ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 38,No. 5,Sep. 2023,pp. 995－1016
DOI:10. 16337/j. 1004-9037. 2023. 05. 001

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

音频隐写方法综述:从传统到深度学习

张雄伟,葛晓义,孙　蒙,宋宫琨琨,李　莉

(陆军工程大学指挥控制工程学院,南京 210007)

摘　要: 数字音频作为网络空间中广泛应用的媒体,是承载秘密信息的良好载体,常被用来构建实时性强、复杂度低、不可感知性好的隐蔽通信。音频隐写作为确保网络信息安全和数据保密通信的关键技术手段之一,正受到越来越多学者的关注。本文对音频隐写方法的发展脉络进行了系统性梳理。首先,介绍了音频隐写的基本内容,对问题描述、常用数据格式、工具和评价指标等进行总结。其次,按照嵌入域的不同,将传统音频隐写方法分为时域方法、变换域方法和压缩域方法,并分析其优缺点;根据隐写载体的不同,将基于深度学习的隐写方法划分为嵌入载体式、生成载体式和无载体式音频隐写,并对这3种音频隐写方法进行了对比分析。最后,指出了当前音频隐写进一步的研究方向。

关键词: 音频隐写;信息隐藏;隐蔽通信;深度学习;短时傅里叶变换

中图分类号: TN912　　**文献标志码:** A

An Overview of Audio Steganography Methods: From Tradition to Deep Learning

ZHANG Xiongwei, GE Xiaoyi, SUN Meng, SONG GONG Kunkun, LI Li

(College of Command & Control Engineering, Army Engineering University of PLA, Nanjing 210007, China)

Abstract: As a widely used medium in the cyberspace, digital audio serves as an excellent cover for carrying secret information and is often employed in the construction of covert communication systems that prioritize real-time performance, low complexity, and imperceptibility. Audio steganography, one of the key techniques for ensuring network information security and confidential communication, has attracted increasing attention from scholars. This paper presents a systematic review of the development context of audio steganography methods. Firstly, we introduce the basic contents of audio steganography, and summarize the problem description, evaluation indicators, common data formats, and tools. Secondly, according to different embedding domains, traditional audio steganography methods are classified into time domain methods, transform domain methods and compression domain methods, and their advantages and disadvantages are analyzed. Furthermore, based on different steganographic covers, the deep learning-based steganography methods are categorized into embedding cover-based, generating cover-based, and coverless audio steganography, then the three steganography methods are compared and analyzed. Finally, suggestions for further research directions in audio steganography are pointed out.

Key words: audio steganography; data hiding; covert communication; deep learning; short time Fourier transform

---

基金项目: 国家自然科学基金(62071484,62371469);江苏省优秀青年基金(BK20180080)。
收稿日期: 2023-07-24;**修订日期:** 2023-08-26

引　言

在开放的网络环境中,传输音频、图像、视频等媒体可能会遭到攻击者的破坏,给政治、医疗、金融等方面的信息安全和数据通信带来严重的危害。尽管采用加密的方法可以防止攻击者读取信息,但加密并不能隐藏秘密通信的存在,从而可能会让攻击者采取阻断、破坏、干扰等暴力手段破坏秘密信息。同时,随着计算机硬件和算法的迭代,通过深度学习算法可能会推断出解密的密钥,更甚至跳过秘密数据挖掘情报。信息隐藏是一种将秘密信息隐藏在多媒体信息中的方法,让攻击者无法觉察秘密信息的存在,能够从源头上保护秘密信息的传递[1]。数字水印与信息隐藏具有密切相关性,不同之处在于信息隐藏是将秘密信息嵌入到载体中,达到隐蔽通信的目的;而数字水印是将版权信息嵌入载体中,通常包含数据的来源、状态或发送方的信息[2]。根据嵌入载体的不同,可将信息隐藏分为音频隐写、图像隐写、视频隐写和文本隐写等。音频作为人类交流的主要方式和常见的媒体信息,是天然的隐写载体,在数据通信、信息安全和数字水印等具有广泛应用需求,因此音频隐写具有重要的研究意义。然而,在音频中隐藏信息是一个巨大的挑战,因为人类听觉系统(Human auditory system, HAS)在一个很宽的频率范围内动态工作,介于20～20 000 Hz之间,对于添加的"噪声"非常敏感,并且人类的听觉系统比视觉系统更敏感[3],因此嵌入秘密信息而导致载体内容的变化很容易引起注意。同时,由于音频信号的维数比视频信号小,音频中可嵌入的数据量要低于视频中嵌入的数据量。

自音频隐写算法研究以来,国内外学者提出许多高效的算法,尤其是随着深度学习算法在音频领域的不断发展,给音频隐写带来了新的机遇,促使音频隐写出现了一系列新算法、新架构。隐写分析则是与隐写博弈的技术,借助分析载体的具体统计特征,从而判断载体中有没有嵌入相关秘密信息,而且还可估计具体的嵌入量信息,提取获得隐藏信息。该技术被广泛应用于军事情报侦察、网络攻防、互联网犯罪跟踪以及搜查证据等方面[4],也是一种提高音频隐写安全性的对抗手段[5]。AlSabhany等[6]和Dutta等[7]分别就嵌入过程和嵌入域不同对音频隐写的方法进行了系统的综述,并对音频隐写的数据集和评价指标等进行了详细的统计,并在最后对音频隐写分析进行了简要的梳理。张卫明等[8]对国内外多媒体隐写研究进展进行了总结和对比,其中对近两年音频隐写的国内外研究现状进行了梳理。

基于深度学习的音频隐写起步晚、文献相对少,因此在已有的音频隐写综述中未对基于深度学习的音频隐写方法进行总结分类。就现阶段而言,根据嵌入载体不同,可以对已提出的基于深度学习的音频隐写方法进行总结分类。本文不仅综述了音频隐写的相关内容,而且根据发展历程和具体方法,对音频隐写方法进行了系统的划分和梳理,并对基于深度学习的方法进行了总结。

1　音频隐写概述

音频隐写是指利用数字音频或语音作为载体来隐藏秘密信息,主要是在音频信号的时域、频域、小波域以及音频流(Voice of internet protocol, VoIP)中嵌入,关注嵌入信息的隐蔽性与不可感知性[9],以便保护这些数据免受未经授权者的访问。本节对音频隐写的问题描述、音频类型和工具进行了梳理,并在最后对音频隐写常见的评价指标进行了归纳总结。

1.1　问题描述

音频隐写的问题描述如下:给定秘密信息$M=\{m_1, m_2, \cdots, m_i, \cdots, m_K\}$和音频载体$C=\{c_1, c_2, \cdots, c_j, \cdots, c_L\}$,其中$i\in K$,$j\in L$,$K$与$L$分别表示秘密信息和音频载体的数量,$m_i$代表一条秘密信息或秘密信息的一部分,$c_j$代表一条音频隐写载体。隐写过程如图1所示。音频隐写的任务是要通过隐写算法Emb(·)将秘密信息$M=\{m_1, m_2, \cdots, m_K\}$逐个隐藏在音频载体中,即:$\mathrm{Emb}(m_i, c_j)\rightarrow s_j$,$S=\{s_1, s_2, \cdots, s_j, \cdots, s_L\}$表示含密载体,可以在公开的网络信道中传输。接收方通过提取算法Ext(·)从接

收到的含密载体$S'=\{s'_1,s'_2,\cdots,s'_j,\cdots,s'_L\}$中提取隐写的秘密信息,即$\mathrm{Ext}(s'_j)\rightarrow m'_i$。

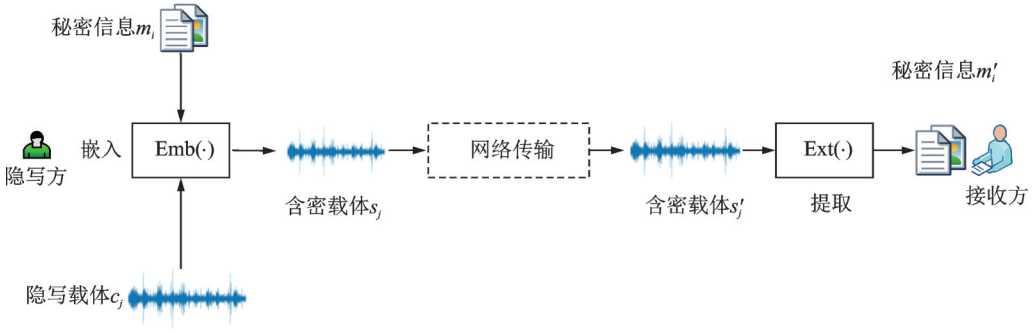

图1 音频隐写一般过程

Fig.1 General process of audio steganography

1.2 音频数据与音频隐写工具

对常用数据集梳理可以发现,音频隐写模型大都选择自定义的音频数据进行评估,只有少部分模型使用公开数据集进行实验[6]。主要采用TIMIT[10]、NOIZEU[11]、GTZAN[12]和CORPORA[10]等数据集。对当前音频隐写采用的音频类型进行梳理分析,可以发现提出的方法中常用的音频格式分别是WAV、VoIP、MP3(Moving picture experts group audio layer Ⅲ)、AMR(Adaptive multi-rate)、AAC(Advanced audio coding)、AU和MIDI(Musical instrument digital interface)。其中最常见的载体格式是WAV、VoIP和MP3,采用AMR、AU和MIDI的方法较少,具体信息如表1所示。

表1 常用音频格式

Table 1 Common audio formats

类型	格式	特点	适用性
WAV	未压缩	真实记录自然声波形,声音不失真,数据量大	Windows
AU	未压缩	品质音频高、兼容性强、稳定性高	Unix、Java
MIDI	N/A	轻量级、可编辑性、兼容性、音色丰富	多用于音乐制作和演奏等
MP3	有损压缩	对不同频段采用不同的压缩率,压缩后占用空间小	多平台适用,常用移动设备
AAC	有损压缩	编解码器质量高,性能高	多平台适用
AMR	有损压缩	压缩比较大,满足移动通讯需求,对于通话效果较好	多用于移动设备
VoIP	有损压缩	适用网络通讯,减少失真	智能手机与电脑在内的联网接入设备

与此同时,一些针对不同音频格式的音频隐写软件也应运而生。在Hayati等[13]研究的基础上,添加了部分软件,并增加了相对应的算法,具体如表2所示。可以发现当前音频隐写软件主要针对MP3和WAV的音频格式,且大都采用LSB算法。当前针对音频隐写软件的好坏并没有统一的评判标准,但可以通过代码静态测试、动态测试及黑盒测试等对隐写软件的准确度进行检测。常用的隐写软件主要有MP3Stego、SilentEye和DeepSound。

MP3Stego是在压缩过程中将信息隐藏在MP3文件中,具体来说,是采用基于量化步长修改的方法将数据隐藏在MP3文件的奇偶校验块中。SilentEye和DeepSound可以面向WAV格式的音频,以及JPEG和BMP格式的图片,能够利用AES算法对秘密信息进行加密,且具有集成新隐写算法优势。DeepSound不仅具有隐写、解码的功能,还具有对多种音频格式实施转换的功能,且支持使用AES-256

表2 音频隐写工具

Table 2 Audio steganography tools

名称	支持格式	算法	开源
MP3Stego	MP3	基于量化步长修改方法	√
S-Tools	WAV	LSB	√
SilentEye	WAV	LSB	√
Hide4PGP	MP3/VOC	LSB	√
stegandotnet	MP3/WAV/MIDI/AU	LSB	√
StegoStick	WAV	LSB	√
Info Stego	MP3	N/A	√
Scram Disk	WAV	N/A	√
DeepSound	MP3/WAV	LSB	√
Steghide	WAV/AU	LSB	√

算法加密秘密信息。另外,DeepSound能够很好地控制嵌入秘密后的音频质量,当秘密信息的大小为音频文件的1/8时,具有较高的质量,当为1/4时,具有正常的质量,当为1/2时,含密音频质量较低。

1.3 评价指标

音频隐写常用的3个主要衡量标准为不可感知性、隐写容量和鲁棒性,三者成三角关系,相互矛盾。当过分地强调某一性能时,则致使另外两项性能下降,因此构建一个好的隐写方法,需要同时考虑3项性能,使不可感知性、隐藏容量和鲁棒性能够达到较好的平衡。

不可感知性常用的评价指标主要有信噪比(Signal-to-noise ratio, SNR)、峰值信噪比(Peak signal-to-noise ratio, PSNR)、客观等级差异(Objective difference grade, ODG)和均方误差(Mean squared error,MSE)等。SNR、PSNR和MSE主要表示含密音频的失真度,数值越大表示失真越小,同时不可感知性也越好。根据信息隐藏标准,隐写后的SNR值不小于20 dB[14]。

ODG是通过模拟人耳系统对比分析参考信号和测试信号得出,在音频隐写中参考信号和测试信号分别是隐藏前后的信号,主要用于衡量不可感知性,其数值范围为[−4,0],−4表示表示音频质量劣,−2～0之间表示音频质量相对较好。因此,ODG的数值越大,不可感知性越好,评价等级如表3所示。

隐写容量是指在一个音频载体中最多能嵌入的信息,是在满足不可检测性的条件下分析隐写容量的极限。通常在嵌入多个秘密消息后,再次分析对比不可感知性的评价指标。

鲁棒性是指隐藏的秘密信息抵抗各种信号处理的能力。比特误码率(Bit error rate, BER)常用于衡量音频隐写的鲁棒性,是指错误比特数与总比特数之比,BER值越小,则提取的秘密信息越准确。在音频隐写中,说话人身份属性也常被用作评价鲁棒性的指标,常用等错误率(Equal error rate, EER)来衡量是否保留了说话人的身份信息,EER越低,则说明音频中保留的说话人身份信息越完整。

表3 ODG评价等级表

Table 3 ODG evaluation scale

ODG	描述	音频质量
0	不可察觉	优
−1	可察觉但不刺耳	良
−2	轻微刺耳	中
−3	刺耳	差
−4	非常刺耳	劣

2 传统音频隐写方法

本节根据秘密信息嵌入域不同,将传统音频隐写方法分为时域、变换域和压缩域方法,再根据所采用的技术

将其进行详细划分,具体方法划分如图2所示。

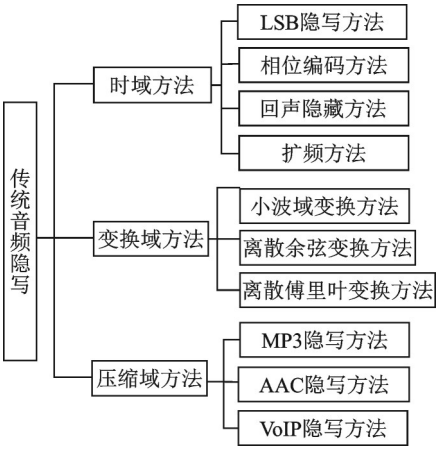

图2 传统音频隐写方法

Fig.2 Traditional audio steganography methods

2.1 时域方法

2.1.1 LSB隐写方法

最低有效位(Least significant bit, LSB)隐写方法作为音频隐写中最简单的方法,具有嵌入容量大,复杂度低的优点。具体来说,LSB隐写方法是指将载体音频采样值的最不重要位替换为秘密信息,以达到信息隐藏目的,通常只利用一个最低有效位。为了提高隐写容量,也有利用两个最低有效位的方法。将最低位替换为秘密信息的二进制位的具体方法如图3所示。

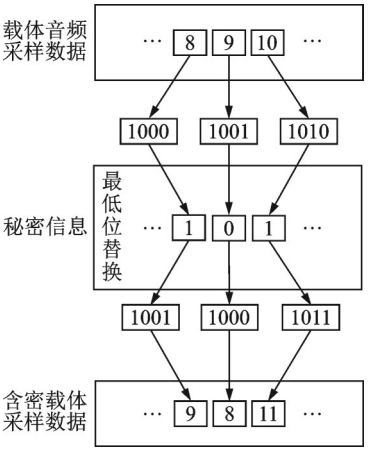

图3 LSB隐写方法

Fig.3 LSB steganography

尽管LSB隐写方法具有简单、高效的优点,但是将最低有效位直接替换为二进制秘密信息的方法存在着极大的安全问题,通常将密码学与隐写相结合来提高安全性,即将秘密信息加密后再嵌入隐写载体中。Gambhir等[15]使用RSA算法将秘密信息加密成密文,然后使用LSB方法将密文隐藏在音频信号中。Mishra等[16]将秘密信息转换成ASCII码,然后通过遗传算法在载体中找到嵌入秘密信息的最佳位置,最后利用LSB方法将秘密信息对应的ASCII码嵌入载体中。Nassrullah等[17]提出基于LSB的高效音频隐写方法,通过利用载体在隐藏容量和失真率之间进行平衡来提高隐写性能,并能根据秘密信息大小、载体大小和SNR自适应地确定每个音频样本的隐藏比特数。该方法主要是通过2个过程使得载体的利用最大化:(1)估计载体的容量,并找到一个嵌入比;(2)每个音频载体按照嵌入比和容量嵌入若干消息位。Rakshit等[18]提出一种基于模式LSB将图片信息嵌入至音频中的方法,利用视觉密码技术将图片划分为多个不可见的子图,然后利用模式LSB音频隐写方法将多个子图分别隐藏在不同的音频文件中。

2.1.2 相位编码隐写方法

相位编码是利用HAS对相位失真的低灵敏性,通过改变相位值来实现,对噪音攻击具有较高鲁棒性,其感知性取决于变化的严重程度和嵌入率[19]。Bender等[1]通过将初始段的相位替换为表示数据的参考相位来修改,取得较好的SNR,但频率分量之间相位差的变化,则可能产生失真。Dong等[20]提出了两种相位编码方法,对MP3压缩也具有鲁棒性,其中一种是根据数据修改音频频谱的一对(或多个)频率分量的相对相位进行隐藏。Malik等[21]提出了一种在隐写载体音频中使用一组具有不同参数的全通滤波器(Allpass filters, ApF)致使可控相位修改的方法,选择滤波器参数对嵌入信息进行编码,具有较大的隐藏容量,也能具备较好的不可感知性和鲁棒性。当前相位编码的隐写方法大部分是基于静态嵌入,缺乏自适应性,且采用顺序嵌入的方法,这些都会导致隐写效果差。Alsabhany等[22]针对隐写性能不平衡、缺乏自适应性和动态分配的问题,提出自适应多级相位编码的方法。在不改变原始音频的情况下,利用自适应嵌入策略增加嵌入率和数据容量,并且在嵌入过程中考虑了数据的插入和删除等情况,保证了数据的可靠性和稳定性,可以有效防御抖动、低通滤波器等攻击。

2.1.3 回声隐写方法

回声隐写是利用HAS具有屏蔽特性,即时域掩蔽效应,在隐写载体音频中加入不同延迟的回声,从而将秘密信息嵌入到隐写载体中,具有较好的不可感知性以及简单易实现的优点[1,23]。Oh等[24]提出了一种回声嵌入方法,能够在隐藏载体音频质量不恶化的情况下嵌入高能量回波,使其对常见的信号处理修改具有较强的鲁棒性和抗篡改性。Erfani等[25]提出了一种基于回声隐藏的音频水印技术,通过在隐藏载体音频上增加短共振,将秘密信息插入其中,对常见信号处理攻击的鲁棒性较好。Ghasemzadeh等[26]针对之前方法的不安全性,通过伪随机方式改变回波的参数,提高了回声隐藏的安全性。具体来说,其根据私钥生成两个伪随机序列,使用第1个序列将要嵌入的秘密信息生成为不同的长度,从而提高了安全性,因为在不知道长度的情况下恢复秘密信息是困难的。第2个序列用于通过改变延迟值为每个段生成不同的回声库,为了提取密钥,将密钥输入伪随机数发生器,该序列用于重新生成每个片段及其回声库的长度。然后,计算各节段的倒谱,并比较其在2个指标上的值。针对传统回声隐写存在安全性、嵌入限制和音频结构等问题,Wang等[27]分别提出了利用稀疏子空间聚类(Sparse subspace clustering, SSC)实现了基于听不见的回声隐藏语音水印方法和利用音频信号中重复结构的时-频(Time-frequency, T-F)特性的回声隐藏方法。他们将原始音频信号转换为高维T-F表示,然后将具有相似T-F特征的时间帧聚类到同一子空间中,将原始音频分解为子空间的联合,每个子空间对应于一个时域子信号。将成对和相反的回波核应用于能量平衡子信号进行水印嵌入,在提取过程中根据T-F相似性恢复子空间,并利用倒谱分析提取水印。

2.1.4 扩频隐写方法

扩频最初是为了增强通信的鲁棒性和保证信息的传递而提出的一种通信方法。将扩频概念引入音频信息隐藏形成了扩频隐写方法,该方法通过将窄带秘密信息信号在较宽的频率范围内进行扩展实现信息隐藏,可以分为直接序列扩频和跳频扩频两种方法。Matsuoka[28]提出了一种在心理声学模型的基础上,在数据频域进行掩蔽的扩频方法。在音频信号中引入相移来降低每个子带与PN(Pseudo-random noise)信号的相关性。在对复合信号进行解扩时,可以方便地从音频中检测出嵌入的数据信号,产生的复合信号具有抗噪声的优点。Kuznetsov等[29]研究了在音频中使用直接扩频技术隐藏信息的方法,探索了5种不同的产生扩频序列的方式对音频隐写的影响。

2.2 变换域方法

变换域方法是在变换后的隐写载体域内嵌入秘密信息,再将含密载体通过逆变换为时域进行传输,具体流程如图4所示。常用的变换域方法有快速傅里叶变换(Fast Fourier transform, FFT)、离散余弦变换(Discrete cosine transform, DCT)和离散小波变换(Discrete wavelet transform, DWT)等。

Gopalan[30]提出了一种在隐写载体的倒谱域中嵌入秘密信息的方法,音频帧的复倒谱被定义为该帧复对数的逆FFT的频谱。Fallahpour等[31]提出一种基于FFT变换的音频隐写方法,具有低复杂度和实时性的特性,该方法在5~15 kHz之间的频率带中幅度略有修改,并且使用幅度小于阈值的频率进行嵌入。Dieu等[32]认为基于时域的音频隐藏方法对于常见的信号处理攻击来说往往较弱,认为使用基于变换的方法以增加计算复杂性为代价提供了更好的感知质量和抵抗常见攻击的鲁棒性,提出了一种基于FFT变换的方法,使用FFT将隐写载体从时域转换到频域,通过修改频域的幅度来嵌入秘密信息。

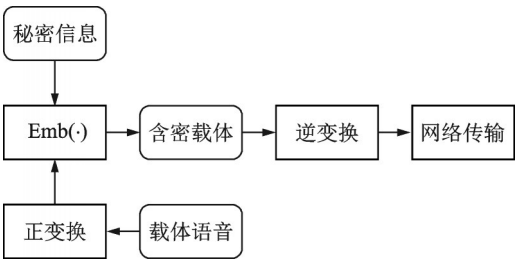

图4 变换域方法处理流程

Fig.4 Flowchart of transformation domain method

陈红松等[33]针对基于扩频方法存在的盲提取问

题,提出了一种基于DCT的盲音频隐写算法。Charfeddine等[34]提出一种利用DCT将隐写载体转换到频域,从而将水印隐藏其中的方法。Kanhe等[11]提出一种基于浊音和清音的音频隐写方法,通过更改浊音和清音的DCT系数,可以在无声帧中嵌入更多信息。

在基于变换域的音频隐写算法中,DWT是用于数据隐藏的最佳变换方法,能够准确给出时域音频信号的频域特征[7]。葛倩蓉等[35]认为基于DCT的音频隐写在嵌入秘密信息的过程会导致音频质量的下降,首先对隐写载体进行DWT变换,并利用一个序列作为同步信号,然后将同步信号和经过Arnold变换的秘密信息嵌入到低频系数的均值中,能够获得更好的音频质量。文献[36-39]先对隐写载体进行DWT变换,再利用时域中的LSB方法、扩频方法等进行秘密信息的嵌入,取得比原方法更大的隐藏容量和更高的音频质量。

2.3 压缩域方法

当前网络上传播的音频文件,大多是经过压缩的文件。利用压缩的音频文件进行隐写不仅能够缓解传输带宽问题,也能降低隐写通信的风险。然而,音频压缩算法已经充分地去除了音频数据的冗余和"无用信息",因此基于压缩域的音频隐写算法相对更难。常用的音频压缩格式是MP3和AAC,其中MP3最为流行。基于IP的语音传输(VoIP)作为一种网络语音通话技术,广泛应用于智能手机和计算机的多种社交媒体中,其中采用的语音编码是通过多种有损压缩的方法(如波形编码、参数编码、混合编码)对语音数据进行压缩,因此可将VoIP隐写方法归为压缩域方法。

2.3.1 MP3隐写方法

Atoum等[40]将在MP3文件中隐写的方法总结为压缩前嵌入信息、压缩中嵌入信息和压缩后嵌入信息3种方法,并在之后对MP3隐写中不同LSB技术组合进行比较分析[41]。通过分析可以发现,压缩前嵌入信息会造成信息丢失,提取度低,鲁棒性差;压缩中嵌入通常会存在低容量的问题。当前MP3隐写算法主要是在压缩后嵌入,无需编解码过程且效率高,以Huffman编码算法为代表[42]。Huffman编码算法具有高透明性和安全性的特点,但隐藏容量相对较小,敖珺等[43]新增10对Huffman码字,又与大容量MP3比特流音频隐写算法[44]融合,使得隐藏容量大幅增加。Yi等[45]针对当前的MP3音频隐写方法存在安全性弱、嵌入率低和自适应性差的问题,提出了自适应霍夫曼编码映射方法,根据音频信号的心理声学模型对编码表进行动态修改,从而提高隐写嵌入的效率和隐蔽性,并且采用自适应的方法,针对不同的音频信号进行优化,以提高算法的适应性。Yang等[46]从隐写分析角度出发,考虑到当前大多音频隐写方法难以有效抵挡基于QMDCT(Quantified modified discrete cosine transform)系数的分布扰动提出先进的MP3隐写分析方法,提出了一种将QMDCT系数的linbits作为嵌入域的自适应双嵌入MP3隐写方案,秘密信息以二进制STC(Syndrome-trellis codes)通过LSB嵌入到每一层。第1层使用的代价函数是利用掩蔽效应设计的。在第2层中,根据第1层的嵌入结果对第2层的代价函数进行修正,并根据双层嵌入结果对linbits进行修正。

2.3.2 AAC隐写方法

研究者认为,AAC具有更好的压缩率和音质[47],但AAC音频的适用性较低,因此AAC隐写方法的研究相对较少,王昱洁等[48]将基于AAC隐写方法分为未压缩的时域信号、频域实数信号以及频域量化值中的3种,针对时域或频域量化过程会损失部分信息和造成提取错误的问题,提出一种基于MDCT(Modified discrete cosine transform)量化系数的以AAC为载体的隐写方法。Li等[49]提出了一种基于遗传算法的AAC音频隐写算法,通过自适应调节MDCT系数并利用AAC压缩技术隐藏消息,主要是采用遗传算法对AAC压缩过程中的MDCT系数进行自适应调节,采用AAC压缩技术实现隐写,避免了在音频质量和文件大小之间的平衡,提高了隐蔽性和抗攻击性。Ren等[50]认为现有AAC隐写算法无法保持Huffman码字直方图分布和帧内、帧间QMDCT系数的统计特性,导致隐写在统计特性上

出现安全性低的问题,提出一种多视角统计失真的安全AAC隐写方法。该方法基于多视角统计失真,结合Huffman码字的统计分布、QMDCT的帧内和帧间统计相关性以及心理声学模型中的频率感知掩蔽,同时采用随机扰动和编解码结构对隐写信息进行保护,提高了隐写质量和安全性。

2.3.3 VoIP隐写方法

VoIP在网络社交媒体中被广泛地应用,因此VoIP所采用的压缩音频也是作为音频信息隐藏的主要载体,但因其低冗余性而成为研究的重难点。刘小康等[51]对基于VoIP隐写和隐写分析方法进行了全面的回顾和总结,并将VoIP隐写方法总结分类。本节在此基础上,根据现有的参考文献,将VoIP隐写方法中基于音频有效载荷的隐写方法划分为基于线性预测系数参数、基于固定码本参数和基于自适应码本参数的隐写方法,如图5所示。

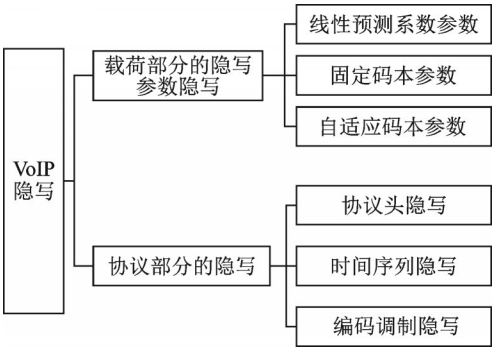

图5 VoIP隐写方法分类

Fig.5 Classification of VoIP steganography

田晖等[52]认为应当充分挖掘不同位置的载体的差异性进行VoIP隐写,能够提高不可感知性,提出一种利用LSB隐写方法的基于可量化性能分级的自适应IP语音隐写方法。孙鑫昊等[53]本着隐写载体改动越小,不可感知性越好的原则,提出一种基于最短欧式距离替换码元的VoIP隐写方法,具有较高的隐写效率和较好的语音质量。Abd Ali等[54]为了解决实时性语音中高嵌入容量的难题,提出一种同时使用互联网低比特率编解码器(Internet low bit rate codec,iLBC)和G.711编解码器对秘密语音进行编码。iLBC作为VoIP信源编解码器来压缩秘密嵌入语音载体中,并在有损信道(丢包的网络环境)上对语音质量进行校正,同时采用G.711编码器在同步时间段(每20 ms)对载体语音进行压缩,以满足VoIP的要求。

对上述传统音频隐写方法的对比分析如表4所示。首先,在时域方法中LSB隐写方法是最简单的且具有较高的嵌入容量,但是安全性相对较低。同时,回声隐藏方法也很简单,易于实现,但是在隐写分析中成本较高。其次,变换域方法具有较高的隐写容量和鲁棒性,不可感知性好,但是存在计算复杂度高的问题。最后,压缩域方法因为数据冗余小,使得隐写容量相对较小,但是都具有较好的不可感知性。因此,在具体使用中应当充分考虑复杂度、安全性和鲁棒性等实际情况,可以将两种或两种以上的

表4 传统音频隐写方法对比

Table 4 Comparison of traditional audio steganography methods

嵌入域	方法	优点	缺点
时域	LSB隐写方法	算法简单,易于实现、计算复杂度低、隐藏容量大	鲁棒性差,抗检测能力弱
	相位编码隐写方法	鲁棒性好、可以有效调节不可感知性	具有一定的相似度
	回声隐写方法	算法简单、隐藏效果好、不产生噪声、能够实现盲检测	隐藏容量小、提取效果差、信道噪声影响大
	扩频隐写方法	鲁棒性好、不可感知性好	隐藏容量小、算法相对复杂
变换域	FFT隐写方法	稳健性好、隐写容量大	不可感知性差
	DCT隐写方法	隐写容量大	计算复杂度高
	DWT隐写方法	隐写容量大、不可感知性好	计算复杂度高
压缩域	MP3隐写方法	不可感知性好	隐写容量小
	AAC隐写方法	抗隐写分析较好、鲁棒性好	高比特率下的隐写容量小
	VoIP隐写方法	实时性好、隐写方法灵活、隐写区域多、抗检测性好	隐写容量小

方法结合使用,从而在简单的情况下嵌入大量的数据,具有较高的安全性,但其本质还是人工选择冗余进行秘密信息的嵌入。因此,仍存在以下问题:(1)需要通过人工的方式选择合适的冗余进行秘密信息的嵌入,耗费大量的人力、物力及时间;(2)人工选择冗余进行秘密信息的嵌入,会导致含密载体存在严重的改动痕迹;(3)基于深度学习的音频隐写分析方法的快速发展,给传统的音频隐写方法带来前所未有的挑战。

3 基于深度学习的音频隐写

深度学习具有较强的特征表示能力,在语音增强[55]、说话人验证系统攻击[56]和语音欺骗检测[57]等任务中发挥着巨大的作用。深度学习的应用推动了信息隐藏的快速发展,基于深度学习的图像隐写取得较好的效果,并出现多种应用场景下的方法[58-60],但是这些模型对数字音频并不适用。同时,基于深度学习音频隐写分析的良好性能[61-63]却给传统音频隐写带来了巨大的挑战。

近年来,部分学者针对音频独有的特点,提出基于卷积神经网络(Convolutional neural network, CNN)[64]、生成对抗网络(Generative adversarial networks, GAN)[65]和递归神经网络(Recurrent neural networks, RNN)[66]等多种深度学习的音频隐写方法,促进了音频隐写的快速发展。本节对基于深度学习的音频隐写的发展进行梳理,分析并总结已有的基于深度学习的音频隐写方法,根据现有基于深度学习的音频隐写方法的载体情况,将基于深度学习的音频隐写方法分为嵌入载体式音频隐写、生成载体式音频隐写和无载体式音频隐写。

3.1 嵌入载体式音频隐写

基于深度学习的嵌入载体式音频隐写是指在数字音频上通过深度学习的方法完成秘密信息的嵌入和提取,根据设计方法分为Encoder-Decoder结构、自动学习嵌入代价和基于对抗样本的方法。

3.1.1 Encoder-Decoder结构

Encoder-Decoder结构的隐写方法是利用训练好的深度神经网络将秘密信息在音频载体中进行隐写和提取,仅需要训练所采用的模型。Kreuk等[67]认为基于深度学习的图像隐写模型不适用于音频隐写,利用门控卷积神经网络(Gated convolutional neural network, GCNN)[68]进行编码与解码,提出基于深度神经网络的音频隐写模型。该模型将短时傅里叶变换(Short time Fourier transform,STFT)和逆STFT作为网络内的可微层,从而对网络训练施加了重要约束。

模型的目标是从接收的$\hat{C}$中恢复时域波形,进而获取音频中的秘密信息,但仅从STFT的振幅中恢复$\hat{C}$不是一个简单的问题[69],因为相位不匹配会导致失真。为克服这种相位恢复,Griffin等[70]提出了一种经典交替投影算法,但这种方法会产生带有明显伪影的载波,且这种方法无法作为中间层构造一个端到端的隐写模型。因此,通过S和S'的损失函数来解决上述问题,如图6所示,C为载体音频,M

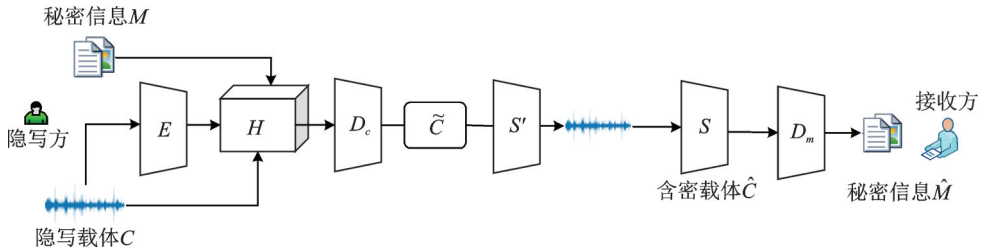

图6 基于GCNN的音频隐写方法(Hide & Speak)

Fig.6 GCNN-based audio steganography method (Hide & Speak)

为秘密信息，E表示载体音频编码器，则$H=[E(C),C,M]$为三者的连接，通过含密载体解码器$D_c(\cdot)$得到含密载体频谱$\tilde{C}=D_c(H)$，再利用c的相位$\angle c$通过$S'(\tilde{C},\angle C)$得到语音，$S(\cdot)$为STFT。接收方将接收到的语音经过$\tilde{C}=S(S'(\tilde{C},\angle C)$得到含密载体$\tilde{C}$，$S'()$为逆STFT，最后通过$D_m(\cdot)$秘密信息解码器获取重建秘密信息$\hat{M}=D_m(S)$，则损失函数为

$$L(C,M)=\lambda_c\left\|C-\hat{C}\right\|+\lambda_m\left\|M-\hat{M}\right\| \tag{1}$$

式中λ_c、λ_m分别表示权值系数。

Takahashi等[71]认为Hide & Speak方法[67]尽管证实了传输过程中引入部分失真仍具有较好的鲁棒性，但是难以在一个混合音频中的单个音频中隐藏信息，例如音乐的单个乐器轨道，并从混合音频中源分离的音频恢复秘密信息。针对上述问题，文献[71]提出了一种基于DNN的源混合和分离的音频隐写(Source mixing and separation robust audio steganography, MSRAS)模型，如图7所示。m_i、c_i ($i=1,2,\cdots,n$)分别表示秘密信息和载体语音，通过编码器后可得到含密载体c_i，接收方对接收到的混合语音经过源分离的操作得到$\tilde{c}_i$，最后通过解码器获得重建的秘密信息$\tilde{m}_i$。该方法为避免相位重建的问题，将音频隐藏在时域内，与其他未知的音频混合，通过源分离的方法提取秘密信息；其主要针对音乐，使创作者能够独立地在声源中隐藏信息，实现音乐的版权保护。

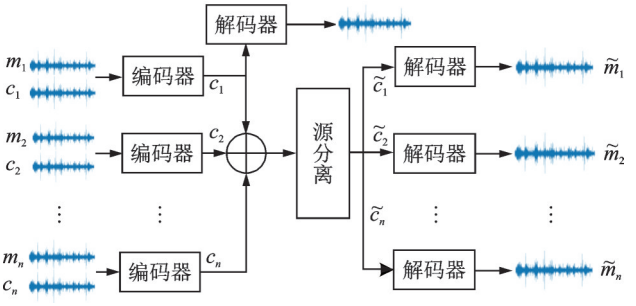

图7 MSRAS模型

Fig.7 MSRAS model

Cui等[72]考虑到当前基于深度学习的音频隐写方法无法将图片直接隐写在音频中，提出一种可以在音频中嵌入图片信息的音频隐写方法(Deep neural network based image-to-audio steganography, DITAS)。该方法没有直接隐藏秘密图像，而是利用多级网络将秘密图像的残差分阶段逐步嵌入到音频载体中，利用基于U-Net[73]省略全连接的2D全卷积神经网络构成编码器和解码器，编码器对需要隐藏的图像进行编码，将其添加到音频载体STFT的频域中。在隐藏过程中，残差随着阶段的增加而变得更加稀疏，这不仅使有效载荷容量的控制更加灵活，而且残差的稀疏性使得隐藏更加容易。同时，为了确保音频的低失真和恢复图像的质量，采用以下损失函数表示

$$L(s,s',C,C')=\beta\|s-s'\|+(1-\beta)\left\|C-\hat{C}\right\|_2 \tag{2}$$

式中：$\beta\in[0,1]$；s、s'分别表示隐藏图像和提取后的秘密图像；C、C'分别表示载体音频和含密音频，并分别利用MSE和均方绝对误差(Mean absolute error, MAE)对音频和图像进行训练。

通过STFT表示音频可以自然地将图像隐藏在音频中，也能够保留局部性，取得较好的效果。Geleta等[74]认为STFT同时具有幅度和相位变换，在重建音频时存在相位重建的问题，而短时离散余弦变换(Short-time discrete cosine transform, STDCT)只有实数变换。因此，为了避免相位重建问题，提出了一种基于STDCT的新型残差网络结构的音频隐写模型PixInWav，具体框架如图8所示。

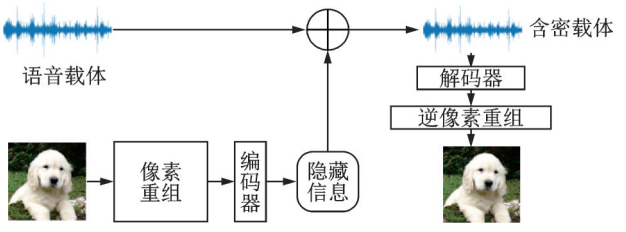

图8 PixInWav模型

Fig.8 PixInWav model

DITAS方法利用多级网络将图像的残差分阶段逐步嵌入到多个音频载体中,则隐藏一张图像就需要多个音频载体。PixInWav方法则是将1幅图像嵌入1个音频载体中,为了很好地将图像颜色信息嵌入语谱图中,在三通道的RGB图像上添加一个零通道,并进行2像素 × 2像素重组操作[75],将这4个通道重新排列到空间维度,将颜色信息分布到空间域。在解码网络端提取图像时,通过逆像素重组将空间信息重新排列回颜色通道。在损失函数上,PixInWav方法相较于DITAS方法增加了音频上的软动态时间规整(Soft dynamic time warping, Soft-DTW)损失[76]作为损失函数的一项。实验表明,增加DTW损失函数会在图像上引入很小的失真,这一边变化对人类而言并不易察觉;相反去除该项后,音频信噪比下降超过10 dB。

Alonso等[77]在Geleta等[74]基础上提出一种新的具有强鲁棒性的音频隐写方法。该方法通过修改原方法的损失函数,将STDCT 替换为STFT,并在编码过程中引入冗余进行纠错,提高了隐写的鲁棒性和感知透明性。

3.1.2 自动学习嵌入代价的方法

利用Encoder-Decoder结构的隐写方法可以较好地在音频载体中嵌入和提取秘密信息,同时也可以设计更好的嵌入失真代价,使秘密信息嵌入后,含密音频的隐写失真最小,减少因嵌入音频带来的异常[78]。尤其是GAN的提出,模型如图9所示,推动了基于深度学习的信息隐藏的发展。Yang等[79]利用GAN实现音频隐写在时域内的最优嵌入,提出了一种基于GAN的音频隐写最佳嵌入方法。该方法由基于U-Net的生成器、嵌入模拟器和鉴别器组成。通过生成器和鉴别器之间的对抗性训练,可以训练生成器学习音频样本的嵌入概率。实验结果表明,该框架能够自动学习自适应嵌入概率,比传统的LSB方法和基于AAC的音频隐写方法具有更好的安全性。

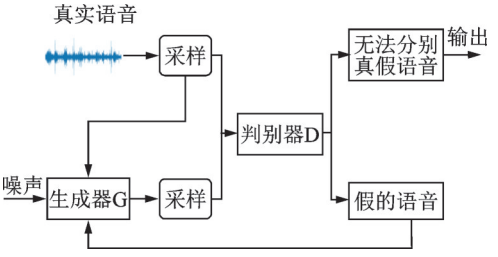

图9 GAN模型

Fig.9 GAN model

Jiang等[80]将生成对抗训练应用于音频隐写任务,提出一种基于深度卷积GAN的智能嵌入应用的轻量级生成式音频隐写模型DCGAN,并且该模型的参数小于5 MB,可用于物联网中的许多智能设备,其模型如图10所示。该模型与文献[67]相似,利用STFT将秘密音频和载体音频从时域变换为频域作为编码器的输入,利用ISTFT将含密载体和秘密信息转换为音频。在编码器和解码器均采用卷积神经网络CNN作为基础机构,基于CNN的隐写分析器作为鉴别器,从而获取具有更高隐蔽性的含密音频。

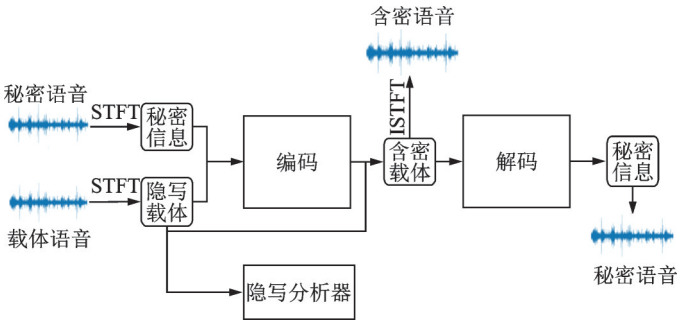

图10 基于深度卷积GAN的音频隐写

Fig.10 Audio steganography based on deep convolutional GAN

岳峰等[81]认为当前基于GAN的音频隐写方法忽略了隐写容量和不可感知性的高要求,仅是针对抗隐写能力而言,因此提出了一种基于批处理归一化(Batch normalization,BN)优化SNGAN的自适应隐写方法(Batch normalization optimized spectral normalization GAN, BNSNGAN)。该方法包含编码器、解码器和隐写分析器,且在GAN的生成器和判别器中分别利用了频谱归一化(Spectral normalization,SN)[82]和BN[83],通过3部分的协同学习,使音频隐写在嵌入容量、不可感知性和抗隐写分析上实

现均衡。在音频信号预处理中,该方法利用音频的时域补零方法使得秘密音频和载体音频的复数矩阵能够级联,导致音频仍是完全嵌入。

张国富等[84]针对音频隐写的不可感知性和抗检测性差的问题,提出一种基于MIDI和GAN的音频隐写方法。该方法利用Music21工具包构建带有索引的MIDI音符字典,利用GAN网络的生成器、提取器和判别器网络进行训练。具体步骤:(1)根据秘密音频得到对应的浮点数;(2)利用生成器对上述浮点数进行处理得到载密MIDI;(3)利用提取器对含密MIDI进行解密。

3.1.3 基于对抗样本的方法

对抗样本[85]是根据目标机器学习模型的梯度添加一些不易察觉的噪声而产生的扰动输入,利用巧妙设计的对抗样本可以成功地欺骗模型。

Wu等[86]认为当前基于CNN的分类器容易被对抗样本所欺骗,提出了一种基于对抗样本的时域音频隐写方法,与那些高度依赖现有的嵌入代价的图像隐写方法[87]不同,包括嵌入代价的初始化、迭代过程中更新的策略,以及从所有迭代的临时结果中得到最终嵌入代价的方式都是不同的。该方法可以从一个固定甚至随机的嵌入代价开始,利用对抗性攻击迭代更新初始嵌入代价,直到获得较好的安全性能。

Chen等[88]针对基于深度学习的音频隐写分析给传统音频隐写带来的问题,提出一种利用对抗样本对传统方法嵌入的含密载体进行再次训练,达到欺骗隐写分析器的目的,如图11所示。首先采用传统音频隐写LSBM方法将秘密消息嵌入音频载体,然后在含密音频载体上加入扰动,构建噪声隐写音频,最后将含密载体通过训练好的隐写分析仪进行误分类。扰动在对抗过程中不断优化,旨在寻求最佳扰动,以增强含密载体的不可感知性和不可检测性。该方法在训练过程中,首先找到一个合适的扰动来欺骗隐写分析器,然后再集中优化扰动,使得最小化损失函数,获取更好的优化效果。在音频隐写的实际应用中,可以在音频载体上先进行对抗性攻击过程获得的最优扰动,构建具有对抗性的音频载体,然后再根据应用场景选择合适的音频隐写算法,将秘密消息嵌入对抗性音频载体生成含密载体,从而获取较好的感知质量和不可检测性。

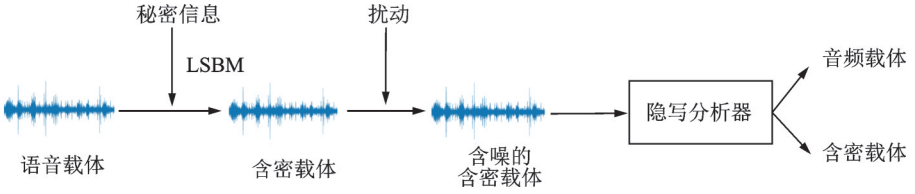

图11 基于对抗样本的音频隐写

Fig.11 Audio steganography based on deep adversarial example

对嵌入载体式音频隐写方法的代表模型进行对比,如表5所示,音频隐写主要在时域和频域中隐写,时频域转换主要是通过STFT进行(PixInWav为STDCT),隐藏的秘密信息主要为音频,评价指标以SNR为主,数据集主要是采用TIMIT。

3.2 生成载体式音频隐写

生成载体式音频隐写方法不需要事先给定音频载体,而是利用深度神经网络生成适合隐写的音频载体,然后在生成的载体上完成秘密信息的嵌入和提取。相较于嵌入载体式音频隐写方法,生成载体式音频隐写多了生成载体的过程,经过现在声码器技术的快速发展,利用声码器可以生成较高质量的语音。

Chen等[89]提出一种将深度学习方法和传统方法相结合的生成载体式音频隐写方法,利用深度学习

表5 嵌入载体式音频隐写模型对比

Table 5 Comparison of embedded cover-based audio steganography models

模型	嵌入域	时频变换	秘密类型	模型特点	评价指标	数据集
Hide & Speak[67]	频域	STFT	音频	基于门控卷积神经网络的音频隐写模型,不可感知性好,且可以在单个音频中隐藏多个秘密信息	SNR	TIMIT YOHO
DCGAN[80]	频域	STFT	音频	基于GAN的音频隐写模型,具有少量参数的轻量级模型,可用于物联网的许多设备中	SNR	TIMIT Librispeech
U-NetGAN[79]	时域		音频	利用GAN实现音频隐写在时域内的最优嵌入,抗隐写分析能力强	Error rate	UME-ERJ WSJ
CNNAE[86]	时域		音频	基于对抗样本的音频隐写模型,不依赖于现有的隐写成本,抗隐写分析能力强	Accuracy	TIMIT
BNSNGAN[81]	时域		音频	可以实现任意长度秘密音频的嵌入,并且在不可感知性,隐写容量和抗检测性上有较好的均衡	SNR ODG BER	TIMIT Librispeech
MSRAS[71]	时域		音频	对源混合和分离具有鲁棒性的音频隐写,秘密信息被单独隐藏在某些源中,并再与其他源混合和源分离后能够准确恢复秘密信息,鲁棒性强	SNR	MUSDB18
DITAS[72]	频域	STFT	图片	利用多级网络将图片隐藏到音频中的多模态隐写方法,有效载荷容量控制灵活,隐藏容易	MSE PSNR	TIMIT LJ Speech VOC2012
PixInWav[74]	频域	STDCT	图片	基于STDCT的多模态隐写方法,能够在不影响隐写载体质量的情况下独立编码图像,并可以离线编码图像	SNR SSIM	FSDnoisy18K ILSVRC2012
MIDI-GAN[84]	时域		音频	突破有载体隐写在不可感知性和抗隐写检测性的限制,将秘密信息转化为MIDI音频,从而提高载密音频的有效性安全性	MOS	MIDI
PixInWav2[77]	频域	STFT	图片	对PixInWav模型损失函数的修改,STDCT替换为STFT,以及在编码过程中引入冗余进行纠错等增强鲁棒性	SSIM PSNR SNR	ILSVRC2012 FSDnoisy18k
LSBMAE[88]	时域			在LSBM方法嵌入得到的含密音频载体上加入扰动并通过训练好的隐写分析仪进行误分类,具有高感知的含密载体,并且抗检测性强	PSNR SNR Accuracy	TIMIT UME

的方法生成与真实语音相似的语音作为音频载体,再利用传统隐写方法将秘密信息嵌入其中,并训练一个判别器来逼近生成的语音载体和真实语音的相似性,通过训练隐写分析器逼近含密载体和隐写载体的相似性。该音频隐写方法分别利用GAN生成音频载体,LSBM嵌入秘密信息,如图12所示。首先,将隐写载体音频作为GAN生成器的输入,生成不可区分的隐写载体音频。然

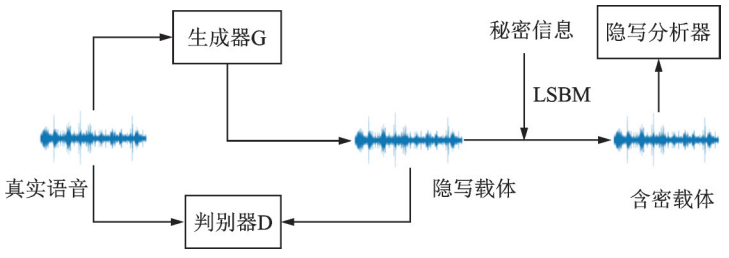

图12 基于GAN的生成载体式音频隐写

Fig.12 Generating cover-based Audio steganography based on GAN

后，利用传统的隐写算法LSBM[1]将秘密信息隐藏到生成的隐写载体音频中，得到含密音频。最后，将含密音频传递给训练好的隐写分析器，将其误分类为原始载体音频。为了提高不可检测性，分别利用构建的两个损失函数来限制模型的收敛，一是通过原始隐写载体和生成隐写载体之间的相似性，二是通过含密载体和原始隐写载体之间的相似性。实验结果表明，与传统音频隐写方法相比，所提音频隐写方法具有较好的优越性，但是并未与其他深度学习的隐写方法对比。

进一步，Chen等[90]基于文本合成语音算法WaveGlow[91]和WaveNet[92]提出了两种分布保持的可证安全隐写算法。在嵌入秘密信息阶段，两种方法均是先将输入文本通过语谱图生成模型转换为Mel谱图，基于WaveGlow的方法是将秘密信息映射为高斯向量，然后与Mel谱图一起输入到WaveGlow中产生含密载体音频，而基于WaveNet的方法是先将Mel谱图通过WaveNet得到样本分布，然后利用自适应算术解码将秘密信息嵌入到样本分布中获取含密载体音频。在提取秘密信息阶段，均是先利用语音识别技术将语音识别为文本，再按照嵌入阶段的方法将文本转换为Mel谱图，基于WaveGlow的方法是利用其本身的可逆性，通过Mel谱图获取秘密信息的高斯分布，再通过秘密信息映射的可逆性将高斯分布转换为秘密信息。WaveNet的方法是利用WaveNet将Mel谱图生成相同的分布，在具有相同分布和含密载体音频的情况下，使用自适应算术编码提取秘密信息。同时，针对利用STC和SPC（Steganographic polar codes）编码的隐写方法存在秘密信息嵌入的计算复杂度较高的问题，Chen等[93]又提出一种基于深度生成模型的载体可重现隐写方法，分别利用文本生成语音和图像的技术生成隐写载体实现了载体可重现的隐写。利用音频作为载体的隐写方法与前一工作相似，先将输入文本通过语谱图生成模型转换为Mel谱图，通过WaveGlow的方法获取隐写载体音频，然后利用利用自适应算术解码将秘密信息嵌入到隐写载体获取含密载体音频。提取阶段的方法与基于WaveNet的方法相似。

3.3 无载体式音频隐写

生成载体式音频隐写和嵌入式音频隐写方法本质上都是在音频载体中嵌入秘密信息，尽管都取得了较好的效果，但是不可避免地会对载体造成损伤，致使能够被隐写分析器检测到。无载体式音频隐写并不是不需要载体，是指在隐写前不需要载体，而是由隐写模型根据秘密信息直接生成含密载体。

Yang等[94]提出一种基于RNN的生成载体音频的隐写模型（AAG-Stega），根据需要嵌入的秘密比特流自动生成高质量的含密音频。在音频生成过程中，可以根据每个音符的条件概率分布对其进行合理编码，然后根据位流控制音频生成。同时，可以通过精细调节编码部分来控制信息嵌入率，从而实现隐蔽性和隐藏容量的同步优化。具体来说，AAG-Stega模型利用长短时记忆网络（Long short-term memory，LSTM）[95]将音频建模为每个时刻条件概率的乘积，再利用注意力机制[96]解决梯度消失的问题，同时获取每个输出相对于该步骤获得的关注度，然后与当前步骤LSTM的输出连接，最后生成含密音频，如图13所示。

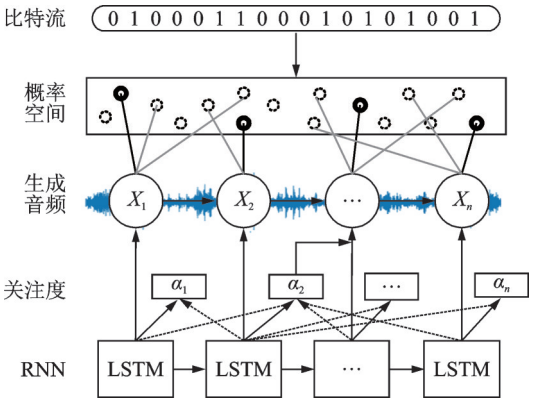

图13 基于RNN的无载体音频隐写

Fig.13 Coverless audio steganography based on RNN

Li等[97]提出一种基于GAN的无载体音频隐写方法，利用音频合成模型Wave GAN[98]作为生成模块的基础，并将输入的秘密音频直接生成为含密载体音频实现生成式隐写，如图14所示。通过生成器的核心部分（线性层、五组转置卷积及其激活函数）生成音频，并增加后处理部分降低噪声提高生成的隐写音频质量。判别器的结构与生成

器的结构相反,秘密音频输入到生成器后,经过核心部分后将秘密音频输入到后处理部分,得到最终的隐写音频。判别器由5组Conv1d、1个线性层和它们之间的激活函数组成;并且在每个Conv1d之间增加一个相位混洗操作,随机改变音频的相位,可以去除周期性的噪声影响。因此,判别器能够更准确地判断,生成器生成的音频听起来更真实。

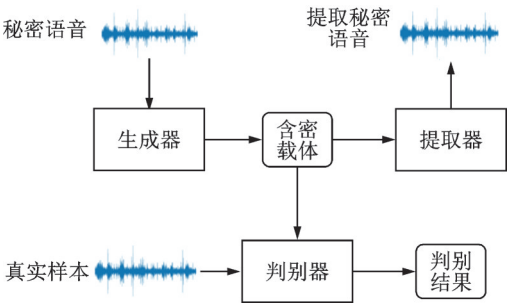

图14 基于GAN的无载体音频隐写

Fig.14 Coverless audio steganography based on GAN

3.4 分析对比

近3年来,基于深度学习的音频隐写取得了较好的效果,利用深度学习通过嵌入、构造或生成含密载体,不仅不需要手动寻找冗余,而且获得了较高的性能。通过分析对比当前基于深度学习的音频隐写方法,能够掌握和了解其优缺点,便于下一步开展具体的研究和实际应用。当前基于深度学习的音频隐写方法分析对比如表6所示。但就目前的隐写模型而言,仍存在以下不足:

(1) 生成载体式音频隐写模型少、生成音频质量差。基于生成载体式的隐写模型高度依赖于生成的音频质量,从而提高嵌入秘密信息的安全性,但基于RNN生成的音频存在质量不高的问题。在下一步研究中,可以利用GAN生成清晰、高质量的载体音频。

(2) 安全性差。基于Encoder-Decoder结构的方法通常具有大容量隐写的特点,导致难以抵抗隐写分析模型的检测。

(3) 音频的损伤。上述方法往往会通过隐写载体、含密载体、秘密信息以及恢复的秘密信息来约束损失函数,还会经过池化和归一化的操作,使得不能无损地从含密载体中恢复出原始音频。

(4)可解释性差。通过神经网络嵌入秘密信息,难以理解其中蕴含的详细情况,容易产生信任问题。

表6 基于深度学习的音频隐写方法对比

Table 6 Comparison of audio steganography methods based on deep learning

方法	实现过程	优点	缺点
生成载体式音频隐写	利用LSTM和Attention、GAN生成载体音频,再通过哈夫曼树等传统隐写方法实现秘密信息的嵌入与提取	无需准备音频载体,能够直接生成音频载体	受模型及训练过程的影响,生成音频载体质量不高
嵌入载体式音频隐写	利用CNN、GAN等自动学习实现秘密信息的嵌入与提取	自动学习嵌入代价、秘密信息的嵌入与提取	计算效率低,秘密音频的损伤
无载体式音频隐写	利用GAN模型根据秘密信息直接生成含密音频	能够根据秘密音频直接生成含密载体,抗检测能力强	受秘密音频的限制,隐写容量低

4 进一步研究方向

数字音频广泛应用于网络数字媒体中,是理想的隐写载体,同时音频隐写经过了一定的发展,取得了较好的效果。但目前该领域的研究工作还相对较少,尤其是基于深度学习的音频信息隐藏仍处于起步阶段,进一步研究方向包括以下方面:

(1)基于深度学习的隐写方法。一方面,基于深度学习的音频隐写算法也仅在时域和频域中进行

隐写，对于压缩域的基于深度学习的隐写算法还未涉及，可以在音频压缩域隐写算法中，通过神经网络学习嵌入位置，实现大容量安全隐写。另一方面，基于深度学习的音频隐写方法中采用的音频类型还比较单一，仍是以WAV格式为主，VoIP类型语音作为网络中常用的语音，是社交媒体中面向公开信道更好的载体，但因低冗余性而成为研究的重难点，已有多种传统的隐写方法[53-54]，但针对社交媒体中常用的VoIP类型语音的深度学习隐写方法还未涉及。

(2)鲁棒音频隐写方法。当前各类社交媒体均可以实现音频对话与传输，但往往需要对音频数据进行转码操作。秘密信息嵌入的载体类型可能与传输的类型不同，或者传输过程中需要编码与解码的操作，而这一类的操作都会带来一定的噪音，甚至使得接收方无法较好地提取秘密信息。因此，在设计音频隐写方法时，应当充分考虑鲁棒性，更甚至需要结合社交媒体的转码特点，设计贴合实际应用的音频隐写方法，实现可在社交媒体中进行可靠传输的鲁棒隐写。另外，当秘密信息为音频时，对鲁棒性的要求更高，不仅希望提取的秘密信息准确，也希望提取的秘密语音中说话人的个性特征完整。

(3)轻量级隐写方法。为了获取较好的隐写效果和良好的不可检测性，通常会通过深度网络模型将秘密信息嵌入载体中，并设置较深的网络模型去提取秘密信息，这样会增加计算复杂度和模型参数量。当前更多的是在小型可移动智能设备中应用社交媒体的公开信道发送信息，因此音频隐写模型也应当适应该类设备和软件，就需要模型有计算复杂度更小的轻量级隐写算法。当前基于深度学习的音频隐写方法仍是采用传统CNN的方法[80]，通过设置较浅的深度来满足要求，但是为了隐藏性能更好和模型更轻量，可以展开轻量级结构的研究[99-100]，以及借鉴语音增强[101]和语音识别[102]中轻量化模型进行隐写的嵌入和提取。

(4)行为安全的音频隐写方法。在公开信道上传输含密载体时，不仅需要考虑鲁棒性和抗检测性，也需要确保隐写载体的内容是合理的，对于第三方而言是正常的、可理解的[90]。生成载体式和无载体式隐写方法中含密载体的音频与真实音频具有很高的相似度，但在内容上仍无法确保是具有逻辑性的。因此，需要紧贴实际应用场景，在考虑安全的前提下，要确保载体内容的合理性和逻辑性。

(5)语音实时隐蔽通信系统。音频隐写最终需要面向实际应用，即利用音频隐写构建隐蔽通信系统。隐蔽通信系统中常常将通信速率作为评价指标之一[103]，音频隐写构建的隐蔽通信系统不仅希望传输速率更高，对于音频这种需要分帧和加窗进行预处理的载体也希望在嵌入和提取秘密时具备更高的速率，即具备良好的实时性。面向公开信道构建基于音频隐写的隐蔽通信系统，不仅希望能够具备良好的不可见性和鲁棒性等优点，也需要具备较好的实时性，即确保能够实时嵌入、传输及提取秘密信息。

5 结束语

本文以数字音频作为隐写载体构建隐蔽通信为前提，从音频隐写研究的角度出发，首先面向音频隐写的基本内容，对音频隐写的问题描述、数据格式、工具以及对抗技术的隐写分析进行了详细的总结。然后根据音频隐写的发展历程将音频隐写方法进行了详细的划分，对传统音频隐写方法和基于深度学习的音频隐写方法分别进行了分析、评价和对比，并指出各类方法的优缺点。最后，在讨论了传统音频隐写方法和基于深度学习隐写方法当前存在问题的基础上，对现阶段音频隐写进一步研究方向进行了总结，并从多角度预测音频隐写任务未来的发展方向。

参考文献：

[1] BENDER W, GRUHL D, MORIMOTO N, et al. Techniques for data hiding[J]. IBM Systems Journal, 1996, 35(3/4):

313-336.

[2] HARTUNG F, KUTTER M. Multimedia watermarking techniques[J]. Proceedings of the IEEE, 1999, 87(7): 1079-1107.

[3] CVEJIC N. Algorithms for audio watermarking and steganography[M]. Oulu: University of Oulu, 2004.

[4] 张卫明, 田辉. 信息隐藏技术及应用[J]. 网信军民融合, 2017, 6(3): 75-77.

[5] GHASEMZADEH H, KAYVANRAD M H. Comprehensive review of audio steganalysis methods[J]. IET Signal Processing, 2018, 12(6): 673-687.

[6] ALSABHANY A A, ALI A H, RIDZUAN F, et al. Digital audio steganography: Systematic review, classification, and analysis of the current state of the art[J]. Computer Science Review, 2020, 38: 100316.

[7] DUTTA H, DAS R K, NANDI S, et al. An overview of digital audio steganography[J]. IETE Technical Review, 2020, 37(6): 632-650.

[8] 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 Weiming, WANG Hongxia, LI Bin, et al. Overview of steganography on multimedia[J]. Journal of Image and Graphics, 2022, 27(6):1918-1943.

[9] MBL D, MMA J. Highly transparent steganography model of speech signals using efficient wavelet masking[J]. Expert Systems with Applications, 2012, 39(10): 9141-9149.

[10] JANICKI A, MAZURCZYK W, SZCZYPIORSKI K. Steganalysis of transcoding steganography[J]. Annals of Telecommunications, 2014, 69(7/8): 449-460.

[11] KANHE A, AGHILA G. DCT based audio steganography in voiced and un-voiced frames[C]//Proceedings of the International Conference on Informatics and Analytics. New York, NY, USA: Association for Computing Machinery, 2016: 1-4.

[12] MOHAMMED D Y, DUNCAN P J, AL-MAATHIDI M M, et al. A system for semantic information extraction from mixed soundtracks deploying MARSYAS framework[C]//Proceedings of 2015 IEEE 13th International Conference on Industrial Informatics (INDIN). Cambridge, UK: IEEE, 2015: 1084-1089.

[13] HAYATI P, POTDAR V, CHANG E. A survey of steganographic and steganalytic tools for the digital forensic investigator [C]//Proceedings of Workshop of Information Hiding and Digital Watermarking. [S.l.]: [s.n.], 2007: 1-12.

[14] SU Z, ZHANG G, YUE F, et al. SNR-constrained heuristics for optimizing the scaling parameter of robust audio watermarking[J]. IEEE Transactions on Multimedia, 2018, 20(10): 2631-2644.

[15] GAMBHIR A, KHARA S. Integrating RSA cryptography & audio steganography[C]//Proceedings of 2016 International Conference on Computing, Communication and Automation (ICCCA). Greater Noida, India: IEEE, 2016: 481-484.

[16] MISHRA A, JOHRI P, MISHRA A. Audio steganography using ASCII code and GA[C]//Proceedings of 2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions) (ICTUS). Dubai, United Arab Emirates: IEEE, 2017: 646-651.

[17] NASSRULLAH H A, FLAYYIH W N, NASRULLAH M A. Enhancement of LSB audio steganography based on carrier and message characteristics[J]. Journal of Information Hiding & Multimedia Signal Processing, 2020, 11(3): 126-137.

[18] RAKSHIT P, GANGULY S, PAL S, et al. Securing technique using pattern-based LSB audio steganography and intensity-based visual cryptography[J]. Computers, Materials & Continua, 2021, 67(1): 1207-1224.

[19] DJEBBAR F, AYAD B, ABED-MERAIM K, et al. Unified phase and magnitude speech spectra data hiding algorithm[J]. Security and Communication Networks, 2013, 6(8): 961-971.

[20] DONG X, BOCKO M F, IGNJATOVIC Z. Data hiding via phase manipulation of audio signals[C]//Proceedings of 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing. Montreal, QC, Canada: IEEE, 2004: V-377.

[21] MALIK H M A, ANSARI R, KHOKHAR A A. Robust data hiding in audio using Allpass filters[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2007, 15(4): 1296-1304.

[22] ALSABHANY A A, RIDZUAN F, AZNI A H. The adaptive multi-level phase coding method in audio steganography[J]. IEEE Access, 2019, 7: 129291-129306.

[23] 唐升, 侯榆青, 卢艳玲, 等. 回声隐藏技术研究进展[J]. 电声技术, 2006(3): 37-41.

TANG Sheng, HOU Yuqing, LU Yanling, et al. Research development of echo hiding technology[J]. Audio Engineering, 2006(3): 37-41.

[24] OH H O, SEOK J W, HONG J W, et al. New echo embedding technique for robust and imperceptible audio watermarking [C]//Proceedings of 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Salt Lake City, UT, USA: IEEE, 2001: 1341-1344.

[25] ERFANI Y, SIAHPOUSH S. Robust audio watermarking using improved TS echo hiding[J]. Digit Signal Process, 2009, 19 (5): 809-814.

[26] GHASEMZADEH H, KAYVANRAD M H. Toward a robust and secure echo steganography method based on parameters hopping[C]//Proceedings of 2015 Signal Processing and Intelligent Systems Conference (SPIS). Tehran, Iran: IEEE, 2015: 143-147.

[27] WANG S, YUAN W, UNOKI M. Multi-subspace echo hiding based on time-frequency similarities of audio signals[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2020, 28: 2349-2363.

[28] MATSUOKA H. Spread spectrum audio steganography using sub-band phase shifting[C]//Proceedings of 2006 International Conference on Intelligent Information Hiding and Multimedia. Pasadena, CA, USA: IEEE, 2006.

[29] KUZNETSOV A, ONIKIYCHUK A, PESHKOVA O, et al. Direct spread spectrum technology for data hiding in audio[J]. Sensors, 2022, 22(9): 3115.

[30] GOPALAN K. Audio steganography by modification of cepstrum at a pair of frequencies[C]//Proceedings of 2008 9th International Conference on Signal Processing. Beijing, China: IEEE, 2008: 2178-2181.

[31] FALLAHPOUR M, MEGÍAS D. High capacity method for real-time audio data hiding using the FFT transform[C]//Proceedings of Advances in Information Security and Its Application. Berlin, Heidelberg: Springer, 2009: 91-97.

[32] DIEU H B, HUY N X. An improved technique for hiding data in audio[C]//Proceedings of 2014 Fourth International Conference on Digital Information and Communication Technology and Its Applications (DICTAP). Bangkok, Thailand: Springer, 2014: 149-153.

[33] 陈红松, 王禹, 余乾圆, 等. 一种基于离散余弦变换的盲性音频隐写算法研究[J]. 信息网络安全, 2013(9): 60-63.
CHEN Hongsong, WANG Yu, YU Qianyuan, et al. Research for a DCT-based blind audio steganography algorithm[J]. Netinfo Security, 2013(9): 60-63.

[34] CHARFEDDINE M, EL'ARBI M, BEN AMAR C. A new DCT audio watermarking scheme based on preliminary MP3 study[J]. Multimedia Tools and Applications, 2014, 70(3): 1521-1557.

[35] 葛倩蓉, 李梦超, 曾毓敏. 一种基于小波域的同步均值量化音频隐写算法[J]. 信息化研究, 2012, 38(4): 18-21.
GE Qianrong, LI Menghchao, ZENG Yumin. Synchronize audio information hiding algorithm based on wavelet domain and mean quantization[J]. Informatization Research, 2012, 38(4): 18-21.

[36] CVEJIC N, SEPPANEN T. A wavelet domain LSB insertion algorithm for high capacity audio steganography[C]//Proceedings of 2002 IEEE 10th Digital Signal Processing Workshop, 2002 and the 2nd Signal Processing Education Workshop. Pine Mountain, GA, USA: IEEE, 2002: 53-55.

[37] SHIRALI-SHAHREZA S, MANZURI-SHALMANI M T. High capacity error free wavelet domain speech steganography [C]//Proceedings of 2008 IEEE International Conference on Acoustics, Speech and Signal Processing. Las Vegas, NV, USA: IEEE, 2008: 1729-1732.

[38] HEMALATHA S, DINESH ACHARYA U, RENUKA A, et al. Audio steganography in discrete wavelet transform domain [J]. International Journal of Applied Engineering Research, 2015, 10(16): 37544-37549.

[39] EL-KHAMY S E, KORANY NOHAO, EL-SHERIF M H. Robust image hiding in audio based on integer wavelet transform and chaotic maps hopping[C]//Proceedings of 2017 34th National Radio Science Conference (NRSC). Alexandria, Egypt: IEEE, 2017: 205-212.

[40] ATOUM M S, IBRAHIMN S, SULONG G, et al. Exploring the challenges of MP3 audio steganography[C]//Proceedings of 2013 International Conference on Advanced Computer Science Applications and Technologies. Kuching, Malaysia: IEEE,

2013: 156-161.

[41] ATOUM M S. A Comparative study of combination with different LSB techniques in MP3 steganography[C]//Proceedings of Information Science and Applications. Berlin, Heidelberg: Springer, 2015: 551-560.

[42] 高海英. 基于Huffman编码的MP3隐写算法[J]. 中山大学学报(自然科学版), 2007, 46(4): 32-35.
GAO Haiying. The MP3 steganography algorithm based on Huffman coding[J]. Acta Scientiarum Naturalium Universitatis Sunyatseni, 2007, 46(4): 32-35.

[43] 敖珺, 李睿, 张涛. 基于MP3格式的语音隐写算法[J]. 桂林电子科技大学学报, 2016, 36(4): 315-320.
AO Jun, LI Rui, ZHANG Tao. Voice hiding algorithm based on MP3 format[J]. Journal of Guilin University of Electronic Technology, 2016, 36(4): 315-320.

[44] 刘秀娟, 郭立. 大容量MP3比特流音频隐写算法[J]. 计算机仿真, 2007(5): 110-113.
LU Xiujuan, GOU Li. High capacity audio steganography in MP3 bitstreams[J]. Computer Simulation, 2007(5): 110-113.

[45] YI X, YANG K, ZHAO X, et al. AHCM: Adaptive huffman code mapping for audio steganography based on psychoacoustic model[J]. IEEE Transactions on Information Forensics and Security, 2019, 14(8): 2217-2231.

[46] YANG Y, YU H, ZHAO X, et al. An adaptive double-layered embedding scheme for MP3 steganography[J]. IEEE Signal Processing Letters, 2020, 27: 1984-1988.

[47] BRANDENBURG K. MP3 and AAC explained[C]//Proceedings of Audio Engineering Society Conference: High-Quality Audio Coding. [S.l.]: Audio Engineering Society, 1999.

[48] 王昱洁, 郭立, 王翠平. 一种以AAC压缩音频为载体的隐写方法[J]. 小型微型计算机系统, 2011, 32(7): 1465-1468.
WANG Yujie, GUO Li, WANG Cuiping. Steganography method for advanced audio coding[J]. Journal of Chinese Computer Systems, 2011, 32(7): 1465-1468.

[49] LI C, ZHANG X, LUO T, et al. Audio steganography algorithm based on genetic algorithm for MDCT coefficient adjustment for AAC[C]//Proceedings of 2020 IEEE International Symposium on Multimedia (ISM). Naples, Italy: IEEE, 2020: 111-112.

[50] REN Y, CAI S, WANG L. Secure AAC steganography scheme based on multi-view statistical distortion (SofMvD)[J]. Journal of Information Security and Applications, 2021, 59: 102863.

[51] 刘小康, 田晖, 刘杰, 等. IP语音隐写及隐写分析研究[J]. 重庆邮电大学学报(自然科学版), 2019, 31(3): 407-419.
LIU Xiaokang, TIAN Hui, LIU Jie, et al. Survey for voice-over-IP steganography and steganalysis[J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2019, 31(3): 407-419.

[52] 田晖, 郭舒婷, 秦界, 等. 基于可量化性能分级的自适应IP语音隐写方法[J]. 电子学报, 2016, 44(11): 2735-2741.
TIAN Hui, GUO Shuting, QIN Jie, et al. Adaptive voice-over-IP steganography based on quantitative performance ranking[J]. Acta Electronica Sinica, 2016, 44(11): 2735-2741.

[53] 孙鑫昊, 王开西. 基于最短欧氏距离替换码元的VoIP隐写算法[J]. 计算机工程与应用, 2022, 58(13): 128-134.
SUN Xinhao, WANG Kaixi. Codeword replacement based on shortest Euclidean distance for VoIP steganography[J]. Computer Engineering and Applications, 2022, 58(13): 128-134.

[54] ABD ALI B Q, SHAHADI H I, KOD M S, et al. Covert VoIP communication based on audio steganography[J]. International Journal of Computing and Digital Systems, 2022, 11(1): 821-830.

[55] LI Y, ZHANG X, SUN M. A unified speech enhancement approach to mitigate both background noises and adversarial perturbations[J]. Information Fusion, 2023, 95: 372-383.

[56] 张雄伟, 张星昱, 孙蒙, 等. 说话人验证系统攻击方法的研究现状及展望[J]. 数据采集与处理, 2021, 36(5): 831-849.
ZHANG Xiongwei, ZHANG Xingyu, SUN Meng, et al. Attack methods in speaker verification system: The state of the art and prospects[J]. Journal of Data Acquisition and Processing, 2021, 36(5): 831-849.

[57] 张雄伟, 李嘉康, 孙蒙, 等. 语音欺骗检测方法的研究现状及展望[J]. 数据采集与处理, 2020, 35(5): 807-823.
ZHANG Xiongwei, LI Jiakang, SUN Meng, et al. Speech anti-spoofing: The state of the art and prospects[J]. Journal of Data Acquisition and Processing, 2020, 35(5): 807-823.

[58] 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, 2023, 45(1): 372-390.

[59] JING J, DENG X, XU M, et al. HiNet: Deep image hiding by invertible network[C]//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, QC, Canada: IEEE, 2021: 4713-4722.

[60] ZHU J, KAPLAN R, JOHNSON J, et al. HiDDeN: Hiding data with deep networks[C]//Proceedings of Computer Vision—ECCV 2018. Cham: Springer International Publishing, 2018: 682-697.

[61] 李敬轩, 胡润文, 阮观奇, 等. 基于手工特征提取与结果融合的CNN音频隐写分析算法[J]. 计算机学报, 2021, 44(10): 2061-2075.
LI Jingxuan, HU Runwen, RUAN Guanqi, et al. A CNN based audio steganalysis algorithm by manual feature extraction and result merging[J]. Chinese Journal of Computers, 2021, 44(10): 2061-2075.

[62] REN Y, LIU D, LIU C, et al. A universal audio steganalysis scheme based on multiscale spectrograms and DeepResNet[J]. IEEE Transactions on Dependable and Secure Computing, 2023, 20(1): 665-679.

[63] TIAN H, QIU Y, MAZURCZYK W, et al. STFF-SM: Steganalysis model based on spatial and temporal feature fusion for speech streams[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2023, 31: 277-289.

[64] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]//Proceedings of the 25th International Conference on Neural Information Processing Systems—Volume 1. Red Hook, NY, USA: Curran Associates Inc., 2012: 1097-1105.

[65] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//Proceedings of the 27th International Conference on Neural Information Processing Systems— Volume 2. Cambridge, MA, USA: MIT Press, 2014: 2672-2680.

[66] MIKOLOV T, KARAFIAT M, BURGET L, et al. Recurrent neural network based language model[C]//Proceedings of Eleventh Annual Conference of The International Speech Communication Association. Makuhari, Chiba, Japan: ISCA, 2010: 1-4.

[67] KREUK F, ADI Y, RAJ B, et al. Hide and speak: Towards deep neural networks for speech steganography[C]//Proceedings of Interspeech 2020. Shanghai, China: ISCA, 2020: 4656-4660.

[68] DAUPHIN Y N, FAN A, AULI M, et al. Language modeling with gated convolutional networks[C]//Proceedings of the 34th International Conference on Machine Learning. Sydney, NSW, Australia: JMLR.org, 2017: 933-941.

[69] JAGANATHAN K, ELDAR Y C, HASSIBI B. STFT phase retrieval: Uniqueness guarantees and recovery algorithms[J]. IEEE Journal of Selected Topics in Signal Processing, 2016, 10(4): 770-781.

[70] GRIFFIN D, LIM J. Signal estimation from modified short-time Fourier transform[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1984, 32(2): 236-243.

[71] TAKAHASHI N, SINGH M K, MITSUFUJI Y. Source mixing and separation robust audio steganography[EB/OL]. (2022-02-18). https://arxiv.org/abs/2110.05054v1.

[72] CUI W, LIU S, JIANG F, et al. Multi-stage residual hiding for image-into-audio steganography[C]//Proceedings of 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Barcelona, Spain: IEEE, 2020: 2832-2836.

[73] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation[C]//Proceedings of Medical Image Computing and Computer-Assisted Intervention. Cham: Springer International Publishing, 2015: 234-241.

[74] GELETA M, PUNTÍ C, MCGUINNESS K, et al. Pixinwav: Residual steganography for hiding pixels in audio[C]//Proceedings of 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Singapore: IEEE, 2022: 2485-2489.

[75] SHI W, CABALLERO J, HUSZÁR F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016: 1874-1883.

[76] CUTURI M, BLONDEL M. Soft-DTW: A differentiable loss function for time-series[C]//Proceedings of the 34th International Conference on Machine Learning. Sydney, NSW, Australia: JMLR.org, 2017: 894-903.

[77] ALONSO J R, GELETA M, PONS J, et al. Towards robust image-in-audio deep steganography[EB/OL]. (2023-05-14). https://arxiv.org/abs/2303.05007.

[78] 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 Zhangjie, WANG Fan, SUN Xingming, et al. Research on steganography of digital images based on deep learning[J]. Chinese Journal of Computers, 2020, 43(9): 1656-1672.

[79] YANG J, ZHENG H, KANG X, et al. Approaching optimal embedding in audio steganography with GAN[C]//Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Barcelona, Spain: IEEE, 2020: 2827-2831.

[80] JIANG S, YE D, HUANG J, et al. SmartSteganogaphy: Light-weight generative audio steganography model for smart embedding application[J]. Journal of Network and Computer Applications, 2020, 165: 102689.

[81] 岳峰, 朱慧, 苏兆品, 等. 基于BN优化SNGAN的自适应音频隐写[J]. 计算机学报, 2022, 45(2): 427-440.
YUE Feng, ZHU Hui, SU Zhaopin, et al. An adaptive audio steganography using BN optimizing SNGAN[J]. Chinese Journal of Computers, 2022, 45(2): 427-440.

[82] MIYATO T, KATAOKA T, KOYAMA M, et al. Spectral normalization for generative adversarial networks[C]//Proceedings of International Conference on Learning Representations. Vancouver, BC, Canada: OpenReview.net, 2018.

[83] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]//Proceedings of the 32nd International Conference on International Conference on Machine Learning. [S.l.]: JMLR.org, 2015: 448-456.

[84] 张国富, 史志远, 苏兆品, 等. 基于MIDI和对抗生成网络的音频隐写方法和系统: CN 115440234[P]. 2022-11-08.

[85] SZEGEDY C, ZAREMBA W, SUTSKEVER I, et al. Intriguing properties of neural networks[EB/OL]. (2014-02-19). https://arxiv.org/abs/1312.6199.

[86] WU J, CHEN B, LUO W, et al. Audio steganography based on iterative adversarial attacks against convolutional neural networks[J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 2282-2294.

[87] TANG W, LI B, TAN S, et al. CNN-based adversarial embedding for image steganography[J]. IEEE Transactions on Information Forensics and Security, 2019, 14(8): 2074-2087.

[88] CHEN L, WANG R, DONG L, et al. Imperceptible adversarial audio steganography based on psychoacoustic model[J]. Multimedia Tools and Applications, 2023, 82(17): 26451-26463.

[89] CHEN L, WANG R, YAN D, et al. Learning to generate steganographic cover for audio steganography using GAN[J]. IEEE Access, 2021, 9: 88098-88107.

[90] CHEN K, ZHOU H, ZHAO H, et al. Distribution-preserving steganography based on text-to-speech generative models[J]. IEEE Transactions on Dependable and Secure Computing, 2022, 19(5): 3343-3356.

[91] PRENGER R, VALLE R, CATANZARO B. WaveGlow: A flow-based generative network for speech synthesis[C]//Proceedings of 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Brighton, United Kingdom: IEEE, 2019: 3617-3621.

[92] VAN DEN OORD A, DIELEMAN S, ZEN H, et al. WaveNet: A generative model for raw audio[EB/OL].(2016-09-12). http://arxiv.org/abs/1609.03499.

[93] CHEN K, ZHOU H, WANG Y, et al. Cover reproducible steganography via deep generative models[J]. IEEE Transactions on Dependable and Secure Computing, 2023, 20(5): 3787-3798.

[94] YANG Z, DU X, TAN Y, et al. AAG-Stega: Automatic audio generation-based steganography[EB/OL]. (2018-09-10). https://arxiv.org/abs/1809.03463v1.

[95] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.

[96] CHO K, VAN MERRIENBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for

statistical machine translation[EB/OL]. (2014-03-18). https://arxiv.org/abs/1406.1078.

[97] LI J, WANG K, JIA X. A coverless audio steganography based on generative adversarial networks[J]. Electronics, 2023, 12(5): 1253.

[98] DONAHUE C, MCAULEY J, PUCKETTE M. Adversarial audio synthesis[EB/OL]. (2019-02-09). https://arxiv.org/abs/1802.04208v3.

[99] IANDOLA F N, HAN S, MOSKEWICZ M W, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size [EB/OL]. (2016-11-04). http://atxiv.org/abs/1602.07360.

[100] HOWARD A, SANDLER M, CHU G, et al. Searching for MobileNetV3[EB/OL]. (2019-05-01). https://arxiv.org/abs/1905.02244.pdf.

[101] JIA X, LI D. TFCN: Temporal-frequential convolutional network for single-channel speech enhancement[EB/OL]. (2022-01-03). https://arxiv.org/abs/2201.00480.

[102] HAN W, ZHANG Z, ZHANG Y, et al. ContextNet: Improving convolutional neural networks for automatic speech recognition with global context[EB/OL]. (2020-05-16). https://arxiv.org/abs/2005.03191v2.

[103] 戴跃伟, 刘光杰, 曹鹏程, 等. 无线隐蔽通信研究综述[J]. 南京信息工程大学学报(自然科学版), 2020, 12(1): 45-56.
DAI Yuewei, LIU Guangjie, CAO Pengcheng, et al. Covert wireless communication: A review[J]. Journal of Information Engineering University, 2020, 12(1): 45-56.

作者简介:

张雄伟(1965-),男,教授,研究方向:智能语音处理和信息内容安全,E-mail:xwzhang9898@163.com。

葛晓义(1995-),**通信作者**,男,博士研究生,研究方向:智能信息处理与信息隐藏,E-mail:lgd_gxy@163.com。

孙蒙(1984-),男,副教授,研究方向:智能语音处理和机器学习,E-mail:sunmeng@aeu.edu.cn。

宋宫琨琨(1989-),男,讲师,研究方向:智能语音处理和阵列信号处理,E-mail:sgkk@nuaa.edu.cn。

李莉(1977-),女,副教授,研究方向:信息内容安全,E-mail:754583546@qq.com。

(编辑:张黄群)