中图法分类号： **文献标识码：** A **文章编号：** 1006-8961(XXXX)XX-0001-19

论文引用格式： Chen Yaoyi, Wang Na, Peng Yanchun, Chen Jiahao, Qiao Pengxu, Wang Wei, Qin Chuan. Screen Content Image Dataset for Screen-shooting Resistant Watermarking — SCID[J/OL]. Journal of Image and Graphics, XXXX:1-19. DOI: 10.11834/jig.250661.(陈尧一, 王娜, 彭彦淳, 陈嘉豪, 乔蓬旭, 王伟, 秦川. 面向抗屏摄水印的屏幕内容图像数据集—SCID[J/OL]. 中国图象图形学报,XXXX:1-19. DOI: 10.11834/jig.250661.)[DOI:10.11834/jig.250661]

面向抗屏摄水印的屏幕内容图像数据集—SCID

陈尧一[1],王娜[1],彭彦淳[1],陈嘉豪[1],乔蓬旭[1],王伟[2],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. 中国人民解放军海军军医大学,上海 200433

摘 要：目的 抗屏摄鲁棒水印技术通过嵌入算法将秘密信息嵌入载体图像,当发现图像被侵权时可经提取算法还原信息,实现版权保护。但目前对屏幕内容图像的版权保护方面缺少专用数据集,导致基于自然图像数据集训练的模型在屏幕内容场景中应用时易出现视觉质量下降的问题。**方法** 针对保护屏幕内容图像的抗屏摄鲁棒水印任务,本文构建了一个新的数据集。在不同的操作系统下,以全屏和窗口两种显示形式,采集了网页类应用10 303张、聊天类应用823张、编程类应用636张、工程制图类应用2 294张、线上会议类应用676张和办公类应用2 369张等不同主题的图像,最终建立了一个包含17 101张图像的面向抗屏摄水印的屏幕内容图像数据集(screen content image dataset, SCID)。**结果** 所构建的数据集涵盖图片、文本等多样化的屏幕内容类型,并选用StegaStamp、MBRS、PIMoG、HiFiMSFA和MTVDGAN五个典型深度学习水印嵌入方法,在SCID与自然图像数据集上训练,并进行多组对比实验。特别地,在屏摄攻击实验中,我们设置不同的光照(50Lux、100Lux、150Lux)、拍摄角度(Up/Down30°和15°、0°、Left/Right30°和15°)、拍摄距离(20cm、30cm、40cm)、显示亮度(45%、60%、75%)和不同设备组合,对各个模型进行测试。实验结果表明,基于自然图像数据集训练的水印模型在SCID上进行水印嵌入测试时,含水印图像的峰值信噪比较其在自然图像数据集上的测试结果下降2~4dB;而基于SCID训练的水印模型在自然图像数据集上测试时,含水印图像的视觉质量保持稳定。在鲁棒性实验中,基于SCID训练的模型在数字攻击和真实屏摄攻击条件下的水印提取准确率与基于自然图像数据集训练模型的测试结果相当,性能差异较小,表现出良好的泛化能力。**结论** 构建了一个面向抗屏摄鲁棒水印的屏幕内容图像数据集,通过大量的对比实验表明了该数据集在屏幕内容版权保护场景下的有效性,该工作可为屏幕内容保护的抗屏摄鲁棒水印技术研究提供有力支撑。为便于学术同行复现与验证,本文构建的数据集将在论文录用后公开,届时将在文中补充完整下载地址。

关键词： 抗屏摄鲁棒水印;版权保护;深度学习;屏幕内容图像;数据集

Screen Content Image Dataset for Screen-shooting Resistant Watermarking — SCID

Chen Yaoyi[1], Wang Na[1], Peng Yanchun[1], Chen Jiahao[1], Qiao Pengxu[1], Wang Wei[2], Qin Chuan[1*]

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

2. Naval Medical Center of PLA, Second Military Medical University, Shanghai 200433, China

收稿日期： 2025-12-30;**修回日期：** 2026-04-13

* **通信作者：** 秦川 qin@usst.edu.cn

基金项目： 国家自然科学基金项目(62571333,62502312,62172280);上海市扬帆计划项目(24YF2730200);上海市教育委员会人工智能驱动的研究范式与学科发展专项

Supported by: National Natural Science Foundation of China (62571333, 62502312, 62172280); Shanghai Sailing Program (24YF27 30200); Shanghai Municipal Education Commission Project of AI-Driven Research Paradigm and Disciplinary Development

Abstract: **Objective** Screen-shooting resistant watermarking technology is an effective copyright authentication technique which has received widespread attention in recent years. It utilizes a pre-designed embedding algorithm to embed secret information into the cover image. When the copyright of the image is found to be infringed, the corresponding extraction algorithm can be used to extract the secret information, thereby achieving copyright protection. However, screen content images as a predominant medium for information transmission, the widespread use of screen-shooting enables low-cost duplication of screen content, posing severe challenges to copyright protection. Although screen-shooting resistant robust watermarking technology has emerged as an effective solution for copyright authentication and infringement tracing, existing research faces a critical problem: the lack of dedicated screen content image datasets. Current deep learning-based watermarking models are primarily trained on natural image datasets such as ImageNet, COCO, and MIR-Flickr. These natural image datasets focus on real-world scenes with rich textures and complex color distributions, which differ fundamentally from screen content images characterized by text-dominated content, large uniform color backgrounds, and vector-based elements like lines and diagrams. This domain gap leads to significant visual quality degradation (e. g. , visible artifacts) when models trained on natural images are applied to screen content. Additionally, existing screen-related datasets are designed for quality assessment tasks and contain a large number of noise-processed images, making them unsuitable for training robust watermarking models that require accurate simulation of real-world screen content scenarios. To address these issues, this study aims to construct a large-scale, high-quality dedicated screen content image dataset to support the development of screen-shooting resistant robust watermarking technologies, thereby bridging the performance gap between natural image and screen content watermarking applications and enhancing the practicality of copyright protection for screen-based information. **Method** A screen content image dataset for screen-shooting resistant watermarking (SCID) was constructed specifically for screen-shooting resistant robust watermarking tasks. The dataset was categorized into six functional themes based on common usage scenarios: webpage applications, chat applications, programming environments applications, engineering drawings applications, online meetings applications, and office applications. For webpage applications images, diverse sources were collected through search engines, including official websites, social media platforms, open-source project repositories, and large language model conversation interfaces, covering text-heavy pages, image-rich content, and interactive interfaces. Chat applications images included interfaces from popular communication software such as QQ and WeChat, as well as public account push and conversation interfaces. Programming applications images captured code displays and runtime results from various development platforms. Engineering drawing applications images consisted of both 2D blueprints and 3D model renderings, covering mechanical, architectural, and electrical design scenarios. Online meeting applications images were collected from remote collaboration tools like Tencent Meeting and FeiShu, including live lectures, video conferences, and remote desktop control scenes. Office images were captured from common office software (Microsoft Office and WPS), including Word documents, Excel spreadsheets, PDF files, and PPT presentations. In total, the SCID contains 17 101 high-resolution images, integrating text, images, diagrams, and video frames to simulate both daily and professional screen usage scenarios. **Result** To validate the effectiveness of SCID, five deep learning watermarking methods (StegaStamp, MBRS, PIMoG, HiFiMSFA and MTVDGAN) were selected for comparative experiments, and the performance of the models was evaluated from three key dimensions: visual quality, robustness against digital attacks, and robustness against real screen-shooting attacks. For visual quality (assessed by PSNR and SSIM), models trained on natural image datasets showed a significant drop of 2~4 dB in PSNR when tested on SCID, indicating obvious visual artifacts in screen content watermarking. In contrast, models trained on SCID maintained stable PSNR and SSIM when tested on natural image datasets, demonstrating that SCID-trained models retain excellent visual quality across domains without additional fine-tuning. In digital attack experiments (including random cropping, JPEG compression, Gaussian blur, Gaussian noise, median filtering, and salt-and-pepper noise), the accuracy difference (AD) of SCID-trained models was consistently better than that of natural image-trained models. This indicates that SCID-trained models have smaller performance fluctuations when transferred between screen content and natural images, reflecting stronger versatility. For real screen-shooting attacks, although the SCID trained model did not show significant performance improvement compared to the model trained on natural image datasets, the fluctuation range of AD was within 0. 1%, that the performance fluctuation range is still within an acceptable range. These results collectively confirm that SCID not only

improves the visual quality of screen content watermarking but also maintains strong robustness against both digital and real-world screen-shooting attacks, while ensuring excellent generalization to natural images. **Conclusion** This study addresses the lack of dedicated datasets for screen-shooting resistant watermarking by constructing the large-scale SCID, which covers 17 101 images across six practical themes. Comparative experiments using three watermarking methods demonstrate that SCID effectively resolves the visual quality degradation issue of natural image-trained models when applied to screen content, while enabling models to retain stable performance on natural images. The dataset's diverse content and realistic scenario simulation enhance the generalization and practicality of watermarking models, providing critical data support for the development of screen content copyright protection technologies. Additionally, SCID can serve as a benchmark dataset for screen-shooting resistant watermarking research, promoting standardized evaluation and technological innovation in the field. This work contributes to advancing copyright protection for digital screen content and provides a foundation for addressing infringement and information leakage challenges in cross-media transmission scenarios. To facilitate replication and verification by academic peers, the dataset constructed in this paper will be made public upon acceptance of the paper, and the complete download link will be provided in the article at that time.

Key words: Screen-shooting resistant watermarking; copyright protection; deep learning; screen content image; dataset

0 引　言

随着多媒体技术和移动拍摄设备技术的快速发展，数字图像的生成和传输方式不断演进(Qian等，2021)，作为数字信息隐藏的重要分支，数字水印技术(Chang等，2017)已成为图像版权保护领域的关键技术之一。以屏幕—拍摄为代表的跨媒介数据传输场景为例(Li等，2024；He等，2024；Xiao等，2024)，用户仅需使用移动终端对屏幕内容进行拍照，即可近乎零成本地获取图像数据，这对屏幕内容的版权保护提出了严峻挑战。抗屏摄水印技术能够通过设计专门的水印嵌入算法，将水印信息以人眼不可感知的方式嵌入到图像中，当图像版权受到侵犯时，可利用相应的提取算法恢复嵌入的水印信息，从而证明图像的版权所有。

为了增强抗屏摄水印的鲁棒性，传统方法通常使用离散余弦变换(Fares等，2021；Mahto等，2023；Wen等，2025)、离散傅里叶变换(Tian等，2013；Li等，2021；Zheng等，2023)等变换域方法，通过调制变换系数来嵌入水印信息。此外，研究者们还提出了多种变换域和混合域水印方法以进一步增强鲁棒性(Onn等，2025；Fikri等，2025；Gen等，2025)。然而这些方法高度依赖于人工设计的嵌入和提取规则，虽然在特定的情况下具有较好的鲁棒性，但在面对多样化、复杂化的真实屏摄场景时，往往缺乏足够的泛化能力。

近年来，深度学习技术的发展为抗屏摄鲁棒水印研究提供了新的解决思路。通过构建不同结构的编解码网络(Wu等，2023；Guo等，2024)，并在训练阶段中引入多种噪声攻击进行对抗性训练，可获得性能更优，适应性更强的抗屏摄水印模型。因此，基于深度学习的抗屏摄水印方案已逐渐成为当前研究的主流方向。然而，现有方法在模型训练阶段主要依赖ImageNet(Li F等，2009)、COCO(Lin等，2014)等自然图像数据集，仍存在以下两个关键问题：1)自然图像数据集以人物、动物、建筑等显示场景为主，通常具有纹理丰富、像素分布无序等特点，而屏幕内容则以文字、线条和图形界面元素为主体，往往包含大面积同色背景，纹理复杂度较低，像素分布更加规则。二者在统计特性和视觉结构上的显著差异，使得在自然图像数据集上训练得到的水印嵌入模型，在应用于屏幕内容图像时容易产生明显的视觉伪影，从而降低含水印图像的视觉质量。2)已有部分研究构建了屏幕内容图像相关数据集，但这类数据集通常包含大量经噪声处理或失真模拟后的图像，其主要用途在于屏幕内容质量评估，而非版权保护场景下的抗屏摄鲁棒水印模型训练，因此并不适合作为专用训练数据集。

针对上述问题，本文构建了一个全新的面向抗屏摄鲁棒水印任务的屏幕内容图像数据集(screen content image dataset，SCID)。该数据集在不同操作系统环境下采集，涵盖文字、图片、工程制图、视频会议等多种主题内容，力求真实模拟实际应用中具有版权价值的屏幕内容。

本文构建的SCID数据集共包含17 101张图像，

旨在为抗屏摄鲁棒水印模型训练提供专用数据支撑。主要创新点如下：

1）在 Windows、Linux 和 MacOS 三种操作系统环境下，分别以全屏与窗口两种显示形式采集数据，涵盖网页类应用10 303张、聊天类应用823张、编程类应用636张、工程制图类应用2 294张、线上会议类应用676张以及办公类应用2 369张，共计17 101张屏幕内容图像，主题类型丰富，覆盖典型应用场景。

2）大量实验结果表明，在自然图像数据集上训练得到的水印模型，在SCID上进行水印嵌入测试时，含水印图像的视觉质量明显下降，其峰值信噪比（Peak Signal to Noise Ratio，PSNR）相较于在对应自然图像数据集上的测试结果降低约2～4dB；而在SCID上训练得到的水印模型，在自然图像数据集上进行测试时，含水印图像的PSNR和结构相似性（Structure Similarity Index Measure，SSIM）均保持稳定，表明基于SCID训练的模型能够以更高的视觉质量对屏幕内容进行保护。

3）在鲁棒性实验中，基于SCID训练的水印模型在数字攻击和真实屏摄攻击条件下，其水印提取准确率与基于自然图像数据集训练模型的测试结果相当，性能差异较小，表明该模型在保持鲁棒性的同时具备良好的泛化能力。

1 相关工作

本节首先介绍近年来基于深度学习的数字水印方法及抗屏摄鲁棒水印技术的研究进展，随后分析了现有用于抗屏摄鲁棒水印模型训练的自然图像数据集及其面临的现实局限，为后续屏幕内容图像数据集的构建以及相关实验研究奠定了理论基础。

1.1 基于深度学习的数字水印方案

近年来，基于深度学习的数字水印技术取得了快速发展，在鲁棒性与隐蔽性之间的权衡方面取得了显著进展（Zheng等，2021）。Zhu等人（2018）提出了首个端到端训练框架Hidden，其设计的“编码器-噪声层-解码器”架构（Encoder - Noise Layer - Decoder，END）为后续研究奠定了基础。围绕该框架，相关研究主要从噪声层设计和编解码器结构优化两个方向展开。在噪声层设计方面，为增强模型对JPEG压缩的鲁棒性，Jia等人（2021）提出了MBRS方法，通过引入小批次真实JPEG与模拟JPEG的混合噪声层，有效提升了模型在JPEG压缩条件下的鲁棒性，同时对多种其他失真类型亦表现出较好的抵抗能力，该方法随后成为抗屏摄水印研究中的常用对比方案。在编解码器结构设计方面，为增强编码器与解码器之间的协同性，Fang等人（2022b，2023）提出在编码器前引入解码器，并通过共享解码器权重加强编码与解码过程之间的耦合关系，从而提升水印嵌入效果；同时，利用可逆网络结构进一步增强了编解码器的耦合性。Sun等人（2025）提出了一种双解码器结构，分别用于处理未受噪声攻击和受攻击的水印图像，从而在一定程度上规避了噪声不可微问题。

上述方法主要面向数字传输场景进行设计，而屏幕拍摄所引入的真实失真具有更强的复杂性和不确定性，这也为抗屏摄鲁棒水印技术的研究带来了更大的挑战。

1.2 抗屏摄鲁棒水印方案

在真实应用环境中，屏幕—拍摄过程所产生的失真机制复杂多样，如何通过噪声层设计准确刻画真实屏摄失真，成为抗屏摄水印研究中的核心问题。Tancik等人（2019）提出的StegaStamp方法将屏幕拍摄过程分解为一系列图像操作，通过优化噪声层建模提升水印的抗屏摄能力。Fang等人提出的PIMoG方法（2022a）通过保留透视变换失真、光照失真、摩尔纹失真以及高斯噪声等关键噪声层，对模型性能进行了系统优化。Wengrowski等人（2019）提出利用神经网络模拟屏幕拍照过程（Light Field Messaging，LFM），通过拍摄数据集训练屏摄失真模拟函数，并将其作为噪声层引入模型训练，从而显著提升了水印模型的抗屏摄能力。然而，LFM方法对大规模拍摄数据的依赖较强，难以保证失真模拟网络的通用性。为此，Fang等人（2023）引入小样本学习策略，以提升失真模拟网络的多样性，从而增强模型对新型失真的鲁棒性。此外，Guo等人（2024）设计了双分支编码器结构，并引入最小可视察误差（Just Noticeable Distortion，JND）约束函数辅助训练；也有研究工作（Zhang等，2020；Li等，2023）提出了通用性深度隐藏（Universal Deep Hiding，UDH）框架，通过调整编码器输入形式，将原先同时输入载体图像和水印信息的方式，改为仅输入水印信息，从而生成仅包含水印信息的信息膜。

综上所述，现有方法在应对屏幕拍摄场景下的多种失真方面已取得一定进展，但这些方法在训练阶段大多依赖自然图像数据集，缺乏面向屏幕内容保护的专用数据集支持，限制了抗屏摄水印技术在屏幕内容版权保护场景中的应用效果。本文构建的数据集正是针对上述问题提出的解决方案。

1.3 抗屏摄鲁棒水印模型训练使用的数据集

目前，抗屏摄鲁棒水印模型训练时使用的数据集主要包括 ImageNet、COCO 和 MIRFlicker25K：

1）ImageNet 数据集目前包含超过 1 400 万张高分辨率图像，覆盖约 2.2 万个 WordNet 同义词集类别，其中超过 100 万张图像附有边界框标注，可直接用于图像分类、目标检测与定位等任务。

2）COCO 数据集由微软于 2014 年提出，当前包含约 32.8 万张图像，强调物体与场景之间的上下文关系，已成为目标检测、实例分割、全景分割、人体关键点检测和图像字幕生成等任务的标准数据集。

3）MIRFlicker25K（Huiskes 等，2008）是从 Flickr 平台采集并筛选的 2.5 万张真实世界图像，侧重体现社交图像中的标签噪声特性与多模态学习需求，广泛应用于标签去噪、跨模态检索、弱监督学习和零样本识别等研究方向。

上述三种自然图像数据集在不同抗屏摄水印方法的训练集与测试集设置中如表 1 所示。

表 1 用于训练和测试不同抗屏摄水印方法的数据集

Table 1 Datasets for training and testing on different screen-shooting resistant watermarking methods

方法名称	训练集	测试集
StegaStamp（Tancik 等，2019）	MIRFlicker	ImageNet
MBRS（Jia 等，2021）	ImageNet	COCO
SSRIW（Wang 等，2025）	COCO	ImageNet
FPSMark（Chen 等，2025）	COCO	MIRFlicker
PIMoG（Fang 等，2022a）	COCO	COCO
HiFiMSFA（Zhang 等，2025）	MIRFlicker	MIRFlicker
MTVDGAN（Gao 等，2025）	ImageNet	ImageNet

2 数据集构建

针对现有抗屏摄鲁棒水印研究中普遍依赖自然图像数据集、缺乏面向屏幕内容保护专用数据支撑的问题，本文围绕屏幕内容图像的典型视觉特性与实际应用场景，设计并构建了一个面向抗屏摄鲁棒水印任务的屏幕内容图像数据集。该数据集从数据采集环境、内容主题类型以及数据规模等方面进行系统规划，力求真实反映屏幕显示内容在版权保护场景中的实际分布特征。

2.1 数据集的构建与采集主题

本文所构建的数据集以真实用户使用场景为核心导向，综合考虑不同操作系统及应用软件显示形态的差异性，以确保数据在内容覆盖和视觉特征上的全面性与代表性。数据集基于当前主流桌面操作系统（Windows、Linux、MacOS）进行采集，涵盖用户日常生活与工作中高频使用的典型系统环境。在应用软件显示形态方面，同时包含全屏模式与窗口模式，力求真实还原用户实际操作过程中的屏幕视觉呈现特征。各主题类别的具体采集范围及数据规模如表 2 所示，相关内容说明如下：

表 2 数据集采集主题与样本规模统计

Table 2 Statistics of collection themes and /sample size in our dataset

类型	示例软件	数量（张）
网页类应用	Edge、Google、Firefox 等	10 303
办公类应用	Microsoft Office、WPS 等	2 369
工程制图类应用	AutoCAD、SolidWorks	2 294
聊天类应用	QQ、微信等	823
线上会议类应用	腾讯会议、飞书等	676
编程类应用	Visual Studio、Pycharm 等	636

1）网页类应用：该类别旨在模拟用户获取信息的典型使用场景。通过 Edge、Chrome、Firefox 等主流浏览器，围绕地图导航、学术文献、社区论坛及电子商务等多种主题关键词进行检索，以覆盖网页内容在结构与视觉呈现方面的多样性。采集内容包括各类官方网站的首页及二级页面，包含大篇幅图像与视频的多媒体网页，大语言模型交互界面，微博、知乎等社交媒体与技术社区页面，GitHub、Gitee 等开源项目托管平台及在线代码沙盒界面，以及中国知网、Google Scholar 等在线文献阅读与下载界面。本类别共采集屏幕内容图像 10 303 张。

2）办公类应用：该类别模拟用户日常办公场景。数据采集涵盖 Microsoft Office 与 WPS 等办公软件的

界面,包括Word文档(含图文混排、表格、公式等)、PPT演示文稿(含不同版式与动画预览)、PDF文件(含扫描版与可编辑版)以及PNG、JPG等图像文件的查看界面。本类别共采集屏幕内容图像2 369张。

3)工程制图类应用:该类别用于模拟用户进行工程图纸设计与查看的典型场景。数据采集涵盖AutoCAD软件中的2D工程图纸(如机械零件图、建筑施工图)以及SolidWorks软件中的3D模型图纸(如产品装配图、三维效果图),充分体现尺寸标注、图层管理、视角切换等专业操作特征。本类别共采集屏幕内容图像2 294张。

4)聊天类应用:该类别用于模拟用户之间进行即时通信的典型使用场景,采集对象涵盖QQ、微信等主流即时通讯软件的单人对话与群聊界面。同时,还包括微信公众号、QQ公众号的消息推送界面及图文内容展示界面,覆盖文字、图像、文件、表情等多种消息类型在不同交互形式下的显示场景。本类别共采集屏幕内容图像823张。

5)线上会议类应用:该类别模拟用户进行线上视频会议交流的场景。数据采集涵盖腾讯会议、Zoom、飞书等主流远程协作软件的会议界面,包括视频通话窗口、屏幕共享界面、课件展示界面及远程控制界面,充分体现多参会者视频窗口、会议控制面板、实时批注等典型场景特征。本类别共采集屏幕内容图像676张。

6)编程类应用:该类别用于模拟用户进行代码编写、调试及运行结果展示的典型场景。数据采集基于Visual Studio Code、PyCharm等主流编程环境,涵盖Python、Java、C++等多种编程语言的代码编辑界面、调试界面和部分程序的运行结果展示界面,包含语法高亮、代码折叠、控制台输出等典型编程场景特征。本类别共采集屏幕内容图像636张。

2.2 数据集的可视化展示

图1展示了数据集的部分图像示例,其中前七列为全屏状态下的图像,最后一列为窗口化图像示例。通过对不同主题的图像采集,本数据集覆盖了大量典型使用场景,不仅模拟了用户日常生活中的操作,如视频浏览、图片搜索及相关问题查询等,也涵盖了工作场景中的关键应用,例如代码或图纸的隐私保护以及重要视频会议等保密场景。此外,

本数据集在网页类采集中包含地图、图片、视频等多种搜索结果,也混入了部分自然图像,这有助于提升训练模型在其他数据集上的泛化能力。综上所述,与常见的自然图像数据集相比,本数据集来源于屏幕内容,类型和来源更加多样,同时兼具部分自然图像特征,能够更好地满足不同用户需求。利用该数据集训练得到的水印模型,不仅能够实现屏幕内容的高效水印嵌入与提取,也能够在自然图像数据集上保持良好性能,具备较强的泛化能力。

3 数据集验证与评估

为了系统验证所构建的SCID在抗屏摄鲁棒水印模型训练中的有效性,本节设计多组对比实验。通过选取主流深度学习水印嵌入方法,分别在SCID与常用自然图像数据集上进行训练与测试,从视觉质量和鲁棒性两个核心维度进行量化评估,全面验证数据集的性能支撑能力。

3.1 实验设置

本实验选取了五种基于深度学习的水印嵌入方法:StegaStamp、MBRS、PIMoG、HiFiMSFA和MTVDGAN,分别在COCO、MIRFlicker、ImageNet以及本文构建的SCID数据集上进行性能比较。每个数据集划分为训练集10 000张、验证集1 000张、测试集100张。所有载体图像的分辨率均为128×128。在嵌入容量方面,StegaStamp和MTVDGAN嵌入64 bits,HiFiMSFA嵌入50 bits,MBRS与PIMoG嵌入30 bits。所有模型均在NVIDIA GeForce RTX 4060 GPU上进行训练。

3.2 视觉质量对比

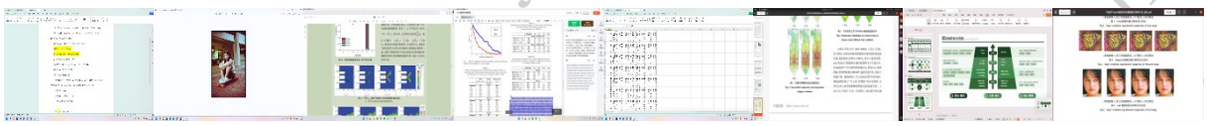
本小节对水印嵌入后图像的视觉质量进行系统评估。通过对比在SCID上训练的模型在自然图像数据集上的表现与在自然图像数据集上训练的模型在SCID上的表现,并结合PSNR和SSIM两类经典视觉质量评价指标,分析不同训练数据集对模型生成含水印图像视觉质量的影响。

PSNR是图像压缩与重建等领域中常用的客观质量评价指标,用于衡量重建图像与参考图像之间的像素级误差,其通过均方误差MSE进行定义,计算公式如下:

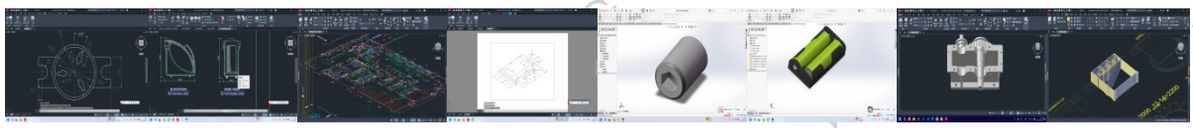
$$\mathrm{PSNR}=10\cdot\log_{10}\left(\frac{MAX_I^2}{\mathrm{MSE}}\right),\tag{1}$$

式中,MAX_I是表示图像像素的最大数值,MSE=

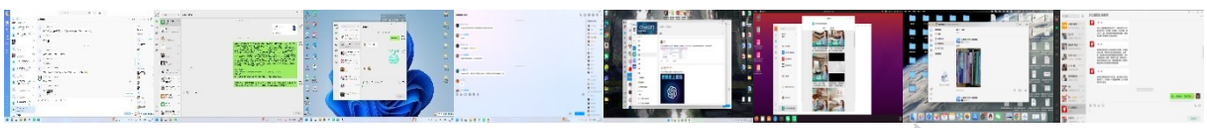
（a） 网页类应用的图像示例



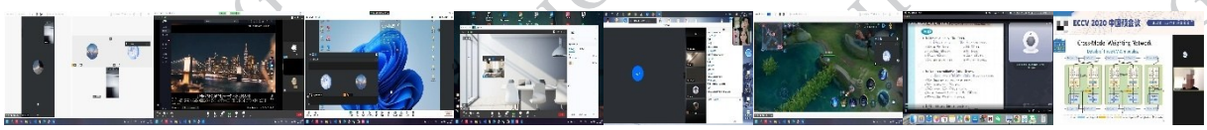
（b） 办公类应用的图像示例



（c） 工程制图类应用的图像示例



（d） 聊天类应用的图像示例



（e） 线上会议类应用的图像示例

（f） 编程类应用的图像示例

((a) Webpage class; (b) Office class; (c) Engineering drawing class; (d) Dialog class; (e) Online meetings class; (f) Programming class)

图1 数据集的图像示例

Fig. 1 Image examples in our dataset

$\frac{1}{H\times W}\sum_{i=0}^{H-1}\sum_{j=0}^{W-1}[I(i,j)-J(i,j)]^2$，其中$H$和$W$为图片的高度和宽度。PSNR数值越大，表示两幅图像之间的像素差异越小，含水印图像的视觉失真程度越低。

SSIM通过从亮度、对比度与结构三个维度构建数学模型，量化评估图像x和图像y的相似性，其计算公式如下：

$$\mathrm{SSIM}(x,y)=\frac{(2\mu_x\mu_y+c_1)(2\sigma_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)},\tag{3}$$

式中，μ_x是x的平均值，μ_y是y的平均值，σ_x是x的标准差，σ_y是y的标准差，σ_{xy}是x和y的协方差，$c_1=(k_1L)^2$，$c_2=(k_2L)^2$是用来维持稳定的常数，L是像素值的动态范围，k_1=0.01，k_2=0.03。SSIM取值范围为[0，1]，数值越接近1，表示图像结构保留越完整，视觉质量越高。

表3给出了五种模型的PSNR测试结果。在StegaStamp的测试结果中，第一行对应于在COCO训练集上训练的StegaStamp模型，其在COCO测试集与SCID测试集上的PSNR分别为23.989 dB和其在SCID测试集与COCO测试集上的PSNR分别为23.091 dB和23.313 dB，相比之下PSNR提升了0.222 dB。可以观察到，在COCO上训练的模型，当测试集由COCO切换至SCID时，PSNR出现显著

表3　不同模型的含水印图像在不同训练—测试数据集上的PSNR

Table 3　PSNR of watermarked images from different models on different training-testing datasets

模型	训练集—测试集	PSNR d_1(dB)	训练集—测试集	PSNR d_2(dB)	d_2-d_1 (dB)
StegaStamp (64 bits)	COCO-COCO	23.989	COCO-SCID	19.987	−4.002
	SCID-SCID	23.091	SCID-COCO	23.313	**0.222**
	MIRFlicker-MIRFlicker	24.110	MIRFlicker-SCID	22.170	−1.940
	SCID-SCID	23.091	SCID-MIRFlicker	23.231	**0.139**
	ImageNet-ImageNet	22.749	ImageNet-SCID	20.687	−2.062
	SCID-SCID	23.091	SCID-ImageNet	22.979	**0.112**
MBRS (30 bits)	COCO-COCO	24.563	COCO-SCID	20.384	−4.179
	SCID-SCID	27.047	SCID-COCO	27.686	**0.639**
	MIRFlicker-MIRFlicker	26.711	MIRFlicker-SCID	21.180	−5.531
	SCID-SCID	27.047	SCID-MIRFlicker	28.033	**0.986**
	ImageNet-ImageNet	28.310	ImageNet-SCID	21.407	−6.903
	SCID-SCID	27.047	SCID-ImageNet	28.666	**1.619**
PIMoG (30 bits)	COCO-COCO	28.053	COCO-SCID	25.717	−2.336
	SCID-SCID	24.739	SCID-COCO	24.832	**0.093**
	MIRFlicker-MIRFlicker	22.178	MIRFlicker-SCID	23.042	**0.864**
	SCID-SCID	24.739	SCID-MIRFlicker	24.520	−0.219
	ImageNet-ImageNet	24.795	ImageNet-SCID	24.044	−0.751
	SCID-SCID	24.739	SCID-ImageNet	24.500	**0.239**
HiFiMSFA (64 bits)	COCO-COCO	26.936	COCO-SCID	18.810	−8.126
	SCID-SCID	26.871	SCID-COCO	26.425	**−0.446**
	MIRFlicker-MIRFlicker	25.944	MIRFlicker-SCID	20.349	−5.595
	SCID-SCID	26.871	SCID-MIRFlicker	25.693	**−1.178**
	ImageNet-ImageNet	27.819	ImageNet-SCID	21.036	−6.783
	SCID-SCID	26.871	SCID-ImageNet	26.727	**−0.144**
MTVDGAN (50 bits)	COCO-COCO	25.547	COCO-SCID	22.376	−3.171
	SCID-SCID	25.019	SCID-COCO	25.103	**0.084**
	MIRFlicker-MIRFlicker	26.839	MIRFlicker-SCID	23.532	−3.307
	SCID-SCID	25.019	SCID-MIRFlicker	25.011	**−0.008**
	ImageNet-ImageNet	25.583	ImageNet-SCID	23.105	−2.478
	SCID-SCID	25.019	SCID-ImageNet	24.969	**−0.050**

下降，表明模型在屏幕内容图像上的水印嵌入会引入较明显的视觉失真；相反，在SCID上训练的模型，当测试集由SCID切换至COCO时，PSNR不仅未出现下降，反而呈现小幅提升，视觉质量保持较为稳定。StegaStamp的其余两组对比结果同样呈现出一致规律，即在自然图像数据集上训练的模型在不同测试集上测试时PSNR均明显降低，而在SCID上训练的模型在不同测试集上的PSNR则保持稳定或略有提升。此外，表3中MBRS、PIMoG、HiFiMSFA和MTVDGAN四种模型的测试结果与StegaStamp模型表现一致，进一步验证了上述结论。具体而言，在自然图像数据集上训练的模型，当测试于SCID时，

表4　不同模型的含水印图像在不同训练—测试数据集上的SSIM

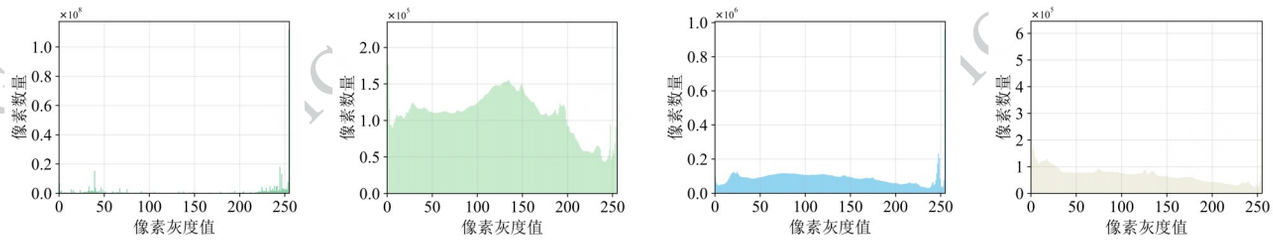
Table 4　SSIM of watermarked images from different models on different training-testing datasets

模型	训练集—测试集	SSIM s_1	训练集—测试集	SSIM s_2	s_2-s_1
StegaStamp (64 bits)	COCO-COCO	0.911	COCO-SCID	0.896	**-0.015**
	SCID-SCID	0.940	SCID-COCO	0.905	-0.035
	MIRFlicker-MIRFlicker	0.946	MIRFlicker-SCID	0.965	**0.019**
	SCID-SCID	0.940	SCID-MIRFlicker	0.895	-0.045
	ImageNet-ImageNet	0.927	ImageNet-SCID	0.946	**0.019**
	SCID-SCID	0.940	SCID-ImageNet	0.901	-0.029
MBRS (30 bits)	COCO-COCO	0.906	COCO-SCID	0.860	-0.046
	SCID-SCID	0.966	SCID-COCO	0.972	**0.006**
	MIRFlicker-MIRFlicker	0.940	MIRFlicker-SCID	0.908	-0.032
	SCID-SCID	0.966	SCID-MIRFlicker	0.964	**-0.002**
	ImageNet-ImageNet	0.954	ImageNet-SCID	0.925	-0.029
	SCID-SCID	0.966	SCID-ImageNet	0.967	**0.001**
PIMoG (30 bits)	COCO-COCO	0.908	COCO-SCID	0.913	**0.005**
	SCID-SCID	0.840	SCID-COCO	0.834	-0.006
	MIRFlicker-MIRFlicker	0.845	MIRFlicker-SCID	0.843	**-0.002**
	SCID-SCID	0.840	SCID-MIRFlicker	0.823	-0.017
	ImageNet-ImageNet	0.839	ImageNet-SCID	0.839	**0**
	SCID-SCID	0.840	SCID-ImageNet	0.812	-0.028
HiFiMSFA (64 bits)	COCO-COCO	0.940	COCO-SCID	0.910	-0.030
	SCID-SCID	0.937	SCID-COCO	0.939	**-0.002**
	MIRFlicker-MIRFlicker	0.944	MIRFlicker-SCID	0.929	**-0.015**
	SCID-SCID	0.937	SCID-MIRFlicker	0.920	-0.017
	ImageNet-ImageNet	0.946	ImageNet-SCID	0.909	-0.037
	SCID-SCID	0.937	SCID-ImageNet	0.938	**-0.001**
MTVDGAN (50 bits)	COCO-COCO	0.932	COCO-SCID	0.923	-0.009
	SCID-SCID	0.941	SCID-COCO	0.936	**-0.005**
	MIRFlicker-MIRFlicker	0.921	MIRFlicker-SCID	0.915	-0.006
	SCID-SCID	0.941	SCID-MIRFlicker	0.941	**0**
	ImageNet-ImageNet	0.924	ImageNet-SCID	0.910	-0.014
	SCID-SCID	0.941	SCID-ImageNet	0.932	**-0.009**

PSNR普遍下降了2~4 dB，说明该类模型在屏幕内容图像上进行水印嵌入时更容易产生可感知的视觉伪影；而在SCID上训练的模型，当测试于自然图像数据集时，其PSNR波动均控制在1 dB以内，显示出更强的跨数据集稳定性。

图2给出了数据集的像素灰度分布。可以看出，SCID的像素灰度值高度集中于特定区间，呈现出明显的“像素集中化”特征，而其余三类自然图像数据集的灰度分布则更为离散。由于PSNR对像素级偏差高度敏感，在自然图像数据集上训练的模型应用

会被显著放大，从而引发MSE急剧增加并导致



（a） SCID （b） COCO （c） ImageNet （d） MIRFlicker

图2 数据集的像素灰度分布图

Figure. 2 Pixel grayscale distribution of each dataset

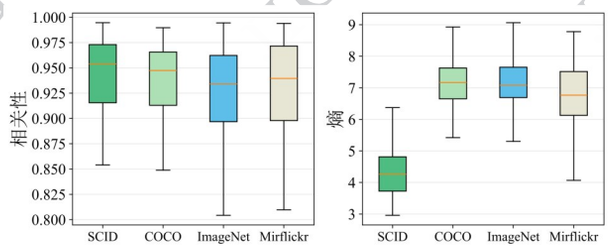


图3 灰度共生矩阵特征对比图

Figure. 3 Comparison of GLCM features

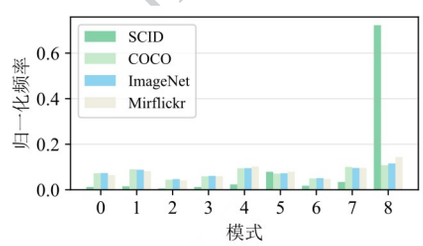


图4 局部二值模式直方对比图

Figure. 4 Comparison of LBP rectangles

PSNR明显下降；相比之下，在SCID上训练的模型能够有效约束纯色区域的像素偏移，即便应用于纹理更为丰富的自然图像，也不会引入显著的像素误差，从而维持视觉质量的整体稳定性。

表4给出了五种模型的SSIM测试结果。在StegaStamp的测试结果中，第一行对应于在COCO训练集上训练的StegaStamp模型，其在COCO测试集与SCID测试集上的SSIM分别为0.911和0.896，SSIM下降幅度为0.015。第二行对应于在SCID训练集上训练的StegaStamp模型，其在SCID测试集与COCO测试集上的SSIM分别为0.940和0.905，SSIM下降幅度为0.035。可以观察到，无论是在COCO上训练的模型，还是在SCID上训练的模型，其在不同测试数据集上得到的SSIM差值均较小，且未表现出显著差异。StegaStamp的其余两组对比实验结果同样呈现出一致规律，即无论模型训练于自然图像数据集还是SCID数据集，在跨数据集测试时，SSIM仅出现小幅波动，整体保持稳定。此外，表4中MBRS、PIMoG、HiFiMSFA和MTVDGAN四种模型的测试结果与StegaStamp模型表现一致，进一步验证了上述结论。总体而言，在自然图像数据集上训练的水印模型与在SSIM指标上的差异均不明显。

图3给出了各数据集的灰度共生矩阵（Gray Level Co-occurrence Matrix，GLCM）相关性对比结果。可以看出，SCID的GLCM相关性特征平均值达到0.95，显著高于其余三类自然图像数据集，表明SCID中图像纹理在结构方向上具有更强的一致性。同时，SCID的GLCM熵平均值明显低于自然图像数据集，说明其纹理复杂度相对较低。图4展示了不同数据集的局部二值模式（Local Binary Patterns，LBP）直方图分布情况。可以观察到，SCID的LBP模式高度集中于少数类别，其中模式8的占比超过70%，而其余三类自然图像数据集在各LBP模式上呈现较为均匀的分布特征。这表明SCID的图像结构模式更加单一，具有显著的“强规则结构”特征。由于SSIM指标对图像整体结构较为敏感，在自然图像数据集上训练的模型应用于屏幕内容图像时，并未对其结构信息造成明显破坏，因此SSIM仅出现小幅波动。同时，自然图像数据集较低的GLCM相关性反映出其结构更加无序，从而在一定程度上提升了模型对结构扰动的容错能力，使得模型在跨数据集应用时不会引入额外的结构破坏，进一步缩小了SSIM差值。

综上所述，SCID所具有的“像素集中化”特征使得在其上训练的模型在PSNR指标上表现出更好的稳定性，而其“强规则结构”特征则有助于保障

图像结构信息的完整性。因此，无论模型训练于SCID还是自然图像数据集，其在不同测试数据集上的SSIM指标均能够保持相对稳定。

3.3 数字攻击实验对比分析

本小节对嵌入水印后的鲁棒性进行数字攻击对比实验。实验选取的数字攻击类型包括随机裁剪（Crop）、JPEG压缩、高斯滤波（Gaussian Filter，GF）、高斯噪声（Gaussian Noise，GN）、中值滤波（Median Filter，MF）以及椒盐噪声（Salt-and- pepper noise，SP），具体攻击参数设置如表5所示。

实验采用准确率差值（Accuracy Difference，AD）作为评价指标，其计算公式如下：

$$\mathrm{AD} = BAR_1 - BAR_2 = \frac{n_1}{l} - \frac{n_2}{l}, \tag{4}$$

式中，l为原始水印信息的比特位数，n_1和n_2是分别表示在“训练集与测试集来源不同”和“训练集与测试集来源相同”场景下提取水印与原始水印匹配的正确比特位数，BAR_1和BAR_2分别对应上述两种场景的比特准确率。AD数值越大，表明模型在跨数。据集测试时的性能越好；反之则说明模型泛化能力较弱。

图5给出了StegaStamp模型的测试结果。其中，

表5 数字攻击的类型和强度

Table 5 Types and strength of digital attacks

攻击类型	攻击强度
随机裁剪	(50%, 60%, 70%, 80%, 90%)
JPEG压缩	(50, 60, 70, 80, 90)
高斯滤波	(1, 3, 5, 7, 9)
高斯噪声	(0.02, 0.04, 0.06, 0.08, 0.1)
中值滤波	(1, 3, 5, 7, 9)
椒盐噪声	(0.02, 0.04, 0.06, 0.08, 0.1)

绿色箱体表示在COCO训练集上训练的StegaStamp模型，在不同攻击类型与攻击强度下分别在COCO测试集和SCID测试集上得到的AD；蓝色箱体表示在SCID训练集上训练的StegaStamp模型，在相同攻击条件下分别在SCID测试集和COCO测试集上得到的AD，横坐标中的BAR表示该攻击下两个模型的平均准确率。箱体在纵轴上的位置高度反映了模型在不同数据集上测试时的性能差异，位置越高表示跨数据集性能提升越明显；箱体长度反映了模型性能的稳定性，箱体越短说明模型在不同测试集上的性能波动越小。可以观察到，大多数攻击条件下训练的模型在跨数据集数字攻击测试中，其鲁棒性优于在自然图像数据集上训练的模型。图5中其余两组对比结果以及图6至图9的实验结果同样呈现出一致的趋势。即便在部分攻击条件下，蓝色箱体的位置未显著高于绿色箱体，其箱体长度仍明显短于绿色箱体。例如，在图7（c）中的JPEG压缩、随机裁剪和高斯滤波攻击场景下，尽管绿色箱体和蓝色箱体的AD均低于0，但蓝色箱体的长度明显更短，表明在SCID上训练的PIMoG模型在SCID测试集上的性能下降幅度显著小于在MIRFlickr上训练的模型。上述结果进一步说明，在SCID上训练的模型在测试于自然图像数据集时，能够有效保证水印提取性能的稳定性，避免出现大幅性能退化。实验结果表明，基于SCID训练得到的水印模型在应用于其他自然图像载体时，其水印鲁棒性仍可保持较高水平，从而验证了所构建数据集在数字攻击场景下的良好泛用性。

3.4 屏摄攻击实验对比分析

本小节通过屏摄攻击对嵌入水印后的鲁棒性进行对比实验。本节实验中设计了五种不同的屏摄环境，每种屏摄环境的设置说明如下：

1）不同光照条件（Illuminance）：保持相机的拍摄角度垂直于屏幕上显示的含水印图像，且与显

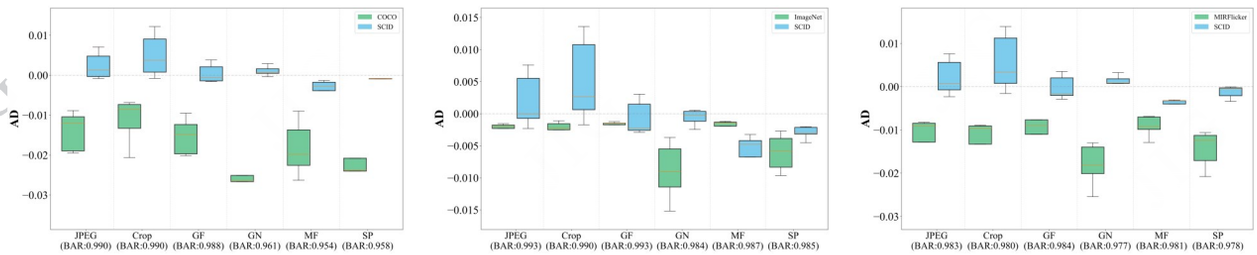
光照水平分别为50Lux、100Lux和150Lux进行拍摄；

2）不同拍摄角度条件（Angle）：保持相机的光学镜头与显示器之间的中心距离为30cm，拍摄角度分别设置为上下俯仰30°和15°、左右偏移30°和15°与0°偏转九种角度进行拍摄，其中Up为仰拍，Down为俯拍，Left为左偏移拍摄，Right为右偏移拍摄；

3）不同拍摄距离条件（Distance）：保持相机的拍摄角度垂直于显示器上的含水印图像，拍摄距离分别设置为20cm、30cm和40cm进行拍摄；

4）不同显示亮度条件（Brightness）：保持相机的拍摄角度垂直于屏幕上显示的含水印图像，且与显示器保持30cm的固定距离，分别选用45%、60%和75%的显示亮度进行拍摄；

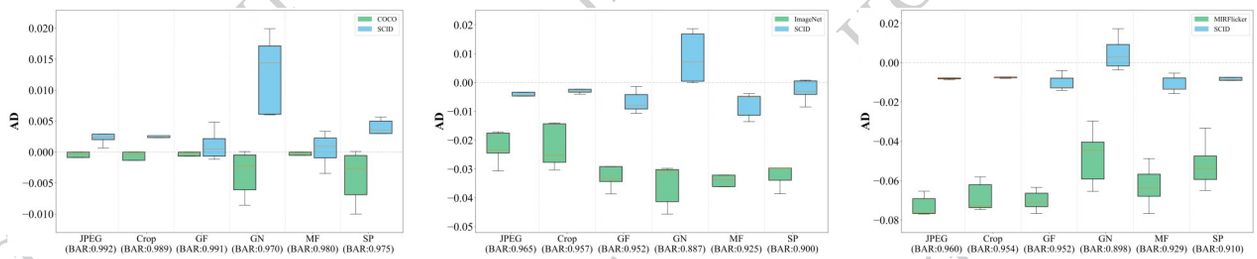
5）不同设备组合条件（Equipment_Group）：保持相机的拍摄角度垂直于屏幕上显示的含水印图像，且与显示器保持30cm的固定距离，选用HKC



（a） COCO-SCID （b） ImageNet-SCID （c） MIRFlicker-SCID

图5 StegaStamp在自然图像和SCID数据集上的数字攻击对比

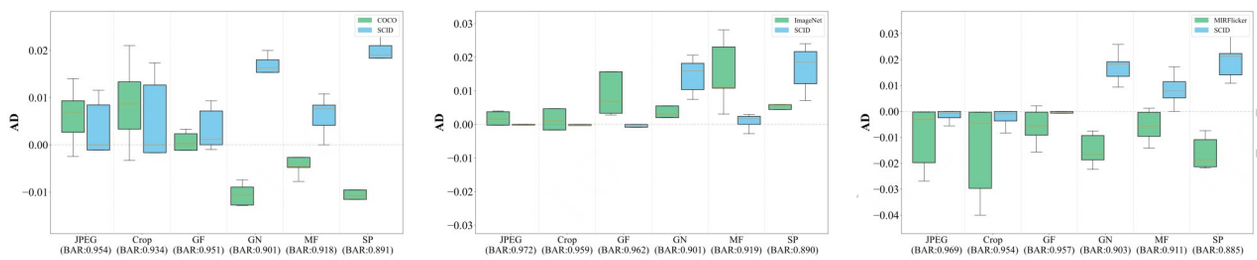
Fig. 5 Comparison of StegaStamp under digital attacks on natural image and SCID datasets



（a） COCO-SCID （b） ImageNet-SCID （c） MIRFlicker-SCID

图6 MBRS在自然图像和SCID数据集上的数字攻击对比

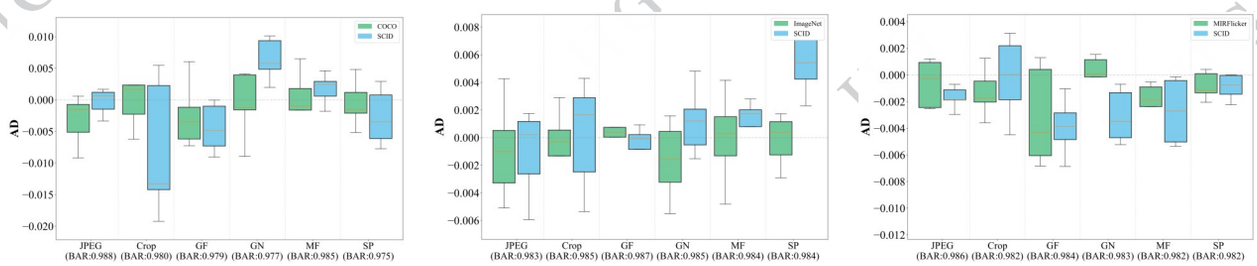
Fig. 6 Comparison of MBRS under digital attacks on natural image and SCID datasets



（a） COCO-SCID （b） ImageNet-SCID （c） MIRFlicker-SCID

图7 PIMoG在自然图像和SCID数据集上的数字攻击对比

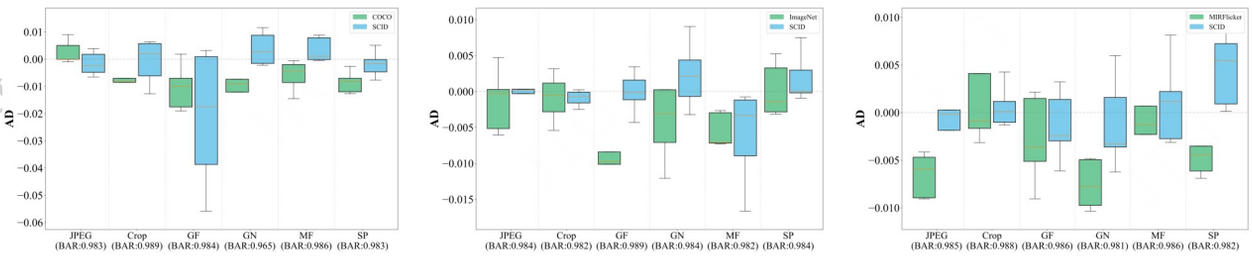
Fig. 7 Comparison of PIMoG under digital attacks on natural image and SCID datasets



（a） COCO-SCID （b） ImageNet-SCID （c） MIRFlicker-SCID

图8 HiFiMSFA在自然图像和SCID数据集上的数字攻击对比

Fig. 8 Comparison of HiFiMSFA under digital attacks on natural image and SCID datasets



（a） COCO-SCID　　（b） ImageNet-SCID　　（c） MIRFlicker-SCID

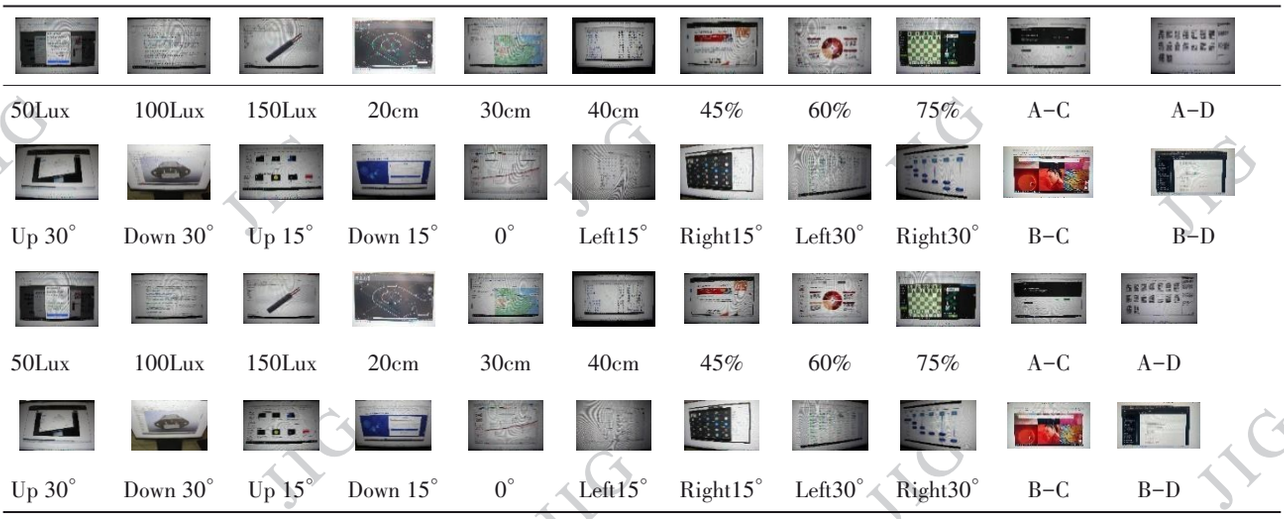
图9　MTVDGAN在自然图像和SCID数据集上的数字攻击对比

Fig. 9　Comparison of MTVDGAN under digital attacks on natural image and SCID datasets

图10　图11自然图像数据集训练模型在SCID测试集上水印嵌入后的不同屏摄结果

SCID训练模型在SCID测试集上水印嵌入后的不同屏摄结果

Fig. 10　Fig. 11Screen shooting results of watermarked SCID images under different conditionsScreen shooting results of watermarked SCID images under different conditions using the model trained on SCID

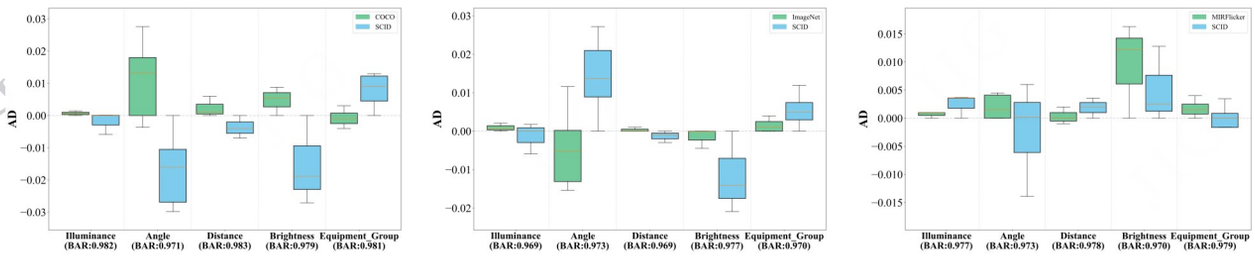


G24H1显示器（Monitor-A）和AOC Q24G2显示器（Monitor-B）作为图像显示设备，选用佳能EOS 60D（Camera-C）和iPhone 12 Pro Max（Camera-D）作为图像拍摄设备进行拍摄。

针对每组不同的训练集—测试集组合，各采集20张图像，同时为了模拟真实屏幕拍摄场景，我们将模型生成得到的残差图叠加在原始尺寸的图像上，再经定位与校正处理后进行水印提取。实验仍采用准确率差值（AD）作为评价指标。

图10和图11分别展示了不同条件下在自然图像数据集和在SCID上训练的模型对屏幕内容图像嵌入水印后的屏摄结果。可以看出，相较于在SCID上训练的模型，在自然图像数据集上训练的模型对屏幕内容图像嵌入水印后，实际屏摄得到的图像中呈现出更加明显的视觉伪影。

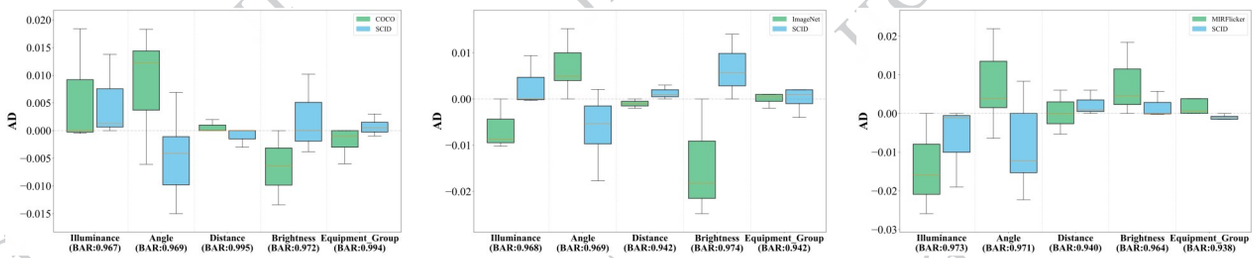
图12给出了StegaStamp模型的测试结果。其中，绿色箱体表示在COCO训练集上训练的StegaStamp模型，在不同的拍摄环境下分别在COCO测试集和SCID测试集上得到的AD；蓝色箱体表示在SCID训练集上训练的StegaStamp模型，在相同屏摄条件下分别在SCID测试集和COCO测试集上得到的AD，横坐标中的BAR表示该攻击下两个模型的平均准确率。可以观察到，绿色箱体和蓝色箱体在纵轴位置上整体持平，但大部分蓝色箱体的长度长于绿色箱体，表明在自然图像数据集上训练的模型在跨数据集屏摄攻击测试中，其鲁棒性与在SCID上训练的模型相比未表现出显著的优劣差异，且前者在跨数据集屏摄攻击测试中稳定性更佳。然而，从图14和图16可以观察到相反的趋势：蓝色箱体的长度大部分小于绿色箱体，表明在SCID上训练的模型在该跨数据集屏摄攻击测试时模型的稳定性比在自然图像数据集上训练的模型更好。



（a） COCO-SCID （b） ImageNet-SCID （c） MIRFlicker-SCID

图12 StegaStamp在自然图像和SCID数据集上的屏摄对比

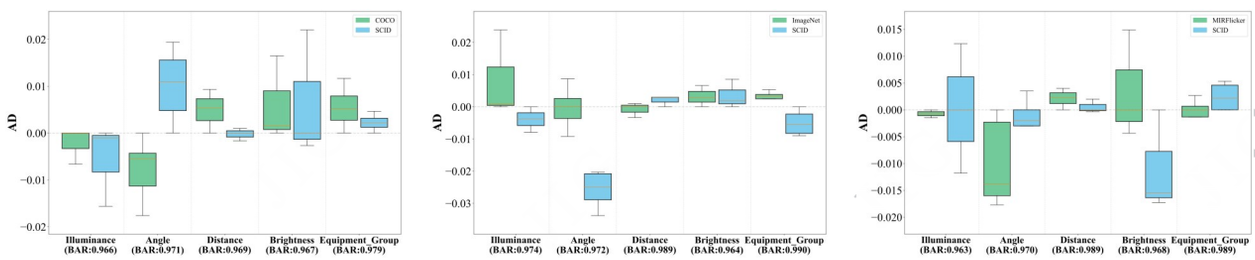
Fig. 12 Comparison of StegaStamp under Screen shooting on Natural Image and SCID Datasets



（a） COCO-SCID （b） ImageNet-SCID （c） MIRFlicker-SCID

图13 MBRS在自然图像和SCID数据集上的屏摄对比

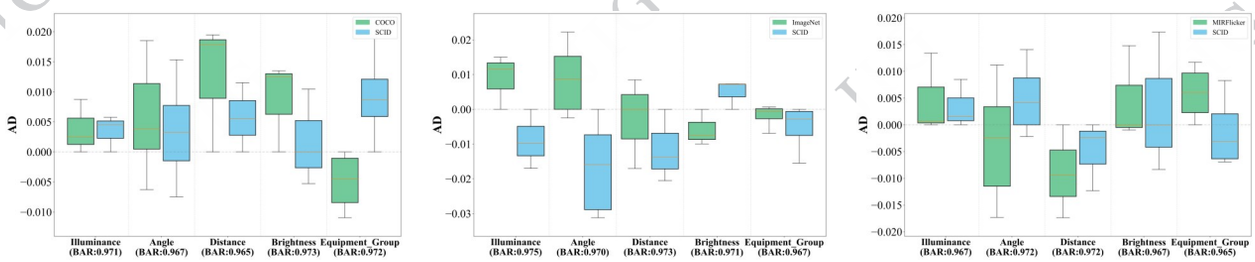
Fig. 13 Comparison of MBRS under Screen shooting on Natural Image and SCID Datasets



（a） COCO-SCID （b） ImageNet-SCID （c） MIRFlicker-SCID

图14 PIMoG在自然图像和SCID数据集上的屏摄对比

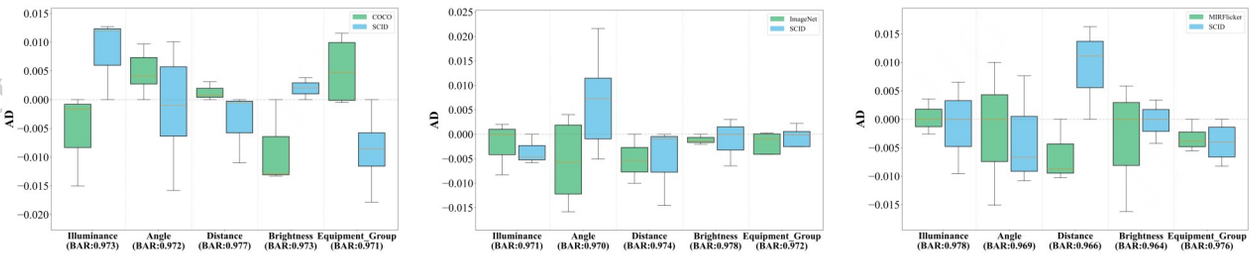
Fig. 14 Comparison of PIMoG under Screen shooting on Natural Image and SCID Datasets



（a） COCO-SCID （b） ImageNet-SCID （c） MIRFlicker-SCID

图15 HiFiMSFA在自然图像和SCID数据集上的屏摄对比

Fig. 15 Comparison of HiFiMSFA under Screen shooting on Natural Image and SCID Datasets



（a） COCO-SCID （b） ImageNet-SCID （c） MIRFlicker-SCID

图16 MTVDGAN在自然图像和SCID数据集上的屏摄对比

Fig. 16 Comparison of MTVDGAN under Screen shooting on Natural Image and SCID Datasets

表6 不同模型的含水印图像在SCID的不同类别上的PSNR和SSIM

Table 6 PSNR and SSIM of watermarked images from different models on different categories of SCID

模型	类别	PSNR (dB)	SSIM
StegaStamp (64 bits)	All	23.091	0.940
	C_1	23.066	0.942
	C_2	22.608	0.941
	C_3	22.912	0.938
MBRS (30 bits)	All	27.047	0.966
	C_1	28.579	0.983
	C_2	28.281	0.978
	C_3	28.251	0.982
PIMoG (30 bits)	All	24.739	0.840
	C_1	24.841	0.849
	C_2	24.104	0.841
	C_3	24.849	0.845
HiFiMSFA (64 bits)	All	26.871	0.937
	C_1	26.880	0.939
	C_2	25.894	0.931
	C_3	26.632	0.930
MTVDGAN (50 bits)	All	25.019	0.941
	C_1	24.943	0.936
	C_2	24.522	0.935
	C_3	24.593	0.940

根据上述实验结果，可以得到如下结论：StegaStamp和MBRS方法在编码器的设计上直接将水印信息和载体图像作为输入，未引入额外的图像处理或引导模块，使得模型在训练时仅需关注如何在整幅图像中嵌入水印信息，而无需考虑图像特定部位的嵌入效果。因此，在自然图像数据集上训练的模型在对屏幕内容图像进行水印嵌入时，虽然会引入更为显著的视觉伪影，但这些伪影在一定程度上有助于提取网络更准确地恢复水印信息。相比之下，在SCID上训练的模型受SCID的“像素集中化”特性影响，其嵌入水印时产生的视觉伪影较小，尤其在自然图像数据集上，水印信息更容易受到复杂纹理特征的干扰，从而降低提取网络的准确率。PIMoG方法在模型训练的水印嵌入阶段，引入BDCN网络（He等，2020）生成边缘掩码，引导模型将水印信息嵌入于图像的边缘区域。对四个数据集进行Canny边缘检测后统计可得，SCID的边缘密度为0.041 8，COCO的边缘密度为0.168 6，ImageNet的边缘密度为0.174 6，MIRFlicker的边缘密度为0.152 0。结果表明，SCID的边缘特征显著少于其他自然图像数据集。因此，在SCID上训练得到的PIMoG在对自然图像进行水印嵌入时，能够更充分地利用自然图像中丰富的边缘特征进行信息嵌入，同时不影响提取网络的性能。类似地，HiFiMSFA方法在模型设计中加入多尺度显著特征注意力模块，MTVDGAN方法在模型设计中加入Multi-Token-ViT（MTV）机制机制，两者均能够在训练过程中将水印嵌入在图像的纹理丰富区域，结合3.2节中的部分分析，相较于SCID，自然图像具有更丰富的纹理信息，故在SCID上训练得到的MTVDGAN在对自然图像进行水印嵌入时，同样能够更充分地利用自然图像中丰富的纹理区域进行信息嵌入，不影响提取网络的性能。

综上所述，在SCID上训练的模型对不同数据集进行水印嵌入后，经屏摄处理再进行水印提取，其提取准确率虽有所下降，但降幅均控制在0.1%以内。

结合第3.2节中对视觉质量的分析结果，相较于在自然图像数据集上训练的水印模型在对屏幕内容图像嵌入水印时出现的视觉质量下降问题，在SCID上训练的模型对自然图像进行水印嵌入能够较好地

表7　不同模型的含水印图像在SCID的不同类别上的数字攻击对比

Table 7　Comparison of Digital Attacks on Watermarked Images from different models across different categories of SCID

模型	类别	JPEG	Crop	GF	GN	MF	SP
StegaStamp (64 bits)	All	0.972	0.983	0.974	0.983	0.973	0.979
	C_1	0.984	0.977	0.975	0.970	0.976	0.975
	C_2	0.985	0.975	0.970	0.980	0.985	0.978
	C_3	0.965	0.983	0.971	0.970	0.965	0.973
MBRS (30 bits)	All	0.978	0.972	0.987	0.976	0.971	0.977
	C_1	0.977	0.972	0.973	0.960	0.975	0.973
	C_2	0.987	0.971	0.984	0.981	0.980	0.989
	C_3	0.976	0.989	0.975	0.976	0.962	0.971
PIMoG (30 bits)	All	0.980	0.986	0.974	0.982	0.984	0.973
	C_1	0.970	0.982	0.986	0.972	0.974	0.982
	C_2	0.979	0.980	0.987	0.975	0.969	0.982
	C_3	0.968	0.974	0.983	0.971	0.983	0.964
HiFiMSFA (64 bits)	All	0.980	0.970	0.984	0.969	0.98	0.963
	C_1	0.963	0.960	0.97	0.969	0.965	0.989
	C_2	0.977	0.966	0.976	0.971	0.972	0.972
	C_3	0.98	0.967	0.986	0.979	0.960	0.984
MTVDGAN (50 bits)	All	0.986	0.985	0.981	0.983	0.980	0.975
	C_1	0.982	0.973	0.961	0.974	0.987	0.980
	C_2	0.984	0.966	0.973	0.977	0.976	0.970
	C_3	0.977	0.972	0.985	0.980	0.974	0.971

保持原有视觉质量，表明其具有更好的泛化能力。同时，PIMoG、HiFiMSFA和MTVDGAN的对比实验结果进一步说明，基于边缘特征驱动与结合注意力模块或MTV机制的水印方法在SCID上训练，可获得视觉质量更高且鲁棒性更强的水印模型，为后续基于深度学习的抗屏摄鲁棒水印技术研究提供有益参考。

3.5　数据集的类别均衡性分析与分层评测

本小节结合视觉质量对比、数字攻击与屏摄攻击实验，对该数据集各图像类别开展系统性评测，通过对比模型在SCID不同类别上的视觉质量与鲁棒性实验结果，分析SCID各类别样本对模型训练效果的影响规律。本小节实验仍选用PSNR、SSIM以及比特准确率（BAR）作为评价指标。根据表2的各类别图像数量分布，将数据集分成头部类C_1（含网页类共10 303张），中部类C_2（含工程制图类和办公类，共4 663张）和尾部类C_3（含聊天类、编程类和线上会议类，共2 135张）三组数据，并整体（ALL）进行对比实验，详细分布如图17所示。在

视觉质量和数字攻击对比实验中，从每组样本中随机抽取100张图像开展测试；屏摄攻击实验则从每组中随机抽取20张图像，其余条件与3.3节和3.4节一致。

表6给出了在SCID上训练得到的五种模型的PSNR和SSIM测试结果。在StegaStamp的测试结果中，ALL、C_1、C_2和C_3分组的PSNR依次为23.091dB、23.006dB、22.608dB和22.912dB，SSIM依次为0.940、0.942、0.941和0.938。可以观察到，模型在

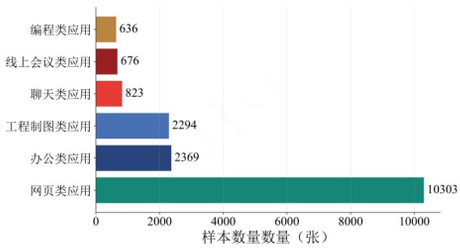


图 17　不同应用类别样本数量分布

Figure. 17　Distribution of sample sizes across different application categories

四个类别上的PSNR和SSIM无显著波动，

表明模型在这四组类别图像的含水印图像能够保持视觉质量稳定。此外，表6中其余四种模型的测试结果与StegaStamp的表现一致。

表7、表8分别为不同模型在SCID各类别上的数字攻击、屏摄攻击对比实验测试结果（各攻击类型下均取平均比特准确率）。可以观察到，各模型在四组样本上的水印提取准确率均保持稳定，表明模型在SCID不同类别图像上生成的含水印图像可维持稳定的比特准确率。

根据上述实验结果可得出结论：尽管SCID中六个类别的样本数量存在显著差异，但水印模型在SCID上训练时，未因数据集内部的样本分布差异产生训练异化现象；即在SCID上完成训练的模型，在该数据集各类别图像上均能保持稳定的视觉质量与比特准确率。

4　结　论

目前，面向屏幕内容图像的公开标准数据集仍较为匮乏，现有多数基于深度学习的水印方法主要依赖自然图像数据集进行训练，导致所构建模型在屏幕内容图像场景下难以维持稳定性能。因此，如

表8　不同模型的含水印图像在SCID的不同类别上的屏摄对比

Table 8　Comparison of Screen shooting on Watermarked Images from different models across different categories of SCID

模型	类别	Illuminance	Distance	Angle	Brightness	Equipment_Group
StegaStamp (64 bits)	All	0.982	0.973	0.975	0.969	0.979
	C_1	0.980	0.984	0.974	0.970	0.974
	C_2	0.978	0.963	0.985	0.961	0.960
	C_3	0.984	0.975	0.978	0.975	0.964
MBRS (30 bits)	All	0.990	0.989	0.985	0.970	0.984
	C_1	0.971	0.985	0.980	0.983	0.972
	C_2	0.970	0.988	0.979	0.974	0.977
	C_3	0.981	0.980	0.984	0.969	0.973
PIMoG (30 bits)	All	0.973	0.984	0.981	0.967	0.976
	C_1	0.964	0.984	0.976	0.961	0.974
	C_2	0.980	0.971	0.969	0.970	0.990
	C_3	0.987	0.976	0.961	0.980	0.989
HiFiMSFA (64 bits)	All	0.966	0.973	0.982	0.980	0.962
	C_1	0.983	0.982	0.988	0.968	0.960
	C_2	0.961	0.981	0.969	0.974	0.989
	C_3	0.971	0.977	0.963	0.964	0.972
MTVDGAN (50 bits)	All	0.987	0.990	0.978	0.974	0.975
	C_1	0.975	0.972	0.977	0.989	0.976
	C_2	0.965	0.963	0.982	0.961	0.979
	C_3	0.984	0.975	0.977	0.975	0.988

何实现对屏幕内容图像的高质量水印嵌入与稳定提取,已成为亟需解决的重要问题。针对上述问题,本文构建并发布了一个大规模屏幕内容图像数据集,涵盖网页类10 303张、聊天对话框类823张、编程环境类636张、工程制图类2 294张、线上会议类676张以及应用界面类2 369张,共计17 101张图像。该数据集在内容多样性与样本规模上均具备显著优势,为神经网络模型的有效训练提供了充分支撑。大量实验结果表明,相较于在自然图像数据集上训练得到的抗屏摄鲁棒水印模型,基于所构建数据集训练的模型在屏幕内容图像场景下能够实现更优的视觉质量与更稳定的水印嵌入与提取性能,同时展现出良好的泛化能力。本文的研究工作为屏幕内容图像的安全保护提供了有力支撑,有助于推动抗屏摄鲁棒水印技术在屏幕内容保护领域的进一步发展与实际应用。

参考文献(References)

Chang C and Shen J. 2017. Features classification forest: a novel development that is adaptable to robust blind water-marking techniques. IEEE Transactions on Image Process- ing: 26(8): 3921-3935 [DOI: 10.1109/TIP.2017.2706502]

Chen M, Liao X, Fang H, Guo J, Chen Y and Wu X. 2025. Flexible Partial Screen-shooting Watermarking with Prova-ble Robustness. IEEE Transactions on Circuits and Syste-ms for Video Technology: IEEE: 1-1 [DOI: 10.1109/TCSVT.2025.3585739]

Deng J, Dong W, Socher R, Li L, Li K and Li F F. 2009. Imagenet: A large-scale hierarchical image database. IEEE conference on computer vision and pattern recogni- tion. IEEE: 248-255 [DOI: 10.1109/CVPR.2009.5206848]

Fang H, Jia Z, Ma Z, Chang E and Zhang W. 2022a. PIMoG:An effective screen-shooting noise-layer simulation for de-ep-learning-based watermarking network. Proceedings of the 30th ACM international conference on multimedia: ACM MM: 2267-2275 [DOI: 10.1145/3503161.3548049]

Fang H, Jia Z, Qiu Y, Zhang J, Zhang W and Chang E C. 2022b. De-END: Decoder-driven watermarking network. IEEE Transactions on Multimedia: IEEE TMM: 25: 7571-7581 [DOI: 10.1109/TMM.2022.3223559]

Fang H, Qiu Y, Chen K, Zhang J, Zhang W and Chang E C. 2023. Flow-based robust watermarking with invertible noi-se layer for black-box distortions. Proceedings of the AAAI conference on artificial intelligence: AAAI: 37(4): 5054-5061 [DOI: 0.1609/aaai.v37i4.25633]

Fares K, Khaldi A, Redouane K and Salah E. 2020. DCT & DWT based watermarking scheme for medical informationsecurity. Biomedical Signal Processing and Control: 66:102403 [DOI: 10.1016/j.bspc.2020.102403]

Fikri C, Nugroho F A, Apriyansyah B ans Fakhreldin M. 2025. Dual watermarking based on human visual characteristi-cs with IWT-SVD. IJACI: International Journal of Advan-ced Computing and Informatics: 1(1): 1-12 [DOI: 10.71129/ijaci.v1.i1.pp1-12]

Gao G, Ding Y, Feng T, Fu Z and Shi Y. 2025. MTVDGAN: Multi-Token-ViT Dense GAN for Robust Screen-Shooting Watermarking [J]. IEEE Transactions on Information Foren-sics and Security: 11804-11815 [DOI: 10.1109/TIFS.2025.3626146.]

Gen L C, Avivah S N and Muzahid A J M. 2025. Image wa-termarking for ensuring image integrity and robust copy- right protection based on discrete wavelet transform. IJACI: International Journal of Advanced Computing and Infor-matics: 1(1): 28-38 [DOI: 10.71129/ijaci.v1.i1.pp28-38]

He J, Zhang S, Yang M, Shan Y and Huang T. 2020. BDCN: Bi-Directional Cascade Network for Perceptual Ed-ge Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence: PP(99): 1-1 [DOI: 10.1109/TPAMI.2020.3007074]

He M, Feng B, Guo Y, Weng J and Lu W. 2024. Camera-sh-ooting resilient watermarking on image instance level. IEEE Transactions on Circuits and Systems for Video Tech- nology, 34(11), 10874-10887 [DOI:10.1109/TCSVT.2024.3411816]

Huiskes M J and Lew M S. 2008. The MIR-flickr retrieval evaluation. Proceedings of the 1st ACM international con-ference on Multimedia information retrieval: 39-43 [DOI: 10.1145/1460096.1460104]

Jia Z, Fang H and Zhang W. 2021. Mbrs: Enhancing robust- ness of dnn-based watermarking by mini-batch of real andsimulated jpeg compression. Proceedings of the 29th ACMinternational conference on multimedia: ACM MM: 41-49 [DOI: 10.1145/3474085.3475324]

Li L, Bai R, Zhang S, Chang CC and Shi M. 2021. Screen- Shooting Resilient Watermarking Scheme via Learned In- variant Keypoints and QT. Sensors: 21(19): 6554 [DOI: 10.3390/s21196554]

Li X, Guo D, Zhuo X, Yao H and Qin C. 2023. Carrier-inde-pendent screen-shooting resistant watermarking based on information overlay superimposition. Chinese Journal of Network and Information Security: 9(3): 135-149 (李晓萌, 郭玳豆, 卓训方, 姚恒, 秦川. 2023. 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, 9(3): 135-149) [DOI: 10.11959/j.issn.2096-109x.2023045]

Li Y, Liao X and Wu X. 2024. Screen-shooting resistant wa- termarking with grayscale deviation simulation. IEEE Tra-nsactions on Multimedia, 26: 10908-10923 [DOI: 10.1109/TMM.2024.3415415]

Lin T Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, et al. 2014. Microsoft coco: Common objects in con- text. European conference on computer vision: LNIP: 740-755 [DOI: 10.1007/978-3-

319-10602-1_48]

Mahto D K, Singh A K, Singh K N, Singh O P and Agrawal A K. 2024. Robust copyright protection technique with high-embedding capacity for color images. ACM Transac- tions on Multimedia Computing, Communications and Applications: 20(11): 1-12 [DOI: 10.1145/3580502]

Onn N C, Avivah S N, Alam M M and Ahmad A. 2025. Se-cure and Effective Image Integrity and Copyright Protec- tion Using Two-Layer Authentication with Integer WaveletTransform. IJACI: International Journal of Advanced Com-puting and Informatics: 1(1): 13-27 [DOI: 10.71129/ijaci.v1.i1.pp13-27]

Qian K, Lu Y, Yang Z, Zhang K, Huang K, Cai X, et al. 2021. AIRCODE: Hidden Screen-Camera Communication on an Invisible and Inaudible Dual Channel. 18th USENIX Symposium on Networked Systems Design and Implemen- tation: NSDI 21: 457-470 [EB/OL] https://www.usenix.org/system/files/nsdi21-qian-kun.pdf

Sun N, Fang H, Lu Y, Zhao C and Ling H. 2025. END^2: Robust Dual-Decoder Watermarking Framework Against Non-Differentiable Distortions. Proceedings of the AAAI Conference on Artificial Intelligence: 39(1): 773-781 [DOI: 10.1609/aaai.v39i1.32060]

Tancik M, Mildenhall B and Ng R. 2020. Stegastamp: Invisi- ble hyperlinks in physical photographs. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition: CVPR: 2117-2126 [DOI: 10.48550/arXiv.1904.05343]

Tian H, Zhao Y, Ni R and Qin L. 2013. LDFT-Based Water-marking Resilient to Local Desynchronization Attacks. IEEE Trans Cybern: 2190-2201 [DOI: 10.1109/TCYB.2013.2245415]

Wang J, Kang X, Li W, Geng J, Miao Y and Chen Y. 2025. Toward imperceptible and robust image watermarking aga-inst screen-shooting with dense blocks and CBAM. App- lied Intelligence, 55(7): 645 [DOI: 10.1007/s10489-025-06496-0]

Wengrowski E and Dana K. 2019. Light field messaging with deep photographic steganography. Proceedings of the IEEE/CVF conference on computer vision and pattern recogni- tion: CVPR: 1515-1524 [DOI: 10.1109/CVPR.2019.00161]

Wen W, Hu J, Xiao X, Qiu B and Zhang Y. 2025. RSIRW- DCF: Dual-Channel Feature-Guided Robust Watermarking Against Dual Attacks for Remote Sensing Image. IEEE Transactions on Consumer Electronics: 1-1 [DOI: 10.1109/TCE.2025.3598821.]

Wu J, Li X and Qin C. 2023. Screen-shooting robust water- marking with end-to-end neural network. Journal of Imageand Graphics: 28(12):3713-3730 (吴嘉奕, 李晓萌, 秦川. 2023. 面向屏幕拍摄的端到端鲁棒图像水印算法. 中国图象图形学报, 28(12): 3713-3730) [DOI: 10.11834/jig.221141]

Zhang C, Benz P, Karjauv A, Sun G and Kweon I. 2020. Udh: Universal deep hiding for steganography, watermark- ing, and light field messaging. Advances in Neural Infor- mation Processing Systems: 33: 10223-10234 [DOI: 10.5555/3495724.3496581]

Zhang Y, Ni J and Su W. 2025. HiFiMSFA: robust and high-Fidelity image watermarking using attention augmented de-ep network. IEEE Signal Processing Letters: 781-785 [DOI: 10.1109/LSP.2025.3535216]

Zheng G, Hu D, Ge H and Zheng S. 2021. End-to-end image steganography and watermarking driven by generative ad- versarial networks. Journal of Image and Graphics, 26(10): 2485-2502 (郑钢, 胡东辉, 戈辉, 郑淑丽. 2021. 生成对抗网络驱动的图像隐写与水印模型. 中国图象图形学报, 26(10): 2485-2502) [DOI: 10.11834/jig.200404]

Zhu J, Kaplan R, Johnson J and Li F F. 2018. Hidden: Hid- ing data with deep networks. Proceedings of the European Conference on Computer Vision: ECCV: 657-672 [DOI: 10.48550/arXiv.1807.09937]

作者简介

陈尧一,男,硕士研究生,主要研究方向为抗屏摄鲁棒水印技术。E-mail:232210442@st.usst.edu.cn

秦川,通信作者,男,教授,主要研究方向为多媒体信息安全、人工智能安全和数字图像处理。 E-mail:qin@usst.edu.cn

王娜,女,副教授,主要研究方向为多媒体信息安全和数据压缩。E-mail:wna@usst.edu.cn

彭彦淳,男,本科生,主要研究方向为抗屏摄鲁棒水印技术和图像水印攻击。 E-mail:2335061119@st. usst.edu.cn

陈嘉豪,男,本科生,主要研究方向为抗屏摄鲁棒水印技术。E-mail:2435055208@st.usst.edu.cn

乔蓬旭,女,本科生,主要研究方向为抗屏摄鲁棒水印技术。E-mail:2435053604@st.usst.edu.cn

王伟,男,教授,主要研究方向为医学图像处理、人工智能以及其在医学大数据中的应用。 E-mail:wwangfd@fudan.edu.cn