

基于倒谱图像的语音回声隐写分析方法

唐军豪, 杜庆治, 龙 华, 邵玉斌, 李一民

(昆明理工大学信息工程与自动化学院, 昆明 650500)

摘 要: 语音信号回声隐写后其倒谱系数会在回声延迟处产生峰值, 传统回声隐写分析主要采用倒谱系数的统计特征作为隐写检测特征, 然而在低回声幅度时隐写信号倒谱系数的峰值并不明显, 基于统计特征的方法检测性能并不理想。本文将倒谱分析与图像识别技术结合, 提出了一种基于倒谱图像的语音回声隐写分析方法, 对语音信号分帧加窗后进行倒谱计算, 然后以时间为横轴, 倒谱序列点为纵轴, 倒谱系数幅值为灰度级生成倒谱图像, 将生成的倒谱图像作为隐写检测的输入, 采用残差神经网络作为分类器进行回声隐写分析。实验结果表明, 在3种经典回声隐写算法上低回声幅度时检测准确率分别达到98.2%、98.6%和96.1%, 本文方法在低回声幅度时检测准确率相较传统回声隐写分析方法有较大提升, 解决了传统回声隐写分析方法在低回声幅度检测效果不佳的问题。

关键词: 回声隐写; 语音隐写分析; 倒谱; 图像识别; 残差神经网络

中图分类号: TP391 **文献标志码:** A

Speech Steganalysis Method for Echo Hiding Based on Image of Cepstrum

TANG Junhao, DU Qingzhi, LONG Hua, SHAO Yubin, LI Yimin

(Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China)

Abstract: After echo hiding, the cepstrum coefficient of a speech signal will peak at the echo delay. The traditional echo hiding steganalysis mainly uses the statistical characteristics of the cepstrum coefficient as the steganalysis feature. However, the peak value of the cepstrum coefficient of the steganography signal is not obvious when the echo amplitude is low, and the detection performance of the method based on the statistical characteristics is unsatisfactory. This paper combines cepstrum analysis with image recognition technology, and proposes an steganalysis method for speech echo hiding based on cepstrum image. The speech signal is divided into frames and windowed for cepstrum calculation. Then, the image is generated with time as the horizontal axis, cepstrum sequence points as the vertical axis, and cepstrum coefficient amplitude as the gray level. The generated cepstrum image is used as the steganalysis input, and residual neural network is used as the classifier for echo hiding steganalysis. The experimental results show that the detection accuracy of the three classical echo hiding algorithms reaches 98.2%, 98.6% and 96.1% respectively at low echo amplitude. The detection accuracy of this method at low echo amplitude is greatly improved compared with the traditional echo hiding steganalysis method, which solves the problem that the traditional echo hiding steganalysis method has unsatisfactory detection effect at low echo amplitude.

Key words: echo hiding; speech steganalysis; cepstrum; image recognition; residual neural network

引言

随着数字媒体和网络技术的飞速发展,大量多媒体数据充斥着人们的生活,这也促进了隐写术的发展。语音隐写技术是一种将秘密信息隐藏在语音载体中的方法,语音隐写分析技术则是一种判断语音文件中是否含有秘密信息的技术。语音回声隐写技术是一种非压缩域的语音隐写技术,通过在载体语音中引入有规律的回声来隐藏秘密信息。Bender等^[1]首次提出通过引入不同延迟的单回声进行秘密信息的隐藏,之后Oh等^[2]提出了双极性回声隐藏算法, Kim等^[3]提出了双向回声隐藏算法。回声隐写具有良好的不可感知性,得到了研究人员的广泛关注,文献[4-9]均是回声隐写技术与其他信息处理技术结合,提出的一系列隐写算法。

回声隐写技术的广泛应用也促进着回声隐写分析技术的发展,针对回声隐写进行检测的方法分为通用语音隐写分析和专用回声隐写分析,在通用语音隐写分析上,Geetha等^[10]提出将语音的豪斯多夫距离(Hausdorff distance)高阶统计量作为隐写分析特征,利用决策树作为分类器进行语音隐写分析,在回声隐写算法上的检测准确率达到90%。Kocal等^[11]利用DVV(Delay vector variance)特征对录音语音进行隐写分析,但在回声隐写算法上的检测效果并不理想,在隐写信噪比为40 dB时,检测准确率只有78%。Luo等^[12]设计了一种基于时域和梅尔频域改进的语音隐写分析特征集来对一些典型的隐写进行检测,在回声隐写算法上检测准确率达到82%。上述通用语音隐写分析方法并不针对回声隐写,检测性能并不理想。在专用回声隐写分析上,杨榆等^[13]提出了一种基于回声隐藏的VDS(Variants of difference of sum of cepstrum)隐写分析算法,通过构造倒谱和差分方差统计量VDS作为隐写分析的特征,实验结果显示该算法在单回声核回声隐写上的检测准确率为88%。王昱洁等^[14]提出了一种基于功率倒谱特征的隐写分析方法,以平均功率倒谱的差分方差和平坦度作为统计特征,该隐写分析方法在3类经典回声隐写算法上检测准确率相较之前的研究有一定提升,在回声幅度较低时检测准确率仍有待提升。Qi等^[15]提出了将倒谱系数、梅尔倒谱系数和直方图等特征作为隐写分析特征。Wang等^[16]提出将语音功率倒谱的偏度和峰度作为隐写分析特征,实验表明该方法在高回声幅度时检测准确率较高,但在低回声幅度时检测准确率只有84.4%。王杰等^[17]提出了一种基于卷积神经网络的回声隐写分析方法,提取语音的幅度谱系数矩阵作为隐写分析的浅层特征。邓伟^[18]提出将语音功率倒谱的峰度、偏度、平坦度、1~8阶差分的方差作为回声隐写分析的特征,在低回声幅度时检测准确率在90%左右。

上述语音回声隐写分析方法有一定效果,但这些方法一般基于语音的倒谱或功率倒谱的统计特征,然而倒谱和功率倒谱系数本身就有一些起伏,在回声幅度较低时回声隐写信号倒谱和功率倒谱中出现的峰值很低,基于统计特征作的方法检测性能并不理想。针对上述问题,本文提出了一种基于倒谱图像的回声隐写分析方法,根据回声隐写信号在倒谱中出现峰值的位置相对固定这一规律,利用倒谱系数构造倒谱图像,将倒谱图像作为隐写检测的输入,不同时间帧回声隐写信号倒谱出现的峰值在倒谱图像中构成一条不连续的线状结构,利用残差神经网络作为分类器进行回声隐写的检测,在传统回声隐写、双极性回声隐写和双向回声隐写上进行实验,实验结果表明,本文提出的方法初步解决了低回声幅度下回声隐写检测准确率不高的问题。

1 回声隐写算法

基于回声的隐写技术利用了人耳听觉系统对语音的时域掩蔽特性,通过引入回声进行信息隐藏,假设秘密信息是由0和1构成的比特序列,文献[1]提出用满足掩蔽效应的时延 d_0 和 d_1 分别表示比特0和比特1,将载体语音分成长度相等的若干语音帧,每个语音帧通过引入一个时延的回声来隐藏1 bit的秘密信息,若隐藏的信息为“0”,则时延为 d_0 ,若隐藏的信息为“1”,则时延为 d_1 ,最后将每个语音帧串联组成新的语音信号,实现回声隐写。在语音中引入单个回声的数学模型为

$$y(n) = s(n) + \alpha s(n - d) \tag{1}$$

式中: $y(n)$ 为引入回声后的语音信号; $s(n)$ 为原始语音信号; α 为回声的幅值; d 为回声的延迟。式(1)可以表示为卷积的形式

$$y(n) = s(n) * h(n) \tag{2}$$

$$h(n) = \delta(n) + \alpha \delta(n - d) \tag{3}$$

式中 $h(n)$ 称为回声核,式(3)也被称为传统回声核(记作K1),传统回声隐写如图1所示。

文献[2]提出引入不同极性和不同时延的回声进行回声隐写,称为双极性回声隐写,这种方式能够使回声能量更高,提高了隐写算法的鲁棒性,其回声核形式为

$$h(n) = \delta(n) + \alpha_1 \delta(n - d_1) - \alpha_2 \delta(n - d_2) \tag{4}$$

式(4)被称为双极性回声核(记作K2),双极性回声核如图2所示。

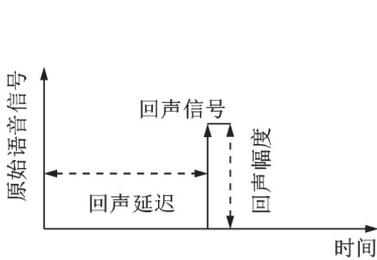


图1 传统回声核示意图

Fig.1 Schematic diagram of echo hiding with traditional kernel

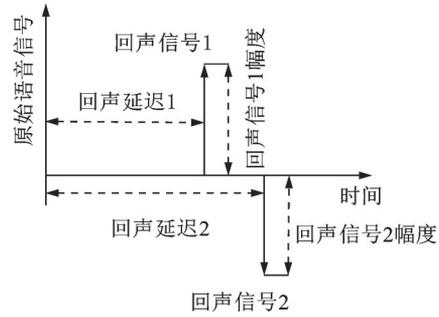


图2 双极性回声核示意图

Fig.2 Schematic diagram of echo hiding with positive and negative kernels

文献[3]利用人耳听觉的前掩蔽效应,提出了一种双向回声隐写,其回声核表达式为

$$h(n) = \delta(n) + \alpha \delta(n - d) + \alpha \delta(n + d) \tag{5}$$

式中: $\delta(n + d)$ 称为前向回声核; $\delta(n - d)$ 称为后向回声核。式(5)被称为双向回声核(记作K3)。通过引入前向回声,双向回声隐写后倒谱的峰值更加明显,在利用双向回声核进行回声隐写时可以降低回声的幅值,双向回声核如图3所示。

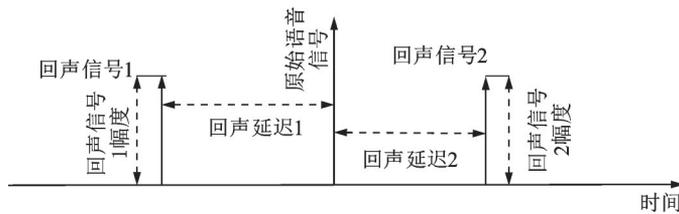


图3 双向回声核示意图

Fig.3 Schematic diagram of echo hiding with backward and forward kernels

2 回声隐写分析方法

2.1 倒谱分析

引入回声后语音信号是原始语音信号与回声核的卷积,从时域或频域分析语音引入回声后的变化比较困难,因此利用卷积同态滤波技术进行处理,将卷积信号变成加性信号^[19],可以采用传统的复倒谱

分析进行回声的检测,但要克服复数计算的问题,一般采用实倒谱分析(本文简称为倒谱),倒谱的定义为

$$C(k) = \text{IFFT}[\log|\text{FFT}(x(n))|] \quad (6)$$

当语音引入回声时,即将式(2)代入式(6),此时隐写语音的倒谱为

$$\begin{aligned} C_e(n) &= \text{IFFT}[\ln|\text{FFT}(s(n)*h(n))|] = C(n) + \text{IFFT}[\ln|\text{FFT}(h(n))|] = \\ &= C(n) + \text{IFFT}\left[\ln\left(\frac{H(j\omega) + H^*(j\omega)}{2}\right)\right] = C(n) + \text{IFFT}[\ln(1 + \alpha \cos \omega d)] = \\ &= C(n) + \text{IFFT}\left(\alpha \cos \omega d - \frac{\alpha^2 \cos^2 \omega d}{2} + Q\right) \approx C(n) + \\ &= \text{IFFT}\left[\alpha(e^{j\omega d} + e^{-j\omega d}) - \frac{\alpha^2}{4}(e^{j2\omega d} + e^{-j2\omega d}) - \frac{\alpha^2}{4}\right] \approx C(n) + \\ &= \left\{-\frac{\alpha^2}{4}\delta(n) + \alpha[\delta(n+d) + \delta(n-d)] - \frac{\alpha^2}{4}[\delta(n+2d) + \delta(n-2d)]\right\} \end{aligned} \quad (7)$$

由式(7)可知,语音引入回声后倒谱在回声延迟 d 的倍数处产生峰值,其中在回声延迟 d 处的峰值最大,由于引入回声的幅度 $\alpha < 1$,在回声延迟的其他倍数 d 处出现的峰值较小。图4给出了一段语音引入延迟为64、幅度为0.5的回声后倒谱出现的变化,由于倒谱计算具有对称性,图中仅给出倒谱的前半段。

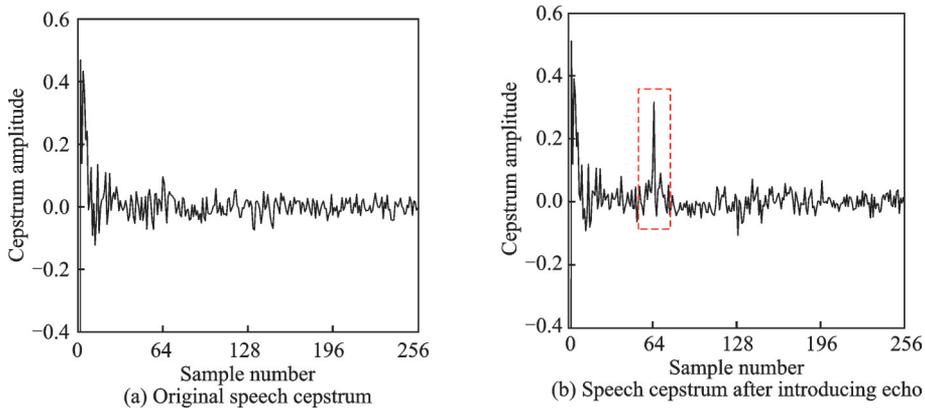


图4 语音引入回声后倒谱对比

Fig.4 Comparison of speech cepstrum after echo introduction

2.2 回声隐写分析框架

语音进行回声隐写后,其倒谱系数在回声延迟处出现峰值,研究者一般将倒谱或功率倒谱的多阶差分的方差、平坦度、峰度和偏度等标量作为隐写分析特征,由于低回声幅度回声隐写时语音倒谱或功率倒谱出现的峰值较低,上述这些特征作为隐写检测特征并不稳定。针对上述问题,本文将语音信号处理与图像识别技术结合,提出了一种基于倒谱图像的回声隐写分析方法。首先对语音进行分帧加窗,计算所有时间帧的倒谱,再生成以时间为横轴,倒谱样点序列为纵轴,倒谱系数幅值为灰度级的倒谱图像,将倒谱图像作为回声隐写检测模型的输入,采用残差神经网络作为分类器进行回声隐写分析,本文算法流程图如图5所示。

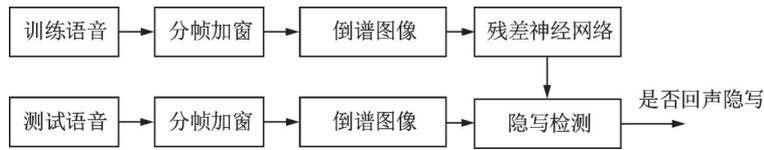


图5 回声隐写分析算法流程图

Fig.5 Schematic diagram of echo hiding steganalysis method

2.3 检测分帧

回声隐写时将语音进行分帧,在每个语音帧引入不同延迟的回声来隐藏秘密信息。在进行隐写检测时,并不清楚隐写分帧的长度,在生成倒谱图像时,需要先对语音数据进行分帧,为了保证回声隐藏的不可感知性,回声隐写帧长一般大于256个采样点^[14],在进行隐写检测时将分帧长度设置为256,此时检测语音帧可能位于隐写语音帧内,也可能位于隐写语音帧之间,如图6所示。



图6 检测分帧长度与隐写分帧长度示意图

Fig.6 Schematic diagram of detection and steganography framing length

以单回声核为例,假设代表秘密信息“0”和秘密信息“1”的回声延迟分别为32个采样点和64个采样点,回声幅度为0.5。若检测语音帧位于隐写语音帧内(即图6中检测语音帧1和3),此时检测语音帧只包含一种回声延迟的回声,由2.1节可知,检测语音帧倒谱在一个回声延迟处出现峰值。倒谱计算具有对称性,图7展示了检测语音帧1~4的倒谱前半段。检测语音帧1和3的倒谱如图7(a,c)所示,为方便观察,将倒谱前8个系数置零;若检测语音帧位于隐写语音帧之间,相邻隐写语音帧隐藏的秘密信息不同(即检测语音帧2),那么检测语音帧倒谱在两个回声延迟都可以产生峰值,峰值可能出现在两个回声延迟处中的一个位置上,也可能都出现,图7(b)展示了检测语音帧2的倒谱,此时倒谱在两个回声延迟处均出现了峰值;若检测语音帧位于隐写语音帧之间,相邻隐写语音帧隐藏的秘密信息相同时(即检测语音帧4),此时检测语音帧包含一种回声延迟的回声,故检测语音帧倒谱只在一个回声延迟处出现峰值,如图7(d)所示。

2.4 倒谱图像生成

由2.3节分析可知,不论回声隐写分析的检测帧位于隐写信号的任何位置,回声隐写后语音倒谱系数都会在回声延迟处产生峰值,而回声隐写引入的延迟是固定的,不同时间帧倒谱的峰值在同一回声延迟处可以构成一条不连续的线状结构,根据这一规律本文提出将倒谱构造为二维图像的形式,以宏观的形式分析回声隐写。倒谱图像生成包括5个步骤:

步骤1 对语音信号 $x(n)$ 进行分帧和加窗,帧长为256个采样点,帧移为0,窗函数选用汉明窗,处理后得到第 i 帧语音数据为 $x_i(n)$,其中 $1 \leq i \leq F$, F 表示帧数。

步骤2 对第 i 帧语音信号 $x_i(n)$ 进行离散傅里叶变换,有

$$S_i(k) = \sum_{n=0}^{N-1} x_i(n) e^{-j\frac{2\pi}{N}kn} \quad (8)$$

式中 N 为傅里叶变换的点数。

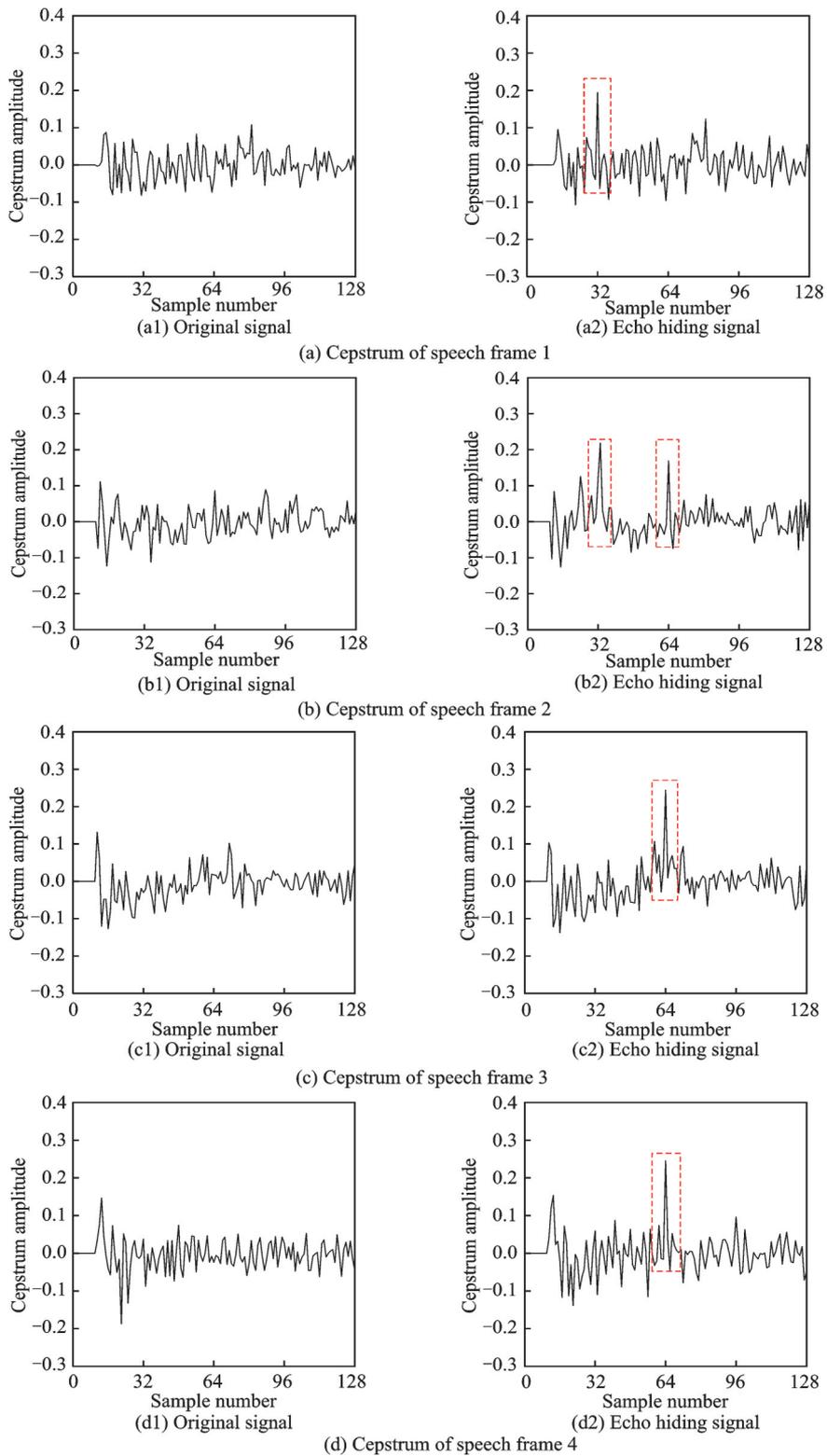


图7 不同语音帧回声隐写前后倒谱对比

Fig.7 Speech cepstrum comparison after echo hiding for different speech frames

步骤3 对 $S_i(k)$ 取绝对值, 并取对数, 有

$$P_i(k) = \ln|S_i(k)| \tag{9}$$

步骤4 对 $P_i(k)$ 进行逆离散傅里叶变换, 得到每一帧语音信号的倒谱, 有

$$C_i(m) = \frac{1}{N} \sum_{k=0}^{N-1} P_i(k) e^{j\frac{2\pi}{N}km} \quad m=0, 1, \dots, N-1 \tag{10}$$

舍弃与语音自身相关的首尾各8个倒谱系数值, 由于倒谱计算具有对称性, 只需要倒谱的前半段即可。

步骤5 以时间 n 为横轴, 倒谱采样点 m 为纵轴, 倒谱值 C 表示灰度级构成二维图像矩阵 $I = [C/\max(C)] \times 255$, 进行3类经典回声隐写后, 原始语音倒谱图像和隐写后语音倒谱图像如图8所示, 其中秘密信息为随机二进制序列, 语音长度为3 s。

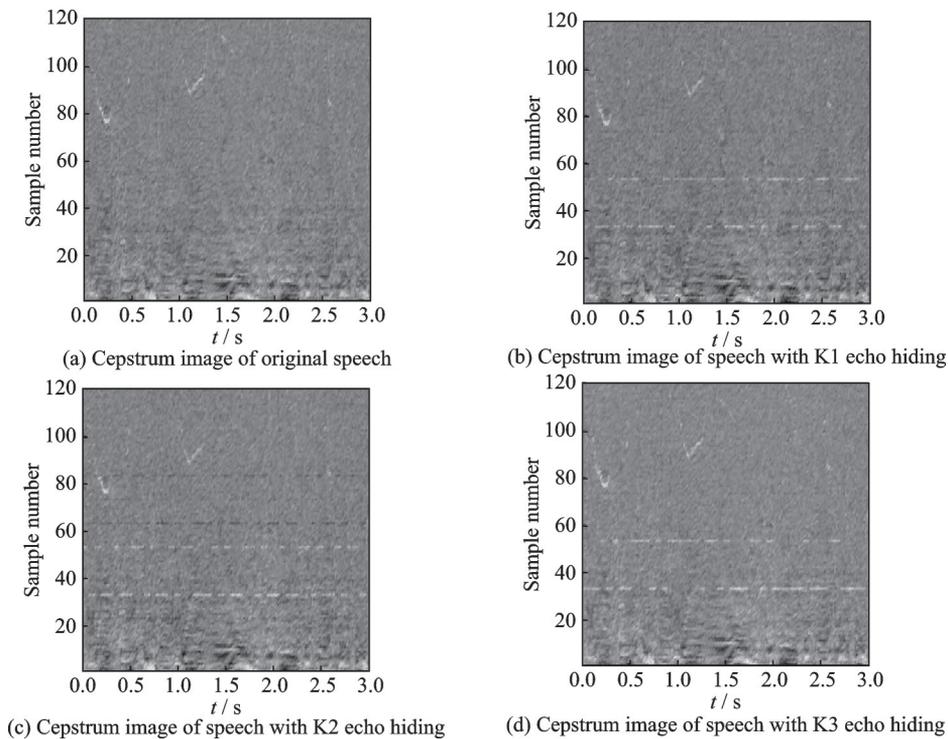


图8 不同类型回声隐写前后倒谱图像

Fig.8 Cepstrum image of speech with different echo hiding

由图8可以看出, 在语音进行回声隐写后, 语音的倒谱图像出现了由不同时间帧倒谱在回声延迟处的峰值构成的线状结构, 利用这一明显的视觉特征, 去除图像中坐标等干扰信息, 采用图像识别的方法对倒谱图像进行识别, 从而实现回声隐写的检测。

2.5 残差神经网络

本文将倒谱图像去除坐标信息后作为回声隐写分析的检测特征, 采用残差神经网络^[20]进行检测分类。残差神经网络是一种强大又简单的CNN架构, 其内部的残差块使用了跳跃连接的结构, 如图9所示。图中 $F(x)$ 表示残差块中卷积层的非线性映射, x 表示残差块

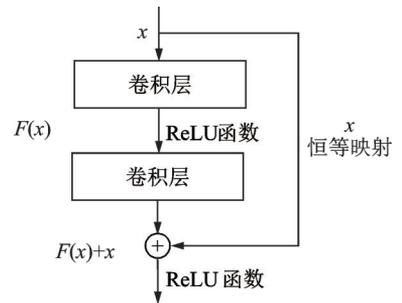


图9 残差块结构图

Fig.9 Basic structural unit of the residual neural network model

的输入,ReLU函数的表达式为

$$\sigma(x)=\begin{cases} 0 & x < 0 \\ x & x \geq 0 \end{cases} \quad (11)$$

残差块的输出为

$$H(x)=\sigma(F(x)+x) \quad (12)$$

残差块的结构使得原来2层卷积层的输出 $F(x)$ 变成了 $F(x)+x$,通过跳跃连接的结构,反向传播时信号可以无衰减地传递,缓解了深层网络中梯度弥散和精度下降的问题,能够达到更深的网络深度,在回声幅度较低时也能对倒谱图像进行比较准确的判别。本文使用34层的残差神经网络(Resnet34)作为分类器,Resnet34的网络结构如表1所示,包括1个常规卷积层(conv1),1个池化层(max pool),16个残差块(conv2_x、conv3_x、conv4_x、conv5_x),1个平均池化层(Average pool)和1个全连接层(Fully connected layer,FC)。

表1 残差神经网络结构

网络层名称	输出尺寸	网络层参数
conv1	112×112	$7 \times 7, \text{stride} = 2$
max pool	56×56	$3 \times 3, \text{stride} = 2$
conv2_x	56×56	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$
Average pool	1×1	$1 \times 1, \text{stride} = 2$
FC	1×1	1 000

3 实验结果分析

3.1 实验数据

为测试本文提出的回声隐写分析算法的检测性能,本文在VoxForge^[21]公共语料库上进行实验,VoxForge是一个收集多个语种的开源语音库,包含了French、German、English、Spanish、Italian及Russian等多个语种不同说话人的语音录音。本文选取French、German、English、Spanish和Italian作为所用数据,每条语音均采用16 kHz采样,16 bit量化,PCM编码,语音长度3 s。每个语种选取1 000条语音(利用伪随机数确定语音序号选择语音),共5 000条语音进行回声隐写分析实验,隐写参数如表2所示。对选取的5 000条实验样本中的2 500条分别进行不同回声核不同回声幅度的回声隐写,另外2 500条不进行隐写,共得到9个回声隐写数据集,每个回声隐写数据集的分布均如表3所示。

表2 回声隐写参数设置

Table 2 Parameter settings of echo hiding

隐写方法	隐写分段长度	回声幅度	回声延迟	隐写内容	嵌入率/%
K1	256~1 024个采样点之间随机选取	$\alpha = 0.3, 0.4, 0.5$	20~50个采样点之间随机选取	随机	100
K2		$\alpha_1, \alpha_2 = 0.1, 0.3/0.2, 0.4/0.3, 0.5$		二值序列	
K3		$\alpha = 0.1, 0.2, 0.3$			

3.2 实验设计

本文采用Pytorch1.4深度学习框架,在Windows 10操作系统上使用Resnet34网络作为隐写分析分类器,epoch大小设置为30,batchsize为16,网络初始学习率为0.000 1,使用交叉熵损失函数,采用Adam优化器和ReLU激活函数。

在本文中对隐写数据集进行实验时,分别选取2 000条隐写语音段和2 000条未隐写语音段作为训练集,剩余1 000条语音段作为测试集,将隐写分析准确率作为评价指标,有

表3 每个回声隐写数据集的分布

Table 3 Distribution of each echo hiding steganography dataset

语音分布	训练集		测试集		总共/条
	隐写语音数量/条	未隐写语音数量/条	隐写语音数量/条	未隐写语音数量/条	
French	400	400	100	100	1 000
German	400	400	100	100	1 000
English	400	400	100	100	1 000
Spanish	400	400	100	100	1 000
Italian	400	400	100	100	1 000
总共/条	2 000	2 000	500	500	5 000

$$ACC = \frac{TN + TP}{TN + FP + TP + FN} \times 100\% \quad (13)$$

式中: ACC表示回声隐写检测准确率;TN为不含秘密信息语音检测正确的个数;TP为含秘密信息语音检测正确的个数;FP为含秘密信息语音检测错误的个数;FN为不含秘密信息语音检测错误的个数。

3.3 实验测试与分析

实验1 验证本文提出的基于倒谱图像的回声隐写分析方法的有效性。将实验的语音数据进行预处理,生成倒谱图像,将倒谱图像作为输入采用Resnet34网络训练得到回声隐写分析的检测模型,检测时同样对语音数据进行预处理生成倒谱图像,使用训练好的检测模型对倒谱图像进行判别。另外与文献[14]提出的基于功率倒谱差分的方差和平坦度的方法,文献[18]提出的基于功率倒谱差分的方差、平坦度、偏度和峰度的方法进行对比实验,检测准确率如表4所示。从表4可以看出,回声隐写分析方法在不同回声隐写算法上的检测效果不同,同一种回声隐写方法回声幅度不同检测效果也不同。一般来说,回声幅度越低检测准确率越低,本文方法在传统回声隐写、双极性回声隐写和双向回声隐写上低回声幅度时检测准确率达到98.2%、98.6%和96.1%,相比其他两种方法均有较大提升。文献[14]采用功率倒谱多阶差分的方差以及平坦度来描述语音回声隐写后功率倒谱产生的峰值,在回声幅度较低时,功率倒谱出现的峰值幅度较低,并且功率倒谱本身也有起伏,功率倒谱的差分和平坦度描述峰值并不准确,导致检测效果不佳。文献[18]在文献[14]的基础上增加峰度和偏度来描述功率倒谱的峰值,在回声幅度时较低时检测效果得到了一定提升,但功率倒谱的峰度和偏度在低回声幅度时描述功率倒谱的峰值仍然受峰值低、功率倒谱本身起伏的影响。

回声隐写后倒谱产生的峰值是一个明显的视觉特征,在低回声幅度时,尽管峰值不够明显,但每一个检测分段中出现的峰值均位于回声延迟处,这些峰值可以在倒谱图像中构成线状结构,残差神经网络能够达到更深的网络深度,具有强大的图像分类和物体识别的能力,采用残差神经网络对倒谱图像进行检测分类,可以实现低回声幅度回声隐写的检测。

实验2 测试本文方法在不同回声隐写方法和不同回声幅度下训练得到的模型的通用性。在实

表4 不同隐写分析方法检测准确率对比

Table 4 Detection accuracy comparison of different steganalysis methods %

回声隐写方法	回声幅度	隐写分析方法		
		文献[14]	文献[18]	本文方法
K1	0.3	84.2	89.2	98.2
	0.4	94.3	95.8	99.6
	0.5	97.1	98.9	100
K2	0.3,0.1	86.3	88.1	98.6
	0.4,0.2	94.2	94.8	99.2
	0.5,0.3	97.8	98.2	100
K3	0.1	74.2	82.6	96.1
	0.2	96.5	96.9	98.4
	0.3	100	100	100

际应用中,对于待检测的语音,检测者往往不知道其使用了何种回声核何种回声幅度。对此,本文进行交叉实验,对于3.1节中9个不同回声核不同回声幅度下的隐写数据集,固定一个回声核的训练集训练网络模型,将训练好的网络模型检测其他回声核的测试集,实验结果如表5所示。

表5 交叉实验准确率
Table 5 Accuracy of cross-over testing %

隐写方法	回声幅度	K1			K2			K3		
		0.3	0.4	0.5	0.3,0.1	0.4,0.2	0.5,0.3	0.1	0.2	0.3
K1	0.3	—	97.6	98.4	97.6	97.8	97.5	95.2	96.7	96.8
	0.4	97.8	—	98.6	97.1	97.4	98.2	87.2	96.6	98.8
	0.5	96.9	99.2	—	94.6	96.8	97.9	78.8	94.4	100
K2	0.3,0.1	96.3	95.9	95.8	—	96.2	96.5	93.3	95.6	96.8
	0.4,0.2	94.2	95.2	96.7	95.5	—	98.5	84.7	94.2	96.6
	0.5,0.3	93.5	96.6	99.2	94.2	99.2	—	70.2	94.8	95.2
K3	0.1	94.8	95.6	97.6	95.7	95.7	96.8	—	96.9	96.9
	0.2	94.1	95.7	97.9	95.2	95.5	97.2	86.0	—	97.8
	0.3	89.6	95.0	96.2	88.3	94.8	98.0	67.8	95.6	—

由表5可以看出,回声幅度较大时数据集训练得到的网络模型,对回声幅度较小时数据集的检测准确率相对较低,倒谱图像出现的线状结构较深时训练,网络提取的特征层次相对较浅,对线状结构较浅的倒谱图像并不能较好地检测出来。在实际应用时,选择3类经典回声隐写中低回声幅度的数据进行模型训练,3种经典回声隐写低回声幅度隐写数据训练得到的模型可以对其他两种回声隐写进行检测。

实验3 测试倒谱图像与功率倒谱图像的性能差异。在本文中采用倒谱构造图像,同样功率倒谱也可以构造图像,功率倒谱^[22]的定义为

$$E_{pc}(k) = \left\{ \text{IFFT} \left[\ln \left| \text{FFT}(x(n)) \right|^2 \right] \right\}^2 \quad (14)$$

将语音数据构造成倒谱图像与构成功率倒谱图像进行对比实验,在3种经典回声隐写下检测准确率如表6所示。由表6可知,倒谱图像与功率倒谱图像检测性能基本持平,但是在K3回声隐写回声幅度为0.1时,功率倒谱图像作为检测输入却突然失效了。对比式(6)倒谱的定义与式(12)功率倒谱的定义,可以得知:倒谱是对幅度谱取对数,而功率倒谱是对功率谱取对数,在取对数后倒谱与功率倒谱为倍数关系,对检测影响不大,主要差异是由于功率倒谱在逆傅里叶变换后还进行了平方计算,该处理使得回声幅度较大时功率倒谱中出现的峰值相比倒谱更加明显,但是在回声幅

表6 倒谱图像与功率倒谱图像检测准确率对比

Table 6 Detection accuracy comparison of cepstrum image and power cepstrum image %

回声隐写方法	回声幅度	准确率	
		倒谱图像	功率倒谱图像
K1	0.3	98.2	98.6
	0.4	99.6	99.4
	0.5	100.0	100
K2	0.3,0.1	98.6	98.1
	0.4,0.2	99.2	99.2
	0.5,0.3	100.0	99.2
K3	0.1	96.1	60.8
	0.2	98.4	98.8
	0.3	100.0	99.5

度较小时,功率倒谱出现的峰值也会更小,K3隐写下回声幅度较大时和较小时倒谱图像与功率倒谱图像的差异如图10所示。

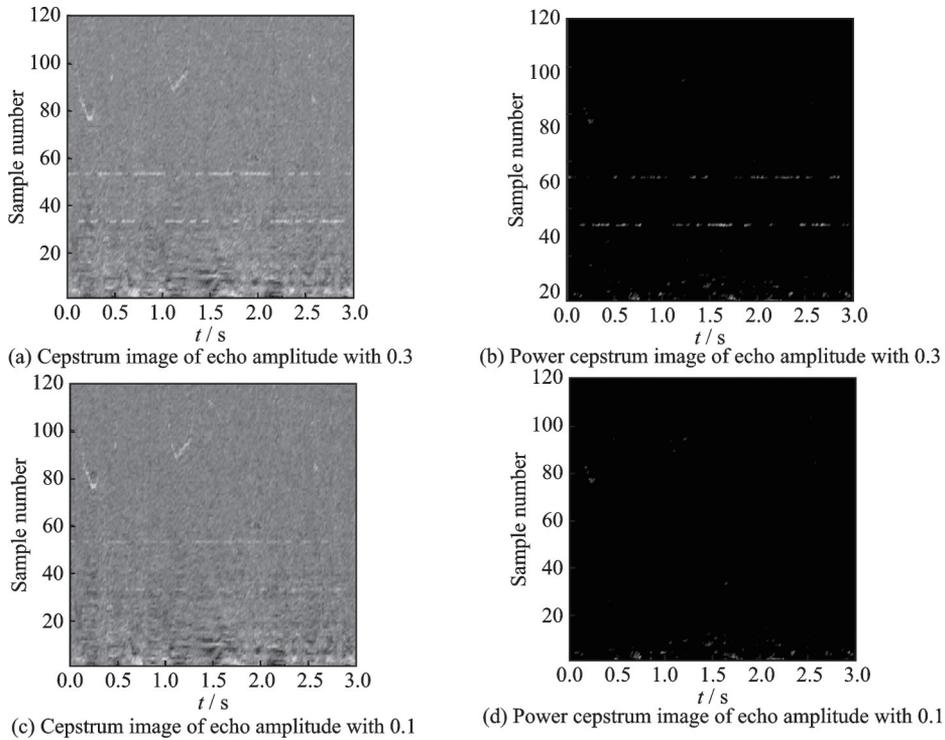


图10 K3隐写下语音倒谱图像与功率倒谱图像

Fig.10 Cepstrum image and power cepstrum image of speech after K3 steganography

由图10可以看出,在K3隐下回声幅度为0.3时,倒谱图像与功率倒谱图像都能够观察到线状结构,在K3隐写下回声幅度为0.1时,倒谱图像依然能够观察到线状结构,而功率倒谱图像中的线状结构却很难被观察到,只能看到功率倒谱本身就存在的一些高频率点,检测信息已经基本丢失。K3隐写下回声幅度为0.1时将功率倒谱图像作为检测输入,检测结果混淆矩阵如图11所示。由图11可知,500条含秘密信息语音有392条被检测为不含秘密信息语音,这是由于功率倒谱图像在K3隐写下低回声幅度时其线状结构很难被检测出来。由上述分析可知,在进行回声隐写分析时,应该选择倒谱构造图像作为回声隐写检测的输入。

4 结束语

本文针对传统回声隐写分析方法在低回声幅度检测性能的不足,提出了一种基于倒谱图像的语音回声隐写分析方法。实验分析证明,本文方法在3种经典回声隐写上低回声幅度时隐写检测准确率有较大提升。然后进行了交叉实验,测试了本文方法训练得到的网络模型之间的通用性。实验表明不同回声核、不同回声幅度训练得到的网络模型能够通用,具备实际应用的可能。最后分析了倒谱构造图

	检测	含秘密信息语音	不含秘密信息语音
实际			
含秘密信息语音		108	392
不含秘密信息语音		0	500

图11 功率倒谱图像在低回声幅度K3隐写混淆矩阵

Fig.11 Confusion matrix of detection with low amplitude under K3 steganography using power cepstrum image as detection input

像与功率倒谱构造图像的区别,实验结果表明选用倒谱构造图像在低回声幅度时更能够保留隐写检测的特征。下一步工作的重点是带噪语音的隐写分析。

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] 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*. [S.l.]: IEEE, 2001: 1341-1344.
- [3] KIM H J, CHOI Y H. A novel echo-hiding scheme with backward and forward kernels[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2003, 13(8): 885-889.
- [4] VENUGOPAL A, JAYADAS C K. Audio security system based on echo hiding and audio inpainting[J]. *International Journal of Creative Research Thoughts*, 2018, 6(2): 931-946.
- [5] RAHMAD C, MUHIQQIN I, PATMA T S, et al. Echo hiding method for video file message security[C]//*Journal of Physics: Conference Series*. [S.l.]: IOP Publishing, 2019: 066034.
- [6] THAN M, SIN S. Secure data transmission in MP3 file using LSB and echo hiding[J]. *International Journal of Advanced Research in Computer Science*, 2019, 10(4): 45.
- [7] WANG S, YUAN W, WANG J, et al. Inaudible speech watermarking based on self-compensated echo-hiding and sparse subspace clustering[C]//*Proceedings of 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. [S.l.]: IEEE, 2019: 2632-2636.
- [8] 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.
- [9] RAFIEE H, FAKHREDANESH M. Presenting a method for improving echo hiding[EB/OL].(2021-02-12)[2022-10-21].<https://arxiv.org/paper/2102/2102.06774.pdf>.
- [10] GEETHA S, ISHWARYA N, KAMARAJ N. Audio steganalysis with Hausdorff distance higher order statistics using a rule based decision tree paradigm[J]. *Expert Systems with Applications*, 2010, 37(12): 7469-7482.
- [11] KOCAL O H, YÜRÜKLÜ E, DİLAVEROĞLU E. Speech steganalysis based on the delay vector variance method[J]. *Turkish Journal of Electrical Engineering and Computer Sciences*, 2016, 24(5): 4129-4141.
- [12] LUO W, LI H, YAN Q, et al. Improved audio steganalytic feature and its applications in audio forensics[J]. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 2018, 14(2): 1-14.
- [13] 杨榆, 雷敏, 钮心忻, 等. 基于回声隐藏的 VDSC 隐写分析算法[J]. *通信学报*, 2009, 30(2): 83-88.
YANG Yu, LEI Min, NIU Xinxin, et al. VDSC steganalysis algorithm based on echo hiding[J]. *Journal on Communications*, 2009, 30(2): 83-88.
- [14] 王昱洁, 郭立, 王翠平. 基于功率倒谱统计特征的语音回声隐写分析方法[J]. *中国科学技术大学学报*, 2010, 40(10): 1074-1081.
WANG Yijie, GUO Li, WANG Cuiping. An audio steganalysis method for echo hiding based on statistical features of power cepstrums[J]. *Journal of University of Science and Technology of China*, 2010, 40(10): 1074-1081.
- [15] QI Y C, ZHANG J N, WANG W L. Time-spread echo hiding steganalysis algorithm based on ensemble learning[C]//*Proceedings of 2011 International Conference on Network Computing and Information Security*. [S.l.]: IEEE, 2011: 426-429.
- [16] WANG Y, WEN H, JIAN Z, et al. Steganalysis on positive and negative echo hiding based on skewness and kurtosis[C]//*Proceedings of 2014 9th IEEE Conference on Industrial Electronics and Applications*. [S.l.]: IEEE, 2014: 1235-1238.
- [17] 王杰, 王让定, 严迪群, 等. 基于卷积神经网络框架的回声隐藏检测方法[J]. *计算机应用*, 2020, 40(2): 375-380.
WANG Jie, WANG Rangding, YAN Diqun, et al. Detection method for echo hiding based on convolutional neural network framework[J]. *Journal of Computer Applications*, 2020, 40(2): 375-380.
- [18] 邓伟. 音频隐写分析方法研究[D]. 南京: 东南大学, 2021.

DENG Wei. Research on audio steganalysis[D]. Nanjing: Southeast University, 2021.

[19] 韩纪庆. 音频信息处理技术[M]. 北京: 清华大学出版社, 2007: 44-46.

HAN Jiqing. Audio information processing technology[M]. Beijing: Tsinghua University Press, 2007: 44-46.

[20] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.[S.l.]: IEEE, 2016: 770-778.

[21] VoxForge.org. Free speech recognition (Linux, Windows and Mac) [EB/OL]. (2018-12-19). <http://www.voxforge.org/>.

[22] YAN B, SUN S H, LU Z M. Improved echo hiding using power cepstrum and simulated annealing based synchronization technique[C]//Proceedings of the 2003 International Conference on Machine Learning and Cybernetics. [S.l.]: IEEE, 2003: 2142-2147.

作者简介:



唐军豪(1998-),男,硕士研究生,研究方向:语音信号处理、隐写分析,E-mail: mailtangjh@163.com。



杜庆治(1977-),通信作者,男,高级实验师,研究方向:无线通信、信息安全,E-mail: 57960748@qq.com。



龙华(1963-),女,博士,教授,研究方向:无线通信、信号处理。



邵玉斌(1970-),男,教授,研究方向:音频信号处理、自然语言处理。



李一民(1963-),男,博士,教授,研究方向:图像处理、信息处理、计算机视觉。

(编辑:刘彦东)