Article

A Fast Selection Based on Similar Cross-Entropy for Steganalytic Feature

Ruixia Jin ¹, Xinquan Yu ² , Yuanyuan Ma ^{3,*}, Shuang Yin ² and Lige Xu ³
¹ SanQuan College, Xinxiang Medical University, Xinxiang 453003, China; jinruixia@sqmc.edu.cn

² College Software, Henan Normal University, Xinxiang 453007, China; xinquanyu106@163.com (X.Y.); 13525006200@163.com (S.Y.)

³ College Computer and Information Engineering, Henan Normal University, Xinxiang 453007, China; xulige980714@163.com

* Correspondence: 121100@htu.edu.cn; Tel.: +86-186-3738-8061

Abstract: The mutual confrontation between image steganography and steganalysis causes both to iterate continuously, and as a result, the dimensionality of the steganalytic features continues to increase, leading to an increasing spatio-temporal overhead. To this end, this paper proposes a fast steganalytic feature selection method based on a similar cross-entropy. Firstly, the properties of cross-entropy are investigated, through the discussion of different models, and the intra-class similarity criterion and inter-class similarity criterion based on cross-entropy are presented for the first time. Then, referring to the design principles of Fisher's criterion, the criterion of feature contribution degree is further proposed. Secondly, the variation of the cross-entropy function of a univariate variable is analyzed in principle, thus determining the normalized range and simplifying the subsequent analysis. Then, within the normalized range, the variation of the cross-entropy function of a binary variable is investigated and the setting of important parameters is determined. Thirdly, the concept of similar cross-entropy is further presented by analyzing the changes in the value of the feature contribution measure under different circumstances, and based on this, the criterion for the feature contribution measure is updated to decrease the complexity of the calculation. Remarkably, the contribution measure criterion devised in this paper is a symmetrical structure, which equitably measures the contribution of features in different situations. Fourth, the feature component with the highest contribution is selected as the final selected feature based on the result of the feature metric. Finally, based on the Bossbase 1.01 image base that is a unique standard and recognized base in steganalysis, the feature selection on 8 kinds of low and high-dimensional steganalytic features is carried out. Through extensive experiments, comparison with several classic and state-of-the-art methods, the method designed in this paper attains competitive or even better performance in detection accuracy, calculation cost, storage cost and versatility.

Keywords: steganalysis; feature selection; similar cross-entropy; fast selection; inter-class similarity

Citation: Jin, R.; Yu, X.; Ma, Y.; Yin, S.; Xu, L. A Fast Selection Based on Similar Cross-Entropy for Steganalytic Feature. *Symmetry* **2021**, *13*, 1564. <https://doi.org/10.3390/sym13091564>

Academic Editor: Tomohiro Inagaki

Received: 20 July 2021

Accepted: 20 August 2021

Published: 25 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 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

Image steganography [1–4] refers to the use of some algorithm to embed secret information in an image for covert communication. In this regard, the original image is called the cover image and the image embedded with secret information is called the stego image. On the one hand, steganography ensures the privacy of subscribers and special confidential units, but on the other hand, it can be used by unscrupulous elements to compromise public security [5–7]. This has led to the advent of steganalysis [8–11], which uses feature extraction algorithms to analyze image characteristics and then uses classifiers to distinguish between cover images and stego images, thereby safeguarding national and public security. Common steganalysis algorithms are: Jan et al. [11] proposed 548-D CC-PEV feature (the PEV features enhanced by Cartesian Calibration), Holub et al. [12] proposed 8000-D DCTR feature (Discrete Cosine Transform Residual feature), Song et al. [9] proposed 17000-D GFR feature (jpeg

Rich model utilizing Gabor Filters), Kodovský et al. [13] proposed 22510-D CC-JRM feature (Cartesian Calibrated JPEG domain Rich Model), Fridrich et al. [14] proposed 34671-D SRM feature (full Spatial domain Rich Model).

Nevertheless, along with the rapid development of adaptive steganography [15–17], in order to improve the detection accuracy of stego images, steganalysis needs to extract features from different scales and orientations [18–20], leading to an increasing number of feature dimensions [13,21,22], resulting in huge computational and storage overheads.

To efficiently solve the problem of distinguishing stego images from cover images, researchers have devised feature selection methods [23–25]. Depending on the application of the methods, the existing feature selection methods can be divided into: specific feature selection methods and general selection methods.

The specific feature selection method [25,26] means that the algorithm only works on one or some steganalytic features with weak generalization. For example, Yang et al. [25] proposed a feature subspace selection method based on Fisher's criterion (SSFC), which first calculates the Fisher value and probability value of individual feature components, and then calculates the weight of each feature component, finally, selects the feature component whose probability value is proportional to the weight as the final selected feature, which improves the detection accuracy of GFR features to a certain extent. Yu et al. [26] proposed a multi-scale feature selection method for steganalytic feature GFR (SRGS), which first used the SNR criterion to measure the uselessness of features and removed the useless features, then innovated the Relief algorithm to measure the importance of features and select the important features, and finally took the important features as the final selected features. The experiment verified that the algorithm reduced a certain number of dimensions and improved detection accuracy at the same time.

The generic feature selection method [24,27–32] means that the algorithm achieves excellent results for most existing steganalytic features. For example, Qin et al. [27] devised a principal component analysis-based feature selection method (PCA-based), which first calculates the mean value of each feature component, and then calculates the covariance matrix as well as the eigenvalues. Then arranges the feature components in descending order according to the eigenvalues, and finally selects the feature components of a specified dimension as the final selected features according to the requirements. Wang et al. [28] devised a comprehensive criterion-based feature selection method (CGSM), which was guided by the disparity function and Person coefficients, and first selected the feature components with large disparity, on this basis, then removed the interference of redundant features, reducing the feature dimensionality while slightly improving the detection accuracy for stego images. Ma et al. [24] proposed a feature selection method based on decision rough set α -positive domain approximation, which first applied rough set theory to the steganalytic feature selection, then using the attribute separability measure (ASM) criterion to measure the divisibility of feature components, which was then extended to measure the divisibility of feature vectors, and finally, the final dominant features were selected based on the classifier, and the experiments demonstrated that the method significantly reduced the dimensionality of some features.

Even though the above feature selection methods have attained certain results, there are still problems such as the feature dimensionality is still high, the selection time is still long and the detection accuracy is still low, which limits its application in practice [33–35].

In order to solve the above problems, this paper attempts to devise a fast selection method for steganalytic features based on similarity cross-entropy (FSCE). Specifically, firstly, the properties of cross-entropy are investigated, the intra-class similarity criterion and inter-class similarity criterion are investigated and proposed, followed by a further feature contribution measure criterion regarding the model of Fisher's criterion. Secondly, the variation of cross-entropy functions of univariate and binary variables is analyzed separately to determine the problem of setting important parameters. Finally, the feature component with high contribution is selected as the final selected feature based on the results of the feature metric.

Remarkably, the feature selection in image steganalysis differs from the regular feature selection method with two manifestations: on the one hand, we use two symmetrical images, i.e., cover and stego images, during training and testing. On the other hand, in feature selection, we analyze two sets of symmetrical features, i.e., cover and stego features, for the convenience of calculation.

In order to verify the effectiveness and efficiency of FSCE, a large number of experiments are carried out in Bossbase 1.01 image database [36] (That contains 10,000 gray image pictures whose size is 512×512 .), which is the only standard and recognized image base in the steganalysis field. This includes: Firstly, a comparison was made with the features selected under different thresholds to determine the correctness of the final selection threshold in this paper. Secondly, a comparison was made with the features selected by the original steganalytic features as well as randomly selected features of the same dimension. Finally, a comparison was made with several classical and state-of-the-art fast feature selection methods. The effectiveness, efficiency and generality of FSCE are verified by the above large number of experiments.

The rest of this paper is organized as follows. Section 2 introduces the related work. In Section 3, the variation of univariate and binary cross-entropy is investigated and analyzed, while the concept of similar cross-entropy is then devised for the first time, then the FSCE method is then introduced; A series of comparative experiments are conducted in Section 4 to verify the effectiveness, efficiency and generality of the FSCE. Finally, a summary of the whole paper is presented.

2. Materials and Methods

To devise an appropriate and rational feature selection method, in this section, a bit of background knowledge needed for the methods in this paper is presented. In particular, Section 2.1 introduces the Fisher criterion, which is the most popular today. In addition, Section 2.2 presents the concept of information entropy, from which the principle of cross-entropy and its properties are introduced and presented.

2.1. Fisher Criterion

The Fisher criterion was first introduced to the field of image steganalysis by Yang et al. [25], which uses the idea of “intra-class aggregation and inter-class dispersion” to measure the separability of feature components for the selection of important features. Since the feature selection process requires measuring the performance of each feature component, which is related to the rate of change of feature values and the statistical dispersion in different feature classes. In addition, the Fisher criterion, which is the classical method for measuring feature discrimination in pattern recognition, considers not only deviation between different classes of images but also dispersion of features in each class of images. As a result, it is widely used in feature selection [25,29]. The formula is as follows.

$$Fisher(f_i) = \frac{(\mu_{f_i}^c - \mu_{f_i}^s)^2}{(\sigma_{f_i}^c)^2 + (\sigma_{f_i}^s)^2} \quad (1)$$

where, f_i represents the i th feature component, $Fisher(f_i)$ represents the Fisher value of f_i , $\mu_{f_i}^c$ and $\mu_{f_i}^s$ represent the mean value of f_i in cover class and stego class, respectively. $\sigma_{f_i}^c$ and $\sigma_{f_i}^s$ represent the standard deviation of f_i in cover class and stego class, respectively. $(\mu_{f_i}^c - \mu_{f_i}^s)^2$ represents the inter-class distance of f_i , $(\sigma_{f_i}^c)^2 + (\sigma_{f_i}^s)^2$ represents the intra-class distance of f_i . Remarkably, the larger the $Fisher(f_i)$, the greater the contribution of the f_i to distinguishing the stego image from the cover images, i.e., the more likely it is to be important.

2.2. Cross-Entropy

In 1948, Sharon introduced the concept of information entropy, which solved the problem of quantifying information. The basic principle is that if a feature has a greater

effect on the clustering of data, the more information it carries, i.e., the greater its entropy value. The formula for this is as follows.

$$HI(X) = -\sum_{i=1}^n p_i \log_2(p_i) \quad (2)$$

where, $X = [X_1, X_2, \dots, X_{n-1}, X_n]$, X_i represents the i th feature, $HI(X)$ represents the information entropy of X , p_i represents the probability values of X_i . Notably, the larger the $HI(X)$, the more information the X carries, i.e., the more likely the features in the X are to be useful.

Even though information entropy can measure the information carried by a single feature set, it cannot address the discrepant information of two feature sets. For this reason, cross-entropy [37,38] was created and is formulated as follows.

$$H(P, Q) = -\sum_{i=1}^n p_i \ln(q_i) \quad (3)$$

where, $H(P, Q)$ represents the cross-entropy of P to Q , p_i represents the i th value of P , q_i represents the i th value of Q . Remarkably, the larger the $H(P, Q)$, the less information there is about the difference between P and Q .

3. FSCE

Feature selection, also known as feature dimensionality reduction, aims to reduce the number of features while maintaining or even improving the detection accuracy of stego images and speeding up feature selection. In order to measure the contribution of a single feature component, this paper considers the cross-entropy principle as a guide to constructing a similar cross-entropy-based feature contribution criterion. Among them, Section 3.1 devises a criterion for the feature contribution measure, which provides a sound basis for the selection of feature components with a high contribution. Section 3.2 discusses the problem of setting some important parameters. In addition, Section 3.3 presents the overall process of the algorithm with performance analysis. Section 3.4 illustrates the advantages of the FSCE.

3.1. Contribution Probing

Drawing on the idea of the construction for Fisher's criterion, in this section, we attempt to construct intra-class similarity and inter-class similarity criteria for the feature components utilizing the cross-entropy principle. Specifically, Section 3.1.1 introduces some important notations in this paper. Section 3.1.2 proposes the construction of the intra-class similarity criterion. Section 3.1.3 introduces the construction of the inter-class similarity criterion, and Section 3.1.4 introduces the criterion for the feature contribution measure.

3.1.1. Symbol Description

$F = [F^c, F^s]^T = [f_1, f_2, \dots, f_i, \dots, f_{N-1}, f_N]$ represents all steganalytic feature in this case. $F^c = [f_1^c, f_2^c, \dots, f_i^c, \dots, f_{N-1}^c, f_N^c]$ and $F^s = [f_1^s, f_2^s, \dots, f_i^s, \dots, f_{N-1}^s, f_N^s]$ represent all cover and stego features, respectively. $N = \|F^c\| = \|F^s\|$ represents the feature dimensions, i.e., the number of feature components contained. $f_i = [f_i^c, f_i^s]^T$ represents the set of all eigenvalues of the i th feature component. $f_i^c = [f_{i,1}^c, f_{i,2}^c, \dots, f_{i,j}^c, \dots, f_{i,M-1}^c, f_{i,M}^c]$ represents the cover feature of f_i . $f_i^s = [f_{i,1}^s, f_{i,2}^s, \dots, f_{i,j}^s, \dots, f_{i,M-1}^s, f_{i,M}^s]$ represents the stego feature of f_i . $M = \|f_i^c\| = \|f_i^s\|$ represents the total number of cover/stego images.

3.1.2. Construction of Intra-Class Similarity Criterion

Based on the cross-entropy principle in Section 2.2, we suppose utilizing this principle to assess the intra-class similarity of a single feature component. Since there are two classes of cover features and stego features in a single feature component, and the eigenvalues of the two are not identical. In order to more appropriately measure the similarity relationship within two different categories of a single feature component, this paper defines the two

categories in cross-entropy as the model of Equation (4) and follows Equation (5) to measure the intra-class similarity $DIn(f_i)$ of a single feature component.

Remarkably, we discuss two cases in Equation (4), i.e., when discussing the intra-class similarity of cover features, both P and Q in Equation (3) are f_i^c (i.e., illustrated in the upper part of Equation (4)), while the cross-entropy of P and Q , i.e., $H(P, Q) = H(f_i^c, f_i^c)$ is the intra-class similarity of cover features. Correspondingly, when discussing the intra-class similarity of stego features, both P and Q in Equation (3) are f_i^s (i.e., illustrated in the lower part of Equation (4)), while the cross-entropy of P and Q , i.e., $H(P, Q) = H(f_i^s, f_i^s)$ is the intra-class similarity of stego features.

$$\begin{cases} f_i^c \triangleq P \text{ and } f_i^c \triangleq Q, \text{ When } f_i^c \leftrightarrow f_i^c \\ f_i^s \triangleq P \text{ and } f_i^s \triangleq Q, \text{ When } f_i^s \leftrightarrow f_i^s \end{cases} \quad (4)$$

$$DIn(f_i) = H(f_i^c, f_i^c) + H(f_i^s, f_i^s) \quad (5)$$

where, f_i represents the i th feature component, $DIn(f_i)$ represents the intra-class similarity of f_i , $H(f_i^c, f_i^c)$ represents the cross-entropy within the cover features, $H(f_i^s, f_i^s)$ represents the cross-entropy within the stego features. Remarkably, the smaller the $DIn(f_i)$, the better the intra-class aggregation of f_i , i.e., the greater the contribution of f_i .

3.1.3. Construction of Inter-Class Similarity Criterion

Similarly, for the construction of the inter-class similarity criterion, this paper measures the similarity between the cover feature values and the stego feature values in the feature components. However, because it is apparent that cross-entropy does not satisfy the exchange law, i.e., $H(P, Q) \neq H(Q, P)$. Then this paper treats the cover image class and the stego image class in a single feature component as P and Q , respectively, as illustrated in Equation (6). Hence the inter-class similarity of a single feature component is measured using Equation (7).

Remarkably, we discuss two cases in Equation (6), i.e., when the cross-entropy of cover to stego is measured, with P and Q in Equation (3) being f_i^c and f_i^s , respectively (i.e., illustrated in the upper part of Equation (6)). Correspondingly, when the cross-entropy of stego to cover is measured, with P and Q in Equation (3) being f_i^s and f_i^c , respectively (i.e., illustrated in the lower part of Equation (6)):

$$\begin{cases} f_i^c \triangleq P \text{ and } f_i^s \triangleq Q, \text{ When } f_i^c \leftrightarrow f_i^s \\ f_i^s \triangleq P \text{ and } f_i^c \triangleq Q, \text{ When } f_i^s \leftrightarrow f_i^c \end{cases} \quad (6)$$

$$DIt(f_i) = H(f_i^c, f_i^s) + H(f_i^s, f_i^c) \quad (7)$$

where, f_i represents the i th feature component, $DIt(f_i)$ represents the inter-class similarity of f_i , $H(f_i^c, f_i^s)$ represents the cross-entropy of the cover features to the stego features, $H(f_i^s, f_i^c)$ represents the cross-entropy of the stego features to the cover features. Remarkably, the larger the $DIt(f_i)$, the better the inter-class dispersion of f_i , i.e., the greater the contribution of f_i .

3.1.4. Feature Contribution Metric

Based on the above intra-class similarity criterion and inter-class similarity criterion, to observe the contribution of individual features more easily, we tried to combine them together, we then devised a feature contribution measure criterion shaped akin to Fisher's criterion [29], whose formula is shown in Equation (8).

$$FCS(f_i) = \frac{\overleftarrow{DIt(f_i)}}{\overrightarrow{DIn(f_i)}} = \frac{\overleftarrow{H(f_i^c, f_i^s) + H(f_i^s, f_i^c)}}{\overrightarrow{H(f_i^c, f_i^c) + H(f_i^s, f_i^s)}} \quad (8)$$

$$\overleftarrow{DIt(f_i)} = \overleftarrow{H(f_i^c, f_i^s) + H(f_i^s, f_i^c)} = \overleftarrow{H(f_i^c, f_i^s)} + \overleftarrow{H(f_i^s, f_i^c)} \quad (9)$$

$$\overleftarrow{H}(f_i^c, f_i^s) = \sum_{j=1}^{M/2} \left(-\mu_1 f_{i,j}^c \ln(\mu_2 f_{i,j}^s) \right), \quad \overleftarrow{H}(f_i^s, f_i^c) = \sum_{j=1}^{M/2} \left(-\mu_1 f_{i,j}^s \ln(\mu_2 f_{i,j}^c) \right) \quad (10)$$

$$\overrightarrow{DIn}(f_i) = \overrightarrow{H}(f_i^c, f_i^c) + \overrightarrow{H}(f_i^s, f_i^s) = \overrightarrow{H}(f_i^c, f_i^c) + \overrightarrow{H}(f_i^s, f_i^s) \quad (11)$$

$$\overrightarrow{H}(f_i^c, f_i^c) = \sum_{j=1}^{M/2} \left(-\mu_3 f_{i,j}^c \ln(\mu_4 f_{i,j}^c) \right), \quad \overrightarrow{H}(f_i^s, f_i^s) = \sum_{j=1}^{M/2} \left(-\mu_3 f_{i,j}^s \ln(\mu_4 f_{i,j}^s) \right) \quad (12)$$

where, $FCS(f_i)$ represents the overall contribution of f_i , $\overleftarrow{DIt}(f_i)$ is a deformation of $DIt(f_i)$ standing for the inter-class similarity of f_i which can be calculated from Equations (9) and (10). $\overleftarrow{DIt}(f_i)$ is a deformation of $DIn(f_i)$ standing for the intra-class similarity of f_i which can be calculated from Equation (11) and Equation (12). Since we are only exploring the variation of $\frac{DIt(f_i)}{DIn(f_i)}$, we need to restrict the trend of the cross-entropy, for which we have introduced four parameters: μ_1, μ_2, μ_3 and μ_4 . For simplicity, we set μ_1, μ_2, μ_3 and μ_4 to take only the values 1 and -1 (in fact, they could take other values, but in this case, the simpler -1 and 1 are chosen because they are only used to change the trend) to control the reasonableness of $FCS(f_i)$. Yet, the specific deformation patterns and the results of the changes are discussed and analyzed in Section 3.2. Notably, the larger the $FCS(f_i)$, the more useful the f_i is for detecting stego images, i.e., the more it should be retained.

3.2. Parameter Setting

For the Equation (8) proposed in this paper, in order to satisfy the thought of “intra-class aggregation and inter-class dispersion”, we must set the size of the μ_1, μ_2, μ_3 and μ_4 parameters appropriately to ensure that the inter-class similarity does not conflict with the intra-class similarity, so as to better determine the extent to which this feature component contributes to the classification.

To this end, we first investigated the variation of the simpler monadic function $H(X, X) = -x_i \ln(x_i)$ in $[0, 1]$, which takes the shape of $H(f_i^c, f_i^c)$ in intra-class similarity, and it is easy to see that $H(X, X)$ shows an increasing and then decreasing trend within $[0, 1]$, with the extreme point coordinates being $(\frac{1}{e}, \frac{1}{e})$. Based on this statistic, and for simplicity of calculation, we intend to normalize the steganalytic feature values via Equation (13) to restrict them to between $[0, \frac{1}{e}]$, where $H(X, X)$ is monotonically increasing. From this, it is given that $\mu_3 = 1, \mu_4 = 1, \overrightarrow{H}(f_i^c, f_i^c) = H(f_i^c, f_i^c), \overrightarrow{H}(f_i^s, f_i^s) = H(f_i^s, f_i^s), \overrightarrow{DIn}(f_i) = DIn(f_i)$. Since $H(f_i^c, f_i^c)$ is monotonically increasing within $[0, \frac{1}{e}]$, $\overrightarrow{H}(f_i^c, f_i^c)$ is monotonically increasing, i.e., $\overrightarrow{DIn}(f_i)$ is monotonically increasing. Then we just need to make sure that $\overleftarrow{DIt}(f_i)$ is monotonically decreasing in $[0, \frac{1}{e}]$ to perfectly satisfy the “intra-class aggregation, inter-class dispersion” principle.

$$f_{i,j}^{c/s} = \begin{cases} \frac{f_{i,j}^{c/s} - \min(F^c, F^s)}{e \times (\max(F^c, F^s) - \min(F^c, F^s))}, & \max(F^c, F^s) \neq \min(F^c, F^s) \\ \frac{f_{i,j}^{c/s} - \min(F^c, F^s)}{e}, & \max(F^c, F^s) = \min(F^c, F^s) \end{cases} \quad (13)$$

where, f_i represents the i th feature component, $f_{i,j}$ denotes the j th value of f_i taken.

Then, on the basis of the above normalization, we investigated the variation of the binary function $\overrightarrow{H}(X, Y) = -\mu_1 x_i \ln(\mu_2 y_i)$ in $[0, \frac{1}{e}]$. Since μ_1 and μ_2 each take two values, there are four cases of $\overrightarrow{H}(X, Y)$, but since $\overrightarrow{H}(X, Y)$ is meaningless when $\mu_2 = -1$, only two

cases exist. Subsequently, we set the change step for both X and Y to 0.01 and plotted Figure 1 using the meshgrid(), the subplot() and the surf() in Matlab 2016.

Figure 1. Variation of binary cross-entropy when μ_1 and μ_2 take different values. (a–c) represent the variation of $\overrightarrow{H(X,Y)}$ at different viewpoints when $\mu_1 = -1$ and $\mu_2 = 1$, respectively. (d–f) represent the variation of $\overrightarrow{H(X,Y)}$ at different viewpoints when $\mu_1 = 1$ and, $\mu_2 = 1$ respectively.

From Figure 1, when $\mu_1 = -1$ and $\mu_2 = 1$, the larger X is, the larger $\overrightarrow{H(X,Y)}$ is, when Y is certain. In addition, the larger Y is, the smaller $\overrightarrow{H(X,Y)}$ is, when X is certain. Yet when X and Y increase simultaneously, $\overrightarrow{H(X,Y)}$ is not monotonically decreasing, i.e., it does not satisfy our objective. Then when $\mu_1 = 1$ and $\mu_2 = 1$, the larger X is, the smaller $\overrightarrow{H(X,Y)}$ is, when Y is certain. In addition, the larger Y is, the larger $\overrightarrow{H(X,Y)}$ is, when X is certain. Yet when X and Y increase simultaneously, $\overrightarrow{H(X,Y)}$ is not monotonically decreasing, i.e., such a case again does not satisfy our objective.

To this end, after reviewing the literature, we found that one could try to replace $\ln(x)$ with e^{-x} [39] (From Figure 2, it can be seen that $\ln(x)$ and e^{-x} vary in opposite trends within $[0, \frac{1}{e}]$), which led to the concept of similar cross-entropy, as shown in Equation (14). Correspondingly, Equations (8) and (10) can be transformed into Equations (15) and (16), respectively.

$$SH(P, Q) = -\sum_{i=1}^n p_i e^{-q_i} \quad (14)$$

$$\overleftarrow{FCS}(f_i) = \frac{\overleftarrow{SH}(f_i^c, f_i^s) + \overleftarrow{SH}(f_i^s, f_i^c)}{\overleftarrow{SH}(f_i^c, f_i^c) + \overleftarrow{SH}(f_i^s, f_i^s)} = \frac{\overleftarrow{SH}(f_i^c, f_i^c) + \overleftarrow{SH}(f_i^s, f_i^s)}{\overleftarrow{SH}(f_i^c, f_i^c) + \overleftarrow{SH}(f_i^s, f_i^s)} \quad (15)$$

$$\overleftarrow{SH}(f_i^c, f_i^s) = \sum_{j=1}^{M/2} \left(-\mu_1 f_{i,j}^c e^{-\mu_2 f_{i,j}^s} \right), \quad \overleftarrow{SH}(f_i^s, f_i^c) = \sum_{j=1}^{M/2} \left(-\mu_1 f_{i,j}^s e^{-\mu_2 f_{i,j}^c} \right) \quad (16)$$

where, $SH(P, Q)$ represents the similar cross-entropy between P and Q . Since e^{-x} and $\ln(x)$ vary in opposite trends within $[0, \frac{1}{e}]$ (as shown in Figure 2), the smaller the $\overleftarrow{FCS}(f_i)$, the better the usefulness. Nevertheless, at this point we still need to determine the values of μ_1 and μ_2 , so we have investigated the variation of $\overleftarrow{SH}(X, Y) = -\mu_1 x_i e^{-\mu_2 y_i}$ within $[0, \frac{1}{e}]$.

Subsequently, we set the change step for both X and Y to 0.01 and plotted Figures 3 and 4 using the subplot() and the surf() in Matlab 2016B.

Figure 2. Variation of cross-entropy and similar cross-entropy within $[0, \frac{1}{e}]$ (a,b) represent the variation of $\ln(x)$ and e^{-x} , respectively. In addition (c,d) represent the variation of cross-entropy and similar cross-entropy, respectively.

Figure 3. Variation of binary similar cross-entropy when μ_1 and μ_2 take different values. (a–c) represent the variation of $\overline{SH}(X,Y)$ at different viewpoints when $\mu_1 = -1$ and $\mu_2 = -1$, respectively. (d–f) represent the variation of $\overline{SH}(X,Y)$ at different viewpoints when $\mu_1 = -1$ and $\mu_2 = 1$, respectively.

From Figures 3 and 4, when $\mu_1 = 1$, $\mu_2 = -1$ and X and Y increase simultaneously, $\overleftarrow{SH}(X,Y)$ is monotonically decreasing. This means that it thoroughly met the fundamental requirements of our design guidelines.

Overall, after investigating and analyzing the values of μ_1 , μ_2 , μ_3 and μ_4 , we have finally determined that $\mu_1 = 1$, $\mu_2 = -1$, $\mu_3 = 1$ and $\mu_4 = 1$. Therefore, Equation (15) can be transformed into Equation (17).

$$\begin{aligned} \overleftarrow{FCS}(f_i) &= \frac{\overleftarrow{SH}(f_i^c f_i^s) + \overleftarrow{SH}(f_i^s f_i^c)}{\overleftarrow{SH}(f_i^c f_i^c) + \overleftarrow{SH}(f_i^s f_i^s)} = \frac{\overleftarrow{SH}(f_i^c f_i^c) + \overleftarrow{SH}(f_i^s f_i^s)}{\overleftarrow{SH}(f_i^c f_i^c) + \overleftarrow{SH}(f_i^s f_i^s)} \\ &= \frac{\sum_{j=1}^{M/2} (-f_{i,j}^c e^{f_{i,j}^s}) + \sum_{j=1}^{M/2} (-f_{i,j}^s e^{f_{i,j}^c})}{\sum_{j=1}^{M/2} (-f_{i,j}^c e^{-f_{i,j}^c}) + \sum_{j=1}^{M/2} (-f_{i,j}^s e^{-f_{i,j}^s})} \end{aligned} \quad (17)$$

where, M represents the total number of cover/stego images, and $M/2$ represents the number of cover/stego images used for testing. Notably, the smaller the value of $\overleftarrow{FCS}(f_i)$, the greater the contribution of f_i , i.e., the more important it is to be selected.

Figure 4. Variation of binary similar cross-entropy when μ_1 and μ_2 take different values (a–c) represent the variation of $\overleftarrow{SH}(X,Y)$ at different viewpoints when $\mu_1 = 1$ and $\mu_2 = -1$, respectively. (d–f) represent the variation of $\overleftarrow{SH}(X,Y)$ at different viewpoints when $\mu_1 = 1$ and $\mu_2 = 1$, respectively).

3.3. Overall Process and Performance Analysis

The FSCE algorithm principally consists of the following details. Firstly, normalize the eigenvalues to restrict their range to $[0, \frac{1}{e}]$. Secondly, calculate the similar cross-entropy of cover to stego and stego to cover, respectively, using Equation (16) and the determined μ_1 and μ_2 , and thus calculate the inter-class similarity $\overleftarrow{DI}(f_i)$. Then calculate the similar cross-entropy of cover to cover and stego to stego, respectively, using Equation (16) and the determined μ_3 and μ_4 , and thus calculate the intra-class similarity $\overleftarrow{DI}(f_i)$. Thirdly, calculate $\overleftarrow{FCS}(f_i)$ using Equation (17). Finally, rank the feature components according to $\overleftarrow{FCS}(f_i)$ in ascending order and select the smaller feature component of $\overleftarrow{FCS}(f_i)$ as the final selected feature.

To illustrate in more detail the working principle of the FSCE method, we give a specific algorithm based on the major steps outlined above, which is shown in Algorithm 1.

Analyzing Algorithm 1, it can be seen that the complexity of the FSCE algorithm depends mainly on the calculation process of $\overleftarrow{FCS}(f_i)$, and the time complexity of this part is $O(NM)$, where N is the total dimensions of the features and M represents the total number of cover/stego images. In contrast, the time complexity of the more popular classifier-based feature selection method is $O(mLMN^2) + O(mLN^3)$ [24,25], where L is

the number of classifiers, M is the number of image samples in the training set, N is the dimensions of feature in the test set, and m is the number of cycles.

Algorithm 1: FSCE.

Input: Original steganalytic features $F = [f_1, f_2, \dots, f_i, \dots, f_{N-1}, f_N]$, and dimension selected Ds

Output: Final selected feature $F' = [f_1', f_2', \dots, f_i', \dots, f_{n-1}', f_n']$ and corresponding column number *columns*

- (1) **For** $i = 1 : N$ **do**
 - (2) **For** $j = 1 : M$ **do**
 - (3) Normalize the eigenvalues to within $[0, \frac{1}{e}]$ using Equation (13).
 - (4) **END For**
 - (5) **END For**
 - (6) Let $\overline{FCS}(f_i) = \text{zeros}(1, N)$.
 - (7) **For** $i = 1 : N$ **do**
 - (8) Calculate $\overleftarrow{SH}(f_i^e, f_i^e)$ and $\overleftarrow{SH}(f_i^e, f_i^e)$ using the Equation (16), u_1 and u_2 .
 - (9) Calculate $\overleftarrow{DIt}(f_i)$ using the Equation (9).
 - (10) Calculate $\overleftarrow{SH}(f_i^e, f_i^e)$ and $\overleftarrow{SH}(f_i^e, f_i^e)$ using the u_3 and u_4 .
 - (11) Calculate $\overleftarrow{DIn}(f_i)$ using the Equation (11).
 - (12) Calculate $\overline{FCS}(f_i)$ using the Equation (17).
 - (13) **END For**
 - (14) Arrange the steganalytic feature components in ascending order based on the $\overline{FCS}(f_i)$ values from Step(12) and acquire $\bar{F} = [\bar{f}_1, \bar{f}_2, \dots, \bar{f}_i, \dots, \bar{f}_{N-1}, \bar{f}_N]$.
 - (15) Set the dimensions to be selected Ds .
 - (16) **For** $i = 1 : N$ **do**
 - (17) Select the top Ds feature components as the final features based on the ranking results of Step(14).
 - (18) **END For**
 - (19) Return $F' = [f_1', f_2', \dots, f_i', \dots, f_{n-1}', f_n']$ and *columns*.
-

The comparison shows that the time complexity $O(NM)$ of the FSCE method is significantly lower than the time complexity $O(mLMN^2) + O(mLN^3)$ of the integrated classifier-based feature selection method.

3.4. The Merits of FSCE

The merits of the FSCE method can be summarized as follows.

Firstly, an innovative improvement in the normalization range simplifies the algorithm design process. Compared to the traditional method of normalizing features to $[0, 1]$, normalizing to $[0, \frac{1}{e}]$ permits the $H(X, X) = -x_i \ln(x_i)$ to be monotonically increasing in this interval, simplifying the subsequent analysis and making it easier to determine μ_1, μ_2, μ_3 and μ_4 .

Secondly, cross-entropy, which is applicable to image steganalysis, is investigated for the first time and a feature contribution measure criterion is constructed, which is similar to Fisher's criterion. Based on the advantage that cross-entropy can determine the difference information of two classes, we classify the models of different cases and then propose the intra-class similarity criterion and inter-class similarity criterion. Based on which we further propose the feature contribution measure criterion with reference to the design principle of Fisher's criterion.

Thirdly, the concept of similar cross-entropy is theoretically proposed and proved, based on which the complexity of the calculation is considerably reduced. In determining the values of the parameters, we analyzed the variation of the feature contribution measure under different situations and found that it did not meet the original intention of our

design. For this reason, after searching for a large amount of literature, we proposed the concept of similar cross-entropy and updated the feature contribution measure criterion based on it. The analysis revealed that the new feature contribution measure criterion not only meets the requirements of the design but also decreases the computational complexity by changing the logarithmic operation into an exponential operation.

Fourth, FSCE is more general. Through a series of experiments in Section 4, it is found that the FSCE method is relatively effective in selecting many steganalytic features, attaining the goal of reducing the dimensions of feature by 40% while maintaining or even improving the detection accuracy of stego images.

Finally, the FSCE method has a low time complexity. With the performance analysis in Section 3.3, we find that the comparison demonstrates that the time complexity $O(NM)$ of the FSCE method is significantly lower than the time complexity $O(mLMN^2) + O(mLN^3)$ of the integrated classifier-based feature selection method. From this, one can find that FSCE is more efficient, which enables it to be used in time-critical applications.

4. Experiment

To verify the performance of the FSCE method proposed in this paper, in this section, we conduct experiments on the selection of different image steganalytic features. Specifically, Section 4.1 describes the experimental setup; Section 4.2 compares the features selected with different thresholds to determine the correctness of the final selection threshold in this paper; Section 4.3 compares the original features as well as randomly selected features to verify the effectiveness of FSCE; And finally, Section 4.4 compares the features with several classical and state-of-the-art feature selection methods to verify the efficiency and generality of FSCE.

Remarkably, all experiments in this paper were performed based on Matlab 2016B. It deserves mentioning that all algorithms are executed on a PC with 4 Intel(R) Core (TM) i7-8700 @ 3.20GHz CPUs, 8Gb memory.

4.1. Experiment Setup

The images used in this paper are taken from the only recognized image library in the steganography and steganalysis, BOSSbase 1.01 (http://dde.binghamton.edu/download/ImageDB/BOSSbase_1.01.zip, accessed on 7 November 2014), which contains 10,000 grey-scale images. To acquire the image steganalytic features, we performed the following operations on downloaded images.

- (1) Set a specified quality factor QF and then transform the PGM images in Bossbase 1.01 into the JPEG images of a certain QF.
- (2) Set the embedding rate *Payload*, and then use the steganography algorithm to embed secret information into the JPEG images to acquire the stego images under the current *Payload*.
- (3) Based on the set QF and *Payload*, use the steganalysis algorithm to extract the corresponding steganalytic features for the cover/stego images.
- (4) Depending on the steganography algorithm, steganalysis algorithm, QF and *Payload* (whose specific settings are shown in Table 1), by repeating (1)–(3), we will eventually construct a steganography detection image library containing 80,000 cover images and 400,000 stego images, and acquire a library containing 8 different steganalytic features.

Meanwhile, we continued to train and test the sample data along with the fisher linear discriminant (FLD) integrated classifier [40] by selecting 5000 cover and stego images as the training set and the remaining 5000 cover and stego images as the testing set, and then calculated the detection accuracy using Equation (18).

$$\overline{P_A} = 1 - \overline{P_E} \quad (18)$$

$$\overline{P_E} = \min_{P_{FA}} [P_{FA} + P_{MD}(P_{FA})] / 2 \quad (19)$$

where, $\bar{P}_A$ denotes the average detection accuracy, $\bar{P}_E$ denotes the average detection error rate, P_{FA} denotes the false alarm rate, P_{MD} denotes the false positive rate. To ensure that the experimental results could be fair and reliable, we took the average detection accuracy of the ten-fold cross-check as the final result of this feature selection method.

The experiments in this paper consist of three main parts.

- (1) Comparison experiments with features selected under different thresholds
- (2) Comparison experiments with original features and randomly selected features
- (3) Comparison experiments with several classical and state-of-the-art feature selection methods

Table 1. Specific experimental setup.

Source	BOSSbase 1.01	Number of cover images	10,000
Size	512×512	Number of stego images	$10,000 \times 5$
Colour	Gray-scale	QF	95, 75
Formats	JPEG	Payload	0.1, 0.2, 0.3, 0.4, 0.5
Training images	5000 pairs	Steganography Algorithm	nsF5 [4], SI-UNIWARD [7], S-UNIWARD [7]
Testing images	5000 pairs	Steganalysis Algorithm	GFR [9], DCTR [12], CC-JRM [13], SRM [14]
Total $10,000 \times (5 + 1) \times 8 = 480,000$			

QF represents the quality factor. *Payload* represents the the embedding rate of the steganography algorithm.

4.2. Comparison Experiments with Features Selected under Different Thresholds

In order to determine the value of D_s in the FSCE method, in this subsection, we conducted a mass of experiments on the image steganalytic features extracted in Section 4.1, and then determined the relatively appropriate D_s based on the experimental results under different D_s . As for the range of D_s and the iteration step, after analyzing a large amount of literature, we found that most of the existing feature selection methods reduce the dimensionality to 70%, while a few feature selection methods can reduce the dimensionality of some features to about 50%. For example, the SRGS algorithm reduces the GFR feature to roughly 50%, the CGSM algorithm reduces the GFR feature to 65% and the CC-PEV feature to 67%. steganalysis- α reduces the J + SRM (union of SRMQ1 (SRM with the fixed quantization $q = 1c$) and CC-JRM) dimension to about 70%. To make the FSCE method more generalizable and effective for most of the steganalytic features, the range of D_s is specified between $0.5N$ and $0.7N$ in this paper, and the iteration step is $0.02N$. In fact, based on this strategy (setting the same dimensions selected for different Payloads of the same feature), the dominant feature can be selected more efficiently. At the same time, this helps to propose new feature extraction methods. In general, the results of feature selection under different D_s are shown in Table 2.

As can be seen from Table 2, the proposed FSCE achieves excellent selection results for different steganalytic features. For example, for the F_1 feature, when *Payload* = 0.1, the detection accuracy of $D_s = 0.64N$ is the highest, which is improved by 0.29% compared to the original feature. For the F_2 feature, when *Payload* = 0.2, the detection accuracy of $D_s = 0.56N$ is the highest, which is improved by 1.20% compared to the original feature. For the F_4 feature, when *Payload* = 0.1, the detection accuracy of $D_s = 0.54N$ is the highest, which is improved by 0.93% compared to the original feature. For the F_5 feature, when *Payload* = 0.1, the detection accuracy of $D_s = 0.68N$ is the highest, which is improved by 0.24% compared to the original feature. For the F_6 feature, when *Payload* = 0.5, the detection accuracy of $D_s = 0.62N$ is the highest, which is improved by 0.23% compared to the original feature.

In summary, based on a combination of the feature dimensions selected and the detection accuracy of the stego images, we found that FSCE was better at selecting many features when $D_s = 0.58N$. For example, for the F_2 feature, when *Payload* = 0.1, the detection accuracy of FSCE is improved by 0.86% compared to the original feature, and when *Payload* = 0.2, the detection accuracy of FSCE is also improved by 1.20%. For the F_4 feature, when *Payload* = 0.1, the detection accuracy of FSCE is improved by 0.92% compared to the original feature and when *Payload* = 0.3, the detection accuracy of FSCE is also improved by 0.80%. For the F_5 feature, when *Payload* = 0.2, the detection accuracy of FSCE is improved by 0.18% compared to the original feature. For the F_6 feature, when *Payload* = 0.5, the detection accuracy of FSCE is

improved by 0.22% compared to the original feature. For the F_7 feature, when $Payload = 0.3$, the detection accuracy of FSCE is improved by 0.24% compared to the origin-feature.

Table 2. Experimental results of FSCE at different values of D_s .

F	P	Q	Origin	0.5N	0.52N	0.54N	0.56N	0.58N	0.6N	0.62N	0.64N	0.66N	0.68N	0.7N
F ₁		D	8000	4000	4160	4320	4480	4640	4800	4960	5120	5280	5440	5600
	0.1	P _A	0.5239	0.5221	0.5213	0.5226	0.5258	0.5260	0.5253	0.5253	0.5268	0.5243	0.5259	0.5264
	0.2	P _A	0.5256	0.5226	0.5240	0.5242	0.5268	0.5286	0.5271	0.5286	0.5285	0.5287	0.5287	0.5277
	0.3	P _A	0.5407	0.5357	0.5372	0.5369	0.5385	0.5412	0.5414	0.5394	0.5400	0.5397	0.5393	0.5397
	0.4	P _A	0.5745	0.5666	0.5669	0.5678	0.5684	0.5727	0.5723	0.5719	0.5752	0.5730	0.5736	0.5732
	0.5	P _A	0.6300	0.6203	0.6237	0.6256	0.6280	0.6302	0.6298	0.6295	0.6297	0.6283	0.6299	0.6293
F ₂		D	8000	4000	4160	4320	4480	4640	4800	4960	5120	5280	5440	5600
	0.1	P _A	0.7209	0.7269	0.7284	0.7267	0.7294	0.7295	0.7284	0.7289	0.7286	0.7286	0.7286	0.7260
	0.2	P _A	0.7246	0.7343	0.7336	0.7317	0.7366	0.7366	0.7342	0.7328	0.7346	0.7316	0.7327	0.7326
	0.3	P _A	0.7338	0.7432	0.7426	0.7432	0.7442	0.7447	0.7437	0.7417	0.7421	0.7419	0.7406	0.7418
	0.4	P _A	0.7557	0.7616	0.7609	0.7622	0.7626	0.7631	0.7634	0.7622	0.7626	0.7622	0.7605	0.7614
	0.5	P _A	0.7864	0.7874	0.7901	0.7893	0.7926	0.7924	0.7908	0.7903	0.7911	0.7908	0.7902	0.7907
F ₃		D	8000	4000	4160	4320	4480	4640	4800	4960	5120	5280	5440	5600
	0.1	P _A	0.8496	0.8278	0.8311	0.8342	0.8421	0.8438	0.8489	0.8478	0.8473	0.8476	0.8469	0.8486
	0.2	P _A	0.9913	0.9789	0.9833	0.9856	0.9876	0.9892	0.9913	0.9907	0.9902	0.9907	0.9909	0.9908
	0.3	P _A	0.9993	0.9980	0.9987	0.999	0.9994	0.9993	0.9993	0.9994	0.9993	0.9994	0.9993	0.9993
	0.4	P _A	0.9998	—	—	—	—	—	—	—	—	—	—	—
	0.5	P _A	0.9999	—	—	—	—	—	—	—	—	—	—	—
F ₄		D	8000	4000	4160	4320	4480	4640	4800	4960	5120	5280	5440	5600
	0.1	P _A	0.7884	0.7972	0.7964	0.7977	0.7965	0.7976	0.7954	0.7964	0.7963	0.7965	0.7964	0.7951
	0.2	P _A	0.9607	0.9648	0.9649	0.9652	0.9656	0.9655	0.9652	0.9652	0.9653	0.9654	0.9644	0.9646
	0.3	P _A	0.9939	0.9945	0.9946	0.9948	0.9947	0.9947	0.9948	0.9948	0.9947	0.9945	0.9945	0.9945
	0.4	P _A	0.9998	0.9981	0.9982	0.9983	0.9983	0.9983	0.9982	0.9983	0.9983	0.9982	0.9983	0.9983
	0.5	P _A	0.9991	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9991	0.9991	0.9992	0.9992	0.9992
F ₅		D	17,000	8500	8840	9180	9520	9860	10,200	10,540	10,880	11,220	11,560	11,900
	0.1	P _A	0.5168	0.5155	0.517	0.5165	0.5173	0.5174	0.517	0.518	0.5166	0.5187	0.5192	0.5151
	0.2	P _A	0.5205	0.5199	0.5208	0.5218	0.5215	0.5223	0.5222	0.5209	0.5227	0.5229	0.5222	0.5201
	0.3	P _A	0.5388	0.5365	0.5370	0.5367	0.5380	0.5389	0.5381	0.5383	0.5384	0.5385	0.5375	0.5388
	0.4	P _A	0.5738	0.5732	0.5738	0.5738	0.5741	0.5754	0.5758	0.5744	0.5747	0.5745	0.5748	0.5761
	0.5	P _A	0.6292	0.6268	0.6275	0.6290	0.6275	0.6294	0.6287	0.6275	0.6280	0.6290	0.6288	0.6283
F ₆		D	17,000	8500	8840	9180	9520	9860	10,200	10,540	10,880	11,220	11,560	11,900
	0.1	P _A	0.5035	0.5051	0.5034	0.5045	0.5054	0.5056	0.5042	0.504	0.5036	0.5052	0.5044	0.5033
	0.2	P _A	0.5245	0.5242	0.5237	0.5259	0.5252	0.5257	0.5251	0.5248	0.5241	0.5231	0.5254	0.5244
	0.3	P _A	0.5603	0.5610	0.5612	0.5593	0.5602	0.5614	0.5616	0.5616	0.5611	0.5604	0.5612	0.5602
	0.4	P _A	0.6125	0.6120	0.6130	0.6126	0.6108	0.6139	0.6134	0.6145	0.6116	0.6134	0.6114	0.6110
	0.5	P _A	0.6752	0.6735	0.6739	0.6739	0.6753	0.6774	0.6758	0.6775	0.6758	0.6754	0.6742	0.6752

Table 2. Cont.

F	P	Q	Origin	0.5N	0.52N	0.54N	0.56N	0.58N	0.6N	0.62N	0.64N	0.66N	0.68N	0.7N
F ₇		D	22,510	11,255	11,705	12,155	12,605	13,055	13,506	13,956	14,406	14,856	15,306	15,757
	0.1	P _A	0.529	0.5288	0.5295	0.5305	0.531	0.5305	0.5298	0.5302	0.5307	0.5303	0.5304	0.5303
	0.2	P _A	0.5345	0.5350	0.5346	0.5344	0.5346	0.5357	0.5358	0.5358	0.5364	0.536	0.5356	0.536
	0.3	P _A	0.538	0.5398	0.5406	0.5400	0.5397	0.5404	0.5390	0.5404	0.5407	0.5398	0.5401	0.5394
	0.4	P _A	0.5475	0.5467	0.5476	0.5463	0.5485	0.5484	0.5473	0.5483	0.5485	0.5472	0.5475	0.5466
	0.5	P _A	0.5705	0.5688	0.5693	0.5707	0.5715	0.5724	0.569	0.5710	0.5721	0.5703	0.5711	0.5715
F ₈		D	34,671	17,335	18,028	18,722	19,415	20,109	20,802	21,496	22,189	22,882	23,576	24,269
	0.1	P _A	0.5988	0.5957	0.5957	0.5972	0.5974	0.5982	0.5934	0.5967	0.5979	0.5986	0.598	0.5991
	0.2	P _A	0.6795	0.6799	0.6788	0.6802	0.6800	0.6802	0.6788	0.6797	0.6801	0.6792	0.6804	0.6807
	0.3	P _A	0.745	0.7447	0.7440	0.7444	0.7439	0.7451	0.7457	0.7438	0.7444	0.7448	0.7458	0.7457
	0.4	P _A	0.7938	0.7935	0.7935	0.7937	0.7944	0.7947	0.7944	0.7947	0.7943	0.7935	0.7942	0.7943
	0.5	P _A	0.8375	0.8365	0.8369	0.8373	0.8386	0.8387	0.8375	0.8379	0.8381	0.8378	0.8386	0.8374

F₁~F₈ represent the SI-UNIWAD-DCTR-95, SI-UNIWAD-DCTR-75, nsF5-DCTR-95, nsF5-DCTR-75, SI-UNIWAD-GFR-95, SI-UNIWAD-GFR-75, SI-UNIWAD-CC-JRM-95 and S-UNIWAD-SRM-75 feature, respectively. *N* represents the total dimensionality of the original feature. *P* represents the *Payload*. *D* represents the feature dimension selected. Remarkably, the darker the background colour, the higher the detection accuracy at this point. “—” means that the detection accuracy of this original feature is already very high and no further feature selection is needed for it.

4.3. Comparison Experiments with Original Features and Randomly Selected Features

To verify the effectiveness of the FSCE, in this subsection we compare the features selected by the FSCE with the original features and the randomly selected features. Notably, the dimensionality of the “randomly selected features” is the same as the dimensionality of the features selected by the FSCE, so as to demonstrate the effectiveness of the FSCE. The results of the experiments are shown in Table 3.

As can be seen from the table, the FSCE-feature has the best performance compared to the Random-feature and Origin-feature. Specifically, FSCE reduces the dimensionality by 42% while maintaining or even improving the detection accuracy. For example, for the F₁ feature, when *Payload* = 0.5, FSCE improved the detection accuracy by 0.94% compared to the Random-feature. For the F₂ feature, when *Payload* = 0.2, FSCE improved the detection accuracy by 3.85% compared to the Random-feature and by 1.20% compared to the Origin-feature. For the F₃ feature, when *Payload* = 0.1, FSCE improved the detection accuracy by 5.91% compared to the Random-feature. For the F₄ feature, when *Payload* = 0.1, FSCE improved the detection accuracy by 4.65% compared to the Random-feature and by 0.92% compared to the Origin-feature. For the F₅ feature, when *Payload* = 0.4, FSCE improved the detection accuracy by 0.65% compared to the Random-feature and by 0.16% compared to the Origin-feature. For the F₆ feature, when *Payload* = 0.5, FSCE improved the detection accuracy by 0.82% compared to the Random-feature and by 0.22% compared to the Origin-feature. For the F₇ feature, when *Payload* = 0.3, FSCE improved the detection accuracy by 0.48% compared to the Random-feature and by 0.24% compared to the Origin-feature. For the F₈ feature, when *Payload* = 0.4, FSCE improved the detection accuracy by 0.40% compared to the Random-feature.

For other *Payloads*, FSCE also achieved excellent results compared to Random-feature and Origin-feature, thus verifying the effectiveness of FSCE.

Table 3. Comparison results of FSCE-feature with Origin-feature and Random-feature.

<i>Payload</i>	<i>Feature Quality</i>	<i>D</i>	<i>F₁</i> <i>P_A</i>	<i>F₂</i> <i>P_A</i>	<i>F₃</i> <i>P_A</i>	<i>F₄</i> <i>P_A</i>	<i>D</i>	<i>F₅</i> <i>P_A</i>	<i>F₆</i> <i>P_A</i>	<i>D</i>	<i>F₇</i> <i>P_A</i>	<i>D</i>	<i>F₈</i> <i>P_A</i>
0.1	<i>Origin</i>	8000	0.5239	0.7209	0.8496	0.7884	17,000	0.5168	0.5035	22,510	0.5290	34,671	0.5988
	<i>Random</i>	4640	0.5232	0.6920	0.7847	0.7511	9860	0.5151	0.5021	13,055	0.5270	20,109	0.5969
	<i>FSCE</i>	4640	0.5260	0.7295	0.8438	0.7976	9860	0.5174	0.5056	13,055	0.5305	20,109	0.5989
0.2	<i>Origin</i>	8000	0.5256	0.7246	0.9913	0.9607	17,000	0.5205	0.5245	22,510	0.5345	34,671	0.6795
	<i>Random</i>	4640	0.5240	0.6981	0.9681	0.9395	9860	0.5180	0.5218	13,055	0.5311	20,109	0.6764
	<i>FSCE</i>	4640	0.5286	0.7366	0.9892	0.9655	9860	0.5223	0.5257	13,055	0.5357	20,109	0.6802
0.3	<i>Origin</i>	8000	0.5407	0.7338	0.9993	0.9939	17,000	0.5388	0.5603	22,510	0.5380	34,671	0.7450
	<i>Random</i>	4640	0.5367	0.7074	0.9979	0.9907	9860	0.5346	0.5565	13,055	0.5356	20,109	0.7413
	<i>FSCE</i>	4640	0.5412	0.7447	0.9993	0.9947	9860	0.5389	0.5614	13,055	0.5404	20,109	0.7451
0.4	<i>Origin</i>	8000	0.5745	0.7557	0.9998	0.9998	17,000	0.5738	0.6125	22,510	0.5475	34,671	0.7938
	<i>Random</i>	4640	0.5674	0.7298	—	0.9974	9860	0.5689	0.6067	13,055	0.5442	20,109	0.7907
	<i>FSCE</i>	4640	0.5744	0.7631	—	0.9993	9860	0.5754	0.6139	13,055	0.5484	20,109	0.7947
0.5	<i>Origin</i>	8000	0.6300	0.7864	0.9999	0.9991	17,000	0.6292	0.6752	22,510	0.5705	34,671	0.8375
	<i>Random</i>	4640	0.6208	0.7626	—	0.9989	9860	0.6182	0.6692	13,055	0.5665	20,109	0.8354
	<i>FSCE</i>	4640	0.6302	0.7924	—	0.9992	9860	0.6294	0.6774	13,055	0.5724	20,109	0.8387

“Bolded” indicates the highest detection accuracy in the current situation. Remarkably, F_1 – F_4 are placed in a box because they have the same feature dimension. F_5 – F_6 are similar. Significantly, the meanings of the abbreviations in this table are consistent with those in Table 2.

4.4. Comparison Experiments with Several Classical and State-of-the-Art Feature Selection Methods

To validate the efficiency and generality of FSCE, we conducted comparison experiments with PCA-based method [27], SRGS method [26] and CGSM method [28], where PCA-based method is more classical, SRGS is a more novel method for specific feature selection and CGSM is a more novel method for general feature selection. The results of the comparison between the algorithm proposed in this paper and the above three methods are shown in Table 4.

From Table 4, it is clear that the performance of the proposed FSCE method is superior compared to the PCA-based method, SRGS method and CGSM method. Specifically, for example, for the F_1 feature, when $Payload = 0.5$, the detection accuracy of FSCE is further improved by 13.00% compared to PCA-feature, and by 1.32% compared to SRGS-feature, while FSCE further reduces the dimensionality by 2363-D (about 29.54%) on SRGS-feature. In addition, a further improvement of 0.22% in detection accuracy compared to CGSM-feature. For the F_2 feature, when $Payload = 0.5$, the detection accuracy of FSCE is further improved by 29.19% compared to PCA-feature, 4.41% compared to SRGS-feature and 1.09% compared to CGSM-feature, while the dimensionality is further reduced by 1109-D (about 13.86%). For the F_3 feature, when $Payload = 0.2$, the detection accuracy of FSCE is further improved by 48.77% compared to PCA-feature, 0.28% compared to SRGS-feature, and the dimensionality is further reduced by 1623-D (about 20.29%) compared to CGSM-feature. For the F_4 feature, when $Payload = 0.3$, the detection accuracy of FSCE is further improved by 49.29% compared to the PCA-feature, and the dimensionality is further reduced by 894-D (about 11.18%) compared to the SRGS-feature while maintaining comparable detection accuracy, and compared to the CGSM-feature a further improvement of 0.10%, while the dimensionality was further reduced by 2194-D (about 27.43%).

Table 4. Comparison results for FSCE, PCA-based, SRGS and CGSM.

Payload	Feature	F_1		F_2		F_3		F_4	
	Quality	D	P_A	D	P_A	D	P_A	D	P_A
0.1	Origin	8000	0.5239	8000	0.7209	8000	0.8496	8000	0.7884
	PCA	4640	0.5001	4640	0.5012	4640	0.5009	4640	0.5006
	SRGS	2150	0.5193	7542	0.7213	4534	0.8335	5499	0.7868
	CGSM	6999	0.5234	6850	0.7153	6810	0.8252	6999	0.7866
	FSCE	4640	0.5260	4640	0.7295	4640	0.8438	4640	0.7976
0.2	Origin	8000	0.5256	8000	0.7246	8000	0.9913	8000	0.9607
	PCA	4640	0.5000	4640	0.5008	4640	0.5015	4640	0.5013
	SRGS	4339	0.5293	7252	0.7271	4389	0.9864	5524	0.9617
	CGSM	6278	0.5239	6591	0.7184	6263	0.9895	6962	0.9605
	FSCE	4640	0.5286	4640	0.7366	4640	0.9892	4640	0.9655
0.3	Origin	8000	0.5407	8000	0.7338	8000	0.9993	8000	0.9939
	PCA	4640	0.5002	4640	0.5003	4640	0.5023	4640	0.5018
	SRGS	5913	0.5398	5056	0.7366	4420	0.9993	5534	0.9945
	CGSM	5671	0.5389	6356	0.7283	5068	0.9991	6834	0.9937
	FSCE	4640	0.5412	4640	0.7447	4640	0.9993	4640	0.9947
0.4	Origin	8000	0.5745	8000	0.7557	8000	0.9998	8000	0.9998
	PCA	4640	0.5005	4640	0.4997	—	—	4640	0.5023
	SRGS	7051	0.5681	2355	0.7302	—	—	5814	0.9982
	CGSM	4956	0.5720	6080	0.7513	—	—	6568	0.9979
	FSCE	4640	0.5744	4640	0.7631	—	—	4640	0.9993
0.5	Origin	8000	0.6300	8000	0.7864	8000	0.9999	8000	0.9991
	PCA	4640	0.5002	4640	0.5005	—	—	4640	0.5045
	SRGS	7003	0.6170	2323	0.7483	—	—	5473	0.9991
	CGSM	4570	0.6280	5749	0.7815	—	—	6105	0.9990
	FSCE	4640	0.6302	4640	0.7924	—	—	4640	0.9992

“Bolded” indicates the highest detection accuracy in the current situation. Significantly, the meanings of the abbreviations in this table are consistent with those in Table 2.

From the Table 5, for the F_5 feature, when $Payload = 0.5$, the detection accuracy of FSCE is further improved by 12.87% compared to PCA-feature, 0.19% compared to SRGS-feature, and further reduced by 3681-D (about 21.65%) compared to CGSM-feature while maintaining comparable detection accuracy. For the F_6 feature, when $Payload = 0.5$, the detection accuracy of the FSCE is further improved by 17.54% compared to the PCA-feature and by 0.25% compared to the SRGS-feature, while the dimensionality is further reduced by 6183-D (about 36.37%). For the F_7 feature, when $Payload = 0.5$, the detection accuracy of the FSCE is further improved by 7.18% compared to the PCA-feature and by 0.37% compared to the SRGS-feature, while the dimensionality is further reduced by 6482-D (about 38.13%). For the F_8 feature, when $Payload = 0.1$, the detection accuracy of FSCE is further improved by 9.76% compared to PCA-feature, 0.59% compared to SRGS-feature, while the dimensionality is further reduced by 2112-D (about 6.10%), and further improved compared to CGSM-feature by 0.18%, while the dimensionality is further reduced by 8046-D (about 23.21%). For other situations, FSCE has achieved equally excellent results.

Table 5. Comparison results for FSCE, PCA-based, SRGS and CGSM.

Payload	Feature Quality	F_5		F_6		F_7		F_8	
		D	P_A	D	P_A	D	P_A	D	P_A
0.1	Origin	17,000	0.5168	17,000	0.5035	22,510	0.5290	34,671	0.5988
	PCA	9860	0.5000	9860	0.5002	13,055	0.5003	20,109	0.5013
	SRGS	4508	0.5144	8437	0.5032	8713	0.5353	22,221	0.5930
	CGSM	15,794	0.5147	14,548	0.5032	15,244	0.5324	28,155	0.5971
	FSCE	9860	0.5174	9860	0.5056	13,055	0.5305	20,109	0.5989
0.2	Origin	17,000	0.5205	17,000	0.5245	22,510	0.5345	34,671	0.6795
	PCA	9860	0.5001	9860	0.5002	13,055	0.4998	20,109	0.5021
	SRGS	7054	0.5210	14,581	0.5243	12,423	0.5394	21,577	0.6685
	CGSM	15,363	0.5222	12,813	0.5243	13,143	0.5373	20,731	0.6740
	FSCE	9860	0.5223	9860	0.5257	13,055	0.5380	20,109	0.6802
0.3	Origin	17,000	0.5388	17,000	0.5603	22,510	0.5356	34,671	0.7450
	PCA	9860	0.5002	9860	0.5006	13,055	0.5000	20,109	0.5038
	SRGS	8208	0.5363	15,520	0.5607	14,350	0.5419	21,316	0.7318
	CGSM	14,813	0.5375	11,310	0.5605	11,330	0.5413	14,166	0.7274
	FSCE	9860	0.5389	9860	0.5614	13,055	0.5404	20,109	0.7451
0.4	Origin	17,000	0.5738	17,000	0.6125	22,510	0.5475	34,671	0.7938
	PCA	9860	0.5005	9860	0.5008	13,055	0.5005	20,109	0.5041
	SRGS	8578	0.5742	14,029	0.6109	17,813	0.5503	21,425	0.7825
	CGSM	14,121	0.5743	10,255	0.6088	9946	0.5490	9506	0.7608
	FSCE	9860	0.5754	9860	0.6139	13,055	0.5484	20,109	0.7947
0.5	Origin	17,000	0.6292	17,000	0.6752	22,510	0.5705	34,671	0.8375
	PCA	9860	0.5007	9860	0.5020	13,055	0.5006	20,109	0.5068
	SRGS	8580	0.6275	16,043	0.6749	19,537	0.5687	21,750	0.8239
	CGSM	13,541	0.6294	10,233	0.6766	8773	0.5706	7239	0.7954
	FSCE	9860	0.6294	9860	0.6774	13,055	0.5724	20,109	0.8387

“Bolded” indicates the highest detection accuracy in the current situation. Significantly, the meanings of the abbreviations in this table are consistent with those in Table 2.

Table 6 illustrates the feature selection time of the FSCE method in comparison with the PCA-based and the SRGS and CGSM method. From the table, it can be seen that the feature selection time of the FSCE method proposed in this paper is significantly shorter than that of the PCA-based, SRGS and CGSM method. Specifically, For F_1 features, when $Payload = 0.5$, FSCE takes only 2.93 s to select the features, a further reduction of 125.58 s compared to the PCA-based method, 765.21 s compared to the SRGS method, and 5.45 s compared to the CGSM method. For F_2 features, when $Payload = 0.1$, FSCE takes only 2.88 s to select the features, a further reduction of 124.22 s compared to the PCA-based method, 810.13 s compared to the SRGS method, and 5.37 s compared to the CGSM method. For

F_3 features, when *Payload* = 0.2, FSCE takes only 2.84 s to select the features, a further reduction of 125.55 s compared to the PCA-based method, 525.49 s compared to the SRGS method and 5.81 s compared to the CGSM method. For F_4 features, when *Payload* = 0.3, FSCE takes only 2.80 s to select the features, a further reduction of 124.24 s compared to the PCA-based method, 681.51 s compared to the SRGS method and 5.54 s compared to the CGSM method. For F_5 features, when *Payload* = 0.5, FSCE takes only 6.22 s to select the features, a further reduction of 264.54 s compared to the PCA-based method, 1698.08 s compared to the SRGS method and 22.75 s compared to the CGSM method. For F_6 features, when *Payload* = 0.4, FSCE takes only 6.38 s to select the features, a further reduction of 264.59 s compared to the PCA-based method, 1863.42 s compared to the SRGS method and 29.47 s compared to the CGSM method. For F_7 features, when *Payload* = 0.5, FSCE takes only 8.30 s to select the features, a further reduction of 301.99 s compared to the PCA-based method, 2486.4 s compared to the SRGS method and 95.42 s compared to the CGSM method. For F_8 features, when *Payload* = 0.2, FSCE takes only 17.44 s to select the features, a further reduction of 391.16 s compared to the PCA-based method, 2568.86 s compared to the SRGS method and 17.60 s compared to the CGSM method. For other situations, FSCE has achieved equally excellent results.

Table 6. Comparison of feature selection time for FSCE, PCA-based, SRGS and CGSM.

<i>P</i>	Feature Method	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8
Time of Selecting Feature (s)									
0.1	PCA	127.49	127.1	127.87	127.19	272.17	274.07	314.23	408.6
	SRGS	271.17	813.01	542.54	663.17	1117.07	1002.79	1152.21	2586.29
	CGSM	8.65	8.27	9.12	10.40	28.85	37.96	77.47	35.04
	FSCE	2.95	2.88	2.81	2.95	6.22	5.89	8.12	17.44
0.2	PCA	127.51	126.33	128.39	127.63	272.66	275.65	310.81	414.14
	SRGS	529.64	721.53	528.33	651.4	1453.06	1537.18	1507.82	2537.18
	CGSM	8.27	8.19	8.65	10.33	22.05	34.45	79.38	37.91
	FSCE	2.96	2.89	2.84	2.90	6.20	6.38	7.93	17.83
0.3	PCA	127.74	126.54	127.60	127.04	272.81	271.10	315.25	415.16
	SRGS	697.29	600.23	519.90	684.31	1537.49	1790.01	1726.11	2529.80
	CGSM	9.83	8.31	8.63	8.34	28.84	35.15	79.94	37.09
	FSCE	2.92	2.81	2.84	2.80	6.28	6.36	7.93	23.75
0.4	PCA	128.59	127.06	127.71	126.75	265.68	270.97	311.63	407.24
	SRGS	723.55	283.77	525.92	639.65	1668.66	1869.75	2226.19	2525.94
	CGSM	8.35	8.27	8.18	8.78	22.07	35.81	97.26	98.65
	FSCE	2.95	2.81	2.87	2.98	6.20	6.38	7.88	15.17
0.5	PCA	128.51	127.74	127.27	126.91	270.76	272.81	310.29	415.56
	SRGS	768.14	297.59	531.77	656.5	1704.28	1824.51	2494.69	2570.46
	CGSM	8.38	9.50	8.08	12.52	28.97	35.49	103.72	117.89
	FSCE	2.93	2.82	2.83	2.87	6.22	6.37	8.30	15.07

“Bolded” indicates the shorter feature selection time in the current situation. Significantly, the meanings of the abbreviations in this table are consistent with those in Table 2.

5. Conclusions

To decrease the feature dimensions and the spatio-temporal overhead, this paper presents a fast selection method for image steganalytic features based on similar cross-entropy (FSCE). Specifically, firstly, the innovative improvement of the normalization range simplifies the analysis of the changing trend of binary information entropy, and lays the foundation for the overall algorithm design process. Secondly, the cross-entropy applicable to image steganalysis is investigated, and a feature contribution measure criterion shaped akin to Fisher’s criterion is constructed. Thirdly, after analyzing the constructed feature contribution criterion, it is found that the criterion has certain shortcomings. For this reason, after reviewing a large amount of literature, the concept of similar cross-entropy is

presented for the first time and after analyzing its change trend, the feature contribution criterion is innovated and its reliability is verified theoretically. As a result, the contribution degree of each feature component is better measured. Finally, the feature component with the highest contribution is selected as the final selected feature. Based on the above operations, FSCE has exceptional usability, which facilitates its use in real-world applications with strict memory footprint constraints and high-efficiency requirements.

The effectiveness and efficiency of FSCE have been demonstrated through extensive experiments on the only standard and widely utilized Bossbase 1.01 image library. For example, for the F_2 feature, when *Payload* = 0.5, the detection accuracy of the FSCE is further improved by 4.41% compared to the SRGS-feature. For the F_3 feature, when *Payload* = 0.1, the detection accuracy of the FSCE is improved by 5.91% compared to the Random-feature. For the F_4 feature, when *Payload* = 0.3, the detection accuracy of FSCE is further improved by 49.29% compared to PCA-feature. For the F_8 feature, when *Payload* = 0.5, the detection accuracy of FSCE is further improved by 4.33% compared to CGSM-feature.

For the future, we will continue to devote ourselves to steganography and steganalysis, focusing on two aspects: On the one hand, analyzing the properties of each deleted feature component lays the foundation for new, more secure steganography techniques. On the other hand, investigating the characteristics of each retained feature component lays the foundation for new effective, efficient and low-dimensional steganalysis techniques.

Author Contributions: Conceptualization, R.J. and X.Y.; methodology, X.Y. and Y.M.; software, validation, X.Y.; formal analysis, Y.M.; writing—original draft preparation, X.Y.; writing—review and editing, X.Y. and S.Y.; visualization, X.Y. and L.X.; project administration, X.Y. All authors have read and agreed to the published version of the manuscript.

Funding: Please add: This work was supported in part by the National Natural Science Foundation of China under Grant 62002103, in part by the Key Scientific and Technological Project of Henan Province under Grant 202102210165, in part by the Promotion Special (Soft Science) Project of Henan Province under Grant 202400410088 and in part by the Key Scientific Research (Soft Science) Project of Higher Education Institutions of Henan Province under Grant 19A880030.

Institutional Review Board Statement: Not Applicable.

Informed Consent Statement: Not Applicable.

Data Availability Statement: The images used in this paper are taken from the Bossbase 1.01 image library: http://dde.binghamton.edu/download/ImageDB/BOSSbase_1.01.zip (accessed on 7 November 2014).

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

References

1. Zhou, W.; Zhang, W.; Yu, N. A New Rule for Cost Reassignment in Adaptive Steganography. *IEEE Trans. Inf. Forensics Secur.* **2017**, *12*, 2654–2667. [CrossRef]
2. Denemark, T.; Fridrich, J. Steganography with Multiple JPEG Images of the Same Scene. *IEEE Trans. Inf. Forensics Secur.* **2017**, *12*, 2308–2319. [CrossRef]
3. Tomáš, F.; Judas, J.; Fridrich, J. Minimizing Additive Distortion in Steganography using Syndrome-Trellis Codes. *IEEE Trans. Inf. Forensics Secur.* **2011**, *6*, 920–935.
4. Fridrich, J.; Tomáš, P.; Jan, K. Statistically Undetectable JPEG Steganography: Dead Ends, Challenges, and Opportunities. In Proceedings of the 9th Workshop on Multimedia & Security, Dallas, TX, USA, 20–21 September 2007; pp. 3–14.
5. Filler, T.; Fridrich, J. Gibbs Construction in Steganography. *IEEE Trans. Inf. Forensics Secur.* **2010**, *5*, 705–720. [CrossRef]
6. Holub, V.; Fridrich, J. Designing Steganographic Distortion using Directional Filters. In Proceedings of the 2012 IEEE International Workshop on Information Forensics and Security, Costa Adeje-Tenerife, Spain, 2–5 December 2012; pp. 234–239.
7. Vojtěch, H.; Fridrich, J.; Denemark, T. Universal Distortion Function for Steganography in an Arbitrary Domain. *EURASIP J. Inf. Secur.* **2014**, *1*, 1–13.
8. Tomáš, P.; Bas, P.; Fridrich, J. Steganalysis by Subtractive Pixel Adjacency Matrix. *IEEE Trans. Inf. Forensics Secur.* **2010**, *5*, 215–224.
9. Song, X.; Liu, F.; Zhang, Z.; Yang, C.; Luo, X.; Chen, L. 2D Gabor Filters-Based Steganalysis of Content-Adaptive JPEG Steganography. *Multimed. Tools Appl.* **2016**, *76*, 26391–26419. [CrossRef]
10. Tomas, P.; Fridrich, J. Merging Markov and DCT Features for Multi-Class JPEG Steganalysis. In Proceedings of the Security, Steganography, and Watermarking of Multimedia Contents IX, San Jose, CA, USA, 28 January 2007; pp. 650501–650503.

11. Jan, K.; Fridrich, J. Calibration Revisited. In Proceedings of the 11th ACM Workshop on Multimedia and Security, New York, NY, USA, 7–8 September 2009; pp. 63–74.
12. Holub, V.; Fridrich, J. Low-Complexity Features for JPEG Steganalysis using Undecimated DCT. *IEEE Trans. Inf. Forensics Secur.* **2015**, *10*, 219–228. [[CrossRef](#)]
13. Kodovský, J.; Fridrich, J. Steganalysis of JPEG Images using Rich Models. In Proceedings of the Media Watermarking, Security, and Forensics, Burlingame, CA, USA, 23–25 January 2012; pp. 83030A-1–83030A-13.
14. Fridrich, J.; Kodovsky, J. Rich Models for Steganalysis of Digital Images. *IEEE Trans. Inf. Forensics Secur.* **2012**, *7*, 868–882. [[CrossRef](#)]
15. Ghasemzadeh, H. Calibrated Steganalysis of Mp3stego in Multi-Encoder Scenario. *Inf. Sci.* **2019**, *480*, 438–453. [[CrossRef](#)]
16. Zhang, W.; Zhang, Z.; Zhang, L.; Li, H.; Yu, N. Decomposing Joint Distortion for Adaptive Steganography. *IEEE Trans. Circuits Syst. Video Technol.* **2017**, *27*, 2274–2280. [[CrossRef](#)]
17. Denemark, T.; Boroumand, M.; Fridrich, J. Steganalytic Features for Content-Adaptive JPEG Steganography. *IEEE Trans. Inf. Forensics Secur.* **2016**, *11*, 1736–1746. [[CrossRef](#)]
18. Zhang, R.; Zhu, F.; Liu, J.; Liu, G. Depth-Wise Separable Convolutions and Multi-Level Pooling for an Efficient Spatial CNN-Based Steganalysis. *IEEE Trans. Inf. Forensics Secur.* **2020**, *15*, 1138–1150. [[CrossRef](#)]
19. Ye, J.; Ni, J.; Yi, Y. Deep Learning Hierarchical Representations for Image Steganalysis. *IEEE Trans. Inf. Forensics Secur.* **2017**, *12*, 2545–2557. [[CrossRef](#)]
20. Sedighi, V.; Fridrich, J. Histogram layer, Moving Convolutional Neural Networks Towards Feature-Based Steganalysis. *Electron. Imaging* **2017**, *7*, 50–55. [[CrossRef](#)]
21. Holub, V.; Fridrich, J.; Denemark, T. Random Projections of Residuals as an Alternative to Co-Occurrences in Steganalysis. In Proceedings of the Media Watermarking, Security, and Forensics 2013, Burlingame, CA, USA, 22 March 2013; p. 86650L.
22. Holub, V.; Fridrich, J. Random Projections of Residuals for Digital Image Steganalysis. *IEEE Trans. Inf. Forensics Secur.* **2013**, *8*, 1996–2006. [[CrossRef](#)]
23. Boroumand, M.; Fridrich, J. Applications of Explicit Non-Linear Feature Maps in Steganalysis. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 823–833. [[CrossRef](#)]
24. Ma, Y.; Luo, X.; Li, X.; Bao, Z.; Zhang, Y. Selection of Rich Model Steganalysis Features Based on Decision Rough Set α -Positive Region Reduction. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *29*, 336–350. [[CrossRef](#)]
25. Yang, C.; Zhang, Y.; Wang, P.; Luo, X.; Liu, F.; Lu, J. Steganalysis Feature Subspace Selection Based on Fisher Criterion. In Proceedings of the 2017 IEEE International Conference on Data Science and Advanced Analytics, Tokyo, Japan, 19–21 October 2017; pp. 514–521.
26. Yu, X.; Ma, Y.; Jin, R.; Xu, L.; Duan, X. A Multi-Scale Feature Selection Method for Steganalytic Feature GFR. *IEEE Access* **2020**, *8*, 55063–55075. [[CrossRef](#)]
27. Qin, J.; Sun, X.; Xiang, X.; Niu, C. Principal Feature Selection and Fusion Method for Image Steganalysis. *J. Electron. Imaging* **2009**, *18*, 033009.
28. Wang, Y.; Ma, Y.; Jin, R.; Liu, P.; Ruan, N. Comprehensive Criteria-Based Generalized Steganalysis Feature Selection Method. *IEEE Access* **2020**, *8*, 154418–154435. [[CrossRef](#)]
29. Lu, J.; Liu, F.; Luo, X. Selection of Image Features for Steganalysis Based on the Fisher Criterion. *Digit. Investig.* **2014**, *11*, 57–66. [[CrossRef](#)]
30. Ma, Y.; Xu, J.; Zhang, Y.; Yang, C.; Luo, X. W2ID Criterion-Based Rich Model Steganalysis Features Selection. *Chin. J. Comput.* **2021**, *44*, 724–740.
31. Chen, Y.; Chen, Y.; Yin, A. Feature Selection for Blind Image Steganalysis using Neighborhood Rough Sets. *J. Intell. Fuzzy Syst.* **2019**, *37*, 3709–3720. [[CrossRef](#)]
32. Davidson, J.; Jalan, J. Feature Selection for Steganalysis using the Mahalanobis Distance. In Proceedings of the Media Forensics and Security II, San Jose, CA, USA, 18–20 January 2010; p. 754104.
33. Boroumand, M.; Chen, M.; Fridrich, J. Deep Residual Network for Steganalysis of Digital Images. *IEEE Trans. Inf. Forensics Secur.* **2019**, *14*, 1181–1193. [[CrossRef](#)]
34. Denemark, T.; Holub, V.; Cogranne, R.; Fridrich, J. Selection-Channel-Aware Rich Model for Steganalysis of Digital Images. In Proceedings of the IEEE International Workshop on Information Forensics and Security 2014, Atlanta, GA, USA, 3–5 December 2014; pp. 48–53.
35. Yuan, H.; Li, J.; Lai, L.; Tang, Y. Low-Rank Matrix Regression for Image Feature Extraction and Feature Selection. *Inf. Sci.* **2020**, *522*, 214–226. [[CrossRef](#)]
36. Patrick, B.; Tomáš, F.; Tomáš, P. “Break Our Steganographic System”: The Ins and Outs of Organizing BOSS. In Proceedings of the International Workshop on Information Hiding, Berlin, Germany, 18–20 May 2011; pp. 59–70.
37. Shore, J.; Johnson, R. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Trans. Inf. Theory* **1980**, *26*, 26–37. [[CrossRef](#)]
38. Pieter, T.; Dirk, P.; Shie, M.; Reuven, Y. A tutorial on the cross-entropy method. *Ann. Oper. Res.* **2005**, *134*, 19–67.
39. Wang, D.; Li, Q.; Zhang, J. A New Method to Analyze Evidence Conflict. *Control Theory Appl.* **2011**, *28*, 839–844.
40. Kodovsky, J.; Fridrich, J.; Holub, V. Ensemble Classifiers for Steganalysis of Digital Media. *IEEE Trans. Inf. Forensics Secur.* **2012**, *7*, 432–444. [[CrossRef](#)]