ICS 35.240
L 60

团　　体　　标　　准

T/ISC 0058—2024

文本图像篡改检测系统技术要求

Standard for Text Image Tampering Detection System

2024-9-3 发布　　2024-10-3 实施

中　国　互　联　网　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。

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。

本文件由中国互联网协会提出并归口。

本文件起草单位：

中国信息通信研究院、上海合合信息科技股份有限公司、中国图象图形学会、中国科学技术大学、深圳大学、上海交通大学、华南理工大学、南开大学、北京智游网安科技有限公司、蚂蚁科技集团股份有限公司。

本文件主要起草人：

王景尧、吴荻、李玮、郭丰俊、丁凯、宋宏宇、陆大公、金连文、谢洪涛、王裕鑫、李斌、李昊东、金耀辉、薛洋、高学、胡梦婷、原国杰、刘健、张天翼、刘菁菁、唐佳伟、陈倩华、曹海啸、冯艺卓、何梦醒、张家珲、常天恩。

引　言

科技的发展使得图像逐渐成为重要的信息传递手段，人们逐渐将纸质文本以数字图像的形式进行信息传递，给人们提供了方便的同时，也带来的了极大的安全隐患。

文本图像篡改指的是对包含文本内容的图像（如卡证、文档、截图等）进行篡改。目前，图像篡改检测研究主要集中在检测自然图像中被篡改的物体。相比之下，文本图像篡改表现出不同的特征，例如篡改主要集中在文本区域且篡改痕迹难以察觉等。这些特征为图像取证带来了新的挑战。

人们或通过图像编辑工具进行图像编辑，或以物理篡改更加低成本地进行图像篡改，以达到人眼难以区分或人工智能机器难以辨认的目的。

文本图像篡改检测技术规范的目的是，为文本图像篡改检测技术的发展、应用和推广提供指导和支持，以解决实际问题并确保技术的可靠性和有效性。建立文本图像篡改检测技术规范具有必要性和重大意义。

文本图像篡改检测系统技术要求

1 范围

本文件规定了文本图像篡改检测系统的技术要求。

本文件适用于拍照证照、扫描证照、拍照文档、扫描文档、截图等五大类文本图像篡改检测系统的设计与评估。

2 规范性引用文件

本文件无规范性引用文件。

3 术语和定义

下列术语和定义适用于本文件。

3.1 文本图像 Text Image

文本图像是通过某种方式将纸质文本数字化而得到的以图像格式存储的数据,可供用户电子阅读，可供计算机进行相应的信息处理。

3.2 文本图像篡改 Text Image Tampering

文本图像篡改是对文本图像的字符（串）、文本行（列）、图像区域进行未经授权的修改或操纵，以改变文本图像的内容、外观或含义的过程。

注：字符（串）包含符号、数字、字母、拼音音标、希腊字母、英文音标、简体中文、繁体中文、藏文、英文或其它语言。图像区域包括文字，人像、图标等。仅修改图像呈现质量而不改变图像内容信息的操作不属于文本图像篡改。

3.3 文本图像篡改检测 Text Image Tampering Detection

文本图像篡改检测是对文本图像内的字符（串）、文本行（列）、图像区域进行检测并判别文本图像是否存在篡改的过程（篡改手段详见附录 13.1）。

注：字符（串）包含符号、数字、字母、拼音音标、希腊字母、英文音标、简体中文、繁体中文、藏文、英文或其它语言。图像区域包括文字，人像、图标等。仅修改图像呈现质量而不改变图像内容信息的操作不属于文本图像篡改。

3.4 物理篡改 Physical Tampering

物理篡改是指直接在检测主体上进行篡改，再经过拍照或扫描等操作转为文本图像。

3.5 数字篡改 Digital Tampering

数字图像篡改是指将纸质文本转为数字图像后再进行篡改活动。

3.6 物理攻击 Physical Attack

物理攻击是指在生成文本图像之前，对主体进行干扰，从而造成文本图像篡改检测系统失灵。

3.7 数字攻击 Digital Attack

数字攻击是指直接在文本图像上添加干扰，从而造成文本图像篡改检测系统失灵。

3.8 准确率 Accuracy

在特定数据集中检测所有无篡改文本图像和篡改文本图像的准确程度的测量指标。

3.9 误检率 False Positive Rate

在特定数据集中无篡改文本图像误检程度的测量指标。

3.10 召回率 True Positive Rate or Recall

在特定数据集中检测篡改文本图像召回程度的测量指标，即检测篡改文本图像的准确程度的测量指标。

3.11 均值交并比 mean Intersection over Union

在特定数据集中定位篡改区域像素级准确度的测量指标。

3.12 可交换图像文件格式 Exchangeable image file format

记录数码照片的属性信息和拍摄数据。

4 缩略语

TP	True Positive	真阳
TN	True Negative	真阴
FP	False Positive	假阳
FN	False Negative	假阴
Acc	Accuracy	准确率
FPR	False Positive Rate	误检率
TPR/Recall	True Positive Rate	召回率
IoU	Intersection over Union	交并比
mIoU	mean Intersection over Union	均值交并比
Exif	Exchangeable image file format	记录数码照片的属性信息和拍摄数据

5 系统输入/输出信息

5.1 系统输入信息

文本图像输入应支持以下要求：

1、 支持对包含但不限于JPG、PNG、BMP、TIFF、WEBP、单帧GIF格式等常见格式作为输入；

2、 支持对包含但不限于PDF等不可编辑内容的常见格式转为文本图像；

3、 可支持对视频进行抽帧并转为文本图像；

4、 文本图像主体分辨率不小于64×64像素；

5、 支持包含不同语言的文本图像作为输入；

6、 支持原始图像输入，不需要对原始图像进行裁切。

5.1.1 文本图像获取方式

文本图像输入应支持拍照、扫描、截屏等常见文本图像获取方式。

检测系统应支持以下现象产生的文本图像。

5.1.1.1 拍照

对于拍照获取的文本图像，应支持以下要求：

1、 应支持手机相机、单反相机等不同拍照设备获取的拍照图像；

2、 应支持不同角度拍照获取的拍照图像；

3、 应支持一定拍照距离拍照获取的拍照图像；

4、 应支持光照强度过强或过暗条件下拍照获取的拍照图像；

5、 应支持由拍照设备导致色彩失真的拍照图像；

6、 应支持拍照获取的拍照图像内存在背景干扰；

7、 应支持由拍照设备产生噪声的拍照图像；

8、 应支持由拍照设备产生镜像翻转的拍照图像；

9、 应支持存在透视的拍照图像；

10、 应支持存在非检测内容遮挡的拍照图像；

11、 应支持纸质文本存在弯曲或弯折的拍照图像。

5.1.1.2 扫描

对于扫描获取的文本图像，应支持以下要求：

1、 应支持由扫描设备导致的色彩失真的扫描图像；

2、 应支持由不同扫描设备生成的彩色扫描图像；

3、 应支持由扫描设备的光源质量导致亮度不均匀的扫描图像；

4、 应支持由扫描设备生成的低质量扫描图像；

5、 应支持纸质文本存在弯曲或弯折的拍照图像；

6、 应支持由拍照图像转为扫描文件的扫描图像；

7、 应支持扫描时纸张摆放的角度倾斜导致图像倾斜的扫描图像；

5.1.1.3 截屏

对于截屏获取的文本图像，应支持以下要求：

1、 应支持电子设备导致分辨率低、图像质量差的截屏图像；

2、 应支持存在背景干扰的截屏图像；

3、 应支持由其他主体遮挡导致待检测主体不完整的截屏图像；

4、 应支持存在水平翻转或垂直翻转的截屏图像。

5.1.2 文本图像类型

文本图像输入应支持支持以下文本图像类型：

1、 支持证件类型的文本图像；

2、 支持文档类型的文本图像；

3、 支持截图类型的文本图像；

单一类型的文本图像，文本图像应支持以下要求：

1、 应支持不同国家的文本图像；

2、 应支持不同语言的文本图像；

3、 应兼容同一证件或文档类型下不同版本的文本图像。

以身份证件为例：

1、 应支持不同国家不同格式的身份证；

2、 应支持不同国家或同一国家的不同使用语言的身份证；

3、 应支持不同版本的身份证，如居民身份证、临时身份证等。

5.2 系统输出信息

结果输出形式

1、 输出结果宜采用JSON格式文件；

2、 输出文档应包含图片类型，图片篡改检测结果、图片篡改检测结果可视化、篡改区域坐标及坐标格式、篡改区域检测置信度等信息。输出文件参考样例如下：

表1 文本图像篡改分类与定位文档输出形式参考样例

JSON 文档	说明
"type": "figure"	图片类型
"tamperScore": 0.008219	图片篡改检测结果
"image": "/9j/4AAQSkZJRgABAQAAAQABAAD/2"	图片篡改检测结果可视化
"exif": [	Exif 信息输出
"artist": ["Lin"]	编辑者
"software":[" Photoshop"]	编辑软件
"Datatime": ["2023-07-21 09:11:54"]]	编辑时间
"DateTimeDigitized": ["2023-07-21	图像写入时间
09:10:23"]]	
"locations": [	文本图像篡改区域检测结果
"points": [	文本图像篡改区域定位坐标

[115.0,	左上角横坐标
364.6],	左上角纵坐标
[216.0,	右下角横坐标
277.0]],	右下角纵坐标
″confidence″: [0.99344], tamperType:[“”]]	置信度分数

6 文本图像篡改检测

6.1 文本图像篡改分类

文本图像篡改分类模型框架支持多种模型，包含但不限于单一文本图像篡改分类模型（如身份证文本图像篡改分类模型），通用证件/文档/截图文本图像篡改分类模型，以及通用文本图像篡改分类模型。文本图像篡改分类系统应支持对输入的文本图像做整体判断，检测文本图像是否存在篡改，应检测的内容如下：

1、 应支持字符（串）的篡改分类；

2、 应支持文本行的篡改分类；

3、 应支持图像区域的篡改分类。

分类指标：Acc、FPR、Recall，计算方法见10.1。性能要求见10.2。
检测内容可进一步划分。

6.1.1 证件篡改分类

证件文本图像篡改分类应支持以下内容：

1、 应支持文字和数字内容篡改分类，包含但不限于时间、编码、身份信息等内容；

2、 应支持图像区域篡改分类，包含但不限于人像、印章等内容；

3、 应支持二维码、条形码篡改分类；

4、 应支持字体和排版篡改分类。

6.1.2 文档篡改分类

文档文本图像篡改分类应支持以下内容：

1、 应支持文字和数字内容的篡改分类；

2、 应支持图像和图表的篡改分类；

6.1.3 截图篡改分类

截图文本图像篡改分类应支持以下内容：

1、 应支持文字和数字内容的篡改分类；

2、 应支持图像和图表的篡改分类；

6.2 文本图像篡改定位

文本图像篡改定位系统应支持以下功能：

1、 支持图像是否为篡改分类；

2、 支持图像篡改区域定位。

其中：

- 分类指标为Acc、FPR、Recall，计算方法见10.1；
- 定位的指标指标为mIoU，计算方法见10.1。

性能要求见10.2。

依据文本图像类型，检测内容可进一步划分。

6.2.1 证件篡改定位

证件文本图像篡改定位应支持以下内容：

1、 应支持文字和数字内容篡改定位，包含但不限于时间、编码、身份信息等内容；

2、 应支持图像区域篡改定位，包含但不限于人像、印章等内容；

3、 应支持二维码、条形码篡改定位；

4、 应支持字体和排版篡改定位。

6.2.2 文档篡改定位

文档文本图像篡改检测应支持以下内容：

1、 应支持文字和数字内容的篡改定位；

2、 应支持图像和图表的篡改定位；

6.2.3 截图篡改定位

截图文本图像篡改检测应支持以下内容：

1、 应支持文字和数字内容的篡改定位；

2、 应支持图像和图表的篡改定位。

7 测试数据集

7.1 测试数据集的标注和格式

数据集标注分为分类标签和定位标签两种：

1、 分类标签支持：

A. 二值标签：一种通用的分类标注方式，将真实文本图像标注为值1，将篡改文本图标注为值0。

B. 多标签：根据实际应用场景，如用户希望知道篡改文本图像使用了哪些篡改手段，可使用多标签，以标注文本图像。对于真实文本图像，将无篡改类别标注为值1，其余标注为0。对于篡改文本图像，将使用的篡改类别标注为值1，其余标注为值0。标注示例如下表：

表2 数据集多标签标注示例

文本图像	无有篡改	贴片篡改	涂抹篡改		复制移动篡改	拼接篡改
真实图像	1	0	0	...	0	0
篡改图像	0	1	0		1	0

2、 定位标签：（应该和分类标签共同标注，分一级标签和二级标签）

A、应支持掩码：掩码是一个二值图像，与对应的文本图像具有相同的大小；文本图像中篡改区域在掩码中对应的数值为1，未篡改区域对应的数值为0；（示例见附录10）

B、应支持边界框：用于描述篡改区域的位置，每一个篡改区域的位置由四个数值[x1,y1,x2,y2]组成，分别表示篡改区域的左上角横坐标、左上角纵坐标、右下角横坐标和右下角纵坐标。（示例见附录11）

7.2 测试数据集的难度和多样性

为了确保文本篡改检测技术能落地于应用场景中，数据集要求如下：

1、 样本分辨率：数据集可包含大小不一的样本，图像主体分辨率最小为64×64像素。不设最大分辨率；

2、 背景复杂度：背景包含具有干扰信息，如文字、人像等；

3、 遮挡：可遮挡图像中主体的部分区域，但不能完全遮挡；

4、 光照变化：包含光照变化差异大的图像；

5、 图像主体形变：包含应拍摄、扫描角度，或纸质文本存在弯曲或弯折，从而导致文本图像主体形变的样本。

为了应对不同的应用场景，需确保数据集的多样性：

1、 样本数量：数据集中包含的样本数量越多，涵盖的情况就越多，数据集的多样性就越高。

2、 样本类别：数据集中包含的样本类别越多，涵盖的情况就越多，数据集的多样性就越高。例如，数据集可包含现存的所有篡改类型的样本。

3、 样本变化：数据集中包含的样本变化越多，涵盖的情况就越多，数据集的多样性就越高。例如，文本图像获取方式，文本图文类型，光照和色彩变化，图片背景变化等。

4、 样本来源：数据集中包含的样本来源越多，涵盖的情况就越多，数据集的多样性就越高。例如，从公开数据集获取样本，从合成应用工具生成数据等。

7.3 数据集的公开性和可重复性

为了保证研究的可信度和可重复性，数据集应该具有公开性和可重复性。可公开与重复使用的数据应符合以下要求：

1、 数据集的许可证：数据集的许可证应该明确规定数据集的使用方式和条件，以便其他人可以了解数据集的使用限制和要求。

2、 数据集的格式和文档：数据集应该提供详细的文档和说明，以便其他人可以了解数据集的结构和内容。同时，数据集应该以标准格式或常见格式进行发布，以便其他人可以方便地使用和处理数据。

3、 数据集的访问和下载：数据集应该提供公开的访问和下载方式，以便其他人可以方便地获取数据集。同时，数据集的下载应该是可重复的，以便其他人可以在相同的条件下重复实验和验证结果。

4、 数据集的更新和维护：数据集应该定期进行更新和维护，以便保持数据集的完整性和准确性。同时，数据集的更新应该具有可追溯性，以便其他人可以了解数据集的变化和更新内容。

8 应用丰富度

8.1 类型完备度

文本图像篡改检测应符合以下要求：

1、 应完全支持拍照、文档、截屏三种文本图像获取方式；

2、 每种文本图像类型建议支持五种以上文本图像篡改检测，如证件照，建议至少支持身份证、港澳通行证、护照、驾照和营业执照的文本图像篡改检测。

3、 建议至少支持两种以上语言的文本图像篡改检测。

9 系统成熟度

9.1 易用性

文本图像检测系统应支持以下需求：

1、 无代码操作：用户无需操作代码，仅需在Web页面或端侧上交互以完成文本图像篡改检测相关工作；

2、 可视化面板：系统应支持结果可视化，处理状态可视化。将处理进度、检测结果显示在Web页面或移动端；

3、 文档输出：系统可支持将检测结果以图片或文本信息方式导出。

9.2 安全性

文本图像检测系统应满足以下要求：

1、 隐私保护：确保篡改检测系统对于用户隐私信息的保护，不泄露或滥用用户上传的图像和数据。用户在上传图像前，应当被告知其数据将如何使用并获得用户的知情同意。系统应提供详细的知情同意书，明确说明数据的收集、处理和存储方式，以及用户的权利和数据安全保障措施。

2、 鲁棒性：系统应具备对不同图像质量和不同攻击类型的鲁棒性，以确保在各种条件下都能进行准确的篡改检测。不同攻击类型包括：缩放、压缩、软件传输等。

3、 可信度评估：系统应能够对检测结果进行可信度评估，判断检测结果的准确性和可靠性。

4、 对抗性防御：针对对抗性攻击，系统应具备相应的防御机制，能够有效识别和抵御对抗性篡改，包括对抗训练、异常检测、模型增强等。

5、 安全通信：确保图像传输和处理过程中的安全性，采用加密技术和安全通信协议，防止数据泄露和中间人攻击，具体可参考安全通信标准EN50159。

6、 跨平台兼容性：系统应支持在不同平台上进行图像篡改检测，包括PC、移动设备和云端等，保证跨平台的兼容性和一致性。

7、 审计与日志记录：系统应记录用户操作、图像处理过程和检测结果，以便进行安全审计和追踪。

8、 用户认证和权限控制：确保只有授权用户才能访问和使用篡改检测系统，防止未经授权的访问和操作。

10、更新与维护：及时更新系统，修复漏洞和错误，以保持系统的稳定性和安全性。

9.3 产品部署

1、 安全性要求：确保在部署过程中和运行时能够保护用户数据和隐私。防止安全漏洞和攻击，包括网络攻击、数据泄露等。

2、 性能要求：在不同硬件平台和网络环境下应具有稳定的性能。对于计算密集型任务，需要优化算法和硬件选择，以提高计算速度。根据不同应用场景，可选择单一文本图像检测类型，以确保检测精度。

3、 可伸缩性要求：应该能够在不同规模的用户量下保持良好的性能。考虑并发用户数、数据量等因素，确保产品能够水平扩展。

4、 硬件支持：应该能够在不同的硬件设备上运行，如CPU，GPU，Atlas服务器处理器等。

5、 跨平台支持：应该能够在不同操作系统和设备上运行，如Windows、Linux、iOS、Android等。确保在不同浏览器、手机型号等下都能正常展示和运行。

6、 合规性要求：针对特定行业或地区的法规和规定，满足相应的合规性要求。

10 评价标准

10.1 评价指标

10.1.1 文本图像篡改分类指标

采用测评数据对检测系统进行测试，统计测评数据的TP、TN、FP、FN值：

1、 TP：无篡改文本图像预测正确数量；

2、 TN：篡改文本图像预测正确数量；

3、 FP：无篡改文本图像预测错误数量；

4、 FN：篡改文本图像预测错误数量。

10.1.1.1 准确率（Accuracy）

在特定数据集中检测所有文本图像准确程度的测量指标，Acc指标计算如下：

$$Acc = \frac{TP + TN}{TP + FP + FN + TN} \times 100\%$$

10.1.1.2 误检率（False Positive Rate）

在特定数据集中无篡改文本图像误检程度的测量指标，FPR指标计算如下：

$$FPR = \frac{FP}{TP + FP} \times 100\%$$

10.1.1.3 召回率（Recall）

在特定数据集中检测篡改文本图像准确程度的测量指标，Recall指标计算如下：

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

10.1.2 文本图像篡改定位指标

采用测评数据对检测系统进行测试，统计预测的篡改区域与实际篡改区域的IoU值（检测结果和真值标注之间交集面积占并集面积的比例）：

$$\mathrm{IoU} = \frac{\mathrm{A} \cap \mathrm{B}}{\mathrm{A} \cup \mathrm{B}}$$

式中：

A ——预测的篡改区域；

B ——实际的篡改区域；

然后计算mIoU：

$$\mathrm{mIoU} = \frac{1}{\mathrm{N}} \sum_{\mathrm{n}=1}^{\mathrm{N}} \mathrm{IoU}_{\mathrm{n}}$$

式中：

N ——测评数据集中篡改区域总数量；

10.2 性能要求

运行速度：对于不超过10M的单张图像处理时间≤2秒。

检测精度：文本图像篡改检测精度要求如下，对抗攻击方式见6.2：

表3 文本图像篡改分类精度要求

功能	对抗攻击	文本图像类型系统	准确率	误检率	召回率
文本图像篡改分类		单一证件类型	99.5%	0.5%	99.5%
	√		99%	1.0%	99%
		单一文档类型	99.5%	0.5%	99.5%
	√		99%	1.0%	99%
		单一截图类型	99.5%	0.5%	99.5%
	√		99%	1.0%	99%
		通用证件类型	99%	1.0%	99%
	√		98%	1.0%	98%
		通用文档类型	99%	1.0%	99%
	√		98%	1.0%	98%
		通用截图类型	99%	1.0%	99%
	√		98%	1.0%	98%
		通用类型	95%	1.0%	95%
	√		95%	1.0%	95%

表4 文本图像篡改定位精度要求

功能	对抗攻击	文本图像类型系统	均值交并比
文本图像篡改定位		单一证件类型	0.95
	√		0.95
		单一文档类型	0.95
	√		0.95
		单一截图类型	0.95
	√		0.95
		通用证件类型	0.85
	√		0.85
		通用文档类型	0.80
	√		0.80
		通用截图类型	0.80
	√		0.80
		通用类型	0.70
	√		0.70

10.3 测评方法

1、 测评数据量要求：

A. 不同文本图像类型（证件、文档和截图）的文本图像均不小于1000张；

B. 同一文本图像类型下，至少包含五个不同类型数据，每种数据不少于300张。

C. 不同对抗攻击方式（物理攻击、数字攻击，见6.2）的文本图像均不小于300张；

D. 不同篡改方式（物理篡改、数字图像篡改，见6.1）的文本图像均不小于300张。

2、 测评数据要求：

A. 测评文本图像分辨率不小于64×64像素。

3、 测评方法：

A. 统计测评数据的检测结果，根据11.1计算文本图像篡改分类与定位指标；

B. 统计单张文本图像在各个硬件支持下的检测时间；

C. 达到11.2所示的要求。

附录

（规范性附录/资料性附录）

1、热力图输出：预测的篡改区域与未篡改区域由明显的颜色差异，输出样式示例如下，预测的篡改区域由蓝色、绿色、黄色、橙色和红色表明。预测的篡改区域越接近红色表明该区域篡改的可能性越高，越接近蓝色表明该预测篡改的可能性偏小。

图1 从左到右分别为输入文本图像、文本图像篡改检测输出样例

2、边界框输出：在输出图片中，采用边界框（通常为矩形框）标注篡改区域，输出样式示例如下，红色矩形框下标有篡改区域预测置信度，置信度越高该区域篡改的可能性越高。

图2 从上到下分别为输入文本图像、文本图像篡改检测输出样例

3、物理遮挡篡改示例如下，红色矩形框内区域为贴片篡改：

图3 贴片篡改示例图

4、复制移动篡改示例如下，其中掩码为二值图像，用于标注篡改图像的篡改区域：

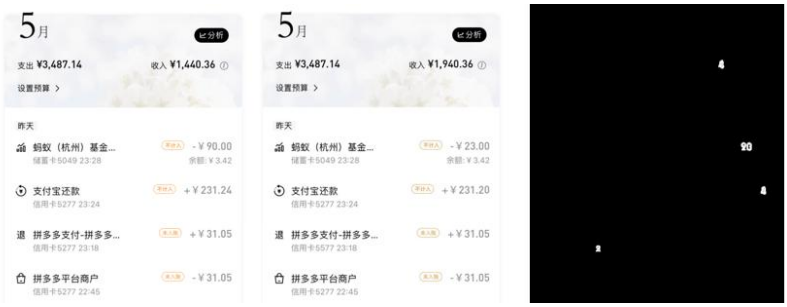

图4　复制移动示例，从左到右分别是原始图像、篡改图像和掩码

5、拼接篡改示例如下：

图5 拼接示例，从左到右分别是原始图像、原始图像、篡改图像和掩码

6、擦除篡改示例如下：

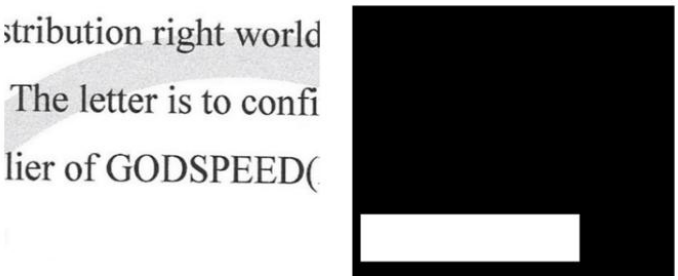

图6 擦除示例，从左到右分别是篡改图像和掩码

7、添加生成篡改示例如下：

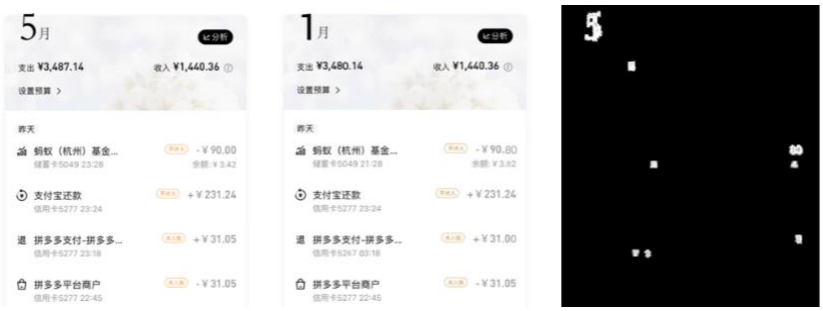

图7 添加生成，从左到右分别是原始图像、篡改图像和掩码

8、几何变形攻击示例如图所示。

图8 从左到右分别是原始文本图像、压缩攻击后的文本图像，旋转攻击后的文本图像

9、文本区域模糊攻击示例如图所示。

图9 从左到右分别是原始文本图像、文本区域模糊攻击后的文本图像

10、掩码的示例如下图，掩码中白色部分表示数值为1，黑色部分表示数值为0。

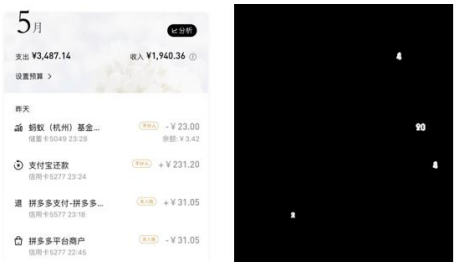

图10 掩码标签示例，从左到右分别为文本图像和对应的掩码

11、边界框标签示例如下图。

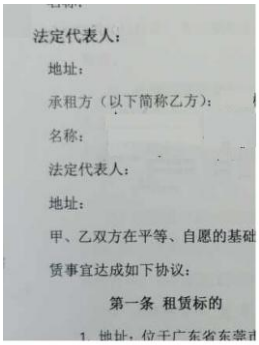

image_00001.jpg

JSON文件

{"name": image_00001.jpg,
"region":[33.0, 26.0, 149.0, 51.0]}

{"name": image_00002.jpg,
.....

图11 边界框标签示例，从左到右分别为文本图像和对应的边界框标签

12、EXIF信息检测

可交换图像文件格式（Exchangeable image file format，Exif），是专门为数码相机的照片设定的文件格式，可以记录数码照片的属性信息和拍摄数据。 Exif 可以附加于 JPEG、TIFF 等文件之中，为其增加有关数码相机拍摄信息的内容和索引图或图像处理软件的版本信息。其信息示例如表 2 所示：

表 5 Exif 信息

Exif 定义名	中文定义名	备注
ImageDescription	图像描述	
Artist	作者	
Make	生产商	
Model	型号	相机型号
Orientation	方向	
XResolution	水平方向分辨率	
YResolution	垂直方向分辨率	
ResolutionUnit	分辨率单位	

Software	软件	
.....		
DateTime	日期和时间	照片最后修改时间
DateTimeOriginal	拍摄时间	照片拍摄时间
DateTimeDigitized	数字化时间	照片被写入时间

文本图像篡改检测系统支持检测 EXIF 信息，并将相关信息输出给用户，作为用户参考文本图像是否存在篡改的依据。输出的 EXIF 信息需包含：作者、软件、日期和时间和数字化时间。

13、系统流程

文本图像篡改检测系统流程如下：

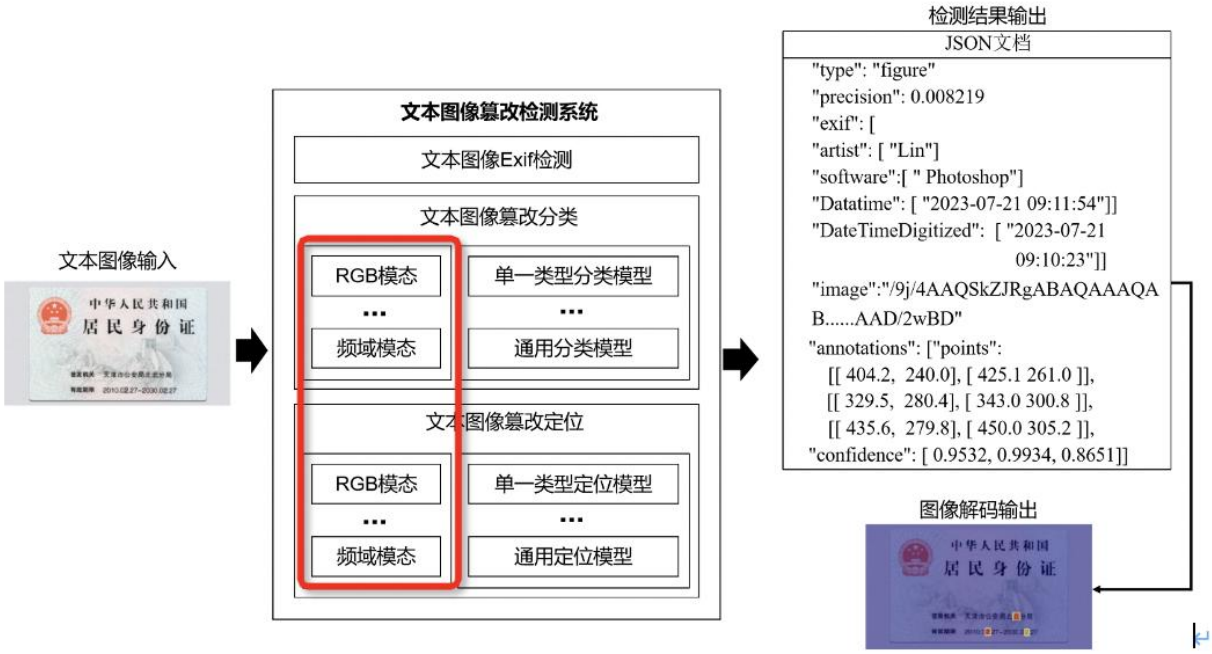

图12、文本图像篡改检测系统流程

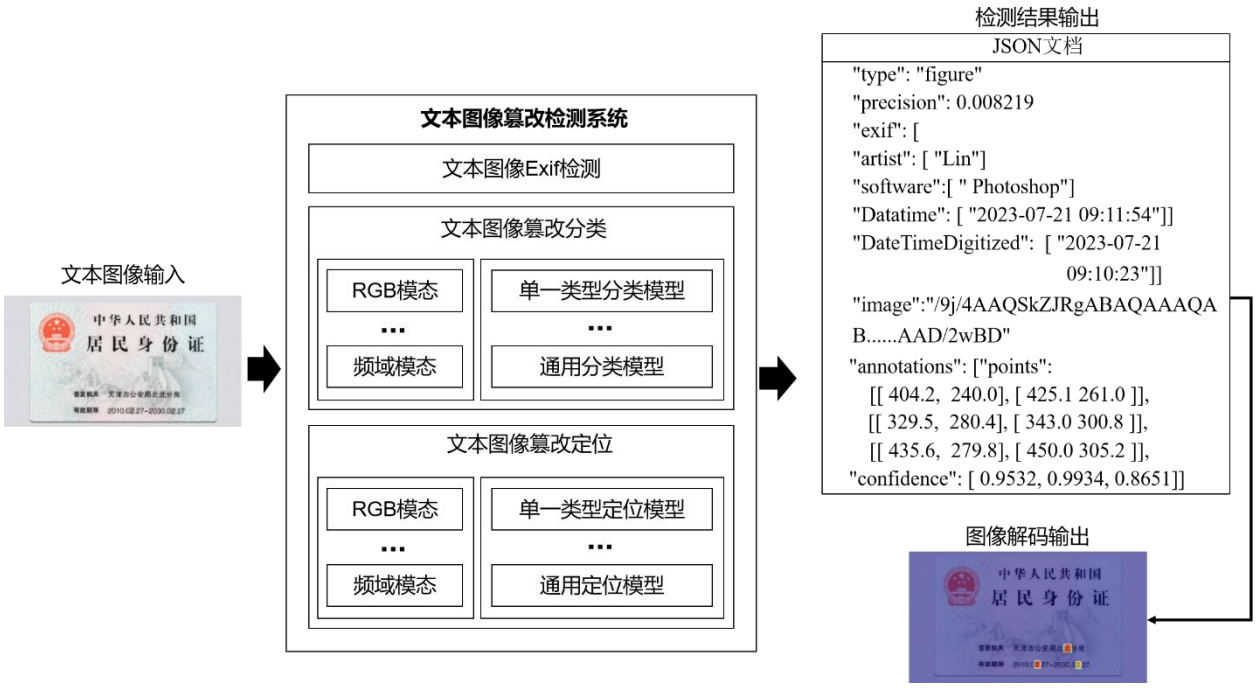

图 13 文本图像篡改检测系统流程

系统流程包含文本图像输入、文本图像篡改检测系统和检测结果输出。

1、 文本图像输入要求见5.1；

2、 文本图像篡改检测系统应包含文本图像Exif信息检测、文本图像篡改分类和文本图像篡改定位三大功能。

Exif信息检测：检测文本图像是否使用过图像编辑软件并获取相应信息；

文本图像篡改分类：检测图像整体是否存在篡改，要求见6.1；

文本图像篡改定位：对于存在篡改的文本图像，定位篡改区域，要求见6.2；

3、 检测结果输出如图1中JSON内容，要求见5.2。

13、系统流程

文本图像篡改攻击方式分为物理攻击和数字攻击两大类，篡改检测系统应适配以下攻击以避免误判。

13.1图像篡改类型

文本图像篡改包括如下物理篡改和数字篡改。

13.1.1物理篡改

文本图像篡改检测应支持以下物理篡改类型：

1、 物理遮挡通过贴片、涂抹等活动对检测主体部分内容进行篡改。（示例见附录3）

2、 物理拼接：通过将多个纸质文本拼接以生成新的文本。

13.1.2数字篡改

数字图像篡改检测应支持用计算机技术，包括但不限于软件、深度学习模型等方式，生成的以下篡改类型：

1、 复制移动：将文本图像中一个或多个区域复制并移动到另一个区域。（示例见附录4）

2、 拼接：将文本图像中的一个或多个区域拼接到另一张文本图像中。（示例见附录5）

3、 擦除：将文本图像中的一个或多个区域删掉，并用视觉上和谐的内容填充。（示例见附录6）

4、 擦除添加：将文本图像中的一个或多个区域擦除后，重新添加生成字体、大小等相似的不同文本或图像上去。

5、 添加生成：直接在文本图像不同目标区域添加生成文本或图像。（示例见附录7）

6、 深度合成：将文本图像通过图层合成或AIGC生成的方式进行信息修改。

13.2文本图像篡改攻击

13.2.1物理攻击

篡改检测系统应能对抗以下物理攻击：

1、 光学干扰攻击：利用光线投射、反射等手段以改变检测主体的光照条件，从而影响文本图像的质量；

2、 翻拍攻击：通过拍摄电子设备上显示的图像，以生成新的文本图像；

3、 材质攻击：通过用不同材质的物体进行遮挡，从而影响文本图像的质量；

13.2.2数字攻击

篡改检测系统应能对抗以下数字攻击：

1、 JPEG压缩、VAE压缩等压缩攻击方式：因有损压缩导致文本图像信息丢失。（示例见附录8）

2、 几何变形攻击方式：通过包含但不限于水平翻转、旋转、裁切、尺度变换、广义几何变形等方式改变文本图像的原有的统计特性。（示例见附录8）

3、 图像增强处理攻击方式：通过包含但不限于低通滤波、锐化、直方图修正、Gamma矫正、颜色量化、复原等图像增强处理改变文本图像的色彩、对比度。

4、 噪声攻击：通过添加包含但不限于高斯噪声、椒盐噪声或周期性噪声等噪声以模糊图像细节，混淆文本内容。

5、 文本区域模糊攻击：通过对文本区域进行局部模糊，以掩盖篡改的痕迹。（示例见附录9）

6、 在社交工具传输中发生的文本图像变化：文本图像在经过社交工具传输通常会经过压缩和解压流程、压缩和尺度变换等操作，可能会产生人眼无法察觉的变化。

7、 对抗攻击：通过对样本进行微小改动生成新的样本，来对抗篡改器的分类和检测。