

团 体 标 准

T/GDCSA XXX-XXXX

信息技术 智能相册产品性能评测 指标及方法

Information technology—performance evaluation indicators and methods
for smart photo album products

(征求意见稿)

2024-XX-XX 发布

2024-XX-XX 实施

XXXXXXXXXX

发 布

目 次

前言	II
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 评测环境要求	3
4.1 所需软件与硬件	3
4.2 所需素材	3
5 基本测试流程	3
6 指标及评测方法	4
6.1 评测指标体系	4
6.2 图像分类	4
6.3 图像增强	5
6.4 图像分割	6
参考文献	8

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。

本文件由中移互联网有限公司提出。

本文件由广东省网络空间安全协会归口管理。

本文件起草单位：XXXX…。

本文件主要起草人：XXXX。

信息技术 智能相册产品性能评测指标及方法

1 范围

本文件规定了智能相册类产品或功能的指标要求与评测方法，包括评测环境、流程、指标项与评测方法。

本文件适用于智能相册类产品或具有类似功能点的产品的性能评估。

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。

GB/T 22239-2019 信息安全技术网络安全等级保护基本要求

GB/T 41864-2022 信息技术 计算机视觉 术语

GB/T 41867-2022 信息技术 人工智能 术语

3 术语和定义

下列术语和定义适用于本文件。

3.1

智能相册 smart album

智能相册是一种利用智能算法来管理、处理照片的相册软件。

3.2

图像分类 image classification

根据图像的主要内容或特征，将图像划分到某个特定的预定义类别中。

3.3

图像超分辨率 image super-resolution

一种通过低分辨率图像构建高分辨率图像的方法。

3.4

图像去模糊 image deblurring

去除运动、散焦、降采样、压缩等原因造成的图像模糊进而改善图像质量的方法。

3.5

图像增强 image enhancement

任何基于主管观察和偏好设计的改善图像质量的过程。

3.6

图像分割 image segmentation

将图像分成若干具有相似性质的区域的过程。

3.7

指标 indicator

能用来估计或预计另一度量的一种度量。

3.8

混淆矩阵 confusion matrix

又名误差矩阵，是表示精度评价的一种标准格式。

3.9

正类 positive

二分类问题中期望得到的类别，与之对应的类别为负类。

3.10

负类 negative

二元分类中和正类相对类别。

3.11

真正类 (TP) true positive

二分类问题中被正确判断为正类的样本。

3.12

假正类 (FP) false positive

二分类问题中被错误判断为正类的样本。

3.13

真负类 (TN) true negative

二分类问题中被正确判断为负类的样本。

3.14

假负类 (FN) false negative

二分类问题中被错误判断为负类的样本。

3.15

准确率 accuracy

预测正确的结果占总样本的比例。

3.16

精准率 precision

又名查准率，是真正类与真负类样本占总样本的比例。

3.17

召回率 recall

又名查全率或真阳率，是真正类样本占有所有正类样本的比例。

3.18

F1 分数 f1-score

精准率和召回率的调和平均数。

3.19

交并比 intersection over union, IoU

图像交集部分与并集部分的面积比值。

4 评测环境要求

4.1 所需软件与硬件

- a) 具备图像分类、图像超分辨率等智能算法的相册软件或类似软件。
- b) 支持运行前述软件的硬件平台。
- c) 如软硬件的运行需要互联网连接，还应准备相应的网络环境。

4.2 所需素材

所需素材应包含：

- a) 测试图像，具体图像要求见下文测试流程。
- b) 上述测试图像所对应的标签、原始图像或标准图像。

5 基本测试流程

5.1 功能确认：分析智能相册的功能点，按照下列规则将其分类至对应的类别。

智能相册软件通常包含相册分类功能，可能包含智能抠图、图像增强等其他智能算法功能。测试前先按照功能特点将其分为以下三类：

图像分类型：将图像依据主体物体类别归类到某一预定义的类别中。常见的相册分类属于此功能类别，人物识别、人脸识别等也属于此类别。

图像增强型：任何基于主管观察和偏好设计的改善图像质量的功能。常见的“照片修复”、“模糊图片清晰”等功能输入此类别。

图像分割型：将图像依据某种因素分割成不同的区域。常见的“智能抠图”、“智能去背景”属于此类别。

5.2 性能测试：按照智能相册的功能点所属的类型，分别开展测试。

图像分类型：

- a) 测试时应当准备一定数量的测试图像并记录下其实际所属的类别，确保每种类别图像的数量超过10张。
- b) 将准备好的测试图像导入至智能相册软件，并开启分类功能。
- c) 待图像分类完毕后，记录每张图像的编号、实际类别、相册分类的类别。

图像增强型：

- a) 图像增强型在准备测试图像时根据增强前后分辨率是否会变化分成两类。如果分辨率不变，在

选定一批图像后，将其称之为参考图像。之后使用通过对参考图像添加高斯模糊和噪点的方式得到相应的输入图像。如果分辨率会发生改变，如使用了图像超分辨率技术的增强功能，则应当先测试增强前后的分辨率变化情况。假设一张分辨率为 $m \times n$ 的图像，在处理完成后分辨率为 $M \times N$ ，这时可以先准备一批分辨率为 $M \times N$ 的图像作为参考图像，之后将其缩放至 $m \times n$ ，作为输入图像。参考图像应当选取内容差异较大的图像，数量大于30张。

- b) 使用智能相册软件中图像增强的功能依次处理输入图像。处理后的图像称之为输出图像。记录每张参考图像对应的输入图像和输出图像。

图像分割型：

- a) 准备一批主体明确、类型不同的图像作为原始图像。人工处理图像，精确地确定图像中主体的边缘，得到掩膜图像。数量大于30张。
- b) 使用智能相册的“智能抠图”等图像分割类功能处理原始图像，得到处理后的图像作为输出图像。
- c) 记录每张原始图像对应的掩膜图像和输出图像。

5.3 结果统计：按照下文指标与测评方法部分的说明，进行主观和客观指标的评分。

6 指标及评测方法

6.1 评测指标体系

通话智能翻译产品质量评测指标体系见表1。

表 1 通话智能翻译产品质量评测指标体系

类别	指标	评测方法
图像分类	Macro-F1分数	测试计算
图像增强	图像质量主观评价	主观评估
	峰值信噪比评分	测试计算
	结构相似度评分	测试计算
图像分割	图像分割主观评价	主观评估
	像素准确度评分	测试计算
	像素交并比评分	测试计算

6.2 图像分类

6.2.1 Macro-F1 评分

将测试图像实际所属的类别称之为真实值，将智能相册分类的结果称之为预测值。对于每一个分类类别，根据真实值与预测值将测试分成如表2的四类。

表2 图像分类混淆矩阵表

		预测值	
		正类 Positive	负类 Negative
真实值	正类 Positive	真正类 TP	假负类 FN
	负类 Negative	假正类 FP	真负类 TN

根据每一类的混淆矩阵，计算单独一类的精准率、召回率，并计算得到单独一类的 F1 分数，见式 (1)、式 (2)、式 (3)：

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \dots\dots\dots (1)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \dots\dots\dots (2)$$

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \dots\dots\dots (3)$$

假设相册可识别 C1 至 Cn 等 N 类事物，则 Macro-F1 分数为 N 类 F1 分数的平均值，该指标评分见式 (4)：

$$\text{评分} = \text{MacroF1} \times 100 = \frac{\sum_{i=1}^N \text{F1}_{C_i}}{N} \times 100 \dots\dots\dots (4)$$

6.3 图像增强

6.3.1 图像质量主观评价

将图像增强后的输出图像与原始的参考图像进行对比，依据人的主观感受进行评分，见表 3。

表 3 图像质量主观评分规则表

评分	主观评分标准
100 分	输出图像与参考图像基本一致。
80 分	输出图像的质量相对输入图像而言有较大的改善，仍存在一些模糊、噪点。
60 分	输出图像的质量相对输入图像而言有一定的改善，但是存在较多的模糊、噪点等未去除。
40 分	输出图像的质量相对输入图像而言稍有改善，但是存在很多的模糊、噪点未消去、或生成了新的影响质量的不利因素。
0 分	输出图像未见任何感官上的改善，或者出现感官上的劣化。

参照表 3 的标准，在 0-100 分的范围内进行评分。

6.3.2 峰值信噪比评分

将图像增强后的输出图像与原始的参考图像转化为灰度图像，分别为图像 K 和图像 I，以像素为单位，图像的尺寸是 M×N，则峰值信噪比的定义见式 (5)：

$$\text{PSNR} = 20 \times \log_{10} \left(\frac{\text{MAX}_I}{\sqrt{\text{MSE}}} \right) \dots\dots\dots (5)$$

其中 MAX_I 是参考图像上取值最大的像素点的取值。MSE 的定义见式 (6)：

$$MSE = \frac{1}{M \times N} \sum_0^{M-1} \sum_0^{N-1} \|I(i,j) - K(i,j)\|^2 \dots\dots\dots (6)$$

根据计算得到的 PSNR 值，评分规则见式（7）：

$$\text{评分} = \begin{cases} 100, & \text{PSNR} \geq 40 \\ 60 + \frac{(\text{PSNR} - 30) \times 40}{10}, & 40 > \text{PSNR} \geq 30 \\ \frac{(\text{PSNR} - 20) \times 60}{10}, & 30 > \text{PSNR} \geq 20 \\ 0, & \text{PSNR} < 20 \end{cases} \dots\dots\dots (7)$$

6.3.3 结构相似性指数评分

将图像增强后的输出图像与原始的参考图像转化为灰度图像，分别为 x 和 y，则 SSIM 计算方法见式（8）：

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \dots\dots\dots (8)$$

$$C_1 = (k_1 \times L)^2$$

$$C_2 = (k_2 \times L)^2$$

其中 μ_x 和 μ_y 分别表示两张图片的均值， σ_x 和 σ_y 为 x 和 y 的协方差， σ_{xy} 表示 x 和 y 之间的协方差， k_1 、 k_2 为常数，分别取 0.01 和 0.03。L 表示图像像素的最大取值，对于一般的 8 位深度的图像， $L = 2^8 - 1 = 255$ 。

SSIM 值取值范围为[-1, 1]，当值为 1 时代表 x 与图像完全一致。该指标的评分方法见式（9）：

$$\text{评分} = \begin{cases} SSIM \times 100, & SSIM > 0 \\ 0, & SSIM \leq 0 \end{cases} \dots\dots\dots (9)$$

6.4 图像分割

6.4.1 图像分割主观评价

评价图像分割的边缘是否够流畅、完整，方法见表 4。

表 4 图像分割主观评分规则表

评分	主观评分标准
100 分	输出图像边缘流畅无毛刺，与参考的掩膜一致。
80 分	输出图像边缘比较流畅、毛刺较少，与掩膜基本一致。
60 分	输出图像边缘流畅程度一般，存在较多毛刺，与掩膜相比整体形状有局部细节差异。
40 分	输出图像边沿不流畅、毛刺较多，形状与掩膜相比有出入，主体有部分缺失或者存在额外内容。
0 分	输出图像和掩膜有较大差异，遗漏或额外增加较大面积的内容。

参照表 4 的标准，在 0-100 分的范围内进行评分。

6.4.2 像素准确率评分

遍历图像，根据对应位置的像素是否包含在输出图像及掩膜中，划分混淆矩阵：

表5 像素分类混淆矩阵表

		输出图像	
		存在 Positive	不存在 Negative
掩膜图像	存在 Positive	真正类 TP	假负类 FN
	不存在 Negative	假正类 FP	真负类 TN

像素准确度PA定义见式（10）：

$$PA = \frac{TP+TN}{TP+TN+FP+FN} \dots\dots\dots (10)$$

该指标评分方法见式（11）：

$$\text{评分} = PA \times 100 \dots\dots\dots (11)$$

6.4.3 像素交并比评分

遍历图像，根据对应位置的像素是否包含在输出图像及掩膜中，划分混淆矩阵（同上节）。交并比IoU的计算方法见式（12）：

$$IoU = \frac{|A \cap B|}{|A \cup B|} = \frac{TP}{TP + FP + FN} \dots\dots\dots (12)$$

该指标评分方法见式（13）：

$$\text{评分} = IoU \times 100 \dots\dots\dots (13)$$

参 考 文 献

- [1] GB/T 41864-2022 信息技术 计算机视觉 术语
 - [2] GB/T 41867-2022 信息技术 人工智能 术语
 - [3] T/CESA1026-2018 人工智能 深度学习算法评估规范
 - [4] 周志华. 机器学习. 清华大学出版社. 2016
-