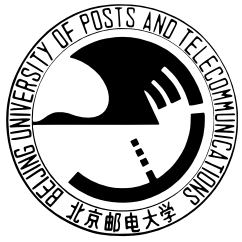
密级：　　保密期限：

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

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百分之 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, 是一个从全局特征、细节特征到深度特征的过程，具体来讲: 全局表观特征：包括 Intensity，图像 PCA，BIF 生物启发式特征，局部二值模式 LBP（Local binary patterns），加窗傅立叶变换（Gabor）等。细节特征如：主动外观模型 AAM（Active Appearance Model），纹理，肤色，人脸形状，sift 特征等。深度学习特征：如 CNN,DNN 中网络的不同层卷积输出。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

在这一章中主要介绍涉及人脸属性的一些深度学习基本常识和常见的计算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及训练方式、多机多卡训练的工作和优化、网络前馈的工程加速等。这些工作都是我整个研究生生涯花费了大量的精力去理解并且思考的，在很多地方也总结了一些看法和规律，在我的研究生科研和工作过程中起到了非常重要的作用。

2.1 卷积神经网络的基础操作和训练

卷积神经网络一般是指是针对以共享式多通道卷积操作为代表的一连串数学操作计算组合的一个总称，因为卷积计算是其中的主要计算过程和核心特征提取方式，而计算的过程往往需要加入一步非线性的激活函数用以增加整个计算过程对于非线性过程的模拟，和人体的神经元结构非常相似，所以从计算科学的角度称为卷积神经网络。由于卷积神经网络具有参数多，训练数据广的特点，难以通过正常的线性代数和微积分求得最优解，一般会使用反向传播算法进行训练，也称梯度下降算法。

在图像识别领域中，卷积神经网络和普通的神经网络，玻尔兹曼感知机等很像：都是把输入数据最后转化成输出；都要使用输入并进行点积运算；都使用可以学习参数的神经元；神经元都含有非线性激活函数；在最后都加入分类的损失函数等等。

而卷积神经网络，借鉴于其常见的多维向量点乘的结构的设计，使其对于图像的 2d 结构具有良好的亲和性，利用这个特点，基于图像的卷积神经网络结构层出不穷，各自的识别效果是不尽相同。接下来，笔者将从基本网络结构的组成、常用非线性激活函数、常用初始化参数方法、卷积神经网络的训练与优化四个方面来介绍卷积神经网络的相关内容。

2.1.1 卷积神经网络结构的基本组成

层是卷积神经网络结构的基本组成单位，不同的网络结构中都是使用类似或者基础的层来进行搭建完成，少则三、四层，多达成百上千，但无论是深度还是广度的扩增，都是通过增加层的使用来完成。

2.1.1.1 卷积层

卷积层是卷积神经网络最重要的层，因为卷积层承担着从图像到高层语义的转化任务，实际使用中也占据了整个网络计算量中的绝大部分。甚至毫不客气地讲，对于卷积层的实现和优化好坏，决定着一个深度学习框架的工作使命和存在意义。（最直观的例子就是最著名的卷积神经网络框架 caffe 就是因为对于卷积的实现做得好，训练和测试的速度较快，从而取代了 cuda convnet 和 CXXnet 成了主流，也是目前影响最深的深度学习框架。）

具体来讲：每一个卷积层都会使用 N 个不同参数的卷积滤波器核，每一个卷积滤波器核对整个输入特征图进行卷积操作，在这个卷积过程中是只是用那一个核的参数，也就是所谓的参数共享。参数层面: 卷积滤波器核超参数包括：

1. N 卷积滤波器核数目
2. $Kernel_h$,$Kernel_w$ 卷积滤波器核高度和宽度
3. $Stride_h$,$Stride_w$ 卷积在高、宽维上的步长
4. $PadH$,$PadW$ 对于高、宽二维上对于 feature map 的空白补足

输入输出参数:

1. $Coutput$,$Houtput$,$Woutput$ 分别为输出通道数、输出高度、输出宽度
2. $Cinput$,$Hinput$,$Winput$ 分别为输入通道数、输入高度、输入宽度

其中输出参数由如下公式确定：

$$\begin{aligned} Woutput &= \frac{Winput - KernelW}{StrideW} + 1 \\ Houtput &= \frac{Hinput - KernelH}{StrideH} + 1 \\ Coutput &= N \end{aligned} \tag{2-1}$$

卷积层的内部参数包括权重 W 和偏置 b,

1. W 是一个维数是 $Coutput * Cinput * KernelH * KernelW$ 多维数组.
2. b 是维数为 $Coutput * 1$ 的数组。

卷积的基本计算公式如下：

$$x_j = f\Big(\sum_{i\in(convscale)} x_i * W_i + b_j\Big) \tag{2-2}$$

convscale 是指对应一次卷积操作中对应的 $Coutput * kernel_h * kernel_w$ 范围内的输入图片和对于第 j 层的 W 和 b. 在计算机对于卷积的实现操作过程中因为参数和输入的特征都是按照数组的方式进行储存，所以可以简单的认为是对于一定长度的一维数组进行点乘操作，实际上为了减少内存的读取所带来的延时，很多快速算法都实用类类似的想法。具体计算的示意图如下：

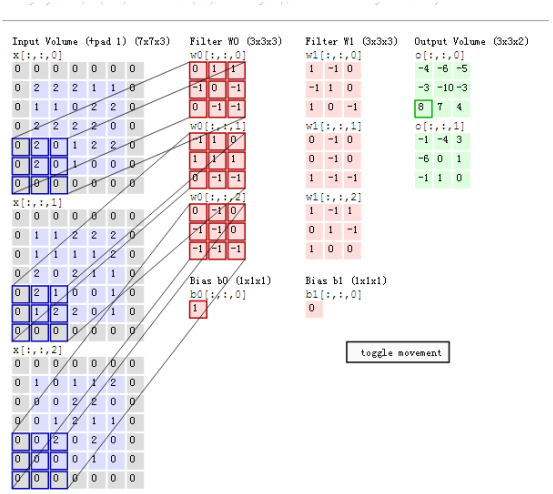


图 2-1　卷积操作的示意图

除此之外，很多不同形式的卷积更加注重对于图像信息提取的卷积形式也被提出，比如放射卷积，形变卷积等，他们都不再局限于对于固定范围内特征值进行卷积，而是着重于在输入特征图上更多具有影响力的范围内进行输入位置的选取。

2.1.1.2　池化层

池化（Pooling）层可以相比于信号处理中的采样操作，一般对于输入的特征图片图进行降采样操作，从而起到去除噪声，增加网络的旋转不变性和唯一不变性，提升整体的训练效果。事实上池化层有时也被用作上采样，也就是 uppooling，通过输入的特征图进行等间隔复制的方式得到原尺度多倍的输出，但是随着卷积神经网络的慢慢发展，人们开始主要使用反卷积的方式来实现上采样的操作，以至于 uppooling 的操作慢慢不被人所熟知。

池化层的超参数包括：

1. KernelH,KernelW 高宽方向上的池划范围大小

2. StrideH,StrideW 高宽方向上的池化步长

输入输出参数包括：

1. Channel,Height,Width 分别为输入通道数、输入高度、输入宽度

2. Coutput,Houtput,Woutput 分别为输出通道数、输出高度、输出宽度

其中输出参数由如下公式确定：

$$\begin{aligned} Woutput &= \frac{Winput - KernelW}{StrideW} + 1 \\ Houtput &= \frac{Hinput - KernelH}{StrideH} + 1 \\ Coutput &= Cinput \end{aligned} \tag{2-3}$$

池化层的计算公式如下：

$$x_j = poolmethod(x_i) \qquad i \in (poolsacale) \tag{2-4}$$

pool scale 和 conv scale 类似，是指一次池化操作中对应的 $kernel_h * kernel_w$ 范围 poolmethod 代表的是具体的下采样函数。通常有三种类型，一种是最大型池化，一种是均值型池化，还有是随机型池化。其中最大型池化使用最为广泛，因为其最能够体现池化层对于平移和旋转操作的鲁棒性，而均值池化层主要用在特征归一化操作和最后高位信息的整合上面, 以最常见的 max-pooling 操作为例，具体的计算示意图如下：

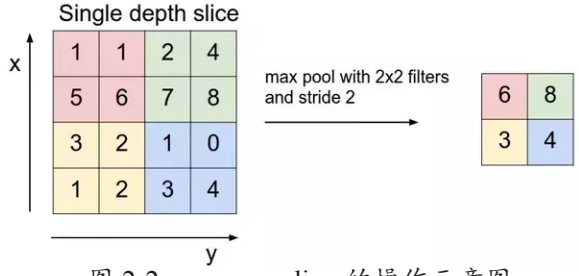


图 2-2　max-pooling 的操作示意图

和卷积层的发展类似，pooling 层也慢慢发展出很多不同的分支，除了前面提到的 up pooling 层之外，在物体检测中存在着使用非常广泛的 roi-pooling 和空间金字塔 pooling（也就是 sppnet）

2.1.1.3 全连接层

全连接（Fully-connected，FC）层是普通机器学习方法中使用最多的层，包括SVM，随机森林，boosting 种处处可见与之呼应的操作，简单来讲输入的类似于一维向量通过一个矩阵做矩阵乘法运算得到另一个一维向量的过程。

全连接层使用过程中需要设定输出大小即可，其自身的参数是一个输入大小乘以输出大小的矩阵，如果采用带有偏置的算法则还需要一个和输出长度相同数目的偏置项，总结来讲全连接层的参数包括：

1. Fc_input 输入大小
2. Fc_output 输出大小
3. Fc_W 模型参数 W
4. Fc_b 模型参数偏置项

具体的计算过程可以通过公式进行表达：

$$Fc_j = X_i \ W_j^T + bj \tag{2-5}$$

从实现的角度上看，可以比较明显的看出全连接层的整体实现可以通过矩阵乘法来完成，通过和卷积层类似的方法将所有的特征输入扩展成多维的等长向量也就是矩阵然后通过，矩阵乘法来快速的将全连接层的相关问题进行实现。如下是输入为 800，输出为 500 的全连接层示意图。

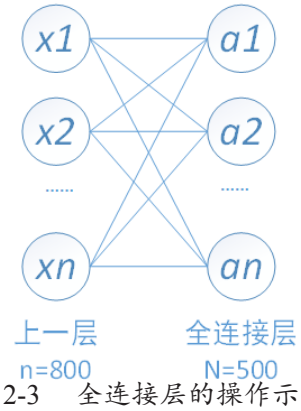


图 2-3 全连接层的操作示意图

但是需要注意的是，在使用过程中，全连接层的的参数量是 n^2 也就是说全链接层很容易出现参数量过大的问题。比较著名的案例就是 VGGnet 种最后的三个 FC

层，几乎占据了整个网络一半的参数量，实际中完全可以使用 1*1 卷积等类似的操作对其进行替代以减少参数量。

2.1.1.4 归一化层

归一化层指的是对每一层的输出进行标准化操作，使得下一层的输入保持一个较为稳定的分布。常用的标准化层有局部区域归一化（Local Region Normalization，LRN）层，批量标准化（BatchNomalization，BN）层等。这里着重介绍一下 BatchNormalization 层：

Batch Normalization 也就是批量规范化，在网络训练时，对每个 mini-batch 的特征在卷积之后做一次规范化，使得其输出的特征为零均值和 1 方差的，在加入两个动态学的参数，尺度（scale) 和偏移景（shift), 用于还原最初的输入，从而保证整个网络的表征能力。实际上，BN 层的引入本质上是为了使得每一个卷积层的输入数据的分布统一化，即保证相同均值与方差。另外一个主要的原因则足防止梯度消失，通过将输入归一化到固定的均俏和方差, 使得原本尺度很小的特征相应变火，后馈计算时其梯度也相应增大，从而很好的防止梯度弥散。

BN 层的参数包括 global status（使用网络参数中的均值方差还是根据输入数据重新计算）、moving average fraction（每次计算的累加方式）、eps（为了防止方差为了 0 加入的偏置项）

Input: Values of x over a mini-batch: $\mathcal{B} = \{x_{1...m}\}$;
Parameters to be learned: γ, β
Output: $\{y_i = \mathrm{BN}_{\gamma,\beta}(x_i)\}$

$$\mu_{\mathcal{B}} \leftarrow \frac{1}{m}\sum_{i=1}^{m} x_i \qquad \text{// mini-batch mean}$$

$$\sigma_{\mathcal{B}}^2 \leftarrow \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_{\mathcal{B}})^2 \qquad \text{// mini-batch variance}$$

$$\widehat{x}_i \leftarrow \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}} \qquad \text{// normalize}$$

$$y_i \leftarrow \gamma \widehat{x}_i + \beta \equiv \mathrm{BN}_{\gamma,\beta}(x_i) \qquad \text{// scale and shift}$$

图 2-4 BN.png

BN 的提出在神经网络的发展过程中具有非常重要的意义，大大改善了神经网络的收敛问题，也可以说从某种程度上改进了神经网络的整体效果。

2.1.1.5 损失函数 loss 层

loss function 是神经网络中用来衡量网络预测和真实值之间的误差情况，最常用的决策层损失函数是 Softmax 损失函数和欧几里得损失函数。

Softmax 损失函数主要用于多分类任务，其具体的损失函数表达为：

$$l(y,z) = -log\left(\frac{e_y^z}{\sum_{j=1}^{m} e_j^z}\right) \tag{2-6}$$

其中 m 表示分类的类别总的数目，y 表示标签，z 表示网络预测的类别。也就是说将所有类别的预测值取他们的指数值求和，然后判断实际标签中样本在其中所占的比重，并将其取 log 作为 loss 函数和优化的值。

欧几里得损失函数主要用于回归任务, 具体的回归损失函数如下：

$$l(y,z) = (z-y)^2 \tag{2-7}$$

其中 y,z 同 softmaxloss 的含义相同。

除了这两种常用的 loss 函数之外，还有比如交叉熵损失函数，smooth L1 损失函数等，都是在分类和检测中经常使用的损失函数。

2.1.2 卷积神经网络常用激活函数

卷积输出之后通常会使用激活函数进行非线性激活，从而增强网络的模拟变换能力，不然只是线性变化的组合可以涵盖的空间非常有限. 下图总结了神经网络中经常使用的激活函数：

Name	Formula	Time
sigmoid	$y = 1/(1+e^{-x})$	1986
tanh	$y = (e^{2x}-1)/(e^{2x}+1)$	1986
ReLU	$y = \max(0, x)$	2010
SoftPlus	$y = \ln(e^x+1) - \ln 2$	2011
LReLU	$y = \max(x, \alpha x), \alpha \approx 0.01$	2011
maxout	$y = \max(W_1 x + b_1, W_2 x + b_2)$	2013
APL	$y = \max(0, x) + \sum_{s=1}^{S} a_i^s \max(0, -x + b_i^s)$	2014
VLReLU	$y = \max(x, \alpha x), \alpha \in 0.1, 0.5$	2014
RReLU	$y = \max(x, \alpha x), \alpha = \text{random}(0.1, 0.5)$	2015
PReLU	$y = \max(x, \alpha x), \alpha$ is learnable	2015
ELU	$y = x, \text{if } x \geq 0, \text{else } \alpha(e^x - 1)$	2015

图 2-5 激活函数的具体表达式以及出现时间

其中比较重要的是 sigmoid 函数、relu 函数、以及 prelu 函数，下面是他们的函数曲线图：

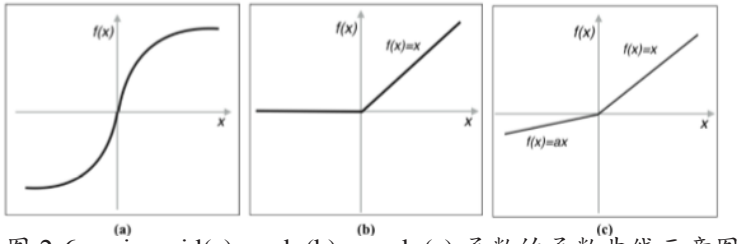


图 2-6　sigmoid(a)、relu(b)、prelu(c) 函数的函数曲线示意图

2.1.2.1　sigmoid 函数

sigmoid 函数是神经网络最早使用的激活函数，从函数的特性上可以看到其能把输出映射到区间 (0,1)：若输入趋于负无穷, 则趋近于 0；若输入趋于正无穷，则输出趋近于 1。近年来由于梯度下降法的使用增加，sigmoid 在数值较大的情况下，导数趋近于 0，这样导致的后果则是对应的梯度也会慢慢消失，导致训练的过程变得缓慢难以得到正常的收敛效果。

2.1.2.2　relu 函数

RELU（Rectified Linear Unit）激活函数，中文也称修正式线性激活函数。由于自身的非饱和特性，修正式线性单元极大程度的加速了深度卷积网络的收敛。但是修正式线性单元也存在一个很大的问题。在训练的时候，修正式线性单元比较脆弱并且可能“死亡”。举个例子来说，当一个很大的梯度流过修正式线性单元的神经元的时候，可能会导致梯度更新到一种特别的状态，在这种状态下神经元将无法被其他任何数据点再次激活。如果这种情况发生，那么从此所以流过这个神经元的梯度将都变成 0。也就是说，这个单元在训练中将不可逆转的死亡，而这样会导致数据多样化的丢失。

2.1.2.3　prelu 函数

prelu 函数在负轴上加入了固定大小斜率的 a, 从而确保梯度不会因为突然死亡的问题而导致网络崩溃。

2.1.3 卷积神经网络常用的参数初始化方法

神经网络求解的是局部最小值，一个好的参数初始化方法能使得卷积神经网络收敛且收敛的更快。常用的卷积神经网络初始化方法有如下几种。

1. 常数初始化: 使用固定的常数初始化每个参数，常用来初始化的常数一般比较小，通常为 0。常数初始化方法通常对偏置项所使用。

2. 均匀分布初始化: 假设参数服从在区间 [l,h] 上的均匀分布，进而为参数进行初始化。通常为权重参数使用。Xavier 等在 2010 年提出的 Xavier 初始化方法 [30] 就是一种均匀分布初始化方法 Xavier 初始化方法能够使得每一层的输出方差尽量相等，从而让网络中的信息更好的流动。

3. 高斯分布初始化：为 0 方差的高斯分布，σ 为参数进行初始化。通常为权重参数使用。σ 可以是人为制定也可以是通过输入输出计算得到。实验证明，对于较深的卷积神经网络，MSRA 初始化方法比 Xavier 初始化方法更容易收敛。

4. gabor 初始化，gabor 初始化的方法是根据 gabar filter 的参数直接作为神经网路的参数，一般在网络的第一层进行使用。由于其具有天然的图像滤波特性，所以很多时候可以固定参数。不对其就行学习，从而达到减轻训练时间和负担的作用。

2.1.4 卷积神经网络的训练与优化

通常对于包含 N 个数据的数据集 D ，优化的损失函数可以写成：

$$J(W,b)=\frac{1}{N}\sum_{i=1}^{N}l(y_i,z)+\lambda\Phi(W) \tag{2-8}$$

以具有动量的 SGD 梯度下降方法为例：

$$\begin{aligned} V_{t+1}&=\mu V_t-\alpha\bigtriangledown l(W_t)\\ W_{t+1}&=W_t+V_{t+1} \end{aligned} \tag{2-9}$$

其中 t+1 是迭代的当前轮数，W 是需要更新的参数，$a\bigtriangledown l(W_t)$ 是目标损失函数对于 W_t 的偏导数，V_t 是上一次参数的更新量，V_{t+1} 是本次的参数更新量，μ 是动量值，α 是学习率。

卷积神经网络的训练主要使用基于梯度的反向传播（Backpropagation，BP）算法。假设卷积神经网络一共有 N 层，记作 $L_1 \cdots L_n$，y 代表样本标签，z 代表每一层的输出，a 代表每一层输出的激活值，δ 代表每一层传回的梯度值，W 为权重，b 为偏置项，则 BP 算法步骤如下：

1. 进行前馈运算，利用前向传导公式，得到 $L_1 \cdots L_n$ 的激活值

2. 对 LN 层，计算损失函数对应的偏导值

3. 对于第 i 层 Li，i = N−1,...,2，计算输出的梯度

4. 依次计算每一层参数的梯度

5. 利用公式 2-19 更新参数值。

2.2 多机多卡策略对于网络训练的提升

在大型数据集上进行训练的现代神经网络架构可以跨广泛的多种领域获取可观的结果，领域涵盖从语音和图像认知、自然语言处理、到业界关注的诸如欺诈检测和推荐系统这样的应用等各个方面。但是训练这些神经网络模型在计算上有严格要求。尽管近些年来 GPU 硬件、网络架构和训练方法上均取得了重大的进步，但事实是在单一机器上，网络训练所需要的时间仍然长得不切实际。幸运的是，我们不仅限于单个机器：大量工作和研究已经使有效的神经网络分布式训练成为了可能。多机多卡训练也可以理解为分布式训练，但是由于传统的分布式训练主要基于 cpu 进行计算。而现代深度学习的分布式框架学习，gpu 的使用是其中不得不实现的一个部分，所以多机多卡的形容更加贴切。

关于多机多卡的训练策略和普通的单机训练相面临更多的挑战，作为研究生生涯中非常感兴趣的一部分，也是整个神经网络实验的基础研究部分，将我对于多吉多卡的一些调研和感悟简单做介绍：

2.2.1 并行模式

多机多卡训练一般有两种模式：数据并行和模型并行。

模型并行（model parallelism），在分布式系统中的不同机器分别负责在单个网络的不同部分计算——例如每层神经网络可能会被分配到不同的机器。

数据并行（data parallelism），不同的机器有着整个模型的完全拷贝；每个机器只获得整个数据的不同部分。计算的结果通过某些方法结合起来。

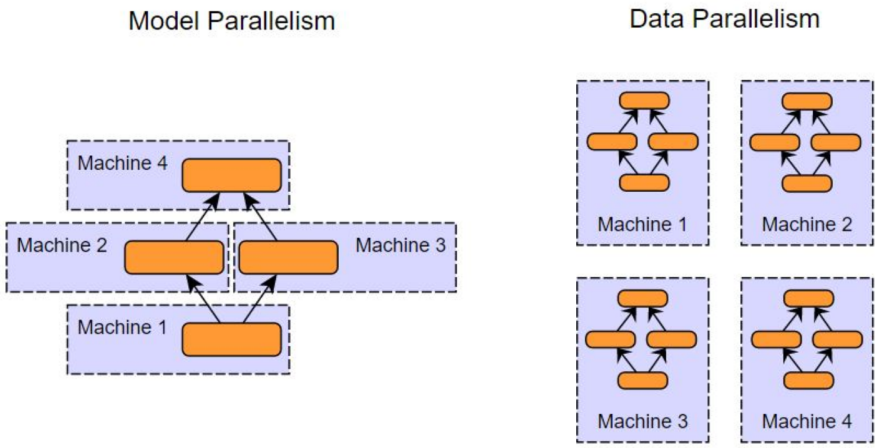


图 2-7 模型并行与数据并行示意图

当然，这些方法并不是互相排斥的。想象一个多 GPU 系统的集群，我们可以对每个机器使用模型并行（将模型分拆到各个 GPU 中），并在机器间进行数据并行。尽管在实践中模型并行可以取得良好的效果，但数据并行毫无争议是分布式系统中最适的方法，而且也一直是更多研究的焦点。实现性、容错性和好的集群利用率让数据并行比模型并行更加简单。分布式训练中的数据并行方法在每一个 worker machine 上都有一套完整的模型，但分别对训练数据集的不同子集进行处理。数据并行训练方法均需要一些整合结果和在各工作器（worker）间同步模型参数的方法。

2.2.2 参数更新方式

对应着数据并行训练方式，如何对于参数进行更新也就成了非常关键的问题，目前比较主流的参数更新方法有两种：参数平均化的更新方式，异步随机梯度下降的方式。

参数平均化：参数平均化是概念上最为简单的数据并行方法。使用参数平均时，训练按照如下方式执行：

1. 根据模型配置随机初始化网络参数
2. 将现有的参数的一个副本分配给每一个 worker machine
3. 在该数据的一个子集上对每一个 worker 进行训练

4. 从每一个 worker 的平均参数上设立一个全局参数

5. 当还需要处理更多数据时，回到第 2 步

同参数平均化相似的方法是：基于更新的数据并行化（‘update based’data parallelism）。两者的基本区别在于：我们不会将参数从 worker 传递给参数服务器，而是传递更新（例如：梯度柱型的学习率和动量（gradients post learning rate and momentum））当我们放松同步更新的要求时，基于更新的数据并行变得越来越有趣（毫无疑问它更有用）。即在更新参数的变化值被计算的时候就应用于参数向量（而不是等待所有 worker 的 $N \geq 1$ 次迭 1 代）。这就催生了异步随机梯度下降算法

异步随机梯度下降： 异步的随机梯度下降故顾名思义，根据一定的算法对于多个 machine 的更新的过程进行权衡处理，但是保持训练机器的训练过程不中断，可以预见异步随机梯度下降有两个主要优点：首先，潜在可能在整个分布式系统中获得更高的通量：worker 可以将更多时间花在执行有用的计算而不是等待参数平均化步骤完成。其次，worker 有可能可以集成来自其它 worker 的信息（参数更新），这比使用同步（每 N 个步骤）更新更快通过在参数向量中引入异步更新，也引入了一个新问题，也就是过期梯度问题（stale gradient problem）。过期梯度问题很简单：梯度（更新）的计算需要时间，在一个 worker 完成这些计算并将结果应用于全局参数向量前，这些参数可能已经更新过许多次了，所以如何设定这些更新的策略也是现在多机多卡的研究热点所在。

2.2.3 基于机器学习框架的多机多卡训练

2.2.3.1 caffe 中的多机多卡

caffe 是由贾扬清开发的轻量级 C++ 深度学习框架，因为其出现时间早，计算速度快等优势被 Alex 用于训练 imagenet 并且大获成功，从而大获成功，是影响比较大的学习框架之一。作为用户，在 caffe 中实现多卡训练比较简单，主需要在命令行中设置参数 -gpu gpu_ID 就可以选择希望占用的 GPU。在综合了代码和测试效果来看,caffe 是采用了数据并行，参数平均化更新的策略，实际测试中也可以很明显的看到默认 0 卡的显存占用更大，而且存在一定的间隔内，显卡中的使用率是会有明显空缺的，除去数据读取的原因之外，参数同步也是其中重要的一部分原因。

在 nVidia 开源的 nVidia-caffe 中，则大胆的采用了异步的 SGD 参数更新方式，可以很明显看到不仅训练的速度快，而且由于其训练和测试都被分摊到了单独的显卡，

而不是都在 0 卡上进行同步，导致其训练的所占用的显存要比普通的 caffe 版本更加小很多，但是所带来的问题也同样明显，在训练的过程中可以明显看出其收敛的稳定性不够出色，同样的训练参数和训练数据在普通的 caffe 中训练，可以收敛 loss 曲线也非常平稳，但是 nVidia-caffe 中却迟迟不能收敛，而且网络的训练也不够稳定，易出现崩坏的情况。

再来看基于分布式的多机多卡训练，在 intel/caffe 中提供了多机多卡的训练方式，使用的方式稍微复杂，需要安装 ansible，然后在用来训练的机器上配置好主机的 SSH 公钥验证，然后使用 ansible 统一安装 caffe 所需要的软件同时编译 caffe，设置好同步的文件夹，然后就可以选择需要的训练配置文件和数据就可以了。使用 mpi 命令配合执行 caffe train 命令就可以实现多机分布式训练网络。

综合来看，从工程上来讲，目前的分布式的多机训练其实是建立以往分布式训练的基础之上，有很强的 spark 和 Hadoop 烙印，对于多卡训练，大多数的实现都是基于 nvidia-NCCL（Nvidia Collective multi-GPU Communication Library）的多 GPU 通信库。NCCL 是 Nvidia Collective multi-GPU Communication Library 的简称，它是一个实现多 GPU 的 collective communication 通信（all-gather, reduce, broadcast）库，Nvidia 做了很多优化，以在 PCIe、Nvlink、InfiniBand 上实现较高的通信速度。

而从算法上来看，无论是同步训练还是异步训练其实都有其各自的挑战，对于同步训练来讲，虽然一定程度上加速了网络的训练速度，但神经网络在 batch 数目较大情况下的优化其实不是一帆风顺，而且随着 cluster 的数目增加，同步时延会成为性能关键，树形和环形拓扑都会成为其下一步的改进方向。

实际上异步算法更加受到人们的期待，就是因为其一定程度上摆脱了同步时延限制，能够实现了硬件性能的线性加速。从其他机器学习的时间和发展过程来说，如果深度学习的使用得到广泛的应用，那么异步算法的优化就会是大势所趋，因为对于实际应用的成本压缩和人们的对于性能的追求决定了算法的方向。

2.3 网络前馈速度的优化

网络前馈又被称为网络的 inference，一般是指经过一定的数据训练，对于输入数据和输出结果具有一定正向的科学计算过程，可能这样的说法不是很准确，因为很多时候为了获得更为理想的判断结果，往往会采取人工检查和机器过滤的两种方式进行结合，那么在我的工作之中主要是对于机器过滤中对于所需要使用的科学计算过程所进行的一些速度上的优化。具体算法的表述形式可以参照 2.1 节中的神经网

络的基础算法过程进行对比。

2.3.1 卷积计算的优化方式

对于卷积层的优化有三个方向，对于直接计算卷积的过程进行优化，使用第三方的加速库对卷积计算进行加速、和结合卷积的快速算法进行优化。笔者的主要工作贡献集中在第一种、和第二种。第三种因为非常重要也非常有趣，也稍作介绍作为未来工作的目标

2.3.1.1 直接进行卷积的算法优化

CPU 直接计算卷积：正如之前的基础知识中介绍的，普通的卷积操作需要对于输入特征中的 N，C，H，W 四维数据中进行提取，然后和卷积参数中 Coutput,Cinput,kernelH,KernelW 各个维数中的参数进行分别点乘，通常的写法使用 for 循环进行完成的话，需要使用 7 个 for 循环来完成：（加入 for 循环的计算图）这种粗暴的实现方式会使用 N*Cinput*Coutput*Woutput*Houtput*KernelW*kernelH 次乘法和加法操作，并且读取相同次数的内存。

使用向量指令集来加速卷积的直接计算方式：向量指令集的一个最大的优点是它能够允许软件传递大量的并行任务给硬件，而只需要一条很短的指令即可。在 SSE4 指令集中可以一次进行 4 次乘法操作，而在 AVX512 指令集中，可以一次完成 16 个 float 的乘法。结合这一思想，可以有效加速对于卷积中的乘法操作。尤其在移动端的优化计算种，

Winograd 卷积算法：使用线性代数分解的方式将一些固定卷积核尺寸大小卷积操作如 2x2,3x3 等，分解成多个具有固定参数的小矩阵相乘的方式，从而大幅度提升了卷积的速度。

MEC：im2col+gemm 的改进版，在减少内存的同时顺便可以提升一些速度。

2.3.1.2 借助第三方的计算库对于卷积计算进行优化

1. imcol+gemm+blas, 最常见的卷积快速算法是 imcol+gemm，（gemm 是矩阵乘法的简称）, 在卷积层的介绍中，可以看出笔者把卷积的每一个输出值都用多个输入数据和卷积层参数值相称的求和的方式表示，也就是向量积，一共需要做 Coutput*Houtput*Woutput 次这样的向量乘法，而且每次向量乘法的维数都一致，所以可以通过矩阵乘法来实现相关操作。具体的操作可以参见下图

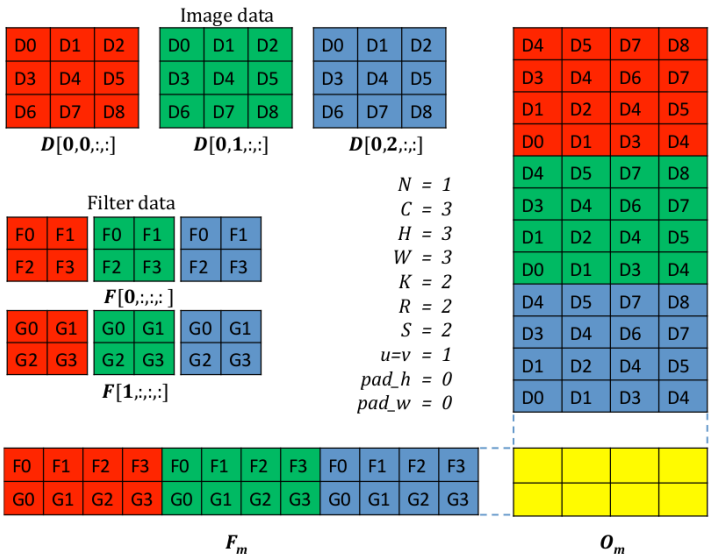


图 2-8 im2col 的操作示意图

矩阵乘法可以借助 openblas、MKL、altblas 等的线性代数优化库，可以有效地提升时间。

2. 使用 MKLDNN 中的卷积层实现, 用于深度神经网络的英特尔（R）数学内核库（英特尔（R）MKL-DNN）是用于加速英特尔（R）体系结构上的 DL 框架的深度学习（DL）应用程序的开源性能库。英特尔（R）MKL-DNN 包括高度矢量化和线程化构建模块，用于实现具有 C 和 C ++ 接口的卷积神经网络（CNN）。

3. 使用 NNPACK 中的 FFT 卷积操作，NNPACK 也是神经网络计算的加速包，基于傅立叶变换和 Winograd 变换的快速卷积算法，优势在于没有附加的依赖库，非常适合于移动端的开发。

4. 使用 CUDNN 和 TensorRT 中的卷积实现，cudnn 和 tenorRT 是使用在 gpu 上的算法加速库，都是 nVidia 为了加速神经网络而开发的闭源库，但是可以通过下载现有的公开库进行使用，cudnn 和 tensorrt 是所有加速库中能够获得加速比最高选择，原因在于其对于 GPU 的出色使用。

2.3.2 不同网络层的合并

事实上，随着神经网络的不断发展，卷积神经网络的层的种类其实是不断扩充的，有很多在随后的岁月中被发明并且广泛的使用，如 BatchNorm 层、relu 层等等，

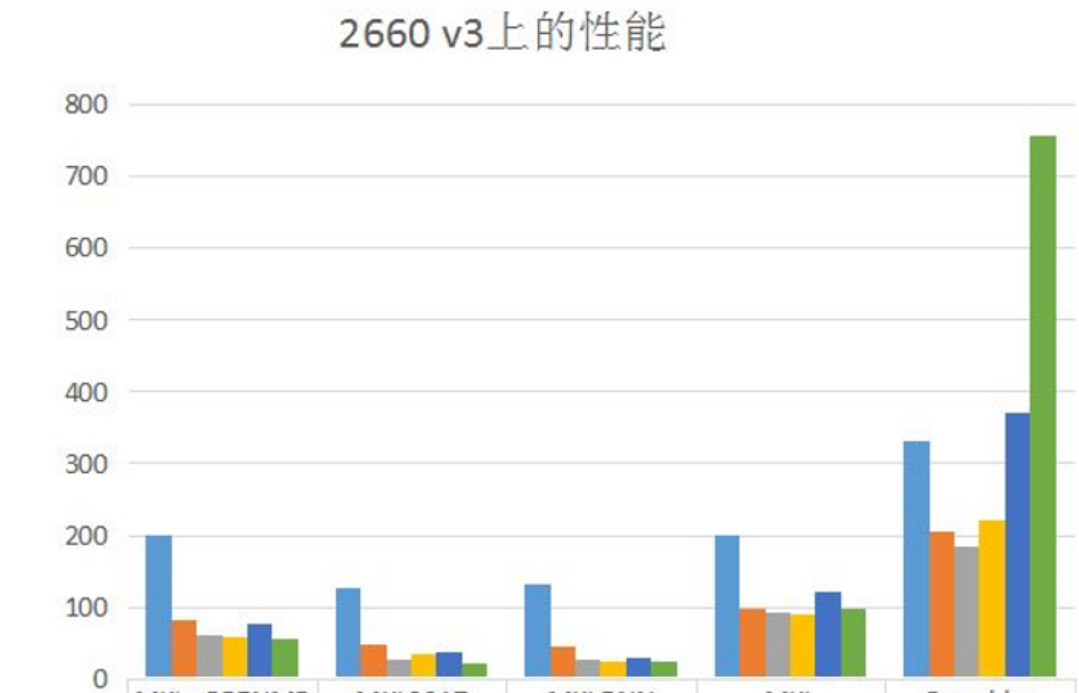
这些网络层在图像识别的发展过程中都起到了非常关键的作用，也极大加速了网络训练的收敛速度和预测效果，但是在工程的生产实践过程之中，很多网络结构在inference 的过程中显得非常冗余，也就是说他们可以和其他的操作进行合并。

1. scale 层和 batch norm 层的合并, 在 BN 层的介绍之中，在 BatchNorm 层将输入的特征分布转换均值为 0，方差为 1 的同分布之后，还需要连接一个 scale 层，让网络重新学习数据的分布，这在训练的过程中，确实确实很有效，但是在前馈的过程中，可以把 scale 层的参数和 batchNorm 中的除方差一步结合起来，从而减少计算量和多余的内存使用。

2. BN 层和卷积层的合并, 接着上面的优化方向，既然 scale 层可以和 BatchNorm 层，相互合并，那么同样的道理，我们将 batchnorm 层中储存的方差的值和卷集中的 weights 的值相互合并，将 batchNorm 中的 bias 除以 scale 之后和卷积中的 bias 相互合并，就完全可以在卷积层一层的实现过程中完成所有的计算，而不用再次使用 batchnorm 层。

3. 卷积层和 relu 层的合并,relu 层实质上就是一个符号函数，在卷积完成之后，使用使用一个符号函数就可以避免为 relu 层重新新建层。

2.3.3 本章小结



	MKL+OPENMP	MKL2017	MKLDNN	MKL	Openblas
1	199	125	130	199	329
4	81	47	44	96	205
8	59	25	26	91	182
16	57	33	24	88	220
32	76	36	28	119	369
不设置	55	20	24	96	755

图 2-9 MKL、MKL2017、MKLDNN、openblas 加速方法的具体速度

第三章 人脸多属性属性识别的架构

本章主要介绍人脸属性识别任务数据库和一些人脸属性识别中常用的方法，并且根据这些方法的弱点和问题，提出改进方法改进并进行实验。包括人脸属性中输入图片对于识别效果的影响，改进网络结构对于属性性能的提升，设计面向多数据分布和多属性分类的神经网络框架、以及对于网络输出置信度模块的构建。

3.1 人脸属性性质分析

3.1.1 人脸属性的类别

人脸属性的数据标注的环境各有不同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）, 其中标签往往具有很多种表示和性质，比如相对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）但总体分为有序性与无序性，整体性与局部性等，具体包括：

1. 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。;

2. 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）

3. 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;

4. 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。

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。

3.1.2 多属性标签表示形式

在训练的过程中，人脸属性通常以分类或者回归问题的形式出现，但是在多属性识别的任务中，通常使用标签编码或者多标签回归的方式。

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。

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。

3.1.3 属性之间的相互联系

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。

3.2 人脸属性数据库简介

这一章主要对于对于具体的数据库进行介绍：**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。

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。

LFWA：LFWA 是另一个无约束的人脸属性数据库，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），以及与 CelebA 数据库中相同的 40 个属性注释。

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。LAPAge2015 是一个无约束的脸部数据库，用于在 ICCV 2015 上发布的视在年龄估计。该数据库包含 4,699 张脸部图像，每

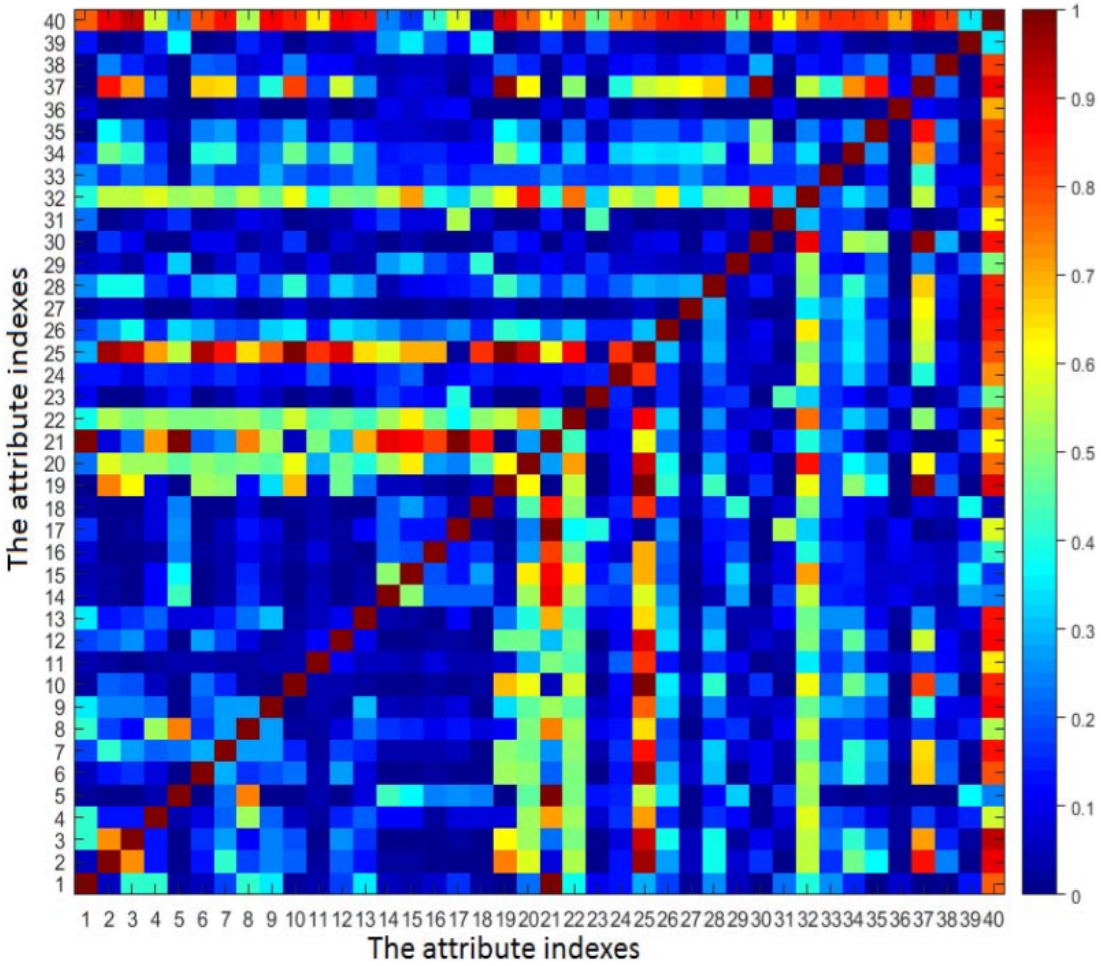


图 3-1 基于共享神经网络特征和 SVM 分类器的人脸属性识别

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。由于年龄信息的测试不可用，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。

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

表 3-1　celeA 中的属性表

属性序号	属性	属性序号	属性
|1|	5OClockShadow	|21|	GrayHair
|2|	Male	|22|	Sideburns
|3|	ArchedEyebrows	|23|	BigLips
|4|	MouthSlightlyOpen	|24|	Smiling
|5|	BushyEyebrows	|25|	BigNose
|6|	Mustache	|26|	StraightHair
|7|	Attractive	|27|	Blurry
|8|	NarrowEyes	|28|	WavyHair
|9|	BagsUnderEyes	|29|	Chubby
|10|	NoBeard	|30|	WearEarrings
|11|	Bald	|31|	DoubleChin
|12|	OvalFace	|32|	WearHat
|13|	Bangs	|33|	Eyeglasses
|14|	PaleSkin	|34|	WearLipstick
|15|	BlackHair	|35|	Goatee
|16|	PointyNose	|36|	WearNecklace
|17|	BlondHair	|37|	HeavyMakeup
|18|	RecedingHairline	|38|	WearNecktie
|19|	BrownHair	|39|	HighCheekbones
|20|	RosyCheeks	|40|	Young

MORPH II 数据库，其他数据库主要包含真实场景下的人脸图像。

可以看到人脸属性的数据集其实比较庞大，如果都能够充分利用，可以获得与使用单一数据集更加出色的识别效果。但是每个数据集之间的数据不同，数量不同，标注不同实际使用中往往使用先训练一个再训练一个的流程，非常耗时而且不能够保证模型效果在原有数据集上保持良好的效果。实际是用来看，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。

类似的问题对于年龄这一属性更严峻一点，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。

3.3　基于传统特征的人脸属性识别

基于传统特征的人脸属性识别往往采用特征提取和分类器结合的方式，其中较为经典的是基于 DIF 特征的人脸属性识别是经典的属性学习方法，在 morphII 上一度取得了非常优秀的实验结果，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。具体结构见下图：

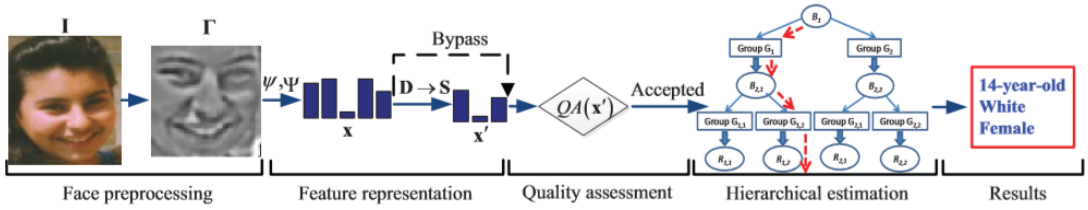


图 3-2　基于 DIF 特征的人脸属性识别

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。

基于 DIF 特征的属性识别方法是经典的基于传统特征和分类器的属性识别方法，即使在现在，特征融合、层级分类器建设等操作依然具有一定的借鉴意义

问题与不足：但存在一定的问题，例如，层级分类器确实能够提升分类的效果，但是复杂都明显过高，并不简洁，使用基于传统滤波器和表层信息的图像特征，需要大量的特征筛选和过滤工作。而且总结来讲是 DIF 系统还处在各个部分的分开设计，整个系统并不是处在一个整体性学习的状态。需要较多的人工干预和训练才能得到较好的效果。

3.4 基于共享神经网络特征和最大间隔分类器的人脸属性识别

随着深度学习方法的提出，深度学习的特征慢慢取代了传统的手工设计的特征，结合深度学习中经常使用的分类器，得到了更高的效果。具有代表性的是基于级联 CNN 网络和 SVM 分类器的识别方法，使用两个 CNN 框架 Lnet 和 Anet 进行级联学习，其中 Lnet 负责检测图片中的人脸，Anet 针对于 Lnet 中检测人脸使用交叉熵 loss 进行训练，为了提升识别的准确性使用 SVM 对于 ANet 中的特征进行训练。最后由 SVM 分类器输出具体的人脸属性预测值。其中不难发现器图片的标签就采用的是标签化编码的方式。

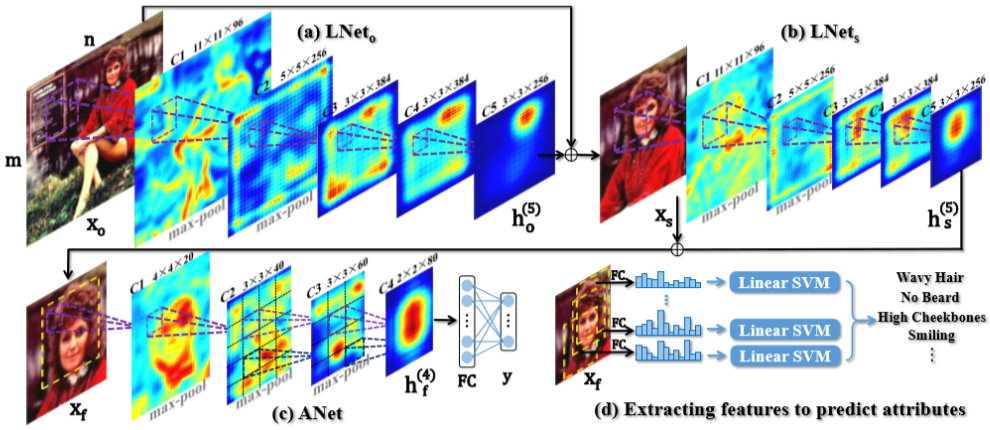


图 3-3 基于共享神经网络特征和 SVM 分类器的人脸属性识别

问题与不足：其实在训练的过程中，神经网络就已经可以对于属性进行预测，但是识别的效果却并不如 SVM 训练的结果，说明在这一框架中，神经网络对于不同属性的自网络决策层和 loss 函数设计的不够好。从图中可以看到，不同的人脸属性之间都是用的同样的 FC 层全连接而来的。不能够体现出属性整体和局部特性。

3.5　基于共享特征和子任务模块的端到端的人脸属性学习

机器学习神经网络中的端到端，一般是指输入原始数据，输出最后结果的过程。对于人脸属性的识别过程来讲：如何解决人脸属性多任务的输出是解决问题的关键，目前比较主流的方法是使用网络共享单元和网络子任务模块相互组合的方式，具体来讲可以参考下图：对于不同的人脸识别任务，如果希望能够在一个框架中通过端

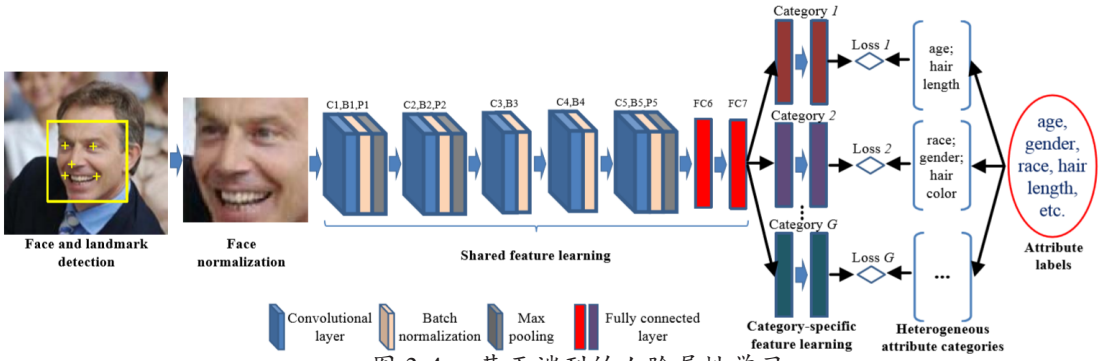


图 3-4　基于端到的人脸属性学习

到端的方式解决，而且具有一定可用性，就需要将不同属性识别任务中重复性学习的工作整合成共享模块，然后不同的人脸属性模块再基于共享模块单独对于自身的学习任务单独进行学习。端到端基于共享模块的学习方法可以很大程度上加快过程中的使用效率，同时简化了训练流程。而类别特定的子模块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，基于共享单元和网络子任务模块的方式在保持低计算代价的同时，实现了更高精度的属性估计精度，使其在许多人脸识别应用中具有价值。

问题与不足：基于共享模块和子网模块的识别方式虽然看上去简化了识别的过程，但却是以减少了模型训练的先验规则为代价。也就是说把更多的学习过程交给了模型本身，那么模型的学习过程一方面取决于模型自身设计的学习能力和使用的训练方法，而更重要的是训练数据的选择和实用。

尤其对于端到端网络来讲，由于整体网络结构变得封闭和训练方式的固定化，训练数据的选择和训练过程中对于真实环境的模拟就变得尤为重要。但真实场景中的数据具有较大的变化，包括数据的场景来源不同，数据中样本的比例不均匀，数据自身的姿态和角度变化等。训练样本的选取和预处理过程成了和网络结构选择一样重要的问题。

表 3-2　SA-softmax 的置信度判断对照表

无法识别类概率	模型自评置信度
80-100	判定完全无法识别
60-80	判定难以识别
40-60	判定可以猜测但不对结果负责
20-40	判定基本正确
0-20	判定模型输出非常自信

3.6　使用 SA-sfotmax 进行模型稳定化输出

模型的稳定性输出体现在很多方面，如常见的连续变化的视频，不同设备收集的图像等。在这些场景中，网络可能偶尔会发生判断错误的情况，影响用户的体验。虽然很多时候可以采用平滑的策略进行弥补，但是网络本身的稳定性性能是整个识别系统的关键。为了提升属性识别模型的稳定化输出，我提出了基于带有自评功能的 softmax（Self-assessment softmax SA-softmax），以及相应的动态标签（Dynamic tag，Dt）训练方法来增强网络的稳定性。

3.6.1　SA-softmax 的改进形式

传统的 softmax with loss 面对 N 个输入类别，或有 N 元数字作为输入，同时输入的标签是从 0 N-1 的一个整形数字用以表示具体的 ground truth。通过 softmax 操作，计算出每个类别的相对概率值。将对应标签的概率输出取负对数作为损失函数和优化的对象。

SA-softmaxLoss 在分类结果上将网络分类概率输出由 N 个，改为 N+1 个，第 N+1 维值定义为无法识别类（unrecognizable class）。

分类 label 由原来的一元数字标签变为一个四元的整形数组分别用于储存图片在过去训练过程中出现的次数、图像被判正确的次数、图像真实的标签，图像目前的标签。

训练的过程中使用动态标签（Dynamic tag，Dt）训练方法，不断更新训练图片的类别。

测试的过程中，对于具体的结果按照无法识别概率的大小，对于网络的置信度进行评估，而每种置信度的情况下，会对其余的类别重新使用 softmax 进行概率获取，以便输出。

SA-softmax 其他的设计包括：

3.6.2 动态调整标签的训练方法

对于 softmax 简单的加入一类未知类，还远远不够，至少训练的样本都没有，这分支的梯度永远都是 0，不会对网络产生训练作用。于是在现有的框架下，我们设计了如下的样本生成策略：1. 首先需要在不加入未知类的模式下，将分类网络训练至收敛。2. 将训练收敛的网络模型的 softmax 改成 SA-softmax，即输出的预测类别增加一个未知类 3. 正常训练 10 个 epoch，在这个过程之中，每次都更新训练图片的标签四元组中的出现次数和判断正确次数。4. 在图片出现的第 11 次，验证前 10 次确率是否低于 0.5，并将图片的现有标签改为 N+1 类，同时清空出现次数和被判断正确的次数。5. 之后图片每被训练 10 次，检查图片现有标签和实际标签是否一致，如果一致且训练 10 次准确率高于 0.5，那么说明被模型学习的很好；如果一致但训练 10 次准确率低于 0.5，将图片标签标记成 N+1 类；如果不一致且准确率低于 0.5，将模型的标签改回原来的标签；如果不一致但准确率高于 0.5，说明这类图片确实无法识别，模型认识到自己的局限性。

可能出现的状况有：模型不能够将经常判错的图片归并到无法识别类，导致图片的标签总在原有标签和无法识别类之间摇摆。解决办法是将无法识别类的分类 loss weight 提高，也就是提高其对应的后馈比重，对其进行强行分类。

但这种做法对于标签标错的正常样本并不能产生作用，所以每次检查图像标签在未识别类的时候都会把图片单独存放起来，方便检查

3.7 实验设置

3.8 人脸属性识别的技术改进

在上一节中，总结了在人脸属性任务的主流方法和发展过程，可以看出随着深度学习的发展和端到端学习在模式识别过程中的发展，很多问题都得到了改善，但依然存在着一些主流问题和场景困境一直存在，我仔细对于相关问题进行了思考总结。

1. 问题 1：人脸数据库的标注各不相同，使用怎样的框架才能将不同的数据库都充分利用起来。

2. 问题 2：使用怎样的数据输入和数据处理方式和训练方式，才能发挥深度学习的特性。

3. 问题 3：提高模型输出的稳定性，减轻网络错误输出的偶然性和影响。

为了解决上述问题，设计了以下策略来解决:

3.8.1 数据集并行训练的方式

对于不同的数据集来讲，预处理的方式往往类似，也就是说输入图片虽然不同，但是输入的格式是一致的（事实上即使不同也没有什么问题，关键是图片数据不同），但是因为标签不同，导致在一个网络结构中无法进行统一的训练。那么不妨就按照多个单独的网络对于图像进行训练，每个网络在特征提取阶段采用相同参数的全卷及网络结构进行特征提取，但每一层的特征图会单独进行存储，训练的时候每个数据集都按照自己的数据结构特性为了拟合 loss 层的设计，采用不同的子网络结构。（画图）

3.8.2 人脸矫正固定输入格式和数据增强扩充样本

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。（加入变换公式）

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：

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）

简单介绍一下人脸矫正的过程：

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）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。在具体介绍人脸属性的任务过程中，首先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。

3）人机性能对比

人机性能比对是指和实际的人眼评估进行比较，这也是人脸识别领域中经常使用的指标。在人和机器的性能对比过程中，发现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。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。

第四章 对抗生成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

早在 2014 年，人们对于神经网络技术的研究非常狂热的同时，也有一部分理智的科学家认为神经网络的输出判断具有非常高的风险，输出具有非常高的不稳定性，所谓数据集上的准确率超越人类不过是一场谎言，为了戳穿这一谎言，他们在神经网络判断正确的图片上简单加了一些噪声，对于人类来说根本没有察觉图像的变化，但是在神经网络却完全将其判断成另外一种物体。同时科学家们宣称这样极具欺骗性的图片并非偶然得到，而是可以量产的，比如通过 GAN 网络。

借助于博弈论中的零和博弈思想（在零和博弈中，游戏玩家之间的利益总和是固定的，即一方获得收益，另一方就要承担损失。）Goodfellow 极具想象力的提出了可以通过搭建两个对抗的网络，各自的目的就是降低对方的准确率，或者说提升对方 loss。通过这样非常具有竞争性的训练过程，最够提升两个网络的性能。具体来讲：在对抗生成网络中，玩家的角色会分别有生成模型 (generative model) 和判别式模型 (discriminative model) 充当. 生成模型 G 捕捉样本数据的分布，判别模型 D 是一个二分类器，估计一个样本来自于训练数据（而非生成数据）的概率。G 和 D 可以是线性代数的算法操作组合，也可以是神经网络的网络模型，都可以理解成或者定义成非线性函数。通过不断调整 G 和 D，直到 D 不能把事件区分出来为止。在调整过程中，需要：优化 G，使得它尽可能的让 D 混淆优化 D，使得它尽可能的能区分出假冒的东西当 D 无法区分出事件的来源的时候，可以认为，G 和 M 是一样的。从而，就获得了能够以假乱真的数据。

而在不断的发展中 GAN 网络有了更多的应用和算法分支

4.1.1 非监督图片的生成

4.1.2 图片超像素

4.2 对抗神经网络在人脸属性中的应用

从上面对抗生成网络的演变和发展来看，对抗生成网络很明显并不能像正常的 CNN 网络一样对于具体的模式识别任务，但是作为探究 CNN 生成原理的一部分，对抗生成网络主要是希望能够了解 CNN 能够从图像中学习到什么样的信息，怎样学习

的，并且能否以较为直观的形式也就是生成图像来表示出来，（尽管学习到的东西很多时候并不能够以图像的形式进行展现）。在本论文的实验后期，我们想到了使用对抗生成网络来探究一下 CNN 对于不同分布的数据集之间的学习能力。

4.2.1 人脸属性的监督式学习困境

首先引入一个经典的模式识别场景，泛化能力的问题: 在 Hu 的工作种，他发现一个在很多实验中都会出现的问题，使用 MTL 的人脸属性框架进行人脸属性识别的过程中，具同样 40 个属性标签的两个数据集 lfwA 和 celeA，两个在各自数据集上训练之后的模型，在各自数据集上的准确率都很高，但是在对方的测试集效果都比较糟糕。（加入 hu 的 lfw 和 celeA 的实验对比表）如何进行改善呢? 我们针对于这种情况设计了这样的思路：问题引出：对于相同的网络模型，使用相同的训练方法，在不同数据集中的训练之后，对自身数据集的测试集准确率要远远高于其他数据集的测试集。问题分析：首先这不是一个过拟合问题，因为对于数据集中训练集和测试集的准确率较高，所以网络的训练没有问题。但是对于不同数据集的测试集准确率很低，所以推测问题的出现是因为数据的分布不同

尝试解决办法：首先我们先假定网络模型容量可以容纳两个数据的分布（数据的分布可能不满足线性加法，但是应该满足集合性合并不减的特性，所以假定两种数据的分布集合会比原来更大，所以对于网络容量的要求会更大），既然数据的分布不同，就应当减少数据分布对于模型训练带来的影响。

第一种方法就是将两个数据集合并训练，如果标签相同，那么可以简单的将两个数据集合并成一个数据集训练，也可以首先在一个数据集上份训练，再经过另一个数据集 finetune，又或者采用上一章所提到的主干网路参数共享，不同数据集分别使用一个网络支线进行训练。都可以直观地学习到两个数据集之间的数据分布。往往就可以取得较好的效果，有效的提高在不同数据集上准确率的表现。缺点：最致命的缺点就在于不同数据集的准确率提高，但是难以保证在自身的数据集上数据的准确性。即使采用较小的学习率谨慎的进行 finetune, 对于不同任务的训练过程也即将面临着大量的手动干预，还是处于一个监督学习的框架之中。对此我们决定使用类似于迁移学习的方式来完成这个任务，并且结合 gan 网络来完成我们的任务。具体思路是这样的：从上面对于 gan 网络的介绍中可以发现，

4.2.2 使用 GAN 网络生成训练数据以扩充监督式学习方法

在之前对于 GAN 网络的介绍中，可以发现 GAN 网络最初是来证明神经网络算法对于数据分布具有一定的局限性。而慢慢发展，人们并不在乎神经网络是否对于数据分布有一定的的局限性，而狂热的希望能够通过 GAN 网络获得以假乱真的机器生成图片。似乎人们觉得如果机器能创造他，那机器肯定可以了解他，那么识别他也是轻而易举。于是乎这种炫酷，但是有一定投机取巧性质的思路不仅开始影响最初使用 GAN 网络探究神经网络有效性的本意，也影响着各种识别任务的传统数据 + 模型的预测方式。针对于这种非常具有诱惑力的尝试方式，即通过使用噪声实现对于特定图像的无标注成本转换，我们制定了如下的网络结构

首先实现了从 100 维的噪声生成数字图像，

然后实现了 32*32 的物体图像，

甚至实现了具有一定属性人脸图像。

但尽管证明 GAN 网络确实能够自动的生成具有一定真实图片特征的图片但是生成图片的效果还是没有较好的方式来进行控制，远远不能达到以假乱真的程度。与此同时使用 GAN 网路生成图片，然后加入训练数据中的思想，还是局限于监督学习的框架，需要具有标注的网络图像。

4.2.3 结合 GAN 超像素实现迁移学习

在发现 GAN 网络其实并不能直接从噪声生成具有一定训练意义的图片之后，我们并没有气馁。在参考了很多具有使用意义的 GAN 网络工作之后，决定从超像素的方向重新研究。

第五章 总结与展望

5.1 全文总结

通观全文，与其说实在研究如何提高人脸属性的识别准确率，倒不如说是在各种偏离基础算法使用场景的情况下，解决一个又一个出现的问题，包括为了能够提高训练速度，在训练中的不同框架，尝试探究多机多卡。为了在实际测试中，具有较高的反馈和实用价值，在基础的神经网络操作中，对于基本算法的加速和改进。为了适应针对数据集训练和评测的这种模式，设计针对于多种属性标签，多种数据集的网络架构。为了对于不同的场景数据分布存在偏差的问题，针对于背景的变化，使用gan网络对于人脸图片进行了所谓的场景人脸重构。在这个过程中，不仅对于模式识别的基本算法有所掌握，同样也印证着发现问题，分析问题和解决问题的思路。从学术研究的贡献来看，其实并没有提出提别的足够具有改变行业性的算法，更像是对于现有算法更加深入和具体化的改进。从整个研究任务的完成上，主要的思考点和实现的标准有两方面：一方面从底层实现上探究在目前计算机水准上，热衷于探究对于算法的实现能够具有怎样的加速方法，让算法的实现和使用变得快速化。另一方面在常规的网络构建上思考如何能够构建更加具有端到端的特性，让机器学习的相关问题分析和解决流程变得更加简洁化。仅此而已。

总体自我评价来看，还需要更多

5.2 未来展望

本文所介绍的人脸属性识别属于图像识别种基于监督学习的分支，同时也是非常具有代表性的任务之一，类似的人物包括物体识别中的物体性质识别，如经典的鱼种类识别等。所以人脸属性识别的进展需要依托于整个图像识别的基础技术进展和图像数据库的建设。而图像识别领域基础技术的进展其实更加依托于更加基础计算机科学的进步与发展，细节小到晶体管的制造工艺，计算机内存和缓存的读取速度，处理器的主频提升，布局大到整个体系结构的变革，冯诺依曼体系的变革，量子计算机的进化等，都会对于模式识别算法有着较为深远的影响。

除了对于底层科学的依赖，现实生活中的应用落地也同样具有重大意义，比如慢慢成熟的人脸识别，自动驾驶等新兴技术行业，无一不是有基础的模式识别技术

发展而来，但是却无一不在现实的产业结构中引发巨大热潮，让实验室的算法走出实验室出现在人们的现实生活种，可以极大激励人们对于人工智能的探索的热情和改变人类生活状态的前行动力。并且在实际生活中慢慢探索图像识别的规律，加速人工智能领域的快速发展。

致　谢

感谢自己，感谢所有人.

攻读学位期间发表的学术论文目录