中国矿业大学计算机学院

信息安全新技术报告

报告题目：_____生成式隐写研究_____

班　　级：_____信安 19-1 班_____

姓　　名：_____江一川_____

学　　号：_____08193041_____

任课老师：_____张艳群_____

2022.10

摘要

隐写术通常将秘密信息以不可见的形式隐藏到载体中, 从而通过传递含密载体实现隐蔽通信. 嵌入式隐写方案通过修改载体将秘密信息嵌入其中, 但会不可避免地改变载体的统计特性, 因此难以抵抗各类隐写分析工具的检测. 为了解决此问题, 生成式隐写方案以秘密信息为驱动直接生成含密载体. 相比于嵌入式隐写方案, 生成式隐写方案针对现有基于统计特征的隐写分析方法具有较好的抗检测性能, 因此逐渐成为信息隐藏领域的研究热点. 本文首先对两类生成式隐写方案进行详细地描述和分析, 包括: 1) 图像生成式隐写方案, 2) 文本生成式隐写方案; 其次, 通过实验详细分析和对比了各种图像生成式隐写方法的性能; 最后, 本文分析了当前生成式隐写方案仍然存在的问题, 并提出相应的解决方案和展望未来的发展方向.

关键词：生成式隐写;信息隐藏;数字水印;隐蔽通信;隐写分析;数字取证

目录

1 引言

隐写术(Steganography)通常将秘密信息以不可见的形式隐藏到载体中以实现隐蔽通信的目的.作为另一种常用的保障数据安全的技术,数据加密(Encryption)通常将秘密信息加密为无意义的密文形式, 但这也暴露了这些数据的重要性, 使得秘密信息容易遭到第三方的怀疑和拦截.与加密技术相比,隐写术不仅可以保护秘密信息的内容安全, 而且可以确保隐蔽通信的行为安全.

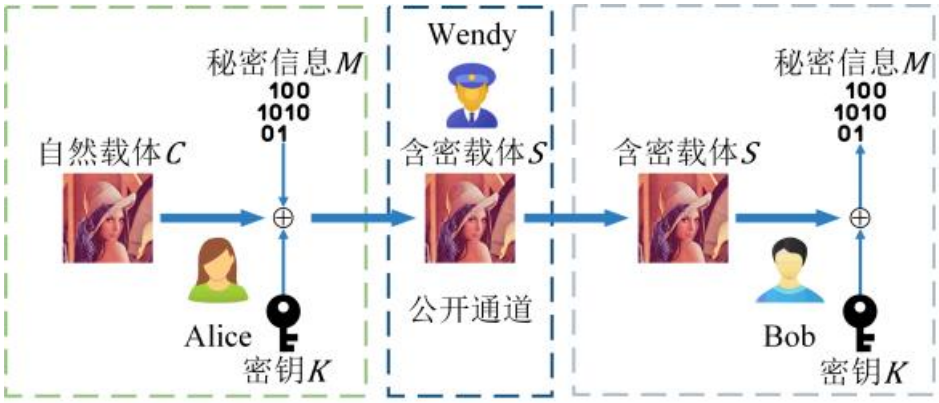

图 1 囚徒模型示意图

随着数字多媒体数据(如图像、视频、音频、文本等)的广泛使用,基于多媒体数据的隐写方案受到了广泛关注. 近二十多年来, 研究者们提出了大量的多媒体隐写方法.

在早期的隐写研究中, 研究者们大多聚焦于嵌入式隐写方案, 其中图像是嵌入式隐写方案中最常用的载体.作为图像嵌入式隐写的经典方案, 最低有效位(Least Significant Bit, LSB)隐写方法[4-6]通过修改图像载体中每个像素的最低有效位嵌入秘密信息. 然而,此类方法通常以相同的概率修改载体的每一个像素,容易导致含密载体大量失真,使得含密载体与自然载体在统计特性上的差异较大.为了减少含密载体的失真, 一些研究者通过编码的方式压缩秘密信息以减少需要修改的像素数量[7-9];一些研究者手工设计关于图像像素修改的失真函数,以最小化嵌入失真为目标,从而自适应地选择载体修改的位置[10-17];一些研究者使用神经网络学习出失真函数来代替手工方式设计的失真函数, 进一步降低了含密载体的失真度[18, 19].

受图像嵌入式隐写方法启发, 研究者们随后提出了基于文本和音频的嵌入式隐写方法. 例如: 基于文本格式的隐写方法[20-22], 基于文本内容的隐写方法[23-25], 基于音频时域的隐写方法[26-28], 基于音频频域的隐写方法[29-31]等.

作为隐写术的天生敌手,隐写分析(Steganalysis)技术旨在检测载体中是否存在隐藏的秘密信息.早期的隐写分析器从含密载体中提取统计特征[32-34], 然后将其输入到分类器中进行训练, 利用训练后的分类器判断载体中是否存在秘密

信息. 另一些研究者使用卷积神经网络(Convolutional Neural Networks, CNN)[35, 36]从载体中自动学习出更加有效的隐写分析特征[37]. 由于现有的嵌入式隐写方案不可避免地改变了载体的统计特性,为隐写分析提供了重要的检测依据,从而难以逃避隐写分析工具的成功检测.

为了抵抗隐写分析,研究者们提出了全新的隐写思路, 即“生成式隐写”, 具体是指:以秘密信息为驱动,直接“构造”或“生成”含密载体.近年来,研究者们提出了大量的多媒体生成网络模型,包括:循环神经网络(Recurrent Neural Network, RNN)[38,39]、变分自编码器(VariationalAuto-Encoder, VAE)[40, 41]、生成对抗网络 (Generative Adversarial Network, GAN)[42, 43] 、 流 模 型 (Flow-based Model)[44-46]等.这些模型生成的多媒体数据足以达到“以假乱真”的程度.这些性能强大的生成模型为生成式隐写的发展提供了良好的基础.

相比嵌入式隐写方案, 生成式隐写方案没有对现有载体进行任何修改, 因此可以较好地抵抗各类基于统计特性的隐写分析工具的检测; 同时, 生成的多媒体数据在网络中占比快速增长. 根据 Gartner①发布的消息, 预计 2025 年生成多媒体数据在网络中占比将达到 10%.因此采用生成的多媒体数据作为含密载体不容易引起怀疑与攻击. 由于以上因素,生成式隐写逐渐成为信息隐藏领域的研究热点, 引起了研究者们大量关注.

根据生成的含密载体类型,本文将现有的生成式隐写方案分为以下四大类: 1) 图像生成式隐写方案, 2) 文本生成式隐写方案, 3) 音频生成式隐写方案, 4) 社交网络行为生成式隐写方案;根据隐写方式的不同,每类生成式隐写方案又可以进一步细分为若干个子类.生成式隐写分类如图 2.

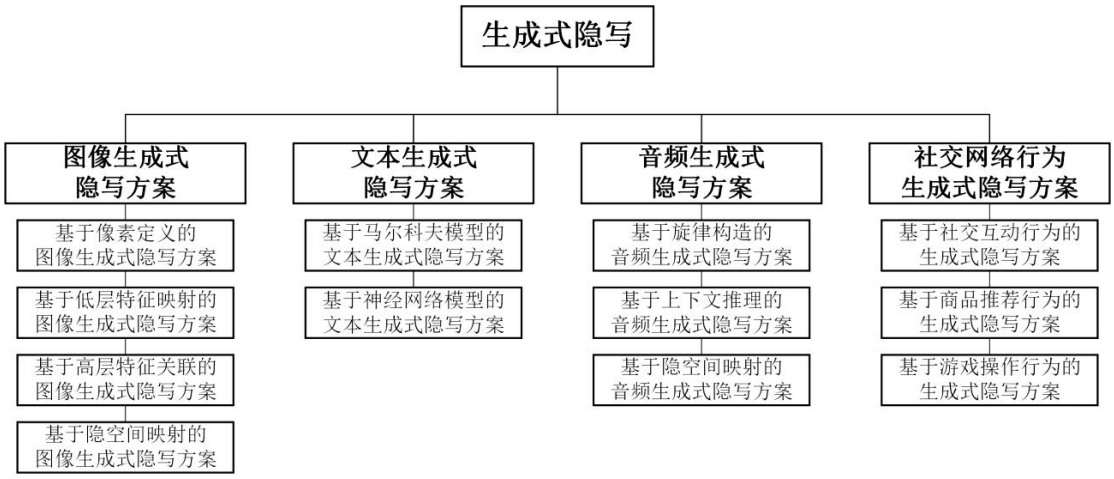

图 2 生成式隐写方案分类

本文在第2-3章详细描述了图像生成式隐写方案,总结和分析这些方法的优缺点; 第 4 章通过实验比较了各种图像成式隐写方法的性能; 第 5 章总结了目前生成式隐写方法存在的问题; 第 6 章展望未来的发展方向.

2 图像生成式隐写方案

根据秘密信息表达方式的不同，本文将现有的图像生成式隐写方案分为:基于像素定义的图像生成式隐写方案,基于低层特征映射的图像生成式隐写方案,基于高层特征关联的图像生成式隐写方案，以及基于隐空间映射的图像生成式隐写方案,如表 1 所示.

表 1 图像生成式隐写方案的分类对比

方案类别	主要思路	抗隐写分析性能	隐写容量	提取率	鲁棒性
基于像素定义的图像生成式隐写方案	将秘密编码为图像的像素值从而构造整幅图像；或者先映射为图像中部分指定位置像素的值，再补全图像中剩余部分	高	高 (9.8×10^{-3} ~4.303±0.850*bpp*)	高 (58%~100%)	低
基于低层特征映射的图像生成式隐写方案	将秘密信息映射为图像低层特征，生成具有此特征的含密图像	高	中 (2.0×10^{-3} ~3.28×10^{-2}*bpp*)	高 (70%~100%)	中
基于高层特征关联的图像生成式隐写方案	将秘密信息与图像高层特征相关联，生成具有此特征的含密图像	高	低 (5.069×10^{-5}~ 6×10^{-2}*bpp*)	高 (97.64%~100%)	高
基于隐空间映射的图像生成式隐写方案	通过可逆神经网络学习隐空间和图像空间之间的可逆映射函数，然后将秘密信息编码到隐空间向量，利用映射函数生成含密图像	高	高 (7.813×10^{-3}~8*bpp*)	高 (84%~100%)	中

2.1 基于像素定义的图像生成式隐写方案

基于像素定义的图像生成式隐写方案主要有两类. 一类方法将秘密信息编码为图像的像素值从而构造整幅图像；另一类方法先将秘密信息映射图像中部分指定位置像素值然后利用图像生成技术补全图像的剩余部分.

Yang 等[47]将秘密信息编码为像素值并构成整幅含密图像. 该方法使用像素卷积神经网络(Pixel Convolutional Neural Networks, PixelCNN)[48]对像素之间的依赖关系建模，从而可以根据已经生成像素获得当前待生成像素在[0,255]范围的取值概率分布.PixelCNN 通过每个像素的条件概率分布的乘积对图像进行建模:

$$p(x) = \prod_{i=1}^{n^2} p(x_i \mid x_1, x_2, \ldots\ldots, x_{i-1}) \tag{1}$$

其中$p(x_i \mid x_1, x_2, \ldots\ldots, x_{i-1})$是第$i$个像素$x_i$的条件概率分布；然后，使用拒绝采样的策略[49]，在该像素的概率分布中进行重复采样直到采样像素值满足以秘密信息为约束条件的值,并将该值作为待生成像素的值，这样使得秘密信息编码到该像素中;最后，按以上图像像素生成方法，从空白图像的以左上角到右下角为顺序逐个生成像素，最终构造出整幅含密图像.

Zhang 等[50]在 Yang 等[47]的基础上提出了 PixelStega 方法. Pixel-Stega 方法使用性能更为先进的 PixelCNN++[51]代替 PixelCNN，并通过基于算术编码的采样策略，根据秘密信息的内容和当前待生成像素的概率分布，确定对应的像素值. Pixel-Stega 方法的隐写模型如图 3 所示.由于使用了基于算术编码的采样策略，

Pixel-Stega 方法能够在含密图像熵值较大的区域隐藏较多的信息,而在熵值较小的区域隐藏较少的信息,即能够自适应地隐藏秘密信息,从而在给定的隐写负荷下,一定程度上提升了图像的生成质量.

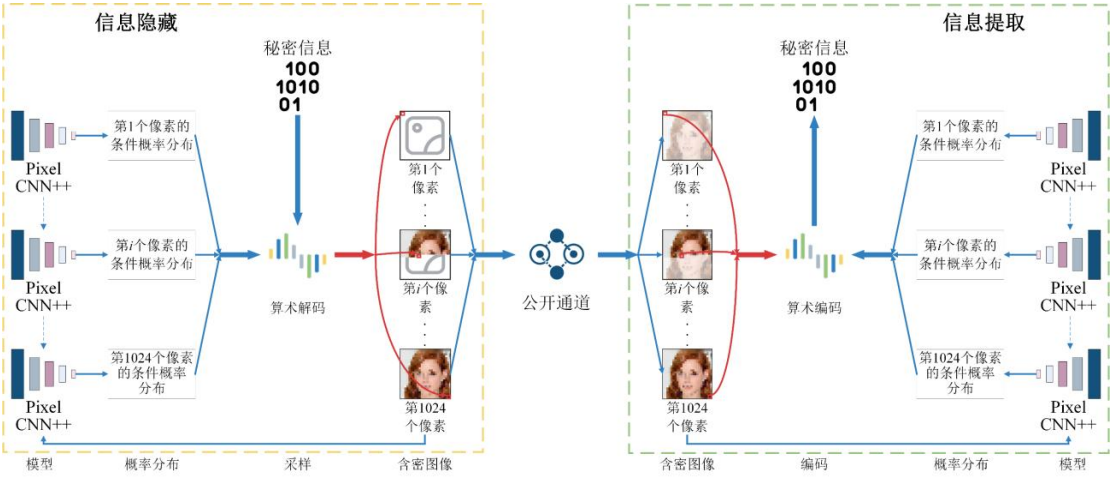

图 3 Pixel-Stega 隐写模型

以上两种方法对像素之间的依赖关系建模, 但它们根据已生成的像素生成当前像素来构造含密图像, 忽略了图像的全局信息, 使得含密图像难以保持语义的合理性. 因此, 以上两种方法生成的含密图像质量较低.

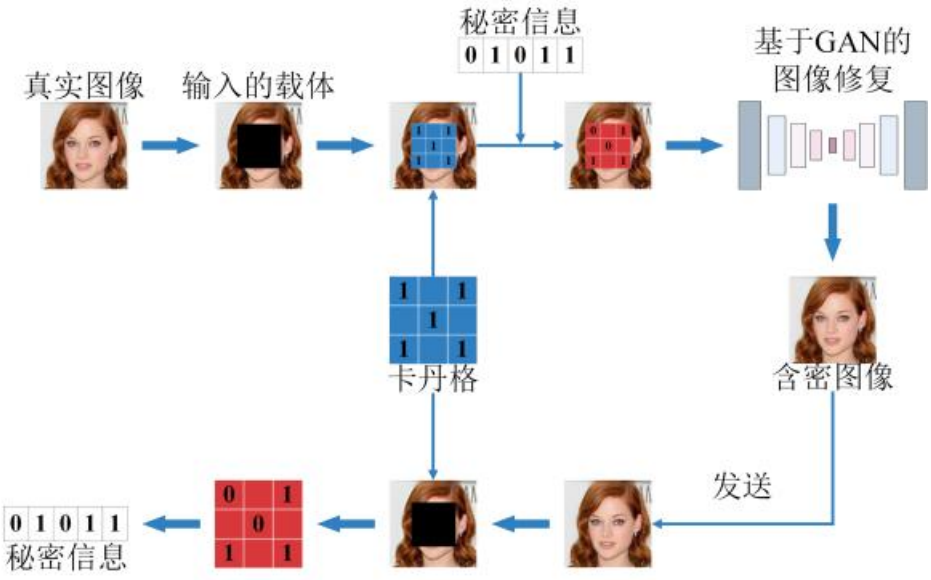

图 4 Liu 等[52]隐写模型

卡丹格(Cardan Grille)是一种西方早期用于隐藏信息的简单网格. 发送者以卡丹格为掩模, 将秘密信息预先写到书信等文本载体上的卡丹格网格指定位置, 然后根据已写好的秘密信息将其余文本补充完整, 从而在传递载体时不引起他人的怀疑, 接收者可以利用卡丹格为掩模从文本指定位置提取出秘密信息. 受卡丹格隐藏信息思路的启发, Liu 等[52]提出了一种利用卡丹格实现基于像素定义

的图像生成式隐写方法. 如图 4 所示, 该方案将含有受损区域的图像作为载体并设计了一个与图像受损区域相同形状的卡丹格掩膜, 将秘密信息填充到卡丹格掩膜中标记为“1”的区域; 然后将填充秘密信息后的受损图像输入到 DCGAN 中, 以填充的秘密信息保持不变为条件, 自动生成和恢复出受损区域以获得含密图像. 在秘密信息提取时, 将含密图像与卡丹格掩膜结合, 直接从卡丹格掩膜中的标记为“1”的区域中提取秘密信息. 该方法在信息隐藏过程中, 受损区域的恢复受到秘密信息需保持不变以及与未受损区域需保持语义一致的双重条件约束. 因此, 如果缩小未受损区域面积, 将弱化恢复受损区域时需与未受损区域保持语义一致性的约束, 有利于恢复受损区域时秘密信息保持不变, 从而提高秘密信息提取率. 按照以上思路, 一些研究者缩小未受损区域面积, 然后利用卡丹格将秘密信息隐藏在相应的位置, 并通过 DCGAN 补全受损区域, 从而达到了理想的秘密信息提取率[53, 54].以上方法以图像未受损区域作为参考信息, 能够使得补全的受损区域与未受损区域语义保持一致. 因此, 与 Yang 等[47]和 Pixel-Stega 隐写方法相比, 该方法所生成的含密图像质量有所提高.

表 2 基于像素定义的图像生成式隐写各方法对比

代表方法	主要思路	优点	缺点
Yang 等[47]	在使用 PixelCNN 生成像素的过程中, 通过拒绝采样的策略, 将秘密信息编码为对应像素值	隐写容量较高(1*bpp*)	图像的生成质量较低, 鲁棒性较低
Pixel-Stega[50]	在使用 PixelCNN++生成像素的过程中, 通过基于算数编码的采样策略, 将秘密信息编码为对应像素值	相比 Yang 等[56]方法, 图像的生成质量有所提高	图像的生成质量仍然较低, 鲁棒性较低
Liu 等[52]	利用卡丹格实现像素定义, 通过 DCGAN 修复受损区域以生成含密图像	相比 Yang 等[56]和 Pixel-Stega[50]方法, 图像的生成质量相对较好	秘密信息提取率较低, 鲁棒性较低
隐写方法[53][54]	在 Liu 等[61]方法基础上, 进一步缩小未受损区域, 实现基于卡丹格的生成式隐写	相比 Liu 等[61]方法, 该方法秘密信息的提取率有所提高(95%~100%)	鲁棒性较低

基于像素定义的图像生成式隐写各方法的对比如表 2 所示.由于隐写过程中没有对载体进行修改, 基于像素定义的图像生成式隐写方案能够有效抵抗现有基于统计特征的隐写分析工具的检测. 此外, 该方案将秘密信息映射为像素, 图像中的像素所能承载的信息量较大, 因此该方案的隐写容量相对较高. 然而, 图像在传输过程中有可能受到各种各样的攻击(如重压缩、添加噪声等), 这些常见攻击会对图像像素值的影响较大, 因此基于像素定义方法的鲁棒性普遍较低.

2.2 基于低层特征映射的图像生成式隐写方案

为了解决基于像素定义的图像生成式隐写的鲁棒性较低问题,部分研究者提出了基于低层特征映射的图像生成式隐写方案.该类方案将秘密信息映射为图像的低层特征(如纹理、轮廓等),然后生成具有此特征的含密图像.

Otori 等[55]提出了一种根据秘密信息生成纹理图像的隐写方法. 该方法首先建立局部二值模式(Local Binary Pattern, LBP)算子[56]的特征值与纹理中彩色点

之间的映射规则. 通过该映射规则,将秘密信息等值的 LBP 特征映射为相应的彩色点,并绘制在空白图像中的固定位置上;然后, 根据选定的图像纹理, 利用纹理合成技术对载体图像上的空白区域进行补全, 最终生成含密图像. 由于LBP 算子的特征值和彩色点对各种常见的图像攻击不敏感,因此该方法针对打印扫描和重拍摄等常见图像攻击具有较好的鲁棒性;然而, 由于 LBP 特征值维度较低,承载的信息容量受限, 因此该方法隐写容量偏低.

Xu 等[57]基于大理石纹理合成技术提出了 stego-texture 方法. 如图 5(a)所示,在隐写过程中,该方法首先将秘密字符显式地绘制在空白图像上; 然后根据秘密字符的颜色、曲率、方向、字号等属性添加背景线条得到背景图像, 生成的背景线条可以在不遮挡秘密信息的同时, 生成视觉自然的纹理图像, 生成的背景图像如图 5(b)所示; 最后, 如图 5(c)所示, 通过置乱操作把背景图像映射为最终的含密图像, 其中使用的置乱操作由七种可逆几何形变函数排列组合构成,而且置乱操作的参数也会被隐藏到构造的含密图像中; 接收者可以提取在含密图像中置乱操作的参数将隐写图像恢复为背景图像, 从而恢复出秘密信息, 如图 5(d)所示.

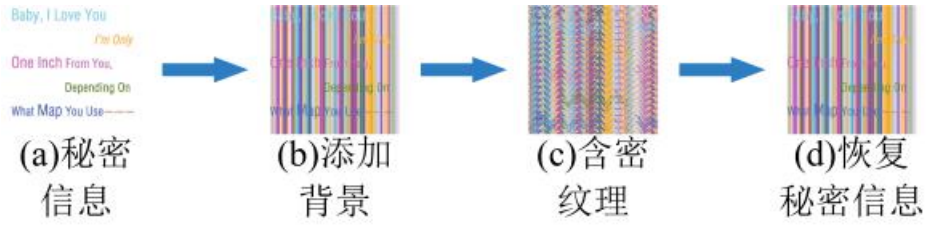

图 5　stego-texture 隐写模型

为了提升隐写容量,Wu 等[58]根据秘密信息选择相应的纹理图像块, 拼接成整幅纹理图像作为含密图像. 该方法选定一个图像纹理块,将此源纹理块分割并拓展为一系列候选图像块; 然后按照候选块之间的均方误差(Mean-Square Error, MSE)值排序,根据排序值划分到候选列表对应位置, 将每一候选列表映射为不同类型的秘密信息片段(候选列表建立的规则如图 6 所示); 接着根据秘密信息从相应的图像块候选列表中,选择相应的图像块依次填充到载体图像中的空白区域,从而生成整幅含密图像. 此外, 基于可逆纹理合成技术, 该方法可以从含密图像中准确恢复出源图像块并提取秘密信息.

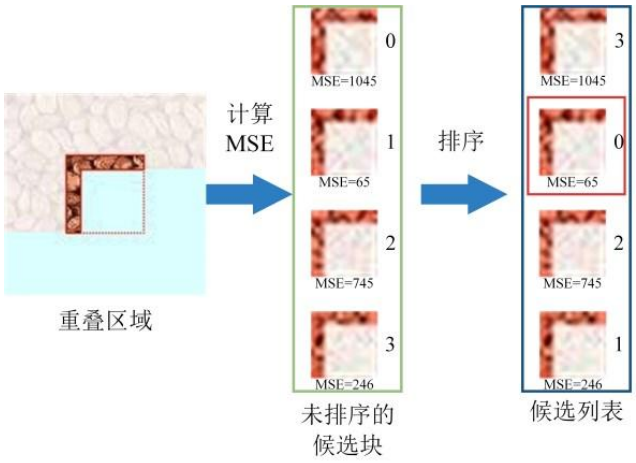

图 6　Wu 等[58]候选图像块列表的建立

指纹图像作为一种特殊的纹理图像，被广泛使用于各类实际应用中.Li 等[59]提出了以秘密信息为驱动构造指纹图像的隐写方法. Larkin 等[60]认为，任意一张自然指纹图像都可以被分解为偏置、幅度、全息相位和噪声四个部分，其中全息相位由螺旋相位和连续相位组成. 由于在螺旋相位中螺旋的位置与指纹图像中细节点的位置一致且随机性较强,Li 等[59]提出的方法将秘密信息及其纠错码[61]映射为螺旋相位中螺旋的位置并据此构建出螺旋相位；然后使用基于 Garbor 滤波的指纹生成模型[62]构建连续相位；最后将螺旋相位和连续相位合成为全息相位,并经过加噪、渲染等后期处理得到最终的高质量指纹图像. 由于指纹细节点的位置相对比较稳定，接收者可以直接根据细节点的位置反推出秘密信息，实现准确的秘密信息提取.

基于低层特征映射的图像生成式隐写各方法的对比如表 3 所示.由于图像的纹理、轮廓等低层特征相对稳定,不容易受到图像攻击的影响.因此相比基于像素定义的生成式隐写方法,基于低层特征映射的生成式隐写方法显著地提高了鲁棒性. 然而，与像素相比，纹理和轮廓所能承载的信息容量较低.因此,基于低层特征映射的生成式隐写方案隐写容量普遍低于基于像素定义的生成式隐写方案的隐写容量.

表 3　基于低层特征映射的图像生成式隐写各方法对比

代表方法	主要思路	优点	缺点
Otori 等[55]	将秘密信息编码为彩色点并绘制在空白载体图像上，然后补全载体图像剩余部分	具有较好的鲁棒性	隐写容量偏低，生成的纹理图像容易引起攻击者怀疑
stego-texture [57]	根据秘密信息的属性构造背景图像，并通过置乱操作把背景图像映射为含密纹理图像	具有较好的鲁棒性	生成的纹理图像容易引起攻击者怀疑
Wu 等[58]	根据秘密信息选择相应的纹理图像块构造出整幅含密纹理图像	相比 Otori 等[64]和 stego-texture 方法，隐写容量有所提高(3.28×10^{-2} *bpp*)，具有较好的鲁棒性	生成的纹理图像容易引起攻击者怀疑
Li 等[59]	根据秘密信息构造螺旋相位，并生成相应的指纹图像	具有较好的鲁棒性	生成的指纹图像容易引起攻击者怀疑
Zhou 等[63]	将秘密信息映射为轮廓信息并作为 GAN 的输入构造出相应的含密图像	能够生成自然图像具有较好的隐蔽性，具有较好的鲁棒性	隐写容量较低

2.3 基于高层特征关联的图像生成式隐写方案

为了进一步提升含密图像的生成质量,一些研究者提出了基于高层特征关联的图像生成式隐写方案. 该类方案将秘密信息编码为图像的特定高层特征, 并生成符合该特征的含密图像.

图像的语义信息作为常用的高层特征,被研究者应用于图像生成式隐写任务中. Cao 等[66]提出了一种基于动漫角色生成的图像生成式隐写方法. 该方法主要包括秘密信息与属性标签转换模块、图像生成模块、图像质量评估模块. 首先将秘密信息转换为二进制字符串;然后通过 LSTM 模型将秘密信息转换为动漫角色的属性标签集合(如发型、发色、瞳色等), 以该属性标签为作为 GAN 网络的输入条件生成动漫角色图像;最后,评估生成图像的质量,选择质量较高的图像作为含密图像. 接收者从含密图像中提取动漫角色的标签, 并将其转换为秘密信息.

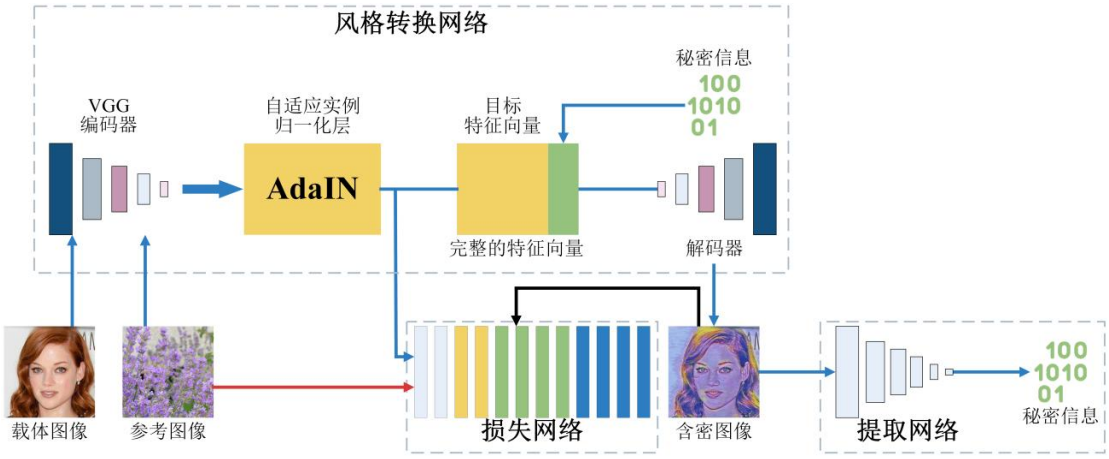

图 7　STNet 隐写模型

Wang 等[68]提出了基于风格迁移的深度隐写模型 STNet(Style Transformation Network for Deep Image Steganography),在图像风格转换的过程中将秘密信息隐藏到生成图像的风格特征中. 如图 7 所示,STNet 方法由风格转换网络、提取网络和损失网络三部分组成, 其中基于 VGG-19 模型[69]设计了编码器、解码器和损失网络. 在风格转换网络中, 发送方首先将载体图像和参考图像输入 VGG 编码器以获得载体图像内容特征和参考图像风格特征; 然后, 使用自适应实例归一化(Adaptive Instance Normalization, AdaIN)层[70]对载体图像内容特征和参考图像风格特征归一化, 以获得新的特征; 最后, 将秘密信息与获得的新特征拼接成为完整的特征, 然后输入到解码器中从而获得含密图像; 损失网络是预训练的 VGG-19 网络, 通过评估含密图像的内容特征和风格特征与输入图像的差异, 确保含密图像的内容与风格特征与输入图像一致; 提取网络由六个包含批归一化和 Leaky ReLU 激活函数[71]的卷积层构成, 接收者将含密图像输入到提取网络便可以提取秘密信息.STNet 整体的损失函数由内容损失L_C, 风格损失L_S, 以及秘密信息重建损失L_M, 经过加权构成, 其可表示为:

$$L_{STNet} = L_C + \lambda L_S + \mu L_M \quad (2)$$

其中λ和μ分别是风格损失和秘密信息重建损失的权重的系数.

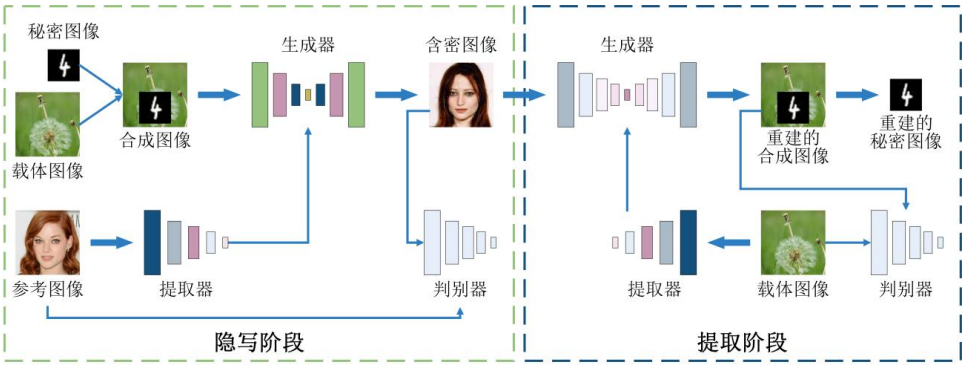

图 8　Li 等[73]隐写模型

循环一致性生成对抗网络(Cycle-consistency Generative Adversarial Networks, Cycle-GAN)[72]可以实现两个不同图像域的转换, 其结构是由两个生成器和两个判别器组成.受 Cycle-GAN 在风格迁移相关研究的启发, Li 等[73]提出了一种基于风格迁移的图像生成式隐写方法,通过图像风格的迁移网络, 将秘密图像与另一幅图像合成为一幅指定风格的含密图像,如图 8 所示. 该方法在隐写阶段和提取阶段各设计了一个生成器、一个提取器以及一个判别器.其中, 这两个阶段的提取器与判别器结构相同,分别采用了 VGG-16[69]和 DCGAN 结构. 在隐写阶段, 该方法首先将秘密图像放置在载体图像上组成合成图像, 然后将该合成图像和参考图像输入到 Cycle-GAN 的生成器中获得含密图像,这样使得该含密图像的风格与参考图像风格一致;在提取秘密图像的过程中,该方法使用提取器从原载体图像提取特征,将该特征和含密图像作为 Cycle-GAN 的另一生成器为输入,生成和恢复出合成图像.并且在恢复时需要融合原始载体图像的特征, 使恢复的合成图像与原始载体图像风格一致.

基于高层特征关联的图像生成式隐写的各方法的对比如表 4 所示.相比纹理、轮廓等图像低层特征, 图像高层特征更加稳定, 更不容易受到图像攻击(如添加噪声等)的影响, 因此基于图像高层特征关联的生成式隐写方法鲁棒性更高. 然而, 从图像中抽象出来的高层特征所能承载的信息量较低,因此基于高层特征关联的生成式隐写方案的隐写容量会明显低于基于像素定义以及基于低层特征映射的生成式隐写方案.

表 4 基于高层特征关联的图像生成式隐写各方法对比

代表方法	主要思路	优点	缺点
Cao 等[66]	将秘密信息转换为动漫角色的属性标签集, 并利用 GAN 生成动漫角色图像	秘密信息提取率较高(100%), 鲁棒性较高	隐写容量非常低
SSS-GAN [67]	将秘密信息与图像的语义标签信息构建映射关系	秘密信息提取率较高(100%), 鲁棒性较高	隐写容量非常低
STNet [68]	提取载体图像内容特征和参考图像风格特征, 与秘密信息拼接得到的特征向量输入解码器中, 将载体图像转换为具有参考图像风格特征的含密载体图像	秘密信息提取率较高(99.8%), 鲁棒性较高	相比 Cao 等[75]和 SSS-GAN [67]方法, 隐写容量有所提高但仍然偏低
Li 等[73]	将秘密图像与另一张图像合成得到合成图像, 然后转换为具有参考图像风格特征的含密载体图像	秘密信息提取率较高(97.64%), 鲁棒性较高	相比 Cao 等[75]和 SSS-GAN [67]方法, 隐写容量有所提高但仍然偏低

2.4 基于隐空间映射的图像生成式隐写方案

一些研究者发现自然图像通常服从特殊的复杂分布, 通过神经网络能够学习到图像空间分布与某隐空间分布(如高维高斯分布等)的映射规则[44, 46]. 根据以上研究成果启发, 基于隐空间映射的图像生成式隐写方案通过构建秘密信息与隐空间向量的映射规则, 将秘密信息映射为隐向量并转换为相应的图像. 该图像可以作为含密图像从而实现隐蔽通信.

Hu 等[74]提出了一种将秘密信息映射为隐向量的图像生成式隐写方法.如图 11 所示, 该方法首先将秘密信息编码为 DCGAN 生成器的输入低维噪声向量, 利用该生成器生成载体图像;并训练 CNN 模型作为提取器, 确保提取的噪声向量与原始的噪声向量一致,从而恢复出秘密信息. 训练好的生成器和提取器共同实现了在隐空间和图像空间之间相互映射.

由于 DCGAN 生成的一些图像不够自然, Li 等[75]使用 WGAN-GP[76]代替 DCGAN 以生成更加真实的含密图像.

Zhang 等[77]提出了一种图像生成式隐写网络 GSN(Generative Steganography Network), 不同于现有的大多数方法将秘密信息映射为隐向量然后利用该隐向量生成含密图像,该方法在通过隐向量生成图像的过程中隐藏秘密信息. GSN 由生成器、判别器、隐写分析器和提取器组成.在隐写阶段, 发送方首先将隐向量输入生成器获得特征图,然后将二进制秘密信息通过数据合并操作(data merging operation)添加到特征图中,从而可以利用添加秘密信息后的特征图生成含密图像; 最后, 发送方同时将生成器生成的载体图像和含密图像通过无损信道传递给接收方. 在提取阶段, 接收方使用提取器从含密图像中提取秘密信息.

GAN 模型训练时, 隐空间的维数通常远低于图像空间的维数,而提高隐空间的维度将会导致模型难以训练. 因此, Hu 等[74]隐写方法和 Li 等[75]隐写方法的隐写容量受到了限制. 由于 GSN 方法在通过隐向量生成图像的过程中隐藏秘密信息,避免了隐向量低维度或图像语义标签低信息量的限制,能够极大地提升隐写容量和含密图像生成质量. 然而,由于 GAN 模型不能实现隐空间到图像空间的可逆变换, 需要额外训练 CNN 模型提取器, 因此 Hu 等[74]隐写方法、Li 等[75]隐写方法和 GSN 方法的秘密信息的提取准确率需进一步提高.

Liu 等[78]提出了基于图像解耦自编码器(Image Disentanglement Autoencoder for Steganography, IDEAS)的隐写方法. 在隐写过程中,发送方将秘密信息映射为隐向量, 并将其转换为图像的结构特征; 然后发送方将该结构特征与随机采样获得的图像纹理特征作为生成器的输入以生成含密图像. 在秘密信息提取的过程中, 接收方首先训练编码器从含密图像中恢复结构特征; 然后训练解码器,根据结构特征恢复隐向量;最后将该隐向量逆向映射为秘密信息. 由于图像的结构特

征较为稳定，因此 IDEAS 提升了秘密信息提取的准确率.此外,由于加入了随机采样得到的纹理特征,IDEAS 可以生成具有多样性的含密图像，可以避免针对特定的秘密信息只能生成单一的含密图像的问题,从而提升了隐写的安全性.

IDEAS 方法采用 Autoencoder 网络实现秘密信息的提取任务.然而，由于该过程涉及池化和归一化操作，容易丢失图像内部的细节信息. 因此，该方法依旧不能保证秘密信息的精确提取.

基于隐空间映射的图像生成式隐写的各方法的对比如表 5 所示. 基于隐空间映射的图像生成式隐写方案根据秘密信息从隐空间中采样,可以获得较高的隐写容量.常见的图像攻击有可能使得隐空间的向量值受到一定程度的影响，与高层特征关联的图像生成式隐写方案相比，该方案的鲁棒性有所降低.然而，值得注意的是,基于隐空间映射的生成式隐写方案的模型训练目标是将高斯分布拟合为自然图像在图像空间中的分布，而现有的方法构建的隐向量都是服从高斯分布的,实际上自然图像对应的隐向量分布只是近似的高斯分布，设计隐空间的检测器工具可以根据此现象仍然有可能检测到隐写的存在，将一定程度上影响隐写信息的安全性.

表 5 基于隐空间映射的图像生成式隐写各方法对比

代表方法	主要思路	优点	缺点
Hu 等[74]	将秘密信息映射为噪声隐向量，分别训练 DCGAN 生成器生成含密图像和提取器提取秘密信息	能生成质量较好的含密图像	生成的含密图像类型比较单一，且无法实现秘密信息准确提取
Li 等[75]	同时训练生成器和提取器，并使用 WGAN-GP 代替 DCGAN	能生成质量较好的含密图像	生成的含密图像类型比较单一，且无法实现秘密信息准确提取
GSN[77]	将隐向量输入生成器获得特征图，然后将二进制秘密信息张量通过数据合并操作加入到特征图中利用加入秘密信息后的特征图生成含密图像	隐写容量较高(8*bpp*)，能生成质量较好的含密图像	生成的含密图像类型比较单一，且无法实现秘密信息准确提取
IDEAS[78]	秘密信息映射为隐向量，将其转换为图像的结构特征并与随机采样的纹理特征输入到 Autoencoder 的编码器生成含密图像，然后含密图像输入到 Autoencoder 的解码器提取秘密信息	隐写容量较高(7.813×10^{-3} *bpp*)，提取率较高(100%)，能够生成种类丰富的含密图像	无法实现秘密信息准确提取
S2IRT[79]	将秘密信息编码为隐向量，通过 Glow 模型转换为含密图像，可以通过 Glow 模型逆变换提取秘密信息	隐写容量较高(4*bpp*)，提取率较高(100%)	生成的含密图像类型比较单一

由于具有较好的抗隐写分析性能，图像生成式隐写方案已经成为生成式隐写领域的研究热点.研究者们基于不同的映射规则将秘密信息转换为合适的数据类型从而生成含密图像.然而，虽然现有的图像生成式隐写方案所使用的生成模型能够生成较高质量的含密图像，但也经常生成一些低质量图像. 特别是在隐写荷载较大的情况下，图像生成质量不高.因此，现有的图像生成式隐写方法的图像生成质量仍然不够稳定.

3 文本生成式隐写方案

早期的文本隐写方案大多数是通过修改文本实现信息隐藏的嵌入式隐写方

案[25, 80].然而,相比图像载体,文本载体的数据量小,存在的冗余修改空间有限,因此修改文本的嵌入式隐写方式不利于大量秘密信息的隐写.随着自然语言处理技术的发展,文本生成模型生成的文本质量越来越高,这为文本生成式隐写的发展奠定了坚实的基础.本节根据生成文本所使用生成模型的不同,将现有的文本生成式隐写方案分为以下两类: 基于马尔科夫模型的文本生成式隐写方案和基于神经网络模型的文本生成式隐写方案.文本生成式隐写方案的分类对比如表 6 所示.

表 6 文本生成式隐写方案的分类对比

方法类别	主要思路	优点	缺点
基于马尔科夫模型的文本生成式隐写方案	对自然文本中每个单词出现的频率进行统计，获得单词的概率，然后根据秘密信息选择相应概率的单词生成含密文本	隐写容量较大	难以完全统计单词条件概率并建立理想的文本生成模型，难以生成质量较高的含密文本
基于神经网络模型的文本生成式隐写方案	利用神经网络学习自然文本的生成模型并自动生成文本，在生成过程中根据秘密信息选择相应条件概率的单词生成含密文本	隐写容量较大，并提升了含密文本的生成质量	当隐写荷载较大时，难以保证含密文本的生成质量

3.1 基于马尔科夫模型的文本生成式隐写方案

马尔科夫模型可以根据自然文本中邻近单词 的出现频次对文本进行建模.因此,早期的文本生成式隐写方法大多使用马尔科夫模型生成含密文本.该类方案首先通过马尔科夫模型对自然文本中每个单词出现的频率进行统计, 近似获得单词的概率, 然后根据秘密信息选择相应概率的单词以生成含密文本, 在生成文本的过程中实现隐写.

部分研究者通过对单词进行哈夫曼编码以提高文本的生成质量.Yang 等[82]提出了一种基于马尔科夫模型和哈夫曼编码的文本生成隐写方法.如图 9 所示,在使用马尔科夫模型生成文本的基础上,对每个单词使用哈夫曼树进行动态编码,根据秘密信息选择相应的单词以生成含密文本.哈夫曼编码可以依据单词的条件概率分布来构造平均长度最短的编码, 使编码概率较高的单词编码长度更短.该方法在编码过程中充分考虑了每个单词的条件概率分布, 从而在一定程度上提高了隐写容量.

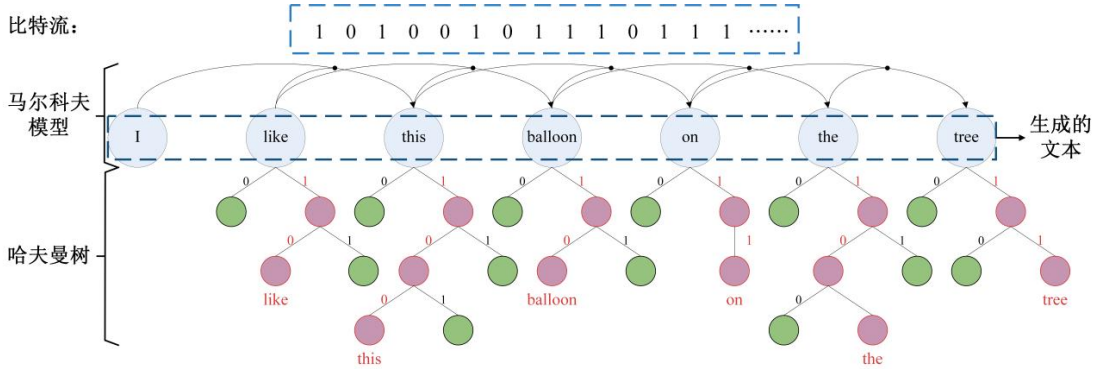

图 9 Yang 等[82]隐写模型

部分研究者认为生成某种特殊题材的文本能够使得生成模型生成更高质量的文本.中国古代宋词作为一种特殊题材，每首词都有固定的音律和句长,生成一首词相当于根据特定的音律选择合适的单词.Luo 等[83]提出了基于词的文本生成式隐写方法(Ci-Based Steganography Methodology, Cistega).Cistega 方法使用马尔科夫模型生成文本，并在预定音律的约束下,选择合适的单词构成含密文本.在隐写过程中，该方法首先确定初始单词、生成词的音律以及候选池的容量，其中候选池用于存储符合设定音律的当前单词；然后，根据单词的条件概率，选择符合设定音律的单词存储到候选池中；最后对候选池中的单词进行编码，选出编码与秘密信息比特流匹配的单词以构成含密文本，从而生成具有设定音律的含密文本. 生成特殊题材的文本虽然能够获得更高的生成质量,但是由于题材的特殊性,所以难以获取足够大的语料库,可能会导致生成过程中出现断链,即生成某一个单词后没有合适的下一单词能够被选择.虽然 Cistega 方法通过降低阈值、缩短生成长度、直接选取高频次的策略能在一定程度上缓解以上问题,但不能彻底解决该问题.除古诗词以外,个人笔记[84]、幽默短文[85]、互联网协议文本[86, 87]等特殊类型的文本同样也被用于生成式隐写.

表 7 基于马尔科夫模型的文本生成式隐写各方法对比

代表方法	主要思路	优点	缺点
Shniperov 等[81]	在使用马尔科夫模型生成文本过程中，根据秘密信息选择相应概率值的单词以生成含密文本	隐写容量较高	含密文本生成质量不高
Yang 等[82]	在使用马尔科夫模型生成文本过程中，对每个单词进行哈夫曼编码，根据秘密信息选择相应的单词以生成含密文本	相比 Shniperov 等[80]方法，进一步提高了隐写容量(每个单词隐藏 1-4 位)	含密文本生成质量不高
Cistega[83]	在使用马尔科夫模型生成文本过程中，在预定音律的约束下，根据单词的条件概率选择合适的单词构成含密文本	生成诗歌等特定题材的文本具有更高的生成质量	含密文本生成过程中容易出现断链

基于马尔科夫模型的文本生成式隐写方法的对比如表 7 所示.基于马尔科夫模型的文本生成式隐写方案能够在一定程度上可以保证生成含密文本的质量.然而，该方案只通过计算前一个或多个单词出现时当前单词出现的频率来计算该单词条件概率,即只考虑当前单词之前的几个单词，而不是所有单词,因此马尔科夫模型不能准确地对自然文本建模.此外，在文本生成式隐写过程中,马尔科夫模型并没有计算文本中单词实际的条件概率，而是利用统计频率近似作为条件概率.因此，马尔科夫模型难以获得准确的条件概率以及理想的文本生成模型，从而难以生成质量较高的含密文本.

3.2 基于神经网络模型的文本生成式隐写方案

随着深度学习技术的发展，神经网络模型成为解决文本生成领域各种问题的主流方法，更多的研究者聚焦基于神经网络模型的文本生成式隐写方案.该方案通常利用神经网络从大量真实文本样本学习文本模型，在生成过程中根据秘密信息选择相应条件概率的单词，从而实现秘密信息的隐写任务.

Zhou 等[93]提出了一种基于随机候选池的文本生成式隐写方法.该方法不选择概率值最大的单词组成候选池, 而是随机选择概率值在一定范围的单词组成候选池, 有效缓解了生成文本与自然文本的统计偏差. 在隐写过程中, 该方法利用 LSTM 对当前单词进行预测, 获得当前单词的条件概率分布, 然后随机选择概率值在一定范围的单词组成候选池并过滤一些低频单词, 最后根据秘密信息从候选池中选择相应的单词生成含密文本. 由于组成候选池是随机选择概率值在一定范围的单词并过滤了低频单词, 避免了只选择概率值最大的单词组成候选池所导致生成的含密文本低频单词较多的问题,从而保证了生成的文本分布尽可能接近真实文本分布. 因此该方法一定程度上提高了含密文本的生成质量.

为了解决生成长含密文本时易出现的语义不连贯和语义错误等生成文本质量问题,Cao 等[94]提出了 PPLM-Stega 隐写方法, 该方法基于 PPLM 文本生成模型(Plug and Play Language Model, PPLM)[95].所使用的 PPLM 能够以较少的计算资源生成基于属性模型的合理可控文本.PPLM-Stega 首先改进了 PPLM, 在 PPLM 的输出层之前增加了隐写层以生成含密文本.在传输秘密信息时,PPLM-Stega 首先根据通信各方之间共享的密钥来确定含密文本的主题.然后, 使用 PPLM 生成符合主题且具有语义一致性的长可读含密文本.在文本生成过程中,隐写层首先确定具有最高概率的候选词的概率,然后计算其他候选词的概率与最高概率候选词
的比例.如果该比例大于预设的阈值,则将其添加到可嵌入的候选词池(Embeddable Candidate Word Pool,ECWP),根据 ECWP 的大小来确定可隐藏的秘密信息比特数.最后,根据编码后的秘密信息,选择当前的单词, 从而生成含密文本.PPLM-Stega 能够在生成语义连贯可读性高的长文本同时, 具有较高的隐写容量.

Yi 等[96]基于 BERT 生成模型和 Gibbs 采样策略[97]提出了一种文本生成式隐写方法 ALiSa(Acrostic Linguistic Steganography Based on BERT and Gibbs Sampling), 该方法直接将秘密文本的单词隐藏到生成文本的指定位置中以隐蔽地传递秘密文本. 在隐写过程中, 发送方首先根据秘密文本中的单词和给定的位置密钥, 利用 BERT 生成模型生成初始文本, 在生成的初始文本中的特定位置单词使用掩膜代替, 并且 BERT 为每个掩膜位置计算了单词的条件分布概率; 然后根据每个掩膜位置的条件概率和 Gibbs 采样策略, 从秘密文本中选择合适的单词, 放置到文本的掩模位置; 最后通过多次迭代计算, 利用 BERT 模型在保证生成质量的条件下生成含密文本. 在提取过程中, 接收者可以根据位置密钥直接从含密文本中提取秘密信息.

最近, 一些研究者利用神经网络对古诗词进行建模, 提出了基于神经网络的

古诗词生成式隐写方法. Qin 等[98]提出了一种基于绝句生成的文本生成式隐写方法. 该方法利用基于注意力机制的 Seq2Seq 模型[99]生成符合设定主题和音律的绝句, 并在生成绝句的过程中将秘密信息隐藏到绝句中, 从而生成含密文本. 在训练绝句生成模型过程中, 该方法将诗句的平仄和音律信息作为约束, 使得模型能够生成高质量的文本.在隐写阶段, 发送方将秘密信息分为两段, 根据第一段秘密信息确定主题词和音律等参数, 将参数输入到生成模型生成候选诗句集; 然后, 根据第二段秘密信息从候选诗句中选择相应的诗句以构成含密文本. 在提取阶段, 接收方使从含密文本中提取相关参数, 从而根据参数提取第一段秘密信息, 将参数输入到生成模型以生成候选诗句集; 然后,从候选诗句集中检索到含密文本的每句候选诗句, 根据候选诗句在候选诗句集中的序号提取到第二段秘密信息, 最终拼接两段秘密信息获得完整的秘密信息.

Qin 等[100]提出了 SongNet 方法, 首先构建秘密信息与宋词的音律信息的映射关系,然后基于改进的 BERT 模型, 在利用设定的音律生成宋词的过程中, 从而实现秘密信息在宋词中的隐藏. 在隐写阶段, 发送者将秘密信息分为两段, 首先, 发送方根据第一段秘密信息确定待生成宋词的词牌信息和格式; 然后,根据第二段秘密信息确定待生成宋词的关键字、韵律和押韵字符, 从而利用改进的 BERT 模型根据确定的词牌信息生成具有相应音律信息的含密文本. 在提取阶段, 接收方首先提取含密文本中词牌信息, 在词牌信息表检索以获得对应的词牌格式, 从而可以根据词牌格式模板恢复第一段秘密信息; 然后利用提取的词牌格式从含密文本中提取出关键字、韵律以及押韵字符以恢复第二段秘密信息; 最后将两段秘密信息拼接获得完整的秘密信息.

基于神经网络模型的文本生成式隐写的各方法的对比如表 8 所示.随着深度学习技术的快速发展,文本生成式隐写方案的生成质量得到了显著的提升.

相比传统的文本嵌入式隐写方案,文本生成式隐写方案能够抵抗隐写分析工具的检测,从而保证了含密文本传递过程中的隐蔽性.然而,现有的模型难以对自然文本完美地建模,导致所生成的含密文本与自然文本的统计特性仍然存在一定差距,含密文本生成质量有待进一步提高, 尤其当隐写负载较高时, 难以保证含密文本的生成质量.

表 8 基于神经网络模型的文本生成式隐写各方法对比

代表方法	主要思路	优点	缺点
RNN-Stega[88]	利用 RNN 对下一个单词进行预测，将预测的单词的条件概率编码为二进制比特流，以此选择对应的单词，将整个秘密信息隐藏到生成的文本中	隐写容量较高(每个单词隐藏 1-5 位)	没有充分考虑自然文本与生成的含密文本之间的总体统计分布差异，文本生成质量相对较低
GAN-TStega[89]	利用 GAN 网络对不同类型文本数据集的对抗训练以生成高质量文本，在文本生成过程中隐藏秘密信息	隐写容量较高(每个单词隐藏 1-5 位)，在 RNN-Stega[87]基础上，进一步提高文本生成质量	生成的含密文本的语义是随机的、不可控的
VAE-Stega[91]	从隐空间中采样一个隐向量，在该隐向量的约束下，使用 AutoEncoder 的解码器根据秘密信息选择对应的候选单词生成含密文本	隐写容量较高(每个单词隐藏 1-5 位)，在 RNN-Stega[87]基础上，进一步提高文本生成质量	生成的含密文本的语义是随机的、不可控的
Zhou 等[93]	利用 LSTM 对当前单词进行预测，随机选择条件概率值在一定范围的单词组成候选池，从而根据秘密信息在候选池中选择相应单词生成含密文本	隐写容量较高(每个单词隐藏 1-3 位)，在 RNN-Stega[87]基础上，进一步提高文本生成质量	生成的含密文本的语义是随机的、不可控的
PPLM-Stega[94]	计算其他候选词的概率与最高概率候选词的比例，选择该比例大于预设的阈值的单词添加到候选池，并根据秘密信息从候选池中选择相应的单词生成含密文本	隐写容量较高，可以生成长文本	在高隐写荷载的条件下，难以保证含密文本的生成质量
ALiSa[96]	在使用 BERT 生成文本的过程中，根据生成文本中指定位置的单词分布条件概率和秘密信息中选择相应的单词，隐藏到文本的指定位置中以生成含密文本	隐写容量较高	采样策略需进一步优化
Qin 等[98]	将秘密信息分为两段，首先根据第一段秘密信息生成候选诗句，然后根据第二段秘密信息从候选诗句中选择相应的诗句以构成含密文本	隐写容量较高，可生成高质量的绝句诗	只能生成特殊题材的文本，适用性较低
SongNet[100]	将秘密信息分为两段，首先根据第一段秘密信息确定待生成宋词的词牌信息，然后根据第二段秘密信息确定待生成宋词的关键字、韵律和押韵字符，从而生成具有相应音律信息的含密文本	隐写容量较高(每个词句隐藏 18-21 位)，可生成高质量的宋词	只能生成特殊题材的文本，适用性较低

4 实验对比与分析

本章主要对具有代表性的图像生成式隐写方案的实验结果进行对比与分析.根据各隐写方法的隐写容量、秘密信息的提取率以及在各种噪声(如高斯噪声、椒盐噪声)攻击下的提取率,本文对基于像素定义的图像生成式隐写方案、基于低层特征映射的图像生成式隐写方案、基于高层特征关联的图像生成式隐写方案以及基于隐空间映射的图像生成式隐写方案的实验结果进行分析.表 11-14 为部分图像生成式隐写方案的隐写算法性能对比.

4.1 基于像素定义的图像生成式隐写方案

Pixel-Stega 方法使用 MNIST[119]、FreyFaces[50]以及 CIFAR-10[120]数据集与 Yang 等[47]的隐写方法进行了对比实验.为了进行定量评估,使用 Pixel-Staga 生成了 5,000 幅载体图像和 5,000 幅含密图像并使用 Yang 等[47]生成了 5,000 幅含密图像.根据实验结果，相比 Yang 等[47]的隐写方法在三个数据集上均为 1.0bpp 的隐写容量, Pixel-Stega 方法的隐写容量最高可以达到 4.3bpp. 虽然 Pixel-Stega 方法在 MNIST 数据集上的隐写容量低于 Yang 等[47]方法，但从整体上看，Pixel-Stega 方法的隐写容量大于 Yang 等[47]方法的隐写容量. 这是因为 Pixel-Stega 方法基于图像像素熵自适应地隐藏的秘密信息. 由于 MNIST 数据集中的图像均为黑白图像，大多数像素的熵较低，而 Frey Faces 和 CIFAR-10 数据集中的图像更加多样化，它们的像素具有较大的熵, Pixel-Stega 方法可以充分隐藏秘密信息. 实验结果表明, Pixel-Stega 方法能够显著提高秘密信息的隐写容量.

针对 Liu 等[52]的隐写方法使用 LFW 数据集[121]进行相关实验.该方法的受损区域由受损图像中未受损部分和秘密信息填充部分组成,受损区域的面积约占整个图像区域的 12.5%, 当隐写容量为 2.5×10-3bpp 时, 提取率约为 58%; 在隐写方法[53]的实验中, 受损区域完全是由秘密信息填充部分组成, 受损区域的面积占整个图像区域的 90%, 当隐写容量为 5×10-1bpp 时, 提取率约为 95%; 而隐写方法[54]使用 CelebA 和 LSUN[122]数据集进行相关实验,与隐写方法[53]相同, 该方法受损区域完全是由秘密信息填充部分组成,受损区域的面积占整个图像区域的 95%, 将 20 字节的秘密信息隐藏到卡丹格中, 最大的隐写容量约为 9.8×10-3bpp, 最大的提取率为 100%. 实验结果表明, 对于使用卡丹格对像素进行预定义的图像生成式隐写方法, 缩小未受损区域的面积有利于提升秘密信息的提取率.

表 9 基于像素定义的图像生成式的性能对比

代表方法	数据集	隐写容量	提取率	噪声攻击后的提取率
Yang 等[47]	MNIST	1.0000±0.0000*bpp*	100%	41.76%
	Frey Faces	1.0000±0.0000*bpp*	100%	52.96%
	CIFAR-10	1.0000±0.0000*bpp*	100%	48.62%
Pixel-Stega[50]	MNIST	0.5840±0.1879*bpp*	100%	48.07%
	Frey Faces	4.0479±0.1403*bpp*	100%	58.39%
	CIFAR-10	4.3028±0.8502*bpp*	100%	50.81%
Liu 等[52]	LFW	2.5×10^{-1}*bpp*	58%	32.94%
隐写方法[53]	LFW	5×10^{-1}*bpp*	95%	57.65%
隐写方法[54]	CelebA、LSUN	9.8×10^{-3}*bpp*	100%(最大)	60.18%

基于像素定义的图像生成式隐写方案的性能对比如表 9 所示. 以上方法的隐写容量整体上处于较高的水平, 这是因为基于像素定义的图像生成式隐写方案将秘密信息映射为像素, 而图像中的像素所能承载的信息量较大. 然而, 由于像素值对噪声攻击的较为敏感, 以上方法在受到攻击后的提取率会大大降低. 因此, 此类方法的鲁棒性较低, 难以抵抗各种噪声攻击.

4.2 基于低层特征映射的图像生成式隐写方案

Otori 等[55]的隐写方法用 EPSON PX-G5100 彩色喷墨打印机在超薄的 A4 纸上在 2×2 英寸的正方形区域内打印 200×200 像素的数据编码后的纹理图像,并打开手机的摄像头以 480×640 像素的微距模式拍摄打印的图像.通过对 4 种不同纹

理的例子进行 10 次测试,该方法可以隐藏 200-800 位信息,因此其最大隐写容量约为 2×10-3bpp,经过打印和拍摄攻击后提取率约为 90%.

Stego-texture 方法使用 7 种大理石纹理图案的样本以隐藏秘密信息,其将秘密信息转换为纹理图像,并从 512×512 像素缩小到 180×180 像素,然后把该图像编码为 777,600 位的二进制字符串以隐藏到大理石纹理图案中. 经过检测, 最大隐写容量约为 6×10-3bpp, 最大提取率为 70%.

Wu 等[58]的隐写方法通过对比四种不同的源纹理图案作为测试图案得到实验结果,结果显示源纹理图案的分辨率越大,其能提供的总隐写容量越小.该方案中每幅图像的隐写容量最大可达 34,398 位,换算得到容量为 3.28×10-2bpp, 并且能够准确地提取信息.与 Otori 等[55]和 stego-texture 方法相比, 该方法隐写容量得到了显著提高.

Li 等[59]的隐写方法根据特定的阈值构造了 1,000 幅指纹图像,分别具有 300×300 和 500×500 像素两种尺寸.其中每个指纹图像都拥有二值化、稀疏化和灰度三种不同形式.实验结果表明,虽然该方法的提取率可以达到 100%,但指纹图像中细节点的数量有限,为了保证图像的质量,隐写容量受到了限制,约为 1.7×10-3bpp.

Zhou 等[63]的隐写方法在实验中建立了包含 500 幅分辨率为 256×256 像素的彩色真实山脉图像库作为训练集;使用训练好的轮廓−图像可逆变换模型,在每个轮廓点中隐藏长度不同秘密信息,生成 10,000 幅含密图像作为测试集.具体来说,在每个轮廓点中分别隐藏 1-8 bit 长度不同的秘密信息,从而生成对应的 8 类含密图像,每类含密图像生成 1,250 幅,一共得到 10,000 幅含密图像.由于从作为显式特征的轮廓信息到图像的映射过程更易于学习和训练, 因此, 与现有的生成式图像隐写方法相比, Zhou 等[63]的隐写方法很容易训练出相应的图像生成网络和秘密信息提取网络, 从而可以获得较高的隐写容量(当生成图像图像的尺寸为 256×256 时, 隐写容量约为 4×10-3bpp)和秘密信息提取的准确率(98.55%).

表 10 基于低层特征映射的图像生成式的性能对比

代表方法	数据集	隐写容量	提取率	噪声攻击后的提取率
Otori 等[55]	4 种不同纹理图像	$2\times10^{-3}bpp$ (打印和拍摄攻击后)	90% (打印和拍摄攻击后)	-
stego-texture[57]	7 种大理石纹理图案的样本	$6\times10^{-3}bpp$ (最大)	70% (最大)	65.27%
Wu 等[58]	四种不同的源纹理图案	3.28×10^{-2} bpp (最大)	100% (最大)	82.48%
Li 等[59]	1,000 幅指纹图像	$1.7\times10^{-3}bpp$ (最大)	100% (最大)	100%
Zhou 等[63]	500 幅彩色真实山脉图像库	$4\times10^{-3}bpp$	98.55%	79.33%

基于低层特征映射的图像生成式隐写方案的性能对比如表 10 所示. 根据各类方法在攻击后的最大提取率可以得到以下结论. 图像的纹理、轮廓等特征与像素相比更加稳定, 从而可以提升在噪声攻击下的鲁棒性. 然而, 其所能承载的信息容量却有所降低, 导致该方案的隐写容量普遍低于基于像素定义的生成式隐写方案的隐写容量.

4.3 基于高层特征关联的图像生成式隐写方案

Cao 等[66]的隐写方法使用从 Getchu①上采集的动漫头像进行训练,将 $N \times N$ 个较小的动漫角色组成含密图像.当 $N = 1$ 时,每个含密图像可以表达 14 位秘密信息. 由于利用 GAN 网络生成的动漫角色以 8×8 的形式输出更为常见,因此该方法将 N 设定为 8,相应的隐写容量为每个载体可以隐藏 896 位信息(896 bits/carrier).

为了验证方法的可行性,SSS-GAN 方法使用 MNIST[119]、CIFAR-10[120]和 CIFAR-100[120]数据集来训练模型.该方法等效于将 m 位秘密信息映射到图片中,在实验中将 m 设置为 6.SSS-GAN 方法的隐写容量取决于含密图像中包含的语义标记的数量, 根据实验结果,该方法的隐写容量超过 7.3×10^{-4}bpp. 并且,由于构建了秘密信息和图像语义信息之间的映射关系,SSS-GAN 方法可以在不同图像数据集的训练以达到模型的收敛,从而可以使提取器能够准确提取秘密信息.

STNet 方法使用 COCO 数据集[123]作为载体图像的数据集, 并将从 wikiart.org 获取的图片数据集作为参考图像的数据集. 将载体图像和参考图像的大小调整为 512×512 像素, 在图像中随机裁剪大小为 256×256 像素的区域. 为评估秘密信息的提取率, STNet 方法随机选择了 10,000 幅载体图像和参考图像, 生成 10,000 幅含密图像作为测试图像.实验结果显示,STNet 方法可以成功提取 99.8%的秘密信息,并且能够在每个像素隐藏 6×10-2 位信息生成任意大小的含密图像.

Li 等[73]的隐写方法使用 MNIST 数据集[119]中的图像作为秘密图像,并收集了 30,000 幅包含一朵花的图像作为载体图像,从 seepretty_anime_face 和 faces_datasets 数据集[73]中选择了 50,000 幅漫画图像作为反差图像.为了评估隐写方法的性能,在实验中采用三种不同大小(7×7,14×14,28×28 像素)的秘密图像,并且将载体图像和参考图像调整为 256×256 像素. 实验结果显示,该方案对噪声和滤波攻击有良好的鲁棒性.由于 MNIST 数据集是 0-9 数字的集合,可以表示 10 个数,当秘密图像的大小为 28×28 像素, 构建的含密图像为 256×256 大小时,恢复的秘密图像的最大准确率为 97.64%,其最大的隐写容量为 5.068×10^{-5}bpp.

基于高层特征关联的图像生成式隐写方案的 性能对比如表 11 所示. 从整体上看,由于图像高层特征比图像低层特征更加稳定, 不容易受到图像攻击(如添加

噪声等)的影响,因此该方案在受到噪声攻击后的提取率与原提取率差异较小, 从而具备较高的鲁棒性. 然而, 从图像中抽象出的高层特征所能承载的信息量较少,该隐写方案的的隐写容量会明显低于基于像素定义以及基于低层特征映射的生成式隐写方案.

表 11 基于高层特征关联的图像生成式的性能对比

代表方法	数据集	隐写容量	提取率	噪声攻击后的提取率
Cao 等[66]	Getchu 上采集的动漫头像	896 bits/carrier	100%	96.38%
SSS-GAN[67]	MNIST、CIFAR-10、CIFAR-100	$7.3\times10^{-4}bpp$	100%	99.96%
STNet[68]	COCO	$6\times10^{-2}bpp$	99.8%	98.64%
Li 等[73]	MNIST、seepretty_anime_face、faces_datasets	5.069×10^{-5} *bpp* (最大)	97.64%	95.39%

4.4 基于隐空间映射的图像生成式隐写方案

GSN 方法使用 CelebA 数据集和 LSUN 中的卧室场景图像子数据集. 实验结果表明, 虽然随着隐写荷载的增加, 秘密信息的提取准确率和生成质量均有所下降, 但是该方法仍然可以提供最大 8bpp 的隐写容量以及 97.53%的秘密信息提取准确率.

IDEAS 方法使用的数据集分别来自 LSUN 数据集[122]中的卧室和教堂场景图像子数据集以及 FFHQ 数据集[125]中的人脸图像子数据集. 每一子集包括 70,000 幅随机选择的图像,并将这些图像归一化为 256×256 像素的图像.根据实验结果,由于 IDEAS 方法可以利用图像结构特征的稳定性提升秘密信息的提取率,在三个子数据集上,IDEAS 方法均获得最高 100%的提取率以及每幅图像 1536bits(7.813×10^{-3}bpp)的隐写容量.

S2IRT 方法在 CelebA-HQ 数据集上进行实验,该数据集由 30,000 幅从 CelebA 数据集选出的高分辨率人脸图像.在实验中,每一幅图像均被缩放为 256×256 像素的大小以训练 Glow 模型.实验结果表明,由于位置编码的使用以及 Glow 模型提供的隐空间与图像空间可逆映射功能,S2IRT 方法的秘密信息提取率在隐写容量从 0.1bpp 到 4bpp 的范围内均能保持非常高的水平,最大的提取率为 100%. 此外,通过改进 S2IRT 方法的编码方式而形成的 SE-S2IRT 方法,其使用独立编码的策略代

替位置编码的策略对秘密信息进行编码,可以避免隐向量中单一元素的改变而导致的对整体秘密信息提取的影响,从而可以提升隐写方法的鲁棒性.SE-S2IRT 方法在各种随机噪声攻击下,秘密信息提取率仍然可以达到 91%.

基于隐空间映射的图像生成式的隐写方法性能对比如表 15 所示. 由于该隐写方案的实现基于从隐空间采样的方式, 一方面可以获得较高的隐写容量, 另一方面各种随机噪声会影响隐向量的值, 从而导致该方案与基于高层特征关联的图像生成式隐写方案相比有所降低.

表 12 基于隐空间映射的图像生成式隐写方案的性能对比

代表方法	数据集	隐写容量	提取率	噪声攻击后的提取率
Hu 等[74]	CelebA、Food101	$2.4\times10^{-2}bpp$ ($\sigma=1$)	96%	89.83%
		$7.3\times10^{-2}bpp$ ($\sigma=3$)	89%	86%
Li 等[75]	CelebA	$2.4\times10^{-2}bpp$ ($\sigma=1$)	98%	90.17%
		$7.3\times10^{-2}bpp$ ($\sigma=3$)	84%	82.50%
GSN[77]	CelebA、LSUN	$8bpp$	97.53%	88.62%
IDEAS[78]	LSUN、FFHQ	7.813×10^{-3} bpp	100%	90.42%
S2IRT[79]	CelebA-HQ	$4bpp$	100%	91%

5 存在的问题

相比传统的嵌入式隐写方案,生成式隐写方案针对现有基于统计特征的隐写分析方法具有较好的抗检测性能, 然而仍然存在以下问题.

(1)在高隐写荷载条件下, 秘密信息的提取难以达到完全无损.为了从含密载体中提取秘密信息, 基于 GAN 模型的隐空间映射的生成式隐写方法通常利用全卷积神经网络结构来设计和训练秘密信息提取器[74, 75]. 然而, 随着隐写的荷载增加, 提取器的网络模型在训练的过程中难以收敛得到全局最优解, 导致提取器提取秘密信息的准确性大大降低. 为了解决该问题, 另一些研究者基于可逆模型如 Glow 等流模型设计了隐写算法[79], 这类模型不仅可以实现秘密信息的隐写,也支持秘密信息的直接提取. 然而,以基于隐空间映射的图像生成式隐写方案为例,虽然流模型理论上是可逆的, 但是将隐空间向量映射到图像空间再重新映射回隐空间后, 所得到的隐向量与原隐向量存在一定的差异. 这是因为流模型是将连续的隐空间和离散的图像数据建立映射关系,隐向量映射为图像数据时会产生超出固定范围的异常数据元素并有所损失, 这样会显著影响秘密信息的准确提取.

(2)含密载体生成质量的稳定性较低.生成式隐写方案通过模型生成含密载体,与嵌入式隐写方案相比,含密载体的质量较差,而且并不能保证每个含密载体都有较高质量,即生成的含密载体的质量存在稳定性较差的问题. 其原因主要包括以下两点.首先,目前大多数的生成式隐写方案是基于 GAN 和流模型等生成网络模型实现的.GAN 模型存在训练过程不稳定的问题如梯度消失和模式崩溃;流模型是在高维度的连续隐空间和连续数据空间之间建立可逆映射关系,而多媒体数据通常是离散, 将连续的隐向量映射为离散的多媒体数据将会存在映射误差. 因此,这两类模型均难以保证含密载体生成质量的稳定性. 其次,现有的生成网络模型的生成含密载体能力是建立在庞大数据集上充分训练为基础,但实际用于生成网络模型训练的数据集大小通常受限,影响生成网络模型的训练效果,进而影响含密载体的生成质量的稳定性.

(3)非空间域的抗隐写分析性能有待提高.现有的生成式隐写方案在多媒体空间域上具有较好的抵抗隐写分析能力, 但在隐空间域和通信上下文域的抗隐写分析性能有待提高.1) 隐空间域: 现有的生成网络在训练阶段, 通常将服从高斯分布的隐向量映射为多媒体数据. 基于隐空间映射的生成式隐写方案,首先将秘密信息映射为隐空间的隐向量, 然后将该隐向量输入到训练后的生成网络模型以生成含密载体.然而, 在秘密信息映射过程中, 由秘密信息映射得到的隐向量难以保持高斯分布, 因此攻击者可以将含密载体逆变换为隐向量,在隐空间域可以通过检测该隐向量是否符合高斯分布来判断载体是否存在隐写行为;2) 通信上下文语义域: 当发送方与接收方进行多次隐蔽通信时, 通常需要多次传递含密载体数据. 然而, 现有的隐写方法大多数没有考虑通信数据上下文语义关联性,那么攻击者可以通过分析通信上下文语义关联从而轻易地检测出通信载体数据是否有可能包含秘密信息.因此为了保证含密载体图像通信的安全性, 不仅需要考虑多媒体空间域的抗隐写分析性能, 还需要考虑非空间域包括隐空间域、通信上下文域等其他域的抗隐写分析性能. 如何同时保证生成式隐写方案在各个域同时保证具有较好的抗隐写分析性能, 仍然是一个具有挑战性的问题.

6 总结与展望

相比于嵌入式隐写方案, 生成式隐写方案不需要对现有的载体进行修改, 而是以秘密信息为驱动直接生成含密载体, 因此针对现有基于统计特征的隐写分析方法具有较好的抗检测性能, 成为信息隐藏领域具有前景发展方向. 本文根据隐藏秘密信息载体的类别,将生成式隐写分类为图像生成式隐写方案、文本生成式隐写方案、音频生成式隐写方案和社交网络行为生成式隐写方案,分别对其中图像生成式隐写方案和文本生成式隐写方案进行了细分,并对其中的方法进行分

析和总结.然后, 本文通过大量实验, 着重对图像生成式隐写方案的性能进行了对比和分析.此外,本文总结了现有的生成式隐写存在的问题,包括秘密信息的提取难以达到完全无损、含密载体生成质量的稳定性较低、非空间域的抗隐写分析性能有待提高.

针对第 5 章所提出的问题,提出相应的解决方案和展望未来的发展方向.

(1)针对在高隐写荷载条件下秘密信息的难以精确提取问题, 拟采用基于流模型映射误差校正的生成式隐写方案.由于流模型在构建训练隐空间和多媒体空间的可逆映射时存在大量的映射误差,拟在原始流模型的映射中增加一对校正函数. 一方面, 在连续隐向量映射到离散多媒体数据过程中, 学习一个可逆的校正函数对生成的多媒体数据进行校正;另一方面, 在离散多媒体数据映射到连续隐向量映射反过程中, 使用校正函数的反函数对原始生成的多媒体数据进行恢复.以上校正函数可以确保超出固定范围的异常多媒体数据校正到固定的范围内,而对其余数据元素进行微调,利用其反函数可以准确地恢复原始图像数据和隐向量, 从而保证在高隐写荷载条件下秘密信息的精确提取.

(2)针对含密载体生成质量的稳定性不足的问题, 拟采用基于深度自注意力变换(Transformer)网络的生成式隐写方案.与卷积网络和循环网络等网络类型相比,基于 Transformer 的模型由于引入了自注意力模块,可以自动地捕获用于多媒体内容的全局依赖关系[131, 132],Transformer 网络尤其是视觉自注意力变换(Vision Transformer, ViT)网络[133]在各种计算机视觉领域上表现出强大的性能.由于目前用于含密载体生成的生成器大多利用深度卷积网络来实现含密载体的生成,而较小的卷积核很难捕获多媒体数据的有效特征[134].采用 ViT 网络设计生成式隐写模型的生成器, 可以更加有效捕获多媒体数据的全局相关性,从而提高多媒体数据生成质量和稳定性.为了进一步提高多媒体数据生成质量和稳定性,拟采用自监督学习方法和训练数据自动增强方法,以高效的方式来解决现有生成式隐写方法的训练数据集不足和训练不充分的问题.

(3)针对非空间域的抗隐写分析性能不足的问题,拟采用以下解决方案:1) 为了提高生成式隐写方案在隐空间域的抗隐写分析性能,应确保秘密信息映射的隐向量仍是服从高斯分布的.为此, 并非将秘密信息直接编码为隐向量的元素值,而是拟将秘密信息编码为隐向量元素的位置排列顺序, 而元素排列位置的变化将不会改变隐向量的高斯分布特性, 因此能够有效保持隐向量的高斯分布特性,从根本上提高了生成式隐写方法在隐空间域的抗隐写分析能力. 2) 为了解决生成的含密多媒体载体在通信上下文环境中语义合理性的问题,拟从已经传递的多媒体数据(图像、文本、语音等)序列中提取语义信息,并用 LSTM 建立语义序列自动生成模型,将可以得到与真实语义序列统计特性基本一致的语义序列自动生成

模型.然后利用该生成模型,根据已经传递的多媒体数据序列预测当前待传递的多媒体语义信息, 然后生成相应语义的含密载体多媒体数据用于隐蔽通信, 有效保证了含密多媒体载体在通信上下文环境中语义合理性, 提高了生成式隐写方法在上下文语义域的抗隐写分析能力.

(4)为了进一步提高现有的生成式隐写方案的隐写容量, 拟设计秘密信息到多媒体数据的高效可逆转换方式. 例如: 由于 Zhou 等[63]的隐写方法在图像的一维轮廓上进行隐写的隐写容量有限,而图像的二维轮廓相比一维轮廓信息承载量更大.因此,在后续的研究中,可以将秘密信息转换为图像的二维轮廓, 将该二维轮廓输入到生成网络模型中生成相应的含密载体图像,从而提高隐写容量.

(5)为了进一步验证生成式隐写方法的安全性,将研究面向生成式隐写的安全证明模型.随着各类隐写分析工具的发展, 研究者们针对隐写的安全性通常以大量实验的方式进行评价.然而,这些实验数据来说明隐写方法针对某一种或几种隐写分析工具具有抵抗能力,难以从实验上验证对现有其他隐写分析工具和未知的隐写分析工具具有较好的抵抗能力. 因此,需要研究如何从理论的角度证明隐写方法的安全性.可证明安全隐写指通过一定的理论推导的方式证明隐写方法是具有安全性的. 随着生成数据的越来越普及,其分布特性可以用规范的分布如高维高斯分布拟合并表达,这样为可证明安全隐写的发展提供了数学基础.因此, 面向生成式隐写的安全证明模型是信息隐藏领域值得关注的研究方向.

参考文献

[1] Jamil Tariq. Steganography: The art of hiding information in plain sight. IEEE potentials, 1999, 18(1):10-12

[2] Wang Shuo-Zhong, Zhang Xin-Peng, Zhang Kai-Wen. Steganography and steganalysis: Information warfare technology in the internet age. Tsinghua University Press, 2005

[3] Simmons Gustavus J. The prisoners' problem and the subliminal channel// Advances in Cryptology. 1984:51-67

[4] Van Schyndel Ron G, Tirkel Andrew Z, Osborne Charles F. A digital watermark// Proceedings of 1st international conference on image processing. 1994:86-90

[5] Bender Walter, Gruhl Daniel, Morimoto Norishige, Lu Anthony. Techniques for data hiding. IBM systems journal, 1996, 35(3.4):313-336

[6] Mielikainen Jarno. Lsb matching revisited. IEEE signal processing letters, 2006, 13(5):285-287

[7] Fridrich Jessica, Soukal David. Matrix embedding for large payloads. IEEE Transactions on Information Forensics and Security, 2006, 1(3):390-395

[8] Zhang Xinpeng, Wang Shuozhong. Dynamical running coding in digital steganography. IEEE signal processing letters, 2006, 13(3):165-168

[9] Willems Frans Mj, Van Dijk Marten. Capacity and codes for embedding information in gray-scale signals. IEEE Transactions on Information Theory, 2005, 51(3):1209-1214

[10] Filler Tomáš, Judas Jan, Fridrich Jessica. Minimizing additive distortion in steganography using syndrome-trellis codes. IEEE Transactions on Information Forensics and Security, 2011, 6(3):920-935

[11] Pevný Tomáš, Filler Tomáš, Bas Patrick. Using high-dimensional image models to perform highly undetectable steganography// International workshop on information hiding. 2010:161-177

[12] Holub Vojtěch, Fridrich Jessica. Digital image steganography using universal distortion// Proceedings of the first ACM workshop on Information hiding and multimedia security. 2013:59-68

[13] Li Bin, Wang Ming, Huang Jiwu, Li Xiaolong. A new cost function for spatial image steganography// 2014 IEEE International Conference on Image Processing (ICIP). 2014:4206-4210

[14] Sedighi Vahid, Cogranne Rémi, Fridrich Jessica. Contentadaptive steganography by minimizing statistical detectability. IEEE Transactions on Information Forensics and Security, 2015, 11(2):221-234

[15] Zhang Weiming, Zhang Zhuo, Zhang Lili, Li Hanyi, Yu Nenghai. Decomposing joint distortion for adaptive steganography. IEEE Transactions on Circuits and Systems for Video Technology, 2016, 27(10):2274-2280

[16] Liao Xin, Yu Yingbo, Li Bin, Li Zhongpeng, Qin Zheng. A new payload partition strategy in color image steganography. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 30(3):685-696

[17] Su Wenkang, Ni Jiangqun, Hu Xianglei, Fridrich Jessica. Image steganography with symmetric embedding using gaussian markov random field model. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 31(3):1001-1015

[18] Tang Weixuan, Tan Shunquan, Li Bin, Huang Jiwu. Automatic steganographic distortion learning using a generative adversarial network. IEEE signal processing letters, 2017, 24(10):1547-1551

[19] Yang Jianhua, Ruan Danyang, Huang Jiwu, Kang Xiangui, Shi Yun-Qing. An embedding cost learning framework using gan. IEEE Transactions on Information Forensics and Security, 2019, 15:839- 851

[20] Low Steven H, Maxemchuk Nicholas F, Brassil Jack T, O'gorman Lawrence. Document marking and identification using both line and word shifting// Proceedings of INFOCOM'95. 1995:853-860

[21] Fu Bing. Research on text information hiding algorithms based on Unicode coding parity. Fujian Computer, 2008, 24(12):66-66 (付兵. 基于字符 unicode 编码奇偶性的文本信息隐藏算法研 究. 福建电脑, 2008, 24(12):66-66)

[22] Yang De-Ming, Guo Sheng. Data hiding method based on word document. Computer Applications and Software, 2015, 32(5):314- 318 (杨德明, 郭盛. 基于 word 文档的数据隐藏方法. 计算机应 用与软件, 2015, 32(5):314-318)

[23] Chang Ching-Yun, Clark Stephen. Practical linguistic steganography using contextual synonym substitution and a novel vertex coding method. Computational linguistics, 2014, 40(2):403- 448

[24] Zhang Jianjun, Wang Lucai, Lin Haijun. Coverless text information hiding method based on the rank map. Journal of Internet Technology, 2017, 18(2):427-434

[25] Yang Xiao, Li Feng, Xiang Ling-Yun. Synonym substitutionbased steganographic algorithm with matrix coding. Journal of Chinese Computer Systems, 2015, 36(6):1296-1300 (杨潇, 李峰, 向凌云. 基于矩阵编码的同义词替换隐写算法. 小型微型计算机系统, 2015, 36(6):1296-1300)

[26] Gopalan Kaliappan. Audio steganography using bit modification// 2003 International Conference on Multimedia and Expo. ICME'03. Proceedings (Cat. No. 03TH8698). 2003:I-629

[27] Gruhl Daniel, Lu Anthony, Bender Walter. Echo hiding// International Workshop on Information Hiding. 1996:295-315

[28] Erfani Yousof, Siahpoush Shadi. Robust audio watermarking using improved ts echo hiding. Digital Signal Processing, 2009, 19(5):809-814

[29] Paillard Bruno, Mabilleau Philippe, Morissette Sarto, Soumagne Joël. Perceval: Perceptual evaluation of the quality of audio signals. Journal of the Audio Engineering Society, 1992, 40(1/2):21-31

[30] Gang Litao, Akansu Ali N, Ramkumar Mahalingam. Mp3 resistant oblivious steganography// 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 01CH37221). 2001:1365-1368

[31] Djebbar Fatiha, Hamam Habib, Abed-Meraim Karim, Guerchi Driss. Controlled distortion for high capacity data-in-speech spectrum steganography// 2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing. 2010:212-215

[32] Fridrich Jessica, Goljan Miroslav. On estimation of secret message length in lsb steganography in spatial domain// Security, steganography, and watermarking of multimedia contents VI. 2004:23-34

[33] Chen Chunhua, Shi Yun Q. Jpeg image steganalysis utilizing both intrablock and interblock correlations// 2008 IEEE International Symposium on Circuits and Systems. 2008:3029-3032

[34] Pevny Tomáš, Bas Patrick, Fridrich Jessica. Steganalysis by subtractive pixel adjacency matrix. IEEE Transactions on Information Forensics and Security, 2010, 5(2):215-224

[35] Qian Yinlong, Dong Jing, Wang Wei, Tan Tieniu. Learning and transferring representations for image steganalysis using

convolutional neural network// 2016 IEEE international conference on image processing (ICIP). 2016:2752-2756

[36] Xu Guanshuo, Wu Han-Zhou, Shi Yun-Qing. Structural design of convolutional neural networks for steganalysis. IEEE signal processing letters, 2016, 23(5):708-712

[37] Xu Guanshuo, Wu Han-Zhou, Shi Yun Q. Ensemble of cnns for steganalysis: An empirical study// Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security. 2016:103-107

[38] Zaremba Wojciech, Sutskever Ilya, Vinyals Oriol. Recurrent neural network regularization. arXiv preprint arXiv:1409.2329, 2014

[39] Lipton Zachary C, Kale David C, Elkan Charles, Wetzel Randall. Learning to diagnose with lstm recurrent neural networks. arXiv preprint arXiv:1511.03677, 2015

[40] Kingma Diederik P, Welling Max. Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114, 2013

[41] Doersch Carl. Tutorial on variational autoencoders. arXiv preprint arXiv:1606.05908, 2016

[42] Goodfellow Ian, Pouget-Abadie Jean, Mirza Mehdi, Xu Bing, Warde-Farley David, Ozair Sherjil, Courville Aaron, Bengio Yoshua. Generative adversarial nets. Advances in neural information processing systems, 2014, 27

[43] Creswell Antonia, White Tom, Dumoulin Vincent, Arulkumaran Kai, Sengupta Biswa, Bharath Anil A. Generative adversarial networks: An overview. IEEE signal processing magazine, 2018, 35(1):53-65

[44] Dinh Laurent, Krueger David, Bengio Yoshua. Nice: Non-linear independent components estimation. arXiv preprint arXiv:1410.8516, 2014

[45] Dinh Laurent, Sohl-Dickstein Jascha, Bengio Samy. Density estimation using real nvp. arXiv preprint arXiv:1605.08803, 2016

[46] Kingma Durk P, Dhariwal Prafulla. Glow: Generative flow with invertible 1x1 convolutions. Advances in neural information processing systems, 2018, 31

[47] Yang Kuan, Chen Kejiang, Zhang Weiming, Yu Nenghai. Provably secure generative steganography based on autoregressive model// International Workshop on Digital Watermarking. 2018:55- 68

[48] Van Oord Aaron, Kalchbrenner Nal, Kavukcuoglu Koray. Pixel recurrent neural networks// International conference on machine learning. 2016:1747-1756

[49] Hopper Nicholas J, Langford John, Ahn Luis Von. Provably secure steganography// Annual International Cryptology Conference. 2002:77-92

[50] Zhang Siyu, Yang Zhongliang, Tu Haoqin, Yang Jinshuai, Huang Yongfeng. Pixel-stega: Generative image steganography based on autoregressive models. arXiv preprint arXiv:2112.10945, 2021

[51] Salimans Tim, Karpathy Andrej, Chen Xi, Kingma Diederik P. Pixelcnn++: Improving the pixelcnn with discretized logistic mixture likelihood and other modifications. arXiv preprint arXiv:1701.05517, 2017

[52] Liu Jia, Zhou Tanping, Zhang Zhuo, Ke Yan, Lei Yu, Zhang Minqing. Digital cardan grille: A modern approach for information hiding// Proceedings of the 2018 2nd International Conference on Computer Science and Artificial Intelligence. 2018:441-446

[53] Liu Jia, Ke Yan, Lei Yu, Li Jun, Wang Yaojie, Han Yiliang, Zhang Minqing, Yang Xiaoyuan. The reincarnation of grille cipher: A generative approach. arXiv preprint arXiv:1804.06514, 2018

[54] Wang Yaojie, Yang Xiaoyuan, Liu Wenchao. Generative image steganography based on digital cardan grille// International Conference on Security and Privacy in New Computing Environments. 2020:343-355

[55] Otori Hirofumi, Kuriyama Shigeru. Texture synthesis for mobile data communications. IEEE Computer graphics and applications, 2009, 29(6):74-81

[56] Mäenpää Topi, Pietikäinen Matti. Texture analysis with local binary patterns//Handbook of pattern recognition and computer vision. World Scientific, 2005: 197-216

[57] Xu Jiayi, Mao Xiaoyang, Jin Xiaogang, Jaffer Aubrey, Lu Shufang, Li Li, Toyoura Masahiro. Hidden message in a deformation-based texture. The Visual Computer, 2015, 31(12):1653-1669

[58] Wu Kuo-Chen, Wang Chung-Ming. Steganography using reversible texture synthesis. IEEE Transactions on Image Processing, 2014, 24(1):130-139

[59] Li Sheng, Zhang Xinpeng. Toward construction-based data hiding: From secrets to fingerprint images. IEEE Transactions on Image Processing, 2018, 28(3):1482-1497

[60] Larkin Kieran G, Fletcher Peter A. A coherent framework for fingerprint analysis: Are fingerprints holograms? Optics express, 2007, 15(14):8667-8677

[61] Reed Irving S, Solomon Gustave. Polynomial codes over certain finite fields. Journal of the society for industrial and applied mathematics, 1960, 8(2):300-304

[62] Cappelli Raffaele, Erol A, Maio D, Maltoni D. Synthetic fingerprint-image generation// Proceedings 15th International Conference on Pattern Recognition. ICPR-2000. 2000:471-474

[63] Zhou Zhi-Li, Wang Mei-Min, Yang Gao-Bo, Zhu Jian-Yu, Sun Xing-Ming. Generative steganography method based on autogeneration of contours. Journal on Communications, 2021, 42(9):144-154 (周志立，王美民，杨高波，朱剑宇，孙星明. 基于轮廓自动生 成的构造式图像隐写方法. 通信学报, 2021, 42(9):144-154)

[64] Hochreiter Sepp, Schmidhuber Jürgen. Long short-term memory. Neural computation, 1997, 9(8):1735-1780

[65] Isola Phillip, Zhu Jun-Yan, Zhou Tinghui, Efros Alexei A. Imageto-image translation with conditional adversarial networks// Proceedings of the IEEE conference on computer vision and pattern recognition. 2017:1125-1134

[66] Cao Yi, Zhou Zhili, Wu Qm, Yuan Chengsheng, Sun Xingming. Coverless information hiding based on the generation of anime characters. EURASIP Journal on Image and Video Processing, 2020, 2020(1):1-15

[67] Zhang Zhuo, Fu Guangyuan, Ni Rongrong, Liu Jia, Yang Xiaoyuan. A generative method for steganography by cover synthesis with auxiliary semantics. Tsinghua Science and Technology, 2020, 25(4):516-527

[68] Wang Zihan, Gao Neng, Wang Xin, Xiang Ji, Liu Guanqun. Stnet: A style transformation network for deep image steganography// International Conference on Neural Information Processing. 2019:3-14

[69] Simonyan Karen, Zisserman Andrew. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014

[70] Huang Xun, Belongie Serge. Arbitrary style transfer in real-time with adaptive instance normalization// Proceedings of the IEEE international conference on computer vision. 2017:1501-1510

[71] Gu Jiuxiang, Wang Zhenhua, Kuen Jason, Ma Lianyang, Shahroudy Amir, Shuai Bing, Liu Ting, Wang Xingxing, Wang Gang, Cai Jianfei. Recent advances in convolutional neural networks. Pattern recognition, 2018, 77:354-377

[72] Zhu Jun-Yan, Park Taesung, Isola Phillip, Efros Alexei A. Unpaired image-to-image translation using cycle-consistent adversarial networks// Proceedings of the IEEE international conference on computer vision. 2017:2223-2232

[73] Li Qi, Wang Xingyuan, Wang Xiaoyu, Ma Bin, Wang Chunpeng, Shi Yunqing. An encrypted coverless information hiding method based on generative models. Information Sciences, 2021, 553:19- 30

[74] Hu Donghui, Wang Liang, Jiang Wenjie, Zheng Shuli, Li Bin. A novel image steganography method via deep convolutional generative adversarial networks. IEEE Access, 2018, 6:38303- 38314

[75] Li Jun, Niu Ke, Liao Liwei, Wang Lijie, Liu Jia, Lei Yu, Zhang Minqing. A generative steganography method based on wgan-gp// International Conference on Artificial Intelligence and Security. 2020:386-397

[76] Gulrajani Ishaan, Ahmed Faruk, Arjovsky Martin, Dumoulin Vincent, Courville Aaron C. Improved training of wasserstein gans. Advances in neural information processing systems, 2017, 30

[77] Wei Ping, Li Sheng, Zhang Xinpeng, Luo Ge, Qian Zhenxing, Zhou Qing. Generative steganography network. arXiv preprint arXiv:2207.13867, 2022

[78] Liu Xiyao, Ma Ziping, Ma Junxing, Zhang Jian, Schaefer Gerald, Fang Hui. Image disentanglement autoencoder for steganography without embedding// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022:2303-2312

[79] Zhou Zhili, Su Yuecheng, Wu Qm, Fu Zhangjie, Shi Yunqing. Secret-to-image reversible transformation for generative steganography. arXiv preprint arXiv:2203.06598, 2022

[80] Huang Ding, Yan Hong. Interword distance changes represented by sine waves for watermarking text images. IEEE Transactions on Circuits and Systems for Video Technology, 2001, 11(12):1237- 1245

[81] Shniperov Alexey Nikolaevich, Nikitina Ka. A text steganography method based on markov chains. Automatic Control and Computer Sciences, 2016, 50(8):802-808

[82] Yang Zhongliang, Jin Shuyu, Huang Yongfeng, Zhang Yujin, Li Hui. Automatically generate steganographic text based on markov model and huffman coding. arXiv preprint arXiv:1811.04720, 2018

[83] Luo Yubo, Huang Yongfeng, Li Fufang, Chang Chinchen. Text steganography based on ci-poetry generation using markov chain model. KSII Transactions on Internet and Information Systems (TIIS), 2016, 10(9):4568-4584

[84] Desoky Abdelrahman. Notestega: Notes-based steganography methodology. Information Security Journal: A Global Perspective, 2009, 18(4):178-193

[85] Desoky Abdelrahman. Jokestega: Automatic joke generation–based steganography methodology. International Journal of Security and Networks, 2012, 7(3):148-160

[86] Huang Yongfeng, Liu Chenghao, Tang Shanyu, Bai Sen. Steganography integration into a low-bit rate speech codec. IEEE

Transactions on Information Forensics and Security, 2012, 7(6):1865-1875

[87] Huang Yong Feng, Tang Shanyu, Yuan Jian. Steganography in inactive frames of voip streams encoded by source codec. IEEE Transactions on Information Forensics and Security, 2011, 6(2):296-306

[88] Yang Zhong-Liang, Guo Xiao-Qing, Chen Zi-Ming, Huang YongFeng, Zhang Yu-Jin. Rnn-stega: Linguistic steganography based on recurrent neural networks. IEEE Transactions on Information Forensics and Security, 2018, 14(5):1280-1295

[89] Yang Zhongliang, Wei Nan, Liu Qinghe, Huang Yongfeng, Zhang Yujin. Gan-tstega: Text steganography based on generative adversarial networks// International Workshop on Digital Watermarking. 2019:18-31

[90] Yang Zhongliang, Wang Ke, Li Jian, Huang Yongfeng, Zhang YuJin. Ts-rnn: Text steganalysis based on recurrent neural networks. IEEE signal processing letters, 2019, 26(12):1743-1747

[91] Yang Zhong-Liang, Zhang Si-Yu, Hu Yu-Ting, Hu Zhi-Wen, Huang Yong-Feng. Vae-stega: Linguistic steganography based on variational auto-encoder. IEEE Transactions on Information Forensics and Security, 2020, 16:880-895

[92] Devlin Jacob, Chang Ming-Wei, Lee Kenton, Toutanova Kristina. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018

[93] Zhou Xuejing, Peng Wanli, Yang Boya, Wen Juan, Xue Yiming, Zhong Ping. Linguistic steganography based on adaptive probability distribution. IEEE Transactions on Dependable and Secure Computing, 2021

[94] Cao Yi, Zhou Zhili, Chakraborty Chinmay, Wang Meimin, Wu Qm Jonathan, Sun Xingming, Yu Keping. Generative steganography based on long readable text generation. IEEE Transactions on Computational Social Systems, 2022

[95] Dathathri Sumanth, Madotto Andrea, Lan Janice, Hung Jane, Frank Eric, Molino Piero, Yosinski Jason, Liu Rosanne. Plug and play language models: A simple approach to controlled text generation. arXiv preprint arXiv:1912.02164, 2019

[96] Yi Biao, Wu Hanzhou, Feng Guorui, Zhang Xinpeng. Alisa: Acrostic linguistic steganography based on bert and gibbs sampling. IEEE signal processing letters, 2022, 29:687-691

[97] Gelfand Alan E, Smith Adrian Fm. Sampling-based approaches to calculating marginal densities. Journal of the American statistical association, 1990, 85(410):398-409

[98] Qin Chuan, Wang Meng, Si Guang-Wen, Yao Heng. Constructive information hiding with chinese quatrain generation. Chinese Journal of Computers, 2021, 44(4):773-785

[99] Sutskever Ilya, Vinyals Oriol, Le Quoc V. Sequence to sequence learning with neural networks. Advances in neural information processing systems, 2014, 27