
人脸识别综述^{*}

学生姓名¹

¹(东南大学 计算机科学与工程学院,南京 211189)

A Survey on Face Recognition^{*}

Student Name¹⁺

¹(School of Computer Science Engineers, Southeast University, Nanjing 211189, China)

+ Corresponding author: Phn: +1850098xxxx, E-mail: 2311990815@qq.com

Abstract: Starting in the sixties, face recognition has become one of the most researched topics in computer vision and biometrics. Traditional methods based on hand-crafted features and traditional machine learning techniques have recently been superseded by deep neural networks trained with very large datasets. In this paper we provide a comprehensive and upto-date literature review of popular face recognition methods including both traditional (geometry-based, holistic, feature-based and hybrid methods) and deep learning methods.

Key words: face recognition; machine learning; neural networks; computer vision; biometrics;

^{*}Supported by ...;

作者简介: 我...

摘要： 从 60 年代开始到现在，人脸识别一直是计算机视觉和生物识别研究的主题。传统的基于手动设计提取特征和传统机器学习的方法已经被使用大量数据训练出的深度神经网络所取代。本文给出一个从传统（基于几何、整体、基于特征和混合方法）到深度学习热门的人脸识别方法的综合论述。

关键词： 人脸识别；机器学习；神经网络；计算机视觉；生物识别；

1 引言

人脸识别是指能够从图像、视频中识别或验证主体（人）身份的技术。第一个人脸识别算法开始于 70 年代早期[1], [2]。从那以后，人脸识别的准确度一直在不断的提高，目前人脸识别的生物识别方法常常比其他传统成熟的生物识别（比如指纹、虹膜识别[3]）更受欢迎。这是因为人脸识别有着其独特的优点。其中一个就是人脸识别不涉及到隐私侵犯的问题，比如，指纹识别需要识别对象把手指放到指纹识别传感器上；虹膜识别需要识别对象的眼睛距离摄像机很近；而声音识别需要识别对象的声音足够响亮。而现在的人脸识别系统仅要求识别主体在摄像机的视野范围内即可（假设主体与摄像机在一个合理的距离内）。这使得人脸识别是一个最好的用户友好型生物识别方法。这意味着人脸识别拥有更广泛的应用前景，它可以被部署在那些用户不太配合的地方，比如视频监控系统。另外，人脸识别在门禁系统、安检系统、身份识别、社交媒体等方面也有着广泛的应用。

非限制条件下的人脸识别是最具挑战的生物识别方法之一，这种部署在现实生活中的人脸识别拥有很高的可变性（这种人脸图像通常称为自然脸，faces in-the-wild）。这些变化通常包括头部姿势、年龄、遮挡、光照和表情等。如图 1 所示。

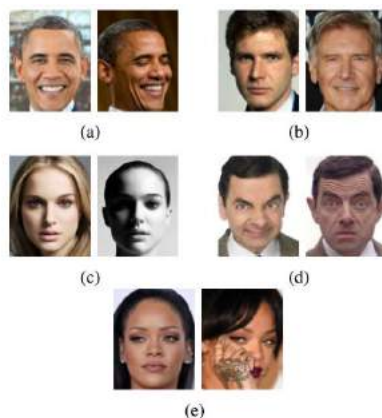


图 1：几种典型的自然脸。(a)头部姿势，(b)年龄，(c)光照，(d)表情，(e)遮挡。

多年以来人脸识别技术有了很大的发展。传统基于手工提取特征的边界、纹理描述器，与机器学习相结合的主成分分析（PCA）、线性判别分析（LDA）和支持向量机（SVM）。在非限制条件下产生的变化使得人工提取特征非常困难，以至于研究人员只能针对某种具体的变化设计专门的方法，比如年龄不变性方法[4],[5]；姿势不变性方法[6]，光照不变性方法[7],[8]等。最近，传统的人脸识别方法已经被基于卷积神经网络（CNNs）的深度学习的方法所取代。深度学习的主要优点是它可以通过一个很大的数据集学习到一个最好的特征表示。这种包含现实变化的自然脸可以通过网络来进行大规模的收集[9][10][11][12][13][14][15]。使用这些数据训练出来的基于 CNN 的人脸识别方法已经获得了很高的精确度，这是应为它们可以通过训练的真人脸图片学习到稳定的特征。另外，随着 CNNs 被用于解决许多其他的计算机视觉难题，深度学习方法在计算机视觉上的应用也同样加快了人脸识别技术的研究，比如物体的检测、识别和分割，光学字符识别、表情分析和年龄估计等。

人脸识别系统通常由以下几块组成：



图 2：人脸识别的组成部分。

1) 人脸检测 (face detection)。使用人脸检测器在图像中检测出脸的位置，如果检测到就用边框标记出来，返回边框的位置坐标。如图 3a 所示。

2) 人脸对齐 (face alignment)。人脸对齐是使用一组参考点来定位图像中的固定的几个位置来完成人脸图像的缩放和裁剪。在简单的 2D 人脸对齐中，通常需要使用人脸标志 (landmark) 检测器来找出一组人脸的标志位置与参考点相配合以找出最好的仿射变换 (affine transformation)。图 3b 和 3c 表示使用一组相同的参考点来实现人脸的对齐。更为复杂的 3D 人脸对齐算法 (如[16]) 可以实现人脸的转正，即可以改变人脸的位置使之面向正前方。

3) 人脸表征 (face representation)。在人脸表征阶段，人脸图像的像素值被转换成一个紧凑可判别的特征向量，也称作模板。事实上同一个主体的不同人脸图像应该映射成相似的特征向量。

4) 人脸匹配 (face matching)。在人脸匹配环节，通过比较两个人脸模板来获得一个表示属于同一主体的可能性的相似值。

人脸表征可以认为是人脸识别系统中最重要的一部分，也是本文第二部分文献综述关注的重点。

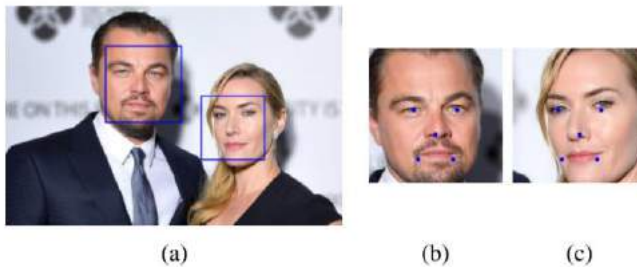


图 3：(a)用人脸检测器圈出人脸。(b)和(c)参考点和人脸对齐。

2 文献综述

早期的人脸识别关注的方法是使用图像处理技术来匹配简单描述的人脸几何特征。即使这些方法只能在非常苛刻的条件下才能工作，但是他们显示出了使用计算机实现人脸自动识别的可能性。之后，像主成分分析 PCA 和线性判别分析 (LDA) 等统计子空间方法变得越来越流行。这些方法因为使用整个面部区域作为输入而称作整体(holistic)方法。同时，计算机视觉领域的进步导致局部特征提取器的发展，它可以描述一幅图像中不同区域的纹理。这种通过匹配人脸图像的局部特征来实现的人脸识别称作基于特征的人脸识别方法 (feature-based method)。整体方法和基于特征的方法得到进一步的发展，然后又结合起来形成混合的方法。直到深度学习出现并成为计算机视觉应用包括人脸识别的主要方法以前，基于混合方法的人脸识别系统依然代表着人脸识别的最高水平。本文后面的部分给出了前面几种方法中一些最具代表性研究工作的概述。

2.1 基于几何的方法 (Geometry-based Methods)

70 年代早期 Kelly[1]和 Kanade[2]博士论文被认为是最早关于自动人脸识别的研究工作。他们提出使用专门的边界和轮廓探测器来找出一系列的人脸标志，再测量它们之间的相对位置和距离。早期的系统的精确度仅在很小的数据库上演示 ([1]中使用了 10 个主体的数据库而[2]中使用了 20 个人的数据库)。在[17]中，一种类似[2]的基于几何的方法与另一种把人脸图像描述为梯度图像的方法做了比较。作者表示梯度图

像的方法比基于几何的方法识别准确度好，但是基于几何的方法速度更快内存使用更少。使用人脸标志和它们之间的几何关系进行人脸识别在[18]中作了详尽的研究。具体来说，他们提出了一种基于测量两组人脸标志之间的普鲁克距离(Procrustes distance)方法和另一中基于测量人脸标志之间距离比例的方法。作者认为虽然其他能够提取更多信息的方法（如整体方法）可以达到更高的精确度，该基于几何的方法速度更快而且可以和其他方法结合成混合方法。由于在 3D 标志中编码了深度信息[20][21]，基于几何的 3D 人脸效率更高，这已经得到了证明。

基于几何的人脸识别在早期的人脸识别研究中是至关重要的。它可以作为一个该综述后面所描述的先进方法的更快替代选择或与它们结合起来。



图 4：从 ORL 人脸数据库中计算出的本征脸，按差异从大到小排序的前 5 个。

2.2 整体方法 (Holistic Methods)

整体识别方法是用整个人脸区域的信息来描述人脸特征，通常是把人脸图像投射到一个低维空间来过滤掉不相关的细节和其他影响识别的变量。这种类别中，其中一个比较流行的方法是使用主成分分析 (Principal Component Analysis, PCA)。在[22][23]首先提出这种想法，把 PCA 应用在一组训练的人脸图像，以获得可以表示数据分布差异的本征向量，由于该本征向量与人脸的相似性通常也叫做本征脸。如图 4 所示。该方法通过把被识别对象的人脸图像投射到由本征脸构成的子空间中来获取本征脸线性组合的权重。在[24]中使用了这种想法，通过比对识别对象的权重和数据存储的权重来输出人脸识别的结果。在[25]中提出了基于概率的版本，即使用贝叶斯算法来分析图像差异。这种方法使用两个本征脸集合分别构造了对象内部 (intra-person) 和对象之间 (inter-person) 的变量模型。在原始的本征脸方法的基础上还提出了很多变体版本。例如[27]提出的基于核方法的非线性扩展的 PCA，即核 PCA[26]；独立成分分析 (ICA) [28]，在[29]中提出的一种可以捕捉像素间的高阶相关性的 PCA 泛化；在[30]中提出的基于 2D 图像矩阵的 2 维 PCA，而不是 1D 向量。

基于 PCA 的方法存在一个问题，就是投影时会最大化训练集中所有图像的差异。这表明顶端的本征向量可能会对识别准确度产生负面影响，因为它们可能是与识别任务无关的对象内的 (intra-person) 变量(比如光照、姿势和表情)。在[32]中提出了基于线性判别分析 (LDA) 的整体方法，也称作 Fisher 判别分析，用来解决[33][34][35][36]中的问题。LDA 的主要思想是使用类标签找到一个投影矩阵 W 来最大化类间差异同时最小化类内差异：

$$W^* = \arg \max_W \frac{|W^T S_b W|}{|W^T S_w W|} \quad (1)$$

这里的 S_w 和 S_b 类内 (within-class) 和类间 (between-class) 散度矩阵，定义如下：

$$S_w = \sum_k \sum_{x_j \in C_k} (x_j - \mu_k)(x_j - \mu_k)^T \quad (2)$$

$$S_b = \sum_k (\mu - \mu_k)(\mu - \mu_k)^T \quad (3)$$

这里的 x_j 表示数据样本， μ_k 是类 C_k 的均值， μ 是全局均值， K 是数据库中的类号，通过计算分离矩阵 $S = S_w^{-1} S_b$ 的本征向量可以求出公式 (1) 的解。与 PCA 类似，LDA 也可以通过选择最大本征值对应的本征向量的子集来实现降维处理。虽然在人脸识别中 LDA 与 PCA 相比更具有优势，但是单纯基于 LDA 方法的人脸识别很可能因为类内的散度矩阵估计不准而产生过拟合现象[35][36]，特别是在使用高维度小样本集来训练的时候容易发生。在极端情况下 S_w 变成奇异矩阵并且 W 不可解[33]。因此，通常先使用 PCA 的方

法来对数据进行降维处理，再进行 LDA 处理[33][35][36]。LDA 同时也扩展到使用核方法的非线性情况[37][38]和基于概率的 LDA[39]。

支持向量机（Support Vector Machines, SVMs）的方法常常被用于整体人脸识别中。在[40]中用图像差异来训练 SVM，把人脸识别转变成二分类的问题来处理。具体说就是两个包含同类中不同图像的差异和不同类之间的差异集合（这种公式与[25]提出的基于概率的 PCA 很相似）。另外，可以对传统 SVM 公式增加一个参数来控制整个系统的操作点。在[40]中，为每一个类训练了一个单独的 SVM，他使用了基于 PCA 投影和 LDA 投影分别训练得到的 SVM 进行实验，结果是与使用 PCA 投影训练的简单欧氏距离相比，这种方法效果更好，因为 LDA 已经编码了识别面部所需的判别信息。

在[42]中何晓飞提出了一种与 PCA 和 LDA 相关的局部保留投影（Locality Preserving projection, LPP）的方法。PCA 和 LDA 都是在最大化差异和判别信息时保留图像空间的整体结构，而 LPP 的目标确实保留图像空间的局部结构。这意味着由 LPP 得到的投影图像在 LPP 子空间中与临近点具有相似的局部信息。比如，同一个人开口和闭口的不同图像使用 LPP 将会投影成相似点，但是这在 PCA 和 LDA 中却是不可能的。这种方法在多数数据集中明显优于 PCA 和 LDA。在[43]中使用正交向量来做 LPP 获得了更好的效果。

另一类流行的整体识别方法是基于人脸的稀疏表示（Sparse Representation of Faces）。在[44]中第一提出基于稀疏表示分类（sparse representation-based classification, SRC），使用训练图像的线性组合来表示人脸的想法：

$$\mathbf{y} = \mathbf{A}\mathbf{x}_0 \quad (4)$$

这里 \mathbf{y} 是测试图像， \mathbf{A} 是包含所有训练图像的矩阵， \mathbf{x}_0 是一个稀疏系数的向量。通过执行稀疏表示，大多数非零系数归属到正类的训练图像。在测试时归属于各类的系数用来重构图像，获得最小重构误差的分类被认为是正确的那个分类。这种方法可以通过对该线性组合增加一项稀疏误差系数 \mathbf{e}_0 来增强对噪音、遮挡等损坏图像的处理：

$$\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{e}_0 \quad (5)$$

这里的非零项 \mathbf{e}_0 对应图像中损坏的像素。为了增强人脸识别的健壮性和减少计算的复杂性，有很多其他改进方法被提出，比如，在[45]中提出的使用判别的 K-SVD 算法选择更为差别紧凑的训练图像集来重构图像；在[46]中提出使用马尔可夫随机场对遮挡空间连续性的先验假设进行建模，以此来扩展 SRC；在[47]中提出通过对图像中的每个像素点增加权重来获得更好的重构图像。

最近，受基于概率的 PCA[25]的启发，有人提出了联合贝叶斯的方法[48]。该方法把人脸图像描述为表示同一个人（intra-person）的和不同人（inter-person）的两个独立的高斯变量之和。这种方法在具有挑战性的 LFW（Labeled Faces in the Wild）数据库[49]达到了 92.4% 的识别准确度。这是该数据库使用整体方法达到的最高准确度记录。

由大量文献提出的人脸识别方法中可以证明，整体人脸识别方法对于现实世界人脸识别系统的发展至关重要。在下面的部分将讨论一类流行的可以替代整体方法的基于特征的人脸识别。

2.3 基于特征的方法

基于特征的人脸识别是一种提取并放大大人脸图像不同区域的特征差异。与计算脸部特征几何关系的基于几何的人脸识别不同，基于特征的人脸识别的关注重点是提取人脸的特征差异而不是计算特征的几何关系（从技术上讲，基于几何的方法可以看作基于特征方法的一个特例，因为许多基于特征的方法也放大提取特征的几何关系）。当人脸图像出现表情、光照等局部变化时，基于特征的方法比整体方法更具有优势。比如，同一个人的两张人脸图像，一张是张开眼睛的，另一张是闭着眼睛的，在基于特征的人脸识别中仅仅与眼睛相关的特征向量的系数不同，然而在整体人脸识别中所有的特征向量系数都可能不同。另外在基于特征的方法中很多描述器针对不同的变化（如缩放、旋转和平移）都被设计成不变的。

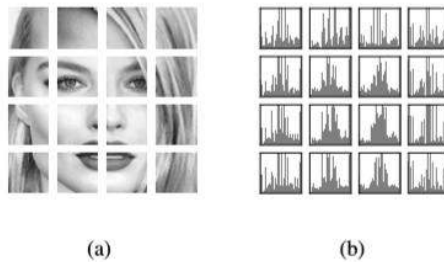
第一个基于特征的人脸识别是在[50]中提出的模块化本征脸方法（modular eigenfaces method），是原始本征脸技术的一种扩展。该方法时把 PCA 独立的运用到人脸图像不同的局部区域来产生一组本征特征

(eigenfeature)。虽然在[50]中表示本征特征和本征脸通常可以达到相同精确度，但是当只有很少的本征向量可以使用时，本征特征的精确度会更好。

在[51]中提出了双边特征(binary edge features)的方法。他们的贡献是改善了[52]中用来比较二进制图像的 Hausdorff 距离。Hausdorff 距离是通过考虑一个集合中的点到另一个集合中的点的最小距离的最大值来测量两个点集的邻近性。[51]中提出改进的 Hausdorff 距离要求一个集合中的点与另一个集合中的点必须是邻近点。这个性质可以使得该方法在小的、非刚性局部扭曲图像上更健壮。作为这个方法的变种，[53]提出使用线边缘图(line edge maps, LEMs)进行人脸表示。LEMs 提供了一种紧凑型的人脸表示，因为边界被编码成线段，即只使用了终点坐标。另外为了配合该方法作，者还提出了线段的 Hausdorff 距离。该距离不适合用来匹配不同方向的直线，更适合直线的平移，并把两个 LEMs 中的直线数量的差异度量结合起来。

另一种流行的特征人脸识别方法是弹性约束图匹配(elastic bunch graph matching, EBGM)[54]，是[55]中的一种动态链结构(dynamic link architecture)的扩展。在这种方法中使用节点图来表示一张人脸。该节点包含从一组预定义的人脸标志周围提取的小波系数[56]。在训练中，通过手工定位训练图像的节点的叠加来生成人脸约束图像(face bunch graph, FBG)。当出现一张新的测试人脸时就会生成一个新的图像，并通过在 FBG 模型中搜索最相似的节点来匹配人脸标志。两张图像通过测量他们的图像节点的相似度来进行比较。在[59]中提出该方法的另一个版本是使用方向梯度直方图(histograms of oriented gradients, HOG)[57][58]代替小波特征，由于 HOG 描述器更适应光照、旋转、小位移的变化，该方法比原始的 EBGM 表现更好。

随着局部特征描述在其他计算机视觉应用的发展[60]，基于特征的人脸识别方法变得越来越流行。在[61]中独立提取局部 LBP 描述器的直方图，并连接成一个全局特征向量。如图五所示。另外，他们使用带权卡方距离(Chi-square distance)来测量两个特征向量 a 和 b 的相似性：



图五：(a)人脸图像被分成 4×4 的局部区域。(b)从每个局部区域计算出来的 LBP 描述器直方图。

$$\chi^2(a, b) = \sum_i \frac{w_i(a_i - b_i)^2}{a_i + b_i} \quad (6)$$

这里 w_i 是控制特征向量第 i 个元素贡献的权重。如[62]中所示，很多这种方法的变体别提出用来来提高人脸识别的准确度，或者用来处理其他相关的工作，如人脸检测、人脸表情分析和人口分类。比如在[63][64]中提出的 LGBP 描述器，即从 Gabor 特征图中提取的 LBP 描述器；在[65]中提出的旋转不变性 LBP 描述器，即把傅里叶变换应用到 LBP 直方图中。在[66]中提出了局部导数模式(local derivative pattern, LBP)通过编码方向模式特征来提取高阶局部信息的 LBP 的变体。

尺度不变特征转换(SIFT)描述器[67]也被广泛用于人脸识别。在[68]中提出了三种不同的方法用遍历人脸图像来匹配 SIFT 描述器：(i) 计算所有 SIFT 对之间的距离并用最小距离作为相似度的分数；(ii) 与 (i) 相似，但是眼睛和嘴巴周围的 SIFT 描述器是独立比较的，并且两个最小距离的平均值作为相似度的分数；通过一个规则的网格来计算 SIFT 描述器，并使用相应描述器对之间的平均距离作为相似度分数。在[69]中提出了一个相关的方法，使用加速强化特征(speeded up robust features SURF)[70]代替 SIFT。在

这项工作中作者发现通过规则的网格进行密集特征提取能达到更好的结果。在[71]中提出了两个 SIFT 的变体，即根据他们的尺度来去除不可靠的关键点的 volume-SIFT 和在大尺度和人脸边界附近发现关键点的 partial-descriptor-SIFT。与原始的 SIFT 相比，这两种方法都提高了人脸识别的准确度。

一些基于特征的方法致力于从训练样本中学习局部特征。例如，在[72]中用非监督学习技术（K-means[73]，PCA 树[74]个随机映射树[74]）把脸部局部微结构编码成一组离散编码。然后离散编码在不同的人脸区域组成直方图。在[75]中提出了类似 LBP 基于学习的描述器。具体来说，这个描述器是由两部分组成的，一个是把局部 3×3 区域的中心像素减到邻近的像素所产生的差异模式；另一个是计算差异模式的高阶统计的高斯混合模型的训练。在[76]中提出了带有学习步骤的 LBP-like 描述器。在这项工作中，使用了 LDA 学习一个增强对图像差异的判别能力的滤波器，同时也学习到了可以反映每一个局部区域的邻近像素对差异模式贡献的权重。

对于不同的版本，基于特征的方法以显示比整体方法更稳定。但是却丢失了一些整体方法的优点，如丢掉了非判别信息和更紧凑的表示方式。下面将讨论结合这两种方法的混合方法。

2.4 混合方法

混合方法结合了整体和基于特征方法的技术。在深度学习流行之前，大多数最高水平的人脸识别系统都是基于混合方法的。一些混合方法仅仅使用了两种不同的技术，这两种技术根本不做什么互动。例如之前提到的[50]中的模块化本正脸，作者使用本征脸和本征特征相结合的表示方法来实验，结果比使用任何单一的一种方法达到的准确度都高。然而最流行的混合方法都是提取局部特征（如 LBP, SIFT）并把它们投影到低维空间或判别子空间（使用 PCA 或 LDA）。如图 6 所示。

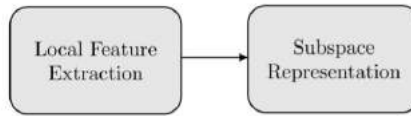


图 6: 典型的混合人脸表示

在[77][78][79]中提出了一些 Gabor 小波特征与不同子空间相结合的方法。在这些方法中把一幅图像不同方向和尺度的 Gabor 核做卷积，再把输出连接成一个特征向量。然后对该特征向量进行降采样降低它的维度。在[77]中使用了[80]提出的增强线性判别模型对该特征向量作进一步处理。在[78]中使用 PCA 之后又用了 ICA 对特征向量进行降采样，并使用了[80]中的概率论证模型对两张图像是否属于同一个人进行分类。在[79]中使用了含有多项式核的 PCA 核增强特征向量的高阶统计。所有的混合方法都比单独使用 Gabor 小波特征的精确度要高。

在许多混合方法中 LBP 描述器已经成为关键的部件。在[81]中把一幅图像分割成多个不重复的区域，并在多种方案中提取 LBP 描述器。每个区域的 LBP 系数被连接成区域特征向量并投影到 PCA+LDA 子空间中去。在[82]中把该方法应用到彩色图像中了。在[83]中把拉普拉斯 PCA, 一种 PCA 的扩展，应用到 LBP 描述器中，结果比标准的 PCA 和核 PCA 表现更好。在[84]中使用了两个新颖的 LBP 的 patch 版本，即 three-patch LBP (TPLBT) 和 four-patch LBP (FPLBP)，与 LDA 和 SVMs 结合起来。这两种方法可以通过编码相邻片像素的相似性来增强人脸识别的准确度。最近，在[85]中提出了通过在人脸标志周围密集提取多尺度 (multi-scale) LBP 描述器来进行高维人脸表示。这个高维(十万维)特征向量使用 PCA 降到 400 维，然后最终的判别特征向量使用联合贝叶斯习得。在[85]的实验中高维特征提取从 1 千维增加到 10 万维可以提高 6-7% 的人脸识别的准确度。这种方法的缺点是对如此大规模的维度进行降维需要很高的计算代价。因此，他们提出使用稀疏线性投影矩阵 B 解下式优化问题对 PCA 和联合贝叶斯转换进行近似计算：

$$\min_B \|Y - B^T X\|_2^2 + \lambda \|B\|_1 \quad (7)$$

这里的第一项是高维特征向量矩阵 Y 和其在低维特征向量上的投影矩阵 X 之间的重构误差；第二项是执行

稀疏性的投影矩阵； λ 是平衡两项贡献的参数。最近又提出了基于判别高斯处理浅变量模型（即高斯脸[86]）的多任务学习方法。这种方法扩展了[87]提出的高斯过程方法并结合了计算更高效的核 LDA 版本来从 LBP 描述器中学习人脸表示，该描述器可以从多资源域中发掘数据。使用这种方法在 LFW 数据库可以达到 98.52% 的准确度。这足以和深度学习的准确度相媲美。

一些混合方法还提出了使用不同的局部特征相组合的方法。例如，在[88]中使用了 Gabor 小波和 LBP 特征。作者认为两种特征采集了互补的信息。尽管 LBP 描述器只能采集到很少的表面细节，Gabor 小波特征却能在更大尺度范围内编码人脸外形。PCA 被独立应用到包含 Gabor 小波系数的特征向量和 LBP 系数来减少它们的维度。最终的人脸表示是通过连接两个 PCA 转换的特征向量并应用名为核判别通用向量[89]，一个类似核 LDA 的子空间方法，来获得的。在[90]中提出了另一种使用 Gabor 小波和 LBP 特征的方法。在这种方法中对包含 LGBP 描述器直方图[64]的区域使用 PCA+LDA 来表示人脸。在[8]中提出了多特征系统用来处理不良光照条件下的人脸识别。这项工作中有三项贡献：(i) 一个预处理过程用来减少光照变化的影响；(ii) 一个 LBP 的扩展，叫局部三重模式（local ternary patterns, LTP），它判别能力更强，在均匀区域受噪音的影响更小；(iii) 结合了一组 Gabor 小波和 LBP/LTP 特征核 LDA 的结构，用来对规范和融合进行评分。在一个相关的方法中[91]提出了一个对模糊很有效的新颖的描述器，他把局部相位量化（local phase quantization, LPQ）[92]扩展成多多尺度的（MLPQ）。另外在核 LDA 框架中使用了核融合技术把 MLPQ 描述器和 MLBP 描述器结合起来。在[5]中提出了年龄不变性的人脸识别系统，该系统是基于 SIFT 的密集提取和结合了新的多特征判别分析（MFDA）的多尺度 LBP 描述器。该 MFDA 技术使用自由子空间采样构建多个低维特征子空间并为 LDA 袋选训练样本子集，为了增强表示的判别能力该 LDA 包含分类边界附近的类间对。在[95]中也使用了密集 SIFT 描述器作为纹理特征并结合了以人脸标志对相关距离形式的外形特征。这种外形和纹理的结合再使用多 PCA+LDA 变换作进一步处理。

总结一些这个部分，与图 6 描述的过程不同的其他混合方法已经评论过了。在[96]中，通过训练属性和明喻二分类 SVM 分类器使用低位局部特征（如图像在 RGB 和 HSV 颜色空间、边界数量和梯度方向的强度）来计算高位视觉特征。属性分类器检测脸部可描述的属性，比如性别、种族和年龄。另一方面，明喻分类通过测量面部的不同部分与有限的一组参考对象的相似性来检测不可描述的属性。为了比较两张图像，两张图像明喻分类器和属性的输出都要送入到 SVM 分类器中。在[97]中提出了一种与[96]中明喻分类器相似的方法。主要的不同是，[97]中使用了大量的一对一的简单分类器而不是[96]中复杂的一对多的分类器，并且使用 SIFT 描述器作为低位特征。在[98]中提出了两种用于人脸鉴定的度量的学习方法。第一个叫逻辑判别度量学习（logistic discriminant metric learning, LDML），基于两个正例（属于同一个主体）之间的距离应该小于两个负例（属于不同的主体）之间的距离的思想。第二个叫边缘化 kNN（marginalised kNN, MkNN），使用 k 最近邻分类器来发现两个比较向量的附近能构成多少个正邻对。两种方法都是使用在脸部固定位置（嘴角、眼睛和鼻子）计算的 SIFT 描述器的向量对来进行训练。

混合方法提供了最好的整体方法和基于特征的方法。它们的主要限制是选择可以完全提取识别面部所需信息的良好特征。一些方法通过组合不同种类的特征来克服这个问题，然而另外一些人却引入了学习步骤来提高特征的判别能力。下面讨论的深度学习方法通过训练端到端点的系统把这种想法又向前推了一步，这些系统可以学习大量适合识别任务的特征。

2.5 深度学习人脸识别方法

深度学习人脸识别最常用的方法是卷积神经网络（CNNs）。深度学习的主要优点是它可以通过使用大量的训练数据来学习到一个稳定的、适用各种变化的人脸模型。对于光照、姿势、表情、年龄等变化，CNNs 通过大量的训练数据就能够很好的识别出来。缺点就是它需要包含各种变化的数据来训练以便能归纳出未知的样本。幸运的是几个包含自然脸图像的大规模人脸数据库最近被发布到公共领域供 CNN 模型训练使用[9][10][11][12][13][14][15]。神经网络除了可以学习判别特征，还可以用来降低维度，被训练成一个分类器或用作度量学习方式。CNNs 是点对点的训练系统，不需要与其他具体的方法结合使用。

CNN 人脸识别模型可以通过不同的方法来训练出来。其中一个就是把识别作为分类问题来处理，每一张人脸对应一个类。训练之后，这个模型可以通过丢弃分类层或使用先前层作为人脸表示来识别不在训练集出现的人脸[99]。在深度学习文献中，通常这些特征称作瓶颈（bottleneck）特征。通过第一阶段的训练之后，可以使用其他技术优化瓶颈特征来进一步训练该模型。比如，使用联合贝叶斯建模[9]或者使用不同的损失函数来微调 CNN 模型[10]。另一种常用学习人脸模型的方法是通过优化两张脸[100][101]或三张脸[102]之间的距离度量来直接学习瓶颈特征。

使用神经网络进行人脸识别的想法并不新鲜，早在 1997 年 Lin S.H.就提出基于概率决策的神经网络（PBDNN）[103]用于检测人脸、定位眼睛、识别人脸。PBDNN 人脸识别把每一个训练主题分割成两个完全关联的子网以减少隐藏单位的数量同时避免过拟合。使用强度和边界特征分别训练两个 PBDNN 联合输出最终的分类决策结果。在[104]中提出了使用自组织映射（self-organising map, SOM）和卷积神经网络相结合的方法。自组织映射[105]是一种非监督学习的神经网络，它把输入数据投影到低维空间来保留输入空间的拓扑特性。这就是说，在输入空间邻近的样本在输出空间也是邻近的。值得注意的是这两种方法都没有使用点对点的训练方式（[103]中使用了边界特征，[104]中使用了 SOM），因此它们都是浅层的神经网络。在[100]中提出了点对点的人脸识别 CNN，该方法使用对比损失函数[106]来训练一个 Siamese 结构。对比损失在执行一个度量学习过程，该过程最小化同一主体的特征向量之间的距离，同时最大化不同主体的特征向量之间的距离。这种使用小样本训练出来的 CNN 结构依然是浅层的神经网络。

表 1：公共的大规模人脸数据库

Dataset	Images	Subjects	Images per subject
CelebFaces+ [9]	202,599	10,177	19.9
UMDFaces [14]	367,920	8,501	43.3
CASIA-WebFace [10]	494,414	10,575	46.8
VGGFace [11]	2.6M	2,622	1,000
VGGFace2 [15]	3.31M	9,131	362.6
MegaFace [13]	4.7M	672,057	7
MS-Celeb-1M [12]	10M	100,000	100

由于当时只能够获得较少的训练集并且神经网络的容量很小，上面的方法都没有获得突破性的进展。直到后面神经网络模型的规模变大可以使用大量的数据[107]来训练时，第一个基于深度学习的人脸识别[99][9]才达到最高水平。特别是 Facebook 的 DeepFace 就是一个基于大规模 CNN 的人脸识别，在 LFW 上的识别准确度高达 97.35%，错误率比之前的最高水平低了 27%。作者使用 4030 个主体的 440 万张人脸结合 softmax 激活函数与 cross-entropy 损失函数（softmax loss）训练出一个 CNN。这项工作有两个创新点：（i）基于显性 3D 人脸建模的高效人脸对齐系统；（ii）包含局部连接层的 CNN 结构[108][109]，与常规的卷积层不同，它可以从图像的每个区域学习到不同的特征。最近 DeepID 系统通过训练建立在包含 10 个区域、三个尺度、RGB 或灰色通道的补片上的 60 个不同的 CNNs 获得了类似的结果。在测试中，从每个补片和它的水平翻转副本中提取 160 个瓶颈特征构成 19200 维特征向量（ $160 \times 2 \times 60$ ）。与[99]相似，该 CNN 结构也使用了局部连接层。通过 CNNs 提取的 19200 维特征向量训练一个联合贝叶斯分类器[48]获得了证明结果。这个系统使用了 10177 个名人的 202599 张人脸图像数据库进行训练。

影响基于 CNN 的人脸识别的因素有三个：训练数据、CNN 结构、损失函数。大多数深度学习的应用中都需要大量的训练数据来防止出现过拟合现象。通常每一类的训练样本越多 CNN 训练出来的分类器准确度就越高。这是因为 CNN 模型暴露在更多的同类变量之中学习到的特征就越强大。然而我们更关注的是提取那些可以泛化主题的特征，而不是出现在训练集中的特征。因此用于人脸识别的数据集需要包含很大数量的主体，这样 CNN 模型才有可能暴露在更多的类间变量中。在[110]中研究了数据库中的主体数量对识别准确度的影响。在这项工作中，一个大型的数据库按照每个主体人脸图像的多少从小到大排序。然后用主体数量渐增的不同子集对 CNN 进行训练。拥有图像最多的前 10000 个主体训练出的准确度最高。

增加更多的主体会降低识别准确度，因为每个额外的主题分配的图像数量会很少。在[111]中做了另一项研究，广度数据库是不是优于深度数据库，或则反之亦然（如果一个数据库比另一个数据库包含更多的主体就被认为是广度的，如果一个数据库每个主体包含的图像比另一个更多就认为是深度的）。从这项研究中可以得出结论，相同数量的图像，广度数据库提供的准确度更高。作者认为是由于广度数据库包含更多的类间变量，因此在未知的主体上有更好的泛化性能。表一给出了用于人脸识别 CNN 训练的常用公共数据库。

人脸识别的 CNN 结构一直受到 ImageNet 上的大规模视觉识别挑战（ILSVRC）所达到的最高准确度的启发。例如在[11]中使用了一个 16 层的 VGG 网络[112]，在[10]中使用了类似但较小的这种网络。在[102]中国探索了两种不同的 CNN 结构：VGG 样式的网络[112]和 GoogleNet 样式的网络[113]。尽管两种网络的准确度相差不大，但是 GoogleNet 样式的网络的参数要小 20 倍。特别是最近，许多物体识别任务，包括人脸识别[115][116][117][118][119][120][121]，更倾向于选择残差网络（ResNets）[114]。ResNets 的主要创新是引进了使用捷联学习残差映射的块，如图 7 所示。使用残差可以允许训练更深的结构，因为这有利于信息的穿层

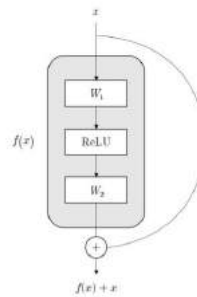


图 7：在[114]中提出的原始残差块

流动。在[121]中对不同的 CNN 结构做了深入的研究。权衡准确度、速度和模型规模，最好的是带有类似[122]提出的残差块的 100 层 ResNet 结构。

为基于 CNN 方法的训练选择损失函数一直是人脸识别最活跃的研究领域。即使使用 softmax loss 训练的 CNNs 一直非常成功[99][9][10][123]，但是这种损失函数对不在训练集中的主体泛化性能不好。这是因为 softmax loss 擅长在增加类间差异的训练集中学习特征并分类，但是减少类内变化却未必这样。为了减轻这种问题，提出了几种方法。一种简单的方法是使用如联合贝叶斯的判别子空间方法来优化瓶颈特征，如[9][124][125][126][10][127]中的做法。另一种方法是使用度量学习。例如，在[100][101]中对比损失仅用作监督信号并于分类损失[124][125][126]相结合。人脸识别中最流行的一种度量学习方法是三元损失函数[128]，在[102]中被首先用于人脸识别任务。三元损失的目的是通过边界把正对之间的距离从负对之间的距离中分离出来。更正式的讲，对于每个三元 i 需要满足下面的条件[102]：

$$\|f(x_a) - f(x_p)\|_2^2 + \alpha < \|f(x_a) - f(x_n)\|_2^2 \quad (8)$$

这里 x_a 是一个固定图像， x_p 是同一个主体的图像， x_n 是不同主体的图像， f 是模型所学习的映射， α 是施加在正对与负对之间的边界。在实际应用中，使用三元损失训练 CNN 要比 softmax 损失收敛的慢，这是因为大量的三元组（或者对比损失对）需要覆盖到整个训练集上。尽管这个问题训练中可以通过选择硬三元组（即不满足边界条件的三元组）来得到改善[102]，而通常的做法是先使用 softmax 损失进行训练，然后在使用三元组对瓶颈特征进行微调[11][129][130]。一些三元损失的其他版本被提出。例如，在[129]中使用点积代替欧式距离作为相似性的度量；在[130]中提出了基于概率的三元损失；在[131][132]中提出了三元损失的修改版，最小化正得分和负得分的标准差。在[133]中提出了一个用于学习判别特征的中心损失函数。中心损失的目标是最小化瓶颈特征间的距离和它们相应的类中心距。通过使用 softmax 和中心损失联合训

显示由 CNN 学习到的特征能够有效的增加类间差异 (softmax loss) 同时减少类内差异 (中心损失)。中心损失还有效率高的有点, 同时对比损失和三元损失更易于执行, 因为在训练中不需要形成对或者三元组。另一个相关的度量学习方法是[134]提出的区间损失, 目的是为了提高在不平衡数据库中的训练。区间损失由两部分组成。在每次批训练中, 一个是最小化同类中 k-最大距离样本类内损失, 另一个是最大化最小类中心距的类间损失。通过使用这些极端情况, 中心损失从每个类中使用相同的信息, 忽视每个类中能够获得样本数量。与中心损失类似, 区间损失需要与 softmax 损失结合使用以避免损失被降到 0[133]。

组合不同的损失函数所带来的一个困难是如何正确的平衡它们。最近提出了一些修改 softmax 损失的方法, 以便不需要与其他损失函数结合就可以学习判别特征。其中一个增加瓶颈特征判别能力的方法是特征规范化[115][118]。比如, [115]提出规范特征使它们拥有共同的 L_2 -norm, [118]提出规范特征使它们拥有 0 均值和相同的差异。一个非常成功的研究是在 softmax 损失的每一类决策边界上引入余边[135]。为了简化, 考虑使用 softmax 损失的二分类情况。在这种情况下, 每一类的决策边界通过下式给出 (如果偏置为 0):

$$\|\mathbf{x}\|(\|\mathbf{W}_1\| \cos \theta_1 - \|\mathbf{W}_2\| \cos \theta_2) = 0 \quad (9)$$

这里 \mathbf{x} 是特征向量, \mathbf{W}_1 和 \mathbf{W}_2 是对应每一类的权重, θ_1 和 θ_2 是 \mathbf{x} 分别与 \mathbf{W}_1 和 \mathbf{W}_2 的夹角。在等式 9 中引入乘数余边 m , 这两个决策边界变得更严格了:

$$\|\mathbf{x}\|(\|\mathbf{W}_1\| \cos m\theta_1 - \|\mathbf{W}_2\| \cos \theta_2) = 0 \text{ for class 1} \quad (10)$$

$$\|\mathbf{x}\|(\|\mathbf{W}_1\| \cos \theta_1 - \|\mathbf{W}_2\| \cos m\theta_2) = 0 \text{ for class 2} \quad (11)$$

如图 8 所示, 余边使类间分的更开, 并且类内更紧密。根据余边如何结合损失提出了几个可供替换的方法[116][119][120][121]。例如, 在[116]中权重向量别规范化拥有统一的范数, 这样决策边界仅依赖于 θ_1 和 θ_2 角。在[119][120]中提出了附加的 cosine 余边。与乘数余边[135][116]相比, 附件余边更容易执行和优化。在这项工作中, 除了规范化权重向量, 还对特征向量进行[115]中的正规化并缩放。在[121]中提出了一种可供选择的附加余边, 它除了保持[119][120]的优点还有一个更好的几何解释, 因为余边被附加到角度上而不是 cosine 上。表 II 总结了带有余边不同 softmax 损失版本的决策边界。这些方法都代表着目前人脸识别的最高水平。

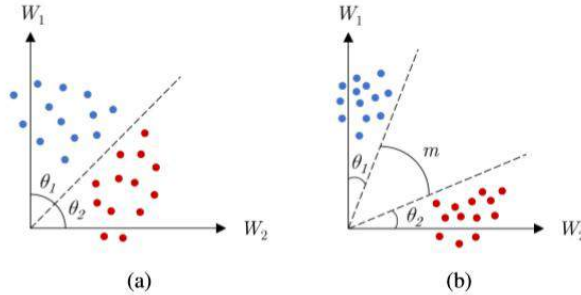


图 8: 在两类决策边界上引入 m 边界的效果。(a) softmax 损失 (b) 带有余边的 softmax 损失。

表 II: 带有余边 softmax 损失不同版本的决策边界。注意这仅是二分类情况下类 1 的决策边界。

Type of softmax margin	Decision boundary
Multiplicative angular margin [116]	$\ \mathbf{x}\ (\cos m\theta_1 - \cos \theta_2) = 0$
Additive cosine margin [119], [120]	$s(\cos \theta_1 - m - \cos \theta_2) = 0$
Additive angular margin [121]	$s(\cos(\theta_1 + m) - \cos \theta_2) = 0$

3 总结

随着计算机视觉的应用，我们见证了人脸识别的发展，由手工设计提取特征的传统人脸识别方法所达到的最高水平，在几年前已被基于 CNNs 的人脸识别所超越。实际上，基于 CNN 的人脸识别由于其超高的准确度已成为人脸识别的主流。另外，它还可以直接通过更大规模的神经网络和更大的训练数据来获得更高的准确度。然而收集大规模已标记的人脸图像是非常昂贵的，同时过深层的神经网络训练和部署起来都很慢。针对第一个问题[136]提出了很有前景的生成对抗网络（generative adversarial networks, GANs）。最近使用 GANs 处理人脸图像的工作主要有：人脸特征操纵（facial attributes manipulation）[137][138][139][140][141][142][143][144][145][146]、人脸表情编辑（facial expression editing）[147][148] [142]、新身份生成[149]、人脸转正（face frontalisation）[150][151]和人脸老化（face ageing）[152][153]。期望可以通过这些方法可以获得更多的训练图像，而不需要大规模已标记的人脸图像；针对第二个问题有人提出了更高效的移动网络（MobileNets）[154][155]，可以用在移动的计算能力有限的实时人脸识别设备上[156]。

致谢 在此,对薛老师表示感谢.

References:

- [1] M. D. Kelly, "Visual identification of people by computer,," tech. rep., STANFORD UNIV CALIF DEPT OF COMPUTER SCIENCE, 1970.
- [2] T. KANADE, "Picture processing by computer complex and recognition of human faces," PhD Thesis, Kyoto University, 1973.
- [3] K. Delac and M. Grgic, "A survey of biometric recognition methods," in 46th International Symposium Electronics in Marine, vol. 46, pp. 16–18, 2004.
- [4] U. Park, Y. Tong, and A. K. Jain, "Age-invariant face recognition," IEEE transactions on pattern analysis and machine intelligence, vol. 32, no. 5, pp. 947–954, 2010.
- [5] Z. Li, U. Park, and A. K. Jain, "A discriminative model for age invariant face recognition," IEEE transactions on information forensics and security, vol. 6, no. 3, pp. 1028–1037, 2011.
- [6] C. Ding and D. Tao, "A comprehensive survey on pose-invariant face recognition," ACM Transactions on intelligent systems and technology (TIST), vol. 7, no. 3, p. 37, 2016.
- [7] D.-H. Liu, K.-M. Lam, and L.-S. Shen, "Illumination invariant face recognition," Pattern Recognition, vol. 38, no. 10, pp. 1705–1716, 2005.
- [8] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," IEEE transactions on image processing, vol. 19, no. 6, pp. 1635–1650, 2010.
- [9] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1891–1898, 2014.
- [10] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," arXiv preprint arXiv:1411.7923, 2014.
- [11] O. M. Parkhi, A. Vedaldi, A. Zisserman, et al., "Deep face recognition,," in BMVC, vol. 1, p. 6, 2015.
- [12] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, "Ms-celeb-1m: A dataset and benchmark for large-scale face recognition," in European Conference on Computer Vision, pp. 87–102, Springer, 2016.
- [13] A. Nech and I. Kemelmacher-Shlizerman, "Level playing field for million scale face recognition," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3406–3415, IEEE, 2017.
- [14] A. Bansal, A. Nanduri, C. D. Castillo, R. Ranjan, and R. Chellappa, "Umdfaces: An annotated face dataset for training deep networks," in Biometrics (IJB), 2017 IEEE International Joint Conference on, pp. 464–473, IEEE, 2017.
- [15] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "Vggface2: A dataset for recognising faces across pose and age," arXiv preprint arXiv:1710.08092, 2017.

- [16] T. Hassner, S. Harel, E. Paz, and R. Enbar, "Effective face frontalization in unconstrained images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4295–4304, 2015.
- [17] R. Brunelli and T. Poggio, "Face recognition: Features versus templates," *IEEE transactions on pattern analysis and machine intelligence*, vol. 15, no. 10, pp. 1042–1052, 1993.
- [18] J. Shi, A. Samal, and D. Marx, "How effective are landmarks and their geometry for face recognition?," *Computer vision and image understanding*, vol. 102, no. 2, pp. 117–133, 2006.
- [19] I. L. Dryden and K. V. Mardia, *Statistical shape analysis*, vol. 4. Wiley Chichester, 1998.
- [20] F. Daniyal, P. Nair, and A. Cavallaro, "Compact signatures for 3d face recognition under varying expressions," in *Advanced Video and Signal Based Surveillance, 2009. AVSS'09. Sixth IEEE International Conference on*, pp. 302–307, IEEE, 2009.
- [21] S. Gupta, M. K. Markey, and A. C. Bovik, "Anthropometric 3d face recognition," *International journal of computer vision*, vol. 90, no. 3, pp. 331–349, 2010.
- [22] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces," *Josa a*, vol. 4, no. 3, pp. 519–524, 1987.
- [23] M. Kirby and L. Sirovich, "Application of the karhunen-loeve procedure for the characterization of human faces," *IEEE Transactions on Pattern analysis and Machine intelligence*, vol. 12, no. 1, pp. 103–108, 1990.
- [24] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of cognitive neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
- [25] B. Moghaddam, W. Wahid, and A. Pentland, "Beyond eigenfaces: Probabilistic matching for face recognition," in *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, pp. 30–35, IEEE, 1998.
- [26] B. Schölkopf, A. Smola, and K.-R. Müller, "Kernel principal component analysis," in *International Conference on Artificial Neural Networks*, pp. 583–588, Springer, 1997.
- [27] K. I. Kim, K. Jung, and H. J. Kim, "Face recognition using kernel principal component analysis," *IEEE signal processing letters*, vol. 9, no. 2, pp. 40–42, 2002.
- [28] P. Comon, "Independent component analysis, a new concept?," *Signal processing*, vol. 36, no. 3, pp. 287–314, 1994.
- [29] M. S. Bartlett, "Independent component representations for face recognition," in *Face Image Analysis by Unsupervised Learning*, pp. 39–67, Springer, 2001.
- [30] J. Yang, D. Zhang, A. F. Frangi, and J.-y. Yang, "Two-dimensional pca: a new approach to appearance-based face representation and recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 26, no. 1, pp. 131–137, 2004.
- [31] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in *Applications of Computer Vision, 1994., Proceedings of the Second IEEE Workshop on*, pp. 138–142, IEEE, 1994.
- [32] R. A. Fisher, "The statistical utilization of multiple measurements," *Annals of Human Genetics*, vol. 8, no. 4, pp. 376–386, 1938.
- [33] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 19, no. 7, pp. 711–720, 1997.
- [34] K. Etemad and R. Chellappa, "Discriminant analysis for recognition of human face images," *JOSA A*, vol. 14, no. 8, pp. 1724–1733, 1997.
- [35] W. Zhao, A. Krishnaswamy, R. Chellappa, D. L. Swets, and J. Weng, "Discriminant analysis of principal components for face recognition," in *Face Recognition*, pp. 73–85, Springer, 1998.
- [36] W. Zhao, R. Chellappa, and P. J. Phillips, *Subspace linear discriminant analysis for face recognition*. Citeseer, 1999.
- [37] S. Mika, G. Ratsch, J. Weston, B. Schölkopf, and K.-R. Müller, "Fisher discriminant analysis with kernels," in *Neural networks for signal processing IX, 1999. Proceedings of the 1999 IEEE signal processing society workshop.*, pp. 41–48, Ieee, 1999.
- [38] Q. Liu, R. Huang, H. Lu, and S. Ma, "Face recognition using kernelbased fisher discriminant analysis," in *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, pp. 197–201, IEEE, 2002.
- [39] S. Ioffe, "Probabilistic linear discriminant analysis," in *European Conference on Computer Vision*, pp. 531–542, Springer, 2006.
- [40] P. J. Phillips, "Support vector machines applied to face recognition," in *Advances in Neural Information Processing Systems*, pp. 803–809, 1999.
- [41] K. Jonsson, J. Kittler, Y. Li, and J. Matas, "Support vector machines for face authentication," *Image and Vision Computing*, vol. 20, no. 5-6, pp. 369–375, 2002.

- [42] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using laplacianfaces," *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 3, pp. 328–340, 2005.
- [43] D. Cai, X. He, J. Han, and H.-J. Zhang, "Orthogonal laplacianfaces for face recognition," *IEEE transactions on image processing*, vol. 15, no. 11, pp. 3608–3614, 2006.
- [44] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 2, pp. 210–227, 2009.
- [45] Q. Zhang and B. Li, "Discriminative k-svd for dictionary learning in face recognition," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pp. 2691–2698, IEEE, 2010.
- [46] Z. Zhou, A. Wagner, H. Mobahi, J. Wright, and Y. Ma, "Face recognition with contiguous occlusion using markov random fields," in *Computer Vision, 2009 IEEE 12th International Conference on*, pp. 1050–1057, IEEE, 2009.
- [47] H. Jia and A. M. Martinez, "Face recognition with occlusions in the training and testing sets," in *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*, pp. 1–6, IEEE, 2008.
- [48] D. Chen, X. Cao, L. Wang, F. Wen, and J. Sun, "Bayesian face revisited: A joint formulation," in *European Conference on Computer Vision*, pp. 566–579, Springer, 2012.
- [49] 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," tech. rep., Technical Report 07-49, University of Massachusetts, Amherst, 2007.
- [50] A. Pentland, B. Moghaddam, T. Starner, et al., "View-based and modular eigenspaces for face recognition," 1994.
- [51] B. Takacs, "Comparing face images using the modified hausdorff distance," *Pattern Recognition*, vol. 31, no. 12, pp. 1873–1881, 1998.
- [52] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the hausdorff distance," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 15, no. 9, pp. 850–863, 1993.
- [53] Y. Gao and M. K. Leung, "Face recognition using line edge map," *IEEE transactions on pattern analysis and machine intelligence*, vol. 24, no. 6, pp. 764–779, 2002.
- [54] L. Wiskott, N. Krüger, N. Kuiger, and C. Von Der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 19, no. 7, pp. 775–779, 1997.
- [55] M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C. Von Der Malsburg, R. P. Wurtz, and W. Konen, "Distortion invariant object recognition in the dynamic link architecture," *IEEE Transactions on computers*, vol. 42, no. 3, pp. 300–311, 1993.
- [56] T. S. Lee, "Image representation using 2d gabor wavelets," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 10, pp. 959–971, 1996.
- [57] W. T. Freeman and M. Roth, "Orientation histograms for hand gesture recognition," in *International workshop on automatic face and gesture recognition*, vol. 12, pp. 296–301, 1995.
- [58] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1, pp. 886–893, IEEE, 2005.
- [59] A. Albiol, D. Monzo, A. Martin, J. Sastre, and A. Albiol, "Face recognition using hog-ebgm," *Pattern Recognition Letters*, vol. 29, no. 10, pp. 1537–1543, 2008.
- [60] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.
- [61] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.
- [62] D. Huang, C. Shan, M. Ardabilian, Y. Wang, and L. Chen, "Local binary patterns and its application to facial image analysis: a survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 41, no. 6, pp. 765–781, 2011.
- [63] W. Zhang, S. Shan, H. Zhang, W. Gao, and X. Chen, "Multi-resolution histograms of local variation patterns (mhlvp) for robust face recognition," in *International Conference on Audio-and Video-Based Biometric Person Authentication*, pp. 937–944, Springer, 2005.

-
- [64] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, "Local gabor binary pattern histogram sequence (lgbphs): a novel non-statistical model for face representation and recognition," in *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, vol. 1, pp. 786–791, IEEE, 2005.
 - [65] T. Ahonen, J. Matas, C. He, and M. Pietikäinen, "Rotation invariant image description with local binary pattern histogram fourier features," in *Scandinavian Conference on Image Analysis*, pp. 61–70, Springer, 2009.
 - [66] B. Zhang, Y. Gao, S. Zhao, and J. Liu, "Local derivative pattern versus local binary pattern: face recognition with high-order local pattern descriptor," *IEEE transactions on image processing*, vol. 19, no. 2, pp. 533–544, 2010.
 - [67] D. G. Lowe, "Object recognition from local scale-invariant features," in *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, vol. 2, pp. 1150–1157, Ieee, 1999.
 - [68] M. Bicego, A. Lagorio, E. Grosso, and M. Tistarelli, "On the use of sift features for face authentication," in *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on*, pp. 35–35, IEEE, 2006.
 - [69] P.Dreuw,P.Steingrube,H.Hanselmann,H.Ney,andG.Aachen,"Surface: Face recognition under viewpoint consistency constraints.," in *BMVC*, pp. 1–11, 2009.
 - [70] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *European conference on computer vision*, pp. 404–417, Springer, 2006.
 - [71] C. Geng and X. Jiang, "Face recognition using sift features," in *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pp. 3313–3316, IEEE, 2009.
 - [72] Z. Cao, Q. Yin, X. Tang, and J. Sun, "Face recognition with learningbased descriptor," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pp. 2707–2714, IEEE, 2010.
 - [73] S. Lloyd, "Least squares quantization in pcm," *IEEE transactions on information theory*, vol. 28, no. 2, pp. 129–137, 1982.
 - [74] Y. Freund, S. Dasgupta, M. Kabra, and N. Verma, "Learning the structure of manifolds using random projections," in *Advances in Neural Information Processing Systems*, pp. 473–480, 2008.
 - [75] G. Sharma, S. ul Hussain, and F. Jurie, "Local higher-order statistics (lhs) for texture categorization and facial analysis," in *European Conference on Computer Vision*, pp. 1–12, Springer, 2012.
 - [76] Z. Lei, M. Pietikäinen, and S. Z. Li, "Learning discriminant face descriptor," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 2, pp. 289–302, 2014.
 - [77] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Transactions on Image processing*, vol. 11, no. 4, pp. 467–476, 2002.
 - [78] C. Liu and H. Wechsler, "Independent component analysis of gabor features for face recognition," *IEEE transactions on Neural Networks*, vol. 14, no. 4, pp. 919–928, 2003.
 - [79] C. Liu, "Gabor-based kernel pca with fractional power polynomial models for face recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 26, no. 5, pp. 572–581, 2004.
 - [80] C. Liu and H. Wechsler, "Robust coding schemes for indexing and retrieval from large face databases," *IEEE Transactions on image processing*, vol. 9, no. 1, pp. 132–137, 2000.
 - [81] C.-H. Chan, J. Kittler, and K. Messer, "Multi-scale local binary pattern histograms for face recognition," in *International conference on biometrics*, pp. 809–818, Springer, 2007.
 - [82] C.-H. Chan, J. Kittler, and K. Messer, "Multispectral local binary pattern histogram for component-based color face verification," in *Biometrics: Theory, Applications, and Systems, 2007. BTAS 2007. First IEEE International Conference on*, pp. 1–7, IEEE, 2007.
 - [83] D. Zhao, Z. Lin, and X. Tang, "Laplacian pca and its applications," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pp. 1–8, IEEE, 2007.
 - [84] L. Wolf, T. Hassner, and Y. Taigman, "Descriptor based methods in the wild," in *Workshop on faces in 'real-life' images: Detection, alignment, and recognition*, 2008.
 - [85] D. Chen, X. Cao, F. Wen, and J. Sun, "Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pp. 3025–3032, IEEE, 2013.

- [86] C. Lu and X. Tang, "Surpassing human-level face verification performance on lfw with gaussianface.," in AAAI, pp. 3811–3819, 2015.
- [87] R. Urtasun and T. Darrell, "Discriminative gaussian process latent variable model for classification," in Proceedings of the 24th international conference on Machine learning, pp. 927–934, ACM, 2007.
- [88] X. Tan and B. Triggs, "Fusing gabor and lbp feature sets for kernelbased face recognition," in International Workshop on Analysis and Modeling of Faces and Gestures, pp. 235–249, Springer, 2007.
- [89] H. Cevikalp, M. Neamtu, and M. Wilkes, "Discriminative common vector method with kernels," IEEE Transactions on Neural Networks, vol. 17, no. 6, pp. 1550–1565, 2006.
- [90] S. Shan, W. Zhang, Y. Su, X. Chen, and W. Gao, "Ensemble of piecewise fda based on spatial histograms of local (gabor) binary patterns for face recognition," in Pattern Recognition, 2006. ICPR 2006. 18th International Conference on, vol. 3, IEEE, 2006.
- [91] C. H. Chan, M. A. Tahir, J. Kittler, and M. Pietikainen, "Multiscale local phase quantization for robust component-based face recognition using kernel fusion of multiple descriptors," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 5, pp. 1164–1177, 2013.
- [92] E. Rahtu, J. Heikkilä, V. Ojansivu, and T. Ahonen, "Local phase quantization for blur-insensitive image analysis," Image and Vision Computing, vol. 30, no. 8, pp. 501–512, 2012.
- [93] T. K. Ho, "The random subspace method for constructing decision forests," IEEE transactions on pattern analysis and machine intelligence, vol. 20, no. 8, pp. 832–844, 1998.
- [94] L. Breiman, "Bagging predictors," Machine learning, vol. 24, no. 2, pp. 123–140, 1996.
- [95] D. S'aez-Trigueros, H. Hertlein, L. Meng, and M. Hartnett, "Shape and texture combined face recognition for detection of forged id documents," in Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2016 39th International Convention on, pp. 1343–1348, IEEE, 2016.
- [96] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Attribute and simile classifiers for face verification," in Computer Vision, 2009 IEEE 12th International Conference on, pp. 365–372, IEEE, 2009.
- [97] T. Berg and P. N. Belhumeur, "Tom-vs-pete classifiers and identitypreserving alignment for face verification.," in BMVC, vol. 2, p. 7, Citeseer, 2012.
- [98] M. Guillaumin, J. Verbeek, and C. Schmid, "Is that you? metric learning approaches for face identification," in Computer Vision, 2009 IEEE 12th international conference on, pp. 498–505, IEEE, 2009.
- [99] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1701–1708, 2014.
- [100] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," in Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, vol. 1, pp. 539–546, IEEE, 2005.
- [101] H. Fan, Z. Cao, Y. Jiang, Q. Yin, and C. Doudou, "Learning deep face representation," arXiv preprint arXiv:1403.2802, 2014.
- [102] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 815–823, 2015.
- [103] S.-H. Lin, S.-Y. Kung, and L.-J. Lin, "Face recognition/detection by probabilistic decision-based neural network," IEEE transactions on neural networks, vol. 8, no. 1, pp. 114–132, 1997.
- [104] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face recognition: A convolutional neural-network approach," IEEE transactions on neural networks, vol. 8, no. 1, pp. 98–113, 1997.
- [105] T. Kohonen, "The self-organizing map," Neurocomputing, vol. 21, no. 1-3, pp. 1–6, 1998.
- [106] J. Bromley, I. Guyon, Y. LeCun, E. S'ackinger, and R. Shah, "Signature verification using a" siamese" time delay neural network," in Advances in Neural Information Processing Systems, pp. 737–744, 1994.
- [107] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, pp. 1097–1105, 2012.
- [108] K. Gregor and Y. LeCun, "Emergence of complex-like cells in a temporal product network with local receptive fields," arXiv preprint arXiv:1006.0448, 2010.

- [109] G. B. Huang, H. Lee, and E. Learned-Miller, "Learning hierarchical representations for face verification with convolutional deep belief networks," in *Computer Vision and Pattern Recognition (CVPR)*, 2012 IEEE Conference on, pp. 2518–2525, IEEE, 2012.
- [110] E. Zhou, Z. Cao, and Q. Yin, "Naive-deep face recognition: Touching the limit of lfw benchmark or not?," *arXiv preprint arXiv:1501.04690*, 2015.
- [111] A. Bansal, C. Castillo, R. Ranjan, and R. Chellappa, "The do's and don'ts for cnn-based face verification," *arXiv preprint arXiv:1705.07426*, vol. 5, 2017.
- [112] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [113] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, et al., "Going deeper with convolutions," *Cvpr*, 2015.
- [114] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
- [115] R. Ranjan, C. D. Castillo, and R. Chellappa, "L2-constrained softmax loss for discriminative face verification," *arXiv preprint arXiv:1703.09507*, 2017.
- [116] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "Sphereface: Deep hypersphere embedding for face recognition," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 2017.
- [117] Y. Wu, H. Liu, J. Li, and Y. Fu, "Deep face recognition with center invariant loss," in *Proceedings of the on Thematic Workshops of ACM Multimedia 2017*, pp. 408–414, ACM, 2017.
- [118] A. Hasnat, J. Bohné, J. Milgram, S. Gentric, and L. Chen, "Deepvisage: Making face recognition simple yet with powerful generalization skills," *arXiv preprint arXiv:1703.08388*, 2017.
- [119] H. Wang, Y. Wang, Z. Zhou, X. Ji, Z. Li, D. Gong, J. Zhou, and W. Liu, "Cosface: Large margin cosine loss for deep face recognition," *arXiv preprint arXiv:1801.09414*, 2018.
- [120] F. Wang, W. Liu, H. Liu, and J. Cheng, "Additive margin softmax for face verification," *arXiv preprint arXiv:1801.05599*, 2018.
- [121] J. Deng, J. Guo, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," *arXiv preprint arXiv:1801.07698*, 2018.
- [122] Y. Yamada, M. Iwamura, and K. Kise, "Deep pyramidal residual networks with separated stochastic depth," *arXiv preprint arXiv:1612.01230*, 2016.
- [123] X. Wu, R. He, and Z. Sun, "A lightened cnn for deep face representation," in *2015 IEEE Conference on IEEE Computer Vision and Pattern Recognition (CVPR)*, vol. 4, 2015.
- [124] Y. Sun, Y. Chen, X. Wang, and X. Tang, "Deep learning face representation by joint identification-verification," in *Advances in neural information processing systems*, pp. 1988–1996, 2014.
- [125] W.-S. T. WST, "Deeply learned face representations are sparse, selective, and robust," *perception*, vol. 31, pp. 411–438, 2008.
- [126] Y. Sun, D. Liang, X. Wang, and X. Tang, "Deepid3: Face recognition with very deep neural networks," *arXiv preprint arXiv:1502.00873*, 2015.
- [127] J.-C. Chen, V. M. Patel, and R. Chellappa, "Unconstrained face verification using deep cnn features," in *Applications of Computer Vision (WACV)*, 2016 IEEE Winter Conference on, pp. 1–9, IEEE, 2016.
- [128] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *Journal of Machine Learning Research*, vol. 10, no. Feb, pp. 207–244, 2009.
- [129] S. Sankaranarayanan, A. Alavi, and R. Chellappa, "Triplet similarity embedding for face verification," *arXiv preprint arXiv:1602.03418*, 2016.
- [130] S. Sankaranarayanan, A. Alavi, C. D. Castillo, and R. Chellappa, "Triplet probabilistic embedding for face verification and clustering," in *Biometrics Theory, Applications and Systems (BTAS)*, 2016 IEEE 8th International Conference on, pp. 1–8, IEEE, 2016.
- [131] B. Kumar, G. Carneiro, I. Reid, et al., "Learning local image descriptors with deep siamese and triplet convolutional networks by minimising global loss functions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5385–5394, 2016.

- [132] D. S. Trigueros, L. Meng, and M. Hartnett, “Enhancing convolutional neural networks for face recognition with occlusion maps and batch triplet loss,” *Image and Vision Computing*, 2018.
- [133] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, “A discriminative feature learning approach for deep face recognition,” in *European Conference on Computer Vision*, pp. 499–515, Springer, 2016.
- [134] X. Zhang, Z. Fang, Y. Wen, Z. Li, and Y. Qiao, “Range loss for deep face recognition with long-tail,” *arXiv preprint arXiv:1611.08976*, 2016.
- [135] W. Liu, Y. Wen, Z. Yu, and M. Yang, “Large-margin softmax loss for convolutional neural networks,” in *ICML*, pp. 507–516, 2016.
- [136] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” in *Advances in neural information processing systems*, pp. 2672–2680, 2014.
- [137] A. B. L. Larsen, S. K. Sønderby, H. Larochelle, and O. Winther, “Autoencoding beyond pixels using a learned similarity metric,” *arXiv preprint arXiv:1512.09300*, 2015.
- [138] G. Perarnau, J. van de Weijer, B. Raducanu, and J. M. Álvarez, “Invertible conditional gans for image editing,” *arXiv preprint arXiv:1611.06355*, 2016.
- [139] A. Brock, T. Lim, J. M. Ritchie, and N. Weston, “Neural photo editing with introspective adversarial networks,” *arXiv preprint arXiv:1609.07093*, 2016.
- [140] W. Shen and R. Liu, “Learning residual images for face attribute manipulation,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1225–1233, IEEE, 2017.
- [141] Y. Lu, Y.-W. Tai, and C.-K. Tang, “Conditional cyclegan for attribute guided face image generation,” *arXiv preprint arXiv:1705.09966*, 2017.
- [142] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, “Stargan: Unified generative adversarial networks for multi-domain image-to-image translation,” *arXiv preprint arXiv:1711.09020*, 2017.
- [143] W. Yin, Y. Fu, L. Sigal, and X. Xue, “Semi-latent gan: Learning to generate and modify facial images from attributes,” *arXiv preprint arXiv:1704.02166*, 2017.
- [144] Z. Shu, E. Yumer, S. Hadap, K. Sunkavalli, E. Shechtman, and D. Samaras, “Neural face editing with intrinsic image disentangling,” in *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*, pp. 5444–5453, IEEE, 2017.
- [145] G. Lample, N. Zeghidour, N. Usunier, A. Bordes, L. Denoyer, et al., “Fader networks: Manipulating images by sliding attributes,” in *Advances in Neural Information Processing Systems*, pp. 5969–5978, 2017.
- [146] Z. He, W. Zuo, M. Kan, S. Shan, and X. Chen, “Arbitrary facial attribute editing: Only change what you want,” *arXiv preprint arXiv:1711.10678*, 2017.
- [147] Y. Zhou and B. E. Shi, “Photorealistic facial expression synthesis by the conditional difference adversarial autoencoder,” *arXiv preprint arXiv:1708.09126*, 2017.
- [148] H. Ding, K. Sricharan, and R. Chellappa, “Exprgan: Facial expression editing with controllable expression intensity,” *arXiv preprint arXiv:1709.03842*, 2017.
- [149] C. Donahue, A. Balasubramani, J. McAuley, and Z. C. Lipton, “Semantically decomposing the latent spaces of generative adversarial networks,” *arXiv preprint arXiv:1705.07904*, 2017.
- [150] R. Huang, S. Zhang, T. Li, R. He, et al., “Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis,” *arXiv preprint arXiv:1704.04086*, 2017.
- [151] L. Tran, X. Yin, and X. Liu, “Representation learning by rotating your faces,” *arXiv preprint arXiv:1705.11136*, 2017.
- [152] G. Antipov, M. Baccouche, and J.-L. Dugelay, “Face aging with conditional generative adversarial networks,” *arXiv preprint arXiv:1702.01983*, 2017.
- [153] Z. Zhang, Y. Song, and H. Qi, “Age progression/regression by conditional adversarial autoencoder,” in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, 2017.
- [154] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, “Mobilenets: Efficient convolutional neural networks for mobile vision applications,” *arXiv preprint arXiv:1704.04861*, 2017.

- [155] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “Inverted residuals and linear bottlenecks: Mobile networks for classification, detection and segmentation,” arXiv preprint arXiv:1801.04381, 2018.
- [156] S. Chen, Y. Liu, X. Gao, and Z. Han, “Mobilefacenets: Efficient cnns for accurate real-time face verification on mobile devices,” arXiv preprint arXiv:1804.07573, 2018.