上海交通大学硕士学位论文

跨年龄的人脸识别

|  |  |  |
| --- | --- | --- |
| 硕士研究生 | ： | 吴长虹 |
| 学号 | ： | 115032910024 |
| 导 师 | ： | 苏剑波教授 |
| 专业 | ： | 控制科学与工程 |
| 答 辩 日 期 | ： | 2018年1月 |

Dissertation Submitted to Shanghai Jiao Tong University for the Degree of Master

CROSS-AGE FACE RECOGNITION

|  |  |
| --- | --- |
| Candidate： | Wu Changhong |
| Student ID: | 115032910024 |
| Supervisor： | Prof.Su JianBo |
| Specialty： | Automation |

**School of Electronics and Electric Engineering**

**Shanghai Jiao Tong University**

**Shanghai, P.R. China**

**January,2018**

跨年龄的人脸识别

摘 要

随着计算机视觉技术的发展，尤其是近两年人工智能、深度学习相关技术的推动，人脸识别技术取得了突飞猛进的进展，然而针对背景光照复杂，人脸姿态、表情以及年龄变化的情况，人脸识别技术依旧面临诸多挑战。相比于对其他干扰情况的研究，抗年龄干扰的人脸识别相关研究还相对较少。然而在诸多应用场景中，实现对年龄变化鲁棒的人脸识别系统非常有必要。随着年龄的变化，人脸在形状、纹理等方面会发生巨大改变，如何处理年龄变化带来的干扰是跨年龄人脸识别的关键。

本文以实现跨年龄的人脸识别为目标，紧紧围绕提取对年龄鲁棒的人脸特征为主题。不管人的年龄特征如何变化，与人身份相关的那部分特征是稳定不变的，然而直接采用通用的人脸识别方法提取的特征中不可避免的还包含年龄带来的干扰信息。现有的跨年龄的人脸识别方法大都没有主动考虑年龄的干扰信息，合理地将其区分开。本文引入并设计年龄估计来辅助人脸识别任务去选择对年龄变化鲁棒的身份特征，分别从传统的特征子空间学习角度和基于深度学习的角度进行探索，有效的将年龄干扰信息独立开，本文的主要工作包括如下几点：

**1）非年龄相关的身份特征子空间学习算法**

本文在传统的手工特征提取基础上，引入身份字典和年龄字典将人脸特征编码到两个独立的子空间，年龄估计结果指导年龄特征子空间学习对年龄敏感的信息，由此将年龄干扰信息有效地独立开，身份识别任务指导身份子空间学习人脸特征中相对稳定的身份信息，同时引入类别约束保证学到的特征具有足够的判别能力。由此在只包含身份特征的子空间可实现跨年龄的人脸识别。本文从特征可视化和实际应用的角度证明了该方法的有效性。特别需要指出的是本文提出的方法中引入的年龄字典也可替换为表情、姿态等其他干扰的字典，因此可以方便拓展到类似的问题上。

**2）基于深度卷积网络的跨年龄的人脸识别算法**

本文利用深度卷积网络自主学习对年龄变化鲁棒的身份特征，基于已有的通用人脸数据库训练得到的模型底座进行优化，克服跨年龄的人脸数据集数据量小的局限。特别提出从特征图可视化、神经元能量分布等角度，指导基于深度卷积网络的多任务学习的网络结构设计。引入的年龄估计任务在低层级与人脸识别任务共享人脸边缘、纹理的特征提取过程，在特征选择过程中辅助人脸识别任务去选择对年龄变化鲁棒的那部分特征，保证身份识别任务能够提取到干净的身份特征，此外，在任务的目标函数中引入了Center Loss约束，充分保证身份特征的判别能力。在流行的跨年龄人脸数据库上的实验证明了该方法的有效性。

**3）基于多任务学习的跨年龄的人脸识别框架**

本文分别从传统的子空间学习和深度学习的角度设计多任务学习的跨年龄的人脸识别框架，引入的年龄估计任务提取年龄敏感的特征，辅助身份识别任务获取对年龄变化鲁棒的人脸特征。基于这两种方法提取得到的身份特征，采用计算特征Cosine距离和最近邻分类器获得最终人脸识别的结果，在MORPH和FGNET数据集上的实验证明本文基于多任务学习框架的跨年龄人脸识别的有效性。相比于单任务学习，引入的年龄辅助任务能够提高跨年龄的人脸识别准确率，同时模型的泛化能力也得到了提高。

关键词：人脸识别，年龄干扰，特征选择，子空间学习，多任务学习

CROSS-AGE FACE RECOGNITION

ABSTRACT

Recently, great advancement has been shown in face recognition due to the development of the computer vision and deep learning strategy. However, face recognition accuracy is still limited by large intra-class variations such as lighting, pose, expression and age. Face recognition with age variation has not achieved adequate attention compared with other sources of variations. Actually, cross-age face recognition is required in many practical applications. The sharp and texture of a face is subject to remarkable changes caused by the aging process. How to deal with the changes of face appearance across age is the key to tackle the cross-age face recognition.

In this thesis, the goal is to construct a robust cross-age face recognition system and the focus is to extract age invariant face feature. The person specific feature is relatively stable across age, which is one of the requirements that face recognition must meet as the identity authentication. It is inevitable the extracted face feature still contains the age variations if using the general face recognition methods. In this paper, the age estimation task is innovatively introduced to guide the feature selection for face recognition by reducing the chance of choosing age sensitive features. Both the traditional subspace learning and the deep learning perspectives are explored. The main work of this paper can be described as follows:

1. **Age invariant face feature space learning**

In this thesis, the identity and age dictionaries are introduced to encode the hand-crafted face features onto two separated subspace. The age estimation task guides the age feature subspace to catch the age sensitive feature, which help separate the age variations from the identity specific features. The face recognition task guides the identity feature subspace to catch the stable identity specific features. Moreover, the label matrix constraints are introduced to ensure the discriminative ability of the learned subspace. Thus, the cross-age face recognition can perform on the identity feature subspace. The experiments, together with feature visualization demonstrate the effectiveness of the proposed method.

1. **Cross-age face recognition based on deep convolution network**

In this thesis, deep convolution network are introduced to learn the age invariant face feature. The model is based on a pre-trained model which may overcome the shortcomings of the limited number of the cross-age face databases. Feature visualization and neuron energy distribution are used to guide the design of the deep convolution network architectures. The age estimation share the low level texture feature with the face recognition task and guide the face recognition to choose the age insensitive feature. Center loss constraints are introduced to improve the discriminative ability of the identity feature. The performance is evaluated by the experiments.

3) **Propose the cross-age face recognition framework based on multi-task learning**

From the perspective of traditional subspace learning and deep learning, a multi-task cross-age face recognition framework is proposed. The age estimation is introduced to extract age sensitive features, which guide the face recognition task to focus on age insensitive features. Cosine distance and the nearest neighbor classifier are used for face recognition. Extensive experiments on two well-known face aging datasets: MORPH and FGNET demonstrate the effectiveness of the proposed framework. Compared with the single task, the age estimation task can help improve the generalization ability of the mode and enhance the rate of the cross-age face recognition.

KEY WORDS: face recognition, age invariance, feature selection, space

learning, multi-task learning

目 录

[第一章 绪论 3](#_Toc503877023)

[1.1 引言 3](#_Toc503877024)

[1.2 研究背景 4](#_Toc503877025)

[1.2.1 应用背景 4](#_Toc503877026)

[1.2.2 技术背景 5](#_Toc503877027)

[1.3 国内外研究现状 7](#_Toc503877028)

[1.3.1 人脸识别的研究现状 7](#_Toc503877029)

[1.3.2 跨年龄的人脸识别研究现状 9](#_Toc503877030)

[1.3.3 多任务学习的研究现状 11](#_Toc503877031)

[1.4 本文的主要研究工作 12](#_Toc503877032)

[1.5 本文的主要章节安排 13](#_Toc503877033)

[第二章 人脸检测与特征点定位 14](#_Toc503877034)

[2.1 引言 14](#_Toc503877035)

[2.2 人脸检测 14](#_Toc503877036)

[2.2.1 Viola-Jones人脸检测器 15](#_Toc503877037)

[2.2.2 MTCNN人脸检测器 18](#_Toc503877038)

[2.3 人脸特征点定位 20](#_Toc503877039)

[2.3.1 组合回归树人脸定位算法 21](#_Toc503877040)

[2.3.2 TCDCN人脸特征点定位 24](#_Toc503877041)

[2.4 人脸归一化 25](#_Toc503877042)

[2.4.1 仿射变换 26](#_Toc503877043)

[2.4.2 人脸姿态校正 27](#_Toc503877044)

[2.5 本章小结 28](#_Toc503877045)

[第三章 非年龄相关的身份特征子空间学习 29](#_Toc503877046)

[3.1引言 29](#_Toc503877047)

[3.2 非年龄相关的身份特征子空间学习 31](#_Toc503877048)

[3.2.1 人脸图像特征分解 31](#_Toc503877049)

[3.2.2 基于身份字典和年龄字典的联合表示 32](#_Toc503877050)

[3.2.3 子空间判别能力的约束 33](#_Toc503877051)

[3.2.4 模型求解 34](#_Toc503877052)

[3.3 跨年龄的人脸识别框架 36](#_Toc503877053)

[3.3.1 人脸局部特征提取 37](#_Toc503877054)

[3.3.2 人脸身份特征比对 38](#_Toc503877055)

[3.4 实验结果与分析 39](#_Toc503877056)

[3.4.1 数据集 39](#_Toc503877057)

[3.4.2 实验设置 39](#_Toc503877058)

[3.4.3 特征可视化实验 40](#_Toc503877059)

[3.4.4 FGNET数据集上的实验 40](#_Toc503877060)

[3.4.5 MORPH数据集上的实验 42](#_Toc503877061)

[3.5 本章小结 43](#_Toc503877062)

[第四章 基于卷积神经网络的跨年龄人脸识别 44](#_Toc503877063)

[4.1 引言 44](#_Toc503877064)

[4.2 基于多任务深度卷积神经网络的跨年龄人脸识别 47](#_Toc503877065)

[4.2.1 VGG网络基础模型 47](#_Toc503877066)

[4.2.2 跨年龄人脸识别的网络结构设计 49](#_Toc503877067)

[4.2.3 跨年龄人脸识别的目标函数设计 53](#_Toc503877068)

[4.3 基于Cosine距离和KNN的跨年龄人脸识别 59](#_Toc503877069)

[4.4 实验结果分析 60](#_Toc503877070)

[4.4.1 实验设置 60](#_Toc503877071)

[4.4.2 MORPH数据集上的实验 60](#_Toc503877072)

[4.4.3 FGNET数据集上的实验 64](#_Toc503877073)

[4.5 本章小结 66](#_Toc503877074)

[第五章 总结与展望 67](#_Toc503877075)

[5.1 全文总结 67](#_Toc503877076)

[5.2 研究展望 68](#_Toc503877077)

[参 考 文 献 70](#_Toc503877078)

[致 谢 80](#_Toc503877079)

[攻读硕士学位期间已发表或录用的论文 81](#_Toc503877080)

[攻读学位期间参与的项目 82](#_Toc503877081)

# 绪论

## 引言

随着社会的发展和科学技术水平的不断提高，安全可靠的身份认证正成为社会广泛的需求。相比于传统的身份认证方式，如密码、钥匙、身份证等，生物特征识别技术具有更好的便携性、安全性和防伪造性。生物特征识别技术即利用人体的生理特征来对人的身份进行认证的技术，它需要满足如下几个要求[1]：1）普遍性，该生物特征必须是每个人都拥有的特征；2）唯一性，该生物特征对于每个体人都是独一无二的，足以区分不同的个体；3）可度量性，该生物特征是可以通过度量获得的；4）稳定性，该生物特征在一段时间内能够保持相对稳定。常见的生物特征识别的方式有静脉识别、虹膜识别、指纹识别、声纹识别、人脸识别等。相比于其他识别方式，人脸识别具有便捷安全、成本低廉、可靠性强、用户友好等优点，在智能监控、门禁考勤、安全安保、人机交互等领域已获得了广泛的应用。

人脸识别技术近年来已经取得了巨大的进展，但对于存在较大干扰的情况，如光照环境的变化、人脸姿态、表情的变化以及年龄带来的影响，人脸识别的效果依然不是很理想[2]。越来越多的研究者开始关注这些类内干扰对人脸识别的影响。相比于对其他干扰情况的研究，针对年龄干扰的人脸识别研究还相对较少，但是跨年龄的人脸识别系统在很多应用场合有着广泛的需求。随着年龄的变化，人脸的形状（几何特征）和纹理（统计特征）都会发生很大的改变，不同人的脸部特征改变不尽相同，影响这些人脸特征变化的因素也有很多[3]：一方面是基因遗传的影响，人脸的脸型以及皱纹等变化可能与自身的基因有联系；一方面是外部环境的影响，不同的生活方式、居住环境也会产生不同的年龄特征变化。这些年龄引起的变化不可避免地会降低人脸识别的性能，如何处理一个人在不同环境、不同年龄段下的人脸特征变化是跨年龄的人脸识别需要集中解决的问题[2]。

虽然人脸的特征会随着年龄的变化发生不可逆的改变，但是与之身份相关的某些生物特征是相对稳定的[4-5]，这也是人脸识别作为生物特征识别技术的一种所必须满足的条件之一。现有的基于判别模型的跨年龄人脸识别算法缺乏对年龄干扰信息的有效区分，本文以如何获取不受年龄干扰的身份特征为核心展开研究。考虑到人的每个年龄阶段，都有其对应的人脸纹理信息，如皱纹等，这是跨年龄的人脸识别任务不需要的特征，而另一方面年龄估计的任务又需要获取到这部分年龄敏感的信息。所以本文利用年龄估计作为辅助任务指导人脸识别任务去选择对年龄鲁棒的人脸特征，将不同年龄段的人脸特征都投影到两个子空间去，年龄子空间包含人脸随着年龄变化的纹理特征，身份子空间只包含稳定的身份特征，由此在身份子空间下进行人脸识别，可有效的排除年龄带来的干扰。考虑到上述子空间学习的过程是分步完成的，无法保证身份子空间的特征是完备的，所以本文又提出基于深度卷积网络的跨年龄的人脸识别算法，实现端到端的特征学习，用年龄估计的辅助任务去指导网络学习只与身份相关的特征，由此获取对年龄鲁棒的人脸特征进行身份识别。

## 研究背景

### 应用背景

人脸识别作为生物特征识别的一种，以其安全便捷、交互方便等特性在人们的日常生活中发挥巨大的作用。然而人脸的形状和纹理特征会随着年龄不断发生变化，当系统中录入的人脸数据与实际认证时的人脸数据存在较长时间间隔时，人脸识别的效果不可避免地会受到影响。近年来，越来越多的国内外的研究机构和学者开始关注跨年龄的人脸识别问题，不仅因为其可以从理论上进一步推动人脸识别技术的发展，更因为在实际生活中它也有着广泛的应用需求。概括来说，跨年龄的人脸识别有但不仅限于以下几个应用[2,4-5]：

**1.政府工作：大规模的身份登记、认证系统**

在很多依靠身份证、护照、驾照等证件进行人证比对的应用场景中，通常将证件持有人当前获取到的人脸图像与证件上的照片进行比对来判断人证是否匹配。进行身份认证时采集的人脸图像与证件上的人脸图像不可避免会存在时间差，年龄变化带来的干扰信息很大程度会影响识别的效果。这类大规模的身份登记工作更换注册照片过程又是极其庞杂繁琐的。如果能够实现跨年龄的人脸识别，那么就可以避免频繁更新系统中的照片，提升认证系统的使用效率和识别效果。

**2.公益领域：寻找失踪儿童**

寻找丢失儿童是社会广泛关注的问题，人工寻找周期长、难度大，通过将各种公共场所以及社交网站获取的人脸图像与失踪儿童的人脸图像数据进行比对，可以从技术角度减小寻找的难度，但是考虑到寻找周期往往较长，比对的人脸图像与实际系统获得的图像往往存在较长的时间间隔，显然，跨年龄的人脸识别技术必然是一项重要的支撑技术。

**3.公共安全领域：追捕逃犯**

在公安刑侦等应用中，可以通过车站、商场、马路等公共场所的摄像头捕获的人脸图像数据与公安机关相关的数据库进行比对，锁定嫌疑人，但是对于一些逃逸多年的嫌犯，他们的人脸特征会有意无意地发生一定的改变，而数据库中还是嫌犯曾经的人脸图像，常用的人脸识别算法在该种情况下不再完全适用，所以，实现对年龄变化鲁棒的人脸识别系统非常有必要。

### 技术背景

跨年龄的人脸识别主要包括四个步骤：人脸检测、人脸特征点定位及归一化、人脸身份特征的提取及表达、人脸分类识别，如图1-1所示。在针对跨年龄的人脸识别任务上，在特征提取及表达环节更需要关注对年龄变化鲁棒的身份特征。

人脸

检测

特征点定位及归一化

分类

识别

身份特征的提取及表达

图1-1人脸识别框架

Fig.1- Framework of face recognition

#### 1.2.2.1 人脸检测

人脸检测是所有人脸识别任务的第一步，主要是判断输入的图像中是否存在人脸，如果存在，检测出人脸在图像中位置及大小。人脸检测的结果直接影响后续步骤的进行，如果检测有误，则后续的人脸识别也会受到不利的影响。人脸检测曾经是作为人脸识别的附属研究内容。随着大人流中的人数统计等应用的逐渐普及，人脸检测目前已发展成为作为一个独立的研究方向，并逐渐已经取得了诸多有效的研究成果。传统的人脸检测方法有：基于先验知识的肤色模型的高斯模型[6]、混合高斯模型[7]、直方图模型[8]、灰度模板匹配模型[9]和变形模板匹配模型[10]；基于统计特征的AdaBoost算法[11]和支持向量机SVM算法[12]。

#### 1.2.2.2 人脸特征点定位及归一化

人脸检测获取到的人脸图像的大小、位置、角度是不定的，如果不采取合适的归一化措施，最终识别的结果也会受到影响，所以一般进行人脸检测后会对人脸图像进行人脸特征点定位和归一化操作。人脸特征点定位就是在上述人脸检测的基础上定位出人脸显著的特征点，如眉毛、眼睛、鼻子、嘴巴等。常见的人脸特征点定位主要有四类：基于主动外观模型[13]和主动形状模型[14]；基于局部模型[15]；基于回归方法[16]和基于深度学习[17]的方法。归一化操作就是在人脸特征点定位的基础上，根据特征点的先验信息，如两眼的位置在同一水平线上等，将人脸图像通过平移、旋转和缩放归一化到正脸的情况，通过归一化的过程可以保证人脸的特征点在相对一致的位置，方便后续的特征提取和分类过程。

#### 1.2.2.3人脸特征的提取和表达

人脸特征的提取和表达是人脸识别任务中非常关键的一步，提取和表达的特征不仅要刻画不同人的人脸图像之间的差别，还要保证同类的人在该特征表达下具有相似的表征。针对跨年龄的人脸识别，人脸特征的提取和表达需要对同一个人的不同年龄的人脸图像具有一定的鲁棒性和稳定性。特征提取是从人脸图像的像素点中挖掘有价值的几何结构信息和纹理信息。基于结构信息的提取需要精确的脸部特征点位置，一般较难实现；传统的基于纹理信息的特征提取方法有全局统计特征和局部统计特征。全局特征有Eigenface[18]、Fisherface[19]等；局部特征有Gabor[20]、LBP[21] 、HOG[22]等。相比几何特征，基于纹理信息的特征维度高计算复杂，但是鲁棒性更好。特征表达是对提取的特征进行变化和选择，在降低维度的同时提高特征的判别能力，主要有PCA[23]、LDA[24]等。随着深度学习的发展，基于深度卷积网络的特征提取表征能力更强[41-45]，正逐渐成为主流的特征提取方法。针对跨年龄的人脸识别任务，在人脸特征的提取和表达过程中需要考虑如何处理年龄的干扰信息对人脸识别的影响。

#### 1.2.2.5 人脸分类识别

人脸分类识别的主要任务是根据上述获取到的特征确定人的身份，根据应用场景的不同，人脸识别的过程可以分为：1:1，1:N两种模式。1:1模式也称为人脸验证，主要是针对人证匹配的应用场景，需要根据比对特征的距离设置阈值判断是否属于同一个人；1:N模式主要是在人脸数据集中匹配当前识别对象的识别模式，本文的跨年龄人脸识别属于该模式。这两种模式的分类过程都涉及对特征距离的度量，常用的度量有：Cosine距离、马氏距离、欧氏距离等。1:N模式的人脸识别过程还涉及到分类器的设计，常用的有：最近邻/k-近邻(1/KNN)[25-26]、支持向量机(SVM)[12]等。但是，一定程度上，特征提取与表达的效果决定了分类结果的上限，精准的距离度量和有区分度的分类算法只能帮助不断去逼近这个上限。

综上所述，人脸特征的提取和表达是所有人脸识别任务的关键，如何有效地获取对年龄鲁棒且具有足够区分的人脸特征是跨年龄的人脸识别的重要部分，这也是本文的主要着眼点所在。

## 国内外研究现状

### 人脸识别的研究现状

自20世纪60年代Chan和Bledsoe创建最早的人脸识别系统以来，人脸识别相关研究已有五十多年的发展历史[28]。近两年，随着深度学习技术的发展，人脸识别的相关研究迎来了新一波的热潮。

人脸识别技术主要包括人脸检测、特征点定位及归一化、特征提取及表达和分类识别四个部分。因为人脸检测、特征点定位和归一化以及分类部分具有较强的独立性，可分别进行研究。本文重点是寻找对年龄变化鲁棒的人脸特征。因此，这里主要根据特征提取方式的不同，介绍基于传统的特征提取及表达方式和基于深度学习的特征提取方式的研究现状。

#### 1.3.1.1传统的特征提取及表达

传统的手工特征提取主要分为几何特征和统计特征，统计特征根据特征提取的范围不同又可分成全局统计特征和局部统计特征。

几何特征利用人脸面部特征点的位置、几何形状关系构成的特征[29-32]，比如利用人脸各部分（眉毛、眼睛、鼻子、嘴巴等）之间的距离、角度等参数。这类基于几何特征的方法提取的人脸特征维度小，计算复杂度低，识别过程相对简单容易。但是该方法也存在较明显的缺点，几何特征容易受到姿态、表情的干扰，准确度不够，且精准的面部特征点定位依然是一个开放的技术难题，后续的研究不再单纯只关注该类特征的提取。

全局统计特征是对整张人脸图像的像素灰度值进行统计和分析。主要的方法有：Eigenface[18]、Fisherface[19]。Eigenface主要是利用主成分分析法PCA获取均方差最小意义下的特征，将人脸图像投影到低维的空间；Fisherface将人脸图像投影到最大化类间散度同时最小化类内散度的特征子空间。虽然Fisherface弥补了Eigenface的缺陷，但当人脸图像背景光照复杂，人脸姿态、表情以及年龄发生变化时，这两种方法的性能都会出现明显的下降。

相比全局统计特征而言，局部统计特征对人脸图像的类内干扰更加鲁棒。经典的局部特征有：Gabor[20]、局部二值模式(Local Binary Pattern，LBP)[21]、方向梯度直方图(Histogram of Oriented Gradient, HOG) [22]等。Gabor特征是一种符合人眼生物学原理的局部统计特征，主要提取人脸图像的空间频率、位置和方向等结构信息，但提取的特征维度高、计算复杂；相比而言，局部二值模式特征其特征维度低，计算速度快受到广泛关注，局部二值模式特征统计图像中像素点与其邻域像素点的差值关系，得到该像素的编码模式作为特征。局部二值模式是Ojala等人首先提出[21]，随后衍生出多种类型的改进方法[33-35]。HOG特征主要是计算图像局部区域的梯度方向直方图来描述局部特征，但其容易受到噪声的干扰。

经过特征提取步骤以后，为了降低特征维度并增加特征的区分度，还需要根据任务对特征进行合适的变换和选择。

常见的特征选择的方法有主成分分析法(Principal Component Analysis，PCA)、白化主成分分析法(Whitened Principal Component Analysis, WPCA)、 线性判别分析法(Linear Discriminant Analysis, LDA)以及 AdaBoost 方法。PCA，WPCA以及LDA通过无监督或者有监督的方式将特征映射到低维的空间，同时提取特征的主元信息，保证特征的区分度。AdaBoost算法级联多个弱分类器构成一个强分类器，通过大量正负样本的学习训练出一系列特征的位置和权重，达到特征选择的目的。随着范数求解理论和技术的发展，稀疏表达[37]也逐渐成为特征选择的方式之一：通过对字典样本进行学习得到具有区分能力的特征或者特征块，从而实现特征选择。

#### 1.3.1.2基于深度卷积网络的特征提取及表达

传统的手工特征提取受人为设计的规则的制约，一定程度上会制约其识别的精度，基于深度学习的特征提取方式直接从原始图像学习与任务完成相关的具有判别能力的特征，越来越引起人们的关注。在海量人脸数据、大规模图形处理器(GPU)的支撑下，随着深度模型的进一步优化，人脸识别在速度和精度上已远远超越人类。

二十世纪60年代，Hubel和Wiesel受猫的脑皮层中局部感知神经元的启发提出了卷积神经网络(Convolutional Neural Networks，CNN)[38]的理论和技术。1989年，Lecun等人把神经网络误差反向传播算法应用于手写数字识别[39]，2006年，Geoffrey Hinton提出深度置信网络(DeepBeliefNetworks,DBN)[40]，2012年Geoffrey Hinton的学生Alex在2012年的视觉领域竞赛ILSVRC上，在百万级的ImageNet数据集上将识别率从传统的70%提升到80%，随后深度卷积网络逐渐广泛应用于图像特征提取，在图像检测分类等问题上取得了惊人的效果，包括R-CNN[41]、GoogleNet[42]、VGG[43]、ResNet[44]、DenseNet[45]。在人脸识别领域，深度卷积神经网络也获得了广泛的应用[47-55]。卷积神经网络（CNN）是一种带有卷积结构的、具有多个隐层的神经网络，它通过局部连接、权值共享等策略，根据训练任务的目标函数，逐层将输入数据转化为非线性、更具表征意义的特征。研究结果表明，相比于传统的手工特征，基于深度卷积网络得到的人脸特征表达信息更丰富，具有更多人脸特征中的重要特性，比如它对人脸身份及属性具有很好的选择性，在局遮挡问题上鲁棒性更高[46]。这些特征是通过大量的人脸图像训练自主获得的，不需要额外的约束。Facebook 人工智能实验的Taigman等人训练了一个8层的卷积神经网络DeepFace在LFW数据库上人脸识别的正确率达到97.35%[47]；香港中文大学汤晓鸥实验室提出了DeepID[48]、 DeepID2[49]和DeepID2+[50]，DeepID2在类别监督信号的基础上加上了身份验证信号，在LFW数据集上的人脸识别率达到了99.15%，已经超过人类在该数据集上的识别率97.5%[51]。DeepID2+进一步探索卷积神经网络，发现神经元适度稀疏后，识别效果依然能保持，高层的神经元具有特征选择的功能。Google提出了faceNet[52]，用特征三元组损失替换了softmax损失，在LFW上获得99.63%的识别率。不同于图像分类任务，人脸识别任务除了需要考虑训练过程的分类损失，更需要关注所学特征对同类和不同类别之间的区分度，越来越多的研究工作开始关注度量学习这一方向，如 triplet-loss[52], N-pair[53],Center Loss[54]以及A-softmax[55]等。深度学习已成为人脸识别研究领域的热点，越来越多的新算法和新方法不断涌现，在性能上也逐渐超过传统的人脸识别方法。

### 跨年龄的人脸识别研究现状

近年来，对人脸图像中年龄影响的研究已经相对较多，但大部分工作主要集中在年龄估计[56-65]和衰老过程的模拟上[66-69]。对跨年龄的人脸识别任务研究相对较少。已有的关于跨年龄的人脸识别研究主要分为两个方向：生成模型角度[2,70-72]和判别模型角度[4-5,74-78]。生成模型的出发点是构建一个人脸衰老的模型，将待识别的人脸图像转化到待比对人脸图像对应的年龄段上，在此基础上再进行人脸识别。 判别模型的出发点是提取非年龄相关的人脸特征，然后再进行分类识别。下面分别从这两个角度介绍传统的机器学习方法中跨年龄的人脸识别研究现状。

生成模型[2,70-72]充分利用年龄影响过程的渐进性，构建一个年龄影响下的人脸特征变化模型。该模型能够利用已有年龄的人脸图像生成出其他年龄段的人脸图像，即可以将待测试和库中人脸图像通过该模型转化到同一个年龄段，从而去除年龄带来的影响，在此基础上再进行人脸识别。生成模型的跨年龄人脸识别主要分为两个模块：年龄模拟模型构建模块以及人脸识别模块。文献[2,71]基于人脸形状和纹理的 PCA 系数建立一个年龄变化模型来模拟年龄的影响，在人脸识别之前将待测人脸图像转变到与待比对人脸图像相同年龄下的人脸。生成模型构建的过程就是一个模拟年龄变化对人脸特征影响的过程。但是考虑到不同的年龄状态人脸老化的过程也不一样，比如:0-18岁的变化主要是体现在人脸形状的变化上，大于18岁的变化主要体现在人脸的纹理信息上[2]，此外构建这样的一个年龄老化模型需要较为干净的训练数据和准确的年龄估计做支撑，且模型的生成过程依赖大量的参数估计，年龄影响的仿真结果容易失真，计算的复杂度高[4-5]。因此，近年来关于跨年龄的人脸识别研究，主要从判别模型的角度考虑的，本文提出的模型也是属于判别模型的范畴。

判别模型[4-5,74-78]充分利用人脸图像中与身份相关的某些特征在年龄变化过程中的稳定性，从提取非年龄相关特征的角度来克服年龄带来的干扰。文献[74]使用梯度方向金字塔 GOP 作为非年龄相关的特征。文献[75]结合MLBP 和 SIFT 两种局部特征作为非年龄相关的人脸特征，并引入基于随机采样的线性判别分类器（RS-LDA）来进一步提升分类的性能。文献[73]使用一个较大的参考数据集（ Cross-Age Celebrity Dataset）将原始的人脸图像特征编码到一个非年龄相关的参考空间，同时引入最大池化聚合的方法保证一个人不同年龄的两张人脸能获得相似的编码特征。判别模型的核心是将年龄的干扰信息独立出来，只提取与身份信息相关的人脸特征。然而上述文献 [74-77] 只是从提取人脸图像中具有判别能力的特征角度出发，并没有主动分析年龄干扰对人脸特征的影响，合理地将年龄的干扰信息与人脸身份特征进行区分，导致最终提取的特征中不可避免地还含有年龄的干扰信息[4]。文献[4-5]主动考虑到图像特征的不同部分对于最终识别效果的影响，在一个线性概率判别模型中同时引入了年龄和身份的潜在因子，学习到一个身份子空间和一个年龄子空间，但是他们都对模型中身份和年龄因子以及随机噪声做了高斯分布的假设，且未对其合理性加以说明。

随着深度学习的发展，基于卷积神经网络的人脸识别获得了令人瞩目的效果，在跨年龄的人脸识别这个任务上，相关的研究也已展开，Wen等人[79]在文献[4-5]所提方法的基础上将传统的HOG特征提取方式替换为深度卷积网络，并在网络中加入隐变量层来指导非年龄相关的身份特征的学习；Zheng等人[80]利用深度学习网络获得共享的人脸特征，用该特征减掉用来做年龄估计的年龄信息，由此获得对年龄鲁棒的身份特征。

本文的主要出发点是主动考虑年龄的干扰信息，利用年龄估计和身份识别任务之间的关系，获取干净且具区分能力的身份特征。

### 多任务学习的研究现状

多任务学习已经在机器学习[81-82]和计算机视觉领域[83-84]获得了广泛的研究。多任务学习的方式包括但不限于以下几种：联合学习、自主学习、借助辅助任务的学习。作为一种归纳迁移的方法，多任务学习通过利用隐含在多个相关任务训练信号中的信息，达到提高主任务或者多个任务学习性能的目的。Caruana[86]认为相比于单任务的学习，多任务学习能够学到适用多任务的特征，同时能够提高模型的准确率、泛化能力，具体的原因可以从以下几个方面解释[85]：

1)不同的任务有不同的噪声模式，通过学习多个任务可以对噪声模式进行平均，从而获得更具泛化能力的模型；

2)当单个任务数据量小，噪声严重时，模型很难区分与任务相关和不相关的特征，额外的相关任务可以提供特征选择的依据，帮助当前任务去关注有影响的特征；

3)如果某个任务A不太容易学到某个特征，而另一个任务B正好对该特征比较敏感，那么可以在A任务的基础上引入B任务来进行模型窃听，即训练B任务来预测哪些是最重要的特征，从而让A任务在训练过程中关注这些特征。

4)多任务学习引入了归纳偏置，能起到正则化的作用，能够降低模型的复杂度，减小过拟合的风险。

大多数多任务学习是基于任务之间存在相关性的假设，那么如何去描述这种相关性就是非常重要的一个问题。传统多任务学习多采用正则化机制来刻画任务之间的关系，各个任务有各自的模型和参数，通过对任务间的不同加稀疏约束正则项来建模任务的关系，可以称之为参数的软共享。文献[82,87]学习不同的任务之间共享的特征，通过一个参数矩阵来刻画任务之间的关系，通过L1、L2或者迹范数约束保证矩阵的低秩。文献[88]通过训练数据集学习一个相关性矩阵来描述任务的相关性；文献[89]也是引入正则项约束矩阵刻画人脸验证和年龄验证这两个任务特征之间的关系。

相比于传统的多任务学习，基于深度神经网络的多任务学习更多采用参数硬共享的方式，即多任务共享隐层的特征，保留与任务相关层的参数。文献[84,90]提出多任务卷积神经网络同时解决人脸检测、特征点定位以及姿态估计任务，这几个任务浅层的纹理特征是共享的，但是又有各自的任务层。文献[91]在用深度网络解决人脸特征点定位任务的同时引入人脸属性分类作为辅助任务，文献[92]在人脸多姿态图像生成的任务上加入人脸识别辅助任务，保证生成的其他姿态人脸图像中依然含有该个体的身份信息。文献[93]证明了这种共享参数的过拟合风险比执行单个任务的参数过拟合风险要小。相比于传统机器学习，多任务学习在深度卷积网络上的应用大都没有考虑正则项的应用，更多的是利用任务的相关性获取共享的低层级的参数，对于任务联系不那么紧密或者联系比较复杂的情况，单纯地去学习共享的参数空间并不能很好利用多任务学习的优势，多任务学习是如何在深度卷积网络起作用并没有给出统一的方法。所以对于类似年龄估计和身份识别这样的多任务学习，比起学习共享的参数空间，我们更应该关注任务之间的相互交互模式。

## 本文的主要研究工作

由上述研究现状可知，现有的两类跨年龄的人脸识别方法都存在一些缺陷，基于生成模型的方法核心是构建年龄老化的生成模型，但是构建过程复杂且很容易失真，基于判别模型的方法以提取对年龄鲁棒的身份特征为核心，但是现有方法缺少对年龄干扰信息的主动考虑，个别文献虽然考虑了人脸特征的不同构成，但是缺乏行之有效的特征选择方法，无法有效的将年龄干扰信息独立开。

本文进一步从判别模型的角度考虑，深入探索如何有效的将年龄干扰信息从稳定的身份特征的独立开。主动引入年龄估计的任务来获得人脸特征中对年龄敏感的信息，辅助人脸识别任务选择对年龄鲁棒的身份信息，并分别从子空间学习和基于深度卷积网络的多任务学习两个角度进行了探索。不同于传统的子空间学习方法，本文将人脸特征映射到两个独立的子空间，年龄估计任务指导年龄子空间学习对年龄敏感的信息，身份子空间只包含与身份相关的特征信息；相比于基于深度卷积网络学习单一的人脸特征，本文设计多任务卷积网络结构，年龄估计过程辅助人脸识别学习对年龄鲁棒的身份特征。最后分别基于上述的两种特征学习方法，采用Cosine距离和最近邻分类的方法实现跨年龄的人脸识别，并在通用的跨年龄人脸识别库上验证本文所提方法的有效性。

## 本文的主要章节安排

综上所述，本文的章节安排如下：

**第一章：绪论**

本章介绍论文的研究背景，目的和意义，从传统的特征提取与表达和深度学习的角度综述了现有的人脸识别技术，介绍了跨年龄的人脸识别以及多任务学习的国内外研究现状，最后阐述了本文的主要研究内容以及各个章节的安排。

**第二章: 人脸检测和特征点定位**

本章介绍人脸识别的预处理技术，包括人脸检测、特征点定位以及姿态校正三个部分，其中人脸检测主要介绍了Viola-Jones的人脸检测方法和基于深度学习的人脸检测方法，人脸特征点定位主要介绍了基于组合回归树和基于多任务学习的MTCNN方法。最后介绍了利用仿射变换对人脸进行校正的方法。

**第三章: 非年龄相关的身份特征子空间学习算法**

本章介绍非年龄相关的身份特征子空间学习算法，引入了两个字典：身份字典和年龄字典，将人脸特征分别编码到两个独立的子空间，并对这两个子空间的表征能力进行约束，对应的年龄子空间包含了做年龄估计需要的年龄敏感信息，由此将年龄干扰信息排除出去，在身份子空间可以进行跨年龄的人脸识别。同时，本章给出非年龄相关的身份子空间模型的求解方法，将多字典的求解过程转化为凸优化的问题，简单高效。在流行的跨年龄人脸数据库上的实验证明了基于非年龄相关的身份特征子空间学习算法的有效性。

**第四章: 基于深度卷积神经网络的跨年龄的人脸识别**

本章主要从深度学习的角度处理跨年龄的人脸识别任务，引入年龄估计任务辅助识别任务在特征学习的过程中选择对年龄鲁棒的身份特征。首先从特征图可视化以及神经元能量角度探索了多任务学习的网络结构设计，随后基于特征学习的过程设计多任务学习的目标函数，充分保证所学的身份特征具有足够的判别能力，最后基于上述学习到的鲁棒的身份特征，结合Cosine距离度量和最近邻分类器给出人脸识别的结果，在MORPH和FGNET数据集上的实验证明了该方法的有效性。

**第五章：总结与展望**

本章总结论文的主要研究工作和创新点，并从对抗生成网络、迁移学习、多任务学习等角度对未来跨年龄的人脸识别研究进行了展望。

# 人脸检测与特征点定位

## 引言

跨年龄的人脸识别是一个相对比较复杂的任务，在对图像提取特征进行分类之前需要对输入的图像进行预处理，以达到预先排除一部分无关干扰信息的目的。在人脸识别任务中，图像的预处理主要包括人脸检测、特征点定位以及归一化操作，这些预处理操作是人脸识别任务中的第一步，对后续的识别效果也有重要影响[94-95]。

人脸检测的目的是在给定的图像中判断是否存在人脸，存在时则需给出人脸在图像中的位置及大小。因为跨年龄的人脸识别系统中，输入的图像可能不存在人脸或者存在多个人脸，如果不采用人脸检测直接进行人脸识别，不仅增加了计算复杂度而且加大了跨年龄人脸识别的难度。

人脸特征点定位是在上述人脸检测的基础上，对候选的人脸区域标注面部特征点的过程。特征点一般会选取边缘、角点等具有明显区分度的部位，人脸中多为眼睛、鼻子、嘴巴等位置[16,90,96]。人脸特征点定位为人脸校正和归一化提供了基础，通过将输入人脸图像的特征点归一化到相似的位置，可以为特征比对提供依据。

综上，人脸检测、特征点定位以及归一化等是跨年龄的人脸识别的非常重要的预处理技术。本章将具体介绍与跨年龄的人脸识别相关的预处理的技术实现。

## 人脸检测

对于人类而言，人脸检测是一个比较简单的任务，因为从小就收集了很多信息关于如何去辨认一张人脸。然而对机器而言，这却是一个比较困难的任务。在过去的十多年中，随着人脸数据集的不断扩增和检测算法的不断发展，人脸检测技术已经获得了重大进步。传统的人脸检测方法[97]大致可以分为：基于先验知识的方法、基于外观的方法、基于模板匹配的方法和基于特征不变的方法。2001年，Viola和Jones等人[98]提出了基于级联分类器的人脸检测算法，不仅检测准确率高而且检测速度快，很快成为人脸检测经典算法被广泛使用。随着深度学习技术的发展，基于多任务深度卷积神经网络的人脸检测算法也逐渐成为主流方法，包括Zhang等人[91]的DCNN人脸检测器、Zhang等人[99]的MTCNN人脸检测器等，下面介绍上述两类经典的人脸检测方法的实现进行介绍。

### Viola-Jones人脸检测器

Viola-Jones人脸检测器通过Harr-Like特征和Adaboost算法训练级联分类器，在准确率和实时性上均有很好的表现[98]。下面主要对直接影响其性能的三部分进行介绍。

#### 2.2.1.1 Haar-Like 特征及积分图

人脸检测时，需要一个子窗口在待检测图片中滑动，子窗口每经过一个位置，就计算该区域的图像特征，然后利用训练好的级联分类器对该部分特征进行筛选，如果该特征被所有强分类器鉴定为人脸，该区域则被判定为人脸。Viola-Jones人脸检测器采用Haar-Like特征描述图像每个区域，如图2-1所示。

a b c d e

图2-1 Haar-Like 特征

Fig.2-1 Haar-Like rectangle feature

将图2-1中的矩形放到人脸区域上，黑色矩形与白色矩形的像素和的差，即为该区域的Haar-Like特征值。Haar-Like特征值反映了人脸局部区域之间的特征明暗关系，比如眼睛区域比周围区域更暗一些等，为区分某区域是否是人脸特征提供了有效的信息。Lienhart R.等[100]对Haar-like特征进行了扩展，大致分为如下四种：边缘特征、线性特征、圆心环绕特征以及特定方向特征。通过改变这些特征模板在图像中的位置和大小，可以在很小的检测窗口中获得非常多的矩形特征，如何快速计算这些特征就显得非常重要。

Viola-Jones人脸检测器用积分图快速计算Harr-Like特征，如式（2-1）：

 （2-1）

其中是在位置处的原始像素值，是在处的积分图值。对图2-2矩形ABCD计算Haar-Like特征值时可以通过积分图变得非常高效[98]。计算结果如式（2-2）所示：

. （2-2）

x

C

D

A

B

y

图2-2 积分图示意图

图2-2 积分图计算图示

Fig.2-2 Illustration of the integral image

由此，Haar-like特征值只与特征矩形端点的积分图相关，可通过两个或四个矩阵像素和的差求得，能在常数时间内完成。

#### 2.2.1.2 Adaboost 算法

除了特征提取过程，分类的速度也是影响检测速度的重要因素。AdaBoost算法是非常实用的自适应Boosting算法[101]，针对同一个训练集训练一系列不同的弱分类器，最后通过集合的方法，把一系列弱分类器构造成一个更为强大的最终分类器，不仅准确率高，复杂度也相对较低。本质上，AdaBoost算法是一种迭代算法，结构如下图2-3所示：

图2-3 Adaboost的结构框图

Fig.2-3 The structure diagram of Adaboost

图2-3中，每个弱分类器将依据不同的权重α组合起来，共同决定最终的分类器 。一个完整的AdaBoost算法流程如下所示：

（1）训练样本，首先对每一个训练样本的权值进行初始化， 其中 。

（2）总共进行K轮的迭代，第k轮的学习过程如下：

① 基于权值为 的训练样本训练一个弱分类器。

② 计算第k个弱分类器的分类误差率：

(2-3)

③ 根据分类误差率计算弱分类器所占的权重：

(2-4)

④ 更新训练样本的权重系数：

. (2-5)

⑤ 重复①-④的过程。

（3）将上述得到的K个弱分类器根据各自的权重系数进行线性组合，得到强分类器：

. (2-6)

Viola-Jones人脸检测器中对于弱分类器的定义如下：

(2-7)

其中x是子窗口输入的图像特征， 是阈值，当特征值大于该阈值时，弱分类器的输出值是1，表示子窗口对应的是人脸，否则就是非人脸的窗口。弱分类器的训练时要寻找合适的阈值，使得最终分类的误差最小。

#### 2.2.1.3 级联结构

级联结构是Viola-Jones人脸检测中非常重要的部分，通过将简单的分类器级联成强分类器，非人脸的区域会逐步被排除出去，最后输出待检测的人脸区域。一方面级联结构增强了分类器的性能，另一方面一些不含人脸的候选窗口在经过级联结构的初级分类器时就会被排除，极大提高了人脸检测的效率[98]。分类器级联结构如图2-4所示。

1

2

T

T

T

F

F

F

子窗口

判断为人脸

判断为非人脸

…

图2-4 分类器的级联结构

Fig.2-4 The cascade structure

Viola-Jones人脸检测器通过积分图、Adaboost分类算法以及级联结构获得了巨大成功并对后续的人脸检测算法产生了深远的影响。

### MTCNN人脸检测器

自2006年起，随着深度学习技术的爆发[40]，目标检测的相关研究获得了跨越式的发展。2013年，R-CNN[41]的出现把基于深度学习的目标检测带到一个新的起点，启发和衍生了一系列相关的工作。R-CNN方法抛弃了传统的滑动窗口模式，使用了候选窗口生成来代替它，相比于Viola-jones人脸检测器的多分类器级联逐步排除的方法，R-CNN可以看成是不断提取出人脸候选区域的加法式方案。基于CNN学习的特征提取方法相比于手工设计特征Haar-Like更加契合数据的分布和目标任务。基于R-CNN系列方式的检测方法一般分为四步：1、生成候选窗口；2、CNN提取候选窗口特征；3、训练分类器对上述提完特征的窗口进行分类；4、对判别为人脸的窗口回归修正位置。

相比于传统的Viola-jones人脸检测器，虽然R-CNN在精度上获得巨大提升，但是基于深度卷积网络的特征提取方式和采用的候选窗口方法增加了计算的复杂度，因而检测速度相对较慢。随后的Fast R-CNN[102]直接计算整张图像对应的卷积特征，针对每个单独的候选窗口，在整张图的卷积特征图上获取窗口对应的特征，从而避免了R-CNN 每个窗口单独计算可能造成的重复计算问题，大大提升了检测速度。Faster R-CNN[103]采用多任务学习的方式，让候选窗口生成、分类及回归共享特征层，极大减少了计算量，提升了效率。

针对人脸检测这一特定任务，Cascade CNN[104]采用类似Viola-Jones的级联结构，每一级卷积神经网络提取特征后进行分类，在FDDB数据集上检测效果和检测速度均有很好的表现，MTCNN[99]采用Multi-task的人脸检测框架，同时检测人脸和特征点位置，在速度和精度上都比R-CNN系列等通用检测方法有很大提升，也是采用类似的3个级联CNN的结构，实现由粗到精的检测效果。

第一阶段：P-Net是一个全卷积网络，初步生成候选窗，利用边框回归向量修正候选窗，如图2-5所示。

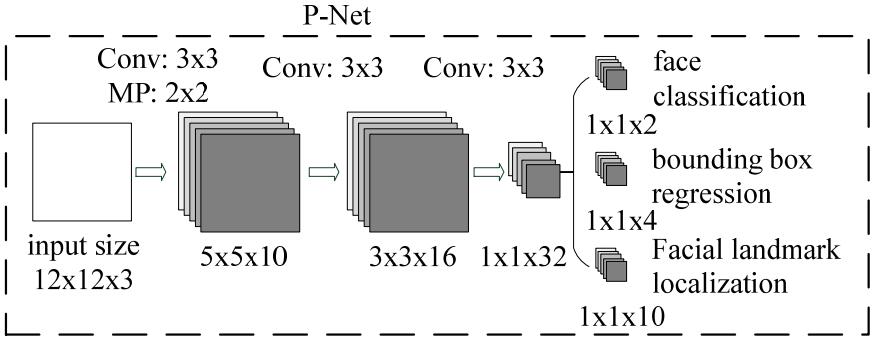


图2-5 P-Net

Fig.2-5 The P-Net structure

第二阶段：R-Net进一步修正候选窗，将P-Net得到的候选窗送入R-Net网络，进一步去除掉不含人脸的窗口，如图2-6所示。。

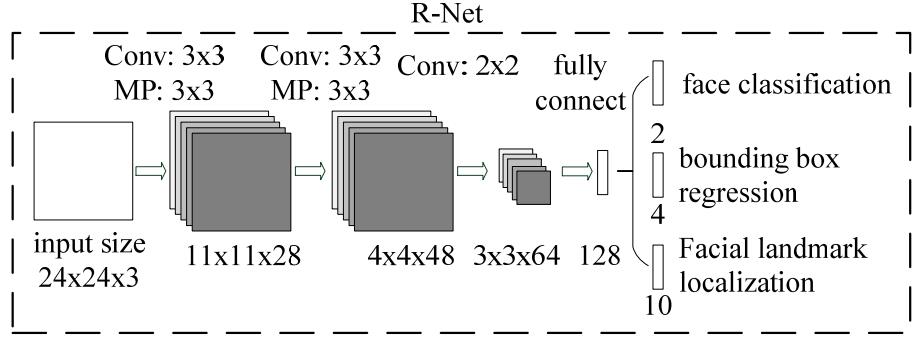


图2-6 R-Net

Fig.2-6 The R-Net structure

第三阶段：O-Net输出精选的人脸框和特征点位置，如图2-7所示。

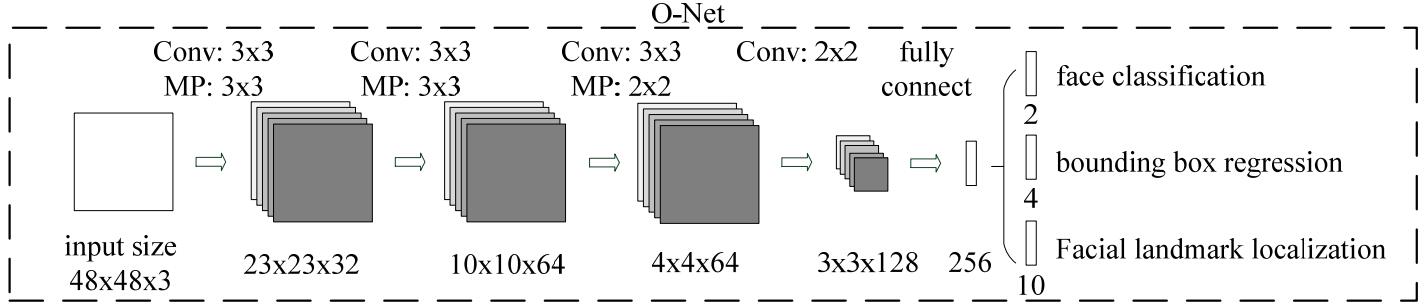


图2-7 O-Net

Fig.2-7 The O-Net structure

MTCNN需要学习三个任务：人脸检测、人脸框位置回归以及人脸特征点定位。针对不同的任务，对应的目标函数不同。

1）人脸检测任务

 (2-8)

2）人脸框位置回归

 (2-9)

3）人脸特征点定位

 (2-10)

MTCNN有类似Viola-Jones的级联结构，实现了候选窗由粗到细的挑选，同时多任务的学习可以相辅相成，使得该方法的检测性能比传统的方法好很多，同时网络结构相比[CascadeCNN](http://blog.csdn.net/shuzfan/article/details/50358809)略深但参数更少，所以在速度和[CascadeCNN](http://blog.csdn.net/shuzfan/article/details/50358809)[104]也相差不多。虽然基于深度网络的检测器目前检测精度已远远超过传统方法，但相对而言它的复杂度也很高，速度相对较慢，所以后续的进一步发展还要依赖模型的简化和加速。

## 人脸特征点定位

人脸特征点定位是在人脸候选框中标注出人脸特征点位置的过程，如图2-8所示，人脸特征点一般选择眼睛、鼻子、嘴巴以及人脸轮廓等人脸中具有显著特点的位置。

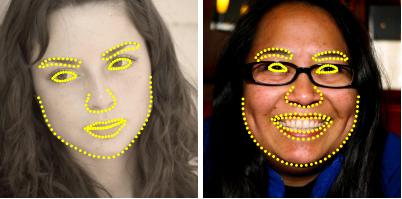


图2-8人脸特征点定位示意图

Fig.2-8 face alignment

人脸特征点定位是相对独立的一个问题，人脸特征点定位方法可以分为[17]：基于局部块分类的方法[105-109]和直接预测特征点位置（形状参数,形状回归）方法[3,110-112]。基于局部特征快分类的方法根据人脸特征点的局部特征训练分类器，用滑动窗口扫描图像，因为没有考虑人脸全局信息，所以定位的效果比较一般[112]；传统的预测形状参数的方法有主动外观模型[14]、主动形状模型[13]，该方法能够利用人脸的全局特征建立通用模型，参数多、计算复杂度高；基于回归的方法近些年也逐步被采用[110-112],该方法主要利用人脸的全局特征信息且不需要大量的迭代，所以相对更加鲁棒有效。组合回归树人脸特征点定位算法[113]作为基于回归方法的典型代表，在定位的准确率和速度上都有显著的优势。

传统的手工提取特征的方式可能不足以准确定位特征点，Sun等人[17]用3个级联的卷积神经网络在整个人脸区域中提取高层次特征，用于人脸特征点定位。文献[114-115]也是基于卷积神经网络采用由粗到精的特征点定位算法。上述方法均需要复杂的级联网络结构作为支撑。Zhang等人[84]以人脸属性作为辅助任务，采用一个独立的卷积神经网络(TCDCN)实现了人脸特征点定位，获得了相对更好的定位效果。

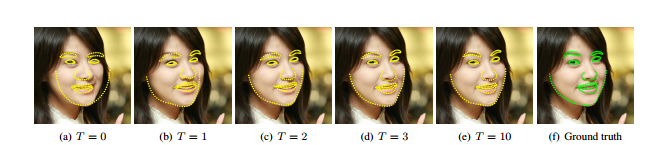
下面对上述提到的基于回归树的人脸特征点算法和TCDCN人脸特征点算法进行介绍，这两种方法分别作为基于回归方法和基于深度学习方法的典型代表，在定位速度和精度上均有很好的表现。

### 组合回归树人脸定位算法

2014年，Vahid提出了基于级联回归树的人脸特征点定位算法[113],单张人脸的定位速度能达到1ms，定位的精度也能得到很好的保证。对于人脸特征点定位，一方面希望通过与形状精确相关的特征来预测人脸形状，一方面又期望依赖可靠的人脸形状来提取有效的特征。该级联回归树方法通过迭代级联的方法很好地解决了该问题。人脸图像由一组刻画当前人脸形状的特征点来表示，根据该形状向量提取人脸图像的特征值获得形状向量的变化，再基于新的形状向量在图像中提取特征，一直重复直至收敛。具体的算法如下：

#### 2.3.1.1 级联回归模型

令 为人脸图像的第个特征点坐标，向量 表示人脸图像中个面部特征点，即为该人脸图像对应的形状模型。假设为模型当前对的估计。每个级联回归器 输出当前的人脸形状估计与根据图像预测得到的形状之间的增量，将该增量叠加到当前形状估计来获得新的形状估计，通过不断迭代，最终收敛到目标位置。

 （2-11）

(a)T=0 (b)T=1 (C)T=2 (d)T=3 (e)T=10 (f)真实结果

图2-9组合回归树人脸定位示意图

Fig.2-9 face alignment with an ensemble of regression trees

根据人脸图像中提取的特征和当前人脸形状估计，每一级的回归器给出人脸形状增量的预测，不断迭代得到的人脸形状越来越准确，对应提取的特征与形状越来越相关。回归器的设计使用了梯度提升树的算法。下面具体介绍该回归算子的训练过程。

#### 2.3.1.2 级联结构中的回归算子

训练数据定义为， 是第i个脸部图像， 是其形状向量。为了学习级联中的第一个回归量，我们从训练数据创建包含人脸图像、初始形状估计以及人脸形状更新值的三元组,,其中:

（2-12）

 （2-13）

 （2-14）

假设是单张人脸图像中特征点数量，则三元组的数量。三元组中每个图像的初始形状估计均来自中的无放回采样。可以通过平方误差和损失和梯度提升树求解回归函数。更新训练的三元组，令，得到下一个级联回归量：

 （2-15）

, （2-16）

上述过程不断迭代，直到学习到满足要求的个级联的回归量。回归算子的具体学习算法如2-1所示：

**算法 2–1 级联回归算子的学习**  
**输入:** 训练数据， 学习率

1： 初始化：

2： **for** *k* = 1*,..., K* **do**

3： **for** *i* = 1*,..., N* **do**

4： 

5： **end for**

6： 根据回归值**获得弱回归方程   
 7： 更新   
 8： **end for**

**输出**：回归算子 

每个回归量使用梯度提升方法拟合残差， 在回归树的每个节点，基于两个像素的强度差值做决策。选取平均形状中位于位置和的两个像素点，对于任意形状的人脸图片，我们都能找到对应这两个位置的点。令作为面部特征最靠近点的索引，定义它与的偏移量为：

 （2-17）

对于人脸图像的人脸形状向量，与平均形状中位置对应的可以由式（2-18）得到:

（2-18）

其中是放缩系数和，对应平均形状到的旋转矩阵，这两个参数可以由形变后人脸面部特征点坐标之间的平方差的和最小求得，如式（2-19）求得，：

 （2-19）

每次节点的分裂都关系到三个参数训练过程中的决策采用式(2-20)实现:

. （2-20）

为了训练一个回归树，每个节点随机生成一组候选分离点，通过贪心算法，最小化平方和误差选择最优的分裂结果θ\*。Q是一个节点的训练样本集合，目标优化方程为：

 （2-21）

其中表示在进行分裂θ时样本集合，是梯度提升算法计算的残差向量，的计算如式（2-22）所示：

 （2-22）

每个节点是基于一对像素强度差值作为特征，相比于单个像素，能更好地适应环境光线的变化。然而这种方法的缺点是候选的分离选择数量庞大，如果不进行遍历搜索，很难发现合适的θ。通过考虑图像的结构信息可以一定程度上减少候选的集合，因此引入如下的一个先验：

 （2-23）

该先验知识直接约束两个像素点之间的距离，优先选择距离近的像素对。

### TCDCN人脸特征点定位

文献[84]在进行人脸特征点定位的同时加上对头部姿态和人脸相关属性的学习，比如当一个小孩在笑时，他嘴巴可能会张开很大，通过对这些潜在人脸属性的探索和挖掘能够一定程度上提高人脸特征定位的准确性，下面具体介绍该模型。

TCDCN算法的网络结构如图2-10所示，网络的输入是40\*40的灰度人脸图像，特征提取部分包括四层卷积层，三层下采样层以及一层全连接层，每个卷积层包含多个特征图，采用|tanh|作为激活函数，下采样层采用最大池化的方式，全连接层连接第四层获得一个100维的共享特征向量，该特征向量可以用来做人脸特征点定位，同时也能做人脸属性分类等辅助任务。

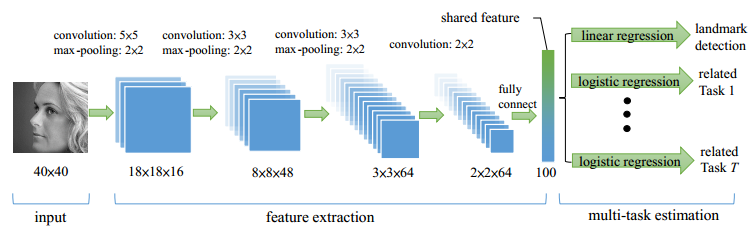


图2-10 TCDCN人脸定位示意图

Fig.2-10 face alignment with TCDCN

传统的多任务学习通过联合学习多个相关任务提升整体的效果。假设有个任务，第个任务的训练数据集为，分别对应特征向量和标签，多任务学习的目标函数如式（2-24）所示：

 （2-24）

其中是输入的共享特征和权重的函数，是损失函数，针对特征点定位任务，该损失函数可以是最小化平方差，针对人脸属性分类任务，该损失函数可以是交叉熵损失函数。是惩罚权重的正则项。

不同于传统的基于整体优化的多任务学习，该方法的主要任务是优化主要任务特征点定位任务，人脸属性只是作为辅助任务，所以最终的损失函数定义如式（2-25）所示：

 （2-25）

其中第一项是特征点定位的损失函数，；第二项是各个人脸属性分类任务的softmax损失函数，控制不同任务的比重；第三项是权重正则项，其中。通过卷积神经网络获得所有任务的共享特征，再根据任务的不同选择线性回归或者分类处理方式。

针对不同任务训练的难易程度不一样，该方法采用了“early-stop”的训练策略，在辅助任务过拟合之前即停掉该任务的训练。训练刚开始，主任务和辅助任务同时训练，提高泛化能力，避免主任务陷入局部极小值，当辅助任务收敛到最优时，即不再对主任务有所贡献时即终止掉辅助任务的学习。该方法在性能上比单任务特征点定位效果要好，也比基于cascaded CNN[17]的方法准确度更高，耗时也相对较少。

## 人脸归一化

由于人脸检测方法对人脸图像的大小、人脸姿态等具有鲁棒性，所以经过人脸检测过程获得的人脸图像在尺度、人脸角度等方面往往是多样的（如图2-11所示），所以在进行人脸识别任务之前需要在不改变人脸特征点分布模型的基础上，将输入的人脸图像归一化到一个统一的标准，以方便后续的特征提取及分类过程，虽然部分基于深度学习的方法在训练过程中对人脸归一化已没有严格要求[48,52]，但是在实际识别阶段，经过归一化后的人脸图像比对比直接采用原始人脸图像效果要更好一些。常用的人脸几何归一化方法是仿射变换。

图2-11 原始输入图像人脸姿态多样化

Fig.2-11 Different pose of the face images

### 仿射变换

仿射变换能够将二维坐标线性变换到另一个二维坐标，在变换的过程中，二维图像间的相对位置不会改变，即原来平行的两条线依然平行，相交线的夹角依旧不变，同时原来为直线的经过线性变化后依然是直线，直线上的线段长度比例不变。这种“平行性”和“平直性”能够保证经过变换后的二维图像不会发生畸变。仿射变换可以由一些基本变换组合而成：旋转(rotation)、平移(translation)、缩放(scale)、错切(shear)，如下图2-12所示：

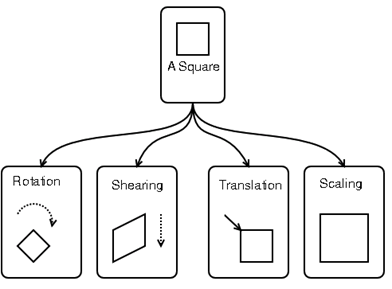


图2-12仿射变换组成

Fig.2-12 The composition of affine transformation

仿射变换矩阵刻画了原始图像和新坐标系下的图像之间的映射关系，通过仿射变换的矩阵T，可以将二维图像（如图2-13）在原坐标系中的特征点变化到新坐标系下的，如式（2-26）所示:

m

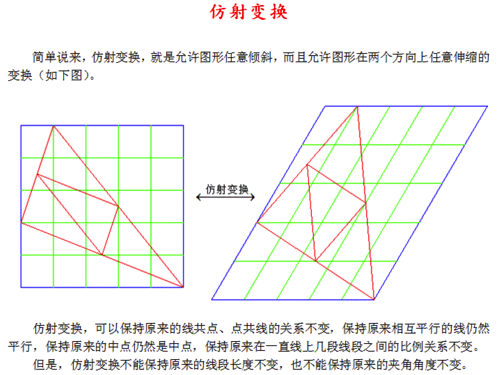


图2-13仿射变换示意图

Fig.2-13 Affine transformation

 (2-26)

如果已知原始图像和仿射变换后图像的特征点坐标，也可以根据多组原图和仿射变换后的图片对应的坐标反推出仿射变换的矩阵。

### 人脸姿态校正

人脸校正需要刻画原始图像和归一化以后图像之间的映射关系，根据计算机视觉的成像原理，该映射满足仿射变换的关系。但是类似上述的映射关系矩阵A受成像设备及环境的影响，一般情况下无法直接获得。所以针对任意环境下采集的人脸图像，进行仿射变换之前需要对仿射变换矩阵的参数进行估计。如图2-14所示，需要将人脸图像(a)归一化到与模板人脸(b)一致的大小，同时人脸特征点的位置也要完全一致。

1. (b)

图2-14人脸姿态校正示意图

Fig.2-14 Face image correction

为了方便求解仿射变换的参数，用旋转中心、旋转角度以及缩放因子S来刻画映射矩阵，令，上述式(2-26)的仿射变换矩阵T可以表示为：

 (2-27)

通常情况，这些参数可通过人脸特征点位置的先验信息来估计，根据先验知识，人的两只眼睛差不多在同一水平位置，通过计算待归一化的人脸与模板人脸两眼水平线的角度可以获得旋转角度参数，通过计算两眼水平的距离可以计算归一化前后的缩放因子。由此，可以以人眼的位置为旋转中心，获得仿射变换后的人脸图像。针对上述图2-11中的人脸图像（FGNET数据集），我们经过仿射变换得到对应的归一化人脸图像，为了方便后续工作，归一化后输出对应的灰度图，如图2-15。

图2-15人脸姿态校正后的图像

Fig.2-15 The face images after affine transformation

## 本章小结

本章主要介绍了与跨人脸识别任务相关的三个部分：人脸检测、人脸特征点定位和人脸校正。其中人脸检测技术主要介绍了经典的Viola-Jones人脸检测算法以及当前最新的基于深度卷积网络的MTCNN人脸检测算法。人脸特征点定位部分介绍了基于级联组合回归树的方法和基于多任务学习的TCDCN人脸特征点算法，基于级联组合回归树的方法不仅准确率高而且速度快，TCDCN人脸特征点算法利用辅助任务有效地提升了人脸特征点定位的准确率。人脸校正部分主要介绍了基于仿射变换的人脸归一化技术，该方法利用人脸的先验信息，计算方便。总之，本章主要关注了与跨年龄人脸识别相关的预处理技术，为后续更好地进行人脸识别提供基础，此外人脸检测及特征点定位中基于卷积神经网络的多任务学习也为后续的研究提供了理论基础。

# 非年龄相关的身份特征子空间学习

## 引言

近年来，人脸识别技术已经取得了巨大的进展，然而当光照环境、人脸姿态、表情以及年龄发生较大变化时，通用的人脸识别技术依然存在诸多挑战。越来越多的研究学者开始关注这些复杂情况的人脸识别，相比于其他干扰情况的研究，针对抗年龄干扰的人脸识别研究还相对较少。随着年龄的变化，人脸的形状和纹理会发生较大的改变[75]，如图3-1所示，如何去处理这部分变化的特征是跨年龄人脸识别的关键。



5岁 6岁 11岁 16岁 18岁

23岁 27岁 30岁 35岁 39岁

图3-1 FGNET 数据库中同一个人不同年龄的人脸图像

Fig.3-1 Example images of one subject in different ages in the FG-NET database

现有的跨年龄人脸识别方法主要有两种：生成模型方法和判别模型的方法。生成模型[2,70-72]利用年龄影响过程的渐进性，构建一个年龄影响下的人脸特征变化模型，将待测人脸图像转换到与待比对的人脸图像一样的年龄，并在此基础上进行人脸识别。该方法的核心是建立一个准确的年龄变化模型。然而这模拟年龄变化的过程需要准确的年龄估计和干净的训练数据做支撑，且计算复杂度相对较高、构建的模型很容易失真[4-5]。因此近些年关于跨年龄的人脸识别研究主要是从判别模型的角度考虑的，本章对此类方法进行深入的研究。

判别模型[4-5, 74-78]的核心是利用人脸身份特征在年龄变化过程中的稳定性，从提取非年龄相关的人脸特征角度来实现跨年龄的人脸识别。然而已有的文献[74-77]只是从提取人脸图像中具有判别能力的特征角度出发，没有主动分析年龄干扰对人脸特征的影响，并合理地将年龄的干扰信息与人脸身份特征进行区分，导致最终提取的特征中不可避免地还含有年龄的干扰信息[4-5]。

人脸的特征会随着年龄的变化发生很大的改变，但是与之身份相关的某些特征是相对稳定的[4-5]。一方面，跨年龄的人脸识别任务希望获得这部分对年龄鲁棒的身份特征，另一方面，年龄估计任务希望获得对年龄敏感的特征，而人脸图像特征一定程度上可以看成是年龄敏感的特征与身份相关的特征的组合。基于上述分析，本章的策略是要对这两部分特征进行有效的区分，将人脸图像中的年龄信息与身份信息编码到两个独立的子空间，在只包含身份信息的子空间中进行人脸识别，以根本消除年龄干扰对人脸识别的影响。

传统的子空间学习算法大都是基于原始图像特征构建整体的优化目标来求投影矩阵，如主成分分析法(PCA)、线性判别分析法 (LDA) 等，没有考虑图像特征的不同部分对于最终识别效果的影响。文献[4-5]在线性概率判别模型中引入年龄和身份的潜在因子，同时学习到一个身份子空间和一个年龄子空间，但是模型中的身份、年龄因子以及随机噪声都是基于高斯分布的假设，其合理性并未给以说明。文献[116]用稀疏表示的方法将人脸特征划分为与类别相关和与类别无关的部分，在此基础上再进行子空间学习，但稀疏表示与子空间的分步学习忽略了优化目标的整体性。

稀疏表示[126]的原理是信号可以由一组过完备的字典和稀疏系数进行表示。对于存在较大干扰的人脸识别场景，单一的字典学习[118]并不能有效地区分开干扰信息与类别信息。文献[119-121]引入字典的转化和约束项保证学习到的字典能够将与类别相关的特征很好地区分出来。文献[122]将信号分解为与类别是否相关的部分和噪声部分，在此基础上再进一步求解。但上述的字典学习过程忽略了对编码系数分类能力的约束。此外，文献[123]提出当存在较大类内干扰或者单个类别样本数目不足时，样本不能完全由字典中某个特定类别的样本进行表示，使用联合表示比稀疏表示更加合理。

基于上述分析，本章提出基于非年龄相关的身份特征子空间学习(Age Invariant Feature Space Learning，AIFSL)的方法，尽可能地将年龄干扰信息从身份特征中分离开，同时保证身份特征具有足够的判别能力。该方法将人脸图像特征看成由四部分组成：身份信息部分，年龄干扰部分、平均脸部分以及随机噪声部分，引入身份字典和年龄字典分别对身份特征和年龄特征进行编码。考虑到跨年龄的人脸识别场景下单个类别的样本相对较少，使用联合表示来学习两个独立子空间。在学习过程中引入重构误差约束和类标监督约束项[124]。识别阶段，可以通过计算待比对图像在身份空间的编码系数的距离进行分类。

## 非年龄相关的身份特征子空间学习

年龄干扰对人脸识别过程的影响是不可忽略的，如果能把对年龄变化敏感的特征与表述身份的特征区分开，则有望获得对年龄变化鲁棒的身份特征。

### 人脸图像特征分解

将人脸图像看成由以下四部分组成：平均脸部分，身份特征部分，年龄干扰部分及随机噪声部分。身份特征部分对应人脸特征中随着年龄变化相对稳定的部分，年龄干扰部分对应随着年龄不断变化的特征，平均脸部分对应所有样本的平均特征，随机噪声部分即为不能由字典表示的一些噪声信息。人脸图像的特征分解效果如图3-2所示，其中由于随机噪声的不可描述性而省略了随机噪声部分。



年龄子空间

身份子空间

平均脸

图3-2 人脸图像分解示意图

Fig.3-2 Visualization of face component separation

假设训练集由类身份类别组成，第类训练样本表示为，其中表示第类训练样本的第个样本。 对应训练集样本矩阵，其中。训练集中每个训练样本的人脸特征都可以表示为以下四部分的线性组合：身份特征部分，年龄干扰部分，平均脸部分，噪声部分。 因此，训练样本集中的单张人脸图像特征的分解可以表示为：

（3-1）

### 基于身份字典和年龄字典的联合表示

这里引入身份字典和年龄字典分别对训练样本矩阵的人脸特征的身份和年龄干扰部分进行编码。为训练样本集身份部分对应的编码系数矩阵。对应的年龄部分的编码系数矩阵。表示第类样本的身份部分的编码系数矩阵，其中对应第类样本中第个样本身份部分的编码系数。同样，对应第类样本年龄部分的编码系数矩阵，其中对应第类样本中第个样本年龄部分的编码系数。

身份字典构成了与类别相关的子空间的一组基，对应的编码系数即为第类第个样本的身份特征部分在该子空间的坐标。年龄字典构成了类内年龄干扰信息子空间的一组基，对应的编码系数即第类第个样本的年龄部分在该子空间的坐标，字典编码示意图如图3-3所示。

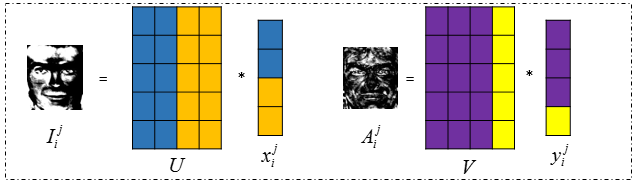


图3-3身份字典与年龄字典编码示意图

Fig.3-3 Coding with identity and age dictionaries

基于上述对人脸图像特征分解的分析，结合重构误差约束和正则项约束构成子空间学习的目标函数。文献[123]提出在训练样本单个类别样本不足或样本含较大类内干扰的情况下，采用范数比稀疏表示更加有效。由于本章研究的跨年龄的人脸识别中，单个类别的训练样本是有限的，待测样本并不能完全由字典中某个特定类别的样本进行表示，因此这里用范数来对编码系数进行约束：

 （3-2）

其中表示矩阵的Frobenious范数，第一项对应信号的重构误差项，和分别对应式(3-1)中人脸图像的身份特征部分和年龄特征部分。式(3-2)中对应的平均脸可以由训练集样本特征求平均得到：， 其中 ， 表示训练样本集的样本数量。为了均衡重构误差项和稀疏约束项，引入 两个参数。

### 子空间判别能力的约束

身份字典和年龄字典张成了对应的身份子空间和年龄子空间，一方面我们需要保证子空间中身份特征，具有足够的判别能力来进行人脸识别，另一方面，通过对年龄子空间的约束，即保证年龄特征足以能够完成年龄估计的任务，由此达到将与年龄相关的特征完全从人脸特征中剥离出来，进一步确保身份子空间获取的身份特征是足够干净的，因此我们引入身份和年龄的类标约束充分保证对应编码系数的判别能力，如图3-4所示。

身份特征

年龄特征

人脸图像特征

身份类别约束

年龄类别约束

身份字典和年龄字典编码

图3-4 对子空间学习过程进行约束

Fig.3-4 Constraints on subspace learning

根据文献[124]， 引入作为为身份特征的类标矩阵，当第个训练样本(字典)来自第个人时，该类标矩阵对应的第行，第列的元素对应为1，该列的其他元素为0；引入 为年龄特征的类标矩阵，当第个训练样本来自第个年龄阶段时，该类标矩阵对应的第行，第列的元素对应为1，该列的其他元素为0。对于整个训练集，定义两个线性的转化函数将编码得到的身份特征矩阵和年龄特征矩阵转化到身份和年龄特征的判别空间，和分别对应身份和年龄的转化矩阵。由此，加入类标矩阵约束的训练模型如下：

 （3-3）

其中对应身份特征的类标约束项，它能保证同一身份类别的人编码系数是相似的。同样，对应年龄特征的类标约束项，它能保证同一年龄的人编码系数是相似的。，分别对应身份和年龄类标矩阵约束项的正则参数，通过调整的值，可以调节两个子空间判别能力的权重比。

从特征描述角度， 项用人脸的平均脸信息、身份特征、年龄特征以及随机噪声来重构整个人脸图像特征，对应于人脸中的身份信息和年龄信息的总和。

从任务完成的角度出发，为了让包含完成人脸识别任务所需要的身份信息，让包含年龄识别所需要的年龄信息，引入了对编码系数和的约束: ()和() ，使得、分别具有足够的判别能力来区分不同的身份类别和年龄类别，由此，对于身份信息的表示能力得到了保证，对于年龄信息的表示能力得到了保证。因此，学到的字典和能够将人脸图像中的身份部分和年龄部分进行有效的区分，使得编码的系数和能够完成人脸识别和年龄识别的任务。

### 模型求解

非年龄相关特征子空间的目标函数（式(3-3)）对于字典、，类标转换矩阵、 和编码系数、整体的求解而言是非凸的，然而当、 固定时，只含、、、变化量的目标函数是凸函数；当、、、 固定时，只含、变化量的目标函数是凸函数。因此，模型的求解可以划分为两个子问题：固定字典和类标矩阵，更新编码系数；固定编码系数，更新字典和类标矩阵，交替更新，直至最终收敛。

首先，固定，更新，式(3-3)简化为：

 （3-4）

为了方便更新，令，，式(3-4)可以改写为如下形式：

**** （3-5）

可以依据文献[125]的字典更新方法进行逐类更新，字典和可根据行向量的维度从 中分解得到。

其次，固定，更新，式(3-3)简化为：

 （3-6）

类似地，式(3-6)可以转换成如下形式：

 （3-7）

其中，，，采用文献[ 125 ]中字典更新的算法来求解式(3-7)。

再次，固定，更新，式(3-3)简化为：

 （3-8）

式(3-8)可以转换成如下形式：

 （3-9）

式(3-9)是一个凸函数，因此身份特征的编码系数可以直接计算得到：

 （3-10）

最后，固定，更新，的更新过程与的近似。

 （3-11）

综上，身份、年龄字典的求解如算法3-1所示。

**算法3-1：**身份字典和年龄字典求解算法

**输入：**个身份类别的训练样本矩阵，参数。

**输出：**身份字典和年龄字典。

①初始化

计算平均脸特征，参照3.2.2节;

计算类标矩阵和，参照3.2.3节;

对训练样本分别按照身份和年龄类别进行PCA降维，初始化身份字典和年龄字典;

初始化编码系数矩阵，式(3-2) (假设 )可简化为式(3-12)， 可直接求得。

 （3-12）

初始化转换矩阵：， 。

②固定， 使用文献[125]中的字典更新方法更新式(3-5)中的。

③固定， 使用文献[125]中的字典更新方法更新式(3-7)中的。

④固定，根据式(3-10)更新。

⑤固定，根据式(3-11)更新。

⑥迭代条件判断

达到预设的迭代次数或损失小于阈值后停止迭代，否则转到②。

## 跨年龄的人脸识别框架

图像输入

人脸检测、定位和校正

人脸局部特征提取

非年龄相关的身份子空间学习

非年龄相关的身份特征

训练

测试

身份识别

图3-5 跨年龄的人脸识别框架

Fig.3-5 The framework of cross-age face recognition

基于非年龄相关的身份特征子空间的跨年龄人脸识别框架如图3-5所示，主要包括人脸图像的预处理、局部特征的提取以及非年龄相关的身份特征提取(灰色区域)三部分。其中非年龄相关的身份特征提取需要根据训练集的学习来获得具有区分能力的身份字典和年龄字典，如图3-6所示，在测试过程，由学习到的身份字典和年龄字典将待比对的人脸图像特征编码到两个独立的子空间，身份子空间对应的编码系数即为非年龄相关的身份特征，基于该编码系数的距离可以实现跨年龄的人脸识别。

训练集

++

=

=

身份特征

年龄特征

平均脸

原图



+



+



+



… … … …

子空间学习

年龄字典 系数 身份字典 系数

\*

\*

Gallery 集

Probe 集

身份分类



=

=

+

+

+

+

=

训练

测试

+

身份编码系数

图3-6非年龄相关的身份特征子空间学习算法

Fig.3-6 Age Invariant face feature space learning

### 人脸局部特征提取

相比于全局特征，局部特征在人脸图像特征多尺度、多方向的表示上更加有效，方向梯度直方图(HOG)特征是已有的局部特征描述算法中效果相对效果较好的一种[22]。 HOG特征的主要思想是将图像分成一个个小的细胞单元(cell)，然后统计每个细胞单元的像素梯度直方图，再将多个细胞单元组合将一个大的块(block)，如图3-7，将2\*2的cell组成一个大的块，将块里面所有细胞单元的直方图特征串联起来构成块的特征描述子，块与块之间有重叠，最后将所有块的特征串联起来构成最终的HOG特征描述子。

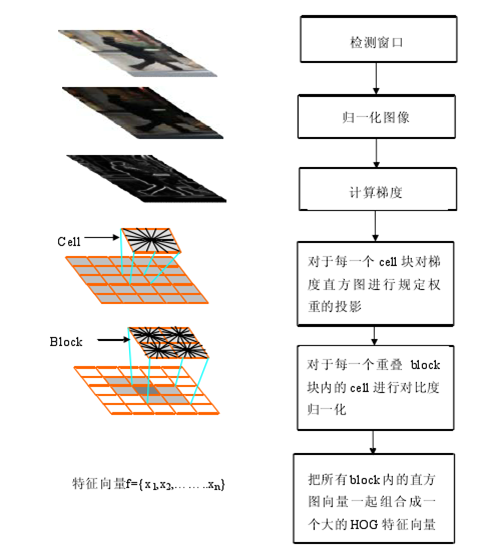


图3-7方向梯度直方图特征

Fig.3-7 Histogram of Oriented Gradient feature

为了便于与已有的跨年龄的人脸识别方法进行比较，本章采用HOG 特征作为特征描述子，具体的设置如表3-1 所示。

表3-1 HOG特征描述算子参数设置

Table 3- The parameters of HOG

|  |  |  |
| --- | --- | --- |
| 块的大小 | 重叠因子 | 方向数 |
| 10 | 0.9 | 9 |

在特征提取之前，需要对所有的人脸图片按照第二章节的人脸特征点定位和校正部分进行预处理，考虑到提取的HOG特征维度很高，需要对特征进行降维，这里采用PCA来获得较低维度的特征。

### 人脸身份特征比对

对单个人脸图片提取完HOG特征并采用PCA降维后，获得表示人脸局部特征的较低维度的特征向量。该特征向量可以由学习到的身份字典和年龄字典 编码到对应的身份子空间和年龄子空间，编码系数 即为非年龄相关的身份特征。

 (3-13)

由此，，可直接由式（3-13）计算得到：

 (3-14)

为了证明所提模型的有效性，本章直接计算样本非年龄相关的身份特征两两之间的距离，采用最近邻算法做分类。

## 实验结果与分析

### 数据集

在跨年龄的人脸识别任务上，使用广泛的数据集有：FGNET[128]、MORPH[127]数据集，具体信息如表3-2 所示：

表3-2 FGNET数据集和MORPH II数据集比较

Table 3-2 Comparison of the FGNET database and MORPH database

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 数据库 | 样本数目 | 每个类别  样本数 | 数据库龄范围 | 单个人的最大年龄跨度 |
| FGNET | 1002 | 12.2 | 0-69 | 0-45 |
| MORPH II | 55134 | 4.1 | 16-77 | 0-12 |

FGNET数据库是第一个被用来做与年龄相关任务的数据库，该数据集包含1002张图片，82个不同的身份类别，数据集的年龄跨度从0岁到69岁，且同一个人的年龄跨度相对较大，最大为从0岁到45岁。

MORPH数据库包含两个版本，MORPH I只有625个人，1690张人脸图像，MORPH II数据库相对较大，有55134张人脸图像，13618个人，已有的相关研究大都在MORPH II上进行的实验。相比FGNET数据集，MORPH II虽然图片数更多，但是平均每个人只有四张图片，每个人的年龄跨度相对较小，且包含较大的姿态、表情干扰。

### 实验设置

本章在MORPH II和FGNET数据库上验证所提的模型的有效性，所有训练集划分别分成不同的年龄组，每5岁的年龄跨度作为一个年龄组。非年龄相关的身份子空间学习算法中涉及到以下的参数：PCA降维以后的维度，类标约束的参数以及正则项的参数。对于PCA降维度，本章将从100提高到2000，当维时识别性能最好。对于类标约束参数和正则项参数，本章进一步采用交叉验证的方法来寻找最优值,最终确定FGNET集参数 ，MORPH集。

### 特征可视化实验

为了从实验角度初步证明本章所提模型能够将身份特征和年龄特征有效的区分开，本章先直接提取FGNET数据集中人脸图像的原始像素，采用非年龄相关的身份特征子空间学习算法学习一组身份字典和年龄字典，用该身份字典和年龄字典去编码一个人不同年龄段的人脸图像，获得对应的身份编码系数和年龄编码系数。通过身份字典和身份编码系数重构出该人的身份特征，年龄字典和年龄编码系数重构出该人对应年龄段的年龄特征，一个人不同年龄的人脸图像的身份特征和年龄特征的可视化结果如图3-8所示，第一行对应原始人脸图像，第二行是平均脸，第三行是对应身份子空间的特征部分，第四行是年龄子空间中的特征部分。

平均脸

身份特征

年龄特征



2岁 10岁 18岁 28岁 43岁

原图

图3-8子空间学习的可视化效果

Fig.3-8 Visualization of the subspace

由可视化的结果可以看出，由字典表示的身份特征部分确实包含了人脸身份的相关信息，由字典表示的年龄特征部分包含了人脸年龄相关的纹理特征。年龄估计的类别约束充分保证了年龄子空间中包含足够的年龄敏感信息用于年龄估计，由此将年龄干扰信息不会投影到身份子空间，保证了身份子空间的特征只与人的身份信息相关。

### FGNET数据集上的实验

依据文献[4-5,75]中训练集和测试集的划分，本章采用留一法(leave-one-person-out)来验证模型的有效性。FGNET 中82个人的样本均参与实验。81 个人的样本用来做训练，另外一人的样本用来测试。对每个待测试的类别随机挑选10对样本，每对中的第一张图像与其他81个类别的样本组成gallery 集，另一张人脸图像作为probe 集。

表3-3给出了实验的比较结果， 本章提出的方法识别率为79.8%，比其他已发表的方法效果更好些，此外，基于非年龄相关的身份特征子空间学习模型的人脸识别比直接提HOG特征效果好很多。

表3-3 本章方法与已有方法在FGNET数据集上的比较

Table 3-3 Comparison with other methods on FGNET database

|  |  |  |  |
| --- | --- | --- | --- |
|  | 方法 | 数据库(类别数，图像数) Probe集和Gallery集中 | 识别率 |
| Geng et al.[57] | 生成模型： 构建人脸随年龄变化的2D 模型，依据该年龄影响模型将待测人脸图像转变到待比对图像对应年龄下的图像再做识别。 | FGNET(10，10) | 38.1% |
| Park et al.[2] | 生成模型： 基于人脸的形状和纹理特征构建三维的人脸年龄变化模型来弥补年龄带来的影响。 | FGNET(82，1002) | 37.4% |
| Li et al. [75] | 判别模型： 结合多个局部特征LBP，SIFT等 作为非年龄相关的特征，采用基于随机采样的线性判别分类器(RS-LDA)构建判别模型。 | FGNET(82，82) | 47.5% |
| Gong et al. [4] | 判别模型： 基于身份因子和年龄因子的隐因子分析概率判别模型。 | FGNET(82，82) | 69.0% |
| Gong et al. [5] | 判别模型： 提出基于最大熵的特征算子(MEFD)并采用概率匹配模型直接计算两个人脸图像中身份特征部分相似的概率。 | FGNET(82，82) | 76.2% |
| 本章方法 | 直接提取HOG特征，采用最近邻算法做分类。 | FGNET(82，82) | 53.7% |
| 判别模型： 非年龄相关的身份特征子空间学习算法 | FGNET(82，82) | 79.8% |

图(3-9)给出正确识别的部分人脸图像，第一行为待测试的人脸图像，第二行对应识别结果，与正确的结果一致。

图(3-10)给出错误识别的部分人脸图像。第一行为待测试的人脸图像，第二行对应识别结果，第三行对应真实的结果。可以发现，由于年龄带来巨大的干扰，识别结果比真实结果更加接近测试的人脸图像。



18岁 5岁 36岁 58岁 61岁 6岁

2岁 28岁 12岁 18岁 24岁 35岁

待测试图像

识别的正确结果

图3-9 FGNET 数据库中识别正确的部分人脸图像

Fig.3-9 Examples of successful retrieve results in FGNET

待测试图像



15岁 23岁 2岁 14岁 19岁 14岁

9岁 8岁 2岁 67岁 13岁 10岁

9岁 4岁 8岁 22岁 6岁 3岁

识别结果

正确结果

图3-10 FGNET 数据库中识别错误的部分人脸图像

Fig.3-10 Examples of failed retrieve results in FGNET

### MORPH数据集上的实验

本节在MORPH Album2数据库上进行了实验，依据文献[4-5,75]对训练集和测试集的划分，随机选取10000人作为测试集，每个人选取年龄跨度最大的两张图片，年龄较小的10000张人脸图片作为gallery 集，年龄较大的作为probe集，再另外选取其他的1000人作为训练集，实验结果如表3-4所示。

本章所提方法针对的是年龄跨度较大的人脸识别，因为MORPH数据库单个人的训练样本较少，年龄跨度很小，且包含较大的姿态、表情的干扰，并不能充分展现本章所提方法的优越性，尽管如此，本章所提方法依然有相对较好的识别效果， 比直接提HOG特征效果好很多。

表3-4 本章方法与已有方法在MORPH数据集上的比较

Table 3-4 Comparison with other methods on MORPH database

|  |  |
| --- | --- |
| 方法 | 识别率 |
| Park et al.[2] | 79.80% |
| Du et al.[72] | 79.24% |
| Klare et al.[74] | 79.08% |
| Otto et al.[77] | 81.27% |
| Li et al.[75] | 83.90% |
| Gong et al.[4] | 91.14% |
| Chen et al.[73] | 92.80% |
| Gong et al.[5] | 92.26% |
| Li et al.[129] | 87.13% |
| HOG | 43.65% |
| 本章方法(HOG+AIFSL+Euclidean) | 88.43% |
| 本章方法(HOG+AIFSL+Cosine) | 89.57% |

## 本章小结

本章提出了一种非年龄相关的身份特征子空间方法，该方法将人脸特征编码到两个独立的子空间，在只包含身份信息的身份子空间进行人脸识别可有效地克服年龄干扰的影响。不同于通用的人脸识别模型，本章主动考虑年龄干扰对人脸识别的影响，直接去除年龄特征存在一定的困难，但是通过年龄识别的任务则很容易获取到这部分特征。本章引入了两个字典：身份字典和年龄字典，在字典学习过程中同时考虑对身份类别和年龄类别的约束，由此学到的身份字典能够对人脸特征中的身份部分进行表示，同时能够保证经过该身份字典的编码，同类人的编码系数是近似的，不同类的人是能区分开的。学到的年龄字典能够保证捕捉到的年龄敏感信息能完成年龄估计的任务，由此将年龄干扰信息有效的区分开，保证身份特征是足够干净的。在MORPH和FGNET数据库上的实验结果证明了本章方法的有效性，此外，本章引入的年龄字典可以被替换为表情、光照等其他干扰字典，有望用来解决类似的类内干扰较大的人脸识别问题。

# 基于卷积神经网络的跨年龄人脸识别

## 引言

上一章提出了非年龄相关的身份特征子空间学习算法，其主要思想是主动考虑年龄特征对人脸识别的影响，在提取完局部特征以后，引入两个字典：身份字典和年龄字典，将人脸特征编码到两个独立的子空间去，在身份子空间实现了跨年龄的人脸识别。然而这样先提特征再进行特征选择的分步过程，因为手工特征的局限，没有直接与任务的完成以及完成的效果进行关联，很大程度上限制了最终人脸识别的效果。本章希望能够将人脸特征的提取过程与跨年龄的人脸识别任务建立直接联系，实现“端到端”的特征学习，同时主动考虑年龄带来的干扰对人脸识别的影响，引入年龄估计任务，辅助人脸识别任务去选择与年龄无关的稳定的人脸身份特征。

传统特征提取算法主要有Gabor[20]、LBP[21]、HOG[22]等，这些特征均为人为设计的特征提取模式，无法保证获取到所有与任务相关的特征，一定程度上也决定了最终分类结果的上限，所以本章选择采用“端到端”的深度学习来克服这一局限，图4-1是传统方法与深度学习方法的框架对比。

输入

特征提取

特征

特征选择

输出

传统机器学习算法框架

输入

深度学习算法

输出

深度学习算法框架

图4-1传统的机器学习算法与深度学习框架比较

Fig.4-1 Comparison of traditional machine learning and deep learning framework

基于卷积神经网络的深度学习方法通过对任务构建合理的目标函数指导网络去学习完成任务所需的特征，克服了传统特征提取的局限，但是在跨年龄的人脸识别领域研究相对较少，主要原因就是缺乏足够多的包含年龄信息的人脸数据库用以提供训练，已有的年龄数据集FGNET、MORPH等跨年龄的人脸图像也相对较少。文献[79]用大规模的人脸数据集训练一个“底座”，然后再在年龄数据集训练过程中引入身份隐变量模型指导卷积神经网络学习非年龄相关的身份特征，然而隐变量模型主要依赖因子高斯分布的假设，文章并未给以合理的说明。本章类似地采用在大的人脸数据集上先进行预训练，在此基础上再对身份特征和年龄特征进行合理的拆分。

人脸识别任务期望获得对年龄等其他干扰鲁棒的身份特征，另一方面年龄估计任务期望获得对年龄变化敏感的特征，一定程度上，这两个任务分别期望获取到人脸图像中随着年龄变化相对稳定和不断变化的特征，而这两部分特征又能够重构出每个年龄阶段人脸的整个特征。所以本章希望在身份识别任务的基础上引入年龄估计的任务，帮助身份识别任务去选择与年龄干扰无关的特征。基于深度学习算法的多任务学习思想在人脸检测[90-91]，特征点定位[84]，行人检测[83]，属性估计[130]等任务上获得了应用，然而在人脸识别相关任务上的研究还相对较少。TCDCN[84]将特征点检测和识别眼镜、笑脸、性别和姿态任务结合起来，多个任务相辅相成，共同提高识别性能；CJCRF[131]将特征点检测和表情识别任务结合起来，特征点的检测任务可以帮助表情识别任务更好地关注特征点的变化。这类多任务学习主要是通过共享基础的卷积层获得低层次的纹理信息，通过定义各自的任务层获得与各自任务相关的高层次的纹理特征，多个任务之间相互提示，可以学到单个任务容易忽视的重要特征。此外通过使用包含在相关任务的约束信息中的知识也能一定程度上提高模型的泛化性能[85]。

不同的是，我们这里的身份识别任务和年龄估计任务对高层次特征的选择方向不是一致的，所以我们需要更加关注这两个任务之间的关系。在类似的抗姿态干扰的问题上，文献[92]将人脸姿态的重构或者识别和人脸识别任务结合起来，文献[80]也是从多任务学习的角度解决跨年龄的人脸识别，但直接用学到的人脸特征减去对应的年龄特征缺乏足够的说服力。我们这里引入年龄估计作为辅助任务，在获得共享的特征基础上，年龄估计任务约束身份任务选择与年龄变化无关的特征，如图4-2所示，基于深度卷积网络CNN获得了人脸图像的关于这两个任务共享的特征，随后特征选择层的映射矩阵对上述的共享特征进行挑选，年龄估计任务选择与年龄相关的特征，同时指导身份识别任务不要选择该部分的特征，两者相辅相成，人脸识别任务能够挑选到对年龄鲁棒的只与身份相关的特征。

含身份、年龄标签的人脸图像

共享特征

辅助

身份识别

年龄估计

身份特征

年龄特征

图4-2基于卷积神经网络的跨年龄人脸识别框架

Fig.4-2 The framework of cross-age face recognition based on deep convolution networks

深度卷积网络通用的物体分类模型一般采用Softmax loss作为目标损失函数，但是在人脸识别任务中只在类别上进行约束并不能保证同类之间能够聚拢，不同的类别特征能够分开。DeepID2[49]在softmax目标函数的基础上叠加上了样本两两之间是否属于同类的验证信号；Google facenet[52]更进一步，直接去掉Softmax层, 对同类样本和异类样本之间的距离进行约束。然而上述两种目标函数均需要在训练过程中考虑对应的采样算法，对训练策略也有很高的要求。文献[54]在Softmax的基础上，引入Center loss，在保证类间分离的同时对每个样本到类中心的距离进行约束，不需要额外构造训练的样本对，也保证了特征的区分度。本章在人脸识别和年龄估计任务上在考虑对类别做区分的同时也会充分考虑特征的区分度。

本章从多任务学习的角度来解决跨年龄的人脸识别任务，在主任务人脸识别的基础上加上辅助任务年龄估计，由深度卷积网络获取到共享的人脸特征之后，年龄估计任务指导任务层的参数矩阵去对应挑选与年龄估计相关的特征，同时帮助身份识别任务排除掉该部分的特征。由于这两个任务有不同的噪音模式，同时学习这两个任务可以增加模型的泛化能力。模型的目标函数中除了对类别进行约束,还对特征的聚拢程度进行了约束，充分保证同类人的映射特征在特征空间中能够聚拢，不同类的人能够最大限度地分开。由此，基于多任务学习的深度神经网络的能够学到较为干净且具有表征能力的身份特征。在识别阶段，通过计算两幅人脸图像的身份特征之间的距离度量，采用KNN分类器给出最终人脸识别结果。

## 基于多任务深度卷积神经网络的跨年龄人脸识别

本章希望引入年龄估计的任务辅助网络学习对年龄鲁棒的身份特征，实现跨年龄的人脸识别，下面主要从网络结构、目标函数设计以及模型求解三方面进行介绍。

### VGG网络基础模型

本章采用的基础深度卷积网络是VGG[43]，VGG-Net是牛津大学visual geometry group（VGG）组的Karen Simonyan 和Andrew Zisserman提出的，该模型在2014年ILSVRC竞赛上获得了定位任务一名、分类任务第二名。VGG网络的输入是固定大小（224\*224）的RGB图像，网络主体由5个卷积组组成，如图4-3所示，每个卷积组里面包括两个或者三个卷积层，每一个卷积组的后面接一个最大池化层，达到降低特征维度，减少网络参数的目的，经过降采样的特征图具有平移、旋转和缩放不变性。经过5个卷积组和5个最大池化层后，接3个全连接层，最后是softmax分类层。

Conv2\_1

Conv1\_2

Conv1\_1

Pooling

Conv2\_2

Pooling

Conv3\_1

Conv3\_2

Conv3\_3

Pooling

Conv4\_1

Conv4\_2

Conv4\_3

Pooling

Conv5\_1

Conv5\_2

Conv5\_3

Pooling

全连接-1

Block1

Block2

Block3

Block4

Block5

全连接-3

全连接-2

图4-3 VGG 网络示意图

Fig.4-3 VGG Convolutional Neural Networks

VGG网络每个卷积层(conv)采用的是 3\*3大小的卷积核，滑动步长(stride)和扩充(padding)边缘的大小均为1。3\*3大小的卷积核是包括上下左右以及中心像素的最小尺寸，能够提取到特征图中更密集丰富的特征，同时两个卷积层堆叠对应的感受野是5\*5，三个3\*3的卷积层堆叠对应的感受野是7\*7，相比5\*5或者7\*7的卷积层，其参数量要相对少很多。

网络每个卷积组后面接的最大池化层滤波器的大小为2\*2, 对特征图中每一个2\*2大小的特征进行聚合统计，取该区域中的最大值作为该部分的特征，输出的特征图变为原始特征图的四分之一大小，减少了网络模型的参数，同时增加了后续网络层的感受野。

网络的每个卷积层和全连接层后面接激活函数是ReLU，对提取的特征进行非线性映射，增加网络对特征的非线性表达能力。在网络的前两个全连接层采用了dropout方法，防止网络的过拟合。本章主要采用的是VGG-16版本，具体的参数如表4-1所示：

表4-1 VGG-16网络参数配置

Table 4-1 The configuration of VGG-16 Network parameters

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 输入图像 | 3\*224\*224 | | | | |
| Block设置 | Conv 层 | Conv 层 | Conv 层 | Pooling层 | 输出特征 |
| Block 1 | 卷积核：3\*3  核数量：64  步长：1  扩充：1 | 卷积核：3\*3  核数量：64  步长：1  扩充：1 | \_\_\_\_\_\_\_\_\_\_\_ | 类型:MAX  核大小：2\*2  步长：2 | 112\*112\*64 |
| Block 2 | 卷积核：3\*3  核数量：128  步长：1  扩充：1 | 卷积核：3\*3  核数量：128  步长：1  扩充：1 | \_\_\_\_\_\_\_\_\_\_\_ | 类型:MAX  核大小：2\*2  步长：2 | 56\*56\*128 |
| Block 3 | 卷积核：3\*3  核数量：256  步长：1  扩充：1 | 卷积核：3\*3  核数量：256  步长：1  扩充：1 | 卷积核：3\*3  核数量：256  步长：1  扩充：1 | 类型:MAX  核大小：2\*2  步长：2 | 28\*28\*256 |
| Block 4 | 卷积核：3\*3  核数量：512  步长：1  扩充：1 | 卷积核：3\*3  核数量：512  步长：1  扩充：1 | 卷积核：3\*3  核数量：512  步长：1  扩充：1 | 类型:MAX  核大小：2\*2  步长：2 | 14\*14\*512 |
| Block 5 | 卷积核：3\*3  核数量：512  步长：1  扩充：1 | 卷积核：3\*3  核数量：512  步长：1  扩充：1 | 卷积核：3\*3  核数量：512  步长：1  扩充：1 | 类型:MAX  核大小：2\*2  步长：2 | 7\*7\*512 |
| 全连接层1 | 4096个输出节点 | | | | 1\*1\*4096 |
| 全连接层2 | 4096个输出节点 | | | | 1\*1\*4096 |
| 全连接层3 | 1000个输出节点 | | | | 1\*1\*1000 |
| softmax | | | | | |

### 跨年龄人脸识别的网络结构设计

一方面，不同的人在相同的年龄段具有相似的年龄信息，如皱纹等，另一方面，同一个人在不同的年龄段与其身份相关的信息是稳定的。在跨年龄的人脸识别任务中需要去获取这部分稳定不变的特征。直接用深度卷积网络获得的特征不可避免的还包含年龄的相关干扰信息。然而这部分年龄相关的干扰信息很容易从年龄估计的任务中获取，因此本章设计的网络希望用年龄估计的任务辅助网络去学习只含身份信息的人脸特征。

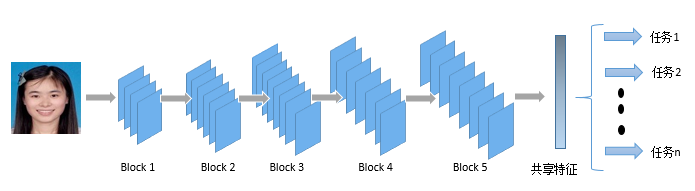


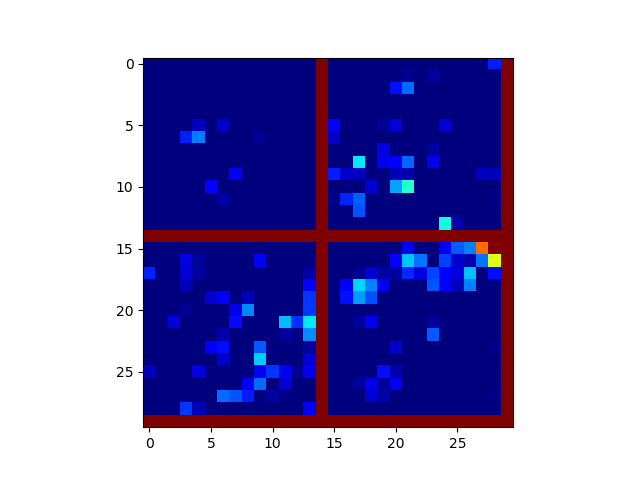
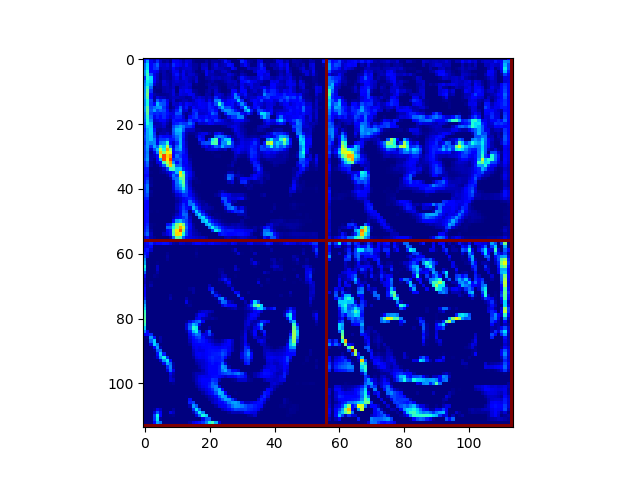
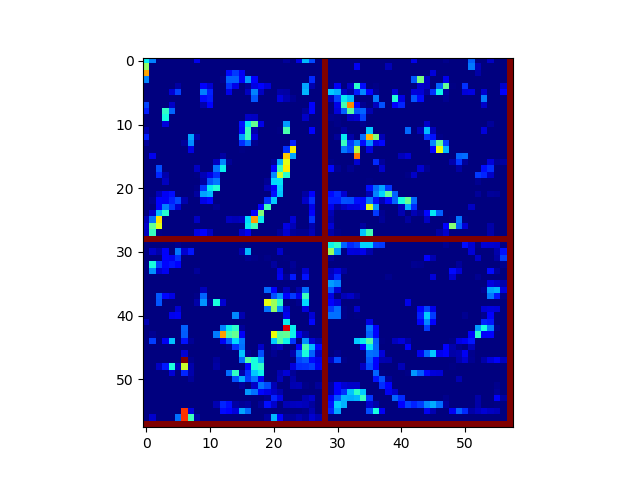
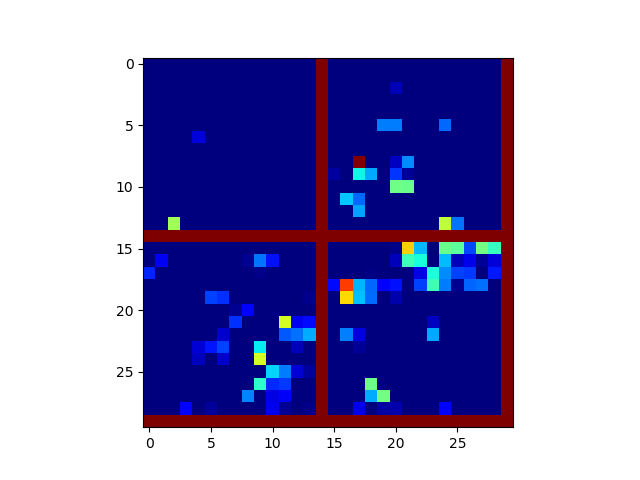
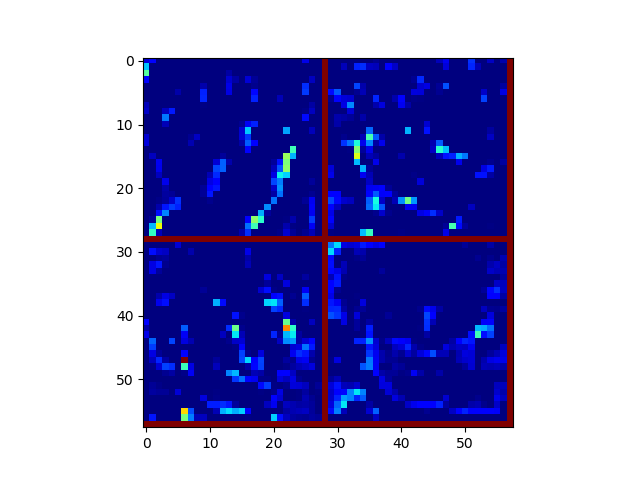
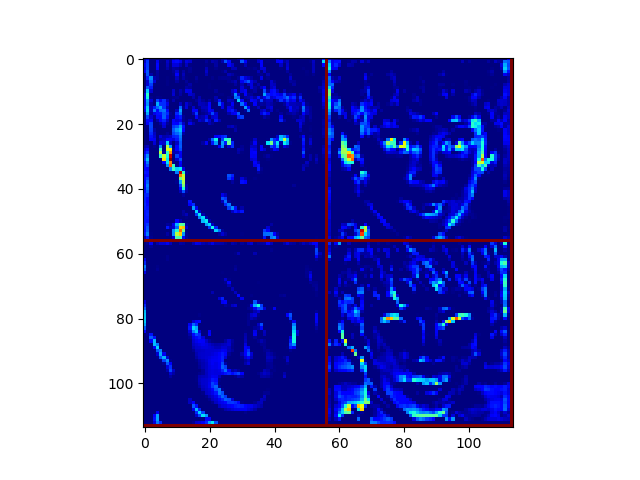
图4-4 基于深度卷积网络的多任务学习的经典结构

Fig.4-4 The structure of multi-task learning based on deep convolution network

本章的网络结构设计首先要保证网络能够自主学习与身份识别对应的身份特征，学习与年龄估计对应的年龄特征，已有的基于多任务学习的深度卷积网络，大都让网络特征输出层学习一个共享的特征，然后基于任务的不同在分类层进行特征选择，选择与各自任务相关的特征，如图4-4所示。然而这样设计的网络结构只适用于具有相同学习方向的多任务特征，此外，即使这些多任务具有共享的低层级纹理信息，但是高层的抽象语义特征是否也一致有待证明。

针对年龄特征和身份特征的学习，学习一个完全一致的高层级的共享特征可能会增加后续特征选择的难度，我们需要考虑设计合适的网络去学习这两个特征。

参考人工特征的设计准则（HOG、LBP、SIFT），不管是提取身份特征还是年龄特征，都需要关注人脸的轮廓、边缘、角点以及低层次表征像素区域关系的特征。这部分一定程度上可以通过深度卷积网络通过共享一致的网络结构和参数来获得，但是高层级的抽象特征能否通过一致的方式获得，已有的众多基于深度卷积网络的多任务学习并未给予说明[84,90-91,130-131]。针对VGG网络，用基于身份识别任务训练的模型和基于年龄估计任务训练的模型分别对一张人脸图像提取特征，图4-5给出了低层级的Block2,Block3,Block4输出的特征图，针对不同的任务Block2、Block3输出的低层次的边缘特征具有很大的相似性，而Block4提取的特征开始逐渐有一些差异。



身份识别

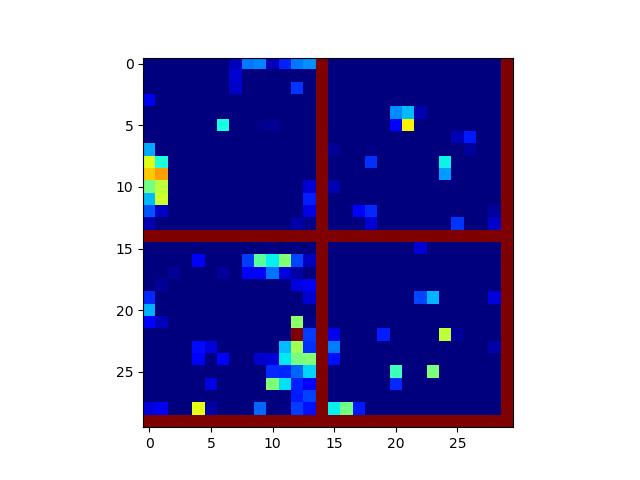
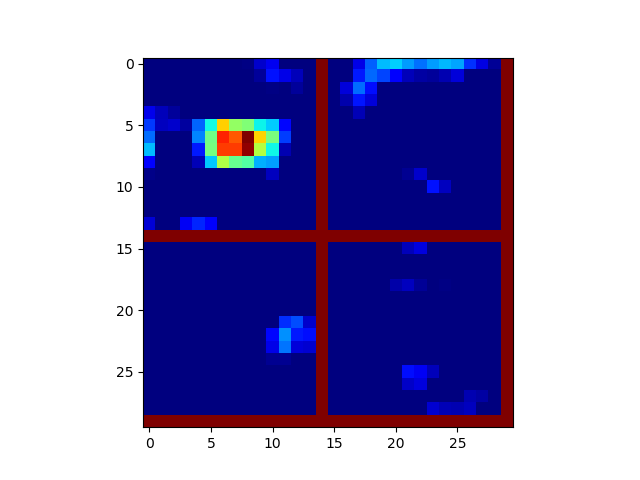
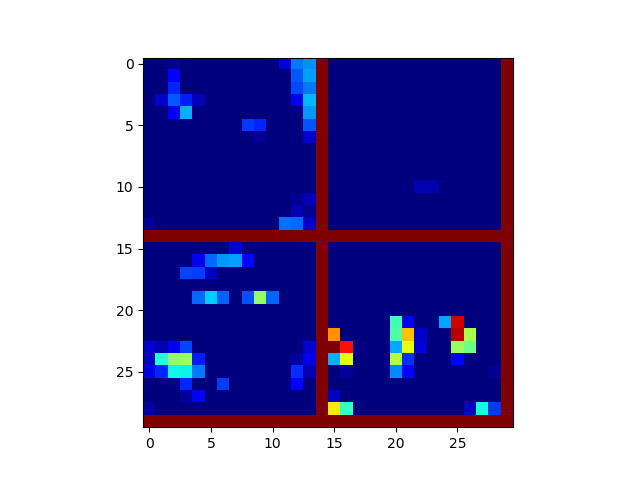
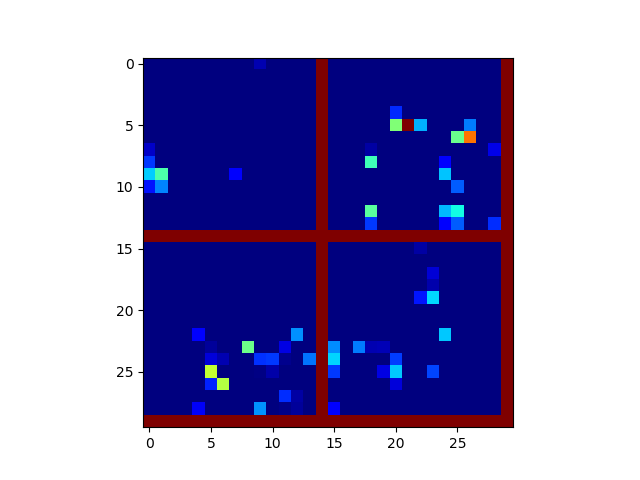
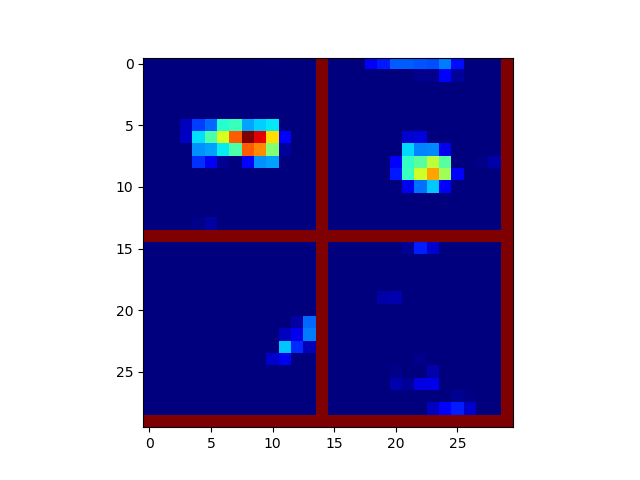
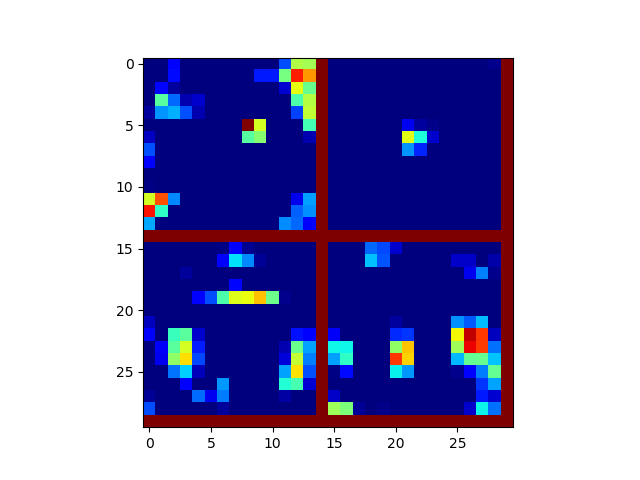
年龄估计

Block2 Block3 Block4

图4-5不同任务低层级网络可视化对比效果图

Fig.4-5 Visualization of the low level layers’ feature maps from different tasks

针对网络的第5个Block输出的特征，如图4-6所示，可以清晰的发现针对不同的任务，其抽象意义的特征还是存在较为明显的区别的。



年龄估计

身份识别

con5\_1 con5\_2 con5\_3

图4-6不同任务高层级(Block5)网络可视化对比效果图

Fig.4-6 Visualization of the high level layers’ feature maps from different tasks

上述的可视化分析已经一定程度上说明，为了让网络自主学习对应的身份特征和年龄特征，单纯地只在特征输出层，即分类层的前一层（fc7层）提取完全一致的共享特征存在一定的局限性。因为从Block4结束，con5\_1卷积层开始，这两个任务所希望获取的语义特征已存在明显差异，如果提取一个完全一致的共享特征，无疑会增加后续在任务层进行特征选择的难度。

结合上述对网络结构各层特征图的分析，初步证明了基于多任务学习的网络能够通过低层级参数共享的网络层提取人脸图像的边缘、轮廓等初级纹理信息，同时满足人脸识别和年龄估计的任务；对于更高层级的与任务相关的抽象特征则无法完全保证可以由共享的网络参数获得，因此本章对于高层级的卷积网络层设计有两种配置，即分别考虑在VGG网络的第5个Block后获得共享的高层级抽象语义特征，第4个Block后获得统一的基础纹理特征后再各自提取与任务相关的抽象语义特征。

如果采用上述共享的卷积网络配置，已有的多任务学习均是在第三个全连接层选择与各自任务相关的特征，针对本章的任务，这样的设计也存在一些局限，为了说明该问题，参考文献[132]中对特征重要程度的描述，本章引入能量函数来描述某一神经元对特征的表示能力。令第层的第个神经元与第层的第j个神经元之间的权重参数为，n为第层神经元的总数，则图4-7中第i个神经元所对应的能量值：

 （4-1）

K

K+1

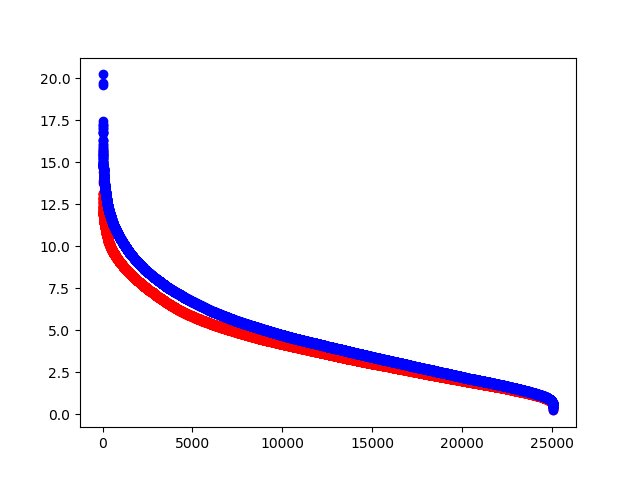
图4-7神经元能量函数定义

Fig.4-7 The definition of neuron energy function

越大，说明该神经元重要程度越大，对特征的响应能力越强，与任务的相关度越大。根据神经元能量值的定义，则第k层的神经元能量分布可以获得。我们对上述身份识别和年龄估计任务网络的pool5层以及fc6层神经元的能量分布函数进行了统计，横轴为对应每层神经元，纵轴为神经元的能量值，为了方便观察，将能量值从大到小进行排序，如下图4-8所示：

pool5层神经元能量分布

能量值



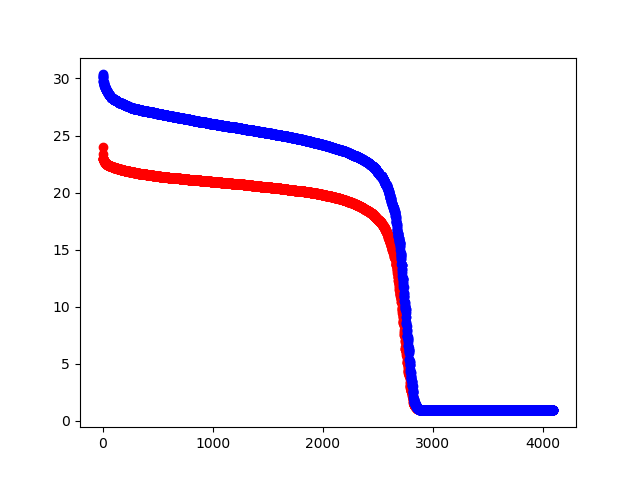
神经元

年龄估计

身份识别

fc6层神经元能量分布

能量值



神经元

身份识别

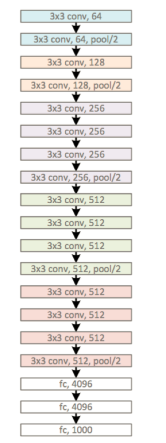
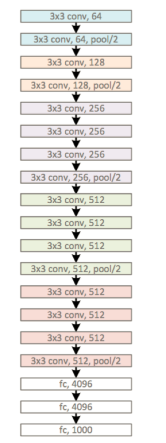
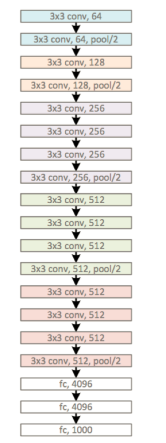
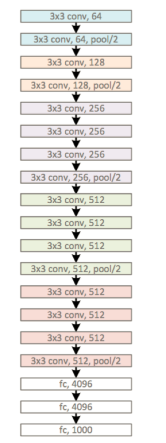
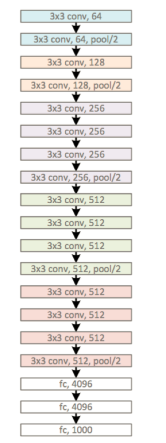
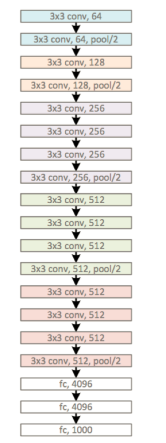
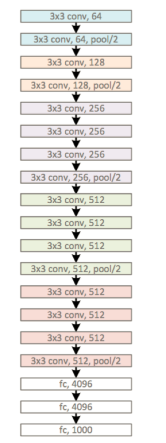
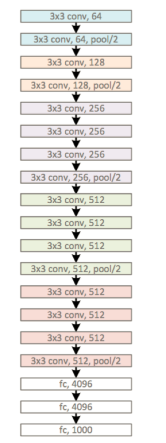
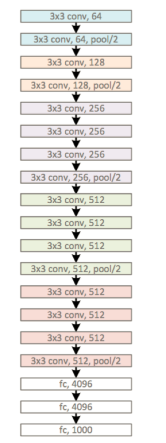
年龄估计

图4-8 不同任务神经元能量分布图

Fig.4-8 Neurons energy distribution of different tasks

神经元能量分布的不同，说明了身份识别和年龄估计对于统一的共享特征图特征响应是不一致的，即两者关注的与任务相关的特征是不一样的，从pool5层到第一个全连接层(fc6)之间神经元能量的分布已不一致，即两个任务所选择的特征已出现差异，第二个全连接层（fc6层）到第三个全连接层两个任务关注的特征差异越来越大，因此针对上述的共享配置情况(即Block5获得一致的特征图)，最理想的情况是从fc6即可开始选择与各自任务相关的特征，为了后续的实验比对，我们也设置了从fc7层以及最后的任务层做特征选择的情况。

综上，为了后续能够进一步从实验证明我们上述的分析，本章一共设计了4种多任务学习的网络结构，其中VGG\_AIFR\_A对应在Block5开始选择各自任务相关的特征；VGG\_AIFR\_B，VGG\_AIFR\_C，VGG\_AIFR\_D分别是从fc6层、fc7层以及任务层做特征选择的情况，其中VGG\_AIFR\_A，VGG\_AIFR\_B两种情况是通过上述分析相对合理的配置情况，VGG\_AIFR\_C，VGG\_AIFR\_D是为了与已有的通用的多任务方法进行比较而设置的。



VGG\_AIFR\_A VGG\_AIFR\_B VGG\_AIFR\_C VGG\_AIFR\_D

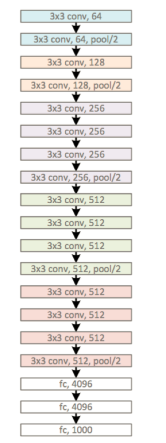


图4-9跨年龄的人脸识别网络结构设计

Fig.4-9 The networks of cross-age face recognition

### 跨年龄人脸识别的目标函数设计

上一节主要对基于深度卷积网络的跨年龄人脸识别的网络结构进行讨论，让网络能够自主学习对应的身份特征和年龄特征，本小节主要从两个方面对任务的完成进行约束，一是引入年龄估计任务辅助身份识别学习干净的身份特征，二是保证所学的身份特征对不同的人具有足够的区分度。下面主要从特征选择角度、特征分类以及Center loss三个方面介绍文章的目标函数设计。

#### 4.2.3.1 特征选择

如何通过年龄估计任务辅助人脸识别任务去挑选合适的特征是本章需要考虑的，在类似的抗姿态干扰问题上，文献[132]通过隐式的学习保证多任务的目标函数从共享特征中选择与自身任务相关的特征，实验证明了引入姿态干扰任务可以帮助身份识别任务选择只与身份相关的特征；文献[80]通过显式的约束方式，即用学到的共享特征减去年龄估计学到的年龄特征来获得非年龄相关的特征。本章这里考虑的是隐式的学习方式，如图4-10所示。

给定包含N张图片的训练集：，包含C个身份类别，K个年龄类别，其中 对应输入的第i张人脸图像，对应第张图片的身份标签和年龄标签。令为卷积神经网络学到的某张图片身份和年龄共享的特征，为卷积网络对特征进行选择之前的卷积核参数及偏置，对应从输入图像到共享特征的非线性映射，则学到的共享特征可以表示为：

 （4-2）

定义特征选择层身份识别任务的映射矩阵和偏置向量, ，则由上述共享特征经过身份权重矩阵的映射获得对应的身份特征：

 （4-3）

类似的定义特征选择层年龄估计任务的权重矩阵和偏置向量, ，则由上述共享特征经过特征选择层的权重矩阵的映射获得对应的年龄特征：

 （4-4）

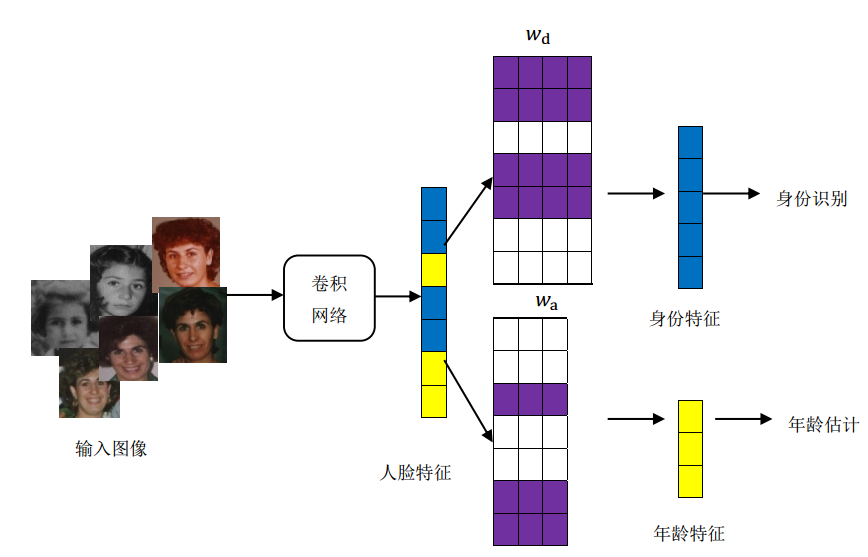


图4-10跨年龄的人脸识别特征选择示意图

Fig.4-10 Feature selection of cross-age face recognition

#### 4.2.3.2 特征分类

根据4.2.2节的网络结构模型，从共享特征中选择出身份特征与年龄特征后，模型VGG\_AIFR\_A、VGG\_AIFR\_B、VGG\_AIFR\_B需要再经过各自任务相关的卷积层或者全连接层，身份任务和年龄任务的参数分别为，获得对应的身份特征进入softmax分类层，计算该身份特征对应所属的身份类别：

 （4-5）

softmax分类层将输入身份特征映射到（0,1）之间的概率值，方便进一步分类，最终分类的结果可以根据最大化softmax的输出获得：

 （4-6）

同样，获得对应的年龄特征进入softmax年龄分类层，计算该年龄特征对应所属的年龄类别：

 （4-7）

最终分类的结果可以根据最大化softmax的输出获得：

   （4-8）

通过上述特征选择层的映射矩阵，原始的共享特征将身份特征和年龄特征有效的区分开，由此可实现跨年龄的人脸识别。给定训练集T，为了学习到对应的卷积神经网络特征选择层的权重矩阵以及其他层的网络参数（包括共享特征提取过程的权重参数以及各个任务层自身的参数），本章所提的模型需要最小化身份分类交叉熵损失和年龄分类的交叉熵损失之和：

 （4-9）

其中控制年龄估计和身份识别任务之间的权重比， 分别为真实的身份和年龄类别，当时，模型只包含身份识别的任务。

共享特征

年龄特征

身份特征



身份识别

年龄估计



身份1 身份2 身份3

年龄1 年龄2 年龄3

图4-11跨年龄的人脸识别特征分类示意图

Fig.4-11 Feature classification of cross-age face recognition

#### 4.2.3.3 Center Loss约束

上述softmax分类的交叉熵损失只是关注图像属于哪个类别，对单个分类任务中特征的可分离程度没有给予足够的关注。通常我们期望学习到的特征不仅可分，而且能保证不同类间之间有足够的区分度，同时提取到的特征要具有广泛的表征能力，能够依据此特征提取模型去处理不包含在训练数据标签中的类别。如图4-12所示，单纯地采用softmax分类能够对类别进行约束，类似左侧的效果，但实际人脸识别应用中，实际训练和测试的人多为不同的类别，所以在保证不同类别的特征能够区分开的同时还需要保证同类样本的特征距离要尽量远小于异类样本之间的距离，同类的尽量往类的中心聚拢。因此，我们引入Center Loss[54]对特征进一步约束。

(a)可分离的特征 (b)具有判别能力的特征

图4-12 特征判别能力的比较

Fig.4-12 Comparison of the separable and discriminative feature

每个类别的数据都有类内的中心Center，Center loss出发点是惩罚每个类别的样本与该类别中心的距离，使得同类的样本尽量聚集在一起。在上述特征选择的基础上， 为第i张图片的身份特征值，为所属类别的中心：

 （4-10）

同样，我们对年龄特征也做类似的约束：

 （4-11）

由此，最终的损失函数可以描述为：

 （4-12）

其中控制Cneter loss所占的比重。为了方便后续求解的描述，将Softmax loss的参数与前面各自的任务层参数统一描述为。

#### 4.2.3.4 模型求解

Softmax loss有两个输入，一个是样本的真实标签，一个是预测的结果。针对身份识别的Softmax loss，求解过程如下：

令，, 其中：

 （4-13）

由此可通过式(4-14)直接求解:

 （4-14）

年龄估计的Softmax loss求解过程类似。

针对Center loss的更新，每次迭代过程中，根据该分类特征值的平均值进行更新，为了防止Center变化剧烈，采用权重系数控制Center的学习率，每一次迭代计算当前数据与Center的距离，然后把这个距离以梯度形式叠加到Center上。

针对身份识别任务，Center loss具体更新过程如下所示，每一次迭代，根据每个batch计算每一类特征的中心值（该种情况下，并不是每一类的特征中心都会被计算），当\*条件满足时，，否则。

 （4-15）

 （4-16）

 （4-17）

类似的年龄估计的求解方法类似，。

基于深度卷积网络的跨年龄人脸识别算法如算法4-1所示。

**算法4-1 基于深度卷积网络的跨年龄人脸识别算法**

输入：训练数据.初始的卷积层参数，特征选择层的参数为特征选择后的卷积层或者全连接层参数，超参数，以及学习率，迭代次数 。

输出：卷积层参数特征选择层参数特征选择后各任务相关层参数.

1: 

2: 计算联合损失函数

3：分别计算两个任务的反向传递的误差,.

4: 更新各自任务层参数.

5: 更新各自任务的类中心.

6：更新特征选择层的参数.

7：更新卷积层的参数.

8：重复1-8,直至收敛。

## 基于Cosine距离和KNN的跨年龄人脸识别

在模式识别领域中，k最近邻算法（k-Nearest Neighbor, KNN）是一种非参的机器学习算法，既可以用来解决回归问题，也可以解决分类问题[25]。做回归问题的结果是取决于最近K个值的均值，做分类问题的结果主要取决于最近k个值所属类别的“投票表决”。如图4-13所示是K最近邻算法用来做分类问题的一个案例，正方形和三角形代表样本的不同类别，中间的圆形是测试的样本。k取3时，测试样本最近邻为两个三角形和一个正方形，根据多数表决的结果，最终测试样本属于三角形所属类别，类似的，如果k取5时，测试样本最近邻为三个正方形和两个三角形，根据多数表决的结果，最终测试样本属于正方形所属类别。

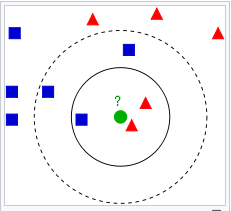


图4-13 KNN分类示意图

Fig.4-13 The visualization of KNN algorithm

KNN分类结果的优劣主要取决于k值和样本间相似性的度量，一般通过交叉验证方法来选择k值，欧式距离，Cosine距离以及其他距离度量可以用作样本间的相似性度量。KNN算法简单有效，不需要额外的训练，训练时间复杂度为0，在实际人脸识别应用系统中也多采用此分类方式。

本章采用基于Cosine距离度量的分类方式，该距离度量更加关注两个特征向量在方向上的差异，针对学习到的两个特征向量，其对应的cosine距离为：

 （4-18）

在本章的跨年龄人脸识别任务中，由上述深度卷积网络的倒数第二层，获取到对应的身份特征和年龄特征以后，采用Cosine距离作为度量方式，直接采用KNN做分类。

## 实验结果分析

### 实验设置

在训练过程中，所有训练图像经过第二章节的人脸图像预处理，并进行数据增强操作，包括水平翻转，裁剪归一化到224\*224 大小，如图4-14所示，针对图像的每一个像素值，减去所有训练集计算得到的RGB均值。

（a）原图 （b）归一化后

图4-14人脸图像归一化

Fig.4-14 Normalization of face image

模型是在Caffe上运行的，batch的大小为8，因为基础的VGG网络在人脸数据库上进行了预训练，所以本章的学习率初始设置为1e-4，根据loss的下降情况调整衰减步长，经过多次参数比较实验，身份与年龄的Softmax损失权重设置为0.7和0.3，Center loss权重均设置为0.008。采用随机梯度下降(SGD)方法更新参数。为了评估识别的效果，在学到的身份特征的基础上采用cosine距离和最近邻方法做分类。

### MORPH数据集上的实验

MORPH II数据集包含55134张人脸图像，13618个人，每个人的年龄跨度相对较小些，且包含较大的姿态、表情干扰。我们将数据集拆分成训练集和测试集。训练集10000人，其中随机选出一部分样本作为验证集，测试集1000人，每人两张人脸图像，其中年龄较小的作为gallery集，年龄较大的作为probe集。

#### 4.4.2.1 不同网络结构的比较

结合4.2.2节对网络结构的分析，这里从实验角度进一步验证上述的分析，VGG\_AIFR\_A对应在Block5开始选择各自任务相关的特征；VGG\_AIFR\_B，VGG\_AIFR\_C，VGG\_AIFR\_D分别是从fc6层、fc7层以及任务层做特征选择的情况。

表4-2不同网络结构的比较

Table 4-2 Comparison of different networks architecures

|  |  |  |
| --- | --- | --- |
| 模型 | 训练集准确率 | 验证集准确率 |
| VGG\_AIFR\_D | 99% | 88.6% |
| VGG\_AIFR\_C | 99% | 88.7% |
| VGG\_AIFR\_A | 99% | 89.0% |
| VGG\_AIFR\_B | 99% | 91.3% |

实验结果显示VGG\_AIFR\_A，VGG\_AIFR\_B这两种结构的效果最好，也应证了上述对多任务的网络结构分析。因此，本章的后续实验继续使用VGG\_AIFR\_B这个模型。

#### 4.4.2.2不同模型的比较

本章基础的VGG网络采用的是Softmax loss对类别进行约束，在此基础上我们引入Center loss对特征的区分度进一步约束，保证同类的特征聚在一起，这里分别用基础的VGG网络单独对身份任务进行训练(VGG\_Id)，加上年龄估计作为辅助任务进行的训练(VGG\_Id+Age)和对多任务学习损失函数加上Center loss训练(VGG\_Id+Age+CenterLoss)三种模型在MORPH数据集上进行训练，在验证集上测试，横轴为迭代次数，纵轴为识别率，效果如下图4-16所示：

图4-16不用模型在验证集上的识别效果

Fig.4-16 Comparison of different models on Validation set

图4-16表明，在身份识别任务中引入年龄估计任务，相比于单任务的人脸识别，在验证集上识别效果更好，在softmax Loss的基础上加上Center Loss进一步提高了特征的判别能力。表4-3给出了不同模型在Gallery、Probe集上的比对结果。相比于VGG\_baseline，本文所提方法在识别率上有了很大提高。加入年龄估计任务可以提高模型的泛化能力，辅助人脸识别任务进行特征选择。引入的Center loss可以提高模型的识别效果。

表4-3不同模型在Gallery集\Probe集上的比较

Table 4-3 Comparison of different models on gallery and probe sets

|  |  |  |  |
| --- | --- | --- | --- |
| 方法 | Top1 | Top3 | Top5 |
| VGG\_Baseline | 85.0% | 90.5% | 92.4% |
| VGG\_Id | 87.3% | 92.0% | 93.6% |
| VGG\_Id+Age | 88.4% | 92.5% | 94.1% |
| VGG\_Id+Age+CenterLoss | 89.7% | 92.9% | 94.6% |

#### 4.4.2.3 与现有方法的比较

为了进一步说明本文所提方法的有效性，我们与已有文献的方法进行了比较，如表4-4所示。

表4-4 本章方法与已有方法在MORPH数据集上的比较

Table 4-4 Comparison with other methods on MORPH database

|  |  |
| --- | --- |
| 方法 | 识别率 |
| Park et al.[2] | 79.80% |
| Du et al. [72] | 79.24% |
| Klare et al.[74] | 79.08% |
| Otto et al.[77] | 81.27% |
| Li et al.[75] | 83.90% |
| Gong et al. [4] | 91.14% |
| Chen et al. [73] | 92.80% |
| Gong et al.[5] | 92.26% |
| Li et al.[129] | 87.13% |
| HOG | 43.65% |
| HOG+AIFSL+Euclidean | 88.43% |
| HOG+AIFSL+Cosine | 89.57% |
| VGG\_baseline | 85.00% |
| 本章方法 | 89.70% |

从表中可以发现，VGG\_baseline模型在MORPH数据集上只有85%的识别率，但是比传统的基于HOG特征的识别效果要好很多，充分证明了深度卷积网络在特征提取能力上的优势。此外本章方法获得89.70%的识别效果，比基础的VGG网络效果要好很多，也比基于非年龄相关的身份子空间学习算法识别效果要好。此外，考虑到如下两个方面，一方面VGG基础网络本身在MORPH数据集上的表现不是特别突出，另一方面考虑到该数据库上相对而言姿态、表情变化更加复杂，本章所提方法的优势不能完全显示出来，尽管如此，本章方法该数据库上也取得较好的识别效果。



测试集



Rank-1

真实结果

29岁 35岁 54岁 44岁 48岁 28岁

26岁 23岁 43岁 39岁 41岁 25岁

图4-17 MORPH 数据库中识别正确的部分人脸图像

Fig.4-17 Examples of true retrieve results in MORPH



Rank-1

测试集

真实结果

27岁 25岁 27岁 35岁 27岁 26岁

25岁 24岁 30岁 20岁 24岁 24岁

25岁 23岁 25岁 33岁 25岁 24岁

图4-18 MORPH 数据库中识别错误的部分人脸图像

Fig.4-18 Examples of failure retrieve results in MORPH

图4-17给出了部分识别正确的图像，第一行是要Probe集中要进行测试比对的人脸图像，第二行是在Gallery集中检索的排在Rank-1的识别结果。

图4-18给出了部分识别错误的人脸图像，第一行是待比对的人脸图像，第二行为对应的识别出来Rank-1的人脸图像，第三行为检索的库中实际与之对应的人脸图像，可以发现，识别出来的Rank-1的人脸图像一定程度上相比于真实的结果可能更接近测试集人脸图像。

### FGNET数据集上的实验

FGNET数据集包含来自82个人的1002张人脸图像，每个人的年龄跨度都比较大，依据文献[4-5,75]采用的留一法，我们将本章所提方法与已有方法进行了比较，如表4-5所示。

表4-5 本章方法与已有方法在FGNET数据集上的比较

Table 4-5 Comparison with other methods on FGNET database

|  |  |
| --- | --- |
| 方法 | 识别率 |
| Geng et al.(2008) [57] | 38.1% |
| Park et al.(2010) [2] | 37.4% |
| Li et al.(2010) [75] | 47.5% |
| Gong et al.(2013) [4] | 69.0% |
| Gong et al.(2015) [5] | 76.2% |
| HOG | 53.7% |
| 非年龄相关的身份子空间 | 79.8% |
| VGG-baseline | 78.8% |
| 本章方法 | 81.2% |

本章所提方法获得了相对较高的识别效果81.2%，明显高于VGG的基础方法，也比其他已有方法的识别效果要好很多。该数据集的年龄跨度相比于MORPH数据集要大很多，随着年龄的变化，人脸在形状和纹理上都出现很大差别，这极大地增加了识别的难度。实验充分证明了本文所提方法在跨年龄的人脸识别任务上的有效性。

7岁 11岁 1岁 49岁 36岁 18岁

0岁 4岁 18岁 20岁 40岁 8岁



测试集

Rank-1

真实结果

图4-19 FGNET 数据库中识别正确的部分人脸图像

Fig.4-19 Examples of true retrieve results in FGNET

其中识别正确的部分人脸图像如图4-19所示，第一行是要Probe集中要进行测试比对的人脸图像，第二行是在Gallery集中检索的排在Rank-1的识别结果。



Rank-1

测试集

真实结果

26岁 9岁 2岁 17岁 18岁 14岁

17岁 5岁 4岁 20岁 37岁 3岁

17岁 4岁 8岁 8岁 22岁 1岁

图4-20 FGNET 数据库中识别错误的部分人脸图像

Fig.4-20 Examples of failure retrieve results in FGNET

识别错误的部分人脸图像如图4-20所示。第一行是待比对的人脸图像，第二行为对应的识别出来Rank-1的人脸图像，第三行为检索的库中实际与之对应的人脸图像。该数据集年龄跨度相对较大，识别错误的结果在人脸形状和纹理上都随着时间发生了较大的改变。

%

图4-21 FGNET和MORPH数据集年龄分布

Fig.4-21 The distribution of the datasets with different ages.

考虑到MORPH数据集和FGNET数据集在年龄的分布上不太一致，FGNET数据集上超过61%的人小于16周岁。但是对于MORPH数据集，所有的人脸图像都是大于16周岁的。我们比较了这两个数据集人脸图像的年龄分布，如图4-21所示，横轴为年龄的分布，纵轴是各个年龄段所占的百分比。

我们进一步了验证本章所提方法在不同年龄分布下的有效性，表4-6给出了本章方法在FGNET数据集上不同年龄段的识别结果。

表4-6 FGNET数据集上不同年龄段的识别结果

Table 4-6 Performance of different age groups on FG-NET

|  |  |  |  |
| --- | --- | --- | --- |
| 年龄段 | 样本数量 | VGG\_baseline | 本章方法 |
| 0-4 | 193 | 39.38% | 40.41% |
| 5-10 | 218 | 66.05% | 72.01% |
| 11-16 | 201 | 64.68% | 72.63% |
| 17-24 | 182 | 84.62% | 85.71% |
| 25-69 | 208 | 92.79% | 93.26% |
| 0-16 | 612 | 62.42% | 65.69% |
| 17-69 | 390 | 90.26% | 92.05% |

表4-6的结果显示，本章所提方法相比于VGG\_baseline在各个年龄阶段都有提升。此外，我们该发现，0-16岁相比于17-69岁年龄段的人脸识别效果要差很多，相比于17-69岁的年龄段，人脸从0岁到16岁这一阶段在形状和纹理上变化要大很多，这一实验结果与人脸老化的自然规律比较吻合，也对我们以后的跨年龄人脸识别研究有一定的启发意义。

## 本章小结

本章基于多任务深度卷积网络实现跨年龄的人脸识别，主要有以下几个特点：首先引入深度卷积网络自主学习身份特征和年龄特征，克服了手工特征的局限，基于大数据训练的模型底座进行优化，克服了年龄数据集相对较小的约束；其次对多任务学习的网络结构进行分析，分别从特征图及参数的可视化、神经元的能量分布等角度为网络结构的设计提供指导，并针对本文任务的特点给出优选的网络结构；最后从任务完成的角度出发设计网络学习的目标函数，一方面充分利用年龄估计获取到的年龄特征辅助身份识别任务去捕捉相对干净的人脸特征，另一方面引入的Center loss损失函数进一步保证了身份特征的判别能力。在MORPH和FGNET数据集上的实验证明了本章所提模型的有效性，针对其他类似情况的抗姿态干扰的人脸识别任务也可采用类似的解决方法。

# 总结与展望

## 全文总结

随着计算机视觉技术的发展，尤其是近两年人工智能、深度学习相关技术的推动，人脸识别技术取得了突飞猛进的进展，然而针对背景光照复杂，人脸姿态、表情以及年龄变化的情况，人脸识别技术依旧面临诸多挑战。相比于其他干扰情况，针对年龄干扰对人脸识别影响的研究还相对较少。然而在诸多应用场景中，如大量的登记识别问题、寻找丢失儿童、监控追捕逃犯等，实现对年龄变化鲁棒的人脸识别系统非常有必要。随着年龄的变化，人脸在形状、纹理等方面会发生巨大改变，这种改变与人的基因、生活习惯和环境等有着千丝万缕的联系，此外不同的年龄阶段特征的改变形态也不一致。如何去处理年龄变化带来的干扰是跨年龄人脸识别的关键。

本文以实现跨年龄的人脸识别为目标，紧紧围绕提取对年龄鲁棒的身份特征，不管人的年龄特征如何变化，人脸特征中与人身份相关的那部分是稳定不变的。然而直接采用通用的人脸识别方法提取的特征中不可避免的还包含年龄带来的干扰信息。本文提出了引入年龄估计任务来辅助人脸识别任务去选择对年龄变化鲁棒的身份特征的策略，学习身份字典和年龄字典将手工提取的人脸特征编码到两个独立的子空间，在只包含身份信息的子空间实现跨年龄的人脸识别。考虑到手工特征提取的局限，本文进一步采用端到端的深度卷积网络来提取人脸特征，年龄估计任务关注于人脸特征中随着年龄变化的那部分，人脸识别任务关注人脸特征中稳定不变的那部分，年龄估计的辅助任务在特征选择过程中辅助人脸识别任务去选择对年龄变化鲁棒的那部分特征，由此实现跨年龄的人脸识别。在通用的包含年龄信息的人脸数据库上验证了本文提出的两种方法的有效性。综上所述，本文的主要工作包括：

1. **非年龄相关的身份特征子空间学习算法**

传统的子空间学习算法大都基于原始图像的特征构建一个统一的优化目标，没有考虑图像特征的部分对最终识别的影响。本文将人脸图像特征看成由四部分组成：身份信息部分、年龄干扰部分、平均脸部分以及随机噪声部分，引入身份字典和年龄字典对人脸的身份特征和年龄特征进行编码，在字典的学习过程中同时考虑类别的约束，保证由这两个字典张成的空间对人脸的身份特征和年龄特征具有很好的区分度。年龄子空间包含了用于年龄估计的年龄相关的特征，由此将人脸特征中的年龄干扰信息从身份子空间中剥离开，保证身份特征对年龄变化的鲁棒性。本文从特征可视化和实验的角度证明了该方法的有效性。此外，该方法中引入的年龄字典也可替换为表情、姿态等其他干扰的字典，可拓展到类似的问题上。

1. **基于深度卷积网络的跨年龄的人脸识别算法**

考虑到手工特征提取的局限性，本文利用深度卷积网络自主学习对年龄变化鲁棒的身份特征，基于已有的通用人脸数据库训练得到的模型底座进行优化，克服跨年龄的人脸图像数据集不足的困境。创新性地从特征图可视化、神经元能量分布等角度指导基于深度卷积网络的多任务学习的网络结构设计。引入的年龄估计任务在低层级与人脸识别任务共享人脸边缘、纹理的特征提取过程，在特征选择过程中辅助人脸识别任务去选择对年龄变化鲁棒的那部分特征，保证身份识别任务能够提取到干净的身份特征，此外，在多任务学习的目标函数中引入了Center Loss约束，充分保证身份特征的判别能力，由此实现了跨年龄的人脸识别。

1. **基于多任务学习的跨年龄的人脸识别框架**

本文分别从传统的子空间学习方法和基于深度学习方法的角度探索跨年龄的人脸识别算法。在两个方法中均采用了多任务学习的方式，引入的年龄估计任务提取年龄敏感的特征，辅助身份识别任务获取对年龄变化鲁棒的人脸特征。基于这两种方法提取得到的身份特征，本文采用计算特征Cosine距离和KNN分类器获得最终人脸识别的结果，在MORPH和FGNET数据集上证明了本文基于多任务学习框架的跨年龄的人脸识别的有效性。相比于单任务学习，引入的年龄辅助任务能够提高跨年龄的人脸识别准确率，同时模型的泛化能力也得到了提高。

## 研究展望

跨年龄的人脸识别是一个极富挑战的问题，对于该问题的研究不仅能够让人脸识别技术更广泛的应用于人们的日常生活中，而且对人脸识别技术的发展具有重要的推动作用。

本文从提取鲁棒的身份特征角度出发，提出了非年龄相关的身份子空间算法将人脸特征编码到两个独立的子空间，提出了基于多任务的深度卷积网络算法，利用年龄估计任务辅助人脸识别任务选择对年龄鲁棒的身份特征。虽然这些方法取得了一定的效果，但是跨年龄的人脸识别研究无论是从思想还是方法上都还有很多可以研究的地方，结合自身的研究过程，我认为后续的研究可以从以下几个方面展开：

**1）基于对抗生成网络的跨年龄的人脸识别**

本文主要是从判别模型的角度实现的跨年龄的人脸识别。起初研究跨年龄的人脸识别问题的学者多采用生成模型的角度，但是传统的生成模型在刻画人脸老化过程中往往参数量庞大，计算复杂，生成的人脸老化模型很容易失真，导致后续的诸多研究转向了判别模型的角度。随着对抗生成网络的提出[133]，准确、逼真的人脸老化模型的构建方法有了新的突破点，跨年龄的人脸识别可以利用该方法构建准确的人脸老化模型，将待比对的人脸图像转化到对应的年龄阶段，在此基础上在进行识别，可有效地克服年龄变化带来的影响。

**2）基于迁移学习的跨年龄的人脸识别**

跨年龄的人脸识别研究相对较少一个很大的原因就是跨年龄的人脸数据库相对较少。而基于深度卷积网络的人脸识别方法需要大量的数据来学习样本的特征，迁移学习能够一定程度上解决这个问题。不同种族、不同地区的人脸数据集在样本的分布上存在不同，但是人脸的老化过程则是相似的，如果能够利用迁移学习方法，则可以获得通用的跨年龄的人脸识别模型，一定程度上能够克服跨年龄的人脸图像数据量不足的问题。

**3）基于多任务学习的人脸识别**

本文已经从多任务学习的角度进行了尝试，越来越多的文献也在关注针对人脸检测、特征点定位、人脸属性分类等问题上多任务学习的应用。相比于传统的多任务学习中正则项约束的使用，基于深度学习的多任务学习并没有相对统一的方法去刻画任务之间的关系，对于任务间特征的关系的探索也相对较少。人脸特征除了刻画一个人的身份信息中，还包括性别、年龄等信息，如果能够将这些辅助任务特征的学习与人脸识别构建到统一的学习框架中，则将进一步提高人脸识别的鲁棒性。

1. **提取更加鲁棒的身份特征**

传统的特征提取方式大都是提取人脸图像较低层级的纹理信息，虽然深度学习能够根据任务的不同从大量的数据中提取到高层级的抽象语义特征，但是如果数据量有限或者任务的目标函数刻画不精准，则提取的特征可能无法满足任务完成的要求，此外，人脸图像采集的环境多变，不可控的干扰也很多，因此更加鲁棒的身份特征提取依然是值得研究的问题。

参 考 文 献

[1] 孙冬梅, 裘正定. 生物特征识别技术综述[J]. 电子学报, 2001, 29(s1):1744-1748.

[2] Park U, Tong Y, Jain A K. Age-invariant face recognition [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2010, 32(5):947-954.

[3] Cootes T F, Taylor C J. Constrained active appearance models[C]. IEEE International Conference on Computer Vision, 2001:748-754 vol.1.

[4] Gong D, Li Z, Lin D, et al. Hidden Factor Analysis for Age Invariant Face Recognition [C]. IEEE International Conference on Computer Vision, 2013:2872-2879.

[5] Gong D, Li Z, Tao D, et al. A maximum entropy feature descriptor for age invariant face recognition [C].IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2015:5289-5297.

[6] Terrillon J C, Fukamachi H, Akamatsu S, et al. Comparative Performance of Different Skin Chrominance Models and Chrominance Spaces for the Automatic Detection of Human Faces in Color Images[C]. IEEE International Conference on Automatic Face and Gesture Recognition, 2000:54.

[7] Jones M J, Rehg J M. Statistical color models with application to skin detection [J]. International Journal of Computer Vision, 2002, 46(1):81–96.

[8] Abdel-Mottaleb M, Elgammal A. Face detection in complex environments from color images[C]. IEEE International Conference on Image Processing, 1999, 3: 622-626.

[9] Govindaraju V, Srihari S N, Sher D B. A computational model for face location[C]. IEEE International Conference on Computer Vision, 1990: 718-721.

[10] Yang G, Huang T S. Human face detection in a complex background [J]. Pattern Recognition, 1994, 27(1):53–63.

[11] Viola P, Jones M. Rapid Object Detection using a Boosted Cascade of Simple Features[C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003:I-511-I-518 vol.1.

[12] Guyon I, Weston J, Barnhill S, et al. Gene Selection for Cancer Classification using Support Vector Machines[J]. Machine Learning, 2002, 46(1):389–422.

[13] Cootes T F, Taylor C J. Constrained active appearance models[C]. IEEE International Conference on Computer Vision, 2001:748–754 vol.1.

[14] Cootes T F, Taylor C J. Active Shape Model Search using Local Grey-Level Models: A Quantitative Evaluation[C].BMVC. 1993, 93: 639-648.

[15] Belhumeur P N, Jacobs D W, Kriegman D J, et al. Localizing parts of faces using a consensus of exemplars[C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2011:545– 552.

[16] Cao X, Wei Y, Wen F, et al. Face alignment by Explicit Shape Regression[C].IEEE Conference on Computer Vision and Pattern Recognition, 2012:2887–2894.

[17] Sun Y, Wang X, Tang X. Deep convolutional network cascade for facial point detection[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2013: 3476-3483.

[18] Turk M, Pentland A. Eigenfaces for Recognition[J]. Journal of Cognitive Neuroscience, 1991, 3(1):71-86.

[19] Kamran Etemad, Rama Chellappa. Discriminant analysis for recognition of human face images[J]. Journal of the Optical Society of America A, 1997, 14(14):125-142.

[20] Gordon G G. Face Recognition from Frontal and Profile Views[J]. Proceedings of the International Workshop on Automatic Face- and Gesture-Recognition, IWAFGR 95, 1995:47--52.

[21] Ojala T, Pietikainen M, Harwood D. Performance evaluation of texture measures with classification based on Kullback discrimination of distributions[C]. International Conference on Pattern Recognition, 1994. Vol. 1.

[22] Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection[C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005:886-893.

[23] Jolliffe I T. Principal Component Analysis[J]. Springer Berlin, 1986, 87(100):41–64.

[24] Welling M. Fisher Linear Discriminant Analysis[J]. Department of Computer Science, 2007, 16(94):237–280.

[25] Weinberger K Q, Blitzer J, Saul L K. Distance metric learning for large margin nearest neighbor classification[C].Advances in neural information processing systems. 2006: 1473-1480.

[26] Keller J M, Gray M R, Givens J A. A fuzzy K-nearest neighbor algorithm[J]. IEEE Transactions on Systems Man & Cybernetics, 1985, SMC-15(4):580–585.

[27] Lawrence S, Giles C L, Tsoi A C, et al. Face recognition: a convolutional neural-network approach.[C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1996:217-222.

[28] Chan H, Bledsoe W W. A man-machine facial recognition system: some preliminary results[J]. Panoramic Research Inc., Palo Alto, CA, USA1965, 1965.

[29] Kaufman G J, Breeding K J. The Automatic Recognition of Human Faces from Profile Silhouettes[J]. IEEE Transactions on Systems Man & Cybernetics, 1976, 6(2):113-121.

[30] Takeo K. Picture processing by computer complex and recognition of human faces[D]. PhD thesis, Kyoto University, 1973.

[31] Harmon L D, Khan M K, Lasch R, et al. Machine identification of human faces[J]. Pattern Recognition, 1981, 13(2): 97-110.

[32] Brunelli R, Poggio T. Face recognition through geometrical features[C]. IEEE International Conference on Computer Vision, 1992: 792-800.

[33] Tan X, Triggs B. Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions[J]. IEEE Transactions on Image Processing, 2010, 19(6):1635.

[34] Guo Z, Zhang L, Zhang D, et al. Hierarchical multiscale LBP for face and palmprint recognition[C]. IEEE International Conference on Image Processing, 2010:4521-4524.

[35] Chen D, Cao X, Wen F, et al. Blessing of Dimensionality: High-Dimensional Feature and Its Efficient Compression for Face Verification[C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2013:3025-3032.

[36]Wright J, Yang A Y, Ganesh A, et al. Robust Face Recognition via Sparse Representation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2008, 31(2):210-227.

[37] Wright J, Ma Y, Mairal J, et al. Sparse Representation for Computer Vision and Pattern Recognition[J]. Proceedings of the IEEE, 2010, 98(6):1031-1044.

[38] Hubel D. H, Wiesel T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex[J]. The Journal of Physiology, 1962, 160(1):106.

[39]Lecun Y, Boser B, Denker J S, et al. Backpropagation Applied to Handwritten Zip Code Recognition[J]. Neural Computation, 1989, 1(4):541-551.

[40] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets[J]. Neural computation, 2006, 18(7): 1527-1554.

[41] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C].Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.

[42] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]. Proceedings of the IEEE conference on computer vision and pattern recognition.2015:1-9.

[43] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[44] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C].Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[45] Huang G, Liu Z, Weinberger K Q, et al. Densely connected convolutional networks[J]. arXiv preprint arXiv:1608.06993, 2016.

[46] Huang G B, Mattar M, Berg T, et al. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments[C].Workshop on faces in'Real-Life'Images: detection, alignment, and recognition. 2008.

[47] Taigman Y, Yang M, Ranzato M A, et al. Deepface: Closing the gap to human-level performance in face verification[C].Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 1701-1708.

[48] Sun Y, Wang X, Tang X. Deep learning face representation from predicting 10,000 classes[C].Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 1891-1898.

[49]Sun Y, Chen Y, Wang X, et al. Deep learning face representation by joint identification-verification[C].Advances in neural information processing systems. 2014: 1988-1996.

[50] Sun Y, Wang X, Tang X. Deeply learned face representations are sparse, selective, and robust[C].Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 2892-2900.

[51] Kumar N, Berg A C, Belhumeur P N, et al. Attribute and simile classifiers for face verification[C]. Proceedings of the IEEE Conference on Computer Vision, 2009: 365-372.

[52] Schroff F, Kalenichenko D, Philbin J. FaceNet: A unified embedding for face recognition and clustering[C]. .Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015:815-823.

[53]Sohn K. Improved deep metric learning with multi-class n-pair loss objective[C].Advances in Neural Information Processing Systems. 2016: 1857-1865.

[54] Wen Y, Zhang K, Li Z, et al. A discriminative feature learning approach for deep face recognition[C].European Conference on Computer Vision. Springer International Publishing, 2016: 499-515.

[55] Liu W, Wen Y, Yu Z, et al. SphereFace: Deep Hypersphere Embedding for Face Recognition[J]. arXiv preprint arXiv:1704.08063, 2017.

[56] Fu Y, Huang T S. Human age estimation with regression on discriminative aging manifold[J]. IEEE Transactions on Multimedia, 2008, 10(4): 578-584.

[57] Geng X, Zhou Z H, Smith-Miles K. Automatic age estimation based on facial aging patterns[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(12): 2234-2240.

[58] Guo G, Fu Y, Dyer C R, et al. Image-based human age estimation by manifold learning and locally adjusted robust regression[J]. IEEE Transactions on Image Processing, 2008, 17(7): 1178-1188.

[59] Guo G, Mu G, Fu Y, et al. Human age estimation using bio-inspired features[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.2009: 112-119.

[60] Young H Kwon, Lobo N D V. Age Classification from Facial Images[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1999:762-767.

[61] Lanitis A, Draganova C, Christodoulou C. Comparing different classifiers for automatic age estimation[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2004, 34(1): 621-628.

[62] Ramanathan N, Chellappa R. Face Verification Across Age Progression[C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005:462-469.

[63] Ramanathan N, Chellappa R. Face verification across age progression[J]. IEEE Transactions on Image Processing, 2006, 15(11): 3349-3361.

[64] Wang J, Shang Y, Su G, et al. Age simulation for face recognition[C]. IEEE International Conference on Pattern Recognition, 2006:913-916..

[65] Zhou S K, Georgescu B, Zhou X S, et al. Image based regression using boosting method[C]. IEEE International Conference on Computer Vision, 2005:541-548 Vol. 1.

[66] Lanitis A, Taylor C J, Cootes T F. Toward automatic simulation of aging effects on face images[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(4): 442-455.

[67] Suo J, Zhu S C, Shan S, et al. A Compositional and Dynamic Model for Face Aging[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2010, 32(3):385.

[68] Suo J, Chen X, Shan S, et al. Learning long term face aging patterns from partially dense aging databases[C]. IEEE International Conference on Computer Vision, 2009:622-629.

[69] Tsumura N, Ojima N, Sato K, et al. Image-based skin color and texture analysis/synthesis by extracting hemoglobin and melanin information in the skin[J]. ACM Transactions on Graphics (TOG), 2003, 22(3): 770-779.

[70] Lanitis A, Taylor C J, Cootes T F. Toward automatic simulation of aging effects on face images[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(4):442-455.

[71] Geng X, Zhou Z H, Smithmiles K. Automatic age estimation based on facial aging patterns [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(12):2234.

[72] Du J X, Zhai C M, Ye Y Q. Face aging simulation and recognition based on NMF algorithm with sparseness constraints [J]. Neurocomputing, 2013, 116(10):250-259.

[73]Chen B C, Chen C S, Hsu W H. Cross-Age Reference Coding for Age-Invariant Face Recognition and Retrieval [C]. European Conference on Computer Vision. Springer, Cham, 2014:768-783.

[74] Klare B, Jain A K. Face recognition across time lapse: On learning feature subspaces [C]. International Joint Conference on Biometrics. IEEE, 2011:1-8.

[75] Li Z, Park U, Jain A K. A Discriminative Model for Age Invariant Face Recognition [J]. IEEE Transactions on Information Forensics & Security, 2011, 6(3):1028-1037.

[76] Ling H, Soatto S, Ramanathan N, et al. Face Verification Across Age Progression Using Discriminative Methods [J]. IEEE Transactions on Image Processing, 2006, 15(11):3349.

[77] Otto C, Han H, Jain A. How does aging affect facial components? [C].IEEE International Conference on Computer Vision, 2012:189-198.

[78] Li Z, Gong D, Li X, et al. Aging Face Recognition: A Hierarchical Learning Model Based on Local Patterns Selection [J]. IEEE Transactions on Image Processing, 2016, 25(5):2146.

[79] Wen Y, Li Z, Qiao Y. Latent Factor Guided Convolutional Neural Networks for Age-Invariant Face Recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016:4893-4901.

[80] Zheng T, Deng W, Hu J. Age Estimation Guided Convolutional Neural Network for Age-Invariant Face Recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017:503-511.

[81] Argyriou A, Evgeniou T, Pontil M. Convex multi-task feature learning[J]. Machine Learning, 2008, 73(3): 243-272.

[82] Gong P, Zhou J, Fan W, et al. Efficient multi-task feature learning with calibration[C]. International conference on Knowledge discovery and data mining, 2014: 761-770.

[83] Tian Y, Luo P, Wang X, et al. Pedestrian detection aided by deep learning semantic tasks[C].Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 5079-5087.

[84] Zhang Z, Luo P, Chen C L, et al. Facial Landmark Detection by Deep Multi-task Learning[C]. European Conference on Computer Vision. 2014:94-108.

[85] Ruder S. An overview of multi-task learning in deep neural networks[J]. arXiv preprint arXiv:1706.05098, 2017.

[86] Caruana R. Multitask learning[M].Learning to learn. Springer US, 1998: 95-133.

[87] Obozinski G, Taskar B, Jordan M. Multi-task feature selection[J]. Statistics Department, UC Berkeley, Tech. Rep, 2006, 2.

[88] Lin K, Xu J, Baytas I M, et al. Multi-Task Feature Interaction Learning[C]. International Conference on Knowledge Discovery and Data Mining, 2016:1735-1744.

[89] Du L, Ling H. Cross-age face verification by coordinating with cross-face age verification[C]. IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2015:2329-2338.

[90] Chen D, Ren S, Wei Y, et al. Joint cascade face detection and alignment[C].European Conference on Computer Vision. Springer, Cham, 2014: 109-122.

[91] Zhang C, Zhang Z. Improving multiview face detection with multi-task deep convolutional neural networks[C]. Applications of Computer Vision. IEEE, 2014:1036-1041.

[92] Yim J, Jung H, Yoo B I, et al. Rotating your face using multi-task deep neural network[C].Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 676-684.

[93] Baxter J. A Bayesian/Information Theoretic Model of Learning to Learn via Multiple Task Sampling[J]. Machine Learning, 1997, 28(1):7-39.

[94] Zafeiriou S, Zhang C, Zhang Z. A survey on face detection in the wild: Past, present and future [J]. Computer Vision & Image Understanding, 2015, 138:1–24.

[95] Wu B, Ai H, Huang C, et al. Fast rotation invariant multi-view face detection based on real adaboost[C].IEEE International Conference on Automatic Face and Gesture Recognition, 2004:79–84.

[96] Xiong X, Torre F D L. Supervised Descent Method and Its Applications to Face Alignment[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013:532–539.

[97] Yang M H, Kriegman D J, Ahuja N. Detecting Faces in Images: A Survey[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2002, 24(1):34–58.

[98] Viola P, Jones M. Rapid Object Detection using a Boosted Cascade of Simple Features[C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003:I-511-I-518 vol.1.

[99] Zhang C, Zhang Z. Improving multiview face detection with multi-task deep convolutional neural networks[C]. IEEE Winter Conference on Applications of Computer Vision (WACV), 2014: 1036-1041

[100] Lienhart R, Maydt J. An extended set of Haar-like features for rapid object detection[C]. IEEE International Conference on Image Processing, 2002:I-900-I-903 vol.1.

[101] Meir R, Rätsch G. An introduction to boosting and leveraging[M].Advanced lectures on machine learning, 2003: 118-183.

[102] Girshick R. Fast R-CNN[C]. IEEE International Conference on Computer Vision, 2015:1440-1448.

[103] Ren S, Girshick R, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137.

[104] Li H, Lin Z, Shen X, et al. A convolutional neural network cascade for face detection[C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2015:5325-5334.

[105] Amberg B, Vetter T. Optimal landmark detection using shape models and branch and bound[C]. IEEE International Conference on Computer Vision, 2011: 455-462.

[106] Belhumeur P N, Jacobs D W, Kriegman D J, et al. Localizing parts of faces using a consensus of exemplars[J]. IEEE transactions on pattern analysis and machine intelligence, 2013, 35(12): 2930-2940.

[107] Gu L, Kanade T. A generative shape regularization model for robust face alignment[J]. European Conference on Computer Vision, 2008: 413-426.

[108] Liang L, Xiao R, Wen F, et al. Face alignment via component-based discriminative search[J]. European Conference on Computer Vision, 2008: 72-85.

[109] Zhu X, Ramanan D. Face detection, pose estimation, and landmark localization in the wild[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2012: 2879-2886.

[110] Zhang Z, Zhang W, Liu J, et al. Facial landmark localization based on hierarchical pose regression with cascaded random ferns[C]. International Conference on Multimedia, 2013:561-564.

[111] Yang H, Patras I. Sieving regression forest votes for facial feature detection in the wild[C].Proceedings of the IEEE International Conference on Computer Vision. 2013: 1936-1943.

[112 ] Cootes T F, Ionita M C, Lindner C, et al. Robust and Accurate Shape Model Fitting Using Random Forest Regression Voting[C]. European Conference on Computer Vision, 2012:278-291.

[113] Kazemi V, Sullivan J. One millisecond face alignment with an ensemble of regression trees[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2014:1867–1874.

[114] Zhou E, Fan H, Cao Z, et al. Extensive Facial Landmark Localization with Coarse-to-Fine Convolutional Network Cascade[C].IEEE International Conference on Computer Vision Workshops, 2014:386-391.

[115] Zhang J, Shan S, Kan M, et al. Coarse-to-Fine Auto-Encoder Networks (CFAN) for Real-Time Face Alignment[C]. European Conference on Computer Vision, 2014:1-16.

[116] Zhang L, Zhu P, Hu Q, et al. A linear subspace learning approach via sparse coding [C]. IEEE International Conference on Computer Vision, 2012:755-761.

[117] Wright J, Ma Y, Mairal J, et al. Sparse Representation for Computer Vision and Pattern Recognition [J]. Proceedings of the IEEE, 2010, 98(6):1031-1044.

[118] Aharon M, Elad M, Bruckstein A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation [J]. IEEE Transactions on signal processing, 2006, 54(11): 4311-4322.

[119] Liu H D, Yang M, Gao Y, et al. Bilinear discriminative dictionary learning for face recognition [J]. Pattern Recognition, 2014, 47(5):1835-1845.

[120] Sun Y, Liu Q, Tang J, et al. Learning discriminative dictionary for group sparse representation [J]. IEEE Transactions on Image Processing, 2014, 23(9):3816-3828.

[121] Zhao Y, Su J. A structured sparse learning approach for efficient facial feature description [C]. IEEE International Conference on Information and Automation, 2013:870-874.

[122] Jiang X, Lai J. Sparse and Dense Hybrid Representation via Dictionary Decomposition for Face Recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(5):1067-79.

[123] Zhang L, Yang M, Feng X. Sparse representation or collaborative representation: Which helps face recognition? [C]. International Conference on Computer Vision, 2011:471-478.

[124] Jiang Z, Lin Z, Davis L S. Label Consistent K-SVD: Learning A Discriminative Dictionary for Recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(11):2651-64.

[125] Yang M, Zhang L, Yang J, et al. Metaface learning for sparse representation based face recognition [C]. IEEE International Conference on Image Processing, 2013:1601-1604.

[126] Wright J, Yang A Y, Ganesh A, et al. Robust face recognition via sparse representation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(2):210.

[127] Jr K R, Tesafaye T. MORPH: A Longitudinal Image Database of Normal Adult Age-Progression [C]. IEEE International Conference on Automatic Face and Gesture Recognition, 2006:341-345.

[128] FG-NET Aging Database [Online]. Available: [http:.www.fgnet.rsunit.com/](http://www.fgnet.rsunit.com/)

[129] Li Z, Gong D, Li X, et al. Learning compact feature descriptor and adaptive matching framework for face recognition [J]. IEEE Transactions on Image Processing, 2015, 24(9):2736-45.

[130] Abdulnabi A H, Wang G, Lu J, et al. Multi-Task CNN Model for Attribute Prediction[J]. IEEE Transactions on Multimedia, 2016, 17(11):1949-1959.

[131] Wu Y, Ji Q. Constrained Joint Cascade Regression Framework for Simultaneous Facial Action Unit Recognition and Facial Landmark Detection[C]. Computer Vision and Pattern Recognition, 2016:3400-3408.

[132] Yin X, Liu X. Multi-Task Convolutional Neural Network for Pose-Invariant Face Recognition[J]. IEEE Transactions on Image Processing, 2018,27(2):964-975.

[133] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]. International Conference on Neural Information Processing Systems, 2014:2672-2680.

致 谢

时光飞快，转眼我的研究生生涯也即将步入尾声，回首这两年多经历的点点滴滴，心中感慨万千。在这期间，我一点点感受自己的成长和蜕变，而这一切都离不开一直在给予我指导、鼓励、关心和支持的老师、同学和家人。在此，我要向他们表达最衷心的感谢。

首先要感谢我的导师苏剑波教授对我的教育和指导，是苏老师引领我走进学术研究的殿堂。在研究方向的把握上，苏老师用渊博的知识、深厚的学术积淀和敏锐的学术洞察力指导我去做“有学术品味的”的研究，在每一次的工作汇报上，苏老师用严谨的态度和务实的做法教育我认真完成每一次的“出手”，对我的研究工作和论文写作给予了很全面的指导。苏老师是一位学术眼光独到，思想深邃且师德高尚的老师，从他那里，我不仅学到了科研学术的知识，更学到了很多做人做事的道理，这对我以后的工作和生活是一笔宝贵的财富。在此，我要向苏老师表示深深的感谢和祝福！

其次，要感谢RCIR实验室可爱的师兄师姐师弟师妹们，每天一起科研、交流学习的日子是非常充实快乐的。感谢实验室的马哲、陈叶飞、黄瑶、向国菲、项祯帧、吴琛、倪辉、张瑞、胡永盛等师兄师姐对我科研和生活上的指导和鼓励，尤其感谢陈叶飞师兄，每一次遇到学术上的困难和问题时，都耐心的给我解决的思路和建议，帮助我度过难关。感谢朱铭德、夏凌翔两位同僚的支持和陪伴，感谢蒋晨之、魏旭宏、陈若冰、俞经睿、戴毅、王岚、白栋林等师弟师妹们对我工作和生活的支持和鼓励。RCIR是一个团结有爱的集体，不管是科研上还是生活上，我都收获很多！

还要感谢上海交通大学B1503291班的全体同学，感谢能和你们一起步入交大，一起成长，一起奋斗，一起毕业。感谢我的室友李晶玲、李听枫、李雨慧，谢谢你们给予我生活上的照顾和包容。感谢自动化系的老师们，谢谢你们给我学习和生活上的帮助和支持。

最后，感谢我的父母及家人，你们一直是我坚强的后盾，我会带着您们的鼓励和期许继续努力，勇往直前！

攻读硕士学位期间已发表或录用的论文

[1] 吴长虹, 苏剑波, 陈叶飞. 抗年龄干扰的人脸识别. 电子学报（已录用）

[2] Changhong Wu, Jianbo Su, “A unified framework for age invariant face recognition and age estimation”, Proceedings of 2017 Chinese Intelligent Automation Conference(CIAC),2017, pp. 623-630. (EI检索)

攻读学位期间参与的项目

[1] 国家自然科学基金重点项目 “泛在信息制造环境下的机器人群智视觉交互与优化控制”， 2015.8-至今

[2] 上汽集团横向项目“汽车驾驶员疲劳检测”， 2016.6-2017.6