

**2023届毕业设计(论文)**

**课 题 名 称: 卷积神经网络手写英文字母识别研究**

**课 题 名 称（英文）： Research on Handwritten English Letter Recognition Based on Convolutional Neural Network**

**学 生 姓 名： 王妮妮 学 号： 2019082526**

**专 业 名 称： 计算机科学与技术\_\_\_\_\_\_\_\_\_\_**

**指 导 教 师： 唐琦哲 职 称： 讲师\_\_\_\_**

**所 在 学 院： 信息工程学院\_\_\_\_\_\_\_\_\_\_\_\_\_**

**完 成 日 期： 2023 年 3 月 26 日**

**教务处制表**

**卷积神经网络对手写字母识别的研究**

**摘要**：随着当前世界形势的不断变化以及我国国民经济的高速健康发展，通过模式识别技术实现大量信息的自动识别，以方便人们的生活和提高工作的效率是顺应时代发展的必然趋势。为和科技时代发展的趋势相适应、更好地优化我们的日常生活方式，自动检测各类文字信息、快速而准确地执行信息的存储和识别等过程对我们来说尤为重要，将脱机手写字符识别真正应用于实际生活和工作当中，这并不是口头说说就能办到的事，实现的过程必然面临着许多的问题和挑战，为顺应科技时代要求，我们要克服这重重困难。我们的日常生活逐渐离不开信息技术，验证码识别、车牌号识别等应用在当今社会较为常见，相比较印刷体的识别手写字体形式多样、相似性较低，因此手写字母的识别要求更高，但其在现实生活中应用也并不少见，手写体的识别在人机交互中扮演着重要角色，在社会生活中也显得尤为重要，可以通过手写体的识别来批改学生日常作业等，减少老师的工作量，提高工作效率。

**关键词：**卷积神经网络，深度学习，PyTorch，可视化

**Research on Handwritten English Letter Recognition Based on Convolutional Neural Network**

**Abstract**:With the continuous changes in the current world situation and the rapid and healthy development of China's national economy, it is an inevitable trend to adapt to the development of the times to achieve automatic recognition and processing of a large amount of information through pattern recognition technology, in order to continuously facilitate people's lives and improve work efficiency. In order to adapt to the development trend of the technology era and better optimize our daily lifestyle, it is particularly important for us to automatically detect various types of text information, quickly and accurately perform information storage and recognition processes. Our goal is to truly apply offline handwritten character recognition to real life and work. This is not something that can be done verbally, and the implementation process inevitably faces many problems and challenges, But to comply with the requirements of the technological era, what we need to do is overcome these difficulties. Our daily life is increasingly inseparable from information technology. Applications such as verification code recognition and license plate number recognition are more common in today's society. Compared to printed recognition, handwritten font recognition has diverse forms and low similarity, so the recognition requirements for handwritten letters are higher. However, its application in real life is not uncommon. Handwritten recognition plays an important role in human-computer interaction, and it also appears particularly important in social life, Handwriting recognition can be used to correct students' daily homework, reduce the workload of teachers, and improve work efficiency. Therefore, research on improving recognition accuracy is beneficial to life and development.

**Keywords:** Convolutional neural network,Deep learning, PyTorch,Visualization

**目 录**

[第一章 绪 论 1](#_Toc31984)

[1.1 选题的意义 1](#_Toc602)

[1.2 研究现状及发展趋势 2](#_Toc26219)

[1.2.1 研究现状 2](#_Toc8126)

[1.2.2 发展趋势 3](#_Toc15328)

[1.3 手写字母识别研究难点 4](#_Toc25876)

[1.4 本章总结 4](#_Toc11996)

[第二章 卷积神经网络的相关研究 5](#_Toc1739)

[2.1 卷积神经网络 5](#_Toc3798)

[2.1.1 深度学习简介 5](#_Toc30750)

[2.1.2 卷积神经网络简介 5](#_Toc25915)

[2.1.3 反向传播算法 5](#_Toc23418)

[2.1.4 激活函数的必要性 6](#_Toc22185)

[2.2 卷积神经网络的结构 6](#_Toc18242)

[2.2.1 输入层 7](#_Toc6968)

[2.2.3 激活层 11](#_Toc2439)

[2.2.4 池化层 12](#_Toc11707)

[2.2.5 全连接层（fully-connected layer） 13](#_Toc10924)

[2.2.6 输出层 14](#_Toc4706)

[2.3 本章总结 14](#_Toc11797)

[第三章 数据准备与预处理 15](#_Toc22314)

[3.1 数据集简介 15](#_Toc18763)

[3.2 数据预处理 17](#_Toc17655)

[3.3 卷积神经网络对手写字母识别的设计 17](#_Toc19770)

[3.4 本章总结 19](#_Toc11851)

[第四章 系统设计与实现 20](#_Toc6778)

[4.1 系统运行环境 20](#_Toc28922)

[4.2 模型读取并测试自己的数据 20](#_Toc190)

[4.3 本章总结 23](#_Toc10836)

[第五章 总结与展望 24](#_Toc10976)

[5.1 总结 24](#_Toc22919)

[5.2 展望 24](#_Toc1056)

[参考文献 25](#_Toc19032)

# 绪 论

本章通过对卷积神经网络对字符识别相关的文献和研究资料进行查阅，介绍分析了卷积神经网络对手写英文字母的识别算法,包括其背景、在现实当中的意义、研究现状、发展趋势及研究过程所存在的疑难点。

## 1.1 选题的意义

模式识别技术是人工智能的一个重要分支，它利用计算机算法和数学模型，从数据中自动地识别和提取出各种规律和特征，从而实现自动化的数据分析和决策。随着科技的不断进步和社会的快速发展，我们所处的信息时代中，数据已经成为一种非常宝贵的资源。而模式识别技术则是实现对这些数据的自动化处理和分析的关键工具之一。例如，我们可以利用模式识别技术来自动化地识别图像、语音、文本等数据中的特征和模式，从而实现自动化的图像识别、语音识别、自然语言处理等应用。在当前的世界形势和我国国民经济高速健康发展的背景下，模式识别技术的应用前景非常广阔。它可以帮助企业、政府、医疗机构等实现更高效的数据分析和决策，提高生产力和效率，为人们的生活带来更多的便利和创新。同时，模式识别技术也是未来人工智能发展的重要方向之一，将在人机交互、智能机器人、自动驾驶等领域带来更多的创新和突破。因此，积极推进模式识别技术的研究和应用，不仅是顺应时代发展的必然趋势，也是促进我国科技进步和经济发展的重要举措[1]。

为和科技时代发展的趋势相适应、更好地优化我们的日常生活方式，自动检测各类文字信息、快速而准确地执行信息的存储和识别等过程对我们来说尤为重要，我们的目标是将脱机手写字符识别真正应用于实际生活和工作当中，这并不是口头说说就能办到的事，实现的过程必然面临着许多的问题和挑战，但顺应科技时代要求，我们要做的就是克服这重重困难。我们的日常生活逐渐离不开信息技术,验证码识别、车牌号识别等应用在当今社会较为常见，相比较印刷体的识别手写字体形式多样、相似性较低，因此手写字母的识别要求更高，但其在现实生活中应用也并不少见，手写体的识别在人机交互中扮演着重要角色，在社会生活中也显得尤为重要，可以通过手写体的识别来批改学生日常作业等，减少老师的工作量，提高工作效率，因此对识别正确率的提高研究有利于生活和发展[4]。

脱机手写字符识别是一项非常重要的技术，它可以实现从手写字符图像中自动识别和提取出字符信息，从而方便人们的生活和工作。要将脱机手写字符识别真正应用于实际生活和工作当中，可以从以下几个方面入手：

1.提高识别准确率：脱机手写字符识别的准确率对于其应用的成功非常关键。要提高识别准确率，可以从优化算法、增加训练数据、改善图像质量等方面入手，不断完善和提升算法和技术的性能。

2.应用场景扩展：脱机手写字符识别技术的应用场景非常广泛，可以应用于身份证、银行卡、签名、信件等多种场景。要将其应用到更多的场景中，需要根据不同场景的特点和需求，进行适当的算法调整和模型优化，以提高识别准确率和性能。

3.结合其他技术：除了脱机手写字符识别技术本身外，还可以结合其他相关技术，如OCR（光学字符识别）技术、图像处理技术等，以进一步提高识别准确率和性能。

4.推广普及：最终实现脱机手写字符识别技术的应用，还需要大力推广普及。可以通过开展科普宣传活动、推广相关应用产品等方式，让更多人了解和使用这项技术，从而进一步提高其应用价值和社会影响力[1]。

## 1.2 研究现状及发展趋势

### 1.2.1 研究现状

总体概括：手写字母识别要做的就是将图片中输入的样本字母与对应的标准数字检测然后进行模式匹配，它的最终识别输出结果是模式匹配中最大类似程度的样本字母。在识别字母的过程中，要注意选择样本特征提取方法、按照需求设计符合要求的分类器，特征提取方法和分类器的设计越好，识别系统的所达到的效果和性能越优。近年来卷积神经网络渐渐被人们所熟知，卷积神经网络渗透多个方面，包括基于农学方面植物种类的识别、针对害虫的识别；基于FPGA卷积神经网络的人脸识别；基于轻量卷积神经网络的车牌定位识别；还有化学地质方面对矿石识别；物理方面对分子散射成像材料识别方法的研究等。除此之外生物、艺术等方面也大有体现。然而正是因为其包含面广、覆盖范围大，研究方面也大有不同。本篇文章致力于Pytorch利用CNN卷积神经网络对手写字母的识别，与普通字母识别的不同之处在于普通字母识别的识别范围有限，只能识别固定像素范围内的图片，显然使用卷积神经网络能大大节省人力，使识别更加智能方便。由于对手写字母的识别不存在语义相关性，而一般情况下我们对所识别的准确性要求较高，所以研究时对其识别精确需要达到更高的水平。

当前研究成果：随着最近几年人工智能相关概念的火爆，神经网络、深度学习等概念被人熟知，但是这些东西不是最近才出现的，神经网络自上个世纪就开始了研究发展。神经网络的第一个模型（即MP模型）在1943年被提出，MP模型为神经网络的研究提供了基本数学理论[2]；而后Hopfield模型、Boltzmann模型、BP算法等在上世纪80年代被相继提出，在解决许多实际问题时起到了重要作用，该算法对今天的研究也有着现实意义[1]。80年代末，LeCun等人结合反向传播算法提出了第一个卷积神经网络模型，研究了卷积神经网络对手写邮政编码的识别，这种方法与传统识别技术不同的是，它的图像预处理方式更加简单，减轻了特征提取的工作量；秉持着提高手写字符识别准确度的原则，LeCun等人在第一个卷积神经网络模型的基础上，经过改进，他们的模型在美国邮政总局提供的邮政编码数字数据的测试结果表明：错误率仅为1%，拒绝率约为9%；1998年，他们将手写数字识别的各种方法在标准的手写数字识别基准上进行比较，结果表明他们的网络优于其他所有模型，经过多年的研究和迭代，最终发展成为LeNet-5[3]。

后来，逐层预训练、深度置信网络（DBN）、层叠自动去噪编码机（SDA）等网络模型也相继被提了出来，至此，模型提取特征的能力得到了有效的提高。针对卷积神经网络的识别，是随时代发展的必然趋势和要求，为不断提高生活的便利性和工作的效率，王松在他的论文中对卷积神经网络的手写字符识别方法进行了研究，通过分析Nist和Chars 74K 数据集，对其中异常数据进行规范化处理，提出了一种基于Goole Inception V2-a模型的手写识别方法[2]。何国对在他的论文中提出了基于注意力机制的图卷积文本分类模型，传统的图卷积神经网络文本分类模型利用与其连接的词汇顶点更新文档顶点信息，这种分类的聚合方式没有充分考虑到文档顶点对每个与其所连接的词汇的异同，然而每个词汇对文档的重要程度是不同的[3]。文章中提出了一种轻量级的图注意力机制，避免了计算资源占用过大的问题，并搭建了基于注意力机制的图卷积文本分类模型引导模型关注相关性更大的词汇，从特征的多样性的角度来提取特征，通过词共现的方法构建文本数据中样本间的关联关系，用来提升文本分类的性能。屈喜文等人提出了又一种卷积神经网络，该模型实现了编码器与解码器的结合，增加了网络深度。

总结及其不足之处：手写字母识别主要分为联机手写字母识别和脱机手写字母识别，相较于脱机手写字母识别，联机手写字母识别可以通过手写的笔顺和起势等来作出判断，难度较低，因此联机手写字母的识别比脱机手写字母的识别准确率更高。而印刷体书写与手写字母相比整体较为规范，毫无疑问，其识别准确率高于手写字母的识别。因此，在当前字母和字符识别的研究中，印刷体和联机手写字母的识别技术相对来说发展较为成熟，但脱机手写字符识别还需要进一步研究[4]。

### 1.2.2 发展趋势

近年来随着卷积神经网络的不断发展，卷积神经网络的研究在多个领域取得了重大突破，也可以作为多种研究学习的有效工具。卷积神经网络识别与传统的字符识别技术相比，过程更加简洁、识别也更加出色[2]。因此，本文通过相关文献研究，在所整理的The Chars 74K EnglishImg和EMNIST dataset手写字母数据集上，基于原有的卷积神经网络模型对其识别进行研究，以求能够提高卷积神经网络对手写英文字母识别的准确率。

## 1.3 手写字母识别研究难点

国内外对手写英文字母的识别研究仍存在很多难点和挑战需要解决克服，具体如下：

1. 数据集。由于人们的书写习惯因人而异，且收集和提取的数据涵盖面不一定足够广泛，也并不一定具有普遍化的特点，因此，可查询到的数据集数量不足以做大量研究。
2. 书写的多样性。字母笔顺简单且结构相似，受手写字母书写的多样性和相似性影响，这对手写字母的识别无疑是一大难题。
3. 使用场合的严格性。涉及字母及数字的场合多为其代表特殊含义或信息，失之毫厘，差之千里，故对手写字母的识别准确度要求较高，模型检测更加严格，因此对识别模型的精确度要求更高[3]。

## 1.4 本章总结

本章通过对卷积神经网络对字符识别相关的文献和研究资料进行查阅，介绍分析了卷积神经网络对手写字母算法的背景、研究现状、发展趋势及研究过程所存在的疑难。模式识别技术是人工智能的一个重要分支，它利用计算机算法和数学模型，从数据中自动地识别和提取出各种规律和特征，从而实现自动化的数据分析和决策。随着科技的不断进步和社会的快速发展，我们所处的信息时代中，数据已经成为一种非常宝贵的资源。而模式识别技术则是实现对这些数据的自动化处理和分析的关键工具之一。

# 卷积神经网络的相关研究

## 2.1 卷积神经网络

### 2.1.1 深度学习简介

深度学习是一项十分具有前景的技术，通过研究人员长时间的研究，已经在许多地方获得重大的成果。深度学习自身具有十分强大的学习能力，并且在模型训练参数不断的调整下，往往能够找到当中的信息规律。深度学习随着多年的发展也出现了一些时序预测相关的模型，长短期记忆网络更是深度学习时序预测的典型代表。相比深度学习在其它地方的应用成果，这种技术在预测算法这一方面也许还不能解决各种各样的问题，不过，人工智能并不是一直原地踏步，相反，人工智能具有极强的生命力和广阔的发展前途，而传统算法却是越来越具有局限性，所以相信随着时间的发展将会有越来越多的人去尝试利用深度学习来获得更好的预测效果，弥补传统预测算法的各种短板[15][7]。

### 2.1.2 卷积神经网络简介

卷积用通俗的话说，就是在把图片的一些特征提取出来。在神经网络中，保持了层级网络结构，不同层级使用不同的形式（运算）与功能，其使用前一层的输出值乘以每条边的weight（权重），得到的值记为后一层的输入层，首先把图片矩阵和卷积矩阵展平，变成一维向量的值，可以把卷积成的这个值看成是与对应节点的权重，而其他不参与卷积的节点的边的权值可以看成0[4]。卷积之后得到的结果就是卷积层神经元的值。而最大池化起到让图片中的特征更加突出的作用，最大池化的过程可以执行多次，每次都会得到一张比之前更小的图片。而全连接层中我们把一个多维的矩阵“加工”变成一个一维的向量（或被称为数组）作为全连接层的输入。而后面就是一个单纯的全连接神经网络的训练。卷积神经网络（CNN）被定义为采用监督方式训练的一种面向两维形状不变性识别的特定多层感知机，是[深度学习](https://baike.baidu.com/item/%E6%B7%B1%E5%BA%A6%E5%AD%A6%E4%B9%A0/3729729?fromModule=lemma_inlink" \t "https://baike.baidu.com/item/%E5%8D%B7%E7%A7%AF%E7%A5%9E%E7%BB%8F%E7%BD%91%E7%BB%9C/_blank)（deep learning）的代表算法之一 ，它还有“平移不变人工神经网络（SIANN）”的名称[9]。卷积神经网络主要在图像识别和物品分类上应用比较广泛。

### 2.1.3 反向传播算法

1986年Williams等人在《Nature》上首次提出了著名的反向传播算法，我们现在所使用的深度网络模型大多都是在这个网络的基础上发展出来的。反向传播算法从后向前计算参数梯度值、是神经网络中加速计算参数梯度值的方法，但计算机中真实的计算过程是一个更加模块化的过程。此算法的优势是大幅降低了模型训练所需要的时间[5]。其学习过程由如下信号的正向传播与误差的反向传播两个过程组成[3]。

正向传播时，输入层传入输入样本，经各隐层逐层处理，即每个神经元都对它的输入和权重相乘再相加，然后加上偏置，最后使用非线性激活函数，重复此过程，直至传递到输出层。最终输出输出层的结果，这个过程称为正向传播（Forward Propagation）。在神经网络中，每个神经元接收来自前一层神经元的信号，经过一定的计算后，再将信号传递到下一层神经元。在正向传播过程中，每层神经元都有一组权重参数，用来控制信号的传递和计算，同时每个神经元还有一个偏置项，用来调整神经元的激活状态。输入样本经过输入层后，每个神经元将输入信号与其权重参数相乘，然后加上偏置项，再经过激活函数处理，最终输出该神经元的激活值。这个过程在每层神经元中都重复进行，直到输出层产生最终的结果。这个结果可以是分类问题中的预测类别，也可以是回归问题中的数值预测结果。正向传播过程完成后，可以将输出结果与实际结果进行比较，计算损失函数，并用反向传播算法更新网络参数，进一步优化网络性能。如果得到的输出不符合期望的输出，则进行反向传播阶段。反向传播（Backward Propagation）是一种用于训练神经网络的算法，它通过将损失函数关于网络参数的梯度反向传播，逐层更新网络参数，从而使网络逐渐逼近最优解。反向传播算法是基于链式法则（Chain Rule）来计算网络参数梯度的，它将输出层的误差信号沿着网络逐层反向传播，并将误差信号分摊给每个神经元，然后根据误差信号来更新每个神经元的权重参数和偏置项。具体来说，在反向传播过程中，首先计算输出层的误差信号，然后逐层向前计算每个神经元的误差信号，最终将误差信号传递到输入层。在每层神经元中，根据误差信号和激活函数的导数，计算该神经元的权重参数和偏置项的梯度，并使用梯度下降法来更新这些参数。通过反复迭代正向传播和反向传播过程，可以不断优化网络的性能，使其能够更准确地预测目标值。需要注意的是，反向传播算法要求神经网络中的每个神经元都是可微的，这意味着神经元的激活函数必须是可导的。另外，反向传播算法可能会出现梯度消失或梯度爆炸的问题，在实际应用中需要注意处理这些问题。

它的训练过程主要是为了找到损失函数最小的权重。公式如下：

（2.1）

梯度下降：求得每个参数相对于损失函数的偏导，用偏导数乘以学习率得到需要改变的量，用原权重减去改变的量，得到新的权重。再进行正向传播，计算新的损失值，重复此过程使得损失函数为零或在设定的范围，完成训练。公式如下：

 （2.2）

### 2.1.4 激活函数的必要性

在无激活函数的情况下，神经元的输出就是输入与对应参数的乘积之和，或许再加一个偏置，但是这些都是线性的，感知机的堆叠依然是一个线性的函数。实际场景中很多问题不是线性可分的，激活函数的作用就是让神经元的输出有一定的非线性性质[7]。

## 2.2 卷积神经网络的结构

卷积神经网络是为了专门应对一个2D的图片数据开发出来的，一个二维的灰度的图片数据每一个元素上的数值用（0-255）或者（0-1）的数值表示，用[b,h,w,1]来表示有b张照片，h行w列，维度为一维，一维表示单通道的一个图片，而对于一个彩色的图片，可以用三个这样的通道，表示为[b,h,w,3]，每个通道分别表示为R、G、B，将三个这样子分开化的矩阵合在一起称作最开始的image，一个基本的卷积神经网络有数据输入层、卷积计算层、ReLU激励层、池化层和全连接层构成，以LeNet模型为例[3]。 LeNet模型结构如图2-1所示：

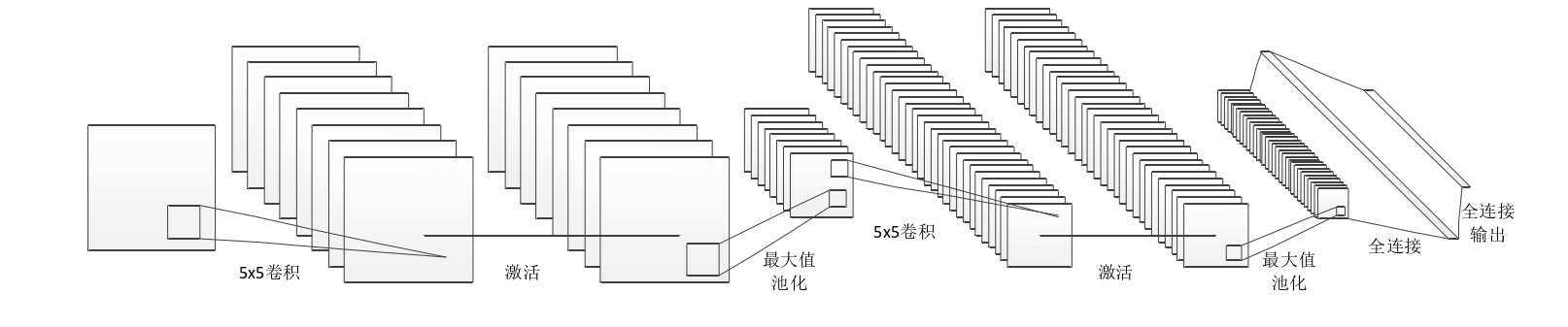


图2-1 LeNet模型结构

### 2.2.1 输入层

卷积神经网络（Convolutional Neural Network，CNN）和传统神经网络/机器学习一样，需要对输入的数据进行预处理。数据预处理的主要原因是为了提高网络的性能和训练效率，同时避免数据的不良影响，使得网络更容易收敛和泛化。

常见的三种数据预处理方式包括：

1.去均值（Mean subtraction）：将输入数据的各个维度中心化到0，即将数据的均值减去每个样本的特征值。这种处理方式可以去除数据集中的平移因素，使得训练过程更稳定，收敛速度更快。

2.归一化（Normalization）：将输入数据的各个维度的幅度归一化到同样的范围，使得不同的特征具有相同的重要性，避免数据范围大的输入在模式分类中的作用偏大，而数据范围小的作用偏小。常用的归一化方法包括将输入数据缩放到0和1之间或将数据缩放到均值为0，方差为1的标准正态分布。

3.PCA/白化（PCA/Whitening）：用PCA降维，能够去掉特征与特征之间的相关性，使得数据更加独立，从而提高网络的泛化能力。而白化则是在PCA的基础上对转换后的数据每个特征轴上的幅度进行归一化，使得每个特征的方差相等，从而提高特征的区分度和分类性能。

需要注意的是，不同的数据预处理方法适用于不同的数据类型和应用场景，选择合适的预处理方法可以提高网络的性能和准确度。此外，在预处理过程中还需要注意对数据的标准化、去除异常值和填充缺失值等问题。去均值化和归一化操作如图2-2所示：

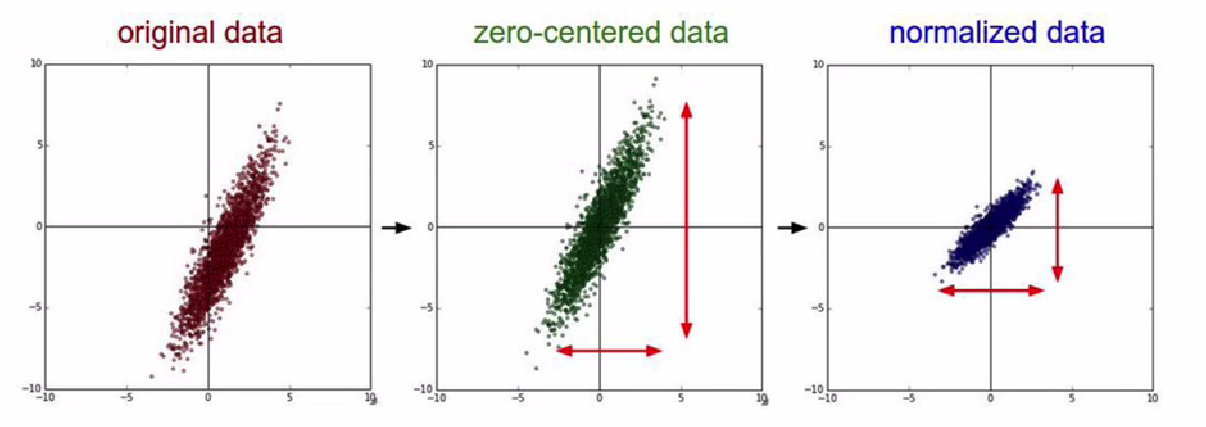


图2-2 去均值和归一化

PCA和白化操作如图2-3所示：

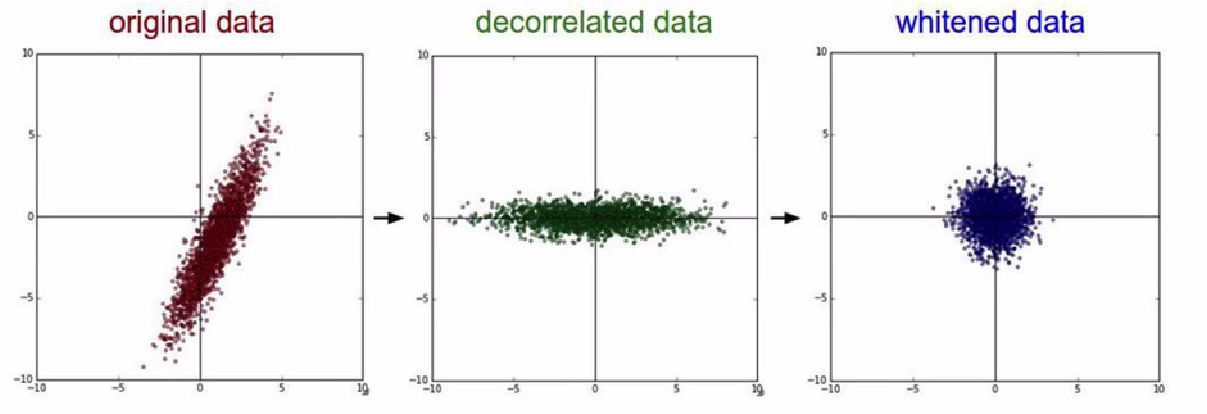


图2-3 PCA&白化

实际上在卷积神经网络当中，去均值和归一化是最常用的数据预处理方式，因为它们能够有效地提高模型的性能和训练效率。去均值的操作可以使数据的中心化到0，从而消除数据集中的平移因素，使得模型更容易收敛。归一化的操作则可以将不同特征的数据幅度统一到同一个范围内，使得不同的特征具有相同的重要性，避免数据范围大的输入在模式分类中的作用偏大，而数据范围小的作用偏小。这些操作可以使得模型更加稳定，提高泛化性能，并且在训练过程中降低了过拟合的风险。相比之下，PCA和白化的操作更加复杂和计算密集，而且在实际应用中也较少使用。PCA可以通过降维操作去掉特征之间的相关性，但是它可能会损失一些有用的信息，并且在实际应用中常常会被替代为更加高效的卷积操作。白化则需要进行复杂的计算，并且可能会引入噪声，因此在实际应用中使用较少。因此，去均值和归一化通常是最好的选择，它们可以有效地提高模型的性能和训练效率，并且可以在实践中轻松应用。在使用数据集时，对数据特征进行去均值和归一化的预处理是非常常见和重要的步骤。

2.2.2 卷积层

卷积层经历卷积、池化、激活的过程，得到的特征图进行展平进入神经网络，得到预测的概率，通过反向传播的方式不断调整这些参数，以达到预测值和真实值最接近的情况，完成网络的训练。卷积层训练参数时，参数初始时随机的，但在经过训练之后计算机会帮我们找到能够体现它的特征的一些值，比如从左上到右下，卷积时我们发现，特征图保留了原始图像的特征，特征图在原始图像区域的值趋近于1，但在其他没有体现特征图位置的值都很低，把卷积核的特征进行了提取。

1.卷积核

卷积层主要是由卷积核对图像进行卷积操作。输出数据的通道数是由卷积核个数所决定的。1\*1卷积核对输入数据的通道做简约，每个1\*1卷积核相当于在输入数据的通道上做了一个降维（经过一个神经元个数为1的全连接层），从而相当于大幅度降低了特征图的数量，但不影响特征图的结构。卷积层主要是提取输入数据的特征，卷积层内部包含多个卷积核（滤波器），每卷积神经网络中的每个卷积核都具有相同的形状，通常为(h,w,d)，其中h、w分别对应上一层输入的长和宽，d则对应上一层输入的通道数。这组数据的保证能够确保输出特征图的形状一致，从而使得卷积操作具有平移不变性，也就是说，如果输入的图像发生了平移变化，输出的特征图不会发生变化，这也是卷积神经网络具有良好的平移不变性的原因之一。由于每个卷积层所具有的卷积核不同，因此得到的特征映射关系也不同，从而产生了不同的特征图，这些特征图能够提取出不同的特征，例如边缘、纹理、角点等等，以此保证了卷积神经网络的多样性和表达能力，也是其在图像处理领域中得到广泛应用的重要原因之一。因此，卷积神经网络中的卷积核具有相同的形状，并且每个卷积层所具有的卷积核不同，这样能够保证提取出多种多样的特征，从而提高了卷积神经网络的表达能力和分类性能[2]。每个卷积核都有一组权重和一个偏置项，可以视为与前一层的神经元连接的权重共享的一组参数。卷积核的大小决定了每个神经元在前一层输入中所处理的区域大小，这也是卷积神经网络中卷积操作的核心。因为卷积核中的参数是共享的，所以可以显著减少需要学习的参数数量，从而提高网络的效率和泛化能力。卷积操作如图2-4所示：

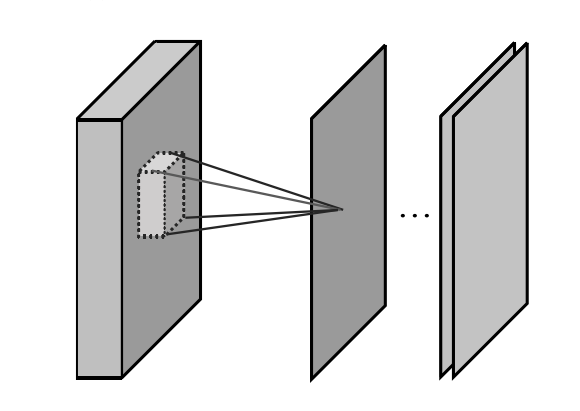


图2-4 卷积操作

卷积核按照从左向右、从上向下的顺序，以一定的步长在卷积核对应的矩阵上依次遍历卷积层的每一个输入，将滤波器所覆盖的子空间V上的长、宽、通道数的乘积（即h\*w\*d）个像素（神经元）转化为当前层上的一个神经元，直到遍历到输入矩阵的右下角为止。卷积层基于局部感受野和权值共享的特点既提取了输入的所有的有效特征，又有效的减少了模型的参数量，这样就减少了模型的计算量、减轻了模型过拟合的影响。卷积操作的计算公式为[3]：

 （2.3）

其中，为第层的第个神经元；为求第层第个神经元所对应的感受野；为第-1层的感受野内的第神经元；为第层滤波器矩阵的第个权值；b为每个滤波器的偏置，而且不同的滤波器可以设置不同的偏置[2]。这样得到该卷积核在该位置的输出值。输出值被称为特征图，特征图的每个元素对应着卷积核在对应位置的输出值，而特征图的维度则由卷积核的数量决定。在卷积层中，每个卷积核的参数是共享的。

卷积神经网络（Convolutional Neural Networks，CNNs）在基础的线性卷积操作之外，还可以使用一些更为复杂的卷积操作，包括：

1）平铺卷积（Dilated Convolution）：在标准卷积操作中，卷积核的每个位置只覆盖输入图像中的一个像素。而在平铺卷积中，卷积核在输入图像上“跳过”一些像素，从而扩大了感受野（receptive field）的大小，使网络能够更好地捕捉到不同尺度的特征。

2）反卷积（Transposed Convolution）：反卷积可以将低维特征映射扩展到高维空间。在反卷积操作中，卷积核被用来对输出图像进行上采样，从而恢复出原来输入图像的空间分辨率。

3）扩张卷积（Deformable Convolution）：在标准卷积中，卷积核的形状和大小是固定的。而在扩张卷积中，卷积核的形状和大小可以根据输入图像的特征动态调整，从而使网络能够更好地适应不同形状和大小的目标。

这些复杂的卷积操作可以进一步提高卷积神经网络的性能，使其能够更好地应对各种视觉任务[16]。

1. 卷积层参数

卷积层输出特征图的尺寸是由卷积核大小、步长和填充这三个超参数共同决定的。具体来说：

1）卷积核大小（kernel size）：卷积核是卷积层的参数之一，决定了卷积核在输入图像上滑动时覆盖的区域大小。通常情况下，卷积核的大小是正方形或长方形，例如 3x3、5x5 等。

2）步长（stride）：卷积核每次在输入图像上滑动的步长决定了输出特征图的尺寸。通常情况下，步长是水平和垂直方向上相等的，例如步长为 1 或 2。

3）填充（padding）：在进行卷积操作时，在输入图像的边界处可以使用填充来保留边界信息。填充可以是 0 填充（zero padding）或者边界填充（edge padding），通常用于保持输出特征图与输入图像大小相同或者保证输出特征图的大小不会减小太多。

这些超参数的选择会影响卷积层输出特征图的大小和形状，进而影响神经网络的性能。因此，在设计卷积神经网络时，需要根据具体的任务和数据集对这些超参数进行调整和优化。

当卷积层设置越多时，特征图的尺寸会越小，这是由卷积核的交叉计算推出的结论。卷积核可以在输入张量上进行零填充、均值填充和重复边界值填充等填充操作。

填充依据其层数和目的可分为有效填充、半填充、全填充和任意填充这四类。

1）半填充（Half Padding）：半填充是一种在输入图像的边缘上填充一半像素的方式，以保持卷积操作前后输入和输出特征图的尺寸相同。半填充通常用于等长卷积，其卷积核大小为奇数，步长为 1。

2）全填充（Full Padding）：全填充是一种为输入图像的每个像素都添加足够多的填充像素的方式，以保证每个像素在卷积操作中都被访问相同的次数。全填充通常用于宽卷积，其卷积核大小和输入图像大小相同，步长为 1。

3）有效填充（Valid Padding）：有效填充是一种不在输入图像边缘上填充像素的方式，也就是没有填充。有效填充通常用于窄卷积，其卷积核大小小于输入图像的大小。

4）任意填充（Arbitrary Padding）：任意填充是一种人为设定的填充方式，其填充量介于有效填充和全填充之间，不常用。

总的来说，填充是为了在卷积操作中保留输入图像的边缘信息或者控制输出特征图的大小而进行的。不同的填充方式对应不同的应用场景和卷积操作。

其中，窄卷积（Narrow Convolution）是一种使用有效填充（valid padding）的卷积操作。在有效填充的卷积中，卷积核在输入图像上进行滑动时，不会在输入图像的边缘进行填充，而是只考虑卷积核覆盖到的部分像素。因此，有效填充的卷积操作不会改变输入图像的尺寸，输出特征图的尺寸会随着卷积核的大小和步长而变化。

在窄卷积中，卷积核的大小通常要小于输入图像的大小，因此输出特征图的尺寸会比输入图像的尺寸小。这种卷积操作可以在不改变输入图像尺寸的情况下，提取出输入图像中的局部特征信息，并对输入图像进行特征提取和降维，从而减少网络参数和计算量，提高网络的效率和泛化能力。窄卷积常用于卷积神经网络中的卷积层和池化层，以提取和压缩输入图像的特征信息[12]。窄卷积输出的特征图尺寸为：

 （2.4）

### 2.2.3 激活层

激活层（Activation Layer）是卷积神经网络中的一种重要组成部分，其主要作用是对卷积层的输出进行非线性变换，以提高网络的表达能力和泛化性能。激活函数是激活层的核心组成部分，它可以将输入信号映射到一个非线性空间中，从而使得神经网络能够学习更加复杂的特征表示。

常用的激活函数包括sigmoid函数、ReLU函数、LeakyReLU函数、ELU函数等。其中，ReLU函数是目前使用最广泛的激活函数，它在处理大量数据时具有很好的性能表现，并且能够缓解梯度消失问题。

除了提高模型的表达能力和泛化性能外，激活函数还具有一定的正则化效果，可以帮助防止模型出现过拟合的情况。因此，在卷积神经网络中，激活函数通常会被加入到每个卷积层之后。[3]。ReLU激活函数的公式是：

（2.5）

第层第个神经元的激活值可表示为[2]：

 （2.6）

此函数把所有小于零的数置为零，所有大于零的数置为它本身，这种激活方式可以提高计算效率同时进行一些非线性处理。ReLU激活函数图像如图2-5所示：

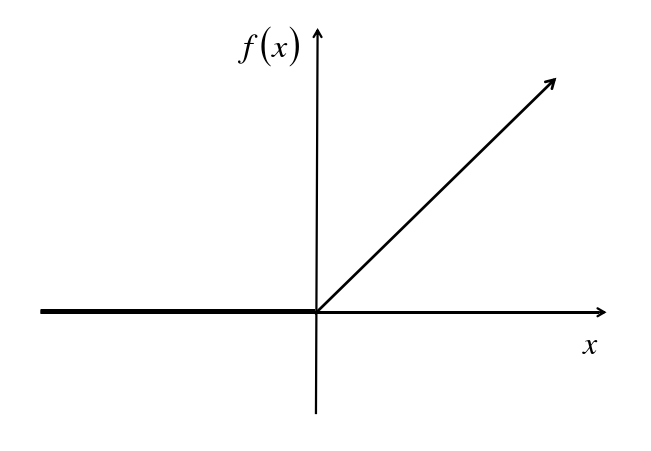


图2-5 ReLU激活函数

### 2.2.4 池化层

池化层（Pooling Layer）是卷积神经网络中的一种常见的操作，主要用于对卷积层输出的特征图进行降采样，以减小特征图的大小并增强特征的鲁棒性。池化操作通常通过滑动一个固定大小的池化窗口（Pooling Window）在特征图上进行，将池化窗口内的特征进行聚合并输出到下一层。

池化操作通常有两种方式：最大池化（Max Pooling）和平均池化（Average Pooling）。其中，最大池化操作选择池化窗口内的最大值作为输出，通常被用于提取特征的最显著部分。平均池化操作则选择池化窗口内的平均值作为输出，通常被用于提取特征的整体趋势。在最大池化中，池化操作会选择每个区域中的最大值作为该区域的输出，因此也被称为“最大池化”。在平均池化中，池化操作会选择每个区域中的平均值作为该区域的输出，因此也被称为“平均池化”。除了最大池化和平均池化外，还有一些其他类型的池化方法，如L2池化、随机池化等。池化操作有助于减小特征图的尺寸，降低网络的计算复杂度，同时可以提高特征的鲁棒性和不变性，减少过拟合的风险。但是需要注意的是，过多的池化操作会导致特征信息的损失，因此在实际应用中需要根据具体问题进行合理的池化策略设计[3]。

虽然池化层可以降低特征图的大小，但它并不是进行特征选择和信息过滤的唯一方法。在实际应用中，一些卷积神经网络也会使用其他的特征选择和信息过滤方法，例如Dropout等。

与卷积层类似，选取池化区域时，池化层由池化大小、步长和填充控制。若步长为1，则在卷积操作时每次走一格，为二则每次走两格，相对来说，若卷积步长越大，则最后的输出越小，步长越小，遍历越仔细，输出越慢，输出尺寸计算如下（加边可以捕捉到边缘信息）：

（2.7）

卷积需要训练参数，但池化不需要。

均值池化计算公式为：

 （2.8）

最大池化计算公式为：

 （2.9）

其中，n表示池化区域中神经元的数量；Ci表示池化区域C中的第个元素。

最大池化操作如图2-6所示，图2-6是以2为固定步长的2\*2的池化滤波器对4\*4的矩阵进行最大池化遍历，得到2\*2的矩阵输出。

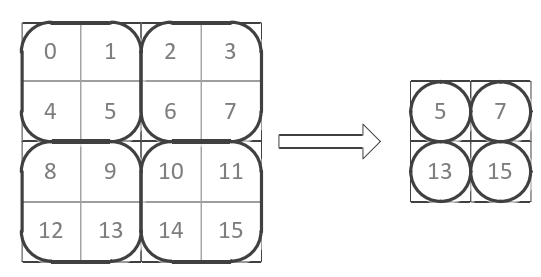


图2-6 最大池化操作

### 2.2.5 全连接层（fully-connected layer）

全连接层适用于做向量的连接转换，负责把大维度的向量转化维小维度的向量，只需要搭配相应的全连接层。全连接层的作用是将卷积层和池化层输出的特征图进行展平后，对提取的特征进行组合和转换，得到最终的分类结果。全连接层中的每个神经元与前一层中的每个神经元都有连接，因此全连接层中的参数量非常大，需要大量的计算资源和时间进行训练。为了避免这种问题，一些卷积神经网络中使用了全局平均池化层代替全连接层，将特征图每个通道的所有值取平均后直接输出。这种做法可以大大减少参数量和计算量，同时也不会影响网络的分类性能[7]。一个简单的两层全连接网络如图2-7所示：

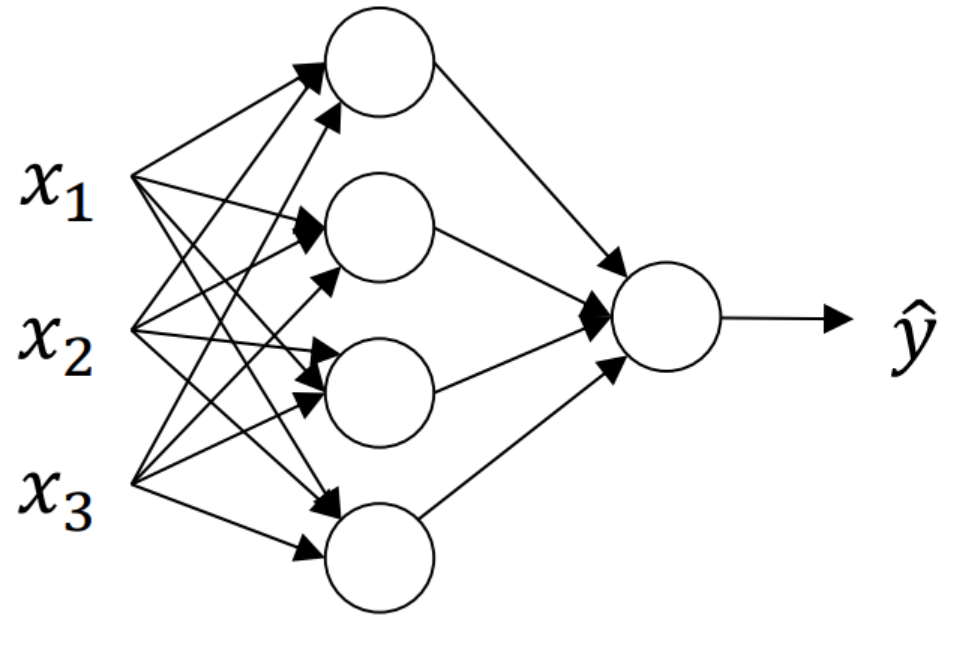


图2-7 一个简单的两层全连接网络

### 2.2.6 输出层

输出层的结构和工作原理与传统前馈神经网络中的输出层相同，主要用于对模型最终的输出进行处理和转换。输出层的神经元个数通常与分类问题中的类别数相同，每个神经元对应一个类别，并输出该样本属于该类别的概率值。在二分类问题中，输出层只需要一个神经元，其输出值为样本属于正类的概率值，另一个类别的概率值则为1减去正类概率值。不同类型的问题通常需要使用不同的输出层结构和激活函数来得到合适的输出结果。下面列举几种常见的问题和相应的输出层类型：

1.二分类问题：通常使用sigmoid激活函数，输出值范围在0到1之间，表示样本属于某一类别的概率。

2.多分类问题：通常使用softmax激活函数，输出值表示每个类别的概率，所有输出值的和为1。

3.回归问题：通常使用无激活函数的线性层作为输出层，输出层的值即为回归结果。

4.目标检测问题：通常使用边界框回归和分类两个输出层。边界框回归输出预测的边界框坐标，分类输出预测物体所属的类别。

5.语言模型问题：通常使用softmax激活函数，输出值表示下一个词的概率分布，常见的语言模型包括循环神经网络和Transformer模型。

值得注意的一点是，隐藏层和输出层节点的输出，都是由上一层的输出、对应的权重、偏置以及激活函数决定[2]。

## 2.3 本章总结

本章通过对卷积神经网络结构的概念介绍，对输入层、卷积层、池化层、全连接层、输出层这几种基本的层结构进行介绍、给出简单图示并阐述其计算原理，使我们对卷积神经网络的工作流程更加了解，此外，对卷积神经网络工作的细节了解更加透彻。

# 第三章 数据准备与预处理

## 3.1 数据集简介

手写英文字母的对时间的要求不是很高，无论是过去的手写字母还是现在的手写字母，都能够反映出写者的手写特点，于是，手写字母的数据集可用性比较强，需要解决的问题主要是尺寸和图片可用性问题，对于自己创建的数据集来说，大小不一且可用性不强，需要经过裁剪、筛选等操作，这样不仅极其麻烦而且浪费时间，最后得到的数据集数量还会大大减少。因此，我采用的是现成开源数据集The Chars 74K EnglishImg和EMNIST dataset，由于数据集数量的有限性，但多样的数据集能够使训练出的模型更好地训练，于是我采用的数据集是EMNIST和The Chars 74K的融合，使训练好的模型具有更好的泛化效果，数据集图片尺寸大小为28\*28。部分数据集展示如图3-1、3-2、3-3所示：

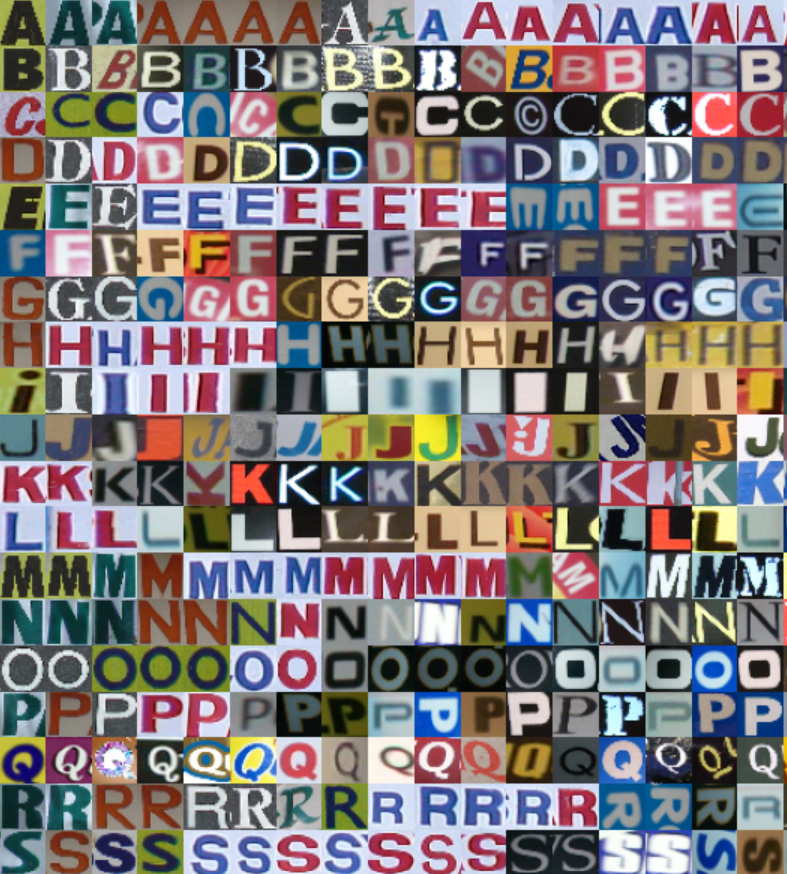


图3-1 数据集展示（1）

EMNIST数据集包含多个不同的数据类别，但由于此次只用到手写英文字母数据集，所以只需要emnist-letters部分，由于下载好的数据集为.idx3-ubyte格式，在pycharm中进行识别时并不能读取这种格式的文件，首先在代码中我们要先将其文件解析成为图片格式，再考虑训练和测试的问题。

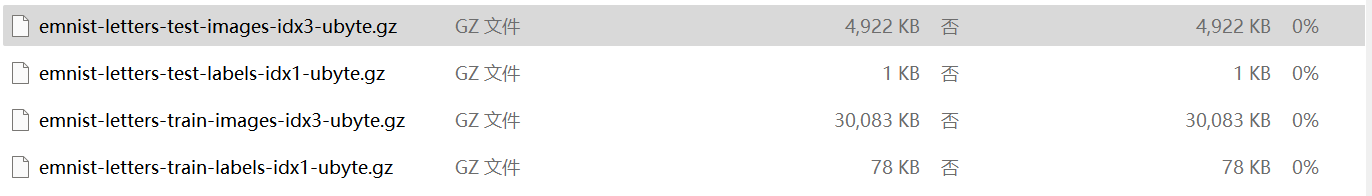


图3-2 数据集展示（2）



图3-3 数据集展示（3）

解析后的标签信息用分别用1-26来对应不区分大小写的英文字母A-Z，同时，需要将解析后的数据集标签根据一并解析的txt文本里的标签信息进行一一标注，并将不同的字母存放在不同的文件夹下对数据进行分类。数据的读取采用的是pytorch里的ImageFolder类。文件夹分类如图3-4所示：

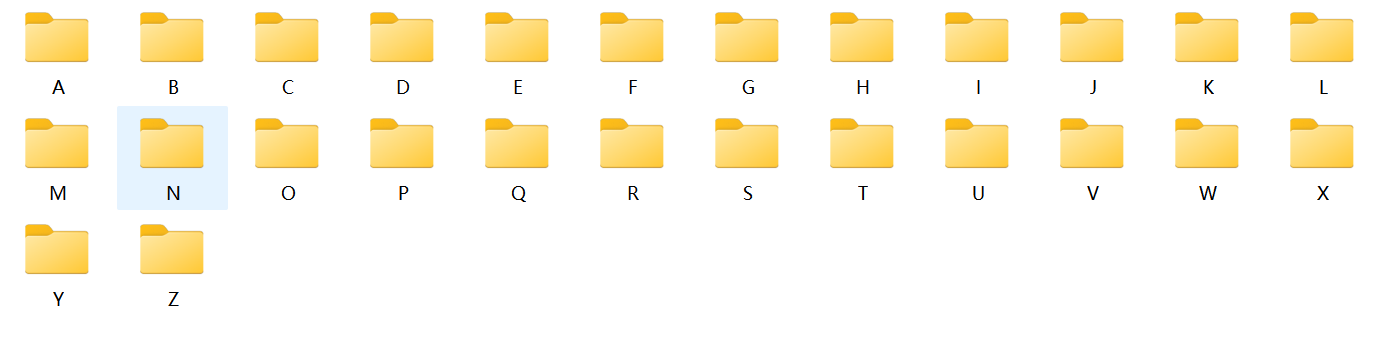


图3-4 文件夹分类

## 3.2 数据预处理

为了将此数据集运用到长短期记忆网络中进行模型训练和测试，首先需要对下载到的原始数据集进行预处理工作。首先定义超参数（训练批次、训练的最小规模和学习率），使用torchvision.datasets下的ImageFolder类构造数据集、使用shuffle来确定是否在构建批次时随机选取数据、设置PyTorch中数据读取的一个重要接口，能够将自定义的数据集封装成一个Batch Size大小的Tensor用于后面的训练，再将transform作为数据集的自定义预处理函数。

## 3.3 卷积神经网络对手写字母识别的设计

在卷积神经网络对手写字母的识别中，识别过程可分为：数据采集、建立模型、模型训练和模型测试。最主要的就是根据所输入数据来做出分类判断。本文由此对目前所公开的数据集模型进行展开，设计了一个可以识别所需识别图片的模型。本次训练字母识别使用的网络是基于LeNet-5结构的网络，采用卷积层+池化层+全连接层+全连接层+全连接层的网络结构。

训练流程图如图3-6所示：

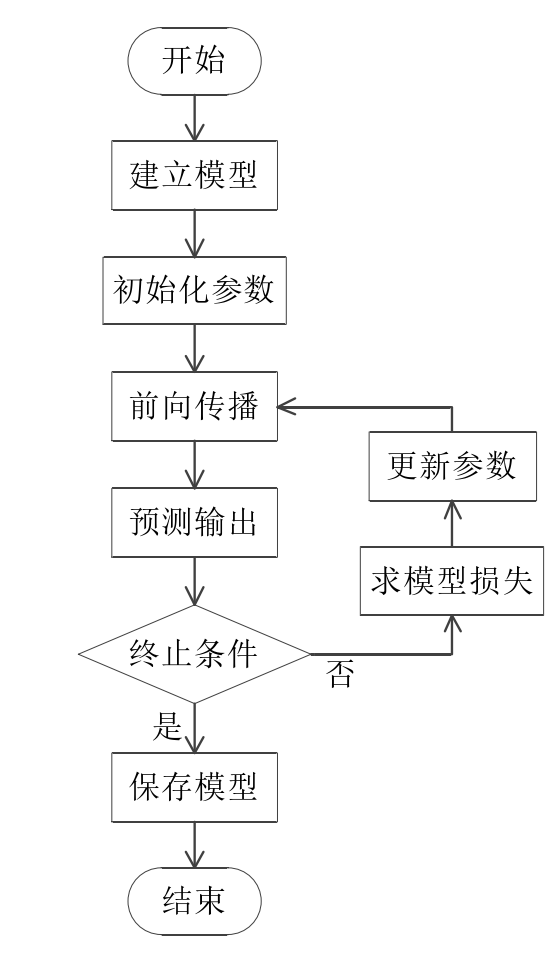


图3-5 训练流程图

测试流程图如图3-7所示：

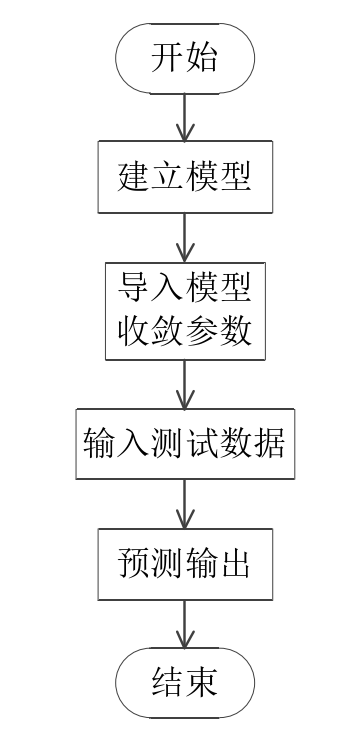


图 3-6 测试流程图

训练过程中，本文根据定义的Epoch进行循环，每次Epoch分为若干个step，首先需要将测试集传入网络测试，但网络不接受ImageFolder类，只接受Tensor类数据，我采用的解决办法是：将测试集也封装为DataLoader类，batch设置为整个测试集大小，在遍历时将图像与标签分别读取，这样就没有问题了。

可视化训练过程如图3-7所示：



图3-7 训练精度

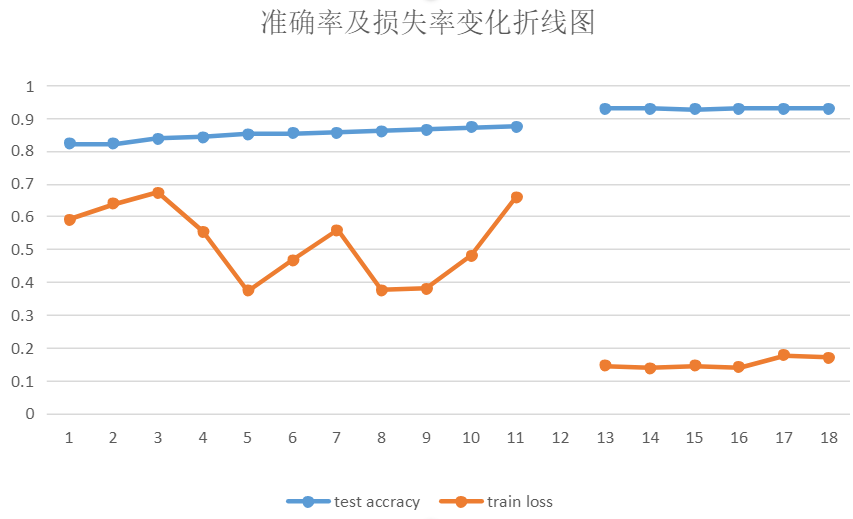


图3-8 训练第一次与训练多次精度折线图

## 3.4 本章总结

本章完成了对整个卷积神经网络对手写字母的识别算法的设计和总体思路的详细叙述，并且对卷积神经网络对手写字母的识别算法的总体流程和数据来源进行了介绍，其中包括对原始数据进行了数据预处理，对手写字母识别模型的设计等流程。最后通过不断地模型评估和模型优化，选择了性能较为优秀的模型进行对手写字母的识别。

# 第四章 系统设计与实现

## 4.1 系统运行环境

模型运行系统信息如图4-1所示：

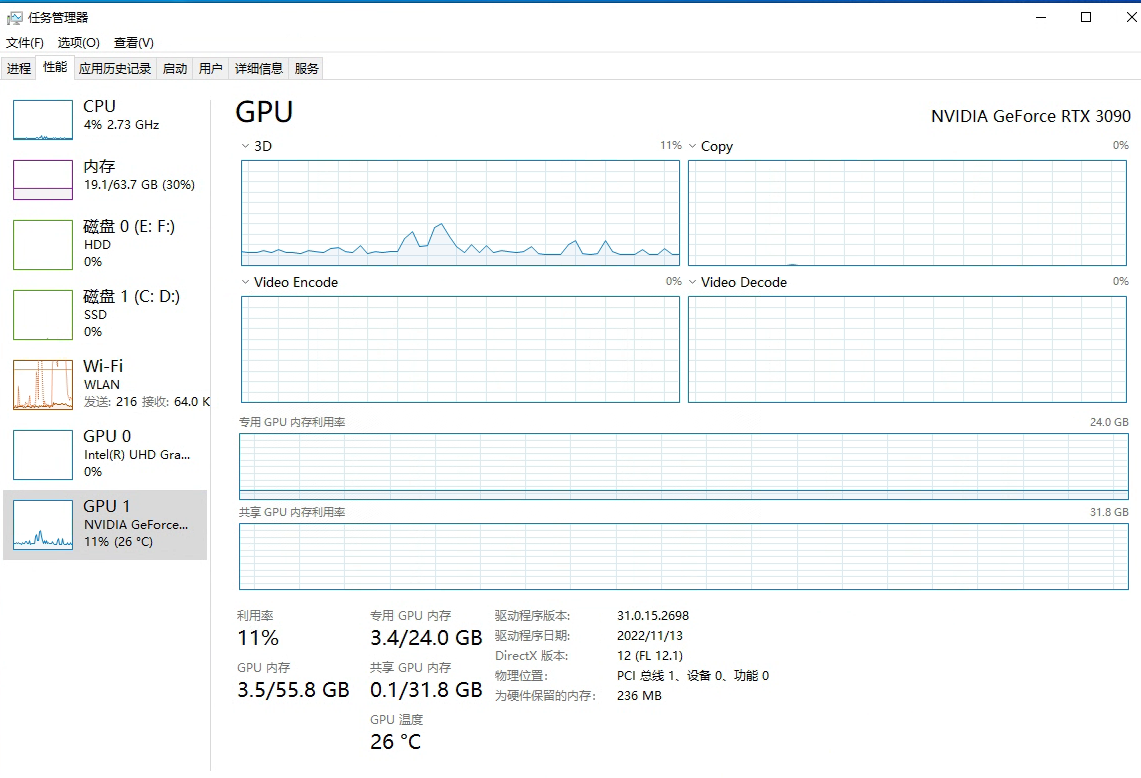


图4-1 系统信息

### **4.2 模型读取并测试自己的数据**

首先读取权重，输入待预测图像，提供自己的数据进行测试时，需进行转换为单通道、转换为28\*28尺寸、转换为张量、添加两个维度并把一个维度的灰度映射在（0，1）之间的操作。随后进行可视化部分并输出预测结果，此时需注意在输出预测结果时应将数字转换为对应的ASCAII字符。测试过程如下：

将处理好的测试图片放入文件中，此处我采用的是字母H，输入层如图4-2所示：

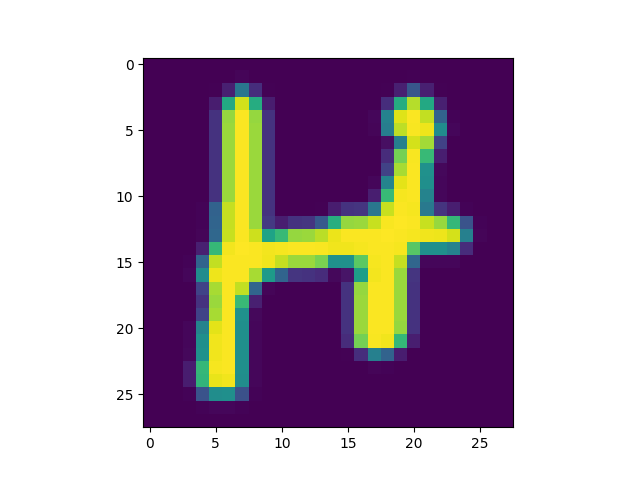


图4-2 输入（手写体）

卷积层经历卷积、池化、激活的过程，得到的特征图进行展平进入神经网络，得到预测的概率，通过反向传播的方式不断调整这些参数，以达到预测值和真实值最接近的情况，完成网络的训练。卷积层训练参数时，参数初始时随机的，但在经过训练之后计算机会帮我们找到能够体现它的特征的一些值，比如从左上到右下，卷积时我们发现，特征图保留了原始图像的特征，特征图在原始图像区域的值趋近于1，但在其他没有体现特征图位置的值都很低，把卷积核的特征进行了提取。卷积层测试如图4-3、图4-4所示：

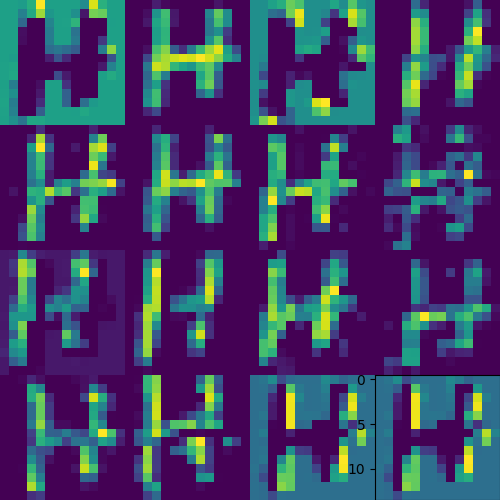


图4-3 卷积层1

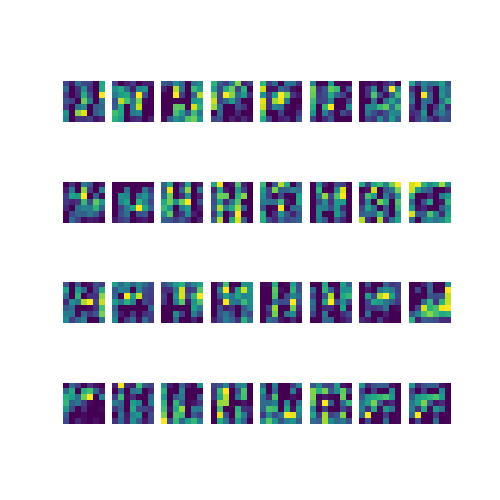


图4-4 卷积层2

全连接层适用于做向量的连接转换，负责把大维度的向量转化维小维度的向量，只需要搭配相应的全连接层。全连接层的作用是将卷积层和池化层输出的特征图进行展平后，对提取的特征进行组合和转换，得到最终的分类结果。全连接层中的每个神经元与前一层中的每个神经元都有连接，因此全连接层中的参数量非常大，需要大量的计算资源和时间进行训练。全连接层如图4-5所示：

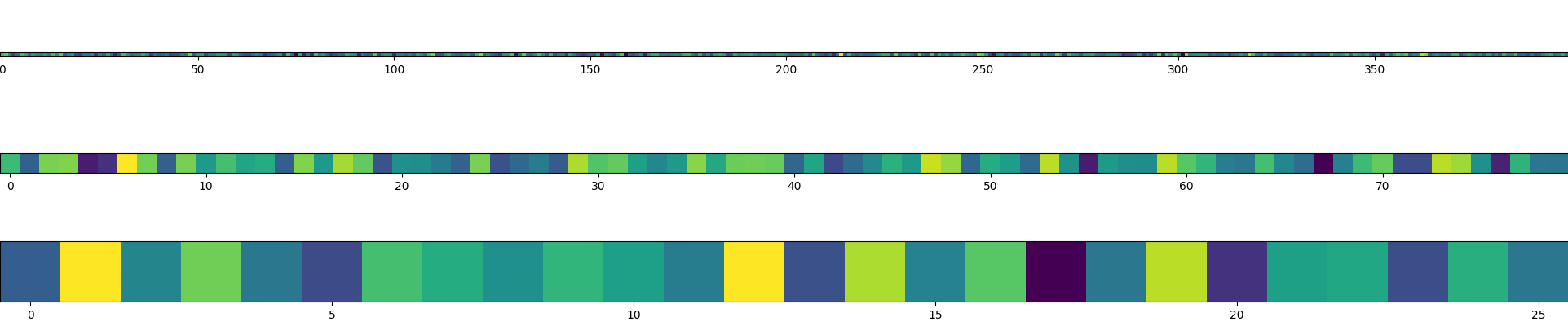


图4-5 全连接层

测试结果输出如图4-6、图4-7所示：

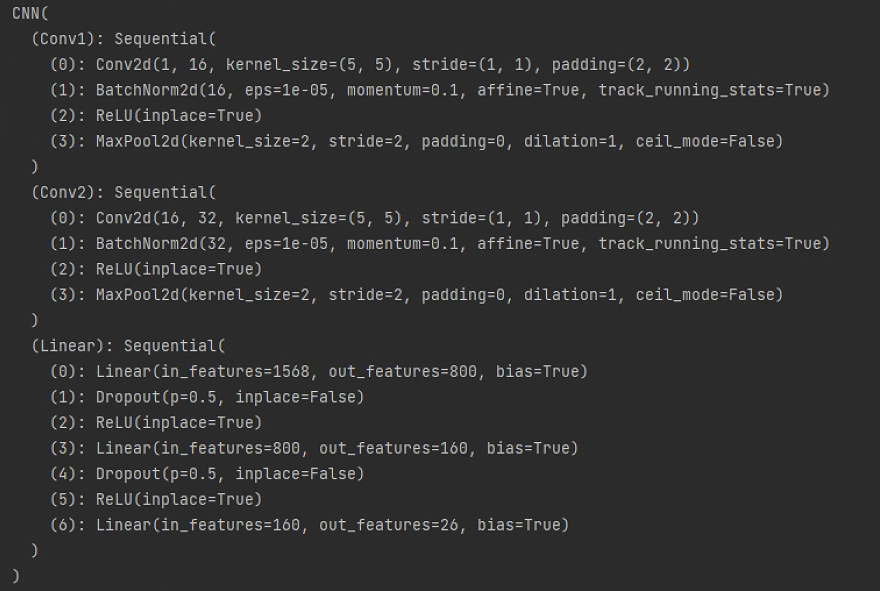


图4-6 CNN模型结构

训练完成后，识别的精确度达到93%，如图4-7所示，能够基本准确地识别字母，识别结果输出如图4-8所示：

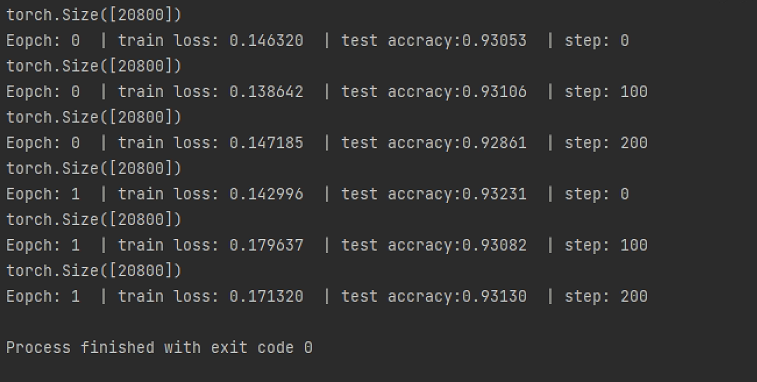


图4-7 识别精度

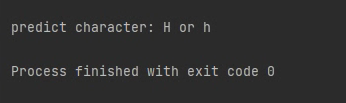


图4-7 输出识别结果

## 4.3 本章总结

本章分析了系统的硬件和软件运行环境，并介绍了卷积神经网络对手写英文字母识别的主要功能、操作流程，并对整个系统的运行进行了展示。

# 第五章 总结与展望

## 5.1 总结

在如今快速发展的信息化和智能化时代，字符识别作为模式识别中最重要的一个分支，受到了许多研究学者的研究。由于手写英文字母本身存在的复杂性、多样性和相似性，手写字母在识别的准确度的提升一直是我们需要不断研究改进的[12]。因此，本文在卷积神经网络的基础上，围绕数据集、模型结构和特征分析，对手写英文字母的识别进行研究，具体的内容可以分为以下三点：

1.对数据集数据的处理。通过对EMNIST和The Chars 74K数据集中数据的观察和分析发现，数据集中有些图像为异常数据，并不符合模型训练的标准，因此本文先对数据集进行了筛选、融合等预处理，此过程有利于模型的特征提取和收敛效果的提升。

2.网络搭建、模型结构的设计。本文结构类似于LeNet-5的结构，由卷积层、池化层和三层全连接层组成。

3.训练与模型保存。为了可视化网络在测试集上的效果，需要将测试集传入网络测试，本文使用将测试集也封装为类的方法，使遍历时将图像与标签分别读取。

最后进行模型读取和测试，相较于原始LeNet-5模型，本文所描述的模型对手写字母的识别有更高的灵活性和可行性。

## 5.2 展望

本文设计的以卷积神经网络为核心的手写英文字母模型，虽然能够进行对手写英文字母进行识别，但是仍然存在着些许弊端，有待提高：

1. 本系统的数据集数据量有限，面对更加多样化的手写字母，对其识别捉襟见肘，后续可以使用更多更广的数据集进行训练。

2. 该模型对手写英文字母的识别准确率仅93%左右，识别准确率还有待进一步提高。

# 参考文献

1. 洪奇峰,施伟斌,吴迪等.深度卷积神经网络模型发展综述 [J]. 软件导刊 ,2019(4):84 -88.
2. 王松.基于卷积神经网络的手写字符识别[D].西安：长安大学，2019：4-8.
3. 何国对.基于图卷积神经网络的文本分类研究[D].南宁：南宁师范大学，2021：5-23.
4. 范望，韩俊刚，苟凡，李帅[A].卷积神经网络识别汉字验证码.西安邮电大学研究生学院, 2018，54（3）.DOI：10.3778/j.issn.1002-8331.1706-0304
5. 孙红，陈强越.融合BERT词嵌入和注意力机制的中文文本分类[J/0L].小型微型计算机系统：1-6[2021-03-25].
6. 徐昭洪,刘宇,全吉成等.基于 VGG16 预编码的遥感图像建筑物语义分割 [J]. 科学技术与工程 ,2019,19(17) ： 250 -255.
7. 马建红，刘亚培，刘言东，陶永才，石磊，卫琳.CGGA:一种CNN与并行门控机制混合的文本分类模型[J].小型微型计算机系统，2021，42（03）：516-521.
8. 陈龙,郄小美,黄信静等.手写体数字字符识别算法仿真比较研究[J].实验室研究与探索 ,2017,36(1):9397.
9. 闫鹏，郑雪峰，朱建勇，肖赟泓.一种优化的K-NN文本分类算法[J].计算机科学，2009，36（10）：217-221.
10. 茹晓青 ，华国光,李丽宏等.基于形变卷积神经网络的手写体数字识别研究[J].微电子学与计算机 ，2019,36(4) :47-51.
11. 王建仁,马鑫,段刚龙等.边缘智能背景下的手写数字识别计算机应用 ,2019,39(12) ：358-3555.
12. 魏明俊，许道云 ，徐梦珂.带PCA 卷积的稀疏表示图像分类[J].算法计算机工程与应用，2017, 53(14)：155-160.
13. 赵耀.基于关联规则的文本分类研究[D].河北大学，2020.
14. Wu Y, Zhang K, Wu D, et al. Person re-identification by multi-scale feature representation learning with random batch feature mask[J]. IEEE Transactions on Cognitive and Developmental Systems, 2020.
15. Yuan C, Wu Y, Qin X, et al. An effective image classification method for shallow densely connected convolution networks through squeezing and splitting techniques[J]. Applied Intelligence, 2019, 49(10): 3570- 3586.
16. 高磊，洪奔奔，姚青岐. 基于卷积神经网络的随机手写数字识别[J]. 软件，2018，39（9）：74-78.
17. 苏金树，张博锋，徐昕.基于机器学习的文本分类技术研究进展[J].软件学报，2006（09）：1848-1859.