

CRNN 是由卷积神经网络和循环神经网络结合而成的，主要是为了解决图像的序列问题，可以应用在对于文本字的识别。CRNN 之所以能应用在文本识别领域主要是由于文本行的信息长度是不定长的，而 CRNN 采取了 CTCLoss 恰好能够解决该问题。CRNN 最大的贡献就是把卷积神经网络做提取的特征能力，与循环神经网络提取的序列特征能力相结合。

在识别汉字字符的时候，对于文本序列的识别用该文本行的上下文信息要比单独处理单个的汉字字符信息更有优势。例如，如果单独对一个文字进行识别的话，可能会识别出它的形近字，识别天字的话可能会识别出夭字的可能性很大，但是把天子放在一句话中来进行识别，就像识别“蓝蓝的天空”中，就会识别的十分准确。也就是说有些书写不规范的形近的汉字单独识别的时候并不能很清楚的判断出来，但是放在整个文本行去结合上下文信息的话就容易区分出来了。输入图片在输入到 CRNN 网络模型中并不是直接就经过循环神经网络来进行特征提取的，而是先经由卷积神经网络提取特征后，会先把提取到的特征进行维度上面的调整，之后再输入到模型中。

深度神经网络中的隐藏层都是在水平方向进行延伸的，但是这样的话就有一个缺点，就是没有考虑时序的变化，而循环神经网络则不同，循环神经网络关注隐藏层中每个神经元在时间维度上的不断成长与进步，它会延伸在时间的维度上面，并且建立时序上的关联性，这里的层级扩展并不是指神经元数量增加了而是体现在隐藏层中不同时刻的状态，根据时序的需要隐藏层的时间关联既可以是全连接形式的，即每一个隐藏层之间的神经元节点全部相连，也可以是自己对自己相连接。循环神经网络的结构图如下所示：

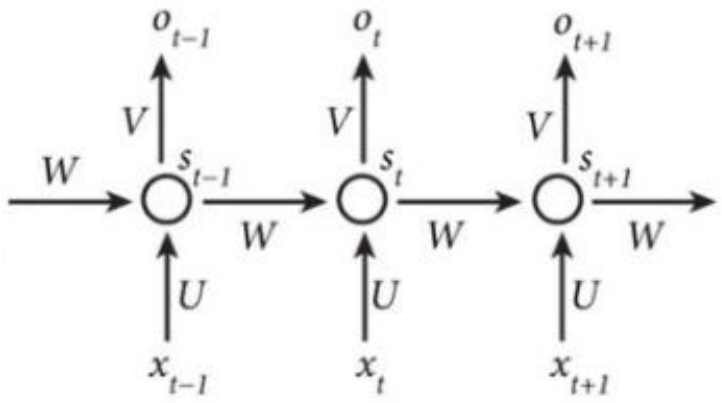

图 3.3　循环神经网络模型结构

Figure 3.3　Recurrent Neural Network Structure Diagram

x 作为特征输入到模型中，x_t 为 t 时刻的输入，从上图可以看出，循环神经网络和卷积神经网络不同之处在于，卷积神经网络是把全部特征进行输入，不管特征之间的联系。循环神经网络是把所有特征按照顺序进行输入，并且把 t-1 时刻的输出结果和 t 时刻的特征做融合。在文本的识别方面，由于句子之间的联系性，可以把一句话中的每个字都分为时序特征，输入到模型中去，从而获得文本的前后文之间的关联性。

虽然循环神经网络可以处理时序上的特征，但也有一定的缺点，在一句话很长的时候，刚开始的字符信息很难传输到最后面的字符信息，也就是循环神经网络的记忆不具有长期性，而长短期记忆网络（LSTM）可以解决这一缺点，该模型提出了门控机制，分别为遗忘门（接连输入的位置，使用非线性函数把强特征保留，弱特征丢弃，主要功能是有选择性的保留特征信息）、输入门（连接着输入特征，将经由选择门输出的特征进一步进行特征选择，把强特征保留，弱特征丢弃）、输出门（隐藏状态以及当前输入的特征信息先经过非线性激活函数以确定隐藏状态应携带的信息，之后把新的状态输入到下一个时序特征中），LSTM 网络结构如下图 3.4 所示：

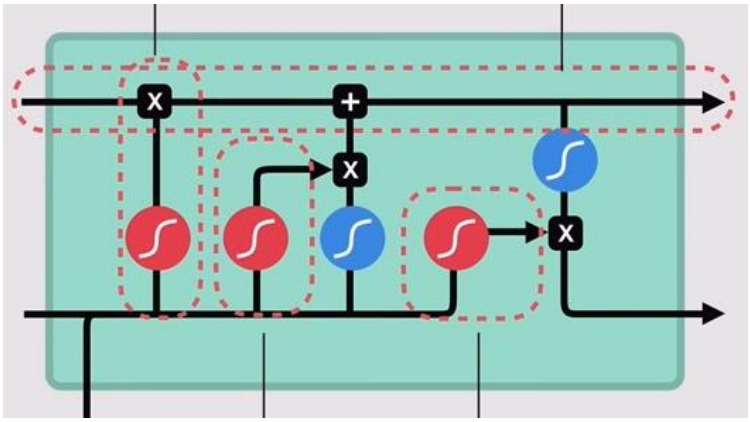

图 3.4　LSTM 模型结构

Figure 3.4　LSTM Model Structure

3.4　网络模型的设计

网络模型一共由三部分组成：首先是特征提取模块主要目的是为了提取输入图的特征信息、其次是文本检测模块主要目的是检测出文本图像中的文本框，最后是文本识别模块主要目的是对检测出来的文本框进行字符识别，CNN 特征提取模块采用的 ResNet 网络结构和空洞卷积结构，主要用来提取图像特征，经过 CNN 提取到的特征图再经过 Text Proposal Network 模块提取候选框，之后分为两部分，进入 RPN 模块的属于文本检测部分，经过 RPN 对候选框进行边框回归，之后和 Text Proposal Network 模块提取到的候选框进行修正输入到 BiLSTM（双向长短期记忆网络）模块来进行文本识别，网络模型结构如下图 3.5 所示。

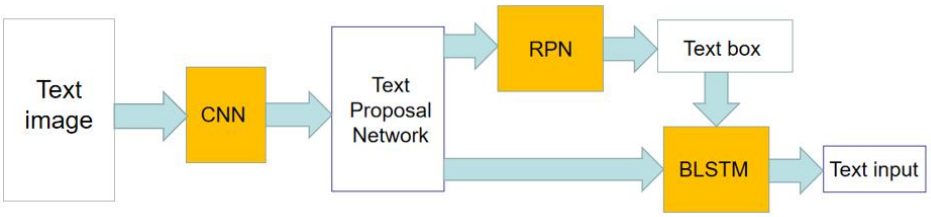

图 3.5　网络模型结构

Figure 3.5　Network Model Structure

现有的 OCR 技术大部分是把文本检测和文本识别分开进行，这样做的缺点是要计算两遍卷积，运行速度较低。本文提出的模型，把文本识别和文本检测放在一起进行，卷积的参数是共享的，只需要进行一次卷积，大大提高了运行速度。

3.4.1 特征提取模块

在卷积神经网络中，会把低层的特征和高层的特征做融合，达到多尺度特征的目的，可以提高网络模型的性能。对于前几层特征图来说，进行卷积的次数较少，分辨率较大，因此低层特征图所富含的语义信息较少，但是富含的位置信息以及细节信息较丰富。高层特征经过较多次数的卷积，拥有较高的语义信息，像 Yolo V3 网络模型中，做法是将输入图像经过 8 倍下采样之后的特征图和 16 倍下采样之后的特征图上采样之后进行特征融合进行一次输出，之后在和 32 倍的特征图上采样之后在进行一次特征融合进行输出。这样把低层特征和高层特征做融合的目的就是为了把低层的空间信息以及细节信息与高层特征的语义信息做融合，不仅可以提高网络模型的分类能力还能提高网络模型对于位置信息的准确率。除了这种融合高低层特征信息方法除外，也可以每次进行一次下采样都进行输出预测，这两种方法都是为了做多尺度预测。

特征融合按照融合的先后顺序，可以分类为早融合和晚融合。

早融合：顺序是先把早期的输入特征图就进行融合，之后在融合后得到的新的特征图上面进行训练，特征图融合的方式有两种操作，分别为 concat 以及 add 操作。

（1）concat：将两个特征图再通带维度上做融合，但是两个特征图的长和宽必须是一致的，通道可以不相同。相当于把两个特征图进行拼接操作，若两个特征图的长和宽都是 w*h，通道数量分别为 c1 和 c2，那么 concat 之后的特征图大小为 w*h*(c1+c2)。

（2）add：两个特征图做 add 操作的前提条件是它们的长和宽以及通道数量全部一致，把两个特征图的每个通道上面的像素对应相加得到新的特征图，若像素值相加后超过 255 则改成 255，相当于并行策略。

晚融合也有两种形式，分别为以下方式：

（1）将不同尺度的特征图分别输出进行预测，也就是说每进行卷积过后都

会在指定的层数之间将输出特征图作为输出，相当于做多尺度特征，最后将所有的输出结果按照比例进行判断预测，如 SSDNet, Multi-scale CNN(MS-CNN)网络模型等。

（2）对特征进行金字塔融合。

本文特征提取模块采取的是 ResNet 网络结构，如下图 3.6 所示。网络经过四次下采样之后，在经过两次上采样，这里采用转置卷积的方法进行上采样。最终得到的特征图的分辨率为原始输入图像的 1/4。这里采取特征融合方法，即分别把 4 倍、8 倍、16 倍下采样之后的特征图和后面上采样之后的特征图做通道上的融合。这样把低层特征和高层特征做融合的目的就是为了把低层的空间信息以及细节信息与高层特征的语义信息做融合，不仅可以提高网络模型的分类能力还能提高网络模型对于位置信息的准确率。除了这种融合高低层特征信息方法除外，也可以每次进行一次下采样都进行输出预测，这两种方法都是为了做多尺度预测。

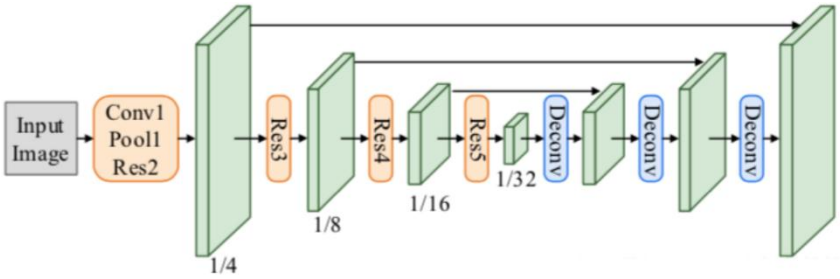

图 3.6　特征提取模块

Figure 3.6　Feature extraction module

每一个特征图上对应的感受野越大表示能够接触到原始输入图像的范围就越大，同时也就意味着富含更丰富的语义信息。相反，感受野的值越小就表示富含的局部细节信息较为丰富，语义信息不足。在卷积的过程中，分辨率和感受野是乘反比的，分辨率小的情况下对于边框的定位是不利的，为了解决这个问题，引入了空洞卷积。空洞卷积是在原始的特征图中添加空洞，即每隔 r（空洞率）个像素填充 0，从而达到增大感受野的效果。和正常卷积相比较，引入空洞卷积的目的就是为了减少下采样的同时增加特征图对应的感受野，从而提高文本检测

的准确率，加入空洞率的卷积核的大小如下公式 3.1 所示：

$$K_{new} = K_{old}(K_{old} - 1)(R - 1) \quad \text{...（3.1）}$$

3.4.2 RPN 网络结构

RPN 第一次出现是在 Faster RCNN 网络模型中，主要作用就是为了提取候选框并对对预选框进行修正，在 Faster RCNN 之前，对候选框进行提取的方法一般采用 Selective Search，这个是比较传统的方法，是选用一个框在输入图上面进行上下左右的滑动，比较耗时。RPN 的引入，把物体检测的整个流程融入到了一个网络模型中去，只需要进行一次前向传播就能得到结果。RPN 的运行机制如下图 3.7 所示：

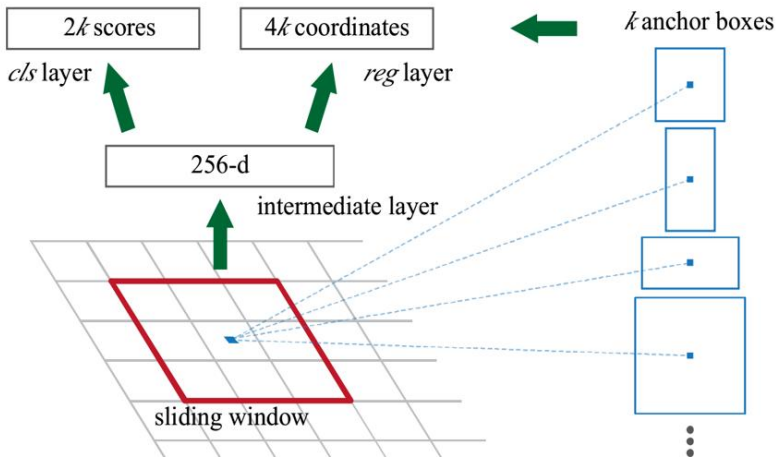

图 3.7 RPN 模块

Figure 3.7 RPN Module

RPN 的输入特征图是在 backbone 的输出特征，特征图进入 RPN 网络中首先经过一个 3*3 的卷积层和一个 Relu 激活函数，之后分成两个分支，一个分支进行 1*1 的卷积后在经过 softmax 层得到的是分类是前景还是背景。另一个分支也

经过 1*1 的卷积后得到的是坐标的偏移量，以获得更准确的区域。

如下图 3.8 所示。其中 Region proposal 是提取到的候选框，根据特征图和原图的比例映射到特征图中，再经过卷积核池化操作，最终输入到全连接层看属于前景还是背景，把背景去除掉，最后针对前景框进行边框修正。

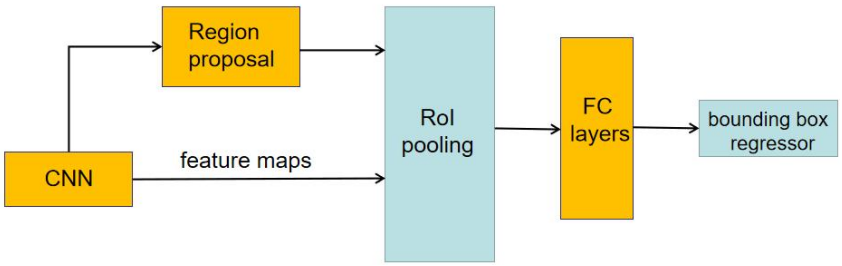

图 3.8　候选框的提取

Figure 3.8　Extraction of Candidate Boxes

本文采取的是基于 anchor 进行修正，因为文本框一般都属于长条形状的，所以预先设置 anchor 的长宽比，之后再基于 anchor 的基础上进行边框修正，这样仅仅计算偏移量而不是物体的位置大大降低了优化难度。边框修正实例图如图 3.9 所示：

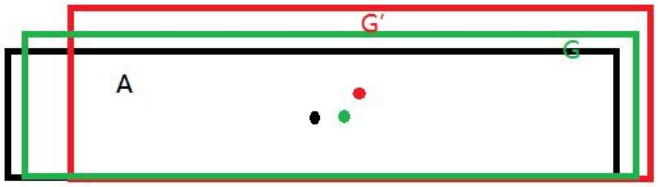

图 3.9　Anchor 的设计

Figure 3.9　Design of Anchor

对于图 3.7 来说，一共有三个边框，分别为：黑色的是 anchor，红色的是模型预测出来的边界框，绿色的为文本图片上真实标注的边界框。如果预测出来的框和真实标注的框越接近说明网络模型的效果越好，这里使用均方差损失函数来

对边界框做回归。对于边界框的坐标使用(x,y,w,h)四个值来表示，代表中心的坐标以及宽高。

先做平移：

$$\begin{aligned} \mathrm{G}_{\mathrm{x}}^{'} &= A_w * d_x(A) + A_x \\ \mathrm{G}_{\mathrm{y}}^{'} &= A_h * d_y(A) + A_y \end{aligned} \quad \text{...（3.2）}$$

再做缩放：

$$\begin{aligned} \mathrm{G}_{w}^{'} &= A_w * \exp(d_w(A)) \\ \mathrm{G}_{h}^{'} &= A_h * \exp(d_h(A)) \end{aligned} \quad \text{...（3.3）}$$

3.4.3 双向长短期记忆网络（BiLSTM）

文本检测和普通的物体检测是有所区别的，文本检测的目标要小，十分密集都是细长条的检测，并且前后文本之间是互相联系的，比如说一个单词在不同的语义之间会有不同的意思，而目标检测的每个目标都是互相独立的，两者之间是没有联系的，因此在文本检测以及文本识别的时候都可以结合文本的上下文信息，因此对于文本不光要提取图片的空间特征，同时引入文本的序列特征可以提高模型的精度，本文采取 BiLSTM（双向长短期记忆网络）来提取文本的序列信息，如下图 3.10 所示。

本文把提取到的特征结合时间序列特征来训练文本识别的算法，文本检测网络提取到的文本边框映射到卷积神经网络提取到的特征中去，先把维度转变为序列特征，之后输入到 BiLSTM 中去，结合上下文本信息，输出为预测的文本序列，这里引入 CTC Loss 损失函数，主要解决不定长序列对齐的问题。

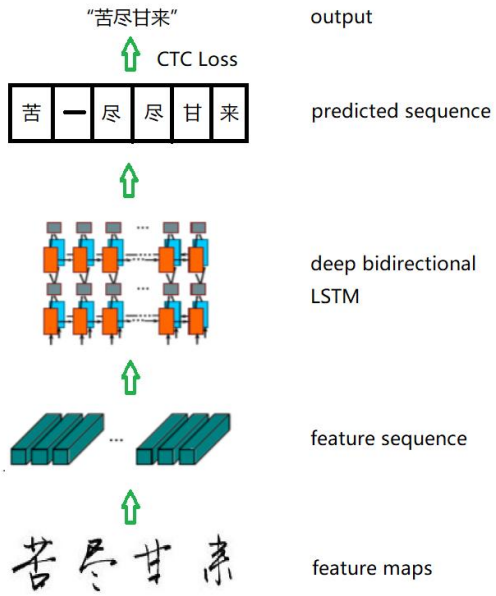

图 3.10　BLSTM 的设计

Figure 3.10　Design of BiLSTM

3.5　多任务损失函数

该网络把检测和识别集于一体，再训练的时候，文本检测的时候使用的是交叉熵损失，在文本识别的时候使用的是 CTC Loss。在文本识别的输入与输出 GT 文本很难在单词上对齐,在预处理的时候对齐是非常困难的,但是如果不对齐而直接训练模型的话,由于字符距离的不同,导致模型很难收敛。采用 CTC Loss 主要是为了解决上述问题。

CTC(Connectionist Temporal Classification) Losss 的中文名字为链接时序分类，这要用来解决深度神经网络中的预测值和输入值的真实标签长度不对齐的问题，CTLLoss 的特点在于不需要强制对标签进行对齐并且标签的长度是可变的，只需要输入序列和标签就可以进行训练。CTCLoss 一般用在文字识别、语音识别、验证码识别等领域。

CTCLoss 的输入来自 LSTM 的输出，对于 LSTM 给定输入 x 的情况下，输出 l 的概率为：

$$p(l|x) = \sum_{\Pi \in B^{-1}(l)} p(\Pi|x) \quad \text{...（3.4）}$$

其中$\Pi \in B^{-1}$（l）代表所有经过 B 变换后是 l 的路径Π。其中，对任一条路径Π都有：

$$pp(\Pi|x) = \prod_{t=1}^{T} y^{t}_{\Pi t}, \forall \Pi \in L^{T} \quad \text{...（3.5）}$$

注意这里的$y^{t}_{\Pi t}$中的Π,小标 t 表示Π路径的每个时刻。T 是根据输入数据而自己定的。CTCLoss 在网络训练的过程中，利用本身的梯度对长短期记忆网络里面的权重进行改变，使$\Pi \in B^{-1}$（l）的时候，p(l|x)取值为最大。

3.6 段落的识别

对于图片汉字识别的流程都是从文本行的左侧开始，到右侧结束。如下图 3.11 所示的图片：

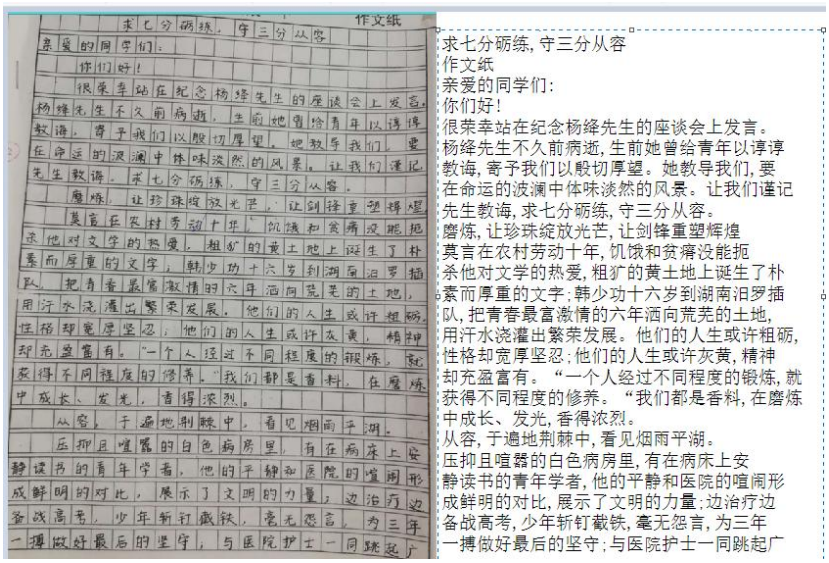

图 3.11 识别结果展示

Figure 3.11 Recognition Results Are Displayed

上图左侧为上传的手写体文本图片，右侧为识别结果。可以看出来此结果有一个缺点，就是只能按照行来识别，原始图片上面的每一个文本行单独进行识别，并不会把段落识别出来，所以识别出来的排版并没有段落，而是按照文本行的顺

序排列。针对这一问题，可以根据网络模型输出每一行的文本框的坐标信息进行对段落的检测，检测出来段落后在调整识别结果的输出格式。

在检测模型中，网络的输入是 C（类别数）+ confinde（置信度）+ 4*3（4 是[x,y,w,h]，3 是锚框的数量），其中[x,y,w,h]为每个文本框的空间信息，x,y 代表文本行所在的框的中心坐标，w 代表文本行所在的框的宽度，h 代表文本行所在的框的高度。因此可以根据 x,y,w,h 四个信息计算出每个文本框的最左侧的坐标信息，得到这个信息之后每个文本行的最左侧信息也就可以计算出来。有了所有文本行的最左侧坐标信息之后，根据每个文本行最左侧坐标信息的插值就可以判断出该文本行是否属于新的一段，对识别结果进行分段。每行的坐标信息如下图 3.12 所示：

{'words_result': [{'location': {'top': 2, 'left': 168, 'width': 309, 'height': 40}, 'words': '求七分砺练,守三分从容'}, {'location': {'top': 0, 'left': 509, 'width': 68, 'height': 19}, 'words': '作文纸'}, {'location': {'top': 32, 'left': 57, 'width': 179, 'height': 27}, 'words': '亲爱的同学们:'}, {'location': {'top': 66, 'left': 109, 'width': 96, 'height': 25}, 'words': '你们好!'}, {'location': {'top': 96, 'left': 102, 'width': 517, 'height': 51}, 'words': '很荣幸站在纪念杨绛先生的座谈会上发言。'}, {'location': {'top': 128, 'left': 42, 'width': 574, 'height': 52}, 'words': '杨绛先生不久前病逝,生前她曾给青年以谆谆'}, {'location': {'top': 161, 'left': 39, 'width': 577, 'height': 51}, 'words': '教诲,寄予我们以殷切厚望。她教导我们,要'}, {'location': {'top': 195, 'left': 38, 'width': 580, 'height': 48}, 'words': '在命运的波澜中体味淡然的风景。让我们谨记'}, {'location': {'top': 228, 'left': 38, 'width': 474, 'height': 45}, 'words': '先生教诲,求七分砺练,守三分从容。'}, {'location': {'top': 264, 'left': 93, 'width': 526, 'height': 43}, 'words': '磨炼,让珍珠绽放光芒,让剑锋重塑辉煌'}, {'location': {'top': 296, 'left': 89, 'width': 529, 'height': 43}, 'words': '莫言在农村劳动十年,饥饿和贫瘠没能扼'}, {'location': {'top': 326, 'left': 25, 'width': 592, 'height': 46}, 'words': '杀他对文学的热爱,粗犷的黄土地上诞生了朴'}, {'location': {'top': 358, 'left': 24, 'width': 596, 'height': 47}, 'words': '素而厚重的文字;韩少功十六岁到湖南汨罗插'}, {'location': {'top': 393, 'left': 22, 'width': 582, 'height': 43}, 'words': '队,把青春最富激情的六年洒向荒芜的土地,'}, {'location': {'top': 427, 'left': 20, 'width': 601, 'height': 43}, 'words': '用汗水浇灌出繁荣发展。他们的人生或许粗砺,'}, {'location': {'top': 462, 'left': 19, 'width': 597, 'height': 41}, 'words': '性格却宽厚坚忍;他们的人生或许灰黄,精神'}, {'location': {'top': 495, 'left': 16, 'width': 603, 'height': 41}, 'words': '却充盈富有。“一个人经过不同程度的锻炼,就'}, {'location': {'top': 530, 'left': 15, 'width': 607, 'height': 41}, 'words': '获得不同程度的修养。“我们都是香料,在磨炼'}, {'location': {'top': 567, 'left': 15, 'width': 349, 'height': 30}, 'words': '中成长、发光,香得浓烈。'}, {'location': {'top': 602, 'left': 72, 'width': 498, 'height': 37}, 'words': '从容,于遍地荆棘中,看见烟雨平湖。'}, {'location': {'top': 638, 'left': 70, 'width': 546, 'height': 41}, 'words': '压抑且喧嚣的白色病房里,有在病床上安'}, {'location': {'top': 673, 'left': 5, 'width': 616, 'height': 42}, 'words': '静读书的青年学者,他的平静和医院的喧闹形'}, {'location': {'top': 710, 'left': 6, 'width': 615, 'height': 42}, 'words': '成鲜明的对比,展示了文明的力量;边治疗边'}, {'location': {'top': 746, 'left': 5, 'width': 611, 'height': 41}, 'words': '备战高考,少年斩钉截铁,毫无怨言,为三年'}, {'location': {'top': 784, 'left': 4, 'width': 605, 'height': 41}, 'words': '一搏做好最后的坚守;与医院护士一同跳起广'}], 'words_result_num': 25, 'log_id': 1448135188275873305}

图 3.12　文本行的坐标信息

Figure 3.12　Coordinate Information For a Line of Text

根据坐标信息把识别结果进行分段之后的结果如下图 3.13 所示：

求七分砺练,守三分从容
亲爱的同学们:
　　你们好!
　　很荣幸站在纪念杨绛先生的座谈会上发言。杨绛先生不久前病逝,生前她曾给青年以谆谆教诲,寄予我们以殷切厚望。她教导我们,要在命运的波澜中体味淡
先生教诲,求七分砺练,守三分从容。
　　磨炼,让珍珠绽放光芒,让剑锋重塑辉煌。
　　莫言在农村劳动十年,饥饿和贫瘠没能扼杀他对文学的热爱,粗犷的黄土地上诞生了朴素而厚重的文字;韩少功十六岁到湖南汨罗插队,把青春最富激情的六
用汗水浇灌出繁荣发展。他们的人生或许粗砺,性格却宽厚坚忍;他们的人生或许灰黄,精神
却充盈富有。“一个人经过不同程度的锻炼,就获得不同程度的修养。“我们都是香料,在磨炼中成长、发光,香得浓烈。
　　从容,于遍地荆棘中,看见烟雨平湖。
　　压抑且喧嚣的白色病房里,有在病床上安静读书的青年学者,他的平静和医院的喧闹形成鲜明的对比,展示了文明的力量;边治疗边备战高考,少年斩钉截铁,
一搏做好最后的坚守;与医院护士一同跳起广

图 3.13　分段后的结果展示

Figure 3.13　The Results Are Displayed After Segmentation

3.7 实验结果与分析

3.7.1 实验环境

本次实验条件：处理器 CPU 为 Intel i7 8700K，GPU 采用的是 Nvidia RTX3090 Advance 12GB 的显存容量，选用 Pytorch 作为深度学习框架。

3.7.2 评价指标

文本检测模块采用的衡量指标为评价精度均值（mAP)。mAP 被广泛的用到目标检测任务中，是各类 PR 曲线的平均值。PR 曲线是由 Recall 以及 Precision 组成的，分别为横坐标以及纵坐标。因为针对于目标检测问题，每张图片上面都有一种或多种不同类别的物体，当评估网络模型性能的时候，用于图像分类的指标精度并不能直接应用于此，这就是为什么需要 mAP。当然文本检测也是属于检测问题，mAP 指标同样适用。

1. 真实标签

在对模型进行评估的时候，除了要选择合适的评价指标外，还需要知道测试集的真实值即测试集的真实标签。只有知道了测试集的真实标签，才能够拿真实标签和网络模型的预测值进行比较，从而判断网络模型的性能。对于文本识别问题，真实标签就相当于汉字文本图片中的文字所在的区域的边界框以及该边界框内的文字是什么。

2. mAP 的含义及计算

使用训练好的网络模型对新输入的图片进行预测的时候，会产生大量的结果，这些结果有的并不是我们想要的，只有那些置信度（confidence score）很高的结果才是我们需要的。因此可以采用一种后处理的方法，去把那些置信度很低的结果排除掉，这里使用的方法是对置信度做阈值筛选。网络模型的输出结果是很多带有边界框的预测结果，并且每一个结果都会带有一个置信度，首先需要检测每个输出结果是否正确，采用的方法是 Iou。

IoU 指交兵比，也就是预测的边框和真实边框的交集得到的值去除以并集得到的值。如下图 3.14 所示：

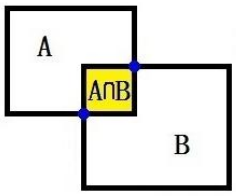

图 3.14　IOU

Figure 3.14　IOU

假设白色 A 的区域为预测出来的边框，白色区域 B 为真实的边框，那么黄色区域就是区域 A 和区域 B 的 IoU。

为了计算精确率（precision）和召回率（recall），我们必须找出真正例（TP）、假正例（FP）、真负例（TN）和 假负例（TN）。如下表 3.1 所示：

表 3.1　正反例表

Table 3.1　Positive and Negative columns

真实情况	预测结果	
	正例	反例
正例	TP	FN
反例	FP	TN

得到正反例表之后我们就可以计算出来精确率和召回率，PR 曲线的横坐标即召回率。纵坐标为精确率。AP 的值可以通过 PR 曲线下的面积计算得到，而 mAP 就是 AP 的值除以类别数。

3.7.3　实验结果分析

进行训练的时候将所有的输入文本图像的分辨率大小调整为 600*600，若输入图像的分辨率大于 600*600 的话采取图像金字塔来进行调整。若输入图像的分辨率小于 600*600，用最 600 除以长和宽的最大值得到比例系数，之后再用小值乘上比例系数进行扩大图像。训练刚开始的时候把学习率设置为 1，学习率并不

是固定不变的而是随着训练的轮数进行不断的衰减，设置衰减系数为 0.1，根据这个衰减系数每经过一个轮次就对学习率进行一次改变。Batch_Size 选用 32，太大的话容易造成显存爆炸，太小的话会造成结果很差，经过实验验证选择 32 效果较好。因为在网络设计的时候，每层卷积层过后都会加入 BN 层，因此没有使用 Dropout。梯度下降算法采用的是 Adam 算法，训练测试曲线图如下图 3.15 所示：

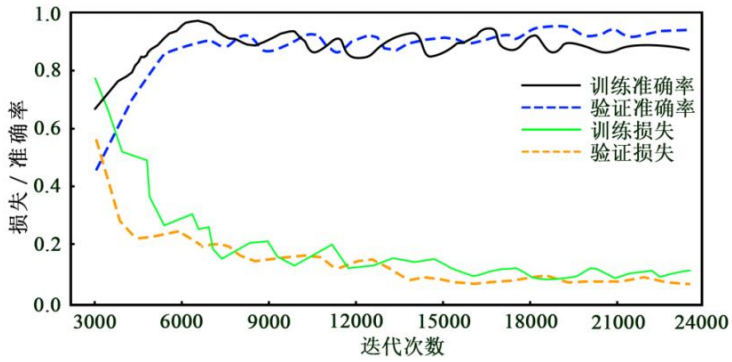

图 3.15　实验曲线图

Figure 3.15　Experimental Graph

在 4000 轮迭代后，损失函数开始有明显下降，18000 轮迭代后，损失函数逐渐停止下降，此时验证准去率大于训练的准确率。

为了说明本文所设计的网络模型的有效性，除了在检测结果方面和其他网络模型做了对比实验，还在识别速度方面和其他模型做了对比实验，其中，Model 为本文所提出的网络模型，如下表 3.2 所示：

表 3.2　实验结果对比表

Table 3.2　Comparison Table of Experimental Results

Model	Precision	Accuracy	Rate
CTPN	0.76	0.72	16fps
EAST	0.80	0.73	20fps
SegLink+CRNN	0.84	0.80	14fps
PixelLink+CRNN	0.82	0.81	16fps

TextBoxes+CRNN	0.86	0.74	23fps
Model	0.88	0.78	36fps

实验结果可以看到，相比于其他文本识别算法，本文提出的基于端到端的文本识别算法相比于其他的文本识别的算法在准确率和速度上均有提升。

3.8 本章小结

本章主要对手写体汉字识别网络模型的设计做了详细的介绍，从数据集的来源以及数据增强处理，到文本识别和文本检测算法模型的设计思路，以及最后的网络模型的实验测试。

第一节主要介绍了该网络模型所使用的数据集的来源，并且针对手写体汉字的数据集所设计的数据增强的方法。第二节介绍了关于文本检测的算法的设计，并且借鉴了 CTPN 网络模型进行设计，文本检测的目的主要是为了检测出文本图片中的文本框所在的位置。第三节介绍了文本识别算法，把 CTCLoss 引入到算法中，解决了不定长度的标签对齐的问题，同时还引入了循环神经网络，结合文本的上下文信息来增加对手写汉字文本识别的准确率。第四节介绍了该网络模型的特征提取模块，把文本检测和文本识别融合到一个特征检测模型中，只进行一次前向传播就能够输出结果。第五节介绍了该模型所采用的损失函数为多任务损失函数，最终的损失是文本检测和文本识别的损失加权和得到的结果，第六节是基于段落的识别，目前文本识别算法是不能做到把段落给识别出来的，本文基于文本检测出来的每行文本框信息的坐标差来实现对段落的一个分类。最后一节主要是对手写汉字识别算法的一个实验测试，先提出了衡量指标之后采用测试集对该网络模型进行测试，测试结果表明，相比于其他文本识别算法，本文提出的基于文本识别算法相比于其他的文本识别的算法在准确率和速度上均有提升。

第 4 章　基于文本识别的手写汉字识别平台的设计与实现

前面几章依次介绍了有关深度学习的概念、手写汉字识别网络模型的搭建以及为了解决汉字识别效率慢而把文本检测和文本识别融合在一个模型中的思路及实现。在这些前期准备工作完成的基础上，本章主要内容是对手写体汉字识别平台的设计与实现工作，系统平台主要是基于中小学生上传的语文作文图片进行手写体汉字的识别。并完成了及用户注册、登录、上传图片、识别图片、结果展示、数据保存、错误反馈一整套识别流程的开发任务。系统功能示意图如下图 4.1 所示。

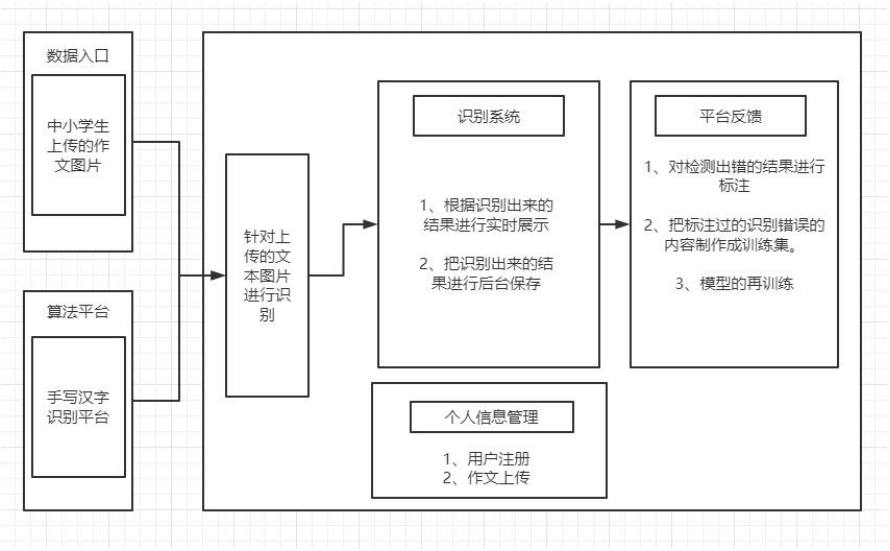

图 4.1　系统功能示意图

Figure 4.1　Schematic Diagram of System Functions

4.1　系统的需求分析

本小节主要是针对于系统平台的各项需求做分析，看平台都需要哪些需求，并且在下一节针对本小节所提出的需求进行设计。

4.1.1　系统的功能性需求分析

基于文本识别的手写汉字识别平台需要实现的主要功能如下所示：

1. 用户登录功能

该功能是平台最基本的功能，使用者可以创建自己的账户并登录，设计登录功能的目的是为了使使用者每次进行作文识别的时候都会把识别的结果保留在自己的账户中，方便下次查看。登陆界面如下图 4.2 所示：

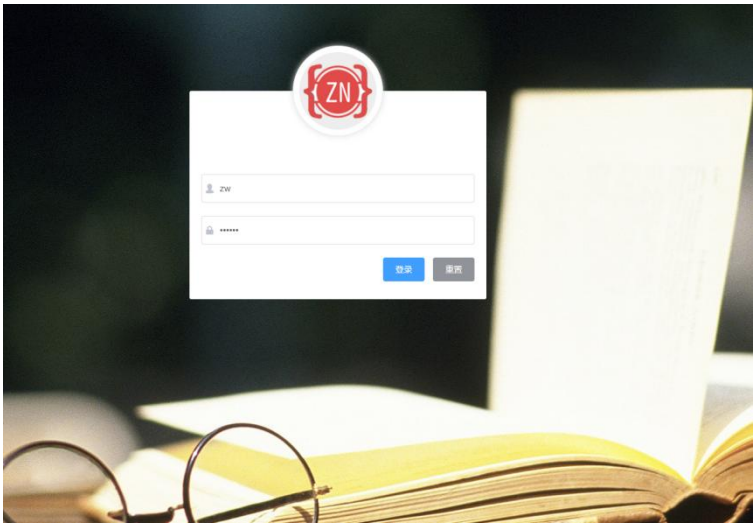

图 4.2　用户登录

Figure 4.2　The User Logs On

2. 用户上传图片功能

该功能的作用是用户上传待识别的手写汉字图片，供算法的识别。可以分为单张上传图片或者一次上传多张图片两种方式。也就是说不仅在上传模块的时候可以单张作文图片的上传，还可以一次性上传多张作文图片，当上传多张作文图片的时候，平台会对所有的作文图片进行汉字识别，并按照上传作文图片的顺序进行结果展示。这样做的好处就是消除使用者多次上传图片带来的不便性。

3. 文本图书识别功能

此功能为平台的主要核心功能，目的就是将用户上传的文本图片经过算法把图片中的汉字给识别出来。首先算法模型需要经过训练，之后保留训练好的参数，最后根据新上传的图片进行识别。训练流程图如下图 4.3 所示：

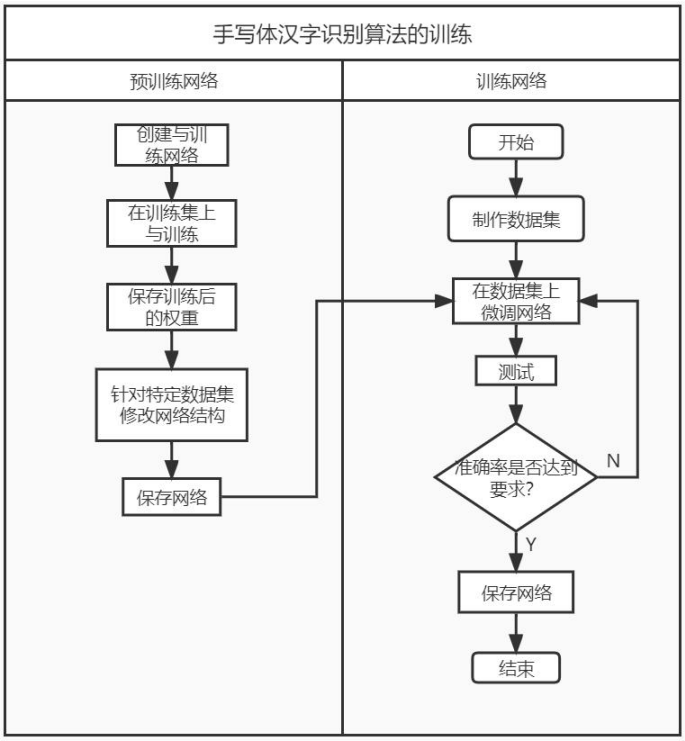

图 4.3　训练流程图

Figure 4.3　Training Flow Chart

测试流程图如下图 4.4 所示：

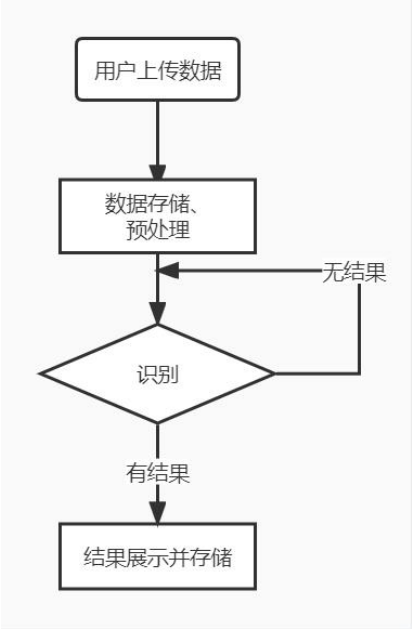

图 4.4　测试流程图

Figure 4.4　Test Flow Chart

4. 结果展示功能

该功能主要是为算法识别出来的汉字进行实时的展示，除了展示外，为了方便用户使用，还可以一键复制识别出来的汉字字符。

5. 错误反馈功能

针对用户上传的问题图片经过识别后，用户可以从展示的结果中对比原始文本图片，若有识别错误的地方。用户可以上传错误给平台，针对用户上传的错误制作新一轮训练网络模型的数据集，然后在训练网络的时候，把这些识别错误的数据集增加权重，使网络训练的时候更多的关注容易出错的地方，这样做的目的是提高网络模型的泛化能力。

4.1.2 非功能性需求

非功能需求是除功能需求之外必要的，主要目的是为了增加用户对该平台使用的体验感以及系统平台的安全性。

系统平台的非功能需求主要包含了性能和安全性，关于性能只要是从系统平台针对于用户请求的反映时间、登录页面之后的渲染时间以及算法对于资源的利用率等。关于算法对硬件资源的利用率主要从两点分析，一是分析 GPU 的利用率是否充分，如果能够充分利用 GPU 资源的话，对于手写汉字图片的识别的效率会快很多，二是分析平台对于 CPU 的利用率。

对于系统平台的安全性需求分析，主要目的是为了保证该系统是否存在安全隐患，有能力应对非法入侵，最重要的是需要保护每个使用该系统平台的个人信息的安全。

4.2 系统平台的设计

上节内容主要是针对平台的需求功能做分析，本节内容主要是针对整个平台的搭建，平台包含的有数据层、算法模型、服务层、应用层，各层之间相互依赖且又独立。

数据层主要是对用户输入作文图片的预处理以及数据存储的功能，主要作用是为算法模型提供数据的支撑。算法模型包括了模型的训练以及部署，是本平台的核心模块，作用就是对作文图片进行识别。服务层包含平台的搭建以及数据的

访问和平台的部署，系统架构图如下图 4.5 所示：

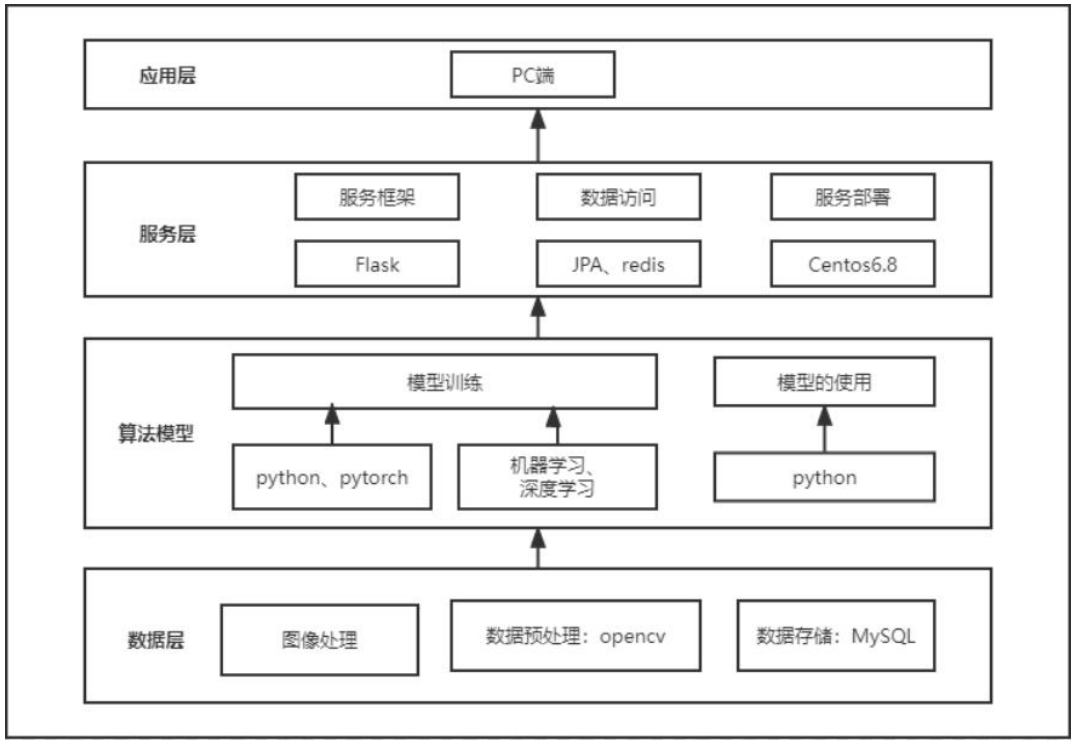

图 4.5　系统架构图

Figure 4.5　System Architecture Diagram

4.3　平台实现

4.3.1　平台开发环境

平台使用的开发语言以及手写汉字识别算法的开发语言都是使用的 Python 语言，数据的存储使用的是 MySQL 工具，网络模型的搭建采用的是 Pytorch 框架搭建，并使用 Flask 框架把汉字识别算法封装成接口，前端页面采用的是 VUE 框架搭建。

4.3.2　登录模块

该模块是平台的最基础模块，平台不支持游客登录，必须先创建自己的账户之后才能登录，用户注册账户之后，平台的数据库会保存用户的基本信息，之后会根据用户输入的账户识别账号和密码是的输入是否符合规定，若不符合规定或者找不到用户输入的账户密码就会提示用户输入错误，需重新输入。在这里采用

的是 Https 协议来提高账户的安全性能。

4.3.3 用户上传模块

该模块的功能主要是用户进行上传文本图片。用户登录网站之后，可以选择自己的作文图片进行上传。不仅在上传模块的时候可以单张作文图片的上传，还可以一次性上传多张作文图片，当上传多张作文图片的时候，平台会对所有的作文图片进行汉字识别，并按照上传作文图片的顺序进行结果展示。这样做的好处就是消除使用者多次上传图片带来的不便性。上传功能如下图 4.7-4.8 所示：

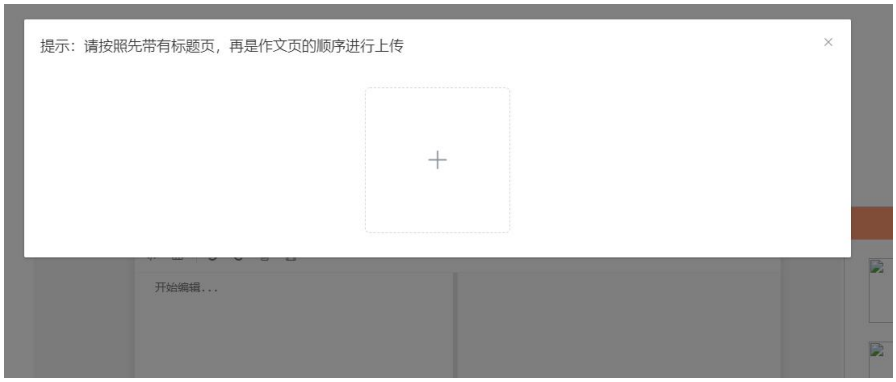

图 4.6 上传图片界面

Figure 4.6 Upload Image Interface

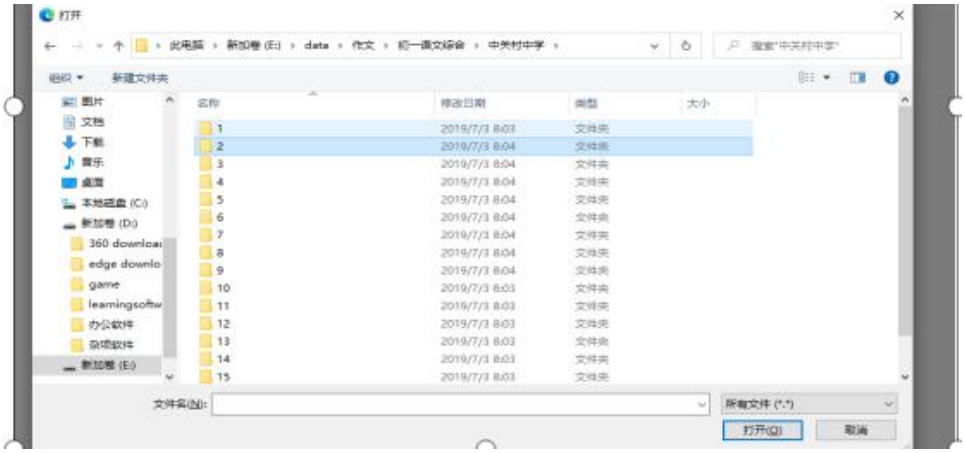

图 4.7 选择图片

Figure 4.7 Select Image

4.3.4 算法识别模块

算法识别模块对于整个系统平台来说是十分重要的，因为该模块设计是否完善决定着系统平台是否完善。用户把自己需要识别的手写汉字图片上传至系统平台后，系统平台会把图片作为输入，输入到算法模型中去，算法模型就会对文本图片进行处理之后就开始对图片进行一系列的特征提取操作了。经过一系列的卷积层和池化层提取手写汉字图片的特征，之后采用文本检测算法把图片中的每一行文本检测出来进行边框标注，使用 PRN 网络结构对标注的文本行的再次进行修正，得到更准确的文本行信息。最后使用卷积层和循环神经网络层标注出来的每一个文本行进行汉字识别，得出结果，并把识别结果展示出来。识别效果如下图 4.9 所示：

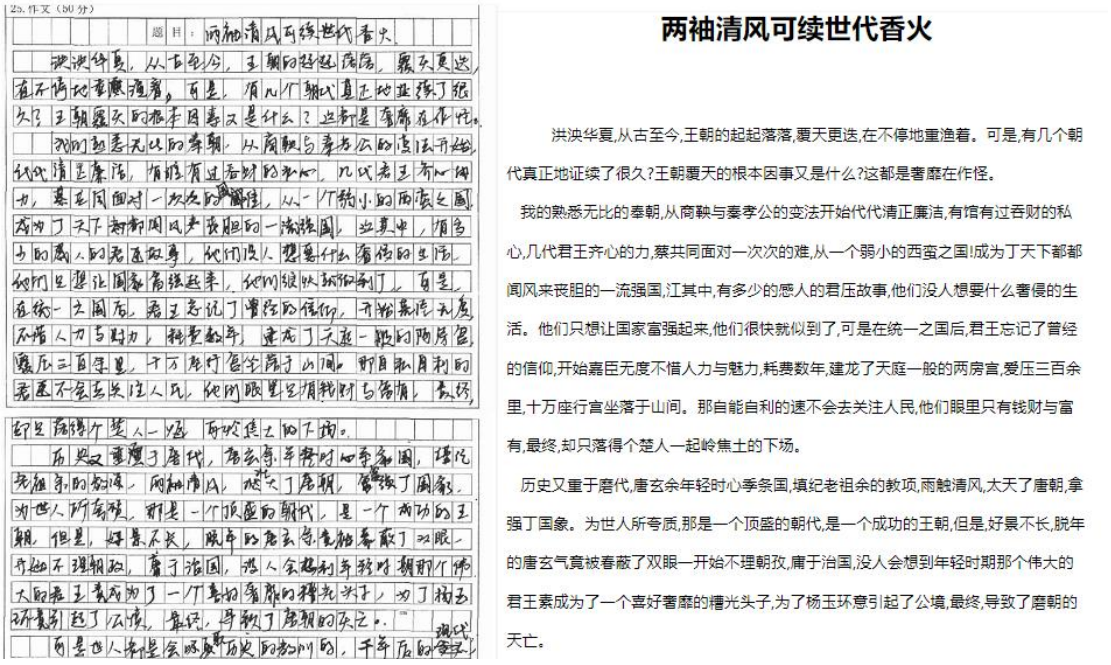

图 4.8 手写汉字图片识别结果

Figure 4.8 Handwritten Picture Recognition Results

4.3.5 日志收集模块

关于日志收集模块主要分为两个模块分别为算法的日志收集以及系统平台的日志收集。算法的日志收集主要是对手写汉字图片进行识别的时候都会把识别时间、识别次数、识别结果等信息保存下来，并且把算法对文本图片识别的结果存储到数据库中，供用户访问历史识别记录，如果识别失败的话会返回识别失败的原因，供开发者以后的查看研究。系统平台的日志收集主要收集用户的功能请求等信息，通过日志收集模块，可以将该平台的使用情况进行具体的分析，为以后的系统更新有很大的帮助。

4.4 本章小结

本章主要是对手写汉字识别平台的各项功能的需求分析以及各功能模块的设计。第一节从平台的功能需求开始分析分析了平台的具体需求，梳理了手写汉字识别平台的各项功能，并根据平台的各方面需求都做了详细的分析。第二节主要是针对功能需求设计了手写汉字识别平台的整体架构。其中包含了四分部，分别是：底层数据层、中间算法层、上层的服务层以及顶层的应用层。之后对每一层结构做详细的介绍，以及各功能模块的设计。最后一节是针对于平台的各模块的设计。

平台功能

技术架构图

第 5 章 系统测试

在系统进行开发的过程中，对于系统的测试是必不可少的，本章节主要是通过各种案例来对手写体汉字识别系统进行测试，包含功能性测试以及非功能性测试两个模块。功能测试主要是对手写汉字识别平台的各种功能是否按照功能的需求进行设计，是否达到要求。关于系统的非功能性测试，主要目的是为了检查系统的安全性能是否达标，以及是否可以给到用户一个良好的体验。

5.1 系统测试环境

5.1.1 硬件环境

本系统的算法模型的训练以及测试平台主要采用的是一台搭载 Nvidia Titan XP 显卡的服务器，因为 Nvidia 旗下的显卡对 cuda 的支持性较高，采用 cuda 可以对网络模型的训练加速，本论文也是采用 GPU 加速训练和测试，使用的 cuda 版本为 9.0，cudnn 版本为 7.05。关于系统的测试采用的是一台客户机，先把手写汉字识别模型在服务器上面训练、测试完成并达到理想效果之后，再把算法模型部署在服务器上面，使用的是 Ubuntu 环境，部署框架采用的是 Python 的 Flask 框架，Python 作为开发语言，Flask 主要用于 Web 端部署接口，深度学习环境采用的是 Pytroch。系统测试时候客户机访问部署在服务器上面的网络模型进行汉字图片识别。算法部署配置的详细参数如下表 1 所示：

表 5.1 算法部署配置表

Table 5.1 Algorithm Deployment Configuration

硬件配置参数	配置
处理器	10 核，2.8GHz，20 线程
内存	32G
硬盘	2TB
显卡	Titan XP
操作系统	Ubuntu18

5.1.2 软件环境

软件开发用到的工具主要有 Python、Flask、Pytorch、MySQL 等，具体的软件开发工具版本如下表所示：

表 5.2 软件开发环境和版本表

Table 5.2 Software Development Environment and Version table

软件	版本
Python	3.7
Flask	2.0
Pytorch	1.71
Torchvision	0.8.2
MySQL	5.8
系统平台	Ubuntu18

5.2 系统的功能性测试

功能测试主要是对手写汉字识别平台的各种功能是否按照功能的需求进行设计，是否达到要求。在测试过程中，把手写汉字识别平台当作一个不清楚内部原理的黑盒子，即只关心平台的各项功能是否实现。对平台的每一个接口进行详细严格的测试，测试结果如下表 5.3-5.6 所示：

表 5.3 用户个人信息相关功能测试表

Table 5.3 Functional Test Forms Related to User Personal Information

功能模块	需要做的操作	模块设计的预期	功能是否完善
测试登录功能是否完善	前端写入账号和密码，点击按钮登录	假若是新用户的话，可以选择登录界面里面的注册账户，用户可以注册自己的账户，同时也可以使用手机号验证码登录。如果不是新用户的话，直接输入账号密码登录	完善

续表 5.3 用户个人信息相关功能测试表

Table 5.3 Functional Test Forms Related to User Personal Information

功能模块	需要做的操作	模块设计的预期	功能是否完善
测试用户个人基本信息设置功能是否完善	可以根据用户的输入来填写或修改用户的基本的信息，例如：姓名、年级、历史上传图片等。	登录账号之后可以点击个人信息，查看自己的基本信息，如果是新用户的话可以选择性填写自己的基本信息。同时也有更改基本信息的功能	完善
测试用户账号修改功能是否完善	输入旧密码、新密码、确认新密码	当用户想要修改密码的时候，选择更改密码按键，然后输入自己的账号，输入成功过后，会提示用户输入两次新密码。修改成功过后必须重新登录。假若用户忘记了账号密码也可以通过用户绑定的手机号信息或者通过自己设置的密保问题进行更改	完善
测试管理员功能是否完善	需要添加管理员账户，并且设置管理员账户信息	该功能是需要系统里面添加管理员，只有管理员能够使用此功能。当管理员登录自己的管理员账号后，可以对所有的用户进行删除或者修改的操作，并且管理员的每次操作都会记录在后台日志中	完善

表 5.4　文本识别功能测试表

Table 5.4　Text Recognition Function Test

功能模块	需要做的操作	模块设计的预期	功能是否完善
测试图片上传功能是否完善	点击上传按钮	用户点击上传按钮后，会显示出上传图片功能框，可以单张图片进行上传，也可以一次进行多张图片的上传。当上传完成后，会显示上传成功的窗口。	完善
测试文本识别功能是否完善	在上传图片后，会产生文本识别的按钮	根据用户上传的图片进行文本识别，如果识别成功会返回识别出来的字符信息。当用户上传多张图片的时候，会根据用户上传图片的顺序，依次返回字符信息。	完善
测试结果展示功能是否完善	识别结果自动展示	当图片识别完成后，会直接显示识别出来的字符信息	完善
测试数据存储功能是否完善	识别结果出来后，点击结果保存按钮	提示保存成功窗口，并且在个人信息中的历史识别记录中可以找到以往识别的信息。只会保存用户点击保存的内容，如果用户没有点击保存的话就默认不保存。	完善
测试查询历史识别记录功能是否完善	点击识别历史记录按钮	会展示出所有的历史识别记录，随便点击一个，就会把此结果给呈现出来。	完善

表 5.5 结果反馈功能测试表

Table 5.5 Result Feedback Function Test Form

功能模块	需要做的操作	模块设计的预期	功能是否完善
识别结果反馈功能	若识别错误，点击上传按钮。若识别正确，不用点击	用户根据识别出来的结果进行判断，若识别错误的话，对错误进行标注之后上传，会显示上传成功。	完善
制作数据集功能	点击数据集上传按钮	此功能只有管理员模式登录才会有。根据用户把上传的识别错误的图片生成新的数据集。	完善

5.3 系统的非功能性测试

关于文本识别的手写汉字识别平台的非功能性测试的目的主要是为了检测该系统平台对于用户的请求，做出的反映时间是否在一定范围内，除此之外还要检测该系统平台的安全性是否达到要求。

1. 系统平台的请求反映时间测试

本系统综合性能指标要求系统页面加载数据并渲染的平均响应时间不超过 3 秒，同时手写汉字算法模型针对于文本图片识别的流程不能超过 2 秒。为了测试系统平台的反映时间不超过规定时间，测试了将近 300 条请求，并根据这些请求算出了平均时间，平台的请求反映平均时间表如下表 5.6 所示：

表 5. 6 平台的请求反映平均时间表

Table 5.6 Platform's Requests Reflect the Timeline

测试功能	测试结果
请求过后页面平均渲染时间	800ms
图片视频算法平均运算时间	1.2s

2.系统的安全性能测试

关于系统的安全性能测试主要目的是为了检测该系统是否存在安全隐患，是否有能力应对非法入侵，并保证用户的个人信息的安全性。测试结果如下表 5.8 所示：

表 5.7　系统的安全性能测试表

Table 5.7　The Security Performance Test Table of The System

测试功能	测试次数	测试结果
用户登录功能	100	通过
用户的权限功能	100	通过
用户访问历史数据功能	100	通过

5.4　本章小结

本章对手写汉字识别平台做了全面的测试说明，并且使用多种案例对手写汉字识别系统进行测试。先介绍了关于手写汉字识别算法所采用的开发工具以及各工具的版本，之后介绍了对该算法平台的开发时候所用到的硬件参数以及开发工具即版本号。接着用各种案例对平台的所有功能进行测试，分别测试了平台的登录功能、信息修改功能、账号密码更改功能、管理员删除账户功能、图片上传功能、文本识别功能、结果展示功能、数据存储功能、历史记录查询功能，通过测试结果得到平台的各功能实现了设计时候的各种需求，能够让用户有一个良好的体验。

第 6 章　总结与展望

6.1　本文总结

本文从手写汉字识别的研究背景、发展现状以及功能需求出发，分析了深度学习在手写汉字识别方向的具体实现，应用深度学习技术实现了基于文本识别的手写汉字识别平台。分析了传统的汉字识别以及现在主流的文本识别模型在针对手写体汉字以及速度上面的不足，应用卷积神经网络和循环神经网络相结合的方法，提出了一个把文本检测和文本识别集一体的神经网络模型，来提升汉字识别的效率。针对手写体汉字提出了一种数据增强方法，来提高网络模型的识别的准确率。在平台方面对中小学生的手写中文作文进行识别，平台集用户登录、文本图片上传、文本图片识别、识别结果展示、识别结果保存等一系列功能均已实现。

本文针对于手写体汉字识别的研究，主要的工作及总结包含以下几个方面：

1. 针对深度学习在图像识别上面的应用，研究了卷积层、池化层、激活函数、网络的训练等一些列知识。

2. 针对于手写体汉字，使用翻转、平移、旋转、不同程度的扭曲等数据增强方法，来扩大数据集，提高网络模型的鲁棒性。

3. 在网络模型设计方面，设计了针对于手写体汉字图片的特征提取网络模型，引入了双向长短期记忆网络，增加网络模型对文本上下文信息的特征提取，从而提高了识别的准确率。

4. 经过大量的对比实验，验证得到本文提出的模型在针对手写体汉字的识别上面识别的准确率和效率都达到了一定的提升。

5. 从平台的功能需求开始分析，设计了集用户登录、文本图片上传、文本图片识别、识别结果展示、识别结果保存等模块。

虽然本文设计出来的手写体汉字识别平台能够实现对手写体汉字的有效识别，但是只是针对于文本图片排版规则的图片。对排版不规范的文本图片进行识别的效果不太理想，平台中的识别算法还需要进一步的研究。

6.2 后续展望

本论文已经实现了对手写体汉别的识别，但是本文所识别的手写体汉字图片都是针对于按照规范从左到右、从上到下书写的，类似于手写作文图片等，对于这类的手写汉字图片本文所设计的算法都可以正常的识别，但是对于非规范的图片，例如票据等排版的图片还是不能够做的十分精准的识别。原因在于，本文算法是基于图片中的文本行训练的，因此对于排版没有固定顺序的文本图片识别效果较差。考虑在接下来的研究工作中，主要以实现在不同排版的文本识别算法为主进行研究。

针对于排版没有固定顺序的文本图片，目前考虑先采用传统的图像处理操作，先把图片中所有的文本块全部分割出来，之后对每个文本块进行单独的识别，该方案是否有效还要进一步进行验证。后期的研究工作主要围绕该方法进行算法的改进，真正的实现手写汉字文本图片的智能识别。

参考文献

白琮,黄玲,陈佳楠,潘翔,陈胜勇.面向大规模图像分类的深度卷积神经网络优化[J].软件学报,2018,29

边亮,屈亚东,周宇.双向特征融合的快速精确任意形状文本检测[J].电子与信息学报,2021,43

陈站,邱卫根,张立臣.基于改进 inception 的脱机手写汉字识别[J].计算机应用研究, 2020 37(04):22-29.

邓丹. PixelLink：基于实例分割的自然场景文本检测算法[D].浙江大学,2018.

邓冠玉. 基于表征学习的不规则场景文本检测与识别研究及系统实现[D].北京邮电大学,2021.

陈鹏飞,应自炉,朱健菲,商丽娟.面向手写汉字识别的残差深度可分离卷积算法[J].软件导刊,2018,17(11).

戴津. 基于 MSER 的文本检测方法研究[D].天津师范大学,2014.

范丽丽,赵宏伟,赵浩宇,胡黄水,王振.基于深度卷积神经网络的目标检测研究综述[J].光学精密工程,2020,28(05):1152-1164.

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.

盖荣丽,蔡建荣,王诗宇,仓艳,陈娜.卷积神经网络在图像识别中的应用研究综述[J].小型微型计算机系统,2021,42(09):1980-1984.

高学,金连文,尹俊勋,黄建成.一种基于支持向量机的手写汉字识别方法[J].电子学报,2002(05):651-654.

高灿. 基于卷积神经网络的脱机手写汉字识别系统研究[D].安徽理工大学,2017.

韩雨晴. 基于数据增强技术的古籍文字识别方法研究[D].厦门理工学院,2021.

黄洋. 基于深度学习的脱机手写汉字识别技术研究[D].重庆邮电大学,2019.

刘崇宇,陈晓雪,罗灿杰,金连文,薛洋,刘禹良.自然场景文本检测与识别的深度学习方法[J].中国图象图形学报,2021,26(06):1330-1367.

刘欢. 基于深度学习的发票图像文本检测与识别[D].华中科技大学,2019.

林恒青.基于深度卷积神经网络的脱机手写汉字识别系统的设计与实现[J].湖北理工学院学

报,2019,35(02):31-34.

李彦冬. 基于卷积神经网络的计算机视觉关键技术研究[D].电子科技大学,2017.

李翌昕,马尽文.文本检测算法的发展与挑战[J].信号处理,2017,33(04):558-571.

彭国雯. 基于深度学习的场景文字检测算法的融合技术研究[D].河南大学,2019.

彭鹏. 基于 CNN 卷积神经网络的车牌识别研究[D].山东大学,2020.

孙巍巍. 基于深度学习的手写汉字识别技术研究[D].哈尔滨理工大学,2017.

杨伟东,田永祥,万峰,王炜.基于深度学习的车载屏幕文本检测与识别研究[J].光电子·激光,2021,32(04):395-402.

袁晨晖. 深度卷积神经网络的迁移学习方法研究与应用[D].南京邮电大学,2020.

杨超杰. 基于深度学习的文本检测与识别技术研究[D].哈尔滨工业大学,2019.

严春满,王铖.卷积神经网络模型发展及应用[J].计算机科学与探索,2021,15(01):27-46.

张珂,冯晓晗,郭玉荣,苏昱坤,赵凯,赵振兵,马占宇,丁巧林.图像分类的深度卷积神经网络模型综述[J].中国图象图形学报,2021,26(10):2305-2325.

张顺,龚怡宏,王进军.深度卷积神经网络的发展及其在计算机视觉领域的应用[J].计算机学报,2019,42(03):453-482.

朱晓慧,钱丽萍,傅伟. 图像数据增强技术研究综述[J]. 软件导刊,2021,20(05):230-236.

周飞燕,金林鹏,董军.卷积神经网络研究综述[J].计算机学报,2017,40(06):1229-1251.

CTPNTian Z, Huang W, He T, et al. (2016) Detecting Text in Natu-ral Image with Connectionist Text Proposal Network. In: Leibe B., Matas J., Sebe N., Welling M. (eds) Computer Vision – EC- CV 2016. ECCV 2016. Lecture Notes in Computer Science, 9912. Springer, Cham.

Dandan Wu. Text recognition technology for natural scenes[A]. The International Society for Applied Computing (ISAC)、Tokyo University of Science、Japan and Cisco Networking Academy.Proceedings of 2021 4th International Conference on Data Science and Information Technology (DSIT 2021)[C].The International Society for Applied Computing (ISAC)、Tokyo University of Science、Japan and Cisco Networking Academy:,2021:6.

Epshtein B, Ofek E, Wexler Y.Detecting text in natural scenes with stroke width transform (2010), in IEEE Computer Vision and Pattern Recognition (CVPR)

Fouda, Y.M. A Robust Template Matching Algorithm Based on Reducing Dimensions.[J]. Journal

of Signal and Information Processing, 2015(6), 109-122.

Gang Wang. Design and Implementation of English Text Recognition System under Robot Vision[A]. Hubei Zhongke Institute of Geology and Environment Technology.Proceedings of 2020 International Conference on Computer Science and Communication Technology (ICCSCT 2020)[C].Hubei Zhongke Institute of Geology and Environment Technology:,2020:7.

Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.

Jiong Zhang. OCR-based Aircraft Maintenance Report Data Structuring[A]. International Association of Applied Science and Engineering (IAASE).Conference Proceedings of 2019 International Conference on Advanced Information Science and System (AISS 2019)[C].International Association of Applied Science and Engineering (IAASE):,2019:5.

Kaite Xiang. Store Sign Text Recognition for Wearable Navigation Assistance System[A]. Asia Pacific Institute of Science and Engineering.Proceedings of 2019 3rd International Conference on Machine Vision and Information Technology (CMVIT 2019)[C].Asia Pacific Institute of Science and Engineering:,2019:10.

Liang Z. and Shi P. A metasynthetic approach for segmenting handwritten Chinese character strings[J]. Pattern Recognition Letters, 2005, 26(10): 1498-1511

Qu Yuanyuan. Neural-network-based Approach to Detect and Recognize Distorted Text in Images with Complicated Background[A]. The International Society for Applied Computing.Proceedings of 2021 3rd International Conference on Advanced Information Science and System (AISS 2021)[C].The International Society for Applied Computing:,2021:7.

致　谢

回首读研的三年，我经历了一段美好的研究所生活，也经历了在北京大数据与认知智能实验室忙碌的生活。我经历了起伏，但也收获了很多。在我低迷的时候，有身边朋友的热心帮助，有也家人的关怀和问候，是陪伴在我身边的家人、老师、朋友给了我信心和鼓励，让我在这三年的时光里面能够克服困难，努力钻研，无论是知识的积累还是思想的成熟，我都达到了新的高度，毕业论文的完稿也为我的硕士生涯画上了一个完整的句号，然而这三年中发生的点点滴滴和美好的回忆仍历历在目。

感谢付老师，在研二进入北京基地后，付老师就一直带领我们进行学术研究，在我们需要帮助的时候，及时给予我们帮助。每次会议都会教会我们要熟练，了解管理和被管理，学习为人处世。在论文创作过程中，付老师向我们提出了非常重要的指导性建议，也给予了我们很多学术上的帮助。无论在我们的生活还是学习中遇到困难的时候，付老师总是会及时的教导我们并且对我们给予帮助。在此，我向老师表示感谢，感谢您在我的研究生学习生涯中的帮助与指导！于老师是我的第一导师，于老师严谨的治学态度和低调务实的做人理念都让我受益匪浅。无论是科研中遇到的各种瓶颈或者是论文书写过程中遇到的困难和疑问，他们都能耐心的给予我指导，帮我走出困境，因此在临毕业之际，我也将谨记两位老师的教诲，成为一个对社会有贡献的人。感谢张老师在这三年内对我和我们班同学的照顾，张老师作为我们班的班主任，在生活上面给与了我和我们班同学不少关心于照料。在校近三年的时间里面，每位老师都在不竭余力的帮助我和我身边的同学，无论是在生活方面还是在学习方面都会给与我们宝贵的建议，我在这里向每位老师发出真诚的感谢。

感谢实验室里的李旭师兄等人在科研工作上面给与给我的支持，在我遇到不懂的技术的时候，李旭师兄会对我一一讲解，直到我全部掌握为止，除此之外，每次遇到新的技术的时候，李师兄总是会带领我进行研究。感谢和我同级的陈永泽、高豪等人。我们一起经历风雨的时光令人难忘。感谢舍友赵钰晨、刘晨旭、

郝振良在生活中对我的关心和理解。

最后，我要感谢我的父母，感谢你们对我无微不至的关心以及耐心教导。只要有你，我就不怕大的风雨，因为我知道，不管我在哪里，还有人会一直关心我，陪伴我，谢谢你在我身后默默的保护。