计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.190600103

图像隐写分析算法研究概述

彭 伟[1] 胡 宁[2] 胡璟璟[1]
1 国防科技大学计算机学院 长沙 410073
2 广州大学网络空间安全研究院 广州 510006

摘 要 图像隐写技术可以在互联网上传输各种数字图片中隐藏的敏感或秘密信息,在过去二十多年中得到了快速的发展,并被不法分子用来交换可能危害社会安全的信息。为消除这些危害,相应发展了各种图像隐写分析技术。通过检查可疑图片中隐藏的秘密信息,图像隐写分析可以提供数字法理证据。在图像隐写算法发展现状分析的基础上,将图像隐写分析算法分为专用和通用隐写分析算法两大类,对图像隐写分析技术进行了介绍和归纳。在专用算法方面,分别介绍了针对特定图像隐写算法和针对特定图像类型的图像隐写分析途径。在通用算法方面,介绍了基于图像特征的图像隐写分析方法的一般流程,归纳总结了图像隐写分析常用的几类图像特征。通过回顾图像隐写分析的已有工作,分析了图像隐写分析中采用的技术,包括基于机器学习的分类方法、特征选择方法等。最后,对图像隐写分析的未来研究发展方向做了简要的讨论。

关键词:信息隐藏;图像隐写分析;图像隐写分析算法;机器学习;特征选择
中图法分类号 TP309.2

Overview of Research on Image Steganalysis Algorithms

PENG Wei[1], HU Ning[2] and HU Jing-jing[1]
1 College of Computer, National University of Defense Technology, Changsha 410073, China
2 Institute of Cyber-space Security, Guangzhou University, Guangzhou 510006, China

Abstract Image steganography is the technique to hide sensitive or secret data in digital pictures transmitted on the Internet. It has gone through fast development during the past two decades, and is utilized by criminals including terrorists to exchange information which may threaten social security. Many kinds of image steganalysis techniques have been developed to fight back the threat. By examining the secret information hidden in the suspicious images, image steganalysis can provide digital forensic evidence. This paper firstly gave a survey on the research status of algorithms of image steganography, then introduced and summarized the image steganalysis techniques by classifying them into two categories: specialized algorithms and generalized algorithms. For specialized algorithms, the approaches designed for specific image steganography algorithms and specific image types are introduced respectively. For generalized algorithms, the general procedures of image steganalysis based on image features are described. Then several classes of image features used for image steganalysis are summarized. Furthermore, the techniques used in general image steganalysis including machine learning-based classification and feature selection are analyzed by reviewing the existing research work on image stenanalysis. At last, a brief discussion on future research directions of image steganalysis is presented.

Keywords Information hiding, Image steganalysis, Image steganalysis algorithm, Machine learning, Feature selection

1 引言

网络通信除了通过加解密来保证数据安全以外,有时还需要隐藏用户的通信行为,以保护敏感用户的隐私信息。信息隐藏是一种兼顾数据安全和通信行为的隐藏技术,随着互联网的发展其得到了发展和应用。信息隐藏通过将隐秘信息隐藏在语音、图像、文本、视频、通信协议等载体之中,较好地保护了用户数据的安全。由于互联网上图像的广泛可获取性以及人类在图像上的视觉冗余特性,图像是一种很适合进行信息隐藏的载体。过去若干年中,针对图像隐写技术开展了大量研究,取得了许多研究成果,提出了多种基于图像的信息隐藏算法。图像隐写技术除了被合法使用外,也可能被不法分子或敌对势力等用来从事恐怖犯罪活动或间谍活动。为了防止图像隐写技术被不法分子利用而给社会和国家带来危害,需要针对性地展开图像隐写分析技术的研究和应用。

隐写分析(Steganalysis)指采用各种技术手段对数字载体进行分析,从中检测和提取出隐藏的信息。隐写分析一般可以分为几个阶段:首先检测判断目标对象中是否存在隐秘信息;其次识别目标对象所使用的隐写算法;然后确定隐写算法所使用的参数,包括定位隐秘信息的嵌入位置和信息量大小;最后从目标对象中提取出隐写数据,供进一步的解密使用。由于提取隐写信息的难度较大,目前的研究大多还停留在隐藏信息的检测和判断上。如何从大量普通的信息中检测出可能藏有秘密的信息,是一项很复杂的技术工程。

近年来,研究人员提出了大量针对数字图像的隐写方法[1-5]。相比之下,图像隐写分析的研究落后于图像隐写技

通信作者:彭伟(wpeng@nudt.edu.cn)

术，主要原因在于针对数字图像的隐写分析存在以下几方面的困难[6]：1)隐写嵌入的低扰动性。隐写嵌入对原图像的修改通常很小，视觉上不会产生明显改变。许多较先进的图像隐写算法均努力保持主要的图像统计特征不发生明显变化，大大增加了图像隐写分析的难度。2)图像类型和隐写算法的多样性。由于编码方法的不同，产生了各种各样的图像格式类型，如 JPEG，GIF，TIFF，PNG 等。同时，基于不同的研究思路提出了多种多样的图像隐写算法。这些都增加了图像隐写分析的复杂性和难度。3)隐写数据和过程的随机性。隐写数据通常经过加密后变为了随机性很强的数据，加上图像隐写过程中有意设置的随机性，使得难以从图像中检测出隐写嵌入的数据。

根据适用范围不同，一般将图像隐写分析分为针对特定对象或特定隐写术的分析方法和针对一大类目标对象的通用型分析方法。前者针对特定类型的图像或特定的图像隐写算法，能够利用图像类型或图像隐写算法的特点进行设计，因此检测准确性较高，但应用范围受限。通用型分析方法一般针对一大类图像或图像隐写算法进行分析设计，目标是在不确定具体隐写算法的情况下检测出隐写带来的图像变化，通常通过从图像中提取多个统计特征来进行分析，可适用的范围较大，但检测准确性较难提高。

对隐写分析技术的性能评价主要从准确性、适用性、实时性等方面来考虑。准确性指正确地检测出目标载体中是否存在隐秘信息的成功率，或者指隐藏消息检测、识别或估计的精度。适用性或通用性指隐写分析方法是否适用于大量的图像隐写术和图像类型。实时性指隐写分析算法的计算开销或运行速度，在线隐写分析系统或面向大数据量的隐写分析系统一般要求较好的实时性。

本文首先对图像隐写算法进行简要的介绍，然后分类介绍当前在图像隐写分析算法方面的研究现状，最后对研究现状进行归纳总结并探讨图像隐写分析方法的发展趋势。

2 图像隐写算法

从是否改变载体或原始图像数据来看，图像隐写算法分为嵌入型隐写算法和无载体隐写算法。

2.1 嵌入型图像隐写算法

嵌入型隐写算法直接将隐秘信息嵌入到图像载体中，隐写容量较高，是当前研究最多的途径。嵌入型隐写算法可以根据不同的标准进行进一步细分，如根据嵌入域可分为空域算法和变换域算法，根据图像维度可分为二维图像隐写算法和三维图像隐写算法等[7]。为提高图像隐写的安全性，在图像隐写算法中可以使用一些自适应技术，如各种机器学习或人工智能的技术。

2.1.1 基于嵌入域的隐写算法

根据隐秘信息的嵌入域可以将图像隐写算法分为空域算法和变换域算法。

(1)空域算法

空域算法通过直接修改载体图像的像素等来嵌入隐秘信息，其思想较简单、实现较容易，因此较早得到研究。这类算法包括 LSB(Least Significant Bit)算法[8]、PVD(Pixel Value Differencing)算法[9]、BPCS(Bit Plane Complexity Segmentation)算法[10]等。LSB 算法的思想在于，图像像素的最低位只表示很微小的信息，改变像素的最低位不会引起视觉上的明显变化。LSB 算法简单易实现，但容易被图像直方图和 RS(Regular Singular)隐写分析[11]攻击，因而后续提出了一些改进算法，如 LSB 匹配算法[12]、动态补偿 LSB 隐写算法、LSB 匹配重访(LSB Matching Revisited)算法等。PVD 算法通过比较相邻像素的差值来嵌入隐秘信息，在平滑性上强于 LSB 算法。PVD 算法可以与 LSB 算法相结合来设计更有效的图像隐写方法。除了直接修改像素值以外，根据图像类型的不同，还可以使用多个图像的位平面、调色板数据等来隐藏信息。在嵌入隐秘信息时有两种策略：顺序和分散。在顺序策略下，在图像像素中顺序地嵌入隐秘信息。在分散策略下，隐秘信息嵌入图像的位置由某种随机顺序确定，使得隐秘信息分散嵌入到整个图像中。

(2)变换域算法

变换域算法将隐秘信息嵌入到图像的变换域数据中，例如 JPEG 图像中的 DCT(Discrete Cosine Transform)系数。在变换域嵌入和解码隐藏信息比在空域复杂得多，这使得变换域图像隐写算法具有更好的安全性。同时，变换域数据较少受到图像压缩、剪切、旋转等的影响，可以应对图像传输中对图像的变换分析。变换域算法的嵌入容量通常小于空域算法的嵌入容量。

变换域算法包括基于 DCT、DFT(Discrete Fourier Transform)、DWT(Discrete Wavelet Transform)、IWT(Integer Wavelet Transform)、CWT(Complex Wavelet Transform)的算法以及在它们基础上的衍生算法。DCT 是常用的将图像从空域变换到频域的有效办法，图像隐写时将隐藏信息嵌入到 DCT 系数中，主要适用于 JPEG 图像。基于 DCT 的算法在图像隐写时需要考虑 DCT 系数的分布，以取得良好的隐写效果。研究中常用作比较的有 F5 算法[13]、Outguess 算法[14]等。DFT 也是图像处理中常用的技术，图像隐写时可将隐藏信息嵌入到 DFT 的系数中。当使用 DCT 和 DFT 系数进行隐写时，常用 LSB 算法的思想在系数最低位嵌入隐秘信息。由于小波变换的系数可同时表示图像的空间和频率特征，在嵌入隐秘信息时，基于小波变换的方法可以取得更好的隐写不可感知性和隐藏信息的不可恢复性[14]。在使用变换域系数嵌入隐秘信息时，需要保持系数分布特征不出现明显改变。例如，基于模型的隐写算法[15]可保持图像 DCT 系数的边缘统计分布，从而可对抗针对边缘统计分布的隐写分析。基于 DCT 的算法通常要跳过值为 0 的 DCT 系数，以避免图像特征出现明显改变。此外，压缩感知方法也被用来设计有效的变换域图像隐写算法[16]。

2.1.2 基于图像维度的隐写算法

从图像维度上看，已有算法可分为二维和三维图像隐写算法。二维算法中，隐秘数据被嵌入到载体图像的二维平面中。隐写对象可以是空域的像素值或变换域的系数值。二维图像隐写的方法也可以直接推广应用到三维图像中，最直接的就是在三维图像的各个平面分别应用二维图像隐写方法。为了取得更好的隐写安全性，需要针对三维图像的特点设计专门的三维图像隐写算法。三维图像隐写可以在几何域、拓扑域或表示域实现[17]。在三维几何模型中，隐秘信息可被嵌入到图像的节点数据中。这种几何域的嵌入容量高于拓扑域和表示域的图像隐写方法。相比二维图像，三维图像具有更

丰富的特征，因此三维图像隐写算法更为复杂，但安全性和嵌入容量也更高。

2.1.3　图像隐写的自适应技术

自适应技术的主要目的是在图像隐写时，通过自适应地选择嵌入位置、隐写计算方法、嵌入位的多少等来使得隐写后的图像具有较高的安全性。从研究思路来看，这些技术可分为基于区域的隐写术、基于人类视觉系统(Human Vision System)的隐写术、基于机器学习和人工智能方法的隐写术等。

基于区域的隐写术主要寻找图像中适合嵌入信息的区域来完成隐写。这些区域一般是图像细节较为复杂丰富而不是较平滑的区域，对这些区域的改变不会引起视觉上的可观察变化[18]。同时，在嵌入隐秘信息时还可根据局部图像特征自适应地选择嵌入信息的位数。限制图像隐写的区域会带来嵌入容量一定程度的降低，因此通常需要在嵌入容量和隐写安全性之间取得折中。基于人类视觉系统的隐写术主要利用人类视觉上的冗余性等特点来实现图像隐写[19-20]。人类视觉有许多盲点，在某些背景下难于区分图像的细节，对图像平滑区域色彩强度的微小改变不敏感，难于感知复杂图像区域的较小变化。因此，可以利用人类视觉的这些特点来自适应地选择图像隐写的区域，再利用空域或变化域的方法来完成隐秘信息的嵌入。机器学习和人工智能方法可以较好地在保持图像隐写安全性的同时提高嵌入容量。常用的机器学习方法包括支持向量机(SVM)、决策树、遗传算法(GA)、神经网络、模糊逻辑方法等。

2.2　无载体图像隐写算法

无载体隐写算法不直接改变图像载体数据，而是通过图像的特征属性等来表示要传输的隐秘信息，因此可以对抗各种针对嵌入型隐写算法的隐写分析技术，其隐秘性更强。Zhou 等提出一种基于图像灰度值的无载体隐蔽通信方案[21]，利用图像不同区域灰度值大小差异来表征要传输的隐蔽信息。Wu 等提出基于灰度级梯度共生矩阵的图像隐写方法[22]。Ruan 等提出一种基于 GIF 图像的无载体隐蔽通信方法[23]，基本思想是分析 GIF 图像载体的拓展模块属性，根据属性的特性，将其按照一定规则映射到隐秘消息空间。Zhang 等提出一种基于 DCT 和 LDA 主题分类的无载体图像隐写方法[24]。Duan 等提出一个用于无载体图像隐写的生成模型[25]，通过生成模型将要传输的隐秘图像变换为一个普通的图像，从而达到隐写的目的。由于载体的属性能表征的信息很有限，因此无载体隐写技术嵌入容量较小，研究成果尚不多，但由于其良好的抗隐写分析能力正得到越来越多的研究。

3　图像隐写分析算法

图像隐写分析的思路有：1)寻找图像数据由于隐写嵌入而发生的特殊改变或特殊模式，例如图像数据某些字段可能发生有规律的变化；2)寻找图像统计特征的变化，这类方法采用统计判别的思想，即从图像中计算统计特征信息，并据其判决图像是否为载密图像。大多数隐写分析方法属于统计判别类的方法，其检测前提是隐写嵌入信息在一定程度上将改变图像数据的全局统计特性。从适用范围来看，图像隐写分析可分为专用和通用隐写分析算法两大类。

3.1　专用隐写分析算法

3.1.1　针对特定隐写算法的隐写分析

LSB 算法是较早提出的图像隐写算法，针对该算法的专用隐写分析方法较为成熟，产生了许多研究成果。这些隐写分析技术主要利用了 LSB 隐写嵌入带来的图像特征的改变。通过分析图像像素最低位平面的某些异常特性可以检测是否存在隐秘信息。LSB 替换隐写会使图像 DCT 系数值的相邻对值(Pairs of Values)的取值趋于一致，从而可以通过衡量相邻对值的接近程度来判断载体中是否含有秘密信息，基于此提出了 χ^2 算法[26]。该算法的具体实现为：1)计算每组对值的均值作为“预测分布”，然后与目标载体的实际对值取值分布进行比较，最后计算两者的 χ^2 统计量来得出目标载体中包含秘密信息的概率。算法在隐秘消息顺序嵌入时较为有效，当隐秘信息随机分散嵌入时效果不好；2)采用空域 LSB 隐写嵌入数据后图像系数的空间相关性会降低，基于此提出了针对空域 LSB 替换隐写的 RS 分析方法[11]，其主要思想如下：首先将图像像素按一定顺序等分为若干像素组，然后将像素值按照 LSB 替换前后判别函数的取值分为规则类、异常类和不可使用类，最后对每类的个数进行统计，依据替换前后每类个数的变化来构造方程，从而计算得出嵌入的秘密信息量；3)通过提取局部相关采样点的高维联合特征值也可以进行隐写分析[27]，该方法利用图像在横向、纵向的多个像素点作为隐写分析的基本单位，具有比 RS 分析方法更高的检测率。

LSB 匹配算法比 LSB 替换算法有更高的安全性，但图像经过该算法嵌入信息后统计特性也会有改变。针对 LSB 匹配算法提出了一种基于噪声高阶统计特性的隐写分析方法[28]，该方法首先在 DCT 域提取噪声的绝对中心矩，然后用 FLD 分类器进行训练分类，取得了较高的检测率。BPCS 算法亦能被统计分析方法攻击[29]。

早期比较有名的图像隐写算法有 F5，Outguess，EzStego 等，针对这些算法提出了一些隐写分析算法，分别使用了不同的图像特征和检测模型[30]。针对 F5 算法提出了一种基于平方误差最小准则的隐写分析算法[31]。该算法将载密图像解压到空域后进行上 4 行左 4 列裁剪，以估计原始载体图像，再将其与载密图像进行 DCT 系数直方图的比较分析，以此为依据进行载密图像判断，进一步估计出载密图像的嵌入容量。基于 0/1 系数组合差异也可对 F5 隐写算法的载密图像进行分析[32]，该算法提取同一分块内、水平和垂直相邻的两个分块内相邻频率位置 0/1 系数组合差异的 12 维特征，并运用支持向量机进行分类。

3.1.2　针对特定图像类型的隐写分析

JPEG 是最常用的图像类型，针对 JPEG 图像提出了许多隐写算法，如 Jsteg，JPHide 算法等。Jsteg 算法实质上是将空域 LSB 隐写算法引入到 DCT 变换域，因此同样可以采用 χ^2 算法进行检测分析。其他 JPEG 隐写算法一般都会对系数相邻对值趋于一致这一统计特性进行补偿，因此 χ^2 算法基本没有效果。扩展 χ^2 算法[33]不直接对整幅图像进行检测，而是对图像进行分段检测，通过提供更丰富的检测数据，能够针对 JPHide 等嵌入位置随机的 JPEG 图像隐写算法进行分析，并能够粗略估算嵌入信息量。

基于 Jsteg 的嵌入特点以及 DCT 系数的对称分布特性，提出了一种估计信息嵌入率方法[34]，其主要原理是分析隐写图像 DCT 系数直方图的分布。在 χ^2 检测法的基础上进一步提出了 CA(Category Attack)和 GCA(Generalized Category

Attack)检测法[35],能够检测 Jsteg 和 JPHide 算法随机嵌入的隐藏信息,具有较高的检测正确率。许多 JPEG 隐写分析算法采用了一种剪切重压缩(Crop and Recompress)的校正方法[36]。该方法先对 JPEG 图像解压缩,然后剪切前面 4 行和左边 4 列,最后用与原图像相同的质量因子来压缩图像。由于该图像的统计特性与载体图像的统计特性接近,因此一般用它作为隐写分析的预测图像。

3.2 通用隐写分析算法

通用隐写分析不针对某种特定的隐写算法,而是寻求一些隐写嵌入会带来影响的图像特征,使用这些特征来训练学习和识别分类,因此具有较好的适用性,能够用来检测多种隐写算法。通用隐写分析算法一般有两个阶段:首先将一些对多种隐写算法都具有一定区分能力的特征向量提取出来,然后通过各种分类器(如支持向量机、神经网络或贝叶斯分类器等)对这些特征向量进行训练和分类检测。其一般流程如图 1 所示。

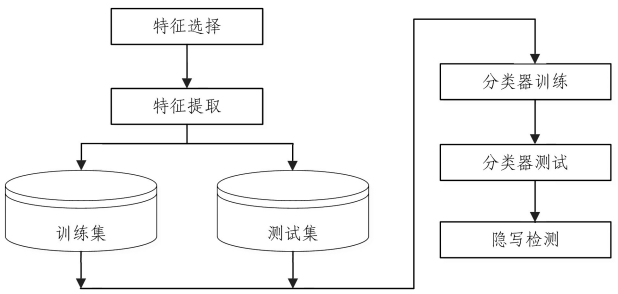

图 1 基于特征的隐写分析方法的一般流程

Fig. 1 General flow of feature-based image steganalysis

由于通用图像隐写分析方法需要利用图像的某些特征以及某些基于机器学习或人工智能的分类器,因此其训练分类器的过程很重要,训练过程中所使用特征的有效性、维度等因素均对隐写分析准确性有较大影响。分类器模型通过两类样本训练得到,即正常图像样本和隐写嵌入后的载密图像样本,根据所使用的分类器对应的有监督学习方法进行训练。由于互联网上有大量的图像或图片,有些网站还提供了专门的图片数据集,因此可以比较容易地获得大量的训练样本。

由于图像本身内容的复杂性和多样性,不同图像的统计特征概率分布具有较大变化。理想的隐写分析特征要能够表征正常图像和载密图像的明显区别,即正常图像与隐写图像的统计特征分布差异要尽可能大。当前已有一些隐写分析特征,根据图像格式的不同可分为针对空域图像的特征和针对变换域图像的特征,从特征的复杂性上可分为低维特征和高维特征。

空域图像包含 BMP、PNG、PGM 等格式,该类格式图像直接存储像素的灰度值。空域图像隐写分析常用的特征如下[3,6]。

(1)相邻像素相关性特征:空域图像的相邻像素之间通常具有一定的相关性,隐写嵌入相当于是在原始图像上增加了噪声扰动,因此将对相邻像素间的相关性带来一定程度的破坏。针对空域的图像隐写分析可通过检测像素间相关性特征的破坏情况来实现。空域图像的特征可采用信号分解的方式来获得,如利用正交镜像滤波器(Quadrature Mirror Filters)、小波分解等方法在各个频带上提取均值、偏差、方差等作为隐写检测的特征[37-38]。He 等定义了图像位平面的 32 个相关性特征,采用支持向量机进行图像隐写分析[39]。

(2)统计矩特征:以马尔可夫转移模型的特征为主的统计矩特征是空域图像隐写分析的主流特征。该类隐写分析方法将图像相邻像素的变化建模为马尔可夫过程,使用图像数据计算马尔可夫概率转移矩阵,并以此为特征进行隐写检测。马尔可夫类特征中具有代表性的特征如 SPAM 特征[40],通过在差分图像上统计二阶马尔可夫概率转移矩阵来得到。通过使用更多的滤波器并计算高阶共生矩阵,空域图像的特征可扩展至高维特征,通过特征的组合和选择使用,使用高维特征可取得比低维特征更好的检测性能,代表性的特征有 SRM[41]和 PSRM[42]。SRM 使用了多种滤波器和多种量化因子来提取特征并进行组合,特征维度高达上万维。Whitaker 等利用 SRM 和 PSRM 的研究表明,使用同一场景下拍摄的部分内容重叠的图片,可有助于检测利用它们进行隐写的载密图片[43]。

(3)图像质量度量特征:隐写嵌入相当于在图像中添加了噪声数据,总会产生或多或少的统计异常,使得图像质量降低。因此,可以采用多变量分析选择一些图像质量度量作为区分原始图像和隐写图像的特征。Avcibas 等从图像空域与 DFT 域提取了 10 个表征图像质量的度量,通过方差分析选择相对有效的度量来组成检测的特征向量集,然后用多元回归分析进行分类,判定图像中是否存在隐写信息[44]。

JPEG 是最常见的变换域图像格式,JPEG 图像存储的基本信息包括图像分块 DCT 变换和经过量化后的 DCT 系数。与空域的统计矩特征类似,从 JPEG 图像存储的 DCT 系数中也可提取相关统计特征。将图像分块内相邻的 DCT 系数变化建模成马尔可夫过程,从而可以提取出 JPEG 马尔可夫特征[45]。基于 JPEG 图像分块 DCT 变换的特点,提出了 274 维的 JPEG 隐写分析特征 PEV274[46]。PEV274 由多种统计量组成,包括 DCT 系数的全局直方图和某些频段的直方图、多种 DCT 系数的共生矩阵等。PEV274 采用了校正技术(Calibration)来增强图像特征对隐写嵌入的敏感性。在 PEV274 基础上提出了 PEV548 特征[47],通过采用拼接操作来代替相减,使得特征维度增加了一倍。Guan 波提出的隐写分析算法使用 243 维特征和支持向量机,适用于对 F5、Outguess 等 JPEG 图像隐写的检测[48]。后续的多种 JPEG 高维隐写分析特征也采用了校正或其扩展的思想,如 CC-JRM[49]、DCTR[50]、GRF[51]和 Wang 等提出的特征集[52]等,使得隐写检测精度有了较大提升。在特征提取前,自适应图像隐写分析(AIS)算法使用了人工蜂群(Artificial Bee Colony)技术来选择最优的图像区域[53]。

通用隐写分析技术使用多个统计特征来检测隐写术对图像带来的变化,这些特征通常比专用隐写分析方法使用的统计量更复杂,通常难以通过设置阈值的方式来完成检测,而需要借助训练分类器模型来识别隐写图像。分类器模型一般来自于统计机器学习领域。当隐写分析特征维度较低时,使用支持向量机、神经网络等分类器就可实现,其中以使用高斯核和多项式核的支持向量机较为普遍。例如,Fridrich 提出 CFB(Calibrated Feature Based)隐写分析算法[54],使用剪切后重压缩的校正方法,将图像 DCT 系数一阶和二阶统计特征及空域图像的块效应特征相结合,再用 SVM 分类器来训练并分类。Fu 等提出一种基于经验概率转移矩阵的通用检测技

术[55]，使用经验概率转移矩阵来分别反映嵌入秘密信息前后块内 DCT 系数相关性和块间 DCT 系数相关性的变化，在此基础上提取了 200 维特征，采用 SVM 分类器进行训练和分类。Dong 等通过一种 BFS(Boosting Feature Selection)的特征聚合选择方法从许多待选的高阶特征(包括高阶小波分解图像统计特征、特征函数矩等)中选择相对合适的特征组成特征向量，再用 SVM 来训练和分类检测[56]。Xu 提出一种基于 DCT 系数二阶统计特征的 JPEG 图像隐写分析方法[57]，该方法首先对 JPEG 图像提取共生矩阵特征和差分特征共 50 维特征向量，然后利用 SVM 进行隐写分析。Wang 等提出一种用于隐写分析的快速支持向量分类算法 FC-SS2LM，通过构造最小超球体和双边最大间隔隐写分析模型，使检测模型既能准确构造分类边界又能考虑不同隐写样本的分布特点，从而达到兼顾检测准确性和通用性的目的[58]。Babu 等对图像隐写分析中用到的一些特征提取和分类技术进行了归纳总结[59]。

随着深度学习在图像识别领域的优异表现，卷积神经网络(CNN)逐渐成为一种重要的图像隐写分析方法。Xu 等构建了一个 5 层的 CNN 模型进行隐写分析，针对 S-UNIWARD 和 HILL 隐写算法取得了较高的检测准确率[60]。Ye 等提出的 CNN 模型在 S-UNIWARD、HILL、WOW 算法上也取得了较高的检测准确率[61]。由于深度残差网络(DRN)可以解决 CNN 网络模型的退化问题，因此基于深度残差网络的神经网络技术被用于图像隐写分析[62-63]。CNN 和 DRN 的实验结果都表明，针对空域图像隐写技术，基于 CNN 和 DRN 的方法能取得比基于富模型的方法更好的图像隐写分析性能。

新型的高安全性隐写技术需要使用高维特征完成隐写分析，此时传统的支持向量机分类器难以达到检测性能要求。Qin 针对 JPEG 图像提出一种基于贝叶斯决策的图像隐写分析算法[64]，该算法采用了基于 DCT 域、Markov 模型、多向概率转移矩阵等的多个特征值，将多个特征值分成 5 个特征子集进行局部分类，所得分类结果再采用贝叶斯最小风险决策融合理论得出最终结果，可提高隐写检测准确度。集成 FLD(Fisher Linear Discriminant)分类器的方法中[65]，多个子分类器分别使用部分特征进行判别，再用多个子分类器的结果进行投票决策，可以解决对于高维特征容易过学习的问题，可以处理大量的数据，且具有较高的训练速度。当前的通用型隐写分析方法大多采用了高维特征和集成分类器。

高维特征可以增加特征的丰富程度，是增强隐写检测精度的有效方法。但在特征数量增加的同时，会不可避免地增加一些无效特征，或是具有冗余信息的特征。这些特征在分类器训练时将可能使得训练的分类器模型精度降低、分类器容易过学习等。因此，需要从原始的特征集合中选择有效性高、冗余性小的特征子集。针对高维特征和集成分类器的特征选择具有以下几个难点[6]：

(1)特征选择方法一般采用有监督学习的方法进行，在样本数量有限的条件下，选择特征子集难以避免高维特征带来的过学习问题。

(2)难以与集成分类器结合。集成分类器需要使得多个子分类器对不同的样本检测结果具有多样性，因此子分类器使用的特征应具有多样性和有效性。传统的特征选择方法只考虑选择特征的有效性，没有考虑集成分类器对特征多样性的要求，也没有与集成分类器进行综合优化设计。

针对特征选择问题提出了基于线性规划的特征选择方法[6]，该方法将特征选择过程形式化为一个优化问题，通过求解特征选择模型中每个特征对应的权值参数，再据此进行特征选择。Ma 等提出一种基于决策粗糙集-positive 区域缩减的隐写分析特征选择方法[66]，针对 GFR 等富模型特征取得了较好的特征精简性能。人工智能方法也被用于图像隐写分析的特征选择中，例如使用自适应粒子群优化算法来完成特征选择[67]。

结束语　图像隐写分析是一个理论性强、实践难度大的研究领域。隐写嵌入的低扰动性、图像类型和隐写算法的多样性、隐写数据和过程的随机性给图像隐写分析带来了极大的挑战。到目前为止，基于统计的隐写分析方法取得了比其他途径更好的性能，但隐写分析仍然面临许多挑战性问题，例如特征选择问题、如何设计更好的基于机器学习的隐写分析方法问题、隐写分析方法的通用性问题、隐写参数的估计问题[68]等。尽管近年来取得了一些研究成果，未来图像隐写分析还将随着各种隐写技术的发展而不断发展。

未来图像隐写分析技术的研究方向有：

(1)研究并提出更有效的图像隐写检测特征。嵌入型图像隐写方法本质上相当于在普通图像中添加了扰动或噪声，噪声的随机性会使得图像的某些特征发生改变。依据图像类型的不同，需要进一步研究普通图像对噪声敏感的特征，包括空域和频域等不同域的多维特征，利用这些特征来更准确地检测经过隐写的载密图像。相应地，需要研究合适的特征选择方法，以避免特征维数过多带来的过学习问题。

(2)利用新的机器学习或人工智能方法来提高图像隐写的检测性能。过去的图像隐写检测主要使用了支持向量机、卷积神经网络等机器学习方法。近年来，随着深度学习技术在各个领域的广泛应用，机器学习方法得到了较大的发展，在生成对抗网络(GAN)、深度强化学习等方面有许多新成果。利用这些新成果有可能发展出新的图像隐写检测算法，进一步提高图像隐写检测的精确性[69]。

(3)需要开展无载体图像隐写分析的研究。过去的图像隐写分析工作绝大部分都是针对嵌入型图像隐写方法的，无载体图像隐写方法不改变载体图像内容，因此已有的图像隐写分析方法对无载体图像隐写方法基本上是无效的。未来随着无载体图像隐写方法的发展与应用，亟需开展针对无载体图像隐写的检测分析研究。

参 考 文 献

[1] LIU H X, XIA C H. Overview of Steganalytic Algorithm to Digital Images [J]. Computer Engineering and Design, 2006, 27(1): 21-25.

[2] WANG S Z, ZHANG X P, ZHANG W M. Recent Advances in Image Based Steganalysis Research [J]. Chinese Journal of Computers, 2009, 32(7): 1247-1263.

[3] ZHANG J, XIONG F, ZHANG D. Overview on Image Steganalysis Technology [J]. Computer Engineering, 2013, 39(4): 165-168.

[4] DONG J, QIAN Y L, WANG W. Recent Advances in Image Steganalysis [J]. Journal of Image and Signal Processing, 2017, 6(3): 131-138.

[5] KARAMPIDIS K, KAVALLIERATOU E, PAPADOURAKIS G. A Review of Image Steganalysis Techniques for Digital Forensics [J]. Journal of Information Security and Applications, 2018, 40: 217-235.

[6] GUAN Q X, ZHU J, ZHAO X F, et al. Image Steganalysis Based on Linear Programing Feature Selection and Ensemble Classifier [J]. Journal of Cyber Security, 2018, 3(1): 83-94.

[7] KADHIM I J, PREMARATNE P, VIAL P J, et al. Comprehensive Survey of Image Steganography: Techniques, Evaluations, and Trends in Future Research [J]. Neurocomputing, 2019, 335: 299-326.

[8] PROVES N, HONCYMAN P. Hide and Seek: An Introduction to Steganography [J]. IEEE Security & Privacy, 2003, 1(3): 32-44.

[9] PAN F, LI J, YANG X. Image Steganography Method based on PVD and Modulus Function [C]// Proceedings of the 2011 International Conference on Electronics, Communications and Control (ICECC), 2011: 282-284.

[10] KAWAGUCHI K, EASON R O. Principle and Application of BPCS Steganography [C]// Proceedings of SPIE Multimedia Systems and Applications. Boston, 1998: 464-472.

[11] FRIDRICH J, GOLJAN M, DU R. Detecting LSB Steganography in Color and Gray-Scale Images [J]. IEEE Multimedia, 2001, 8(4): 22-28.

[12] Proves N. Defending Against Statistical Steganalysis [C]// 10th USENIX Security Symposium. 2001: 24-25.

[13] WESTFELD A. F5-A Steganographic Algorithm: High Capacity Despite Better Steganalysis[C]// Proceedings of 4th International Information Hiding Workshop. Pittsburgh, 2001: 289-302.

[14] KUMAR V, KUMAR D. A Modified DWT-based Image Steganography Technique [J]. Multimedia Tools and Applications, 2018, 77(11): 13279-13308.

[15] SALLEE P. Model-based Steganography [C]// Proceedings of the International Workshop on Digital Watermarking. LNCS, vol. 2939, Springer, 2003: 154-167.

[16] SHAFEE S, RAJAEI B. A Secure Steganography Algorithm Using Compressive Sensing based on HVS Feature [C]// Proceedings of the 2017 Seventh International Conference on Emerging Security Technology. IEEE, 2017: 74-78.

[17] GIRDHAR A, KUMAR V. Comprehensive Survey of 3D Image Steganography Techniques [J]. IET Image Processing, 2018, 12(1): 1-10.

[18] RABIE T, KAMEL I. High-capacity Steganography: a Global-adaptive-region Discrete Cosine Transform Approach [J]. Multimedia Tools and Applications, 2017, 76(5): 6473-6493.

[19] HONG W. Human Visual System based Data Embedding Method Using Quadtree Partitioning [J]. Signal Processing: Image Communication, 2012, 27(10): 1123-1133.

[20] XIAO J J, LU Q. Adaptive Steganography Algorithm Based on Vision Effect [J]. Journal of Test and Measurement Technology, 2012, 26(1): 9-14.

[21] ZHOU Z, SUN H, HARIT R, et al. Coverless Image Steganography without Embedding [C]// International Conference on Cloud Computing and Security, LNCS, vol. 9483. Springer International Publishing, 2015: 123-132.

[22] WU J, LIU Y, DAI Z, et al. A Coverless Information Hiding Algorithm Based on Grayscale Gradient Co-occurrence Matrix [J]. IETE Technical Review, 2018, 35(sup1): 23-33.

[23] RUAN S H, QIN Z C. Coverless Covert Communication based on GIF Image [J]. Communications Technology, 2017, 50(7): 160-167.

[24] ZHANG X, PENG F, LONG M. Robust Coverless Image Steganography Based on DCT and LDA Topic Classification [J]. IEEE Transactions on Multimedia, 2018, 20(12): 3223-3238.

[25] DUAN X, SONG H, QIN C, et al. Coverless Steganography for Digital Images Based on a Generative Model [J]. Computers, Materials & Continua, 2018, 55(3): 483-493.

[26] WESTFELD A, PFITZMANN A. Attacks on Steganographic Systems [C]// Proc. of International Workshop on Information Hiding (IH'99). Springer-Verlag, LNCS, 1999: 61-76.

[27] KWANGSOO L, JUNG C, LEE S, et al. New Steganalysis Methodology: LR Cube Analysis for the Detection of LSB Steganography [C]// Proc. of International Workshop on Information Hiding (IH'05). Springer-Verlag, LNCS, 2005: 312-326.

[28] ZHANG T, PING X. A New Approach to Reliable Detection of LSB Steganography in Natural Images [J]. Signal Processing, 2003, 83(10): 2085-2093.

[29] ZHANG X P, WANG S Z. Statistical Analysis Against Spatial BPCS Steganography [J]. Journal of Computer Aided Design & Computer Graphics, 2005, 17(7): 1625-1629.

[30] ZIOU D, JAFARI R. Efficient Steganalysis of Images: Learning is Good for Anticipation [J]. Pattern Analysis Applications, 2014, 17(2): 279-289.

[31] FRIDRICH J, GOLJAN M, HOGEA D. Steganalysis of JPEG Images: Breaking the F5 Algorithm [C]// Proc. of International Workshop on Information Hiding (IH'02). LNCS, 2578, 2002: 310-323.

[32] HAN X D, PING X J, ZHANG T. Steganalysis Based on the Differences of Coefficient Combinations of 0, 1 for Detecting F5 Steganography [J]. Journal of Information Engineering University, 2009, 10(2): 184-187.

[33] PROVOS N, HONEYMAN P. Detecting Steganographic Content on the Internet [C]// Proc. of ISOC NDSS'02. 2002: 408-412.

[34] ZHANG T, PIRLG X. A Fast and Effective Steganalytic Technique Against Jsteg-like Algorithms [C]// Proc. of 2003 ACM Symposium on Applied Computing. ACM Press, 2003: 307-311.

[35] LEE K, WESTFELD A, LEE S. Generalized Category Attack - Improving Histogram-based Attack on JPEG LSB Embedding [C]// Proc. of 9th Information Hiding Workshop. Springer, LNCS, 2007: 35-48.

[36] FRIDRICH J, GOLJAN M, HOGEA D. New Methodology for Breaking Steganographic Techniques for JPEGs [C]// Proc. of IS&T/SHE Electronic Imaging: Security and Watermarking of Multimedia Contents V. SPIE, 2003: 143-155.

[37] LYU S, FARID H. Detecting Hidden Messages Using Higher-Order Statistics and Support Vector Machines [C]// Proc. IH'02, Springer-Verlag. LNCS, 2002: 340-354.

[38] FARID H, LYU S. Higher-order Wavelet Statistics and Their Application to Digital Forensics [C]// Computer Vision and

Pattern Recognition Workshop (CVPRW'03). 2003:94-94.

[39] HE J H,LIANG X P,LI J Q,et al. Image Steganalysis Based on Bit-Plane Statistical Correlation Using Support Vector Machine [J]. Acta Scientiarum Naturalium Universitatis Sunyatseni, 2004,43(sup2):17-20.

[40] PEVNY T,BAS P,FRIDRICH J. Steganalysis by Subtractive Pixel Adjacency Matrix [J]. IEEE Transactions on Information Forensics and Security,2010,5(2):215-224.

[41] FRIDRICH J,KODOVSKY J. Rich Models for Steganalysis of Digital Images [J]. IEEE Transactions on Information Forensics and Security,2012,7(3):868-882.

[42] HOLUB V,FRIDRICH J. Random Projections of Residuals for Digital Image Steganalysis[J]. IEEE Transactions on Information Forensics and Security,2013,8(12):1996-2006.

[43] WHITAKER J M,KER A D. Steganalysis of Overlapping Images [J]. Proceedings of SPIE, vol. 9409, Media Watermarking, Security, and Forensics,2015,94090X.

[44] AVCIBAS I,MEMON N,SANKUR B. Steganalysis Using Image Quality Metrics [J]. IEEE Transactions on Image Processing,2003,12(2):221-229.

[45] SHI Y,CHEN C,CHEN W. A Markov Process based Approach to Effective Attacking JPEG Steganography [C]// Proc. of 8th International Workshop on Information Hiding (IH'2006). Springer,LNCS,2007:249-264.

[46] PEVNY T,FRIDRICH J. Merging Markov and DCT Features for Multi-Class JPEG Steganalysis [C]// Proc. of SPIE Electronic Imaging. Photonics West,2007:3-4.

[47] KODOVSKY J,FRIDRICH J. Calibration revisited [C]// Proceedings of the 11th ACM Workshop on Multimedia and Security (MM&Sec'09). New York,ACM,2009:63-74.

[48] GUAN J B. Research and Implementation of Steganalysis for JPEG Images [D]. Guilin:Guilin University of Electronic Technology,2013.

[49] KODOVSKY J,FRIDRICH J. Steganalysis of JPEG Images Using Rich Models [C]// Proc. of SPIE Electronic Imaging, Media Watermarking. Security, and Forensics,2012:1-13.

[50] HULOB V,FRIDRICH J. Low-Complexity Features for JPEG Steganalysis Using Undecimated DCT [J]. IEEE Transactions on Information Forensics and Security,2015,10(2):219-228.

[51] 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'15). New York:ACM,2015:15-23.

[52] WANG C,FENG G. Calibration-based Features for JPEG Steganalysis Using Multi-level Filter [C]// Proc. of IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC 2015). 2015.

[53] SAJEDI H. Adaptive Image Steganalysis [J]. Multimedia Tools and Applications,2018,77(13):17269-17284.

[54] FRIDRICH J. Feature-based Steganalysis for JPEG Images and Its Implications for Future Design of Steganographic Schemes [C]// Proc. of 6th Information Hiding Workshop. Springer, LNCS,2004:67-81.

[55] FU D,SHI Y Q,ZOU D,et al. JPEG Steganalysis Using Empirical Transition Matrix in Block DCT Domain [C]// International Workshop on Multimedia Signal Processing (MMSP'2006). 2006:310-313.

[56] DONG J,WANG W,TAN T N. Multi-Class Blind Steganalysis Based on Image Run-Length Analysis [C]// Proc. of International Workshop on Digital Watermarking (IWDW'09). LNCS, 2009:199-210.

[57] XU M. Steganalysis for JPEG Image Based on SVM [D]. Changsha:Hunan University,2012.

[58] WANG L N,WANG H S,ZHAI L M,et al. A Blind Steganalytic Method to Detect JPEG Image Steganography [J]. Journal of Wuhan University (Nature Science Edition),2018,64(3):217-224.

[59] BABU J,RANGU S,MANOGNA P. A Survey on Different Feature Extraction and Classification Techniques Used in Image Steganalysis [J]. Journal of Information Security,2017,8(3): 186-202.

[60] XU G,WU H,SHI Y. Structural Design of Convolutional Neural Networks for Steganalysis [J]. IEEE Signal Processing Letters,2016,23(5):708-712.

[61] YE J,NI J,YI Y. Deep Learning Hierarchical Representations for Image Steganalysis [J]. IEEE Transactions on Information Forensics and Security,2017,12(11):2545-2557.

[62] WU S,ZHONG S,LIU Y. Deep residual learning for image steganalysis [J]. Multimedia Tools and Applications, 2018, 77(9):10437-10453.

[63] GAO P X,WEI L X,LIU J,et al. Image Steganalysis Based on Deep Residual Neural Network [J]. Computer Engineering and Design,2018,39(10):3045-3049.

[64] QIN B. JPEG Images Steganalysis Research Based on Bayes Decision [D]. Shenyang:Northeastern University of China,2011.

[65] KODOVSKY J,FRIDRICH J,HOLUB V. Ensemble Classifiers for Steganalysis of Digital Media [J]. IEEE Transactions on Information Forensics and Security,2012,7(2):432-444.

[66] MA Y,LUO X,LI X,et al. Selection of Rich Model Steganalysis Features Based on Decision Rough Set α-Positive Region Reduction [J]. IEEE Transactions on Circuits and Systems for Video Technology,2019,29(2):336-350.

[67] ADELI A,BROUMANDNIA A. Image Steganalysis Using Improved Particle Swarm Optimization Based Feature Selection [J]. Applied Intelligence,2018,48(6):1609-1622.

[68] WU M Q,ZHU Z L,JIN S Y. Secret Key Estimation in Sequential Steganography Based on the Laplacian Model [J]. Computer Engineering & Science,2008,30(2):9-14.

[69] CHAUMONT M. Deep Learning in Steganography and Steganalysis from 2015 to 2018 [M]. Draft, Montpellier University, 2019.

PENG Wei, research fellow at department of cyber security, college of computer, national university of defense technology, is a senior member of CCF. His main research interests include techniques of computer networks and network security.