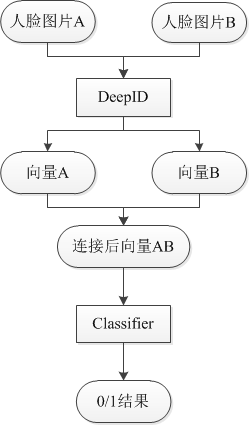
# DeepID人脸识别算法

## 问题引入及算法流程

http://blog.csdn.net/stdcoutzyx/article/details/42091205

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达到的效果都是在LFW数据集上，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。该数据集太小，很多identities都只有一张人脸，5000个人只有13000张图片。所以 DeepID引入了外部数据集CelebFaces和CelebFaces+，每次模型更新都会使用更大的数据集，这在后面介绍DeepID时再细说。

卷积神经网络在DeepID中的作用是是学习特征，即将图片输入进去，学习到一个 160维的向量。然后再这个160维向量上，套用各种现成的分类器，即可得到结果。DeepID之所以有效，首先在于卷积神经网络学习到的特征的区分能力 比较强，为了得到比较强的结果，DeepID采取了目前最常用的手法——增大数据集，只有大的数据集才能使得卷积神经网络训练的更加的充分。增大数据集有 两种手法，第一种手法，就是采集好的数据，即CelebFaces数据集的引入。第二种手法，就是将图片多尺度多通道多区域的切分，分别进行训练，再把得 到的向量连接起来，得到最后的向量。DeepID的算法流程如下：



在上述的流程中，DeepID可以换为Hog，LBP等传统特征提取算法。Classifier可以是SVM，Joint Bayes，LR，NN等任意的machine learning分类算法。

在引入外部数据集的情况下，训练流程是这样的。首先，外部数据集4:1进行切分，4那份用来训练DeepID，1那份作为训练DeepID的验证集；然后，1那份用来训练Classifier。这样划分的原因在于两层模型不能使用同一种数据进行训练，容易产生过拟合。

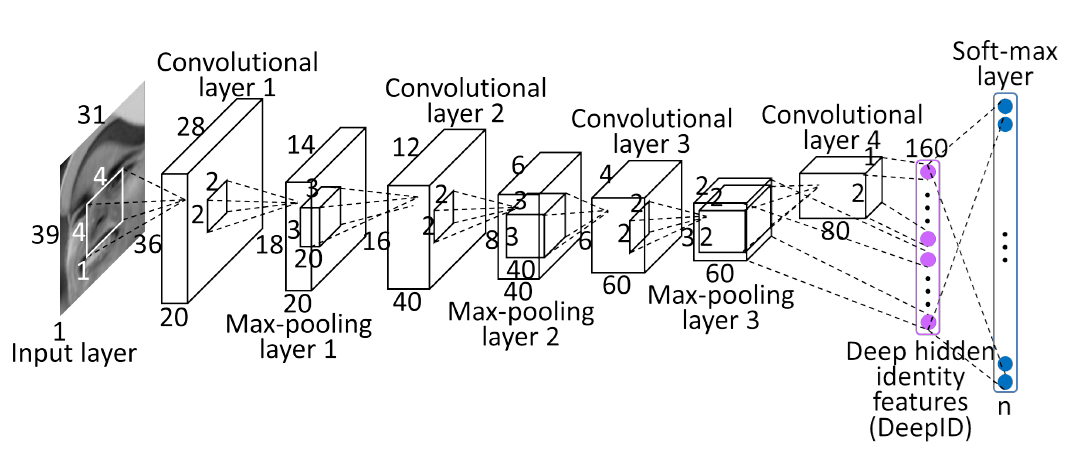
如此，想必大家对DeepID的应用场景已经熟悉了，下面开始讲三代DeepID的进化。

## DeepID

在这里，我假定大家对卷积神经网络已经有了基本的认识，如果没有的话，出门左转看我这篇blog：卷积神经网络http://blog.csdn.net/stdcoutzyx/article/details/41596663。

### DeepID网络结构

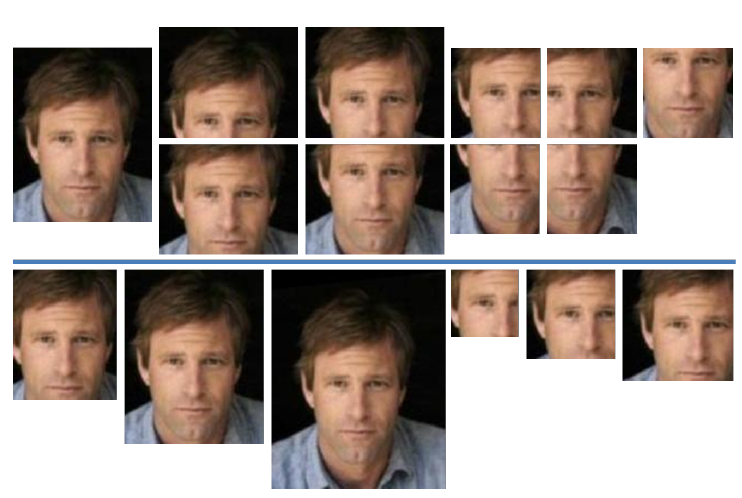
DeepID是第一代，其结构与普通的卷积神经网络差不多。结构图如下：



该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。

### DeepID实验设置

实验中，人脸图片的预处理方式，也就是切分方式的样例如下：



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（也就是上图这样的数据）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。

### 实验结论

* 使用multi-scale patches的convnet比只使用一个只有整张人脸的patch的效果要好。
* DeepID自身的分类错误率在40%到60%之间震荡，虽然较高，但DeepID是用来学特征的，并不需要要关注自身分类错误率。
* 使用DeepID神经网络的最后一层softmax层作为特征表示，效果很差。
* 随着DeepID的训练集人数的增长，DeepID本身的分类正确率和LFW的验证正确率都在增加。

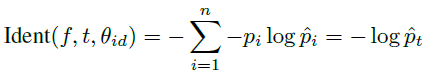
这就是DeepID第一代。

## DeepID2

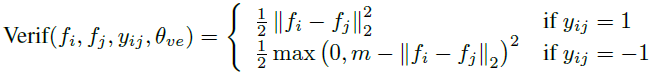
DeepID2相对于DeepID有了较大的提高。其主要原因在于在DeepID的基 础上添加了验证信号。具体来说，原本的卷积神经网络最后一层softmax使用的是Logistic Regression作为最终的目标函数，也就是识别信号；但在DeepID2中，目标函数上添加了验证信号，两个信号使用加权的方式进行了组合。

### **两种信号及训练过程**

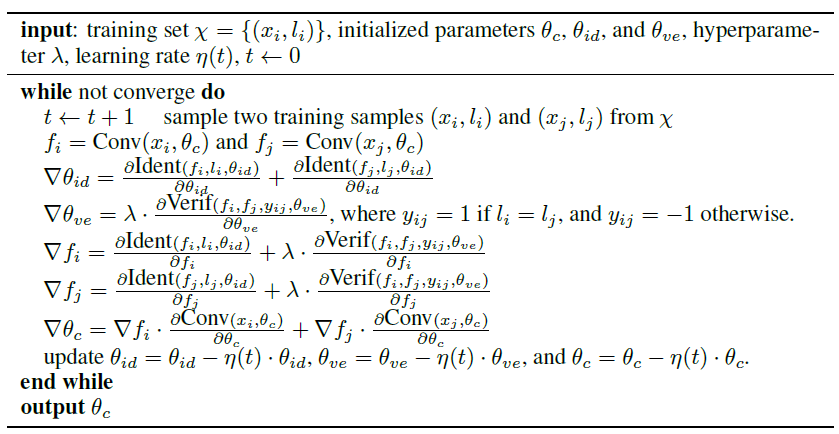
识别信号公式如下：



验证信号公式如下：



由于验证信号的计算需要两个样本，所以整个卷积神经网络的训练过程也就发生了变化，之前是将全部数据切分为小的batch来进行训练。现在则是每次迭代时随机抽取两个样本，然后进行训练。训练过程如下：



在训练过程中，lambda是验证信号的加权参数。M参数时动态调整的，调整策略是使最近的训练样本上的验证错误率最低。

## 3.2 实验设置

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。

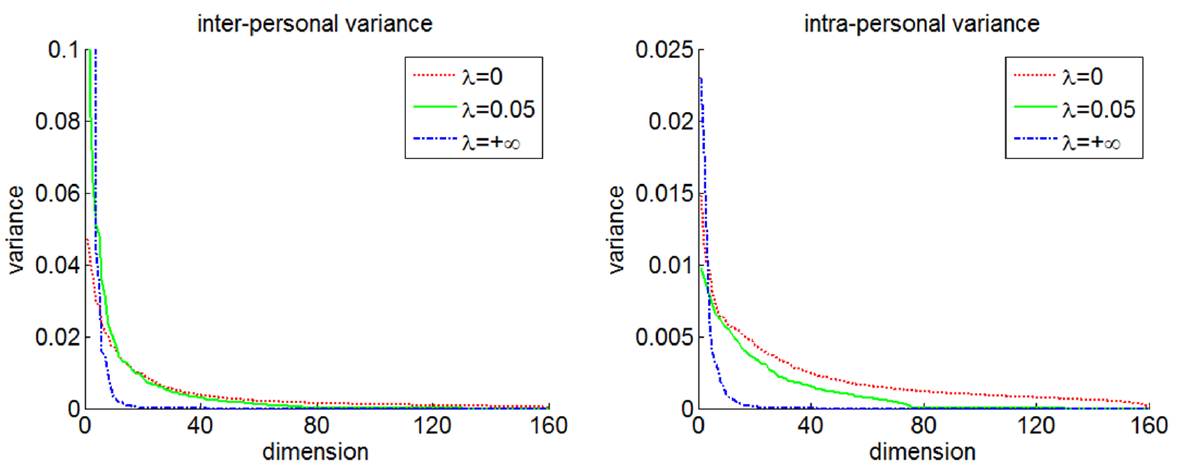
DeepID2使用的外部数据集仍然是CelebFaces+，但先把 CelebFaces+进行了切分，切分成了CelebFaces+A(8192个人)和CelebFaces+B(1985个人)。首先，训练 DeepID2，CelebFaces+A做训练集，此时CelebFaces+B做验证集；其次，CelebFaces+B切分为1485人和500人 两个部分，进行特征选择，选择25个patch。最后在CelebFaces+B整个数据集上训练联合贝叶斯模型，然后在LFW上进行测试。在上一段描述 的基础上，进行了组合模型的加强，即在选取特征时进行了七次。第一次选效果最好的25个patch，第二次从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

其中，选取的25个patch如下：



## 3.3 实验结论

* 对lambda进行调整，也即对识别信号和验证信号进行平衡，发现lambda在0.05的时候最好。使用LDA中计算类间方差和类内方差的方法进行计算。得到的结果如下：



可以发现，在lambda=0.05的时候，类间方差几乎不变，类内方差下降了很多。这样就保证了类间区分性，而减少了类内区分性。如果lambda为无穷大，即只有验证信号时，类间方差和类内方差都变得很小，不利于最后的分类。

* DeepID的训练集人数越多，最后的验证率越高。
* 对不同的验证信号，包括L1，L2，cosin等分别进行了实验，发现L2 Norm最好。

# 4 DeepID2+

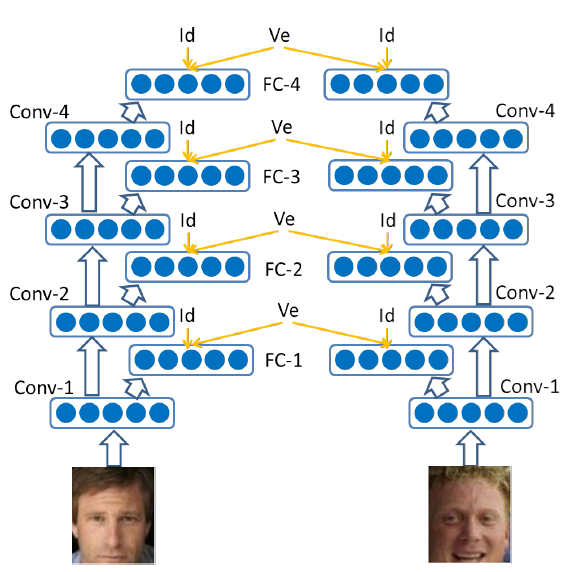
DeepID2+有如下贡献，第一点是继续更改了网络结构；第二点是对卷积神经网络进 行了大量的分析，发现了几大特征，包括：+ 神经单元的适度稀疏性，该性质甚至可以保证即便经过二值化后，仍然可以达到较好的识别效果；+ 高层的神经单元对人比较敏感，即对同一个人的头像来说，总有一些单元处于一直激活或者一直抑制的状态；+ DeepID2+的输出对遮挡非常鲁棒。

## 4.1 网络结构变化

相比于DeepID2，DeepID2+做了如下三点修改：

* DeepID层从160维提高到512维。
* 训练集将CelebFaces+和WDRef数据集进行了融合，共有12000人，290000张图片。
* 将DeepID层不仅和第四层和第三层的max-pooling层连接，还连接了第一层和第二层的max-pooling层。

最后的DeepID2+的网络结构如下：



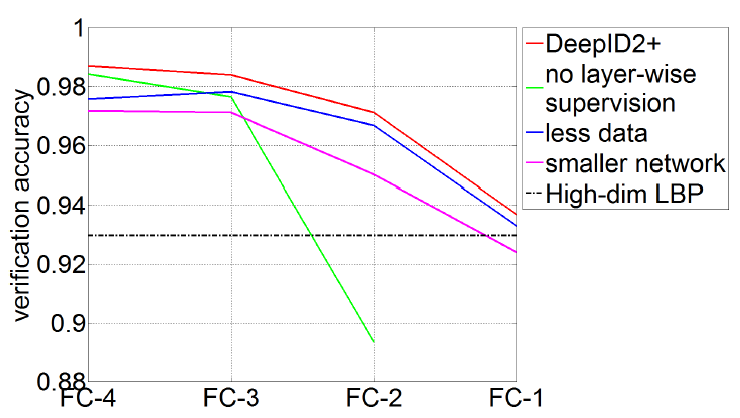
上图中，ve表示监督信号（即验证信号和识别信号的加权和）。FC-n表示第几层的max-pooling。

## 4.2 实验设置

训练数据共有12000人，290000张图像。其中2000人用于在训练DeepID2+时做验证集，以及训练联合贝叶斯模型。

## 4.3 实验结论

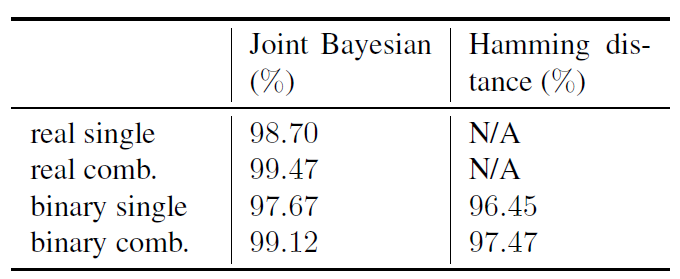
分别使用FC-n进行实验，比较的算法包括DeepID2+、只有从FC-4反向传播下来进行训练的模型、使用少量数据的、使用小的特征向量的模型。结果如下：



DeepID2选取了25个patch，DeepID2+选取了同样的25个patch，然后抽取的特征分别训练联合贝叶斯模型，得到的结果是DeepID2+平均比DeepID2提高2%。

## 4.4 适度稀疏与二值化

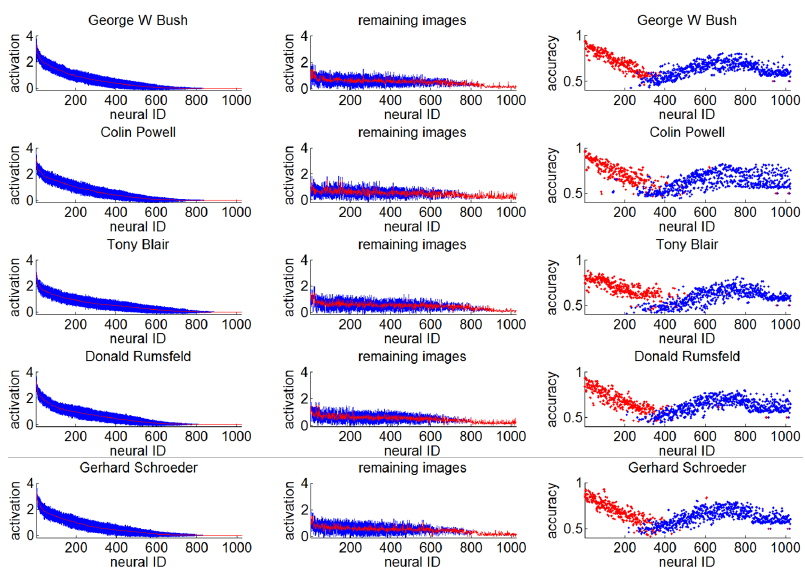
DeepID2+有一个性质，即对每个人，最后的DeepID层都大概有半数的单元是激活的，半数的单元是抑制的。而不同的人，激活或抑制的单元是不同的。基于此性质。使用阈值对最后输出的512维向量进行了二值化处理，发现效果降低有限。



二值化后会有好处，即通过计算汉明距离就可以进行检索了。然后精度保证的情况下，可以使人脸检索变得速度更快，更接近实用场景。

## 4.5 特征区分性

存在某个神经单元，只使用普通的阈值法，就能针对某个人得到97%的正确率。不同的神经单元针对不同的人或不同的种族或不同的年龄都有很强的区分性。在这里，对每个单元的激活程度进行由高到低排序，可以得到下图所示：

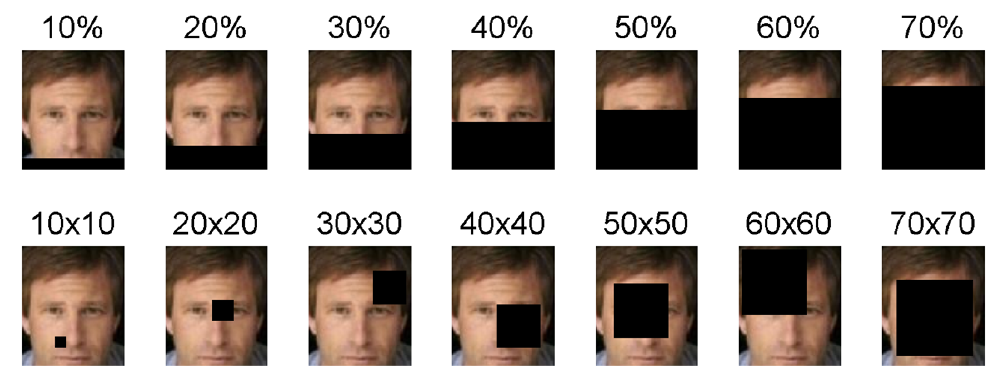


上图只是其中一张图示，还有针对种族、年龄等的激活分析。此处不赘述。

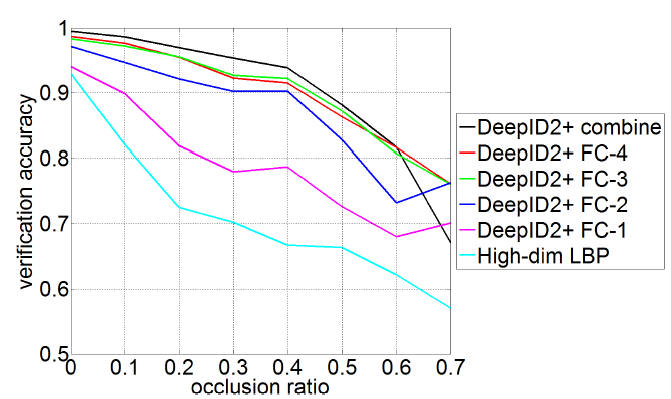
但值得说的是，这种分析方法对我们很有启发。卷积神经网络的输出的含义是什么，很难解释，通过这种方法，或许可以得到一些结论。

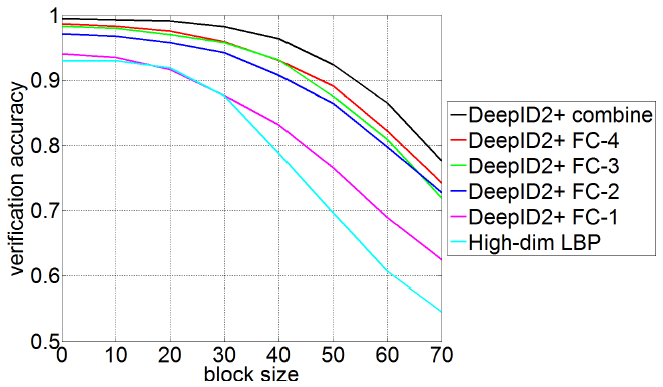
## 4.6 遮挡鲁棒性

在训练数据中没有遮挡数据的情况下，DeepID2+自动就对遮挡有了很好的鲁棒性。有两种方式对人脸进行多种尺度的遮挡，第一种是从下往上进行遮挡，从10%-70%。第二种是不同大小的黑块随机放，黑块的大小从10×10到70×70。



结论是遮挡在20%以内，块大小在30×#30以下，DeepID2+的输出的向量的验证正确率几乎不变。





# **总结**

至此，DeepID的三代进化史就讲完了。简单的说一下我的感受。

首先是卷积神经网络的作用，虽说之前听说过卷积神经网络既可以分类，也可以学习特征，但ImageNet上的卷积神经网络都是分类的，这次终于见到不关注分类错误率而关注特征的卷积神经网络。

其次，卷积神经网络的改进方式，无非如下几种：增大网络深度和宽度，增加数据，将网络隐含层连接到前面几层来，添加其他的信号。

再次，也是最重要的，就是DeepID在发展过程中对输出向量的分析，尤其是DeepID2+，神经网络的各个单元一直是无法解释的，但这次作者不仅试图去发现规律，还基于规律做出了一些改动，比如二值化。

最后，卷积神经网络的鲁棒性真的很厉害。

# **参考文献**

* [1] 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.
* [2] 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.
* [3] Sun Y, Wang X, Tang X. Deeply learned face representations are sparse, selective, and robust[J]. arXiv preprint arXiv:1412.1265, 2014.