

湖州师范学院

## 2019 届毕业设计（论文）

课 题 名 称： 基于深度学习的手写汉字质量评价方法

课 题 名 称（英文）： Handwritten Chinese character  
quality evaluation method based on deep learning

学 生 姓 名： 李思琦 学 号： 2019082108

专 业 名 称： 计算机科学与技术

指 导 教 师： 李重 职 称： 教授

所 在 学 院： 信息工程学院

完 成 日 期： 2023 年 4 月 19 日

## 基于深度学习的手写汉字质量评价方法

**摘要:**近年来,由于中国传统文化的不断演变以及科技的飞速进步,手写汉字的质量评价已经成为当今社会的一个重要组成部分,它可以用于各种领域,如信息收集、商务交流。由于汉字的庞大、复杂的结构、众多的相似之处,再加上不一定的书写方式,使得汉字的识别极具挑战性,而且评估其书写质量更是一项艰巨的任务。然而,随着深度学习的飞速进步,可以将复杂的函数表达清晰,并且可以准确地计算出样本的概率分布,被普遍地运用到了图像识别的各个方面。通过深度学习理论,我们开发了一种基于深度学习的手写汉字评价方法,旨在提高人们对汉语的使用效率。

为了解决传统的手写汉字识别精度较差的问题,本文提出了一种基于注意力的深度可分离卷积神经网络,将传统的网络模块用 MBConv 替代,并利用注意力机制进行特征权重提取,提高了传统的手写汉字识别精度。同时针对当前手写汉字质量评价标准中出现的差异性,我们开发了一种以相似性权重来评价汉字质量。首先,我们确定一种特定的汉字种类,将构成一个由这些种类的汉字构成的评估小组,每个小组都包含一些具体的汉字,我们可以根据这些小组的特征来给予它们不同的等级,以实现汉字质量的精准评估。为了验证本文提出的算法的有效性,我们用中科院自动化研究所的 CASIA-HWDB1.1 手写汉字数据集进行实验验证,实验结果表明,本文提出的方案准确率和精确率得到显著改善,与原有的 mAP 相比,其准确度提高了9.04%,精确率提高了3.07%。同时,针对本文提出手写汉字质量评价方式,将检查的结果与传统的方法进行了比较,本文提出的手写汉字评价的准确度和可信度都较高于行业内平均标准,有效地解决了当前行业内手写汉字评价效率及准确率不高的问题。

**关键词:**深度学习, 手写汉字, 质量评价, 深度可分离网络

# Handwritten Chinese character quality evaluation method based on deep learning

**Abstract:** In recent years, due to the continuous evolution of Chinese traditional culture and the rapid progress of science and technology, the quality evaluation of handwritten Chinese characters has become an important part of today's society, which can be used in various fields, such as information collection and business communication. Due to the huge, complex structure, numerous similarities of Chinese characters, and the uncertain writing methods, the recognition of Chinese characters is extremely challenging, and it is a difficult task to evaluate their writing quality. However, with the rapid progress of deep learning, complex functions can be clearly expressed, and the probability distribution of samples can be accurately calculated. Deep learning is widely applied to all aspects of image recognition. Through deep learning theory, we developed a handwritten Chinese character evaluation method based on deep learning, aiming to improve the efficiency of people's use of Chinese.

In order to solve the problem of poor accuracy of traditional handwritten Chinese character recognition, this paper proposes an attention-based deep separable convolutional neural network. The traditional network module is replaced with MBConv, and the attention mechanism is used to extract feature weights, which improves accuracy of traditional handwritten Chinese character recognition. We also develop a similarity weight to evaluate the quality of handwritten Chinese characters in response to the variability that appears in current handwritten Chinese character quality evaluation criteria. First, we identify a specific kind of Chinese characters that will constitute an evaluation group consisting of these kinds of Chinese characters, each group contains some specific Chinese characters, and we can give them different grades according to the characteristics of these groups in order to achieve accurate evaluation of Chinese character quality. In order to verify the effectiveness of the algorithm proposed in this paper, we use the CASIA-HWDB1.1 handwritten Chinese character dataset from the Institute of Automation, Chinese Academy of Sciences for experimental validation. The experimental results indicate that the accuracy and precision rate of the scheme proposed in this paper are significantly improved, with a 9.04% increase in accuracy and 3.07% increase in precision rate compared with the original mAP. Meanwhile, for the handwritten Chinese character quality evaluation method proposed in this paper, the results of the check were compared with the traditional method. The accuracy and credibility of the handwritten Chinese character evaluation proposed in this paper were higher than the average standard in the industry, which effectively solved the current problem of low efficiency and accuracy of handwritten Chinese character evaluation in the industry.

**Keywords:** Deep learning, handwritten Chinese characters, quality evaluation, deep separable networks

## 目 录

|                                 |    |
|---------------------------------|----|
| 第一章 前言 .....                    | 1  |
| 1.1 研究背景及意义 .....               | 1  |
| 1.1.1 研究背景 .....                | 1  |
| 1.1.2 研究意义 .....                | 1  |
| 1.2 国内外研究现状 .....               | 1  |
| 1.3 发展趋势 .....                  | 2  |
| 1.4 本文研究内容 .....                | 2  |
| 1.5 本文组织结构 .....                | 3  |
| 第二章 相关技术与理论 .....               | 4  |
| 2.1 神经网络技术 .....                | 4  |
| 2.1.1 神经网络 (NN) .....           | 4  |
| 2.1.2 卷积神经网络 (CNN) .....        | 5  |
| 2.2 手写汉字识别技术分析 .....            | 6  |
| 2.2.1 传统手写汉字识别技术 .....          | 6  |
| 2.2.2 基于卷积神经网络的手写汉字识别技术 .....   | 7  |
| 第三章 基于注意力机制的深度可分离网络手写汉字识别 ..... | 8  |
| 3.1 手写汉字识别基本流程 .....            | 8  |
| 3.1.1 手写汉字基本流程 .....            | 8  |
| 3.1.2 手写汉字图像预处理 .....           | 8  |
| 3.2 手写汉字识别卷积神经网络结构设计 .....      | 9  |
| 3.2.1 深度可分离卷积 .....             | 9  |
| 3.2.2 深度可分离卷积网络结构设计 .....       | 10 |
| 3.2.3 注意力机制 .....               | 11 |
| 3.3 模型训练及结果分析 .....             | 12 |
| 3.3.1 实验环境 .....                | 12 |
| 3.3.2 数据集 .....                 | 13 |
| 3.3.3 实验结果及分析 .....             | 13 |
| 第四章 手写汉字质量评价 .....              | 17 |
| 4.1 手写汉字质量评价组构建 .....           | 17 |
| 4.2 手写汉字质量评价标准 .....            | 17 |
| 4.3 基于相似度匹配的手写汉字质量评价方法 .....    | 18 |
| 4.3.1 模糊综合评价 .....              | 18 |
| 4.3.2 一对多评价方式 .....             | 19 |
| 4.4 手写汉字质量评价系统的实现 .....         | 22 |
| 4.5 手写汉字质量评价实验结果分析 .....        | 23 |
| 4.5.1 数据集 .....                 | 23 |

|                  |    |
|------------------|----|
| 4.5.2 实验结果 ..... | 23 |
| 第五章 总结与展望 .....  | 27 |
| 5.1 总结 .....     | 27 |
| 5.2 展望 .....     | 27 |
| 参考文献 .....       | 28 |
| 致谢 .....         | 30 |

## 第一章 前言

本节探讨了基于深度学习的手写汉字质量评价的研究背景及意义，国内外研究现状、当前手写汉字质量评价的发展趋势，以及本文的全文安排。

### 1.1 研究背景及意义

#### 1.1.1 研究背景

随着现代社会的进步，以及人类对新知识的迅速掌握，使得文字的使用范围和方式也在大大扩大，从原本的纸质文件转变成了可以在电脑上快速、便捷地进行的数据库<sup>[1]</sup>，以满足当今世界的信息共享和交流的需求。采取图像识别技术来取代传统的人工输入，可以显著减少文本记忆所需要花费的时间和精力，同时也可以避免可能发生的错误。特别是对于全球使用量最高的中国汉语，采取这种技术可以有效地提高记忆效率，从而更好地满足社会对于智慧化记忆的需求。汉字识别技术不仅可以用于文本输入，还可以被广泛地应用于各种领域<sup>[2]</sup>，如银行支付、电子信函、考试、身份验证。

根据识别目标的类型，我们把汉字识别分为两类：印刷文字和手绘文字。其中，印刷文字的结构比较严谨，识别困难，相关的研究也比较丰富<sup>[3]</sup>，在日常的生活当中也得到了普遍的使用。而手绘文字则是通过扫描仪或摄像机捕捉的离线文字，它们的形状、位置、颜色以及其他外部条件的变化，使得能够提供准确信息的特征数量相当有限。随着科技的发展，汉字的繁体、异体、异义以及书写者的技巧，以及其他因素，如笔画的变化、拼音的变化、拼凑的方式的变化，都让人们对汉字的认知变得越来越艰巨。

#### 1.1.2 研究意义

随着技术的发展，许多学术界都开始关注和探索手写汉字的识别<sup>[4]</sup>，其中包括图像处理、模型认知、人工智能、数字信号处理以及其他多种技术。这些技术的发现为解决当今社会中复杂的文字问题提供了有力的支持，也促使许多领域的发展。尽管近年来在手写汉字的识别方面取得一定的进展，但由于许多技术尚未完全解决，其在现代社会中的运行还存在一定的挑战。因此，进一步深入探索该技术的发展，将会为我们带来更加丰富的知识，以及更好的社会服务。

近些年来，深度学习理论取得了长足的发展，它以人类大脑的神经系统为基础<sup>[5]</sup>，采用多层网络结构，对数据信息进行深度学习，从而实现了计算机视觉、图像识别、图像分类、自然语言处理等领域的高效运算，取得了显著的成果。学术界已经开始利用深度学习技术来重新审视和改进手写汉字识别算法，这一技术的应用已经超越了传统的特征提取方法，取得了显著的成果。因此，利用深度学习技术来解决手写汉字识别问题，具有十分重要的研究价值。

### 1.2 国内外研究现状

当今，在手写输入的辨认方法中，各种辨认方法都各具特色，它们各具特点，但是其辨认效果却大致如此。目前，K最近邻方法、神经网络、隐马尔可夫模型、贝叶斯分类器以及支持向量机(Support Vector Machine, SVM)都是广泛应用的方法<sup>[6]</sup>。在应用mqdf这种改良的二次判别函数时，首先需要将原始的图片经过多种不同的预处理技术，包含图像二值化、规则化、线性规则化和非线性规则化，以减少图片的畸变，让它们能够更加准确地表达出来，然后再通过特征提取方法，将具备一定维度的特征向量转换为可以识别的信息，从而实现mqdf的高效应用。在给出分析结果时，应该考虑到不同的环境，因此，在确保分析结果的准确性的前提下，应该采取线性判别分析方法(Linear Discriminant Analysis, LDA)和主成分分析(Principal Component Analysis)等多种方法<sup>[7]</sup>，对分析结果提供有效的降维处理，并

利用 MQDF 分类器对其加以划分。MQDF 分类器具备优异的特征表示性，然而，它的对于复杂的数据的性能却不尽如人意，因此，许多研究人员纷纷致力于将其作为一种替代的解决方案，并发表了大量的相关论著。本文旨在探讨如何将 MQDF 的参数估计技术应用于模型识别，以及如何将其转换成更加有效的手写自识别技术<sup>[8]</sup>。我们采用一种全面的技术，从多个训练集的参数分布中筛选出最佳的参数，从而实现更加准确的分类效果。

本文旨在探讨一种新的、有效的二次语言学习方法，它将传统的最小距离、多项式核函数，以及支撑向量机等技术有机地融入到语音识别中，以期达到更高的语音识别精度。此外，本文还提出了一种基于卷积神经网络的手写汉字识别系统<sup>[9]</sup>，它可以有效地收集到大量的词汇，而且可以有效地检测到每个词汇的含义，从而达到更好的语音识别效果。本研究开发的一种全新的特征矩阵，利用梯度特征克服高维向量散点矩阵的复杂性，显著减少手写字的辨认失败的可能。此外，本研究还采用非循环图支持向量机，对手写汉字进行准确的辨认，在处理少量字符的情况下表现良好，而在处理复杂的大规模字符时，其辨认速度明显减缓，从而影响其准确性。近年来，深度学习理论的发展推动了 KDLA 子空间学习算法的发展，它采用了最先进的主成分分析技术，可以大大改善信息提取的精度，从而极大地改善了字符识别的性能。这种算法的发展，使得它能够极大地改善人类的工作流程，并且深深地影响到了人工智能的发展。

### 1.3 发展趋势

评估手写汉字的质量不仅仅局限于遵守书写规范，还应该考虑它的美观度、笔画质量、结构和章法等方面。传统的评估方式往往会受到教师主观因素的影响，因此很难准确地反映出学生的书写水平。评估手写汉字的质量是一项重要的工作，它需要使用计算机来衡量和评估手写汉字的质量。传统的方法是通过教师的经验和观察来判断手写汉字的整体结构和笔画<sup>[10]</sup>，并将其与标准汉字进行比较，以得出评估。使用计算机辅助技术来评估手写汉字的书写质量，可以通过分析手写汉字的特征和模板汉字的特征来判断两者之间的差异，如果差异较小且两者相似度较高，则表明手写汉字的书写质量较高。该评估方法的基本框架可以参考图 1-1。

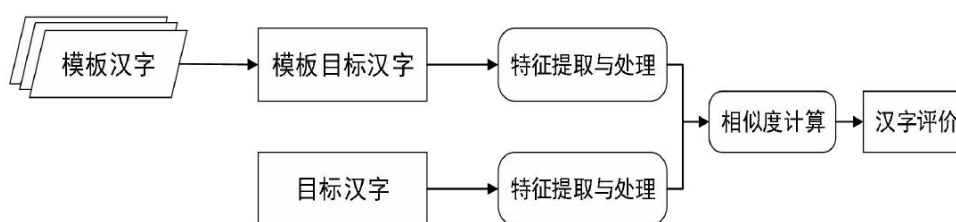


图 1-1 汉字书写质量框图

近年来，随着技术的发展，人们发现，采用学习的技术可以显著改善手写汉字的识别精度。目前，中国的这种技术已取得长足的进步，其中的许多产品都受到广泛的认可和赞誉<sup>[11]</sup>。2012 年，百度开始探索和研究学习，经过多次测试，它已经获得了显著的巨大成功，尤其是在语言和图形辨识方面。2013 年，百度创建了 DL 学习研究所，随后，腾讯、阿里和 360 也纷纷投入到这一领域，建设起自己的研究团队和实验室。2019 年 3 月，腾讯优图的 OCROCR 技术获得突破性进展，90% 的手写汉字和 95% 的手写英文单词的识别精度都超过同期的其他竞争者，“华为云智能 OCR 数据化产品”更是在 5 月，以其出色的技术表现，荣膺“2019 数博会领先科技成果奖”的荣誉，成为同类产品中的佼佼者。

### 1.4 本文研究内容

这一部分讨论的重点是如何通过深度学习技术来解决手写汉字的问题。我们的研究重点是设计一个

优化的卷积神经网络，以便更好地理解人类的语言能力，从而更有效地帮助人们学习汉语。采用基于相似性的质量评估技术，我们可以根据汉字的特征和语言特点，确定不同的汉字，然后构成一个特定的评估小组，以此来衡量每个汉字的表现，最终得到每个汉字的总体得分。

## 1.5 本文组织结构

本文各个章节组织如下：

第1章为前言。首先对手写汉字质量评价的研究背景进行介绍，重点阐述了手写汉字评价的研究背景及意义，以及当前国内外对于手写汉字评价的研究现状，对于当前形势的介绍。最后说明了各部分章节安排及研究内容。

第2章为相关技术及理论基础。对本文用到的理论知识进行梳理介绍，包括神经网络、卷积神经网络、手写汉字识别相关技术。

第3章为基于注意力机制的深度可分离网络手写汉字识别。包括手写汉字的基本流程，手写汉字卷积神经网络结构设计以及本文相关的模型训练及结果分析。

第4章为手写汉字质量评价。包括手写汉字质量评价组构建、本文提出的手写汉字质量评价标准以及基于相似度匹配的手写汉字质量评价方法，最后针对本文提出的手写汉字进行实验验证及结果分析。

第5章是本文的结论部分。



## 第二章 相关技术与理论

### 2.1 神经网络技术

#### 2.1.1 神经网络 (NN)

神经网络是人工神经网络的简称,如图 2-1 表示人脑神经元的基本结构图,主要是受到生物大脑中的中枢神经元联结方式的影响,所发展演化出来的一种模糊计算的神经网络计算模型,主要用来进行对相应的数学模型问题中较为复杂函数的近似逼近于数据拟合。

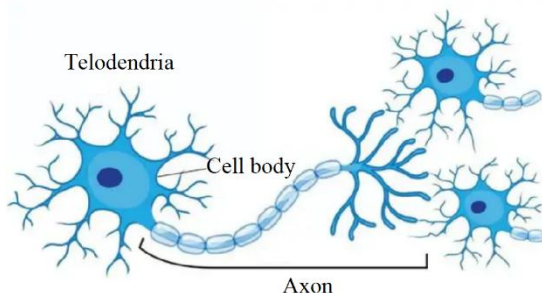


图 2-1 人脑神经元结构

神经网络主要是受人类脑部构造的影响,将大脑中的神经元结构通过类比模拟转换为人工神经网络,再通过和大脑结构相类似的拓扑构造转换成神经元,并把这些人工神经元连结为一个很像神经网的网络结构,就如同神经元上不同信息之间的传导方式那样,在神经网络中的人工神经元也具有同样的功用。神经元被普遍认为是整个神经网络系统中的最小单元,主要功用在于把整个神经网络系统中的某些输入输出信息内容,以函数为主要载体,并对它加以逻辑处理和传递<sup>[12]</sup>。下图 2-2 为一种典型人脑神经元结构,从图中可以看出单个神经元在针对信息数据处理的过程中的一些数据处理方式。

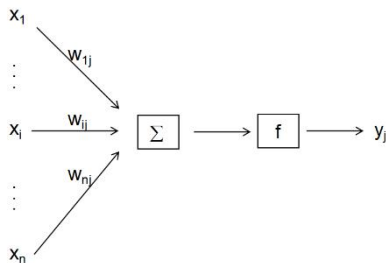


图 2-2 单个神经元模型

当有  $m$  个具体的神经元特征内容输入到神经元模型中时,各个特征的具体内容  $x_i$  会被加持一个权重  $w_i$ ,通过将其中具有不同权重的特征具体内容进行求和后,得到相应数据权重值,并且在权重值的基础上加上一个全局偏置向量  $b$ ,同时将这部分偏置向量进行求和运算后通过激活函数  $f$  加以数据预处理,由此得到的神经网络模型的运算结果  $y$ 。这一个神经网络模型再处理信息内容的主要方式可以通过公式 2-1 进行描述:

$$y = f\left(\sum_{i=1}^n w_i x_i - \theta\right) \quad (2-1)$$

单个的神经元结构过于简单,只能解决某个简单的问题。通过将多个独立的神经元结构融为一体,构建出一种复杂的神经网络,从而能够处理更多的挑战,并且能够构建出一种完整的、高效的模式,就像图 2-3 中的那样,它由输入、隐藏和输出三个阶段构成。将一些重复的信息从输入层传递到输出层,

而在接下来的时间里，我们可以利用隐藏层来实现有效的信息收集，从而减少信息的复杂性，提升信息的质量<sup>[13]</sup>。所以，多层次的神经网络可以解决更加复杂的问题，也能够表述更加全面的特征内容，具有更加强的分析能力。

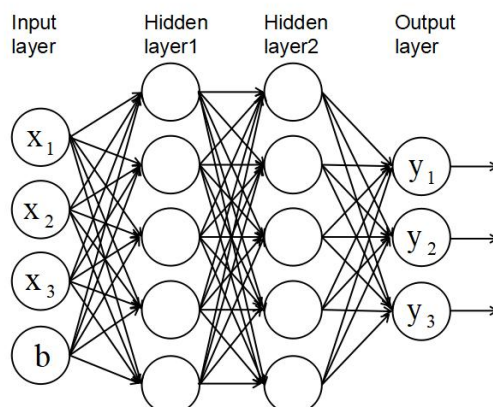


图 2-3 多隐层神经网络模型

但是针对单一神经网络结构是非常简单的，只能够以单一的问题作为目标导向。通过把众多的神经元组织结构加以链接，可以建立起较为复杂的神经网络模型结构，由此可以处理较为复杂的问题。如图 2-3 所示，神经网络模型主要包括以下三个部分结构：输入层、隐藏层和输出层。输入层的主要作用是将某些特性信息内容进行输入，而不是直接对它进行逻辑处理。隐藏层可以根据当前的任务处理的需求方式，从而确定出要隐藏的数据信息内容的总量，所以对于多隐藏层的神经网络结构相比于单一的神元网络层结构，功能和结构上更加全面和复杂，同时可以通过对所提供的特性信息进行全方位实现对隐式特性信息的挖掘，从而获取了更强的模型学习能力。

### 2.1.2 卷积神经网络 (CNN)

CNN，也被称为卷积神经网络，是一种基于局部链接、数据权重共享以及多卷积层的深度学习模型。它的诞生源于 1962 年<sup>[14]</sup>，当时 Sub-I 们发现猫的视觉皮质细胞能够捕捉外界信息，并将其转化为可以用于解决复杂问题的机器学习算法。野机制的工作原理是通过对外界刺激的特殊处理，使神经元能够接受其所控制的范围内的信号，从而形成一种独特的反应。在视觉神经网络系统中，神经元的信息单元受到外界刺激的影响，并通过视网膜上的感受器将信号传输到大脑皮层，从而形成一系列的反应，最终导致人们的行为。在味觉神经网络模型中，与视觉神经网络类似，接收到的信号主要来源于舌头上的味觉神经元的刺激，只有在接触到味觉神经单元，才能够激发该神经元。

卷积神经网络是一种复杂的机器学习模型<sup>[15]</sup>，它由多个层次组成，其中最基本的是卷积层、池化层、全连接层，它们之间的关系是一种复杂的、有机的、有序的、有效的、能够实现多种功能的机器学习模型。与传统的前馈神经网络模型不同的是，卷积层之间的计算比神经网络拥有更少的参数。模型通过利用卷积层和池化层获取有效的信息单元，同时可以自动获取数据分析中的特征向量，可以自动获取数据的一些特征向量，减少计算的复杂程度，提高处理的精确程度。

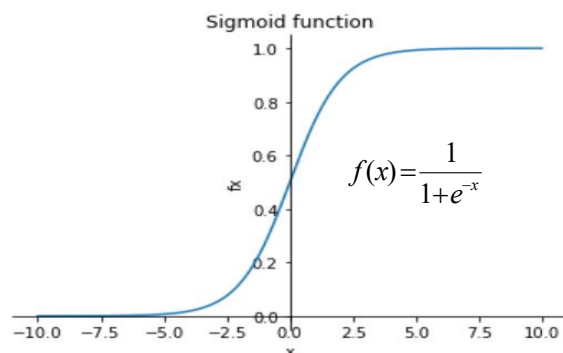


图 2-4 sigmoid 函数

卷积神经网络技术可以有效地解决复杂的图像识别、目标定位、图像分割以及其他相关任务，其中，卷积层的应用尤为重要，可以有效地提高识别准确率，从而实现更加精确的结果。卷积神经网络长处是在目标参数数量这方面，其架构的参数数量少，具备易计算、易存储、易处理图像的特点。卷积神经网络由卷积层、最大集合层、丢弃层以及一系列完整的感知器组成，它们共同构成了一个复杂的网络架构。

## 2.2 手写汉字识别技术分析

### 2.2.1 传统手写汉字识别技术

多年来，随着技术的发展，传统的手写汉字识别技术已经发展出一套完善的工作原则，其中包含多个关键环节<sup>[16]</sup>，比如：图像预处理、特征抽取、特征压缩和分类器的构建。其中，以梯度特征的应用最具代表性，可参照图 2-5。

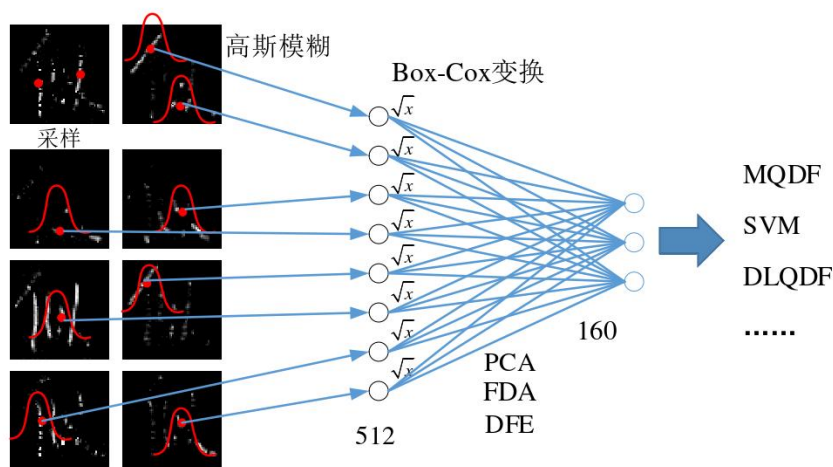


图 2-5 传统手写汉字识别方法流程图

为了更好地识别汉字，我们需要对其进行预处理。这些步骤通常包括灰色校正、尺码校正等。接下来，我们需要从汉字的梯度信息集中抽取重要的信息，例如图 2-5。我们可以通过分析这些信息集，找出最能体现汉字的关键部分，并利用 Box-Cox 算法将其转换为更高斯的模糊信息。通过主成分分析（Principal Component Analysis, PCA）、线性判别分析和鉴别特征提取，可以大大减少数据的复杂程度，从而极大地改善计算效率。在确定了需要使用哪种分类算法之后，我们可以考虑使用 MQDF、QLQDF 和 SVM 这些传统的算法。

### 2.2.2 基于卷积神经网络的手写汉字识别技术

通过采用卷积神经网络，可以有效地识别手写汉字，其结构包括：输入、卷积、池化、全连接以及输出。为了实现这一目的，可以通过收集具有特定标记的汉字样本，并经过精心的训练，以获得最佳的识别结果，从而实现手写汉字的准确识别。

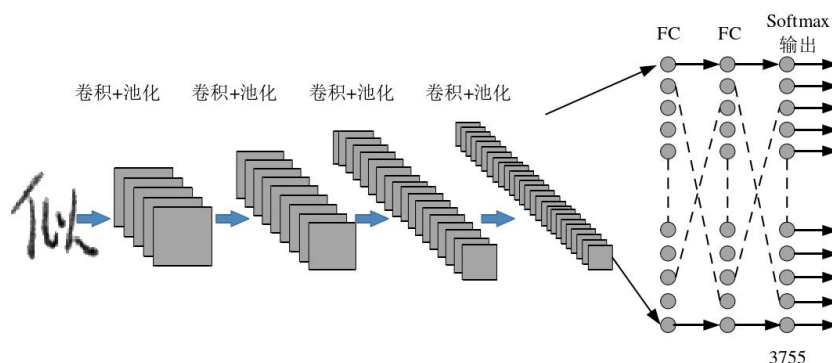


图 2-6 基于卷积神经网络的手写汉字识别示意图

通过卷积神经网络，可以有效地识别出手写的汉字，而无需人工干涉，因此，该网络可以将汉字的原始信息转换成可以被机器识别的数据，并且可以根据机器的识别结果，对这些数据进行有效的分类，而无需人工干涉。该网络的运算流程由两个步骤组成：模型训练、预测。模型的前向传播意味着从一个输入信号中获取更多信息，并利用这些信息来构建一个更优秀的预测系统。例如，当一个汉字图像被投递给卷积神经网络时，它会被卷缩并从中提炼出一些关键信息，并被转换为一个更加精确的预测结果。这些预测结果会被转换为一个更加优秀的预测系统，并被 Softmax 等软件工具识别。

在使用卷积神经网络进行手写汉字识别的情况下，首先，我们必须对其进行一层一层的前向计算，以便将数据一层一层地传递至输出层，从而得出正确的预测结果。为了实现这一目标，我们必须精心设计一个合理的卷积神经网络架构。在构建复杂的模型时，许多关键的超参量都必须考虑，比如：全连络层的神经元种类、各层的卷积核的多少，以及激活的函数的类型，这些都必须通过反复的测试来得出。

### 第三章 基于注意力机制的深度可分离网络手写汉字识别

虽然《中华字海》和北京国安咨询设备公司的汉字字库都记载了 85568 个和 91251 个的汉字，但是由于这些记载的汉字的复杂性，很难准确地将它们分开，因此，在这种情况下，准确地分辨每一个汉字显得尤其重要。根据 3000 个汉字的数量，我们可以轻松地辨认出 99% 的汉字<sup>[17]</sup>。因此，我们将重点研究 GB2312-80 标准所提供的 3755 个汉字，以便更好地实现手写汉字的辨认。在这篇文章里，我们将介绍手写汉字辨认的基础步骤和预处理方法，并通过卷积神经网络的方法，提高汉字辨认的效率。具体而言，我们将采用汉字的八方向梯度特征，并结合深度可分离卷积方法，来提高汉字辨认的精确性。通过多组卷积神经网络的应用，我们可以准确地辨认出手写的汉字，并且可以将其与其他文本区分。

#### 3.1 手写汉字识别基本流程

##### 3.1.1 手写汉字基本流程

根据图 3-1，手写汉字的识别方法由三个步骤组成：首先是对原始汉字的图像进行预处理，然后是八方向的梯度特征的抽取，最后是使用卷积神经网络进行模型的构建。该模型具有强大的特征抽取与模型划分的能力，从而实现对汉字的准确识别。鉴于梯度特征可以显著地优化传统的手写汉字的分析，因此，我们将其应用于卷积神经网络，以获得更准确地分析结果。然而，由于卷积过程需要将多个信息卷积起来，存在较大的冗余，从而降低了分析的准确率，因此，我们采用深度可分离卷积技术来解决这一问题，以达到更优的分析结果。

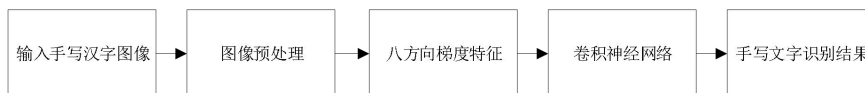


图 3-1 基于卷积神经网络的手写汉字识别流程

##### 3.1.2 手写汉字图像预处理

在实际应用中，由于采集的原始汉字图像容易受到多种外界因素的影响，从而无法直接用于识别任务。为了解决这个问题，我们采取了图像预处理技术，其中包括对比度增强和尺寸归一化，从而有效地抑制外界因素的影响，从而提高识别准确性。

###### 1. 对比度增强

通过使用灰度变换<sup>[18]</sup>，我们能够提高图像的对比度 $[0,255]$ ，从而避免由于光线和采样器的原因造成的细节模糊和明暗度差的情况。通过使用灰度变换，我们能够更好地理解并捕捉图片的细节，从而更好地处理和分析数据。

$$D(x,y) = \frac{(I(x,y) - I_{\min}) \times 255}{I_{\max} - I_{\min}} \quad (3-1)$$

其中  $I_{\max}$  和  $I_{\min}$  为原图像的最大、最小灰度像素值， $I(x,y)$  为原图像像素值， $D(x,y)$  为目标图像像素值。

###### 2. 尺寸归一化：

为满足卷积神经网络的输入要求，一般需要调整网络输入图片的尺寸为固定尺寸。在经典的 MNIST 手写数字识别中，输入图像通常被归一化为 28x28 大小，已能够取得良好的识别效果。输入图片尺寸过大会增加网络的训练负担，过小则影响网络的识别性能，考虑到汉字图像比数字更加复杂，

本文将汉字图像归一化为  $56 \times 56$  大小，并且在图像四周各添加四个空白像素，使汉字内容居于图像正中，最终的输入图片尺寸为  $64 \times 64$ 。图像尺寸归一化方法很多，本文采用简单的线性归一化方法，使用最近邻插值策略，坐标映射公式为：

$$\begin{cases} x' = \frac{w_2}{w_1}x \\ y' = \frac{h_2}{h_1}y \end{cases} \quad (3-2)$$

其中， $w_1$ 和 $h_1$  为原图像的宽度、高度， $w_2$ 和 $h_2$  为目标图像的宽度、高度， $(x,y)$  为原图像坐标， $(x',y')$  为目标图像坐标。

随机挑选 CASIA-HWDB1.1 数据集中一张汉字图片“似”，对比度增强及尺度归一化效果如图 3-2 所示，图片边缘黑框起辅助显示作用。

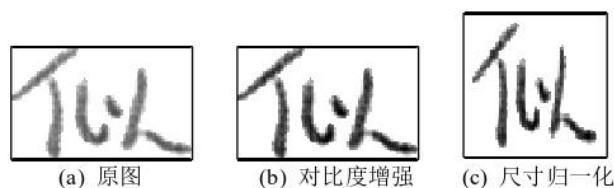


图 3-2 图像预处理前后效果对比图

## 3.2 手写汉字识别卷积神经网络结构设计

自 1998 年 LeNet-5 卷积神经网络问世以来，它的结构发展出了多种多样的形式，例如 AlexNet、VGGNet、GoogLeNet 和 ResNet，它们在不同的应用场景中表现出色，但是，由于它们的结构限制，它们并不适合用于手写汉字的识别。卷积神经网络是一种复杂的机器学习技术，它可以模拟人类的思维和行为，并且可以根据特定的应用场景和需求，自动调整网络的复杂度，以达到最佳的识别效果。本文将探讨如何利用卷积神经网络的输入、方式和复杂度，来实现对手写汉字的准确识别。

### 3.2.1 深度可分离卷积

在卷积神经网络中，卷积技术发挥着极其重要的作用。它拥有出色的特征抽取性<sup>[19]</sup>，而且与传统的全连接方法相比，它可以节省更多的参数。然而，在传统的卷积方法中，由于各种因素的存在，使得各种信息之间的耦合程度较高，从而导致网络的复杂性和模型的准确性受到一定的限制。这篇文章采用了一种新的方法，即利用深度可分割的方法，把数据集划分为两部分。首先，对于单独的数据集进行处理，然后利用  $1 \times 1$  的方法对多条数据进行处理，并把它们组合到一起，从而得到一张有意义的特征表。这种方法能够有效地降低数据集中的相似程度，从而提升模型的分类能力。

如果我们要求输入一张  $64 \times 64 \times 3$  的图像，其中  $64 \times 64$  表示图像尺寸，3 表示图像通道数，那么我们可以采用两种卷积方法来实现。首先，我们可以使用 256 个  $3 \times 3$  的卷积核来分层，然后，我们可以使用 256 个  $1 \times 1 \times 3$  的卷积核来加权，最终得到一张  $64 \times 64 \times 256$  的大小的图片。图 3-3 展示了深度卷积与传统卷积的显著差异。



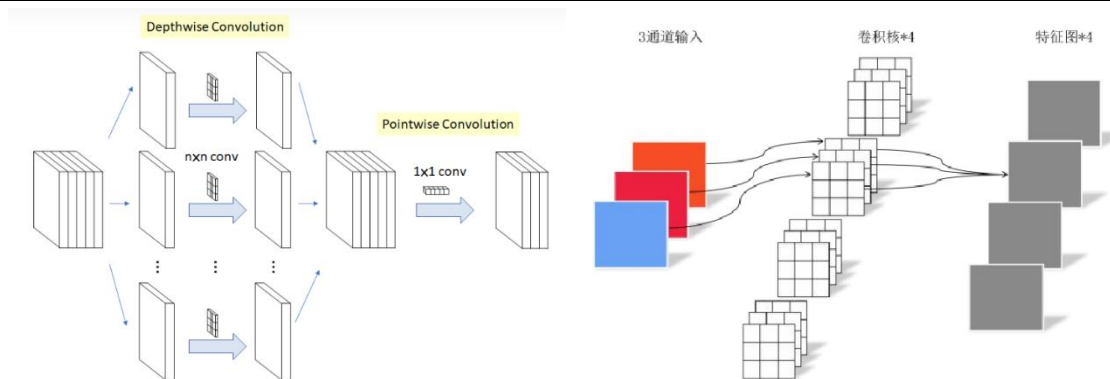


图 3-3 深度可分离卷积网络结构图和常规卷积网络

经过分层卷积处理后，我们可以尝试使用  $1 \times 1$  维卷积来实现跨通道信息融合，这种方法可以将不同通道的图像放置在同一空间中，从而更好地利用它们的信息。此外，这种方法还可以提高图像的维度，使得它更加精确。

### 3.2.2 深度可分离卷积网络结构设计

深度卷积技术可以分为分离卷积和单独卷积，前者包括沿通道和单独的卷积，而后者则是沿着单独的通道和单独的卷积进行的。

(1) 通过逐通道卷积（DW），我们可以将一张  $32 \times 32 \times 3$  的图片转换为三个特征图像，这一过程需要使用一个滤波器，将每个通道的特征图像进行卷积，以确保特征图像的数量与输入通道的数量一致。具体的转换步骤可以参考图 3-4。

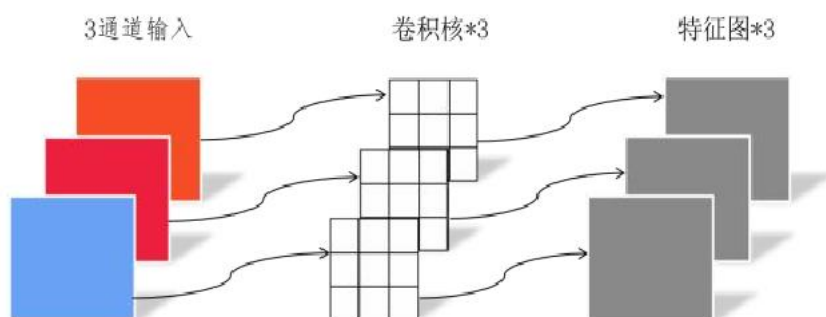


图 3-4 逐通道卷积过程

从图 3-4 可以看出，逐通道卷积操作只能获取输入图像的一部分，无法对其进行拓展，而且 DW 只能针对每一层的深度进行卷积，因此无法充分利用不同层次的图像数据。所以要使用逐点卷积对逐通道卷积操作得到的特征图片进行重新整合。

(2) 通过逐点卷积（PW），可以有效地将多个特征图像进行深度分析<sup>[20]</sup>，从而构建出一个更加精确的特征图。PW 的滤波器尺寸为  $1 \times 1 \times M$ ，其中  $M$  表示前一层的通道数量，通过这种方式，可以有效地提取出更多的特征信息，从而构建出更加精确的特征图。通过调整卷积核的数量，我们可以获得与输出特征图数量相等的结果，如图 3-5 所示。

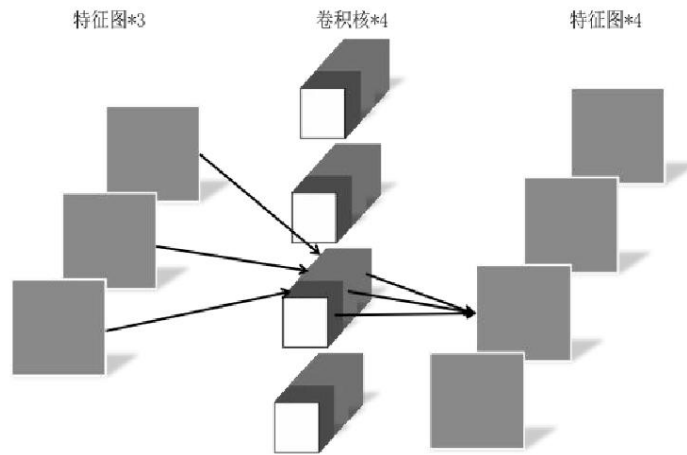


图 3-5 逐点卷积过程

### 3.2.3 注意力机制

在不同的时间节点，时间序列的多维特征会对预测值产生不同的影响。本研究把注意力机制和当前的研究进行结合来获取特征的权重分布。注意力机制在初始的阶段应用在了语言处理领域，比如机器翻译等。在注意力机制当中，包括关键信息和其他信息这两个不同的部分。主要是处理一些形容词、名词，对于图片层次来说，例如在处理图像的时候，不会像电脑一样逐步地分析每一个不同的像素，而是提取最关键的部分。比如，如果想分辨出这幅图中是什么动物，会根据动物的具体面部特征进行识别，进而再补充一些其他的信息，比如背景，还有动物的神态等等。所以在处理图像的时候，将信息分成了关键信息以及其他信息两个不同的部分。

需要注意的是，不管是分析文本还是处理图像的时候，都会存在一定的分析顺序。对于文本信息来说，首先是需要按照规定的语法对其进行解读，保证分析结果的准确性。处理图像时，RNN 技术可以帮助我们更好地理解场景的先后顺序，从而更有效地将图像分类。它可以将图像按照整体到局部，或者局部到整体的顺序，与人脑的思维方式相似，并且通过将注意力机制与机器学习算法相结合，使得机器学习算法与人脑的处理方式更加接近。

SE 网络利用其独创的通道注意力机制，可以根据多个特征通道的相似度，以及它们之间的联系，进行自动化的重复标记，从而增强其学习效率。这种方法利用 Squeeze 和 Excitation 技术，可以根据多个特征的重复标记，从而更好地识别出每个特征的价值。这种神经网络使用了一种神经机器学习的技术，它能够识别出当前任务的关键特征，根据这些关键特征的强弱，选择最优的进行处理，而忽略那些次优的。SE 模型的架构可以参考图 3-6。

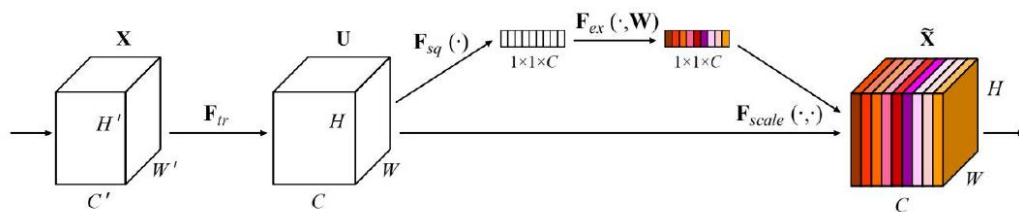


图 3-6 通道注意力机制

Squeeze 是一种用于压缩输入特征图的方法，它可以将特征图中的每个元素都转换成一个实数，从而使得整个特征图的分布更加完整。这种方法的基础是全局平均池化，可以有效地提高模型的准确性和



可靠性。Excitation 操作是一种基于原循环神经网络的门机制，它可以根据参数  $w$  来确定特征图中每一个特征通道的权重，从而更好地描述它们之间的关联性。

Excitation 是一种用于识别和分析数据集上各个元素间相互影响的方法。它使用两个完整的模型，分别用于描述数据集上各个元素的相似性和独立性。它可以帮助我们更好地理解数据集上的元素，从而更准确地预测数据集上的元素间的相似性。它使用两个模型，分别用于描述数据集上的元素，以及它们对数据集的影响。Reweight 是一种用于重建和改进输入特征的方式。它使用 Excitation 算子来计算出各个特征之间的相似性，然后使用乘积算子来计算出各个特征之间的相似性，从而使整张图片的各个方面都能够被精确地标记。

当我们将注意力集中到一组向量上时<sup>[21]</sup>，编码器就会产生一系列的序列，而根据不同的权重，解码器则会对这些序列进行精确的计算，从而实现信息的传输。具体的计算公式如下所示：

译码器隐含值与解码器单元隐含值计算出注意力权重  $\alpha_t^i$

$$\alpha_t^j = \frac{\exp(\text{score}(s_t, h_i))}{\sum_{j=1}^n \exp(\text{score}(s_t, h_j))} \quad (3-3)$$

在上述公式中，函数将编码器隐含的数值和解码器隐含的数值作比较，进行统一处理，具体计算过程如下所示：

$$\text{score\_addition}(s_t, h_i) = V^T \tan h(W_a [s_t; h_i]) \quad (3-4)$$

$$\text{score\_multiplication}(s_t, h_i) = s_t^T W_a h_i \quad (3-5)$$

由此可计算到注意  $i$  权重  $a$ ，再通过加和计算能够得到上下文语义向量  $c_t$

$$c_t = \sum \alpha_i^t \times h_i \quad (3-6)$$

译码器此时刻的隐含值  $h_j$  又向量  $c_t$  的最大注意力权重为：

$$a_t = f(c_t, h_i) = \tan h(W_c [c_t; h_i]) \quad (3-7)$$

下一个计算单元以输入变量来接受相应的注意力向量为：

$$y_t = f(h_t, y_t, c_t) \quad (3-8)$$

在公式中， $v_t$  为权重矩阵<sup>[22]</sup>， $\alpha_t^i$  译码器的第  $i$  个单元对应的解码器第  $t$  个单元的值， $c_t$  不同的单元经过编译器处理之后输出的向量。注意力机制，详细的说就是不同的输出目标，经过函数处理之后进行输出。在上述计算方式中，不仅仅包含了译码器全部的信息，而且还包含隐含单元的详细信息，通过训练之后，可以从模型中看出输入序列不同部分的重要程度。

### 3.3 模型训练及结果分析

#### 3.3.1 实验环境

对于本研究来说，是采用 Tensorflow 编写的数据库为基础<sup>[23]</sup>，Tensorflow 有很大的学习能力和自动求导能力，可以快速构建深度神经网络。利用 Python 语言的 numpy 和 pandas 包来对实验的数据进行分类，具体的配置如表 3-1 所示。在构建模型之前，按照三比一比一的比例对时间进行设置，时间序列数据长度设置为 24，训练周期设置为 500，批设置为 256，在训练过程中加入 Early Stop，当模型误差在超过一百个周期后没有改善就停止实验。采用 Adam 优化器，其中超参数设置为  $\alpha = 0.9, \beta =$

0.999,  $\epsilon = 1e-8$ , 学习率为  $1e-4$ 。学习率不管是过大或者是过小都会影响最后的实验结果。这个优化器可以为每个参数提供合适的学习率, 从而保证模型输出最优解。

表 3-1 实验硬件和软件环境表

| 配置环境 | 版本  |
|------|---|
| CPU  | Inter(R) Xeon(R) CPU E5-2620 v4 @ 2.10GHz     |
| 内存   | 32G   |
| GPU  | NVIDIA TITAN Xp@1.6 GHz 12GB                  |
| 存储   | 16TB HDD, 以 RAID5 模式运行                        |
| 操作系统 | Ubuntu 16.04.10 64 位                          |
| 编程语言 | Python 3.9 64 位                               |
| 软件包  | Tensorflow2.3.0, numpy 1.15.4 , pandas 0.22.0 |

### 3.3.2 数据集

研究中使用的 CASIA-HWDB1.1 是中科院自动化研究所的一个重要工具, 它收录了多种不同类型的文档, 其中包含了 CASIA、MNIST、HCL200 以及 CASIA-HWDB 等多个版本<sup>[24]</sup>。我们的研究采用的 CASIA-HWDB1.1 文档收录了来自 300 个不同作者的独立文献。这个文档收录的文献共计 171 个英文数学符号, 以及 3755 个普通文言文。通过对 64 个不同尺寸、64 个像素的脱机手写汉字的数据进行处理, 我们可以从中抽取出 512 维的特征向量, 并且可以从图 3-7 中看到一些类似的汉字的实际情况。

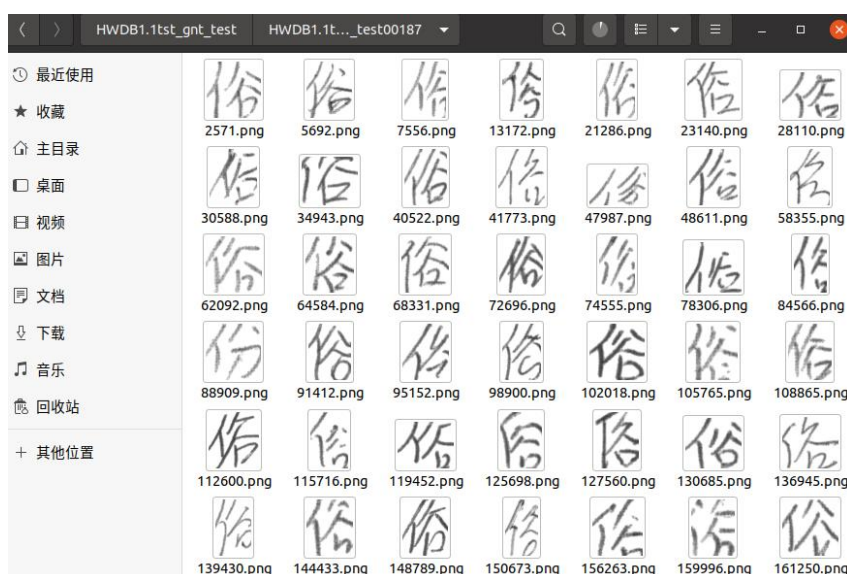


图 3-7 CASIA-HWDB1.1 数据集

### 3.3.3 实验结果及分析

根据表 3-2 的数据, 我们发现一级 3755 类手写汉字的情况。

随着技术的发展, 7 层的 LeNet-5 和 16 层的 VGGNet、22 层的 GoogLeNet 等模拟技术的出现, 使得模拟的复杂程度大大提高, 这也就意味着, 在提高模拟的准确率和可靠性方面, 深度神经网络的表现明显比浅层的更胜一筹。

根据表 3-2 的数据, 尽管采用的是原始的八方向梯度特征图像, 随着模型的深度和宽度的不断变化, Net5 的分类能力也会有显著的改善。在八方向梯度特征图像的情况下, Net5 的识别精度达到 94.95%,

而在 5 个卷积神经网络模型中, Net2 的精度比 Net1 高出 1.89%, Net3 的精度比 Net2 高出 0.97%, Net4 的精度比 Net3 高出 0.52%, Net5 的精度比 Net4 高出 0.29%, 但随着卷积神经网络的复杂化, 其精度的改善也越来越明显, 最终接近饱和点。而从表 3-3 可以看出, 深层卷积的精度也在不断改善, 但其所需的训练成本也在不断攀升。经过实验证明, 无论是传统的卷积还是深度可分离卷积, 其每一次的模型训练所需的时间都非常接近, 八方向的梯度信息的输出比起原始数据更快。

表 3-2 常规卷积时不同手写汉字识别准确率结果

| 网络模型 | 原图输入   | 八方向梯度特征图像输入 |
|------|--------|-------------|
| Net1 | 90.12% | 91.28%      |
| Net2 | 92.05% | 93.17%      |
| Net3 | 93.38% | 94.14%      |
| Net4 | 93.77% | 94.66%      |
| Net5 | 94.16% | 94.95%      |

表 3-3 网络训练迭代周期及训练时间统计

| 网络模型   | Net1 | Net2 | Net3 | Net4 | Net5  |
|--------|------|------|------|------|-------|
| 训练时间/秒 | 20   | 23   | 24   | 29   | 39    |
| 迭代周期/次 | 30   | 30   | 30   | 30   | 30    |
| 总时间/小时 | 3.98 | 4.78 | 5.67 | 9.97 | 16.38 |

从 Net1-Net4 四个模型来看, 它们在每个阶段的训练速度几乎没有区别, 在每个阶段只需 30 次迭代即能达到预定目标。然而, 在整个网络中, Net5 模型表现得尤为优秀, 它在每个阶段的训练速度和迭代次数都比 Net1 高了 3.3 倍, 这主要归功于它比卷积神经网络更容易实现。随着 Net5 的发展, 对于模型的复杂性的提升, 使得训练变得越来越艰巨。为了达到最佳的结果, 我们必须考虑到模拟的成本和汉字的识别精度, 而这两者必须得到有效的协调。

根据图 3-8 的数据, 可以看出, 当采用深度可分离卷积方式进行手写汉字识别时, 其准确率明显优于采用常规卷积方式, 尤其是在不同复杂度的卷积神经网络中, 前者的准确率更高, 而后者的准确率则更低。在原始图像中, 深度可分离卷积的准确率比传统卷积的准确率高出 0.42%, 而在八方向梯度输入的情况下, 这种差异更加明显, 其准确率也有所提高。

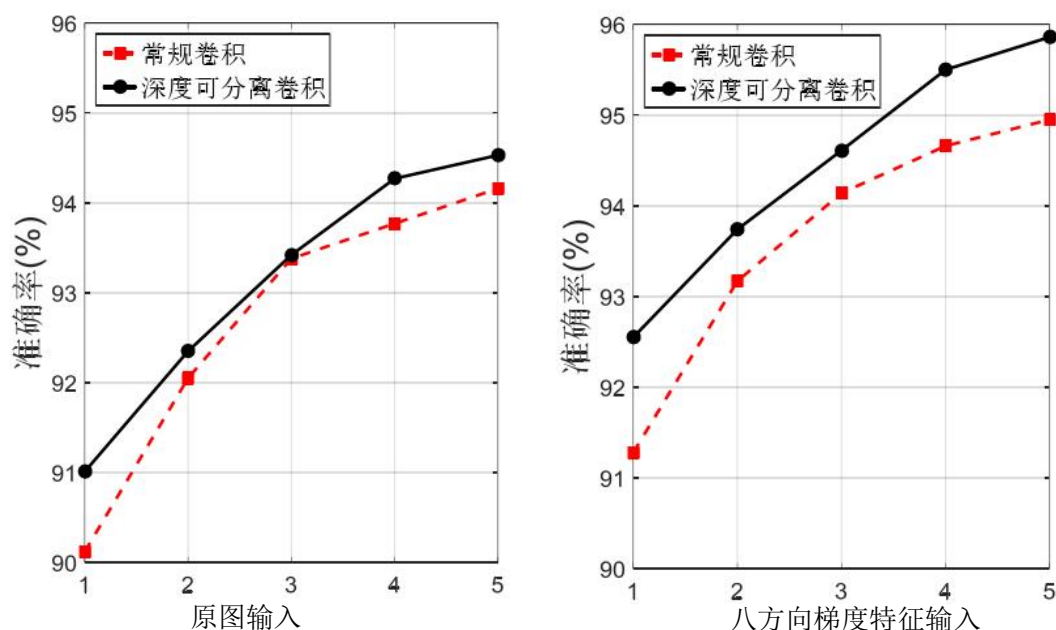


图 3-8 不同卷积方式时手写汉字识别准确率对比图

从这里可以清楚地发现,卷积技术能够显著降低图像中各个通道的相互影响,从而提高识别精确率。而且,深度卷积技术不仅能够降低网络的复杂程度,而且能够大大缩短卷积层的参数数,从而降低模拟的存储空间。

通过对 EfficientNet 网络结构的比较,我们发现,当它被拓展到 CASIA-HWDB1.1 手写汉字库时,其性能有了显著提升,具体情况可参见表 3-4。

表 3-4 基于注意力机制的深度可分类网络与改进前识别对比结果

| 网络模型                    | 识别结果   |
|-------------------------|--------|
| CNN                     | 0.7763 |
| FasterR-CNN             | 0.7931 |
| LSTM                    | 0.8345 |
| Efficien                | 0.8973 |
| Efficient-Net           | 0.9132 |
| Efficient-Net+Attention | 0.9652 |

根据表格数据, EfficientNet 网络的识别准确率达到 91.32%, 而将其中的 MBConv 结构替换为基于洗牌算法的 shuffler 网络, 其识别准确率更是达到了 95.013%; 此外, 在 shuffle 网络基础上, 再加入注意力模型, 提出的 Efficient-Net+Attention 网络, 其识别准确率也达到了 96.652%, 这些都证明了 EfficientNet 的有效性和可靠性。与传统的 EfficientNet 网络结构相比, 新一代网络识别算法的识别准确率显著提升。

在这篇论述中, 我们通过使用洗牌算法来改善 EfficientNet 的 MBConv 模型, 从而解决了它在高精度方面的不足。我们使用的 shuffler 模型可以有效地改善这种模型, 从而大幅度提高识别算法的精度。经过重新设计, EfficientNet 的 SE 模块能够更好地捕捉和处理多个数据集之间的关系, 这样就能够更加

精细地分析出每个数据集的内容，并且能够更好地处理多个数据集之间的关系。经过这种方法，我们发现，EfficientNet 的汉字识别能力得到显著的提高。

## 第四章 手写汉字质量评价

### 4.1 手写汉字质量评价组构建

在评估准备阶段，我们需要将需要比较的集合缩小，因为如果不缩小，每次只需要对一个字进行评估，就会导致整个数据集中的所有字都需要进行相似度匹配，这会带来巨大的工作量，而且效率也会降低。评价组是一种经过筛选的基础集合，它由两部分组成：一部分是通过对上一步手写汉字的检测和识别的结果，以确定自然环境中的待测汉字的位置，并预测出相应的评价字图，这一部分叫做预测集合；另一部分则是用来进行评估的候选集合。

尽管识别准确率不能保证百分百正确，尤其是当涉及到近似字的时候，例如“己”、“已”、“巳”，但是为了减少识别准确率对评价的不利影响，我们可以采取扩展集合的方式，将评价标准候选集合增加一部分，以便更好地识别出错误的字，从而避免漏选的情况发生。

### 4.2 手写汉字质量评价标准

这个汉字质量评价标准字库由书法家庞中华创建，它包含了 3323 个一级汉字，并以 png 格式呈现。每个字都有一个独特的编号，如图 4-1 所示。

|   |   |   |   |   |
|---|---|---|---|---|
| 上 | 韦 | 中 | 书 | 半 |
| 北 | 卞 | 旧 | 归 | 且 |
| 由 | 申 | 甲 | 曲 | 串 |
| 木 | 术 | 札 | 杈 | 极 |
| 杯 | 林 | 板 | 杞 | 枝 |
| 栈 | 相 | 桓 | 检 | 楷 |
| 忆 | 忙 | 忱 | 恍 | 忧 |
| 怀 | 忡 | 忤 | 快 | 怙 |
| 怖 | 怜 | 怕 | 恼 | 悦 |

图 4-1 手写汉字初值评价标准字库样图

确定一个准确的字帖评价标准是一个棘手的问题，本文采用的是书法初级考试标准，它的划分更加明确、合理，重点关注笔画、结构和章法，因此，我们在评估字帖时，会从这三个方面入手，即笔画形态、整体结构、汉字排列等，并通过比较这三个方面的相似度，最终得出字帖的质量评价结果，这个结果将由三个方面的相似度共同加权计算得出。根据下面的公式，可以计算出相似度。

$$S = W_1 S_1 + W_2 S_2 + W_3 S_3 \quad (4-1)$$

其中， $S_1$  代表笔画形体相似度， $S_2$  代表整体结构相似度， $S_3$  代表汉字排布相似度， $W_1, W_2, W_3$  表示其对应权重值。

### 4.3 基于相似度匹配的手写汉字质量评价方法

在本文中，我们提到的两种用来衡量手写汉字质量的方法：一种是一对多评价方式，另一种则是基于标准库的评价方式。前者使用的是评估集，使用余弦相似性来计算骨骼和边缘的优劣。需要指的是，这个评估不仅仅局限在单个汉字的比较上。第二种评估方案采用的是一种新的技术，它使用一种专门的软件来衡量一种汉字的优劣。该软件使用的文本来源于一系列的硬笔书法经典字帖，并且能够根据文本的整体结构和笔画的相似性来衡量文本的优劣。使用两种不同的评估方案可以更精确地衡量文本的优劣。

第一种方法的优势在于，可以根据多种可用的信息，不局限于字帖的标准图像，待评价样本与评价集中该字图像进行相似性对比，从多个角度，结合多种可用的信息，以此作为一种个性化的书写评估，从而提供一个全面的、客观的、可靠的结论。然而，第一种方法的弊端在于，当可用的信息较少时，待评价样本质量指标的准确性可能受到一些影响，因此，我们需要对其进行一次模糊综合评估，以检验采用评价集得出的评价结果是否准确。若该评估结果并非准确，则需要使用第二种方法来对其进一步地检验。具体流程如图 4-2 所示：

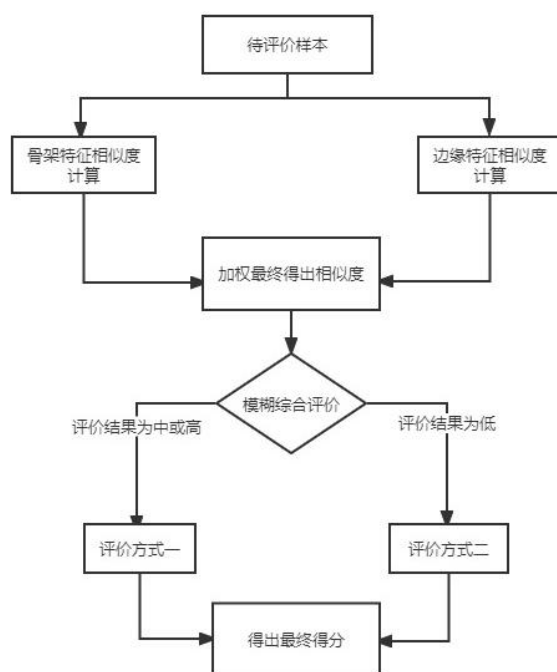


图 4-2 评价流程图

#### 4.3.1 模糊综合评价

采用模糊综合评价方法，可以根据不同的评价要求，从而实现从单一的评价角度到复杂的评价角度的全面评估，从而更好地实现定性评价、定量评价的综合运用。

有关模糊综合评价步骤如下：

**Step1:** 通过多个可以反映待评估对象特征的指标，我们构建了一个多维度的评估因子集  $U$ 。该因子由边缘特征和骨架特征的相似度两部分组成。

**Step2:** 通过构建评价等级集  $V$ ，我们可以将待评价对象的各种可能结果综合起来，形成一个完整的等级集，其中包括高、中、低三个等级，以此来衡量其可靠性。

**Step3:** 确定评判隶属矩阵  $R$ ,  $R$  表示以及评价矩阵，这个矩阵是由多个单个评价因素向量

$R_i(i=1,2,3,\dots,m)$  构成, 如公式 (4-2) 所示, 其中  $R_i$  表示第  $i$  个评价因素  $u$  的评价集合, 是评价等级集  $V$  模糊子集。

Step4: 综合评价模型为  $B = A \circ R$ , 进行模糊矩阵的合成运算, 这里的运算 采用的是主因素决定型, 公式如下所示:

$$b_j = (a_i \wedge r_{ij}) \quad (4-2)$$

Step5: 计算最终评价结果  $Z = B \times F$ , 其中  $F$  表示数据集  $(f_1, f_2, f_3, \dots, f_n)$ , 其中,  $f_n$  表示第  $n$  级评价的内容, 对得到的评价结果集合进行归一化处理, 再通过最大隶属度原则选出集合中的最大的结果。

#### 4.3.2 一对多评价方式

本文采用两种评价方式是为了使每个待评价样本都能够得到可靠的质量评价结果, 提高该方法的通用性。汉字质量评价由于汉字书写本身带有的个性化风格, 使得对它的评价很难形成一个统一的标准。传统的评价方式是通过特征提取再与标准字帖进行对比从而得出结果, 这就是待评价样本与相对应标准字帖进行一次对比的过程, 这种方式存在一定的局限性。而本文所使用的一对多评价方式, 它不是与标准字帖做对比, 而是与所建立的评价集合进行相似匹配, 从而预测出更加准确的评价结果。

一对多评价方式是在上一步模糊综合评价结果在高或中时才会进行结果计算的评价方式, 而模糊综合评价就像是该方式计算前的一个确认步骤, 如果模糊综合评价结果为低则会判定第一种评价方式得出的结果不够准确, 而采用基于标准库, 也就是标准字帖的一对一评价方式对待评价样本进行评价。一对多评价通过待评价样本与评价集合使用余弦相似性来计算骨骼和边缘的优劣, 加权得出最终的相似度, 最后得出质量评价结果。

##### 1. 骨架特征提取

在书写汉字质量评价中, 骨架特征是汉字特征的最重要的组成部分, 汉字的个性化可以说就是由手写汉字的骨架特征和轮廓来体现的。提取汉字骨架就是对汉字进行细化。细化是将图像的线条从多像素宽度减少到单位像素宽度过程的简称, 通常将细化结果描述为“骨架化”、“骨架提取”, 如图 4-3 所示:

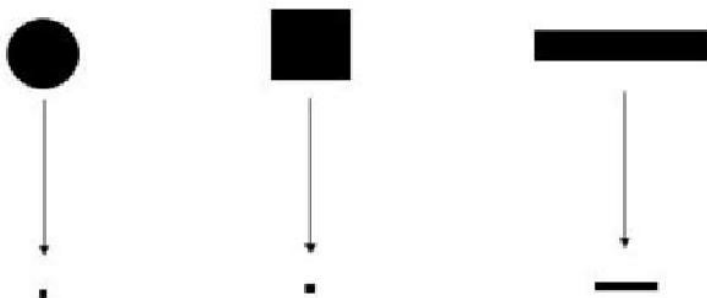


图 4-3 简单图形的骨架示意图

由上图可以看出细化后的汉字仍然保留了基本的骨架特征, 并不会影响质量评价结果, 同时又使得计算机更加容易辨别和处理, 存储量较原文字的二值化点阵大为下降, 减小了后继处理的工作量。为了能够得到较好的汉字的形状特征和笔画之间的拓扑结构信息, 我们需要选用一种合适的细化算法。目前, 迭代算法、跟踪算法和极值算法都是比较流行的图像细化算法, 本文采用的是迭代算法中的索引表细化算法。基于索引表的细化算法是根据每个目标像素点的八邻域的像素值来决定是否将目标像素点置为背



景点，八邻域像素值的各种情况事先根据一定的规则列为一索引表。如图 4-4 所示，像素点  $P_1$  是否被置为背景点取决于它的八邻域  $P_2, P_3, P_4, \dots, P_8, P_9$  的像素值，可表示为：

$$P_{1_{final}} = \begin{cases} 0, & (P_1 = 0) \\ f(P_2, P_3, P_4, \dots, P_8, P_9), & (P_1 = 1) \end{cases} \quad (4-3)$$

|       |       |       |
|-------|-------|-------|
| $P_9$ | $P_2$ | $P_3$ |
| $P_8$ | $P_1$ | $P_4$ |
| $P_7$ | $P_6$ | $P_5$ |

图 4-4 点  $P_1$  的八邻域示意图

其中  $f(P_2, P_3, P_4, \dots, P_8, P_9)$  表示索引表的值，每一组可能的  $P_2, P_3, P_4, \dots, P_8, P_9$  像素值会对应一个  $P_{1_{final}}$  的值，共有  $2^8=256$  项。如图 4-5 所示为八邻域权值分布，在细化过程中，每一个  $3 \times 3$  的模板与其相应的权值相乘求和得到索引值，根据得到的索引值到索引表中相应的位置找到对应的数据并判定是否将其置为背景点。

|    |       |     |
|----|-------|-----|
| 1  | 2     | 4   |
| 8  | $P_1$ | 16  |
| 32 | 64    | 128 |

图 4-5 点  $P_1$  的八邻域权值分布

算法的具体步骤如下：

- (1) 开始扫描每个像素点，扫描顺序为从上往下，从左往右；
- (2) 当扫描到的当前像素点的像素值为 1 时，则获取以它为中心的  $3 \times 3$  区域内的八邻域的像素值与定义的模板中的权值进行相乘求和，得到索引值  $K$ ；如果像素值不是 1，则转向第一步继续扫描下一个像素点；
- (3) 根据得到的索引值  $K$  到索引表中相应的位置找到对应的数据，如果是 1，则像素值保持不变，如果是 0，则将其像素值置为 0；
- (4) 在将图像从头到尾扫描完成后，如果在这次扫描过程中有修改过像素值，则跳到步骤(1)，进行下一轮扫描，否则整个细化过程结束。

如图 4-6 为细化过程图，左侧图为未经处理过的图像，中间图为经过一次细化后的结果，右侧图为经过细化过程的最终结果。



图 4-6 细化过程图

## 2. 边缘特征提取

本文的边缘特征提取算法选用的是 canny 算子边缘特征提取的算法，具体步骤如下：

1) 采用高斯滤波器对图像进行处理，也就是选择 Gauss 滤波函数的标准方差，以及领域大小，对图像  $f(x)$  进行操作，从而获得更平滑的图像，起到抑制噪声的作用。

2) 在此基础上，对经过滤波处理后的图像每个像素梯度的幅值和方向进行计算。

3) 对梯度幅值进行非极大值抑制运算。图像上幅值  $M$  的大小并不能对边缘点进行确定，通过采用非极大值抑制的方式才能确定边缘点，细化幅值图像中的屋脊带，保留在幅值中局部变化最大的点。对图像中的幅值需要利用梯度方向来进行计算，当梯度方向上的非最大值点的赋值为零时，这点即为边缘。

4) 用双阈值算法检测和连接边缘。对于非极大值进行抑制幅值时，可以得到边缘的阵列图像，结果会受到阈值大小的影响。将这些边缘点相连形成边缘特征，如图 4-7 为提取效果图。

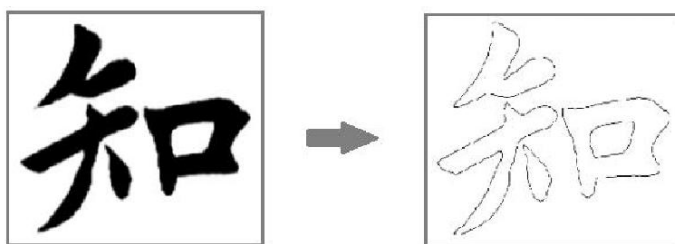


图 4-7 边缘特征提取效果图

## 3. 余弦函数特征相似度

本文中待评价样本与评价集合之间相似度的计算是使用余弦相似性来计算骨骼和边缘的优劣，加权得出最终的相似度，最后得到质量评价分数。余弦相似度的计算是使用向量来表示每个像素的灰度值，再通过余弦函数来计算两个向量的余弦值。

公式 (4-4) 为基本的余弦函数表达式。

$$\cos \theta = \frac{a^2 + b^2 - c^2}{2ab} \quad (4-4)$$

公式 (4-5) 为计算向量  $a, b$  和它们之间距离的公式，其中  $x, y$  分别表示两点的横坐标和纵坐标。

$$|a| = \sqrt{x_1^2 + y_1^2} \quad |b| = \sqrt{x_2^2 + y_2^2} \quad |c| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (4-5)$$

公式 (4-6) 为由平面二维像素点的余弦相似度计算公式推出的多维余弦相似度计算公式。

$$\cos \theta = \frac{\sum_{i=1}^n (x_i * y_i)}{\sqrt{\sum_{i=1}^n (x_i)^2 * \sum_{i=1}^n (y_i)^2}} \quad (4-6)$$

在得到待评价图像和评价集合各图像的相似度后，使用经验公式 (4-7) 得出最后的分数。

$$S = x_0 + \partial \frac{\sum_{i=1}^n (x_i - x_0) \frac{y_i}{y_i + y_0}}{n} \quad (4-7)$$

其中，S 表示最终的得分， $x_0$  为限定的基准分数，这个基准分数是待评价图像与评价集合各图像相似度最高的图像的分数，n 表示评价集合参与相似度评价的图像数量， $x_i$  和  $y_i$  分别表示第 i 个样本的分数和相似度， $\partial$  为调节参数。

#### 4.4 手写汉字质量评价系统的实现

下面将介绍手写汉字质量评价系统的整体的流程图，系统主要分为两个模块，分别为识别模块和评价模块。系统总流程图如图 4-8 所示。

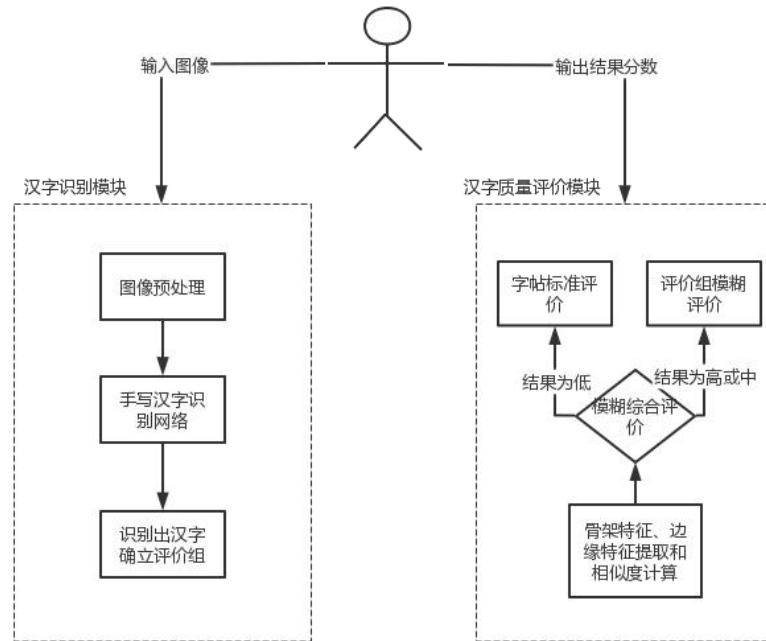


图 4-8 系统流程图

## 4.5 手写汉字质量评价实验结果分析

### 4.5.1 数据集

我们已经从互联网搜索到 100 个汉字和 300 个图片，用于评估手写汉字的质量。我们的评估方法是使用电脑自动评估，它的评估结果是用表格的方法呈现的。我们还使用人工评估的方法，它的评估得分与人工评估的得分相当，我们希望这些信息有助于我们更好地评估这种方法的精度，如图 4-9 所示。

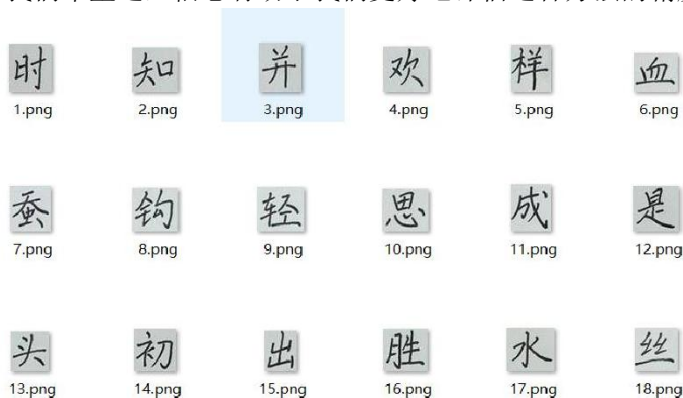


图 4-9 手写汉字初值评价标样本图

### 4.5.2 实验结果

#### 1. 骨架特征提取测试

手写汉字的骨架特征提取使用了索引表算法和数学形态算法两种算法，实验结果如表 4-1 所示。

表 4-1 两种骨架特征提取算法实验结果

| 二值化汉字图像 | 索引表法 | 数学形态算法 |
|---------|------|--------|
| 言       | 言    | 言      |
| 修       | 修    | 修      |
| 先       | 先    | 先      |

通过上表可以看出两种算法都可以有效地提取出汉字的骨架特征，但是两种算法之间提取的骨架存在差异，统计两种算法提取出的骨架的断裂次数，如表 4-2 所示。

表 4-2 两种算法的提取骨架断裂次数

| 二值化汉字图像 | 索引表法断裂 | 数学形态算法断裂 |
|---------|--------|----------|
| 言       | 0      | 0        |
| 修       | 0      | 1        |
| 先       | 1      | 4        |

通过上表可以看出由索引表算法提取的骨架基本上不存在断裂,较好的保存了汉字的形状特征和笔画之间的拓扑结构信息,而数字形态算法提取的骨架断裂次数较高,相较而言,索引表算法在提取汉字骨架上的表现更加优异。

## 2. 标准库字帖的相似度测试

采用第二种评价方式,即基于标准库的相似度评价时,我们使用九宫格分割处理方式,对每一个分割部分通过 Hu 矩和皮尔逊相关系数计算出各部分相似度,加权得到最后评价分数,这种方法使得评价结果更加精准。如图 4-10 所示为“好”字的骨架九宫格分割,先计算每个部分的 Hu 矩,再与标准字帖对应部分的 Hu 矩进行皮尔逊相关系数的计算,最后加权得到相似度。

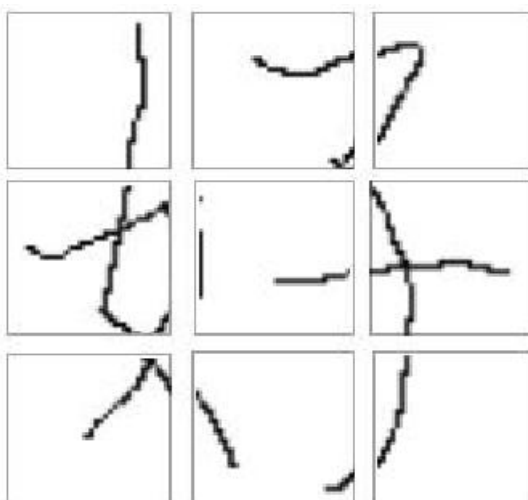


图 4-10 “好”字的骨架九宫格分割

表 4-3 基于 Hu 矩的骨架“九宫格”的相似度计算结果

| 标准字帖 | 骨架九宫格   | 待评价字体 | 骨架九宫格   | 相似度 (%) |
|------|---|-------|---|---------|
| 好    |    | 好     |    | 74.33   |
| 世    |    | 世     |    | 77.65   |
| 深    |  | 深     |  | 68.33   |

上表通过对待评价字体和标准字帖进行九宫格分割处理，Hu 矩和皮尔逊相关系数计算加权得到的相似度值基本符合真实情况，该算法具有良好的适用性。

### 3. 系统整体测试

为了检验手写汉字的质量是否与人的主观评价相一致，我们对多名测试者进行了评估，其中包括 2 名教师和 2 名项目团队学生，根据评估标准进行打分，最终将这些分数相加，以得出最终的平均分。通过与计算机提供的评分进行对比，我们可以看到表 4-5 中的详细信息。

表 4-5 评分表结果

| 编号  | 学生 1  | 学生 2  | 老师 3  | 老师 4  | 计算机 5       |
|-----|-------|-------|-------|-------|-------------|
| 写 同 | 63/56 | 65/52 | 61/53 | 61/50 | 63.21/56.23 |
| 俗 俗 | 54/32 | 53/33 | 52/34 | 52/36 | 56.32/32.89 |
| 兵 兵 | 57/31 | 54/34 | 53/35 | 53/38 | 51.38/35.99 |
| 荷 荷 | 73/43 | 76/46 | 75/53 | 74/48 | 70.89/41.89 |
| 兵 兵 | 76/43 | 77/47 | 76/63 | 75/47 | 76.32/40.26 |
| 于 于 | 80/54 | 88/58 | 85/54 | 86/50 | 80.45/51.79 |
| 荷 荷 | 65/41 | 61/49 | 61/51 | 69/49 | 61.59/48.90 |

通过对比，我们发现，在处理手写汉字时，电脑和人类的评分并没有太多的区别。这表明，这种方法的评估结果非常准确。为了更好地分析这种方法的效率，我们运用 SPSS 分析软件来研究它们的相互影响。经过 SPSS 软件的检验，发现人类的主观和计算机的客观评估之间存在着 0.156 的相关性，而且这种相关性的显著性也达到了 0.243（当  $S_g$  小于 0.33 时，这种结论是非常重要的），因此这个系统的评估是可信的。

## 第五章 总结与展望

### 5.1 总结

由于汉字的多样性和复杂性，它们的识别历史上一直被认定为比较棘手的问题，但近年来，它们也被广泛应用于模型识别领域，并受到了越来越多的关注。特别是当今，手写汉字的识别不仅仅局限于移动设备的输入，而且也被广泛应用于各种邮件和包裹的查询，从而发挥出它们的独特优势。近年来，由于深度学习的飞速发展，它的应用范围不断扩大，特别是在计算机视觉和模式识别的领域，它的优势显而易见。尤其是在汉字识别，它不再局限于使用现存的网络模型，而是充分考虑到汉字的内涵，从而取得更好的效果。汉语的文化源远流长，它的基础是丰富的文化元素，比如笔画和偏旁。随着时间的推移，这些元素的结合变得越来越稳定，形成了一套完整的文化体系。目前，人们正致力于探索更好地使用这些元素来实现更高效的文化认知。总结本文完成工作如下：

为了解决传统的梯度特征和深度可分离网络的识别精度受限的问题，本文提出了一种基于注意力分组卷积方法，将 MBConv 模块与传统的梯度特征相互取代，并采取洗牌算法，通过注意力机制设置相应的权重值，实现多个分组的有效融合，从而大大改善了汉字的识别精度。针对手写汉字质量评价，我们确定一种特定的汉字种类，构成一个由这些种类的汉字构成的评估小组，每个小组都包含一些具体的汉字，我们可以根据这些小组的特征来给予它们不同的等级。我们将所提的模型应用于中科院自动化研究所的 CASIA-HWDB1.1 手写汉字数据集中，经过多次测试，其准确度和可信度得到显著改善，与原有的 mAP 相比，其准确度和可信度均有 9.04% 的改善。通过使用深入学习方法，我们对手写汉字的质量进行了检查，并将检查的结果与传统的方法进行了比较，以确保该系统的准确度和可信度。

### 5.2 展望

本文在研究手写汉字识别的问题上，利用注意力机制对汉字信息中较为重要部分分配较高权重值，由此可切分出合适的模板以及利用手写汉字特征值挑选最合适的模板得到评价文本，但存在不同的手写汉字特征偏差值词性意义相同，因此并不能对于所有的手写汉字都能达到一个较高的标准，后续会考虑到用一些词典处理的方式对错误率较高的内容进行较真

同时本文用于实验的数据量的数量不够。利用深度可分离的神经网络实现手写汉字模板提取需要大量的训练数据。而目前应用的数据集较小且仅局限于当前手写汉字评价领域中，并且每个训练数据的质量不同，使得实验结果存在局限。



## 参考文献

- [1] 杨朋波, 桑基韬, 张彪, 冯耀功, 于剑. 面向图像分类的深度模型可解释性研究综述[J]. 软件学报, 2023, 34(01): 230-254.
- [2] Zhao Yuliang, Zhang Xinyue, Fu Boya, Zhan Zhikun, Sun Hui, Li Lianjiang, Zhang Guanglie. Evaluation and Recognition of Handwritten Chinese Characters Based on Similarities[J]. Applied Sciences, 2022, 12(17): 8521-8521.
- [3] 刘萌, 齐孟津, 詹圳宇, 曲磊钢, 聂秀山, 聂礼强. 基于深度学习的图像-文本匹配研究综述[J/OL]. 计算机学报: 1-32[2023-04-23].
- [4] 李学龙, 赵致远. 像素级语义理解: 从分类到回归[J]. 中国科学: 信息科学, 2021, 51(04): 521-564.
- [5] 邱祥庆, 刘德喜, 万常选, 李静, 刘喜平, 廖国琼. 文本情感原因自动提取综述[J]. 计算机研究与发展, 2022, 59(11): 2467-2496.
- [6] Yu Wanbo, Li Yaosheng, Peng Hongwei, Zhang Li. Image iterative method for handwritten Chinese character recognition[J]. Journal of Physics: Conference Series, 2020, 1684(1): 12101-12107.
- [7] Liu ChengLin, Sako Hiroshi, Fujisawa Hiromichi. Discriminative learning quadratic discriminant function for handwriting recognition[J]. IEEE transactions on neural networks, 2004, 15(2): 430-444.
- [8] 周正扬, 詹恩奇, 郑建彬, 胡华成. 局部关联度最优的手写汉字骨架提取[J]. 中国图象图形学报, 2017, 22(06): 833-841.
- [9] Li Zhiyuan, Yi Xiao, Qi Wu, Min Jin, Lu Huaxiang. Deep template matching for offline handwritten Chinese character recognition[J]. The Journal of Engineering, 2020, (4): 120-124.
- [10] Dan Yongping, Li Zhuo. Particle Swarm Optimization-Based Convolutional Neural Network for Handwritten Chinese Character Recognition[J]. Journal of Advanced Computational Intelligence and Intelligent Informatics, 2023, 27(2): 165-172.
- [11] 张重生, 陈杰, 李岐龙, 邓斌权, 王杰, 陈承功. 深度对比学习综述[J]. 自动化学报, 2023, 49(01): 15-39.
- [12] 李雷孝, 孟闯, 林浩, 高静, 王慧. 基于图像增强与深度学习的安全带目标检测[J]. 计算机工程与设计, 2023, 44(02): 417-424.
- [13] 韩彦岭, 沈思扬, 徐利军, 王静, 张云, 周汝雁. 面向深度学习图像分类的 GPU 并行方法研究[J]. 计算机工程, 2023, 49(01): 191-200.
- [14] 张涛, 唐华, 张甜甜. 基于深度神经网络训练与优化算法的图像分类[J]. 计算机仿真, 2023, 40(02): 284-287.
- [15] 王小鹏, 魏统艺, 房超, 朱生阳. 自适应非局部空间约束与 K-L 信息的模糊 C-均值噪声图像分割算法[J]. 控制理论与应用, 2022, 39(07): 1261-1271.
- [16] 王振彪, 覃亚丽, 王荣芳, 郑欢浙. 基于残差特征聚合的图像压缩感知注意力神经网络[J]. 计算机科学, 2023, 50(04): 117-124.
- [17] 肖雪, 李成城. 手写汉字评价方法研究进展[J]. 计算机工程与应用, 2022, 58(02): 27-42.
- [18] 徐翰文, 张闯, 陈苏婷. 基于视觉显著性的图像特征提取算法[J]. 计算机应用, 2022, 42(S2): 72-78.
- [19] 李盼盼, 王朝立, 孙占全. 基于注意力机制的多特征融合人脸活体检测[J]. 信息与控制, 2021, 50(05): 631-640.
- [20] 阴敬方, 朱登明, 石敏, 王兆其. 基于引导对抗网络的人体深度图像修补方法[J]. 系统仿真学报, 2020, 32(07): 1312-1321.
- [21] 左一帆, 方玉明, 马柯德. 深度学习时代图像融合技术进展[J]. 中国图象图形学报, 2023, 28(01): 102-117.
- [22] 王佳欣, 朱志亮, 邓小明, 马翠霞, 王宏安. 基于深度学习的草图分割算法综述[J]. 软件学报, 2022, 33(07): 2729-2752.

- [23] 江俊君,李震宇,刘贤明.基于深度学习的单目深度估计方法综述[J].计算机学报,2022,45(06):1276-1307.
- [24] 王军敏,樊养余,李祖贺.基于深度卷积神经网络和迁移学习的纹理图像识别[J].计算机辅助设计与图形学学报,2022,34(05):701-710

## 致谢

时光飞逝，四年的大学时光就在不经意中即将结束了。人生最大的财富，莫过于岁月留下的痕迹。

首先，我要感谢我的导师李重老师、胡海龙老师，从论文定题到论文最终完成，给予了我极大的帮助，在论文的每一个环节都非常认真和严谨地给予我指导和建议，这样的责任心和工作态度使我深受感动和鼓舞，成为我顺利完成论文的动力。在此，我对我的导师表示衷心的感谢！

感谢大学期间我的同学们、室友，四年的相伴不长不短，但我要感谢你们四年来对我的帮助和带给我的快乐。

感谢我的家人，感谢多年来的包容和理解，是你们让我知道了很多是非道理，也明白了每个人都有自己的或大或小的人生故事，接下来在人生道路上该怎样更好的演绎自己的角色，做好自己的本分。

感谢审阅这篇文章的各位老师，你们的指导和批评是我今后继续深造和学习的动力。

亲情、友情永远是我生活和学习不断奋进的最大动力源泉。不管人生的下一站将驶向何方，在湖师四年的读书生活是令我永远难忘的。