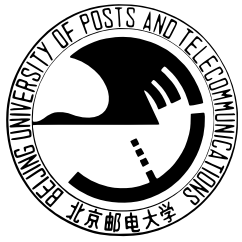
密级：　　保密期限：

北京郵電大學

硕士学位论文



题目：基于对抗神经网络的人脸图片属性识别与生成

学　　号：2015140024

姓　　名：于志鹏

专　　业：电子与通信工程

导　　师：董远

学　　院：信息与通信工程学院

二〇一七年十一月三十日

独创性（或创新性）声明

本人声明所呈交的论文是本人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢中所罗列的内容以外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得北京邮电大学或其他教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

申请学位论文与资料若有不实之处，本人承担一切相关责任。

本人签名：_____ 日期：_____

关于论文使用授权的说明

本人完全了解并同意北京邮电大学有关保留、使用学位论文的规定，即：北京邮电大学拥有以下关于学位论文的无偿使用权，具体包括：学校有权保留并向国家有关部门或机构送交论文，有权允许学位论文被查阅和借阅；学校可以公布学位论文的全部或部分内容，有权允许采用影印、缩印或其它复制手段保存、汇编学位论文，将学位论文的全部或部分内容编入有关数据库进行检索。（保密的学位论文在解密后遵守此规定）

本学位论文不属于保密范围，适用本授权书。

本人签名：_____ 日期：_____

导师签名：_____ 日期：_____

基于对抗神经网络的人脸图片属性识别与生成

摘 要

在模式识别与多媒体搜索领域，神经网络近五年新兴技术，但是凭借着简洁、有效、易训练等优势迅速在图像处理领域得到了广泛的应用。尤其是在人脸相关的领域，卷积神经网络的出现极大提升了人脸识别的准确率，并且尤其而上增加了人们对于探究人脸属性的期待，甚至想象。传统的人脸属性识别，往往采用对于人脸特征进行提取，在特征工程上进行分类器选择和调整。即使在神经网络飞速发展的前提之下，人们依然非常执着和热衷于将不同的神经网络结构下的人脸特征进行提取，然后按照特征工程的方式构建属性识别系统。然而随着端到端学习的兴起，人们对于整合识别的流程经过有了新的认识，简化任务流程设定合理的调整策略，让原本基于分治算法的子任务问题解决模态和神经网络较强的学习能力相互结合，尽量能够直接给出具体识别任务的答案，是目前领先的思想和发展方向。

另一方面，作为神经网络另类演变，对抗生成网络初期是为了探究神经神经网络的内部构造原理。但是随着对抗生成网络的不断进化，神经网络在图像重建领域开始有了出色的表现。例如超像素、图像风格，数据合成等。但是随着图片分布理念的兴起，迁移学习的理念作为非监督学习与监督学习之间的过渡，具有较高的使用价值和较低的转换成本。而对抗生成网络对于图像分布的良好表现能力，在迁移学习上的应用也很值得被探索。

本文主要研究两个方面的问题，一方面是调整网络的结构，结合端到端的设计思想和优化方式，尝试对于人脸图片的属性识别准确性提升。另一方面结合对抗生成网络对于人脸图片进行生成，不断优化合成数据的真实度和广泛性，能够结合迁移学习的思想提升机器对于人脸属性更好的理解能力。具体的贡献工作包括：设计带自控提升的智能学习策略，

使用多机多卡的训练策略加速训练的过程，使用指令集和超线程优化前馈过程快速生成图片，调整人脸图片的生成网络增强人脸图片的合成效果，通过使用合成数据弥补不同场景数据对于学习支持的不足，进而在不同场景下人脸属性的识别准确率。

关键词： 对抗生成网络 人脸属性 迁移学习 多机多卡 前馈优化

GENERATIVE ADVERSARIAL NETWORK BASED FACE ATTRIBUTE RECOGNITION AND REGENERATION

ABSTRACT

In the field of pattern recognition and multimedia search, the neural network has been emerging for nearly five years, but it has been widely applied in the field of image processing due to its advantages of concise, effective and easy training. Especially in the face-related fields, the appearance of convolutional neural network greatly enhances the accuracy of face recognition, and in particular, it increases people's expectations for exploring face attributes, and even imagines them. Traditional face recognition, face feature extraction is often used in the feature engineering classifier selection and adjustment. Even under the premise of the rapid development of neural network, people are still very persistent and keen to extract facial features under different neural network structures, and then construct the attribute recognition system according to the way of feature engineering. However, with the rise of end-to-end learning, people have a new understanding of the process of integration and recognition, simplify the task flow to set a reasonable adjustment strategy, so that the original sub-task based on divide-and-conquer algorithm to solve the problem of modal and neural network Strong learning ability combined with each other, as far as possible to give specific answers to the identification task, is the leading thinking and direction of development.

On the other hand, as an alternative to neural networks, the early days of confrontation generation networks were to explore the internal structural principles of neural networks. However, with the evolution of adversarial networks, neural networks have begun to perform well in the field of image reconstruction. Such as super-pixel, image style, data synthesis and so on. However, with the rise of the concept of image distribution, the concept of relocated learning as a transition between unsupervised learning and supervised learning has high

value of use and low conversion cost. However, the ability of confrontation-generating networks to perform well in image distribution is also worth exploring in migration learning.

This paper mainly studies two aspects of the problem. On the one hand, it adjusts the structure of the network, and combines the end-to-end design and optimization methods to try to improve the accuracy of attribute recognition of face images. On the other hand, combining anti-generated network to generate face images and continuously optimizing the authenticity and universality of the synthesized data, it is possible to improve the machine's ability of comprehension of face attributes in combination with the idea of migration learning. Specific contributions include designing intelligent learning strategies with self-controlled ascension, accelerating the training process using multi-machine and multi-card training strategies, using instruction sets and hyperthreading to optimize the feedforward process to quickly generate pictures, and adjusting the generation of face pictures to enhance the network The synthetic effect of face pictures can make up for the lack of learning support for different scene data by using synthetic data and the recognition accuracy of face attributes in different scenes.

KEY WORDS: GAN face attribue transfer leearning Multi-machine multi-card Feedforward optimization

目 录

表格索引

插图索引

符号对照表

$(\cdot)^*$	复共轭
$(\cdot)^{\mathrm{T}}$	矩阵转置
$(\cdot)^{\mathrm{H}}$	矩阵共轭转置
$\mathbf{X}$	矩阵或向量
$\mathscr{A}$	集合
$\mathscr{A} \times \mathscr{B}$	集合 $\mathscr{A}$ 与集合 $\mathscr{B}$ 的 Cartesian 积，即 $\mathscr{A} \times \mathscr{B} = \{(a,b) : a \in \mathscr{A}, b \in \mathscr{B}\}$

第一章 绪论

1.1 课题研究的背景与意义

自从人类第一次开始在石板上有意识的划刻开始，图像这一最早的原始信息传递媒介就开始以各种各样的形式在人类的信息传递过程中发挥着非彼寻常的作用，正所谓“一图胜千言”,“耳听为虚，眼见为实”说明的正是这个道理。随着图像的表现形式不断的发展和图像数据的日益增加，如何提取图像中所包含的海量信息以及信息的分析使用成了现代多媒体，人工智能，自动化控制等多个领域都亟需解决的问题。人脸识别是图像信息中中具有身份信息生物特征的一部分，可以广泛的用在安防，娱乐，多媒体等领域。而如果说人脸识别是识人，辩人。那么人脸属性识别就可以说是“相面算命”了，比如说人机交互中的表情识别互动，又比如说视频播放网站中限制级视频对于低龄观众限制，尤其在用户数据统计的过程中一张人脸图片，就可以识别出用户性别，年龄，是否戴眼镜，基本面部特征，发型状态等信息，从而可以进行个性化的业务分析与定制推荐，这在逐渐强调个性化发展的社会中具有很高的市场。近些年来图像领域人工智能迅猛发展，神经网络技术因为简单，高效，对于数据适应性强等特性，在各种图像识别的领域大规模训练使用，取得了非常好的效果。人脸识别受相关技术的影响，也有了很大的进步，从慢慢接近人眼的便是效果，到不断超越，以至于后来的百万级人脸搜索 99% 的准确率，可以说正在慢慢朝着实用化的技术发展。但是人脸属性作为另一项研究领域，却总是不温不火，无论是准确率还是实际使用都有一定的发展空间。其中主要的问题在于网络结构上对于人脸属性多样性兼容问题，以及人脸属性任务对于人脸图片数据要求的复杂和严格性，人脸属性种类繁多，且人脸场景分布极为复杂，标注工作难度较大，且歧义性较大。因此，本文旨在在深度学习对于图像识别任务有较大推动的今天，研究网络结构数据分布对于属性识别的影响。具体分为属性识别的网络结构探索和不同数据分布对于属性数据的提高效果。

1.2 国内外研究现状

从时间角度来看，基于人脸图像的多种人脸属性预测估计在上世纪 90 年代就开始，1990 年，MIT 的 Cottrell 和 Metcalfe 把基于 Auto-Encoder 的特征降维用于性别

和表情识别；1999 年，塞浦路斯学院的 Lanitis 构建了 FGNET 年龄估计数据库（共 82 人，1002 张图像），当时用 PCA 做特征提取；2006 年，北卡的 Ricanek 和 Tesafaye 构建了首个大规模年龄、性别、种族数据库 MORPH(1.3 万人，5.5 万图像)；2008 年，哥大的 Kumar 等人构建了包含 10 个属性（后在期刊文章里扩展到 60 多个）的大规模名人数据库 PubFig（共 200 人，6 万张图像）仅部分公开，提取了手工设计特征，之后对每个属性训练 SVM;2010 年，MIT 的 Pho 等人首次研究了基于普通摄像头的非接触式心率估计，这是“由表及里”的一次突破；2015 年，中科院计算所 VIPL 研究组首次研究了人与机器在属性识别上的性能差异（可控），并发现机器在年龄、性别和种族的识别上已经可以超过人类；NIST 组织了年龄和性别预测方面的评测竞赛，并且出了一个报告概括了领域相关工作；此外，香港中文大学汤老师组构建了大规模互联网名人的 40 个属性数据集 celeA.（20 万图像）由此可见，研究工作的时间跨越度并非很大，但是，各方面工作的丰富性和多样性还是令人瞩目的。从特征的表示方法来看：包括从全局表观特征（Intensity，PCA，Gabor，LBP，BIF 生物启发式特征），到细节特征（纹理，肤色，人脸形状）以及近年来将尝试用的一些深度特征如 CNN,DNN 特征。其中是一个不断演变但是也时有结合的过程。而从特征分类方法上来看：研究的任务形式也从单任务学习（常用方法：每个属性训练一个分类器）慢慢演变到多标签学习（回归目标不仅是数，而是向量形式）而后根据不同的细粒度额精确化需求，发展出层级式的分类器（由粗到细，特别适用于年龄分类，如先确定年龄范围，再进行具体年龄分类）和多任务学习（多任务限制玻尔兹曼机，多任务 CNN 等等）总结来看，是一个从手工设计特征到深度特征、从组合式的学习到端到端学习、从 STL 到 MTL（从单任务学习到多任务学习）的发展过程。

人脸视觉属性学习并不简单，特别是在非可控的真实场景下。影响因素有以下几个方面：传感环境（尤其在室外）的不可控性以及人物的不配合性，这会引起姿态、光照、遮挡等多种因素的影响；属性之间的相关性以及差异性；属性数量的增多引起内存消耗的增加，因此需高效的模型

1.3 本文的工作与贡献

1.3.1 研究内容

在本文的主要研究的内容有两个，第一项是结合人脸属性的性质探究在人脸属性在深度学习技术下的表现。其中包括人脸属性数据的总结与整理，规划人脸属性的标注类型，单任务模型下的人脸属性的表现，多任务模式下人脸属性的表现等，主

要的衡量指标是在不同模型组合和模型策略的情况下人脸属性模型的准确率。另一方面，是针对于现实环境中图片采集的不可控制性，使用对抗生成网络来模拟不同场景的人脸数据，并且探究如何使用迁移学习的思想来对提高人脸属性对于不同场景泛化能力。在这一任务中，除了最终对于人脸属性的准确率提升之外，人脸图片的生成质量也是衡量得指标之一。

1.3.2 主要贡献

在这项研究工作之中主要贡献包括研究内容上的工作和一定的工程优化工作，具体如下：研究上的工作：基于 Alexnet、残差网络和 inception 结构对于人脸属性的单任务和多任务网络进行设计。设计具有网络输出置信评估的模块，或者说网络结构，用于网络对于自身的输出的把控，以及提高实际使用的方便和准确性。设计相应的对抗生成网络，用来对于不同场景的人脸及逆行生成和模拟使用超像素网络的思想来提升对抗生成网络的图像生成质量。使用生成的网络图像的输出图片提升网络在位置数据上的人脸属性准确性。工程上的工作：设计多机多卡的训练，来提升模型的训练速度在前馈的过程之中，使用多线程、指令集等优化方式，提升模型输出的速度，和图片生成的速度。

1.3.3 论文的组织结构

第二章：笔者主要介绍涉及人脸属性在深度学习技术种一些基本常识和常见的操作和笔者对相关瓶颈操作的一些优化。具体包括：在卷积神经网络的基础操作中介绍所谓卷积操作的多种实现和使用方式介绍，激活函数的具体使用，常见的网络参数初始化方法和网络训练相关细节。在多机多卡的部分介绍，在多卡训练中数据的同步和分发方式，模型参数的更新策略，多机训练中需要注意的一些关键选项配置等。在网络前馈的优化部分会介绍一些实用性非常强的快速卷积算法，对于网络中常见操作的一些合从而提升速度，以及一点计算机图计算理论种基于静态图的速度优化方式和内存节省技巧。

第三章笔者主要介绍针对于在人脸属性所进行的一些实验和创新的过程的一些相关工作：常见的人脸属性数据集：包括 FG-NET，MOROH II，CelebA，LFWA，ChaLearn LAP and FotW 等单任务模型下人脸属性的实验过程和结果。多任务模型下人脸属性的实验过程和结果。人脸属性中网络能力自评模块的设计的实验过程和结果。

第四章笔者会介绍如何使用对抗生成网络对于不同场景下的人脸进行学习并且根据噪声生成人脸图片。使用超像素的方式对于人脸图片进行一定程度上的效果增强和场景迁移。通过结合迁移之后的人脸图像进行学习可以方便的改进人脸属性中由于数据分布不同导致准确率下降情况

第五章笔者主要对实验过程做一个综合性的概述并且自我评价一下整个实验过程中出现的问题和解决问题的方法。回顾在解决问题种反应的一些现实层面的现象以及个人对这些现象出现的原因和结果的思考。当然也包含一点关于未来和未解决工作的思索。

第二章 卷积神经网络的相关技术介绍

2.1 卷积神经网络的基础操作和训练

2.2 多机多卡策略对于网络训练的提升

2.3 网络前馈速度的优化

第三章 人脸多属性属性识别的架构

在具体介绍人脸属性的任务过程中，首先对于对于人脸属性的一些常见问题做简单的介绍：人脸属性识别的输入一般为具体的 RGB 图片，同时至少带有人脸检测输出的人脸框以及用于人脸矫正的 landmark，实际实验证明，经过矫正的人脸对于和人脸姿势无关的属性具有很好的提升。简单介绍一下人脸矫正的过程：人脸矫正顾名思义：就是将不够“端正”人脸调整到标准的大小，位置和姿态，这样可以让人脸都在同样的环境下进行比较，人的面部姿态一般会从 roll(平面旋转), pitch(左右侧脸) 和 yaw(抬头低头) 三个维度来描述。平面旋转很容易处理，只需将图片旋转一个角度调整至水平即可。而侧脸和低头处理起来比较有挑战，但通过放射变化也可以较好的解决。经过人脸矫正之后，不同的算法和模型其实是对人脸矫正之后的图片或者说一定 3*H*W 维数值分布在 (0-255) 的向量空间进行各种线性和非线性计算，最后输出图片对应的属性分类标签的过程。

（加一个简单的流程图）

下面分别从人脸数据库，人脸属性的单模型预测、人脸属性的多任务预测几个角度对个人研究相关的人脸属性识别领域进行一下介绍，并且对于自身的创新工作和相关任务进行说明。

3.1 人脸属性数据库的分析与简介

人脸属性的数据库是根据人脸的标签进行标注和构建的的，其中标签往往具有很多种表示和性质，具体包括：有序性与无序性：无序性：无序性的属性有两个或两个以上的类别（值），但在类别之间没有内在的顺序。例如，种族是具有多个类别的名义属性，例如黑色，白色，亚洲等，并且这些值（类别）没有内在排序。有序性：有序性的属性具有明确的变量排序。例如，一个人的年龄，通常从 0 到 100，是不平均的。（实际上，年龄不仅是相互独立的存在，在不同的年龄标签中，具有一定钟形的分布）整体性与局部性。整体性：整体性标签描述了整个人脸的特征，诸如年龄，性别，种族等局部性：和整体性标签相反，局部性描述了部分人脸的特征，例如：尖鼻子，大嘴唇等。类似的对于人脸属性的分析还有很多，比如限制性场景和非限制性场景（如固定摄像头拍摄和日常采集的场景），相对的性标注和绝对属性标

注（如颜值数据标注之间只有相互的高低，但没有绝对的属性标签）本文中也主要根据上面的人脸属性的性质来设计网络和分析问题。在早期，数据集通常只有一种属性的标注，比如之前提到的 FG-NET，它包含 82 个目标的 1002 张图像，只有年龄属性。近来，很多人脸数据集都有多属性标注。比如，MORPH、celeA 等。下面对相关的数据集进行详细的介绍 **MOROH II**：MORPH 是一个大型的 mugshot 图像数据库，每个数据库都有相关的元数据，包含三个标注属性：年龄（有序），性别（无序）和种族（无序）。通过调查 MORPH Album II（MORPH II）上的所有三个属性估计任务，其中包含大约 78K 的超过 20K 个主题的图像。在 MORPH II 上的结果五等分数据进行交叉验证。todo MOrph 加入示例图片 **CelebA**：CelebA 是一个大型的人脸属性数据库拥有超过 10 万个身份的 200K 个名人图像，每个人拥有 40 个属性注释。该数据集中的图像在姿态，表情，种族，背景等方面存在较大的变化，使得面部属性估计具有挑战性。此外，由于有 40 个属性标注，CelebA 数据库在特征学习效率方面对联合属性估计算法提出了挑战。CelebA 的结果按照 [23] 中提供的协议报告。

（加入 celeA 属性表 todo）

LFWA：LFWA 是另一个无约束的人脸属性数据库 [23]，其中包含来自 LFW 数据库的脸部图像（5,749 个主题的 13,233 张图像）[52]，以及与 CelebA 数据库中相同的 40 个属性注释。按照 [23] 中提供的方案报告 LFWA 的结果。

Chalearn LAP and FotW：ChaLearn 挑战系列从 2011 年开始，在促进人们视觉或多模式分析方面取得了非常成功的成果 [53]。LAPAge2015 是一个无约束的脸部数据库，用于在 ICCV 2015.5 上发布的视在年龄估计。该数据库包含 4,699 张脸部图像，每个平均年龄至少由 10 个不同的用户估算。数据库被分割为 2,476 张图像进行训练，1,136 张图像进行验证，1,087 张图像进行测试 [51]。由于年龄信息的测试不可用，我们遵循 [17] 的协议，主要使用 validation 集进行测试。FotW 数据库是通过收集来自互联网的公开可用图像创建的，其中包含两个数据集，一个用于辅助分类，另一个用于性别和微笑分类。FotW 数据集分别包含 5,651,2,826 和 4,086 幅用于训练，验证和测试的面部图像; 每个都用七个二进制附件属性注释（见表 5（a））。FotW 性别和笑容数据集分别由 6171 个，3086 个和 8505 个面部图像组成，用于训练，验证和测试; 每个都注明三元性别（男性，女性，不确定）和二元微笑的属性。我们遵循相同的测试协议在 FotW 上报告结果。

LFW+：LFW+ 是扩展了 LFW 数据库 [52]，从无约束的人脸图像研究联合属性估计（年龄，性别和种族）。由于 LFW 数据库中的年轻受试者（例如 0-20 岁年龄段）

的数量非常少（根据 MTurk 工作者提供的标签，在 5,749 名受试者中仅有 209 该年龄段的受试者），LFW 数据库通过收集 2,466 使用 Google 图片搜索服务，在 0-20 岁年龄范围内的受试者的脸部图像不受约束。具体而言，我们首先使用“baby”，“小孩”和“青少年”等关键词从 Google 图片中找到约 5000 幅感兴趣的图片。然后将 Viola-Jones [54] 人脸检测器应用于生成一组候选人脸。最后，我们手动删除了错误的脸部检测以及大多数似乎超过 20 个的主题。扩展的 LFW 数据库（LFW +）包含约 8,699 个受试者的约 15,699 个无限制面部图像。对于每个脸部图像，要求三个 MTurk 工作者提供他们的估计年龄，性别和种族。表观年龄被定义为三次估计的平均值，性别和种族由多数票决定。在 LFW + 上的结果用五折，主题独特的交叉验证方案报道。

这些数据库可以根据所使用的注释方法分为三类：（i）具有名义和有序属性的数据库（MORPH II 和 LFW +），（ii）具有二进制属性的数据库（CelebA，LFWA 和 FWW）和（iii）具有单个属性的数据库（LAPAge2015）。我们可以看到，除了 MORPH II 数据库，其他五个数据库主要包含无约束的人脸图像。对这些数据库的属性估计评估可以提供真实应用场景下系统性能的见解。

3.2 人脸属性识别的单任务模型

基于人脸属性的单任务模型 STL（下称单任务模型），顾名思义就是给定一个图像，建立一个模型去对一个属性进行学习。这不仅是人脸属性识别任务中的常见做法，也是整个模式识别领域基础的框架模式，。作为一个分类问题，一些常见的模式识别方法也被应用其中，例如主动外观模型 AAM（Active Appearance Model），局部二值模式 LBP（Local binary patterns），加窗傅立叶变换（Gabor）等，这些常规的做法，总体来讲还是遵循特征提取工程再加上分类器模型的流程，包括特征相关性的筛选，不同模型的融合等等。但是很多时候根据固定模式提取的特征往往不够具有代表性，与识别任务的关联性不够高。于是大家开始着力想寻找一些相关性更高的方法包括，Fu 等将流形学习方法引入年龄估计；另外，Guo 等提出了生物启发的特征方法，Hu 提出了统计信息特征（Dif，Demographic informative feature）的概念；但是时至今日，CNN 特征在图像领域的出色表现，让人们对于单任务模型的使用和理解有了非常大的提高，我们根据日常的使用的经验和学术界普遍的做法总结了一套非常简练有效的的框架：在介绍这套框架之前，首先对于 hu 的 DIF 方法做一下简单的介绍，让大家对于常规的单属性任务模型有所了解，并且作为对比：hu 的基本框架

概述如下：前端为特征提取阶段，旨在提取对属性有判别力的特征，而不是完全无监督的。后端连接一个层级式的分类器，用于属性学习。（Todo hu 算法的框架，截图）

其中有几个主要部分：DIF（Demographic informative features）特征提取，层级式分类器，人机对于单属性预测任务的对比 1）DIF 特征 DIF（Demographic informative features）是基于 BIF（生物启发式特征）的。比如，输入一个人脸部件，先用 Gabor 滤波器提特征（12 个尺度，8 个方向），再做一些池化操作，以减小特征图的数目和维度（6 个尺度，8 个方向），将得到的特征串成一个 4280 维的长向量，用来做之后的分类等任务。总体上还是一个无监督的特征处理方法。所以之后，又对此工作做了改进，旨在不仅能够抓住图像细节，还能减小冗余性，提高特征与最终识别任务的相关性。这一部分主要引入一些特征学习工作，从之前的特征集中不断特征子集，挑选出最相关的特征，比如：学习一个新的特征子空间（如 LDA），基于 Boosting 的特征选择。2）层级分类器的建设层级分类器主要针对年龄。比如，首先进行年龄组分类（针对数据集设定阈值），在此按是否超过 18 岁分为两类；低于 18 岁的一类再判断是否低于 7 岁，再分为两类，然后低于 7 岁再进行回归得到具体的年龄数值，以此类推，先一层一层地通过多个分类器树形展开得到具体的人脸年龄段，然后在具体的人连年龄分段中及进行回归。hu 的实验证明，这种层级式的分类方式要优于直接分类方法。3）人机性能对比在人和机器的性能对比，hu 当时规模最大的数据集来衡量并对比人和机器的性能。数据集包含以下几个方面：FG-NET，年龄估计；MORPH（2000 张图片），年龄、性别、种族估计；PCSO（2000）张图片，年龄、性别、种族估计实验显示这种 dif 的方法取得了当时最好的结果，具有最小误差，且具有非常好的演示和出色的数学模型和理论推导。在人和机器的性能对比，可以看出机器识别能力的绝对误差要小于人类。当在做年龄估计时，算法估计偏差比较平衡。而人类往往会将年龄估计偏高。这里是误差分析，我们发现，虽然总体上机器性能高于人，但是机器会犯一些偏离实际较大的低级错误，这也是很多学习算法的共同问题。年龄估计，实验表明，算法对真实年龄和人类标注的表观年龄的估计偏差并不大。总体来讲机器的表现可圈可点。

但是需要注意的是在这个过程中，这个工作得益于作者的精心调试和改进，技术细节较为复杂很可能一个步骤做不好就整个系统崩溃，同时因为图片数据库和过多人工干预导致了一定程度上的局部最优解，在真实场景中，难以取得良好的效果，

下面介绍神经网络中的单模型预测框架，得益于现代神经网络的出色表现，单属性预测模型的 pipeline 得到了极大的简化，同时结合端到端的设计思路和数据量的

增加可以很好的提升单属性模型的预测效果。CNN 算法下的单模型输出预测：首先经过图像预测处模块，将图像通过一些基础的图形变换转到统一的形变空间中，常见的操作包括图像识别任务中的空间颜色变换，尺度统一化，多尺度变换，多位置截取等。在人脸属性的人物之中，我们经常采用的方式是人脸 alignment，也就是根据人脸检测输出的人脸边框位置和 landmark，通过仿射变换，将人脸图像中的关键点映射到图像中的标准位置。

然后设计网络结构作为图像特征网络提取模块，这一部分往往有两条规则可以遵循从而有效的搭建神经网络结构，第一规则是根据现有的经典神经网络结构进行改进，比如 alexnet，googlnet，resnet 等，这样做有两方面考虑，一部分是因为这些网络在实际使用中“久经考验”，体现出了良好的收敛性能，另一方面由于类似的神经网络在科研的过程中使用的人数和场景比较多，在搭建和调参上会有很多共同的地方可以互相交流，也方便不同方法的比较。所以总结来讲，其实如果主要的研究问题不在网络结构后对于识别任务的影响上，一般还是会使用业界通用的网络结构。第二条规则就是在自己设计附加的网络结构过程中，也要符合一定的网络特性，包括结构上的自洽，设计之中不能产生模块之间不匹配的情况，比如同一层网络输入大小相互有差别，网络操作参数设置不合理等初级问题等，这一点看上去很简单，实际上出错的几率非常大。针对于图像任务的标签选取不同的损失优化函数，常见针对于非连续的数据标签如分类问题，可以选取 softmax cross entropy loss 的集合，亦可以针对于每个分类标签设置为多个二分类的问题，然后多个的二分类的标签联合训练使用交叉熵 loss 进行训练，当然这两种 loss 本身具有很多相似的地方，且在类别中只有两类的时候，具有相同的表达形式，但由于 softmax with cross entropy loss 的简洁形式，往往再多分类问题中选取这种损失函数。我们也做了一些相关性的对比实验，发现整个模型的效果和时间都较之前有了很大的提升。（插入图表 todo）

3.3 人脸属性识别的多任务模型

如同上文提到的，人脸属性数据库的构建慢慢从一个单属性数据库的建设变为多属性。那么模式识别的任务也从对于单个的人脸属性进行识别，变为对于人脸图片的多种属性进行预测，如果针对每一个属性都完全独立出来，设计一个模型，那么其模型复杂度过大。因此，能否设计一个模型来实现多属性的识别呢？答案是肯定的，也是可行的。近几年人脸多属性识别的任务同样具有很多的进展，使得人脸多属性识别在相关数据集和场景中有了很大的进展。下面介绍一下具体情况和我的

贡献：基于单任务的多属性学习

首先可以考虑将人脸属性多任务转化为人脸单任务的做法，可以对于多属性的人脸标签分为以下几种情况：方法一：标签编码：将多属性标签组合进行编码（比如，将一岁亚洲男性标记为 001，将一岁非洲男性标记为 002 等），将多属性问题转化为分类编码问题，也就是单一属性。局限性：但是，对于属性数目较多的情况，这种方式会引起数据的组合爆炸。因此，该方法只适合属性数目很小的情形。

方法二：多标签回归通过回归的方法，使预测的特征向量与 Groud-truth 属性向量的损失越来越小，二者趋向接近，由此得到预测的特征向量。局限性：在提特征阶段，虽然有几十个属性，但用的都是同样的特征，未考虑不同属性的相关性和差异性。但是无论如何设计，强行转化为单任务的算法方式，不仅非常生硬不能很好拟合属性之间和属性内部的分布关系，而且看上去也非常不优雅。

基于多任务的属性学习：正如上文提到的，属性之间具有非常大的异构性，但是作为人脸特征，它们同时在很多表现过程之中，也有很多共同的地方，那么在设计的过程中我们更倾向于用的是单框架多任务方式。这也利用属性之间的相关性，包括正相关和负相关等来进行互相补足；同时多任务的方式设计也应对属性之间的异质性，比如年龄是可量化的，而种族是类别化的，这就需要不同的处理方式。我们对 CelebA 数据集的 40 个属性做了成对的 co-occurrence 计算，它揭示了，属性的相关性是普遍存在的，且我们认为它对属性学习有所帮助。（加入 co-occurrence 的图片）

不得不提的是，在完成这一任务的过程中，有两种方法对我产生了一定的影响，一个是 Ziwei Liu 和 Ping Luo 在 2015 年 ICCV 上发表的 Deep Learning Face Attributes in the Wild，另一个是 Hu Han 在 2017 年同样是 T-PAMI 上发表的 Heterogeneous Face Attribute Estimation: A Deep Multi-Task Learning Approach。这两篇论文分别代表了前 CNN 和后 CNN 时代，人脸属性多任务识别的不断演变和进展。

ZiWei Liu 提出了一个新的在非限制性场景下进行属性预测的深度学习框架。通过级联两个神经网络，LNet 和 ANet，它们与属性标签一起进行细调，但是预训练方式不同.LNet 通过大规模的一般对象类别预先训练用于人脸定位，而 ANet 通过大量的人脸识别进行预训练以进行属性预测。这个框架（1）它展示了人脸定位（LNet）和属性预测（ANet）的性能如何通过不同的预培训策略改善。（2）它揭示了尽管 LNet 的滤波器只是使用像素级别属性标签进行 finetune，但是它们在整个图像上的响应映射具有强烈的脸部位置指示。这个事实使得能够训练 LNet 用于仅具有图像级别注释的面部定位，但没有所有属性识别工作所需的面部边界框或地标。（3）也证明

了 ANet 的高级隐含神经元经过大规模人脸识别预训练后自动发现语义概念，经过与属性标记的微调后，这些概念得到了极大的丰富。每个属性都可以用这些概念的稀疏线性组合来解释。（加入 LIU ZiWEI 的图）Hu 提出了一个深度多任务学习（DMTL）方法来联合估计来自单个人脸图像的多个异构属性。所提出的 DMTL 包括一个早期阶段的所有属性的共享特征学习，然后是异类属性类别的特定类别特征学习（见图 2）。考虑单个卷积神经网络（CNN）中的属性相关性和属性异质性，对有序与无序性和整体与局部异质性的共同考虑导致四种类型的子网络：整体 - 名义，整体 - 顺序，局部 - 名义和局部 - 顺序。每种类型子网的损失函数的选择仍然取决于子网是名义上的还是有序的，共享特征学习自然地利用任务之间的关系来实现强健的和有区别的特征表示。类别特定的特征学习旨在对共享特征进行精细调整，以便对每个异构属性类别进行最优估计。由于有效的共享特征学习和类别特定的特征学习，所提出的 DMTL 在保持低计算代价的同时，实现了有希望的属性估计精度，使其在许多人脸识别应用中具有价值。值得一提的是，HU 解决了一个我在 Ziwei 的工作中一直迷惑的点，那就是对于属性的学习率动态调整的问题，解决的办法也很简单，就是使用简单的两层全连接网络，对于不同性质的子网络进行单独学习，他们认为图像空间和标签空间有一个高度非线性的关系，可以表示成 F，数据可以表示为 D，X（图像），Y（属性）（加入 Hu 网络结构图和高度非线性关系的图，加入 celeA co-occurrence 图）Hu 的工作非常全面，在所以的数据集上都进行了评测，同时也一定程度上探究了人脸属性对于不同数据集的泛化能力。（加入 HU celeA 测试 LFW 的表格）

3.4 对于不同标注数据集中人脸数据的充分利用

上面的两个工作其实已经解决了神经网络对于人脸属性的多任务学习问题，但是他们都有一个共同的缺点那就是，对于人脸数据的利用程度还不够，例如，虽然在各个数据集上 Hu 都进行了一定的评测，但是很明显，不同数据集的结果互相之间具有很大的差距，使用 celeA 训练的数据对于 lfwA 的数据效果并不好，实际上加入 lfwA 的数据训练就可以提升相关 lfwA 上的准确率。但需要注意的是 LFWA 的数据量远小于 celeA 的数据量，合起来训练，两个数据库之前的差异分布其实并不能得到特别好的弥补，训练的准确率还是不能和单独使用 lfwA 相比。类似的问题更严峻一点，对于年龄这一属性，不同的数据库标注是不一样的，在 morph 中是连续的标签，但是在 adience 数据集上，年龄的标注是 7 个单独的类别，如果强行进行 label 的转换就会存在很多不匹配的现象，无法对其进行测试，但根据 adience 重新 finetune，那么

就会存在类似的数据匹配和模型输出改变的问题。为了解决这一问题，只有改变思维的限制，那么能否改善现有的框架，使其对于任何标签的属性数据都能够并行的

3.5 人脸属性识别中的网络能力自评估模块的设计

第四章 对抗神经网络在人脸属性中的应用

4.1 对抗生成网络相关技术的介绍

4.2 对抗神经网络在人脸属性中的应用

4.3 对抗神经网络对于属性识别中提升

第五章 总结与展望

5.1 全文总结

5.2 未来展望

附录 A 不定型（0/0）极限的计算

定理 A.1 （L'Hospital 法则 ） 若

1. 当 $x \to a$ 时，函数 $f(x)$ 和 $g(x)$ 都趋于零；
2. 在点 a 某去心邻域内，$f'(x)$ 和 $g'(x)$ 都存在，且 $g'(x) \neq 0$；
3. $\lim\limits_{x\to a} \dfrac{f'(x)}{g'(x)}$ 存在（或为无穷大），

那么

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)}. \tag{A-1}$$

证明： 以下只证明两函数 $f(x)$ 和 $g(x)$ 在 $x = a$ 为光滑函数的情形。由于 $f(a) = g(a) = 0$，原极限可以重写为

$$\lim_{x\to a} \frac{f(x) - f(a)}{g(x) - g(a)}.$$

对分子分母同时除以 $(x - a)$，得到

$$\lim_{x\to a} \frac{\dfrac{f(x) - f(a)}{x - a}}{\dfrac{g(x) - g(a)}{x - a}} = \frac{\lim\limits_{x\to a} \dfrac{f(x) - f(a)}{x - a}}{\lim\limits_{x\to a} \dfrac{g(x) - g(a)}{x - a}}.$$

分子分母各得一差商极限，即函数 $f(x)$ 和 $g(x)$ 分别在 $x = a$ 处的导数

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \frac{f'(a)}{g'(a)}.$$

由光滑函数的导函数必为一光滑函数，故 (A-1) 得证。 □

致　谢

感谢 Donald Ervin Knuth.

攻读学位期间发表的学术论文目录

期刊论文

[1] **Zhang San** , Newton I, Hawking S W, et al. An extended brief history of time[J]. Journal of Galaxy, 2079, 1234(4): 567–890. （SCI 收录，检索号：786FZ）.

会议论文

[2] McClane J, McClane L, Gennero H, et al. Transcript in Die hard[A]. // Proc. HDDD 100th Super Technology Conference (STC 2046)[C]. Eta Cygni, Cygnus: 2046: 123–456. （EI 源刊）.

专利

[3] 张三 , 李四. 一种进行时空旅行的装置 [P]. 中国: 1234567, 2046–01–09.