

RESEARCH

Open Access



# Robust steganography in practical communication: a comparative study

Tong Qiao<sup>1\*</sup> , Shengwang Xu<sup>1</sup>, Shuai Wang<sup>1</sup>, Xiaoshuai Wu<sup>2</sup>, Bo Liu<sup>3</sup>, Ning Zheng<sup>1</sup>, Ming Xu<sup>1</sup> and Binmin Pan<sup>1</sup>

\*Correspondence:  
tong.qiao@hdu.edu.cn

<sup>1</sup> School of Cyberspace,  
Hangzhou Dianzi University,  
Hangzhou, China

<sup>2</sup> College of Computer Science  
and Electronic Engineering,  
Hunan University, Changsha,  
China

<sup>3</sup> Chinese Aeronautical  
Establishment, Beijing, China

## Abstract

To realize the act of covert communication in a public channel, steganography is proposed. In the current study, modern adaptive steganography plays a dominant role due to its high undetectability. However, the effectiveness of modern adaptive steganography is challenged when being applied in practical communication, such as over social network. Several robust steganographic methods have been proposed, while the comparative study between them is still unknown. Thus, we propose a framework to generalize the current typical steganographic methods resisting against compression attack, and meanwhile empirically analyze advantages and disadvantages of them based on four baseline indicators, referring to as capacity, imperceptibility, undetectability, and robustness. More importantly, the robustness performance of the methods is compared in the real application, such as on Facebook, Twitter, and WeChat, which has not been comprehensively addressed in this community. In particular, the methods modifying sign of DCT coefficients perform more superiority on the social media application.

**Keywords:** Steganography, Robustness, Undetectability, Social network

## 1 Introduction

Steganography, one type of information hiding, is a science and art of transmitting secret bits in a covert channel [1, 2]. Different from watermarking technique for protecting copyrights [3], steganography conceals the existence of covert communication by embedding message into media as cover. Conventional media, including text [4], image [5], audio [6], video [7], are widely used for steganography, among which images are the most popular.

Image steganography algorithms are generally divided into two categories: non-adaptive steganography and adaptive steganography [8–15]. Non-adaptive steganographic algorithms such as Least Significant Bit replacement (LSBr) and Least Significant Bit matching (LSBm) steganography, only randomly embed secret message by modifying pixels/coefficients of the image, where selection of embedding position is nearly independent of image content. On the contrary, adaptive steganographic algorithms with minimizing a distortion function can achieve high undetectability. Typically, spatial-domain algorithms such as HUGO [16], WOW [17], HILL [18], and S-UNIWARD [19] select the appropriate embedding region according to the texture complexity of the

image; similarly, frequency-domain algorithms embed secret message by adaptively modifying DCT coefficients including UED [20], UERD [21], J-UNIWARD [19], and SI-UNIWARD [19]. In an alternative manner, the performance of adaptive strategy has been increasingly improved by minimizing non-additive distortion in steganography [22–24].

However, it hardly holds true that adaptive steganography can extract secret information entirely and correctly when stego image suffers attack from a *dirty* channel [25], such as compression, cropping, scaling, even adding noise. The dirty channel denotes a public channel where the transmitted stego image might be attacked by the server provider. The goal of steganography is to achieve covert communication between sender and receiver, and the secret message can be completely reconstructed even in the dirty channel. In this context, the design of robust steganographic algorithm has attracted more attention in the community of image steganography.

Some algorithms have already made breakthroughs. Zhang et al. first proposed a series of robust steganographic methods [26–28]. The core idea behind those algorithms is that the mapping relationship among features from cover source, which basically remains stable, should be mainly addressed. For instance, the magnitude and position relationship of DCT coefficients among inter or intra blocks are utilized. That is because the relationship basically remains its characteristic before and after compression. In the following, we simply generalize the methods [26–28]) in the early stage:

- DCRAS [26]: First, the DCT coefficients are divided into  $8 \times 8$  blocks. Next, through comparing the magnitude of target DCT coefficient and the mean of its neighbouring ones among the same sub-band, the relationship is mapped to a set of cover elements  $\{0, 1\}$ . Specifically, if the target DCT is larger than the mean of its neighbouring ones, cover element with 1 is extracted; otherwise, cover element with 0 is acquired. Finally, the secret bits encoded by RS are embedded into cover elements to generate the stego elements, which help us to acquire robust stego image.
- FRAS [27]: First, the robust regions are located by adopting Harris–Laplacian features [29]. Next, on the basis of DCRAS, the cover elements can be obtained, which further strikes the balance between undetectability and robustness. Thus, the performance of FRAS is superior to DCRAS.
- DMAS [28]: Relying on JPEG compression quantization table, with the help of dither modulation, the cover elements can be successfully extracted. Through taking full advantage of compression characteristics, the constructed embedding domain is more stable than its predecessors, leading to that the modification magnitude of DCT coefficients is smaller. Thus, both robust and undetectable performance is remarkably improved.

In the early stage, people pay more attention on the robust domain. While in the current stage, for instance, Tao et al. [30] proposed a novel robust steganographic algorithm via intermediate images based on coefficient adjustment. Based on the transmission channel matching, the authors of [31] repeatedly compressed the image to reduce the impact caused by JPEG compression of the channel, that directly improves the robustness of the transmitted stego image. In addition, Zhu et al. proposed a robust steganography

method by modifying sign of DCT coefficients [25]. Recently, robust steganographic algorithms have been further developed based on generalized dither modulation and expanded embedding domain [32], and based on the strategy of adaptive dither adjustment [33]. Encouragingly, the aforementioned algorithms perform very well when resisting attacks by JPEG compression while ensuring a high accuracy of correctly extracting secret information.

It should be addressed that in this paper, we are not prepared to list all the state of the arts while mainly revisit the typical methods in the current stage. That is because the methods in the current stages (lack of comprehensive comparison) improve the performance of the methods in the early stage. Then, the rest of this paper is organized as follows. In Sect. 2, we first propose to formulate the problems of current robust steganographic algorithms; we mainly overview three typical algorithms, which is followed by experimental results in Sect. 3. Finally, the concluding remarks are drawn in Sect. 4.

## 2 Typical robust steganography

### 2.1 Framework and metric

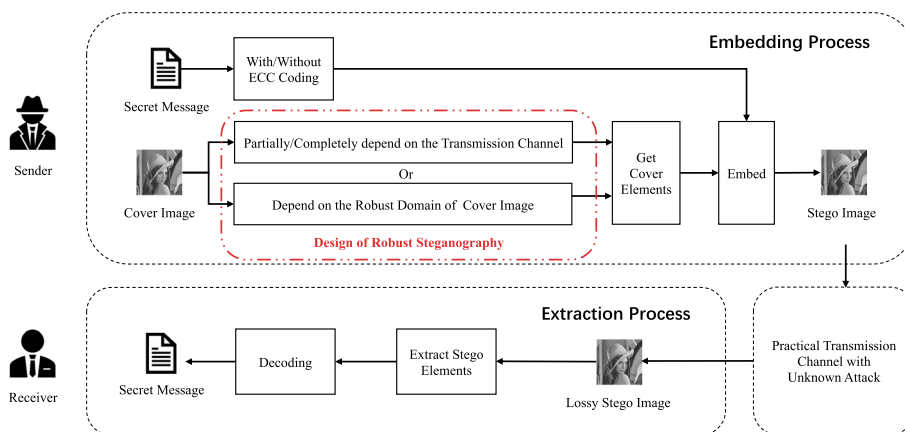
In the current study, the simulated channels in the laboratory are often used to attack stego images instead of real transmission channels. Various channels might offer different attacks, among which JPEG compression is widely-adopted. In this context, we propose to address some limitations of current arts.

First, some steganographic methods heavily depend on the quality factor (QF) of JPEG compression from the transmission channel, where the property of transmission channel is required in advance. Therefore, if the transmission channel becomes unknown, the robustness of steganography cannot be guaranteed, even resulting in complete failure of secret information extraction.

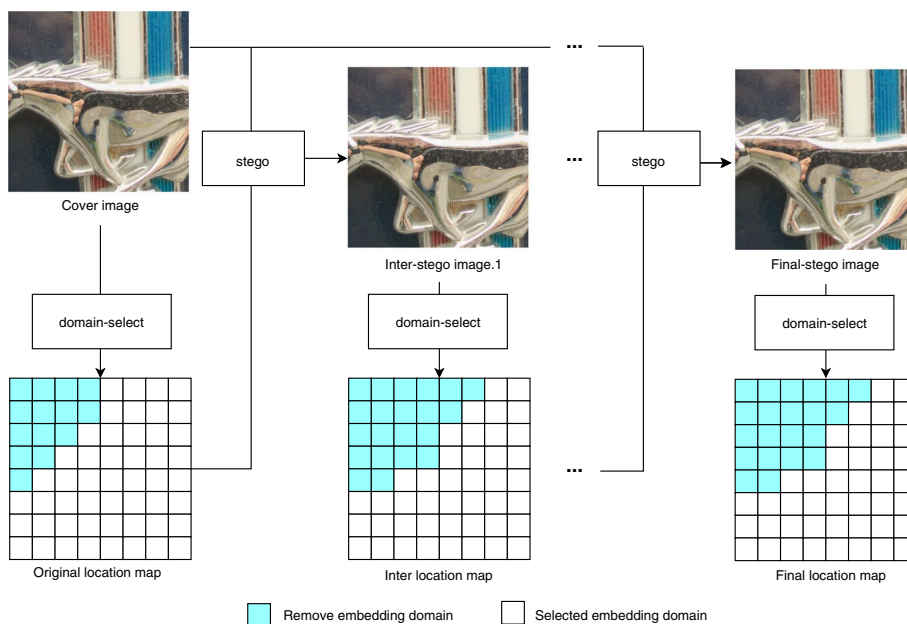
Second, many robust steganographic methods only consider a single attack, which is unreal in practical communication. In particular, one method can resist well against JPEG compression while fail to deal with the case of re-scaling. In addition, robust steganography relying on the correlation feature between neighboring pixels (or DCT coefficients) probably fails when the relationship is destroyed by multiple attacks from the transmission channel.

To our knowledge, to complete the task of robust steganography, most of current studies are prone to sacrifice the embedding capacity for improving the security of covert communication. For clarity, it is proposed to generalize a framework of robust steganography as Fig. 1 reports. For a sender, she/he would like to embed the secret message into a cover image, in which cover elements are selected for embedding generally relying on the design of robust steganography, leading to a stego image; for a receiver, the secret message can be successfully extracted by carrying out the extraction process even in a dirty channel, in which lossy compression unavoidably happens. Moreover, for instance, we also provide a use case of robust steganography in Fig. 2, by iteratively simulating procedure of JPEG compression, the robust DCT coefficients from the 64 subbands, which always remain unchanged before and after compression, are selected.

Besides, four indicators are used to measure the overall performance of different algorithms:



**Fig. 1** Framework for covert communication in a dirty channel; this framework has been adopted by most current robust steganography



**Fig. 2** Use case of robust steganography

- Capacity: the stego image carries the maximum number of the secret bits.
- Imperceptibility: the stego image is indistinguishable from the cover image in human visual perception.
- Undetectability: the stego image can bypass the detection from modern advanced steganalysis.
- Robustness: the stego image capable of successfully transmitting secret bits, is immune to known or unknown post-processing attacks.

Without loss of generality, the technique of steganography needs to move from the laboratory to the real world for a wide range of applications. Then the fashionable social network replaces the simulated transmission channel for cover communication, such as Facebook,

Twitter, and WeChat, in which various unknown attacks are adopted to threaten stego images. To address those challenges, the study of robust steganography should mainly focus on the robustness of stego image itself (active scheme), rather than relying on the transmission channel (passive scheme). That is because the vulnerability of transmission channel based algorithm is possibly exposed when social network ceaselessly adjusts its corresponding attack strategy such as QFs of JPEG compression (see Subsection 3.6 of Sect. 3). Next, we generally overview three typical robust steganographic methods, and provide the comprehensive comparison in the numerical experiments.

In general, two mainstream categories of robust steganography are proposed. The first one mainly relies on the transmission channel, see [30, 31] for instance; the second one mainly relies on the robust domain, see [25] for instance. In this section, we mainly introduce the main ideas of three typical robust steganographic algorithms [25, 30, 31], and highlight the advantages and limitations. The representative steganographic methods we chose are chronologically proposed to deal with the problem of robustness of covert communication, dependent of different strategies.

### 2.2 Transmission channel matching

The main idea of [31] is to adapt the image to the transmission channel, that is, to repeatedly compress the image with the same QF to reduce extraction error, leading to the improved robustness of the stego image. It is worth noting that before transmission, a “cover image” has to be optimized.

As Fig. 3 illustrates, the principal of TCM (Transmission Channel Matching) is actually that the strategy of multiple JPEG compression helps us reduce/eliminate errors caused by quantization and rounding, that is straightforward formulated as

$$S = \text{round} \left[ F_{IDCT} (Q \times D) \right], \tag{1}$$

where  $D$  denotes the original DCT coefficients from a JPEG image,  $Q$  the quantization step, and  $F_{IDCT}(\cdot)$  represents the inverse DCT operation. After rounding, we can obtain the de-compressed image  $S$  in the spatial domain. Then we generate the new DCT matrix  $D^{(1)}$  by transforming  $S$  to DCT coefficients:

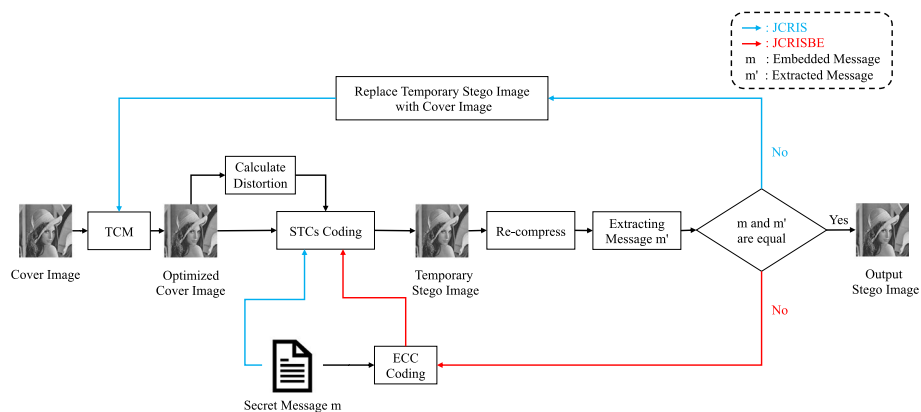


Fig. 3 Scheme of transmission channel matching [31]

$$\mathbf{D}^{(1)} = \text{round} \left[ F_{\text{DCT}}(\mathbf{S}) / \mathbf{Q} \right], \quad (2)$$

where  $F_{\text{DCT}}(\cdot)$  denotes DCT operation. Then, let us define  $N_{\text{diff}}(\mathbf{D}, \mathbf{D}^{(1)})$  as the number of different DCT coefficients between  $\mathbf{D}$  and  $\mathbf{D}^{(1)}$ . Since that the effect on DCT coefficients from JPEG compression attack is mainly caused by rounding and quantization operation, the number  $N_{\text{diff}}$  can be further decreased (to zero for optimal solution) by multiply compressing an image.

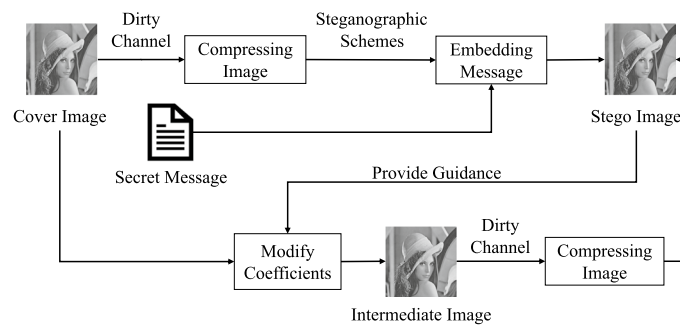
Furthermore, let us extend our description of TCM by defining DCT coefficients as  $\mathbf{D}$  from a cover image  $\mathbf{X}$ ; meanwhile the QF pre-set by the dirty channel, denoted as  $\mathbf{Q}_{\text{channel}}$ , is prior-known (easily acquired by up-loading and down-loading the same image). Before applying TCM operation, the size of the image should be calibrated, matching to the channel. First, it is proposed to obtain  $\mathbf{D}^{(1)}$  by  $\mathbf{Q}_{\text{channel}}$ . If  $N_{\text{diff}}(\mathbf{D}, \mathbf{D}^{(1)}) = 0$ , the cover image  $\mathbf{X}$  is used for embedding; on the contrary,  $\mathbf{X}'$  re-constructed by  $\mathbf{D}^{(1)}$  is re-compressed to obtain  $\mathbf{D}^{(2)}$ . Next, in the case of  $N'_{\text{diff}}(\mathbf{D}^{(1)}, \mathbf{D}^{(2)}) = 0$ ,  $\mathbf{X}'$  is used for embedding; otherwise, we calculate the ratio  $\frac{N_{\text{diff}}}{N'_{\text{diff}}}$ . When the ratio is greater than 0.98, the original cover image  $\mathbf{X}$  is used for embedding. If the above conditions cannot be met, in virtue of Eqs. (1) and (2), we re-do the aforementioned procedure using the twice-compressed image until the optimal cover image is acquired.

During the embedding process, two benchmark adaptive steganographic algorithms for JPEG image are adopted, namely J-UNIWARD [19] and UED [20]. More importantly, to further improve the robustness, the mechanism of ECC (Error Correction Coding) is established, leading to two specific schemes. In particular, JCRIS (JPEG Compression Resistant Solution) embeds the secret message without ECC into the optimized cover image. If the secret message of the stego image cannot be correctly extracted, the stego image needs to be re-compressed again as TCM operation until the secret message are correctly extracted. Different from JCRIS, prior to embedding, JCRISBE (JPEG Compression Resistant Solution with Bch code) first needs to encode the secret message.

JCRIS has the higher embedding capacity while the advantages of JCRISBE are the stronger robustness and higher security. Moreover, the robustness of stego image is improved by adopting both two algorithms. Experimental results show that the correct rate of the secret message extraction can reach more than 99%. It should be noted that The TCM-based robust steganography works very well in the dirty channel with stable property. However, when the dirty channel arises the compression attack by using randomly-selected QF, the method probably fails.

### 2.3 Modifying coefficients of cover image

Tao et al. [30] proposed a robust steganographic scheme in another manner. The main idea behind the method focuses on Modifying Coefficients of Cover Image (MCCI) in virtue of the DCT coefficients of the stego image. Figure 4 illustrates the general framework of MCCI. In fact, MCCI successfully simulates the procedure of compression attack in a dirty channel as that of generating stego image. In this well-designed framework, the stego image generated by compression is completely the same as the stego images directly generated by using a modern adaptive embedding algorithm, such as J-UNIWARD [19] or UERD [21].



**Fig. 4** Scheme of modifying coefficients of cover image [30]

Given that a cover image is used for embedding, let us compress the cover image with a prior previously known QF in a dirty channel. Next, we apply some adaptive steganographic scheme to embed a secret message into the compressed cover image for generating stego one. Meanwhile, the DCT coefficients of the original cover image are adjusted according to the stego image for obtaining an intermediate image. Finally, the intermediate image are transmitted in the dirty channel, leading to a new stego image. This method ensures that the stego image generated by compressing the intermediate image is completely identical to the one generated by directly embedding at the compressed cover image. Moreover, the secret message can be correctly extracted at the receiver side. It is worth noting that the key step of MCCI is to generate the correct immediate image.

Given DCT coefficient matrix  $\mathbf{I}$  from an intermediate image, coefficient matrix  $\mathbf{D}$  from a cover image  $\mathbf{X}$  and  $\mathbf{D}'$  from its corresponding stego one  $\mathbf{Y}$ , we are trying to find an  $\alpha$  to make the following equation hold:

$$\mathbf{I} = \mathbf{D} + \alpha, \tag{3}$$

where

$$\alpha = \arg \min_{\epsilon} \left| \left[ \mathbf{D} + \epsilon \right] \cdot \frac{\mathbf{Q}_{\text{cover}}}{\mathbf{Q}_{\text{channel}}} - \mathbf{D}' \right| \tag{4}$$

where  $\alpha$  denotes a dithering map whose elements are integers.  $\epsilon$  denotes a set of integers, quantization step  $\mathbf{Q}_{\text{cover}}$  from cover image and  $\mathbf{Q}_{\text{channel}}$  from the dirty channel. In such tricky manner, we can first acquire an intermediate image based on  $\mathbf{I}$  with the equivalent  $\mathbf{Q}_{\text{cover}}$ , and then generate the stego image  $\mathbf{Y}$  via transmitting the intermediate image in the dirty channel.

Although MCCI has remarkable advantage of correct extraction of secret message, it possibly fails to resist JPEG compression attack when the dirty channel with a metabolic QF or double compression.

### 2.4 Modifying sign of DCT coefficients

When the JPEG image is transmitted in a dirty channel, the sign of DCT coefficients is not easy to be changed, which can be utilized for designing a robust steganographic method. Zhu et al. [25] proposed a scheme by Modifying Sign of DCT Coefficients



(MSDC) (see Fig. 5 for illustration). First of all, the suitable DCT coefficients are selected and optimized for embedding. While instead of original DCT coefficients, the cover elements are generated based on the sign of DCT coefficients, referring to as 1 (positive value) or 0 (negative value). Next, relying on the STCs, the stego elements (mapping to the sign of DCT coefficients from stego image) are acquired, which follow the same rule of generating cover element. Meanwhile, the secret message is encoded by ECC before embedding. Then, the key step of MSDC is to design the optimal cost function for STCs coding.

To design well-performed cost function, it is proposed to borrow the idea from J-UNIWARD [19] for achieving better undetectability, where the complex texture areas are usually assigned small values. Furthermore, in virtue of DCT coefficient  $d_{i,g}$ , the cost function of MSDC can be formulated as

$$\rho(d_{i,g}) = \begin{cases} |d_{i,g}|, & d_{i,g} \leq |\alpha|, \rho^J(d_{i,g}) < \beta \\ |d_{i,g}|^\gamma, & d_{i,g} > |\alpha|, \rho^J(d_{i,g}) < \beta \\ \rho^J(d_{i,g}), & \rho^J(d_{i,g}) \geq \beta \end{cases} \quad (5)$$

where  $\rho^J$  represents the cost function of J-UNIWARD.  $g$  and  $i$  respectively represent the  $g$ -th DCT block and the  $i$ -th position sorted by zig-zag in the DCT block. Here,  $\alpha$  is used to limit the maximum of the DCT coefficients  $\mathbf{D}$  for embedding;  $\beta$  represents a threshold, and  $\gamma > 1$  serves as the gain.

For clarity, the STCs is also given, which can be formulated by

$$\mathbf{s} = F_{\text{STCs}}(\mathbf{c}, \rho(\mathbf{D}), \mathbf{m}, \mathbf{H}) \quad (6)$$

where  $\mathbf{c}$  denotes the cover elements, the secret message  $\mathbf{m}$ , and stego elements  $\mathbf{s}$ . We select the parity-check matrix  $\mathbf{H}$  as a key. Finally, we compare the cover elements with the stego elements which are used for embedding. If two values are identical, the DCT coefficient remains unchanged; otherwise, the sign of the DCT coefficient is flipped.

Obviously, MSDC has strong robustness due to that the sign nearly remains invariant after compression. Moreover, the algorithm preferentially selects non-zero DCT coefficients with small absolute values for modification. However, as the payload increases, the number of modified DCT coefficients also increases, leading to the enlarged extraction error or perceptual distortion.

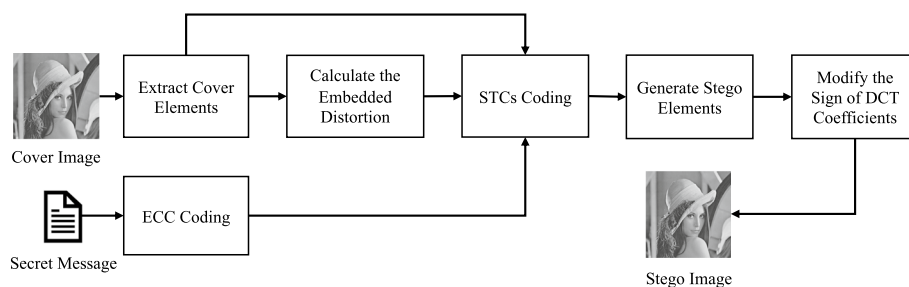


Fig. 5 Scheme of modifying sign of DCT coefficients [25]



### 3 Comparison experiment and analysis

#### 3.1 Experiment setup

In this section, we will conduct the comprehensive comparison experiments, in which the current benchmark methods TCM (Transmission Channel Matching), MCCI (Modifying Coefficients of Cover Image), and MSDC (Modifying Sign of DCT Coefficients) are compared based on the indicators: capacity, imperceptibility, undetectability, and robustness. The baseline dataset BOSSbase ver. 1.01 [34] contains 10,000 grayscale raw images in PGM format with  $512 \times 512$  pixels. First of all, all images are compressed with QF 75, 85, and 95; the payload used for embedding ranges from 0.05 to 0.1 bpnzac (bit per non-zero AC DCT coefficients). Meanwhile, we use STCs [35] encoding to minimize the distortion embedding. Besides, it should be noted that for fair comparison, the embedding strategy of three methods refers to the prior-art J-UNIWARD.

#### 3.2 Capacity performance

The capacity refers to the maximum number of secret bits embedded into the image. In the frequency domain, the capacity is measured by the number of non-zero AC coefficients for embedding. Specifically, all 10,000 images are compressed with  $QF = \{75, 85, 95\}$ , serving as cover source. Meanwhile, the JPEG compression attack from the simulated dirty channel uses QF 75. As QF is decreased, the number of non-zero AC coefficients reduces. It is worth noting that the comparison experiments are conducted under the same condition. Among three methods, the capacity of MCCI is the number of non-zero AC coefficients of the cover image while the capacity of the others is constrained by the corresponding algorithms (see Table 1).

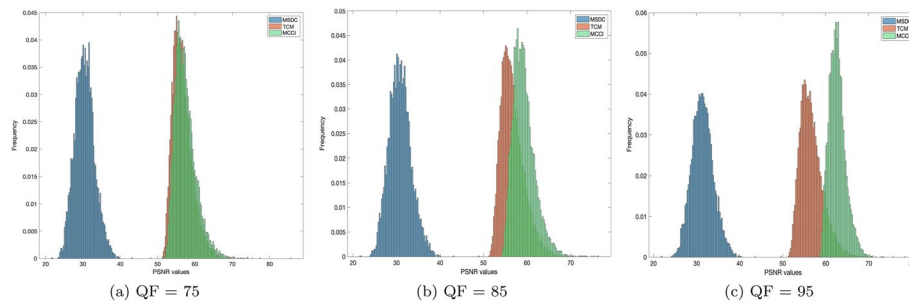
The number of non-zero AC coefficients changes when QF changes. Accordingly, MCCI has the largest capacity, where all non-zero AC coefficients can be used as cover elements for embedding. While both TCM and MSDC have to select robust DCT coefficients as cover elements for embedding. Since that TCM needs to multiply compress the transmitted images with the same QF for reserving the invariance of DCT coefficients, the similar capacity appears among original cover images with different QFs. As a matter of fact, TCM and MSDC sacrifice to some extent capacity for guaranteeing robustness.

#### 3.3 Imperceptibility performance

The stego image should remain perceptual distortion as small as possible, referring to imperceptibility. In this section, we select stego images with 0.1 payload and  $QF = \{75, 85, 95\}$ . To evaluate the imperceptibility of different stego images, the baseline PSNR (Peak Signal to Noise Ratio) serves as a metric. The higher the PSNR value is, the less perceptual distortion the stego image suffers.

**Table 1** Comparison of average capacity

Quality factor	75	85	95
<b>Method</b>			
MSDC	41529	28748	31525
TCM	33387	33362	33440
MCCI	41592	56206	98065



**Fig. 6** Normalized histogram comparison of PSNR values from 10,000 images

**Table 2** Average PSNR comparison

Quality factor Method	MSDC	TCM	MCCI
75	30.4625	56.6977	57.4343
85	30.8160	56.6986	59.2747
95	31.3941	56.6959	62.6871

As shown in Table 2, the average PSNR values of 10,000 images are compared. MSDC has the lowest average PSNR value among different QFs. MCCI has the highest PSNR value, that is close to TCM. With increasing QF, the average PSNR of three methods basically has the incremental trend. In addition, Fig. 6 shows the normalized histogram of PSNR values from 10,000 images. It can be obviously observed that MSDC has the lowest PSNR values among three QFs. When  $QF = 75$ , the histograms of TMC and MCCI are nearly overlapped, meaning the very similar PSNR values. It should be noted that the overall distribution of TMC basically remains unchanged.

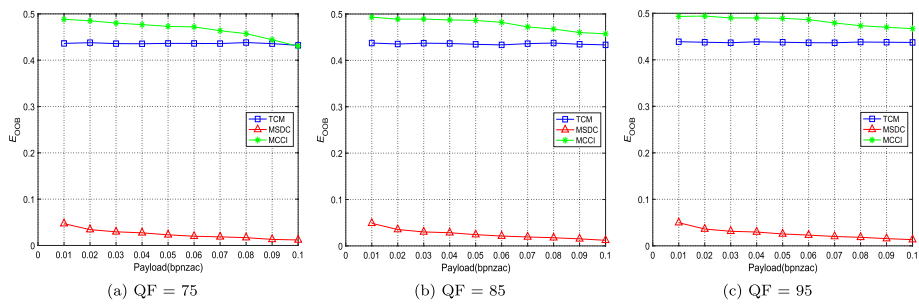
MSDC modifies the sign of the DCT coefficient, leading to the most changes of the image content compared to the others. Thus, the smallest PSNR values are from MSDC. If the selection of DCT coefficients is further optimized, we believe the the PSNR values of MSDC can be enhanced. Due to that TCM generates stego images by repeatedly compressing cover images with different QF, the PSNR values are very similar among three methods.

### 3.4 Undetectability performance

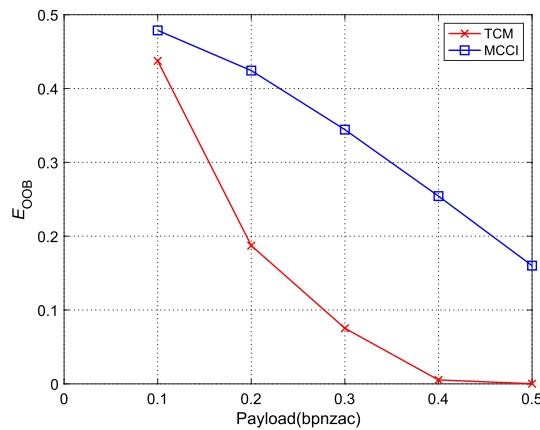
In this section, we evaluate the undetectability of three methods by using the benchmark steganalytic detector. The ensemble classifier [36], equipped with the rich model features DCTR-8000D [37], is our optimal choice. We randomly select 5,000 stego images and 5,000 cover images for training; the remaining images are used for testing. The  $P_E$  (classification error rate) is defined as

$$P_E = \min \frac{P_{FA} + P_{MD}}{2}, \quad (7)$$

where  $P_{FA}$  is the false alarm rate and  $P_{MD}$  is the missed detection rate. In this paper, we use the ensemble's "out-of-bag" (OOB) error  $E_{OOB}$  instead of  $P_E$ , since that  $E_{OOB}$  is an unbiased estimate of the  $P_E$  (see [21]).



**Fig. 7** Undetectability comparison with payloads from 0.01 to 0.1



**Fig. 8** Undetectability of TCM and MCCI with payloads from 0.1 to 0.5

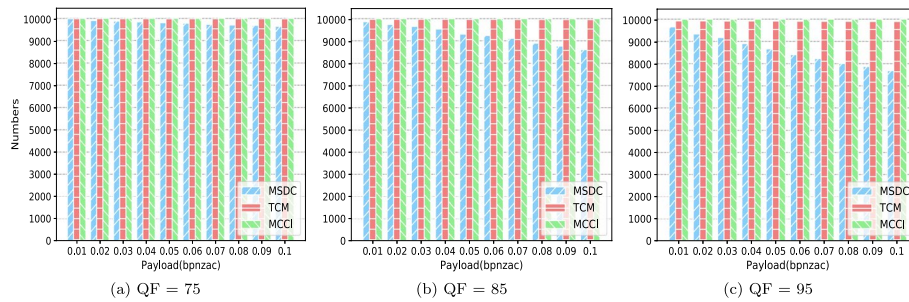
As shown in Fig. 7, if the small payloads (not larger than 0.1) are adopted for evaluation, MSDC has the lowest  $E_{OOB}$  while MCCI has the highest undetectability. Basically, the changes of QF cannot impact the undetectability among three methods.

Moreover, we propose to compare the undetectability of TCM and MCCI using more payloads {0.1, 0.2, 0.3, 0.4, 0.5} (see Fig. 8 for illustration). Due to the poor undetectability of MSDC, we remove the results of comparison experiments. TMC usually chose non-zero AC coefficients with strong robustness as cover elements. Therefore, the capacity is limited, leading to that as payload increases, TMC has to use the high-cost DCT coefficients. On the contrary, in the framework of MCCI, the stego image is the same as the stego one generated directly by the prior-art J-UNIWARD. Therefore, It has the optimal undetectability. Among three methods, MSDC has a relatively low undetectability, that is kind of different from the results of [25], where the CCPEV [38] and CCJRM [39] features are used. Nevertheless, MSDC changes the sign of the DCT coefficients, directly leading to the unsatisfactory undetectability.

To enrich our experiments, it is proposed to illustrate the undetectable performance of methods [26–28] in the early stage. Here, it is proposed to evaluate the performance by using the ensemble detector equipped with CCPEV [38] features. As Table 3 illustrates, DCRAS performs worst while FRAS to some degree improves the performance due to that the cover elements are optimally selected in the robust domain. It can be observed that DMAS performs best based on the embedding domain constructed by

**Table 3** Undetectable performance ( $E_{OOB}$ ) of methods in the early stage

Payload	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
DCRAS	0.4282	0.3378	0.2655	0.1979	0.1383	0.0997	0.0752	0.0706	0.0411	0.0256
FRAS	0.4595	0.3819	0.3387	0.2590	0.1809	0.1373	0.0961	0.0862	0.0646	0.0576
DMAS	0.4660	0.4330	0.3985	0.3590	0.2980	0.2555	0.2030	0.1630	0.1245	0.0955



**Fig. 9** Number of images that can completely reconstruct secret bits

dither modulation. In this case, the modification of DCT coefficients caused by embedding tends to the minimum compared to the others. Nevertheless, the methods in the early stage provide a new manner to design robust steganographic scheme, where the robust domain is proposed by referring to robust watermarking, that more or less guides the design of the following methods in the current stage.

### 3.5 Robustness performance

In this section, we propose to adopt extraction error rate for comparing the robustness performance of three methods, which can be calculated by

$$R_{\text{error}} = \frac{n_{\text{error}}}{n_{\text{msg}}}, \tag{8}$$

where  $n_{\text{error}}$  denotes the number of incorrectly extracted bits, and  $n_{\text{msg}}$  is the total number of secret bits. The smaller  $R_{\text{error}}$  means the stronger robustness. Alternatively, we define the number of correct extraction, which represents the total number of images that can completely reconstruct the secret information.

As Fig. 9 reports, all secret bits can be perfectly correctly extracted in MCCI. That is because the stego image generated by MCCI is the same as the stego image directly generated by modern adaptive steganography. And the number of correct extraction in TCM is also satisfying. While MSDC cannot perform very well when QF equals 75 as the cases of  $QF = \{85, 95\}$ . In addition, Fig. 10 illustrates the  $R_{\text{error}}$  results of three compared methods. As we expected, the performance of both MSDC and MCCI is satisfying, nearly approaching to zero error. However, the  $R_{\text{error}}$  of TCM is slightly larger than the others among three QFs.

TCM is prone to multiply compress the image before transmission. It is always trying to remove images that are not suitable for embedding by setting an empirical threshold.

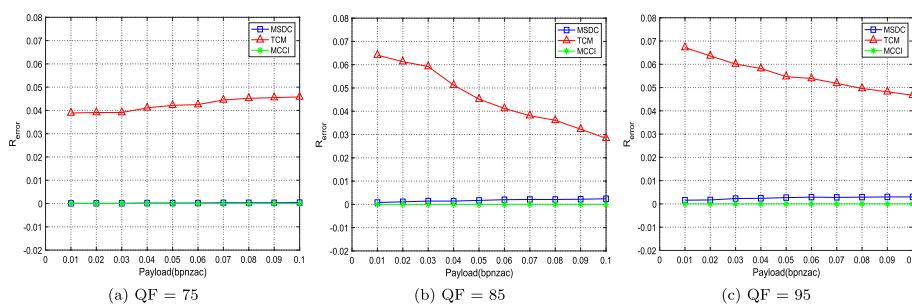


Fig. 10 R<sub>error</sub> performance of resisting JPEG compression

Table 4 Robustness performance on Facebook

Method	MSDC	MCCI	TCM	J-UNIWARD	UERD
<b>Results</b>					
$R_c$	47/50	0/50	0/50	0/50	0/50
$R_{error}$	0.0043	0.3537	0.5013	0.4593	0.4600

In our experiment, strictly as the setting of [31], all the images are used for embedding, which unavoidably contains the unsatisfactory samples (unsuccessfully excluded), directly resulting in that the  $R_{error}$  increases. However, the number of those samples is not very large, that can explain the better performance in Fig. 9.

### 3.6 Practical performance on social network platform

In this section, we will evaluate the practical performance of three baseline methods on Social Network Platform (SNP), such as Facebook, Twitter, and WeChat. Also, two baseline modern adaptive steganographic schemes, namely J-UNIWARD and UERD, are also compared. First, we randomly select 50 images in the BOSSbase, and compress them with QF 85. Then we embed secret message by payload 0.1. All the stego images are uploaded to Facebook or Twitter, and then downloaded; otherwise, the stego images are transmitted through WeChat. The robustness performance is measured by  $R_c$  and  $R_{error}$  (see Eq. (8)).  $R_c$  denotes the ratio of the number of images in which the secret message can be completely extracted to the number of all the tested stego images.

As Table 4 reports, it can be seen that by adopting MSDC, the secret message can be completely extracted among 47 stego images, meaning that MSDC performs very well on Facebook. However, all the images of MCCI or TCM fail, where the results of them are not as good as that in the simulated channel (see Fig. 10). In addition, as we expected, both two non-robust steganographic schemes are invalid for covert communication in the real application.

To our knowledge, it hardly holds true that the social network platform, such as Facebook, only adopts simply once JPEG compression, where the metabolic dirty channel probably remarkably impacts the robust steganography heavily relying on the transmission channel such as MCCI and TCM. Prior to embedding, MCCI or TCM needs to learn the property of the dirty channel, such as QF. While the learned property will be invalid as the dirty channel continuously changes. On the contrary, the dependence of

MSDC on the channel is minimal, which is nearly immune to the changes of the transmission channel, leading to the best performance among three methods. In fact, in our early observation (at the time both MCCI and TCM were published in 2018), Facebook indeed fixes the QF as 71. In 2019, the strategy of Facebook was changed, in which original images with QF not larger than 85 are re-compressed with QF 71; otherwise, the re-compression is proceeded with various QFs (see [32]). However, in our recent observation (till the date of paper submission), Facebook completely breaks the prior rule of compression, that continuously changes QF even though the images have the same original QF before uploading. Meanwhile, when the downloaded images from Facebook are uploaded again, the compression strategy is still unpredictable.

Moreover, we conduct the experiments on Twitter. When comparing the transmitted image via repeatedly uploading and downloading, it can be clearly observed that the image is post-processed by updating the strategy of compression. As Table 5 illustrates, TCM fails to complete the task of robust steganography. That is because the requirement (the ratio  $\frac{N_{\text{diff}}}{N_{\text{diff}}} > 0.98$ ) of transmission channel cannot be met as multiple uploading and downloading. Besides, the modern adaptive steganographic schemes are still invalid. On the contrary, MSDC without heavily relying on the channel performs best. Compared with the results on Facebook, the  $R_{\text{error}}$  of MSDC decreases while the number of correct extraction to some extent is reduced. It is worth noting that the DCT coefficient (carrying secret bit) with its absolute value equal to 1, more easily changes to zero (caused by compression from Twitter) at the receiver side, that directly leads to the incorrect extraction of secret bits of MSDC.

Next, it is proposed to conduct our experiments on WeChat, which is one of the most popular instant-messaging App. Here, we consider two models for covert communication, that is original model and non-original model. In the original model, the image is sent with selection of *original image*; In the non-original model, the image is directly sent with default selection. As we expected, in the original model, among five compared schemes, the secret bits can be perfectly extracted from all testing images (see Table 6). By checking the DCT coefficients and QF before and after transmission on WeChat, we cannot observe any changes; meanwhile, the capacity of image is reduced. Thus, we infer that the unknown lossless compression happens in the original model. Specifically, lossless compression is a manner of compression without losing any data, where the original data can be perfectly reconstructed from the compressed data (see [40] for details). Unfortunately, as Table 7 illustrates, in the non-original model, all the methods fail to transmit stego images on WeChat. That is because WeChat non-maliciously attacks the images using multiple post-processing manners except compression.

Recently, some new robust steganographic methods are proposed successively. We select the typical methods designed based on the proposed framework. It is worth

**Table 5** Robustness performance on Twitter

Method	MSDC	MCCI	TCM	J-UNIWARD	UERD
<b>Results</b>					
$R_C$	35/50	0/50	fail	0/50	0/50
$R_{\text{error}}$	0.0016	0.4226	fail	0.4991	0.4993

**Table 6** Robustness performance on WeChat (original image)

Method	MSDC	MCCI	TCM	J-UNIWARD	UERD
<b>Results</b>					
$R_C$	50/50	50/50	50/50	50/50	50/50
$R_{\text{error}}$	0	0	0	0	0

**Table 7** Robustness performance on WeChat (non-original image)

Method	MSDC	MCCI	TCM	J-UNIWARD	UERD
<b>Results</b>					
$R_C$	0/50	0/50	0/50	0/50	0/50
$R_{\text{error}}$	0.1297	0.4807	0.5008	0.4977	0.4979

**Table 8** Robustness performance of color images on Facebook

Method	ESS	SSR	DMMR	SRJS	PQ-UNIWARD
<b>Results</b>					
$R_C$	50/50	49/50	50/50	0/50	0/50
$R_{\text{error}}$	0	0	0.2224	0.5773	0.4958

**Table 9** Robustness performance of color images on Twitter

Method	ESS	SSR	DMMR	SRJS	PQ-UNIWARD
<b>Results</b>					
$R_C$	50/50	49/50	50/50	0/50	0/50
$R_{\text{error}}$	0	0	0.2307	0.5749	0.4958

noting that ESS (Enhanced Sign Steganography) [12] and SSR (Sign Steganography Revisited) [13] are both MSDC-based, DMMR (Dither Modulation and Modification with Re-compression) [41] is TCM-based, and SRJS (Secure Robust JPEG Steganography) [42] is MCCI-based. Besides, we add the updated classical steganography PQ-UNIWARD [43] for comparison. Since that the color images are usually transmitted on SNPs, we intend to carry out the experiments on ALASKA dataset [44], where 50 testing color images are selected for verifying the effectiveness of the compared methods. In Tables 8 and 9, by observation, the robust steganography methods can successfully extract the secret images while the traditional steganography PQ-UNIWARD cannot complete the task of robust steganography. Besides, although SRJS performs very well in the simulated settings, it unavoidably exposes its limitation in the practical SNPs.

On SNP such as Facebook, Twitter, and WeChat, the transmission channel, that is characterized as *black box*, behaves in the more complicated system than the dirty channel in the simulated settings which only consider image compression attack. In virtue of our empirical analysis in this subsection, in the case that the channel is not completely acquainted, it is suggested to adopt the scheme relying on robust domain such as MSDC



while carefully adopting the scheme heavily relying on the transmission channel such as MCCI and TCM.

#### 4 Conclusion

In this paper, we mainly overview three current typical robust steganographic methods. More importantly, the performance of them is comprehensively compared relying on the baseline indicators, referring to as capacity, imperceptibility, undetectability, and robustness. Also, both strength and limitation of three methods are empirically analyzed. Although three methods perform very well, none of them can achieve the best results in all four indicators. Nevertheless, in the simulated transmission channel, MCCI is basically our optimal choice for robust steganography. That is because the stego image generated by MCCI is the most similar to the stego one by modern adaptive steganography such as J-UNIWARD. In the practical application, for instance covert communication on Facebook, it is no doubt that MSDC performs best in the aspect of robustness.

Moreover, based on the empirical results and analysis, we list the following studies for improving current algorithms:

- 1 To obtain better undetectability performance, the design of robust steganography should combine with modern adaptive scheme, together with its optimized cost function.
- 2 To further improve the robustness performance, the selection of cover elements is of importance, in which the property of DCT coefficients including their adjacent relationship can be further dug out. Also, the other known attacks, such as re-scaling and noise-adding, or more unknown attacks, should be addressed.
- 3 In an alternative manner, the robustness performance can be improved by to some extent reducing capacity. In the real scenario, we segment a set of secret message into several subsets, which are embedded into multiple images rather than using only one image.
- 4 The carrier may not only be limited to DCT coefficients or pixels. We can also investigate other ways of hiding the secret message into the image files, the property of which remains unchanged in covert communication.

In the future study, we will mainly focus on improving the robustness of the steganography method to ensure that the transmitted stego images can resist against known and unknown attacks from social network. Moreover, we address again that moving the steganography from the laboratory to the real world is of importance.

#### Abbreviations

LSBr	Least Significant Bit replacement
LSBm	Least Significant Bit matching
HUGO	Highly Undetectable steGo
WOW	Wavelet Obtained Weights
S-UNIWARD	Spatial UNiversal WAvelet Relative Distortion
UED	Uniform Embedding Distortion
UERD	Uniform Embedding Revisited Distortion
J-UNIWARD	JPEG Ulversal WAvelet Relative Distortion
SI-UNIWARD	Side Information UNiversal WAvelet Relative Distortion
RS	Reed-solomon
DCRAS	DCT Coefficients Relationship based Adaptive Steganography
FRAS	Feature Regions based Adaptive Steganography

DMAS	Dither Modulation based Adaptive Steganography
JPEG	Joint Photographic Experts Group
QF	Quality factor
TCM	Transmission Channel Matching
ECC	Error Correction Coding
JCRIS	JPEG Compression Resistant Solution
JCRISBE	JPEG Compression Resistant Solution with Bch code
MCCI	Modifying Coefficients of Cover Image
MSDC	Modifying Sign of DCT Coefficients
STC	Syndrome-trellis codes
PGM	Portable graymap file format
PSNR	Peak Signal to Noise Ratio
OOB	Out-of-bag
SNP	Social Networking Platform
ESS	Enhanced Sign Steganography
SSR	Sign Steganography Revisited
DMMR	Dither Modulation and Modification with Re-compression
SRJS	Secure Robust JPEG Steganography

#### Acknowledgments

Not applicable

#### Author contributions

TQ, SX, and SW designed and implemented the study and wrote the paper. XW, BL, NZ, MX and BP gave advice and revised the paper. The authors read and approved the final manuscript.

#### Funding

This work was supported by Zhejiang Provincial Natural Science Foundation of China (No. LZ23F020006).

#### Availability of data and materials

The datasets used and analyzed in the current study are available from the corresponding author upon reasonable request.

#### Declarations

#### Competing interests

The authors declare that they have no competing interests.

Received: 14 August 2020 Accepted: 25 October 2023

Published online: 31 October 2023

#### References

1. P.C. Mandal, I. Mukherjee, G. Paul, B. Chatterji, Digital image steganography: a literature survey. *Inf. Sci.* (2022). <https://doi.org/10.1016/j.ins.2022.07.120>
2. M. Hussain, A.W.A. Wahab, Y.I.B. Idris, A.T. Ho, K.-H. Jung, Image steganography in spatial domain: a survey. *Signal Process. Image Commun.* **65**, 46–66 (2018)
3. M. Asikuzzaman, M.R. Pickering, An overview of digital video watermarking. *IEEE Trans. Circuits Syst. Video Technol.* **28**(9), 2131–2153 (2017)
4. T.-Y. Liu, W.-H. Tsai, A new steganographic method for data hiding in Microsoft word documents by a change tracking technique. *IEEE Trans. Inf. Forensics Secur.* **2**(1), 24–30 (2007)
5. B. Li, S. Tan, M. Wang, J. Huang, Investigation on cost assignment in spatial image steganography. *IEEE Trans. Inf. Forensics Secur.* **9**(8), 1264–1277 (2014)
6. Y. Huang, C. Liu, S. Tang, S. Bai, Steganography integration into a low-bit rate speech codec. *IEEE Trans. Inf. Forensics Secur.* **7**(6), 1865–1875 (2012)
7. D. Xu, R. Wang, Y.Q. Shi, Data hiding in encrypted H. 264/AVC video streams by codeword substitution. *IEEE Trans. Inf. Forensics Secur.* **9**(4), 596–606 (2014)
8. M. Hussain, A.W.A. Wahab, A.T. Ho, N. Javed, K.-H. Jung, A data hiding scheme using parity-bit pixel value differencing and improved rightmost digit replacement. *Signal Process. Image Commun.* **50**, 44–57 (2017)
9. A.A. Zakaria, M. Hussain, A.W.A. Wahab, M.Y.I. Idris, N.A. Abdullah, K.-H. Jung, High-capacity image steganography with minimum modified bits based on data mapping and LSB substitution. *Appl. Sci.* **8**(11), 2199 (2018)
10. Y. Zhang, X. Luo, Y. Guo, C. Qin, F. Liu, Multiple robustness enhancements for image adaptive steganography in lossy channels. *IEEE Trans. Circuits Syst. Video Technol.* **30**(8), 2750–2764 (2019)
11. Y. Wang, M. Tang, Z. Wang, High-capacity adaptive steganography based on LSB and hamming code. *Optik* **213**, 164685 (2020)
12. T. Qiao, S. Wang, X. Luo, Z. Zhu, Robust steganography resisting jpeg compression by improving selection of cover element. *Signal Process.* **183**, 108048 (2021)
13. X. Wu, T. Qiao, Y. Chen, M. Xu, N. Zheng, X. Luo, Sign steganography revisited with robust domain selection. *Signal Process.* **196**, 108522 (2022)

14. K. Zeng, K. Chen, W. Zhang, Y. Wang, N. Yu, Improving robust adaptive steganography via minimizing channel errors. *Signal Process.* **195**, 108498 (2022)
15. G. Xie, J. Ren, S. Marshall, H. Zhao, R. Li, A novel gradient-guided post-processing method for adaptive image steganography. *Signal Process.* **203**, 108813 (2023)
16. T. Pevný, T. Filler, P. Bas, Using high-dimensional image models to perform highly undetectable steganography. In: *International Workshop on Information Hiding*, (Springer, 2010), pp. 161–177
17. V. Holub, J. Fridrich, Designing steganographic distortion using directional filters. In: *2012 IEEE International Workshop on Information Forensics and Security (WIFS)*, (IEEE, 2012), pp. 234–239
18. B. Li, M. Wang, J. Huang, X. Li, A new cost function for spatial image steganography. In: *2014 IEEE International Conference on Image Processing (ICIP)*, (IEEE, 2014), pp. 4206–4210
19. V. Holub, J. Fridrich, Digital image steganography using universal distortion. In: *Proceedings of the First ACM Workshop on Information Hiding and Multimedia Security*, (ACM, 2013), pp. 59–68
20. L. Guo, J. Ni, Y.Q. Shi, Uniform embedding for efficient jpeg steganography. *IEEE Trans. Inf. Forensics Secur.* **9**(5), 814–825 (2014)
21. L. Guo, J. Ni, W. Su, C. Tang, Y.-Q. Shi, Using statistical image model for jpeg steganography: uniform embedding revisited. *IEEE Trans. Inf. Forensics Secur.* **10**(12), 2669–2680 (2015)
22. T. Denemark, J. Fridrich, Improving steganographic security by synchronizing the selection channel. In: *Proceedings of the 3rd ACM Workshop on Information Hiding and Multimedia Security*, (2015), pp. 5–14
23. B. Li, M. Wang, X. Li, S. Tan, J. Huang, A strategy of clustering modification directions in spatial image steganography. *IEEE Trans. Inf. Forensics Secur.* **10**(9), 1905–1917 (2015)
24. W. Zhang, Z. Zhang, L. Zhang, H. Li, N. Yu, Decomposing joint distortion for adaptive steganography. *IEEE Trans. Circuits Syst. Video Technol.* **27**(10), 2274–2280 (2016)
25. Z. Zhu, N. Zheng, T. Qiao, M. Xu, Robust steganography by modifying sign of DCT coefficients. *IEEE Access* **7**, 168613–168628 (2019)
26. Y. Zhang, X. Luo, C. Yang, D. Ye, F. Liu, A jpeg-compression resistant adaptive steganography based on relative relationship between DCT coefficients. In: *2015 10th International Conference on Availability, Reliability and Security*, (IEEE, 2015), pp. 461–466
27. Y. Zhang, X. Luo, C. Yang, F. Liu, Joint jpeg compression and detection resistant performance enhancement for adaptive steganography using feature regions selection. *Multimed. Tools Appl.* **76**(3), 3649–3668 (2017)
28. Y. Zhang, C. Qin, W. Zhang, F. Liu, X. Luo, On the fault-tolerant performance for a class of robust image steganography. *Signal Process.* **146**, 99–111 (2018)
29. J.-S. Tsai, W.-B. Huang, Y.-H. Kuo, M.-F. Horng, Joint robustness and security enhancement for feature-based image watermarking using invariant feature regions. *Signal Process.* **92**(6), 1431–1445 (2012)
30. J. Tao, S. Li, X. Zhang, Z. Wang, Towards robust image steganography. *IEEE Trans. Circuits Syst. Video Technol.* **29**(2), 594–600 (2018)
31. Z. Zhao, Q. Guan, H. Zhang, X. Zhao, Improving the robustness of adaptive steganographic algorithms based on transport channel matching. *IEEE Trans. Inf. Forensics Secur.* **14**(7), 1843–1856 (2018)
32. X. Yu, K. Chen, Y. Wang, W. Li, W. Zhang, N. Yu, Robust adaptive steganography based on generalized dither modulation and expanded embedding domain. *Signal Process.* **168**, 107343 (2020)
33. F. Li, K. Wu, C. Qin, J. Lei, Anti-compression jpeg steganography over repetitive compression networks. *Signal Process.* (2020). <https://doi.org/10.1016/j.sigpro.2020.107454>
34. P. Bas, T. Filler, T. Pevný, break our steganographic system: the ins and outs of organizing boss. In: *International Workshop on Information Hiding*, (Springer, 2011), pp. 59–70
35. T. Filler, J. Judas, J. Fridrich, Minimizing additive distortion in steganography using syndrome-trellis codes. *IEEE Trans. Inf. Forensics Secur.* **6**(3), 920–935 (2011)
36. J. Kodovsky, J. Fridrich, V. Holub, Ensemble classifiers for steganalysis of digital media. *IEEE Trans. Inf. Forensics Secur.* **7**(2), 432–444 (2011)
37. V. Holub, J. Fridrich, Low-complexity features for jpeg steganalysis using undecimated DCT. *IEEE Trans. Inf. Forensics Secur.* **10**(2), 219–228 (2014)
38. T. Pevný, J. Fridrich, Merging markov and dct features for multi-class jpeg steganalysis. In: *Security, steganography, and watermarking of multimedia contents IX*, International Society for Optics and Photonics vol. 6505, 650503 (2007).
39. J. Kodovský, J. Fridrich, Steganalysis of jpeg images using rich models. In: *Media Watermarking, Security, and Forensics 2012*, International Society for Optics and Photonics vol. 8303, 83030 (2012).
40. K. Sayood, *Introduction to data compression* (2017)
41. Z. Yin, L. Ke, Robust adaptive steganography based on dither modulation and modification with re-compression. *IEEE Trans. Signal Inf. Process.* **7**, 336–345 (2021)
42. W. Lu, J. Zhang, X. Zhao, W. Zhang, J. Huang, Secure robust jpeg steganography based on autoencoder with adaptive bch encoding. *IEEE Trans. Circuits Syst. Video Technol.* **31**(7), 2909–2922 (2020)
43. J. Butora, Y. Yousofi, J. Fridrich, How to pretrain for steganalysis. In: *Proceedings of the 2021 ACM Workshop on Information Hiding and Multimedia Security*, (2021), pp. 143–148
44. R. Cogranne, Q. Giboulot, P. Bas, The alaska steganalysis challenge: A first step towards steganalysis. In: *Proceedings of the ACM Workshop on Information Hiding and Multimedia Security*, (2019), pp. 125–137

## Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Tong Qiao** received the B.S. degree in Electronic and Information Engineering in 2009 from Information Engineering University, Zhengzhou, China, and the M.S. degree in Communication and Information System

in 2012 from Shanghai University, Shanghai, China, and the Ph.D. degree in University of Technology of Troyes, Laboratory of Systems Modelling and Dependability, Troyes, France, in 2016. Since 2020, he works as an Associate Professor in School of Cyberspace from Hangzhou Dianzi University. His current research interests focus on steganalysis and digital image forensics.

**Shengwang Xu** received the B.S. degree in Information Security from Hangzhou Dianzi University, Hangzhou, China, in 2021. Since 2021, he has been with the Laboratory of Internet and Network Security, Hangzhou Dianzi University. His current research interests focus on AI security.

**Shuai Wang** received the B.S. degree in network engineering from East China Jiaotong University, Nanchang, China, in 2018. Since 2018, he has been with the Laboratory of Internet and Network Security, Hangzhou Dianzi University. His current research interests focus on image steganography and steganalysis.

**Xiaoshuai Wu** received the B.S. degree in 2019 from Nanyang Institute of Technology and the M.S. degree in 2022 from Hangzhou Dianzi University. Currently, he is pursuing the Ph.D. degree at Hunan University. His research interests include data hiding and AI security.

**Bo Liu** received his Ph.D. degree from Xi'an Jiaotong University in 2010. He is currently a Researcher of Chinese Aeronautical Establishment. His research interest is AI security.

**Ning Zheng** received his M.S. degree from Zhejiang University, in 1987. He is currently a Full Professor of Hangzhou Dianzi University. His research interest is digital forensics.

**Ming Xu** received his M.S. and Ph.D. degrees from Zhejiang University, in 2000, and 2004, respectively. He is currently a Full Professor of Hangzhou Dianzi University. His research interest is digital forensics.

**Binmin Pan** received the B.S. degree in Information Security from Hangzhou Dianzi University, Hangzhou, China, in 2022. His current research interests focus on image steganography and steganalysis.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

---

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)

---