

团体标准

T/CSMT XXXX—202X

微光场景图像质量评价方法构建要求

Building requirements for Image quality valuation method in low light scene

（送审稿）

（本稿完成时间：2025 年 4 月）

202X-XX-XX 发布

202X-XX-XX 实施

中国计量测试学会 发布

目 次

前 言 1

1 范围 2

2 规范性引用文件 2

3 术语、定义和缩略语 2

 3.1 术语和定义 2

 3.2 缩略语 2

4 图像质量评价方法总体要求 3

5 图像质量评价方法构建要求 3

 5.1 构建流程 3

 5.2 评测方案建模 3

 5.3 样本库构建 4

 5.4 算法模型训练 5

 5.5 评测方案封装 5

6 图像质量测试方法 6

 6.1 评价对象图像要求 6

 6.2 测试步骤 6

 6.3 结果评价 6

附 录 A （资料性） 人脸图像质量评价方法应用 7

 A.1 人脸图像质量评测方案 7

 A.2 人脸图像质量指标计算方法 7

 A.3 人脸图像质量评价工具封装 9

 A.4 人脸图像质量测试方法 10

前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别这些专利的责任。

本文件由中国测试学会提出并归口。

本文件起草单位：中国计量科学研究院、深圳中国计量科学研究院技术创新研究院、华为技术有限公司、上海交通大学、中移（杭州）信息技术有限公司

本文件主要起草人：徐英莹、戴列峰、陈赤、翟广涛、余峰、王鹏飞、李博文、李林芝、许一骅、刘松、汪艳

本文件为首次发布。

微光场景图像质量评价方法构建要求

1 范围

本文件规定了微光场景下的图像主观感受质量的量化评价方法的构建要求,包括图像质量评价方法总体要求,图像质量评价方法构建要求和图像质量测试方法。

本文件对图像采集设备最终输出的图像质量进行评价,用于量化人眼对图像质量的主观感受,不涉及图像采集设备的成像方案(包括硬件实现或软件增强等)。

本文件适用于构建微光场景下图像质量的量化评价方法,以及测试方法,可应用于图像采集设备的设计、开发、验收等。在正常光照度场景下的图像质量评价方法的构建和测试也可参照执行。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中,注日期的引用文件,仅该日期对应的版本适用于本文件;不注日期的引用文件,其最新版本(包括所有的修改单)适用于本文件。

SJ/T 11900.1 超高清视频图像质量 第1部分:主观评价

3 术语、定义和缩略语

3.1 术语和定义

3.1.1

微光场景 low light scene

指环境光照度低于10lux的室内和室外场景。

3.1.2

目标对象图像质量 target object image quality

指在不同场景的抓拍图像中主要关注目标的图像质量,如安防监控场景的人脸目标、交通管理场景的车牌目标等,根据此类目标对象的场景需求给出的符合人眼主观评价感受的质量量化结果。

3.2 缩略语

下列缩略语适用于本文件。

AGGD: 非对称广义高斯分布 (Asymetry Generalized Gaussian Distribution)

BRISQUE: 无参考空间域图像质量评估算法 (Blind Referenceless Image Spatial Quality Evaluator)

CPBD: 模糊检测的累积概率 (Cumulative Probability of Blur Detection)

DoG: 高斯差分图 (Difference of Gaussians)

GGD: 广义高斯分布 (Generalized Gaussian Distribution)

JNB: 正好可分辨模糊 (Just Noticable Blur)

MOS: 主观平均分数 (Mean Opinion Score)

- MSCN: 平均减去对比度归一化系数 (Mean Substracted Contrast Normalized Coefficients)
- NSS: 自然场景统计 (Natural Scene Statistics)
- SROCC: 斯皮尔曼等级相关系数 (Spearman Rank-Order Correlation Coefficient)
- SSCQE: 单刺激连续质量评价法 (Single Stimulus Continuous Quality Evaluation)
- SVR: 支持向量回归 (Support Vector Regression)

4 图像质量评价方法总体要求

- 图像质量评价方法应满足如下要求:
- a) 应能对图像的全图质量或图像中具体目标对象 (人脸、车牌等) 的图像质量进行评价, 评价结果符合人眼主观感受;
 - b) 图像质量评价结果应可量化, 输出打分结果, 可采用五分制、百分制等相对分值或绝对分值;
 - c) 图像质量评价方法应封装成工具包或应用软件, 具备操作界面和可执行功能。

5 图像质量评价方法构建要求

5.1 构建流程

图像质量评价方法构建流程如下图所示, 并满足如下要求:

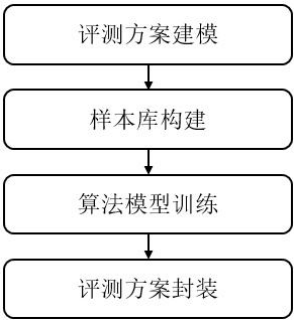


图 1 图像质量评价方法构建流程

- a) 支持对图像质量评价对象 (全图或人脸、车牌等目标对象) 进行各维度指标评测和整体质量指标评测, 评测维度覆盖评价对象的基本特征;
- b) 评测方案支持AI运算推理能力, 应用至少一种深度学习和机器学习框架;
- c) 用于训练评测方案的样本库数据应覆盖不同的评价对象类型和不同微光环境光照条件。

5.2 评测方案建模

5.2.1 评价对象分析

评测方案应具备对评价对象图像质量指标的可解释性。应根据评价对象的具体图像特点, 将整体图像质量指标分为多个维度指标分别进行评价, 各维度指标评测结果加权融合得到评价对象的整体评测结果。其框架如下图所示。

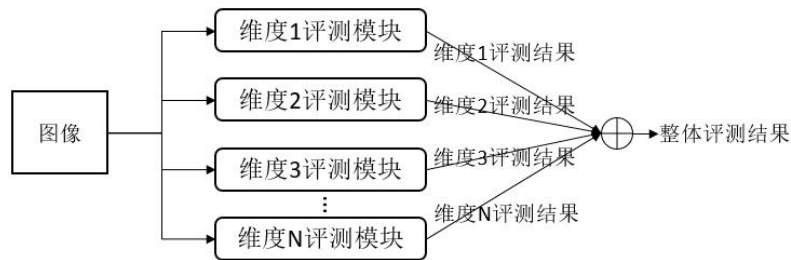


图 2 图像质量评测方案框架

图像质量可参照SJ/T 11900.1中的相关要求，设置清晰度、图像噪声、白平衡、色彩饱和度等维度。人脸图像质量评测方案见附录A.1。

5.2.2 维度评测方案

各个维度指标采用传统计算机视觉算法与深度学习算法相结合的方式进行评测，通过加权融合得到该维度的评测结果，如下图所示，并满足如下要求。

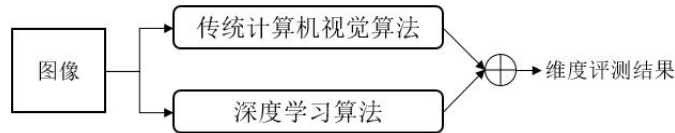


图 3 维度指标评测方案框架

- 传统计算机视觉算法：根据各维度的特殊性选择合理的传统计算机视觉算法，输出结果作为深度学习算法输入辅助推理或是与深度学习算法输出加权得到最终维度评测结果；
- 深度学习算法：应具备一定的普适性，宜选取ResNet为代表的卷积神经网络结构和以Swin Transformer^[2]为代表的Transformer网络结构，如Swin-B、ResNet-152等模型结构。模型的参数量应不小于20M，计算量FLOPs应不小于5G。

人脸图像质量各维度评价方案见附录A.2。

5.3 样本库构建

5.3.1 样本库数据要求

样本库应满足如下要求：

- a) 测试数据不应小于10000张；
- b) 测试数据应覆盖不同厂家设备、不同设备型号、不同评价对象类型；
- c) 测试数据应覆盖0.1lux至10lux的不同微光环境照度，色温覆盖2600K至9000K色温每个照度和色温下的数据数量差异不超过10%。

5.3.2 主观打分要求

样本库测试数据应进行主观打分，打分人员应满足如下要求：

- a) 应覆盖专业人员和“无经验”观察者，总人数不少于20人（其中专业人员约占10%，“无经验”观察者包含青年、中年、老年人员）；
- b) 打分人员应进行系统性打分培训，测试合格后方可进行参与打分活动。

进行打分的全量数据应包含10%的重复图像，用于打分完毕后进行打分筛选。打分结果采用SSCQE方法进行筛选^[1]，舍弃与均值存在显著偏差和前后不一致的打分结果。筛选后选取所有评价人员的MOS作为主观评测分数，公式如下：

$$MOS = \frac{1}{M} \sum_{i=1}^M r_i$$

其中，M是观察者的数量， r_i 是第*i*个观察者对图像的评分。

5.4 算法模型训练

5.4.1 数据集

将样本库分为训练数据集、验证数据集与测试数据集进行深度学习算法迭代训练：

——训练数据集：用于训练深度学习模型，对模型中的参数进行迭代优化进而使模型输出得分逼近主观打分结果；

——验证数据集：用于训练过程中评估深度学习模型的当前性能，判断是否可终止训练；

——测试数据集：用于深度学习模型训练完毕后，对其进行性能测试评估。

5.4.2 权重拟合

模型训练宜使用预训练权重，冻结大部分参数，仅开放最后2~3层参数进行训练。使用各维度MOS分数及对应图像进行训练得到相应维度评测方案的算法模型。

应用于评价对象整体图像质量的各维度权重值使用最小二乘法进行拟合，其计算公式如下。

$$L = \sum_{i=1}^n (y_i - \sum_{j=1}^m \omega_j f_j(x_i))^2$$

其中n为训练集样本个数，m为评价对象维度个数， ω_j 为评价对象各维度权重值， $f_j(x_i)$ 为评价对象各维度评测分数， y_i 为样本主观整体评测分数。

该拟合得到的权重仅代表构建样本库时所选择的打分人员整体对于各维度在整体维度中所占比重的认知，可根据场景需求进行选择修改。

5.4.3 评测性能要求

图像质量评测方案通过SROCC进行性能评价，其计算公式如下：

$$SROCC = 1 - \frac{6 \sum_{i=1}^N d_i^2}{N(N^2 - 1)}$$

SROCC对两个目标数组（即主观打分数组和算法打分数组）的秩次大小作线性相关分析，取值为-1~1，等于1时，表明两个数组的先后顺序完全一致。N表示样本数量， d_i 表示第*i*幅图像主观质量分数排名与算法质量分数排名的差值。

图像质量整体指标评测结果和各维度指标评测结果的SROCC均应大于0.8。

5.5 评测方案封装

评测方案应封装成为可执行文件，提供可视化界面，并支持如下功能：

- a) 应支持在主流操作系统上的安装、运行；
- b) 应支持图像的本地导入、存储、播放以及导出评价统计数据，宜支持联网导入图像；
- c) 支持对单张图片和批量图片进行评价打分和统计分析输出；

d) 宜支持各维度的选择和权重设置。

6 图像质量测试方法

6.1 评价对象图像要求

根据图像的不同评价对象（全图、人脸、车牌等）选择相应的图像质量评价方法，进行评价的图像应满足如下要求：

- a) 进行目标对象图像质量评价应满足图像中目标对象的要求，包括目标对象的完整性及像素要求，如人脸图像质量评测要求图像中人脸宽度不小于40像素，车牌图像质量评测要求车牌长边不小于100像素等；
- b) 评价图像采集设备在微光场景固定光照度的图像质量时，批量导入的图像应在相同光照度环境下采集获取，且数量应不少于100张；
- c) 对比评价不同设备的图像质量时，不同设备应在相同的光照度环境下，对同一批目标对象采集获取。

6.2 测试步骤

图像质量评价按如下步骤进行测试：

- a) 将需要评价的图像保存在指定位置；
- b) 在图像质量评价工具中设置图像导入路径和评价结果导出保存路径；
- c) 如需要，可修改各维度权重；
- d) 执行图像质量评价工具，获取评价结果。

6.3 结果评价

对于单张图像的图像质量评价，打分结果即图像质量评价结果。

- 对于图像采集设备的图像质量进行评价，可通过批量图像的平均值和打分分布来评价图像质量结果：
- 各维度及整体维度评测分数平均值。平均值越高，说明该维度或是整体图像质量越好，如下图中设备B；
 - 各维度及整体维度评测分数分布。当分数在分值区间覆盖范围越小，说明该维度或是整体图像质量越稳定，如下图中设备B。

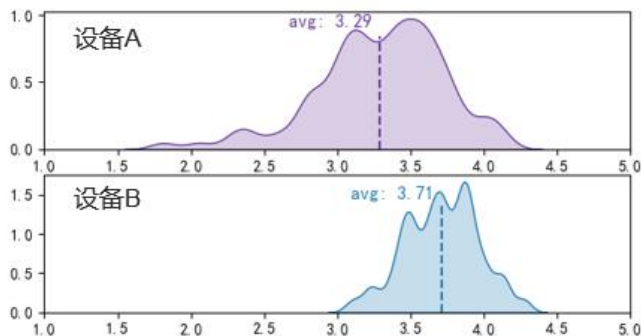


图 4 不同设备评测结果直方图示意图

附 录 A
(资料性)
人脸图像质量评价方法应用

A.1 人脸图像质量评测方案

根据人脸的主观评价感受，人脸图像质量可分为噪声、纹理细节、自然感、亮度对比度五个维度，其评测方案框架如下图。

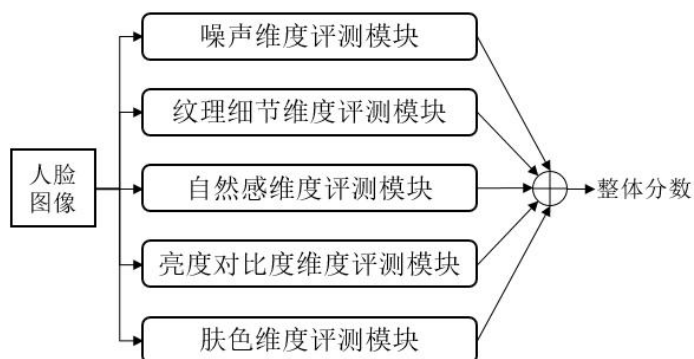


图 5 人脸指标计算模型框图

人脸图像质量评价公式如下：

$$F = \sum \omega_i F_i, i \in \{T, N, F, C, L\}$$

其中 F_N 为噪声指标、 F_T 为纹理细节指标、 F_F 为自然感指标、 F_L 为亮度对比度指标， F_C 为肤色指标， ω 为各指标权重。

根据5.4.2的要求，各维度算法分数与主观整体分数应用最小二乘法拟合得到的最优值： $\omega_N = 0.074$ ， $\omega_T = 0.414$ ， $\omega_F = 0.302$ ， $\omega_L = 0.153$ ， $\omega_C = 0.057$ 。实际使用时可依据具体场景侧重点调整。

A.2 人脸图像质量指标计算方法

A.2.1 传统计算机视觉算法

A.2.1.1 人脸噪声指标 F_N

按下列方法进行测试计算人脸噪声指标：

- a) 使用差分高斯图DoG^[3]方式，模拟这个多空间频率通道的过程，该方法需要使用不同标准差生成的DoG图像可以当做多空间频率通道。在图像处理中，将两幅图像在不同参数下的高斯滤波结果相减，生成DoG图。过程如下所示：

$$\text{DoG} \triangleq G_{\sigma_1} - G_{\sigma_2} = \frac{1}{\sqrt{2\pi}} \left(\frac{1}{\sigma_1} e^{-\frac{x^2+y^2}{2\sigma_1^2}} - \frac{1}{\sigma_2} e^{-\frac{x^2+y^2}{2\sigma_2^2}} \right)$$

- b) 在得到一系列的DoG结果后，按照 32×32 的分辨率进行图像切块，并进行奇异值分解；
- c) 分解所得到的奇异值结果中，噪声图像相较于原始（无噪声）图像会有较高的奇异值。基于此原理，对图像的切块分解奇异值后，选取最小的10个奇异值相加作为该切块的特征，然后每张待评价图像选取特征最小的9个块作为9维特征，基于所提取的特征结果进行评分结果的映射。

A. 2. 1. 2 人脸纹理指标 F_T

人脸纹理指标评估图像中人脸的模糊程度，使用CPBD的计算方法^[4]进行检测计算。通过sobel算子进行边缘检测，可以同时获得边缘像素 e_i 及对应角度方向，对于每个边缘像素可以借助角度方向计算边缘宽度 $w(e_i)$ ，利用模糊概率心理学曲线的原理计算每一个边缘像素的模糊概率如下：

$$P_{blur} = 1 - \exp\left(-\left|\frac{w(e_i)}{w_{JNB}(e_i)}\right|^\beta\right), \beta = 3.6$$

其中，JNB根据对比度 C 分别设置不同的值，如下：

$$w_{JNB}(e_i) = \begin{cases} 5, & C \leq 50 \\ 3, & C \geq 51 \end{cases}$$

心理感知临界点 $P_{JNB} = 0.63$ ，因此最终CPBD的计算如下，当CPBD值超过0.63时，则认为该图像清晰度较好：

A. 2. 1. 3 CPBD = $P(P_{blur} \leq P_{JNB})$ 人脸自然感指标 F_F

自然感采用BRISQUE方法^[5]进行计算，统计特征越接近高斯分布则认为图像越合理、越自然，主要计算MSCN。

- a) 对于图像中的像素 (i, j) ，其与相邻像素具有较高的相关性，因此采用如下的局部亮度归一化过程减弱相关性：

$$\hat{l}(i, j) = \frac{I(i, j) - \mu(i, j)}{\sigma(i, j) + C}$$

其中， $\mu(i, j)$ 和 $\sigma(i, j)$ 分别是局部均值和标准差， $C=1$ ，用于防止分母为零。

- b) 特征提取：首先提取MSCN系数的统计特征，包括形状参数 α 和方差 σ ；其次计算MSCN系数在四个方向上的成对乘积，并拟合AGGD分布；最后提取AGGD的形状参数、不对称参数和左右方差。
- c) 将上述特征汇总为一个特征向量。对于每个尺度，提取18个特征（2个来自MSCN的GGD拟合，16个来自成对乘积的AGGD拟合）。通常，BRISQUE会在两个尺度上提取特征，最终得到36个特征。
- d) 使用预训练的支持向量回归SVR模型对特征向量进行回归，得到图像质量得分。

A. 2. 1. 4 人脸亮度对比度指标 F_L

亮度对比度主要受图像的强度影响，首先提取图像亮度的统计信息如下：

- a) 均值 $\mu = \frac{1}{N} \sum_{i=1}^N I(i)$
- b) 标准差 $\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (I(i) - \mu)^2}$
- c) 偏度 $\gamma = \frac{\frac{1}{N} \sum_{i=1}^N (I(i) - \mu)^3}{\sigma^3}$ （衡量图像强度分布的对称性）
- d) 峰度 $\kappa = \frac{\frac{1}{N} \sum_{i=1}^N (I(i) - \mu)^4}{\sigma^4} - 3$ （衡量图像强度分布的尖锐度或平坦度）

此外，再计算图像强度的熵 $H = -\sum_{i=1}^N I(i) \log_2(I(i))$ 。

其中 N 代表像素总个数。

基于自然场景统计NSS原理^[6]，理想的自然图像的强度分布也趋近于一个高斯分布，因此针对均值、标准差和峰度三个结果，通过已有大数据拟合的NSS高斯分布中计算结果符合理想自然图像的概率，

$$P(\mu) = \frac{1}{\sqrt{2\pi\sigma_\mu^2}} \exp\left(-\frac{(\mu - \mu_0)^2}{2\sigma_\mu^2}\right)$$
$$P(\sigma) = \frac{1}{\sqrt{2\pi\sigma_\sigma^2}} \exp\left(-\frac{(\sigma - \sigma_0)^2}{2\sigma_\sigma^2}\right)$$
$$P(\gamma) = \frac{1}{\sqrt{2\pi\sigma_\gamma^2}} \exp\left(-\frac{(\gamma - \gamma_0)^2}{2\sigma_\gamma^2}\right)$$

将所计算的均值、标准差和峰度代入上述式子可以得出图像为理想自然图像的可能性，进而衡量图像的亮度对比度质量。

A. 2. 1. 5 人脸肤色指标F_C

首先将图像从RGB色彩空间转换至LAB色彩空间，此后计算过程参照保真度的BRISQUE方法过程。

A. 2. 2 深度学习算法

人脸图像质量各维度统一采用深度学习算法的Swin-B模型结构，采用分层的视觉Transformer架构，通过移位窗口机制来高效地处理图像数据。其参数量为88M，以224尺寸的图像作为输入时计算量为15. 4G FLOPs。

A. 3 人脸图像质量评价工具封装

人脸图像质量评价工具界面如下图所示，评价结果采用5分绝对分制。纹理、噪声、自然感、亮度对比度、肤色等指标可选，相应权重可单独设置。



图 6 人脸图像质量评价工具 UI 界面

A.4 人脸图像质量测试方法

A.4.1 人脸图像要求

用于人脸图像质量评价的图像应满足如下要求：

- a) 图像中人脸宽度应在40像素值以上；
- b) 图像中的人脸不应有五官遮挡的大角度人脸图像；
- c) 如需对比不同设备的人脸图像质量，需保证各设备在同光照环境下采集，并且至少各采集100张以上的人脸图像。

A.4.2 人脸图像质量评价测试步骤

按如下步骤进行人脸图像质量测试：

- a) 将人脸图像放入待评测文件夹路径。
- b) 打开人脸图像质量评价工具，选择待评测文件夹导入人脸图像数据。
- c) 如对各维度权重有场景需求，可修改对应维度的权重参数。如没有，直接执行评价工具。
- d) 在评价结果文件夹中查看评价结果，其中包含各人脸各子指标分数、整体指标分数的excel表格及其他结果文件。

参考文献

- [1] Series B. Methodology for the subjective assessment of the quality of television pictures[J]. Recommendation ITU-R BT, 2012, 500(13).
 - [2] Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 10012-10022.
 - [3] Joshi P, Prakash S. NR-IQA for noise-affected images using singular value decomposition[J]. IET Signal Processing, 2019, 13(2): 183-191.
 - [4] Narvekar ND, Karam L J. A no-reference image blur metric based on the cumulative probability of blur detection (CPBD) [J]. IEEE Transactions on Image Processing, 2011, 20(9): 2678-2683.
 - [5] Mittal A, Moorthy A K, Bovik A C. No-reference image quality assessment in the spatial domain[J]. IEEE Transactions on image processing, 2012, 21(12): 4695-4708.
 - [6] Fang Y, Ma K, Wang Z, et al. No-reference quality assessment of contrast-distorted images based on natural scene statistics[J]. IEEE Signal Processing Letters, 2014, 22(7): 838-842.
-