

上海交通大学

SHANGHAI JIAO TONG UNIVERSITY

学士学位论文

BACHELOR'S THESIS



论文题目： 人脸识别算法研究

学生姓名： 赵彤

学生学号： 5120369050

专 业： 信息安全

指导教师： 郭捷

学院(系)： 电子信息与电气工程学院

人脸识别算法研究

摘要

人脸识别是当今人工智能的热门研究领域，它可以应用于军事、安全、金融、商业等多个领域中。但人脸图像的非刚性、易变性、相似性与用户配合度都使该课题成为本领域最为复杂的研究课题之一。本文旨在研究复杂背景中的人脸识别相关算法，并利用深度学习技术对相关算法进行设计和改进，得到具有一定实用价值的算法或模型。

作为人脸识别算法的基础，本论文首先对人脸定位算法开展研究，主要包括人脸检测、关键点检测与人脸对齐三个步骤。针对人脸检测算法，本文分别研究了基于 Haar 特征与基于 LBP 特征的 AdaBoost 算法，并在 MUCT 数据集上进行了大量的仿真实验与对比分析。为进一步对检测到的人脸图像进行归一化处理，本文还分别实现了基于深度学习的人脸关键点检测算法与基于仿射变换的人脸对齐算法，并在现有的大规模非限制性人脸数据集上取得了很好的测试效果。

然后，本论文对人脸识别算法开展研究，分别针对身份查找与身份确认两种应用场景进行算法设计。针对分类粒度较高的身份查找问题，本文提出了基于转化学习的人脸模型训练方法，与基于特征脸与支持向量机的传统机器学习算法以及普通卷积神经网络相比，本模型体现出了较明显的识别优势。以该模型为基础，本文还实现了基于卷积神经网络特征提取的身份确认算法与基于对称卷积神经网络的身份确认算法，并在多个公开人脸数据集上取得了较好的测试效果。

在以上研究的基础上，本论文还将人脸识别技术创新性的引入到人脸属性（包括年龄、性别、种族等）的识别算法研究中，借助一个简洁的新型卷积神经网络结构，通过合并层降低模型复杂度，可以有效地解决多种属性的识别问题，大量的仿真实验结果表明，该算法在真实环境下的数据集上有着突出表现，从而进一步拓展了人脸识别技术的应用。

关键词：人脸定位，人脸身份识别，人脸属性识别，深度学习

RESEARCH ON FACE RECOGNITION ALGORITHMS

ABSTRACT

Face recognition is a hot field of artificial intelligence. It can be used in many fields, like: military, security, finance, commerce and so on. However, face images are non-rigid, mutable but similar to some extent, which made it one of the most difficult research subjects in this field. This paper aims to study related algorithms of face recognition with complex backgrounds. We realize and compare some of the most effective algorithms. We combine these original algorithms with deep learning and obtain several models which can be used in practical applications.

As the base of face recognition algorithm, this paper first focuses on face localization algorithm, including face detection algorithm, facial landmark detection algorithm, face alignment algorithm. For face detection algorithm, we realize two cascade AdaBoost classifiers with Haar-like feature and local binary pattern (LBP) respectively. Comparing their results based on MUCT database, we analysis the differences between these two algorithms as well as their applications. In order to normalize facial images further, we implement the facial landmark detection based on deep learning and the face alignment algorithm based on affine transformation, which shows a good result on large-scale unrestricted face databases.

And then, we do researches on face recognition algorithm. It contains two aspects: one-to-all face identification algorithm and one-to-one face verification algorithm. For the face identification with high granularity, this paper proposes a face model training method based on transfer learning. Compared with traditional machine learning algorithm based on Eigenface with support vector machine and original convolutional neural network, our model embodies apparent advantages. Based on this model, we also design and realize the verification algorithm based on convolutional neural network feature and the verification algorithm based on Siamese network. Both methods show good results on LFW database.

After above studies, we introduce face recognition technology to facial attribution recognition algorithm creatively. We focus on age, gender, and race attribution. We propose a new network structure with merge layer to reduce its computation complexity, which could be used on resolve many kinds of facial attribution recognition problems. The results of many experiments show that our method can be applied in databases from real environment, thus expand the application of face recognition technology.

Key words: face localization, face recognition, facial attribution recognition, deep learning

目 录

第一章 绪论	1
1.1 课题的研究背景及意义	1
1.2 国内外研究现状	2
1.3 课题研究内容	2
1.4 论文结构安排	3
第二章 静态图片中的人脸定位算法研究	5
2.1 AdaBoost 人脸检测算法	5
2.1.1 AdaBoost 分类器原理	5
2.1.2 基于 Haar 特征的 AdaBoost 人脸检测算法	7
2.1.3 基于 LBP 纹理特征的 AdaBoost 人脸检测算法	9
2.1.4 实验仿真与性能对比	13
2.2 基于深度学习的人脸关键点检测算法	14
2.2.1 人脸关键点	14
2.2.2 卷积神经网络原理	15
2.2.3 模型结构	16
2.2.4 实验仿真与性能对比	18
2.3 人脸对齐算法	18
2.3.1 仿射变换	18
2.3.2 实验仿真与性能对比	19
2.4 本章小结	20
第三章 静态图片中的人脸识别算法研究	21
3.1 人脸识别算法概述	21
3.2 基于人脸识别的身份查找算法	21
3.2.1 基于特征脸与支持向量机的身份查找算法	22
3.2.2 基于卷积神经网络的身份查找算法	23
3.2.3 基于卷积神经网络的身份查找算法的改进——转化学习	24
3.2.4 实验仿真与结果分析	25
3.3 基于深度学习的身份确认算法	28
3.3.1 基于特征提取的身份确认算法	28
3.3.2 基于对称卷积神经网络的身份确认算法	29
3.3.3 实验仿真与结果分析	31
3.4 本章小结	32
第四章 静态图片中的人脸属性识别算法研究	33
4.1 人脸年龄识别算法概述	33
4.2 基于分类的年龄识别神经网络	34
4.2.1 模型结构	34
4.2.2 Softmax 损失函数	34
4.2.3 实验仿真与结果分析	35
4.3 基于回归的年龄识别神经网络	36



4.3.1 基于分类模型的改进卷积神经网络	36
4.3.2 基于多尺度的年龄识别神经网络	37
4.3.3 实验仿真与结果分析	40
4.4 基于不平衡数据集的人脸性别与种族识别神经网络	41
4.4.1 模型结构	41
4.4.2 Hinge 损失函数	42
4.4.3 实验仿真与性能对比	42
4.5 本章小结	43
第五章 结论	45
5.1 结论	45
5.2 展望	46
谢辞	52

第一章 绪论

随着智能监控与图片识别系统在日常生活中的广泛应用，生物识别技术成为一项热门的计算机技术研究领域。生物识别技术主要运用人体具有的生理特征（如：人脸、虹膜、指纹等）和行为表现（如：声音、笔迹、动作等），结合影像学、光学、分析学、统计学等先进技术手段，来进行生物身份的判别与认定。在自然界中，每个生物个体都有其独特的生物属性，具有唯一性、可认证性、终身不变性、血缘遗传性等，较之传统认证技术，生物识别技术有着很大的优势，发展前景良好。

人脸识别技术是生物识别技术最为重要的研究领域之一，其借助计算机技术，通过分析静态图片或视频帧分析与确认个体身份。人脸识别技术在安全领域有着重要的作用，可以结合其他生物识别技术完成特定人物监控、身份确认等工作。同时，人脸识别技术也有着重要的商业与市场价值，是大数据时代的前沿科学领域。

1.1 课题的研究背景及意义

所谓人脸识别，是指用摄像机、照相机、摄像头等多媒体技术采集包含人脸区域的图像或视频流，运用计算资源，基于人脸部的特征与信息进行人物身份识别与认证的一种生物识别技术，通常也被称为面部识别技术或人像识别技术。

人脸识别在很多领域有着广阔的应用前景，例如：身份验证系统、公安机关罪犯筛查系统、档案管理系统、银行与海关监控系统、商业数据分析系统等多方面任务。与其他生物识别技术（如：指纹识别、虹膜识别、视网膜识别等）相比，人脸识别技术因其具有数据的易收集性、易处理性，自然性和不易被测个体察觉的特点，更适合现今流行的大规模计算，得到鲁棒性与准确率良好的模型。

人脸识别技术的研究起步较早，但至今为止，人脸识别技术仍尚未成熟，较难运用于对准确率、安全性要求极高的系统中。人类从婴儿时期开始学习人脸识别的能力，直至成年，其识别准确率仍因人而异，存在上限。因此，构建一个鲁棒性好，准确率高的计算机自动人脸识别系统更是一项极具挑战性的任务。首先，人脸为一个非刚性物体，不同的表情、形态（如：正脸、侧脸等）、服饰配件（如：眼镜、头巾、帽子、护袖等）、发型、完整度等变化因素很大程度上影响着算法的性能。人脸识别技术也因计算机视觉中普遍存在的光线、角度、复杂背景、遮挡、模糊等问题而面临严峻挑战。

二十世纪六十年代，Beldsoe 与 Helen Chan、Charles Bisson 共同提出使用计算机识别人脸。但由于他的策略源于自定义算法并且不允许大量出版，因此并没有得到广泛关注。直到二十世纪八十年代，各行各业对于人脸识别系统有着极为迫切的需求，科学家们再次聚焦于人脸识别算法。1988 年，Kirby 和 Sirovish 应用主成分分析算法将经过正规化和对齐的人脸映射成了少于一百维的向量，是人脸识别算法研究史的里程碑。1991 年，Turk 和 Pentland 发现通过使用特征脸技术，残差可应用与在图片中寻找人脸位置，使实时人脸自动识别系统的实现成为可能。尽管这种方式仍然受到环境因素的影响，但它很大程度上促进了自动人脸识别技术的进一步发展。2001 年 1 月，Super Bowl 通过实验证明了可以通过获取监控视频系列图片，与数据库中的电子图片进行比较完成实时人脸识别系统，吸引了公众与媒体的广泛关注。2006 年，Geoffrey Hinton 提出了将非监督的逐层贪心训练算法

用于卷积神经网络的训练,使得深度学习得以在各领域应用^[1]。计算机视觉进入了新的时代,人脸识别算法也得到快速的发展和提高。

1.2 国内外研究现状

近 20 年来,研究者在人脸识别领域有了长足的进步。前期其注意力主要集中于在有严格条件限制的小型人脸数据集(一万张左右)上实现正脸图片识别任务的特定策略,而后期,大家的注意力更多的转向到基于大数据集(十万张以上)进行的多姿态、弱限制条件、实时人脸识别任务。目前常用的人脸识别算法可分为基于传统模型的算法与基于深度学习的算法。

基于传统模型的算法中最重要的问题就是如何选取人脸的特征。目前较为常用的特征有基于纹理的局部二值特征(Local Binary Pattern)^[2,3,4,5,7],基于边缘滤波的 Gabor 特征^[6,7,8,9,12],基于傅里叶变换的离散余弦变换(Discrete Cosine Transform)^[10,11,12],基于特征点的 SIFT 特征^[13,14,15]等。由于单一特征的不稳定性与易破坏性,多特征融合的方法被提出并得到广泛应用^[16,17,18]。由于人脸特征的维数通常较大,降维技术在该领域也有着广泛的应用,常见的降维方法有可用于构造特征脸的主成分分析法(Principal Component Analysis)^[19,20,21]和线性判断式分析法(Linear Discriminant Analysis)^[22,23,24]。

基于深度学习的算法在人脸识别领域有着广泛的应用,它使得人脸识别可以真正用于实际的工程中。当前人脸识别领域流行的应用如 Face++、How old、DeepID^[25]、DeepFace^[26]等,均是深度学习的成功应用案例。越来越多的基于卷积神经网络的模型结构被提出^[27,28,29,30]。卷积神经网络还可以通过非凸函数的拟合而产生鲁棒性更强的高级图像特征^[31,32,33]。借助 GPU 加速功能,卷积神经网络可以充分利用大规模的已标注数据进行训练,得到有效的图片特征表达。随着网络结构的不断优化和层数的不断加深,人脸识别算法的准确率不断提高,为实现其应用价值奠定了良好的基础。

随着人脸识别技术的成熟与推广,更多的需求被提出。作为人脸识别的拓展应用,人脸属性识别受到了学者的广泛关注。如 Gil Levi 等人将卷积神经网络应用于年龄与性别识别^[34], Antipov 等人将集成的卷积神经网络应用于性别识别^[73], Santoni 等人将其应用于种族识别^[36]。不同的问题给人脸识别带来了更多的活力。

1.3 课题研究内容

本次课题的研究内容为人脸识别算法,共分为三个部分:人脸定位算法、人脸识别算法、人脸属性识别算法。

人脸定位算法旨在从给定图片中精确定位人脸,进行目标的归一化。本部分将从基于 AdaBoost 的人脸检测算法、基于深度学习的人脸关键点定位算法、基于仿射变换的人脸对齐算法三方面进行研究。本文将对三种算法进行编程实现,对比找到更符合实际应用场景的算法,为后续算法提供支持。

人脸识别算法包括身份查找和身份认证两种机制。身份查找旨在将输入的一张人脸图片与后台数据库中所有人进行对比,完成身份匹配任务。身份确认旨在对比两张输入图像,判定其是否来自于同一个人。对于身份查找任务,本文以基于特征脸和支持向量机的身份查找算法作为基准,通过卷积神经网络算法、基于卷积神经网络的转化学习方法对结果进行提升。对于身份确认任务,本文将尝试基于卷积神经网络特征的距离分类算法、对称神经网络等当前最为热门的深度学习方法,并通过实验对结果进行比较,以期得到可以实际应用的人脸识别模型。

针对人脸属性识别算法，本文将从年龄、性别、种族三方面进行讨论。卷积神经网络可以有效的解决分类与回归两种问题。通过在传统模型基础上进行改进，新的模型结构被用于人脸属性识别问题，有效的降低了计算复杂度。得到的模型将在公开的非限制性数据集上进行有效性验证，与当前最好的结果进行对比。

本文所涉及到的算法实现逻辑与最终系统架构如图 1-1 所示。

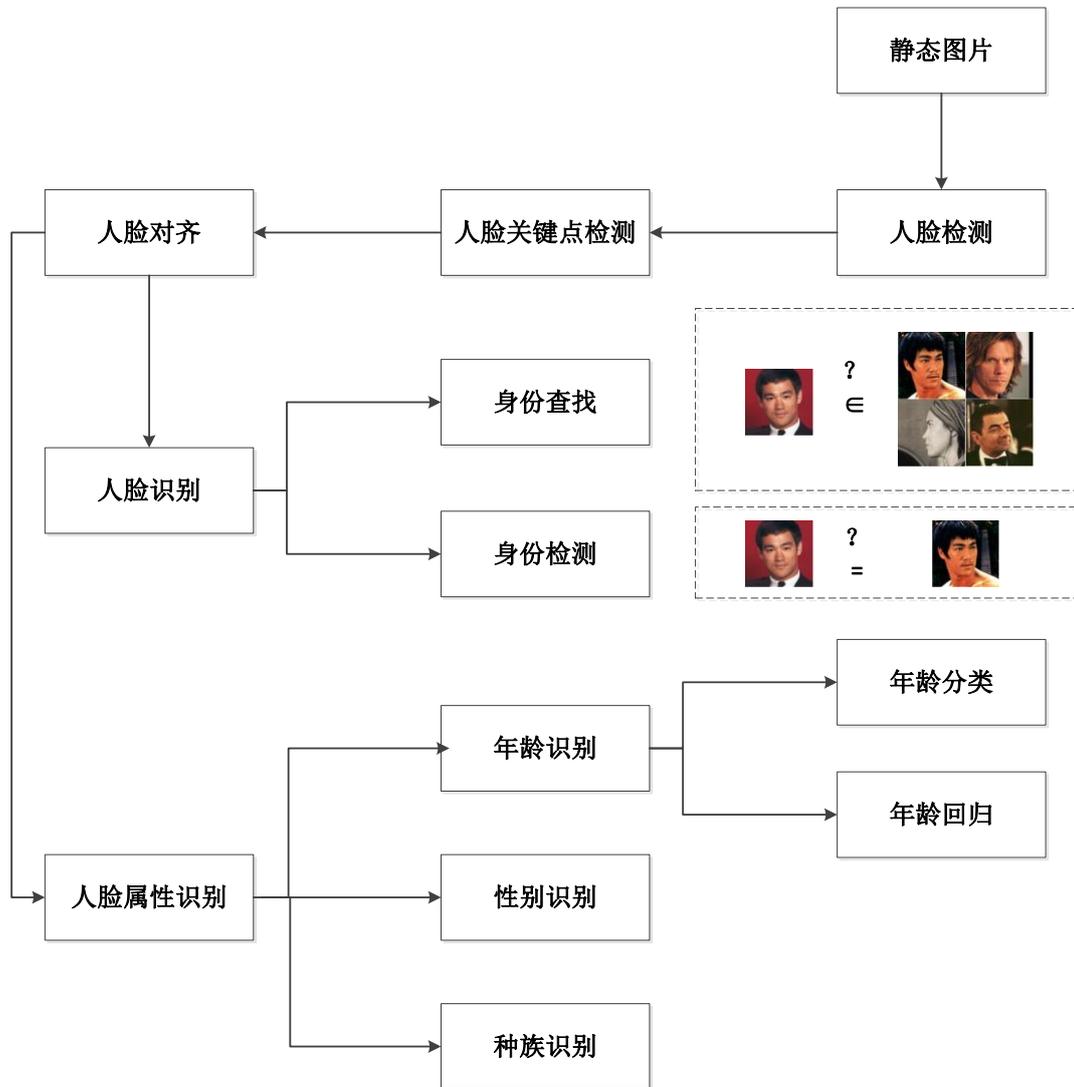


图 1-1 人脸识别系统设计框架

1.4 论文结构安排

本文的结构安排如下：

第一章简述了本文的研究课题、研究意义、国内外研究现状与课题研究内容。

第二章介绍了人脸定位算法中的人脸检测算法、人脸关键点定位算法与人脸对齐算法。针对每个算法，本文使用了一种或多种方法进行编程实现，通过对比明确了本次课题采用的方案。

第三章从身份查找、身份验证两方面实现了人脸识别算法。本章在传统算法的基础上引入了深度学习算法，并对卷积神经网络模型进行不断进行改进，在现有的公开复杂人脸

数据集上达到了很好效果。

第四章阐述了基于深度学习的人脸属性算法。本文所提出的新型简洁网络结构可以适应分类与回归两种问题，得到了用于年龄、性别、种族判断的高性能模型。

第五章是对本文的总结与展望，具体说明了本文的贡献与不足，并对之后的工作进行了规划。

第二章 静态图片中的人脸定位算法研究

人脸预处理是所有人脸识别任务的基础工作，建立一个可以有效完成人脸检测与对齐的系统可以有效的提高后续算法的效率。本次课题中需要实现的人脸预处理任务的具体步骤为：在图片中进行全局搜索，准确定位包含人脸信息的矩形区域；通过关键点检测算法对五官进行定位，得到人脸的几何特征；根据提取到的特征进行人脸比例与尺寸的归一化。因此，本章将分别从人脸检测、关键点定位、人脸对齐三个方面展开。

2.1 AdaBoost 人脸检测算法

大多数人脸检测算法被看做是一个二元分类问题。尽管它看起来易于实现，但却在实际问题中受到诸多因素的挑战，例如：外貌的差别、人脸位置与角度的变化、饰品与背景的遮挡等。同时，作为其他工作的基础，人脸检测算法对实现效率也有着很高的要求。有效的特征提取、性能良好的分类器是任务的关键。本次课题分别采用 Haar 特征与 LBP 特征，基于 AdaBoost 分类器进行算法实现。

2.1.1 AdaBoost 分类器原理

AdaBoost 是 Adaptive Boosting 的简写，在 1995 年由 Yoav Freund 和 Robert Schapire 提出^[37]。AdaBoost 的主要思想为通过对简单的弱分类器的线性组合级联构建一个强分类器，如图 2-1 所示。它的自适应的特点体现在，前级分类器结果错误的样本会在之后的训练中加强，而加权后的训练集中的所有样本都将被用于下一级弱分类器的训练。随着弱分类器数目的增加，分类器在训练集上的准确率将不断提高，直到达到预设的可接受错误率或达到指定的最大迭代轮数。

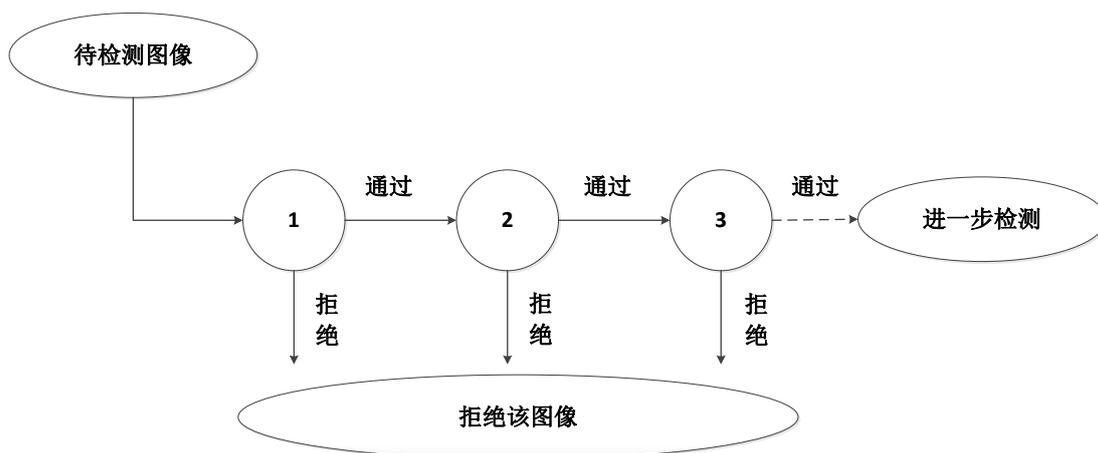


图 2-1 AdaBoost 分类器测试流程

算法流程如下^[38]：

给定训练数据集 $T = \{(x_1, y_1), \dots, (x_m, y_m)\}$ ，其中样本 $x_i \in x$ ，样本空间 $x \subset R^n$ ，

$y_i \in \{-1, +1\}$

(1) 初始化训练样本的权值分布:

$$D_1(i) = \frac{1}{m} \quad (2-1)$$

其中 $i = 1, \dots, m$

(2) 对于 $t = 1, \dots, T$:

a. 通过权值分布 D_t 训练弱分类器, 得到弱假设 $h_t: x \rightarrow \{-1, +1\}$

b. 计算 h_t 在训练样本上的分类错误率

$$\varepsilon_t = Pr_{i \sim D_t}[h_t(x_i) \neq y_i] \quad (2-2)$$

由此表达式可知, h_t 在训练样本上的错误率 ε_t 即为被 h_t 错误分类的样本的权值之和

c. 选择

$$\alpha_t = \frac{1}{2} \ln\left(\frac{1 - \varepsilon_t}{\varepsilon_t}\right) \quad (2-3)$$

d. 计算 $i = 1, \dots, m$, 更新参数

$$D_{t+1}(i) = \frac{D_t(i) e^{-\alpha_t y_i h_t(x_i)}}{Z_t} \quad (2-4)$$

其中 Z_t 是规范化因子, 它使得 D_{t+1} 成为一个概率分布。

(3) 组合弱分类器

$$f(x) = \sum_{t=1}^T \alpha_t h_t(x) \quad (2-5)$$

得到最终的分类器为

$$H(x) = \text{sign}(f(x)) = \text{sign}\left(\sum_{t=1}^T \alpha_t h_t(x)\right) \quad (2-6)$$

AdaBoost 的主要优势在于它的特征选择机制, 弱分类器不需要任何先验知识。限定特征搜索采取贪心策略, AdaBoost 可以有效地选择特征以最小化分类错误率。在训练之前, AdaBoost 不需要设定超参数 (迭代轮数 T 除外)。与其他学习算法相比, AdaBoost 在全局搜索的速度上占有较大优势, 使得其在各类检测算法中起着至关重要的作用。

2.1.2 基于 Haar 特征的 AdaBoost 人脸检测算法

较之于基于像素点的检测算法，基于特征的检测算法有着更稳定的性能和更快的速度。特征可以针对不同问题的专业知识对大量的数据进行编码压缩，对信息进行有效的表示，提高算法的检出率。

Haar-like 特征于 1997 年被 Oren 和 Papageorgiou 用于行人检测^[39]。2001 年，Viola 和 Jones 两位学者在 AdaBoost 算法的基础上，使用 Haar-like 小波特征和积分图的方法进行人脸检测，取得了很好的效果^[40]。因此基于 Haar 特征的 AdaBoost 人脸检测算法又被成为 Viola-Jones 检测器。2002 年，Rainer Lienhart 和 Jochen Maydt 在原有的 Haar-like 特征上进行了拓展，最终形成了现在广泛使用的 Haar-like 特征^[41]。

(1) Haar-like 特征

Haar-like 特征按照特征模板分为三类，分别为：边界特征、线特征、中心特征。特征模板由白色和黑色两种矩形组成。

Haar-like 特征共有 14 种，如图 2-2 所示。

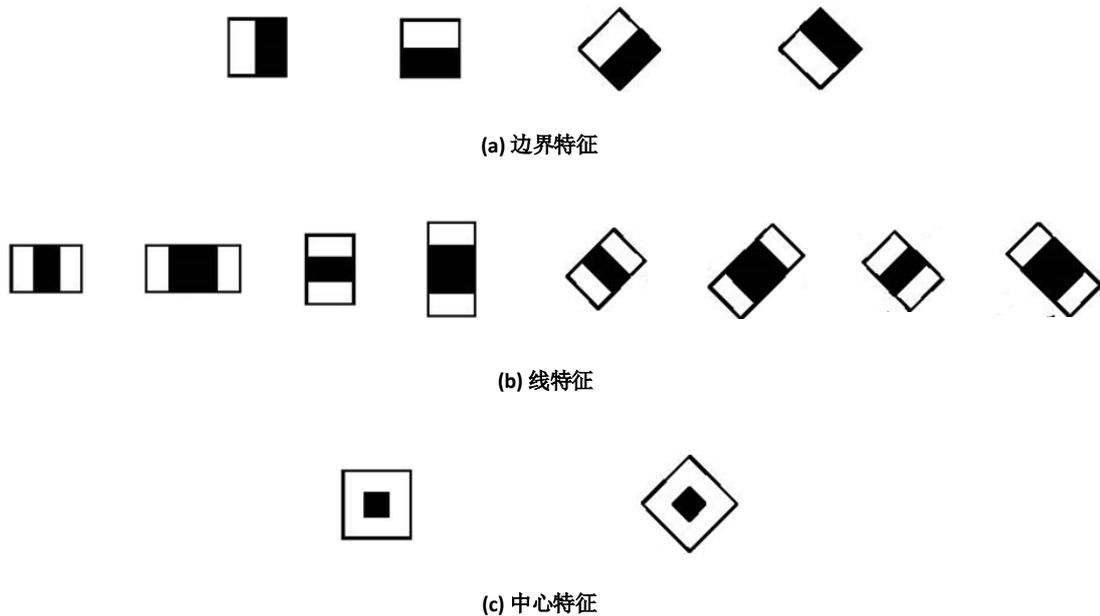


图 2-2 Haar 特征模板

模板的特征值定义如下：

- 对于边界特征，特征值为白色矩形区域内的像素和减去黑色矩形区域内的像素和。
- 对于线特征，特征值为白色矩形区域内的像素和减去两倍的黑色矩形区域内的像素和。
- 对于中心特征，特征值为白色矩形区域内的像素和减去三倍的黑色矩形区域内的像素和。

模板的特征值可以有效的反应图像特定区域的灰度变化情况，例如：眼部轮廓颜色深于面颊、鼻梁两侧区域深于鼻梁。效果如图 2-3 所示。



图 2-3 有效的人脸 Haar 特征

随着特征模板的尺寸和位置的变化，可在搜索图像子窗口中得到数量极多的特征。为了实现快速计算并有效提取特征，Haar-like 特征需要借助积分图进行计算。

(2) 积分图

积分图是一种可以加速矩形特征计算的对图像的中间表示形式。它只需要对图像进行一次遍历便可以计算出图像中所有区域的像素和。其主要思想为：在内存中记录图像以左上角和每个像素点为对角线形成的矩形区域内的像素之和。当需要计算 Haar-like 特征时，可以对其进行直接对数组元素进行索引，而不需要重新计算。积分图可以支持多尺度特征提取，并能在常数时间内完成每次计算，对检测速度有极大的提高。

积分图生成算法如下：

- a. 以 $S(i,j)$ 代表行方向像素点的累加和，并初始化 $S(i,-1) = 0$
- b. 以 $ii(i,j)$ 代表积分图像，并初始化 $ii(-1,j) = 0$
- c. 按行扫描图像，以递归方式计算每个像素点 (i,j) 对应的 $S(i,j)$ 和 $ii(i,j)$ ，其

中：

$$S(i,j) = S(i,j-1) + f(i,j) \tag{2-7}$$

$$ii(i,j) = ii(i-1,j) + s(i,j) \tag{2-8}$$

- d. 直至图像扫描至右下角，积分图 ii 构建完成

利用积分图，图像中指定矩形区域的特征值均可通过加减运算得到，如图 2-3 所示：

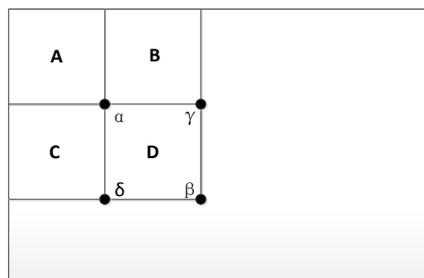


图 2-3 利用积分图计算 Haar 特征

将 D 的四个顶点分别标记为 α 、 β 、 γ 、 δ ，则矩形区域 D 的像素和为：

$$\text{sum}(D) = \text{ii}(\alpha) + \text{ii}(\beta) - \text{ii}(\gamma) - \text{ii}(\delta) \quad (2-9)$$

(3) Viola-Jones 检测器

Viola-Jones 检测器分为训练与检测两个部分，系统流程如图 2-4 所示。训练过程在 2.1.1 中已经给出，在此不再赘述。

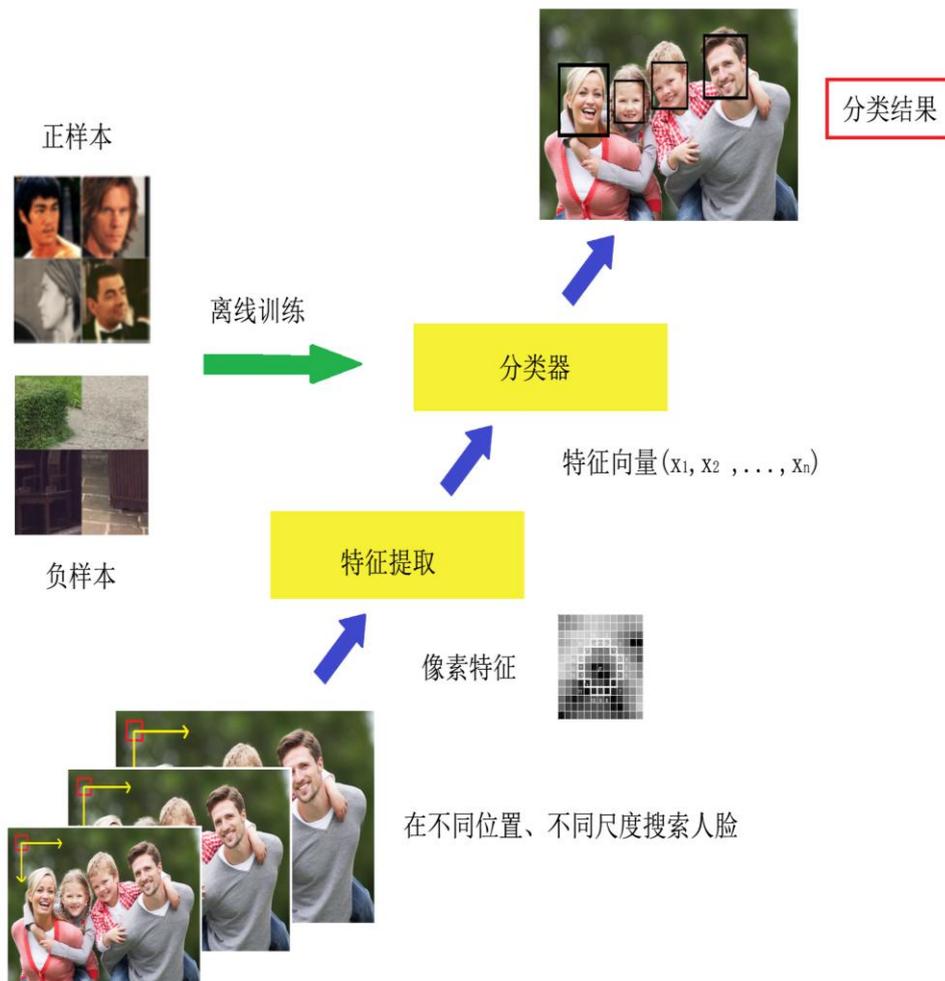


图 2-4 Viola-Jones 检测器系统流程图

在对图像进行检索的过程中，输入样本的尺寸一般都会比训练样本大很多，因此需要不断扩大检测窗口的尺寸。

检测开始时，设定检测窗口与样本保持一致，按照设定的步长（即每次移动的像素点个数）按照向左向下的方向进行移动，直至历遍整个图像，并标记出可能存在人脸的图像区域。完成一次遍历后，按照设定的放大比例对检测窗口进行放大，重复上述过程，直至检测窗口尺寸超过原图像的一半。在完成全部遍历过程后，对检测到的人脸区域进行合并与过滤，得到最终的目标区域。

2.1.3 基于 LBP 纹理特征的 AdaBoost 人脸检测算法

基于 Haar 特征的 AdaBoost 算法使人脸检测领域得到了突破性的进展，然而随着实际

应用的增加, 这种方法也暴露出一些问题。一方面, Haar-like 矩形特征过于简单, 使得弱分类器效果不佳, 在级联过程中, 算法很难进一步提高强分类器的准确率。另一方面, Haar-like 特征的特征空间极大, 使得训练过程耗时较长。因此提出了基于 LBP 纹理特征的 AdaBoost 人脸检测算法^[42]。

(1) LBP 特征的原理

LBP 算子^[43]是性能最好的纹理特征之一, 它被广泛的应用于各种应用之中。它有着高区别度, 低计算复杂度, 同时它对单调的灰度级别变化并不敏感, 使得其能在各种光照条件下保持稳定的性能。

原始的 LBP 算子定义在图像中的每一个 3*3 的窗口中, 通过比较计算中心像素点与周围 8 个像素点的值的大小对中心像素点进行二元编码。图 2-5 通过实例展示了如何计算 LBP 特征值。

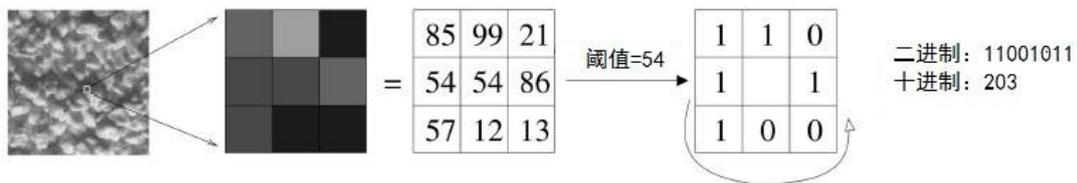


图 2-5 LBP 特征值的计算

中心像素点的 LBP 特征值定义如下:

$$LBP(x_c, y_c) = \sum_{p=0}^{p-1} 2^p s(i_p - i_c) \quad (2-10)$$

上式中, (x_c, y_c) 表示中心像素点的坐标, 其灰度值为 i_c , 3*3 邻域的灰度值为 i_p , $s(x)$ 为符号函数, 即:

$$s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{else} \end{cases} \quad (2-11)$$

为了能够提取多维度的 LBP 特征, 圆形 LBP 算子被进一步提出^[43]。原始算子中的正方形区域被扩展为半径不同的圆形区域。定义邻域为一组平均分布在半径为 R 的圆上平均分布的 P 个采样点, 当采样点没有落在像素点上时, 使用双线性插值求得对应值。在图 2-6 中, 表达式(P,R)表示 P 个采样点在半径为 R 的进行采样。

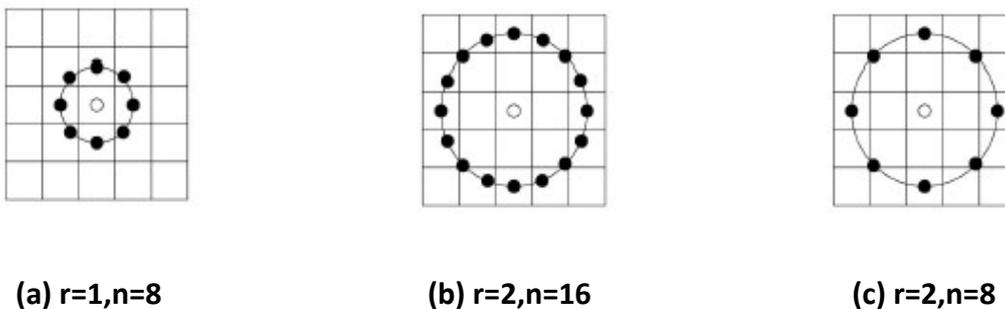


图 2-6 圆形 LBP 算子

由上述可知，LBP 算子虽然对于灰度有很好的稳定性，但却对于旋转敏感。为了使其具有旋转不变性，Ojala 等人提出了 LBP 旋转不变模式，即将求得的二进制数字进行循环移位，以其最小值作为该中心点的 LBP 特征值。如图 2-7 所示，8 种情况的 LBP 特征值均为 00001111 (15)。

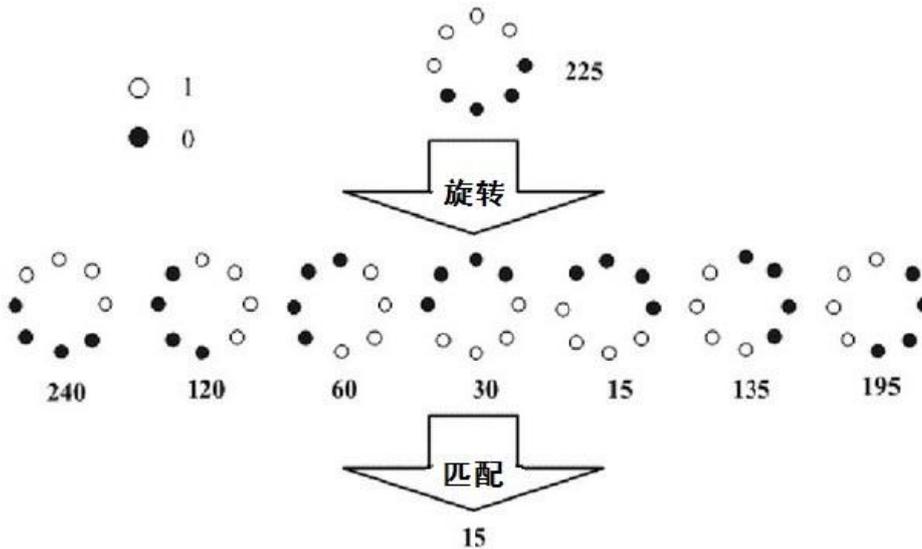


图 2-7 LBP 算子的旋转不变性

随着半径与采样点数目的不断增加，LBP 算子所产生的特征值空间也成指数倍增大。这样的二值模式对信息的提取和表达极为不利。为了解决此问题，Ojala 等人提出了 LBP 等价模式，对 LBP 算子进行降维操作。当一个 LBP 特征值最多只有两次 0/1, 1/0 的跳变时，就把其成为是一个等价 LBP 特征值。例如：00000000 (0 次跳变)、01110000 (2 次跳变)、11001111 (2 次跳变) 均为等价特征值，但 11001001 (4 次跳变)、01010011 (6 次跳变) 均不是等价特征值。

在计算 LBP 直方图时，所有等价特征值将被映射到一个固定的单值上。实验证明，LBP 等价模式有着极好的性能和出色的表现。

(2) 人脸 LBP 描述符

LBP 在特征分类问题中，由 LBP 特征值被映射成直方图，通过比较直方图的相似性来决定分类结果。与此相反，简单的将人脸的 LBP 纹理特征映射到直方图将会损失掉大量的位置和纹理信息。因此，更好的方法是通过 LBP 运算提取每一个局部区域的特征，然后将他们组合成一个全局特征描述符。这一方法有两点好处：其一，基于局部特征的提取方法可以对全局描述进行限制，使得其对位置、光照更加鲁棒；其二，基于局部特征的提取方法可以包含更多的信息，使其更具有代表性。

人脸 LBP 描述符由 Ahonen 在 2016 年提出^[44]。人脸图像被分割成局部区域并使用 LBP 等价算子分别提取特征。所有局部区域的特征拼接成为人脸的全局描述符，如图 2-8 所示。

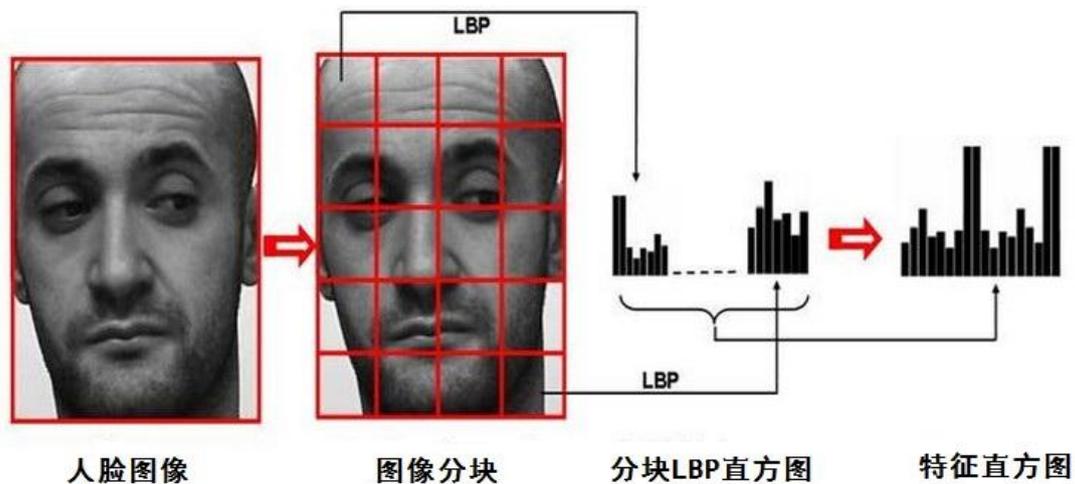


图 2-8 人脸 LBP 描述符的提取过程

人脸 LBP 直方图可以对人脸的信息进行三个级别的有效表示：LBP 特征值包含了样本的图案信息，局部直方图包含了样本的区域信息，整体直方图包含了样本的位置信息。需要注意的是在使用 LBP 特征描述符时，全局直方图不需要进行归一化，他们也不需要拥有同样的大小或尺寸，不同的区域允许部分重叠。

(3) 基于 LBP 特征的 AdaBoost 人脸检测算法

在 AdaBoost 算法中，模型选择在正负样本分布概率上最小化分类误差的弱分类器。在基于 LBP 特征的 AdaBoost 算法中，检测器在每个 24×24 的窗口中搜索人脸。在每一级的检查过程中，每个弱分类器检查某一特征点上的 LBP 特征值，并根据匹配程度给出一个可正可负的权重加至累加器，直至此级所有弱分类器检查完毕。当一级弱分类器全部衡量完成之后，将累加器的值与本级的阈值作比较：低于阈值，则拒绝此区域；高于阈值，则进行下一级的检查。直至所有级检测完毕。流程图如图 2-9 所示。

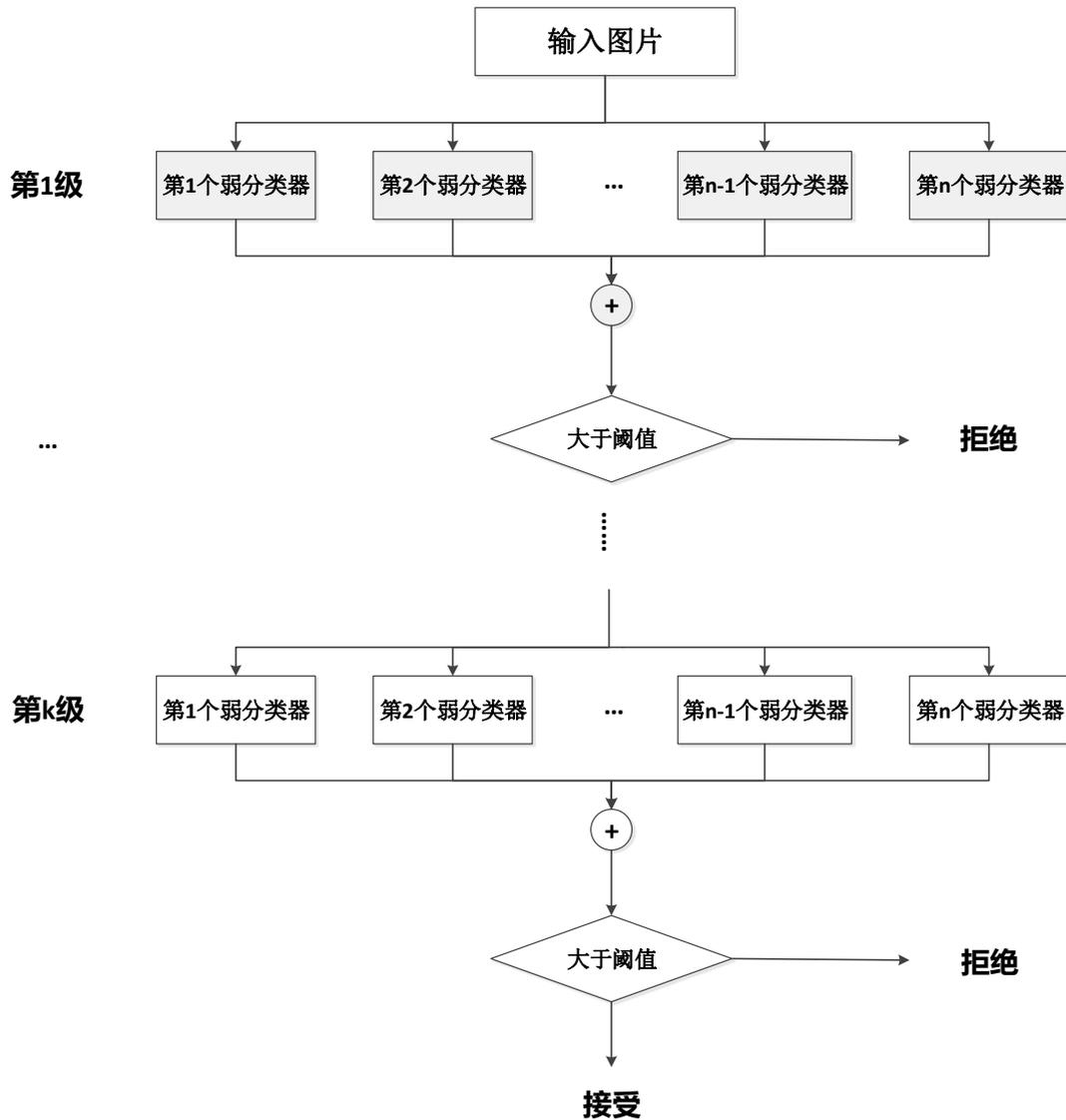


图 2-9 基于 LBP 特征的 AdaBoost 人脸检测算法流程图

2.1.4 实验仿真与性能对比

以上所述的两种算法虽然均基于 AdaBoost 学习算法，但由于所采用的特征不同，分类器训练与检测的效率有所区别。为了从速度和准确率两方面评价两种方法，本课题选用 MUCT 人脸数据库^[45]，本数据库包含 3755 张包含人脸信息的图片，它们采集自不同光照条件、年龄、种族与性别，非常具有代表性。部分图片如图 2-10 所示。



图 2-10 MUCT 数据库

通过调用 opencv 提供的已经训练好的 haarcascade 与 lbpcascade, 我们得到了以下结果:

表 2-1 AdaBoost 人脸检测实验结果

衡量指标	Haar-like	LBP
Hits	3635	3569
Misses	55	106
False Detects	63	77
Multi-hit	5	3
Time	4min2.060s	1min12.511s
Accuracy	96.804%	95.047%

通过实验可以观察到, 虽然基于 LBP 的 AdaBoost 检测器的人脸检出率略低于 Haar-like 特征, 但是速度却比 Haar Cascade 快近 3 倍, 对于实时系统来讲有着十分重要的意义。由于本次课题只关注静态图片中的人脸识别问题, 故采用准确率相对较高的 Haar Cascade。

2.2 基于深度学习的人脸关键点检测算法

人脸关键点检测是诸多人脸分析算法必不可少的组成部分。通过人脸检测得到包含人脸区域的样本之后, 我们需要找到其面部特征 (如: 眼角、眉毛、嘴角、鼻尖等的位置), 以对人脸的尺寸和位置进行归一化。关键点位置识别是否准确决定了人脸对齐的质量, 进而影响了之后一系列的人脸相关算法。

传统的人脸关键点检测算法有基于模板的检测方法^[46,47,48], 基于特征的检测方法^[49,50,51]等。近年来, 随着深度学习理论的不完善和应用的拓展, 卷积神经网络给计算机视觉的很多领域带来了革命性的突破。本课题中, 将采用卷积神经网络对人脸关键点进行建模。

2.2.1 人脸关键点

检测面部关键点是一项十分有挑战性的任务。每个个体的面部特征都有很大的区别, 此外, 即便是同一个个体, 面部特征也会随着位置、大小、姿态、角度和光照条件而改变。

人脸关键点检测问题是一个标准的多标签回归问题, 每一个需要预测的关键点都由一个有序实数对 (x, y) 给出, 标定该点在图像二维平面中的直角坐标。对于每张图片, 共给出 15 个关键点, 分别代表人脸的以下位置: 左眼中心、右眼中心、左眼内眼角、左眼外眼角、右眼内眼角、右眼外眼角、左眉眉梢、左眉眉尖、右眉眉梢、右眉眉间、鼻尖、左嘴角、右嘴角、上唇中心、下唇中心。图 2-11 为一个标出关键点的样本图片。

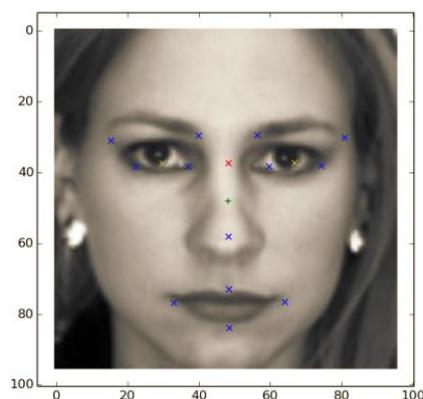


图 2-11 标有人脸关键点的样本

2.2.2 卷积神经网络原理

卷积神经网络 (Convolutional Neural Networks) 是受生物学启发而创造的多层感知器模型。1968 年 Hubel 和 Wiesel 通过研究猫的视觉皮层发现其包含一类结构复杂的细胞。这类细胞对于视野内的子区域十分敏感，这些子区域被命名为感受野 (Receptive Field)^[52]。感受野通过排列覆盖整个视觉区域。这些细胞相当于输入空间的局域滤波器，十分适合提取自然图像的空间区域相关信息。

受生物学启发，卷积神经网络通过邻接层之间的神经元传递区域空间信息。如图 2-12 所示，第 m 层神经元的输入来源于第 $m-1$ 层神经元输出的一个子集，这样神经元就有了空间连续的视野域。

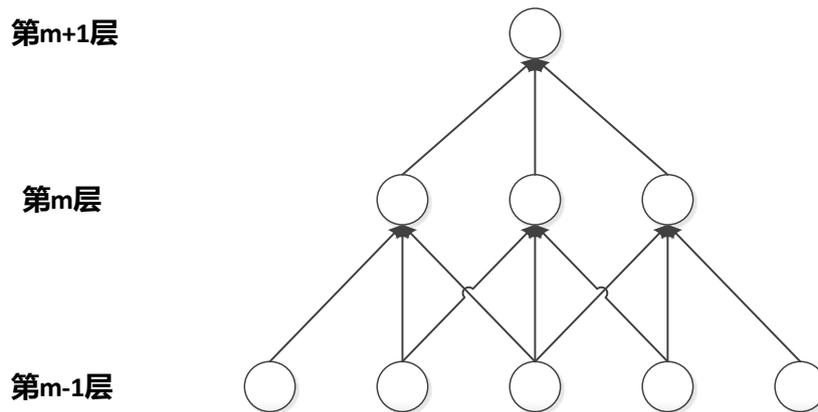


图 2-12 卷积神经网络中神经元的连接方式

一个完整的卷积神经网络模型由不同数量不同类型的层所组成。人脸关键点检测过程中用到的层如下：

(1) 卷积层 (Convolutional Layer)

卷积层是卷积神经网络的核心组成部分，它有着巨大的计算量。卷积层的参数为一组可学习的滤波器，即为卷积核 (Kernel)。对于一个输入，每个卷积核分别于其进行卷积，生成一组对应的特征空间 (Feature Map)，作为下一层的输入。卷积层的超参数有：卷积核的大小 (Kernel Size)、卷积核的个数 (Kernel Number)、跨度 (Stride) 和补零 (zero-padding)。它们决定了卷积层的输出空间维度。

(2) 池化层 (Pooling Layer)

池化层通常在连续的卷积层之间出现，它的主要功能是通过减少特征空间的维度从而减少网络的参数，降低计算复杂度，防止过拟合现象的出现。池化分为均值池化 (Mean Pooling) 和最大值池化 (Max Pooling) 两种类型，本课题中使用的均为最大值池化。它通过提取一个矩形区域内的最大值，从而使用一个元素代表矩形区域，实现了对特征空间的减采样。

(3) 激活层 (Active Layer)

激活层运算按元素进行操作，输入空间与输出空间维数相同。他们通过将输入进行函数变换映射到一定范围的空间中去，减少异常值的出现。常用激活函数有：Sigmoid 函数、TanH 函数、ReLU 函数等。

(4) 全连接层 (Fully-connected Layer)

全连接层的神经元与上一层的所有激活元素相连接。因此它们的运算过程可以看做是

一个矩阵乘法之后加一个偏移量。

(5) 损失函数层 (Loss Layer)

损失函数通过计算网络输出与真实标签的差别确定后向反馈的梯度，对整个网络进行后向传导调整参数。常见损失函数有：Softmax、Hingeloss、Infogain、Euclideanloss 等。在回归问题中，通常使用 Euclidean 损失函数，计算公式如下：

$$\frac{1}{2N} \sum_{i=1}^N \|x_i^1 - x_i^2\|_2^2 \quad (2-12)$$

2.2.3 模型结构

通过不断增加模型的复杂度，可以找到模型容量适中且准确率较高的模型。在进行人脸关键点检测的任务中，将基于以下三种模型进行网络的训练与测试。

(1) 具有一个单隐层的神经网络

单隐层神经网络由三个基本结构组成：输入层、隐层、输出层。其中：

- 输入层为训练集的图片与正确标签。
- 隐层由一个全连接层与一个激活层组成，全连接层通过对输入层进行线性变换得到一个指定长度的特征向量，特征向量中的每个值都由输入层经线性变换得到。激活层采用 ReLU 函数，即为

$$y = \max(0, x) \quad (2-13)$$

函数图像如图 2-13 所示。其作用在于对特征向量进行非线性变换，激活活跃神经元。

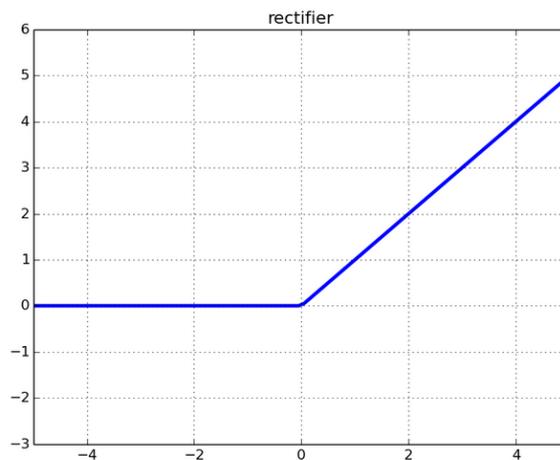


图 2-13 ReLU 函数图像

- 输出层：输出层为一个具有 30 个输出的全连接层，用于得到 15 对人脸关键点的 x、y 坐标

网络的损失函数由欧几里得距离得到，并通过随机梯度下降算法找到后向反馈的方向，对参数进行修正。总体网络结构如图 2-14 所示。

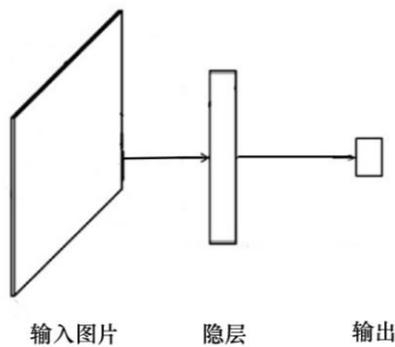


图 2-14 单隐层神经网络结构图

(2) 卷积神经网络

卷积在神经网络中的应用使得深度学习在计算机视觉领域取得了极大的进步^[53]。卷积层不同于全连接层，它使用了一些技巧去减少自身的参数，但同时保持着较高的性能。对于一个图像，2D 空间卷积可以在提取图像信息的同时保留其位置信息，使得网络得到更高级的特征。

本次实验所用到的网络结构为三层卷积神经网络。一个典型的卷积结构包括卷积层、激活层、池化层，和一个 Dropout 层。前三层的功能已在上一小节进行叙述。Dropout 层的使用是为了防止网络出现严重的过拟合现象，增加训练模型的通用性。

总体网络结构如图 2-15 所示：

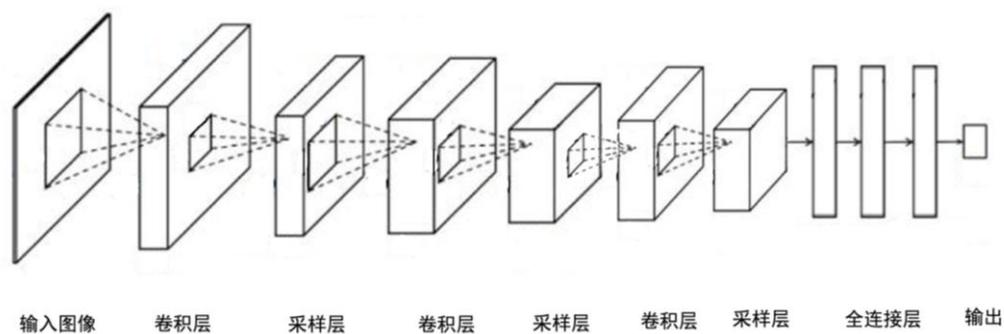


图 2-15 本实验网络结构图

(3) 专家模型

通过观察人脸关键点的分布，可将 15 个关键点分成 6 类，分别对应不同的面部器官。通过分别训练不同的专家模型，可以有效提高模型的准确率，同时能对训练数据进行充分的利用。本课题将 15 个关键点分为以下 6 类，见表 2-2。

表 2-2 关键点分组表

分组序号	关键点
1	左眼中心、右眼中心
2	鼻尖
3	左嘴角、右嘴角、上唇中心

分组序号	关键点
4	下唇中心
5	左眼内眼角、左眼外眼角、右眼内眼角、右眼外眼角
6	左眉眉尖、左眉眉梢、右眉眉尖、右眉眉梢

每个专家模型的网络结构与上述卷积神经网络结构基本保持一致，唯一的区别在于最后一层的全连接层输出由当前模型关键点的个数决定。

2.2.4 实验仿真与性能对比

人脸关键点检测算法采用 Kaggle 的 Face Keypoints Detection 比赛所提供的数据集。训练集包括 7049 张灰度图片，图像尺寸为 96*96。其中，只有 2140 张图片标注完整，带有 15 对关键点坐标，其余图片均由信息缺失。故在单隐层神经网络与卷积神经网络的训练过程中，只用到了 2140 张标注完整的图片。而在专家模型中，每个网络会用到本分组关键点标注完整的所有数据。

验证集由 1783 张标注完整的图片组成，图像预处理方式与训练集保持一致。训练过程涉及到的参数见表 2-3。

表 2-3 模型学习参数与训练结果

参数	单隐层神经网络	卷积神经网络	专家模型
优化算法		随机梯度下降	
初始学习率		0.01	
网络损失	0.00374897	0.00172055	0.00091650
均方根误差	2.940	1.979	1.453

2.3 人脸对齐算法

许多研究都表明，人脸对齐在人脸识别系统中起着重要的作用。人脸对齐的实现基于人脸关键点的检测。结合图像中的人脸关键点信息，算法通过为每个点选择理想的坐标，进行仿射变换将人脸映射到预设的平均脸模型之中。

2.3.1 仿射变换

在几何上，仿射变换被定义为两个可以保持点、线与平面空间关系的二维平面之间的变换。由于仿射变换是一个线性变换，任意一对平行直线在经过仿射变换后仍然保持平行直线关系，但相交线之间的夹角、点之间的距离则均有可能发生改变。仿射变换可用如下公式表示：

$$\begin{cases} x' = ax + by + m \\ y' = cx + dy + n \end{cases} \quad (2-14)$$

其矩阵表示为：

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} a_2 & a_1 & a_0 \\ b_2 & b_1 & b_0 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (2-15)$$

通过仿射变换，可以实现包括平移、旋转、缩放、翻转、错切这五种几何变换功能^[54]。如图 2-17 所示：

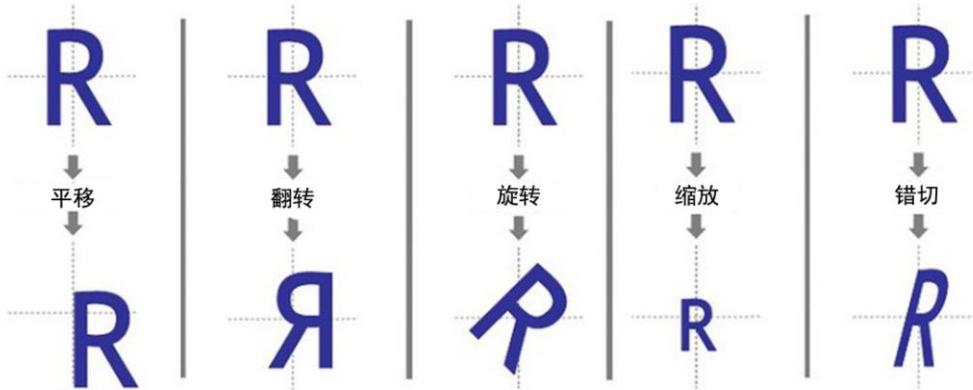


图 2-17 仿射变换的五种功能

2.3.2 实验仿真与性能对比

借助已经得到的人脸关键点坐标，可以将未归一化的人脸图像仿射变换到定义好的模型上^[55]。Eidiingerr 和 Enbar 等人在^[56]中使用了基于人脸关键点的放射变换进行人脸对齐。经过对齐，人脸图片质量的到较大幅度的提高，图 2-18 给出了部分图片处理前后的对比图。

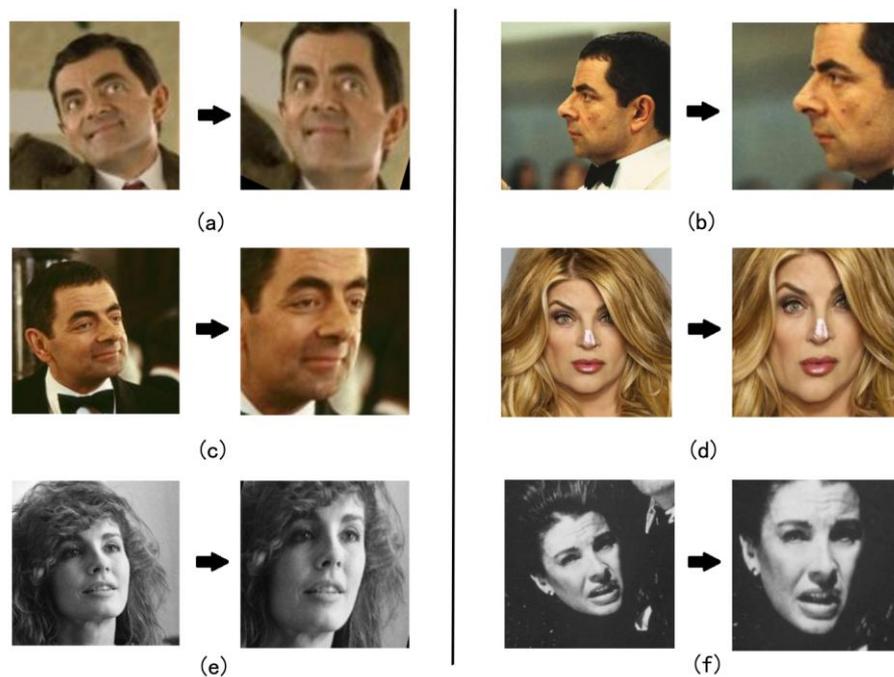


图 2-18 处理前后对比图

2.4 本章小结

本章对人脸定位处理算法做了较为全面的研究与讨论，具体涉及：人脸检测算法、人脸关键点检测算法、人脸对齐算法，并通过实验对每个算法进行了实现与测试。

在人脸检测的研究中，本章分别介绍了基于 Haar-like 特征和基于 LBP 纹理特征的 AdaBoost 算法。两种算法虽然使用相同的学习方式，但却在效率和准确率上有所区别。通过在 MUCT 数据集上进行实验，可以观察到基于 Haar-like 特征的 AdaBoost 算法虽然速度较慢，但准确率和误报率相对较低。由于本次实验关注静态图片中的人脸识别问题，故可离线处理所有实验样本，因此最终选用基于 Haar-like 特征的 AdaBoost 算法作为人脸检测器。

不同于传统的模型或特征匹配检测机制，本章使用了三种不同的神经网络结构训练人脸关键点模型，分别为：单隐层神经网络、包含三个卷积层的神经网络、专家网络。实验表明，随着网络的不断加深与模型结构的优化，模型在验证集上的正确率逐渐增加，性能得到较大幅度的改善。同时，也验证了深度学习在人脸关键点检测算法中的可行性。

本章所讨论的人脸对齐算法以几何仿射变换作为基础，通过以上两个步骤所得到的信息将样本映射到标准人脸模型上。实验证明，人脸对齐算法可以有效的对人脸图像进行正规化，达到了较好的图像定位处理的效果。

结合以上研究与实验结果，本课题最终使用的人脸定位处理算法构架如图 2-19 所示：

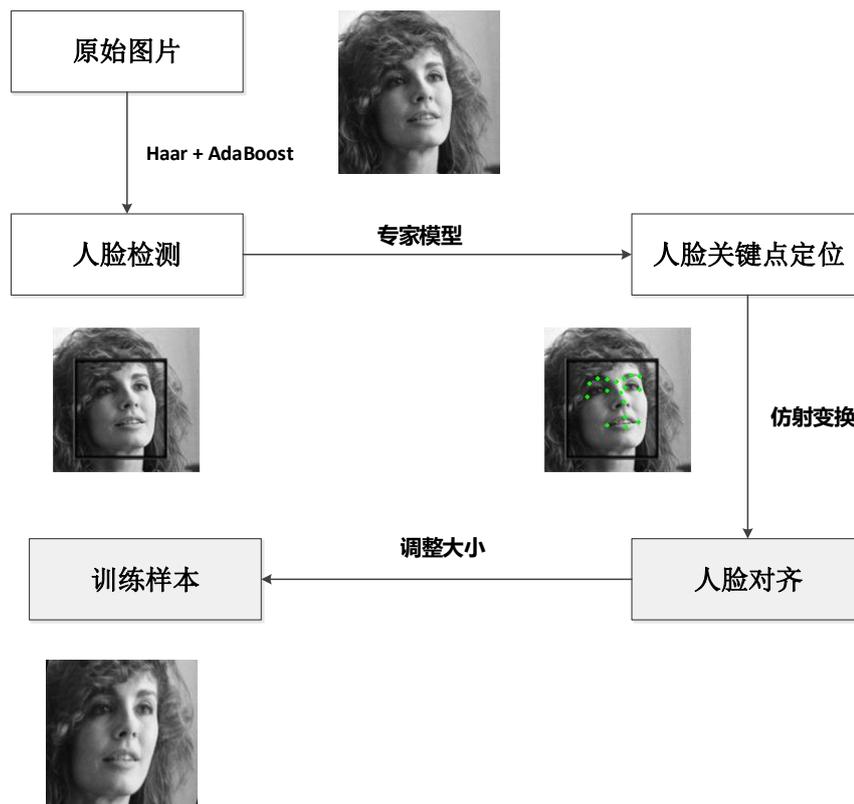


图 2-19 人脸定位算法总体构架

第三章 静态图片中的人脸识别算法研究

人脸识别技术是指借助计算机技术对图像或视频中的人脸视觉信息进行分析，以达到身份鉴别或身份查找的目的。它经常被使用于安全系统中，并且可以和其他生物信息识别技术（如：指纹识别、虹膜识别等）相结合。近几年来，众多商业系统也利用此技术作为营销工具，创造市场利润。

3.1 人脸识别算法概述

人脸识别技术于 20 世纪 60 年代开始兴起，但由于技术所限一直停留在科学研究阶段。直至 80、90 年代，计算机技术与影像技术的发展使人脸识别技术得到了本质上的突破，真正可以用于投入实际使用的算法不断被科学家提出。

目前人脸识别算法可分为以下几类：

(1) 基于几何的人脸识别算法^[57]：通过分析面部关键点特征和他们的几何关系分析识别人脸。这种识别算法通常依赖于人脸关键点检测算法，且对人脸的位置、角度、姿态等变量极为敏感。

(2) 基于特征的人脸识别算法^[58]：通过使用特征算子提取面部特征来提取特征向量，比较空间距离从而达到识别人脸的目的。常用特征有：LBP、SIFT、SURF、Gabor 等。

(3) 基于统计学的人脸识别算法^[59,60]：有时被称为基于模型的人脸识别算法。借助统计学原理总结训练集的知识，将人脸信息汇总到一个模型中，用于后续的使用。常见方法有：支持向量机、贝叶斯模型、迭代决策树等。

(4) 基于深度学习的人脸识别算法^[61]：使用卷积神经网络对大规模数据进行学习，通过复杂的网络结构提取人脸的高级特征，得到最终的人脸识别模型。深度学习在人脸识别算法中的应用使得最终的准确率有大幅度的提高，是目前 state of art 的算法。

人脸识别算法根据其任务目标的不同而分为两类：

(1) 基于人脸识别的身份查找算法 (Face Identification)：身份查找需要将输入的一张人脸图像与后台数据库中的所有图像进行对比，完成身份匹配任务。此种算法对应一对多进行图像比较的过程。

(2) 基于人脸识别的身份确认算法 (Face Verification)：身份确认指给定的两张人脸图像，通过分析对比其面部特征的相似程度而判定图像是否来自同一个人的技术。此种算法对应一对一进行图像比较的过程。

本章的后续小节将分别对实现两种不同任务的人脸识别算法进行探究。

3.2 基于人脸识别的身份查找算法

对于数据库已经确定的人脸识别问题，可以借助人脸识别算法完成身份查找任务。算法目标可简化为：已知数据库中存储有 N 个不同身份的人，通过对比测试样本与数据库中每个人的特征相似性，找到与其最匹配的身份并返回。该算法可以用于考勤系统、监控系统、门禁系统等多种实际应用系统中。

本课题将分别研究基于特征脸 (Eigenfaces) 的身份查找算法与基于卷积神经网络 (Convolutional Neural Network) 的身份查找算法。为了进一步提高准确率，将借助 VGG

提供的人脸检测模型作为预训练模型进行再次训练，得到更加具有普适性的网络模型。

3.2.1 基于特征脸与支持向量机的身份查找算法

(1) 特征脸

特征脸于 1987 年被 Sirovich 和 Kirby 两位学者用于计算机视觉领域^[62]。1991 年，Matthew 和 Alex 两人将其应用于人脸识别算法中^[63]，并证明其为一种可工程化的算法。特征脸通过提取人脸图像所组成的高维向量空间概率分布的协方差矩阵而得到。特征脸是用于构建协方差的所有图片所组成空间的一组基底 (Base)。

特征值的提取算法如下：

令 $X = \{x_1, x_2, \dots, x_n\}$ 为 N 个训练样本所构成的观察向量，

a. 计算平均值 μ

$$\mu = \frac{1}{N} \sum_{i=1}^n x_i \quad (3-1)$$

b. 计算协方差矩阵 S

$$S = \frac{1}{N} \sum_{i=1}^n (x_i - \mu)(x_i - \mu)^T \quad (3-2)$$

c. 计算特征值 λ_i ，和 S 的特征向量 v_i

$$Sv_i = \lambda_i v_i \quad i = 1, 2, \dots, n \quad (3-3)$$

d. 按照特征值降序排列特征向量。选择 K 个最大的特征值所对应的向量作为主成分。得到 K 个特征值之后，训练样本将由得到的 K 个基底所重建：

$$x = Wy + \mu \quad (3-4)$$

其中， $W = (v_1, v_2, \dots, v_k)$ 。

(2) 支持向量机

支持向量机于 1995 年由 Cortes 和 Vapnik 提出^[64]。由于其完备的理论基础和良好的实验效果，支持向量机成为机器学习领域的热门学习算法。它通过构建一个或一组在高维空间或无穷维空间的超平面而对回归或分类问题进行监督学习。

本次实验用到的是以径向基函数(Radial Basis Function)为函数核的非线性支持向量机。通过对样本进行非线性映射，可以将部分在原空间线性不可分的向量转化为线性可分的向量，通常对于模式识别问题有着更好的效果。

对于两个输入空间中的样本 x 与 x' ，其核函数定义为：

$$K(x, x') = e^{-\frac{\|x-x'\|^2}{2\sigma^2}} \quad (3-5)$$

可以注意到，RBF 核函数随着距离的增大而减小，且值域为[0,1]。在训练过程中，支持向量机通过最大化下述公式去最大化不同类样本之间的距离，从而将样本分类的目标。

$$D(\alpha) = \left(\sum_{i=1}^n \alpha_i \right) - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (3-6)$$

使其满足：

$$\sum_{i=1}^n \alpha_i y_i = 0 \text{ and } 0 \leq \alpha_i \leq C \quad (3-7)$$

在测试时，对于测试样本 x ，

$$h(x) = \text{sign} \left(\sum_{x_i \in SV} \alpha_i y_i K(x_i, x) + b \right) \quad (3-8)$$

(3) 基于特征脸的人脸识别算法

算法步骤如下：

- a. 将所有训练样本投影到特征向量所组成的子空间
- b. 将待测试样本投影到特征向量所组成的子空间
- c. 运用支持向量机训练样本，得到模型参数
- d. 通过模型对测试样本进行运算，获得分类标签，得到结果

3.2.2 基于卷积神经网络的身份查找算法

近年来随着社交网络的兴起，越来越多的人脸图像出现在网络中。在搜索机器的帮助下，大规模的极具差异的人脸图像被下载并整理成人脸数据库，供学术研究所用。大量的数据有利于训练得到更加具有统计特性、更加普适的模型，但却也无法避免的增加了计算复杂度，对计算资源形成了极大的挑战。传统学习模型，如支持向量机、主成分分析、线性判别式分析等方法均无法支持大规模训练。深度神经网络的出现为解决这一问题提供了很好的方案。

(1) Alexnet

在 2012 年的 ImageNet 比赛中，Alex Krizhevsky 等人凭借自己设计的网络一举拿下冠军^[65]。他们的网络训练样本由分属 1000 类的 120,000,000 张图片组成，最终 top-1 错误率为 37.5%，top-5 准确率为错误率为 17.0%，他们的网络结构被称为 Alexnet。经过大量实验表明，Alexnet 对分类问题有着很好的普适性，因此本课题也采用 Alexnet 作为基本网络结构，并根据训练集的大小、标签数目等做一定的调整。

Alexnet 的基本结构如图 3-1 所示。

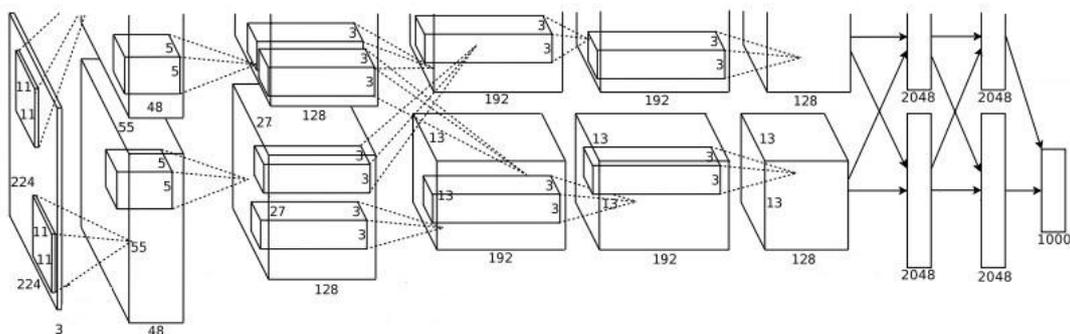


图 3-1 Alexnet 的基本结构

如上述结构图所示，网络包括五个卷积层和三个全连接层，最后一个全连接层的输出是一个 1000 维的向量，softmax 层接受向量并根据分布判断 1000 类 label 的最终概率值。网络通过最大化类间的对数似然函数与正确标签间分布的相似性而后向传播调整参数。

整个训练过程由两个 GPU 合作完成。其中，第 2、4、5 个卷积层分别只与同属于 GPU 的前一层连接。第 3 个卷积层与所有全连接层则与两个 GPU 的所有上层神经元相连，这样保证了两个 GPU 既各有分工，又同时保持了网络的连通性。

(2) 基于 Alexnet 的改进神经网络结构

由于 Alexnet 的训练集为百万级，故网络参数极多，训练耗时较长。为了使之适应本课题所采用的 CASIA 数据集，对传统的 Alexnet 进行了以下改进：

- 网络输入样本大小改为 64*64 的 RGB 图像
- 因网络输入减小，五个卷积层中卷积核分别减小至 7、3、5、3、3。
- 使用单 GPU 进行训练，第 1、2 个全连接层输出改为 2048
- 由于数据集最终分 925 类，因此最后一个全连接层输出改为 925

最终网络结构如图 3-2 所示。



图 3-2 基于 Alexnet 的改进网络结构

3.2.3 基于卷积神经网络的身份查找算法的改进——转化学习

转化学习 (Transfer Learning) 是当前深度卷积网络研究领域的热门话题。在实际生活中，很少有人拥有规模足够大、质量足够高的训练集去从头训练一个神经网络，因此，在一个规模巨大的数据集 (如 ImageNet) 上进行预训练，然后再根据自己的需求进行 finetune 成为了研究人员更为青睐的方法。为了弥补数据量不足的缺陷，本课题采用 Visual Geometry Group 训练的 VGG_Face 模型作为预训练模型^[66]，沿用 VGG 的网络结构进行 finetune。

(1) VGG Face

Visual Geometry Group 受到深度神经网络的启发，旨在基于大数据训练一个可以有效提取人脸特征的卷积神经网络。其训练样本包括 2622 个人的 260,000,000 张图片，数量远大于现有的公开人脸数据集。



图 3-3 VGG Face 的网络结构

VGG Face 的网络结构如图所示。虽然大体网络结构仍然沿用自 Alexnet，但 VGG Face 仍然有其较为独特的地方。首先，卷积层的卷积核大小均为 3*3，较小的卷积核可以更好的

提取图像的空间信息。其次，每个卷积层的卷积核数目递增，以抽取更多的高级特征。网络的最后一层全连接层输出改为 2622，输入给 Softmax 层进行分布概率判断。Visual Geometry Group 公开了他们训练好的模型，本课题将以此模型作为预训练模型，为模型进行参数初始化。

(2) 模型转化学习

基于大数据进行训练的 VGG Face 模型对人脸特征有着较好的提取功能，因此在转化学习的过程中，将关闭所有卷积层与前两层全连接层的学习功能，只对最后一层全连接层进行 finetune。网络训练的结构与预训练保持一致，除最后一层全连接层输出更改为 925 并进行参数初始化，输出向量代表需要判断的 995 个身份标签的概率分布。模型结构如图 3-4 所示。

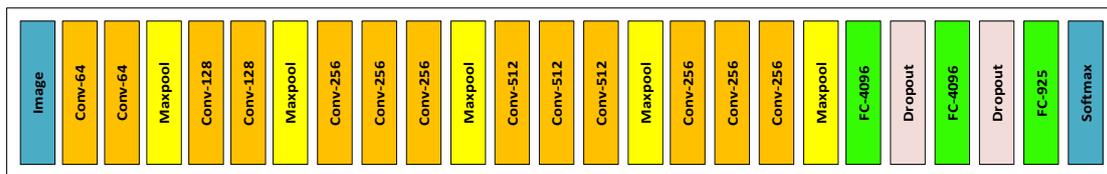


图 3-4 本实验所用的网络模型

在训练过程中，学习速率需要降低至原来的百分之一。

3.2.4 实验仿真与结果分析

鉴于传统学习模型与神经网络模型容量的区别，本实验分为两部分进行。

(1) 基于 LFW 数据集的传统学习方法

Labeled Faces in the Wild (LFW) 数据集^[67]是为非限制条件下的人脸识别问题所收集整理的数据集。数据集共有 13000 张从网络上收集的人脸图片，来自 5749 个不同人物，每张图片以人名作为标签。其中，单类样本总量超过 50 张的人共 12 个。为确保支持向量机学习到有效信息，训练与测试过程中只是用了这 12 个标签下的图片，共 1560 张图片，其中 1170 张作为训练集，390 张作为验证集。

通过选取特征值最大的 150 个特征向量，可以重构 12 张特征脸。重构图像如图 3-5 所示。

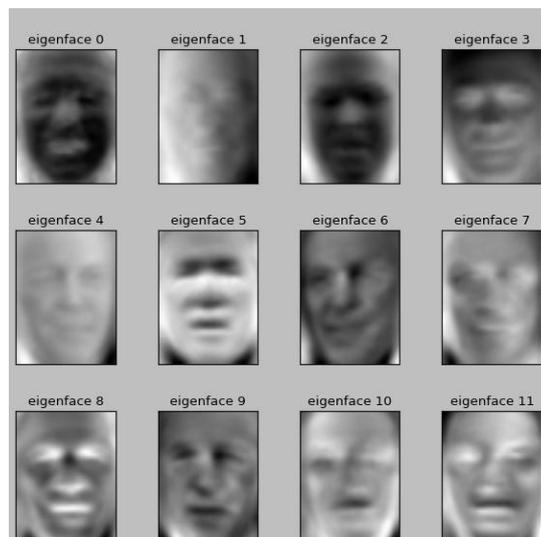


图 3-5 重构的特征脸图像

训练结束后，使用验证集对模型进行验证，结果如下表所示：

表 3-1 验证集结果

Label	precision	recall	F1-score	support
Ariel Sharon	0.75	0.75	0.75	16
Colin Powell	0.79	0.80	0.80	66
Donald Rumsfeld	0.88	0.85	0.87	27
George W Bush	0.79	0.93	0.86	140
Gerhard Schroeder	0.78	0.58	0.67	24
Hugo Chavez	0.76	0.76	0.76	17
Jacques Chirac	0.75	0.30	0.43	10
Jean Chretien	0.68	0.93	0.79	14
John Ashcroft	1.00	0.69	0.81	16
Junichiro Koizumi	1.00	0.73	0.84	11
Serena Williams	0.88	0.78	0.82	9
Tony Blair	1.00	0.80	0.89	40
avg/total	0.83	0.82	0.81	390

最终模型准确率为 83%。由于支持向量机模型容量有限，当训练样本过多时，将没有办法直接求得解析解，故本课题未再增加样本数量。可以观察到，传统模型在大数据的处理方面有着极大的缺陷。

(2) 基于 CASIA 数据集的神经网络方法

CASIA WebFace 数据集^[68]是一个包含有 494,414 张人脸图片的开源数据集，数据采集自 10,575 个不同的人，对训练卷积神经网络有着极大的帮助。

学习到有效特征，影响分类结果，因此在创建 LMDB 时，去掉了所有样本不足 100 的组别，最终得到一个拥有包含 925 个不同标签的 146,436 张图片，其中 145,511 张（80%）作为训练集，36,390 张（20%）作为验证集。训练集与验证集中的每个标签样本比例如图 3-5 所示。

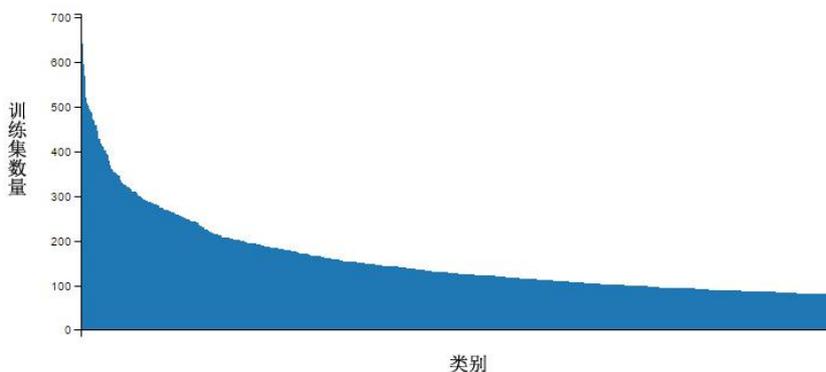


图 3-5(a) 训练集各标签样本数量

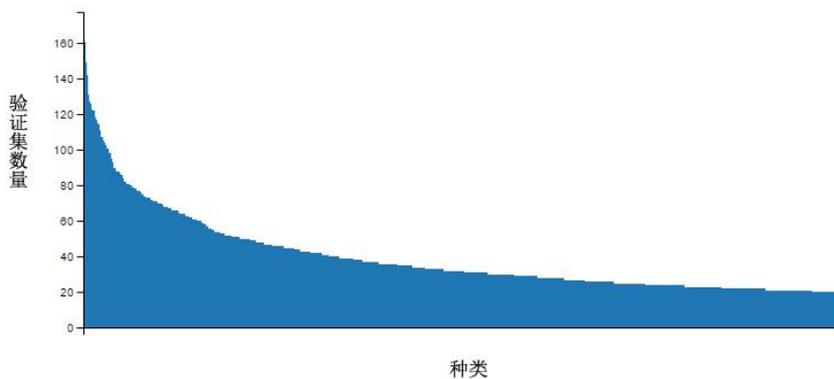


图 3-5(b) 验证集各标签样本数量

训练集平均脸图像如图 3-6 所示：

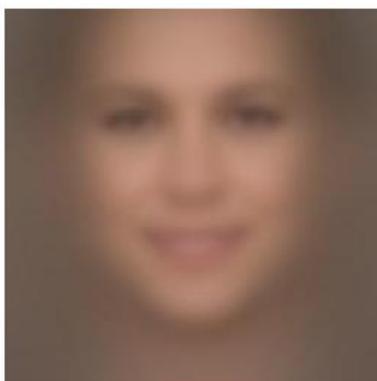


图 3-6 训练集平均脸

基于 Alexnet 的网络训练过程如图 3-7 所示。最终模型的准确率为 69.9258%。

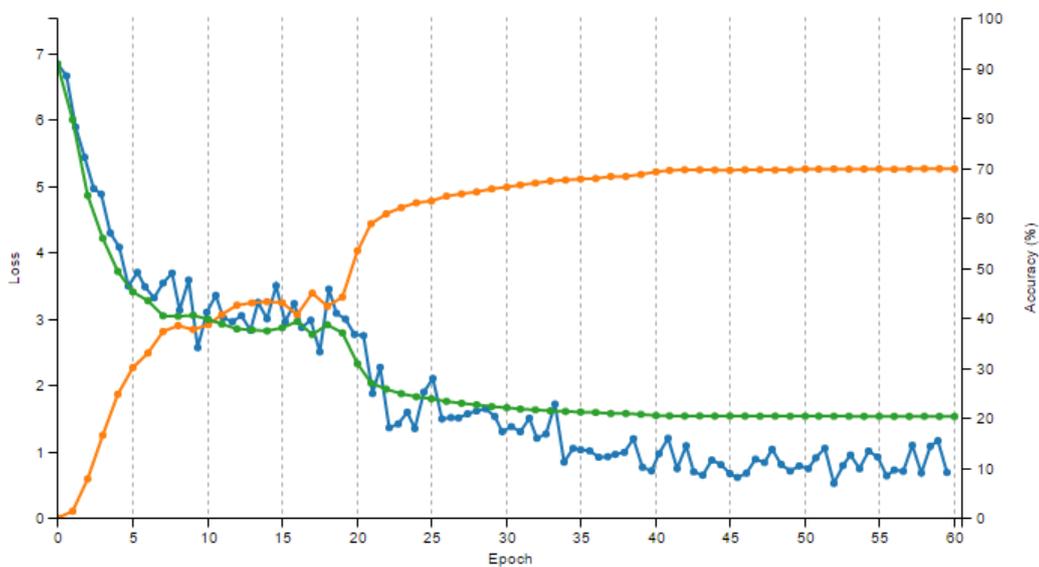


图 3-7 本实验模型训练曲线

(3) 基于 CASIA 的转化学习算法

转化学习所用到的数据集和(2)完全保持一致。网络初始化除最后一个卷积层外全部使用 VGG Face 提供的 caffemodel, 最后一个卷积层则做具有高斯分布的随机初始化。所有卷积层的参数被固定, 不再进行学习, 而三个全连接层则打开学习参数, 重新学习组合特征。由于网络直接进行转化学习, 故学习率降为 0.001, 原来的十分之一, 以降低对网络结构可能造成的破坏。

VGG 网络输入较大, 参数较多, 极其耗费内存和时间。本次实验在 GeForce GTX TITAN X 上进行, 共耗时 18 小时 12 分钟, 训练过程如图 3-8 所示。最终 925 分类网络准确率达到 83.1%, 充分体现了卷积神经网络的优势。

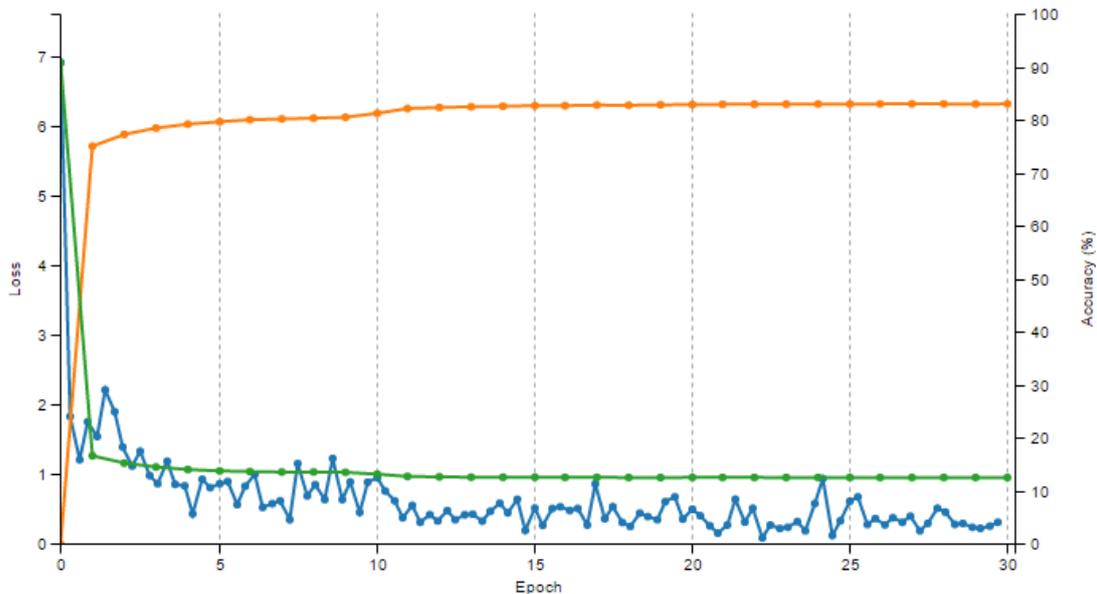


图 3-8 本实验模型训练曲线

3.3 基于深度学习的身份确认算法

身份确认算法是卷积神经网络最成功的应用之一。它用于判断两个人脸图像是否来自同一个人, 即进行图像的一对一匹配。用于身份确认的卷积神经网络通常被分成两组。

第一组方法借助多分类问题所训练的卷积神经网络提取面部特征向量, 然后通过对比向量进行二元判断。这种方法通常基于数据分布均衡的假设, 较难适应复杂的情况。

第二组方法旨在与直接优化匹配/非匹配样本对的误差。此方法可以有效地解决未知样本的分类问题。

本小节将分别针对两种算法进行研究, 并通过实验对比分析其性能。

3.3.1 基于特征提取的身份确认算法

随着卷积神经网络在识别算法中的广泛应用, 更多的人开始关注 CNN 特征在工程问题上的应用。2014 年 Girshick 等人提出了 R-CNN, 他们通过监督学习将神经网络训练成为一个黑盒特征提取器, 而后经过支持向量机进行分类。他们的系统在 PASCAL VOC 2012 数据集上取得了 state of art 的结果, 同时也证明了 CNN 特征的有效性, 引起了人们的广泛关注。

和其他传统特征(如: SIFT、SURF、HOG、LBP 等)相比, CNN 特征难以通过人类的知识进行理解, 但却有着数学意义。CNN 特征的提取基于网络的预训练, 根据需要完成任务的不同, 卷积神经网络可以通过调整参数最小化损失函数, 使其卷积层成为专用特征

提取器。同时，卷积神经网络可以提取比传统特征更高级别的特征，在应用中发挥着极大的作用。

身份确认算法通常用于判断网络之前没有见过的两张包含人脸信息的图像是否来源于同一个个体。这就要求网络有能力判断没有经过训练的数据，给出精确可靠的结果。通过训练一个可以提取丰富人脸特征的网络，对未知图片进行有效的特征提取，是解决身份确认问题的一个很好的办法。

得到两个测试样本的 CNN 特征之后，需要用距离函数计算其在同一空间中的距离，并找到阈值进行判断。若距离超过阈值，则认定两个样本来源于不同的两个人；否则，认为其来源于一个人。常用的距离函数有以下三种：

(1) L1 归一化距离 (L1 Normalized Distance)

L1 归一化距离首先对两个需要对比的空间向量求 L1 范数，然后分别使用其 L1 范数进行归一化，最后用欧几里得距离计算两个归一化向量之间的距离。计算公式如下：

$$D_{L1}(a, b) = \sum_{i=1}^n \left(\frac{a}{\|a\|_1} - \frac{b}{\|b\|_1} \right)^2 \quad (3-9)$$

(2) L2 归一化距离 (L2 Normalized Distance)

L2 归一化距离通过 L2 范数对两个需要对比的空间向量进行归一化，然后在用欧几里得距离计算两个归一化向量之间的距离。计算公式如下：

$$D_{L2}(a, b) = \sum_{i=1}^n \left(\frac{a}{\|a\|_2} - \frac{b}{\|b\|_2} \right)^2 \quad (3-10)$$

(3) 余弦距离 (Cosine Distance)

余弦距离用于衡量两个向量在内积空间的夹角的余弦值。当其夹角为 0° 时，余弦值为 1。其计算公式为：

$$\cos(a, b) = \frac{\sum_{i=1}^n a_i b_i}{\sqrt{\sum_{i=1}^n a_i^2} \sqrt{\sum_{i=1}^n b_i^2}} \quad (3-11)$$

然后用以下公式将余弦距离归一到 [0, 1]：

$$\text{similarity}(a, b) = 0.5 + 0.5 * \cos(a, b) \quad (3-12)$$

3.3.2 基于对称卷积神经网络的身份确认算法

对称网络 (Siamese Network) 用于从数据中学习相似性矩阵，它通过将样本通过函数映射到一个可以通过 L1/L2 范数衡量其距离的目标空间而到达判断一对输入的对比结果的目的。这种网络特别是适用于训练集中每一类样本较少，且验证集中的类别在训练集中从未出现过的情况。在模型学习过程中，网络通过最小化区别损失函数 (Discriminative Loss Function) 调整相似性矩阵，使得相匹配的样本之间的距离范数足够小，不匹配的样本之间的距离范数足够大。对称网络的特点十分符合身份确认算法的需求，是一个极为成功的应用案例。

人脸身份确认算法的关键在于将疑似图像与已知图像进行对比，决定拒绝还是接受此图像。它有两个衡量标准：假接受率 (False Accepts Rate)、假拒绝率 (False Rejects Rate)。一个好的系统应该能同时最小化这两个参数。通过卷积神经网络将原始图片的像素点进行非线性变换到一个低维空间。学习相似性矩阵的对称卷积网络将两个结构完全相同、共享

权值的网络并联，通过 Contrastive Loss Function 进行优化。

(1) 对称卷积神经网络

典型的对称卷积神经网络结构如图 3-9 所示。

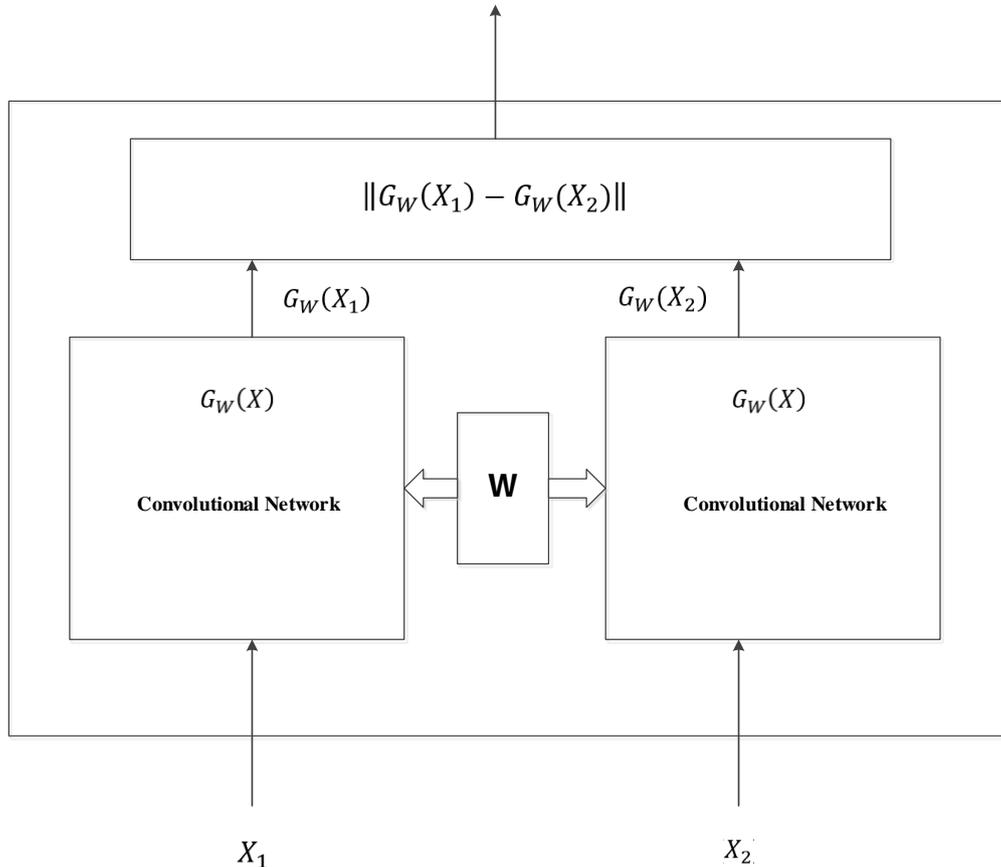


图 3-9 对称卷积神经网络结构

令 X_1 与 X_2 为输入的一对图像， Y 为二元标签，当 X_1 与 X_2 来自同一个人时 $Y=0$ ，否则 $Y=1$ 。 W 为共享参数向量，即为学习过程中需要调整的参数。令 $G_w(X_1)$ 与 $G_w(X_2)$ 为 X_1 与 X_2 在低维空间中的映射向量，则系统可将 $E_w(X_1, X_2)$ 看做是能量函数，用于衡量其距离，定义如下：

$$E_w(X_1, X_2) = \|G_w(X_1) - G_w(X_2)\| \quad (3-13)$$

(2) 对比损失函数 (Contrastive Loss Function)

对比损失函数用于优化对称卷积神经网络的参数。其计算公式如下：

$$E = \frac{1}{2N} \sum_{n=1}^N (y)d^2 + (1-y)\max(\text{margin} - d, 0)^2 \quad (3-14)$$

其中，

$$d = \|a_n - b_n\|_2 \quad (3-15)$$

(3) 网络结构

为了将原始数据映射到一个低维空间，并且通过学习得到一个相似度矩阵，需要使用

两个结构完全相同且参数共享的神经网络。通过对神经网络进行端到端的训练，可以将低级的像素级特征转化为高级的卷积神经网络特征。训练过程中所使用的模型需要经过预训练以保证可以获得有效的特征。

在两次训练时，使用预训练好的模型对两个并行的神经网络进行初始化。对比损失函数用于连接两个网络的最后一层，以计算两个特征向量的相似度，给出最终的分类结果。

本次实验基于 VGG Face 模型进行，具体网络结构如图 3-10 所示：

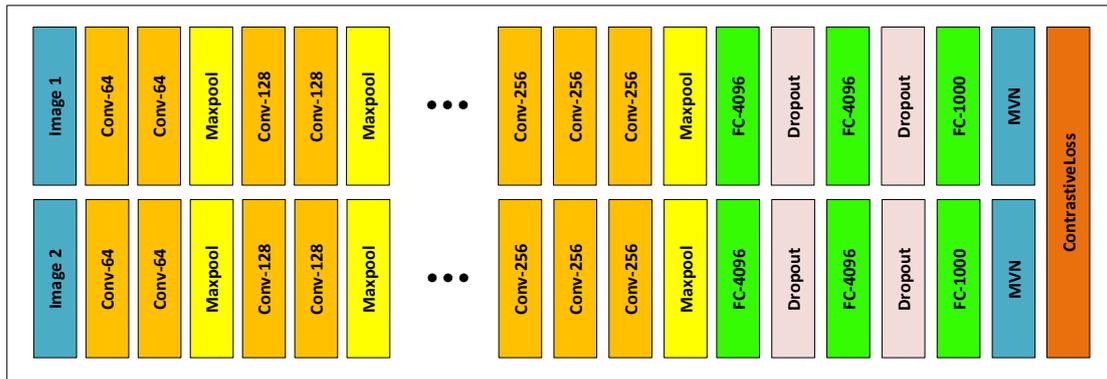


图 3-10 本实验所用网络结构

3.3.3 实验仿真与结果分析

(1) 基于特征提取的身份确认算法

如本章节所述，VGG Face 是基于大规模数据训练的，专门用于完成人脸识别任务的网络。同时，它也是一个很好的 CNN 特征提取器，因此本次实验选用 VGG Face 所提供的 caffemodel 作为特征提取器，选取 LFW 数据集作为验证集。

首先，LFW 中的所有图片经过对齐之后被送入 caffemodel，提取网络的最后一层全连接层，共 2622 维。测试图像对由 LFW 数据集中 verification 的标准测试集给出，共 3000 对正样本，即来自于同一个人的图像对；3000 对负样本，即来自不同人的图像对。分别计算其 L1 归一化距离、L2 归一化距离、余弦距离。

得到每对样本之间的距离后，需要找到其最合适的阈值。本次实验通过绘制 AUC 曲线来找到阈值，计算准确率。

接受者工作特性曲线（Receiver Operating Characteristic Curve）可以较好的评价二元分类问题的准确率，找到合适的阈值。曲线的横纵坐标分别为假阳性率（False Positive Rate）和真阳性率（True Positive Rate），其曲线表示随着阈值的变化两个指标的变化情况。ROC 曲线与 $y=0, x=1$ 两条曲线所围成的图形面积被定为 ROC 曲线下的面积 (Area Under Curve)，它可以很好的反应一个二元分类器的效果。

根据 ROC 曲线，可以通过以下公式求得准确率：

$$\text{accuracy} = \frac{1 + \text{TPR} - \text{FPR}}{2} \quad (3-16)$$

实验所得结果如表 3-2 所示：

表 3-2 不同距离函数的准确率

距离函数	AUC Value	Accuracy
L1 归一化距离	0.9676	90.52%
L2 归一化距离	0.9702	90.93%
余弦距离	0.9702	90.93%

(2) 基于对称卷积神经网络的身份确认算法

实验使用 CASIA 数据集，但与之前不同的是，由于本次需要借助神经网络完成二分类任务，故训练集为从 CASIA 数据集中随机选取的 100,000 个来自同一个人/来自不同人的图像对。验证集取法相同，数量为 20,000 对。

实验借助 VGG Face 的网络结构，其 caffemodel 作为预训练模型初始化两个子卷积神经网络的参数。子卷积神经网络输出后通过均值归一层进行归一化。由于最后一层输出为 1000 维，两个特征向量间的最大距离为 31.622，故设置对比损失函数的最大间距为 62，以尽可能的使两个子网相似。

训练持续 30 epoch，最终 loss 为 0.014763，在验证集上准确率为 97.15%。

3.4 本章小结

本章介绍了静态图片中的人脸识别算法。常见的人脸识别算法根据其目的分为两种：身份查找算法与身份确认算法。

针对身份查找算法，本文分别使用了三种方法实现：基于特征脸和支持向量机的传统学习方法、基于 AlexNet 的卷积神经网络、基于 VGG Face 的转化学习。传统学习方法往往无法利用大数据进行学习，模型容量有限，较难实现规模的增加和精度的提升。卷积神经网络作为深度学习的主要算法，在大规模的数据上体现出极大的优势。实验证明，有效的预训练可以使网络更容易学习到质量较高的特征，使得网络精度得到提高。最终经过 Fine-tune 的 VGG Face 网络在 CASIA 数据集上的 925 类分类网络精度达到 83.1%。

针对身份确认算法，本文使用了两种基于深度学习的方法实现：基于 CNN 特征的确认算法、基于对称卷积神经网络的算法。实验证明，经过预训练的卷积神经网络提取到的全连接层的特征可以对数据进行有效的距离比较，根据阈值进行结果的判定。对称神经网络是通过神经网络学习一对图片的高级特征之间距离关系的网络。它可以通过对比损失函数最小化距离误差，调整网络参数。

第四章 静态图片中的人脸属性识别算法研究

人脸属性识别算法是建立在人脸识别算法基础上的热门课题，它有着更加广泛的应用与前景。本章关注的人脸属性有：年龄、性别和种族。

人脸年龄识别的目的是通过人脸图像的生物学特征判断其真实年龄。作为人脸属性识别最重要的一部分，年龄识别在各种相关应用中都扮演着重要的角色，如：基于年龄的访问控制系统、年龄自适应的需求推送系统、多线索身份认证系统、数据挖掘系统等。

人脸性别识别在社交领域中起着极为重要的作用。自动性别识别可以应用于智能用户界面显示、视频监控、商业人数统计等场合。随着计算机视觉系统的应用不断增加，性别识别功能成了所有应用必不可少的一部分。

不同种族的人脸通常包含着不同的信息。由于人种的不同所产生的模型不稳定性经常会出现在实际应用中。为了解决这一问题，最好的办法是在进行其他人脸算法前先判断个体的种族信息，根据此信息决定使用特定的模型。

本章将基于深度学习对人脸属性识别算法进行研究，并进行实验分析与对比。同时，本章给出了新的属性识别模型网络结构，经实验证明，该网络对于提升属性识别网络的精度有着较好的效果。

4.1 人脸年龄识别算法概述

近年来，年龄识别算法引起了研究人员的广泛关注，不同的算法也陆续被提出，但直到现在年龄识别都还是一个较为困难的课题。首先，每个人的真实年龄为其骨龄，其外貌年龄通常与真实年龄相符合，但浮动却难以避免。其次，人的年龄同时受内部因素（如：基因）与外部因素（如：工作环境、生活方式、健康情况）的影响。最后，年龄是一个连续的变化过程，外貌特征很难在较小的年龄组之间被观察到。但随着年龄数据库的不断增大，深度学习使得我们可以得到更加稳定可靠的年龄识别模型，模型准确率不断提高。本小节将基于深度学习对人脸年龄识别算法进行研究。

人脸年龄识别问题可以分为两类：以获得年龄分组为目标的年龄识别问题，与以获得真实年龄为目标的年龄识别问题。

(1) 以获得年龄分组为目标的年龄识别问题是一个典型的多分类问题，典型的解决方法为为每一个年龄组的图片找到一组面部特征，然后通过标准分类器进行识别。所提取的年龄特征可能为该年龄组的普遍外形特征或平均脸，也可能为一些生理特征（如皱纹）。该类问题的准确率很大程度上受年龄分类标准影响，而分类标准则由人类判断年龄的先验知识所决定。

(2) 以获得真实年龄为目标的年龄识别问题是一个传统的回归问题，将年龄的真实标签作为一个连续自然数，通过损失函数进行拟合，是解决该问题的关键。基于回归的年龄识别神经网络需要更少的经验知识，但该任务更具挑战性，也较难达到更高的精度。

下面将对这两种问题进行分别讨论。

4.2 基于分类的年龄识别神经网络

4.2.1 模型结构

Gil Levi 和 Tal Hassner 在 2015 年提出了一个适用于完成年龄分组与性别分类的卷积神经网络模型^[34]。模型结构如图 4-1 所示：

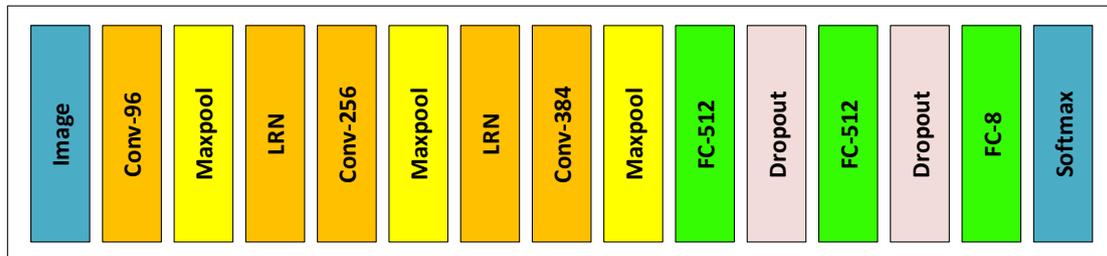


图 4-1 Gil Levi 等人提出的网络结构

网络整体框架由三个卷积层和三个全连接层组成。此网络参数与传统的 Alexnet、VGG 等相比参数更少，结构更加简单，带来的好处是可以降低网络出现过拟合的风险，并且符合较小数据集的样本拟合规律。

网络输入为三通道的 RGB 图像，经过预处理的训练集首先被减采样至 256*256，然后在输入网络后被随机裁剪为 227*227 的样本被送入网络。三个卷积层的卷积核大小分别为 7、5、3，每个卷积层后均有池化层与激活层进行减采样与特征激活。卷积层得到的特征向量将依次经过两个输出维度为 512 维的全连接层，最后被送入输出维度为 8 的第三个卷积层，用于得到每个分类的概率分布。

最后一个卷积层的输出作为最终的概率向量，将被送入 Softmax 损失函数层进行损失函数的计算。后向反馈的参数由损失函数的梯度所决定，向上传递进行网络参数的修正。

4.2.2 Softmax 损失函数

Softmax 损失函数是卷积神经网络训练多分类问题是最常用的损失函数。Softmax 损失函数层计算输入的 Softmax 函数的多项式逻辑回归，它等同于一个 Softmax 和一个多项式回归损失函数层的叠加，但是它能提供更好的数值稳定性。

一个具有 k 个输出的 Softmax 的函数形式可表示为：

$$h_{\theta}(x^{(i)}) = \begin{bmatrix} p(y^{(i)} = 1|x^{(i)}; \theta) \\ p(y^{(i)} = 2|x^{(i)}; \theta) \\ \dots \\ p(y^{(i)} = k|x^{(i)}; \theta) \end{bmatrix} = \frac{1}{\sum_{i=1}^k e^{\theta_j^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \dots \\ e^{\theta_k^T x^{(i)}} \end{bmatrix} \quad (4-1)$$

其对应的 Softmax 损失函数为：

$$J(\theta) = -\frac{1}{m} \left[\sum_{i=1}^m \sum_{j=1}^k 1\{y^{(i)} = j\} \log \frac{e^{\theta_j^T x^{(i)}}}{\sum_{i=1}^k e^{\theta_i^T x^{(i)}}} \right] \quad (4-2)$$

在实际的训练过程中，为防止参数方差过大，需要在损失函数后加入正则项，故损失函数变为：

$$J(\theta) = -\frac{1}{m} \left[\sum_{i=1}^m \sum_{j=1}^k 1\{y^{(i)} = j\} \log \frac{e^{\theta_j^T x^{(i)}}}{\sum_{i=1}^k e^{\theta_i^T x^{(i)}}} \right] + \frac{\lambda}{2} \sum_{i=1}^k \sum_{j=0}^n \theta_{ij}^2 \quad (4-3)$$

其梯度方向计算公式为：

$$\nabla_{\theta_j} J(\theta) = -\frac{1}{m} \sum_{i=1}^m [x^{(i)} (1\{y^{(i)} = j\} - p(y^{(i)} = j|x^{(i)}; \theta))] + \lambda \theta_j \quad (4-4)$$

4.2.3 实验仿真与结果分析

本实验采用数据集为 Adience 人脸数据库。Adience 数据库用于实践年龄与性别的分类问题。该数据库手机自由智能设备上传至 Flickr 的照片，这些照片没有经过人工过滤，背景各异，质量差别较大，反映了真实世界中的人脸图像的多变性与复杂性。

全部的 Adience 数据集共包括来自 2284 名用户的 26,000 张人脸图像。年龄分组共 8 组，其标签分别为：0-2、4-6、8-13、15-20、25-32、38-43、48-53、60+。官方提供了经过人脸对齐的版本，为了提高网络的准确率，本次实验使用经过人脸对齐的数据集。数据集的年龄分布如表 4-1 所示。

表 4-1 Adience 数据集年龄分布

	0-2	4-6	8-13	15-20	25-32	38-43	48-53	60-	Total
Male	745	928	934	734	2308	1294	392	442	8192
Female	682	1234	1360	919	2589	1056	433	427	9411
Both	1427	2162	2294	1653	4897	2350	825	869	19487

在训练过程中，整个数据集被分为两个部分，其中训练集共 15590 张图片，验证集共 3897 张图片。训练集平均脸图像见图 4-2。

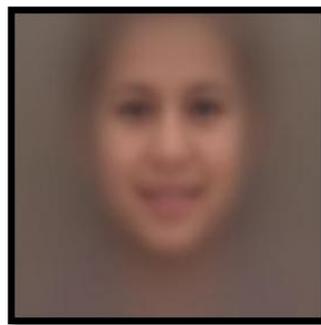


图 4-2 Adience 数据集平均脸

网络的优化算法使用 AdaGrad 算法，其迭代规则为：

$$(W_{t+1})_i = (W_t)_i - \alpha \frac{(\nabla L(W_t))_i}{\sqrt{\sum_{t'=1}^t (\nabla L(W_{t'}))_i^2}} \quad (4-5)$$

网络训练过程如图 4-3 所示，最终的网络准确率稳定在 80%。

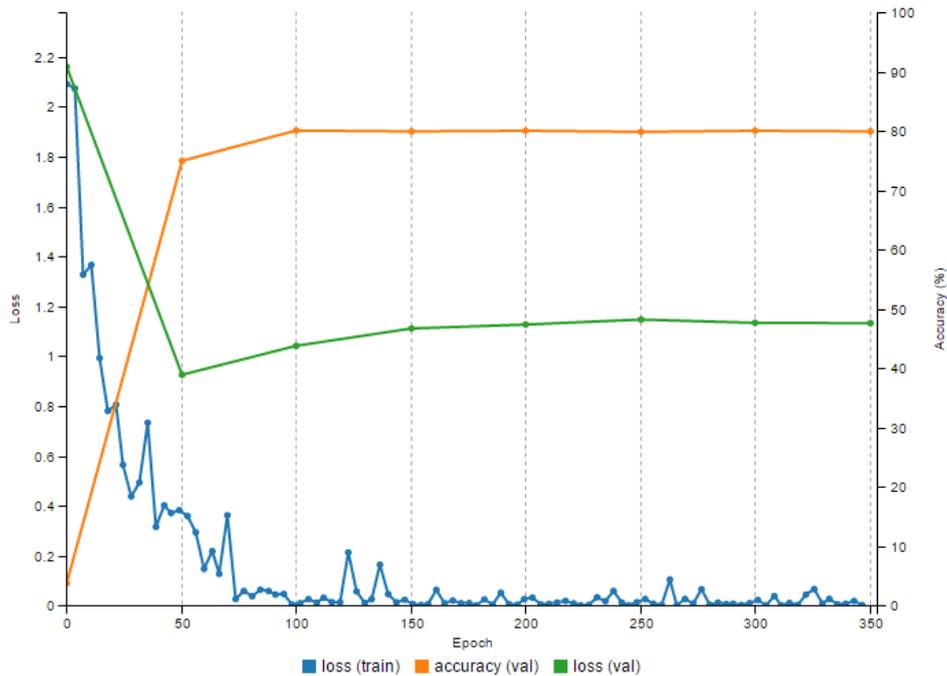


图 4-3 本模型训练曲线

4.3 基于回归的年龄识别神经网络

4.3.1 基于分类模型的改进卷积神经网络

(1) 模型结构

在回归问题中用到的网络模型由上一小节所提到的分类问题中的模型改进而成。模型改进的思路为：在保证模型的高准确率的同时降低过拟合的风险，减少网络的参数^[35]。改进后的网络模型如图 4-4 所示。

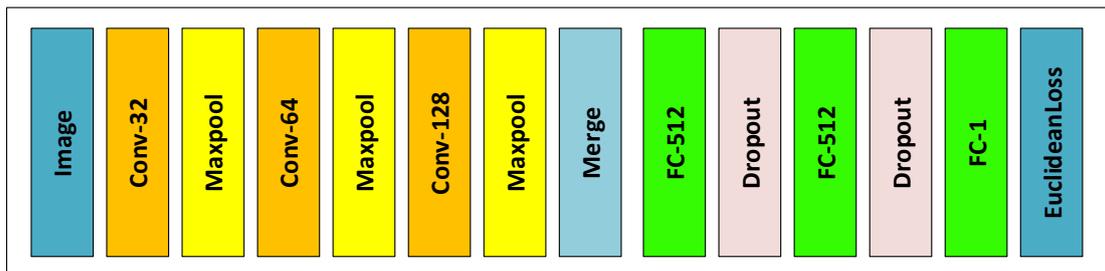


图 4-4 本模型网络结构

模型的主结构依然由三个卷积层与三个全连接层组成，但模型新加入了合并层（Merge Layer）以减少模型的规模。每个卷积层之后均跟随有一个 ReLU 层、一个池化层。图像输入为尺寸为 227*227 的三通道 RGB 图像。样本在送入网络之前每个通道会分别减去平均脸的平均像素值。

合并层是一个按元素进行操作的层，它旨在通过减少网络的参数降低整个网络的计算复杂度。可以观察到第三个卷积层的输出维度极大，因此整个网络的计算量有 90%集中在第一个全连接层的计算操作上。合并层通过将卷积层输出进行拆分，然后再通过加和方

式进行合并而将输出维度降低为原来的一半，从而有效的降低计算复杂度。合并层的操作示意图如图 4-5 所示。

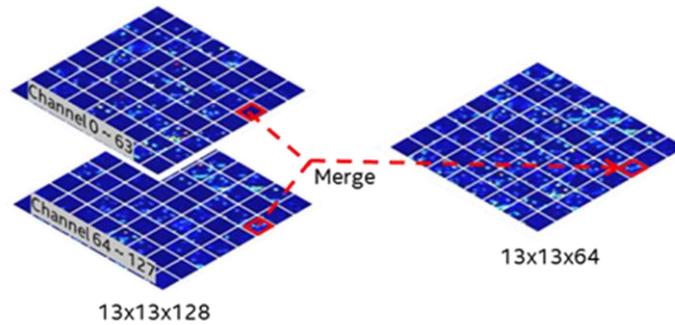


图 4-5 合并层示意图

(2) Euclidean 损失函数

上述网络的输出将被送入 Euclidean 损失函数层进行损失函数的计算。Euclidean 损失函数层是卷积神经网络在完成最小方差回归任务时最常见的损失函数层。它计算两个输入向量在欧式空间中的距离的平方，计算公式如下：

$$E = \frac{1}{2N} \|\hat{y}_n - y_n\|_2^2 \quad (4-6)$$

在回归问题中，常用平均绝对误差 (Mean Absolute Error) 作为对模型结果的衡量标准。其计算公式为：

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_i - y_i| = \frac{1}{n} \sum_{i=1}^n |e_i| \quad (4-7)$$

其中， f_i 是预测值， y_i 是真实值。需要注意的是，平均绝对误差需要建立在统一尺度的数据上。因此在使用前需要将输入进行尺度归一化。

4.3.2 基于多尺度的年龄识别神经网络

通过人脸图像得到人的属性信息是计算机视觉领域中极具挑战性的任务之一。年龄有外界因素和内部因素共同决定，且情况因人而异，因此比其他人脸识别任务更注重细节的把握。人类有着很好的观察和捕捉面部细节（如：皱纹、肤色等）的能力，但对于计算机来讲，只通过学习一张人脸图像远远不能达到目的。因此，越来越多的研究学者开始采用多尺度学习来进一步提高模型的辨识能力^[69,70,71]。

当图像尺度不同时，人脸可以显示出不同的结构与特征。同时，将不同尺度结合是一种得到对尺度鲁棒的模型的简单而有效的方法。

(1) 多尺度图片处理

人脸关键点是多尺度识别问题的重要组成部分，特别是对于非限制环境下的人脸识别问题。在得到关键点的准确信息之后，可以有效的对人脸进行姿态、角度等的纠正。多尺度图片处理的原理为：使用关键点坐标将完整的人脸图像切分成不同的区域，将所有局部人脸图像作为网络的输入进行处理。

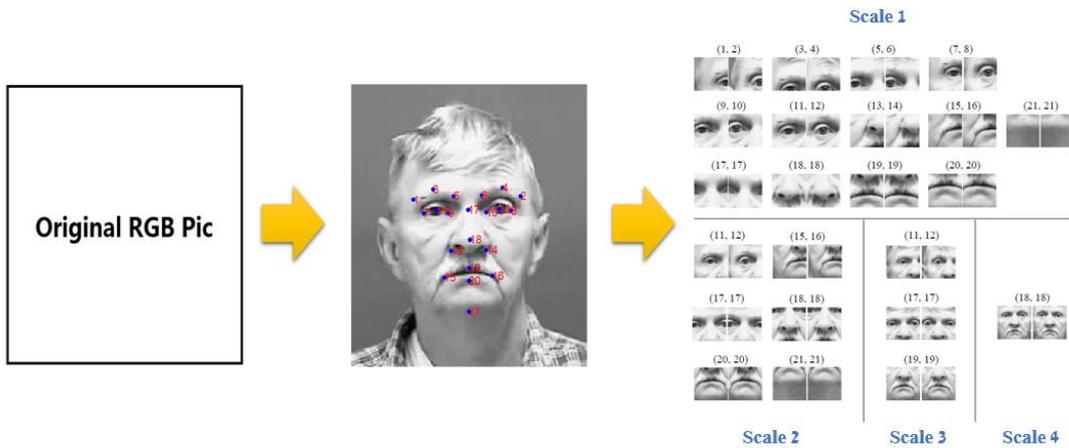


图 4-8 局部图像生成流程图

图像的处理过程为：

- 将包含完整人脸信息的图像转化为灰度图像。此步骤的目的是避免 RGB 颜色空间的不稳定性对年龄识别的干扰。
- 得到面部的 21 个关键点信息。
- 分别将 (17,19) 间的距离全局归一化到 60, 42, 30, 22, 生成 4 个不同的尺度
- 按照图 4-8 所示，以每个坐标点为中心提取 48*48 的局部区域图像
- 按照图 4-8 所示，将每对对称的局部区域图像配对，进入同一个训练通道经过预处理，最终每个样本得到 4 个尺度共 21 组局部区域图像对。

(2) 模型结构

网络结构虽然随着所用到的图像对数量的不同而变化，但其基本结构均为：N 个并行的卷积网络分别提取每组图像对的特征，一个合并层整合所有的卷积特征映射图，然后再送入连续的全连接层进行特征整合，最终得到结果。一个典型的网络结构如图 4-9 所示。

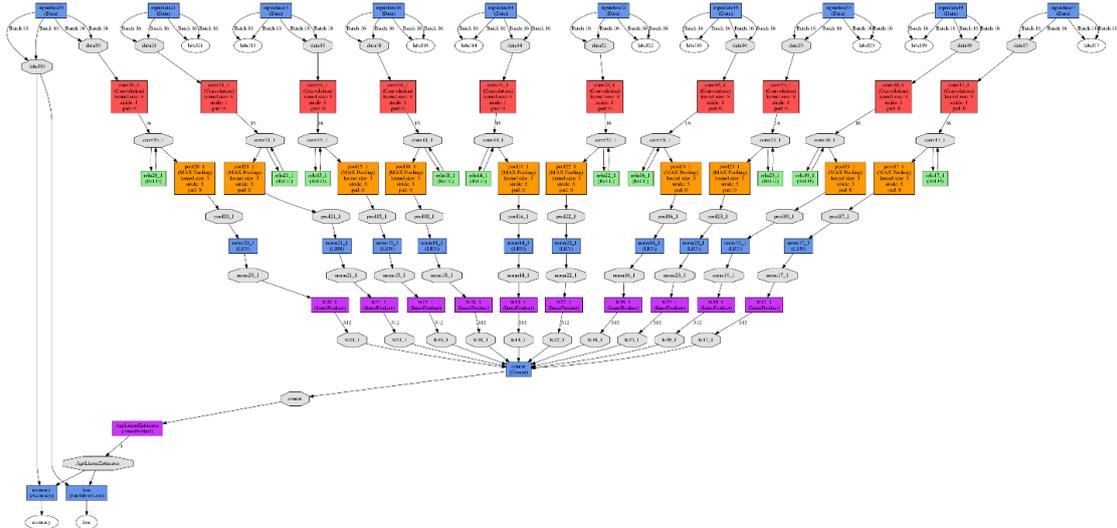


图 4-9 多尺度卷积网络结构示意图

a. 卷积层结构分析

每个子网络的结构如图 4-10 所示。

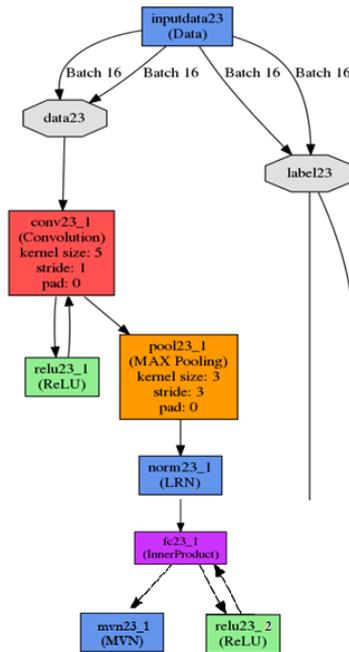


图 4-10 子网络结构示意图

为了避免计算的复杂度过高，每个子网的结构应尽可能简单，故子网由一个卷积核为 5 的卷积层，一个池化层，两个激活层，一个局部响应归一层（Local Response Normalization Layer）、一个全连接层，和一个均值归一层（Mean-Variance Normalization Layer）。均值归一层是为了在子网整合时平衡所有子网的权重，给每组图像相同的重视程度。

b. 特征整合网络结构

每个子网提取特征后，将会把特征映射图输入特征整合网络，进行最终的标签预测。特征整合网络结构如图 4-11 所示。

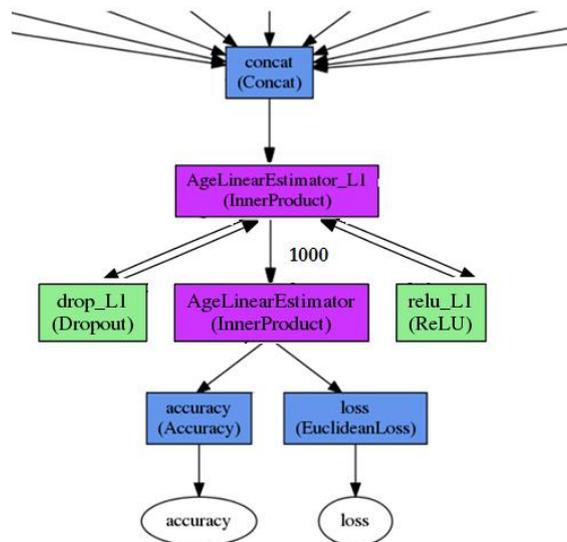


图 4-11 特征整合网络结构图

为了有效的结合不同子网的特征，两个全连接层用于输出维度分别为 1000、1 的特征向量，对于第一个全连接层，同时使用 dropout 层与激活层以避免出现过拟合现象。最终的损失函数为欧几里得损失函数。

4.3.3 实验仿真与结果分析

(1) 基于分类模型的改进卷积神经网络

实验所用数据集为 MORPH Album 2 数据集^[72]，它包括来自 13000 多人的 55000 张图片，每张图像均带有真实年龄标签，年龄范围为 16 到 77 岁。然而，数据集的数量并不平均，特别是 50 岁以上的人极少。年龄分布如图 4-6 所示：

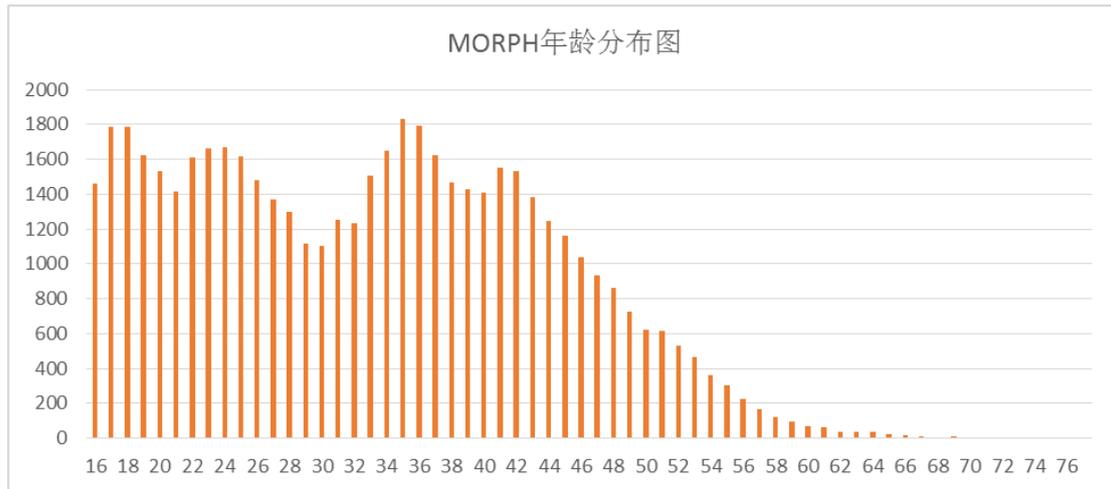


图 4-6 MORPH 年龄分布图

训练过程中，训练集包含 44,000 张图片，验证集包含 11,000 张图片。训练过程准确率曲线如图 4-7 所示：

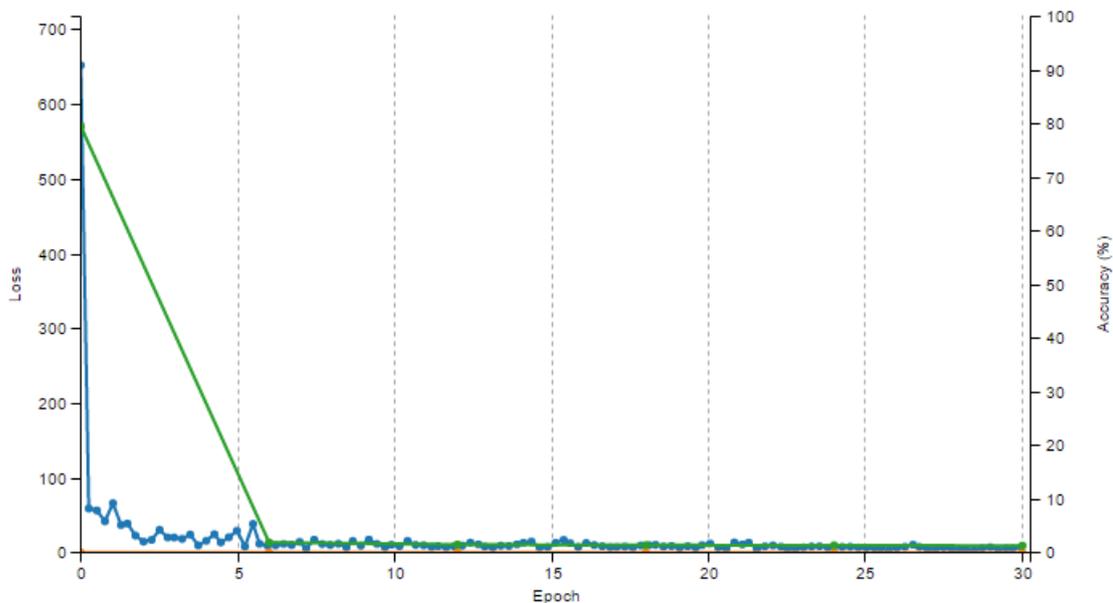


图 4-7 本模型训练曲线

最终网络的 loss 为 9.15674，通过 Caffe 提供的 pycaffe 接口进行验证，模型在验证集的平均绝对误差为 3.23 岁。

(2) 基于多尺度的卷积神经网络

本实验同样基于 MORPH Album 2 数据集，训练集与验证集比例如图 4-12 所示。

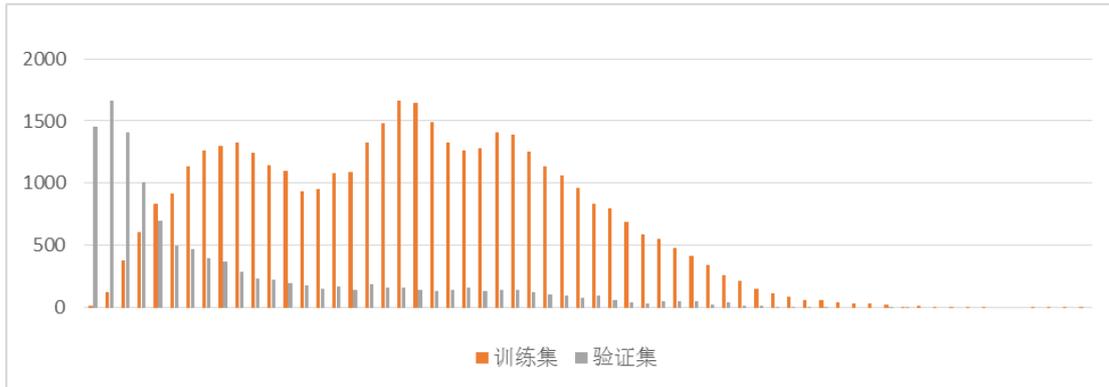


图 4-12 数据划分比例

为了探究不同尺度的训练集对结果的影响，本次实验共测试了 4 个网络，分别为：Scale 1、Scale 4、Scale 2+3+4、Scale 1+2+3+4。实验结果如表 4-2 所示。

表 4-2 不同尺度网络训练结果

尺度	子网个数	网络损失	平均绝对误差
Scale 1	13	11.7956	3.8210
Scale 4	1	18.4633	4.7997
Scale 2+3+4	10	9.5972	3.3108
Scale 1+2+3+4	23	8.4250	3.1223

4.4 基于不平衡数据集的人脸性别与种族识别神经网络

目前常见的可用于人脸性别识别的数据集往往具有类间样本比例不平衡的问题。如 MORPH Album 2 数据集男女比例为 5.5:1，LFW 数据集男女比为 2.8:1。对于种族识别问题，目前没有包括东方人数据的大规模数据集，因此只能搜集图片并进行整理用于训练与验证，这个使用 Softmax 损失函数的神经网络带来了很大的挑战。由于 Softmax 损失函数在学习过程中，会向着降低损失的方向后向传播，因此整个网络会对训练集较多的类别进行偏移，使得模型难以在实际生活中投入使用。因此需要引进新的损失函数——Hinge 损失函数。本小节主要介绍基于 Hinge 损失函数的神经网络在人脸性别与种族识别问题上的应用，并通过实验验证其可行性。

4.4.1 模型结构

用于性别与种族识别的神经网络的基本结构与年龄分类网络保持一致，但为了降低计算复杂度，模型在进入全连接层前加入了合并层。总体结构如图 4-13 所示。

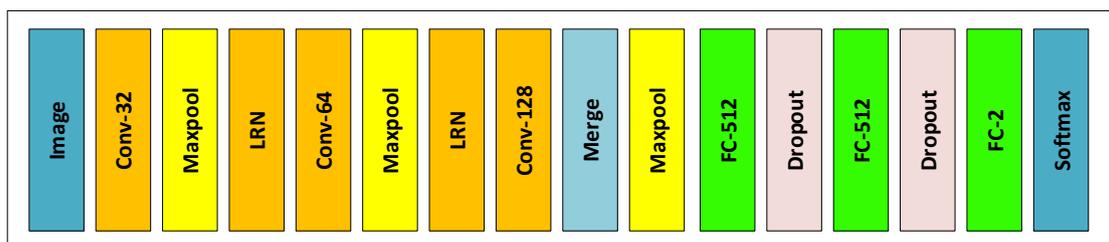


图 4-13 本实验模型结构

网络主要结构为 3 个连续的卷积层加 3 个全连接层。由于两个网络均为二分类问题，故网络的输入改为 96*96 的三通道 RGB 图像，最后一个全连接层输出为 2，用于表示标签的可能性概率分布。

4.4.2 Hinge 损失函数

Hinge 损失函数是一个常见的用于训练分类网络的损失函数。它继承了支持向量机的思想，用于最大间距分类问题。函数计算公式为：

$$E = \frac{1}{N} \sum_{n=1}^N \sum_{k=1}^K [\max(0, 1 - \delta\{l_n = k\}t_{nk})]^p \quad (4-8)$$

$$\delta(\text{condition}) = \begin{cases} 1 & \text{if condition} \\ -1 & \text{otherwise} \end{cases} \quad (4-9)$$

其中，p 表示函数使用的范数，常见的有：L1 范数、L2 范数、L2 支持向量机等。本次实验中均使用 L2 范数。

4.4.3 实验仿真与性能对比

性别识别网络在 MORPH Album 2 数据集上进行实验。训练集与验证集分布如下表所示。

表 4-3 训练集与验证集的数据分布表

集合	男性	女性	总和
训练集	37316	6793	44109
验证集	9329	1696	11025

采用上述结构进行训练，最终网络准确率为 95.14%，网络训练曲线如图 4-14 所示。

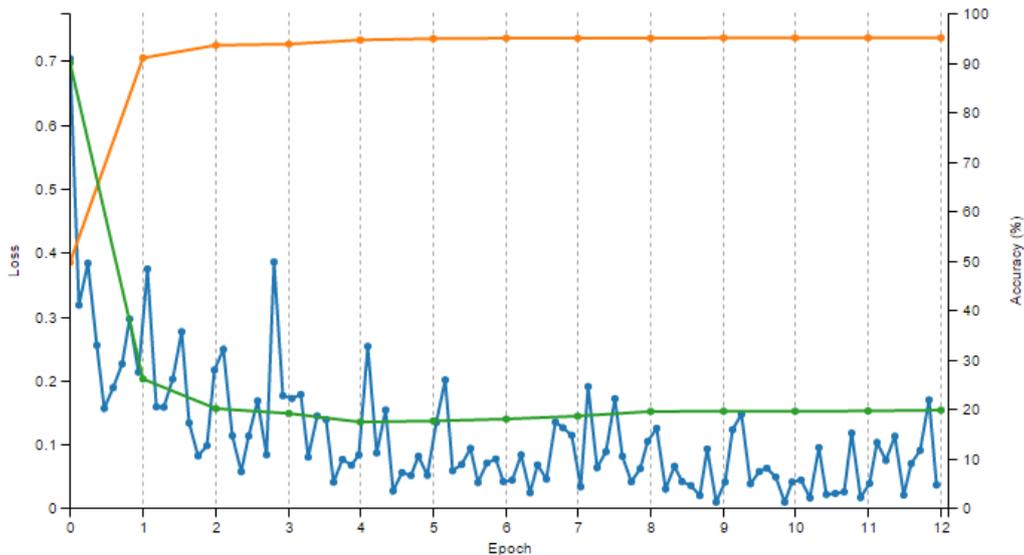


图 4-14 模型训练曲线

对于种族识别问题，由于现有数据库匮乏东方人脸图片，故采用在社交平台收集图片结合现有西方人数据库整合而成。模型为一个二分类网络，标签为：东方人、西方人。数据集比例如表 4-4 所示。

表 4-4 训练集与验证集数据分布表

集合	东方人	西方人	总和
训练集	22000	18000	40000
验证集	8000	6000	14000

训练结果如下图所示，最终网络准确率为 99.16%。

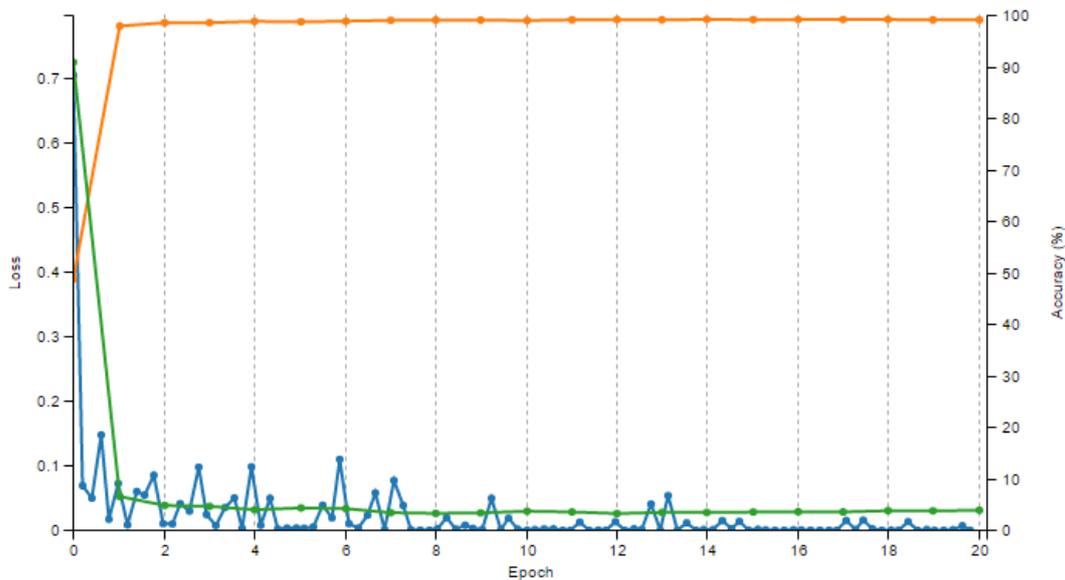


图 4-15 模型训练曲线

4.5 本章小结

本章主要研究了通过深度学习的方法完成人脸属性识别任务的方法，其中包括：人脸年龄识别、人脸性别识别和人脸种族识别。

人脸年龄识别任务分为基于年龄分类的识别任务和基于真实年龄回归的识别任务两种形式。

对于基于年龄分类的识别任务，本章介绍了 Gil Levi 等人提出的 Age CNN 的网络结构，并借助 Caffe 深度学习框架进行了网络的还原与再训练。网络在 Adience 数据集上表现良好，8 分类问题准确率达到 80%。

对于基于真实年龄回归的识别任务，本章首先在 Age CNN 上进行了结构的调整与改进，并通过使用欧几里得损失函数替换 Softmax 损失函数，使得网络可以用于训练回归模型。为了降低网络复杂度，提出了合并层的概念，并在 Caffe 中通过 Slice Layer 和 Eltwise Layer 的串联实现其功能。合并层有效的提高了模型的平均绝对误差，在 MORPH Album 2 数据集上 MAE 到达 3.23。为了让网络学习到更多人脸的局域信息，本章实现了基于多尺度的卷积神经网络。通过结合人脸关键点坐标，对人脸进行不同尺度的归一化和不同中心点的切割，形成了 23 对不同尺度不同位置的区域图像。经试验，提出的网络结构在数据集上表现良好，包含 23 个子网的网络在 MORPH Album 2 数据集上 MAE 达到了 3.12，在该数据集的分数上处于绝对的领先地位。

针对人脸性别识别问题，采用了 Age CNN 的改进结构，加入合并层进行计算量的简化，并使用 Hinge 损失函数解决非均衡数据集的优化问题。模型最终在 MORPH Album 2 数据集

上达到了 95.14% 的准确率。

针对人脸种族识别问题，采用与性别识别网络相同的结构。在网络数据与开放数据库融合的新数据集上，网络准确率高达 99.16%，有着极好的表现。

第五章 结论

5.1 结论

人脸识别是当今生物识别技术的热门研究领域，它可以应用于军事、安全、金融、商业等多领域中。本文研究了人脸识别相关算法，实现并对比了现今领域中的最为有效、常见的多种算法。在原有算法的基础上，本文结合了基于卷积神经网络的深度学习方法进行改进，得到多个可实际应用的模型，在许多较为有挑战性的非限制环境下的人脸数据集上取得了较好的结果。

本文所研究的算法共分为三个部分：人脸定位算法、人脸识别算法、人脸属性识别算法。

(1) 人脸定位算法

在人脸定位算法部分，本文分析了人脸检测算法、人脸关键点定位算法与人脸对齐算法。首先分别实现了基于 Haar、LBP 的 AdaBoost 人脸检测算法，并在 MUCT 数据集上实验。通过对结果进行对比，分析得到了两种算法的优劣与不同的适应场合。在得到人脸区域的基础上，本文实现了基于深度学习的人脸关键点定位。为了提高关键点预测的精度，六个专家模型被分别运用于不同面部区域进行识别。根据得到的关键点信息，仿射变换可以将人脸图像进行尺寸归一化和位置标准化。

(2) 人脸识别算法

在人脸识别算法部分，本文从两个方面进行了研究：一对多的身份查找算法、一对一的身份确认算法。身份查找作为传统的分类问题，本文首先实现了基于特征脸与支持向量机的身份查找算法，作为后续实验的基准。为了增加模型容量，得到具有应用价值的算法，本文使用了卷积神经网络建立模型。基于 CASIA Webface 数据库训练的卷积神经网络模型得到了较高的准确率。进一步通过对预训练好的模型进行转化学习，最终模型在 995 类分类问题中准确率达到 83.1%。

对于身份确认算法，本文实现了两种算法——基于卷积神经网络特征提取的确认算法与基于对称卷积神经网络的身份确认算法。前者借助用于分类问题的卷积神经网络完成特征提取，通过计算距离，与阈值相比较得到二元分类结果。后者借助对比损失函数训练卷积神经网络。两种方法在 LFW 数据集上均可达到 95% 以上的准确率。

(3) 人脸属性识别算法

在人脸属性识别部分，本文分别对年龄、性别、种族识别进行了研究。年龄识别问题可分为基于分类的年龄识别与基于回归的年龄识别。本文借助了一个新的网络结构，并使用合并层降低其计算复杂度。通过实现多尺度的卷积网络，模型在 MORPH Album 2 数据集上的平均绝对误差为 3.11，是目前 State-of-the-art 的结果。

用于性别与种族的神经网络模型结构以年龄分类模型作为基础，引入了 Hinge 损失函数以解决数据集不平衡的问题。性别模型在 MORPH Album 2 数据集上达到了 95% 的准确率，种族模型也在私人数据集上达到了 99% 的准确率。

本文的主要贡献在于以下三个方面：

(1) 本文对人脸识别问题中涉及到的各种算法做了深入的研究，通过大量实验对每种

算法进行了详细的分析与对比，明确了不同算法的适用场景。

(2) 将基于卷积神经网络的深度学习应用于各种人脸识别相关算法中，利用现有的公开数据集进行大规模数据的学习，训练得到可以应用于真实场景的模型。

(3) 对传统卷积神经网络进行改进，引入了专家模型、转化学习、多尺度学习等方法，针对不同问题和不同情况使用不同的损失函数对网络进行优化，对准确率有很大程度的提高。

5.2 展望

本课题研究了当前人脸识别领域的各种相关算法，并将其与深度学习算法相结合得到了可以投入实际应用中的解决方案，但本课题仍然有一些问题与不足需要进一步的研究与实践。

首先，本文对人脸定位算法的研究只针对于正脸情况。所实现的关键点定位算法对侧脸效果较差，会导致人脸对齐算法的失效。为解决这一问题，需要在进行关键点定位前对人脸姿态进行预先判定，为侧脸关键点定位建立新的模型或算法。

其次，目前公开数据集大多收集自西方社交网络，大部分样本均采集自西方人，使得模型在对东方人进行识别时，准确率有所下降。

最后，由于本次课题研究时间较短，没能将所有算法进行整合做成一个端到端的系统。这也是本次课题结束之后的一个改进方向，希望可以让本次实验中的算法投入到实际使用中，发挥其价值与作用。

参考文献

- [1] Hinton, Geoffrey E., and Ruslan R. Salakhutdinov. "Reducing the dimensionality of data with neural networks." *Science* 313.5786 (2006): 504-507.
- [2] Devi, O. Rama, et al. "Robust Rule Based Local Binary Pattern for Face Recognition." *International Journal of Advanced Research in Computer Science* 6.3 (2015).
- [3] Ren, Shaoqing, Xudong Cao, Yichen Wei, and Jian Sun. "Face alignment at 3000 fps via regressing local binary features." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1685-1692. 2014.
- [4] Qi, Xianbiao, et al. "Pairwise rotation invariant co-occurrence local binary pattern." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 36.11 (2014): 2199-2213.
- [5] Shyam, Radhey, and Yogendra Narain Singh. "Face recognition using augmented local binary pattern and Bray Curtis dissimilarity metric." In *Signal Processing and Integrated Networks (SPIN), 2015 2nd International Conference on*, pp. 779-784. IEEE, 2015.
- [6] Chai, Zhenhua, et al. "Gabor ordinal measures for face recognition." *Information Forensics and Security, IEEE Transactions on* 9.1 (2014): 14-26.
- [7] Abdulrahman, Muhammad, Tajuddeen R. Gwadabe, Fahad J. Abdu, and Alaa Eleyan. "Gabor wavelet transform based facial expression recognition using PCA and LBP." In *Signal Processing and Communications Applications Conference (SIU), 2014 22nd*, pp. 2265-2268. IEEE, 2014.
- [8] Lu, Xiaojun, Lingmei Kong, Mengzhu Liu, and Xiangde Zhang. "Facial Expression Recognition Based on Gabor Feature and SRC." In *Biometric Recognition*, pp. 416-422. Springer International Publishing, 2015.
- [9] Cheng, Yu, et al. "A fast and robust face recognition approach combining Gabor learned dictionaries and collaborative representation." *International Journal of Machine Learning and Cybernetics* 7.1 (2016): 47-52.
- [10] Sharif, Muhammad, et al. "Face recognition using edge information and DCT." *Sindh University Research Journal-SURJ (Science Series)* 43.2 (2015).
- [11] Krisshna, NL Ajit, et al. "Face recognition using transform domain feature extraction and PSO-based feature selection." *Applied Soft Computing* 22 (2014): 141-161.
- [12] Dosodia, Priya, Amarjeet Poonia, Sandeep K. Gupta, and Shubh Lakshmi Agrwal. "New Gabor-DCT Feature Extraction Technique for Facial Expression Recognition." In *Communication Systems and Network Technologies (CSNT), 2015 Fifth International Conference on*, pp. 546-549. IEEE, 2015.
- [13] Ge, Wei, Lijuan Cai, and Weida Zhan. "Face recognition with illumination, pose and expression variations using SIFT method." (2015).
- [14] Susan, Seba, et al. "Fuzzy match index for scale-invariant feature transform (SIFT) features with application to face recognition with weak supervision." *Image Processing, IET* 9.11 (2015): 951-958.
- [15] Quy, Nguyen Hong, et al. "3D human face recognition using sift descriptors of face's feature regions." *New Trends in Computational Collective Intelligence*. Springer International Publishing, 2015. 117-126.

- [16] Pong, Kuong-Hon, and Kin-Man Lam. "Multi-resolution feature fusion for face recognition." *Pattern Recognition* 47.2 (2014): 556-567.
- [17] Yan, Haibin, et al. "Multi-feature multi-manifold learning for single-sample face recognition." *Neurocomputing* 143 (2014): 134-143.
- [18] Li, Hongjun, Nicola Nobile, and Ching Y. Suen. "Face recognition based on discriminative dictionary with multilevel feature fusion." *Artificial Neural Networks in Pattern Recognition*. Springer International Publishing, 2014. 252-263.
- [19] Deng, Weihong, et al. "Transform-invariant PCA: A unified approach to fully automatic facealignment, representation, and recognition." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 36.6 (2014): 1275-1284.
- [20] Li, Jiadi, Zhenxue Chen, and Chengyun Liu. "Low-resolution Face Recognition of Multi-scale Blocking CS-LBP and Weighted PCA." *International Journal of Pattern Recognition and Artificial Intelligence* (2016).
- [21] Agrawal, Samiksha, and Pallavi Khatri. "Facial expression detection techniques: based on Viola and Jones algorithm and principal component analysis." In *Advanced Computing & Communication Technologies (ACCT), 2015 Fifth International Conference on*, pp. 108-112. IEEE, 2015.
- [22] Hafez, Samir F., Mazen M. Selim, and Hala H. Zayed. "2D Face Recognition System Based on Selected Gabor Filters and Linear Discriminant Analysis LDA." *arXiv preprint arXiv:1503.03741* (2015).
- [23] He, Guan-nan, Ze-hua Li, Fei-fei Sun, Ren-can Nie, and Hong-wei Ding. "LDA Based on PCNN for Gaussian Noisy Face Recognition." In *Information Science and Control Engineering (ICISCE), 2015 2nd International Conference on*, pp. 282-285. IEEE, 2015.
- [24] Huang, Jun, et al. "An MPCA/LDA Based Dimensionality Reduction Algorithm for Face Recognition." *Mathematical Problems in Engineering* 2014 (2014).
- [25] Sun, Yi, et al. "Deepid3: Face recognition with very deep neural networks." *arXiv preprint arXiv:1502.00873* (2015).
- [26] Taigman, Yaniv, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. "Deepface: Closing the gap to human-level performance in face verification." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1701-1708. 2014.
- [27] Parkhi, O.M., Vedaldi, A. and Zisserman, A., 2015. Deep face recognition. *Proceedings of the British Machine Vision*, 1(3), p.6.
- [28] Ding, Changxing, and Dacheng Tao. "Robust face recognition via multimodal deep face representation." *Multimedia, IEEE Transactions on* 17.11 (2015): 2049-2058.
- [29] Wu, Xiang, Ran He, and Zhenan Sun. "A Lightened CNN for Deep Face Representation." *arXiv preprint arXiv:1511.02683* (2015).
- [30] Sun, Yi, Xiaogang Wang, and Xiaoou Tang. "Deep learning face representation from predicting 10,000 classes." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1891-1898. 2014.
- [31] Savalle, Pierre-André et al. "Deformable part models with cnn features." *European Conference on Computer Vision, Parts and Attributes Workshop*. 2014.
- [32] Razavian, Ali, et al. "CNN features off-the-shelf: an astounding baseline for recognition." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2014.

- [33] El Khayari, Hachim, and Harry Wechsler. "Face Recognition across Time Lapse Using Convolutional Neural Networks." *Journal of Information Security* 7.03 (2016): 141.
- [34] Levi, Gil, and Tal Hassner. "Age and gender classification using convolutional neural networks." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2015.
- [35] Yi Yang, Feng Chen, et al. "Video System for Human Attribute Analysis using Compact Convolutional Neural Network." *ICIP*. 2016.
- [36] Santoni, Mayanda Mega, et al. "Cattle Race Classification Using Gray Level Co-occurrence Matrix Convolutional Neural Networks." *Procedia Computer Science* 59 (2015): 493-502.
- [37] Freund, Yoav, et al. "An efficient boosting algorithm for combining preferences." *The Journal of machine learning research* 4 (2003): 933-969.
- [38] Schapire, Robert E. "Explaining adaboost." *Empirical inference*. Springer Berlin Heidelberg, 2013. 37-52.
- [39] Oren, Michael, Constantine Papageorgiou, Pawan Sinha, Edgar Osuna, and Tomaso Poggio. "Pedestrian detection using wavelet templates." In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pp. 193-199. IEEE, 1997.
- [40] Viola, Paul, and Michael J. Jones. "Robust real-time face detection." *International journal of computer vision* 57.2 (2004): 137-154.
- [41] Lienhart, Rainer, and Jochen Maydt. "An extended set of haar-like features for rapid object detection." In *Image Processing. 2002. Proceedings. 2002 International Conference on*, vol. 1, pp. I-900. IEEE, 2002.
- [42] Li, Stan Z., et al. "Illumination invariant face recognition using near-infrared images." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 29.4 (2007): 627-639.
- [43] Ojala, Timo, Matti Pietikäinen, and Topi Mäenpää. "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24.7 (2002): 971-987.
- [44] Ahonen, Timo, Abdenour Hadid, and Matti Pietikainen. "Face description with local binary patterns: Application to face recognition." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 28.12 (2006): 2037-2041.
- [45] Milborrow, Stephen, John Morkel, and Fred Nicolls. "The MUCT landmarked face database." *Pattern Recognition Association of South Africa* 201, no. 0 (2010).
- [46] Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. *PAMI* 23(6), 681–685 (2001)
- [47] Yu, X., Huang, J., Zhang, S., Yan, W., Metaxas, D.N.: Pose-free facial landmark fitting via optimized part mixtures and cascaded deformable shape model. In: *ICCV*. pp. 1944–1951 (2013)
- [48] Zhu, X., Ramanan, D.: Face detection, pose estimation, and landmark localization in the wild. In: *CVPR*. pp. 2879–2886 (2012)
- [49] Zhao, Sanqiang, Yongsheng Gao, and Baochang Zhang. "Gabor feature constrained statistical model for efficient landmark localization and face recognition." *Pattern Recognition Letters* 30.10 (2009): 922-930.
- [50] Lim, Resmana, and M. J. T. Reinders. "Thiang," Facial Landmark Detection using a Gabor Filter Representation and a Genetic Search Algorithm." *Proceeding, (SITIA'2000)*, Graha Institut Teknologi Sepuluh Nopember. 2000.

- [51] Milborrow, Stephen, Tom Bishop, and Fred Nicolls. "Multiview active shape models with SIFT descriptors for the 300-W face landmark challenge." *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 2013.
- [52] Hubel, David H., and Torsten N. Wiesel. "Receptive fields and functional architecture of monkey striate cortex." *The Journal of physiology* 195.1 (1968): 215-243.
- [53] LeCun, Yann, L. D. Jackel, L. Bottou, A. Brunot, C. Cortes, J. S. Denker, H. Drucker et al. "Comparison of learning algorithms for handwritten digit recognition." In *International conference on artificial neural networks*, vol. 60, pp. 53-60. 1995.
- [54] Dong, Ping, and Nikolas P. Galatsanos. "Affine transformation resistant watermarking based on image normalization." In *Image Processing. 2002. Proceedings. 2002 International Conference on*, vol. 3, pp. 489-492. IEEE, 2002.
- [55] Hamouz, Miroslav, et al. "Feature-based affine-invariant localization of faces." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 27.9 (2005): 1490-1495.
- [56] Eiding, Eran, Roe Enbar, and Tal Hassner. "Age and gender estimation of unfiltered faces." *Information Forensics and Security, IEEE Transactions on* 9.12 (2014): 2170-2179.
- [57] Xu, Chenghua, Yunhong Wang, Tieniu Tan, and Long Quan. "Automatic 3D face recognition combining global geometric features with local shape variation information." In *Automatic face and gesture recognition, 2004. Proceedings. Sixth IEEE international conference on*, pp. 308-313. IEEE, 2004.
- [58] Ahonen, Timo, Abdenour Hadid, and Matti Pietikäinen. "Face recognition with local binary patterns." *Computer vision-eccv 2004*. Springer Berlin Heidelberg, 2004. 469-481.
- [59] Zhao, Wenyi, et al. "Discriminant analysis of principal components for face recognition." *Face Recognition*. Springer Berlin Heidelberg, 1998. 73-85.
- [60] Yang, Jian, et al. "Two-dimensional PCA: a new approach to appearance-based face representation and recognition." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 26.1 (2004): 131-137.
- [61] Lawrence, Steve, et al. "Face recognition: A convolutional neural-network approach." *Neural Networks, IEEE Transactions on* 8.1 (1997): 98-113.
- [62] Sirovich, Lawrence, and Michael Kirby. "Low-dimensional procedure for the characterization of human faces." *Josa* 4.3 (1987): 519-524.
- [63] Turk, Matthew A., and Alex P. Pentland. "Face recognition using eigenfaces." In *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*, pp. 586-591. IEEE, 1991.
- [64] Cortes, Corinna, and Vladimir Vapnik. "Support-vector networks." *Machine learning* 20.3 (1995): 273-297.
- [65] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." In *Advances in neural information processing systems*, pp. 1097-1105. 2012.
- [66] Parkhi, Omkar M., Andrea Vedaldi, and Andrew Zisserman. "Deep face recognition." *Proceedings of the British Machine Vision* 1.3 (2015): 6.
- [67] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. *Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments*. University of Massachusetts, Amherst, Technical Report 07-49, October, 2007.
- [68] Dong Yi, Zhen Lei, Shengcai Liao and Stan Z. Li, "Learning Face Representation from

- Scratch". arXiv preprint arXiv:1411.7923. 2014.
- [69] Lin, Dahua, and Xiaoou Tang. "Recognize high resolution faces: From macrocosm to microcosm." In Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, vol. 2, pp. 1355-1362. IEEE, 2006.
- [70] Zhu, Pengfei, et al. "Multi-scale patch based collaborative representation for face recognition with margin distribution optimization." Computer Vision–ECCV 2012. Springer Berlin Heidelberg, 2012. 822-835.
- [71] Yi, Dong, Zhen Lei, and Stan Z. Li. "Age estimation by multi-scale convolutional network." Computer Vision--ACCV 2014. Springer International Publishing, 2014. 144-158.
- [72] Karl Ricanek Jr and Tamirat Tesafaye, "MORPH: A Longitudinal Image Database of Normal Adult Age-Progression," IEEE 7th International Conference on Automatic Face and Gesture Recognition, Southampton, UK, April 2006, pp 341-345.
- [73] Antipov, Grigory, Sid-Ahmed Berrani, and Jean-Luc Dugelay. "Minimalistic CNN-based ensemble model for gender prediction from face images." Pattern Recognition Letters 70 (2016): 59-65.

谢辞

我还清晰的记得四年前与交大的第一次邂逅，那个拖着重重的箱子，在偌大的校园里迷失方向的短发女生，一定不知道自己将在这里经历怎样的成长与蜕变。四年时光一瞬即逝，虽然一路走来难免踌躇不知所措，偶尔失落妄自菲薄，但不忘初心，不曾放弃，是送给自己最好的毕业礼物。很庆幸自己选择了交大，四年，在这里遇到的每一位兢兢业业的老师、每一位朝夕相处的同学，都在我的成长中扮演了重要的角色。

站在毕业的门口，首先想感谢我深爱的父母。一路走来，是他们的支持和鼓励，让我有勇气承担责任，面对困难。每一次失落和无助，家都是我最温暖的港湾。年龄越长越大，离父母的距离却越来越远，能陪伴的时间也越来越少，但家，永远是心中最深的牵挂。

感谢母校上海交通大学，在这里我收获了太多太多。这里充满了机遇与挑战，充满了青春与奋斗。严谨而求实的学术氛围，让我找到自己喜欢的研究领域；开放而多样化的国际化分氛围，让我有机会在异国继续我的学业。在交大，我遇到了太多优秀的人，他们给了我前进的动力，是我奋斗的榜样。

特别想要感谢邱卫东老师，让我有机会在大一下学期进入实验室这样一个温馨的团体，在每一次的重要决定前给我的悉心的指导，在我申请过程中无私的帮助。感谢我的毕业设计指导老师，郭捷老师。从大二的 PRP 到大三的视频项目组再到大四的毕设，郭老师带我从一个计算机视觉领域的小白成长起来。本次毕设过程中，郭老师从毕业设计方向与题目的确定、算法设计思路与方法、论文撰写与修改等方面都做了细致与耐心的指导与帮助。在本次毕业设计过程中，我学到了很多，也更加坚定我专心科研的信心。感谢项目指导老师，黄征老师。每一次遇到技术问题时和黄老师的讨论都让我受益匪浅。同时我很敬佩黄老师的科研能力与对科研的热情，是我学习的榜样。十分感谢陈恭亮老师在我申请法国学校的过程中给我的所有无私的帮助和鼓励，让我有勇气去追求自己的梦想。

感谢实验室里的每一个师兄师姐，让我感到无限的温暖与归属感。从电群二号楼到软件大楼，变了的是周围的环境，不变的是陪在身边的人。感谢我的同窗，四年的时间见证了我们的奋斗，大家一起努力成为更好的自己是一件再美好不过的事情。感谢我的室友，虽然性格迥异但却能像亲姐妹一样相处，一起走过弥足珍贵的四年时光。

最后，感谢各位专家和教授能在百忙之中抽出时间来参加我的论文评审和答辩。

RESEARCH ON FACE RECOGNITION ALGORITHMS

Face recognition is a hot field of artificial intelligence. With the help of computer technology, we can analyze static images or dynamic videos to recognize one's identity. Face recognition system has a very important application in security fields. Combining with other biometric identification technology, it can be applied to complete monitoring tasks and identifying tasks of certain persons. Meanwhile, face recognition technology can be widely used in finance and commerce. To sum up, it is an advanced scientific field of the big data era.

However, face images are non-rigid, mutable but similar to some extent, which made it one of the most difficult research subject in this field. This paper aims to study related algorithms of face recognition with complex backgrounds. We realize and compare some of the most effective algorithms. We combine these original algorithms with deep learning and obtain several models which can be used in practical applications. They show good results on many challenging large-scale unrestricted face databases.

The algorithms studied in this paper are divided into three parts: face localization algorithm, face recognition algorithm, and facial attribution detection algorithm.

(1) Face localization algorithm

Face localization algorithm is to locate precisely frontal face from the given image and normalize it to a mean face image. It is the foundation of all kinds of face recognition tasks. A strong face localization algorithm could give a higher performance to furthering algorithms. In this part, we analysis face detection algorithm, facial landmark detection algorithm and face alignment algorithm. First of all, we realize two cascade AdaBoost classifiers with Haar-like feature and local binary pattern (LBP) respectively. Comparing their results based on MUCT database, we analysis the differences between these two algorithms as well as their applications. To be precise, the classifier based on Haar-like feature has a little higher accuracy than that based on LBP feature, but its speed is a bit lower. This experiment focuses on static images, so we choose the one based on Haar-like feature.

After getting the bounding box of facial images, we use a convolutional neural network to detect facial landmarks. In order to have a higher accuracy, six expert models are applied to identify different facial region. Each model contains three convolutional layers and three fully connected layers. According to the key-point information, affine transformation will be used to normalize the resize of facial images and standardize their positions.

(2) Face Recognition Algorithm

In this part, the paper does researches from two aspects: one-to-all face identification algorithm and one-to-one face verification algorithm.

Face identification is a traditional classification problem. It aims to recognize a person who is already included in our database. First of all, we implement the algorithm based on Eigenface and support vector machine (SVM), which is the baseline of the following experiments. To increase our model capacity and to get an applicable algorithm, we adopt a convolutional neural

network whose model construction is proposed by AlexNet. Comparing with traditional methods, it shows a great advantage. After using pre-trained VGG FACE model and fine-tuning on CASIA WebFace Database, we obtain an accuracy of 83.1% in the case of 995-class classification problem. Our experiment proves the pre-trained model makes the network learn higher qualified feature easily.

Face verification is a binary classification problem. It aims to judge whether two input images are from same person. To realize face verification algorithm, the paper implement two methods: the verification algorithm based on convolutional neural network feature and the verification algorithm based on Siamese network. The former method extracts image feature from a pre-trained convolutional neural network. The pre-trained model we used is a VGG FACE model fine-tuned on CASIA WebFace Database. The feature we used is from last fully connected layer. The features are used to calculate distances and compared to a threshold to decide the binary classification result. We try three functions to calculate distances: L1 normalized function, L2 normalized function and Cosine function. Both the second and the third give a good result. The latter method uses the contrastive loss function to train a convolutional neural network. The Siamese network consists two parallel convolutional neural networks with same weights. To learn how to judge the binary classification result, we use pre-trained VGG FACE model to initialize parameters. During the process of fine-tuning, the parameters of convolutional layers are fixed. The final layer – contrastive loss function – is to minimize the distance error and to adjust the whole network. Both methods give an accuracy over 95% on LFW database, which is really competitive to other algorithms.

(3) Facial attribution algorithm

Facial attribution task is established on face recognition task. It is widely used in many applications. In this part, we focus on three attributions: age, gender and race.

The target of age recognition is to estimate one's real age according to its biological characteristics shown on his/her facial image. Age recognition task includes regression task and classification task. This paper proposes a new network construction, which has a merge layer to reduce its computation complexity. It is used in both age classification model and age regression model. Then we introduce a multi-scale convolutional neural network. There are 4 scales to generate 23 local patches of facial regions. 23 subnet accept different inputs and are joint by a concat layer. Our final model achieves a mean absolute error of 3.11 on MORPH Album 2 Database, which is the state-of-the-art result on this database.

Gender recognition is very important in social networks. It is useful for intelligent user interface displays, for video surveillance, for people counting, and so on. Facial images from different races usually include different information. The instability caused by race variation always affects our models' performance in practical applications.

The construction of the convolutional neural network used for gender and race classification is similar to the one for age classification. In order to deal with the imbalance of gender and race dataset, we introduce Hinge loss function. It is encouraged by the theory of support vector machine. Our gender model achieves an accuracy of 95% on MORPH Album 2 database. Our race model achieves an accuracy of 99% on our private database.

This paper focuses on all aspects of face recognition problem, which includes face

localization, face recognition, and facial attribution recognition. The main contribution of this paper is the following:

(1) We do many experiments of face recognition tasks and go into more detail on different algorithms. Through these results, we give an exhaustive analysis and a detailed comparison among all algorithms. We make it clear to applicative scenarios of diverse methods.

(2) We apply deep learning based on convolutional neural network to many kinds of face recognition algorithms. After testing them on many challenging large-scale unrestricted face databases, we get models which could be used in real-world scenarios.

(3) We introduce many skills to improve the performances of traditional neural networks, such as expert models, transformative learning, and multi-scale learning. Aiming at various tasks and situations, we use various loss functions to optimize our models, which give a great extent of improvement to final accuracies.

(4) We propose a new compact convolutional neural network construction with a merge layer. It can be applied to train an efficient classification and regression model. It can also be completed easily with the help of open source code - Caffe. These models give an excellent performance on relating databases.

This paper studies algorithms of face recognition with complex background and unrestricted environment. Combining deep learning based on convolutional neural network, we get solutions which could be used to practical applications. But there are still some further points to resolve.

First of all, our paper focus on face localization algorithms on frontal faces. Our realized facial landmark detection algorithm is not very helpful to un-frontal faces, which will lead to the failure of following face alignment algorithm. To solve this problem, we need to train two models, one based on frontal face and the other based on un-frontal face. Then we could predict face positions before detecting facial landmark. According to the result, we can choose to use the model built for frontal faces or the model built for un-frontal faces.

Then, most of existing public databases are collected from western social network. Their images are usually taken by western people. Our models are trained on these databases. So when they are used to predict results of eastern people, the accuracy will go down. If we could collected more data of eastern people, our model will be more practical in China.

Finally, since the study time is a little short, we failed to contribute an end-to-end system to combine all facial algorithms. This is also an improving direction after this research. We hope that we can apply our algorithms into one system which has a role to play in our daily life.