JOURNAL OF IMAGE AND GRAPHICS

主办：中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2022
01
VOL.27

ISSN1006-8961
CN11-3758/TB

数字图像/视频内容安全

中国图象图形学报

刊名题字：宋健 | 月刊（1996年创刊）

主管单位 中国科学院
主办单位 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

主　　编 吴一戎
编辑出版 《中国图象图形学报》编辑出版委员会
通信地址 北京市海淀区北四环西路19号
邮　　编 100190
电子信箱 jig@aircas.ac.cn
电　　话 010-58887035
网　　址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号
总 发 行 北京报刊发行局
订　　购 全国各地邮局
海外发行 中国国际图书贸易集团有限公司
（邮政信箱：北京399信箱　邮编：100048）
印刷装订 北京科信印刷有限公司

第27卷第1期（总第309期）
2022年1月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

Journal of Image and Graphics

Title inscription: Song Jian | Monthly, Started in 1996

Superintended by Chinese Academy of Sciences
Sponsored by Aerospace Information Research Institute, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Wu Yirong
Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics
Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China
Zip code 100190
E-mail jig@aircas.ac.cn
Telephone 010-58887035
Website www.cjig.cn

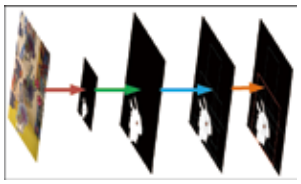
Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation (P.O.Box 399, Beijing 100048,P.R.China))
Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB
ISSN 1006-8961
CODEN ZTTXFZ

国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元

2022年1月 第27卷 第1期（总第309期）
中国图象图形学报
Zhongguo Tuxiang Tuxing Xuebao

目次



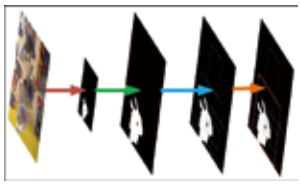
检测小篡改区域的U型网络(第0176页)

结合半张量积压缩感知的可验证图像加密(第0215页)

实离散分数Krawtchouk变换及其在数字图像水印中的应用(第0252页)

Volume 27, Number 1
Published January 16, 2022

CONTENTS
JOURNAL OF IMAGE AND GRAPHICS



U-Net for detecting the small forgery region(P0176)

Semi-tensor product compression sensing integrated to verifiable image encryption method(P0215)

Real discrete fractional Krawtchouk transform with an application to image watermarking(P0252)

E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035

中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS

中图法分类号:TP391　　**文献标识码**: A　　**文章编号**: 1006-8961(2022)01-0262-15

论文引用格式: Sun S, Zhang W M, Fang H and Yu N H. 2022. Automatic generation of Chinese document watermarking fonts. Journal of Image and Graphics,27(01):0262-0276(孙杉,张卫明,方涵,俞能海. 2022. 中文水印字库的自动生成方法. 中国图象图形学报,27(01):0262-0276)[DOI: 10.11834/jig.200695]

中文水印字库的自动生成方法

孙杉,张卫明*,方涵,俞能海

1. 中国科学技术大学网络空间安全学院, 合肥　230027; 2. 中国科学院电磁空间信息重点实验室, 合肥　230027

摘　要: **目的**　文档水印技术是一种用以解决文档泄密溯源的信息隐藏技术。传统的基于字库的文档水印方案需要手动生成字库,极大地影响了水印的使用效率。为此本文设计了一种基于自动生成字库的鲁棒文档水印方案。**方法**　该方法由一个端到端的编码—解码器结构的自动字库生成网络、一个字符筛选嵌入端和一个神经网络提取端组成,可自动完成变形字库的生成,而后进行水印的嵌入和提取。为了抵抗传输过程中可能存在的失真,在编码器和解码器之间加入可导噪声层用以模拟失真过程,使得水印模型获得对应的鲁棒性。**结果**　本文方法在含252个中文字符的真实文档中嵌入252 bit水印信息,与其他文档水印方法的视觉质量和鲁棒性进行了对比。结果表明,相对于现有的基于字符特征的中文文档水印方法,本文方法的峰值信噪比(peak signal to noise ratio,PSNR)、结构相似性(structural similarity,SSIM)和主观质量评分分别提升了11.68 dB、0.08和5.8%,说明其有更好的视觉质量。对于数字信道传输场景,本文方法达到了与其他方法大致相当的性能;对于打印扫描场景,本文方法在三号、四号、小四号和五号字体下的水印提取率分别提升了2.4%、3.07%、1.34%和0.02%,在打印、扫描分辨率失配的场景下也具有较好性能,说明其在抗打印扫描上具有更高的鲁棒性。**结论**　与基于人工设计字库的中文字符水印相比,本文方法充分利用了字符的几何特征并且能够自动生成字库,降低了中文文档水印方案的复杂度。

关键词:文档水印;深度学习;中文字库生成;抗数字失真;抗打印扫描

Automatic generation of Chinese document watermarking fonts

Sun Shan, Zhang Weiming*, Fang Han, Yu Nenghai

1. School of Cyber Security, University of Science and Technology of China, Hefei　230027, China;

2. Key Laboratory of Electromagnetic Space Information, Chinese Academy of Sciences, Hefei　230027, China

Abstract: **Objective** The copyright protection has been the hotspot with the amount of digital documents increased dramatically. In order to protect the document copyright and locate the source of the leaked document, watermarking technology innovation for documents has been widely focused on. The protection can be realized via adding invisible digital watermark information (e.g., device number, date, etc.) to the document. To realize the traceability of document leakage, the leaked source can be located by extracting the watermark from the document once the watermarked document is leaked. Meanwhile, the current watermarking technology can also act as a deterrent which effectively reduce the occurrence of the leaking events. The current document watermarking methods can be divided into five categories: document structure based methods, natural language processing based methods, grid pattern based methods, image based methods and font based methods. Among them, the font based methods guarantee the best performance in the view of robustness and transparency.

收稿日期:2020-12-14;**修回日期**:2021-02-04;**预印本日期**:2021-02-11

*通信作者:张卫明　zhangwm@ustc.edu.cn

基金项目:国家自然科学基金项目(62072421,U1636201,62002334)

Supported by: National Natural Science Foundation of China (62072421, U1636201, 62002334)

The main idea of such methods is representing the watermark information into the characteristics of the fonts (e. g., the size, shape or brightness) while the modified fonts maintain the high visual consistency with the original one. The robustness, transparency, capacity as well as the integrity can be achieved simultaneously. However, the existing font based methods need to design the modification features manually, and cannot automatically generate the new fonts. For the Chinese character system which contains a large number of characters, such methods will cause a labor cost workload and severely less efficiency. To overcome such drawbacks, this research proposes an automatic font generation based robust document watermarking scheme. **Method** The framework of such scheme is comprised of an end-to-end encoder-decoder structure automatic font generation network, a character selection embedder and a neural network based extractor. With the designed font generation network, the deformed font library is further utilized for embedding the watermark generated automatically. Meanwhile, a differentiable noise layer is complemented between the encoder and the decoder to simulate the distortion process in order to realize the robustness against different distortions, so that the encoder can learn better features to create the new font and the decoder can be trained to be adaptive to such distortions. This research designs a combined noise layer that can effectively simulate the common distortions via the common distortions in digital transmission channels (e. g. , screenshots, scaling, Gaussian noise and JPEG compression) . The whole font generation network consists of four parts: encoder, noise layer, decoder and adversarial recognizer. The encoder receives the watermark information and the carrier character image to generate the encoded character image. The noise layer adds noise to the encoded image to generate the noisy image. In particular, several of six simulated noise layers (identity mapping layer, scaling layer, translation layer, rotation layer, Gaussian noise layer and Gaussian blur layer) are randomly opted as a combined noise layer at each iteraction. The decoder receives the noisy image and outputs the corresponding watermark label. The adversarial recognizer tries to detect whether the current image is a carrier character image or an encoded one, which aids to improve the visual quality of the generated font. The encoder provides training samples for the extractor to ensure better extraction performance, and the extractor guides the generation direction of encoder to create better character images. The two coordinated modules make the generated font with higher visual quality and stronger robustness. Based on the well trained font generation, the network has generated the watermarked font library via feeding them with an original font library and different watermark signal. Each character in the font library can corresponds to different perturbation, which can be decoded to different watermark signal further. Hence, based on the generated font library, the corresponding character in the codebook is sorted out in the watermark embedding stage according to the current watermark information to replace the current character in the input document. The watermark information can be embedded into the whole original document and generate the watermarked document in this way. In the extraction stage, the whole document is firstly divided into some single character images by character segmentation after receiving the distorted watermarked document transmitted with digital channel. Each character is sent to the pre-trained decoder in the watermarking generation stage. The watermark information embedded in the current character can be accurately extracted. The characters in the document undergo the transformation distortion from digital signal to analog signal (A-D) process and from analog signal to digital signal (D-A) process, and the image quality is greatly reduced in the context of the print-scan scene. Simultaneously, such process that contained various image attacks cannot be accurately simulated by differentiable distortions, so the robustness against print-scanning distortions should be considered as a target. To achieve such robustness, this proposal is a fine-tuning scheme for the extractor which can effectively train the extractor to be adaptive to the print-scanning distortions. Specifically, the font generation model is fixed as a pre-training network and a set of following documents are embedded with watermarks based on the pre-trained embedder. The real print-scanning process is conducted on such documents to generate the distorted image library. Based on the distorted image and its watermark, the decoder is fine-tuned to be adaptive to the distorted features further. The robustness against print-scanning distortion can be achieved. **Result** This scheme embeds 252 bits of watermark into a real document containing 252 Chinese characters in comparison of the visual quality and robustness with other document watermarking methods. The results show that the peak signal to noise ratio (PSNR), structural similarity (SSIM) and subjective quality scores of the proposed scheme are higher with 11. 68 dB, 0. 08 and 5. 8% respectively, demonstrating the qualified visual quality of the proposed scheme. For the robustness, the watermark extraction rate of the scheme is 100% under the screenshot and scaling, and the performance under JPEG compression and Gaussian noise is approximately equivalent to that of

other methods. For the print-scan scene, the watermark extraction rates of the scheme in the font size of three, four, small four and five are realized 2.4%, 3.07%, 1.34% and 0.02%, respectively. The qualified performance is achieved under different mismatched printing and scanning qualities as well, which indicates that the scheme has higher robustness in resisting the print-scanning distortions. **Conclusion** Compared with the existing Chinese character watermarking methods based on the manually designed font library, the proposed scheme can automatically generate the tagged Chinese font library that is similar to the target font visually, which effectively reduces the complexity of the font generation. The experimental results show that the proposed document watermarking scheme has presented better visual quality and embedding capacity. In addition, the proposed scheme maintains the strong robustness against digital editing channel as well as the print-scanning channel. However, the scheme is not suitable yet for print-shooting and screen-shooting process at present. Future research will be concentrated on how to design robust document watermarking schemes for these two scenes.

Key words: document watermarking; deep learning; Chinese font generation; anti digital distortion; anti print-scanning distortion

0 引 言

数字水印是信息隐藏技术的重要分支,可以用于数字产品版权保护、原始数据的真伪鉴别、数据侦测与跟踪等。数字水印技术是利用多媒体文件的冗余性以及人类感知系统的不敏感性,对多媒体文件进行一些修改使得水印成为载体不可分割的一部分,使人们很难分辨出原始和嵌入水印后的多媒体文件之间的区别。同时,在嵌入水印后的多媒体文件遭受一些故意或非故意的攻击后,仍能从失真文件中提取出水印信息或是证明水印信息的存在。根据水印嵌入的载体形式,可以将数字水印技术分为图像水印(Fang 等,2019;Su 和 Chen,2018;陈怡 等,2019;邓小鸿 等,2014)、视频水印(Kelkoul 等,2018;李淑芝 等,2015)和文本水印(Brassil 等,1995)等。本文主要研究文本水印技术,旨在解决文本的泄密溯源问题。

目前,很多重要信息,如各类公文、方案和报告等都是通过文本形式传播。随着无纸化办公的广泛应用,如何保护这些文档不被泄露成为一个重要的安全问题。现存的通过加密、权限控管和身份认证等手段对文件进行安全管理的方法已经比较完善。但是一旦文档被截屏或打印到纸质介质上时,就很难对泄密行为进行监管,当文档遭到泄露后,相关负责人员很难找到泄露源头。为了解决泄密溯源的问题,本文使用文档水印方案,通过向文本中添加不可见的数字水印(如设备号、日期等),一旦含水印文本被泄密,即可从泄密文件中提取出水印信息,从而实现文档信息的溯源。这种溯源能力还可以起到震慑作用,从而减少泄密事件的发生。

由于文档载体相对于其他载体冗余信息要小得多,在文档中嵌入不可见水印比较困难。目前的文档水印方法主要可以分为以下几类:

1)基于文档结构的文档水印算法(Brassil 等,1995,1999;Huang 和 Yan,2001)。这类方法是最早的一类文档水印方法,通过轻微地调整文档结构(如字间距、行间距等)来实现水印数据的嵌入。但这类方法无法抵抗简单的几何变化攻击,而且嵌入的信息量少。

2)基于自然语言处理的文档水印算法(Atallah 等,2001;张宇 等,2005)。这类方法通过调整句子的语法结构,或者通过等价信息替换而不改变句子的意义完成水印嵌入。Atallah 等人(2001)根据具体的水印信号对语句结构的顺序进行调整来插入水印。张宇等人(2005)通过修改语义和句法结构(如同义词替换、主动式变被动式等)嵌入水印。这类方法的鲁棒性非常强,但水印的嵌入过程改变了文档的内容,破坏了文档的完整性,对于政府机构文件、法律文件等使用场景来说是不可接受的。

3)基于网格图案的文档水印算法(Suzaki 和 Suto,2005;Takahashi 等,2008;Briffa 等,2010)。这类方法在文档上叠加一层底纹代表水印信息,可以用于任何种类的文本文档,嵌入容量很大,也可抵抗多次复印攻击。但是这类方法大多属于可见水印(Suzaki 和 Suto,2005;Takahashi 等,2008),视觉效果较差;而少数的属于不可见水印(Briffa 等,2010),鲁棒性较差。

4)基于图像的文档水印算法(Wu 和 Liu,2004;亓文法 等,2008;艾平,2014)。该类方法首先将文

档转化成文本二值图像,以像素翻转的方式嵌入水印,主要应用于抗打印扫描场景。但这类方法只能针对二值文档图像,而且很容易导致笔画边界不平整,降低了文档的视觉质量。此外,该类方法无法实现文档水印的实时嵌入。

5)基于字符几何特征的方法(刘东 等,2007;Xiao 等,2018;Qi 等,2019)。这是一类更为直观的方案,通过修改字符的大小、形状和亮度等特征表达水印信息。随着深度学习的飞速发展,出现了一些基于深度学习方法的水印方案。FontCode(Xiao 等,2018)利用字形流形学习生成新的字体代表水印编码,再使用卷积神经网络(convolutional neural networks,CNN)对水印模式进行识别。但是该方法只适用于字母和数字,无法生成具有复杂字形的字体。Qi 等人(2019)针对中文字符特征手工建立特殊的字体库,利用模板匹配实现对水印模式的提取。该方法有很好的视觉质量和很高的鲁棒性,但是需要手动设计字库的修改模式,对于使用中文这种具有庞大数量字符的文字系统来说工作量是巨大的,给实际使用带来了很大的不便。

各类文档水印方法的优缺点如表 1 所示,从表中可以看出,基于字符几何特征的文档水印方法在视觉效果、鲁棒性和容量这 3 个方面达到了较好的折中,而且不会改变文档的内容,适合用于文档溯源场景。但是目前现存的基于字符几何特征的中文文档水印还需要人工设计修改特征,不能自动生成字符。为了充分利用字符的几何特征并且能够自动生成字库,本文提出了一种基于自动生成字库的中文文档水印方案。该方案的自动生成字库模块主要由一个端到端的编码—解码器结构的神经网络组成。首先编码器利用字符图片生成变形字体图像,期间利用掩码指导生成函数使得生成的字符扰动集中到字符轮廓上。而后根据生成的字体图像转化为字库,并通过水印中不同的比特信息进行字符的选择完成水印的嵌入,同时利用在形变字体生成时训练好的解码端提取水印信息。与现有的方法相比,本方案最终生成的含水印文档具有较好的视觉质量,其水印提取率在各种数字信道失真场景和打印扫描场景下均保持很高的鲁棒性和稳定性。综上所述,本文主要有以下 3 点贡献:

1)提出了一种端到端网络框架,由一个编码—解码器联合网络组成,能自动地生成与目标字体相近的中文变形字体图像,以便后续转化成字库,同时能够实现水印的嵌入和消息的提取。

2)提出了一种掩码指导失真函数的方案,有效提升了生成变形字体字库的视觉质量。

3)实验证明了本水印方案在数字噪声信道和打印扫描场景下的溯源有效性。

表 1　各类文档水印方法比较

Table 1　Comparisons of various document watermarking methods

文档水印方法	视觉效果	鲁棒性	容量	完整性	实时性	复杂度
基于文档结构	√	×	×	√	√	√
基于自然语言处理	√	√	√	×	√	√
基于网格图案	×	√	√	√	√	√
基于图像	×	√	√	√	×	√
基于字符几何特征	√	√	√	√	√	×
本文方法	√	√	√	√	√	√

注:“√”表示该方法效果较好,“×”表示该方法效果不好。

1　相关工作

1.1　传统的抗打印扫描文档水印方法

传统的抗打印扫描文档水印方法大多是将文档转化为二值图像再进行水印的嵌入。由于打印扫描设备型号、老化程度和人工操作的不同,文档图像将会产生不同程度的失真。为了实现抗打印扫描水印算法,需要找到一种打印扫描前后都不受影响的特征值,即利用打印扫描前后的不变量来指导抗打印扫描水印算法的设计与实现。亓文法等人(2008)通过大量实验找到了打印扫描过程的不变量,即当前字符黑色像素点数量与平均字符黑色像素点数量之间的比值 μ ,并以这个特征量作为抗打印扫描文档水印算法中嵌入水印的依据。

艾平(2014)在 MinWu(Wu 和 Liu,2004)像素翻转模型的基础上,提出了一种考虑像素块 3 种度量因素(平滑度、连通性和距离度量)的加权像素反转模型。根据 3 种度量因素来确定像素翻转策略,并制定像素翻转性分数的打分策略。通过翻转分数

比较高的像素点,使得打印扫描不变量 μ 满足某种约束关系,这里使用奇偶量化将 μ 修改为量化步长 K 的奇数倍或偶数倍来代表水印编码。在水印提取过程中,将含水印的文本图像打印到纸质介质上后,再对其进行扫描,得到灰度文档图像。经过字符切分后计算出打印扫描不变量 μ',最后通过奇偶量化对嵌入区域中的字符进行水印提取。

这类方法是通过像素翻转实现文档特征值的修改,因此在水印嵌入前需要将文档转化为二值图像,这直接降低了文档的图像质量。在可翻转像素点不足的情况下,文档嵌入水印后也会产生明显的笔画断连和不平整。

1.2 基于字符几何特征的英文文档水印方法

在 FontCode(Xiao 等,2018)中,作者根据水印信息对英文字符的几何特征进行一些微小的扰动生成水印文档。嵌入端字符扰动的方法是基于字符结构实现的(Campbell 和 Kautz,2014),而提取端是通过对扰动模式的识别进行水印信息恢复。字符扰动模型包括字符匹配、流形生成和字体生成 3 个部分。利用现有字体的多样性,该模型将所有不同的字体中的每个字符单独进行轮廓匹配,而后根据这些高维轮廓特征生成低维流形。字型流形是一种利用高斯过程潜在变量生成模型构成低维空间的映射,同时也能通过对低维空间的修改映射回高维特征,进而推断和获得新的字体。

在 FontCode 水印嵌入的预处理阶段,首先利用多种字体库生成字体生成模型中的低维流形,并且为常用字体构造一个扰动符号的水印码表。然后根据水印信息选择水印码表中的扰动类型,对原始的文本文档进行字符替换从而完成水印的嵌入。在含水印文档图像经过跨媒介传输后,通过光学字符识别(optical character recognition,OCR)检测得到每个字符的边界框,完成字符分割。而后将单个字符图像送入一个简单的卷积神经网络得到当前水印标签,组合所有的水印标签即可提取出完整的水印信息。

这个方法适用于大多数文档类型,而且人眼几乎分辨不出字符修改的扰动,对各种跨媒介传输都有很好的鲁棒性。但是此方法只能适用于具有简单轮廓的字符集合,比如字母或数字。对于类似中文这种具有复杂形状的字符,此方法无法生成高质量的字体流形进而嵌入水印信息。

1.3 基于字符几何特征的中文文档水印方法

由于字体流形生成矢量字库方法的局限性,上述基于英文字符几何特征的文档水印方法难以应用于中文字符。Qi 等人(2019)提出了一种抗打印扫描的中文文档水印方法。在水印嵌入的预处理阶段,该方法通过适当改变字符的拓扑结构,设计字符生成特殊矢量字体库,并根据每个不同的字形代表不同的水印信息位建立了水印码表,生成对应的字符模板。在打印文本文档时,该方法通过终端监控服务程序拦截文本文档内容,而后根据要嵌入的水印信息序列,动态替换含编码的字符嵌入水印信息。在经过跨媒介传输后,对文档进行字符切割,而后利用快速归一化互相关方法(Hii 等,2006)对当前字符和模板字符进行模板匹配得到水印信息。

这类方法的性能依赖于变形字库的生成过程,既要保证人眼对字符的几何改变不敏感,又要保证文档在跨媒介传输后仍具有鲁棒性,所以生成高质量的中文字库至关重要。而设计一款中文字体,至少需要六千多个字,甚至能达到一万多字。若仅对常用字生成含水印字符,会限制该类方法的水印容量和适用范围。数量庞大的汉字给中文字库的制作带来了巨大挑战,同时字体的设计和操作需要大量的专业知识,这将耗费大量的人力物力。为了解决这一问题,本文提出了一种可以自动生成变形字库的中文文档水印方法,利用神经网络强大的学习能力自动指导字符扰动的方向,从而实现端到端的水印嵌入和提取。

2 基于自动生成字库的中文文档水印方法

本文提出一种基于变形字库的中文文档水印方案,包括 3 个阶段。第 1 阶段是水印字库生成与解码器训练阶段,第 2 阶段是水印嵌入阶段,第 3 阶段是水印提取阶段,具体流程如图 1 所示。在第 1 阶段中,本文提出了一种基于深度神经网络的变形字体生成网络,能自动生成中文变形字体并保证文字特征在数字和打印扫描信道下的鲁棒性。利用编码—解码器网络生成字符图像,而后通过一些专业的转换工具将字符图像转为字库格式的文件类型(如 ttf、ttc 文件等),从而实现字库的生成过程。生成字库的同时训练解码器作为水印提取端。在此过程中,本文在编码—解码器联合网络之间加入噪声层以保证文档水印方法的鲁棒性。在第 2 阶段水印嵌入时,

将原始文档分割为单字符，依据水印的信息实现字库中文字的选择，生成含水印的文档。在第 3 阶段中，含水印的文档经过截屏或打印扫描后，通过文档校正和字符分割，将字符图像送入阶段 1 中训练好的解码器进行水印信息的提取。文档在泄露后，一旦被捕获，就能提取出水印从而实现信息来源的追踪。

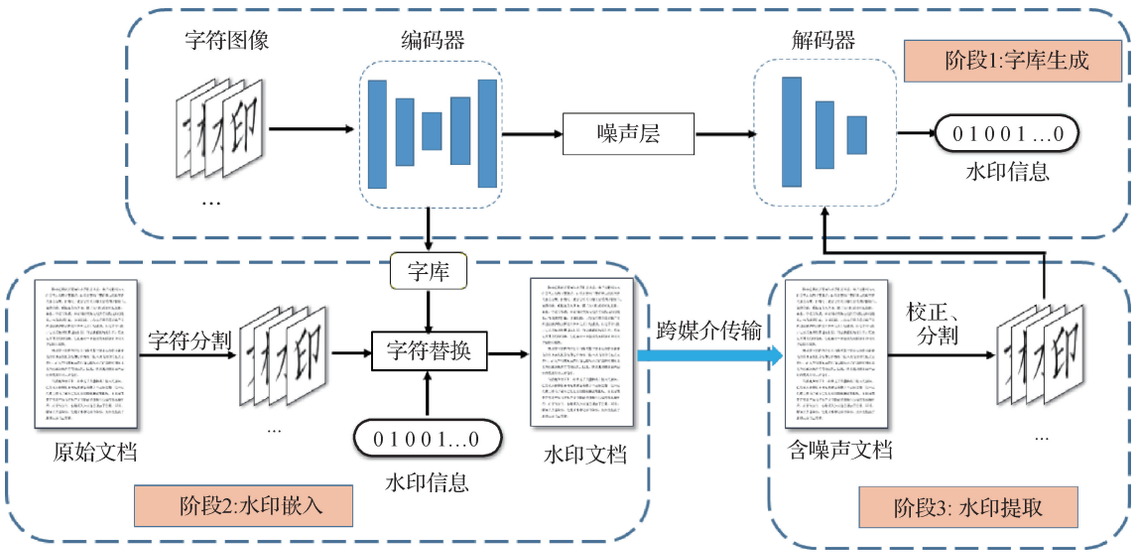


图 1　跨媒介传输的文档水印溯源流程

Fig. 1　Tracing process of document watermarking in cross media transmission

2.1　数字传输信道下的文档水印方法

针对常见的数字传输信道下的失真（如截屏、放缩、高斯噪声和 JPEG 压缩），本文设计了一个基于可导噪声层的变形字体生成网络结构来自动生成字库，同时结合嵌入算法与提取算法，提出一种能抵抗数字传输信道的文档水印方案。本文方法的整体框架可以分为水印字库生成与解码器训练、水印嵌入和水印提取 3 个阶段。

2.1.1　水印字库生成与解码器训练

在水印字库生成与解码器训练阶段，通过一个编码—解码器网络自动生成含水印的字符图像并完成水印提取器的训练过程，这是本文水印方案的核心部分。

1）网络架构。本网络模型包括 4 个部分，分别是：编码器 E 、噪声层 N 、解码器 D 和对抗识别器 A ，网络架构如图 2 所示。该网络生成新字库的同时进行水印提取端的训练。字库生成过程为提取端提供训练样本，提取端用于指导字符图像的生成方向，两个模块相互协同，使网络生成视觉质量更高且鲁棒性更强的变形字体。编码器 E 接收水印信息 $\boldsymbol{W}$ 和载体字符图像 $\boldsymbol{I}_{\mathrm{C}}$ 后生成编码字符图像 $\boldsymbol{I}_{\mathrm{E}}$ ；噪声层 N 对编码字符图像 $\boldsymbol{I}_{\mathrm{E}}$ 进行加噪生成含噪字符图像 $\boldsymbol{I}_{\mathrm{N}}$ ；解码器 D 对加噪字符图像 $\boldsymbol{I}_{\mathrm{N}}$ 解码，提取水印标签 $\boldsymbol{W}'$ ；对抗识别器 A 检测当前图像 $\boldsymbol{I}' \in \{\boldsymbol{I}_{\mathrm{C}},$

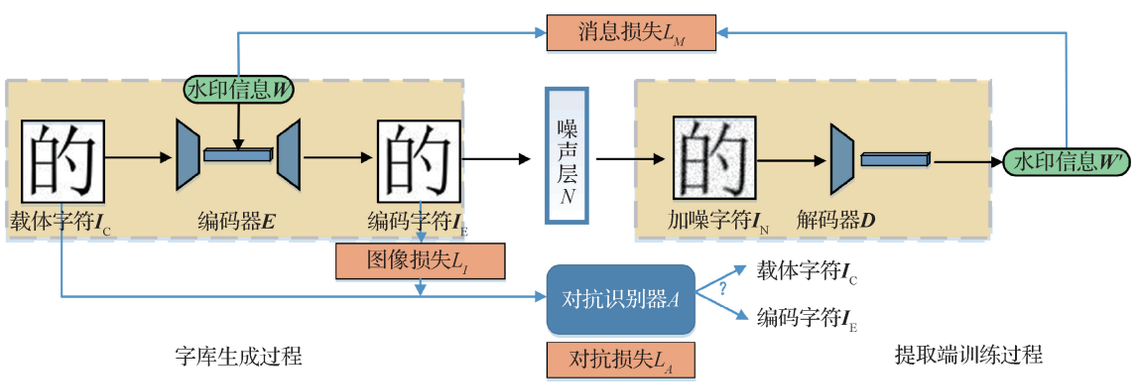


图 2　水印字库生成与解码器网络架构

Fig. 2　Network architecture of generation of watermarked fonts and decoder

$\boldsymbol{I}_{\mathrm{E}}\}$ 是载体字符图像还是编码字符图像。编码器和解码器通过联合训练,最小化载体字符图像 $\boldsymbol{I}_{\mathrm{C}}$ 和编码字符图像 $\boldsymbol{I}_{\mathrm{E}}$ 之间的图像损失 L_{I}、编码水印信息 $\boldsymbol{W}$ 和预测水印标签 $\boldsymbol{W}'$ 之间的消息损失 L_{M}、对抗识别器能否检测到编码字符图像 $\boldsymbol{I}_{\mathrm{E}}$ 的对抗损失 L_{A}。通过该网络架构,编码器 E 和解码器 D 的联合训练可用于生成含水印字符图片以自动构造字库,同时提取水印信息;对抗识别器 A 用于提升字库生成的视觉质量;噪声层 N 用于提升网络抗攻击鲁棒性。

2)编码器。编码器首先对输入的载体字符图像 $\boldsymbol{I}_{\mathrm{C}}$ 进行卷积,生成隐藏层的高维特征,然后将水印信息复制并拼接在卷积后的高维特征向量上,与此同时,将 $\boldsymbol{I}_{\mathrm{C}}$ 也拼接在高维向量中,之后通过反卷积操作生成编码后的字符图像 $\boldsymbol{I}_{\mathrm{E}}$。

编码器的网络结构如图 3 所示,对于给定尺寸为 $C \times H \times W$ 的载体字符图像 $\boldsymbol{I}_{\mathrm{C}}$,编码器对输入图像应用带有 64 个输出滤波器的 6 个残差块,生成 $64 \times H \times W$ 的图像中间形式。然后在空间上复制水印信息 $\boldsymbol{W}$ 以形成一个 $L \times H \times W$ 消息量,L 是水印长度。把得到的消息量连同原始图像添加到中间形式,连接成一个 $(64+L+C) \times H \times W$ 的特征图。然后编码器应用带有 64 个输出滤波器的 6 个残差块,输出形状为 $64 \times H \times W$ 的特征图。最后经过一个卷积核为 1×1、步长为 1 的卷积层,生成形状为 $C \times H \times W$ 的编码字符图像 $\boldsymbol{I}_{\mathrm{E}}$。

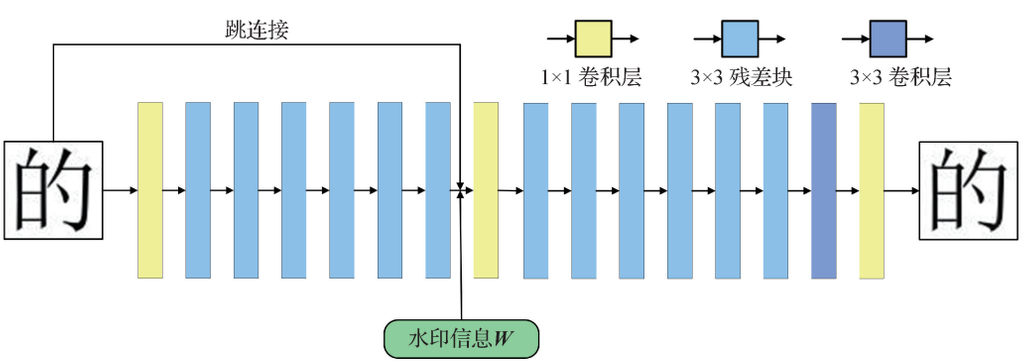


图 3　编码器网络结构

Fig. 3　Network architecture of encoder

为了达到不可见水印的目的,编码字符图像 $\boldsymbol{I}_{\mathrm{E}}$ 在视觉上应该和载体字符图像 $\boldsymbol{I}_{\mathrm{C}}$ 相似。本文方法除了使用像素级别的损失函数衡量 $\boldsymbol{I}_{\mathrm{E}}$ 和 $\boldsymbol{I}_{\mathrm{C}}$ 之间的相似性外,仍需保证图像的修改区域集中在字符笔划的连接处,以保证修改更符合笔顺的连续性。为了实现这样的功能,本文方法在网络中加入掩码指导的失真函数。由此本方法将水印定义为文本边缘区域,通过形态学图像处理中的图像膨胀得到字符图像的修改区域,进而得到掩码图像 $\boldsymbol{I}_{\mathrm{M}}$,其中可修改部分置为 1,不可修改部分置为 100。膨胀操作 $\oplus$ 使用指定的结构元素对 $\boldsymbol{I}_{\mathrm{C}}$ 进行膨胀,该结构元素决定像素邻域的形状,膨胀操作的定义为

$$\boldsymbol{I}_{\mathrm{C}} \oplus \boldsymbol{S} = \{\boldsymbol{x} \mid (\boldsymbol{S}^{V})_{x} \cap \boldsymbol{I}_{\mathrm{C}} \neq \boldsymbol{\Phi}\} \tag{1}$$

式中,$\boldsymbol{x}$ 表示集合平移的位移量,$\boldsymbol{S}^{V}$ 表示对结构元素 $\boldsymbol{S}$ 做关于其原点的反射得到的集合。$(\boldsymbol{S}^{V})_{x}$ 表示在目标图像 $\boldsymbol{I}_{\mathrm{C}}$ 上将 $\boldsymbol{S}^{V}$ 平移 $\boldsymbol{x}$,$(\boldsymbol{S}^{V})_{x}$ 与目标图像 $\boldsymbol{I}_{\mathrm{C}}$ 至少有一个非零公共元素相交时,对应的原点位置所组成的集合即为用 $\boldsymbol{S}$ 膨胀 $\boldsymbol{I}_{\mathrm{C}}$ 的结果。在实验中,本文首先对字符图像 $\boldsymbol{I}_{\mathrm{C}}$ 的黑白像素进行翻转,采用 3×3 的卷积核作为结构元素进行膨胀。本文将在 3.2 节通过实验说明加入掩码对字库生成质量的影响。

3)噪声层。为了使本文的模型能够保持对各种图像畸变的鲁棒性,本文考虑了常见的图像失真,并设计了相应的可导噪声层来模拟失真过程,设计的噪声层包括:

(1) 恒等映射层,保持编码字符图像 $\boldsymbol{I}_{\mathrm{E}}$ 不变;

(2) 缩放层,随机缩放编码字符图像 $\boldsymbol{I}_{\mathrm{E}}$;

(3) 高斯加噪层,对编码字符图像 $\boldsymbol{I}_{\mathrm{E}}$ 进行高斯加噪;

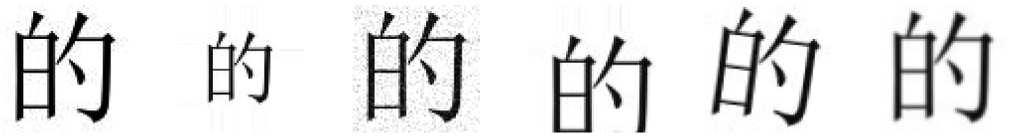
(4) 平移层,随机平移编码字符图像 $\boldsymbol{I}_{\mathrm{E}}$;

(5) 旋转层,随机旋转编码字符图像 $\boldsymbol{I}_{\mathrm{E}}$;

(6) 高斯模糊层,对编码字符图像 $\boldsymbol{I}_{\mathrm{E}}$ 进行高斯模糊。

各种噪声层对应的加噪字符图像如图 4 所示。

本文随机选择以上 6 种噪声层中的一种或几种为一次迭代的总噪声层，噪声层接收编码后的字符图像 $\boldsymbol{I}_{\mathrm{E}}$ 作为输入并输出失真后的字符图像 $\boldsymbol{I}_{\mathrm{N}}$ 。本文将通过实验证明使用噪声层模拟噪声失真，对抗缩放、高斯噪声和 JPEG 压缩等常见数字失真的有效性。



(a) 原始字符　(b) 缩放　(c) 高斯加噪　(d) 平移　(e) 旋转　(f) 高斯模糊

图 4　不同噪声层生成的加噪字符图

Fig. 4　Noisy character images generated by different noise layers ((a) original character; (b) scaled character; (c) Gaussian noisy character; (d) translated character; (e) rotated character; (f) Gaussian blurred character)

4) 解码器。解码器相当于一个字符识别的网络，对编码后加噪的字符图像进行解码，通过网络训练恢复噪声字符图像 $\boldsymbol{I}_{\mathrm{N}}$ 中的水印信息。解码器网络结构如图 5 所示，该解码器输入含噪字符图像，通过残差块和降采样操作生成具有 L 维特征通道的中间表示形式，然后使用一个平均池化层和一个全连接层，生成预测的字符标签，用于表达提取的水印消息 $\boldsymbol{W}'$ 。

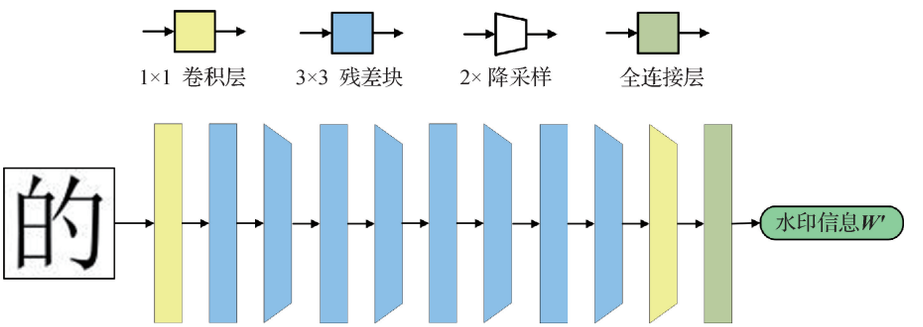


图 5　解码器网络结构

Fig. 5　Network architecture of decoder

5) 对抗识别器。对抗识别器用于判断输入图像是编码字符图像 $\boldsymbol{I}_{\mathrm{E}}$ 还是载体字符图像 $\boldsymbol{I}_{\mathrm{C}}$ 。编码—解码器和对抗识别器两个模块作为生成模型和判别模型互相进行对抗训练，使得编码器生成的含水印字符更加真实。本文的对抗识别器使用 PatchGAN(Isola 等，2017) 的结构。

6) 损失函数。本文在编码—解码器联合训练网络中定义了 6 个损失函数。

(1) 在编码器 E 中，最小化载体字符图像 $\boldsymbol{I}_{\mathrm{C}}$ 和编码字符图片 $\boldsymbol{I}_{\mathrm{E}}$ 之间的图像损失 L_{I} ，即

$$L_{\mathrm{I}}(\boldsymbol{I}_{\mathrm{C}},\boldsymbol{I}_{\mathrm{E}}) = \|(\boldsymbol{I}_{\mathrm{C}} - \boldsymbol{I}_{\mathrm{E}}) \times \boldsymbol{I}_{\mathrm{M}}\|_2^2 / (C \times H \times W) \tag{2}$$

式中，$\boldsymbol{I}_{\mathrm{M}}$ 表示掩码图像。

(2) 在解码器 D 中，最小化编码水印信息 $\boldsymbol{W}$ 和字符解码标签代表的水印信息 $\boldsymbol{W}'$ 之间的消息损失 L_{M} 为

$$L_{\mathrm{M}}(\boldsymbol{W},\boldsymbol{W}') = \|\boldsymbol{W} - \boldsymbol{W}'\|_2^2 / L \tag{3}$$

(3) 在对抗网络中，编码器会试图欺骗对抗识别器 A ，使得对抗识别器无法区分出 $\boldsymbol{I}_{\mathrm{C}}$ 和 $\boldsymbol{I}_{\mathrm{E}}$ ，最终利用生成损失 L_{G} 提升生成的编码字符图像 $\boldsymbol{I}_{\mathrm{E}}$ 的质量，即

$$L_{\mathrm{G}}(\boldsymbol{I}_{\mathrm{E}}) = \log(1 - A(\boldsymbol{I}_{\mathrm{E}})) \tag{4}$$

(4) 相反，有了 $\boldsymbol{I}_{\mathrm{C}}$ 和 $\boldsymbol{I}_{\mathrm{E}}$ ，对抗识别器 A 会努力对其进行二分类，通过最小化载体字符图像 $\boldsymbol{I}_{\mathrm{C}}$ 和编码字符图像 $\boldsymbol{I}_{\mathrm{E}}$ 的分类损失 L_{C} 得以实现，即

$$L_{\mathrm{C}} = \log(1 - A(\boldsymbol{I}_{\mathrm{C}})) + \log(A(\boldsymbol{I}_{\mathrm{E}})) \tag{5}$$

(5) 编码—解码器模块总的损失定义为

$$L_2 = \lambda_1 L_{\mathrm{I}}(\boldsymbol{I}_{\mathrm{C}},\boldsymbol{I}_{\mathrm{E}}) + \lambda_2 L_{\mathrm{G}}(\boldsymbol{I}_{\mathrm{E}}) + \lambda_3 L_{\mathrm{M}}(\boldsymbol{W},\boldsymbol{W}') \tag{6}$$

式中，λ_1 、λ_2 和 λ_3 表示控制损失的相对权重，同时对抗识别器也参与其中。

在网络训练阶段，同时训练编码—解码器模块

和对抗识别器模块。

2.1.2　水印嵌入

在生成水印字库后,为常用字体构造一个含水印信息的字符码本。该字体中的每个字符对应几种不同的轻微扰动,用这些字的编号代表信息。若生成2个变形字则每个字代表1 bit,若生成4个则每个字可代表2 bit,而原始字符不代表信息。本文在实验中设置生成2种形变字符,构成如图6所示的字符水印码表,字符标签代表水印信息。对于字符形变,人眼几乎不可识别,但是解码器仍然可以识别。所有形变字符根据相应的水印信息生成完整的字符水印码表。在水印嵌入时,根据当前的水印信息,在输入文档中选择水印码表中对应的字符对当前字符进行替换。这样即可把水印信息编码到整个原始文档当中生成含水印文档。

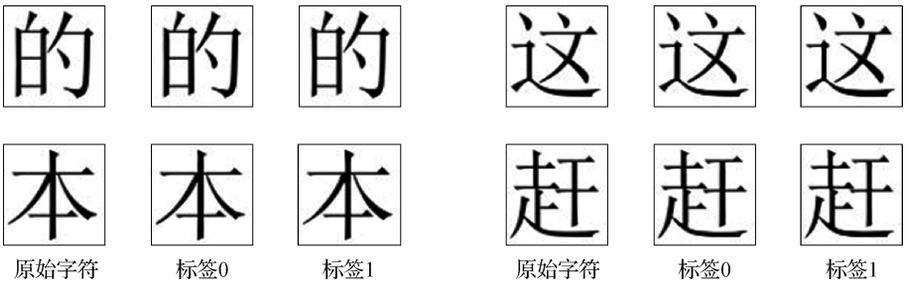


图6　字符水印码表

Fig. 6　Codebook of watermarked characters

2.1.3　水印提取

在水印提取阶段,需要从当前文档每个字符中恢复字符编码标签以代表水印信息,水印提取过程如图7所示。在收到数字信道传输下失真的文档水印后,首先通过字符分割将整个文档分为单字符图像。而后将当前字符送入水印字库生成阶段训练好的网络中,也就是编码—解码器结构的解码器中。通过解码器识别即可提取出当前字符嵌入的水印消息。

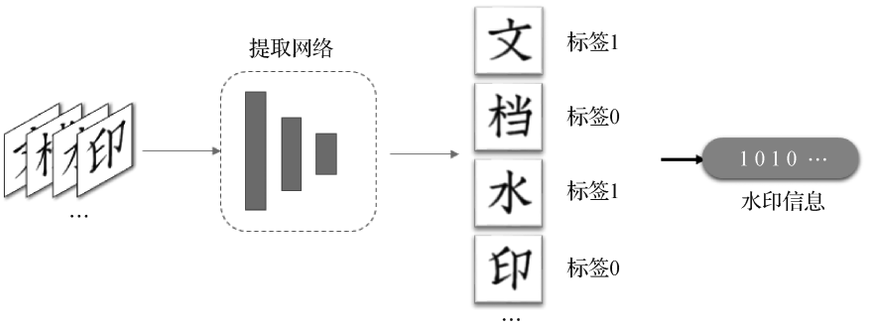


图7　水印提取过程

Fig. 7　Watermark extraction process

2.2　打印扫描场景下的文档水印方法

对文档进行打印扫描时,文档中的字符经历了数字信号到模拟信号再到数字信号的转化失真。在这个过程中含有大量不均匀的采样和量化操作,会造成原始文件的像素失真和几何失真,导致图像质量大幅下降。而且不同型号的打印扫描设备以及设备的老化程度,都会给文档带来不同程度的失真,导致无法精确地进行失真过程的模拟,这也使得2.1节提到的抵抗数字信道传输的文档水印方案的适应性较差。为此本文提出一种提取端微调方案,在端到端网络的基础上,针对打印扫描场景,仅对字库生成网络进行微调,其他部分均保持不变,使得水印提取的准确率大幅度上升。

打印扫描场景下的文档水印字库生成模型框架如图8所示,本模型将上节中的字库生成模型作为预训练的网络,然后对网络中所提取的新的特征进行训练。这个过程相当于一种迁移学习,也就是微调网络。在微调的过程中,首先通过上述文档水印

模型自动生成编码字符图像集，然后将这些字符图片合成完整文档后进行打印扫描。将打印扫描文档进行预处理和字符切割后作为数据集传入解码器，单独调整解码器。本文通过实验证明了使用较小的数据集对网络进行微调即可大幅度提高网络在打印扫描场景下文档水印的提取率。

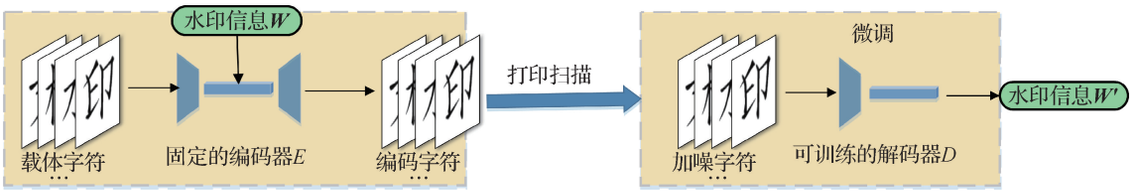


图 8　打印扫描场景下的文档水印网络架构

Fig. 8　Document watermarking network architecture in print-scan scene

3　实验结果

3.1　实验设置

针对数字信道传输和打印扫描这两种应用场景，实验对本文中文文档水印模型进行了性能评估。实验过程中采用随机生成的二进制序列作为水印信息，打印扫描设备是 HP OfficeJet Pro 8720，分辨率均设置为 600 dpi。实验中采用的数据集是中文宋体字符图像，包含常用的 3 000 个汉字，单幅字符图像大小为 64 × 64 像素。测试集是具有 252 个中文字符的真实文档，在本实验中，每个字符嵌入 1 bit 水印信息，因此整个文档共嵌入 252 bit 信息。

本文从以下 3 个指标来衡量文档水印模型：1）视觉质量，即生成的水印图像和原始图像的相似度；2）鲁棒性，即图像遭受截屏失真和打印扫描失真后，水印提取的成功率；3）容量，即整个文档中嵌入的水印信息比特。

3.2　掩码失真的重要性

本文方法在编码器中分别使用掩码和不使用掩码指导的损失函数生成含水印的编码字符，其他结构均保持不变，效果对比如图 9 所示，图 9（a）表示输入的载体字符图像 $\boldsymbol{I}_{\mathrm{C}}$，图 9（b）表示使用的掩码图像 $\boldsymbol{I}_{\mathrm{M}}$，图 9（c）表示使用掩码图像 $\boldsymbol{I}_{\mathrm{M}}$ 生成的编码字符图像 $\boldsymbol{I}_{\mathrm{C}}$，图 9（d）表示未使用掩码图像生成的编码字符图像 $\boldsymbol{I}_{\mathrm{E}}'$。从图中可以看出，未使用掩码图像的编码器生成的字符图像的文字部分出现断连和伪影，这些改变对于人眼来说是非常明显的。而使用掩码图像的生成字符很好地保证了笔划的连续性，表现为仅改变了某些部分的粗细，生成结果和载体字符图像非常相似，有很好的视觉效果，应用到文档中人眼几乎不可区分。所以在接下来的实验中均使用掩码指导的损失函数来生成字库。

（a）原始字符　（b）掩码图像　（c）使用掩码生成字符　（d）未使用掩码生成字符

图 9　使用掩码和未使用掩码生成效果对比

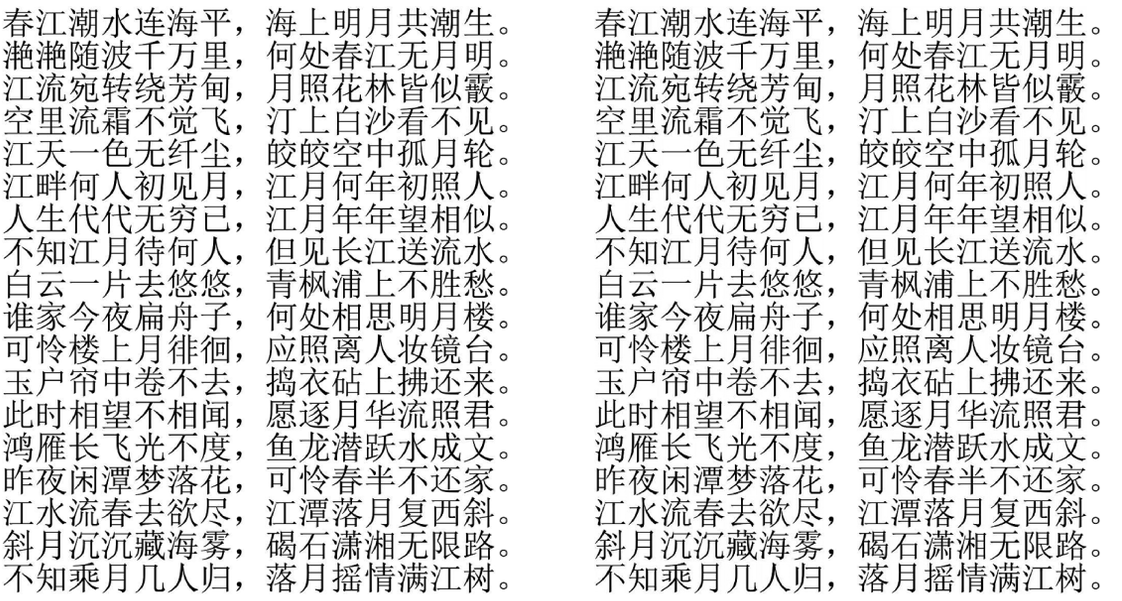
Fig. 9　Comparison of the effect of using mask and not using mask

（（a）original character；（b）mask image；（c）character using mask；（d）character not using mask）

3.3　视觉质量评估

本文对真实的中文文本文档嵌入水印，用以测试本文方法的效果。图 10（a）是没有进行水印嵌入的原始图像，共有 252 个中文字符。在本文方法的字体生成过程中，每个字符嵌入 1 bit 信息，共嵌入 252 bit 水印信息。图 10（b）显示了使用本文方法嵌入水印后的文档效果。可以看出，文档嵌入水印后的视觉质量非常好，人眼几乎不能分辨出与原始文本图像的差别。

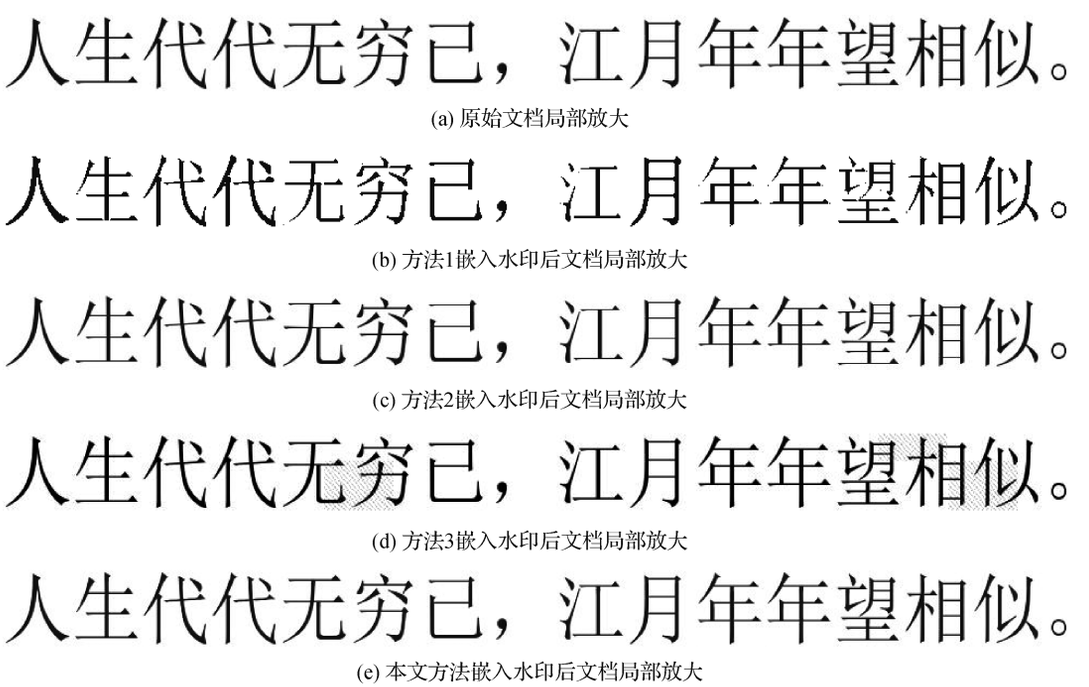
图 11 展示了本文方法和方法 1（亓文法 等，2008）、方法 2（Qi 等，2019）以及方法 3（Fang 等，2019）嵌入水印后的局部文档效果比较。方法 1 中的量化步



(a) 原始文档　　(b) 含水印文档

图 10　嵌入水印前后的文档视觉效果对比

Fig. 10　Comparison of document visual quality before and after embedding watermark ((a) before; (b) after)



(a) 原始文档局部放大

(b) 方法1嵌入水印后文档局部放大

(c) 方法2嵌入水印后文档局部放大

(d) 方法3嵌入水印后文档局部放大

(e) 本文方法嵌入水印后文档局部放大

图 11　不同方法嵌入水印前后文档局部放大对比

Fig. 11　Comparison of local enlarged drawings of documents embedded watermark by different methods

((a) original; (b) method #1; (c) method #2; (d) method #3; (e) proposed scheme)

长 K 设置为 0.15。从图中可以看出,方法 1 嵌入水印后的字符出现了笔画的残缺、断连和毛刺,这是由于此方法基于文档二值图像,像素点只有翻转和不翻转这两种状态,修改任意像素点对视觉的影响都是巨大的。

方法 2 的生成图像的视觉质量相比之下要好很多,这是由于利用了字体设计的专业知识改变字符的拓扑结构,生成人眼不易觉察差别的高质量字库。

但是数量庞大的汉字给中文字库的制作带来了巨大挑战。同时在设计字库时,需要考虑怎样修改字符使得文本在失真后仍能识别出这些变化模式,这个设计过程具有较强的主观性和不确定性。

方法 3 是目前效果较好的图像水印方法,但用于文档水印时生成的字符图像背景出现了大面积的块状浅灰色噪点,这些噪点在文档打印出来后更为明显。这说明了图像水印方法难以直接应用到文档图像,还需要针对文档的特征设计特殊的文档水印方案。此外,这类基于图像的方法需要首先将文档转化为图像,不能进行实时嵌入,而基于字库生成的文档水印方法克服了这一问题。在图 11 中,本文方法嵌入水印后文档局部放大图和原始文档局部放大图几乎没有差异,在编码—解码器联合训练阶段,字符解码网络向字符生成网络提供了字符生成的修改方向,自动设计出机器更容易识别的水印模式,采用的基于掩码的失真设置也限制了对字符图像的修改均分布在字符轮廓上以生成更高质量的字库。

本文用每个字符承载的平均比特数(bits per character,BPC)来衡量水印嵌入方法的容量,用峰值信噪比(peak signal to noise ratio,PSNR)和结构相似性(structural similarity,SSIM)来衡量水印嵌入后文档图片的视觉质量。本文方法与方法 1、方法 2 以及方法 3 的容量和视觉质量的定量比较如表 2 所示。方法 1 需要把某些字符单独分组出来作为调整字符,导致这些字符无法嵌入水印,所以水印容量比其他方法稍低。方法 2 对某些结构简单的字符,比如“一”,无法修改其拓扑结构也就无法嵌入水印,这些少量的字符在嵌入水印时称为无效字符。方法 3 中每个 64×64 像素块可嵌入 1 bit 的水印,相当于每字符嵌入 1 bit 消息。本文的方法对任何结构的字符均可生成水印字符,故每字符均也可嵌入 1 bit 水印。从表 2 中可以看出,本文方法的 PSNR 和 SSIM 值显著高于其他两种文档水印方法,略高于图像水印方法。虽然由于文档图像的特殊性,PSNR 和 SSIM 不足以完全衡量文档的视觉效果,但也一定程度上说明了本文提出方法的生成水印文档图片与原文档最为相似。

此外,本文还采用了主观质量评价,即问卷调查的方式比较这 4 种方法的视觉质量。问卷随机选取了 20 个中文汉字,包含具有复杂字形和简单字形的字符图片,将原始宋体字符图像和文档水印方法生成的含水印字符图像进行对比。对两者的相似性程度进行打分,5 为最相似,1 为最不相似。本文的调查收回了 30 份有效问卷,问卷的结果如表 3 所示。从问卷的结果来看,方法 3 的视觉质量最差,再次说明了图像水印方法不适合直接用于文档图像。方法 1 的视觉效果也比较差,本文方法的得分比方法 2 稍高,具有最好的视觉质量。

表 2　水印容量和视觉质量比较

Table 2　Comparisons of embedding capacity and visual quality

方法	BPC	PSNR/dB	SSIM
方法 1	0.94	17.30	0.88
方法 2	0.99	16.94	0.91
方法 3	**1**	28.57	0.96
本文	**1**	**28.62**	**0.99**

注:加粗字体为每列最优值。

表 3　文档水印视觉质量问卷结果

Table 3　Questionnaire results of visual quality of document watermarking

	方法 1	方法 2	方法 3	本文方法
平均相似性分数	2.48	4.16	2.12	**4.45**

注:加粗字体为最优值。

3.4　针对数字噪声的鲁棒性评估

针对数字噪声的应用场景,这里对本文方法和方法 1、方法 2、方法 3 进行了鲁棒性评估比较,水印的提取效果如表 4 所示。这里的数字噪声包括最常见的截屏、高斯加噪、JPEG 压缩和缩放。可以看出,方法 3 在各种数字噪声中水印提取率均为 100%,说明这类图像水印方法对于各类图像处理均有很强的鲁棒性。本文方法在最典型的数字噪声截屏场景下提取率也达到了 100%;在高斯噪声和 JPEG 压缩的失真场景下与其他 3 种方法取得了几乎相当的效果。本文方法在不同缩放大小的场景下均可正确提取水印。方法 1 在缩放比例小于 1 的失真场景下水印的提取效果较差。该实验验证了本方法在各种数字噪声的场景下水印的提取率都比较稳定。

表 4　数字噪声下水印提取效果比较

Table 4　Comparison of watermark extraction rate under digital distortion

/%

	方法 1	方法 2	方法 3	本文
截屏	89.83	98.94	**100**	**100**
高斯噪声 $\sigma=5$	**100**	**100**	**100**	99.80
高斯噪声 $\sigma=10$	**100**	**100**	**100**	99.80
JPEG 压缩 $Q=50$	**100**	**100**	**100**	99.21
JPEG 压缩 $Q=80$	**100**	**100**	**100**	99.80
缩放 50	67.37	**100**	**100**	**100**
缩放 80	81.78	98.40	**100**	**100**
缩放 200	**100**	**100**	**100**	**100**

注：加粗字体为每行最优值。

3.5　针对打印扫描场景的鲁棒性评估

针对打印扫描场景对不同字号的文档的水印提取效果进行比较，结果如表 5 所示。本文方法微调网络的数据集只有 500 幅字符图像，共训练 5 轮。为了保证方法 1 的视觉质量，本文选用的量化步长为 0.15，这导致了水印的提取率显著低于其他方法。加大量化步长可以提高抗打印扫描水印提取率，但是会导致文档视觉质量进一步大幅度降低。为了保证鲁棒性和视觉质量的折中，需要谨慎选择量化步长。方法 2 的水印提取效果在不同字号下均较好。相比于方法 2，本文方法在字号增大时，打印扫描下的水印提取率会略微上升。方法 3 在大字号的打印扫描场景下鲁棒性很高，而在小四号、五号这类小字号场景下无法提取出水印信息。通过比较实验得知，本文的文档水印方法使用几种不同的常用字号在打印扫描场景下的效果均最好。

表 5　打印扫描场景下不同字号的文档水印提取效果比较

Table 5　Comparison of watermark extraction rate of different font sizes in print-scan scene

/%

方法	字号大小			
	三号	四号	小四号	五号
方法 1	89.41	95.76	82.20	88.98
方法 2	96.81	95.74	96.28	96.81
方法 3	96.35	98.44	59.36	57.81
本文	**99.21**	**98.81**	**97.62**	**96.83**

注：加粗字体为每列最优值。

此外，本文针对不同的打印质量和扫描质量进行了交叉提取效果对比。实验选取了最常用的打印分辨率：600 dpi、1 200 dpi 和扫描分辨率：400 dpi、600 dpi，文档字号为四号，水印提取效果如表 6 所示。表中第 1 行表示打印质量/扫描质量，单位为 dpi。从表中可以看出，相比于其他方法，本文方法提取率均为最佳或与最佳结果接近。在打印分辨率为 1 200 dpi、扫描分辨率为 400 dpi 的场景下，由于分辨率相差较大，各方法的水印提取效果均降低。而在其他交叉场景下，本文方法水印提取率均在 98% 以上。因此，通过比较实验说明，本方法在打印和扫描分辨率不匹配的情况下，也优于其他 3 种方法。

表 6　不同打印质量和扫描质量场景下的文档水印提取效果比较

Table 6　Comparison of watermark extraction rate of different printing and scanning dpi

/%

方法	打印质量/扫描质量(dpi)			
	600/400	600/600	1 200/400	1 200/600
方法 1	94.68	95.76	**92.02**	97.34
方法 2	89.41	95.74	84.32	91.10
方法 3	92.19	98.44	53.75	90.63
本文	**98.02**	**98.81**	**91.67**	**98.41**

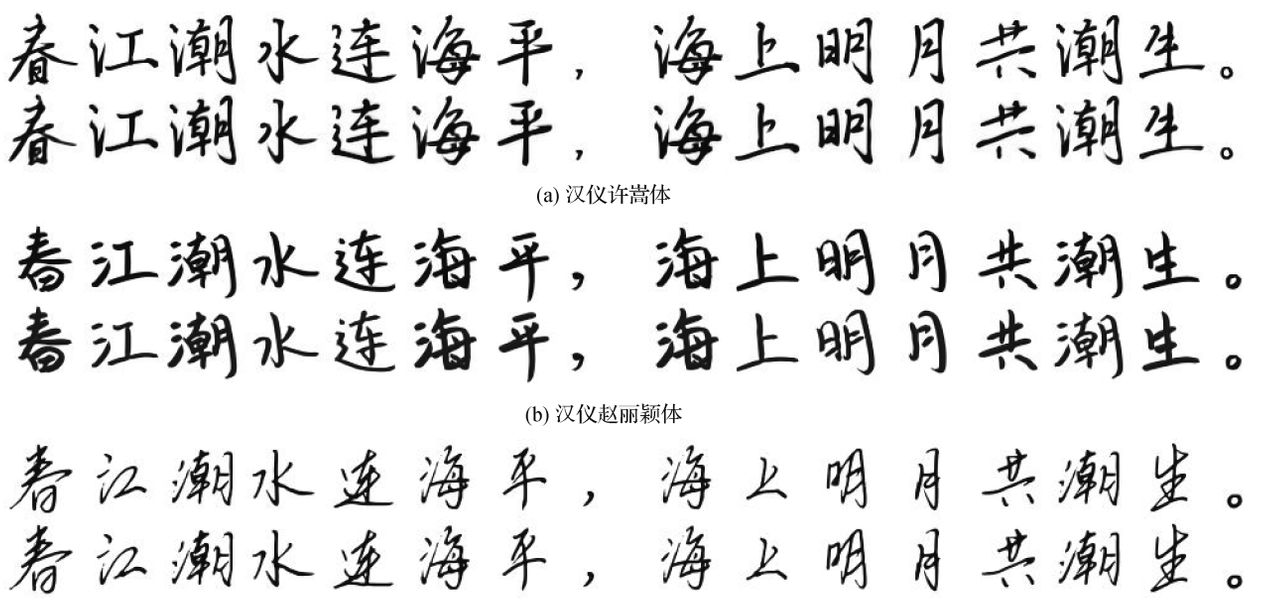
注：加粗字体为每列最优值。

3.6　针对手写风格字体库的文档水印

为了验证本文文档水印方法的通用性，将本文方法应用于风格化手写字体上。这里使用了 3 种手写字库，分别是汉仪许嵩体、汉仪赵丽颖体和钟齐山文丰手写体，嵌入水印后的效果如图 12 所示。图中第 1 行表示原始手写体字符，第 2 行表示嵌入水印后的字符。从图中可以看出，尽管手写字体比印刷字体的笔迹更复杂，但本文方法在生成手写字体文档水印方面仍取得了良好的视觉质量。我们也通过水印提取实验验证了该方法同样具有良好的鲁棒性。

4　结　论

利用文档水印技术，可以对文档的泄密源头进行有效追踪，但目前的中文文档水印方法存在复杂度较高、不能自动生成字库的问题。因此，本文提出了一种基于编码—解码器网络的自动生成中文字库的文档水印算法。其中，字库生成网络只需输入目



(a) 汉仪许嵩体

(b) 汉仪赵丽颖体

(c) 钟齐山文丰手写体

图 12　手写字体原始文档和含水印文档对比

Fig. 12　Comparison of document visual quality of different handwritten fonts before and after embedding watermark ((a) font HYXuSongJ; (b) font HYZhaoLiYingJ; (c) font HYZhongQiShanJ)

标字体字符图像作为训练样本,即可生成水印编码的字符图像进而生成中文字库,并同步完成水印提取器的训练。而后根据当前水印信息和字库进行字符替换完成水印嵌入,将含噪水印字符送入训练好的提取器即可实现水印的提取。

实验表明了本文方法与其他方法相比,无论是主观评价还是定量分析,均具有更好的视觉质量。对于截屏等常见的数字失真场景,本文方法均能保持 99% 以上的水印提取率。对于打印扫描场景,本文方法在不同字号和多种打印扫描分辨率下,相比于其他方法均具有更好的提取效果。此外,本文还探索了针对手写体风格字库的文档水印效果,表明了本文方法的通用性。

尽管本文方法在数字传输和打印扫描场景下都具有较好的鲁棒性,但目前还不适用于打印拍照和屏幕拍照的场景。这是由于拍照过程对文档的影响更剧烈,难以通过加噪网络对字符失真进行准确模拟。随着智能手机的普及,对文档进行打印拍照和屏幕拍照将愈加普遍,因此针对这两个场景设计鲁棒的文档水印算法是未来可以研究的方向。

参考文献(References)

Ai P. 2014. Research of Text Watermarking Algorithm to Resist Printing and Scanning. Xi'an: Xidian University (艾平. 2014. 抗打印扫描文本水印算法研究. 西安: 西安电子科技大学)

Atallah M J, Raskin V, Crogan M, Hempelmann C, Kerschbaum F, Mohamed D and Naik S. 2001. Natural language watermarking: design, analysis, and a proof-of-concept implementation//Proceedings of the 4th International Workshop on Information Hiding. Pittsburgh, USA: Springer-Verlag: 185-199 [DOI: 10.1007/3-540-45496-9_14]

Brassil J, Low S, Maxemchuk N and O'Gorman L. 1995. Hiding information in document images//Proceedings of the 29th Annual Conference on Information Sciences and Systems. Baltimore, USA: Johns Hopkins University: 482-489

Brassil J T, Low S and Maxemchuk N F. 1999. Copyright protection for the electronic distribution of text documents. Proceedings of the IEEE, 87(7): 1181-1196 [DOI: 10.1109/5.771071]

Briffa J A, Culnane C and Treharne H. 2010. Imperceptible printer dot watermarking for binary documents//Proceedings Volume 7723, Optics, Photonics, and Digital Technologies for Multimedia Applications. Brussels, Belgium: SPIE: 77230M [DOI: 10.1117/12.854708]

Campbell N D F and Kautz J. 2014. Learning a manifold of fonts. ACM Transactions on Graphics, 33(4):#91 [DOI: 10.1145/2601097.2601212]

Chen Y, Li Z, Zhang J and Wang G M. 2019. Robust watermarking algorithm for diffusion weighted images. Journal of Image and Graphics, 24(9): 1434-1449 (陈怡, 李智, 张健, 王国美. 2019. 弥散加权图像的鲁棒水印算法研究. 中国图象图形学报, 24(9): 1434-1449) [DOI: 10.11834/jig.180672]

Deng X H, Chen Z G and Mao Y M. 2014. Lossless watermarking algorithm for medical image's tamper detection and recovery with high quality. Journal of Image and Graphics, 19(4): 583-591 (邓小鸿, 陈志刚, 毛伊敏. 2014. 基于无损水印的医学图像篡改检测和高质量恢复. 中国图象图形学报, 19(4): 583-591) [DOI: 10.11834/jig.20140413]

Fang H, Zhang W M, Zhou H, Cui H and Yu N H. 2019. Screen-shooting resilient watermarking. IEEE Transactions on Information Forensics and Security, 14(6): 1403-1418 [DOI: 10.1109/TIFS.2018.2878541]

Hii A J H, Hann C E, Chase J G and Van Houten E E W. 2006. Fast normalized cross correlation for motion tracking using basis functions. Computer Methods and Programs in Biomedicine, 82(2): 144-156 [DOI: 10.1016/j.cmpb.2006.02.007]

Huang D and Yan H. 2001. Interword distance changes represented by sine waves for watermarking text images. IEEE Transactions on Circuits and Systems for Video Technology, 11(12): 1237-1245 [DOI: 10.1109/76.974678]

Isola P, Zhu J Y, Zhou T H and Efros A A. 2017. Image-to-image translation with conditional adversarial networks//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE: 5967-5976 [DOI: 10.1109/CVPR.2017.632]

Kelkoul H, Zaz Y, Tribak H and Schaefer G. 2018. A robust combined audio and video watermark algorithm against cinema piracy//Proceedings of the 6th International Conference on Multimedia Computing and Systems (ICMCS). Rabat, Morocco: IEEE: 1-4 [DOI: 10.1109/ICMCS.2018.8525979]

Li S Z, Zhang X, Deng X H and Wu X Y. 2015. Reversible video watermarking algorithm for H.264/AVC based on mode feature. Journal of Image and Graphics, 20(10): 1285-1296 (李淑芝, 张翔, 邓小鸿, 吴晓燕. 2015. 基于模式特征的 H.264/AVC 可逆视频水印. 中国图象图形学报, 20(10): 1285-1296) [DOI: 10.11834/jig.20151001]

Liu D, Sun M and Zhou M T. 2007. A text digital watermarking technology based on graph theory. Journal of Computer Research and Development, 44(10): 1757-1764 (刘东, 孙明, 周明天. 2007. 基于图论的文本数字水印技术. 计算机研究与发展, 44(10): 1757-1764)

Qi W F, Guo W, Zhang T, Liu Y X, Guo Z M and Fang X F. 2019. Robust authentication for paper-based text documents based on text watermarking technology. Mathematical Biosciences and Engineering, 16(4): 2233-2249 [DOI: 10.3934/mbe.2019110]

Qi W F, Li X L, Yang B and Cheng D F. 2008. Document watermarking scheme for information tracking. Journal on Communications, 29(10): 183-190 (亓文法, 李晓龙, 杨斌, 程道放. 2008. 用于信息追踪的文本水印算法. 通信学报, 29(10): 183-190) [DOI: 10.3321/j.issn:1000-436X.2008.10.026]

Su Q T and Chen B J. 2018. Robust color image watermarking technique in the spatial domain. Soft Computing, 22(1): 91-106 [DOI: 10.1007/s00500-017-2489-7]

Suzaki M and Suto M. 2005. A watermark embedding and extracting method for printed documents. Electronics and Communications in Japan (Part III: Fundamental Electronic Science), 88(7): 43-51 [DOI: 10.1002/ecjc.20142]

Takahashi Y, Yamada T, Ebisawa R, Fujii Y and Tezuka S. 2008. Information embedding method for home printing of certifications//Proceedings of the 10th International Conference on Advanced Communication Technology. Gangwon, Korea (South): IEEE: 2116-2120 [DOI: 10.1109/ICACT.2008.4494206]

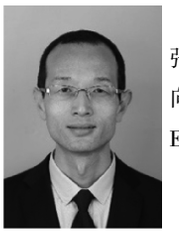
Wu M and Liu B D. 2004. Data hiding in binary image for authentication and annotation. IEEE Transactions on Multimedia, 6(4): 528-538 [DOI: 10.1109/TMM.2004.830814]

Xiao C, Zhang C and Zheng C X. 2018. FontCode: embedding information in text documents using glyph perturbation. ACM Transactions on Graphics, 37(2):#15 [DOI: 10.1145/3152823]

Zhang Y, Liu T, Chen Y H, Zhao S Q and Li S. 2005. Natural Language Watermarking. Journal of Chinese Information Processing, 19(1): 56-62, 70 (张宇, 刘挺, 陈毅恒, 赵世奇, 李生. 2005. 自然语言文本水印. 中文信息学报, 19(1): 56-62, 70) [DOI: 10.3969/j.issn.1003-0077.2005.01.009]

作者简介

孙杉, 1996 年生, 女, 硕士研究生, 主要研究方向为数字水印。
E-mail: shansun@mail.ustc.edu.cn



张卫明, 通信作者, 男, 教授, 主要研究方向为多媒体内容安全与人工智能安全。
E-mail: zhangwm@ustc.edu.cn

方涵, 男, 博士研究生, 主要研究方向为数字水印与图像处理。E-mail: fanghan@mail.ustc.edu.cn
俞能海, 男, 教授, 主要研究方向为多媒体内容安全、多媒体内容检索与视频处理。E-mail: ynh@ustc.edu.cn