人脸验证DeepID总结

**摘要：**

如今，深度学习方兴未艾，大数据风起云涌，各个领域都在处于使用深度学习进行强突破的阶段。人脸识别也不例外。香港中文大学的团队使用卷积神经网络学习特征，将之用于人脸识别的子领域人脸验证方面，取得了不错的效果。连发三箭，皆中靶心，使用的卷积神经网络已经改进了多次，破竹之势节节高。在这里将DeepID神经网络的几代进化史进行总结与实验重现，学习其中的精髓。

DeepID所应用的领域是人脸识别的子领域----人脸验证，就是判断两张图片是不是同一个人。人脸验证问题很容易就可以转化成人脸识别问题，人脸识别就是多次人脸验证。DeepID达到的效果就是在LFW数据集上（但是训练过程由于LFW数据集同一个identities都只有一张人脸，5000个人只有13000左右张图片，所以选用的是CelebFaces和CelebFaces+的dataset，其中每张脸的数据集不止一个），该数据集是wild人脸数据集，即没有经过对其的人脸，背景变化比较大。

卷积神经网络在DeepID中的作用是学习特征，即将图片输入，学习到一个160维的向量。然后在这个160维的向量上，套用各种现成的分类器，即可得到结果。DeepID之所以如此有效，首先在于通过卷及神经网络学习到的特征泛化能力较强，为了得到比较强的结果，DeepID采用了目前最常用的方法----增大数据集，只有大数据集才能使卷及神经网络训练的更加充分。增大数据集有两种方法：第一种就是采集好的数据，即CelebFaces数据集的引入。第二种是将图片多尺度多通道多区域的划分，再把得到的向量连接起来得到最后的向量。

DeepID总体的算法流程：

1. 首先使用CelebFaces的数据集训练一个卷积神经网络
2. 利用这个卷积神经网络计算数据的DeepID
3. 利用计算得到的DeepID通过分类器将其分类完成验证

在实际的算法运行过程中，数据集分成4:1份，4份用来训练网络，1份用来验证数据集，然后1那份训练分类器，因为不能使用同一份数据进行训练，以免产生over-fitting。

**一、神经网络，深度神经网络与卷积神经网络**

神经网络技术起源于上世纪五、六十年代，当时叫感知机（perceptron），拥有输入层、输出层和一个隐含层。输入的特征向量通过隐含层变换达到输出层，在输出层得到分类结果。早期感知机的推动者是Rosenblatt。但是，Rosenblatt的单层感知机有一个严重得不能再严重的问题，对于计算稍微复杂的函数其计算力显得无能为力。

随着数学的发展，这个缺点直到上世纪八十年代才被Rumelhart、Williams、Hinton、LeCun等人发明的多层感知机（multilayer perceptron)克服。多层感知机，顾名思义，就是有多个隐含层的感知机。我们看一下多层感知机的结构：

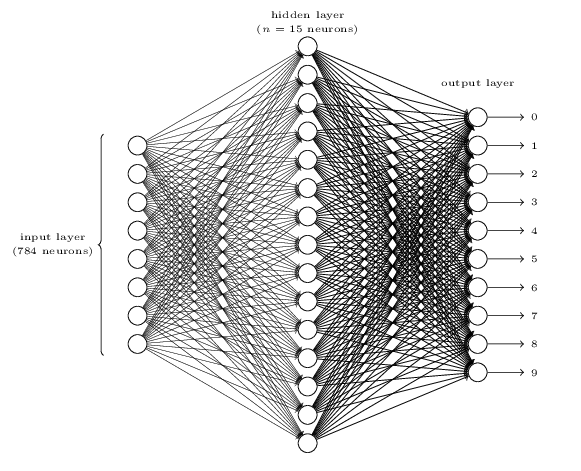


图1-1.多层感知机

多层感知机可以摆脱早期离散传输函数的束缚，使用sigmoid或tanh等连续函数模拟神经元对激励的响应，在训练算法上则使用Werbos发明的反向传播BP算法。对，这就是我们现在所说的神经网络( NN)！多层感知机解决了之前无法模拟异或逻辑的缺陷，同时更多的层数也让网络更能够刻画现实世界中的复杂情形。多层感知机给我们带来的启示是，神经网络的层数直接决定了它对现实的刻画能力——利用每层更少的神经元拟合更加复杂的函数。

即便大牛们早就预料到神经网络需要变得更深，但是有一个梦魇总是萦绕左右。随着神经网络层数的加深，优化函数越来越容易陷入局部最优解，并且这个“陷阱”越来越偏离真正的全局最优。利用有限数据训练的深层网络，性能还不如较浅层网络。同时，另一个不可忽略的问题是随着网络层数增加，“梯度消失”现象更加严重。具体来说，我们常常使用 sigmoid 作为神经元的输入输出函数。对于幅度为1的信号，在BP反向传播梯度时，每传递一层，梯度衰减为原来的0.25。层数一多，梯度指数衰减后低层基本上接受不到有效的训练信号。

随着2006年Hinton利用预训练方法缓解了局部最优解问题，将隐含层推动到了7层[1]，神经网络真正意义上有了“深度”，由此揭开了深度学习的热潮。这里的“深度”并没有固定的定义——在语音识别中4层网络就能够被认为是“较深的”，而在图像识别中20层以上的网络屡见不鲜。为了克服梯度消失，ReLU、maxout等传输函数代替了 sigmoid，形成了如今深度神经网络的基本形式。单从结构上来说，全连接的深度神经网络和上图的多层感知机是没有任何区别的。值得一提的是，后来出现的高速公路网络（highway network）和深度残差学习（deep residual learning）进一步避免了梯度弥散问题，网络层数达到了前所未有的一百多层 ！

如图1-1所示，我们看到全连接深度神经网络的结构里下层神经元和所有上层神经元都能够形成连接，带来的潜在问题是参数数量的膨胀。假设输入的是一幅像素为1K\*1K的图像，隐含层有1M个节点，光这一层就有10^12个权重需要训练，这不仅容易过拟合，而且极容易陷入局部最优。另外，图像中有固有的局部模式（比如轮廓、边界，人的眼睛、鼻子、嘴等）可以利用，显然应该将图像处理中的概念和神经网络技术相结合。此时我们可以祭出题主所说的卷积神经网络CNN。对于CNN来说，并不是所有上下层神经元都能直接相连，而是通过“卷积核”作为中介。同一个卷积核在所有图像内是共享的，图像通过卷积操作后仍然保留原先的位置关系。卷积层之间的卷积传输的示意图如下：

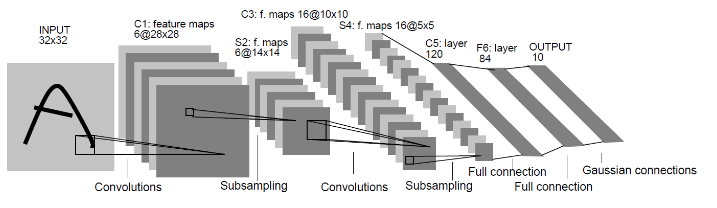


图1-2.卷积神经网络

通过一个例子简单说明卷积神经网络的结构。假设我们需要识别一幅彩色图像，这幅图像具有四个通道 ARGB （透明度和红绿蓝，对应了四幅相同大小的图像），假设卷积核大小为 3∗3 共使用100个卷积核w1到w100（从直觉来看，每个卷积核应该学习到不同的结构特征）。用w1在ARGB图像上进行卷积操作，可以得到隐含层的第一幅图像；这幅隐含层图像左上角第一个像素是四幅输入图像左上角 3∗3 区域内像素的加权求和，以此类推。同理，算上其他卷积核，隐含层对应100幅“图像”。每幅图像对是对原始图像中不同特征的响应。按照这样的结构继续传递下去。CNN中还有 max-pooling 等操作进一步提高鲁棒性。

我们注意到，对于图像，如果没有卷积操作，学习的参数量是灾难级的。CNN之所以用于图像识别，正是由于CNN模型限制了参数的个数并挖掘了局部结构的这个特点（与直觉相符，图像的局部相关性被利用）。顺着同样的思路，利用语音语谱结构中的局部信息，CNN照样能应用在语音识别中。在普通的全连接网络或CNN中，每层神经元的信号只能向上一层传播，样本的处理在各个时刻独立，因此又被成为前向神经网络(Feed-forward Neural Networks)。

**二、Deep Learning Face Representation from Predicting 10,000 Classes**

这个系列的第一篇文章，发表在2015年之前，但是为了整个报告的连贯和完整，仍然将其加入到报告中，因为本文是后面三篇文章的基础。首先将文章的主要内容做详细总结与介绍。

本文提出要通过深度学下来学习一套高级特征（简称DeepID）进行面部验证。文章中认为可以通过深入了解DeepID挑战性多样的面部识别任务，同时可以推广到其他任务（如验证）以及训练集中看不到的新身份。而且，DeepID的泛化预测能力随着脸部训练集的增加而增加。 DeepID特征取自深卷积网络（ConvNets）的最后隐藏层神经元激活。当作为分类器学习识别约10000个面孔身份训练集并配置为沿着特征提取层次不断减少神经元数量，这些深ConvNets通过少量隐藏神经元逐渐在顶层形成简介的身份相关的特征。提出的特征是从提取的各个部分的面部区域形成互补和过度完整表示。任何先进的分类器都可以基于这些高层次的表现来学习验证。 在LFW上的wild数据集上实现了97.45%的正确率。

因为本文的网络结构与普通的卷积网络结构相似，先重复一下普通卷积神经网络的要点。

1. 局部感知。卷积神经网络有两种神器可以降低参数数目，第一种神器叫做局部感知野。一般认为人对外界的认知是从局部到全局的，而图像的空间联系也是局部的像素联系较为紧密，而距离较远的像素相关性则较弱。因而，每个神经元其实没有必要对全局图像进行感知，只需要对局部进行感知，然后在更高层将局部的信息综合起来就得到了全局的信息。网络部分连通的思想，也是受启发于生物学里面的视觉系统结构。视觉皮层的神经元就是局部接受信息的（即这些神经元只响应某些特定区域的刺激）。如图2-1所示：左图为全连接，右图为局部连接。在右图中，假如每个神经元只和10×10个像素值相连，那么权值数据为1000000×100个参数，减少为原来的万分之一。而那10×10个像素值对应的10×10个参数，其实就相当于卷积操作。

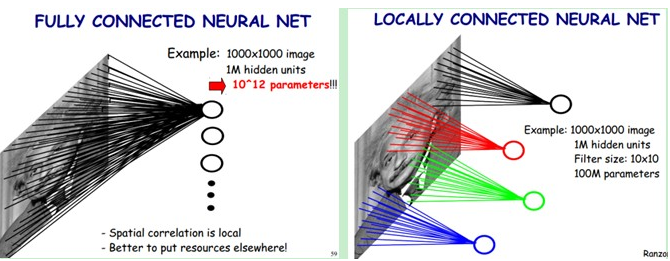


图2-1.局部感知

1. 参数共享。但其实这样的话参数仍然过多，那么就启动第二级神器，即权值共享。在上面的局部连接中，每个神经元都对应100个参数，一共1000000个神经元，如果这1000000个神经元的100个参数都是相等的，那么参数数目就变为100了。怎么理解权值共享呢？我们可以这100个参数（也就是卷积操作）看成是提取特征的方式，该方式与位置无关。这其中隐含的原理则是：图像的一部分的统计特性与其他部分是一样的。这也意味着我们在这一部分学习的特征也能用在另一部分上，所以对于这个图像上的所有位置，我们都能使用同样的学习特征。更直观一些，当从一个大尺寸图像中随机选取一小块，比如说 8x8 作为样本，并且从这个小块样本中学习到了一些特征，这时我们可以把从这个 8x8 样本中学习到的特征作为探测器，应用到这个图像的任意地方中去。特别是，我们可以用从 8x8 样本中所学习到的特征跟原本的大尺寸图像作卷积，从而对这个大尺寸图像上的任一位置获得一个不同特征的激活值。如图2-2所示，展示了一个3×3的卷积核在5×5的图像上做卷积的过程。每个卷积都是一种特征提取方式，就像一个筛子，将图像中符合条件（激活值越大越符合条件）的部分筛选出来。

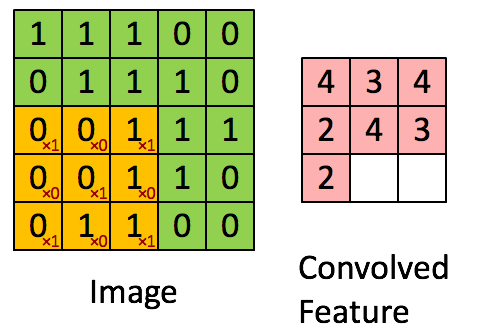


图2-2.参数共享

1. 多卷积和。上面所述只有100个参数时，表明只有1个10\*10的卷积核，显然，特征提取是不充分的，我们可以添加多个卷积核，比如32个卷积核，可以学习32种特征。在有多个卷积核时，如图2-3所示：

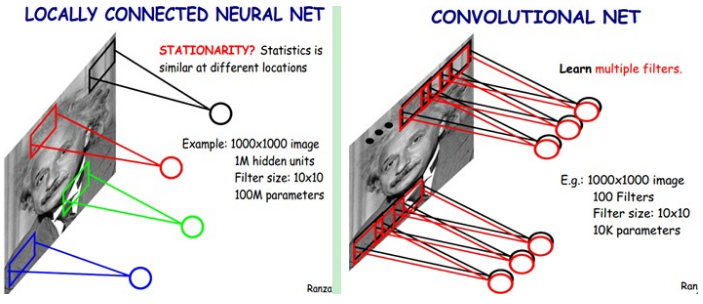


图2-3.多核卷积

右图中不同颜色表示通过不同卷积核生成的不同的图像的map，表示图像的不同特征。

1. Pooling。在通过卷积获得了特征 (features) 之后，下一步我们希望利用这些特征去做分类。理论上讲，人们可以用所有提取得到的特征去训练分类器，例如 softmax 分类器，但这样做面临计算量的挑战。例如：对于一个 96X96 像素的图像，假设我们已经学习得到了400个定义在8X8输入上的特征，每一个特征和图像卷积都会得到一个 (96 − 8 + 1) × (96 − 8 + 1) = 7921 维的卷积特征，由于有 400 个特征，所以每个样例 (example) 都会得到一个 7921 × 400 = 3,168,400 维的卷积特征向量。学习一个拥有超过 3 百万特征输入的分类器十分不便，并且容易出现过拟合 (over-fitting)。为了解决这个问题，首先回忆一下，我们之所以决定使用卷积后的特征是因为图像具有一种“静态性”的属性，这也就意味着在一个图像区域有用的特征极有可能在另一个区域同样适用。因此，为了描述大的图像，一个很自然的想法就是对不同位置的特征进行聚合统计，例如，人们可以计算图像一个区域上的某个特定特征的平均值 (或最大值)。这些概要统计特征不仅具有低得多的维度 (相比使用所有提取得到的特征)，同时还会改善结果(不容易过拟合)。这种聚合的操作就叫做池化 (pooling)，有时也称为平均池化或者最大池化 (取决于计算池化的方法)。形象的操作就是在一个二维矩阵上通过分块求Max降维。

文章中提出一种行之有效的方式提取出人脸图像中的高水平特征，处理流程大致如图2-4所示。

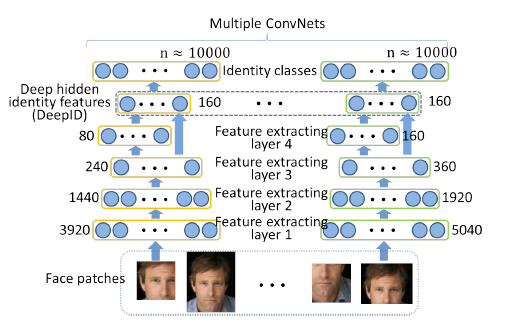


图2-4.DeepID处理流程

ConvNets学习将所有可用于由身份进行训练的面部进行分类，最后的隐藏层神经元激活作为特征（称为深度隐藏的身份特征或DeepID）。每个ConvNet都采用face patch作为输入，并提取底层中的局部低级特征。特征数量沿着特征提取级联不断减小，而在顶层中形成全局和高级特征。在级联结束时采集高度紧密的160维DeepID，其包含丰富的身份信息，并直接预测更大数量（例如10 000）的身份类别。同时分类所有身份而不是训练二进制分类器是基于两个考虑因素。首先，将训练样本预测为许多类中的一个，而不是执行二进制分类更为困难。这个具有挑战性的任务可以充分利用神经网络的超级学习能力来提取人脸识别的有效特征。第二，它隐含地增加了ConvNets的强正则化，这有助于形成可以很好地分类所有身份的共享隐藏表示。因此，学习的高级特征具有很好的泛化能力，不能超过一小部分的训练面。我们约束的DeepID比他们预测的身份类别显着更少，这是学习高度紧凑和歧视性特征的关键。我们进一步连接从各种面部区域提取的DeepID，以形成互补和过分的代表。学习的特征可以很好地推广到测试中的新身份，这在训练中看不到，并且可以容易地与任何最先进的面部分类器（例如，联合贝叶斯分类）集成以进行面部验证。

文章的方法仅使用弱对齐面即可实现LFW面部验证精度达97.45％，几乎与人类表现97.53％相当。我们也观察到随着训练身份数量的增加，验证性能稳步提高。虽然训练阶段的预测任务变得更具挑战性，但学习特征的歧视和泛化能力却增加。通过更多的训练数据找到一个方法，以便将来提高准确性。

具体的网络结构如图2-5所示。

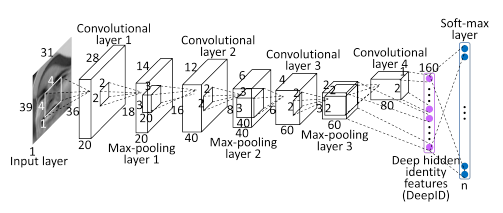
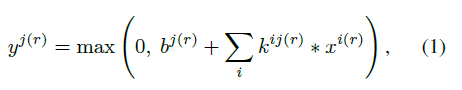


图2-5.DeepID网络结构

该结构与普通的卷积神经网络的结构相似。整个深层ConvNets包含四个卷积层（最大池）以分层提取特征，其次是完全连接的DeepID层和指示身份类的softmax输出层。图像patch的输入为39×31×k，图像块为31×31×k，彩色图的k = 3，灰度图为k = 1。图2-5显示了ConvNet的详细结构，该结构采用39×31×1输入并预测n（例如，n = 10 000）标识类别。当输入尺寸变化时，接下来网络层中的map的高度和宽度将相应改变。 DeepID层的尺寸固定为160，而输出层的尺寸根据其预测的类别数量而有所不同。特征数量沿着特征提取层次不断减少，直到形成高度紧凑和预测特征的最后一个隐藏层（DeepID层），其中仅使用少量但是具有大量信息的特征来预测更多的身份类。

卷积操作的表达式为

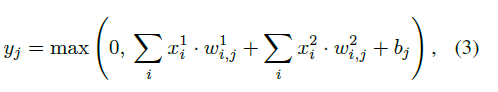


其中xi和yi分别是第i个输入图和第j个输出图。k ij是第i个输入映射和第j个输出映射之间的卷积核。\*表示卷积。b j是第j个输出图的偏置。我们使用ReLU非线性（y = max（0，x））隐藏的神经元，这显示出比S形函数更好的修复能力。文章的ConvNets的较高卷积层的权重在局部共享，以了解不同地区的不同中级或高级特征。等式(1)中的r表示共享权重的局部区域。在第三卷积层中，权重在每2×2区域中局部共享，而第四卷积层中的权重是完全非共享的。Max-pooling的公式如下：



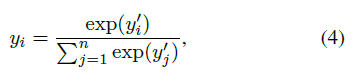
其中第i个输出图中的每个神经元yi在第i个输入图xi中的s×s个非重叠局部区域上。

DeepID的最后一个隐藏层完全连接到第三和第四卷积层（在最大池之后），使得它看到多尺度特征（第四卷积层中的特征比第三卷中的特征更global）。这对于特征学习至关重要，因为沿着级联的连续下采样之后，第四卷积层包含太少的神经元并且成为信息传播的瓶颈。在第三个卷积层（称为跳过层）和最后一个隐藏层之间添加旁路连接减少了第四个卷积层可能的信息丢失。最后一个隐藏层执行函数：



其中x 1，w 1，x 2，w 2分别表示第三和第四卷积层中的神经元和权重。它早期通过非线性ReLU组合了前两个卷积层的特征。

ConvNet输出是一个n路softmax，预测n个不同身份的概率分布。



其中将160个DeepID特征xi线性组合为神经元j的输入，yj为其输出。ConvNet通过使用第t个目标类来最小化log（yt）来学习。 随机梯度下降用于通过反向传播计算的梯度。

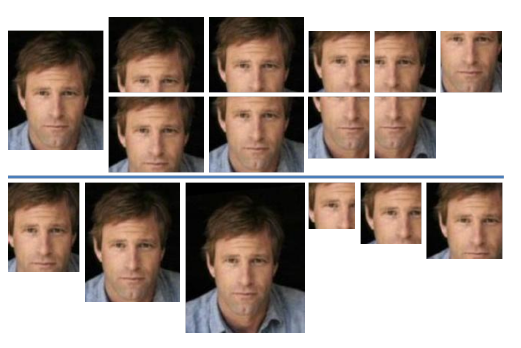


图2-6 特定面部区域

特征提取：文章通过Sun等人提出的面部检测方法来检测五个面部标志，包括两个眼睛中心，鼻尖和两个嘴角。根据两个眼睛中心和两个嘴角的中点，通过相似变换在global对齐脸部。特征是从10个区域，三个尺度和RGB或灰色通道的60个面部斑块中提取出来的。图2-6标出了两个特定面部区域的十个面部区域和三个尺度。我们训练了60个ConvNets，每个ConvNets每个从一个特定的patch提取两个160维的DeepID向量，并且它的水平Flip对应。一个特殊情况是两个眼睛中心和两个嘴角附近的斑块，它们不会自动闪烁，但是与它们对称的斑块（例如，以左眼为中心的补片的Flip对应物是通过以补片为中心在右眼）。 DeepID的总长度为19,200（160×2×60），可以进行最终的面部验证。

文章的面部验证使用在一些论文中表现的很成功的贝叶斯联合技术。提取的面部特征通过两个独立高斯变量表示：



我们用 μ 来表示identity，用 ε 表示intra-personal variation（包括：光线，姿态，表情等变化）。两个变量 μ 和 ε 分布服从两个高斯分布：N(0, Sμ) 和 N(0, Sε)。简要来说，上面的表达式包括附加的一些假设可以作为一个人脸的先验知识。

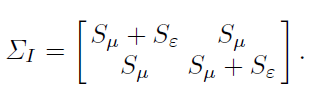
有了上面的先验知识，就可以得到一个均值为0的高斯联合分布{x1，x2}。若 μ 和 ε 是相互独立的，那么我们就可以得到两张人脸特征的协方差如下



1. 在 Hi的假设前提下：（两张人脸相同）

此时，μ1 和 μ2 是相同的，并且ε1 和 ε2 是独立的。

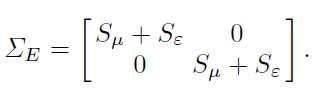
因此P(x1, x2|Hi)分布的协方差矩阵就可以按照如下计算：



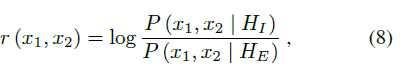
2. 在 He的假设前提下

此时，μ 和ε 都是独立的。

因此P(x1, x2|He)分布的协方差矩阵就可以按照如下计算：



Sμ 和 Sε在文章中只提到了使用EM算法可以进行求解。在此去google了下，大概就是期望最大化的一个算法具体选择使用的时候去详细了解即可。然后就可以计算似然比:



文章还提到训练了一个神经网络进行验证，并将其与贝叶斯联合进行比较，以了解其他模型是否也可以从提取的特征中学习，以及多少特征和良好的面部验证模型分别对性能有贡献。神经网络包含一个输入层，采用DeepID，一个局部连接的层，一个完全连接的层，以及一个表示面部相似性的单个输出神经元。输入功能分为60组，每组包含从特定ConvNet的特定补丁对提取的640个特征。同一组中的特征是高度相关的。本地连接层中的神经元只连接到一组特征，以了解其本地关系并同时减少特征维度。第二个隐藏层与第一个隐藏层完全相连，以学习global关系。单输出神经元完全连接到第二隐藏层。隐藏的神经元是ReLU，输出神经元是S形神经元。神经网络结构的图示如图2-7所示。它具有38400个输入神经元，每个pathc具有19200个DeepID特征，以及以下两个隐藏层中的4800个神经元，第一隐藏层中的每80个神经元本地连接到60组输入神经元。

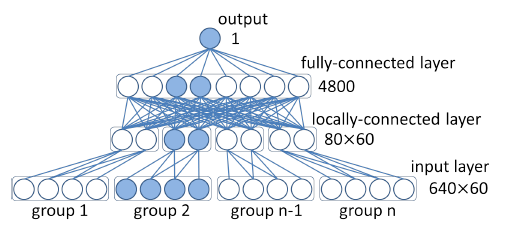


图2-7 分类神经网络

Dropout Learning被用于所有隐藏的神经元。输入神经元不能被删除，因为学习的特征是紧凑和分布式表示（代表具有极少数神经元的大量身份），并且必须彼此协作来表示身份。 另一方面，由于梯度扩散，学习不衰退的高维特征是困难的。为了解决这个问题，我们首先训练60个子网，每个子网具有单个组的特征作为输入。 一个特定的子网络如图2-7所示。然后，我们使用子网络的第一层权重来初始化原始网络的那些权重，并调整第一层权重被削减的原始网络的第二层和第三层。

文章的实验部分在LFW数据集上评估算法，揭示了弱对齐面部图像上面部验证的最新技术。虽然LFW包含5749个人，但只有85个拥有超过15个图像，4069个人只有一个图像。训练身份分类人员的人数不足，这是不够的。所以，作者在CelebFaces上培训了模型，并在LFW上进行了测试。 CelebFaces包含来自互联网的5436名名人的87,662张脸部图像，平均每人约16张。 LFW和CelebFaces中的人是相互排斥的。

文章随机选择Celeb-Faces的80％（4349）人学习DeepID，并利用剩下的20％的人学习面部验证模型（联合贝叶斯或神经网络）。对于特征学习，ConvNets被监督以从特定类型的面部patch和他们的Flipped对手同时分类4349人。我们随机选择每个培训人员的10％图像以生成验证数据。在每个训练时期之后，我们观察前1个验证集错误率，并选择提供最低的验证集合的模型。

在面部验证中，文章中的特征维度在PCA联合贝叶斯模型学习之前被减少到150。性能几乎保留在广泛的维度上。在测试中，通过将联合贝叶斯似然比与训练数据中优化的阈值进行比较，对每个面部进行分类。

而且验证了将第三卷积层（最大池）之后直接连接到最后一个隐藏层（DeepID层）的神经元的有效性，使得它看到第三和第四卷积层特征，形成所谓的多层卷积层，规模转换。它还导致从卷积层到DeepID层（如图2-1所示）的特征数量减少，这有助于后者学习更高级别的特征，以便以更少的神经元良好地表示面部身份。图2-8比较了60个ConvNets的顶级验证集错误率，可以将4349个身份类别进行分类，无论是否有跳过层。错误率越低，隐藏的特征越好。允许DeepID集中多种功能可以将验证错误平均减少4.72％。实际上，当将DeepID从60个ConvNets连接起来并使用Joint Bayesian进行面部验证时，它的最终面验证精度从95.35％提高到96.05％。

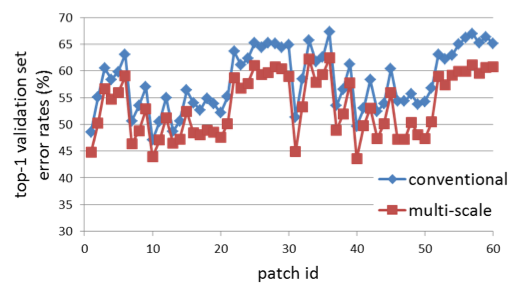


图2-8 不同方法的错误率

同时分类大量的身份是学习歧视性和紧凑隐藏特征的关键。为了验证这一点，我们增加身份等级训练指数（并相应地输出神经元数）从136到4349，同时固定所有以前的层中的神经元数（DeepID保持为160维）。我们观察到ConvNets（通过前1个验证集错误率测量）的分类能力以及通过增加的身份类别获得的面部验证（通过测试集验证精度测量）的学习隐藏表示的有效性。输入是在这个实验中覆盖整个脸的单个补丁。如图2-9所示，当身份类别加倍时，联合贝叶斯和神经网络在验证精度上线性提高，改进是显著的。

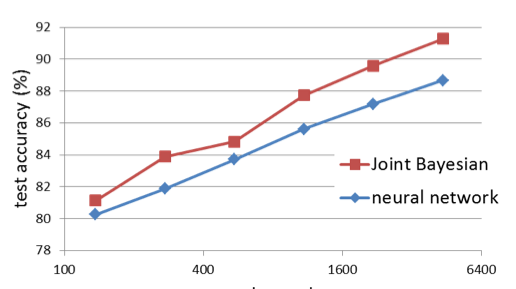


图2-9 训练集数目增大的趋势

当身份等级从136增加到4349次的32倍时，联合贝叶斯和神经网络的准确度分别提高了10.13％和8.42％，平均分别为2.03％和1.68％。同时，验证集错误率下降，即使预测的类是最后隐藏层神经元数十倍，如图2-10所示。这种现象表明，ConvNets可以从分类每个身份学习并形成共享的隐藏层的可以很好地分类所有身份的表示。更多的身分分类帮助学习更好的隐藏层的表示，可以区分更多的人（歧视），而不增加特征长度（紧凑）。相对于指数增长的训练数据，测试精度的线性增加表明，如果更多的身份可用，我们的特征将进一步改善。

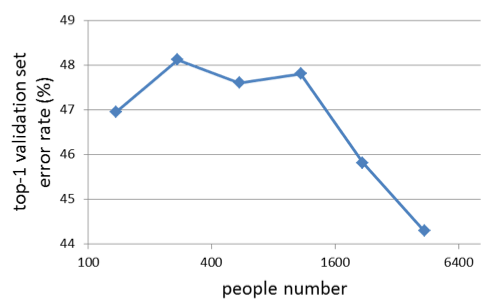


图2-10 测试精度

从4349训练身份中获取并从LFW测试对中提取的160维DeepID的示例显示在图2-11中。我们发现相同身份的脸部倾向于具有更常见的激活的神经元（积极特征位于相同位置）比起那些不同的身份。所以学习到的特征可以提取身份信息。

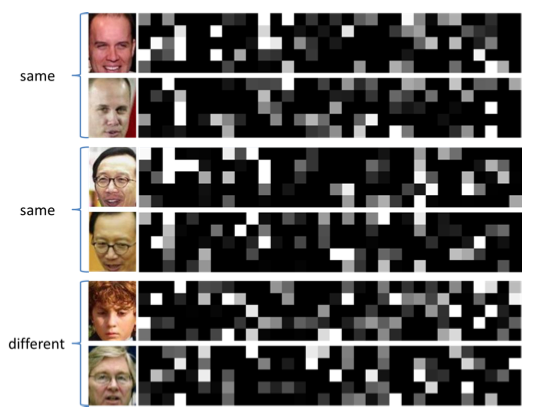


图2-11 不同的人的特征

我们还将4349维分类输出作为面部验证的特征进行测试。联合贝叶斯只能在这些特征上达到大约66％的准确度，而神经网络失败，其中将所有面部对作为正或负对。每个类都有这么多类和少量样本，分类输出是多样的和不可靠的，因此不能用作特征。

文章还评估了从各种面部斑块提取的多少组合特征将有助于性能。用k patches（k = 1,5，15，30，60）的特征训练面部验证模型。不可能计算所有可能的patch组合，所以选择最具代表性的patch。报告表现最好的single patch (k = 1), the global color patches in a single scale (k = 5), all the global color patches (k = 15), all the color patches (k = 30), and all the patches (k = 60)。如图2-12所示，从各个区域，尺度和颜色通道添加更多特征可以一直提高性能。组合60个patch分别比联合贝叶斯和神经网络的最佳单个patch提高了4.53％和5.27％。使用联合贝叶斯和神经网络分别达到96.05％和94.32％的精度。曲线表明，如果提取更多的特征，可以进一步提高性能。

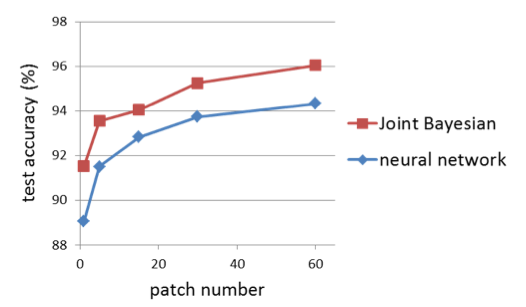


图2-12 k patch的实验结果

文章最后得到的结论：学习有效的高级特征，揭示面部验证的身份。这些功能是建立在深层ConvNets的特征提取层次之上的，并且从多尺度中级功能中总结出来。通过用少量隐藏变量表示大量不同的身份，获得了高度紧凑和歧视的特征。从不同面部区域提取的特征是互补的，并进一步提高了性能。它在LFW上获得了97.45％的正面验证精度，而仅需要弱对齐面部图像。

如果更多的身份可用于增加训练阶段的预测维度，则可以了解更加紧凑和歧视性的DeepID。作者希望有更大的训练集，进一步提升我们的表现。并将最近的此领域的工作做了简要比较和介绍。

关于本论文的重现相关的内容放到到报告的结尾，使用的深度学习框架caffe。有关caffe框架的介绍和使用同样放在报告结尾处。

**三、Deep Learning Face Representation by Joint Identiﬁcation-Veriﬁcation**

这个系列的第二篇论文，在第一篇的网络基础上进行了修改，提高了脸部验证的正确率。

文章认为脸部识别的关键问题是寻找到更加有效的特征表示，减少同一张脸的内部变化（由于光线，年龄等），扩大不同脸之间的差异。**文章提出了在深度学习中使用面部识别和验证两种信号作为监督来解决问题。**深层识别功能（DeepID2）通过精心设计的深卷积网络学习。面部识别任务通过从不同身份提取的DeepID2分开来增加个人间的差异，而面部验证任务通过从相同身份提取的DeepID2相结合来减少个人内部变化，这两者都是面部识别所必需的。学习的DeepID2的特征可以很好地推广到训练数据中看不到的新身份。在具有挑战性的LFW数据集上，实现了99.15％的面部验证精度。与LFW的最佳深度学习成果相比，错误率显着降低了67％。

显然当以不同的姿势，照明，表情，年龄和遮挡物呈现时，相同身份的面孔可能会有很大的不同。相同身份之间的这种变化可以压倒由于身份差异引起的变化，并使脸部识别具有挑战性，特别是在无约束条件下。因此，减少个人差异，同时扩大不同人之间的差异是人脸识别中的永恒话题。它可以追溯到早期的子空间人脸识别方法，如LDA，贝叶斯面和统一子空间。（补充，LDA：线性判别式分析(Linear Discriminant Analysis, LDA)，也叫做Fisher线性判别(Fisher Linear Discriminant ,FLD)，是模式识别的经典算法，它是在1996年由Belhumeur引入模式识别和人工智能领域的。性鉴别分析的基本思想是将高维的模式样本投影到最佳鉴别矢量空间，以达到抽取分类信息和压缩特征空间维数的效果，投影后保证模式样本在新的子空间有最大的类间距离和最小的类内距离，即模式在该空间中有最佳的可分离性。因此，它是一种有效的特征抽取方法。使用这种方法能够使投影后模式样本的类间散布矩阵最大，并且同时类内散布矩阵最小。就是说，它能够保证投影后模式样本在新的空间中有最小的类内距离和最大的类间距离，即模式在该空间中有最佳的可分离性。）然而，这些模型受到其线性或浅结构的限制，而个体间和个体间的变化是复杂的，高度非线性的，并且在高维图像空间中才能被观察到。

在这项工作中，文章提出深入学习提供了更强大的工具来处理两种类型的变体。由于其深厚的架构和较大的学习能力，通过分层非线性映射可以获得面部识别的有效特征。文章认为，通过同时使用两个监控信号（即面部识别信号和验证信号）学习这些特征是非常重要的，而学习的特征被称为深度识别验证特征（DeepID2）。识别是将输入图像分类为大量的身份类别，而验证是将一对图像归类为属于相同的身份（即二进制分类）。在训练阶段，给出具有识别信号的输入面部图像，其DeepID2特征在学习层次非线性特征表示的顶层隐藏层中提取，然后通过另一个函数g（DeepID2）映射到大量身份中的一个）。在测试阶段，学习的DeepID2功能可以推广到其他任务（如面部验证）和训练数据中看不到的新身份。识别监控信号倾向于将DeepID2分解为不同的标识，因为它们必须被分类到不同的类中。因此，学习的功能将具有丰富的身份相关或个人间的差异。然而，识别信号对从相同身份提取的DeepID2具有相对较弱的约束，因为不同的DeepID2可以通过函数g映射到相同的身份。当DeepID2功能被推广到新任务和测试中的新身份时，会导致问题，其中g不再适用。我们通过使用额外的面部验证信号来解决这个问题，这要求从相同身份提取的每两个DeepID2向量彼此靠近，而从不同身份提取的那些被保留。对DeepID2强大的每个元素的约束可以有效地减少个人内部的变化。另一方面，单独使用验证信号（即，一次仅区分一对DeepID2）在使用识别信号（即，一次区分数千个身份）时提取身份相关特征方面并不有效。因此，这两个监督信号强调特征学习中的不同方面，应该一起使用。

为了从不同方面对脸部进行表征，文章从各种面部区域和分辨率中提取互补的DeepID2特征，并且在PCA尺寸减小之后连接形成最终特征表示。

PCA（Principal Component Analysis）是一种常用的数据分析方法。PCA通过线性变换将原始数据变换为一组各维度线性无关的表示，可用于提取数据的主要特征分量，常用于高维数据的降维。降维当然意味着信息的丢失，不过鉴于实际数据本身常常存在的相关性，我们可以想办法在降维的同时将信息的损失尽量降低。

总结一下PCA的算法步骤：

设有m条n维数据。

1）将原始数据按列组成n行m列矩阵X

2）将X的每一行（代表一个属性字段）进行零均值化，即减去这一行的均值

3）求出协方差矩阵C

4）求出协方差矩阵的特征值及对应的特征向量

5）将特征向量按对应特征值大小从上到下按行排列成矩阵，取前k行组成矩阵P

6）Y=PX即为降维到k维后的数据

由于学习到的DeepID2功能在不同身份之间是多样化的，而在同一身份内是一致的，所以使得脸部识别变得更加容易。使用学习的特征表示和最近提出的面部验证模型，在具有挑战性和广泛研究的LFW数据集上实现了最高的99.15％的面验证精度。这是第一次提出只有脸部区域验证其精度与人类的99.20％的准确度相一致的机器，对于包括脸部区域和大背景区域在内的整个LFW脸部图像来说，它们被准确验证。

那么如何在卷积神经网络中中使用这两种监督信号就成为了主要问题。使用变化的深度卷积神经网络（Deep ConvNets）来学习特征。深层ConvNets中的卷积和pooling操作专门用于从局部的低级特征到global高级别层次地提取视觉特征。我们的深层ConvNets采用与上篇论文中相似的结构。它包含四个卷积层，其中第一个是最大池。为了学习多种高级功能，我们不需要在较高卷积层的整个特征图上进行权重分配。特别地，在我们深层ConvNets的第三卷积层中，神经元权重在每2×2个局部区域中被局部共享。在更合适地称为局部连接层的第四卷积层中，权重在神经元之间完全不共享。 ConvNet在特征提取级联的最后一层提取160维DeepID2向量。要学习的DeepID2层完全连接到第三和第四卷积层。由于第四卷积层提取出比第三卷卷更多的全局特征，所以DeepID2层采用多尺度特征作为输入，形成所谓的多尺度ConvNets 。我们对卷积层和DeepID2层中的神经元使用整流线性单元（ReLU）。 ReLU具有比大型训练数据集的S形模型更好的能力。用于提取DeepID2的ConvNet结构的图示如图3-1所示，给定大小为55×47的RGB输入。当输入区域的大小改变时，后面图层中的map大小将相应地改变。 DeepID2提取过程表示为f = Conv（x，θc），其中Conv是ConvNet定义的特征提取函数，x是输入面patch，f是提取的DeepID2向量，θc表示ConvNet参数要学习。

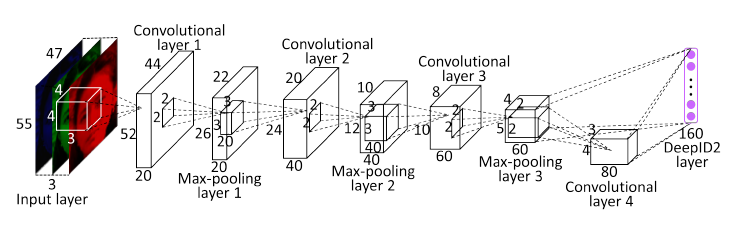
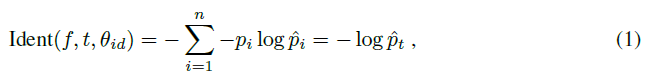
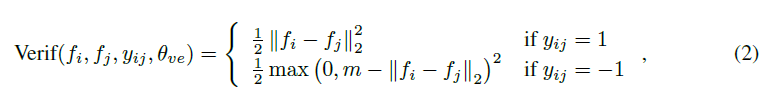


图3-1 DeepID2的学习网络

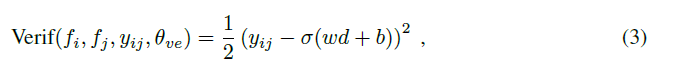
DeepID2功能在两个监控信号下学习。 第一个是面部识别信号，其将每个脸部图像分类为n（例如，n = 8192）个不同身份之一。 通过使用n-way softmax层跟随DeepID2层来实现识别，该层输出n个类的概率分布。 训练网络通过最小化交叉熵损失，我们称之为识别损失来实现。 它被表示为



其中f是DeepID2向量，t是目标类，θid表示softmax层参数。p i是目标概率分布，其中对于所有i，其中p i = 0，除了目标类t的p t = 1。 p i是预测的概率分布。 为了正确地对所有类进行同步分类，DeepID2层必须形成区别性身份相关的特征（即具有大的个人间变化的特征）。第二个是面部验证信号，它鼓励从相同身份的脸部提取的DeepID2是类似的。验证信号直接规范了DeepID2，可以有效降低个人差异。常用的约束包括L1 / L2范数和余弦相似度。我们采用Hadsell等最初提出的L2规范的以下损失函数以降低维数

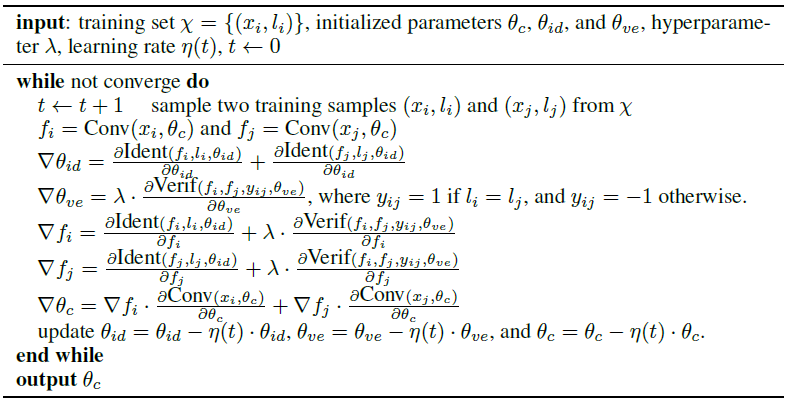


其中f i和f j是比较从两个脸部图像提取的DeepID2向量。y ij = 1表示fi和fj来自相同的身份。在这种情况下，它使两个DeepID2向量之间的L2距离最小化。 y ij = -1表示不同的身份，方程 （2）要求距离大于边际m。θve = {m}是在验证损失函数中要学习的参数。 基于L1规范的损失函数可以具有相似的公式。 在文献[2]中使用余弦相似度



其中是DeepID2向量之间的余弦相似度，ωve = {w,b}是可学习的缩放和移位参数， 是sigmoid函数，yij是两个比较的脸部图像是否属于相同身份的二进制对象。 在实验中对所有三种损失函数进行评估和比较。

目标是在特征提取函数Conv（）中学习参数c，而id和ve只是在训练过程中传播识别和验证信号的参数。在测试阶段，只有c用于特征提取。 数通过随机梯度下降算法进行更新。识别和验证梯度由超参数α加权。公（2）是一个特殊情况，不能通过梯度下降来更新，因为它的渐变始终是非负的。相反，我们在训练中自适应地更新m，这是最近训练样本的最低验证误差的阈值。学习算法总结如下



评估上面描述的特征学习算法。DeepID2特征嵌入到面部对齐，特征提取和面部验证的常规面部验证处理中。 文章首先使用最近提出的SDM算法来检测21个面部地标。然后根据检测到的地标，通过相似变换将面部图像全局对齐。根据全球对齐的面孔和面部地标的位置，裁剪了400个面部贴片，其位置，尺度，颜色通道和水平翻转情况各不相同。 因此，通过总共200个深度ConvNc提取400个DeepID2向量，每个深度ConvNets被训练为分别在每个面部的一个特定面部patch及其水平翻转的对应物上提取两个160维DeepID2向量。

为了减少大量的DeepID2特征之间的冗余，让系统实用，文章采取了前向贪婪算法[3]选择少量有效和补充的DeepID2向量（文章实验中为25），这节省了大部分测试期间的特征提取时间。图3-2示出了所有选择的25个patch，从中提取了25个160维度的DeepID2向量并且连接到4000维DeepID2向量。通过PCA进一步压缩4000维矢量进行面部验证。



图3-2 不同的patch输入

文章基于提取的DeepID2尝试了用于面部验证的联合贝叶斯模型。联合贝叶斯已经成功地用于模拟两个面孔相同或不同的人的联合概率，上文中有所提及。它将面部的特征表示f模型作为个体间和个体间变化的总和，或者，其中两者和被建模为高斯分布，并根据训练数据估计。通过对数似然比检验实现面部验证，其中分子和分母分别是给定个体间或个体间变异假设的两个脸部的联合概率。

至此，第二篇文章的主要理论基本完成，接下来简要介绍文章中提到的实验。

文章在无约束条件面部验证实际标准测试集LFW数据集上报告面部验证结果。它包含13 233张从互联网收集的5749个身份的脸部图像。为了比较的目的，算法通常报告LFW中6000个给定面对上的平均面验证精度和ROC曲线。因为LFW中的大多数身份只有一个脸型。需要依靠更大的外部数据集进行训练，使用CelebFaces +数据集进行训练，其中包含202 599张，总共10177个从互联网收集的个身份。 CelebFaces +和LFW中的人是相互排斥的。从CelebFaces +（称为CelebFaces + A）中随机采样的8192个身份的脸部图像中学习了DeepID2的特征，而1985个（称为CelebFaces + B）的剩余脸部图像用于以下特征选择和学习面部验证模型（联合贝叶斯）。在CelebFaces + A上学习DeepID2时，使用CelebFaces + B作为验证集来决定学习率，训练时期和超参数之后，CelebFaces + B被分为1485个身份的训练集和500个身份的验证集合用于特征选择。最后，使用所选的DeepID2在整个CelebFaces + B数据上训练联合贝叶斯模型并对LFW进行测试。

实验中我个人认为比较关键的是通过调节λ来观察识别和验证信号对面部验证的影响，并试图分析。λ为0或时只有一个信号起到监督的作用，图3-3表现了两种信号都不能单独很好地表征脸部特征信号，要结合在一起使用。

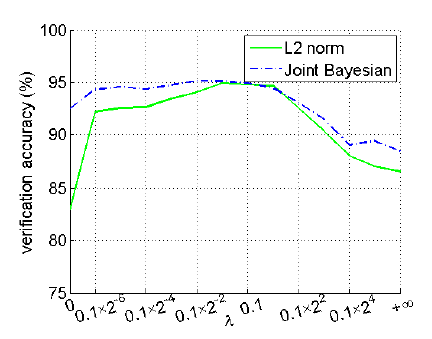


图3-3 对两种信号比重的分析

文章试图解释这个现象：可以从个人差异与不同身份之间差异，可以通过LDA近似。根据LDA，个人差异的散布矩阵为，其中xi拔表示第i个身份的平均特征，x拔是整个数据集的平均特征，ni是第i个身份的图像数量。身份之间差异可以使用这个散布矩阵表征，其中Di是第i个身份的特征集，xi是相应的平均值，c是不同身份的数量。个人差异和身份之间差异是相应散点矩阵的特征值，如图3-4所示，相应的特征向量表示不同的变化模式。特征差异的大小和多样性在认识上都是重要的。如果所有的特征差异都集中在少数特征向量上，则表明个体之间或个人之间变化的多样性较低。分别使用λ=0.05和。每个给定的特征差异λ通过相应的平均特征方差归一化。

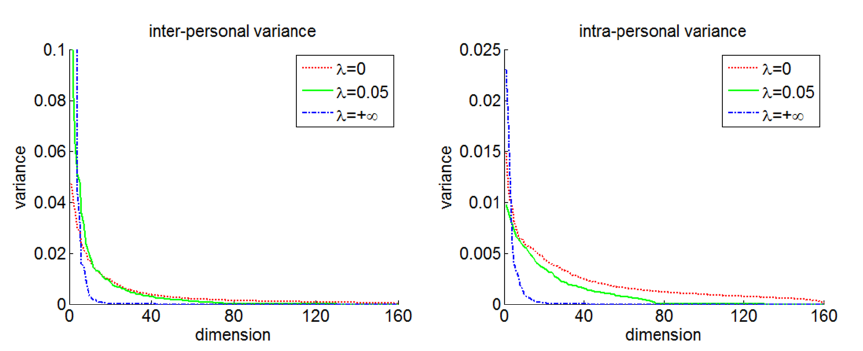


图3-4 不同参数的特征差异

当仅使用识别信号（λ= 0）时，学习的特征包含了两种不同的个体间和个体间变化，如图3-4中红色曲线的长尾所示。虽然不同的个人差异有助于区分不同的身份，但是大小多样的个人内部变化是噪音，并使面部验证困难重重。当识别和验证信号以适当的加权（λ= 0.05）使用时，个体间变化的多样性保持不变，而几个主要方向的变化变得更大，如左侧的绿色曲线所示相对于红色来说。同时，个体差异在多样性和幅度上均有所下降，如右图中绿色曲线所示。因此，个人和个人之间的变化在使面部验证更容易的方向上发生变化。当λ进一步向上增加时，个体之间和身份之间的差异仅在几个主要方向上变化，因为没有识别信号，不能形成不同的特征。个人差异的差异很小，区别不同的身份变得困难重重。因此，性能显着降低。

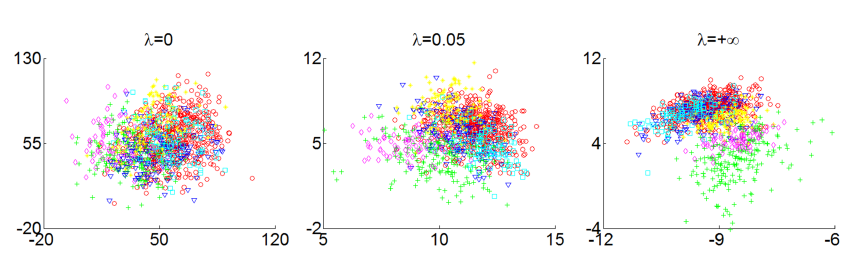


图3-5 聚类的分簇

图3-5显示了分别用λ= 0,0.05和+∞学习的特征的两个PCA维度。这些特征来自LFW中面部图像数量最多的六个身份，并以不同的颜色标记。该图进一步验证了我们的观察结果。当λ= 0（左））时，由于个人差异较大，不同的聚类混合在一起，尽管聚类中心实际上是不同的。当λ增加到0.05（中间）时，个人内部变化显着降低，并且簇变得可区分。当λ进一步朝向（右）方向增加时，尽管个人内部变化进一步减小，但是集群中心也开始崩溃，并且一些集群变得显着重叠（如图3-5中的红色，蓝色和青色集群），使聚类变得很难区分。

文章为了了解识别信息中包含的身份信息如何影响学习的特征，尝试在训练期间用于识别的指数增加的身份数（从32到8192），而验证信号始终从所有8192训练身份生成。图3-6显示了所学习的DeepID2（从L2范数和联合贝叶斯得到的）的验证精度如何随着识别信号中使用的身份数量而在测试集上变化。它表明识别大量（例如，8192）的身份是学习有效的DeepID2表示的关键。 这一观察结果与上文分析相符。越来越多的身份提供更丰富的身份信息，并有助于形成具有不同个人差异的DeepID2，使不同身份的类别中心更加可区分。

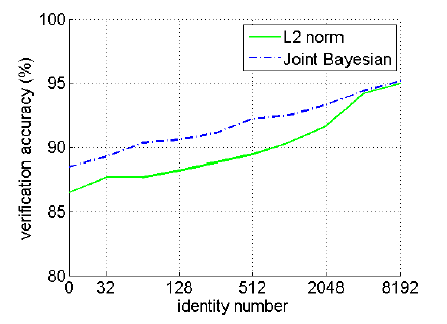
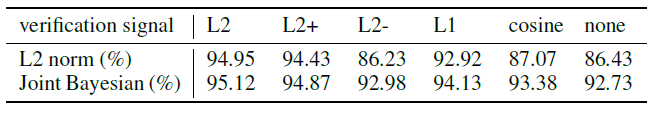


图3-6 增大身份训练数量影响

如上文所述，中等强度的验证信号主要是减少个人差异的影响。为了进一步验证这一点，文章将所有样本对的L2范数验证信号进行比较，那些仅限制正或负样本对，分别表示为L2 +和L2-。也就是说，L2 +仅减少相同身份的DeepID2之间的距离，而L2-仅增加不同身份的DeepID2之间的距离。通过L2范数和联合贝叶斯分别测量的学习的DeepID2在测试集上的面部验证精度如下表所示。它还与L1范数和余弦验证信号进行比较，以及没有验证信号（无）。所有比较的识别信号是相同的（对8192个身份进行分类）。



使用L2 +验证信号学习的DeepID2功能仅比用L2学习的功能略差。如此验证之下，L2验证信号在特征学习中几乎没有帮助，并且提供与使用无验证信号几乎相同的结果。这是一个强有力的证据，证明信号的效果主要是减少个人内部的变化。另一个观察结果是，除了识别信号之外，一旦添加了验证信号，通常都能改善面部验证精度。然而，L2规范比其他比较验证度量更好。这可能是因为所有其他约束都比L2弱，并且在减少个人内部变化方面效果较差。例如，余弦相似性仅限制角度，而不是大小。

以上就是第二篇文章的大致内容，相对于第一篇来说，主要在目标函数上加入了验证信号，使得同一张脸的内部差异减小，将效果提升到了当时的最佳。

**四、Deeply learned face representations are sparse, selective, and robust**

本篇文章在DeepID2的基础上，提出了更深层的结构，同时对卷积网络进行了大量分析，发现了几个重要特征。

在面部识别已经达到甚至超越人眼识别率的时候，人们开始更加想要知道深度神经网络到底学习到了什么特征使得性能表现的如此优越。在识别领域，有许多关于视觉神经如何处理面部区分的研究。在这些研究的背景下，文章分析了人造神经网络中神经元的行为，试图解释人脸识别过程中在深层网络，什么信息在神经元中被编码，以及它们对破坏程度的强大容忍度（鲁棒性）。

研究是基于文章提出的高性能深层神经网络（ConvNet的改进），被称为DeepID2 +。通过增加隐藏表示的维度并对早期的卷积层增加监督，改进了最先进的DeepID2网络。最好的单一DeepID2 +网络（采用原始和水平位移的脸部图像作为输入）在LFW上获得98.70％的验证精度（DeepID2为96.72％）。结合25个DeepID2 +网络在多个基准上实现了最新的效果：面向验证的LFW为99.47％（DeepID2为99.15％），封闭和开放式LFW的95.0％和80.7％ （通过WST分别为82.5％和61.9％），面孔验证（35％）为93.2％，而DeepFace 为91.4％）。

随着最先进的深层ConvNets和广泛的经验评估，文章发现神经激活的三个属性对于高性能至关重要：稀疏性，选择性和鲁棒性。

DeepID2 +的神经激活中度稀疏。如图4-1所示，对于输入面部图像，顶部隐藏层中的大约一半的神经元被激活。另一方面，每个神经元在大约一半的脸部图像上被激活。这种稀疏分布可以最大化深网的辨别力以及图像之间的距离。不同的身份有不同的神经元子集激活。相同身份的两个图像具有相似的激活模式。这促使文章将顶层隐藏层中的神经反应二值化，并使用二进制代码进行识别。其结果令人惊讶。其在LFW上的验证精度仅略微下降1％以下。它对大面积搜索具有重大影响，因为节省了巨大的存储和计算时间。这也意味着二进制激活模式比深层神经网络中的激活幅度更重要。

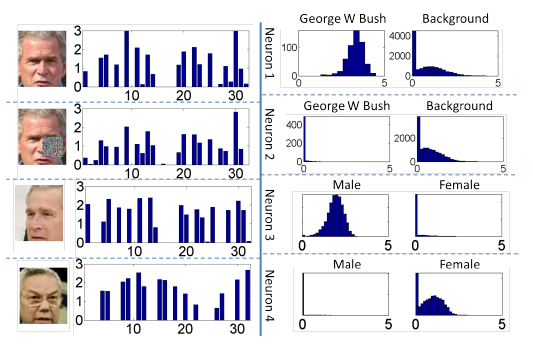


图4-1 稀疏，选择性和鲁棒性

与稀疏相似，还观察到，较高层中的神经元对身份和身份相关属性具有高度的选择性。当身份（可以在训练数据之外）或属性被呈现时，可以识别不断激发的神经元的子集，并且还可以发现不断被抑制的另一个神经元子集。来自这两个子集中的任一个的神经元具有关于该身份或属性的存在/不存在的强烈暗示，并且文章的实验表明单个神经元对于特定身份或属性具有高识别精度。换句话说，神经激活对身份和属性具有稀疏性，如图4-1所示。尽管DeepID2 +没有被教导在训练中区分属性，但它隐含地了解了这样一个高层次的概念。直接使用DeepID2 +学习的面部表征，导致身份相关属性的分类精度高于广泛使用的手工功能，如高维LBP 。

实证研究表明，较高层中的神经元对于人脸识别中的图像损伤要比手工制作的特征（如高维LBP或下层神经元）更加鲁棒。如图4-1所示，当面部图像被部分遮挡时，其二进制激活模式保持稳定，尽管幅度可能改变。我们推测原因可能是更高层的神经元捕获全局特征，对局部变化较不敏感。 DeepID2 +由正常的网络脸部图像集训练，并且没有将人工遮挡模式添加到训练集中。

文章的DeepID2+网络是从DeepID2网络中改进得到的，与DeepID2相比，DeepID2 +做了三个改进。首先，在四个卷积层中的每一个中使用128个特征map。最终的特征表现从160到512个维度。第二，训练数据通过合并CelebFaces +数据集，WDRef数据集和一些从LFW排除的新收集的身份来扩大。大型的DeepID2 +网络训练集有来自12,000个身份的大约290,000张脸部图像，而来自用于训练DeepID2网络的8,000个身份的160,000张图像。第三，在DeepID2网络中，监控信号仅被添加到连接到第三和第四卷积层的一个完全连接的层，而较低的卷积层只能通过从较高层反向传播的梯度得到监视。通过将512维全连接层连接到四个卷积层（最大池）之后来加强监督，在n = 1,2,3,4时表示为FC-n，并对这四个完全连接层进行监控，连接层与识别验证监控信号同时进行，如下页图4-2所示,以这种方式，监控信号变得越来越接近早期卷积层，并且更有效。

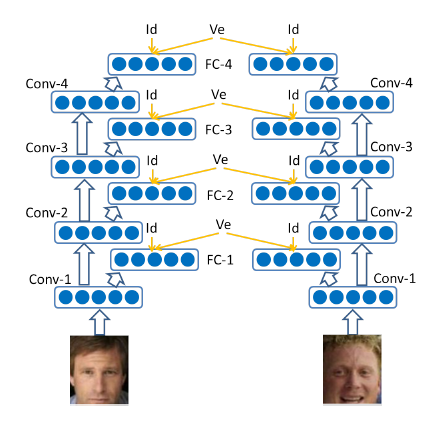


图4-2 DeepID2+

为了验证改进，文章基于DeepID2 +网络中n = 1,2,3,4的四个512维特征向量（神经激活）FC-n中的每一个，了解用于面部验证的联合贝叶斯模型。联合贝叶斯（不包括LFW人员）中接受了2000人数据集的训练，这2000个身份不用于训练DeepID2 +网络，并在LFW的6 000名给定面部对中进行面部验证。当训练DeepID2 +网络以确定学习率和训练迭代时，这2000个身份也作为验证集。将DeepID2 + net与三个分别删除了三个改进之一的网络进行比较。如图4-3所示，对于没有分层监督的网络，只有FC-4的梯度反向传播到卷积层。具有较少训练数据的网络，仅使用来自CelebFaces +的训练数据。较小的网络，卷积层中的特征图的数量与DeepID2中的特征映射的数量相同，并且在n = 1,2,3,4中，将160维特征用于FC-n。所有被比较的网络在固定区域覆盖整个脸部。可以清楚地从图4-3中看到三个方面的改进。

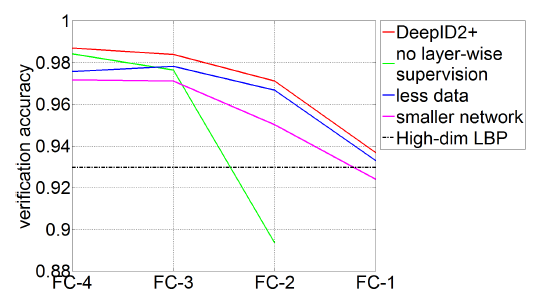


图4-3 三个方面改进的对比

为了与DeepID2网络进行比较，文章训练了25个DeepID2 +网络，采用了DeepID2选择的相同的25个脸部区域，如图3-2所示。提取了DeepID2 +的FC-4层的特征，以训练贝叶斯联合模型进行验证。 LFW面部验证任务的25个深度对比之间的比较如图4-4所示。 DeepID2 +网络在DeepID2平均提高了约2％的精度。当分别从25个脸部区域提取的FC-4层特征及其与25个DeepID2 +网络的水平对照相结合时，文章分别在LFW和YouTube Faces数据集上实现了99.47％和93.2％的面部验证精度。

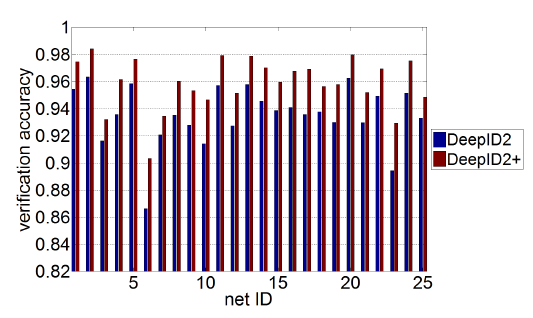


图4-4 对比直方图

下表是与两个数据集上各种方法的最佳结果的精度比较。

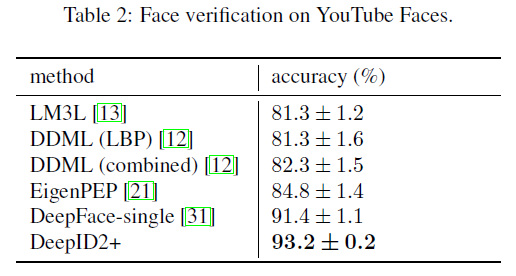
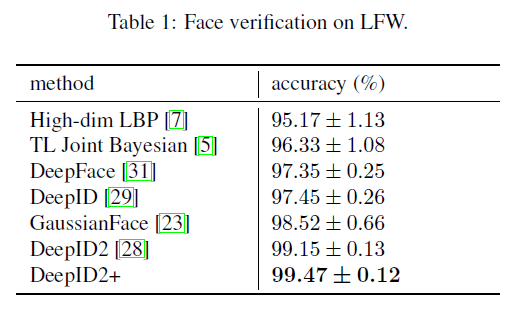


图4-5是ROC的比较。文章的DeepID2 +网络在两个数据集上都胜过以前的所有结果。LFW和YouTubeFaces中有几个错误标签的测试脸对。修正后，文章的脸部验证精度在LFW上提高到99.52％，在YouTube上增加到93.8％。

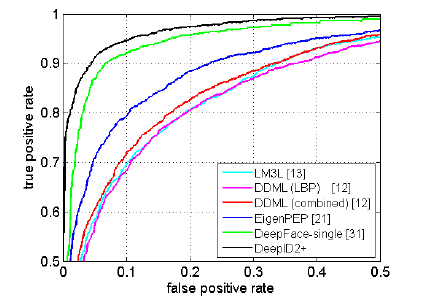
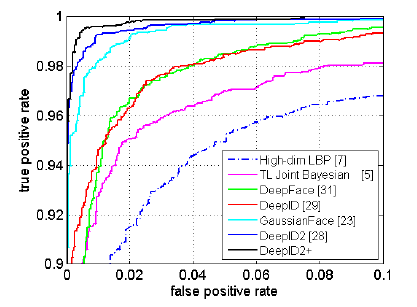


图4-5 ROC的对比

对于每个图像，大约有一半被激活（具有正激活值）的神经元，并且对于每个神经元，大约一半的图像被激活，这意味着神经激活是适度稀疏的。图像的适度稀疏使不同身份的脸部最大限度地区分开来，而神经元的适度稀疏使他们具有最大的辨别能力。文章通过在自己的验证数据集（图4-6左边）中计算46 594个图像中的每个图像上激活的神经元的直方图，以及激活每个神经元的图像数量的直方图来验证这一点（图4-6右侧）。以整个面部区域作为输入，评估基于单个DeepID2 +网络中的FC-4层神经元。与FC-4层的所有512个神经元相比，图像上激活的神经元数量的平均值和标准偏差为292±34，而与所有46,594个验证图像相比，图像数量的平均值和标准偏差每个神经元被激活的是26,565±5754，两者都以所有神经元/图像的一半近似为中心。

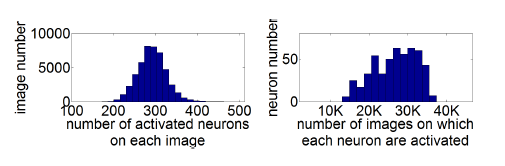


图4-6 对比直方图

文章进一步验证激活模式，即神经元是否被激活，比精确的激活值更重要。将神经激活转换为二进制代码，通过阈值和比较LFW的面部验证能力与原始表示的面部验证能力。如下表表示，当与联合贝叶斯联合时，二进制表示法分别为1％或更低的精度（分别为97.67％和99.12％，当单个网络或组合25个网络）。更有趣的是，即使通过直接计算汉明距离，二进制代码仍可以通过单个网络或组合25个网络分别达到96.45％和97.47％的精度。这表明神经元的激发或抑制状态已经包含了大多数歧视性信息。二进制代码对存储是经济的，而且使得图像搜索更加快速，这是一个未来工作的有趣方向。

接下来文章讨论了身份和属性的选择性。

1. 神经元的辨别力

文章测试了两个二分类任务的DeepID2 +特征。首先是将一个人的脸部图像与所有其他人或背景的脸部图像进行分类。然后是将脸部图像分类为具有属性或没有属性。DeepID2 +特征分别来自整个脸部区域的单个DeepID2 +网络的FC-4层及其水平平移的对应图像。实验在LFW下进行，身份在训练期间对DeepID2 + net是不可见的。LFW被随机分为两个子集，报告其交叉验证精度。文章还比较了具有各种特征尺寸的高维LBP特征。如图4-7所示，DeepID2 +特性在属性分类中显着优于LBP（DeepID2 +具有良好的身份分类结果并不奇怪）。图4-8显示了仅选择一个最佳特征的身份和属性分类精度。为不同的身份（属性）选择不同的最佳特征。使用单个功能（神经元），DeepID2 +对于某些身份和属性达到约97％。这就是DeepID2 +特征是身份和属性选择的证据。显然LBP没有。

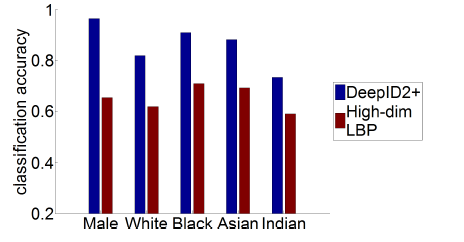
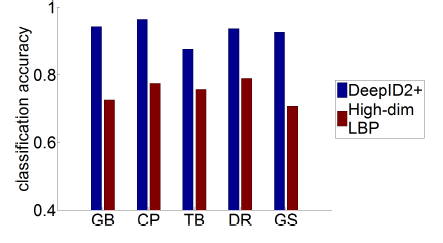


图4-8

1. 激活与抑制神经元

文章发现对身份和面部属性的区分是由于某些身份或属性上的神经元的激发和抑制模式。例如，当看到乔治•布什（George Bush）变得抑制，或者一个神经元可能会被西方人激活，而对亚裔人士抑制，一个神经元可能会被激活。图4-9左表示比较DeepID2 +神经激活对属于特定单个身份（左列）和剩余图像（中间列）的图像的平均值和标准偏差，以及显示每个给定身份与每个特定身份的每个神经元分类的准确性相比于剩余图像（右列）。在LFW中具有最多脸部图像的五个身份被评估（其他身份具有相似的结果）。所有三列中图像的评估身份的平均神经激活对神经顺序进行排序。对于每个给定的身份，存在强烈兴奋的神经元（例如，具有小于200 ID的神经元）或被抑制（例如具有大于600 ID的神经元）。对于激发的神经元，其激活分布在更高的值，而其他图像在这些神经元上具有显着较低的平均值。因此，兴奋性神经元可以很容易地将身份与其他身份区分开来，这是通过右栏列出的图形中具有小神经元的红点所示的高分类精度来验证的。

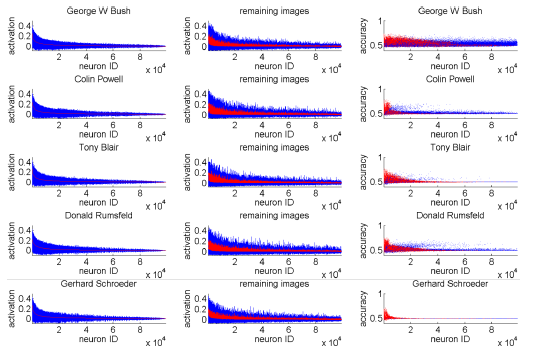
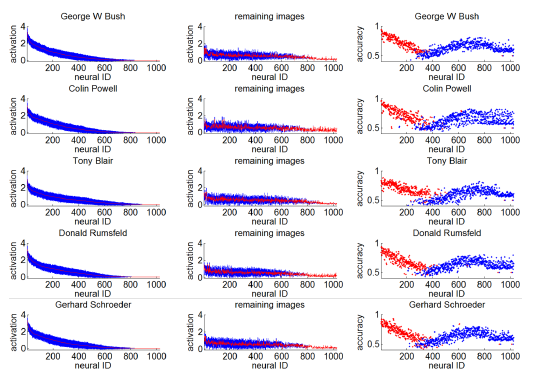


图4-9

对于位于中间的神经元（例如，具有约400 ID的神经元），它们在给定身份上的激活分布与其他身份上的激活分布大部分重叠。对于给定的身份，它们具有弱的识别能力，由两种颜色的连接点附近的红色和蓝色点的低精度证实。这些神经元的激发或抑制状态有很大的不确定性。

当平均激活进一步降低（例如，高于600 ID的神经元）时，神经元显示出抑制性质，并且与其他特征相比较，对于给定身份很少被激活。这些抑制性神经元也具有相对较高分类精度的鉴别能力。

然而，如图4-9右图所示，在LBP特征上不能发现类似的现象。给定身份的LBP功能的范围和剩余的图像对于所有功能重叠。与DeepID2 +神经激活相比，LBP特征具有低得多的分类精度，其中大部分在50％随机猜测线上累积。

图4-10左图比较包含特定属性（左列）和剩余图像（中间列）的面部神经激活的范围，以及显示每个属性与其余图像（右列）区分的每个神经元分类的准确性。与身份相似，下列和更高级别的神经元对于属性的选择表现出如图所示，包括男性，白人，黑人，亚裔，儿童，高级，秃头和灰色头发。这些属性与身份有区别。其他属性的选择性相对较弱，如印度，青年，中年，黑发，白发和褐发（未显示）。这些属性在视觉上是不明确的或对身份的区分较少。例如，印度人有时看起来像亚洲人，而且经常看到在青年和中年时拍摄的身份相同的身份，或以不同的头发颜色拍摄

图4-10右图比较了与左图相同的属性集合的LBP特征的范围和每特征分类精度。给定属性和剩余图像的LBP特征范围与所有特征重叠，分类精度累积在50％随机猜测线上。

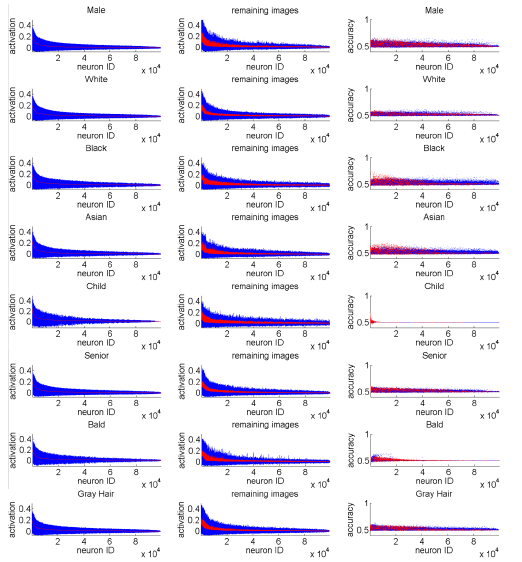
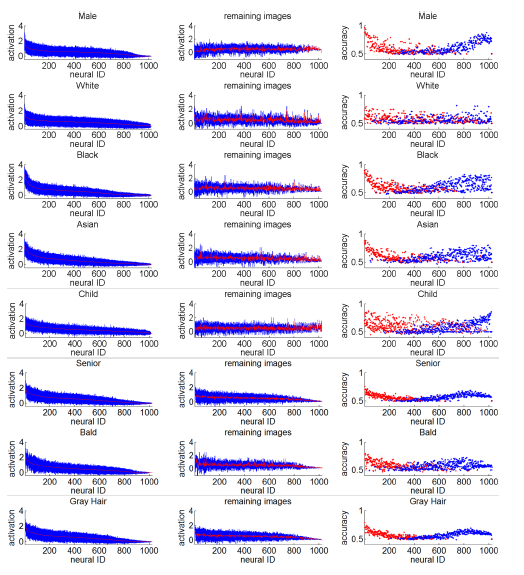


图4-10

1. 神经激活分布

图4-11和4-12示出了在给定身份或属性上的神经激活的直方图的示例。图4-11第一行还显示了五个随机选择的神经元的所有图像上的直方图。对于每个神经元，其激活的大约一半是零（或接近零），另一半具有更大的值。相比之下，给定身份的直方图表现出很强的选择性。一些神经元对于给定的身份不断被激活，激活直方图分布在正值中，如图4-11中每个身份的直方图的第一行所示而其他一些被持续抑制，其中激活直方图累积在零或小的值，如每个身份的第二行直方图所示。

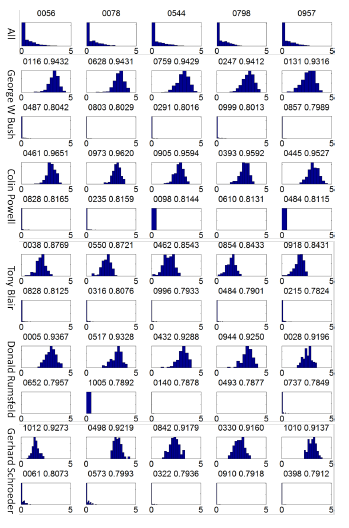
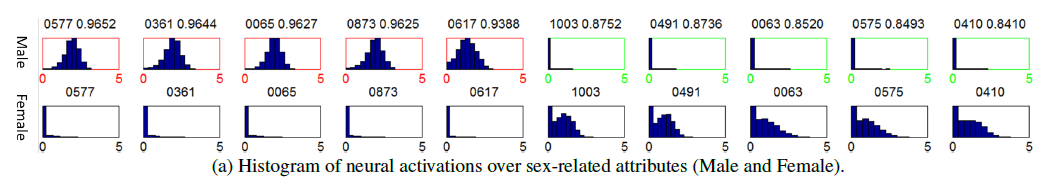
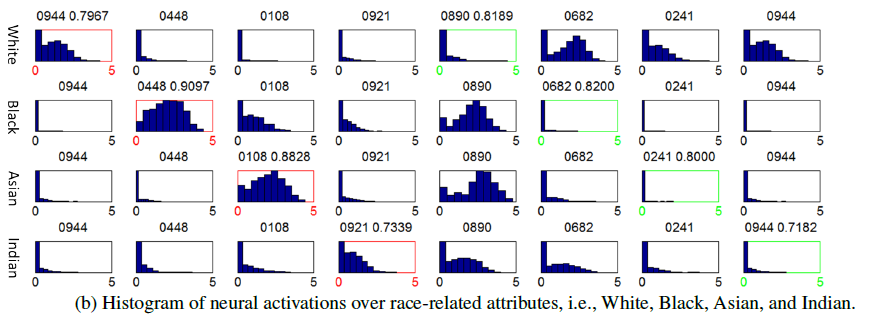
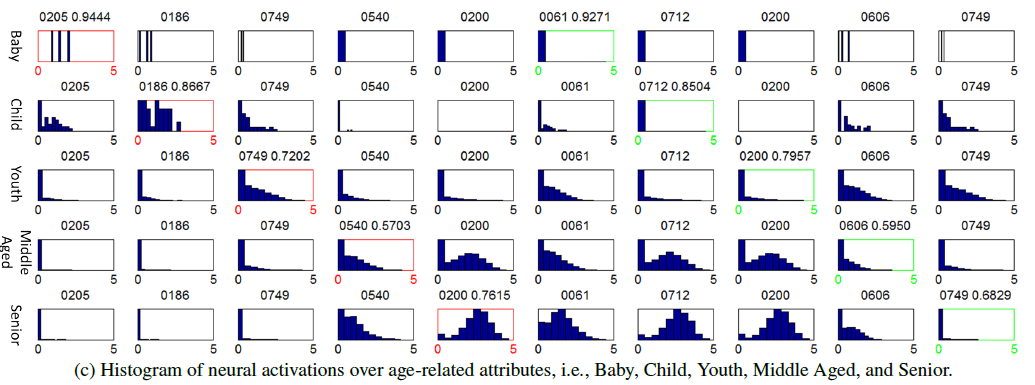


图4-11

对于属性，在4-12a，4-12b，4-12c和4-12d中，我们分别显示了几个相关属性，即与性别，种族，年龄和头发相关的单个神经元的直方图。神经元选择为兴奋性（红色）或抑制性（绿色），并且可以最佳地分类每行左侧所示的属性。如这些图所示，神经元对于某些属性表现出强烈的选择性（被激活或抑制），其中对于给定属性，神经元被激活（抑制），同时对同一类别中的其他属性被抑制（激活）。







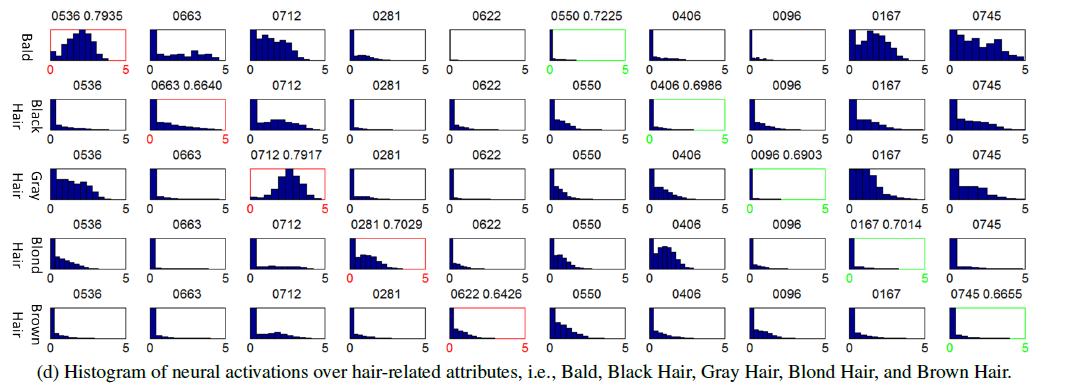


图4-12

最后讨论了DeepID2+网络学习到的特征具有鲁棒性（将脸部遮挡部分然后验证）这一性质。第一个设定，面部被局部遮挡10％到70％的区域，如图4-13第一行所示。第二设定，如图4-13第二行所示，面部被10×10〜70×70像素的随机块遮挡。在实验中，DeepID2 +网络和联合贝叶斯模型在训练集中的原始脸部图像上被学习，没有任何人工添加的约束，而遮挡面仅用于测试。文章还测试了高维LBP特征加上联合贝叶斯模型进行比较。图4-14比较了具有不同程度部分遮挡的LFW测试集上的DeepID2 +和LBP特征的面部验证精度。将DeepID2 +特征从单个DeepID2 +网络中的FC-1到FC-4层，并以整个面部区域作为输入。文章还使用25个DeepID2 +网来评估整个人脸识别系统，比较的高维LBP特征是从21个面部地标中提取的99，120个维度。如图4-14，LBP的表现显着下降，即使轻微的10％-20％的遮挡。相比之下，对于具有两个卷积（FC-2，FC-3和FC-4）的DeepID2 +特征，性能在很大的范围内缓慢下降。当40％的面部被遮挡（除了FC-1层）之后，DeepID2 +的面部验证精度仍然高于90％，而LBP特征的性能已经降到70％以下。当关键眼睛区域被遮挡时，DeepID2 +的表现只会降低50％以上。它还表明，较高层（其应该是更global分布）的特征对于遮挡更为鲁棒，而LBP和FC-1都是局部特征，对遮挡敏感。结合从25个脸部区域提取的DeepID2 +网络实现了最强的鲁棒性，具有93.9％的面部验证精度，40％的遮挡和88.2％的精度，即使只显示额头和头发。

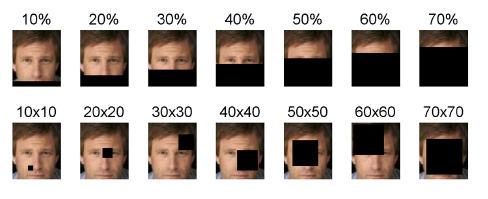


图4-13

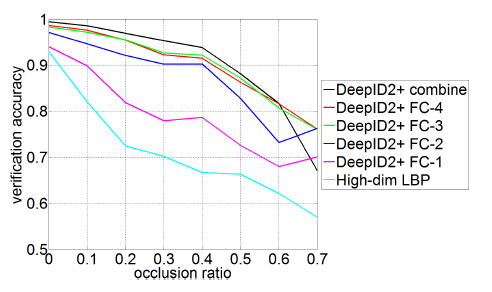


图4-14

文章还对具有随机块遮挡的面部图像进行DeepID2 +和LBP特征的面部验证，n×n块大小分别为n = 10到70。这种设定是具有挑战性的，因为待验证的两个面中的遮挡区域的位置通常是不同的。因此，在像素距离的意义上，同一人的图像将会看起来有很大的不同。图4-15显示了比较结果，当块大小大于20×20时，LBP特征的准确性开始下降，而DeepID2 +特征（FC-1除外）在较大范围内保持了性能。具有50×50块的遮挡，LBP功能的性能已经下降到70％左右，而单个DeepID2 +网络的FC-4层仍具有89.2％的准确性，而25个DeepID2 +网络的组合更高达92.4 ％ 准确性。而且，浅层FC-1层的特征行为更接近LBP特征。上述实验表明，深层结构使神经元对图像的破坏更为鲁棒。这种鲁棒性在深层ConvNets中是固有的，不需要明确的建模。

最后研究了FC-4层神经元对具有不同程度的部分和随机块遮挡的单个身份的图像的平均激活。神经元根据每个身份的原始图像的平均激活进行排序。对于两种类型的遮挡，激活模式保持大致不变，除非存在很大比例区域的遮挡。

本文对三种性质的深入研究使得面部识别的深层神经网络具有一定的理论佐证，同时对其他图像领域的研究有一定的启发。

**五、DeepID3: Face Recognition with Very Deep Neural Networks**

在上一节中介绍的技术已经基本达到了目前人脸验证的较高的识别率，本文的主要意义在于结合最新的很深层的网络结构应用在本领域，在图像识别领域取得的成功能否应用于脸部识别。提出了两种非常深的神经网络架构，称为DeepID3，用于人脸识别。这两种架构由VGG和GoogLeNet 中提出的堆叠卷积和初始层重建，使其适合面对识别。在训练期间，中间和最终特征提取层都加入了联合面部鉴定验证监控信号。所提出的两种架构的综合分别达到了99.53％的LFW面验证精度和96.0％的LFW等级1面部识别精度。并在最后给出了对LFW面部验证结果的进一步讨论。

DeepID3有两种不同的结构，分别为DeepID3 net1，DeepID3 net2。相对DeepID2+，它的层数更多，网络更深。同时还借鉴了VGG net和GoogLeNet，引入了inception layer，这个主要是用在了DeepID3 net2里面。网络中还出现了连续两个conv layer直接相连的情况，这样使得网络具有更大的receptive fields和更复杂的nonlinearity，同时还能限制参数的数量。

为了比较，文章简要介绍了以前提出的DeepID2 +网络架构。如图5-1所示DeepID2 + net具有三个卷积层，其次是最大池（仅在局部区域中的第三个运输层中的神经元共享权重），随后是一个本地连接的层和一个完全连接的层。 联合识别验证监控信号被添加到最后一个完全连接的层（从中提取最终特征以进行面部识别）以及从中间池层分出的一些完全连接的层，以更好地监督早期特征提取流程。

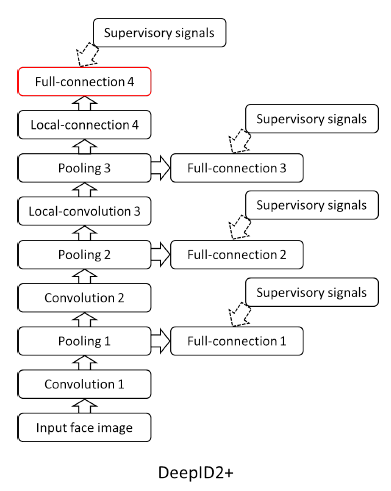


图5-1 DeepID2+的结构图

所提出的DeepID3网络继承了DeepID2 +网络的一些特征，包括最后几个特征提取层中的非共享神经元权重，以及将监控信号添加到早期层的方式。 然而， DeepID3网络显着更深，具有10到15个非线性特征提取层。特别地，文章提出两个DeepID3的网络架构，DeepID3 net1和DeepID3net2，如图5-2所示。DeepID3网络的深度是由于在每个池层之前堆叠多个卷积/初始层。连续卷积/初始化有助于形成具有较大接收场和更复杂的非线性特征的特征，同时限制参数数量。

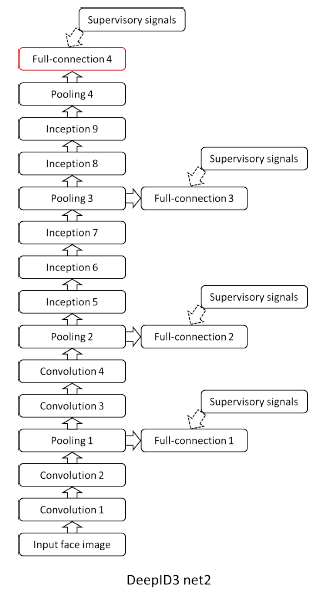
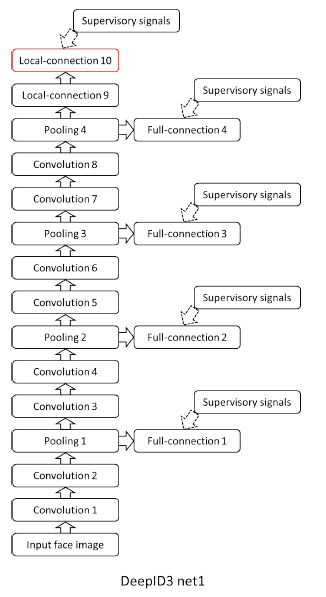


图5-2 DeepID3

提出的DeepID3 net1在每个池化层之前采用两个连续的连续层。VGG网相比，在从中间层分支出来的多个全连接层中添加了更多的监控信号，这有助于学习更好的中级特征，并使得优化非常深的神经网络更容易。前两个卷积层被局部连接的层代替。使用非共享参数，顶层可以形成具有减少的特征维度的更具表现力的特征。DeepID3 net1的最后一个本地连接层用于提取最终的特征，而无需额外的完整连接层。

DeepID3 net2从每两个连续的连续层开始，随后是DeepID3 net1中的一个池层，同时在后续特征提取阶段采用初始层：在第三个池层和两个初始层之前有三个连续的初始层在第四个池层之前。联合识别验证监控信号在每个池层之后的完全连接的层上添加。

在提出的两种网络架构中，rectified linear non-linearity 被用于所有除了集合层之外，并且最终特征提取层添加了dropout learning。尽管具有显著的深度，但由于每层中的特征图数量有限，所以DeepID3网络比VGG网络或GoogLeNet在一般对象识别中提出的要小得多。

分别再DeepID3 net1 ，DeepID3 net2 和DeepID2+上进行了对比试验，所提出的DeepID3网络在LFW面部验证和识别任务上实现了最先进的性能。然而，当LFW中的一些错误的标签被纠正时，DeepID3与DeepWD2 +在LFW面部验证上的改进就消失了。所以只能期望将来通过更大规模的训练数据对更深层的神经网络的有效性进一步研究。

**六、DeepID的重现**

首先介绍深度学习框架caffe（Convolutional Architecture for Fast Feature Embedding）,是由BVLC开发的基于C++/CUDA/Python实现的卷积神经网络，提供了面向命令行、Matlab和Python的绑定接口。

Caffe依赖解析：

1. ProtoBuffer是由Google开发的一种可以实现内存与非易失存储介质（如硬盘文件）交换的协议接口。用户只需要建立统一的参数描述文件（proto），然后利用protoc编译就能让协议细节等关键部分代码自动生成，节省了大量的开发、调试时间。
2. BLAS（Basic Linear Algebra Subprograms），卷积神经网络中用到的数学计算主要是矩阵、向量的计算，Caffe调用了BLAS中的相应方法。
3. HDF5（Hierarchical Data File）是美国国家高级计算应用中心为了满足各种领域研究需求而研制的一种能高效存储和分发科学数据的新型数据格式。它可以存储不同类型的图像和数码数据的文件，并且可以在不同类型的机器上传输，同时还有统一处理这种文件格式的函数库。
4. LDMB和LEVELDB：LDMB（Lightning Memory-Mapped Database Manager）--闪电般的内存映射数据库管理器，在Caffe中的作用主要是提供数据管理，将形形色色的原始数据（如JPEG图片、二进制数据）转换为统一的Key-Value存储，便于Caffe的DataLayer获取这些数据。LEVELDB库是Caffe早期版本使用的数据存储方式，由Google开发。它是一种持续的键值对存储方式，键和值可以为任意字节数组。键的存储顺序可以由用户定义的比较函数决定。

Caffe模型和优化以纯文本的模式定义，而不是以代码的形式。不过以模块化的组织形式使得新任务和配置要求具有灵活性和扩展性。通过逐层定义（layer-by-layer）的方式定义一个网络（Nets）。网络从数据输入层到损失层自下而上地定义整个模型。Caffe使用blobs结构来存储、交换和处理网络中正向和反向迭代时的数据和导数信息：Blob是Caffe的标准数组结构，它提供了一个统一的内存接口。Layer是Caffe模型和计算的基本单元，Net是一系列layers和其连接的集合。Blob详细描述了信息是如何在layer和net中存储和交换的。

Solving（求解方法）单独配置，以解耦模型的建立与优化过程。

从数学意义上来说blob是按C风格连续存储的N维数组。

Layer是Caffe模型的本质内容和执行计算的基本单元。Layer可以进行很多运算，如：convolve（卷积）、pool（池化）、inner product（内积），rectified-linear和sigmoid等非线性运算，元素级的数据变换，normalize（归一化）、load data（数据加载）、softmax和hinge等losses（损失计算）。一个layer通过bootom 接收数据，通过top 输出数据。每一个layer都定义了三种重要运算：setup（初始化设置），forward（前向传播），backward（反向传播）。

与大多数机器学习模型一样，在caffe中，学习是由一个损失函数驱动的（通常也被成为误差、代价或者目标函数）。一个损失函数通过将参数集（即当前的网络权值）映射到一个可以标识这些参数“不良程度”的标量值来学习目标。因此，学习的目的是找到一个网络权重的集合，使得损失函数最小。

Solver通过协调Net的前向推断计算和反向梯度计算，来对参数进行更新，从而达到减小loss的目的。Caffe模型的学习被分为两个部分：由Solver进行优化、更新参数，由Net计算出loss和gradient。

Solver用于优化过程的记录、创建训练网络（用于学习）和测试网络（用于评估），通过forward和backward过程迭代地优化和更新参数。

实际上我们要做的deepid第一篇的重现实验主要工作就是编写proto文件，然后通过数据进行训练，测试即可。当然，配置和使用Caffe的过程就是了解学习整个Caffe框架的过程。

有关proto配置文件在最后的附录中。

根据net的proto定义文件，Caffe训练的网络如下图6-1所示：

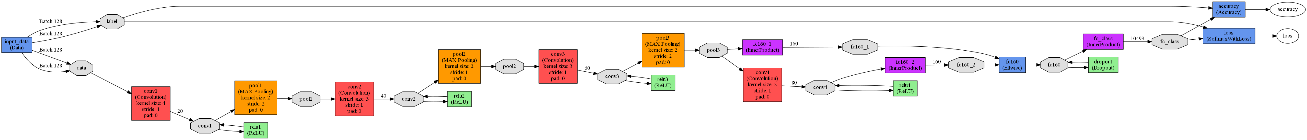
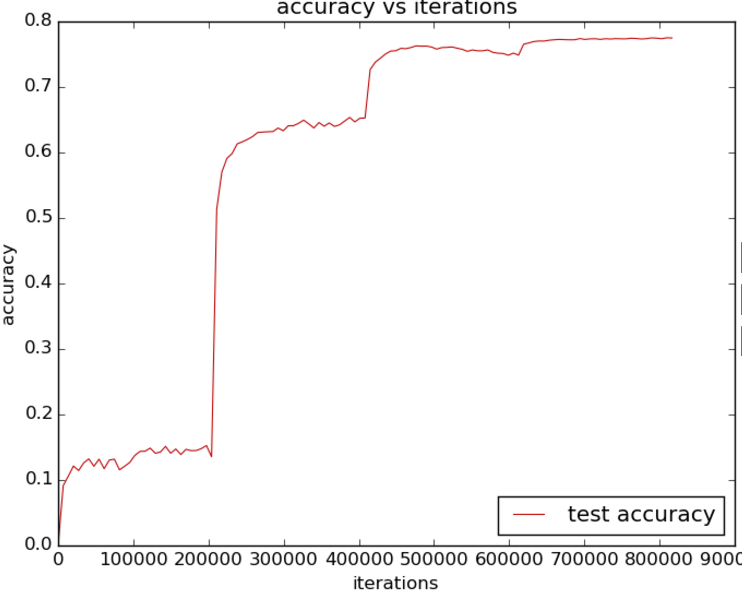
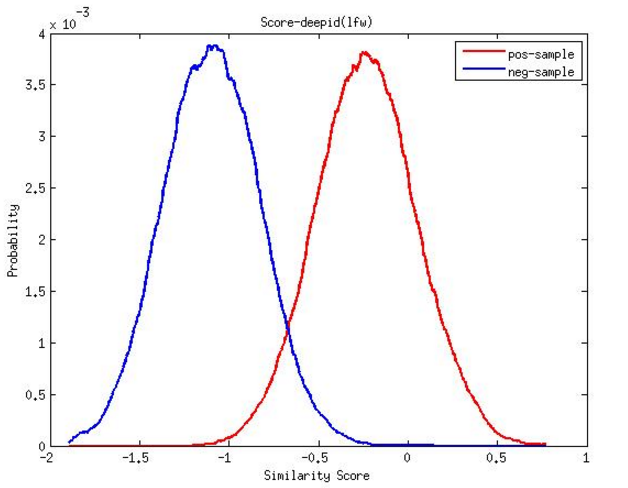
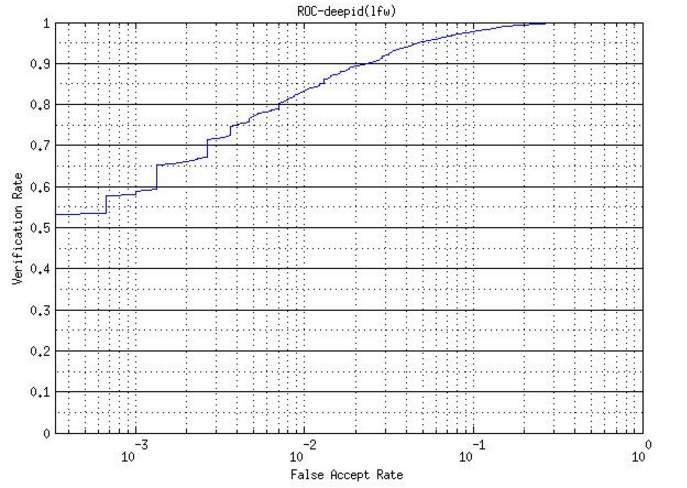


图6-1 DeepID的网络结构

在该CeleA训练集下训练，通过贝叶斯方法判断是否同一个人，相关图像如下，正确率0.947，并没有达到文献中那么高，分析后我认为是因为数据的选取和网络设计的问题，鉴于时间问题未来再进行改进。



**七、总结**

通过阅读这个系列的文献，首先为了更好地理解文献，重新学习了神经网络以及深度神经网络的内容，加深了理解。然后在人脸验证领域的应用并达到了人眼识别率以上的结果让人对机器学习的前景充满向往。在具体应用中需要使用几个深度学习的框架，分析对比之后选用了Caffe这个框架，因为虽然它配置和安装并不是最简单的，不过它基于Google开源的ProtoBuff，网络搭建很容易，并且文档较多，同时在计算机视觉领域Caffe足以满足绝大多数应用。但是虽然在很多领域机器学习近年来大放异彩，也要保持理性的态度，不是所有的问题目前都适合使用机器学习的方法进行解决，比如网络的性能优化和资源配置，在大量数据的获取上就存在一定障碍。总体来说，这次研读文献让我对机器学习的精彩世界有了一定的直观认识，很有收获。

参考文献：

[1] Hinton G E, Salakhutdinov R R. Reducing the Dimensionality of Data with Neural Networks[J]. Science, 2006, 313(5786):504-507.

[2] V. Nair and G. E. Hinton. Rectified linear units improve restricted Boltzmann machines. In

Proc. ICML, 2010.

[3]. T. Zhang. Adaptive forward-backward greedy algorithm for learning sparse representations.

IEEE Trans. Inf. Theor., 57:4689–4708, 2011.