Robust Data Hiding for High Fidelity JPEG Images Over Social Networking Platforms

Rakesh Kumar, Savina Bansal, and R. K. Bansal

Abstract High-quality images are prevalent on social media, presenting opportunities for covert communication through image steganography. However, platforms often recompress images to save on bandwidth and storage, corrupting hidden data. This paper proposes a novel robust steganography method for high fidelity JPEG images. The technique meticulously selects discrete cosine transform (DCT) coefficients and uses an adaptive embedding algorithm that minimally modifies those coefficients to encode secret messages. For the test cases examined, the proposed approach demonstrates the ability to reliably recover hidden information with 96% accuracy even after recompression from quality 95 to 70, significantly outperforming prior arts. The method also provides goodness in imperceptibility and statistical undetectability across evaluated payload sizes. The results on benchmark datasets indicate potential of the proposed method for practical steganography applications.

Keywords JPEG · Robust steganography · Data hiding · SNP

1 Introduction

With the proliferation of multimedia, communication through images on social networking platforms (SNPs) has become a common practice. On average, studies indicate that over 14 billion images are collectively shared on various SNPs every day [1]. This presents both opportunities and challenges for covert communication through images. Steganography, the science of hiding secret information within digital medias like images, video, and audio, has attracted significant research attention in recent times [2, 3]. With images being prevalent, steganography offers covert communication through images. A secret message can be covertly transmitted through an insecure public channel without raising suspicions. This has diverse use cases spanning defence, intelligence agencies, cybercrime, secure communication, e-office, and copyright protection. As communication is happening over insecure

R. Kumar (✉) · S. Bansal · R. K. Bansal
Department of ECE, GZSCCET-MRSPTU, Bathinda, Punjab, India
e-mail: kumarrakesh0791@gmail.com

N. Hasteer et al. (eds.), *Adaptive Intelligence*, Lecture Notes in Electrical Engineering 1280, https://doi.org/10.1007/978-981-97-9045-6_29

channel, imperceptibility and undetectability become crucial. The generated stego image must be invisible to any visual detection and resistant to steganalysis tools that expose hidden data. Along with undetectability, reliability is another key requirement—the hidden data should survive impairments posed by post-processing like recompression and other channel effects during transmission. This is termed as robust image steganography.

The Joint Photographic Experts Group (JPEG) format is widely adopted over internet for image storage, owing to its efficient compression capabilities. Commonly used SNPs like Facebook, WhatsApp, LinkedIn, etc., often recompress the uploaded images with lower quality factors to save server storage and bandwidth. This recompression poses a significant challenge in achieving robust steganography. Therefore, designing steganography methods that are robust to JPEG recompression has become an active area of research. While most of the research in robust steganography concentrates on Upward Robustness (UR) [4–7], Downward Robustness (DR) [8] tends to be overlooked and requires more attention. This emphasis is crucial, especially considering that the recompression performed by SNPs falls under DR scenario. This paper introduces a novel robust JPEG steganography scheme tailored for DR scenario. The following sections review existing literature and its limitations, outline the proposed algorithm, describe experimental procedures, and present performance evaluations. Through this investigation, we aim to contribute to the field of robust steganography and advance the understanding of its potential applications in information security and privacy.

2 Related Work

With the proliferation of JPEG images, significant research efforts have focused on advancing steganographic techniques tailored for the JPEG format. JPEG steganography embeds secret messages by modifying the DCT coefficients in frequency domain. While traditional spatial domain approaches like LSB manipulation-based [9–11] can be adapted to the DCT domain, they are susceptible to statistical steganalysis attacks. Consequently, more sophisticated adaptive techniques were developed that dynamically adjust embedding based on image characteristics to evade detection. A major advancement in JPEG steganography came with the introduction of syndrome trellis codes (STCs) [12]. STCs utilize binary linear convolutional codes represented by a parity-check matrix to provide near-optimal embedding efficiency while minimizing distortion. This led to a popular framework involving the joint design of an STC-based embedding scheme and a distortion function. The distortion function calculates the cost of modifying each DCT coefficient to guide embedding. Many distortion functions have been proposed in literature [13–15], with UNIWARD [16] and UERD [17] being particularly effective and popular. These schemes balance the imperceptibility and payload constraint well while exhibiting resistance against steganalysis. However, a key limitation is that these algorithms assume the stego image will reach the receiver undistorted. This assumption fails

when sharing images over SNPs, recompression applied to uploaded images potentially corrupts any hidden data. To address this limitation, robust steganography algorithms have been developed.

Most existing DCT-modification-based robust steganography techniques are designed for UR scenario, where the cover image has lower quality factor than the recompressed image [4–7]. However, a common real-world use case is overlooked—users sharing high-quality digital photographs that get recompressed at lower quality by SNPs. This DR scenario presents distinct challenges that existing methods struggle to handle effectively. The literature reviewed underlines a research gap in current techniques, and to show this, we employ two well-known existing techniques- sign steganography (SS) [5] and dither modulation adaptive steganography (DMAS) [4], for comparative analysis. This work seeks to bridge this gap by proposing a novel robust steganography scheme for high-quality JPEG images shared on social networking platforms. Aiming to withstand the impending recompression at lower quality settings. The following sections presents the proposed methodology and comprehensive experimentation and evaluation.

3 Proposed Work

This section outlines proposed method for robust image steganography aimed at bolstering robustness against JPEG recompression. The approach involves two key steps: selection of robust cover elements and content-adaptive embedding.

3.1 Selection of Robust Cover Elements

A critical aspect of the proposed algorithm is the careful selection of cover elements. Through analysis of image characteristics and quantization tables, patterns and relationships are uncovered that guided in identifying robust cover locations $\boldsymbol{\Psi}$. Based on this, a selection criterion is employed represented by the binary function as

$$\boldsymbol{\Psi} = \begin{cases} 1 \text{ if } \operatorname{mod}\left(\boldsymbol{Q}^{(2)}, \boldsymbol{Q}^{(1)}\right) = 0 \text{ and } \operatorname{mod}\left(\frac{\boldsymbol{Q}^{(2)}}{\boldsymbol{Q}^{(1)}}, 2\right) = 1 \\ 0 \qquad\qquad\qquad\qquad \text{otherwise} \end{cases} \tag{1}$$

This function uses cover quantization table $\boldsymbol{Q}^{(1)}$ and prior knowledge of channel quantization table $\boldsymbol{Q}^{(2)}$ to identify optimal locations for embedding. It produces a binary 8×8 locations matrix, indicating robust DCT coefficients for embedding in each 8×8 DCT block. As the selection is solely based on quantization tables, it needs to be performed only once, and the matrix can be reused for all DCT blocks, reducing computational overhead. Notably, the knowledge of these selected locations is required later during extraction of hidden message.

3.2 Adaptive Embedding Algorithm

The embedding algorithm encodes the secret message bits into the selected DCT coefficients of cover image with minimal modifications. Represented as a binary string, the secret message $\mathbf{m}$ consists of randomly generated 0 s and 1 s. Algorithm 1, presented as pseudocode, outlines the proposed Content Adaptive Robust Embedding (CARE) algorithm. Utilizing a content-adaptive approach, CARE embeds the secret message within the judiciously selected DCT coefficients specified by $\boldsymbol{\Psi}$. It calculates embedding offsets ζ_i, and adjusts coefficients based on the secret message content. The resulting modified coefficients ($\hat{\mathbf{d}}^{\mathrm{C}}$) produce a robust stego image ($\mathbf{S}$).

Algorithm 1 - Content **A**daptive **R**obust **E**mbedding strategy

Input: Cover image ($\mathbf{C}$), robust locations $\boldsymbol{\Psi}$, message $\mathbf{m}$, quantization table $\mathbf{Q}^{(2)}$;
Output: Robust stego image ($\mathbf{S}$);

1. Obtain DCT coefficients $\mathbf{d}^{\mathrm{C}}$ from the cover image $\mathbf{C}$.
2. Select robust DCT coefficients $\mathbf{d}^{\mathrm{C}}_{\boldsymbol{\Psi}}$ using the locations specified by $\boldsymbol{\Psi}$.
3. Apply proposed embedding algorithm to get modified DCT coefficients $\hat{\mathbf{d}}^{\mathrm{C}}_{\boldsymbol{\Psi}}$:
   $\hat{\mathbf{d}}^{\mathrm{C}}_{\boldsymbol{\Psi}}$= CARE ($\mathbf{d}^{\mathrm{C}}_{\boldsymbol{\Psi}}, \mathbf{Q}^{(2)}, \mathbf{m}$)
   a. For each selected DCT coefficients $\mathrm{d}^{\mathrm{C}}_{\psi}$, calculate embedding offset ζ_i:
   $$\zeta_i = \mathrm{Q}_i^{(2)} \times \left(\frac{\mathrm{Q}_i^{(2)} - \mathrm{mod}\left(\mathrm{d}^{\mathrm{C}}_{\psi}, \mathrm{Q}_i^{(2)}\right)}{2\mathrm{Q}_i^{(2)}} \right)$$
   b. If the secret message content is 1:
   If $\mathrm{mod}\left(\left|\left[\mathrm{d}^{\mathrm{C}}_{\psi}, \mathrm{Q}_i^{(2)}\right]\right|\right) = 1$ **then** $\hat{\mathrm{d}}^{\mathrm{C}}_{\psi} = \mathrm{d}^{\mathrm{C}}_{\psi}$;
   Else, If $\mathrm{mod}\left(\left|\left\lfloor \mathrm{d}^{\mathrm{C}}_{\psi}, \mathrm{Q}_i^{(2)} \right\rfloor\right|\right) = 0$ **then** $\hat{\mathrm{d}}^{\mathrm{C}}_{\psi} = \mathrm{d}^{\mathrm{C}}_{\psi} + \zeta_i$;
   Else, If $\left|\left[\mathrm{d}^{\mathrm{C}}_{\psi}, \mathrm{Q}_i^{(2)}\right]\right| - \left|\left\lfloor \mathrm{d}^{\mathrm{C}}_{\psi}, \mathrm{Q}_i^{(2)} \right\rfloor\right| \geq Th$ **then** $\hat{\mathrm{d}}^{\mathrm{C}}_{\psi} = \mathrm{d}^{\mathrm{C}}_{\psi} - 1 + \zeta_i$;
   Else $\hat{\mathrm{d}}^{\mathrm{C}}_{\psi} = \mathrm{d}^{\mathrm{C}}_{\psi} - \zeta_i$;
   c. If the secret message content is 0:
   If $\mathrm{mod}\left(\left|\left[\mathrm{d}^{\mathrm{C}}_{\psi}, \mathrm{Q}_i^{(2)}\right]\right|\right) = 0$ **then** $\hat{\mathrm{d}}^{\mathrm{C}}_{\psi} = \mathrm{d}^{\mathrm{C}}_{\psi}$;
   Else, If $\left|\left[\mathrm{d}^{\mathrm{C}}_{\psi}, \mathrm{Q}_i^{(2)}\right]\right| - \left|\left\lfloor \mathrm{d}^{\mathrm{C}}_{\psi}, \mathrm{Q}_i^{(2)} \right\rfloor\right| \geq Th$ **then** $\hat{\mathrm{d}}^{\mathrm{C}}_{\psi} = \mathrm{d}^{\mathrm{C}}_{\psi} - 1 + \zeta_i$;
   Else, If $\mathrm{mod}\left(\left|\left\lfloor \mathrm{d}^{\mathrm{C}}_{\psi}, \mathrm{Q}_i^{(2)} \right\rfloor\right|\right) = 0$ **then** $\hat{\mathrm{d}}^{\mathrm{C}}_{\psi} = \mathrm{d}^{\mathrm{C}}_{\psi} + \zeta_i$;
   Else $\hat{\mathrm{d}}^{\mathrm{C}}_{\psi} = \mathrm{d}^{\mathrm{C}}_{\psi} - \zeta_i$;
4. Replace the modified DCT coefficients to obtain $\hat{\mathbf{d}}^{\mathrm{C}}$:
5. Generate the stego image $\mathbf{S}$ using $\hat{\mathbf{d}}^{\mathrm{C}}$.

At the receiver's end, the concealed message is retrieved by examining the parity of quantized DCT coefficients at specified locations $\boldsymbol{\Psi}$. Even parity denotes a hidden message of 0, while odd parity signifies a hidden message of 1. Subsequent sections evaluate the proposed method's performance through extensive experimentation and analysis.

4 Experimental Setup and Performance Metrics

For experiments, benchmark BOSSbase 1.01 image dataset [18] is used, containing 10,000 PGM images of size 512 × 512 pixels. 100 cover images are randomly selected and converted to JPEG format with quality factor (QF) 95, simulating high fidelity images typically shared on SNPs. Stego images are generated by embedding secret message using the proposed algorithm for varying payloads from 1 to 10 percent (P1-P10) of total non-zero AC DCT coefficients (nzAC). Existing schemes, DMAS [4] and SS [5] are used for comparison. To emulate JPEG recompression by SNPs, a MATLAB script is used to recompress stego images with quality factors ranging from 70 to 85 with step of 5. All experiments are performed on MATLAB platform, running on an i7 Intel Core processor with 8 GB RAM and results are averaged over 100 selected images.

To test goodness of the proposed scheme, robustness, undetectability, and imperceptibility are used as comparative performance metrics. *Robustness* measures the integrity of extracted message recovered from recompressed stego images and the embedded message. It is quantified as message extraction error rate (*MER*), given as

$$MER = \frac{number\ of\ incorreclty\ extracted\ bits}{length\ of\ embedded\ msg} \times 100 \tag{2}$$

Undetectability refers to the difficulty of distinguishing stego images from cover images using statistical analysis methods. To evaluate undetectability of the proposed method, an ensemble classifier [19] is employed in combination with DCTR [20] steganalyzer. The classification error rate (*CER*) between cover and stego images is calculated using Eq. (3) as

$$CER = \frac{E_{FP} + E_{FN}}{2} \tag{3}$$

Here, E_{FP} is the false positive error (misclassifying cover as stego image) and E_{FN} is the false negative error (mis-detecting a stego as cover). A higher *CER* is desirable as it implies stronger undetectability.

Imperceptibility is the ability of hiding data that is perceptually indistinguishable to human vision. Two well-known metrics PSNR and SSIM are used to test imperceptibility of the proposed work.

5 Performance Evaluation and Discussion

5.1 Robustness Evaluation

The proposed method demonstrates superior robustness against message extraction error rates compared to DMAS and SS steganography schemes, as shown in Fig. 1a–d. Particularly, in Fig. 1d, depicting the MER when the stego image (QF 95) is recompressed to QF 70, the proposed scheme maintains a low error rate of 3.4% for P1 and stabilizing at 3.2% for P10. In contrast, DMAS exhibits elevated error rates, with an overall MER ranging from 48 to 50% for payloads P1-P10. The SS scheme demonstrates MER fluctuations between 10 and 21% across payloads. These trends persist across other tested scenarios also. Clearly, even as payload size increases the proposed method consistently maintains low error rates, demonstrating its efficiency in handling varying data capacities without significant performance degradation.

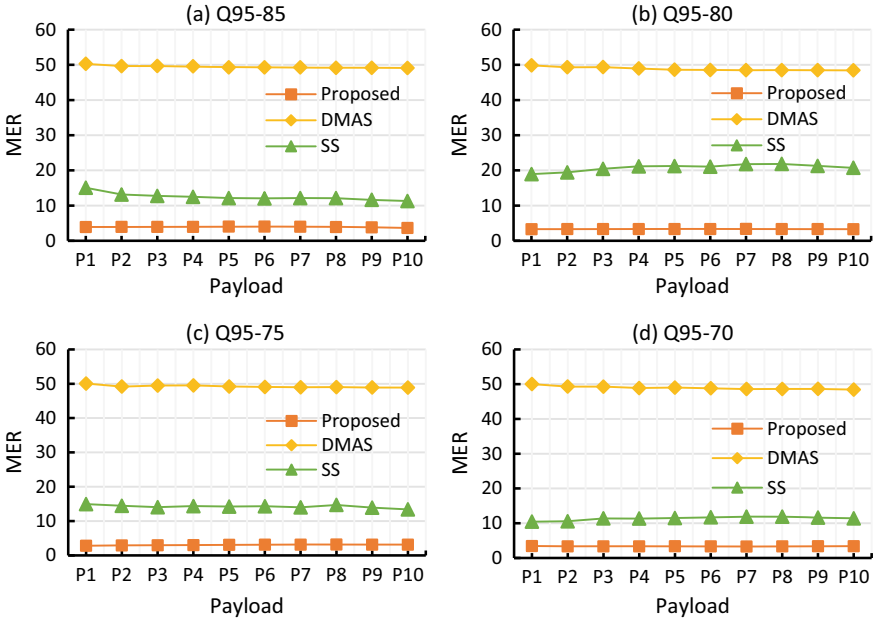

Fig. 1 Robustness evaluation of proposed algorithm for different tested scenarios

5.2 *Undetectability Evaluation*

The proposed method demonstrates strong statistical undetectability, clearly surpassing DMAS and SS schemes as payload size increases. As illustrated in Fig. 2a–d, the proposed approach sustains high classification error rates (CER) from P1 to P10. In QF 95–85 case (Fig. 2a), the proposed method achieves a CER of 0.615 at P1, slightly decreasing to 0.404 at P10. In contrast, GMAS CER drops from 0.455 at P1 to 0.23 at P10, while SS exhibits 0.683 to 0.22 CER. Similar trends are observed for other tested scenarios underscoring the proposed method's undetectability against steganalysis. Notably, the proposed method maintains high undetectability even as the payload increases, showcasing its adaptability and capability to be used in practical cases.

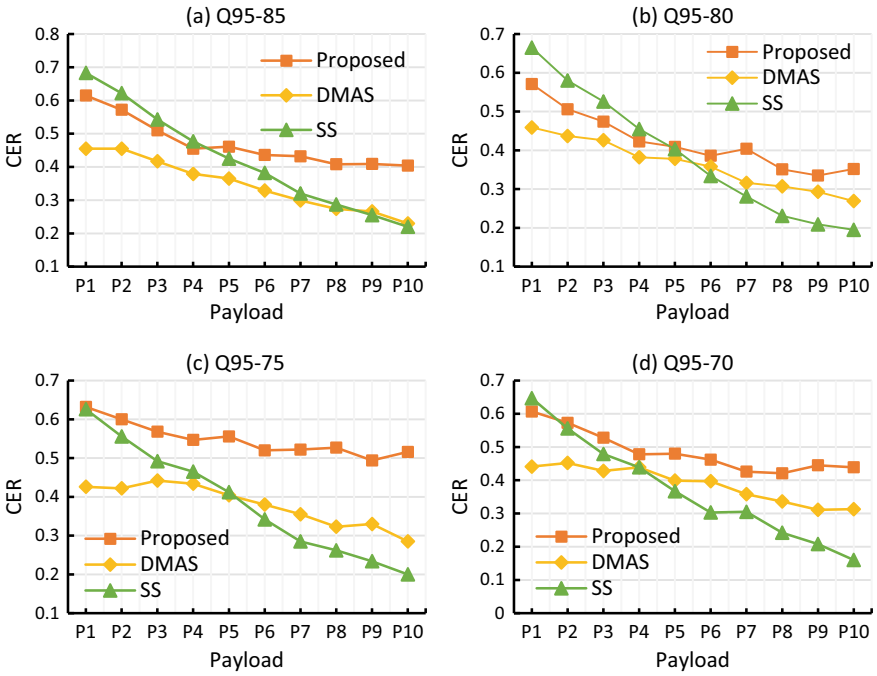

Fig. 2 Undetectability evaluation of proposed work for different tested scenarios

5.3 Imperceptibility Evaluation

The proposed method demonstrates higher imperceptibility across varying payloads. PSNR values, outlined in Table 1, attest to the method's consistent high performance. The proposed scheme maintains PSNR values, ranging from 66.8 dB at P1 payload to 57.5 dB at P10, surpassing the values exhibited by both DMAS (61.3 dB at P1 to 56.6 dB at P10) and SS (58.7 dB at P1 to 46.6 dB at P10) methods. Notably, the proposed algorithm achieves SSIM values of 0.9999 at P1 and 0.9993 at P10, outperforming both comparative schemes. This underscores proposed method's ability to embed confidential data with minimal discernible alterations to the cover image.

6 Conclusion

This paper introduces a novel robust JPEG steganography technique tailored for high fidelity images typically shared on social networking platforms. By strategically selecting robust coefficients and using content-adaptive embedding secure data concealment is ensured. Extensive experiments showcase proposed method's superiority in terms of robustness, imperceptibility, and undetectability. The technique maintains reliable message extraction, stronger undetectability, and high imperceptibility, validating its efficacy for practical steganography applications. Future work will focus on increasing payload capacity, improving resilience against advanced steganalyzers, and broadening evaluations with diverse datasets to better mimic real-world social media scenarios.

Table 1 PSNR and SSIM comparison of proposed algorithm

	Scheme	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
PSNR	DMAS	61.3	60.5	59.8	59.2	58.7	58.2	57.8	57.4	57.0	56.6
	SS	58.7	55.1	53.1	51.6	50.4	49.4	48.6	47.8	47.2	46.6
	Proposed	**66.8**	**63.8**	**62.1**	**60.8**	**59.9**	**59.2**	**58.7**	**58.2**	**57.8**	**57.5**
SSIM	DMAS	0.9997	0.9996	0.9996	0.9995	0.9994	0.9994	0.9993	0.9992	0.9991	0.9991
	SS	0.9997	0.9993	0.9989	0.9984	0.9979	0.9974	0.9969	0.9963	0.9957	0.9951
	Proposed	**0.9999**	**0.9998**	**0.9997**	**0.9997**	**0.9996**	**0.9995**	**0.9994**	**0.9994**	**0.9993**	**0.9993**

References

1. How many photos are there? Statistics & Trends (2023). https://photutorial.com/photos-statistics/. Accessed 20 Nov 2023
2. Rahman S, Uddin J, Zakarya M, Hussain H, Khan AA, Ahmed A, Haleem M (2023) A comprehensive study of digital image steganographic techniques. IEEE Access 11:6770–6791. https://doi.org/10.1109/ACCESS.2023.3237393
3. Sahu AK, Sahu M (2020) Digital image steganography and steganalysis: a journey of the past three decades. Open Comput Sci 10:296–342
4. Zhang Y, Zhu X, Qin C, Yang C, Luo X (2018) Dither modulation based adaptive steganography resisting jpeg compression and statistic detection. Multimed Tools Appl 77:17913–17935. https://doi.org/10.1007/s11042-017-4506-3
5. Zhu Z, Zheng N, Qiao T, Xu M (2019) Robust steganography by modifying sign of DCT coefficients. IEEE Access 7:168613–168628
6. Kumar R, Bansal S, Bansal RK (2023) A resilient approach to robust jpeg steganography: ensuring hidden data integrity. In: International conference on integrated intelligence and communication systems (ICIICS). IEEE, Kalaburagi, India, pp 1–6
7. Duan X, Li B, Yin Z, Zhang X, Luo B (2023) Robust image steganography against lossy jpeg compression based on embedding domain selection and adaptive error correction. http://arxiv.org/abs/2304.13297
8. Zeng K, Chen K, Zhang W, Wang Y, Yu N (2023) Robust steganography for high quality images. IEEE Trans Circuits Syst Video Technol:1–1
9. Kaur S, Bansal S, Bansal RK (2019) A data security approach based on steganography 3
10. Kaur, Bansal S, Bansal RK (2021) Image steganography for securing secret data using hybrid hiding model. Multimed Tools Appl 80:7749–7769
11. Kaur S, Bansal S, Bansal RK (2014) Steganography and classification of image steganography techniques. In: International conference on computing for sustainable global development (INDIACom). IEEE, New Delhi, India, pp 870–875
12. Filler T, Judas J, Fridrich J (2011) Minimizing additive distortion in steganography using syndrome-trellis codes. IEEE Trans Inf Forensic Secur 6:920–935
13. Bansal S, Bansal RK, Kumar R (2022) A novel upgraded uniform embedding technique for JPEG steganography. In: Nagar AK, Jat DS, Marín-Raventós G, Mishra DK (eds) Intelligent sustainable systems. Springer Nature, Singapore. pp 723–730. https://doi.org/10.1007/978-981-16-6309-3_68
14. Wang Y, Zhang W, Li W, Yu N (2021) Non-additive cost functions for jpeg steganography based on block boundary maintenance. IEEE Trans Inf Forensic Secur 16:1117–1130. https://doi.org/10.1109/TIFS.2020.3029908
15. Su W, Ni J, Li X, Shi Y-Q (2018) A new distortion function design for jpeg steganography using the generalized uniform embedding strategy. IEEE Trans Circuits Syst Video Technol 28:3545–3549. https://doi.org/10.1109/TCSVT.2018.2865537
16. Holub V, Fridrich J, Denemark T (2014) Universal distortion function for steganography in an arbitrary domain. EURASIP J Inf Secur 2014:1
17. Guo L, Ni J, Su W, Tang C, Shi Y-Q (2015) Using statistical image model for jpeg steganography: uniform embedding revisited. IEEE Trans Inf Forensic Secur 10:2669–2680. https://doi.org/10.1109/TIFS.2015.2473815
18. Bas P, Filler T, Pevný T (2011) Break our steganographic system: the ins and outs of organizing boss. In: Filler T, Pevný T, Craver S, Ker A (eds) Information hiding. Springer, Berlin, Heidelberg, pp 59–70
19. Kodovsky J, Fridrich J, Holub V (2012) Ensemble classifiers for steganalysis of digital media. IEEE Trans Inf Forensic Secur 7:432–444
20. Holub V, Fridrich J (2015) Low-complexity features for jpeg steganalysis using undecimated DCT. IEEE Trans Inf Forensic Secur 10:219–228