



浙江工业大学

本科毕业论文(设计)

外文翻译及原稿

译文题目: MagFace:人脸识别和质量评估的通用表示

原稿出处: CVPR 2021

学 院: 计算机科学与技术学院、软件学院

专 业: 软件工程(中外合作办学)

班 级: 20192019 软件工程(中外合作办学)01

学 号: 201906150117

学生姓名: 沈浪

指导老师: 李小薪

提交日期: 2023 年 02 月

MagFace: 人脸识别和质量评估的通用表示

摘要 人脸识别系统的性能随着人脸的可变性的增加而降低。之前的工作通过在预处理中监测人脸质量或随着人脸特征预测数据的不确定性来缓解这个问题。本文提出了 MagFace, 这是一类学习通用特征嵌入的损失, 其大小可以衡量给定人脸的质量。在新的损失下, 可以证明, 主题越容易被识别, 特征嵌入的大小就增加的越多。此外, MagFace 引入了一种自适应机制, 通过将简单的样本拉到类中心, 同时将困难的样本推到类中心, 来学习结构良好的类内特征分布。此外, MagFace 引入了一种自适应机制, 通过将简单的样本拉到类中心, 同时将困难的样本推离类中心, 来学习结构良好的类内特征分布。这可以防止模型在有噪声的低质量样本上过度拟合, 并提高野外的人脸识别能力。在人脸识别、质量评估和聚类方面进行的大量实验表明, 该算法优于现有技术。该代码可在<https://github.com/IrvingMeng/MagFace>上获取。

关键词: 计算机科学-计算机视觉和模式识别

1、引言

在野外识别人脸是困难的, 主要是由于在无约束环境中获得的人脸图像表现出很大的可变性。这种可变性与图像采集条件 (如照明、背景、模糊度和低分辨率)、面部因素 (如姿势、遮挡和表情) 或部署的人脸识别系统^[1] 的偏差有关。为了应对这些挑战, 在无约束环境下 (如监控视频), 大多数相关的人脸分析系统包括三个阶段: 1) 人脸采集, 从一组原始图像或从视频流中捕获最适合识别的人脸图像; 2) 特征提取, 从每张人脸图像中提取判别表征; 3) 面部应用, 将参考图像与给定的图库或聚类人脸匹配为同一人的组。

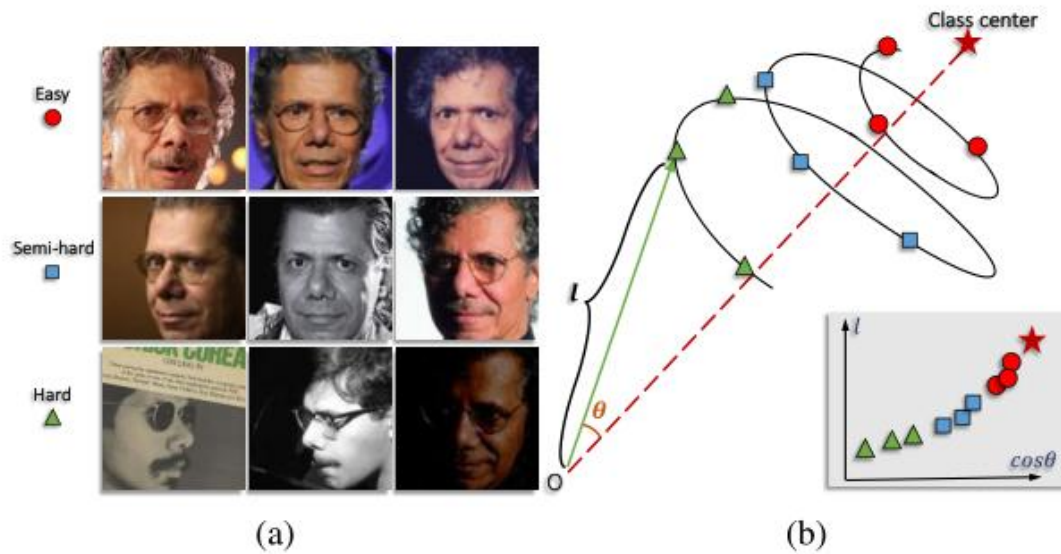


图 1: MagFace 通过将更容易的样本拉到类中心并将其推离原点 o 来学习 (a) 随机人脸 (b) 通用嵌入。正如我们的实验所示并得到数学证明的支持, 归一化前的幅度 l 随着特征到其类中心的余弦距离的增加而增加, 因此揭示了每个人脸的质量 l 越大, 样本被识别的可能性越大。

为了在第一阶段获得最优参考图像, 通常对每个检测到的人脸采用人脸质量评估技术^{[2][3]}。虽然理想的质量分数应该是人脸识别性能的指示, 但早期的大部分工作^{[4][5]} 是基于亮度、失真和姿态角等人类可理解的因素来估计质量的, 这可能不会直接有利于第二阶段的人脸特征学习。或者, 基于学习的方法^{[2][6]} 用人工或人工标记的质量值训练质量评估模型。这些方法容易出错, 因为缺乏对质量的明确定义, 人们可能不知道整个系统的最佳特性。

为了在第二阶段获得较高的端到端应用性能, 过去几年出现了各种度量学习^{[7][8]} 或分类损失^{[9][10][11][12][13][14][15]}。这些作品学习将每个人脸图像表示为嵌入潜在空间的确定性点, 而不考虑人脸固有的方差。然而在现实中, 像图 1a 这样的低质量或大姿态的图像普遍存在, 其面部特征模糊或缺失考虑到这些挑战, 嵌入点的巨大变化是不可避免的, 从而导致错误识别。例如, 先前最先进的^[16] 在 IJB-C 上报告的性能远低于 LFW。最近, 信心感知方法^{[16][15]} 提出将每张人脸图像表示为潜在空间中的高斯分布, 其中分布的平均值估计最有可能的特征值, 而方差显示特征值的不确定性。尽管性能有所提高, 但这些方法试图将人脸特征学习与数据噪声建模分离。因此, 在架构中引入额外的网络块来计算每个图像的不确定性级别。这增加了训练过程的复杂性, 并增加了推理的计算负担。此外, 不确定度测量不能直接用于比较人脸特征的常规度量。

本文提出 MagFace 来学习一种通用的、具有质量意识的人脸表示方法。MagFace 的设计遵循两个原则: 1) 给定同一主体不同质量等级的人脸图像 (如图 1a

所示),它寻求学习类内分布,其中高质量的图像保持在类中心附近,而低质量的图像分布在边界附近。2) 在计算人脸特征的同时,改变现有的推理架构来测量人脸质量的成本最小。为了达到上述目标,我们选择了幅度,即特征向量方向的独立性质,作为质量评估的指标。MagFace 的核心目标不仅是扩大类间距离,而且要保持如图 1b 所示的类内锥形结构,即模糊样本被推离类中心,并被拉到原点。这是通过在训练过程中自适应地降低模糊样本的权重,并在 MagFace 损失中奖励学习到的特征向量来实现的。综上所述,MagFace 在两个方面改进了之前的工作:

1. 在人脸识别问题中,MagFace 首次探索了与特征向量方向和幅度相关的两个性质的完整集合,而以往的工作往往通过对特征进行归一化而忽略了幅度的重要性。通过广泛的实验研究和坚实的数学证明,我们表明,大小可以揭示人脸的质量,并且可以与识别的特征捆绑在一起,而不涉及任何质量标签。

2. MagFace 在角度方向上明确地在结构上分布特征(如图 1b 所示)。通过根据样本的识别硬度动态分配角边缘,MagFace 可以防止模型在有噪声和低质量的样本上过拟合,并学习更适合识别和聚类目的的结构良好的分布。

2、相关的工作

2.1 人脸识别

近年来,深度卷积人脸识别技术取得了突破性进展。一些成功的系统,如 DeepFace^[17],DeepID^[18],FaceNet^[19] 在人脸识别和验证方面表现出了令人印象深刻的性能。除了大规模的训练数据和深度网络架构外,CNN 的主要进步来自于训练损耗的进化。早期的工作大多依赖于基于度量学习的损耗,包括对比损耗^[20]、三重态损耗^[19]、n 对损耗^[8]、角损耗^[21] 等。在大规模数据集上,基于嵌入的训练方法由于人脸三组数的组合爆炸而效率低下。因此,深度人脸识别的研究主要集中在设计更高效和有效的基于分类的损失。Wen 等^[22] 开发了一个中心损失来学习每个身份的中心,以增强类内紧凑性。L2-softmax^[10] 和 NormFace^[23] 研究了归一化操作的必要性,在特征和权值上都应用了 L2 归一化约束。从那时起,一些基于角边缘的损耗,如 SphereFace^[11],AM-softmax^[24],SV-AM-Softmax^[25],CosFace^[13],ArcFace^[26],逐渐将各种基准测试的性能提高到新的水平。最近,AdaptiveFace^[27],AdaCos^[28] 和 FairLoss^[29] 引入了自适应边际策略,以自动调优超参数,并在训练期间生成更有效的监督。与我们的方法相比,所有这些工作都倾向于通过归一化特征向量来抑制量级对损失的影响

2.2 面部质量评估

人脸图像质量是实现高性能人脸识别系统 [4] 的重要因素。传统的方法,如 ISO/IEC 19794-5 标准^[4],ICAO 9303 标准,Brisque^[30],Niqe^[31] 和 Pique^[32],从基于图

像的方面 (例如, 失真, 照明和遮挡) 或基于主题的测量 (例如, 附件) 描述质量。基于学习的方法, 如 FaceQNet^[33] 和 Best-Rowden^[2], 通过在人类评估和基于相似的标签上训练的网络回归质量。然而, 这些质量标签容易出错, 因为人类可能不知道识别系统的最佳特征, 因此不能考虑所有适当的因素。近年来, 提出了几种基于不确定性的方法, 利用特征的不确定性来表达人脸质量。SER-FIQ^[34] 将图像多次转发到有 dropout 的网络中, 通过提取特征的变化来测量人脸质量。基于信心感知的人脸识别方法提出将每张人脸图像在潜在空间中表示为高斯分布, 并学习特征值中的不确定性。虽然这些方法像我们一样以无监督的方式工作, 但它们需要额外的计算成本或网络块, 这使得它们在传统人脸系统中的使用变得复杂。

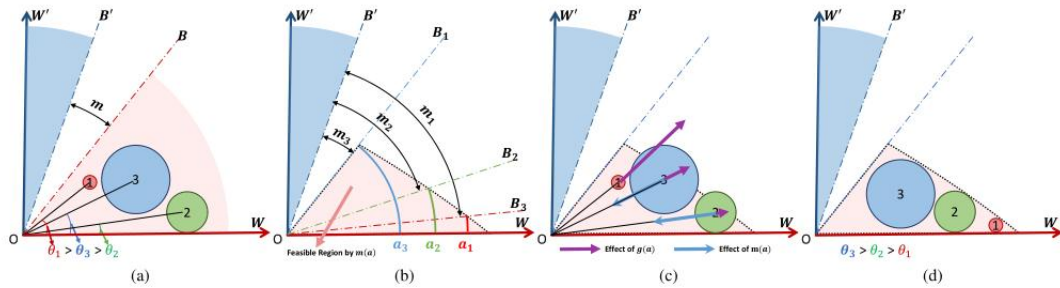


图 2: ArcFace 和 MagFace 优化的特征空间 (未归一化) 的几何解释。(a) ArcFace 优化的两类分布, w 和 w' 是类中心, 它们的决策边界 B 和 B' 由加性裕度 m 分隔。圈 1、2、3 代表 w 类的三种质量递减样本。(b) MagFace 引入 $m(ai)$, 根据特征大小动态调整边界, 最终得到一个新的可行区域。(c) $g(ai)$ 和 $m(ai)$ 的影响。(d) MagFace 的最终特征分布。彩色观看效果最好。

2.3 人脸聚类

人脸聚类利用未标记的数据将它们聚成伪类。传统的聚类方法通常以无监督的方式工作, 例如 K-means^[35]。DBSCAN^[36] 和分级集群。近年来提出了几种基于图卷积网络 (GCN) 的监督聚类方法。例如, LGCN^[37] 执行推理并推断子图中对之间链接的可能性。Yang et al.^[38] 设计了两个图卷积网络, 命名为 GCN-V 和 GCN-E, 分别用于估计顶点的置信度和边的连通性。我们的目标不是开发聚类方法, 而是改进用于聚类的特征分布结构。

3、方法论

在本节中, 我们首先回顾 ArcFace^[14] 的定义, 它是人脸识别中最常用的损失之一。在对 ArcFace 进行分析的基础上, 推导出目标并证明了 MagFace 的关键特性。最后, 从特征量级的角度对 softmax、ArcFace 和 MagFace 进行了比较。

3.1 Arcface 重申

训练损失在人脸表征学习中起着重要作用。在各种选择中 (参见^[39] 最近的调查), ArcFace^[14] 可能是在学术界和工业界应用中最广泛采用的, 因为它易于实现, 并且在许多基准测试中具有最先进的性能。假设我们有 N 个人脸样本 $\{f_i, y_i\}_{i=1}^N$ 的 N 个身份的训练批次, 其中 $f_i \in \mathbb{R}^d$ 表示从神经网络的最后一个完全连接层计算的 d 维嵌入, $y_i = 1, \dots, n$ 是其相关的类标签。ArcFace 和其他变体通过优化超球面流形上的特征嵌入来改善传统的 softmax 损失, 其中学习的人脸表示更具鉴别性。将 f_i 与第 j 类中心 $w_j \in \mathbb{R}^d$ as $w_j^T f_i = \|w_j\| \|f_i\| \cos \theta_j$ ArcFace 的目标可以表述为

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

其中 $m > 0$ 表示附加角边缘, s 为缩放参数。

尽管 ArcFace 在加强类内紧凑性和类间差异方面表现出色, 但它使用的角边缘惩罚 m 是质量不确定的, 并且在无约束场景下, 类内分布的结果结构可能是任意的。例如, 让我们考虑图 2a 所示的场景, 其中我们有由圆大小表示的三个质量级别的同一类人脸图像: 半径越大, 特征表示越不确定, 人脸识别就越困难。因为 ArcFace 使用统一的边距 m , 一个类中的每个图像共享相同的决策边界, 即 $B: \cos(\theta + m) = \cos(\theta')$ 相对于相邻类。三种样本都可以停留在可行区域 (图 2a 的阴影区域) 的任意位置, 不会受到角边缘的任何惩罚。这导致了不稳定的类内分布, 例如, 高质量的人脸 (类型 1) 停留在边界 B , 而低质量的人脸 (类型 2 和 3) 更接近中心 w 。这种不稳定性会损害野外识别的性能以及其他面部应用, 如人脸聚类。此外, 硬样本和噪声样本很难停留在可行区域内, 模型可能会对它们过拟合, 从而导致样本权重过大。

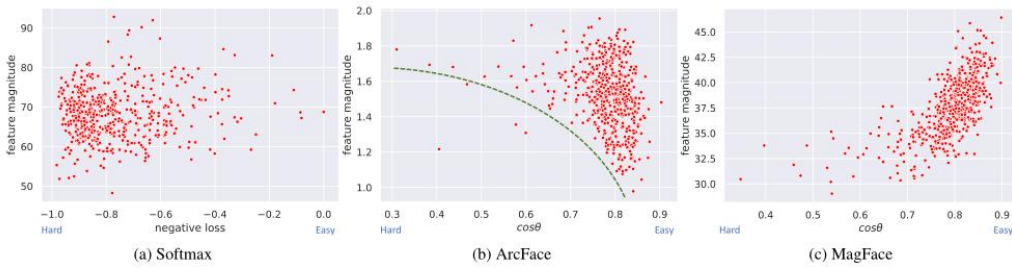


图 3: 特征大小的可视化和识别难度。在 MS1M-V2 上训练模型, 使用上一次迭代的 512 个样本进行可视化。对于 Softmax, 我们使用负损失来显示硬度, 而对于 ArcFace 和 MagFace, 我们使用 θ 的余弦值 (特征与其类中心之间的角度)。

3.2 Magface

基于上述分析,之前基于余弦相似度的人脸识别损失在固定的边界 m 之外缺乏更细粒度的约束,这导致类内结构不稳定,特别是在无约束的情况下(如图 2a),每个受试者的人脸变异性很大。为了解决上述问题,本节提出了 MagFace,这是一种将质量度量编码到人脸表示中的新框架。与之前的工作不同,我们要求额外的不确定性项,我们通过优化幅度 $a_i = f_i$ 来追求极简设计,而不对每个特征 f_i 进行标准化。我们的设计有两个主要优势:1) 我们可以继续使用基于余弦的度量,这已经被大多数现有的推理系统广泛采用;2) 通过同时强化其方向和大小,习得的人脸表征对野外人脸的可变性更具鲁棒性。据我们所知,这是第一个将特征大小统一为人脸识别质量指标的工作。

在定义损失之前,让我们首先介绍两个与 a_i 相关的辅助函数,大小感知角边缘 $m(a_i)$ 和正则器 $g(a_i)$ 。 $m(a_i)$ 的设计遵循一个自然的直觉:对于高质量的样本 x_i ,它们应该集中在簇中心 w 周围的一个小区域,具有较高的确定性。通过假设大小和质量之间存在正相关关系,因此,如果 x_i 的大小 a_i 很大,我们就会在 $m(a_i)$ 方面对 x_i 进行更多的惩罚。为了更好地理解,图 2b 可视化了对应不同幅度值的边缘 $m(a_i)$ 。与 ArcFace(图 2a) 相比,由 $m(a_i)$ 定义的可行区域相对于类中心 w 的特征大小有一个缩小的边界。因此,该边界将低质量样本(图 2c 中的圈 2 和圈 3)拉到被惩罚风险较低的原点。然而,对于图 2c 中的圆 1 这样的高质量样本,仅由 $m(a_i)$ 形成的结构是不稳定的,因为它们在可行区域内有较大的自由度。因此,我们引入正则化 $g(a_i)$ 来奖励具有较大幅度的样本。将 $g(a_i)$ 设计为关于 a_i 的单调递减凸函数,将每个样本推向可行域边界,将高质量样本(圆 1)拖向类中心 w ,如图 2d 所示。简而言之,MagFace 扩展了 ArcFace (Eq. 1),使用幅度感知边缘和正则化器,通过优化来提高类间样本的多样性和类内样本的相似性:

$$L_{Mag} = \frac{1}{N} \sum_{i=1}^N L_i$$

$$L_i = -\log \frac{e^{s \cos(\theta_{y_i} + m(a_i))}}{e^{s \cos(\theta_{y_i} + m(a_i))} + \sum_{j \neq y_i} e^{s \cos \theta_j}} + \lambda_g g(a_i)$$

超参数 λ_g 被用来权衡分类损失和正则化损失。

MagFace 的设计不仅遵循直观动机,而且有理论保证。假设模 a_i 有界于 $[l_a, u_a]$, 其中 $m(a_i)$ 为严格增凸函数, $g(a_i)$ 为严格减凸函数, λ_g 足够大,我们可以证明(具体要求和证明见补充)在优化 L_i / a_i 时, MagFace 损失的以下两个性质始终成立:

收敛的性质。对于 a_i 属于 $[l_a, u_a]$, L_i 是一个严格凸函数,它有唯一的最优解 a_i^* 。

单调性。最优 ai^* 是随着到其类中心的余弦距离减小和到其他类的余弦距离增大而单调增加的。

对于 ai , 算法的收敛性保证了算法的唯一最优解和快速收敛。特征量的单调性表明了识别的难度, 因此可以作为人脸质量的度量标准。

3.3 特征幅度分析

为了更好地理解 MagFace 损失的影响, 我们在广泛使用的 MS1M-V2^[14] 数据集上进行了实验, 并对收敛时的训练样例研究了特征幅度与其与类中心相似度之间的关系, 如图 3 所示。

Softmax。经典的基于 softmax 的损失是关于深度人脸识别的先驱工作的目标。从图 3a 中观察到, 在没有明确的量级限制的情况下, 每个样本的负损失值几乎与其量级无关。如所指出的, softmax 倾向于创建径向特征分布, 因为 softmax 损失作为 max 算子的软版本, 特征大小的缩放不影响其类的赋值。为了消除这种影响, 建议使用标准化特征将有益于该任务。

ArcFace。当 $m(ai) = m, g(ai) = 0$ 时, ArcFace 可视为 MagFace 的特例。如图 3b 所示, $\cos(\theta)$ 与类中心相似度较大的高质量样本在幅度上变化较大。这一证据呼应了我们在 ArcFace 中由固定角边缘定义的不稳定结构上的动机, 以获得简单的样本。另一方面, 对于难以识别的低质量样本 ($\cos(\theta)$ 较小), 固定的角裕度决定了幅值需要足够大才能拟合在可行区域内 (图 2a)。因此, 图 3b 中虚线所示的特征幅度 w.r.t. 的面部质量有一个递减的下限。

MagFace。与 ArcFace 相比, 我们的 MagFace 根据其大小优化了自适应边缘和正则化特征。在此损失下, 从图 3c 可以清楚地观察到, 特征幅值与其与类中心的余弦相似度之间存在很强的相关性。右上角的例子是质量最好的例子。当量值变小时, 示例就会更偏离类中心。这种分布有力地支持了这样一个事实, 即 MagFace 学习到的特征幅度是衡量面部质量的一个很好的指标。

4、实验

在本节中, 我们将研究提出的 MagFace 在三个重要的人脸任务上: 人脸识别、质量评估和人脸聚类。补充部分的 C 节介绍了边缘分布与识别性能之间关系的烧蚀研究。

4.1 人脸识别

数据集。原始的 MS-Celeb-1M 数据集^[12] 包含大约 1000 万张 100,000 个身份的图像。然而, 它包含了大量的噪声人脸图像。相反, 我们使用 MS1M-V2^[14] (5.8M 张图像, 85k 个身份) 作为我们的训练数据集。我们采用 LFW^[40]、CFPFP^[41]、AgeDB-30^[42]、CALFW^[43]、CPLFW^[44]、IJBB^[45]、IJB-C^[46] 作为评测基准。根据 ArcFace 中的设置, 所有图像都对齐到 112×112 。

Method	LFW	CFP-FP	AgeDB-30	CALFW	CPLFW
Softmax	99.70	98.20	97.72	95.65	92.02
SV-AM-Softmax [42]	99.50	95.10	95.68	94.38	89.48
SphereFace [20]	99.67	96.84	97.05	95.58	91.27
CosFace [40]	99.78	98.26	98.17	96.18	92.18
ArcFace [9]	99.81	98.40	98.05	95.96	92.72
MagFace	99.83	98.46	98.17	96.15	92.87

Table 1: Verification accuracy (%) on easy benchmarks.

基线。我们重新实现了最先进的基线，包括 Softmax, SV-AM-Softmax^[47], SphereFace^[48], CosFace^[13], ArcFace^[14]。骨干设备为 ResNet100。我们为每个模型使用推荐的超参数，例如，ArcFace 的 $s = 64$, $m = 0.5$ 。

训练。我们用随机梯度下降法在 81080tis 上训练模型。学习率从 0.1 开始初始化，在 10、18、22 个 epoch 除以 10，在第 25 个 epoch 停止训练。权重衰减设置为 $5e-4$ ，动量为 0.9。我们只通过随机水平翻转来增加训练样本。对于 MagFace，我们将大小的上界和下界固定为 $l_a = 10$, $u_a = 110$ 。取 $M(a_i)$ 为线性函数，取 $g(a_i)$ 为双曲线。 $m(a_i)$ 、 $g(a_i)$ 和 λg 的详细定义请参见附录 B2 节。最后，我们的平均裕度以及其他超参数都与 ArcFace 一致。

LFW、CFP-FP、AgeDB-30、CALFW 和 CPLFW 的结果。我们直接使用 ArcFace 采用的对齐图像和协议，并在表 1 中展示我们的结果。我们注意到性能几乎饱和。ArcFace 在 LFW、CFP-FP 和 CPLFW 上分别提高了 0.03%、0.14%、0.54%，在 AgeDB-30 和 CALFW 上分别下降了 0.12%、0.22%。MagFace 获得了整体最好的结果，在五个基准测试中分别超过 ArcFace 0.02%、0.06%、0.12%、0.19% 和 0.15%。

IJB-B/IJB-C 检测结果。IJB-B 数据集包含 1845 个主题，21.8K 静态图像和来自 7011 个视频的 55K 帧。作为 IJB-B 的扩展，IJB-C 数据集涵盖约 3,500 个身份，共 31,334 张图像和 117,542 个无约束视频帧。在 1:1 验证中，IJB-B 的阳性/阴性匹配数为 10k/8M，IJB-C 为 19k/15M。我们报告 $FAR=1e-6$ 、 $1e-5$ 和 $1e-4$ 的 tar，如表 2 所示。

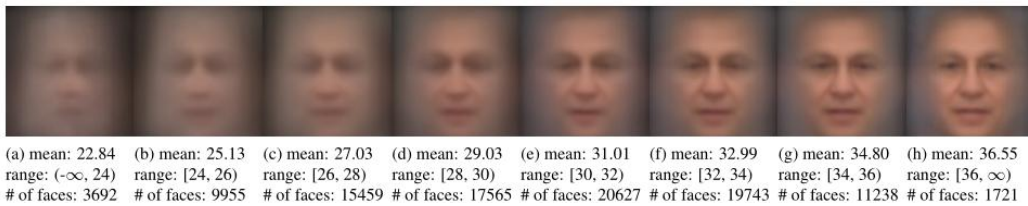


图 4: 从 IJB-C 数据集采样的 100k 张图像的平均人脸的可视化。每个平均脸对应于一组基于 MagFace 学习的特征的大小级别的脸。

Method	IJB-B (TAR@FAR)			IJB-C (TAR@FAR)		
	1e-6	1e-5	1e-4	1e-6	1e-5	1e-4
VGGFace2* [6]	-	67.10	80.00	-	74.70	84.00
CenterFace* [44]	-	-	-	-	78.10	85.30
CircleLoss* [32]	-	-	-	-	89.60	93.95
ArcFace* [9]	-	-	94.20	-	-	95.60
Softmax	46.73	75.17	90.06	64.07	83.68	92.40
SV-AM-Softmax [42]	29.81	69.25	84.79	63.45	80.30	88.34
SphereFace [20]	39.40	73.58	89.19	68.86	83.33	91.77
CosFace [40]	40.41	89.25	94.01	87.96	92.68	95.56
ArcFace [9]	38.68	88.50	94.09	85.65	92.69	95.74
MagFace	40.91	89.88	94.33	89.26	93.67	95.81
MagFace+	42.32	90.36	94.51	90.24	94.08	95.97

图 5: 在困难基准上的 V 验证精度 (%).“*”为引用原论文的结果。

我们实现的 ArcFace 与原始论文相当,例如,我们在 FAR=1e-4 处的 tar 与作者在 IJB-B 和 IJB-C 上的差异分别为 -0.11% 和 +0.14%。与基线相比,我们的 MagFace 在除 IJB-B 上的 FAR=1e-6 外的所有 FAR 标准中仍然是顶级的,因为 TAR 在 FP 数量很小时对噪声非常敏感。与 CosFace 相比, MagFace 在 TAR@FAR=1e-6、1e-5、1e-4 上的增益分别为 0.50%、0.63%、0.32%,在 IJB-C 上的增益分别为 1.30%、0.99%、0.25%。IJB-B 和 IJB-C 分别比 ArcFace 提高了 2.23%、1.38%、0.24% 和 3.61%、0.98%、0.07%。这一结果证明了 MagFace 在更具挑战性的基准测试中的优势。值得一提的是,当一个身份存在多张图像时,可以通过按幅度加权的特征聚合来进一步提高平均嵌入。例如,在 FAR=1e-6 时, MagFace+ 优于 MagFace 1.41%/0.98%, FAR=1e-5 时优于 MagFace 0.48%/0.41%, FAR=1e-4 时优于 MagFace 0.18%/0.16%。

4.2 面部质量评估

在这一部分中,我们研究了表 2 中提到的预训练 MagFace 模型的定性和定量性能,以进行质量评估。

平均面的可视化。我们首先从 IJB-C 数据库中抽取 10 万张图像,并根据特征幅度将其分为 8 组。我们在图 4 中可视化每组的平均面孔。可以看出,当幅度增加时,相应的平均面揭示了更多的细节。这是因为高质量的面孔更倾向于正面和独特。这意味着 MagFace 特征的大小是一个很好的质量指标。

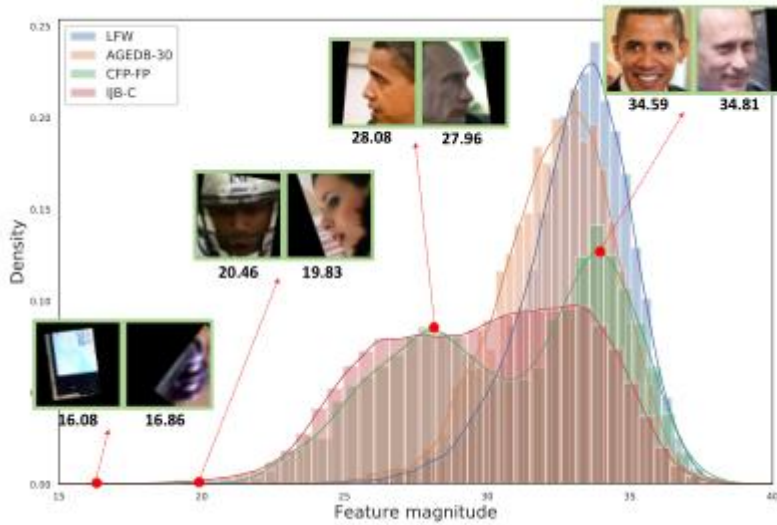


图 6: 不同数据集上的幅度分布。

数据集的样本分布。图 5 绘制了不同基准相对于 MagFace 幅度的样本直方图。我们观察到 LFW 是噪声最小的一个, 其中大多数样本都是大的。由于年龄变化较大, 与 LFW 相比, AGEDB-30 的分布略有左移。对于 CFP-FP, 在 28 和 34 左右有两个震级峰, 分别对应于正面和侧面。鉴于面部质量的巨大差异, 我们可以得出结论, IJB-C 比其他基准更具挑战性。对于 a 相似于 15 星等的图像 (更多的例子可以在补充中找到), 没有人脸或非常嘈杂的人脸可以观察。当特征幅值从 20 增加到 40 时, 面部有明显的变化趋势, 从侧面, 模糊和遮挡, 到更正面和更有特色。总的来说, 这个数字使我们相信 MagFace 是一个有效的工具来排名人脸图像根据他们的质量。

基线。我们选择了三种类型的六个基线进行定量质量评价。Brisque^[30], Nique^[31] 和 Pique^[32] 是基于图像的质量指标。FaceQNet^[33] 和 SER-FIQ^[34] 是基于人脸的。对于 FaceQNet, 我们采用了作者发布的模型。对于 SER-FIQ, 我们使用了论文中产生最佳性能的“同模型”版本。按照作者的设置, 我们设置 $m=100$ 将每个图像转发 100 次, 其中退出在推理中处于活动状态。作为相关工作, 我们重新实现了最近的 DUL^[15] 方法, 该方法可以估计人脸特征的不确定性。

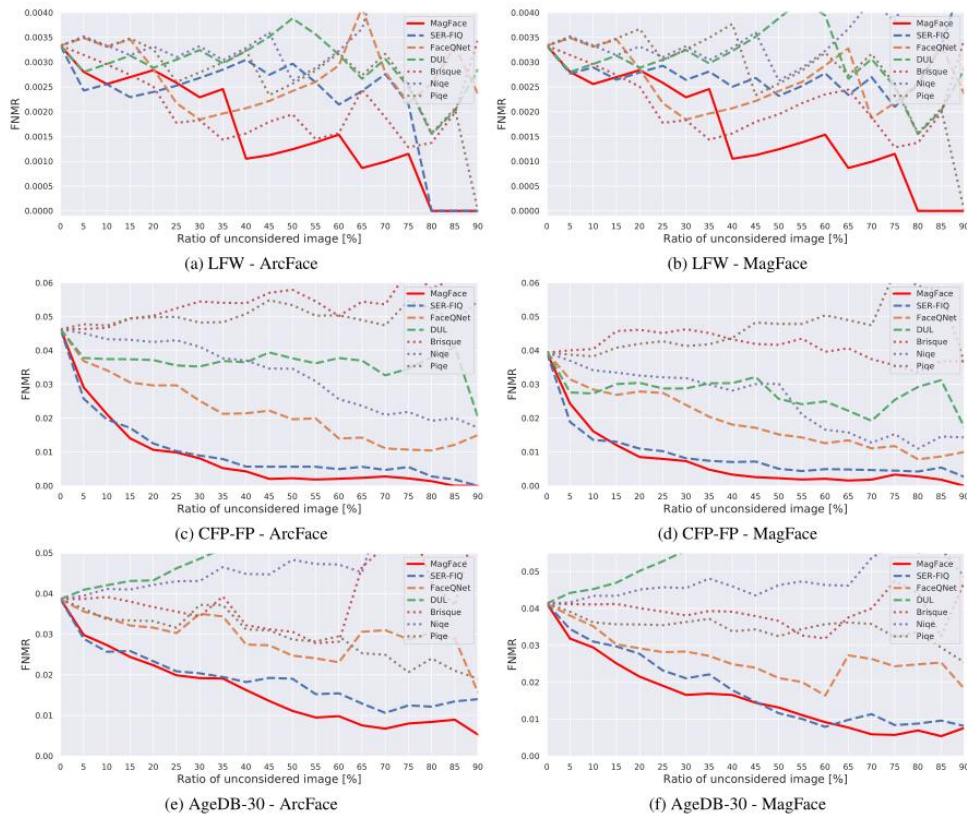


图 7: 使用两种评估模型 (ArcFace 和 MagFace) 预测的人脸质量值的人脸验证性能。该曲线显示了在假不匹配率 (FNMR) 方面拒绝低质量人脸图像的有效性。彩色观看效果最好。

评价指标。根据之前的工作，我们通过误差-拒绝曲线评估 LFW/CFPFP/ AgeDB 的质量评估，其中不考虑预测质量最低的图像，并在剩余图像上计算错误率。当验证误差不断减小，而未考虑图像的比例不断增加时，error - versus-reject 曲线表示良好的质量估计。为了计算用于验证的特征，我们采用 ArcFace* 以及表 2 中的 MagFace 模型。

人脸验证结果。图 6 显示了不同质量方法在假匹配率 (FMR) 阈值为 0.001 时报告的假不匹配率 (FNMR) 的误差-拒收曲线。总的来说，我们有两个高水平的观察结果。1) CFP-FP 和 AgeDB-30 上的曲线比 LFW 上的曲线平滑得多。这是因为 CFP-FP 和 AgeDB-30 由姿势和年龄变化更大的面孔组成。有效地删除低质量的人脸可以在这两个基准测试中更有利于验证性能。2) 无论从 ArcFace(左列) 还是 MagFace(右列) 计算特征，MagFace 幅值对应的曲线在不同基准上始终是最底的。这表明 MagFace 量级的性能在数据集和面部特征上都很好地泛化。然后我们分析了每种方法的质量性能。1) 基于图像的质量指标 (Brisque^[30], Nique^[31], Pique^[32]) 在大多数情况下会导致相对较高的误差，仅凭图像质量不适合进行广义的人脸质量估计。面部的因素 (如姿势、遮挡和表情) 和模型偏差没有被这些算法

覆盖,但可能在面部质量评估中发挥重要作用。2) 基于人脸的方法 (FaceQNet^[33] 和 SER-FIQ^[34]) 在大多数情况下优于其他基线。特别是, SER-FIQ 在验证错误率方面比 FaceQNet 更有效。这是因为 SER-FIQ 构建在已部署的识别模型之上, 因此其预测更适合验证任务。然而, SEQFIQ 的计算代价是二次的 w.r.t.。相反, 计算幅度的开销可以忽略, 这使得所提出的 MagFace 在许多实时场景中更加实用。此外, MagFace 的训练不需要对人脸质量进行明确的标注, 不仅耗时, 而且容易出错。3) 最后, 不确定度方法 (DUL) 在 CFP-FP 上表现较好, 但在 AgeDB-30 上, 随着未考虑图像比例的增加, 验证误差较大。这可能说明 DUL 中数据方差的高斯假设过于简化, 导致模型不能很好地泛化到不同类型的质量因子。

4.3 人脸聚类

在本节中, 我们进行了关于人脸聚类的实验, 以进一步研究 MagFace 学习的特征表示的结构。

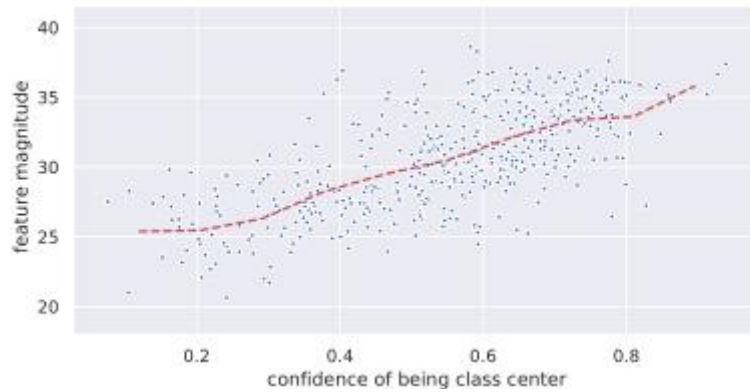


图 8: 来自 IJB-B-1845 的 500 个样本的 MagFace 星等的可视化 w.r.t. 它们作为类中心的置信度。

基线。我们通过将 MagFace 和 ArcFace 的特征与各种聚类方法相结合, 比较了它们的性能。为了公平比较, 我们在训练期间将两个模型的超参数约束为一致 (例如, $s=64$, 平均裕度 0.5)。评价采用四种聚类方法: K-means^[35]、AHC^[49]、DBSCAN^[36] 和 L-GCN^[37]。对于非确定性算法 (K-means 和 AHC), 我们报告了 10 次运行的平均结果。对于 L-GCN, 我们在 CASIAWebFace^[50] (来自 10k 个人的 0.5M 图像) 上训练模型, 并遵循论文^[37] 中的推荐设置。

基准。我们采用 IJB-B 数据集作为基准, 因为它包含 7 个子任务的聚类协议, 这些子任务的地面真相身份数量不同。之后, 我们评估三个最大的子任务, 其中身份数分别为 512、1024 和 1845, 样本数分别为 18,171、36,575 和 68,195。标准化互信息 (NMI) 和 BCubed F 测度^[51] 被用作评估度量。

Method	Net	IJB-B-512		IJB-B-1024		IJB-B-1845	
		F	NMI	F	NMI	F	NMI
K-means [21]	ArcFace	66.70	88.83	66.82	89.48	66.93	89.88
	MagFace	66.75	88.86	67.33	89.62	67.06	89.96
AHC [17]	ArcFace	69.72	89.61	70.47	90.54	70.66	90.90
	MagFace	70.24	89.99	70.68	90.67	70.98	91.06
DBSCAN [11]	ArcFace	72.72	90.42	72.50	91.15	73.89	91.96
	MagFace	73.13	90.61	72.68	91.30	74.26	92.13
L-GCN [43]	ArcFace	84.92	93.72	83.50	93.78	80.35	92.30
	MagFace	85.27	93.83	83.79	94.10	81.58	92.79

图 9: 聚类基准的 f-得分 (%) 和 NMI(%).

结果。表 3 总结了聚类结果。我们可以观察到,从 Kmeans 到 L-GCN 的更强的聚类方法可以提高整体的聚类性能。对于集群和协议的任何组合, MagFace 在 F-score 和 NMI 指标方面总是比 ArcFace 获得更好的性能。这种一致性的优势说明 MagFace 特征更适合于聚类。注意,我们为聚类保留了相同的超参数。使用 MagFace 的改进必须来自于它更好的类内特征分布,在类中心周围的高质量样本更有可能被分离到不同的类中。

我们进一步探讨了特征量与类中心置信度之间的关系。根据^[38]中提到的思想,每个样本作为类中心的置信度是根据人脸特征定义的邻居结构来估计的。局部连接密集且纯的样本具有较高的置信度,而连接稀疏或位于多个聚类边界的样本置信度较低。从图 7 不难看出,在 IJB-B-1845 基准上, MagFace 星等与类中心置信度呈正相关。这一结果反映了 MagFace 特征表现出预期的类内结构,即高质量样本分布在类中心附近,而低质量样本则远离类中心。

5、总结

在本文中,我们提出了 MagFace 来学习统一的特征来进行人脸识别和质量评估。通过将模糊样本从类中心推离, MagFace 改善了以前基于边缘的人脸识别工作的类内特征分布。充分的理论和实验结果表明, MagFace 可以同时获取输入人脸图像的质量。作为一个通用框架, MagFace 可以潜在地扩展到其他分类任务,如细粒度物体识别,人员再识别。此外,所提出的探索特征大小的原则为估计其他对象的质量铺平了道路,例如, reid 中的人的身体或活动分类中的动作片段。

参考文献

- [1] TERHRST P, KOLF J N, DAMER N, et al. SER-FIQ: Unsupervised Estimation of Face Image Quality Based on Stochastic Embedding Robustness[C] // Computer Vision and Pattern Recognition. 2020.
- [2] BEST-ROWDEN L, JAIN A K. Learning Face Image Quality From Human Assessments[J]. IEEE Transactions on Information Forensics and Security, 2018, 13 : 3064 – 3077.
- [3] SCHLETT T, RATHGEB C, HENNIGER O, et al. Face Image Quality Assessment: A Literature Survey[J], 2020.
- [4] JTC/SC 1. Information technology-biometrics-biometric data interchange formats-Part5 : Face image data[J], 2008.
- [5] STRUCTURE-LDS L D. MACHINE READABLE TRAVEL DOCUMENTS[J], .
- [6] HERNANDEZ-ORTEGA J, GALBALLY J, FIERREZ J, et al. FaceQnet: Quality Assessment for Face Recognition based on Deep Learning[J], 2019.
- [7] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: A unified embedding for face recognition and clustering[J]. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015 : 815 – 823.
- [8] SOHN K. Improved deep metric learning with multi-class N-pair loss objective[C] // Neural Information Processing Systems. 2016.
- [9] YUAN Y, YANG K, ZHANG C. Feature Incay for Representation Regularization[J], 2017.
- [10] RANJAN R, CASTILLO C D, CHELLAPPA R. L2-constrained Softmax Loss for Discriminative Face Verification[J], 2017.
- [11] LIU W, WEN Y, YU Z, et al. SphereFace: Deep Hypersphere Embedding for Face Recognition[C/OL] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017 : 6738 – 6746.
<http://dx.doi.org/10.1109/CVPR.2017.713>.
- [12] GUO Y, ZHANG L, HU Y, et al. MS-Celeb-1M: A Dataset and Benchmark for Large-Scale Face Recognition[C] // European Conference on Computer Vision. 2016.
- [13] WANG H, WANG Y, ZHOU Z, et al. CosFace: Large Margin Cosine Loss for Deep Face Recognition[J]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018 : 5265 – 5274.

- [14] DENG J, GUO J, ZAFEIRIOU S. ArcFace: Additive Angular Margin Loss for Deep Face Recognition[J]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018 : 4685–4694.
- [15] CHANG J, LAN Z, CHENG C, et al. Data Uncertainty Learning in Face Recognition[J]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020 : 5709–5718.
- [16] SHI Y, JAIN A K. Probabilistic Face Embeddings[J], 2019.
- [17] TAIGMAN Y, YANG M, RANZATO M, et al. DeepFace: Closing the Gap to Human-Level Performance in Face Verification[C] // IEEE Conference on Computer Vision Pattern Recognition. 2014.
- [18] SUN Y, LIANG D, WANG X, et al. DeepID3: Face Recognition with Very Deep Neural Networks[EB]. 2015.
- [19] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: A Unified Embedding for Face Recognition and Clustering[J]. IEEE, 2015.
- [20] CHOPRA S, HADSELL R, LECUN Y. Learning a similarity metric discriminatively, with application to face verification[C] // 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). 2005.
- [21] JIAN W, FENG Z, WEN S, et al. Deep Metric Learning with Angular Loss[EB]. IEEE, 2017.
- [22] WEN Y, ZHANG K, LI Z, et al. A Discriminative Feature Learning Approach for Deep Face Recognition[J], 2016.
- [23] WANG F, XIANG X, CHENG J, et al. NormFace: L_2 Hypersphere Embedding for Face Verification[J]. arXiv, 2017.
- [24] WANG F, CHENG J, LIU W, et al. Additive Margin Softmax for Face Verification[J/OL]. IEEE Signal Processing Letters, 2018, 25(7) : 926–930.
<http://dx.doi.org/10.1109/LSP.2018.2822810>.
- [25] WANG X, WANG S, ZHANG S, et al. Support Vector Guided Softmax Loss for Face Recognition[J], 2018.
- [26] DENG J, GUO J, ZAFEIRIOU S. ArcFace: Additive Angular Margin Loss for Deep Face Recognition[J], 2018.
- [27] LIU H, ZHU X, LEI Z, et al. AdaptiveFace: Adaptive Margin and Sampling for Face Recognition[J]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019 : 11939–11948.

- [28] ZHANG X, ZHAO R, QIAO Y, et al. AdaCos: Adaptively Scaling Cosine Logits for Effectively Learning Deep Face Representations[J]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019 : 10815 – 10824.
- [29] LIU B, DENG W, ZHONG Y, et al. Fair Loss: Margin-Aware Reinforcement Learning for Deep Face Recognition[J]. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019 : 10051 – 10060.
- [30] SUN T, ZHU X, PAN J-S, et al. No-Reference Image Quality Assessment in Spatial Domain[C] // International Conference on Genetic and Evolutionary Computing. 2014.
- [31] MITTAL A, SOUNDARARAJAN R, BOVIK A C. Making a “Completely Blind” Image Quality Analyzer[J]. IEEE Signal Processing Letters, 2013, 20 : 209 – 212.
- [32] VENKATANATH N, PRANEETH D, BH. M C, et al. Blind image quality evaluation using perception based features[J]. 2015 Twenty First National Conference on Communications (NCC), 2015 : 1 – 6.
- [33] HERNANDEZ-ORTEGA J, GALBALLY J, FIERREZ J, et al. FaceQnet: Quality Assessment for Face Recognition based on Deep Learning[J]. 2019 International Conference on Biometrics (ICB), 2019 : 1 – 8.
- [34] TERHORST P, KOLF J N, DAMER N, et al. SER-FIQ: Unsupervised Estimation of Face Image Quality Based on Stochastic Embedding Robustness[J]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020 : 5650 – 5659.
- [35] LLOYD S P. Least squares quantization in PCM[J]. IEEE Trans. Inf. Theory, 1982, 28 : 129 – 136.
- [36] ESTER M, KRIEGEL H-P, SANDER J, et al. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise[C] // Knowledge Discovery and Data Mining. 1996.
- [37] WANG Z, ZHENG L, LI Y, et al. Linkage Based Face Clustering via Graph Convolution Network[J]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019 : 1117 – 1125.
- [38] YANG L, CHEN D, ZHAN X, et al. Learning to Cluster Faces via Confidence and Connectivity Estimation[J]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020 : 13366 – 13375.

- [39] DU H, SHI H, ZENG D, et al. The Elements of End-to-end Deep Face Recognition: A Survey of Recent Advances[J]. ACM Computing Surveys (CSUR), 2020, 54 : 1 – 42.
- [40] HUANG G B, MATTAR M A, BERG T L, et al. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments[C] //. 2008.
- [41] SENGUPTA S, CHEN J-C, CASTILLO C D, et al. Frontal to profile face verification in the wild[J]. 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), 2016 : 1 – 9.
- [42] MOSCHOGLOU S, PAPAIOANNOU A, SAGONAS C, et al. AgeDB: The First Manually Collected, In-the-Wild Age Database[J]. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 : 1997 – 2005.
- [43] ZHENG T, DENG W, HU J. Cross-Age LFW: A Database for Studying Cross-Age Face Recognition in Unconstrained Environments[J]. ArXiv, 2017, abs/1708.08197.
- [44] ZHENG T, DENG W. Cross-Pose LFW : A Database for Studying Cross-Pose Face Recognition in Unconstrained Environments[C] //. 2018.
- [45] WHITELAM C, TABORSKY E, BLANTON A, et al. IARPA Janus Benchmark-B Face Dataset[J]. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 : 592 – 600.
- [46] MAZE B, ADAMS J C, DUNCAN J A, et al. IARPA Janus Benchmark - C: Face Dataset and Protocol[J]. 2018 International Conference on Biometrics (ICB), 2018 : 158 – 165.
- [47] WANG X, WANG S, ZHANG S, et al. Support Vector Guided Softmax Loss for Face Recognition[J]. ArXiv, 2018, abs/1812.11317.
- [48] LIU W, WEN Y, YU Z, et al. SphereFace: Deep Hypersphere Embedding for Face Recognition[J]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017 : 6738 – 6746.
- [49] JAIN A K, DUBES R C. Algorithms for Clustering Data[C] //. 1988.
- [50] YI D, LEI Z, LIAO S, et al. Learning Face Representation from Scratch[J]. ArXiv, 2014, abs/1411.7923.
- [51] AMIGÓ E, GONZALO J, ARTILES J, et al. A comparison of extrinsic clustering evaluation metrics based on formal constraints[J]. Information Retrieval, 2009, 12 : 613.