基于多特征融合和BiLSTM的语音隐写检测算法

苏兆品[1,2,3,4],张 羚[1],张国富[1,2,3,4],岳 峰[1,4]

(1. 合肥工业大学计算机与信息学院,安徽合肥 230601;2. 大数据知识工程教育部重点实验室(合肥工业大学),安徽合肥 230601;
3. 智能互联系统安徽省实验室(合肥工业大学),安徽合肥 230009;
4. 工业安全应急技术安徽省重点实验室(合肥工业大学),安徽合肥 230601)

摘 要: 针对传统互联网低比特率编解码器(internet Low Bit Rate Codec,iLBC)语音隐写主要集中在线性频谱频率系数矢量量化、码本搜索矢量量化或增益量化的单个阶段,难以应对多阶段下的联合隐写检测等问题,提出一种基于多特征融合和双向长短时记忆(Bi-Directional Long Short-Term Memory,BiLSTM)网络的iLBC语音隐写检测算法.通过分析隐写对不同阶段参数带来的影响,提取线性频谱频率系数矢量量化、码本搜索矢量量化和增益量化过程中的多种隐写特征,并分别输入到相应的BiLSTM检测网络,最后将各检测网络的结果进行融合,得到最终隐写检测结果.实验表明,所提算法可以实现多阶段下的联合隐写检测,而且在语音时长较短时,仍能取得优异的检测结果,平均检测准确率达到了90%以上.

关键词: 联合隐写检测;互联网低比特率编解码器;双向长短时记忆网络;隐写特征提取;多特征融合

基金项目: 安徽省重点研究与开发计划(No.202004d07020011,No.202104d07020001);教育部人文社会科学研究青年基金项目(No.19YJC870021);广东省类脑智能计算重点实验室开放课题(No.2020B121201001);中央高校基本科研业务费专项资金项目(No.PA2021GDSK0073,No.PA2021GDSK0074)

中图分类号: TP309 **文献标识码:** A **文章编号:** 0372-2112(2023)05-1300-10

电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20220553

A Speech Steganalysis Algorithm Based on Multi-Feature Fusion and BiLSTM

SU Zhao-pin[1,2,3,4], ZHANG Ling[1], ZHANG Guo-fu[1,2,3,4], YUE Feng[1,4]

(1. *School of Computer Science and Information Engineering*, *Hefei University of Technology*, *Hefei*, *Anhui* 230601, *China*;
2. *Key Laboratory of Knowledge Engineering with Big Data* (*Hefei University of Technology*), *Ministry of Education*, *Hefei*, *Anhui* 230601, *China*;
3. *Intelligent Interconnected Systems Laboratory of Anhui Province* (*Hefei University of Technology*), *Hefei*, *Anhui* 230009, *China*;
4. *Anhui Province Key Laboratory of Industry Safety and Emergency Technology* (*Hefei University of Technology*), *Hefei*, *Anhui* 230601, *China*)

Abstract: The traditional internet low bit rate codec (iLBC) based speech steganography mainly focuses on a single stage of the linear spectrum frequency coefficient vector quantization, the codebook search vector quantization, or the gain quantization, which is difficult to deal with the multi-stage joint steganalysis. To this end, an iLBC speech steganalysis algorithm based on the multi-feature fusion and the bi-directional long short-term memory (BiLSTM) network is proposed. Specifically, the impact of steganography on iLBC parameters is first analyzed in the linear spectrum frequency coefficient vector quantization process, the dynamic codebook search process, and the gain quantization process. Then, multiple steganographic features in the above three stages are extracted and input to three different detection models based on BiLSTM, respectively. Finally, a fusion strategy is presented to merge the detection results of each model. Experimental results show that the proposed algorithm can achieve multi-stage joint steganalysis and good detection results with an average detection accuracy of more than 90%, even if the speech duration is short.

Key words: joint steganalysis; internet low bit rate codec; bi-directional long short-term memory network; steganographic feature extraction; multi-feature fusion

Foundation Item(s): Anhui Provincial Key Research and Development Program (No.202104d07020011, No.202104d07020001); MOE (Ministry of Education in China) Project of Humanities and Social Sciences (No.19YJC870021); Guangdong Provincial Key Laboratory of Brain-Inspired Intelligent Computation (No.2020B121201001); Fundamental Research

收稿日期:2022-05-14;修回日期:2022-06-24;责任编辑:覃怀银

Funds for the Central Universities (No.PA2021GDSK0073, No.PA2021GDSK0074)

1 引言

隐写术是将秘密信息隐藏在文本、图像、音频、视频等公开载体中,以实现秘密信息和通讯行为的双重隐蔽. iLBC(internet Low Bit Rate Codec)是一种专为包交换网络通信设计的语音编解码器,解决了语音传输中网络丢包严重影响通话质量的实际问题,在实时通信系统(如电话系统、视频会议、语音流和及时消息等)领域得到了广泛的应用. 因此,面向iLBC的语音隐写成为近年来信息隐藏领域的一个研究热点. Wu等[1]基于量化索引调制(Quantization Index Modulation, QIM)方法在iLBC编码过程中的动态码本搜索阶段通过构建二叉树的方式将码本划分为左子树和右子树,提出一种固定码本(FixedCcodeBook, FCB)隐写方法,不仅提升了隐写容量,还提升了语音质量. Huang等[2]基于线性频谱频率(Linear Spectrum Frequency, LSF)系数量化进行iLBC语音隐写,用秘密信息控制码本的搜索范围,从而实现了一种QIMC(QIM-Controlled)隐写方法. Su等[3]提出了一种基于增益量化的隐写方法(Gain Quantization based Steganography, GQS),通过对增益量化表的合理划分嵌入秘密信息,在保证不可感知性的前提下追求更好的抗检测性. 苏兆品等[4]提出一种分层隐写(Hierarchical Steganography, HS)算法,可根据隐写质量需求自适应的选择不同的隐写位置.

然而,隐写是一把双刃剑,当被不法分子恶意使用时,会给国家与社会的安定带来巨大威胁. 作为隐写的对抗技术,隐来检测可分析音频中是否含有秘密信息. 早期的音频隐写检测多采用混合统计方法[5]. 随着深度学习技术的快速发展,基于深度学习的隐写分析方法通过提取隐写与非隐写的音频数据的深度特征,可得到更好的检测结果. Lin等[6]提出了一种有效的在线隐写分析方法来检测QIM隐写术. Gong等[7]针对自适应多速率编码语音的FCB域隐写方法,提出一种基于循环神经网络和卷积神经网络的隐写分析器SRCNet(Steganalysis based on Recurrent Convolutional Networks),通过结合时域和空域两方面的相关性取得了更好的隐写分析性能. Ren等[8]提出了一种通用的音频隐写分析器SpecResNet(Residual Network of Spectrogram),利用语谱图作为通用特征结合深度残差网络进行隐写分析. Yang等[9]利用注意机制来解决压缩流中基于QIM隐写术的隐写分析问题,并设计了一种基于轻量级神经网络的快速相关性提取模型FCEM(Fast Correlation Extract Model). 此外,为了满足在线隐写分析,Yang等[10]在RNN-SM的基础上使用一个隐藏层来提取载波码字之间的相关性,设计了一种快速VoIP流隐写分析方法. 李望望[11]分析了GQS帧内、相邻帧、跨帧的相关性,提出了一种基于长短时记忆网络(Long Short-Term Memory, LSTM)的GQS专用隐写分析器G-LSTM. 需要指出的是,虽然上述方法可对iLBC语音隐写进行有效的检测,但只能针对某一个编码过程. 当面临多阶段联合隐写时,检测效果有限.

2 iLBC语音编码

iLBC编码支持20 ms和30 ms两种帧长度编码,iLBC编码流程图如图1所示. iLBC编码器的输入数据被分为若干帧,每帧包含160/240(20 ms/30 ms)个采样点,编码器流程描述如下.

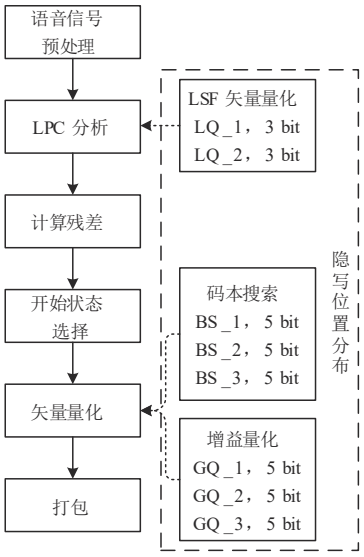

图1　iLBC编码过程与隐写位置分布示意图

(1) 将每帧分为4/6个子帧,每个子帧包含40个采样点. 30 ms帧进行两次10阶的线性预测系数(Linear Predictive Coefficient, LPC)分析,20 ms帧进行一次10阶LPC分析,得到相应LPC系数.

(2) 将LPC系数转换为LSF系数,并对LSF系数进行量化、内插以得到各个子帧的LSF系数,且由各子帧LSF系数得到各子帧对应的分析器,再由分析器对各子帧进行预测,得到各子帧的残差信号.

(3) 找到残差信号中两个连续能量最大的子帧,然后选取首或尾较大的连续57/58个样点作为开始状态.

(4) 利用差分脉冲编码调制对初始状态进行标量量化,其结果作为编码输出的一部分. 将初始状态存入码本存储区,以构成动态码本的初始值,用于对剩余样点的矢量量化.

(5) 对于剩余残差的量化,量化顺序如下:包含有

初始状态的两个连续子帧中剩余的23/22个样点;时间轴上处于初始状态之后的各个子帧;时间轴上在初始状态之前的各个子帧,矢量量化每次搜索的码本范围是动态码本,其中存储了已经被解码的对象,并且随着新的解码结果更新动态码本.

(6) 对编码结果进行打包处理.

根据上述流程,面向iLBC的语音隐写主要集中在LSF系数矢量量化、码本搜索矢量量化以及增益量化三个阶段,如FCB[1]、QIMC[2]、GQS[3]、HS[4]等隐写方法.以30 ms帧为例,从图1可以看到,在LSF系数矢量量化阶段,在两个阶段的量化过程中均能嵌入3 bit;在动态码本搜索过程,5个矢量分别进行3阶段的搜索,每个阶段可以嵌入5 bit;在增益量化过程,同样需要对5个矢量分别进行3阶段的增益系数量化,每个阶段可以嵌入5 bit.

3 iLBC语音隐写检测算法

基于多特征融合和BiLSTM的音频隐写检测算法流程如图2所示.首先,将原始语音流音频和含密语音流进行特征提取,接着分别将LSF系数量化索引、码本量化索引和增益系数量化索引分别输入模型,经过长短时记忆网络处理,得到子模型结果后进行融合,得出最终的判别结果.

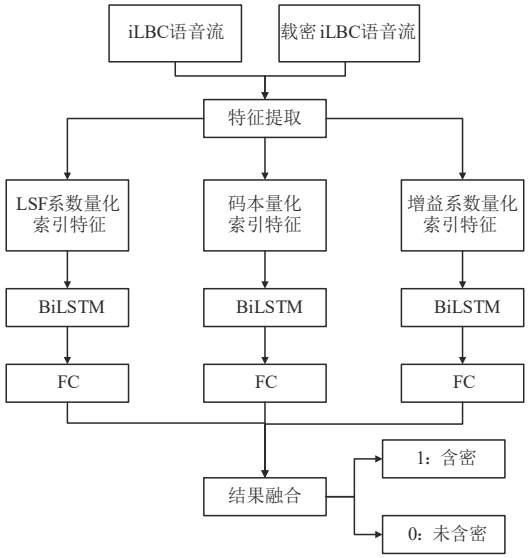

图2 iLBC语音隐写检测算法流程图

3.1 多种语音隐写特征的提取

iLBC编码的不同阶段参数对语音编解码的功能各不一样,隐写对各个阶段参数之间相关性产生的影响也不一样,所以本文以30 ms帧为例,分别提取三个阶段的隐写特征.

对于每一编码后的语音帧,隐写特征包括:

(1) LSF系数矢量量化的特征.在两个阶段的量化过程中均有3个量化索引值,第一阶段记为LQ_11、LQ_12和LQ_13,第二阶段记为LQ_21、LQ_22和LQ_23.图3是20组时长为1 s的语音在LSF系数矢量量化阶段隐写的特征差异.可以看出,在每对原始和载密语音上,隐写前后特征值差异较大且具有大量奇异值,表明特征值发生了明显的变化.因此,LSF量化索引值可以作为LSF系数矢量量化阶段隐写检测的特征.

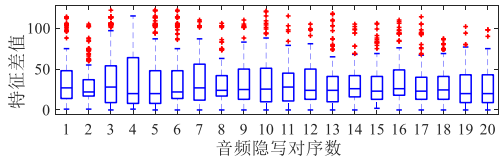

图3 LSF系数索引隐写前后的特征差异

(2) 动态码本搜索的特征.5个矢量分别进行3阶段的搜索,每个阶段会产生5个量化索引值,第一阶段记为BS_11、BS_12、BS_13、BS_14和BS_15,第二阶段记为BS_21、BS_22、BS_23、BS_24和BS_25,第三阶段记为BS_31、BS_32、BS_33、BS_34和BS_35.图4是20组1 s音频在码本搜索阶段隐写的特征差异.可以看出,在每对原始和载密音频上,隐写前后的特征值差异很大,且具有奇异值,表明特征值发生了明显的变化.因此,码本搜索矢量量化索引特征可以作为动态码本搜索阶段的隐写检测特征.

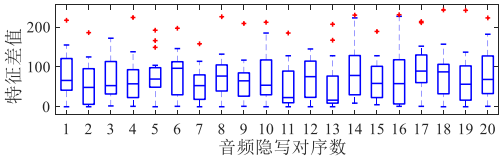

图4 码本索引隐写前后的特征差异

(3) 增益量化过程的特征提取.增益量化阶段的隐写需要对5个矢量分别进行3阶段的增益系数量化,每个阶段有5个增益量化系数索引,第一阶段记为GQ_11、GQ_12、GQ_13、GQ_14和GQ_15,第二阶段记为GQ_21、GQ_22、GQ_23、GQ_24和GQ_25,第三阶段记为GQ_31、GQ_32、GQ_33、GQ_34和GQ_35.图5是20组1 s音频在增益系数量化阶段隐写的特征差异.可以看出,在几乎每对原始和载密音频上,隐写前后的特征值差异明显,且有奇异值,表明特征值发生了明显变化.因此,增益量化索引可以作为增益量化过程隐写检测的特征.

3.2 BiLSTM网络结构设计

语音具有很强的时序性,即“上下文”相关性,而

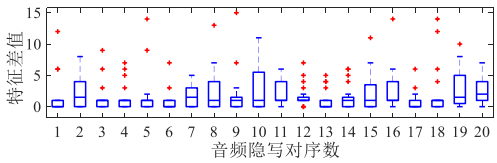

图5　增益系数量化索引隐写前后的特征差异

BiLSTM适合处理时序数据,能够充分考虑上下文信息,正向LSTM捕获序列的历史信息,反向LSTM捕获序列的未来信息[12]. 因此,本文设计了如图6所示的BiLSTM网络架构. 每个BiLSTM单元结构如图7所示,由一个正向LSTM和一个反向LSTM组成,在结构模块中具有输入门、输出门和遗忘门3个乘法结构,可以对输入的信息提取深度隐写特征.

为确保足够高的检测精度以及泛化能力,避免出现过度拟合现象,隐藏层数量通常是输入特征维度两倍的大小. 由于隐写特征均为一维数据,因此本文采用二层BiLSTM网络,正向更新过程可以表示为:

$$H^{+}=\mathrm{LSTM}^{+}(H_{t-1},X_t) \tag{1}$$

具体来说,计算过程如下:

$$C_t'=\tanh(W_{xc}X_t+W_{hc}H_{t-1}+b_c) \tag{2}$$

$$i_t=\sigma(W_{xi}X_t+W_{hi}H_{t-1}+W_{ci}C_{t-1}+b_i) \tag{3}$$

$$f_t=\sigma(W_{xf}X_t+W_{hf}H_{t-1}+W_{cf}C_{t-1}+b_f) \tag{4}$$

$$C_t=f_tC_{t-1}+i_tC_t' \tag{5}$$

$$o_t=\sigma(W_{xo}X_t+W_{ho}H_{t-1}+W_{co}C_{t-1}+b_o) \tag{6}$$

$$H_t=o_t\tanh(C_t) \tag{7}$$

同样,反向更新过程表示为$H^{-}=\mathrm{LSTM}^{-}(H_{t+1},X_t)$,计算过程参考正向更新. 网络输出结果表示为:

$$y_t=W_{yh}H^{+}+W_{yh}'H^{-}+b_y \tag{8}$$

其中,σ是sigmoid函数,$\tanh(\cdot)$是双曲正切函数,C_t和C_t'分为表示临时记忆单元值和当前时刻的记忆单元值,i_t、f_t、o_t和H_t分别表示输入门、遗忘门、当前输出门以及隐藏层的值,y_t是网络的输出结果,b_y为偏置.

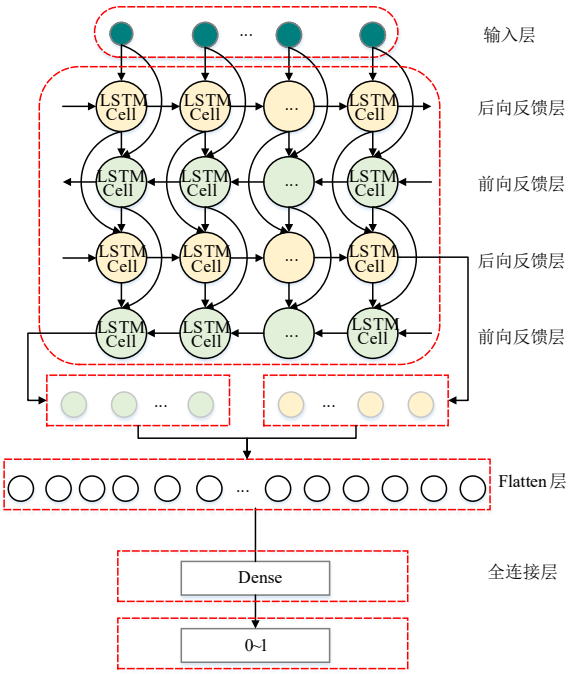

图6　BiLSTM网络结构

本文针对三个阶段的隐写设计了三个检测网络,LSF-BiLSTM、CB-BiLSTM和GQ-BiLSTM. 它们结构相同,但由于每个网络具有不同的输入,其参数略有不同. 表1给出了三个网络的结构参数,其中n表示iLBC语音帧数.

3.3　结果的融合

对于任一iLBC语音流,LSF-BiLSTM、CB-BiLSTM

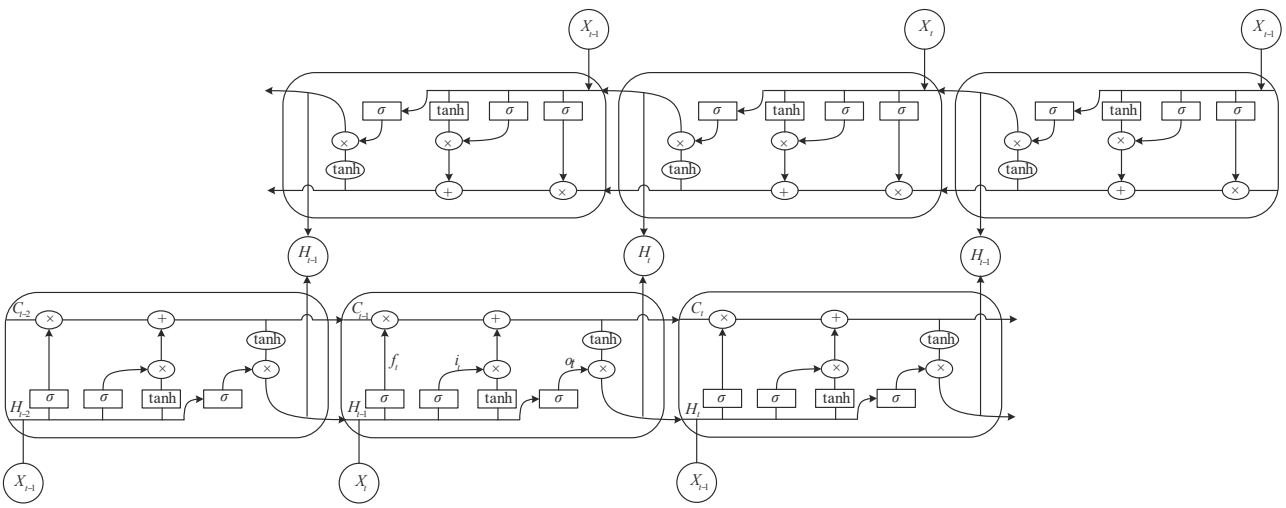

图7　BiLSTM网络单元

表1 各模型网络参数

		BiLSTM1	BiLSTM2	Flatten	Dense
LSF-BiL-STM	输入	6, *n*, batch	12, *n*, batch	12, *n*, batch	12×*n*×batch
	输出	12, *n*, batch	12, *n*, batch	12×*n*×batch	1
CB-BiL-STM	输入	15, *n*, batch	30, *n*, batch	30, *n*, batch	30×*n*×batch
	输出	30, *n*, batch	30, *n*, batch	30×*n*×batch	1
GQ-BiL-STM	输入	15, *n*, batch	30, *n*, batch	30, *n*, batch	30×*n*×batch
	输出	30, *n*, batch	30, *n*, batch	30×*n*×batch	1

和GQ-BiLSTM均会得到一个检测结果,这些结果可能存在不一致.但对于隐写检测来说,需要确认音频载体中是否含有秘密信息,所以需要对不同特征检测网络的结果进行融合[13],具体描述如下:

(1) 如果三个特征检测网络都判定待检测音频样本中未含有秘密信息,则融合判定结果"0",表示此音频样本没有隐写.

(2) 如果有一个或者一个以上网络判别为此音频样本含有秘密信息,则融合判别结果为"1",即该音频样本被隐写.

4 实验结果与分析

为了验证本文算法(简称MFSNet)的有效性,将MFSNet与SRCNet[7]、SpecResNet[8]、FCEM[9]、G-LSTM[11]四种已有隐写分析方法进行对比,考虑对多种隐写方法(FCB[1],QIMC[2],GQS[3]和HS[4])在不同样本长度和不同嵌入率情况的检测.

4.1 数据集和参数设置

测试的中文和英文语音数据集来源于https://github.com/fjxmlzn/RNN-SM,样本格式均是8 kHz采样、16 bit量化的标准PCM信号.为了对比的充分性,将iLBC不同帧长(30 ms和20 ms)均进行对比实验.

截取不同时长、不同语种的音频样本各10 000条,每个隐写分析器分别针对不同的隐写方法在做隐写分析实验时,分为两步.首先采用10%~100%不同嵌入率隐写8 000条载密样本,和原始未隐写的8 000条原始样本训练并保存模型.用剩余的2 000条对模型的判别准确率做实验,分别隐写嵌入率为10%、20%、40%、60%、80%、100%的语音样本各2 000条作为测试集,测试模型的准确率.所有对比方法的代码均基于Python编写,并在Intel(R) Core(TM) i5-8500 CPU @ 2×3.00 GHz、RAM 16.0 GB、Windows 10操作系统的个人PC上进行测试,batch-size大小设置为32.

嵌入率是每帧实际嵌入的秘密信息比特数和每帧可嵌入总比特数的比值.通常情况下,嵌入率越低代表嵌入的秘密信息越少,就越难以检测其是否隐写,对于低嵌入率隐写的检测对隐写分析器也是一个很大的挑战.

4.2 不同时长下隐写检测效果的对比

FCEM、SRCNet、G-LSTM、SpecResNet和MFSNet隐写分析器在30 ms的中文样本上对各类隐写样本做不同时长满嵌时的隐写分析实验,检测结果如表2所示.

表2 不同分析器在30 ms帧的中文语音样本上的检测率

隐写方法	隐写分析器	0.1 s	0.3 s	0.5 s	0.7 s	1 s
QIMC	FCEM	1	0.972	1	0.997	1
	SRCNet	0.024	0.354	0	0	0.009
	G-LSTM	0.588	0.547	0.57	0.628	0.595
	SpecResNet	0.456	0.574	0.606	0.572	0.623
	MSFNet	**0.948**	**0.948**	**0.959**	**0.988**	**0.959**
FCB	FCEM	0.153	0.17	0.002	0.021	0.051
	SRCNet	0.995	0.942	1	1	0.999
	G-LSTM	0.278	0.505	0.004	0.001	0.008
	SpecResNet	0.488	0.568	0.636	0.565	0.654
	MSFNet	**0.992**	**1**	**1**	**1**	**0.997**
GQS	FCEM	0.153	0.17	0.002	0.021	0.051
	SRCNet	0.033	0.376	0.002	0	0.003
	G-LSTM	0.814	0.826	0.817	0.914	0.944
	SpecResNet	0.48	0.584	0.617	0.567	0.633
	MSFNet	**0.845**	**0.938**	**0.995**	**0.995**	**0.998**
HS	FCEM	0.153	0.17	0.002	0.021	0.051
	SRCNet	0.996	0.938	1	1	1
	G-LSTM	0.618	0.513	0.79	0.83	0.869
	SpecResNet	0.526	0.577	0.607	0.586	0.677
	MSFNet	**0.996**	**1**	**1**	**1**	**1**

可以看出,FCEM隐写分析器在QIMC隐写的时长为0.1 s样本上,检测准确率就可以达到100%,而且随着时长的增加一直保持较高的准确率,但在其余隐写方法隐写的样本上,隐写检测率最高只能达到51%,且随着时长的增长变化很小. SRCNet隐写分析器虽然在FCB和HS方法隐写的不同时长样本上检测率都在93%以上,但是在其余隐写方法上检测率非常低,且随着时长的增加变化也很小,SRCNet对HS方法满嵌时有较高检测率,这是因为HS方法在满嵌时会影响码本搜索阶段的矢量索引. G-LSTM隐写分析器在GQS隐写样本上随着时长的增加准确率由81.4%逐渐升至94.4%,在HS隐写的样本上随着时长的增长也大致呈现出上升的趋势,而在其余样本上却未能随着时长呈现出更高的检测率. SpecResNet的检测率随着时长的增长缓慢增加,但在各类隐写样本上检测率均不高于70%. 本文MFSNet在各类隐写样本上几乎都能达到90%以上的检测准确率率,且大致呈现出随着时长的增长检测率逐渐上升的趋势.

4.3　短时语音隐写检测的有效性

在网络通信中,对短时语音的隐写检测是非常重要的. 为了体现出所提出MFSNet的有效性,在0.1 s的语音中进行测试,表3至表6分别给出了不同隐写分析算法在不同帧长和不同嵌入率下的检测结果.

表3　在0.1 s长、30 ms帧的中文语音样本上的检测率

隐写分析器	隐写方法	嵌入率					
		0.1	0.2	0.4	0.6	0.8	1
FCEM	QIMC	0.998	1	1	1	1	1
	FCB	0.153	0.153	0.153	0.153	0.153	0.153
	GQS	0.153	0.153	0.153	0.153	0.153	0.153
	HS	0.153	0.153	0.153	0.153	0.153	0.153
	平均值	**0.364**	**0.365**	**0.365**	**0.365**	**0.365**	**0.365**
SRCNet	QIMC	0.034	0.029	0.024	0.036	0.032	0.024
	FCB	0.932	0.958	0.993	0.992	0.994	0.995
	GQS	0.03	0.03	0.031	0.032	0.027	0.033
	HS	0.023	0.027	0.027	0.027	0.052	0.996
	平均值	**0.255**	**0.261**	**0.269**	**0.272**	**0.276**	**0.512**
G-LSTM	QIMC	0.553	0.54	0.542	0.589	0.577	0.588
	FCB	0.167	0.219	0.237	0.231	0.264	0.278
	GQS	0.5	0.541	0.541	0.771	0.802	0.814
	HS	0.5	0.561	0.669	0.505	0.457	0.618
	平均值	**0.43**	**0.465**	**0.497**	**0.524**	**0.525**	**0.575**
SpecResNet	QIMC	0.463	0.451	0.471	0.439	0.477	0.456
	FCB	0.469	0.483	0.482	0.482	0.472	0.488
	GQS	0.497	0.482	0.491	0.495	0.489	0.48
	HS	0.532	0.517	0.535	0.532	0.506	0.526
	平均值	**0.424**	**0.435**	**0.431**	**0.453**	**0.487**	**0.499**
MSFNet	QIMC	0.84	0.955	0.971	0.977	0.962	0.948
	FCB	0.966	0.969	0.987	0.993	0.997	0.992
	GQS	0.646	0.637	0.658	0.804	0.829	0.845
	HS	0.615	0.59	0.647	0.748	0.757	0.996
	平均值	**0.767**	**0.788**	**0.816**	**0.881**	**0.886**	**0.945**

表4 在0.1 s长、20 ms帧的中文语音样本上的检测率

隐写分析器	隐写方法	嵌入率					
		0.1	0.2	0.4	0.6	0.8	1
FCEM	QIMC	0.959	0.962	0.978	0.998	0.999	1
	FCB	0.013	0.013	0.013	0.013	0.013	0.013
	GQS	0.013	0.013	0.013	0.013	0.013	0.013
	HS	0.013	0.013	0.013	0.013	0.013	0.013
	平均值	**0.25**	**0.25**	**0.254**	**0.259**	**0.26**	**0.26**
SRCNet	QIMC	0.002	0.003	0.001	0.003	0	0.003
	FCB	0.987	0.991	1	1	1	0.999
	GQS	0.005	0.008	0.004	0.007	0.004	0.009
	HS	0.005	0.01	0.008	0.016	0.41	0.988
	平均值	**0.25**	**0.253**	**0.203**	**0.257**	**0.354**	**0.502**
G-LSTM	QIMC	0.44	0.438	0.442	0.472	0.502	0.462
	FCB	0.159	0.078	0.109	0.117	0.133	0.122
	GQS	0.416	0.564	0.743	0.854	0.844	0.811
	HS	0.51	0.559	0.7	0.682	0.627	0.719
	平均值	**0.381**	**0.41**	**0.499**	**0.518**	**0.527**	**0.529**
SpecResNet	QIMC	0.528	0.524	0.531	0.549	0.528	0.531
	FCB	0.495	0.525	0.534	0.587	0.617	0.556
	GQS	0.487	0.499	0.492	0.52	0.498	0.51
	HS	0.485	0.487	0.501	0.518	0.552	0.523
	平均值	**0.499**	**0.509**	**0.515**	**0.544**	**0.549**	**0.53**
MSFNet	QIMC	0.862	0.925	0.948	0.964	0.961	0.967
	FCB	0.996	0.996	0.999	0.999	1	1
	GQS	0.637	0.712	0.794	0.882	0.876	0.872
	HS	0.69	0.719	0.837	0.796	0.909	1
	平均值	**0.796**	**0.838**	**0.895**	**0.91**	**0.937**	**0.96**

可以看出,FCEM对QIMC隐写的样本检测率都在90%以上,且随着嵌入率的增大而增高,但对于其它方法隐写的样本都低于20%;SRCNet对FCB隐写的样本检测率在93%以上,随着嵌入率的增大逐渐增大,对HS隐写的样本只有当嵌入率达到100%时才可以达到90%以上,有较高的检测率;G-LSTM对GQS和HS隐写的样本检测率随着嵌入率的增大大致呈现出逐渐增大的趋势,而对其他隐写方法的样本检测率较低,且不随嵌入率的增大而增大. SpecResNet对0.1 s的音频样本的检测率在60%以下,且随着嵌入率的增大检测率有的增大并不明显. 因此,这些方法无法都无法覆盖所有的隐写样本.

本文MFSNet对各隐写方法均具有较好的检测结果. 具体来说,MFSNet随着隐写率的增加,对FCB、QIMC、GQS、HS的检测率逐渐提升. 由于GQS隐写对

表5　在0.1 s长、30 ms帧的英文语音样本上的检测率

隐写分析器	隐写方法	嵌入率					
		0.1	0.2	0.4	0.6	0.8	1
FCEM	QIMC	0.908	0.989	0.998	0.999	0.997	1
	FCB	0.003	0.003	0.003	0.003	0.003	0.003
	GQS	0.003	0.003	0.003	0.003	0.003	0.003
	HS	0.003	0.003	0.003	0.003	0.003	0.003
	平均值	**0.227**	**0.247**	**0.25**	**0.25**	**0.249**	**0.25**
SRCNet	QIMC	0.008	0.013	0.019	0.018	0.014	0.017
	FCB	0.959	0.959	0.996	1	0.997	0.999
	GQS	0.019	0.016	0.021	0.018	0.019	0.018
	HS	0.014	0.015	0.018	0.014	0.029	0.996
	平均值	**0.25**	**0.251**	**0.264**	**0.275**	**0.265**	**0.508**
G-LSTM	QIMC	0.3	0.335	0.371	0.356	0.37	0.417
	FCB	0.46	0.61	0.7	0.59	0.68	0.84
	GQS	0.449	0.676	0.889	0.897	0.95	0.886
	HS	0.295	0.588	0.434	0.416	0.509	0.426
	平均值	**0.376**	**0.62**	**0.599**	**0.565**	**0.627**	**0.642**
SpecResNet	QIMC	0.45	0.509	0.489	0.492	0.546	0.512
	FCB	0.418	0.43	0.429	0.47	0.482	0.527
	GQS	0.417	0.393	0.407	0.402	0.425	0.444
	HS	0.412	0.406	0.4	0.448	0.495	0.512
	平均值	**0.424**	**0.435**	**0.431**	**0.453**	**0.487**	**0.499**
MSFNet	QIMC	0.712	0.963	0.98	0.983	0.971	0.951
	FCB	0.946	0.945	0.988	0.988	0.994	0.992
	GQS	0.522	0.726	0.887	0.905	0.907	0.871
	HS	0.396	0.55	0.56	0.583	0.671	0.99
	平均值	**0.647**	**0.796**	**0.854**	**0.865**	**0.886**	**0.951**

增益量化索引之间的相关性影响最弱,HS是在增益量化和矢量量化的层进行自适应隐写,而且在层内未满时运用混沌序列挑选隐写位置,进一步对模型的训练加大了难度,导致检测准确率较低,但仍高于FCEM、SRCNet、G-LSTM和SpecResNet.

值得注意的是,由于FCEM隐写分析器是针对LSF系数量化阶段的隐写提出的隐写分析器,对FCB、GQS和HS隐写的样本检测率都保持同一个值,这是因为这三种隐写方法的隐写位置都位于LSF系数量化之后,无论如何隐写都不会影响LSF系数量化索引值,所以检测率保持一致且不随嵌入率增大发生变化. 此外,当0.1 s音频在嵌入率达到100%时,本文MFSNet隐写检测方法在30 ms帧的第三帧、20 ms帧的第五帧即可检测出是否存在隐写,而且每帧的检测时间不高于0.1 ms. 因此,MFSNet算法具有较好的实时性,可以满足在线语音隐写检测的需求.

表6 在0.1 s长、20 ms帧的英文语音样本上的检测率

隐写分析器	隐写方法	嵌入率					
		0.1	0.2	0.4	0.6	0.8	1
FCEM	QIMC	0.982	0.991	0.996	0.998	1	1
	FCB	0.005	0.005	0.005	0.005	0.005	0.005
	GQS	0.005	0.005	0.005	0.005	0.005	0.005
	HS	0.005	0.005	0.005	0.005	0.005	0.005
	平均值	**0.249**	**0.252**	**0.287**	**0.253**	**0.254**	**0.254**
SRCNet	QIMC	0.001	0	0	0.003	0.002	0.006
	FCB	0.99	0.989	0.999	0.998	0.999	1
	GQS	0.004	0.003	0.005	0.003	0.004	0.004
	HS	0.005	0.004	0.007	0.009	0.49	0.997
	平均值	**0.25**	**0.249**	**0.253**	**0.253**	**0.374**	**0.502**
G-LSTM	QIMC	0.408	0.468	0.489	0.531	0.524	0.518
	FCB	0.135	0.122	0.103	0.087	0.099	0.099
	GQS	0.446	0.691	0.745	0.809	0.91	0.709
	HS	0.569	0.541	0.858	0.793	0.635	0.721
	平均值	**0.39**	**0.308**	**0.549**	**0.555**	**0.542**	**0.512**
SpecResNet	QIMC	0.452	0.422	0.51	0.502	0.346	0.356
	FCB	0.396	0.41	0.44	0.452	0.449	0.433
	GQS	0.392	0.388	0.4	0.376	0.375	0.381
	HS	0.386	0.38	0.388	0.389	0.399	0.408
	平均值	**0.407**	**0.4**	**0.435**	**0.43**	**0.392**	**0.395**
MSFNet	QIMC	0.781	0.956	0.962	0.982	0.991	0.996
	FCB	0.978	0.98	0.996	0.998	0.997	0.999
	GQS	0.571	0.706	0.763	0.82	0.889	0.769
	HS	0.6	0.603	0.866	0.8	0.863	1
	平均值	**0.735**	**0.811**	**0.897**	**0.9**	**0.935**	**0.941**

5 结论

针对已有iLBC语音隐写检测方法对多阶段隐写难以达到理想检测结果的问题,本文提出了一种基于多特征融合和BiLSTM的iLBC语音隐写检测算法,通过分析不同阶段隐写对语音的影响,提取各隐写域的特征,设计双向长短时记忆网络训练多个检测模型,并将各模型结果进行融合. 对比实验结果表明,本文提出的MFSNet算法不仅对多阶段隐写具有较好的检测结果,而且对短时语音也能实现有效的检测. 由于MFSNet隐写分析器是基于目前常用的QIM隐写方法设计的多阶段隐写检测算法,对于基于扩频、patchwork等技术的隐写方法还存在检测率不高的问题. 因此,未来的工作重点是设计更为通用的iLBC语音隐写检测方法.

参考文献

[1] WU Z, SHA Y. An implementation of speech steganography for iLBC by using fixed codebook[C]//IEEE International Conference on Computer and Communications. Chengdu: IEEE Press, 2016: 1970-1974.

[2] HUANG Y, TAO H, XIAO B, et al. Steganography in low bit-rate speech streams based on quantization index modulation controlled by keys[J]. Science China Technological

Sciences, 2017, 60(10): 1585-1596.

[3] SU Z, LI W, ZHANG G, et al. A steganographic method based on gain quantization for iLBC speech streams[J]. Multimedia Systems, 2020, 26(2): 223-233.

[4] 苏兆品, 张羚, 张国富. 低比特率语音流大容量分层隐写方法[J/OL]. 中国图象图形学报, 2022, DOI: 10.11834/jig.210307.
SU Zhao-pin, ZHANG Ling, ZHANG Guo-fu. High-capacity hierarchical steganography in a low-bit rate speech codec[J/OL]. Journal of Image and Graphics, 2022, DOI: 10.11834/jig.210307. (in Chinese)

[5] LIU Q, SUNG A H, QIAO M .Temporal derivative-based spectrum and mel-cepstrum audio steganalysis[J]. IEEE Transactions on Information Forensics and Security, 2009, 4(3): 359-368.

[6] LIN Z, HUANG Y, WANG J. RNN-SM: Fast steganalysis of VoIP streams using recurrent neural network[J]. IEEE Transactions on Information Forensics and Security, 2018, 13(7): 1854-1868.

[7] GONG C, YI X, ZHAO X, et al. Recurrent convolutional neural networks for AMR steganalysis based on pulse position[C]//ACM Workshop on Information Hiding and Multimedia Security, Paris: ACM Press, 2019: 2-13.

[8] REN Y, LIU D, LIU C, et al. A universal audio steganalysis scheme based on multiscale spectrograms and deep-ResNet[J/OL]. IEEE Transactions on Dependable and Secure Computing, 2022. DOI: 10.1109/TDSC.2022.3141121.

[9] YANG H, YANG Z, BAO Y, et al. FCEM: A novel fast correlation extract model for real time steganalysis of VoIP stream via multi-head attention[C]//International Conference on Acoustics, Speech and Signal Processing, Barcelona: IEEE, 2020: 2822-2826.

[10] YANG H, YANG Z, BAO Y, et al. Fast steganalysis method for VoIP streams[J]. IEEE Signal Processing Letters, 2020, 14: 286-290.

[11] 李望望. 面向iLBC语音流的隐写与隐写分析技术研究[D]. 合肥: 合肥工业大学计算机与信息学院, 2019.

[12] 张浩, 胡昌华, 杜党波等. 多状态影响下基于Bi-LSTM网络的锂电池剩余寿命预测方法[J]. 电子学报, 2022, 50(3): 619-624.
ZHANG H,HU C,DU D,et al. Remaining useful life prediction method of lithium-ion battery based on Bi-LSTM network under multi-state influence[J]. Acta Electronica Sinica, 2022, 50(3): 619-624. (in Chinese)

[13] 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 J, HU R, RUAN G, 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. (in Chinese)

作者简介

苏兆品　女，1983年8月出生于山东省菏泽市. 现为合肥工业大学计算机与信息学院副教授、硕士生导师. 获安徽省自然科学奖1项. 在国内外发表学术论文40余篇. 中国电子学会会员编号:E190027825M.
E-mail: szp@hfut.edu.cn

张　羚　女，1995年4月出生于甘肃省武威市. 硕士研究生，主要研究方向为音频隐写和隐写分析.
E-mail: 1772950753@qq.com

张国富(通讯作者)　男，1979年3月出生于安徽省合肥市. 现为合肥工业大学计算机与信息学院教授、硕士生导师. 主要研究方向为联盟博弈、进化计算、音频安全.
E-mail: zgf@hfut.edu.cn

岳　峰　男，1981年2月出生于安徽省合肥市. 现为合肥工业大学计算机与信息学院副研究员、硕士生导师. 主要研究方向为软件工程、信息安全.
E-mail: yuefeng@hfut.edu.cn