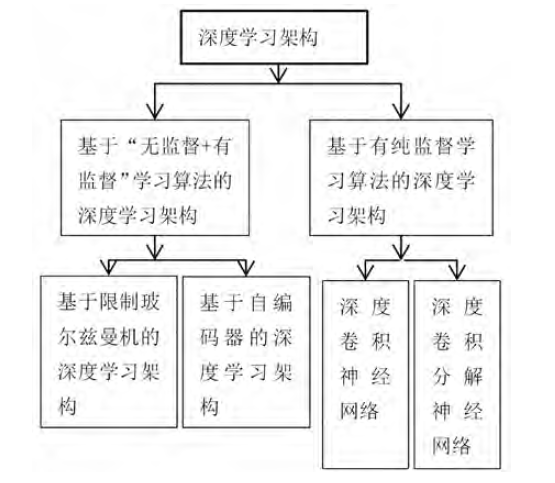
调研报告

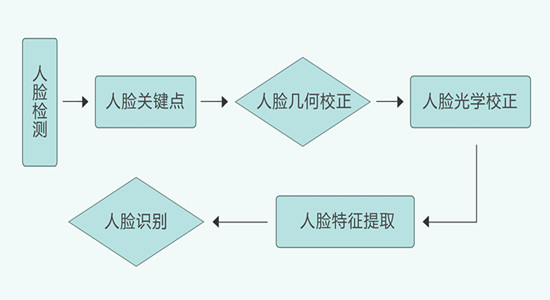
目前人脸识别的主要研究方法有3种。① 基于模板匹配的方法。模板匹配是指固定模板匹配和形变模板匹配。固定模板匹配是通过计算人脸图像和分布模型之间特征向量的差值来识别人脸。形变模板匹配是通过参数化的简单人脸图元模型来识别人脸；② 基于主分量分析的方法。主分量分析的实质是对人脸图像进行压缩，用一个低维向量来代替人脸图像，该向量包含原始图像的重要信息即特征脸，利用该方法可以重建和识别人脸；③ 基于神经网络的方法。神经网络是一种基于样本统计模型的方法，通过监督学习方法对人脸样本进行训练，并根据提取的特征构建分类器，从而对测试样本进行人脸识别。一般在实际运用中，会综合以上几种方法。

自从2006年Hiton提出深度学习的概念以来，对于深度学习的研究便广泛进行，在理论和运用方面都有巨大进展。深度学习的主要框架如下。无监督+有监督的有受限波尔兹曼机和自动编码机两种框架。自动编码机又拓展为稀疏自动编码机（降低隐层维度）和降噪自动编码机（加入随机噪声）。纯有监督的主要是卷积神经网络。在实际的运用中主要还是采用深度卷积神经网络的模式。



人脸识别的技术运用广泛，在追踪嫌犯、养老生存认证、考生身份验证、金融银行、交通、社区、景点检票、门禁、考勤等各种各样的社会化场景中都可以看到人脸识别技术的应用。

当今的人脸识别系统虽然包含无数细节和各种工程技巧，但大的系统框架不外乎如下图所示的流程：一张人脸图片输入后，需要先找到人脸的位置（人脸检测），然后在这个基础上定位人脸关键点的位置（如眼睛中心或嘴角等），每个系统关键提取的数量相差很大，有的只有左右眼睛中心两个点，有的可能多达近百个点。这些点的位置一是用来做人脸的几何校正，即把人脸通过缩放、旋转、拉伸等图像变化变到一个比较标准的大小位置。这样待识别的人脸区域会更加规整，便于后续进行匹配。同时，现在的实际系统一般也都配有人脸光学校正模块，通过一些滤波的方法，去除一些对光照更加敏感的面部特征。在做完这些预处理之后，就是从人脸区域提取各种特征，包括LBP、HOG、Gabor等。最终相关的特征会连接成一个长的特征向量（Feature Vector），然后匹配出人脸的相似度，根据相似度的大小，系统会判定两张图片到底是不是一个人。人脸识别技术是一个系统链条较长，较为有技术门槛的领域。因为这条流水线的每个环节可能都会严重影响最终系统性能，所以一个好的人脸识别技术需要在各个环节上追求细节，最终才有可能在最后的人脸识别精度上有出色的表现。



接下来介绍几个具有代表性的研究人脸识别的企业及主要算法。

Deepface：Deepface是facebook公司研发的人脸识别算法。DeepFace利用面部多点的稠密对齐，借助于3D模型对有姿态的人脸进行校正，同时利用一个9层深度卷积神经网络在400万规模的人脸库上训练了一个深度人脸表示模型，在LFW数据集上取得了97.25%的平均精度。

DeepID：DeepID是香港中文大学汤晓鸥课题组发明的一种人脸识别算法。孙祎团队研发的技术和产品已经在当今备受关注的互联网金融领域被应用于“人脸在线身份认证”，让用户无需面签即可通过人脸识别在手机端认证身份，继而开户、转账、借贷，这项身份认证服务已达到了百万级/天的调用量。其经过了三代的发展。第一代DeepID结构与普通的卷积神经网络的结构相似，但是在隐含层，也就是倒数第二层，与Convolutional layer 4和Max-pooling layer3相连，鉴于卷积神经网络层数越高视野域越大的特性，这样的连接方式可以既考虑局部的特征，又考虑全局的特征。DeepID2在DeepID的基础上添加了验证信号。具体来说，原本的卷积神经网络最后一层softmax使用的是Logistic Regression作为最终的目标函数，也就是识别信号；但在DeepID2中，目标函数上添加了验证信号，两个信号使用加权的方式进行了组合。第三代DeepID2+有如下贡献，第一点是继续更改了网络结构；第二点是对卷积神经网络进行了大量的分析，发现了几大特征，包括：1神经单元的适度稀疏性，该性质甚至可以保证即便经过二值化后，仍然可以达到较好的识别效果；2高层的神经单元对人比较敏感，即对同一个人的头像来说，总有一些单元处于一直激活或者一直抑制的状态；3DeepID2+的输出对遮挡非常鲁棒。

Facenet：FacenNet是Google开发的深度学习算法。与其他的深度学习方法在人脸上的应用不同，FaceNet并没有用传统的softmax的方式去进行分类学习，然后抽取其中某一层作为特征，而是直接进行端对端学习一个从图像到欧式空间的编码方法，然后基于这个编码再做人脸识别、人脸验证和人脸聚类等。

FaceNet算法有如下要点：

去掉了最后的softmax，而是用元组计算距离的方式来进行模型的训练。使用这种方式学到的图像表示非常紧致，使用128位足矣。

元组的选择非常重要，选的好可以很快的收敛。

三元组比softmax的优势在于 ：

softmax不直接，（三元组直接优化距离），因而性能也不好。

softmax产生的特征表示向量都很大，一般超过1000维。

商汤科技：SenseTime （商汤科技），是中国一家致力于计算机视觉和深度学习原创技术的创新型科技公司，总部位于北京。

Face++：Face++TM是北京旷视科技有限公司旗下的新型视觉服务平台，Face++TM平台通过提供云端API、离线SDK、以及面向用户的自主研发产品形式，将人脸识别技术广泛应用到互联网及移动应用场景中，人脸识别云计算平台市场前景广阔。和蚂蚁金服共同开发的Smile to Pay，可以进行扫脸支付。Face++和蚂蚁金服今年开发的人工智能机器人——蚂可，于6月30日和“水哥”进行了一次网红脸识别的比赛。

近日，CVPR2016刚刚召开，会议上深度学习基本占据了主导，成了计算机视觉研究的标配。以下是关于人脸识别的几篇文章的概括。

商汤科技这次提交的其中一篇文章为Joint Training of Cascaded CNN for Face Detection。文章考虑了一个如何联合训练检测器的问题，可以认为是卷积神经网络方案中的一个，但是不局限于某个特定的 CNN 检测方法。不管是人工设计的特征的方法，还是基于卷积神经网络的检测算法，往往都需要做级联。而级联的缺点是不太好直接联合训练，这恰恰破坏了卷积神经网络可以端到端训练的性质。由于传统的级联训练往往是单步求解最优，得到的结果可能没有多步联合最优的效果好。为此，文章提出了一种可以联合训练整个级联分类器的一种方式。得益于神经网络可以反向传播的性质，给出了如何联合训练整个级联分类器。并且，我们说明了如何在简单的 cascade CNN 和更复杂一点的 faster-RCNN 上都可以用这种方式来联合训练，并且可以取得非常优异的性能。 特别是当网络的规模比较小的时候，这种联合训练的方法提升非常大。这个工作做完之后，我们进一步大幅度提升了检测的速度和精度，并把完全基于 CNN 的方法做到了包括 ARM，CPU 和 GPU 各个平台上，并且跑到了非常好的速度。如今，商汤科技基于深度学习的人脸检测技术适配于多个品牌不同型号的手机上，支持着几十个各类 APP 的人脸分析功能；在单核 CPU 上可以跑到上百 FPS，并每天支持着大量的金融人脸认证服务；在单个 GPU 上可以实时的处理超过 12 路以上的高清视频，并运用到很多的安防监控系统中；甚至不久的将来，还会集成到相机前端的人脸抓拍芯片中。

汤晓鸥课题组的Sparsifying Neural Network Connections for Face Recognition：提出通过稀疏神经网络的连接来提升人脸识别的效果。稀疏卷积网络通过迭代的方法来学习，每次将一个额外的层进行稀疏化，将整个模型在之前迭代所得的权重下再训练。一个重要的发现，一开始就使用稀疏卷积网络训练无法得到很好地人脸识别效果，但使用之前训练好的密集模型来初始化一个稀疏模型对学习人脸识别的有效特征很关键。本文还提出来一种基于相关性的权重选择标准，并从经验上证实了它每次迭代中从之前学习的模型中选择有益连接的有效性。选择合适的稀疏模型（大概为密集模型的26-76%），使用同样的数据集，它提升了最新的DeepID2+模型的人脸识别效果。他只需要12%的原来的参数就能达到基准模型的表现。

研究发现，神经元之间的相关系数比权重的大小或二阶导能更好地衡量神经元连接的重要性。

经过这几年的研究，在现有的人脸数据库如LFW中，现有的算法已经达到几乎完美的水平，识别能力已经超过人眼，但这并不能代表在现实中会有这样的效果。于是提出创建更大更有挑战性的数据库The MegaFace Benchmark。The MegaFace Benchmark: 1 Million Faces for Recognition at Scale：新建的数据库，包含690000人的1百万张照片。参加测试的算法，能够在LFW数据库（相当于只有10个干扰项）上达到95%以上的识别率，在这个有1M的干扰项中只达到35-75%的识别率，LBP和联合贝叶斯算法作为基准则降到了不足10%。如谷歌的FaceNet和FaceN这样在大的数据集上进行训练的算法，在MegaFace中测试时的效果更好。而且在Face-Scrub数据集中，在18M照片中训练的FaceN达到了和在500M照片中训练的FaceNet相当的效果。验证集和注册集之间人物的年龄差别对于识别影响很大。验证集和注册集之间的动作变化导致识别率减低，对于大规模的数据尤其重要。

WIDER FACE: A Face Detection Benchmark。汤晓欧课题组也提出了一个用于人脸识别的数据库---WINDER FACE,比现有的数据库大10倍多，包含丰富的注释，包括地点，姿势，活动类别，和脸部边框。用这个数据库训练，当前最好的算法也只能达到70%的正确率。这个数据库更像现实中的人脸采集状况，也是将来研究的方向。

本次CVPR最佳论文Deep Residual Learning for Image Recognition。这是微软方面的最新研究成果， 在第六届ImageNet年度图像识别测试中，微软研究院的计算机图像识别系统在几个类别的测试中获得第一名。之前的研究发现，卷积网络的深度越深，训练效果越好。本文是解决超深度CNN网络训练问题，152层及尝试了1000层。本文主要解决的是随着深度的增大，训练误差也增加的问题。求取残差映射 F(x):= H(x)-x，那么原先的映射就是 F(x)+x，对原先的 平原网络（plain network）上加入快捷连接（shortcut connections）构成残差网络，以此随着网络深度的增加检测效果也更好。

Pose-Aware Face Recognition in the Wild。只要是来研究在对于不同的人脸姿态的识别，不同于其他的方法，要么通过训练大量的数据来希望找到一个模型来学习姿态的不变性，要么将图像标准化为一个正面姿态，该方法是使用多种姿态确定模型和提取的脸部图像来解决姿态的变化。用训练多个深度卷积神经网络来学习不同角度的人脸图像。

Face Alignment Across Large Poses: A 3D Solution。也是研究人脸对齐的文章。他主要考虑了人脸姿态特别偏的情况，会到时人脸的一些特征点看不到。于是他们提出一种三维密集脸部模型的方法，构建人的三维人脸，而不是用离散的特征点来对其，这样解决了局部特征看不到的问题。然后用卷积神经网络来拟合三维人脸模型。

FACE ALIGNMENT BY DEEP CONVOLUTIONAL NETWORK WITH ADAPTIVE LEARNING RATE。是在2016年ICASSP会议上的一篇文章。主要是研究使用深度卷积网络来进行人脸对齐。该方法是基于数据增加和自适应学习速率的对于人脸特征点的检测。数据增加解决缺少图像的问题，自适应学习来得到定位的最优解。该网络可以学习全局的高级特征，直接预测人脸特征点的坐标。各项测试表明，该方法效果很好，尤其在复杂环境，姿势，光照等情况下。

文章DEEP PAIR-WISE SIMILARITY LEARNING FOR FACE RECOGNITION。作者提出了一种新的人脸识别的模型。主要考虑到很多情况下充足的数据无法提供，那么现有的算法无法发挥很好地效果。该方法的输入时成对的图片，不像现有的深度学习模型来学习单个图片的特征，再进行分类，该方法是直接学习两张图片的相似性，直接作出判断。该算法通过融合两张图片，输入卷积网络，提取特征，但并不进行分类，而是直接计算相似系数，来判断两张图片是否是一个人。主要用来识别同一个人的人脸的不同姿势。

针对现有的人脸识别数据库做如下总结：

1.ImageNet

这个数据集包含约120万张训练图像、5万张验证图像和10万张测试图像，分为1000个不同的类别。每年的ILSVRC竞赛就是基于这个数据集。

2. CIFAR-10

CIFAR-10数据集包含60000个32\*32的彩色图像，共有10类。有50000个训练图像和10000个测试图像。

数据集分为5个训练块和1个测试块，每个块有10000个图像。测试块包含从每类随机选择的1000个图像。训练块以随机的顺序包含这些图像，但一些训练块可能比其它类包含更多的图像。训练块每类包含5000个图像。

3. CIFAR-100

数据集包含100小类，每小类包含600个图像，其中有500个训练图像和100个测试图像。100类被分组为20个大类。每个图像带有1个小类的“fine”标签和1个大类“coarse”标签。

4．COCO

是微软组件的一个新的图像识别、分割、和字幕数据集，包含300000多张图片，超过2百万个例子，80多类，每张图有5个注释，有100000人的关键点。

5.LFW

由美国马萨诸塞大学阿姆斯特分校计算机视觉实验室整理完成，共13233幅图像，其中5749个人，其中1680人有两幅及以上的图像，4069人只有一幅图像。图像为250\*250大小的JPEG格式。绝大多数为彩色图，少数为灰度图。

大多数图像都是由Viola-Jones的检测器得到之后，被裁剪为固定大小，有少量的，人为地从false positive中得到。

6. WIDER FACE

汤晓欧课题组也提出了一个用于人脸检测的数据库---WINDER FACE,比现有的数据库大10倍多，包含丰富的注释，包括地点，姿势，活动类别，和脸部边框。用这个数据库训练，当前最好的算法也只能达到70%的正确率。这个数据库更像现实中的人脸采集状况，其中的人脸形状小，有遮挡，姿势比较极端，也是将来研究的方向。

7. The MegaFace Benchmark：新建的数据库，包含690000人的1百万张照片。

8. CASIA-WebFace：包含10575人的494414张图片，公开使用。

9.SFC：Facebook的人脸数据库，包含4030人的4400000张照片，不公开

10. Youtube Face(YTF): 3425videos/1595人，非限制场景、视频，来源于Youtube上的名人，公开。

11. FaceNet:谷歌的数据库，拥有超过10M人的超过500M的照片，来源于互联网，不公开

12. IJB-A IAPRA：500人的25813张照片

13. FaceScrub：530人，约100k张图片，非限制场景

14. MultiPIE：337个人的不同姿态、表情、光照的人脸图像，共750k+人脸图像，限制场景人脸识别

15. CAS-PEAL：1040个人的30k+张人脸图像，主要包含姿态、表情、光照变化，限制场景下人脸识别

16. Pubfig：200个人的58k+人脸图像，非限制场景下的人脸识别

17. FDDB：2845张图片中的5171张脸，标准人脸检测评测集

针对现有的人脸识别算法做如下归纳：

1.Deepface：

Deepface是facebook公司研发的人脸识别算法。其实现的基本流程为：分为如下几步：首先是人脸对其，通过以下流程实现：

a. 人脸检测，使用6个基点

b. 二维剪切，将人脸部分裁剪出来

c. 67个基点，然后Delaunay三角化，在轮廓处添加三角形来避免不连续

d. 将三角化后的人脸转换成3D形状

e. 三角化后的人脸变为有深度的3D三角网

f. 将三角网做偏转，使人脸的正面朝前。

g. 最后放正的人脸

h. 一个新角度的人脸

总体上说，这一步的作用就是使用3D模型来将人脸对齐，从而使CNN发挥最大的效果。

第二步：人脸表示

经过3D对齐以后，形成的图像都是152×152的图像，输入到上述网络结构中，进行卷积神经网络处理，该结构的参数如下：

Conv：32个11×11×3的卷积核

max-pooling: 3×3， stride=2

Conv: 16个9×9的卷积核

Local-Conv: 16个9×9的卷积核，Local的意思是卷积核的参数不共享

Local-Conv: 16个7×7的卷积核，参数不共享

Local-Conv: 16个5×5的卷积核，参数不共享

Fully-connected: 4096维

Softmax: 4030维

第三步：人脸表示归一化

对于输出的4096-d向量：

先每一维进行归一化，即对于结果向量中的每一维，都要除以该维度在整个训练集上的最大值。

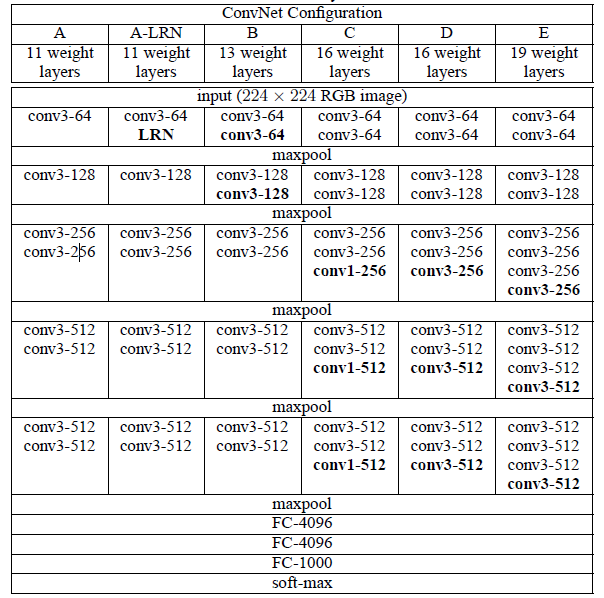
每个向量进行L2归一化

最后进行分类：

得到表示后，使用了多种方法进行分类：直接算内积；加权的卡方距离；使用Siamese网络结构。

2.VGG

模型框架为：输入为尺寸224X224的RGB图，由A-E5个卷积网络构成，深度由浅到深。所有的卷积核都是用很小的3x3,步长为1，只有一个是1x1的卷积核，可视为输入通道的线性变换。一共有5个池化层，不是每个卷积层后都有池化层。池化用2x2像素的窗口，步长为2.。最后三层是全连接层，前两层有4096个通道，最后一层有1000个通道分别对应1000个类别，最后一层用softmax分类。每个隐层用ReLU做输出函数，整个网络都不用LRN，因为没效果，还占内存消耗和计算时间。



训练的目的是为了最优化多项逻辑回归，通过基于BP算法的mini-batch gradient descent来实现。训练通过权重衰减和对前两层全连接层的dropout regularisation来调整参数。学习速率初始设为0.01，并以10倍减少当正确率不再提高时。一共衰减了三次，学习到370K次迭代后停止。

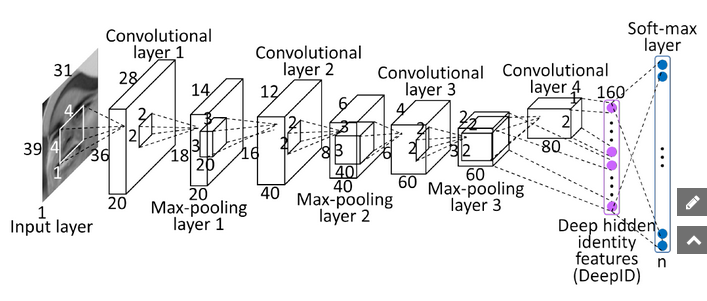
初始化网络的权重：首先用随机初始化训练结构A，因为网络较浅；训练其他深的结构，就用A来初始化他们的前四层和最后三个全连接层，中间层随机初始化，不改变预初始化的学习速率，让他们随着学习改变。随机初始化用，使用0平均和10-2的方差的标准正态分布。

有两种确定训练图像大小的方法，出于速度的考虑，采取单一大小的方式，并定为384.

3．DeepID系列

DeepID：

结构：由四层卷积神经网络构成，前三层后跟池化层，第三层池化层和第四层卷积层一起全连接够成最后的DeepID层，提取图片特征。最后用Soft-max来分类，结构如下：

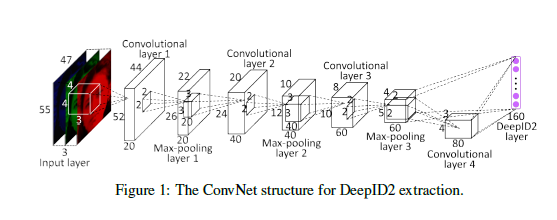


实现：

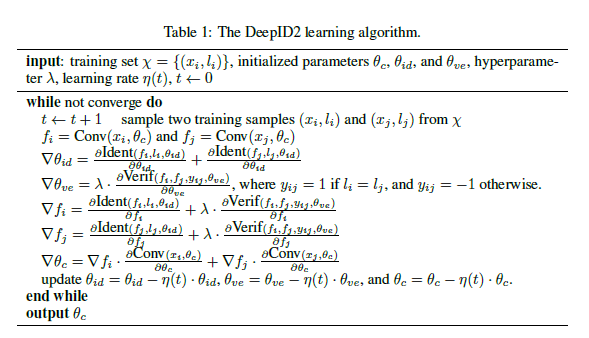
在DeepID的实验过程中，使用的外部数据集为CelebFaces+，有10177人，202599张图片；8700人训练DeepID，1477人训练Joint Bayesian分类器。切分的patch数目为100，使用了五种不同的scale。每张图片最后形成的向量长度为32000，使用PCA降维到150。如此，达到97.20的效果。使用某种Transfer Learning的算法后，达到97.45%的最终效果。

DeepID2:结构框架：输入为55x47的RGB图，网络包含4个卷积层，前三个后面跟着池化层。为了学到不同的高层特征，高层卷积网络权值不共享。尤其，第三层卷积网络权值只在2x2的局部区域共享，第四层本地连接层完全不共享。最后的一层是第三和第四两层卷积网络的全连接层，提取了160维的向量。使用ReLU作为激活函数。

另外，加上了face identification signal和face verification signal两个监督信号。face identification signal通过在DeepID2层之后加上n路的softmax层，通过训练最小化交叉熵损失，来确保正确分类。那么DeepID2层就要找到最具类间区分度的特征，这样就最大化了类间差距。face verification signal用来认准两个人是否是同一个人。通过L1/L2标准化或余弦相似性来衡量，通过训练，使DeepID2层对于同一个人尽量显示一致的特征，以此来减小类内差距。结构如下图：



训练过程：首先初始化权重参数，从训练集中随机取两个样例输入网络，计算各自的输出，采用随机梯度下降法来更新各参数，进行迭代，知道模型收敛，输出各权重参数。如下图：



人脸验证：

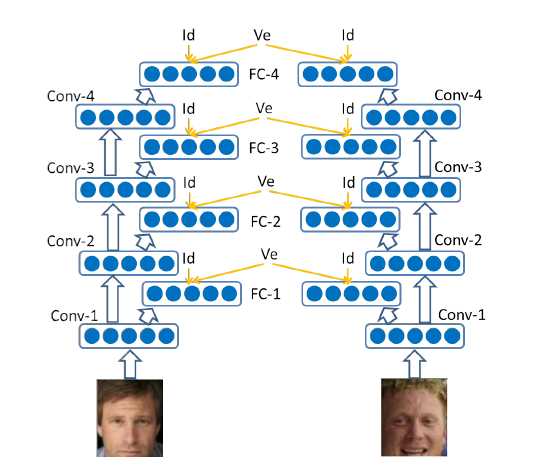
首先使用SDM算法对每张人脸检测出21个landmarks，然后根据这些landmarks，再加上位置、尺度、通道、水平翻转等因素，每张人脸形成了400张patch，使用200个CNN对其进行训练，水平翻转形成的patch跟原始图片放在一起进行训练。这样，就形成了400×160维的向量。

这样形成的特征维数太高，所以要进行特征选择，不同于之前的DeepID直接采用PCA的方式，DeepID2先对patch进行选取，使用前向-后向贪心算法选取了25个最有效的patch，这样就只有25×160维向量，然后使用PCA进行降维，降维后为180维，然后再输入到联合贝叶斯模型中进行分类。

DeepID2使用的外部数据集仍然是CelebFaces+，但先把CelebFaces+进行了切分，切分成了CelebFaces+A(8192个人)和CelebFaces+B(1985个人)。首先，训练DeepID2，CelebFaces+A做训练集，此时CelebFaces+B做验证集；其次，CelebFaces+B切分为1485人和500人两个部分，进行特征选择，选择25个patch。最后在CelebFaces+B整个数据集上训练联合贝叶斯模型，然后在LFW上进行测试。在上一段描述的基础上，进行了组合模型的加强，即在选取特征时进行了七次。第一次选效果最好的25个patch，第二次从剩余的patch中再选25个，以此类推。然后将七个联合贝叶斯模型使用SVM进行融合。最终达到了99.15%的结果。

DeepID2+:

结构：和DeepID2基本类似，区别在于将前三层的池化层和第四层卷积网络都和最后一层进行了全连接，每层提供128维的特征，一共组成512维的DeepID2+向量，对其添加face identification signal和face verification signal两个监督信号，等于是对每层都进行监督，增强了监督；还有就是加大了训练集，训练集将CelebFaces+和WDRef数据集进行了融合，共有12000人，290000张图片。结构如下：

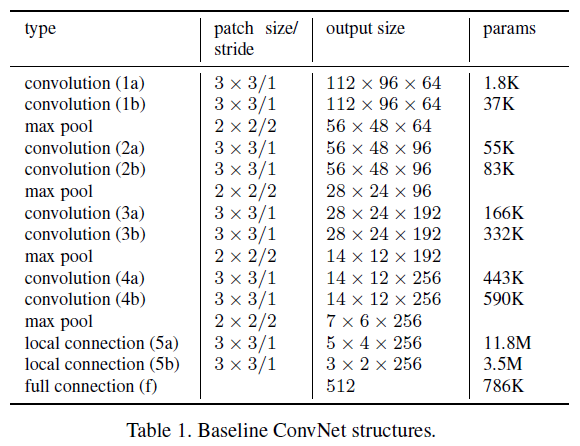


4. Sparsifying Neural Network Connections for Face Recognition

研究发现对于有挑战的人脸识别问题，神经元相关性的指标，用神经元联系的重要性衡量比权重的量级或二阶导更好。

结构：

先训练一个基准模型：基准模型和VGG类似，每两个卷积层跟一个池化层，主要区别在于最后两个卷积层替换为两个局部连接层。这是为了学习人脸不同部分的不同特征。第二个局部连接层后跟着一个512维的全连接层，全连接层的特征用来进行人脸识别。结构如下：



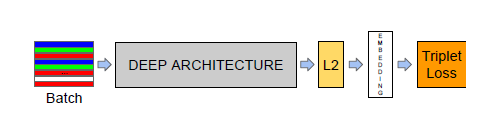
把Joint identification-verification supervisory signal加入到最后的全连接层来学习特征表述，来区分不同人脸，识别出同一个人脸，同样的监督信号也加入到之前的一些层，在学习阶段增强监督。用ReLU做所有卷积，局部连接和全连接层的激活函数。将30-50%丢弃率的Dropout learning用于最后一层局部连接层和全连接层。用同样的数据，单个的这个模型已经比单个DeepID2+效果更好。

对基准模型进行稀疏卷积化：从最后的全连接层开始，逐层稀疏化网络。因为最后几层有大量参数的冗余，减少他们不会影响网络的效果。用前一次的密集网络的权重参数来初始化此次的网络，再进行稀疏化。直到所有层都稀疏化。

5.FaceNet

FaceNet是谷歌提出的人脸识别算法。

结构：从图像中获取输入快，经过卷积神经网络来获取特征，但并没有使用softmax来分类，而是经过L2归一化后，用训练好的三元组计算三元损失，直接判断图片是否是一个人。结构如下图：



实施过程：

三元损失由三部分组成，需要被判断的图像（anchor），属于同一个人的另一张图像（positive），和不是同一个人的一张图像(negative).训练目的就是为了让所有anchor和positive之间的距离都要比anchor和negative之间的距离都要短，那么就能正判断出图中的人是否是同一个。

而三元组的选择很重要，选的好就能迅速收敛。文章提出两种方法：

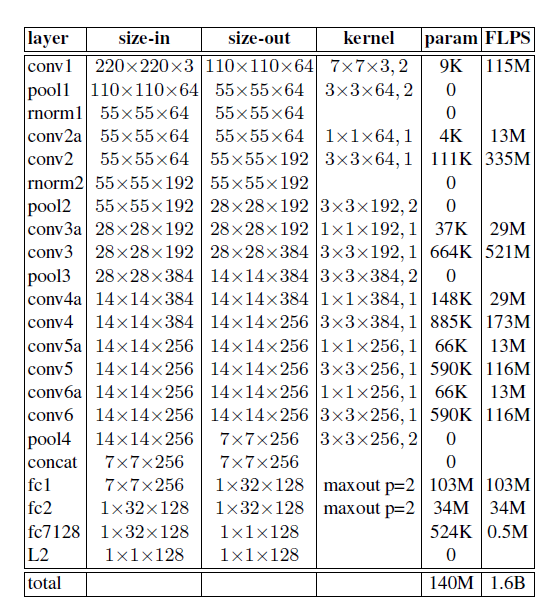
1.每N步线下在数据的子集上生成一些triplet

2.在线生成triplet，在每一个mini-batch中选择hard pos/neg 样例。

使用线上生成时，为了使mini-batch中生成的triplet合理，生成mini-batch的时候，保证每个mini-batch中每个人平均有40张图片。然后随机加一些反例进去。在生成triplet的时候，找出所有的anchor-pos对，然后对每个anchor-pos对找出其hard neg样本。这里，并不是严格的去找hard的anchor-pos对，找出所有的anchor-pos对训练的收敛速度也很快。

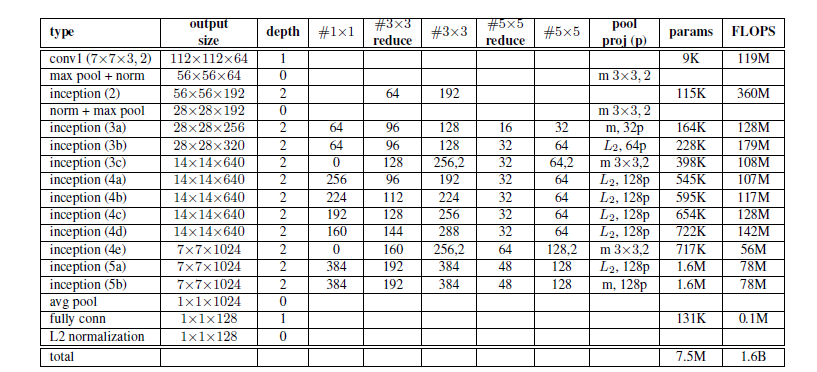
文章选择了两种深度卷积神经网络：

1. 第一种是Zeiler&Fergus架构，22层，140M参数，1.6billion FLOPS称之为NN1



2第二种是GoogleNet式的Inception模型。模型参数是第一个的20分之一，FLOPS是第一个的五分之一。针对Inception模型，提出了两个小模型NNS1和NNS2可以运用到手机上。保持结构一样，减少输入尺寸，提出NN3和NN4.

NN2模型如下：



参考文献：

1. Karen Simonyan，Andrew Zisserman. VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION

2. Yi Sun，Xiaogang Wang，Xiaoou Tang. Sparsifying Neural Network Connections for Face Recognition

3. Sun Y, Wang X, Tang X. Deep learning face representation from predicting 10,000 classes[C]//Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on. IEEE, 2014: 1891-1898.

4. Sun Y, Chen Y, Wang X, et al. Deep learning face representation by joint identification-verification[C]//Advances in Neural Information Processing Systems. 2014: 1988-1996.

4.Sun Y, Wang X, Tang X. Deeply learned face representations are sparse, selective, and robust[J]. arXiv preprint arXiv:1412.1265, 2014.

5. Taigman Y, Yang M, Ranzato M A, et al. Deepface: Closing the gap to human-level performance in face verification[C]//Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on. IEEE, 2014: 1701-1708.

6. Schroff F, Kalenichenko D, Philbin J. Facenet: A unified embedding for face recognition and clustering[J]. arXiv preprint arXiv:1503.03832, 2015.

7. Ira Kemelmacher-Shlizerman,Steven M. Seitz ,Daniel Miller,Evan Brossard. The MegaFace Benchmark: 1 Million Faces for Recognition at Scale

8. Iacopo Masi,Stephen Rawls. Pose-Aware Face Recognition in the Wild

9. Kaiming He Xiangyu Zhang .Deep Residual Learning for Image Recognition

10. Hongwei Qin, Junjie Yan,Xiu Li, Xiaolin Hu .Joint Training of Cascaded CNN for Face Detection