

# 监控视频下的人脸图像质量分析技术<sup>①</sup>



张 雪, 郭佳昕

(北京工业大学 软件学院, 北京 100124)  
通信作者: 张 雪, E-mail: zxbjut@163.com

**摘 要:** 监控场景下人脸图像质量分析技术的研究具有重要意义. 由于监控视频采集到人脸模糊、头部角度不正、被其他物体遮挡等的低质量图像进入识别系统会造成识别准确率下降. 为了解决上述问题, 通过实验, 研究了监控场景下影响图像质量的两个重要因素: 人脸角度、图像清晰度. 设计了基于聚类的人脸图像质量分析算法, 提出了人脸图像质量分数的计算公式, 实验表明该技术能够有效过滤监控视频下采集到的低质量图像, 进而提高人脸识别系统的准确率.

**关键词:** 监控视频; 人脸角度; 人脸识别; 聚类; 人脸图像; 图像质量分析

引用格式: 张雪, 郭佳昕. 监控视频下的人脸图像质量分析技术. 计算机系统应用, 2022, 31(9): 313-318. <http://www.c-s-a.org.cn/1003-3254/8696.html>

## Quality Analysis Technology for Face Images in Surveillance Videos

ZHANG Xue, GUO Jia-Xin

(School of Software, Beijing University of Technology, Beijing 100124, China)

**Abstract:** The research on quality analysis technology for face images in surveillance scenes is of great significance. Since the low-quality images collected from surveillance videos have blurred faces and incorrect head angles and are subjected to occlusion by other objects, the input of them into the recognition system can result in lower identification accuracy of the system. To solve the above problems, this work studies two important factors that affect image quality in surveillance scenes through experiments, namely, the face angle and image clarity. Thus, a quality analysis algorithm for face images based on clustering is designed, and a calculation method for scoring the quality of face images is proposed. The experiment proves that the technology can effectively filter the low-quality images collected in the surveillance videos and improve the accuracy of the face recognition system.

**Key words:** surveillance video; face angle; face recognition; clustering; face images; image quality analysis

## 1 引言

### 1.1 研究背景

人脸识别系统目前被广泛应用在监控场景中, 从监控视频中快速并准确获取人脸身份信息对于公安破案等具有重要意义<sup>[1-3]</sup>. 人脸图像质量分析作为人脸识别过程中的关键一环, 不同场景选择不同的图像质量分析方法对识别的准确率有很大的影响, 特别是在监控场景下, 由于拍摄的人脸图像含有大量被其他物体遮挡、模糊、头部姿态不正以及无法判断人脸等的低

质量图像, 若这些图像进入识别系统会导致系统准确率下降, 严重影响识别效果<sup>[4,5]</sup>. 因此在监控场景中预先采用合理的图像质量分析方法对提高人脸识别系统准确率有重要应用价值.

### 1.2 研究现状

人脸图像质量分析作为图像质量分析的一部分, 在保留传统的图像质量分析方法基础上, 加入人脸特有的因素. 2009年, 国际标准化组织制定了关于人脸图像质量的 ISO/IEC 标准, 对多种参数做出了规定<sup>[6]</sup>. 基

① 收稿时间: 2021-12-14; 修改时间: 2022-01-12, 2022-01-28; 采用时间: 2022-02-15; csa 在线出版时间: 2022-06-07

于该标准国内外学者对于人脸图像质量分析做了很多研究. 整体研究方法分为两种: 基于深度学习的人脸图像质量分析和基于多因素融合的人脸图像质量分析. 前者从大量数据中学习获取人脸质量关注的特征并进行质量判断. 在提取特征向量后对数据进行学习, 质量好的样本离中心距离较近, 不好的样本离中心距离尽量远, 最后根据样本到中心的距离, 得到质量分数. 该方法2015年开始出现, 随后国外相继出现了使用 ResNet-50<sup>[7]</sup>、AlexNet<sup>[8]</sup>、GoogLeNet<sup>[9]</sup> 等训练人脸图像质量模型. 近两年出现的有 SER-FIQ 利用随机嵌入鲁棒性来估计人脸图像质量<sup>[10]</sup>. 而这种方法普遍存在一个问题就是没有针对具体的场景进行训练, 虽然最终得到的质量模型在识别系统上有较好的鲁棒性, 但若是使用在含有大量低质量图像为主导的视频监控场景中, 模型效果并没有较为突出的表现, 并且这样训练得到的模型无法适应特定场景, 一旦使用场景改变, 就需要重新训练, 耗费大量的时间. 后者通过分析单个的质量因素, 例如光照、清晰度、遮挡等, 对其分配权重进行加权融合, 最终得到一个质量分数, 该分数就是人脸图像的质量分数. Nasrollahi 等<sup>[11]</sup> 分别对人脸角度、图像亮度、锐度和分辨率逐一计算, 通过赋予这 4 个因素一定比例的权重, 得到图像的质量分数, 进而对人脸图像质量进行评价. Fourney 等<sup>[12]</sup> 为 人脸提取了 6 个特征: 人脸姿势、亮度、锐度、人体皮肤的存在、分辨率, 最终将这 6 个参数加权得到质量分数. 然而不同场景下指标的选取、权重的分配都会对质量分析结果产生影响.

鉴于上述研究缺陷和问题本文选取人脸角度 (*pitch*、*roll*、*yaw*) 和图像清晰度作为人脸图像质量分析的两个关键指标. 针对这两个指标分别采用合适的方法训练得到每个指标下的人脸图像质量分数, 并设计基于聚类的人脸图像质量分析算法, 对这两个指标分配权重进行加权融合, 获得人脸图像质量分数的计算公式. 最后将所研究的人脸图像质量分析技术应用人脸识别系统中, 证明该技术能够有效过滤监控场景中的低质量图像, 提高识别系统的准确率<sup>[13]</sup>.

## 2 基于聚类的人脸图像质量分析

本文分析过程主要包含 3 大模块: 人脸角度计算、图像清晰度计算、基于聚类的人脸图像质量分析系统. 最后得到人脸图像质量分数. 算法流程图如图 1.

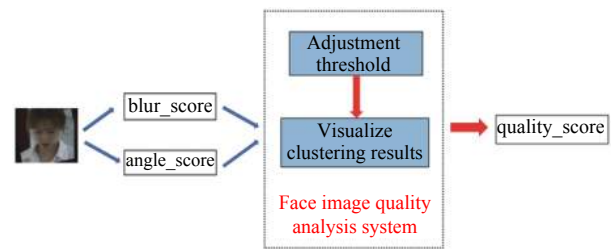


图 1 算法流程

首先分别训练人脸角度模型和图像清晰度模型, 得到对应的分数值, 之后设计基于聚类的人脸图像质量分析算法, 并设计聚类系统, 在前端调节两者的权重后可可视化聚类结果, 计算聚类纯度, 最终确定人脸图像质量的计算公式.

### 2.1 人脸角度

(1) 概述: 人脸角度指的是根据一张二维的人脸图像, 计算出这个人在实际三维空间中的面部朝向. 即输入一张人脸图片, 输出表示方位的 3 个旋转角度 (*pitch*, *yaw*, *roll*), 其中, *pitch* 表示俯仰角 (关于 *x* 轴的旋转角度), *yaw* 表示偏航角 (关于 *y* 轴的旋转角度), *roll* 表示翻滚角 (关于 *z* 轴的旋转角). 将一张二维图像的特征映射到三维模型上效果如图 2 所示<sup>[14,15]</sup>.



图 2 人脸角度的三维效果图

(2) 本文使用方法: 由于人脸角度估计与头部姿态估计的输入输出值相同, 本文中使用了头部姿态估计的经典方法, 利用先分类后回归的思想, 使用 CNN 预测 *pitch*、*yaw*、*roll* 三个值. 如图 3 所示. 使用 ResNet 作为主干网络, 分别全连接 3 个 fc 层, 每个层单独预测. 首先对 3 个欧拉角分类, 将角度范围从  $[-99, +99]$  以 3 为间隔划分, 划分成 66 个区间. 分类得到的结果做回归, 即 Softmax 将其映射为概率值, 根据 Softmax 定义, 此时所有的概率值和为 1, 求出期望后, 网络的输出可以被映射到  $[0, 99]$  这个区间范围内, 乘以 3 减去 99, 这个区间范围就被映射到了  $[-99, +99]$  这个区间范

围<sup>[16,17]</sup>. 然后计算回归的 Loss, 使用 MSE Loss 与 BCE Loss 按照一定权重加权求和, 对最终的 Loss 梯度反向, 就完成了整个过程, 每个欧拉角的 Loss 定义为:

$$L = H(y, \hat{y}) + \alpha \cdot MSE(y, \hat{y}) \quad (1)$$

其中,  $H$ 表示交叉熵函数,  $\alpha$ 表示权重系数,  $MSE$ 为平方误差损失函数.

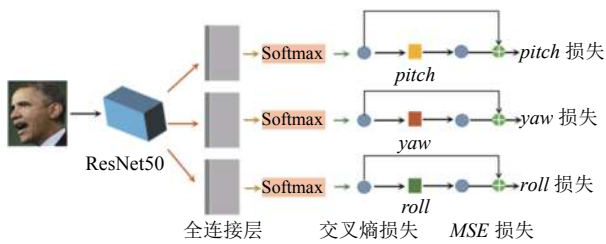


图3 人脸角度算法流程图

(3) 模型训练: 使用 ALFW2000 数据集, 包含 2000 多张不同的头部姿势图像, 训练集和测试集按照 7:3 的比例划分. 主要参数有:

- 1) 学习率 0.00001;
- 2) 一次训练所取的样本数即 batch size 为 16;
- 3) epochs 为 20;
- 4) 权重系数  $\alpha$  设置为 1 时, 损失最小.

## 2.2 图像清晰度

(1) 概述: 图像清晰度和模糊度是相对的, 一张人脸图像清晰度越高, 则模糊度越低. 反之如此. 因此使用图像模糊度算法进行研究. 作为质量评价中的重要因素之一, 其方法主要分为两种情况: 一种是有参考算法, 即根据已有的图像, 来判断当前图像是否模糊; 另一种是无参考的算法, 判断图像是否模糊; 常见的图像清晰度算法主要有: Brenner 梯度函数, 它是最简单的梯度评价函数, 它的原理是计算相邻两个像素灰度差的平方; Tenengrad 梯度函数采用 Sobel 算子分别提取水平和垂直方向的梯度; Laplacian 梯度函数与 Tenengrad 梯度函数基本一致, 用 Laplacian 算子替代 Sobel 算子即可; SMD (灰度方差) 函数当完全聚焦时, 图像最清晰, 图像中的高频分量也最多, 故可将灰度变化作为聚焦评价的依据, 常用的很多, 但都是在最简单的基础上进行优化和变换.

(2) 本文使用方法: 由于摄像头下采集到的图片参差不齐, 将人脸图像的清晰度作为人脸识别其中一项关键的指标很有必要. 本文使用 OpenCV 结合拉普拉斯方差算法计算人脸的模糊度进而得到图像清晰度分

数. 拉普拉斯算子是各向同性微分算子, 主要用来测量图像的三阶导数, 突出图片中强度快速变化的区域, 一个二维图像函数的拉普拉斯变换是各向同性的二阶导数<sup>[18]</sup>. 定义为:

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \quad (2)$$

其中,  $\nabla^2$ 表示拉普拉斯算子,  $f$ 表示空间标量函数,  $x$ 和  $y$ 表示图像像素的两个方向. 对图片进行灰度处理后, 将图片中的某一通道用拉普拉斯掩模做卷积运算后计算方差<sup>[19]</sup>. 基于图像模糊度进行训练后, 得到每个图像对应的分数值, 值越小图像越清晰, 0 表示非常清晰, 1 表示非常模糊.

## 2.3 基于聚类的人脸图像质量分析系统

本文设计了基于聚类的人脸图像质量分析算法, 该算法在传统的人脸识别流程的基础上加入图像质量分析和聚类, 如图 4.

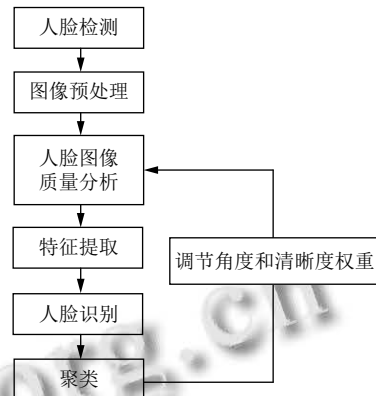


图4 基于聚类的人脸图像质量分析算法

为了方便可视化聚类结果, 开发了基于聚类的人脸图像质量分析系统, 通过调节图像清晰度  $blur$  和人脸角度的权重, 对人脸图像聚类, 聚类的纯度越高则当前阈值下的人脸识别准确率越高. 纯度的计算公式如下:

$$Purity = m_i / m \quad (3)$$

其中,  $m_i$ 表示为同一人的组数,  $m$ 表示总组数, 本文  $m$  固定为 200. 纯度可表示为聚类的准确率, 即人脸识别的准确率.

系统主要涉及两大部分的功能:

1) 簇内聚类. 该模块的功能主要支持调节欧式距离, 计算各人脸图像到簇心之间的距离, 使用 K-means 聚类算法将相似的人聚在一类, 并随机可视化前 200 组聚类结果的人脸图像. 在每一个组图像下设置了



判断该组人脸图像是否是同一人的选择框,如图5所示.选择完成后即可计算出对应距离下的聚类纯度.



图5 簇内聚类结果

2) 簇间聚类.在进行完上述的簇内聚类后,已经筛选出聚类纯度较高的簇内距离,接着调节簇间的阈值,此处引入人脸图像质量的两个因素,图像清晰度 $blur$ 和人脸角度.通过调节这两个的权重值,将筛选后的结果随机分为两组对比,观察这两组的纯度,纯度较高的清晰度和角度阈值即为最终分配的权重.图6是簇间聚类结果.



图6 簇间聚类结果

### 2.4 人脸图像质量分数

一个较为完整的人脸识别流程要经过以下几个步骤:人脸检测、人脸对齐、提取特征、人脸识别.本文在进行人脸检测后,先根据检测到的人脸关键点进行人脸对齐校准(face alignment)操作,校准之后的人脸 $roll$ 就全部变成0,因此在进行质量分析时,暂时不考虑 $roll$ 的分数和权重.下面重点介绍人脸质量分数的计算流程:

步骤1.运行完人脸角度模型,得到人脸的俯仰角、偏转角以及翻滚角,将角度转化为分数值如下:

$$a = 10.0/3.0 \tag{4}$$

$$b = -7.0/3.0 \tag{5}$$

$$\begin{cases} yaw_{score} = a \times \cos \frac{yaw}{180\pi} + b \\ pitch_{score} = a \times \cos \frac{pitch}{180\pi} + b \\ angle_{score} = 0.5 \times yaw_{score} + 0.5 \times pitch_{score} \end{cases} \tag{6}$$

其中, $a$ 、 $b$ 为参数. $yaw_{score}$ 是人脸图像 $yaw$ 角转化后对应的分数值, $pitch_{score}$ 是人脸图像 $pitch$ 角转化后对应的分数值, $angle_{score}$ 是人脸角度的分数值.一张人脸角度最正的分值是 $pitch$ 、 $roll$ 全部为0,此时基于人脸角度的图片分析,该图像质量是高的.

步骤2.运行清晰度模型并得到对应的 $blur$ 分数值.

步骤3.簇内聚类.采用本文研究的基于聚类的人脸图像质量分析技术,对人脸识别结果聚类,选择合适的聚类距离,观察聚类效果.

步骤4.簇间聚类.调节前两步得到的 $angle_{score}$ 和 $blur$ 权重阈值,并计算聚类纯度.

步骤5.通过实验分析确定权重分别为0.6和0.4,最终得到图像质量分数 $quality_{score}$ .计算公式如下:

$$quality_{score} = 0.6 \times blur_{score} + 0.4 \times angle_{score} \tag{7}$$

其中, $blur_{score}$ 是关于图像清晰度的分数值, $angle_{score}$ 是关于人脸角度的分数值.

由表1可见,在融入基于聚类的人脸图像质量分析算法后,同一批数据集在簇内距离为0.45、0.5、0.6时调节 $blur$ 和 $angle$ 的权重计算得出的聚类纯度均在0.6和0.4时最高,因此当为 $blur$ 和 $angle$ 分配权重分别为0.6和0.4时,对识别准确率的影响最大.

表1 不同参数下人脸图像的聚类纯度

距离	清晰度权重	角度权重	纯度 (%)
0.45	0.3	0.7	77
	0.4	0.6	80
	0.5	0.5	83
	<b>0.6</b>	<b>0.4</b>	<b>88.5</b>
	0.7	0.45	84
0.50	0.3	0.7	79.5
	0.4	0.6	83
	0.5	0.5	85
	<b>0.6</b>	<b>0.4</b>	<b>89.5</b>
	0.7	0.3	86
0.60	0.3	0.7	85
	0.4	0.6	87.5
	0.5	0.5	89.5
	<b>0.6</b>	<b>0.4</b>	<b>93</b>
	0.7	0.3	91

### 3 实验结果和分析

#### 3.1 数据集

为验证该研究技术的有效性,本文使用某公司2021年2月到2021年5月之间采集到的监控视频,对视频进行抽帧,抽帧方法选取常用的帧差法,即对相邻两帧作差分运算来获得运动目标轮廓.抽出的图像在进行简单的检测和识别后,最终包含有6 000 000张人脸图像,为了进行实验对比,得到更客观准确的实验结果,本文将数据分为10个批次,每个批次含有600 000张图像,其中训练集470 000张,测试集130 000张进行验证.

#### 3.2 实验结果

对一批数据设置不同的阈值并对该阈值下的结果聚类可视化,随机展示200组聚类结果,计算聚类纯度.通过研究发现,加入人脸图像质量分析算法后,图像的纯度相比没有质量过滤的纯度有很大的提高.如表2.

表2 不同聚类距离下的人脸图像质量分析结果

距离	方法	纯度(%)
0.45	质量分析前	73.5
	质量分析后	86.5
0.50	质量分析前	78.5
	质量分析后	89.5
0.60	质量分析前	84
	质量分析后	93

由表2可见,在对一批数据测试后分别对比未加入人脸图像质量分析算法和本文研究的基于聚类的人脸图像质量分析算法的纯度结果,本文研究的技术使聚类纯度平均提高了11.67%,该方法在一批数据集上表现较好.

在10批数据上分别进行测试后,采用基于聚类的人脸识别技术明显优于未加入质量分析的情况,聚类纯度明显提升,因此本文得到的人脸图像质量计算公式具有一定的适用性.如图7、图8、图9分别是10批簇内距离在0.45、0.5、0.6时加入和未加入图像质量分析技术的结果.

#### 3.3 分析对比

由于本文研究的人脸图像质量分析技术的目的是提高人脸识别系统的准确率,此处的人脸识别系统采用某公司的门禁系统.在未加入人脸图像质量分析前人脸识别Top10准确率79.56%,采用本文技术,目前识别准确率已经达到95.98%.结果见表3.

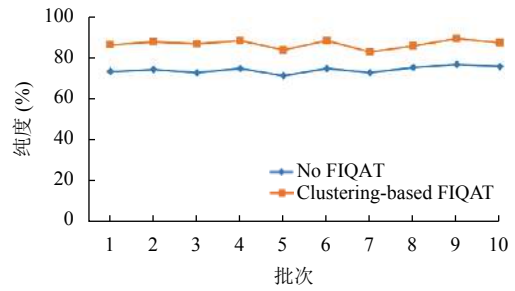


图7 簇内距离在 0.45 时质量分析前后对比

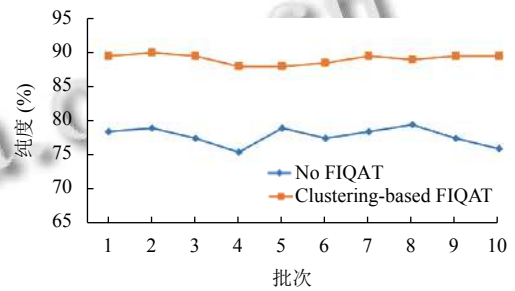


图8 簇内距离在 0.50 时质量分析前后对比

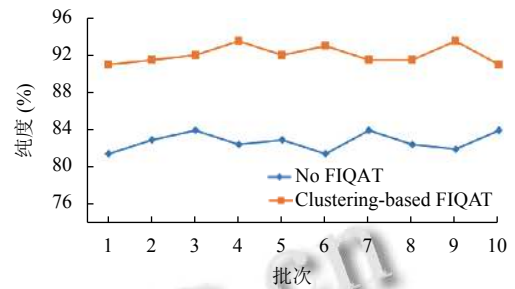


图9 簇内距离在 0.60 时质量分析前后对比

表3 加入质量分析前后人脸识别准确率比较(%)

方法	Top1	Top10
质量分析前	69.32	79.56
质量分析后	91.27	95.98

将本批从监控场景下获取到的数据集在基于AlexNet、GoogLeNet、CNN的人脸质量模型上测试后分别计算Top1和Top10,虽然这些模型在公开数据集上有较好的表现,但在监控场景中,采用本文提出的人脸图像质量计算公式应用在识别系统中得到的结果优于其他方法.如图10.

### 4 结论与展望

本文研究了基于监控视频下的人脸图像质量分析技术,该技术可有效过滤低质量图像进入识别系统,提高人脸识别准确率.该研究技术对于监控下的人脸识别

有很大意义。目前人脸图像质量分析的研究已经进入深度学习阶段,但在特定场景下深度学习的方法未必达到较好的效果,相信日后会有更好的方法解决此问题。

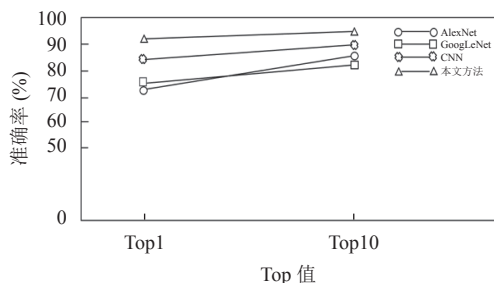


图10 不同人脸图像质量分析方法在识别系统的表现

### 参考文献

- Li XF. Development status and changes of the face recognition market. *China Security & Protection*, 2020, S1: 39–42.
- Zhang KP, Zhang ZP, Li ZF, *et al.* Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 2016, 23(10): 1499–1503.
- Jesorsky O, Kirchberg KJ, Frischholz RW. Robust face detection using the Hausdorff distance. *Proceedings of the 3rd International Conference on Audio- and Video-based Biometric Person Authentication*. Halmstad: Springer, 2001. 90–95.
- 夏盼盼. 行人检测中有效脸判定与检索模型研究 [硕士学位论文]. 南昌: 江西财经大学, 2017.
- Fang GN, Hu QH, Fang SY, *et al.* Face image quality valuation in video stream and its application in classroom attendance system. *Computer Applications and Software*, 2018, 35(10): 140–146, 251.
- Grother PJ, Quinn GW, Phillips PJ. Report on the evaluation of 2D still-image face recognition algorithms. NIST interagency report 7709. Gaithersburg: NIST, 2010. 106.
- He KM, Zhang XY, Ren Q, *et al.* Deep residual learning for image recognition. *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas: IEEE, 2015. 770–778.
- Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 2017, 60(6): 84–90.
- Szegedy C, Liu W, Jia YQ, *et al.* Going deeper with convolutions. *Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition*. Boston: IEEE, 2014. 1–9.
- Terhörst P, Kolf JN, Damer N, *et al.* SER-FIQ: Unsupervised estimation of face image quality based on stochastic embedding robustness. *Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle: IEEE, 2020. 5650–5659.
- Nasrollahi K, Moeslund TB. Face quality assessment system in video sequences. *Proceedings of the 1st European Workshop on Biometrics and Identity Management*. Roskilde: Springer, 2008. 10–18.
- Fourney A, Laganiere R. Constructing face image logs that are both complete and concise. *Proceedings of the 4th Canadian Conference on Computer and Robot Vision (CRV'07)*. Montreal: IEEE, 2007. 488–494. [doi: 10.1109/CRV.2007.20]
- 王畅, 崔荣一, 金璟璇, 等. 基于人脸聚类的视频中人脸图像优选方法的研究. *激光与光电子学进展*, 2019, 56(20): 201002.
- Schlett T, Rathgeb C, Henniger O, *et al.* Face image quality assessment: A literature survey. *ACM Computing Surveys*, 2021. [doi: 10.1145/3507901]
- 代欢欢, 谭生祥, 张旺. 头部姿态估计方法研究综述. *现代计算机*, 2021, 7: 130–134. [doi: 10.3969/j.issn.1007-1423.2021.07.028]
- Ruiz N, Chong E, Rehg JM. Fine-grained head pose estimation without keypoints. *Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. Salt Lake City: IEEE, 2018. 2074–2083.
- Bulat A, Tzimiropoulos G. Binarized convolutional landmark localizers for human pose estimation and face alignment with limited resources. *Proceedings of 2017 IEEE International Conference on Computer Vision*. Venice: IEEE, 2017. 3726–3734.
- 林增敏, 洪朝群, 庄蔚蔚. 融合人脸跟踪和聚类的人脸图像去重方法. *计算机科学*, 2020, 47(11A): 615–619.
- Pech-Pacheco JL, Cristobal G, Chamorro-Martinez J, *et al.* Diatom autofocusing in brightfield microscopy: A comparative study. *Proceedings of the 15th International Conference on Pattern Recognition*. Barcelona: IEEE, 2000. 314–317.

(校对责编: 牛欣悦)