

载体独立的抗屏摄信息膜叠加水印算法

李晓萌¹, 郭玳豆¹, 卓训方², 姚恒¹, 秦川¹

(1. 上海理工大学光电信息与计算机工程学院, 上海 200093; 2. 上海数据交易所有限公司, 上海 201203)

摘要: 金融安全作为国家安全的重要组成部分, 是经济平稳健康发展的重要基础。数字图像水印技术在金融信息安全方面发挥着巨大作用, 其中, 抗屏幕拍摄水印算法成为数字图像水印技术新的研究重点。如何兼顾水印图像的视觉质量和鲁棒性是抗屏幕拍摄鲁棒水印算法需要解决的重要问题。现有的水印方案一般通过对载体图像进行修改来实现水印不可见的目的, 这种水印嵌入方式不具有普适性。为此, 提出了一种新的基于深度学习的端到端抗屏幕拍摄鲁棒水印算法。作为端到端网络架构的一部分, 编码器网络的输入是随机二进制字符串, 经过网络训练后根据输入的水印信息生成相应的水印信息膜, 可以附加在任意的载体图像上。使用数学方法模拟了屏幕拍摄过程中可能产生的失真, 使得经过网络学习后的模型具有抵抗屏幕拍摄噪声的能力。增加了基于图像的恰可察觉差损失来进一步提升水印图像的视觉质量。此外, 为了更灵活地平衡水印图像的视觉质量和鲁棒性, 在训练阶段设计了一个嵌入超参数, 通过改变嵌入超参数的大小, 就可以得到适合不同场景的模型。为了验证所提算法的有效性, 针对视觉质量和鲁棒性做了多种实验。实验结果表明, 与目前的主流算法相比, 使用所提算法生成的水印图像具有更好的视觉质量, 并且在不同距离、不同角度、不同光照条件和不同设备的鲁棒性实验中均可以准确还原原始水印信息。

关键词: 深度学习; 鲁棒水印; 屏幕拍摄; 视觉质量; 恰可察觉差

中图分类号: TP393

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.2023045

Carrier-independent screen-shooting resistant watermarking based on information overlay superimposition

LI Xiaomeng¹, GUO Daidou¹, ZHUO Xunfang², YAO Heng¹, QIN Chuan¹

1. School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

2. Shanghai Data Exchange Corporation, Shanghai 201203, China

Abstract: Financial security, an important part of national security, is critical for the stable and healthy development

收稿日期: 2023-03-17; 修回日期: 2023-04-23

通信作者: 秦川, qin@usst.edu.cn

基金项目: 国家自然科学基金(U20B2051, 62172280, 62172281); 上海市自然科学基金(21ZR1444600)

Foundation Items: The National Natural Science Foundation of China (U20B2051, 62172280, 62172281), The National Natural Science Foundation of Shanghai (21ZR1444600)

引用格式: 李晓萌, 郭玳豆, 卓训方, 等. 载体独立的抗屏摄信息膜叠加水印算法[J]. 网络与信息安全学报, 2023, 9(3): 135-149.

Citation Format: LI X M, GUO D D, ZHUO X F, et al. Carrier-independent screen-shooting resistant watermarking based on information overlay superimposition[J]. Chinese Journal of Network and Information Security, 2023, 9(3): 135-149.

of the economy. Digital image watermarking technology plays a crucial role in the field of financial information security, and the anti-screen watermarking algorithm has become a new research focus of digital image watermarking technology. The common way to achieve an invisible watermark in existing watermarking schemes is to modify the carrier image, which is not suitable for all types of images. To solve this problem, an end-to-end robust watermarking scheme based on deep learning was proposed. The algorithm achieved both visual quality and robustness of the watermark image. A random binary string served as the input of the encoder network in the proposed end-to-end network architecture. The encoder can generate the watermark information overlay, which can be attached to any carrier image after training. The ability to resist screen shooting noise was learned by the model through mathematical methods incorporated in the network to simulate the distortion generated during screen shooting. The visual quality of the watermark image was further improved by adding the image JND loss based on just perceptible difference. Moreover, an embedding hyperparameter was introduced in the training phase to balance the visual quality and robustness of the watermarked image adaptively. A watermark model suitable for different scenarios can be obtained by changing the size of the embedding hyperparameter. The visual quality and robustness performance of the proposed scheme and the current state-of-the-art algorithms were evaluated to verify the effectiveness of the proposed scheme. The results show that the watermark image generated by the proposed scheme has better visual quality and can accurately restore the embedded watermark information in robustness experiments under different distances, angles, lighting conditions, display devices, and shooting devices.

Keywords: deep learning, robust watermarking, screen-shooting, visual quality, just noticeable difference

0 引言

金融领域产生的数据量巨大且内容丰富多样, 如何保证数据安全与个人信息是新时代面临的新风险与挑战。在当今信息时代, 随着计算机网络技术、多媒体信息技术的飞速发展, 人们习惯于将文件以数字的形式进行存储与分享。随着数字多媒体资源在互联网中广泛传播与使用, 人们在享受便捷的同时, 信息安全问题^[1]逐渐暴露出来, 其中就包括金融数字媒体内容安全^[2]。这些媒体内容很可能被未经授权者下载并分发, 这将严重损害内容所有者的利益甚至危害国家金融系统安全。如何保证这些多媒体资源的安全性成为亟待解决的问题。

数字水印技术能够以一种不可察觉的方式将标识信息嵌入原始图像, 通过提取这些信息可识别媒体内容的所有权, 或跟踪数字媒体内容的分发情况, 达到版权保护的目的。然而, 随着手机、相机、打印机等数码产品的普及, 数字文件不仅仅会通过网络或者硬件存储设备对文件进行传输, 还会在屏幕拍摄、打印拍摄的过程中发生。大多数传统的数字水印方案仅可以抵抗数字环境下的攻击, 如缩放、旋转、JPEG 压缩等。不同于

数字传输信道, 跨媒介传输信道会引入更加随机且复杂的噪声。与数字环境相比, 真实环境中涉及的噪声更为复杂。具体来说, 屏幕显示和相机成像涉及两次 D/A 和 A/D 转换, 这一过程不仅受到显示器类型、摄像头素质等硬件设备的影响, 还会受环境光源、拍摄角度、拍摄距离等人为因素的影响。所以相比数字噪声, 真实的屏幕拍摄噪声对图像的改变是巨大的, 这对水印信息的准确提取提出了新的挑战。因此, 设计一种能够抵抗屏幕拍摄信道传输的鲁棒水印算法具有重要意义。图 1 展示了抗屏幕拍摄水印的应用流程。



Figure 1 Application procedure of screen-shooting watermarking

此外, 现有的水印算法通常是载体图像和水印信息相融合生成一张残差图再附加在载体图像上, 将信息编码为高频信息的形式添加到载体图像上, 以此达到水印不可见的目的。这类一张载体图像对应一张残差图的嵌入方式, 实用性不强。

本文目标为让编码器只编码秘密信息,而与载体图像无关。生成的信息膜可以附加在任意的载体图像上,相比之前的工作减少了很多工作量,并且更具实际应用价值。

针对上述问题,本文提出了一种新的抗屏幕拍摄鲁棒水印方案,主要贡献如下。

1) 提出了一种基于深度学习的端到端的抗屏幕拍摄水印框架,该方案可以实现与载体图像无关的嵌入方式,生成的信息膜可以附加在任意图像上。

2) 引入了一个嵌入超参数,有效兼顾水印图像的鲁棒性和视觉质量。同时引入图像的恰可察觉差(JND, just noticeable difference)损失,进一步提升了水印图像的视觉质量。

3) 通过大量真实环境下的拍摄屏幕实验,包括不同角度、不同距离、不同设备、不同光照条件等,有效验证了本文提出的抗屏摄水印算法在含水印图像视觉质量、鲁棒性以及实用性方面优于已有算法。

1 相关工作

鲁棒水印算法主要分为两类:基于传统图像处理的鲁棒水印算法与基于深度学习^[3]的鲁棒水印算法。

Nakamura 等^[4]在 2004 年第一次提出了适用于摄像头的数字水印算法。他们提出了一种基于摄像头捕获图像的快速水印检测算法。该算法包括两个过程,一个是校正图像的几何畸变,另一个是从校正后的图像中提取水印信息。该算法对小的几何失真具有鲁棒性,此外,还介绍了一种定量评价方法,可以评价水印检测的可靠性。Pramila 等^[5]在 2018 年提出了一种增强拍摄角度鲁棒性的水印方案,使用了一种基于角度旋转的图像编码方式,借鉴 Chou 等^[6]的算法,使用 JND 来控制伪随机图案的嵌入强度,提高了水印图像的视觉质量。实验结果表明,该算法具有较强的鲁棒性,可以在拍摄角度不超过 60°的情况下正确提取水印信息。Fang 等^[7]分析了拍摄屏幕过程中产生的畸变,提出了一种基于强度的尺度不变特征转换(I-SIFT)算法,该算法可以精确定位嵌入区域;在提取方面,设计了一个基于交叉验

证的提取算法。该算法在屏幕拍摄过程的鲁棒性和提取效率方面都有显著提高。随后, Fang 等^[8]又提出了一种基于底色的屏幕拍摄弹性水印方案,为了满足摄像机拍摄过程的鲁棒性和水印的透明性,采用了交换 DCT 系数的方法来嵌入水印。与自然图像相比,文本图像的纹理更加简单,为了更好地同步水印区域,他们提出了一种基于翻转的水印同步算法来排列水印底色,从而保证了底色的强自相关性和对称性,利用行间距区域和嵌入块的对称性,提出了一种文本区域干扰的补偿方法,可以大大降低文本区域可能产生的噪声和失真。受可见光通信(VLC)的启发, Cui 等^[9]提出了一种称为 UnseenCode 的抗屏幕拍摄水印方案,利用人眼无法感知高频闪烁的特性,通过构造一个恒定亮度的融合对,在图像上叠加一个信息模板,然后以高速率在屏幕上闪烁,实现信息的隐形嵌入。在提取方面,他们分析了摄像机曝光过程,提出了一种基于彩色图像交叉分量相关性的提取算法。

近几年来,深度卷积神经网络以其优越的性能,成为研究热点,同时为抗屏幕拍摄水印技术提供了新思路。目前已经有许多学者开始探索使用深度学习^[10-14]来实现数字图像水印算法。Zhu 等^[15]提出了首个端到端的可训练的抗打印拍摄神经网络模型 Hidden,其可以同时应用在隐写术和数字水印上,在满足隐藏信息容量的情况下,既具有不错的隐蔽性,还具有鲁棒性。Liu 等^[16]提出了一种新颖的两阶段可分离深度学习水印框架。在第一阶段,采用多层特征编码策略来训练编码器,该编码器可以在不受任何噪声干扰的情况下将水印信息冗余地编码到输入图像中;在第二阶段,根据第一阶段预先训练的编码器结合真实的失真攻击生成一批噪声图像,利用这批图像对解码器进行训练,达到增强解码器鲁棒性的目的。受到 HiDDeN 以及图像分割网络 U-Net^[17]的启发, Tancik 等^[18]将 U-Net 网络用于端到端框架的水印生成步骤,在提取步骤中使用了 STN^[19]作为矫正网络,以增强模型对于图像形变的鲁棒性,同时使用生成对抗网络监督编码器的训练,以最小化水印图像的感知失真。Wengrowski 等^[20]提出了一个端到端的光场信息的图像隐写算法,

他们首先生成一个大规模的“屏幕-摄像头”数据集来训练模拟网络,然后将该网络添加到噪声层进行训练,以实现良好的抗屏幕拍摄鲁棒性。Fang等^[21]提出了一种基于深度模板的屏幕拍摄弹性水印算法。在嵌入方面,本文首先设计了两个用于信息嵌入和定位的模板,这两个模板是利用人类视觉系统的特殊属性实现的。在提取方面,本文提出了一种新颖的两级深度神经网络,它由一个辅助增强子网络和一个分类子网络组成。该方案生成的水印图像可以实现较好的视觉质量,但是鲁棒性一般。

一般来说,基于深度学习的抗屏幕拍摄水印方案都是通过将水印信息与载体图像相融合的方式生成水印图像。然而,这种方式实用性不强,因此,本文设计了一种新的端到端水印框架,不需要修改载体图像,当需要在多张载体图像中嵌

入相同的信息时,只需要利用编码网络生成一次信息膜,相比之前的研究减少了很多工作量。

2 端到端网络架构

本文算法的端到端架构主要包括编码网络、噪声网络、解码网络和鉴别器网络4个模块。待嵌入信息输入编码网络中生成信息膜,附加在载体图像上生成水印图像。利用噪声网络对水印图像进行攻击,使模型学习抵抗屏幕拍摄噪声的能力。随后,将附加噪声的水印图像输入解码网络以提取水印信息。同时,使用一个鉴别器网络来监督编码网络的训练,两者进行对抗训练,以最小化水印图像的视觉感知。为了更灵活地权衡水印图像的鲁棒性和透明性,本文在训练过程中设置了一个嵌入超参数。所提抗屏摄鲁棒水印端到端网络架构如图2所示。

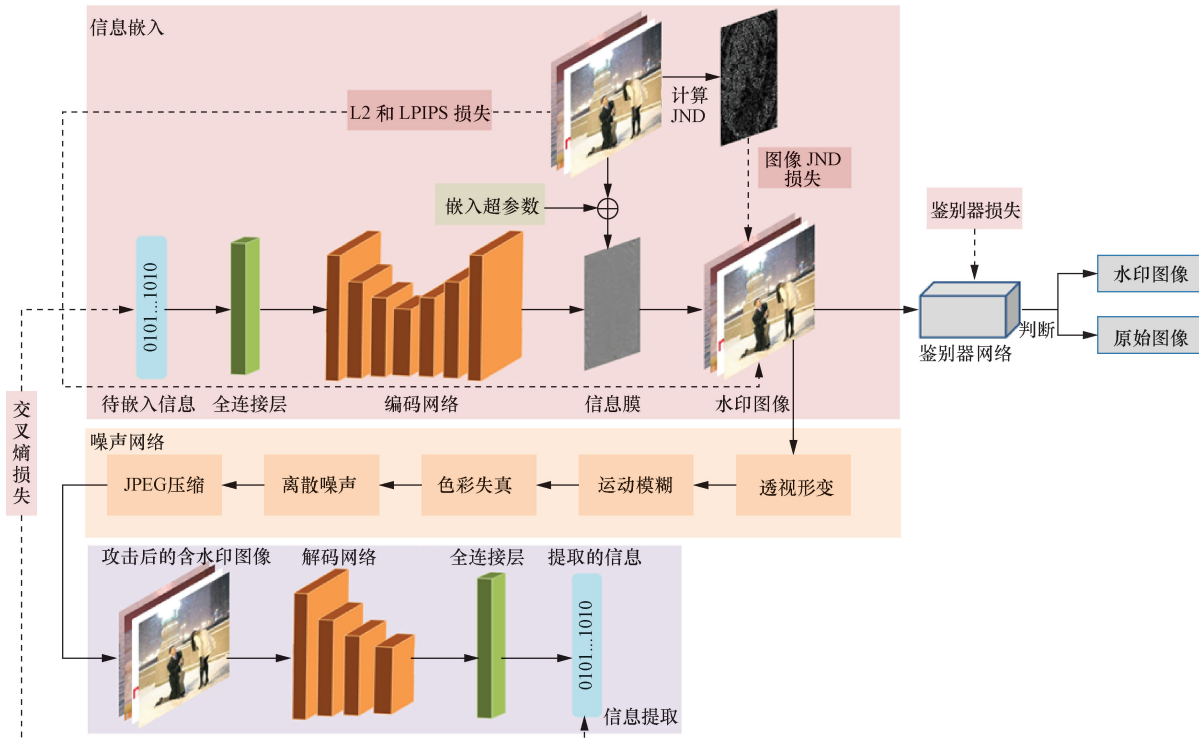


图2 所提抗屏摄鲁棒水印端到端网络架构

Figure 2 The proposed end-to-end framework of screen-shooting resistant image watermarking

2.1 编码网络

本文选择与 U-Net 相似的网络架构作为编码器,具体的编码器网络结构如图3所示。编码网络的输入是一个长度为 L 的随机二进制字符串。首先将水印信息通过一个线性全连接层进行预处

理,一维的水印信息被扩充形成一个 $50 \times 50 \times 3$ 的张量,然后上采样形成 $400 \times 400 \times 3$ 的张量。通过一系列的深度卷积计算和特征学习,输出一张仅含有水印信息的信息膜。实际使用时,只需要用训练得到的信息膜直接叠加到任意载体图像就可

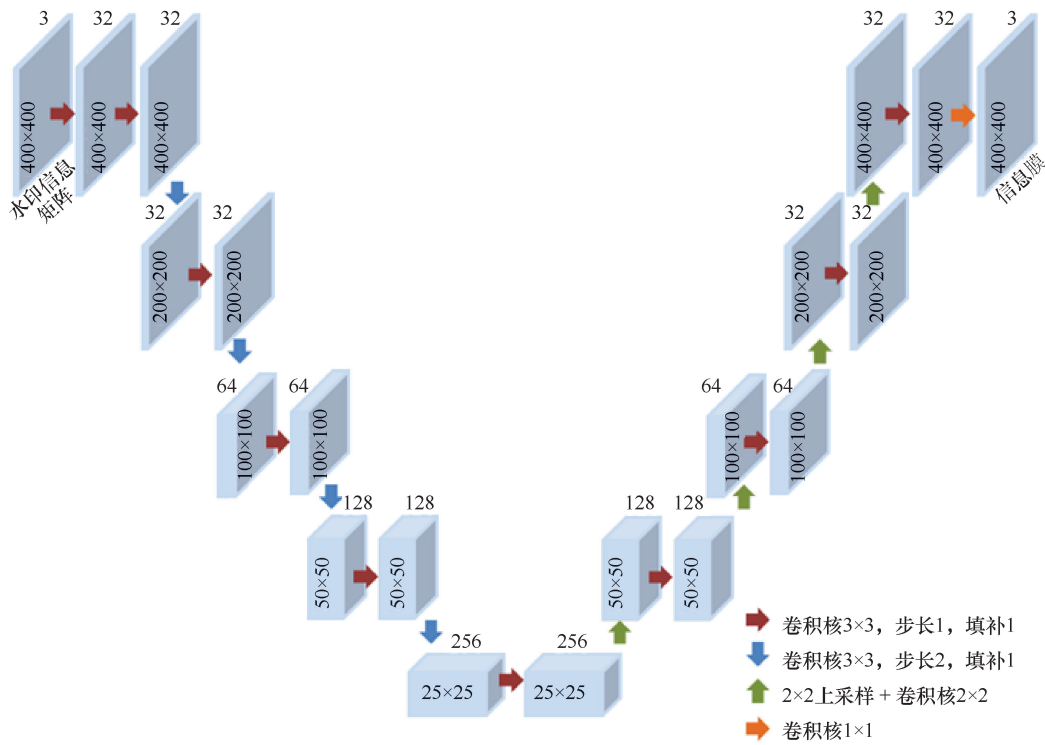


图3 编码器网络结构

Figure 3 The architecture of watermarking encoder

以实现水印的嵌入。图4中显示了一些信息膜的示例，其中图4(a)~图4(c)分别是输入水印信息为“xm”“zy”“ab”时模型生成的信息膜。模型训练时采用信息膜与水印图像相加的嵌入方式来生成水印图像，同时在信息膜前乘上一个嵌入超参数。

$$I_{\text{en}} = I_o + \alpha \cdot I_m \quad (1)$$

其中， I_o 表示原始载体图像， I_{en} 表示水印图像， I_m 表示带有水印信息的信息膜， α 表示嵌入超参数。实验证明，训练过程中通过自适应地改变 α 的大小，可以有效调整模型所生成的水印图像的视觉质量，这样可以帮助生成适合不同场景和需求的水印模型。在训练过程中设计了相应的损失函数来监督水印图像的视觉质量，这一过程的损失函数将在第3节中详细解释。水印图像视觉质量实验的具体细节见第4节。

2.2 噪声网络

为了增强水印图像对于现实世界噪声的鲁棒性，在训练过程中，本文参考StegaStamp^[18]的方法在编码器和解码器之间设置噪声层，以近似现实世界中物理显示和摄像机成像过程中引起的各

种类型的失真，模拟真实的屏幕拍摄噪声过程如图5所示。

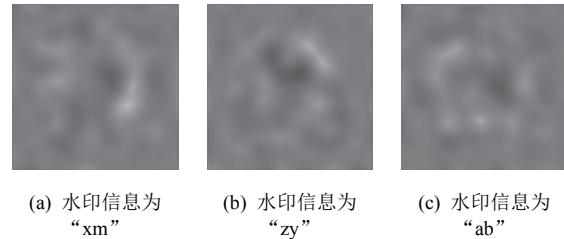


图4 信息膜示例

Figure 4 Example of information overlay



图5 模拟真实的屏幕拍摄噪声过程

Figure 5 Simulation process for real world screen-shooting noise

2.2.1 透视变形

摄像头拍摄屏幕的过程中，很难保证没有角度变化，因此拍摄出来的图片很容易出现透视变形。使用随机单应性透视变换可以模拟这种噪声。假设原图的4个角的坐标为 (a_1, b_1) ， (a_2, b_2) ， (a_3, b_3) ， (a_4, b_4) ，则可以通过以下公式得到变换

后的新坐标 (a'_1, b'_1) , (a'_2, b'_2) , (a'_3, b'_3) , (a'_4, b'_4) 。

$$a'_i = \omega(a_i - b_w, a_i + b_w) \quad (2)$$

$$b'_i = \omega(b_i - b_w, b_i + b_w) \quad (3)$$

其中, ω 表示取其两个参数中的随机值, $i=1,2,3,4$, a_w 和 b_w 可以通过设置超参数 h , 并用以下公式进行计算。

$$a_w = \omega(0,1) \cdot h \quad (4)$$

$$b_w = \omega(0,1) \cdot h \quad (5)$$

最后, 通过双线性插值法获得透视变换后的图像, 记为 I_p 。

2.2.2 运动模糊

在屏幕拍摄过程中, 摄像机的移动和不准确的自动对焦都可能导致图像模糊^[22], 可以借鉴随机轨迹生成算法^[23]来模拟运动模糊。本文设置了一个参数来控制模糊核的方差, 生成一个宽度在 3 到 7 像素之间的直线高斯模糊核 G , 并将随机旋转角度控制在 0 到 2π 之间, 记 $G'(x', y')$ 为旋转后的高斯核, 计算过程如下:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (6)$$

$$G'(x', y') = G(x, y) \quad (7)$$

其中, (x, y) 表示距卷积核中心的坐标, (x', y') 表示旋转后的卷积核的坐标, 旋转所用的公式如下:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} \quad (8)$$

因此, 模糊的图像可以表示为

$$I_b(i, j) = \sum_{m=1}^7 \sum_{n=1}^7 I_p(i+m-3, j+n-3) \cdot G'(x', y') \quad (9)$$

其中, (i, j) 表示 I_b 的坐标, (x', y') 表示旋转后的高斯卷积核的坐标。

2.2.3 随机离散噪声

摄像机成像时会引入的噪声通常分为光子噪声、暗噪声和散粒噪声^[24]。因此, 噪声函数的响应可以用与信号相关的标准差来描述^[25]。本文采用高斯噪声 (Ng) 模型。

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (10)$$

$$I_d = I_b + N_g \quad (11)$$

本文使用标准偏差 $\sigma \sim U[0, 0.03]$ 计算成像噪声。将创建的噪声添加到原始图像的空间域中, I_d 是添加随机离散噪声后的图像。

2.2.4 色域噪声

显示器显示和摄像机拍摄是一系列数字到模拟和模拟到数字转换的过程, 这会使数字图像产生不同程度的色彩上的失真, 包括色调偏移、亮度调整和饱和度变化。这些失真可以通过模拟色彩区域的随机色彩变换来实现。

图像色调整: 为每个色彩域加上随机的空间域补偿。色彩补偿由随机范围设定在 -0.1 到 0.1 之间的参数乘上 I_d 的 RGB 分量生成, 记为 I_h 。

图像亮度调整: 调整图像亮度和对比度属于像素变换操作, 本文使用线性变换生成亮度调整的值, 公式如下。

$$I_c(i, j) = \lambda \cdot I_h(i, j) + \beta \quad (12)$$

其中, β 是改变亮度的参数, 它的值在 -0.3 到 0.3 之间, λ 是改变对比度的参数, 它的值在 0.5 到 1.5 之间。

饱和度调整: 首先将图像从 RGB 色域转换到 YUV 色域, 并且选择 Y 分量作为图像亮度的参考值, 标记为 I_y , 之后使用线性公式组合原图与 Y 分量。

以上步骤可以综合为

$$I_f = (1+t) \cdot (I_c + I_h) + t \cdot I_y \quad (13)$$

其中, t 的取值 $[-0.1, 0.1]$, 用于控制色相变化的强度。

2.2.5 JPEG 压缩

使用数码相机拍摄后, 图像通常会以 JPEG 格式保存, 其主要原理是将图像分为 8×8 的数据块, 并对它们进行离散余弦变换(DCT), 然后四舍五入为最接近的整数来量化所得的系数。因此, 可以利用 Shin^[26]的方法, 用分段函数逼近量化步长。

$$z(w) = \begin{cases} w^3, & |w| < 0.5 \\ w, & |w| \geq 0.5 \end{cases} \quad (14)$$

其中, $z(w)$ 是 JPEG 压缩后的结果, w 表示该点的像素值。

2.3 解码网络

解码器主要用于将编码图像下采样为一列二进制字符串, 本文解码器的网络架构如图 6 所示。为了提高在摄像头捕获和校正编码图像期间对角度变化的鲁棒性, 本文采用空间变换网络 (STN) [17] 将带有噪声的编码图像输入解码器。由于 STN 需要一个线性层来拟合透视变换的权重, 所以使用了几个卷积层和 ReLU 层来产生图像的空间变换参数。根据空间变换参数, STN 会将图像进行仿射变换, 修正形变。使用 STN 可以使水印提取层对屏幕拍摄过程产生的变形更加鲁棒。最后, 解码器结合一个全连接层和一个 ReLU 层来产生与水印数据长度相同的二进制输出。

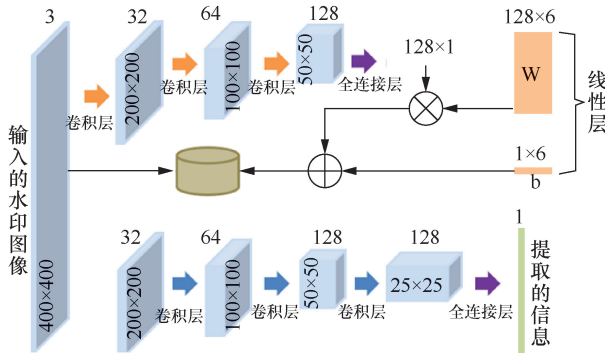


图 6 解码器的网络结构
Figure 6 The architecture of watermarking decoder

2.4 鉴别器网络

为了提高编码图像的质量, 实现较低的感知失真, 本文使用一个鉴别器与编码器进行对抗训练。鉴别器网络由 5 个卷积层和最大池化层组成, 其输出是一个二进制数字 0/1, 指示图像是否带有水印。在训练过程中, 当鉴别器不再能够区分原始图像和编码图像时, 训练过程达到平衡。作为整体损失的一部分, 本文使用 Wasserstein 损失 [27] 来监督判别器。

3 多任务损失函数设计

抗屏幕拍摄水印算法是一个多任务流程模型, 其中包括水印嵌入网络、水印提取网络、鉴别器网络的训练, 整体的训练目标是尽可能小地

改变图像, 并同时保证正确提取水印信息。为了同时达成这些目标, 需要设计一个多任务的损失函数对整个流程的所有训练目标进行整合、统一训练, 其中包括水印嵌入网络对于图像质量的监督损失函数、水印提取网络的水印信息损失函数、鉴别网络判断图像是否带有水印的损失函数。

3.1 图像质量损失函数

当水印信息嵌入原始图像 I_o 中时, 本文的主要目标是保证高解码鲁棒性, 同时保证编码后的水印图像 I_{en} 具有较好的视觉质量。本文首先将 I_o 和 I_{en} 都转换到 YUV 色域, 并使用 L_2 损失函数评估这两个图像的相似度。

$$L_2(I_o, I_{en}) = \|I_o - I_{en}\|^2 / (X \cdot Y \cdot Z) \quad (15)$$

其中, $(X \cdot Y \cdot Z)$ 代表了图像的尺寸。

然而, L_2 损失是通过像素到像素的方式计算两个图像的相似性, 忽略了图像的结构信息。因此, MSE 的相似性评估结果有时会与人眼的视觉感知不一致。在此基础上, 本文引入图像感知损失函数进一步对图像质量进行监督, 可以表示为 $LPIPS(I_o, I_{en})$ 。

JND 用于表示人眼不能察觉的最大图像失真, 体现了人眼对图像改变的容忍度。在图像处理领域, JND 可以用来度量人眼对图像中不同区域失真的敏感性。本文可以通过以下公式来计算图像的 JND。

$$JND(i, j) = \delta_1 \cdot (k_1(p(i, j), q(i, j)) + \delta_2) + k_2(l(i, j)) \quad (16)$$

其中, 使用 k_1 函数和 k_2 函数分别对空间掩蔽分量和基于背景亮度的视觉阈值进行建模。

$$k_1(p(i, j), q(i, j)) = q(i, j) \cdot g_1 \cdot (p(i, j)) + g_2 \cdot (p(i, j)) \quad (17)$$

其中, $q(i, j)$ 表示像素点 (i, j) 周围 4 个方向的亮度改变量的最大平均值, 计算式如下。

$$q(i, j) = \max_{k=1,2,3,4} \{|\text{grad}_k(i, j)|\} \quad (18)$$

$$\text{grad}_k(i, j) = \frac{1}{16} \sum_{a=1}^5 \sum_{b=1}^5 f(i-3+a, i-3+b) \cdot W_k(a, b) \quad (19)$$

其中 $W_k(a, b)$ 是用于检测纹理的高通滤波器。

$$W_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 3 & 8 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ -1 & -3 & -8 & -3 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (20)$$

$$W_2 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 8 & 3 & 8 & 0 \\ 1 & 3 & 0 & -3 & -1 \\ 0 & 0 & -3 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{bmatrix} \quad (21)$$

$$W_3 = \text{transpose}(W_2) \quad (22)$$

$$W_4 = \text{transpose}(W_1) \quad (23)$$

$g_1(p(i, j))$ 和 $g_2(p(i, j))$ 是与背景亮度相关的函数，它们分别表示 k_1 的斜率与与能见度阈值轴的交点，计算式如下：

$$g_1(p(i, j)) = p(i, j) \cdot 0.0001 + 0.115 \quad (24)$$

$$g_2(p(i, j)) = \delta - p(i, j) \cdot 0.1 \quad (25)$$

其中， $p(i, j)$ 表示背景亮度的平均值，通过一个带权的低通滤波器 $M(a, b)$ 实现。

$$l(i, j) = \frac{1}{32} \sum_{i=1}^5 \sum_{j=1}^5 M(a, b) \quad (26)$$

$$M = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 & 1 \\ 1 & 2 & 0 & 2 & 1 \\ 1 & 2 & 2 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \quad (27)$$

式(28)中的 k_2 用来计算表示背景亮度的视觉阈值。为了方便表示，将 $l(i, j)$ 用变量 t_n 来代替。

$$k_2(t_n) = \begin{cases} T_0 \cdot \left(1 - \sqrt{\frac{t_n}{127}}\right) + 3, & t_n \leq 127 \\ \varphi \cdot (t_n - 127) + 3, & t_n > 127 \end{cases} \quad (28)$$

其中， T_0 表示背景灰度值为 0（即黑色）时的可见度阈值， φ 表示背景亮度较高的情况下对函数进行建模的函数的斜率^[6]。经过上述公式的计算，如图 7 所示，可以得到原始图像的 JND 图像，很难区分原始图像、生成的 JND 图像及其与原始图

像的组合。



图 7 原始图像与其对应的 JND 图像示例
Figure 7 An example of image JND map

受到 JND 的启发，本文构造了一个基于 JND 的损失函数用于监督水印图像的视觉质量。

$$L_{I-jnd} = \text{LPIPS}(L_I, \tau \cdot I_{jnd}) \quad (29)$$

其中 L_{I-jnd} 表示该损失函数的损失值，LPIPS 函数用于计算生成的水印图像与原始图像的 JND 的感知相似度， τ 是一个超参数，用于调整 JND 的权重。

3.2 水印信息损失函数

为了保证水印图像中嵌入的信息能够被正确解码，即所提算法的鲁棒性，本文引入损失函数来监督解码网络的训练。将原始水印信息表示为 W_o ，将被攻击的水印图像输入解码网络后的解码信息表示为 W_{de} 。这两组信息的表示方式都是二进制的，解码网络的最终输出可以看作每个比特为 0 或 1 的概率。本文使用交叉熵损失来计算两组信息之间的差异。

$$L_w(W_o, W_{de}) = - \sum_{n=1}^{W_l} W_o(n) \log(W_{de}(n)) \quad (30)$$

其中， $W_o(n)$ 代表原始水印信息的第 n 位， $W_{de}(n)$ 代表提取信息的第 n 位， W_l 代表信息长度。信息熵越大意味着两段信息的差异越大，故该网络的训练目标是将该损失函数的输出值尽可能地小。

3.3 鉴别器网络损失函数

鉴别器网络的输出是一个二进制数，指示图像是否带有水印。输入原始图像 I_o 进入网络后的输出为 $D(I_o)$ ，输入水印图像 I_{en} 进入网络后的输出为 $D(I_{en})$ ，则鉴别器网络的损失函数可以表示为

$$L_D(I_o, I_{en}) = D(I_o) - D(I_{en}) \quad (31)$$

在训练过程中， $D(I_o)$ 的值无限趋近于 0， $D(I_{en})$ 的值无限趋近于 1，因此该网络的训练目

标是使得式(31)的值越来越小。

综上所述，本文的端到端网络架构的多任务损失函数最终可以定义为

$$L(I_o, I_{en}, I_{jnd}, W_o, W_{de}) = s_1 \cdot L_2(I_o, I_{en}) + s_2 \cdot L_{I-jnd} + s_3 \cdot \text{LPIPS}(I_o, I_{en}) + s_4 \cdot L_w(W_o, W_{de}) + s_5 \cdot L_D(I_o, I_{en}) \quad (32)$$

其中， s_1 、 s_2 、 s_3 、 s_4 、 s_5 是可以调节的超参数。

3.4 训练细节

本文方案的首要目标是保证解码器提取信息的准确率。因此在训练开始时，本文将 s_1 、 s_2 、 s_3 和 s_5 的值初始化为0，优先保证解码器的鲁棒性，当解码器的位提取精度接近100%时， s_1 、 s_2 、 s_3 和 s_5 的值逐渐线性增加，最终分别稳定在2、1.5、0.5和1，保证了水印图像的视觉质量和鉴别器的有效性。值得注意的是， s_4 始终是一个常量，以确保鲁棒性，本文将其设置为1.5。为了模拟数字图像在屏幕拍摄过程中可能发生的形变，本文利用透视变形对原始图像和形变图像进行映射，这会使模型学习在图像边缘添加干扰水印信息，导致水印图像的边缘产生阴影。本文使用余弦下降的方法增加边缘处 L_2 损失的权重来减轻这种不良影响。

4 实验结果与分析

本节比较了本文方案和一些已有的代表性方案，并从视觉质量、鲁棒性等方面评价了整体性能。

4.1 实验设置

模型训练过程中采用COCO数据库^[28]作为原始图像的数据集。所有实验图像的大小调整为 400×400 ，测试图像被排除在训练集之外。在所有的性能评估实验中，如水印图像的鲁棒性和视觉质量，本文将嵌入消息长度定为32 bit。在水印嵌入前，利用BCH(32,16)的纠错编码对水印信息进行处理，可以纠正3 bit错码。详细地说，在32 bit中，16 bit是水印信息位，16 bit是纠错位。并且所有的实验都是在Nvidia Titan RTX GPU以及2.30 GHz/12核心Xeon Gold 5118 CPU的环境下进行的。真实环境下的屏幕拍摄实验使用的显示器和拍摄设备分别是Macbook Pro和iPhone XS。

对于水印图像的视觉质量，本文使用峰值信

噪比(PSNR)、结构相似性指数(SSIM)和学习感知图像块相似度(LPIPS)作为评价指标。对于这些指标，PSNR和SSIM的值越大，LPIPS的值越小，意味着图像的视觉质量越好(较低的失真)。另外，对于真实环境和数字环境下的鲁棒性测试，本文设计了两个指标来评估模型的性能，即没有纠错编码的直接提取的平均准确率(ε_1)以及经过纠错编码还原后的平均准确率(ε_2)，计算式如下：

$$\varepsilon_1 = \frac{\sum_{n=1}^N N_{bc}}{N \cdot L_{bs}} \quad (33)$$

$$\varepsilon_2 = \frac{N_c}{N} \quad (34)$$

其中， N_{bc} 是每个提取的比特串中正确恢复的比特数， L_{bs} 是每个比特串的长度， N 是测试图像数量， N_c 是可正确提取水印信息的图像数量。

4.2 含水印图像视觉质量比较

实验中，本文选取同样是基于深度学习的StegaStamp^[18]作为对比对象。为了测试训练模型时所设置的嵌入超参数的有效性，实验过程中多次改变嵌入超参数 α 的大小，分别训练了模型($\alpha=1.5, 1.6, 1.7$)，并使用模型生成的信息膜与载体图像直接叠加完成水印嵌入，同时观察其对图像质量的影响。实验选用COCO图像库的103张图像(不包括训练集的图像)作为载体图像，统一设置嵌入信息的长度为32 bit，实验结果如表1所示。

表1 含水印图像的平均PSNR、SSIM和LPIPS
Table 1 Average PSNR, SSIM and LPIPS of watermarked images

算法	PSNR/dB	SSIM	LPIPS
StegaStamp ^[18]	27.24	0.902 6	0.0511
本文算法($\alpha=1.5$)	28.64	0.967 2	0.009 9
本文算法($\alpha=1.6$)	30.56	0.968 0	0.009 3
本文算法($\alpha=1.7$)	31.76	0.974 0	0.007 5

从表中可以看出，本文算法($\alpha=1.5$)相比StegaStamp，PSNR的平均值提高了1.4 dB，SSIM的平均值提高了0.064 6，LPIPS的平均值降低了0.041 2，由此可见，本文算法有效地提升了水印图像的视觉质量。图8列出了原图以及4种算法

训练的模型输出,对比图 8(b)和图 8(c)可以看到,本文算法生成的水印图像在图像大块纯色区域,阴影减少,较好地保留了图像的细节,水印图像的视觉质量有明显提升。对比图 8(c)、图 8(d)、图 8(e)可以看出,使用嵌入超参数为 1.7 的模型所生成的水印图像视觉质量最好,但是鲁棒性有一定程度的下降。



图8 含水印图像视觉质量比较

Figure 8 Comparisons of visual quality for watermarked images

4.3 鲁棒性比较

为了更加客观全面地评价本文算法的鲁棒性,本文对模型生成的水印图像进行了大量实验,包括数字环境下的鲁棒性实验和真实拍摄屏幕的鲁棒性实验,所有的鲁棒性实验均选择嵌入超参数 $\alpha=1.5$ 的网络模型。

4.3.1 数字环境下的鲁棒性实验

本文首先考虑了数字环境的情况,即水印图像不需要真实环境下的屏幕拍摄操作。实验中,本文选取了5种常见的数字信道噪声作为攻击手段,分别是缩放攻击、JPEG压缩、高斯模糊、裁剪攻击、边缘遮盖。本文从 COCO 数据库中选择了 103 个图像(不包括训练集的图像)作为水印嵌入的原始载体图像,并计算提取精度 ε_1 和 ε_2 ,实验细节如下。

缩放攻击:在真实的屏幕拍摄过程中,拍摄距离的远近会导致图像产生不同程度的缩放。本

文对实验图像进行了不同比例的缩放来模拟这种攻击。在实验过程中,分别选取 4/5、3/4、3/5、1/2、2/5、1/4 和 1/5 作为缩放比例,对应的图像尺寸分别为 320×320 、 300×300 、 240×240 、 200×200 、 160×160 、 100×100 和 80×80 。

高斯模糊:高斯核的大小直接决定了对图像的破坏程度,故在实验中分别选用了长宽为 13、15、17、19 和 21 的高斯核,作为添加噪声运算时的卷积核。

JPEG 压缩:在数字环境下,为了节省存储空间并增加传输效率,图像数据极少会采用无损编码的方式进行存储,而通常采用 JPEG 格式对图像进行。压缩生成的水印图像在经过网络传输后,也可能被网页渲染器或者文件传输器二次压缩,因此算法对 JPEG 压缩攻击的鲁棒性十分重要。影响 JPEG 压缩后图像质量的参数是质量因子,在实验中本文选取 40、50、60、70、80 和 90 作为重新存储图像的 JPEG 质量因子取值。

裁剪攻击:实际的屏幕拍摄过程中,经常会有异物遮挡带有水印的图像,导致拍摄到的图像中心缺失。为了验证模型对裁剪攻击的鲁棒性,在数字环境下,本文选用不同尺寸大小的黑色方块区域,遮盖水印图像的中心部分。实验中,本文选用 20×20 、 40×40 、 60×60 、 80×80 、 100×100 和 120×120 的黑色像素方块遮盖嵌入水印后的图像的中心位置。

边缘遮盖:在屏幕拍摄过程中,相比中心遮挡物,由于拍摄角度产生的边缘遮挡发生的概率更大,与中心裁剪做法相反,实验中,本文采用四周加边框的方式为图像增加边缘遮挡,遮挡时使用的像素值 RGB 为 (0, 0, 0),加框的厚度分别为 10、20、30、40 和 50 个像素。

以上 5 类数字噪声攻击的实验结果如表 2 所示。从表 2 中不难看出,所提模型可以抵抗数字环境下的多类噪声,在裁剪和遮挡攻击下也能较为准确地提取信息,并且随着噪声强度的增大,模型的提取准确率也较为稳定。

4.3.2 真实屏摄环境下的鲁棒性实验

相比数字环境,真实环境下拍摄屏幕实验所产生的叠加噪声更加复杂且随机,实验结果更能反映算法的鲁棒性。实验中,本文从测试数据集

中随机选取 12 张图像作为原始载体图像，在叠加含有水印的信息膜后显示在屏幕上并使用智能手机进行拍摄，实验环境如图 9 所示。为了更全面地测试本文算法的鲁棒性，本文分别进行了不同拍摄距离、不同拍摄角度、不同拍摄环境、不同拍摄设备和不同显示设备的鲁棒性实验。

表 2 不同数字噪声攻击下的提取准确率
Table 2 Extraction accuracy with different digital noise attacks

攻击类型	攻击强度	本文算法	
		ε_1	ε_2
缩放攻击 (缩放比例)	4/5	97.15%	96.12%
	3/4	97.15%	95.15%
	3/5	97.18%	97.09%
	1/2	97.00%	94.17%
	2/5	96.70%	94.17%
	1/4	96.42%	93.20%
	1/5	95.55%	89.47%
高斯模糊 (高斯核大小)	13	97.03%	96.12%
	15	96.54%	96.12%
	17	95.18%	92.23%
	19	93.05%	80.58%
	21	90.96%	67.96%
JPEG 压缩 (质量因子大小)	40	96.30%	90.29%
	50	96.63%	94.17%
	60	96.75%	96.12%
	70	96.69%	96.12%
	80	96.91%	95.15%
	90	96.88%	97.09%
裁剪攻击 (裁剪区域尺寸)	20 × 20	97.09%	97.09%
	40 × 40	96.94%	94.17%
	60 × 60	97.21%	98.06%
	80 × 80	96.91%	97.09%
	100 × 100	96.18%	97.09%
	120 × 120	94.45%	93.20%
边缘遮盖 (遮盖宽度)	10 pt	95.93%	90.29%
	20 pt	95.27%	87.38%
	30 pt	95.33%	90.29%
	40 pt	94.66%	89.32%
	50 pt	94.57%	82.52%



图 9 屏幕拍摄实验环境
Figure 9 Screen-shooting experimental environment

不同拍摄距离的鲁棒性实验：真实环境下的拍摄过程中，拍摄者与显示屏上的图像的距离不是恒定不变的，因此拍摄出来的图像会存在一定程度上的缩放。另外，随着摄像头的移动，在拍摄过程中会出现摩尔纹现象^[29]，使水印信息提取的难度加大。图 10 显示了不同拍摄距离下的屏幕拍摄图像及其对应的透视校正图像示例。实验中，改变拍摄距离对两种基于深度学习的算法进行鲁棒性测试。本文使用经过纠错编码还原后的平均准确率 ε_2 作为评价标准，实验结果见表 3。从中可以发现，StegaStamp 模型在近距离拍摄的情况下，提取准确率降低，这是由于近距离拍摄会导致摩尔纹现象更加严重，而本文提出的算法在 20 cm 的距离下依旧可以实现水印信息的正确提取，对于抵抗摩尔纹攻击的能力更强。



图 10 不同拍摄距离下的屏幕拍摄图像示例
Figure 10 Screen shooting images at different shooting distances

表 3 不同拍摄距离下的提取准确率
Table 3 Extraction accuracy with different shooting distances

拍摄距离/cm	ε_2	
	StegaStamp ^[18]	提出的算法
20	83.33%	100.00%
35	100.00%	100.00%
55	100.00%	100.00%

不同拍摄角度的鲁棒性实验：屏幕拍摄过程中，拍摄角度的偏差是无法避免的，这将会导致实际拍摄到的图像比真实显示的图像出现不同程度的形变，相比原图，失真程度更大，因此，对不同拍摄角度的鲁棒性进行测试，可以更真实地反映出在实际拍摄过程中算法的鲁棒性。如表 4 所示，在实验中选择了 45° 、 30° 和 15° 作为拍摄位置与水印图像所在位置的夹角大小，并且用正负号分别代表从左侧和右侧拍摄。在实际测试中，光照强度被控制在 150 lux，每次的拍摄距离设定

为 35cm。图 11 显示了不同拍摄角度下的屏幕拍摄图像及其对应的透视校正图像示例。本文与最先进的抗拍摄水印算法做了比较，其中 Kang 等^[30]、Nakamura 等^[4]和 Pramila 等^[31]是基于传统图像处理的算法。实验结果如表 4 所示，从中可以看到，两种基于深度学习的算法，在鲁棒性上远远高于传统的基于图像处理的算法。本文提出的算法在不同角度的拍摄下也可以保证鲁棒性，通过纠错编码后可以准确地还原原始水印信息。

表 4 不同拍摄角度下的提取准确率
Table 4 Extraction accuracy with different shooting angles

拍摄角度	ε_1	ε_1	ε_1	ε_1	ε_2	ε_1	ε_2
	Kang 等 ^[30]	Pramila 等 ^[4]	Nakamura 等 ^[31]	StegaStamp ^[18]		本文算法	
+45°	58.89%	55.28%	83.69%	99.22%	100.00%	97.19%	100.00%
+30°	72.17%	57.00%	86.81%	99.22%	100.00%	97.19%	100.00%
+15°	80.47%	54.88%	77.73%	99.48%	100.00%	97.19%	100.00%
0°	67.50%	55.95%	84.56%	99.74%	100.00%	96.56%	100.00%
-15°	74.31%	64.03%	86.13%	99.74%	100.00%	98.13%	100.00%
-30°	73.52%	63.48%	83.69%	99.74%	100.00%	98.13%	100.00%
-45°	80.08%	71.39%	84.17%	98.69%	100.00%	97.19%	100.00%

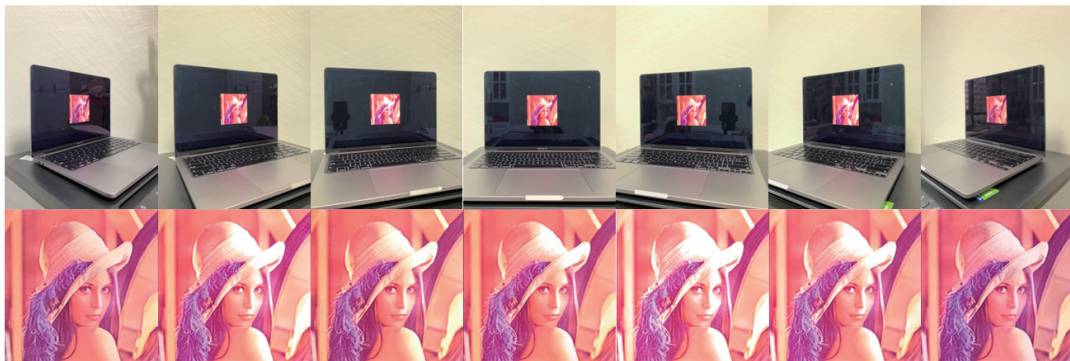


图 11 不同角度下的屏幕拍摄图像示例
Figure 11 Screen-shooting images at different shooting angles

不同光照条件的鲁棒性实验：在不同的光照环境下进行拍摄，会对图像引入不同的亮度、对比度和色相变化。与此同时，在低光照条件下拍摄的图像需要更长的曝光时间，会造成图像抖动，并且高感光度还会带来更多的噪点，这将引入一系列复杂的失真。为了验证模型不同光照条件下的鲁棒性，本文分别在光照强度为 50 lux、150 lux 和 500 lux 的环境下统一使用 iPhone XS 在 35 cm

的距离下进行屏幕拍摄实验，图 12 显示了不同光照强度下的屏幕拍摄图像及其对应的透视校正图像示例，实验结果如表 5 所示。可以看出，不论是在低光照条件下还是在强光照条件下，本文模型都具有较好的鲁棒性。

不同显示设备的鲁棒性实验：市面上的显示器类型多种多样，衡量显示器性能的参数主要包括像素点距、分辨率、色彩纯净度等。不同型号

的显示器呈现出的图像色彩、色温、清晰度等各不相同。此外，不同显示设备的屏幕拍摄实验对于测试模型的鲁棒性和普适性具有重要意义。实验中，本文统一使用 iPhone XS 作为拍摄设备，在 150 lux 光照环境、35 cm 拍摄距离下分别使用 MacBook Pro 2019、Dell G15 5511 和 AOC I2490PXZ 显示图像，图 13(a)、图 13(b)、图 13(c) 为不同设备下的屏幕拍摄图像示例。从表 6 中可以看出，在不同的显示设备下，本文的模型配合 BCH 纠错码可以准确还原水印信息。



图 12 不同光照强度下的屏幕拍摄图像示例

Figure 12 Screen shooting images at different illumination

表 5 不同光照环境下的提取准确率

Table 5 Extraction accuracy with different illumination

光照强度/lux	ε_1	ε_2	ε_1	ε_2
	StegaStamp ^[18]		本文算法	
50	100.00%	100.00%	98.75%	100.00%
150	99.74%	100.00%	96.56%	100.00%
500	99.74%	100.00%	98.75%	100.00%

表 6 不同设备下的提取准确率

Table 6 Extraction accuracy with different devices

设备	名称	ε_1	ε_2
拍摄设备	iPhone XS	96.56%	100%
	Xiaomi 11	97.44%	100%
	HUAWEI P40 Pro+	98.18%	100%
显示设备	MacBook Pro 2019	96.56%	100%
	Dell G15 5511	98.18%	100%
	AOC I2490PXZ	98.32%	100%

不同拍摄设备的鲁棒性实验：不同拍摄设备的像素分辨率和光圈大小会显著影响拍摄图像的质量。像素分辨率越高，光圈越大，拍摄出的图像质量越好，失真程度越小。实验中，本文统一使用 MacBook Pro 作为显示设备，分别使用

iPhone XS、Xiaomi 11 和 Huawei P40 Pro+在 150 lux 光照环境、35 cm 距离下进行拍摄，图 13(d)、图 13(e)、图 13(f)为不同拍摄设备下的屏幕拍摄图像示例。从表 6 中可以看出，本文算法在不同的拍摄设备下，可以在纠错解码后准确地提取水印信息。



图 13 不同设备下的屏幕拍摄图像示例

Figure 13 Screen-shooting images at different shooting devices

5 结束语

本文提出了一种抗屏幕拍摄攻击的端到端鲁棒图像水印算法。为了提高模型的普适性，本文的编码器经过训练后，通过输入信息可以生成通用的信息膜，这种不需要载体图像参与的嵌入方式更具有实用意义，在一定程度上减少了工作量，提高了水印嵌入效率。此外，根据图像的恰可察觉差，设计了图像的 JND 损失，结合多任务损失函数进一步监督编码图像的视觉质量，有效减少图像的失真。同时，在训练过程中引入了一个嵌入超参数，通过调整嵌入超参数可以更加灵活地编码图像的视觉质量和鲁棒性，自适应地训练出适合不同场景的模型。大量的实验表明，本文算法的整体性能优于一些先进算法。

参考文献：

- [1] 李凤华. 信息技术与网络空间安全发展趋势[J]. 网络与信息安全学报, 2015, 1(1): 8-17.
- [2] LI F H. Development trends of the information technology and cyberspace security[J]. Chinese Journal of Network and Information Security, 2015, 1(1): 8-17.
- [3] 靳玉红. 大数据环境下互联网金融信息安全防范与保障体系研

- 究[J].情报科学, 2018, 36(12): 134-138.
- JIN Y H. Internet financial information security and protection systems in big data environment[J]. Information Science, 2018, 36(2): 134-138.
- [3] 王馨雅, 华光, 江昊, 等. 深度学习模型的版权保护研究综述[J]. 网络与信息安全学报, 2022, 8(2): 1-14.
- WANG X Y, HUA G, JIANG H, et al. Survey on intellectual property protection for deep learning model[J]. Chinese Journal of Network and Information Security, 2022, 8(2): 1-14.
- [4] NAKAMURA T, KATAYAMA A, YAMAMURO M, et al. Fast watermark detection scheme for camera-equipped cellular phone[C]//Proceedings of the 3rd International Conference on Mobile and Ubiquitous Multimedia. 2004: 101-108.
- [5] PRAMILA A, KESKINARKAUS A, SEPPÄNEN T. Increasing the capturing angle in print-cam robust watermarking[J]. Journal of Systems and Software, 2018, 135: 205-215.
- [6] CHOU C H, LI Y C. A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile[J]. IEEE Transactions on Circuits and Systems for Video Technology, 1995, 5(6): 467-476.
- [7] FANG H, ZHANG W, ZHOU H, et al. Screen-shooting resilient watermarking[J]. IEEE Transactions on Information Forensics and Security, 2018, 14(6): 1403-1418.
- [8] FANG H, ZHANG W, MA Z, et al. A camera shooting resilient watermarking scheme for underpainting documents[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 30(11): 4075-4089.
- [9] CUI H, BIAN H, ZHANG W, et al. Unseencode: Invisible on-screen barcode with image-based extraction[C]//IEEE INFOCOM 2019-IEEE Conference on Computer Communications. 2019: 1315-1323.
- [10] BALUJA S. Hiding images in plain sight: deep steganography[C]//Advances in Neural Information Processing Systems, 2017, 30.
- [11] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, (1).
- [12] TANG W, TAN S, LI B, et al. Automatic steganographic distortion learning using a generative adversarial network[J]. IEEE Signal Processing Letters, 2017, 24(10): 1547-1551.
- [13] HU D H, WANG L, JIANG W J, et al. A novel image steganography method via deep convolutional generative adversarial networks[J]. IEEE Access, 2018, 6.
- [14] VOLKHONSKIY D, NAZAROV I, BURNAEV E. Steganographic generative adversarial networks[C]//Twelfth International Conference on Machine Vision (ICMV 2019). 11433: 991-1005.
- [15] ZHU J, KAPLAN R, JOHNSON J, et al. Hidden: Hiding data with deep networks[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 657-672.
- [16] LIU Y, GUO M, ZHANG J, et al. A novel two-stage separable deep learning framework for practical blind watermarking[C]//The 27th ACM International Conference. 2019.
- [17] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation[C]//18th International Conference on Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015. 2015: 234-241.
- [18] TANCIK M, MILDENHALL B, NG R. Stegastamp: invisible hyperlinks in physical photographs[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 2117-2126.
- [19] JADERBERG M, SIMONYAN K, ZISSERMAN A. Spatial transformer networks[C]//Advances in Neural Information Processing Systems. 2015: 28.
- [20] WENGROWSKI, ERIC, KRISTIN DANA. Light field messaging with deep photographic steganography[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 1515-1524.
- [21] FANG HAN, DONGDONG CHEN, QIDONG HUANG, et al. Deep template-based watermarking[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 31(4): 1436-1451.
- [22] BORACCHI G, FOI A. Modeling the performance of image restoration from motion blur[J]. IEEE Transactions on Image Processing, 2012, 21(8): 3502-3517.
- [23] KUPYN O, BUDZAN V, MYKHAILYCH M, et al. Deblurgan: blind motion deblurring using conditional adversarial networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 8183-8192.
- [24] KIM W, LEE S H, SEO Y. Image fingerprinting scheme for print-and-capture model[C]//Advances in Multimedia Information Processing-PCM 2006. 2006: 106-113.
- [25] NUUTINEN M, OITTINEN P. Simulation of printing and image capture for linking applications[J]. Graphic Arts in Finland, 2009, 38(1): 1-13.
- [26] SHIN R, SONG D. Jpeg-resistant adversarial images[C]//NIPS 2017 Workshop on Machine Learning and Computer Security. 2017: 1-8.
- [27] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein generative adversarial networks[C]//International Conference on Machine Learning (PMLR). 2017: 214-223.
- [28] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft coco: common objects in context[C]//13th European Conference on Computer Vision-ECCV 2014. 2014: 740-755.
- [29] SUN Y, YU Y, WANG W. Moiré photo restoration using multiresolution convolutional neural networks[J]. IEEE Transactions on Image Processing, 2018, 27(8): 4160-4172.

- [30] KANG X, HUANG J, ZENG W. Efficient general print-scanning resilient data hiding based on uniform log-polar mapping[J]. IEEE Transactions on Information Forensics and Security, 2010, 5(1): 1-12.
- [31] PRAMILA A, KESKINARKAUS A, SEPPÄNEN T. Toward an interactive poster using digital watermarking and a mobile phone camera[J]. Signal, Image and Video Processing, 2012(6): 211-222.



卓训方(1973-), 男, 上海人, 博士, 上海数据交易所有限公司研究员, 主要研究方向为数据流通与安全。

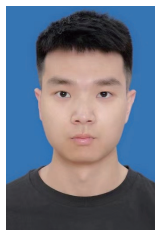
[作者简介]



李晓萌(1997-), 女, 安徽淮北人, 上海理工大学硕士生, 主要研究方向为深度学习和信息隐藏。



姚恒(1982-), 男, 安徽芜湖人, 博士, 上海理工大学副教授, 主要研究方向为多媒体信息安全和模式识别。



郭玳豆(1993-), 男, 河南安阳人, 上海理工大学博士生, 主要研究方向为深度学习和数字水印。



秦川(1980-), 男, 安徽芜湖人, 博士, 上海理工大学教授、博士生导师, 主要研究方向为多媒体信息安全和AI安全。