

兰州大学

毕 业 论 文

(本 科 生)

论文题目 (中文) 基于迁移学习的盲文字符图像识别方法研究

论文题目 (外文) The Research on Braille Character Image
Recognition Based on Transfer Learning

学 生 姓 名 石嘉

导师姓名、职称 高若宇 高级实验师

学生所属学院 信息科学与工程学院

专 业 教育技术学

年 级 2014

兰州大学教务处

诚信责任书

本人郑重声明：本人所呈交的毕业论文（设计），是在导师的指导下独立进行研究所取得的成果。毕业论文（设计）中凡引用他人已经发表或未发表的成果、数据、观点等，均已明确注明出处。除文中已经注明引用的内容外，不包含任何其他个人或集体已经发表或在网上发表的论文。

本声明的法律责任由本人承担。

论文作者签名：_____

日 期：_____

关于毕业论文（设计）使用授权的声明

本人在导师指导下所完成的论文及相关的职务作品，知识产权归属兰州大学。本人完全了解兰州大学有关保存、使用毕业论文的规定，同意学校保存或向国家有关部门或机构送交论文的纸质版和电子版，允许论文被查阅和借阅；本人授权兰州大学可以将本毕业论文的全部或部分内 容编入有关数据库进行检索，可以采用任何复制手段保存和汇编本毕业论文。本人离校后发表、使用毕业论文或与该论文直接相关的学术论文或成果时，第一署名单位仍然为兰州大学。

本毕业论文研究内容：

☐ 可以公开

☐ 不宜公开，已在学位办公室办理保密申请，解密后适用本授权书。

（请在以上选项内选择其中一项打“√”）

论文作者签名：_____

导师签名：_____

日 期：_____

日 期：_____

基于迁移学习的盲文字符图像识别方法研究

摘 要

盲文作为一种特殊的文字，其应用范围局限于视障人士。盲文字符的识别对不懂盲文的成年视障人士和明眼人来说具有重要意义。本论文针对盲文字符图像的识别展开研究，提出一种基于迁移学习的方法对盲文字符进行识别。与传统的深度学习算法不同，迁移学习避免了对深层次卷积神经网络进行长达数周的训练，也不需要高性能的 GPU 和庞大的数据集。

首先，本文创建了盲文字符数据集，其中包括盲文字符 0-9，共 10 种盲文字符，每一种包括 240 张盲文图像。然后用迁移学习的方法对这 10 种盲文字符图像进行识别。实验结果显示，该方法取得了 54.02% 的正确率。

该方法具有比较好的理论意义和实践价值，一方面填补了基于深度学习方法对盲文字符图像进行识别的空缺；另一方面促进了明眼人与视障人士更好地交流与沟通，对构建和谐社会具有较好意义。

关键词：盲文字符；图像识别；迁移学习；深度学习

The RESEARCH ON BRAILLE CHARACTER IMAGE RECOGNITION BASED ON TRANSFER LEARNING

Abstract

As a special kind of language, Braille is quite limited to visually impaired people. This paper studies the recognition of braille character images and proposes a method based on transfer learning to recognize braille characters. Unlike traditional deep learning algorithms, transfer learning avoids weeks of training for deep convolutional neural networks and does not require high-performance GPUs and large data sets.

First, this paper created a Braille character data set that included Braille characters 0-9, a total of 10 Braille characters, each containing 240 Braille images. Then use the transfer learning method to identify the 10 Braille character images. The experimental results show that the method achieved a correct rate of 54.02%.

This method has relatively good theoretical significance and practical value. On the one hand, it fills in the vacancies based on deep learning method to recognize Braille character images; On the other hand, it promotes better communication and communication between discerning people and visually impaired people and has a good meaning in building a harmonious society.

Key words: Braille characters; Image Recognition; Transfer learning; Deep learning

目 录

摘 要.....	I
Abstract.....	II
第一章 绪论.....	1
1.1 研究背景.....	1
1.2 研究现状.....	2
1.3 主要研究内容.....	5
第二章 盲文字符图像识别模型.....	6
2.1 前期工作.....	6
2.2 结构.....	6
2.3 训练方法.....	9
第三章 盲文图像数据集.....	10
3.1 数据集设计.....	10
3.2 数据集开发.....	11
第四章 实验方法及结果.....	12
4.1 实验方法.....	12
4.2 实验环境.....	13
4.3 实验过程.....	13
4.4 实验结果及分析.....	16
第五章 总结与展望.....	18
参考文献.....	19
附 录.....	21
论文（设计）成绩.....	25

第一章 绪论

1.1 研究背景

根据第二次全国残疾人抽样调查数据,据推算 2006 年全国各类残疾人总数为 8296 万人。其中视力残疾人口为 1233 万人,约占残疾人总人口数的 14.86%。视力残疾的问题给视障人士的个人生活和学习工作带来诸多不便和困难。盲文作为一种语言工具,是视障人士日常生活、学习文化知识、获取信息的重要手段,对其知识习得、认知发展和社会交往有重要作用。

但是,盲文作为一种特殊的文字符号系统,它的使用范围一直以来只是局限于视力残疾人士。根据统计数据表明,懂盲文的视障人士占视力残疾总人数的 7%,这就说明大多数的视障人士是无法阅读盲文的。另外,盲聋哑学校的教师,阅读现行盲文的水平不高。只有不到半数的老师认为自己能很熟练或者是比较熟练地阅读现行盲文。教师主要在制作试卷讲义和批改作业时阅读和刻印盲文,在布置作业的过程中也会有所涉及,但使用比例并不是很高^[1]。

由此可以得出,通过一种方法把盲文转化为文字,具有重要的理论价值和现实意义。一方面,可以帮助没有系统学习过盲文的成年视障人士阅读盲文;另一方面,也可以辅助明眼人更好地与视障人士交流,如帮助盲聋哑学校的教师提高批改作业的效率、帮助盲文工作者将盲文资料数字化等。

2012 年, Krizhevsky et al.^[2] 凭借 Alex-Net 在 ImageNet 的大规模视觉识别挑战赛中以超越第二名近 11% 的准确率夺冠,使卷积神经网络成为了图像识别领域的焦点。随着人们对卷积深度网络的不断改进,这种方法在解决图像识别和目标检测问题上显示出优越的结果。

本文提出一种基于迁移学习的图像识别方法,对盲文字符进行识别。与传统的深度学习方法相比较,迁移学习并不用训练整个神经网络。该方法避免了对神经网络的长时间训练和对计算机内存的高性能要求,同时也能实现较高的识别率。实验证明,与传统的图像识别方法如 SVM、kNN、BPNN¹ 相比较,也具有比较明显的优势。

¹ SVM 是 Support Vector Machine 的缩写,译为支持向量机; kNN 是 k-NearestNeighbor 的缩写,译为最邻近分类算法; BPNN 即 Back Propagation Neural Network,译为 BP 神经网络或反向传播神经网络。

1.2 研究现状

1.2.1 盲文字符识别

盲文是由法国人路易斯·布莱尔于 1824 年发明改造而来，现已广泛作为视障人士交流学习的工具。在中国，大约有 1263 万视力残疾人士，但是其中只有 80 多万可以阅读盲文。所以说，对盲文进行识别对成年视障人士和进行盲文编辑工作的明眼人具有重要意义。

盲文由 6 点组合而成，分为三行两列，共有 63 种（不包括空白）不同的组合。虽然盲文的组合仅有 63 种，但是配合相应的规则可以表达出丰富的语义。值得注意的是，不同国家对盲文的使用规则是不尽相同的。有的国家甚至有几种不同编码规则同时使用，如中国的现行盲文和汉语双拼盲文。

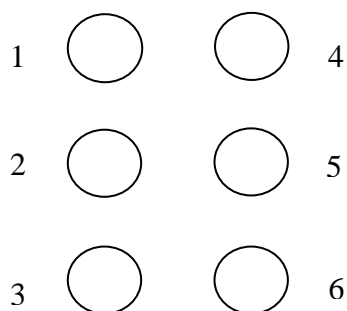


图 1.1 盲文字符单元

盲文字符的识别具有重要意义，人们已经用不同的方法对盲文字符识别展开研究。1988 年，Dubus et al.^[3]利用基本的图像处理算法，可以将刻印的盲文转化成墨印的文本。但是这种方法，很难准确地确定盲文图像的阈值。1994 年，Mennens et al.^[4]在一定程度上解决了图像中的扭曲和阴影问题。但是该方法需要用到专业的扫描仪，并且无法解决点状布局中的扭曲问题。Ritchings et al.^[5]使用平板扫描仪扫描盲文文档，运用字符分割和识别来定位盲文单元。该方法使用固定网格去解决字符的定位问题，这种方法无法应对图像噪点、图像伸缩和文档破旧等问题。1995 年，Hentzschel et al.^[6]通过不同光照条件获得两种盲文图像阴影，通过分析阴影的差异来判断该位置是否有盲点。该方法可以解决图像中的噪点问题和渐变阴影问题。但是该方法无法自动识别盲点之间的距离。1997 年，Oyama et al.^[7]将双面盲文图像曝光于 LED 灯下，分析盲点的高光与阴影图案可以区分凸点和凹点。1999 年，Ng C et al.^[8]通过边缘增强、二值化、边缘检测和特征提取等算法，可以将双面盲文进行分割，然后把盲点分成单元格进行模式匹配，并

把它转化为英文。该方法获得比较不错的识别效果，但是没有对网格变形问题进行进一步解释。

2001 年，论文^[9,10]使用一种手提式的扫描仪获取盲文图像，并且使用 DSP 算法对图像进行识别。对于该方法，扫描仪可以获得较好的图像质量，缺点是扫描仪很难保持垂直于盲文线，而且往往偏离理想的方向。2004 年，Antonacopoulos et al.^[11]提出一种相对灵活的网格，允许不同行之间的字符位置的变化。该方法为每一行盲文构建网格，然后将单个网格组合起来形成整个页面。2005 年，Néstor et al.^[12]发明了一款基于动态阈值处理、自适应盲文网格和文本到语音技术的盲文阅读软件 BrailLector。2006 年，Namba et al.^[13]提出一种基于细胞神经网络的方法对盲文图像进行识别，该方法通过对盲文图像进行预处理、特征提取，然后进行识别。最后，该方法取得了 87.9% 的识别率。2007 年，Zhang et al.^[14]介绍了一种新的系统，可识别带嵌入式相机的手机拍摄的照片中的盲文字符。

2010 年，Al-Salman et al.^[15]应用基于 Beta 分布的分割方法来估计阈值。Li et al.^[16]介绍了一种使用 SVM¹从图像识别盲文字符的系统。该方法可以获得低于 5% 的错误率。并且利用 SVM 分类器，可以比传统的图像处理算法更好地处理背景噪声。但是该方法对每页盲文进行处理大约需要 20 分钟时间，这是相当耗时的。2012 年，Alsalman et al.^[17]提出了一种盲文图像分割的新方法，使用类间方差和 Gamma 分布混合来解决盲文图像的非对称直方图问题。运用该方法寻找最佳估计阈值，然后利用这个阈值分割盲文图片。

通过以上文献可以发现，传统的盲文字符识别过程基本分为两步：第一步，对图像进行预处理，分割图像为单个盲文字符；第二步，通过设置的阈值判断网格相应的位置是否有盲点，进行模式匹配转化为文字。可以看出，传统的方法对盲文字符进行识别需要人工对阈值进行判断，如果设置不当可能会出现较差的识别结果。通过深度学习的方法，本论文可以将该过程自动化。通过训练数据找到盲文字符的特征，然后利用该模型对盲文字符进行识别。这可能会是一个比较好的识别方法。

1.2.2 迁移学习

在机器学习和数据挖掘算法中有一个主要的假设，训练数据和测试数据必须在一个特征空间并且同分布，然而在现实应用中，这个假设可能是不成立的^[18]。在这种情况下，如果本文使用迁移学习的方法，可能会得到比较好的效果。

¹ Support Vector Machine，简称 SVM，中文译为支持向量机

人们发现，半监督分类^[19,20,21,22]提出通过训练大部分未标记的数据和小部分标记的数据，如果标记的数据太少，可能无法训练出一个好的分类器。从 1995 年起，人们对迁移学习（transfer learning）的关注越来越多，但是它们有着不同的名字。如 learning to learn，终身学习，知识迁移，感应迁移，多任务学习等^[23]。1997 年，Caruana et al.^[24]提出一个多任务学习框架，该方法在同时对不同任务进行学习。2005 年，美国国防高级研究计划局给迁移学习的定义是：一个能够识别并将以前的任务中学到的知识和技能应用于新任务的系统。与多任务学习不同的是，该定义明确了迁移学习关注的只是目标任务，而不是同时学习源任务和目标任务。

图 1.2 反映了传统的机器学习与迁移学习的不同之处^[18]。我们可以看到，传统的机器学习技术重新开始学习各个任务；而迁移学习是，在后面的任务有较少的高质量训练数据时，把先前任务学习到的知识迁移到后面的任务。

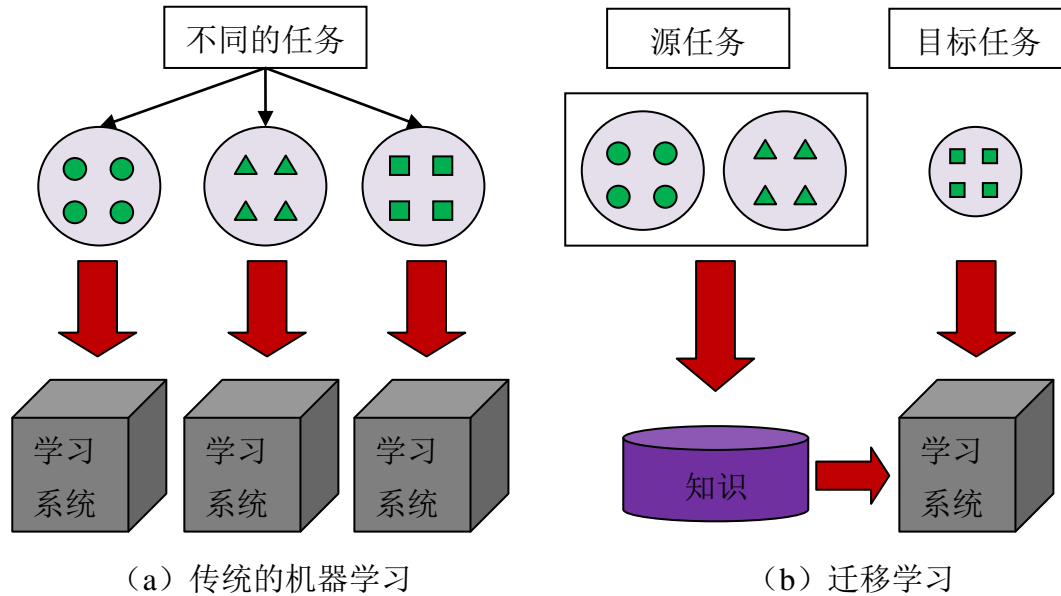


图 1.2 传统的机器学习与迁移学习的差异

最近，迁移学习也成功应用于解决实际问题。2013 年，Donahue et al.^[25]运用迁移学习的方法，在相应的数据集获得最好的识别率。2016 年，Long et al.^[26]提出一种深度网络域适应¹的新方法，可以同时从源域中的标记数据和目标域中的未标记数据中学习自适应分类器和可转换特征。该方法在标准的域适应数据集上成为最佳。

¹ 域适应，domain adaptation，根据^[18]的分类，域适应是属于迁移学习的一种，标记的数据仅限于源域。

1.3 主要研究内容

本文主要的研究内容为，基于迁移学习的方法对盲文字符图像进行识别。本论文的研究目标见表 1.1。

表 1.1 研究目标

类型	标准
盲文字符图像识别	盲文字符-ASCII
识别字符数	10
准确率	尽可能高

本文首先通过深度学习模型对源数据集¹进行预训练，然后对该模型进行调整，利用调整后的模型对目标数据集²进行再训练，进而调整分类器的参数，最后预测盲文字符。迁移学习可以把解决源问题的模型利用到解决目标问题上，这是因为该深度学习模型已经可以很好地进行图像的特征提取。

图 1.3 描述了整体的实验流程，主要包括以下几部分工作：

- (1) 构建盲文字符数据集。
- (2) 选择迁移学习的模型。
- (3) 利用迁移学习模型对该数据集进行再训练。
- (4) 盲文字符的预测。

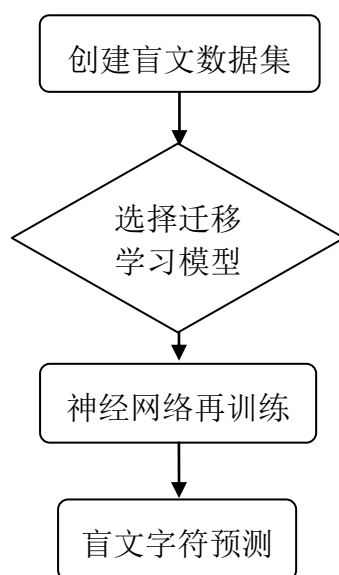


图 1.3 迁移学习模型

¹ 源数据集为现有的图像分类数据集，如 MNIST 数据集，ImageNet 数据集等。

² 目标数据集为要解决的图像分类数据集，这里为盲文字符图像数据集。

第二章 盲文字符图像识别模型

2.1 前期工作

现代图像处理的神经网络模型有数以百万计的参数。从零开始训练需要大量标记的训练数据和庞大的计算能力，即使用 GPU 进行训练，训练时长也可长达数百小时。迁移学习可以简化这一过程，它通过获取一个已经在相关任务训练过的模型，然后新模型中重用该模型。

虽然这种方法没有重新训练整个模型好，但是对于拥有中等数量训练数据的问题来说却是可行的。这种方法在笔记本上使用 CPU 运算需要 70 分钟左右（本论文），使用 GPU 加速运算会大大缩短该时间。

论文^[25]分析了在半监督的多任务框架中的深度特征的使用。经过训练的 Alex-Net 模型可以学习具有足够代表能力和泛化能力的特征，以使用简单线性分类器执行语义视觉区分任务。这验证了迁移学习是具有可行性的。

2.2 结构

本论文利用迁移学习解决盲文字符识别时，所采用的深度学习模型是谷歌的 Inception-v3。论文^[27]⁹实验表示，即使在低至 79×79 分辨率的图像上，该模型也能获得比较好的分类结果，这正是本文采用这个模型的原因。Inception-v3 模型由 Inception-v1 发展而来，本小节解释了该模型的发展过程以及描述了该模型的框架。

2.2.1 Inception-v1

提高深度神经网络性能的最直接的办法是扩大规模，一方面是增加深度网络的深度，另一方面可以增加每一层的宽度。但是这种方法也会带来一些问题，首先是若在训练数据有限的情况下，扩大规模通常意味着增加参数，而过多的参数会使模型过拟合（over-fitting）。其次模型扩大，可能会导致梯度弥散，这会使计算资源低效利用。2014 年，Szegedy et al.^[28]提出解决这两个问题的关键是，如何用密集成分近似地替代最优的局部稀疏结构。一方面该方法通过在中间层增加两个辅助分类器来解决扩大深度带来的梯度弥散现象，另一方面该方面并联了不

同大小的卷积核来扩大模型的宽度，同时还运用了 1×1 的卷积核减少特征图（feature map）的厚度，进行降维之后的卷积运算量将会线性减少。

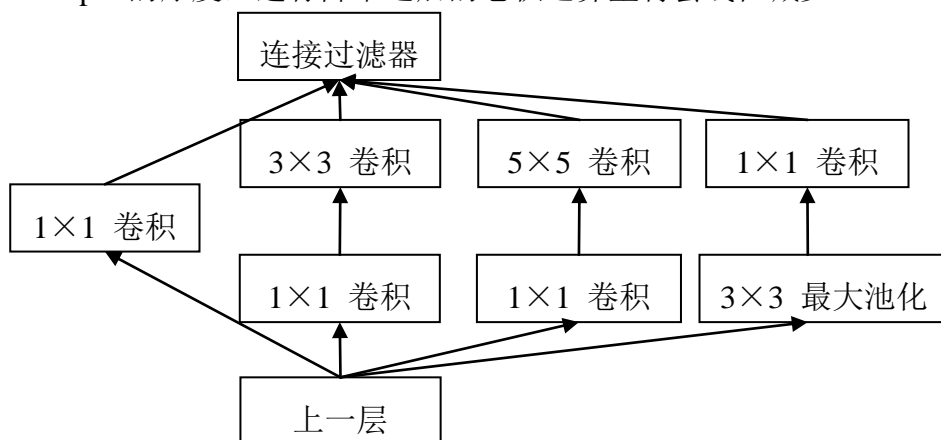


图 2.1 Inception-v1 模块

2.2.2 Inception-v2

2015 年 3 月，Ioffe et al.^[29]提出了一种可以加速深度网络训练的方法批量标准化（Batch Normalization），它可以减少内部神经元的数据分布发生变化，使每一层的输出规范化。批量标准化可以让我们使用更高的学习速率以及降低我们对初始化的干预程度。

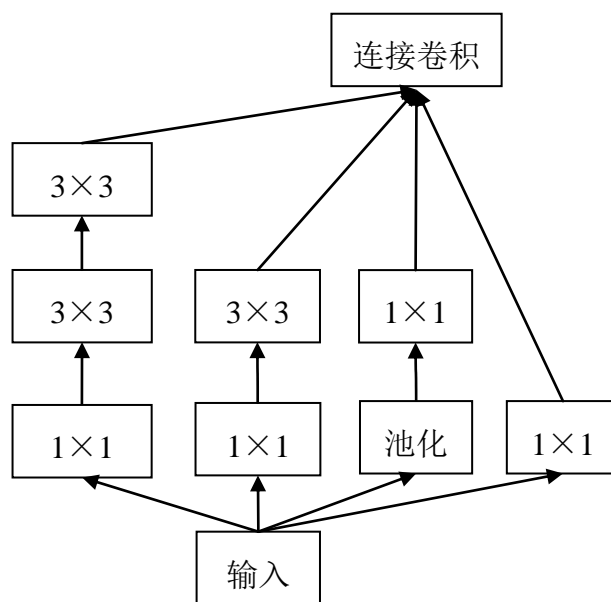


图 2.2 5×5 的卷积被 3×3 替换

2.2.3 Inception-v3

2015 年 12 月, 论文^[27]提出通过分解卷积和积极正则化来提高增加的运算的效率, 从而扩大网络的规模。一方面, 该方法将卷积核替换为更小的叠加的卷积核(如将 5×5 的卷积核替换为两个 3×3 的卷积核), 另一方面该方法尝试把 $n \times n$ 的卷积核分解为非对称的卷积核叠加起来(如 $1 \times n, n \times 1$)以提高运算效率。

2.2.4 神经网络结构

Inception-v3 的网络结构见表 2-1。值得注意的是, Inception-v3 把传统 7×7 的卷积替换为 3 个 3×3 的卷积, 这样做的目的是为了减少较大的卷积核带来的运算量的增加。对于这个神经网络的 Inception 部分, 有 3 个传统的 Inception 接收 $35 \times 35 \times 288$ 的输入, 输出 $17 \times 17 \times 768$ 。接下来连接了 5 个分解后的 Inception, 如图 0-1, 输出 $8 \times 8 \times 1280$ 。然后连接了两个 Inception, 如图 0-2, 输出 $8 \times 8 \times 2480$ 。

表 2.1 Inception-v3 的网络结构

类型	过滤器大小/步长 或标记	输入大小
卷积	$3 \times 3/2$	$299 \times 299 \times 3$
卷积	$3 \times 3/1$	$149 \times 149 \times 32$
卷积 填充	$3 \times 3/1$	$147 \times 147 \times 32$
池化	$3 \times 3/2$	$147 \times 147 \times 64$
卷积	$3 \times 3/1$	$73 \times 73 \times 64$
卷积	$3 \times 3/2$	$71 \times 71 \times 80$
卷积	$3 \times 3/1$	$35 \times 35 \times 192$
3×Inception	如图 2.1	$35 \times 35 \times 288$
5×Inception	如图 0.1 ¹	$17 \times 17 \times 768$
2×Inception	如图 0.2 ²	$8 \times 8 \times 1280$
池化	8×8	$8 \times 8 \times 2480$
线性	未归一化的概率	$1 \times 1 \times 2480$
Softmax	分类器	$1 \times 1 \times 1000$

¹ 见附录。

² 见附录。

2.3 训练方法

图 2.3 反映了迁移学习的具体过程，该过程分为几个阶段：

第一阶段解析磁盘上的所有图像，计算并存储每个图像的瓶颈值¹。这个倒数第二层已经被训练输出一组值，这些值足以让分类器用来区分它被要求识别的所有类。这个经过本文训练过的最后一层能够在新问题上工作的原因是，它已经训练地足够强大，已经可以区分 ImageNet 上的 1000 种类，同样可以区分新的分类问题。由于在训练过程中会多次用到每张图像，所以将图像的瓶颈值保存到磁盘上会是一个好的选择。

第二个阶段通过瓶颈层的输出训练模型的最后一层。与 CNN 训练过程一样，我们会看到一系列输出，其中包括训练准确度²、验证准确度³和交叉熵⁴。代码默认会训练 4000 步，每一步需要从训练集中随机挑选 100 张图像。

第三个阶段通过反向传播（back-propagation）过程调整最后一层的权值，以使损失值达到最小化。

第四个阶段输出最终的测试准确度⁵。该准确度可以在一定程度上反映该模型在本课题的应用效果如何。

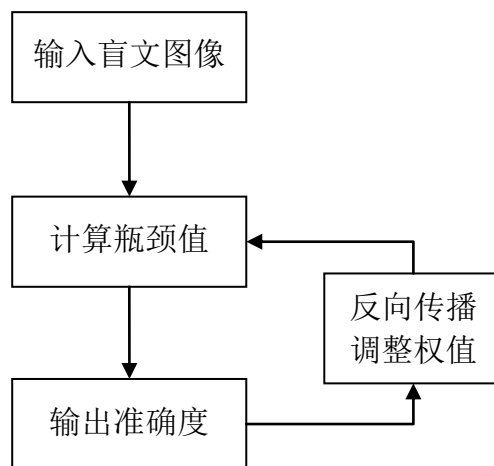


图 2.3 迁移学习具体过程

¹ '瓶颈'（Bottleneck）是一个非正式术语，我们经常称最后一个输出层之前的那一层为瓶颈层。

² 训练准确度（Test accuracy）是在当前训练批次（batch）中，标有正确类的图像占的百分比。

³ 验证准确度（Validation accuracy）是从不是当前批次的集合中随机选择一些图像，测试其准确度。

⁴ 交叉熵（Cross entropy）是一种损失函数（Loss function），可以让我们看到学习过程的进展情况。常见的损失函数有 Multiclass SVM 损失函数，Softmax 损失函数等。

⁵ 测试集与训练集和验证集的交集为空。

第三章 盲文图像数据集

3.1 数据集设计

盲文字符共有 64 种，关于盲文数据集，本论文一方面让每种字符的图像尽可能多，因为这样可以使神经网络训练的更好，这里本文设计了 240 种。另一方面，本文模拟真实的拍摄情景，在不同时间使用不同设备拍摄不同角度的盲文字符图像，这样可以使系统更具有鲁棒性¹。具体的数据集设计方案见表 3.1。

本论文在 6 种不同背景的纸上手工打印了盲文字符，每页纸上各包含 64 种盲文字符。其中第一种纸张为标准的盲文打印纸，其余 5 种为本文挑选的不同背景的背景的纸张。本文之所以这样选择是想模拟真实的打印环境，盲校学生可能使用杂志纸作为学习用纸，这种纸张的背景往往具有复杂性。本文期望通过选择不同背景的纸张以获得比较好的识别效果。纸张样式如图 3.1。



图 3.1 纸张样式

本论文选择在室内和室外进行拍摄，在室外分别在阴影下和阳光下进行拍摄，在室内选择在普通灯管照明状态下和自然光状态下进行拍摄。本文拍摄的角度分为正面、上倾约 45° 、下倾约 45° 、左倾约 45° 和右倾约 45° 。

本文选择的拍摄设备为佳能 550D 相机和普通智能手机。

表 3.1 数据集设计

拍摄方案	种类数
纸张	6
设备	2
角度	5
室内/室外	2
光线条件	2

¹ 鲁棒是 Robust 的音译，也就是健壮和强壮的意思，在图像识别问题中指的是健壮性。

3.2 数据集开发

按照数据集的设计方法，作者和付鹏飞于3月26日在兰州市盲聋哑学校手打了6页盲文字符，每页包括63种盲文字符。图3.2是在盲文图书馆手工打印盲文。



图 3.2 手打盲文

本论文在4月至5月期间，分4次采集了盲文字符图像。采集时间分别为4月11日，4月16日，4月19日和5月11日。拍摄完成后，用Photoshop软件的切片工具对图像进行切割，把整页的盲文字符图像分割成单个的盲文字符图像。如图3.3。



图 3.3 切割后的盲文字符

本文共切割了240页盲文字符图像，每一页切割10种盲文字符，至此整个盲文字符数据集已经开发完成，其包括2400张盲文字符图像。

按照迁移学习训练过程的需要，本文对这2400张盲文字符图像重新进行了组织和命名，详见附录。组织方式为，按照字符类型创建文件夹，如盲文数字0对应ASCII码48，本文则创建一个名为48的文件夹，里面存放240种切割好的盲文数字0的图像。本文还对这240种盲文进行了命名，格式为文件夹名_编号，编号从0-239。

第四章 实验方法及结果

4.1 实验方法

在实验过程中，本论文重新训练 Softmax 层，使它可以识别盲文字符。通过反向传播调整最后一层的参数。在实验最后，输入测试集，输出测试准确度。由第三章可以知道，每种盲文字符共有 240 种。

本文选择训练集、验证集和测试集的比例分别为 80%、10% 和 10%。学习率初始化为 0.01，训练步长为 30000 步。每一批次的训练集中有 100 张盲文字符图像，验证集有 100 张盲文字符图像。本文在最后输出测试的准确率，实验结果的准确率取 5 次训练结果的均值。这 5 次训练本文均采用相同的模型，且不对输入图像进行预处理。¹

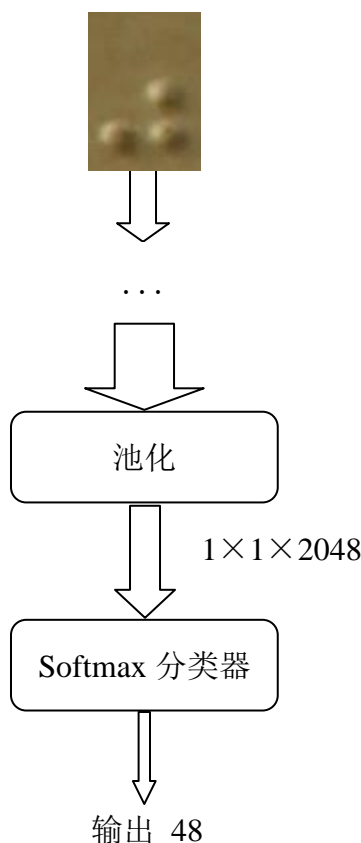


图 4.1 盲文字符识别

¹ 代码网址 <https://github.com/shijiatongxue/Braille-Characters-Recognition-Based-on-Transfer-Learning>

4.2 实验环境

本实验在 Windows7 操作系统下搭载了 Python，并通过 pip 方式安装了深度学习的开源库 TensorFlow¹。具体的实验环境配置见表 4.1。

表 4.1 实验环境

类型	具体配置
硬件	Intel Core i7, 4G 内存
	NVIDIA GeForce 840M, 2G 显存
软件	Windows 7
	Python3.6
	TensorFlow1.7.0

4.3 实验过程

本论文在盲文字符数据集上进行了实验，盲文字符数据集为 10 个单独的文件夹，里面各包含有一种盲文字符，数量为 240 个。这里本文选择的是从盲文数字 0-9，ASCII 码值为 48-57。这些盲文字符的长和宽大小在 40 像素至 200 像素之间。本文还用传统的图像分类算法 SVM、kNN、BPNN 对相同数据集进行了识别。^[30]

迁移学习的关键代码见下。

```
python retrain.py --image_dir braille --saved_model_dir /tmp/saved_models/
```

这部分的代码的功能是启动训练过程，--image_dir 参数选择盲文数据集所在的位置，--saved_model_dir 参数选择存储模型的位置。

```
python label_image.py
--graph=/tmp/output_graph.pb
--labels=/tmp/output_labels.txt
--input_layer=Mul
--output_layer=final_result
```

¹ TensorFlow 是由 Jeff Dean 领头的谷歌大脑团队基于谷歌内部第一代深度学习系统 DistBelief 改进而来的通用计算框架。由于 TensorFlow 对深度学习算法的灵活支持以及它较好的发展趋势，本论文选择 TensorFlow 作为盲文字符识别的计算框架。

```
--input_mean=128
```

```
--input_std=128
```

```
--image=images/test.jpg
```

训练完成后，通过上面的代码预测盲文字符。--graph 参数选择模型图的存储位置，--labels 参数存储盲文字符标签，它通过读取盲文字符图像所在文件夹的名称获取，--image 参数选择预测图像的位置。

训练部分的关键函数和代码见下。

```
prepare_file_system()
```

```
create_model_info()
```

```
if not model_info:
```

```
    tf.logging.error('Did not recognize architecture flag')
```

```
    return -1
```

```
create_image_lists()
```

```
class_count = len(image_lists.keys())
```

```
if class_count == 0:
```

```
    tf.logging.error('No valid folders of images found at ' + FLAGS.image_dir)
```

```
    return -1
```

```
if class_count == 1:
```

```
    tf.logging.error('Only one valid folder of images found at ' +
```

```
                      FLAGS.image_dir +
```

```
                      ' - multiple classes are needed for classification.')
```

```
    return -1
```

```
should_distort_images()
```

```
maybe_download_and_extract()
```

```
tf.global_variables_initializer()
```

```
for i in range(FLAGS.how_many_training_steps):
```

```
    get_random_cached_bottlenecks()
```

```
    sess.run()
```

```
    train_writer.add_summary()
```

```
    if (i % FLAGS.eval_step_interval) == 0 or is_last_step:
```

```
        tf.logging.info('%s: Step %d: Train accuracy = %.1f%%' %
```

```
                        (datetime.now(), i, train_accuracy * 100))
```

```
        tf.logging.info('%s: Step %d: Cross entropy = %f' %
```

```

                                (datetime.now(), i, cross_entropy_value))
    get_random_cached_bottlenecks()
    tf.logging.info('%s: Step %d: Validation accuracy = %.1f%%'
(N=%d)' %
                                (datetime.now(), i, validation_accuracy * 100,
                                len(validation_bottlenecks)))

    FLAGS.intermediate_store_frequency
run_final_eval()
save_graph_to_file()
export_model()

```

这部分代码的核心在于，首先通过 `create_image_lists()` 函数创建盲文字符图像列表，把每个文件夹下的盲文字符图像按照 `testing_percentage` 参数和 `validation_percentage` 参数对图像进行分配，输出一个字典，包括标签、训练数据、验证数据和测试数据。它的分配用到了 `hash` 算法，它获取的对象是图像的文件名。通过 `hash` 函数获得图像的 `hash` 值后，可以获得该图像在散列表的位置，然后通过比较大小决定该图像属于训练集、验证集或是测试集。通过 `hash` 算法，可以获得一个稳定的分配结果。另外 `create_image_lists()` 函数还用到了一个正则表达式 `re.sub(r'^a-z0-9+', '', dir_name.lower())`，它可以读取文件夹的名字，并把非字母和数字的字符去掉，并把这个值当做该文件夹的标签。然后通过 `for` 循环对每一步进行训练，这里用到了 `get_random_cached_bottlenecks()` 函数，它的作用是随机从磁盘获取图像的瓶颈值。训练最后，通过输出相应的准确度。

另外，在训练开始之前，可以根据实验效果对超参数进行调整。如通过 `--how_many_training_steps` 参数对步长进行调整，`--learning_rate` 参数对学习率进行设置，`--validation_batch_size` 控制每一步验证集的图像数量。

预测部分的关键函数和代码见下。

```

load_graph()
load_labels()
top_k = results.argsort()[-5:][::-1]
for i in top_k:
    print(labels[i], results[i])

```

预测部分的相关代码，通过已经训练好的模型，输入预测图像，输出分类的结果。这部分的代码用到了 `argsort()` 函数，可以把分类的结果从小到大排列。

4.4 实验结果及分析

本文在实验过程中，数据集共被独立重复训练了 5 次，最终的识别准确度取这 5 次的均值。本文在测试集上的识别准确度分别为 53.5%、53.9%、54.8%、53.5% 和 54.4%，取均值为 54.02%。通过图 4.2 可以发现识别准确度在 54% 上下波动，这说明实验结果趋于稳定，该模型在盲文数据集进行迁移学习，识别准确度大约为 54%。

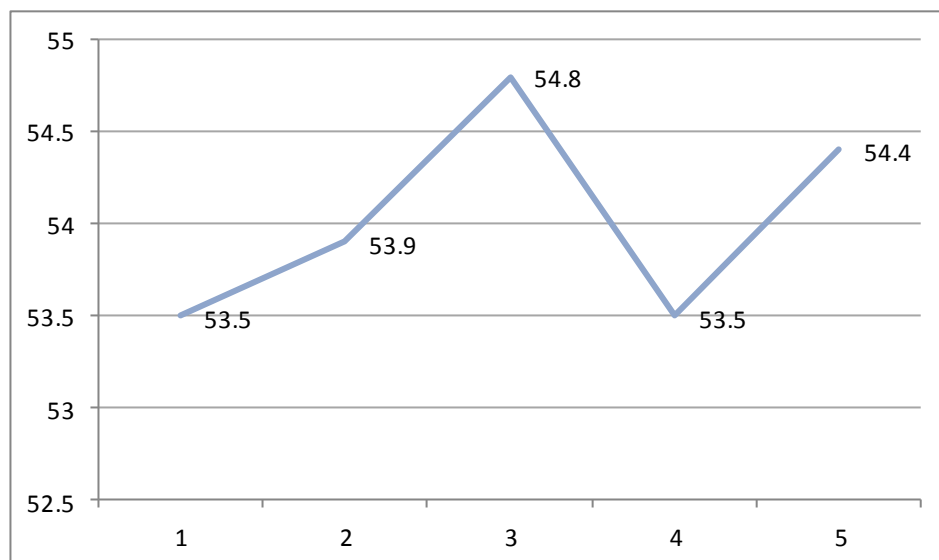


图 4.2 五次重复实验结果

在实验过程中，本文默认的训练步长为 4000 步，但是由于这时交叉熵比较高 (>1.0)，训练的模型还有提升的空间，于是我们把训练步长延长至 30000 步，这时交叉熵的值降了下来。本文进行一次步长为 30000 的训练，训练时长大约为 70 分钟，与传统的深度学习方法相比较，这个训练时长还是相当短的。另外，本文还尝试对图像进行预处理，如随机剪切、随机增减亮度等，实验发现这对识别准确度没有明显影响。实验时，本文还对初始的学习率进行了微调，发现对实验结果也没有显著影响。

关于数据集部分，在 2018 年 4 月和 5 月，作者和付鹏飞按照盲文字符数据集的设计方案进行拍摄和切割。我们共采集了四次，实验过程中，本文分别用迁移学习模型对不同数量的盲文数据集进行训练。实验发现，我们在切割了 160 张盲文图像时，测试的准确度达到了 52% 左右。加入新的一组盲文图像（30 张）后，测试的准确度下降为 47% 左右。数据集增大之后，测试的准确度应该有所提升，这里却有下降的趋势。本文分析，这可能是由于切割的盲文图像没有代表性

所致，由于每一种盲点都有其固定位置，但是由于人工切割的时候，可能造成错位的问题，所以导致代表性变差。这与我们设计的拍摄角度的差异有所不同，拍摄角度可能会造成一部分盲点的位置有点变化，但是总体来说，是可以辨认出来的。如图 4.3，问题盲文为切割不规范的盲文字符，它的上方留有大量背景，这会造成盲点在图像中相对位置错误。本文分析它是使模型识别率降低的主要原因。当然，也有可能是该部分盲文字符图像太少，与已经训练的模型有所冲突，导致识别结果变差。

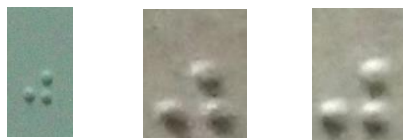


图 4.3 问题盲文、倾斜盲文和正常盲文

本论文还通过传统的图像分类方法对该盲文数据集进行识别，识别的结果见图 4.3。从图中可以看出，迁移学习获得了 54.02% 的准确度，比传统的 BPNN 和 SVM 方法要高出很多。kNN 方法获得了 48.89% 的准确度，迁移学习的识别准确度比该结果要高出约 5%。

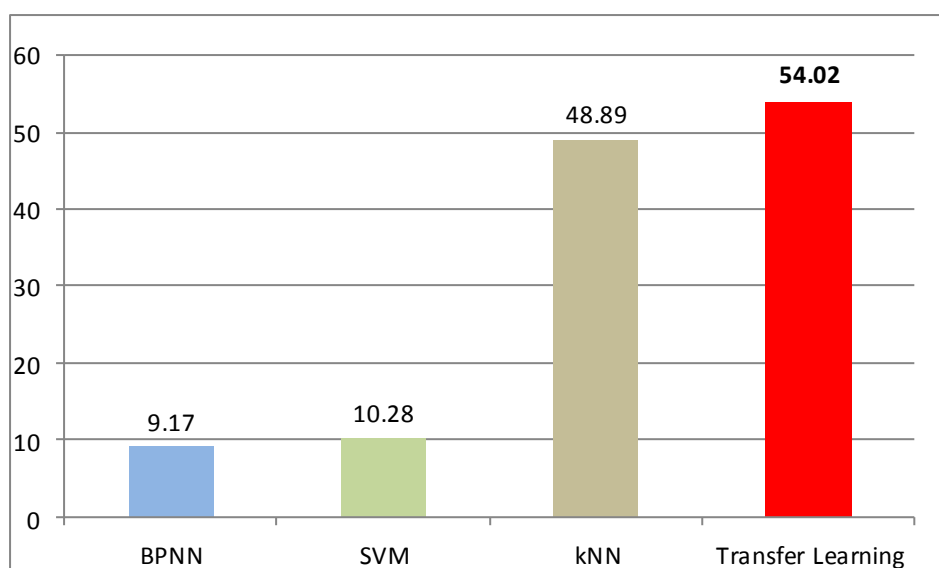


图 4.4 多种方法识别结果

实验证明，迁移学习与传统的图像分类方法相比，具有一定优势。这说明基于迁移学习的方法还是比较适合对盲文字符图像进行识别的。

第五章 总结与展望

在对深度学习方法进行研究的基础上,提出一种基于迁移学习的方法对盲文字符图像进行识别。传统的图像识别算法可以对盲文图像进行识别,但是需要依据经验对图像进行处理和操作,处理时间也比较长,存在盲目性和复杂性。基于迁移学习的盲文字符识别方法很好地克服了以上缺点。

同时,本方法也存在诸多不足之处:

(1) 数据集过小。由于时间有限,本文只建立了 10 种盲文图像的数据集,每种字符有 240 张图像,所以整个盲文字符数据集只有 2400 张图像。

(2) 数据集是否具有代表性是不确定的。本论文在创建数据集的时候尽量避免非字符因素带来的差异,所以选择了 6 种纸张进行了手工打印,并且选择在不同的光照条件下和利用不同的拍摄角度获取盲文字符图像。但是这种方法是否具有科学性是未知的。

(3) 迁移学习所采用的模型不具有多样性。本方法是利用 Google 的 Inception-v3 模型进行迁移学习,没有对比其它的 CNN 模型。所以,本文所采用的模型不一定是对盲文字符进行识别的最好的模型,本文所采用的方法只具有利用迁移学习对盲文字符进行识别的可行性的参考价值。

值得注意的是,笔者还对 CIFAR-10 数据集进行了测试。实验结果显示,即使是选取相同大小的数据集,测试准确度也能达到大约 70%。这要远大于本文获取的准确度(约为 54%)。

对于未来的工作,主要在于如何通过迁移学习的方法,提高盲文字符的识别率。一方面可以创建更加具有代表性的数据集;另一方面可以尝试用不同的模型进行迁移学习,或者用 CNN 的方法进行盲文字符识别。

盲文字符识别具有很大的研究价值和现实意义,如何把整页盲文图像转化为分段落的文字,也是研究者需要努力的方向。

参考文献

- [1] 程黎, 顾定倩, 刘艳虹, 等. 我国盲文使用状况的调查研究[J]. 语言文字应用, 2013(2):42-48.
- [2] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]// International Conference on Neural Information Processing Systems. Curran Associates Inc. 2012:1097-1105.
- [3] Dubus J P, Benjelloun M, Devlaminck V, et al. Image processing techniques to perform an autonomous system to translate relief braille into black-ink, called: Lectobrilie[C]// Engineering in Medicine and Biology Society, 1988. Proceedings of the International Conference of the IEEE. IEEE, 1988:1584-1585 vol.4.
- [4] Mennens J, Tichelen L V, Francois G, et al. Optical recognition of Braille writing using standard equipment[J]. IEEE Transactions on Rehabilitation Engineering, 1994, 2(4):207-212.
- [5] Ritchings R T, Antonacopoulos A, Drakopoulos D. ANALYSIS OF SCANNED BRAILLE DOCUMENTS[J]. 1994.
- [6] Hentzschel T W, Blenkhorn P. An optical reading system for embossed braille characters using a twin shadows approach[J]. Journal of Microcomputer Applications, 1995,18(4):341-354.
- [7] Oyama Y, Tajima T, Koga H. Character recognition of mixed convex - concave braille points and legibility of deteriorated braille points[J]. Systems & Computers in Japan, 1997,28(2):44-53.
- [8] Ng C M, Ng V, Lau Y. Regular feature extraction for recognition of Braille[C]// International Conference on Computational Intelligence and Multimedia Applications, 1999. Iccima '99. Proceedings. IEEE, 1999:302-306.
- [9] Murray I, Dias T. A portable device for optically recognizing Braille. I. Hardware development[C]// Intelligent Information Systems Conference, the Seventh Australian and New Zealand. IEEE, 2001:129-134.
- [10] Murray I, Dias T. A portable device for optically recognizing Braille - Part II: software development[M]. 2001.
- [11] Antonacopoulos A, Bridson D. A Robust Braille Recognition System[C]// Document Analysis Systems VI, International Workshop, Das 2004, Florence, Italy, September 8-10, 2004, Proceedings. DBLP, 2004:533-545.
- [12] Néstor Falcón C M T J. Image Processing Techniques for Braille Writing Recognition.[J]. Lecture Notes in Computer Science, 2005,3643:379-385.
- [13] Namba M, Zhang Z. Cellular Neural Network for Associative Memory and Its Application to Braille Image Recognition[C]// International Joint Conference on Neural Networks. IEEE, 2006:2409-2414.

- [14] Zhang S, Yoshino K. A Braille Recognition System by the Mobile Phone with Embedded Camera[C]// International Conference on Innovative Computing, Informatio and Control. IEEE Computer Society, 2007:223.
- [15] Al-Salman A M S, El-Zaart A, Al-Suhaibani Y, et al. An Efficient Braille Cells Recognition[C]// International Conference on Wireless Communications NETWORKING and Mobile Computing. IEEE, 2010:1-4.
- [16] Li J, Yan X. Optical Braille character recognition with Support-Vector Machine classifier[C]// International Conference on Computer Application and System Modeling. IEEE, 2010:V12-219-V12-222.
- [17] Alsalman A, El-Zaart A, Al-Salman S, et al. A novel approach for Braille images segmentation[C]// International Conference on Multimedia Computing and Systems. IEEE, 2012:190-195.
- [18] Pan S J, Yang Q. A Survey on Transfer Learning[J]. IEEE Transactions on Knowledge & Data Engineering, 2010, 22(10):1345-1359.
- [19] Xiao Jin Z. Semi-Supervised Learning Literature Sur-vey[J]. 2007, 37(1):63-77.
- [20] Nigam K, McCallum A K, Thrun S, et al. Text Classification from Labeled and Unlabeled Documents using EM[J]. Machine Learning, 2000, 39(2-3):103-134.
- [21] Blum A. Combining labeled and unlabeled data with co-training[C]// Conference on Computational Learning Theory. 1998:92-100.
- [22] Joachims T. Transductive Inference for Text Classification using Support Vector Machines[C]// Sixteenth International Conference on Machine Learning. Morgan Kaufmann Publishers Inc. 1999:200-209.
- [23] Thrun S, Pratt L. Learning to Learn[M]. Springer US, 1998.
- [24] Caruana R. Multitask Learning[J]. Machine Learning, 1997, 28(1):41-75.
- [25] Donahue J, Jia Y, Vinyals O, et al. DeCAF: a deep convolutional activation feature for generic visual recognition: International Conference on International Conference on Machine Learning, 2014[C].
- [26] Long M, Wang J, Jordan M I. Unsupervised Domain Adaptation with Residual Transfer Networks[J]. 2016.
- [27] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the Inception Architecture for Computer Vision[J]. Computer Science, 2015:2818-2826.
- [28] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[J]. 2014:1-9.
- [29] Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[J]. 2015:448-456.
- [30] Shiyu Mou. Image Classification with 5 methods[EB OL]. https://github.com/Fdevmsy/Image_Classification_with_5_methods, 2017.

附 录

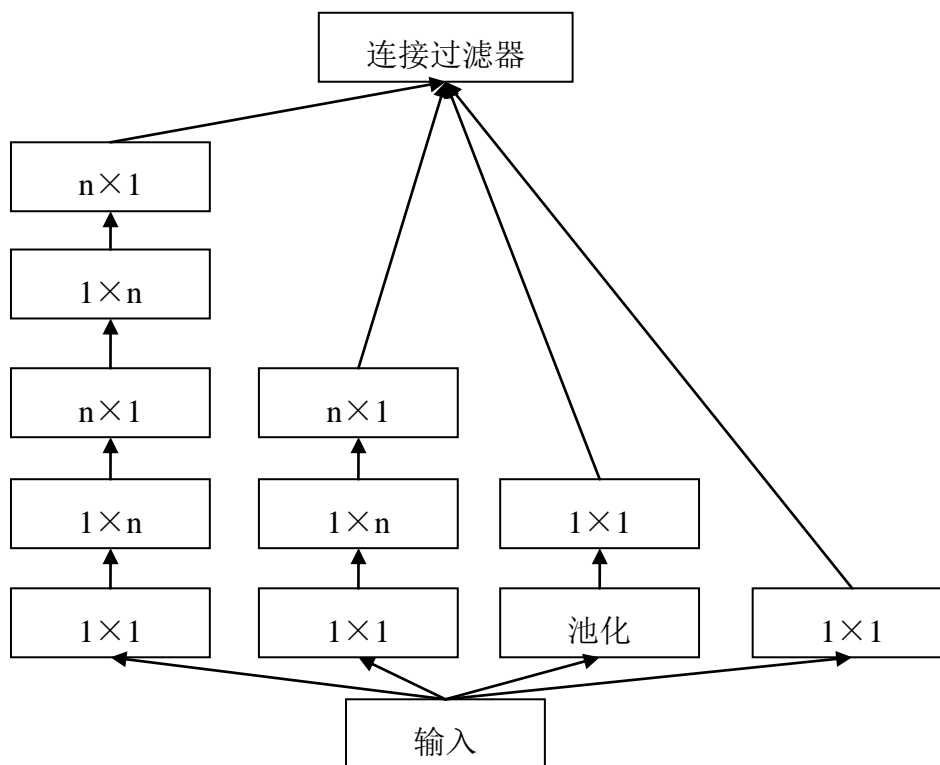
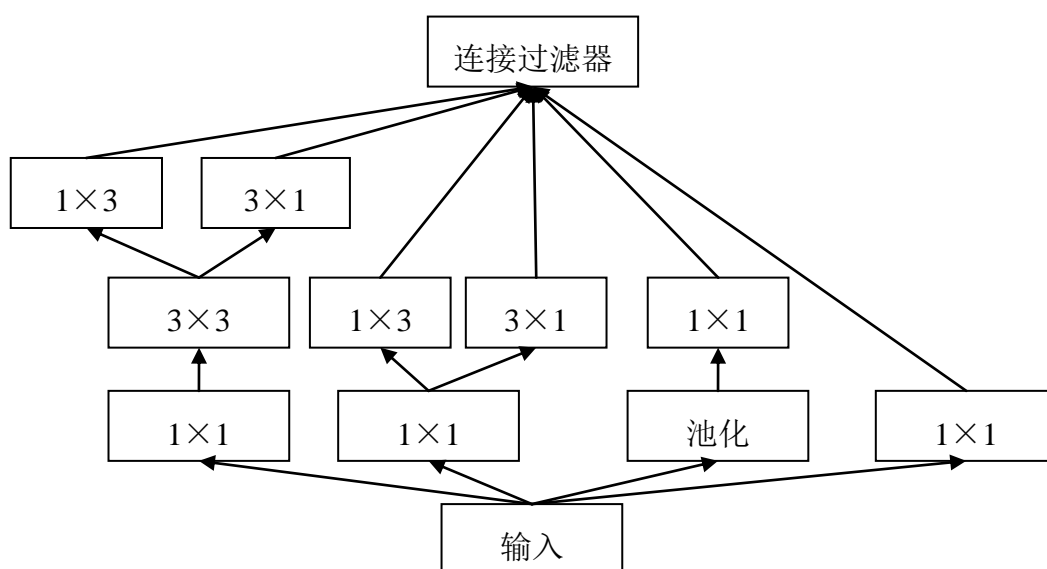
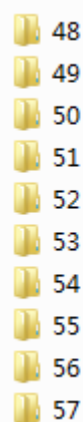
图 0.1 $n \times n$ 分解之后的 Inception

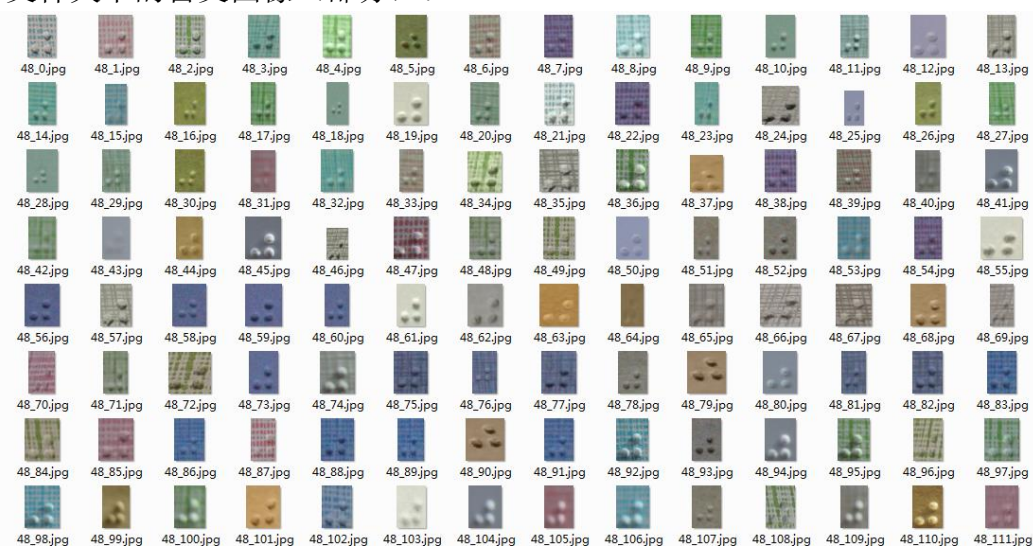
图 0.2 扩大滤波器组输出的 Inception

盲文数据集

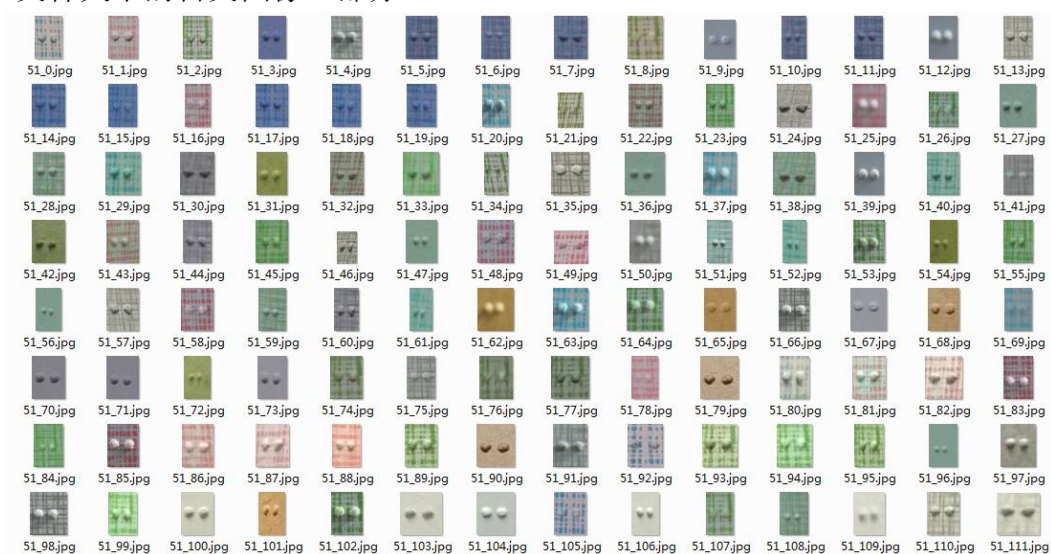
当前目录: D:\python3\braille



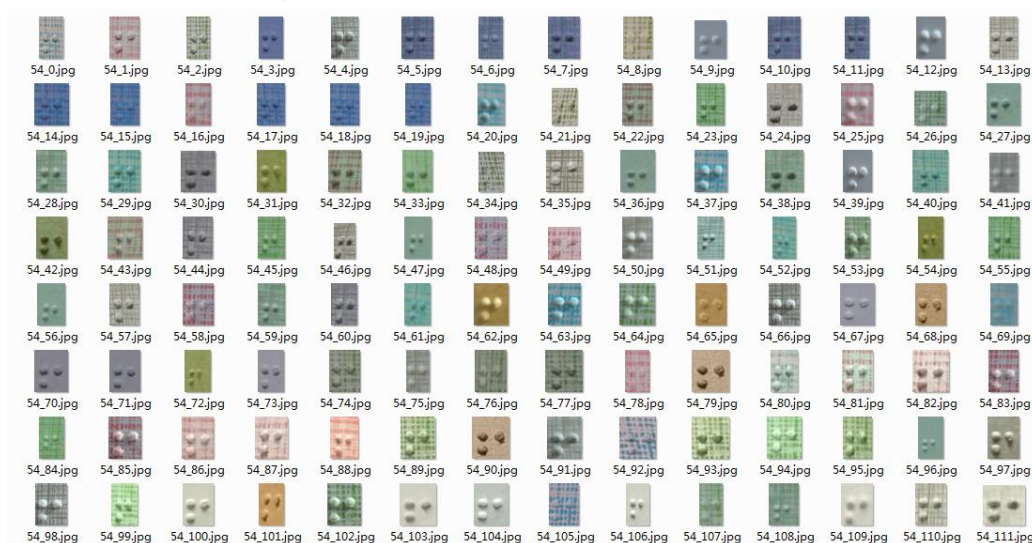
48 文件夹下的盲文图像（部分）：



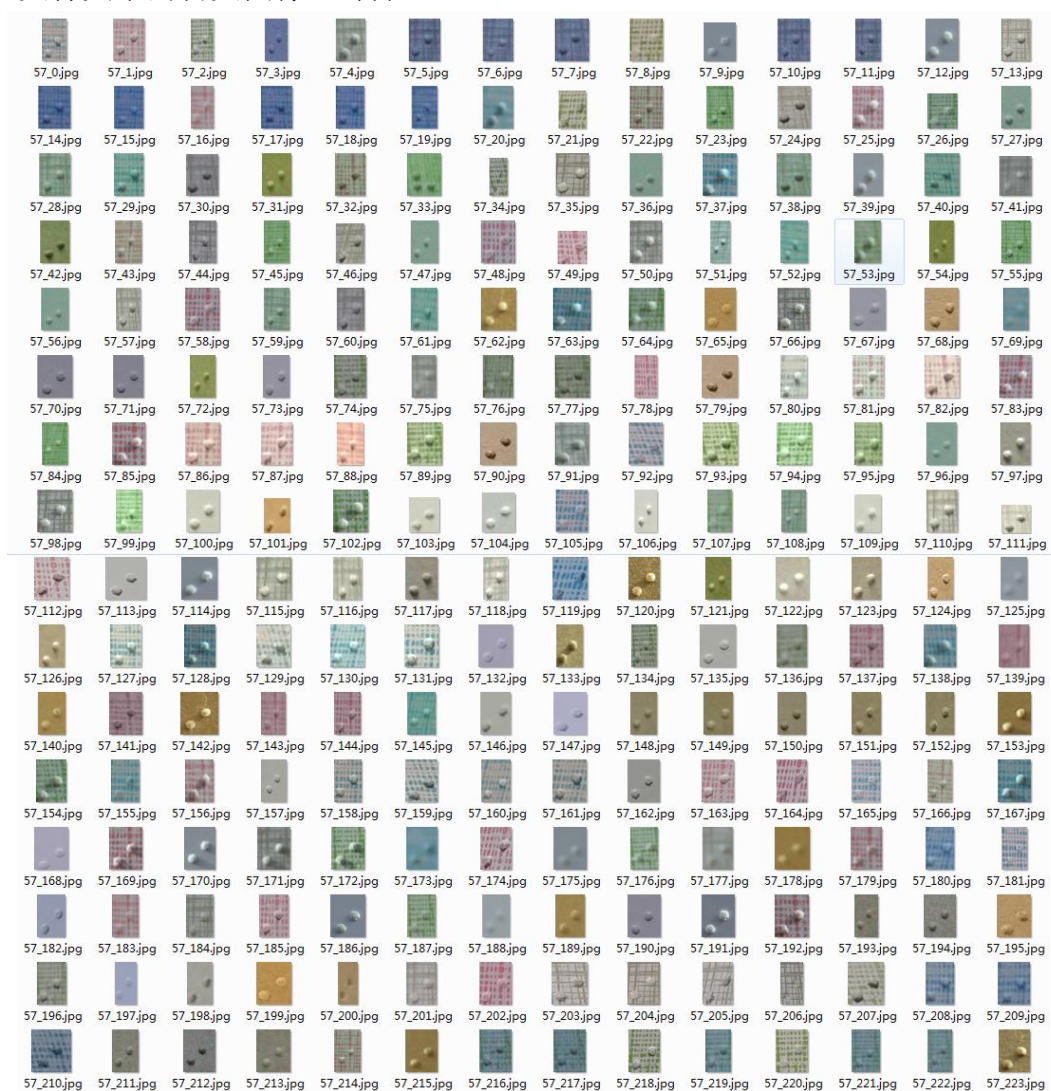
51 文件夹下的盲文图像（部分）



54 文件夹下的盲文图像（部分）：



57 文件夹下的盲文图像（部分）：



致 谢

在本人毕业论文即将完成之际，衷心感谢以下帮助过我的人。

感谢我的指导老师高若宇老师。您在我的毕业论文完成期间给予了我很大的帮助，时刻关注着我的毕设进展和身心状况，在此表示衷心的感谢。

非常感谢苏伟老师，也非常有幸参与苏伟老师的课题组。在这半年的学习期间，我感受到了苏伟老师对工作的热情和对学生的认真负责，也初步学习到了科研的精神和方法。

感谢兰州市盲聋哑学校的刘波老师，您在我的课题方面给予了很大的帮助。

感谢所引用参考文献的作者们，你们在图像识别和盲文字符识别所做的工作对我有莫大的启迪。你们对学术严谨的态度深深地感动了我。

感谢兰州大学，学校自强不息独树一帜的学术精神和良好的学习生活环境让我受益匪浅。

感谢我的同学和室友们，四年的朝夕相处，荣辱与共，你们是我在大学期间难忘的记忆。

最后我要感谢我的家人，是你们一直在背后支持我，鼓励我，不让我有太大的压力和烦恼，我爱你们。感谢在我考研期间一直陪伴我的女友，是你让我有良好的心情去迎接一个又一个挑战。

谢谢你们。

论文（设计）成绩

导师评语

建议成绩 _____

指导教师（签字）_____

答辩小组意见

答辩委员会负责人（签字）_____

成绩 _____

学院（盖章）_____

年 月 日