

基于自同步机制的抗屏摄水印方法

姚意智^{1),2)} 廖鑫^{1),2)} 欧博^{1),2)} 胡峰松¹⁾ 秦拯¹⁾

¹⁾(湖南大学网络空间安全学院 长沙 410082)

²⁾(岳麓山工业创新中心 长沙 410082)

摘要 屏幕拍摄为信息获取提供了便捷途径,但也使得未经授权盗摄行为日益严重,而抗屏摄水印技术能针对这一现象,有效对盗摄图像进行溯源追踪。然而,在屏摄过程中所产生的异步噪声,导致屏摄图像难以实现与原水印图像的像素级同步。虽然现有方法通过选取水印区域四角进行透视矫正以实现水印同步,但手动操作耗时费力,且解码准确率对四角定位精度依赖性强。针对上述问题,本文提出一种基于自同步机制的抗屏摄水印方法,以实现无需透视矫正的水印解码。首先,为了深入模拟屏摄失真,我们设计了双态屏摄噪声层,通过建模异步噪声并结合像素噪声,模拟屏摄场景下的失真特性;为规避透视矫正中的误差传播,先通过理论推导验证自同步机制,再构建预测同步掩码的同步网络,为解码器提供同步信息,从而实现无透视矫正的水印解码。此外,为提升水印图像的视觉质量,在多阶段、多损失的自同步屏摄水印训练框架中加入能量分布相似性损失约束,并在编码器中引入卷积块注意力机制。实验结果表明本文方案实现了水印自同步,从而消除解码对定位过程的强依赖性,在各种屏摄场景下的水印提取鲁棒性优于现有方法,且水印图像的视觉质量也有显著提高。

关键词 数字水印;版权保护;屏幕拍摄;水印同步;自同步机制

中图分类号 TP309 **DOI号** 10.11897/SP.J.1016.2026.00934

Self-Synchronization Mechanism Based Screen-Shooting Resilient Watermarking

YAO Yi-Zhi^{1),2)} LIAO Xin^{1),2)} OU Bo^{1),2)} HU Feng-Song¹⁾ QIN Zheng¹⁾

¹⁾(College of Cyber Science and Technology, Hunan University, Changsha 410082)

²⁾(Yuelushan Center for Industrial Innovation, Changsha 410082)

Abstract Screen shooting provides a convenient means for information acquisition, but it also leads to the increasing prevalence of unauthorized screen-shooting activities. Screen-shooting resilient watermarking technology can effectively trace the source of pirated images captured via screens to protect digital copyrights and intellectual property. However, the asynchronous noise—stemming from perspective distortion and device shaking—makes it extremely difficult to achieve pixel-level synchronization between the screen-shooting image and the original watermarked image. Although existing methods achieve watermark synchronization through perspective correction by selecting the four corners of the watermark region, manual operation is time-consuming and labor-intensive, and decoding accuracy is highly dependent on the positioning precision of the four corners. Crucially, any localization error propagates globally during

收稿日期:2025-08-13;在线发布日期:2026-01-08。本课题得到国家自然科学基金区域创新发展联合基金重点项目(U22A2030)、国家重点研发计划项目(2024YFF0618800,2022YFB3103500)、湖南省自然科学基金杰出青年基金项目(2024JJ2025)、湖南省重点研发计划项目(2024AQ2027,2025AQ2022)、湖南省重大科技攻关项目(2025QK2008)、岳麓山工业创新中心项目(2025YCH0222)资助。姚意智,硕士研究生,主要研究领域为抗屏摄水印和多媒体信息安全。E-mail: yyz@hnu.edu.cn。廖鑫(通信作者),教授,博士生导师,中国计算机学会(CCF)高级会员,主要研究领域为信息内容安全和人工智能安全。E-mail: xinliao@hnu.edu.cn。欧博,副教授,博士生导师,中国计算机学会(CCF)会员,主要研究领域为信息安全和计算机视觉。胡峰松,副教授,硕士生导师,主要研究领域为多媒体交互技术和3D可视化技术。秦拯,教授,博士生导师,中国计算机学会(CCF)会员,主要研究领域为网络信息安全和隐私保护。

correction, following a “large near, small far” pattern that severely compromises the reliability of watermark extraction. To tackle the above problems, this paper proposes a screen-shooting resilient watermarking method based on a self-synchronization mechanism, enabling watermark decoding without the need for perspective correction. This approach decouples the strong dependency between decoding and the positioning process to achieve high robustness. First, we design a dual-state screen-shooting noise layer to simulate distortions, which models the distortion characteristics by simulating asynchronous noise and combining it with pixel noise. The asynchronous noise simulation includes four critical steps—scaling, translation, perspective distortion, and background filling—to precisely reflect the variations in camera distance, position, and angle. The pixel noise encompasses variations in brightness, contrast, and saturation, alongside Gaussian noise, motion blur, and specialized lighting and moire noise. Second, the self-synchronization mechanism is verified through theoretical derivation, proving that correction errors inevitably diffuse across the entire image. We construct a synchronization network utilizing the HRNet architecture to maintain high-resolution features. This network is trained to predict synchronization masks, which directly identify the watermark region to provide vital synchronization information for the decoder. Additionally, an energy distribution similarity (EDS) loss constraint is integrated into the multi-stage, multi-loss training framework. The EDS loss ensures that the watermark residual matches the original image’s energy distribution, effectively suppressing green-spot artifacts in smooth regions. A convolutional block attention mechanism (CBAM) is also introduced into the encoder to focus embedding on areas that are human-imperceptible and noise-resistant. Detailed ablation studies confirm the indispensability of each module, showing that the removal of CBAM notably leads to a decrease in extraction accuracy while maintaining similar visual metrics. Furthermore, it is verified that while the EDS loss is essential for preserving subjective visual quality, the self-synchronization mechanism serves as the fundamental guarantee for the model to effectively resist asynchronous noise and ensure stable convergence. Experimental results show that the proposed method significantly outperforms existing methods like StegaStamp, PIMoG and SSDS in extraction robustness under various scenarios. The method achieves an average extraction accuracy (ACC) of over 96.4% across diverse distances and angles. Notably, the system is highly efficient, with a total synchronization and decoding time of only 0.022 seconds, meeting real-time requirements while achieving a substantial lead in visual quality with a peak signal-to-noise ratio (PSNR) of 38.89 dB and a structure similarity index measure (SSIM) of 0.989.

Keywords digital watermark; copyright protection; screen shooting; watermark synchronization; self synchronization mechanism

1 引言

随着移动端摄像功能的普及,屏幕拍摄已成为信息快速获取与传播的重要方式,在学术分享、商务会议和个人汇报等场景中得到广泛应用。这种便捷的信息获取手段,凭借其高效、低成本的优势,广泛应用于日常生活与工作的各个环节,但也衍生出大量的图像偷拍盗摄行为。未经授权的屏摄行为使得版权

保护、内容溯源等需求面临严峻挑战,而水印技术作为一种在图像中嵌入版权信息以实现版权标识与追踪的手段,为解决该问题提供了潜在解决方案。

现有的抗屏摄图像水印技术旨在让水印图像在经过图1(a)的过程后仍能准确提取水印,而水印的抗屏摄鲁棒性关键在于设计一个贴合实际屏摄失真的噪声层。Tancik与Jia等人^[1-2]通过数学建模屏摄失真环境,可以显著增强水印信息对屏摄失真的鲁棒性。Fang等人^[3]提出了PIMoG的噪声层,只模拟对水印

图像影响最大的三个部分(透视畸变、光照失真以及摩尔纹失真)就可以让水印对屏摄过程具有良好的鲁棒性。基于此,Li等人^[4]更进一步对屏幕特征建模,加入了针对屏幕的灰度偏移噪声,进一步提升了水印信息在屏摄场景下的提取准确率。除了数字模拟的方式以外,一些研究利用深度神经网络的强大拟合能力来模拟屏幕拍摄的退化过程。Wengrowski等人^[5]开发了一种相机-显示器转换函数CDTF,使深度学习网络在CDTF上端到端训练水印嵌入和恢复,从

而具备抵抗屏摄失真的能力。Sun等人^[6]通过初级恢复与精细增强两阶段的渐进式处理,分别实现基础特征重建和细节特征优化。上述研究虽然建模了屏摄过程中水印图像与屏摄图像的像素级映射关系,但并未关注到图像屏摄过程产生的像素位置变化,即异步噪声。其源于拍摄角度偏移导致的透视畸变、设备抖动引发的局部像素偏移等,会破坏水印信号的空间连续性,导致像素级映射关系失效,进而严重影响后续水印解码的准确性。

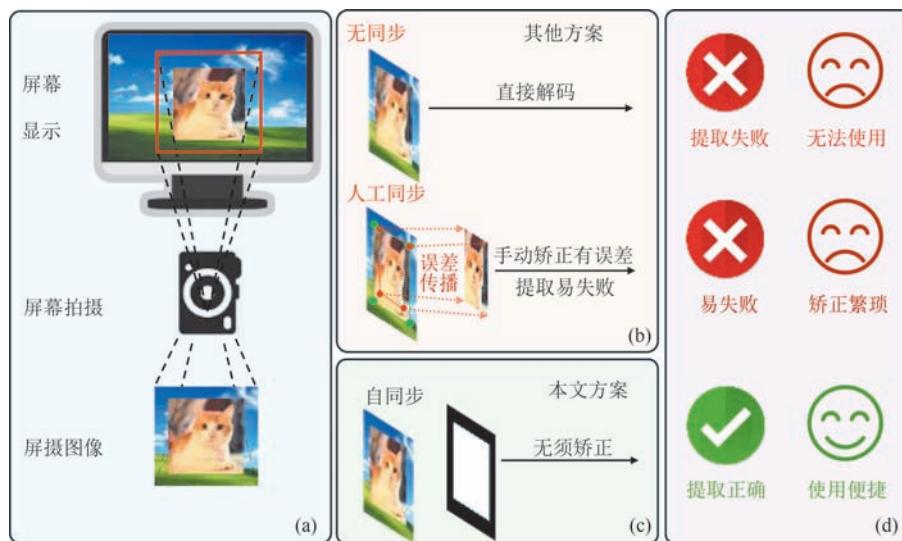


图1 本文方案与其他方案的区别

现有抵抗异步噪声的方法一般是通过手动对屏摄图像进行透视矫正,该过程如图1(b)所示。这种方案使用繁琐,费时费力。为了解决这一问题,研究人员提出了一些自动化的解决方案,比如添加可见的定位边框^[1-2,7]。其中,Tancik等人^[1]使用白色边框包裹水印图像,继而让预训练的BiSeNet来分割出水印区域,再采用凸四边形进一步拟合水印区域找到图像的四角坐标做透视矫正。但是,添加的可见边框严重影响了水印图像的外观,因此越来越多的研究人员开始探究无边框场景下对屏摄图像中的水印区域进行自动化定位。Zhao等人^[8]设计了一个双分支双头DBDH网络,用于精确定位图像中不可见的嵌入区域。Jia等人^[9-10]提取嵌入在图像中的不可见标记来定位嵌入子区域的四个顶点,从而纠正透视失真。Zhu等人^[11]使用Lite-HRNet和DUE网络作为定位网络和编码器,进一步提高了定位效率和视觉质量。然而,上述方法都依赖于透视矫正来进行同步,即都依赖于定位网络精确的四角定位。当四角坐标定位存在误差,其误差随透视矫正传播至

图像的每一个像素而导致解码失败。因此,解耦水印解码与同步的强依赖关系,设计一种无需矫正的抗屏摄水印方法成为了亟待解决的难题。

基于上述分析,本文提出了一种基于自同步机制的抗屏摄水印方法,借助像素-异步的双态屏摄噪声层,通过水印编解码器的端到端训练实现高屏摄鲁棒性。本文还使用数学推导的方式探寻透视矫正过程中的误差传播方式,得出误差传播“近大远小,全图扩散”的规律。为从根本上避免不准确的透视矫正所带来的误差传播,我们在异步噪声模拟过程中生成标签掩码用以监督训练同步网络,继而将同步网络预测的掩码作为水印同步信息输入至解码器,使其在无透视矫正下解码水印。此外,本文在编码器中加入卷积块注意力模块(Convolutional Block Attention Module,CBAM),并在多阶段、多损失训练中加入能量分布损失(Energy Distribution Similarity,EDS),旨在进一步提升水印图像的视觉质量。本文的主要贡献如下:

(1)本文提出了一种双态屏摄噪声层,将像素噪

声与异步噪声结合对屏摄过程进行了更为精确的模拟。其中,将异步噪声模拟划分为缩放、平移、透视畸变以及背景填充四个步骤,准确地模拟屏摄时的异步过程。

(2)本文提出了一种自同步机制,创新性地实现了无需透视矫正的水印解码方式。我们首先通过数学推导论证了矫正误差的传播特性,证明了自同步机制的必要性。在自同步机制下,解码器利用同步网络预测的同步掩码获取水印同步信息,实现无透视矫正的水印解码以规避误差传播。

(3)本文设计了一个多阶段、多损失的自同步抗屏摄水印训练框架。该框架受多阶段、多损失约束,通过逐步增加的损失函数复杂度与任务难度,实现高视觉质量与强鲁棒性的平衡。此外,我们在编码器中加入卷积块注意力机制、在损失约束中加入能量分布损失,进一步提升视觉质量。

(4)本文进行一系列实验来验证所述模块的有效性与训练框架的优越性。实验结果表明,我们的方法在视觉质量和抗屏摄鲁棒性上均优于现有方法。

2 相关工作

2.1 数字水印技术

传统的数字水印技术通常基于信号处理、信息论和密码学的原理,通过手工设计的算法或规则实现水印嵌入和提取。数字水印最早由 Ron van Schyndel 在 1994 年定义^[12],随后被广泛应用于图像^[13-14]、音频^[15-16]和视频^[17]。数字图像水印主要分为空域和变换域方法,空域方法通过对图像像素的转换或修改来将水印信号嵌入到数字图像中^[18-19],而变换域方法则利用特定域变换方法将水印应用到变换域中^[20-24]。Bamatraf 等人^[18]利用最低有效位 LSB 进行嵌入,得到了高视觉质量的水印图像。Kumar 等人^[19]将最低有效位嵌入应用到彩色视频帧中,并增强了其对一些数字噪声攻击的鲁棒性。最低有效位嵌入的空域方法往往极易受到统计学分析,因此 Joseph 等人^[21]将离散余弦变换与离散小波变换结合应用到图像水印中,提升了图像的安全性。Bei 等人^[22]利用人类视觉系统 HVS 在频域与小波域中选择最佳的水印嵌入位置。Fares 等人^[23]在频域与小波域的中频区域进行水印嵌入,保护了医学图像的隐私安全。Peng 等人^[24]利用多方向差分的直方图揭示了不同图像在高频域的统计特征,大幅度地提升了下采样操作的检测准确率以及针对 JPEG 压缩的鲁棒

性。近年来,深度学习技术的飞速发展数字图像水印领域取得了突破性进展,出现了一系列基于深度神经网络 DNN 的水印模型。基于深度学习的水印技术,其典型架构通常包含编码器、噪声层与解码器三个核心组件。Zhu 等人^[25]率先将神经网络应用于数字图像水印,为该领域奠定了重要基础。在此基础上,Liu 等人^[26]提出了一种两阶段可分离深度学习框架 TSDL 专门用于盲水印任务。针对水印对攻击的适应性问题,Ahmadi 等人^[27]提出了基于深度网络的残差扩散水印框架 ReDMark。该框架通过将各类攻击模拟为可微分网络层,有效支持端到端训练过程,从而使水印能够更好地适配攻击特征。Chen 等人^[28]提出了用于模拟 JPEG 压缩的深度学习 JSNet,通过可微分形式对 JPEG 压缩过程进行建模,使其能够融入端到端训练流程。Jia 等人^[29]通过小批次的真实和模拟 JPEG 压缩增强了水印对 JPEG 压缩的鲁棒性,同时提出水印信息扩散块进一步提升编码器性能。Yin 等人^[30]构造无损载体以确定鲁棒嵌入域,有效维持了安全性和鲁棒性的平衡。此外,Luo 等人^[31]采用通道编码策略提升水印鲁棒性,但同时也导致嵌入容量有所降低。而 Fei 等人^[32]则聚焦于生成对抗网络的水印嵌入,将预训练的 CNN 水印解码模块集成到生成器输出端,并引入水印损失项,实现了生成图像的不可见水印嵌入。但上述数字水印技术均不能应对复杂的屏摄场景,从而无法准确提取水印信息。

2.2 抗屏摄深度水印技术

与传统的数字失真不同,屏摄场景下的噪声复杂多样,带来了新的挑战。为解决这一问题,Tancik 与 Jia 等人^[1-2]通过对屏幕拍摄过程失真进行数学建模,构建可微的噪声层提升水印对屏幕拍摄过程的鲁棒性。Fang 等人^[33]将传统水印方案的模板水印方法与深度学习的解码器相结合,首次提出了基于模板的深度水印算法。Fang 等人^[7]还总结了如屏摄、打印等跨物理介质的四个属性(透明度,效率,鲁棒性和适应性),明确了跨物理介质水印研究的目标方向。Gao^[34]设计了基于风格迁移的屏摄仿真网络,并基于此网络构建了一个全面的噪声层,以实现屏幕拍摄失真的鲁棒性。但是,其方法需要基于大量的训练数据且相较于数学模拟的方法更容易过拟合。Zhong 等人^[35]提出了无监督的训练方式,不需要任何人的干预和注释的先验信息训练抗屏摄水印。传统水印发展历程表明,在频域中嵌入水印有独特的优势。因此,一些研究通过将水印研究方向转移到频域中^[36-38]。Lu 等

人^[36]率先将频域引入抗屏摄水印中,以提高水印图像的视觉质量。Zheng等人^[37]在频域上引入了一种退化模型来表征摩尔纹对图像的影响。Fu等人^[38]提出了一个小波域级联网络增强水印图像视觉质量,并采用逆向恢复模块提升抗屏摄鲁棒性。然而,上述抗屏摄水印技术没有考虑屏摄过程的异步噪声,且均依赖于精准的手动定位方式。

3 本文方案

如图2所示,本文提出的基于自同步机制的抗

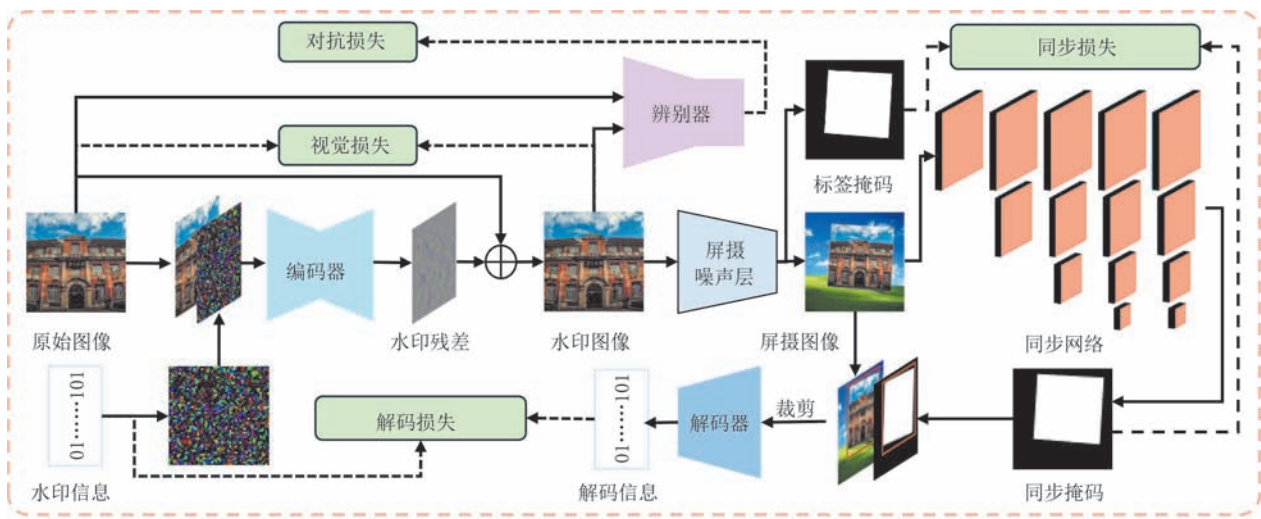


图2 基于自同步机制的抗屏摄水印方法

为提升编码器性能,我们对常用的水印编码网络 U-Net 进行了改进,在 U-Net 的跳跃连接中加入了卷积块注意力模块 CBAM,旨在将水印编码到人眼难以察觉且不易受到屏摄噪声攻击的区域。卷积块注意力模块 CBAM 包括通道注意力模块 CAM 和空间注意力模块 SAM^[39]。具体来说,通道注意力模块会自动选择重要的特征通道并抑制不相关的通道,而空间注意力模块通过关注关键区域而忽略不太重要的区域来增强空间特征。CBAM 的输入由高分辨率特征和通过上采样获得的低分辨率特征组成。输入特征图 F_{in} 首先采用 CAM 模块计算通道注意力图 M_c ,然后使用 SAM 模块计算空间注意力图 M_s ,最后得到输出特征图 F_{out} 。具体过程如下:

$$F_{out} = ((F_{in} \otimes M_c) \otimes M_s) \quad (2)$$

总之,在水印嵌入任务中,卷积块注意力模块让水印信号优先嵌入在既难以受屏摄噪声攻击影响又难以被人眼感知的位置,从而保证水印图像的鲁棒

性。屏摄水印方法包括5个部分:编码器、双态屏摄噪声层、同步网络、解码器与判别器。该方法使用多损失约束进行多阶段训练,将复杂任务逐步拆解,最终逐步实现视觉质量与鲁棒性的平衡。

3.1 编码器

编码器是深度水印方法中的重要组成部分,负责将水印信息不可见地编码至原始图像中,既保证水印图像的视觉质量,也保证水印提取的鲁棒性。编码器 E 将 m 位水印信息 w 编码到尺寸为 $H \times W$ 的原始图像 $I_{原始}$ 中,过程如下:

$$I_{水印} = I_{原始} + E(I_{原始}, w) \quad (1)$$

性和不可感知性。

3.2 双态屏摄噪声层

现有方法的噪声模拟主要聚焦在像素扰动,而对于异步噪声则仅限于关注透视畸变再矫正后的像素值损失。然而,这并不能反映实际屏摄情况,因为相机拍摄的过程中难以准确捕捉屏幕上显示的水印图像,水印图像往往只占据屏摄图像的一部分;此外,屏摄的水印图像四周也可能出现复杂的背景。因此,为了以更逼真的方式模拟屏幕拍摄过程,必须对此过程进行更精细的模拟。如图3所示,本文将屏摄噪声划分为两类,一类是像素噪声,具体表现为对应像素点的像素值变化,例如亮度/对比度扰动、饱和度扰动、高斯噪声、运动模糊以及光照噪声和摩尔纹噪声;另一类是异步噪声,具体表现为像素点位置变化,例如缩放、平移、透视畸变和背景填充。结合两类噪声模拟的双态屏摄噪声层可对整个屏摄过程进行精细建模,让编/解码器及同步网络在端到端训练中对真实屏摄过程更具鲁棒性。

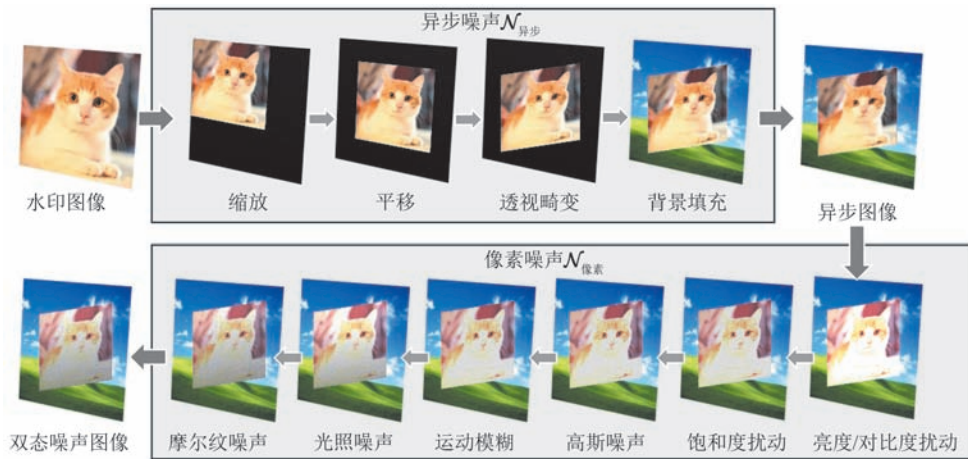


图3 双态屏摄噪声层

3.2.1 异步噪声模拟 $\mathcal{N}_{异步}$

异步噪声的划分主要基于屏摄过程中导致像素位置异步偏移的核心物理因素。实际屏摄时，相机与屏幕的相对距离变化会直接引发水印区域在图像中的尺寸缩放，因此缩放是模拟异步的第一步。相机拍摄位置的水平或垂直偏移会导致水印区域在画面中发生平移，需要平移步骤来复现这一现象。异步噪声破坏水印的关键在于透视畸变，其源于相机在拍摄时的角度倾斜。此外，屏摄图像中水印区域通常不会充满整个画面，周围必然存在无关的背景，背景填充正是为了模拟这种屏摄场景。这四个步骤依次对应了距离、位置、角度及画面占比这四个影响屏摄异步性的关键维度，共同实现对屏摄过程中像素位置变化的全面且精准的模拟。

在屏摄场景下，受屏摄角度与距离影响，屏摄过程将改变水印图像的像素位置，产生异步噪声。如图3所示，我们对异步过程进行了精确模拟，采用缩放、平移、透视畸变、背景填充模拟水印图像在屏摄场景下的异步噪声。对于一个长为 H 、宽为 W 的水印图像， (x, y) 位置的像素通过缩放因子 $S_{缩放}$ 可以计算缩放图像对应像素位置 $(x_{缩放}, y_{缩放})$ ，其中 $0 \leq x < W, 0 \leq y < H$ ，计算过程如下：

$$x_{缩放} = x \times S_{缩放} \quad (3)$$

$$y_{缩放} = y \times S_{缩放} \quad (4)$$

然后，对缩放后的图像进行随机像素距离的平移，但仍要保持平移后水印图像区域完整。假设水平平移 T_x 个像素距离、垂直平移 T_y 个像素距离，且 $0 \leq x_{缩放} + T_x < H, 0 \leq y_{缩放} + T_y < W$ 。平移后的像素位置计算如下：

$$x_{平移} = x_{缩放} + T_x \quad (5)$$

$$y_{平移} = y_{缩放} + T_y \quad (6)$$

紧接着对图像四角进行扰动模拟透视畸变产生的失真，我们假定缩放和平移后的图像四角坐标为 $(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)$ ，使用 $\Delta x_i, \Delta y_i$ ($i = 1, 2, 3, 4$) 代表四角坐标在 x 轴和 y 轴上的扰动像素距离。扰动后畸变图像 $I_{畸变}$ 的对应像素坐标为 $(x_i^{畸变}, y_i^{畸变})$ ，计算过程如下：

$$x_i^{畸变} = x_i \pm \Delta x_i \quad (7)$$

$$y_i^{畸变} = y_i \pm \Delta y_i \quad (8)$$

最后，我们设计了一个标签掩码可以将背景 $I_{背景}$ 与畸变图像 $I_{畸变}$ 融合得到最终的异步图像 $I_{异步}$ 。背景图像在训练集中随机选择，并拉伸为尺寸 $H \times W$ 。我们对与水印图像同尺寸的纯白图像做相同的异步操作，以记录异步图像中水印图像区域的分布情况。生成的二值标签掩码 $M_{标签}$ 中白色区域表示异步图像的水印图像区域，黑色区域则表示异步图像的背景区域。基于标签掩码 $M_{标签}$ ，整个融合过程如下：

$$I_{异步} = I_{畸变} \times M_{标签} + I_{背景} \times (1 - M_{标签}) \quad (9)$$

3.2.2 像素噪声模拟 $\mathcal{N}_{像素}$

像素噪声模拟则通过数学方式模拟水印图像中像素点在屏摄前后的数值变化。首先，我们调整不同的亮度/对比度和饱和度来反映屏幕显示的色差与不同光照条件下的屏摄失真，然后添加了高斯噪声模拟传感器噪声。在此基础上，我们采用线性模糊操作模拟相机移动过程中造成的模糊，以上像素噪声模拟基本沿用 StegaStamp^[1] 的噪声层。最后针对屏摄场景下的光照与摩尔纹噪声，我们沿用了 PIMOG^[3] 的模拟方法。

我们的双态屏摄噪声层设计为先添加异步噪声再添加像素噪声的方式。在按下拍摄键前，受相机

与屏幕中水印图像的距离、角度影响,异步噪声将发生在像素噪声产生前。当按下拍摄按键,已经遭受异步噪声攻击的图像再由于光照条件、屏幕-相机硬件设备影响产生像素噪声。最终基于对实际屏幕拍摄过程的拆解,将异步噪声放在像素噪声前,使得噪声添加顺序更真实,具体如下:

$$I_{\text{屏幕}} = \mathcal{N}_{\text{像素}}(\mathcal{N}_{\text{异步}}(I_{\text{水印}})) \quad (10)$$

3.3 自同步机制

本章围绕自同步机制,先通过3.3.1的数学推导,分析透视矫正的误差传播规律,指出其“近大远小,全图扩散”的普遍规律,以表明自同步机制的必要性;3.3.2则基于此提出自同步解码方式,由同步网络预测同步掩码,解码器再借助该掩码提供的同步信息实现无透视矫正解码,从而规避误差传播。

3.3.1 透视矫正的误差传播

在现有抗屏摄水印方案中,往往通过手动或自动方式定位水印图像四角进行透视矫正。现有方法无论是手动还是自动,对四角的定位往往存在误差,在透视矫正过程中误差会传播到每一个像素,对水印造成破坏。

设屏摄图像中水印区域的四角齐次坐标为 $p_i = (u_i, v_i, 1)^\top$,对应目标点为 $q_i = (x_i, y_i, 1)^\top$,其中 $i = 1, 2, 3, 4$,且 p_i, q_i 的三维分量为1,表示其处于同一尺度下。当四角定位存在误差时,实际使用的源坐标为 $p'_i = p_i + \Delta p_i$,其中 $\Delta p_i = (\Delta u_i, \Delta v_i, \Delta w_i)^\top$ 为误差向量。假设源坐标的误差为像素级的微小偏移,只影响二维位置而不影响尺度,即误差向量三维分类 $\Delta w_i \approx 0$ 。对于无误差的透视矫正而言,存在单应性矩阵 \mathbf{H}_0 满足 $q_i = \mathbf{H}_0 p_i$ 。通过水印区域四角的齐次坐标矩阵 $\mathbf{P} = [p_1, p_2, p_3, p_4]$ 和对应目标点坐标矩阵 $\mathbf{Q} = [q_1, q_2, q_3, q_4]$,可以计算 $\mathbf{H}_0 = \mathbf{Q}\mathbf{P}^{-1}$ 。含误差的单应性矩阵 $\mathbf{H}' = \mathbf{H}(p_i + \Delta p_i)$,用泰勒公式一阶展开可近似为 $\mathbf{H}' \approx \mathbf{H}(p_i) + \Delta \mathbf{H} = \mathbf{H}_0 + \Delta \mathbf{H}$,其中 $\Delta \mathbf{H} = \sum_{i=1}^4 \frac{\partial \mathbf{H}}{\partial p_i} \Delta p_i$ 是一阶误差项。因 \mathbf{H}' 是给定误差下求解的最优结果,即在存在输入误差情况下,保持源平面到目标平面的近似关系 $q_i \approx \mathbf{H}' p'_i$,代入 p'_i, \mathbf{H}' 则有

$$q_i \approx (\mathbf{H}_0 + \Delta \mathbf{H})(p_i + \Delta p_i) \approx \mathbf{H}_0 p_i + \mathbf{H}_0 \Delta p_i + \Delta \mathbf{H} p_i + \Delta \mathbf{H} \Delta p_i \quad (11)$$

由于 $\mathbf{H}_0 p_i = q_i$ (真实关系),两边减去 q_i ,再忽略

$\Delta \mathbf{H} \Delta p_i$ 这个二阶小量(在泰勒展开的线性近似中,高阶小量可忽略不计),可得

$$\Delta \mathbf{H} p_i \approx -\mathbf{H}_0 \Delta p_i \quad (12)$$

通过带误差的单应性矩阵 \mathbf{H}' 计算的目标点与真实目标点的差值可计算 Δq_i :

$$\begin{aligned} \Delta q_i &= \mathbf{H}' p_i - q_i \\ &= (\mathbf{H}_0 + \Delta \mathbf{H}) p_i - \mathbf{H}_0 p_i \\ &= \Delta \mathbf{H} p_i \end{aligned} \quad (13)$$

最终联合公式(12)与公式(13)可得

$$\Delta q_i \approx -\mathbf{H}_0 \Delta p_i \quad (14)$$

将单应性矩阵 $\mathbf{H}_0 = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{pmatrix}$ (将第三行

归一化)代入 $q_i = \mathbf{H}_0 p_i$,则对应目标点 $q_i = (x_i, y_i, 1)^\top$ 的变换为

$$\begin{aligned} \mathbf{H}_0 p_i &= \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{pmatrix} \begin{pmatrix} u_i \\ v_i \\ 1 \end{pmatrix} = \\ &= \begin{pmatrix} h_{11}u_i + h_{12}v_i + h_{13} \\ h_{21}u_i + h_{22}v_i + h_{23} \\ h_{31}u_i + h_{32}v_i + 1 \end{pmatrix} = \begin{pmatrix} X_i \\ Y_i \\ W_i \end{pmatrix} \end{aligned} \quad (15)$$

因此我们得出

$$\begin{cases} X_i = h_{11}u_i + h_{12}v_i + h_{13} \\ Y_i = h_{21}u_i + h_{22}v_i + h_{23} \\ W_i = h_{31}u_i + h_{32}v_i + 1 \end{cases} \quad (16)$$

计算的非齐次坐标需除以第三分量 W_i ,可得目标点 q_i 的二维坐标系坐标 $(x_i, y_i) = \left(\frac{X_i}{W_i}, \frac{Y_i}{W_i}\right)$ 。再将 \mathbf{H}_0 代入公式(14)计算 $\Delta q_i = (\Delta x_i, \Delta y_i, 0)^\top$,则 $\Delta x_i, \Delta y_i$ 具体表示为

$$\begin{cases} \Delta x_i = -\frac{h_{11}\Delta u_i + h_{12}\Delta v_i}{h_{31}u_i + h_{32}v_i + 1} = -\frac{h_{11}\Delta u_i + h_{12}\Delta v_i}{W_i} \\ \Delta y_i = -\frac{h_{21}\Delta u_i + h_{22}\Delta v_i}{h_{31}u_i + h_{32}v_i + 1} = -\frac{h_{21}\Delta u_i + h_{22}\Delta v_i}{W_i} \end{cases} \quad (17)$$

公式(16)表示,目标点坐标受透视变换的影响与尺度 W_i 相关。在齐次坐标转换为非齐次坐标时, w 表示为齐次坐标系 (x, y, w) 与对应二维坐标系 $\left(\frac{x}{w}, \frac{y}{w}\right)$ 的缩放关系。对应到透视变换中, $\frac{1}{w}$ 表示为点到透视中心(相机)的相对深度,当 $\frac{1}{w}$ 较大时,表示点靠近透视中心,坐标变换后尺度被放大。这个规律符合相机透视变换中“近大远小”的特性。由

公式(17)可知,分母 W_i 决定了误差传播中的放大倍数,当 W_i 较小时,相同的源误差 $\Delta u_i, \Delta v_i$ 会导致目标点误差 $\Delta x_i, \Delta y_i$ 显著增大。因此,“近大远小”的特性同样适用于解释误差传播的普遍规律:在带误差的单应性矩阵 H' 作用下,靠近相机的像素点因 W_i 大而目标点误差小,远离相机的像素点因 W_i 小而目标点误差大。此外,当透视矫正误差存在时,矫正图像中的所有像素点都会因 H' 作用下产生误差,即误差不可避免地“全图扩散”。

3.3.2 基于自同步机制的解码

通过剖析透视变换矫正过程的误差传播机制,我们发现矫正误差会引发显著的误差传播,矫正后再解码的方式存在进一步破坏水印信息的可能。因此,本文提出了基于自同步机制的解码方式,由水印同步网络预测同步掩码,为解码器提供同步信息帮助其解码水印。鉴于 HRNet 在语义分割任务中的突出表现^[40-41],我们选择其作为本文的同步网络。该网络通过并行分支结构维持高分辨率特征图的传递,避免了传统网络因多次下采样导致的细节丢失。屏摄过程中的异步噪声常造成水印区域的边缘模糊,而 HRNet 的高分辨率分支可精准追踪细微的空间信息,做到边缘的精确预测。同时,我们将同步问题视为像素值的二分类问题,由于编码水印过程引入了水印图像和背景图像之间的特征差异,低分辨率分支提取的语义信息能精确将屏摄图像中的每个像素划分为水印图像像素和背景图像像素。最终,借助二值标签掩码 $M_{\text{标签}}$ 进行端到端训练,同步网络可以学习到水印图像与背景图像的差异,从而预测同步掩码 $M_{\text{预测}}$ 。

在解码前,我们对屏摄图像和同步掩码进行相同的裁剪拉伸,尽可能削减背景填充的影响。首先,对同步掩码进行连通区域分析,将同步掩码分解为多个独立区域。连通区域分析过程中,采用了高效且易于实现的 Two-Pass 算法,通过两次扫描同步掩码:先为前景像素分配临时标记,并记录等价关系;再统一相同连通区域的标记,从而完成所有连通区域的识别。最终得到多个独立区域 L :

$$L = \text{Two-Pass}(M_{\text{预测}}) \quad (18)$$

其中 $L \in \{0, 1, \dots, K\}$ 是标记矩阵, K 是连通区域数量, $L=0$ 表示背景。然后计算每个区域的面积并选择最大区域,这一步是排除同步掩码中的小连通区域影响:

$$\begin{aligned} \text{面积}_k &= \sum_{i,j} (L(i,j) = k) \\ k^* &= \arg \max_{k=1}^K \text{面积}_k \end{aligned} \quad (19)$$

再根据最大连通区域 k^* 计算裁剪边界:

$$\begin{aligned} y_1 &= \min\{i: L(i,j) = k^*\}, \\ y_2 &= \max\{i: L(i,j) = k^*\}, \\ x_1 &= \min\{j: L(i,j) = k^*\}, \\ x_2 &= \max\{j: L(i,j) = k^*\} \end{aligned} \quad (20)$$

最后执行裁剪并拉伸成固定的尺寸 $H \times W$, 过程如下:

$$\begin{aligned} I_{\text{裁剪}\&\text{拉伸}} &= \text{拉伸}_{H \times W}(I_{\text{屏摄}}[y_1:y_2, x_1:x_2]) \\ M_{\text{裁剪}\&\text{拉伸}} &= \text{拉伸}_{H \times W}(M_{\text{预测}}[y_1:y_2, x_1:x_2]) \end{aligned} \quad (21)$$

我们将裁剪和拉伸后的屏摄图像与掩码拼接成四通道张量作为解码器的输入,如图4所示。解码器由多个 ConvBNReLU (卷积 Conv、批归一化 Batch Normalization 和激活函数 ReLU)、SE 块 (Squeeze and Excitation Block) 与全连接层组成,输出解码水印信息。因此,解码器可以获得输入图像中水印区域的先验知识,从而可以正确解码水印 w' 。其次,解码器的输入为未经矫正处理的图像,从而避免了上述不精确矫正产生的误差传播。同步网络 Sync 与解码器 D 的自同步解码过程如下:

$$M_{\text{预测}} = \text{Sync}(I_{\text{屏摄}}) \quad (22)$$

$$(I_{\text{裁剪}\&\text{拉伸}}, M_{\text{裁剪}\&\text{拉伸}}) = \text{裁剪}\&\text{拉伸}(I_{\text{屏摄}}, M_{\text{预测}}) \quad (23)$$

$$w' = D(I_{\text{裁剪}\&\text{拉伸}}, M_{\text{裁剪}\&\text{拉伸}}) \quad (24)$$

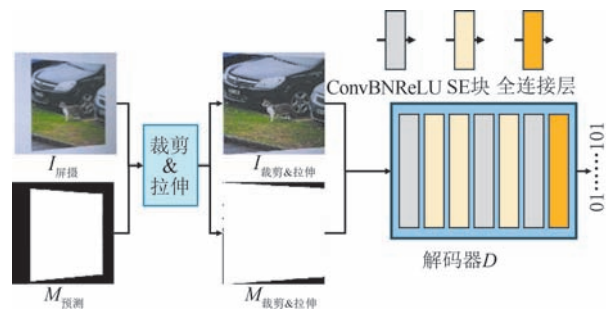


图4 自同步解码过程

3.4 多损失约束的多阶段训练框架

为实现鲁棒性、视觉质量与水印容量的平衡,有必要对复杂问题进行分解,采用多损失约束的多阶段训练框架。为此,我们设计四个训练阶段,各阶段将逐步提升损失函数的复杂度,并融入难度更高的任务。

第一阶段的训练目标在于确保在无噪声情况下正确的编/解码水印信息。在这个阶段,我们仅仅更新编码器与解码器的参数,并使用二分类交叉熵(Binary Cross Entropy, BCE)来计算信息损失 $L_{\text{信息}}$:

$$L_{\text{一阶段}} = L_{\text{信息}} = \text{BCE}(\omega, \omega') \quad (25)$$

在第二阶段中添加了3.2节中设计的双态屏摄噪声层并开始更新同步网络的参数,目的是确保水印信息的正确同步与解码。这一阶段中,我们加入了由二分类交叉熵损失与交并比(Intersection Over Union, IoU)损失构成的同步损失 $L_{\text{同步}}$ 。正如上一节所述,我们将水印同步问题视为像素二分类问题,因此使用二分类交叉熵来优化同步网络。此外,加入的交并比损失也有助于模型收敛,进一步提升同步网络的精度。显而易见的是,良好的同步是正确解码的前提,因此我们使用 μ_1, μ_2 来控制同步损失与信息损失的权重,并给予同步损失更大的权重。综上,第二阶段的损失函数 $L_{\text{二阶段}}$ 如下:

$$L_{\text{同步}} = \text{BCE}(M_{\text{预测}}, M_{\text{标签}}) + (1 - \text{IoU}(M_{\text{预测}}, M_{\text{标签}})) \quad (26)$$

$$L_{\text{二阶段}} = \mu_1 L_{\text{信息}} + \mu_2 L_{\text{同步}} \quad (27)$$

第三阶段目标在正确提取水印的基础上逐步恢复视觉质量。我们设置了一个混合的视觉损失 $L_{\text{视觉}}$,包括均方误差(mean square error, MSE)、感知损失LPIPS^[42]、结构相似性(Structure Similarity Index Measure, SSIM)损失、能量分布相似性EDS损失。为了达到水印信号在水印图像中肉眼不可见的视觉效果,我们设计让编码器生成的水印残差与原始图像具有相似的能量分布。若水印残差的能量分布与原始图像差异显著,则叠加后会打破原始图像的能量平衡,导致水印图像的能量分布异常,从而易被人眼识别。因此,我们加入能量分布相似性损失 L_{EDS} 作为视觉损失的一部分。

对于给定的水印残差 $I_{\text{残差}}$ 与原始图像 $I_{\text{原始}}$,我们对输入张量进行平方操作将像素值转换为能量值,再使用全1的卷积核(窗口大小设为 $W_{\text{窗口}}$)计算局部能量总和,最终返回局部能量分布图,具体如下:

$$E(I)_{i,j} = \sum_{m=1}^{W_{\text{窗口}}} \sum_{n=1}^{W_{\text{窗口}}} [I(i+m, j+n)]^2 \quad (28)$$

其中, $W_{\text{窗口}}$ 是窗口大小, $E(I)_{i,j}$ 表示位置 (i, j) 处的局部能量。计算出能量分布图 $E(I)$ 再展平为向量 $e(I)$ 计算余弦相似度。能量相似度损失 L_{EDS} 是基于 $I_{\text{残差}}$ 与 $I_{\text{原始}}$ 能量分布图之间的余弦相似度计算的,具

体过程如下:

$$L_{\text{EDS}} = \text{余弦相似度}(e(I_{\text{残差}}), e(I_{\text{原始}})) = \frac{e(I_{\text{残差}}) \cdot e(I_{\text{原始}})}{\|e(I_{\text{残差}})\| \cdot \|e(I_{\text{原始}})\|} \quad (29)$$

最终,我们按 $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ 对各视觉损失进行加权相加,得到的 $L_{\text{视觉}}$ 可表示为

$$L_{\text{视觉}} = \lambda_1 \text{MSE}(I_{\text{原始}}, I_{\text{水印}}) + \lambda_2 \text{LPIPS}(I_{\text{原始}}, I_{\text{水印}}) + \lambda_3 (1 - \text{SSIM}(I_{\text{原始}}, I_{\text{水印}})) + \lambda_4 (1 - L_{\text{EDS}}(I_{\text{原始}}, I_{\text{残差}})) \quad (30)$$

设视觉损失的权重为 μ_3 ,则第三阶段的总损失 $L_{\text{三阶段}}$ 计算如下:

$$L_{\text{三阶段}} = \mu_1 L_{\text{信息}} + \mu_2 L_{\text{同步}} + \mu_3 L_{\text{视觉}} \quad (31)$$

第四阶段加入对抗损失,通过对抗损失 $L_{\text{对抗}}$ 进一步减少可见伪影提高视觉质量。此阶段我们将判别器加入到训练过程中,利用判别损失 $L_{\text{判别}}$ 优化判别器判断是否为含水印的图像,而生成器则在 $L_{\text{判别}}$ 优化下尽可能欺骗判别器从而生成更高质量的水印图像。因此,设对抗损失权重为 μ_4 ,第四阶段的总损失 $L_{\text{四阶段}}$ 定义如下:

$$L_{\text{对抗}} = \text{BCE}(I_{\text{水印}}, 0) \quad (32)$$

$$L_{\text{判别}} = \text{BCE}(I_{\text{原始}}, 0) + \text{BCE}(I_{\text{水印}}, 1) \quad (33)$$

$$L_{\text{四阶段}} = \mu_1 L_{\text{信息}} + \mu_2 L_{\text{同步}} + \mu_3 L_{\text{视觉}} + \mu_4 L_{\text{对抗}} \quad (34)$$

4 实验与分析

在本节中,我们将首先介绍实验设置,然后对本文方案进行视觉效果与抗屏摄鲁棒性对比分析,证明本文所述方案在视觉效果与抗屏摄鲁棒性上的优越性,最后通过消融实验证明本文改进模块的作用。

4.1 实验设置

实验采用COCO2017数据集作为训练样本,包含118 287张图像。实验设定图像尺寸 $H, W=400$,随机生成的水印信息长度 $m=100$ 。测试集由ADE20K、COCO2017、lvis2017、Open Image以及VOC2012五个公开数据集各随机抽取100张样本组成,覆盖多种场景与图片类型。训练集与测试集图像均统一缩放到 400×400 的分辨率。屏摄实验中,所使用的显示设备为“TL160ADMP03显示屏”和“ENVISION G249G显示屏”,拍摄设备为“小米15智能手机”和“苹果13智能手机”。其中,“TL160

ADMP03显示屏”与“ENVISION G249G显示屏”，分辨率分别为 2560×1600 和 1920×1080 ，二者均统一采用sRGB色域显示测试内容。此外，“小米15智能手机”和“苹果13智能手机”均采用主摄像头进行拍摄实验，主摄的像素大小为50 MP和12 MP。

训练过程基于英伟达RTX5090进行，总计完成350 k次迭代。其中第一阶段、第二阶段、第三阶段和第四阶段的迭代次数分别为5 k、15 k、280 k、50 k。这种分配是因各阶段任务难度递增：第一阶段仅需实现无噪声下基础编解码，5 k迭代足以完成；第二阶段引入噪声层和同步网络训练，15 k迭代可达成初步协同；第三阶段任务复杂，需优化多维度视觉损失以平衡鲁棒性与视觉质量，故分配280 k迭代使网络充分收敛；第四阶段通过对抗损失微调视觉效果，通过50 k迭代实现精细优化。针对编码器、同步网络和解码器的参数优化，采用Adam优化器及其默认超参数；对于判别器，则采用RMSprop优化算法。为获取最优性能，通过网络搜索将超参数设置为： $\mu_1 = 15$ ， $\mu_2 = 50$ ， $\mu_3 = 8$ ， $\mu_4 = 5$ ， $\lambda_1 = 1$ ， $\lambda_2 = 2$ ， $\lambda_3 = 2$ ， $\lambda_4 = 1$ ， $W_{窗口} = 11$ 。双态屏摄噪声层中，异步噪声强度在特定范围内随机变化，其强度范围设置如下： $S_{缩放} \in [0.5, 1]$ ， $T_x \in [0, W \times (1 - S_{缩放})]$ ， $T_y \in [0, H \times (1 - S_{缩放})]$ ， $\Delta x_i \in \{0.1H\}$ ， $\Delta y_i \in \{0.1W\}$ 。像素噪声的强度设置沿用StegaStamp^[1]与PIMoG^[3]的相关配置。

为客观评估本文方案的性能，采用可量化指标进行分析。在容量方面，比较每像素平均嵌入比特率(Bits Per Pixel, BPP)；在视觉质量方面，采用峰值信噪比PSNR和结构相似性SSIM作为评价指标^[43]，其值越高表明与原图的相似度越高，即视觉质量越优。此外，我们还添加了感知损失LPIPS来评估图像质量，其使用预训练的深度学习提取图像特征，然后计算这些特征之间的距离，以评估图像之间的感知相似度。因此，LPIPS越低则图像质量越好^[42]；针对屏摄场景的鲁棒性，采用平均水印提取准确率(Accuracy, ACC)进行评估。为评估本文方案的优越性，我们选取了StegaStamp^[1]、RIHOOP^[2]、PIMoG^[3]以及SSDS^[4]作为对比方案。其中，StegaStamp^[1]结合通用的预训练定位模型BiSeNet，作为自动定位水印图像四角并矫正的方案代表。由于其定位依赖于纯白的边框，我们统一将屏摄过程中水印图像的背景设置为纯白。在屏摄实验中，所有方案解码前均不经过人工处理，比较其屏摄鲁棒性。

4.2 视觉质量对比与分析

因为水印方法在嵌入容量、视觉质量和鲁棒性之间存在权衡关系，我们采用与大部分方法相同的尺寸与水印长度，即100比特。在此条件下，本文方案取得了最优的视觉指标，如表1所示。而对于PIMoG^[3]方法，虽然其BPP高于本文方案，但水印嵌入的图像尺寸仅为128，限制了其应用场景。

表1 不同方法视觉质量的指标对比

指标	容量	尺寸	BPP	PSNR	SSIM	LPIPS
StegaStamp ^[1]	100	400	0.000 63	27.85	0.904	0.0910
RIHOOP ^[2]	100	400	0.000 63	28.60	0.936	0.0796
PIMoG ^[3]	30	128	0.001 83	36.21	0.985	0.0233
SSDS ^[4]	100	400	0.000 63	34.17	0.982	0.0356
本文方案	100	400	0.000 63	38.89	0.989	0.0086

为更直观地展示本文方案在视觉质量上的优越性，我们分别选择了纹理细节丰富和匮乏的代表性图像，以对比水印图像的视觉效果。具体地，分别展示了原始图像、水印图像以及水印残差图像。其中，残差图像为原始图像与水印图像差值绝对值的5倍。为了更清楚地展现本文方案的优越性，我们还对两张图的部分区域进行放大，进一步展示了水印嵌入模式的差异。可视化结果如图5所示，不难看出所有方案都倾向于将水印嵌入到图像的高频区域，即细节纹理丰富的区域。水印信号在这些区域不易被人眼观察到，因此所有方案都可以无感地在原图1上嵌入水印。然而，对于纹理细节较少的原图2，其他方案无法抑制对平滑背景区域的修改，出现了肉眼可见的伪影。而本文方案在原图2的嵌入依旧保持了最小限度的背景修改，从而对于此类图像，本文方案的水印图像表现出最优的视觉效果。当切换为原图2时，方案间的性能差异显著显现：其他方案均未能有效抑制对平滑背景区域的修改。具体表现为，这些方案在平滑区域嵌入水印，导致平滑背景的像素值发生可察觉的改动，进而产生肉眼可见的伪影。相比之下，本文方案在原图2中实现了对平滑背景区域的最小限度控制。从残差图像可以清晰看出，本文方案的水印残差与原图具有相似的能量分布，在视觉统一性和自然度上表现最优。

综合上述实验结果，本文方案在核心视觉质量指标上均显著优于其他对比方案，展现出最优的客观视觉指标。从视觉质量可视化实验结果来看，在纹理丰富的图像中，各方案虽均能实现水印的无感嵌入，但在纹理匮乏的平滑区域图像中，唯有本文方

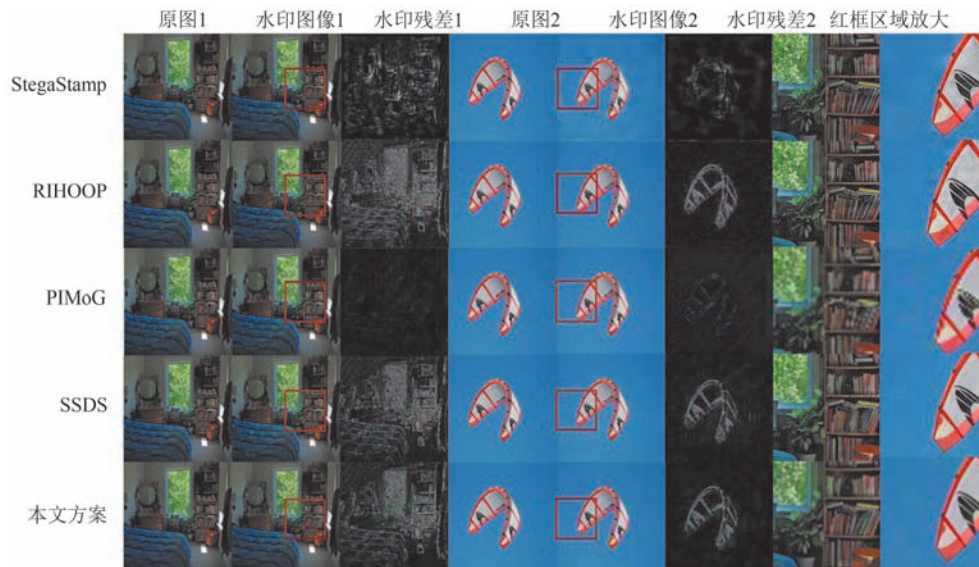


图5 视觉质量可视化对比

案通过精准控制水印嵌入能量,有效避免了平滑背景的视觉伪影,展现出优异的视觉质量。

4.3 抗屏摄鲁棒性对比分析

4.3.1 不同距离的屏摄鲁棒性实验

为验证本文方案在屏幕拍摄场景中对距离变化的鲁棒性,实验选取“TL160ADMP03显示屏”与“小米15智能手机”两款设备作为测试终端(除不同设备对实验外均采用此设置),通过控制拍摄距离变量开展对比分析。实验设置中,屏摄距离梯度依次为20 cm、30 cm、40 cm、50 cm及60 cm,通过控制屏幕显示的图像大小,确保水印区域在屏摄图像中的占比恒定为80%(除不同占比的屏摄实验外均采用此设置),以排除区域占比差异对评估结果的干扰。表2的实验数据表明,所提方案在20-60 cm的测试范围内展现出优异的鲁棒性,水印提取的准确率均值始终保持在97%以上。其中,30 cm处取得最高准确率,该距离既能保证图像拍摄清晰度,又可避免因过近拍摄导致的相机失焦模糊问题;而40 cm处出现最低准确率,因为在该设备对下40 cm处拍摄产生了最明显的摩尔纹干扰,对水印恢复过程造成一定影响。对比方案中,仅StegaStamp^[1]可以基本提取出水印信息,但在较近的距离下与本文方案的性能差距明显。综合来看,本文方案在各测试距离下均保持了稳定且优异的性能表现,显著优于其他对比方案,充分验证了其在不同屏摄距离下的鲁棒性优势。

4.3.2 不同角度的屏摄鲁棒性实验

为评估本文方案在屏摄场景中对角度变化的鲁棒性,实验将拍摄设备固定在距显示图像40 cm处,

表2 不同距离下的屏摄实验

距离	20 cm	30 cm	40 cm	50 cm	60 cm
StegaStamp ^[1]	88.5%	82.3%	93.0%	95.1%	94.7%
RIHOOP ^[2]	57.3%	54.2%	51.7%	48.9%	50.6%
PIMoG ^[3]	64.0%	68.7%	59.3%	48.3%	54.7%
SSDS ^[4]	70.1%	70.3%	67.5%	60.1%	57.2%
本文方案	98.2%	98.5%	97.7%	98.21%	98.4%

通过改变拍摄设备中心与显示设备连线和屏幕法线的夹角进行测试。其中水平与垂直方向的角度变化均以从左向右及从下到上为正方向,角度梯度设置为15°,覆盖从-45°至45°的6个测试角度,实验结果如表3所示。本文方案在水平和垂直方向各角度下均保持了最高的水印提取准确率,远超其他对比方案。由于角度变化越大将伴随越严重的透视畸变,本文方案准确率基本呈现出随拍摄角度变大而降低的趋势,但仍保持在93%以上的高准确率水平。对比方案中,RIHOOP^[2]和PIMoG^[3]在无人工矫正条件下无法正确提取水印信息;SSDS^[4]在低角度屏摄下仅能部分恢复水印;而StegaStamp^[1]借助BiSeNet实现四角定位与透视矫正,虽在较低角度屏摄场景下具备一定鲁棒性,但在大角度屏摄场景下由于过大的四角定位误差基本丧失水印提取能力。上述结果充分验证了所提方案在较大角度变化范围内,能有效抵御屏摄透视畸变影响。

4.3.3 不同占比的屏摄鲁棒性实验

针对水印区域占比的影响,实验设定拍摄距离为40 cm,采用0°正面拍摄方式,通过缩放屏幕显示图像的尺寸调节水印区域占比。实验中,占比梯度

表3 不同角度下的屏摄实验

角度	水平方向						竖直方向					
	-45°	-30°	-15°	15°	30°	45°	-45°	-30°	-15°	15°	30°	45°
StegaStamp ^[1]	52.4%	79.4%	91.8%	92.1%	77.5%	54.0%	58.5%	77.3%	85.8%	91.0%	87.5%	52.9%
RIHOOP ^[2]	49.8%	50.5%	49.2%	50.2%	49.1%	51.8%	50.7%	49.3%	50.1%	48.5%	50.3%	52.0%
PIMoG ^[3]	50.2%	49.7%	50.8%	49.0%	50.3%	51.7%	49.5%	50.6%	49.9%	50.2%	49.3%	48.8%
SSDS ^[4]	57.4%	68.3%	71.3%	68.1%	67.9%	49.5%	52.7%	63.5%	65.2%	68.4%	61.2%	59.0%
本文方案	95.8%	95.9%	98.1%	97.9%	96.4%	95.4%	96.4%	97.3%	97.9%	96.7%	95.0%	93.8%

设为10%，测试范围覆盖50%至90%，共完成5组对照实验，结果如表4所示。观察结果可知，本文方案在不同水印区域占比条件下均保持了优异的稳定性，水印提取准确率始终维持在96%以上。相比之下，对比方案中除StegaStamp^[1]外，其余方案对占比条件的适应性较差：当水印区域占比低于80%时，这些方案基本丧失水印提取能力。StegaStamp^[1]虽在60%至80%的占比范围内表现出较强的提取性能，但在图像占比接近满画幅(90%)或占比较低(50%)的极端情况下，因图像存在某角定位缺失的问题，导致水印几乎无法提取。这种定位失效现象显著拉低了其在极端条件下的提取准确率，体现出明显的不稳定性。综上，本文方案在全范围占比条件下均展现出卓越的鲁棒性，使用者无需进行图像矫正或变焦放大操作，即可实现水印的准确稳定解码。

表4 不同占比下的屏摄实验

占比	50%	60%	70%	80%	90%
StegaStamp ^[1]	72.9%	96.1%	98.5%	93.0%	69.3%
RIHOOP ^[2]	51.2%	49.3%	50.9%	57.3%	64.6%
PIMoG ^[3]	48.6%	51.9%	53.3%	64.0%	94.8%
SSDS ^[4]	47.9%	49.3%	55.8%	70.1%	97.2%
本文方案	97.0%	97.5%	98.1%	97.7%	96.3%

4.3.4 不同设备的屏摄鲁棒性实验

为验证本文水印方案的设备泛化能力，选取两款显示终端(TL160ADMP03显示屏、ENVISION G249G显示屏)与两款拍摄设备(小米15智能手机、苹果13智能手机)组成四组设备对进行屏摄实验。实验严格控制拍摄条件：设备间距固定为40cm，水平拍摄角度设置为30°。具体设备对组合及实验结果如表5所示。实验数据显示，本文方案在所有设备对组合下均展现出卓越的鲁棒性。我们还对四组设备对下的水印提取准确率进行统计分析，计算其平均提取准确率达96.9%，方差为0.515。这一方差充分说明，不同设备对组合对水印提取准确率的影响极小，数据分布高度集中于均值附近，最大差值仅为1.7个百分点。

上述量化分析结果有力表明，本文方案在面对不同品牌、类型的显示终端与拍摄设备时，均能保持稳定且优异的水印提取性能，具备极强的泛化能力。

表5 不同设备对下的屏摄实验

设备对	TL160ADMP03	ENVISION G249G	均值	方差
小米15	96.4%	97.7%	97.1%	0.845
苹果13	96.0%	97.5%	96.7%	1.125
均值	96.2%	97.6%	96.9%	-
方差	0.080	0.020	-	0.515

本文方案在距离、角度、占比的屏摄实验中展现出极佳的鲁棒性，在不同设备对的实验中展现出极强的泛化性。我们总结了不同屏摄实验下各方法的统计学数据，如表6所示计算距离、角度、占比的屏摄实验中的平均值、最大差值以及方差等数据以评估各方案的稳定性。数据显示，本文方案在各个屏摄实验下均取得了最高的平均值，即具有最佳的性能。在此基础上，本文方案在不同距离和不同占比下取得了最小的极差和方差，说明其水印在此屏摄场景下极具稳定性，不易受距离变化或占比变化影响。而在不同角度的屏摄实验中，由于RIHOOP^[2]和PIMoG^[3]几乎无法在不同角度下提取水印，因此本文方案的极差和方差稍逊于RIHOOP^[2]和PIMoG^[3]，但提取准确率的平均值大幅领先其他方案。综上，本文方案在屏摄场景下的提取准确率不仅具有最高的平均值，且在不同距离、角度和占比条件下均表现出极小的性能波动和极佳的稳定性，显著优于其他对比方案，充分验证了其在复杂屏摄环境中的综合优势。

4.4 单变量噪声影响分析

单变量噪声影响分析旨在通过控制变量法，单独探究双态屏摄噪声层中各噪声分量(异步噪声与像素噪声的细分类型)对水印提取性能的影响。通过量化不同噪声强度下的平均水印提取准确率，验证本文方案在各噪声分量干扰下的性能边界。

表6 不同屏摄实验下各方法的稳定性指标

稳定性	不同距离屏摄			不同角度屏摄			不同占比屏摄		
	平均值	最大差值	方差	平均值	最大差值	方差	平均值	最大差值	方差
StegaStamp ^[1]	90.7%	12.8%	23.2	75.0%	39.7%	237.2	86.0%	29.2%	151.6
RIHOOP ^[2]	52.5%	8.4%	8.6	50.1%	3.5%	1.0	54.7%	15.3%	32.1
PIMoG ^[3]	59.0%	20.4%	50.4	49.9%	2.9%	0.6	62.5%	46.2%	287.2
SSDS ^[4]	65.0%	13.1%	29.0	62.7%	21.8%	43.3	64.06%	49.3%	336.4
本文方案	98.2%	0.8%	0.1	96.4%	4.3%	1.6	97.3%	1.8%	0.4

4.4.1 异步噪声单变量影响分析

异步噪声表现为像素位置的空间变化,包含缩放、平移、透视畸变及背景填充四个部分。为探究各分量的独立影响,实验采用单变量控制法,具体设置如下:缩放实验聚焦于缩放分量的影响,控制无透视畸变且水印区域位于图像中心,以水印区域在图像中的占比为调节参数,每间隔10%设置一个梯度,实验范围设定为50%至90%;平移实验以“无透视

畸变、水印区域占比为图像的70%”为基础条件,仅调节平移分量(以右下为正方向),水印区域沿对角线方向平移,平移间隔设为15个像素;透视畸变实验控制水印区域占比为图像的70%且处于中心位置,仅调节透视畸变分量,通过随机方向扰动图像四角实现透视畸变,扰动强度以四角坐标的扰动距离为量化指标,每间隔8个像素设置一个梯度,最大扰动像素距离为32。上述实验结果如图6所示。

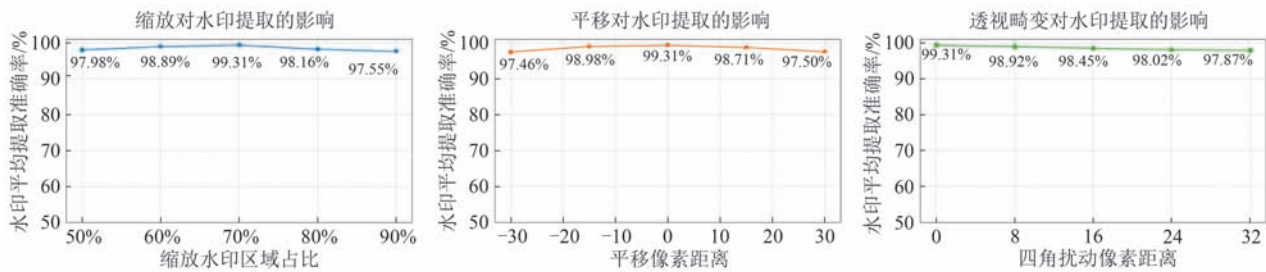


图6 异步噪声单变量影响实验

本文方案在各种单变量的异步噪声影响下均能保持在97%以上的高水平提取准确率,且受噪声强度变化影响小,提取准确率极差不超过2%。从缩放和平移实验可以看出,缩放至70%尺寸并置于图像中心可以获得最佳的水印提取准确率。而透视畸变实验表现出水印提取准确率会受四角扰动影响,扰动像素距离越大,准确率越低,但总体保持水印提取的高鲁棒性。

为探究背景填充对水印提取的影响,我们还在“无透视畸变、水印区域占比为图像的70%且处于图像中心”的实验条件下,使用不同背景进行实验。实验分为填充纯白背景、填充纯黑背景、填充任意RGB颜色背景以及测试集图像背景四组,实验结果如表7所示。替换背景为图像后,结果仅降1.4%,仍保持较高水平。

表7 背景填充对水印提取的影响

背景	纯白背景	纯黑背景	纯色背景	图片背景
ACC	99.3%	99.1%	98.5%	97.9%

综上所述,本文方案在大幅的缩放噪声、不同距离的平移干扰、严重的透视畸变以及复杂的背景填充下,均能通过自同步机制对水印进行同步和提取,充分验证了其在异步噪声影响下的鲁棒性与稳定性。

4.4.2 像素噪声单变量影响分析

在像素噪声的单变量影响实验中,针对不同类型噪声下:对于亮度/对比度及饱和度噪声,以原始参数值1为基准,通过x倍因子进行调节(如亮度实验中,-0.5的噪声强度表示将整体亮度降至原水平的50%,0.5的噪声强度则表示将整体亮度提升至原水平的150%);对于高斯噪声,以图像像素值归一化至0-1区间为前提,将噪声强度设定为高斯分布的标准差(例如0.1的噪声强度对应应在0-1像素值范围内叠加标准差为0.1的高斯噪声);对于运动模糊,采用运动模糊核的尺寸作为噪声强度指标,最大核尺寸设为11;对于光照噪声,随机选取线性光源与点光源,将噪声强度定义为光照处理过程中亮

度变化的倍数;对于摩尔纹噪声,其强度表示为摩尔条纹与图像叠加时的权重占比。最终的实验结果如图7所示,总体显示出水印提取准确率随噪声强度增加而降低的趋势。从实验结果来看,不同类型像素噪声对水印提取准确率影响各异:亮度/对比度噪声中,亮度变化相较于对比度变化对水印的影响更为显著,强度 ± 0.5 时,亮度影响准确率降至81.95%。饱和度噪声随强度增加,准确率缓慢下降,在极端情况下(强度为 ± 0.5 时)仍能保持93.62%的水印提取准确率;高斯噪声与运动模糊

在极强的噪声强度下仍能保持接近90%的水印提取准确率,反映出其具有一定的抵抗能力。光照噪声实验里,准确率随强度上升基本保持稳定,水印提取准确率在噪声影响下仅有最大3%的下降,说明方案对光照变化具备较强的鲁棒性;而摩尔纹噪声在一定强度内,本文方案的抵抗能力强,超过一定限度(如噪声强度大于0.2)时,水印提取准确率出现骤降。整体而言,该方案在各像素噪声影响下,水印提取准确率均保持较高水平,展现出良好的提取性能。

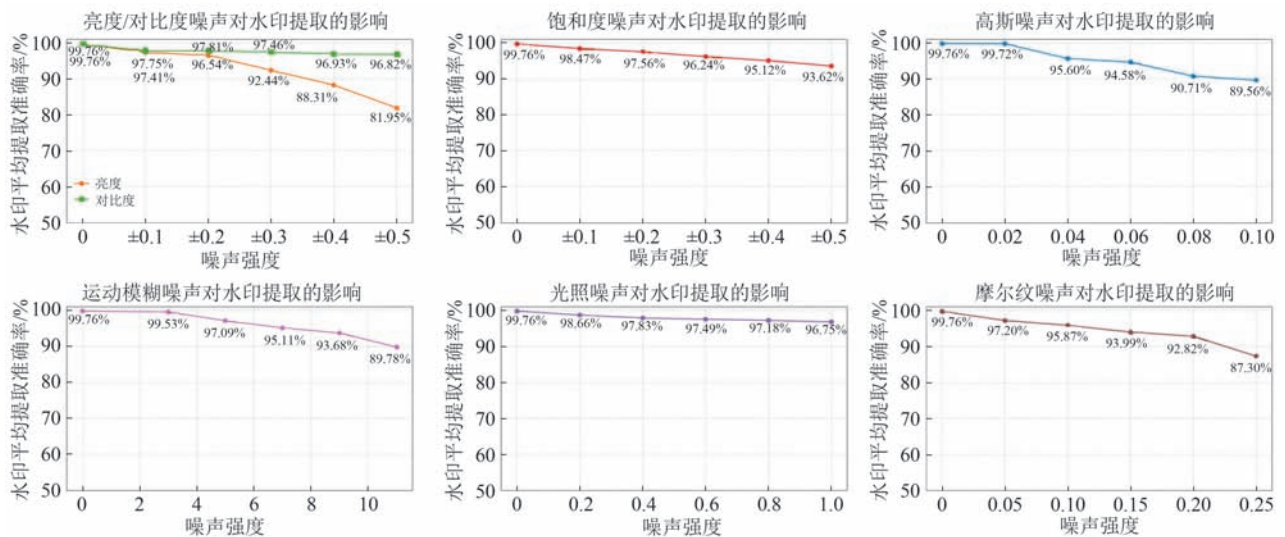


图7 像素噪声单变量影响实验

4.5 消融实验

为深入剖析本文所提方案中各核心模块的具体作用与贡献,特设计并开展消融实验。实验采用控制变量法,在保持测试集、测试环境及评价指标(PSNR、SSIM、ACC)一致的前提下,通过依次移除方案中的关键模块(CBAM模块、EDS损失及自同步机制),将所得消融模型与完整方案模型的性能进行对比分析,从而量化各模块对方案整体性能的独立影响,验证方案的有效性及各模块的必要性。其中,ACC指标的测试基于统一的屏摄环境:设定拍摄距离为40cm,水平角度为 30° ,水印区域占屏摄图像的80%。各消融方案的具体实施方式如下:对于去除CBAM模块的消融模型,将编码器还原为基础U-Net结构;对于去除EDS损失的消融模型,将视觉损失函数中EDS损失的权重系数设为0;对于去除自同步机制的消融模型,先将同步损失的权重系数设为0,再移除解码器中同步掩码的输入。最终实验结果如表8所示,去除CBAM模块后,模型在

达到相近视觉质量的同时,屏摄水印的提取准确率出现一定程度的下降。在视觉指标PSNR和SSIM上,各消融方案与完整方案相比变化幅度较小,其中去除EDS后的消融模型具有最高的视觉指标。当去除自同步机制后,模型在第二阶段引入异步噪声的情况下无法实现有效收敛,其准确率降至50%左右,这表明自同步机制在抵抗异步噪声方面具有不可或缺的作用。需要说明的是,更高的视觉指标并不等同于更高的视觉质量,在相似的视觉指标下,可通过可视化水印图像与残差图像更直观地感受消融模型视觉质量的变化。

因此,除上述量化指标外,我们亦对各个消融模

表8 消融实验的量化指标

	PSNR	SSIM	ACC
本文方案	38.89	0.989	96.4%
去除CBAM	38.11	0.989	91.7%
去除EDS	40.58	0.994	94.9%
去除自同步	-	-	-

型编码的水印图像进行可视化对比,展示了各消融方案的水印图像与水印残差图像。如图8所示,可视化对比的3个水印图像均包含较多低频部分(水印图像1和2的蓝色天空、水印图像3的墙壁),这对水印嵌入构成极高挑战。显而易见的是,去除EDS损失后,三个水印图像的低频区域均出现明显绿色斑点伪影。EDS损失的核心功能在于使编码器生成的水印残差与原始图像保持相似能量分布---通过计算两者局部能量分布图的余弦相似度,约束残

差能量分布贴合原始图像的能量特征。当EDS损失被去除时,水印残差的能量分布不再受到有效约束,与原始图像的能量平衡被打破,对于包含大面积低频区域(如蓝色天空、墙壁)的图像而言,残差能量的异常分布会直接体现在叠加后的水印图像中。而含有EDS损失的方案中,水印图像的低频区域更为纯净,像素修改集中在高频区域,实现了残差图像与原始图像能量分布上的统一,从而获得了主观上更好的视觉效果。

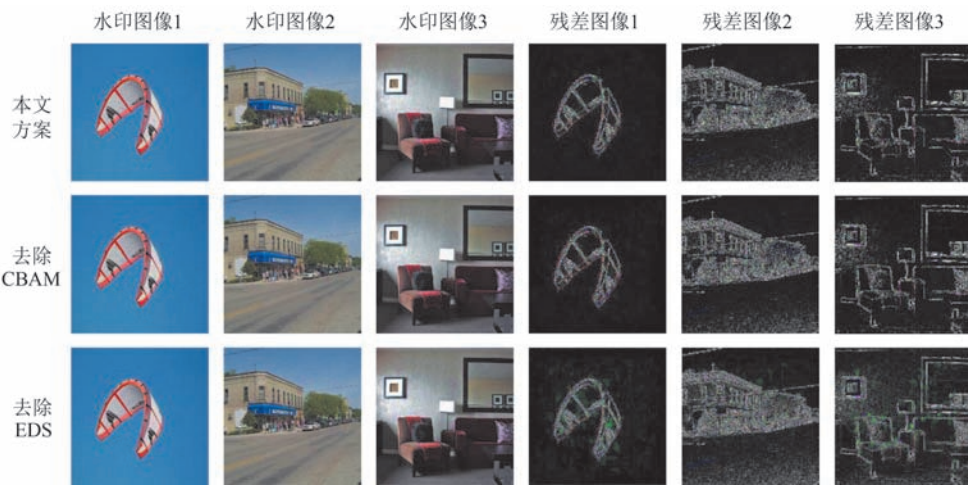


图8 消融实验的可视化对比

综上所述,本文方案中提出的各核心模块均具有其独特的必要性与功能。CBAM模块提升了编码器性能,在轻微提升视觉质量的基础上,大幅提升了屏摄时的水印提取准确率,是保障水印提取性能的关键组件。EDS损失虽会使PSNR、SSIM等客观视觉指标略有降低,却通过约束残差能量分布有效抑制低频区域伪影,对提升水印图像视觉质量具有不可替代的作用;自同步机制是模型抵抗异步噪声、实现有效收敛的核心保障,其缺失会导致水印提取准确率骤降至不可用的水平。三者协同作用,共同保障了本文方案在高视觉质量与强鲁棒性的平衡。

4.6 参数规模与推理耗时

本节聚焦参数规模与推理耗时展开分析,核心研究编码器、同步网络与解码器三大模块的关键性能指标。参数规模统计计算了三个核心模块的可训练参数量,以反映模型部署时的内存占用需求;推理耗时测试时,使用英伟达RTX5090显卡,取1000次连续推理的平均耗时,并排除首次加载模型的初始化耗时,以体现模型实时处理能力。具体统计结果如表9所示。

从参数规模看,本方案编码简单而同步与解码

表9 核心模块的参数规模与推理耗时统计

核心模块	参数数量	参数占比	推理耗时
编码器	1,971,863	1.48%	0.0021 s
同步网络	65,846,401	49.52%	0.0168 s
解码器	65,345,677	49.00%	0.0052 s

复杂,呈现出嵌入轻量化与同步解码精准的特点。从推理耗时看,水印提取,包括同步与解码,总耗时仅0.022秒,可以做到毫秒级响应,满足实时处理需求。综上,本方案在参数规模上兼顾轻量化部署与功能需求,在推理耗时上具备毫秒级响应的实时性,能适配抗屏摄图像水印溯源的实际应用场景。

5 总 结

本文围绕屏幕拍摄场景下的图像盗摄溯源问题展开研究,针对现有抗屏摄水印方案采用透视矫正来抵抗异步噪声的弊端;定位误差通过透视矫正过程传播至每个像素而影响解码准确率,提出了一种基于自同步机制的抗屏摄水印方法。该方法首先对异步噪声建模,并结合像素噪声设计双态屏摄噪声

层,全面模拟屏摄场景下的失真特性。其次,本文还通过理论推导论证透视矫正作为同步方式的弊端,并设计自同步机制规避误差传播。在自同步机制下,同步网络预测同步掩码,为解码器提供同步信息实现无透视矫正的水印解码。此外,本文方案在编码器中引入卷积块注意力机制,并加入能量分布相似性损失,通过多阶段、多损失的自同步屏摄水印训练框架提升水印图像的视觉质量。大量的实验结果验证了该方案在视觉质量上的优越性,以及在各种屏摄场景下的鲁棒性。相关的消融实验也进一步证实了所提各核心模块的有效性和必要性,为抗屏摄水印技术的发展提供了新的思路。

然而,当前技术还存在显著的分辨率适配缺陷,现有方法因依赖透视矫正实现水印同步,而透视变换矩阵的生成又需以原始图像尺寸作为关键先验信息,从而导致其仅能应对预设的固定分辨率图像。本文提出的基于自同步机制的抗屏摄水印方法,因无需依赖透视矫正过程,天然规避了对原始图像尺寸先验的需求,具备解决任意分辨率图像水印提取问题的核心优势。因此,我们的未来工作聚焦于将本文方法扩展于任意分辨率图像的水印方法中,进一步完善任意分辨率水印提取的技术体系,打破现有方法的分辨率局限。

参 考 文 献

- [1] Tancik M, Mildenhall B, Ng R. StegaStamp: Invisible hyperlinks in physical photographs//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2020: 2117-2126
- [2] Jia J, Gao Z, Chen K, et al. RIHOOP: Robust invisible hyperlinks in offline and online photographs. IEEE Transactions on Cybernetics, 2022, 52(7): 7094-7108
- [3] Fang H, Jia Z, Ma Z, et al. PIMoG: An effective screenshooting noise-layer simulation for deep-learning-based watermarking network//Proceedings of the 30th ACM International Conference on Multimedia. Lisboa, Portugal, 2022: 2267-2275
- [4] Li Y, Liao X, Wu X. Screen-shooting resistant watermarking with grayscale deviation simulation. IEEE Transactions on Multimedia, 2024, 26: 10908-10923
- [5] Wengrowski E, Dana K. Light field messaging with deep photographic steganography//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019: 1515-1524
- [6] Sun Jing, Sun Fu-Qi, Hao Shi-Jie, et al. Two-Stage Low-Light Image Enhancement Based on Wavelet Transform, 2025, 48(5): 1188-1211 (in Chinese)
(孙静, 孙福奇, 郝世奇等. 一种基于小波变化的两阶段照度图像增强方法. 计算机学报, 2025, 48(5): 1188-1211)
- [7] Fang H, Chen D, Wang F, et al. TERA: Screen-to-camera image code with transparency, efficiency, robustness and adaptability. IEEE Transactions on Multimedia, 2022, 24: 955-967
- [8] Zhao C, Ling H, Xie S, et al. DBDH: A dual-branch dual-head neural network for invisible embedded regions localization//Proceedings of the 2024 International Joint Conference on Neural Networks. Yokohama, Japan, 2024: 1-7
- [9] Jia J, Gao Z, Zhu D, et al. Learning invisible markers for hidden codes in offline-to-online photography//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA, 2022: 2263-2272
- [10] Jia J, Gao Z, Yang Y, et al. Hidden barcode in sub-images with invisible locating marker. ACM Transactions on Multimedia Computing Communications and Applications, 2024, 20(10): 1-24
- [11] Zhu L, Fang Y, Zhao Y, et al. Lite localization network and DUE-based watermarking for color image copyright protection. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(10): 9311-9325
- [12] Van Schyndel R G, Tirkel A Z, Osborne CF. A digital watermark//Proceedings of the 1st International Conference on Image Processing. Austin, USA, 1994: 86-90
- [13] Hsu C T, Wu J L. Hidden digital watermarks in images. IEEE Transactions on Image Processing, 1999, 8(1): 58-68
- [14] Hernandez J R, Amado M, Perez-Gonzalez F. DCT domain watermarking techniques for still images: Detector performance analysis and a new structure. IEEE Transactions on Image Processing, 2000, 9(1): 55-68
- [15] Cox I J, Kilian J, Leighton F T, et al. Secure spread spectrum watermarking for multimedia. IEEE Transactions on Image Processing, 1997, 6(12): 1673-1687
- [16] Yue Feng, Zhu Hui, Su Zhao-Pin, et al. An Adaptive Audio Steganography Using BN Optimizing SNGAN. Chinese Journal of Computers, 2022, 45(2): 42-440 (in Chinese)
(岳峰, 朱慧, 苏兆品等. 基于BN优化SNGAN的自适应音频隐写. 计算机学报, 2022, 45(2): 42-440)
- [17] Langelaar G C, Setyawan I, Lagendijk R L. Watermarking digital image and video data: A state-of-the-art overview. IEEE Signal Processing Magazine, 2000, 17(5): 20-46
- [18] Bamatraf A, Ibrahim R, Salleh Mohd N B M. Digital watermarking algorithm using LSB//Proceedings of the 2010 International Conference on Computer Applications and Industrial Electronics. Kuala Lumpur, Malaysia, 2010: 155-159
- [19] Kumar S, Yadav A K, Gupta A, et al. RGB image steganography on multiple frame video using LSB technique//Proceedings of the 2015 International Conference on Computer and Computational Sciences. Greater Noida, India, 2015: 226-231
- [20] Li Ji-Yu, Fu Zheng-Jie, Wang Fanet al. Canny-Gauss Universal Domain Image Steganography Algorithm. Chinese Journal of Computers, 2024, 47(1): 213-230 (in Chinese)
(李季瑀, 付章杰, 王帆等. Canny-Gauss通用域图像隐写算法.

- 计算机学报, 2024, 47(1): 213-230)
- [21] Joseph H, Rajan BK. Image security enhancement using DCT & DWT watermarking technique//Proceedings of the 2020 International Conference on Communication and Signal Processing. Chennai, India, 2020: 940-945
- [22] Bei Y L, Qiao S, Liu M X, et al. A color image watermarking scheme against geometric rotation attacks based on HVS and DCT-DWT//Proceedings of the 2018 International Conference on Security, Pattern Analysis, and Cybernetics. Jinan, China, 2018: 343-347
- [23] Fares K., Khaldi A, Redouane K, et al. DCT & DWT based watermarking scheme for medical information security. Biomedical Signal Processing and Control, 2021, 66: 102403
- [24] Peng An-Jie, Zeng Hui, Kang Xiang-Gui. Resampling forensics based on multi-directional difference. Science in China (Information Sciences), 2016, 46(5): 627-642 (in Chinese)
(彭安杰, 曾辉, 康显桂. 基于多方向差分的重采样取证技术. 中国科学:信息科学, 2016, 46(5): 627-642)
- [25] Zhu J, Kaplan R, Johnson J, et al. HiDDeN: Hiding data with deep networks//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany, 2018: 682-697
- [26] Liu Y, Guo M, Zhang J, et al. A novel twostage separable deep learning framework for practical blind watermarking//Proceedings of the 27th ACM International Conference on Multimedia. Nice, France, 2019: 1509-1517
- [27] Ahmadi M, Norouzi A, Karimi N, et al. ReDMark: Framework for residual diffusion watermarking based on deep networks. Expert Systems with Applications, 2020, 146: 113157
- [28] Chen B, Wu Y, Coatrieux G, et al. JSNet: A simulation network of JPEG lossy compression and restoration for robust image watermarking against JPEG attack. Computer Vision and Image Understanding, 2020, 197-198: 103015
- [29] Jia Z, Fang H, Zhang W. MBRS: Enhancing robustness of DNN-based watermarking by mini-batch of real and simulated JPEG compression//Proceedings of the 29th ACM International Conference on Multimedia. New York, USA, 2021: 41-49
- [30] Yin Xiao-Lin, Lu Wei, Zhang Jun-Hong, et al. Robust JPEG steganography based on lossless carrier and robust cost. Journal of Image and Graphics, 2022, 27(1): 238-251 (in Chinese)
(尹晓琳, 卢伟, 张俊鸿等. 无损载体和鲁棒代价结合的 JPEG 图像鲁棒隐写. 中国图象图形学报, 2022, 27(1): 238-251)
- [31] Luo X, Zhan R, Chang H, et al. Distortion agnostic deep watermarking//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2020: 13545-13554
- [32] Fei J, Xia Z, Tondi B, et al. Supervised GAN watermarking for intellectual property protection// Proceedings of the 2022 IEEE International Workshop on Information Forensics and Security. Shanghai, China, 2022: 1-6
- [33] Fang H, Chen D, Huang Q, et al. Deep template-based watermarking. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(4): 1436-1451
- [34] Gao G, Chen X, Li L, et al. Screen-Shooting Robust Watermark Based on Style Transfer and Structural Re-Parameterization. IEEE Transactions on Information Forensics and Security, 2025, 20: 2648-2663
- [35] Zhong X, Huang P C, Mastorakis S, et al. An automated and robust image watermarking scheme based on deep neural networks. IEEE Transactions on Multimedia, 2021, 23: 1951-1961
- [36] Lu J, Ni J, Su W, et al. Wavelet-based CNN for robust and high-capacity image watermarking//Proceedings of the 2022 IEEE International Conference on Multimedia and Expo. Taipei, China, 2022: 1-6
- [37] Zheng B, Yuan S, Yan C, et al. Learning frequency domain priors for image demoiring. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(11): 7705-7717
- [38] Fu L, Liao X, Guo J, et al. WaveRecovery: Screenshooting watermarking based on wavelet and recovery. IEEE Transactions on Circuits and Systems for Video Technology, 2025, 35(4): 3603-3618
- [39] Woo S, Park J, Lee J Y, et al. CBAM: Convolutional block attention module//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany, 2018: 3-19
- [40] Wang J, Sun K, Cheng T, et al. Deep highresolution representation learning for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(10): 3349-336
- [41] Chen Shan-Juan, Yu Yun-Long, Li Ying-Ming. Research Status and Analysis of Few-Shot Semantic Segmentation. Chinese Journal of Computers, 2024, 47(10): 2417-2451 (in Chinese)
(陈善娟, 于云龙, 李英明. 小样本语义分割研究现状与分析. 计算机学报, 2024, 47(10): 2417-2451)
- [42] Zhang R, Isola P, Efros A, et al. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 586-595
- [43] Yan Jie-Bin, Fang Yu-Ming, Liu Xue-Lin, et al. A Survey on Recent Advances in Video Quality Assessment, 2023, 46(10): 2196-2224 (in Chinese)
(鄢杰斌, 方玉明, 刘学林等. 视频质量评价研究综述. 计算机学报, 2023, 46(10): 2196-2224)



LIAO Xin, Ph. D., professor. His main research interests include information content security and artificial

YAO Yi-Zhi, M. S. candidate. His main research interests include screen-shooting resilient watermarking and multimedia information security.

intelligence security.

OU Bo, Ph. D., associate professor. His main research interests include information security and computer vision.

HU Feng-Song, Ph. D., associate professor. His main research interests include multimedia interaction technology and 3D visualization technology.

QIN Zheng, Ph. D., professor. His main research interests include network information security and privacy protection.

Background

In this paper, we address a critical challenge in multimedia security: screen-shooting resilient watermarking. The proliferation of digital displays and high-resolution mobile cameras has made unauthorized screen capture a primary vector for content leakage. Digital watermarking offers a covert solution for tracing these leaks, yet existing methods are fundamentally constrained. They typically rely on error-prone four-corner localization to correct perspective distortions. Crucially, any localization error propagates globally during homography correction, severely compromising the reliability of watermark extraction.

To overcome these limitations, this paper introduces a self-synchronization mechanism based screen-shooting resilient watermarking that eliminates the need for perspective correction. We design a synchronization network that predicts synchronization mask, directly identifying the watermark region in the captured image and enabling the decoder to perform robust extraction without geometric alignment. To better simulate real-world

conditions, we construct a dual-state screen-shooting noise layer that models both asynchronous and pixel-level distortions. Furthermore, our multi-stage, multi-loss training framework incorporates an energy distribution similarity loss to enhance visual quality and uses convolutional block attention module to improve embedding performance. Experimental results demonstrate that our method significantly outperforms existing approaches in both visual quality and extraction robustness under real-world screen-shooting conditions.

This work was supported by the National Natural Science Foundation of China (Grant No. U22A2030), the National Key R&D Program of China (Grant Nos. 2024YFF0618800, 2022YFB3103500), Hunan Provincial Funds for Distinguished Young Scholars (Grant No. 2024JJ2025), Hunan Provincial Key Research and Development Program (Grant Nos. 2024AQ2027, 2025AQ2022), Hunan Provincial Major Science and Technology Program (Grant No. 2025QK2008), and Project of Yuelushan Center for Industrial Innovation (Grant No. 2025YCII0222).