Journal of Image and Signal Processing 图像与信号处理, 2017, 6(3), 131-138
Published Online July 2017 in Hans. http://www.hanspub.org/journal/jisp
https://doi.org/10.12677/jisp.2017.63016

Hans汉斯

Recent Advances in Image Steganalysis

Jing Dong[1,2], Yinlong Qian[1,3*], Wei Wang[1]

[1]Center for Research on Intelligent Perception and Computing, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing
[2]State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing
[3]Department of Automation, University of Science and Technology of China, Hefei Anhui
Email: *ylqian@mail.ustc.edu.cn

Received: Jun. 1st, 2017; accepted: Jun. 18th, 2017; published: Jun. 21st, 2017

Abstract

In recent years, steganalysis has become an important research direction in information security. With rapid development, numerous methods have been proposed to solve the steganalysis problem. This article aims to review recent advances in image steganalysis to provide useful information to the researchers in this field. It first summarizes recent progress in traditional handcrafted feature based methods, and then introduces the deep learning based steganalysis, which is a new trend in steganalysis. Finally, the article summarizes the future trends and challenges in steganalysis.

Keywords

Steganography, Steganalysis, Universal Steganalysis, Pattern Recognition, Deep Learning

图像隐写分析研究新进展

董　晶[1,2]，钱银龙[1,3*]，王　伟[1]

[1]中科院自动化研究所模式识别国家重点实验室智能感知中心，北京
[2]中科院信息工程研究所信息安全国家重点实验室，北京
[3]中国科学技术大学自动化系，安徽 合肥
Email: *ylqian@mail.ustc.edu.cn

收稿日期：2017年6月1日；录用日期：2017年6月18日；发布日期：2017年6月21日

摘　要

隐写分析是信息安全领域一个很重要的研究方向。随着研究的快速发展，已经有大量的隐写分析方法提

*通讯作者。

文章引用: 董晶, 钱银龙, 王伟. 图像隐写分析研究新进展[J]. 图像与信号处理, 2017, 6(3): 131-138.
https://doi.org/10.12677/jisp.2017.63016

出。本文的目的是对近几年图像隐写分析领域的新进展和新思路进行梳理和总结，以给领域内的研究者提供参考。文章首先对传统基于人工特征的隐写分析新进展进行总结，进而重点介绍了隐写领域中新提出的基于深度学习的隐写分析方法。最后，文章对隐写分析的研究面临的挑战以及研究趋势进行了讨论。

关键词

隐写术，隐写分析，通用隐写分析，模式识别，深度学习

1. 引言

随着信息技术的飞速发展以及互联网的广泛普及，数字媒体已经成为军事、商业等机构以及个人获取和传递信息的重要载体。但与此同时，由于互联网中的数字通信容易受到窃听、恶意干扰等活动的威胁，人们比以往任何时候更加关注通信安全问题。传统的加密技术将信息转换成密文实现对信息内容的隐藏。但缺点是由于加密后的密文是无意义的乱码，很容易引起攻击者的注意，从而导致信息被干扰或拦截，甚至破解。在这种背景下，一种新的通信安全观念逐渐形成：通信安全不仅意味着隐藏信息的内容而且要求隐藏信息传输行为的存在性。因此，具有“伪装”特点的更具隐蔽性的数字隐写术(Steganography)得到了越来越多的关注。

数字隐写术的基本原理是利用图像、视频、音频、文本等数字载体中存在的人类感知系统不敏感的冗余信息的特性，将待传递的秘密信息隐藏到该冗余信息中，并借助载体的传输来传递秘密信息。由于嵌入信息后的载密载体表面上和普通载体一样，使得可能的攻击者很难觉察到秘密信息的存在，从而保证了信息安全隐蔽的传输。近年来，随着隐写术的迅速发展，大量的隐写算法被提出，人们无需太多的专业知识便可以方便的从互联网上下载免费的隐写软件来实现信息的隐蔽通信。

然而隐写术是一把“双刃剑”，其在保障人们互联网通信安全的同时，也为怀有恶意企图或不当目的组织或个人提供了便利。事实上，近些年来，关于隐写术被用于间谍、恐怖袭击、违法犯罪等活动的新闻时有报道。非法或恶意使用数字隐写术的行为已经给国家、公司及个人的信息安全与通信安全安全带来了严重危害。在这种情况下，如何有效的对隐写术的使用进行监督，即时防止或阻截隐写术的恶意或非法应用，成为各国军事、安全等部门的迫切需求。正因为如此，隐写分析(Steganalysis)作为隐写术的对抗技术应运而生，并得到了各国政府、科研机构的重视。隐写分析是通过对载体的统计特性进行分析，判断载体中是否隐藏有额外的信息甚至估计信息嵌入量、获取隐藏信息内容的技术。其研究对于防止机密信息泄露、打击恐怖主义与犯罪活动、维护互联网安全等方面都具有十分重要的意义。

上世纪九十年代末期美国的乔治梅森大学的 Neil F. Johnson 最早开始隐写分析的研究。随后达特茅斯学院、麻省理工大学、纽约州立大学、普渡大学、新泽西理工学院、Wet Stone 公司、IBM 公司、Microsoft 公司等机构先后开展了这一方向的研究，并大都得到了美国国防部、国家安全局等政府部门的大力支持。另外英国、德国、法国、俄罗斯、日本、芬兰等国也积极的投入的研究，并有重要成果发表。目前领域内国外的著名研究专家有 Jessica Fridrich、Andrew D. Ker、Yun Q. Shi、Hany Farid、Niels Provos、Andreas Westfeld 等。在国内，中国科学院自动化研究所、中科院信息工程研究所、北京电子技术应用研究所、

清华大学、中山大学、中国科学技术大学、深圳大学、北京邮电大学、大连理工大学、天津大学、上海大学、同济大学、解放军信息工程大学、湖南大学等多家高校和研究所都开展了这方面的研究，并得到了国家“863”、“973”、国家自然科学基金等项目或基金的资助。领域内的国际著名期刊及会议有：IEEE Transactions on Information Forensics and Security (TIFS), Information Hiding and Multimedia Security (IH&MMSec), Media Watermarking, Security, and Forensic (MWSF), International Workshop on Information Forensics and Security (WIFS), International Workshop on Digital Watermarking (IWDW)等。国内的全国信息隐藏暨多媒体信息安全学术大会截止 2016 年 10 月已成功召开了 13 届。

目前的隐写分析研究领域通常将隐写分析看成一个二分类问题，目标是区分正常载体和含密载体。虽然人类感知器官无法进行区分，研究者可以借助机器学习、模式识别等工具，并通过对图像统计特性进行分析来实现有效判别。根据应用范围的不同，隐写分析可以分为两种类型：专用隐写分析(Specific Steganalysis)和通用隐写分析(Universal Steganalysis)。专用隐写分析利用某种隐写术中存在的特定的缺陷或漏洞进行针对性的攻击，检测准确率较高。但其局限性在于只对特定的一种或一类隐写术有效。由于目前的隐写算法千变万化，且新的算法层出不穷，针对性的隐写分析方法很难满足实际应用需求。在这种情况下，不针对特定隐写算法的通用隐写分析技术受到了更多的重视，并成为了目前隐写分析领域主流的研究方向。通用隐写分析的基本思想是利用隐写算法对载体图像统计特性造成的改变中存在的共性来进行检测，可以对多种隐写术进行攻击，更接近实际应用需求。

隐写分析作为信息安全领域内的重要研究方向，虽然仍是一个相对年轻的研究领域，但已有大量的研究论文发表。目前领域内也出现了一些对前期研究工作的进行总结的综述和概括文章。然而由于研究发展较快，近年来不断涌现新的研究成果，除了传统的基于人工特征的方法外，最近出现的基于深度学习的通用隐写分析方法的提出给隐写分析研究注入了新的活力。这种情况下，有必要对最近几年的研究新进展进行及时的归纳和梳理，给研究者提供参考。隐写术的数字载体中以图像的使用最为广泛，目前基于图像的隐写术和隐写分析的研究也更受关注。本文主要讨论以图像为载体的通用隐写分析技术的最新研究动态。第 2 节讨论基于人工特征的通用隐写分析近年来代表性的进展，第 3 节重点讨论新兴的基于深度学习的通用隐写分析技术。第 4 节是总结和展望。

2. 基于人工特征的隐写分析进展

传统的图像通用隐写分析方法一般包括特征提取和训练分类器两个步骤。其中，隐写分析中的特征是对正常图像和载密图像具有区分能力的统计量。而分类器则是机器学习中常用的 SVM、集成分类器(Ensemble Classifier) [1]等可训练优化的类别判别工具。两个步骤中，特征表达是研究中的关键问题，其对检测性能起到决定性的作用。在传统的方法中，特征表达主要依赖于人工设计，且特征设计的基本思想是找到隐写操作前后图像中具有明显差异的统计量。

早期的隐写分析特征通常是统计矩、特征函数质心等相对简单的统计量，且特征维度较低，通常只有几十维。代表性的有 Harmsen 和 Pearlman [2]提出的基于直方图特征函数质心(histogram characteristic function center of mass, HCFCOM)特征，Ker 等[3]引入校准技术后提出的邻接直方图特征函数质心(adjacency histogram characteristic function center of mass, AHCFCOM)特征，Lyu 和 Farid[4]提出的基于小波分解的高阶统计量特征，Shi 等[5]提出的基于图像小波系数直方图各阶统计矩的 CF 特征，Goljan 等[6]提出的基于小波绝对矩(wavelet absolute moment, WAM)特征，Xuan 等[7]提出的小波系数直方图统计矩特征，Chen 等[8]提出的基于经验矩阵的特征，以及 Dong 和 Tan[9]提出的基于游程的特征等。但随着图像隐写术的不断发展，新提出的隐写算法可以保持更复杂的图像统计特性，例如近几年提出的 HUGO [10], SUNIWARD [11], WOW [12], HILL-CMD [13], MiPOD [14]等内容自适应隐写术可以自动的将隐秘信息

嵌入到纹理、噪声丰富的图像区域，从而保持复杂的图像高阶统计特性。隐写术的快速进步使得早期提出的基于简单统计量的特征很难取得有效的检测效果。

为了对抗更先进的自适应隐写术，隐写分析领域特征设计过程中需要考虑更复杂的图像统计特性，并且特征也逐渐朝着复杂化、高维化发展。目前，基于对图像邻域复杂相关性进行建模的高阶统计量特征称为隐写分析领域的主流特征。该类特征提取方法包括如下两步：第一步是预处理，即对相邻的像素或者 DCT 系数用不同种类高通滤波核进行滤波操作得到不同的残差图像；第二步分别从各个残差图像中计算描述邻近像素或系数间相关性的共生矩阵[15] [16] [17]、投影直方图[18]、LBP (Local Binary Pattern) [19]等作为特征，进而通过组合不同残差图像上得到的特征得到最终的高维特征集。其中高通滤波预处理较大程度上抑制了低频图像内容的影响，增强了隐写分析更关注的隐写噪声成分，使得后续的特征表达更有效。而通过组合从不同残差中提取的特征，可以融合不同残差中的互补信息，从而捕捉到更丰富的隐写分析相关的邻域相关性等复杂统计特性，使得检测性能有较大提升。该类特征可统称为“富模型”特征，其中“富”字体现在其通过组合相异的子特征集使得特征可以表达丰富的统计特性。同时该类特征这些特征在维度上相比于早期的特征有了大幅度的提升，例如代表性的 SRM (Spatial Rich Model) [17], PSRM (Projection Spection Rich Model) [18]等特征维度均超过了 10,000 维。

在“富模型”特征研究基础上，近几年基于人工特征的隐写分析研究的新趋势是在“富模型”特征的提取过程中引入选择信道(Selection Channel)先验信息来进一步提升对自适应隐写术的检测能力。其中选择信道指的是自适应隐写术中的用到的修改概率，其用于在秘密信息嵌入过程中选择适合嵌入的像素或频域系数位置，从而保证隐写的安全。

Tang 等[20]首先提出了利用选择信道信息的自适应隐写分析(Adaptive Steganalysis)，且主要针对的是 WOW 隐写算法。其思想是首先从图像中筛选出可疑区域，然后只从可疑区域中提特征。筛选可疑区域的方式是首先用 WOW 算法中的方法计算每个像素的嵌入代价(embedding cost) ρ_{ij}，ρ_{ij} 越小表示该像素被修改的概率越大。通过一定的比例 p 选出 ρ_{ij} 较小的像素集作为可疑区域，其中 p 为可调参数。最终在筛选出的可疑区域上计算残差并提取共生矩阵作为特征。该方法对 WOW 的检测相对传统特征提取方式有了一定的提升。

Denemark 等[21]在 Tang 等工作的基础上提出了一种新的通用的自适应隐写分析特征。该方法所提的特征可看成是传统 Rich Model 的变体。提特征所用到的残差、量化截短等操作均与传统 Rich Model 中的相同。不同点在于共生矩阵的计算。传统 Rich Model 在从残差上计算四维共生矩阵时，同等的看待每个四元组(d_1, d_2, d_3, d_4)的贡献，即每次四元组出现时对应的共生矩阵中的 bin 值加 1，其中 d_1, d_2, d_3, d_4 为残差图像中的元素。而该文的作者认为应当对图像中每个元素对最终特征的贡献有所区分，修改概率大的位置应赋予更大的权重。该方法中，每次四元组出现时，对应的共生矩阵中的 bin 值加四元组元素对应的修改概率 β_{ij} 的最大值，其中 β_{ij} 是根据相应隐写术中的计算方式得到。该方法相比 HUGO、WOW、SUNIWARD 等自适应隐写术的检测效果较传统方法都有所提升。

3. 基于深度学习的隐写分析

虽然基于人工设计特征的通用隐写分析技术近些年取得发展较快，但仍面临诸多困难及挑战。首先，人工设计特征是一种非常费力、启发式的方法，有效特征的设计选取更多的依赖于人的经验，且需要花费大量的时间精力。其次，作为互相对抗的两种技术，在隐写分析发展的同时，隐写术也在不断地进步。新型的隐写术层出不穷，且新提出的隐写术往往可以保持图像中更复杂的统计特性，这也给隐写分析提出了越来越高的要求。为了进行有效的检测，隐写分析特征也需要考虑更复杂的统计特性，特征设计难度不断加大。

深度学习是近些年机器学习学科中的一个新兴的研究领域，该类方法通过构建一个多层的非线性结构单元组成的网络结构模型进行优化训练来挖掘隐含在数据内部的复杂的关系，从而自动的从数据中学到有效的特征表达，同时大大减少对人的经验和精力的需求。自 2006 年 Geoffrey Hinton 等人[22]在 Science 上提出的基于深度信念网络(deep belief network, DBN)的深度学习模型以来，深度学习得到了机器学习以及相关应用领域的极大关注，并迅速成为研究热点。随着深度学习技术的迅速进展，新的深度学习方法不断被提出，代表性的有深度波兹曼机(Deep Boltzmann Machines, DBM) [23]，深度自动编码器(Deep Autoencoder, DAE) [24]，卷积神经网络(Convolutional Neural Networks, CNN) [25]，生成对抗网络(Generative Adversarial Nets) [26]等。并且基于数据驱动的学习模式的深度学习方法已在计算机视觉、语义分析、语音识别、自然语言处理等众多机器学习相关应用领域取得了成功的应用，并颠覆了这些领域基于“人造特征”的传统范式。在这些工作的启发下，已有工作开始关注利用深度学习工具从特征学习的角度解决隐写分析中最为关键的特征表达问题。

Qian 等[27]于 2015 年提出了基于深度学习的隐写分析框架。文章中选取卷积神经网络(Convolutional Neural Network, CNN)这一代表性的深度学习模型，并结合隐写分析的特点，提出了基于 CNN 的适应于隐写分析的特征学习模型。该模型包括一个图像预处理层(Image processing layer)，数个卷积层(Convolutional layer)及数个全连接层，其中图像预处理层用隐写分析中常用的高通滤波器对图像进行预处理以增强隐写噪声，卷积层实现对图像特征的抽取，全连接层则用以对图像进行分类。与传统基于人工设计特征方法不同的是，该方法将特征提取和分类整合到一个网络框架下，并通过反向传播方法自动的优化两个步骤中的参数，从而同时实现特征学习和分类。

继 2015 年 Qian 等人的工作之后，基于深度学习的隐写分析得到了更多地关注，陆续有新的基于深度学习的隐写分析成果发表。其中一部分工作主要集中在网络模型结构的改进上。Pibre 等[28]提出了一种新的基于适应于隐写分析的 CNN 的网络结构模型。该网络模型中，同样包含了高通滤波预处理模块用于增强图像隐写噪声，使得后面的网络模块可以学习到有效的隐写分析特征表达。与 Qian 等人网络中用了 5 层卷积层不同，该文章的网络只用了两层卷积层，同时增加了每层卷积层中 feature map 的数量。即相比 Qian 的网络增加了宽度，减小了深度。同时该网络去掉了 pooling 操作，作者认为 pooling 操作中的缩放操作会起到平滑噪声的效果，不利于隐写分析。但实验表明，作者提出的网络模型取得良好检测性能的前提是限定隐写算法的嵌密钥针对不同图像保持不变。值得注意的是，实际情况中，对于不同的图像，嵌入密钥一般是不同的。该种情形下，作者提出的网络结构模型检测性能有大幅下降。Xu 等[29]提出的基于 CNN 的适应于隐写分析的网络结构模型中也首先用与 Qian 等的工作类似的方式对图像进行高通滤波预处理以增强隐写噪声。不同点在于，作者在预处理之外的网络结构单元的选择上，借鉴了更多的其他应用领域中取得成功的新提出的 CNN 结构单元，具体包括：1) 第一层卷积层后用了 Absolute Activation (ABS)层以利用残差图像的对称性；2) 用了 Batch Normalization；3) 用了 1*1 卷积及 Global Average Pooling；4) 激活函数选取上，部分层用了 tanh 函数。实验结果表明，该网络模型取得了稍优于 Rich Model 的检测结果。

另一部分基于深度学习的隐写分析研究则关注在利用迁移学习(Transfer Learning)、模型融合、正则化手段等方式进一步提升模型表达能力以及泛化性能。文献[30]借鉴迁移学习的思想改善基于 CNN 的隐写分析方法时检测低嵌入率载密图像时 CNN 较难训练的问题。事实上，低嵌入率图像的检测的难点在于嵌入信息量小，从而对图像统计特性的改变相对较小，因此正常图像和载密图像间的差异很小，更难区分。作者用 CNN 模型进行训练时，发现当训练集中的载密类图像(stego)的嵌入率较低时，检测效果较差，甚至 CNN 模型不能收敛。针对该问题，作者提出通过迁移 CNN 高嵌入率载密图像数据集上学到的特征

中的先验信息，增强低嵌入率载密图像数据集上的特征学习，从而提升 CNN 模型对低嵌入率载密图像的检测性能。作者认为尽管高嵌入率和低嵌入率隐写操作在对图像的统计特性会分别造成不同的影响，但是其对图像邻近像素相关性的影响模式具有相似之处。而同时这些模式(pattern)在高嵌入率载密图像中很容易被捕获，进而可以通过迁移这些信息来促进对低嵌入率载密图像的检测。其具体的做法是首先将 CNN 模型在由正常图像(cover)和高嵌入率载密图像(stego)组成的训练集上进行预训练，然后在由正常图像和低嵌入率载密图像组成的训练集上进行进一步微调训练(Fine tuning)，最后用于低嵌入率载密图像检测。实验表明该方案可以有效促进 CNN 模型在低嵌入率图像上的训练，并提升检测性能。文献[31]提出了基于正则化(Regularized)CNN 隐写分析模型，其利用传统手工设计特征(如 SRM，maxSRM 等)中的先验信息对 CNN 模型进行正则化约束，降低 CNN 训练中的过学习问题，从而提高模型的隐写检测性能。作者认为传统隐写分析特征中的有效的全局统计信息中的不太容易被卷积网络结构获取，并提出通过模型正则化的方式将该类信息利用到 CNN 网络训练中，促进 CNN 学习到更有效的隐写分析特征表达。其做法是在 Qian 等 2015 年提出的网络基础上，在目标函数中引入了辅助损失函数作为正则项，其中辅助损失函数为均方差函数，用以计算辅助特征与学习到的特征之间的损失。文献[32]研究通过对不同的基于 CNN 的模型通过 ensemble 方法进行融合，进一步提升 CNN 模型的隐写检测性能。该工作主要从两个方面进行考虑：1) 对于同一待测嵌入算法及嵌入率，如何训练不同的基于 CNN 的模型以用于融合；2) 如何进行融合。对于第一方面，文中作者通过调整 CNN 中每层 pooling 操作时的起始位置，可以得到不同的 pooling 输出，进而得到不同的模型输出。作者认为该种方式可以尽可能恢复在 Pooling 中丢失的有用信息。对于第二点，文中作者考虑了直接对不同 CNN 模型的输出概率进行投票融合，以及提取不同 CNN 最后一层 pooling 层输出，并组合得到特征集，用 Ensemble Classifier 训练两种方式。

4. 问题与展望

近年来隐写分析得到了快速的发展，并不断有新的成果涌现。本文着重讨论了近几年图像通用隐写分析领域发展情况，包括传统基于人工特征的方法的新进展以及基于深度学习在隐写分析方法新思路。今后隐写分析的进一步发展可以从以下几个方面考虑：

1) 特征表达仍将是隐写分析中的研究重点。基于深度学习的方法的提出开辟了隐写分领域的新范式，也将是接下来的研究重心所在。但目前基于深度学习的隐写分析研究仍是在初步阶段，仍有很多问题需要进一步的考虑与解决。首先，目前的基于深度学习的隐写分析框架中仍需要人工设计的高通滤波核进行图像预处理，未来是否能不需要依赖人工预处理，实现端到端学习是一个值得继续研究的问题；其次，目前基于深度学习的隐写分析模型的检测性能仍没有完全超过目前最好的基于人工设计特征的方法，有待进一步提高；最后，目前提出的针对隐写分析的深度学习模型相对较简单，未来能否结合迁移学习，增强学习，对抗生成网络等深度学习领域内的研究热点提出更适合隐写分析的模型也值得关注。

2) 目前的通用隐写分析技术仍需要用到隐写算法、嵌入量等先验信息，例如在构建分类器时仍需要与待检测隐写术对应的载密图像进行训练，并不能真正做到“盲”检测。考虑到新的隐写算法层出不穷，因此研究独立于隐写术的通用隐写分析十分重要。

3) 面向实际应用的隐写分析，特别是社交网络平台大数据环境下的隐写分析研究是一个很值得关注的问题。目前的隐写分析研究大多局限在特定的小规模数据库等实验室条件下。而实际应用环境下隐写分析面临的则是社交网络背景下大数据、异质图像源等复杂的图像数据，实验室条件下取得成功的方法往往很难成功的进行实际应用。目前面向实际应用的隐写分析研究较少，但需求迫切，尚待进一步的研究。

4) 目前的隐写分析研究目标主要停留在检测隐写存在性这一层次上，而更高层次的定位隐写嵌入位

置、提取出隐写信息等目标目前仍没有有效的方法，其也是今后隐写分析需要努力的方向。

基金项目

论文得到北京市自然科学基金(No. 4164102)，国家自然科学基金(No. 61502496, No. 61303262, No. U1536120, No. U1636201)和国家重点研发计划(No. 2016YFB1001003)的资助。

参考文献 (References)

[1] Kodovsky, J., Fridrich, J. and Holub, V. (2012) Ensemble Classifiers for Steganalysis of Digital Media. *IEEE Transactions on Information Forensics and Security*, **7**, 432-444. https://doi.org/10.1109/TIFS.2011.2175919

[2] Harmsen, J.J. and Pearlman, W.A. (2003) Steganalysis of Additive-Noise Modelable Information Hiding. *SPIE Proceedings*, **5020**, 131-142. https://doi.org/10.1117/12.476813

[3] Ker, A.D. (2005) Steganalysis of LSB Matching in Grayscale Images. *IEEE Signal Processing Letters*, **12**, 441-444. https://doi.org/10.1109/LSP.2005.847889

[4] Lyu, S. and Farid, H. (2002) Detecting Hidden Messages Using Higher-Order Statistics and Support Vector Machines. Springer, Berlin.

[5] Shi, Y.Q., *et al*. (2005) Effective Steganalysis Based on Statistical Moments of Wavelet Characteristic Function. *International Conference on Information Technology: Coding and Computing*, **1**, 768-773. https://doi.org/10.1109/itcc.2005.138

[6] Fridrich, J., Goljan, M. and Holotyak, T. (2006) New Blind Steganalysis and Its Implications. *Proc Spie*, **6072**, 1-13.

[7] Xuan, G.R., *et al*. (2005) Steganalysis Based on Multiple Features Formed by Statistical Moments of Wavelet Characteristic Functions. Springer, Berlin.

[8] Chen, X., *et al*. (2006) Blind Image Steganalysis Based on Statistical Analysis of Empirical Matrix. *International Conference on Pattern Recognition*, **3**, 1107-1110.

[9] Dong, J. and Tieniu, T. (2008) Blind Image Steganalysis Based on Run-Length Histogram Analysis. *IEEE International Conference on Image Processing*, **2008**, 2064-2067. https://doi.org/10.1109/icip.2008.4712192

[10] Pevn`y, T., Filler, T. and Bas, P. (2010) Using High-Dimensional Image Models to Perform Highly Undetectable Steganography. Springer, Berlin.

[11] Holub, V. and Fridrich, J. (2012) Designing Steganographic Distortion Using Directional Filters. *IEEE International Workshop on Information Forensics and Security* (*WIFS*), **2**, 234-239. https://doi.org/10.1109/WIFS.2012.6412655

[12] Holub, V. and Fridrich, J. (2013) Digital Image Steganography Using Universal Distortion. *Proceedings of the First ACM Workshop on Information Hiding and Multimedia Security*, Montpellier, 17-19 June 2013, 59-68. https://doi.org/10.1145/2482513.2482514

[13] Sedighi, V., Cogranne, R. and Fridrich, J. (2016) Content-Adaptive Steganography by Minimizing Statistical Detectability. *IEEE Transactions on Information Forensics and Security*, **11**, 221-234. https://doi.org/10.1109/TIFS.2015.2486744

[14] Li, B., Wang, M., Li, X., Tan, S. and Huang, J. (2015) A Strategy of Clustering Modification Directions in Spatial Image Steganography. *IEEE Transactions on Information Forensics and Security*, **10**, 1905-1917. https://doi.org/10.1109/TIFS.2015.2434600

[15] Fridrich, J., *et al*. (2011) Breaking HUGO—The Process Discovery. *Information Hiding-international Conference*, **6958**, 85-101. https://doi.org/10.1007/978-3-642-24178-9_7

[16] Jan, K. and Fridrich, J. (2011) Steganalysis in High Dimensions: Fusing Classifiers Built on Random Subspaces. *Proceedings of SPIE—The International Society for Optical Engineering*, **7880**, 181-197.

[17] Fridrich, J. and Kodovsky, J. (2012) Rich Models for Steganalysis of Digital Images. *IEEE Transactions on Information Forensics and Security*, **7**, 868-882.

[18] Holub, V. and Fridrich, J. (2013) Random Projections of Residuals for Digital Image Steganalysis. *IEEE Transactions on Information Forensics and Security*, **8**, 1996-2006. https://doi.org/10.1109/TIFS.2013.2286682

[19] Shi, Y.Q., Sutthiwan, P. and Chen, L. (2012) Textural Features for Steganalysis. *International Workshop on Information Hiding*, **7692**, 63-77.

[20] Tang, W., Li, H., Luo, W. and Huang, J. (2014) Adaptive Steganalysis against WOW Embedding Algorithm. In: Uhl, A. Katzenbeisser, S. Kwitt, R. and Piva, A., Eds., 2*nd ACM IH&MMSec*, ACM Workshop on Information Hiding & Multimedia Security, Salzburg, 91-96. https://doi.org/10.1145/2600918.2600935

[21] Denemark, T., Sedighi, V., Holub, V., Cogranne, R. and Fridrich, J. (2014) Selection-Channel-Aware Rich Model for Steganalysis of Digital Images. *IEEE International Workshop on Information Forensics and Security*, **2015**, 48-53. https://doi.org/10.1109/wifs.2014.7084302

[22] Hinton, G.E. and Salakhutdinov, R.R. (2006) Reducing the Dimensionality of Data with Neural Networks. *Science*, **313**, 504-507. https://doi.org/10.1126/science.1127647

[23] Salakhutdinov, R. and Hinton, G. (2009) Deep Boltzmann Machines. *Journal of Machine Learning Research*, **5**, 1967-2006.

[24] Larochelle, H., Bengio, Y., Louradour, J. and Lamblin, P. (2009) Exploring Strategies for Training Deep Neural Networks. *The Journal of Machine Learning Research*, **10**, 1-40.

[25] Cun, Y.L., Bottou, L., Bengio, Y. and Haffner, P. (1998) Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE*, **86**, 2278-2324. https://doi.org/10.1109/5.726791

[26] Goodfellow, I.J., *et al*. (2014) Generative Adversarial Nets. *International Conference on Neural Information Processing Systems*, **3**, 2672-2680.

[27] Qian, Y., Dong, J., Wang, W. and Tan, T. (2015) Deep Learning for Steganalysis via Convolutional Neural Networks. *IS&T/SPIE Electronic Imaging*, **9409**, 94090J-94090J-10.

[28] Pibre, L., Jérôme, P., Ienco, D. and Chaumont, M. (2016) Deep Learning is a Good Steganalysis Tool when Embedding Key is Reused for Different Images, even if there is a Cover Source-Mismatch. *Media Watermarking, Security, & Forensics*, **4**, 79-95.

[29] Xu, G., Wu, H.Z. and Shi, Y.Q. (2016) Structural Design of Convolutional Neural Networks for Steganalysis. *IEEE Signal Processing Letters*, **23**, 708-712. https://doi.org/10.1109/LSP.2016.2548421

[30] Qian, Y., Dong, J., Wang, W. and Tan, T. (2016) Learning and Transferring Representations for Image Steganalysis Using Convolutional Neural Network. *2016 IEEE International Conference on Image Processing* (*ICIP*), **2016**, 2752-2756. https://doi.org/10.1109/ICIP.2016.7532860

[31] Qian, Y.L., Dong, J., Wang, W. and Tan, T.N. (2015) Learning Representations for Steganalysis from Regularized CNN Model with Auxiliary Tasks. *International Conference on Communications, Signal Processing, and Systems* (*CSPS*), Chengdu, 23-24 October 2015, 629-637.

[32] Xu, G.S., Wu, H.Z. and Yun, Q.S. (2016) Ensemble of CNNs for Steganalysis: an Empirical Study. *Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security*, **2016**, 103-107. https://doi.org/10.1145/2909827.2930798