ICS 35.240
CCS L70

团　体　标　准

T/CESA 1199—2022

人工智能 智能字符识别技术规范

Artificial intelligence—Technical specification for intelligent character recognition

2022-06-30 发布　　2022-06-30 实施

中国电子工业标准化技术协会　发布

目　　次

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。

本文件由中国电子技术标准化研究院提出。

本文件由中国电子技术标准化研究院、中国电子工业标准化技术协会归口。

本文件起草单位：中国电子技术标准化研究院、腾讯云计算（北京）有限责任公司、华为技术有限公司、深圳云天励飞技术股份有限公司、四川云从天府人工智能科技有限公司、西安深信科创信息技术有限公司、美的集团（上海）有限公司、阿里云计算有限公司、北京百度网讯科技有限公司、浙江大华技术股份有限公司、北京旷视科技有限公司、杭州海康威视数字技术股份有限公司、华为云计算技术有限公司、上海计算机软件技术开发中心、上海依图网络科技有限公司、上海商汤智能科技有限公司、深圳市矽赫科技有限公司、马上消费金融股份有限公司、北京九章云极科技有限公司、西北工业大学、上海人工智能研究院有限公司。

本文件主要起草人：董建、马珊珊、刘海涛、杨晓光、刘皓、张小宝、徐洋、杨雨泽、王小叶　、王彭、郑文先、代翔、李军、李继伟、田福康、胡蓉、脱立恒、郭嘉、姚聪、杨志博、章成全、杨烨华、李笑如、陈媛媛、熊剑平、程淼、梅敬青、付英波、程战战、钮毅、谢泽澄、符海芳、郝阳阳、陈敏刚、马泽宇、赵春昊、梁鼎、武焕、洪鹏达、洪宝璇、李云峰、刘志强、方磊、毛玉婷、王鹏、王冀、宋海涛、王资凯。

人工智能 智能字符识别技术规范

1 范围

本文件确立了智能字符识别技术参考框架，规定了功能要求和性能要求，描述了对应的测试方法。

本文件适用于智能字符识别产品和服务的设计、开发、应用和测试评价。

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 5271.12—2000 信息技术 词汇 第12部分:外围设备

3 术语和定义、缩略语

3.1 术语和定义

GB/T 5271.12—2000界定的以及下列术语和定义适用于本文件。

3.1.1

光学字符识别 optical character recognition

一种字符识别，它使用光学手段鉴别图形字符。

[来源：GB/T 5271.12—2000，12.01.53]

3.1.2

智能字符识别 intelligent character recognition

一种基于深度学习的光学字符识别技术。对印刷文字、手写文字、表格、公式符号以及文档结构要素进行识别和编码。

3.1.3

文本检测 text detection

对图像上字符（串）、文本行（列）位置进行定位的过程。

注：字符（串）包括数字，符号，英文，中文或其他语言文本。

3.1.4

文本识别 text recognition

对图像上字符（串）、文本行位置进行识别的过程。

注：字符（串）包括数字，符号，英文，中文或其他语言文本。

3.1.5

文本信息提取 text information extraction

对图像上识别出的文本，进行排序、合并、自然语言处理等操作，使其转换为结构化信息的过程。

3.2 缩略语

下列缩略语适用于本文件。

AI：人工智能（artificial intelligence）BMP 位图（bitmap）

GIF：图像互换格式（graphics interchange format）

ICR：智能字符识别（intelligent character recognition）

JPEG：联合图像专家组（joint photographic experts group）

OCR：光学字符识别（optical character recognition）

PNG：便携式网络图形（portable network graphics）

PDF：可携式文件格式(portable document format)

TIFF：标签图像文件格式（tag image file format）

WER：词错误率（word error rate）

4 智能字符识别系统框架

基于人工智能技术的字符识别系统框架见图1。

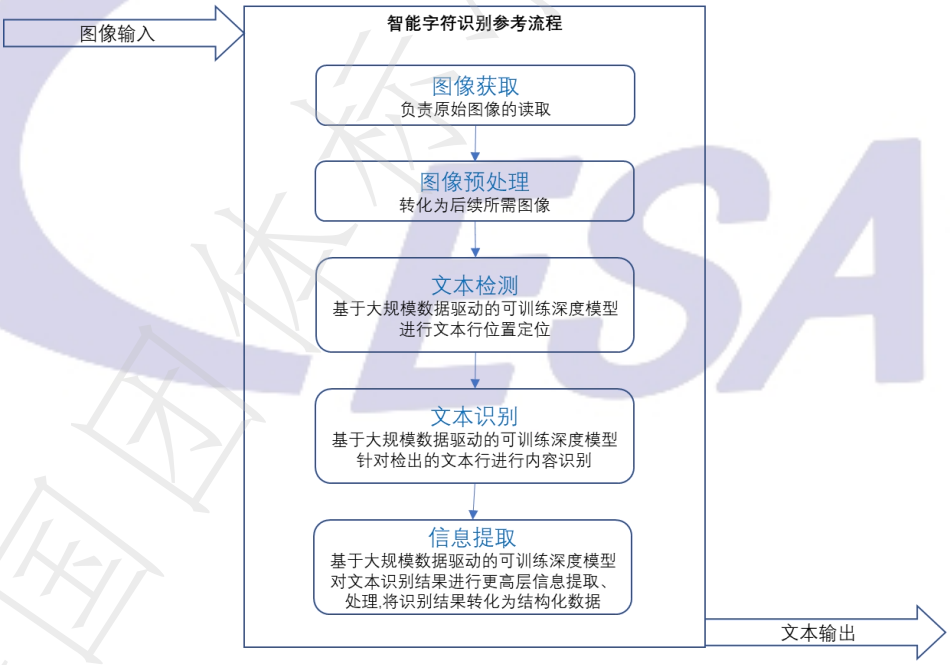

图 1 智能字符识别系统框架

ICR将给定图像中的点或像素字符转换为数字编码信息，包括图像获取、图像预处理、文本检测、文本识别、信息提取五个模块。

a) 图像获取模块负责图像、视频截图中图像的读取；

b) 图像预处理模块负责对从图像获取模块接收到的图像原始数据，将其转换成满足文本检测、文本识别等输入要求的图像，其中包括质量增强、版面分析和质量筛选等功能；

c) 文本检测模块负责对于给定图像（包括原始图像、图像中间样本等）进行字符和文本行检测；

d) 文本识别模块负责对于给定文本图像，可包括原始图像、图像中间样本、文本检测出的图像区域等，进行字、词和文本行的内容识别；
e) 信息提取模块依据版面分析、自然语言处理等手段将基于文本检测和文本识别结果，转换为结构化数据，以及识别结果的矫正。

5 功能要求

5.1 图像采集

图像采集功能应符合以下要求：
a) 支持对包含但不限于 JPEG、GIF、PNG、TIFF、BMP、PDF 等常见的图片格式进行读取。图片分辨率支持范围应该包含 128×128 dpi～4096×4096 dpi；
b) 支持对包括但不限于自然场景 卡证、票据、文档、表单等常见文本场景文字的检测和识别。

5.2 图像预处理

对获取到的图像进行预先处理，使图像便于后续的检测和识别符合以下要求：、
a) 增强图像质量，应对图像进行几何变换、畸变校正、修剪、数据格式转换等操作；采用滤波、超分辨率等技术手段，在不破坏图像边缘、轮廓等原有细节的条件下对噪声进行抑制；
b) 版面分析，应根据适用场景有效的检测并区分出文字段落区域、图片区域、表格、图表、公式、图章、二维码等不同类别的元素；
c) 质量筛选，宜对图像成像质量及图像完整性进行评价和判别，过滤无法正常识别的低质量和完整度不足的图像，如带有反光、暗光、防伪标识等干扰、以及关键角点缺失等完整度不足的图像。

5.3 文本检测

5.3.1 字符检测

在原始图像或图像中间样本识别从预定义范围的字符符合以下要求：
a) 应支持设置待检测字符类型范围，如：Unicode 字符集；
b) 应检测出预先定义范围内的字符类型，包含但不限于：中文简体、中文繁体以及英语、阿拉伯语、俄语等西文；宜支持藏语、蒙语、维语等少数民族语言，数字、特殊符号及其组合等；
c) 应在检测结果中包含字符在图像中的位置信息；
d) 宜支持对所检测图像中的最小、最大字符大小的设置，如：8 px～256 px。

5.3.2 文本行检测

对原始图像或图像中间样本进行文本行检测符合以下要求：
a) 应定位出图像中文字块的位置，位置信息支持水平矩形、旋转矩形、不规则四边形以及多轮廓点等形式；
b) 应根据位置信息将含有文本行的区域，通过算法，如：仿射、最小外接矩形等归一化算法，处理成规则的图像数据；
c) 应支持对所需检测图像分辨率的设置，如：128×128 dpi～4096×4096 dpi；
d) 宜支持不同语言种类的检测，如对中文、英文、混合语种的检测；支持印刷体和手写体的混合模式、不同字体类型大小、不同角度倾斜、不同程度遮挡物等情况的文字区域检测。

5.4 文本识别

对图片中的文本行检测区域进行定位后，对检测区域内文本内容进行识别，应符合以下要求：

a) 对印刷文字和手写文字的识别；
b) 对字符、字母、混合语种中的文字内容进行识别；
c) 对数字、数学公式以及特殊符号的识别；
d) 英文识别的最小尺寸为 16 x16px，中文识别的最小尺寸为 32x32px；
e) 对方向有旋转的文字，支持文字与水平轴＜±15°夹角偏转；
f) 支持对如中、日、韩文等有比较多竖排文字呈现的文字识别。

5.5 信息提取

信息提取是基于文本检测和文本识别结果，将嵌入其中的结构化信息或非结构化信息自动提取转换为结构化数据，应符合以下要求：

a) 可对文本中的特定词汇进行纠错；
b) 可根据特定的语言上下文的关系，对识别结果进行校正。对于需要校正的字段，应支持定义校正规则，并依据校正规则进行处理，如日期、地址、金额类等；
c) 可支持对文本版式结构的还原，包括但不限于标题、章节、段落、图表、脚注、页眉、页脚等版本格式；
d) 可支持识别表格区域行列信息，并对表格区域结构单元信息进行恢复还原。

6 性能要求

6.1 文本检测性能要求

6.1.1 交并比（IoU）

交并比是用来评价文本目标框和文本预测框之间的重合度。计算公式如式1，

$$\mathrm{IoU}=\frac{area(B_p\cap B_{gt})}{area(B_p\cup B_{gt})} \quad (1)$$

式中：

Bp——预测的矩形框区域；

Bgt——标注的矩形框区域。

6.1.2 精确率（PR）

精确度包括字符精确度、单词精确度和字段精确度。其中：

a) 字符精确率：适合中文 ICR 评测，字符包括单个文字以及标点符号；
b) 单词精确率：适合英文 ICR 评测，单词以空格分隔；
c) 字段精确率：适合卡证类、发票类、车牌 ICR 评测，能提取到结构化的字段信息。

精确率用于衡量正确检测出的字符（串）框数量占所有检测出的字符（串）框数量的比例。计算方法见公式2：

正确检测是指预测框与目标框的IoU不小于0.5。

$$PR=\frac{C}{M}\times 100\% \quad (2)$$

式中：

PR ——精确率；

C ——正确检测出的字符（串）框数量；

M ——检测出的字符（串）框总数量。

6.1.3 召回率（RR）

召回率包括字符召回率、单词召回率和字段召回率。召回率用于衡量正确检测出的字符（串）框数量占图像上真实存在的字符（串）框数量的比例。计算方法见公式3：

$$RR = \frac{C}{N} \times 100\% \quad \text{（3）}$$

式中：

RR ——召回率；

C ——正确识别的字符（串）框数量；

N ——图像上应该被正确识别的字符（串）框总数量。

6.1.4 F Score

F Score为精确率和召回率的调和平均,平衡因子（权重）取1。计算方法见公式4：

$$FScore = (1+\beta^2)\frac{PR \times RR}{\beta^2(PR+RR)} \quad \text{（4）}$$

式中：

FScore——精确率和召回率的调和平均；

β ——平衡因子，β=1时，精确率和召回率权重相同；

PR ——精确率；

RR ——召回率。

6.1.5 AP 测度

AP测度为在不同IoU阈值情况下，不同召回率下的平均精确率。以召回率RR为横轴，精确率PR为纵轴，可以得到不同IoU阈值下的精确率-召回率曲线。通常，IoU阈值较低时，精度低,召回高，IoU阈值较高时，精度高,召回低，这样可以得到一条类似双曲线的函数。之后对该曲线进行平滑处理，即该曲线上的每一个点，精确率的值取该点右侧最大的精确率的值。绘制出平滑后的精确率-召回率曲线后，取横轴0-1的10等分点的精确率的值，计算其平均值作为最终AP测度的值，它的计算公式见公式5：

$$AP = \frac{1}{11} \sum_{RR \in \{0, 0.1 \ldots 1.0\}} PR_{smooth}(RR) \quad \text{（5）}$$

式中：

AP ——不同召回率下的平均精确率；

RR ——召回率；

PR_{smooth}——平滑后的精确率。

6.1.6 不同场景下文本检测性能要求

电子/扫描、拍照、自然街景、网络、多语音等场景下的文本检测性能要求应符合表1的要求。

表 1　文本行检测要求

场景类型	场景描述	精确率	召回率	F Score	AP
电子/扫描	使用数字格式或扫描得到带有文本的图像	≥95%	≥95%	≥95%	≥90%
拍照	使用相机拍照得到带有文本的图像	≥90%	≥90%	≥90%	≥85%
自然街景	使用自然街景中带有文本的图像	≥70%	≥75%	≥70%	≥65%
网络	使用网络获取到带有文本的图像	≥80%	≥80%	≥80%	≥75%
多语言	使用多语言图片中带有文本的图像	≥70%	≥60%	≥60%	≥55%

6.2　文本识别性能要求

6.2.1　精确率(PR)

精确率用于衡量正确识别出的字符（串）数量占所有检测出的字符（串）数量的比例。计算方法见公式6：

$$PR = \frac{C}{M} \times 100\% \quad \text{（6）}$$

式中：

PR ——精确率；

C ——正确识别的字符（串）数量；

M ——识别的字符（串）总数量。

注1：若引擎将两个单词之间的空格漏掉，则两个单词都算识别错误。

注2：字段中如有一个错误识别的文字则整个字段算作识别错误。

6.2.2　编辑距离

编辑距离表示一个字符串修改为和另外一个字符串一致，总共需要修改的字符数。编辑距离越大表示两字符串之间的差异越大。编辑距离包括全图编辑距离、最小编辑距离和平均编辑距离。

归一化编辑距离可以衡量两个字符串之间的相似性，在编辑距离的基础上加入归一化操作可规避字符串长度带来的指标差异。计算方法见公式 7：

$$\mathrm{Norm} = 1 - \frac{1}{N}\sum_{i=1}^{N} D\left(s_i, \hat{s}_i\right) \Big/ \max\left(s_i, \hat{s}_i\right) \quad \text{（7）}$$

式中：

Norm ——归一化编辑距离；

N ——文本行的总数；

s_i ——预测的文本内容；

$\hat{s}_i$ ——真实文本内容；

$\max\left(s_i, \hat{s}_i\right)$——$s_i$ 和 $\hat{s}_i$ 的最大长度；

$D\left(s_i, \hat{s}_i\right)$——$s_i$ 与 $\hat{s}_i$ 的编辑距离，描述了两个字符串的相似度，定义为从一个字符串变换到另一个字符串所需要的最少操作数。

例如，有一个字符串 a='love',b='lolpe'.那么计算a和b的编辑距离，就是要算出从a变化到 b需要经过多少个步骤。

a) love->lolve(插入 l)

b) lolve->lolpe(用 v 替换成 p)

那么ab之间的编辑距离为2。

6.2.3 词错误率

词错误率（Word Error Rate，WER）是一种基于编辑距离的评价文本识别准确率的指标。而在评价文本，通常采用此错误率，该指标的定义为公式8：

$$\mathrm{WER} = \frac{EDITDIS(label, pred)}{LENGTH(label)} \times 100\% \quad \text{……（8）}$$

式中：

EDITDIS(label,pred) ——表示标签label与预测的结果pred之间的编辑距离；

LENGTH(label)——表示标签的字符数。

6.2.4 不同场景下文本识别性能要求

印刷文字、手写文字等场景下的文本识别性能要求应符合表2的要求。

表 2 文本行识别要求

场景类型		单字符精确率	文本行精确率	编辑距离
印刷文字	中文	≥96%	≥75%	≥78%
	数字	≥97%	≥85%	≥88%
	英文	≥98%	≥85%	≥88%
	特殊字符	≥95%	≥85%	≥88%
手写文字	签名[a]、批注	≥90%	≥80%	≥83%
	一般手写文字	≥80%	≥65%	≥68%
[a] 字迹清晰、非艺术字体				

7 测试方法

7.1 测试流程

智能字符识别系统的测试流程见图2。

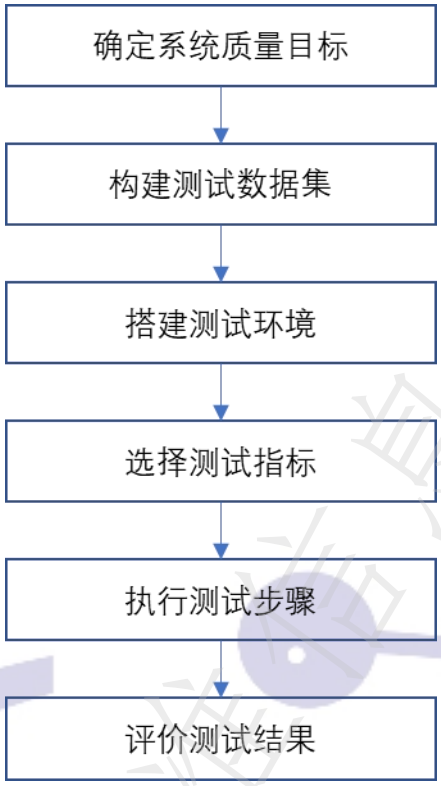

图 2 智能字符识别测试流程

7.2 确定系统质量目标

应运用以下步骤确定智能字符识别系统的质量目标：

a) 场景分析：分析智能字符识别系统的应用场景、运行环境与使用流程，既要考虑系统正常使用的情况，也要考虑可预见的异常情况；

b) 风险分析：根据智能字符识别系统的不同应用场景，分析误识别与漏识别可能出现的风险，分析针对字符识别系统可能产生的对抗攻击手段；

c) 确定系统质量目标：根据系统的应用场景和风险，确定智能字符识别系统的质量目标，包括：

1) 确定系统功能有效性、性能、兼容性、维护性、可移植性、训练数据集的质量、对抗样本的影响、对应用场景数据的鲁棒性、可解释性、安全性的指标要求；

2) 确定测评指标评价的准则。

7.3 构建测试数据集

在测试开始前，应根据不同场景制作测试数据集。采集数据要均衡，避免场景单一、字体单一、文字信息单一，尽量均衡覆盖常用汉字和各类字符。

测试场景及对应的测试数据集要求如下：

a) 印刷文字场景：测试数据集应包括但不限于卡证类、票据类、车牌类和文档类数据；每种类型的测试数量应不少于 200 张；样本图片类型应包括不同拍摄角度、不同光线场景；样本字符应包括中文简/繁体、生僻字、英文、特殊字符、多语言字符；

b) 手写文字场景：测试数据集应包括作文类、试卷类、批注类数据；每种类型的测试数量应不少于 200 张；样本图片类型应包括不同手写字体、不同版面类型、和不同拍摄光线及可能出现的遮挡、涂改、污损等；样本字符应包括中文简/繁体、生僻字、英文、特殊字符、多语言字符；

c) 其他文字场景：除了常规场景，也需要考虑一些数据增强场景。例如：加噪、图像压缩、旋转、图像缩放等。该阶段需要完成数据集的采集，数据清洗，数据标注，标注结果校验的工作。保证测试数据的完整、标注数据的准确性。

7.4 搭建测试环境

根据被测的智能字符识别服务所需要的软硬件参数构建出完整的软硬件环境，保证被测服务在环境中运行正常。

软件提供与扫描仪的接口，如扫描仪驱动软件。硬件配置如影像扫描仪、传真机或任何摄影器材等设备。

若无法复现出测试服务需要的软硬件环境，则要能够通过其他方式支撑服务的运行，并且人为可控因运行环境带来的测试差异。

7.5 选择测试指标

根据制定的系统质量目标，选择第6章中描述的若干测试指标作为测试目标。

7.6 执行测试步骤

确定被测服务的应用场景（例如：自然街景或电子扫描、手写体或印刷体、检测服务或识别服务），然后检出相应的测试集数据按照指定的请求协议，获取每个测试图片的服务处理结果。

将得到的服务处理结果转化为指定的数据文件格式。

根据第6章给出的不同场景的指标统计方式，结合被测服务的应用场景，使用指标统计工具，计算出具体测试场景的指标。

7.7 评价测试结果

文本检测性能测试结果应符合6.1.6中表1，文本识别性能测试结果应符合6.2.4中表2。