计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2018,27(11):71−77 [doi: 10.15888/j.cnki.csa.006608]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于 CNN 的监控视频中人脸图像质量评估①

王　亚, 朱　明, 刘成林

(中国科学技术大学 信息科学技术学院, 合肥 230027)
通讯作者: 王　亚, E-mail: wangyaya@mail.ustc.edu.cn

摘　要: 在公共安全领域, 监控视频中的人脸识别技术是不可或缺的技术, 成为研究热点. 而监控视频中低质量的人脸图像会大大降低整个人脸识别系统的识别准确率, 系统难以更广泛地被投入实际使用. 本文提出了一种基于 CNN 的人脸图像质量评估方法. 通过对 Alexnet 模型进行改进, 将网络中的多个卷积层与全连接层连接, 从而提取不同尺度的图像特征. 通过端到端的训练过程, 预测人脸图像质量分数. 另外, 采用人脸识别算法来标定人脸图像的质量分数, 使质量分数能更有效地筛选出适合识别算法的图像. 在 Color FERET 数据集上实验表明, 本文方法能够准确地对人脸图像进行质量评估. 而在实际采集的监控视频数据集上实验表明, 本文方法能筛选出高质量的人脸图像用作后续人脸识别, 提高人脸识别准确率.

关键词: CNN; 人脸图像质量评估; 监控视频

引用格式: 王亚,朱明,刘成林.基于 CNN 的监控视频中人脸图像质量评估.计算机系统应用,2018,27(11):71–77. http://www.c-s-a.org.cn/1003-3254/6608.html

Face Image Quality Assessment in Surveillance Videos Using CNN

WANG Ya, ZHU Ming, LIU Cheng-Lin

(School of Information Science and Technology, University of Science and Technology of China, Hefei 230027, China)

Abstract: Face recognition in surveillance videos is an essential technology in public security and has gotten more and more attention. But it is a little hard for the face recognition systems to be integrated into real application due to the low recognition rate caused partly by low face image quality. This study proposes a method of face image quality assessment using CNN. The proposed net, modified from the Alexnet, connects intermediate convolution layers to fully connect layer, to get multiple image features. Then, face image quality scores can be gotten from proposed net which is trained by end to end. In addition, a face image quality metric is used to relate the quality with the face recognition algorithm. Experiments on Color FERET datasets show that the proposed algorithm is able to elevate the face image quality exactly. Further experiments on a video surveillance dataset (collected by ourselves) show that the proposed method can select high quality face image for face recognition, leading to significant improvements in recognition accuracy.

Key words: CNN; face image quality assessment; surveillance videos

智能视频监控系统利用计算机视觉、图像处理等技术自动对监控视频内容进行识别分析与理解, 是视频监控系统的发展趋势. 而人脸识别模块是智能视频监控系统中的重要组成部分. 经过近 60 年的发展, 可限制条件下人脸识别方法已逐渐成熟, 成果众多[1]. 但是基于监控视频的人脸识别技术仍面临许多挑战. 一方面, 监控环境中存在光照、背景等不断变化; 另一方面, 视频中的人是自由行动的. 因此监控视频中采集到

① 基金项目: 国家重大科技专项 (2017ZX03001019)
Foundation item: National Science and Technology Major Project of China (2017ZX03001019)
收稿时间: 2018-03-26; 修改时间: 2018-04-24; 采用时间: 2018-04-27; csa 在线出版时间: 2018-10-24

的人脸经常会存在光照或姿态或表情变化大, 甚至由于运动而模糊的低质量人脸图像. 虽然很多方法[2–4]被提出来以增强人脸识别算法对低质量图像的鲁棒性, 但是很明显, 大多数识别算法在高质量的人脸图像上会实现更好的效果[5]. 以人脸验证为例, 在 2010 年由 NIST 组织的 MBE 中, 在高质量人脸数据库上测试时, 人脸验证错误率为 0.3%[6], 而在非限制数据集 LFW[7] 上的错误率不少于 18%[8]. 将低质量的人脸图像用于人脸识别, 不仅会降低整个系统的人脸识别率, 而且由于人脸识别过程中特征计算复杂, 造成计算资源浪费. 解决此问题的方法之一就是进行人脸质量评估, 筛选出高质量的人脸图像用于后续识别.

本文接下来的结构安排如下: 第 1 节介绍了常见的人脸图像质量评估方法; 第 2 节详细阐述了本文方法的算法思想及步骤; 相关实验设置及结果分析在第 3 节进行被说明; 第 4 节总结了本文的工作.

1 人脸图像质量评估相关方法

图像质量评估方法可分为主观评价和客观评价两种. 主观评价是以人为视觉感知为主, 辅以事先制定的一些评价尺度[9], 又可分为绝对评价和相对评价. 绝对评价时, 评估者直接按照视觉感受给出图像质量判断或分数. 而相对评价是给出一组图像, 评估者进行相对比较, 按照从低到高的分类做出评估. 主观评价方法符合人的主观感受, 但它耗时耗力, 且容易受评估者本身的专业背景、动机和情绪等主观因素影响. 客观评价方法是根据人类视觉特性, 利用数学算法, 对图像质量做出客观量化的评价值, 能方便地被集成到实际的相关系统中.客观评价方法主要用到方差、梯度、信息熵、峰值信噪比、均方误差等技术指标, 根据对参考图像的依赖程度, 可分为全参考、半参考和无参考图像质量评估方法.

人脸图像质量评估属于图像质量评估的一个分支, 既要考虑传统图像质量评价中所关注的因素, 如图像对比度、清晰度、光照等, 又要考虑人脸所特有的因素, 如姿态、表情、遮挡等. 2005 年, 国际标准化组织制定了关于人脸图像质量的标准, 对多种参数做出了规定[10]. 在此基础上, 许多人脸图像质量评估方法被提出来. 大致可以分为两种, 基于多因素融合的方法和基于全局学习的方法. 在多因素融合方法中, 先单独分析某种图像特性, 如对比度、光照、人脸姿态等, 然后加权融合形成最后的质量分数. 例如, Nasrollahi 等将姿态、对比度、亮度、分辨率的分数进行加权融合[11]; Castro 将对称性分析与两种对比度度量方法相结合[12]; 一种基于人脸位置、图像对比度、清晰度和明亮度的综合人脸图像质量评价方法被蒋刚毅等人提出[13]. 邹国锋等先对包含人脸的原始图像进行第一级评价, 再提取人脸有效区域进行第二级评价[14]. 多因素融合大方法易受各因素影响, 一个因素评价失误将导致整个评价出现偏差. 并且, 各因素难以被全面考虑, 且对人脸图像质量的影响权重难以确定. 另外, 需要分别进行多个因素的质量评估, 计算较为繁琐. 基于全局学习的方法是指通过训练学习自动将各个因素进行融合得到质量分数. Ozay 使用一个贝叶斯网络来拟合图像特征和质量分数之间的关系[15]. Deng[16]利用三个人脸图像存在明显差异的数据集, 采用学习排序的方法进行质量评估, 效果显著. 基于全局学习的方法操作更为简便, 特征自动融合, 质量评估结果相对更为可靠.

CNN(卷积神经网络) 具有强大的特征学习能力, 通过端到端的训练, 逐层得到由简单到抽象的特征, 在识别、检测、跟踪等各个领域取得了许多突破性的成果. Kang 等人[17]便提出了一个简单的 CNN 结构用作图像质量评估, 并取得了非常好的效果. Liu 等人[18]首先利用 VGG 网络提取大量图像特征, 再通过稀疏字典学习选择有效的特征, 最后用 SVR 回归预测人脸图像质量分数. 在自制监控数据集上实验发现, 此方法能够挑选到高质量的人脸图像, 从而提高识别率. 但学习步骤较为繁琐, 且每一个步骤都是独立的, 没有端到端的统一调整.

2 基于 CNN 的监控视频中质量评估

监控视频中的人脸图像具有光照、姿态、表情等多种变化, 传统的图像质量评估方法难以对其进行全方面的评估. CNN 的泛化能力显著优于其他方法, 因此本文提出了一种基于 CNN 的监控视频中人脸图像质量评估方法.

2.1 网络结构

网络模型来源于 Alexnet 模型[19], 并对其进行了改进. Hariharan 等人证明, 使用 CNN 网络中多个层次的特征更有利于完成特定任务[20]. 而且, 低水平特征对于图像质量有很重要的影响. 而 CNN 的中间卷积层包含了大量的边缘、几何等低水平特征. 于是, 将 Alexnet 的

中间卷积层与全连接层进行连接, 从而融合简单特征与抽象特征.

图 1 是 Alexnet 的基本结构, 包括 5 个卷积层和 3 个全连接层. 为了实现多尺度特征融合, 把 Conv2、Conv3、Conv4 的输出特征图分别先进行池化, 并裁剪至与 Conv5 输出尺寸一致, 再连同 Conv5 一起与 fc6 层连接. 最后, SVR 函数被选为损失函数. 如图 2 所示.

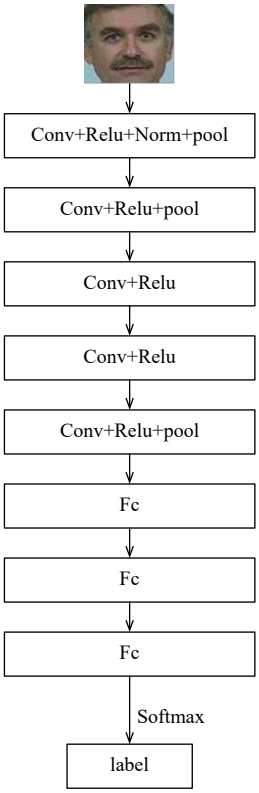


图 1 Alexnet 网络结构

具体参数为:

输入: 227×227 的三通道图像;

Conv1: 96 个 11×11 的卷积核, 步长为 4; 最大池化, 核大小为 3, 步长为 2;

Conv2: 256 个 5×5 的卷积核, 步长为 1, 填充值为 2; 最大池化, 核大小为 3, 步长为 2;

Conv3: 384 个 3×3 的卷积核, 步长为 1, 填充值为 1;

Conv4: 384 个 3×3 的卷积核, 步长为 1, 填充值为 1;

Conv5: 256 个 3×3 的卷积核, 步长为 1, 填充值为 1; 最大池化, 核大小为 3, 步长为 2;

新增加两个 pool 层: 最大池化, 核大小为 3, 步长为 2;

Fc6 和 Fc7 分别输出 4096 维特征, Fc8 输出 1 维的质量分数.

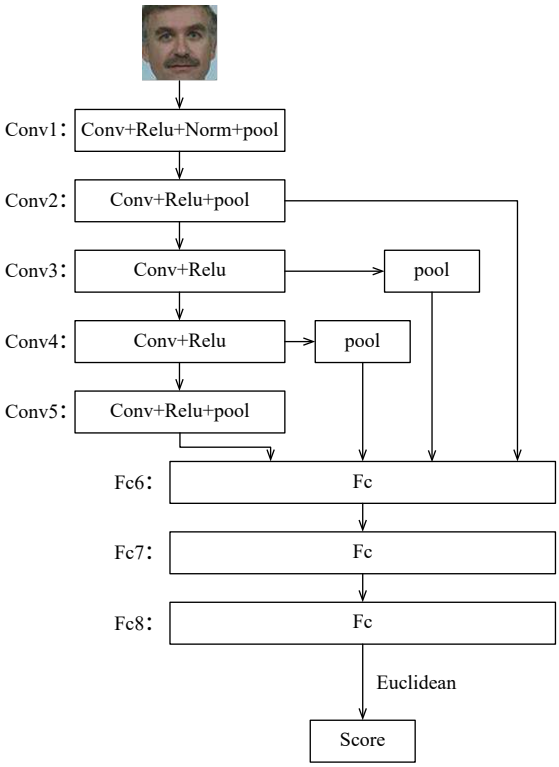
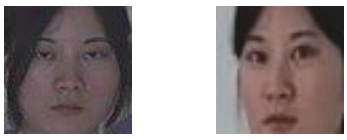


图 2 本文网络结构

2.2 结合人脸识别算法的训练样本标注

传统的人脸图像质量评估方法使用了人类视觉系统的先验知识, 于是得到的人脸图像质量分数只与人的视觉感受相一致. 但实际上, 评估人脸图像质量时, 应考虑到人脸识别系统本身的运行机制, 将质量分数与识别算法联系起来. 例如, 图 3(a) 是一张光照偏暗的正脸图像, 图 3(b) 是一张光照均匀的侧脸图像. 对于一个对人脸姿态鲁棒性很好的识别算法来说, 图 3(b) 的质量分数自然要高于图 3(a). 而若识别算法对光照并不敏感, 则图 3(a) 比图 3(b) 更适合用于识别.



(a) 光照偏暗的正脸 (b) 光照正常的侧脸

图 3 不同变化因素的人脸图像

因此, 本文采用具体的人脸识别算法来对人脸图像进行质量分数的标定. 在这里, 我们选择 VGGFace 模型[21]加余弦相似度的识别算法, 以余弦相似度作为质量分数. 同时, 这种标定方式也解决了训练 CNN 模

型所需大量带标签训练样本的问题. 具体操作步骤如图 4 所示.

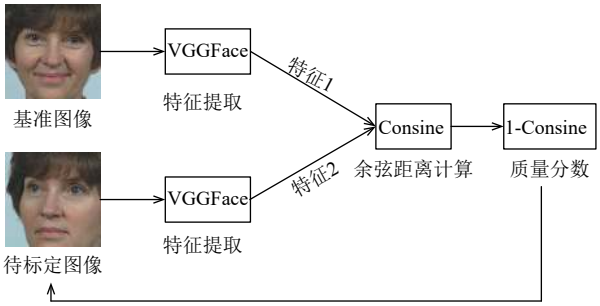


图 4 结合识别算法的人脸图像标定过程

3 实验与分析

3.1 实验数据集

Color FERET 数据集[22]是由美国 Feret 项目组收集的人脸数据库, 包含 994 个类别多姿态、光照的人脸图像, 共 11 338 张. 其可分为三部分: fa 子集是统一光照的正脸图像集; fb 子集也是统一光照的正脸图像, 但其表情与 fa 集有差距; 其他是各种姿态变化的人脸图像. PIE 人脸数据库[23]由美国卡耐基梅隆大学创建, 在严格控制条件下采集了 41 368 张包含姿态、光照变化的人脸图像.

另外, 本文通过学校监控系统平台, 收集到实际场景下的监控视频数据集. 数据集中包含 183 个类别, 每类别有 100 张左右的人脸图像, 包括姿态、光照、表情、分辨率等多种变化因素. 平均尺寸在 56×56 左右. 图 5 其中的一些人脸图像示例.

图 5 实际监控视频中的人脸图像样例

3.2 人脸图像质量评估实验

为了验证本文方法对人脸图像质量评估的准确性, 综合 Color FERET 和 PIE 数据集, 进行了以下实验.

3.2.1 数据准备

为了增加训练样本数量以及样本多样性, 先对数据集进行增强. 将 fb 子集中的每张图片做如下变换: (1) 水平与竖直方向分别平移 0, ±2, ±4, ±6, ±8 个像素; (2) 平面内正逆时针分别旋转 0°, ±10°, ±20°, ±30°; (3) 分别在 0.7~1.3 不同尺度截取人脸; (4) 先将人脸图像缩放为原来的 0.25, 0.75, 1.25, 1.75 倍, 再恢复到原尺寸. 这四种变换为数据库分别加入了对齐误差和清晰度变化. Color FERET 中包含有 15°, 22.5°, 45°, 67.5°不同姿态的人脸图像, 被用来评估姿态对图像质量的影响. 而针对光照对人脸图像质量的影响, 选取了 PIE 中光源角度为 54°~67°的人脸图像. 最后, 形成一个大约包含 35 万张不同光照、姿态、分辨率的人脸图像数据集. 部分人脸图像如图 6~图 8 所示.

图 6 Color FERET 中人脸姿态变化样本示例

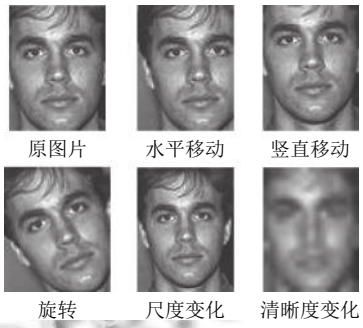


图 7 Color FERET 中 fb 子集人脸图像变化示例

图 8 PIE 中人脸图像光照变化样本示例

用上文提到的标定方法对数据集进行标定. 对于 Color FERET 数据集, fa 被选为基准图像, 而 PIE 数据集中正面光照条件下采集的图像被选为基准图像. 图 9 展示了部分标定结果. 由图可知, 被水平或竖直移动变换的图像质量分数依然较高, 而被旋转或尺度缩小的图像质量分数变低. 这与本文所采用的人脸识别算法对于人脸水平或竖直移动敏感度较小, 而对旋转及尺度变化较为敏感的特性相符合. 从而说明了本文的图像质量分数标定方法与识别算法的本身特性联系紧密.

图 9 结合识别算法的样本标注示例

3.2.2 模型训练

将数据集分成训练集、验证集和测试集三部分. 训练前, 对所有图片进行归一化, 并统一尺寸为 227×227. 训练时, 采用了调优的方法. 初始参数模型是已在 Imagenet 数据库训练好的 Alexnet 模型[21], 然后重新学习新的全连接层. 设置新的学习率为 0.0001, batchsize 为 64, 采用 SGD 优化方法, 迭代了 12 000 次, 大约 6 个 epoch, 最终达到收敛.

3.2.3 结果展示及分析

将训练好的人脸图像质量评估模型在测试集上进行测试, 部分测试图像的质量分数如图 10 所示. 由图可知, 对于姿态、光照、清晰度或光照变化, 我们模型给出的质量分数都能有效进行区分.

(a) 不同表情、姿态与清晰度的人脸图像质量评估结果

(b) 不同光照的人脸图像质量评估结果

图 10 人脸图像质量评估结果

图像质量评估算法常见评价指标有 LCC (线性相关系数) 和 SROCC (秩相关系数). LCC 描述算法评价值与参考值之间的相关性, 从而衡量了算法预测的准确性. SROCC 衡量算法预测的单调性. 在测试集上计算这两项指标, 并与只考虑单一因素的评估方法进行比较, 如表 1 所示. 从表中可以发现, 基于 CNN 学习的方法比根据单一因素评估的方法效果好很多, 而本文对 Alexnet 进行改进后, 效果更加提升.

表 1 Color FERET 测试集上的 LCC 和 SROCC

方法	LCC	SROCC
清晰度	0.0027	0.0014
对比度	0.3877	0.3689
对称性	0.4618	0.4906
Pan 的方[24]	0.7784	0.7991
Alexnet	0.8521	0.8739
改进的 Alexnet	0.8634	0.9124

3.3 基于质量评估的监控视频中人脸验证实验

为了验证本文方法能够提高监控视频中人脸识别系统识别率, 在自制的监控视频数据集上进行了实验.

首先, 用训练好的模型评估监控视频中的人脸图像质量, 结果如图 11 所示. 虽然整体质量分数偏低 (这与监控视频中人脸图像本身质量低也是相符的), 但依然能对姿态 (图 11(a)、(d)、(e)、(f))、清晰度 (图 11(a)、(c))、表情 (图 11(a)、(b)) 变化进行区分. 另外, 由于人脸检测算法存在误差, 检测出一些非人脸图像. 但是其质量分数非常低, 可以通过质量分数将其剔除 (图 11(g)、(h)).

图 11 监控视频中人脸图像质量评估结果

然后, 将质量评估模块加入人脸识别系统中, 简单流程图如图 12 所示. 先对监控视频进行人脸检测及跟踪, 得到同一个人的一连串人脸图像, 再对这些人脸图像进行质量评估并由高到低进行排名, 分别选出其中质量排名为 1, 2, 4, 8, 16 以及所有人脸图像进行后续人脸识别, 识别准确率如图 13 所示. 由图可看出, 对于改进后的 Alexnet, 当选择质量排名为前 8 的人脸图像进行人脸识别时, 识别率最高达到 91%. 若不进行质量

评估, 而将所有人脸图像全部用于识别, 识别率只有64%. 从而证明了本文提出的质量评估方法能提高监控视频中人脸识别准确率.

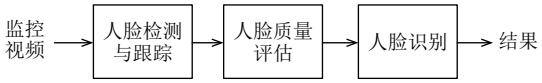


图 12 加入质量评估模块的人脸识别系统

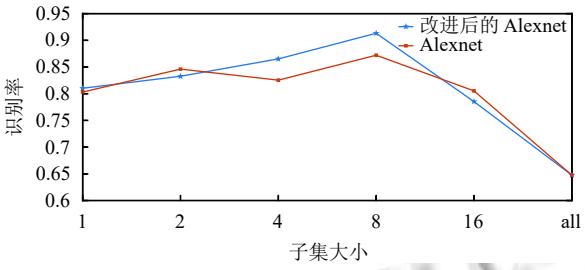


图 13 选择不同质量排名的人脸图像进行识别的结果

4 结论与展望

本文提出了一种基于 CNN 的监控视频中人脸图像质量评估方法. 主要有两点: 一是传统图像评估方法大多只考虑到部分因素对图像的影响, 且融合过程需人为设计. 本文通过将 Alexnet 的中间卷积层与全连接层连接, 自动融合多尺度特征进行图像质量评估; 二是网络训练需要大量带标签样本,人工进行标定耗时耗力, 且标定结果与人的视觉系统相一致, 而脱离了实际人脸识别系统. 因此采用结合人脸识别算法的方法自动标定. 实验证明, 本文方法能够对姿态、表情、光照、清晰度变化引起的图像质量变化给予准确的评估, 筛选出高质量的人脸图像, 提高识别准确率.

参考文献

1 Zhao W, Chellappa R, Phillips PJ, *et al*. Face recognition: A literature survey. ACM Computing Surveys (CSUR), 2003, 35(4): 399–458. [doi: 10.1145/954339.954342]

2 Wiskott L, Krüger N, Kuiger N, *et al*. Face recognition by elastic bunch graph matching. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7): 775–779. [doi: 10.1109/34.598235]

3 Blanz V, Vetter T. Face recognition based on fitting a 3D morphable model. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(9): 1063–1074. [doi: 10.1109/TPAMI.2003.1227983]

4 Wright J, Yang AY, Ganesh A, *et al*. Robust face recognition via sparse representation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(2): 210–227. [doi: 10.1109/TPAMI.2008.79]

5 Grother P, Tabassi E. Performance of biometric quality measures. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(4): 531–543. [doi: 10.1109/TPAMI.2007.1019]

6 Grother PJ, Quinn GW, Phillips PJ. Report on the evaluation of 2D still-image face recognition algorithms. NIST Interagency Report, 2010, 7709: 106. [doi: 10.6028/NIST.IR.7709]

7 Huang GB, Ramesh M, Berg T, *et al*. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. University of Massachusetts Amherst Technical Report. Amherst, Massachusetts: University of Massachusetts Amherst, 2007.

8 Fan HQ, Cao ZM, Jiang YN, *et al*. Learning deep face representation. arXiv preprint arXiv:1403.2802, 2014.

9 魏政刚, 袁杰辉, 蔡元龙. 图象质量评价方法的历史、现状和未来. 中国图象图形学报, 1998, 3(5): 386–389. [doi: 10.3969/j.issn.1006-8961.1998.05.008]

10 Sang JT, Lei Z, Li SZ. Face image quality evaluation for ISO/IEC standards 19794-5 and 29794-5. Proceedings of the 3rd International Conference on Advances in Biometrics. Alghero, Italy. 2009. 229–238.

11 Nasrollahi K, Moeslund TB. Face quality assessment system in video sequences. Proceedings of the First European Workshop on Biometrics and Identity Management. Roskilde, Denmark. 2008. 10–18.

12 Rúa EA, Castro JLA, Mateo CG. Quality-based score normalization and frame selection for video-based person authentication. Proceedings of the First European Workshop on Biometrics and Identity Management. Roskilde, Denmark. 2008. 1–9.

13 蒋刚毅, 黄大江, 王旭, 等. 图像质量评价方法研究进展. 电子与信息学报, 2010, 32(1): 219–226.

14 邹国锋, 傅桂霞, 李震梅, 等. 融合二级评价指标的人脸图像质量评价方法. 山东大学学报 (工学版), 2016, 46(2): 6–13.

15 Ozay N, Tong Y, Wheeler FW, *et al*. Improving face recognition with a quality-based probabilistic framework. Proceedings of 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. Boston, MA, USA. 2009. 134–141.

16 Chen JS, Deng Y, Bai GC, *et al*. Face image quality assessment based on learning to rank. IEEE Signal

Processing Letters, 2015, 22(1): 90–94. [doi: 10.1109/LSP.2014.2347419]

17 Kang L, Ye P, Li Y, *et al*. Convolutional neural networks for no-reference image quality assessment. Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA. 2014. 1733–1740.

18 Liu GR, Xu Y, Lan JP. No-reference face image assessment based on deep features. Proceedings Volume 9971, Applications of Digital Image Processing XXXIX. San Diego, CA, USA. 2016. 9971: 99711S.

19 Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, NV, USA. 2012. 1097–1105. [doi: 10.1145/3065386]

20 Hariharan B, Arbeláez P, Girshick R, *et al*. Hypercolumns for object segmentation and fine-grained localization. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA. 2015. 447–456.

21 Parkhi OM, Vedaldi A, Zisserman A. Deep face recognition. Proceedings of British Machine Vision Conference. Swansea, UK. 2015. 6.

22 Phillips PJ, Moon H, Rizvi SA, *et al*. The FERET evaluation methodology for face-recognition algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(10): 1090–1104. [doi: 10.1109/34.879790]

23 Sim T, Baker S, Bsat M. The CMU pose, illumination, and expression database. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(12): 1615–1618.

24 Pan CH, Ni BB, Xu Y, *et al*. Recognition oriented facial image quality assessment via deep convolutional neural network. Proceedings of the International Conference on Internet Multimedia Computing and Service. Xi'an, China. 2016. 160–163.