mathematics

MDPI

Article

Pooled Steganalysis via Model Discrepancy

Jiang Yu [1,*], Jing Zhang [1,*] and Fengyong Li [2]

[1] Faculty of Business Information, Shanghai Business School, Shanghai 200235, China
[2] College of Computer Science and Technology, Shanghai University of Electric Power, Shanghai 201306, China; fyli@shiep.edu.cn
* Correspondence: jiangyu@sbs.edu.cn (J.Y.); zhangj@sbs.edu.cn (J.Z.)

Abstract: Pooled steganalysis aims to discover the guilty actor(s) among multiple normal actor(s). Existing techniques mainly rely on the high-dimension and time-consuming features. Moreover, the minor feature distance between cover and stego is detrimental to pooled steganalysis. To overcome these issues, this paper focuses on the discrepancy of the statistical characteristics of transmitted multiple images and designs a model-based effective pooled steganalysis strategy. Facing the public and monitored channel, without using the feature extractions, pooled steganalysis collects a set of images transmitted by a suspicious actor and use the corresponding distortion values as the statistic representation of the selected image set. Specifically, the normalized distortion of the suspicious image set generated via normal/guilty actor(s) is modelled as a normal distribution, and we apply maximum likelihood estimation (MLE) to estimate the parameter (cluster center) of the distribution by which we can represent the defined model. Considering the tremendous distortion difference between normal and stego image sets, we can deduce that the constructed model can effectively discover and reveal the existence of abnormal behavior of guilty actors. To show the discrepancy of different models, employing the logistic function and likelihood ratio test (LRT), we construct a new detector by which the ratio of cluster centers is turned into a probability. Depending on the generated probability and an optimal threshold, we make a judgment on whether the dubious actor is normal or guilty. Extensive experiments demonstrate that, compared to existing pooled steganalysis techniques, the proposed scheme exhibits great detection performance on the guilty actor(s) with lower complexity.

Keywords: pooled steganalysis; batch steganography; distortion distribution; guilty actor; discrepancy

MSC: 94A99

check for updates

Citation: Yu, J.; Zhang, J.; Li, F. Pooled Steganalysis via Model Discrepancy. *Mathematics* **2024**, *12*, 552. https://doi.org/10.3390/math12040552

Academic Editors: Cheonshik Kim and Ki-Hyun Jung

Received: 19 January 2024
Revised: 7 February 2024
Accepted: 8 February 2024
Published: 11 February 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Batch steganography is a technology to embed secret information into multiple cover media with slight modifications to avoid detection. Conversely, pooled steganalysis aims to disclose covert communication by monitoring and analyzing the digital media transmitted on public channel [1]. In general, batch steganography employs spatial images or JPEG images selected from a given image set or downloaded from the internet to achieve the delivering of secret data [2]. Pooled steganalysis involves an eavesdropper (Warden) aiming to disclose the covert communication by observing the overall behaviors of images transmitted via one or more actors [3,4]. Generally, after collecting several images from public channels [5,6], the typical steganalysis method is to represent the images with features through which pooled steganalysis determines whether the suspicious actor is normal/guilty [7]. In [8], based on agglomerative hierarchical clustering and novel features, blind and universal steganalysis is designed to achieve the preferable detection of the guilty actor(s) [9].

As the adversary, batch steganography methods consider the universal cases in which more actors transmit messages with multiple images [10]. To ensure security, batch

steganography selects the most complex images as covers, while other smooth images are not used. Normally, a steganographer achieves the maximum security via two consecutive steps. The first critical step is designing a distortion function on the local region of a pixel or JPEG coefficient [11], where the complex region is assigned on a smaller value and vice versa. Therefore, the distortion function called embedding cost can be seen as the statistical representation of the cover and could reveal the complexity of the cover image. The second key procedure Syndrome-Trellis codes (STC) minimizes the total embedding distortion [12,13]. Hence, based on the distortion function and STC, there are many novel methods, such as WOW (Wavelet Obtained Weights), SUNIWARD (Spatial Universal Wavelet Relative Distortion) [14], MVGG (Multivariate Generalized Gaussian), HILL (High-pass Low-pass and Low-pass) [15], JUNIWARD (JPEG Universal Wavelet Relative Distortion), UERD (Uniform Embedding Revisited Distortion) [16] and so on [17–19]. Conversely, there are many important features to counter the steganography [20–25].

From the perspective of security, if a guilty actor hopes to spread a fixed-length message across multiple covers, relying on a distortion function, batch steganography evaluates the complexity of a given individual image and assigns unequal embedding capacity to the given cover. Obviously, to achieve higher security, the more complex image loads more messages, and vice versa. Based on the distortion-based mechanism, many cover-selection approaches are designed. In [26], the cover-selection and payload-allocation are unified to obtain a novel batch steganography method. Considering the practical scenarios, Wang et al. [27] proposed a practical secure cover selection strategy. Besides the above-mentioned methods, researchers have designed many batch steganography schemes [28,29].

Obviously, the straightforward way to conquer batch steganography is pooled steganalysis which focuses on the feature difference between normal and guilty actor(s), while few efforts have been dedicated to the statistical model of multiple images. To the best of our knowledge, modern steganalysis can roughly be classified into two categories supervised and unsupervised. The supervised method consists of three consecutive parts: feature representation, model construction (classifier training) and testing. For another type of method, existing techniques always choose some novel features as the statistical description. Despite demonstrated effectiveness, clustering-based pooled steganalysis may suffer from two limitations. First, for smaller payloads, the minor feature distance between the cover and stego image is lower for the Warden to make the optimal decision. Second, for high-dimensional features, the time complexity of feature extraction is rather high. Inspired by the successful model-based cases in steganalysis [30–35], we aim to design an effective pooled steganalysis to counter batch steganography. To provide a better illustration of the related work on steganalysis, we summarize in Table 1.

Table 1. Summary of existing steganalysis method.

Method	Feature	Classifier	Supervised/Unsupervised	Complexity	Detection
SPAM [20]	Need	Ensemble classifier	Supervised	Low	Image-level
SRM [21]	Need	Ensemble classifier	Supervised	High	Image-level
DCF [22]	Need	Ensemble classifier	Supervised	High	Image-level
TLBP [23]	Need	Ensemble classifier	Supervised	High	Image-level
Clustering [9]	Need	No	Unsupervised	High	Actor-level
Ours	No	Logistic function, likelihood ratio test (LRT)	Unsupervised	Low	Actor-level

In this paper, we propose an unsupervised pooled steganalysis scheme to achieve excellent detection performance on batch steganography, and the schematic representation of the proposed method is shown in Figure 1. To decrease the algorithm complexity, in our scheme, we remove feature extraction in our scheme. For a fixed payload, the distortion set of a certain suspicious image set is modeled as a normal distribution with appropriate model parameters. To simplify, we use maximum likelihood estimation (MLE) to estimate

the cluster center of the defined distribution. Meanwhile, employing the logistic function and the likelihood ratio test (LRT), the ratio of cluster centers of the different distributions reveals the discrepancy of distributions, through which we make a judgment whether the actor is a normal or guilty one. Extensive results validate that the proposed scheme shows great detection ability on the guilty actor(s), especially for the smaller parameter cases. The contributions of this article are listed as follows:

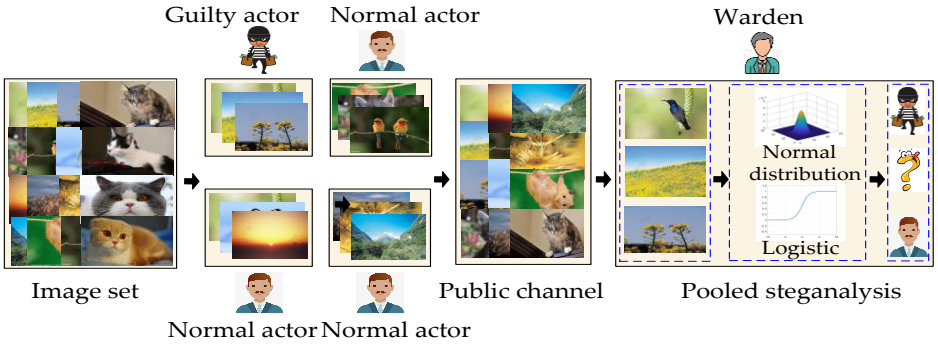


Figure 1. Schematic representation of the proposed method.

(1) **Without feature extraction:** We abandon the feature extraction operation employed in contemporary steganalysis. Traditional steganalysis is denoted as a binary classification problem, which means that the steganalyst should have full or partial knowledge of the steganographic mechanism and the distribution of cover sources. Moreover, the dimensionality of novel rich features is always large, which implies that the complexity of feature extraction and the testing phase is rather high. Therefore, the higher complexity resists the practical implication of traditional pooled steganalysis.

(2) **Model-based strategies:** To make the right judgment on the behavior of the suspicious actor(s), we collect multiple images transmitted on the public channel. After obtaining the normalized distortion of collected images with a certain distortion function, we model the distortion with a normal distribution and the model parameters are estimated via maximum likelihood estimation (MLE). The proper judgment will be given based on the model discrepancy of different image sets.

(3) **New detector:** Our detector relies on the logistic function and likelihood ratio test (LRT). Since we consider the pooled steganalysis as an unsupervised classification problem, the model discrepancy of different image sets can be expressed and evaluated using the logistic function and likelihood ratio test (LRT). Compared with other pooled steganalysis methods, our scheme exhibits excellent detection performance.

The remaining parts of this article are organized as follows. Section 2 illustrates the basic theory of steganography, batch steganography and the typical pooled steganalysis. In Section 3, we provide detailed description of the proposed scheme including the construction of the distortion of the collected image set and the optimal detector. Later, the extensive experimental results and the corresponding discussions are presented in Section 4. Finally, we provide the summarizations and conclusions in Section 5.

2. Related Work

This section includes three basic descriptions of content-adaptive steganography, cover selection strategy and pooled steganalysis.

2.1. Content-Adaptive Steganography

To achieve high undetectability, content-adaptive batch steganography aims to conceal the message with minimal embedding distortion. Generally, depending on the image format, the embedding distortion of a certain element (pixel or DCT coefficient) can be expressed using the surrounding elements. For simplification, in this paper, we restrict the discussed cover to spatial images.

Assuming an original image set $\mathbf{U}$ contains m images, each image $\mathbf{X}_k$ includes n pixels $\{x_k(1), x_k(2), \ldots, x_k(n)\}$, where $1 \leq k \leq m$. Employing the predefined distortion function, the embedding cost $\rho_k(l)$ of each pixel $x_k(l)$ can be computed, where $1 \leq l \leq n$. Suppose the flipping probability $p_k(l)$ of $x_k(l)$ is

$$p_k(l) = \frac{e^{-\lambda\rho_k(l)}}{1 + e^{-\lambda\rho_k(l)}} \tag{1}$$

under the constraint of the payload α_k of $\mathbf{X}_k$, the theoretical minimal steganographic distortion d_k is calculated with STC

$$d_k = \sum_{l=1}^{n} p_k(l)\rho_k(l), \tag{2}$$

where λ is a positive parameter used to make (3) satisfy

$$-\sum_{l=1}^{n}\{p_k(l)\log_2 p_k(l) + [1 - p_k(l)]\log_2[1 - p_k(l)]\} = \beta_k. \tag{3}$$

Here, β_k is equal to $\alpha_k n$.

Normally, the content of an image is different, and, when the payload α_k of each selected image is fixed, the corresponding distortion is distinct. It implies that, for the given payload, the image is more complex, and the embedding distortion is lower, and vice versa for the case of a less complex one. Therefore, under the constraint of payload, we deem that the embedding distortion value can reveal the texture of individual images Here, we hope to construct the distortion-based model which is used to represent the statistical distribution of an arbitrary image set.

2.2. Batch Steganography

Batch Steganography considers the practical scenario in which the steganographer already has a large number of innocent cover images. The optimal pattern of transmitting the secret information is to split it into many small pieces and embed the segmented data into some selected images. The chosen images always possess complex regions that are highly correlated with security (undetectability). From this perspective, batch steganography depends on the complexity of multiple images to achieve better security. Nowadays, many kinds of cover selection strategies exist, such as image complexity [26], image similarity [28] and changeable DCT (discrete cosine transform) coefficients. In [29], by measuring and maintaining the specific MMD (maximum mean discrepancy) distance, wang et al. proposed a batch steganography scheme to realize the optimal cover selection. Considering the steganographic changes in the transform domain, relying on the Contourlet transform (CT), Subhedar et al. designed a new cover selection scheme [30]. Except for CT, there are also other multiscale transformations employed to achieve the cover selection, such as curvelet transform and so on. Considering the real network environment, wang et al. presented a new cover selection method to resist the image processing [27]. To combine the cover selection and payload assignment, wang et al. proposed a gradual approach to minimize total embedding distortion.

2.3. Pooled Steganalysis

Pooled steganalysis (Warden) aims to discover the guilty actors among multiple normal actors. As stated in Section 2.2, every actor (normal or guilty) will transmit large groups of cover objects over a public channel. To provide a better statistical description of actors, we assume that all cover object sets are sampled from the same cover source. Obviously, the selected cover image set can be seen as a subset of the original distribution of the cover source, which motivates us to design the model-based pooled steganalysis and the detailed scheme is provided in Section 3.

More generally, batch steganography and pooled steganalysis are described and modeled as a game between Alice and Warden. Ker et al. presented three pooled strategies which are simple nonparametric tests, the average component statistic and a generalized maximum likelihood ratio test, obtaining better detection performance on LSB (Least Significant Bit) Replacement [6]. Considering the better representation of feature, Ker et al. proposed a new paradigm for steganalysis through clustering in which authors extracted the features, calculated the MMD distance between actors, and clustered actors using agglomerative clustering [9]. Since the universal feature is rich, it should be fragile in a mismatch channel and the extraction complexity of tens or hundreds of gigabytes of data is large. Pevný et al. constructed a new steganalysis method in which, through linear projection, the so-called scheme *calibrated least squares* is designed to project the feature and make it sensitive to stego content yet insensitive to cover variation [8].

Based on the analysis of batch steganography and pooled steganalysis, when an eavesdropper (Warden) pays special attentions to the public channel, we believe the selection results of multiple covers may arouse the Warden's vigilance.

3. Proposed Method

The section presents a model-based strategy to find the guilty actor among multiple normal actors. In the proposed method, we mainly construct the model of the normalized distortion values computed from any suspicious image set and create a new detector with the logistic function and LRT. Importantly, the model discrepancy is described as the ratio of model parameters through which the created detector can determine whether the actor is normal or guilty.

3.1. Architecture of the Proposed Method

According to the previous discussions, we aim to rely on the model divergence to reveal the sign of guilty actors. When the Warden monitors a public channel, he will collect multiple images. Therefore, the model can be constructed based on the statistical representation of chosen images. Considering the highly complex feature extraction, unlike other pooled steganalysis, we directly calculate the normalized distortion of selected Images, and the model is built based on the obtained distortion set. To show the model difference, we employ the logistic function and LRT to design a new detector through which the appropriate detection results can be obtained. The detailed architecture is shown in Figure 2.

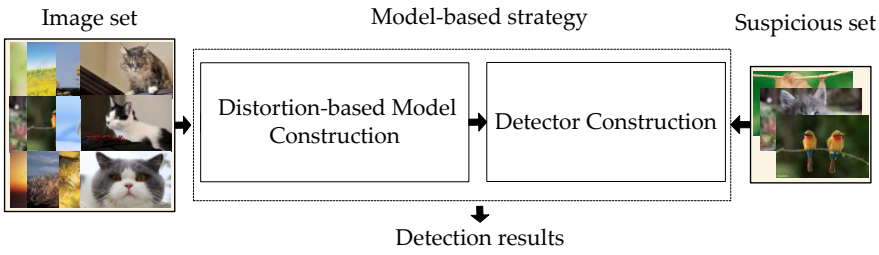


Figure 2. Detailed architecture of the proposed method.

As illustrated in Figure 2, our model-based strategy contains two consecutive parts denoted as distortion-based model construction and new detector construction. The input of the proposed method is the image source $\mathbf{U} = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_m\}$ including diverse images, such as spatial images, JPEG images and so on. The aim of the distortion-based model is to use the Gaussian distribution to model the statistical representation of multiple distortion set computed from selected images. Here, the model parameters are obtained using maximum likelihood estimation (MLE). Clearly, for the collected image set transmitted via the normal actor and guilty actor, the corresponding constructed models will possess different model parameters (mean and variance). Inspired by the success of the likelihood ratio test (LRT), we combine the logistic function and LRT to design a new detector through

which we can depend on the model discrepancy to make a proper judgement on whether the actor is normal or guilty. The detailed descriptions of each part are given as follows.

3.2. Image Set

As stated in the subsection above, the optimal cover should contain heterogeneous categories. Generally, the original image source **U** is drawn according to a probability distribution $\mathcal{Y}$, and the image subset selected for conveying data is sampled from this cover source **U**. After the embedding operation, the stego image subset can also be seen as the samples generated from a stego source **Q** sampled from a distribution $\mathcal{N}$. In our personal view, due to special cover selection strategies, the chosen image set must exhibit some particular statistical characteristics.

For a normal image probability distribution (without special design), the emergence probability of images is different. Regardless of the image format, the number of complex images is high, and the corresponding value of complex images is relatively low. When a batch steganographer utilizes a content adaptive strategy to select images as optimal covers, the created image set must contain many complex images. From a probability perspective, the emergence probability of the complex image subset is high. However, if we choose a uniform selection to generate the cover set, the distribution should follow the normal image set in which the number of complex images is small. To illustrate this phenomenon, our ultimate purpose is to employ the distortion-based image model to compute the generation probability of a certain image subset, such as a randomly/selectively generated image set. To obtain an accurate probability description, we assume each image is independent and identically distributed (IID). It means the joint probability distribution of the image subset can be presented using the product of the individual probability of each image.

Suppose the defined distortion model is $\overline{f}_\theta$ and the distortion set of the image subset is $\overline{\mathbf{D}}$, for a given image subset owing t images, the occurrence probability P_θ of the distortion subset $\overline{\mathbf{D}}$ is described as

$$P_\theta\left(\overline{\mathbf{D}}\right) = \prod_{i=1}^{t} \overline{f}_\theta\left(\overline{d}_i\right). \tag{4}$$

Assume there are two image distortion sets $\overline{\mathbf{D}}(1)$ and $\overline{\mathbf{D}}(2)$ created using the based-distortion and random selection, the occurrence probabilities are defined as

$$P_\theta\left(\overline{\mathbf{D}}(1)\right) = \prod_{i=1}^{t} \overline{f}_\theta\left(\overline{d}_i(1)\right), \tag{5}$$

and

$$P_\theta\left(\overline{\mathbf{D}}(2)\right) = \prod_{i=1}^{t} \overline{f}_\theta\left(\overline{d}_i(2)\right). \tag{6}$$

According to the selection strategy of batch steganography, $\overline{d}_i(1)$ is smaller than $\overline{d}_i(2)$ under the defined distribution, $\overline{f}_\theta\left(\overline{d}_i(1)\right)$ is also smaller than $\overline{f}_\theta\left(\overline{d}_i(2)\right)$. Therefore,

$$P_\theta\left(\overline{\mathbf{D}}(1)\right) < P_\theta\left(\overline{\mathbf{D}}(2)\right), \tag{7}$$

It implies that, for a normal channel, the occurrence probability of the guilty actors is lower than that of the normal actors.

3.3. Image Set Model Generation

Normally, the content of an image is different and, when the payload of each selected image is fixed, the corresponding distortion is distinct. It means that, for the given payload, the image is more complex, and the corresponding embedding distortion is lower, and vice versa for the case of a less complex one. Therefore, under the constraint of payload, we deem that the embedding distortion value can reveal the complexity of individual

images. Here, we aim to construct the distortion-based model, which is used to represent the statistical distribution of an arbitrary image set.

Under the framework of batch steganography, the payloads α_k of each image are different. With three defined equations and the distortion function F, d_k is computed. For simplification, assume the payload of $\mathbf{U}$ as $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, \ldots, \alpha_m\}$, the embedding distortion set is $\mathbf{D} = \{d_1, d_2, \ldots, d_m\}$ which is represented by the distribution θ parameterized using θ. Since the most images are less complex, the great majority of $\mathbf{D}$ are small. We hypothesize that $\mathbf{D}$ is characterized by a normal distribution and the probability mess function (pmf) is given as

$$f_\theta(d_k) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(d_k - \mu)^2}{2\sigma^2}\right), \forall d_k \in \mathbb{R}. \tag{8}$$

Here, θ stands for the model parameter involving two important parameters μ and σ, which stand for the expectation and standard deviation of the normal distribution.

To show the accuracy of the proposed distortion-based model, we utilize a typical distortion function and employ the defined normal distribution to fit the statistical distribution of distortion values which are computed from BOSSbass with two relative payloads 0.2 and 0.5. According to the fitting results shown in Figure 3, it is clearly that the fitting results are relatively optimal.

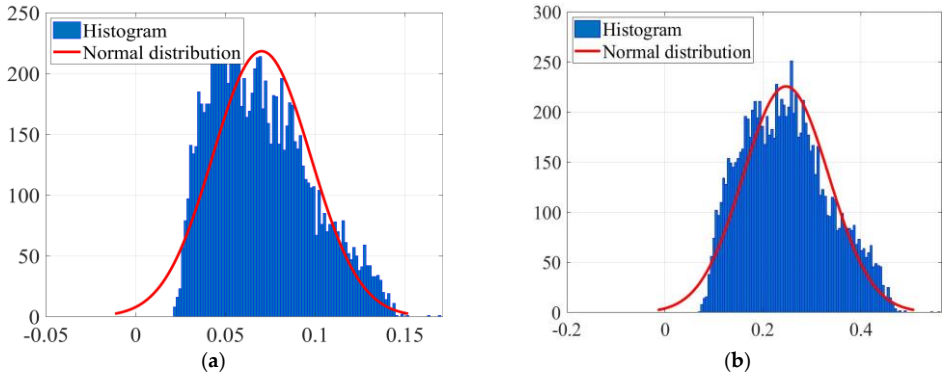


Figure 3. The fitting results of BOSSbass for two payloads, where (**a**) 0.2 and (**b**) 0.5.

For batch steganography, based on the distortion and image selection policy, a series of ideal images are chosen from the image set $\mathbf{U}$. Suppose batch steganography selects t images to construct the image subset $\overline{\mathbf{U}} = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_t\}$ to load H bits, which are constrained under the condition

$$\sum_{i=1}^{t} \alpha_i t = H, \tag{9}$$

where $1 \leq t \leq m$. Denote the corresponding distortion subset as $\overline{\mathbf{D}} = \left\{\overline{d}_1, \overline{d}_2, \ldots, \overline{d}_t\right\}$, following the same hypothesis, the statistical distribution of $\overline{\mathbf{D}}$ is represented through $\mathcal{Q}_\theta$ which is also described using a one-dimension normal distribution with $\overline{\mu}$ and $\overline{\sigma}$. The corresponding pmf is described as

$$\overline{f}_\theta\left(\overline{d}_i\right) = \frac{1}{\overline{\sigma}\sqrt{2\pi}} \exp\left(-\frac{\left(\overline{d}_i - \overline{\mu}\right)^2}{2\overline{\sigma}^2}\right), \forall \overline{d}_i \in \mathbb{R}, \tag{10}$$

where $1 \leq i \leq t$. In Figure 4a, with a payload of 0.5 and $t = 1000$, we provide the fitting result of the selected image set. Clearly, with proper parameters $\overline{\mu}$ and $\overline{\sigma}$, the distribution $\overline{f}_\theta\left(\overline{d}_i\right)$ is almost fitted using a normal distribution.

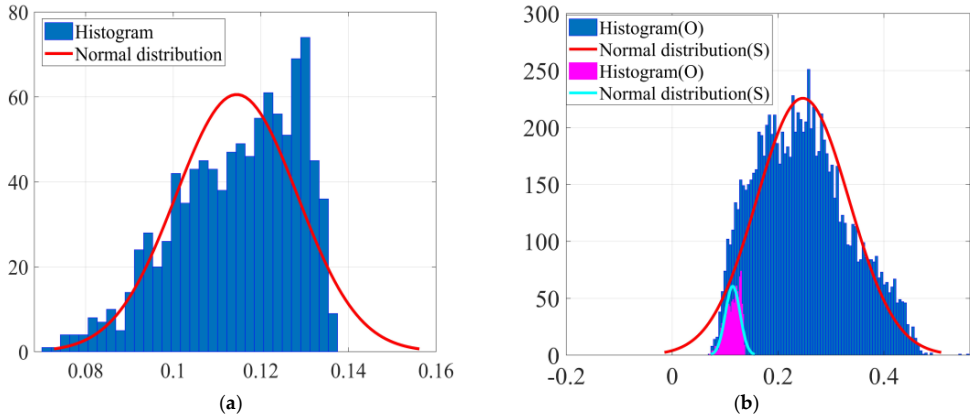


Figure 4. The fitting results of the selected image set (**a**) and the comparison of fitting results (**b**) between the selected image set and BOSSbass for a payload of 0.5.

To show the distribution difference between the original image set (BOSSbase) and the selected image set, both with the same payload, the comparison result of the model difference is listed in Figure 4b, where the two letters "O" and "S" stand for the original image set and the selected image set, respectively. According to the above discussions, we know that, for a given original image set, the embedding distortion of selected images is small. Therefore, compared with the distribution of the original image set, the distribution $\overline{f}_\theta\left(\overline{d}_i\right)$ of selected image set is located at the tail of the distribution $f_\theta(d_i)$. Here, we intend to utilize the discrepancy of distributions to differentiate the normal and guilty actor(s) (batch steganography).

3.4. Statistical Distortion-Based Detector

The existing statistical model-based detectors are designed based on the hypothesis test, in which, for a suspicious image, the detector makes a judgment with the distortion-based statistical representation (model). Different from previous methods, our detector consists of two steps. The first procedure is the logistic function, which has achieved significant success in deep learning and the second is the Likelihood Ratio Test (LRT). Combining the logistic function and LRT, the model discrepancy is turned into a probability, which is mapped into a decision value with a proper threshold.

Assume a suspicious image set is $\widetilde{\mathbf{U}}$, and the corresponding distortion subset is $\widetilde{\mathbf{D}}$. Under two hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, the detector gives a judgement whether the image set is generated by a guilty actor or a normal actor. The hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ are defined as

$$\begin{cases} \mathcal{H}_0 : \widetilde{\mathbf{U}} = \mathbf{C} \text{ is a normal image set} \\ \mathcal{H}_1 : \widetilde{\mathbf{U}} = \mathbf{S} \text{ is a selected image set} \end{cases}. \tag{11}$$

Generally, the statistical test is achieved using a mapping function $\delta : \mathbb{Z}^t \mapsto \{\mathcal{H}_0, \mathcal{H}_1\}$, and the optimal detection performance is obtained by satisfying the Neyman-Pearson bi-criteria. For a given False Positive Rate (FPR), the powerful test $\delta(\cdot)$ is given as

$$\delta\left(\widetilde{\mathbf{D}}\right) = \begin{cases} \mathcal{H}_0 \text{ if } \Lambda\left(\widetilde{\mathbf{D}}\right) = \dfrac{\widetilde{f}_\theta\left[\widetilde{\mathbf{D}}\right]}{f_\theta^\alpha[\mathbf{D}]} < \tau \\ \mathcal{H}_1 \text{ if } \Lambda\left(\widetilde{\mathbf{D}}\right) = \dfrac{\widetilde{f}_\theta\left[\widetilde{\mathbf{D}}\right]}{f_\theta^\alpha[\mathbf{D}]} \geq \tau \end{cases}, \tag{12}$$

where Λ and τ represent the Likelihood Ratio (LR) and decision threshold (see [33] for details).

As we all know, if the expectation and standard deviation are given, a normal distribution can be fixed. Moreover, the location of the distribution is determined using the expectation. For simplification, we transfer the distribution difference into the mean difference, and we use the maximum likelihood estimation (MLE) to compute the estimator of expectations of distributions. With theoretical distributions $f_\theta(d_i)$ and $\widetilde{f}_\theta\left(\widetilde{d}_i\right)$, two estimators are defined as

$$\hat{\mu} = E(\mathbf{D}) = \sum_{k=1}^{m} f_\theta(d_k) \cdot d_k, \tag{13}$$

and

$$\hat{\widetilde{\mu}} = E\left(\widetilde{\mathbf{D}}\right) = \sum_{k=1}^{t} \widetilde{f}_\theta\left(\widetilde{d}_k\right) \cdot \widetilde{d}_k, \tag{14}$$

where $\hat{\mu}$ and $\hat{\widetilde{\mu}}$ are the unbiased estimators of μ and $\widetilde{\mu}$. In fact, the construction pattern of the distortion set $\mathbf{D}$ is random. Therefore, the unbiased estimators $\hat{\mu}$ can be seen as a random statistic. For another subset $\widetilde{\mathbf{D}}$, even considering the image selection strategy, the expectation estimator $\hat{\widetilde{\mu}}$ is also regarded as a random statistic. Based on the analysis of LRT, we define the mean ratio function Ψ as

$$\Psi\left(\widetilde{\mathbf{D}}\right) = \frac{\hat{\widetilde{\mu}}}{\hat{\mu}}. \tag{15}$$

Since $\widetilde{\mathbf{U}} \neq \phi$, $\Psi > 0$. Meanwhile, when $\widetilde{\mathbf{U}} = \mathbf{U}$, $\Psi = 1$. Therefore, $\Psi \in (0, 1]$. Then, we employ the logistic function to turn the ratio of mean into the posterior probability $P(\cdot)$, and the powerful test $\delta(\cdot)$ is redefined as

$$\delta\left(\widetilde{\mathbf{D}}\right) = \begin{cases} \mathcal{H}_0 \text{ if } P\left(Y = 0 \middle| \Psi\left(\widetilde{\mathbf{D}}\right)\right) = \frac{1}{1+e^{-(a\cdot\Psi(\widetilde{\mathbf{D}})+b)}} < \tau \\ \mathcal{H}_1 \text{ if } P\left(Y = 1 \middle| \Psi\left(\widetilde{\mathbf{D}}\right)\right) = \frac{1}{1+e^{-(a\cdot\Psi(\widetilde{\mathbf{D}})+b)}} \geq \tau \end{cases}, \tag{16}$$

where a and b are the tune parameters. $Y = \{0, 1\}$ is the label of a suspicious image set. It indicates that, when the probability is smaller than the threshold τ, the detector gives a judgement that $\widetilde{\mathbf{U}}$ is generated via the random mode (normal actor). However, when the value is larger than τ, we can make judgement that $\widetilde{\mathbf{U}}$ is created through a guilty actor. The detailed algorithm is described in Algorithm 1.

Algorithm 1 Pooled steganalysis via model discrepancy

Input: Original image set $\mathbf{U} = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_m\}$; distortion function F; collected image subset $\widetilde{\mathbf{U}} = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_t\}$; payload set $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, \ldots, \alpha_m\}$;
Output: A new detector $\delta(\cdot)$; decision on the suspicious actor(s)

(1) Compute the distortion set $\{d_1, d_2, \ldots, d_m\}$ of the image set $\mathbf{U} = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_m\}$ with the constraint of payload set $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, \ldots, \alpha_m\}$;
(2) Use Gaussian distribution to model the distortion set $\mathbf{D} = \{d_1, d_2, \ldots, d_m\}$ and estimalte the model parameters $\hat{\mu}$ and $\hat{\sigma}$;
(3) Calculate the distortion set $\widetilde{\mathbf{D}} = \left\{\widetilde{d}_1, \widetilde{d}_2, \ldots, \widetilde{d}_t\right\}$ of $\widetilde{\mathbf{U}} = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_t\}$;
(4) Introduce Gaussian distribution to model the distortion set $\widetilde{\mathbf{D}} = \left\{\widetilde{d}_1, \widetilde{d}_2, \ldots, \widetilde{d}_t\right\}$ and estimalte the model parameters $\hat{\widetilde{\mu}}$ and $\hat{\widetilde{\sigma}}$;
(5) Employ LRT and logistic function to create the new detector $\delta(\cdot)$;
(6) Give the decision on the suspicious actor(s) who transmit the image subset $\widetilde{\mathbf{U}} = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_t\}$.

4. Experimental Results

This section shows the effectiveness of our proposed scheme in terms of the security and complexity.

4.1. Experimental Setup

To present the detection ability of the guilty actor(s), we use the typical image set BOSSbass [36] as the original image set abbreviated as OI, which comprises a total of $10,000$ 512×512 uncompressed grayscale images. Specifically, we apply two novel batch steganographic schemes (guilty actors proposed in [26] and [27] respectively) to choose t images ($t = 1000$) and construct the selected image set named as SI. Each image of SI is considered as the proper image to load data with higher security. In the past decades, to achieve the security, researchers have proposed many distortion functions, among which SUNIWARD and HILL are the most important algorithms. Therefore, we choose SUNIWARD and HILL to testify the effectiveness of our scheme. Using HILL and SUNIWARD, we obtain 8000 stego images with four payloads 0.05, 0.1, 0.2, and 0.3. On the opposite, the normal actors randomly select the 1000 images from original image set OI. To imitate the behavior of a normal actor, four stego image set are created with the same four payloads.

For SI and OI, cover and stego images are segmented into training sets and testing sets of equal size. After the dividing operation, we obtain two training image sets, including 500 selected/5000 original images, respectively. Meanwhile, the remaining parts of SI and OI are employed as the testing image set. At the stage of training, we utilize the training set to construct the most powerful test $\delta(\cdot)$ employing two optimal tune parameters (a and b). To obtain the parameters of the powerful test $\delta(\cdot)$, 500 selected images are segmented into non-overlapped groups $\overline{\mathbf{U}}(v)$ with parameter N, which is the number of images delivered by the guilty actors, where $v \in [1, 2, \ldots, \lfloor 500/N \rfloor]$. It implies that there are totally $\lfloor 500/N \rfloor$ guilty actors. Generally, the number of guilty actors is less than that of normal actors. For simplification, we assume the ratio between the number of normal actors and guilty actors is R, where $R \in (0, \lfloor 500/N \rfloor)$. Therefore, there are $R\lfloor 500/N \rfloor$ normal actors. With the training set, we can acquire the optimal parameters a and b.

To provide a reasonable comparison, we employ clustering-based pooled steganalysis as the comparison method, in which there are two versions comprised of two features known as the spatial rich model with the fixed quantization step 1 named as SRMQ1 and SPAM (subtractive pixel adjacency matrix). The two corresponding strategies are denoted as "Clustering-SRMQ1" and "Clustering-SPAM". Combining the proposed method, we obtain three pooled steganalysis in total. Based on each decision of single image subset $\overline{\mathbf{U}}(v)$ and ensemble mechanism [37], the final performance is evaluated using the detection error

$$P_{\mathrm{E}} = \min_{P_{\mathrm{FA}}} \frac{1}{2}(P_{\mathrm{FA}} + P_{\mathrm{MD}}), \tag{17}$$

where P_{FA} (false alarm rate) represents the rate at which a normal actor is judged as a guilty actor and P_{MD} (missed detection rate) stands for the rate at which a guilty actor is regarded as a normal actor.

4.2. Security Evaluation

With the key parameters N and R, two image sets, SI and OI, can be divided into multiple image subsets holding different security. To give fair testing, N and R are set to be large and small respectively. For the large case, we designate three parameter combinations denoted as "A", "B" and "C" representing $\{R = 30, N = 20\}$, $\{R = 40, N = 50\}$ and $\{R = 60, N = 100\}$. According to the reported results in Table 2, we observe that all testing values of the clustering-based and the proposed method are zero. It indicates that, if the parameters are large, three pooled schemes can obtain excellent detection performance and, under the condition of different payloads, they can find all the guilty actors on the public channel. The only shortcoming is that, based on the detection results in reported in Table 2, we cannot

differentiate which one is better. However, when N and R are small, the clustering-based methods are inferior to our proposed scheme at most embedding payloads.

Table 2. Detection results P_E of three pooled steganalysis with large parameters.

Payloads \ Methods	Clustering-SRMQ1			Clustering-SPAM			Proposed		
	A	B	C	A	B	C	A	B	C
0.05	0	0	0	0	0	0	0	0	0
0.1	0	0	0	0	0	0	0	0	0
0.2	0	0	0	0	0	0	0	0	0
0.3	0	0	0	0	0	0	0	0	0

Especially, N and R are set to be 2 and 3. The testing results for two novel batch steganography methods are shown in Figures 5 and 6. Compared with "Clustering-SRMQ1" and "Clustering-SPAM", the average improvements of the proposed strategy on HILL across four payloads are about 27.34%, 12.71%, 3.75% and 3.46%. For SUNIWARD, the improvements on average are 27.62%, 15.95%, 6.57% and 1.42%. Therefore, our scheme exceeds the clustering-based methods at most cases. Moreover, since we remove the feature extraction, the complexity of our scheme is lower than the that of the comparison methods.

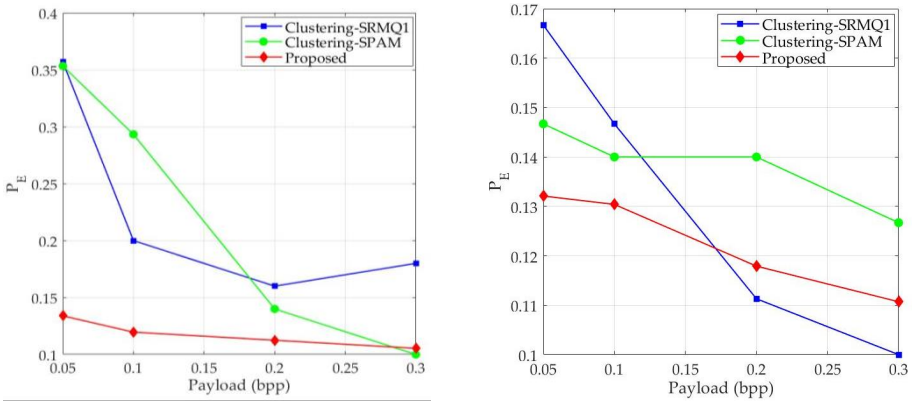


Figure 5. Detection comparison for method in [26] of two steganography methods between the proposed model discrepancy and clustering-based schemes, where (**a**) HILLand (**b**) SUNIWARD.

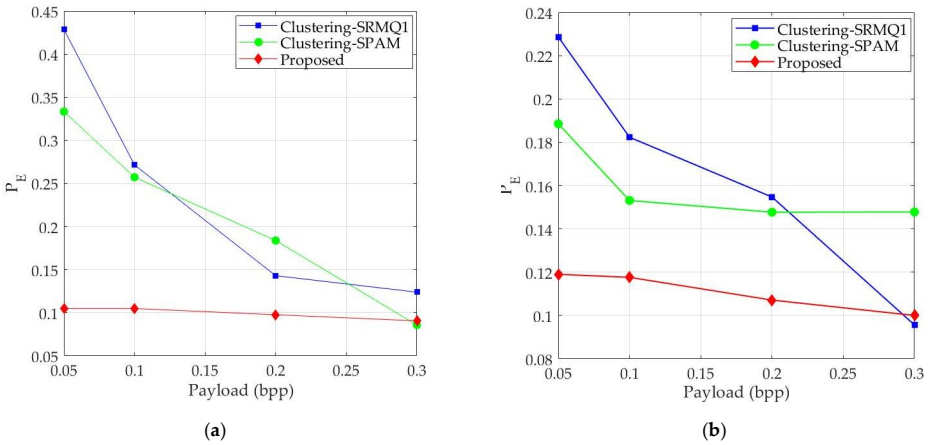
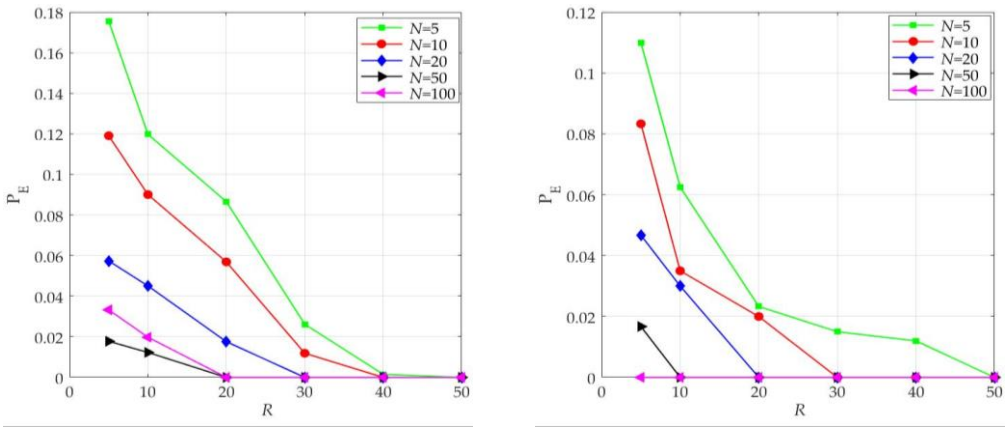


Figure 6. Detection comparison for method in [27] of two steganography methods between the proposed model discrepancy and clustering-based schemes, where (**a**) HILL and (**b**) SUNIWARD.

To provide a clear illustration of the effectiveness of batch steganography, two key parameters, N and R, are set to be diverse values. In detail, N is equal to 5, 10, 20, 50 and 100. When N varies, another parameter R is equal to 5, 10, 20, 30, 40, and 50. With all 60 parameter combinations, we obtain a total of 60 detection results on two novel content-adaptive steganography methods HILL and SUNIWARD, for payload 0.2, and all results are shown in Figure 7. Clearly, when these two parameters are set to be larger, all detection accuracies are almost zero. In reality, the demand for larger parameter values is easily guaranteed. Therefore, our scheme can achieve significant improvement under different conditions.



(**a**) (**b**)

Figure 7. Detection results for the proposed method with different parameters and two distortion functions, where (**a**) HILL and (**b**) SUNIWARD.

In some realistic scenarios, if multiple actors (normal or guilty) transmit cover objects over an insecure and public channel, the delivered data may be mixed. In fact, if the collected image set just contains only the stego images, the computed distortion mean is notably smaller than the corresponding value calculated from a normal actor. Combining the defined model and new detector, we can depend on the model discrepancy to judge whether the actor is normal or guilty. Meanwhile, the mean is seen as the macro statistics of the image set. When some images are corrupted, the mean value is not affected by the corrupted samples and our proposed scheme can effectively deal with this situation.

4.3. Time Complexity

Since the time complexity is highly related to the application, time complexity should be considered into the designing of the algorithm. Therefore, without feature extraction and the cluster processing, our proposed scheme directly uses the statistical representation (model) to describe the selected images. Meanwhile, the complexity of the training processing of the constructed new detector is rather low. In summary, combining two key points correlated with the time complexity, we think the proposed strategy has low time complexity. To provede a fair comparison, we calculate the processing time of each image of three methods and present the results in Table 3. Clearly, for each judgment, the consumed time (algorithm complexity) of the clustering-based method is extremely larger than that of our proposed method, especially for the high-dimension feature version "Clustering-SRMQ1".

Table 3. The comparison of algorithm complexity among three methods.

Methods	Clustering-SRMQ1	Clustering-SPAM	Proposed
Time (s)	46.281	1.948	0.1252

4.4. Further Comparison

To further verify the effectiveness of our scheme, we use the clustering-based methods to incorporate more features in this section. Another group of experiments is carried out on the color image set UCID [38] including 1338 color image of size 512×384, and all the color images are converted into gray versions. With the proposed cover-selection method in [26], 100 and 200 images are selected as optimal covers. For the selected images, using two typical steganographic methods, SUNIWARD and HILL, we can create 16 stego image sets across four relative payloads 0.1, 0.2, 0.3 and 0.4. Then, for the original and stego image sets, we extract two novel steganographic features SRM and LBPF [39]. Especially, N and R are set to be 2 and 3. Applying the clustering-based pooled steganalysis strategy, we can obtain the detection performance on the given cover-selection method with clustering-SRM and clustering-LBPF. As a comparison method, we use the proposed method to perform security detection on the generated stego image sets and all the comparison results are listed in Figure 8, in which "100" and "200" represent the number of the selected covers. According to the reported results, our scheme exhibits excellent detection performance across all the payloads.

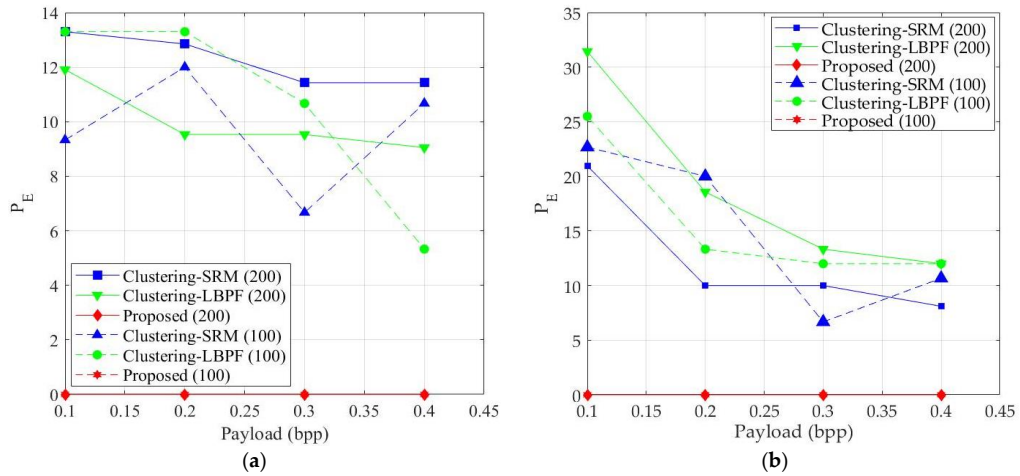


Figure 8. Detection comparison for the method in [26] of two steganography methods between the proposed model discrepancy and clustering-based schemes, where (**a**) SUNIWARD and (**b**) HILL.

5. Conclusions

This article presents a novel pooled steganalysis aimed at identifying the guilty actor on the public channel. The main aim of the proposed method is to employs the model-based strategy to make a judgment about whether a suspicious image set belongs to an abnormal actor or a normal one. Considering the cover selection mechanism, we collect several images from the monitored channel and obtain the corresponding distortion set of the chosen image set. Furthermore, the distribution of the distortions of the image set is modeled as a normal distribution along with the proper estimated parameters (slope and bias). Importantly, the discrepancy of distribution of the abnormal/normal actor is then represented by the difference in cluster centers. Using the logistic function and LRT, we design a new detector through which we make an exceptional detection performance on multiple highly secure batch steganography approaches. The extensive results show the effectiveness and low-complexity of our proposed scheme. For future study, the dynamic image generation processing will be considered.

Author Contributions: Conceptualization, J.Y.; methodology, J.Y.; software, J.Y. and J.Z.; validation, J.Y. and J.Z.; formal analysis, J.Y. and F.L.; investigation, J.Y. and F.L.; resources, F.L.; data curation, F.L.; writing-original draft preparation, J.Y.; writing-review and editing, F.L.; visualization, J.Y. and J.Z.; supervision, J.Z.; project administration, J.Y. and F.L.; funding acquisition, J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Humanities and Social Science Fund of Ministry of Education under Grant 23YJC790187.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhang, X. Behavior steganography in social network. In Proceedings of the 2012 International Hiding Multimedia Signal Process (IIH-MSP), Piraeus-Athens, Greece, 18–20 July 2012; pp. 4206–4210.
2. Ker, A.; Bas, P.; Böhme, R.; Cogranne, R.; Craver, S.; Filler, T.; Fridrich, J.; Pevný, T. Moving steganography and steganalysis from the laboratory into the real world. In Proceedings of the ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec), Montpellier, France, 17–19 June 2013; pp. 45–48.
3. Cachin, C. An information-theoretic model for steganography. *Inf. Comput.* **2004**, *192*, 41–56. [CrossRef]
4. Farid, H.; Lyu, S. Steganalysis using higher-order image statistics. *IEEE Trans. Inf. Forensics Secur.* **2016**, *1*, 111–119.
5. Hu, M.; Wan, H. Image steganalysis against adversarial steganography by combing confidence and pixel artifact. *IEEE Signal Process. Lett.* **2023**, *30*, 987–991. [CrossRef]
6. Ker, A. Batch steganography and pooled steganalysis. In Proceedings of the 8th Information Hiding Workshop, Alexandria, VA, USA, 10–12 July 2006; pp. 265–281.
7. Barni, M.; Tondi, B. The source identification game: An information-theoretic perspective. *IEEE Trans. Inf. Forensics Secur.* **2013**, *8*, 450–463. [CrossRef]
8. Pevny, T.; Ker, A. The challenges of rich features in universal steganalysis. In *Media Watermarking, Security, and Forensics*; SPIE: Philadelphia, PA, USA, 2013; Volume 8665, pp. 203–217.
9. Ker, A.; Pevny, T. A new paradigm for steganalysis via clustering. In *Media Watermarking, Security, and Forensics III*; SPIE: Philadelphia, PA, USA, 2011; Volume 7880, pp. 312–324.
10. Evsutin, O.; Kokurian, A. Cover selection for steganographic embedding. In Proceedings of the 2006 IEEE International Conference on Image Process (ICIP), Atlanta, GA, USA, 8–11 October 2006; pp. 117–120.
11. Fridrich, J.; Goljan, M.; Lisonek, P.; Soukal, D. Writing on wet paper. *IEEE Trans. Signal Process.* **2005**, *53*, 3923–3935. [CrossRef]
12. Filler, T.; Judas, J.; Fridrich, J. Minimizing additive distortion in steganography using syndrome-trellis codes. *IEEE Trans. Inf. Forensics Secur.* **2011**, *6*, 920–935. [CrossRef]
13. Filler, T.; Judas, J.; Fridrich, J. Minimizing embedding impact in steganography using trellis-coded quantization. In *Media Forensics and Security II*; SPIE: Philadelphia, PA, USA, 2010; Volume 7541, pp. 38–51.
14. Holub, V.; Fridrich, J.; Demark, T. Universal distortion function for steganography in an arbitrary domain. *EURASIP J. Inf. Secur.* **2014**, *2014*, 1. [CrossRef]
15. Li, B.; Wang, M.; Huang, J.; Li, X. A new cost function for spatial image steganography. In Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP), Paris, France, 27–30 October 2014; pp. 4206–4210.
16. Guo, L.; Ni, J.; Su, W.; Tang, C.; Shi, Y. Using statistical image model for JPEG steganography: Uniform embedding revisited. *IEEE Trans. Inf. Forensics Secur.* **2015**, *10*, 2669–2680. [CrossRef]
17. Su, W.; Ni, J.; Li, X.; Shi, Y. A new distortion function design for JPEG steganography using the generalized uniform embedding strategy. *IEEE Trans. Circuits Syst. Video Technol.* **2018**, *28*, 3545–3549. [CrossRef]
18. Wang, Z.; Qian, Z.; Zhang, X.; Yang, M.; Ye, D. On improving distortion functions for JPEG steganography. *IEEE Access* **2018**, *6*, 74917–74930. [CrossRef]
19. Sedighi, V.; Cogranne, R.; Fridrich, J. Content-adaptive steganography by minimizing statistical detectability. *IEEE Trans. Inf. Forensics Secur.* **2015**, *11*, 221–234. [CrossRef]
20. Pevný, T.; Bas, P.; Fridrich, J. Steganalysis by subtractive pixel adjacency matrix. *IEEE Trans. Inf. Forensics Secur.* **2010**, *5*, 215–224. [CrossRef]
21. Fridrich, J.; Kodovský, J. Rich models for steganalysis of digital images. *IEEE Trans. Inf. Forensics Secur.* **2012**, *7*, 868–882. [CrossRef]
22. Feng, G.; Zhang, X.; Ren, Y.; Qian, Z.; Li, S. Diversity-based cascade filters for JPEG steganalysis. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *30*, 376–386. [CrossRef]
23. Li, B.; Li, Z.; Zhou, S.; Tan, S.; Zhang, X. New steganalyitic features for spatial image steganography based on derivative filters and threshold LBP operator. *IEEE Trans. Inf. Forensics Secur.* **2017**, *13*, 124201257.
24. Holub, V.; Fridrich, J. Low-Complexity Features for JPEG Steganalysis Using Undecimated DCT. *IEEE Trans. Inf. Forensics Secur.* **2015**, *10*, 219–228. [CrossRef]
25. Li, F.; Zhang, X.; Cheng, H.; Yu, J. Digital image steganalysis based on local texture feature and double dimensionality reduction. *Secur. Commun. Netw.* **2016**, *9*, 729–736. [CrossRef]
26. Wang, Z.; Zhang, X. Joint cover-selection and payload-allocation by steganographic distortion optimization. *IEEE Signal Process. Lett.* **2018**, *25*, 1530–1534. [CrossRef]
27. Wang, Z.; Zhang, X.; Qian, Z. Practical cover selection for steganography. *IEEE Signal Process. Lett.* **2020**, *27*, 71–75. [CrossRef]

28. Wang, Z.; Feng, G.; Shen, L.; Zhang, X. Cover selection for steganography Using image similarity. *IEEE Trans. Dependable Secure Comput.* **2023**, *20*, 2328–2340. [CrossRef]
29. Wang, Z.; Zhang, X. Secure cover selection for steganography. *IEEE Access.* **2019**, *7*, 57857–57867. [CrossRef]
30. Subhedar, M.; Mankar, V. Curvelet transform and cover selection for secure steganography. *Multimedia. Tool Appl.* **2018**, *77*, 8115–8138. [CrossRef]
31. Cogranne, R.; Zitzmann, C.; Retraint, F.; Nikiforov, I.V.; Cornu, P.; Fillatre, L. A local adaptive model of natural images for almost optimal detection of hidden data. *Signal Process.* **2014**, *100*, 169–185. [CrossRef]
32. Zitzmann, C.; Cogranne, R.; Fillatre, L.; Nikiforov, I.V.; Retraint, F.; Cornu, P. Hidden information detection based on quantized laplacian distribution. In Proceedings of the 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Kyoto, Japan, 25–30 August 2012; pp. 1793–1796.
33. Qiao, T.; Zitzmann, C.; Cogranne, R.; Retraint, F. Detection of JSteg algorithm using hypothesis testing theory and a statistical model with nuisance parameters. In Proceedings of the 2nd ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec), Salzburg, Austria, 11–13 June 2014; pp. 3–13.
34. Thai, H.T.; Cogranne, R.; Retraint, F. Statistical model of quantized DCT coefficients: Application in the steganalysis of JSteg algorithm. *IEEE Trans. Image Process* **2014**, *23*, 1980–1993. [CrossRef] [PubMed]
35. Qiao, T.; Zitazmann, C.; Retraint, F.; Cogranne, R. Statistical detection of JSteg steganography using hypothesis testing theory. In Proceedings of the 2014 IEEE International Conference on Image Process (ICIP), Paris, France, 27–30 October 2014; pp. 5517–5521.
36. DDE Download. 2023. Available online: http://dde.binghamton.edu/download/ (accessed on 2 September 2023).
37. Kodovský, J.; Fridrich, J.; Holub, V. Ensemble classifiers for steganalysis of digital media. *IEEE Trans. Inf. Forensics Secur.* **2020**, *7*, 432–444. [CrossRef]
38. Schaefer, G.; Stich, M. UCID: An uncompressed color image database. In Proceedings of the Storage Retrieval Methods Applications for Multimedia 2004, San Jose, CA, USA, 20–22 January 2004; pp. 472–481.
39. Chakraborty, G.; Jalar, A.S. A novel local binary pattern based blind feature image steganography. *Multimedia Tools Appl.* **2020**, *79*, 19561–19574. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.