

# 基于特征偏移的人脸图像质量评价及影响因素研究

彭悦, 杨红雨, 刘艳丽

(四川大学计算机学院, 成都 610065)

**摘要:** 针对低质量人脸图像阻碍识别系统性能提升的问题, 本文提出了一种无参考的人脸图像质量评价方法, 并使用该方法评估了不同类型的图像退化对人脸图像质量的影响程度. 该方法使用一种集群卷积网络结构, 模拟人脸图像退化过程中的特征偏移, 根据特征偏移量和图像信息量之间的相关性, 完成人脸图像质量分数计算. 使用遗传算法对构成集群网络的网络单元进行筛选, 可使用更小网络规模实现同等性能. 以质量评价算法为工具进行实验, 研究评估了不同图像退化类型对人脸识别的影响, 为指导今后人脸质量相关研究得出了有益结论. 在主流人脸数据集上进行的实验证明, 通过筛选数据库中低质量分数的人脸图像, 可以进一步提升有人脸识别系统的性能, 且识别率提升表现出良好稳定性. 该方法复杂度低, 无需训练, 与 FaceQNet 等最新方法相比, 在 FNMR 和 EER 指标上显示出明显优势.

**关键词:** 卷积网络; 人脸识别; 图像质量评价; 遗传算法

**中图分类号:** TP391

**文献标识码:** A

**DOI:** 10.19907/j.0490-6756.2022.033003

## Face image quality evaluation based on feature offset and research on influencing factors

PENG Yue, YANG Hong-Yu, LIU Yan-Li

(College of Computer Science, Sichuan University, Chengdu 610065, China)

**Abstract:** Aiming at the problem that low-quality face images hinder the performance improvement of recognition systems, we propose a no reference face image quality evaluation method to evaluate the impact of different types of image degradation on the quality of face images. This method uses a cluster convolutional network structure to simulate the feature offset in the face degradation process, and face image score is calculated based on the correlation between the feature offset and the amount of image information. The genetic algorithm is used to select the network units, so that the same performance can be achieved with a smaller network scale. Using the evaluation algorithm as a tool for experiments, the impact of different image degradation types on face recognition is evaluated and studied, some useful conclusions are drawn for guiding future research on face quality. Experiments conducted on mainstream face data sets show that by filtering low-quality face images in the database, the performance of the existing face recognition system can be further improved, and the improvement of the recognition rate shows good stability. This method is low in complexity and does not require training. Compared with the latest methods such as FaceQNet, it shows obvious advantages in FNMR and EER indicators.

**Keywords:** Convolutional network; Face recognition; Image quality evaluation; Genetic algorithm

收稿日期: 2021-09-23

基金项目: 国家自然科学基金(61972271)

作者简介: 彭悦(1992—), 男, 河北保定人, 硕士研究生, 研究方向为计算机视觉. E-mail: 2019223040093@stu.scu.edu.cn

通讯作者: 刘艳丽. E-mail: yanliliu@scu.edu.cn

## 1 引言

人脸图像识别是最广泛使用的生物特征识别方法之一<sup>[1]</sup>,这是因为它具有很高的公众接受度,且在非配合状态下仍可进行采集匹配<sup>[2]</sup>,不仅在门禁、交通、考勤、支付和社交媒体等日常生活场景广泛使用,在安防、流控、刑侦等公共安全领域也具有重要地位。

近年来,随着深度学习在计算机视觉领域的广泛应用,其在人脸相关任务上已取得卓越性能。得益于网络结构和损失函数等方面的调整创新,最新的深度学习方法在主流人脸数据集上可取得 99% 以上的精度<sup>[3]</sup>。人脸识别领域的研究重心逐渐由识别模型转移到对人脸数据本身的分析上来。

人脸图像是对人脸的采样,其采样质量是影响人脸识别系统性能的重要因素,在采样条件受控情况下,当前的人脸识别系统能够实现高精度性能。然而在非配合、不受限、采集环境差和图像采集人员技术水平差等情况下,人脸在多个维度上呈现出丰富、剧烈的变化,这会导致识别精度的显著降低。这些变化因素包括图像获取条件(例如照明,背景,模糊和低分辨率),以及人脸因素(例如姿势,遮挡,化妆,配饰和表情),这些变化因素会导致识别性能显著下降。不同图像退化方式对人脸识别的影响程度不同,研究其差异,对指导提高人脸识别系统性能和相关领域研究具有积极意义。

## 2 相关研究

随着人脸图像相关领域的蓬勃发展,关于人脸质量和识别置信度的研究逐渐成为新的热门领域,论文数量展现出逐年增多的趋势<sup>[4]</sup>,同时可以观察到,深度学习技术正逐步取代传统手工算法,成为本领域的主流方法,在各分支均取得了不错进展。

最新的国内外研究成果包括,文献[5]利用随机辍学技术获得人脸随机嵌入,根据图像随机嵌入鲁棒性评价人脸质量;文献[6]结合人类视觉特性,设计 4 种梯度卷积模板,结合人脸特征点,提取特征点附近的人脸梯度信息,使用回归方法预测质量分数;文献[7,8]根据 ICAO 标准选择标准人脸图像,通过计算与标准人脸图像的距离,预测人脸图像质量,批量生成带有质量标签的人脸数据集,在预训练人脸识别模型上进行迁移学习计算质量得分;文献[9]提出了一种基于排名的学习方法,他们使用基于排名损失的多项式核和训练权重定义了

一个线性质量评估方程;文献[10]创建了 FIQD,一个“人脸图像照明质量数据库”,具有 224 个人,733 张图像,200 种不同的照明模式及其质量分数,然后使用该数据训练基于 ResNet-50<sup>[11]</sup>的模型以估计照明质量;文献[12]将图像灰度、色彩、构图等全局特征与剪裁后的人脸特征进行融合,预测带有人脸图像的美学得分;文献[13]对人脸图像损失类型进行识别并利用回归获得质量分数。

## 3 本文模型

本文提出一种新的人脸图像质量评价算法,通过网络模型模拟人脸图像退化过程中出现的特征偏移,根据特征偏移量和图像信息量之间的相关性,完成人脸质量分数自动评价。使用遗传算法对构成模型的网络单元进行筛选,用更小网络规模实现同等性能。

本文贡献可概括如下:将特征偏移概念引入人脸图像质量评价,通过微调人脸识别网络权重,模拟图像退化形成的特征偏移,利用其与人脸质量的相关性,提出基于特征偏移的人脸质量评价算法框架。

采用遗传算法自动筛选关键特征,实现网络规模压缩和计算复杂度简化。使用本文提出的质量评价算法研究高斯模糊、运动模糊、高斯噪声、椒盐噪声等多种常见类型图像退化对人脸识别的影响差异。

### 3.1 主要思想

在当前主流人脸识别方法中,通常使用二维或三维人脸进行特征提取,通过特征比对和阈值设置完成识别或验证任务。人脸图像,是对原始人脸的采样,采样过程必然伴随一定误差和特征损失,误差和特征损失来源包括且不限于:维度损失,离散化损失,光照不均,光照不足,表情造成的面部扭曲,化妆,佩戴装饰品,面部遮挡,图像采集设备造成的模糊和噪声等。

如图 1 所示,假设每张人脸映射到特征空间里的唯一点,代表其独一无二的身份特征,那么上述原因造成的误差和特征损失反映到特征空间里,即为不同程度的特征点偏移<sup>[14]</sup>。类似地,如果对面脸图像重采样,将导致信息进一步丢失,特征点发生二次偏移。为方便读者理解特征偏移概念,对其做出如下定义:

$$S(I) = d(F(I'), F(I)) \quad (1)$$

其中, $I$  为人脸图像; $I'$  为重采样(退化)人脸图像;

$F$  为特征提取函数(此处即为用于人脸特征提取的神经网络),特征偏移即为退化人脸特征与原人脸特征之间的距离。

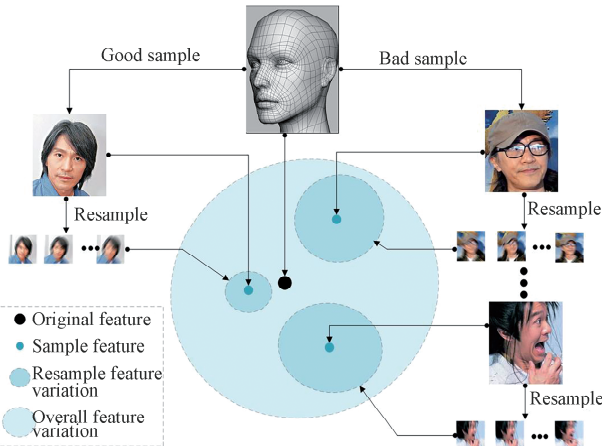


图 1 人脸采样与特征偏移示意  
Fig. 1 Face sampling and feature variation

文献[15]从“数据不确定性”角度出发,认为单张人脸图像提取的特征服从高斯分布,均值代表最接近特征真值的点,方差代表人脸图像质量,方差越小,类内紧凑型和类间可分性越好。同理,在本文语境中,特征点偏移类似上述方差的概念,图像质量越高,包含信息量越大,偏移幅度越小;图像质量越差,包含信息量越小,偏移幅度越大。受此思想启发,对特征提取网络进行权重微调,可模仿采样过程中图像退化造成的误差和特征丢失,从而导致特征点偏移。为叙述简洁,后文我们将微调后的网络称作“特征偏移网络”。

如果人脸图像质量差,即特征信息损失多,那么输入特征偏移网络后,造成的特征偏移大(因为图像包含有价值的人脸信息少),即称之具有低质量;反之,如果图像采样质量高,特征信息损失少,

那么输入特征偏移网络后,造成的特征偏移小(因为图像包含有价值的人脸信息多),即称之具有高质量。

对特征提取网络随机进行多次权重微调,可形成特征偏移网络集群。利用特征偏移网络集群(包含  $m$  个特征偏移网络),可得到  $m$  个偏移特征。使用统计学度量方法计算集群网络造成的平均特征偏移,即可代表图像质量,并与人脸识别表现高度相关。

3.2 模型结构

特征偏移网络及网络集群. 将预训练的人脸识别卷积神经网络去掉分类层,作为无偏移的特征提取网络,本文使用预训练的 ResNet-18<sup>[11]</sup>模型作为无偏移特征网络,其结构如图 2 所示。然后随机生成  $N$  个特征选取模板,使用模板屏蔽最后一个全连接层部分权重,即可生成  $N$  个特征偏移网络,从而模拟人脸图像退化过程中部分特征信息的丢失,  $N$  个特征偏移网络共同构成集群卷积网络,集群卷积网络结构如图 3 所示。

layer name	output size	ResNet-18
Conv1	112×112	7×7, 64, stride 2
Conv2_x	56×56	3×3 max pool, stride 2
		$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix} \times 2$
Conv3_x	28×28	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix} \times 2$
Conv5_x	14×14	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix} \times 2$
Conv4_x	7×7	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix} \times 2$
	1×1	Average pool, 1000-d fc softmax
FLOPs		$1.8\times10^9$

图 2 ResNet-18 网络结构  
Fig. 2 ResNet-18 network structure

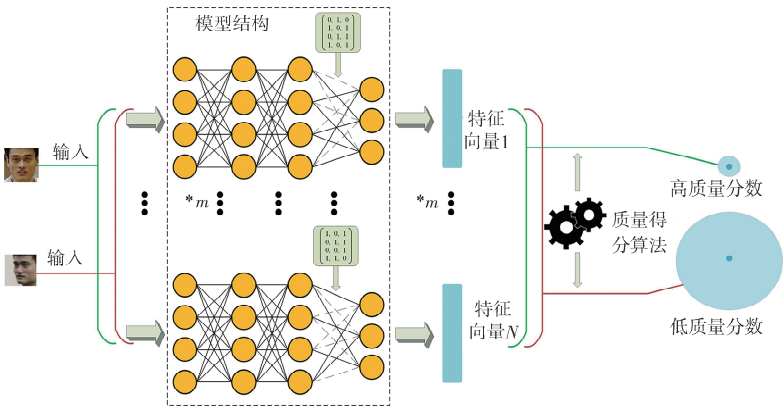


图 3 集群卷积网络结构  
Fig. 3 Structure of cluster convolutional

### 3.3 特征偏移计量

使用平均欧式距离对偏移量进行计量.  $I$  为输入的人脸图像,  $m$  为偏移特征个数,  $f_i, f_j (1 \leq i \leq j \leq m)$  为图像  $I$  得到的  $m$  个偏移特征, 计算  $m$  个偏移特征两两之间欧氏距离并求和, 因为求和号中共有  $\frac{m^2-m}{2}$  个距离量, 因此求得每张人脸平均特征偏移如下.

$$S(I) = \frac{2}{m^2 - m} \sum_{1 \leq i < j \leq m} d(f_i, f_j) \tag{2}$$

将特征偏移进行归一化操作, 得到取值区间为 0~1 的人脸图像质量分数. 计算公式如下:

$$P(I) = 2 \times (\sigma(-\frac{2}{m^2 - m} \sum_{i < j} d(f_i, f_j)) - 0.5) \tag{3}$$

其中,  $P(I)$  代表图像  $I$  的质量分数;  $\sigma$  为 sigmoid 函数;  $f_i$  为集群网络中第  $i$  个特征偏移网络提取的特征向量. 负号体现特征偏移与人脸质量的负相关关系.

### 3.4 特征选取

最终提取的特征向量中, 不同特征对完成识别任务的贡献率不同. 举例说明, 脸型、瞳距等特征比皮肤纹理对于身份识别有更大贡献. 因此应筛选对识别任务影响更大的关键特征, 通过影响这些关键特征, 可提升采样质量评估效率.

本文使用遗传算法对特征偏移网络进行筛选. 选取偏移量尽可能大的子网, 使质量评估更为高效. 每个特征偏移子网, 均由一个独特的特征偏移模板生成, 子网与模板是一一对应的(图 3), 因此对特征偏移模板的筛选, 即为对特征偏移网络的筛选, 我们使用特征偏移模板作为染色体编码对象. 每个模板都是由一系列 0 和 1 的布尔值变量组成的  $m \times n$  矩阵( $m$  为倒数第二层维度,  $n$  为最后一层维度), 可将其展开为  $m \times n$  维向量, 作为遗传算法的染色体单元, 如图 4. 初代染色体为满足二项分布的随机值, 二项分布的概率取值为 0.5, 通过遗传算法筛选迭代, 最终得到目标个体.

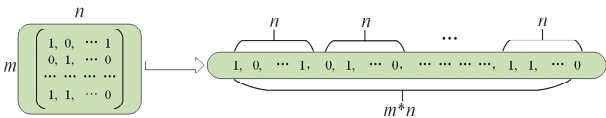


图 4 遗传算法染色体编码

Fig. 4 Genetic algorithm chromosome coding

从 VGGFace2 数据集中随机选取 100 张人脸形成迷你子集, 用于遗传算法筛选特征. 为筛选关

键特征的优化目标为:

$$F_x^* = \operatorname{argmax}_{F_x} (\sum_{i=1}^{100} \|F_x(I_i) - F(I_i)\|^2) \tag{4}$$

式中  $F_x(I_i)$  为特征偏移网络输出的特征;  $F(I_i)$  为无偏移特征, 优化目标为找到特征偏移量最大的  $F_x$  函数, 即特征偏移网络, 这里为全面体现正负偏移量, 求差后取 2 范数.

适应度函数设计为

$$\text{Fitness} = \frac{\sum_{i=1}^{100} \|F_x(I_i) - F(I_i)\|^2}{\max_{1 \leq x \leq 100} (\sum_{i=1}^{100} \|F_x(I_i) - F(I_i)\|^2)} \tag{5}$$

分子为待计算个体(网络)形成的特征偏移量, 求和号的 100 代表迷你集的 100 个人脸样本. 分母为当前代次表现最好的网络, 即特征偏移最大的网络的偏移量, 因此取  $\max_x$ .  $x$  取值范围 0~100, 代表当前代次 100 个个体(网络), 每个个体(这里指特征偏移网络)的适应度, 就是它的特征偏移量与当前代次最大特征偏移量的比值, 取值范围是 0%~100%, 以此作为个体(特征偏移网络)的生存适应度. 然后按照适应度执行遗传算法, 淘汰不良个体.

### 3.5 图像退化

本文选取了几种常见图像退化方式进行模拟, 以研究图像退化对人脸识别的影响. 主要研究对象分为高斯模糊、运动模糊、高斯噪声、椒盐噪声四种单独类型退化方式和高斯模糊加高斯噪声、高斯模糊加椒盐噪声、运动模糊加高斯噪声、运动模糊加椒盐噪声四种叠加退化方式. 每种退化方式分为 20 个层级, 退化程度由轻到重逐步加深. 例如将高斯模糊的高斯核尺寸逐级加大, 逐级提高高斯噪声的噪声添加比例, 具体设置将在第 4.4 节介绍. 将上述 8 种退化方式各 20 个层级分别应用于 LFW 数据处理, 生成 160 个模拟图像退化的实验数据集.

## 4 实验及分析

本节介绍了实验设置和结果, 在主流数据集上, 与最新的方法进行了对比实验, 对结果进行了定性和定量分析, 并补充了对算法一致性的说明.

### 4.1 数据集

LFW 数据集主要测试人脸识别的准确率, 该数据库从中随机选择了 6000 对人脸组成了人脸辨识图片对, 其中 3000 对属于同一个人 2 张人脸照



片, 3000 对属于不同的人每人 1 张人脸照片. 测试过程 LFW 给出一对照片, 询问测试中的系统两张照片是不是同一个人, 系统给出“是”或“否”的答案. 通过 6000 对人脸测试结果的系统答案与真实答案的比值可以得到人脸识别准确率. 这个集合被广泛应用于评价 face verification 算法的性能.

VGGFace2<sup>[16]</sup>是一个大规模人脸识别数据集, 该数据集是从谷歌中下载的, 包含不同姿态、年龄、光照和背景的人脸图片, 其中约有 59.7% 的男性. 除了身份信息之外, 数据集还包括人脸框, 5 个关键点、以及估计的年龄和姿态. 共包含 331 万图片, 9131 个 ID, 平均每个 ID 图片个数为 362.6. 这个数据集有以下几个特点: (1) 人物 ID 数量大, 且每个 ID 包含的图片个数也较多; (2) 覆盖大范围的姿态、年龄和种族; (3) 噪声数据少. 实验图像均使用 Retinaface<sup>[17]</sup>进行预处理.

为研究不同图像损失类型对人脸识别的影响, 我们使用高斯模糊、运动模糊、高斯噪声、椒盐噪声 4 种常见图像损失方式和“高斯模糊加高斯噪声”、“运动模糊加高斯噪声”、“高斯模糊加椒盐噪声”、“运动模糊加椒盐噪声”4 种交叉损失方式, 每种方式从弱到强分为 20 个层级, 对 LFW 数据集进行模拟退化处理, 作为数据扩充辅助研究. 因为每种图像退化方式对图像质量的影响曲线不同, 为使曲线上各点分布均匀, 退化参数为手动设置并经过反复测试调整, 具体数值如下:

高斯模糊: 高斯核尺寸第一级设置为  $3 \times 3$ , 之后边长逐级加 4.

运动模糊: 运动方向为随机方向, 模糊卷积核为正方形, 尺寸设置为  $[2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32]$ .

高斯噪声: 高斯噪声是符合高斯分布的加性噪声, 为不影响图像整体亮度, 噪声均值设置为 0, 方差为未做归一化处理的像素值方差, 取值区间为  $[0 \sim 255^2]$ , 本实验第一级设置为 3, 之后每级加 3.

椒盐噪声: 噪声添加比例为  $[0.0005, 0.0010, 0.0020, 0.0040, 0.0080, 0.012, 0.016, 0.020, 0.028, 0.036, 0.044, 0.052, 0.060, 0.068, 0.078, 0.088, 0.098, 0.108, 0.118, 0.128]$ .

## 4.2 实验设置

模型使用无监督算法, 不需要进行单独训练, 构成集群卷积网络的每个单元网络只需加载预训

练的模型参数, 并使用特征偏移模板对网络进行参数重构. 对于集群网络中的单元网络个数  $N$ , 根据我们的经验取  $N=10$  可以较好平衡精度和性能. 具有类似网络结构的方法<sup>[5]</sup>推荐子网个数为 100, 相比之下, 本文通过遗传算法筛选特征后  $N$  取 10 即可较好地平衡精度和性能, 进一步精简了网络结构.

在特征筛选部分, 遗传算法筛选出 10 个个体, 即 10 个特征偏移网络, 最终形成目标集群网络. 这些个体使用布尔值, 按照  $p=50\%$  的筛选率随机初始化 100 个个体, 持续 300 个代次, 交叉组合率 80%, 变异率 2%, 代次筛选率 10%.

为了横向比较本文方法, 实验引入来自计算机视觉社区的三个基线方法: *Brisque*<sup>[18]</sup>, *Niqe*<sup>[19]</sup> *Piqe*<sup>[20]</sup>, 均为无参考图像质量评价. 另外引入两个学术界最新的人脸质量评价方法 *FaceQnet*<sup>[7]</sup>, *Best-Rowden*<sup>[21,22]</sup>, 均采用预训练模型参与实验.

## 4.3 图像退化

本文选取了几种常见图像退化方式进行模拟, 以研究图像退化对人脸识别的影响. 主要研究对象分为高斯模糊、运动模糊、高斯噪声、椒盐噪声四种单独类型退化方式和高斯模糊加高斯噪声、高斯模糊加椒盐噪声、运动模糊加高斯噪声、运动模糊加椒盐噪声四种叠加退化方式. 每种退化方式分为 20 个层级, 退化程度由轻到重逐步加深. 例如将高斯模糊的高斯核尺寸逐级加大, 逐级提高高斯噪声的噪声添加比例, 具体设置将在第 4.4 节介绍. 将上述 8 种退化方式各 20 个层级分别应用于 LFW 数据处理, 生成 160 个模拟图像退化的实验数据集.

## 4.4 度量方法

为了评估算法性能, 我们遵循<sup>[23]</sup>的方法, 使用错误率与拒绝曲线. 曲线横轴为拒绝掉低质量人脸图像的比率, 纵轴为剩余人脸图像的错误识别率. 理想的情况是, 当拒绝一部分质量最低的人脸图像后, 人脸识别系统在剩余图像上的错误识别率应降低, 并且随着拒绝率的不断提升, 剩余图像的错误识别率不断降低. 具体的, 我们使用 FNMR 错误非匹配率, EER 等错误率两项指标评价算法性能, 因为这些错误率在国际标准中被指定用于生物特征验证评估<sup>[24]</sup>.

在研究图像退化对人脸识别影响的问题上, 我们使用等质量分数识别率曲线和识别率直方图进行分析. 对经过图像退化处理的数据集, 使用本文

中的人脸图像质量算法逐张人脸图像计算质量分数,并求得数据集平均得分,同时使用 resnet18 预训练模型加 arcface 损失函数,测试每个数据集的验证精确率。

4.5 实验结果及分析

为直观展示算法有效性,我们在 LFW 和 VG-GFace2 两个当前主流人脸图像数据集上进行了质

量分数预测. 绘制总体分数分布图(图 5),并列出了最坏和最好的人脸(图 6). 在两个数据集上都可以观察到,得分最高的人脸几乎都具有“无遮挡,光照均匀,五官清晰,姿态端正”等共同特征;得分最低的人脸均具有“佩戴物遮挡,头发、手部遮挡,表情夸张,姿态不端”其中一个或多个影响身份识别的问题。

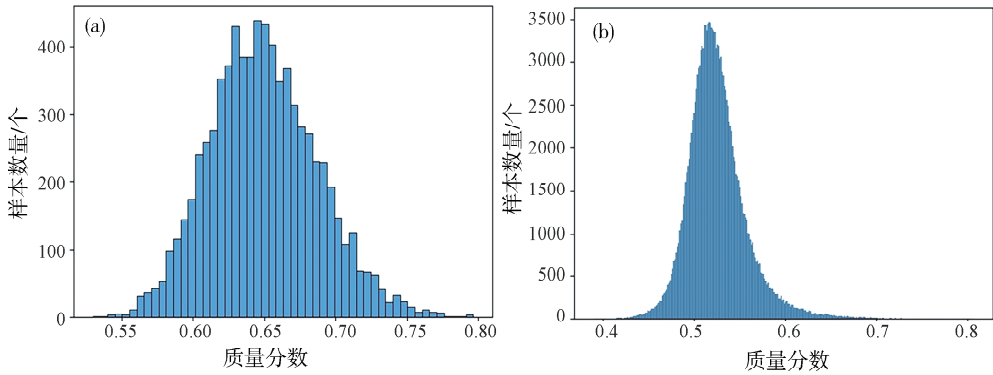


图 5 LFW 数据集质量分数分布(a); VGGFace2 数据集质量分数分布(b)

Fig. 5 Quality score distribution of LFW dataset (a); Quality score distribution of VGGFace2 dataset (b)

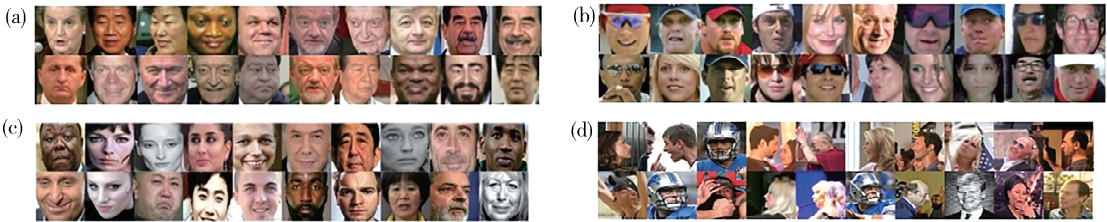


图 6 LFW 数据集得分最高的 20 张(a)、最低的 20 张(b)人脸;  
VGGFace2 数据集得分最高的 20 张(c)、最低的 20 张(d)人脸

Fig. 6 The 20 faces with the highest scores (a) and the 20 faces with the lowest scores (b) in the LFW dataset; the 20 faces with the highest scores (c) and the 20 faces with the lowest scores (d) in the VGGFace2 dataset

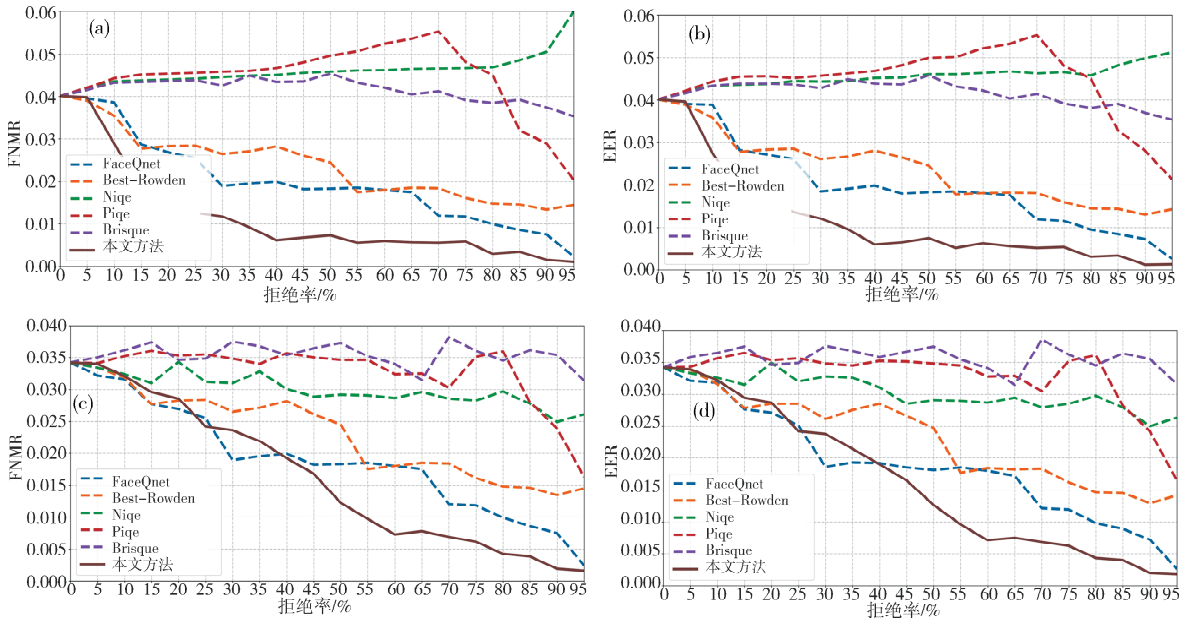


图 7 LFW 数据集错误率与拒绝曲线(a,b);VGGFace2 数据集错误率与拒绝曲线(c,d)

Fig. 7 Error versus reject curves of LFW dataset (a,b); Error versus reject curves of VGGFace2 dataset (c,d)

LFW 数据集分数分布普遍高于 VGGAFace2 数据集,且分布范围更广,这是合乎预期的,因为 VGGFace2 数据包更多无约束照片,人物呈现大量年龄、表情、姿态、光照变化,且背景更加复杂,对身份识别具有更大挑战。

对三个基线方法和两个学术界最新方法进行

了定量实验,为公平比较算法性能,采用 4.4 节介绍的错误率拒绝曲线,如图 7. 使用所有参与实验的方法,对数据集全部人脸预测质量得分,然后按照一定比率拒绝得分最低的部分人脸,计算剩余人脸的识别错误率,逐渐提高拒绝率,重复此过程,最终绘制完整曲线。

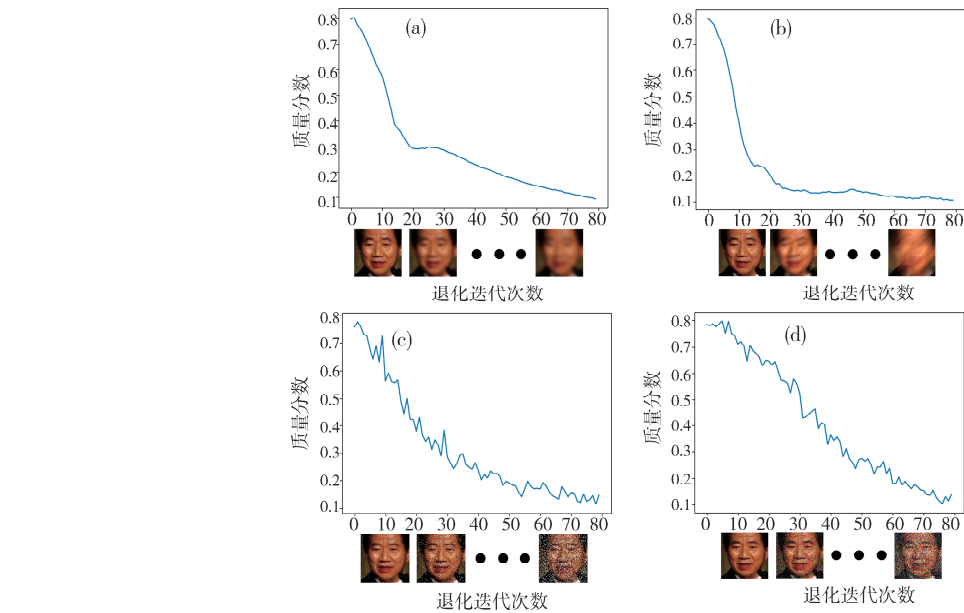


图 8 4 种图像退化-质量分数曲线  
(a) 高斯模糊;(b) 运动模糊;(c) 高斯噪声;(d) 椒盐噪声  
横轴为退化迭代次数,纵轴为人脸图像质量得分

Fig. 8 Four kinds of image degradation-quality score curve

(a) Gaussian blur; (b) motion blur; (c) Gaussian noise; (d) salt and pepper noise

可以观察到,本文方法在两个数据集上均表现出了较好性能. 通过拒绝低质量人脸图像,实现了识别精度的提升,错误率进一步下降, FNMR、EER 两个指标均证明了算法有效性. 且随着拒绝率逐步提升,错误率可以稳步下降,体现了算法具有良好稳定性. 绝大多数情况下,本文方法的曲线位于对比方法的曲线下方。

为进一步验证算法稳定性,我们通过手工图像退化进行补充验证,如图 8. 使用高斯模糊,运动模糊,椒盐噪声三种常见图像退化方式,对人脸图像进行手工退化操作,并使用我们的方法计算人脸图像质量. 可以观察到,随着退化程度加深,同幅图像四种退化方式的分数均表现出连续下降,同样证明了图像质量分数与可识别率之间的高度相关性。

从识别率质量分数曲线(如图 9)观察到,随着图像退化程度逐渐加深,数据集平均质量分数下降,且在平均质量分数相同情况下,不同图像退化类型的识别率体现出差异性,具体地,在单一退化

方式中,对识别率影响程度由小到大依次为:高斯噪声,高斯模糊,运动模糊,椒盐噪声. 注意到橙绿两条曲线在后半段出现交替,但考虑到此时图像质量已经严重破坏,没有考察价值,并不影响上述结论. 在椒盐噪声最后几个退化级别上出现了质量分数异常(绿色曲线),具体表现为:随着图像退化级别加深,质量分数不再降低,反而有小幅提高. 这可

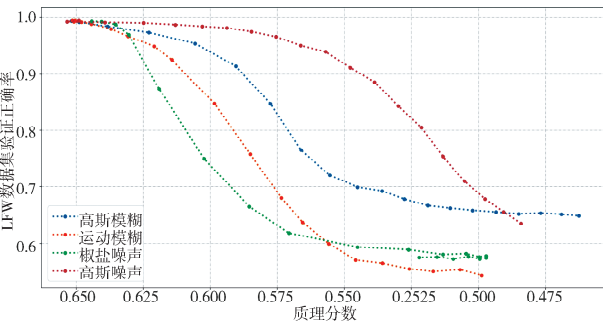


图 9 验证率-质量分数曲线

Fig. 9 Verification rate-quality score curve

能是因为椒盐噪声具有较大随机性,本文算法正是

利用随机特征偏移工作的. 当随机噪声增大到一定程度时, 算法误把部分噪声当作有效信息处理, 造成了算法波动. 虽不影响总体结论, 但有待进一步研究改进.

图 9 中, 每条曲线代表一种图像退化方式, 包含 20 个退化级别的验证率-平均质量分数构成的 20 个数据点. 可以观察到不同退化方式对图像质量的影响是不同的.

质量分数直方图如图 10 所示, 横坐标为 20 个图像退化级别, 1 级退化程度最轻, 20 级退化程度最重. 从图 10 可知, “模糊+噪声”两种退化类型叠加的方式比单一模糊或单一噪声的退化类型对识

别率影响更大. 在 2 种模糊类型和 2 种噪声类型形成的 4 个交叉组合实验中, 均证明了这一规律. 虽然 4 种退化类型对识别率的影响程度不一, 但叠加方式强于单一方式的规律始终没变, 这在图像退化程度较浅的前几个层级数据中尤其明显. 这是因为, 此时人脸图像质量和识别率较高, 图像退化对于人脸身份信息的损失收益大. 在退化等级大于 10 以后, 由于图像质量严重受损, 验证率已接近 50% 的极限, 图像退化对图像质量的影响已非常有限, 上述规律变得不明显, 甚至出现反常, 此类现象属于极端情况, 没有考察价值, 不影响整体结论.

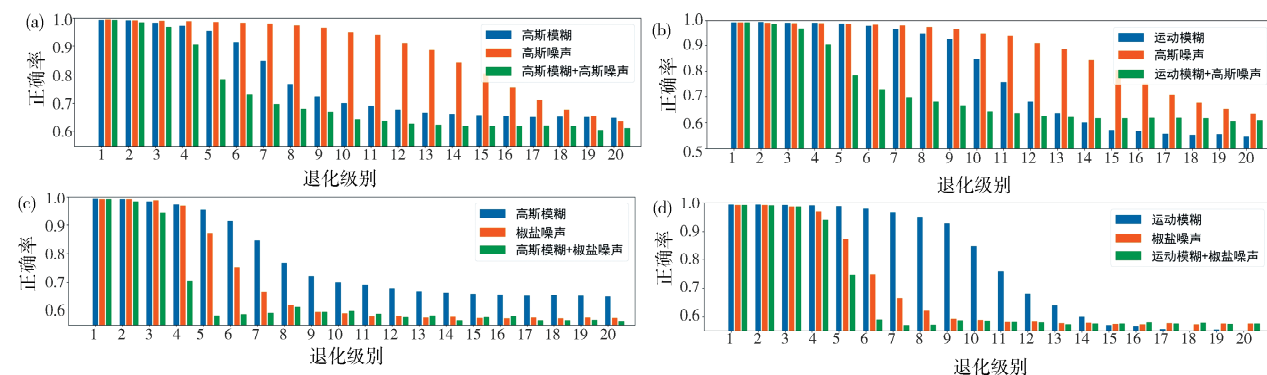


图 10 验证率直方图

(a) 高斯模糊+高斯噪声; (b) 运动模糊+高斯噪声; (c) 高斯模糊+椒盐噪声; (d) 运动模糊+椒盐噪声

Fig. 10 Verification accuracy histogram

(a) Gaussian blur + Gaussian noise; (b) motion blur + Gaussian noise;

(c) Gaussian blur + salt and pepper noise; (d) motion blur + salt and pepper noise

## 5 结 论

针对低质量人脸图像阻碍识别系统性能提升的问题, 本文提出了一种无参考的人脸图像质量评价方法. 分析阐述了利用特征偏移与图像信息量的相关性进行质量分数计算的思想. 提出了一种集群卷积网络模型, 用于人脸图像质量分数自动预测. 使用遗传算法对网络模型进一步优化, 在保持同等性能前提下大幅压缩了模型复杂度. 通过在主流人脸数据集上进行的实验, 定性和定量分析证明了算法具有优越性, 通过筛选数据库中低质量得分的人脸图像, 可以进一步提升现有人脸识别系统的性能, 且识别率提升表现出良好稳定性. 模拟了几类常见图像退化方式, 对原始人脸数据库进行数据扩充, 并使用本文提出的质量评价算法为工具进行实验, 研究评估了不同类型图像损失对人脸识别的影响, 为指导今后人脸质量相关研究得出了有益结论.

## 参考文献:

[1] Wang C P, Wei W, Zhang J S, *et al.* Robust face recognition via discriminative and common hybrid dictionary learning [J]. Appl Intell, 2018, 48: 156.

[2] Tripathi B K. On the complex domain deep machine learning for face recognition [J]. Appl Intell, 2017, 47: 382.

[3] Grother P, Ngan M, Hanaoka K. Ongoing face recognition vendor test (frvt) part 2: Identification [EB/OL]. (2018-11-27) [2021-08-08]. <https://www.nist.gov/publications/ongoing-face-recognition-vendor-test-frvt-part-2-identification>.

[4] Schlett T, Rathgeb C, Henniger O, *et al.* Face image quality assessment: a literature survey [EB/OL]. (2015-05-26) [2021-08-01]. <https://arxiv.org/abs/2009.01103>.

[5] Terhorst P, Kolf J N, Damer N, *et al.* SER-FIQ: unsupervised estimation of face image quality based on stochastic embedding robustness [C]//Proceed-

- ings of the 2020 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). [S. l. ]: IEEE, 2020.
- [6] Lu Y, Li W, Ning X, *et al.* Face quality assessment based on local gradient [C]//Proceedings of the 2020 IEEE International Conference on Artificial Intelligence and Information Systems (ICAIS). [S. l. ]: IEEE, 2020.
- [7] Hernandez-Ortega J, Galbally J, Fierrez J, *et al.* FaceQnet: quality assessment for face recognition based on deep learning [C]//Proceeding of the 12th IAPR International Conference on Biometrics (ICB). Crete, Greece: IEEE, 2019.
- [8] Hernandez-Ortega J, Galbally J, Fierrez J, *et al.* Biometric quality: review and application to face recognition with faceqnet [EB/OL]. (2021-02-28) [2021-10-08]. <https://arxiv.org/abs/2006.03298>.
- [9] Chen J, Deng Y, Bai G, *et al.* Face image quality assessment based on learning to rank [J]. IEEE Signal Proc Let, 2015, 22: 90.
- [10] Zhang L, Zhang L, Li L. Illumination quality assessment for face images: a benchmark and a convolutional neural networks based model [C]//Proceeding of the International. Conference on Neural Information Processing (ICONIP). Guangzhou, China: Springer International Publishing, 2017.
- [11] He K, X Zhang, S Ren, *et al.* Deep residual learning for image recognition [EB/OL]. (2015-12-10) [2021-08-08]. <https://arxiv.org/abs/1512.03385>.
- [12] Celona L, Schettini R. A genetic algorithm to combine deep features for the aesthetic assessment of images containing faces [J]. Sensors, 2021, 21: 1307.
- [13] Chen S Y, Zhang H L, Yang J M. A face quality evaluation method based on DCNN [C]//Proceedings of the 2020 Chinese Control and Decision Conference. [S. l. ]: IEEE, 2020.
- [14] Chang J, Lan Z, Cheng C, *et al.* Data uncertainty learning in face recognition [C]//Proceeding of the Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE, 2020.
- [15] Shi Y C, Jain A K, Kalka N D. Probabilistic face embeddings [C]//Proceedings of the IEEE International Conference on Computer Vision. [S. l. ]: IEEE, 2019.
- [16] Cao Q, Shen L, Xie W, *et al.* Vggface2: A dataset for recognising faces across pose and age [C]//Proceedings of the 2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018). [S. l. ]: IEEE, 2018.
- [17] Deng J, Guo J, Ververas E, *et al.* Retinaface: single-shot multi-level face localisation in the wild [C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). [S. l. ]: IEEE, 2020.
- [18] Mittal A, Moorthy A K, Bovik A C. No-reference image quality assessment in the spatial domain [J]. IEEE T Image Process, 2012, 21: 4695.
- [19] Mittal A, Soundararajan R, Bovik A C. Making a completely blind image quality analyzer [J]. IEEE Signal Proc Let, 2013, 20: 209.
- [20] Venkatanath N, Praneeth D, Maruthi C B, *et al.* Blind image quality evaluation using perception based features [C]//Proceedings of the 2015 21st National Conference on Communications (NCC). [S. l. ]: IEEE, 2015.
- [21] Best-Rowden L, Jain A K. Learning face image quality from human assessments [J]. IEEE T Inf Foren Sec, 2018, 13: 3064.
- [22] Best-Rowden L, Jain A K. Automatic face image quality prediction [EB/OL]. (2017-06-02) [2021-08-01]. <https://arxiv.org/abs/1706.09887>.
- [23] Grother P, Tabassi E. Performance of biometric quality measures [J]. IEEE T Pattern Anal Mach Intell, 2007, 29: 531.
- [24] Tripathi B K. On the complex domain deep machine learning for face recognition [J]. Appl Intell, 2017, 47: 382.

#### 引用本文格式:

中 文: 彭悦, 杨红雨, 刘艳丽. 基于特征偏移的人脸图像质量评价及影响因素研究[J]. 四川大学学报: 自然科学版, 2022, 59: 033003.

英 文: Peng Y, Yang H Y, Liu Y L. Face image quality evaluation based on feature offset and research on influencing factors [J]. J Sichuan Univ: Nat Sci Ed, 2022, 59: 033003.