

杭州电子科技大学

硕士学位论文

题目：基于深度学习的中国绘画图像分类研究

研究生 黄雪丽

专业 计算机技术

指导教师 王小华 教授 杨冰 副教授

完成日期 2017 年 3 月

杭州电子科技大学硕士学位论文

基于深度学习的中国绘画图像分类研究

研 究 生： 黄雪丽

指导教师： 王小华 教授 杨 冰 副教授

2017 年 3 月

**Dissertation Submitted to Hangzhou Dianzi University
for the Degree of Master**

Chinese Painting Image Classification Research based on Deep Learning

Candidate: Huang Xueli

Supervisor: Prof. Wang Xiaohua ,Yang Bing

March,2017

杭州电子科技大学

学位论文原创性声明和使用授权说明

原创性声明

本人郑重声明： 所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品或成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。

申请学位论文与资料若有不实之处，本人承担一切相关责任。

论文作者签名： 日期： 年 月 日

学位论文使用授权说明

本人完全了解杭州电子科技大学关于保留和使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权单位属杭州电子科技大学。本人保证毕业离校后，发表论文或使用论文工作成果时署单位名称仍然为杭州电子科技大学。学校有权保留送交论文的复印件，允许查阅和借阅论文；学校可以公布论文的全部或部分内容，可以允许采用影印、缩印或其它复制手段保存论文。（保密论文在解密后遵守此规定）

论文作者签名： 日期： 年 月 日

指导教师签名： 日期： 年 月

摘 要

随着数字化技术的发展,实现中国绘画的精准分类及快速查找等功能,对中国绘画图像数字化的建立具有重要意义。传统的中国绘画识别系统主要有特征提取和分类两个步骤,其中特征提取主要是根据个人经验完成的,虽然能够对中国绘画进行分类,但仍存在细节信息容易丢失,模型泛化能力低等问题。寻找一种能够自动、高效识别绘画的技术是目前及未来研究的热点。

绘画分类作为图像分类的一种,其分类的难点在于特征的表达。而深度学习可以通过多隐层的网络模型对大数据进行训练,实现对数据特征的更好表达,从而提高分类结果的准确度。因此,针对中国绘画图像分类所出现的难题,本文对基于深度学习的中国绘画图像分类算法进行了研究。

本文首先提出了一种基于深度信念网络的方法,实现对中国绘画图像的分类。该方法一方面通过采用深度信念网络所忽略的二维结构信息实现对高阶统计信息的提取,另一方面将卷积操作应用到网络结构中,可以在降低噪音的同时增强对原始信号特征的表达。同时还在隐层限制波尔兹曼机结构中引入了最大概率模型和稀疏正则化算法,从而很好地实现了对概率的推理和超完备现象的弱化。最后本文通过实验验证了此方法在中国绘画图像分类上具有可行性。

随后本文提出了一种基于卷积神经网络的方法,对中国绘画图像进行分类。该算法首先针对过拟合问题,提出了一种改进的合成少数类过采样技术,以达到对数据良好的扩增,然后将扩增后的数据直接输入到卷积神经网络中,经过隐藏层的卷积和亚采样,并在结构中利用 ReLu+Sigmoid 替代传统的 Sigmoid+Sigmoid 激活函数,最终提取出可更好地表示绘画图像的特征。实验结果表明,与传统分类方法相比,该方法在中国绘画图像分类上具有更好的分类效果。

关键词: 中国绘画图像分类,深度学习,卷积深度信念网络,合成少数类过采样技术,卷积神经网络

ABSTRACT

With the development of digital technology , it is significantly meaningful to realize the accurate classification and fast searching of Chinese painting for the establishment of painting images system. The traditional Chinese painting recognition system contains two main steps, which are feature extraction and classification. Feature extraction is based on personal experiences, which leads to two problems including loss of detail information and low generalization ability of math model. To discover an automatic and efficiency recognized technology will be a research hotspot in the future.

Painting classification, as one type of images classification, its research difficulties come from the feature extraction. Deep learning methods could learn better feature representation through training large amounts of data compared with traditional shallow learning model, which improves accuracy of classification in many applications. Therefore, in view of the problems of Chinese painting image classification, this paper studies the Chinese painting image classification algorithm based on depth learning.

In this paper, a method based on DBN (Deep Belief Network) is firstly proposed to classify the Chinese painting. On the one hand, the CDBN model takes the two-dimensional structure information ignored in DBN model into account. On the other hand, it uses the convolutional operation, which makes the CDBN model be robust to noise and then learns good features containing statistical properties. At the same time, we introduce probability model and sparse regularization into CRBM (Convolutional Restricted Boltzmann Machine), which improves the ability of reasoning of CRBM model and weakens the model over-complete, and further enhances the performance of painting classification.

Then, we classify Chinese painting images based on a CNN (Convolutional Neural Network). We introduce an improved SMOTE (Synthetic Minority Over-sampling Technique) to overcome the over-fitting problem. The technique preprocesses the input raw data and then directly input the amplified data into CNN model, after the convolution and sub sampling of hidden layer, we replace traditional sigmoid + sigmoid function with ReLu (Rectified linear units) + sigmoid function to

extract better efficiency painting images representation. The experimental results prove that the method is effective in Chinese painting images classification task than traditional.

Key Words: Chinese painting classification, deep learning, convolutional deep belief network, synthetic minority over-sampling technique, convolutional neural network

目 录

摘 要.....	I
ABSTRACT.....	II
目 录.....	IV
第一章 绪论.....	1
1.1 研究背景.....	1
1.2 研究目的和意义.....	1
1.3 研究现状.....	2
1.3.1 中国绘画图像分类的研究现状.....	2
1.3.2 深度学习的研究现状.....	3
1.4 论文主要工作.....	4
1.5 论文组织架构.....	4
第二章 相关理论基础.....	6
2.1 引言.....	6
2.2 图像分类概述.....	6
2.2.1 图像分类的概念.....	6
2.2.2 图像特征提取的流程.....	7
2.2.3 分类器的类型.....	9
2.3 浅层学习方法.....	11
2.3.1 浅层学习概述.....	11
2.3.2 浅层学习局限性.....	12
2.4 深度学习方法.....	14
2.4.1 深度学习概念.....	14
2.4.2 深度学习训练过程.....	16
2.5 深度学习的模型.....	17
2.5.1 前馈深度网络.....	18
2.5.2 反馈深度网络.....	20
2.5.3 双向深度网络.....	21
2.6 本章小结.....	24
第三章 基于 DBN 的中国绘画图像分类算法.....	25
3.1 引言.....	25

3.2 DBN 的构成.....	25
3.3 卷积限制波尔兹曼机 CRBM.....	25
3.4 卷积深度信念网络 CDBN.....	27
3.4.1 最大概率池模型.....	27
3.4.2 稀疏正则化.....	28
3.4.3 CDBN 模型结构.....	29
3.5 实验结果与分析.....	30
3.5.1 实验所用数据库及目的.....	30
3.5.2 实验评价标准.....	30
3.5.3 实验设计及结果分析.....	31
3.6 本章小结.....	34
第四章 基于 CNN 的中国绘画图像分类算法.....	35
4.1 改进的 SMOTE 扩增方法.....	35
4.2 数据的生成.....	36
4.2.1 样本的聚类.....	36
4.2.2 样本的生成.....	36
4.2.3 聚类个数的设置.....	37
4.3 基于 Sigmoid+ReLU 的 CNN.....	37
4.3.1 激活函数 Sigmoid+ReLU.....	37
4.3.2 CNN 网络结构.....	38
4.4 实验结果与分析.....	38
4.5 本章小结.....	40
第五章 总结与展望.....	41
5.1 工作总结.....	41
5.2 未来展望.....	42
致谢.....	44
参考文献.....	45
附 录.....	49

第一章 绪论

1.1 研究背景

中国绘画是我国艺术界的璀璨珍宝之一，不仅承载着深厚的文化底蕴，还传达了中华民族的社会意识和审美观念。我们将中国的绘画称为“中国画”，并简称“国画”。主要指在绢、宣纸、帛上绘画并加以装裱，可以说国画是中国的传统绘画，具有很高的文化价值。随着信息技术的高速发展，人们迫切希望应用一种机器识别技术对中国绘画进行分类，从而实现中国绘画的数字化。

对于图像分类这个研究领域，一直处在人们的研究热点中，所应用的技术现如今已经很成熟。虽然中国绘画也是图像的一种，但其领域的很多技术还有待研究。由于中国绘画所涵盖的语义信息更加丰富和抽象，所以应用到传统图像的分类方法不一定适用于中国绘画^[1]。传统的中国绘画分类技术主要依靠浅层学习，采用的是人工设定特征方式。显然这种方式在处理复杂分类问题时，其泛化能力会受到很大的限制，且在设计大量的人工特征时，不仅选取费时费力，而且还需要选取人员具备启发式的专业知识，这在很大程度上靠的是运气和经验。深度学习能够自动地学习数据的特征，不需人为干涉，就能通过有监督或者无监督的方式将底层的特征抽象为高层特征，从而更好地实现数据的分类^[2]。又因为深度学习在绘画方面的研究还处于起步阶段，所以该方法无疑将成为绘画分类研究的热点。

基于深度学习的分类研究，主要通过模拟人的视觉信息处理机制，经过特征在不同层的变换，使该特征从原始空间的表示变换到一个新的空间^[3]。中国绘画就是应用该技术，自动地学习中国绘画的特征，不需人工标注或者人为设定特征，如颜色、纹理、结构等，通过层次网络结构的特征表达进行分类。

1.2 研究目的和意义

随着电子信息的快速发展以及多媒体技术的广泛应用，如何使数字化的图像资源成为一种有效的信息媒体，是目前及未来研究的热点。依据人们对多媒体讯息的了解，按其特质分为语音、音频、图像、视频和文本信息等。对于现在这个可视化年代，图像和视频可当之无愧地成为其重要的表现形式。但图像作为一种携有结构化信息的载体，其检索方式要比鉴于文本的检索方式困难的多^[4]。

在研究图像时，其分类问题一向是大家关注的热门话题。好的图像分类，不仅有助于图像数字化的建立，还可以方便人们管理、检索、查阅图像等。毋庸置疑，中国绘画属于图像的一种表现形式，但同其他常见自然场景下的图像还是有区别^[5]。首先绘画的材料制作一般为纸质的，由作者用毛笔等绘画工具完成，这就造成了在本质上易受自然、人为等因素的影响，而不易保存。其次对于绘画作品的分类，一般是由人工标注的，存在个人主观因素强及标注过程繁琐等问题。最后名画著作的存储不是聚集在某一个专场，而是分布在世界各地，这给人们异地欣赏、作品共享带来了不便，同时也不利于绘画作品的管理。

综上所述，实现绘画作品数字化，不仅能够满足人们关于绘画作品电子信息化的需求，还可以对数字化作品博物馆的构建及其管理等起到非常好的辅助作用。其中有效的索引、浏览和检索作品是当前迫切需要解决的问题，而处理好这个问题关键的一步是实现绘画图像的分类。在科学研究方面，绘画作品的分类不同于传统的图像分类，它是一种非自然场景下的产物，具有着本身的特性，所以传统的图像分类方法不一定适用于绘画图像的分类。若能实现一种精确、高效的绘画图像分类方法，将有助于中国画作品数字化的研究，能够帮助绘画作品的长久保存，促进分类标准的客观公正，同时能够解决异地共享、管理等问题。

1.3 研究现状

1.3.1 中国绘画图像分类的研究现状

虽然当前在图像的检索、分类及自动语义标注等区域已做了很多的探究，但针对中国绘画的探究并不多。中国绘画与西方绘画不同^[6]，在上色、触笔、色泽、轮廓等角度，具备自己的特色，主要体现的是“以形写神”。特别是在表现透视要领时，可以不限在固定的视域中，依照作家的感觉和需求，转移作画的立脚点，在作品上展示出可见和不可见的景物。这就使得那些应用在物体识别的方法、图像基于内容的检索技术、图像低层视觉特征描述子等在中国绘画图像上的应用受到一定的限制。另一方面，中国绘画作品与普通数字图像所表达的确切语义是不一样的，它自身所表示的语义信息具有一定的抽象性，不着重于物体的光色变化、焦点透视和物体外部的写实。对于绘画作品相关研究的历程如下：

James.Z.Wang 等人^[7]在 2003 年时使用多尺度小波变换的特征系数去表征中国书画的纹理特性，然后采用二维多尺度隐马尔科夫模型(2DMHMM)实现对不同作家的作品分类；张丹青等人^[8]在 2004 年时采用一种框架模型完成了对中国绘画的分类和检索；蒋树强^[9]在 2005 年时通过使用决策树对提取到的颜色、纹

理、形状等层次化的特征进行处理,实现国画和非国画的分类,通过应用 SVM 实现工笔和写意画的分类;Shuqiang 等人^[10]在 2006 年时采用边缘大小直方图来表示中国绘画图像边缘的稀疏度和粒度,从而实现工笔画和写意画的分类;鲍鸿^[11]在 2012 年时通过 BOW 构建类别相关的视觉单词频率直方图,并应用 SVM 对提取到的 color-shift 特征按照人物、花鸟、山水进行分类。

由上文描述可知,对于中国绘画图像的分类,人们一般采用的是浅层的学习方法,而在深度学习方面的应用并不多。在查阅众多参考文献及资料后,得出以下关于中国绘画方面引入的深度学习方法:Crowley E J 等人^[11]采用 CNN 方法进行了从自然图像中识别各种目标的研究。Meijun Sun 等^[12]提出了一种基于混合稀疏的 CNN 方法来提取中国水墨画的笔触特征,按作者对其划分,并取得了很好的效果。由此可见,在中国绘画领域中采用深度学习方法,将会有远大的应用前景。

1.3.2 深度学习的研究现状

当前深度学习在人工智能领域已获得了重大突破,如语音识别、图像处理、自然语言处理等,这是因为近年来随着互联网的快速发展,使得网络上的数据迅速膨胀,如何处理好这些数据是当前研究的热点。前文对绘画作品的研究中,所用的方法可以称之为浅层学习方法,易受系统计算函数复杂度的限制。深度学习与浅层学习方法的最大不同在于它可以自动地从大数据中学习特征,而不是选用手工设计特征。毋庸置疑,提取出好的特征表达一直是提高机器识别或分类的关键,但在过去由于存在训练速度及收敛等问题,使得浅层学习处于统治地位。浅层学习中的特征,由于主要是靠人工设计的,所以需要设计者具备一定的先验知识,这样就造成了所提取到的特征具有一定的主观性,且设计出一个被大家认可的具有广泛性的有效特征表达难度系数比较大,允许出现的参数调节不易太多。所以当数据规模比较大时,应用浅层学习很难体现出其优势,而深度学习的出现,可以很好地解决该问题^[13]。

深度学习作为人工智能领域的新秀在图像识别领域取得了一定的成果,1989 年 Hinton 提出的深度信念网在图像识别上取得了很好的结果^[14];2006 年 Hinton 提出了一种以非监督贪心训练方法为基础的 DBN,给深度神经网络的应用带来了希望^[15];2012 年 Hinton 率领他的学生应用深度卷积神经网络在 ImageNet 图像库中,获得了令世人震惊的分类效果^[16];2013 年在 ImageNet ILSVRC 比赛中,排名前 20 的小组所使用的技术都为深度学习,其中第一名 Rob Fergus 小组通过对网络结构的进一步优化,采用卷积网络使错误率降到 11.197%^[17];2014 年

GooLeNet 通过增加 CNN 的深度使得分类的错误率降到了 6.656%，从而获得 ImageNet ILSVRC 比赛的冠军^[18]。

尽管深度学习在 ImageNet 上取得了很大的成功，但对于绘画图像来说是否能够取得相应的结果，还需要人们进一步验证。同时它本身也存在一定的问题，例如由于预测误差的传播方向是从顶层到底层的，若网络结构比较深，当该误差传到底层时，就变得很小，难以使底层参数进行更新。另一方面由于深度学习是自动地学习特征，所以训练数据在大的样本下才可以取得好的效果^[19]。随着深度学习研究的不断深入，如何将其应用到较小的训练集，以拓宽它的应用范围；如何探索出高效的、性能好的深度学习网络，是现在及未来研究的重大课题。

1.4 论文主要工作

本文方法是从目前信息技术的发展状况和中国绘画分类研究现状与发展趋势切入的。通过查阅国内外相关文献和参考资料，对当前中国绘画图像分类及深度学习的研究做了系统性的概括与总结，研究目的是运用深度学习的自动学习特性，提取出能够更好地表征数据的特征，从而最终改善中国绘画分类的效果。基于此，论文的主要工作描述如下：

(1)我们首先剖析了当前应用在中国绘画图像分类的方法以及这些方法所存在的问题，然后详细介绍了中国绘画图像分类和深度学习的研究现状和发展趋势，并通过比较现有的中国绘画分类方法，确定了本文的研究方向。

(2)在学习了 DBN 后，提出了一种 CDBN 方法。实现对中国绘画图像的分类，该方法通过应用无监督的学习方式来提取局部特征中的高阶信息，然后将 DBN 与卷积相结合，以实现局部特征提取与权值共享，同时利用 DBN 的概率最大池方法来降低数据维数，通过采用稀疏正则化来防止数据超完备化。最后通过实验来验证此方法在中国绘画图像分类上可以取得良好的分类结果。

(3)在研究过卷积神经网络 CNN 以后，本文提出了一种基于 CNN 的中国绘画图像分类法。该方法首先针对过拟合问题，应用改进的 SMOTE 算法来扩增数据，然后将扩增后的数据直接输入到 CNN 中，经过隐藏层的卷积和亚采样，并 ReLu+Sigmoid 替代传统的 Sigmoid+Sigmoid 激活函数，最终提取出可更好地表示数据的特征。最后实验结果表明了此方法在中国绘画图像分类上具有很好的可行性。

1.5 论文组织架构

本文主要由五大部分组成，具体组织结构如下：

第一章为绪论，首先简述了本文的研究背景、目的和意义，然后详述了中国绘画图像分类、深度学习的研究现状，最后对本文的主要工作和论文结构做了概述。

第二章是本文相关理论的基础部分，首先介绍了关于图像分类的一些方法；接着引出了属于该范畴内的中国绘画图像分类，并指出了它与常见的自然图像分类的区别；最后详细阐述了深度学习方法以及其相关内容，如概念、训练过程、常见模型等。

第三章研究 CDBN 在中国绘画图像分类中的应用，通过在 DBN 中引用卷积和最大值池化模型及稀疏正则化方法，来提高其学习中国绘画特征的能力，通过实验，可得出该方法在中国绘画图像分类上的结果要好于传统 CNN 和 DBN 方法。

第四章，通过对 CNN 的研究和分析，本章提出了一种基于 CNN 的中国绘画图像分类法。首先针对过拟合问题，提出了一种改进的 SMOTE 方法，通过该方法扩增数据，然后将扩增后的数据直接输入到 CNN 中，经过隐藏层的卷积和亚采样，并采用 ReLu 和 Sigmoid 激活函数去提取表征中国绘画图像的特征，最后对其进行分类。从实验结果可以看出，该算法大大的提高了分类的准确率。

第五章为论文的总结和展望部分，总结了本文的研究方法，然后对本文方法所存在的缺陷进行分析，并给出了后续的改进思路。

第二章 相关理论基础

2.1 引言

近几年来深度学习在机器学习领域获得了重大突破,使得其应用前景变得十分广泛。虽然该方法在图像识别、语音识别等领域均取得了突破性的进展,但是目前在中国绘画图像分类领域尚处于起步阶段,很多技术还不够成熟。而在图像分类中,大部分应用的是由人工设计的特征,其特征的选取既耗时耗力,又需要具备专业知识和经验的人员,由此为自动学习一些特征,引入了源于人工神经网络的深度学习方法。若能同样将深度学习成功地引入到中国绘画图像分类中,将具有重大意义。

2.2 图像分类概述

2.2.1 图像分类的概念

图像分类是一种通过应用分类技术对提取出的目标物特征(一般用特征向量表示)进行分类的技术。模式识别是通过应用计算机技术模拟人的辨识能力,更具体地说通过计算机完成对多种事物或对象的分析、描述、判别和辨别。所以说图像分类属于模式识别,其分类的过程就是模式识别过程,是目视判读的延续和发展。图 2-1 是图像的识别过程,通过预处理图像的输入数据提升图像的质量,然后分割图像以提取出感兴趣的目标区域,最后对目标物进行特征提取,并选择合适的方法进行分类^[20]。现对各个阶段分别作如下简要的讲解。

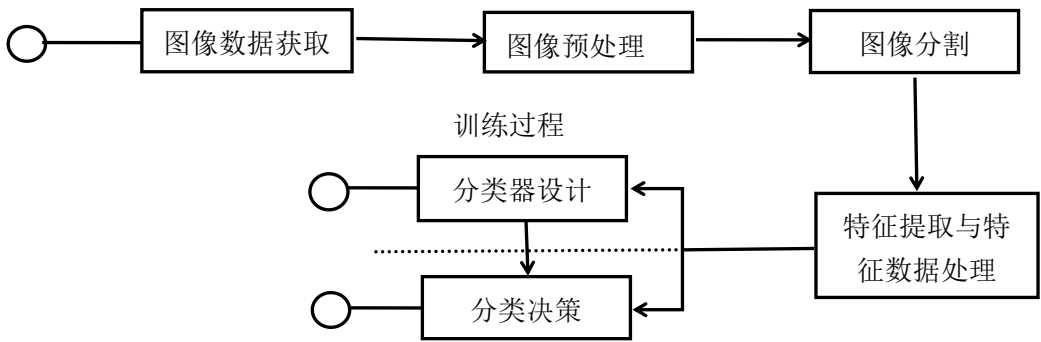


图 2-1 图像识别过程

1)图像数据获取

获取数据的方式是在经过对原数据的收集和量化后,用矩阵或向量的形式表示。通常输入对象信息有三种:二维图像,如文字、图片、超声影像等;一维图像,如脑电图、心电图、机械振动波形、地震波、电磁波等;物理参数和逻辑值,如医学诊断中的各种数据。

2)图像预处理

因为在原始图像的信号中经常存有噪音和变形问题,所以需要预处理,如锐化、增晰、灰度级修正、复原等,以提升图像质量,为后续分割和识别做准备。

3)图像分割

图像分割是从目标中提取出感兴趣部分,如边缘检测、点检测、区域生长、分类合并等。在识别分类过程中,该阶段是为后续的特征提取和识别做准备的,对分类效果有着重要的影响。

4)特征提取与特征数据处理

图像识别阶段中的特征描述是通过各种描述方法来实现的,如颜色描述、纹理描述、直方图描述等。在识别的过程中,此过程为特征的形成阶段。原始特征有可能维数很大,或者存在冗余,致使后续的分类识别效果不好。特征数据的处理是对图像的原始特征进行去粗存精的过程,包含特征提取、选择以及对特征数据的更高层次的组织等。

5)分类决策

分类决策通过采取统计算法将被识别目标判为某类。常用的方法是从样本训练集上寻找一种判决准则,使得在应用该准则后,将对分类不好的因素降为最低,即提供某种意义上最优的算法,根据对象的特征向量区分出该对象所属的模式。

随着图像分类技术的不断进步,使得越来越多的领域开始使用该技术,如遥感探测、场景分类、气象预感等,从图像识别的过程可知特征提取和分类在图像分类中占有重要地位,所以下面我们主要讲述的是图像特征提取的流程和分类器的类型。

2.2.2 图像特征提取的流程

好的特征提取是图像分类取得良好效果的必要条件,所以特征的选取对于图像分类来说是很重要的。特征的选取方法主要为:对原始特征的获取,如颜色特征采用颜色直方图、颜色矩等方法,纹理特征采用小波变换、随机场模型法等,形状特征采用 Hough 变换、几何矩特征法等;线性特征法,如 PCA、FLDA 等;非线性特征提取法,如核方法、流行学习等;特征袋模型,也叫词袋模型;稀疏编码。其中经常需要从许多初始特征中选取一些对分类器测度有益的特征子集,

而相关程度高的初始特征或噪声大的初始特征通常在去掉后,可以改善分类器的性能。所以特征的提取可以看作一个不断剔除无用特征和聚合相关特征的过程,最终可以使得分类器的性能达到最好^[21-23]。下面以两个初始特征情形为例,分析其特征提取的流程。

1)计算每类的特征均值

$$\begin{cases} \hat{\mu}_{xj} = \frac{1}{N_j} \sum_{i=1}^{N_j} x_{ij} \\ \hat{\mu}_{yj} = \frac{1}{N_j} \sum_{i=1}^{N_j} y_{ij} \end{cases} \quad (2-1)$$

在式(2-1)中,两式结果是基于训练样本的估计值, N_j 指第 j 类的样本数, x_{ij} 和 y_{ij} 是第 j 类中的第 i 个样本的两个特征分量。

计算特征方差, 第 j 类两个特征方差估计值:

$$\begin{cases} \hat{\sigma}_{xj}^2 = \frac{1}{N_j} \sum_{i=1}^{N_j} (x_{ij} - \hat{\mu}_{xj})^2 \\ \hat{\sigma}_{yj}^2 = \frac{1}{N_j} \sum_{i=1}^{N_j} (y_{ij} - \hat{\mu}_{yj})^2 \end{cases} \quad (2-2)$$

理想环境中, 类别相同的全部对象的相应特征分量一般是非常相近的, 即(2-2)中的特征方差在理想情况下的估计值很小。

2)计算特征相关系数

$$\hat{\sigma}_{xyj} = \frac{[\frac{1}{N_j} \sum_{i=1}^{N_j} (x_{ij} - \hat{\mu}_{xj})(y_{ij} - \hat{\mu}_{yj})]^2}{\hat{\sigma}_{xj}^2 \hat{\sigma}_{yj}^2} \quad (2-3)$$

上式 $\hat{\sigma}_{xyj}$ 是第 j 类 x, y 两个特征分量之间的相关系数估计值, 取值范围 $[-1, 1]$ 。若值为 0, 说明二者之间没有相关性, 可定为两个独立特征分量; 若接近 1, 表明二者之间的相关性极强, 可以合并或者舍去其一。若接近-1, 说明一个特征分量与另一个的负值成正比, 只是符号差别而已。

3)计算类间距离: 类间距离是区别两类性能的重要指标。对特征分量 x 来讲, 不同类(i 和 j), 其类间距离为

$$\hat{D}_{xij} = \frac{|\hat{\mu}_{xi} - \hat{\mu}_{xj}|}{\sqrt{\hat{\sigma}_{xi}^2 + \hat{\sigma}_{xj}^2}} \quad (2-4)$$

同理特征分量 y 有同样的估算。显然, 类间距越大其特征的质量越好。

4)特征的聚合(降维): 把两个特征融合为一个特征, 可减少特征空间的维数。利用一个线性函数 $z = ax + by$ 能够将两个特征 x, y 聚合为一个新的特征 z 。系数

a, b 决定了两个特征各自在聚合中所占的比例。只需对特征系数(即特征的缩放倍数)稍加限制: $a^2 + b^2 = \cos^2 \theta + \sin^2 \theta = 1$, 则有

$$z = x \cos \theta + y \sin \theta \quad (2-5)$$

如此, 原来两个系数 a, b 变成一个变量 θ 。原来的二维特征变量变成了一维特征, 同时应选择使得类间距最大的特征进行聚合。

2.2.3 分类器的类型

图像经由特征选取后, 通常需要判别图像之间的近似性, 从而为图像识别工作提供最终的识别结果^[24,25]。而近似性判别通常需要借助分类器, 常见的分类器可分成以下四种:

1)统计分类器

统计分类器的形成是建立在概率统计理论的基础之上, 通过决策函数对目标进行分类。按照决策函数的组成可将该模型分为两大类型: 基于似然函数的分类器和基于集群分析方式的分类器。

2)结构分类器

在图像分类中, 常常根据物体自身结构特点来对它进行识别或分类, 如生物特征中的指纹、面孔、染色体的识别分类, 文字的判别等等。这类根据模式或者物体自身结构与结构的不同来区分或者识别分类的方法就是结构分类法。这种分类方法一般是基于形式语言理论, 按它的结构可以分解成模式基元或者子模式, 对于模式基元, 通常使用文法的形式来描述之间的连接关系。

3)模糊分类器

模糊分类器通过采用模糊集理论, 实现对模糊信息的采集和分析, 然后对数据进行分类。该方法的思想与人类智能思维相接近, 是机器学习领域中值得研究的重大课题之一。常见方法有最大隶属原则法、就近原则法、模糊聚类分析法。

4)神经网络分类器

ANN 的产生源于对生物神经网络结构和功能的效仿, 它的模型、拓扑结构学习和训练的算法等都是基于生物神经元系统, 具有联想存储、高速寻找最优解、自动学习等的优点, 适合于解决分类问题, 在图像分类或识别中得到了大量的推广^[26]。图 2-2 是神经网络分类器的基本框架图, 首先算出匹配度, 再将匹配度输送至第二级, 分类器第二级再对第一级做出反馈, 通过一些算法来训练网络中相应的权值, 循环执行该步骤, 直至得出期望值。

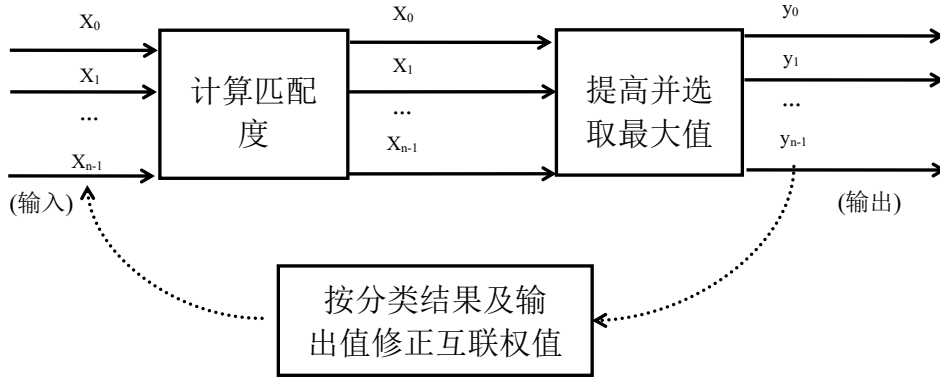


图 2-2 神经网络分类器图

ANN 通过某种方法将需要处理的单元进行连接，并对连续的输入作出状态响应。按照神经元的连接方式，该网络可分成前馈神经网络、反馈神经网络、层内互联神经网络、全互联神经网络^[27]。结构如图 2-3 所示。

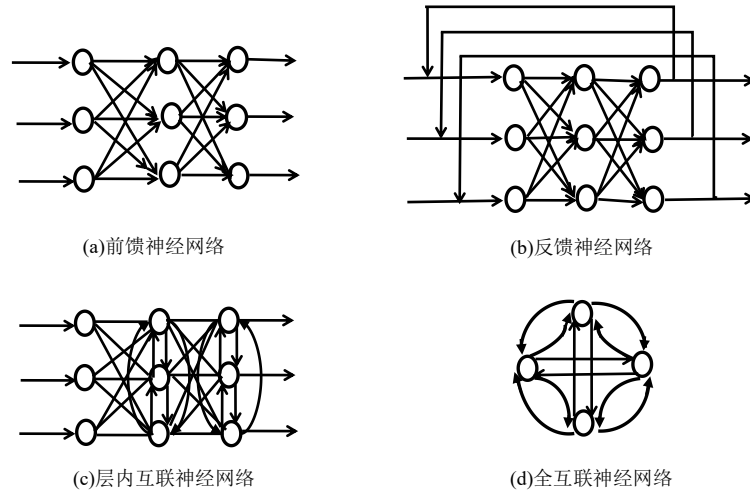


图 2-3 神经网络结构模型

神经网络与模式识别有着密切的联系，按照学习方法的不同，神经网络可分成监督学习神经网络和非监督学习神经网络两种。典型的监督学习神经网络是先利用已知分类结果的样本对网络进行训练，接着利用学习过的网络对新的样本进行分类。非监督学习神经网络应用的是一种无指导的学习方式，通过对数据的结构和数值进行归纳，自动地学习数据的内部组成，这同模式识别中的聚类分析是一样的。

分类位于图像识别的最后一个环节，其算法的选取对目标物分类识别的准确率有很大的影响，除了上述几种分类器，人们也会用到基于核函数的 SVM 及 Adaboost 算法等，至于选择哪种分类器，应根据具体的研究对象和研究内容等做决定。

2.3 浅层学习方法

2.3.1 浅层学习概述

应用在ANN的反向传播BP(Back Propagation)方法是最具有代表性的浅层学习方法之一，同时它的出现给机器学习带来了新的转机，兴起了一股研究基于统计模型的热潮。在这之后，陆续提出了各式各样的浅层学习方法，如最大熵法、K-Means、Boosting 等。这些模型的结构基本上有一层或没有隐藏节点，应用BP在训练时计算梯度，然后应用梯度下降法在参数空间寻找最优解。由于该模型经常应用凸代价函数，在进行理论分析时相对简单，训练方式也便于掌握，所以在很多应用中取得了成功。相比之下，由于神经网络在进行理论分析时难度比较大，同时训练方法需要很多经验和技巧，所以该方法并没有获得高度的重视^[28]。

传统的中国绘画图像分类涉及到了计算机视觉、机器学习、图像检索以及绘画艺术等多门学科，常见的方法基于浅层学习，虽然这些方法已取得了一定的成就，但在某些方面还存在一些局限性。图像识别中在处理一些特定的问题时，常常会在采集到数据之后进行一些特征提取。浅层学习中这些特征提取算法主要依靠的是人们对这些数据研究发现的规则，如常见的SIFT、HOG之类的特征。但在设计这些特征的同时，也随之带来了来一些问题，如特征的选取需要选取人员具备相关的专业知识且耗时耗力，一定程度来说靠的是履历和运气。另一方面这些特征的选取，所涵盖的语义信息不够全面。在中国绘画图像特征的提取中，常常提取的是基于底层的视觉特征，即浅层学习到的特征。这些特征的学习，只是人为设定的某些特征，往往会忽视中层语义所涵盖的特征，且这些中层特征对于特征的良好表达可以起到很好的促进作用，但对于人工而言是十分困难的^[29-30]。图2-4显示了某些数据在应用浅层方式提取特征时所忽视的一些中层信息。

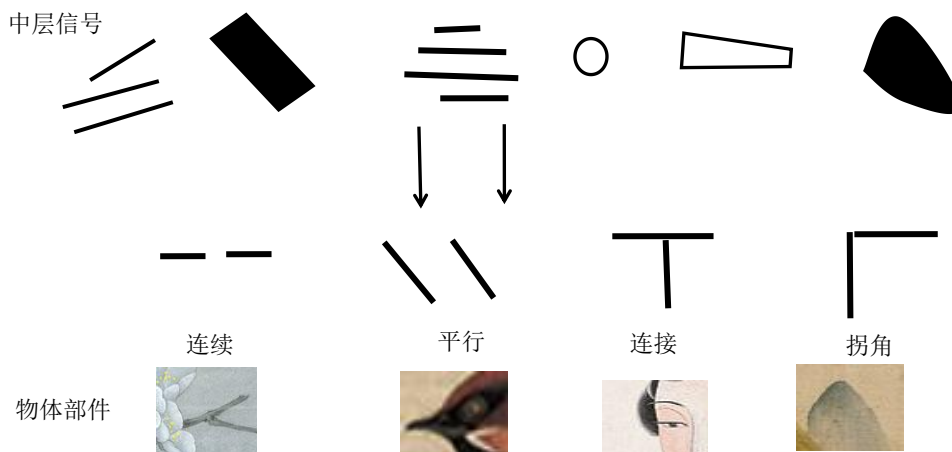


图 2-4 中层特征

2.3.2 浅层学习局限性

尽管 ANN 被叫做多层感知机，但其实它只有一层隐藏节点，是浅层学习中的经典方法之一。从最早由 Frank Rosenblatt 提出的具备单层计算单元的感知器 (Perception) 开始，到后来由 Minsky 的论证，挖掘出感知器并不能很好地解决许多的基本问题，验证了初期的浅层结构无法有效地解决非线性问题，但可以有效地对简单的结构进行计算^[31]。所以通过单个神经元构成的单层感知器，只能用来处理线性可分的二类型问题，其基本模型如图 2-5 所示。

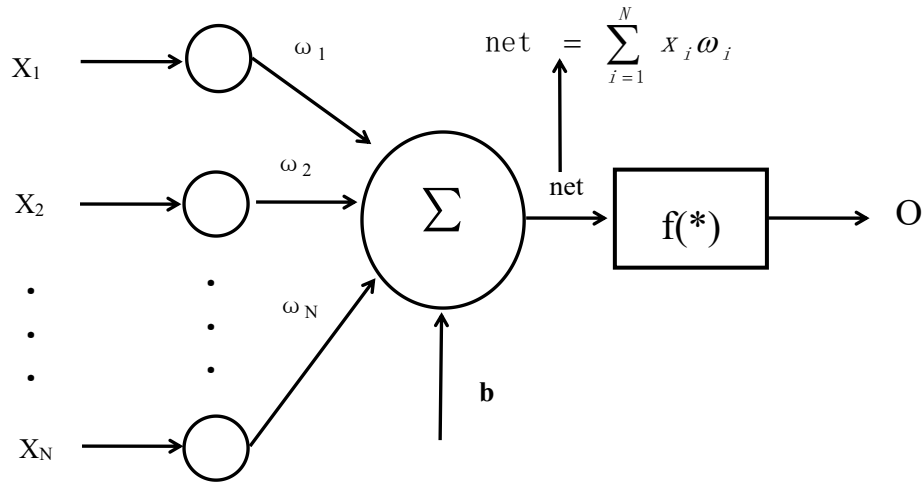


图 2-5 神经元的模型

在神经网络中，单层感知器只可以做近似分类，不能处理好线性不可分的问题；感知器的激活函数由于使用的是阈值函数，所以在输出的时候只有两种取值，这使得在扩展分类的种类上受到了限制；网络在输入的样本奇异时需要用很长的时间去学习，且其学习效果只对单层有效。Braverman M 在他的研究中，论证了许多浅层结构不能表示复杂函数^[32]，其计算模型如下图 2-6 所示。

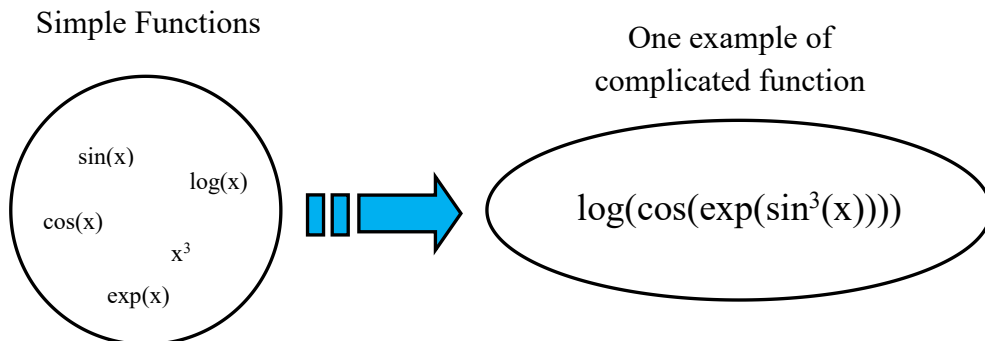


图 2-6 浅层模型复杂函数的表示

由上图可知，浅层学习在表达复杂函数时，不但增大了计算的复杂度，而且在表征数据的内在信息时，描述能力不充足。面对如今这样一个大数据时代，当

采用浅层学习方式对其进行训练时，会出现以下问题：实现复杂函数的表示比较困难，复杂分类的广泛性应用受限。于是探寻出一种能够从大数据中发掘出更多有价值信息的一种模型，是有必要的。深度学习的出现，刚好可以解决此问题，这是因为深度学习采用的是层次化的网络结构，可以将低层的特征逐层抽象到高层特征，从而更好地表现语义或意图，实现复杂函数的逼近^[33]。

在机器识别中好的特征表达是成功的关键，并且在识别过程中计算和测试所需时间大部分用在特征提取环节，当前所采用的特征样式普遍是由人工设计的，且在采用人工设计的特征进行融合时，达到好的分类或识别效果比较困难。例如 Gehler 在 LP- β 多核学习(Multiple Kernel Learning, MKL)的实验中，在数据集 Caltech-256 上先通过将 39 个不同的特征进行融合(如 PHOG、SIFT、Region 等)，然后对目标物进行识别，其训练出的结果如图 2-7 所示，验证了 MKL 在普通特征上表现能力有限^[34]。

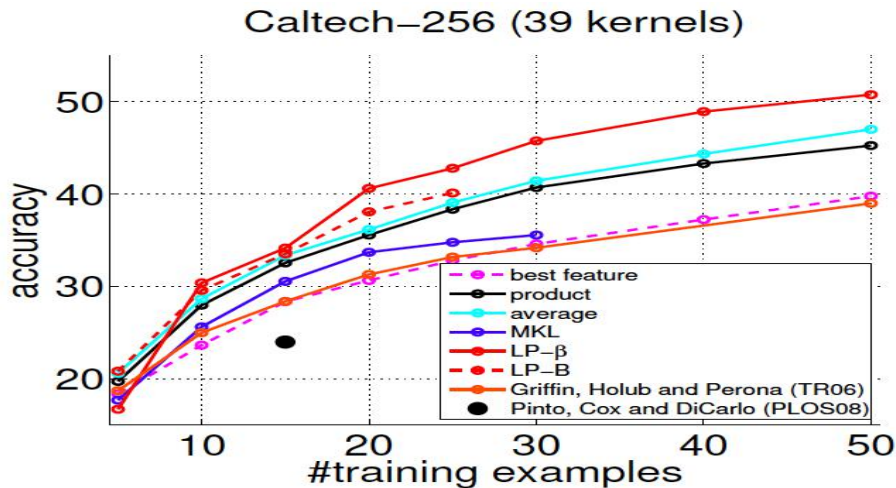


图 2-7 不同特征进行融合的识别效果

既然特征占据如此重要的地位，那么是不是特征愈多愈好呢？答案是否定的。虽然通常特征愈多，表示的信息也愈多，使得准确度有所提高，但随之也会带来一些问题^[35]，如使计算变得更加复杂，探索的空间增大，使可训练数据在所有的特征上表现稀疏等。且当特征达到了一定的数量，其准确率就不会再有所提升，如图 2-8 所示。所以要用多少个特征表示比较适合，需通过不断学习来判定，这对人工而言是非常困难的。

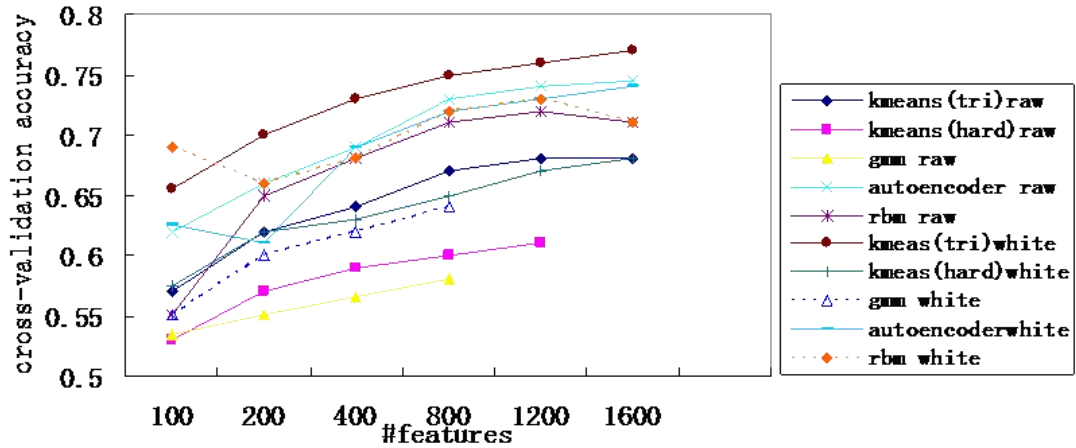


图 2-8 不同特征数采用不同方法的分类效果

2.4 深度学习方法

2.4.1 深度学习概念

深度学习又称为特征学习或者无监督特征学习，最早于 2006 年由 G. E.Hinton 等人^[15]提出，使用给定的训练方法将样本数据训练为包含多层的深度网络结构，从而能够更精准的刻画数据和具有更强的特征学习能力。一般的神经网络中的权值是随机初始化的，这就使得网络经常在局部最小值处收敛。但是深度学习对初始化网路权值进行了优化，采用的是无监督预训练方法，之后再对权值进行调整，克服了此问题。深度学习通过学习一种深层非线性网络结构，将原始的数据特征通过逐层学习方式映射到新的特征空间，从而能够自动化的对层次特征表示进行学习；而浅层学习通常是采用人工设计的方式提取特征的，且学习到的特征不具有层级的结构，仅属于单层的特征，在特征的分类和可视化方面没有深度学习有优势。所以深度学习的思想可简要概述为设计多个层，上一层的输出作为下一层的输入，并经由该方式实现对输入信息的分级表述。这种算法思维与人脑识别物体的过程相类似，先提取目标物的边缘信息，在对此进行更高层次的抽象化表示，进而形成更复杂的视觉形状，识别出该目标物^[36]。这两种方式的学习流程如下图所示。

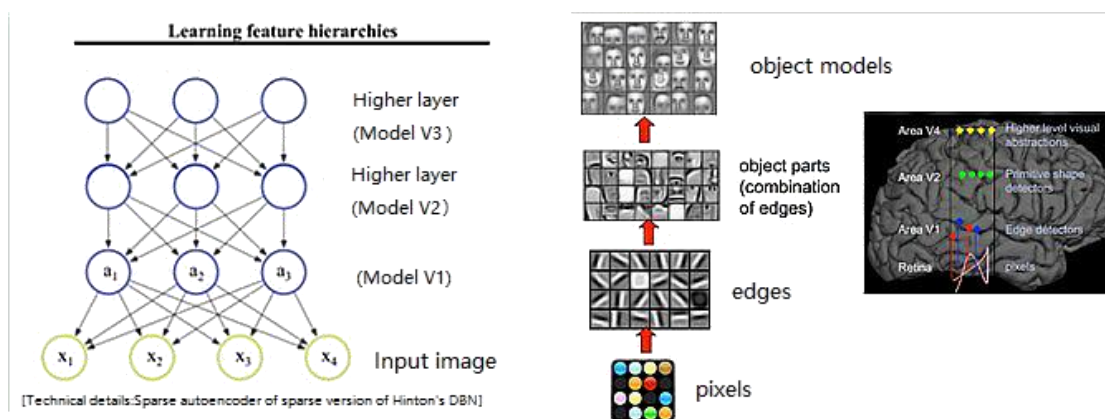


图 2-9 深度学习与人脑视觉皮层学习的流程图

深度学习的本质是从大数据中训练参数并构造出含有多个隐藏层的学习框架，进而获得可更好表达数据的特征，提高预测或分类的精确度。深度学习较浅层学习具有很多独特的特点，如前者含有多隐藏层的网络结构；采用逐级选取特征的方法把原始数据投射到新的特征空间，以易于预测或分类；可以从大数据中提取出具有更加丰富的信息特征表达。一般而言，在对于输入的原始数据在空间中不能直接被分开的情况下，会将数据引入到一个间接特征空间，该空间是线性可分的，例如自动学习，手工编码以及 SVM 等。后两种方式需要具有很高的专业知识才能驾驭，并且对计算资源的需求也十分巨大，因此并不适合做高维输入空间。而第一种方式则可通过深度学习来实现。神经网络的上下层之间存在着非线性变换关系，在将多层神经网络结构运用在深度学习时，通常采用“逐层初始化”的方式来解决对网络进行训练时出现的困难，从而实现对输入更复杂函数的表示^[37]，其实现方式如下图 2-10 所示。

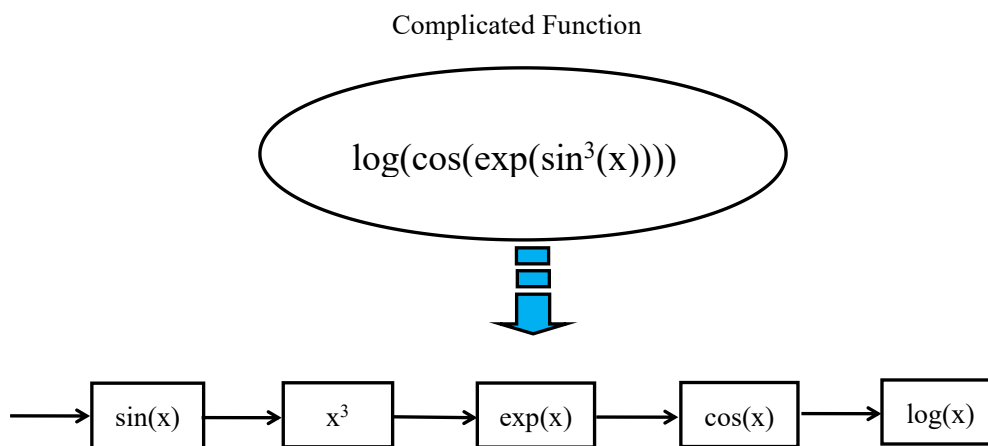


图 2-10 深度学习实现复杂函数逼近流程图

深度学习是神经网络的拓展，可以根据使用者的需求来确认网络层数。深度神经网络不仅结构与传统的不同，其训练方法的区别也很大，传统的神经网络的参数初始值是随机设定的，常采用 BP 计算网络的残差并训练网络，直到收敛。这种

方法在网络加深时,就会出现弊端,如由于目标函数初始值的随机设置,造成网络收敛于局部最小值;网络结构的深化,使得残差向上传播时损失比较大,最终致使梯度出现扩散的现象。所以传统的方法大都是对浅层有效而对深度网络作用不大。深度学习选用的是逐层贪婪无监督训练法,前一层的训练结果作为下一层的输入数据,最后当训练完所有层后,采用有监督的学习模式对整个网络进行调参。传统的神经网络模型是浅层学习方式,其隐藏层一般不会超过两层,而深度学习方法的隐藏层通常是两层以上,两种结构模型^[38]表示如下图 2-11 所示。

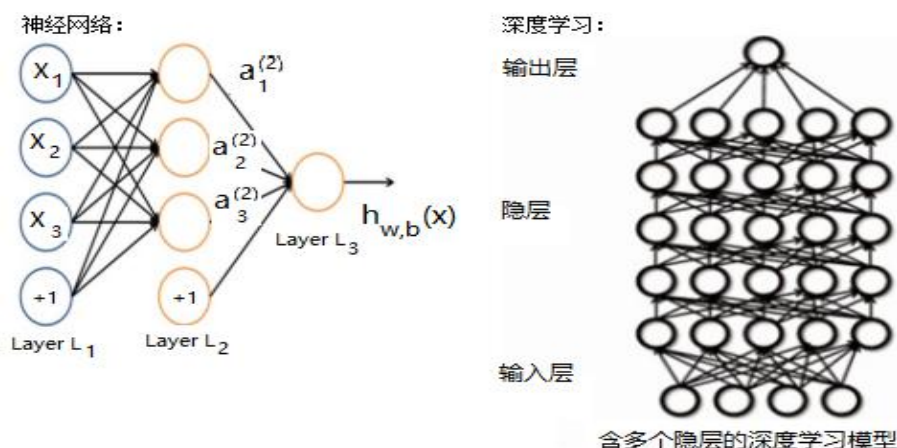


图 2-11 神经网络和深度学习模型

2.4.2 深度学习训练过程

在深度学习中不再采用神经网络中的 BP 算法而采用无监督学习算法对其网络进行训练^[39],具体原因:反馈调整时,由顶层到底层的梯度会愈来愈稀疏,导致误差的校正信号越来越小;选用随机初始化参数的方式会导致基于梯度的训练收敛于局部极值,不能找出代表全局网络的最优解,且随着网络层数的增加,其泛化性能降低,致使得出的训练结果不好,甚至还不如只有一两层的浅层神经网络;BP 算法需要采用有标签的数据进行网络训练,但现实中很多数据是没有标签的,如下图 2-12 所示。

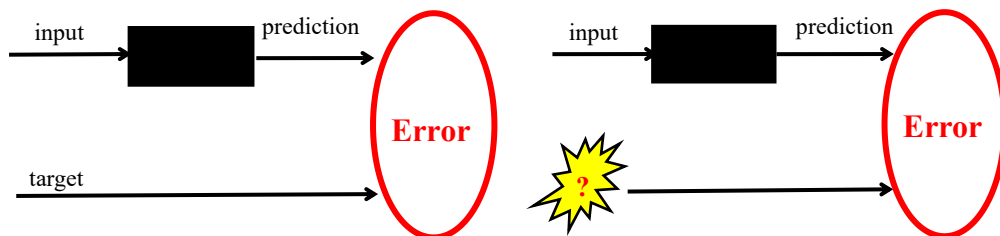


图 2-12 数据有无标签的训练方式

深度学习的训练过程主要分为以下两个环节^[40]:

1)由下而上的无监督学习:该环节实际上是一个通过采用无监督学习方式来

初始化网络参数的过程,与传统神经网络随机初始化参数方法不同,通过该方法可以使得到的初始值与全局最优更接近,所以可以达到更好的识别或分类效果。在该过程中需要分两步完成,第一步需要先逐层地构建单层的神经元;第二步应用 wake-sleep 算法对网络进行逐层调优,且每次仅调整一层,其流程如图 2-13 所示。

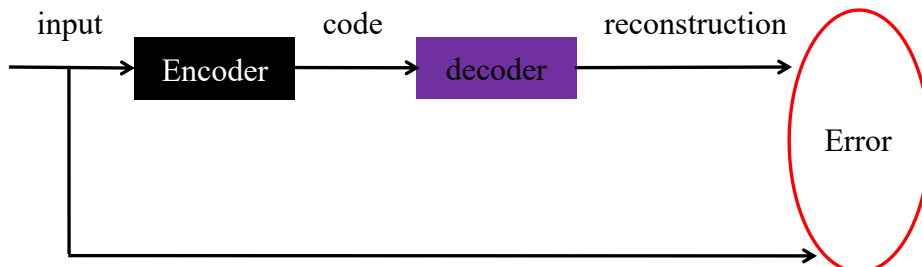


图 2-13 wake-sleep 算法流程图

wake-sleep 算法翻译为中文是醒-睡的意思,由此可以看出它主要分为 wake 和 sleep 两个部分,每个部分的简要描述如下:

①wake: 又称为认知过程,通过向下的 input(输入特征)和上层的 encoder(认知)权重生成各层的 code(抽象表示),然后应用当前(decoder)权重生成一个 reconstruction(重建信息),计算 reconstruction 和 input 之间的残差,并通过梯度下降法修正每层之间的下行 decoder。可简要概述为“一个从现实到概念的过程”。

②sleep: 又称为生成过程,经由上层概念(code)与向下的生成权重(decoder)创建下层的状态,再使用 encoder(认知权重)生成一种抽象的表达。然后计算上层概念与新建抽象数据之间的残差,并运用梯度下降法修正层与层之间的认知权重(encoder)。可简要概述为“一个从概念到现实的过程”。

2)由上向下的监督学习:当完成第一步训练后,把一个分类器(如 SVM、Bayes 等)加在网络结构的最顶端,然后采用有监督的学习方法,对带有标签的数据进行训练和调参。其学习方式的结构示意图如下图 2-14 所示。

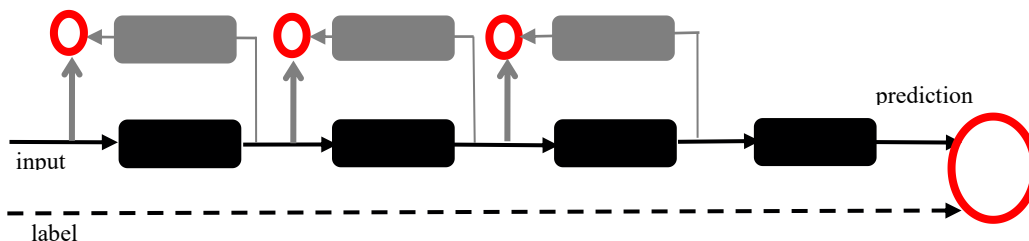


图 2-14 由下向上的监督学习流程图

2.5 深度学习的模型

人的视觉系统对信息的处理是分级的,对于摄取对象,首先是从他的边缘特征到具体的外形(或其他目标),然后到更高层次的目标特性,也就是将底层特征

结合为高层次的特征,使得特征从低到高的表现方式愈来愈抽象。深度学习借鉴此方法,通过堆叠多个单层网络来实现模型的搭建,按照这个单层网络中所采用的编码与解码方式,可将此网络分为仅有编码部分、仅有解码部分、两者皆有。编码器通过一种从下到上的方式,把数据特征从输入层映射到隐藏层;解码器为了能够使运行的结果尽量靠近原始输入,采用的是将特征隐藏层映射到输入层。通过堆叠多个编码器组成前馈深度网络 FFDN,同理将多个解码器进行堆叠可以得到一个反馈深度网络 FBDN 模型。而将多个编码器和解码器一起进行堆叠就可以获得一个双向深度网络(bi-directional deep networks, DDN)模型,其中在该模型中的每层网络中可以只有编码或解码的过程,也可以两者都包含在内^[41]。由于本文主要研究的是基于 DBN 和 CNN 的中国绘画图像分类,所以将主要介绍这两方面的内容。

2.5.1 前馈深度网络

在前馈神经网络中,信息是单向流动的,不存在封闭的环路,从输入层经过一个或者多个隐藏层到输出层。多层感知机与 CNN 是常用的前馈深度网络模型,这里以 CNN 为例进行讲解。CNN 是神经认知类的扩展,主要用来识别位移、缩放及其他变形的二维图像而特殊设计的一个多层感知器。每层 CNN 由三个阶段组成,分别为卷积、非线性变换、下采样^[42],其每阶段具体功能描述如下:

(1)卷积阶段:对输入信号进行特定模式的观测是通过提取信号的不同特征来实现的,这种观测模式即是卷积核,为了使同一个输入的特征图在权值上达到共享,使所有的卷积核都可以检验到特征图中任意位置的指定特征,采用不一样的卷积核来实现操作,进而更好地提取出输入特征图上的不一样的特征。

(2)非线性阶段:一般应用非线性变换的方法从卷积阶段获得的特征中进行选取,从而规避线性方式表现能力不足等问题。在该过程中直接把从卷积阶段所获取的特征当作输入数据,并通过非线性映射 $R=h(y)$ 组成计算层。在传统 CNN 中,非线性操作通常选用 sigmoid、tanh 等饱和非线性函数,而近些年的 CNN 所应用的非线性操作更多是不饱和非线性函数 ReLU。采用 ReLU 的网络结构在对梯度下降进行训练时,相比传统饱和非线性函数具有更好的收敛速率,所以在全面训练网络时,其训练速率也远远超过传统的方法。

(3)下采样阶段:独立地操作每个特征图像,经常有两种运用方法,一种是最大池化(max pooling),另一种是平均池化(average pooling)。图像经过池化后,虽然输出的特征图分辨率减少了,但该操作拥对原数据较好的畸变容忍力。使特征映射的输出平移和其他形式的变形的敏感度下降的作用,并降低了上层的计算

复杂度。部分 CNN 可以在没有下采样操作的情况下，通过在卷积过程中配置卷积核(Kernel)窗口滑动步长至少为 2 的做法实现减少分辨率的效果。

CNN 的核心主旨是通过把局部感受野、权值共享以及时间或空间的下采样架构思想相联结，对于位移、缩放及其他形式扭曲具有不变性。通常可直接将图像输入到网络，规避了在传统识别算法中图像的重构和繁杂特征的选取过程。在输出层通常选用的连接方式是线性全连接，分类算法为 softmax，其中减弱网络模型的复杂度是通过权值共享和局部感受野来实现的^[43]。究其根本就是一种从输入到输出的映射，其训练过程主要分为以下两个步骤。

第一步，前向传播：

a)从样本集中取出一个样本 (X, Y_p) ， X 是输入向量， Y_p 是理想输出。

b)计算相应的实际输出 O_p 。

在网络结构中当前层的输入数据来自于上一层的输出，当前层的输出是通过激活函数对输入的数据计算而得，依次这样向后传递，直到最后一层。当前层的输出方法如式(2-6)所示。

$$X^l = f(W^l X^{l-1} + b^l) \quad (2-6)$$

上式 X 是当前输出层， l 代表层数， f 代表选取的当前层的激活函数(常采用 sigmoid 或双曲正切 tanh 函数)， W 为当前层的权值， b 为当前层的一个偏置。

在前向传播的过程中，经过轮流计算卷积层和亚采样层，来完成特征的学习，实现对输入数据较高的畸变容忍能力。通常 C 层(Convolution Layers，卷积层)是特征提取层，通过对前一层的输出特征图(feature-maps)卷积和函数激活，生成该层的输出结果。其公式如式(2-7)所示。

$$X_j^l = f\left(\sum_{i \in M_j} X_i^{l-1} * k_{ij}^l + b_j^l\right) \quad (2-7)$$

每个卷积层后面都紧跟一个子采样层 S (Sub-sampling Layers)，以实现局部平均和子抽样，将卷积层输出的 feature-maps 作为该层的输入数据，然后由公式(2-8)得出该层的输出 feature-maps。

$$X_j^l = f(\beta_j^l \text{down}(X_j^{l-1}) + b_j^l) \quad (2-8)$$

其中，down(.)为下采样函数， β 和 b 分别为乘性偏置和加性偏置。该层一般会对输入 $n \times n$ 的 maps 求区域和，在两个维度上使得输出的结果分别缩减了 n 倍。

第二步，后向传播：

a)计算实际输出 O_p 与其相对应 Y_p 的差值；

b)反向传播极小误差，继而对权的矩阵进行调整。

在反向传播进行调参时，经常采用平方误差代价函数，在 N 个训练样本 c 个类别中，平方误差代价函数如公式(2-9)所示。

$$E^N = \frac{1}{2} \sum_{n=1}^N \sum_{k=1}^c (t_k^n - y_k^n)^2 \quad (2-9)$$

其中， E 为输出的误差， t_k^n 指第 n 个样本所对应第 k 维的标签， y_k^n 指第 n 个样本相对应的第 k 个输出。

2.5.2 反馈深度网络

反馈网络区别于前馈网络，不编码输入的信号，经过反卷积或者学习数据的基反解输入信号，其中反卷积网络、层次稀疏编码网络是其常见模型。DN 观念与 CN 相似，只是在框架组成和实现方式上不同。CN 所获得的多层信息是经由一种由底向上的方式，通过对输入数据卷积、非线性变换和下采样三个步骤来实现。而 DN 采用的方式为由顶向下，经由对滤波器所学习到的卷积特征进行融合，实现对输入信息的重建^[44,45]。层次稀疏编码与 DN 类似，不同之处在于前者选用矩阵乘积的形式分析图像，后者选用的是卷积矩阵。

(1)单层反卷积网络：一种经过先验式的学习，稀疏分解和重建信息的正则化方法。

(2)反卷积网络：通过堆叠多层的单层反卷积网络构成。该网络一边对滤波器组实行学习，一边推断特征图，其中处于 L 层的特征图和滤波器是通过反卷积计算分析 $L-1$ 层相应特征图和滤波器获得。在训练 DN 时，应用一组具有差别性的 y 信号去训练，轮流优化滤波器组 f 和特征图 z 。训练起始于第一层，在优化时选用贪心算法逐层向上，且对每层之间进行独立的优化。能够经由最优化输入信号来推算特征，而非采用编码器求近似，如此可更好地表征隐藏层，便于对目标信息的分类或再建，是它的一大亮点。

(3)自动编码器：在一个神经网络中，假定输出等同输入(理想状态下)，而后经过训练调解参数获得每层的权重，由此可获得输入数据的几种表达方式，也就是特征的表达。自动编码就是将自动学习获得的特征添加在原有的特征上，以此来提升对目标分类或识别的准确率。若是再加入少许的限制条件，就能够获取新的深度学习方法。例如将稀疏性制约加入到 AE 上，便可获得稀疏自动编码器 SAE。

(4)稀疏自动编码器：SAE 同 CNN 的区别在于它应用的是一种无监督学习方式。这种模型的基本组成单元是 AE，输入的特征 X 经过一定的条例和训练方法进行编码，使用低维向量来表示原始的特征。AE 首先组建与传统神经网络结构相似的层次模型，同时也假定其输出 Y 等于输入 X ，然后通过重复调整参数获得

最终的网路参数值。若在 AE 的基础上,应用稀疏性来制约隐藏层的神经元,即使大部分节点的值 0 或者近似为 0,就可获得 SAE。为了确保网路结构具有一定的价值性,隐层神经元的个数通常情况下要小于所输入的数据个数。如 PCA,通过选取消除冗余的方法来降低空间维数,从而能够实现使用更少的特征来尽量描绘完好的数据信息。实践证明将学习到的多种隐层特征协同原始特征,能够很好地提升算法的准确度^[46]。

AE 参数训练方式有多种,可以选用任意持续化的训练方式对参数进行训练。但因其模型组织不偏向生成型,所以不可以在确认模型的合理性时采用联合概率等这种定量方式。在优化深度学习算法中,稀疏性约束之所以愈来愈重要,是因为跟深度学习的特点有很大关系。太多的训练参数可能致使训练流程变复杂,使输入数据的维数要远低于输出维数,从而导致生成很多额外数据讯息。添加一种稀疏性的制约,不仅可以让学习到的特征更有意义,还可以让这种结构与人脑神经元的稀疏特性相符。

2.5.3 双向深度网络

双向网路由编码器和解码器组成,所以它是通过联合前馈网路和反向网路的方法进行训练的。每层在训练时采纳一种贪心算法,通过计算输入信号 IL 和权值 w 来生成 $IL+I$,并将它传送到下一层。同理经由 $IL+I$ 作为输入信号与等同的权值 w 产生重构信号 IL ,并将其投射到输入层,每一层的网路经过连续缩短 IL 与 IL 之间的误差来进行训练。每层网路在历经预训练以后,网路经由反向迭代误差法对权值进行全面的调节。近似反馈网路的训练方法中,对各层网路的预训练可看作是对输入信息再建编码和解码的过程。常见的双向深度网路有深度玻尔兹曼机、DBN、栈式自编码器等^[47],此处以 DBN 为例,它包含多个隐藏单元,主要由 RBM 和 Bayes 构成,其中 RBM 模型结构如图 2-15 所示。

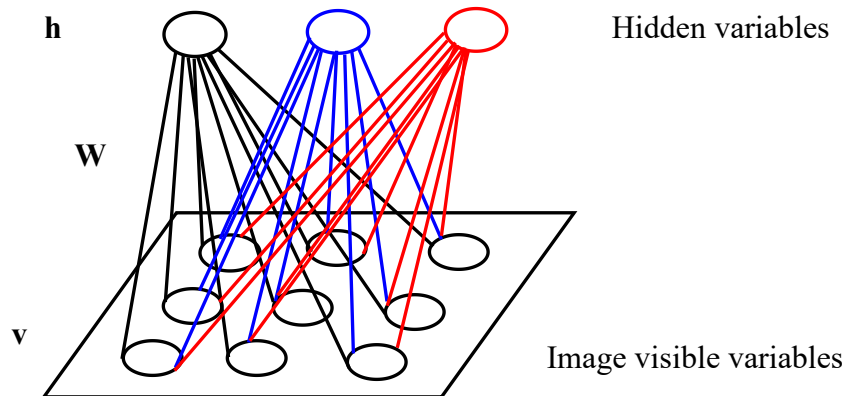


图 2-15 RBM 模型结构图

RBM 是一种随机的递归神经网络,从图 2-15 可以看出该模型是假设在一个

二部图中，包含一个可视层 v 和一个隐藏层 h ，通过权值 w 对这两层进行连接，且在同一层中节点之间无连接。假定每个节点都符合随机二值(0,1)，全概率分布 $p(v,h)$ 符合 Boltzmann 的条件。RBM 之所以能够成为深度学习的一种，原因是在这个二部图的结构中，在已知 v 的条件下，存在于隐藏层的每个节点是独立的，即可得出隐藏层节点的概率：

$$p(h|v) = p(h_1|v) \cdots p(h_n|v) \quad (2-10)$$

同样我们也可以由 h 推出 v ，且处在 v 层的节点也是独立的。这样在全部的 v 、 h 均符合 Boltzmann 分布时，我们能够在输入 v 后，应用 $p(h|v)$ 获得隐层 h 的概率，这之后再经过参数的调解，尽量能够实现由隐藏层得出可视层与该隐藏之前的可视层一致，那么可以称此隐藏层为可视层的另一种表现形式。由此可将隐层的数据作为下一层可视层的输入，如此隐层和可视层交替进行。该模型学习的流程如下图 2-16 所示：

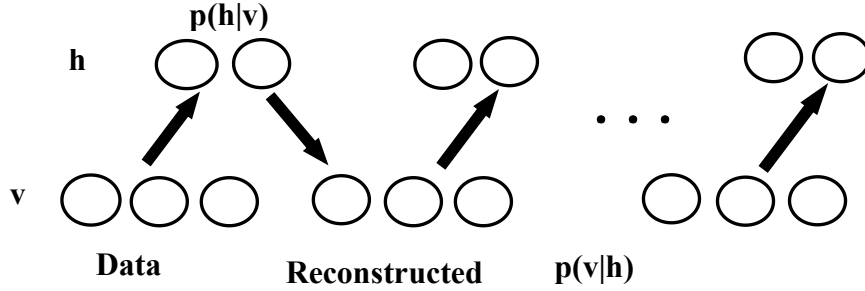


图 2-16 RBM 学习流程图

RBM 如何确定可视节点与隐节点的权值，这需要做一些数学分析，具体描述为在 RBM 模型中，RBM 可视层与隐藏层节点 $\{v,h\}$ 之间的联合组态能量函数如公式(2-11)所示：

$$E(v,h;\theta) = -\sum_{ij} w_{ij} v_i h_j - \sum_i b_i v_i - \sum_j a_j h_j \quad (2-11)$$

上式 $\theta = \{w,a,b\}$ 是参数的合集， w 是位于 v 、 h 之间的权重，向量 a 与 b 分别是可视层和隐藏层的偏置值。如此某组态能够由 Boltzmann 分布和其能量来确定联合概率分布，其概率为：

$$p_{\theta}(v,h) = \frac{1}{Z(\theta)} \exp(-E(v,h;\theta)) = \frac{1}{Z(\theta)} \prod_{ij} e^{w_{ij} v_i h_j} \prod_i e^{b_i v_i} \prod_j e^{a_j h_j} \quad (2-12)$$

其中 $Z(\theta) = \sum \exp(-E(v,h;\theta))$ 为归一化因子，即配分函数。由于存在于隐藏层的节点之间是相互独立的，所以可推出可视层节点的条件概率为：

$$p(v|h) = \prod_i p(v_i|h) \quad (2-13)$$

然后对上述公式进行因式分解，可得出基于指定隐藏层，可视层第 i 个节点是 1 或 0 的概率为：

$$p(v_i = 1 | h) = \frac{1}{1 + \exp(-\sum_j w_{ij} h_j - b_i)} \quad (2-14)$$

在指定可视层 v 的条件下，同样可得出隐藏层的概率：

$$p(h | v) = \prod_j p(h_j | v) \quad (2-15)$$

$$p(h_j = 1 | v) = \frac{1}{1 + \exp(-\sum_i w_{ij} v_i - a_j)} \quad (2-16)$$

训练 RBM 网络参数的过程可以看成是估量最大数似然函数下的模型。所给定的样本集符合独立分布条件： $D = \{v^{(1)}, v^{(2)}, \dots, v^{(N)}\}$ ，需要学习参数 $\theta = \{w, a, b\}$ 求解最大似然函数 $L(\theta)$ 的估计：

$$L(\theta) = \frac{1}{N} \sum_{n=1}^N \log p_{\theta}(v^{(n)}) - \frac{\lambda}{N} \|w\|_F^2 \quad (2-17)$$

通过最大化 $L(\theta)$ 求解 RBM 参数，对最大似然函数 $L(\theta)$ 求偏 w 导数，如下公式所示：

$$\frac{\partial L(\theta)}{\partial w_{ij}} = E_{p_{data}}[v_i h_j] - E_{p_{\theta}}[v_i h_j] - \frac{2\lambda}{N} w_{ij} \quad (2-18)$$

RBM^[48]通过使用对比散度算法来提升训练速度。其基本思想是用吉布斯采样，把训练数据作为输入的初始状态，经过少数次数的状态转移，并对转移后的状态进行估算。RBM 在训练的过程中，原始数据的数目一般作为可视层数据的大小，隐藏层的大小通过人为设定。具体步骤如下：

1. 初始化

- (1) 给定训练样本集合 $S(|S| = n_s)$;
- (2) 设定训练周期、学习率、CD-k 参数分别为 J 、 η 、CD-k;
- (3) 指定单元数量在可见层是 n_v ，隐藏层是 n_h ;
- (4) 初始化偏置向量 a 、 b 和权值矩阵 w 。

2. 训练

for $iter=1, 2, \dots, J$ Do

{

1. 调用 $CDK(k, S, RBM(w, a, b); \Delta w, \Delta a, \Delta b)$ ，生成 $\Delta w, \Delta a, \Delta b$;

2.刷新参数: $w=w+\eta(\frac{1}{n_s}\Delta w)$, $a=a+\eta\frac{1}{n_s}\Delta a$, $b=b+\eta\frac{1}{n_s}\Delta b$ 。

}

若添加隐藏层的个数, 可以获得深度玻尔兹曼(Deep Boltzmann Machine, DBM); 当在可视层附近的区域采用 Bayes 网络, 较远处采用 RBM 时, 就构成了深度信念网络^[49]。这两种模型的结构图如图 2-17 所示。

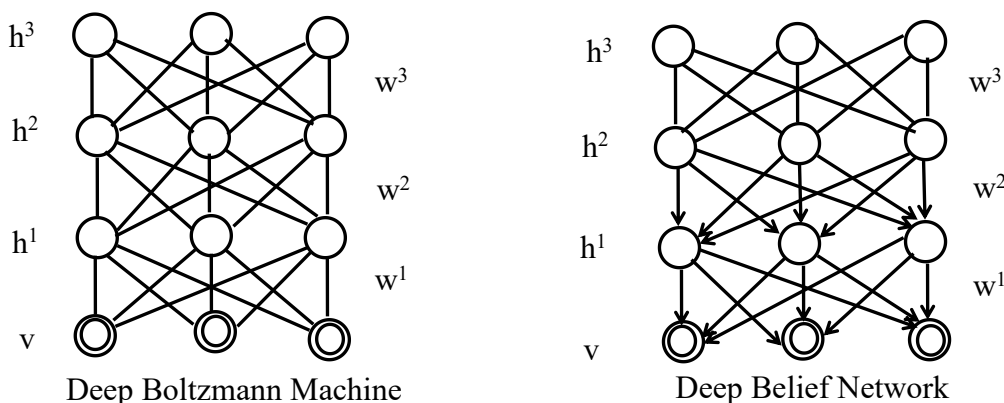


图 2-17 DBM 和 DBN 两种模型结构图

DBN 将深度学习和特征学习进行了融合, 是一种生成概率模式, 所建立的联合分布存在于观察数据和数据标签之间, 这种方式与神经网络的传统判别模式是不同的。DBN 的训练分为两个部分: 预训练(pre-training)、微调(fine-tuning)。

预训练阶段: 对每层 RBM 训练的方式采用的是逐层训练, 当该层是可见层时, 它的输入数据来自于低它一层的隐层输出。

微调阶段: 在最后一层的 BP 网络应用有监督的学习方式进行训练, 对整个 DBN 的权值通过逐层向后的方式计算出实际输出与期望输出的误差, 从而实现对权值的微调。

由上可知, RBM 的训练过程实际上可看成是对多层 BP 网络权值的初始化, 并由此解决了 BP 在对权值随机初始化时, 所造成的训练时间过长和易导致局部最优的问题。

2.6 本章小结

本章节首先对图像的分类及相关技术做了简要概述, 然后分别对浅层学习和深度学习作了简要介绍, 并诠释了引入深度学习的缘由。在这之后又阐述了深度学习的训练过程及常见模型, 其中详细阐述了 CNN 和 DBN, 为下文所提出的改进算法奠定基础。

第三章 基于 DBN 的中国绘画图像分类算法

3.1 引言

卷积深度信念网络 CDBN 是 DBN 的一个拓展。由于 DBN 的输入数据只是简单的一维向量模式，没有考虑到数据中所蕴含的二维结构信息。CDBN 不仅考虑了 DBN 所忽略的信息，而且还通过卷积操作，降低了网络参数的计算复杂度，从而更好地对数据进行特征提取。此外利用最大概率池化模型、稀疏正则化使得学习到的特征能够更好的应对图像的变形扭曲和避免模型超完备化，从而学习到更加有效的深度特征，实现良好的识别效果。本章首先对 CDBN 的相关概念做介绍，然后引出本文方法，最后通过实验验证本文方法。

3.2 DBN 的构成

前文已对 DBN 作了简要介绍，这里主要介绍的是 DBN 的构成，其结构图见图 2-17。首先多层神经元可构成 DBN，并将这些神经元分为显性和隐性。其中显性神经元作为输入的接收端口，隐性神经元用作为特征提取(特征检测器)的窗口。由于 DBN 的组成元件是 RBM，所以它是逐层训练的，每层用向量式的数据推断隐藏层，接着将该隐藏层当作下一层的输入数据。

由上述可知 DBN 的整体网络结构经过训练神经元之间的权重，可使训练数据依据最大概率模式生成。这种生成模式不仅可以有助于 DBN 对特征的识别和数据的分类，还可以用来生成数据。但是当图像是高维的图像时，要求算法能够合理的建模，计算简便；另一方面由于对象常常分布在图像的局部，所以要求特征表示对输入的局部变换具有不变性。面对这样的情景，只用 DBN 是很难处理好的。而由 CRBM 构成的 CDBN，不仅可以通过 CRBM 模型区达到生成模型的不变性和对高阶数据统计性的学习，还可利用卷积操作实现网络计算复杂度的降低和对畸变具有较好的容忍力等。如此，我们可以将卷积操作引入到 DBN 模型中，实现局部不变性和对高阶统计特征的学习。

3.3 卷积限制波尔兹曼机 CRBM

在前文中我们已对 RBM 做了介绍，这里重点介绍一下 CRBM。对于受限玻

尔兹曼机，图像的每个像素跟隐层每个结点进行全连接，训练参数非常之大，模型复杂度较高。而且对于二维图像的处理，我们希望算法在学习特征的初始阶段尽可能保留二维结构信息。卷积限制波尔兹曼机 CRBM 与 RBM 类似^[50]，只不过 CRBM 的隐含层与可视层之间的权值是共享的，使得这种网络结构与生物的神经网络更加相似，从而降低了模型的复杂度，减少了参数的个数，而且该网络结构融入了卷积的特性，使其具备平移不变性，如移位、大小等比缩放、旋转等方式，同时还具有较好的抗噪、抗干扰能力。下图 3-1 展示的是一个采用最大池化的 CRBM。 v 为可视层，代表输入数据，其大小为 $N_v \times N_v$ 。 H 表示隐藏层，包含 k 组，每组用二进制单元矩阵 $N_h \times N_h$ 进行表示，这样共计 $N_h^2 \times k$ 个单元。每组是在输入的数据经过一个 $N_w \times N_w$ 的滤波器形成的($N_w = N_v - N_h + 1$)，该滤波器的权值在隐含层中进行共享，每个隐藏组有一个偏置 b_k ，这样所有的可视层单元共享一个偏置 c 。 p 为池化层，包含 k 组，每组二进制单元有 $N_p \times N_p$ 个。在每组中， p 通过一个小的整数(如 2 或 3) C 来对隐藏层的表示进行压缩，这样隐含层 h 将被分成大小为 $C \times C$ 的块，从每块中提出一个二进制的池化层单元 P_k ，所以 $N_p = N_h / C$ 。

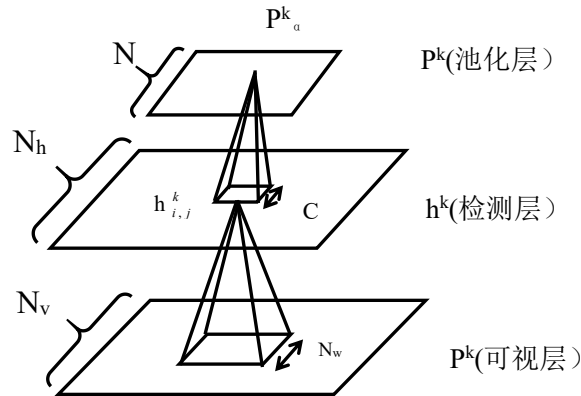


图 3-1 CRBM 应用最大池化的架构图

CRBM 与 RBM 的能量函数相类似，可以定义为公式 3-1：

$$E(v, h) = - \sum_{k=1}^k \sum_{j=1}^{N_h} \sum_{r,s=1}^{N_w} v_{i+r-1, j+s-1} - \sum_{k=1}^k \sum_{i,j=1}^{N_h} h_{ij}^k - c \sum_{i,j=1}^{N_v} v_{ij} \quad (3-1)$$

进一步化简得：

$$E(v, h) = - \sum_{k=1}^k h^k \bullet (\tilde{w}^k \times v) - \sum_{k=1}^k b_k \sum_{i,j} h_{i,j}^k - c \sum_{i,j} v_{ij} \quad (3-2)$$

其中 $h_{i,j}^k$ 指隐藏层中第 k 个子层的单元， b_k 指其第 k 个子层的偏置， w^k 指第 k 个卷积核。同理，该模型的隐层和显层的条件概率与 RBM 类似，分别如下公式所示：

$$p(h_{ij}^k = 1 | v) = \sigma((\tilde{w}^k \times v)_{ij} + b_k) \quad (3-3)$$

$$p(v_{ij} = 1 | h) = \sigma((\sum_k w^k \times h^k)_{ij} + c) \quad (3-4)$$

其中 σ 为 s 型函数。通过吉布斯采样反复推断隐层，重建可视层。CRBM 的训练过程与 RBM 一样，可以参照上文。

3.4 卷积深度信念网络 CDBN

本文中所引用的 CDBN 的网络架构与 DBN 类似，不同的是我们将卷积限制波尔兹曼机作为基本单元进行叠加，并将最大概率池模型 (probabilistic max-pooling) 加入到 CRBM 的顶层，然后用类似 DBN 的贪婪方式逐层训练网络。

3.4.1 最大概率池模型

为了能对大的图像进行好的描述，通常的做法是聚合统计位于不同区域的聚合特征，如在 CNN 中，虽然在卷积过后，可以减少不少的训练参数，但它的输出向量的维数很大，对后面分类器的设计带来了困难。而池化 pooling 操作能够计算一个位于图像区域中代表某个特定特征值的最大值(或平均值)，一方面缩小特征图，简化网络计算复杂度。另一方面提取出主要的特征，从而实现特征的压缩，改良我们所期望的结果(不易过拟合)。主要的池化操作可以分为最大值池化(max pooling)和均值池化(mean pooling)，其一般操作流程(以 max pooling 为例)如下图 3-2 所示。

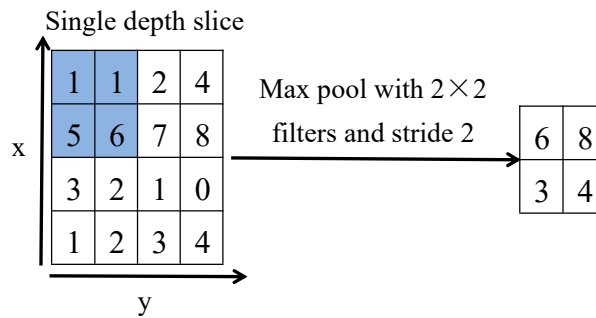


图 3-2 最大值池化(max pooling)

在上图中 max pooling 是在每一个区域中寻找最大值，应用的滤波器大小为 2×2 ，移动的步长 stride 为 2，该图的右侧代表的是最后从原特征图中提取到的主要特征图。由于池化操作缩小了特征图，就可能会使网络结构的准确度受到影响，不过我们可以通过加大网络模型的深度来减弱该现象。

由于最大值池化操作是确定的，而且其前向操作不可逆，不能够很好地进行概率推理，所以在本文中用的是最大概率池模型(Probabilistically Max-pooling)，如图 3-1。该模型结构的方向是向前的，有 K 组单元存在于检测层和池化层 ($k \in 1, 2 \dots k$)， $N_p \times N_p$ 组二进制单元存在于池化层，其中检测层映射到池化层块的大小为 $C \times C$ 。在检测层的池化块以及与他相对应的池化单元潜伏着一个限制条件，即最多有一个单元在检测层被激活，而在池化层单元被激活的条件是当且仅当激活了一个检测单元。此时 CRBM 最大池化的能量函数可定义为：

$$E(v, h) = - \sum_k \sum_{i,j} (h_{i,j}^k (w^k \times v)_{i,j} + b_k h_{i,j}^k) - c \sum_{i,j} v_{i,j}, \quad (3-5)$$

$$\text{when } \sum_{(i,j) \in B_\alpha} h_{i,j}^k \leq 1, \forall k, \alpha.$$

检测层 h 的第 k 组接受的输入是从可视层 v 由下到上的信号输入，其计算公式为：

$$I(h_{ij}^k) = b_k + (\tilde{w}^k \times v)_{ij} \quad (3-6)$$

继而分块池化这个隐藏层的单元，如果设块 α 中的一个隐单元是 h_{ij}^k ，那么此时的条件概率是：

$$P(h_{i,j}^k = 1 | v) = \frac{\exp(I(h_{i,j}^k))}{1 + \sum_{(i',j') \in B_\alpha} \exp(I(h_{i',j'}^k))} \quad (3-7)$$

$$P(p_\alpha^k = 0 | v) = \frac{1}{1 + \sum_{(i',j') \in B_\alpha} \exp(I(h_{i',j'}^k))} \quad (3-8)$$

在公式(3-7)中，隐层激活概率是从可见层获得的，公式(3-8)的概率属于与此相对应的池化单元抑制概率。在整个池化后的检测单元中经过(3-7)的一个被激活概率的单元后，最少有一个被激活的概率为：

$$P(pool_{on}) = P(anyhide_{on}) = \sum P(on) = \sum_{(i',j') \in B_\alpha} \frac{\exp(I(h_{i',j'}^k))}{1 + \sum_{(i',j') \in B_\alpha} \exp(I(h_{i',j'}^k))} \quad (3-9)$$

由于只存有两种状态，即抑制和被激活，所以这两种状态的概率之和为 1，也就是公式(3-9)和(3-8)相加为 1。

3.4.2 稀疏正则化

在训练 CRBM 时，计算确切对数似然估计的梯度比较困难，这时采用对比散度 CD 可以有效地帮助解决该问题。但是训练时，实际上表示数据的大小要比输入时的尺寸大的多，所以可以说我们的模型是过完备的。过完备的学习模型在

学习的时候具有一定的风险性,在这里我们采取的解决方法是将稀疏惩罚项添加到网络结构的运算中,表示只有一小部分的单位应该相对于给定的刺激被激活。其培训的目标是满足以下公式(3-10):

$$L_{\text{likelihood}}(w, b, c) + L_{\text{sparsity}}(w, b, c) \quad (3-10)$$

使 w, b, c 最小化。其中 $L_{\text{likelihood}}$ 是一个负对数似然函数,用来衡量 CRBM 算法与输入分布数据的近似程度, L_{sparsity} 是一个惩罚项,用来约束隐层单元的平均稀疏激活度。这种稀疏正则化(sparsity regularization)具有限制网络规模的能力,往往能够更好地表示数据的特征。

3.4.3 CDBN 模型结构

本文中所采用的算法 CDBN 是由上述采用最大池化的 CRBM 堆叠而成,并在 CRBM 中引入稀疏正则化以缓解模型过度完备。由于 CDBN 同 DBN 一样,采用的是一种无监督的贪婪逐层算法,所以过多的隐藏层会导致训练时间过长,所以本文中的 CDBN 模型结构采用的是两层 CRBM。其模型结构如下图 3-3 所示:

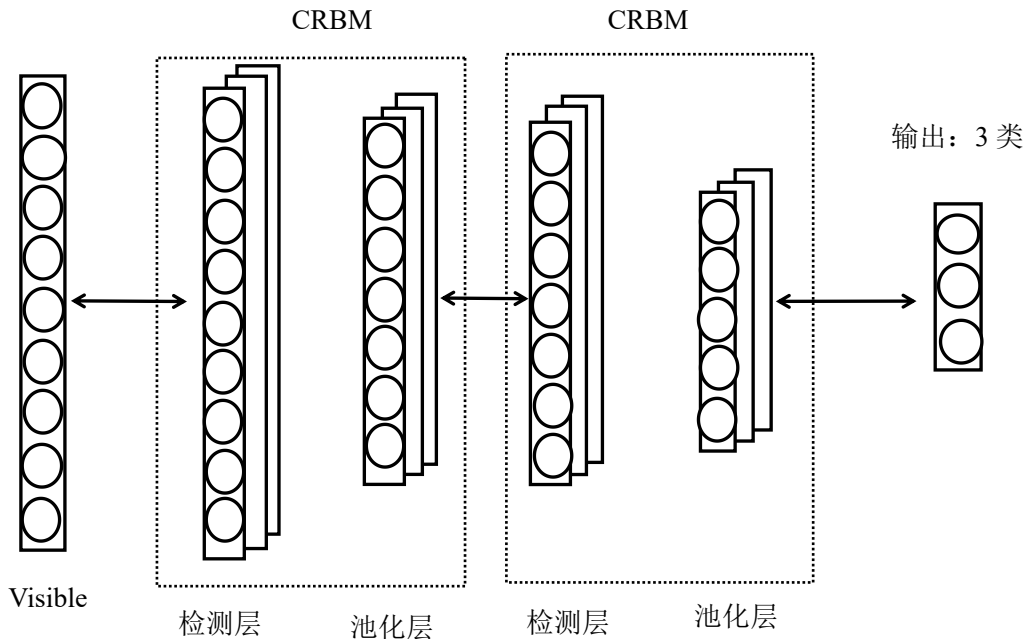


图 3-3 CDBN 模型结构

在 CDBN 的训练过程中,首先将要输入的数据转化为向量模式,然后再经过隐藏层的卷积和汇聚操作的交替进行,提取出原始样本的低层特征表示。引入多层的 CRBM 可以弥补没有自顶向下的输入而造成的损失。在此基础上再进行高层特征的提取。该过程进行的是无监督学习方式,应用 CDBN 可以提取出数

据中的高维统计特征。其中将输出的每层池化结果直接输入到下一层 CRBM，当训练完一层网络后，其权值不做改动并在被激活后作为下一层的输入数据。以此类推，用该种方式训练网络。

3.5 实验结果与分析

3.5.1 实验所用数据库及目的

由于中国绘画分布比较广泛，分类不明确，每类数量有限，且互联网上没有专业的中国绘画系统管理机构供人们欣赏与查阅，所以实现对中国绘画的分类具有重大的意义。本文根据绘画题材的不同，将其分为花鸟、人物、山水三大类。其前期数据的划分是由几位具有硕士学位的研究生进行标注的，然后再对每位标注者的结果作总结。由于花鸟、人物、山水这三类所具有的特性差别比较显著，所以在人工鉴别时就比较容易，出错率就会比较小，最后总结出来的数据分类结果基本上是一致的。

由于文中深度学习所分析的对象是基于大数据的，且中国绘画的分布不集中、数量有限等特性，所以试验中的数据是经过多种途径得来的，主要是来自于《珍好画业》的友情提供和网页的手动搜索。实验中的数据花鸟、人物、山水共 2400 幅，每类 800 幅。其中不同种类的划分是手动归类统计而成，样例表示如图 3-4 所示。



图 3-4 样例示意图

3.5.2 实验评价标准

在图像分类中，测量标准一般有准确率(precision)、召回率(recall)、ROC 曲线等，本文中所用的标准是准确率，此方法的计算公式如下：

$$\begin{aligned}
 \text{precision} &= \frac{\text{ImageNumber}_{\text{true}}}{\text{ImageNumber}_{\text{true}} + \text{ImageNumber}_{\text{false}}} \times 100\% \\
 &= \frac{\text{ImageNumber}_{\text{true}}}{\text{ImageNumber}_{\text{all}}} \times 100\%
 \end{aligned} \tag{3-11}$$

在上式中, $\text{ImageNumber}_{\text{true}}$ 代表测试数据集中取得分类正确的图片数量, $\text{ImageNumber}_{\text{false}}$ 表示测试错误的图片数量, $\text{ImageNumber}_{\text{all}}$ 表示测试数据的总量。为检验实验结果的客观性, 输入的绘画图像为从总的数据中随机抽取的, 分类结果的正确率采用的方法是交叉验证法 CV(Cross Validation)。CV 的主要思想是将原始数据分为两组, 分别代表训练集 T 和验证集 V。这里先通过 T 训练分类器, 然后用 V 去测试由 T 所训练出的模型, 从而验证实验的分类性能。本文中的实验结果是对测试集上数据进行验证的, 且其分类结果取决于 n 次实验结果准确率的均值, 即平均准确率。

3.5.3 实验设计及结果分析

实验中的训练集个数为 2300, 测试集的个数是 100。对输入的数据进行归一化预处理(如大小、灰度化等), 实验中输入数据的大小为 32×32 , CRBM 的层数为 2 层, 每层的学习率分别为 1, 0.05, 稀疏度都为 0.03, 卷积核的大小分别为 9×9 , 7×7 , 卷积层的特征映射个数分别为 18、32, 池化大小都为 2×2 。

实验 1: 不同网络参数对分类结果的影响

1) 不同 Sparsity(稀疏度)值对实验结果的影响

在本次实验中, 采用了两层 CRBM, 所以将会在每层都有 Sparsity 值的设置。经过多次实验可得当两层 Sparsity 的值都取 0.03 时, 分类效果达到最好。所以这里将设其中一层的 Sparsity 值为固定值 0.03, 另一个作为变量, 查看当其取不同值时, 所对分类结果造成的影响。其中 Sparsity1 代表第一层 CRBM 的 Sparsity 值, 第二层用 Sparsity2 表示。其实验效果如图 3-5 所示。

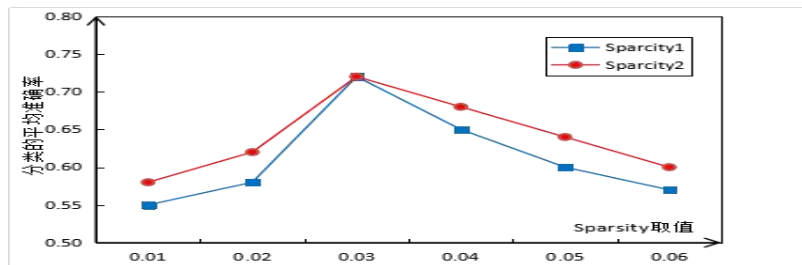


图 3-5 不同 Sparsity 值对实验结果的影响

由上图可知, Sparsity 的取值会影响到实验的分类结果, 这说明采用稀疏正则化在对于网络数据因超完备而导致的退化问题起到了一定的作用; 另一方面第

一层 Sparsity 的取值会影响到第二层的特征提取,这说明第一层的特征提取很重要,它训练的好坏会影响到后面网络层的训练;还可以看出当两层 CRBM 的 Sparsity 值当都取 0.03 时,分类结果达到最好,说明了此处是本实验可取最佳结果的位置。

2)不同卷积核的大小对实验结果的影响

图 3-6 为当两层 CRBM 的其中一层的卷积核大小不变,另一层取值不同时,采用本文方法对实验分类结果的影响。其中设 Kernel1、Kernel2 分别代表第一层和第二层卷积核的大小,当它们作为不变量时,其取值分别为 9×9 , 7×7 。其实验结果如图 3-6 所示。

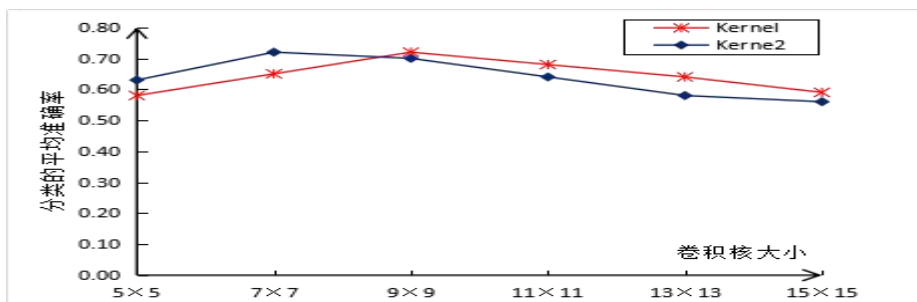


图 3-6 不同卷积核大小对实验结果的影响

由上图可以看出,两层的卷积核都符合随着值的增大而使实验结果分类的准确率先提高后降低的情形,这有可能是因为当卷积核较大时会引起网络出现过拟合现象。另一方面可以看出,Kernel1 的取值会影响到 Kernel2 的分类效果,这是因为 Kernel2 是在 Kernel1 的基础上再次进行特征提取,在 Kernel1 取值恰当时,会使实验的分类准确度有所提升。

3)不同迭代次数对实验结果的影响

CDBN 的最终网络运算权值是由网络训练的多次迭代而获取的,所以其迭代次数的取值会对实验中对国画三类的分类结果造成的一定影响,图 3-7 中 epoch1、epoch2 分别代表位于两层 CRBM 的迭代次数,其中 epoch1、epoch2 作为不变量时其取值都为 40。线性 epoch1 和 epoch2 表示随相对应迭代次数的变动而绘制出的结果变化趋势图。图 3-7 展示了当其中一层的迭代次数不变,另一层迭代次数变动时对实验结果准确率的影响。

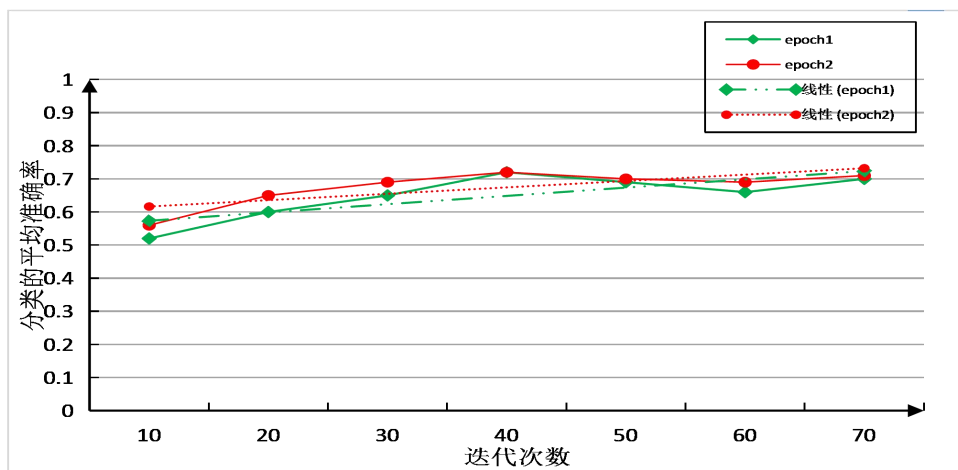


图 3-7 不同迭代次数的实验结果

由上图可知随着迭代次数的增加，实验的分类准确率会随之提升，当迭代次数达到一定数目时，其准确率的上升幅度变小。这是因为网络参数会随着迭代次数的增加而不断优化，致使分类准确率得以提升，当达到一定的次数时，网络参数呈现收敛的状态，使得参数的变动不会太大，分类效果也达到最佳。同两层卷积核大小的相互影响一样，epoch2 在 epoch1 取得适当值的基础上，其实验效果要好于 epoch1。

实验 2：不同算法在本数据上的实验结果

表 3-1 显示的是当分别采用传统 CNN、DBN、单层 CRBM 与双层 CRBM 方法时，对数据进行分类所取得的实验结果。这四个实验所用数据都为本文中的数据库，其中训练集和测试集分别为 700，100。

表 3-1 不同算法分类结果

采用算法	CNN	DBN	单层 CRBM	两层 CRBM
测试准确率/%	64.1	60.5	64.2	72.3
收敛步数	190	200	90	60
每次迭代时间/s	895	30	25	30

由上表可知，本文中所采用的两层 CRBM 性能要优于其他三种方法，具体原因如下：DBN 的实验效果低于 CNN 是因为两者的学习方式不同，DBN 采用的是无监督学习方式，在训练网络时不会给其类别标签进行学习，而 CNN 是有监督的学习方式，会将其标签进行参考；CDBN 的分类效果高于 DBN 的原因是 DBN 忽略了数据的结构化信息，而 CDBN 不仅将此信息考虑在内，还加入了卷积操作，很好地保留了图像的领域信息；两层 CRBM 的分类率高于单层的，是因为深度的增加可以更好地表征数据特征，提升识别率，但随之训练所用的时间也会增加。

3.6 本章小结

本章节针对传统 DBN 没有将图像的二维结构考虑在内的情况,提出了一种基于 CDBN 的中国绘画图像分类方法,同时在结构中采用了最大概率池模型来进行概率推理,并应用稀疏正则化来化解模型过完备而致使计算变复杂的问题。在实验中通过不同算法验证了本文所提算法具有一定的可行性,通过参数的不同设置,证明了不同参数的设置会影响实验结果的准确率。

第四章 基于 CNN 的中国绘画图像分类算法

本章节提出了一种基于 CNN 的中国绘画图像分类法。首先针对过拟合问题，应用了一种改进的算法 SMOTE 来扩增数据，然后将扩增后的数据直接输入到 CNN 中，经过隐藏层的卷积和亚采样，并采用校正线性单元 ReLu、S 形生长曲线 Sigmoid 替代传统的 Sigmoid 激活函数，提取的数据能更好地表示其特征。实验结果阐明：本文方法比传统分类法能够在中国绘画图像分类上取得更好的分类效果。

4.1 改进的 SMOTE 扩增方法

过拟合就是为了使实际输出与期望输出相一致，而致使实际输出过程中的计算变得过度复杂。产生过拟合的原因常常是由于在训练某个模型时，所训练的数据有可能数量不足^[51]，即无法通过训练数据估计整个数据的分布或者是过度地对模型进行训练。而 CNN 之所以具有提高分类或预测的准确率，是因为它可以通过训练海量的数据构建出一个含有多隐藏层的模型，从而学习到能够更好地表征数据的特征。由于绘画图像数量有一定的局限性，故在文中采纳一种改进 SMOTE 的方法来扩增数据集。

SMOTE 主要原理是将若干值插在少数距离较近的样本间，生成少量新的类样本，进而使少量样本的数量增多，提高少量类样本的分类精度^[52]。但该方法生成的样本会散落在稠密区域，不能很好地解决当前实验样本稀疏分布问题。本文提出的改进的 SMOTE 算法适当填充了类内间的稀疏区域，首先对每类样本分别进行聚类，将每类样本聚成 k 类，然后在到每个聚类的一个随机样本的路径上对每个已有的样本进行随机插值使得原始的每类样本扩展 k 倍，生成后的样本集是原始样本集的 $k+1$ 倍。根据需要样本集的大小来调整 k 值，如原始样本集大小为 n ，将原始样本集扩展一倍则将 k 设为 1 即可。该方法与添加噪声、随机复制等添加样本的方法相比，能更好地防止过拟合。因为添加噪声、随机复制等方法虽然在一定程度上增加了样本数量，但是该方法同时也增添了一些额外的训练数据，使得分类器的构建时间变长，另一方面所增加的类样本涵盖的有用信息不多，对于过拟合问题不能很好的解决。

4.2 数据的生成

4.2.1 样本的聚类

运用 K-means 聚类算法把每类样本归为多个子集，在收缩范围的同时可以快速选出重要的样本^[53]。但由于类质心初始化的情况会影响 K-means 聚类算法的结果，所以会使结果相对不稳定。经理论及实践验证，融合多个分类器的结果可以提高分类的正确率^[54]。因此在本文中为了使得到的聚类结果相对稳定，采用了一种融合多分类器的思想。所采取的方法是在每次初始化聚类时其质心不同，然后选择用户投票的方法融合 Q 次聚类结果，其中用矩阵 $A=(a_{ij})$ 代表聚类的结果 (n 为样本数， a_{ij} 为样本 i, j 在同一聚类中出现的聚类次数)，如果 $a_{ij}/Q > 0.5$ ，那么将这 2 个样本归到同一聚类中^[55]。

4.2.2 样本的生成

每个样本均有一个自身所属的聚类，在自身聚类内和其他随机聚类样本之间进行随机插值。如图 4-1 所示， X_i 为一个已有样本属于 C 聚类， P_1, P_2, P_3 分别为样本到 A, B, C 聚类路径上随机生成的扩增样本 ($P_3 \in C$)，提升了其当前类的分布密度， P_1, P_2 则随机填充到相对稀疏区域。适当设置扩增数据的比例，得到的样本集能够更好地模拟数据的现实分布。这样既兼顾了 SMOTE 原生算法的功能，同时也扩展了对每类样本相对稀疏区域的填充。生成过程如下：

- 1) 计算每个样本 X_i 到每个聚类一个随机样本 R_i 的距离 d ;
- 2) 根据 $P=X_i+rd$ 进行线性插值，生成样本 P 。其中， r 为一个与 X_i 维度相同的随机矩阵。

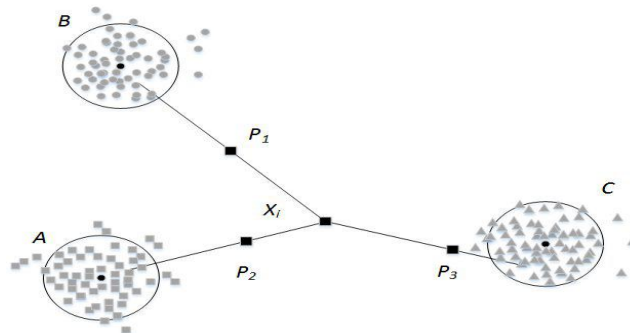


图 4-1 数据扩增示意图

4.2.3 聚类个数的设置

由于每个样本在每个聚类的方向上都会随机生成一个样本, k 个聚类将原有样本增加 k 倍, 而过多的扩增数据会破坏数据的分布特性, 减弱模型的泛化能力。 k 值较小时, 聚类的结果无法很好地确定类内稀疏区域, 所以本文选择从生成的样本中随机挑选一定数量的数据加入训练。

图 4-2 展示了本文根据绘画题材将绘画分为花鸟、山水、人物 3 大类的扩增数据。由此图可以看出, 生成的图像仍保有原图像的主要特征, 使训练数据变得更丰富。

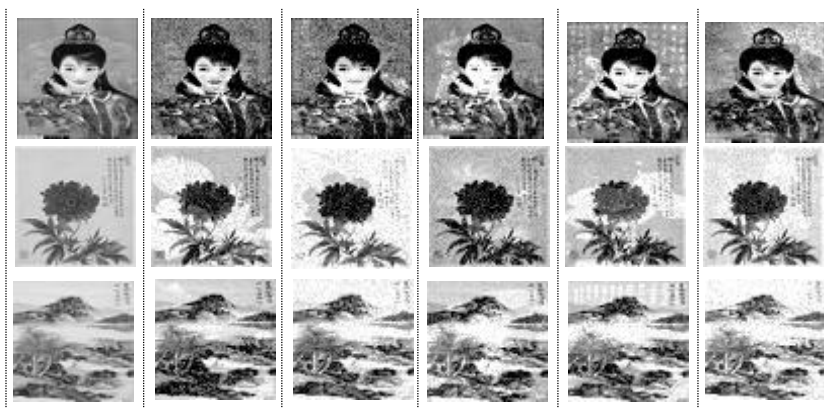


图 4-2 数据扩增样本示例

4.3 基于 Sigmoid+ReLU 的 CNN

4.3.1 激活函数 Sigmoid+ReLU

传统神经网络的激活函数通常为 Sigmoid 或 tanh 函数。基于数学的角度, 非线性的 Sigmoid 函数能够在信号的特征映射空间取得很好的效果, 原因是这样的方式可以使得信号增益在中央区域较大, 两翼区域较小。而从神经学的角度出发, 这样的中央区域和两翼区域可以看做是神经元的兴奋态和抑制态。因此在对数据进行学习时, 最好能够将一些重要的特征汇聚在中央区, 而非重要的汇聚在两翼区。以校正函数 $\max(0, x)$ 为例的线性函数 ReLu 正是属于该模型, 同时使用该函数可以降低梯度训练深度网络时的残差衰减问题。从输出层反向传播梯度 Grad 时, 误差 Error 在各层都要乘以该层的输入神经元值 x 、激活函数 $\text{Sigmoid}(x)$ 的一阶导数。即 $\text{Grad} = \text{Error} \cdot \text{Sigmoid}'(x) \cdot x$ 这是一个双端饱和的 Sigmoid 系函数, 该函数会出现 2 个问题: ① $\text{Sigmoid}'(x) \in (0, 1)$ 导数缩放; ② $x \in (0, 1)$ 或 $x \in (-1, 1)$ 饱和值缩放。

在这种情况下, 途径神经网络的每一层时, 误差都会成倍衰减, 造成网络学

习的速度降低，这是因为反向传播在进行多层递推时，所得到的梯度会随之不停地衰减，甚至消失。但 ReLu 却可以很大程度地提高训练速率，这是因为它的饱和度和只在一端，而且其梯度值为 1，使得梯度可以在反向传播时速度加快。所以文中卷积层应用的激活函数是 ReLu，输出层为 Sigmoid。同时在我们的亚采样层没有用到激活函数，通过采用平均值池化操作来降低网络模型计算的复杂度。

4.3.2 CNN 网络结构

研究表明，CNN 的网络结构的选取会影响分类结果的性能，影响其分类的优劣因素有很多，如输入数据的大小、网络的深度选取以及训练参数的设置等问题。本文主要是通过调整卷积神经网络的网络结构来优化网络，即对网络纵向深度和横向宽度进行调整。深度上，调整卷积层与亚采样层的堆叠层数；宽度上，调整网络各层的特征图数目以及每层平面的大小。经过多次调整，最终筛选出 CNN 的输入图像大小为 32×32 ，最佳结构为 6 层：1 层输入，2 层卷积，2 层亚采样层和 1 层全连接，如图 4-3 所示。

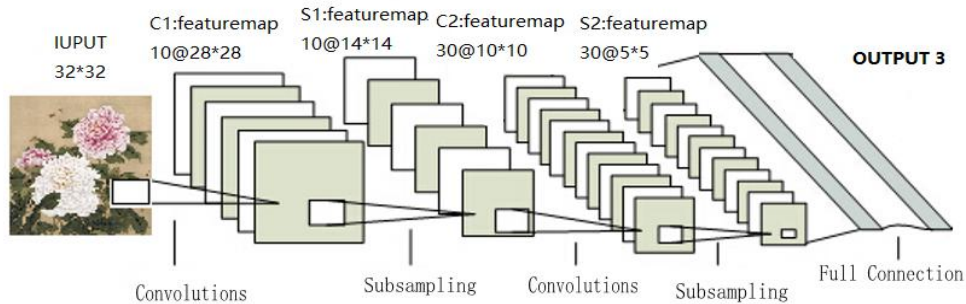


图 4-3 CNN 网络结构设定

4.4 实验结果与分析

为验证本文所提方法的有效性，首先采用本文方法对不同扩增数据进行测试，以此来验证采用改进的 SMOTE 扩增数据可以提高分类的准确率；然后采用未改进的 CNN、神经网络 (Neural Networks, NN)、DBN 对本实验的数据进行测试，并将结果与本文的方法进行对比；最后将本文方法与浅层学习方法的分类结果进行对比，验证本文方法的可行性。

实验 1：不同扩增数据集对分类结果的影响

为验证本文数据扩增的有效性，应用本文方法对 4 组不同程度的扩增数据进行实验。将原始样本分为训练集和测试集，每类的训练集是从 800 幅样本中随机抽取 780 幅，余下的 20 幅作为测试集。本实验中原始集用 A(780)来表示，训练

集与测试集同幅度扩增为原始集的 i 倍($i=0,2,3,4,5$), 用 $B(780 \times i)$ 表示, 总的训练集用 $A+B$ 表示, 实验结果如表 4-1 所示。拟合度即训练集预测准确率是从训练集 ($A+B$) 幅绘画数据中随机抽取 $20i$ 幅作为训练集的预测样本测试得到的准确率, 平均准确率为在测试集中测试得到的准确率。

表 4-1 不同扩增数据集对分类的影响

扩增倍数	原数据集	扩增数据集	训练集	拟合度/%	平均准确率/%
0	A(780)	B(780×0)	A+B(780)	93.48	66.33
2	A(780)	B(780×2)	A+B(2340)	94.33	82.22
3	A(780)	B(780×3)	A+B(3120)	94.50	87.66
4	A(780)	B(780×4)	A+B(3900)	94.97	90.33
5	A(780)	B(780×5)	A+B(4680)	94.51	91.55

如果训练集上预测得出的结果好, 而在测试集上预测得出的结果较差, 这种现象就体现为过拟合。由表 4-1 可以看出, 使用原生样本集训练, 拟合度为 93.48%, 分类平均准确率为 66.33%, 有明显的过拟合现象。通过改进的 SMOTE 方法对原数据集进行扩增, 扩增了原始数据的 2 倍时, 拟合度为 94.33%, 平均准确率达到 82.22%, 当扩增了原始数据的 5 倍时, 拟合度为 94.51%, 平均准确率达到 91.55%, 与此方法应用在原始数据集中的结果相比, 分类准确率得到显著提高, 且显著得解决了过拟合问题。同时可以看出不同扩增比对分类效果影响不一, 随着扩增数量的增大, 对实验结果的影响有减弱趋势, 这是由于该数据生成算法只关注于对类内稀疏区域的填充, 并未实现对类间稀疏区域的处理, 不能扩展类别的边界。

实验 2: 本文方法与 CNN、NN、DBN 的实验对比

分别运用本文方法、NN、DBN 及传统 CNN 算法在相同的数据集上对比测试, 训练集从花鸟、人物、山水样本集中各随机抽取 780 幅, 余下各 20 幅作为测试集。实验测试的数据集分别为扩增前与扩增后的数据, 扩增前的训练集为 2340 (780×3), 测试集共为 60 (20×3), 扩增后的数据集为扩增前的 6 倍, 训练集和测试集分别为 14040 (2340×6) 和 360 (60×6), 实验结果如表 4-2 所示。

表 4-2 不同深度结构学习方法的分类结果比较

算法类别	训练数据集	平均正确率/%	收敛步数	训练时间/s
本文算法	2 340	66.33	170	943.5
	14 040	91.55	60	3 778.2
CNN 算法	2 340	65.00	200	896.0
	14 040	84.20	150	3 804.0
NN 算法	2 340	78.00	220	30.8
	14 040	85.76	180	201.6
DBN 算法	2 340	61.60	200	30.0
	14 040	73.94	180	201.6

从表 4-2 可以看出, 利用改进的 SMOTE 扩增数据, 在本文算法以及传统 CNN, NN 与 DBN 中, 都可以显著地提高学习方法的识别率, 且本篇所提方案的分类性能显著强于其他方案, 体现了本文方案的可行性。但由于数据扩增后, 要训练的数据大大增加, 导致了训练的时间变长。由于不一样的学习方案训练数据的方式各异, 使得各个训练方式消耗的时间也就各不相同; 另一方面由于数据扩增可以更好地提取出比较全面的特征, 使其收敛步数有所减少, 从而提高分类性能。

实验 2: 本文方法与浅层学习方法的对比

鲍鸿^[1] 在他的论文中提出图像分类性能对分类器的选取影响不大, 因此在与其它浅层学习方法进行对比时, 所采用的分类器统一为 SVM, 其中提取的人工特征分别为尺度不变特征转换 SIFT、梯度直方图 HOG、基于 HSV 的全局颜色直方图、Color-SIFT。浅层实验中的训练集是随机从 800 幅花鸟、山水、人物样本集中抽出 700 幅, 余下作为测试集。本文方案的测试数据来自于将原始测试数据扩增 5 倍后的数据, 实验准确率如表 4-3 所示。

表 4-3 基于不同特征描述子与本文方法分类准确率的比较

采用算法	SIFT	HOG	HSV	Color+SIFT	本文算法
平均准确率/%	68.30	67.30	68.70	73.00	91.55

从表 4-3 可以知道, 本篇所述方案的分类准确率显然优于基于浅层学习方案的准确率。同采用人工定义规则来取得特征的方案作比较, 本文的优势在于采用大数据自动学习事物特征的方式更好地描述出了数据丰富的内在特性, 从而提高了分类的准确性。

4.5 本章小结

本章节主要介绍为解决传统的绘画图像分类方法的局限性问题, 在 CNN 的基础上提出了一种新的分类方法。通过对比不同扩增集的实验结果和不同深度的学习方法的实验结果, 可以分析出本文方法在对中国绘画图像分类上具有较好的效果。

第五章 总结与展望

5.1 工作总结

数字化图像的建立,不仅是当代产品电子信息化发展的需求,更是便于人们检索、查阅与管理图像的基本要求。而中国绘画图像作为图像的一种,因其自身具有很高的艺术价值、收藏价值以及欣赏价值,所以该领域数字化的建立,将会对图像数字化的成立及我国艺术界领域的相关研究产生重大的影响。其中实现中国绘画图像的分类,是中国绘画数字化建立的重要课题之一。本文通过总结中国绘画图像分类常用方法,得出当前一般应用浅层学习的方式来对其进行分类。虽然这种方法取得了一定的成绩,但是由于这种方法存在如下缺陷:在有限样本和计算单元的情况下,其表征复杂函数的能力有限,针对复杂分类的问题其泛化能力受限;由工设定的特征,不仅繁琐和设计时比较麻烦,而且对于中层特征表示十分困难。而深度学习的多层非线性网络结构,不仅能够很好地实现对复杂函数的逼近,还可以自动地学习特征,并通过逐层训练的方法实现对中层特征的良好表达。由此,本文提出了一种基于深度学习的中国绘画图像分类。

首先本文针对 DBN 在对网络进行训练时,尚未将数据空间领域的信息考虑在内,提出了一种将卷积与 DBN 相结合的一种算法 CDBN。该方法一方面利用 CRBM 达到使生成模型的变换具有不变性,且较易地将数据转换到高维空间的目的。另一方面利用卷积操作实现计算复杂度降低和高度的畸变容忍性等目的。同时在本方法中还引入了最大池化模型和稀疏正则化,实现了对概率的很好推理和减弱过完备现象的目的。由此对中国绘画图像进行特征提取和分类,通过在实验中将本方法与传统 CNN、DBN 作对比,可得出本文算法的分类准确率要高于他们,说明了本文方法具有一定的研究价值。通过分析不同参数的取值对实验结果的影响,可知实现网络的全局收敛度与这些参数的设置有关。

然后针对传统 CNN 算法在激活函数选取时,所存在的不足之处,提出了应用 Relu+Sigmoid 替代原先的 Sigmoid+Sigmoid 激活函数,从而可以实现对局部有效区很好的响应;另一方面,在 CNN 对输入数据进行训练时,由于数据不够充足而出现的过拟合现象,提出了一种改进的 SMOTE 算法,来达到能够很好对数据进行扩增的目的,从而改善数据特征表现的力度。最后通过在实验中与不同浅层学习方法对分类结果的对比,可以得出本方法对中国绘画图像的分类准确度要比浅层学习方法的高些,验证了本文方法的可行性。同时从不同参数取值对实

验结果所造成的影响，可知参数的选取对于网络训练能够取得较好结果的重要性。

最后对本文所应用的方法及所起的作用做了相关的总结，指出了研究该课题的意义所在，并对未来的发展方向及本文中所需要改进的地方，进行了简要的阐述。

5.2 未来展望

随着大数据的到来，在语音、图像、文本等各个领域都掀起了一场应用深度学习的浪潮，这是因为该算法的设计思路逐渐成熟，使原始数据从训练出发，经过一个端到端的模型，然后即可直接输出最终结果。这种模型不仅可以使自动学习数据特征变得更加简单，而且由于深度学习中的每一层都可以为了最终的任务来调整自己，最终实现各层之间的通力合作，可以大大提高任务的准确度。另一方面从中国绘画当前所处现状来说，将深度学习引入到其分类中，将会对于其数字化的实现提供很大的帮助。

尽管本文应用深度学习在中国绘画分类上取得了一定的成果，但是由于深度学习本身具有高复杂度、高抽象性的特点，且本人研究的时间、涉及面等有限，对于应用在中国绘画图像分类上的深度学习，还有很多地方值得探究。就本文而言，还可以从以下几方面加以改进：

1)本文第三章是基于 DBN 的中国绘画图像分类法，虽然应用卷积和稀疏正则化实现了对于中国绘画高维特征的提取和缓解训练过度完备化，但是由于训练的数据较大，使得训练的时间较长，这点我们可以尝试使用并行处理、GPU 等技术进行加速。同时还可以引入权值调整时，应用优化的方法提高训练速度，如退化法、粒子群法等。

2)在本文第四章中提出了基于 CNN 的中国绘画图像分类法，利用改进的 SMOTE 减弱过拟合现象，在卷积层、输出层分别采用 ReLu、Sigmoid 作为激活函数，然后分类。虽然与其他方法相比，本方法的效果显著，但由于绘画背景及与主题无关的信息如印章、不同类别的融合等，对分类结果具有一定干扰性，而本算法并未将其考虑在内，后期可在本文基础上，通过显著性等方法提出主题，进而提高分类性能。

3)本文仅对 DBN 及 CNN 在中国绘画分类上进行了相关研究，并未将深度学习的其他模型考虑在内，如稀疏自动编码 SAE、堆叠自动编码等。后期可以将深度学习的其他方法以及这些方法与其他方法的结合进行研究，如 CNN 与 SVM 的结合、DBN 在特征提取时进行特征融合等。

4)本文实验所用的数据库样本数量有限,对于深度学习在大数据上的优势并未完全展示出来,后期可以扩大数据或者分类的种类进行进一步的探索和研究。

致谢

两年多的研究生生活就快要结束了，在这期间我过的非常充实并学习到了很多知识。不仅在导师的带领下，提升了自身的科研能力，还在研二阶段担任了学院研会主席，从不同方位锻炼了自己。历史虽已成为过往，但不会在我脑海里淡忘，在此我要向那些给予我帮助和关心的所有人致以深深地谢意。

首先，我要衷心的感谢我的导师王小华教授！从有幸成为王老师的学生开始，不管是在学业还是生活，都受到了王老师的悉心教导和热心的帮助，使我在研究生阶段取得了较大的进步。同时我还要感谢我的指导老师杨冰老师！从踏入“人工智能与信息处理”实验室开始，我便在她的带领下，参与了她所负责的研究课题《基于语义认知和视觉显著性模型的绘画艺术图像分类方法研究》和《基于艺术风格的绘画图像分类研究》，它们分别来自于国家自然科学基金青年基金和浙江省自然科学基金青年基金项目，本论文就是在这些项目的基础上进行开展的。在对论文的研究过程中，杨老师提供了许多启发性和建设性的意见和指导，使我能够较为顺利地完成。同时，杨老师还经常在百忙之中给予我耐心的指导，并与我探讨解决方案和修改意见，这对于我论文的进展有很大的促进作用。导师们严谨的教学态度以及科研作风，都让我受益颇丰。

其次我要感谢我们实验室的另外几位老师，他们分别是姚金良老师、黄孝喜老师、王荣波老师和谌志群老师。他们不仅向我们树立了良好的科研作风，同时还能够每次在实验室学术讨论会上，根据我们的汇报情况，指出我们的不足之处并给予合理的意见，以助我们在科研的道路上得以更好的发展。

再次我要感谢我们实验室已经毕业的师兄师姐们，感谢他们在学习和生活中给予的帮助与关心，你们的热心和帮助让我倍感温暖，让我在研究生阶段寻找出了一条方向更加明确的道路，能够与你们相识，深感荣幸。我还要感谢跟我一起经历了这两年多读研阶段的同窗孙小雪、陈浩、张天宇，与他们相识，不仅能够互相学习，还可以在生活上相互帮助和支持。除此之外，我还要感谢我可爱的室友李瑞敏、蔡扬、项璐露，他们的陪伴与默默地支持，使得我的生活和学业更加精彩。最后，我还要感谢韩鹏程同学，他在我探索深度学习时，给了我很多好的建议和指导，加快了我论文的完成进度。在此，祝福你们鹏程万里，事业有成。

我还要感谢我的家人，因为有了你们才有了我现在的成绩，你们对我的殷切希望是我不断前进的动力，我要祝你们身体健康，幸福快乐。

最后再次衷心感谢所有帮助、鼓励、支持我的老师、同学、家人和朋友。

参考文献

- [1] 鲍泓. 基于视觉感知的中国画图像语义自动分类研究[D]. 北京交通大学, 2012.
- [2] Bar Y, Levy N, Wolf L. Classification of Artistic Styles Using Binarized Features Derived from a Deep Neural Network[M]// Computer Vision - ECCV 2014 Workshops. 2015:71-84.
- [3] Liu P H, Su S F, Chen M C, et al. Deep learning and its application to general image classification[C]// International Conference on Informative and Cybernetics for Computational Social Systems. IEEE, 2015.
- [4] Liu G J, Liu Y, Guo M Z, et al. Non-negative Locality-Constrained Linear Coding for Image Classification[M]// Intelligence Science and Big Data Engineering. Image and Video Data Engineering. Springer International Publishing, 2015.
- [5] Zhang D, Pham B, Li Y. Modelling Traditional Chinese Paintings for Content-Based Image Classification and Retrieval[C]// International Multimedia Modelling Conference. IEEE Computer Society, 2004:258.
- [6] 杨冰, 许端清, 杨鑫等. 基于艺术风格相似性规则的绘画图像分类[J]. 浙江大学学报工学版, 2013, 47(8):1486-1492.
- [7] Li J, Wang J Z. Studying Digital Imagery of Ancient Paintings by Mixtures of Stochastic Models[J]. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2004, 13(3):340-53.
- [8] Lu G, Gao Z, Qin D, et al. Content-Based Identifying and Classifying Traditional Chinese Painting Images[C]// Image and Signal Processing, 2008. CISP '08. Congress on. 2008:570-574.
- [9] 蒋树强. 视觉媒体语义自动提取关键技术研究[D]. 中国科学院计算技术研究所, 2005.
- [10] Jiang S, Huang Q, Ye Q, et al. An effective method to detect and categorize digitized traditional Chinese paintings[J]. Pattern Recognition Letters, 2006, 27(7):734-746.
- [11] Crowley E J, Zisserman A. In Search of Art[M]// Computer Vision - ECCV 2014 Workshops. 2015:54-70.

- [12]Sun M, Zhang D, Ren J, et al. Brushstroke based sparse hybrid convolutional neural networks for author classification of Chinese ink-wash paintings[C]// IEEE International Conference on Image Processing. 2015.
- [13]Bengio Y. Learning Deep Architectures for AI[J]. Foundations & Trends® in Machine Learning, 2009, 2(1):1-127.
- [14]Hinton G. Deep Belief Nets[M]. Springer US, 2011.
- [15]Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets.[J]. Neural Computation, 2006, 18(7):1527-1554.
- [16]Krizhevsky A, Sutskever I, Hinton G E. ImageNet Classification with Deep Convolutional Neural Networks[J]. Advances in Neural Information Processing Systems, 2012, 25(2):2012.
- [17]Sermanet P, Eigen D, Zhang X, et al. OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks[J]. Eprint Arxiv, 2014.
- [18]He K, Zhang X, Ren S, et al. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification[J]. 2015:1026-1034.
- [19]Schmidhuber J. Deep Learning in neural networks: An overview.[J]. Neural Networks, 2014, 61:85-117.
- [20]Nath S S, Kar J, Mishra G, et al. A survey of image classification methods and techniques[C]// International Conference on Control, Instrumentation, Communication and Computational Technologies. IEEE, 2014:554-557.
- [21]Roomi M, Vasuki P. A survey of SAR image Classification Methods and Techniques for improving classification Performances[C]// Icicic. 2012:592-597.
- [22]Shao L. A survey on feature based image retrieval techniques[J]. 2011.
- [23]Sonka M, Hlavac V, Ceng R B D M. Image Processing, Analysis, and Machine Vision[M]// Image processing, analysis, and machine vision /. Thomson Learning, 2008:685–686.
- [24]Yang H P. Computer Digital Image Processing Application and Research[J]. Advanced Materials Research, 2014, 926-930:3389-3392.
- [25]Gonzalez R C, Woods R E. Digital image processing[M]. Pearson/Prentice Hall, 2007.
- [26]Yao X. Evolutionary Artificial Neural Networks[J]. International Journal of Neural Systems, 2014, 4(03):539--567.
- [27]Koprinkova-Hristova P, Mladenov V, Kasabov N K. Artificial Neural Networks[J]. European Urology, 2012, volume 40(1):245.

- [28]Eladel A, Ejbali R, Zaied M, et al. Deep learning with shallow architecture for image classification[C]// International Conference on High PERFORMANCE Computing & Simulation. 2015.
- [29]Liu C, Jiang H. Classification of traditional Chinese paintings based on supervised learning methods[C]// IEEE International Conference on Signal Processing, Communications and Computing. IEEE, 2014:641-644.
- [30]Gatys L A, Ecker A S, Bethge M. A Neural Algorithm of Artistic Style[J]. Computer Science, 2015.
- [31]Apolloni B, Ghosh A, Alpaslan F N, et al. Machine Learning and Robot Perception[J]. Springer Berlin, 2013, 7.
- [32]Braverman M. Poly-logarithmic Independence Fools AC^0 Circuits[J]. 2009, 57(5):3-8.
- [33]Najafabadi M M, Villanustre F, Khoshgoftaar T M, et al. Deep learning applications and challenges in big data analytics[J]. Journal of Big Data, 2015, 2(1):1-21.
- [34]Gehler P, Nowozin S. On feature combination for multiclass object classification[J]. 2009, 30(2):221-228.
- [35]Sun B Y, Zhang X M, Li J, et al. Feature fusion using locally linear embedding for classification.[J]. IEEE Transactions on Neural Networks, 2010, 21(1):163-168.
- [36]Schmidhuber J, Meier U, Ciresan D. Multi-column deep neural networks for image classification[C]// IEEE Conference on Computer Vision & Pattern Recognition. 2012:3642-3649.
- [37]Zhong S H, Liu Y, Liu Y. Bilinear deep learning for image classification[C]// International Conference on Multimedia 2011, Scottsdale, Az, Usa, November 28 - December. 2011:883-884.
- [38]Zhang Y, Shang C. Combining Newton interpolation and deep learning for image classification[J]. Electronics Letters, 2015, 51(1):40-42.
- [39]Sutskever I, Martens J, Dahl G, et al. On the importance of initialization and momentum in deep learning[C]// International Conference on Machine Learning. 2013.
- [40]Hall M L. Deep Learning[M]. VDM Verlag Dr. Müller, 2011.
- [41]Goodfellow I. Deep learning of representations and its application to computer vision[J]. 2015.

- [42]Swietojanski P, Ghoshal A, Renals S. Convolutional Neural Networks for Distant Speech Recognition[J]. IEEE Signal Processing Letters, 2014, 21(9):1120-1124.
- [43]Kang L, Ye P, Li Y, et al. Convolutional Neural Networks for No-Reference Image Quality Assessment[J]. 2014:1733-1740.
- [44]Yin B C, Wang W T, Wang L C. Review of deep learning[J]. Beijing Gongye Daxue Xuebao, 2015, 41(1):48-59.
- [45]Telgarsky M. Representation Benefits of Deep Feedforward Networks[J]. Computer Science, 2015, 15(8):1204-1211.
- [46]Merola G M. spca: Sparse Principal Component Analysis[J]. Pattern Recognition Letters, 2013, 34(9):1037–1045.
- [47]Ngoc Khuong Nguyen, Anh-Cuong Le, Hong Thai Pham. Deep Bi-directional Long Short-Term Memory Neural Networks for Sentiment Analysis of Social Data[J]. 2016.
- [48]Hinton G E. A Practical Guide to Training Restricted Boltzmann Machines[M]// Neural Networks: Tricks of the Trade. Springer Berlin Heidelberg, 2012:599-619.
- [49]Lopes N, Ribeiro B. Deep Belief Networks (DBNs)[M]// Machine Learning for Adaptive Many-Core Machines - A Practical Approach. Springer International Publishing, 2015:155-186.
- [50]Huynh T, He Y, Rüger S. Learning Higher-Level Features with Convolutional Restricted Boltzmann Machines for Sentiment Analysis[M]// Advances in Information Retrieval. Springer International Publishing, 2015:447-452.
- [51]Wyk V, Benjamin A. An Analysis of Overfitting in Particle Swarm Optimised Neural Networks[J]. 2015.
- [52]Xu J. IDS Method Based on Improved SVM Algorithm Under Unbalanced Data Sets[M]// Proceedings of the 2012 International Conference on Cybernetics and Informatics. Springer New York, 2013:413-420.
- [53]Bandyopadhyay S, Maulik U. An evolutionary technique based on K-Means algorithm for optimal clustering in, ja:math[J]. Information Sciences, 2002, 146(1–4):221-237.
- [54]Shi X, Manduchi R. A Study on Bayes Feature Fusion for Image Classification[M]. 2003.
- [55]Xie Y, Liu Y, Fu Q. Imbalanced Data Sets Classification Based on SVM for Sand-Dust Storm Warning[J]. Discrete Dynamics in Nature & Society, 2015, 2015(2):1-8.

附 录

一、硕士学习期间发表的论文：

[1] 黄雪丽，王小华，杨冰，姚金良.基于 CNN 的中国绘画图像分类[N].杭州电子科技大学学报.2017

二、硕士学习期间参与的项目

[1] 基于语义认知和视觉显著性模型的绘画艺术图像分类方法研究.国家自然科学基金青年基金(61402143)

[2] 基于艺术风格的绘画图像分类研究.浙江省自然科学基金青年基金(LQ14F020012)

三、主要术语对照表

DBN	深度信念网络
CDBN	卷积深度信念网络
CNN	卷积神经网络
SMOTE	合成少数类过采样技术
ReLU	校正线性单元
PCA	主成分分析
FLDA	Fisher 线性鉴别分析
ANN	人工神经网络
SVM	支持向量机
BP	反向传播法
K-means	K 均值
PHOG	分层梯度直方图
SIFT	尺度不变特征转换
DN	反卷积网络
CN	卷积网络
AE	自动编码器
SAE	稀疏自动编码器
RBM	限制波尔兹曼机
CRBM	卷积限制波尔兹曼机
FFDN	前馈深度网络
FBDN	反馈深度网络