

# 杭州电子科技大学

## 硕士学位论文

题目：面向视觉识别的感知计算模型研究

研究生 王 强

专 业 控制工程

指导教师 范 影 乐 教授

完成日期 2018 年 3 月

杭州电子科技大学硕士学位论文

面向视觉识别的感知计算模型研究

研 究 生：王 强

指导教师：范 影 乐 教 授

2018 年 3 月

Dissertation Submitted to Hangzhou Dianzi University  
for the Degree of Master

# **Research on Visual Recognition Based Perceptual Computing Model**

**Candidate: Wang Qiang**

**Supervisor: Prof. Fan Yingle**

**March , 2018**

# 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各单位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 日

## 摘要

随着高清成像和并行运算硬件能力的提高，基于视觉海量数据的分析和理解技术已经成为模式识别与人工智能领域的研究热点之一。传统视觉识别算法一定程度上也考虑了生物感知计算的原理，例如利用训练海量的样本来动态调整神经网络模型的结构和参数，从而实现出所谓的最优决策。但是它们更多地只是利用了生物感知的一些基本特性，整体上仍是将其作为黑盒来进行模拟研究。事实上生物感知系统所具有丰富的视觉机制，是实现视觉理解和识别的基础。因此本文关注于生物感知系统中的视觉信息流处理机制，试图构建面向视觉识别的感知计算新模型，并以视觉识别中常见的面孔验证和识别为例展开应用研究。本文主要从视觉信息流在视通路上的多路径传递、倒立面孔与正立面孔的独立识别机制，以及基于局部和全局表征信息的多尺度面孔识别展开研究，具体研究内容及成果包括：

(1) 提出了一种基于视觉信息流双视通路的面孔验证模型。构建了底层卷积网络，模拟生物感知系统对视觉信息流的过滤；提出了双视觉通路概念，构建监督信号和识别信号融合的增强网络，实现对视觉识别特征的强化；最后，通过比较证件照片和多姿态面孔照片的特征分布情况，利用联合贝叶斯人脸的方法输出面孔身份的异同。实验结果表明，双视觉通路的设计能够提高面孔验证的准确率，说明了监督信号对面孔特征提取有约束作用。

(2) 提出了一种基于左右半球协作机制的正立和倒立面孔识别新方法。模拟视觉信息流在视通路中的传递和处理过程，首先构建底层神经网络，建立敏感纹理特征以及对称卷积核的机制，实现正立和倒立面孔图像的去除冗余和预处理；接着提出一种基于局部区域提取的池化神经网络层的概念，构建多局部特征融合的网络结构，实现局部信息的压缩提取和融合；最后根据高级视觉皮层中左右半脑协作的特点，提出一种融合整体和局部信息的预测函数。实验结果表明，本文方法能够实现正立和倒立面孔的识别，说明了整体和局部特征的两条视觉通路虽然分别在正立和倒立面孔识别上起了决定性的作用，但它们并不是孤立存在的，两条通路所刻画的面孔信息应该是一种互补式的关系。

(3) 提出了一种基于多层次信息表达的多尺度面孔识别新方法。模拟视网膜和神经节细胞对底层纹理特征的敏感特点，构建多层卷积级联的网络，以达到视觉信息过滤和提炼的目的；考虑到底层信息流对面孔特征表达的裨益作用，构建了基于局部和整体结合的多尺度池化的概念，输出固定大小的特征通道和特征图谱；根据视觉皮层的决策机制，各层次的特征信息融入到全连接层，利用神经元的激活特性判断面孔身份。实验结果表明，本方法可以有效应对多尺度面孔图像识别，从而说明底层纹理特征将有助面孔特征的表达。

**关键词：**视觉感知机制，视觉信息，面孔识别，卷积神经网络

## ABSTRACT

With the high resolution imaging and the hardware ability of parallel computing, the analysis and comprehension technology based on massive visual data has become one of the research focuses in the pattern recognition and artificial intelligence. To some extent, the tradition visual recognition algorithms consider the principle of the biological perception as well, such as using the massive training sample data to modify the structure and parameters for neural network dynamically and realizing so-called optimal decision. But they often only use some basic characters of biological perception and simulate it as a black box overall. Actually the abundant visual mechanisms in biological perception system are the bases to realize the visual comprehension and recognition. Therefore, this thesis focuses on the visual information flow processing mechanism in the biological perception system, tries to establish a new perception computing model which faces to the visual recognition, and takes the common face verification and recognition in visual recognition as an example to carry out the study. The contents such as the multi pathway delivering in the visual path for visual information flow, independent recognition mechanism for upright and upside-down faces and multi-scale face recognition are studied in this thesis. The details and results are as follows:

(1) A visual information flow dual vision pathways based face verification model is proposed. A convolution network at the bottom is established to simulate the filtering of visual information flow by the biological perception system; the concept of dual vision pathways is proposed to construct an enhanced network of supervisory signal and recognition signal fusion and to realize the enhancement of visual recognition features; finally by comparing the identification photo and multi-faceted photos of the spatial distribution, the joint Bayesian face recognition method is used to identify the similarities and differences of faces. The experimental results show that the design of dual vision pathways can improve the accuracy of face verification and the supervisory signal can restraint the face feature extraction.

(2) A new upright and upside-down faces recognition method based on the cooperation mechanism of left and right hemispheres is proposed. To simulate visual information flow transmission and processing in the visual pathway, the bottom neural network and the sensitive texture features and symmetric convolution kernel mechanism are built and the upright and upside-down faces image removal redundancy and preprocessing are achieved; then a concept of pooling neural network layer extracting from local regions is proposed and a multi-local feature

fusion network structure is constructed to compress and extract local information. Finally, according to the characteristics of the left and right hemispheres in the high-level visual cortex, a prediction function which integrates global and local information is proposed. The experimental results show that the proposed method can identify both upright and upside-down faces. The results also indicate that although they play a decisive role in recognizing orthographic and inverted faces, the two visual pathways of global and local features are not isolated but complementary.

(3) A new multi-scale face recognition method based on multi-level information expression is proposed. First of all, the sensory characteristics of retina and ganglion cells to the underlying texture features are simulated and a multi-layer neural network with convolutional pooling structure cascade is constructed to filter and refine the visual information. Secondly, considering the benefit in the bottom information flow expressing the face character, the concept of multi-scale pooling based on local and global integration is constructed and the fixed-size feature channel and feature map is output. Then, according to the decision mechanism of visual cortex, the feature information of each level is integrated into the fully connected layer and the activation characteristics of neurons are used to determine face identity. The experimental results show that this method can be effectively applied to the recognition of multi-scale face images and the results also indicate that the underlying texture features is helpful in the face features expression.

**Keywords:** visual perception mechanism, visual information, face recognition, convolutional neural network

## 目录

摘要.....	I
ABSTRACT.....	II
目录.....	1
第 1 章 绪论.....	3
1.1 研究目的与意义.....	3
1.2 国内外研究现状.....	4
1.3 本文主要工作及创新点.....	5
1.4 论文的组织结构.....	6
第 2 章 视通路以及卷积神经网络.....	8
2.1 视通路概述.....	8
2.1.1 视网膜.....	8
2.1.2 膝状体与初级视觉皮层.....	8
2.1.3 左右半球系统.....	9
2.2 神经元细胞.....	10
2.2.1 局部连接.....	10
2.2.2 激活函数.....	11
2.2.3 池化采样.....	12
2.2.4 Softmax 分类器.....	12
2.3 防止网络过拟合的方法.....	13
2.3.1 Dropout 层.....	13
2.3.2 BN 层.....	14
2.3.3 正则化.....	15
2.4 小结.....	15
第 3 章 基于视觉信息流的双视通路面孔验证模型.....	16
3.1 概述.....	16
3.2 基于视觉信息流的双视通路的面孔验证模型.....	16
3.3 算法原理.....	17
3.3.1 前馈卷积网络.....	17
3.3.2 添加监督信号.....	18
3.3.3 面孔验证分类器.....	20



3.4 实验过程以及分析.....	20
3.4.1 双通道网络结构的性能测试.....	21
3.4.2 控制分类信号和监督信号之间的相关性.....	23
3.5 小结.....	24
第4章 基于视知觉皮质通路的正立和倒立面孔混合识别.....	25
4.1 概述.....	25
4.2 基于视知觉皮质通路的正立和倒立面孔混合识别.....	26
4.3 算法原理.....	27
4.3.1 底层视通路.....	27
4.3.2 局部池化层.....	27
4.3.3 多特征融合.....	28
4.3.4 预测函数.....	28
4.4 实验结果及分析.....	29
4.4.1 单一训练集情况下的识别性能.....	29
4.4.2 混合识别模型中的参数优化.....	31
4.4.3 样本集对识别性能的影响.....	32
4.5 小结.....	33
第5章 基于层次信息融合的多尺度面孔识别新方法.....	34
5.1 概述.....	34
5.2 基于层次信息融合的多尺度面孔识别.....	34
5.3 算法原理.....	35
5.3.1 层级采样.....	35
5.3.2 多层信息表达的全连接层.....	37
5.4 实验过程以及分析.....	38
5.4.1 网络层的采样频率选择.....	39
5.4.2 全局信息和局部信息的关系比.....	40
5.4.3 多尺度样本训练的影响.....	40
5.5 小结.....	41
第6章 总结与展望.....	42
6.1 总结.....	42
6.2 展望.....	43
致谢.....	44
参考文献.....	45
附录.....	45

## 第 1 章 绪论

### 1.1 研究目的与意义

伴随着互联网技术的快速发展，金融、安防和电子商务等领域对目标识别技术的需求也日益增长。高清成像技术和模式识别技术也在历史的浪潮中不断更新迭代，包括以 CCD 为代表的成像设备，以 GPU 等代表的高性能处理单元以及以深度学习为代表的人工智能算法。科研人员也基于此两类技术开展了目标识别技术的应用。例如当前苹果公司生产的 iPhone X 的刷脸支付<sup>[1]</sup>的功能引发了公众的广泛关注，这种支付方式避免了传统加密方式易于丢失的问题，而是采用了面孔等常见生物特征；又例如高速公路上车流量的加剧，利用实时车辆检测和识别技术<sup>[2][3]</sup>可以对车辆道路违法行为进行有效抓拍，起到了预警和防范的作用。综上所述，面向视觉识别的技术具有广阔的应用前景。

研究表明<sup>[4]</sup>，目标识别误差主要由类内误差和类间误差两个方面构成。其中类间误差是不同的识别单位之间本身存在的特征差别；类内误差是同一个识别目标因为光照原因、曝光角度、姿态和拍摄角度等不稳定因素影响造成的差异性。当前的目标识别技术也主要围绕以上述两类误差的研究发展，其主要分为以下两大类<sup>[5][6]</sup>，一个是传统的基于数学特性的视觉识别技术，其研究思路是以物体几何特征作为识别的依据，然而特征的随机性导致视觉识别的不稳定性，故难以推广到实际应用中；另一种技术为当前热门的基于卷积神经网络的视觉识别技术<sup>[7]</sup>能一定程度上解决实际问题，但是昂贵的硬件需求限制了该算法的应用领域，同时识别模型对于内存的巨大消耗限制了跨平台的可移植性。因此面向视觉识别的技术还有更广阔的发展空间。

生物感知视觉系统在大脑中占据了重要地位，其可以帮助人类毫不费力地快速识别目标，然而这对计算机识别来说，并非易事。因此越来越多的科研人员投入到视觉感知系统的研究。例如当前热门的深度学习方法启发自视觉感知系统中神经元细胞，其构建而成的人工神经网络在机器视觉领域中表现出优异的成绩<sup>[8]</sup>，由此可见研究感知计算模型的重要性。但是视觉感知系统的内部机理还有更加广阔的领域尚未被挖掘与利用。现代生理学实验揭示<sup>[9]</sup>，神经元之间相互传递系统不是以权值的方式传递<sup>[10]</sup>，而是利用通过放电的形式传递刺激，也不存在所谓的反向传播的梯度下降机制，也不需要大量的样本来进行大量的参数迭代训练等等特点，这些特点也是当前人工神经网络的局限性所在。现阶段面向视觉识别的感知计算模型的研究一方面采用仿生建模的方式来复现已存在的生理现象，以促进心理学专家在视觉感知领域的研究，另一方面也能为机器视觉领域提供新的思路，因此面向视觉识别的感知计算模型的研究具有重要的研究意义。

## 1.2 国内外研究现状

基于视觉感知的识别技术是利用待识别目标的视觉表征信息来鉴别的生物识别技术，其研究领域主要集成了图像处理、神经网络、机器学习和生物识别机制等等，其应用领域涉及人脸识别、车辆识别和场景识别等等。面孔识别技术在诸多视觉识别的研究中时间较长，最早可以追溯到 Galton 在 Nature 上发表的关于面孔识别的论文。该论文中针对于人类的面孔识别能力阐述和分析，为后续的面孔识别技术奠定了基础。在 1965 年，Bledsoe 提出了人脸识别系统 AFR (Automatic Face Recognition)，由此引发科研人员对面孔识别的研究热潮。当时的目标识别主要围绕识别目标关键的部位的几何特性<sup>[11]</sup>，但是该方法要求研究人员前期需要做大量的标记和测量等人工处理，因此仅适用于实验室内等约束条件下目标识别问题，应用到实际中效果将会表现很差。在 1993 年，Brunelli 和 Poggio 二人的对比实验表明模板匹配方法<sup>[12][13]</sup>在目标识别问题上的性能优于以往基于几何特性的方法，接着 DCT 变换<sup>[14]</sup>、小波变换等浅层特征提取方法相继提出。模版匹配方法的核心思想是将获取的样本平均合成一个基本样本，以特定方法处理过的样本线性组合，作为识别目标的描述，其在车牌识别<sup>[15]</sup>和面孔识别<sup>[16]</sup>中也有一定的应用。同一时期，Truck<sup>[17]</sup>等人提出了基于主成份分析<sup>[18]</sup>的面孔识别方法，在此基础上还演化出 LDA<sup>[19]</sup>和 Kernel LDA 方法。

上述传统的视觉识别方法虽然能在实验室目标识别应用上具有良好的鲁棒性，但是该类人工设计的特征受制于科研人员的日常生活的经验和直观感受的限制，提取后的特征其本质上为浅层特征。在图像预处理和特征提取的过程中，传统的方法不仅可能造成识别目标重要的表征信息丢失，还会人为舍弃有效信息。同时有效的浅层特征则需要人力深入理解待解决问题，经过反复的实验获取得到。因此，传统方法的弊端限制了目标识别技术的发展，也由于生物感知识别系统的诸多良好特性，科研人员也将目光投入到生物视觉领域。

卷积神经网络是深度学习领域应用最广泛的结构，其最初是根据感受野的生物特性构建的仿生网络。Hubel 等人早期通过解剖猫的视觉皮层发现，视觉皮层是以 LGN-V1-V2-V4-IT 这种层级结构形成视觉回路<sup>[20]</sup>。层级结构的感受野细胞如同级联的滤波器将捕捉的视觉区域，以层级激活的方式传递视觉信息，后续的研究工作表明该类型结构适合视觉识别。Lecun 等人利用具有卷积神经网络雏形的 Lenet<sup>[21]</sup>网络来识别手写数字，把数字识别的精读相对以往的神经网络算法有了新的高度，这也是卷积神经网络概念提出最早的文献。但是后续的几年，识别技术并没有在面孔和车辆等识别目标问题上有突破性的进展，尤其是复杂的场景下的识别问题。直到 2012 年，Alex 等人提出的深度卷积神经网络 AlexNet 在 ImageNet<sup>[22]</sup>大赛上在机器视觉领域取得了惊人的突破。由于非饱和的修正线性激活函数抑制了卷积神经网络反向传播的梯度弥散问题，使得网络层次有了较大提升，但是庞大的神经元参数和海量的数据库资源意味着对计算机运算能力的要求极其高，实际应用中难以展开。由于卷积结构在网络训练和测试中需要占用了较大计算资源，因此在后续的研究主要针对卷积操作的优化，相继出

现了基于小卷积核的 VGGnet<sup>[23]</sup>、基于网络嵌网络结构的 Googlenet 系类<sup>[24][25][26][27]</sup>和基于残差网络结构的 ResNet<sup>[28]</sup>。为了满足手机端等硬件需求,科研人员也尝试极度压缩模型,并使得其损失值控制在合理的范围内,例如 Shufflenet<sup>[29]</sup>在深度可分离的卷积思想的基础上,利用输入特征图谱通道均衡混合,以获得更好的全局信息; Mobilenet<sup>[30]</sup>在其基础上利用采用通道拆解和融合的方法来分解卷积,模型大小进一步压缩。在目标检测识别领域, RCNN<sup>[31]</sup>利用卷积网络在候选区域上识别,大大地提高了识别的精度和识别效率,后续的 SPP-NET<sup>[32]</sup>网络保留了原有网络的整体框架,引入空间金字塔池化层,解除网络中的固定尺寸约束,也减少了训练过程中过拟合的现象发生;相继的 Ross Girshick 等人提出 Fast-rcnn<sup>[33]</sup>和 Faster-Rcnn<sup>[34]</sup>,通过将类别判断和位置精调统一用深度网络实现,从而不需要额外的存储。

深度学习算法已经在工业界得到了一定的应用,进一步推动了机器视觉的应用,但是距离广泛普遍的应用仍然有一定差距。为解决这个问题,生物学和人工智能专家通过不断的揭秘视觉感知系统的结果组织以推动人工智能的发展。为了模拟神经元内部的信息传递方式,科学家们在神经元之间编码方式展开了深入地探索。Rullen 和 Cautrais<sup>[35][36]</sup>通过构建多层网络,每次抓取首次放电神经元的信息,利用层层放电的模式,构建了 Spikenet 模式,即模拟感受野神经节细胞 V1 等视通路的过程,根据最后神经元发放的脉冲情况判断在该区域是否存在面孔,证明了视通路机制在目标识别领域具有独特的应用。在 2017 年, Hinton<sup>[37]</sup>等人根据大脑皮层中存在的被称为 Cortical minicolumn 的柱状结构提出了对应的 Capsule 的网络结构,采用向量输出来替代一般的神经网络输出,使得输出的特征对识别目标更加具有表征意义。

### 1.3 本文主要工作及创新点

本文的研究内容主要是围绕视觉感知机制的视觉通路和卷积神经网络模型展开,探索视觉神经机制在面孔识别上的应用。本文首先对视觉通路上的层次结构进行一定的阐述,其涉及到视网膜上感受野性能、初级视觉皮层的方向选择机制以及大脑中枢的左右半脑协作机制等等,进而引入卷积神经网络模型,利用神经元的接受信息和激活等操作对面孔图像进行一定的表征,当中涉及到卷积网络的卷积、池化和激活等操作。首先,增加监督信号参与面孔识别模型的训练,适应日常的面孔验证的实际应用;其次,考虑面孔识别的左右半脑协作机制,构建了基于局部区域提取的池化神经网络层,达到正立和倒立面孔特征差异性表征和识别的目的;最后,考虑到不同的神经网络层的特征信息参与识别过程,提出了多层信息表达的网络结构,能够应对不同尺度的面孔输入场景。

论文具体研究内容以及创新点包括以下几个方面:

(1)提出了一种基于视觉信息流的双视通路的面孔验证模型。构建底层卷积网络,模拟生物感知系统对视觉信息流的过滤;提出了双视觉通路概念,构建监督信号和识别信号融合的增强网络,实现对视觉识别特征的强化;最后,通过比较证件照片和多姿态面孔照片的空间

分布情况，利用联合贝叶斯人脸的方法辨别面孔的身份异同。实验结果表明，双视觉通路的设计能够提高面孔验证的准确率，说明了监督信号对面孔特征提取有约束作用。

(2)提出了一种基于左右半球协作机制的正立和倒立面孔识别新方法。模拟视觉信息流在视通路中的传递和处理过程，首先构建底层神经网络，建立敏感纹理特征以及对称卷积核的机制，实现正立和倒立面孔图像的去除冗余和预处理；接着提出一种基于局部区域提取的池化神经网络层的概念，构建多局部特征融合的网络结构，实现局部信息的压缩提取和融合；最后根据高级视觉皮层中左右半脑协作的特点，提出一种融合整体和局部信息的预测函数。实验结果表明，本文方法能够识别正立和倒立面孔，说明了整体和局部特征的两条视觉通路虽然分别在正立和倒立面孔识别上起了决定性的作用，但它们并不是孤立存在的，两条通路所刻画的面孔信息应该是一种互补式的关系。

(3)提出了一种基于多层次信息表达的多尺度面孔识别新方法。首先，模拟视网膜和神经节细胞对底层纹理特征的敏感的特点，构建多层卷积池化结构级联的神经网络，以达到视觉信息过滤和提炼的目的；其次，考虑到底层卷积对面孔特征表达的有益作用，构建了基于局部和整体结合的多尺度池化的概念，输出固定大小的特征通道和特征图谱；然后，根据视觉皮层的决策机制，各层次的特征信息融入到全连接层，通过神经元的激活特性判断面孔身份。实验结果表明，本方法可以有效适用于多尺度面孔图像的识别，从而说明底层纹理特征将有助面孔特征的表达。

## 1.4 论文的组织结构

本文的组织架构以及各章节的内容安排如下所述：

第一章绪论，首先介绍了目标识别的研究价值，就目标识别技术在当下社会的意义价值做了阐述，并且结合国内外视觉识别技术的难点，对经典的识别技术做了分析，总结基于视觉感知系统的目标识别技术的优势和不足之处；针对现有的目标识别方法的局限性，结合视觉感知机制，有针对性地给出了本文的主要研究内容以及创新点；最后对论文的组织架构和章节安排进行简单的介绍和说明；

第二章首先介绍感知计算模型的基本原理，本文涉及到视觉信息流通路与视觉感知机制基本原理的进行阐述，包括眼球、初级视觉皮层和大脑的组织结构。此外，本文将对神经元细胞建模进行分析，主要包括卷积操作、池化采样和神经元激活等神经元机制，为后续计算模型的改进以及模型构建奠定了理论基础。

第三章主要对基于视觉信息流的双通路面孔验证新方法的研究。研究底层视觉通路的内部机理，构建特征提取网络，并对其参数训练进行一定的阐述；为了对面孔特征有约束作用，引入了双视觉通路的概念，构建监督信号和识别信号融合的特征增强网络；最后，分析面孔特征误差的特点，构建联合贝叶斯的面孔验证分类器，最后达到面孔验证的结果。实验结果表明，笔者将双通路的面孔验证模型，与原系统对比，可以发现引入监督信号的面孔验证模

型可以提高面孔验证的准确率。

第四章提出了一种基于视皮层通路的正立和倒立面孔识别的方法。本文保留原有基于轮廓信息的网络层对识别率的贡献，增加了面孔表面具有表征性的局域纹理信息，设计并且实现了基于局部区域提取的池化神经网络结构，构建了多局部特征融合 **MLPB** 的网络结构，通过模拟左右半脑对整体和局部特征的协作关系，提出一种融合整体和局部信息的预测函数，给出了一种针对正立面孔和倒立面孔混合识别的新模型。实验结果表明，本章的方法不但能提高面孔识别的精度，而且能够实现正立和倒立面孔同时识别的性能。

第五章主要对本文提出的基于层次信息融合的多尺度面孔识别新方法进行研究。考虑到神经网络各层对识别目标的有不同的程度表征，当尺度变化后的输入图像，面孔图像的特征会在一定程度上的丢失，从而将各层的浅层的纹理特征能够投影到视觉决策分析层，针对于经典的面孔识别网络需要对输入的目标图像有固定的尺度大小，构建了固定输输入的池化网络层，由此可以保证决策分析神经网络层的特征的大小固定，从而网络的适应性增强，最后实验结果表明，本章提出的算法不仅解决了面孔识别模型需要固定尺寸输入的弊端，而且具有尺度的泛化能力，从另一方面体现出面孔识别并非是基于单一的视觉通路，而是多视觉通路的存在

第六章是本文内容综合以及对改技术研究和发展的展望。本文将结合视觉神经通路信息流特性，在经典的卷积神经网络的基础上，对特征提取模型和训练模型的改造，解决了日常面孔识别的实际问题并且取得了较好的结果。实验结果说明，生物视觉特征对面孔识别技术有潜在的借鉴意义，也对揭露大脑奥秘有促进作用。

## 第 2 章 视通路以及卷积神经网络

### 2.1 视通路概述

视通路在视觉感知系统中占据了非常重要的地位，也是人类获取信息的重要的渠道。视觉信息从视网膜起最终到达大脑视觉皮层。不同类型的组织细胞分布在不同的组织表面，对应的形成相应的视觉感知机制。经过科学家的研究发现，视觉通路的底层的组织结构<sup>[38]</sup>对图像表面的简单纹理信息较为敏感，随着信息的层级传递，较为后面的神经细胞将会对形状成分的纹理结构信息较为敏感，最后是表征到物体的整体轮廓信息。与此同时，视觉皮层中某些组织被部分损伤并不会影响的生物识别的能力，证明识别并不依靠单个单元，而是许多单元的集体激活的响应。即使失去部分神经元细胞，视觉感知系统同样能够保持目标识别能力。其原因是视觉感知系统中独特的组织形式构成了人类视觉感知系统的稳定性以及接受新事物的高效性。本小节将对本篇论文中所阐述的视觉模块以及其表现的生物学特征进行阐述，视通路的整体结构如图 2.1 所示。

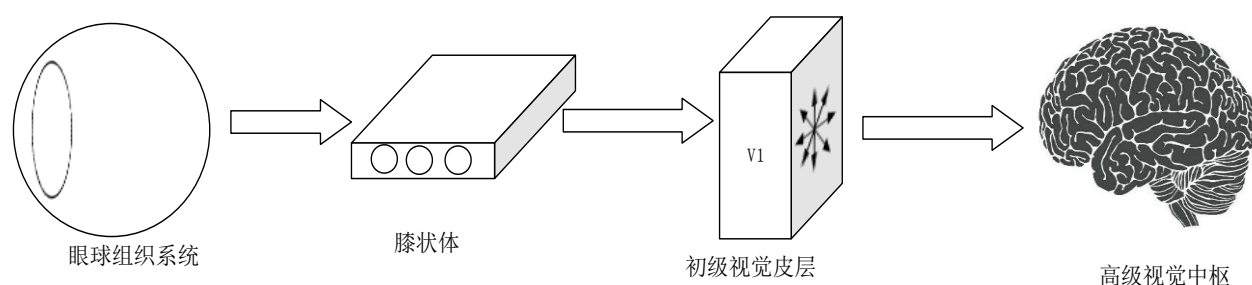


图 2.1 视觉通路的整体结构

#### 2.1.1 视网膜

视网膜是视觉感知系统接受外界刺激的关键节点，在神经科学界被称之为“外周脑”。从神经组织结构来说<sup>[39]</sup>，视网膜的组织细胞从外面到里面依次为视杆细胞和视锥细胞、水平细胞、双极细胞等。光线从角膜进入后将激活位于视网膜后表面的感光细胞，即是视杆细胞和视锥细胞。其中短波（蓝波）视锥细胞对波长 430nm 的关系有最大的反应；中波和长波视锥细胞的感受性峰值移到了更长的波长。事实上，人类虽然具有 2 亿 6 千万感光细胞，但却只有 200 万的神经节细胞，即是视网膜的输出细胞，这个特点表明感光细胞获取的信息在视网膜的中间层进行高效地加工。

#### 2.1.2 膝状体与初级视觉皮层

神经学专家表明，所有的视神经将会终止于皮质结构的位置，但是存在不同类型的视通路。但是研究人员主要关注的是视网膜-膝状体（LGN）-大脑视觉皮层的通路，因为该视通

路包含了超过 90% 的视神经轴突，故被称之为第一视通路。如图 2.2 所示，膝状体的细胞具有中心兴奋-周围抑制或者周围兴奋-中心抑制的同心圆式感受野结构。值得注意的是，初级视皮层的简单细胞可以根据相互邻近侧膝状体细胞感受野相互链接而成。除了对边缘发放出信号之外，简单细胞具有方向选择性质，即是在最优朝向的边缘处产生抑制或者兴奋，但是在垂直朝向上神经元将不会活动。因此视觉感知系统已经对外界接受的刺激进行初步信息加工，并且获取到物体表面简单的纹理信息。

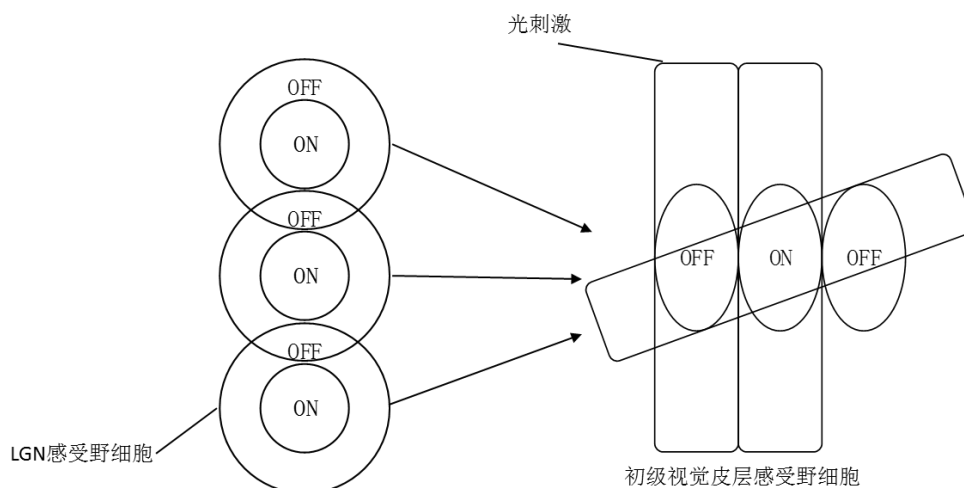


图 2.2 感受野细胞组合特性

### 2.1.3 左右半球系统

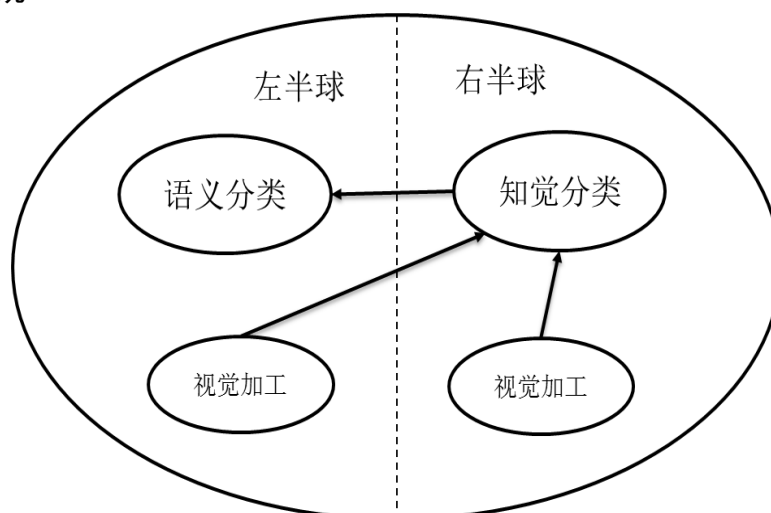


图 2.3 左右半球视觉系统分工

大脑皮层对接受到的电信号刺激处理加工，然后传递到高级视觉中枢决策分析。**Warrington** 提出基于视觉感知识别相关的操作的解剖学模型<sup>[40]</sup>，理论中说明大脑左右两个半球均参与视觉分析。物体视觉识别主要分为知觉表征阶段和语意知识联系两个阶段，两者之间并非一分为二而是相互关联的。右半脑视觉系统进行物体识别的第一阶段，即知觉分类，其要求对接受的视觉刺激以独特来源性和不变来源进行区分；之后的语义分类功能主要位于



左半球视觉系统，左右半球视觉系统分工如图 2.3 所示。

左半球视觉系统主要是分析加工，而右半球视觉系统主要为整体加工，因此被称之为左右半脑协作机制。与此同时，大脑内部错综复杂，物体识别功能的实现并非依靠某个结构单独能够完的，例如最终的语义分类的阶段也有右半脑的参与。本论文的第四章的研究将根据此模块展开分析研究。

## 2.2 神经元细胞

在细胞学层面，细胞作为组织的基本构成单元，其结构特点构成了组织所具有的功能，其中深度学习的基本结构——卷积结构来源于视觉感知机制中的基本单元感受野。在神经元传递信息方面，电生理学实验显示神经元细胞的信息传到形式以电信号为主，即通过细胞膜表面的动作电位形式向前或向后传导。这些网络的传递速率是毫秒级，但是现在神经网络层的传递速率是微秒级。

### 2.2.1 局部连接

感受野概念<sup>[41]</sup>对卷积神经网络的提出有重要作用，其利用局部感受野的结构来接受外界刺激，与图像本身在空间联系上的局部性，由此局部连接层设计的立足点。局部连接的具体思想是：在卷积神经网络中，感受野神经元都只对局部图像进行感知，而不是全局图像的感知，从而局部的视觉信息将在更高层的神经元细胞中综合，通过这样的方式得到图像的全局信息。感受野的计算公式如式(2.1)所示。

$$X_j^l = f(\sum X_j^{l-1} * K_{ij}^l + b_j^l) \quad (2.1)$$

其中 $*$ 表示为卷积操作， $X_j^l$ 表示为第 $l$ 层卷积层输出的第 $j$ 个特征图谱， $K_{ij}^l$ 表示为该层的卷积核的权值参数， $b_j^l$ 表示为卷积层的偏置参数， $f$ 表示为激活函数。

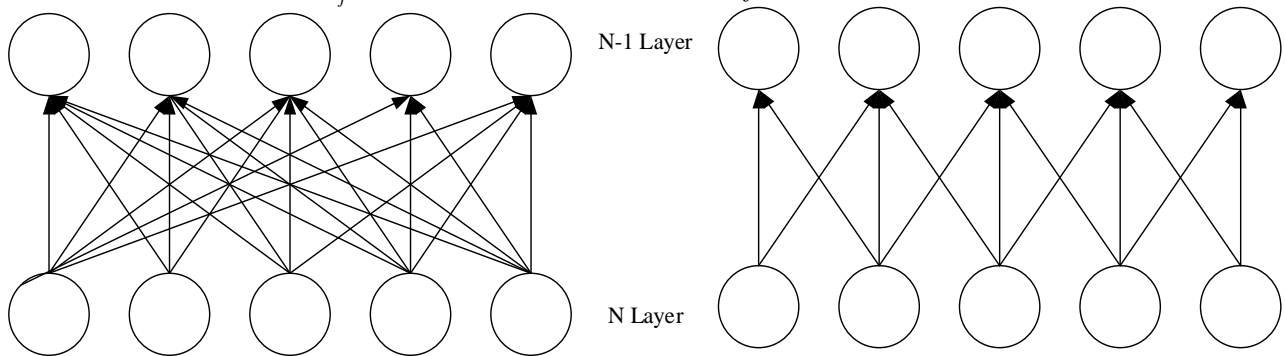


图 2.4 全连接层和局部连接层的神经元连接示意图

局部连接的重要作用是减少权值参数的数量，根据 VC 理论，一定程度上可以减少模型过拟合现象。如图 2.4 所示，在 $N-1$ 层有 $K$ 个节点， $N$ 层为 $L$ 个节点，则此时网络的参数为 $K * L$ ，但是如果采用局部连接的方式，则需要的参数仅仅为 $2L$ 。因此在大规模的卷积神经网络中，利用具有局部连接的卷积网络替换全链接层是压缩模型的重要手段。感受野的卷积核大小不同，感受野的功能也有所不同，例如 $1 \times 1$ 的卷积核在数学上是矩阵变换的作用，还

具有通道升降维度和跨通道的信息融合的能力；较大卷积核（7x7）的感受野细胞可以获取到更加全范围的视觉信息。

### 2.2.2 激活函数

激活函数是在神经节细胞的重要组成部分，其生物学原理为每个神经元细胞接受到周围神经突触传递来信号，若经过细胞体汇合之后的信号大于阈值时，神经元将会被激活，通过树突放出信号。随着神经网络的发展，针对不同的需求激活函数的类型也多样化起来，以下将会介绍常用的激活函数，分别为 Sigmoid 函数、Tanh 函数和 Relu 函数。

- (1) Sigmoid 函数。此函数从数学上分析是一种非线性转移函数，也是较为符合生物特性的激活函数。如图 2.5 所示，其可以将输入的数值转化为 0 到 1 之间的数值。但是工程人员发现该激活函数在大规模网络的参数训练时，容易造成梯度弥散现象，即当输入接近 0 或者 1 时，其函数的梯度将趋近于 0。其数学表达如式(2.2)所示：

$$\sigma(x) = 1 / (1 + e^{-x}) \quad (2.2)$$

- (2) Tanh 函数。如图 2.6 所示，该类函数与 Sigmoid 函数类似，区别在于其函数输出为-1 到 1 之间的数值。但是神经网络反向传播时，该类激活函数也存在梯度弥散的现象，从而也受限与一些特定的场合，其数学表达式如式(2.3)所示。

$$\text{Tanh}(x) = (e^x - e^{-x}) / (e^x + e^{-x}) \quad (2.3)$$

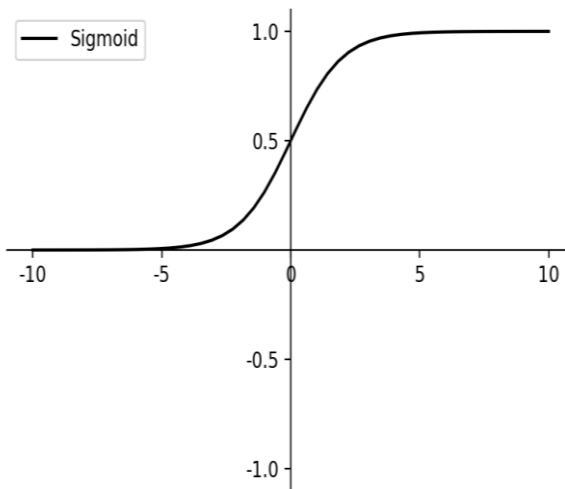


图 2.5 Sigmoid 函数曲线

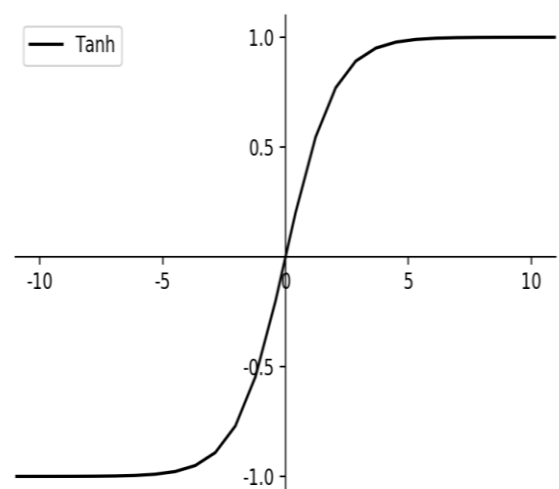


图 2.6 Tanh 函数曲线

- (3) Relu 激活函数。其训练曲线如图 2.7 所示，Relu 相比于 Sigmoid 和 Tanh 能够使神经网络模型在训练过程中快速收敛，也有效缓解了梯度消失的问题。但是在训练过程中时，较大的样本波动可能会出现神经元不可逆转地死亡现象，即权值无法更新。其数学表达式如式(2.4)所示。

$$\text{Relu } f(x) = \max(0, x) \quad (2.4)$$

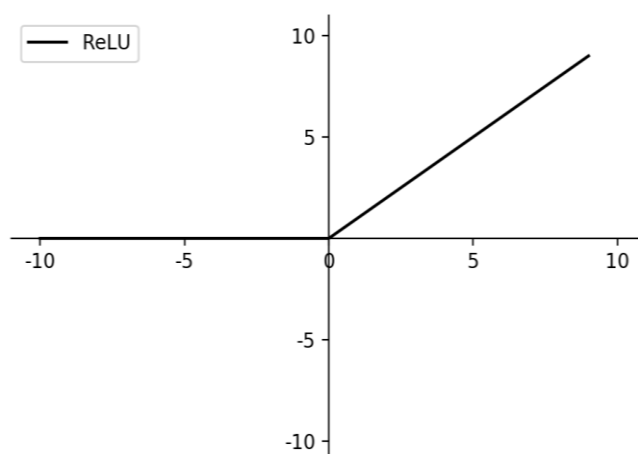


图 2.7 Relu 函数曲线

### 2.2.3 池化采样

池化采样是一种高效的降低数据维度的策略，使得模型具有一定的旋转不变性。较早的经典神经网络采用池化层的方式来实现数据的池化采样，后来由于考虑模型的运行效率的问题，则科研人员将其池化采样操作嵌入到卷积结构中。

池化操作的基本思想是，将卷积特征划分为若干个不相交的区域，通过求这些区的平均值或最大值得到降维后的特征。如图 2.8 所示，针对于  $4 \times 4$  的特征图谱，若采用  $3 \times 3$  的尺度池化，那么下采样得到一个大小  $2 \times 2$  的特征图谱。同时池化采样有以下两个优点：降低模型的复杂度，一定程度上降低模型的过拟合现象；提高对图像位移的鲁棒性，即是当图像发生小的位移的时候，保证提取的特征的不变。

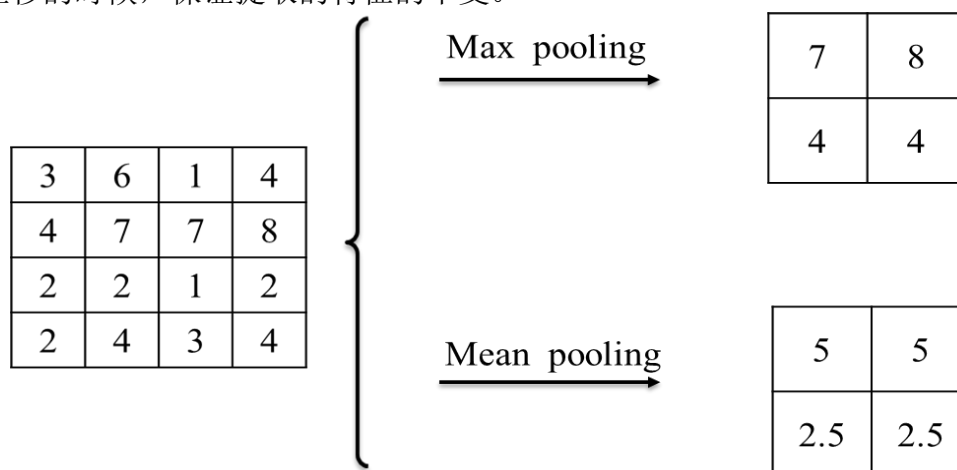


图 2.8 最大池化和平均池化操作

### 2.2.4 Softmax 分类器

图像等输入导入到在卷积层内部，经过多次卷积操作和池化，将输入的特征向量展开成为一维向量分类器分类。考虑到视觉感知机制中的生物学特性，本文所有分类器将都会采用 Softmax，其本质上是逻辑回归函数的推广，以适用于图像的多分类问题。针对于训练的样本

$\{(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}$ , 有  $y^{(i)} \in \{1, 2, \dots, k\}$ , 则输入样本  $x^{(i)}$  分类为  $j$  的概率如式(2.5)所示。

$$p(y^{(i)} = j | x^{(i)}; \theta) = \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^k e^{\theta_l^T x^{(i)}} \quad (2.5)$$

其中,  $\theta_1, \theta_2, \dots, \theta_k \in \mathbb{R}^{n+1}$  是模型的参数, 式中的分母目的是对概率分布做归一化操作, 即  $k$  个分类点将对应  $k$  个概率值, 总概率相加为 1。函数分类器主要应用于卷积模型的测试网络, 测试网络为 Softmax 的分类器的代价函数  $J$  如式(2.6)所示。

$$J(\theta) = -\frac{1}{m} \left[ \sum_{i=1}^m \sum_{j=1}^k \mathbb{1}\{y_i = j\} \log p(y_i = j | x_i) \right] \quad (2.6)$$

其中  $\mathbb{1}\{\}$  表示为示性函数, 即  $\mathbb{1}\{\text{表达式为真}\} = 1$ ,  $\mathbb{1}\{\text{表达式为假}\} = 0$ ,  $m$  表示为批量梯度下降所选用的样本数量。上述代价函数采用交叉熵的形式, 表示了神经网络输出的预测分类结果与真实的样本标签之间的差距。考虑在极端情况下, 当预测值与标签之间的完全一样的时候, 损失函数也将达到最优数值。

除了上述的交叉验证函数之外, 均方误差的代价函数的使用也较为普遍, 如式(2.7)所示。

$$J(W, b; x, y) = \frac{1}{2} \sum_{k=1}^c (t_k - y_k)^2 = \frac{1}{2} \|t - y\|^2 \quad (2.7)$$

其中  $c$  表示为样本的类别,  $W, b$  代表神经网络的权值参数和偏置参数,  $t_k$  表示预测值的第  $k$  维,  $y_k$  表示为标签的第  $k$  维。损失函数的种类会根据实际分类的任务的场景需求选择和修改补充, 无论任何一类损失函数, 卷积网络都会采用梯度下降的方式更新网络中的权值, 最终实现损失函数最小化。但是该过程中将存在模型的过拟合现象风险, 因此一般通过交叉验证的方式来验证选择模型的参数。

## 2.3 防止网络过拟合的方法

### 2.3.1 Dropout 层

大脑中神经元细胞不停地进行新陈代谢, 有时一些神经元细胞发生死亡但是并不影响视觉系统识别的正常工作。同时视觉感知系统对有一定遮挡的识别目标, 也同样具备一定的辨别能力识别。为了模拟此类生物现象, Hinton 提出了 Dropout 的概念, 在神经网络训练阶段让部分特征不参与目标的识别, 考虑到模型收敛的问题, 该结构摆放位置在全链接层后端, 其具体表达式如式(2.8)和式(2.9)所示。

$$y' = r * y \quad (2.8)$$

$$z_i = f(w_i y' + b_i) \quad (2.9)$$

其中  $y$  是全连接层的输出,  $r$  作为控制参与识别的特征占有所有特征的比重, 其取值范围为  $[0, 1]$ ,  $y'$  是经过 Dropout 层后的输出, 当  $z_i$  为经过激活函数后的输出特征。值得注意的是,

具体地在每次前向传播，系数将会设置为 1。如图 2.9 所示，左图为完整的神经网络结构，右图为采用 Dropout 时的网络结构。从右图可以看出，训练过程中半数神经元属于暂时的死亡状态，因此网络迭代多次也意味训练了多个卷积神经网络模型。这也恰恰与生物现象中相符合，即冲击等外界因素导致部分神经元的死亡，但是仍然不影响视觉感知系统的识别能力。

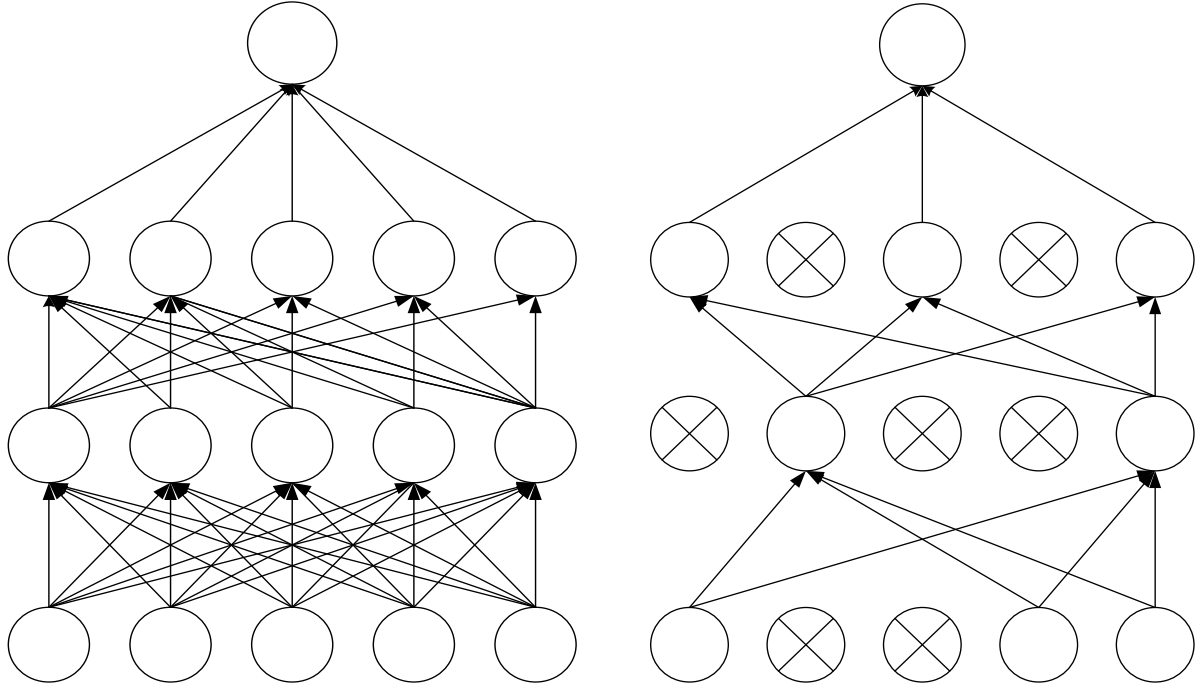


图 2.9 Dropout 功能示意图（左图无 Dropout，右图采用 Dropout）

### 2.3.2 BN 层

BN (Batch Normalization) 层从字面意思上可以理解为批量规范化，其本质上是应对反向传播过程中的梯度问题。其过程为在每次反向传播时，根据输入小的样本数量来应对激活函数来做规范化操作，使得输出的信号的各个维度的均值为 0，方差为 1。假设输入的特征为  $B = \{x_1, x_2, \dots, x_m\}$ ，则特征向量的均值和方差如式(2.10)和式(2.11)所示。

$$\mu_B \leftarrow \frac{1}{m} \sum_{i=1}^m x_i \quad (2.10)$$

$$\sigma_B^2 \leftarrow \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2 \quad (2.11)$$

则特征向量正则化的计算如式(2.12)和式(2.13)所示， $\varepsilon$  本身就是一个较小的数值，存在的意义就是防止方差为 0 出现导致的系统混乱。

$$\hat{x}_i \leftarrow \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} \quad (2.12)$$

$$y_i \leftarrow \gamma \hat{x}_i + \beta \equiv BN_{\gamma, \beta}(x_i) \quad (2.13)$$

其中  $\gamma$ ,  $\beta$  表示为需要学习的参数。BN 操作其作用在非线形映射前和仿射操作后。当神经网络训练时遇到收敛速度很慢或无法收敛的状况，可以尝试 BN 来解决。同时加入 BN 来

加快模型训练速度，提高模型精度。

### 2.3.3 正则化

模型训练最终目的是训练的模型在测试数据集上错误率最小化。正则化手段繁多，下面介绍最常用的 L2 正则，其表达式如式(2.14)所示。

$$C = C_0 + \frac{\lambda}{2N} \sum_w w^2 \quad (2.14)$$

其中  $C_0$  表示为损失函数，另一项表示为惩罚项， $w$  表示为神经网络的权值参数， $\lambda$  是正则系数， $N$  为总体的样本个数。

因此，权值  $w$  和偏置  $b$  的梯度如式(2.15)和式(2.16)所示。

$$\frac{\partial C}{\partial w} = \frac{\partial C_0}{\partial w} + \frac{\lambda}{N} w \quad (2.15)$$

$$\frac{\partial C}{\partial b} = \frac{\partial C_0}{\partial b} \quad (2.16)$$

由此可见，惩罚项对偏置  $b$  没有影响，仅仅对权值  $w$  的更新有作用，因此权值更新的表达式如式(2.17)所示。

$$\begin{aligned} w &= w - \eta \frac{\partial C_0}{\partial w} - \frac{\eta \lambda}{N} w \\ &= (1 - \frac{\eta \lambda}{N}) w - \eta \frac{\partial C_0}{\partial w} \end{aligned} \quad (2.17)$$

由上式可以看出，权值的系数是小于 1，其作用为实现权重的衰减。

## 2.4 小结

本章节主要阐述了本文所需求的理论基础，其中包含关于视觉感知机制的视觉通路的机理和卷积神经网络的相关技术。在第一部分中，笔者主要对视觉通路进行了一定的阐述，包括视网膜、膝状体、初级视觉皮层和大脑。对上述结构的简略描述对面孔识别的卷积神经网络设计有重要的描述作用。在第二部分中，由于本文设计的网络是在卷积神经网络上优化而成，故对卷积网络的基本结构进行一定的阐述，包括卷积层、池化采用层和全连接层模块。在第三部分中，对防止模型过拟合方法的研究，有助于实验快速实现。

## 第3章 基于视觉信息流的双视通路面孔验证模型

### 3.1 概述

对同一身份的面孔会因为不同的姿势、照明、年龄等因素而表现不一样，但是这些因素对面孔识别来说是致命的。同时由于身份差异引起的面孔相似性，使得在非约束条件面孔识别更具有挑战性。因此，减少个体面孔样本内部差异，扩大个体间的差异，是人脸识别中一个永恒的话题。该类思想可以追溯到早期的子空间人脸识别方法如 LDA<sup>[19]</sup>，贝叶斯脸<sup>[42]</sup>和特征子空间<sup>[17]</sup>。例如，LDA 通过使用两个来近似内部和内部的人脸变化；线性子空间是通过寻找到投影方向，使得个体内部和个体间差异的比率最大化。

本文认为面孔识别利用分类信号和监督信号来学习面孔特征是很有必要的。值得说明的是，面孔验证是将一对输入的图像归类是否同一类，即为二元分类问题，然而面孔识别是将输入的图片身份归类，则是一对多的问题。面孔验证模型流程一般分为是特征提取和特征分类两个阶段。在特征提取阶段，给定一个输入待识别的面孔图像，通过前馈卷积神经网络特征的非线性拟合，然后映射到大量的身份识别的决策分析层；在特征分类阶段，前馈神经网络输出的面孔特征通过其他的决策分析函数可以推广到其他任务，即在训练数据中不存在的新身份验证。面孔验证模型的端对端训练，可以提高面孔验证的准确率。

但是相关研究表明<sup>[43]</sup>，识别信号对同一身份的面孔图像特征具有较弱的约束性，因此新面孔特征提取时，不同面孔的面孔特征很大可能映射到相同身份的特征子空间。考虑到视觉感知机制中面孔识别比较学习的方式以提高学习效率，本文提出了一种基于视觉信息流的双通路的面孔识别方法，通过双通路的特征比较产生了外的面孔验证信号来增强特征的约束性，最终实现从属于同一身份的面孔特征向量彼此接近，从属于不同的身份的特征之间是相互远离的，最终证实了强约束性能有效降低个人特征的变化，提高面孔验证的准确率。

### 3.2 基于视觉信息流的双视通路的面孔验证模型

由于卷积神经网络本身拥有的特征提取能力，因此底层视觉通路将会采用卷积层和池化层结合的级联设计而成。模拟视觉感知系统学习和比较的流程，本文将设计两条视觉通路，其中一条视觉通路将输入证件照的图像，另一条视觉通路将会输入多姿态的同一尺度的场景内面孔图像。值得注意的是，视觉系统学习过程中共享一条视通路，即两条视同路的模型共通一套参数的，从而保证两条视觉通路的前馈卷积网络的特征一致的。如图 3.1 所示，为了模拟视觉系统的高级视觉皮层的决策分析，本文将提取后的两条特征向量融合，导入到二元分类器。

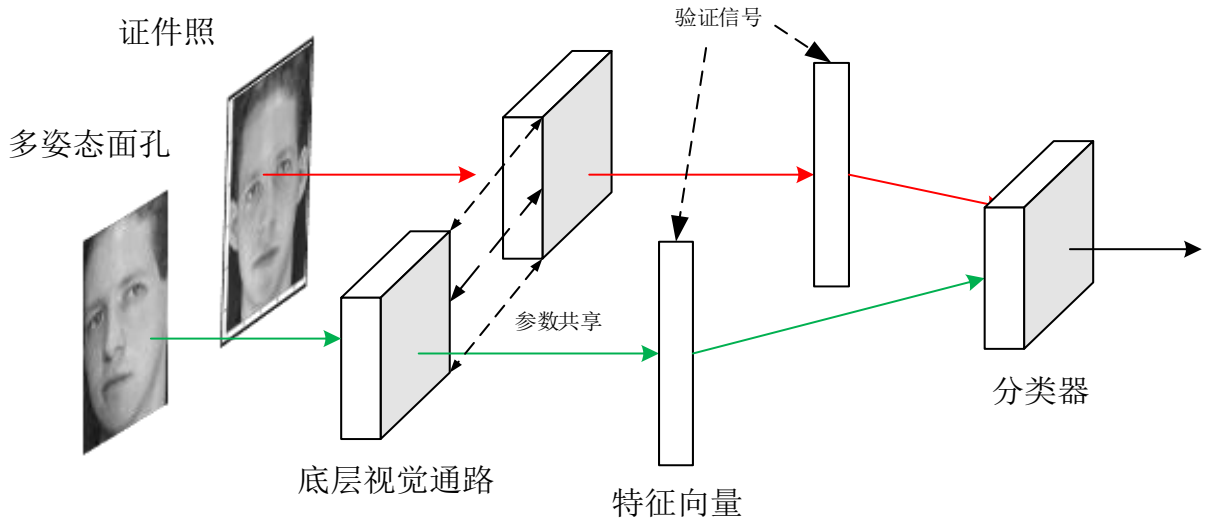


图 3.1 双通路网络结构图

### 3.3 算法原理

#### 3.3.1 前馈卷积网络

卷积神经网络基本结构是特征提取层和特征映射层两层堆式叠加，具有优秀的杂质信息过滤和纹理信息特征提取能力，同时前馈卷积神经的神经元功能符合生物视觉感知机制中的层级编码的机制，即神经元关注某个局部特征或者分类辨别的操作。因而本文的底层特征提取模块将会利用层级构成的卷积神经网络。

前馈神经网络在初始化时内部参数是高斯随机分布的，若此时输入面孔图像，输出的特征将无表征性。因此，工程上卷积神经网络参数是需要以 Softmax 构成的损失函数作为模型训练的衡量标准，使用批量梯度下降的方法并结合反向传播算法实现参数迭代更新，具体步骤如下所示。

(1) 计算输出层  $n_l$  输出的梯度，如式(3.1)所示。

$$\delta^{(n_l)} = -(\nabla_{a^{n_l}} J) \bullet f'(z^{(n_l)}) \quad (3.1)$$

其中  $f$  作为神经元的激活函数， $z$  是仿射求和得到和， $l$  表示网络的层数， $J$  表示损失函数的损失值。

(2)  $l = n_l - 1, n_l - 2, n_l - 3, \dots$ ，则中间变量  $\delta$  如式(3.2)所示。

$$\delta^{(l)} = ((W^{(l)})^T \delta^{(l+1)}) \bullet f'(z^{(l)}) \quad (3.2)$$

则权值  $W$  和偏差  $b$  的偏导数如式(3.3)和式(3.4)所示。

$$\nabla_{W^{(l)}} J(W, b; x, y) = \delta^{(l+1)} (a^{(l)})^T \quad (3.3)$$

$$\nabla_{b^{(l)}} J(W, b; x, y) = \delta^{(l+1)} \quad (3.4)$$

其中， $(x, y)$  为训练样本， $x$  是输入的面孔图像， $y$  是输出的标签。 $a$  表示激活函数的输



出，即是  $a^{(l)} = f(z^{(l)})$ 。

(3) 权值梯度  $\Delta W^{(l)}$  和偏差梯度  $\Delta b^{(l)}$  的更新如式(3.5)和式(3.6)所示。

$$\Delta W^{(l)} := \Delta W^{(l)} + \nabla_{W^{(l)}} J(W, b; x, y) \quad (3.5)$$

$$\Delta b^{(l)} := \Delta b^{(l)} + \nabla_{b^{(l)}} J(W, b; x, y) \quad (3.6)$$

(4) 权值  $W^{(l)}$  和偏差  $b^{(l)}$  更新如式(3.7)和式(3.8)所示。

$$W^{(l)} = W^{(l)} - a \left[ \left( \frac{1}{m} \right) \Delta W^{(l)} \right] + \lambda W^{(l)} \quad (3.7)$$

$$b^{(l)} = b^{(l)} - a \left[ \frac{1}{m} \Delta b^{(l)} \right] \quad (3.8)$$

其中  $\lambda$  表示为神经网络的学习率， $m$  表示批量梯度下降的样本数量。

### 3.3.2 添加监督信号

由于识别信号对同一身份的面孔图像特征具有较弱的约束性，不同面孔的面孔特征很大可能投影到映射到相同身份的特征子空间。因此本文提出了一种基于融合识别信号和监督信号的双通道的面孔识别系统，其基本思想是将证件照的面孔特征作为聚类的核心，通过证件照面孔与多姿态面孔之间产生的损失值作为监督信号，使得两者之间的特征能够相互聚拢最终能够达到约束面孔特征的目的。

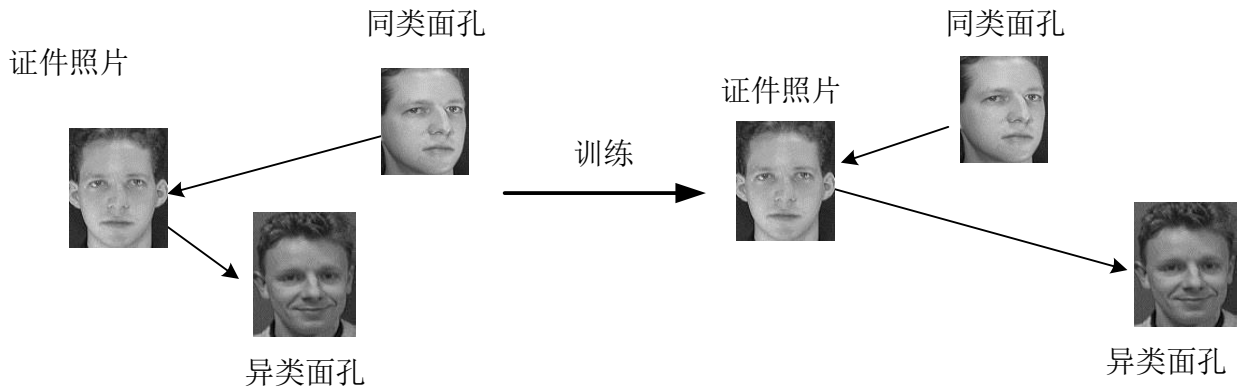


图 3.2 期望的特征分布变化

如图 3.2 所示，将提取的特征任意取两维度投影到二维度平面内部，理想二维特征的分布应为同类的面孔特征聚类在一起，异类面孔之间的类间距较大，因此需要在训练过程中使得异类的特征空间上分布相互远离，相类的面孔特征度量减小。

为了卷积神经网络模型训练能够快速收敛以及考虑到理想状态下的同类面孔特征的差异性较小，故本文将前馈神经网络输出的特征归一化后差值计算，以表示两条视觉通路输入的特征之间的度量关系  $d$ ，表达式如式(3.9)所示。

$$d = \frac{f_i}{\|f_i\|} - \frac{f_j}{\|f_j\|} \quad (3.9)$$

$f_i$  和  $f_j$  分别作为通道 1 和通道 2 输入的提取面孔的表征信息,  $\|f_i\|$  与  $\|f_j\|$  则是两者的泛数。实验过程中发现, 若无归一化操作可能导致模型训练混乱。

以双通路输出的特征之间的度量关系作为依据, 来判别输入的两张面孔特征是否来源于一个人的置信度  $\hat{y}$ , 如式(3.10)所示。

$$\hat{y} = \sigma(wd + b) \quad (3.10)$$

$\sigma$  为 Sigmoid 激活函数, 其为最接近神经元特性的激活函数。 $w$ ,  $b$  分别是激活函数内部的权值参数和偏置参数, 其需要通过样本训练学习得到。值得注意的是,  $\hat{y}$  的取值范围在  $(0,1)$ , 则在输入接近为 0 时, 则预测两张面孔的身份极有可能不一致; 如果输出接近 1, 则预测判定两张面孔的身份可能是一致的。

验证层的损失函数通过真实的标定信息和网络预测的标定信息之间的信号差值为衡量标准,  $verif(f_i, f_j, y_{ij})$  如式(3.11)所示。

$$verif(f_i, f_j, y_{ij}) = \frac{1}{2} \lambda (y_{ij} - \hat{y}_{ij})^2 \quad (3.11)$$

其中,  $y_{ij}$  作为两张输入面孔是否来自于同一个身份的标准, 与取值范围为 0 或者 1, 与  $\hat{y}_{ij}$  的取值范围并不相同, 即当  $y_{ij} = 1$  时, 表明输入的面孔图像来自于同一个身份, 当  $y_{ij} = 0$  时, 表明并不是同一个人面孔图像。本文设计的两者的度量关系为欧式距离, 主要原因是该类设计比较适用于反向传播的计算, 且应用较为广泛。 $\lambda$  是损失信号放大的参数, 输入的特征是做过模型的归一化, 因此监督信号的强度将处于较低水平, 则需要通过放大控制信号的方式来增强面孔特征的约束性。

如图 3.3 所示, 验证网络是两个单通道卷积神经网络耦合而成, 中间添加了验证层以达到模型调优的目的。在模型设计方面, 特征信号也将设计为两条视觉通路, 其中一条视觉通路将会输入证件照的图像, 另一条视觉通路将会输入多姿态的同一尺度的场景内面孔图像。在模型训练方面, 模型每次迭代也分为两个阶段, 第一阶段是验证层的损失信号作为模型的评判标准, 第二阶段是分类层的损失信号作为指标, 实验结果表明经过第一阶段的预训练, 识别网络层的损失值将会趋向于较小的数值, 因此在此阶段主要发挥作用的是验证层的损失函数, 识别网络层的增加为了保证模型训练的稳定性。

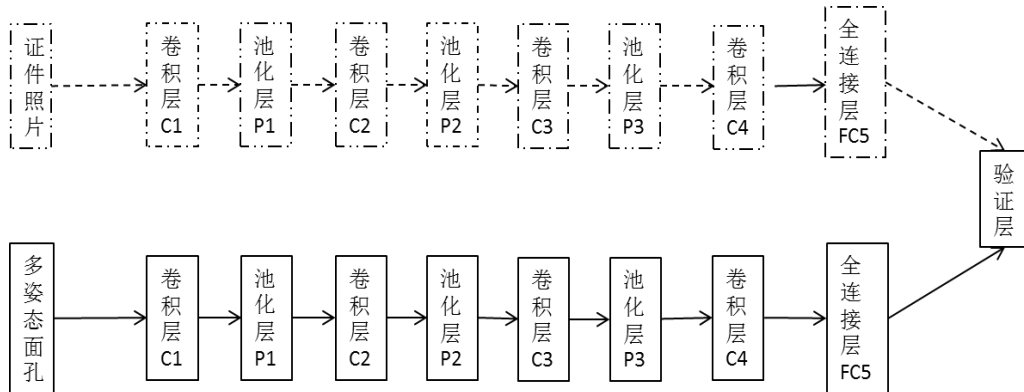


图 3.3 验证增强网络

### 3.3.3 面孔验证分类器

上述的叙述中提到，面孔验证网络中设计了二元验证的分类器，其目的在于增强对提取的面孔特征约束性。事实上，经典的二元分类器并不适合神经网络反向传播算法，因而本文设计了用于监督信号增强的二元分类器。为了模拟视觉识别的类别特性，本章节将采用基于面孔识别类内差异和类间差异的联合贝叶斯人脸理论。面孔特征的差异  $x$  主要由类内差异  $\mu$  和类间差异  $\varepsilon$  两部分组成，类内差异是同一个人不同姿态下提取的差异，类间差异则是不同身份的人本身存在的差异，如式(3.12)所示。

$$x = \mu + \varepsilon \quad (3.12)$$

为了更好地说明，假设双通道的系统提取的面孔特征为  $f_i$  和  $f_j$ ，则两者差异  $\Delta$  表示  $f_i - f_j$ ，鉴别两张面孔为同一个面孔的概率  $S(\Delta)$  如式(3.13)所示。

$$S(\Delta) = P(H_I | \Delta) = \frac{P(\Delta | H_I)P(H_I)}{P(\Delta | H_I)P(H_I) + P(\Delta | H_E)P(H_E)} \quad (3.13)$$

其中  $H_I$  表示这两张图片来源与同一个身份， $H_E$  表示这两张图片来源于不同身份。为了近似计算，上述公式简化如下(3.14)所示。

$$r(f_i, f_j) = \log \frac{P(\Delta | H_I)}{P(\Delta | H_E)} \quad (3.14)$$

其中  $r(f_i, f_j)$  作为参数来判定输入的面孔是否来源于同一个人。若  $r(f_i, f_j)$  的数值极大，即  $P(\Delta | H_I)$  值较大， $P(\Delta | H_E)$  值较小，则判定为来源于同一个人；反之，若  $r(f_i, f_j)$  的数值极大，即  $P(\Delta | H_I)$  值较小， $P(\Delta | H_E)$  值较大，则判定为面孔身份不一致。但是，若两者同时很小，或同时很大，这会造成一个分类错误，结果将会无法判定。

## 3.4 实验过程以及分析

本实验主要分为两个阶段，第一阶段是探求双通路的面孔验证网络的有效性，以识别率作为标准，验证增加监督信号对提取的面孔特征有一定的约束性，有助于提高面孔验证的准确率。第二阶段是探求面孔验证放大信号对面孔验证的影响，验证合适的  $\lambda$  的取值将会对面孔验证有促进作用。

Theano 是为深度学习中处理大型神经网络算法所需的计算而设计的，其表达式是在抽象的意义上定义，编译和后期是用于计算。同时其能够获取设计的网络结构，并使之成为非常高效的代码，在 CPU 或 GPU 上尽可能快地运行。考虑其快速实现和易于模型开发的原因，本章节实验将在 Theano 框架上完成。

实验对象来源于 AT&T 数据库，包含有来自不同状态的 40 个人的 400 张面部图像。每个人的面部图像为 10 张，且这 10 张图像都是互不相同的。他们的位置、旋转、尺寸和表达方式各不相同。方向的变化已经通过旋转来实现。且每个人的样本中，他们的面部表情（睁眼、

闭眼，微笑、不笑等）也是各不相同的。每个图像将转化为  $256 \times 256$  像素阵列，且其灰度值介于 0 到 255 之间。AT&T 数据库面部图像的部分样本如图 3.4 所示。



图 3.4 AT&T 数据库的面部图像部分样本

### 3.4.1 双通道网络结构的性能测试

为了验证双通路模型对面孔特征提取有提升作用，本实验将 PCA 面孔特征提取、单通路的特征提取和本文提出的双视通路特征提取对比，以面孔验证的准确率为标准，采用联合贝叶斯分类器在面孔验证数据集上对比结果。

表 3.1 卷积网络的参数

层数	Conv1	Conv2	Conv3	Conv4
卷积核大小	3x3	3x3	3x3	3x3
池化	2x2	2x2	2x2	2x2
卷积核的数量	20	40	60	80

考虑到实验条件和模型复杂度因素，本文将底层视通路设置为级联结构的卷积神经网络，具体参数如上表 3.1 所示，同时为了模型快速收敛，卷积核内部嵌入批量规范层。为了让实验结果具有代表性，利用 S 折交叉验证的方法计算平均准确率。首先，ORL 面孔数据集基于身份类别按 4: 1 的比例划分，其中 4 份为训练数据集，1 份为测试数据集。训练数据集将会用于模型特征提取模块和面孔验证分类器分类器。测试数据集将样本按照随机选择的原则，算法产生 50 组同类面孔和 50 组异类面孔组。然后，重新划分数据，如此迭代重复 5 次，最后在 5 次评测中平均测试误差。

表 3.2 各个算法的样本测试情况

模型	准确率 %
PCA	83
单通道模型	90
双通道模型	98

从上表 3.2 可以得到, 基于 PCA 的人脸验证方法准确率为 83%, 单通道为 90%, 本设计的双通路模型的面孔验证的准确率达到 98%, 比单通路的面孔识别模型的准确率有了一定的提升, 说明增加监督信号的卷积神经网络模型确实能够对面孔特征具有约束性, 有利于面孔图像的特征, 提高了面孔验证的准确率。各卷积层输出特征图谱如图 3.5 所示, 随着层数的增加, 特征图谱之间的差异性也在增加。卷积层 C1 输出的 20 个特征图谱已经提取了一定的面孔特征, 但是卷积核内部的权值共享的机制存在, 相对的神经元相对位置关系没有发生变化, 因此可以看出特征图谱中还能初步看出面孔的轮廓特征。最大池化操作本质上是原来面孔图像上多个像素点映射到单个像素点上, 类似图像模糊滤波, 因此后续的卷积层输出的特征图谱较为模糊化, 从而卷积神经网络具有一定的旋转不变性。

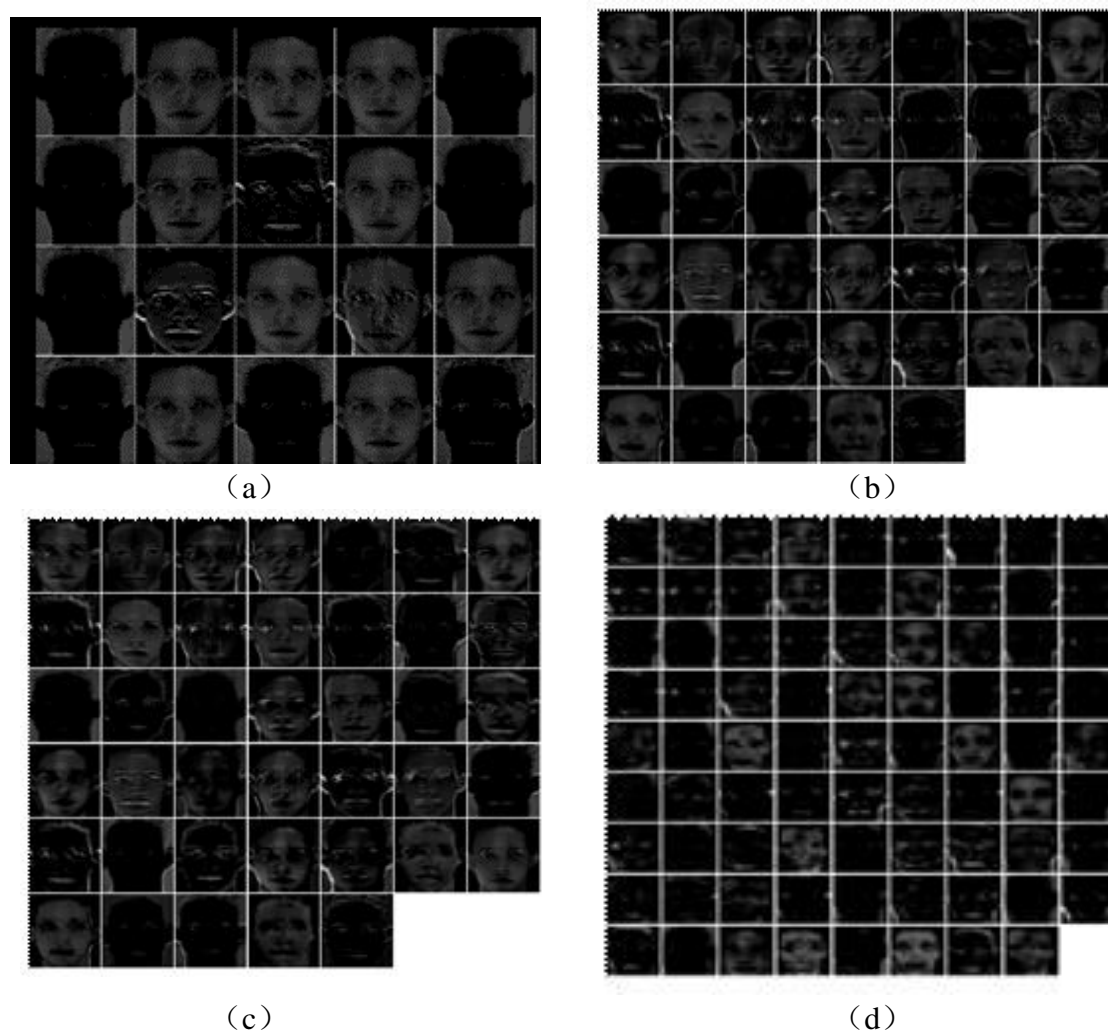


图 3.5 各层输出特征图谱。(a) 第一层卷积层输出特征图谱; (b) 第二层卷积层输出特征图谱; (c) 第三卷积层输出特征图谱; (d) 第四卷积层输出特征图谱

为了进一步探究监督信号对于前馈神经网络的提取的特征有约束作用, 笔者选用一组测试数据集 (包含两个身份的面孔样本) 作为单通路和双通路的面孔识别模型的输入, 输出的多维特征, 选取具有代表性的两维度映射到二维度空间平面内。如图 3.6 所示, 从左至

右分别为单通道和双通道的特征分布，从中可以观察双通路的面孔特征分布比单通路聚拢，说明监督信号对特征提取有益的作用。

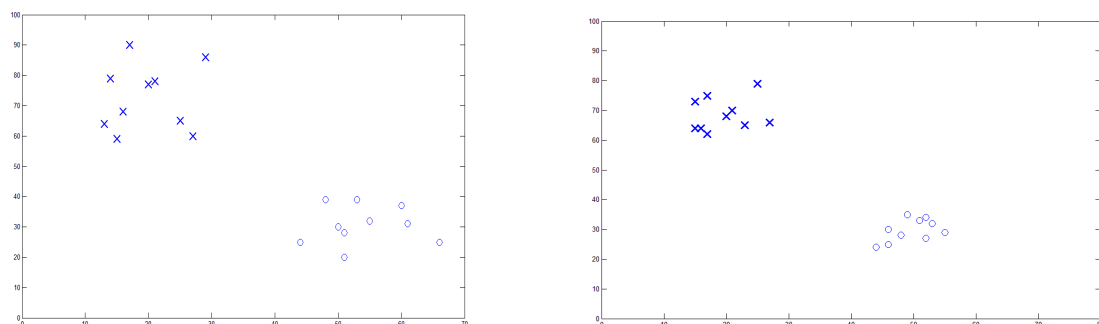


图 3.6 特征点分布图

### 3.4.2 控制分类信号和监督信号之间的相关性

为了调研分类信号和监督信号的关系对特征学习的作用，本实验将参数  $\lambda$  的取值范围为 0 到无穷大。当  $\lambda=0$  时，无监督信号作用，即是单通路的面孔识别模型；随着  $\lambda$  的增加，监督信号将会逐渐影响模型训练；理论上，当  $\lambda$  趋向于无穷大时，即模型中仅仅只有监督信号的作用。对于实验的部分参照， $\lambda$  的参数设置从 0.1 到 2，监督信号的验证强度也逐渐增强，以识别率和召回率作为标准，实验结果如表 3.3 所示。

表 3.3  $\lambda$  参数设置对面孔验证的影响 %

$\lambda$	0	0.2	0.4	0.6	0.8	1	1.2	1.4	1.6	1.8	2	100
识别率	90	93	94	95	96	98	95	94	93	93	93	93

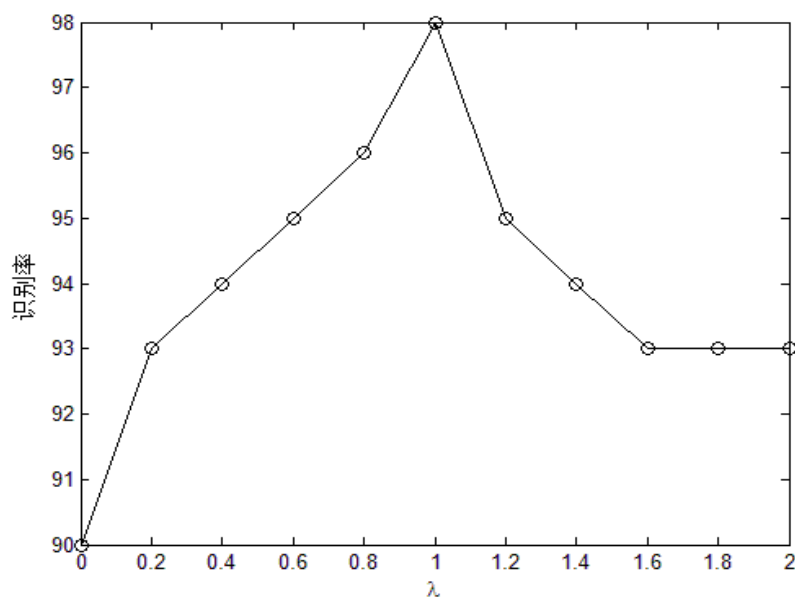


图 3.7  $\lambda$  参数设置对面孔验证识别率的关系图

从图 3.7 中折线趋势可以发现监督信号和识别信号之间相互作用。在  $\lambda=1$  时，模型的识别率表现最优；当  $\lambda=0$  时，识别的准确率为 90%，此时模型内部只有识别信号的作用；随着参数  $\lambda$  的增加，监督信号也将会增强，面孔特征约束性增强，从而类内方差减小，面孔识别的准确率提升；当  $\lambda=100$  时，识别的准确率为 93%，增强网络的约束性没有增强。

### 3.5 小结

本章节主要对提出的基于视觉信息流的双通路面孔验证新方法的研究。研究底层视觉通路的内部机理，构建特征提取网络，并对其参数训练进行一定的阐述；为了对面孔特征有约束作用，引入了双视觉通路的概念，构建监督信号和识别信号融合的特征增强网络；最后，分析面孔特征误差的特点，构建联合贝叶斯的面孔验证分类器，最后达到面孔验证的结果。实验结果表明，笔者将双通路的面孔验证模型，与原系统对比，可以发现引入监督信号的面孔验证模型可以提高面孔验证的准确率。

## 第4章 基于视知觉皮质通路的正立和倒立面孔混合识别

### 4.1 概述

面孔识别是人类与生俱来的一种自然能力，神经生理和心理学的研究表明，面孔识别所对应的大脑皮层激活区域完全不同于物体或文字识别。对面孔识别机制的研究，一方面将有助于促进对于大脑功能机理的理解，另一方面将为图像理解或模式识别技术带来一些新的思路。

由于正立面孔识别是较为常见的情形，因此人们通常以正立面孔识别作为研究问题，认为面孔的尺度或角度问题，完全可以通过对图像的预处理或校正过程进行解决，对于工程技术人员来说，这是一种自然而然的观点。所以人们将注意力集中于正立面孔的特征提取和识别上，陆续提出了一些有效的识别方法，譬如经典的主成分分析方法<sup>[17][18]</sup>、基于监督信息的线性判别方法、基于简单局部纹理信息的非线性辨别方法，近年来陆续也有基于卷积神经网络的方法，例如基于多模型融合的 DeepID 方法<sup>[43][44][45]</sup>，利用监督信号和识别信号结合的方法提升网络的鲁棒性；还有 FaceNet 神经网络模型<sup>[46]</sup>，可实现模型训练的 End-to-End，避免传统方法的多阶段问题；另外神经网络度量函数的优化工作也受到了关注<sup>[47][48]</sup>。上述方法在面孔识别的具体应用中，尤其是卷积神经网络所代表的深度学习多模型融合方法，取得了较好的识别效果。但必须要说明的是，越来越多的神经工程实验结果表明<sup>[49][50][51]</sup>，正立面孔的识别可能依赖于一种组件架构的模式，此时整体信息要大于各局部特征之和；而倒立面孔的识别，并非显著依赖于上述整体信息的表征，反而更可能是一种以眼睛和口鼻嘴为信息源的局部特征表征。面孔识别机理存在着上述两条分别反映整体和局部特征的视皮层感知通路已经被普遍接受，但其中大多数观点认为这两条通路或者两个系统是独立运行的，因此在实践中采用一种分而治之的研究策略。而本文认为正立和倒立面孔绝不是视觉信息上简单倒置的关系，传递处理整体和局部特征的两条视觉通路虽然分别在正立和倒立面孔识别上起了决定性的作用，但它们并不是孤立存在的，两条通路所刻画的面孔信息应该是一种互补式的关系。在利用整体轮廓信息来识别面孔时，不能缺失面孔上具有表征能力的局部区域对面孔识别性贡献。

基于以上分析，本文将考虑多局部特征对识别率的贡献，以视知觉皮层通路作为研究对象，研究左右半球协作机制，构建基于局部和整体信息的双通路面孔识别系统。本文保留原有基于轮廓信息的网络层对识别率的贡献，增加了面孔表面具有表征性的局域纹理信息，设计并且实现了基于局部区域提取的池化神经网络结构，构建了多局部特征融合 MLPB (Multi-Local Pooling Blending) 的网络结构，通过模拟左右半脑对整体和局部特征的协作关系，



提出一种融合整体和局部信息的预测函数，给出了一种针对正立面孔和倒立面孔混合识别的新模型。

## 4.2 基于视知觉皮质通路的双系统面孔混合识别

在视觉感知机制中，视觉通路是从视网膜开始，经过外膝体、初级视觉皮层和高级视觉皮层，从而实现识别视觉信息的功能。著名的倒立效应实验<sup>[52][53][54]</sup>表明，人的视觉系统不仅仅能够识别正立面孔，同时在不加特殊训练的条件下，对倒立面孔具有一定的识别准确率。其原因是视觉系统在加工倒立面孔时，并不只是利用高层次的整体轮廓信息，还必须转向一种更具分析式的、以部分分析作为基础的加工模式，即面孔识别系统不仅仅只存在单条视觉通路的加工方式，至少存在两条及以上的视觉信息加工。

但是，当前基于卷积神经网络的面孔识别模型并没有关注到倒立面孔的识别问题，因为模型仅仅将图像底层的图像纹理特征融合，并转化高层次图像特征，最终是根据高层次的特征来激活函数进行分类，并没有利用到局部分析式的信息特征。

故本文提出一种基于视知觉皮质通路的双系统面孔混合识别模型，其由底层视觉通路和高级视觉皮层两个网络构成。如图 4.1 所示，底层视觉通路主要由卷积层神经网络(CONVNET)构成，对输入的正立和倒立面孔(INPUT)做图像预处理以及特征提取；高级视觉皮层是以左右半球协作机制为核心，并将整体轮廓信息和局部轮廓信息融合预测面孔类别。全局特征直接连接基于整体轮廓信息的全连接层，而局部候选区域将经过 MLPB 融合后，连接基于局部纹理信息的全链接层。SOFTMAX 将会接收上层网络的预测信息，根据权值关系达到同时满足正立面孔和倒立面孔混合识别的目的。

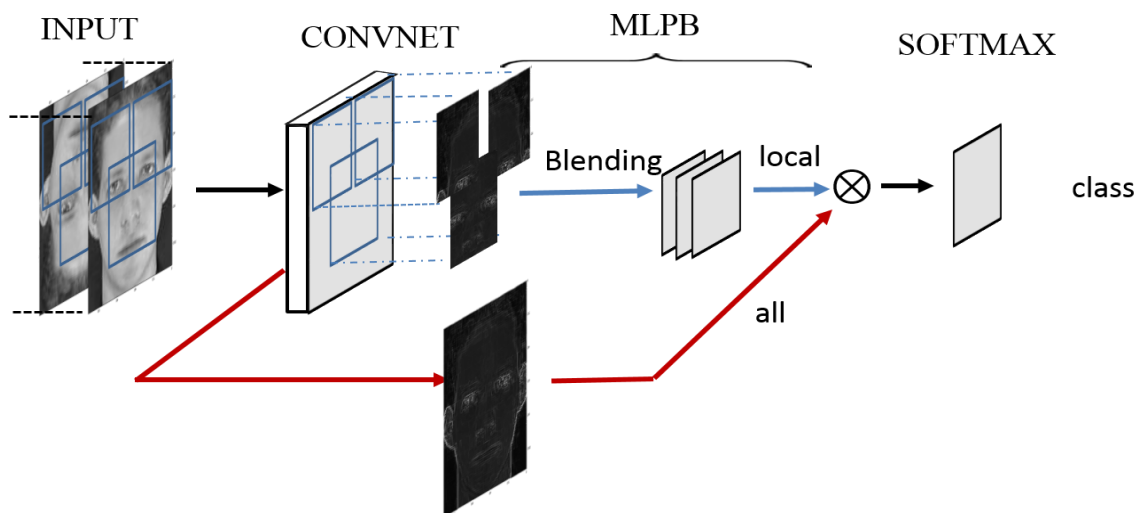


图 4.1 基于视知觉皮质通路的双系统面孔识别模型

## 4.3 算法原理

### 4.3.1 底层视通路

在生理学上，底层视觉信息传递通路有两条，一条是从视网膜-外侧膝状体-初级视皮层 (Primary Visual Cortex, V1) 的视觉，另一条则是视网膜-上丘脑-视皮层。然而绝大多数视觉信息预处理都是由前一条视通路进行传递和处理的，后一条通路主要是确定视觉信息的位置。本文中的面孔图像将通过人脸矫正的方式来将人脸面部的眼耳口鼻等局部区域放在同一位置，以实现后者视觉通路上的功能。

神经生理学表明，眼球接受的刺激，经过预处理和加工后，在 V1 层输出纹理特征。由于卷积神经网络具有杂质信息有效过滤和纹理特征提取能力，故本文将底层的网络设置为卷积神经网络，其基本结构包括特征提取层和特征映射层两层，每个特征提取层后将会紧跟着一个取局部最大和二次提取的特征映射层，该提取结构将会使模型对输入面部图像的抗畸变能力。由于底层的卷积层对面孔表面的纹理特征较为敏感，后续的特征图谱将会是上层特征图谱的一种非线性组合关系<sup>[55][56]</sup>。因此随着网络层数的加深，卷积核输出的特征图谱也将越多，模型对图像的描述能力也越强，从而能够更有效表达面孔特征。

### 4.3.2 局部池化层

传统视觉系统的设计思想是将底层提取特征导入基于整体轮廓信息的单条视觉通路面孔识别模型，因此将不可避免地缺失必要局部特征的参与<sup>[32][33][34]</sup>。为了提取面孔表面局部特征，本文提出了一种多局部特征融合的网络结构，即 MLPB，该结构能在面孔图像经过多层卷积层之后，在特征图谱上提取对应于输入面孔图像表征区域位置的特征，并且将各局部特征融合。局部特征提取示意图如图 4.2 所示。

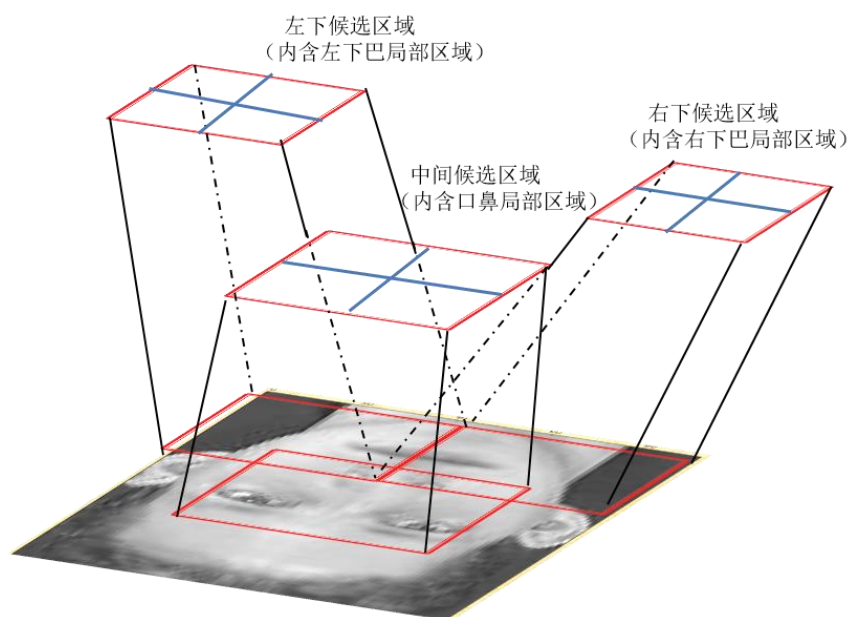


图 4.2 面孔表面局部区域选取图

MLPB 层主要对应于面部三个区域局部区域的特征提取和融合。值得说明的是，为保证正立和倒立面孔表征区域的差异性，正立面孔的选择包含左眼在内的左上区、包含右眼在内的右上区和包含口鼻在内的中间区域作为主要特征区域，同时倒立人脸的特征区域以左下区、右下区以及中间区为主。为有效地转换局部特征，本文提出了基于局部区域提取的池化神经网络层概念，其基本思想将大小为  $L \times L$  的按照  $N \times N$  的步长做切割成小模块，并对每个小模块做最大池化操作（max pooling），输出  $M \times M$  的特征矩阵。 $M$  的计算表达式如式(4.1)所示。

$$M = (L + 2p - k) / N + 1 \quad (4.1)$$

其中  $p$  表示为输入候选区域所增加的像素点个数， $k$  表示为最大池化操作的滤波器大小， $N$  表示为指定滤波器的步长参数。局部区域不同，则其各参数的大小选择也将会不同。但是在下一阶段特征融合时，为简化计算的目的会要求所有输出特征矩阵的长和宽保持对应一致。

输出特征图谱上的每个像素点对应一个最大池化神经元，则每个神经元的操作如式(4.2)所示。

$$f^k(i, j) = \max_{0 \leq m, n < N} \{x^k(a_{roi} + i \cdot N + m, b_{roi} + j \cdot N + n)\} \quad (4.2)$$

其中  $f^k(i, j)$  表示池化后的特征矩阵上的像素点， $x^k$  表示第  $k$  个输入的特征矩阵， $a_{roi}$  和  $b_{roi}$  为候选区域  $Roi$  的左上角点在全局特征矩阵内的坐标。因为反向传播的计算与一般的卷积池化的操作相同，故此处不做赘述。基于局部区域提取的池化神经网络层相比于直接获取原始特征区域的方式将有利于获取更加细节的纹理特征，减少模型参数规模，从而降低模型计算的复杂度，防止模型过拟合的现象发生。

#### 4.3.3 多特征融合

MLPB 网络在局部特征池化之后，需要将其融合。以图 4.2 所示的倒立脸为例，因左下候选区域、右下候选区域和中间候选区域输出的特征图谱之间成相互独立关系，故将三者做连接操作，其数学表达式如式(4.3)所示：

$$F_{Local} = F_{Lower\_L} \otimes F_{Lower\_R} \otimes F_{Center} \quad (4.3)$$

其中  $\otimes$  表示的是连接操作， $F_{Lower\_L}$ 、 $F_{Lower\_R}$ 、 $F_{Center}$  和  $F_{Local}$  分别表示为左下、右下、中间候选区域的特征图谱以及特征图谱融合后的特征图谱。假设经局部池化后， $F_{Lower\_L}$ 、 $F_{Lower\_R}$  和  $F_{Center}$  数据大小分别为  $n \times c_{Lower\_L} \times w \times h$ 、 $n \times c_{Lower\_R} \times w \times h$  和  $n \times c_{Center} \times w \times h$ ，其中， $n$  代表模型进行批量梯度下降时输入面孔图像的数量， $c$  代表各局部特征的通道， $w$  和  $h$  代表特征图谱的宽和高，则  $F_{Local}$  数据大小  $n \times (c_{Lower\_L} + c_{Lower\_R} + c_{Center}) \times w \times h$ 。

#### 4.3.4 预测函数

神经学专家通过面孔失忆症患者和整合性失忆症患者相关实验证明<sup>[57][58]</sup>，存在物体识别的双加工模式，即与左半球相关的分析加工模式，与右半球相关的整体加工模式。同时左右

半球之间存在高层次特征的协作关系：在识别正立面孔时，右半球为起主要作用，左半球起辅助作用；在识别倒立面孔时，将以左半球为主，右半球的特异性加工系统将会失灵。

传统神经网络的预测函数只有考虑整体信息，并没有局部特征模块融合。因此本文提出一种融合整体和局部信息的预测函数，根据输入的特征输出可能性最大的面孔类别，如式(4.4)所示。

$$Class(F) = \arg \max_i (W_i^1 \cdot F_{all} + \lambda \cdot W_i^2 \cdot F_{local} + b_i) \quad (4.4)$$

其中  $Class(.)$  表示预测函数输出的面孔的类别， $W_i^1$  为基于整体轮廓信息的权值矩阵，下标表示矩阵的第  $i$  行，对应于第  $i$  类面孔； $W_i^2$  为基于局部信息的权值矩阵，下标含义同前； $F_{all}$  表示最后一层卷积层输出的全局特征图谱。 $F$  代表最后一层卷积层数输出的特征图谱集合，包括整体和局部的特征图谱， $\lambda$  代表着左右半球的相互协作关系，其数值大小代表着全局信息和局部信息的比值关系，则在正立面孔识别时，因为全局信息的比重较大，即  $\lambda < 1$ ；在倒立面孔识别时，需要强化局部特征信息，即  $\lambda > 1$ 。

## 4.4 实验结果及分析

本实验主要分为两个阶段，第一阶段是比较正立面孔识别模型和混合识别模型，以识别率作为标准，分别在正脸面孔和倒立面孔上做测评，验证倒立面孔识别通路并非孤立存在的，它将有助于提升正立面孔识别性能。第二阶段是探究单一或混合的训练样本和不同的预测函数参数对面孔识别的影响。

本实验应用的神经网络框架将会采用伯克利的团队开发的 CAFFE 开源框架，自框架的开源以来，深刻受学术界和工业界的喜爱的使用。该框架基于 C++/CUDA 架构，支持命令行，Python 和 Matlab 借口；可以在 CPU 模式和 GPU 模式之间可以相互切换，同时在 GPU 并行运算模式可以加快模型训练速度。本文将在 CAFFE 上定义编写自己的模型、最优化设置以及初始化权值，训练结果将会以 Jupyter Notebook 形式再现，并且将用于网络框图、各层可视化和训练过程。

### 4.4.1 单一训练集情况下的识别性能

为验证局部信息将有利面孔识别率的提高，本文将 PCA、LDA、正立面孔识别模型 DeepId 与混合识别模型做识别率上的对比，其中本文模型在保留简化后的 DeepId 网络层和卷积核参数的基础上，添加了基于局部纹理分析的 MLPB 网络。由于 AT&T 数据库中的面孔均是正立的面孔，不含有倒立面孔，从而需要对训练数据和测试数据构造。首先，该数据库包含 40 个人，每个人有 10 张不同姿态的面孔图片，不做处理任何处理，从每个人的库中随机选择 8 张面孔作为训练数集，余下的 2 张作为测试数据。因此正立面孔的训练数据大小为 320 张，正立面孔的测试数据集为 80 张。其次，为了保证倒立样本和正立样本一致性，仅将上阶段正立的训练和测试数据集的面孔上下旋转 180，不做图像大小和样本标签等任何改动，从而构成倒立脸的训练和测试数据集。本实验将 320 张正立面孔数据集训练模型，然后分别在正立和

倒立的测试数据集上做模式识别率的评估，作为对比，将倒立面孔数据集作为训练集，同样也在上述的两类测试集上做模型评估。实验结果如表 4.1 和表 4.2 所示。

表 4.1 算法的平均识别率和测试时间比较

模型	平均识别率 %	单张面孔测试时间 ms
PCA	88	345
LDA	92	905
DeepId	98	1440
混合识别模型	100	1560

从表 (4.1) 中看出，由于模型复杂度的原因，混合识别模型在 I7 6700k CPU 硬件条件下，比传统方法的识别时间略长，但是在保证一定实时性条件下，本文算法依然具有较高的识别率。为体现 GPU 和 CPU 运算效率的差别，混合识别模型在基于 GTX 970 GPU 4G 显存下测试，单张图片测试时间则仅仅需要 126ms。

表 4.2 在单一训练集情况下的算法识别率对比 %

训练方式 模型	正立脸训练		倒立脸训练	
	正立面孔测试	倒立面孔测试	正立面孔测试	倒立面孔测试
PCA	87	7	8	87
LDA	92	6	10	92
DeepId	98	7	14	98
混合识别模型	100	9	14	100

如表 4.2 所示，在单一训练样本的情况下，也就是训练样本要么全部来源于正立脸数据集，要么全部来源于倒立脸数据集，这时进行与训练样本集类别相符的测试，能够取得较好的结果。其中 PCA 的人脸识别方法识别率为 87%，LDA 为 92%，DeepId 为 98%，如果在 DeepId 模型的基础上添加局部特征信息的作用，也就是本文提出的混合模型方法，在测试数据集上的成绩将进一步提高到 100%。这也说明了 MLPB 结构的有效性，面孔局部特征有助于面孔识别精度的提高。表明在正立和倒立面孔识别时，局部和整体轮廓信息的视觉通路并非相互独立的，局部视通路将会在一定程度上强化人脑对正立面孔的识别能力。

观察表 4.2，还可以发现用正立面孔训练集来训练识别的模型，在倒立面孔测试集上的成绩分别为 7% 和 9%，反之的成绩均为 14%，应该说此时识别完全失效。本文将一张正立面孔和对应的倒立面孔导入用正立面孔训练后的混合识别模型中，对比每一层输出的特征图谱和滤波器，如图 4.4 所示。将倒立面孔的特征图谱上下旋转 180，与正立面孔的特征图谱做差值计算，可发现结果接近于 0。接下来的各层卷积层也采用类似处理方式，也能得到相同的结论。说明底层的卷积核具有对称性，对面部的纹理特征响应具有一致性。另外也可以将正立面孔训练 MLPB 的响应图和倒立脸对比，也表明纹理响应具有一致性。上述分析验证了模型

识别性能差的原因应该是由网络末端的全连接层所引起的。因此在模型中，直接影响模型最终分类结果将会是高级视觉皮层。该类情况符合了生物特性，即在视觉通路处理过程中，视网膜、外膝体以及初级视觉皮层对于图像都是进行相同的预加工过程，它们对正立的面孔和倒立的面孔预处理是一致的。因此本实验一方面说明底层视通路上卷积核具有对称性，另一方面说明底层视通路在正立和倒立面识别上并不具有本质上的差异性。

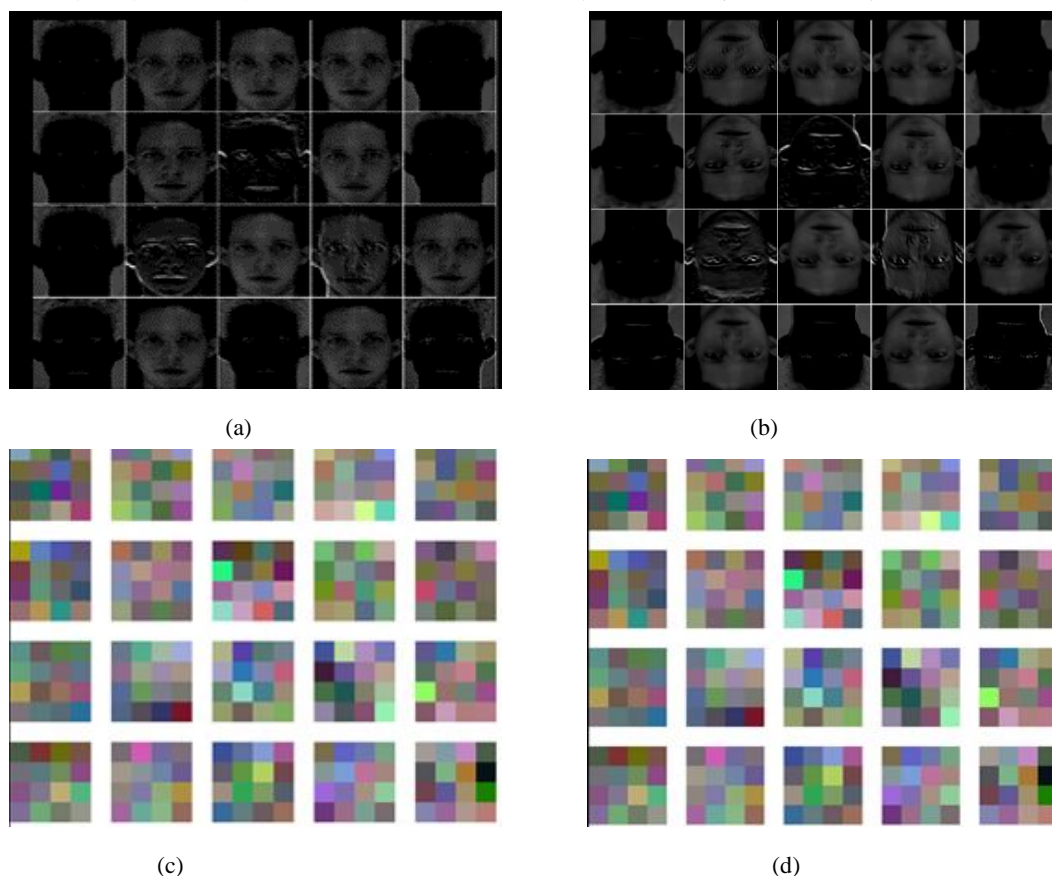


图 4.4 正立和倒立面孔在第一层卷积层的输出特征图谱和滤波器。(a) 正立面孔的第一层卷积层输出特征图谱；(b) 倒立面孔的第一层卷积层输出特征图谱；(c) 正立面孔训练双系统识别的第一层滤波器；(d) 倒立面孔训练双系统识别的第一层滤波器

#### 4.4.2 混合识别模型中的参数优化

先采用单一的训练样本集对混合识别模型进行训练，然后再利用倒立面孔的训练样本集对模型进行训练，探讨整体和局部信息的比重关系对于识别性能的影响。实验中随机选取 320 张正立面孔数据集来训练混合识别模型，然后固定前面四层卷积层的参数以及全局结构的全连接层参数，即接下来的训练时上述 5 层神经网络参数将保持不变，最后用 320 张倒立面孔混合数据集训练，此时模型中只有 MLPB 网络模块的参数才会更新迭代，在损失函数达到局部最优值时停止训练。

为了验证整体和局部特征对面孔识别的作用，参数  $\lambda$  分别设置从 0.1 到 1，局部特征识别率的贡献的比重也逐步增大，在正立和倒立面孔上的测试结果如表 4.3 所示。



表 4.3  $\lambda$  参数设置对面孔识别率的影响 %

$\lambda$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
正立面孔	100	100	100	100	99	98	97	95	94	93
倒立面孔	21	23	30	33	39	46	57	65	75	80

从表 4.3 中反映出全局特征和局部特征在面孔识别率上形成相互干扰。全局特征的全连接层是用正立面孔来训练的，因此正立面孔图像作为输入时，该层输出的结果有益于对正立面孔的识别率，然而 **MLPB** 模块的参数是由倒立面孔数据训练，此时该模块将会失效，从而对正立面孔的识别率造成干扰。如果  $\lambda$  数值增大，即局部特征的贡献比重增大，**MLPB** 模块造成的不利干扰也将增大，从而导致正立面孔的识别准确率将会降低。反之，倒立面孔作为模型的输入导致全局模块失灵，一定程度上影响倒立面孔结果的判定。

从视觉感知机制的角度看来，人类对于倒立面孔的识别，并没有经过特异性训练，即从出生开始用大量正立面孔来做辨认，当然偶尔会有少量的倒立面孔需要辨认。因此即使训练样本少，也并不妨碍人类视觉系统对倒立面孔的识别率。生物学实验证明正常人在正立和倒立脸的识别率分别为 94% 和 82%，即存在人对于倒立脸的区分度不如正立脸的情形。

本文发现倒立面孔作为混合识别模型的输入时，因整体的信息仅仅与正立面孔识别相关，此时不考虑局部特征的模型预测准确率将为 0.1 左右，较低的预测准确率意味着模型在一定程度上处于随机预测的状态。同时本文再把最后一层模型输出的预测率提取后比较，发现模型对每一类面孔的置信度差异接近于 0。因此选择合适的  $\lambda$  将有效地避免整体信息对倒立面孔预测的干扰，此时全局信息对倒立面孔识别贡献将会接近于 0，从而间接地提高了局部特征信息的权重。在正立面孔作为输入时，模型识别率将会提高，即局部信息将会对整体信息产生有益的补充作用。综上考虑，结合视觉感知机制特性以及模型计算的简便性，当  $\lambda$  参数设置为 0.6 具有较好的性能。

#### 4.4.3 样本集对识别性能的影响

为提高混合识别模型在两类测试数据集上的识别率，对模型训练方式改进，首先，用正立面孔样本做模型预训练，并固定卷积和全局特征的参数，其次，不同于上述实验仅用倒立面孔集训练 **MLPB** 模块，而是利用正立和倒立面孔混合样本集训练，其目的是使 **MLPB** 模块同时具有对正立和倒立面孔局部特征的区分能力。

从表 4.4 中发现改进的训练方式，在正立面孔识别率可以达到 100%，有 2% 准确率的提高；同时该模型在倒立面孔识别率达到 93%，较之前仅仅用倒立面孔训练的模型，有了明显的提高。反观第一栏中直接用混合样本训练，识别率非常低，说明这样的训练方式将会导致模型内部结构混乱。训练成功的神经网络面对同一个人的不同人脸输入时，卷积层输出的特征应具有较高的相似性。然而相对于正立面孔来说，倒立面孔经过卷积层后的特征是倒置的，

没有特殊设计直接输入到全连接层，识别的人脸类别将与正立面孔不同。因此直接混合训练将导致预测函数识别混乱，从而损失函数不能收敛。为了避免特征倒置造成识别混乱的现象，本文提出的模型正立和倒立面孔选取的局部特征区域有所不同，在正立面孔提取的局部特征是左眼为主的左上区、右眼为主的右下区和口鼻为主的中间区，倒立面孔提取的局部特征是左下巴为主的左下区、右下巴为主的右下区和鼻子嘴巴为主的中间区。模型参数训练时，同一个人的正脸和反脸的局部特征保持着较高的区分性，使得 **MLPB** 层具备对正立和倒立人脸的区分能力，从而模型的识别准确率保持较高的水平。

表 4.4 不同的样本集训练方式对面孔识别率的影响 %

训练方式 识别的面孔	混合样 本直接训练	正立面孔预训练， 倒立再训练	正立面孔预训练， 混合面孔再训练
正立面孔	15	98	100
倒立面孔	15	46	93

## 4.5 小结

本文通过卷积神经网络对输入图像特征的纹理敏感性的特点，实现对正立面孔和倒立面孔的预处理，考虑到局部的参与识别的生物学机制，设计了基于局部特征的池化神经层，实现了多局部特征融合的网络结构，同时引入了高级视觉皮层中左右半脑相互协作的关系，提出一种融合整体和局部信息的预测函数，重点考察了训练数据因素和局部特征与整体特征的关系因素对正立和倒立面孔识别率的影响。在结果的讨论中，本文提出的混合识别模型比传统的面孔识别模型在本文构建的数据集上的识别率上有了一定程度的提高，最后实现了正立和倒立面孔混合识别的目的。本文提出的正立和倒立面孔混合识别，为面孔识别技术提供了新的研究思路，探讨多视通路融合如何在高级视皮层的图像理解与视觉认知中发挥作用。



## 第5章 基于层次信息融合的多尺度面孔识别新方法

### 5.1 概述

视频摄像头下采集的人脸图像因为远近的关系存在尺度大小的差异性。然而面孔特征提取模型会因为模型的训练和网络结构设计等原因，要求人脸图像尺寸在输入固定大小的。但是当前遇到图像多尺度的输入情况，一般采用的技术手段是将具有差异性的面孔做剪裁或扭曲，但是这种人工改变图像尺寸的策略破坏了原图像的结构性和长宽比例，甚至可能造成面孔图像的信息缺失和增加信息噪声。研究表明<sup>[59]</sup>，即使经过裁剪或者扭曲操作固定了输入图像的尺寸，也意味着深度学习的神经网络模型训练的样本是固定尺寸，因此当目标变形或尺寸变化较大时，则需要重新预处理输入图像或者重现训练网络以适应新图像，否则整个网络或者其他任意尺度的子图像的识别准确率都会降低。然而固定输入尺寸的做法极大地限制了面孔识别在生物视觉感知机制上的发展。

视觉感知系统本身具备识别视野中多尺度目标物体。即使获取的图像尺寸非常小，视通路也同样具有检测识别能力，然而此时高级视觉中枢决策分析的依据并非依赖面部的轮廓信息，例如肤色信息等等。生理学和计算机科学多项研究表明<sup>[60]</sup>，视觉系统内部的神经元细胞将会关注特异化的局部特征，各种不同的特异化特征联觉合成范围较大的特征，如此反复构成了面孔特征。视觉系统内部的神经元将可能输出面孔表面一些局部信息，然而这些局部信息对与面孔特征表达具有重要的意义。

基于以上分析，本文将考虑各层次的局部特征对识别率的贡献，以视知觉皮层通路作为研究对象，研究视觉系统层级编码机制，构建基于层次信息融合的多尺度面孔识别新方法。本文保留基于轮廓信息的特征提取网络层；考虑到底层卷积对面孔特征表达的有益作用，构建基于局部和整体结合的多尺度池化的面孔表征结构 **MRPB (Multi-Roi Pooing Blending)**，输出固定大小的特征通道和特征图谱；然后，根据视觉皮层的决策机制，局部表征信息和全局表征信息融入到全连接层，通过神经元的激活特性判断面孔身份。该模型尽可能减少连续级联下采样操作中丢失的图像信息，拓宽图像特征表达的渠道能够满足适应不同大小的图像输入，同时对图像有一定的表征能力。

### 5.2 基于层次信息融合的多尺度面孔识别

层级编码假说表明，初级特征被整合起来将会形成一个可被知识单元识别的物体。虽然底层视觉通路不同层所提取的人脸特征并不相同，但是同样也可以作为目标识别的参考依据。但是当前基于卷积神经网络的面孔识别模型并不能适应图像多尺度输入的识别问题。不同尺度的图像输入，极为有可能导致部分神经元不能死亡从而提取的表征信息遭到破坏，同时由

于全连接层与卷积层提取后的特征图谱之间的权值矩阵没有自适应性，因此即使提取优秀的表征信息，也会导致系统紊乱。

基于上述分析，本文提出了一种基于层次信息融合的多尺度面孔识别新方法，通过对底层视觉通路进行 **MRLP** 操作以输出特征大小，并且特征将融入到决策分析层以丰富图像面孔的表征信息，从而有效应对视觉识别中图像多尺度输入问题。该整体的基本结构是其由底层视觉通路和高级视觉皮层两个网络构成，如图 5.1 所示，其中底层视觉通路主要由层级结构形式的卷积神经网络，对输入的多尺度的面孔图像预处理以及特征提取；高级视觉皮层是以融合的层级信息为基础的面孔识别。**MRPB** 结构本身设计原因使得各层特征图谱输出固定长度的特征向量，因此部分卷积网络层连接 **MRPB** 结构将会输出固定长度的全连接层，从而 **SOFTMAX** 激活函数根据接收的面孔特征信息表征输出面孔类别，以达到多尺度面孔识别的目的。

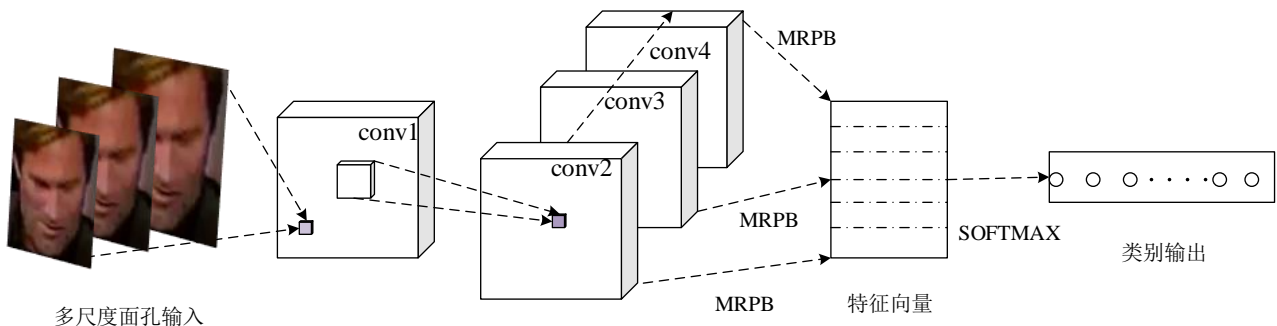


图 5.1 基于多层次信息编码的多尺度面孔识别网络结构

## 5.3 算法原理

### 5.3.1 层级采样

当识别目标进入我们的视野，视觉感知系统认为识别目标是一个整体而不是一开始就剪裁它成为几个不同大小的视图。同样地，大脑不太可能为了检测或定位目标而扭曲所有的目标使适应合适尺寸。因此，大脑在某些更深层处理任意形状的对象时，通过聚合来自前几层的经深度处理过的图像信息和特征。同时面孔图像小到一定程度的时候，并非仅仅能依靠基于整体的高层纹理特征作为整体描述。生物视觉感知系统也将会根据局部信息作为目标识别的参考依据，即当识别目标特征不明显时，初级视觉皮层等皮层提取的一些局部表征信息作为感知计算的参考依据。

结合上述分析，本文将提出了基于局部区域和整体区域结合的多尺度池化神经网络结构概念，**MRPB** 层主要池化采样区域如图 5.2 所示。值得说明的是，为了获取到更加细节的纹理特征，局部候选区域将会选择多尺度采样的形式，以高频率采样挖掘有效特征的目的，其中五个候选的区域包含左眼在内的左上候选区、包含右眼在内的右上候选区、包含口鼻在内的中间候选区域、包含下巴在内的左下候选区和包含下巴在内的右下候选区。

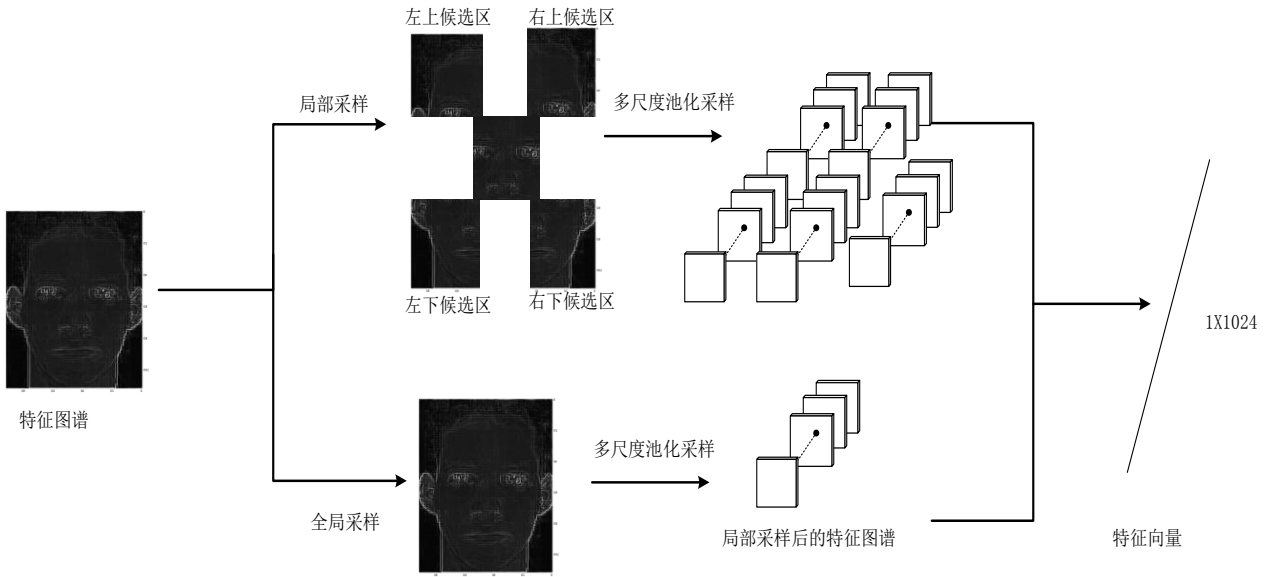


图 5.2 MRPB

多尺度池化采样的基本思想是特征图谱根据多个池化核进行下采样后特征融合，即是大为  $L \times L$  的特征图谱在做模型最大池化操作后，输出大小为  $N \times N$  的特征图谱。其中步长  $n$  与其他参数关系，如式(5.1)所示。

$$N = (L + 2p - k) / (n - 1) \quad (5.1)$$

其中  $L$  为候选区域的长度， $p$  表示为输入候选区域所增加的像素点个数， $k$  表示为最大池化操作的滤波器大小， $N$  表示为指定滤波器的步长参数。多尺度采样的特征图谱将转化为单维的特征向量，最后按序合成单维度的特征向量输入到全连接网络层，具体多尺度池化流程如图 5.3 所示。

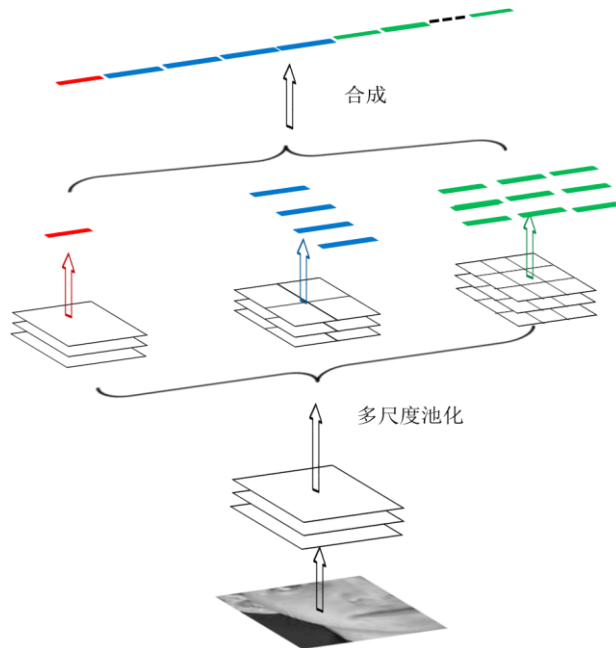


图 5.3 多尺度池化流程

### 5.3.2 多层信息表达的全连接层

特征分析层是卷积神经网络的卷积层-池化层和输出层的中间隐藏层，在神经网络分类决策中起到了关键点的作用，因为其融合了前馈网络中的高级特征，并且在一定程度上减少神经元参数和模型参数，为感知计算模型的决策分析做有益处的补充。但是如果输入的图像较小或者网络层次较为深时，池化输出的特征的维度较小。虽然如此小的维度是高度提炼的面孔图像的特征，但是难免造成信息丢失，造成模型欠拟合现象；如果图像过于大，全连接层将会信息冗余，一方面造成模型的参数过多，另一方面可能造成模型过拟合。

种种脑科学研究迹象表明<sup>[57][58]</sup>，视觉感知系统做视觉识别的决策分析时，并非依靠某个视觉回路特征作为判断的依据，而是多回路协同合作，因此各层次的神经网络特征图谱对面孔图像表征具有重要的意义。为了尽量弥补级联下采样造成的信息丢失，丰富面孔图像的特征信息，本文提出了基于层次信息融合的视觉感知计算公式，即输出的面孔的类别如式(5.2)所示。

$$\text{class}(X_{FC}) = \text{soft max}(X_{FC}) \quad (5.2)$$

其中  $\text{Class}(\cdot)$  表示预测函数输出的面孔的类别， $\text{Softmax}$  为函数的激活，其实本质上是 Sigmoid 激活函数在多分类问题上的推广，且具有良好的生物学特性。其中全连接层的输出  $X_{FC}$  如式(5.3)所示。

$$X_{FC} = \max(0, W_i^1 \cdot F_1 + W_i^2 \cdot F_2 + W_i^3 \cdot F_3 + b_i) \quad (5.3)$$

其中  $F_1$ 、 $F_2$  和  $F_3$  分别是 C2、C3 和 C4 卷积层多尺度池化后生成的列向量， $W_i^1$ 、 $W_i^2$  和  $W_i^3$  为列向量与全连接层 FC5 之间的权值矩阵，下标表示矩阵的第  $i$  行，对应于第  $i$  类面孔； $b_i$  是全连接层的偏差矩阵，函数激励是利用  $\text{Relu}(x)$  进行计算。值得注意的是，考虑到每层提取的面孔特征都可以作为单独面孔识别的能力，因此特征转化后做相加操作，而不是连接操作。如图 5.4 所示，相关实验已经证明第一层的卷积层输出的特征图谱内含有的杂质信息较多，容易造成模型的过拟合，因此并没有输出特征列向量。

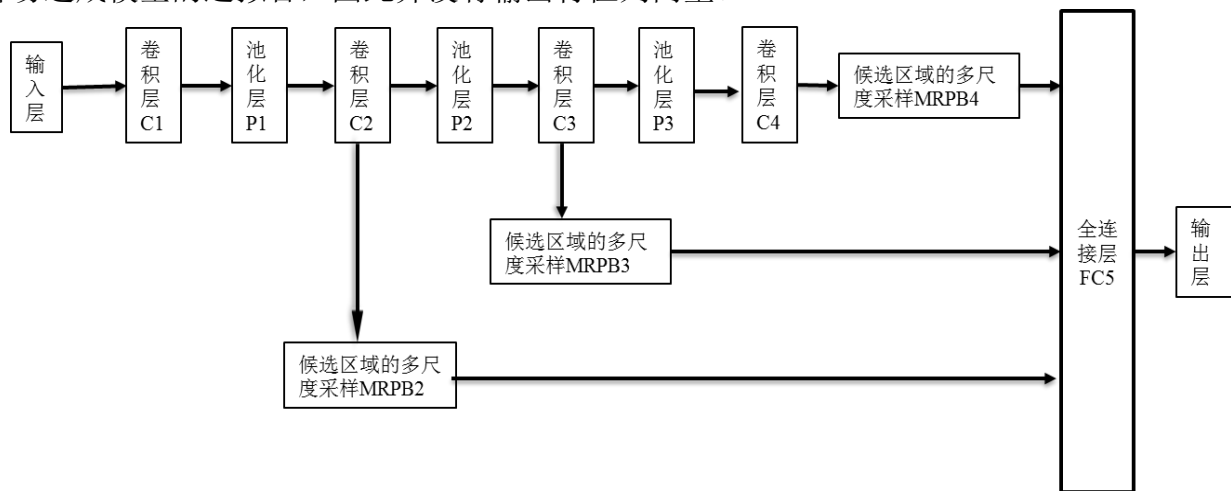


图 5.4 层次编码的池化网络

## 5.4 实验过程以及分析

本实验主要分为三个阶段，第一阶段是比较不同的池化核条件下面孔识别准确率，以验证基于层次信息融合的多尺度面孔识别新方法可以应对图像多尺度问题，多尺度的池化对目标变形带来的变化和空间布局具有较好的鲁棒性；第二阶段是比较仅有全局采样的识别模型和增加局部采样模块的模型的识别率，验证局部多尺度采样模块对面孔特征表达有益作用；第三阶段是比较单尺度样本训练和多尺度样本训练的模型，以多尺度的测试样本的识别率做测试，验证样本多尺度将会提高模型的泛化能力。

考虑到样本来源的一般性，本文选择的面孔样本是 YouTube Faces Database 人脸识别数据集。该数据集来源于视频，以应对于非约束条件下的面孔识别问题。该数据集包含 3425 个视频内的 1595 个不同的人。因此对于某个特征身份的面孔将会有 2.15 个视频，表 5.1 表示样本分布情况。

表 5.1 样本分布情况

视频数量（个）	1	2	3	4	5	6
人的数量（人）	591	471	307	167	51	8

为了增加样本差异性，本文将面孔数据做部分遮挡、图像镜像、对比度变化和亮度变化等 4 个操作，因此单张原始图像将会生成出 40 张图像，将数据库的数量扩大了 40 倍。然后从每个身份面孔样本中随机选择 200 张图片作为面孔识别样本。为了实验样本多尺度的需求，将原始尺度为  $64 \times 64$  调整为  $48 \times 48$  和  $54 \times 54$ ，该类变化只是调整图像的分辨率，并不改变其拓扑结构，样本处理示意图如图 5.5 所示。

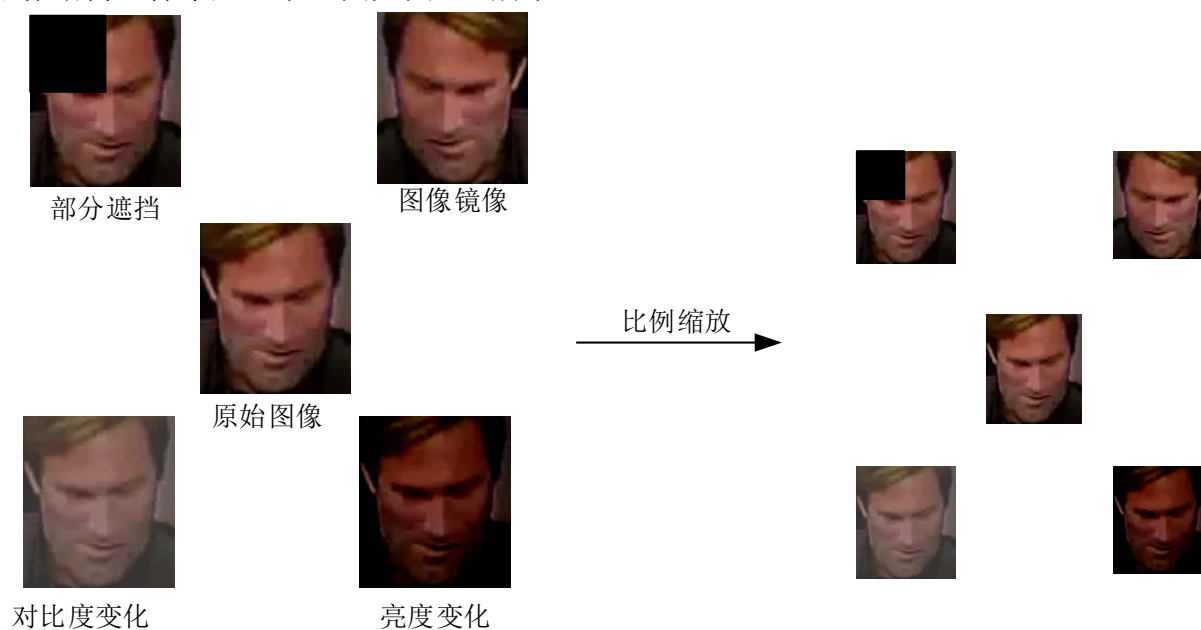


图 5.5 样本处理示意图

### 5.4.1 网络层的采样频率选择

为了证明多尺度池化的神经网络结构的有效性，本文将无多尺度池化面孔识别模型、单尺度池化模型和多尺度池化模型在识别准确率上的对比，其中无尺度池化模型就是普通的卷积神经网络模型，即全连接层仅仅与最后一层卷积层连接；单尺度池化模型是 **MRPB** 池化核大小仅有为  $1 \times 1$ ；多尺度池化模型的池化核选择为  $1 \times 1$  和  $2 \times 2$ ，输出特征融合后与全连接层连接。

本阶段实验将固定输入  $64 \times 64$  大小的面孔样本，在 **CAFFE** 深度学习框架上用批量梯度下降的方式训练模型。在样本处理方面，面孔数据集按照 4: 1 的比例随机划分，即 160 张面孔数据集作为训练数据集，40 张面孔数据集作为验证数据集。重复 10 次后计算平均准确率。测试运行结果如表 5.2 所示，无多尺度池化网络的样本识别准确率为 89%，单尺度训练模型为 85%，多尺度模型的准确为 91%，比一般的卷积神经网络的识别率高 2%，这也说明了基于层次信息融合的多尺度面孔识别新方法的有效性，从而侧面反映了生物视觉感知机制做决策分析时，并非仅仅依靠高级的表征信息，底层视觉通路信息也将会对面孔识别提供有益的补充特征。

表 5.2 样频率对面孔识别的准确率的影响

模型	无尺度池化	单尺度池化	多尺度池化
测试准确率 (%)	89	85	91

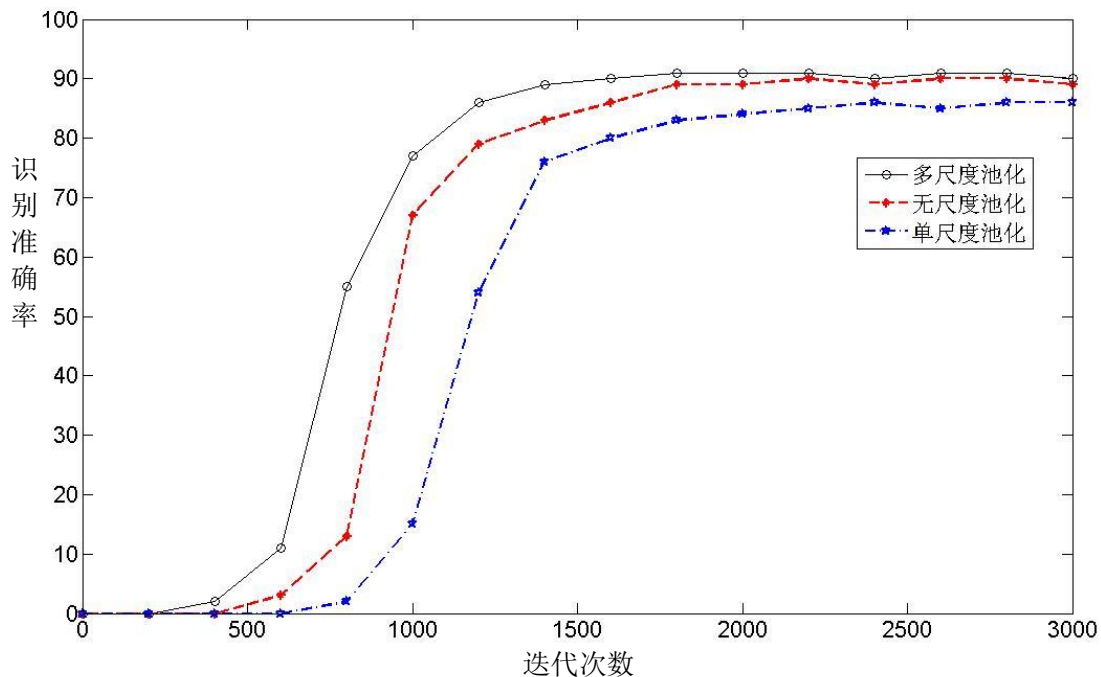


图 5.6 采样频率对面孔识别的准确率的影响

从图 5.6 可以看出, 模型迭代次数较小时候, 模型的识别准确率都处于较低水平; 当迭代次数为 2000 时, 所有面孔识别模型都处于其最佳的水平。从模型收敛的速度来看, 基于局部区域和整体区域结合的多尺度池化神经网络比单尺度池化模型收敛更加快, 证明随着采样频率的增加, 底层通路的视觉信息更加有效地提炼, 从而模型识别准确率也将会提高。当然单尺度池化模型的识别准确率较低, 是因为没有提炼足够的纹理信息造成干扰, 从而影响模型对识别目标空间布局的鲁棒性。

#### 5.4.2 全局信息和局部信息的关系比

为了验证局部多尺度采样模块对面孔特征表达有益作用, 本文将比较仅有全局信息的多尺度池化模型和具有 **MRPB** 结构的多尺度池化模型对比, 其中 **MRPB** 是在全局多尺度采样的基础上的基础上增加了局部多尺度采样模块。

本阶段实验依然固定输入图像大小为  $64 \times 64$  样本, 以 S 折交叉验证的方式试验, 运行结果如表 5.4 所示, 仅有全局采样的识别准确率为 84%, 仅有局部采样的识别准确率为 86%, 利用 **MRPB** 训练模型为 91%, 这也证明了 **MRPB** 结构的有效性, 说明了关注更加纹理细节的表征信息将会对有益于面孔识别的效果。

表 5.3 多尺度池化层加入全连接层的识别效果对比 (%)

采样结构	仅有全局信息	仅有局部信息	<b>MRPB</b>
测试准确率	84	86	91

局部信息的增加可以使得 **MRPB** 结构对遮挡等因素外界因素造成面孔识别变化具有良好的适应性。从视觉感知机制的角度看来, 人类对于较小尺寸的面孔识别情形, 并非仅仅依靠整体的轮廓信息, 而是将会关注到眼睛耳朵等具有表征信息的面孔区域, 也就是说面孔识别将是多视觉通路从不同视角表征的融合结果。

#### 5.4.3 多尺度样本训练的影响

为验证样本多尺度的训练将会提高层次信息融合的多尺度池化模型的泛化能力, 本文将比较单尺寸样本和多尺寸样本来训练模型, 在不同尺寸的面孔数据上做识别率上的对比, 其中单尺寸样本大小为  $64 \times 64$ , 多尺寸的样本大小选择为  $64 \times 64$  和  $48 \times 48$ , 最后在  $54 \times 54$  尺寸的面孔数据集测试识别的结果。

由于本试验采用的深度学习框架的局限性, 网络输入的结构一次训练只能接受固定尺寸的面孔图像, 本文将会设计采用周期性训练模式, 即在第一个训练周期内采用固定尺度的样本输入, 修改模型图像输入的大小, 采用不同尺寸的面孔样本来进行训练, 使得模型能够对其他样本识别有一定的适应性, 如此反复周期的迭代, 最终在损失曲线在趋向于稳定的水平时停止模型的训练。



表 5.4 单尺度模型和多尺度模型识别准确率对比 %

训练方式 \ 图像尺寸	64×64	54×54	48×48
单尺寸样本训练模型	91	89	85
多尺度样本训练模型	93	92	87

实验结果如表 5.4 所示，单尺寸样本训练模型在图像尺寸为  $64 \times 64$  上准确为 91%，相比  $54 \times 54$  的 89% 和  $48 \times 48$  的 85% 都要高；多尺寸样本训练模型在图像尺寸为  $64 \times 64$  上准确为 93%，相比  $54 \times 54$  的 92% 和  $48 \times 48$  的 87% 都要高。

由于多尺度样本在  $64 \times 64$  和  $48 \times 48$  两个尺度的样本上训练，因而具有较好结果。但是，由于测试图像尺度没有在训练图像的样本上出现，因此识别准确率处于较低的水平。单尺度训练模型输入尺寸的多尺度训练比单个尺寸训练的收敛速度更加快，准确率更高。说明多尺度样本的训练可以提高层次信息融合的多尺度面孔识别模型的识别准确率。

卷积核和输入图像之间存在比例关系，由于不同尺寸输入模型的卷积核在数量和感受野大小是保持一致的，因此修改输入图像大小从本质上改变了卷积的神经元接受到面孔信息的细节程度，这与修改卷积核的感受野大小原理是一致的。同时由于多尺度图像内容和空间布局并不发生变化，MRPB 结构一定程度上可以保持提取的信息量一致性。

## 5.5 小结

主要对本章提出的基于层次信息融合的多尺度面孔识别新方法进行研究。考虑到神经网络各层对识别目标有不同的程度表征，当尺度变化后的输入图像，面孔图像的特征会在一定程度上的丢失，从而各层的浅层的纹理特征能够投影到视觉决策分析层，针对于经典的面孔识别网络对输入的目标图像有固定尺度大小的要求，构建了固定输入的池化网络层，由此可以保证决策分析神经网络层的特征尺度，从而增强网络的适应性。实验结果表明，本章提出的算法不仅解决了面孔识别模型需要固定尺寸输入的弊端，而且具有尺度的泛化能力，从另一方面体现出面孔识别并非是基于单一的视觉通路，应该是多视觉通路共同作用后所形成的一种识别能力。



## 第 6 章 总结与展望

### 6.1 总结

随着社会品质的提高以及安防等领域的需求，视觉识别技术已经在模式识别与人工智能领域成为研究热门之一。以面孔识别为例，当前已有的识别技术存在诸多的弊端，例如基于主成份分析的特征提取方法，基于监督信息的线性辨别技术的方法，通常聚焦于数据的数学特征的模版匹配策略；而最近热门的深度学习算法，基于模仿卷积特性的人工设计的神经网络，在海量数据重复训练下，较其他传统的方法基础上有了较大的提高。但是值得说明的是，上述算法并没有考虑到实际的生物视觉识别系统的内在工作原理，而是将作为其作为黑盒来模拟实现，试图利用训练海量的样本来调整模型的参数，从而实现出最优决策。本文研究基于生物视觉感知机制上的视觉识别技术，通过视觉通路的组织功能和细胞特性仿真，对视觉系统决策分析的拓扑结构分析探索，针对与实际的面孔识别问题上应用和解决，构建新型的感知计算模型，为后续基于视觉感知机制的研究和应用起到了新的启发。

本文的具体研究成果如下所示：

(1) 考虑到监督信号对视觉识别模型提取的特征具有约束性，从而经典的识别计算模型在泛化能力上的欠缺。故本文提出了一种基于视觉信息流的双视通路的面孔验证模型。首先，底层的卷积网络将部分高级特征融入到决策分析层，以增强提取后的面孔特征信息；其次，模型训练过程中，在分类信号的基础上，增加监督信号，以强化同类面孔特征的相似性和异类的差异性；最后，通过比较提取证件照片和多姿态面孔照片的空间分布情况，根据联合贝叶斯人脸的方法判断面孔的身份相似性。实验结果表明，应用双通路的面孔设计机制能够提高面孔验证的准确率，以说明监督信号对模型训练有提高作用。

(2) 考虑到面孔识别是左半球和右半球的协作完成，然而传统的神经网络缺少局部结构的分析模块，从而仅仅识别单一的正立面孔图像。因而本文提出了一种基于左右半球协作机制的正立和倒立面孔识别新方法。模拟视觉信息流在视通路  $u$  中的传递和处理过程，首先构建底层神经网络，建立敏感纹理特征以及对称卷积核的机制，实现正立和倒立面孔图像的去除冗余和预处理；接着提出一种基于局部区域提取的池化神经网络层的概念，构建多局部特征融合的网络结构，实现局部信息的压缩提取和融合；最后根据高级视觉皮层中左右半脑协作的特点，提出一种融合整体和局部信息的预测函数。实验结果表明，以 AT&T 数据库为例，本文方法在经典卷积神经网络模型上增加了多局部特征融合的网络结构，识别准确率从 98% 提高到 100%，表明局部信息能够提高对正立面孔识别的能力；同时采用合适的训练数据集，调节融合时整体与局部信息的关系比，该模型对正立和倒立面孔的识别率分别为 100% 和 93%，

表明对正立和倒立面孔识别具有良好的特性。

(3) 考虑到浅层纹理特征也将参与面孔图像的特征, 经典面孔识别方法只能适应固定大小的面孔图像。提出了一种基于多层次信息表达的多尺度面孔识别新方法。首先, 模拟视网膜和神经节细胞对底层纹理特征的敏感的特点, 构建多层卷积池化结构级联的神经网络, 以达到视觉信息过滤和提炼的目的; 其次, 考虑到底层卷积对面孔特征表达的有益作用, 构建全局多尺度特征池化的面孔特征表达结构, 输出固定大小的特征通道和特征图谱; 然后, 根据视觉皮层的决策机制, 局部表征信息和全局表征信息融入到全连接层, 通过神经元的激活特性判断面孔身份。实验结果表明, 在拓展的面孔数据集上, 本方法可以有效适用于多尺度面孔图像的识别, 尤其是尺度较小的面孔识别, 从而说明底层纹理特征将会有助于面孔特征的表达, 提高面孔识别率。

## 6.2 展望

本文在面孔识别技术的实际需求中, 融入了视觉感知系统的机制, 证明了利用了基于视觉通路的进行的面孔识别有效性和可行性。针对于感受野神经元模型, 后续的工作可以考虑感受野的侧向抑制等特性; 视觉感知系统层与层之间是以放电的形式存在的, 后续的工作中可以考虑设计一类新的神经网络的参数更新策略。随着生物试验的不断的展开, 更多的生物视觉特性不断被发现, 其仿真一方面对视觉感知机制的揭露有提升作用, 另一方面, 有益与面孔识别技术的发展。

## 致谢

时光飞逝，转眼间紧张而又充实的研究生生活即将画上句号。在这段学习期间，我得到了很多老师、同学和朋友的关怀和帮忙。在学位论文即将完成之际，我要向所有期间给予我支持、帮忙和鼓励的人表示我最诚挚的谢意。

研究生学业上得以进步，我要由衷地感谢我的导师范影乐教授。范老师学识渊博、治学态度严谨，具有丰富的教书育人经验，在他的悉心指导下，才有我这些科研进展。范老师是学业上的良师，同时亦是生活中的益友。为人亲切的他言行举止儒雅，富有个人魅力，常常为我们生活，学习，就业上给出独到的帮助和建议。

研究生的学习主要是在实验室度过，在此要感谢同窗的 3 位同门郭金朋、张梦楠和胡钧皓，一路与你们共同学习、实验、科研，大家之间的互相帮助让研究生学习富有效率，不显孤独。还要感谢上届的师兄师姐，感谢谢尚周、李康群、郭笑天、李嘉祥、方芳、王典和王迪，在师兄师姐的带领下，我得以迅速进入研究生学习状态，之后的专业学习、论文阅读，以及就业选择方面，你们给了我莫大的帮助，你们宝贵的指导让我少走很多弯路。感谢师弟师妹：张明琦、蒋涯、杜宇华、周涛，实验室有了你们而显得越发有活力，为紧张科研工作注入一份乐趣。

最后还要感谢我的家人对我读研的支持，你们在我身后默默关心、支持，是我前进的最大动力。

## 参考文献

- [1] 宋丹, 黄旭. 生物识别技术及其在金融支付安全领域的应用[J]. 信息安全研究, 2016, 2(01):27-32.
- [2] 傅鹏, 谢世朋. 基于级联卷积神经网络的车牌定位[J]. 计算机技术与发展, 2017, 7(01):1-9.
- [3] 陈光喜, 康春生. 基于深度特征编码的两级车辆检索方法[J]. 计算机应用研究, 2017, 7(09):1-8.
- [4] Roth M W. Survey of neural network technology for automatic target recognition[J]. IEEE Transactions on Neural Networks, 1990, 1(1):28-36.
- [5] Dabbaghchian S, Ghaemmaghami M P, Aghagolzadeh A, et al. Feature extraction using discrete cosine transform and discrimination power analysis with a face recognition technology[J]. Pattern Recognition, 2010, 43(4):1431-1440.
- [6] Phillips P J, Flynn P J, Scruggs T, et al. Overview of the Face Recognition Grand Challenge[C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego: IEEE, 2005:947-954.
- [7] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11):2278-2324.
- [8] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]//International Conference on Neural Information Processing Systems. New York: ACM, 2017:84-90.
- [9] Kang K, Shapley R M, Sompolinsky H. Information tuning of populations of neurons in primary visual cortex[J]. Journal of Neuroscience the Official Journal of the Society for Neuroscience, 2004, 24(15):3726-3734.
- [10] 周涛, 郭义, 郭永明. 针刺电信息的神经编码特性[J]. 中西医结合学报, 2008, 6(12):1300-1304.
- [11] Kaufman G J, Breeding K J. The automatic recognition of human faces from profile silhouettes[J]. IEEE, 1976, 6(2):113-121.
- [12] Bichsel M. Strategies of robust object recognition for the automatic identification of human faces[J]. Thesis Swiss Federal Inst Technol, 1991, 10(2):153-167.
- [13] Brunelli R, Poggio T. Face recognition: features versus templates[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1993, 15(10):1042-1052.
- [14] Cootes, T.F, Edwards, G.J, Taylor, C.J. Active appearance models[J]. Pattern Analysis and Machine Intelligence, 2001, 23(6):681-685.
- [15] 吕润华, 苏婷婷, 马晓伟. BP 神经网络联合模板匹配的车牌识别系统[J]. 清华大学学报, 2013, 53(09):1221-1226.

- [16]Nefian A V. A hidden markov model-based approach for face detection and recognition[C]//PhD dissertation in Electrical Engineering, Georgia: Institute of Technology. 1999:130-141.
- [17]Turk M, Pentland B. Eigenfaces for recognition[J]. Journal of Cognitive Neuroscience, 1991, 3(1):71-80.
- [18]Moore B. Principal component analysis in linear systems: Controllability, observability, and model reduction[J]. IEEE, 1981, 26(1):17-32.
- [19]Lu J, Plataniotis K N, Venetsanopoulos A N. Face recognition using LDA-based algorithms[J]. IEEE, 2003, 14(1):195-202.
- [20]Le Q V. Building high-level features using large scale unsupervised learning[C]//IEEE International Conference on Acoustics. Maryland: IEEE, 2013:8595-8598.
- [21]Lecun Y, Boser B, Denker J S, et al. Backpropagation applied to handwritten zip code recognition[J]. Neural Computation, 1994, 6(4):541-551.
- [22]Jia D, Wei D, R. Socher, et al. ImageNet: A large-scale hierarchical image database[J]. 2009,15(6): 248-255.
- [23]Simonyan K. Very deep convolutional networks for large-scale image recognition[J]. Computer Science, 2014, 7(20):1409-1556..
- [24]Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[J]. IEEE, 2014, 16(1):1-9.
- [25]Sergey I, Christian S. Batch normalization: accelerating deep network training by reducing internal covariate shift[J]. IEEE, 2015, 9(8):448-456.
- [26]Christian S, Vincent V, Sergey I, et al. Rethinking the inception architecture for computer vision[J]. IEEE, 2015, 3(4):2818-2826.
- [27]Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-resnet and the impact of residual connections on learning[J]. IEEE, 2016, 7(20):58-62.
- [28]He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[J]. IEEE, 2015, 10(8):770-778.
- [29]Zhang X, Zhou X, Lin M, et al. An extremely efficient convolutional neural network for Mobile Devices[J]. IEEE, 2017, 1(8):410-418.
- [30]Sifre L, Mallat S. Rigid-motion scattering for texture classification[J]. Computer Science, 2014, 35(59):501-515.
- [31]Girshick R, Donahue J, Darrell T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[J]. 2013, 6(5):580-587.
- [32]He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE, 2014, 37(9):1904-1916.
- [33]Ren S, He K, Girshick R, et al. Towards real-time object detection with region proposal networks[J]. IEEE, 2017, 39(6):1137-1145.
- [34]Girshick R, Microsoft R. Fast region-based convolutional networks for object detection[J]. IEEE International Conference on Computer Vision, 2015, 6(5):1440-1448.

- [35]Rullen R V, Gautrais J, Delorme A, et al. Face processing using one spike per neurone[J]. *Biosystems*, 1998, 48(1-3):229-239.
- [36]Kamaruzaman F, Shafie A A. Recognizing faces with normalized local Gabor features and Spiking Neuron Patterns[J]. *Pattern Recognition*, 2016, 53(C):102-115.
- [37]Sabour S, Frosst N, Hinton G E. Dynamic routing between capsules[J]. *Nature*, 2017, 15(1):44-52.
- [38]Frintrop S, Rome E, Christensen H. Computational visual attention systems and their cognitive foundations[J]. *Transactions on Applied Perception*, 2010, 7(1):1-39.
- [39]Ehinger B. Neurotransmitter systems in the retina[J]. *Retina*, 1982, 2(4):35-42.
- [40]Buchsbaum M, Fedio P. Visual information and evoked responses from the left and right hemispheres[J]. *Electroencephalography & Clinical Neurophysiology*, 1969, 26(3):266-272.
- [41]Olshausen B, Field D. Emergence of simple-cell receptive field properties by learning a sparse code for natural images[J]. *Nature*, 1996, 381(6583):607-616.
- [42]Wang F, Xiang X, Cheng J, et al. L2 Hypersphere Embedding for Face Verification[J]. *IEEE*, 2017, 6(10):555-563.
- [43]Sun Y, Chen Y, Wang X, et al. Deep learning face representation by joint identification verification[C]//Advances in neural information processing systems. Columbus: Advances in neural information processing systems, 2014:1988-1996.
- [44]Sun Y, Wang X, Tang X. Deep Learning Face Representation from Predicting 10,000 Classes[C]// IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE Computer Society, 2014:1891-1898.
- [45]Sun Y, Wang X, Tang X. Deeply learned face representations are sparse, selective, and robust[C]// Computer Vision and Pattern Recognition. Boston: IEEE ,2015:2892-2900.
- [46]Schroff F, Kalenichenko D, Philbin J. A unified embedding for face recognition and clustering[C] // Computer Vision and Pattern Recognition. Boston: IEEE, 2015:815-823.
- [47]Liu W, Wen Y, Yu Z, et al. Large-margin softmax loss for convolutional neural networks[C]. New York: Journal of Machine Learning Research, 2016:507-516.
- [48]Wen Y, Zhang K, Li Z, et al. A discriminative feature learning approach for deep face recognition[C]//European Conference on Computer Vision. Las Vegas: Springer International Publishing, 2016: 499-515.
- [49]Leder H, Goller J, Forster M, et al. Face inversion increases attractiveness[J]. *Acta Psychologica*, 2017, 17(8):25-31.
- [50]Itier R, Taylor M. Face recognition memory and configural processing: A developmental ERP study using upright, inverted, and Contrast-reversed faces[J]. *Neuro*, 2004, 16(3):487-502.
- [51]Hills P J, Mileva M, Thompson C, et al. Carryover of scanning behaviour affects upright face recognition differently to inverted face recognition[J]. *Visual Cognition*, 2017, 6(15): 1-14.
- [52]Schwartz N Z. Reconsidering face specialization and face inversion[J]. *Dissertations & Theses - Gradworks*, 2007, 15(10):66-73.

- [53]Leder H, Bruce V. Feature processing from upright and inverted faces[M]. Berlin Heidelberg: Face Recognition Springer, 1998:547-555.
- [54]Heering A, Rossion B, Maurer D. Revisiting upright and inverted face recognition in 6 to 12-year-old children and adults[J]. Journal of Vision, 2010, 10(7):581-581.
- [55]Zeiler M D, Fergus R. Visualizing and understanding convolutional networks[M]. Zurich: Springer International Publishing, 2014:818-833.
- [56]Sermanet P, Eigen D, Zhang X, et al. Overfeat: Integrated recognition, localization and detection using convolutional networks[J]. IEEE, 2014, 4(6):312-320.
- [57]Helmut Leder, Vicki Bruce. When inverted faces are recognized: The role of configural information in face recognition[J]. Quarterly Journal of Experimental Psychology A Human Experimental Psychology, 2000, 53(2):513-521.
- [58]Bartlett J C, Searcy J. Inversion and configuration of faces[J]. Cognitive Psychology, 1993, 25(3):281-316.
- [59]吴斯. 基于多尺度池化的卷积神经网络人脸识别方法研究[D]. 浙江大学, 2016:20-30.
- [60]Kilner J M, Friston K J, Frith C D. An account of the mirror neuron system[J]. Cognitive Processing, 2007, 8(3):159-166.

## 附 录:

### 作者在读期间发表的学术论文及参加的科研项目

#### 发表（录用）学术论文：

1. 正立和倒立面孔的混合识别研究，中国图象图形学报（一级期刊），录用待发表，1/4

#### 所获的奖项

1. 视频监控下的人脸识别，全国研究生智慧城市技术与创意设计大赛视频挑战赛三等奖，教育部学位与研究生教育发展中心，2016，1/4

#### 参与的科研项目：

1. 基于眼前房角 OCT 影像质量分级的原发性闭角型青光眼辅助自动诊断研究(61501154)，国家自然科学基金项目。
2. 自动化检测中的机器视觉开发模块研发(063113017)，横向课题。