

从统一子空间分析到联合深度学习：人脸识别的十年历程

王晓刚 孙 祎 汤晓鸥
香港中文大学

关键词：深度学习 人脸识别 人脸特征

引言

人脸识别是计算机视觉领域最重要的研究方向之一，有巨大的市场和良好的应用前景。十多年前，人脸识别采用的主流方法是子空间分析。2004年，我们提出了统一子空间分析^[1]方法，利用同一个理论框架对当时最为流行的三种子空间方法——主分量分析子空间 (Principal Component Analysis, PCA)^[2]、贝叶斯子空间 (Bayesian face recognition)^[3] 和线性判别分析子空间 (Linear Discriminant Analysis, LDA)^[4] 进行比较，分析它们的关联与不同，并将三者有机结合达到最佳识别性能。近两年，深度学习的出现使人脸识别取得了突破性进展。本文将对这两类看似完全不同的方法进行对比分析，探讨它们的内在联系，以更好地理解深度学习能成功应用在人脸识别中的原因。

深度学习的初衷是模拟人脑的认知机理。在某些重要的人脸

识别测试集上，深度学习模型已经达到甚至超出了人类的识别水平。这启发我们未来可以在更广的范围内更深入地研究这些模型中神经元响应的性质，并与认知科学的研究成果相结合，相互印证和启发。一些学者将深度学习的成功归因于用一个具有大量参数的复杂模型拟合数据集。而人脸识别的最新研究成果^[7]表明，深度学习得到的人脸特征表达具有手工特征表达所不具备的重要特性，例如它是中度稀疏的、对人脸身份和人脸属性有很强的选择性、对局部遮挡具有良好的鲁棒性。这些特性是通过大数据训练自然得到的，并未对模型加入显式约束或后期处理。这说明深度学习并非单纯拟合数据集，其计算单元有清晰的语义含义，其高识别率可以通过这些特性来解释。

深度学习在人脸识别中的进展

LFW^[5] 是人脸识别领域最受关注的测试集，样本来自互联网的名人人脸照片，用于评估在非可控条件下人脸识别的性能。LFW 的测试任务是判断一对人脸照片是来自同一个人还是不同的人。测试集包含 6000 对人脸图像。经典的人脸识别算法 Eigenface^[2] 在 LFW 中的识别率只有 60%，而最新的深度学习算法的识别率是 99.47%，甚至超过了人眼在此测试集的识别率 (99.25%)^[8]。

表 1 列出了各种深度学习方法在 LFW 人脸确认任务上的识别率。2012 年，勒尼德·米勒 (Learned-Miller) 的研究小组^[9] 率先将深度学习用于 LFW 的人脸识别。他们采用的是无监督的特征学习方法，取得了 87% 的识别率，与当时最好的人脸识别算法相比还有一定差距。2013 年，我们的研究小组采用人脸确认任务作为监督信号，在 LFW 上取得了 92.52% 的识别率^[10]。

CVPR¹ 2014、脸谱 (DeepFace^[11]) 和我们的研究小组 (DeepID^[13]) 都采用人脸辨识作为监督信号, 利用卷积网络^[15] 学习人脸特征, 然后将学到的特征用于 LFW 的人脸确认任务。三者都取得了高于 97% 的识别率, 明显超过了非深度学习的人脸识别算法。但是 DeepFace 和 DeepID 在网络结构等细节上存在差异。DeepFace 采用了 700 万个训练样本, 依靠 73 点的三维人脸图像配准作为预处理; 而 DeepID 只用了 20 万个训练样本和基于 5 点的二维人脸图像配准。它们的训练集和 LFW 在人脸的身份上是不重合的。这意味着 LFW 中的名人对于这些训练得到的深度模型是陌生的。在后续工作中, DeepFace 将训练集扩大到 5 亿, 在 LFW 上的识别率提高到了 98.4%^[12]。我们的研究小组同时将人脸确认和人脸辨识两个任务作为监督信号, 进行联合深度学习, 推出 DeepID2^[14], 仍然利用 20 万个训练样本, 在 LFW 上取得了 99.15% 的准确率。在最新的

研究进展中 (DeepID2+)^[7], 我们将监督信号加到每一层, 采用更大的网络结构, 将训练数据扩大到 45 万, 识别率达到了 99.47%。

深度学习在 LFW 人脸确认任务中的识别率趋于百分之百。2014 年, 阿尼尔·简 (Anil Jain) 的研究小组提出了在 LFW 上人脸辨识任务的测试协议^[16]。这是一对多的匹配, 比一对一的人脸确认难度高很多。该测试协议包含基于封闭集和开放集两种不同的人脸辨识任务。在封闭集上报告的是 Rank-1 识别率, 开放集上报告的是在 1% 的虚警率下 Rank-1 的检测和识别率 (DIR@1% FAR)。表 2 显示了在这个更具挑战性的测试协议下, 不同的深度学习方法之间的差异, 而深度学习和浅层学习之间的差异更加明显。DeepID2+ 比 DeepFace 的 Rank-1 识别率高出了 12.5 个百分点。

统一子空间分析

人脸识别的核心挑战在于如何区分由于姿态、光线、表情、年龄和遮挡等因素引起的类内变化, 和由于身份不同引起的类间变化。这两种变化的分布是高度复杂且非线性的。

用于人脸识别的最流行的子空间分析方法是线性判别子空间。它采用类间散度矩阵描述类间变化, 类内散度矩阵描述类内变化。目的是通过训练得到特征的线性变换, 使类间散度矩阵和类内散度矩阵的比值达到最大。从优化的角度看, 在类内变化保持恒定的前提下, 线性判别子空间找到的线性变换使类中心间的距离达到最大。

统一子空间分析^[1]表明, 贝叶斯人脸识别可以作为线性判别子空间的一个中间步骤, 目的是通过学习类内变化子空间来压缩类内变化, 根据贝叶斯决策理论, 计算两幅人脸图像的差异属于类内变化的后验概率, 作为两幅图像的相似度量。文献 [1] 研究表明, 在通过贝叶斯人脸识别学

表1 深度学习在LFW人脸确认任务中的识别率

方法	用于配准的关键点数	训练样本数	准确率(%)
文献[9]	3	非监督	87
文献[10]	5	8.7628万	92.52
DeepFace ^[11]	73	700万	97.35
DeepFace ^[12]	73	5亿	98.40
DeepID ^[13]	5	20.2599万	97.45
DeepID2 ^[14]	18	20.2599万	99.15
DeepID2+ ^[7]	18	45万	99.47

表2 各方法在LFW人脸辨识任务中的识别率

方法	Rank-1 (%)	DIR@1% FAR(%)
COST-S1 ^[16]	56.7	25.0
COST-S1+S2 ^[16]	66.5	35.0
DeepFace ^[11]	64.9	44.5
DeepFace ^[12]	82.5	61.9
DeepID2 ^[14]	91.1	61.6
DeepID2+ ^[7]	95.0	80.7

¹ IEEE Conference on Computer Vision and Pattern Recognition, IEEE计算机视觉与模式识别会议。

习线性特征变换的过程中,起到关键作用的是反映类内变化的人脸图像对,它们用于计算人脸类内变化子空间。将两幅人脸图像的差异投影到人脸类内变化子空间后的白化过程,可以有效压缩类内变化,放大类间变化,但也会使噪声放大。

统一子空间分析的要点是逐步抑制噪声和类内变化的影响,同时增强类间变化。第一步,以非监督的方式对所有人脸图像进行主分量分析,目的是去除噪声,避免其在后续的步骤中被放大;第二步,基于主分量分析子空间,借鉴贝叶斯人脸识别的方法计算类内变化子空间,通过白化压缩类内变化,放大类间变化;第三步,基于白化后的类内变化子空间,对不同人脸身份类别的中心再次进行主分量分析,进一步扩大类间变化。最终得到的子空间类似于主分量分析子空间,但是它允许自由调整中间步骤中得到的三个不同的子空间的维度。这对于平衡类间变化、类内变化和噪声,优化子空间的人脸识别性能至关重要,因为这些维度决定着统一子空间分析取得了多少对类内变化和噪声的不变性,失去了多少代表类间变化的有用信息。

联合深度学习

统一子空间分析表明了传

统的人脸识别方法是如何区分类内和类间变化的。但是类内和类间这两种变化的分布是复杂且非线性的,无法通过线性子空间的方法将其分离。而深度学习通过多层非线性的特征变换为解决这一难题提供了强有力的工具。DeepID2^[16]通过学习非线性特征变换使类内变化达到最小,而同时使不同身份的人脸图像间的距离保持恒定。

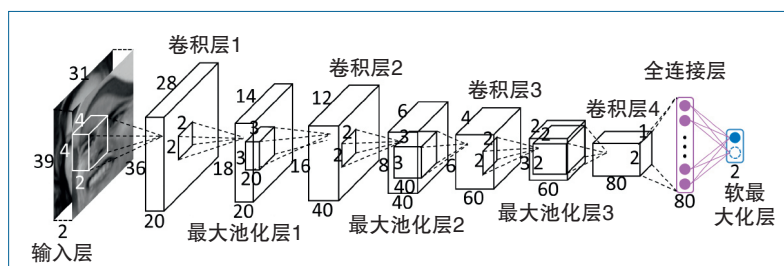


图1 文献[10]提出的通过人脸确认任务学习特征的卷积网络模型

人脸识别分为人脸确认和人脸辨识两种任务。人脸确认是二分类问题,判断两张人脸图像是否属于同一个人,相对简单;人脸辨识是多分类问题,将一张人脸图像分成 N 个身份类别之一,难度随着 N 变大而增加。这两个任务都可以作为监督信号通过深度模型学习人脸特征,也可以联合使用。

通过人脸确认任务学习特征

我们在ICCV² 2013的工作中最早采用人脸确认任务作为监督信号学习人脸特征^[10]。图1显

示了其网络结构。在训练阶段,输入是一对面脸图像,输出是一个二维向量,预测这对面脸图像是否属于同一个人。与普通的卷积网络不同,图1显示的第一个卷积层,输入的两幅图像和若干个“滤波器对”进行卷积,得到反映两幅图相似度的相关特征图。之后相关特征图被当作更高层次的输入,经过若干次正常的卷积和池化(如同普通卷积网络

一样)输出预测。最后一个隐含层包含了80个神经元,是学习得到的人脸特征。文献[10]将人脸图像划分成若干个重叠的区域,针对每个人脸区域训练一个卷积网络,并将多个网络进行融合以提高鲁棒性。实验结果表明,融合最后一个隐含层的特征表示更加有效,因为它保留了更多的关于人脸身份的信息。这也表明训练深度模型的关键在于学习特征表示。去掉卷积网络的输出层后,多个网络的特征表示作为受限玻尔兹曼机的输入来预测最终结果;通过反向传播,可将受限玻尔兹曼机和多个卷积网络视为

² IEEE International Conference on Computer Vision, 国际计算机视觉大会。

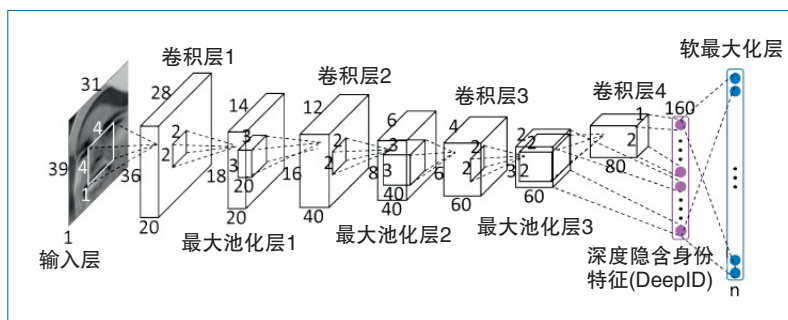


图2 DeepID^[13]中提出的通过人脸辨识任务学习特征的卷积网络模型一个整体联合优化。

DeepID: 通过人脸辨识任务学习特征

DeepID^[13]通过人脸辨识任务学习人脸的特征表示。图2显示了DeepID使用的网络模型。训练集中的每个图像通过卷积网络被分成 N 个身份类别之一（在DeepID中 $N=10000$ ）。卷积网络的输出是一个 N 维向量，该向量只有一个元素是1，其他元素都是0。这意味着通过多层非线性特征变换，人脸被映射到高维单形，不同身份的人脸是等距的

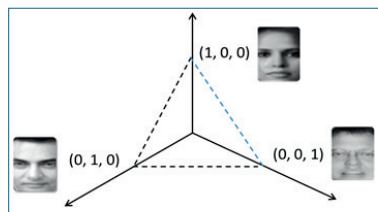


图3 DeepID通过非线性特征变换将不同身份的人脸图像映射到高维单形，使得不同身份的人脸类别间距保持恒定

（如图3所示）。最后一个隐含层包含了160个神经元，是学习得到的人脸特征，它包含了丰富的类间变化信息，而且具有极强的泛化能力。它可以用于识

别训练集中没有出现过的新人，也可以用于其他与人脸识别相关的任务。例如，LFW测试集和DeepID训练集在人脸身份上是没有交集的，虽然DeepID是通过人脸辨识任务进行训练的，但它得到的人脸特征可以直接用于LFW上的人脸确认任务。 N 越大，训练集上的分类任务越难，人脸特征的泛化能力和区分能力反而越强。

与作为二分类问题的人脸确认任务相比，人脸识别任务难度更大，但通过深度模型学习人脸特征的方法对其更加有效。深度学习的难点在于避免训练集上的过拟合问题，更难的训练任务可以避免过拟合。深度学习的重点在于学习特征表示，训练和测试阶段的任务可以不同。因此可以“首先通过一

个更难的任务学习特征，然后利用该特征完成相对简单的识别任务”。人脸确认是一个二分类问题，人工标注的难度较低，但这也意味着每个训练样本所含的信息量较少；人脸辨识是一个多分类问题，人工标注的难度要高很多，每个训练样本的信息量也大很多。从反向传播的角度看，人脸确认任务的输出层只有两个元素传播误差作为反馈，而人脸辨识任务的输出层有 N 个元素传播误差，因此训练更有效。

DeepID2:联合深度学习

DeepID的人脸特征包含丰富的类间变化，也包含影响识别的类内变化。DeepID2^[14]同时利用人脸辨识和人脸确认任务作为监督信号进行联合深度学习。它的目标函数中，除了人脸辨识的

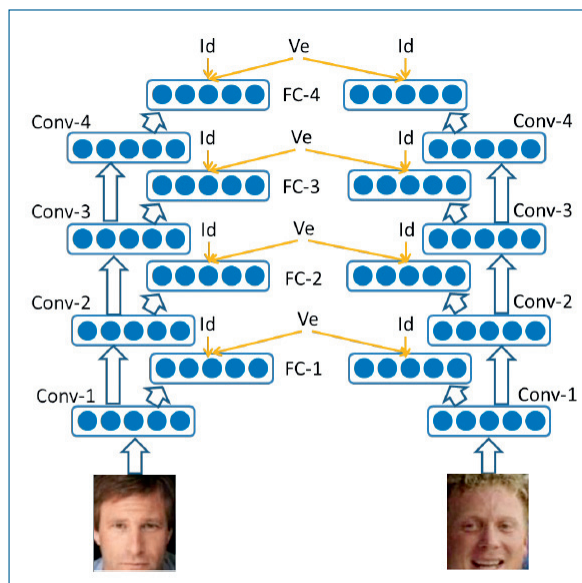


图4 DeepID2+^[7]提出的网络深度模型（它同时采用人脸确认(Ve)和人脸辨识(Id)作为监督信号，并且将监督信号加入到网络的每一层）

预测误差,还包含人脸确认的相关项。如果两幅照片属于不同的身份类别,那么它们的160维人脸特征向量所代表的类内变化就要尽可能地小;否则所代表的类间变化就要尽可能地大。文献[14]中的实验表明,人脸确认监督信号中最有效的是反映类内变化的人脸对,如果将代表类间变化的人脸对去除,识别率几乎不变,因为类间变化已经反映在了人脸辨识的监督信号中。我们在统一子空间分析中也观察到过类似的现象。

DeepID2+通过采用更大的网络结构、更大的训练集(特别是更多的身份类别),以及将监督信号加到每个卷积-池化层中,对DeepID2作了改进,进一步提升了人脸识别的准确性^[7]。图4显示了DeepID2+的网络结构。

统一子空间和联合深度学习的对比

统一子空间分析和联合深度学习都同时利用了人脸辨识和人脸确认两种监督信号。在子空间分析中,人脸辨识信号体现在类间散度矩阵,人脸确认信号体现在类内散度矩阵。统一子空间分析显示,在贝叶斯空间分析中,体现类内变化的人脸“图像对”对于线性特征的学习最有效,而来自不同身份类别的“图像对”体现的作用很小。此现象在DeepID2中也存在。统一子空间的优化目标是在类内变化保持恒

定的条件下,使类间变化最大化;而联合深度学习的优化目标是在类间距离保持恒定的条件下,使类内变化最小化。统一子空间分析通过主分量分析以非监督的学习方式去除噪音,压缩高维图像空间。虽然目前联合深度学习中没有对应的步骤,但是很多研究工作表明,非监督的深度学习也是学习特征的一种重要方式,可以对深度模型进行预训练。在未来工作中,它有可能和联合深度学习进一步结合。从以上可以看出,统一子空间分析和联合深度

学习是紧密相关的。它们的关键不同在于,统一子空间分析学习线性特征变换,识别能力有限;而联合深度学习通过多层非线性变换达到优化目的。

研究DeepID2+人脸特征的性质

为了证实深度学习并非只是用一个复杂的模型去拟合数据,文献[7]进一步研究了DeepID2+的神经元响应具有的特殊性质。研究集中在人脸识别特征的最高

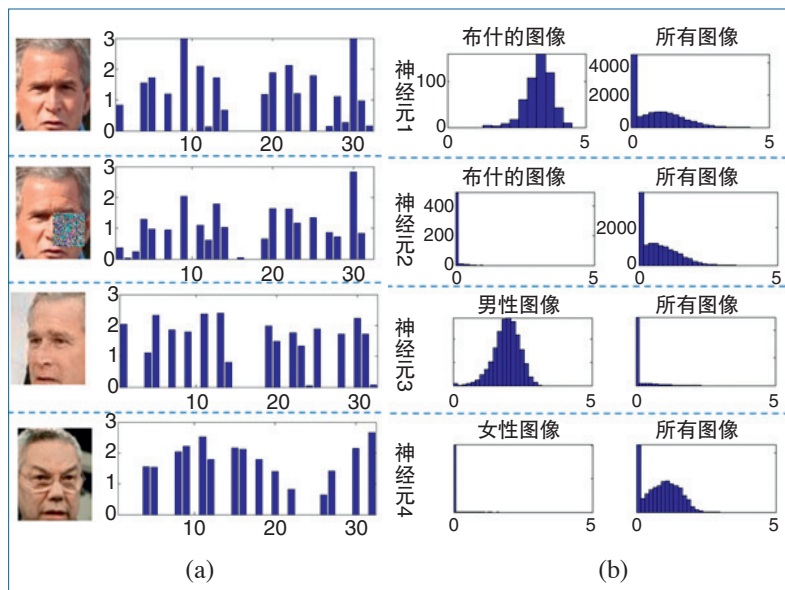


图5 DeepID2+最高隐含层神经元响应的稀疏性以及人脸身份和属性的选择性 (a)DeepID2+最高隐含层神经元对布什和鲍威尔人脸图像的响应: DeepID2+最高隐含层有512个神经元,我们采样32个神经元用于解释。当布什的人脸图像出现局部遮挡和姿态变化时,神经元响应的幅度就会发生一定变化,而响应与否的模式基本保持稳定。布什与鲍威尔人脸图像的响应模式存在明显的差别。(b)选择4个神经元,观察它们在4万张人脸图像、500张布什人脸图像、所有男性人脸图像和所有女性人脸图像上的响应直方图。每个神经元在大约一半的图像上有响应。神经元1几乎对所有的布什图像都有响应,神经元2几乎对所有的布什图像都没有响应,神经元3对绝大多数男性人脸都有响应,神经元4对绝大多数女性人脸图像都没有响应。)

一个隐含层，它被用作人脸的特征表示。

中度稀疏

如图 5 所示，针对输入的一张人脸图像，DeepID2+ 最高隐含层的神经元有的有响应，有的无响应，因此它的特征表示是稀疏的。当人脸出现姿态、遮挡、表情等类内变化时，神经元响应的幅度会发生一定变化，而响应与否的模式基本保持稳定。不同的响应模式会有较大差别。在对神经元的响应二值化后，得到的 0/1 编码在 LFW 的人脸确认任务上仍能达到 99% 以上的识别率^[7]。这意味着，在深度学习中，有价值的信息主要体现在神经元的响应模式中，而非响应幅度。

DeepID2+ 的特征表示是中度稀疏的^[7]。对于一张人脸图像，最高隐含层中大概有一半的神经元是响应的，一半不响应。这样两幅图像的汉明距离³可以达到最大。对于某一个最高隐含层的神经元，大概在一半的图像上有响应，而在剩下的图像上没有响应，可见深度模型的学习能力最强。

对人脸身份和人脸属性的选择性

DeepID2+ 最高隐含层的神经元响应不但稀疏，而且对人

脸身份和人脸属性具有很强的选择性。如图 5(b) 所示，有的神经元对某个人的绝大部分图像都有响应，而有的神经元对某个人的图像几乎都没有响应。这两种神经元对特定的人都有很强的识别能力。文献 [7] 中

类似的现象也发生在人脸属性上。如图 5(b) 所示，神经元 3 对绝大多数男性人脸图像都有响应，而对绝大多数女性人脸图像都没有响应。很多神经元都与人脸属性有清晰的对应关系。图 6 显示了如何将这些对

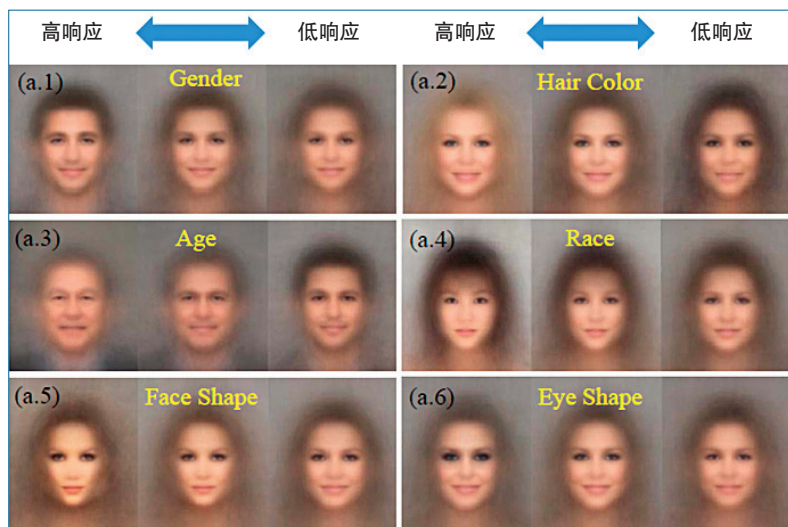


图6 可视化神经元所代表的人脸属性（针对某个神经元的响应，将所有人脸图像分为三组：高度响应、中度响应和低度响应。对每个组的人脸图像取平均，并显示在图中。神经元(a.1)~(a.6)分别代表了性别、头发颜色变化、年龄变化、种族变化、脸型变化和眼型变化。）

的实验结果表明，如果识别任务是将 LFW 的某个名人和其他名人分开，那么最好的单个神经元的识别率可高达 97%。这种特性是其他手工设计的特征所不具备的。例如高维局部二值模式 (Local Binary Patterns, LBP)^[6]（目前人脸识别中最有效的手工特征）中选出的最好的单个特征对，与之对应的单个名人的识别率只有 60%~70%。

应的属性可视化。虽然 DeepID2 只把人脸身份识别作为监督信号进行训练，但将它的特征直接用于 40 种人脸属性识别，仍然可取得很高的准确率，远远超过方向梯度直方图 (Histogram of Oriented Gradient, HOG) 等手工设计的特征的识别率^[16]。

对遮挡的鲁棒性

DeepID2+ 的人脸特征对局

³ 汉明距离是以理查德·卫斯里·汉明的名字命名的。在信息论中，两个等长字符串之间的汉明距离是两个字符串对应位置的不同字符的个数。

部遮挡具有很强的鲁棒性。当存在 10% 随机遮挡的情况时, LFW 的人脸识别率只下降不到 1%; 即使在 40% 的遮挡情况下, 识别率仍能达到 93.9%。但这些随机遮挡的模式只加在测试样本中, 并没有对 DeepID2+ 进行重新训练。与之对比, 随着遮挡的增加, 高维局部二值模式等手工设计特征的识别能力迅速下降。为了探究其原因, 文献 [7] 利用 DeepID2+ 各个卷积-池化层的特征分别进行人脸识别, 发现最低层的特征和高维局部二值模式类似, 也对遮挡非常敏感。这表明, DeepID2+ 对遮挡具有很强鲁棒性的原因在于它提取的是人脸的全局特征, 而 DeepID2+ 的最低层和高维局部二值模式提取的是局部特征。

结语

回顾人脸识别的发展历程, 深度学习给这一领域带来了突飞猛进的发展, 开辟了新的方向。特别是在最近短短一年时间内, 不断取得重要突破。另一方面, 用于人脸识别的深度模型与传统的人脸识别算法, 特别是子空间方法存在着重要的联系。正是过去十几年我们在人脸识别研究中积累的经验引导着深度学习在这一领域不断进步。从某种意义上讲, 联合深度学习是子空间方法在非线性和深度模型上的延续。■



王晓刚

香港中文大学助理教授。主要研究方向为计算机视觉、深度学习等。xgwang@ee.cuhk.edu.hk



孙 祎

香港中文大学博士生。主要研究方向为计算机视觉、深度学习、人脸识别、人脸配准。2010sunyi@gmail.com



汤晓鸥

IEEE 会士。香港中文大学教授。主要研究方向为计算机视觉、模式识别和视频处理。xtang@ie.cuhk.edu.hk

参考文献

- [1] X. Wang and X. Tang. A unified framework for subspace face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004; 26: 1222~1228.
- [2] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 1991; 3(1): 71~86.
- [3] B. Moghaddam, T. Jebara, and A. Pentland. Bayesian face recognition. *Pattern Recognition*, 2000; 33, 1771~1782.
- [4] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. Fisherfaces: recognition using class specific linear project. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997; 19, 711~720.
- [5] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical report, University of Massachusetts, Amherst, 2007.
- [6] D. Chen, X. Cao, F. Wen, and J. Sun. Blessing of dimensionality: Highdimensional feature and its efficient compression for face verification. *IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2013.
- [7] Y. Sun, X. Wang, and X. Tang. Deeply learned face representations are sparse, selective, and robust. arXiv:1412.1265, 2014.
- [8] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Attribute and simile classifiers for face verification. In *IEEE Int'l Conf. Computer Vision*, 2009.
- [9] G. B. Huang, H. Lee, and E. Learned-Miller. Learning hierarchical representation for face verification with convolutional deep belief networks. *IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2012.
- [10] Y. Sun, X. Wang, and X. Tang. Hybrid deep learning for computing face similarities. *IEEE Int'l Conf. Computer Vision*, 2013.
- [11] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. Deepface: Closing the gap to human-level performance in face verification. *IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2014.
- [12] Y. Taigman, M. Yang, M. Ranzato, L. Wolf. Web-Scale Training for Face Identification. arXiv:1406.5266, 2014.
- [13] Y. Sun, X. Wang, and X. Tang. Deep learning face representation from predicting 10,000 classes. *IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2014.
- [14] Y. Sun, X. Wang, and X. Tang.

Deep learning face representation by joint identification-verification. Neural Information Processing Systems, 2014.

[15]Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based Learning Applied to Document Recognition. Proceedings of the IEEE, 1998; 86: 2278~2324.

[16]L. Best-Rowden, H. Han, C. Otto, and et al.. Unconstrained face recognition: Identifying a person of interest from a media collection. TR MSU-CSE-14-1, 2014.

[17]Z. Liu, P. Luo, X. Wang, and et al.. Deep learning face attributes in the wild. arXiv:1411.7766, 2014.