Face detection、recognition and verification review

2016.11.10 徐昆源

相关概念：

1. Face detection : 检测图像中的人脸，并不识别其身份。
2. Face recognition: 在人脸检测的基础上，识别检测到的人脸的身份信息。
3. Face verification: 验证两张人脸是否为同一身份。

# 1 研究简史

## 第一阶段（1964-1990年）

此阶段人脸识别通常只是作为一个一般性的模式识别问题来研究，所采用的主要技术方案是**基于人脸几何结构特征（Geometric feature based）**的方法。此阶段是人脸识别的初级阶段，非常重要的成果不多，也基本上没有获得实际应用。

## 第二阶段（1991-1997年）

1. 提出了**特征脸（其后的很多方法的基础）**方法。
2. 研究者们得出结论：**模板匹配的方法**优于基于特征的方法。
3. 提出**主成分分析（PCA）**对图像表观特征进行降维，然后**采用线性判别分析（LDA）**的方法变换降维后的主成分获得“**尽量大的类间散度和尽量小的类内散度**”。

## 第三阶段（1998-2012年）

在此阶段，**光照、姿态**问题逐渐成为研究热点。引入、提出一系列的方法，包括以SVM、Adaboost为代表的统计学习理论。并且在**2006**年清华大学的方法在**实验室内部拍摄条件、正面姿态、正面光照**的条件下**全面超过人类。**

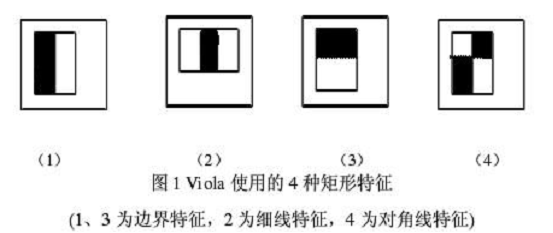
## 第四阶段（2012-至今）

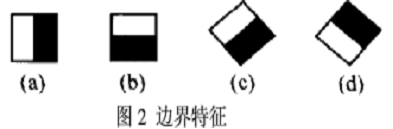
由于**卷积神经网络（CNN）**在图像理解与目标检测识别方面具有突出优势，而且**大量人脸数据库（LFW，CelebFaces，YoutubeFace…）**的形成，**深度卷积网络**开始应用于人脸识别与验证领域，并且实现了非常高的准确率**（LFW上超过99%，超过人类水平）**。可以克服光照、姿态变化的影响。

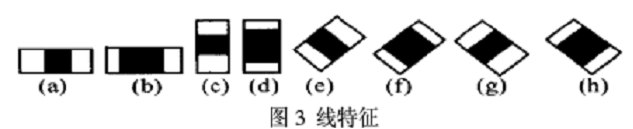
简要介绍一个人脸检测方法和深度学习相关的人脸识别论文与方法。

# 2 Face Detection （Haar 分类器）

## 2.1 Haar-like 特征









特征的计算方法：将上述矩形放入图像，将白色区域的像素和减去黑色区域的像素和得到的即为人脸特征值。

## 2.2 AdaBoost

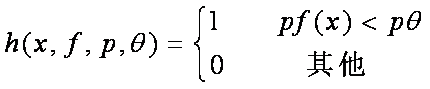
两个概念：

1. 弱学习：一个学习算法对一组概念的识别率只比随机识别好一点。
2. 强学习：一个学习算法对一组概念的识别率很高。

可证：只要有足够的数据，弱学习算法就能通过集成的方式生成任意高精度的强学习算法。

### 2.2.1 弱分类器的孵化

弱分类器的数学公式：



其中为特征，为阈值，p用于控制不等号方向。

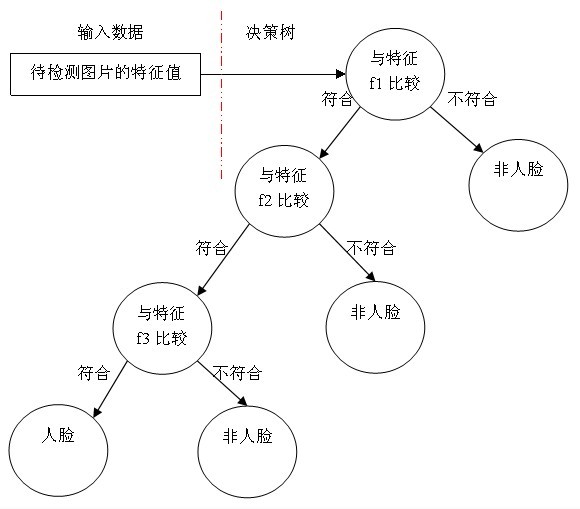
阈值的确定：

1. 对于每个特征，计算所有训练样本的特征值，并将其排序。
2. 扫描一遍排好序的特征值，对排好序的表中的每个元素，计算下面四个值：

* 全部人脸样本的权重的和
* 全部非人脸样本的权重的和
* 在此元素之前的人脸样本的权重的和
* 在此元素之前的非人脸样本的权重的和

1. 最终求得每个元素的分类误差，值最小的元素则为该特征的最优阈值。

弱分类器的形成：



### 2.2.2 强分类器的诞生

强分类器的诞生需要 T 轮的迭代，具体过程：

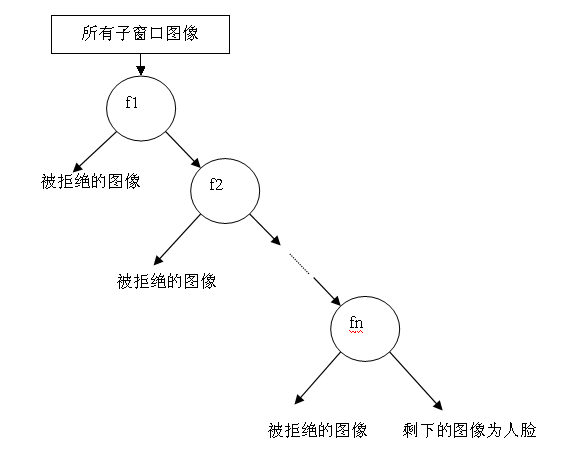
1. 给定训练样本集S ，共 N 个样本，T 为训练的最大循环次数；
2. 初始化样本权重为1/N ，即为训练样本的初始概率分布；
3. 第一次迭代训练N 个样本，得到第一个最优弱分类器；
4. 提高上一轮中被误判的样本的权重；
5. 将新权值的样本和进行新一轮的训练；
6. 循环执行4-5步骤，T 轮后得到T 个最优弱分类器。
7. 组合T 个最有弱分类器得到强分类器，组合方式为：让所有弱分类器按照一定的权重加权判断。

## 2.3 级联强分类器

Haar 分类器是有两个体系，训练的体系和检测的体系。

检测体系：以现实中的一幅大图作为输入，对图片进行多区域、多尺度的检测。

筛选式级联分类器：



级联强分类器策略：将若干个强分类器由简单到复杂排列，希望经过训练使每个强分类器都有较高检测率，而误识率可以放低，比如几乎99%的人脸可以通过，但50%的非人脸也可以通过，这样如果有20个强分类器级联，那么他们的总识别率为0.99^20 =98%，错误接受率也仅为0.5^20=0.0001%

Opencv 集成了此人脸检测方法，还有更好的人脸，时间有限，未研究。

# 3 Face recognition and verification

从2014年以来，基于深度卷积网络的人脸识别方法数次刷新识别正确率。此文介绍以下几篇论文：

1. DeepFace: Closing the Gap to Human-Level Performance in Face Verification.

By FaceBook AI Research and Tel Aviv University 2014

1. Learning Deep Face Representation

By Face++, China 2014

1. Surpassing Human-Level Face Verification Performance on LFW with GaussianFace

By Chinese University of Hong Kong 2014

1. Deep Learning Face Representation from Predicting 10,000 Classes

By Chinese University of Hong Kong 2014

1. Deep Learning Face Representation by Joint Identification-Verification

By Chinese University of Hong Kong 2014

1. Deeply learned face representations are spars, selective, and robust

By Chinese University of Hong Kong 2014

1. FaceNet: A Unified Embedding for Face Recognition and Clustering

By google 2015

1. Deep Face Recognition

VGG model(we are using) 2015

相关数据：

LFW 数据库：13233幅图像，5749个人，其中1680人有两幅以上图像，4069人只有一幅

AlexNet 参数数量：6 千万

GoogleLeNet 参数数量：7 百万

# DeepFace: Closing the Gap to Human-Level Performance in Face Verification

By FaceBook AI Research and Tel Aviv University

期刊：CVPR 2014

## 1 研究问题

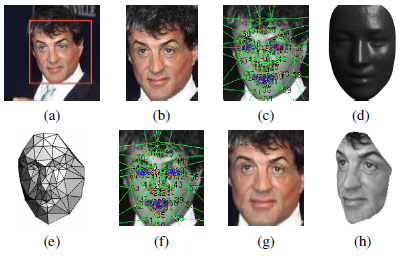
使用3D人脸模型对齐人脸，使用9层深度神经网络结合超大量人脸数据实现高精度人脸识别与人脸验证。

## 2 此文工作

正常的人脸识别流程：人脸检测 —> 人脸对齐 —> 特征表达 —> 分类

此文改进了人脸对齐、特征表达的方法。

### 2.1 人脸对齐



分为如下几步：

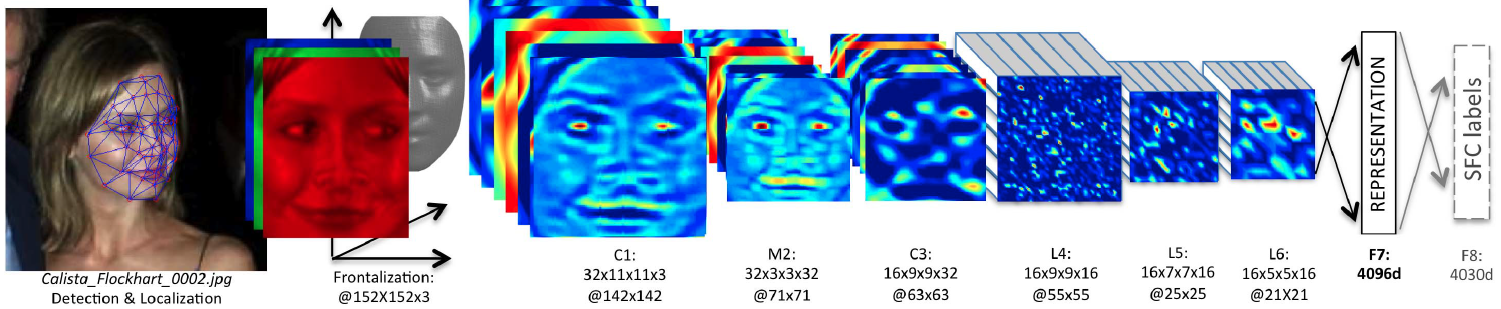
* 1. 检测人脸上的6个基准点
  2. 将这6个点通过旋转、缩放到对应的锚点，得到2D对齐人脸。但是2D对齐不能完成“平面外”的旋转
  3. 在2D对齐人脸上检测67个基准点，将这些点Delaunay三角化，轮廓处添加三角形避免不连续
  4. 将三角化后的人脸转换成3D形状
  5. 三角化后的人脸变为有深度的3D三角网
  6. 将三角网做偏转，是人脸的正面朝前
  7. 最后放正的人脸
  8. 可以变换成不同角度的人脸，但是论文中并没有用到

经过人脸对齐，可以认为人脸的特征固定在某些特定的位置上，便于深度网络的学习。

### 2.2 人脸特征表达

在此之前，人脸特征一直由人造特征表达，如LBP，Haar-Like 等。由于大量人脸数据的获取，基于学习的方法超过了传统的人工特征。此文采用深度神经网络提取特征。

#### 2.2.1 网络结构



经过3D对齐后，形成的图像都是152152，网络结构（共1.2亿个参数）：

* Conv：32个11113的卷积核
* Max-pooling：3 3，stride = 2
* Conv：16个9932的卷积核，到这一层是提取出一些低层特征，比如简单的线或纹理
* Local-Conv：16个9916的卷积核，Local 表示参数不共享
* Local-Conv：16个7716的卷积核，参数不共享
* Local-Conv：16个5516的卷积核，参数不共享
* Fully-connected： 输出为4096维。
* Softmax： 输出为4030维

另外一些细节：激活函数：ReLU max(0,x)、 产生的网络非常稀疏、采用L2 正则化

L4、L5、L6参数不共享的原因：图像是已经对齐的，每个特征处在固定的位置上，不同的人脸区域（眼睛和眼皮）具有不同的特征，不能采用相同的参数；而且这样做L6的输出也是由输入层较大区域（74743）卷积得到的，可以获得全局信息的综合；不采用参数共享，需要训练的参数量大大增加，因此需要很大的数据量，这个条件他们团队是可以达到的，400万张人脸图像，4030个人。

思考：传统的网络使用参数共享，每个卷积核都只提取同一特征，但是这种网络基本上都采用很多个卷积核，可以提取多种特征。在此方法中，由于已经固定了人脸上的特征位置，不同位置特征不同，需要不同参数，没有必要用参数共享但是多卷积核的方式，这个网络得到的特征图较少，可以加快检测速度并减少内存使用。

### 2.3 人脸验证

给定提取出的两个人脸图像特征、，判断是否为同一人.

#### 2.3.1 distance

相似性度量：

其中权重由SVM 训练得到，元素为

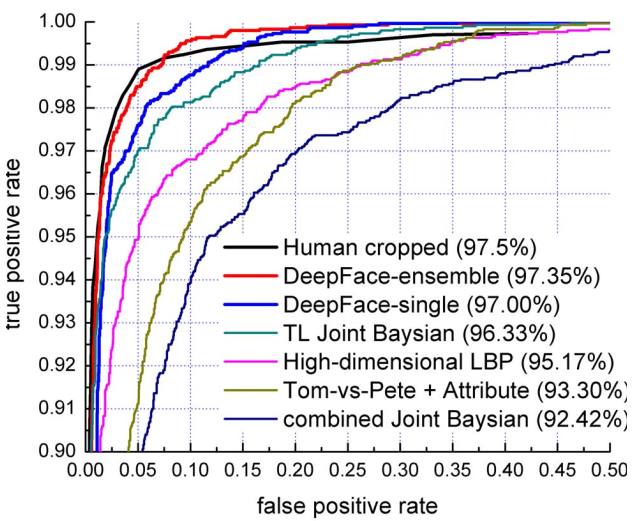
#### 2.3.2 Siamese network

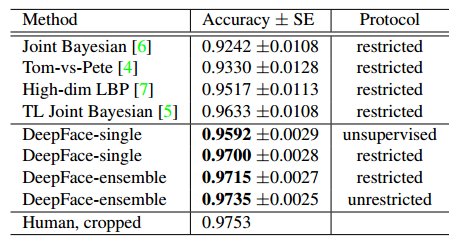
2005年一篇文章提出的网络，特点是接受两张图片作为输入，有如下特点：

* 如果样本对的标记一致，距离很近
* 如果样本对的标记不一致，距离大于margin

距离度量方式：

### 2.4 实验结果（LFW）97.25%





## 3 总结

意义：此文是第一个真正将大数据和深度神经网络应用于人脸识别和验证的方法；人脸识别精度接近人类水平。

缺点：方法非常粗暴，直接采用海量数据训练网络得到特征；需要3D对齐，步骤繁琐复杂。

# Learning Deep Face Representation

By face++ team

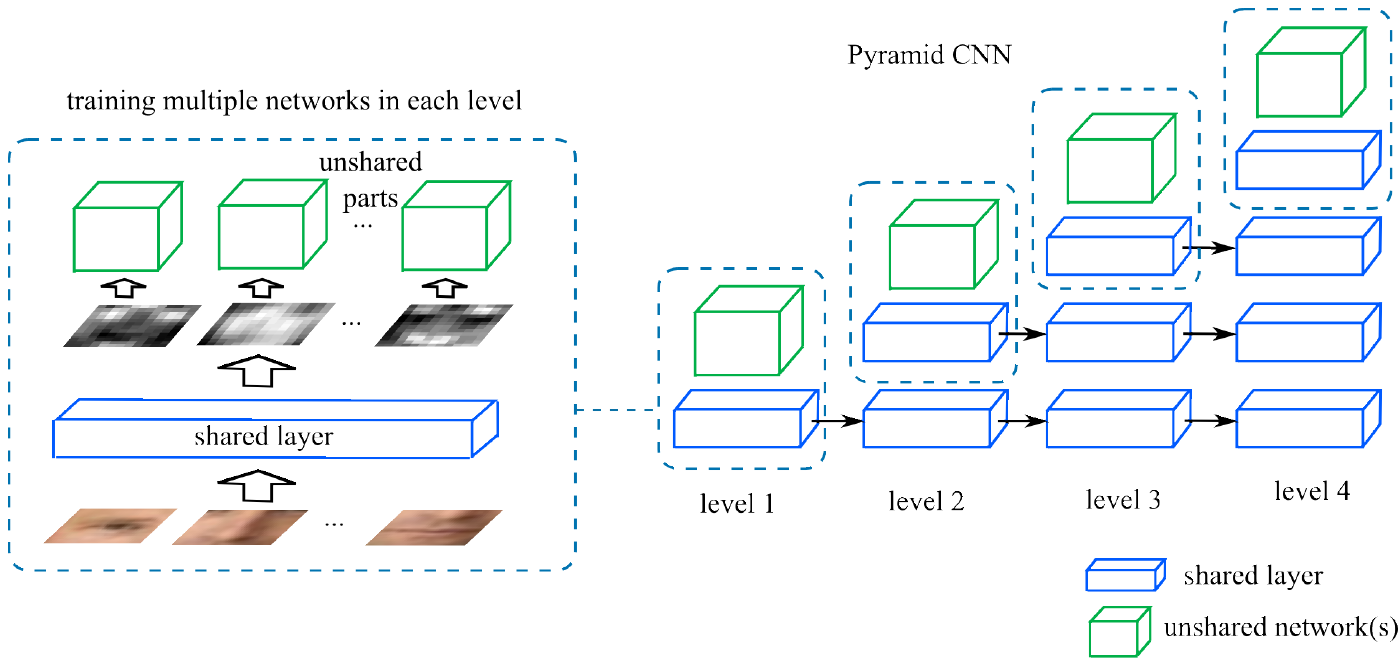
arXiv2014

## 1 研究问题

使用深度神经网络提取特征，提出金字塔神经网络。

## 2 此文工作

### 2.1 金字塔神经网络



网络结构（论文里没有讲清楚具体细节，只给了大概结构）：

* 网络为Siamese network
* 输入为不同的图像patch， 也就是不同的人脸部位
* 整个网络由很多个level 的网络组成
* 每个level 上的网络由共享的层和不共享的层组成，共享的层由之前的level 训练得到，此level 只训练不共享的层
* 每个level 有多个神经网络，它们对应于不同的patch 但是它们的第一层的参数是相同的
* 最后的特征为所有网络输出连接组成（训练时），这样的特征为多尺度特征，也是过完全特征，但是文中说明了过完全对结果有好处，而且特征可以通过PCA 降维，降维后的特征分类识别能力也很强（即使降到8维，正确率也能达到80%以上）
* 检测时使用的特征是最后一个level 的输出

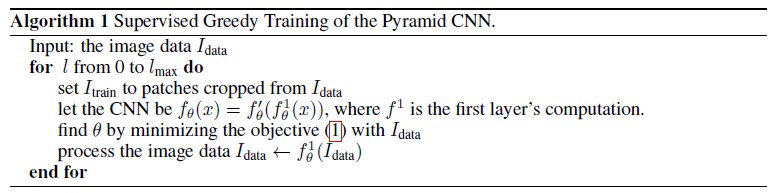
采用金字塔网络的原因：

* 加快网络训练
* 获得多尺度人脸特征

网络训练的损失函数：

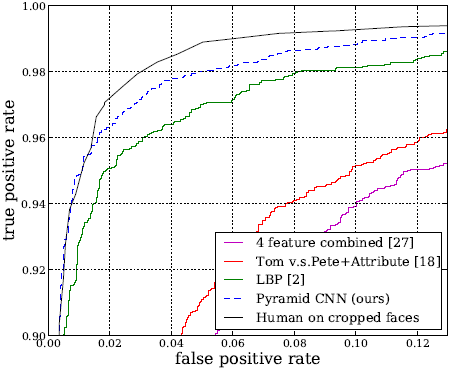
使同一个人的距离相近，不同人的距离相远。

网络的训练方法：

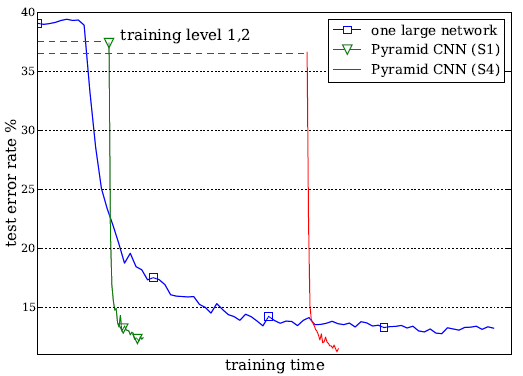


### 2.2 实验结果

准确率（LFW）97.27%



训练时间与准确率：（证明他们的方法可以加快训练）



## 3 总结

优点：定义了一个新的神经网络的结构：金字塔神经网络，并且人脸识别效果相较于传统方法有巨大提高。文章最后总结：由于人脸是高度结构化的图片，所以可以应用金字塔神经网络，但是普通物体具有很大的可变性，不太好应用。

缺点：没有说清楚网络结构和参数，也没有讲清楚训练数据集的来源与大小。

# Surpassing Human-Level Face Verification Performance on LFW with GaussianFace

By Chinese University of Hong Kong

CVPR2014

此文未细读，目前只知道此文是利用高斯过程度量两图片之间的相似性实现高精度人脸验证。

未使用深度神经网络。

训练数据集为2万，实现精度：98.52%

# Deep Learning Face Representation from Predicting 10,000 Classes

By Chinese University of Hong Kong

CVPR2014

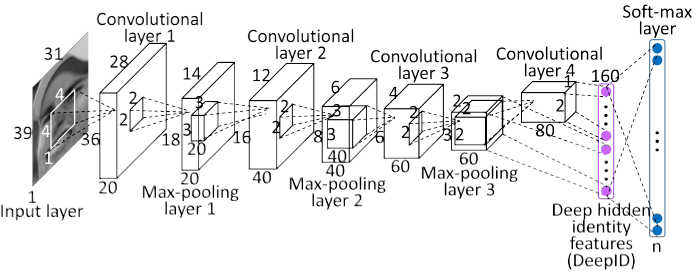
## 1 研究问题

主要通过神经网络学习图像特征，通过分类器验证人脸。

## 2 此文工作

### 2.1 人脸特征表达

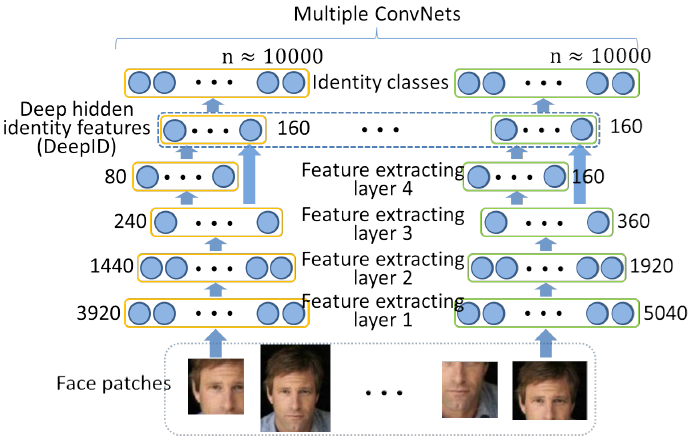
#### 2.1.1 网络结构



* 一共四个卷基层，前三个卷基层后都有Max-pooling
* 在最后一个隐藏层，是由Max-pooling layer3 + Convolutional layer4 全连接而得，为160维。这样的连接方式既考虑到了局部的特征，也考虑到全局特征
* 最后的soft-max训练时使用，检测时不用，而如果使用soft-max作为特征使用，效果很差
* 用于人脸识别和人脸验证的特征是最后一个隐藏层，称为DeepID。

#### 2.1.2 人脸特征





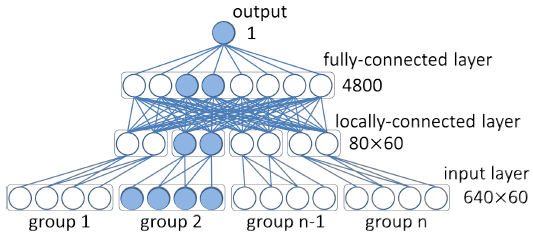
* 将输入图片缩放为三个比例尺大小，并且分为彩色和灰度图像，最后还将图片切分为10个patch， 所以最后输入的patch 数为：3210=60。
* 对每一种patch 都训练一个卷积网络，一共需要训练60个卷积网络。
* 所有patch 输出的特征连接起来，一共为19200（160260），再利用PCA降维到150维用于人脸验证与区分。

### 2.2 人脸验证方法

#### 2.2.1 联合贝叶斯

通过类内方差与类间方差之间的关系判断两个特征是否为一张脸。有一些推导，具体过程不详述。

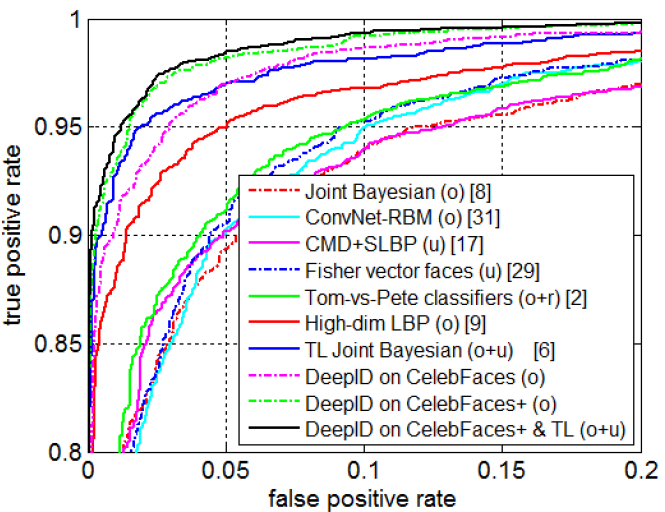
#### 2.2.2 神经网络



分成60个group，每个group对应一个patch，其中有16022=640维。输出为单一特征，表示相似性，由sigmoid生成。

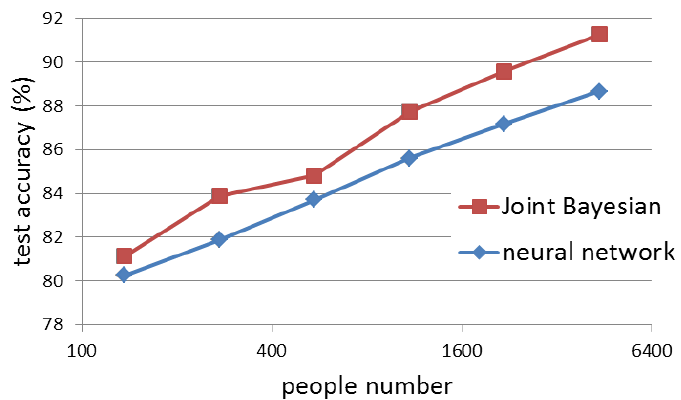
### 2.3 实验与讨论

训练数据：CeleFaces+， 有10177人，202599张图片；8700人训练DeepID，1477人训练联合贝叶斯分类器。准确率达到97.45%



其他一些结论：

* 使用multi-scale patch 的卷积网络效果比只用一张图片好
* 训练样本中的身份数量越多，训练得出的模型越优



## 3 总结

提出将人脸图像分为多个patch， 对每个patch 分别训练神经网络，这样的效果明显优于纯粹输入图片训练。进一步提高了精度。

# Deep Learning Face Representation by Joint Identification-Verification

By Chinese University of Hong Kong

CVPR2014

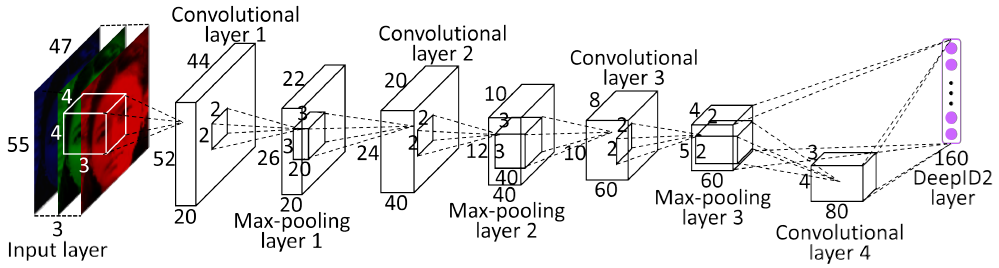
## 1 研究问题

在DeepID 的基础上添加验证信号，实现更高精度的人脸识别与验证。

## 2 此文工作

主旨思想是添加验证信号，减少类内方差，增大类间方差。

### 2.1 网络结构



基本上同DeepID， 但是网络的数量增加到200个。

### 2.2 网络训练

网络的训练由两类信号控制：身份信号 + 验证信号。身份信号指的是对应到训练样本中的身份（ID），验证信号指的是两张图片是否为同一张脸。

身份信号：

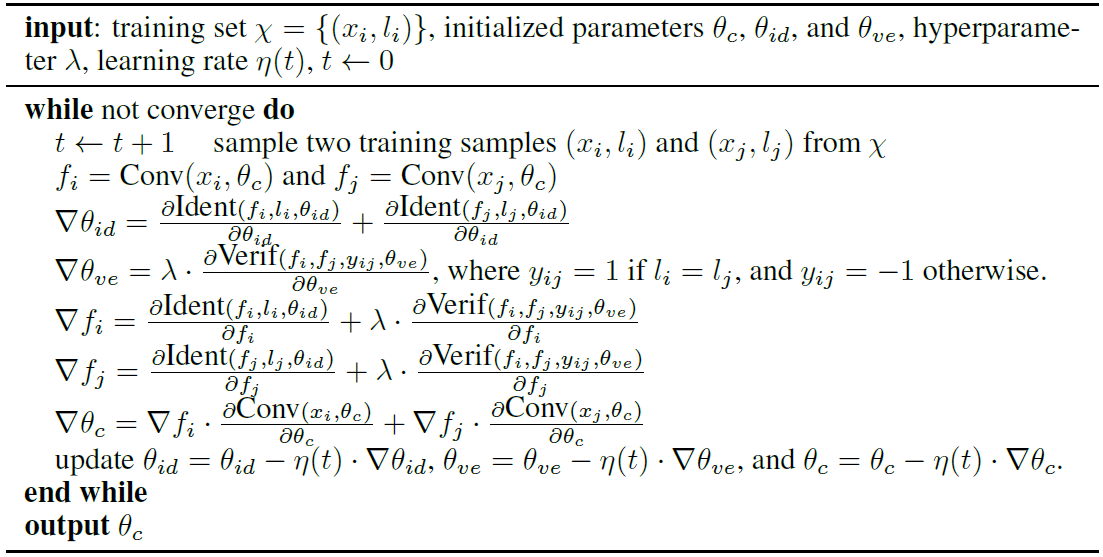
其中， 是DeepID2特征向量， 是目标类别， 表示softmax 参数， 表示预测分布。

验证信号1：

验证信号2：

训练的目标是学习到卷积参数， 和 都只是训练时用来得到反向传播误差的。

训练过程：

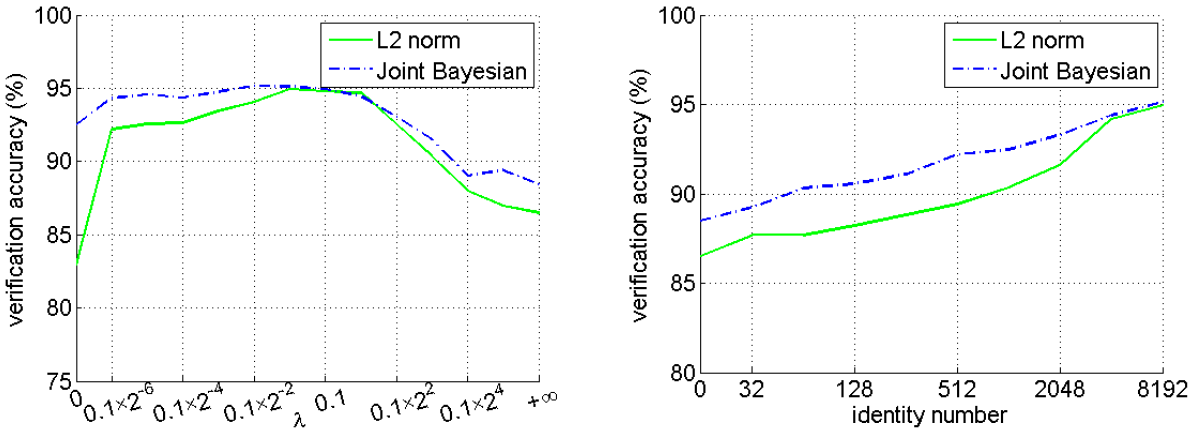


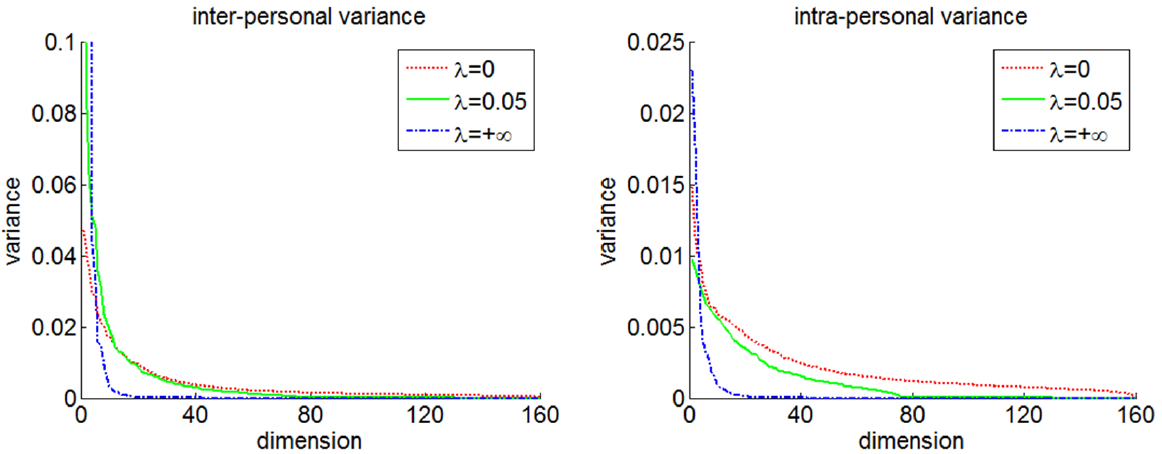
### 2.3 人脸特征

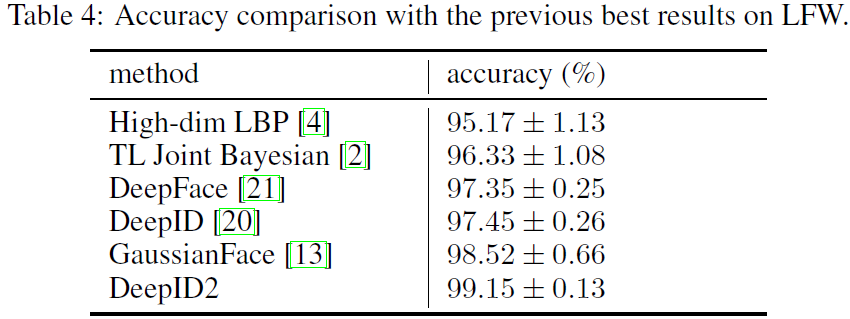
首先提取21个人脸特征点，将人脸对齐，然后通过尺度、切片、位置、通道、水平翻转得到400个patch， 通过200个卷积网络得到200个160维的特征。然后通过前向贪婪算法选取其中的25个patch 的输出作为特征，最后通过PCA降维到180维，即为最终利用的特征DeepID2。最后通过联合贝叶斯或L2范数进行人脸验证。

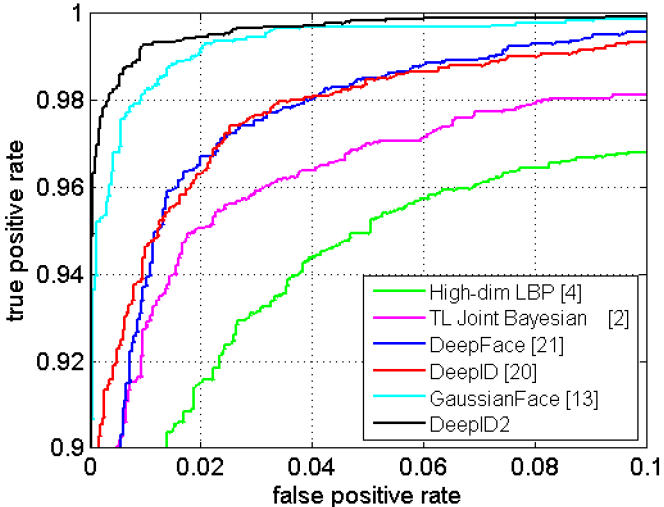
### 2.4 实验与讨论

实验数据：训练数据：CeleFaces+， 有10177人，202599张图片；8192人训练DeepID，1985人训练联合贝叶斯分类器。准确率达到99.15%









## 3 总结

通过添加验证信息，显著提高了人脸验证的水平。再一次刷新了LFW 数据库上的正确率。

# Deeply learned face representations are spars, selective, and robust

By Chinese University of Hong Kong

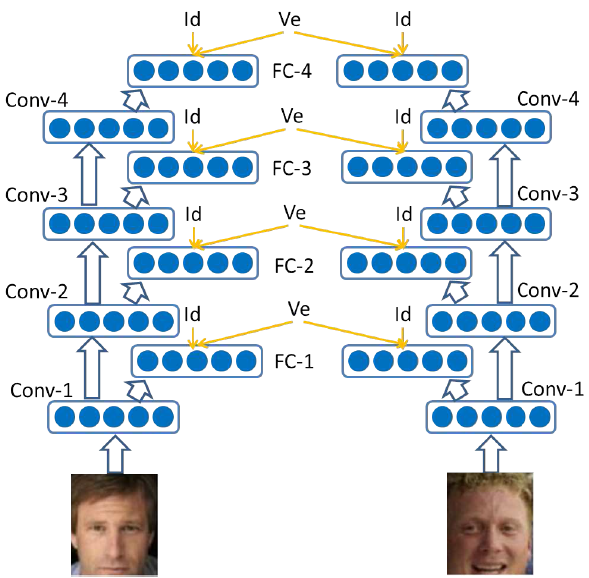
CVPR2014

## 1 研究问题

在DeepID2的基础上进一步改善网络和结果，并分析网络神经元的特性。

## 2 此文工作

### 2.1 网络结构及改进

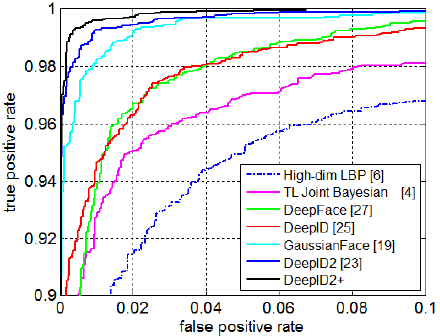
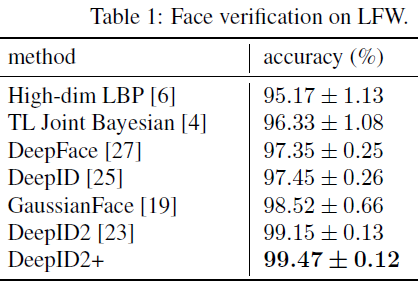


相比于DeepID2的改进：

* 四个卷积层都增加到128个特征图
* 特征输出增加到512维
* 验证信号作用于每一个max-pooling 层特征
* 训练数据增加到1.2万个身份29万张图
* DeepID 为所有max-pooling 层特征相连
* 有非常多的patch， 仍然选取25个patch

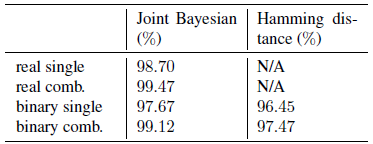
### 2.2 实验与讨论

#### 2.2.1 实验结果（LFW 99.47%）

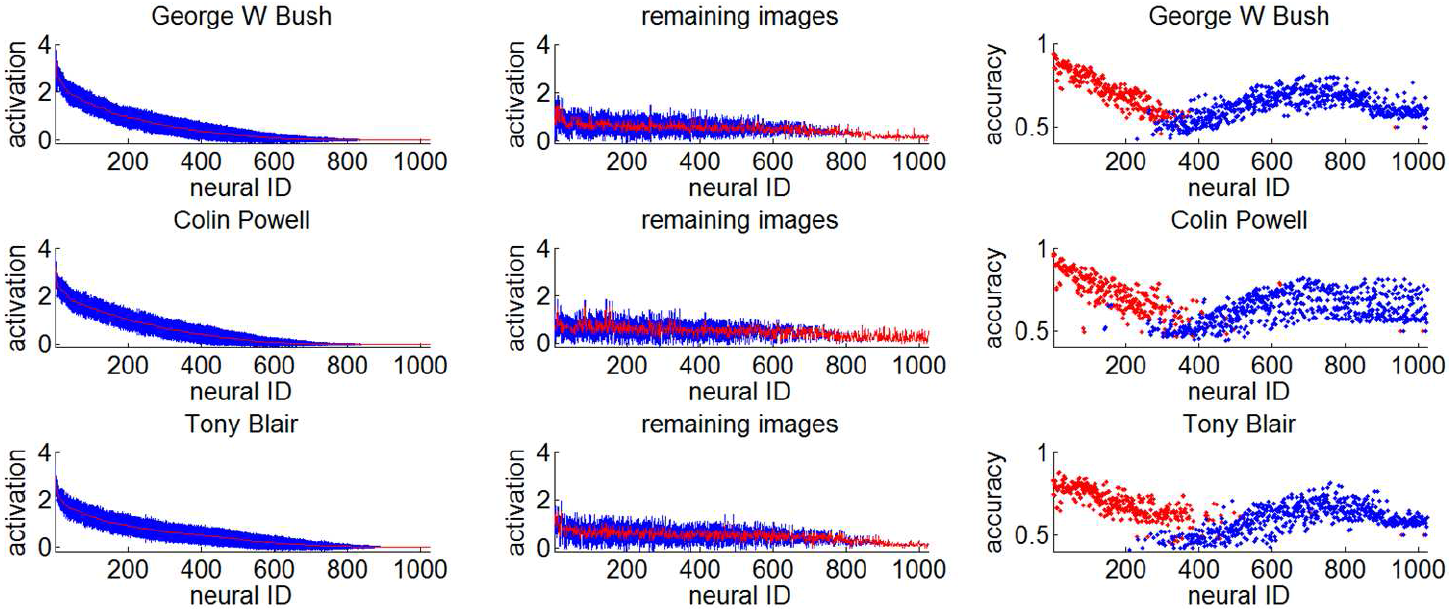
#### 2.2.2 适度稀疏与二值性

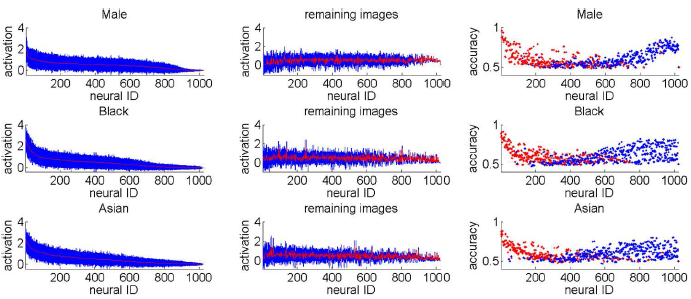
DeepID2+ 对每个人，最后的DeepID 层都大概只有半数的单元是激活的，半数为不激活的。不同的人，激活的单元不同。基于此性质，使用阈值将输出的512维向量进行二值化处理，发现进度降低有限。二值化的好处：通过计算汉明距离就可以进行检索，计算加快。



#### 2.2.3 特征区分性

存在某个神经元，只使用普通的阈值法，就能针对某个人得到97%的正确率。不同的神经元针对不同的人或不同的性别、年龄都具有很强的区分性。

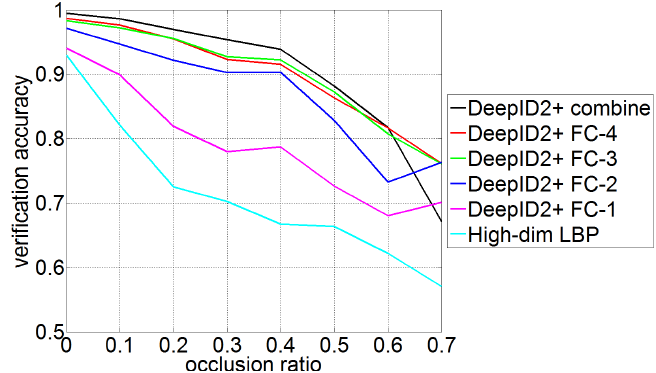


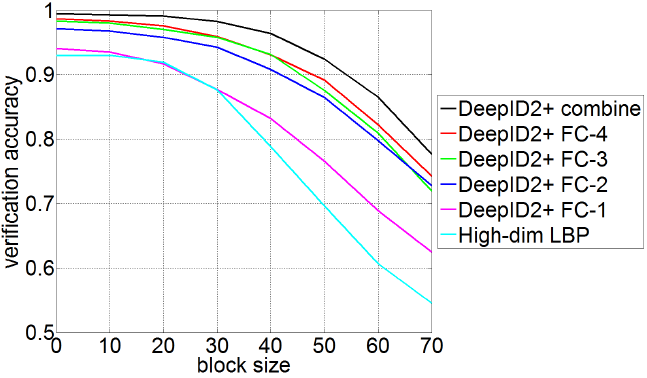


#### 2.2.4 遮挡鲁棒性

在训练数据没有被遮挡的情况下，DeepID2+ 具备对遮挡的鲁棒性，两种遮挡方式：







## 3 总结

进一步改进了DeepID 系列，精度进一步提高。并且探讨验证了输出神经元的适度稀疏性、特征区分性和遮挡鲁棒性。

# FaceNet: A Unified Embedding for Face Recognition and Clustering

By google Inc.

arXiv 2015

## 1 研究问题

人脸识别、人脸验证、人脸聚类

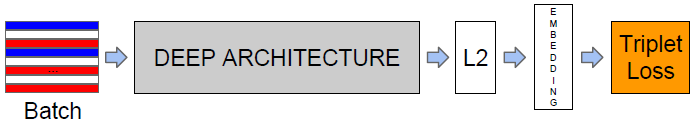
## 2 此文工作

传统的方法：首先利用softmax 损失函数训练网络，然后从网络中抽取某些层或输出层作为特征，然后基于此特征训练分类器，对人脸进行识别、验证。

FaceNet 则认为最后的分类器其实也可以用网络模拟。

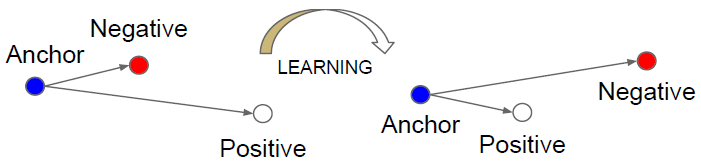
FaceNet 采用三元损失函数训练网络，直接进行端到端学习一个图像到欧式空间的编码方式，然后基于这个编码再做人脸识别、人脸验证和人脸聚类。

### 2.1 模型结构



去掉了最后的softmax ，经过L2归一化，得到特征表示，再基于这个特征计算三元组损失。

### 2.2 Triplet Loss



三元组：anchor、 positive、 negative。 目标：

其中， 为margin， 为anchor 图片的特征…

最小化损失函数：

### 2.3 Triplet Selection

很少的数据就能产生很多的三元组，选取合适的三元组可以加快收敛速度。

最粗暴的方式是对每个样本，找到离它最近的反例和最远的正例，然后优化，但是这样会有两个问题：耗时，比训练时间还长；容易受不好的数据主导。

采用的方法：

* 生成mini-batch， 每个mini-batch 中每个人平均有40张图片，然后随机加入一些反例
* 生成triplet， 不是找到所有的艰难正样本，而是找到所有的anchor-positive 对，然后根据这些特征对，找到semi-hard-negative

### 2.4 网络结构

此文中比较了好几个网络，有些网络很深，参数较多（1.4亿），也有网络较浅，参数较少（4.3百万，此网络可用于手机）。此文重点不在网络，可视为黑箱子。

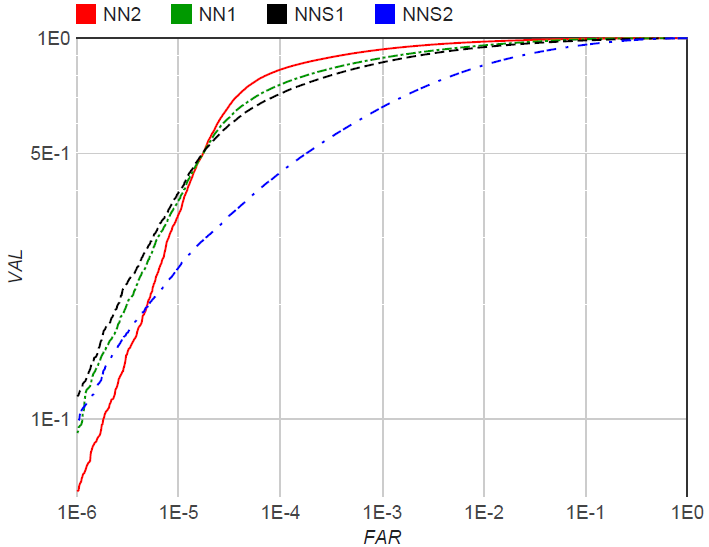
### 2.5 人脸识别、人脸验证、人脸聚类

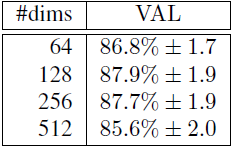
当完成端到端学习到图像到欧式空间的编码，则可以轻松完成以下三个任务：

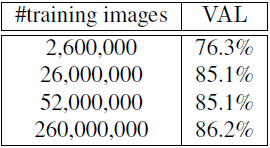
* 人脸识别：成为一个k-NN 问题
* 人脸验证：对两个编码求欧氏距离，与阈值对比
* 人脸聚类：利用像k-means 之类的聚类算法

### 2.6 实验与讨论

训练数据：100-200M张图像，分布在8M个人上。恐怖！







在LFW 上人脸不对齐，效果在98.87%左右，如果使用额外的人脸对齐，效果在99.63%，优于DeepID

## 3 总结

抛弃了softmax， 创造性的采用triplet-loss， 得到特征的欧式编码，使得人脸识别、验证、聚类大大简化。但是其超大量的训练数据他人难以效仿。

# Deep Face Recognition

Visual Geometry Group

Department of Engineering Science

University of Oxford

2015

## 1 研究问题

数据采集、人脸识别与验证。

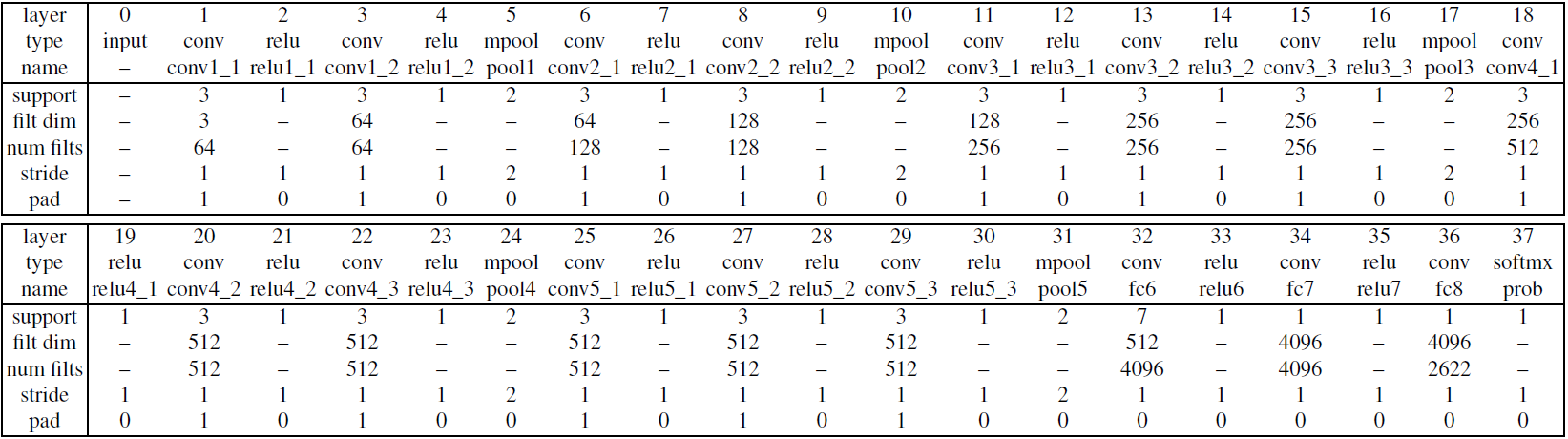
## 2 此文工作

此文有两个主要工作：如何在网络上搜集需要的人脸数据；训练深度网络完成人脸识别与验证。

数据搜集忽略不讲，可自行查阅。

### 2.1 网络结构

此文采用的网络结构为VGGNet， 网络结构如下：

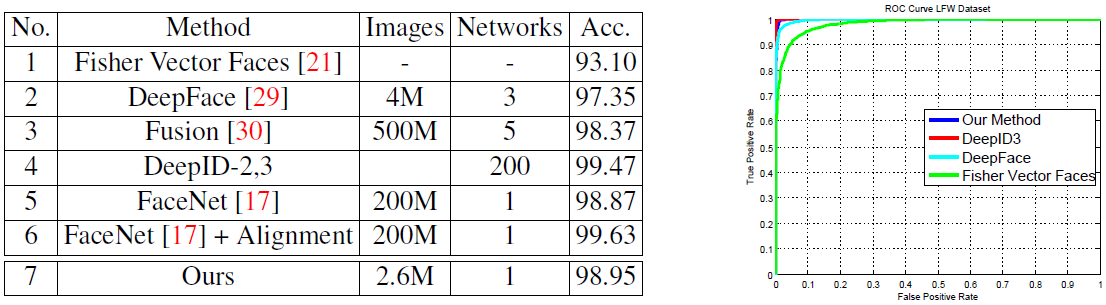


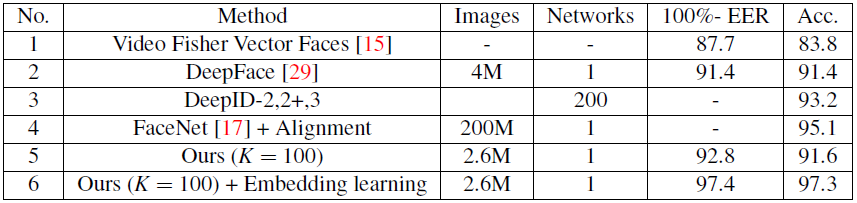
首先使用自己收集到的数据利用softmax 训练，得到的网络即可实现人脸识别。

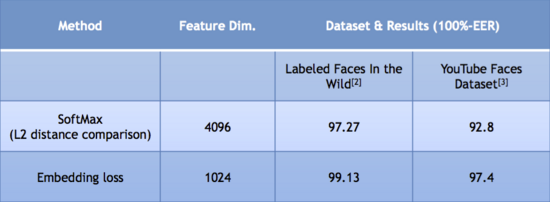
### 2.2 Triplet Loss 进行人脸验证

将上述网络最后输出特征进行L2归一化得到，并将其投影到另一空间。通过最小化以下三元损失函数训练

### 2.3 实验与讨论







此文使用比其他方法少得多的数据得出了更好的结果。

## 3 总结

此文首先介绍了在网络上获取大量的数据的方法，然后在他们自己获得的数据的基础上训练vggface 实现人脸识别， 并将其中最后一层作为特征训练Triplet-Loss 进行人脸验证。在较少的数据的基础上达到了前人海量数据才能达到的效果。

# 总结

人脸识别最新进展：



目前人脸识别与验证的通用方法：

利用深度卷积网络提取图像特征，再利用联合贝叶斯、距离度量、SVM等一系列的分类方法对特征分类验证。

深度网络是手段，特征学习是目的

目前在人脸识别领域，机器基本上超过人类，接下来研究方向主要在于人脸表情识别。