天津科技大学研究生学位论文

(申请硕士学位)

深度学习在手写汉字识别中的应用研究

RESEARCH ON THE APPLICATION OF DEEP LEARNING IN HANDWRITTEN CHINESE CHARACTER RECOGNITION

专 业 名 称：控制工程

指 导 教 师：张春霞 副教授

研 究 生 姓 名：李龙雪

申请学位类别：工程硕士

论文提交日期：2021 年 6 月

分类号：TP273 学校代码：10057
密级：公开 研究生学号：18811701

深度学习在手写汉字识别中的应用研究

Research on the Application of Deep Learning in Handwritten Chinese Character Recognition

工 程 领 域：控制工程

校内指导教师：张春霞 副教授

企业指导教师：张军保 高级工程师

研 究 生 姓 名：李龙雪

申请学位级别：工程硕士

论文提交日期：2021 年 6 月

课 题 来 源：学校自选题目

学位授予单位：天津科技大学

摘　要

汉字识别在我们的生活和工作中被广泛应用，目前手写汉字的识别技术已经很成熟了，但是对于一些特定场合的应用，比如在文字方向、字体和背景都多样化的书法识别中，仍然存在识别率大打折扣的现象。经过研究发现，识别率降低的原因可能是汉字区域和朝向的检测存在一定的问题，本文针对复杂书法字画的应用场合，设计了一个手写汉字识别系统。经过验证，本系统提高了在复杂书法字画中的汉字识别率，而且该系统还可以应用于类似的复杂背景场合以及多文字的汉字识别中。

汉字检测作为汉字识别的第一步，首先，本文提出对汉字区域进行检测的模型是 AdvancedEAST 网络模型，实验中对主要的神经网络 VGG16 进行了设计，用 RoI 层代替其最大的池化层，能够利用 RoI 层对图像进行多尺度变换，最后对图像统一到相同大小的尺寸，并且在实验中改进了损失函数，使用 Dice 损失和 DIoU 损失函数，与之前的网络模型相比，改进后的网络模型可以达到对文本框更好的标定，对汉字区域进行更准确地预测。其次，在实现对汉字区域准确预测的基础上，实验选取经典卷积神经网络 LeNet-5 对汉字进行识别，在网络模型中加入区域加权系数，实现对特征图的不同区域给予不同的关注度，能够让汉字的轮廓变得更加明显，使得对汉字的识别取得了不错的效果，并且对比了文本框标定前后的实验结果，最后，证实了汉字区域的有效检测可以提升网络模型识别汉字的准确率。

本课题将深度学习应用于不同背景情况下手写汉字的检测和识别中，提高了该场合下手写汉字识别的准确率。该系统可以对多种背景情况下的汉字进行识别，例如拍照、截图、复杂场景、书法体、不同人写的相同汉字等，能够给出汉字的内容、定位位置、识别结果的数目、文本框四个顶点的坐标、汉字所在行的置信度以及输入图片的朝向，对比改进前后的实验数据，识别效果得到明显地改善。但是系统仍然存在许多需要改进的地方，本文针对不足之处提出了相应的解决方案，并对未来的研究进行了展望。

关键字：手写汉字识别；深度学习；神经网络；网络算法；汉字检测；损失函数

ABSTRACT

Chinese character recognition is widely used in our lives and work. At present, the recognition technology of handwritten Chinese characters is very mature, but for some specific occasions, such as calligraphy recognition with diverse text directions, fonts and backgrounds, it still exists the recognition rate is greatly reduced. After research, it is found that the reason for the decrease in recognition rate may be a certain problem in the detection of Chinese character area and orientation. This paper designs a handwritten Chinese character recognition system for the application of complex calligraphy and calligraphy. After verification, the system improves the recognition rate of Chinese characters in complex calligraphy and painting, and the system can also be used in similar complex background occasions and multi-text Chinese character recognition.

Chinese character detection is the first step of Chinese character recognition. This paper proposes the AdvancedEAST network model to detect the Chinese character area. In the experiment, the main neural network VGG16 is designed, and the RoI layer is used to replace its largest pooling layer, which can use RoI. The layer performs multi-scale transformation on the image, and finally the image is unified to the same size, and the loss function is improved in the experiment. Using Dice loss and DIoU loss function, compared with the previous network model, the improved network model can reach the text box is better calibrated, and the Chinese character area is predicted more accurately. On the basis of realizing the accurate prediction of the Chinese character area, the experiment selects the classic convolutional neural network LeNet-5 to recognize Chinese characters, and adds the area weighting coefficient to the network model to achieve different levels of attention to different areas of the feature map. The contours of Chinese characters have become more obvious, making the recognition of Chinese characters achieved good results, and comparing the experimental results before and after the text box calibration, it is confirmed that the effective detection of the Chinese character area can improve the accuracy of the network model in recognizing Chinese characters.

This topic applies deep learning to the detection and recognition of handwritten Chinese characters in different backgrounds, and improves the accuracy of handwritten Chinese character recognition in this situation. The system can recognize Chinese characters in a variety of background situations, such as taking photos, screenshots, complex scenes, calligraphy, the same Chinese characters written by different people, etc., and can give the content of the Chinese characters, the location of the location, the number of recognition results, and the text box The coordinates of the four vertices, the confidence of the row of Chinese characters and the orientation of the input picture are compared with the experimental data before and after the improvement, and the recognition effect is greatly

improved. However, there are still many areas to be improved in the system. This article proposes corresponding solutions to the shortcomings, and prospects for future research.
Key words: Handwritten Chinese character recognition, deep learning, neural network, network algorithm, Chinese character detection, loss function

目　录

1 绪 论

1.1 课题研究背景及意义

汉字识别在我国一直都备受关注，并且在我们的生活和工作中被广泛应用。汉字作为中国的母语文字，在中华民族的悠久历史中，古代对朝代的发展和史事几乎都是通过手写汉字来记载的，手写汉字在历史文化的传承中占有非常重要的地位，实现对手写汉字的识别有利于后人了解中国历史的发展，与此同时，能够起到对中国朝代史事和悠久的传统文化传播的作用。随着时代的变迁和发展，汉字逐渐呈现手写体和印刷体汉字，伴随着印刷体汉字的问世，人们在生活和工作中对手写汉字的采用逐渐减少，取而代之的是印刷汉字[1,2]。相比印刷汉字的存在，手写汉字依然有它存在的意义以及它所独有的性质，手写汉字结构不规范，字体和风格都呈现多变性，即便是同一个人写相同的汉字，汉字的大小也不能做到规范化，这就导致了手写汉字的多样性。我们从入学起就开始学习去认识汉字，学习汉字的书写，这是我们学习中国文化必须去经历的，不管汉字是被印刷体化还是数字化，作为中华儿女必须学会书写汉字。传统汉字录入电脑等电子产品是通过人工键盘输入的，这种方式相比汉字识别录入明显效率很低，而且浪费很多的时间和劳动力，远不及通过机器将信息数字化的速度和质量，而且目前汉字通过识别技术进行数字化的准确率也很高。我们积极地研究手写汉字识别，不是为了让其被电子化的产品所取代，而是为了给我们的生活和工作带来便捷，以便节省不必要的时间，所以无论未来生活有多智能化，手写汉字一定不会在中国消失不见，随着中国在国际上的崛起，手写汉字会出现在世界的各个角落[3-6]。

目前，手写汉字识别是基于深度学习的神经网络模型中一大分支，通过对网络模型的不断构建和改进，其能够对手写汉字的处理的识别达到很快的速度和很高的识别率[7]。在我们的生活和工作中随时随地都有可能会用到手写汉字识别软件，例如手写输入法、古书籍的电子化图书馆、快递地址的识别、对考生试卷的识别、电子签名等，除此之外，我们还将汉字识别系统与语音系统进行结合，有利于帮助老年人和盲人对外界信息的获取[8,9]。我们要努力完善手写汉字识别系统，这样才能更好的传递信息，加快人机之间的信息交换，为人类早日实现人机交互奠定基础。生活中传统的汉字传播局限于纸质化的文件，以前主要是通过人工去识别，然后再经人工录入电脑等，在这过程中因为人工录入的方式会产生误差并且工作效率不高。随着信息时代的到来，机器学习迅速蓬勃发展起来，其中深度学习被普遍地应用到工作和生活中，我们可以利用深度学习把纸质化的汉字进行数字化处理，这样就很大程度地提高了录入效率[10]。人工很轻易地就能看出每个人之间对不同手写汉字的区别，可是中国的汉字文化博大精深会存在很多相似的字形和结构复杂的字体，并且每个人的手写习惯不同，最后导致手写汉字不是很标准化，这就会影响识别结果，从而造成在对手写汉字识别过程会出现偏差，所以对手写汉字的识别仍然存在需要解决的问题，而且应用深度学习去识

别手写汉字是近几年的热点，不得不说在利用了深度学习与神经网络等各种算法的结合技术后，研究者们不断提高手写汉字识别的准确率以及速度，由此衍生出的各种识别软件功能强大，应用于生活和工作中极大地节省了人力和时间，给我们带来了便利[11-13]。在深度学习中对于神经网络层次少的模型具有局限性，在进行训练时对功能复杂的算法和函数的处理能力非常有限，例如 SVM（Support Vector Machine）作为最成功的浅层结构，在训练的过程中它只采用一个浅层线性模式分离模型，其对于复杂的分类处理能力显然是有限的，不容易处理一些复杂的情况，比如自然场景文本、碑文、书法以及字迹模糊的图像等[14]。经过研究者不断的钻研，深度学习的模型由层次少的模型结构发展到复杂的多层次的模型结构，在解决手写汉字识别问题的同时，也产生了不同的难点，例如计算量的增加[15]。

社会需要发展，人类需要进步，对于复杂情况下对手写汉字识别的研究，仍然需要我们不停地找寻方法去突破各种瓶颈。所谓复杂情况可以是字迹模糊的古书籍、自然场景下的背景图像、碑文，还可以是古代不同朝代的书法作品等，其中，书法作为中国汉字的一种形式，由于书法的字体不拘一格，导致它的识别要困难得多。在中国文化发展的历史长河中，书法为中国文明增添了浓郁的一笔，同时书法也承载了对传统文化的流传。不同朝代的书法家为世人留下了很多经典作品，这些作品的呈现有利于我们对历史有更多的了解，以便后人更好的学习我国的传统文化。由于受当时社会发展的影响，古代的书法作品不仅存在于书籍中，还有很多作家会把书法刻在石头上，挥笔泼墨写在墙上，这些情况都不利于我们去收集书法字体的数据集。古代书法字体多样，风格迥异，对书法的识别本来就存在很多困难，再加上年代久远字迹模糊，有些情况书法背景更加复杂，这样一来对书法的识别就更加的困难。书法作为手写汉字的一种特殊形式，中华民族传统文化的多样性离不开书法的构成。在数字化、信息化的时代里，研究者们在应用相关技术去识别手写汉字的同时，也在不断完善算法和改进函数以取得突破性地进步，努力攻克识别过程中遇到的困难，争取使其在目标背景不明确、字迹不清晰等复杂情况下的识别也很准确和有效[16, 17]。

目前，通过利用深度学习的优点，研究者们在手写汉字识别方面取得了很好的效果，但深度学习也存在不足之处，所以在运用时往往离不开与各种模型和算法的结合。除此之外，深度学习很难做到具体问题具体分析，对识别的过程中遇到的问题不能有针对性的去解决，对算法和结构改进具有一定的局限性。手写汉字的识别会受到很多因素的影响，例如汉字复杂多样的背景、汉字的构造、字体的风格等，这些存在的问题都造成了识别的困难，对手写汉字的精确率仍然具有提升空间。手写汉字识别作为深度学习重要的分支，我们应继续努力提高其在复杂情况下的准确率，与此同时，伴随着神经网络在深度学习的应用，计算量也随着网络模型层次的增多而在不断的增加，我们在提高正确率的同时，也要注意网络模型训练的速度。随着数字时代的带来，手写汉字识别具有实用价值和应用前景，研究手写汉字的识别将会拓展字符识别的研究范围, 丰富文字识别的研究内容[18, 19]。

综上所述，手写汉字识别研究的进步，可以使深度学习和神经网络不局限于简单场景的识别，也将为深度学习在图像识别上的研究提供有力的模型基础，所以深度学习在手写汉字识别中的应用是一个值得研究并且有价值的课题。

1.2 汉字识别的发展历程

对于汉字识别的研究，国外起步相对来说较早，大概是从 20 世纪 60 年代开始，国外有许多的研究学者为汉字识别的发展做出了贡献，例如：1966 年，BIM 公司的 Casey 和 Nagy 可以识别 1000 个汉字；1977 年，日本已经拥有能够对汉字进行识别的系统，其作为首个识别系统能够识别 2000 个不同类型的汉字；80 年代初，日本又研究出能够对印刷体进行识别的系统，此系统对汉字的识别能够达到 2300 个，识别水平在当时来说已经是非常高的了。相比国外的研究，国内在 20 世纪 80 年代才开始着手研究汉字的识别，随后汉字识别的研究领域在国内迅速发展起来，这与我们生活中离不开汉字有着密切的关系，研究者对汉字的研究逐渐从单字的识别转变到生活和工作中。汉字识别在国内的发展大概分为 3 个阶段：从 1979 年到 1985 年期间，国内主要是处于对汉字识别的探索阶段，国内的研究学者在数字、符号等识别的基础上展开了对汉字识别的研究，探索前进的道路总是艰难的，研究成果不多，仅有少量的模拟识别软件和系统的诞生，但是此阶段为接下来的进步提供了基础，为后续的研究做好了准备工作；从 1986 年到 1988 年期间，国内的汉字识别研究成果有显著增多，实现了对多种字体的识别，并且识别字数达到 6763 个，识别率高速度快，但是对于真实文本的识别率却很低，这三年的发展阶段为后期汉字识别系统应用于生活和工作奠定了基础，对汉字识别技术的研究必将会进入到实用阶段；自 1986 年开始至今，国内的各大高校纷纷加入对汉字识别系统的研制与开发，这些高校研制的汉字识别的软件在此领域中处于先进水平，在用于商业化的软件中占有大部分的市场，代表着汉字识别技术的发展正处于实用阶段[20-22]。

目前，对汉字的识别研究已经不是仅仅对汉字的识别了，现在衍生出很多识别系统的研究，例如：语义识别、场景识别、古书籍识别、识别和提取表格中的汉字、签名鉴别等，这些相关的识别技术已经应用于我们的生活和工作中，对于汉字识别的研究提高了我们的工作效率，给我们的生活带来了便利，节约了时间。随着信息时代的到来，近几年开始将神经网络应用于汉字识别，并且取得了不错的识别效果，但是在手写汉字识别领域仍然存在需要解决的困难，在后续的研究中，我们可以充分利用神经网络以及各种算法，并且可以根据需求对网络进行改进和完善，在未来的时间里有望拓展更多的领域和更大的空间[23, 24]。

1.3 本文的研究内容

本文的研究内容主要是分为两大部分：一方面是对不同背景情况下的汉字区域进行精准的检测，另一方面，以汉字区域的检测为基础，对多种背景下的汉字进行准确地识别。本文总体来说是关于深度学习在手写汉字识别的应用研究，利用神经网络模型对汉字进行检测与识别，并通过各种算法的使用对网络模型起到改善的作用，最后

对实验数据进行分析与总结。

对于复杂背景手写汉字的研究更加困难，因为在进行取样时需要排除很多干扰，例如光照的强弱、字体结构的多样性、背景被遮挡、凹凸不平等，这些都增加了识别的难度。要想对复杂背景的汉字实现更高的识别率，在进行识别时，应首先将汉字在背景图像中检测出来，即在图像中把汉字找出来，对汉字的检测是实现汉字识别的第一步，所以在本文研究中通过设计网络模型，提升汉字的检测率，为下一步汉字实现更高的识别率提供基础。

在本文中对手写汉字的研究无论是单字、语句还是复杂情况，都将是我们前进道路上的垫脚石，我们可以根据原有网络的不足进行设计，改进完善模型的性能，提高识别的准确率，并且达到更快的识别速度。根据本文对汉字进行的检测与识别中遇到的问题，可以具体问题具体分析，提出进一步地解决方案，借助合适的算法和函数，对网络模型的性能进行提升，以达到高效率的识别结果。

1.4 文章组织结构

本文在归纳并分析了近年来汉字识别领域中神经网络模型的基础上，研究了手写汉字的识别，分析了不同神经网络模型和各种算法函数之间的优缺点，在分析网络模型的基础上，改进其不足之处，实现对手写汉字准确有效的识别。文章的主要结构如下：

第一章，绪论，主要介绍了研究手写汉字识别的背景和意义，汉字识别的发展历程，本文的主要研究内容和文章的组织结构。

第二章，基于神经网络的汉字识别架构，主要对比了在识别过程中常用到的神经网络算法，简单阐述了神经网络模型的基本架构，并且介绍了在汉字识别过程中常用到的神经网络模型：LeNet 网络、VGG（Visual Geometry Group）网络以及 ResNet（Residual Network）网络，根据各网络模型的特点，能够解决汉字识别过程中遇到的不同问题，而且还分别介绍了不同情况下的特征提取算法和中文数据集。

第三章，手写汉字识别系统的设计，主要是介绍系统的功能和系统的结构，并对汉字的特征和图像预处理进行了说明与分析。

第四章，手写汉字检测模型的构建，简单阐述了检测模型中用到的损失函数，其中有 L1 损失函数、Smooth L1 损失函数、IoU 损失函数以及改进后的损失函数，同时介绍了不同场景下的汉字检测，其汉字的检测主要是基于 AdvancedEAST 的网络模型，简述了 EAST 网络模型，并在此基础上进行设计改进得到 AdvancedEAST 网络模型，介绍了网络模型中应用到的函数以及其优缺点，最后对实验结果进行分析。

第五章，手写汉字识别系统的实现，本章包括基于卷积神经网络 LeNet-5 的手写汉字模型的构建，并对其实验过程进行了简单说明,最后对实验数据进行了总结，对比不同的实验结果，证实汉字区域的有效检测可以提升网络模型识别汉字的准确率，同时介绍了实验中所涉及的数据集。

第六章，总结与展望，对全文进行了总结，说明了实验中的不足之处，介绍了本

文中的创新点，并对未来的研究工作进行了展望。

第七章，参考文献。

第八章，攻读硕士学位期间发表论文情况。

第九章，致谢。

2 基于神经网络的汉字识别架构

伴随着深度学习的发展，汉字识别系统也在不断完善，其中包括神经网络的运用、网络模型的构建、损失函数的改进等，这些都为准确高效地识别手写汉字提供了架构支持。上一章中对汉字识别的研究背景、研究意义以及发展历程进行了简单阐述，本章针对实际应用需求，为手写汉字识别模型的构建对神经网络展开介绍，简单阐述了几种神经网络的算法和架构，并且介绍了卷积神经网络（Convolutional Neural Networks, CNN），以及常见的几种卷积神经网络模型，其中包括字符识别网络 LeNet-5、VGG 网络以及残差网络 ResNet，并且对特征提取算法和数据集的选择进行了说明。

2.1 神经网络算法

应用深度学习作为研究的方向就需要了解神经网络，因为神经网络是深度学习中网络模型的基础，是必须具备的框架[25,26]。神经网络是模拟人类大脑中的神经，人类的脑神经可以对信息进行存储以及各神经之间可以协同工作，每个神经的结构都是十分简单的并且能够处理的工作是有限制的，但是当数以亿计的单个神经元构成脑神经时，就可以实现非常复杂的操作。深度学习中的神经网络就像脑神经一样，层数越深处理能力越强，可以实现的功能越强大。神经网络算法在根本上就是一种对逻辑规律进行推理演算的过程，采用神经网络模拟人类的脑神经思维，将目标信息用计算机能够清楚的语言符号表达出来，再将其写成可以让计算机进行操作的指令[27-29]。

在神经网络模型中常用到的算法有 K 最近邻算法（k-nearest neighbours，KNN）、批归一化算法（Batch Normalization，BN）、随机梯度下降算法（Stochastic Gradient Descent，SGD）、超图学习算法（Hypergraph Learning）、反向传播算法（Back Propagation，BP）等，通过表格对各算法进行简单总结对比，如表 2-1 所示。

表 2-1 不同算法之间的优缺点对照表

Table 2-1 Comparison table of advantages and disadvantages between different algorithms

神经网络算法	优点	缺点
K 最近邻算法	无监督学习，最简单的神经网络算法[30]，并且可以计算距离、查找近邻以及进行分类，可以实现分类和回归。	运行速度低，数据集样本很大时，运算的时间会变长，并且效率也会变差。
BN 算法[31,32]	提高模型训练的速度，使其加快收敛，增加模型的稳定性，增加分类功能，防止过拟合的正则化，对模型进行调节参数很方便。	不能使用小批量进行训练，会造成均值和方差有很大的偏差，在 RNN 网络中不能被采用。
SGD 算法	训练速度快，可提前获得损失	计算量大，训练时间长，模

	值，将误差进行泛化，即使训练数据集很大，也可对模型进收敛。	型方差很大，很难设定好合适的学习率[33]。
Hypergraph Learning	通过节点特征相似度让网络自我进化，可以不断地迭代构建模型中不存在的关系属性。图像识别中重叠的部分说明图像具有相同的属性，并且重叠部分越多说明两个目标是同一类的可能性越大[34-36]。	需要为数据集构建简单的图,然后用简单图谱聚类作为基础.
BP 算法	有很强的非线性映射、自学习、自适应和容错能力，可以将训练学习的结果应用于新的样本上，可以重新整理误差，提取更多的关键特征。	算法会出现局部极值，导致网络训练失败， 收敛速度慢，对网络结构的选择会影响算法的效率，对训练样本存在依赖[37]。

2.2 神经网络的架构

神经网络是由人工神经元以及它们之间不同的连接方式所构成，其中有两类特殊的神经元：一类用来接收外部的信息，另一类用来输出信息[38]。这样，能够把神经网络看成是将信息由输入到输出的处理系统。神经网络基本架构如图 2-1 所示。

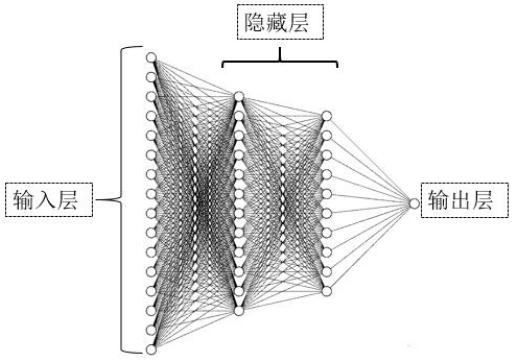


图 2-1 神经网络基本架构图

Fig. 2-1 Neural network basic architecture diagram

2.3 神经网络模型

在深度学习中卷积神经网络 CNN 是最常见的神经网络模型之一，它是根据生物的视觉认知而生成的，并且属于多层的前馈神经网络，和其他普通的神经网络相比，它们都利用反向传播算法对模型进行训练，不同的是卷积神经网络 CNN 的结构能够进行共享权重和局部相连[39, 40]。CNN 的基本框架如图 2-2 所示。

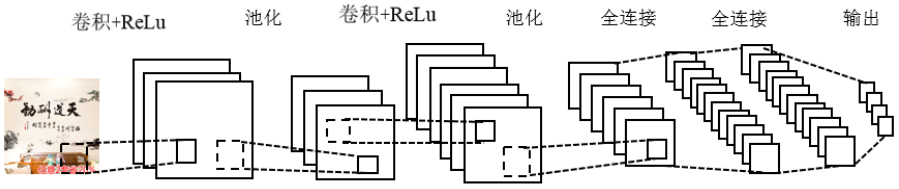


图 2-2 CNN 的基本框架图

Fig. 2-2 The basic framework of CNN

由上图可以看出，卷积神经网络中包括卷积层、池化层和全连接层，其中在 CNN 结构中的卷积层是其所独有的，它的存在是为了获得目标图像的特征，卷积层之间的连接采用的是稀疏连接，这样的连接方式可以增强网络模型的稳定性和泛化性，图中 ReLu 是卷积层的激活函数，池化层被称为下采样层，其出现在卷积层后，这样的连接结构也是 CNN 所特有的，是为了降低图像的分辨率，减少参数，简化模型的计算量，在 CNN 中这种卷积层和池化层交替出现的结构可以提取输入图像的平移不变特征，全连接层的目的就是为了把卷积层提取的特征图进行融合，减少特征对目标进行分类的影响，最后输出对目标的分类[41]。

下面将对几种常见的卷积神经网络进行简单的介绍，包括 LeNet-5 网络、VGG 网络和 ResNet 网络。

2.3.1 LeNet-5 网络

LeNet-5 网络作为用于图像分类的卷积神经网络，最初的 LeNet 网络是由 Yann LeCun 在 1989 年构建的[42]，它的结构中有 2 个卷积层和 2 个全连接层，而 LeNet-5 网络是在 1998 年构建的，它是在 LeNet 网络的基础上加入池化层和沿用学习策略构建的卷积神经网络，其结构同时定义了卷积神经网络的基本结构，用它来识别手写汉字能够达到很高的效率[43]。对于 CNN 网络而言，它可以很好地采用来自目标图像的内部信息，网络内部结构通过局部连接构成，并且在提取特征时进行权重共享，这就使得网络的参数不多，计算量不大。LeNet-5 网络卷积过程如图 2-3 所示。

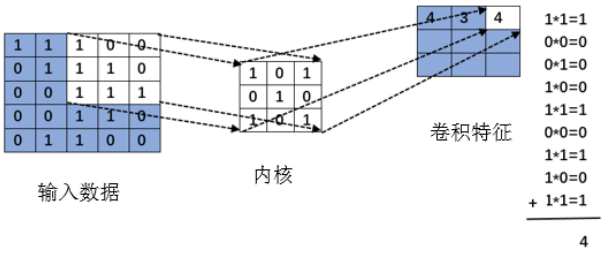


图 2-3 LeNet-5 模型的卷积过程

Fig. 2-3 Convolution process of LeNet-5 model

上图中的卷积过程是以 5×5 的图像为例，滤波器大小为 3×3 的卷积核进行运算[44]。LeNet-5 网络不包括输入层在内，一共有 7 层，分别是 2 层卷积层、2 层池化层和 3 层全连接层，任意一层都能够训练多个特征图，特征图经过卷积运算构成滤波器，

以此来获得输入特征，对于每一张特征图经过不同的卷积层都可以进行特征提取。当网络层数越来越多，图像的尺度也在不断缩小，伴随着图像通道的数量也在增加。

LeNet-5 的 7 层结构模型中最重要的部分是卷积运算，数学公式$(f*g)(n)$称为 f，g 的卷积，其中连续卷积的公式如（2-1）所示。

$$(f*g)(n)=\int_{-\infty}^{\infty} f(\tau)g(n-\tau)d\tau \qquad 式(2-1)$$

离散卷积公式如（2-2）所示。

$$(f*g)(n)=\sum_{\tau=-\infty}^{\infty} f(\tau)g(n-\tau) \qquad 式(2-2)$$

一般情况下，将神经网络中上述形式的函数统称作卷积。

2.3.2 VGG 网络

VGG 网络的结构是相较于 AlexNet 进行改进的[45]，它利用 3*3 的卷积核代替 AlexNet 中的 11*11、7*7 和 5*5 的卷积核，采用小卷积核堆积的形式，多层非线性层能够通过增加网络的深度来确保学习更复杂的模式，对于用到的参数会很少。VGG 网络是由卷积层和全连接层构成的深度卷积神经网络，通常被用来对图像进行特征提取[46, 47]。VGG 的 VGG16 和 VGG19[48]两种结构被广泛用来进行特征提取，它充分体现了卷积神经网络中功能和层数之间的联系。

VGG16 它的规模是 AlexNet 的 2 倍，在 2014 年的 ILSVRC 图像分类竞赛中获得第二名，而且它的泛化性很强，其结构主要有 5 段卷积层，其中每一段中包含 2~3 个卷积层，任意一层的卷积层都拥有相同数目的卷积核，其数目分别为：64,128,256,512,512，由3*3的卷积核和2*2的最大池化层重复叠加形成的，这种小型滤波器组成卷积层的效果比5*5的卷积层要好，采用 3*3 的卷积核代替 5*5 的卷积核如图 2-4 所示。

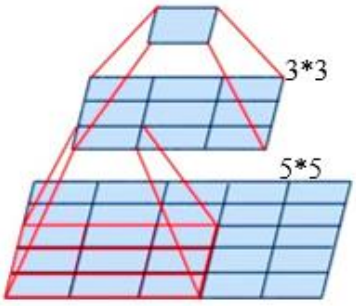


图 2-4 3*3 的卷积核代替 5*5 的卷积核

Fig. 2-4 3*3 convolution kernel instead of 5*5 convolution kernel

在保证拥有一样的感受野的情况下，增加网络的深度，相对来说可以增强网络的性能，而且在这过程中不会改变特征图的分辨率。VGG 网络对图像每进行一次卷积，特征图就变为原来的 1/4，在网络训练时卷积层相对来说消耗的时长比较多，但是在网络的全连接层上主要是消耗参数量，这就导致 VGG 在消耗参数的同时，占用了很

大的内存，使计算量变大，尤其是在第一个全连接层使用了大部分的参数。总体来说，VGG 网络结构简单，而且结构中小滤波器 3*3 的组合比大滤波器 5*5 的卷积层要好，主要是其网络的性能随深度的加深而有所提升。

2.3.3 ResNet 网络

ResNet 网络在 2015 年的 ILSVRC 图像识别比赛中获得第一名，它的规模分别是 AlexNet，VGG16 的 20 倍和 8 倍，其在网络中加入了直连通道，这样的结构可以加快网络的训练，并且模型的准确率也有所提高。ResNet 网络有 18、34、50、101 和 152 这 5 中不同深度的结构，该网络可以解决网络训练退化的问题，网络层数的增加，在解决网络深度带来的负面作用时，也提高了网络的性能，在网络中加入残差架构后网络不会出现退化的现象[49]。ResNet 网络是由基本的网络结构组成的，其基本结构如图 2-5 所示。

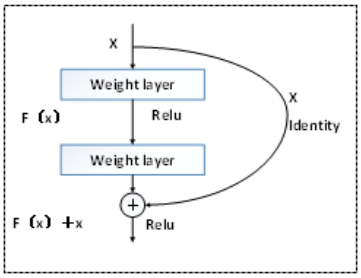


图 2-5 ResNet 网络基本结构图

Fig. 2-5 Basic structure diagram of ResNet network

在 ResNet 网络结构中，需要注意的是残差函数，它的作用是让网络变得更加方便优化，而且当增加网络深度时也可以提高准确率。当输入x经过卷积 Relu 后，将输出的值设为$F(x)$，残差网络会将输出值$F(x)$与输入值x进行相加，此时最终的输出值记为$H(x)$。经过梯度的增加，相比当前梯度，前一层的梯度就会增加输入x的梯度，更深层的梯度可以无障碍的通过，使浅层的参数也可以进行准确有效的训练，这样就可以得到残差模块，其能够用来缓解梯度消失的问题[50]。在 ResNet 网络最初的结构中，它的残差模块中包含 2 个卷积层、1 个跳跃连接、激励函数以及 BN 算法，其残差模块的流程图如 2-6 所示。

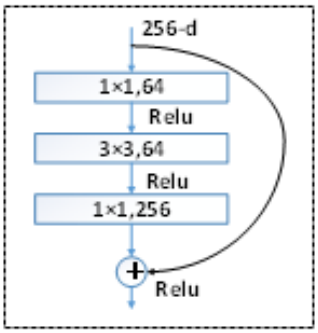


图 2-6 残差模块的流程图

Fig. 2-6 Flow chart of the residual module

ResNet 网络结构经过改变映射方式，使得拟合残差的难度降低了。当残差映射$F(x)$和网络层输入向量x维度相同时，网络层输出向量$H(x)$表述为公式如(2-3)所示。

$$H(x)=F(x)+Wx \qquad \text{式(2－3)}$$

其中W表示权重值。

2.4 特征提取算法

特征提取是要在目标图像中获得重要的信息，排除不必要的因素，以提升对目标分类的准确性。不同的特征提取方法有不同的效果，在图像识别中常用到的特征算法有 LBP、HOG、SIFT、SURT、Gabor、Gist、小波特征等，图像特征包括颜色特征、纹理特征、空间关系特征和形状特征等，其中颜色特征、纹理特征和形状特征反映的都是图像的整体特征，颜色特征描述图像中景物的表面特征，因为颜色对图像的尺度和朝向的改变感受不明显，因此它不适合对目标进行局部区域的捕捉。纹理特征虽然和颜色特征都是描述图像中景物的表面特征，但是其要统计目标区域内的多个像素，而且它具有旋转不变性[51]。空间关系表示目标图像之间在空间中的关系，其对图像的位置和大小的变化感受都很明显，因此它不可以有效地展现出图像中的信息。形状特征分为轮廓特征和区域特征两类，它能够有效地对图像中的目标信息进行检索。

2.4.1 LBP 算法

LBP 算法(Local Binary Patterns)，可以很好的捕捉到图像中的细节，它是属于图像纹理特征提取算法，是一种局部特征[52]。LBP 算法的核心是计算每个像素和周围像素相比的大小关系，它的优点包括运算速度快，对旋转和灰度的变化都可以不改变特征，并且对光照有很强的鲁棒性。LBP 算法定义的公式如（2-4）所示。

$$LBP(x_c,y_c)=\sum_{p=0}^{p-1}2^p s(i_p-i_c) \qquad \text{式(2－4)}$$

公式中，(x_c,y_c)是中心像素，i_p是相邻像素的亮度，i_c是该点的灰度值，其中p是该点邻近像素的个数。其中$s(x)$是符号函数，用来计算相邻位置像素与中间位置像素的差值，其公式定义为（2-5）所示。

$$s(x)=\begin{cases}1, & x\geq 0\\ 0, & x<0\end{cases} \qquad \text{式(2－5)}$$

LBP 算法计算过程如图 2-7 所示。

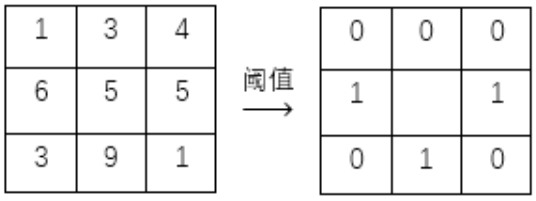


图 2-7 LBP 算法计算过程图

Fig. 2-7 LBP algorithm calculation process diagram

对于邻域像素顺序默认为水平方向左侧为起点，逆时针旋转，由此得到一个二进制值为 00010101，通过计算将其转换成十进制值为 21，即这个中心点(x_c, y_c)的 LBP 值为 21。经过二进制与十进制的转换后，就能够用一个数来表达像素点和邻域的差值关系，因为 LBP 记录的是像素点与邻域像素的差值关系，所以光照变化引起像素的值增加或减少并不会改变 LBP 的大小，特别是在局部的区域，我们可以认为光照对图像造成的像素值变化是单向的，所以 LBP 可以很好的保存图像中像素的差值关系，可以进一步将 LBP 做直方图统计，而这个直方图可以用来作为纹理分析的特征算子。

2.4.2 HOG 算法

HOG 算法（Histogram of Oriented Gradient），其被称为方向梯度直方图，它可以表达出局部区域的特征，其算法存在的意义是能够通过梯度的方向密度准确地表现出图像中局部目标的特征和结构[53]。HOG 特征是在图像的局部区域上操作，计算每个像素的梯度，捕获轮廓信息的同时还能够减弱光照的干扰，它可以对目标大小的多变和光照强度的变化维持很好的不变性。

采用 HOG 算法首先将目标图像分成小的连通区域，这些小的连通区域我们称之为细胞单元，接下来通过采集连通区域中所有的像素点，这样一来就可以获取目标图像特征的方向梯度直方图，通过它们的组合能够得到特征描述器。对方向梯度直方图进行规范化可以使图像目标对光照变化有更好的适应，为了优化 HOG 算法，我们可以将在小连通区域里得到的方向梯度直方图，放入图像中更大的连通区域中进行规范化处理，计算局部直方图在大的连通区域里的概率分布，根据概率值对其进行相应程度的规范化处理。因为 HOG 算法是在图像的小连通区域上进行的，所以它可以对图像目标尺度的多变和光照强度的变化维持很好的不变性，当然上述两种变化在小的连通区域上不会出现，它们只会出现在图像更大的连通区域上。HOG 算法对图像处理的流程图如图 2-8 所示。

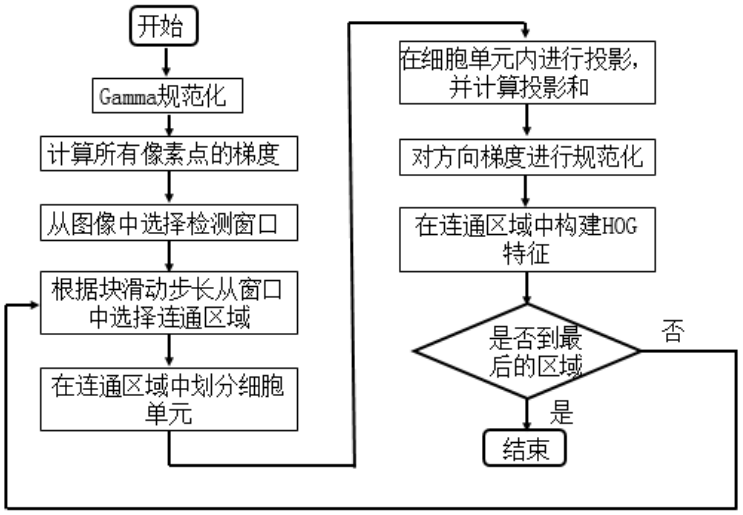


图 2-8 LBP 算法计算过程图

Fig. 2-8 LBP algorithm calculation process diagram

HOG 特征提取的过程会对图像进行分割，将其分割成几个连通的小区域，然后得到这些小区域中的每个像素点的方向梯度，计算其在所有方向上的直方图，每 4 个相邻的小区域组成一个区间，将它们的特征向量加起来，再对目标图像进行检测识别，将全部区间的特征组合起来，这样能够获取目标图像的特征。

当对图像进行规范化处理后，能够降低光照对图像的干扰，因为在图像纹理强度中，局部的表层曝光所占的比率很大，通过压缩处理可以降低图像局部的光照变化和阴影。γ压缩公式为（2-6）所示。

$$I(x,y)=I(x,y)^{\gamma} \qquad \text{式(2－6)}$$

根据图像横纵坐标的梯度方向值，能够得到所有像素点的梯度，并且通过求导能够捕获图像的轮廓，降低光照的干扰。图像中像素点(x,y)的梯度公式为（27-）所示。

$$\begin{aligned} G_x(x,y)&=H(x+1,y)-H(x-1,y)\\ G_y(x,y)&=H(x,y+1)-H(x,y-1) \end{aligned} \qquad \text{式(2－7)}$$

表达式中的$G_x(x,y)$，$G_y(x,y)$，$H(x,y)$分别表示输入图像中像素点(x,y)处的水平方向和垂直方向的梯度以及像素值。像素点(x,y)的梯度幅值公式为（2-8）所示。

$$G(x,y)=\sqrt{G_x(x,y)^2+G_y(x,y)^2} \qquad \text{式(2－8)}$$

而梯度方向的表达式如（2-9）所示。

$$\alpha(x,y)=\tan^{-1}\left(\frac{G_y(x,y)}{G_x(x,y)}\right) \qquad \text{式(2－9)}$$

梯度图像对目标的轮廓进行获取，消除了很多不需要的信息。对于彩色图像其 3 个通道的梯度都能够被计算出来，对于这 3 个通道而言，最大梯度的像素点，其角度也是最大的。

2.4.3 SIFT 算法

SIFT 算法（Scale-invariant feature transform），它能够对目标尺度大小的变化和角度位置的变化维持很好的不变性，属于一种图像局部特征提取算法[54]。SIFT 算法能够解决目标图像的角度多变、尺度不一、光照干扰、噪声、杂物场景、目标遮挡等问题，我们可以利用 SIFT 算法将两张不同拍摄角度的图片中相同部分匹配出来。SIFT 算法的原理就是在相邻的尺度空间上对比所有的采样点与相邻点的大小关系，所谓特征点就是其在尺度空间上的极值点。如图 2-9 所示，中间的采样点与其在同一尺度空间上的 8 个相邻点和上下相邻位置的尺度空间上的 18 个点比较，这样就可以保证采样点是否为所在空间区域上的极值点。如果中间的采样点在这 26 个点中是最大或者最小

值时，则称此采样点是所在空间区域上的特征点。

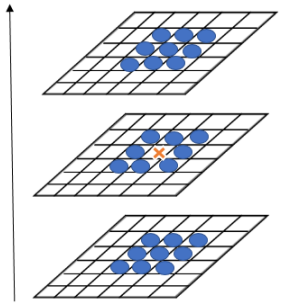


图 2-9 SIFT 算法上的特征点图

Fig. 2-9 Feature point map on SIFT algorithm

高斯卷积核是唯一可以产生多尺度空间的核，所以在 SIFT 算法中可以通过高斯模糊来实现空间上的获取。高斯模糊又称为高斯平滑，属于低通滤波器，一般情况下采用它来降低图像中的噪声，即让图像和正态分布做卷积运算，可以起到模糊图像的作用。高斯分布也叫正态分布，其 N 维空间正态分布G_r公式如（2-10）所示。

$$G(r)=\frac{1}{\sqrt{2\pi\sigma^2}^N}e^{-\frac{r^2}{(2\sigma^2)}} \qquad 式(2-10)$$

公式中，r 是模糊半径，σ 是高斯函数的标准差，σ 值越大，图像越模糊。假设二维模板大小为 a*b，则模板上的元素(x,y)对应的高斯计算公式为（2-11）所示。

$$G(x,y)=\frac{1}{2\pi\sigma^2}e^{\frac{\left(x-\frac{a}{2}\right)^2+\left(y-\frac{b}{2}\right)^2}{2\sigma^2}} \qquad 式(2-11)$$

在二维空间中，高斯曲面的等高线是从中心开始呈高斯分布的同心圆，如图 2-10 所示为二维高斯曲面。图像中每个点都参与构成卷积矩阵，并且和原始图像做变换。所有像素点的值是通过相邻像素的值加权平均得到的，原始像素的值有最大的高斯分布值，相邻像素点距离原始像素越远，其权重越小，通过这种形式模糊图像可以很好地保留边缘效果。

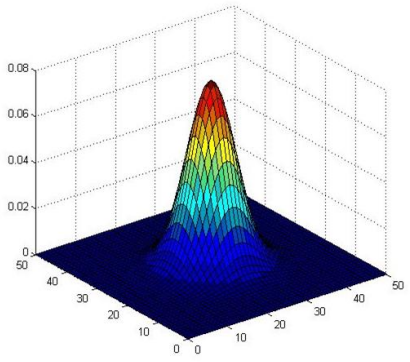


图 2-10 二维高斯曲面图

Fig. 2-10 2D Gaussian surface plot

2.5 数据集的选择

2.5.1 手写汉字数据集

汉字所呈现出来的多样性和背景存在的复杂性，给我们选择合适的数据集带来了困难。对于简单背景的手写汉字的公开数据集有来自北京邮电大学识别实验室收集到的 HCL2000 数据集和中科院自动化研究所模式识别实验室收集的 CASIA-HWDB 数据集。CASIA-HWDB 数据集与 HCL2000 数据集相比，其数据集中的手写汉字更具有说服力，手写汉字的字体随意而且多种变化，在识别上具有一定的难度。对于简单背景下手写汉字的识别，在数据集上的选择均是 CASIA-HWDB1.1 数据集，该数据集由 300 个人手写而成，其中包括了 3755 个一级文字以及 171 个英文数字符号，按照训练集和测试集 4:1的比例，对网络进行训练和测试。手写汉字数据集的示例样本，如图 2-11 所示。

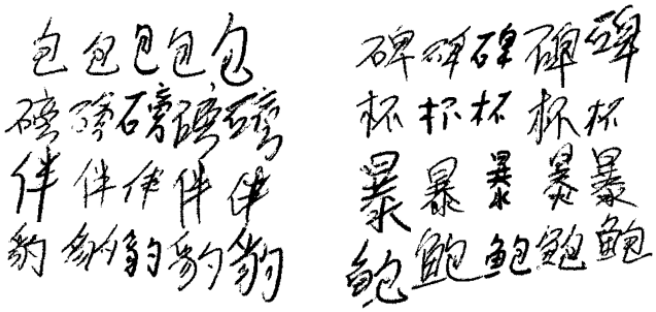


图 2-11 简单背景的汉字样本图

Fig. 2-11 Sample Chinese Characters with simple background

2.5.2 文本检测数据集

以前的复杂背景数据集中 SVHN [55]和 SVT [56]都是采用的 Google 街景的图像[57]，目前，虽然存在很多的数据集，但是其数据集的数目都不足以训练深层的神经网络，例如 MSRA-TD500 数据集[58]中的图像只有 500 张。本实验中复杂场景数据集的选择是中文自然文本数据集 CTW（Chinese Text in the Wild），满足本实验对数据集的要求，在该数据集中具有 1,018,402 个中文字符和 32,285 张图像，数目比此前的数据集都要多，为更加复杂深度学习的训练提供了数据基础[59]。数据集中的手写体汉字示例如图 2-12。

图 2-12 复杂背景的汉字样本图

Fig. 2-12 Chinese character sample map with complex background

数据集中的图像不存在任何特定目的的偏好，其中样本图像中存在水平面、光照强度不同、模糊不清被遮挡、曲面凹凸不平等不同的背景。该数据集对每张图像都标记了中文汉字，并且数据集还标记了汉字的边界框，以及对汉字的字体、背景，是否清楚等情况进行了属性说明。

2.6 本章小结

本章介绍了在深度学习的手写汉字识别系统设计中常用到的神经网络模型，首先简单阐述了模型中会用到的神经网络算法和神经网络的基本架构，然后对几种常用到的卷积神经网络模型的特点进行了总结，其中包括字符识别网络 LeNet-5、VGG 网络以及残差网络 ResNet，然后介绍了特征提取算法，并对模型训练时常采用的数据集进行了简单说明，包括手写汉字数据集和文本检测数据集，本章为手写汉字识别网络模型的构建提供了思路，文章后续会由此章展开对手写汉字识别网络模型进行设计，实现对手写汉字准确高效地识别。

3　手写汉字识别系统的设计

本文所设计的手写汉字识别系统应对现代的深度学习潮流,识别系统以卷积神经网络模型为主体,根据不同的需求对其进行设计和改进，并且融合特征提取算法和损失函数，实现对汉字区域有效地检测和准确地识别。本章简单说明了系统设计的功能和结构，介绍了对图像的预处理，其中包括规范化、降噪和二值化，并对汉字的特征进行了阐述。

3.1 系统总体方案设计

本节从系统的功能和结构展开对手写汉字识别系统的介绍说明。随着互联网时代的发展，手写汉字识别系统在我们的生活和工作中被广泛应用，给我们的生活和工作带来了很多便利，而且近年来深度学习的出现，为解决识别时遇到的问题提供了技术支持，促进了汉字识别领域的飞速发展[60]。系统功能中给出了系统对汉字识别所具备的功能，系统结构给出了识别系统的整体框架，并对识别流程进行了简单说明。

3.1.1 系统功能

根据系统实现对手写汉字的通用识别功能，我们将其系统分成五个方面:（1）图像处理;（2）手写汉字背景区分;（3）汉字区域检测;（4）汉字识别;（5）输出结果。手写汉字通用识别系统的功能框架如图 3-1 所示。

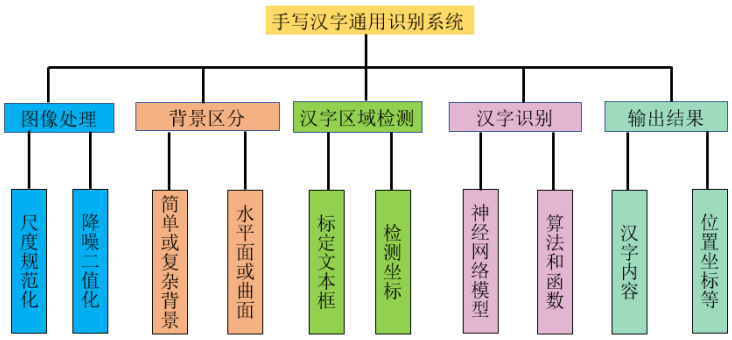


图 3-1 通用识别系统的功能框架图

Fig. 3-1 The functional framework diagram of the universal recognition system

对于输入的目标图像进行处理，主要包括图像的规范化、降噪和二值化；其次，手写汉字背景的区分主要是在网络模型中采用二分类损失函数，实现对背景和汉字区域精准的分类，判断汉字背景图片是否复杂，图像角度是否水平以及汉字区域的范围；对汉字区域进行准确地标定文本框，并输出检测坐标；进行汉字区域检测后将结果输入识别模型中，利用神经网络模型、算法和函数对汉字进行识别；最后输出识别的结果，在反馈结果中包括汉字内容、文本框坐标、置信度、图像朝向以及识别定位时的位置数组等信息。

3.1.2 系统结构

要实现手写汉字通用识别，首先系统结构应该是多模型的，在系统中包括汉字的

检测、识别、特征提取和特征融合等不同的网络和算法，为了实现汉字的检测识别，在对系统进行训练时会用到不同的数据集，进行数据增强和优化学习率等，使其实现对汉字准确高效地识别，其识别系统的基本框架，如图 3-2 所示。

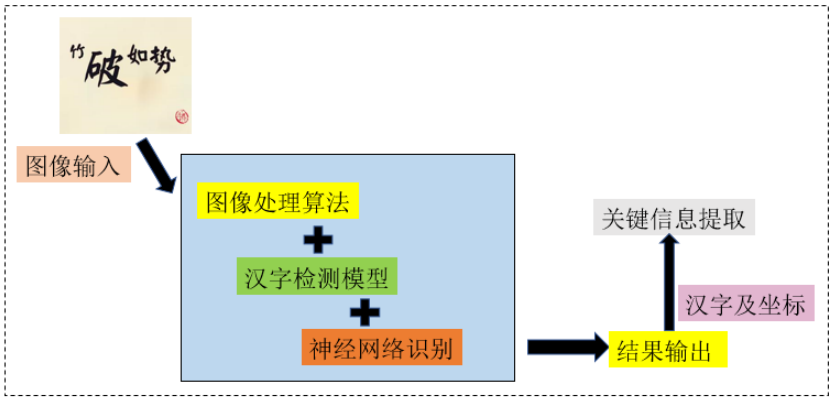


图 3-2 识别系统的框架图

Fig. 3-2 Frame diagram of the recognition system

在识别系统中，对于汉字区域的检测采用改进的 Advanced EAST 的网络框架，因为它不仅可以检测简单背景和复杂背景中的汉字区域，对于曲面汉字的检测也可以达到准确有效；对于图像的特征提取和特征融合，主要采用神经网络模型中的提取和融合算法[61]；汉字识别的过程就是把处理好的二维图像转化为一维的字符串，其中包括对标准汉字识别和不规则的汉字识别的算法。后续的研究中可以根据汉字识别内容的反馈，提取我们所需要的关键信息，比如消费小票、学生的试卷等 ，我们可以根据需求提取消费的金额、商店名、消费商品类型以及时间地点、学生试卷的得分、特定题目的答案等，对信息处理实现自动化会给我们的生活和工作节省不必要的时间和带来很多的便利。

3.2 汉字特征

汉字特征分析在识别和模式匹配中占有重要的位置，并且在深度学习中也是图像处理的重要环节。首先汉字的特征提取可以帮助网络模型在检测文字时，使目标图像简单化，减少计算量，能够明显的显示出目标文字的轮廓；其次对手写汉字的提取和分类，第一步就是需要把文字图像进行二值化，二值化后的图像最关键的一点就是简化了模型训练的过程，提高了训练的速度[62]。假设图像上有 a 个像素点，二值化后图像的像素点为 A×B，其像素点$T(x,y)$表达公式为（3-1）所示。

$$T(x,y)=\begin{cases}0, & \text{像素点为白色}\\ 1, & \text{像素点为黑色}\end{cases} \qquad \text{式}(3-1)$$

黑白二值比 R 的表达公式为（3-2）所示。

$$R=\frac{T(x,y)=1\text{的像素点个数}}{T(x,y)=0\text{的像素点个数}} \qquad \text{式}(3-2)$$

每个汉字的构成都离不开笔画，学习书写汉字时都是从点、横、竖、提等开始的，

所以汉字的最小连笔单位是笔画。不同的人有不同的手写习惯，这就导致了手写汉字风格的多样性。笔画特征可以表示出汉字的局部区域特点，能够很好地表现出人与人之间书写的不同[63]。汉字的结构特点是指笔画之间的组织形式，每个汉字笔画之间不同的组织，就会产生不同特点的汉字，所以手写汉字的结构特点可以很好的显示出每个人书写的习惯[64]。综上，对汉字的特征提取我们可以从笔画和结构两大特点进行分析。

3.2.1 汉字的笔画特征

简单的点、线构成了复杂多样的汉字，首先分析笔画特点，其中笔画长度a_1，不同的人手写汉字的习惯不同，在书写笔画的长短和宽窄上肯定不一样。假设笔画的长度为L，统计其笔画长度上从头到尾的像素点个数，像素点的集合为$B=(b_1,b_2,...,b_n)$，利用像素点的总集合数来统计笔画的长度特征，则长度$a_1=N$。

对于笔画的宽度记为a_2，笔画像素点的长度集合为$B=(b_1,b_2,...,b_n)$，像素点的宽度集合为$W=(w_1,w_2,...,w_n)$，则笔画旳宽度表达式如（3-3）所示。

$$a_2=\frac{1}{n}\sum_{i=1}^{n}w_i \qquad 式(3-3)$$

每个人在书写汉字时，用力的程度是不同的，即使同一个汉字程度也有深浅。用力程度的变化在一定的范围，可以体现出了每个人手写汉字的习惯。力度a_3是利用所有像素点的真实笔画宽度和加权平均后宽度的方差来表示，其表达式如（3-4）所示。

$$a_3=\frac{1}{n}\sqrt{\sum_{i=1}^{n}(w_i-a_2)^2} \qquad 式(3-4)$$

在上面的表述中提到了书写力度的不同，利用手写汉字的竖向和横向笔画表示平均力度，手写汉字中全部的横画宽度集合$W_h=\{h_1,h_2,...,h_n\}$，全部的竖画宽度集合为$W_s=\{s_1,s_2,...,s_m\}$，则汉字的平均力度a_7公式如（3-5）所示。

$$a_7=\frac{\frac{1}{n}\sum_{i=1}^{n}h_i}{\frac{1}{m}\sum_{i=1}^{m}s_i} \qquad 式(3-5)$$

另外，不同手写汉字的字体宽度也是各有不同，相同笔画不同的人去书写，也会导致汉字字体大小不一，字体宽度可以作为识别手写汉字的一个因素。手写汉字宽度的变化率F公式如（3-6）所示。

$$F=\sqrt{\frac{1}{a}\sum_{i=1}^{a}\left(d_i-\frac{1}{a}\sum d_i\right)^2} \qquad 式(3-6)$$

其中，i表示圆心，d_i表示圆内像素点最大半径。

根据笔画特征中像素点开始的坐标位置和终点的坐标位置，我们可以预测得到每

个像素点的行动轨迹，这个轨迹就是笔画的走势，则像素点的轨迹a_4定义如式（3-7）。

$$a_4=\frac{y_n-y_m}{x_n-x_m} \qquad 式(3-7)$$

表达式中，(x_m,y_m)表示像素点在笔画上开始点的坐标，而(x_n,y_n)则表示像素点在笔画上最终点坐标。

同一类型的笔画其轨迹走向大致是相同的，但是这不是印刷字体可以做到一成不变。对于每个人去书写会存在笔画角度问题，这就需要计算笔画开始点和笔划最终点的平均角度a_5，其表达式如（3-8）所示。

$$a_5=\frac{1}{m}\sum_{i=1}^{m}\frac{y_{in}-y_{im}}{x_{in}-x_{im}} \qquad 式(3-8)$$

其中，(x_{im},y_{im})和(x_{in},y_{in})分别为笔画轨迹走向的开始与终点的坐标。

每个人手写汉字习惯的不同，以及构成汉字的笔画具有多变性，其汉字笔画的曲率a_6也是不断在变化，对于检测和识别汉字区域造成了许多细节上处理的难度。笔画像素点的集合为$B=(b_1,b_2,...,b_n)$，假设其中b_i的坐标为$(x_{\varepsilon,}y_{\varepsilon})$，则笔画的曲率变化如公式（3-9）所示。

$$a_6=\frac{\sum_{e=1}^{n-1}\sqrt{(x_{e+1}-x_e)^2+(y_{e+1}-y_e)^2}}{\sqrt{(x_n-x_1)^2-(y_n-y_1)^2}} \qquad 式(3-9)$$

3.2.2 汉字的结构特征

中国汉字字体繁多，其结构特征也是多样的，就像楷书、小篆、行书等更加丰富了汉字的多样性，也为我们识别汉字增加了困难。汉字的基本结构就是在笔画的基础上构建的，在进行手写汉字识别时不仅要了解汉字的笔画特点，还要学习训练汉字的结构特点[65]。不同结构的汉字重心是不同，手写汉字高扁不一，大小也因书写习惯的不同而不确定，重心位置偏低的汉字在结构上比较宽扁，而重心位置偏高的汉字在结构上比较高瘦，根据这个特点汉字结构的重心位置可以成为识别过程中的一个指标，汉字结构的重心位置可以分为 X 轴和 Y 轴。汉字的 X 轴重心Z_x公式为（3-10）所示。

$$Z_x=\frac{\sum_{x=0}^{x=A-1}\sum_{y=0}^{y=B-1}xT(x,y)}{\sum_{x=0}^{x=A-1}\sum_{y=0}^{y=B-1}T(x,y)} \qquad 式(3-10)$$

汉字的 Y 轴重心Z_y公式为（3-11）所示。

$$Z_y=\frac{\sum_{x=0}^{x=A-1}\sum_{y=0}^{y=B-1}yT(x,y)}{\sum_{x=0}^{x=A-1}\sum_{y=0}^{y=B-1}T(x,y)} \qquad \text{式(3－11)}$$

不同人的手写汉字存在长瘦和宽扁之分，所以它们之间的比例是不同的，手写汉字的高宽比 R 表示公式如（3-12）所示。

$$R=\frac{\max\{y_i f(x_i,y_i)\}-\min\{y_i|\ f(x_i,y_i)=1\}}{\max\{x_i f(x_i,y_i)\}-\min\{x_i|\ f(x_i,y_i)=1\}} \qquad \text{式(3－12)}$$

而对于同一个汉字的多种书写结构，我们计算平均高宽比 S ，其表达公式如（3-13）所示。

$$S=\frac{1}{n}\sum_{i=1}^{n}R \qquad \text{式(3－13)}$$

有些人在手写汉字时会有个人喜好的特点，例如按压、连笔等，这些习惯会使汉字在纸张上受力分布不平衡或是在屏幕上进行电子签名时影响识别效果，所以我们能够计算笔墨的颜色分布来预测汉字的结构特征，汉字的结构记为 S 区域，图像记为 G 区域，然后计算区域中笔墨的分布量 m ，则笔墨的分布公式如（3-14）所示。

$$m_1=\sum_{\substack{(x_g,y_g)\in G\\(x,y)\in S}}\max\left\{\left|x-x_g\right|\wedge\left(y=y_g\right),\left|y-y_g\right|\wedge\left(x=x_g\right)\right\} \qquad \text{式(3－14)}$$

(x,y) 和 (x_g,y_g) 分别表示 S 和 G 区域内的点。

3.3 图像预处理

在汉字识别系统中，图像的质量会影响检测的效果，所以在进行特征提取前要对图像进行预处理，以提高模型对手写汉字的识别率。所谓手写汉字图像的预处理是指消除图像中与汉字识别无关的信息，将全部的有效信息保留下来。图像的预处理一般包括：归一化、降噪和二值化，下面将对这几个处理进行简单的介绍。

首先是对图像进行归一化的处理，也被称为规范化，此处理可以将汉字笔画中失真的部分消除掉，使图像在尺度大小、字体结构上呈现规范化，而且这个过程不会改变图像的对比度，将图像进行规范化处理能够使梯度下降得速度加快快，并且使得网络收敛的时间缩短。因为每个人书写的汉字大小不一，而且中国的汉字呈现多样性，所以需要对手写汉字的大小以及结构进行调整。对手写汉字图像采用最大值最小值归一化，其表达公式如（3-15）所示。

$$norm=\frac{x_i-\min(x)}{\max(x)-\min(x)} \qquad \text{式(3－15)}$$

公式中，x_i 是指图像中的像素点值，而 $\max(x)$ 和 $\min(x)$ 分别是指图像中像素点的最大值和最小值。归一化后图像中的信息不会丢失，只是像素的值从 $[0,255]$ 转化为 $[0,1]$

之间。

其次是对图像中的噪声进行消除，在进行手写汉字数据集采集的过程中，肯定会受到外界噪声的影响，而且数据集中的图像成像时光照的强弱也会产生噪声，这些图像中的噪声会影响识别的有效性，为了保证降噪的同时对有效信息的保护，在实验中采用高斯滤波，其属于平滑线性滤波器，可以使图像中的汉字特征更加明显，并且对图像中的分布特征可以有很好的保护作用，因为它对于图像中所有像素进行加权平均。高斯滤波器的表达式如（3-16）所示。

$$G(x,y)=\frac{1}{2\pi\sigma^2}e^{-(x^2+y^2)/2\sigma^2} \qquad \text{式(3－16)}$$

表达式中(x,y)是指像素点的坐标。

最后是对目标图像进行二值化，它可以使图像背景不再复杂，这样一来模型就降低了很多的计算量，并且有利于对图像作进一步的处理，在进行二值化后目标图像中只存在黑白两种颜色，从而使得目标的轮廓更加明显，这是因为像素点的灰度值只有 0 或 255。当像素点的灰度值小于阈值时，此时判定像素点不在检测的范围内，在图像上则表示的是除汉字以外的背景，其表达形式如（3-17）所示。

$$g(x,y)=\begin{cases}255(\text{白}), f(x,y)\geq T\\ 0(\text{黑}), f(x,y)<T\end{cases} \qquad \text{式(3－17)}$$

其中，T 表示最佳的阈值，如上式可知，当像素点的值小于阈值时，此时的灰度值为 0，否则为白，从而实现了汉字与背景的区别。

3.4 本章小结

本章首先对系统的功能和系统的结构进行了分析，对系统中图像的预处理进行了简单介绍，其中包括规范化、降噪和二值化，并对汉字的特征进行了阐述。本章将为接下来手写汉字的检测和识别提供系统设计基础。

4 手写汉字检测模型的构建

在第三章中对手写汉字识别系统进行设计的基础上，本章将展开对手写汉字检测模型的构建，介绍了模型中的损失函数，并对网络模型进行创新改进，以便实现对不同背景情况下汉字准确高效地检测，最后实验数据表明改进后的网络模型，对汉字区域检测的准确率有所提高，并且后续将在此基础上实现对不同背景情况下汉字的精准识别。

4.1 损失函数

所谓损失函数，是指用来衡量模型在训练时产生的误差和错误的函数，它通过反馈值来说明模型训练的效果与原始数据的差别之处，其对于构建识别模型具有重要的作用。在网络模型中，损失函数越小，预测数据越接近真实数据，进而说明网络模型的鲁棒性很好。通过损失函数对网络模型的反馈，我们可以对模型中相应的过程进行改进，有利于更好的对后处理进行分析和总结，以便实现预期的训练结果。本节总结了在网络模型中常用到的损失函数以及针对其不足进行改进后的损失函数，其中包括：L1 损失函数、Smooth L1 损失函数、IoU 损失、GIoU 损失、DIoU（Distance-IoU）损失等。

4.1.1 L1 损失函数

在深度学习中要使用 L1 损失函数，就需要先了解 L1 范数正则化。L1 范数也就是最小绝对偏差，则预测值 $f\left(x_i\right)$ 和真实值 y_i 之间的绝对偏差 S 表达式如（4-1）所示。

$$S=\sum_{i=1}^{n}\left|y_i-f\left(x_i\right)\right| \qquad \text{式(4－1)}$$

L1 范数相比 L2 范数具有很强的鲁棒性，是因为作为最小绝对偏差，它可以处理数据中的异常数值。如果在实验室忽略异常值，有可能导致有效数据的丢失，所以在需要考虑异常值的情况下，可以去选择 L1 损失函数。但是 L1 损失函数并没有很好的稳定性，数据集上一个方向的小变化，就会造成回归线的巨大波动，该损失函数在一个区域上有很多解，跳跃很大使得偏差线倾斜角度增大[66]。使用实验模型中的真实数据和拟合模型对 L1 范数和 L2 范数的稳定性进行比较，其中红色和绿色线条分别表示 L1 和 L2 的网络模型，实线是指数据中不存在异常值，而虚线则是数据中存在异常值。拟合模型如图 4-1 所示。

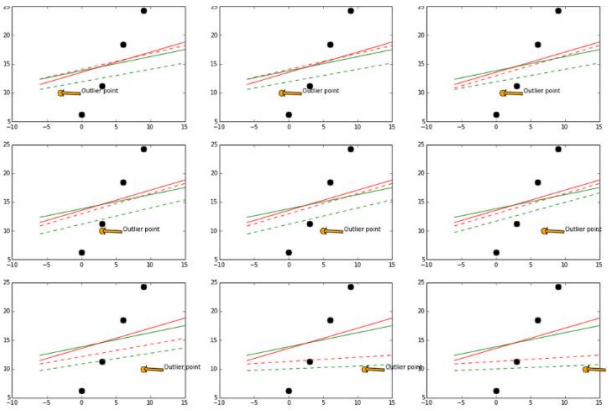


图 4-1 L1 和 L2 网络模型拟合模拟图

Fig. 4-1 L1 and L2 network model fitting simulation diagram

当异常值由左向右移动时，它在左右两边时比在中间时变得更加异常；当异常值位于中间位置时，L1 范数相比 L2 范数来说拟合的还存在很大的变化，由此说明了 L1 损失函数的稳定性较差。L1 范数具有内置特征选择的性质，它会产生稀疏的系数，会存在很多极小的系数以及很少的大系数，这是 L2 范数不具备的特性。

L1 正则化是损失函数的回归形式，对系数进行约束，根据情况进行相应的调整或者减小，使网络模型变得简单，并且提高网络模型的稳定性，这样一来就可以防止过拟合现象的出现。L1 正则化是所有权重的和，其公式如（4-2）所示。

$$L1=\sum_{i=1}^{n}\left|x_i\right| \qquad \text{式(4－2)}$$

损失函数的线性回归公式如（4-3）所示。

$$Y\approx\lambda_0+\lambda_1X_1+\lambda_2X_2+...+\lambda_pX_p \qquad \text{式(4－3)}$$

公式中，Y 是指学习关系，而 λ 是指不同变量 X 的系数预测。在模型中采用正则化，通过调整预测因子，能让模型可以更好的泛化，在减小方差的同时不会造成数据的丢失。L1 正则化可以将训练过程分解为单独的子优化问题，有利于对文本信息的检测以及对语义相关性计算。

4.1.2 Smooth L1 损失函数

根据 L1 损失函数不稳定的特性，后来改进得到了 Smooth L1 损失函数，它首次被提出是在 Fast R-CNN 中，其主要作用就是为了消除梯度爆炸[67]。Smooth L1 损失函数比 L1 损失函数变得更加平滑，可以对远离中心的点以及异常值都不敏感，更加具有鲁棒性， Smooth L1 损失函数的公式如（4-4）所示。

$$smoothed_{L_1}\left(x\right)=\begin{cases}0.5x^2,\left|x\right|<1\\ \left|x\right|-0.5,\left|x\right|\geq1\end{cases} \qquad \text{式(4－4)}$$

由上述公式，可以看出当预测数值与真实数值相差不大于 1 时，因为公式中存在 0.5 的平滑参数，所以模型不会出现梯度爆炸的现象；当预测数值与真实数值大于等

于 1 时，减小损失函数的幂值，对其求导就可以阻止梯度爆炸的出现。L1、L2、Smooth L1 损失函数的曲线图，如图 4-2 所示。

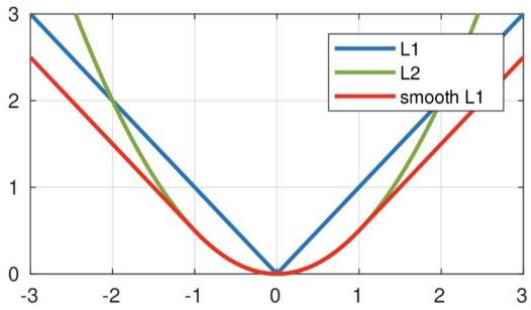


图 4-2 L1、L2 和 Smooth L1 损失函数的曲线图

Fig. 4-2 Graphs of loss functions of L1, L2 and Smooth L1

Smooth L1 损失函数可以在训练中控制梯度的量级，使模型不会出现梯度爆炸的现象。防止实验模型出现梯度爆炸，我们可以减少学习率，也可以采用正则化，检查模型的权重大小，并根据情况对权重进行限制。

4.1.3 IoU、GIoU、DIoU 损失函数

所谓的 IoU 是指交并比，它是计算数据集中对预测物体准确率的标准，在实验中指的是预测文本框和原始文本框的重合率，其表达公式为（4-5）所示。

$$IoU = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \qquad \text{式 (4－5)}$$

IoU 在水平面上的计算简化过程如图 4-3 所示。

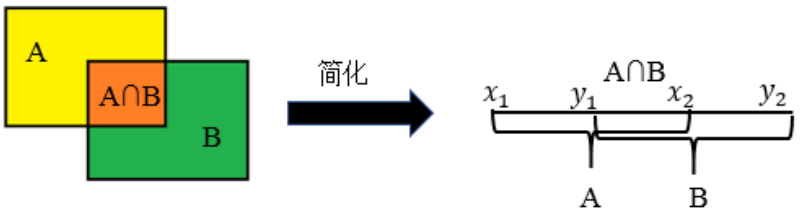


图 4-3 IoU 计算简化过程图

Fig. 4-3 IoU calculation simplified process diagram

采 IoU 损失函数在进行文本框预测的时候，它是将 4 个边界看成一个整体进行回归的，与之不同的是上述 L1 损失是对文本框的 4 个顶点分别求解损失后再相加，4 个顶点之间并没有引入相关性，而且 L1 损失函数更适合训练尺度大的目标。

IoU 损失函数的前项计算过程如表 4-1 所示。

表 4-1 IoU 损失函数的前项计算过程

Table 4-1 The antecedent calculation process of the IoU loss function

Algorithm 1 : *IoU loss Forward*
Input : $\tilde{x}$ as bounding box ground truth
Input : X as bounding box prediction

Output : $\mathbf{L}$ as localization error

for each pixel (i, j) **do**

if $\tilde{x} \neq \mathbf{0}$ **then**

$X = (x_t + x_b) * (x_l + x_r)$

$\tilde{X} = (\tilde{x}_t + \tilde{x}_b) * (\tilde{x}_l + \tilde{x}_r)$

$I_h = \min(x_t, \tilde{x}_t) + \min(x_b, \tilde{x}_b)$

$I_w = \min(x_l, \tilde{x}_l) + \min(x_r, \tilde{x}_r)$

$I = I_h + I_w$

$U = X + \tilde{X} - I$

$IoU = 1/U$

$\mathbf{L} = -\ln(IOU)$

else

$\mathbf{L} = 0$

end

end

表格中的 x，$\tilde{x}$ 分别代表文本框顶点的预测值和原始值，X，$\tilde{X}$ 则分别表示预测文本框的面积和真实文本框的面积，I 表示 $X \cap \tilde{X}$，U 表示 $X \cup \tilde{X}$。

采用 IoU 损失函数对文本进行检测，其可以对图像进行规范化处理，有利于处理多尺寸的目标，对于文本框的检测区域越大，说明损失越多，在实验中我们希望达到的理想效果是预测文本框与真实文本框重合率等于 1，这样对文本的检测才会更加准确和有效。

IoU 损失函数中变量之间是独立的，并且对于目标尺度多变的问题其能保持不变性，这两大特点都是 Smooth L1 损失函数所不具备的，但是 IoU 损失函数也存在不足，当检测文本框与实际文本框不重合时，它不能进一步地学习训练，而且不能反馈重合的大小以及类型[68]。根据上述 IoU 损失函数的缺点，由此衍生出多种损失函数，例如 GIoU 和 DIoU 是在 IoU 的基础上通过改进得到。

GIoU 损失经过改进后保留了 IoU 损失的优点，而且在 IoU 损失原来的基础上增加了一个面积最小的文本框，这个最小矩形框中包含了检测的矩形框和实际的矩形框。相比 IoU 损失而言，GIoU 损失会反馈预测文本框和实际文本框的重合面积，即使两个文本框没有重合的区域，GIoU 损失也会对其进行训练学习[69]。GIoU 损失函数的表达示如（4-6）所示，过程如图 4-4 所示。

$$L_{GIoU} = IoU - \frac{|C - A \cup B|}{|C|} \qquad \text{式(4—6)}$$

上述公式中，C 表示包含预测框和实际框的最小文本框。

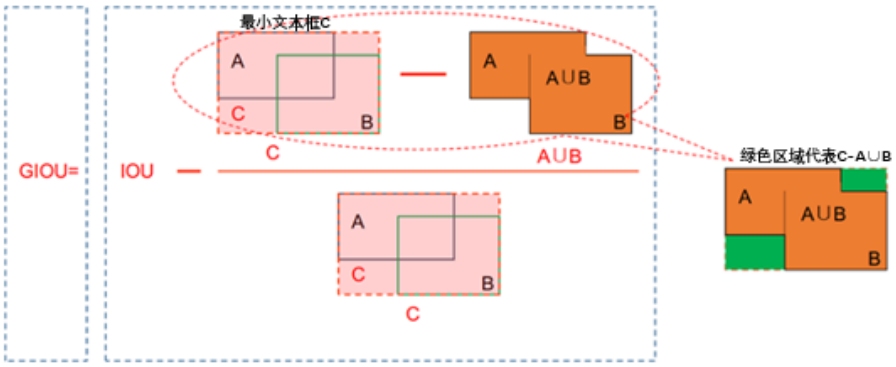


图 4-4 GIoU 损失函数计算过程图

Fig. 4-4 GIoU loss function calculation process diagram

GIoU 损失函数在 x, y 两个垂直方向上，预测效果很差，导致极其不稳定。针对 GIoU 损失的缺点进行改进，通过将预测文本框和实际文本框中心点的距离进行最小化处理，使其可以在垂直方向上可以收敛，改进后称之为 DIoU 损失函数。相对 GIoU 损失而言，DIoU 损失更有利于边界框的回归，它可以反馈文本框的重合率和重合面积的大小，对其进行训练学习，不会出现前两个损失不收敛的现象[70]。

4.2 基于改进 Advanced EAST 的汉字检测

对于简单背景下的手写汉字识别，可以说识别率已经很高了，并且相应的技术非常成熟了，但是对一些复杂背景的手写汉字进行识别时，识别率仍然不能够达到完美。相比简单背景汉字的识别，在复杂情况下的汉字识别首先是对图像中的汉字区域进行检测，然后将其中的汉字区域转化成字符信息，在这过程中我们需要解决很多的问题，例如如何将复杂的背景与手写汉字分离，要确定汉字的坐标位置以及所涉及到的范围是多少，并且在进行数据训练时我们需要充足的数据集，复杂背景的数据集相对较少，对我们前期训练模型带来了不利等。本章将对汉字与复杂背景进行分割，准确找到汉字区域，为下一步汉字的识别提供有效的帮助。

4.2.1 基于 EAST 网络模型的简述

Advanced EAST 是基于对 EAST 网络输出层的改进，其在场景文字识别上的应用比较广泛。EAST 的检测过程先由算法生成样本的候选框，对像素点的语义进行分割，主要是通过卷积神经网络来区别文本框和背景。另一方面在检测过程中会做边框回归，然后再通过采用 NMS 算法搜索局部最大的元素值，其几何形状可以是具有旋转角度的不规则四边形[71-73]。由于 EAST 检测要用到所有预测的像素点，对这些预测的像素点坐标进行加权平均进而计算最终的顶点坐标，其从四边形的另一边预测 2 个顶点，使得文本框内部的像素跨度很大，就导致在对像素进行预测时造成缺陷，这也是 EAST 对长文本检测不准确的原因。EAST 网络模型的框架,如图 4-5 所示。

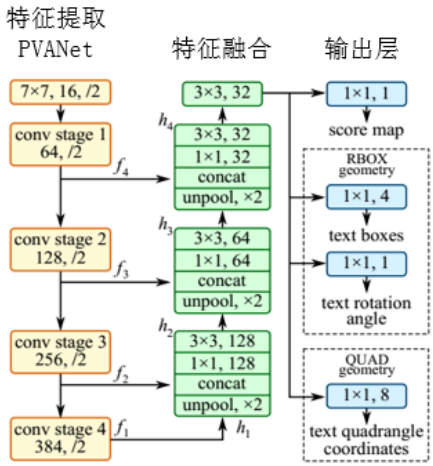


图 4-5 EAST 网络模型的框架图

Fig. 4-5 Frame diagram of EAST network model

EAST 网络是全卷积神经网络模型，先对输入的目标图像进行特征提取，然后根据文本框的得分对特征进行合并融合，后处理部分可以设定预测文本框的阈值，将其应用到所有的预测文本框中，当文本框的得分高于阈值时，才说明预测的文本框是有效地，以及采用局部感知 NMS 对文本框进行选择。

4.2.2 改进网络模型的构建

为了解决 EAST 对长文本检测不准的问题，Advanced EAST 在特征提取层中增加了卷积层的通道，将 EAST 的输出层改成了 7 通道的输出，进而对后面的处理方法也进行了优化。Advanced EAST 的网络结构主体模型是 VGG16，VGG16 的网络结构包括 13 个卷积层和 3 个全连接层，其在经过 4 个阶段的卷积池化交替后，能够获得 4 种尺寸大小不一样的特征图，它们的大小分别是输入图像的 1/4，1/8，1/16，1/32，这 4 个阶段的特征图将作为特征合并的输入[74]。不同大小的特征图可以解决在检测过程中文本行长短不停变化的困难，早期的特征图可以检测短的文本行，而后期的特征图则可以用来检测较长的文本行。Advanced EAST 网络结构的框架，如图 4-6。

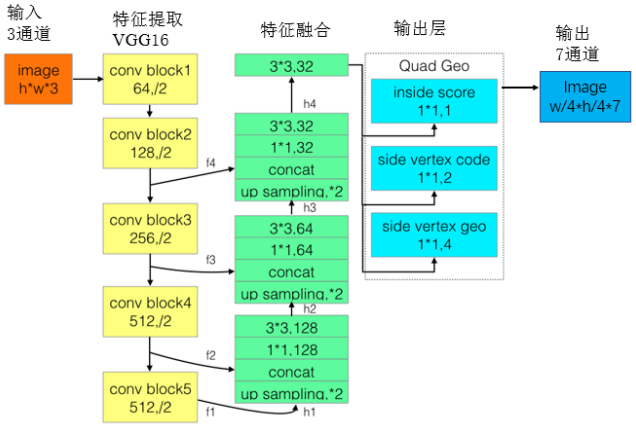


图 4-6 Advanced EAST 网络结构的框架图

Fig. 4-6 Framework diagram of Advanced EAST network structure

由上图的网络框架，可以看出模型的输入是 3 通道的图像，在通过 VGG16 的特征提取以及特征合并后，输出层的图像的尺度大小变为原来的 1/4,而且是 7 通道输出。在 Advanced EAST 模型中通过使用多尺度合并的方法，能够解决在文字检测中遇到的多尺度目标检测。对于上述在特征提取中得到的 1/4，1/8，1/16，1/32 这 4 种不同尺度的特征图，标记为 s_i ，则融合基础 t_i 公式如（4-7）所示。

$$t_i=\begin{cases} unpool\left(m_i\right), & i\le 3 \\ conv_{3\times 3}\left(m_i\right), & i=4 \end{cases} \qquad \text{式(4－7)}$$

融合特征图公式如（4-8）所示。

$$m_i=\begin{cases} s_i, & i=1 \\ conv_{3\times 3}\left(conv_{1\times 1}\left(\left[t_{i-1};s_i\right]\right)\right), & i\ne 1 \end{cases} \qquad \text{式(4－8)}$$

上述表达式中，t_i 是融合基础，m_i 是融合特征图，其中 $\left[t_{i-1};s_i\right]$ 表达的是特征图。根据通道的维度进行融合，我们通过上池化层对特征提取阶段输入的特征图进行尺度扩大处理，将其扩大后再与现阶段的特征图合并，接下来通过 1*1 的卷积减少因融合增加的无用信息，以降低计算量，然后再通过 3*3 的卷积对特征图局部信息进行融合，最后把输出的特征图输入到输出层中。不同大小尺度特征图的感受野不同，当进行汉字区域检测时，感受野太小会造成检测结果的不准确，感受野太大又会造成检测信息的丢失，不利于在图像上对汉字区域进行检测。特征融合能够把尺度大小不一的特征图合并起来，可以实现对目标的多尺度检测。

Advanced EAST 网络的输出层有 7 个通道，分别输出 1 位置信度，预测像素点在文本框内的概率，即像素点是否在标定的文本框内；2 个顶点，预测像素点是否属于文本框边界像素以及顶点是文本框的头部还是尾部，其中我们用 0 代表头部像素点，用 1 代表尾部像素点，预测的像素点构成文本框的形状后，再通过边界像素去预测回归顶点坐标；最后 4 个通道输出 4 位坐标位置，这里坐标位置的真实含义是当前点 $\left(x,y\right)$ 的偏移量。与 EAST 不同的是 Advanced EAST 并不是对所有的像素点进行预测，它是通过边界像素的坐标位置以及头部和尾部的像素点对左上、左下、右上、右下的像素点来预测顶点坐标的，获得矩形中所有像素在图像中的起始坐标，然后得到特征图上大于阈值区域的点映射到原图像分辨时的坐标，根据坐标的偏移量对所有顶点的全部检测值进行加权平均，输出的 4 个顶点作为最终的坐标值，即 8 个坐标值来确定这个文本框。文本框预测过程图，如图 4-7 所示。

图 4-7 文本框预测过程图

Fig. 4-7 Text box prediction process diagram

经过上面的预测过程，文本框预测的效果图，如图 4-8 所示。

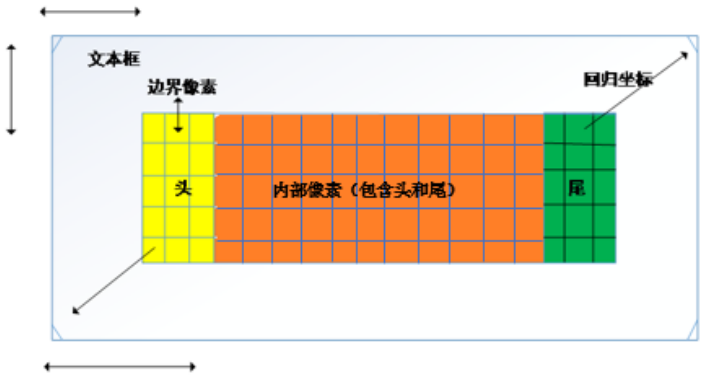


图 4-8 文本框预测的效果图

Fig. 4-8 Effect picture of text box prediction

通过 4 个顶点 a_1、a_2、a_3、a_4 确定的不规则四边形 S_a，先以其中相邻的两边 L_{a01}、L_{a12} 作平行四边形，接下来再通过平行四边形得到外接的最大矩形，过 a_2 作平行于直线 L_{a01} 的线 f_2，同理，过 a_3 作平行于直线 L_{a12} 的线 f_3，由此可以得到一个平行四边形，当然还可以通过作 L_{a23}、L_{a03} 的平行虚线得到另一个平行四边形。计算所有平行四边形的面积，对面积最大的平行四边形作内接矩形，这就构建了文本框。以 L_{a03}、L_{a12} 为对角线分别作平行四边形，利用分割函数将平行四边形分割成三部分，然后各自内接矩形，这样在原来一个矩形的基础上就得到三个矩形。在文本框构建的过程中会涉及到很多数学计算，例如计算四边形的面积、点到直线的距离、矩形的面积、交点坐标等。文本框的坐标位置如下图 4-9 所示。

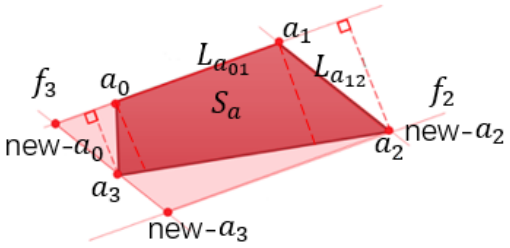


图 4-9 文本框的坐位置图

Fig. 4-9 Sitting position map of the text box

根据上述对模型的表述，我们对网络模型进行解析，过程如图 4-10 所示。

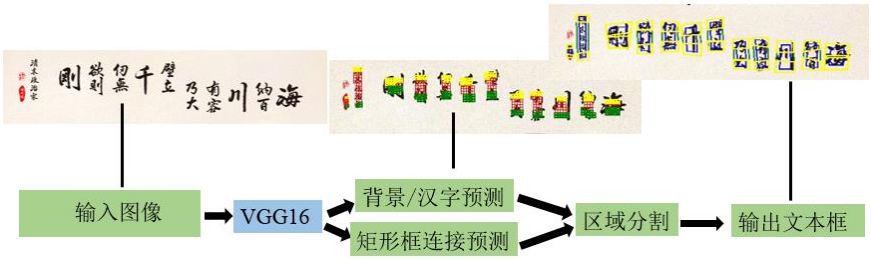


图 4-10 网络模型过程解析图

Fig. 4-10 Network model process analysis diagram

在网络模型训练的过程中，为了生成准确有效的文本框，模型会对目标图像进行 2 个预测，对区分背景和检测目标手写汉字的预测，需要确定检测的手写汉字的坐标和旋转角度。

4.2.3 图像分割函数

复杂背景的文本检测中字形的大小、风格都有很大的差别，由于输出元素的每个像素点到它所在旋转矩形每条边的距离与损失函数有关，这就需要在实验中采用不同的损失函数，主要是分类和回归两类损失函数[75]。网络检测中总损失的计算公式如式（4-9）所示。

$$L = L_s + \lambda_g L_g \qquad \text{式(4—9)}$$

公式中 L_s 表示分类损失，L_g 表示回归损失，其中 λ_g 是用来平衡分类损失和回归损失的，在实验中 λ_g 设置成 1。

判断每个像素点是否属于文本框内，需要用到分类的损失函数，由于文本所在的区域像素的值为 1，其他背景区域像素点值为 0，一般在二分类的任务上，两者的概率和是 1，所以可以仅预测其中的一个概率。在语义分割中一般用交叉熵来做损失函数, 平衡交叉熵的表达式为（4-10）所示。

$$L_s = -\left(\beta \bullet p \bullet \log\left(\hat{p}\right) + \left(1\text{-}\beta\right)\left(1-p\right) \bullet \log\left(1-\hat{p}\right)\right) \qquad \text{式(4—10)}$$

公式中，平衡因子 β 用来平衡正例样本与负例样本的权重，$\hat{p}$ 表示的是网络对样本预测的结果，即网络中样本为正例的概率，而 p 表示的是样本真实的标签，因为是二分类，所以如果样本属于正例，则 $p=1$，否则 $p=0$。对于网络中所有的数据集，其损失函数指的是所有样本点的损失函数的平均值，而平衡交叉熵损失函数对负样本也可以加权[76]。平衡交叉熵可以改善正例样本和负例样本之间权重的不平衡，但其对难易样本的区分不是很充足，所以它不能改善难易样本之间的不平衡。

在实验中我们采用 Dice 损失，它是与区域相关对语义进行分割的损失函数，其比较适用于分析前景区域，尤其是对语义分割中正负样本极度不平衡的情况，因为造成正负样本不平衡的原因就是前景占比小，而且它的收敛速度也比类平衡交叉熵快[77]。

Dice 系数是用来评估不同图像之间相似度的函数，Dice 损失的表达式为（4-11）所示。

$$L_{\mathrm{d}}=1-\frac{2|X\cap Y|}{|X|+|Y|} \qquad \text{式(4－11)}$$

公式中，$|X\cap Y|$表示 X 和 Y 之间的交集，$|X|$和$|Y|$分别表示 X 和 Y 样本的个数，分子中的系数 2 是为了保证L_d 的取值范围为 0 到 1 之间，因为分母的计算中存在重复元素，L_d 在$[0,1]$之间进行取值，而且样本之间的相似度与取值的大小成正比。Dice 损失也有存在不足的地方，它的使用对反方向传播会造成不利的影响，导致训练误差曲线混乱，不易看出收敛的信息，会出现梯度饱和的现象，特别是对于小目标 Dice 损失训练就会不稳定。

在文本框构建的过程中会生成具有旋转角度的旋转矩形以及普通的四边形，对于文字几何形状的预测，要保持其尺度的大小不改变，所以针对旋转矩形和四边形需要采用不同的损失函数。生成的旋转矩形每个像素点都有一个正分值，需要计算像素点到文本框 4 边的距离，而对于普通四边形，文本框中所有像素点的正分值是其与四边形 4 个顶点的坐标偏移。

文本框的边界可以是轴对齐的，也可以是任意方向的，其中轴对齐矩形边界框容易生成，并且使用方便。轴对齐矩形边界框内的点需要满足公式（4-12）。

$$x_{\min}\le x\le x_{\max},\ \ y_{\min}\le y\le y_{\max},\ \ z_{\min}\le z\le z_{\max} \qquad \text{式(4－12)}$$

需要注意的是边界框的中心点 $d=(t_{\min}+t_{\max})/2$，其中 $t_{\min}=[x_{\min},y_{\min},z_{\min}]$，$t_{\max}=[x_{\max},y_{\max},z_{\max}]$。

在矩形边界框中由$t_{\min}$ 指向$t_{\max}$ 的向量称为尺寸向量，则尺度向量 $\vec{p}=t_{\max}-t_{\min}$。尺度向量 $\vec{p}$ 包括矩形边界框的长、宽、高。在矩形边界框中由中心 d 指向$t_{\min}$ 的向量称之为半径向量$\vec{r}$，其表达式如（4-13）所示。

$$\vec{r}=t_{\max}-d=p/2 \qquad \text{式(4－13)}$$

在实验过程中采用$t_{\min}$ 和$t_{\max}$ 去表达矩形边界框，然后在利用$t_{\min}$ 和$t_{\max}$ 去求解 d 、$\vec{p}$ 和$\vec{r}$ 会容易很多。

实验中轴对齐矩形边界框使用基于距离的 DIoU 损失，原来模型中采用的 IoU 损失函数是存在不足的，当检测文本框与实际文本框不重合时，它不能进一步地学习训练，而且不能反馈重合的大小以及类型。而 DIoU 损失更有利于边界框的回归，它可以反馈文本框的重合率和重合面积的大小，对目标进行训练学习，不会出现 IoU 损失不收敛的现象。DIoU 损失相比 IoU 损失而言收敛速度更快，回归更加准确。DIoU 损失是对预测文本框和原始文本框之间进行融合标准化距离，其可以最大程度的保证预测文本框中心和原始文本框重合，而不会产生太大的文本框。真实框与预测框的位置关系，如图 4-11 所示。

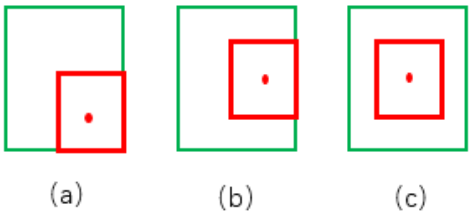


图 4-11 真实框与预测框的位置图

Fig. 4-11 Location map of the real frame and the predicted frame

上图中绿色框表示真实框，而红色框表示预测框。根据真实框和预测框两个框的位置关系，我们分析了 IoU、GIoU 和 DIoU 之间的损失关系，其对比表如 4-2 所示。

表 4-2 IoU、GIoU 和 DIoU 之间的损失对比表

Table 4-2 Loss comparison table between IoU, GIoU and DIoU

损失函数	（a）图	（b）图	（c）图
L_{IoU}	0.75	0.75	0.75
L_{GIoU}	0.75	0.75	0.75
L_{DIoU}	0.81	0.77	0.75

由上表分析可得当 L_{GIoU} 趋近于 L_{IoU} 时，L_{DIoU} 对文本框仍然是可分辨的。由此说明相比 IoU 损失和 GIoU 损失而言，DIoU 损失回归更加准确，有利于边界框的回归。

对于旋转矩形中的 DIoU 损失，其运算公式为（4-14）所示。

$$L_{DIoU} = 1 - IoU + \frac{\rho^2\left(b, b^{gt}\right)}{c^2} \qquad 式(4-14)$$

其中，b，b^{gt} 分别指的是预测文本框和原始文本框的中心点，ρ 指的是 b 和 b^{gt} 之间的欧式距离，c 指的是最小矩形框的对角线距离，这个最小矩形框能够包含预测文本框和原始文本框，而 $R_{DIoU} = \rho^2\left(b, b^{gt}\right)/c^2$ 表示的是惩罚项，DIoU 损失的原理是在 IoU 上增添了惩罚项，它是对两个文本框中心点之间的距离进行最小规范化处理。

DIoU 损失即使与原始文本框没有重合，它依然能够为预测文本框提供移动方向，而且对于两个文本框之间的距离，它可以直接最小规范化处理。DIoU 损失应用于局部感知 NMS 算法中的话，可以使运算结果更加准确和高效。NMS 算法与 IoU 损失的结合使用中，主要是利用 IoU 损失去消除重合的文本框区域，除此之外 DIoU 损失还可以计算两个文本框中心点之间的距离。对于重合率最高的预测文本框，DIoU NMS 表达式定义如（4-15）所示。

$$S_i = \begin{cases} S_i, & IoU - R_{DIoU}\left(M, B_i\right) < \varepsilon \\ 0, & IoU - R_{DIoU}\left(M, B_i\right) \geq \varepsilon \end{cases} \qquad 式(4-15)$$

表达式中，S_i 代表分类的重合率，ε 代表 NMS 的阈值。

余弦角度 θ 的损失计算公式为（4-16）所示。

$$L_\theta\left(\hat{\theta},\theta^*\right)=1-\cos\left(\hat{\theta}-\theta^*\right) \quad \text{式(4－16)}$$

公式中，$\hat{\theta}$指的是对旋转角度的一个预测值，而θ^*是对预测值的标记。由上述对 DIoU 损失和余弦角度θ损失的计算，旋转矩形中的总损失L_g公式如（4-17）所示。

$$L_{\mathrm{g}}=L_{DIoU}+\lambda_\theta L_\theta=1-IoU+\frac{\rho^2\left(b,b^{gt}\right)}{c^2}+\lambda_\theta\left(1-\cos\left(\hat{\theta}-\theta\right)\right) \quad \text{式(4－17)}$$

在普通四边形部分采用 smoothed-L1 损失,它可以对尺度进行规范化处理。其实，所谓的 smoothed-L1 损失是平滑处理之后的 L1 损失函数，L1 损失函数可以通过调整模型解决目标中出现的异常值情况，其具有很好的鲁棒性，但是它的回归线会因为目标集的小变化而产生巨大的波动。smoothed-L1 损失能够让远离中心的点具有鲁棒性，并且对于数据中的异常值不敏感，模型不需要牺牲正常数据的误差来调整异常值。其中，smoothed-L1 损失函数的表达式如（4-18）所示。

$$smoothed_{L_1}\left(x\right)=\begin{cases}0.5x^2,|x|<1\\|x|-0.5,|x|\geq 1\end{cases} \quad \text{式(4－18)}$$

smoothed-L1 损失的导数函数如（4-19）所示。

$$\frac{dsmoothed_{L_1}}{dx}=\begin{cases}x,|x|<1\\\pm 1,|x|\geq 1\end{cases} \quad \text{式(4－19)}$$

从 smoothed-L1 损失的导数函数中也可以看出其不会出现梯度爆炸的情况，因为它的梯度会随x的减小而变小，并且梯度的最大上限值为 1。

对于四边形中的 smoothed-L1 损失增加归一化处理，其所有点的坐标组成一个集合，集合$C_Q=\{x_1,y_1,x_2,y_2,...,x_4,y_4\}$，对于普通四边形的损失$L_g$表达函数为（4-20）所示。

$$L_g=L_Q\left(\hat{Q},Q^*\right)=\min_{\tilde{Q}\in P_{Q^*}}\sum_{\substack{c_i\in Q,\\ \tilde{c}_i\in\tilde{Q}}}\frac{smoothed_{L_1}\left(c_i-\tilde{c}_i\right)}{8\times N_{Q^*}} \quad \text{式(4－20)}$$

在文本检测的过程中会生成无数个预测文本框，需要把文本框置信度的大小当做是权重，根据阈值的大小确定可以留下的文本框，得到文本框的集合后再利用局部感知 NMS 算法进行融合文本框，保留置信度最高的矩形。NMS 算法流程图如 4-12 所示。

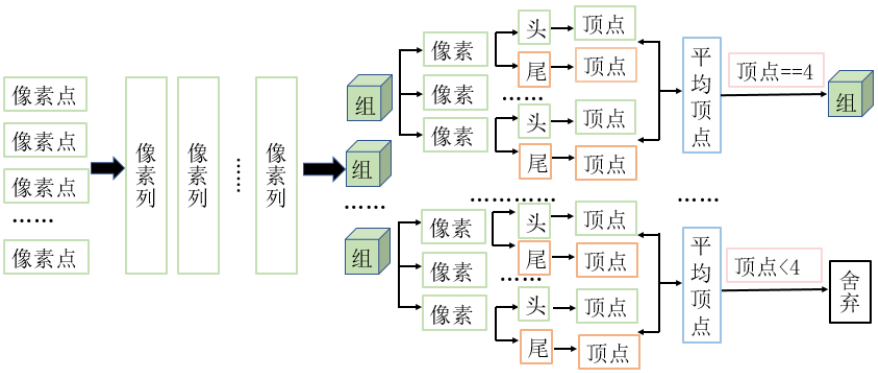


图 4-12 NMS 算法流程图

Fig. 4-12 NMS algorithm flow chart

在文本框中的所有像素点首先由上下组成许多列，这些列左右相连再组成很多分组，其次对每个分组里的像素点进行头尾分类，并且分别将头、尾部区域的像素点的值叠加得出头、尾部的顶点坐标，如上图可以看出输出的是经过加权平均后的顶点坐标，即满足 4 个顶点的 8 个值。在实验中会出现文本框不够 4 个顶点的情况，此时无法形成文本框，模型将放弃对这个文本框的预测，则汉字预测的范围即为被保留的文本框区域。

融合矩形的坐标位置是对两个确定的四边形的置信度大小来进行加权平均的，这种平均的做法可以减少计算量，而且全部文本框的坐标位置都能够被应用到，即使是置信度很低的文本框也可以参与其中，使文本框中尽可能的包含全部信息。在每个区域只能确定一个文本框为局部最大值，与其它区域的局部最大值作比较，预测被选出的文本框在整个区域的大小，实现此预测结果的算法如表 4-3 所示。

表 4-3 NMS 进行基于行融合文本框的算法

Table 4-3 NMS performs an algorithm based on line fusion text box

Algorithm 1 Locality - Aware NMS

```
function NMS LOCALITY (geometries)
    S ← ∅, p ← ∅
    for g ∈ geometries in row first order do
        if p ≠ ∅ ∩ SHOULDMERGE(g, p) then
            p ← WEIGHTEDMERGE(g, p)
        else
            if p ≠ ∅ then
                S ← S ∪ {p}
            end if
            p ← g
        end if
    end if
```

13: **if** $p \neq \emptyset$ **then**
14: $S \leftarrow S \cup \{p\}$
15: **end if**
16: **return** STANDARD NMS(S)
17: **end function**

4.2.4 实验以及数据

首先采用 ICDAR 2015 数据集对图像进行预训练，我们根据训练结果对数据集进行调整，当预测文本框 $0.5 \leq DIoU \leq 1$ 时，则视样本为正例；当预测文本框 $0.1 \leq DIoU < 0.5$ 时，则视样本为负例，在数据集中 25% 为正例样本，75% 为负例样本。该数据集中包含了自然场景文本、广告标语等背景复杂的图像，对于公开的 1500 张图像中测试集有 500 张，其余为训练集。ICDAR 2015 在数据集中又增加了来自 Google Glass 的 1670 张新图片，其包含的数据类型足够为复杂文本的研究提供较为完整的文本数据集[78]。另外，数据集中文本都是利用的 4 个点来标定的，与本实验中应用 4 个顶点标记四边形相对应。

在进行调优时，每一个批处理中放入 M 张图像，在批处理结束后，将预测文本框与预期输出值作对比，然后计算出它们之间的误差，根据误差对模型进行改进，实验模型中的回归损失只处理正例样本。采用 RoI（Region of Interest）池化层[79]代替 VGG16 网络模型的最后一层的最大池化层，它可以对输入多尺度的图像进行规范化处理，将其调整到同样的尺寸大小上。VGG16 网络模型中最后一层的分类输出修改为 21 个，并且采用 DIoU 作为新的回归层，模型实验过程图如 4-13 所示。

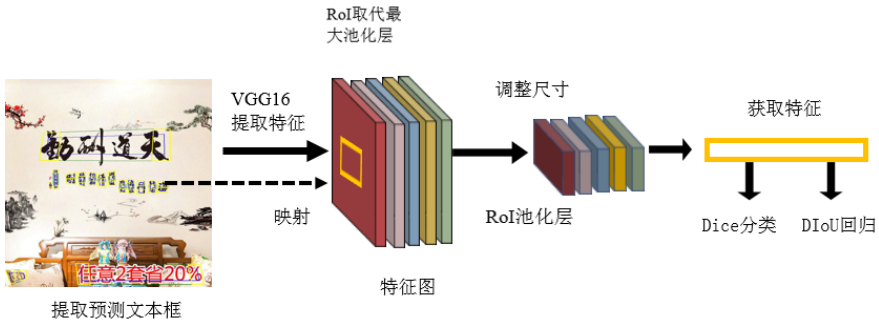


图 4-13 改进后模型实验过程图

Fig. 4-13 Improved model experiment process diagram

在模型输出的 7 通道中,前 3 个通道输出的值,主要是用来判断像素点是否属于文本框,根据情况设定阈值；而后 4 个通道输出的值是检测像素点到文本框的距离,并判断像素点属于文本框的头部还是尾部，对每个头部和尾部像素点检测出的值，对它们进行加权平均就是文本框的边界了，标签生成过程,如图 4-14 所示。

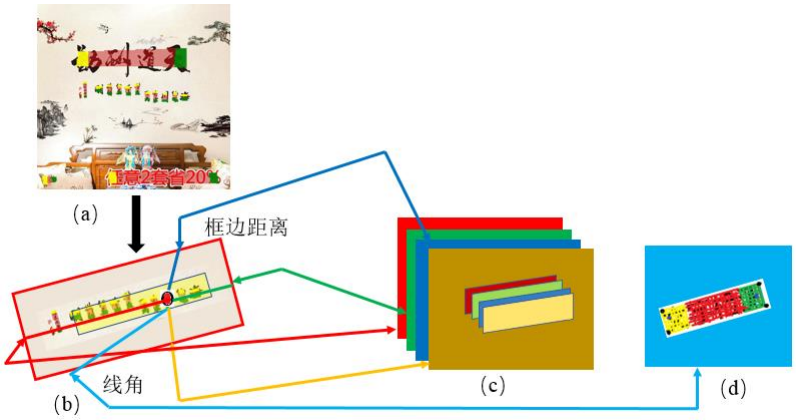


图 4-14 模型标签生成过程图

Fig. 4-14 Model label generation process diagram

上图是表示的是标签生成的过程，其中（a）是对检测到的汉字进行标框；（b）具有旋转角度的文本框的生成；(c)指的是所有像素点到文本框的距离的 4 个通道；(d)表示文本框的旋转角度。网络模型改进前后对检测文本框结果进行对比，图（a）为改进前，图（b）为改进后，示例图如图 4-15 所示。

图（a）改进前检测结果图

Fig.（a）Test result before improvement

图（b）改进后的检测结果图

Fig.（b）Improved test result graph

图 4-15 改进前后的检测结果对比图

Fig. 4-15 Comparison chart of test results before and after improvement

通过检测的对比图，在改进前文本框对文字的标定是不准确也不规范的，预测文字的坐标比较混乱，对文字的检测结果不全面，造成了文字的部分遗漏。对模型进行

重新设计改进后，我们可以看出示例图中文本框的标定较前期相比，对文字的检测更加准确，文本框数量的增加使对汉字区域的检测更加细化，并且提高了对汉字区域的检测率。

在实验中加入了对曲面背景的检测，如上述示例图，可以看出该模型能够很有效的检测出曲面的文字区域，而在文献[80]中采用的是 Polygon NMS，即利用多边形来检测曲面的文字区域，与局部感知 NMS 不同的是多边形 NMS 需要获得 14 个像素点的坐标偏移值，14 个像素点需要 28 个值来确定，并且在模型中需要提供 32 个值以加强对 14 个偏移量的监督，增加了模型的计算量，并且文献中的模型适用于曲面的汉字检测，对其他背景中汉字的检测不够灵活。在本实验过程中采用 4 个顶点 8 个像素值，通过对文本框旋转角度的计算，同样可以对曲面汉字的检测达到准确有效，而且相比计算量大大减少了，可以很灵活的应用到各种背景检测。对于本文中的损失函数在 PASCAL VOC 上进行测试评估，结果评估如表 4-4 所示。

表 4-4 损失函数评估表

Table 4-4 Loss function evaluation table

损失函数	损失评估	相对 IoU 提高
L_{IoU}	46.98	
L_{GIoU}	52.20	4.78%
L_{DIoU}	52.82	6.02%

在上表的数据中，我们可以看出相对前两种损失而言，DIoU 损失更准确一些，比 IoU 高出 6.02%，比 GIoU 高出 1.24%，说明 DIoU 损失更有利于边界框的回归，并且它可以反馈文本框的重合率和重合面积的大小，对其进行训练学习，不会出现前两个损失不收敛的现象。

实验中手写汉字检测的数据集不能够达到对所有情况都包含，所以在模型训练时对比了不同的网络对数据集的测试，以防止出现过拟合与欠拟合的情况。通过 CDAR 2015 数据集中的 500 张测试图像得出测试结果，对于网络模型的性能我们通过召回率、精确率和 F1 得分来进行比较，其中召回率是指真实的样本中正样本被准确分类的数量与真实样本的比率，精确率是指在网络模型进行预测时，它检测到的正样本中真实的正样本所占的比率，而 F1 得分是为了平衡召回率和精确率，让两者能同时到达最高值，其表达形式 F1=2*召回率*精确率/（召回率+精确率）。实验中的测试数据采用平均值，其测试结果如表 4-5 所示。

表 4-5 不同网络的测试描述

Table 4-5 Test description of different networks

基础网络	召回率	精确率	F1 分数
VGG16	0.324	0.5039	0.3945
PVANET[81]	0.302	0.3981	0.3434

PVANET2x	0.340	0.406	0.3701
VGG16+四边形	0.6895	0.7987	0.7401
本文方法	0.7275	0.8046	0.7641

由上表数据能够看出，本章中的检测网络模型在 CDAR 2015 数据集上取得了相对较好的成绩，比原来的基础网络在召回率上分别高出 0.3875 和 0.4255，在精确率上分别高出 0.4006 和 0.4065。但是由于实验数据不足够大，所以对手写汉字存在的其他不同的情况，肯定有没检测到的数据集。后续，我们将对各种汉字图像进行训练测试，以便对网络模型的检测性能进一步地完善。

4.3 文章小结

本章首先列举了网络模型中几种常用到的损失函数，其中包括：L1 损失函数、Smooth L1 损失函数、IOU 损失、GIOU 损失、DIOU 损失、CIOU 损失等。接下来阐述了基于改进 Advanced EAST 的背景与汉字分割的模型设计，首先对 EAST 网络模型进行了简述，其次根据上述 EAST 网络改进后的模型进行构建，最后进行网络模型实验，并对实验数据进行分析，为后续实现手写汉字的识别提供准确高效的检测模型。

5 手写汉字识别系统的实现

在前三章中我们了解到了神经网络模型与各种算法，并且根据需求的不同对网络进行改进，并在模型中选择合适的算法和函数，使其能够实现不同背景情况下对手写汉字准确有效的检测。本章以神经网络模型为基础，实现不同背景情况下对手写汉字的通用识别，简单介绍了基于卷积神经网络 LeNet-5 模型的识别，并对模型实验数据进行了总结与分析。本章通过对识别系统的实现，证实了系统对手写汉字识别的准确性与可行性，另外，对于识别系统设计中遇到的问题以及不足之处，在后续的研究中将会对识别系统作进一步的改进和完善。

5.1 基于卷积神经网络 LeNet-5 模型的识别

5.1.1 LeNet-5 模型的构建

应用深度学习 TensorFlow 框架[82]与 LeNet-5 网络模型并行搭建实现手写汉字识别的功能，实验中需要用到的环境如表 5-1 所示。

表 5-1 网络模型安装程序环境

Table 5-1 Network model installer environment

软件名称	软件版本
Anaconda	5.3.0
TensorFlow	1.10.0
CUDA	10.0.130
Cudnn	7.1.4

根据软件版本将其下载安装好，完成手写汉字识别环境的搭建，在此基础上进行算法的运算，最后实现对数据集中手写汉字的识别。根据第二章对数据集的介绍, 对手写汉字进行识别,首先是选择合适的数据集,接下来将目标图像输入进行预处理，对目标图像完成前期处理后，最后将其输入网络模型中进行手写汉字点的识别，并输出识别效果以检测模型的性能。手写汉字的识别流程如图 5-1 所示。

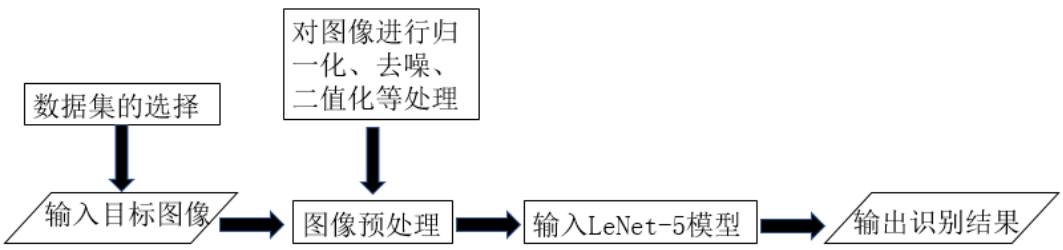


图 5-1 手写汉字的识别流程图

Fig. 5-1 Recognition flowchart of handwritten Chinese characters

基于深度学习的手写汉字识别需要确定选择深度学习框架来搭建模型，其中，TensorFlow 框架很容易进行可视化，并且对数据与模型之间并行很好，处理速度快，

在其框架内可以支持多种神经网络和算法，为手写汉字的识别提供很好的环境。

5.1.2 实验过程及结果分析

基于 LeNet-5 网络模型对得到的特征图进行合并，像素点的合并有利于对汉字的语义进行准确有效的分类[83]。首先对目标图像采取规范化、降噪和二值化等处理，将处理好的64*64的图像输入 LeNet-5 网络，其网络模型中采用的是稀疏连接，在网络模型的第一层卷积层*C*1中会形成 6 个特征图，并且特征图之间实行参数共享，这样大大减少了模型的计算量。根据对 LeNet-5 卷积的运算介绍，特征图通过卷积滤波器获得输入特征，其卷积层的基本结构如图 5-2 所示。

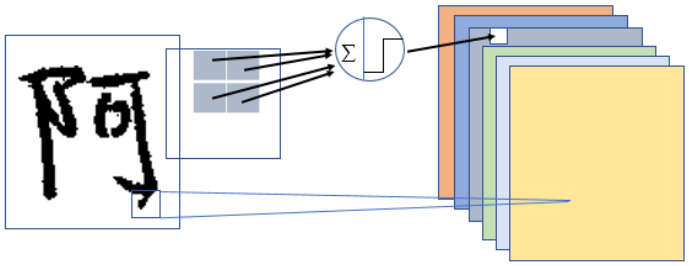


图 5-2 卷积层的基本结构图

Fig. 5-2 Basic structure diagram of convolutional layer

本实验在卷积层中引入区域加权系数，在进行特征提取时可以使汉字的轮廓变得更加明显，不同的汉字之间的结构是不同的，而且同一汉字的不同区域结构也是不同的，所以在实验中加入区域加权系数，对待不同的区域赋予不同的系数[84]。在网络进行训练时不断优化系数，一个区域的加权系数越大，就表示此区域的特征越多。区域加权的表达公式如（5-1）所示。

$$Output = f\left(\sum_{i,j\in M} Input_{i,j} * K * W_{m-i,n-j} + b\right) \qquad 式(5-1)$$

表达式中 Output 和 Input 分别指特征图的输出和输入, K 是指加权系数，W 是指该层卷积核中的特征图，而 m 和 n 分别指卷积核的大小。

第二层*S*2是池化层，将对*C*1卷积层提取的特征图进行下采样，经过池化后可以减少模型的计算量，并且能够防止模型过拟合。接下来就是*C*3卷积层和*S*4池化层，LeNet-5 模型经过卷积层和池化层的交替出现，在共享参数的同时减少了计算量，而且很大程度上降低了过拟合现象的出现。LeNet-5 网络模型的参数如表 5-2 所示.

表 5-2 LeNet-5 网络模型的参数表

Table 5-2 Parameter table of LeNet-5 network model

网络层	输入尺寸	卷积核个数	过滤器尺寸	训练参数	输出尺寸
*C*1	64×64	64	5*5	122304	62×62
*S*2	62×62	64	2*2	5880	32×32
*C*3	32×32	128	5*5	151600	30×30
*S*4	30×30	128	3*3	2000	16×16

$C5$	16×16	256	1*1	48120	14×14
$F6$	14×14	512	3*3	10164	8×8
$F7$	8×8	512	3*3	840	6×6

LeNet-5 模型在采用反向传播算法时会出现梯度消失的问题，因此在实验中采用 ReLU 函数代替 Sigmoid 函数，当函数值变小时，Sigmoid 函数不利于网络的反馈传递，而 ReLU 函数可以解决模型中梯度消失的现象。相比其他函数来说，ReLU 函数可以加速模型的收敛速度，在网络中的表达能力很强。ReLU 函数的表达形式，如图 5-3 所示。

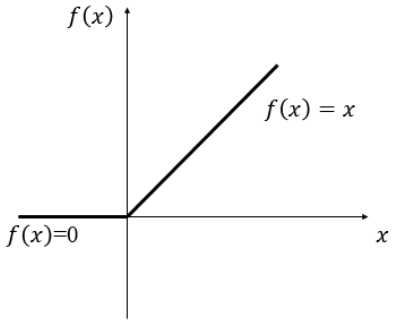


图 5-3 ReLU 函数的曲线图

Fig. 5-3 Graph of ReLU function

由上图可以看出，ReLU 函数是斜坡函数，并且是左饱和函数，这就决定了它能够处理模型中出现的梯度消失现象，而且 ReLU 函数的输出会存在是 0 的情况，使网络变成稀疏连接从而解决过拟合的问题。网络损失函数的曲线如图 5-4 所示。

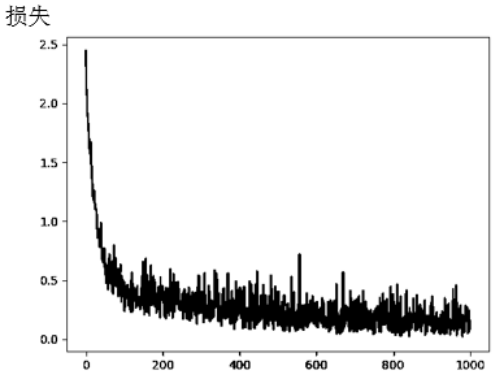


图 5-4 网络损失函数的曲线图

Fig. 5-4 Graph of network loss function

当对网络模型的训练次数增加时，网络的训练损失不断下降，说明此时网络的性能不断在变强，即对汉字的识别准确率在不断提升。实验中在 LeNet-5 模型的卷积层加入区域加权系数，将其应用在手写汉字识别中可以达到很高的准确率。其识别结果示例如图 5-5 所示。

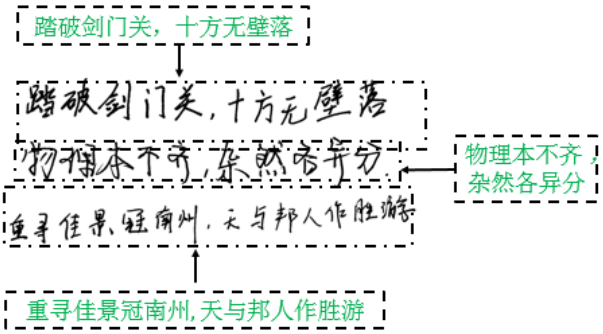


图 5-5 识别结果示例图

Fig. 5-5 Example image of recognition results

本网络与其它常用到的识别算法进行了比较，在进行比较的时候各算法都采用一致的数据集，以确保实验数据具有可比性，其数据比较如表 5-3 所示。

表 5-3 不同算法准确率的比较

Table 5-3 Comparison of accuracy of different algorithms

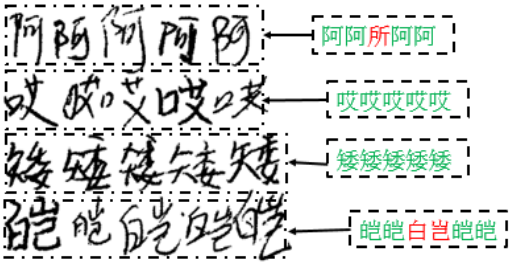
模型	准确率
自适应算法	98.67%
端对端[85]	98.23%
改进 AlexNet	99.75%
本实验方法	99.83%

从上表中可以看出，在采用同一数据集的情况下本实验的模型对汉字的识别率都略高于其他模型，相比其他传统方法，本实验的方法在识别汉字的准确率上占有一定的优势，网络模型在进行特征提取时会根据加权系数的不同，对不同的区域采取不同的关注度，使得汉字结构特征变得更加突出，从而使得提高了汉字识别的准确率。

5.2 系统的实现

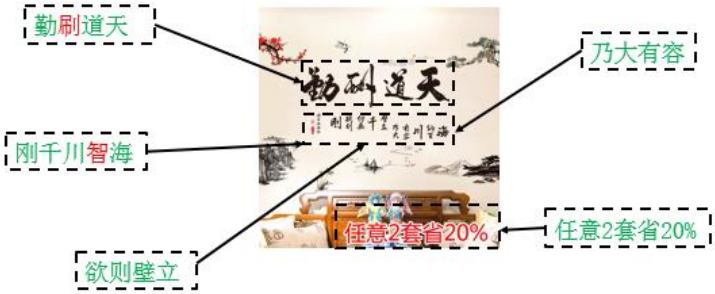
在上节中实现对简单背景下手写汉字的识别，网络模型对手写汉字有很高的识别率，在第四章中实现了对不同背景情况下汉字的检测，改进后模型的检测效果有所提高，使得其对汉字区域的检测准确有效。本节主要是介绍不同背景下汉字的识别，该识别系统适合多场景、多语种、整图汉字的检测和识别，同时也支持生僻字识别，并对识别结果进行采样分析，为后续改进和提高识别系统的性能提供数据对比。

汉字的检测和识别是模式识别中一个非常重要的任务，从自动化办公到数字图书馆经过广泛地研究，可以将汉字的识别分成两大部分：一方面是对书籍等图像中的汉字的检测和识别，另一方面则是对自然场景下汉字的检测和识别，例如广告牌、碑文等[86]。目前，对于前者的实现已经可以达到很高的识别率，在生活和工作中的软件工具已经很成熟了，但是自然场景下汉字的检测和识别仍然存在许多困难，例如字体的多样性、受光照影响大、汉字背景复杂等，这些问题都给检测和识别带来了挑战。在对图像中的汉字区域进行检测前，将图像输入网络模型中进行识别，数据集中部分图像的识别效果如图 5-6 所示。



图（a）相同汉字识别的示例图

Fig.（a）Example image of the same Chinese character recognition



图（b）书法体识别的示例图

Fig.（b）Example image of calligraphy recognition

图 5-6 汉字区域检测前的识别结果示例图

Fig. 5-6 Example image of recognition result before Chinese character area detection

如上图所示的识别结果，绿色表示识别正确，红色表示识别错误。目前手写汉字识别技术虽然已经非常成熟了，但是对于一些特定场合或者一部分汉字的识别仍然存在问题。相比复杂的汉字字体，简单汉字的检测和识别都可以达到很高的准确率，但是有很多汉字结构之间存在着很高的相似度，使得识别难度增加，最后导致识别错误，例如图（a）中‘阿’和‘所’具有一定的相似度，而且‘皑’在识别时被检测成了‘白’和‘岂’两个字，对汉字的手写不是特别规范，最后导致识别出现偏差。对于字体结构复杂的汉字，在网络进行识别匹配的时候容易出现偏差，例如图（b）中‘酬’书写的是书法字体，并且字体相对复杂，网络根据汉字的相似度进行匹配，最后被识别成‘刷’，图中‘纳百’是上下结构进行书写的，两个字结构紧密连接，被检测成一个字体‘智’。

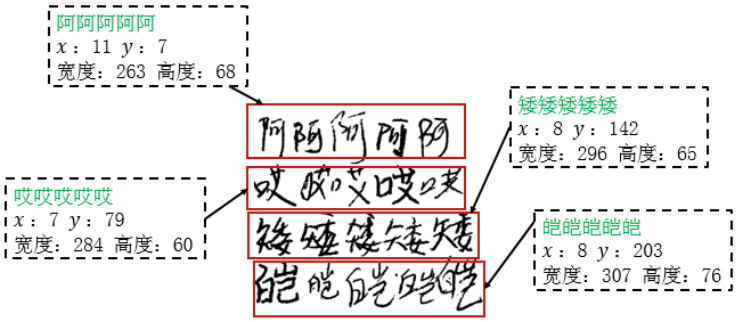
接下来，文章将会通过对汉字区域进行检测标定文本框后，再输入网络模型中进行识别。首先要判断图像中汉字背景的类别，例如：背景是否复杂、检测面是否凹凸、汉字字体的大小、汉字是否被遮挡等，接下来对目标图像中的汉字区域进行检测，对预测出的汉字区域进行文本框的标定，根据对文字区域的判断进行后期汉字的识别，识别结果包括汉字内容及位置数组等。因为数据集的多样性问题，在识别的过程中选取了几种不同类型的汉字进行识别，包括：背景复杂的图像、平面汉字、书法字体、不同人手写的相同汉字以及曲面汉字的识别。识别结果示例图，如图 5-7 所示。

图（a）复杂背景汉字识别的示例图

Fig.（a）Example image of Chinese character recognition in complex background

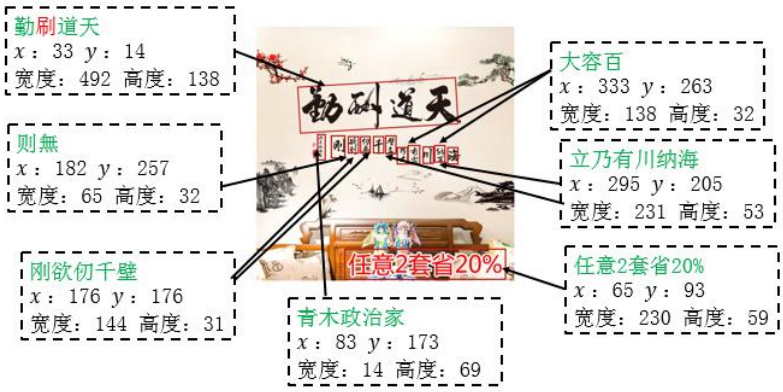
图（b）水平面汉字识别的示例图

Fig.（b）Example image of horizontal Chinese character recognition



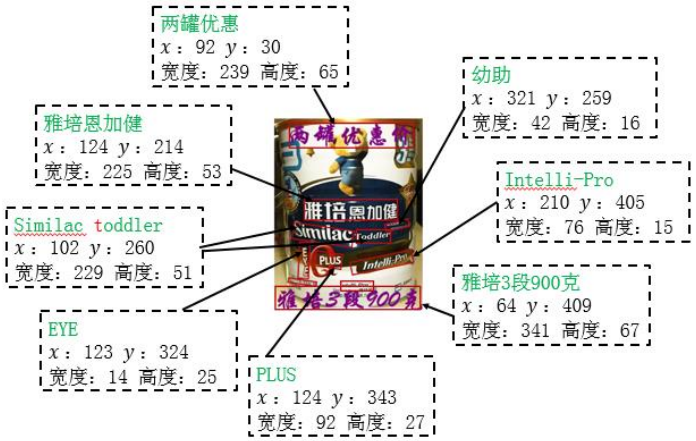
图（c）相同汉字识别的示例图

Fig.（c）Example image of the same Chinese character recognition



图（d）书法字体识别的示例图

Fig.（d）Example image of calligraphy font recognition



图（e）曲面汉字识别的示例图

Fig.（e）Example image of surface Chinese character recognition

图 5-7 不同背景汉字识别结果示例图

Fig.5-7 Examples of Chinese character recognition results in different backgrounds

在汉字识别的反馈结果中包括汉字的内容、定位位置、识别结果的数目、文本框四个顶点的坐标、汉字所在行的置信度以及输入图片的朝向，其中位置数组中 x 和 y 分别表示识别时定位位置的矩形左上顶点的水平坐标和垂直坐标，宽度和高度分别表示识别时定位位置的矩形的宽度和高度，输出的置信度包括行置信度的平均值、最小值和方差三个值，而识别结果的数目是指反馈的汉字元素组的个数，图片朝向是检测图像的方向，其中‘-1’、‘0’、‘1’、‘2’和‘3’分别表示图像未定义、正向、逆时针 90 度、逆时针 180 度和逆时针 270 度。

在实验中对于复杂背景汉字的检测和识别的准确率都很高，如以上示例图中对汉字识别结果的反馈可以看出，当我们对汉字区域进行准确的检测标定文本框后，汉字的识别准确率得到了提升，例如图（c）中‘阿’和‘所’被准确地区分开，识别输出正确结果，而且对于‘皑’也都识别正确。对于字体结构复杂的书法字体，识别结果中仍然存在识别错误的问题，例如图（d）中‘酬’依然被识别成‘刷’，但是图中‘纳百’被精确地识别正确。而在曲面图像将首先是将曲面分割成多个平面对汉字进行检测，最后对汉字的识别出错率不是最高的，例如图（e）中只是出现对英文的大小写识别错误，说明我们将曲面分割成多个平面的思路是有效的。以上是对数据集中部分汉字的示例展示，但是足以证明对汉字区域进行准确的检测有利于提升网络的识别率。在实验的数据集中，根据不同情况统计了汉字识别的数据，其数据如表 5-4 所示。

表 5-4 汉字识别的数据表

Table 5-4 Data sheet for Chinese character recognition

类型	召回率	精确率	F1 得分
复杂背景	0.816	0.852	0.834
遮挡背景	0.638	0.695	0.665
曲面背景	0.751	0.786	0.768

相同汉字	0.807	0.847	0.827

从不同背景下汉字的数据中可以看出，对于不同人书写的相同汉字的召回率并不是最高的，说明对于字体结构不清晰的汉字进行相似度匹配时出现错误，对于网络后续的改进可以反馈与之相似度高的汉字，并且可以在时间上对汉字序列进行识别，进一步对网络进行语义的训练，例如图（d）中的‘天道酬勤’需要对网络进行语义训练。另外，由于数据集的不全面性以及汉字的多样性，在实验中肯定对其他情况存在遗漏，所以需要进行更多的实验积累数据，后期将对实验数据进行更进一步的统计。

5.3 本章小结

本章对汉字识别系统展开了不同的实验，并对实验数据进行了总结与分析，其中包括基于卷积神经网络 LeNet-5 的手写汉字模型的构建，并对其实验过程和结果进行了介绍说明。在系统的实现中，本实验对不同背景情况下的汉字进行了检测和识别，并达到了不错的效果，在反馈结果中可以看到汉字的内容、识别时的定位位置、文本框的顶点坐标、每行的置信度以及图像的朝向。对于识别过程中出现的错误，需要后续对相似度高的汉字进行训练，并对网络进一步实现语义识别，以实现网络对汉字识别达到更高的准确率。

6 总结与展望

6.1 全文总结

本文提出深度学习在手写汉字识别中的应用研究，主要采用神经网络模型对多种背景情况下的手写汉字进行识别，提升了网络模型在不同背景情况下手写汉字识别的精确率。本文的主要工作如下：

1. 阐述了研究手写汉字识别的背景以意义，简单介绍了国内外汉字识别发展历程，对汉字识别中的神经网络算法，神经网络基础架构进行了说明，并且对汉字识别中常用到的神经网络模型进行系统的分析，其中包括 LeNet-5 网络、VGG 网络以及 ResNet 网络，根据各网络模型的特点，能够解决汉字识别过程中遇到的不同问题，而且还分别介绍了不同类型的中文数据集。

2. 本文对多种背景下汉字的检测模型进行了设计，主要是基于 AdvancedEAST 的网络模型，实验中对主要的神经网络 VGG16 进行了改进，用 RoI 层代替其最大的池化层，并且在实验中改进了损失函数，使用 Dice 损失和 DIoU 损失函数，以达到对文本框进行更好的预测。

3.本文利用深度学习 TensorFlow 框架进行经典手写汉字识别模型的搭建，其中在基于 LeNet-5 网络模型的基础上加入区域加权系数，使网络模型对特征图的不同区域给予不同的关注度，能够让汉字轮廓变得更加明显，以此来提高汉字的识别率。经过上述对经典手写汉字识别模型的搭建，本文在此基础上实现了系统对汉字的识别。

4. 综上实验，本文实现了对手写汉字识别系统的设计，该系统可以对不同背景情况下的汉字进行识别，并且在反馈结果中会显示汉字的内容、定位位置、识别结果的数目、文本框四个顶点的坐标、汉字所在行的置信度以及输入图片的朝向，对比不同的实验数据，并对汉字区域检测前后进行了对比实验，结果证实了对汉字区域准确有效的检测可以提升网络模型对汉字识别的准确率。

6.2 论文的创新点

本文的创新点如下：

1. 在复杂背景情况下对汉字的检测模型中，设计了 AdvancedEAST 的网络模型，实验改进了主要网络模型 VGG16，使用 RoI 层代替其最大的池化层，可以利用 RoI 层对图像进行多尺度变换，最后对图像统一到相同大小的尺寸，采用 Dice 损失和 DIoU 损失函数，对文本框进行更好的预测，达到对汉字区域准确的检测。

2. 本文设计的对手写汉字识别的经典模型中，在 LeNet-5 网络模型的基础上加入区域加权系数，使模型提高了对汉字的识别率，并且用激活函数 ReLU 代替 Sigmoid，改善了网络梯度消失的情况，并且加快了网络模型的收敛。

6.3 实验不足之处

1. 本文在基于 LeNet-5 模型的实验，用到了激励函数 ReLU，其虽然收敛速度快，

但是也是非常脆弱的，它对网络的稀疏处理可能会导致模型不能进行有效的学习。

2. 在对汉字进行检测时，文本框数量的变化使得训练参数增多，增加了网络的计算量，导致网络模型在训练时对数据集的训练时间变长，另外，实验中的后处理应用的 NMS 算法，在实验中没有对垂直文本进行系统的训练，造成实验数据不全面。

6.4 展望

本文通过神经网络模型对手写汉字识别系统进行了设计与实现，对提高汉字区域的检测和识别具有一定的实用价值和参考作用。本系统虽然可以提高网络模型对手写汉字识别的准确率，但是后续的研究中仍然存在很多需要改进和完善的地方，后续研究中对本系统的优化方案如下：

1. 在后续的研究中解决汉字区域的位置，可以尝试在前期就对数据增加宽和高，为了对文本框能够更好的预测，在实验中可以对预测框得分进行排序，根据得分情况选择最终的文本框。

2. 由于字体的多样性导致识别中出现错误，在进行网络训练时可以加入语义识别、多角度检测以及相似汉字的反馈，并根据相似度进行排名，以解决汉字相似度带来的误差。

3. 相比简单背景的汉字，背景复杂的情况对汉字的检测和识别就有很多困难需要解决了，首先是对数据集的选择，例如对曲面的汉字检测识别少，导致其数据集也是缺少的，而在训练网络时需要大量的数据集，当然，可以通过网络搜集曲面图像，例如：圆柱建筑物、饮料瓶、足球、硬币等都属于曲面物体，通过拍照可以得到一定数量的曲面图像，但这远远不足以训练我们的模型，所以未来对构建庞大的曲面数据库也是我们要面对的一大棘手问题，以便全面的识别数据给后续的研究带来更加有价值的参考意义。

7 参考文献

[1] Tamhankar PA，Masalkar KD，Kolhe SR．A Novel Approach for Character Segmentation of Offline Handwritten Marathi Documents written in MODI Script [J]．Procedia Computer Science，2020，171：179-87．

[2] Liu C-L，Yin F，Wang D-H，et al．Online and offline handwritten Chinese character recognition：Benchmarking on new databases [J]．Pattern Recognition，2013，46(1)．

[3] Li Z，Wu Q，Xiao Y，et al．Deep Matching Network for Handwritten Chinese Character Recognition [J]．Pattern Recognition，2020，107．

[4] 文中芳，孙新杰．基于大数据下的手写体识别的设计与研发 [J]. 科技风，2020，(03)：32-3．

[5] 罗麟，张非，位一鸣，等．基于卷积神经网络的电力操作票文字识别方法 [J]．浙江：浙江电力，2020，39(04)：68-74．

[6] 丁蒙，戴曙光，于恒．卷积神经网络在手写字符识别中的应用 [J]．软件导刊，2020，19(01)：275-9．

[7] Tao D，Liang L，Jin L，et al．Similar handwritten Chinese character recognition by kernel discriminative locality alignment [J]．Pattern Recognition Letters，2014，35．

[8] 侯一民，周慧琼，王政一．深度学习在语音识别中的研究进展综述 [J]．计算机应用研究，2017，34(08)：2241-6．

[9] 吕亮．基于深度学习的说话人识别方法的研究 [D]．南京：东南大学，2016．

[10] 刘尚林，王佳．大数据下人工智能与书法的互融 [J]．中国书法，2020，(08)：198-9．

[11] 黄洋，谭钦红，施新岚．结合八方向梯度特征和 CNN 的相似手写汉字识别 [J]．信息通信，2019，(04)：5-8．

[12] 张政馗，庞为光，谢文静，等．面向实时应用的深度学习研究综述 [J]．软件学报，2020，v.31(09)：34-57．

[13] Deng L，Yu D，Processing TIS．Deep Learning：Methods and Applications [J]．Foundations，2014，7(3)．

[14] Zhi T，Huang W，Tong H，et al．Detecting Text in Natural Image with Connectionist Text Proposal Network．proceedings of the European Conference on Computer Vision[C]．Amsterdam，The Netherlands：European，2016．

[15] Gal Y，Ghahramani Z．Dropout as a Bayesian Approximation：Representing Model Uncertainty in Deep Learning [J]．JMLRorg，2015．

[16] Prashanth DS，Mehta R，Sharma NJPCS．Classification of Handwritten Devanagari Number-An analysis of Pattern Recognition Tool using Neural Network and CNN [J]．Procedia Computer Science，2020，167：2445-57．

[17] 陈昊，郭海，刘大全，等. 基于 TensorFlow 的手写数字识别系统 [J]. 信息通信，2018，(03)：108-10.

[18] 罗昱成. 场景字符识别综述 [J]. 现代计算机，2020，(04)：32-6.

[19] 余彦超，李绍翔，张开轩，等. 手写识别器的设计与制作 [J]. 科技创新与应用，2020，(12)：42-3.

[20] 任晓文，王涛，李健宇，等. 基于深度学习的异噪声下手写汉字识别的研究 [J]. 计算机应用研究，2019，36(12)：3878-81.

[21] 王文超. 面向大词汇量离线中文手写识别的简约建模方法研究 [D]. 安徽：中国科学技术大学，2019.

[22] 竺博，吴嘉嘉，何春江，等. 人工智能在手写文档识别分析中的技术演进 [J]. 电子测试，2019，(13)：5-8+48.

[23] Shahmoradi S， Shouraki SB. Evaluation of a Novel Fuzzy Sequential Pattern Recognition Tool (Fuzzy Elastic Matching Machine) and its Applications in Speech and Handwriting Recognition [J]. Applied Soft Computing，2017：S1568494617306415.

[24] Choudhury H， Prasanna SRM. Handwriting recognition using sinusoidal model parameters [J]. Pattern Recognition Letters，2018，121.

[25] 杜泽炎，任明武. 一种在低质量图像上提高字符识别率的深度学习框架 [J]. 计算机与数字工程，2019，47(06)：1491-6.

[26] 曾喆昭. 神经网络优化方法及其在信息处理中的应用研究 [D]. 长沙：湖南大学，2008.

[27] Graves A，Fernández S，Gomez F. Connectionist temporal classification：labelling unsegmented sequence data with recurrent neural networks [J]. ACM，2006.

[28] Wu Y C，Yin F，Chen Z，et al. Handwritten Chinese Text Recognition Using Separable Multi-Dimensional Recurrent Neural Network. proceedings of the 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)[C]. Kyoto，Japan ： IAPR，2017.

[29] 王俊淑，张国明，胡斌. 基于深度学习的推荐算法研究综述[J]. 南京：南京师范大学学报(工程技术版)，2018，18(04)：33-43.

[30] 文中芳，孙新杰. 基于大数据下的手写体识别的设计与研发[J]. 科技风， 2020，(03)：32-3.

[31] 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]. 黄石：湖北理工学院学报，2019，35(02)：31-4.

[32] Huang L，Yang D，Lang B，et al. Decorrelated Batch Normalization [J]. IEEE，2018.

[33] 梅啟成，吕文阁. Research on Commodity Image Recognition Based on Deep Learning [J]. 机电工程技术，2018，047(009)：28-31，151.

[34] 魏炳辉，谢晖慧，邓小鸿． 一种多模型超图用于手写汉字识别算法 [J]．计算机应用与软件，2019，36(07)：192-6+201．

[35] Schölkopf B， Platt J， Hofmann T． Learning with Hypergraphs: Clustering，Classification，and Embedding．proceedings of the Advances in Neural Information Processing Systems 19：Proceedings of the 2006 Conference [C]．USA：MIT Press，2007．

[36] Huang S，Elhoseiny M，Elgammal A， et al． Learning Hypergraph-regularized Attribute Predictors．proceedings of the IEEE Computer Society[C]．USA：IEEE，2015．

[37] Sari E Y，Kusrini K，Sunyoto A．Analisis Akurasi Jaringan Syaraf Tiruan Dengan Backpropagation Untuk Prediksi Mahasiswa Dropout [J]． Creative Information Technology Journal，2021，6(2)：85．

[38] Szegedy C，Vanhoucke V，Ioffe S，et al．Rethinking the Inception Architecture for Computer Vision [J]．IEEE，2016：2818-26．

[39] 徐颂民，江玉珍． 基于 TensorFlow 的 CNN 自由手写数字识别研究 [J]．电脑知识与技术，2020，16(13)：11-2．

[40] Xiao X，Jin L，Yang Y， et al． Building fast and compact convolutional neural networks for offline handwritten Chinese character recognition [J]．Pattern Recognition，2017，72．

[41] Ketkar N． Convolutional Neural Networks [J]．Springer International Publishing，2017．

[42] Lecun Y，Boser B，Denker J， et al．Backpropagation Applied to Handwritten Zip Code Recognition [J]．Neural Computation，2014，1(4)：541-51．

[43] Lecun Y， Bottou L． Gradient-based learning applied to document recognition [J]．Proceedings of the IEEE，1998，86(11)：2278-324．

[44] 杨佶．基于深度学习的手写汉字识别技术研究 [D]．沈阳：沈阳师范大学，2019．

[45] Tian Z，Huang WHe T， et al．Detecting Text in Natural Image with Connectionist Text Proposal Network [J]．2016．

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

[47] Zheng Z， Zhang CWei S， et al． Multi-oriented Text Detection with Fully Convolutional Networks [J]．IEEE，2016．

[48] Zihan Y．Classification of picture art style based on VGGNET [J]．Journal of Physics: Conference Series，2021，1774(1)．

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

[50] He K， Zhang X， Ren S， et al． Deep residual learning for image recognition．proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition[C]．USA：IEEE，2016．

[51] Netzer Y，Wang T，Coates A，et al． Reading Digits in Natural Images with Unsupervised Feature Learning [J]．2011．

[52] Kai W，Babenko B，Belongie S．End-to-end scene text recognition．proceedings of the IEEE International Conference on Computer Vision [C]．USA：IEEE，2012，2993-3003．

[53] Hinton G E，Srivastava N，Krizhevsky A，et al．Improving neural networks by preventing co-adaptation of feature detectors [J]．Computer Science，2012，3(4)：págs．212-23．

[54] Cong Y，Xiang B，Liu W，et al．Detecting texts of arbitrary orientations in natural images．proceedings of the Computer Vision & Pattern Recognition[C]．USA：IEEE，2012，1735-1742．

[55] Yuan T L，Zhe Z，Xu K，et al．Chinese Text in the Wild [J]．2018．

[56] 金连文，钟卓耀，杨钊，等．深度学习在手写汉字识别中的应用综述 [J]．自动化学报，2016，42(08)：1125-41．

[57] 李国强，周贺，马锴，等．特征分组提取融合深度网络手写汉字识别 [J]．计算机工程与应用，2020，v.56．No.955(12)：169-74．

[58] 申倬栋，王泽举．不均匀光照条件下二值化图像处理的研究 [J]．电子元器件与信息技术，2020，v.4．No.32(02)：103-4+39．

[59] 周正扬．基于笔画顺序恢复的相似手写汉字识别方法研究 [D]．武汉：武汉理工大学， 2017．

[60] 全志楠，林家骏． 文本无关的小样本手写汉字笔迹鉴别方法 [J]．上海：华东理工大学学报(自然科学版)，2018，44(06)：882-6．

[61] 刘峡壁，贾云得．用于手写体汉字识别的汉字结构模型 [J]．北京：北京理工大学学报， 2003，23(3)：322-．

[62] 耿艳萍，高红斌，任智颖．融合颜色特征和纹理特征的图像检索算法 [J]．无线互联科技，2017，(24)：113-6．

[63] Yaar H，Ceylan M．A new deep learning pipeline to detect Covid-19 on chest X-ray images using local binary pattern， dual tree complex wavelet transform and convolutional neural networks [J]．Applied Intelligence，2020．

[64] Dalal N，Triggs B．Histograms of Oriented Gradients for Human Detection．proceedings of the IEEE Computer Society Conference on Computer Vision & Pattern Recognition [C]．USA：IEEE，2005．

[65] 罗勇，叶正源，陈远知． 高效并行递归高斯 SIFT 算法的实现 [J]．中国科技论

文，2015，10(20)：2382-5+94．

[66] Yong C，Hui Z，Yuan Z，et al．An Improved Regularized Latent Semantic Indexing with L1/2 Regularization and Non-negative Constraints; proceedings of the 2013 IEEE 16th International Conference on Computational Science and Engineering[C]．USA：IEEE，2014．

[67] Shah J， Qureshi I， Deng Y， et al．Reconstruction of Sparse Signals and Compressively Sampled Images Based on Smooth l1-Norm Approximation [J]．Journal of Signal Processing Systems，2017，88(3)：333-44．

[68] Yu J， Jiang Y， Wang Z， et al． Unitbox： An advanced object detection network．proceedings of the Proceedings of the 24th ACM international conference on Multimedia, [C]．Como，Italy：ACM，2016．

[69] Rezatofighi H，Tsoi N，Gwak J，et al．Generalized Intersection Over Union：A Metric and a Loss for Bounding Box Regression． proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)[C]． USA： IEEE/CVF，2019，1062-1071．

[70] Zheng Z，Wang P，Liu W，et al．Distance-IoU Loss：Faster and Better Learning for Bounding Box Regression． proceedings of the AAAI Conference on Artificial Intelligence [C]．New York，USA：AAAI，2020，420-428．

[71] Zhou X，Yao C，Wen H，et al．EAST：An Efficient and Accurate Scene Text Detector [J]．IEEE Conference on Computer Vision，2017．

[72] Neubeck A， Gool L． Efficient Non-Maximum Suppression． proceedings of the International Conference on Pattern Recognition[C]．Kowloon Tong，Hong Kong：International，2006．

[73] Qin H， Yan J， Xiu L， et al． Joint Training of Cascaded CNN for Face Detection．proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [C]．USA：IEEE，2016．

[74] Wen Y，Zhang K， Li Z，et al．A Discriminative Feature Learning Approach for Deep Face Recognition． proceedings of the European Conference on Computer Vision [C]．Amsterdam，The Netherlands：European，2016．

[75] 张婉．融合深度学习的图像分类算法研究 [D]．南京：南京邮电大学，2018．

[76] Cong Y， Xiang B， Nong S， et al． Scene Text Detection via Holistic，Multi-Channel Prediction [J]．2016．

[77] Milletari F，Navab N，Ahmadi S A．V-Net：Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation．proceedings of the 2016 Fourth International Conference on 3D Vision (3DV) [C]．Hong Kong：Fourth International，2016，499-515．

[78] Karatzas D，Gomez-Bigorda L，Nicolaou A，et al．ICDAR 2015 Robust Reading Competition [J].

[79] Fang H S，Cao J，Tai Y W，et al． Pairwise Body-Part Attention for Recognizing Human-Object Interactions [J]．Springer，Cham，2018．

[80] Liu Y，Jin L，Zhang S，et al． Detecting Curve Text in the Wild：New Dataset and New Solution [J]．2017．

[81] Kim K H，Hong S，Roh B，et al．PVANET：Deep but Lightweight Neural Networks for Real-time Object Detection [J]．2016．

[82] 刘芬，隋天宇，王叶群．TensorFlow 平台深度学习寻找最优路径问题研究 [J]．信息技术，2019，(02)：62-5+70．

[83] El-Sawy A，El-Bakry H，Loey M．CNN for Handwritten Arabic Digits Recognition Based on LeNet-5 [J]．Springer International Publishing，2016．

[84] 周有张．基于区域加权 LeNet-5 网络的汉字识别研究 [J]．计算机与数字工程，2020，v.48．No.373(11)：153-7+210．

[85] Baoguang S，Xiang B，Cong Y．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 and machine intelligence，2017，39(11) ．

[86]闫喜亮，王黎明． 卷积深度神经网络的手写汉字识别系统 [J]．计算机工程与应用，2017，53(10)：246-50．

8 攻读硕士学位期间发表论文情况

[1] Chunxia Zhang，Longxue Li，Xudong Li．A Survey of Chinese Character Recognition Research Based on Deep Learning[C]． 7th Annual International Conference on Network and Information Systems，2021．

[2]一种用于检测手写汉字区域的方法[P]．中国专利：202110477950.3，2021-04-30.

9 致　谢

时光匆匆，我的硕士生涯已经接近尾声,伴随结束的还有我的学生时代。在天津科技大学三年时光里，既漫长又短暂的岁月里有酸甜苦辣，但更多的是成长。感谢陪伴我一同成长的亲爱的同学们，也感谢见证我成长的尊敬的张老师。正是因为你们的陪伴与帮助，正是因为老师耐心的指导，我才能克服困难解决问题，顺利的完成学业。

本人的学位论文是在我的恩师张春霞副教授耐心指导下完成的,亲爱的张老师就像是自己的长辈一样,在生活上对我无微不至的关怀,她待人亲切，品德高尚，治学严谨，从课题的选择到论文的完成，在这个过程中张老师不仅教授我学习的技巧，还教授我做人的准则，对我本人来说都是受益匪浅的。在课题进展的过程中，张老师都会耐心听取的汇报，并及时提出需要改进的地方，我每一次取得的进步都离不开张老师的谆谆教导。在此谨向张老师致以崇高的敬意和衷心的感谢。

感谢 419 实验室的师兄师姐和同级的兄弟姐妹们,大家一起分享美食,一起游戏，在 419 实验室的 2 年是我硕士生涯难忘的历程，同时也感谢能与 524 实验室的兄弟姐妹们相遇。

感谢我硕士生涯最珍贵的朋友徐冰同学，她是我读书 20 年来最喜欢的小妹妹，感谢她对我的包容，感谢她能让我依赖和信任，我们彼此分享青春的美好。非常舍不得和她道别，但离别的日子就在眼前，将世间所有的美好都献给她，祝前程似锦！

感谢我的父母对我学业的支持！父母是最爱我的人，也是为我付出最多的人，他们是我前进的动力！我一定好好努力，给父母最好的回报以及更多的陪伴！

由衷的感谢为评阅本论文付出宝贵时间专家和教授们，您们辛苦了！

感谢一路走来关心、帮助过我的所有人，每个阶段的青春岁月都是美好的，因为有大家的参与而变得特别精彩。在以后的生活中我会更加勤奋的努力，争取做到最好，回报社会。祝健康，幸福！

最后，感谢自己！