

密级: 公开



中国科学院大学
University of Chinese Academy of Sciences

硕士学位论文

MP3 音频隐写分析研究

作者姓名: 王运韬

指导教师: 赵险峰 研究员

中国科学院信息工程研究所

学位类别: 工学硕士

学科专业: 信号与信息处理

培养单位: 中国科学院信息工程研究所

2019 年 6 月

Research on MP3 Audio Steganalysis

**A dissertation submitted to
University of Chinese Academy of Sciences
in partial fulfillment of the requirement
for the degree of
Master of Science in Engineering
in Signal and Information Processing
By
Wang Yuntao
Supervisor: Professor Zhao Xianfeng**

Institute of Information Engineering, Chinese Academy of Sciences

June, 2019

中国科学院大学 研究生学位论文原创性声明

本人郑重声明：所呈交的学位论文是本人在导师的指导下独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明或致谢。

作者签名：

日 期：

中国科学院大学 学位论文授权使用声明

本人完全了解并同意遵守中国科学院有关保存和使用学位论文的规定，即中国科学院有权保留送交学位论文的副本，允许该论文被查阅，可以按照学术研究公开原则和保护知识产权的原则公布该论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存、汇编本学位论文。

涉密及延迟公开的学位论文在解密或延迟期后适用本声明。

作者签名：

日 期：

导师签名：

日 期：

摘要

音频应用与音频处理软件的普及极大地促进了音频隐写与音频隐写分析技术的发展。MP3 音频凭借其高压缩率和高品质音质的特性，成为互联网中传播最为广泛的音频格式之一。相比于 WAV 格式的音频，MP3 音频的编码原理更为复杂、公共信道更为丰富、所需的存储空间更小，是非常理想的隐写载体。MP3 音频隐写与隐写分析研究也成为当前多媒体安全领域的重要方向之一。

针对当前 MP3 音频隐写分析算法通用性差、检测精度不足、发展较为滞缓等问题，本论文分别从手工特征设计和深度学习两个角度出发，提出了三种基于 MP3 音频量化修正余弦变换（Quantified Modified Discrete Cosine Transform, QMDCT）系数矩阵的隐写分析算法，可应用于多种 MP3 音频隐写算法的检测分析，在负载率较低时仍能有较为出色的表现。同时，本论文所提算法可在一定程度上解决输入矩阵尺寸失配以及算法与隐写负载率未知条件下的盲检测分析，具有较强的实用性。

本论文的主要工作与创新包含以下五个方面：

1. 提出一种适用于 MP3 音频隐写分析的富高通滤波模型。隐写信号相比于原始的音频信号可以看作是一种加性高频噪声，为了能有效地“放大”隐写信号的痕迹，提升所提特征或网络对隐写信号的敏感性，引入高通滤波器对输入矩阵进行预处理是十分必要的。本论文通过分析多种高通滤波器对 MP3 正常/隐写音频对其 QMDCT 系数矩阵的影响，提出一组适用于 MP3 音频隐写分析的通用富高通滤波模型，从多个角度对隐写信号进行“放大”，有效提升了隐写分析算法的性能。而且，这一滤波模型不仅适用于手工特征设计的隐写分析方法，同时也适用于基于深度学习的隐写分析网络设计。

2. 提出一种基于 QMDCT 系数矩阵多尺度相关性度量的 MP3 音频隐写分析方法。在 MP3 音频编码过程中，大值区和小值区的 QMDCT 系数在 Huffman 编码时具有不同的映射关系。基于这一编码特性，本论文在相邻 QMDCT 系数矩阵相关性度量的基础上，引入了 2×2 和 4×4 码字保持降采样模型，实现了对隐写前后 QMDCT 系数矩阵相关性的多尺度度量，有效提升了基于手工特征设计的 MP3 音频隐写分析算法的检测效果。

3. 提出一种基于 QMDCT 系数矩阵与全卷积神经网络的 MP3 音频隐写分析方法。基于卷积神经网络的 MP3 音频隐写分析方法可以在网络构建完成后自动

地学习到不同类型输入数据的潜在特性，相比于传统的手工特征设计，算法设计更为简便，且性能更为优越。通过分析输入数据类型、卷积核尺寸与个数、池化类型、激活函数类型以及网络结构等若干因素对网络性能的影响，本论文设计了一种性能优越的全卷积神经网络结构，极大地提升了 MP3 隐写音频的检测精度。该网络还能够在一定程度上解决输入数据尺寸失配的问题，具有较强的实用性。

4. 提出一种基于 QMDCT 系数矩阵与多尺度卷积神经网络的 MP3 音频隐写分析方法。为了更好地提升网络的效率，减少模型的存储空间，本论文对卷积神经网络结构做了进一步改进。首先，引入多尺度卷积网络用以实现对输入数据特性的多尺度度量，增强模型在感受野内对局部区域的识别能力。其次，在网络设计过程中以卷积核因式分解的方式减少网络的参数个数。最后，在子网设计中引入残差结构降低网络退化的风险。通过以上改进，网络的检测精度在无明显降低的情况下，模型大小由 1.7G 缩减为 47M，实现了网络参数的轻量化。

5. 构建了一个适用于 MP3 音频隐写分析的公共基础数据集。针对现阶段音频隐写分析领域标准数据集缺乏的问题，本论文通过互联网音频爬取的方式，构建了一个基础音频数据集，数据集包含了标准格式的 WAV 音频 33038 个，音频时长为 10s，并将其分别编码成相应的 MP3 正常/隐写音频对，分别包括当下最为经典的三类 MP3 音频隐写算法，128kbps、192kbps、256kbps、320kbps 等四种常用比特率以及多种隐写负载率，方便后续的各项研究工作的开展。

关键词： MP3，音频，隐写，隐写分析，卷积神经网络

Abstract

The development of audio steganography and steganalysis has been promoted due to the popularization of audio applications and audio processing tools. Because of the high compression ratio with good sound quality, abundant sharing platform and small storage space need, MP3 is regarded as one of the best carriers for audio steganography. Therefore, MP3 audio steganography and steganalysis have become one of the main tasks of multimedia security.

In consideration of the truth that MP3 audio steganalytic algorithms are lack of versatility, of poor detection performance and in a slow development process at present. Three steganalytic algorithms are proposed, and these algorithms are designed based on the quantified modified discrete cosine transform (QMDCT) coefficients matrix of MP3 audio from two aspects of handcrafted features and deep learning-based methods, which are suitable for a variety of MP3 steganographic algorithms, bitrates, and relative payloads. In particular, the detection performance of the three algorithms is still better on the steganalysis of adaptive MP3 steganography. What's more, steganalysis with input size mismatch and blind steganalysis are solved to some extents.

The main contributions and innovations are as follows.

1. A rich high pass filtering module is proposed. Compared with the original audio signal, the steganographic signal can be regarded as additive high-frequency noise. The introduction of high-pass filters is beneficial to “enlarge” the trace of the steganographic signal and boost the sensitivity of the proposed algorithms to the steganographic signal. A group of effective high pass filters is designed according to the diverse influence on the QMDCT coefficients matrix between the MP3 cover and stego pairs. The universal module can be applied to diverse steganographic methods.

2. A steganalytic algorithm based on the multi-scale correlation of QMDCT coefficients matrix is proposed. One Huffman codeword in Big-Value region is corresponding to two QMDCT coefficients, and one Huffman codeword in Count1 region is corresponding to four QMDCT coefficients. In view of the encoding characteristic of MP3 audio, we additionally introduce the 2×2 and 4×4 codeword-aware down-sampling module for the measurement of multiscale correlation characteristics of the

QMDCT coefficients matrix, which effectively improves the detection performance of MP3 audio steganalysis.

3. A steganalytic algorithm based on QMDCT coefficients matrix and convolutional neural networks (CNNs) is proposed. CNN-based steganalytic algorithms are better of the gaining of potential characteristics compared with the traditional hand-crafted features, which leads to a simpler design thinking and a better detection performance. In the design of network structure, we discuss the influence of the type of input data, convolutional kernel size and numbers, the type of pooling, the type of activation function and so on, and the detection accuracy is over 80% in the condition relative embedding rate is less than 2%. The network can be applied to deal with the input data with the size mismatch to some extents, which is more practical compared with other CNN-based methods.

4. A steganalytic algorithm based on QMDCT coefficients and multi-scale CNNs is proposed. To boost the efficiency of CNN-based algorithms and reduce the storage space need of models, we update the structure of the network further. First, multiscale characteristics of the input data are obtained through the construction of multiscale sub-network. And, the ability of the recognition of the local areas in receptive fields is enhanced. The structure of the network is further optimized through convolutional kernel decomposition and skip connection. Thus, the model size if reduced from 1.7G to 47M, and the detection accuracy maintained well.

5. A dataset for MP3 audio steganalysis is constructed. Until now, there is no standard dataset for audio steganalysis. Thus, we construct a basic dataset which consists of 33038 WAV audio files with the duration of 10s. All WAV audio files are obtained via crawler. The WAV audio files are then encoded as corresponding cover/stego pairs, and the dataset includes three typical steganographic algorithms, four common bitrates and plenty of relative payloads, which is facilitated to the development of follow-up work.

Key Words: MP3, Audio, Steganography, Steganalysis, CNNs

目 录

摘要	I
Abstract	III
目录	V
图形列表	IX
表格列表	XI
术语列表	XIII
第1章 绪论	1
1.1 研究背景及意义	1
1.2 音频隐写分析研究现状及发展趋势	3
1.2.1 专用隐写分析方法	3
1.2.2 通用隐写分析方法	6
1.2.3 音频隐写分析未来发展趋势	9
1.3 本文主要研究内容	9
1.4 本文组织结构	10
第2章 相关研究概述	13
2.1 MP3 编解码原理介绍	13
2.1.1 MP3 编码原理	13
2.1.2 MP3 解码原理	15
2.2 MP3 隐写算法介绍	16
2.2.1 MP3Stego 隐写算法	16
2.2.2 HCM 隐写算法	17
2.2.3 EECS 隐写算法	18
2.3 数据集构建	19
2.4 MP3 音频隐写分析算法介绍	20
2.4.1 QMDCT 系数矩阵	20
2.4.2 ADOTP 隐写分析算法	22
2.4.3 MDI2 隐写分析算法	23

第 3 章 基于多尺度相关性度量的 MP3 音频隐写分析研究	25
3.1 引言	25
3.2 算法设计	26
3.2.1 音频富高通滤波模型	26
3.2.2 基于 MP3 编码特性的多尺度相关性度量模型	28
3.2.3 特征优选与降维	31
3.3 实验设计与分析	35
3.3.1 实验设置	36
3.3.2 模块有效性验证	36
3.3.3 隐写音频帧定位	37
3.3.4 实验结果与分析	38
3.4 本章小结	43
第 4 章 基于卷积神经网络的 MP3 音频隐写分析研究	45
4.1 引言	45
4.2 算法设计	46
4.2.1 输入数据类型评估	47
4.2.2 网络结构设计	49
4.2.3 面向输入数据尺寸失配的音频隐写分析	54
4.2.4 基于隐写负载率的迁移学习	56
4.3 实验设计与分析	59
4.3.1 实验设置	60
4.3.2 网络结构优选	60
4.3.3 实验结果与分析	62
4.4 本章小结	66
第 5 章 基于多尺度卷积神经网络的 MP3 音频隐写分析研究	67
5.1 引言	67
5.2 算法设计	69
5.2.1 多尺度卷积结构设计	69
5.2.2 卷积核分解	70
5.2.3 残差网络结构设计	71
5.3 实验设计与分析	72
5.3.1 实验设置	73
5.3.2 子网结构优选	73
5.3.3 MP3 音频隐写分析网络性能评估	75
5.3.4 实验结果与分析	75
5.4 本章小结	83

第 6 章 总结与展望	85
6.1 全文总结.....	85
6.2 下一步工作	86
参考文献	87
附录 A 音频隐写分析数据集	95
附录 B 隐写负载率转换表	97
附录 C 音频分享平台	101
致 谢	103
作者简历及攻读学位期间发表的学术论文与科研成果	105

图形列表

1.1 音频隐写与隐写分析模型示意图	3
1.2 通用隐写分析方法一般架构	7
2.1 MP3 编码原理示意图	13
2.2 QMDCT 系数结构图	14
2.3 大值区 QMDCT 系数与 Huffman 码字映射关系图	14
2.4 小值区 QMDCT 系数与 Huffman 码字映射关系图	15
2.5 MP3 解码原理示意图	15
2.6 MP3 音频隐写算法分类示意图	17
2.7 QMDCT 系数矩阵构造示意图	21
2.8 正常音频与隐写音频的 QMDCT 系数矩阵差异图 (128kbps)	22
2.9 ADOTP 隐写分析算法流程图	23
2.10 MDI2 隐写分析算法流程图	24
3.1 MSC 算法流程图	26
3.2 KV 核滤波效果示意图	27
3.3 正常音频与隐写音频 QMDCT 系数分布差异图 (EECS, 128kbps)	28
3.4 多尺度相关性度量模型示意图	29
3.5 2×2 码字保持降采样模块示意图	30
3.6 2×2 码字保持降采样后各子矩阵 Markov 转移概率矩阵示意图 (128kbps, 水平方向)	30
3.7 正常音频与隐写音频特征分布差异图 (EECS, 128kbps, M 型滤波器)	31
3.8 QMDCT 系数分布图。 (a) 为原始 MP3 音频的 QMDCT 系数分布, (b) - (d) 分别为经过 MP3Stego、HCM 和 EECS ($SPR = 2$) 算法隐写后被修改的 QMDCT 系数分布	32
3.9 对三种隐写算法在不同截断阈值选择下的检测准确率曲线	33
4.1 基于手工特征设计的隐写分析 (A) 与基于卷积神经网络的隐写分析 (B) 对比图	45
4.2 RHFCN 结构示意图 ($16 \times (3 \times 3 \times 9)$ 表示卷积核尺寸为 3×3 , 卷积核个数为 16, 输入特征图通道数为 9。“S1” 表示步长为 1, “SAME” 和 “VALID” 为两种卷积补齐方式, 各方框下为输出特征图的维度。)	46

4.3 Conv 模块示意图	47
4.4 不同数据空间下正常音频与隐写音频的差异图（EECS, 128kbps, $SPR = 2$ ）	48
4.5 用于输入数据评估的网络结构 - Light-RHFCN	48
4.6 神经网络结构示意图	49
4.7 两种卷积模块示意图	50
4.8 四种常用激活函数	52
4.9 带有全连接层的网络结构与全卷积网络结构比较示意图	54
4.10 基于滑动窗口的 MP3 音频隐写分析示意图	55
4.11 机器学习与迁移学习路径搜索比较图	56
4.12 迁移学习分类图	57
4.13 修改的 QMDCT 系数在不同隐写条件下的分布图	58
5.1 以多层感知机为子网的网络结构	67
5.2 RHMSCN 结构示意图。“Sub-Net, 8” 表示当前子网内的各尺寸卷 积核个数均为 8, “S2” 表示步长为 2, 各方框下为输出特征图的维 度）	68
5.3 子网结构示意图	68
5.4 多尺度卷积示意图	69
5.5 卷积核分解示意图	70
5.6 VGG 风格与 ResNet 风格的卷积神经网络示意图	71
5.7 多尺度子网结构示意图（Sub-Net 1）	73
5.8 引入残差结构的多尺度子网结构示意图（Sub-Net 2）	74
5.9 引入卷积核分解和残差结构的多尺度子网结构示意图（Sub-Net 3）	74

表格列表

2.1 不同音频比特率的 QMDCT 系数零区起始索引平均值	21
3.1 各高通滤波器处理后正常音频与隐写音频的差异 (EECS, 128kbps, $SPR = 2$, $T = 7$)	34
3.2 各高通滤波器“贡献度”及排名 (EECS, 128kbps, $SPR = 2$, $T = 7$)	34
3.3 基于后减枝的子滤波器组最优组合模式选择 (EECS, 128kbps, $SPR = 2$, $T = 7$)	35
3.4 实验设置	36
3.5 子模块有效性验证 (EECS, 128kbps, $SPR = 2$)	37
3.6 不同长度音频段对隐写分析算法性能的影响 (EECS, 128kbps, $SPR = 2$)	38
3.7 对 MP3Stego 算法的隐写分析结果	40
3.8 对 HCM 算法的隐写分析结果	41
3.9 对 EECS 算法的隐写分析结果	42
4.1 输入数据类型对 MP3 音频隐写分析网络性能的影响 (EECS, 128kbps, $SPR = 2$)	49
4.2 不同尺寸输入数据的检测准确率 (EECS, 128kbps, $SPR = 2$) ..	55
4.3 基于隐写负载率的迁移学习隐写分析结果 (EECS, 128kbps) ..	58
4.4 实验设置	59
4.5 不同网络结构下的隐写分析准确率 (EECS, 128kbps, $SPR = 2$) ..	61
4.6 对 MP3Stego 算法的隐写分析结果	63
4.7 对 HCM 算法的隐写分析结果	64
4.8 对 EECS 算法的隐写分析结果	65
5.1 实验设置	72
5.2 子网性能评估 (EECS, 128kbps, $SPR = 2$)	75
5.3 MP3 音频隐写分析网络评估 (EECS, 128kbps, BatchSize = 16) ..	75
5.4 对 MP3Stego 算法的隐写分析结果	77
5.5 对 HCM 算法的隐写分析结果	78
5.6 对 EECS 算法的隐写分析结果	79
5.7 用于 MP3 音频盲隐写分析的训练模型说明	80
5.8 混合测试数据集说明	80

5.9 对混合数据集的盲隐写分析结果 (Model 1)	81
5.10 对混合数据集的盲隐写分析结果 (Model 2)	81
5.11 对混合数据集的盲隐写分析结果 (Model 3)	81
5.12 对 HCM 算法的隐写分析检测结果 (128kbps)	82
5.13 对 EECS 算法的隐写分析检测结果 (128kbps)	82
A.1 数据集详细信息	95
B.1 隐写负载率转换表 (MP3Stego)	97
B.2 隐写负载率转换表 (HCM)	98
B.3 隐写负载率转换表 (EECS)	99
C.1 音频分享平台列表	101

术语列表

FPR	False Positive Rate, 虚警率
FNR	False Negative Rate, 漏检率
TPR	True Positive Rate, 真阳性率
TNR	True Negative Rate, 真阴性率
SRM	Spatial Rich Models, 空域富模型
AMR	Adaptive Multi-Rate, 自适应多码率
GRU	Gated Recurrent Unit, 门控循环单元
SVM	Support Vector Machine, 支持向量机
RER	Relative Embedding Rate, 相对嵌入率
PCM	Pulse Code Modulation, 脉冲编码调制
AAC	Advanced Audio Coding, 高级音频编码
HCM	Huffman Code Mapping, 哈夫曼码字映射
STC	Syndrome-Trellis Codes, 校验子格编码
SPR	Steganographic Payload Rate, 隐写负载率
RNN	Recurrent Neural Network, 递归神经网络
LSTM	Long Short-Term Memory, 长短期记忆网络
FFT	Fast Fourier Transformation, 快速傅立叶变换
CNN	Convolutional Neural Network, 卷积神经网络
GAN	Generative Adversarial Network, 生成对抗网络
EECS	Equal Entropy Code Substitution, 等长熵码字替换
MFCC	Mel Frequency Cepstrum Coefficient, 梅尔频率倒谱系数
QMDCT	Quantified Modified Discrete Cosine Transform, 量化修正余弦变换
MP3	Moving Picture Experts Group Audio Layer III, 动态影像专家压缩标准音频层面 3
IQMDCT	Inverse Quantified Modified Discrete Cosine Transform, 反量化修正余弦变换

第1章 绪论

1.1 研究背景及意义

隐写（Steganography），泛指通过将秘密消息隐蔽在可公开的媒体信息中并传递给特定的接收者而难以被第三方察觉其保密通信事实的科学和技术。与传统的密码技术（Cryptography）[1, 2]有所不同，隐写主要是利用人类的视听觉冗余以及图像、音视频等多媒体自身的数据冗余，将秘密信息隐藏在数字媒体中，并通过互联网等公共信道进行传输。随着互联网技术的飞速发展以及各类电子设备、多媒体处理软件的不断普及，隐写技术得到了迅速的发展。现如今，基地组织成员、不法分子以及多国的情报人员等都会利用隐写技术传输秘密消息，对国家政治、军事、经济以及公共安全产生了极大的影响。因此，为了能够对这一行为做出有效的应对，隐写分析（Steganalysis）技术应运而生，并在近年来得到了广泛的关注，被众多国家安全部门应用于实战分析中。

隐写与隐写分析技术不仅会对国家公共安全产生影响，同时也可会影响到我们每一个人的生活。随着科技的发展，网络罪犯也越来越擅长掩盖自己的行踪和隐藏自身犯罪信息。目前，有超过一千种已知的工具可以在图像、音频、视频、网络协议和其他类型的数字载体中隐藏数据。近年来，恶意软件开发人员也一直在积极地将数据隐藏功能集成到恶意代码中，以创建一种被称为无文件恶意软件的高级持久威胁（Advanced Persistent Threat, APT）形式。McAfee 实验室对 2018 年平昌奥运会期间发生的恶意软件攻击进行的一项分析显示，一个恶意的 PowerShell 脚本嵌入在一个 Microsoft Word 文档内的图像中。当用户打开包含图像的 Word 文档时，此恶意脚本便开始执行。调查人员通过调用 stegohunt[3] 工具，识别出文件中所包含的各类潜在恶意或敏感信息的图像，并利用 Wetstone 的隐写分析工具进一步调查所嵌入的数据。

以载体类型为划分依据，隐写分析技术分别包含有图像、视频、音频、文本与网络协议隐写分析。与图像及视频隐写分析技术相比，音频隐写分析，尤其是 MP3 音频隐写分析具有更为广阔的研究空间。

1. MP3 音频是当下最流行的压缩域音频编码格式之一，互联网中 70% 以上的音频资源以 MP3 格式的音频分布与传播。因此，利用 MP3 音频进行秘密消息传输更不易被察觉。

2. MP3 编码原理较为复杂，可探索的隐写空间更为丰富，理论上可以设计

出隐蔽性更高的隐写算法。

3. 与 WAV 格式的音频相比, MP3 音频的压缩比最高可达 10: 1, 所需的传输带宽和存储空间更少; 同时, 与互联网中常用的图像相比, MP3 音频可以嵌入更多的秘密消息; 相比于视频媒体, MP3 音频又具备了大小适中的优点, 更易被用于传播。

4. 根据前期对互联网音频公共分享平台的调研, 目前已统计有包括 instaudio、clyp、reverbnation 等在内的多款国内外音乐分享平台, 此类平台的用户群体大, 支持 MP3 格式的音频上传与下载, 且不会对上传的 MP3 音频进行压缩、转码或重编码, 相比于图像及视频媒体有很大的优势。

然而, 由于 MP3 音频编解码器相对复杂, 且此前的研究人员大多将精力投入至以图像为载体的隐写与隐写分析算法设计中, 现有的 MP3 隐写算法较少, 从而也导致 MP3 音频隐写分析发展相对较为缓慢, 目前仍存在以下问题:

1. 面向音频媒体的隐写分析仍以时域音频载体为主, 不仅时域音频的传播远不及 MP3 音频广泛, 而且基于时域统计特性的隐写分析算法并不适用于 MP3 音频的隐写分析。

2. 针对 MP3 音频的隐写分析仍主要集中在对 MP3Stego[4] 的检测分析, 对于隐写容量大的哈夫曼 (Huffman) 码字替换类算法 [5, 6] 等在内其他 MP3 音频隐写算法的检测能仍有待提升。

3. MP3 音频自适应隐写与自适应隐写分析算法的研究尚处于初级阶段, 还有较大的研究空间。

综上所述, 研究 MP3 音频隐写分析, 其意义在于:

1. 提升 MP3 音频隐写分析算法的有效性。通过分析各类 MP3 隐写算法对 MP3 音频各数据空间的影响, 挖掘其潜在规律, 结合最新的学科态势与技术, 提出合理有效的 MP3 隐写分析算法, 提升 MP3 隐写分析算法的性能, 以有效遏制基于 MP3 音频的非法隐蔽通信行为。

2. 增强 MP3 音频隐写分析算法的实用性。结合模板失配、数据尺寸失配以及盲隐写分析等现实问题, 提升隐写分析系统在面对现实场景的处理能力, 实现算法在真实环境下的高精度、快速检测, 降低 MP3 音频分析算法的平台依赖性, 实现对 MP3 音频隐写分析算法实用性的提升。

3. 促进 MP3 音频隐写算法的发展。如图1.1所示, 音频隐写与隐写分析技术互为攻防, 隐写技术的发展将会促进隐写分析技术的进步, 同时隐写分析技术

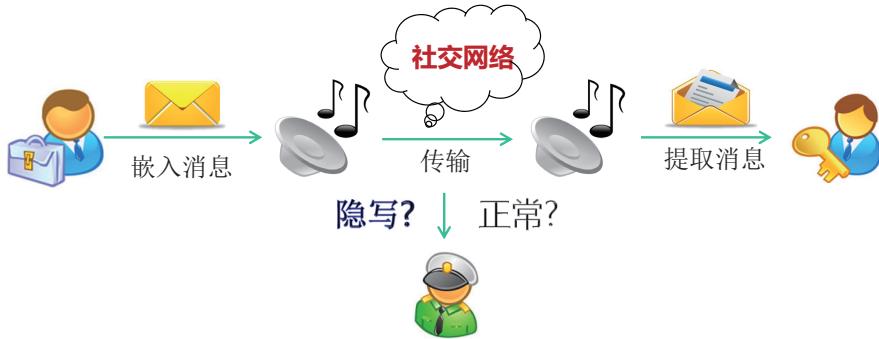


图 1.1 音频隐写与隐写分析模型示意图

Figure 1.1 Diagram of the relationship between audio steganography and steganalysis

的提升又会进一步推动隐写技术的发展。因此，通过 MP3 音频隐写分析的发展，不断促进 MP3 音频及其他压缩域音频隐写技术的进步。

1.2 音频隐写分析研究现状及发展趋势

经过近十多年的发展，现阶段的音频隐写分析算法可以划分为专用隐写分析、通用隐写分析以及基于深度学习的隐写分析等。通用性不断增强，模型复杂度逐步提升，更多的学科知识与技术得以综合运用。

1.2.1 专用隐写分析方法

专用隐写分析主要是针对某一种或某一类隐写算法设计的分析方法，通常具有特征简单、检测准确率高但适用范围窄的特点。由于很多隐写算法的思想和原理在音频和图像中都是通用的，相应的检测分析方法也是类似的，例如，针对最低有效位嵌入（Least Significant Bits Replace, LSBR）的卡方（Chi-Square, χ^2 ）分析法 [7]、正则奇异（Regular and Singular, RS）分析法 [8] 和样本对分析法（Sample Pair Analysis, SPA）[9] 等，在此便不再赘述。此外，本节将主要介绍音频所特有的几类专用分析算法，分别有：回声隐写分析、扩频隐写分析、乘性嵌入隐写分析、MP3Stego 隐写分析以及音频隐写软件分析。

1.2.1.1 回声隐写分析

回声隐写方法提出时间较早，方法研究也很充分，针对回声隐写分析的研究也很多。Zeng 和 Ai 等人 [10] 在对音频信号进行短时分析时发现，隐写后音频的功率倒谱会出现尖峰，由此提出一种基于窗口倒谱在不同时刻的局部最大值的隐写分析特征。在对伪随机编码（Pseudo-Noise, PN）序列的隐写分析中，Wang 和 Hao 等人 [11] 通过计算回声核与待测信号的卷积发现，隐写信号的功率

倒谱和复频谱会在回声核的延时部分出现尖峰。对音频信号分帧后分别计算各个音频帧的功率倒谱和复频谱幅度的偏度，并将偏度向量的峰度作为隐写分析的特征。实验结果表明，基于复频谱幅度的特征优于基于功率倒谱的特征。Xie 和 Cheng 等人 [12] 对回声隐写系统做了进一步的研究，提出了一种基于滑动窗口倒谱行为分析的主动攻击的方法，引入倒谱尖峰位置聚集率（Cepstrum Peak Location Aggregation Rate, CPLAR）特征用于回声隐写分析。这种方法同时还可以确定分割长度，但较难检测出 PN 回声核隐写。杨榆等人 [13] 提出了一种基于倒谱和差分方差统计量（Variants of Difference of Sum of Cepstrum, VDSC）的回声隐写分析法，通过 VDSC 量度量隐写对原始信号的影响。

1.2.1.2 扩频隐写分析

扩频技术最早应用于数字水印领域，后来研究人员也利用该技术来提高隐写方法的鲁棒性。Altun 等人 [14] 提出一种基于边际失真递减原理的形态学隐写分析方法，将音频信号模拟为一阶自回归过程，证明了正常音频与携带测试水印的正常音频之间的汉明距离大于隐写音频与携带测试水印的隐写音频之间的汉明距离，由此实现隐写分析。实验结果表明，在噪声标准差大于 5×10^{-3} 的条件下，检测准确率达到 90% 以上。Gao 等人 [15] 提出一种基于小波变换的扩频隐写分析方法，将隐写模拟为音频加噪过程，将去噪前后的小波系数差值作为特征向量用于训练，由此实现隐写分析。实验结果表明，当隐写负载大于 0.5% 时，检测准确率可达 82.5%。Li 等人 [16] 提出了基于离散小波变换（Discrete Wavelet Transform, DWT）和高斯混合模型（Gaussian Mixture Model, GMM）的直接序列扩频（Direct Sequence Spread Spectrum, DSSS）分析方法，利用 GMM 对 DWT 细节子带系数进行建模，并计算 GMM 的概率密度函数，最后将偏度和峰度作为隐写分析特征。黄昊等人 [17] 从 DSSS 的隐写原理出发，提取失真测度作为特征向量进行训练和检测，并对嵌入容量的干扰提出两种改善策略，平均检测准确率可达 94.2%。谢春辉等人 [18][19] 提出了一种 DSSS 信号 PN 序列的估计方法，实现了扩频隐藏信息的盲提取。实验结果表明，当 $SNR > -5dB$ 时扩频码序列正确估计率接近 100%。当扩频序列长度为 255、嵌入强度为 0.08 时，扩频隐藏信息提取正确率达到 90.73%。

1.2.1.3 乘性嵌入隐写分析

乘性嵌入隐写算法引入的隐写噪声具有可乘性且与隐写载体独立，传统的隐写分析算法无法很好地实现此类隐写算法的检测分析。Qi 等人 [20] 提出对时

域采样点建立对数运算模型，将乘性噪声转换为加性噪声，并在此基础上对信号进行小波变换，由此实现对乘性隐写算法的检测分析。此外，Ru 等人 [21] 通过提取载体信号的统计特性、子带系数以及各个子带的线性预测残差系数实现乘性嵌入隐写算法的分析。

1.2.1.4 MP3Stego 隐写分析

MP3Stego 是过去十年最具代表性的压缩域音频隐写算法之一，主要有三个方面的原因：（1）MP3Stego 是最早的 MP3 音频隐写算法，并具有很好的透明性和编码兼容性；（2）MP3Stego 是一款开源的隐写软件，吸引了众多研究人员的广泛关注和研究，且经常被用作夺旗赛（Capture The Flag, CTF）的题目之一；（3）在一段时期内，MP3Stego 被用作隐写分析方法性能的标准算法。1999 年，Westfeld[7] 最早对 MP3Stego 算法进行了分析，他通过统计分析发现，在经过 MP3Stego 隐写后，part2_3_length 的方差会变大，由此实现对 MP3Stego 的隐写分析。此外，他还分析了不同 MP3 编码器对隐写分析的影响 [22]，并对嵌入消息长度做了近似估计。此后，宋华等人 [23] 通过计算 MP3 文件中 part2_3、Stuffing Bits 的统计量，实现对 MP3Stego 算法的检测分析。Hernandez-Castro 等人 [24] 通过实验发现 MP3Stego 隐写算法会引起编码比特池长度的改变，以比特池长度的相对误差量来实现对 MP3Stego 的检测分析。陈益如等人 [25] 发现 MP3Stego 算法会不同程度地改变 Huffman 码表索引值，从而以索引值的二阶差分值作为隐写分析特征以实现对 MP3Stego 的隐写分析。万威等人 [26] 提出一种基于 Huffman 码表分布特征和音频重编码的 MP3Stego 隐写分析方法。Yu 等人 [27, 28] 利用边信息中 main_data_begin 字段位置的分布来检测 MP3Stego 隐写音频。针对边信息特征在低嵌入率时检测准确率低的问题。李友勇等人 [29] 提出了一种改进方法，将隐写前后的 main_data_begin 字段位置差扩大，以减小特征提取时重压缩估计方法带来的误差。Yan 等人 [30] 提出一种基于量化步长差分统计量的 MP3Stego 隐写分析算法。羊云开等人 [31][32] 提出了一种针对 MP3Stego 算法的主动分析方法，利用突变点分析法实现了对秘密消息长度的估计。Yan 等人 [33] 提出了一种基于比特池长度的 MP3Stego 隐写分析算法，以 MP3 颗粒比特池长度均值和方差的比值作为分析特征，并提出了基于重压缩校正的原始音频载体估计方法，该算法在 0.05% 极低嵌入率下的检测准确率为 93%。余先敏等人 [34] 利用隐写后量化修正余弦变换（Quantified Modified Discrete Cosine Transform, QMDCT）大值区系数的稀疏性，以重压缩前后大值区系数方差的差值作为隐写分析特征，从而实现 MP3Stego 的隐写分析。

1.2.1.5 音频隐写软件分析

隐写软件分析是一种特殊的专用隐写分析方法，其基本思想是利用隐写软件的缺陷和漏洞，分析软件遗留在隐写载体中的痕迹来进行隐写分析。此类分析方法的关键是获取隐写软件标识的鲁棒特征，即软件“特征码”，从而实现对隐写的百分之百检测，与基于病毒库的病毒查杀原理相似。此外，在完成隐写分析的同时还可以获取到所嵌入的文件类型、文件长度、加密方式以及哈希类型等辅助信息，并实现对秘密信息的提取。经过调研发现，现阶段可应用于音频的隐写软件 [35] 有 Xiao Steganography[36]、Invisible Secrets[37]、DeepSound[38]、SilentEye[39]、UnderMP3Cover[40] 以及 MP3Stegz[41] 等。

易小伟等人 [42] 实现了对 Invisible Secrets 软件的隐写分析，可完成包括 BMP、JPG 以及 WAV 等在内的多种文件格式的检测。王让定等人 [43] 提出一种针对 MP3Stegz[41] 的隐写分析方法。UnderMP3Cover 与其他隐写软件有所不同，UnderMP3Cover 是通过 LSB 算法进行连续或不连续的信息嵌入，嵌入完成后在文件中无“特征码”留存，张坚 [44] 等人提出一种改进的 RS 算法对其进行检测分析。汝学民等人 [45] 利用隐写工具的自相关特性，根据音频信号及其线性预测误差的统计特征进行分类器训练，从而实现对已知的三种隐写工具 (Hide4PGP[46]、Stegowav[47] 和 Steghide[48]) 的检测分析。此外，本论文也实现对了 Xiao Steganography、SilentEye 和 DeepSound 三种 WAV 隐写工具的分析，原理与具体实现已共享至 Github。

1.2.2 通用隐写分析方法

通用隐写分析是指在缺乏先验知识的条件下，判断载体中是否隐藏有秘密消息。相比于专用隐写分析具有更强的普适性，可以同时完成多种隐写算法，甚至是未知新算法的检测分析。然而，此类分析方法的检测准确率相比于专用隐写分析方法会有不同程度的下降，且无法实现嵌入消息的提取。通用隐写分析方法根据所分析音频类型的不同可分为时域隐写分析和频域隐写分析，其中时域隐写分析主要是以时域采样点的潜在特性为特征，而变换域隐写分析则是挖掘频域或 QMDCT 域的潜在规律。部分会在时域信号引入较大波动的 MP3 音频隐写算法也可以通过时域隐写分析特性进行检测。一般地，通用隐写分析算法大多基于机器学习方法进行设计，其关键技术分为数据预处理、特征提取与优选，分类器训练与模板选择三个部分，算法架构如图1.2所示。

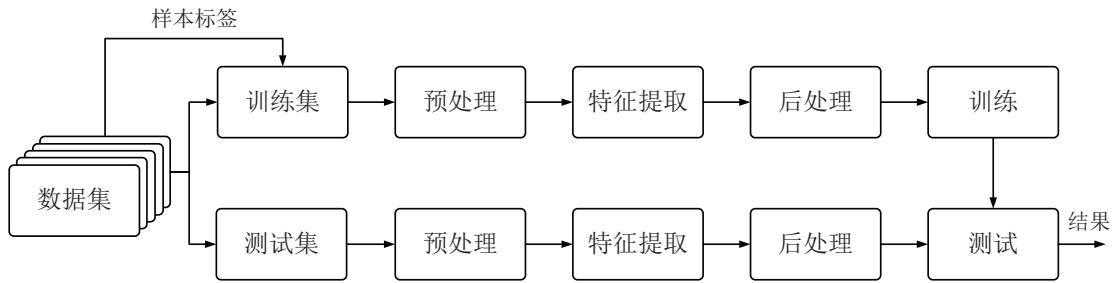


图 1.2 通用隐写分析方法一般架构

Figure 1.2 General architecture of universal steganalytic algorithms

1.2.2.1 时域通用隐写分析特征

汪云路等人在 [49] 中提出一种基于组合质量评估参数的音频盲分析方法。该方法首先进行各类音频质量评估参数的计算，如信噪比、对数似然比等，然后用多维特征选取的方法对组合质量评估参数的种类和数量进行筛选，使用支持向量机（Support Vector Machine, SVM）进行训练和分类，对 DSSS 和常用的变换域信息隐藏方法具有较好的检测效果。Geetha 等人 [50] 提出一种基于豪斯多夫距离高阶统计特性和决策树的隐写分析方法，并通过去噪的方式进行音频校正，可实现对扩频法、回声法等多种时域音频隐写算法的检测分析。Hamzeh 等人 [51] 基于人耳听觉模型提出一种基于反向梅尔频率倒谱（Reversed Mel-Frequency Cepstral Coefficients, RMFCC）统计特性的时域音频隐写分析算法，计算音频信号及其二阶差分信号 RMFCC 系数的均值、标准差、峰度和偏度等一阶统计量，以 SVM 分类器进行分类训练，可实现对 Hide4PGP、StegHide、DSSS 等常见时域隐写算法的检测分析，实验结果相较于传统的 D2-MFCC 特征可提升 17% 以上。Han 等人 [52] 提出一种基于线性预测系数的时域隐写分析方法，可用于 Xiao Steganography、StegoMagic[53] 等时域隐写算法的检测分析。

1.2.2.2 变换域通用隐写分析特征

Qiao 等人 [54] 根据隐写对 MP3 音频 QMDCT 系数域的影响，提出一种针对 MP3Stego 的通用隐写分析方法。提取 MP3 音频 QMDCT 系数各频率子带的均值、方差、峰度和偏度等统计矩特性以及二阶差分系数的累积 Markov 转移概率和邻域累积概率密度作为特征以实现对 MP3Stego 的隐写分析，同时使用广义高斯密度函数函数（Generalized Gaussian Density, GGD）对 QMDCT 系数分布进行建模，以音频复杂度进行度量，从而实现对音频的粗分类来提高检测的准确率。王让定等人 [55] 提出一种基于共生矩阵分析的 MP3 音频隐写检测方法，对 MP3

音频文件进行重编码，分别提取 MP3 音频与其校正音频的 QMDCT 系数矩阵，提取系数矩阵在水平方向、垂直方向、45 度角方向和 135 度角方向的共生矩阵，构造基于共生矩阵统计特性的特征，以实现对多类 MP3 音频隐写算法的检测分析，对基于窗口类型选择的隐写算法的检测准确率可达 99%，对 MP3Stego 算法的检测率为 95%，对基于 Huffman 码表索引选择的隐写算法的检测率为 86%。此后，Jin 等人 [56] 提出一种基于 MP3 音频帧内帧间相关性的隐写分析算法，以下简称 ADOTP (Abs Difference One-Step Transfer Probabilities)，该算法提取 MP3 音频的 QMDCT 系数，计算其一阶绝对值差分系数矩阵的帧内帧间 Markov 单步转移概率，并对所提特征进行优选，可以较好地对 MP3Stego 隐写进行检测分析。在隐写消息长度仅为 10Bytes、128kbps 比特率条件下对 MP3Stego 的准确率仍可达 92% 以上。Ren 等人 [57] 在帧内帧间 Markov 单步转移概率的基础上，引入邻域累积概率密度特征，可实现对 AAC 隐写音频的检测分析，以下简称 MDI2(Multiple Difference between Inter and Intra Frames)。该算法根据 AAC 音频音频帧类型、帧间关系、滤波器的阶数以及滤波方向对检测性能的影响进行特征优选，在 96kbps 和 128kbps 两种比特率下的检测准确率均可达 80% 以上，相比于之前的 AAC 音频隐写分析特征提升 10% 以上。

1.2.2.3 基于深度学习的通用隐写分析

相比于图像隐写分析，深度学习技术在音频隐写分析中的应用还相对较少。Paulin 等人 [58] 最先提出基于深度学习的音频隐写分析算法，算法以深度置信网络 (Deep Belief Network, DBN) [59] 为分类器完成对时域音频的隐写分析。首先提取音频信号的 MFCC 系数，然后将其输入至 DBN 网络进行训练，可实现包括 StegHide, Hide4PGP 和 FreqSteg[60] 在内的三类隐写算法的检测分析，性能相比 SVM 和高斯混合模型 (Gaussian Mixed Models, GMMs) 有较为明显的提升。此外，算法还讨论了 MFCC 系数的阶数对各种分类器检测准确率的影响。Chen 等人 [61] 提出一种基于卷积神经网络 (Convolutional Neural Networks, CNNs) 的时域音频隐写分析算法。以音频的时域采样点为输入数据，通过 CNN 网络实现特征的自动提取，对最低有效位匹配 (Least Significant Bits Match, LSBM) 隐写算法的检测准确提升至 88.85%，相比于此前提出的两种传统手工特征提升了 20-30%。然而，该方法仅在特定的数据集上有较好的表现，泛化能力较差。此后，Wang 等人 [62] 提出首个应用于 MP3 隐写分析的 CNN 网络，明显提升了包含 Huffman 码字映射 (Huffman Code Mapping, HCM) 在内的多种 MP3 隐写算法的检测准确率，对现阶段安全性最高的等长熵码字替换 (Adaptive MP3

Steganography Using Equal Length Entropy Codes Substitution, EECS) 算法的检测精度最高可达 90% 以上, 准确率平均提升 20%, 以下简称 WASDN (Wang Audio Steganalytic Deep Network)。此外, 部分基于深度学习网络也应用于其他压缩音频格式的隐写分析。Lin 等人 [63] 提出一种基于递归神经网络 (Recurrent Neural Network, RNN) 的互联网语音 (Voice of Internet Protocol, VoIP) 隐写分析网络 RNN-SM (RNN-based Steganalysis Model), 相比于传统的 VoIP 隐写分析算法准确率可提升 5% 以上。此后, Yang 等人 [64] 又提出一种基于 CNN 的 VoIP 隐写分析网络, 在低负载条件下, 准确率相比于 RNN-SM 提升 20% 以上。Ren 等人 [65] 也提出一种基于音频声谱图和残差网络的通用隐写分析方法, 可以同时应用于 AAC 及 MP3Stego 的隐写分析。

1.2.3 音频隐写分析未来发展趋势

根据上述介绍可以看出, 近年来国内外在音频隐写分析方面都做了较多的工作, 取得了一定的成绩, 但仍有以下两个方面值得研究和提升。

1. MP3 音频隐写分析方法研究。现阶段的音频隐写分析算法仍是以时域音频隐写的检测分析为主, 然而互联网中 70% 以上的音频资源是以 MP3 格式存在的。而且, 近年来提出的 MP3 音频自适应隐写算法也对 MP3 音频的隐写分析算法提出了新的要求。此外, AAC 音频格式作为 MP3 音频格式的扩展, 其隐写分析算法在一定上是通用的。因此, 进一步对适用于 MP3 音频的隐写分析方法进行研究是十分必要。

2. 基于深度学习的音频隐写分析方法研究。深度学习技术在图像分类、图像分割等计算机视觉领域已经取得了巨大的成功, 我们同样可以将这一先进技术应用于音频隐写与隐写分析技术方法的研究中。深度学习网络可以更好地且更为智能地学习到隐写导致的潜在特性变化, 不仅有助于降低特征设计的难度, 还可以有效提升算法的性能。

1.3 本文主要研究内容

本论文针对 MP3 音频隐写分析中的关键技术进行研究, 主要内容包含以下三个方面:

1. 基于 MP3 音频编码特性的隐写分析算法研究。QMDCT 系数是 MP3 音频中十分重要的参数之一, 以往的 MP3 隐写分析算法在基于 QMDCT 系数矩阵进行特征设计时并未过多考虑 MP3 音频自身的编码特性的影响, 而是直接将已成

功应用在图像隐写分析中的经验移植到 MP3 音频隐写分析的研究中。值得注意的是，在 MP3 音频编码过程中，大值区的 Huffman 码字对应两个 QMDCT 系数，小值区的 Huffman 码字对应四个 QMDCT 系数，这与 JPEG 图像隐写分析中的相位保持（Phase-Aware）特性十分相似。因此，可以基于 MP3 音频这一编码特性对 MP3 音频隐写分析算法进行改进。

2. 基于深度学习的 MP3 音频隐写分析算法研究。MP3 音频隐写分析的本质即为 MP3 音频的二分类，深度学习技术的兴起和发展为 MP3 音频隐写分析带来了福音，现阶段的 MP3 音频隐写分析算法仍以手工特征设计为主，然而手工特征设计则需要极强的理论背景和先验知识，且难以实现对隐写产生的潜在规律变化进行较为全面地捕获。为此，以提升 MP3 音频隐写分析算法的检测性能为目标，研究不同结构的神经网络对 MP3 隐写分析的影响，设计出性能优越、通用性强且实用的隐写分析网络。

3. MP3 音频盲隐写分析研究。真实互联网环境中的 MP3 音频是复杂多样的，不同的数据源、音频比特率、隐写算法、隐写负载率等均会给 MP3 音频隐写分析模型的性能产生不同程度的影响。基于真实环境下 MP3 音频隐写分析的要求，设计合理的解决方案，设计出实用性强的 MP3 音频隐写分析算法。

1.4 本文组织结构

本论文共分为六章，具体内容安排如下：

第一章为绪论，首先介绍了本论文的研究背景及研究意义，简要探讨了音频隐写分析的发展现状，分析了音频隐写分析当下的主要研究方向，并给出本论文的主要研究内容，最后列出全文的组织架构。

第二章分别介绍了 MP3 音频的编解码原理、三种经典的 MP3 音频隐写算法、用于 MP3 隐写与隐写分析算法效果验证的数据集构建以及两种现阶段检测性能较好的压缩域音频通用隐写分析算法。

第三章介绍了一种基于 QMDCT 系数矩阵多尺度相关性度量的 MP3 音频隐写分析方法，提出了一个适用于 MP3 音频隐写分析的富高通滤波模型，并介绍了一种基于 2×2 和 4×4 码字保持的多尺度相关性度量模型。此外，还讨论了隐写音频段长度对分析算法性能的影响，确定了算法可实现的最小隐写音频段的定位粒度。最后通过对比实验分析该分析方法的有效性。

第四章介绍一种基于 QMDCT 系数矩阵和全卷积神经网络的隐写分析方法。分别讨论了输入数据类型、神经网络类型以及网络结构等因素对 MP3 音频隐写

分析网络性能的影响。并提出一种面向尺寸失配的解决方案和一种基于迁移学习的低隐写负载隐写样本分析方案。最后通过对比实验说明分析方法的有效性。

第五章介绍了一种基于 QMDCT 系数矩阵与多尺度卷积神经网络的 MP3 音频隐写分析网络。通过引入多尺度卷积模块、卷积核分解以及跳跃连接优化了网络结构，加快了网络的收敛速度，并在网络性能不降低的前提下实现了网络参数的轻量化。最后通过对比实验说明分析方法的有效性。

第六章是总结与展望，对本论文的工作做了最后的梳理和总结，并对 MP3 音频隐写与隐写分析未来的研究方向做了进一步展望。

第2章 相关研究概述

MP3 的消息嵌入与提取离不开 MP3 编解码器，MP3 音频隐写分析同样如此。为了便于后续内容的理解与展开，本章将主要介绍 MP3 编解码原理、当前最为经典的三种 MP3 音频隐写算法以及写阶段性能最为优越的两种压缩域音频隐写分析算法。同时，为了快速有效地推动音频隐写与隐写分析技术的发展，解决基准数据集缺失的问题，本论文还构建了一个基础音频数据集，本章还将对这一数据集进行简要的介绍。

2.1 MP3 编解码原理介绍

2.1.1 MP3 编码原理

根据 ISO/IEC 11172-3 标准 [66]，MP3 编码主要可分为 4 个模块，多相分析滤波、心理声学模型、位分配循环、比特流格式化，编码流程如图2.1所示。编码后的 MP3 音频文件由音频帧（Frame）组成，每帧包含两个颗粒（Granule），每个颗粒内有一个或多个声道（Channel），每个声道内固定有 576 个 QMDCT 系数。由于双声道的 MP3 音频在网络中传播最为广泛，为了使得实验环境更为真实，本论文所分析的 MP3 音频文件均为双声道，即 Channel = 2。

1. 多相分析滤波：对原始时域采样值进行滤波并将其转换为 32 个子带信号，通过 MDCT 变换，在每个子带内划分出 18 个次子带，由此得到 576 (32×18) 个系数的频域信号。
2. 心理声学模型：根据心理声学模型 II 计算信掩比（Signal-to-Mask Ratio, SMR），从而实现对音频压缩比率的控制，保持听觉不失真。

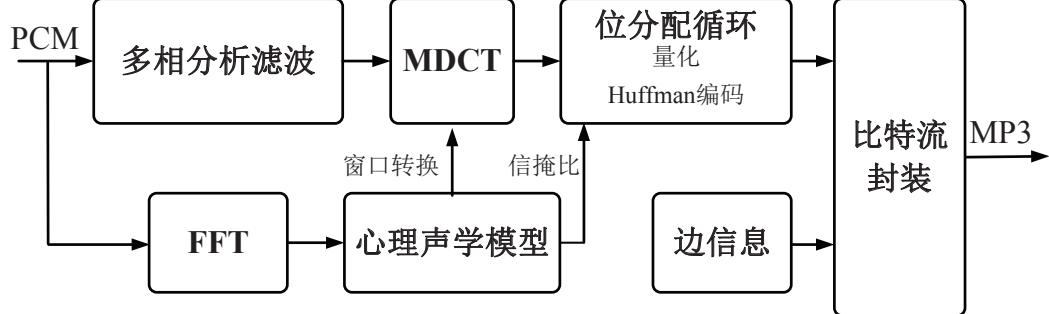


图 2.1 MP3 编码原理示意图

Figure 2.1 Diagram of MP3 encoding

3. 位分配循环：通过量化 MDCT 系数和 Huffman 编码得到二进制比特流。其中，外循环通过掩蔽阈值控制量化噪声的大小，从而得到量化系数；内循环通过 Huffman 码表实现最优的 Huffman 码字的选取，从而完成对 QMDCT 系数的编码。

4. 比特流格式化：将编码后的音频信号与边信息封装为 MP3 码流输出。

QMDCT 系数和 Huffman 码字是 MP3 音频中最为重要的两个参数。编码后的 MP3 音频文件中，90% 以上的内容为 Huffman 码字，Huffman 码字和 QMDCT 系数之间具有完备的映射关系。由于 Huffman 编码为无损编码，因此对 QMDCT 系数的改动和对 Huffman 码字的改动在本质上是相同的。在 MP3 音频帧中，每个声道内包含 576 个 QMDCT 系数，根据频率大小由低到高可依次为大值区 (Big_Value)、小值区 (Count1) 和零区 (Zero)，如图2.2所示。

1. 大值区系数较大，每两个 QMDCT 系数 $< x, y >$ 对应一个 Huffman 码字，其码流结构如图2.3所示。如果系数值为零，则不指定符号位。如果系数值超过 15，则超出的部分由 linbits 位表示。
2. 小值区系数为 -1, 0, 1，每四个 QMDCT 系数 $< x, y, v, w >$ 对应一个 Huffman 码字，码流结构如图2.4所示。小值区无 linbits 位，其余部分与大值区完全相同。
3. 零区的系数全部为 0，无需编码。

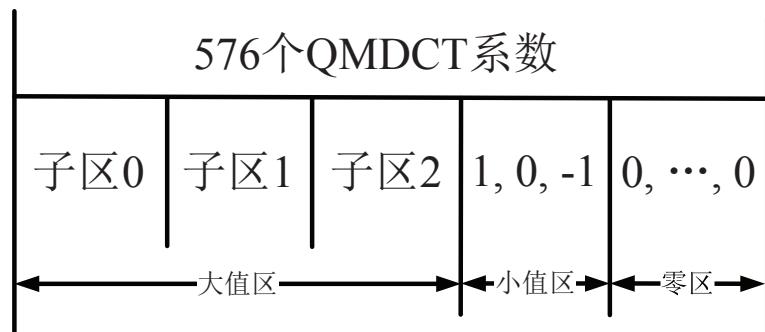


图 2.2 QMDCT 系数结构图

Figure 2.2 Structure of QMDCT coefficients

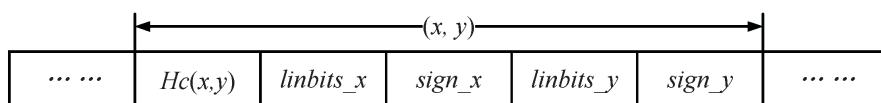


图 2.3 大值区 QMDCT 系数与 Huffman 码字映射关系图

Figure 2.3 Diagram of the relationship between QMDCT coefficients in Big-Value region and Huffman codewords

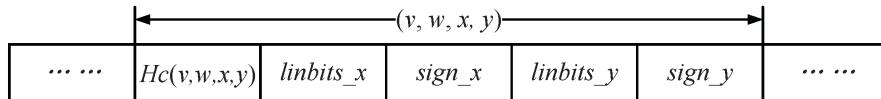


图 2.4 小值区 QMDCT 系数与 Huffman 码字映射关系图

Figure 2.4 Diagram of the relationship between QMDCT coefficients in Count1 region and Huffman codewords

2.1.2 MP3 解码原理

MP3 音频隐写算法基于 MP3 编码得以实现，而与之对应的 MP3 音频隐写分析工作则更多地依赖于 MP3 解码。通过 MP3 解码器可以实现对其各类重要参数的提取，同时也可以判断隐写消息有无成功嵌入。

MP3 解码是其编码的逆过程，解码流程如图2.5。由于 MP3 编码得到的音频文件存储的是 MP3 音频的编码信息和 Huffman 码字流，如果没有对 MP3 解码则无法在设备上对其进行直接进行播放。在 MP3 解码过程中，解码器根据音频帧的同步标识逐帧读取 MP3 数据流，根据存储的编码信息将 Huffman 码字重新还原为 QMDCT 系数，再通过反量化得到 MDCT 系数。随后，分别对两个声道的 MDCT 系数进行立体声处理，按声道重排序，再进行反修正离散余弦变换 (Inverse Modified Discrete Cosine Transform, IMDCT)，由此得到时域信号。最终，通过子带合成得到可以用于播放的时域音频。Huffman 编码为熵编码，根据熵原理此类编码方式不会丢失任何信息。因此，在 MP3 解码过程中所提取的 QMDCT 系数与 MP3 编码过程中得到的 QMDCT 系数完全相同。然而，由于 MP3 编码是有损压缩，解码获得的 WAV 音频与编码前的 WAV 音频是不完全相同的。

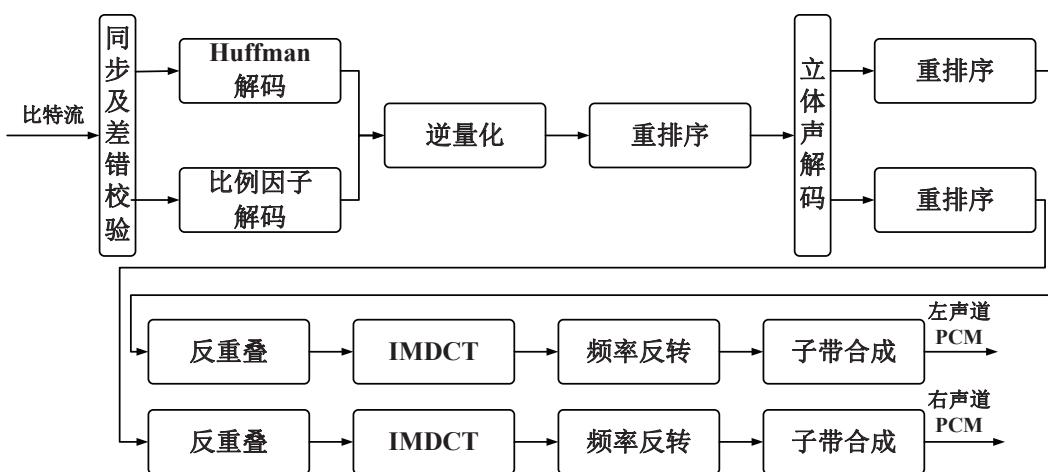


图 2.5 MP3 解码原理示意图

Figure 2.5 Diagram of MP3 decoding

2.2 MP3 隐写算法介绍

不同于 WAV 音频, MP3 音频具有多类数据空间, 如时域、频域、QMDCT 系数域、码字域等, MP3 音频隐写算法的具体实现会受到 MP3 编码器的制约。在 MP3 编码过程中, 需要对量化后的变换样值进行比特分配, 以使得整个量化块最小, 从而实现数据压缩, 而压缩编码前后的数据变换均无损。因此, 根据秘密消息的嵌入位置与压缩编码的关系, MP3 音频隐写算法可分为前置式隐写、内置式隐写和后置式隐写, 如图2.6所示。其中, (1) 前置式隐写是将秘密消息嵌入到压缩编码前的时域或频域信号中, 虽然在设计时可以借鉴很多成熟的时域音频隐写算法, 但此类算法的嵌入容量较小, 同时还会降低算法的不可感知性, 实用性较差。(2) 后置式隐写是将秘密消息直接嵌入至压缩比特流或者 Huffman 码字中, 避免了编码器对消息嵌入的影响。但是由于 MP3 是一种压缩率较高的音频格式, 且 MP3 音频具有较为复杂的码流结构, 对码流结构的随意修改将会导致 MP3 音频解码失败, 因此需要在算法设计时通过一定的补偿手段以保证算法的安全性, 此外算法通常可以获得更大的隐写容量; (3) 内置式隐写算法将信息的嵌入过程与 MP3 压缩编码相结合, 很多图像隐写算法的实现就是利用了这一思想, 如 Jsteg[67]、F5[68] 等, 此类算法能够与压缩编码较好地兼容, 可以获得较好的不可感知性, 但嵌入容量会受到限制。现阶段, 主流的 MP3 音频隐写算法有 MP3Stego, 基于哈夫曼码字映射的 MP3 音频隐写算法 (Huffman Code Mapping, HCM) 和基于等长熵码字替换的 MP3 隐写算法 (Adaptive MP3 Steganography Using Equal Length Entropy Codes Substitution, EECS) 等。MP3Stego 是最早被提出的内置式 MP3 音频隐写算法; HCM 和 EECS 算法是两种基于 Huffman 码字的后置式隐写算法, 隐写容量相比于此前的 MP3 音频隐写算法有明显的提升; 此外, EECS 算法是首个 MP3 自适应隐写算法, 同时兼顾了安全性和嵌入容量, 是现阶段最先进的 MP3 音频隐写算法之一。

2.2.1 MP3Stego 隐写算法

MP3Stego 是一款开源的 MP3 隐写软件, 也是最早的 MP3 隐写算法。MP3Stego 是在 8Hz-MP3[69] 编码器的基础上开发完成的, 根据不同的参数选择即可完成正常 MP3 和隐写 MP3 的制备。MP3Stego 隐写是在 MP3 编码的内循环中完成的, 根据 part_2_3_length 的奇偶性进行秘密消息的嵌入, 当 part_2_3_length 为奇数时表示嵌入信息比特 1, 当 part_2_3_length 为偶数时表示嵌入消息比特 0。在消息嵌入的过程中, 如果量化编码后的 part_2_3_length 不满足嵌入要求, 则需重新

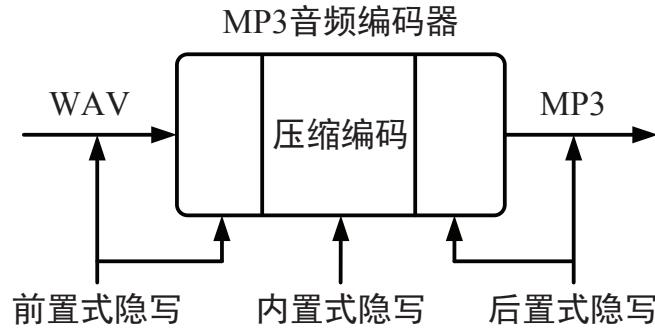


图 2.6 MP3 音频隐写算法分类示意图

Figure 2.6 Classification of MP3 audio steganographic algorithms

调节量化误差。此外，如果 part_2_3_length 不满足量化编码要求，需要先对量化误差进行调节，再重新进行量化编码。使用不同的编码器对 WAV 音频进行压缩会生成的不完全相同的 MP3 文件，因此，在 MP3Stego 的样本制备时，正常载体与隐写载体都需要通过 MP3Stego 编码器进行制备。

由于 MP3Stego 在制备低比特率 MP3 音频时容易崩溃，且在隐写过程中会对待嵌入消息进行加密和压缩，因此本论文对 MP3Stego 编码器做了一定程度的修改。在不影响秘密信息嵌入的前提下移除了加密和压缩模块，并修正了部分漏洞，降低了软件运行报错的风险。

2.2.2 HCM 隐写算法

由上文介绍可知，Huffman 码字在 MP3 压缩比特流中所占比例达到 90% 以上，是非常理想的隐写空间。当隐写前后的 Huffman 码字结构保持一致时，即隐写前后码字长度、符号位以及 Linbits 位均相同，就不会破坏 MP3 的码流结构。HCM 隐写算法就是通过修改 Huffman 码字实现秘密消息嵌入的，具有容量大、隐蔽性高等优点。

在进行信息嵌入前，首先要对 MP3 音频编码中 34 个 Huffman 码表内的码字进行分类。对于一个给定 Huffman 码表 T ，对码字分类的过程共分为三步：(1) 将码表中的码字分类两类，一类是可以被替换的码字 C_V ，一类是不能被替换的码字 C_N ，并在 C_V 中，将码字长度及符号位都相同的码字归位一类。这一步需要通过手工统计完成。(2) 如果码表 T 中共有 M 类码字，各类码字的个数分别为 N_1, N_2, \dots, N_M ，各个码字分别记为 $\{c_{1,0}, c_{1,1}, \dots, c_{1,N_1}\}, \{c_{2,0}, c_{2,1}, \dots, c_{2,N_2}\}, \dots, \{c_{M,0}, c_{M,1}, \dots, c_{M,N_M}\}$ ，并且每一类的码长和符号位分别记为 $\{l_1, s_1\}, \{l_2, s_2\}, \dots, \{l_M, s_M\}$ 。(3) 对每一类码字按照其对应的 QMDCT 系数值大小进行字典排

序，使得修改前后系数值变化的幅度尽可能小，保证隐写算法的安全性。

在码字分类完成后即可进行信息嵌入，具体步骤如下：

1. 对 MP3 音频以帧为单位进行部分解码，跳过所有 C_N 类码字，直到两个码字均为 C_V 类码字。若解码得到的两个码字分别为 $\{c_{i,n_0}, c_{j,n_1}\}$ ，则根据码字所属类别中所包含的码字个数来确定其基底 $\{b_0, b_1\}$ ，由此确定出当前可隐写的位数 R 。其中， $R = \lfloor \log_2^{b_0 \times b_1} \rfloor$ ，对应的十进制值为 w_R 。
2. 计算两个码字所需隐藏的信息 $\{d_0, d_1\}$ 。

$$w_R = d_1 + d_0 \times b_1 \leq b_0 \times b_1 \quad (2.1)$$

$$d_1 = w_R \bmod b_1 \quad (2.2)$$

$$d_0 = ((w_R - d_1)/b_1) \bmod b_0 \quad (2.3)$$

3. 将 $\{d_0, d_1\}$ 以替换的方式隐藏至当前码字对 $\{c_{i,n_0}, c_{j,n_1}\}$ 。

$$v = (n_0 \bmod b_0) - d_0 \quad (2.4)$$

$$c_{in'_0} = \begin{cases} c_{in_0}, & n_0 \bmod b_0 = d_0 \\ c_{i(n_0-v)}, & \text{其他} \end{cases} \quad (2.5)$$

4. 重复步骤 1 - 3 完成所有秘密消息的嵌入。

2.2.3 EECS 隐写算法

EECS 隐写算法是首个自适应 MP3 音频隐写算法，算法基于等长熵码字替换实现秘密消息的嵌入，在嵌入过程中引入了校验子格编码（Syndrome-Trellis Code, STC）[70] 和代价函数构造，大大提升了隐写算法的安全性。首先，算法根据心理声学模型以及映射关系计算待嵌入信息的代价。然后，再根据映射关系将原始的 Huffman 码字映射为二进制比特流。其次，再利用 STC 编码将待嵌信息嵌入到二进制比特流，得到载密的二进制比特流。最后，根据逆映射关系将载密的二进制比特流映射为载密 Huffman 码字。相比于之前 MP3 音频隐写算法，EECS 算法很好地实现了隐写容量和安全性的平衡。尤其是代价函数设计，使得秘密消息的嵌入更为自适应。

在隐写过程中，音频失真主要有两个来源：QMDCT 系数的幅度扰动和心理声学模型。这两个因素会造成音频质量的下降和 Huffman 码字统计特性的变化。人耳对不同频带的声音具有不同的敏感度，通过大量实验得到的绝对听觉阈值描述了在静音环境中，一个纯音需要具备多少能量才能被人耳听见。人耳对各个

频带音频信号的敏感度可以用绝对听觉阈值曲线来表达。绝对听觉阈值曲线可以表示为：

$$T(f) = 3.64 \times \left(\frac{f}{1000}\right)^{-0.8} - 0.65 \times e^{-0.6 \times (\frac{f}{1000} - 3.3)^2} + 10^{-3} \times \left(\frac{f}{1000}\right)^4 \quad (2.6)$$

其中， f 为频率， $T(f)$ 为对应的绝对听觉阈值。人耳对 $T(f)$ 小的音频信号更为敏感，因此，EECS 算法在隐写时为这一频段赋予较大的代价，以减少对低频段信号的修改。同时，为了减少因 QMDCT 系数幅值扰动带来的失真，算法通过曼哈顿距离来度量两个 Huffman 码字之间的距离。

$$d_i^c = |x'_i - x_i| + |y'_i - y_i| \quad (2.7)$$

其中， (x_i, y_i) 和 (x'_i, y'_i) 分别为大值区的 Huffman 码字替换前后对应的 QMDCT 系数对。

最终，EECS 算法的代价函数构造为：

$$\rho_i = \frac{1}{\log_2\left(\frac{2}{t_{2i}+t_{2i+1}} + \sigma\right)} \times d_i^c = \frac{|x'_i - x_i| + |y'_i - y_i|}{\log_2\left(\frac{2}{t_{2i}+t_{2i+1}} + \sigma\right)} \quad (2.8)$$

其中， i 是 Huffman 码字 h_i 在码流中的索引值， t_{2i} 和 t_{2i+1} 是第 $2i$ 和 $2i+1$ 个频率线处的绝对听觉阈值， σ 为一常数，起平滑作用，避免 \log 运算错误。

2.3 数据集构建

由于音频隐写与隐写分析算法的发展相对缓慢，目前还没有类似于图像隐写分析中 BOSS (Break Our Steganographic System) 图像库 [71] 一类的公共标准数据集。因此，为了更好地推动 MP3 音频隐写及隐写分析算法研究的发展，同时为了更好地对本论文所提的各类隐写分析算法进行全面合理的评估，本论文构建了一个基本音频数据集，分别包含 MP3Stego、HCM 和 EECS 等三类 MP3 音频隐写算法；128kbps、192kbps、256kbps、320kbps 等四种常用 MP3 音频比特率以及多种隐写负载率，各种参数下分别有正常/隐写音频 33038 对。所有正常及隐写的 MP3 音频均是由 WAV 音频编码得到，其中 MP3Stego 隐写样本是通过 MP3Stego 编码器编码得到，HCM 和 EECS 隐写工具是基于 lame[72] 二次开发得到。所有的原始 WAV 音频均通过爬虫技术从互联网中爬取得到，包含多种风格，能够实现对现有 MP3 音频隐写分析算法客观、全面、真实的评价，更为详细的信息如表A.1所示。由于 MP3 音频为时序信号，无法通过旋转、反转等常用方式进行图像数据增强手段得以扩充。因此，本论文构建的数据集仍在进一步扩大与完善，以更好地满足基于深度学习的网络构建对海量数据建模的需要。

对于本论文分析的三类 MP3 音频隐写算法，由于 MP3Stego 和 HCM 均为非自适应隐写算法，因此使用隐写音频帧率（Steganographic Frames Rate, SFR）作为其隐写负载率（Steganographic Payloads Rate, SPR）的度量指标，取值依次为 0.1, 0.3, 0.5, 0.8 和 1.0。 $SFR = N_e/N_t$ 。 N_e 和 N_t 分别为隐写音频帧的数量与待检测的音频帧总数。EECS 为自适应隐写算法，在代价函数确定条件下，通过 STC 编码实现对隐写嵌入率大小的调节。 w 和 H 分别为 STC 编码奇偶校验矩阵的宽度和高度。其中， $w = 1/\alpha$ ， α 为 STC 编码中定义的负载率。因此，本论文使用 w 作为 EECS 算法的隐写负载率，取值依次为 2, 3, 4, 5, 6, h 固定为 7。

此外，由于音频载体的原因，MP3 音频隐写算法的相对嵌入率（Relative Embedding Rate, RER）并不能像图像载体那样很好地被固定至诸如 0.1, 0.2 这一类相对常见的数值。但是为了能够更好地与图像隐写分析中的常用指标对应，本论文将上述隐写算法的 SPR 分别转换至隐写速率和 RER，对照结果如表B.1, B.2, B.3所示。

2.4 MP3 音频隐写分析算法介绍

ADOTP 和 MDI2 隐写分析算法是现阶段两种性能较好、适用于压缩域音频隐写分析的算法。这两种算法的基本原理相同，均以 QMDCT 系数矩阵的相关性变化为特征实现 MP3 音频的隐写分析。由于本论文所提出的三种隐写分析算法同样是以 QMDCT 系数矩阵作为分析对象，因此为了便于后续内容的理解，本节将首先介绍 QMDCT 系数矩阵。

2.4.1 QMDCT 系数矩阵

MP3 音频是一维信号，所对应提取的 QMDCT 系数也是一维的。通过构造 QMDCT 系数矩阵，实现了从一维信号向二维信号的转换，图像处理与分析中常用的特征工程手段便可引入至 MP3 音频隐写分析中，拼接形式如图2.7所示。每个音频帧的末尾都会有不同数目的零值系数，这部分系数不会被编码，也不会被嵌入任何隐写消息。音频比特率的不同，每帧所包含的零值系数个数也是不相同的。一般情况下，随着比特率的增加，零区的起始索引值也会逐渐增加。不同的 MP3 音频文件，其零区的起始索引也不完全相同。为此，本论文从数据集随机取出 5000 个 MP3 音频进行分析，四种常用音频比特率所对应的零区起始索引如表2.1所示。虽然每个通道内的大值区、小值区系数个数不完全相同，但是大部分 MP3 音频隐写算法倾向于在大值区嵌入秘密消息，所以只要能充分保留大值

表 2.1 不同音频比特率的 QMDCT 系数零区起始索引平均值

Table 2.1 Mean value of start index of coefficients in Zero region with different bitrates

比特率 / kbps	128	192	256	320
零区起始索引	410	450	500	560

区的 QMDCT 系数就可以较好地保持检测精度。此外，为了减少不必要的运算，同时能够不减少正常音频和隐写音频 QMDCT 系数矩阵之间的差异，本论文用于隐写分析的 QMDCT 系数矩阵尺寸最终选定为 200×400 ，表示为

$$M = \begin{pmatrix} Q_{1,1} & Q_{1,j} & Q_{1,400} \\ \ddots & \ddots & \ddots \\ Q_{i,1} & Q_{i,j} & Q_{i,400} \\ \ddots & \ddots & \ddots \\ Q_{200,1} & Q_{200,j} & Q_{200,400} \end{pmatrix} \quad (2.9)$$

其中，变量 $i \in \{1, 2, \dots, 200\}$ 表示所选择用于隐写分析的通道总数，其范围取决于所选音频帧的个数，如果最小分析单元为 N 帧，则 $i \in [0, 4N]$ 。变量 $j \in \{1, 2, \dots, 400\}$ 表示一个通道内的 QMDCT 系数的个数。本论文出于对精度和计算效率的考虑，选取 50 帧为一个基本检测单元。

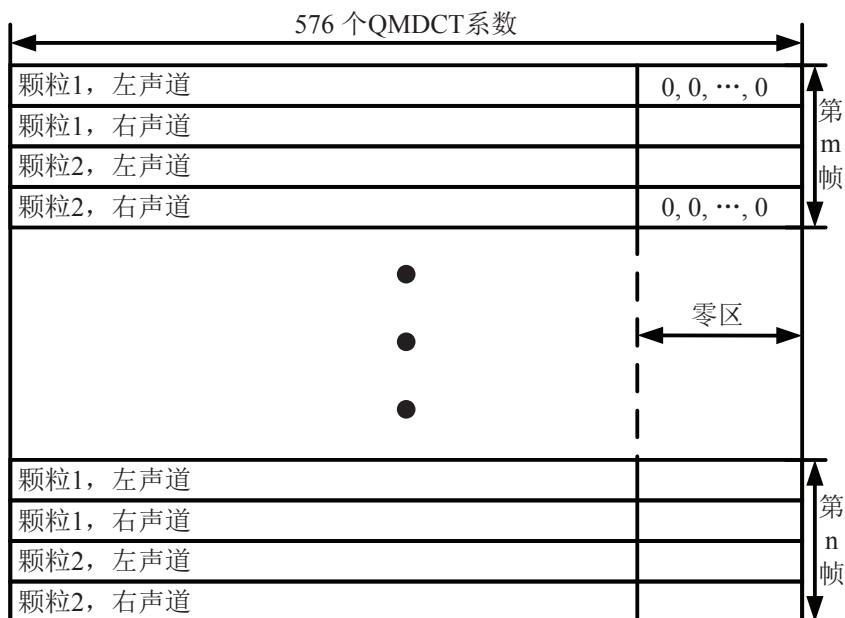


图 2.7 QMDCT 系数矩阵构造示意图

Figure 2.7 Diagram of QMDCT coefficients matrix

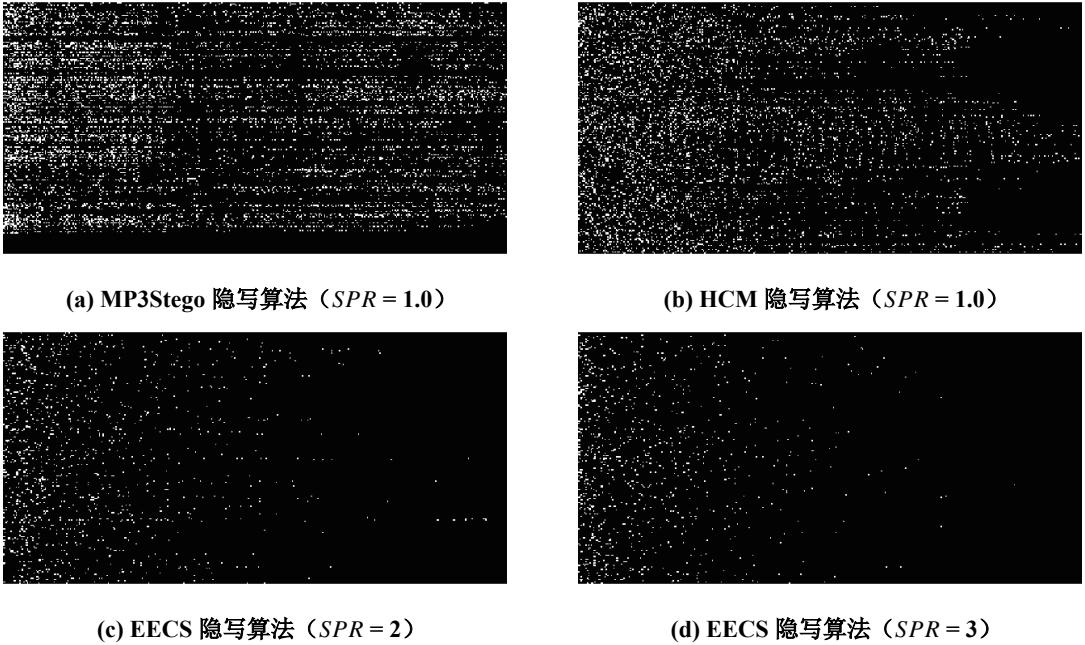


图 2.8 正常音频与隐写音频的 QMDCT 系数矩阵差异图 (128kbps)

**Figure 2.8 Diagram of the difference in QMDCT coefficients matrices
between cover and stego audios (128kbps)**

图2.8中的白色散点为正常音频和隐写音频 QMDCT 系数矩阵中元素差值非零系数所在的位置，即为在隐写中被修改的系数。从图中可以看出，隐写算法对 QMDCT 系数矩阵产生了较为明显的影响，且不同的隐写算法对系数矩阵的影响有较大差异。

2.4.2 ADOTP 隐写分析算法

ADOTP 隐写分析算法的流程如图2.9所示，该算法被提出时是用于 MP3Stego 的隐写分析，首先以一阶绝对值差分 $H(\cdot)$ 对所提取的 QMDCT 系数矩阵进行高通滤波，然后提取差分矩阵的 Markov 单步转移概率 $P_{(\cdot)}$ 作为分类特征，特征优选后被输入到分类器中进行训练，以实现对 MP3Stego 算法的隐写分析。ADOTP 隐写分析算法在嵌入消息长度仅为 10 字节的情况下，对 128kbps 及以上比特率的 MP3 音频的检测准确率仍可达 90% 以上。然而，该算法最大的不足在于高通滤波器 $H(\cdot)$ 的选择，对于以符号位翻转类隐写算法，ADOTP 将失效。本论文在设计过程中也考虑到了这一点，因而避免只使用绝对值差分滤波器。

$$H(\cdot) = |Q_{i+1,j}| - |Q_{i,j}| \quad (2.10)$$

Markov 单步转移概率的计算公式如下：

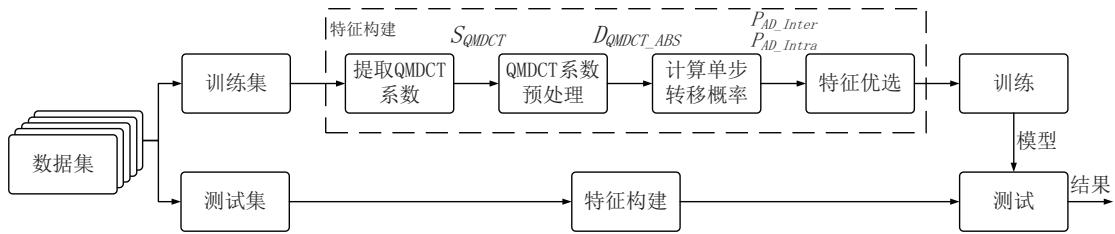


图 2.9 ADOTP 隐写分析算法流程图

Figure 2.9 Flowchart of steganalytic algorithm ADOTP

帧内:

$$P_h(m, n) = \sum_i \sum_j \frac{\delta(Q_{i,j} = x, Q_{i,j+1} = y)}{\delta(Q_{i,j} = x)} \quad (2.11)$$

帧间:

$$P_v(m, n) = \sum_i \sum_j \frac{\delta(Q_{i,j} = x, Q_{i+1,j} = y)}{\delta(Q_{i,j} = x)} \quad (2.12)$$

其中, m, n 为单步转移概率矩阵索引。 $\delta(X = x, Y = y) = \begin{cases} 1, & X = x, Y = y, \\ 0, & \text{其他} \end{cases}$, $x, y \in [-T, T]$, T 是截断阈值。

2.4.3 MDI2 隐写分析算法

MDI2 隐写分析算法的流程如图2.10所示, 该算法最早应用于 AAC 音频隐写算法的检测分析。由于 AAC 作为 MP3 格式的扩展, 与 MP3 具有较为相似的结构, 因此该算法也可直接应用于 MP3 音频隐写分析。首先, 提取音频的 QMDCT 系数矩阵, 分别计算其一阶行差分矩阵、一阶列差分矩阵、二阶行差分矩阵及二阶列差分矩阵。然后, 计算各个差分矩阵的 Markov 单步转移概率 $P_{(.)}$ 以及邻域累积概率密度 $INJ_{(.)}$ 。根据 AAC 音频音频帧类型、帧间关系、滤波器的阶数以及滤波方向对检测性能的影响进行特征优选, 并将优选后的特征输入至分类器中进行训练, 由此实现对 AAC 音频的隐写分析。该分析算法对多种 AAC 隐写算法 [73–75] 都具有很好的检测效果, 检测率整体分布于 85% 以上。

Markov 单步转移概率的计算公式同上, 邻域累积概率密度的计算公式为:

帧内:

$$INJ_h(m, n) = \sum_i \sum_j \frac{\delta(Q_{i,j} = x, Q_{i,j+1} = y)}{N_i \times (N_j - 1)} \quad (2.13)$$

帧间:

$$INJ_v(m, n) = \sum_i \sum_j \frac{\delta(Q_{i,j} = x, Q_{i+1,j} = y)}{(N_i - 1) \times N_j} \quad (2.14)$$

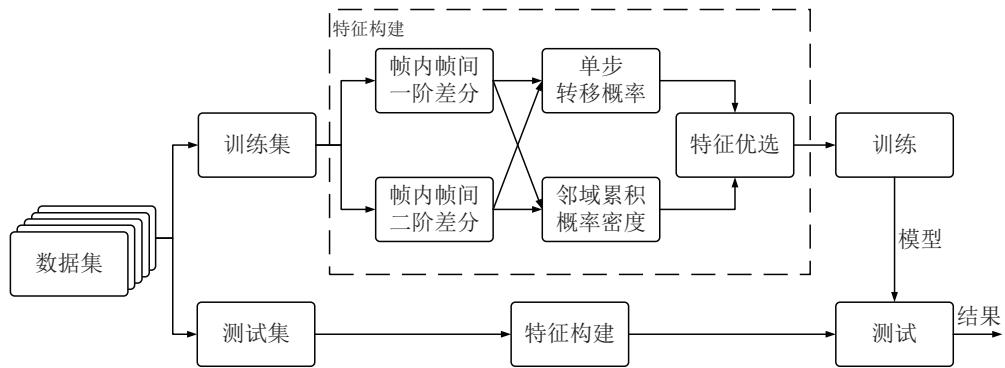


图 2.10 MDI2 隐写分析算法流程图

Figure 2.10 Flowchart of steganalytic algorithm MDI2

其中, N_i 和 N_j 是差分矩阵的总行数和总列数, m, n 为概率密度矩阵索引 $\delta(X = x, Y = y) = \begin{cases} 1, & X = x, Y = y \\ 0, & \text{其他} \end{cases}$ and $x, y \in [-T, T]$, T 是截断阈值。

第3章 基于多尺度相关性度量的MP3音频隐写分析研究

3.1 引言

特征表示是机器学习永恒的主题，隐写分析作为分类任务之一其目前最为重要的问题也在于特征表示，良好的特征表示将对算法的性能起到至关重要的作用。现阶段，MP3音频隐写算法在设计过程中通过与MP3编码原理相结合可以获得较高的隐蔽性，能够很好地保持音频波形、一阶统计特性、频谱图乃至QMDCT系数分布直方图等无明显失真。然而，在当前的隐写算法模型下，二阶及以上的统计度量值仍较难保持。而且，音频是一种时序信号，其编码是逐帧进行的，全局相关性保持变得更为困难。因此，基于MP3音频相关性度量的隐写分析特征设计是一种良好的手工特征设计方法。

手工特征的提取过程一般分为三个步骤：数据预处理、特征提取与优选、分类器训练。提取MP3音频的QMDCT系数矩阵、声谱图或对其进行小波变换，经过预处理后计算其Markov转移概率、邻域累计概率密度、共现矩阵或其他有效相关性度量特性，特征优选后输入至LIBSVM[76]或集成分类器[77]中进行训练和测试，由此实现对MP3音频的隐写分析。基于手工特征设计的MP3音频隐写分析可以在有限的存储和计算资源条件下完成MP3音频的检测分析，具有实用性强、计算效率高、数据依赖性低等优点，仍值得进一步研究和探索。此外，手工特征设计过程中成功的经验也可应用到基于深度学习的MP3音频隐写分析网络设计中。

本章的主要贡献在于提出一种基于QMDCT系数矩阵多尺度相关性度量的隐写分析方法，以下简称MSC（Multiscale Correlations）。算法结合了MP3编码原理与各类MP3隐写算法原理，有效提升了对MP3音频进行隐写分析的精度。此外，富模型[78]已广泛应用于图像隐写分析中，并显著提升了隐写分析算法的性能，受此启发，本章还提出一个适用于MP3音频隐写分析的富高通滤波器模型。该模型用于对QMDCT系数矩阵的预处理，从多个角度“放大”隐写信号的痕迹，提升算法对隐写信号的敏感性，有效提升了算法的检测效果。最后，本章还讨论了待检测音频段长度对隐写分析算法性能的影响，得到了算法可实现的隐写定位粒度。

本章后续内容组织安排如下：3.2节介绍算法设计原理，3.3节是实验设计与分析，3.4节是对本章工作的小结。

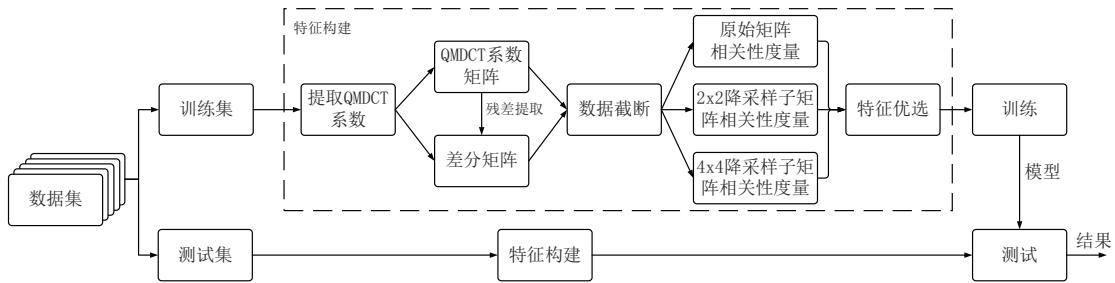


图 3.1 MSC 算法流程图

Figure 3.1 Flowchart of steganalytic algorithm MSC

3.2 算法设计

本章提出一种基于 QMDCT 系数矩阵多尺度相关性度量的 MP3 音频隐写分析算法，算法流程如图3.1所示。首先，提取 MP3 音频的 QMDCT 系数矩阵作为分析数据，矩阵维度为 200×400 ，经过富高通滤波模型进行残差提取后对其进行数据截断。然后，对截断后的矩阵进行 2×2 及 4×4 码字保持降采样，分别计算各矩阵及其子矩阵在水平方向和竖直方向上的 Markov 单步转移概率矩阵，并以此作为 MP3 音频隐写分析的特征。对截断阈值 T 和富高通滤波组合模式进行优选以降低特征维度并提升运算效率。最后，将优选的特征输入到集成分类器中完成训练和测试。

3.2.1 音频富高通滤波模型

相比于 MP3 音频信号，隐写信号较为微弱，可以看作是一种加性高频噪声。音频富高通滤波模型就是基于广义加性噪声模型建立的。

$$S_{i,j} = C_{i,j} + M_{i,j} \quad (3.1)$$

其中， $C_{i,j}$ 和 $S_{i,j}$ 分别表示正常音频和隐写音频的 QMDCT 系数矩阵， $M_{i,j}$ 表示由于秘密信息嵌入引起的扰动。 i 和 j 表示系数矩阵的横纵方向上的索引。

大多数图像隐写算法是将秘密消息嵌入到纹理相对复杂的区域，即像素值变化频繁的区域。在图像处理中，低通滤波器旨在滤除噪声的影响，而高通滤波器则是强化图像的纹理信息，提取残差信息。基于这一原因，各种形式的高通滤波器被广泛应用至图像隐写分析，以放大隐写信号的“痕迹”，如 KV 核 [78]、DCT 核 [79]、Gabor 核 [80]、Gauss 核 [81] 以及 SRM[82, 83] 等。其中，经过 KV 核处理前后的图像分别如图3.2所示。残差计算的引入使得特征对正常音频与隐写音频之间的差异更为敏感，从而可以更好地实现隐写分析。然而，QMDCT 系数矩阵的分布特性与图像、JPEG 系数有较大不同，以上高通滤波器在 MP3 音频

隐写分析中的效果并不理想。因此，本论文设计了一组适用于MP3音频隐写分析的富高通滤波模型。该富滤波模块共包含8个高通滤波器，依次为： $M^\rightarrow, M^\downarrow, A^\rightarrow, A^\downarrow, M^\Rightarrow, M^{\Downarrow}, A^\Rightarrow$ 和 A^{\Downarrow} 。分别对QMDCT系数矩阵以及QMDCT绝对值系数矩阵进行差分运算，从多个角度“放大”隐写信号的痕迹。

各滤波器的数学表达如下所示：

$$M_{m,n}^\rightarrow = Q_{i,j} - Q_{i,j+1} \quad (3.2)$$

$$M_{m,n}^\downarrow = Q_{i,j} - Q_{i+1,j} \quad (3.3)$$

$$A_{m,n}^\rightarrow = |Q_{i,j}| - |Q_{i,j+1}| \quad (3.4)$$

$$A_{m,n}^\downarrow = |Q_{i,j}| - |Q_{i+1,j}| \quad (3.5)$$

$$M_{m,n}^\Rightarrow = Q_{i,j} - 2 \times Q_{i,j+1} + Q_{i,j+2} \quad (3.6)$$

$$M_{m,n}^{\Downarrow} = Q_{i,j} - 2 \times Q_{i+1,j} + Q_{i+2,j} \quad (3.7)$$

$$A_{m,n}^\Rightarrow = |Q_{i,j}| - 2 \times |Q_{i,j+1}| + |Q_{i,j+2}| \quad (3.8)$$

$$A_{m,n}^{\Downarrow} = |Q_{i,j}| - 2 \times |Q_{i+1,j}| + |Q_{i+2,j}| \quad (3.9)$$

其中， $M^\rightarrow, M^\downarrow, A^\rightarrow, A^\downarrow$ 为一阶差分滤波器， $M^\Rightarrow, M^{\Downarrow}, A^\Rightarrow, A^{\Downarrow}$ 为二阶差分矩阵。 m, n 分别为差分矩阵横纵方向上的索引。从图3.3也可以看出，正常音频与隐写音频的QMDCT系数分布，在经过不同高通滤波器处理后，差异均会有不同程度的增加。

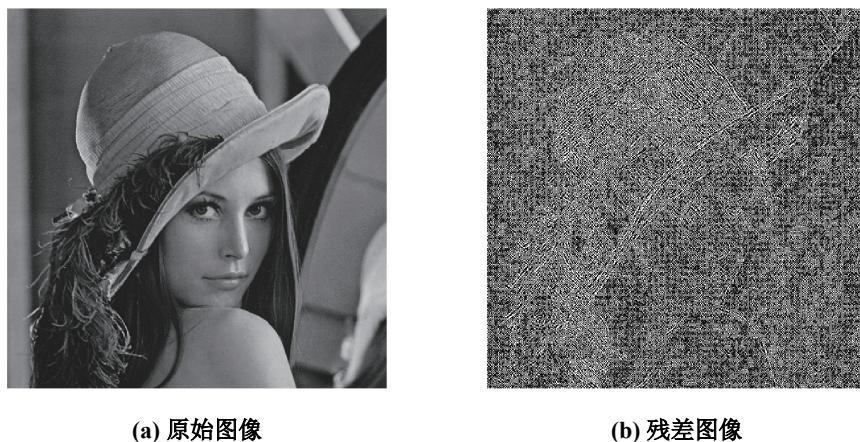


图3.2 KV核滤波效果示意图

Figure 3.2 Diagram of residual image after KV kernel

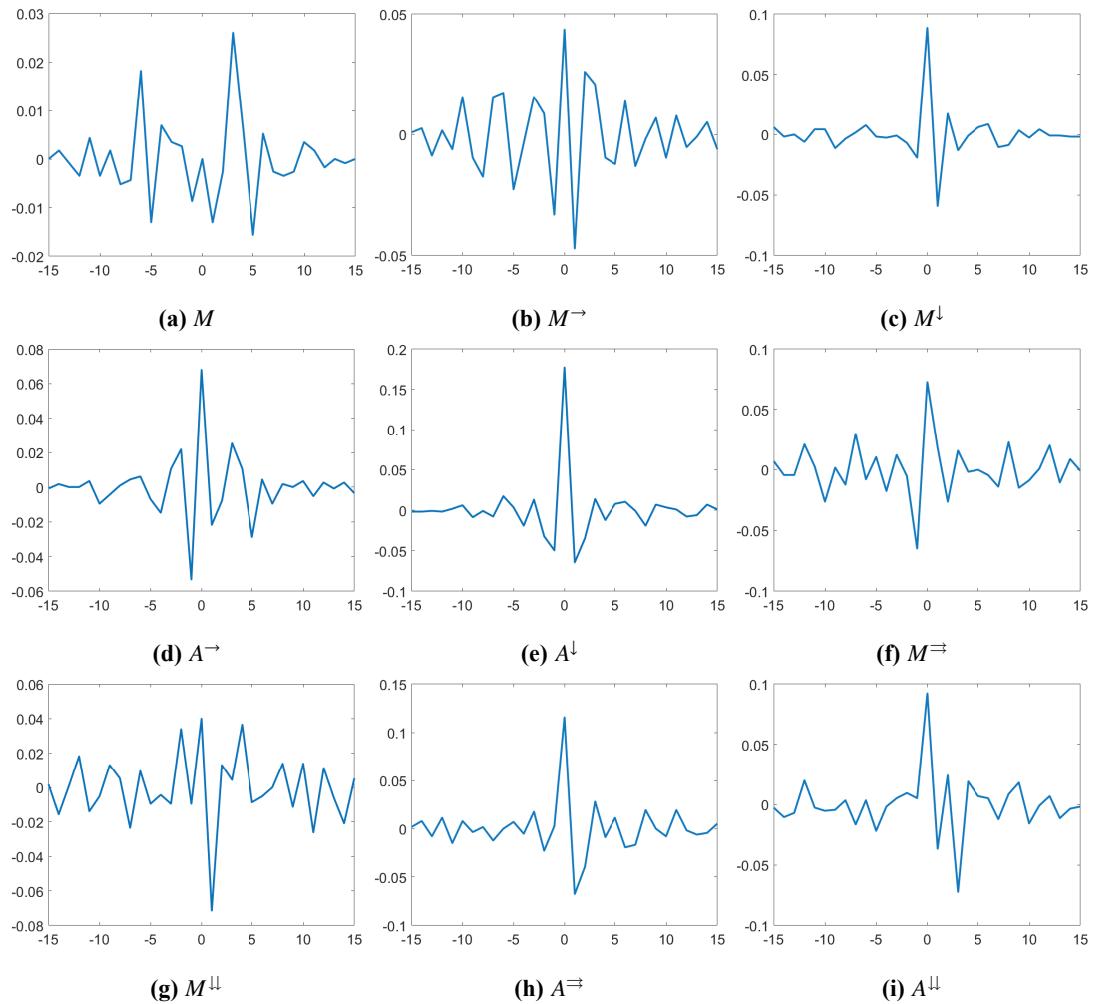


图 3.3 正常音频与隐写音频 QMDCT 系数分布差异图 (EECS, 128kbps)

**Figure 3.3 Difference of QMDCT coefficients distribution between cover and stego audios
(EECS, 128kbps)**

3.2.2 基于 MP3 编码特性的多尺度相关性度量模型

根据上文介绍的 MP3 编码原理，两个大值区的 QMDCT 系数会编码为一个 Huffman 码字，而四个小值区的 QMDCT 系数会编码为一个 Huffman 码字。然而，此前的 MP3 隐写分析方法设计中并未过多考虑这一编码特性。因此，为了能够更好地度量隐写算法对 MP3 音频 QMDCT 系数矩阵产生的影响，本章提出一种基于 MP3 编码特性的多尺度相关性度量模型，如图3.4所示。其中， f 表示音频帧， g 表示颗粒， c 表示声道， h 表示 Huffman 码字。算法在以度量相邻 QMDCT 系数点 Markov 单步转移概率相关性的基础上，分别对 QMDCT 系数矩阵进行 2×2 和 4×4 码字保持降采样，并计算降采样后各个分块子矩阵的 Markov 单步转移概率，用以度量 MP3 隐写对码字级相关性扰动的影响，由此便实现了在 1×1 、 2×2 、 4×4 三种不同的尺度上的相关性度量，图3.5为 2×2 码字保持降采样示

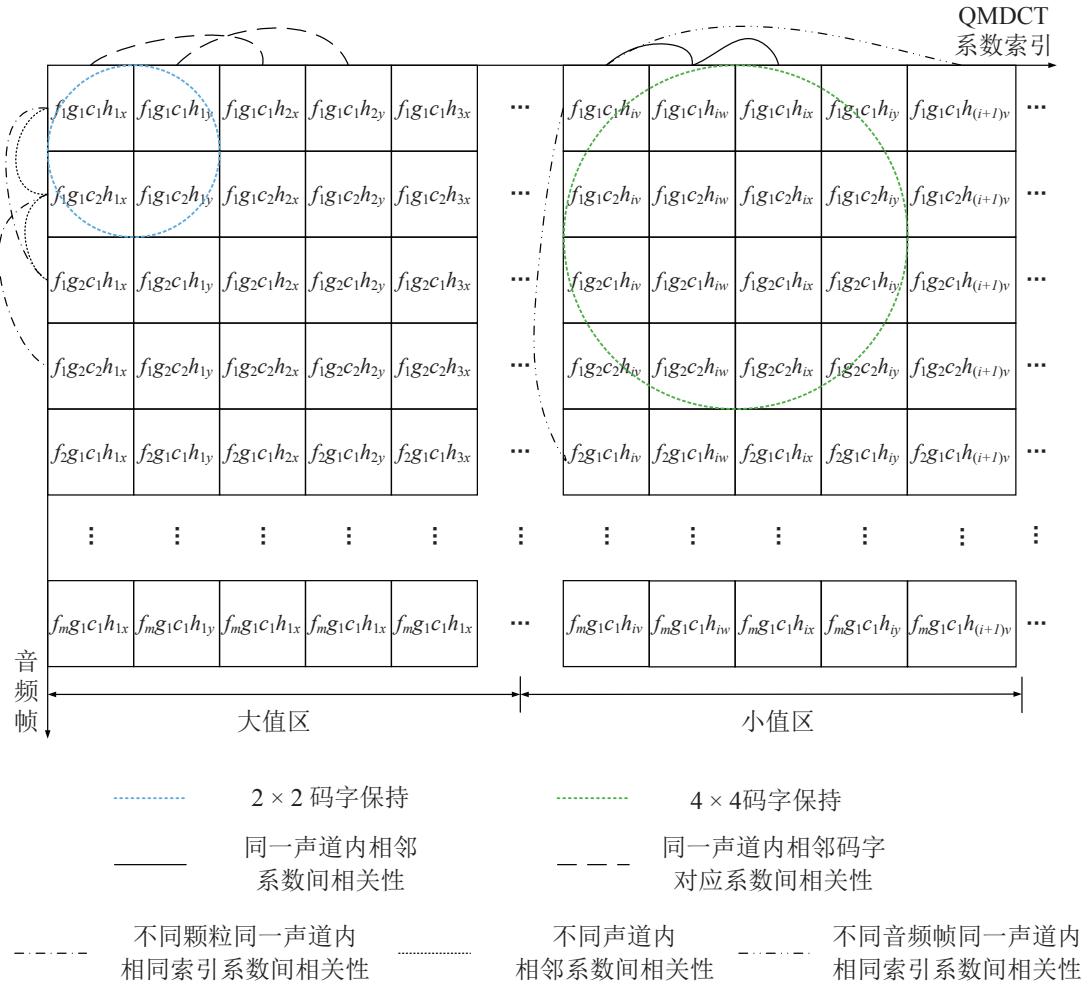


图 3.4 多尺度相关性度量模型示意图

Figure 3.4 Diagram of Multiscale correlation measure module

意图。MSC 算法通过对同一声道内相邻系数间相关性、同一声道内相邻码字对应位置系数间相关性、不同声道内相邻系数间相关性、不同颗粒同一声道内相同索引系数间相关性以及不同音频帧同一声道内相同索引系数间相关性的度量，实现了对由隐写导致的 QMDCT 系数矩阵相关性破坏的多尺度、多方位度量，有效地提升了对 MP3 隐写算法的检测效果。

Markov 单步转移概率的计算公式如3.10和3.11所示。

水平方向：

$$P_h = \sum_i \sum_j \frac{\delta(Q_{i,j} = x, Q_{i,j+1} = y)}{\delta(Q_{i,j} = x)} \quad (3.10)$$

竖直方向：

$$P_v = \sum_i \sum_j \frac{\delta(Q_{i,j} = x, Q_{i+1,j} = y)}{\delta(Q_{i,j} = x)} \quad (3.11)$$

其中, $\delta(X = x, Y = y) = \begin{cases} 1, & X = x, Y = y \\ 0, & \text{其他} \end{cases}$ and $x, y \in [-T, T]$ 。 T 是截断阈值。

通过对特征的可视化发现, 以 2×2 码字保持降采样为例, 降采样后的 4 个子矩阵的 Markov 单步转移概率矩阵具有较为相似的分布特性, 如图3.6所示。因此, 为了降低特征维度, 将 4 个子矩阵求得的转移概率矩阵对应元素相加 (Element-

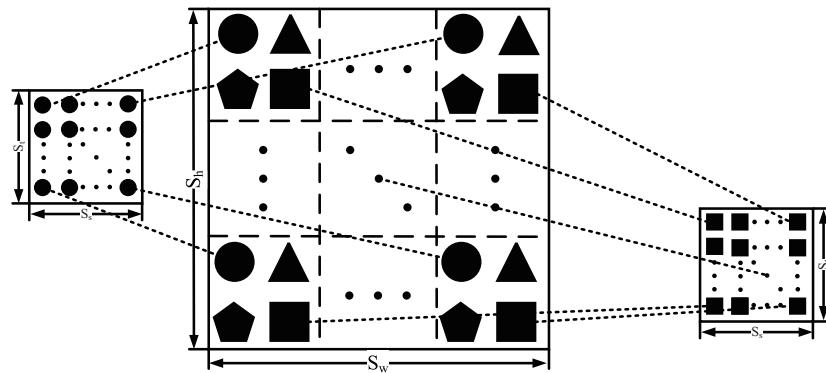


图 3.5 2×2 码字保持降采样模块示意图

Figure 3.5 Diagram of 2×2 codeword-aware down-sampling module

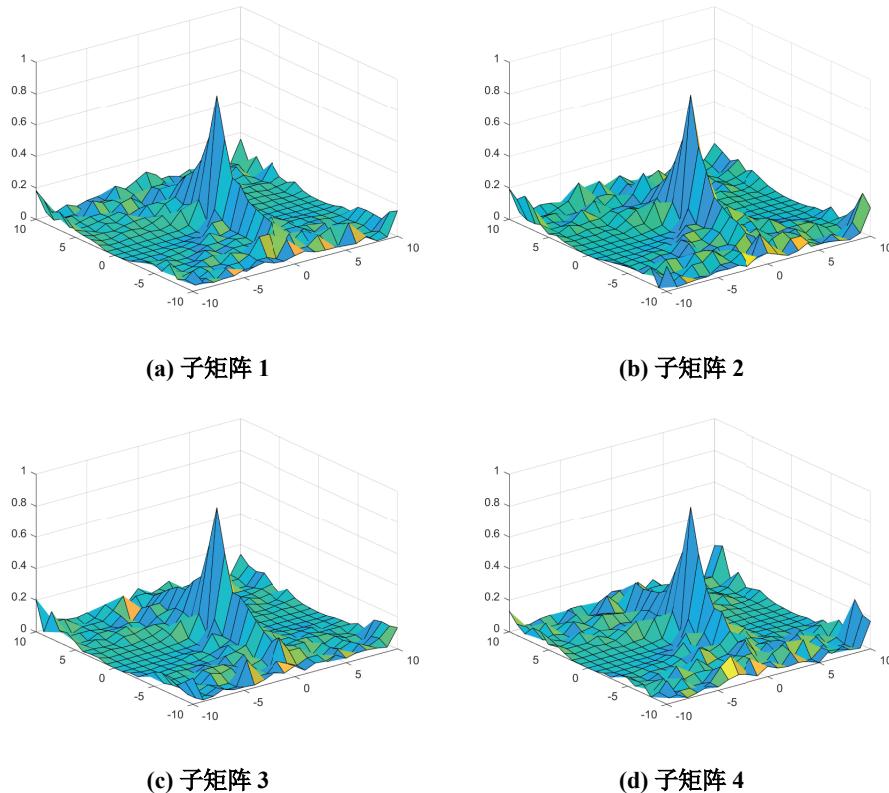


图 3.6 2×2 码字保持降采样后各子矩阵 Markov 转移概率矩阵示意图 (128kbps, 水平方向)

Figure 3.6 Diagram of each sub-matrix after 2×2 codeword-aware down-sampling
(128kbps, in the horizontal)

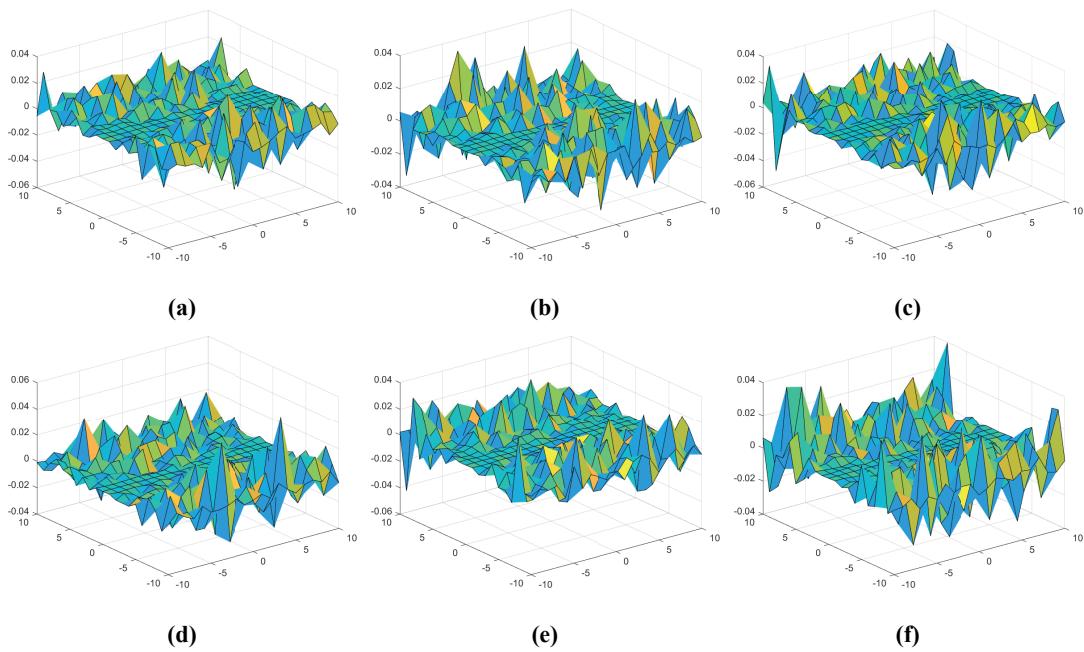


图3.7 正常音频与隐写音频特征分布差异图（EECS, 128kbps, M型滤波器）

**Figure 3.7 Difference of feature distributions between cover and stego audios
(EECS, 128kbps, M filter)**

wise) 后取其算术平均值。 4×4 码字保持降采样模块的特征提取与此同理。

最后,为了能够更好地说明多尺度相关性度量模型的有效性,本论文将对正常音频与隐写音频所提特征在不同尺度下的差异进行可视化,如图3.7所示。其中,3.7a, 3.7b, 3.7c为 1×1 、 2×2 和 4×4 尺度在水平方向上的Markov单步转移概率矩阵差异图;3.7d, 3.7e, 3.7f为竖直方向上的差异图。从图中可以看出, 2×2 和 4×4 码字保持降采样模块的引入,弥补了基于QMDCT系数值点相关性度量的不足,有助于检测算法性能的提升。

3.2.3 特征优选与降维

特征选择是在特征提取完成后一项十分重要的工作,旨在剔除不相关或者冗余的特征以降低有效特征的维度,减少模型训练的时间并且保证模型精度不降低。特征提取是通过空间转换实现对原始数据的降维,而特征选择则是借助统计学方法或者基于机器学习模型本身的特征选择功能实现特征的降维。特征选择的结果需要使用模型去验证,通过不断的重复迭代,最终实现模型提升的目的。本节主要探讨的是对截断阈值的选择和对富高通滤波器组子集的选择。由于截断阈值的选择对最终的检测效果影响更大,因此本节首先固定富高通滤波组,对截断阈值 T 进行选取。然后再固定截断阈值 T ,以“后减枝”的方式对“不重

要”的高通滤波器进行剔除。

3.2.3.1 截断阈值 T 优选

在进行相关性特征提取之前需要先对矩阵进行数据截断，截断阈值 T 的选择不仅会影响到特征的维度，同时也会影响算法的精度。如果截断阈值过大，不仅会大大增加运算时间，还会因为未修改系数点比例的增多而影响检测精度。如果截断阈值过小，则又会导致特征不稳定，难以适应各类算法的要求。由图3.8可以看出，MP3 音频原始 QMDCT 系数的分布集中分布在 $[-15, 15]$ 的范围内，而 MP3Stego、HCM 及 EECS 隐写算法主要修改的系数集中在 $[-10, 10]$ 的范围内。同时，从图3.7中也可以看出，当截断阈值过小时，正常音频与隐写音频所提特征的差异也会减小。因此，在保持原始富高通滤波模型完整的前提下，选取可能的截断阈值 $T \in \{3, 4, 5, 6, 7, 8, 9, 10\}$ 进行分析，实验结果如图3.9所示。从实验

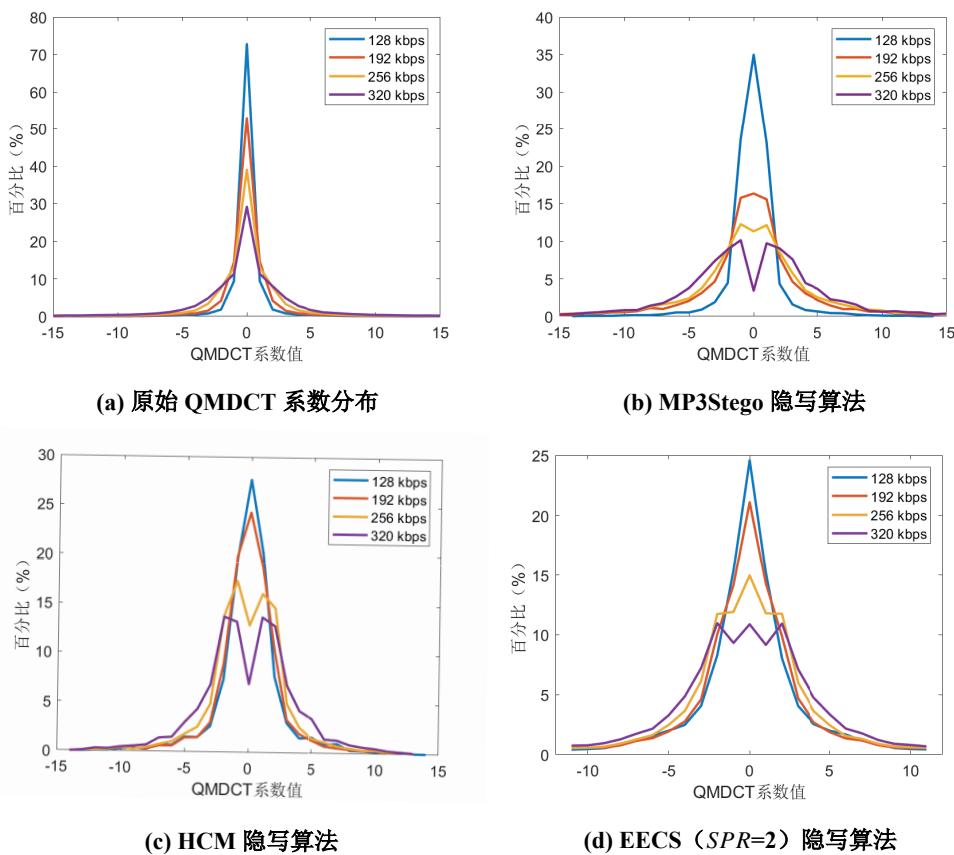


图 3.8 QMDCT 系数分布图。(a) 为原始 MP3 音频的 QMDCT 系数分布，(b) - (d) 分别为经过 MP3Stego、HCM 和 EECS ($SPR = 2$) 算法隐写后被修改的 QMDCT 系数分布

Figure 3.8 Distributions of QMDCT coefficients. (a) is the distribution of the original MP3 audio, (b) - (d) are distributions of modified coefficients through MP3Stego, HCM and EECS ($SPR = 2$) steganographic algorithms

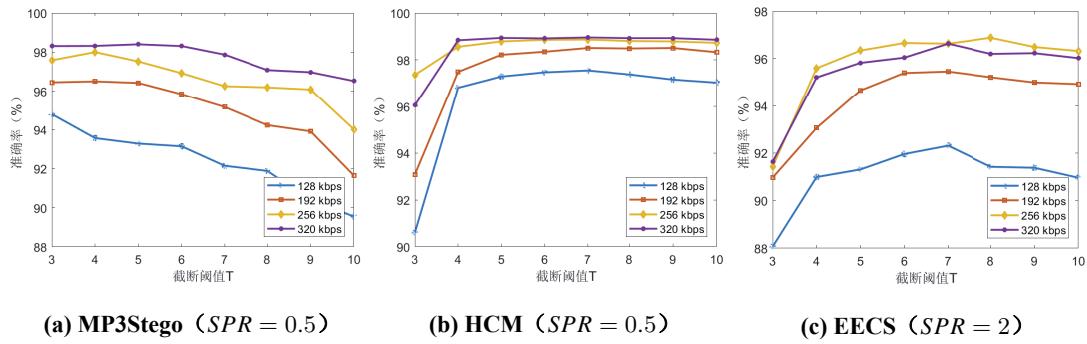


图 3.9 对三种隐写算法在不同截断阈值选择下的检测准确率曲线

Figure 3.9 Curves of detection accuracy of the three steganographic algorithms with different truncated thresholds

结果可以看出，不同的隐写算法和音频比特率对截断阈值 T 的变化呈现出不同的变化趋势。对于 MP3Stego 隐写算法，随着截断阈值 T 的增加，当比特率高于 128kbps 时，检测准确率先略微增加或基本保持不变，然后再下降；而在 128kbps 条件下，检测准确率会随着截断阈值 T 的增加而下降。当 $T < 8$ 时，检测精度的下降幅度仍在可接受的范围内。而对于 HCM 和 EECS 隐写算法，随着截断阈值 T 的增加，准确率均呈现出先上升后下降的趋势，当 $T = 7$ 时达到最大值。因此，截断阈值 T 最终设定为 7。

3.2.3.2 子滤波器组最优组合模式选择

富高通滤波模型是一个通用的残差提取模型，对于 MP3Stego、HCM、EECS 算法的隐写分析，部分高通滤波器移除后，检测精度并不会有较为明显的下降，而且还可以减少特征的维度。影响算法性能的主要因素有输入数据的差异和所提特征的差异，因此本论文使用正常音频与隐写音频 QMDCT 系数矩阵中差异点的比例 d_p ，特征间的欧式距离 $d_e(c, s)$ 和余弦相似度 $d_c(c, s)$ 来度量不同高通滤波器的作用。 d_p 是对正常音频与隐写音频差异性的绝对度量，值越大表明其差异越大； $d_e(c, s)$ 是通过欧式空间中的绝对距离度量所提特征对的差异性，值越大表明其差异越大；余弦相似度是在方向偏离上度量所提特征对的差异性，其值介于 0 到 1 之间，值越接近 0 表明两个向量之间的差异越大。然后，计算各个滤波器的“贡献度” Score，并根据“贡献度”对各个高通滤波器进行降序排列，利用后减枝的方式实现最优子滤波器组的优选，即依次从滤波器组中依次剔除“贡献度”较小的滤波器，直至特征维度和性能指标达到预期要求为止。为了消

除各项指标由于分布不同引起的计算偏差，本文将分别对其进行归一化处理。

$$d_p = \frac{N_{diff}}{200 \times 400} \times 100\% \quad (3.12)$$

$$d_e(c, s) = \sqrt{\sum_{k=1}^n (c_k - s_k)^2} \quad (3.13)$$

$$d_c(c, s) = \frac{\sum_{k=1}^n (c_k \cdot s_k)}{\sqrt{\sum_{k=1}^n c_k^2} \sqrt{\sum_{k=1}^n s_k^2}} \quad (3.14)$$

$$X^* = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (3.15)$$

$$Score = \alpha \times d_p + \beta \times d_e + \gamma \times (1 - d_c) \quad (3.16)$$

其中， N_{diff} 为 QMDCT 系数矩阵中被修改点的个数， c 和 s 分别表示从正常音频与隐写音频中提取的分析特征。 n 为特征的维度， k 为特征索引值。 X 与 X^* 分别为归一化前后的向量。 X_{max} 和 X_{min} 表示向量中的最大值和最小值。 α 、 β 和 γ 均为经验值，在本论文中 $\alpha = 0.1$ ， $\beta = 0.5$ ， $\gamma = 0.4$ 。

以各个高通滤波器对 EECS 隐写算法的影响为例，差异性大小与滤波器排名分别如表3.1、3.2所示。减枝流程与其相应的性能指标如表3.3所示。由上述实验结果可以看出， M^\Rightarrow 和 M^{\Downarrow} 高通滤波器的移除仅使得对 EECS 算法的检测精度下

表 3.1 各高通滤波器处理后正常音频与隐写音频的差异 (EECS, 128kbps, SPR = 2, T = 7)

Table 3.1 Difference between cover and stego audios after the processing of each high-pass filter (EECS, 128kbps, SPR = 2, T = 7)

距离	M	M^\rightarrow	M^\downarrow	M^\Rightarrow	M^{\Downarrow}	A^\rightarrow	A^\downarrow	A^\Rightarrow	A^{\Downarrow}
d_p	4.36	5.60	7.83	8.08	11.17	6.18	7.81	8.16	11.19
$d_e(c, s)$	0.2430	0.2180	0.2566	0.1680	0.1834	0.3477	0.3712	0.2649	0.2653
$d_c(c, s)$	0.9980	0.9988	0.9979	0.9990	0.9989	0.9965	0.9960	0.9979	0.9979

表 3.2 各高通滤波器“贡献度”及排名 (EECS, 128kbps, SPR = 2, T = 7)

Table 3.2 "Contribution" of each high-pass filter and the corresponding rank (EECS, 128kbps, SPR = 2, T = 7)

	M	M^\rightarrow	M^\downarrow	M^\Rightarrow	M^{\Downarrow}	A^\rightarrow	A^\downarrow	A^\Rightarrow	A^{\Downarrow}
Rank	-	6	5	8	7	2	1	4	3
Score	0.3681	0.1679	0.4155	0.0545	0.1509	0.8022	0.9505	0.4407	0.4861

表3.3 基于后减枝的子滤波器组最优组合模式选择 (EECS, 128kbps, SPR = 2, T = 7)

Table 3.3 Optimization of the sub-filters combination based on post-pruning
(EECS, 128kbps, SPR = 2, T = 7)

步骤	移除的滤波器	特征维度	检测指标		
			FPR	FNR	ACC
1	-	12150	7.34	7.63	92.51
2	M^{\Rightarrow}	10800	7.21	7.79	92.50
3	M^{\Downarrow}	9450	7.75	7.25	92.50
4	M^{\rightarrow}	8100	8.44	9.03	91.27
5	M^{\downarrow}	6750	9.03	9.32	90.83
6	A^{\Rightarrow}	5400	12.07	12.67	87.63
7	A^{\Downarrow}	4050	12.03	13.25	87.36
8	A^{\rightarrow}	2700	13.61	14.36	86.02
9	A^{\downarrow}	1350	14.63	15.25	85.06

降了 0.01%，可以将其从滤波器组中移除。 A 型高通滤波器对算法性能的提升作用较为显著，具有较高的“贡献度”，尤为明显的是， A^{\Rightarrow} 滤波器的移除导致准确率下降 3% 以上。然而，考虑到基于符号位反转这一类隐写算法的存在，仍需保留 M 型滤波器的存在。最终，在运算速度和检测性能这一对矛盾的权衡之下，子滤波器组的最优组合模式为： M^{\rightarrow} 、 A^{\Rightarrow} 、 A^{\Downarrow} 、 M^{\downarrow} 、 A^{\rightarrow} 和 A^{\downarrow} 。

此外，在截断阈值选择和富高通滤波模型优选后，还可以通过基于嵌入式 (Embedded)、过滤式 (Filter) 和封装式 (Wrapper) 等常用特征选择方法进一步对所提取特征进行优化，本论文在此不再赘述。

3.3 实验设计与分析

基于上述分析，本节将通过实验验证 MSC 算法的性能。首先，分别验证富高通滤波模型和多尺度相关性度量模型的有效性。其次，讨论待检测音频数据段长度对隐写分析的影响，以确定隐写音频定位的最小粒度。最后，与 ADOTP 和 MDI2 算法进行对比，以评估各分析算法在不同隐写算法、音频比特率及隐写负载率条件下的性能优劣。

表 3.4 实验设置

Table 3.4 Settings of experiments

项目	参数	值
音频信息	帧数	50
	比特率	128 / 192 / 256 / 320 kbps
隐写负载	HCM	$SPR = 0.1 / 0.3 / 0.5 / 0.8 / 1.0$
	MP3Stego	
分类器类型	EECS	$SPR = 2 / 3 / 4 / 5 / 6$
	集成分类器	默认值

3.3.1 实验设置

为了全面地评估本章所提算法的性能, 本论文分别从数据库中选取了各参数条件下 10000 个正常/隐写音频对用于隐写分析, 其中 6000 对用于训练, 4000 对用于测试, 其他各项实验参数设置如表3.4所示。使用虚警率 (False Positive Rate, FPR)、漏检率 (False Negative Rate, FNR) 及准确率 (Accuracy, ACC) 来度量隐写分析算法的性能, 实验结果取 10 次实验的算术平均值, 以减少偶然误差带来的影响。

$$FPR = \frac{FP}{FP + TN} \times 100\% \quad (3.17)$$

$$FNR = \frac{FN}{FN + TP} \times 100\% \quad (3.18)$$

$$ACC = 1 - \frac{FPR + FNR}{2} \times 100\% \quad (3.19)$$

其中, TP 、 FP 、 TN 、 FN 分别为真阳性样本、假阳性样本、真阴性样本和假阴性样本的数量。

3.3.2 模块有效性验证

MSC 算法的主要创新点在于富高通滤波模型和多尺度相关性度量模型的引入, 为了有效说明这两个模块的作用, 分析将其移除后对最终准确率的影响, 实验结果如表3.5所示。

从以上实验结果可以看出:

- 富高通滤波模型的引入能够显著提升隐写分析的准确率。然而, 与图像

表3.5 子模块有效性验证(EECS, 128kbps, SPR = 2)

Table 3.5 Effectiveness of each sub-module (EECS, 128kbps, SPR = 2)

算法说明	检测指标		
	FPR	FNR	ACC
本章所提算法	7.75	7.25	92.50
移除原始QMDCT系数矩阵	10.72	10.20	89.54
移除音频富高通滤波模型	14.63	15.25	85.06
移除所有码字保持降采样模块	10.30	10.88	89.41
移除 2×2 码字保持降采样模块	9.50	9.75	90.38
移除 4×4 码字保持降采样模块	9.00	9.72	90.64

隐写分析不同，原始的QMDCT系数矩阵本身也可以提供十分有用的信息，移除原始的QMDCT系数矩阵会导致准确率有较为明显的下降。因此，在输入数据中保留原始QMDCT系数矩阵可以更好地保证隐写分析算法的性能。

2. 多尺度相关性度量模型同样有助于隐写分析准确率的提高，且 2×2 码字保持降采样模块的作用大于 4×4 码字降采样保持模块。一方面，大值区的系数值更为丰富，被修改的概率更大；另一方面，现有的MP3音频隐写算法以修改大值区的系数为主。此外，系数间的相关性会随着系数间距离的增大而有所下降。

3.3.3 隐写音频帧定位

MP3音频信号是时序信号，任意两个MP3音频的时长可能均不相同。对于恒定比特率的MP3音频，其每帧的长度为26ms，因此一首时长为3分钟的MP3音频，其音频帧总数可到达7000帧，由此形成的QMDCT系数矩阵尺寸为 28000×400 ，数据量远超图像载体。一方面，矩阵尺寸过大会导致检测速度的下降；另一方面，秘密消息的不完全嵌入会降低隐写分析算法的性能，例如一首3分钟的音乐仅在前10s嵌入了秘密消息。因此，为了分析不同音频段大小对隐写分析算法性能的影响，同时为了确定算法所能实现的隐写音频帧的定位粒度，本论文将选取不同高度的QMDCT系数矩阵作为输入数据，分析其对最终检测性能的影响，实验结果如表3.6所示。

从上表中的实验结果可以看出，随着系数矩阵高度的减少，隐写分析的准确率在不断下降。这是由于在检测过程中，随着所选音频帧数的减少，正常音频与

表 3.6 不同长度音频段对隐写分析算法性能的影响 (EECS, 128kbps, SPR = 2)

Table 3.6 Performance of MSC algorithm under different audio length
(EECS, 128kbps, SPR = 2)

QMDCT 系数矩阵高度	音频帧数	时长	检测指标		
			FPR	FNR	ACC
8	2	52 ms	29.55	30.54	69.95
20	5	130 ms	28.66	27.44	71.95
40	10	260 ms	24.22	23.38	76.20
60	15	390 ms	20.40	19.60	80.00
80	20	520 ms	16.51	17.42	83.03
100	25	650 ms	14.14	15.67	85.10
120	30	780 ms	12.57	13.63	86.90
140	35	910 ms	11.06	11.82	88.56
160	40	1040 ms	9.21	9.99	90.40
180	45	1170 ms	8.26	9.29	91.22
200	50	1300 ms	7.75	7.25	92.50
300	75	1950 ms	7.00	5.80	93.60
400	100	2600 ms	5.80	3.40	95.40

隐写音频的差异也将逐渐减小，算法对正常音频与隐写音频的区分能力也在逐步降低。最终，基于计算代价和检测准确率的综合考虑，本论文在进行 MP3 音频隐写分析时使用的 QMDCT 系数矩阵尺寸为 200×400 ，共 50 个音频帧，即 1.3s。此外，从实验结果可以看出，MSC 算法可在仅有 20 个音频帧的条件下实现 83.03% 以上的检测准确率，由此确定算法所能实现的隐写音频帧定位粒度为 500ms（置信度大于 80%）。

3.3.4 实验结果与分析

本小节将通过实验对 MSC 算法、ADOTP 算法和 MDI2 算法的性能进行比较，实验结果如表3.7-3.9所示。从实验结果可以看出，MSC 算法在本论文所分析的各类隐写条件下（三种隐写算法、四种常见比特率以及五种隐写负载率）的检测结果均优于 ADOTP 算法和 MDI2 算法。尤其是对于 EECS 隐写算法的分析，

MSC 算法检测结果的平均值相比于 ADOTP 算法和 MDI2 算法提升 20% 以上。

对于 MP3Stego 隐写算法，由于 MP3Stego 算法对 MP3 音频的 QMDCT 系数矩阵改动较大，因此三种隐写分析在多种嵌入率下的隐写分析结果均较为理想，除比特率为 128kbps, $SPR = 0.1$ 时，ADOTP 算法的检测准确率低于 90%，其余所有的实验结果均在 90% 以上。但是，需要注意的是，在 128kbps 比特率下，三类算法均存在虚警率低，漏检率高的问题；在 192kbps 比特率下，三类算法又变为虚警率高，漏检率低，这一现象本质上是由 MP3Stego 算法较不稳定所导致的。

对于 HCM 算法的隐写分析，算法性能排序为：MSC > MDI2 > ADOTP，但 ADOTP 和 MDI2 算法在低比特率条件下的检测结果难以达到实际使用需求，如比特率为 128kbps, $SPR = 0.1, 0.3, 0.5$ 以及比特率为 256kbps, $SPR = 0.1$ 时，检测结果均未达到 80%。而对于 MSC 算法，仅在 128kbps、192kbps, $SPR = 0.1$ 时检测精度低于 85%，其余参数下的检测结果均在 85% 以上。性能提升较为明显。

对于 EECS 算法的隐写分析，ADOTP 与 MDI2 分析算法即使在比特率为 320kbps 的最大隐写负载条件下，检测准确率仍在 80% 以下。而，MSC 算法在 $SPR = 2$ 和 3 两种较大隐写负载率条件下的检测准确率均在 79% 以上，且漏检率和虚警率相当。特别是，当 $SPR = 2$ 时，MSC 的检测准确率均在 90% 以上。但是三种算法在 $SPR = 4, 5, 6$ 等低隐写负载率条件下的隐写分析效果并不理想，算法性能仍有待提升。

表 3.7 对 MP3Stego 算法的隐写分析结果

Table 3.7 Performance of each steganographic algorithm on the detection of MP3Stego

比特率	SPR	ADOTP			MDI2			MSC		
		FPR	FNR	ACC	FPR	FNR	ACC	FPR	FNR	ACC
128	0.1	11.57	13.21	87.61	7.98	10.70	90.66	7.95	7.85	92.10
	0.3	9.33	9.21	90.73	1.47	14.29	92.12	2.02	10.00	93.99
	0.5	6.47	11.06	91.24	0.97	11.92	93.56	2.45	7.89	94.83
	0.8	1.03	12.68	93.15	1.93	10.36	93.86	1.77	7.19	95.52
	1.0	1.83	10.06	94.06	1.40	9.68	94.46	1.51	5.53	96.48
192	0.1	10.25	9.33	90.21	11.43	5.41	91.58	5.60	4.80	94.80
	0.3	12.54	2.09	92.68	7.74	5.14	93.56	7.99	0.48	95.77
	0.5	6.92	4.04	94.52	6.17	4.68	94.58	6.64	1.56	95.90
	0.8	4.16	3.50	96.17	4.38	3.02	96.30	6.10	1.07	96.42
	1.0	2.38	2.02	97.80	4.98	2.12	96.45	1.39	2.17	98.22
256	0.1	10.81	5.37	91.91	9.45	3.61	93.47	5.31	4.41	95.14
	0.3	5.58	5.42	94.50	5.47	5.53	94.50	5.05	1.51	96.72
	0.5	5.08	2.05	96.44	6.47	3.02	95.26	3.47	2.73	96.90
	0.8	4.88	2.05	96.54	3.77	4.79	95.72	3.70	0.54	97.88
	1.0	2.48	1.52	98.00	5.64	1.46	96.45	1.49	2.01	98.25
320	0.1	6.51	6.61	93.44	5.99	6.81	93.60	5.67	2.35	95.99
	0.3	4.09	3.41	96.25	4.79	5.66	94.78	4.19	2.63	96.59
	0.5	2.53	3.47	97.00	5.12	3.78	95.55	3.20	2.44	97.18
	0.8	2.46	2.03	97.76	4.25	2.80	96.48	2.96	0.51	98.27
	1.0	1.54	0.81	98.83	2.06	4.37	96.79	0.53	1.42	99.03

表3.8 对HCM算法的隐写分析结果

Table 3.8 Performance of each steganographic algorithm on the detection of HCM

比特率	SPR	ADOTP			MDI2			MSC		
		FPR	FNR	ACC	FPR	FNR	ACC	FPR	FNR	ACC
128	0.1	44.26	42.09	56.83	38.32	42.23	59.73	31.00	33.35	67.83
	0.3	33.48	30.12	68.20	28.64	29.88	70.74	12.40	13.35	87.12
	0.5	23.26	18.77	78.98	17.61	15.60	83.39	2.79	2.00	97.61
	0.8	9.82	7.29	91.44	7.91	6.69	92.70	0.33	0.05	99.81
	1.0	4.81	3.84	95.67	4.28	2.53	96.60	0.14	0.11	99.88
192	0.1	39.96	36.14	61.95	35.21	31.80	66.49	20.04	19.61	80.17
	0.3	24.46	22.16	76.69	21.82	21.20	78.49	4.64	4.56	95.40
	0.5	12.80	9.88	88.66	10.47	11.46	89.03	0.61	0.75	99.32
	0.8	4.20	3.65	96.08	3.24	3.14	96.81	0.14	0.03	99.92
	1.0	1.84	1.54	98.31	1.76	1.35	98.44	0.01	0.00	99.99
256	0.1	38.08	34.99	63.47	23.99	21.04	77.49	11.15	11.69	88.58
	0.3	20.59	19.14	80.14	15.17	12.09	86.37	1.85	1.89	98.13
	0.5	9.11	7.99	91.45	6.73	6.03	93.62	0.04	0.29	99.84
	0.8	3.49	2.84	96.84	1.71	1.48	98.41	0.00	0.00	100.0
	1.0	1.56	1.36	98.54	0.94	0.74	99.16	0.00	0.00	100.0
320	0.1	33.16	30.68	68.08	10.02	12.99	88.49	5.84	7.76	93.20
	0.3	19.95	19.32	80.36	7.74	9.13	91.57	1.13	1.92	98.47
	0.5	7.47	8.17	92.18	4.17	4.64	95.59	0.10	0.03	99.94
	0.8	1.76	2.11	98.06	1.43	1.35	98.61	0.00	0.00	100.0
	1.0	0.84	0.11	99.53	0.66	0.90	99.22	0.00	0.00	100.0

表 3.9 对 EECS 算法的隐写分析结果

Table 3.9 Performance of each steganographic algorithm on the detection of EECS

比特率	SPR	ADOTP			MDI2			MSC		
		FPR	FNR	ACC	FPR	FNR	ACC	FPR	FNR	ACC
128	2	31.01	27.01	70.99	28.05	29.27	71.34	7.91	7.80	92.14
	3	41.76	36.22	61.01	38.45	37.69	61.93	20.22	21.65	79.06
	4	44.49	39.73	57.89	42.02	43.01	57.48	30.54	30.48	69.49
	5	45.48	44.35	55.09	46.63	44.44	54.47	34.88	36.45	64.34
	6	48.52	43.75	53.86	50.86	45.74	51.70	39.00	41.06	59.97
192	2	26.77	24.10	74.56	25.11	25.69	74.60	4.35	4.71	95.47
	3	40.27	35.74	61.99	38.20	37.74	62.03	16.77	17.39	82.92
	4	43.84	39.51	58.32	44.16	41.66	57.09	24.65	26.85	74.25
	5	47.79	41.73	55.24	44.40	45.80	54.90	30.66	29.92	69.71
	6	46.18	45.16	54.33	47.69	46.64	52.84	32.36	38.03	64.81
256	2	26.76	24.08	74.58	19.21	21.24	79.78	4.19	4.11	95.85
	3	37.71	36.84	62.72	36.38	30.66	66.48	14.91	17.39	83.85
	4	42.76	39.40	58.92	43.95	38.02	59.01	25.22	24.76	75.01
	5	45.89	42.25	55.93	45.04	40.85	57.06	28.62	31.10	70.14
	6	46.66	44.85	54.24	47.42	42.61	54.98	33.14	36.20	65.33
320	2	26.96	26.21	73.41	22.53	17.58	79.95	3.53	3.86	96.31
	3	37.95	38.59	61.73	31.46	35.00	66.77	15.45	16.44	84.05
	4	42.74	38.76	59.25	35.94	40.38	61.84	24.59	24.33	75.54
	5	45.98	42.48	55.77	41.45	40.84	58.86	29.34	30.23	70.22
	6	46.74	44.32	54.47	43.44	43.89	56.34	33.67	33.45	66.44

3.4 本章小结

本章通过分析多种隐写算法对 MP3 音频 QMDCT 系数的影响，并结合 MP3 音频的编码原理，提出一种基于 QMDCT 系数矩阵多尺度相关性度量的隐写分析方法。在相邻 QMDCT 系数相关性度量的基础上，引入 2×2 和 4×4 码字保持降采样模块，以实现对 QMDCT 系数矩阵相关性的多尺度度量。同时，本章还提出一种适用于 MP3 音频隐写分析的富高通滤波模型，从多个角度“放大”隐写信号的痕迹，有效提升了算法对隐写信号的敏感性，性能相比于以往的隐写分析算法有较为明显的提升，尤其是对现阶段安全性最高的 EECS 隐写算法的检测分析提升更为明显，准确率平均可提升 20% 以上。

此外，本章还分析了待检测音频段长度对隐写分析的影响，实现了 500ms 粒度的隐写音频帧定位，置信度为 80% 以上。

MSC 算法在诸多隐写条件下的检测性能甚至与基于 CNN 的隐写分析算法相媲美，且需要的计算资源更少，具有很强的实用性，但对于更低隐写负载率的隐写分析性能仍有待提升。

第4章 基于卷积神经网络的MP3音频隐写分析研究

4.1 引言

深度学习(Deep Learning)是近年来机器学习领域的一个新兴研究方向,现已被广泛应用于图像分类(Image Classification)、语音识别(Speech Recognition)、目标跟踪(Object Tracking)及语义分割(Semantic Segmentation)等众多领域。在众多优秀的神经网络中,最具代表性的有LeNet[84]、AlexNet[85]、VGG[86]、GoogleNet[87]、ResNet[88]以及DenseNet[89],在ImageNet[90]数据集上的分类能力已经超越了人类。

隐写分析的本质是一个分类任务,其目标是寻找最优的特征表达,以更好地实现对正常载体与隐写载体的区别。因此,深度学习技术同样可以引入至隐写分析算法的设计中。如图4.1所示,基于手工特征设计的隐写分析算法和基于深度学习的隐写分析算法符合相同的框架。两者最本质的区别在于特征表达的不同,基于卷积神经网络的隐写分析算法实现对特征空间的自动寻优,而基于手工特征设计的隐写分析算法需要较强的先验知识和设计经验,所设计的特征也往往难以全面地对正常载体与隐写载体进行区分。到目前为止,卷积神经网络已经在空域和JPEG(Joint Photographic Experts Group)域的图像隐写分析[91–100]中得到了很好的运用,对各类图像隐写算法的检测准确率逐年提升,检测效果从最初的逼近基于富模型与集成分类器的传统分析方法,现在已经实现了完全的超越。此外,各类隐写分析网络在低隐写负载率下也能获得较高的检测精度。然

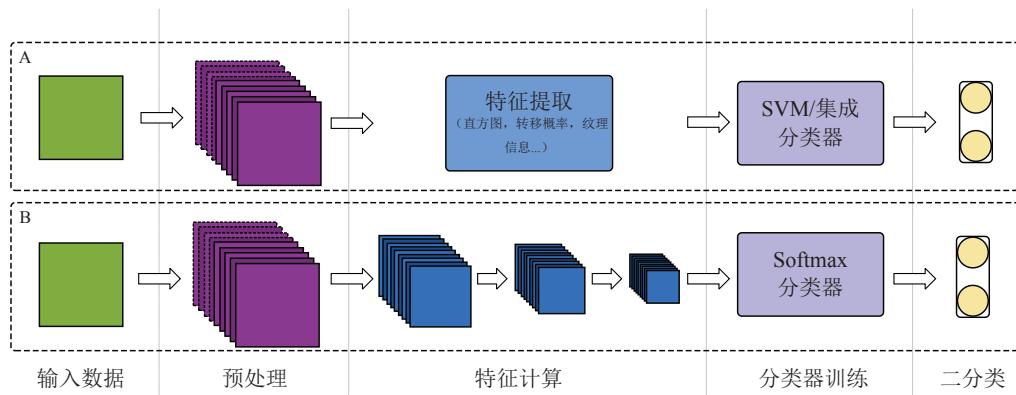


图 4.1 基于手工特征设计的隐写分析 (A) 与基于卷积神经网络的隐写分析 (B) 对比图

Figure 4.1 Comparison of steganalysis based on handcrafted feature design (A) and steganalysis based on CNN (B)

而，深度学习技术现阶段在音频隐写分析中的应用还相对较少，有较大的探索和研究空间。

本章的主要贡献在于提出了一个基于深度学习的 MP3 音频隐写分析设计架构。通过分析不同的输入数据类型、神经网络类型以及网络结构对网络性能的影响，提出一种基于 QMDCT 系数矩阵与 CNN 的 MP3 音频隐写分析网络，以下简称 RHFCN (Rich High-Pass Fully Convolutional Network)。首先，证明了基于 CNN 的隐写分析架构适用于 MP3 音频隐写分析，而且网络的检测性能优于 LSTM 等其他形式的神经网络结构。其次，证明了 QMDCT 系数矩阵是一种更适用于 MP3 音频隐写分析的数据空间。此外，本论文还分别提出一种面向输入数据尺寸失配的解决方案和一种基于迁移学习的低隐写负载率样本隐写分析方案，极大地提升了隐写分析网络的有效性和实用性，对今后压缩域音频隐写分析的发展也具有指导和借鉴意义。

本章后续的内容组织安排如下：4.2节介绍网络设计原理，4.3节是实验设计与分析，4.4节是对本章工作的小结。

4.2 算法设计

本章提出 RHFCN 网络结构如图4.2所示。首先提取 MP3 音频的 QMDCT 系数矩阵，并将其与经过富高通滤波模型处理后的差分矩阵在通道层进行拼接，然后将拼接后的矩阵输入到卷积层中进行特征提取，此处的富高通滤波模型即为第三章提出的适用于 MP3 音频隐写分析的高通滤波器组。RHFCN 网络共包含 8

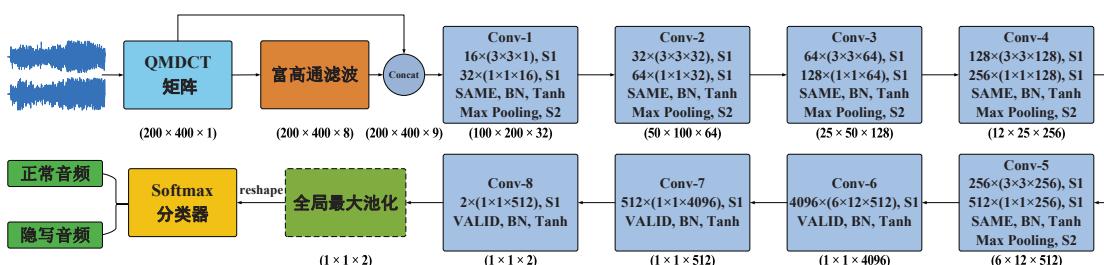


图 4.2 RHFCN 结构示意图 ($16 \times (3 \times 3 \times 9)$ 表示卷积核尺寸为 3×3 ，卷积核个数为 16，输入特征图通道数为 9。“S1”表示步长为 1，“SAME” 和 “VALID” 为两种卷积补齐方式，各方框下为输出特征图的维度。)

Figure 4.2 Diagram of RHFCN. $16 \times (3 \times 3 \times 9)$ represents the size of convolutional kernel is 3×3 , the number of kernels is 16, and the output channels are 9. "S1" means the stride is 1. "SAME" and "VALID" are the two padding method. And, the dimension of feature maps is shown below each block.

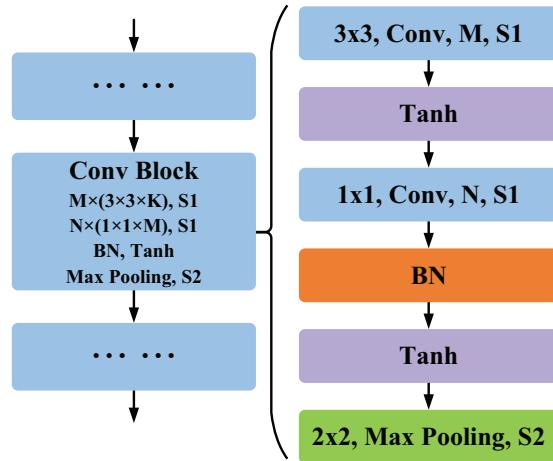


图 4.3 Conv 模块示意图

Figure 4.3 Diagram of Conv block

个卷积模块，其中 Conv1 - Conv5 用于特征提取，其内部结构如图4.3所示。每个卷积模块顺序地包含一个 3×3 卷积（Convolution, Conv）层、一个 1×1 卷积层、一个批标准化（Batch Normalization, BN）层、一个 Tanh 激活函数层以及一个最大池化（Max Pooling）层。Conv6 - Conv8 卷积模块主要用于对特征图的进行信息加以整合，取代了全连接层。经过全局最大池化后将输出特征图进行维度变换，由 $1 \times 1 \times 2$ 变为 1×2 ，输入到 Softmax 分类器中。与之前的许多网络结构不同，训练好的 RHFCN 模型可以接收大于 200×400 的任意尺寸数据的输入，可在一定程度上解决输入数据尺寸失配的问题。

4.2.1 输入数据类型评估

输入数据决定了模型精度的上限，而算法的设计与优化则是尽可能逼近这个上限。清华大学团队也为此提出 ColorNet[101]，用以探索颜色空间对图像分类的重要性。MP3 音频隐写过程中，不仅会对 MP3 音频 QMDCT 系数产生较为明显的影响，同样也会扰动 MP3 音频的其他重要参数及数据空间，如 MFCC 系数域、时域采样点等，如图4.4所示。其中，图4.4a为 MP3 音频的时域波形图，蓝色波形为正常音频的音频波形，红色波形为正常音频与隐写音频波形的差异；图4.4b和4.4d为分别为在梅尔滤波器阶数为 40 和 24 条件下正常音频与隐写音频 MFCC 系数差异的可视化结果；图4.4c为 MP3 音频 QMDCT 系数矩阵的可视化，白色的点表示正常音频与隐写音频的 QMDCT 系数矩阵差值不为零的点。

为了更为全面地分析输入数据类型的选择对 MP3 音频隐写分析网络性能的影响，本论文分别将以上数据输入至相同结构的卷积神经网络中，观察其对隐写

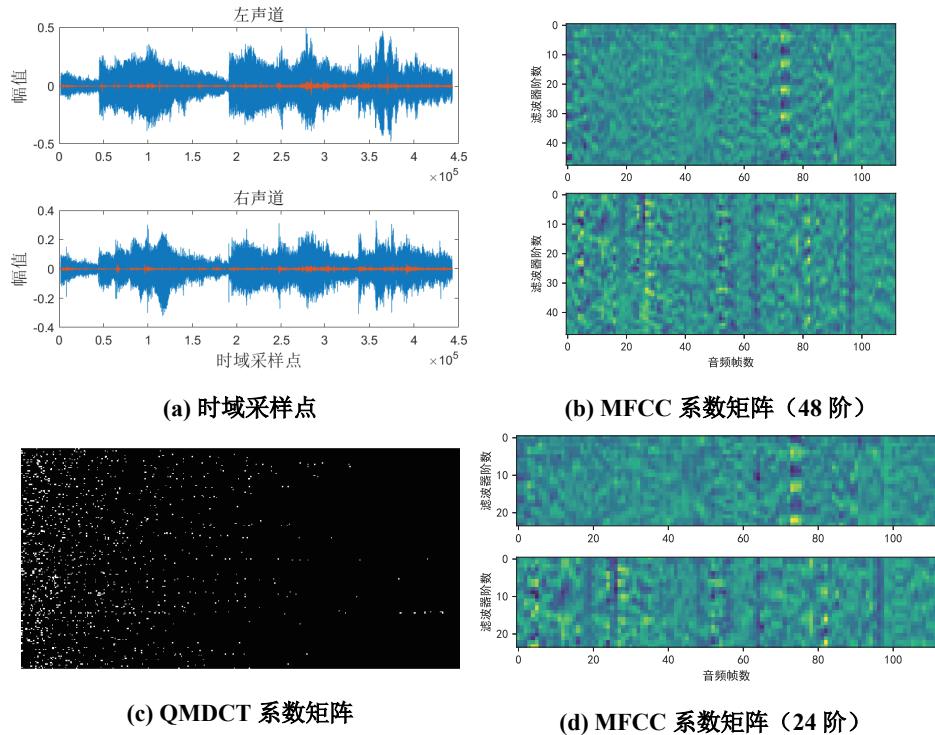


图 4.4 不同数据空间下正常音频与隐写音频的差异图 (EECS, 128kbps, SPR = 2)

Figure 4.4 Difference between cover and stego audios in different data space

(EECS, 128kbps, $SPR = 2$)

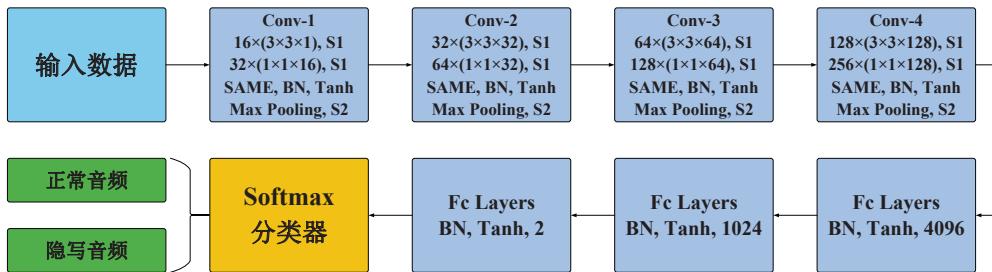


图 4.5 用于输入数据评估的网络结构 - Light-RHFCN

Figure 4.5 Network structure for evaluation of the input data - Light-RHFCN

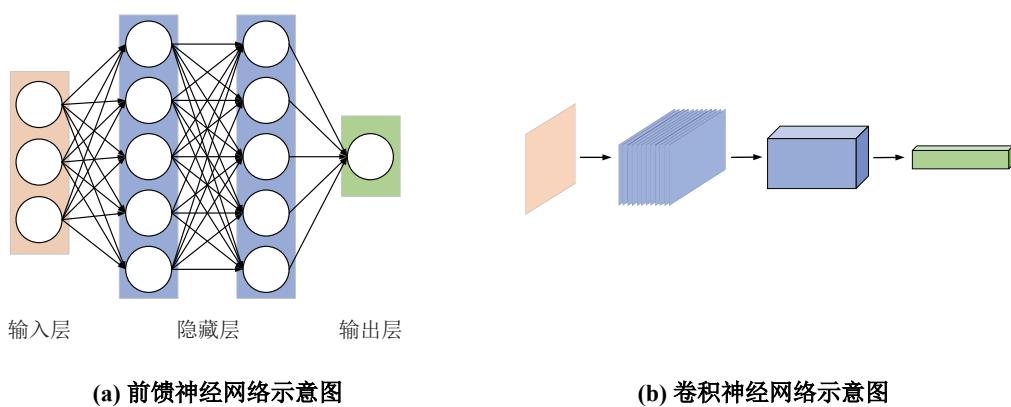
分析网络性能的影响。由于三种输入数据的尺寸不完全相同，RHFCN 网络无法完全适配。为保证训练的正常进行，本论文对网络结构进行了一定程度的调整，调整后的网络结构如图4.5所示。从表4.1中的实验结果可以看出，QMDCT 系数确实是一种良好的数据空间，更适用于 MP3 音频隐写分析。虽然隐写对 MP3 音频的时域采样点和 MFCC 系数域也会产生较大的扰动，但网络无法很好地学习到其潜在规律，这两类数据输入并不利于通用 MP3 音频隐写分析网络的设计。

表 4.1 输入数据类型对 MP3 音频隐写分析网络性能的影响 (EECS, 128kbps, SPR = 2)**Table 4.1 Evaluation of the input data type on the performance of MP3 steganalytic networks
(EECS, 128kbps, SPR = 2)**

数据类型	数据尺寸	准确率 (%)
时域采样点	200×256	66.35
QMDCT 系数矩阵	200×400	90.36
MFCC 系数矩阵 (24 阶)	112×48	76.02
MFCC 系数矩阵 (40 阶)	112×80	78.43

4.2.2 网络结构设计

如图4.6所示，卷积神经网络是一种优秀的前馈神经网络。不同的是，卷积神经网络通过局部感受野和权值共享这两大利器，大大降低了网络中参数的数量，极大地提升了神经网络的实用性，使得沉寂已久的深度学习重新焕发了光彩。卷积神经网络往往具有多个卷积层，每个卷积层的功能不完全相同。一般地，浅层卷积核主要用于对输入数据全局特征的提取，而深层卷积核则是用于对输入数据局部特征的提取，并同时对浅层特征加以整合。经过多层的特征提取、池化和跨通道信息交互等，将所提取的特征输入至 Softmax 分类器中，用于对输入数据的分类。在卷积神经网络设计过程中，卷积核的尺寸与数目、池化类型、BN 层以及激活函数类型等均会对网络的性能产生不同程度的影响。本节将讨论在 RHFCN 网络结构设计过程对以上各因素的考虑和筛选。

**图 4.6 神经网络结构示意图****Figure 4.6 Diagram of neural networks**

4.2.2.1 1x1 卷积核

1×1 卷积核引起人们的重视是源于“Network in Network”(NIN) [102] 结构，NIN 结构将传统的线性卷积核替换为多层感知机 (Multi-Layer Perceptron, MLP)，提高了网络对感受野的表达能力，从而提高了网络的性能。 1×1 卷积核可以看作是一种应用在卷积层内部的全连接层，不仅可以实现对输入特征图的比例缩放，还能以较少的参数实现特征图跨通道的信息交互和信息整合以及卷积核的升维和降维。同时，参数个数的减少也降低了网络过拟合的风险，运算复杂度也随之下降。

如图4.7所示，两个网络的输入、输出特征图维度分别为 $W \times H \times C$, $W \times H \times N$ ，其相应的参数量分别为：

$$P_1 = 3 \times 3 \times C \times M + 1 \times 1 \times M \times N = 9MC + MN \quad (4.1)$$

$$P_2 = 3 \times 3 \times C \times N = 9NC \quad (4.2)$$

通过 1×1 卷积核的引入，在输出特征图维度不改变的情况下，参数量可以减少：

$$P' = P_2 - P_1 = 9C \times (N - M) - MN \quad (4.3)$$

当 $C = 32$, $M = 64$, $N = 128$ 时, $P_1 = 26624$, $P_2 = 36864$, 参数量可以减少为以前的 70%。

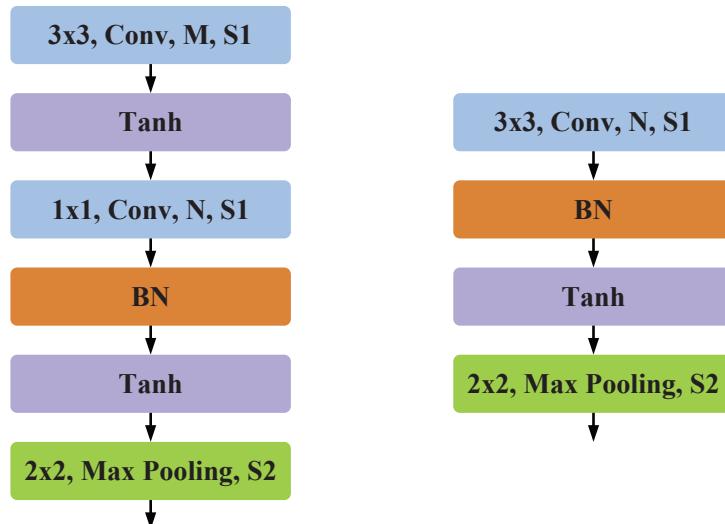


图 4.7 两种卷积模块示意图

Figure 4.7 Diagram of two different convolutional blocks

4.2.2.2 BN层

BN层的引入旨在缓解输入数据的协方差偏移（Covariate Shift）而降低梯度弥散的风险并加快神经网络的训练。在各类机器学习问题中，模型的训练与应用都基于一个经典的假设，即“源空间”和“目标空间”的数据分布是一致的。然而，对于深度卷积神经网络，数据经过各层的卷积、池化等操作后，其输入与输出的分布特性将会出现一定的差异，而且这种差异将会随着网络深度的增加而不断扩大，从而导致网络的泛化能力大大降低，网络的训练难度也会增加。此外，神经网络的训练需要大量数据的支持，而在有限的计算资源条件下需要将数据集划分多个批次输入至网络中进行训练。因此，如果每批次训练样本的分布特性不相同，神经网络还需在每次迭代时重新学习新的不同的分布，这将极大地影响网络的训练。为此，学术界先后提出多种训练数据标准化方式，如Batch Normalization（Batch方向上的标准化，BN），Layer Normalization（Channel方向上的标准化，LN），Instance Normalization（Channel内对每个特征图做标准化，IN）和Group Normalization（在Channel内分组后做标准化，GN）以及Switchable Normalization（将BN、LN、IN加以结合，并让网络学习适宜的标准化方式，SN）等，不同的标准化方式适用于不同的应用场景。在RHFCN设计中，本论文在每个卷积模块内都加入了BN层，以实现对每个Batch内训练数据的标准化。

通过BN层的引入，本论文在网络训练时可以采用更大的学习率，加快了网络的训练。同时，BN层的引入也降低了网络性能对权重和偏置参数初始化方式的依赖，同时还降低了网络过拟合的风险，能够在一定程度上取代此前用于降低网络过拟合风险的 L^2 正则化、Dropout层等策略。但值得注意的是，BN层的效果与批次大小（BatchSize）大小密切相关[103]，BatchSize越大，当前批次训练数据的分布和数据集的分布越接近，训练结果就会越好，收敛速度也会越快。当训练集或测试集的BatchSize为1时，BN层就会失效，在进行测试时尤其需要注意这一点。为此，在进行模型性能验证过程中，本论文分别根据以下两种应用场景对这一问题进行修正。（1）对于大批量样本场景下的隐写分析，本论文使用批次验证方式，保留BN层的is_training=True，即在测试数据集上保留BN层参数的计算。（2）对于单样本场景下的隐写分析，本论文有两种解决方案：（a）将单个样本的检测分析借助先验备用数据融入至批量样本的验证中，再从分析结果中取出对该数据的检测结果，在此过程中仍保留BN层的is_training为True；（b）获得任务训练过程中的BN层参数，将其直接应用于当前待检测样本的隐写分析，同时需要将BN层的is_training设置为False。

BN 层的运算公式为：

$$y = \lambda \frac{x - \mu}{\sigma} + \beta \quad (4.4)$$

其中 x 是输入张量， y 是输出张量， μ 和 σ 分别为当前批次的均值和方差，随着 BatchSize 的增大，不同批次间 μ 和 σ 的差异也将逐渐趋于减小。 λ 和 β 是缩放 (scale)、偏移 (offset) 系数，在网络训练过程中学习得到。

4.2.2.3 激活函数层

激活函数的引入旨在增加网络的非线性因素，以提升网络对各类函数的拟合能力。卷积运算的本质是线性运算，即使是多个卷积层的叠加，其最终输出结果依旧是线性的，等价为一个线性回归模型。因此，我们需要通过激活函数向网络中引入非线性因素，使网络能够逼近任意函数。激活函数通常放置在卷积层或 BN 层之后，用于对输入特征图进行非线性变换。最理想的激活函数为阶跃函数，但是由于函数不连续且不光滑，无法被用于神经网络的训练。为此，人们仿照生物特性，设计出一系列可应用于网络训练的激活函数。常用的激活函数有 Sigmoid、Tanh、ReLU、Leaky ReLU 等，其函数图像如4.8所示。不同于图像数据，QMDCT 系数矩阵中的元素同时包含正值和负值，且关于 y 轴近似对称。因此，

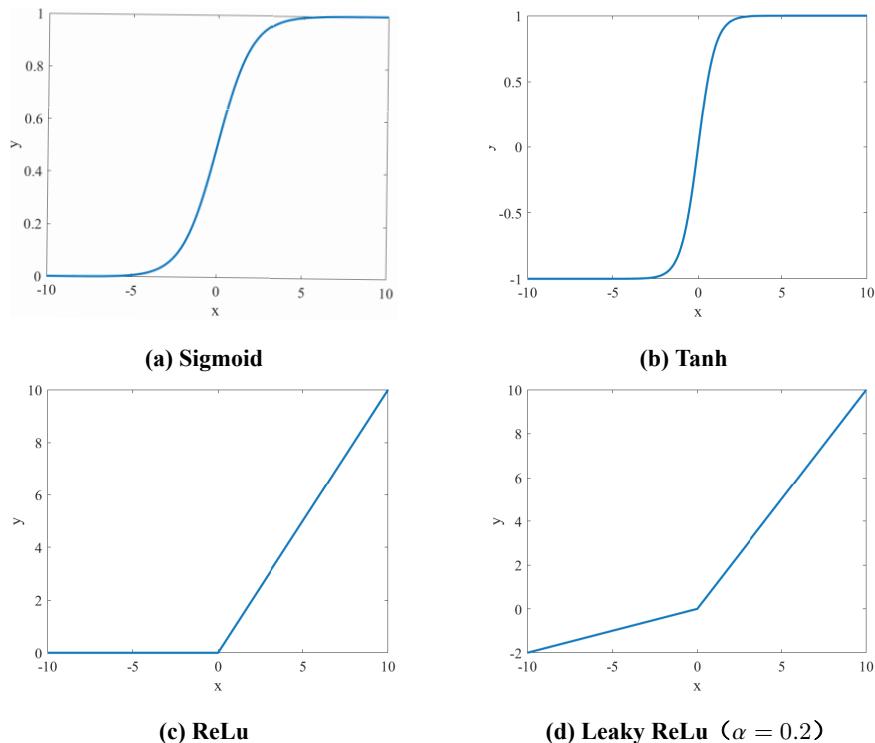


图 4.8 四种常用激活函数

Figure 4.8 Curves of four common activation functions

在 RHFCN 网络设计中使用 Tanh 激活函数函数。

4.2.2.4 池化层

池化层主要用于特征图的降采样，一方面是减少特征图的维度，降低运算复杂度；另一方面是对特征进行压缩，保留更为鲁棒的特征。常用的池化类型包括均值池化、最大池化、随机池化以及卷积池化。

通常来说，基于卷积神经网络进行特征提取的误差主要来自两个方面：（1）邻域大小受限造成的估计值方差增大；（2）卷积层参数误差造成的估计均值偏移。一般地，均值池化可以更多地保留输入数据的背景信息，可以减小第一类误差；而最大池化偏向于保留输入数据的纹理信息，可以减小第二类误差。随机池化则介于两者之间，对输入数据点按照一定的规则赋予其概率值，再根据概率大小进行下采样。从平均意义来看，与均值池化相似；从局部意义来看，与最大池化相似。与前三种池化类型不同，卷积池化则是在卷积过程中将移动步长设置为 2 或更大的数，在特征提取的同时完成降采样。由于隐写噪声往往会嵌入到纹理较为复杂的区域，因此在 RHFCN 网络设计中选择最大池化进行降采样，窗口大小与步长均为 2。

4.2.2.5 正则化

正则化的引入主要是降低网络过拟合的风险。机器学习的核心问题是设计不仅在训练数据集上表现好，而且在新输入上泛化好的算法 [104]。模型在训练数据集上的误差为训练误差，在测试数据集上的误差为测试误差。如果模型的训练误差的测试误差相差过大，则模型过拟合；如果模型训练误差和测试误差均较大，则模型欠拟合。只有当模型的训练误差和测试误差相当，且均在可接受的范围内，这样的模型才是我们需要的。在网络设计中，由于数据集的规模不够大，我们往往面对的都是网络过拟合的问题。此前介绍的 BN 层和 1×1 卷积核都可在一定程度上降低模型过拟合的风险。此外，正则化也是一种常用的策略。在网络训练过程中，本论文使用 L^2 正则化对权重进行参数范数惩罚。

4.2.2.6 全卷积网络结构

在卷积神经网络中，卷积层用于特征提取，将原始数据映射至特征空间；全连接（Fully Connected，FC）层则是用于实现对特征图信息的整合，将学习到的特征映射至标签空间，起到“分类器”的作用。然而，值得注意的是，特征图在被输入至全连接层之前会被拉伸为一个一维向量，虽然输入特征图的数值并未

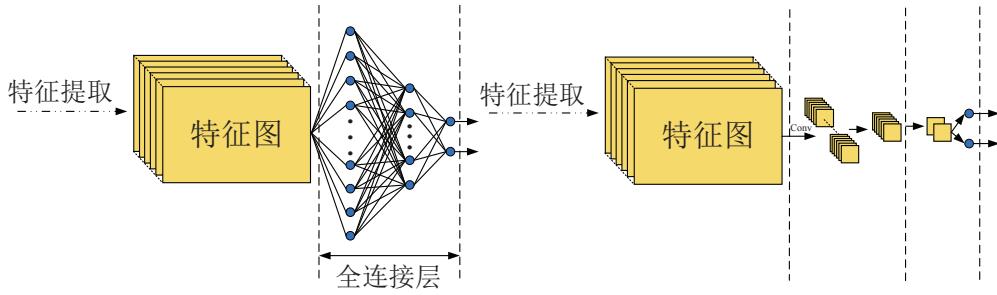


图 4.9 带有全连接层的网络结构与全卷积网络结构比较示意图

Figure 4.9 Comparison of the network with fc layers and the fully convolution network

发生变化，但是其特征图结构和空间关系被破坏，这部分信息便很难被网络利用，影响网络的精度。虽然当全连接层神经元个数足够多时，此类位置关系信息也可以被学习到，但却需要消耗巨大的计算资源。因此，在 RHFCN 结构设计中使用大尺寸的卷积层来替代全连接层，如图4.9所示。在 Conv-6 中，卷积核的尺寸为 6×12 ，与输入特征图的尺寸保持一致（输入数据为 200×400 的系数矩阵条件下）。因此，输出的特征图尺寸为 1×1 ，再使用两层 1×1 卷积层实现对通道数的降维。最终，将 $1 \times 1 \times 2$ 的特征图变换为 1×2 的向量，输入至 Softmax 分类器中完成分类。Softmax 分类器之前的全局最大池化用于对尺寸大于 200×400 的输入数据进行维度归一化，用于接收大于 200×400 的任意尺寸数据的输入。

4.2.3 面向输入数据尺寸失配的音频隐写分析

本论文在3.3.3节讨论了不同长度的音频段对 MSC 隐写分析算法性能的影响，本节将讨论 RHFCN 在面向输入数据尺寸失配时的解决能力。基于卷积神经网络的隐写分析算法与基于手工特征设计的隐写分析算法较大的不同在于对输入数据尺寸变化的敏感性。卷积层对输入特征图的尺寸是不敏感的，只要输入特征图的尺寸大于卷积核的尺寸，卷积运算就可以正常运行。但是，由于全连接层的神经元个数是固定的，特征图与神经元之间的连接数也因此固定。当训练数据的尺寸与测试数据的尺寸不一致时，模型便无法正常被调用。然而，基于手工特征的隐写分析算法对输入数据的尺寸变化并不敏感，即使训练数据与测试数据的尺寸相差很大，只要特征能够正常提取便可完成测试，但精度可能会受到不同程度的影响。

在表情识别、物体识别等图像分类任务中，对于尺寸不匹配的输入数据，常用的解决方法有裁剪 (Cropping)、大小调整 (Resize) 和降采样 (DownSampling)。然而，这三种方法在隐写分析中并不完全适用。由于上述提到的图像分类任务

表 4.2 不同尺寸输入数据的检测准确率 (EECS, 128kbps, SPR = 2)

Table 4.2 Detection accuracies under different sizes of the input data
(EECS, 128kbps, SPR = 2)

	400	430	450	480	500	530
200	95.40	95.64	92.39	89.78	86.37	83.62
250	95.69	95.54	94.43	90.94	90.16	85.07
300	96.61	96.03	94.67	91.20	88.81	82.46
350	97.04	96.17	94.04	89.00	88.13	81.40
400	98.31	96.90	94.04	87.21	85.61	80.60

是基于图像内容进行的，只要待检测的目标相对完整就可以完成分类。但是，隐写分析则是基于媒体数据进行的，且嵌入消息的位置相对随机，裁剪、降采样和大小调整后均有可能造成隐写信号的丢失，影响最终的检测效果。为了解决这一问题，本论文在网络设计过程中借鉴了全卷积网络 [105] 的设计思路，去除了网络的全连接层，并在网络的最后增加了全局最大池化层，实现在测试过程中任意尺寸数据的输入，并能在一定范围内保证隐写分析的准确率。本章在进行 MP3 音频隐写分析时以 50 帧为一个检测单元，然而现实中的 MP3 音频大多在 3-4 分钟左右。因此，在面向真实环境的 MP3 音频隐写分析中，首先需要对提取的 QMDCT 系数矩阵进行裁剪，然后对每个片段单独进行隐写分析，最后对其综合判决，如 4.10 所示。在裁剪过程中，裁剪尺寸是一个着重考虑的问题。为了测试 RHFCN 对不同尺寸输入数据的检测效果，本论文首先使用尺寸为 200×400 的 QMDCT 系数矩阵进行网络的训练，再将其直接应用至尺寸为 200×450 、 250×450

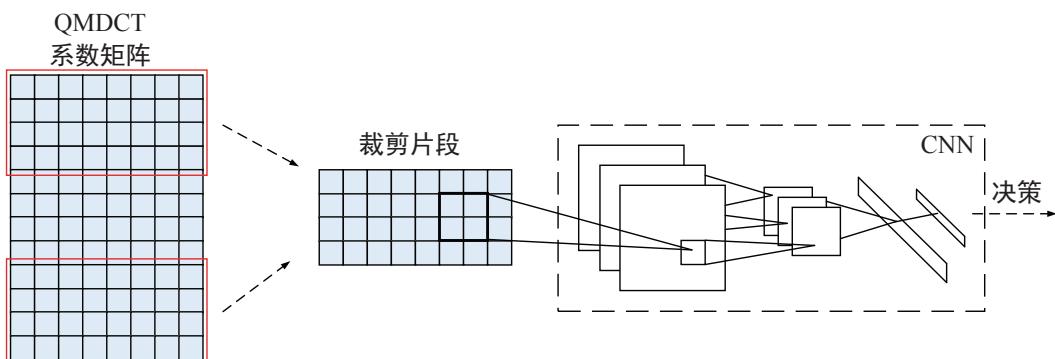


图 4.10 基于滑动窗口的 MP3 音频隐写分析示意图

Figure 4.10 Diagram of MP3 audio steganalysis based on sliding windows

以及 200×500 等其他尺寸的 QMDCT 系数矩阵的测试中，以上所有尺寸的输入数据均是从 400×576 的矩阵中裁剪出来的，实验结果如表4.2所示。

从实验结果可以看出，QMDCT 系数矩阵高度的增加有助于检测准确率的上升；而 QMDCT 系数矩阵宽度的增加则会导致准确率的下降，主要体现在虚警率的上升。这是因为，当 QMDCT 系数矩阵高度增加时，大值区系数增多，隐写的秘密消息也将随之增多，正常音频与隐写音频的差异增大，因此检测准确率会呈现上升趋势。根据第二章中对 QMDCT 系数矩阵的介绍可知，随着音频比特率的增加，零区的起始索引值将逐渐加大，对于比特率为 128kbps 的 MP3 音频，其零区的起始索引值大约在 410 左右。因此，随着 QMDCT 系数矩阵宽度的增加，正常音频与隐写音频的差异先基本保持不变或增加再减少，准确率因此也呈现出相似的变化趋势。QMDCT 系数矩阵宽度的增加虽然会导致准确率的下降，但是基于对计算代价和检测性能的考虑，每帧内“多余”的 QMDCT 系数都将会被丢掉。因此，由于 QMDCT 系数矩阵宽度增加而带来的精度损失可以被忽略。只要输入数据的宽度不超过 500，检测准确率均在 85% 以上，基本满足实际应用需求。

综上所述，RHFCN 可以较为稳定地实现输入数据尺寸失配条件下的隐写分析，而且网络能够在 QMDCT 系数矩阵尺寸增加而隐写信息比例没有极速下降的条件下保持检测精度不减少甚至增加，很好地应对了真实场景下的隐写分析应用需求。

4.2.4 基于隐写负载率的迁移学习

迁移学习（Transfer Learning）就是将已经训练好的模型参数应用到新任务的模型训练中，从而提升新任务的建模速度并提高其性能。一般地，现实生活中

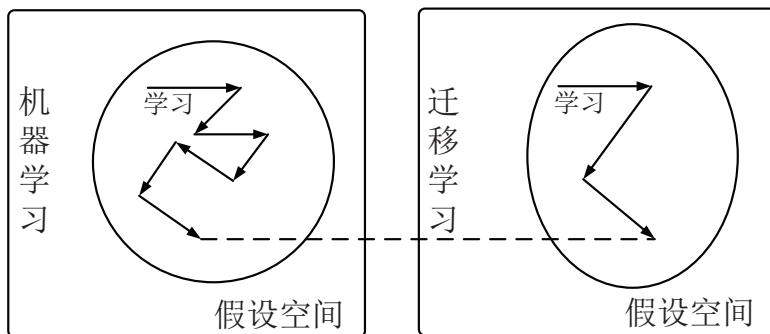


图 4.11 机器学习与迁移学习路径搜索比较图

Figure 4.11 Comparison of conventional machine learning and transfer learning

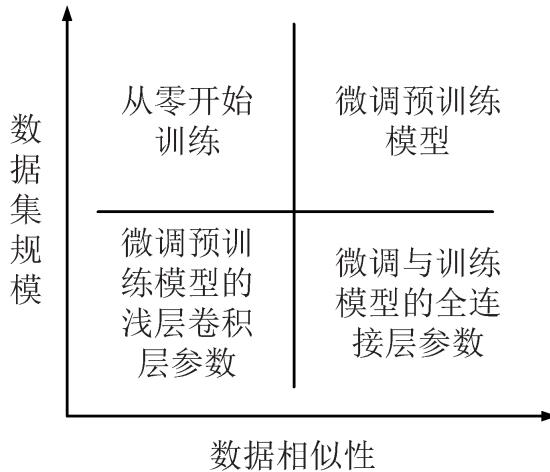


图 4.12 迁移学习分类图

Figure 4.12 Classification of transfer learning

大部分数据和任务都存在一定的相关性，通过迁移学习可以提升新模型的学习效率，避免了由于从零学习而导致的训练周期长和训练难度大等问题。如图4.11所示，相比于从零开始的机器学习，迁移学习可以更快地到达任务所要求的目标点，学习过程更简短。

根据训练数据规模和源数据与目标数据分布特性的相似程度，基于卷积神经网络的迁移学习可以分为四类，如图4.12所示。预训练，就是指预先训练的一个模型或预先训练模型的过程。微调，就是将预训练过的模型作用于自己的数据集，并使参数适应自己数据集的过程。数据集规模相差越大，数据分布的相似性相差越大，迁移学习的难度也越大，需要重新训练的参数也就越多。以图像分类为例，浅层卷积核学习的是图像的全局信息，即图像的颜色、边缘、角点信息等；深层卷积层学习的是图像的局部信息，即当前数据集特有的潜在特性。因此，在数据集规模一定的情况下，如果源数据与目标数据的相似性较高时，只需重新训练其全连接层参数；而随着数据相似性的下降，需要对其更多浅层卷积核进行微调。

影响 MP3 音频隐写分析算法性能的因素主要有隐写算法类型、音频比特率和隐写负载率等。从图4.13可以看出，在隐写算法及音频比特率相同的情况下，在不同的隐写负载率条件下，MP3 音频 QMDCT 系数矩阵中被修改的系数呈现出相似的分布特性。因此，本论文提出一种基于隐写负载率的迁移学习方案，将高隐写负载条件下训练得到的模型用于低隐写负载样本的训练与测试，加快网络的收敛速度，以降低低隐写负载条件下神经网络的训练难度。

为了证明基于隐写负载率的迁移学习的有效性，本论文将 EECS 隐写算法在

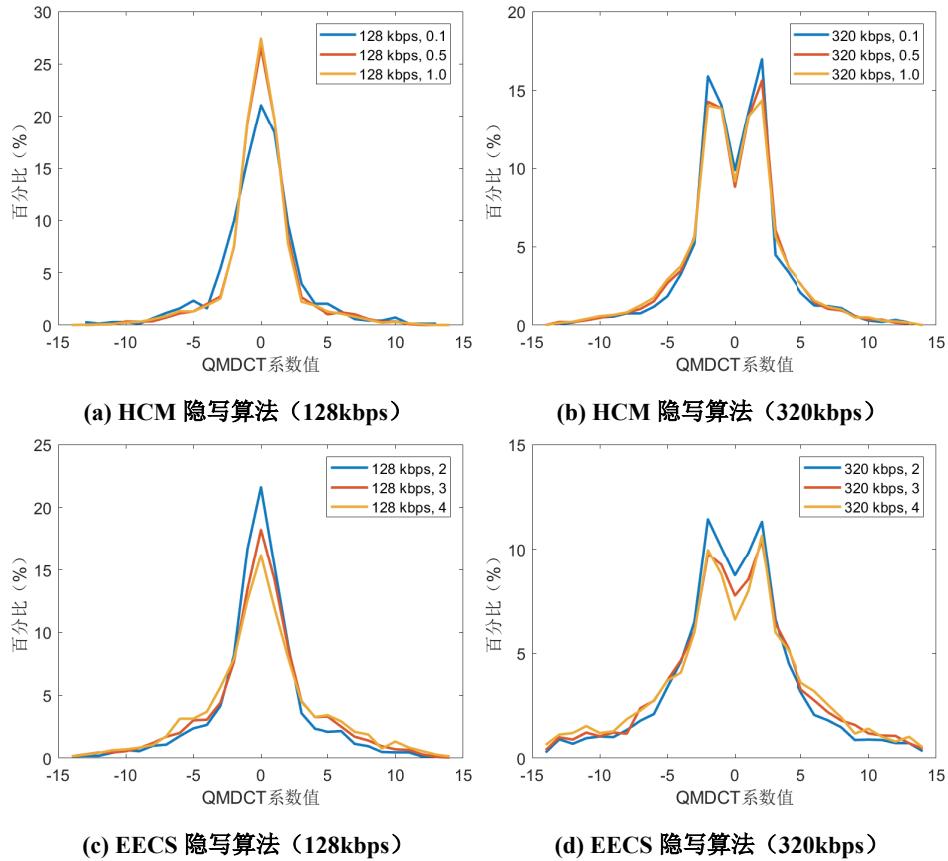


图 4.13 修改的 QMDCT 系数在不同隐写条件下的分布图

Figure 4.13 Distributions of modified coefficients under different steganographic algorithms and payloads

比特率为 128kbps、隐写负载为 $SPR = 2$ 条件下训练得到的模型，以迁移学习和迁移测试两种方式分别应用于 $SPR = 3, 4, 5$ 三种隐写负载下的检测分析，实验结果如表4.3所示。从实验结果可以看出，如果直接将由高隐写负载样本训练得到的模型直接用于低隐写负载样本的检测分析，则会有较大的精度损失，准确率

表 4.3 基于隐写负载率的迁移学习隐写分析结果（EECS, 128kbps）

**Table 4.3 Detection performance of transfer learning based on steganographic payloads
(EECS, 128kbps)**

SPR	从零学习		迁移学习		迁移测试	
	准确率	训练周期	准确率	训练周期	准确率	训练周期
3	89.65	30	89.80	6	86.24	-
4	82.43	42	82.18	13	77.76	-
5	77.80	55	76.36	26	70.90	-

表4.4 实验设置

Table 4.4 Settings of experiments

项目	参数	值
音频信息	帧数	50
	比特率	128 / 192 / 256 / 320 kbps
隐写负载	HCM	$SPR = 0.1 / 0.3 / 0.5 / 0.8 / 1.0$
	MP3Stego	
EECS		$SPR = 2 / 3 / 4 / 5 / 6$
	批次大小	16 (8 正常/隐写音频对)
优化器		Adam [106] ($\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$)
	超参数	
初始学习率		0.001
	学习率更新策略	指数下降 (5000, 0.9)
L^2 正则化系数		0.001
	参数初始化方式	Xavier [107]

将下降3-8个百分点，虽不失为一种折中的方法，但需要考虑应用场景的精度需求。如果将由高隐写负载样本训练得到的模型用于低隐写负载样本的训练，不仅会使得网络更快得收敛，还可在一定程度上提高网络的性能。但需要注意的是，在进行迁移学习时，由于原始模型已经经过多轮学习达到了相对较高的精度，因此应适当降低初试学习率以避免精度的损失。此外，如前文所述，由于浅层卷积核学习的是全局特征，而深层网络学习的是局部特征。因此，可以根据隐写负载的不同，分别固定不同层数的浅层卷积核参数，以减少不必要的训练。

4.3 实验设计与分析

本节将通过实验证明RHFCN结构的有效性。首先，分别讨论卷积核尺寸及大小、激活函数类型，有无富高通滤波模型以及有无BN层等结构变化对网络性能的影响。其次，分别讨论DenseNet、LSTM等网络结构是否适用于基于QMDCT系数矩阵的MP3音频隐写分析。最后，与多种性能较好的MP3音频隐写方法进行对比，以评估各分析算法在不同隐写算法、音频比特率及隐写负载率

条件下性能的优劣。

4.3.1 实验设置

为了全面地评估 RHFCN 网络结构的性能，本论文从数据库中选取了各参数条件下的音频对 33038 个用于实验验证，其中 19200 对用于训练，12800 对用于验证，剩余的 1038 对用于和 ADOTP、MDI2 隐写分析算法进行比对，其余各参数设置如表4.4所示。实验过程中使用 *FPR*、*FNR*、*ACC* 三个指标度量隐写分析算法的性能。测试过程中使用批量测试方式，BN 层在测试阶段保持开启状态，而每次测试过程中数据都会被置乱，每个批次的数据组合均会有所变化，每次测试结果也会有所不同。因此，最终的实验结果取 10 次实验的算术平均值，以减少偶然误差带来的影响。

4.3.2 网络结构优选

由于网络的深度、卷积核的尺寸与个数、激活函数类型、池化方式等均会对最终的网络性能产生不同程度的影响，为了使得网络的性能达到最优，本论文在网络设计过程对其结构进行微调。从表4.5中的实验结果可以看出：

1. 富高通滤波模型的引入可以提升网络的性能，但是与其在传统手工特征设计中的作用相比，显著性有所下降。一方面，神经网络本身可以学习到类似于高通滤波器的残差提取功能；另一方面，QMDCT 系数矩阵本身可以提供较为丰富的信息，特征提取对数据预处理的需求有所下降。但是，网络自身的学习能力依旧是有限的，富高通滤波模型有助于增强网络对隐写信号的敏感性，不仅能够提升网络的性能，同时还可以加快网络的收敛速度。

2. BN 层的引入极大地优化了网络的性能，不仅提升了网络的检测精度，还有效克服了隐写分析网络难以训练的问题。在实验过程中还发现，移除了 BN 层的卷积神经网络在训练过程中较为不稳定，网络在达到局部最优后可能又会重新回到动荡的状态，需要更为精细地对学习率进行调节。

3. 1×1 卷积核的引入不仅减少了模型的参数，降低了网络训练中过拟合的风险，同时也较为明显地提升了网络的检测精度。

4. 池化类型的选择对网络的检测效果影响较小，但是通过实验发现，使用最大池化的网络收敛速度更快。

5. 大尺寸的卷积核虽然可以获得更大的感受野，但同时也增加了感受野内的不相关信息，反而会降低网络的性能。此外，由于卷积核尺寸的增加，连接数和参数个数也随之增长，网络的收敛速度明显减慢。

6. 在激活函数的选择上，相比于 ReLu 和 Leaky ReLu，在当前网络结构中使用激活函数 Tanh 可以获得更好的检测效果，本论文认为这是由于 QMDCT 系数矩阵中同时包含正值和负值以及 QMDCT 系数分布的对称导致的。

7. 全连接层的移除提升了网络的检测效果。

8. 深度学习相比于普通的前馈神经网络，其最大的特色在于网络深度的增加。然而，随着隐藏层的增多，梯度消失的问题也逐渐凸显，反而降低了网络的性能。从表中的实验结果可以看出，简单的卷积层堆叠并不利于网络性能的提升，会引起网络的退化。

9. DenseNet[89] 虽然在图像分类中的表现非常优异，但并不完全适用于 MP3 音频隐写分析，且由于网络结构变得更为负载，相对于 VGG 类型的网络结构训练难度更大。基于 DenseNet 的隐写分析精度下降也体现出正如奥卡姆剃刀原理所表达的“如无必要，勿增实体”。

表 4.5 不同网络结构下的隐写分析准确率 (EECS, 128kbps, SPR = 2)

Table 4.5 Detection performance of different network structure (EECS, 128kbps, SPR = 2)

序号	网络结构微调	准确率 (%)
1	本章提出的网络	95.40
2	移除富高通滤波模型	93.52
3	移除全部 BN 层	74.20
4	移除 1×1 卷积核	88.87
5	将池化类型改为平均池化	94.60
6	将卷积核尺寸由 3×3 替换为 5×5	92.95
7	将激活函数由 Tanh 替换为 ReLu	94.21
8	将激活函数由 Tanh 替换为 Leaky ReLu, $\alpha = 0.2$	94.73
9	取消移除全连接层	93.85
10	加深网络的深度	93.65
11	使用 18 层 Dense-Net 结构	-
12	使用 42 层 Dense-Net 结构	-
13	使用 LSTM 结构 (1024 个神经元)	62.35
14	使用双层 LSTM 结构 (1024 个神经元)	63.12
15	使用 GRU 结构 (1024 个神经元)	60.54

10. 由于 MP3 音频信号为时序信号，本论文也分别尝试了基于 QMDCT 系数与长短期记忆网络（Long-Short Term Memory, LSTM）[108] 和门控循环单元（Gated Recurrent Unit, GRU）[109] 的隐写分析算法。然而，实验结果证明，基于 QMDCT 系数矩阵和 RNN 结构的隐写分析网络效果欠佳，这可能是由于 QMDCT 系数矩阵不具备理想的时序性导致的，也间接印证了 QMDCT 系数矩阵更适合用于基于 CNN 的隐写分析。

4.3.3 实验结果与分析

相比于传统的手工特征设计，本章所提的卷积神经网络 RHFCN 在各类隐写算法、各种音频比特率以及各种隐写负载下的检测准确率均有了明显的提升，已完全实现了对 MP3Stego、HCM 隐写算法的检测分析，对 EECS 隐写算法的检测效果相较于 ADOTP 和 MDI2 算法也有非常显著的提升。RHFCN 的优越性还表现在低隐写负载条件下（如 EECS, 128kbps, $SPR = 3, 4, 5$ ）的检测分析，相比于第三章提出的 MSC 算法，准确率也至少提高 10% 以上。除 128kbps 外，对其余音频比特率隐写音频的检测准确率均在 90% 以上，基本突破了对 EECS 隐写算法的检测分析。此外，从实验结果也可以看出，在基于 RHFCN 的隐写分析中，算法的虚警率和漏检率基本相当。

通过和附录 A 中的隐写负载率转换表进行比对发现，对于 EECS 隐写算法，当 QMDCT 系数矩阵中被修改的点的比例，即相对嵌入率，少于 2% (128kbps, $SPR = 3 \rightarrow 4$) 时，RHFCN 对隐写信号的敏感性将会有较大程度的衰退，准确率将下降 7 个百分点，但与音频比特率的相关性较低，这与第三章提出的 MSC 隐写分析算法有较大的不同。因此，未来 MP3 音频隐写分析算法的设计方向就在于提升相对负载率少于 2% 时的检测准确率提升。

通过对两类隐写分析算法，本论文认为基于手工特征设计的隐写分析算法对经验的要求较高，在设计过程中可能往往只关注到隐写噪声的单一方面的特性，如空间相关性的变化、纹理信息的变化等，对隐写信号的度量不够全面，导致检测精度不足。而基于卷积神经网络的隐写分析算法可以在多个层面、多个角度提取隐写噪声的特性，从而更好地实现对隐写音频的检测。更为重要的是，基于深度学习的隐写分析算法可以实现特征地自主学习和提取，大大简化了隐写分析算法的设计流程。

表4.6 对 MP3Stego 算法的隐写分析结果

Table 4.6 Performance of each steganographic algorithm on the detection of MP3Stego

比特率	SPR	ADOTP			MDI2			RHFCN		
		FPR	FNR	ACC	FPR	FNR	ACC	FPR	FNR	ACC
128	0.1	11.57	13.21	87.61	7.98	10.70	90.66	5.32	6.44	94.12
	0.3	9.33	9.21	90.73	1.47	14.29	92.12	4.05	3.33	96.31
	0.5	6.47	11.06	91.24	0.97	11.92	93.56	1.56	1.14	98.65
	0.8	1.03	12.68	93.15	1.93	10.36	93.86	0.41	0.83	99.38
	1.0	1.83	10.06	94.06	1.40	9.68	94.46	0.00	0.00	100.0
192	0.1	10.25	9.33	90.21	11.43	5.41	91.58	4.82	3.66	95.76
	0.3	12.54	2.09	92.68	7.74	5.14	93.56	2.78	3.54	96.84
	0.5	6.92	4.04	94.52	6.17	4.68	94.58	0.76	0.96	99.14
	0.8	4.16	3.50	96.17	4.38	3.02	96.30	0.00	0.00	100.0
	1.0	2.38	2.02	97.80	4.98	2.12	96.45	0.00	0.00	100.0
256	0.1	10.81	5.37	91.91	9.45	3.61	93.47	3.55	3.81	96.32
	0.3	5.58	5.42	94.50	5.47	5.53	94.50	1.24	0.94	98.91
	0.5	5.08	2.05	96.44	6.47	3.02	95.26	0.00	0.00	100.0
	0.8	4.88	2.05	96.54	3.77	4.79	95.72	0.00	0.00	100.0
	1.0	2.48	1.52	98.00	5.64	1.46	96.45	0.00	0.00	100.0
320	0.1	6.51	6.61	93.44	5.99	6.81	93.60	1.82	2.20	97.99
	0.3	4.09	3.41	96.25	4.79	5.66	94.78	0.60	0.60	99.40
	0.5	2.53	3.47	97.00	5.12	3.78	95.55	0.00	0.00	100.0
	0.8	2.46	2.03	97.76	4.25	2.80	96.48	0.00	0.00	100.0
	1.0	1.54	0.81	98.83	2.06	4.37	96.79	0.00	0.00	100.0

表 4.7 对 HCM 算法的隐写分析结果

Table 4.7 Performance of each steganographic algorithm on the detection of HCM

比特率	SPR	ADOTP			MDI2			RHFCN		
		FPR	FNR	ACC	FPR	FNR	ACC	FPR	FNR	ACC
128	0.1	44.26	42.09	56.83	38.32	42.23	59.73	8.74	8.10	91.58
	0.3	33.48	30.12	68.20	28.64	29.88	70.74	8.41	7.35	92.12
	0.5	23.26	18.77	78.98	17.61	15.60	83.39	2.79	2.00	97.61
	0.8	9.82	7.29	91.44	7.91	6.69	92.70	0.33	0.05	99.81
	1.0	4.81	3.84	95.67	4.28	2.53	96.60	0.14	0.11	99.88
192	0.1	39.96	36.14	61.95	35.21	31.80	66.49	4.97	4.85	95.09
	0.3	24.46	22.16	76.69	21.82	21.20	78.49	1.53	1.67	98.40
	0.5	12.80	9.88	88.66	10.47	11.46	89.03	0.20	0.14	99.83
	0.8	4.20	3.65	96.08	3.24	3.14	96.81	0.00	0.12	99.94
	1.0	1.84	1.54	98.31	1.76	1.35	98.44	0.00	0.00	100.0
256	0.1	38.08	34.99	63.47	23.99	21.04	77.49	2.13	2.73	97.57
	0.3	20.59	19.14	80.14	15.17	12.09	86.37	1.85	1.89	98.13
	0.5	9.11	7.99	91.45	6.73	6.03	93.62	0.14	0.18	99.84
	0.8	3.49	2.84	96.84	1.71	1.48	98.41	0.00	0.00	100.0
	1.0	1.56	1.36	98.54	0.94	0.74	99.16	0.00	0.00	100.0
320	0.1	33.16	30.68	68.08	10.02	12.99	88.49	1.01	1.41	98.79
	0.3	19.95	19.32	80.36	7.74	9.13	91.57	0.00	0.38	99.81
	0.5	11.47	12.17	88.18	4.17	4.64	95.59	0.10	0.03	99.93
	0.8	1.76	2.11	98.06	1.43	1.35	98.61	0.00	0.00	100.0
	1.0	0.84	0.11	99.53	0.66	0.90	99.22	0.00	0.00	100.0

表4.8 对EECS算法的隐写分析结果

Table 4.8 Performance of each steganographic algorithm on the detection of EECS

比特率	SPR	ADOTP			MDI2			RHFCN		
		FPR	FNR	ACC	FPR	FNR	ACC	FPR	FNR	ACC
128	2	31.01	27.01	70.99	28.05	29.27	71.34	4.04	5.16	95.40
	3	41.76	36.22	61.01	38.45	37.69	61.93	11.21	9.49	89.65
	4	44.49	39.73	57.89	42.02	43.01	57.48	17.10	18.05	82.43
	5	45.48	44.35	55.09	46.63	44.44	54.47	22.53	21.87	77.80
	6	48.52	43.75	53.86	50.86	45.74	51.70	25.34	25.67	74.50
192	2	26.77	24.10	74.56	25.11	25.69	74.60	1.41	2.53	98.03
	3	40.27	35.74	61.99	38.20	37.74	62.03	4.10	5.60	95.15
	4	43.84	39.51	58.32	44.16	41.66	57.09	6.45	6.55	93.50
	5	47.79	41.73	55.24	44.40	45.80	54.90	13.15	4.45	91.20
	6	46.18	45.16	54.33	47.69	46.64	52.84	16.70	17.85	82.73
256	2	26.76	24.08	74.58	19.21	21.24	79.78	0.59	1.17	99.12
	3	37.71	36.84	62.72	36.38	30.66	66.48	3.51	3.63	96.43
	4	42.76	39.40	58.92	43.95	38.02	59.01	4.35	4.17	95.74
	5	45.89	42.25	55.93	45.04	40.85	57.06	7.52	7.92	92.28
	6	46.66	44.85	54.24	47.42	42.61	54.98	11.60	11.05	88.68
320	2	26.96	26.21	73.41	22.53	17.58	79.95	0.00	0.10	99.95
	3	37.95	38.59	61.73	31.46	35.00	66.77	3.00	2.98	97.01
	4	42.74	38.76	59.25	35.94	40.38	61.84	3.25	3.67	96.54
	5	45.98	42.48	55.77	41.45	40.84	58.86	6.70	6.86	93.22
	6	46.74	44.32	54.47	43.44	43.89	56.34	8.67	9.83	90.75

4.4 本章小结

本章提出一种基于 QMDCT 系数矩阵和 CNN 的 MP3 音频隐写分析网络 RHFCN，有效地实现了深度学习技术与 MP3 音频隐写分析的有机结合，与基于传统手工特征设计的 MP3 隐写分析算法相比，检测性能得到了显著提升，基本实现了对现阶段已有 MP3 音频隐写算法的检测分析。虽然 96kbps 的 MP3 音频并不常见，但是出于对算法泛化性的考虑，未来也将进一步对低比率音频的 MP3 隐写分析进行探索。

由于输入数据类型、神经网络类型以及神经网络结构对网络性能均会产生不同程度的影响，在对以上因素进行详细分析与讨论后，本章提出一种基于 VGG 架构的全卷积神经网络结构。RHFCN 不仅能够很好地应用于多种 MP3 音频隐写算法、多种比特率以及多种隐写负载条件下的检测分析，还可以在一定程度上解决输入数据尺寸失配的问题，可以有效应对现实场景中任意时长 MP3 音频的隐写分析。

同时，本章还提出了一种基于迁移学习的低隐写负载音频的网络训练方案，即，将高负载隐写条件下训练得到的模型应用于低隐写负载条件下的检测分析，大大缩短了网络的训练周期，有效提升了算法的实用性。

此外，相比于传统手工特征设计，基于 CNN 的 MP3 隐写分析网络虽然在训练阶段对数据资源及计算资源的依赖性较高，但在实际的隐写分析中，其检测速度并不亚于其他各类传统机器学习算法，能够很好地部署于各类检测系统。

第5章 基于多尺度卷积神经网络的MP3音频隐写分析研究

5.1 引言

神经网络结构的优化具体表现为网络性能的提升和网络效率的优化。随着网络层数的逐渐增加，网络结构的日趋复杂，网络的性能也得以逐步提升。随之而来的就是网络的效率问题，主要表现为网络参数的轻量化以及网络训练与测试速度的快速化。网络效率的解决能够让深度学习技术更好地走出实验室，更广泛地应用于移动端设备。SqueezeNet[110]、MobileNet[111]、ShuffleNet[112]、Xception[113]是近年来被提出的四种经典的轻量化网络结构，旨在通过设计更高效的网络计算方式减少网络的参数。另一方面，文献[114]中提出，在不降低网络性能的情况下，臃肿的稀疏网络结构可以被简化。2013年Lin[102]提出NIN，通过在网络内部构建微型网络的方式，即以多层感知机取代单层卷积层，如图5.1所示，增强了模型在感受野内对局部区域的识别能力，大大提升了网络的性能。随后，Google也提出了经典的多尺度卷积神经网络结构GoogleNet[87]，并对其在性能和参数两个方面不断改进，先后发布了GoogleNet-Inception v2[115]、v3[116]、v4[117]，性能及结构日趋完善。

本章的贡献在于提出一种基于多尺度卷积神经网络的MP3音频隐写分析网络，以下简称RHMSCN (Rich High-Pass Multi-Scale Convolutional Network)。在网络设计中分别引入了多尺度卷积、卷积核分解以及残差连接结构，有效提升了网络的性能，并大大减少了网络的参数，从计算资源消耗和模型存储两个方面提升了网络的实用性。同时，本章还提出了一种基于隐写算法及隐写负载率未知条件下的MP3音频隐写分析方案，有效推动了基于深度学习的MP3音频隐写分析

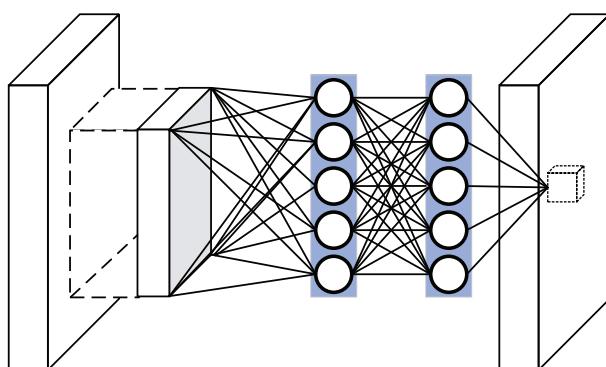


图5.1 以多层感知机为子网的网络结构

Figure 5.1 Network structure with the sub-network of multilayer perceptron

算法走向实际应用。

本章后续的内容组织安排如下：5.2节介绍算法设计原理，5.3节是实验设计与分析，5.4节是对本章工作的小结。

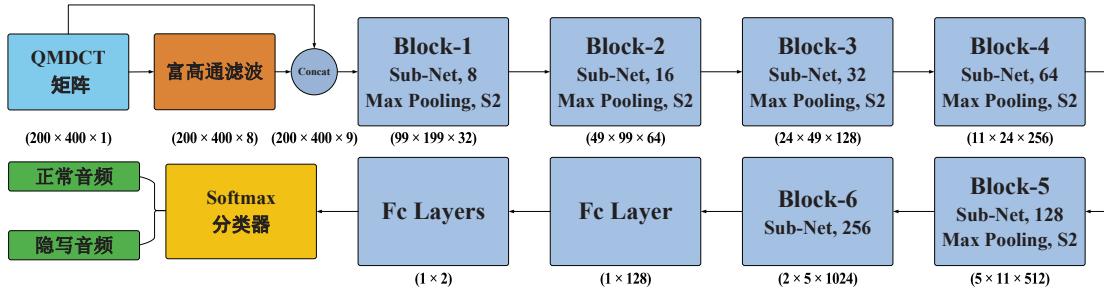


图 5.2 RHMSCN 结构示意图。“Sub-Net, 8” 表示当前子网内的各尺寸卷积核个数均为 8，“S2” 表示步长为 2，各方框下为输出特征图的维度)

Figure 5.2 Diagram of RHMSCN. "Sub-Net, 8" represents the kernel numbers in the current sub-network is 8. "S2" means the stride is 2. And, the dimension of feature maps is shown below each block.

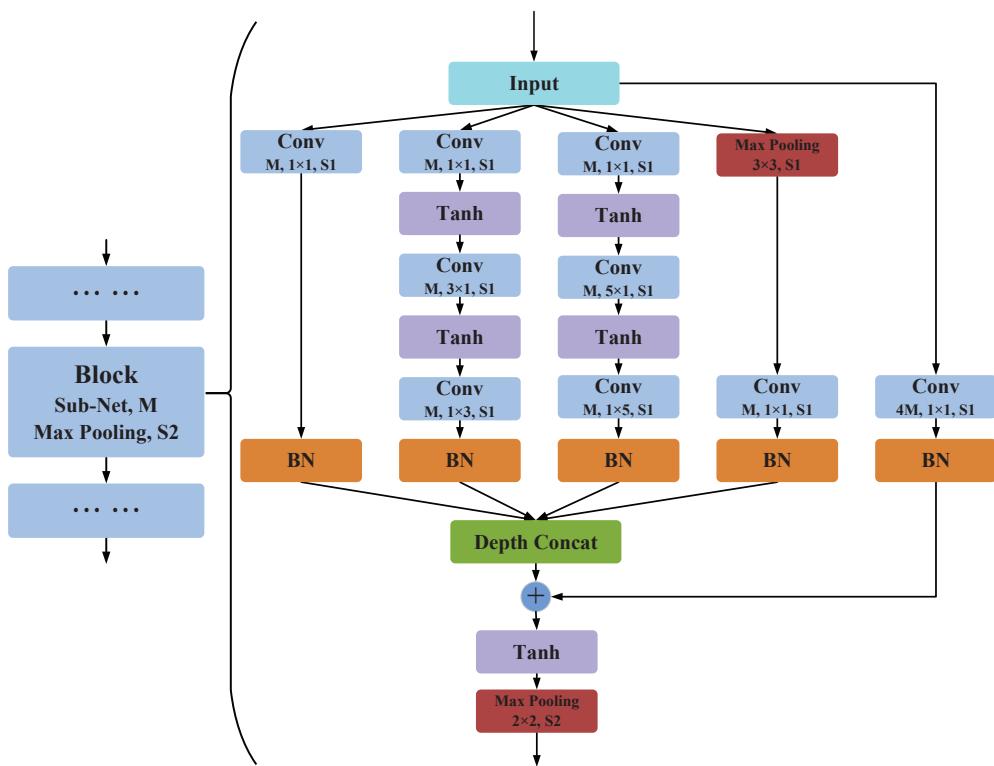


图 5.3 子网结构示意图

Figure 5.3 Diagram of the sub-network

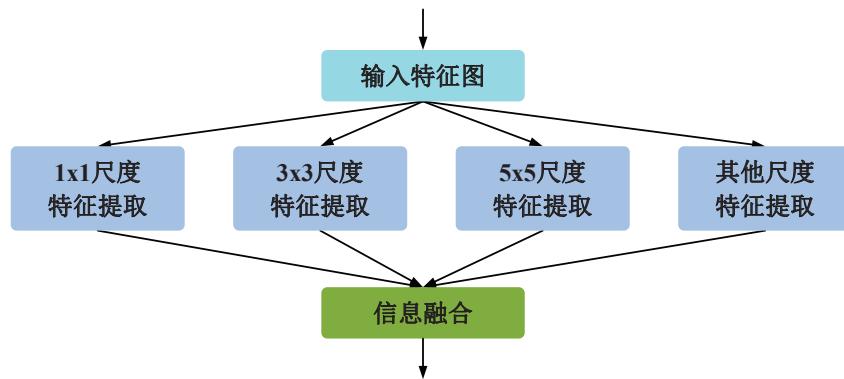


图 5.4 多尺度卷积示意图

Figure 5.4 Diagram of universal multi-scale convolution

5.2 算法设计

RHMSCN 结构如图5.2所示。首先提取 MP3 音频的 QMDCT 系数矩阵，经过富高通滤波模型预处理后，将差分矩阵和原始矩阵在通道层拼接，再将拼接后的矩阵输入到卷积层中进行特征提取。卷积层输出的特征图通过全连接层实现信息整合后输入至 Softmax 分类器，实现对 MP3 音频是否发生隐写的判断。每个卷积模块包含一个子网模块和一个最大池化模块，其中子网结构如图5.3所示。每个子网均包含五个分支，分别在尺寸为 1、3、5 的感受野上进行特征提取，由此实现对输入特征图的多尺度特征提取。其次，利用卷积核分解，使用 $W \times 1$ 和 $1 \times W$ 两个级联的卷积核替代 $W \times W$ 卷积核以减少了网络需要训练的参数个数。同时，在子网设计中还借鉴了 ResNet[88] 的残差连接结构，将子网输出的特征图与输入的特征图按元素相加（Element-Wise），缓解网络的退化问题，降低网络的训练难度。

5.2.1 多尺度卷积结构设计

多尺度卷积结构如图5.4所示。不同大小的卷积核对应了特征图中不同大小的感受野，不同的隐写算法在 QMDCT 系数矩阵不同尺度的感受野上具有不同的表达特性，为了使网络能够同时捕获到不同隐写算法在 QMDCT 系数矩阵不同尺度上的潜在特性，RHMSNC 结构在设计时借鉴了 GoogleNet-Inception 的思想。基于 MP3 音频的编码特性，两个大值区的 QMDCT 系数对应一个 Huffman 码字，四个小值区的 QMDCT 系数对应一个 Huffman 码字。因此，在 RHMSNC 设计中分别使用 1×1 、 3×3 、 5×5 大小的卷积核对输入特征图进行特征提取，并对各分支的输出特征图在通道层进行拼接。

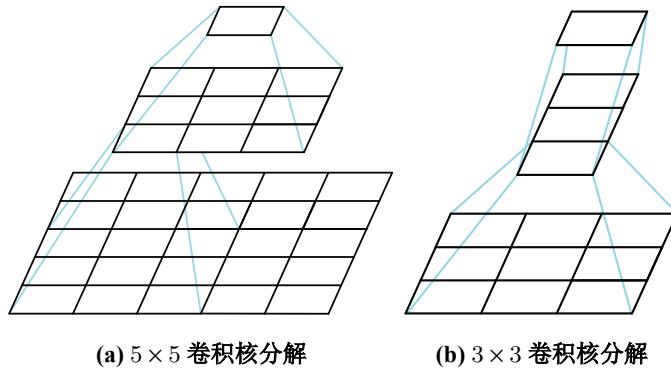


图 5.5 卷积核分解示意图

Figure 5.5 Diagram of convolution decomposition

5.2.2 卷积核分解

卷积核分解的目的在于网络参数的轻量化。大尺寸的卷积核虽然对应了更大的感受野，能够学习到范围更大的信息，但是与之带来的便是参数个数的增加。将一个 3×3 大小的卷积核替换为一个 5×5 大小的卷积核，在卷积核个数不变的情况下，需要学习的权重的个数将会是之前的 $25 \div 9 = 2.78$ 倍。为此，本论文将使用多个小尺寸卷积核级联的方式来替代大尺寸的卷积核。由矩阵性质可知，当矩阵的秩为 1 时，一个 $W \times W$ 的矩阵可以表达为一个 $W \times 1$ 和一个 $1 \times W$ 矩阵的相乘。如 Sobel 算子和 Gauss 算子，其分解过程分别如下所示。

$$K_{Sobel} = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \times \begin{pmatrix} -1 & 0 & 1 \end{pmatrix} \quad (5.1)$$

$$K_{Gauss} = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \times \begin{pmatrix} 1 & 2 & 1 \end{pmatrix} \quad (5.2)$$

文献 [116] 中还提出，两个级联的 3×3 卷积核可以替代一个 5×5 卷积核，如图 5.5a 所示。通过这一变换，在网络性能不降低的情况下，参数个数将减少为此前的 72%。进一步地，如图 5.5b 所示，一个 3×3 的卷积核还可以进一步地被两个级联的 3×1 和 1×3 卷积核取代，最终参数个数便简化至此前的 48%。而在参数量大大减少的同时，卷积层数的增多又增加了激活函数的个数，进一步提升了网络对非线性函数的拟合能力。

5.2.3 残差网络结构设计

随着网络深度的增加，训练的难度也会逐渐加大。阻碍网络训练的因素主要来自两个方面，一个是梯度消失和梯度爆炸，另一个就是网络退化。目前，梯度消失和梯度爆炸已经可以通过BN层的引入被很好地解决。然而，退化问题则需要通过网络结构的更新来解决。从第四章的实验结果也可以看出，对卷积层进行简单地堆叠并不会进一步提升网络的性能，反而有可能导致检测精度的下降。假设有一个 N 层的卷积神经网络，如果将其 $M+1$ 至 N 层的卷积层等价视为一个恒等映射，即 $H_{(M+1)\rightarrow N}(x) = x$ ，那么该网络可以被看作是一个只有 M 层的浅层神经网络。但是，经过实验发现， N 层网络的性能并未优于 M 层网络，即 $H_{(M+1)\rightarrow N}(x) = x$ 的映射关系不易被学习到，从而导致了网络的退化。

为了解决网络的退化问题，He等人[88]提出一种残差网络结构，该网络本质上可以看作是一个差分放大器，网络的基本形式为 $H(x) = F(x) + x$ ，通过跳跃连接（Skip Connection）的方式，将网络的任务由 $F(x) = H(x)$ 转换为 $F(x) = H(x) - x$ 。当 $F(x) = 0$ 时，即为上文提到的恒等映射 $H(x) = x$ 。两者相比，残差网络中映射对输出的变化更为敏感，信息更容易被网络学习到，有助于缓解网络的退化问题。比如，如果神经网络的任务是将1映射为1.1，那么两种网络的映射关系分别为 $H_1(1) = F_1(1) = 1.1$ 和 $H_2(1) = F_2(1) + 1 = 1.1$ 。所以， $F_1(1) = 1.1$ 而 $F_2(1) = 0.1$ 。当网络的输出由1.1变为1.11时，映射 F_1 的相对变

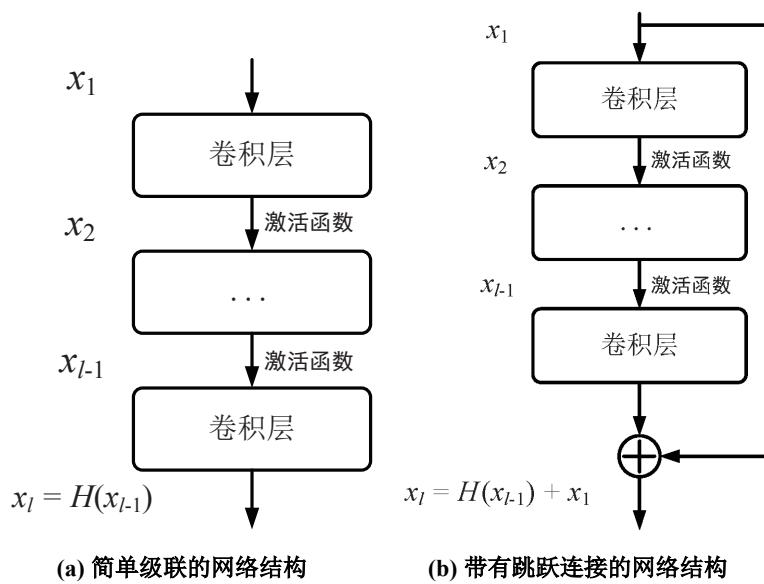


图5.6 VGG风格与ResNet风格的卷积神经网络示意图

Figure 5.6 Diagram of VGG-style network and ResNet-style network

化率为 0.9%，但映射 F_2 的相对变化率却为 10%。因此，在本章网络设计过程中也引入的残差设计的思想。简单级联形式的网络结构和带有跳跃连接形式的网络结构对比如图5.6所示。

5.3 实验设计与分析

本节将通过实验验证基于多尺度卷积神经网络的 MP3 音频隐写分析算法的有效性。首先，分别讨论三种不同的子网结构对网络性能的影响，以分别说明卷积核分解和跳跃连接结构的有效性。其次，基于参数总量、检测准确率、模型大小以及占用显存等多个角度对现有的多个 MP3 音频隐写分析网络进行评估。最后，分别与多种性能较好的 MP3 音频隐写分析算法在隐写算法及隐写负载率已知、隐写算法已知及隐写负载率未知和隐写算法及隐写负载率均未知等三种条件下的性能进行比对，以评估各分析算法性能的优劣。

表 5.1 实验设置

Table 5.1 Settings of experiments

项目	参数	值
音频信息	帧数	50
	比特率	128 / 192 / 256 / 320kbps
隐写负载	HCM	$SPR = 0.1 / 0.3 / 0.5 / 0.8 / 1.0$
	MP3Stego	
EECS		$SPR = 2 / 3 / 4 / 5 / 6$
	批次大小	16 (8 正常/隐写音频对)
优化器		$\text{Adam } (\beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 10^{-8})$
	超参数	
初始学习率		0.001
	学习率更新策略	指数下降 (5000, 0.9)
L^2 正则化系数		0.001
	参数初始化方式	Xavier [107]

5.3.1 实验设置

为了全面地评估 RHMSCN 网络结构的性能，本论文从数据库中选取了各参数条件下的音频对 33038 个用于实验验证，其中 19200 对用于训练，12800 对用于验证，剩余的 1038 对用于和 MSC、RHFCN 隐写分析算法进行比对，其余各参数设置如表5.1所示。实验过程中使用 *FPR*、*FNR*、*ACC* 三个指标度量隐写分析算法的性能。测试过程中使用批量测试方式，BN 层在测试阶段保持开启状态，而每次测试过程中数据都会被置乱，每个批次的数据组合均会有所变化，每次测试结果也会有所不同。因此，最终的实验结果取 10 次实验的算术平均值，以减少偶然误差带来的影响。

5.3.2 子网结构优选

为了验证卷积核分解以及残差结构对网络性能的影响，本论文分别设计了如下三种形式的子网结构。其中，图5.7为朴素的多尺度子网结构示意图，图5.8为引入了跳跃连接的多尺度子网结构示意图，图5.9为引入了卷积核分解和跳跃连接的多尺度子网结构示意图。从表5.2中的实验结果可以看出，由于跳跃连接的引入，网络的性能得以提升。而由于卷积核分解的引入，网络在精度不损失的情况下大大减少了参数量。需要注意的是，由于 RHMSCN 网络的结构相对复杂，训练时间和周期相比于 VGG 风格的 WASDN 和 RHFCN 较长一些。

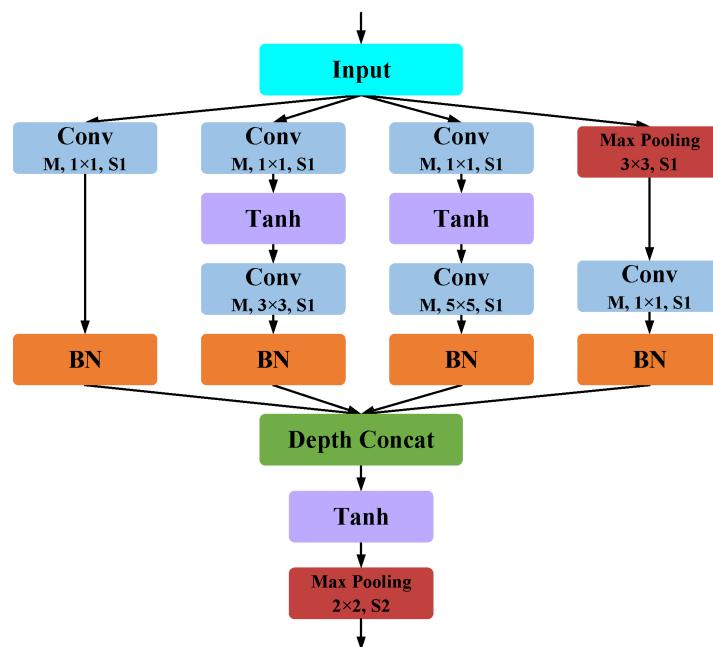


图 5.7 多尺度子网结构示意图 (Sub-Net 1)

Figure 5.7 Diagram of multi-scale sub-network (Sub-Net 1)

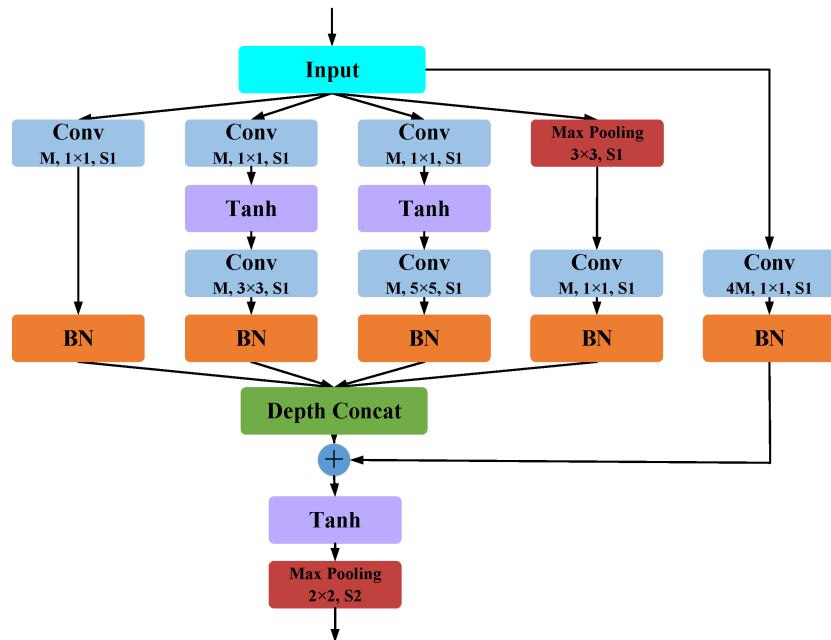


图 5.8 引入残差结构的多尺度子网结构示意图（Sub-Net 2）

Figure 5.8 Diagram of multi-scale sub-network (Sub-Net 2)

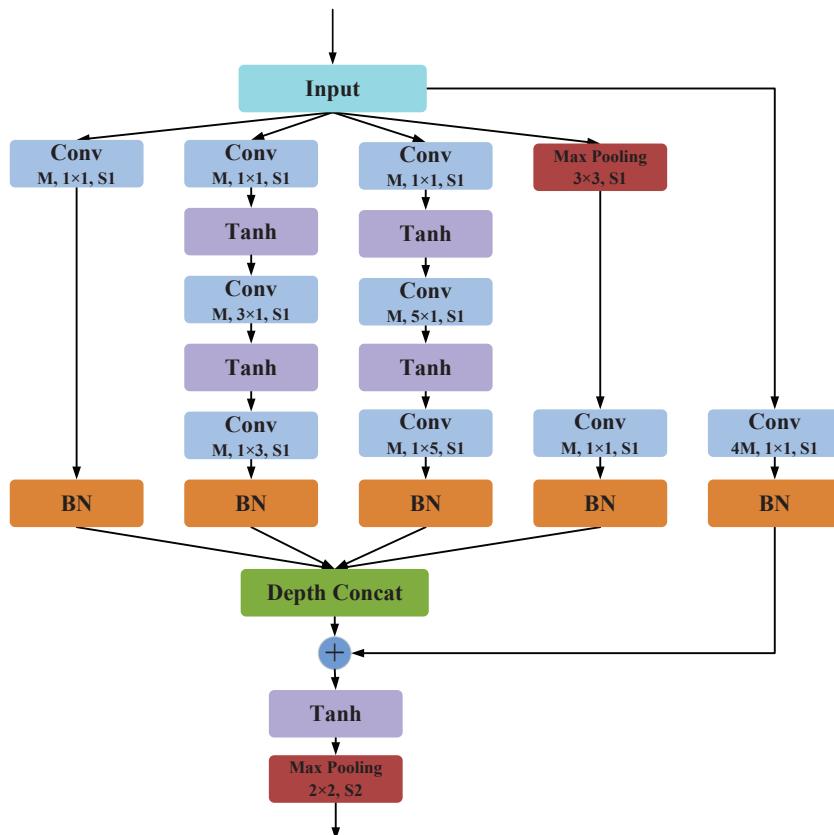


图 5.9 引入卷积核分解和残差结构的多尺度子网结构示意图（Sub-Net 3）

Figure 5.9 Diagram of multi-scale sub-network (Sub-Net 3)

表 5.2 子网性能评估 (EECS, 128kbps, SPR = 2)**Table 5.2 Performance of each sub-network (EECS, 128kbps, SPR = 2)**

子网结构	Sub-Net 1	Sub-Net 2	Sub-Net 3
准确率 (%)	94.68	95.35	95.35
参数个数 (万)	约 3.81k	约 5.69k	约 4.12k

5.3.3 MP3 音频隐写分析网络性能评估

为了对当下性能良好的三种 MP3 音频隐写分析网络 (WASDN、RHFCN、RHMSCN) 结构进行较为全面的评估, 本小节分别从训练周期、检测准确率、网络参数个数、模型大小以及占用显存 (BatchSize = 16) 五个方面进行度量。从表 5.3 中实验结果可以看出, RHFCN 的检测精度最高, 相比于 WASDN 和 MSC 算法提升较为明显; RHMSCN 相比于 RHFCN, 在检测精度不明显降低的情况下, 占用的显存减少 40%、保存模型所需的硬盘空间明显减少, 由 1.7G 缩减为 47M, 更容易应用于移动端应用检测场景。MSC 算法虽然在 $SPR = 2$ 的隐写负载条件下也有较高的检测精度, 但是该特征在低隐写负载条件下的检测效果相对三种 MP3 音频隐写分析网络仍有较大差距, 最终可以根据使用场景的不同对模型进行合理的部署, 以实现对性能和资源消耗的折中。

5.3.4 实验结果与分析

影响 MP3 音频隐写分析算法性能的主要因素有隐写算法类型、音频比特率以及隐写负载率。由于音频比特率是 MP3 音频的固有属性, 可以通过对音频文

表 5.3 MP3 音频隐写分析网络评估 (EECS, 128kbps, BatchSize = 16)**Table 5.3 Performance evaluation of each MP3 audio steganalytic network**

(EECS, 128kbps, BatchSize = 16)

	MSC	WASDN	RHFCN	RHMSCN
准确率 ($SPR = 2$)	92.14	92.39	95.40	95.35
准确率 ($SPR = 3$)	79.06	82.07	89.65	88.60
模型大小 (MB)	1.03	932.29	1764.23	47.20
参数个数 (万)	0.06	约 8k	约 15k	约 4k
占用显存 (MB)	-	4551	7391	4549

件解析直接获取到，因此本小节将主要讨论本论文所提的三种 MP3 音频隐写分析算法，MSC，RHFCN 和 RHMSCN，基于隐写算法和隐写负载率的非盲隐写分析与盲隐写分析。

5.3.4.1 非盲隐写分析性能评估

非盲隐写分析旨在实验室环境下对隐写分析算法性能进行评估。MSC、RHFCN 和 RHMSCN 三种隐写分析算法对 MP3Stego、HCM 和 EECS 隐写算法的检测结果分别如表 5.4、5.5、5.6 所示。从实验结果可以看出，本论文所提的三种隐写分析算法可以很好地应用于 MP3Stego 和 HCM 的检测分析，检测准确率均在 90% 以上，在部分隐写条件下的检测准确率甚至可以达到 100%。对于 EECS 隐写算法的检测分析，MSC 算法在四种比特率的低隐写嵌入率条件下的检测精度均不是很理想，而 RHFCN 和 RHMSCN 两种隐写分析网络的性能相比于 MSC 算法均有较为明显的提升。在高比特率及高隐写负载率条件下的检测精度普遍高于 90%。但是，在低比特率及其对应的低负载率（128kps, $SPR = 4, 5, 6$ ）条件下的检测精度仍有待进一步提升。在不同的嵌入条件下，RHFCN 和 RHMSCN 算法的检测性能整体相当但各有高低，由此也体现出 RHMSCN 结构在网络参数轻量化的同时有效保持了网络的性能无明显衰减。

表5.4 对MP3Stego 算法的隐写分析结果

Table 5.4 Performance of each steganographic algorithm on the detection of MP3Stego

比特率	SPR	MSC			RHFCN			RHMSCN		
		FPR	FNR	ACC	FPR	FNR	ACC	FPR	FNR	ACC
128	0.1	7.95	7.85	92.10	5.32	6.44	94.12	6.14	5.70	94.08
	0.3	2.02	10.00	93.99	4.05	3.33	96.31	3.37	4.59	96.02
	0.5	2.45	7.89	94.83	1.56	1.14	98.65	1.20	1.18	98.81
	0.8	1.77	7.19	95.52	0.41	0.83	99.38	0.54	0.74	99.36
	1.0	1.51	5.53	96.48	0.00	0.00	100.0	0.00	0.00	100.0
192	0.1	5.60	4.80	94.80	4.82	3.66	95.76	4.95	3.41	95.82
	0.3	7.99	0.48	95.77	2.78	3.54	96.84	2.99	3.52	96.74
	0.5	6.64	1.56	95.90	0.76	0.96	99.14	0.48	1.20	99.16
	0.8	6.10	1.07	96.42	0.00	0.00	100.0	0.00	0.00	100.0
	1.0	1.39	2.17	98.22	0.00	0.00	100.0	0.00	0.00	100.0
256	0.1	5.31	4.41	95.14	3.55	3.81	96.32	3.66	3.98	96.18
	0.3	5.05	1.51	96.72	1.24	0.94	98.91	1.64	0.92	98.72
	0.5	3.47	2.73	96.90	0.00	0.00	100.0	0.00	0.00	100.0
	0.8	3.70	0.54	97.88	0.00	0.00	100.0	0.00	0.00	100.0
	1.0	1.49	2.01	98.25	0.00	0.00	100.0	0.00	0.00	100.0
320	0.1	5.67	2.35	95.99	1.82	2.20	97.99	2.02	2.51	97.74
	0.3	4.19	2.63	96.59	0.60	0.60	99.40	0.67	0.48	99.43
	0.5	3.20	2.44	97.18	0.00	0.00	100.0	0.00	0.00	100.0
	0.8	2.96	0.51	98.27	0.00	0.00	100.0	0.00	0.00	100.0
	1.0	0.53	1.42	99.03	0.00	0.00	100.0	0.00	0.00	100.0

表 5.5 对 HCM 算法的隐写分析结果

Table 5.5 Performance of each steganographic algorithm on the detection of HCM

比特率	SPR	MSC			RHFCN			RHMSCN		
		FPR	FNR	ACC	FPR	FNR	ACC	FPR	FNR	ACC
128	0.1	31.00	33.35	67.83	8.74	8.10	91.58	9.63	9.83	90.27
	0.3	12.40	13.35	87.12	8.41	7.35	92.12	9.15	8.19	91.33
	0.5	2.79	2.00	97.61	2.79	2.00	97.61	3.08	2.12	97.40
	0.8	0.33	0.05	99.81	0.33	0.05	99.81	0.21	0.18	99.81
	1.0	0.14	0.11	99.88	0.14	0.11	99.88	0.00	0.00	100.0
192	0.1	20.04	19.61	80.17	4.97	4.85	95.09	4.33	4.91	95.38
	0.3	4.64	4.56	95.40	1.53	1.67	98.40	1.45	1.83	98.36
	0.5	0.61	0.75	99.32	0.20	0.14	99.83	0.19	0.20	99.80
	0.8	0.14	0.03	99.92	0.00	0.12	99.94	0.00	0.00	100.0
	1.0	0.01	0.00	99.99	0.00	0.00	100.0	0.00	0.00	100.0
256	0.1	11.15	11.69	88.58	2.13	2.73	97.57	2.12	2.41	97.74
	0.3	1.85	1.89	98.13	1.85	1.89	98.13	1.73	1.93	98.17
	0.5	0.04	0.29	99.84	0.14	0.18	99.84	0.96	0.77	99.14
	0.8	0.00	0.00	100.0	0.00	0.00	100.0	0.00	0.00	100.0
	1.0	0.00	0.00	100.0	0.00	0.00	100.0	0.00	0.00	100.0
320	0.1	5.84	7.76	93.20	1.01	1.21	98.79	1.35	1.54	98.56
	0.3	1.13	1.92	98.47	0.00	0.38	99.81	0.25	0.19	99.78
	0.5	0.10	0.03	99.94	0.00	0.00	100.0	0.00	0.00	100.0
	0.8	0.00	0.00	100.0	0.00	0.00	100.0	0.00	0.00	100.0
	1.0	0.00	0.00	100.0	0.00	0.00	100.0	0.00	0.00	100.0

表5.6 对EECS算法的隐写分析结果

Table 5.6 Performance of each steganographic algorithm on the detection of EECS

比特率	SPR	MSC			RHFCN			RHMSCN		
		FPR	FNR	ACC	FPR	FNR	ACC	FPR	FNR	ACC
128	2	7.91	7.80	92.14	4.04	5.16	95.40	4.30	5.00	95.35
	3	20.22	21.65	79.06	11.21	9.49	89.65	11.40	11.40	88.60
	4	30.54	30.48	69.49	17.10	18.05	82.43	19.95	18.90	80.58
	5	34.88	36.45	64.34	22.53	21.87	77.80	23.70	23.20	76.55
	6	39.00	41.06	59.97	25.34	25.67	74.50	28.00	30.06	70.97
192	2	4.35	4.71	95.47	1.41	2.53	98.03	2.10	0.45	98.72
	3	16.77	17.39	82.92	4.10	5.60	95.15	5.40	3.40	95.60
	4	24.65	26.85	74.25	6.45	6.55	93.50	7.30	4.85	93.92
	5	30.66	29.92	69.71	13.15	4.45	91.20	10.45	4.85	92.35
	6	32.36	38.03	64.81	16.70	17.85	82.73	14.90	14.80	85.15
256	2	4.19	4.11	95.85	0.69	1.17	99.07	0.52	1.06	99.21
	3	14.91	17.39	83.85	3.51	3.63	96.43	3.44	3.51	96.53
	4	25.22	24.76	75.01	4.35	4.17	95.74	4.00	6.00	95.00
	5	28.62	31.10	70.14	7.52	7.92	92.28	7.41	7.47	92.56
	6	33.14	36.20	65.33	14.60	6.05	89.68	10.00	9.05	90.48
320	2	3.53	3.86	96.31	0.00	0.10	99.95	0.00	0.10	99.95
	3	15.45	16.44	84.06	3.00	2.98	97.01	2.80	3.00	97.10
	4	24.59	24.33	75.54	3.25	3.67	96.54	3.59	3.77	96.32
	5	29.34	30.23	70.22	6.70	6.86	93.22	7.34	7.23	92.72
	6	33.67	33.45	66.44	8.67	9.83	90.75	10.67	8.45	90.44

5.3.4.2 盲隐写分析性能评估

盲隐写分析要求算法在面对未知组合的混合样本时也能有较好的检测效果，从而实现隐写分析系统的在线部署，有效地将隐写分析算法推向实际应用。在进行算法盲隐写性能评估时，本论文将基于两种隐写算法（HCM 和 EECS）与其对应的五种隐写负载率进行讨论，比特率限定为 128kbps。在实验过程中发现，由 EECS 隐写样本为训练数据得到的模型可以直接应用于 HCM 隐写算法的检测分析，当 $SPR \geq 0.3$ 时，仍可以获得 85% 以上的检测准确率。因此，本论文分别构建了三种用于盲分析的训练数据集，详细信息如表5.7所示。分别将由不同数据集训练得到的 Model 1、Model 2 和 Model 3 应用于表5.8中所示的三类混合测试数据集的检测分析以评估不同训练方案对应模型的性能。在所有混合数据集的制备过程中，各参数对应样本均是从测试数据集中等比例随机抽取得到的，以减少因算法和隐写负载差异引起的负面影响。实验结果分别如表5.9、5.10和5.11所示。

从实验结果可以看出，虽然 Model 3 对 Dataset 1 的检测准确率不足 80%，对 Dataset 2 的检测精度相比于 Model 2 也下降了 3-4 个百分点，但是对 Dataset 3 的检测精度可达到 84% 以上，对于全部由 HCM 算法混合得到的样本集，检测准确

表 5.7 用于 MP3 音频盲隐写分析的训练模型说明

Table 5.7 Trained models for MP3 audio blind steganalysis

名称	训练集	数量
Model 1	EECS、128、 $SPR = 2$	10000
Model 2	EECS、128、 $SPR = 2 / 3 / 4 / 5 / 6$	10000
Model 3	EECS、128、 $SPR = 2 / 3 / 4 / 5 / 6$ HCM、128、 $SPR = 0.1 / 0.3 / 0.5 / 0.8 / 1.0$	10000

表 5.8 混合测试数据集说明

Table 5.8 Mixed test dataset

名称	说明	数量
Dataset 1	EECS、128、 $SPR = 2 / 3 / 4 / 5 / 6$	2000
Dataset 2	HCM、128、 $SPR = 0.1 / 0.3 / 0.5 / 0.8 / 1.0$	2000
Dataset 3	Sub-Dataset 1 + Sub-Dataset 2	2000

表5.9 对混合数据集的盲隐写分析结果（Model 1）

Table 5.9 Detection performance of blind steganalysis (Model 1)

数据集	MSC			RHFCN			RHMSCN		
	FPR	FNR	ACC	FPR	FNR	ACC	FPR	FNR	ACC
Dataset 1	12.10	52.60	67.65	21.07	24.80	77.07	20.06	29.84	75.05
Dataset 2	12.10	22.10	82.90	8.37	22.88	84.38	6.55	27.42	83.01
Dataset 3	12.10	36.60	75.65	14.42	24.70	80.44	11.69	29.64	79.33

表5.10 对混合数据集的盲隐写分析结果（Model 2）

Table 5.10 Detection performance of blind steganalysis (Model 2)

数据集	MSC			RHFCN			RHMSCN		
	FPR	FNR	ACC	FPR	FNR	ACC	FPR	FNR	ACC
Dataset 1	29.25	30.30	70.23	19.05	19.76	80.59	20.47	21.99	78.77
Dataset 2	29.10	12.10	79.40	14.21	24.90	80.44	15.22	28.63	78.07
Dataset 3	29.15	20.40	75.23	15.22	24.90	79.94	15.73	25.61	79.33

表5.11 对混合数据集的盲隐写分析结果（Model 3）

Table 5.11 Detection performance of blind steganalysis (Model 3)

数据集	MSC			RHFCN			RHMSCN		
	FPR	FNR	ACC	FPR	FNR	ACC	FPR	FNR	ACC
Dataset 1	16.85	49.75	66.70	25.81	21.77	76.21	31.07	23.03	72.95
Dataset 2	16.65	18.65	82.35	3.53	7.46	94.51	4.23	10.08	92.84
Dataset 3	16.85	34.30	74.42	13.15	17.34	84.76	13.24	19.83	83.46

率更是高达94%以上。因此，本论文认为Model 2更适用于EECS算法在隐写负载率未知条件下的隐写分析，而Model 3则更适用于广泛意义上的盲隐写分析。此外，需要注意的是，RHFCN和RHMSCN在Dataset 1和Dataset 2上的检测效果逊于在Dataset 3上的检测效果，这一现象一方面体现出EECS隐写算法较高的安全性；另一方面也表明基于深度学习的隐写分析算法对输入的训练数据有

着更强的依赖性。但是，相比于 MSC 算法，RHFCN 和 RHMSCN 算法在各类分析场景下的检测效果均有较为明显的提升。RHMSCN 在非盲隐写分析场景下的检测精度相较于 RHFCN 不分伯仲，但是在 MP3 音频盲隐写分析中，RHFCN 的优势更为明显，需要根据应用场景的不同对算法进行合理的选择。综上所述，本论文最终选定 Model 3 用于 MP3 音频盲隐写分析。

此外，本论文还继续将 Model 3 应用于非盲隐写分析场景以对其性能进行评估，从表5.12、5.13中的实验结果可以看出，RHFCN 和 RHMSCN 算法对 HCM 隐写样本仍具有很高的检测精度，即使在 SPR 仅为 0.1 时，分析准确率仍在 80% 以上。而 EECS 隐写算法相对安全，RHFCN 和 RHMSCN 对其检测效果并不理想，仅在 $SPR = 2$ 条件下准确率高于 90%。因此，未来 MP3 音频分析算法设计

表 5.12 对 HCM 算法的隐写分析检测结果 (128kbps)

Table 5.12 Performance on the detection of HCM (128kbps)

SPR	MSC			RHFCN			RHMSCN		
	FPR	FNR	ACC	FPR	FNR	ACC	FPR	FNR	ACC
0.1	23.40	60.00	58.30	15.47	13.73	85.40	19.51	12.86	83.82
0.3	23.40	20.60	78.00	6.49	5.33	94.09	8.14	3.78	94.04
0.5	23.80	2.00	87.10	4.36	3.10	96.27	5.62	3.00	95.69
0.8	23.00	0.00	88.50	1.94	1.55	98.26	3.39	0.29	98.16
1.0	22.40	0.00	88.80	1.45	1.26	98.64	2.52	0.78	98.35

表 5.13 对 EECS 算法的隐写分析检测结果 (128kbps)

Table 5.13 Performance on the detection of EECS (128kbps)

SPR	MSC			RHFCN			RHMSCN		
	FPR	FNR	ACC	FPR	FNR	ACC	FPR	FNR	ACC
2	23.20	10.60	83.10	6.88	4.65	94.23	12.35	6.69	90.84
3	23.80	36.20	70.00	19.28	13.37	83.67	20.30	18.67	80.52
4	23.20	51.60	62.60	30.72	19.48	74.90	39.18	20.58	70.12
5	23.40	61.80	57.40	37.50	28.10	67.20	39.70	26.43	66.94
6	23.20	67.80	54.50	42.15	31.88	62.98	39.00	41.06	59.97

的主要方向仍然在于较低隐写负载率条件下的检测性能提升。此外，虽然在以上各类情况下，MSC 算法的检测效果相比于 RHFCN 和 RHMSCN 算法均有不同程度的下降，但在现有的隐写分析算法中也是较为优秀的。

5.4 本章小结

本章提出一种基于 QMDCT 系数矩阵和多尺度卷积神经网络的 MP3 音频隐写分析算法，通过引入卷积核分解实现网络参数的轻量化，同时通过引入多尺度卷积和跳跃连接实现网络的性能保持。最终，在网络检测精度无明显下降的前提下，参数量大大减少，训练占用的显存以及保存模型所需要的空间都大大降低，有效增强了基于深度学习的 MP3 音频隐写分析算法的实用性。

此外，本章还全面分析了本论文所提的三种隐写分析算法，MSC，RHFCN 和 RHMSCN 算法，在基于隐写算法类型和隐写负载率已知与否条件下的非盲隐写分析与盲隐写分析，并提出一种盲隐写分析方案，检测准确率平均可达到 80% 以上。

第 6 章 总结与展望

6.1 全文总结

MP3 音频隐写分析是多媒体安全与智能分析领域一个十分重要的分支，然而，相比于图像及视频隐写分析，MP3 音频隐写分析发展相对滞缓，具有巨大的研究价值和广阔的研究空间。本论文通过对 MP3 音频编解码原理、MP3 音频隐写和隐写分析算法原理的分析、研究和对比，分别从基于传统手工特征的算法设计和基于深度学习技术的网络结构设计两个角度出发，提出了三种高效且实用的 MP3 隐写分析方法。

本论文的主要工作与创新列举如下。

1. 提出一种适用于 MP3 音频隐写分析的富高通滤波模型。通过分析广泛应用于图像隐写分析中 Rich Model, KV 核等残差提取模型，为了提升隐写分析算法对隐写噪声的敏感性以实现对隐写信号更好的检测，本论文提出一种适用于 MP3 音频隐写分析的通用富高通滤波模型，可以应用于多种 MP3 音频隐写分析算法设计。
2. 提出一种基于 QMDCT 系数矩阵多尺度相关性度量的 MP3 音频隐写分析算法。通过分析 MP3 编码原理、MP3 隐写算法原理以及隐写对 MP3 音频 QMDCT 系数的影响，本论文提出一种基于 QMDCT 系数矩阵多尺度相关性度量的 MP3 音频隐写分析算法 MSC，分别在系数尺度 (1×1) 和 Huffman 码字尺度 (2×2 和 4×4) 对 QMDCT 系数矩阵相关性变化进行度量，检测精度相比于同类型的其他隐写分析算法平均提升 20% 以上。此外，通过分析待检测音频段长度对隐写分析算法性能的影响，确定了 MSC 算法所能实现的隐写音频段定位粒度为 500ms，置信度在 80% 以上。
3. 提出一种基于 QMDCT 系数矩阵与卷积神经网络的 MP3 音频隐写分析算法。通过分析输入数据类型、神经网络类型以及神经网络结构对 MP3 音频隐写分析算法检测精度的影响，本论文提出一种适用于 MP3 音频隐写分析的卷积神经网络结构 RHFCN，与基于手工特征设计的 MP3 音频隐写分析算法相比，检测准确率得到了极为明显的提升，对 EECS 隐写算法在较低比特率条件下的检测准确率相比于 MSC 算法提升 10% 以上。此外，RHFCN 算法也可以在一定程度上接收任意大于 200×400 尺寸的输入数据，且检测精度无明显损失，能够在一定程度上解决输入尺寸失配问题。同时，本论文还提出通过迁移学习解决低隐写负

载条件下的训练难度大，检测困难的问题。

4. 提出一种基于 QMDCT 系数矩阵和多尺度卷积神经网络的 MP3 音频隐写分析方法。针对现阶段基于深度学习技术的 MP3 音频隐写分析网络待训参数多、模型占用空间较大、对计算资源依赖性较强等问题，本论文提出一种轻量级的 MP3 音频隐写分析网络 RHMSCN。通过卷积核分解的形式减少网络中需要训练的参数个数，同时通过多尺度卷积以及跳跃连接的方式维持网络的检测精度不下降，由此得到一个性能优越且计算和存储代价小的卷积神经网络结构。

5. 构建了一个适用于音频隐写分析的基础数据集。音频隐写分析算法性能的验证以及基于神经网络的音频隐写分析算法设计是以完备的数据集为驱动的，而现阶段还没有专门应用于音频隐写与隐写分析的公共数据集。为此，本论文构建了一个包含 33038 个 WAV 音频的基础数据集，并将其编码为相应的正常/隐写音频对，包含 MP3Stego、HCM 和 EECS 等三类经典 MP3 音频隐写算法；128kbps、192kbps、256kbps、320kbps 等四种常用音频比特率以及多种隐写负载率，有助于后续音频隐写与隐写分析研究的进行。

6.2 下一步工作

为了更好地丰富音频隐写分析的理论背景同时推动音频隐写分析技术的落地。本论文工作今后可从以下三个方面继续深入研究。

1. 设计与隐写原理结合更为紧密的音频隐写分析算法。虽然基于深度学习的音频隐写分析技术已经取得了一定的成绩，然而网络性能的提升主要源于网络结构的改进，却并未过多地与音频隐写算法原理以及隐写算法对载体分布特性的影响相结合。因此，结合音频隐写算法和音频编码原理设计通用的音频隐写分析算法是后续一项十分关键的工作。

2. 提升音频隐写分析算法的实用性。虽然基于深度学习的音频隐写分析算法的性能明显优于基于手工特征设计的隐写分析方法，但存在着数据依赖性强、计算成本高、模型文件大等不足。因此，为了进一步提升隐写分析算法的实用性，设计轻量级且性能优越的网络结构、基于少量训练数据的网络结构以及适用于现实场景分析的网络结构都是音频隐写分析可进一步探索的方向。

3. 设计安全性更高的音频隐写算法。借鉴基于深度学习的图像隐写算法，如 CycleGAN[118]、HiDDeN[119]、SSGAN[120] 等，设计更安全的音频隐写算法以带动关于音频隐写分析算法研究的推进。

参考文献

- [1] GOLDREICH O. The Foundations of Cryptography - Volume 1, Basic Techniques[M]. England, UK: Cambridge University Press, 2001.
- [2] GOLDREICH O. The Foundations of Cryptography - Volume 2, Basic Applications[M]. England, UK: Cambridge University Press, 2004.
- [3] WETSTONE-TECHNOLOGIES. StegoHunt[EB/OL]. 2002. <https://www.wetstonetech.com/products/stegohunt>.
- [4] FABIEN P. MP3Stego[EB/OL]. 2002. <http://www.petitcolas.net/steganography/mp3stego>.
- [5] YAN D, WANG R. Detection of MP3Stego Exploiting Recompression Calibration-based Feature[J]. Multimedia Tools and Applications (MTAP), 2014, 72(1): 865–878.
- [6] YANG K, YI X, ZHAO X, et al. Adaptive MP3 Steganography using Equal Length Entropy Codes Substitution[C]//Proceedings of the 16th International Workshop on Digital Forensics and Watermarking (IWDW 2017). Magdeburg, Germany: Springer, 2017: 202–216.
- [7] ANDREAS W, ANDREAS P. Attacks on Steganographic Systems[C]//Proceedings of the 3rd International Workshop on Information Hiding (IH'99). Dresden Germany: Springer, 1999: 61–76.
- [8] FRIDRICH J, GOLJAN M, DU R. Reliable Detection of LSB Steganography in Color and Grayscale Images[C]//Proceedings of the 4th Workshop on Multimedia and Security (MMSec 2001). Ottawa, Ontario, Canada: ACM, 2001: 27–30.
- [9] DUMITRESCU S, WU X, MEMON N D. On Steganalysis of Random LSB Embedding in Continuous-tone Images[C]//Proceedings of the 2002 IEEE International Conference on Image Processing (ICIP 2002). Rochester, New York, USA: IEEE, 2002: 641–644.
- [10] WEI Z, HAOJUN A, RUIMIN H. An Algorithm of Echo Steganalysis based on Power Cepstrum and Pattern Classification[C]//Proceedings of the 2008 International Conference on Audio, Language and Image Processing (ICALIP 2008). Shanghai, China: IEEE, 2008: 1344–1348.
- [11] WANG Y, WEN H, JIAN Z, et al. Steganalysis on Positive and Negative Echo Hiding based on Skewness and Kurtosis[C]//Proceedings of the 9th IEEE Conference on Industrial Electronics and Applications (ICIEA 2014). Hangzhou, Zhejiang, China: IEEE, 2014: 1235–1238.
- [12] XIE C, CHENG Y, CHEN Y. An Active Steganalysis Approach for Echo Hiding based on Sliding Windowed Cepstrum[J]. Signal Processing, 2011, 91(4): 877–889.
- [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.
- [14] ALTUN O, SHARMA G, CELIK M U, et al. Morphological Steganalysis of Audio Signals

- and the Principle of Diminishing Marginal Distortions[C]//Proceedings of the 2005 IEEE International Conference on Acoustics, Speech, and Signal (ICASSP 2005). Philadelphia, Pennsylvania, USA: IEEE, 2005: 21–24.
- [15] GAO S, HU R, ZENG W, et al. A Detection Algorithm of Audio Spread Spectrum Data Hiding[C]//Proceedings of the 4th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM 2008). Dalian, Liaoning, China: IEEE, 2008: 1–4.
- [16] LI C, ZENG W, AI H, et al. Steganalysis of Spread Spectrum Hiding based on DWT and GMM[C]//Proceedings of the 2009 International Conference on Networks Security, Wireless Communications and Trusted Computing (NSWCTC 2009). Wuhan, Hubei, China: IEEE, 2009: 240–243.
- [17] 黄昊, 郭立, 李琳. 基于失真测度的直接扩音频频隐写分析[J]. 中国科学院大学学报, 2008, 25(2): 251–256.
- [18] 谢春辉, 程义民, 陈扬坤. PN 序列估计与扩频隐藏信息分析[J]. 电子学报, 2011, 39(2): 255–259.
- [19] 谢春辉. 音频隐藏分析方法研究[D]. 安徽: 中国科学技术大学, 2011.
- [20] QI Y, YE L, LIU C. Wavelet Domain Audio Steganalysis for Multiplicative Embedding Model[C]//Proceedings of the 2009 International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR 2009). Baoding, Hebei, China: IEEE, 2009: 429–432.
- [21] RU X, ZHANG H, HUANG X. Steganalysis of Audio: Attacking the Steghide[C]// Proceedings of the 4th International Conference on Machine Learning and Cybernetics (ICMLC 2005). Guangzhou, Guangdong, China: IEEE, 2005: 3937–3942.
- [22] BÖHME R, WESTFELD A. Statistical Characterisation of MP3 Encoders for Steganalysis [C]//Proceedings of the 6th Workshop on Multimedia and Security (MMSec 2004). Magdeburg, Germany: ACM, 2004: 25–34.
- [23] 宋华, 幸丘林, 李维奇, 等. MP3Stego 信息隐藏与检测方法研究[J]. 中山大学学报 (自然科学版), 2004, 43(s2): 221–224.
- [24] JULIO H C, JUAN T, ESTHER P, et al. Blind Steganalysis of MP3Stego[J]. Journal of Information Science and Engineering (JISE), 2010, 26(5): 1787–1799.
- [25] 陈益如, 王让定, 严迪群. 基于 Huffman 码表索引的 MP3Stego 隐写分析方法[J]. 计算机工程与应用, 2012, 48(9): 124–126.
- [26] WAN W, ZHAO X, HUANG W, et al. Steganalysis of MP3Stego based on Huffman Table Distribution and Recording[J]. Journal of Graduate University of Chinese Academy of Science, 2012, 29(1): 118–124.
- [27] YU X, WANG R, YAN D. Detecting MP3Stego using Calibrated Side Information Features [J]. Journal of Software (JSW), 2013, 8: 2628–2636.
- [28] YU X, WANG R, YAN D, et al. MP3 Audio Steganalysis using Calibrated Side Information Feature[J]. Journal of Computational Information System (JCIS), 2012, 8(10): 4241–4248.

- [29] 李友勇, 潘峰, 申军伟. 针对 MP3Stego 的一种边信息作为特征的改进分析算法[J]. 小型微型计算机系统, 2015, 36(3): 572–575.
- [30] YAN D, WANG R, YU X, et al. Steganalysis for MP3Stego using Differential Statistics of Quantization Step[J]. Digital Signal Processing, 2013, 23(4): 1181–1185.
- [31] YANG K, WANG R, YAN D, et al. Active Steganalysis for MP3Stego[J]. Journal of Computational Information Systems (JCIS), 2014, 10(15): 6767–6776.
- [32] 羊开云, 王让定, 严迪群, 等. MP3Stego 嵌入密文长度估计[J]. 中国科技论文, 2014(4): 429–433.
- [33] YAN D, WANG R. Detection of MP3Stego Exploiting Recompression Calibration-based Feature[J]. Multimedia Tools and Applications (MTAP), 2014, 72(1): 865–878.
- [34] 余先敏. 压缩域音频隐写分析技术研究[D]. 浙江: 宁波大学, 2012.
- [35] JOHNSON N. STEGANOGRAPHY SOFTWARE[EB/OL]. 2012. <http://www.jjtc.com/Steganography/tools.html>.
- [36] NAKASOFT. Xiao Steganography[EB/OL]. 2006. <https://www.softpedia.com/get/Security/Encrypting/Xiao-Steganography.shtml>.
- [37] EAST-TEC. Invisible Secrets 4[EB/OL]. 2011. <http://www.invisiblesecrets.com>.
- [38] BÁTORA J. DeepSound[EB/OL]. 2008. <http://jpinssoft.net/deepsound>.
- [39] CHOREIN A. SilentEye[EB/OL]. 2011. <https://silenteye.v1kings.io>.
- [40] PLATT C. UnderMP3Cover[EB/OL]. 2002. <http://linus.linuxlab.cs.pdx.edu/cgi-bin/twiki/view/Main/UnderMP3Cover>.
- [41] CHMDZNR. MP3Stegz[EB/OL]. 2008. <https://sourceforge.net/projects/mp3stegz>.
- [42] 易小伟, 李金才, 王运韬, 等. 一种基于 IS4 软件特征的隐藏信息检测及提取方法[P]. 2016-08-26.
- [43] 王让定, 金超, 严迪群, 等. 一种针对 MP3Stegz 的隐写检测方法[P]. 2015-08-05.
- [44] 张坚, 王让定, 严迪群. 一种针对 UnderMP3Cover 的 RS 隐写分析新方法[J]. 电信科学, 2018, 34(4): 68–80.
- [45] 汝学民, 庄越挺, 吴飞. 基于隐写工具的自相关特性进行音频隐写分析[J]. 通信学报, 2006, 27(4): 101–106.
- [46] REPP H. Hide4PGP[EB/OL]. 1996. <http://www.heinz-repp.onlinehome.de/Hide4PGP.htm>.
- [47] PULCINI G. Stegowav[EB/OL]. 1997. <http://www.verrando.com/pulcini>.
- [48] HETZL S. Steghide[EB/OL]. 2003. <http://steghide.sourceforge.net>.
- [49] 汪云路, 程义民, 谢春辉, 等. 基于组合质量评估参数的音频信息隐藏盲检测方法与仿真[J]. 系统仿真学报, 2009, 21(12): 3768–3772.
- [50] 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 (ESA), 2010, 37(12): 7469–7482.
- [51] HAMZEH G, TAJIK K M, KHALIL A M. Audio Steganalysis based on Reversed Psychoacoustic Model of Human Hearing[J]. Digital Signal Processing, 2016, 51: 133–141.

- [52] HAN C, XUE R, ZHANG R, et al. A New Audio Steganalysis Method based on Linear Prediction[J]. *Multimedia Tools and Applications (MTAP)*, 2018, 77(12): 15431–15455.
- [53] SURESH V, SHIBIN.K, ANOOP. StegoMagic[EB/OL]. 2000. <https://www.downloadsource.net/1755462/StegoMagic>.
- [54] QIAO M, SUNG A H, LIU Q. MP3 Audio Steganalysis[J]. *Information Sciences*, 2013, 231: 123–134.
- [55] 王让定, 羊开云, 严迪群, 等. 一种基于共生矩阵分析的 MP3 音频隐写检测方法[P]. 2015-05-20.
- [56] JIN C, WANG R, YAN D. Steganalysis of MP3Stego with Low Embedding-Rate using Markov Feature[J]. *Multimedia Tools and Applications (MTAP)*, 2017, 76(5): 6143–6158.
- [57] REN Y, XIONG Q, WANG L. A Steganalysis Scheme for AAC Audio based on MDCT Difference between Intra and Inter Frame[C]//Proceedings of the 16th International Workshop on Digital Forensics and Watermarking (IWDW 2017). Magdeburg, Germany: Springer, 2017: 217–231.
- [58] PAULIN C, SELOUANI S, HERVET E. Audio Steganalysis using Deep Belief Networks[J]. *International Journal of Speech Technology (IJST)*, 2016, 19(3): 585–591.
- [59] HINTON G E, OSINDERO S, TEH Y W. A Fast Learning Algorithm for Deep Belief Nets [J]. *Neural Computation*, 2006, 18(7): 1527–1554.
- [60] SWANSON E, GANIER C, HOLMAN R, et al. Frequency Domain Steganography[EB/OL]. 2002. https://www.clear.rice.edu/elec301/Projects01/smoke_steg/group.html.
- [61] CHEN B, LUO W, LI H. Audio Steganalysis with Convolutional Neural Network[C]// Proceedings of the 5th ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec 2017). Philadelphia, Pennsylvania, USA: ACM, 2017: 85–90.
- [62] WANG Y, YANG K, YI X, et al. CNN-based Steganalysis of MP3 Steganography in the Entropy Code Domain[C]//Proceedings of the 6th ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec 2018). Innsbruck, Austria: ACM, 2018: 55–65.
- [63] 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 (TIFS)*, 2018, 13(7): 1854–1868.
- [64] YANG Z, YANG H, HU Y, et al. Real-Time Steganalysis for Stream Media based on Multi-Channel Convolutional Sliding Windows[J]. *CoRR*, 2019, abs/1902.01286.
- [65] REN Y, LIU D, XIONG Q, et al. Spec-ResNet: A General Audio Steganalysis Scheme based on Deep Residual Network of Spectrogram[J]. *CoRR*, 2019, abs/1901.06838.
- [66] JISC, SUZUKI T, SAUVAGE S, et al. ISO/IEC 11172-3:1993 Information Technology - Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1,5 Mbit/s - Part 3: Audio[EB/OL]. 1993. <https://www.iso.org/standard/22412.html>.
- [67] UPHAM D. Jsteg[EB/OL]. 1992. <https://zoooid.org/~paul/crypto/jsteg>.

- [68] WESTFELD A. F5-A Steganographic Algorithm[C]//Proceedings of the 4th International Workshop on Information Hiding (IH 2001). Pittsburgh, Pennsylvania, USA: ACM, 2001: 289–302.
- [69] BALLEGOOY A, DOKIC G. 8Hz - MP3[EB/OL]. 1999. <http://www.8hz.com>.
- [70] FILLER T, FRIDRICH J. Minimizing Additive Distortion Functions with Non-Binary Embedding Operation in Steganography[C]//Proceedings of the 2010 IEEE International Workshop on Information Forensics and Security (WIFS 2010). Seattle, Washington, USA: IEEE, 2010: 1–6.
- [71] FILLER T, PEVNY T, REPUBLIC C, et al. BOSS: Be the boss of the Boss, Break Our Steganographic System![EB/OL]. 2011. <http://agents.fel.cvut.cz/boss>.
- [72] LEIDINGER A, DE BRITO R T, HEGEMANN R, et al. LAME MP3 Encoder[EB/OL]. 1998. <http://lame.sourceforge.net>.
- [73] WANG Y, GUO L, WANG C. Steganography Method for Advanced Audio Coding[J]. Journal of Chinese Computer Systems (JCCS), 2011, 32(7): 1465–1468.
- [74] ZHU J, WANG R, YAN D. The Sign Bits of Huffman Codeword-based Steganography for AAC Audio[C]//Proceedings of the 2010 International Conference on Multimedia Technology (ICMT 2010). Ningbo, Zhejiang, China: IEEE, 2010: 1–4.
- [75] 朱杰. 适于 MPEG-2/4 Advanced Audio Coding(AAC) 音频的信息隐藏技术研究[D]. 浙江: 宁波大学, 2011.
- [76] CHANG C C, LIN C J. LIBSVM: A Library for Support Vector Machines[J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2011, 2: 27:1–27:27.
- [77] KODOVSKÝ J, FRIDRICH J, HOLUB V. Ensemble Classifiers for Steganalysis of Digital Media[J]. IEEE Transactions on Information Forensics and Security (TIFS), 2012, 7(2): 432–444.
- [78] FRIDRICH J, KODOVSKÝ J. Rich Models for Steganalysis of Digital Images[J]. IEEE Transactions on Information Forensics and Security (TIFS), 2012, 7(3): 868–882.
- [79] HOLUB V, FRIDRICH J. Low-Complexity Features for JPEG Steganalysis using Undecimated DCT[J]. IEEE Transactions on Information Forensics and Security (TIFS), 2015, 10(2): 219–228.
- [80] SONG X, LIU F, YANG C, et al. Steganalysis of Adaptive JPEG Steganography using 2D Gabor Filters[C]//Proceedings of the 3rd ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec 2015). Portland, Oregon, USA: ACM, 2015: 15–23.
- [81] ZHANG Y, LIU F, YANG C, et al. Steganalysis of Content-adaptive JPEG Steganography based on Gauss Partial Derivative Filter Bank[J]. Journal of Electronic Imaging (JEI), 2017, 26(1): 13011.
- [82] HOLUB V, FRIDRICH J. Random Projections of Residuals for Digital Image Steganalysis [J]. IEEE Transactions on Information Forensics and Security (TIFS), 2013, 8(12): 1996–2006.

- [83] DENEMARK T, SEDIGHI V, HOLUB V, et al. Selection-Channel-Aware Rich Model for Steganalysis of Digital Images[C]//Proceedings of the 2014 IEEE International Workshop on Information Forensics and Security (WIFS 2014). Atlanta, Georgia, USA: IEEE, 2014: 48–53.
- [84] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based Learning Applied to Document Recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278–2324.
- [85] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet Classification with Deep Convolutional Neural Networks[J]. Communications of the ACM, 2017, 60(6): 84–90.
- [86] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition[C]//Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015). San Diego, California, USA: ICLR, 2015.
- [87] SZEGEDY C, LIU W, JIA Y, et al. Going Deeper with Convolutions[C]//Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2015). Boston, Massachusetts, USA: IEEE, 2015: 1–9.
- [88] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016). Las Vegas, Nevada, USA: IEEE, 2016: 770–778.
- [89] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely Connected Convolutional Networks[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017). Honolulu, Hawaii, USA: IEEE, 2017: 2261–2269.
- [90] RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet Large Scale Visual Recognition Challenge[J]. International Journal of Computer Vision (IJCV), 2015, 115(3): 211–252.
- [91] QIAN Y, DONG J, WANG W, et al. Learning and Transferring Representations for Image Steganalysis using Convolutional Neural Network[C]//Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP 2016). Phoenix, Arizona, USA: IEEE, 2016: 2752–2756.
- [92] XU G, WU H, SHI Y. Structural Design of Convolutional Neural Networks for Steganalysis [J]. IEEE Signal Processing Letters (SPL), 2016, 23(5): 708–712.
- [93] CHEN M, SEDIGHI V, BOROUMAND M, et al. JPEG-Phase-Aware Convolutional Neural Network for Steganalysis of JPEG Images[C]//Proceedings of the 5th ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec 2017). Philadelphia, Pennsylvania, USA: ACM, 2017: 75–84.
- [94] YE J, NI J, YI Y. Deep Learning Hierarchical Representations for Image Steganalysis[J]. IEEE Transactions on Information Forensics and Security (TIFS), 2017, 12(11): 2545–2557.
- [95] ZENG J, TAN S, LI B, et al. Large-Scale JPEG Image Steganalysis using Hybrid Deep-Learning Framework[J]. IEEE Transactions on Information Forensics and Security (TIFS), 2018, 13(5): 1200–1214.

- [96] LI B, WEI W, FERREIRA A, et al. ReST-Net: Diverse Activation Modules and Parallel Subnets-based CNN for Spatial Image Steganalysis[J]. IEEE Signal Processing Letters (SPL), 2018, 25(5): 650–654.
- [97] WU S, ZHONG S, LIU Y. Deep Residual Learning for Image Steganalysis[J]. Multimedia Tools and Applications (MTAP), 2018, 77(9): 10437–10453.
- [98] YEDROUDJ M, COMBY F, CHAUMONT M. Yedroudj-Net: An Efficient CNN for Spatial Steganalysis[C]//Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2018). Calgary, Alberta, Canada: IEEE, 2018: 2092–2096.
- [99] WU S, ZHONG S, LIU Y. Residual Convolution Network based Steganalysis with Adaptive Content Suppression[C]//Proceedings of the 2017 IEEE International Conference on Multimedia and Expo (ICME 2017). Hong Kong, China: IEEE, 2017: 241–246.
- [100] BOROUMAND M, CHEN M, FRIDRICH J. Deep Residual Network for Steganalysis of Digital Images[J]. IEEE Transactions on Information Forensics and Security (TIFS), 2019, 14(5): 1181–1193.
- [101] GOWDA S N, YUAN C. ColorNet: Investigating the Importance of Color Spaces for Image Classification[J]. CoRR, 2019, abs/1902.00267.
- [102] LIN M, CHEN Q, YAN S. Network In Network[C]//Proceedings of the 2nd International Conference on Learning Representations (ICLR 2014). Banff, Canada: ICLR, 2014.
- [103] SMITH S L, KINDERMANS P, LE Q V. Don't Decay the Learning Rate, Increase the Batch Size[C]//Proceedings of the 6th International Conference on Learning Representations (ICLR 2018). Vancouver, British Columbia, Canada: ICLR, 2018.
- [104] GOODFELLOW I, BENGIO Y, COURVILLE A. Deep Learning[M]. Cambridge, Massachusetts, USA: MIT Press, 2016.
- [105] LONG J, SHELHAMER E, DARRELL T. Fully Convolutional Networks for Semantic Segmentation[C]//Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2015). Boston, Massachusetts, USA: IEEE, 2015: 3431–3440.
- [106] KINGMA D P, BA J. Adam: A Method for Stochastic Optimization[C]//Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015). San Diego, California, USA: ICLR, 2015.
- [107] GLOROT X, BENGIO Y. Understanding the Difficulty of Training Deep Feedforward Neural Networks[C]//Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS 2010). Chia Laguna Resort, Sardinia, Italy: PMLR, 2010: 249–256.
- [108] HOCHREITER S, SCHMIDHUBER J. Long Short-Term Memory[J]. Neural Computation, 1997, 9(8): 1735–1780.
- [109] CHO K, VAN MERRIENBOER B, GÜLÇEHRE Ç, et al. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation[C]//Proceedings of the 2014

- Conference on Empirical Methods in Natural Language Processing (EMNLP 2014), A meeting of SIGDAT, a Special Interest Group of the ACL. Doha, Qatar: ACL, 2014: 1724–1734.
- [110] IANDOLA F N, MOSKEWICZ M W, ASHRAF K, et al. SqueezeNet: AlexNet-Level Accuracy with 50x Fewer Parameters and <1MB Model Size[J]. CoRR, 2016, abs/1602.07360.
- [111] HOWARD A G, ZHU M, CHEN B, et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications[J]. CoRR, 2017, abs/1704.04861.
- [112] ZHANG X, ZHOU X, LIN M, et al. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices[C]//Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2018). Salt Lake City, Utah, USA: IEEE, 2018: 6848–6856.
- [113] CHOLLET F. Xception: Deep Learning with Depthwise Separable Convolutions[C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017). Honolulu, Hawaii, USA: IEEE, 2017: 1800–1807.
- [114] ARORA S, BHASKARA A, GE R, et al. Provable Bounds for Learning Some Deep Representations[C]//Proceedings of the 31st International Conference on Machine Learning (ICML 2014). Beijing, China: IMLS, 2014: 584–592.
- [115] IOFFE S, SZEGEDY C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[C]//Proceedings of the 32nd International Conference on Machine Learning (ICML 2015). Lille, France: IMLS, 2015: 448–456.
- [116] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the Inception Architecture for Computer Vision[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016). Las Vegas, Nevada, USA: IEEE, 2016: 2818–2826.
- [117] SZEGEDY C, IOFFE S, VANHOUCKE V, et al. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning[C]//Proceedings of the 31st Conference on Artificial Intelligence (AAAI 2017). San Francisco, California, USA: AAAI, 2017: 4278–4284.
- [118] CHU C, ZHMOGINOV A, SANDLER M. CycleGAN, A Master of Steganography[J]. CoRR, 2017, abs/1712.02950.
- [119] ZHU J, KAPLAN R, JOHNSON J, et al. HiDDeN: Hiding Data with Deep Networks[C]// Proceedings of the 15th European Conference on Computer Vision (ECCV 2018). Munich, Germany: Springer, 2018: 682–697.
- [120] SHI H, DONG J, WANG W, et al. SSGAN: Secure Steganography based on Generative Adversarial Networks[C]//Proceedings of the 18th Pacific-Rim Conference on Multimedia (PCM 2017). Harbin, Heilongjiang, China: Springer, 2017: 534–544.

附录 A 音频隐写分析数据集

数据集分别包含了原始 WAV 音频、裁剪后的 WAV 音频、正常 MP3 音频与隐写 MP3 音频、所提取的 QMDCT 系数矩阵、MP3 音频编解码器、隐写编码器与消息提取器以及相关的处理脚本等，具体信息如下。

表 A.1 数据集详细信息

Table A.1 Details of dataset

项目	参数	值
WAV 音频	数量	33038
	声道数	立体声 (2)
	时长	10 s
	比特率	1411kbps
	量化深度	16bits
	采样率	44.1kHz
	类型	歌曲（流行、民谣、摇滚等）
Lame 编码器	版本	Lame-3.99.5
	比特率	128 / 192 / 256 / 320kbps
	其他参数	默认参数
MP3Stego 编码器	版本	MP3Stego-1.1.18
	比特率	128 / 192 / 256 / 320kbps
	其他参数	默认参数
MP3 隐写算法	HCM	$SPR = 0.1 / 0.3 / 0.5 / 0.8 / 1.0$
	EECS	$h = 7$ $SPR = 2 / 3 / 4 / 5 / 6$
QMDCT 系数矩阵	存档格式	txt
	尺寸	$200 \times 576 / 400 \times 576$

附录 B 隐写负载率转换表

表 B.1 隐写负载率转换表 (MP3Stego)

Table B.1 Cross-references of steganographic payloads (MP3Stego)

比特率 / kbps	隐写负载率	隐写速率 / bps	相对嵌入率 (%)
128	0.1	15	2.13
	0.3	46	6.71
	0.5	76	9.36
	0.8	122	17.92
	1.0	152	20.54
192	0.1	15	4.35
	0.3	46	12.60
	0.5	76	15.63
	0.8	122	23.35
	1.0	152	28.83
256	0.1	15	5.85
	0.3	46	17.71
	0.5	76	18.81
	0.8	122	29.40
	1.0	152	32.80
320	0.1	15	6.89
	0.3	46	8.62
	0.5	76	9.65
	0.8	122	12.08
	1.0	152	21.43

表 B.2 隐写负载率转换表 (HCM)

Table B.2 Cross-references of steganographic payloads (HCM)

比特率 / kbps	隐写负载率	隐写速率 / bps	相对嵌入率 (%)
128	0.1	1214	0.86
	0.3	3641	3.04
	0.5	6069	5.30
	0.8	9710	9.10
	1.0	12138	11.64
192	0.1	1368	1.64
	0.3	4103	5.41
	0.5	6839	9.71
	0.8	10924	15.56
	1.0	13678	20.22
256	0.1	1687	2.47
	0.3	5061	7.75
	0.5	8436	13.29
	0.8	13497	21.29
	1.0	16871	26.43
320	0.1	2103	2.72
	0.3	6309	7.92
	0.5	10515	13.31
	0.8	16824	21.31
	1.0	21030	26.51

表 B.3 隐写负载率转换表 (EECS)

Table B.3 Cross-references of steganographic payloads (EECS)

比特率 / kbps	隐写负载率	隐写速率 / bps	相对嵌入率 (%)
128	2	5721	4.06
	3	3775	2.49
	4	2794	1.79
	5	2142	1.32
	6	1580	0.96
192	2	8032	7.66
	3	5325	4.62
	4	3974	3.45
	5	3164	2.58
	6	2626	2.17
256	2	8578	9.87
	3	6894	6.07
	4	5155	4.31
	5	4109	3.33
	6	3404	2.78
320	2	11846	10.51
	3	7865	6.32
	4	5882	4.34
	5	4682	3.33
	6	3893	2.87

附录 C 音频分享平台

表 C.1 音频分享平台列表

Table C.1 List of audio sharing platforms

平台	平台链接
FMA	http://freemusicarchive.org
Clyp	https://clyp.it
Splice	https://splice.com
Vocaroo	https://vocaroo.com
Podcasts	http://podcasts.com
Chirbit	https://www.chirbit.com
Insaudio	https://instaudio.io
Mixcloud	https://www.mixcloud.com
Free Sound	https://freesound.org
YourListen	https://yourlisten.com
SoundCloud	https://soundcloud.com
Looperman	https://www.looperman.com/loop
PureVolume	https://www.purevolume.com
ReverbNation	https://www.reverbnation.com
喜马拉雅 FM	https://www.ximalaya.com

致 谢

感谢导师赵险峰研究员和易小伟老师对我的精心指导和培养，他们的言传身教让我终生受益。

感谢中科院信工所媒体安全与智能分析研究组各位老师和同学的指导与帮助。

感谢父母对我的养育和教导，感谢好朋友们的支持与关心。

感谢参与论文评审和答辩的各位专家教授。

作者简历及攻读学位期间发表的学术论文与科研成果

作者基本情况

王运韬，男，1993年8月出生，山西省临汾人。

2016年6月毕业于西安邮电大学通信工程专业并获得工学学士学位；

2016年9月至今于中国科学院信息工程研究所信息安全国家重点实验室攻读工学硕士学位。

已发表（或正式接收）的学术论文

1. 王运韬, 杨坤, 易小伟, 赵险峰. 基于块内块间相关性的 MP3 隐写分析特征 [C]. 第十四届全国信息隐藏暨多媒体信息安全学术大会 (CIHW 2018). 广东, 中国. 2018: 829-836 (大会优秀论文)
2. **Yuntao Wang**, Kun Yang, Xiaowei Yi, Xianfeng Zhao, Zhoujun Xu. CNN-based Steganalysis of MP3 Steganography in the Entropy Code Domain[C]. Proceedings of the 6th ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec 2018). Innsbruck, Austria, June 20-22, 2018: 55-65. (Best Paper Award, CCF-C, EI 索引)
3. **Yuntao Wang**, Xiaowei Yi, Xianfeng Zhao, Ante Su. RHFCN: Fully CNN-based Steganalysis of MP3 with Rich High-Pass Filtering[C]. Proceedings of the 44th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2019). Brighton, UK, May 12-17, 2019: 2627-2631. (已录用, CCF-B, EI 索引)
4. **Yuntao Wang**, Xiaowei Yi, Xianfeng Zhao. MP3 Steganalysis Based on Joint Point-wise and Block-wise Correlations[J]. Information Sciences, 2019. (一审大修, CCF-B, SCI 索引, 影响因子 4.305, JCR-2)
5. Xiaowei Yi, Kun Yang, Xianfeng Zhao, **Yuntao Wang**, Haibo Yu. AHCM: Adaptive Huffman Code Mapping for Audio Steganography Based on Psychoacoustic Model[J]. IEEE Transactions on Information Forensics and Security (TIFS), 2019. (已录用, CCF-A, SCI 索引, 影响因子 5.824, JCR-2)
6. Yunzhao Yang, **Yuntao Wang**, Xiaowei Yi, Xianfeng Zhao, Yi Ma. Defining Joint Embedding Distortion for Adaptive MP3 Steganography in Sign Bit Domain[C]. Pro-

ceedings of the 7th ACM Workshop on Information Hiding and Multimedia Security(IH&MMSec 2019). Pairs, France, July 3-5, 2019. (已录用, CCF-C, EI 索引)

申请或已获得的专利

1. 易小伟, 李金才, 王运韬, 赵险峰, 于海波, 刘长军. 一种基于 IS4 软件特征的隐藏信息检测及提取方法技术(实质审查的生效, 公开/公告号: CN106845242A)

课题项目参与

1. 国家自然科学基金联合基金项目“缺乏载体先验知识的隐写分析模型与方法研究”
2. 国家自然科学基金联合基金项目“适应多种社交网络信道的安全隐写研究”
3. 国家重点研发计划课题“大数据挖掘技术及系统”
4. 中国科学院战略性先导专项课题“数字媒体安全防护关键技术”

在学期间所获奖励

1. 第十四届全国信息隐藏暨多媒体信息安全学术大会(CIHW2018)优秀论文奖
2. 2018 ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec 2018) Best Paper Award
3. 2017-2018 学年研究生国家奖学金
4. 2017-2018 学年中国科学院信息工程研究所“所长特别奖”
5. 2017-2018 学年中国科学院大学三好学生荣誉称号
6. 2018-2019 学年中国科学院大学三好学生荣誉称号

其他

论文 PDF 版终稿与相关实验代码及说明已共享至 Github

Github: https://github.com/Charleswyt/master_dissertation