Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2024.055735

REVIEW

Digital Image Steganographer Identification: A Comprehensive Survey

Qianqian Zhang[1,2,3], **Yi Zhang**[1,2], **Yuanyuan Ma**[3], **Yanmei Liu**[1,2] **and Xiangyang Luo**[1,2,*]

[1]Information Engineering University, Zhengzhou, 450001, China

[2]Key Laboratory of Cyberspace Situation Awareness of Henan Province, Zhengzhou, 450001, China

[3]College of Computer and Information Engineering, Henan Normal University, Xinxiang, 453007, China

*Corresponding Author: Xiangyang Luo. Email: luoxy_ieu@sina.com

Received: 05 July 2024 Accepted: 05 September 2024 Published: 15 October 2024

ABSTRACT

The rapid development of the internet and digital media has provided convenience while also posing a potential risk of steganography abuse. Identifying steganographer is essential in tracing secret information origins and preventing illicit covert communication online. Accurately discerning a steganographer from many normal users is challenging due to various factors, such as the complexity in obtaining the steganography algorithm, extracting highly separability features, and modeling the cover data. After extensive exploration, several methods have been proposed for steganographer identification. This paper presents a survey of existing studies. Firstly, we provide a concise introduction to the research background and outline the issue of steganographer identification. Secondly, we present fundamental concepts and techniques that establish a general framework for identifying steganographers. Within this framework, state-of-the-art methods are summarized from five key aspects: data acquisition, feature extraction, feature optimization, identification paradigm, and performance evaluation. Furthermore, theoretical and experimental analyses examine the advantages and limitations of these existing methods. Finally, the survey highlights outstanding issues in image steganographer identification that deserve further research.

KEYWORDS

Information hiding; steganalysis; steganographer identification; steganography; covert communication; survey

1 Introduction

Steganography is the techniques of hiding secret information within various forms of digital media, such as images [1], audio [2], video [3], and text files [4], with the aim of achieving covert communication. Although it contributes to secure communications [5], there is still a risk that steganography can be illegal abuse [6–8]. For example, in July 2019, a study by the cybersecurity company, Security Bull, found that nearly 8% of office workers in the UK had used online tools such as steganography or encryption to steal company information. In January 2016, the Russian antivirus company, Doctor Web, revealed that more than 60 games on Google Play could download and execute malicious code concealed within images, which allowed them to stealing user information. The illegal abuse of steganography poses a serious threat to the cyberspace security.

The art of steganalysis is an effective way to antagonize steganography. It detects steganography [9–11], disrupts secret communication [12,13], and extracts secret information [14]. In practice, steganalysis is considerably more challenging than steganography [15]. This is primarily due to the abundance of digital covers and the wide range of embedding methods that can be utilized in steganography, making the detection of secret information within this vast sea of digital media akin to locating a needle in a haystack. Especially with the rapid development of deep learning [16–18], steganography is constantly evolving [19–21], yet corresponding steganalysis methods are noticeably lagging behind. Fig. 1 presents insights into the number of publications about steganography and steganalysis over the last nine years that have been surveyed. As can be seen in the figure, the number of research publications about steganography has been increasing since 2018, while the number of steganalysis has been decreasing. Furthermore, the ratio of steganography to steganalysis still has an elevated trend. Therefore, the development of steganalysis technology plays a vital role in reducing the harm caused by the abuse of illegal steganography and ensuring the security of the state, society and individuals.

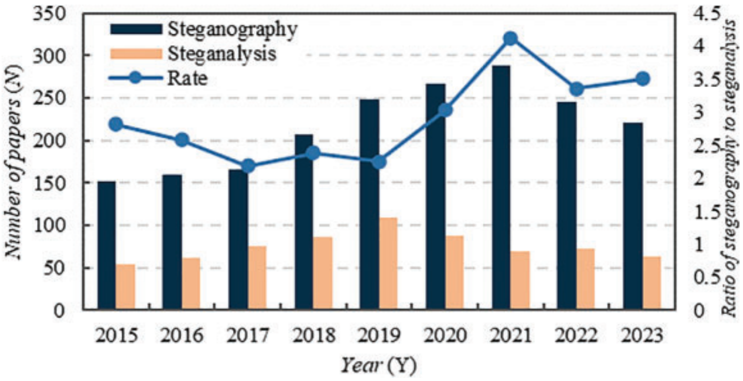

Figure 1: Trend of the number of publications on steganography and steganalysis in the late nine years

The existing digital image steganalysis includes passive steganalysis [22], such as secret information detection [23], secret information extraction [24] and steganographer identification [25], and active steganalysis [26], such as secret information destroy [27,28]. Most of the existing research on steganalysis focuses on secret information detection. Its research methods include traditional supervised machine learning methods [29–31] and deep learning methods such as XuNet [32], YeNet [33], and SRNet [34]. The development of steganalysis can provide reliable support for locating and extracting secret information. However, these methods only give a probabilistic decision on whether an image is a stego, making it difficult to accurately identify and track covert communication behaviors or activities.

In contrast, steganographer identification [35] is an effective passive steganalysis technique that to identify suspected image steganographers and discover illegal covert communication activities [35]. The problem was first posed in 2006. In 2013, Ker et al. [36] listed this problem as one of the "moving steganography and steganalysis from the laboratory into the real world". In practice, steganographer identification technology can detect secret images and discover covert communication source. However, it is still in its early stages. While existing methods for detecting secret information in classical steganography have matured, research methods in this area provide limited references for steganographer identification. On one hand, it is often challenging to obtain information about the specific steganography techniques and payloads used by individuals since each user may employ different image acquisition devices and processing methods. This poses a significant challenge in terms of matching training and testing sets [37]. On the other hand, steganographers may also include cover images along with stego images to deceive steganalysis. In such cases, if information cannot be extracted from a single image alone, the detection result may not provide valuable insights. Therefore, there is an imperative need to develop robust methods for identifying steganographers in complex application scenarios.

With the development of steganalysis technology in the past decade, several surveys have been published [38,39]. For example, Ruan et al. [40] covered the application of deep learning to steganalysis. In 2022, Muralidharan et al. [41] published the survey research "Infinite Competition between Image Steganography and Steganalysis", which considered the interdependencies between steganography and steganalysis from new observations and insights. However, these still focus on steganalysis to detect secret information in images. Although these surveys provide references for steganographer identification, we must systematically sort out existing methods and point out problems in existing research and possible future research. Therefore, this manuscript describes the research in this field, which has served to steganalysis development. By summarizing the steganographer identification methods, the paper helps readers better analyze the current research progress in the field and better understand the current shortcomings and future research direction. The contributions of this paper can be highlighted as:

1. We conduct a comprehensive survey on digital image steganographer identification. Our study aims to systematically review, classify, and compare the performance of the existing methods for steganographer identification.
2. We present a general framework that summarizes the methods proposed in previous literature and evaluate their advantages and limitations based on performance comparisons using reported experimental results from the same datasets.
3. We review studies on various aspects of steganographer identification including data acquisition, feature optimization, identification paradigm, and performance evaluation. Furthermore, we discuss research problems that require further investigation based on our analysis.

The rest of this paper is organized as follows. Section 2 presents several primary concepts and techniques for steganographer identification. Section 3 proposes a general framework for steganographer identification and details the main steps. Section 4 compares the performance of the typical methods theoretically and experimentally. Section 5 discusses the critical research challenges that may need further attention based on the limitations of existing methods, and the last section concludes our work.

2 Primary Concepts and Techniques

2.1 Batch Steganography and Pooled Steganalysis

The classic definition of steganography involves a steganographer aiming to communicate with a passive conspirator over an insecure channel, and an eavesdropper (or Warden) monitoring the channel. The Warden's aim is not to decode the hidden information but merely to deduce its presence. This is steganalysis for the single cover object, i.e., it assumes that each cover object is treated in isolation by both the steganographer and the eavesdropper. However, in practical applications involving considerably large payload, the steganographer should adopt a batch strategy to allocate payload to multiple covers properly.

Batch steganography aims to embed a large number of messages into multiple cover objects while maintaining a satisfactory level of undetectability. Reference [42] posed precisely the problem of batch steganography. In [43], batch schemes for content-adaptive steganography were first mathematically formulated. In specific, when allocating P bits payload on a series of images $x^{(b)}$, $b \in \{1, \ldots, B\}$, the steganographer tries to find an optimal payload-allocation which can minimize the total detectability:

$$\begin{aligned} &\mathbb{R}^{*} = \underset{\mathbb{R}}{argmin} \sum_{b=1}^{B} \mu^{(b)} \left(\mathbb{R}^{(b)}\right)^{2} \\ &s.t. \quad P = \sum_{b=1}^{B} \mathbb{R}^{(b)} \end{aligned} \tag{1}$$

where $R^{(b)}$ denotes the desired payload length allocated to b-th cover image $x^{(b)}$, and $\mu^{(b)}\left(\mathbb{R}^{(b)}\right)$ is the expectation of steganalyzer's detection output for b-th stego image.

Pooled steganalysis, i.e., steganographer identification is a confrontation of batch steganography [44]. In batch steganography, the Warden's task is pooled steganalysis. Reference [43] formulated pooled steganalysis mathematically. Denoting the i-th image with $x^{(i)} = \left(x_{kl}^{(i)}\right)$ and its representation in the feature space as $\boldsymbol{z}^{(i)} = \mathbb{R}^{d}$, the steganographers generate a source of I images $x^{(i)}$, $i = 1, \ldots, I$ that are either all cover or all stego embedded with payloads R_i. The number of images is assumed to be arbitraily large. The Warden inspects a set of B images $x^{(i)}$, $i = 1, \ldots, B$ with a classifier trained with a high-dimensional feature set. Due to the way the features are built and the fact that the test statistic is a projection of high-dimensional features, the Warden's single-image detector output, denoted $\theta^{(i)} = \theta\left(\boldsymbol{z}^{(i)}\right)$, is a sample from a Gaussian distribution $\mathbb{N}\left(0, \sigma^{2}\right)$.

Given $B \geq 1$ images $x^{(i)}$, $i = 1, \ldots, B$, in Warden's pooling bag, the Warden faces the following hypothes is testing problem:

$$\begin{aligned} &H_0: \theta^{(i)} \sim \mathbb{N}\left(0, \sigma^{2}\right), \forall i \\ &H_1: \theta^{(i)} \sim \mathbb{N}\left(\mu_i\left(\mathbb{R}_i\right), \sigma^{2}\right), \forall i \end{aligned} \tag{2}$$

where $\mu_i\left(R_i\right)$ is the expected shift of the detection statistic (over messages) when embedding payload size R_i in $x^{(i)}$. In addition, if it does not impose any assumption on the steganographers' payload spreading strategy, R_i can be different for each image.

2.2 Steganographer Identification

Steganographers often use batch images as covers for covert communication. A steganalyzer can analyze a user's image collection to determine their status as a steganographer by identifying at least one stego image in the collection. Therefore, the problem of identifying steganographers can be described as follows:

Steganographer identification problem formalization. A training set of n users is defined as $U = \{x_i\}_{i=1}^{n}$, where x_i represents a user containing m images. y represents the label corresponding to the user, 0 is the normal user, and 1 is the steganographer. If all images of the user x_i are cover images without secret information, he is a normal user, and his label $y_i = 0$. If the user image set contains a stego image with secret information embedding, he is a steganographer with the label $y_i = 1$. In the test phase, the prediction label $y' \in \{0, 1\}$ is output for user x', which contains multiple images.

2.3 Identification Paradigm

2.3.1 Hierarchical Clustering

Clustering is gathering the more similar and less different samples according to the distance. The goal is to bring similar samples together and different samples separately. Hierarchical clustering is a clustering algorithm that creates a hierarchical nested clustering tree by calculating the similarity between data points of different categories. We introduce several distance measures below:

Let $d(x,\ y)$ for the distance between two objects x and y, and $D(x,\ y)$ for the distance between two clusters X and Y. The single linkage uses the distance between the nearest points in the two clusters:

$$D_{SL} = \min_{x\in X,\ y\in Y} d(x,\ y) \tag{3}$$

and the complete linkage uses the furthest points:

$$D_{CL} = \max_{x\in X,\ y\in Y} d(x,\ y) \tag{4}$$

The single linkage can cause long chains of clusters, whereas complete linkage prefers compact clusters; other agglomerative clustering algorithms are intermediate, including centroid clustering

$$D_{CEN} = \frac{1}{|X|\cdot|Y|}\sum_{x\in X}\sum_{y\in Y} d(x,\ y) \tag{5}$$

and average linkage

$$D_{AL} = \frac{1}{|X\cup Y|^2 - |x\cup Y|}\sum_{x,y\in X\cup Y,x\neq y} d(x,\ y) \tag{6}$$

The input to the basic agglomerative clustering is a distance matrix between objects. The output can be displayed in a dendrogram, a tree of the successive cluster agglomerations using "height" to indicate the distance between clusters being merged.

2.3.2 Local Outlier Factor

The local outlier factor (LOF) [45] is also a distance-based anomaly detection algorithm that can quantify the local deviation of each sample point. Whether the sample point p is abnormal depends not on the local density of p but on comparing the local density of p and its neighbors. Specifically, the reachability distance of p w.r.t. o is defined as $reach\_disk_k(p,\ o) = \max\{d_k(o),\ d(p,\ o)\}$, where $d_k(o)$ is k-th distance of sample point p, $d(p,\ o)$ is a distance between p and o. The local reachability density of p is

$$lrd(p) = \frac{|N_k(p)|}{\sum_{o\in N_k(p)} reach\_disk_k(p,\ o)} \tag{7}$$

where $|N_k(p)|$ is the number of elements in the k-th distance. Thus, the LOF value of p is

$$LOF_k(p) = \frac{1}{|N_k(p)|}\sum_{o\in N_k(p)} \frac{lrd_k(o)}{lrd_k(p)} \tag{8}$$

LOF can capture the degree of sample point p, and the value provided by LOF are interpretable.

2.3.3 Graph Convolutional Network

The Graph Convolutional Network (GCN) [46] proposes a method to extract features from graph data by taking the feature matrix and adjacency matrix as input. The feature matrix contains the initial nodes in the graph, while the adjacency matrix represents their connection relationship. Aggregation operation and node feature updating are crucial in GCN, where low-dimension embeddings for each node are learned by convolving and aggregating information from its neighbors. During each layer's aggregation process, nodes gather information from their previous layer's neighborhood and update their own specific features accordingly.

In GCN, the network transmission from l-th layer to $(l + 1)$-th layer is

$$\boldsymbol{H}^{(l+1)} = f\left(\boldsymbol{H}^{(l)},\ \boldsymbol{A}|\boldsymbol{W}^{(l)}\right) \tag{9}$$

where $\boldsymbol{H}^{(l)}$ is the output of l-th layer, i.e., is the input of $(l + 1)$-th layer. $\boldsymbol{A}$ is adjacency matrix. $\boldsymbol{H}^{(l+1)}$ is the output of $(l + 1)$-th layer. $\boldsymbol{W}^{(l)}$ is the weight matrix that needs to be trained for the feature transformation. In particular, $\boldsymbol{H}^{(0)}$ is the 0-th layer, i.e., is the initial feature matrix.

3 Seganographer Identification Framework

The key to identifying steganographers lies in the distinguishable features of users. We categorize steganographer identification into two main methods: main channel-based and side channel-based, which are based on existing techniques for user feature design. The former distinguishes users by designing image features related to steganography embedding [47,48] in images, while the latter considers user behavioral features [49,50] along with image embedding. Based on this categorization, we propose a general framework for steganographer identification by summarizing the methods presented in existing literature depicted in Fig. 2. It consists of five main steps: data acquisition, feature extraction, feature optimization, identification paradigm, and performance evaluation. The details are as follows:

- Dataset acquisition. Collect batch image data from multiple users, each with varying quantities and types of images.
- Feature extraction. The key steps for accurately identify steganographers include image feature and side channel feature.
- Feature optimization. The important means to improve steganographer identification accuracy include feature calibration, ensemble learning, and dimension reduction for optimized features.
- Identification paradigm. Train a steganalyst to detect or identify steganographers using various learning approaches, including supervised, weakly supervised, unsupervised learning, and other identification paradigms.

3.1 Data Acquisition

In image steganographer identification, each user has a set of images. The inputs are acquired data, including user image data, user behaviour data, and analysis of user attributes, as shown in Fig. 2. As mentioned, the main channel steganographer identification uses image features that can detect steganography embedding. Side channel steganographer identification is to identify steganographers according to user behaviour data. Of course, this all involves the analysis of user attributes.

It is common to simulate user data using image databases commonly used in steganalysis, such as BOSSBase-1.01 [51] and BOWs [52], to evaluate models. In addition, performance evaluation based on real-world datasets is more convincing. Therefore, the acquisition of the real-world datasets is essential for model implementation and performance evaluation. Table 1 lists the real-world user dataset used in existing literature.

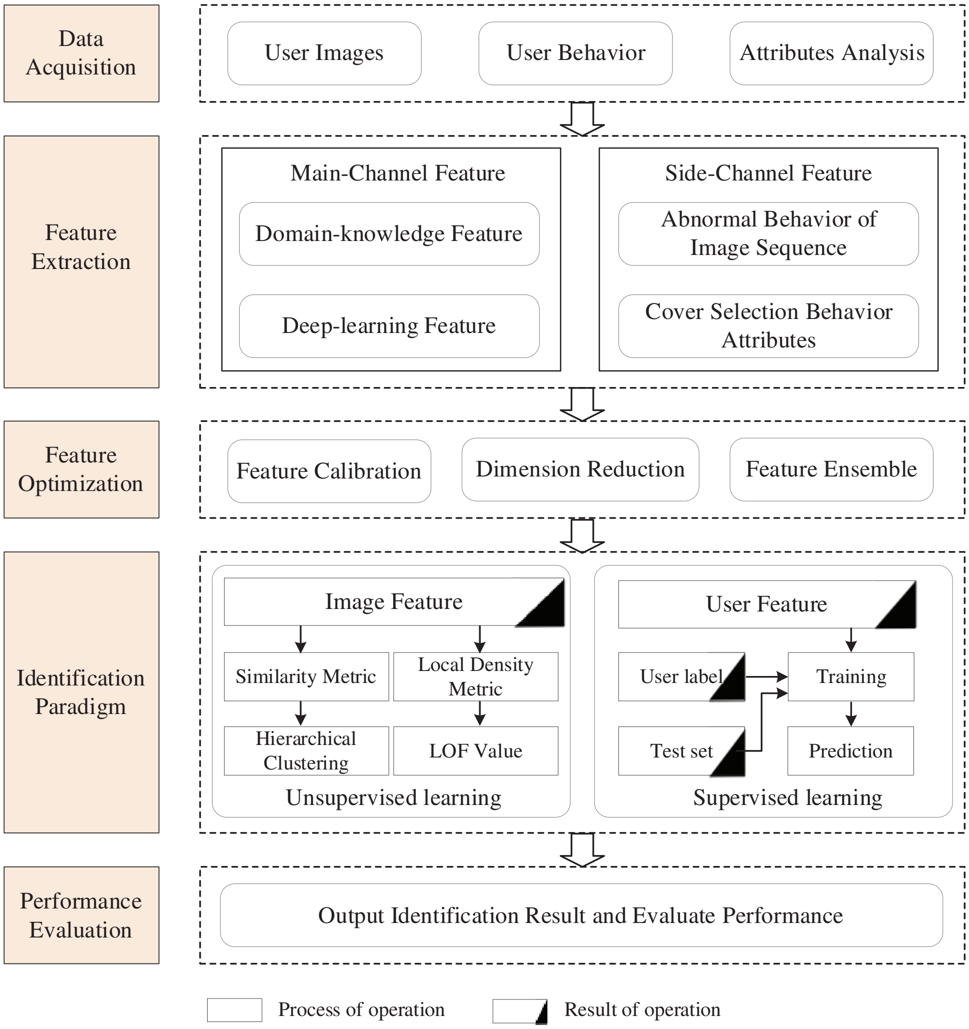

Figure 2: General framework of main channel steganographer identification

Table 1: Data source for steganographer identification

Literature	Data source
Ker et al. [35,53]	Oxford University
Li et al. [54]	Baidu
Li et al. [55,56]	Twitter
Zhang et al. [57,58]	Flickr
Wang et al. [49]	MNIST, CIFAR-10

Ker et al. downloaded over 4 million publicly available JPEG images from real social media sites [35,53]. The crawl was restricted to users who publicly identified themselves with the Oxford University network and the data was then anonymized. Following filtering, they selected a subset of images contributed by 4000 users, with each user providing 200 images. Consequently, they generated an experimental dataset comprising a total of 800,000 images.

Li et al. [54] collected social network image data from an image sharing website (https://image.baidu.com, accessed on 31 August 2024). They downloaded a large quantity of JPEG images from the website and resized them to a size of 1024 × 1024. They recompressed the images using quality of the same factor (QF = 80) to avoid the influence of different quantization matrices on the steganalysis features.

Li et al. [55,56] used Tweepy [59], a public API for Twitter developers, to crawl images from 3000 users on Twitter. And only JPEG images are left. Next, they cropped them to a size of 512 × 512 using center cropping and filtered out users with less than 100 images. Following preprocessing, there were 700 images remaining.

Zhang et al. [57,58] collected images from Flickr (https://www.flickr.com, accessed on 31 August 2024), which is a well-known online social media platform designed for photo management and sharing. They accessed public raw JPEG images by utilizing Flickr's public API, downloading over 400,000 images from 1000 users, with each user contributing between 100 and 600 images.

Wang et al. [49] proposed a method to identify abnormal users who use adversarial attacks among many normal users. The idea and method are similar to steganographer identification except for the identification object. Therefore, the real-world datasets they use also have reference significant for steganographer identification. They used the well-known MNIST [60] and CIFAR-10 [61] image datasets.

3.2 Feature Extraction Methods

The goal of feature extraction is to design features that can distinguish users with effect. Therefore, the separability feature is critical to identify the steganographers accurately. In this section, we introduce the feature extraction method for steganographer identification, and summarize these methods into two categories: main channel features and side channel features, as shown in Fig. 2. The details are as follows.

3.2.1 Main Channel-Based Feature

Image features that distinguish users belong to the main channel features. Both image steganography detection and steganographer identification need image statistical features to make decisions. The former uses the features to detect images, while the latter measures users based on these features. Additionally, the dimension of the features has a greater impact on steganographer identification compared to image steganalysis. Therefore, although there are many relatively mature steganalysis features [62–64] for image classification, they are not suitable for steganographer identification.

We summarize existing image features for steganographer identification. They are categorized into two classes: domain knowledge-based and deep learningbased features. Fig. 3 shows the framework of main channel steganographer identification, where image steganalysis feature extraction is divided into domain knowledge-based and deep learning-based methods. Steganalysis features are often designed on residual images because residual helps suppress image content and amplify steganography noise.

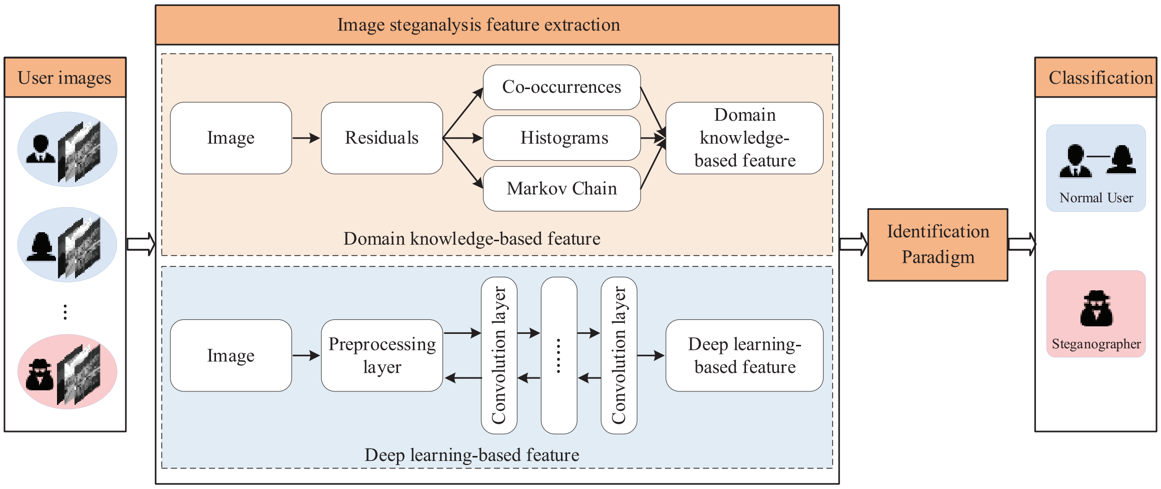

Figure 3: General framework of main channel steganographer identification

A. Domain Knowledge-Based Feature

In 2011, Ker et al. [65] initially proposed clustering-based steganographer identification paradigm using PEV-274 features. The method is more robust because the clustering does not need to train. The PEV-274 [66] was proposed initially for detecting JPEG image steganography. In 2012, Ker et al. [35] proposed a LOF-based steganographer identification method that also utilized this feature. In [54], Li et al. designed a higher-order joint feature with lower dimension and proposed a steganographer identification paradigm based on ensemble clustering [67] using this feature. The higher-order joint feature can effectively capture the modification of the DCT coefficient by JPEG steganography embedding. Due to the development of adaptive steganography [68,69], Li et al. [70] proposed a steganographer identification method using the reduced PEV feature, RPEV-155. This feature is mainly used to identify users who employ adaptive steganography. They sampled and reconstructed the image before extracting the feature, which captured the noise of adaptive steganography to a greater extent.

B. Deep Learning-Based Feature

In deep learning-based steganalysis, multiple convolutional layers extract features, followed by a fully connected layer for classification. Therefore, deep learning-based steganographer identification usually extracts features the network learns after multiple convolutional layers in the deep learning network, as shown in Fig. 3. It should be noted that the performance of these features is closely related to the design and training of the network.

In 2017, Zheng et al. [71] proposed a residual network steganographer identification method, RNSD, which applies deep learning features to steganographer identification for the first time. The backbone is deep residual steganalysis network [72]. To tackle the algorithm mismatch [73,74] in steganographer identification, Zheng et al. [75] proposed a multiclass deep neural network steganographer identification method, MDNNSD. In 2019, inspired by selecting channel-aware steganalysis methods [76], Zheng et al. proposed two embedded probability estimation deep network steganographer detection methods, EPEDN [77] and MEPESD [25], that learn more knowledge about steganography embedding. The difference lies in the embedded probability learning sub-network module. In EPEDN, FCN-8s [78], a fully convolutional network commonly used in image segmentation, is used as this sub-network module. MEPESD adopted the NLDF [79], which is commonly used in saliency detection as the sub-network module. Because labels are not always available in the real world, [80] proposed a Deep Clustering Network for steganographer identification (DCNSD).

C. Discussion

The main channel steganographer identification features are classified into two categories based on feature extraction method: domain knowledge-based features and deep learning-based features.

Through the research status of image feature extraction methods in steganographer identification, this paper analyzes and classifies some existing feature extraction methods, so as to help readers better evaluate feature extraction methods immediately. Fig. 4 lists the current features of main channel-based steganographer identification research and their advantages and limitations. According to previous studies, no validity feature extraction method can be applied to all data sets, and no features is always better than other features. No matter how well a feature extraction method is designed, there will always be application scenarios that are not applicable. For example, the network we train to detect steganographers using J-UNIWARD may not be suitable for identifying users using other steganography, such as UED.

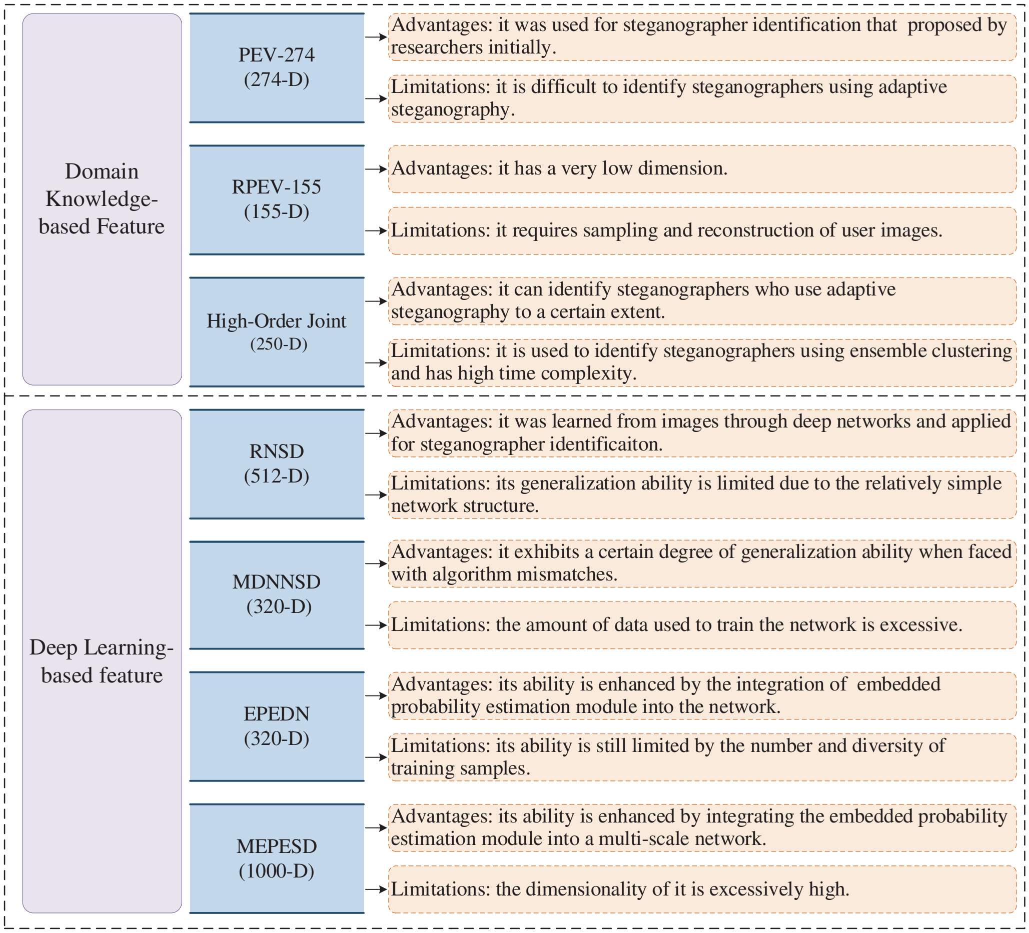
Domain Knowledge-based Feature
PEV-274 (274-D)
Advantages: it was used for steganographer identification that proposed by researchers initially.
Limitations: it is difficult to identify steganographers using adaptive steganography.
RPEV-155 (155-D)
Advantages: it has a very low dimension.
Limitations: it requires sampling and reconstruction of user images.
High-Order Joint (250-D)
Advantages: it can identify steganographers who use adaptive steganography to a certain extent.
Limitations: it is used to identify steganographers using ensemble clustering and has high time complexity.
Deep Learning-based feature
RNSD (512-D)
Advantages: it was learned from images through deep networks and applied for steganographer identificaiton.
Limitations: its generalization ability is limited due to the relatively simple network structure.
MDNNSD (320-D)
Advantages: it exhibits a certain degree of generalization ability when faced with algorithm mismatches.
Limitations: the amount of data used to train the network is excessive.
EPEDN (320-D)
Advantages: its ability is enhanced by the integration of embedded probability estimation module into the network.
Limitations: its ability is still limited by the number and diversity of training samples.
MEPESD (1000-D)
Advantages: its ability is enhanced by integrating the embedded probability estimation module into a multi-scale network.
Limitations: the dimensionality of it is excessively high.

Figure 4: Advantages and limitations of main channel features

3.2.2 Side Channel-Based Feature

In this section, we introduce steganographer identification methods that use behavioral features in the existing literature. We categorize such methods as side channel-based steganographer identification. Fig. 5 shows the framework of side channel steganographer identification. The key to side channel steganographer identification is to find behavioural features that effectively distinguish steganographers from normal users. This involves the analysis of the behavioural attributes of the steganographer. Compared with main channel steganographer identification, there is little research on side channel steganographer identification based on user behaviour characteristics analysis. Existing research includes image sequence and cover selection behavior, described as follows.

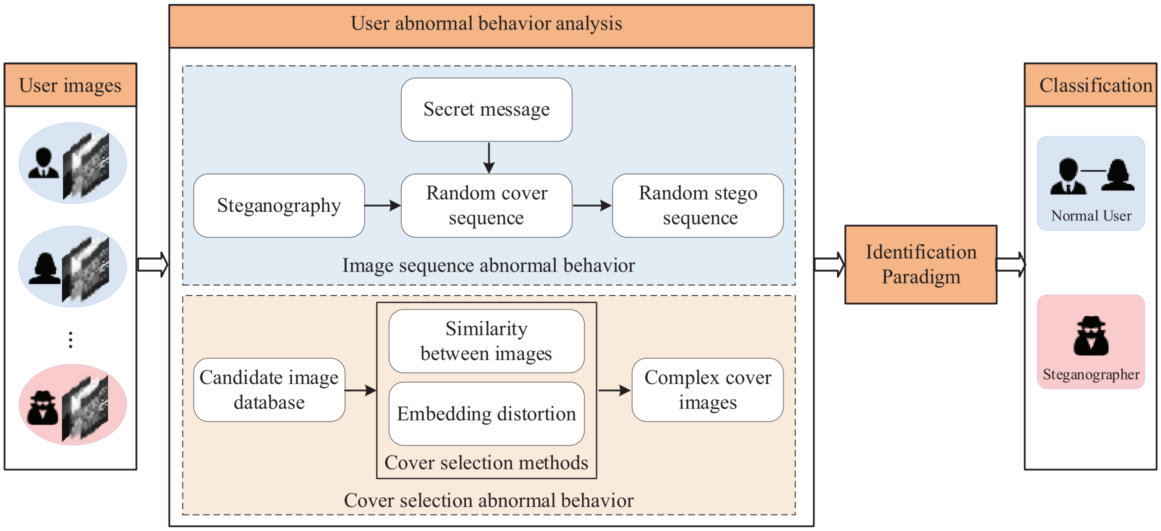

Figure 5: General framework of side channel steganographer identification

A. Image Sequence Behavior Anslysis

In 2018, Li et al. [55] proposed a behavioral separability feature, SIAM (subtractive images adjacent model), between normal users and steganographers. To our knowledge, this is the first research on identifying steganographers using features other than images. SIAM feature design is on the assumption that the images from normal users are usually of a certain sequence and relevance. In contrast, there is usually no such relationship between the images of steganographers because they are more likely to select images suitable for steganography to covert communication. At this point, we can define consistency to describe this sequential relationship of images, as follows:

$$F_{Consistency} = \frac{N_{Total} - N_{Random}}{N_{Total}} \tag{10}$$

where $F_{consistency}$ represents the steganographer's consistency level, N_{Random} and N_{Total} represent the number of random images and total images per user, respectively.

B. Complex Cover Selection Behavior

Steganographers typically choose suitable covers to resist steganalysis attacks. For example, Reference [81] considered images with good visual quality as suitable covers, and some scholars consider the image content and choose images with high texture complexity as covers [50,82]. Based on this, Wang et al. [83] observed that the covers selected by existing cover selection methods [84,85] normally have different characteristics from normal ones, and proposed a steganalysis method to capture such differences. Although the cover selection behaviour is used for image steganalysis in [83], it is also pointed out that the detection results of the images using this method also help in user identification at the same time.

C. Discussion

Actually, there are many differences between steganographers and normal users in terms of behavior. At present, side channel steganographer identification is still in its beginning state, and there is much more to study and explore in this area in the future.

3.3 Feature Optimization Methods

In this section, we introduce the typical feature optimization methods used in existing literature for image steganographer identification, including feature calibration, feature dimension reduction and high dimensional feature ensemble learning. The brief structure is illustrated in Fig. 6, and the details are as follows.

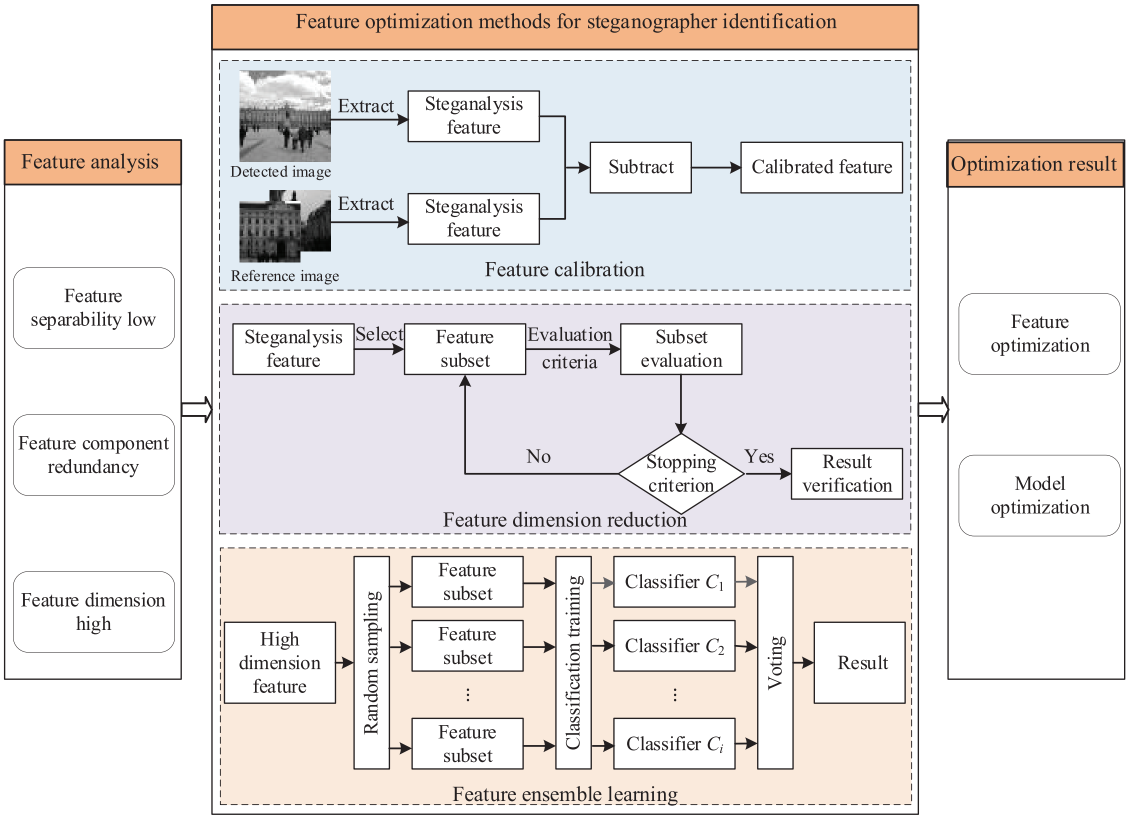

Figure 6: Feature optimization methods

3.3.1 Feature Calibration

In 2020, Li et al. [56] proposed SIAM-C feature for steganographer identification, which are calibrated SIAM [55]. The feature calibration method [86,87] is commonly used in steganalysis to eliminate the differences between different images and noise interference to make the extracted steganalysis features more stable. In [56], an image with similar content to the user's image is selected as a reference image to calibrate the SIAM features.

3.3.2 Feature Dimension Reduction

Image steganalysis has been studied for a longer period as compared to steganographer identification. Several effective single-image steganalysis features have been born. However, these cannot be used for steganographer identification directly. For example, Reference [54] stated that it is nearly impossible to steganographer identification using high-dimensional features such as DCTR (Discrete Cosine Transform Residual) [88] and PHARM [89]. Although the two feature sets are conclusively more sensitive for supervised binary classification, these rich feature sets are formed by a lot of weak features. The weak features contain a large amount of noise, which leads to an inferior performance in steganographer identification. This phenomenon has also been demonstrated in [90]. Based on this, Ma et al. [91] explored a feature selection method for single-image steganalysis.

In main channel steganographer identification, Zhang et al. [92] proposed a steganographer identification method based on steganalysis feature dimension reduction without trying to extracting new image feature. In Section 3.2, we present the experiments of feature dimension reduction methods in steganographer identification.

3.3.3 Feature Ensemble Learning

For the same reasons, Wu [93] proposed a steganographer identification method based on feature bagging [94]. This method can deal with samples with high dimensional features without dimension reduction. Feature bagging randomly extracts subsets from the original high-dimensional feature space and trains multiple feature subspaces using an identification algorithm.

Finally, the prediction results of each sub feature space are integrated.

3.4 Identification Paradigms and Performance Evaluation

Performance evaluation is aiming to evaluate the performance of the steganographer identification algorithm. In this section, we introduce the typical evaluation metrics that widely used in existing literature for steganographer identification.

3.4.1 Performance Evaluation

A. Maximum Mean Difference. Each user corresponds to a set of images and the similarity between users is often evaluated using Maximum Mean Difference (MMD) [65]. The MMD distance can be described as follows:

Given two users, X and Y, and each containing n images. The standardized feature sets for X and Y are denoted as $\{x_i\}_{i=1}^{n}$ and $\{y_i\}_{i=1}^{n}$, respectively. The MMD distance between X and Y can be represented as follows:

$$MMD\left(X,\ Y\right) = \frac{2}{n\left(n-1\right)} \sum_{1 \le i \le j} \left[k\left(x_i,\ x_j\right) - k\left(x_j,\ y_i\right) - k\left(x_i,\ y_j\right) - k\left(y_i,\ y_j\right)\right], \tag{11}$$

where $k(x, y)$ is a kernel function which is pre-defined. The most effective kernel functions are the Linear kernel

$$k(x, y) = x^T y \tag{12}$$

and the Gaussian kernel

$$k(x, y) = \exp\left(-\gamma \|x - y\|^2\right) \tag{13}$$

where parameter γ is the inverse kernel width.

B. Identification Accuracy Rate. Identification accuracy rate AR [54] is calculated as the number of correctly detected steganographic users over the selected total number of steganographic users, i.e.,

$$AR = \frac{N_{correct}}{N_{total}} \times 100\% \tag{14}$$

where $N_{correct}$ is the number of correctly detected steganographers, and N_{total} represents the selected total number of steganographers.

C. Vote Score. The total vote score (V) [56] of actor A_i can be obtained by

$$V(j) = \sum_{m=1}^{M} v(m), \, j = 1, 2, \ldots, N \tag{15}$$

where M is the number of images per user, and N is the number of users under investigation, $v(m)$ represents the ensemble votes yielded by the ensemble classifier.

3.4.2 Hierarchical Clustering-Based Approach

The difference between a steganographer and a normal user is larger than two normal users. Based on this, Ker et al. [65] first proposed a hierarchical clustering-based steganographer identification method in 2011, and the MMD distance was used to measure the distance between users. After several rounds of iterations, the remaining individual in the last iteration is determined to be the steganographer. The hierarchical clustering methods have been used in subsequent steganographer identification studies [71,75,54]. Among them, Li et al. [54] proposed a clustering ensemble-based steganographer identification method for optimal decisions.

3.4.3 LOF-Based Approach

Reference [53] was the earliest research that utilized local outlier factor (LOF) to measure steganographers. In contrast to hierarchical clustering, the LOF-based method can use an abnormal value to weigh the user. For a detailed description and analysis of the LOF, we refer to the original article [45]. The LOF-based identification paradigm provided new ideas for subsequent research [53,56,70,77,95]. For instance, Ker et al. [53] mainly did research on realistic large-scale steganalysis and pointed out that the steganographer is the Outlier. In the content-adaptive selective identification scheme [77], LOF is employed to capture the value of anomaly for each user.

3.4.4 Graph Neural Network-Based Approach

Different from the traditional machine learning-based identification paradigm, such as hierarchical clustering and LOF, in 2020, Zhang et al. [57] first proposed a network learning-based method for steganographer identification. The network learning-based method first constructs the users as graphs using the similarity of the images. Then, the graphs as input and the user labels as output are used to train the steganographer identification model based on Graph Convolutional Neural Network (GNN). GNN is an extension of deep learning methods from structured data to unstructured data, and its core components are aggregation operations and node feature updates. For a detailed description and analysis of the GNN, we refer to the original article [46]. Since then, many network learning-based methods [58,92] have been researched for steganographer identification.

4 Methods Comparison

In this section, we offer a comparative analysis of the representative methods. The details are shown in Table 2. As can be seen from the table, both steganographer identification and steganalysis all require to extract features. However, in addition to extracting image features, steganographer identification can also extract side channel feature, which is different from steganalysis. This is because the goals are different. Steganalysis mainly focuses on the image itself, aiming to detect whether there is steganography embedding in the image. Steganographer identification focuses on the identity of the user behind the steganography, aiming to determine the subject hiding information in the image. Of course, the features that can effectively identify steganographers, especially the behavior features of users, is still in its infancy and requires further research and improvement.

Table 2: Comparison of representative approaches for steganographer identification

Method	Year	Dataset	Type	Feature extraction	Feature optimization	Identification paradigm
PEV_HC [65]	2011	RAW photos	JPEG	DK-based feature	None	Clustering
PEV_SD [53]	2014	Social network site (Oxford University network)	JPEG		None	LOF
HOJ_SD [54]	2016	Social network site (http://image.baidu.com)	JPEG		None	Clustering
RPEV_SD [70]	2017	Social network site (http://image.baidu.com) (http://images.google.com)	JPEG		Dimension reduction	LOF
FSGCN [92]	2023	BOSSBase-1.01/Bows	JPEG		Dimension reduction	GNN
RNSD [71]	2017	BOSSBase-1.01	Spatial	DL-based feature	None	Clustering
MDNNSD [75]	2018	BOSSBase-1.01	Spatial		None	Clustering
EPEDN [77]	2019	BOSSBase-1.01	Spatial		Feature fusion	LOF
MEPESD [25]	2019	BOSSBase-1.01	Spatial/JPEG		None	Gaussian vote
EGCN [57]	2020	BOSSBase-1.01/Flick	Spatial/JPEG		None	GNN
SAGCN [58]	2021	BOSSBase-1.01/Flick	Spatial/JPEG		None	GNN

(Continued)

Table 2 (continued)

Method	Year	Dataset	Type	Feature extraction	Feature optimization	Identification paradigm
MSCNN [95]	2021	BOSSBase-1.01	Spatial		None	LOF
DCNSD [80]	2023	BOSSBase-1.01	Spatial		None	Clustering
SIAM_SD [56]	2018	Twitte	JPEG	SC-based feature	None	Clustering
PSRM [93]	2019	BOSSBase-1.01/UCID	Spatial		None	Clustering
SIAM-C_SD [56]	2020	Twitte	JPEG		Feature calibration	LOF

Note: The hyperlinks in Table 2 were accessed on 31 August 2024; DK-based feature: Domain knowledge-based feature; DL-based feature: Deep learning-based feature; SC-based feature: Side channel-based feature.

The experimental comparisons and analysis of the performance of existing steganographer identification methods, including clustering-based, reduced feature dimension, and graph-based methods, are shown below. It is important to note that, all experiments were performed on the standard dataset BOSSBase1.01, where the frequency domain dataset was compressed using QF = 80. In the training stage, we construct 200 normal users and 200 steganographers, each with a sample of 50 images. The normal user images are cover images, while the steganographer images are stego images. Random sampling with replacement is used as the sampling strategy. In the testing stage, we set 99 normal users and 1 steganographer among a total of 100 users to simulate real-world scenarios. Each user has 50 images.

4.1 Clustering-Based Methods Comparison

Most of the steganographer identification methods use a hierarchical clustering based paradigm. Clustering utilizes the distance or similarity between samples to cluster more similar and less different samples into one class. Hierarchical clustering is a type of clustering algorithm that creates a hierarchical nested clustering tree by calculating the similarity between data points of different categories, aiming to cluster similar samples together and separate different samples.

In clustering-based methods, the performance is closely related to the effectiveness of extracted image features. Fig. 7 describes spatial image steganographer identification comparisons based on clustering according to the experimental results reported in [75] and [71], include the SRMQ1_SD, XuNet_SD, ANSD, RNSD [71], and MDNNSD [75]. Where SRMQ1_SD is the clustering-based steganographer identification method via SRMQ1 [62], which is a well-known spatial rich model with a single quantization step. XuNet_SD is the abbreviation of the clustering-based steganographer identification method based on the network proposed by Xu et al. [32]. ANSD is the clustering-based steganographer identification method via a well-known deep CNN architecture AlexNet [96]. It should be noted that the experiment shows the identification results of steganographers using S-UNIWARD steganography and even embedding strategy.

As seen from Fig. 7, although all of these methods use the clustering paradigm, the identification performance is significantly different when the user uses different payloads of steganography based on various image features. However, in general, the performance of low-dimensional features significantly outperforms that of high-dimension features in steganographer identification. We found that SRMQ1 features have the highest dimension and the worst identification performances. Based on this, we analyze the effect of feature dimension on identification performance, as shown in the following subsection.

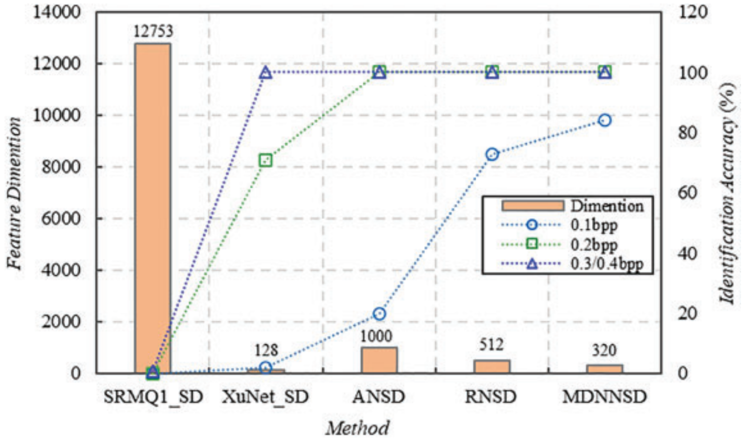

Figure 7: Performance comparison on clustering-based methods

4.2 *Feature Dimension Reduction Comparison*

This section focuses on showing the effect of feature dimensions on identification performance. Fig. 8 shows the identification accuracy of feature selection in different dimensions of CCPEV [87] steganalysis feature. In this case, steganographer use the UED steganography commonly used in the JPEG domain. We select the k-dimension feature components from CCPEV with the highest separability [92] in order of feature separability (k is 50, 100, 150, 200, and 250). The five feature dimensionality reduction of the steganalysis features, together with the 274-D Pev and 548-D CCPev features [87], a total of seven dimensions are used as image features based on the graph neural network as the results of the steganographer identification experiments. The influence of feature dimensions on identification performance is clearly illustrated in Fig. 8.

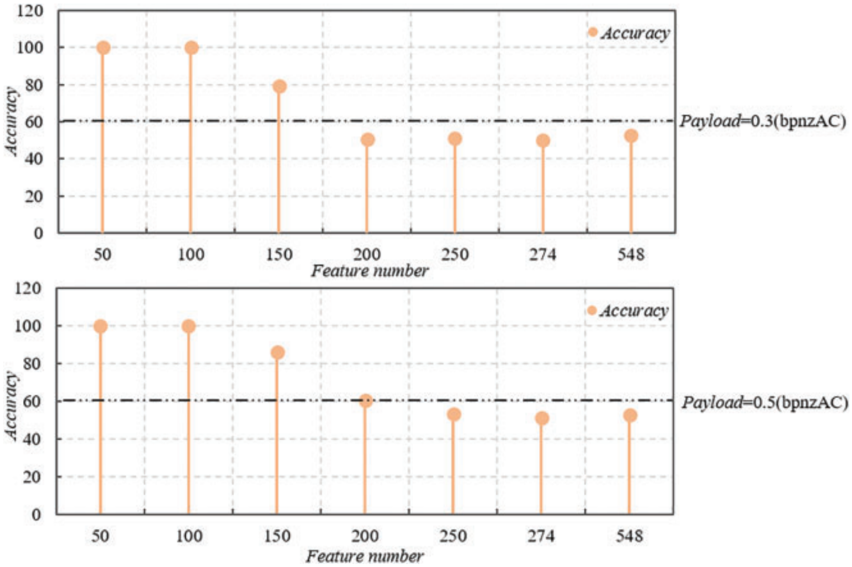

Figure 8: Performance comparison on dimension reduction (The CCPEV features as an example)

Compared to Pev-274 and CCPev steganalysis features, satisfactory identification accuracy is achieved based on the selected low-dimensional features. It is easy to calculate that the dataset is reduced to less than one-tenth of the original, and the complexity of the training set is low, which improves the speed of the classification algorithm and the identification accuracy. At the same time, the selected feature components are quantified in terms of their contribution to classification, perhaps with better readability and interpretability.

4.3 Graph Neural Network-Based Methods Comparison

The experimental results show that the features of different dimensions of steganalysis not only have significant differences in time and space complexity but directly affect the accuracy of model identification. Of course, the key is the effectiveness of the feature selection method at this time.

The previous section demonstrates the method's performance regarding image feature extraction methods, feature dimensions, etc. In addition, the model's design is also crucial for the identification accuracy. The clustering method is used in steganographer identification. On the one hand, it only uses user image statistical feature differences to distinguish users, which leads to a decrease in identification accuracy when the differences between users are minor. On the other hand, the unsupervised learning method is more susceptible to noise, which leads to the instability of the recognition results. The state-of-the-art steganographer identification method uses a geometric deep-learning architecture approach (graph neural network) to represent and recognize steganographers. This work provided new ideas for research in the field of steganographer identification.

Fig. 9 presents the comparative results of conventional clustering-based and GNN-based methods, respectively. Fig. 9a is the result of user utilizing S-UNIWARD steganography, and Fig. 9b shows the result of user utilizing J-UNIWARD steganographer identification. In the spatial domain, the comparisons in the experiments include the SRMQ1 [62] with hierarchical clustering for steganographer identification (SRMQ1SD), the MDNNSD [75] method, the GCN-based [46] method, SAGCN [58] method, and MILGCN [97]. In the JPEG domain, the comparisons include the JRM [98] with hierarchical clustering for steganographer identification (JRM_SD), the PEV [66] with hierarchical clustering for steganographer identification (PEV_SD), the GCN-based [46] method, EGCN [57] method, and FSGCN method [92]. According to the experimental results reported in [57] and [58], the graph neural network-methods are obviously superior to the traditional clustering-based method.

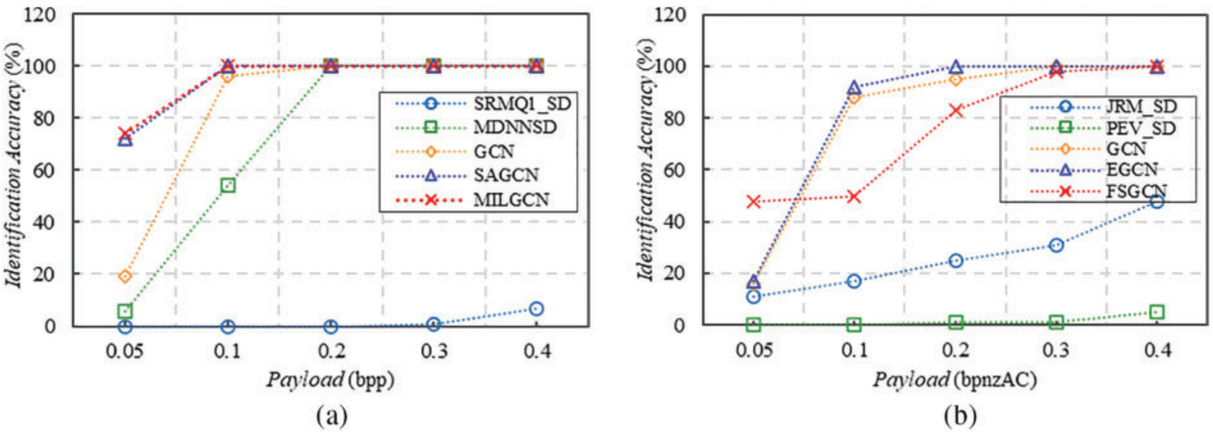

Figure 9: Performance comparison on GNN-based methods. (a) The result of user utilizing S-UNIWARD steganography; (b) The result of user utilizing J-UNIWARD steganography

4.4 Typical Identification Methods Comparison in Frequency Domain

Based on the experimental results reported in the literature, we further present a comprehensive analysis of the typical identification methods in the frequency domain. Fig. 10 shows the experimental results of different methods for identifying steganographers using nsF5 and UED, respectively. Among them, Fig. 10a is the experimental results of users utilizing nsF5 steganography, and Fig. 10a represents the experimental results of users utilizing UED steganography. It should be noted that, the JPEG version of BOSSbase1.01 is utilized to evaluate the proposed method in frequency domain. The comparisons include the PEV [66] with hierarchical clustering for steganographer identification (PEV_Cluster), the PEV [66] with LOF for steganographer identification (PEV_LoF), the GCN-based [46] method (GCN), SAGCN [58], MILGCN [97].

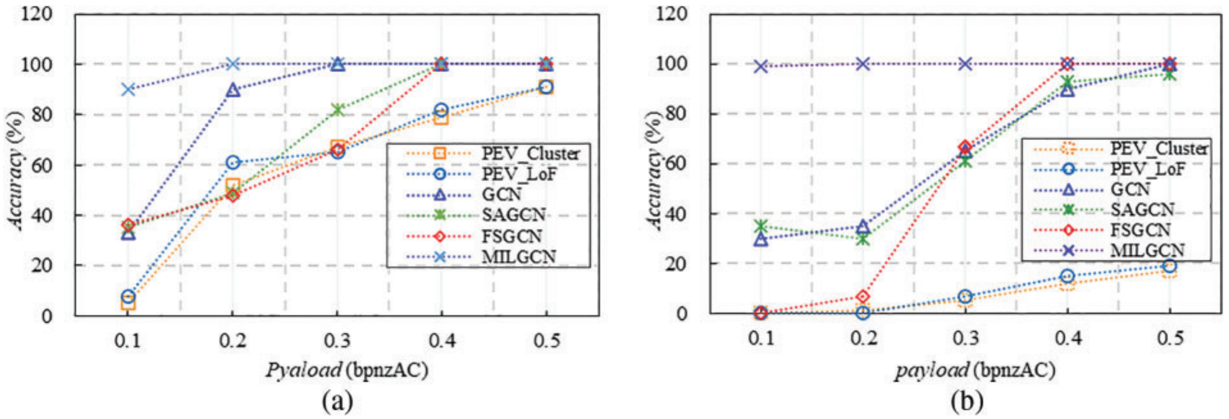

Figure 10: Performance comparison in frequency domain. (a) The results of users utilizing nsF5 steganography; (b) The results of users utilizing UED steganography

From Fig. 10, we can see that graph-based approaches outperform the conventional method that solely relies on statistical features of the image. On one hand, we analyzed that the graph-based method may increase inter-class difference by fusing image features and structural information. On the other hand, multi-layer graph convolutional neural networks possess powerful learning abilities to automatically capture differences between various user types.

4.5 Discusstion

The problem of steganographer identification was first addressed by Ker et al. in 2011. They employed two unsupervised learning methods for steganographer identification, i.e., clustering-based detection and LOF-based detection. It can be said that most of the advanced works reported in the literature are based on them. Unsupervised learning methods do not depend on prior information or labels but instead, distinguish steganographers from normal users by establishing correlations or similarities between data. However, unsupervised learning is highly susceptible to the influence of noise, which can impact its accuracy and reliability. For example, the LOF-based method is incompatible with rich features containing many weak features because the weak features include lots of noise (caused by cover content), resulting in inferior performance.

In contrast, the state-of-the-art GNN-based identification method has good generalization ability, which can somewhat alleviate the problem of cover mismatch. However, the method is still in its infancy. For example, these frameworks are designed for users sharing the same number of images. If the number of images users share is significantly different, upsampling or downsampling is required so that the sampling strategy directly affects the model. Therefore, more questions will likely arise as the study of steganographer identification moves ahead.

In addition to that, it has been shown that the methods of feature processing, such as dimension reduction, can improve the identification performance. In steganographer identification, there are two main reasons for not using high-dimension features. First, the high-dimension feature is usually composed of many weak features that contain a large amount of noise, resulting in inferior performance. Secondly, high-dimension features will affect the distance measurement, making the distance between users smaller, thus affecting the identification accuracy. Feature dimension reduction can exploit salient features and eliminate irrelevant feature fluctuations by representing the discriminative information in a lower dimensional manifold. However, it is crucial to analyze the redundancy of feature components in this method.

5 Future Issues and Open Challenges

This review summarises the recent research on steganographer identification. Although the before-mentioned efforts have yielded substantial results, due to the diversity of available steganography and the the complexity of the social network, there are still several problems that deserve further research.

1. Research on multi-modal features fusion methods for improving classification performance.

In complex and diverse application scenarios, steganographer identification faces increasingly severe challenges. Existing research often relies on single-modal features, which have certain limitations. Strongly representative features can facilitate models in discover patterns and rules in data, thereby improving performance and generalization ability. One of the main challenges we encounter is performing multi-modal feature fusion to obtain richer feature representation for distinguish users. Research on this technique can further advance the development of steganographer identification.

2. Extraction of distinguishable side channel feature.

Steganographers in social networks often exhibit distinct behavioral features compared to normal users, which can used as a side channel. Further exploration of individual behavioral features based on specific application scenarios, such as posting behavior, social interaction patterns, and topic preferences, etc. is recommended. Additionally, combining user behavior with image features for multi-modal information fusion shows promise for future developments in this field. It is important to note that as the number of available features and information increases, the time required for model training and tuning will also increase. Therefore, investigating methods to simplify models while maintaining high performance is essential.

3. Construction of models with strong generalization ability.

The practical challenges in identifying steganographers include time-consuming classification of each user due to the wide variety of media types. Therefore, it is necessary to explore methods for improving model efficiency. Additionally, the diversity of image sources poses an impossible challenge in steganographer identification, leading to a mismatch problem. Hence, future research should focus on Transfer Learning, Domain-Adaptive techniques, and other methods that establish connections between images from different sources to enhance the model's generalization ability and identification accuracy.

4. Active defense of AIGC steganography in social networks.

The development of artificial intelligence has led to the emergence of AI Generated Content (AIGC), which encompasses text, image, audio, and video. However, as AIGC applications become more widespread, concerns regarding privacy and security arise. Malicious user attacks pose a significant threat to the data security of AIGC. These users can manipulate the AIGC model by injecting false data samples to generate misleading or harmful outcomes in the content it produces. They may also exploit generative steganography techniques for covert communication and illegal activities, posing risks to individual privacy and network security. Therefore, researching active defense techniques in steganography is crucial for preventing covert communication failures and tracing AIGC on social networks.

Nowadays, image steganographer identification is still challenging in many aspects, and we highlight some future research directions as follows:

1. Looking for new methods of user feature fusion.

The existing methods often rely on a single type of user data for training and prediction. In the future, it is worth considering fusing and aligning different modal information from users to enhance model performance using multiple modalities.

2. Designing robust identification model.

Although most of the advanced works in the literature are based on two general frameworks for steganographer identification, i.e., clustering-based detection and outlier-based detection, there are significant issues of data imbalance and scarce labels in real-world scenario. In future research, more efforts can be made to address these problems and improve the reliability of identification models.

3. Researching cross-domain covert communication defense.

The potential applications of steganographer identification technology can be explored in various disciplines, enhancing its practical value and effectiveness when combined with cyberspace security, digital forensics, antifraud, and other fields.

6 Conclusion

This survey has extensively reviewed steganographer identification, a pivotal security technique that can provide comprehensive security protection by defending against malicious covert communication. First, we introduce the research background and outline the issue of steganographer identification. Then, a general framework for steganographer identification is introduced, and the existing methods are presented in detail based on this framework. Besides that, the advantages and limitations of these methods are uncovered by comparing them experimentally and theoretically. We find that feature extraction strategies, such as image features or behavioral features, affect the accuracy of identification results, which means that proper feature extraction approaches may lead to better identification performance. Moreover, the latest identification models based on graph neural network perform better. We also find that, in main channel steganographer identification, the dimension of image features significantly affects the identification performance. Therefore, a suitable feature dimension reduction method can obtain good performance. At last, we discuss the possible future issues of steganographer identification, and demonstrate the potential research directions.

Acknowledgement: The authors would like to express our sincere gratitude and appreciation to each other for our combined efforts and contributions throughout the course of this research paper.

Funding Statement: This work is supported by the National Key Research and Development Program of China (No. 2022YFB3102900), the National Natural Science Foundation of China (Nos. 62172435, 62202495 and 62002103), Zhongyuan Science and Technology Innovation Leading Talent Project of China (No. 214200510019), Key Research and Development Project of Henan Province (No. 2211321200), the Natural Science Foundation of Henan Province (No. 222300420058).

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Qiangqian Zhang, Yi Zhang and Xiangyang Luo; data collection and analysis: Qianqian Zhang, Yuanyuan Ma and Yanmei Liu; draft manuscript preparation: Qianqian Zhang. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: In the study, we used the Bossbase1.01 and Bows2 datasets, which are publicly available and can be accessed via the citation links in the paper.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] W. Li, B. Li, W. Zhang, and S. Zhang, "Quaternary quantized gaussian modulation with optimal polarity map selection for JPEG steganography," *IEEE Trans. Inf. Forensics Secur.*, vol. 18, pp. 5026–5040, Apr. 2023. doi: 10.1109/TIFS.2023.3303715.

[2] J. Wu, B. Chen, W. Luo, and Y. Fang, "Audio steganography based on iterative adversarial attacks against convolutional neural networks," *IEEE Trans. Inf. Forensics Secur.*, vol. 15, pp. 2282–2294, Apr. 2020. doi: 10.1109/TIFS.2019.2963764.

[3] P. Fan, H. Zhang, and X. Zhao, "Adaptive QIM with minimum embedding cost for robust video steganography on social networks," *IEEE Trans. Inf. Forensic. Secur.*, vol. 17, pp. 3801–3815, Apr. 2022. doi: 10.1109/TIFS.2022.3215901.

[4] H. Yang, Y. Bao, Z. Yang, S. Liu, Y. Huang and S. Jiao, "Linguistic steganalysis via densely connected LSTM with feature pyramid," in *Proc. IH&MMSec*, Denver, CO, USA, 2020, pp. 2282–2294.

[5] H. V. Desai, "Steganography, cryptography, watermarking: A comparative study," *J. Glob. Res. Comput. Sci.*, vol. 3, no. 12, pp. 33–35, Apr. 2012.

[6] S. Wiseman, "Stegware–Using steganography for malicious purposes," in *Technical Report DS-2017-4*. Malvern, UK: Everfox, 2017.doi: 10.13140/RG.2.2.15283.53289

[7] P. Bak, J. Bieniasz, M. Krzemiński, and K. Szczypiorski, "Application of perfectly undetectable network steganography method for malware hidden communication," in *Proc. Int. Conf. Front. Signal Process. (ICFSP)*, Poitiers, France, 2018, pp. 34–38.

[8] J. Yang, Z. Yang, J. Zou, H. Tu, and Y. Huang, "Linguistic steganalysis toward social network," *IEEE Trans. Inf. Forensic. Secur.*, vol. 18, pp. 859–871, Apr. 2022. doi: 10.1109/TIFS.2022.3226909.

[9] T. Qiao, X. Luo, T. Wu, M. Xu, and Z. Qian, "Adaptive steganalysis based on statistical model of quantized DCT coefficients for JPEG images," *IEEE Trans. Dependable Secur. Comput.*, vol. 18, no. 6, pp. 2736–2751, Apr. 2021. doi: 10.1109/TDSC.2019.2962672.

[10] Y. Xue, L. Kong, W. Peng, P. Zhong, and J. Wen, "An effective linguistic steganalysis framework based on hierarchical mutual learning," *Inf. Sci.*, vol. 586, pp. 140–154, 2022. doi: 10.1016/j.ins.2021.11.086.

[11] Y. Ren, D. Liu, C. Liu, Q. Xiong, J. Fu and L. Wang, "A universal audio steganalysis scheme based on multiscale spectrograms and DeepResNet," *IEEE Trans. Dependable Secur. Comput.*, vol. 20, pp. 665–679, Apr. 2023. doi: 10.1109/TDSC.2022.3141121.

[12] K. Wei, W. Luo, S. Tan, and J. Huang, "Universal deep network for steganalysis of color image based on channel representation," *IEEE Trans. Inf. Forensic. Secur.*, vol. 17, pp. 3022–3036, 2022. doi: 10.1109/TIFS.2022.3196265.

[13] D. Jung, H. Bae, H. -S. Choi, and S. Yoon, "PixelSteganalysis: Pixel-wise hidden information removal with low visual degradation," *IEEE Trans. Dependable Secur. Comput.*, vol. 20, pp. 331–342, Apr. 2019. doi: 10.1109/TDSC.2021.3132987.

[14] J. Li, X. Luo, Y. Zhang, P. Zhang, C. Yang and F. Liu, "Extracting embedded messages using adaptive steganography based on optimal syndrome-trellis decoding paths," *Digit. Commun. Netw.*, vol. 8, pp. 455–465, Apr. 2021. doi: 10.1016/j.dcan.2021.09.005.

[15] W. Shuo, "Recent advances in image-based steganalysis research," (in Chinese), *Chin. J. Comput.*, vol. 32, no. 7, pp. 1247–1263, Apr. 2009. doi: 10.3724/SP.J.1016.2009.01247.

[16] N. Zhong, Z. Qian, Z. Wang, X. Zhang, and X. Li, "Batch steganography via generative network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 1, pp. 88–97, Apr. 2021. doi: 10.1109/TCSVT.2020.2974884.

[17] M. Sewak, "Introduction to deep learning," *Adv. Deep Learn. Eng. Sci.*, vol. 5, pp. 1–22, 2021. doi: 10.1007/978-3-030-66519-7_1.

[18] K. Chen, H. Zhou, Y. Wang, M. Li, W. Zhang and N. Yu, "Cover reproducible steganography via deep generative models," *IEEE Trans. Dependable Secur. Comput.*, vol. 20, pp. 3787–3798, Apr. 2022. doi: 10.1109/TDSC.2022.3217569.

[19] W. Tang, B. Li, M. Barni, J. Li, and J. Huang, "An automatic cost learning framework for image steganography using deep reinforcement learning," *IEEE Trans. Inf. Forensic. Secur.*, vol. 16, pp. 952–967, Apr. 2021. doi: 10.1109/TIFS.2020.3025438.

[20] C. Mou, Y. Xu, J. Song, C. Zhao, B. Ghanem and J. Zhang, "Large-capacity and flexible video steganography via invertible neural network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Vancouver, BC, Canada, 2023, pp. 22606–22615.

[21] D. Huang, W. Luo, M. Liu, W. Tang, and J. Huang, "Steganography embedding cost learning with generative multi-adversarial network," *IEEE Trans. Inf. Forensic. Secur.*, vol. 19, pp. 15–29, Apr. 2024. doi: 10.1109/TIFS.2023.3318939.

[22] J. Zhu, "Passive defense against 3D adversarial point clouds through the lens of 3D steganalysis," 2022, *arXiv:2205.08738v1*.

[23] Y. Ma, X. Yu, X. Luo, D. Liu, and Y. Zhang, "Adaptive feature selection for image steganalysis based on classification metrics," *Inf. Sci.*, vol. 644, Apr. 2023, Art. no. 118973. doi: 10.1016/j.ins.2023.118973.

[24] H. Du, J. Liu, X. Y. Luo, and Y. Zhang, "Extraction method of secret message based on optimal hypothesis test," *IEEE Trans. Dependable Secur. Comput.*, vol. 20, pp. 5265–5277, Apr. 2023. doi: 10.1109/TDSC.2023.3243907.

[25] S. Zhong, Y. Wang, T. Ren, M. Zheng, Y. Liu and G. Wu, "Steganographer detection via multi-scale embedding probability estimation," *ACM Trans. Multim. Comput. Commun. Appl.*, vol. 15, no. 4, pp. 1–23, Apr. 2020. doi: 10.1145/3352691.

[26] C. Francis-Christie and D. Lo, "A combination of active and passive video steganalysis to fight sensitive data exfiltration through online video," in *Proc. IEEE Annu. Comput. Softw. Appl. Conf.*, Atlanta, GA, USA, 2016, pp. 371–376.

[27] Z. Zhu, S. Li, Z. Qian, and X. Zhang, "Destroying robust steganography in online social networks," *Inf. Sci.*, vol. 581, pp. 605–619, Apr. 2021. doi: 10.1016/j.ins.2021.10.023.

[28] Z. Zhu, P. Wei, Z. Qian, S. Li, and X. Zhang, "Image sanitization in online social networks: A general framework for breaking robust information hiding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, pp. 3017–3029, Apr. 2023. doi: 10.1109/TCSVT.2022.3224243.

[29] G. Feng, X. Zhang, Y. Ren, Z. Qian, and S. Li, "Diversity-based cascade filters for JPEG steganalysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, pp. 376–386, Apr. 2020. doi: 10.1109/TCSVT.2019.2891778.

[30] V. Sachnev, N. Sundararajan, and S. Suresh, "A new approach for JPEG steganalysis with a cognitive evolving ensembler and robust feature selection," *Cogn. Comput.*, vol. 15, no. 2, pp. 751–764, Apr. 2023. doi: 10.1007/s12559-022-10087-3.

[31] S. Chhikara and R. Kumar, "Information theoretic steganalysis of processed image LSB steganography," *Multim. Tools Appl.*, vol. 82, no. 9, pp. 13595–13615, Apr. 2023. doi: 10.1007/s11042-022-13931-8.

[32] G. Xu, H. Wu, and Y. Shi, "Structural design of convolutional neural networks for steganalysis," *IEEE Signal Process. Lett.*, vol. 23, no. 5, pp. 708–712, Apr. 2016. doi: 10.1109/LSP.2016.2548421.

[33] J. Ye, J. Ni, and Y. Yi, "Deep learning hierarchical representations for image steganalysis," *IEEE Trans. Inf. Forensic. Secur.*, vol. 12, pp. 2545–2557, 2017. doi: 10.1109/TIFS.2017.2710946.

[34] M. Boroumand, M. Chen, and J. J. Fridrich, "Deep residual network for steganalysis of digital images," *IEEE Trans. Inf. Forensic. Secur.*, vol. 14, pp. 1181–1193, Apr. 2019. doi: 10.1109/TIFS.2018.2871749.

[35] A. Ker and T. Pevny, "Identifying a steganographer in realistic and heterogeneous data sets," in *Proc. Med. Watermarking, Secur., Forensic.*, Burlingame, CA, USA, 2012, 83030N.

[36] A. Ker *et al.*, "Moving steganography and steganalysis from the laboratory into the real world," in *Proc. IH&MMSec*, Montpellier, France, 2013, pp. 45–58.

[37] A. Zakaria, M. Chaumont, and G. Subsol, "Pooled steganalysis in JPEG: How to deal with the spreading strategy," in *Proc. IEEE Int. Workshop Inform. Forensic. Secur. (WIFS)*, Delft, Netherlands, 2019, pp. 1–6.

[38] K. Karampidis, E. Kavallieratou, and G. Papadourakis, "A review of image steganalysis techniques for digital forensics," *J. Inf. Secur. Appl.*, vol. 40, pp. 217–235, Apr. 2018. doi: 10.1016/j.jisa.2018.04.005.

[39] A. Selvaraj, A. Ezhilarasan, S. L. J. Wellington, and A. R. Sam, "Digital image steganalysis: A survey on paradigm shift from machine learning to deep learning-based techniques," *IET Image Process*, vol. 15, no. 2, pp. 504–522, Apr. 2021. doi: 10.1049/ipr2.12043.

[40] F. Ruan, X. Zhang, D. Zhu, Z. Xu, S. Wan and L. Qi, "Deep learning for real-time image steganalysis: A survey," *J. Real Time Image Process*, vol. 17, no. 1, pp. 149–160, 2020. doi: 10.1007/s11554-019-00915-5.

[41] T. Muralidharan, A. Cohen, A. Cohen, and N. Nissim, "The infinite race between steganography and steganalysis in images," *Signal Process*, vol. 201, Apr. 2022, Art. no. 108711. doi: 10.1016/j.sigpro.2022.108711.

[42] A. Ker, "Batch steganography and pooled steganalysis," in *Proc. Int. Workshop Inform. Hiding*, Alexandria, VA, USA, 2006, pp. 265–281.

[43] R. Cogranne, V. Sedighi, and J. Fridrich, "Practical strategies for content-adaptive batch steganography and pooled steganalysis," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, New Orleans, LA, USA, 2017, pp. 2122–2126.

[44] L. Li, W. Zhang, C. Qin, K. Chen, W. Zhou and N. Yu, "Adversarial batch image steganography against CNN-based pooled steganalysis," *Signal Process*, vol. 181, Apr. 2021, Art. no. 107920. doi: 10.1016/j.sigpro.2020.107920.

[45] M. M. Breunig, H. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying density-based local outliers," in *Proc. ACM SIGMOD Conf. Int. Conf. Manage. Data*, Dallas, TX, USA, 2000, pp. 93–104.

[46] T. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*.

[47] W. Tang, B. Li, S. Tan, M. Barni, and J. Huang, "CNN-based adversarial embedding for image steganography," *IEEE Trans. Inf. Forensic. Secur.*, vol. 14, no. 8, pp. 2074–2087, Apr. 2019. doi: 10.1109/TIFS.2019.2891237.

[48] X. Hu, J. Ni, W. Zhang, and J. Huang, "Efficient JPEG batch steganography using intrinsic energy of image contents," *IEEE Trans. Inf. Forensic. Secur.*, vol. 16, pp. 4544–4558, Apr. 2021. doi: 10.1109/TIFS.2021.3109464.

[49] Z. Wang, S. Li, X. Zhang, and G. Feng, "Exploring abnormal behavior in swarm: Identify user using adversarial examples," *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 7, no. 1, pp. 250–260, Apr. 2023. doi: 10.1109/TETCI.2022.3201294.

[50] M. Subhedar and V. Mankar, "Curvelet transform and cover selection for secure steganography," *Multim. Tools Appl.*, vol. 77, no. 7, pp. 8115–8138, Apr. 2018. doi: 10.1007/s11042-017-4706-x.

[51] P. Bas, T. Filler, and T. Pevny, "Break our steganographic system: The ins and outs of organizing boss," in *Proc. Inform. Hiding*, Prague, Czech Republic, 2011. doi: 10.1007/978-3-642-24178-9_5.

[52] A. Piva and M. Barni, "The first bows contest: Break our watermarking system," in *Proc. Secur., Steganograp., Watermarking Multimed. Contents IX*, San Jose, CA, USA, 2007.

[53] A. D. Ker and T. Pevny, "The steganographer is the outlier: Realistic large-scale steganalysis," *IEEE Trans. Inf. Forensic. Secur.*, vol. 9, no. 9, pp. 1424–1435, Apr. 2014. doi: 10.1109/TIFS.2014.2336380.

[54] F. Li, K. Wu, J. Lei, M. Wen, Z. Bi and C. Gu, "Steganalysis over large-scale social networks with high-order joint features and clustering ensembles," *IEEE Trans. Inf. Forensic. Secur.*, vol. 11, no. 2, pp. 344–357, Apr. 2016. doi: 10.1109/TIFS.2015.2496910.

[55] L. Li, W. Zhang, K. Chen, H. Zha, and N. Yu, "Side channel steganalysis: When behavior is considered in steganographer detection," *Multim. Tools Appl.*, vol. 78, no. 7, pp. 8041–8055, Apr. 2019. doi: 10.1007/s11042-018-6582-4.

[56] L. Li, W. Zhang, K. Chen, and N. Yu, "Steganographic security analysis from side channel steganalysis and its complementary attacks," *IEEE Trans. Multim.*, vol. 22, no. 10, pp. 2526–2536, Apr. 2020. doi: 10.1109/TMM.2019.2959909.

[57] Z. Zhang, M. Zheng, S. Hua Zhong, and Y. Liu, "Steganographer detection via enhancement-aware graph convolutional network," in *Proc. IEEE Int. Conf. Multimed. Expo*, London, UK, 2020, pp. 1–6.

[58] Z. Zhang, M. Zheng, S. Zhong, and Y. Liu, "Steganographer detection via a similarity accumulation graph convolutional network," *Neural Netw.*, vol. 136, pp. 97–111, Apr. 2021. doi: 10.1016/j.neunet.2020.12.026.

[59] M. A. Russell and M. Klassen, "*Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More*, 2nd ed. Sebastopol, USA: O'Reilly Media Press, 2013, pp. 137–138.

[60] D. Li, "The MNIST database of handwritten digit images for machine learning research," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 141–142, Apr. 2012. doi: 10.1109/MSP.2012.2211477.

[61] A. Krizhevsky, "Learning multiple layers of features from tiny images," 2009. Accessed: Aug. 31, 2024. [Online]. Available: https://www.cs.utoronto.ca/~kriz/learning-features-2009-TR.pdf

[62] J. J. Fridrich and J. Kodovsky, "Rich models for steganalysis of digital images," *IEEE Trans. Inf. Forensic. Secur.*, vol. 7, no. 3, pp. 868–882, Apr. 2012. doi: 10.1109/TIFS.2012.2190402.

[63] X. Song, F. Liu, C. Yang, X. Luo, and Y. Zhang, "Steganalysis of adaptive JPEG steganography using 2D Gabor filters," in *Proc. ACM Workshop Inform. Hiding Multimed. Secur.*, Portland, OR, USA, 2015, pp. 15–23.

[64] J. Zhang, K. Chen, C. Qin, W. Zhang, and N. Yu, "Distribution-preserving-based automatic data augmentation for deep image steganalysis," *IEEE Trans. Multim.*, vol. 24, pp. 4538–4550, Apr. 2022. doi: 10.1109/TMM.2021.3119994.

[65] A. D. Ker and T. Pevny, "A new paradigm for steganalysis via clustering," in *Proc. Med. Forensic. Secur. III*, San Francisco, CA, USA, 2011.

[66] T. Pevny and J. Fridrich, "Multiclass detector of current steganographic methods for JPEG format," *IEEE Trans. Inf. Forensic. Secur.*, vol. 3, no. 4, pp. 635–650, Apr. 2008. doi: 10.1109/TIFS.2008.2002936.

[67] L. Mauri, B. Apolloni, and E. Damiani, "Robust ML model ensembles via riskdriven anti-clustering of training data," *Inf. Sci.*, vol. 633, pp. 122–140, Apr. 2023. doi: 10.1016/j.ins.2023.03.085.

[68] V. Sedighi, R. Cogranne, and J. Fridrich, "Content-adaptive steganography by minimizing statistical detectability," *IEEE Trans. Inf. Forensic. Secur.*, vol. 11, no. 2, pp. 221–234, Apr. 2016. doi: 10.1109/TIFS.2015.2486744.

[69] W. Zhou, W. Zhang, and N. Yu, "A new rule for cost reassignment in adaptive steganography," *IEEE Trans. Inf. Forensic. Secur.*, vol. 12, pp. 2654–2667, Apr. 2017. doi: 10.1109/TIFS.2017.2718480.

[70] F. Li, M. Wen, J. Lei, and Y. Ren, "Efficient steganographer detection over social networks with sampling reconstruction," *Peer-to-Peer Netw. Appl.*, vol. 11, no. 5, pp. 924–939, Apr. 2018. doi: 10.1007/s12083-017-0603-3.

[71] M. Zheng, S. Zhong, S. Wu, and J. Jiang, "Steganographer detection via deep residual network," in *Proc. IEEE Int. Conf. Multimed. Expo*, Hong Kong, China, 2017, pp. 235–240.

[72] S. Wu, S. Zhong, and Y. Liu, "Steganalysis via deep residual network," in *Proc. IEEE Int. Conf. Parallel Distrib. Syst. (ICPADS)*, Wuhan, China, 2016, pp. 1233–1236.

[73] X. Kong, C. Feng, M. Li, and Y. Guo, "Iterative multi-order feature alignment for JPEG mismatched steganalysis," *Neurocomputing*, vol. 214, pp. 458–470, Apr. 2016. doi: 10.1016/j.neucom.2016.06.037.

[74] D. Meg'ias and D. Lerch-Hostalot, "Subsequent embedding in targeted image steganalysis: Theoretical framework and practical applications," *IEEE Trans. Dependable Secur. Comput.*, vol. 20, no. 2, pp. 1403–1421, Apr. 2023. doi: 10.1109/TDSC.2022.3154967.

[75] M. Zheng, S. Hua Zhong, S. Wu, and J. Jiang, "Steganographer detection based on multiclass dilated residual networks," in *Proc. ACM Int. Conf. Multimed. Retr.*, Yokohama, Japan, 2018, pp. 300–308.

[76] T. Denemark, M. Boroumand, and J. Fridrich, "Steganalysis features for content-adaptive JPEG steganography," *IEEE Trans. Inf. Forensic. Secur.*, vol. 11, pp. 1736–1746, Apr. 2016. doi: 10.1109/TIFS.2016.2555281.

[77] M. Zheng, J. Jiang, S. Wu, S. Zhong, and Y. Liu, "Content-adaptive selective steganographer detection via embedding probability estimation deep networks," *Neurocomputing*, vol. 365, pp. 336–348, Apr. 2019. doi: 10.1016/j.neucom.2019.07.068.

[78] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, 2014, pp. 3431–3440.

[79] Z. Luo, A. Mishra, A. Achkar, J. Eichel, S. Li and P. Jodoin, "Non-local deep features for salient object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Honolulu, HI, USA, 2017, pp. 6593–6601.

[80] E. Amrutha, S. Arivazhagan, and W. Jebarani, "Deep clustering network for steganographer detection using latent features extracted from a novel convolutional autoencoder," *Neural Process. Lett.*, vol. 55, no. 3, pp. 2953–2964, Apr. 2023. doi: 10.1007/s11063-022-10992-6.

[81] O. Evsutin, A. Kokurina, and R. Meshcheryakov, "Approach to the selection of the best cover image for information embedding in JPEG images based on the principles of the optimality," *J. Decis. Syst.*, vol. 27, pp. 256–264, Apr. 2018. doi: 10.1080/12460125.2018.1460163.

[82] H. Sajedi and M. Jamzad, "Cover selection steganography method based on similarity of image blocks," in *Proc. IEEE Int. Conf. Comput. Inform. Technol. Workshops*, Sydney, Australia, 2008, pp. 379–384.

[83] Z. Wang, S. Li, and X. Zhang, "Towards improved steganalysis: When cover selection is used in steganography," *IEEE Access*, vol. 7, pp. 168914–168921, Apr. 2019. doi: 10.1109/ACCESS.2019.2955113.

[84] M. Kharrazi, H. Sencar, and N. Memon, "Cover selection for steganographic embedding," in *Proc. Int. Conf. Image Process.*, Atlanta, GA, USA, 2006, pp. 117–120.

[85] Z. Wang, G. Feng, L. Shen, and X. Zhang, "Cover selection for steganography using image similarity," *IEEE Trans. Dependable Secur. Comput.*, vol. 20, no. 3, pp. 2328–2340, Apr. 2023. doi: 10.1109/TDSC.2022.3181039.

[86] J. Fridrich, M. Goljan, and D. Hogea, "Steganalysis of JPEG images: Breaking the F5 algorithm," in *Proc. Inform. Hiding*, Noordwijkerhout, Netherlands, 2002, pp. 310–323.

[87] J. Kodovsky and J. Fridrich, "Calibration revisited," in *Proc. Workshop Multimed. Secur.*, Princeton, NJ, USA, 2009, pp. 63–74.

[88] V. Holub and J. Fridrich, "Low-complexity features for JPEG steganalysis using undecimated DCT," *IEEE Trans. Inf. Forensic. Secur.*, vol. 10, no. 2, pp. 219–228, Apr. 2015. doi: 10.1109/TIFS.2014.2364918.

[89] V. Holub and J. Fridrich, "Phase-aware projection model for steganalysis of JPEG images," in *Proc. Med. Watermarking, Secur., Forensic.*, San Francisco, CA, USA, 2015, pp. 259–269.

[90] T. Pevny and A. Ker, "The challenges of rich features in universal steganalysis," in *Proc. Med. Watermarking, Secur., Forensic.*, Burlingame, CA, USA, 2013, pp. 203– 217.

[91] Y. Ma, L. Xu, Y. Zhang, T. Zhang, and X. Luo, "Steganalysis feature selection with multidimensional evaluation & dynamic threshold allocation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 3, pp. 1954–1969, Apr. 2024. doi: 10.1109/TCSVT.2023.3295364.

[92] Q. Zhang, Y. Zhang, H. Li, Y. Ma, and X. Luo, "Steganographer identification of JPEG image based on feature selection and graph convolutional representation," (in Chinese), *J. Commun.*, vol. 44, no. 7, pp. 218–229, Apr. 2023. doi: 10.11959/j.issn.1000.

[93] H. Wu, "Feature bagging for steganographer identification," 2018, *arXiv:1810.11973*.

[94] A. Lazarevic and V. Kumar, "Feature bagging for outlier detection," in *Proc. Knowl. Discov. Data Min.*, Chicago, MI, USA, 2005, pp. 157–166.

[95] J. Yang, C. Dong, F. Zhang, M. Lei, and X. Bai, "MSCNN: Steganographer detection based on multi-scale convolutional neural networks," in *Proc. Wireless Algorithms, Syst., Appl.*, Nanjing, China, 2021, pp. 215–226.

[96] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Neural Inform. Process. Syst.*, Lake Tahoe, CA, USA, 2012, pp. 84–90.

[97] S. Zhong and Z. Zhang, "Steganographer detection via multiple-instance learning graph convolutional networks," (in Chinese), *Acta Automatica Sinica*, vol. 50, no. 4, pp. 771–789, Apr. 2024. doi: 10.1145/3065386.

[98] T. Pevny and I. Nikolaev, "Optimizing pooling function for pooled steganalysis," in *Proc. IEEE Int. Workshop Inform. Forensic. Secur.*, Roma, Italy, 2015, pp. 1–6.