

人脸识别算法演化史



SIGAI

微信公众号：SIGAI

268 人赞同了该文章

导言：

本文为人脸识别算法系列专题的综述文章，人脸识别是一个被广泛研究着的热门问题，大量的研究论文层出不穷，文中我们将为大家总结近些年出现的具有代表性的人脸识别算法。请大家关注SIGAI公众号，我们会持续解析当下主流的人脸识别算法以及业内最新的进展。

人脸识别有什么用？

人脸识别的目标是确定一张人脸图像的身份，即这个人是谁，这是机器学习和模式识别中的分类问题。它主要应用在身份识别和身份验证中。其中身份识别包括失踪人口和嫌疑人追踪、智能交互场景中识别用户身份等场景；而身份验证包括身份证等证件查询、出入考勤查验、身份验证解锁、支付等场景，应用场景丰富。就在前不久，北京多家医院借助“黑科技”人脸识别技术阻击“熟脸”的号贩子，降低其挂号率；目前人脸识别还用到了治理闯红灯问题，改善中国式过马路现象。

人脸识别系统的组成

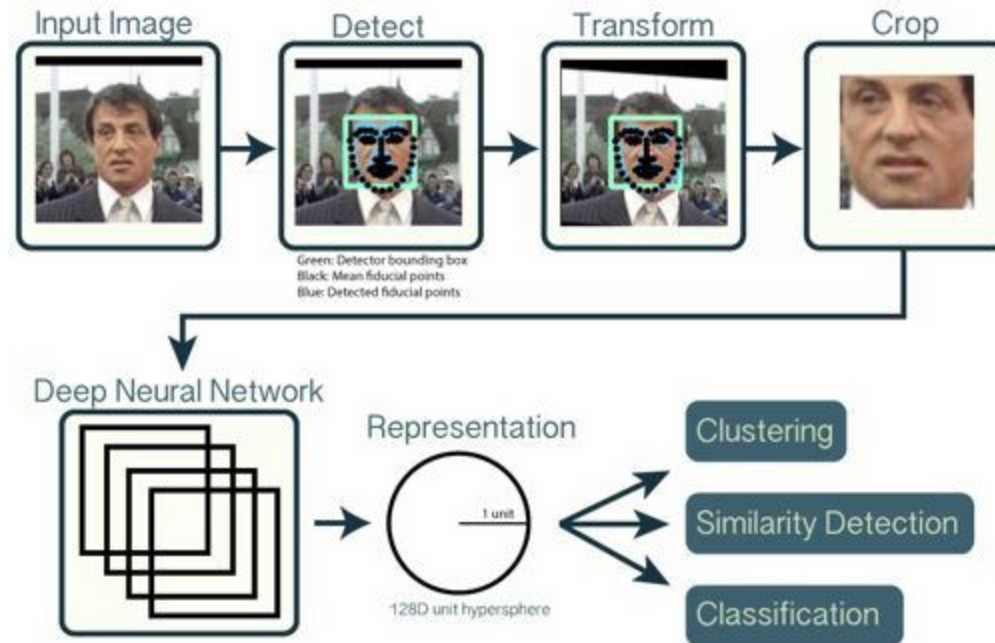
人脸识别算法主要包含三个模块：

人脸检测（Face Detection）

人脸对齐（Face Alignment）

人脸特征表征（Feature Representation）

如下图所示：



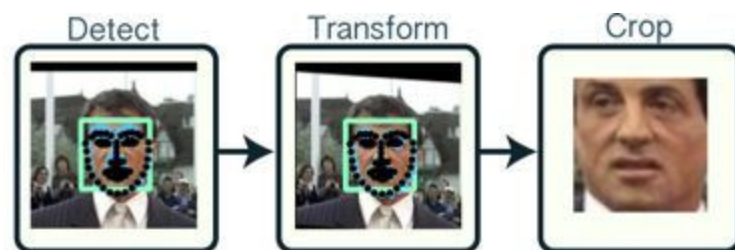
人脸检测

人脸检测用于确定人脸在图像中的大小和位置，即解决“人脸在哪里”的问题，把真正的人脸区域从图像中裁剪出来，便于后续的人脸特征分析和识别。下图是对一张图像的人脸检测结果：



人脸对齐

同一个人在不同的图像序列中可能呈现出不同的姿态和表情，这种情况是不利于人脸识别的。所以有必要将人脸图像都变换到一个统一的角度和姿态，这就是人脸对齐。它的原理是找到人脸的若干个关键点（基准点，如眼角，鼻尖，嘴角等），然后利用这些对应的关键点通过相似变换（Similarity Transform，旋转、缩放和平移）将人脸尽可能变换到标准人脸。下图是一个典型的人脸图像对齐过程：



人脸特征表征

第三个模块是本文重点要讲的人脸识别算法，它接受的输入是标准化的人脸图像，通过特征建模得到向量化的人脸特征，最后通过分类器判别得到识别的结果。这里的关键是怎样得到对不同人脸有区分度的特征，通常我们在识别一个人时会看它的眉形、脸轮廓、鼻子形状、眼睛的类型等，人脸识别算法引擎要通过练习（训练）得到类似这样的有区分度的特征。本系列文章主要围绕人脸识别中的人脸特征表征进行展开，人脸检测和人脸对齐方法会在其它专题系列文章中进行介绍。

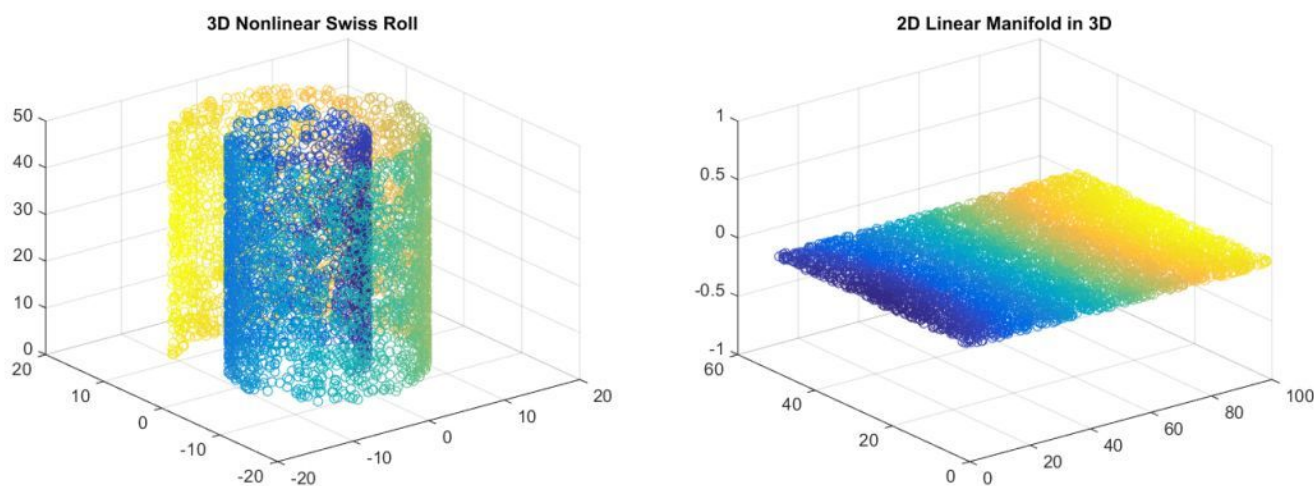
人脸识别算法的三个阶段

人脸识别算法经历了早期算法，人工特征+分类器，深度学习3个阶段。目前深度学习算法是主流，极大的提高了人脸识别的精度，推动这一技术真正走向实用，诞生了如Face++、商汤之类的公司。

早期的算法有基于几何特征的算法，基于模板匹配的算法，子空间算法等多种类型。子空间算法将人脸图像当成一个高维的向量，将向量投影到低维空间中，投影之后得到的低维向量达到对不同的人具有良好的区分度。

子空间算法的典型代表是PCA（主成分分析，也称为特征脸EigenFace）[1]和LDA（线性判别分析，FisherFace）[2]。PCA的核心思想是在进行投影之后尽量多的保留原始数据的主要信息，降低数据的冗余信息，以利于后续的认识。LDA的核心思想是最大化类间差异，最小化类内差异，即保证同一个人的不同人脸图像在投影之后聚集在一起，不同人的脸图像在投影之后被用一个大的间距分开。PCA和LDA最后都归结于求解矩阵的特征值和特征向量，这有成熟的数值算法可以实现。

PCA和LDA都是线性降维技术，但人脸在高维空间中的分布显然是非线性的，因此可以使用非线性降维算法，典型的代表是流形学习[3]和核（kernel）技术。流形学习假设向量点在高维空间中的分布具有某些几何形状，然后在保持这些几何形状约束的前提下将向量投影到低维空间中，这种投影是通过非线性变换完成的。下面是用流形学习进行非线性降维的一个例子：



在这张图中，左边是三维空间中卷曲的一张曲面，不同的颜色代表不同类型的样本，右侧是将这个曲面投影到二维平面后的结果。投影之后各类样本的相对位置关系得到了保留。

核PCA[22]与核LDA[23]作为使用核技术的非线性降维方法，在人脸识别问题上也取得了比线性降维方法更好的结果。

独立成分分析ICA[4]也被用于人脸识别，取得了比PCA更好的效果。**前面介绍的这些算法严重依赖训练集和测试集场景，且对光照、人脸的表情、姿态敏感，泛化能力不足，不具有太多的实用价值。**

隐马尔科夫模型（HMM）也被用于人脸识别问题[5]，和前面这些算法相比，它对光照变化、表情和姿态的变化更鲁棒。另外，卷积神经网络在早期也已经由研究人员被用于人脸识别问题[6]（**We use a database of 400 images of 40 individuals!!! 当时如果研究人员用更多数据的话是不是会大不同呢？**）。

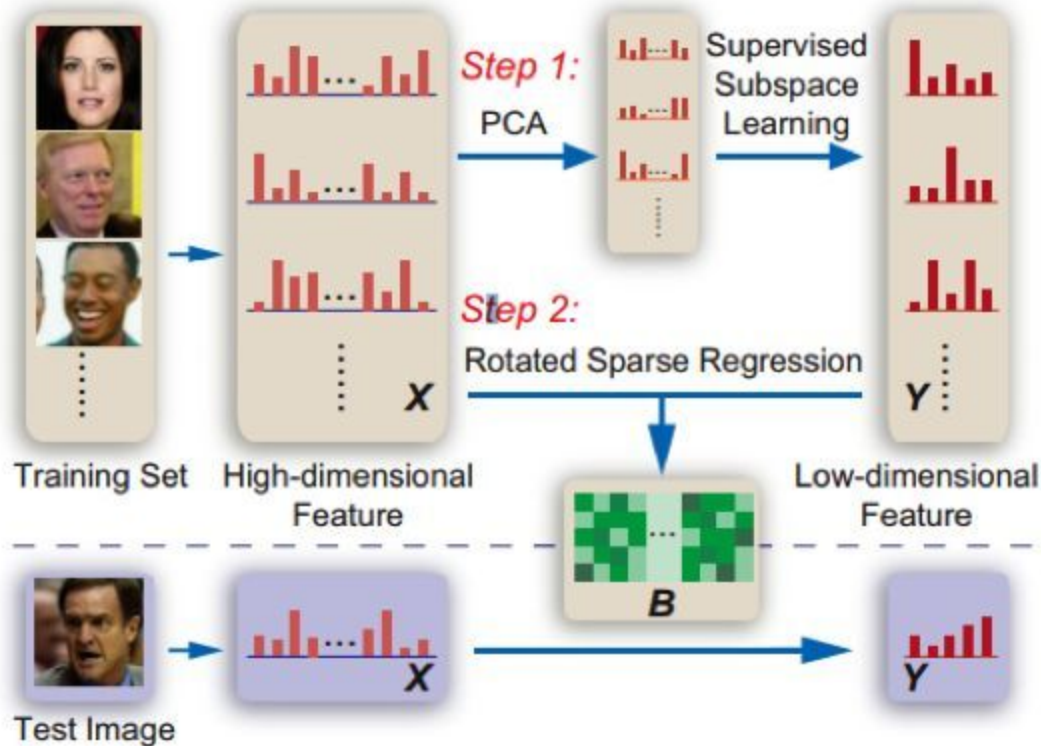
人工特征 + 分类器

第二阶段的人脸识别算法普遍采用了人工特征 + 分类器的思路。分类器有成熟的方案，如神经网络，支持向量机[7]，贝叶斯[8]等。这里的关键是人工特征的设计，它要能有效的区分不同的人。

描述图像的很多特征都先后被用于人脸识别问题，包括HOG、SIFT、Gabor、LBP等。它们中的典型代表是LBP（局部二值模式）特征[9]，这种特征简单却有效。LBP特征计算起来非常简单，部分解决了光照敏感问题，但还是存在姿态和表情的问题。

联合贝叶斯是对贝叶斯人脸的改进方法[8]，选用LBP和LE作为基础特征，将人脸图像的差异表示为相同人因姿态、表情等导致的差异以及不同人间的差异两个因素，用潜在变量组成的协方差，建立两张人脸的关联。文章的创新点在于将两个人脸表示进行联合建模，在人脸联合建模的时候，又使用了人脸的先验知识，将两张人脸的建模问题变为单张人脸图片的统计计算，更好的验证人脸的相关性，该方法在LFW上取得了92.4%的准确率。

人工特征的巅峰之作是出自CVPR 2013年MSRA的"Blessing of Dimisionality: High Dimensional Feature and Its Efficient Compression for Face Verification" [10]，一篇关于如何使用高维度特征在人脸验证中的文章，作者主要以LBP（Local Binary Pattern，局部二值特征）为例子，论述了高维特征和验证性能存在着正相关的关系，即人脸维度越高，验证的准确度就越高。



文中最好的方法在LFW上的精度达到了95.17%，这在当时受到了很大关注，大有成为人脸识别领域新灯塔的迹象，为后续研究者指明未来研究的方向（只可惜好景不长...）。这种融合各种特征形成大杂烩的做法让我们想起了周星驰的电影“国产凌凌漆”里面达文西发明的大杀器-要你命3000：



深度学习

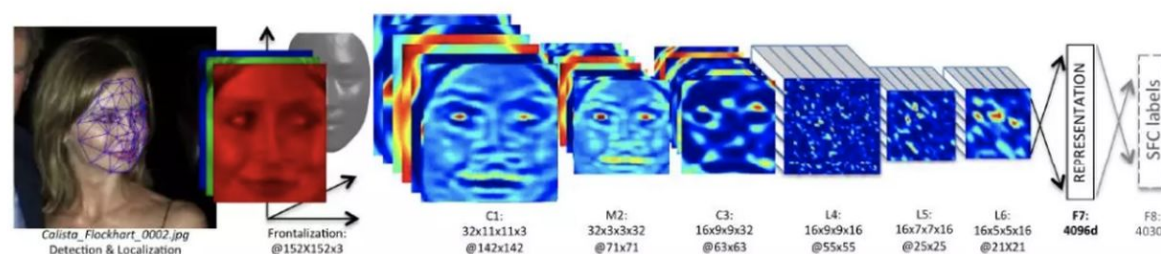
第三个阶段是基于深度学习的方法，自2012年深度学习在ILSVRC-2012大放异彩后，很多研究者都在尝试将其应用在自己的方向，这极大的推动了深度学习的发展。卷积神经网络在图像分类中显示出了巨大的威力，通过学习得到的卷积核明显优于人工设计的特征+分类器的方案。在人脸识别的研究者利用卷积神经网络（CNN）对海量的人脸图片进行学习，然后对输入图像提取出对区分不同人的脸有用的特征向量，替代人工设计的特征。

在前期，研究人员在网络结构、输入数据的设计等方面尝试了各种方案，然后送入卷积神经网络进行经典的目标分类模型训练；在后期，主要的改进集中在损失函数上，即迫使卷积网络学习得到对分辨不同的人更有效的特征，这时候人脸识别领域彻底被深度学习改造了！

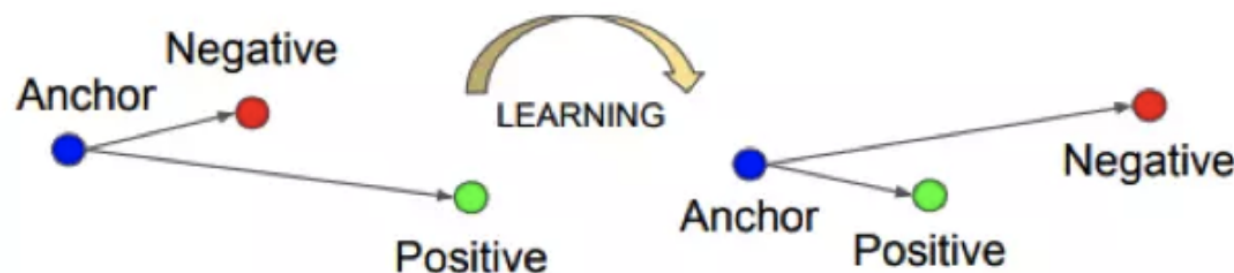
DeepFace[11]是CVPR2014上由Facebook提出的方法，是深度卷积神经网络在人脸识别领域的奠基之作，文中使用了3D模型来做人脸对齐任务，深度卷积神经网络针对对齐后的人脸Patch进行多

类的分类学习，使用的是经典的交叉熵损失函数（Softmax）进行问题优化，最后通过特征嵌入（Feature Embedding）得到固定长度的人脸特征向量。Backbone网络使用了多层局部卷积结构（Local Convolution），原因是希望网络的不同卷积核能学习人脸不同区域的特征，但会导致参数量增大，要求数据量很大，回过头去看该策略并不是十分必要。

DeepFace在LFW上取得了97.35%的准确率，已经接近了人类的水平。相比于1997年那篇基于卷积神经网络的40个人400张图的数据规模，Facebook搜集了4000个人400万张图片进行模型训练，也许我们能得出一个结论：大数据取得了成功！

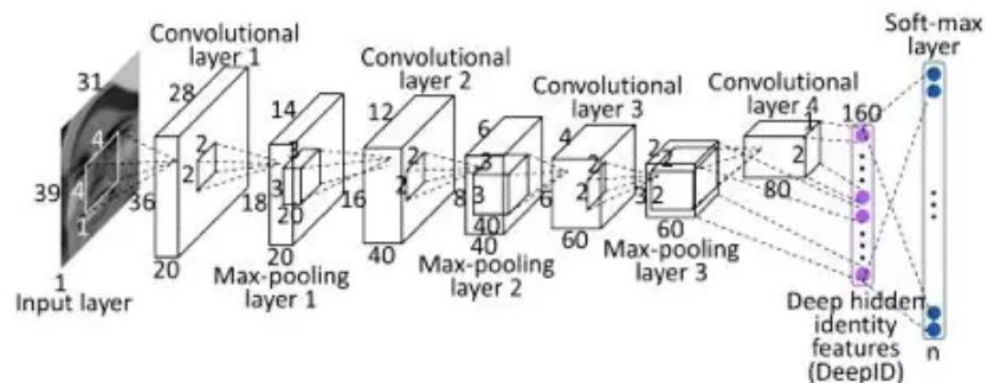


之后Google推出FaceNet[15]，使用三元组损失函数(Triplet Loss)代替常用的Softmax交叉熵损失函数，在一个超球空间上进行优化使类内距离更紧凑，类间距离更远，最后得到了一个紧凑的128维人脸特征，其网络使用GoogLeNet的Inception模型，模型参数量较小，精度更高，在LFW上取得了99.63%的准确率，这种损失函数的思想也可以追溯到早期的LDA算法。



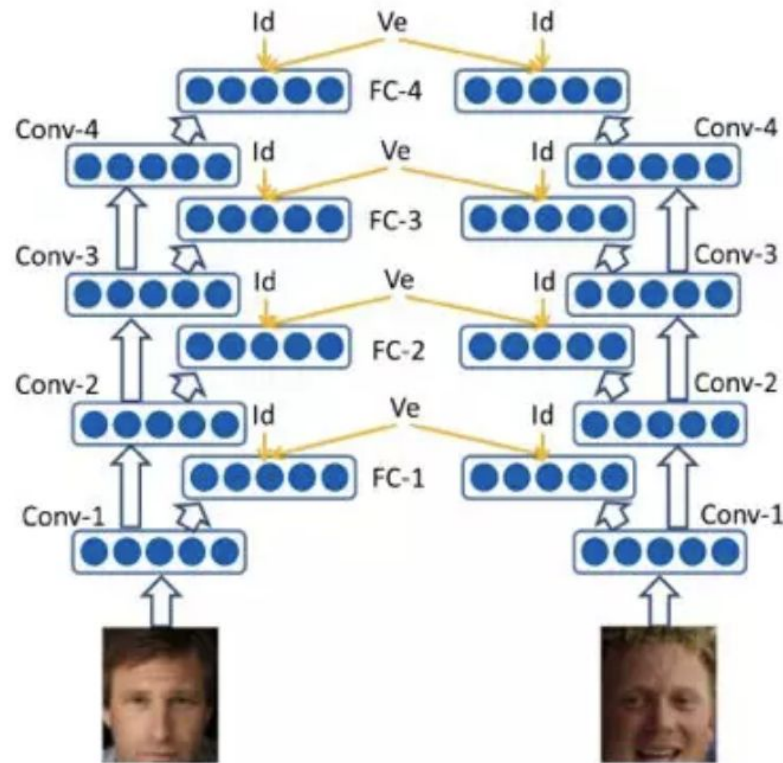
CVPR2014、CVPR2015香港中文大学汤晓鸥团队提出的DeepID系列是一组非常有代表性的工作，其中DeepID1[12]使用四层卷积，最后一层为Softmax，中间为Deep Hidden Identity Features，是学习到的人脸特征表示，并使用Multi-patch分别训练模型最后组合成高维特征，人

脸验证阶段使用联合贝叶斯的方法；通过学习一个多类（10000类，每个类大约有20个实例）人脸识别任务来学习特征，文中指出，随着训练时要预测的人脸类越多，DeepID的泛化能力就越强。



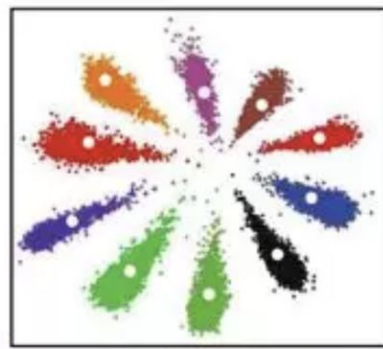
DeepID2[13]在DeepID1的基础上对损失函数部分做了改进，在原有Identification Loss基础上，增加了Verification Loss，其中Verification主要是增加类内的紧致度，而Identification是反应类间的变化。通过提升类间差距，降低类内差距，是训练出来的特征更加适合类似人脸识别的任务。这一思想同样源于早期的LDA算法。

DeepID3[14] 提出了用于人脸识别的两个非常深的神经网络结构（基于VGG和GoogleNet），但识别结果与DeepID2一样，或许当有更多的训练数据时，能够提高性能，需要进一步研究。

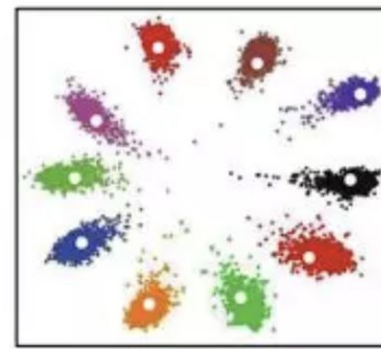


当前人脸识别任务主要是应用在开集识别的情况下，这就需要学习出来的人脸特征有好的泛化能力，而Softmax损失函数本身是用于解决多分类的问题，并没有针对隐含特征层去优化，往往直接训练出来的特征并不具有好的泛化能力。Contrastive Loss和Triplet Loss虽然优化目标很明确很合理，但是需要研发人员具有丰富的数据工程经验（比如OHEM-困难样本挖掘）。能不能有一种端到端的解决方案呢？ECCV2016一篇文章提出了权衡的解决方案。通过添加Center-Loss[16]对特征层进行优化并结合Softmax就能够训练出拥有内聚性良好的特征。该特点在人脸识别上尤为重要，从而使得在很少的数据情况下训练出来的模型也能有不俗的性能。

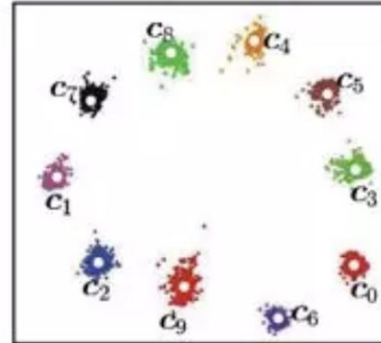
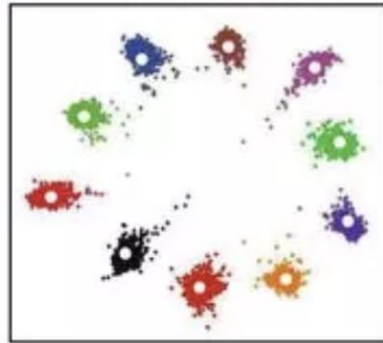
Center-Loss在Softmax的基础上加入了一个维持类别中心的损失函数，并能使特征向所属类别中心聚拢，从而使达到了和Triple Loss类似的效果。



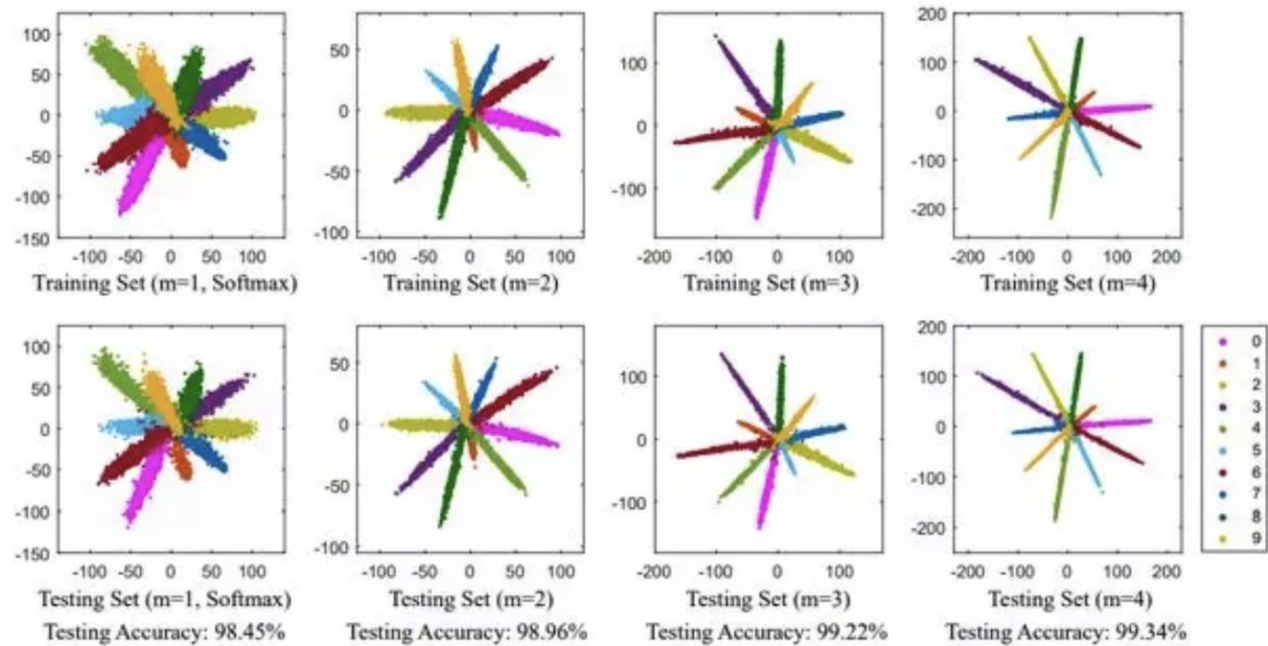
(a) $\lambda = 0.001$



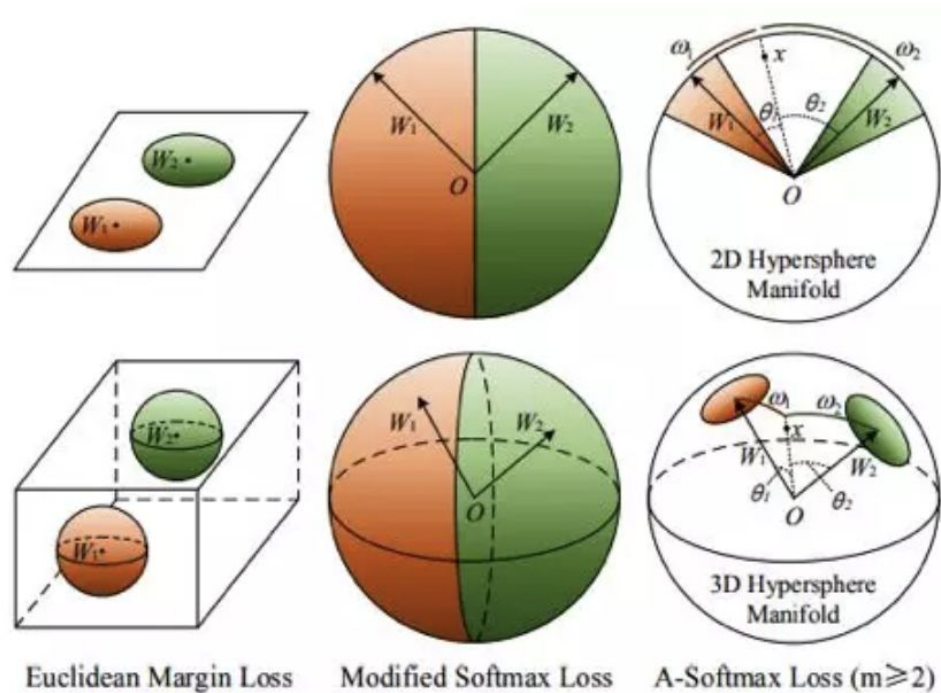
(b) $\lambda = 0.01$



L-Softmax[17]是ICML2016的一篇文章提出的，对Softmax做了改进，在网络设计上将最后一层分类层的偏置项去掉（这一做法在后续的人脸识别损失函数改进中都得到了使用），直接优化特征和分类器的余弦角度，通过人为设定增加了一个角度（Margin）增加了模型的学习难度，借用SVM的思想来理解的话，如果原来的Softmax loss是只要支持向量和分类面的距离大于 h 就算分类效果比较好了，那么L-Softmax就是需要距离达到 mh （ m 是正整数）才算分类效果达到预期。通过这种方式最终使类间距离增大，类内样本更紧凑。



SphereFace[18]提出了A-Softmax，针对L-Softmax做了微小的改进，归一化了权重，可以看成在一个超球面的流形上对样本进行分类判别。

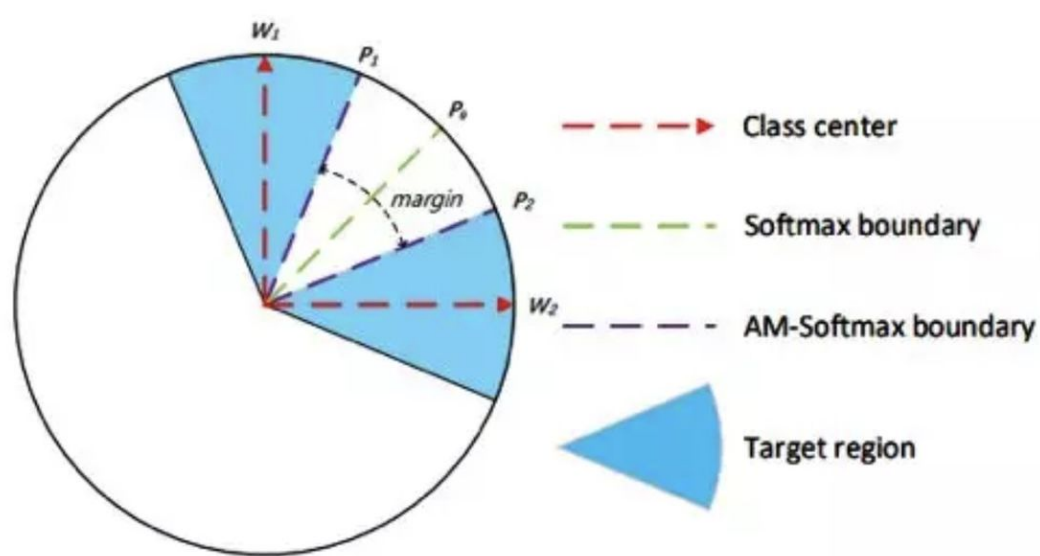


以上的几种方法都没有对嵌入的特征层进行归一化，所以不能看作严格意义上的针对角度的优化。L2-Softmax[19]，将特征向量做了L2归一化，这样做的好处是范数小的特征算出来的梯度会更大；而范数小的特征一般对应的是质量较差的图片。因此，某种程度上特征归一化起到类似于难例挖掘的作用。然而对特征层强行进行L2的约束会导致分类空间太小，导致模型训练困难，Loss值难以下降分类效果不佳。在实际的模型训练中为了便于模型优化作者加入尺度缩放因子将分类的超球空间放大。如下图所示，上边一行的Feature Norm很大，下边一行Feature Norm很小，L2-Softmax能更好的处理下边一行的人脸图片。



后续的AM-Softmax[20]和ArcFace[21]都是针对SphereFace做了改进。

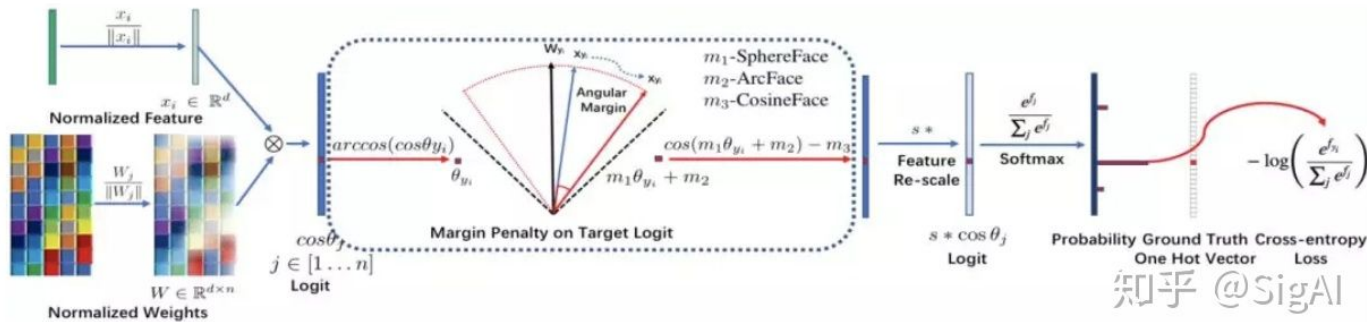
AM-Softmax是把角度裕量从 $\cos(m\theta)$ 改进为 $\cos(\theta) + m\cos(m\theta)$ 改进为 $\cos(\theta) + m$ ，主要的好处是这样改进之后容易收敛。



$$L = -\frac{1}{m} \sum_{i=1}^m \log \frac{e^{s(\cos(\theta_{y_i}) - m)}}{e^{s(\cos(\theta_{y_i}) - m)} + \sum_{j=1, j \neq y_i}^n e^{s \cos \theta_j}}$$

ArcFace可以看做是针对AM-Softmax的改进版本，直接针对角度去加Margin（红线标记的部分），这样做的好处是角度距离比余弦距离在对角度的影响更加直接。ArcFace同时对特征和权重归一化，参考L2-Softmax加入了固定的特征尺度因子S。通过针对性的对网络结构做了改进以及对MegaFace数据集的清洗，作者跑出了令人兴奋的指标。

$$L = -\frac{1}{m} \sum_{i=1}^m \log \frac{e^{\underline{s(\cos(\theta_{y_i} + m))}}}{e^{s(\cos(\theta_{y_i} + m))} + \sum_{j=1, j \neq y_i}^n e^{s \cos \theta_j}}$$



参考文献

- [1] Matthew Turk, Alex Pentland. Eigenfaces for recognition. 1991, Journal of Cognitive Neuroscience.
- [2] Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. Peter N Belhumeur J P Hespanha David Kriegman. 1997 IEEE Transactions on Pattern Analysis and Machine Intelligence.
- [3] He, Xiaofei, et al. Face recognition using Laplacianfaces. Pattern Analysis and Machine Intelligence, IEEE Transactions on 27.3 (2005): 328-340.
- [4] Bartlett M S, Movellan J R, Sejnowski T J. Face Recognition by Independent Component Analysis [J]. IEEE Trans. on Neural Network, 2002, 13(6): 1450-1464
- [5] Ara V Nefian, Monson H Hayes. Hidden Markov models for face recognition. international conference on acoustics speech and signal processing, 1998.
- [6] S Lawrence, C L Giles, Ah Chung Tsoi, Andrew D Back. Face recognition: a convolutional neural-network approach. 1997, IEEE Transactions on Neural Networks.

[7] Guodong Guo, Stan Z Li, Kap Luk Chan. Face recognition by support vector machines. ieee international conference on automatic face and gesture recognition, 2000.

[8] Dong Chen, Xudong Cao, Liwei Wang, Fang Wen, Jian Sun. Bayesian face revisited: a joint formulation. 2012, european conference on computer vision.

[9] Timo Ahonen, Abdenour Hadid, Matti Pietikainen. Face Description with Local Binary Patterns: Application to Face Recognition. 2006, IEEE Transactions on Pattern Analysis and Machine Intelligence.

[10] Dong Chen, Xudong Cao, Fang Wen, Jian Sun. Blessing of Dimensionality: High-Dimensional Feature and Its Efficient Compression for Face Verification. 2013, computer vision and pattern recognition.

[11] Yaniv Taigman, Ming Yang, Marcaurelio Ranzato, Lior Wolf. DeepFace: Closing the Gap to Human-Level Performance in Face Verification. 2014, computer vision and pattern recognition.

[12] Yi Sun, Xiaogang Wang, Xiaoou Tang. DeepID: Deep Learning for Face Recognition. 2014, computer vision and pattern recognition.

[13] Yi Sun, Yuheng Chen, Xiaogang Wang, Xiaoou Tang. Deep Learning Face Representation by Joint Identification-Verification. 2014, neural information processing systems.

[14] Yi Sun, Ding Liang, Xiaogang Wang, Xiaoou Tang. DeepID3: Face Recognition with Very Deep Neural Networks. 2015, Computer Vision and Pattern Recognition.

[15] Florian Schroff, Dmitry Kalenichenko, James Philbin. FaceNet: A unified embedding for face recognition and clustering. 2015, computer vision and pattern recognition.

[16] Yandong Wen, Kaipeng Zhang, Zhifeng Li, Yu Qiao. A Discriminative Feature Learning Approach for Deep Face Recognition. 2016, european conference on computer vision.

- [17] W. Liu, Y. Wen, Z. Yu, and M. Yang. Large-margin softmax loss for convolutional neural networks. In ICML, 2016. 2, 3, 7, 8
- [18] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song. Sphereface: Deep hypersphere embedding for face recognition. arXiv preprint arXiv:1704.08063, 2017. 2, 8
- [19] Ranjan, Rajeev, Carlos D. Castillo, and Rama Chellappa. "L2-constrained Softmax Loss for Discriminative Face Verification." arXiv preprint arXiv:1703.09507 (2017).
- [20] F. Wang, W. Liu, H. Liu, and J. Cheng. Additive margin softmax for face verification. In arXiv:1801.05599, 2018. 1, 2, 3, 4, 9
- [21] Deng J, Guo J, Zafeiriou S. ArcFace: Additive Angular Margin Loss for Deep Face Recognition[J]. arXiv preprint arXiv:1801.07698, 2018.
- [22] Minghsuan Yang, Narendra Ahuja, David J Kriegman. Face recognition using kernel eigenfaces. 2000, international conference on image processing.
- [23] Juwei Lu, Konstantinos N Plataniotis, Anastasios N Venetsanopoulos. Face recognition using kernel direct discriminant analysis algorithms. 2003, IEEE Transactions on Neural Networks.

推荐文章

- [1] [机器学习-波澜壮阔40年 SIGAI 2018.4.13.](#)
- [2] [学好机器学习需要哪些数学知识? SIGAI.4.17.](#)

授权声明:

本图文海报所采用图片经THE YIAN STUDIO授权

SIGAI
2018.04.20

编辑于 2019-10-21

- 计算机视觉
- 机器学习
- 深度学习 (Deep Learning)