基于深度学习的图像信息隐藏方法综述

张汝波[1]，常世淇[1]，张天一[2]

（1. 大连民族大学 机电工程学院，辽宁 大连 116600；2. 北京航空航天大学 网络空间安全学院，北京 100191）

摘　要：基于图像的信息隐藏技术可以实现将信息隐蔽地藏于图像内容中，从而在图片的传输过程中实现保密通信、版权认证等信息安全保护行为，是目前信息安全领域研究的热点之一。本文首先论述了基于深度学习的图像信息隐藏方法的重难点问题；其次，从结构特点、训练特点和应用特点3个角度对基于深度学习的图像隐写方法进行归纳；再次，介绍了领域相关主要数据集和评估指标；然后，总结了图像信息隐藏技术的应用情况；最后，讨论了图像信息隐藏技术的研究方向，为该领域的进一步发展提供见解和建议。

关键词：信息处理技术；信息隐藏；深度学习；图像隐藏；隐写术；数字水印

中图分类号：TP309.7　　**文献标志码**：A　　**文章编号**：1671-5497(2025)05-1497-19

DOI：10.13229/j.cnki.jdxbgxb.20240381

Review on image information hiding methods based on deep learning

ZHANG Ru-bo[1], CHANG Shi-qi[1], ZHANG Tian-yi[2]

(1. *College of Mechanical & Electronic Engineering, Dalian Minzu University, Dalian* 116600, *China*; 2. *School of Cyber Science and Technology, Beihang University, Beijing* 100191, *China*)

Abstract: Image information hiding technology can achieve the goals of confidential communication, copyright authentication and other information security protection behaviors during the transmission of pictures by hiding information covertly in images, which is one of the hotspots of current research in the field of information security. Firstly, we discuss the important and difficult problems of image information hiding methods based on deep learning. Secondly, we discuss the deep learning-based image steganography method from three perspectives: structural features, training features and application features. Then, we introduce the main datasets and evaluation criteria related to the domain and summarize the experimental performance. Next, this paper summarizes the applications of image information hiding techniques. Finally, we discuss the research directions of image information hiding techniques to provide insights and suggestions for the further developments in the field.

Key words: information processing technology; information hiding; deep learning; image information hiding; steganography; digital watermarking

---

收稿日期：2024-04-11.

基金项目：国家自然科学基金项目(62202024)；中央高校基本科研业务费专项项目.

作者简介：张汝波(1963-)，男，教授，博士. 研究方向：智能感知与先进控制. E-mail：zhangrubo@dlnu.edu.cn

通信作者：张天一(1989-)，女，讲师，博士. 研究方向：计算机视觉. E-mail：zhang_tianyi@buaa.edu.cn

0 引 言

图像信息隐藏技术(Image information hiding)又称图像隐写,是一种将秘密信息不可感知地嵌入到数字图像中,必要时再将其从含密的图像中恢复并提取出来的技术[1]。这项技术通常用于隐蔽地传递保密信息,同时也为数字图像版权标识与隐私保护提供技术支持,是信息保护的有力工具。

图像信息隐藏领域中有两个基本概念需要明确。一是"载体图像(Cover image)",指在研究中被嵌入秘密信息的原始图像。二是"隐写图像(Stego image)",指经过编码操作后生成的含密图像。在对图像信息隐藏的性能进行考察时,通常会对不可知性、容量、鲁棒性和安全性提出相应的要求[2]。

传统的图像隐藏方法主要是基于空间域的隐藏方法[3-6]和基于频域的隐藏方法[7-9]。空间域方法通过修改载体图像中的像素值隐藏信息,其中最低有效位方法(Least significant bits,LSB)[10]是最传统的空域隐藏方法。由于图像的光滑区域常出现伪影,隐写分析方法可以很容易地检测到LSB隐藏的秘密信息[11-14]。此外,空间域方法的鲁棒性较差,难抵抗常见的图像攻击,如裁剪、噪声、JPEG压缩[15]等。为提高鲁棒性,研究者们将研究方向转向频域,提出了基于离散余弦变换(DCT)[16]、离散小波变换(DWT)[17]和离散傅立叶变换(DFT)[18]等变换的信息隐藏方法。与空间域隐藏算法相比,频域算法相对更鲁棒和不可检测,且隐写图像质量更高,但有效载荷能力有限,逐渐无法满足容量需求。随着可用数据的大幅度增加,使用深度学习(Deep learning)已经成为一种趋势,并在许多复杂任务处理中被广泛应用。它在图像识别[19]、目标检测[20]、自然语言处理[21]、语音识别[22]、推荐系统[23]、金融领域[24]和医学图像处理[25,26]等领域的应用中都是有力的工具。图像信息隐藏技术同样受益于深度学习方法。现在已经有多种结构的深度学习网络在图像信息隐藏和隐写分析中得到了广泛应用,这些基于深度学习的图像信息隐藏方法实现了秘密信息的成功隐藏且性能较好。研究人员们不断将网络进行变体与优化,以期在隐藏容量、不可知性、鲁棒性和安全性等方面取得进一步提升[27]。本文围绕图像信息隐藏领域近年来深度学习方法的实现及应用情况展开论述。介绍了常用的数据集和主要评价指标,对多种方法进行性能对比,并讨论了需要进一步研究的问题。

1 任务目标及重难点

1.1 图像信息隐藏任务目标

保证秘密信息不被察觉是图像信息隐藏任务的重要目标。该任务不可知性的评价有两个角度:(1)人眼视觉上的不可感知;(2)统计上的低差异。前者是指嵌入秘密信息后的隐写图像与载体图像在人眼视觉上没有任何差别,图像原本的质量与价值不受影响。后者指在一些评价指标的约束下,隐写图像和载体图像之间的差异足够小,直至达到统计标准[28]。值得注意的是,统计失真度量与人类视觉评价并非完全一致[29]。若研究仅通过简单的客观度量去计算图像的失真,可能会造成失真结果的误判,因此需要注意主客观不可知性的权衡。此外,隐写分析技术(Steganalysis)[30-32]的广泛应用和迅速发展对信息隐藏方法的安全性提出了挑战。因此,如何提出具有良好抗隐写分析能力的隐写方法,成为图像信息隐藏研究的新兴目标。信息隐藏技术和隐写分析技术之间互相抗衡、共同促进。许多隐写方法通过结合先进的隐写分析工具进行抗隐写分析能力训练以获得具有更高安全性的隐写模型。

图像信息隐藏技术主要有两个应用类型:隐写术(Steganography)[33]和数字水印(Digital watermarking)[34]。二者有两个主要的不同之处:首先,在嵌入内容上,数字水印通常嵌入的信息与要保护的载体相关联(如版权所属、个人信息),而隐写术则可以隐藏任何不被察觉的需要信息;其次,在应用场景及性能侧重方面,隐写术通常用于进行保密传输[35],被广泛应用于医学、军事等涉及保密通信的场景,因此它更加注重隐蔽性和容量。数字水印常被用来证明图像所有权,一般用于版权鉴定[36,37]、图像保护[38,39]领域。因此,数字水印优先考虑鲁棒性而不是保密性。针对不同应用类型,方法的设计需要进行不同的性能侧重。

1.2 图像信息隐藏的重难点

1.2.1 隐藏容量有局限

在图像中隐藏更多的信息必然会产生更大的图像失真,从而导致图像质量不佳(如伪影或颜色失真)。在数字水印的应用中,版权作品的视觉质

量受到严重影响后会使图像失去自身价值;在保密传输领域,异常的含密图像易遭到第三方怀疑,加剧秘密泄露的风险。针对此问题,很多信息隐藏方法选择把信息量设置为阈值或者较低值以保持图像质量。秘密信息容量的这种局限导致这些方法的实际应用性较差。如何在保证图像质量的情况下提高信息容量,是目前研究的难点之一。

1.2.2　生成式GAN训练复杂度高

生成对抗网络(Generative adversarial network,GAN)是图像信息隐藏的常用技术。GAN的一个核心挑战就是训练过程的不稳定性,生成器和判别器之间的对抗性训练可能会导致模型出现难以收敛的情况。尤其在处理图像信息隐藏这种复杂的任务时,这种不稳定性可能会导致生成的隐写图像质量不一或信息有效载荷保留不足[40]。如何基于GAN进行针对图像隐写任务的合理改进和损失设计,寻找更加稳定的GAN变体[41-43]处理图像信息隐藏任务,也是目前存在的难点之一。

1.2.3　鲁棒性难以提升

网络传输信道的有损处理和图片保存所受到的压缩情况不可避免[44,45],若想要获得有效实用的图像隐写模型,鲁棒性的提升至关重要。目前研究中存在的难点是无法对图像受到的各种噪声实现完全模拟,甚至面临着网络媒体信道噪声未知的情况[46]。若模型在训练中缺乏这种噪声知识,就很难保证秘密信息以较好的准确率被恢复出来。这也是目前众多方法需要深入研究的问题[47]。

2　基于深度学习的图像信息隐藏方法

本文整体上从以下3个角度对近年基于深度学习的图像信息隐藏方法进行介绍,分别为结构特点、训练特点和应用特点。随后进行自整体至细节的深入,具体逻辑框架如图1所示。

2.1　基线模型和结构

本节中,对深度学习图像信息隐藏方法的模型和结构进行归纳。首先,从总体上介绍了方法的整体结构与流程特点。其次,对有性能提升效果的附加模块进行更细分的介绍。

2.1.1　整体网络架构

分阶段解码-编码结构。多数使用深度卷积

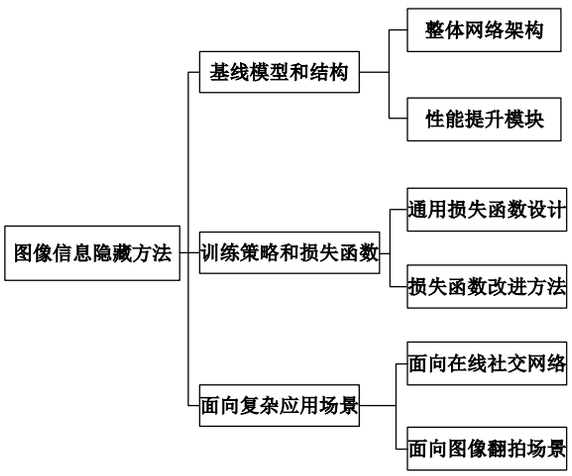

图1　图像信息隐藏方法归纳

Fig. 1　Generalization of image information hiding methods

神经网络的信息隐藏方法很大程度上受到编码器-解码器结构的启发。这类方法将秘密隐藏与恢复作为两个分离的阶段处理,编码网络将秘密信息嵌入载体图像生成隐写图像,解码网络从隐写图像中提取并恢复秘密信息。这种分阶段的解码器-编码器架构的隐写方法流程如图2所示。

Zhu等[44]提出了典型的解码器-编码器网络HiDDeN(Hiding data with deep networks)联合训练解码器和编码器,并使用判别器预测给定图像是否包含编码信息。Rahim等[48]使用损失约束编码器-解码器的联合端到端训练,完成了载体为三通道RGB图像,秘密信息的形式为灰度图或RGB的单通道图的高载荷信息隐藏,实现了在图像中隐藏图像的突破。Zhang等[41]提出了一种使

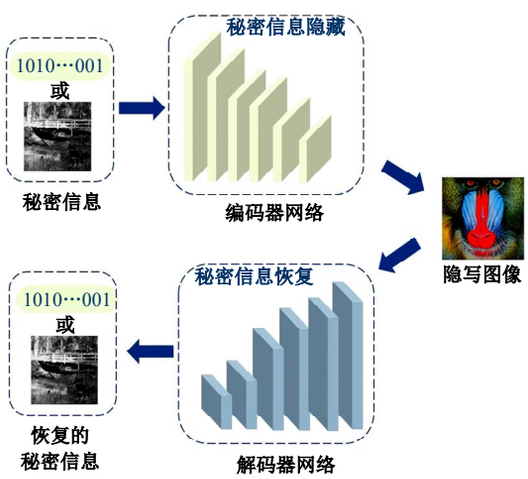

图2　分阶段的图像隐写方法流程

Fig. 2　Process of staged image information hiding method

用GAN在图像中隐藏任意二进制数据的技术SteganoGAN(High capacity image steganography with GANs)。它通过编码和解码两个操作完成在图像中的数据隐藏,通过计算解码正确率、隐写图像与载体图像之间的相似性和隐写图像的真实性,实现对编码器-解码器网络的迭代优化。Kishore等[49]为解决解码器-编码器网络嵌入容量大但解码错误率高的问题,采用SteganoGAN的解码器与编码器结构并基于神经网络对输入的微小扰动的高度敏感性,提出了一种新的隐写方法FNNS(Fixed neural network steganography),实现了更低的解码错误率。此外,Singh等[42]认为很多基于GAN的方法强调不可感知和提取精度而忽视了统计上的不可检出性,提出了双参与者框架StegGAN(Hiding image within image using conditional generative adversarial networks)。具体而言,StegGAN的嵌入网络和提取网络各自由两个子网络组成,即生成器和判别器。其中,嵌入网络的判别器设置为隐写分析网络XuNet(Structural design of convolutional neural networks for steganalysis)[30],通过将图像区分为载体图像或隐写图像与生成器进行博弈并实现纳什均衡,嵌入网络最终可以生成隐写分析器难以区分且高质量的隐写图像;提取网络的生成器从隐写图像中提取出秘密信息,对应的判别器则对真实秘密信息和提取出的秘密信息进行最大限度区分以形成对抗。最终,提取网络可以更准确地从隐写图像中估计出秘密信息。

可逆网络结构。可逆神经网络(Invertible neural networks,INN)最早由Dinh等[50]在2014年提出:给定一个变量y和正向计算$x=f_\theta(y)$,可以通过$y=f_\theta^{-1}(x)$直接恢复y,其中反函数f_θ^{-1}被设计为与f_θ共享相同的参数θ。与试图直接解决模糊逆问题的经典神经网络不同,INN专注于学习正向过程,并使用额外的潜在输出变量捕获丢失的信息[51]。随着INN在多个视觉任务(如图像缩放、图像着色、视频时间动作定位等)中的研究日益深入并展现出良好性能[52],研究者们开始将INN应用于图像信息隐藏领域。基于可逆网络的隐写方法的主要创新之处在于将秘密信息恢复建模为可逆网络结构中秘密隐藏的反向过程,仅训练一次,网络就可以学习到隐藏和恢复所需的所有参数。这种方法不仅提升了对网络的训练效率,而且避免了分阶段隐写方法中隐藏网络和恢复网络之间的松散连接可能导致的颜色失真和纹理复制伪影等安全性问题[53]。这种基于可逆网络的隐写方法流程如图3所示。

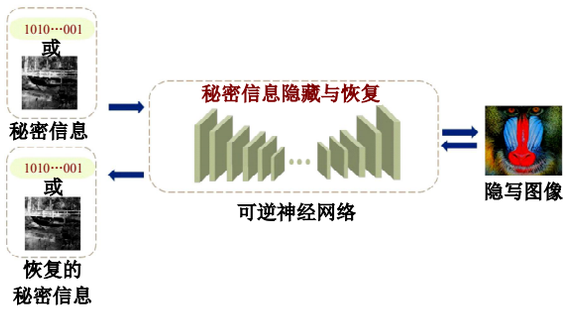

图3 基于可逆网络的图像隐写方法流程

Fig. 3 Process of image information hiding method based on reversible network

2021年,Jing等[53]和Lu等[54]首次将INN应用于图像信息隐藏领域。Jing等[53]提出的基于INN的框架HiNet(Deep image hiding by invertible network)能够实现将全尺寸的秘密图像隐藏到相同大小的载体图像中。HiNet采用DWT(Discrete wavelet transformation)将图像分割成低频和高频子带后再输入可逆模块。同年,Lu等[54]提出的大容量可逆隐写网络(Invertible steganography network,ISN)通过增加隐藏图像的分支通道数将多个秘密图像隐藏到一个载体图像中,显著提高了隐藏容量。但在隐写图像轻微失真的情况下,ISN恢复出的秘密信息和载体图像效果很差。针对ISN和其他基于可逆结构的隐写方法普遍存在的鲁棒性差的问题,Xu等[55]提出的基于流的鲁棒可逆图像隐写框架(Robust invertible image steganography,RIIS)在各种失真下表现出令人满意的鲁棒性。2023年,Guan等[56]开发了一种可逆隐藏神经网络IHNN(Invertible hiding neural network,IHNN),并提出了可逆隐藏框架DeepMIH(Deep invertible network for multiple image hiding)。DeepMIH将INN的想法融入多图像隐藏任务中,以一种新的方式将多个秘密图像隐藏到同一载体图像中,实现了图像的大容量"串联"隐藏。Yang等[57]在HiNet基础上提出了PRIS(Practical robust invertible network for image steganography)。PRIS在图像恢复过程前后分别设置了预增强块和后增强块,目的是减少隐写图像恢复前所受扰动及增强恢复后秘密图像的质量。同时,增强模块还可以削弱可逆网络的严

格可逆性，使其能更好地抵抗噪声。此外，该方法采取三步训练策略以避免增强模块的存在对可逆性的大幅度破坏，在严格的不可逆性和可逆性之间找到平衡。Mou 等[58]提出大容量灵活视频隐写网络（Large-capacity and flexible video steganography network，LF-VSN），该方法通过 INN 实现多个视频的隐写和恢复，可以在一个载体视频中隐藏/恢复 7 个秘密视频。

2.1.2　性能提升模块

前文对方法的整体结构和流程特点进行了介绍，本节将对在整体流程上用于性能提升的附加模块进行进一步介绍。

（1）人为模拟噪声干扰模块。在训练过程中通过人为添加模拟噪声干扰促进模型鲁棒性的提升是很常见的方法。考虑到隐写图像传输过程中受到的噪声干扰，Zhu 等[44]提出可抵抗噪声攻击的模型 HiDDeN，在编码器与解码器之间加入噪声层实现多种不同干扰类型的噪声模拟，包含 Dropout、Cropout、高斯噪声、JPEG 压缩的可微近似等。该方法将隐写图像经由噪声层的模拟攻击生成的含噪隐写图像作为解码器的输入，进行秘密信息提取。这样，即使隐写图像受到一些常见噪声攻击，解码器仍能以高精度恢复秘密信息，同时也证明了可微近似训练可以有效用于鲁棒性模型的训练。Bui 等[59]提出了一种使用自编码器的轻量鲁棒隐写方法（Robust steganography using autoencoder latent space，RoSteALS），在图像编码器和秘密解码器之间插入噪声模型，其中包含 3 种噪声类型：可微加性和线性噪声（亮度、饱和度，对比度）、近似可微噪声（JPEG 压缩）、不可微噪声（spatter、飞溅）。解码器可以在各种数据增强下训练，通过反向传播更新编码器。进一步地，Liu 等[60]指出在以 HiDDeN 为例的这种典型编码器-噪声层-解码器结构的单阶段端到端盲水印架构（The one-stage end-to-end training，OET）中，噪声层中的噪声攻击都是以可微方式进行模拟，这在实践中并不适用。面对一种新的噪声时，OET 并不能很好地应对。此外，在遭受噪声攻击情况下，OET 通常会出现收敛速度慢、隐写图像质量下降的问题。针对这一系列问题，他们提出一种两阶段可分离框架（Two-stage separable deep learning，TSDL）。第一阶段，编码器在没有任何噪声知识的情况下将信息冗余地嵌入载体图像中。第二阶段，解码器在训练时不更新编码器参数，而是在第一阶段基础上根据不同噪声对解码器进行微调。训练期间使用了从 COCO 数据集中随机选择并用图像批处理软件合成的含 5 种攻击类型的黑盒噪声数据集，实现对噪声干扰的模拟。类似的，Yu[61]提出的 AB-DH（Attention based data hiding）使用由原图像和噪声攻击图像混合而成的训练数据集。这种使用带有噪声的样本进行的混合训练，有助于进一步提高鲁棒性。

（2）消息编码模块。在嵌入前对秘密消息进行编码得到更有效的表达，能够在消息传递过程中有效减少消息损失。Luo 等[46]指出，现实中的不可微失真无法在训练过程中被实现。基于此，他们第一次探索了失真不可知深度水印方法的实现。该方法在训练过程中不对失真进行显示建模，而是通过使用卷积神经网络（CNN）生成的含有图像失真扰动的对抗性示例进行训练，并结合信道编码向系统注入冗余实现额外的鲁棒性。具体地，该方法为了向系统中注入额外冗余，用信道编码操作生成较长的二进制消息 X' 去替换原本的消息 X，然后将 X' 传递给水印编码器。使用二进制对称信道（Binary symmetric channel，BSC）对信道失真进行模拟，冗余消息 X' 通过噪声信道传输，以噪声冗余消息 X'_{no} 被水印解码器接收。水印解码器恢复出的消息称为 X'_{dec}，水印系统的误差就是 X' 与 X'_{dec} 差值。最后，信道解码器从损坏的消息 X'_{no} 中恢复出输入的消息。该方法获得了与显式失真训练相当的结果，并且对未知失真具有更好的鲁棒性。冗余信道编码过程如图 4 所示。

Zhang 等[47]提出了另一种消息编码方法。与 Luo 等[46]提出的注入冗余的信道编码不同，该消息编码模块将冗长的二进制消息 M 降维压缩至

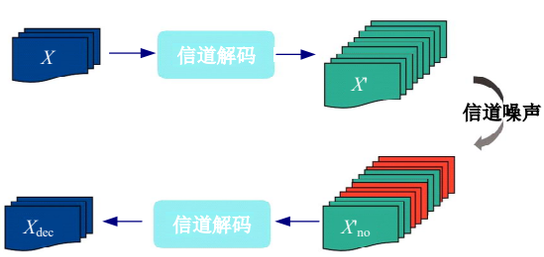

图 4　注入冗余的信道编码

Fig. 4　Channel coding with injected redundancy

一个紧凑的实值空间M_{en}。编码消息M_{en}与解码消息M_{de}之间的差值作为消息解码损失,待编码消息M与重构后的消息M_{out}之间的差值作为消息重构损失,如图5所示。实值域的大空间有利于学习到语义更丰富的消息表示,但是比特数更少。因此,在信息量相同的前提下,干扰更小,鲁棒性更强。

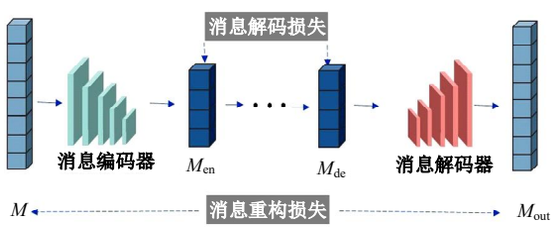

图5 降维压缩的信道编码

Fig. 5 Channel coding for dimensionality reduction compression

(3)嵌入域选择模块。在嵌入秘密信息时,选择更适合的嵌入域可以使信息隐蔽性和鲁棒性等得以提升。Zhang等[47]提出一种新的基于逆梯度注意力(Inverse gradient attention,IGA)的隐藏方法。直观地说,逆梯度注意掩码可以帮助定位对消息嵌入具有鲁棒性的像素点。该方法通过对载体图像中的不同像素点赋予不同的注意力权重,再对权重更高的像素期望嵌入更多信息,以此实现更鲁棒的信息隐藏。相比空间域,在频域进行的信息隐藏对一些常见的图像处理操作(如图像压缩、亮度和对比度调整等)具有更好的鲁棒性,且人眼不敏感的高频区域更加适合隐藏秘密信息。HiNet[53]选择在频域进行信息隐藏工作,采用小波变换将图像分割成低、高频小波子带并将秘密图像隐藏在载体图像的高频区域。这样不仅能更好地将秘密信息融合到载体图像中,还能获得较好的鲁棒性。Yu[61]提出的数据隐藏模型ABDH采用ResNet50作为注意力模型,载体图像输入注意力模块生成注意力掩码。注意力掩码的每个值代表每个像素的"注意力敏感度",这个值被正则化为0~1,接近1时,意味着对应像素的变化将引起明显的视觉检测差异。注意力掩码的值指导秘密信息嵌入的位置集中在视觉差异较弱的区域。原始载体图像、秘密图像和注意力掩码共同作为生成模型的输入,以生成视觉质量更好、秘密信息隐蔽性更强的目标图像。ISN通过叠加隐藏分支的通道数实现大容量多图像隐写。Guan等[56]则认为ISN的这种简单叠加忽视了图像之间的相关性。他们在DeepMIH中提出重要度模块(Important map,IM),在先前图像隐藏结果的基础上指导当前图像的隐藏以充分利用载体图像的隐藏潜力,实现多图像隐藏。Tan等[62]认为信息提取误码率大会导致在实际应用中需要引入更多纠错码,从而降低了有效容量。针对这一问题,他们采用一种信道注意策略,通过建模通道之间的相关性得到各通道的重要程度。该方法通过抑制无用通道的特征,从而削弱网络输出结果中噪声的存在,使信息提取精度得到了提升,间接实现了容量的提升。

(4)安全性提升模块。合理利用隐写分析网络对提升隐写方法的安全性很有帮助。Singh等[42]和Tan等[62]的方法将具有较好性能的隐写分析器XuNet作为鉴别器。XuNet通过对生成图像评分与生成器形成竞争,从而提升生成器性能,使其生成具有更高抗隐写分析能力的隐写图像。同样为提高安全性,Cui等[63]提出了一种创新的思路。他们认为以往的隐写方法都侧重于传输过程中的安全性,但接收端恢复秘密图像之后仍存在隐私泄露的安全风险。为解决这个问题,他们提出的图像隐写框架MIAIS(Multitask identity-aware image steganography)可以在不恢复秘密图像的情况下,直接对隐写图像进行内容识别。

2.2 训练策略和损失函数

2.2.1 通用损失函数设计

在图像信息隐藏领域,几乎所有方法都遵循两对通用的重要差异设计损失。首先是载体图像与隐写图像之间的差异,即隐藏损失。其次是原始秘密信息与恢复出的秘密信息之间的差异,即恢复损失。两种损失的示意如图6所示。

隐藏损失代表着隐藏前后载体图像的退化程度,与不可知性息息相关。隐藏损失的基本设计围绕着两张图像的相似程度,通常用图像失真损失进行度量。常用的有均方误差损失(Mean squared error,MSE)[41,42,59,60,62]和$\mathcal{L}_1$、$\mathcal{L}_2$距离[44,56,64]等。通用设计如下式所示:

$$\mathcal{L}_{con}=\mathrm{MSE}(X,H)$$
$$\mathcal{L}_{con}=\mathcal{L}_{1,2}(X,H) \tag{1}$$

式中:X和H分别为载体图像和隐写图像。

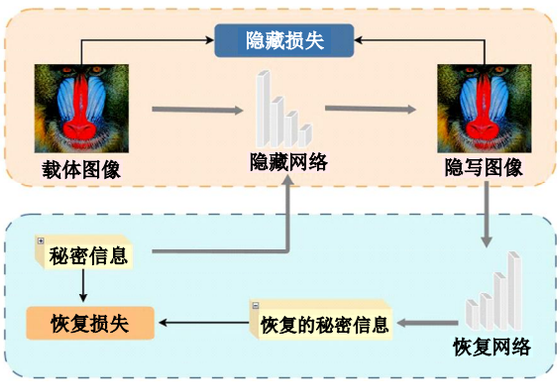

图 6　隐藏损失与恢复损失示意

Fig. 6　Schematic representation of concealing and recovery losses

恢复损失代表着秘密信息在隐藏和恢复操作前后的损失情况，是鲁棒性的直接反映。值得注意的是，在图像信息隐藏中，载体均为图像形式，但秘密信息的形式包含图像和文本两种。针对秘密信息的不同数据形式，恢复损失有不同的设计。当秘密信息是文本信息时，一般通过衡量消息丢失的损失度量，常用的如交叉熵损失（Cross-entropy loss）[41,45,49,62,64,65]；当秘密信息为图像形式时，这种损失同样变成衡量两张图像之间的差异，设计方式与隐藏损失其实是相同的，都是衡量隐写图像的退化程度或者图像相似度。恢复损失常见形式为：

$$\mathcal{L}_{\text{rev}}=\left\|S-S'\right\|_2^2$$

$$\mathcal{L}_{\text{rev}}=\text{CrossEntropy}(M,M') \tag{2}$$

式中：S、M、S'和M'分别为原始秘密图像、秘密信息、恢复后的秘密图像和恢复后的秘密信息。

SteganoGAN[41]使用 MSE 分析隐写图像和载体图像之间的相似度，利用交叉熵损失优化解码精度，并利用判别网络对隐写图像的真实性进行验证，联合优化以上 3 个损失，迭代优化编码器-解码器网络和判别网络。HiDDeN[44]为保证编码后的图像在视觉上与载体图像相似，用图像失真损失（隐写图像和载体图像之间的$\mathcal{L}_2$距离）表征相似性。为保证解码后的消息与编码后的消息相同，HiDDeN 使用原始消息和解码消息之间的$\mathcal{L}_2$距离施加消息失真损失。通过最小化输入消息和图像分布上的这几种损失实现最终优化。Luo 等[46]提出的失真不可知水印框架 Distortion Agnostic Deep Watermarking 为控制编码图像的感知质量和消息损耗，将图像损失、消息损失分别设置为载体图像和隐写图像、解码消息和输入消息之间的$\mathcal{L}_2$损失。在文献[62]中，原始消息M中的每个值都是 0 或者 1，提取出的消息M'的每个元素都是 0～1 内的浮点数。为保证秘密信息的准确恢复，该方法使用二值交叉熵损失最小化M与M'之间的差异。在训练完成并实际应用时，再将M'四舍五入到 0 或者 1 构造实数位序列。DeepMIH[56]将正向隐藏损失定义为隐写图像和载体图像之间的一种差值，这种差值可以用$\mathcal{L}_1$或$\mathcal{L}_2$损失衡量；同样反向恢复损失被定义为恢复出的秘密图像和载体图像之间的$\mathcal{L}_1$或$\mathcal{L}_2$损失。RIIS[55]的特殊之处在于恢复网络将载体图像和秘密信息同时恢复，恢复损失设计为：

$$\mathcal{L}_{\text{rev}}=\left\|X-X'\right\|_2+\left\|S-S'\right\|_2 \tag{3}$$

式中：X、X'分别为原始载体图像和恢复后的载体图像；S和S'分别为原始秘密图像和恢复出的秘密图像。

2.2.2　损失函数改进方法

为了获得性能的进一步提升，研究者们在通用的损失函数基础上又设计了多种损失函数进行更细致的约束。根据所用基线框架和嵌入域的不同，损失函数的设计也会有相应的改进。例如，在频域进行隐藏工作的方法使用的低频小波损失[53,56]、感知损失[42,56,59,62-64]；使用生成对抗思想的信息隐藏方法通常会使用对抗性损失[42,44,60,65]、判别器的分类损失[44,66]、循环对抗损失和不一致损失[61]等提升隐写图像的质量。下面将介绍各方法中以性能提升为目的所进行的损失函数的改进设计。

低频小波损失的设计灵感来自先验知识：隐藏在高频分量中的信息比隐藏在低频分量中的信息更不容易被察觉。因此，在频域进行嵌入的方法倾向于将信息隐藏在高频区域，这就要求方法对低频区域所做的修改尽可能小。低频小波损失通过约束载体图像和隐写图像的低频区域使二者极大相似，从而实现频域隐写的高不可知性。HiNet[53]除了设计常规的隐藏损失和恢复损失，还使用低频小波损失约束隐写图像经小波分解后的低频子带与载体图像的低频小波子带，使二者相似，以此实现将秘密信息隐藏在高频子带中。HiNet 使用的低频小波损失可表示为：

$$\mathcal{L}_{\text{freq}}=\sum_{n=1}^{N}\ell_F\left(\mathcal{H}(x_{\text{cover}}^{(n)})_{\text{LL}},\mathcal{H}(x_{\text{stego}}^{(n)})_{\text{LL}}\right) \quad (4)$$

式中:$\mathcal{H}(x_{\text{cover}}^{(n)})_{\text{LL}}$为隐写图像的低频子带;$\mathcal{H}(x_{\text{stego}}^{(n)})_{\text{LL}}$为载体图像的低频子带;$\ell_F$为载体图像和隐写图像的低频子带之间的差异。Deep-MIH[56]也同样使用了这种损失。

此外,如果想要在不可知性上获得进一步提升,尤其是针对人眼视觉的感知性,部分方法则会使用到感知损失[47]或视觉相似性损失[63]。Tan等[62]为保证载体图像和生成隐写图像在视觉上的不可分辨性,在方法中使用MSE衡量图像失真,并结合感知损失:将载体图像和隐写图像提供给预训练的VGG网络,最小化它们在深度映射之间的差异。他们使用的这种感知损失的计算方法可表示为

$$\mathcal{L}_p=\text{MSE}\left(\text{VGG}(X),\text{VGG}(H)\right) \quad (5)$$

式中:X和H分别为载体图像和隐写图像。

同样,基于条件生成对抗网络的方法Steg-GAN除了使用对抗损失和MSE损失,还使用基于预先训练的VGG16模型的感知损失训练提取器,使其以人类视觉系统(Human visual system,HVS)标准检索隐藏图像,保证了图像的视觉质量和统计不可检出性。Bui等[59]提出的模型RoSteALS使用了MSE结合感知损失LPIPS的损失设计,可表示为

$$\mathcal{L}_{\text{MSE}}=\left\|\gamma(X)-\gamma(H)\right\|^2$$
$$\mathcal{L}_{\text{quality}}=\mathcal{L}_{\text{LPIPS}}(X,H)+\alpha\mathcal{L}_{\text{MSE}} \quad (6)$$

式中:$\gamma(\cdot)$为从RGB空间向感知更均匀的YUV空间的映射函数;α为一个权重常数;X和H分别为载体图像和隐写图像。

侧重于安全性的方法MIAIS[63]为了保证秘密信息在到达接收端后不被泄露,选择不恢复秘密信息直接进行识别的设计。相应地,为保证直接识别的有效性,该方法设计了内容损失和相应的识别损失。内容损失在秘密图像(人脸图像)与隐写图像之间添加相似性约束,要求隐写图像极大保留秘密图像的身份信息,以便直接对隐写图像进行识别而不需要将秘密人脸图像恢复出来,规避掉了人脸信息在接收端泄露的风险。进一步针对人脸验证任务,识别损失采取交叉熵损失和三元组损失相结合,保证人脸图像分类的精确性,从而保证了MIAIS直接识别方法的有效性。此外,注重鲁棒性的方法RoSteALS[59]通过加入多种模拟噪声使秘密解码器可以在各种数据增强下进行训练,同时使用BCE(Binary cross-entropy loss)计算预测秘密和真秘密之间的比特恢复损失,进一步获得高鲁棒性。

2.3 面向复杂应用场景

前文对在基于标准测试数据集场景下的方法进行了总结。本节将对面向实际复杂应用场景时的图像信息隐藏方法进行介绍。

2.3.1 面向在线社交网络

在线社交网络(Online social networks,OSN)为用户提供各种联系和交流的互联网通道,丰富了用户的社交需求。每时每刻都有大量图像被上传到脸书、微信、微博等OSN平台,OSN已经成为存储和传输图像的便捷方式和可行通道。然而,OSN提供的信道中包含的各种有损操作(如裁剪和压缩)给图像隐藏的安全性造成了极大挑战。如何掌握并利用OSN信道的特点并更好地实现基于OSN场景的鲁棒安全的图像隐写,成为研究热点之一。2021年,Sun等[45]提出了一种针对OSN共享的鲁棒高容量水印技术。他们采取Facebook作为代表性OSN,利用Facebook提供的有损信道的精确知识,提出一种DCT域的水印方法。该方法能够做到即使没有任何纠错码,也能对Facebook上的有损操作具有高度鲁棒性,同时满足共享图像的高质量和高嵌入容量。该方法也能成功拓展到其他流行的OSN,如微信和Twitter。2022年,You等[67]提出了用于在OSN上进行隐蔽通信的无覆盖式图像隐写网络CIS-Net(Coverless image steganography network),并认为OSN让信息传递更加便捷,因此利用OSN提供的嘈杂信道传递秘密信息是很好的思路。该方法根据需要传递的秘密信息直接生成一张隐写图像,接收方从这张图像中提取出秘密信息从而实现秘密通信。此外,还发现社交网络上表情包和头像被广泛使用,故而使用这两种常见的形式传递信息,可进一步降低怀疑。CIS-Net最终实现通过生成人脸图像和表情包,在OSN上进行秘密信息的鲁棒传递。应用流程如图7所示。

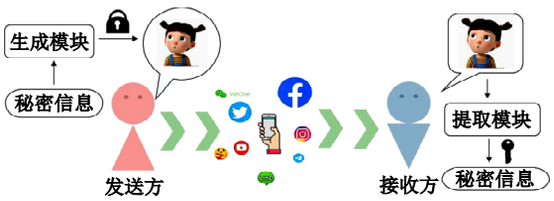

图 7　基于 OSN 实现隐蔽通信流程

Fig. 7　OSN-based implementation of covert communication flow

2.3.2　面向图像翻拍场景

为了获得更高实用性，研究者们努力将图像信息隐藏的应用场景更多地转移到现实生活中。2020 年，Tancik 等[64]设想通过连接互联网的成像系统访问隐藏在物理照片中的数据，他们提出的隐写算法 StegaStamp 能够在隐藏信息不可知的情况下对任意超链接位串进行鲁棒编码和解码。该方法使用一张图像和一个超链接位字符串作为输入，然后通过编码器将位串嵌入目标图像中，生成在感知上与输入图像相同的编码图像。这些编码图像被物理打印（或显示在电子显示器上）并呈现出来，在观感和使用效果上均无异常。用户利用成像系统捕捉包含这些物理照片的相片或视频，系统使用图像检测器对所有图像进行识别和裁剪。最终，解码器处理每个图像以检索出唯一的位串，该位串用于跟踪超链接并检索与图像相关的信息。应用流程如图 8 所示。

文献[64]还在真实情境下进行了鲁棒性训练，在编码器和解码器之间应用一组可微扰动近似由物理显示和相机成像引起的失真。将先前关于合成鲁棒对抗示例的工作如 HiDDeN[44]使用非空间扰动，Deep ChArUco[68]使用空间和非空间扰动训练鲁棒检测器，模拟了透视变形、运动和模糊失焦、颜色失真、相机系统噪声、JPEG 压缩等失真。该做法使该方法在现实应用场景下具有足够的抗干扰鲁棒性，实现了通过物理成像通道健壮地隐藏信息，极大提升了实用性。最后演示了在进行光照、阴影、透视、遮挡及观看距离变化的情况下，对来自摄录视频帧中的超链接进行实时解码。StegaStamp 在纠错后能够稳健地检索 56 位超链接，足够满足在互联网上的每张照片中嵌入一个专属代码。现实场景应用实例如图 9 所示。

图 8　编码超链接翻拍识别

Fig. 8　Encoding hyperlinks and tapping to recognise them

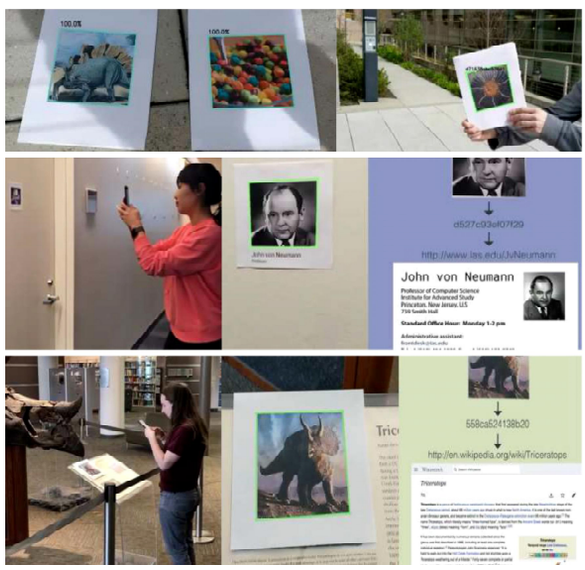

图 9　StegaStamp 在现实场景中的应用

Fig. 9　StegaStamp in real-life scenarios

3　实　验

3.1　数据集介绍

图像信息隐藏对实验数据集没有特殊要求。研究者们可根据任务需求和应用选择合适的数据集进行实验与研究，本节介绍了几种图像信息隐藏领域内的常用数据集。

3.1.1　ImageNet

ImageNet 数据集[69]是目前在深度学习图像领域应用最广泛的数据集。至今，ImageNet 已经发展成为包含 1 400 万张图像，2 万多个类别的大规模视觉识别数据集，每个类别至少包含 500 张图像。该数据集涵盖了各种场景和对象，包括动物、植物、人类、日常用品、建筑等，大小约为 1 TB。除了提供丰富的视觉数据资源，ImageNet 还通过举办挑战赛 ILSVRC（ImageNet large scale visual recognition challenge）推动图像分类、目标检测、图像分割等领域的研究。2012 年，深度学习方法在 ILSVRC 上取得显著的突破，这标志着计算机视觉领域的发展进入新阶段。文献

[42,48,53,54,56,63,70]等均使用ImageNet数据集或其变体与子集。

3.1.2 MS-COCO

MS-COCO (Microsoft common objects in context)数据集[71]是微软于2014年出资标注的大型物体检测和分割数据集。该数据集被广泛应用于目标检测、图像分割等任务,是目前为止拥有语义分割的最大数据集之一。COCO数据集包含80个类别,每个类别至少包含5张图像。该数据集共有超过33万张图片,其中20万张进行了标注,如物体的边界框、关键点等。整个数据大小约为40 GB,压缩后大小约为25 GB。此外,该数据集附带类别注释和分割注释,并且没有预定义的训练和测试分割。使用者可以根据研究主题和便利性对数据集进行自定义划分。文献[41,42,44,49,53,56,60-62]使用了COCO数据集。

3.1.3 DIV2K

DIV2K[72]是在2017年全图像超分辨率挑战赛中引入的单图像超分辨率数据集。该数据集包含1 000张2 K分辨率的图像,其中800张用于训练,100张用于验证,100张用于测试。该数据集可用于训练和评估超分辨率算法在提升图像质量方面的效果。DIV2K数据集的图像涵盖多种不同内容和场景,包括自然风景、人物肖像、动物、建筑等。这些图像都是来自真实世界,具有高度的复杂性和真实性,因此DIV2K数据集能够全面评估超分辨率算法的泛化能力。研究者们可以根据自己的需求利用DIV2K数据集生成不同分辨率的图像,并且能准确地比较不同算法在不同分辨率下的性能表现。DIV2K数据集在[41,42,44,49,53,55-57,64]中被使用。

3.1.4 CelebA

CelebA(CelebFaces attributes)[73]是一个广泛使用的人脸属性数据集,包含大量名人的人脸图像以及相关的属性标注信息。该数据集由香港中文大学多媒体实验室开放提供。CelebA数据集包含10 177个名人身份的202 599张人脸图片。这些图像具有不同的来源、位置、背景和姿态,能够提供多样化的训练数据。因此,CelebA数据集非常适合用于人脸属性标识训练、人脸检测训练以及标记人脸关键点等任务。除了图片本身,CelebA数据集还提供了40种不同的注释,例如戴/不戴眼镜、情绪、发型以及其他配件(如帽子),这些注释可以帮助研究人员更好地理解和分析人脸属性。由于CelebA数据集包含大量真实世界中的人脸图像,它也被广泛应用于图像隐写领域,常见的应用是将人脸图像作为隐藏秘密图像的载体。例如,文献[49,67]利用CelebA数据集作为训练数据集进行人脸隐写图像生成;文献[65]在CelebA数据集上对其生成式隐写网络进行了评估。

3.1.5 其他数据集

为更全面地评估算法性能并使其满足特定的隐藏需求,近年来有许多其他的数据集被用于图像信息隐藏领域。这些数据集根据应用需求不同,包含各种类型的图像和相关的标注信息。隐写分析数据集BOSSbase[30]、手写数字识别数据集MNIST[48]、医疗数据集、卫星数据集[63]、人脸数据集[65,67]、卧室数据集[65]等都较为常见。其中,常用于隐写分析领域算法研究与性能评估的数据集BOSSbase[74]提供了一系列经过隐写术嵌入的隐写图像和对应的原始载体图像。除了用来评估隐写分析方法的性能,这些图像也常在图像信息隐藏领域被使用。MNIST[75]是一个手写数字图像数据集,共有7万张图片,分为一个包含6万个实例的训练集和一个包含1万个实例的测试集。总大小为50 MB,图像均为黑白二值图。宽和高相同,为28像素,以二进制方式存储。表1展示了文献中常用的数据集的细节信息。图10为常

表1 部分数据集汇总

Table 1 Summary of selected datasets

数据集	文献使用	图像数量/张	图像格式	图像尺寸
ImageNet	[42,48,53,54,56,63,70]	超过1 400万	jpg	—
CelebA	[49,65,67]	超过20万	jpg	178×218×3
MS-COCO	[41,42,44,49,53,56,60−62]	超过33万	jpg	—
DIV2K	[41,42,44,49,53,55-57,64]	训练集800 验证集100 测试集100	png	—
BOSS base	[5]	训练集9 074 测试集1 000	tiff	512×512×1
MNIST	[48]	训练集60 000 测试集10 000	idx	28×28×1
LFW	[43]	13 233	jpg	150×150×3
CIFAR-10	[60]	50 000训练集 10 000测试集	—	32×32×3

图 10　常见数据集的图像样例集合

Fig. 10　Sample images of commonly used datasets

见数据集的图像样例集合。

3.2　评价指标

对于图像信息隐藏算法的性能，经常围绕 3 个维度进行评估比较，即隐藏性能、恢复性能和容量。本节对领域内常用的评估指标进行如下介绍。

3.2.1　PSNR

PSNR(Peak signal to noise ratio)即峰值信噪比，是一种评价图像的客观标准。PSNR 值表示处理后图像质量情况，单位是 dB。PSNR 可以用于衡量隐写图像的质量和秘密信息的不可见性。给定宽度为 W，高度为 H 的两幅图像 X 和 Y，PSNR 通过原图与被处理图像之间的均方误差 MSE(Mean Square Error)定义，公式如下：

$$\begin{aligned}&\mathrm{MSE}=\frac{1}{WH}\sum_{i=1}^{W}\sum_{j=1}^{H}\left[X_{i,j}-Y_{i,j}\right]^2\\&\mathrm{PSNR}=10\cdot\log_{10}\frac{\mathrm{MAX}^2}{\mathrm{MSE}}\end{aligned}\tag{7}$$

式中：$X_{i,j}$ 和 $Y_{i,j}$ 分别为图像 X 和 Y 在位置 (i,j) 处的像素值；MAX 为图像的最大像素值，如果像素值由 N 位二进制表示，那么 $\mathrm{MAX}=2^N-1$。以上公式是针对灰度图像的计算方法，如果 X 和 Y 是宽度为 W、高度为 H、通道数为 C 的彩色图像，可通过将公式做如下替换计算：

$$\mathrm{MSE}=\frac{1}{WHC}\sum_{i=1}^{W}\sum_{j=1}^{H}\sum_{k=1}^{C}\left[X_{i,j,k}-Y_{i,j,k}\right]^2\tag{8}$$

式中：$X_{i,j,k}$ 和 $Y_{i,j,k}$ 分别为图像 X 和 Y 在坐标 i、j 以及通道 z 处的像素值[76]。

在图像信息隐藏背景下，式(7)中的图像 X 和 Y 分别对应着载体图像和隐写图像，PSNR 值代表载体图像和隐写图像对之间的差异情况。PSNR 数值越大，代表 MSE 越小，两张图片越相似，图像失真就越小，意味着生成的隐写图像质量很好。PSNR 是使用最为广泛的一种图像客观评价指标，其局限性在于，它是通过计算对应像素点间的误差进行基于误差敏感的图像质量评价，而并未考虑人眼的视觉特性(人眼对空间频率较低的对比差异敏感度较高、人眼对亮度对比差异的敏感度较色度高、人眼对一个区域的感知结果会受到其周围邻近区域的影响等)。因此，PSNR 值的评价结果与人的主观感觉可能会不一致，如图 11 所示。一张 PSNR 值较高的图像，呈现的视觉效果却不尽人意。

3.2.2　SSIM

SSIM(Structural similarity index measure)即结构相似性指标，是基于结构信息退化的图像质量评价方法。SSIM 公式基于样本图像 X 和 Y 之间的 3 个比较衡量：亮度(luminance)、对比度(contrast)和结构(structure)，以更好地适应人类视觉系统，可表示为

$$\begin{cases}l(X,Y)=\dfrac{2\mu_X\mu_Y+c_1}{\mu_X^2\mu_Y^2+c_1}\\c(X,Y)=\dfrac{2\sigma_X\sigma_Y+c_2}{\sigma_X^2+\sigma_Y^2+c_2}\\s(X,Y)=\dfrac{\sigma_{XY}+c_3}{\sigma_X+\sigma_Y+c_3}\end{cases}\tag{9}$$

图 11　视觉上退化较大的人像(a)PSNR 值高于(b)

Fig. 11　Visually degraded portraits with(a) higher PSNR values than (b)

式中:μ_X为X的均值;μ_Y为Y的均值;σ_X^2为X的方差;σ_Y^2为Y的方差;σ_{XY}为X和Y的协方差;$c_1=(k_1L)^2$,$c_2=(k_2L)^2$,$c_3=c2/2$为3个常数,用于避免除数为0;L为图像像素值的范围,即$0\sim2^N-1$。那么,给定两幅图像X和Y,SSIM计算公式如下:

$$\mathrm{SSIM}(X,Y)=\left[l(x,y)^{\alpha}\cdot c(x,y)^{\beta}\cdot s(x,y)^{\gamma}\right] \tag{10}$$

若令α、β、γ为1,则得到常用SSIM计算公式:

$$\mathrm{SSIM}=\frac{(2\mu_X\mu_Y+c_1)(2\sigma_{XY}+c_2)}{(\mu_X^2+\mu_Y^2+c_1)(\sigma_X^2+\sigma_Y^2+c_2)} \tag{11}$$

每次计算时,需要从图片上取一个$N\times N$的窗口,然后不断滑动窗口进行计算,最后取平均值作为全局的SSIM。由公式可知,SSIM返回值在[-1,1],值为1代表两张图像完全相同。在图像信息隐藏任务中,SSIM值越大,表明隐写图像与载体图像具有较高的结构相似性,通常意味着秘密信息具有良好的不可知性。针对不同任务,SSIM的用法也不同。SSIM除了用作评价指标度量两个图像的相似度,还经常被用作损失函数。表2总结了部分以图像为秘密信息的方法的PSNR、SSIM值(基于ImageNet数据集),数据来源于文献[42,48,53,54,56]。

3.2.3 BPP

图像信息隐藏算法的容量一般使用比特每像素(Bits per pixel,BPP)计算。BPP表示在载体图像中每像素用来生成隐写图像的比特数。为了获得高隐藏容量,必须有高BPP值。BPP值计算公式如下:

$$\mathrm{BPP}=\frac{L}{HWC} \tag{12}$$

表2 部分方法的PSNR和SSIM值对比

Table 2 Comparison of PSNR and SSIM values for selected methods

隐写方法	PSNR/dB 隐写-载体/秘密-恢复	SSIM 隐写-载体/秘密-恢复
4bit-LSB	33.68/31.26	0.940 1/0.903 3
StegGAN[42]	42.24/37.17	0.990 5/0.950 8
Rahim等[48]	32.9/36.6	0.96/0.96
ISN[54]	38.05/35.38	0.954/0.955
HiNet[53]	44.6/46.78	0.992 8/0.995 2
DeepMIH[56]	40.31/36.63	0.980 0/0.960 4

式中:L为要隐藏的秘密信息的长度;H、W、C分别为载体图像的高度、宽度、通道数。表3总结了部分方法的BPP值。

表3 部分方法的BPP值对比

Table 3 Comparison of BPP values for selected methods

隐写方法	隐写容量/BPP
HiDDeN[44]	0.203
SteganoGAN[41]	4.4
文献[48]	8

3.2.4 其他指标

出于对隐写图像高质量和实验高效率的追求,各种基于深度学习的方法加入了额外多种指标对隐写方法进行衡量,其中使用较多的有感知相似度LPIPS[64,70]、解码错误率、视觉信息保真度VIF[42]、多尺度结构相似性指数MS-SSIM[42]、通用图像质量指数UQI[42]、消息的比特精度bit accuracy[44,60]、生成图像的视觉质量Fid[65]、隐写图像不可检出性Pe等[65]。

文献[77]表明,广泛使用的PSNR、SSIM、FSIM[78]等指标在判断图片的感知相似度时给出了与人类感知相违背的结论,而相比之下,基于学习的感知相似度度量要更符合人类的感知。由于视觉相似性的概念通常是很主观的,旨在模仿人类的视觉感知,所以真正被我们需要的是一个“感知距离”,以符合人类判断的方式衡量两幅图像的相似程度。其中,近年来被发现在ImageNet中训练的VGG网络的特征作为图像合成的训练损失非常有用。LPIPS的值越低,表示两张图越相似,反之差异越大,是目前除PSNR和SSIM外应用最为广泛的性能指标之一。

3.3 实验性能对比

本节对前文涉及的深度学习方法的性能情况进行归纳。表4展示了部分具有代表性的方法在对应评价数据集上的实验结果。为了方便比较,选择最常用的3个性能指标,并且对涉及多图像隐藏的方法(ISN、DeepMIH、RIIS),仅列出了隐藏单张图像时对应的数值。数据来源自文献[40-44,48-52,54,55,57-59,63,68],—表示来源文献未给出相关数据,*表示该数据集为文献自建数据集。其中,秘密信息形式为二进制的隐写方法不生成恢复的秘密图像,因此仅涉及隐写-载体图像对的PSNR/SSIM值。

表 4　基于深度学习的隐写方法性能比较

Table 4　Performance comparison of deep learning based image information hiding methods

方法名称(来源&年份)	数据集	秘密信息形式	PSNR/dB 隐写-载体/秘密-恢复	SSIM 隐写-载体/秘密-恢复	BPP
HiDDeN[44](ECCV2018)	COCO	二进制	37.27	—	0.203
SteganoGAN[41](2019)	DIV2K	二进制	36.52	0.85	2.63
	COCO		36.33	0.88	4.4
FNNS[49](ICLR2022)	COCO	二进制	30.05	0.71	3
	DIV2K		26.25	0.73	2
	CelebA		27.77	0.72	3
StegGAN[42](Springer2022)	ImageNet	图像	42.24/37.17	0.990 5/0.950 8	—
	COCO		36.26/31.92	0.979 9/0.942 8	—
	DIV2K		34.39/31.85	0.974 4/0.941 0	—
	KODAK		30.97/29.49	0.978 8/0.936 3	—
	SIPI		30.97/29.49	0.978 8/0.936 3	—
ISN[54](CVPR2021)	ImageNet	图像	38.05/35.38	0.954/0.955	—
	Paris StreetView		40.49/43.33	0.980/0.991	—
HiNet[53](ICCV2021)	DIV2K	图像	48.99/52.86	0.9971/0.9992	—
	COCO		46.52/46.98	0.9961/0.9957	—
	ImageNet		44.60/46.78	0.9928/0.9952	—
DeepMIH[56](TPAMI2023)	DIV2K	图像	43.72/41.41	0.9895/0.9801	—
	COCO		40.30/36.55	0.9805/0.9613	—
	ImageNet		40.31/36.63	0.9800/0.9604	—
RIIS[55](CVPR2022)	DIV2K	图像	—/44.19	—	—
	ImageNet		43.97/46.71	—	—
PRIS[57](2023)	DIV2K	图像	41.39/40.71	—	—
	ImageNet		39.90/39.26	—	—
	COCO		37.98/37.16	—	—
	VOC		39.64/39.35	—	—
IGA[47](2022)	COCO	二进制	32.59	—	—
	DIV2K		46.02	—	—
TDSL[60](ACMMM2019)	COCO	二进制	33.51	—	—
ABDH[61](AAAI2020)	Set5、Set14	图像	33.50/30.42	0.964 7/0.948 7	—
UDH[70](NeurLPS2020)	ImageNet	图像	39.13/35.0	0.985/0.976	—
GSN[65](ACMMM2022)	CelebA	二进制	—	—	1～2
Rahim 等[48](ECCV2018)	ImageNet	图像	32.92/36.58	0.96/0.96	8
Luo 等[46](CVPR2020)	COCO	二进制	33.7	—	—
Tan 等[62](IEEE2022)	COCO	二进制	41.84	0.983 2	3
RoSteALS[59](CVPR2023)	CLIC	二进制	32.68±1.75	0.88±0.06	—
	MetFACE		34.46±1.91	0.89±0.07	—
	Stock1K*		33.27±2.32	0.89±0.08	—
StegaStamp[64](CVPR2020)	ImageNet	二进制	28.50	0.905	—

4　图像信息隐藏的应用

4.1　在保密通信中的应用

最早利用隐写术进行保密通信的例子可以追溯到远古时代,如使用隐形墨水和卡登格子法[79]等。从古典隐写术发展到现代隐写术,信息隐藏技术在互联网的基础上产生了较大飞跃。信息的保密传输可以依赖图像信息隐藏技术,隐写术的目的就是掩盖机密信息存在的事实[80],从而躲避攻击者和恶意窃取者的注意,在根本上为保密通信过程加上一层可靠的保护。如在一幅普通图像

中隐藏一幅机密图像,在一段视频流中隐藏各种信息等。Morkel[81]提出根据接收方的数量将安全通信分为自通信、一对一通信和一对多通信三类,利用图像隐写技术实现了每个安全通信类别的应用。Pandey等[82]利用隐写术和图像压缩技术实现了秘密通信以及文本数据提取方法。

4.2 在版权保护中的应用

图像信息隐藏最合适、应用最广泛的领域就是版权保护领域。涉及版权保护的信息隐藏技术统称为数字水印[34]。数字作品极易被无损地复制传播、进行未经授权的修改和共享,因此版权所有者的权益易受到威胁[83]。数字水印技术在图像作品版权保护方面的应用可以分为以下两方面。(1)版权保护:将版权所有者的信息作为水印嵌入要保护的图像中,水印不仅不可感知,还应具有一定的抗干扰能力。如郑钢等[84]提出鲁棒盲水印模型IRBW-GAN(Image robust blind watermark-GAN)以实现产权保护。(2)作品认证:用于图像的真实性认证和完整性保护。如果经检测发现作品中的水印受到破坏,则证明图像被篡改。这类水印一般是脆弱的,需要对信号的改动很敏感,即使载体图像经过很微小的处理,水印也会被破坏或失效[85]。如郑秋梅等[86]针对脆弱水印在图像篡改检测中高定位精度和抗复杂攻击两方面不能同时满足的问题,提出了基于复杂攻击的脆弱水印图像完整性认证算法。

4.3 在医疗图像保护中的应用

随着远程医疗和智能医疗平台的快速发展[87],电子医疗档案面临的安全问题被日益关注。医疗影像是医学诊断中最常用的数据记录与呈现媒介,因此需要对其中的信息完整性和患者隐私进行保护,防止被篡改和盗用[88]。图像信息隐藏技术在不降低影像质量的情况下隐藏患者数据[89,90],同时也能避免医疗影像与纸质诊断分离导致信息不匹配或丢失的情况。Karakus等[91]提出基于最佳像素相似度的医学图像隐写方法,从开放数据库Dicom中获取不同大小的医学图像并将不同身份的医嘱隐藏在这些医学图像中。Priyadharshini等[92]结合密码学和隐写术,首先使用一次性填充算法对医学图像进行加密。其次,利用LSB隐写术将加密后的医学图像植入载体图像中生成隐写图像,再将其送入信道传输。最后,在接收端检索恢复医学图像。Balu等[93]提出了应用于医学影像系统的视频隐写技术,用于将患者信息从一家医院传输到另一家医院。

5 未来方向

深度学习为图像信息隐藏领域的发展带来巨大改变,但目前图像信息隐藏技术尚未成熟。领域内挑战与发展并存,仍有巨大研究潜力。本文对未来图像信息隐藏研究的潜在方向做出展望:

首先,隐写图像的视觉质量可以进一步提升。尤其在用户对容量的需求逐渐增大的情况下,如何不暴露秘密隐藏行为,保证隐蔽传输的安全性依旧是重要方向。

其次,对不同隐写分析方法与不同数据集的泛化能力有待提高。随着基于深度学习的隐写分析技术的发展,隐写方法的安全性被进一步要求。如何使其应对不同隐写分析方法时均能有效保护秘密信息,使用不同类型的载体图像时均能实现性能良好的秘密嵌入,是未来研究的一个重要方向。

最后,实时隐写[57,64]是图像信息隐藏有巨大潜力的研究方向,这涉及抗信道传输噪声[45]、保存与下载的压缩噪声、轻量级设计等[59]问题的进一步研究。视频隐写是另一个有待进一步发展的重要方向,如在视频中隐藏图像或者将一个视频片段隐藏到另一个视频片段中[58]。目前视频隐藏的挑战主要是如何在整个视频时序序列中找到最佳隐藏帧,以及如何在帧之间分配隐藏容量以实现最佳的视频隐藏性能[56]。此外,秘密信息结合视频帧的时序排列问题[58]以及视频画质所受影响的问题都需要进一步研究。针对AI换脸技术的滥用现象,如政治选举诋毁、网络诈骗、威胁勒索、证据造假等[94,95],若有效利用图像信息隐藏技术加强生成式AI管理(如强制其产出携带隐蔽AI标识),从而限制AI生成作品的滥用和未经授权的使用等,也是未来方向之一。

6 总结与展望

图像信息隐藏已经不再是针对某些领域的特殊应用,它在互联网参与下的媒体生活中变得越来越被人们所需要。基于深度学习的图像信息隐藏技术已成为网络安全领域中保护用户隐私和版权信息的关键工具。本文深入分析了这一领域的最新进展,从结构,训练以及应用3个角度进行综

述。结果发现,深度学习极大地推动了图像信息隐藏技术的发展,其中基于编码器-解码器结构的方法最为常见。具体而言,基于生成对抗网络(GAN)的方法因其高安全性和良好的不可知性而广受欢迎。此外,可逆神经网络(INN)在图像信息隐藏中的应用为这一领域提供了新的视角和解决方案。各类方法在鲁棒性、安全性、容量和不可知性等方面取得了显著成就,同时在损失设计和训练策略上也展现了创新性。为应对在线社交网络(OSN)和基于翻拍的现实场景,研究者们通过模拟有损信道噪声和图像扰动有效提升了图像信息隐藏的实用性。本文还总结了该领域内的主要数据集和评价指标,为后续研究提供参考。然而,尽管深度学习在图像信息隐藏领域的应用取得了重大进展,但是领域内仍存在诸多挑战和待解决的问题。未来研究的潜在方向包括进一步提升隐写图像的视觉质量,增强隐写方法对不同隐写分析和数据集的泛化能力以及深入探索实时隐写和视频隐写等。综上所述,深度学习在图像信息隐藏领域具有巨大的潜力,需要同行不断努力,共同推动领域的进一步发展。

参考文献:

[1] 张卫明, 王宏霞, 李斌, 等. 多媒体隐写研究进展[J]. 中国图象图形学报, 2022, 27(6): 1918-1943.
Zhang Wei-ming, Wang Hong-xia, Li Bin, et al. Overview of steganography on multimedia[J]. Journal of Image and Graphics, 2022, 27(6): 1918-1943.

[2] Subramanian N, Elharrouss O, Almaadeed S, et al. Image steganography: a review of the recent advances [J]. IEEE Access, 2021, 9: 23409-23423.

[3] Tsai P, Hu Y C, Yeh H L. Reversible image hiding scheme using predictive coding and histogram shifting [J]. Signal Processing, 2009, 89(6): 1129-1143.

[4] Pan F, Li J, Yang X. Image steganography method based on PVD and modulus function[C]//2011 International Conference on Electronics, Communications and Control (ICECC), Ningbo, China, 2011: 282-284.

[5] 孙曦, 张卫明, 俞能海, 等. 基于空域图像变换参数扰动的隐写术[J]. 通信学报, 2017, 38(10): 166-174.
Sun Xi, Zhang Wei-ming, Yu Neng-hai, et al. Steganography based on parameters′ disturbance of spatial image transform[J]. Journal on Communications, 2017, 38(10): 166-174.

[6] Rustad S, Syukur A, Andono P N. Inverted LSB image steganography using adaptive pattern to improve imperceptibility[J]. Journal of King Saud University-Computer and Information Sciences, 2022, 34(6): 3559-3568.

[7] Fazli S, Moeini M. A robust image watermarking method based on DWT, DCT, and SVD using a new technique for correction of main geometric attacks[J]. Optik-International Journal for Light and Electron Optics, 2016, 127(2): 964-972.

[8] Abadi R Y, Moallem P. Robust and optimum color image watermarking method based on a combination of DWT and DCT[J]. Optik, 2022, 261: No. 169146.

[9] 赵瑶瑶, 李万社. 基于离散小波变换和离散余弦变换的彩色图像水印算法[J]. 应用数学进展, 2021, 10: No. 1096.
Zhao Yao-yao, Li Wan-she. Color image watermarking algorithm based on discrete wavelet transform and discrete cosine transform[J]. Advances in Applied Mathematics, 2021, 10: No. 1096.

[19] Wang R Z, Lin C F, Lin J C. Image hiding by optimal LSB substitution and genetic algorithm[J]. Pattern Recognition, 2001, 34(3): 671-683.

[11] Fridrich J, Goljan M, Du R. Detecting LSB steganography in color, and gray-scale images[J]. IEEE Multimedia, 2001, 8(4): 22-28.

[12] Mandal P C, Mukherjee I, Paul G, et al. Digital image steganography: a literature survey[J]. Information Sciences, 2022, 609: 1451-1488.

[13] Chhikara S, Kumar R. Information theoretic steganalysis of processed image LSB steganography[J]. Multimedia Tools and Applications, 2023, 82(9): 13595-13615.

[14] 付章杰, 李恩露, 程旭, 等. 基于深度学习的图像隐写研究进展[J]. 计算机研究与发展, 2021, 58(3): 548-568.
Fu Zhang-jie, Li En-lu, Cheng Xu, et al. Recent advances in image steganography based on deep learning [J]. Journal of Computer Research and Development, 2021, 58(3): 548-568.

[15] Ehrlich M, Davis L, Lim S N, et al. Analyzing and mitigating jpeg compression defects in deep learning [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, Canada, 2021: 2357-2367.

[16] Khayam S A. The discrete cosine transform (DCT): theory and application[J]. Journal of Michigan State University, 2003, 114(1): 1-31.

[17] Poyueh P Y, Lin H J. A DWT based approach for image steganography[J]. International Journal of Applied Science and Engineering, 2006, 4(3): 275-290.

[18] Ruanaidh J, Dowling W J, Boland F M. Phase watermarking of digital images[J]. Proceedings of 3rd IEEE International Conference on Image Processing, 1996, 3: 239-242.

[19] Li Y. Research and application of deep learning in image recognition[C]//2022 IEEE 2nd International Conference on Power, Electronics and Computer Applications (ICPECA), Shenyang, China, 2022: 994-999.

[20] 徐涛, 马克, 刘才华. 基于深度学习的行人多目标跟踪方法[J]. 吉林大学学报: 工学版, 2021, 51(1): 27-38.
Xu Tao, Ma Ke, Liu Cai-hua. Multi object pedestrian tracking based on deep learning[J]. Journal of Jilin University (Engineering and Technology Edition), 2021, 51(1): 27-38.

[21] Lopez M M, Kalita J. Deep learning applied to NLP [J/OL]. [2024-04-01]. https://arxiv.org/abs/1703.03091

[22] Zhang Z, Geiger J, Pohjalainen J, et al. Deep learning for environmentally robust speech recognition: an overview of recent developments[J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2018, 9(5): 1-28.

[23] 黄立威, 江碧涛, 吕守业, 等. 基于深度学习的推荐系统研究综述[J]. 计算机学报, 2018, 41(7): 1619-1647.
Huang Li-wei, Jiang Bi-tao, Lyu Shou-ye, et al. Survey on deep learning based recommender systems [J]. Chinese Journal of Computers, 2018, 41(7): 1619-1647.

[24] Heaton J B, Polson N G, Witte J H. Deep learning for finance: deep portfolios[J]. Applied Stochastic Models in Business and Industry, 2017, 33(1): 3-12.

[25] 沈权猷, 张小波, 李文豪, 等. U-Net在肺结节分割中的应用进展[J]. 计算机应用, 2023, 43(增刊1): 250-257.
Shen Quan-you, Zhang Xiao-bo, Li Wen-hao, et al. Progress of U-Net applications to lung nodule segmentation[J]. Journal of Computer Applications, 2023, 43(Sup. 1): 250-257.

[26] Alom M Z, Yakopcic C, Hasan M, et al. Recurrent residual U-Net for medical image segmentation[J]. Journal of Medical Imaging, 2019, 6(1): No. 014006.

[27] 付章杰, 王帆, 孙星明, 等. 基于深度学习的图像隐写方法研究[J]. 计算机学报, 2020, 43(9): 1656-1672.
Fu Zhang-jie, Wang Fan, Sun Xing-ming, et al. Research on steganography of digital images based on deep learning[J]. Chinese Journal of Computers, 2020, 43(9): 1656-1672.

[28] 张茹, 刘建毅, 刘功申, 等. 数字内容安全[M]. 北京: 北京邮电大学出版社, 2017.

[29] Sara U, Akter M, Uddin M S. Image quality assessment through FSIM, SSIM, MSE and PSNR—a comparative study[J]. Journal of Computer and Communications, 2019, 7(3): 8-18.

[30] Xu G, Wu H Z, Shi Y Q. Structural design of convolutional neural networks for steganalysis[J]. IEEE Signal Processing Letters, 2016, 23(5): 708-712.

[31] Geng L, Zhang W, Chen H, et al. Real-time attacks on robust watermarking tools in the wild by CNN[J]. Journal of Real-Time Image Processing, 2020, 17: 631-641.

[32] Wei P, Zhu Z, Luo G, et al. Breaking robust data hiding in online social networks[J]. IEEE Signal Processing Letters, 2022, 29: 2682-2686.

[33] Chaumont M. Deep Learning in Steganography and Steganalysis[M]. New York: Academic Press, 2020.

[34] 夏道勋, 王林娜, 宋允飞, 等. 深度神经网络模型数字水印技术研究进展综述[J]. 科学技术与工程, 2023, 23(5): 1799-1811.
Xia Dao-xun, Wang Lin-na, Song Yun-fei, et al. Review of deep neural network digital watermarking technology[J]. Science Technology and Engineering, 2023, 23(5): 1799-1811.

[35] Denis R, Madhubala P. Hybrid data encryption model integrating multi-objective adaptive genetic algorithm for secure medical data communication over cloud-based healthcare systems[J]. Multimedia Tools and Applications, 2021, 80: 21165-21202.

[36] Jiang X H. Digital watermarking and its application in image copyright protection[C]//2010 International Conference on Intelligent Computation Technology and Automation, Changsha, China, 2010: 114-117.

[37] Megías D, Mazurczyk W, Kuribayashi M. Data hiding and its applications: digital watermarking and steganography[J]. Applied Sciences, 2021, 11(22): No. 10928.

[38] 赵洁, 邹天宇, 黄展鹏, 等. 基于手写数字水印的医学图像版权保护研究[J]. 医疗卫生装备, 2016, 37(6): 36-38.
Zhao Jie, Zou Tian-yu, Huang Zhan-peng, et al. Re-

search on copyright protection of medical images based on handwritten digital watermarking[J]. Chinese Medical Equipment Journal, 2016, 37(6): 36-38.

[39] 周娜, 成茗, 贾孟霖, 等. 基于缩略图加密和分布式存储的医学图像隐私保护[J]. 计算机应用, 2023, 43(10): 3149-3155.
Zhou Na, Cheng Ming, Jia Meng-lin, et al. Medical image privacy protection based on thumbnail encryption and distributed storage[J]. Journal of Computer Applications, 2023, 43(10): 3149-3155.

[40] 刘佳, 柯彦, 雷雨, 等. 生成对抗网络在图像隐写中的应用[J]. 武汉大学学报: 理学版, 2019, 65(2): 139-152.
Liu Jia, Ke Yan, Lei Yu, et al. Application of generative adversarial networks in image steganography [J]. Journal of Wuhan University (Science Edition), 2019, 65 (2): 139-152.

[41] Zhang K A, Cuesta I A, Xu L, et al. SteganoGAN: high capacity image steganography with GANs[J/OL]. [2024-04-01]. https: //arxiv. org/abs/1901.03892

[42] Singh B, Sharma P K, Huddedar S A, et al. SteggAN: hiding image within image using conditional generative adversarial networks[J]. Multimedia Tools and Applications, 2022, 81(28): 40511-40533.

[43] Wang Z, Gao N, Wang X, et al. SSteGAN: self-learning steganography based on generative adversarial networks[C] //Neural Information Processing: 25th International Conference, Siem Reap, Cambodia, 2018: 253-264.

[44] Zhu J, Kaplan R, Johnson J, et al. Hidden: hiding data with deep networks[C]//Proceedings of the European conference on computer vision (ECCV), Munich, Germany, 2018: 657-672.

[45] Sun W, Zhou J, Li Y, et al. Robust high-capacity watermarking over online social network shared images[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 31(3): 1208-1221.

[46] Luo X, Zhan R, Chang H, et al. Distortion agnostic deep watermarking[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020: 13548-13557.

[47] Zhang H, Wang H, Cao Y, et al. Robust data hiding using inverse gradient attention[J/OL]. [2024-04-01]. https://arxiv. org/pdf/2011. 10850. pdf

[48] Rahim R, Nadeem S. End-to-end trained CNN encoder-decoder networks for image steganography [C] //Proceedings of the European Conference on Computer Vision (ECCV) Workshops, Munich, Germany, 2018: 723-729.

[49] Kishore V, Chen X, Wang Y, et al. Fixed neural network steganography: train the images, not the network[C]//International Conference on Learning Representations, Vienna, Austria, 2021: 1-10.

[50] Dinh L, Krueger D, Bengio Y. Nice: non-linear independent components estimation[J/OL]. [2024-04-01]. https://arxiv. org/abs/1410. 8516

[51] Ardizzone L, Kruse J, Wirkert S, et al. Analyzing inverse problems with invertible neural networks[J/OL]. [2024-04-01]. https: //arxiv. org/abs/1808.04730

[52] Dinh L, Sohl D J, Bengio S. Density estimation using real NVP[C]//ICLR 2017-Conference Track Proceedings, Toulon, France, 2019: No. 160508803.

[53] Jing J, Deng X, Xu M, et al. HiNet: deep image hiding by invertible network[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, Canada, 2021: 4733-4742.

[54] Lu S P, Wang R, Zhong T, et al. Large-capacity image steganography based on invertible neural networks[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, USA, 2021: 10816-10825.

[55] Xu Y, Mou C, Hu Y, et al. Robust invertible image steganography[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, USA, 2022: 7875-7884.

[56] Guan Z, Jing J, Deng X, et al. DeepMIH: deep invertible network for multiple image hiding[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(1): 372-390.

[57] Yang H, Xu Y, Liu X, et al. PRIS: practical robust invertible network for image steganography[J/OL]. [2024-04-01]. https://www. keyanzhidian. com/doc/detail?id=1010113855433

[58] Mou C, Xu Y, Song J, et al. Large-capacity and flexible video steganography via invertible neural network[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, Canada, 2023: 22606-22615.

[59] Bui T, Agarwal S, Yu N, et al. Rosteals: robust steganography using autoencoder latent space[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, Canada, 2023: 933-942.

[60] Liu Y, Guo M, Zhang J, et al. A novel two-stage separable deep learning framework for practical blind watermarking[C]//Proceedings of the 27th ACM International Conference on Multimedia, Nice, France, 2019: 1509-1517.

[61] Yu C. Attention based data hiding with generative adversarial networks[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(1): 1120-1128.

[62] Tan J, Liao X, Liu J, et al. Channel attention image steganography with generative adversarial networks [J]. IEEE Transactions on Network Science and Engineering, 2021, 9(2): 888-903.

[63] Cui J, Zhang P, Li S, et al. Multitask identity-aware image steganography via minimax optimization[J]. IEEE Transactions on Image Processing, 2021, 30: 8567-8579.

[64] 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, Seattle, USA, 2020: 2117-2126.

[65] Wei P, Li S, Zhang X, et al. Generative steganography network[C]//Proceedings of the 30th ACM International Conference on Multimedia, New York, USA, 2022: 1621-1629.

[66] Liu X, Ma Z, Ma J, et al. Image disentanglement autoencoder for steganography without embedding[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, USA, 2022: 2303-2312.

[67] You Z, Ying Q, Li S, et al. Image generation network for covert transmission in online social network [C]//Proceedings of the 30th ACM International Conference on Multimedia, New York, USA, 2022: 2834-2842.

[68] Hu D, Detone D, Malisiewicz T. Deep charuco: dark charuco marker pose estimation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019: 8436-8444.

[69] Deng J, Dong W, Socher R, et al. Imagenet: a large-scale hierarchical image database[C] //IEEE Conference on Computer Vision and Pattern Recognition, Miami, USA, 2009: 248-255.

[70] Zhang C, Benz P, Karjauv A, et al. Udh: universal deep hiding for steganography, watermarking, and light field messaging[J]. Advances in Neural Information Processing Systems, 2020, 33: 10223-10234.

[71] Lin T Y, Maire M, Belongie S, et al. Microsoft coco: common objects in context[C] //Computer Vision-ECCV 2014: 13th European Conference, Zurich, Switzerland, 2014: 740-755.

[72] Eirikur A, Radu T. NTIRE 2017 challenge on single image super-resolution: dataset and study[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, USA, 2017: 126-135.

[73] Liu Z, Luo P, Wang X, et al. Deep learning face attributes in the wild[C]//Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 2015: 3730-3738.

[74] Bas P, Filler T, Pevný T. Break our steganographic system: the ins and outs of organizing boss[C]//International Workshop on Information Hiding, Berlin, Germany, 2011: 59-70.

[75] Deng L. The mnist database of handwritten digit images for machine learning research best of the web [J]. IEEE Signal Processing Magazine, 2012, 29(6): 141-142.

[76] Setiadi D R I M. PSNR vs SSIM: imperceptibility quality assessment for image steganography[J]. Multimedia Tools and Applications, 2021, 80(6): 8423-8444.

[77] Zhang R, Isola P, Efros A A, et al. The unreasonable effectiveness of deep features as a perceptual metric[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 586-595.

[78] Zhang L, Zhang L, Mou X, et al. FSIM: a feature similarity index for image quality assessment[J]. IEEE Transactions on Image Processing, 2011, 20(8): 2378-2386.

[79] 杨榆, 雷敏. 信息隐藏与数字水印[M]. 北京: 北京邮电大学出版社, 2017.

[80] Pandey B K, Pandey D, Wairya S, et al. Application of integrated steganography and image compressing techniques for confidential information transmission [J]. Cyber Security and Network Security, 2022, 3: 169-191.

[81] Morkel T. Image Steganography Applications for Secure Communication[M]. Pretoria: University of Pretoria (South Africa), 2012.

[82] Pandey D, Wairya S, Mahdawi R S, et al. Secret data transmission using advanced steganography and image compression[J]. International Journal of Non-

linear Analysis and Applications, 2021, 12: 1243-1257.

[83] 张雪锋. 信息安全概论[M]. 北京: 人民邮电出版社, 2014.

[84] 郑钢, 胡东辉, 戈辉, 等. 生成对抗网络驱动的图像隐写与水印模型[J]. 中国图象图形学报, 2021, 26(10): 2485-2502.
Zheng Gang, Hu Dong-hui, Ge Hui, et al. End-to-end image steganography and watermarking driven by generative adversarial networks[J]. Chinese Journal of Image and Graphics, 2021, 26(10): 2485-2502.

[85] 吴汉舟, 张杰, 李越, 等. 人工智能模型水印研究进展[J]. 中国图象图形学报, 2023, 28(6): 1792-1810.
Wu Han-zhou, Zhang Jie, Li Yue, et al. Research progress of artificial intelligence model watermarking [J]. Journal of Image and Graphics, 2023, 28(6): 1792-1810.

[86] 郑秋梅, 刘楠, 王风华. 基于复杂攻击的脆弱水印图像完整性认证算法[J]. 计算机科学, 2020, 47(10): 332-338.
Zheng Qiu-mei, Liu Nan, Wang Feng-hua. Complex attack based fragile watermarking for image integrity authentication algorithm[J]. Computer Science, 2020, 47(10): 332-338.

[87] 张文芳. 远程会诊中的医疗数据水印算法研究[D]. 南京: 南京信息工程大学计算机学院, 2023.
Zhang Wen-Fang. Research on medical data watermarking algorithm in remote consultation[D]. Nanjing: School of Computer Science, Nanjing University of Information Engineering, 2023.

[88] Zeng C, Liu J, Li J, et al. Multi-watermarking algorithm for medical image based on KAZE-DCT[J]. Journal of Ambient Intelligence and Humanized Computing, 2022, 15: 1735-1743.

[89] 刘泳彬, 徐涛, 苏锐豪, 等. 基于数字水印的医学图像信息隐藏系统[J]. 现代计算机, 2019(20): 42-45.
Liu Yong-bin, Xu Tao, Su Rui-hao, et al. A medical image information hiding system based on digital watermarking[J]. Modern Computer, 2019(20): 42-45.

[90] Anand A, Singh A K. Watermarking techniques for medical data authentication: a survey[J]. Multimedia Tools and Applications, 2021, 80: 30165-30197.

[91] Karakus S, Avci E. A new image steganography method with optimum pixel similarity for data hiding in medical images[J]. Medical Hypotheses, 2020, 139: No. 109691.

[92] Priyadharshini A, Umamaheswari R, Jayapandian N, et al. Securing medical images using encryption and LSB steganography[C]//International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), Bhilai, India, 2021: 1-5.

[93] Balu S, Babu C N K, Amudha K. Secure and efficient data transmission by video steganography in medical imaging system[J]. Cluster Computing, 2019, 22(Sup. 2): 4057-4063.

[94] 卢臻. AI换脸骗局频现 人工智能出圈后如何监管?[N]. 通信信息报, 2023-06-14.
Lu Zhen. How to supervise after AI face changing scam occurs frequently? [N]. Journal of Communication Information, 2023-06-14.

[95] 魏哲哲. “AI换脸”, 便利背后有风险[N]. 人民日报, 2023-12-04.
Wei Zhe-zhe. "AI changing faces", convenience behind the risk [N]. People's Daily, 2023-12-04.