

密级：

保密期限：

# 北京邮电大学

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





# 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 and 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，研究工作的时间跨度度并非很大，但是，各方面工作的丰富性和多样性还是令人瞩目的。从特征的表达方法来看：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生成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。当然你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

- 3.1 人脸属性数据库简介
- 3.2 人脸属性识别的单任务模型
- 3.3 人脸属性识别的多任务模型
- 3.4 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  $x \rightarrow a$  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.  $\lim_{x \rightarrow a} \frac{f'(x)}{g'(x)}$  存在 (或为无穷大),

那么

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \lim_{x \rightarrow a} \frac{f'(x)}{g'(x)}. \quad (\text{A-1})$$

**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

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

对分子分母同时除以  $(x - a)$ , 得到

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

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

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

由光滑函数的导函数必为一光滑函数, 故 (??) 得证。 □



## 致 谢

感谢 Donald Ervin Knuth.





## 攻读学位期间发表的学术论文目录