

硕士学位论文

基于深度学习的视频人脸识别方法

**THE VIDEO FACE RECOGNITION  
METHOD BASED ON THE DEEP  
LEARNING**

由清圳

哈尔滨工业大学

2012 年 12 月

国内图书分类号：TP391.9

国际图书分类号：621.3

学校代码：10213

密级：公开

## 硕士学位论文

# 基于深度学习的视频人脸识别方法

硕士研究生： 由清圳

导 师： 丁宇新副教授

申 请 学 位： 工程硕士

学 科、专 业： 计算机技术

所 在 单 位： 深圳研究生院

答 辩 日 期： 2012 年 12 月

授予学位单位： 哈尔滨工业大学

Classified Index: TP391.9  
U.D.C: 621.3

Thesis for the Master Degree of Engineering

# **THE VIDEO FACE RECOGNITION METHOD BASED ON THE DEEP LEARNING**

<b>Candidate:</b>	Qingzhen You
<b>Supervisor:</b>	Associate Prof. Yuxin Ding
<b>Academic Degree Applied for:</b>	Master of Engineering
<b>Speciality:</b>	Computer Technology
<b>Affiliation:</b>	Shenzhen Graduate School
<b>Date of Defence:</b>	December , 2012
<b>Degree-Conferring-Institution:</b>	Harbin Institute of Technology

## 摘 要

本文的视频人脸检测识别方法的基本设计思想是，在给出一段视频文件以及这个视频文件的字幕和剧本之后，可以自动的对视频中的人物进行检测和识别，不需要任何的训练样本。视频人脸检测识别方法主要由四个部分组成：字幕剧本融合部分，人脸检测部分，样本集自动生成部分和基于深度学习的人脸识别部分。本文将深度学习算法引入到了视频人脸识别中来，有两方面的重要意义，一方面，视频人脸的识别要求算法具备一定的抗干扰能力，并且能够保证一定的实时性，本文的实验与分析表明，深度学习算法具备这方面的要求；另一方面，从深度学习算法特性的角度来说，深度学习算法最大的缺点就是构造深度模型需要大量的样本，这很大程度上限制了深度学习算法的应用，然而本文所设计的基于视频的人脸检测模块可以轻松的产生数万、数十万的样本，从而满足了深度学习算法的大样本集要求。

基于深度学习模型的人脸识别部分是整个系统的重点，这一部分主要有两方面的意义：一，经历了视频人脸的检测部分之后，虽然视频人脸集合中人脸的纯度有了很大的提升，但是依然会存在一些杂质，因此必须通过识别模块来进一步的过滤掉人脸集合中的杂质；二，通过视频所得到的帧文件中，经常会出现多张人脸同时出现的情况，在这种情况下，视频人脸的检测部分是无法将说话者与人脸进行对应的，必须通过识别模块才能区分出一个帧中的多个人脸。

基于深度学习模型的人脸识别部分主要包含三个模块：数据预处理模块、深度学习模块和识别模块。数据预处理模块主要由数据整合和构造数据立方体两个部分组成。深度学习模块通过两个具体过程来实现：**RBM** 调节和深度模型的反馈微调。**RBM** 的调节过程是自下而上的各个层间的调节过程，以这种方式来初始化整个深度模型的系统权值，而深度模型的反馈微调，首先进行自下而上的识别模型转换，然后再进行自上而下的生成模型转换，最后通过不同层次之间的不断调节，使生成模型可以重构出具有较低误差的原样本，这样就得到了此样本的本质特征，即深度模型的最高抽象表示形式。经过深度学习模型的处理，可以得到降维之后的样本特征，在此基础上运用识别模块，本文中所采用的识别方法是人工神经网络的识别方法。

**关键词：**人脸检测；肤色模型；深度学习；识别模型；生成模型；人工神经网络

## Abstract

The basic design idea of the video face identification and detection methods is: after the video files and their subtitles and scripts are given, it can automatically detect and identify the characters in the video, does not require any training samples. Video face recognition and detection method mainly consists of four parts: subtitles and screenplay fusion part, face detection portion, the sample set automatically generated part and face recognition part based on deep learning. This paper introduces depth learning algorithm into the video face recognition, there are two aspects' important significance. The one hand, the video face recognition algorithms has certain anti-jamming capability, and can guarantee the real-time, the experiments and analysis show that the depth learning algorithm with these requirements; On the other hand, from the point of view of the characteristics of depth learning algorithm, the biggest drawback of depth learning algorithm is that depth model requires a large number of samples, which largely limits the application of the depth learning algorithm. However, this designed video-based face detection module in this paper can easily generate tens of thousands, hundreds of thousands of samples to meet the large sample set requirements of the depth learning algorithm.

The face recognition part based on the depth learning model is the core of the entire system. The significance of this part consist of two aspects: first, after the video face detection part, although the purity of the human face in the video face collection has been greatly improved, but still there are some impurities, therefore the recognition module must be used to further filter out the impurities in the collection of human face; second, through the frame files obtained from the video, at the same time more than one face occur is possible, and in this case, video face detection section cannot handle the speaker corresponding to the face, the identification module must be used to distinguish more than one face in one frame.

The face recognition part based on depth learning model mainly consists of three modules: data preprocessing module, depth learning modules, and recognition module. Data preprocessing module mainly consist of the data integration and structure data two parts. Depth learning module consists of two parts: RBM regulation and feedback fine-tuning of the depth model. The adjustment process of RBM is the adjustment process between the respective layers of the bottom-up, in this way to initialize the weights of the entire depth model system. The feedback fine tuning of the Depth model, firstly, the bottom-up recognition model conversion, then the top-down generation model conversion, and finally through the continuous adjustment between the different

levels, the generated model can reconstruct the original sample which has a lower error. This essential characteristics of this sample are gotten,  $sp$  is the maximum abstract representation layer of the depth model. After the treatment of deep learning model, the characteristics of the samples after dimensionality reduction can be gotten, and then the identification module is used. This paper uses the artificial neural network method to do the Identification.

**Keywords:** face detection, skin color model, deep learning, recognition model, generated model, artificial neural networks

# 目 录

摘 要 .....	I
Abstract .....	II
第 1 章 绪 论 .....	1
1.1 课题来源 .....	1
1.2 本课题研究的目的及意义 .....	1
1.3 国内外研究现状 .....	2
1.3.1 基于统计的方法 .....	3
1.3.2 基于几何特征的方法 .....	3
1.3.3 人工神经网络的方法 .....	4
1.4 本文主要研究内容 .....	5
第 2 章 视频人脸检测识别方法研究概述 .....	6
2.1 人脸检测Adaboost算法概述 .....	6
2.2 深度学习概述 .....	7
2.2.1 深度学习基础理论 .....	8
2.2.2 深度学习设计模型 .....	10
2.3 人脸识别算法概述 .....	11
2.3.1 BP神经网络 .....	11
2.3.2 支持向量机 .....	13
2.4 本章小结 .....	14
第 3 章 基于深度学习的人脸识别算法 .....	15
3.1 数据整合 .....	16
3.2 构造数据立方体 .....	16
3.3 调节RBM .....	17
3.4 深度模型的反馈微调 .....	19
3.5 本章小结 .....	20
第 4 章 深度学习实验与分析 .....	21
4.1 深度学习模型的训练 .....	21
4.1.1 RBM 训练的实验与分析 .....	21
4.1.2 深度学习反馈微调的实验和分析 .....	22
4.2 深度学习模型的构造和选取 .....	23

4.3 PCA算法和深度学习对比的实验与分析 .....	29
4.3.1 PCA算法基础理论 .....	29
4.3.2 PCA与深度学习的实验分析 .....	30
4.3.3 PCA 与深度学习的对比分析 .....	32
4.4 基于深度学习的BP识别算法的性能分析 .....	35
4.4.1 失衡训练集对BP识别效果影响的实验与分析 .....	35
4.4.2 BP识别算法过拟合现象的实验与分析 .....	39
4.5 本章小结 .....	41
第 5 章 视频人脸检测识别系统 .....	42
5.1 人脸检测模块 .....	43
5.1.1 肤色模型人脸过滤 .....	44
5.1.2 唇色模型人脸过滤 .....	44
5.2 样本集自动生成模块 .....	45
5.2.1 数据采集 .....	45
5.2.2 数据预处理 .....	46
5.3 说话者识别模块 .....	46
5.4 本章小结 .....	47
结论 .....	48
参考文献 .....	49
攻读硕士学位期间发表的论文及其它成果 .....	53
哈尔滨工业大学学位论文原创性声明及使用授权说明 .....	54
致 谢 .....	55



# 第1章 绪 论

## 1.1 课题来源

本课题来自于对深度学习的研究和实验室视频人物标注项目。

## 1.2 本课题研究的目的及意义

深度学习算法的一个重要的特性就是对训练样本集规模要求比较大，在很多利用深度学习处理图像的实验中，实验者为了得到一个较大规模的训练样本集，通常要将一个图片变换不同的姿势，从而衍生出多个样本。这种方式不但低效，而且需要花费实验人员大量的精力和时间。

视频人脸的识别过程，通过一个系统机构可以轻松的生成数万，数十万的样本数据，因此将视频人脸识别与深度学习算法相结合，可以有效的解决深度学习算法样本集规模不足的问题。

同时，视频人物的识别过程中一个非常重要的问题就是实时性问题<sup>[1]</sup>，视频是一个动态的过程，因此对实时性要求很高，这里引入深度学习算法，通过它构建图像的本质特征，可以大大的提高视频人脸识别机构的实时性。

作为一个传统的以视频为基础的人脸识别系统，它通常应该包含人脸检测与追踪模块、人脸特征提取模块<sup>[2]</sup>和人脸识别模块<sup>[3]</sup>三个模块。在这三个模块之中，属于重中之重的就是人脸的识别模块，它的好坏直接决定着整个系统性能的优劣。

人脸识别本质上是一种生物基本特征的识别方法。它在许多方面拥有重要的意义和价值。比如在个人信息管理系统、金融消费验证系统、公共场所监控系统、银行个人用户监测系统、人机对话交互系统等领域拥有着广泛的应用前景。与腕部静脉特征识别、掌纹识别、眼部识别等技术相比，人脸识别技术无论是在样本收集方面还是在识别精度方面都有着自己的优势。人脸作为人类的一个基本的生物特征，在复杂场景的人物识别方面有着不可或缺的重要性，因此人脸识别的深入研究有着重要的理论和实际意义，其中主要体现在三个方面：

(1) 人机交互，传统的人机交互以个人计算机为例，人们主要是通过键盘和鼠标来向计算机输入控制命令，而计算机则通过显示器对人们的命令进行响应。。然而人们希望能够和机器进行更加智能化的交互，以一种更加容易被人类所接受的方式和机器进行交流，人机交互研究的重要意义就在于此。这项研究的最终目的就是希望机器可以和人们进行更自然的沟通，并且帮助人们高效的完成各种工作。为了实现这一目的，机器必须能够理解人们的角色、动作甚至姿态。人脸识

别恰恰是解决这一问题的有效方法<sup>[4]</sup>。

(2) 安全, 目前公共安全问题是全世界各个国家所共同关注的一个重大问题, 美国经历了 911 之后, 各个国家都开始重点关注自己国家的公共安全问题。公共安全的一个重要领域, 就是公共场所的安全问题。人脸识别算法是解决这一问题的有效方法, 通过人脸识别方法, 各国的安全部门可以在各种公共场所, 如飞机场、火车站等地方对那里的流动人员进行监控, 检测和识别危险分子。

(3) 娱乐, 随着科技的发展, 人脸识别技术已经用在了电影制作、互动娱乐等领域中。如很多智能机器可以通过读取人脸的表情来做出不同的响应, 也可以通过人们不同的姿态和动作来进行互动等。

对于普通的人脸识别模块, 整体的识别过程就是输入特定场景中的人脸, 使用人脸数据库存储和识别对应场景下的一张或多张人脸, 其中作为输入的人脸通常就是静止的人脸图像。但对于视频的人脸识别, 输入的不再是静止的人脸图像, 而是一段视频文件, 通过 使用人脸数据库对视频文件中的人脸进行存储和识别, 因为所要处理的对象是视频文件, 因此系统一定要保证人脸检测的实时性<sup>[5]</sup>, 并能够不间断的对视频中的人脸进行连续追踪<sup>[6]</sup>。这里如果忽略视频的实时性要求, 对于视频文件的处理, 完全可以看成是对多幅人脸图像的处理, 此时以多张人脸作为输入来训练系统, 使系统可以检测和识别他们。虽然随着科技的进步, 人脸识别技术发展迅速, 但是在许多复杂场景中, 目前的人脸识别方法还存在着许多不足, 这直接限制了它的应用前景。

为了解决这个问题, 尝试使用深度学习算法, 通过样本训练等方式方法, 得到符合要求的非线性函数集合的构造方式, 以及高层人脸图像的恰当表示, 从一个崭新的角度对人脸进行识别, 使计算机可以深度理解图像人脸的表达意义, 进而完成识别, 这是本课题的研究目的。

### 1.3 国内外研究现状

2000 年以来, 鉴于人脸识别研究的重要意义, 许多机构都在对其进行深入的研究, 比较著名的有麻省理工大学、康奈尔大学以及斯坦福大学等; 国内如北京大学、中科院计算机研究所和哈尔滨工业大学等, 这些大学和研究所在人脸识别领域取得了许多突破, 使人脸识别在检测性能上和识别精度上有了很大程度的提高。

然而尽管在人脸识别的研究取得了很多成果, 形成了许多识别算法, 但是在复杂情境下、光线和姿态不断变化的环境下, 各种算法都存在着很大的缺陷和不足, 而这又大大的限制了这些算法的应用范围。在这些算法中, 绝大部分仅仅考虑到了对图片进行样本特征提取和基于样本特征来进行识别, 并没有通过使计算

机去深入理解图像的方式来最终达到识别的方法。为此基于深度学习的算法被引入到人脸识别中来,例如最近提出的 DEM<sup>[7]</sup>采用能量模型来度量神经网络的稳定性,与此同时采用深度学习算法对目标进行深度表示和学习。由于深度学习算法通过多层次的抽象学习,将图像转换成了不同的表示形式,最终通过高层的抽象表示,来让计算机做出复杂的响应行为,实现智能化人脸识别的目的。自从 2006 年以来,基于深度学习的深度网络已经应用在了很多领域并取得了成功,如模式识别(深度架构网络<sup>[8]</sup>, DBN<sup>[9]</sup>, 稀疏编码<sup>[10]</sup>),降低变量维度<sup>[11]</sup>,运动建模<sup>[12]</sup>,目标分割<sup>[13]</sup>,信息检索<sup>[14]</sup>以及自然语言处理<sup>[15]</sup>等方面。从目前人脸检测识别方法的发展趋势来看主要的方法有以下几类:

### 1.3.1 基于统计的方法

在基于统计的方法中,人脸图像被视为随机向量,因此分析人脸模式可以通过一些统计方法来完成,这类方法均得到了完备的统计学理论的支持,并且在得到较好地发展的基础上,逐渐的产生了很多有效的技术方法。

由 Turk 和 Pentland 提出了特征脸方法。对于每一幅人脸图像的分析,均是按照依次从上到下、从左到右的顺序,从而将人脸图像中所有像素的灰度值组成一个高维的向量,然后再通过主成分分析的方法将高维向量降低维数为低维向量。

随着人们学术研究的不断深入,继而出现了贝叶斯人脸识别方法<sup>[16]</sup>。此算法提出了一种度量方法,该度量方法基于概率的图像相似度,使得人脸图像之间的差异被分为两种:类间和类内差异,其中类间差异代表的是在不同对象之间存在的本质差异,而类内差异则为同一个对象的不同图像之间存在的差异,在此基础上,人们在分析具体的人脸图像差异问题时,将这个差异看成是类间差异和类内差异之和,这样如果类间差异的数值不大于类内差异时,则说明被检测的两张人脸应该是属于同一人物对象。

基于 AdaBoost 算法人脸检测器<sup>[17]</sup>在 2001 年被 Viola 发表出来,这一算法的产生对人脸识别领域产生了重要的影响。在保证达到较高检测率的基础上,使用统计方法领域的 Adaboost 方法,从而实质性的提高了人脸检测的速度。该算法已经成为了目前一种很重要的人脸检测方法,并且使得人脸检测研究有了更加深入的进展。近年来学者们又在此基础上,提出了用于快速进行人脸识别的级联探测器,再次使得人脸检测的实时性和有效性得到了提高。

### 1.3.2 基于几何特征的方法

Bledsoe 提出来的基于几何特征的识别方法,是文献中所记载的最初的人脸识别方法。这一方法最大的特点是将面部特征节点之间的关联比作为特征,通过 KNN 的方法实现人脸识别的目的。通过此识别方法建立起来的人脸识别系统作为一个

半自动的系统，必须通过人手工来定位面部特征点。也正是因为存在人手工的干预，使得这个系统对出现的光照变化和人的姿态变化不足够敏感。

Kanade 首先先来计算面部特征（眼角、鼻孔、嘴巴、下巴等）之间的距离、这些特征之间的角度关系以及其它各种几何关系，继而利用所有的几何关系来进行人脸识别，与简单模板<sup>[18, 19]</sup>匹配的方法相比，此算法在提取人脸特征方面具有一定的优势。早期比较重要的基于几何特征的人脸识别的方法中包含侧影(Profile)识别方法。这一算法的主要思想是在人脸侧影边缘曲线上进行特征提取，以所提取出的特征代表人脸侧影曲线，再接着从中提取出基准点，最后根据这些基准点之间存在的几何特征进行人脸识别。因为侧影识别算法相对来说比较简单并且应用面积比较小，导致研究侧影识别的学者比较少。基于几何特征的方法的优势是非常直观、占用内存少、识别速度快，但是缺点是此类方法提取的特征在一定程度上不太敏感于光照的变化。

### 1.3.3 人工神经网络的方法

通过综合运用人工神经网络的学习与分类能力来对人脸进行识别是将神经网络应用于人脸识别的基本思想<sup>[20]</sup>。人工神经网络的学习能力非常强，许多人类面部的特征和规则很难通过一种有效的方式进行描述，然而通过人工神经网络不断的训练学习过程，这些特征和规则可以被有效的表达出来，因此人工神经网络在处理人脸识别问题上有着一一定的优势。与此同时，人工神经网络也存在着自己的缺陷，与普通方法相比通过直接采用神经网络的方法来对样本的特征进行提取，这一过程通常需要规模巨大的输入神经元集合，而这直接导致了样本在训练的时候要处理大量的参数数据，因而使得整个系统的实现变得困难起来<sup>[21]</sup>。

随着对人工神经网络研究的不断深入，研究人员开始不断的尝试构造含有多个隐藏层（大于等于 2 层）的神经网络，以此来使其可以对更加复杂的客观事物进行深度表示，然而这样的尝试在 2006 年以前都没有取得成功。然而 2006 年和 2007 年发表的 3 篇论文<sup>[22-24]</sup>改变了这种状况，这种状况的改变是通过 Hinton 在 DBNs(Deep Belief Networks)上的不断深入研究所实现的。这三篇文章的主要思想是：

（1）对于客观事物的表达方式可以通过无监督的方式进行学习，这样的学习过程可以被用来首先预训练每一层；

（2）系统的预训练过程本质上就是自下而上的逐层进行无监督训练，后一层紧接着前一层的训练，每一层中训练后所得到的表示将用来作为后一层的输入。

（3）对于系统整体进行无监督的训练过程，首先通过自下而上的识别模型进行逐层表示，然后通过自上而下的生成模型进行重构，在每一层的表示与重构之间不断的进行微调，最终完成对整个系统的训练过程。

人脸作为一种客观事物，对于它的表示十分复杂，而深度学习算法的最大优点之一就是拥有较强的对复杂客观事物的表示能力，因此将深度学习算法应用到人脸识别中来，拥有着重要的理论和实际意义。

## 1.4 本文主要研究内容

本文主要的研究对象就是深度学习模型，对深度学习模型的研究主要从四个方面展开：深度学习模型的训练、深度学习模型的构造和选取、PCA 算法与深度学习模型的对比，基于深度学习的 BP 识别算法的性能分析。

(1)深度学习模型的训练。这部分主要是分析深度学习模型训练的两个过程：RBM 调节过程和深度学习模型的反馈微调过程。对这两部分的分析主要是从深度学习模型的重构能力这一方面进行展开。

(2)深度学习模型的构造和选取。通过采用具有特定属性的样本集合，来训练构造不同的深度模型，通过多个样本集合来形成与其对应的多个深度学习模型，然后采用这些模型进行识别，进而对比分析各个深度学习模型的识别效果，最终选择出性能最优的模型。

(3)PCA 算法与深度学习模型的对比。将深度学习模型与 PCA 进行对比实验，进而分析两种降维方法的优劣。这里针对 PCA 的特性，深度学习模型分别与两种 PCA 方法进行对比，一种是满足一定贡献率的 PCA 算法，另一种是满足降低到指定维度的 PCA 算法。

(4)基于深度学习的 BP 识别算法的性能分析。这部分主要包含两个方面的内容，一个是分析失衡训练集对分类器识别精度的影响；另一个是通过多重交叉验证来对过拟合点进行分析，从而明确深度学习模型对 BP 识别算法的影响。

## 第2章 视频人脸检测识别方法研究概述

本章主要从三个部分介绍了视频人脸检测识别方法的研究概述，分别是：人脸检测 Adaboost 算法概述、深度学习概述和基于深度学习的识别算法概述。在人脸检测 Adaboost 算法概述中主要说明了 Adaboost 算法的基本思想与原理；在深度学习概述中主要介绍了深度学习的基础理论和设计模型；在基于深度学习的识别算法概述中主要介绍了 BP 神经网络和 SVM 两种算法。

### 2.1 人脸检测Adaboost算法概述

Adaboost 算法是一种统计方法，该算法所使用的 Haar 矩形特征总共包括有 4 种形式<sup>[25]</sup>，从图 2-1 中，可以看出各个矩形特征之间的区别。图 2-1 中两个矩形 A 和 B 的特征取值均是白色矩形中的像素值之和减去黑色矩形中的像素值之和，矩形的特征是两翼白色框的加和与中间黑色框的的差，最后矩形 D 的特征为上对角白色框的加和与下对角黑色框的加和做差。由此可以看出，该特征所包含的个数是十分庞大的，举例来说，对于一个大小为 24\*24 的检测器模板，存在的相应矩形框的个数则是有 4 万多个。

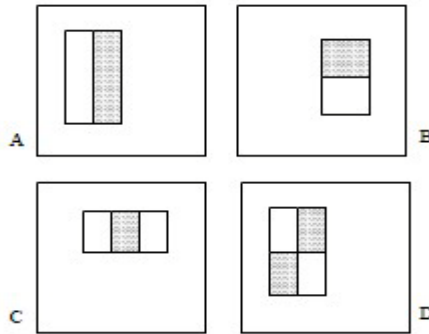


图 2-1 Haar 矩形的四种框结构

可在整型图像的基础上，来快速计算得到 Haar 的特征值，如图 2-2 所示。图中任意一个 $(x,y)$ 像素点的取值定义为公式 (2-1)。

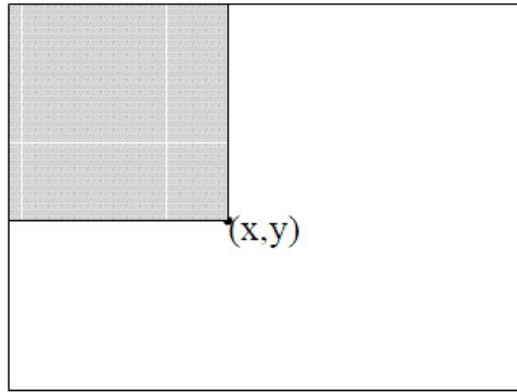


图 2-2 整型图像

$$g(x, y) = \sum_{x' \leq x} \sum_{y' \leq y} f(x', y') \quad (2-1)$$

式中  $(x, y)$ —任意一个像素点；

$f$ —任意图像；

$g$ —任意图像  $f$  的整型图像。

对矩形框内所有的像素点进行求和的计算在整型图像的基础上变成了查表的过程。作为一个非对称的分类问题，在人脸检测问题中，背景中的非人脸样本占据了大多数，并且可以通过很简单的判别规则去除掉大部分的样本，因此采用了级联结构的分类器。该种类型的分类器通过由一系列的单级强分类器级联<sup>[26,27]</sup>得到，如果任何一个输入样本被判断为负面的，那么这个样本就会被过滤掉，只有那些没有被任何分类器过滤掉的输入样本才会被判定为正面的。

## 2.2 深度学习概述

随着 Hinton 在 DBN 上的研究不断的深入，使得他最终通过构建深度神经网络实现了系统学习效率的显著提升。深度学习算法因为使用了多层神经网络，因此它具备更强的表达能力，可以对复杂的客观事物进行描述。针对深度学习算法的这一特点，Hinton 设计了一种算法“greedy layer-wise unsupervised learning algorithm”，通过这种算法来有效的对深度模型进行训练。

这种算法本质上是一种贪婪算法<sup>[28]</sup>，它的基本原理是，首先构造一个拥有多层的人工神经网络，在这个多层模型中，所处在模型的层次越高则说明这一层对可见层输入样本的表示就越抽象，相反如果处于模型的较低层，那么它仅仅能够表示输入样本的低维特征。因此整体说来，这个算法的总体训练的过程就是首先对输入样本进行简单的表示，然后随着所在深度模型层次的不断提升，开始对输入样本进行越来越抽象的表示，最终得到对样本的本质表示的过程<sup>[29,30]</sup>。

## 2.2.1 深度学习基础理论

人脸图像不同人拥有着不同的特征，这些特征复杂而且多变，这成为了人脸识别领域所面临的最大挑战之一，考虑到通过深度学习的方法所构造的深度模型能够对复杂的客观事物进行有效的表示，因此利用深度学习算法为人脸进行建模，并利用建模后的深度模型来对人脸进行识别就拥有着重要的意义。首先通过识别模型来进行正向的转换，使得系统获得对输入样本的各种抽象表示，然后在此基础上，反向构建生成模型，利用生成模型来重构每一层对输入样本的表示，最后在各个层的表示与重构目标之间进行反复调节，最终得到一个符合要求的深度学习模型，并通过此模型来生成对输入样本的本质表示。其过程如图 2-3 所示展示了对于输入样本为一个图片的深度模型的形成过程。

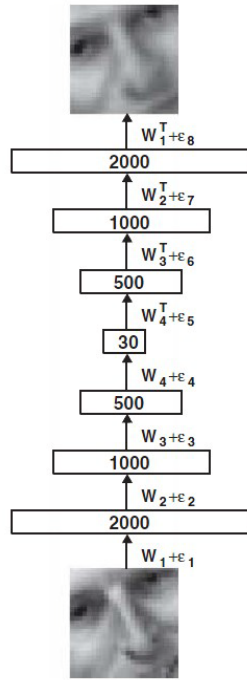


图 2-3 基于人脸图片的深度模型形成过程

对于一个图像来说通过深度学习进行自下而上的深度表示的过程可以通过图 2-4 表达出来，通过这个图可以发现，深度模型的认知过程的输入层就是图像的像素点，然后经历了中间的多层表示，最终得到了对输入样本的本质表示：MAN 和 SITTING。



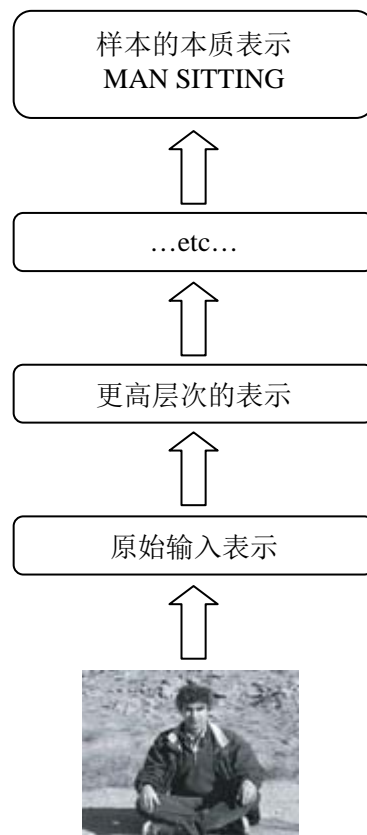


图 2-4 深度学习模型对目标的抽象理解

人脸识别是一个十分复杂而且十分具有挑战性的课题，因此深度学习算法的提出为人们通过构造深度模型的方式来解决人脸识别的问题提供了可能。通过每一层对人脸的抽象表示，以及层与层之间连接权值的不断调整，深度学习模型可以很好的对复杂背景下的人脸进行有效识别。有两点原因使得深度学习算法能够被广泛的应用：

(1) 人类的大脑是一个多层结构。科学研究表明，当人看到普通事物时，视觉皮质会在一系列区域产生神经兴奋，在产生这种神经兴奋的区域里存在一个对输入的代表和一个向下一层传播的生物信号流。在每个层次都有一种对神经兴奋的代表，随着生物信号流的不断传播，代表的形式越来越抽象，最终达到了最高层次的抽象表示。与此同时，在这里人类大脑中的各个中间的代表层是紧密分布的，而靠近可见层的层次则是相对稀疏的。

(2) 人类的认知过程就是一个深度模型的构建过程。人类的学习和认知过程总是先从最原始最简单的知识开始学起的，然后随着学习的不断深入开始去了解更加复杂和抽象的概念。以语言学为例，人类首先学习的是各个单词，然后是会使用由这些单词组成的句子，最后能够通过不同的句子进行组合形成能够表达自己思想的文章。

## 2.2.2 深度学习设计模型

如图 2-5 所示，展示了基本的深度学习模型的调节过程，而本文中的识别模块就是建立在深度学习模型的基础上的。在采用深度学习算法的过程中首先要进行预训练，这个预训练的过程是通过逐层之间以受限制的玻尔兹曼机<sup>[31,32]</sup>为模型进行不断训练实现的。这里所谓的受限制的玻尔兹曼机（RBM）是指一种玻尔兹曼机，它只有一个可见层和一个隐藏层。

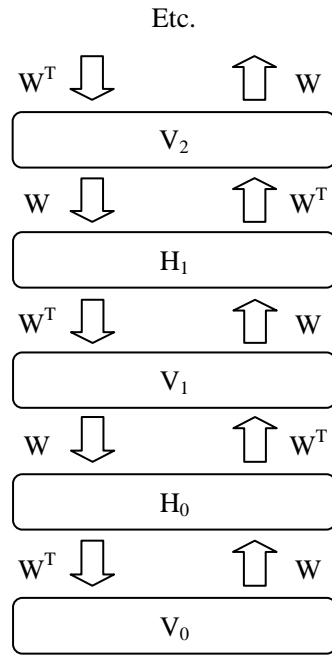


图 2-5 深度学习模型的调节过程

对于一个 RBM 来说，它的可见层  $v$  和隐藏层  $h$  遵循着公式 (2-2) 的联合分布：

$$P(v, h) = \frac{1}{Z} e^{-h'Wv - b'v - c'h} \quad (2-2)$$

式中  $Z$ —联合分布的归一化常量；

$b$ —可见层的偏置；

$c$ —隐藏层的偏置；

$W$ —一层与层之间的权值矩阵。

在公式 (2-2) 中，指数的部分被称为能量函数：

$$energy(v, h) = h'Wv + b'v + c'h \quad (2-3)$$

这里用  $Q(h|v)$  和  $P(v|h)$  表示层与层之间的条件分布，其中，通过使用  $sigm(x) = 1/(1 + e^{-x})$  可以得出它们的公式如式 (2-4) 和式 (2-5)：

$$P(v_k = 1|h) = sigm(-b_k - \sum_j W_{jk} h_j) \quad (2-4)$$

$$Q(h_j = 1|v) = \text{sigm}(-c_j - \sum_k W_{jk} v_j) \quad (2-5)$$

对于受限制的玻尔兹曼机需要特别考虑两种运行模式<sup>[33,34]</sup>：（1）钳制状态，在这种状态下玻尔兹曼机可见层的神经节点会被钳制到某种特定状态，而这个状态是由整体环境所决定的；（2）自由运行状态，在这种状态下无论是可见层还是隐藏层，各个层的神经节点都可以自由运行，不会被限制到特定的状态上。

预训练的过程完成之后，紧接着是基于识别模型和生成模型的反馈微调。首先通过层与层之间的正向链接，进行识别模型<sup>[35]</sup>的构建过程，这一过程结束后将得到对原始样本的各个层次的抽象表示；然后再通过各个层之间的反向连接，来构造生成模型，以此来对各个层的抽象表示进行重构；最后通过不断的对整个模型的各个层参数进行调节来完成整体的无监督训练过程。

最终通过深度学习方法构建了深度学习模型，进而通过这个深度模型很多复杂的概念被逐层的进行了深度表示，最终得到了事物的本质特征。将深度学习算法应用到人脸识别中，通过反复的微调不断的提高对原始样本的表达能力<sup>[36]</sup>，并在自下而上的转化过程中不断的形成对原始信息的更加抽象的表示，最终达到了让机器理解人脸图片信息的目的，从而提高了人脸识别的精度。

通过以上对深度学习模型的理论分析，本文实现了一个深度学习系统，这个系统主要分为两个过程：一，预训练过程，预训练过程的主要目的是确定每层之间的初始权值，采用的方式是逐层进行 RBM 调节，最终找到各个层间的较好的权值；二，微调阶段，通过整体运行深度学习模型，先自下而上进行识别模型的转化，然后再自上而下进行生成模型的转换，最后在原始表示和表示之间进行微调，达到提取样本本质特征的目的。

## 2.3 人脸识别算法概述

通过深度模型，可以得到样本的本质特征，因此本文可以在样本本质特征的基础上做识别，以下是两种常用的识别方法。

### 2.3.1 BP神经网络

BP 神经网络是研究深度学习模型中，比较常用的识别方法。在神经网络中 BP 神经网络是典型的分层结构，样本从输入层进入 BP 神经网络之后，通过隐藏层最终传递到输出层。反馈存在于动态系统，系统一个样本的输出部门影响作用于该元素的输入，基于误差的反馈算法是前馈神经网络通常采用的学习算法，即 BP 算法。BP 算法的整个学习过程可以分为两个部分：正向转换决策部分和反向反馈调节部分。

（1）正向转换决策部分。首先样本进入输入层，转化为第一层的表示，接着这

种表示通过隐藏层的神经元运算后，最终传递到输出层形成最后一层的表示。在整个的运算过程中神经网络各个层之前的权值大小保持不变，每一层的神经节点只会影响和它直接相连的上一层的神经节点，而特定一层的神经节点之间是没有直接联系的。如果输出层所得到的结果没有达到预期的目标，那么整个的训练学习过程就应该从工作信号的正向传递过程转入误差信号的反向传递过程。

(2)反向反馈调节部分。误差信号是由输出层的输出和预期目标运算后所得到的，在反向传递过程中，误差信号开始于输出层，反向通过隐藏层最终传递到输入层。误差信号向后传递每当通过一层时，前一层神经节点的阈值和位于这两层之间的权值和都会被调整。

人工神经网络的本质就是由许多神经节点相互关联所形成的一个综合系统。在这个系统中，虽然每个神经节点内部的结构十分简单，但是这些神经节点却可以构造出一个规模巨大、结构复杂的神经网络系统，通过这个系统对外部的复杂客观事物进行表示和理解。人工神经网络的有两个显著特点：(1)规模庞大的并行分布式结构；(2)神经网络对复杂客观事物的表示和理解能力。这两个显著特点让人工神经网络能够解决一些现在还不能解决的复杂和规模庞大的问题。泛化的过程就是指神经网络通过学习之后，对于不在训练样本集合中的数据也能够产生合理的输出结果。

因此人工神经网络的学习法则应该是：如果神经网络的输出结果和预期相比是错误的，那么就通过神经网络的训练学习过程，使神经网络输出同样错误的可能性降低。首先，初始化整个神经网络的连接权值，通过取在  $(0, 1)$  区间内的随机值的方式来初始化各个权值，将所需要的特征样本作为输入进入网络，网络将输入进行不断的抽象表示，通过每一层的权值运算，最终得到神经网络的输出，然后通过输出的内容与目标进行比对，从而对整个人工神经网络进行反向调节，使得输出与目标不断的接近。采用这种学习方法的神经网络进行多次学习之后，整个神经网络判断的正确率将显著地提高，最终当正确率达到系统设计要求的时候，整体的反复调节过程才停止。当神经网络整体的运行过程结束以后，就意味着神经网络对输入进来的样本特征的学习获得了成功，与此同时这个被成功训练的神经网络已经通过它的连接权值将用于决策的样本特征记录了下来，当下次这个人工神经网络在此被用于样本特征的识别时，就可以基于它的结构与连接权值对样本进行快速的决策。

单一隐藏层的前馈神经网络，又称为三层前馈神经网络，这三层包含：用于样本输入的可见层、用于样本抽象表示的隐藏层和用于决策的输出层。整个神经网络有以下特点：每一层的神经节点只和相邻层的神经节点相互连接，神经节点在同一层内部彼此之间没有连接，同时各个层神经节点之间并没有反馈连接，从

而构成了拥有三层架构的前馈型神经网络。

### 2.3.2 支持向量机

支持向量机 SVM(Support Vector Machine)作为一种可训练的有效分类方法<sup>[37]</sup>。使用一种特定的映射方式，将原始的训练样本映射到高维空间，在高位空间中，去努力搜索线性最佳超平面，用于作为一个决策边界对不同类别进行区分。

Vladimir Vapnik、Bernhard Boser 和 Isabelle Guyon 发表了第一篇有关支持向量机的文章。他们通过对统计学习基础理论的多年研究，提出了一种设计最优准则，用于线性分类器。其演变过程可以从线性可分情况开始，然后逐渐发展到线性不可分的情况，最终发展到使用非线性的函数中来，这种分类器就被称为支持向量机(简称 SVM, Support Vector Machine)。支持向量机针对复杂的非线性决策边界的模拟和建模能力是十分准确的。在与其他模型的对比中支持向量机在克服过拟合问题方面表现的尤为突出。SVM 主要被用于预测和分类，它的主要应用领域包括语音识别，对象识别，图像识别等。

SVM 的主要思想包含两个方面：(1) 针对线性可分情况进行特定处理分析，(2)针对线性不可分情况进行特定处理分析。其中对于线性不可分的情况，通过采用非线性的映射方法，将训练样本从线性不可分的低维空间映射到线性可分的高维空间，进而在高维样本空间中采用之前已经设计好的线性可分处理方法来进行分析。SVM 的一般特征为：

(1)SVM 学习过程可以被看成是一个优化寻找最优解的过程，因此可以采用之前设计好的有效方法去寻找和发现目标函数的全局最小值。而其他的一些学习分类方法（如人工神经网络和规则分类器）都采用一种基于贪心学习的方法来在问题空间进行寻找，这种方法通常仅仅能够得到局部范围的最优解。

(2)SVM 为了对模型进行有效控制通过最大化决策边界边缘的方式来实现。然而，尽管如此，SVM 还需要许多其他的参数，如引入松弛变量和采用核函数等。

(3)SVM 最常应用在解决二类问题上。当推广到多类问题上时，本文希望找出一个最佳的超平面，这里的超平面就是本文所要寻找的决策边界。SVM 来处理多类问题的本质就是寻找最大边缘超平面。根据拉格朗日公式，计算最大边缘超平面的公式为 (2-6)。

$$d(X^T) = \sum_{i=1}^l y_i \alpha_i X_i X^T + b_0 \quad (2-6)$$

式中  $y_i$ —支持向量  $X_i$  的类标号；

$X^T$ —检验向量；

$\alpha_i$  和  $b_0$ —通过 SVM 算法自动确定的数值参数；

1—支持向量的数量。

**SVM** 方法中的非线性映射机制是把线性不可分的样本，从低维特征空间映射到一个高维甚至无穷维的特征空间中，最终使得在原始特征空间中无法实现线性可分的问题，可以在新的特征空间中解决。总的说来，就是通过样本特征维度提升的方式来实现样本特征的线性可分。样本特征维度的提升过程就是把样本特征由低维空间向高维空间做映射，这样做虽然能解决样本特征的线性不可分的问题，但是通常这也会增加运算的复杂度，甚至可能会引起“维数灾难”，正因为这样，**SVM** 的应用范围受到了很大的限制。尽管如此，**SVM** 作为一种独特的解决问题的方法，引起了很多研究者的兴趣。在回归和分类等问题方面，经常会出现一些情况，在这些情况下样本集在低维样本空间无法被线性分解，而被映射到了高维特征空间中之后就可以通过搜索最大边缘超平面的方式实现线性可分。如前所述想高维空间映射的过程很可能会带来计算上的复杂化，为了解决这个难题，**SVM** 采用了核函数的展开定理和计算理论，不需要明确了解非线性映射的公式，因为样本特征是在高维空间中建立的线性可分的学习机构，因此和普通线性模型相比，不仅没有增加计算的复杂度，而且在一定程度上避免了“维度灾难”。

## 2.4 本章小结

本章主要在整体上从三个方面对视频人脸检测识别系统的相关知识进行介绍，分别是人脸检测的Adaboost算法概述、深度学习概述和基于深度学习的识别方法概述。在Adaboost算法概述部分，本章具体说明了Adaboost算法的基本思想和原理。在深度学习概述部分，主要是从深度学习基础理论和深度学习设计模型两个方面来系统的介绍了深度学习理论。在基于深度学习的识别方法概述部分，主要介绍了BP人工神经网络和SVM两种重要的识别方法。

## 第3章 基于深度学习的人脸识别算法

基于深度学习的人脸识别算法的核心是深度学习算法。通过使用深度学习模型，并与上层识别算法相配合将含有多个说话者文件夹中的不同人脸进行识别和分类。本文在第二章已经分析了深度学习的理论设计模型，为了能够将深度学习算法引入到本文所研究的基于视频的人脸识别方法中，本文必须从系统的角度来实现整个深度学习过程。如图 3-1 即为本文所设计的深度学习系统构建。

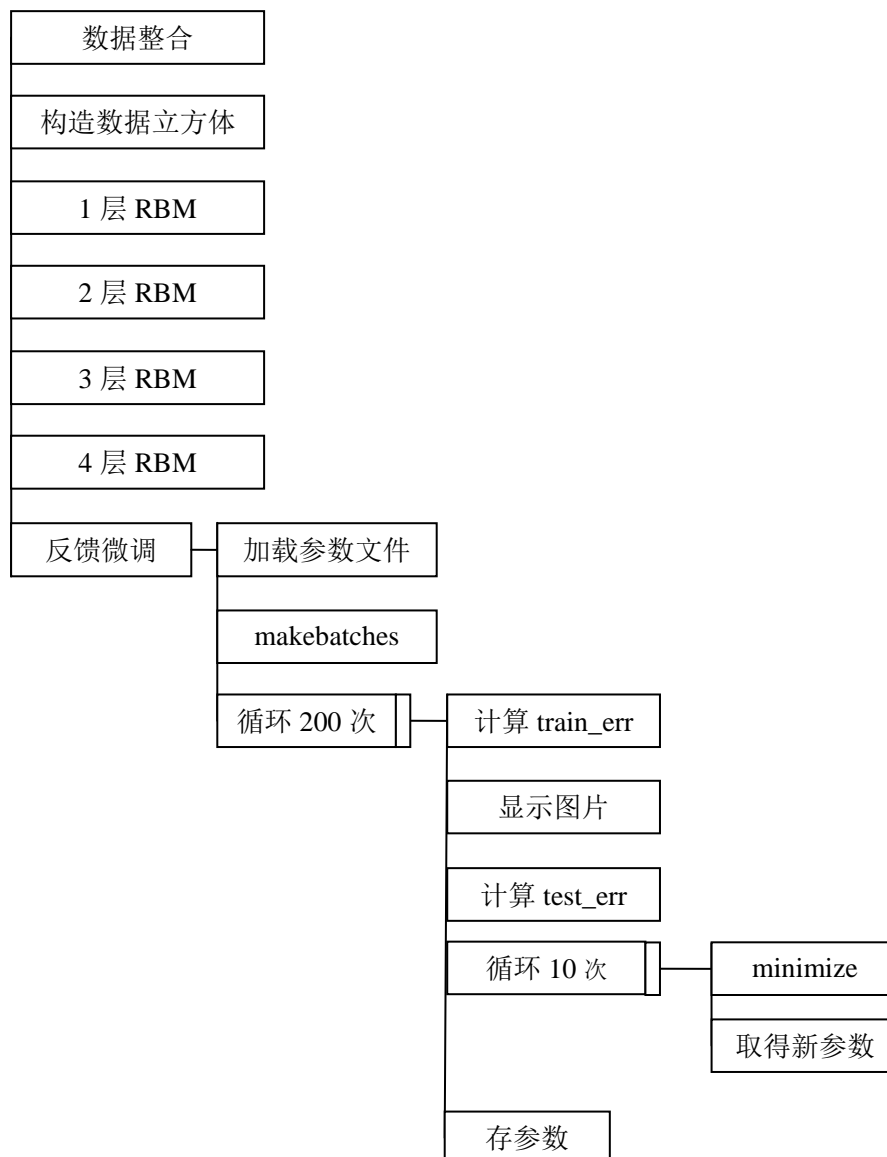


图 3-1 本文所设计的深度学习系统模型

### 3.1 数据整合

通过检测系统提供的标准人脸图片集，将每个图片转化为灰度图片，并把它们的格式从  $28*28$  改变为  $784*1$ 。从而针对每个人的所有图片形成一个二维矩形，其形式如图 3-2 所示。

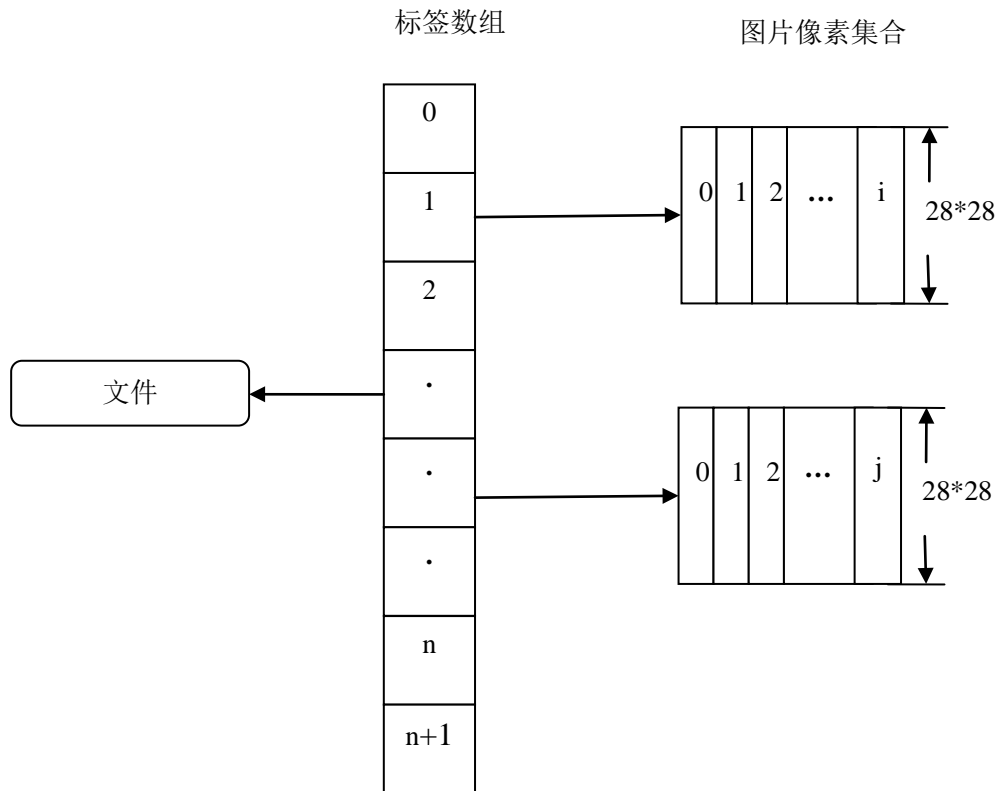


图 3-2 将每个人的所有图片形成一个二维矩形

图 3-2 中，本文将图片从  $28*28$  改变为  $784*1$ ，并形成矩阵，这个矩阵与一个数组相对应。数组的下标对应各个不同的分类。当把所有的训练样本集合全部处理完毕后，本文将整合后的数据存入文件中，便于系统下一阶段的调用。

### 3.2 构造数据立方体

为了便于整个系统对大批量样本进行处理，可以将数据格式化为统一的格式，在这里本文用通过数据整合得到的数据构建数据立方体，将这个立方体作为深度模型读入数据的统一格式，如图 3-3 所示。



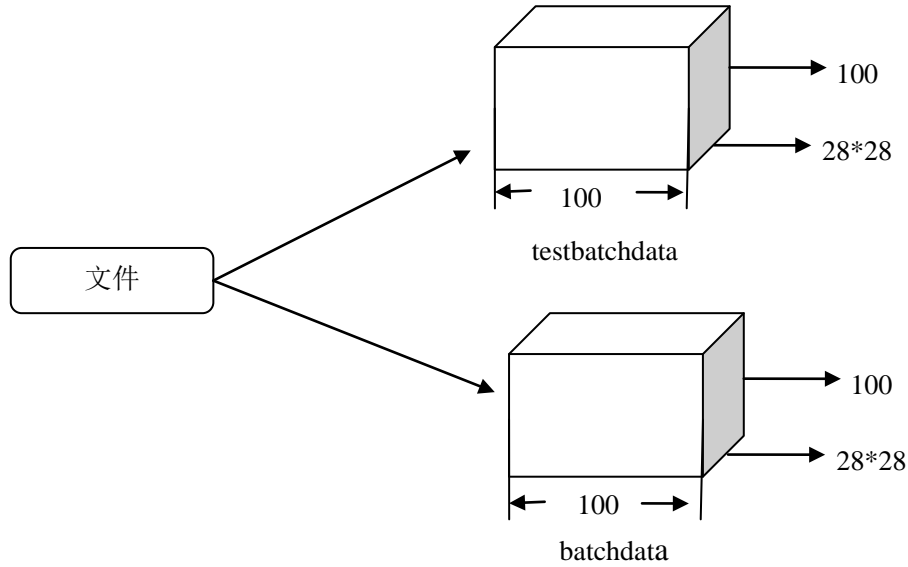


图 3-3 构造用于深度学习模型的数据立方体

图 3-3 中展示的立方体是一个针对 10000 个训练样本所构建的数据立方体。对于一万个样本的情况，本文首先将他们分成 100 个小组，每个小组有 100 个样本，构造立方体的时候，x 轴坐标表示一个小组内不同样本的编号，y 轴坐标表示一个小组中特定一个样本的维度，z 轴表示小组的个数。其中值得注意的是在图 3-3 中所指的文件，就是在数据整合的时候所生成的文件。

### 3.3 调节RBM

RBM 调节是整个深度模型系统非常重要的部分，它以层层递增的方式，利用受限制的玻尔兹曼机模型来调节每个相邻两层之间的权值，通过这种方式来初始化整个深度模型系统的权值。从这个方面来看，RBM 的调节对深度模型至关重要，因为在普通的神经网络的训练过程中，一个最难解决的问题就是如何选择合适的初始化参数来赋值整个神经网络，如果这个参数选择的不好，往往会导致神经网络训练和测试效果的下降。

图 3-4 展示每层之间 RBM 调节的具体过程。首先由可见层向隐藏层转换，经历了这次转换之后，本文以隐藏层为基准进行抽样，得到隐藏层各个节点的状态，再反向由隐藏层向可见层转化，这次转化结束之后，还要进行最后一次的由可见层到隐藏层的转换。之所以要进行三次转换的原因是通过三次的转换，为 RBM 内部的参数调节提供目标。完成了三次转换之后，本文分别得到了可见层和隐藏层的重构目标，通过降低重构对象与原对象的差异来达到调节 RBM 参数的目的。

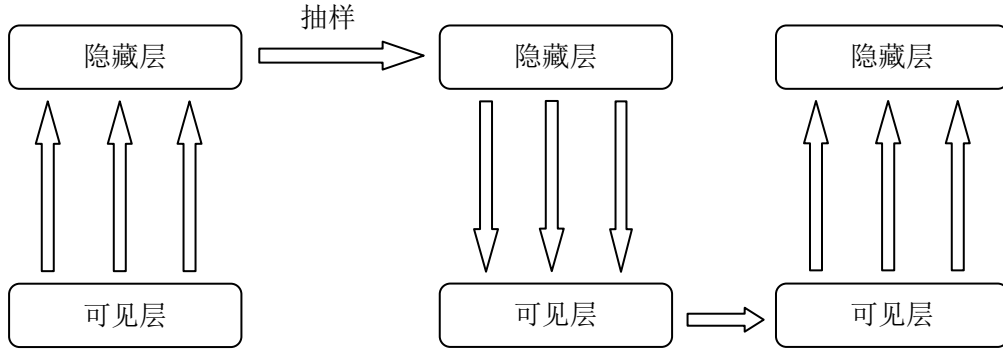


图 3-4 每层之间 RBM 调节的具体过程

基于图 3-4 所描述的 RBM 调节的具体过程, 表 3-1 从算法实现的角度对 RBM 算法进行了清晰的描述。

表 3-1 RBM 调节算法

RBM 调节算法:  $\text{RBM}(W, b, c, v_0)$

$W$  是 RBM 层与层之间的链接权值矩阵

$b$  是 RBM 隐藏层的偏置

$c$  是 RBM 输入层的偏置

$v_0$  是 RBM 训练样本集合中的一个样本。

对于所有隐藏层的神经节点  $i$ :

- 计算  $Q(h_{0i} = 1 | v_0)$ , 也就是进行层与层之间的映射运算  $\text{sigm}(-b_i - \sum_j W_{ij} v_{0j})$
- 依据  $Q(h_{0i} = 1 | v_0)$  进行抽样得到  $h_{0i}$

对于所有可见层的神经节点  $j$ :

- 计算  $P(v_{1j} = 1 | h_0)$ , 即进行层与层之间的映射运算  $\text{sigm}(-c_j - \sum_i W_{ij} h_{0i})$
- 依据  $P(v_{1j} = 1 | h_0)$  进行抽样得到  $v_{1j}$

对于所用隐藏层的神经节点  $i$ :

- 计算  $Q(h_{1i} = 1 | v_1)$ , 即进行层与层之间的映射运算  $\text{sigm}(-b_i - \sum_j W_{ij} v_{1j})$

最后更新链接权值的偏置参数:

- $W = W - \varepsilon(h_0 v_0' - Q(h_1 = 1 | v_1) v_1')$
- $b = b - \varepsilon(h_0 - Q(h_1 = 1 | v_1))$
- $c = c - \varepsilon(v_0 - v_1)$

### 3.4 深度模型的反馈微调

深度模型的反馈微调主要通过三个过程来实现：加载参数文件、构造数据立方体和循环调节。其中加载参数文件和构造数据立方体的过程主要是在做前期的准备工作，而最后的循环调节才是整个深度模型反馈微调机制的重点。图 3-5 描述本文设计的深度模型的整体架构，整个深度模型分为 5 层，随着层次的增加，深度表示的维度逐渐降低，在深度模型的反馈微调阶段，先通过识别模型自底向上转换，到了最高层 layer4 后，再进行从上而下的生成模型的转换，从而生成对各个层次的重构。最终通过对原始表示和重构表示的不断调节实现两者的误差达到可以接受的程度。

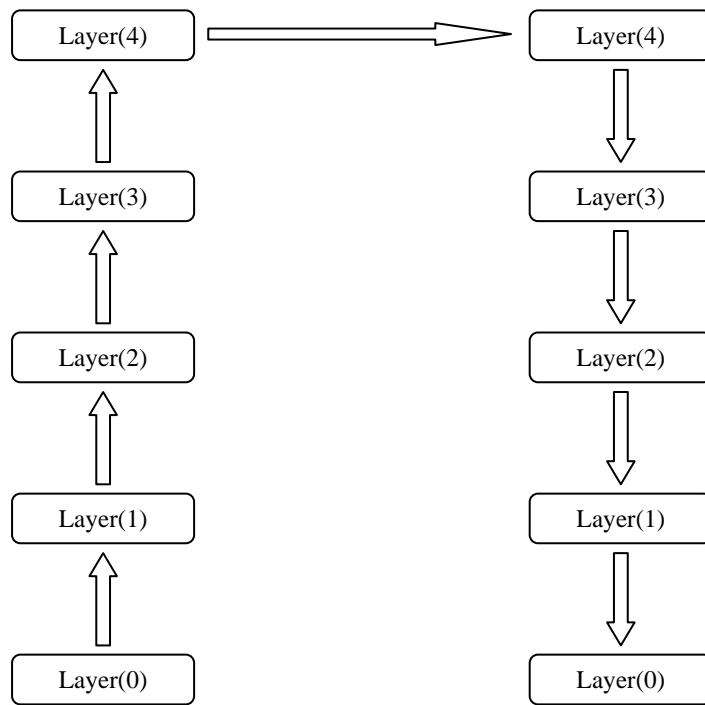


图 3-5 深度模型的反馈微调流程

(1) 加载参数文件。通过图 3-1 所描述的深度学习系统模型可以直观的看到，在深度模型反馈微调执行之前，分别经历了四次的 RBM 调节，通过这四次的调节，使整个深度模型层与层之间拥有了自己的初始链接权值。加载参数文件这一过程的主要任务就是将这些初始链接权值加载到深度学习模型中，作为模型的初始化参数。

(2) 构造数据立方体。这一过程与 3.2 节所描述的构造数据立方体的过程完全一致，之所以要在此重新构造数据立方体的原始是，经历了 RBM 的调节之后，原来所构建的数据立方体中的原始数据被破坏了。

(3) 循环调节。这一过程是整个深度模型反馈微调的主体部分，要对整个系统进行微调，这里采用将自底向上的识别模型和自顶向下的生成模型相结合的方式进行微调。通过识别模型，本文可以得到深度模型对输入样本最初的各个层次上的表示形式，并得到一个深度模型对样本的一个最高抽象表示形式；通过生成模型，本文可以从深度模型的最高抽象表示形式出发，重构深度模型对样本的各个层次的表示。这样做就为原来每个层级的表示提供了调节的目标。通过不同层次之间的不断调节，使生成模型可以重构出具有较低误差的原样本，这样就得到了此样本的本质特征，即深度模型的最高抽象表示形式。

### 3.5 本章小结

深度学习算法作为基于深度学习的人脸识别算法的核心是本章重点阐述的对象。在这部分里本章从数据整合、构造数据立方体、**RBM** 调节和深度学习模型的反馈微调四个方面全面展示了整个深度学习算法的实现过程。其中深度学习模型的反馈微调是整个深度学习算法的核心，它主要由三个过程组成，分别是：加载参数文件、构造数据立方体和循环调节。

## 第4章 深度学习实验与分析

本文的主要研究内容包含深度学习模型的训练、深度学习模型的构造和选取、PCA 算法与深度学习的对比，基于深度学习的 BP 识别算法的性能分析四个部分。本文所设计的实验就是围绕着这四个方面展开的。深度学习模型的训练主要包含两部分：**RBM** 训练和深度模型的反馈微调，这两类实验都是通过使用不同的循环调节次数，来对深度模型的重构能力进行分析。在深度学习模型构造与选取中，通过多组不同的样本集合来构造深度模型，并深入的分析各个深度模型的特性，最终选取出最优的深度模型。PCA 算法与深度学习对比实验中，通过两者的对比来说明深度学习算法性能的优越性。基于深度学习的 BP 识别算法的性能分析中，主要包含两部分：失衡训练集实验和五重交叉验证实验。失衡训练集实验的主要目的是，通过采用失衡训练样本来分别对加入深度模型与没有加入深度模型的分类器进行训练，来最终证明深度模型能够提取出样本的本质特征，而在深度学习的五重交叉验证实验中，主要为了分析深度学习模型对识别算法的影响。

### 4.1 深度学习模型的训练

深度学习模型的训练主要包含两个部分：**RBM** 调节和深度学习模型整体的反馈微调。关于深度学习模型训练的实验就是围绕着这两部分展开的。

#### 4.1.1 RBM 训练的实验与分析

**RBM** 的训练是在整个深度模型构建初期所要完成的工作。**RBM** 的训练过程正如上文中所说的，是一个反复转换的过程。针对 **RBM**，这个实验的主要目的是分析 **RBM** 模型的不同循环调整次数，对实验结果的影响，从而选出一个更好的循环值。

图 4-1 和图 4-2，分别是 **RBM** 模型循环了 1 次和 2 次所重构得到的图像，从图中明显可以看出通过 2 次循环的效果比通过 1 次循环的要好。其中图 4-2 基本描述出了整个人脸的基本轮廓，但是图 4-1 却十分模糊，看不出任何人脸的形状。



图 4-1 RBM 模型循环 1 次重构得到的图像

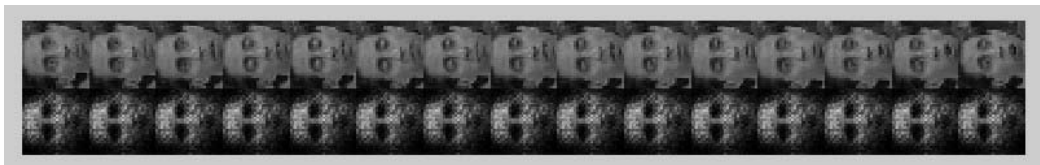


图 4-2 RBM 模型循环 2 次重构得到的图像

图 4-3 到图 4-5，分别是 RBM 模型循环了 5 次、10 次和 30 次所重构得到的图像，从图中可以发现随着循环次数的不断增加，深度模型重构的效果越来越差。其中图 4-3 还勉强可以基本描述出了整个人脸的轮廓，但是图 4-4 却已经变得十分模糊分辨不出人脸的模样；当循环 30 次的时候，利用深度模型所重构的图像更加模糊，只有非常少的一些图像信息，如图 4-5 所示。

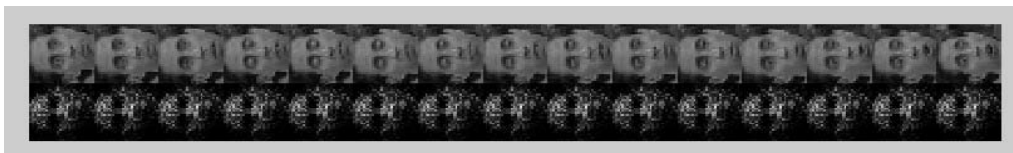


图 4-3 RBM 模型循环 5 次重构得到的图像



图 4-4 RBM 模型循环 10 次重构得到的图像



图 4-5 RBM 模型循环 30 次重构得到的图像

通过这次实验，本文发现深度模型的重构能力和 RBM 调节的循环次数不是呈现着单纯的递增或者递减关系的。而是在开始的时候，随着 RBM 循环次数的增加，重构效果越来越好，但当到了一定程度之后，随着 RBM 循环次数的继续增加，不但重构效果越来越差，而且还会丢失掉原始图像中的很多重要的信息。最终在本文的深度模型中采用十次循环来对 RBM 进行调节。

#### 4.1.2 深度学习反馈微调的实验和分析

在深度学习模型的实验中，本文首先针对深度模型的重构能力进行测试实验。深度模型通过识别模型生成了自下而上的越来越抽象的表示，并利用生成模型重构识别模型中各个层的表示，并不断的循环调整，使重构体与目标体越来越接近。

图 4-6 是图像第一次经历了从下而上识别模型的深度抽象过程，并又经历了自上而下的生成模型的深度重构过程。通过实验可以看出，第一次的深度重构图像还不是很清晰，但是人脸的基本构架已经出来了。

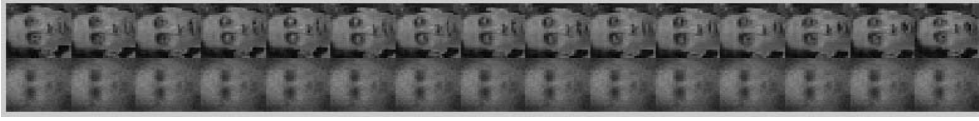


图 4-6 深度模型循环了 1 次之后重构出来的人脸

图 4-7 是循环了 50 次之后重构出来的人脸，图 4-8 是深度模型循环了 150 此重构出来的人脸，图 4-9 是深度模型循环了 200 次的结果，通过这次实验，能够发现，随着微调次数的不断增加，被重构出来的人脸越来越清晰，到迭代两百次的时候，重构出来的人脸和原始图片的人脸的误差已经相当低。这说明将深度学习模型在视频文件的处理上表现出色。



图 4-7 深度模型循环了 50 次之后重构出来的人脸

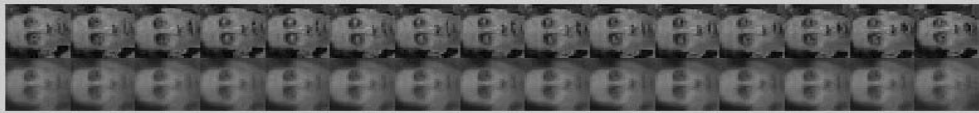


图 4-8 深度模型循环了 150 次之后重构出来的人脸

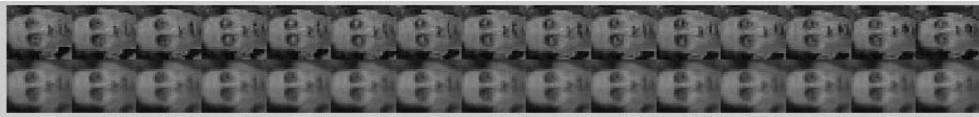


图 4-9 深度模型循环了 200 次之后重构出来的人脸

图 4-10 描述的是在深度模型的训练过程中，重构误差变化的过程。从图中可以明显的发现，随着循环次数的逐渐增加，重构误差在逐渐下降，在循环开始的时候，重构误差的下降非常快，但随着循环次数越来越大，重构误差下降的幅度越来越小，但是依然保持着下降的趋势。最终通过深度模型的不不断调节，使重构误差达到本文预期要求的程度。

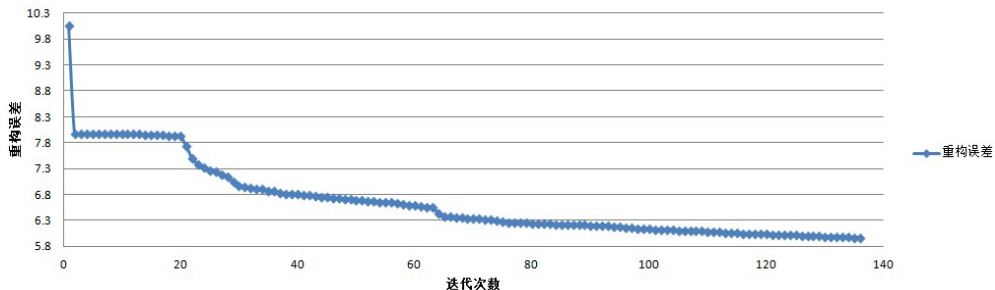


图 4-10 在深度模型的训练过程中重构误差变化的过程

## 4.2 深度学习模型的构造和选取

为了分析深度学习模型的抗干扰能力，实验中构造了不同的样本集合用来生

成深度学习模型。这里的样本集合主要来自于两类视频文件，第一类是《老友记》的视频文件所生成的样本，第二类是《宋飞传》的视频文件所生成的样本。基于这两类视频文件所生成的样本，本文构造了 18 个不同组合的样本集合，其中每个样本集的样本个数均为 9000，以此来生成 18 个深度学习模型。这 18 个样本集可以分为三类：

第一类，《老友记》中有 6 个主要人物，分别是：Chandler、Joey、Monica、Phoebe、Rachel 和 Ross，基于这六个人物形成六个样本集，每个样本集中的元素个数为 9000 个。因此，第一类共 6 个样本集合，可以形成 6 个深度学习模型；第二类，《宋飞传》中有 3 个主要人物，分别是：Elaine、George 和 Jerry，基于这三个人物形成三个样本集合，每个样本集合中的样本总数为 9000 个。同时基于这三个人物进行不同组合。在第二类中一共形成了 9 种不同组合的样本集合，分别是来自《宋飞传》三个主要人物样本集合的不同组合；第三类，《老友记》和《宋飞传》两个视频文件中一共有 9 个主要人物，由这 9 个主要人物的样本集合中各提取 1000 个样本，形成一个新的组合，从《老友记》的六个人物的样本集合中分别提取出 1000 个样本形成又一个新的样本集合，从《宋飞传》的三个人物的样本集合中分别提取出 1000 个样本形成一个新样本集合，因此最后第三类中共有 3 种组合的样本集合。具体各个样本集的组合方式如表 4-1 所示：

表 4-1 不同样本集的组合方式

样本集合名称	组合方式
Chandler	9000 个 Chandler 样本
Joey	9000 个 Joey 样本
Monica	9000 个 Monica 样本
Phoebe	9000 个 Phoebe 样本
Rachel	9000 个 Rachel 样本
Ross	9000 个 Ross 样本
Elaine	9000 个 Elaine 样本
George	9000 个 George 样本
Jerry	9000 个 Jerry 样本
E2_G1	6000 个 Elaine 和 3000 个 George
E2_J1	6000 个 Elaine 和 3000 个 Jerry
G2_E1	6000 个 George 和 3000 个 Elaine
G2_J1	6000 个 George 和 3000 个 Jerry
J2_E1	6000 个 Jerry 和 3000 个 Elaine
J2_G1	6000 个 Jerry 和 3000 个 George
DL_1000_9	9 个人物各 1000 个样本
SF_3000_3	《宋飞传》三个人物，每个人物样本 3000 个
FD_1500_6	《老友记》六个人物，每个人物样本 1500 个



为了得到客观正确的实验结果，本次实验采用了五重交叉验证的方式对每个深度学习模型进行分析。五重交叉验证的实验的具体过程如下：每一类取 2000 个样本，将这 2000 个样本平均分成 5 份每份 400 个，每次抽取一份作为测试样本，余下的 1600 个样本作为训练样本，从而针对每一类进行一次这样的构造过程。为了保证实验的准确性，每一重交叉验证的实验，要重复做三次，然后取三次的平均值作为这一重交叉验证的精度。

如表 4-2 所示，DL\_1000\_9 表示从两个视频的 9 个人物样本集合中，每个提取 1000 个形成的一个样本集；SF\_3000\_3 表示从《宋飞传》的 3 个人物样本集合中，每个提取 3000 个形成一个样本集；FD\_1500\_6 表示从《老友记》的 6 个人物样本集合中，每个提取 1500 个样本形成一个样本集。表中的每个数值单元表示每一重交叉验证，所得到的精度值。

表 4-2 对不同的样本集做每一重交叉验证得到的精度值

	1 样本集	2 样本集	3 样本集	4 样本集	5 样本集
DL_1000_9	0.838750	0.927500	0.851250	0.905000	0.953750
SF_3000_3	0.901250	0.940000	0.886250	0.885000	0.932500
FD_1500_6	0.651250	0.761250	0.820000	0.908750	0.753750

如表 4-3 所示，Chandler、Joey、Monica、Phoebe、Rachel 和 Ross 是《老友记》主要的 6 个人物。这里的人名表达的含义是，由各自人脸样本组成的容量为 3000 的样本集合。

表 4-3 从《老友记》中 6 个人物取人脸样本做每一重交叉验证得到的精度值

	1 样本集	2 样本集	3 样本集	4 样本集	5 样本集
Chandler	0.918750	0.958750	0.878750	0.937500	0.917500
Joey	0.748750	0.928750	0.842500	0.813750	0.805000
Monica	0.716250	0.787500	0.773750	0.775000	0.801250
Phoebe	0.848750	0.890000	0.823750	0.890000	0.918750
Rachel	0.503750	0.776250	0.625000	0.895000	0.793750
Ross	0.831250	0.936250	0.910000	0.946250	0.901250

基于表 4-3 所提供的各个样本集合的精度信息，本文生成了如图 4-11 所示的趋势图。图中，用不同方式标记的曲线分别表示 Chandler、Joey、Monica、Phoebe、Rachel 和 Ross 的精度波动情况。图中的横坐标是识别精度，纵坐标对应的是五重交叉验证的不同样本集。通过比较之后，本文发现，在这所有不同的样本集组合中，Chandler 能够得到更高和更稳定的分类精度。

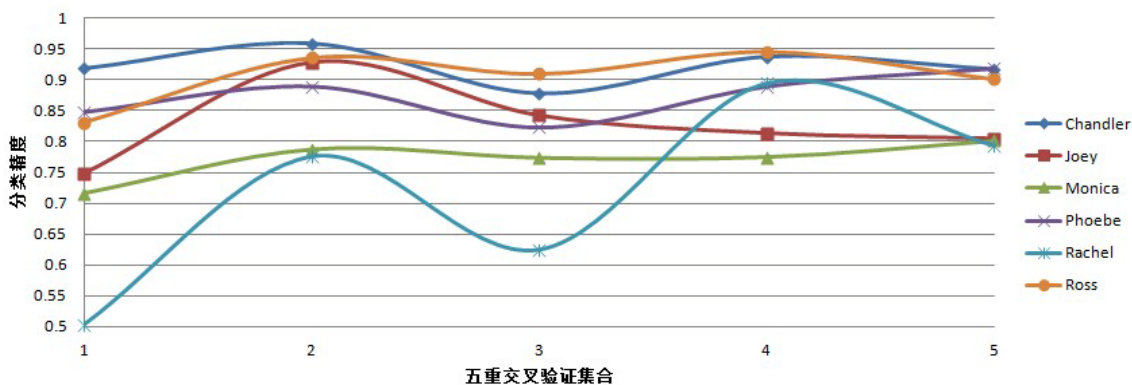


图 4-11 基于表 4-3 所提供的各个样本集合的精度信息所生成的趋势图

如表 4-4 所示，Elaine、George 和 Jerry 是《宋飞传》主要的 3 个人物。这里的人名表达的含义是，由各自人脸样本组成的容量为 3000 的样本集合。其余的表示三者不同组合形成的样本集。

表 4-4 从《宋飞传》中的主要人物取样本集做每一重交叉验证得到的精度值

	1 样本集	2 样本集	3 样本集	4 样本集	5 样本集
Elaine	0.847500	0.895000	0.871250	0.902500	0.842500
George	0.921250	0.937500	0.912500	0.897500	0.878750
Jerry	0.795000	0.873750	0.823750	0.831250	0.871250
E2_G1	0.893750	0.920000	0.835000	0.905000	0.821250
E2_J1	0.703750	0.873750	0.817500	0.881250	0.827500
G2_E1	0.910000	0.862500	0.848750	0.920000	0.912500
G2_J1	0.863750	0.945000	0.852500	0.932500	0.861250
J2_E1	0.760000	0.916250	0.855000	0.926250	0.863750
J2_G1	0.860000	0.941250	0.897500	0.886250	0.943750

如图 4-12 所示，横坐标表示 18 种不同组合形成的样本集，纵坐标表示分类精度。这里用不同的方式标记的五条曲线分别代表着五重交叉验证中的第一到第五重验证集合的精度波动。通过这样的对比观察，来实现选取特定深度学习模型的实验目的。

本次实验要选取的深度学习模型必须能够保证，在不同样本集合下，它的分类精度比较高，并且波动幅度比较小。为了达到这样的要求，本实验计算了 18 个样本集合所形成的深度学习模型在五重交叉验证中的均值和标准差，如图 4-13 中所示，横坐标表示不同的样本组合所形成的深度学习模型，被圆点标记的曲线表示不同深度学习模型精度的均值，被方块标记的曲线表示各个样本集经历了五重交叉验证之后，它们精度的标准差。通过此图可以从整体上观察到各个不同组合样本集所形成的深度学习模型对识别精度的影响。

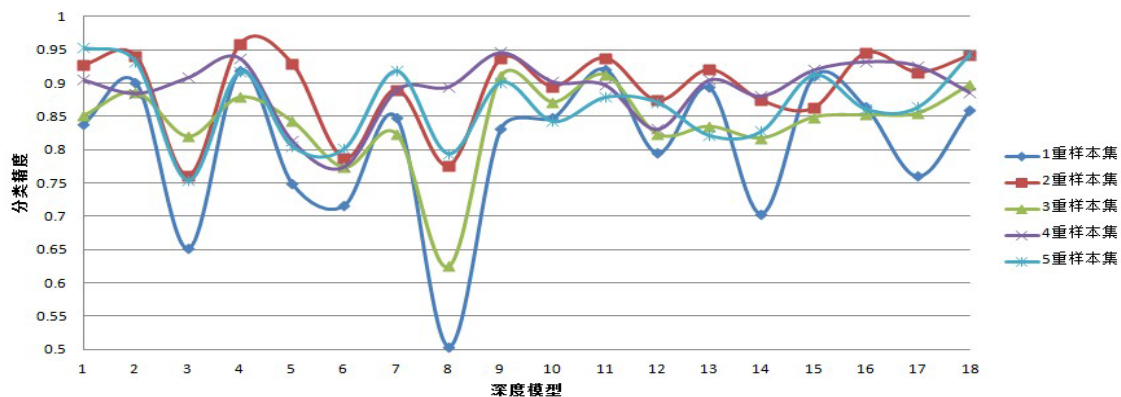


图 4-12 各个样本集的分类精度对比

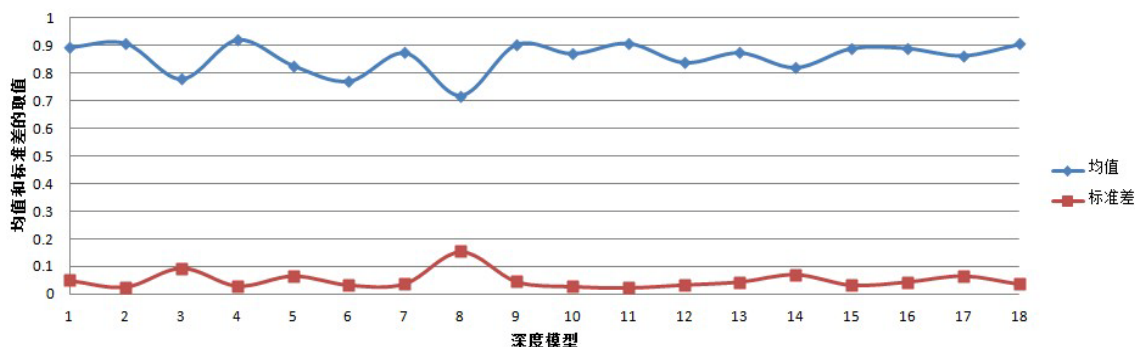


图 4-13 18 个样本集合在五重交叉验证中的均值和标准差

为了有助于本实验更进一步的分析，这里提取出了所有识别精度的均值在 0.9 以上的样本组合进行分析。在本文的实验中，精度均值大于等于 0.9 的深度模型有五组，分别是 SF\_3000\_3、Chandler、Ross、George 和 J2\_G1 这五组样本集所形成的深度模型。这五组满足要求的深度学习模型的均值和标准差如表 4-5 所示。

表 4-5 识别精度均值在 0.9 以上的样本组合的均值和标准差

	SF_3000_3	Chandler	Ross	George	J2_G1
均值	0.9090	0.9223	0.9050	0.9095	0.9057
标准差	0.0007	0.0009	0.0020	0.0005	0.0013

基于表 4-5，本文绘制了这五个样本集均值和标准差变化图像如图 4-14 所示。图中的横坐标表示不同的深度学习模型，用圆点标记的曲线表示的是五个样本集所对应的均值，用方块标记的曲线表示的是它们的标准差，为了便于对比观察，图中将五个样本集标准差的数值统一提升 500 倍。

图 4-14 清楚的展示出，各个样本的均值，和围绕着均值的精度波动情况。在这里 Chandler 样本集合的精度均值虽然在 0.9 以上，但是它的标准差过大，这就说明由 Chandler 样本集生成的深度学习模型受不同训练和测试样本集合的影响很大，因此不适合作为本文要选用的深度学习模型。经过一轮赛选，深度模型样本的选择范围最终缩小到 SF\_3000\_3 和 George 两个样本集上，因为它们两个是排除了

Chandler 集合后精度均值最高的前两个样本集。

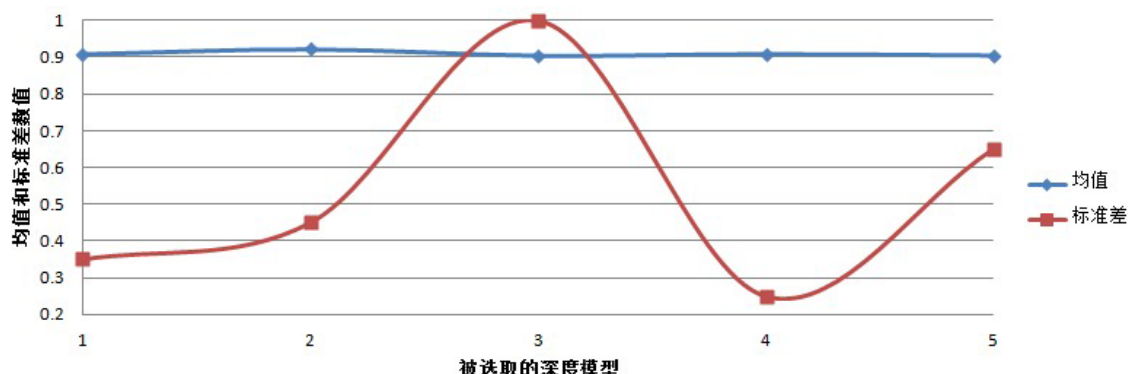


图 4-14 五个样本集合精度均值和标准差变化图像

在这两个样本集合中 George 的精度均值最高，并且标准差最小，似乎应该作为本文深度学习模型的最佳选择，然而经过深入的分析发现，情况并不是这样。在做五重交叉验证的时候，所使用的集合是 George 和 Jerry 的样本集，如表 4-6 展示了 SF\_3000\_3 与 George 两个模型的具体分类情况，其中 S3\_George 和 S3\_Jerry 分别表示基于 SF\_3000\_3 模型分类器对 George 和 Jerry 正确分类的个数，而 S3\_ALL 表示总体正确分类的个数，同样 G9\_George、G9\_Jerry 和 G9\_ALL 表示基于 George 模型对 George、Jerry 和总体正确分类的个数。通过此表可以看出基于 SF\_3000\_3 深度模型对 George 和 S3\_Jerry 的分类是相对均匀的，而基于 George 模型对 George 和 S3\_Jerry 的分类则并不均匀，可以明显的看出对 George 这一类的识别效率过高，这说明由于本实验的训练和测试样本集中拥有 George，才导致基于 George 的深度模型取得了如此好的精度均值和标准差，考虑到本文所选用的深度模型应当对不同样本集有较好的适应能力，因此本文最终选择 SF\_3000\_3 样本集所形成的深度学习模型。

表 4-6 SF\_3000\_3 模型与 George 模型的具体分类情况

	S3_George	S3_Jerry	S3_ALL	G9_George	G9_Jerry	G9_ALL
测试集 1	338	383	721	394	343	737
测试集 2	372	380	752	392	358	750
测试集 3	378	331	709	381	349	730
测试集 4	350	358	708	385	333	718
测试集 5	357	389	746	395	308	703

通过这次实验也可以发现，深度学习模型的抗干扰能力比较强，在 SF\_3000\_3 样本集合中，有三个《宋飞传》的人物。而实验中的五重交叉验证样本集却只有两个人物，在此情况下，SF\_3000\_3 依然表现出了良好的性能，这充分体现出了深

深度学习模型拥有较强的抗干扰能力。然而，深度学习模型的这种抗干扰能力也不是无条件的，当实验中引入了其他视频文件，即《老友记》，发现基于新视频的深度学习模型的表现明显不如基于原始视频的深度学习模型，这表明深度学习模型建立所使用的样本集要与识别样本集相一致才能拥有较高的性能。

### 4.3 PCA算法和深度学习对比的实验与分析

人脸识别中的一个重要部分就是样本特征的有效提取方法。其中线性子空间方法是一种重要的特征提取算法，随着这一算法被不断的深入研究，线性子空间方法被广泛的应用到了许多不同样本集的特征提取中。样本数据经过预处理后通常会得到拥有较高维度的特征向量，因此这就需要通过某种方法来将这些高维向量转化成低维向量，从而减少运算量，并且保证经过转化之后的低维向量能够表达出原始样本的本质特征，线性子空间的方法的重要意义正在于此。线性子空间方法的主要原理就是通过采用线性映射矩阵进行转换的方式来对原始样本进行降维，即公式（4-1）。

$$y = W^T x \quad (4-1)$$

式中  $x$ —原始的高维特征向量；

$y$ —降维后的特征向量；

$W$ —映射矩阵，一般可以通过定义不同的准则函数来寻找映射矩阵  $W$ 。

与 PCA 方法相比，深度模型本身也有着降维的作用<sup>[38]</sup>。原始样本通过深度模型的层层转化，逐渐得到了较高层次的抽象表示，在这一过程中，样本的维度也随着不断的减少，当最终得到了原始样本的本质特征后，也就意味着取得了降维后样本的深度表示形式。

#### 4.3.1 PCA算法基础理论

事实上，在人脸检测与人脸识别中，可以将基于 PCA 的方法划分为两种：第一种是基于自适应 PCA 的方法，第二种则是基于经验 PCA 的方法。在大多数的情况下，首先对人脸图像集进行二元定义，即将人脸图像分为训练集和测试集两种类型的集合。将身份已知的人脸图像集划分到训练集中，对应主分量矩阵  $W$  的 PCA 的投影轴则是根据训练集来确定的，并且也由训练集提供备选身份的样例；同时将假定身份未知的每幅图像划分到测试集中，对算法性能的估计则是通过计算测试集中被正确识别出来的图像比例完成的。根据这样设定的原则，则通过对训练集中被正确识别出的图像比例可以估计算法的性能。在这样的构造原则的基础上，对于训练集来说，其中所包含的身份或图像发生添加或更改的变化，就需要重新训练得到 PCA 的投影轴。

但是，还有另外一种估计 PCA 投影轴的思想。这种思想在引入图库集和查询集两个专用新术语的基础上，进一步提出了训练集的概念，从而形成了对人脸样本集合的三元定义。此处定义的查询集等价于二元定义时的测试集；三元定义中提出的训练集是区别于图库集和查询集的，算法训练则是通过训练集来帮助完成的。为备选身份提供样例则是图库集的功能。依据这样的构造规则，在训练集确定的基础上，PCA 的投影轴也就可以确定了。PCA 投影轴是不会因为被识别身份或代表被识别身份的样例变化而改变的，因此可以用“经验的”来描述这种估计 PCA 投影轴的方式。

数据降维主要使用的是 PCA，对于由一系列图像特征所组成的多维向量来说，多维向量里的元素之间没有可区分性，如果某个元素在全部的例子中都取值为 1，或者说与 1 的区别不大，那么这个元素就是没有区分性的，因此可以将它作为特征来区分，从而导致贡献会非常小。因此本文需要找到变化大的那些元素，即那些方差比较大的维，然后去除掉变化不明显的维，使得特征留下的都是“精品”，而且精简了计算量。如果一个特征是  $k$  维，则说明它的每一维特征与其他维特征均正交（也就是说，在多维坐标系中，坐标轴均是垂直的），那么本文可以通过改变这些维的坐标系，使这个特征能够在某些维上方差大，在某些维上则方差很小。

因此对 PCA 的分析实际上就是求这个投影矩阵的过程，使用这个投影矩阵乘以高维的特征，就能够使得高维特征的维数降低为指定的维数。

寻找  $r$  ( $r < n$ ) 个新变量是 PCA 的目标，使这些变量反映出事物的主要特征，然后可以使原有数据矩阵的规模得到压缩。每个新变量具有一定的实际含义，它们体现的是原有变量的综合效果，是原有变量的线性组合。将这  $r$  个新变量叫做“主成分”，它们是互不相关且正交的，通过它们可以反映得到原来  $n$  个变量的影响。通过主成分分析，然后使数据空间得到压缩，能够在低维空间里将多元数据的特征直观地展现。举例来说，若想将数据的维数从  $RN$  降到  $R3$ ，则需要将多实验条件下、多时间点下的基因表达谱数据（ $N$  维）转换为 3 维空间中的一个点。

#### 4.3.2 PCA 与深度学习的实验分析

为了完成 PCA 实验，制作了一个训练集和一个测试集。训练集由 1000 个 Elaine 的人脸样本和 1000 个 George 的人脸样本构成；测试集由 500 个 Elaine 样本和 500 个 George 样本构成，其中测试集中的样本和训练集中的样本是完全不同的。在 PCA 的实验中，将用向量  $\text{latent}$  表示样本每个特征的贡献率，这里所谓的贡献率是指，样本每个特征对于区分整个样本集的贡献程度，PCA 中所要确定的主成分，就是贡献率最大的样本特征。

通过采用 PCA 的方式对训练样本集进行分析，本文得到了基于这 2000 个样本集的主分量增长曲线，如图 4-15 所示：图中针对  $28 \times 28$  维度的样本，只显示了前



20 个主分量，图中所有主分量的重要性大概占总体的 76% 左右。曲线的形成过程就是各个主分量累加递增的过程。这里的蓝色矩形所对应的就是特征在 latent 中的贡献值。

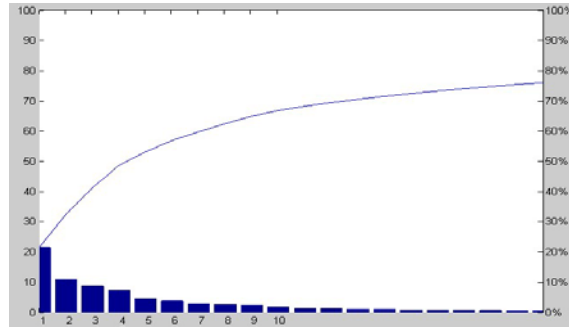


图 4-15 基于 2000 个样本集的主分量增长曲线

除了贡献率 latent 以外，在 PCA 实验中还有两个重要的部分：pca 和 score。其中，pca 是一个  $784 \times 784$  的方阵，实验中所生成的主成分的系数对应于这个方阵中的每一列；这里 score 是一个  $2000 \times 784$  的矩阵，2000 对应着实验中用于训练的 2000 个样本的集合，784 代表每个样本所拥有的特征数量（即样本的维度），它用来表示样本主成分的权重，也就是原来的样本矩阵在主成分空间的映射。

如图 4-16 所示，这是一个三维立方图，图中的每一维代表一个主成分，这里选中的三个主成分分别是在 latent 中贡献率最高的三个。图中的红点表示，2000 个训练样本，是以 score 矩阵的对应值为下标绘制的，因此红点可以理解为训练样本在主成分空间的映射。这里蓝色的射线是以 pca 矩阵的对应值为下标绘制的，射线的长度表示原始特征样本在主成分空间的权重。

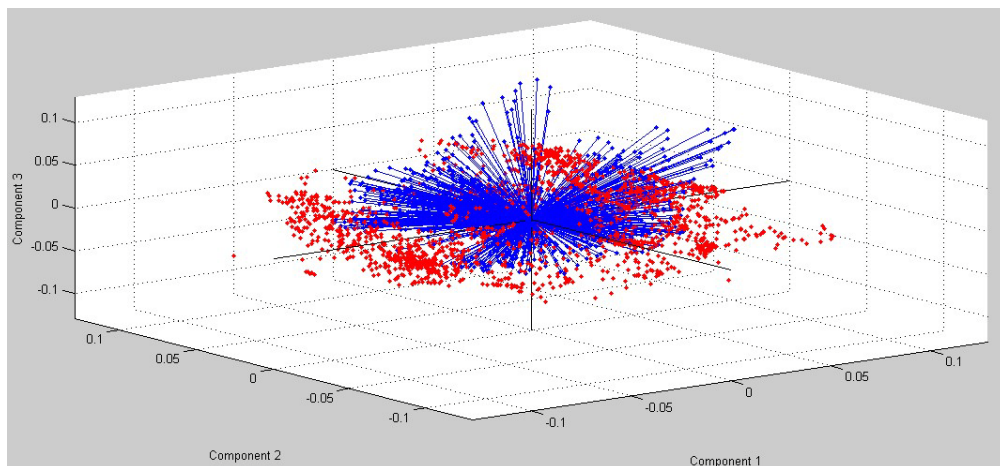


图 4-16 三维立方图

本次实验设置主成分的贡献率要达到 95% 以上，图 4-17 展示了主成分贡献率的增长过程，为了达到 95% 的要求，这次实验使用了 166 个主分量，将样本从原

来的 784 维度降到 166 维度。

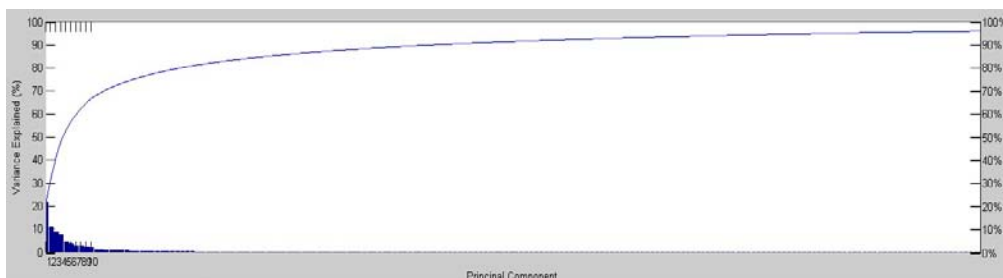


图 4-17 主成分贡献率的增长过程

通过投影变换的 matlab 公式：

$$PCA\_new = X * pca(:,1:n) \quad (4-2)$$

式中  $X$ —原来的样本特征矩阵；

$N$ —在本实验中就是 166（即将原来的样本降维到 166 维度的特征空间）；

$PCA\_new$ —通过降维得到新的样本特征矩阵。

将通过 PCA 降维后的样本作为输入，训练本文的人工神经网络，训练完成后，实验进入测试阶段。本次实验的测试样本由 500 个 George 的人脸样本和 500 个 Elaine 的人脸样本组成。测试样本通过 PCA 降维后，作为神经网络的输入，神经网络以此来进行分类。图 4-18 展示了，经过 2000 个样本训练后的人工神经网络，在测试阶段的表现。图中横轴表示的是每个离散的测试样本，共计 1000 个，纵坐标表示的是经过神经网络分类后所做出的决策。前 500 个样本的理想决策是 1，后 500 个样本的理想决策是 2。其中每个竖线表示一次错分，竖线越密集则说明错分的情况越多。

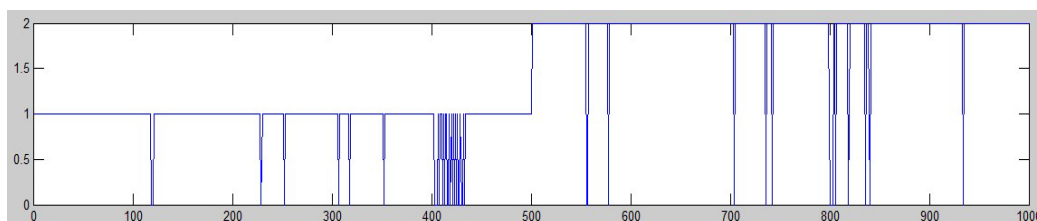


图 4-18 经过 2000 个样本训练后的人工神经网络在测试阶段的表现

### 4.3.3 PCA 与深度学习的对比分析

在 PCA 与深度学习模型的对比实验中，本文采用了五重交叉验证的方法进行对比分析。首先实验中构建一个拥有 4000 个特征样本的集合，这里面有两千个样本是 George 的人脸图像，另外两千个样本是 Jerry 的人脸图像。将每个人物的样本平均分成五份，每份 400 个。然后在每次实验的时候，从每个人物的样本中拿出一份作为测试集合，其余的一千六百份作为训练集合，以此形成五种不同的训练



集和测试集的组合，分别称为一到五样本集。并用这五种不同的组合分别作为 PCA 和深度学习模型的训练集和测试集，从而进行对比实验。

经过五重交叉验证的实验，本文得到表 4-7 所示的结果，表中的内容对应的是最终分类的精度。在表中，第二行表示采用深度学习模型，并将样本特征降维到 30 维后，分类器的识别精度；第三行表示采用 PCA 方法，同样也将样本特征降维到 30 维后的识别精度；第三行表示采用 PCA 方法，在保证 PCA 主分量的贡献率大于等于 0.95 的情况下，分类器的识别精度。

表 4-7 五重交叉验证的实验后得到的最终分类的精度

	1 样本集	2 样本集	3 样本集	4 样本集	5 样本集
深度模型	0.9225	0.94375	0.92125	0.9675	0.93
PCA(30)	0.8575	0.71625	0.52375	0.93125	0.515
PCA(0.95)	0.86875	0.62125	0.89875	0.855	0.89125

图 4-19 是基于表 4-7 而绘制出来的。图中的横坐标表示相应的由五重交叉验证形成的样本集，纵坐标表示最终分类后的精度。图中的被圆点标记的曲线表示通过深度学习模型降维后的识别效果；被方块标记的曲线表示采用 PCA 方法，将样本特征降维到 30 维的情况下，分类器的识别效果；被三角标记的曲线表示采用 PCA 方法，并保证主分量的贡献率大于等于 0.95 的情况下的识别效果。在图中本文发现，经过深度学习模型降维后的识别效果明显好于 PCA 的方法。

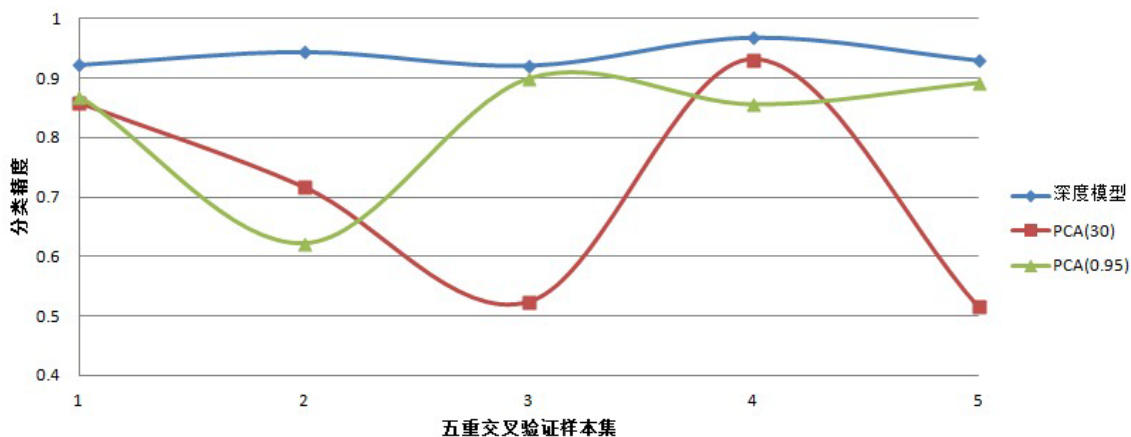


图 4-19 五重交叉验证的实验后得到的最终分类的精度

与此同时，本文还发现了一个现象就是，利用 PCA 对样本特征降维到 30 维后，在不同组合的样本集中，识别效果波动特别大，有的时候甚至比贡献率在 0.95 的 PCA 方法还要好。为了深入的分析这一现象，本文认真分析了各个样本集合的实验数据。

如表 4-8 所示，“贡献率”一行表示，通过使用 PCA 将样本特征降低到 30 维后，对应 30 个主分量的整体贡献率。“所降维度”一行表示，通过 PCA 降维，保

证主分量的整体贡献率大于等于 0.95 的情况下，要将原样本特征降低到的维度值。

表 4-8 对各个样本集合实验数据的分析

	1 样本集	2 样本集	3 样本集	4 样本集	5 样本集
贡献率	0.80973	0.80728	0.80711	0.79846	0.79784
所降维度	164	169	172	178	181

基于表 4-8 可以得出图 4-20 所示的趋势图，图中用圆点标记的曲线表示通过 PCA 算法，在保证 PCA 主分量的贡献率大于等于 0.95 的情况下，分类器的识别精度；用方块标记的曲线表示满足贡献率要求的前提下，不同样本集下样本特征所降低的维度，为了便于观察，这里将所有的维度乘以 0.005。通过这个趋势图，可以发现，在保证主分量贡献率大于等于 0.95 的前提下，不同样本集合所降低的维度波动不大，随着降低维度的提升精度总体呈现轻微上升的趋势。

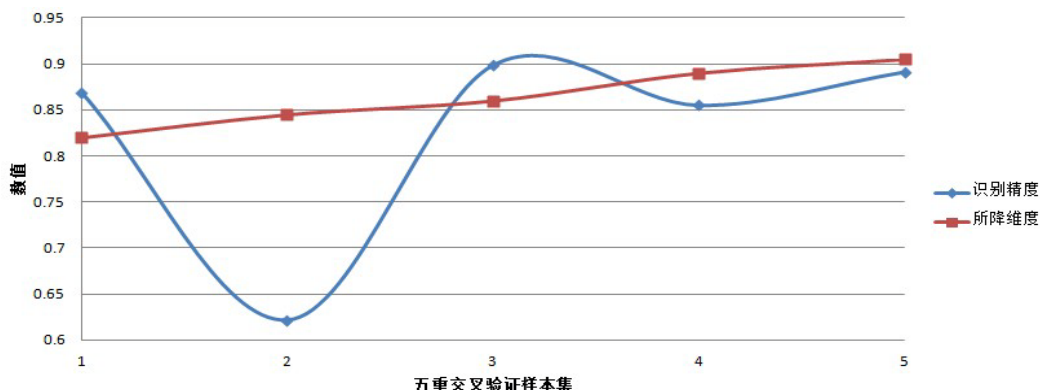


图 4-20 识别精度与所降维度的趋势图

图 4-21 也是基于表 4-8 绘制出来的，图中用圆点标记的曲线表示，通过 PCA 算法，在保证将样本特征降维到 30 维的情况下，分类器的识别精度；用方块标记的曲线表示在满足将原人脸样本特征降低到 30 维的前提下，不同样本集下 PCA 主分量的贡献率。通过这个趋势图，可以发现，在保证将原样本特征降低到 30 维的情况下，各个样本集主分量的贡献率波动不大；与此不同的是，经过降维之后，分类器的精度却波动非常大。经过认真的分析，本文得到结论，出现这种波动现象，包括在图 4-21 第二个样本集上的精度波动现象的原因是由于主分量的特性所引起的。主分量所提取的特征对不同样本集下的分类影响不确定，在一些样本集合，通过主分量降维后，得到了一个新的样本特征空间，在这个空间中，本文的样本集能够被很好的分成不同类，也就是说这样的降维比较适于对本文的分类问题；而在另一些样本集合中，在用主分量进行降维后所形成的新的样本特征空间中，样本集不能够被高效的进行分类，甚至可能比原始的特征空间更难做分类。因为存在这两种情况，所以本文的实验中，会存在精度波动的现象。在精度波动中，降低到 30 维的 PCA 方法明显波动的更厉害，这是由于 30 维所能代表的特征

空间比较小，对不同样本集合的适应能力比较差，因此波动比较频繁，而保证贡献率在 0.95 的 PCA 方法相对波动并不剧烈。

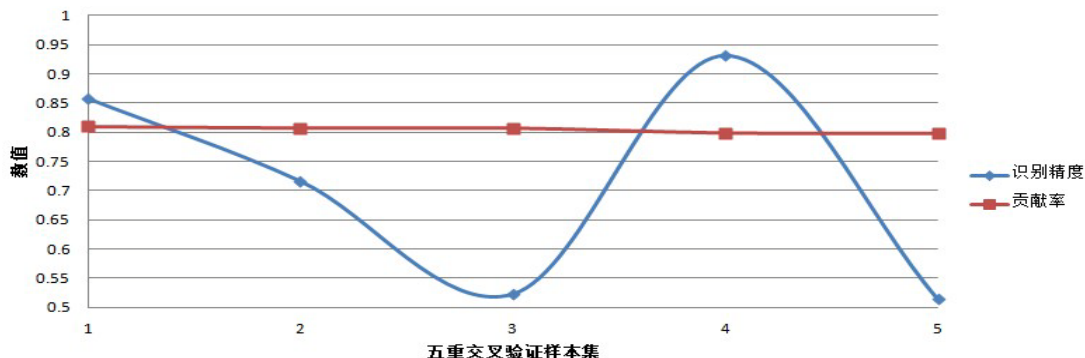


图 4-21 识别精度与贡献率的趋势图

通过PCA与深度学习模型的对比实验可以清楚的发现：首先，通过深度学习模型进行降维后，分类器的识别精度有提高，并且始终高于PCA的两种方法，这体现出了使用深度学习模型降维后用于分类上的优势；其次，考虑到PCA算法在分类问题上的波动性，深度学习模型则表现的比较稳定，这说明深度学习模型对不同样本集合的适应能力比较强，模型自身的稳定性比较高。

## 4.4 基于深度学习的BP识别算法的性能分析

基于深度学习的 BP 识别算法的性能分析实验主要从两方面展开：关于失衡训练集的实验分析和 BP 识别算法过拟合现象的实验分析。

### 4.4.1 失衡训练集对BP识别效果影响的实验与分析

为了完成这次实验，首先设计了五组不同组合的样本集合，这五组集合彼此之间样本的失衡比例不同，具体如表 4-9 所示：

表 4-9 不同组合的失衡样本集

样本集合名称	组合方式
g10+j10	1000 个 George 样本和 1000 个 Jerry 样本
g12+j8	1200 个 George 样本和 800 个 Jerry 样本
g14+j6	1400 个 George 样本和 600 个 Jerry 样本
g16+j4	1600 个 George 样本和 400 个 Jerry 样本
g18+j2	1800 个 George 样本和 200 个 Jerry 样本

1000:1000 数据集合在这里是有 1000 个 George 样本和 1000 个 Jerry 样本组成。图中的曲线正确分类时，前 500 个样本应该被分到 1 类的位置处，后 500 个样本应该被分到 2 类的位置处，出现的下降曲线就是用来描述被错分的情况。

图 4-22 是加入深度模型之后测试样本集合的表现情况，同样图 4-23 是没有加

入深度模型直接使用原始样本进行 BP 训练识别的效果。根据实验本文得出表 4-10 结果。

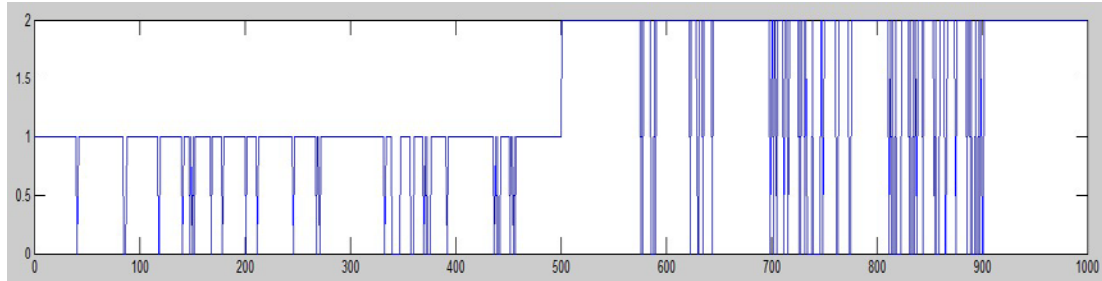


图 4-22 加入深度模型之后测试样本集合的表现情况

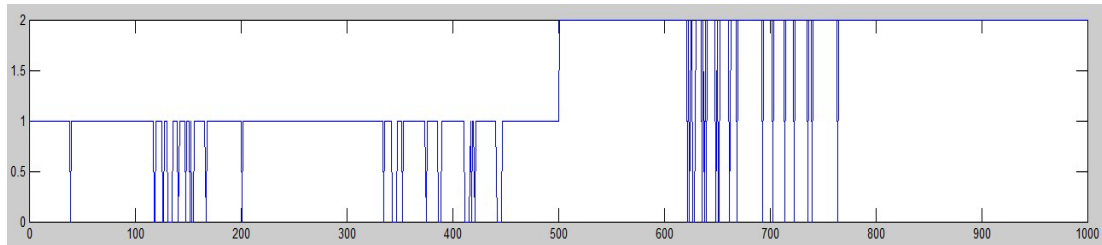


图 4-23 没有加入深度模型直接使用原始样本进行 BP 训练识别的效果

表 4-10 有无深度模型的测试样本集合的表现情况

	训练时间消耗	测试时间消耗	测试准确率
无深度模型	35.3034	0.021122	94%
有深度模型	2.1119	0.0074549	89%

1200:800 数据集合在这里是有 1200 个 Goerge 样本和 800 个 Jerry 样本组成。图 4-24 展示了在使用了深度模型的情况下采用此样本集合训练神经网络后的精度表现，其中实验的具体数据如表 4-11 所示。

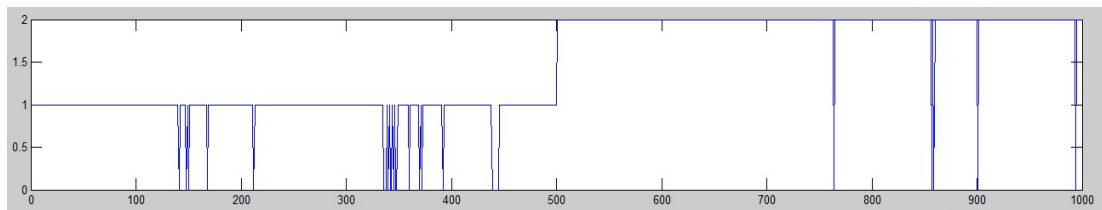


图 4-24 1200:800（1200 个 Goerge 样本和 800 个 Jerry 样本）数据集合的表现

表 4-11 1200:800（1200 个 Goerge 样本和 800 个 Jerry 样本）数据集合的表现

	训练时间消耗	测试时间消耗	测试准确率
无深度模型	22.6259	0.022016	98.6%
有深度模型	7.6541	0.0073757	97%

以下是 1400:600 数据集合。在这里有 1400 个 Goerge 样本和 600 个 Jerry 样本。

图 4-25 展示了使用了深度模型情况下 BP 识别算法的效果，从图中可以看出在此样本集中，BP 的识别效果有明显的提升。

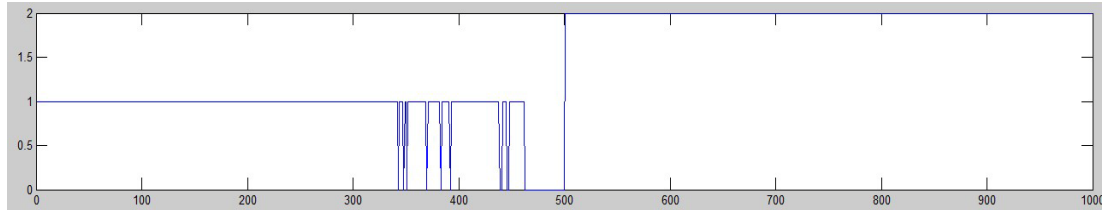


图 4-25 使用了深度模型情况下 BP 识别算法的效果

图 4-26 在没有使用深度学习展示了此样本集的表现。在这组实验中，出现了 BP 网络的训练时间过长的现象，这一现象反映出了，BP 网络在参数初始化的时候，如果用值不当，会导致训练过程很难收敛的问题。通过对 1400:600 样本集合的实验，可以得到表 4-12。

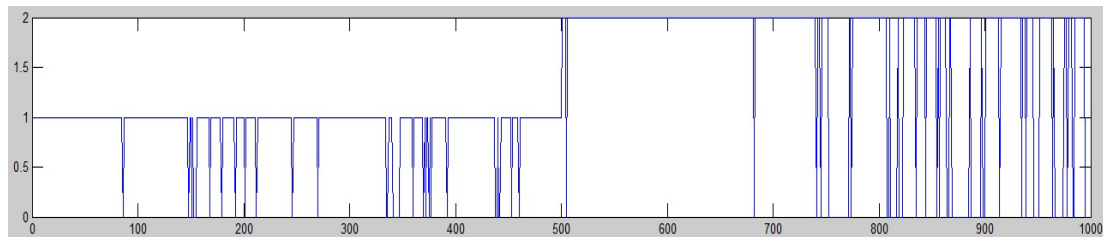


图 4-26 没有使用深度学习情况下 1400:600 样本集的效果

表 4-12 没有使用深度学习展示此样本的表现

	训练时间消耗	测试时间消耗	测试准确率
有深度模型	2.3454	0.0073178	95.1%
无深度模型(1)	468.435	0.019501	82.4%
无深度模型(2)	840.556	0.019901	87.6%
无深度模型(3)	8.542	0.11213	86.29%

整个实验过程采用了五类不同组合的样本集合，分别是 1000:1000、1200:800、1400:600、1600:400 和 1800:200，将这五类集合分别编号为 1 到 5 的五个自然数。通过实验本文得到了不同条件下样本集的精度，如表 4-13 所示。

表 4-13 失衡样本集下系统不同条件精度表

	1 样本集	2 样本集	3 样本集	4 样本集	5 样本集
有深度模型	0.89	0.986	0.951	0.984	0.883
无深度模型	0.94	0.97	0.864	0.938	0.818

基于表 4-13，可以达到图 4-27，图中描绘了不同样本集的精度变化趋势。通过图 4-27 可以发现，当数据训练集合的比例在 1:1 的时候，这个时候没有深度模型的 BP 网络比有深度学习的 BP 神经网络表现出色，但是随着训练样本集逐渐不平衡，有深度模型的 BP 网络表现的越来越好，而没有深度学习的 BP 网络则略微逊色，这说明，当训练样本集失衡的时候，采用有深度模型的 BP 网络能够得到更好的识别效果。这就说明了深度学习模型提取的是目标的本质特征，因此对失衡的样本集合依然能够进行有效分类。于此同时这对本文所设计的系统也有着重要的意义。因为在自动机运行过程中，每次迭代所生成的样本集合的组合一定不可能是完全平衡的，很可能出现失衡的情况，而失衡的样本集合对 BP 神经网络的识别又会产生消极的影响，这里通过使用深度模型可以有效的降低这种影响，从而提高了准确率。

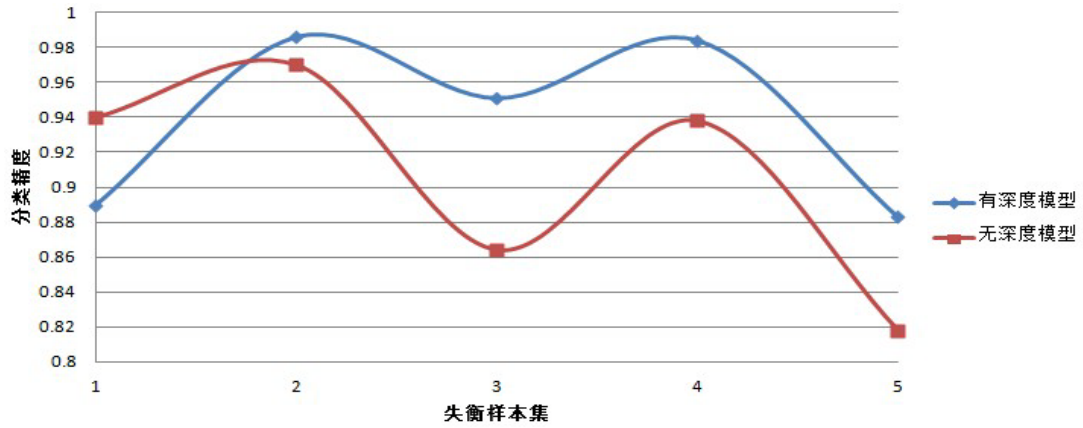


图 4-27 五类样本集的精度对比图

图 4-28 展示了不同样本集合，所产生的时间消耗，这个方面明显可以看出，使用深度学习模型可以有效的缩减训练和测试的时间，从而整体提高系统的实时性。

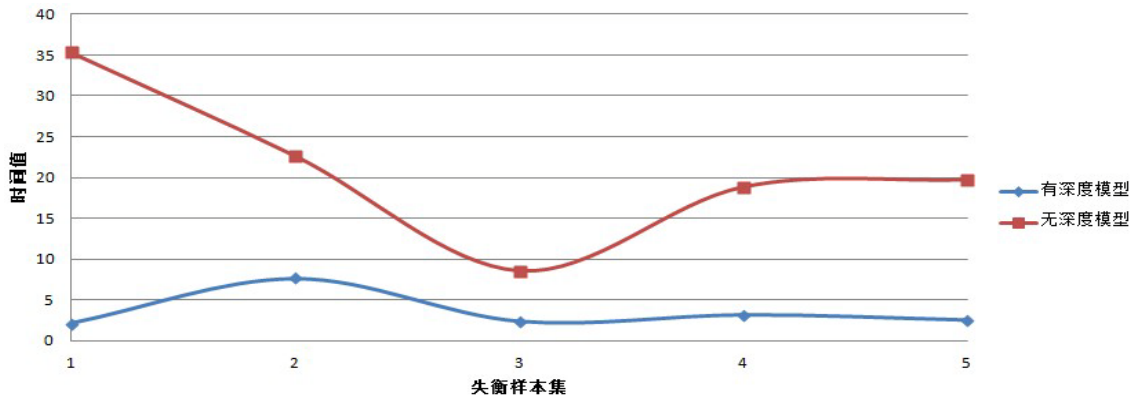


图 4-28 五类样本集的训练时间对比图



#### 4.4.2 BP识别算法过拟合现象的实验与分析

为了分析 BP 识别算法的过拟合现象,本文采用了五重交叉验证的方法进行实验分析。五重交叉验证实验的具体过程如下:分别取 2000 个 George 样本和 2000 个 Jerry 样本,将这两类的每一类平均分成五分,即每份 400 个样本,并分别编号为自然数一到五号分组。每次分别从 George 样本集和 Jerry 样本集抽取一份作为测试样本,每一类余下的 1600 个样本作为训练样本,以此方式构造出五组不同组合的样本集合,具体的组合参见表 4-14。

表 4-14 不同组合的五重交叉验证样本集

样本集合名称	组合方式
1_3200	缺乏两类样本一号分组的训练样本集
2_3200	缺乏两类样本二号分组的训练样本集
3_3200	缺乏两类样本三号分组的训练样本集
4_3200	缺乏两类样本四号分组的训练样本集
5_3200	缺乏两类样本五号分组的训练样本集
1_800	由两类样本一号分组组成的测试样本集
2_800	由两类样本二号分组组成的测试样本集
3_800	由两类样本三号分组组成的测试样本集
4_800	由两类样本四号分组组成的测试样本集
5_800	由两类样本五号分组组成的测试样本集

针对每一类进行五重交叉验证,五重交叉验证结果分别如图 4-29、图 4-30、图 4-31、图 4-32、图 4-33 所示。

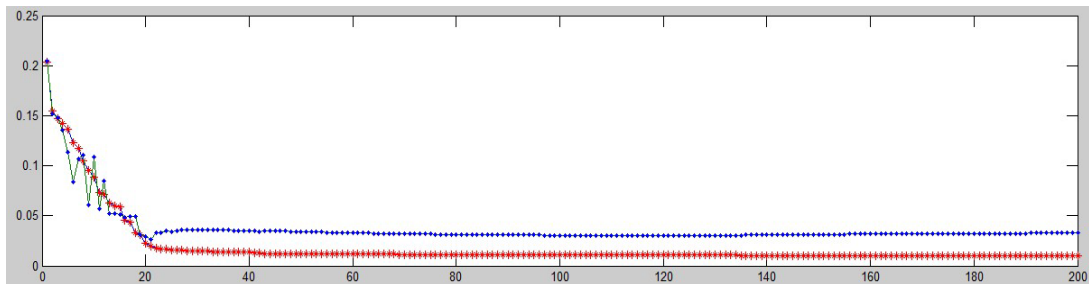


图 4-29 第一次交叉验证

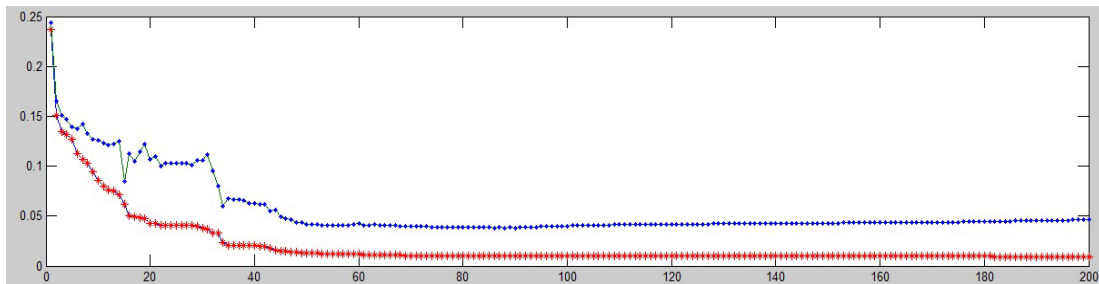


图 4-30 第二次交叉验证

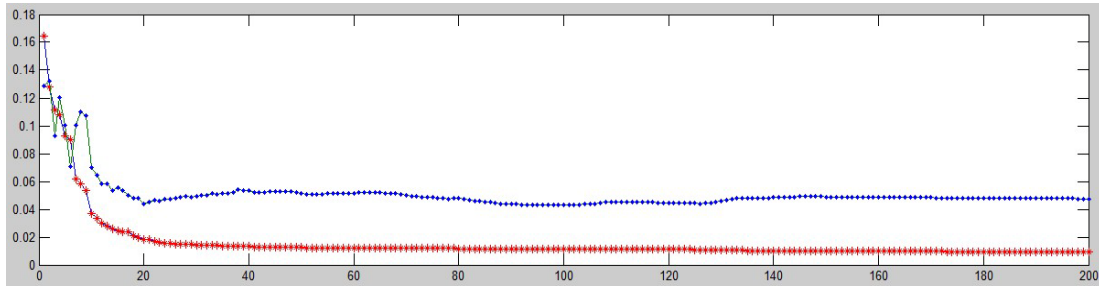


图 4-31 第三次交叉验证

本文发现在做第三次的五重交叉验证的时候，整个迭代过程测试误差从 40 次迭代开始到最后 200 次迭代结束，呈现略微的先下降后体上升的过程，经过分析，本文认为这是因为在训练阶段过分拟合导致的。

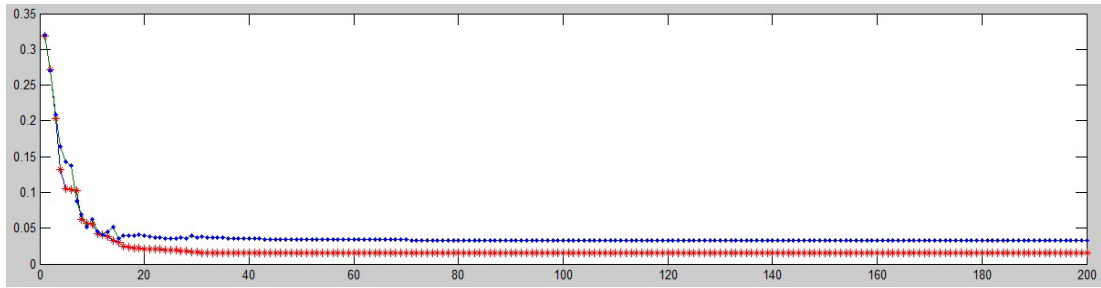


图 4-32 第四次交叉验证

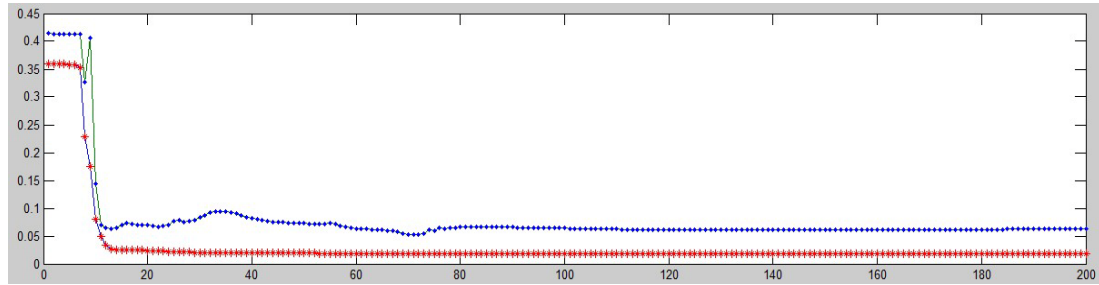


图 4-33 第五次交叉验证

实验中的五重交叉验证是在深度学习模型的基础上完成的，经过实验，通过五重交叉验证得到了如下趋势图。图中横坐标表示 BP 神经网络的迭代次数，纵坐标表示 BP 神经网络每次迭代的均方误差。

均方误差在这里是定义 BP 神经网络表现的一种度量指标，均方误差即为各个测量值误差的平方和的平均值的平方根，测量值的均方误差的计算方法如公式 (4-3) 所示。

$$\sigma = \sqrt{\frac{\varepsilon_1^2 + \varepsilon_2^2 + \varepsilon_3^2 + \dots + \varepsilon_n^2}{n}} = \sqrt{\frac{\sum \varepsilon_i^2}{n}} \quad (4-3)$$

式中  $\varepsilon_1$ 、 $\varepsilon_n$ .....  $\varepsilon_n$ —n 个测量值的误差；



$\sigma$ —测量值的均方误差。

通过五重交叉验证，发现随着迭代次数的不断增加，训练误差和测试误差会逐渐稳定在一个很小的数值范围内。体现出了误差逐渐收敛的特性。同时也发现所有的五重交叉验证到最后，都保持着训练误差低于测试误差的规律。

这次的五重交叉验证是在深度模型的基础上完成的，经过实验本文得出，深度模型基础上的 BP 神经网络表现良好，可以很好的对样本进行识别，同时在识别的效率上也大幅提高。

## 4.5 本章小结

本章主要围绕着四个主要内容展开分别是：深度学习模型的训练、深度学习模型的构造和选取、PCA 算法与深度学习的对比，基于深度学习的 BP 识别算法的性能分析。针对这四个内容进行了多种实验，通过这些实验不但证明了深度学习模型的一些特性，同时有效的证明了深度学习算法在视频人脸识别方法中的重要应用价值。

## 第5章 视频人脸检测识别系统

视频人脸检测识别系统可以根据所提供的视频，剧本和字幕来生成各个人物的样本集合，这个样本集合可以作为之后深度学习的训练样本，进而成为识别部分的训练样本；识别部分训练完成之后，会为系统所生成的人脸样本进行分类，有效的区分出视频人脸中的各个角色。这个系统由四部分构成：字幕剧本融合模块，人脸检测模块，样本集自动生成模块，基于深度学习的人脸识别模块，系统架构如图 5-1 所示。

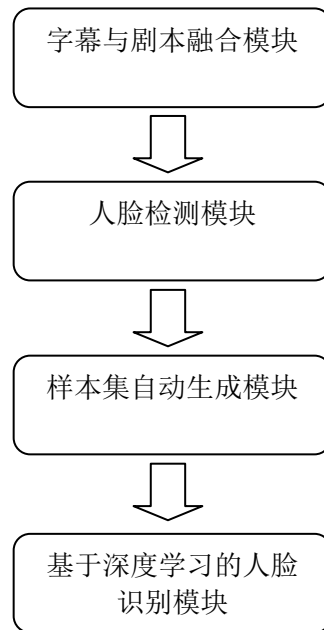


图 5-1 视频人脸检测识别系统架构图

(1)字幕剧本融合模块。视频所具有的一大特点就是，拥有自己的剧本和字幕信息，这些信息在很多识别系统中被忽略了，这里利用了剧本和字幕信息。通过将字幕和剧本信息融合等方法<sup>[39]</sup>，形成本文检测识别过程中所需要的融合文件，此文件为之后的说话者标注与人脸识别提供了重要的依据。

(2)人脸检测模块。基于视频的人脸检测意味着人脸可能存在于十分复杂的背景环境中，人脸的姿势可能存在着很大的变化，因此在如此多变的环境中保证检测的可靠性至关重要。与此同时在视频人脸的检测中，还要保证一个重要特性就是实时性，为了满足这个要求，模块采用 Adaboost 的方法做检测。这样在人脸检测过程中既能考虑到检测人脸的正确性，同时也能保证检测的实时性。通过 Adaboost 算法，以及肤色模型和唇色模型的两层过滤机制实现人脸的准确检测。

(3) 样本集自动生成模块。结合在字幕剧本融合过程中所形成的视频时间区间，分割样本集合，将各个单一人物样本集合合并，形成可供深度学习使用的大型样本集合。这里的人脸识别主要侧重于人脸的辨认，它是一对多进行图像匹配比对的过程，回答的是：你是谁的问题。本文这里采用的是深度学习的方法。利用 DBN 为人脸进行建模。因为采用的是深度学习算法，首先通过设立多个隐藏层构建深度结构，然后在每层中采用非线性算子，以此来提高每个层次的表示能力。最后开始从最底层开始逐层为每层建模，先构建简单的概念，再随着层次的提升逐渐构建更难的概念。

(4) 基于深度学习的人脸识别模块。该模块为本系统的主要部分，它是基于深度学习算法采用 BP 神经网络的方法对视频人脸进行识别，这一模块的核心就是深度学习系统的实现，而 BP 神经网络的识别是建立在深度学习模型的高度抽象表示的基础之上的。通过这一模块的识别，最终实现了对视频人脸的标注<sup>[40]</sup>。

## 5.1 人脸检测模块

人脸检测模块采用了两种过滤技术，通过两层过滤来保证人脸识别的准确性。基于两层过滤的人脸检测系统整体框架如图 5-2 所示。

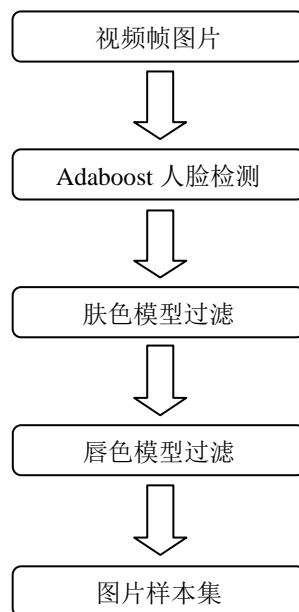


图 5-2 基于两层过滤的人脸检测系统整体框架

人脸过滤检测模块的主要功能是：通过过滤技术将视频帧图片中的人脸准确的检测出来，此模块实现的关键是要保证对人脸检测的准确性和速度性，从人脸检测模型中可以看出这一模块主要分为三个部分：Adaboost 人脸检测、肤色模型人脸过滤和唇色模型人脸过滤。

### 5.1.1 肤色模型人脸过滤

经过 Adaboost 算法粗提取，得到了一个人脸图片集合，在这个集合中并非都是人脸图片，还存在非人脸的错误图片，需要进行进一步的检测和过滤，从而剔除掉这部分错误的图片。这里采用了肤色模型<sup>[41,42]</sup>，首先统计出人脸肤色的阈值特征，进而建立一个肤色模型，最终利用这个肤色模型对人脸图片进行基于像素点的数值分析，将不符合要求的图片过滤掉。

以《宋飞传》视频文件作为检测模块的输入部分，图5-3中展示的是在Adaboost人脸检测基础上使用肤色模型和未使用肤色模型检测模块输出的对比图片（左边是Adaboost算法检测出的人脸，右边是加入肤色模型后检测出的人脸）。通过对比可以发现，使用肤色模型之后，可以有效的过滤掉Adaboost算法所检测出的错误人脸矩形。



图 5-3 使用肤色模型和未使用肤色模型的对比图片

### 5.1.2 唇色模型人脸过滤

肤色模型虽然过滤掉了人脸图片集合中大部分的错误人脸图片，但是经过试验发现，在肤色模型过滤过程中，对于一些和人脸颜色十分近似的物体的过滤效果不是很好，比如，黄色地板和肉色的衣服等。为了克服这一问题，系统引入了唇色模型<sup>[43]</sup>。为了实现唇色模型对错误人脸的过滤，首先进行了嘴部区域提取，这种提取采用的方法是利用嘴部区域在人脸中的几何特征，按照数值比例得到嘴部区域。同时统计人脸中唇色的阈值特征，从而建立唇色模型，最终利用这个唇色模型对经过肤色模型过滤后的人脸图片集合进行数值比对，将那些蕴含在人脸图片集合中的杂质过滤掉。

以《宋飞传》视频文件作为检测模块的输入部分，图5-4中展示的是在使用了Adaboost算法和肤色模型过滤的基础上，使用唇色模型和未使用唇色模型的对比图片（左边是没有使用唇色模型进行过滤，右边是使用了唇色模型进行过滤）。在这组图片中，地板的颜色和人脸的颜色很相近，因此在肤色模型过滤的时候，不能够将这个的错误人脸矩形过滤掉，这时采用唇色模型，可以有效的过滤掉错误的

人脸矩形。



图 5-4 在肤色模型基础上使用唇色模型和未使用唇色模型的对比图片

## 5.2 样本集自动生成模块

样本集自动生成模块的主要任务是自动生成样本集，为了能够达到这样的目的，这一模块必须建立在字幕剧本融合模块和人脸检测模块基础之上。这个模块主要包含两个部分：数据采集和数据预处理。

### 5.2.1 数据采集

在人脸过滤检测完成之后，要将大量的人脸图片流进行分类采集。利用字幕与剧本融合技术，形成了一个融合字幕文件，这个融合字幕文件不但有一段对话的内容和持续时间，还包括了对话的说话者姓名。以融合字幕文件中的时间区间为切割单元，对图片流进行切割。这里的每个时间区间代表一段对话的持续时间，这里被称之为“yes 区间”，两个相邻的“yes 区间”不一定是连续的，如：“yes1 区间”的持续时间 1s-8s，“yes2 区间”的持续时间 10s-15s，这里称两个相邻“yes 区间”的空白为“no 区间”。图片流的切割就是以“yes 区间”和“no 区间”为单位进行的，“yes 区间”被切割的图片集合对应说话者，“no 区间”没有说话者对应。最终采集到的数据是有人名对应的各个独立的人脸图片子集。如图 5-5 所示，描述了基于“yes 区间”和“no 区间”的镜头分割。

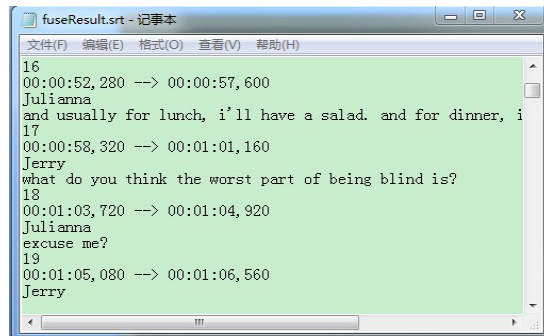


图 5-5 基于“yes 区间”和“no 区间”的镜头分割

## 5.2.2 数据预处理

经历数据采集阶段，将数据分割成了“yes”区间和“no”区间，图 5-6 描述了分割之后形成的各个样本集。

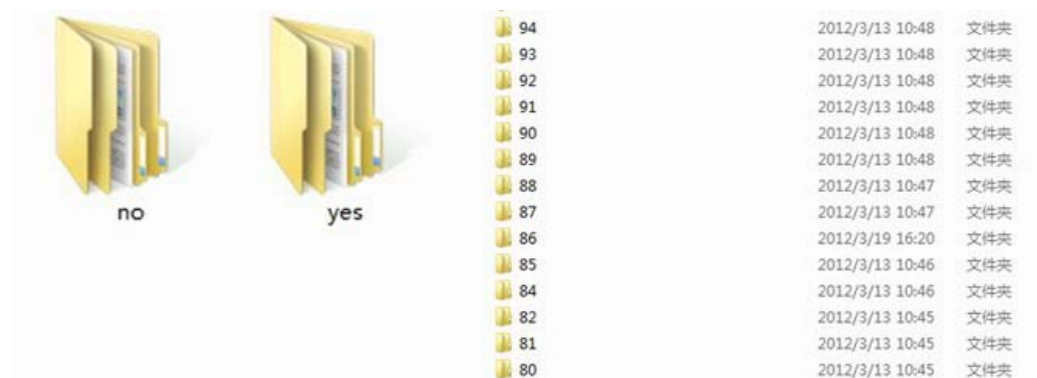


图 5-6 分割之后形成的各个样本集

所有收集到的数据是为深度学习算法而准备的。因此对于收集到的数据要进行预处理，以适应深度学习算法的要求。首先对采集的图片数据进行规格化，这里规格化成 28\*28 的 jpg 图片格式。在数据采集的过程中，数据是以“yes 区间”为单位被收集起来的，每个不同的“yes 区间”它所对应的图片集合可能属于不同的说话者，在预处理阶段，要把所有属于一个说话者的“yes 区间”所对应的集合合并，形成以说话者为单位的新的图片集合。图 5-7 即为经过了数据预处理之后，形成的新样本子集。



图 5-7 数据预处理之后形成的新样本子集

## 5.3 说话者识别模块

通过使用深度学习模型，并与上层识别算法相配合，在系统二次迭代时作为一个独立的过滤层，将含有单一说话者文件夹中的杂质过滤掉，将含有多个说话者文件夹中的不同人脸进行识别和分类。从而形成一个新的样本集。这个样本集所含有的杂质比第一次迭代的时候要少，所含有的样本数量比第一次迭代的时候

要多。同时利用说话者识别模块来对多人脸的帧文件进行人物区分与识别，最终完成对视频文件所有人物的检测与识别工作。

## 5.4 本章小结

本章主要是从整体上介绍了视频人脸检测识别系统的结构。这一系统主要由四个模块构成：字幕剧本融合模块、基于两层过滤的人脸检测模块、针对指定数据格式的样本集自动生成模块和基于深度学习模型的人脸识别模块。其中人脸检测模块在使用了Adaboost算法的基础上使用了肤色模型和唇色模型两层过滤技术。样本集自动生成模块的任务是自动生成样本集，为了能够生成样本集合，此模块必须建立在字幕剧本融合模块和人脸检测模块基础之上，这个模块主要拥有两个功能：数据采集和数据预处理。基于深度学习的人脸识别模块，采用的是深度学习算法，基于此算法通过BP神经网络来对视频文件中的人脸进行过滤和识别。

## 结论

深度学习算法目前是识别和特征提取方面的热点，本文系统的分析了深度学习算法的理论基础和系统实现，并将深度学习算法应用到了视频人脸检测与识别系统之中。本文的主要贡献为以下几点：

(1) 在识别模块中，本文使用了深度学习模型来进行视频人脸样本的特征提取，这种方法相对于 PCA 方法来说，能够提取出人脸样本更本质的特征，从而得到更高的识别精度，同时对不同视频人脸样本集合有着较好的适应能力。

(2) 将视频人物识别与深度学习模型相结合，一方面，提高了视频人脸识别的准确性和稳定性，同时另一方面，视频人脸检测后所形成的大规模样本集合，也为深度学习模型的训练提供了丰富的资源。两者实现了有机的结合和优势的互补。

(3) 在人脸检测部分本文在采用了 AdaBoost 算法的基础上，为人脸建立了肤色模型和唇色模型，使用这两个模型作为过滤模块，使视频人脸样本集合中合格人脸的数量得到了显著提升。

虽然本文在视频人脸检测识别以及深度学习方面取得了一定程度的成绩，但是在某些方面，本文还存在一些缺陷和不足：

(1) 系统整体的代码优化还不够彻底，因此需要进行一次全面的、具体的代码优化过程，以此来使系统更加高效。

(2) 由不同样本组合所构成的深度学习模型对最终的分类有着不同的影响，这些深度模型彼此之间的深层关系还不明确，因此需要进行一个更大规模的深度学习模型的实验，以此来找到不同深度模型之间的深层关系。

将来可以通过不断优化代码，来逐渐的缩短系统的运行时间，通过找出不同深度模型之间的关联来建立一个更加有效的深度学习模型。



## 参考文献

- [1] Lijing Zhang, Yingli Liang. A fast method of face detection in video images[C]. The 2nd IEEE International Conference on Advanced Computer Control (ICACC 2010), 2010, 4: 490-494.
- [2] Iago Landesa-Vázquez, José Luis Alba-Castro. The Role of Polarity in Haar-like Features for Face Detection[C]. 20th International Conference on Pattern Recognition(ICPR 2010), 2010: 412-415.
- [3] Yan Y, Zhang Y J. State-of-the-art on video-based face recognition[M]. Encyclopedia of Artificial Intelligence, 2008: 1455-1461.
- [4] Md. Atiqur Rahman Ahad, Takehito Ogata, Joo Kooi Tan, Hyoungseop Kim. Motion recognition approach to solve overwriting in complex actions[C]. Proceedings of the International Conference on Automatic Face and Gesture Recognition (FGR' 08), 2008, 9: 1-6.
- [5] Liying Lang and Weiwei Gu. Study of Face Detection Algorithm for Real-time Face Detection System[C]. Recent Advances in Electronic Commerce and Security (ISECS 2009), 2009, 2: 129-132.
- [6] Z. Kalal, K. Mikolajczyk, and J. Matas. Face-tld: Tracking-learning-detection applied to faces[C]. 2010 International Conference on Image Processing (ICIP 2010), 2010: 3789-3792.
- [7] Jiquan Ngiam, Zhenghao Chen, and Pang Wei Koh. Learning Deep Energy Models[C]. In proceedings of the 28th International Conference on Machine Learning, 2011.
- [8] H. Larochelle, D. Erhan, A. Courville. An empirical evaluation of deep architectures on problems with many factors of variation[C]. In Proceedings of the Twenty-fourth International Conference on Machine Learning (ICML'07), 2007: 473-480.
- [9] H. Lee, R. Grosse, R. Ranganath. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations[C]. In Proceedings of the Twenty-sixth International Conference on Machine Learning (ICML'09), 2009.
- [10] M. Ranzato, C. Poultney, S. Chopra. Efficient learning of sparse representations with an energy-based model[C]. In Advances in Neural Information Processing Systems 19 (NIPS'06), 2007: 1137-1144.

- [11]R. Salakhutdinov, G. E. Hinton. Learning a nonlinear embedding by preserving class neighborhood structure[C]. In Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics (AISTATS'07), 2007.
- [12]G. Taylor, G. E. Hinton, S. Roweis. Modeling human motion using binary latent variables[C]. In Advances in Neural Information Processing Systems 19 (NIPS'06), 2007: 1345–1352.
- [13]I. Levner. Data Driven Object Segmentation[D]. PhD thesis of Department of Computer Science of University of Alberta, 2008.
- [14]M. Ranzato, M. Szummer. Semi-supervised learning of compact document representations with deep networks[C]. In Proceedings of the Twenty-fifth International Conference on Machine Learning (ICML'08), 2008, 307: 792–799.
- [15]R. Collobert, J. Weston. A unified architecture for natural language processing: Deep neural networks with multitask learning[C]. In Proceedings of the Twenty-fifth International Conference on Machine Learning (ICML'08), 2008: 160–167.
- [16]周龙. 基于朴素贝叶斯的分类方法研究[D]. 安徽大学硕士学位论文. 2006.
- [17]P. Viola, M. Jones. Robust Real-Time Face Detection[J]. Int'l J. Computer Vision, 2004, 57(2): 137-154.
- [18]梁路宏, 艾海舟, 何克忠. 基于多模板匹配的单人脸检测[J]. 中国图像图形学报, 1999, 4(10): 825-830.
- [19]梁路宏, 艾海舟, 肖习攀. 基于模板匹配与支持矢量机的人脸检测[J]. 计算机学报, 2002, 25(1): 22-29.
- [20]L.L. Huang, A. Shimizu, Y. Hagihara. Face detection from cluttered images using a polynomial neural network[J]. Neural computing, 2003, 52: 197-211.
- [21]Adam Coates, Honglak Lee, Andrew Ng. An analysis of single-layer networks in unsupervised feature learning[C]. In Advances in Neural Information Processing Systems, 2010.
- [22]Hinton, G. E., Teh, Y. A fast learning algorithm for deep belief nets[J]. Neural Computation 18, 2006: 1527-1554.
- [23]Yoshua Bengio, Pascal Lamblin, Dan Popovici. Greedy Layer-Wise Training of Deep Networks[C]. Advances in Neural Information Processing Systems 19 (NIPS 2006), 2007: 153-160.
- [24]Marc'Aurelio Ranzato, Christopher Poultney, Sumit Chopr. Efficient Learning of Sparse Representations with an Energy-Based Model[C]. Advances in Neural Information Processing Systems (NIPS 2006), 2007: 792-799.

- [25] I. Landesa-Vázquez and J. L. Alba-Castro. The Role of Polarity in Haar-like Features for Face Detection[C]. Accepted for presentation in 20th International Conference on Pattern Recognition(ICPR 2010), 2010: 412-415.
- [26] Ning Jiang, Wenxin Yu, Shaopeng Tang. Cascade Detector for Rapid Face Detection[C]. 2011 IEEE 7th International Colloquium on Signal Processing and its Applications, 2011: 155-158.
- [27] C. Shen, Jana Zhang. Efficiently Learning a Detection Cascade With Sparse Eigenvectors[C]. IEEE transactions on image processing, 2011, 20(1): 22-35.
- [28] Y. Bengio, P. Lamblin, D. Popovici. Greedy layer-wise training of deep networks[C]. In Advances in Neural Information Processing Systems 19 (NIPS'06), 2007: 153–160.
- [29] Q. V. Le, J. Ngiam, A. Coates. On optimization methods for deep learning[C]. The 28th International Conference on Machine Learning(ICML 2011), 2011.
- [30] J. Ngiam, A. Khosla, M. Kim. Multimodal deep learning[C]. In NIPS Workshop on Deep Learning and Unsupervised Feature Learning, 2010.
- [31] G. Taylor, G. Hinton. Factored conditional restricted Boltzmann machines for modeling motion style[C]. In Proceedings of the 26th International Conference on Machine Learning (ICML'09), 2009: 1025–1032.
- [32] I. Sutskever, G. Hinton, G. Taylor. The recurrent temporal restricted Boltzmann machine[C]. Twenty-Fourth Annual Conference on Neural Information Processing Systems(NIPS 2009), 2009.
- [33] R. Salakhutdinov, G. E. Hinton. Deep Boltzmann Machines[C]. Twelfth International Conference on Artificial Intelligence and Statistics(AISTATS 2009), 2009.
- [34] Salakhutdinov, R., Mnih, A. Hinton, G. Restricted Boltzman Machines for Collaborative Filtering[C]. In Proceedings of the 24th International Conference on Machine Learning, 2007.
- [35] D. Yu, L. Deng, G. Dahl. Roles of pre-training and fine-tuning in context-dependent DBN-HMMs for real-world speech recognition[C]. In Proc. NIPS 2010 Workshop Deep Learn. Unsupervised Feature Learn., 2010.
- [36] Hinton, G. E., B. J. & Neal. The “wake-sleep” algorithm for unsupervised neural networks[J]. Science. 1996(268)1158-1161.
- [37] P. Shih, C. Liu. Face detection using discriminating feature analysis and support vector machine[J]. Pattern Recognition, 2006, 39 (11): 260–276.
- [38] G. E. Hinton, R. Salakhutdinov. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504–507.

- [39]文翰, 黄国顺. 语音识别中 DTW 算法改进研究[J]. 微计算机信息, 2010, 26(7-1): 195-197.
- [40]Y. Wu, W. Hu, T. Wang. Robust speaking face identification for video analysis[C]. In Proc. Pacific Rim Conf., 2007: 665-674.
- [41]Alajel, K.M., Xiang, W.. Face detection based on skin color modeling and modified Hausdorff distance[C]. In Proc. of IEEE Int. Conf. on Consumer Communications and Networking, 2011: 399-404.
- [42]S.L. Phung, A. Bouzerdoun, D. Chai. Skin segmentation using color pixel classification: analysis and comparison[C]. IEEE Trans. Pattern Anal. Mach. Intell., 2005: 27 (1): 148-154.
- [43]Padma Polash Paul, Marina Gavrilova. PCA Based Geometric Modeling for Automatic Face Detection[C]. 2011 International Conference on Computational Science and Its Applications, 2011: 33-38.

## 攻读硕士学位期间发表的论文及其它成果

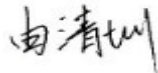
- [1] Yuxin Ding , Bin Zhao, Qingzhen You, Guangren Chai. Object Retrival Based on Visual Word Pairs. 2012 IEEE International Conference on Image Processing(ICIP'2012), 2012.
- [2] Di Zhou, Yuxing Ding, Qingzhen You, Min Xiao. Learning to Rank Documents Using Similarity Information between Objects. ICONIP, 2011.

## 哈尔滨工业大学学位论文原创性声明及使用授权说明

### 学位论文原创性声明

本人郑重声明：此处所提交的学位论文《基于深度学习的视频人脸识别方法》，是本人在导师指导下，在哈尔滨工业大学攻读学位期间独立进行研究工作所取得的成果。据本人所知，论文中除已注明部分外不包含他人已发表或撰写过的研究成果。对本文的研究工作做出重要贡献的个人和集体，均已在文中以明确方式注明。本声明的法律结果将完全由本人承担。

作者签名：



日期： 2013 年 1 月 4 日

### 学位论文使用授权说明

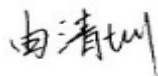
本人完全了解哈尔滨工业大学关于保存、使用学位论文的规定，即：

(1) 已获学位的研究生必须按学校规定提交学位论文；(2) 学校可以采用影印、缩印或其他复制手段保存研究生上交的学位论文；(3) 为教学和科研目的，学校可以将学位论文作为资料在图书馆及校园网上提供目录检索与阅览服务；(4) 根据相关要求，向国家图书馆报送学位论文。

保密论文在解密后遵守此规定。

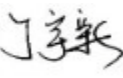
本人保证遵守上述规定。

作者签名：



日期： 2013 年 1 月 4 日

导师签名：



日期： 2013 年 1 月 4 日

## 致 谢

一转眼在哈工大深圳校区两年半的研究生学习生活就要结束了，我的论文已经完成，这篇论文不仅仅饱含着我的辛勤与汗水，同时也蕴含着我的父母、我的导师和我的同学对我的帮助与关心。

首先我要感谢丁宇新导师，丁老师经常主动来到我的桌前，为我指导论文，给我的论文提出了许多的宝贵意见，可以说没有丁宇新老师的细心指导就没有这篇论文的诞生，丁老师对我的恩情我将永生难忘。

与此同时，在我撰写论文的过程中，爸爸时常打来电话问我的进展情况，并在思考方式上启迪我，让我跳出自己固化思想，去更全面、更系统的分析实验。妈妈也经常叮嘱我，在写论文的时候要注意身体，保持自己健康的体魄。

当然我也要感谢我身边的同学们，无论是在平时的科研期间，还是在写论文的时候，他们都给了许多的帮助和关心，使我感受到了团队和集体的力量。

最后，我要感谢我的母校哈工大深圳研究生院，因为您的关怀，我度过了两年半的快乐幸福生活，谢谢你。