ENHANCING IMAGE STEGANOGRAPHY VIA STEGO GENERATION AND SELECTION

Tingting Song, Minglin Liu, Weiqi Luo, Peijia Zheng*

GuangDong Province Key Lab of Information Security Technology
School of Data and Computer Science, Sun Yat-sen University, Guangdong, China
{songtt3,liumlin6}@mail2.sysu.edu.cn, {luoweiqi,zhengpeijia}@mail.sysu.edu.cn

ABSTRACT

Unlike most existing steganography methods which are mainly focused on designing embedding cost, in this paper, we propose a new method to enhance existing steganographic methods via stego generation and selection. The proposed method firstly trains a steganalytic network according to the steganography to be enhanced, and then tries to adjust a tiny part of original embedding costs based on the magnitudes of it and the corresponding gradients obtained from the pre-trained network, and generates many candidate stegos in a random manner. Finally, the method selects a stego according to its image residual distance to cover. Extensive experimental results have shown that the proposed method can siginficantly enhance the security performance of current steganography in spatial domain against four steganalytic classifiers. In addition, comparative analysis between original stegos and the resulting ones with the proposed method are given.

Index Terms— steganography, stego generation, stego selection, adversarial example, steganalysis

1. INTRODUCTION

Typically, design of embedding cost is key issue for the modern steganography both in spatial and JPEG domains. In spatial domain, WOW [1] uses the output of directional high-pass filters to define pixel costs. UNIWARD [2] employs three wavelet directional filters to access texture information and defined costs with the relative changes of wavelet coefficients, and it can be effective both in spatial (S-UNIWARD [2]) and JPEG (J-UNIWARD [2]) domains. HILL [3] uses a high-pass filter and two low-pass filters to calculate cost function. A competitive method MiPOD [4] uses a locally-estimated multivariate Gaussian model to capture the non-stationary character of images. In JPEG domain, UERD [5] and J-UNIWARD [2] are two typical methods. UERD refines the uniform embedding by considering the relative changes of statistical model for digital images. The above mentioned methods are symmetric, meaning that the costs of two directions (i.e., ± 1) are exactly the same for every embedding unit. Recently, some asymmetric methods such as [6, 7, 8, 9] and CNN-based methods (e.g., ADV-EMB [10], MinMax [11] [12], JS-IAE [13]) have been proposed. These methods take the direction of modifications into account and further improve security. For instance, ADV-EMB [10] divides all DCT coefficients into two non-overlapping parts: one for traditional embedding; the other for adversarial embedding. JS-IAE [13] updates existing embedding cost iteratively based on adversarial examples derived from a series of pretrained networks. In addition, the method [14] improves existing symmetric steganography methods by constructing enhance covers. The method [15] modifies some embedding units in stego to enhance the steganography security.

In this paper, we propose a new method to enhance existing steganogrpahic methods via stego generation and selection. The proposed method is inspired by some steganographic methods based on adversarial example (e.g., [10] and [13]) and our previous work [15] based on stego post-processing. In our method, we firstly train a steganalytic network based on the steganography to be enhanced, and then adjust a random part of embedding costs according to the magnitudes of original costs and the corresponding gradients from the pre-trained network via back propagation. In such a way, we can generate many candidate stegos for a given cover. Finally, we select a stego as output which has the smallest residual distance to the corresponding cover. Extensive experimental results show that our method can significantly enhance most current steganographic methods in spatial domain.

The rest of this paper is arranged as follows. Section 2 describes the proposed steganography. Section 3 shows the comparative experimental results and discussions. Finally, the concluding remarks and future works are given in Section 4.

2. PROPOSED METHOD

As illustrated in Fig. 1, the proposed steganography includes three steps, that is steganalytic network training, stego generation and stego selection. We will separately describe them in three following subsections.

*Correspondence author. This work was supported in part by the National Science Foundation of China (61972430, 61672551), in part by the Natural Science Foundation of Guangdong (2019A1515011549)

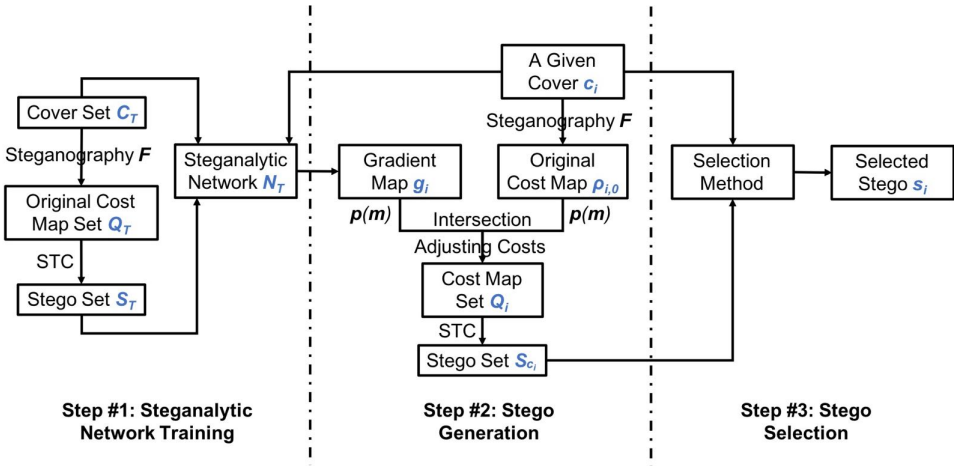


Fig. 1: The framework of the proposed method

2.1. Step #1: Steganalytic Network Training

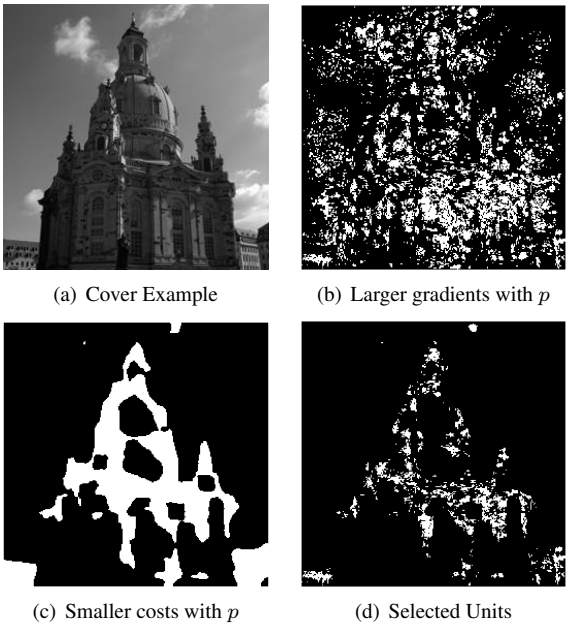
In this step, we firstly collect a cover set C_T, and then use a steganography F to calculate the origianl cost map set, denoted as Q_T. Based on the cost set Q_T, we can obtained the corresponding stego set S_T using STC. By training the cover set C_T and stego set S_T, we finally obtain a CNN based steganalytic network N_T.

2.2. Step #2: Stego Generation

For a cover c_i from the cover set C, where $C \cap C_T = \emptyset, i = 1, 2, 3, \ldots, |C|$, we firstly obtain its original embedding cost map $\rho_{i,0}$ according to the steganography F used in step #1, and then calculate the gradient map g_i via feeding c_i into the pre-trained network N_T in step #1. Finally, we will adjust a part (controlled by a random p, where $p \in [0, 1]$) of original costs m times independently. In each time (p is fixed and may different for each image), two factors are considered. **1) The magnitude of gradients (i.e., g_i).** Embedding units with larger gradients (with p) usually have greater impact on the detection performance of the network according to the properties of network gradient. **2) The magnitude of embedding costs (i.e., $\rho_{i,0}$).** Embedding units with smaller embedding costs (with p) usually introduce less detectable artifacts according to the design of embedding costs.

As illustrated in Fig. 2-(b) and Fig. 2-(c), embedding units with large gradients are usually not consistent with those units with smaller embedding costs. As shown in Fig. 2-(d), we just modify their intersection part in our method. After embedding unit selection, we will adjust embedding costs of selected units using the following formulas:

$$\rho^{+}_{i,j}(x,y) = \begin{cases} \rho^{+}_{i,0}(x,y) & g_i(x,y) < 0 \\ \rho^{+}_{i,0}(x,y) + \alpha & g_i(x,y) > 0 \end{cases} \quad (1)$$



(a) Cover Example (b) Larger gradients with p

(c) Smaller costs with p (d) Selected Units

Fig. 2: Embedding units selection. The payload is 0.4 bpp, $p = 0.2$ and steganography is HILL. In this example, the percentage of selected units (i.e. (d)) is 5.21%

$$\rho^{-}_{i,j}(x,y) = \begin{cases} \rho^{-}_{i,0}(x,y) + \alpha & g_i(x,y) < 0 \\ \rho^{-}_{i,0}(x,y) & g_i(x,y) > 0 \end{cases} \quad (2)$$

where $\rho^{+/-}_{i,j}(x,y)$ denotes the embedding cost at position (x,y) for the j^{th} ($j = 1, 2 \ldots m$) candidate stego; the superscript (i.e., $+$ or $-$) denotes the modification direction; $\alpha > 1$ is the adversarial intensity. Finally, we can obtain $m+1$ generated candidate stegos (including the original one) for any cover c_i, denoted as $s_{i,j} \in S_{c_i}, j = 0, 1, 2 \ldots m$.

Table 1: Average detection accuracy (%) in different steganographic and steganalytic cases. Those values with an asterisk (*) denote the better results in the corresponding cases

Method	Payload	SRM		MaxSRMd2		Deng-Net		SRNET	
		Original	Proposed	Original	Proposed	Original	Proposed	Original	Proposed
WOW	0.1 bpp	56.53	**55.03***	65.27	**61.07***	66.96	**61.78***	66.98	**62.81***
	0.2 bpp	63.82	**62.40***	72.46	**68.55***	77.14	**72.98***	76.43	**72.63***
	0.3 bpp	70.40	**68.20***	77.91	**74.55***	83.36	**80.23***	82.89	**79.33***
	0.4 bpp	76.26	**74.38***	81.97	**79.25***	87.68	**85.31***	86.85	**84.69***
MiPOD	0.1 bpp	54.56	**53.32***	56.24	**54.39***	58.06	**54.39***	58.57	**56.30***
	0.2 bpp	60.00	**58.63***	63.08	**59.53***	68.12	**63.38***	67.37	**63.29***
	0.3 bpp	65.34	**63.20***	68.18	**64.25***	74.85	**69.74***	73.86	**69.73***
	0.4 bpp	70.35	**67.91***	73.16	**69.37***	80.42	**75.93***	78.48	**74.59***
S-UNIWARD	0.1 bpp	55.86	**54.88***	59.61	**56.96***	61.93	**58.70***	61.75	**59.30***
	0.2 bpp	63.16	**61.85***	66.88	**63.21***	72.90	**69.47***	71.25	**68.21***
	0.3 bpp	70.01	**68.33***	72.44	**68.50***	80.69	**77.60***	78.68	**75.36***
	0.4 bpp	75.97	**74.02***	77.48	**73.98***	85.49	**83.97***	83.56	**82.20***
HILL	0.1 bpp	53.60	**52.72***	58.62	**55.33***	61.48	**55.32***	61.48	**56.89***
	0.2 bpp	59.45	**56.93***	65.02	**60.76***	69.96	**64.04***	69.48	**64.53***
	0.3 bpp	64.51	**62.84***	69.84	**65.91***	76.35	**71.77***	75.51	**71.15***
	0.4 bpp	70.10	**68.15***	74.57	**70.96***	80.95	**76.77***	80.03	**76.24***
CMD-HILL	0.1 bpp	52.36	**51.96***	56.71	**54.03***	58.08	**53.29***	59.06	**55.48***
	0.2 bpp	56.03	**55.21***	61.27	**58.12***	66.34	**60.25***	65.75	**61.67***
	0.3 bpp	60.04	**59.10***	65.26	**62.22***	72.19	**66.75***	71.04	**67.27***
	0.4 bpp	64.40	**63.61***	68.89	**66.43***	76.32	**72.13***	75.19	**70.64***

2.3. Step #3: Stego Selection

In step #2, we obtain $m+1$ candidate stegos for any cover c_i. In this step, will select a final stego s_i. The main idea of stego selection is to reduce the image residual distance between cover and the resulting stego. As it did in our previous work [15], we firstly apply a residual function to the cover c_i and $m+1$ candidate stegos to get their residuals $Res(c_i)$, $Res(s_{i,0})$, $Res(s_{i,1})$, ...$Res(s_{i,m})$ using three adaptive high-pass filters $\{B, B^T, B \otimes B^T\}$ of size 7. And then the Manhattan distances between the cover residual $Res(c_i)$ and stego residuals $Res(s_0)$, $Res(s_{i,1})$, ...$Res(s_{i,m})$ are calculated separately, denoted as $d_{i,0}, ..., d_{i,m}$. Finally, we select the stego with the smallest distance among them, denoted as s_i [1].

3. EXPERIMENTAL RESULTS

In our experiments, 20,000 gray-scale images are from BOSSBase-v1.01 [16] and BOWS2 [17]. We firstly resize all images into 256×256 using "imresize" in Matlab with default settings, and then divide them into two non-overlapping parts randomly. The first part contains 10,000 images, which are used to train a steganalytic network (Deng-Net [18] in our experiments) in step #1. The rest 10,000 images are used to generate stegos. Five typical steganographic methods in spatial domain are considered, i.e., WOW [1], MiPOD [4], S-UNIWARD [2], HILL [3] and CMD-HILL [6]. Two hand-crafted feaure sets (i.e., SRM [19], MaxSRMd2 [20]) and two CNN based networks (Deng-Net [18] and SRNet [21]) are used for security evaluation. In the testing stage, 5,000 cover-stego image pairs are used to retrain a classifer, while the rest 5,000 image pairs are used for security evaluation. **To achieve more convincing results, we randomly split the 20,000 images in steps #1-#3 three times and report the average results in our experiments. Furthermore, note that all the following results are evaluated on the re-trained Deng-Net [18] as well as other classifiers.**

There are three important parameters in our method, that is the number m of candidate stegos to be genereated, the adversarial density α, and the range of random parameter p. To obtain a better tradeoff between the effectiveness and time complexity, we fixed $m = 100$ and $\alpha = 2$, and select the parameter p randomly from the continuous uniform distribution $[0.025, 0.125], [0.025, 0.125] \times 2, [0.025, 0.125] \times 3, [0.025, 0.125] \times 4$ for the four payload 0.1, 0.2, 0.3 and 0.4 bpp (bits per pixel) respectively [2].

3.1. Security Performance

The comparative results are shown in Table 1. From Table 1, we obviously observe that the proposed method can improve the steganography security in all cases. Taking HILL at 0.1 bpp for instance, the detection accuracies are 58.62%

[1] Our source codes are available at: https://github.com/stt9621/StegViaStegoGenAndSelect

[2] Our algorithms are implemented on a computer sever with an Intel Core i7-6900K and 4 GPU NVIDIA TITAN X. In step #1, it takes around 8 hours to train a Deng-Net. In steps #2 & #3, it takes around 3 seconds to generate candidate stegos and select a final one.

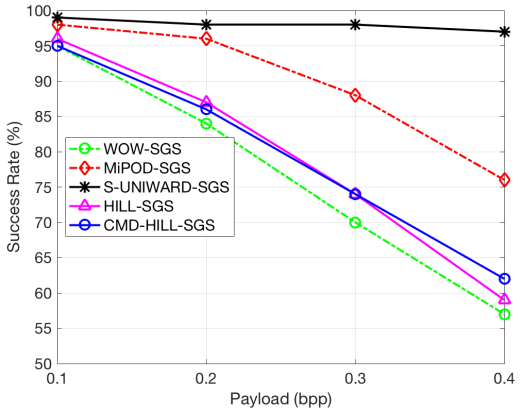


Fig. 3: Success rates for different steganographic methods

and 61.48% for the MaxSRMd2 and SRNET separately, while they become 55.33% and 56.89% after using the proposed method, which means that there are over 3.2% and around 4.5% improvement for the two cases. Note that such an improvement is significant for the modern steganography. Similar results can be found in other cases, including the asymmetric embedding method CMD-HILL [6].

3.2. Success Rate of Stego Generation

For a given cover $c_i \in C$, we generate $m+1$ candidate stegos in step #2, and select a final stego according to their residual distances to its corresponding cover in step #3. If the selected stego is not the original one $s_{i,0}$, we say that we generate a candidate stego with smaller residual successfully. Thus, we define the success rate of stego generation of the proposed method as follows:

$$R_s = \frac{\sum_{\forall c_i \in C} I(F - SGS(c_i) \neq s_{i,0})}{|C|} \tag{3}$$

where the symbol "F-SGS" denotes the enhanced version of the steganography F after using the proposed method via stego generation and selection (SGS). "F-SGS(c_i)" denotes the final selected stego for an input cover c_i ; I is an indicator function; $|C|$ is the number of elements in set C.

The success rates of four steganographic methods at different payloads are shown in Fig. 3. From Fig. 3, we have two following observations: 1) Typically, the success rates for the test steganographic methods would decrease with increasing the embedding payloads ranging from 0.1 bpp to 0.4 bpp. Taking WOW for instance, the success rate is around 95% at 0.10 bpp, while its becomes around 57% at 0.4 bpp. For S-UNIWARD-SGS, however, the success rates are over 95% for all payloads; 2) In addition, for a given payload, the success rates are different for different steganographic methods. The success rate is the highest for S-UNIWARD-SGS, while it is the lowest for the WOW-SGS.

Table 2: Average cost modification rates (%) for different steganography and payloads

	0.1 bpp	0.2 bpp	0.3 bpp	0.4 bpp
WOW-SGS	0.44	0.72	0.76	0.67
MiPOD-SGS	0.39	0.92	1.20	1.03
S-UNIWARD-SGS	0.50	1.38	2.32	3.04
HILL-SGS	0.47	0.78	0.81	0.73
CMD-HILL-SGS	0.38	0.60	0.62	0.63

3.3. Analysis on Cost Modification Rate

To enhance security, our method tries to adjust some parts of original costs with a random parameter p. In this section, we will provide some statistics about cost modification rates.

The average cost modification rates at different steganographic methods and payloads are shown in Table 2. From Table 2, we obviously observe that the modification rates are quite lower, all of them are smaller than 3.1%. In most cases, the modification rates are less than 1%, meaning that over 99% of original costs would not changed at all after using the proposed method, while the security performances can be significantly improved (refer to Table 1 for details).

4. CONCLUSION

In this paper, we propose a very promising method to enhance existing steganograpic methods in spatial domain. The main contributions are as follows.

- We introduce a new framework to enhance steganographic methods via stego generation and selection. The proposed framework is universal, and it is expected to be useful for enhancing other steganography, such as JPEG or audio/video steganography.
- Based on our experiments, we find that by adjusting a tiny part of original costs (e.g., less than 1% in most cases) properly, it is possible to achieve great security improvements for current steganographic methods.

This is our first attempt to combine stego generation and selection to enhance steganogrpahy. Since the proposed framework is flexible, there are many issues worth further studying. For instance, we would employ more pre-trained steganalytic networks (and/or classifers based on handcrafted features) to guide cost modification. We would analyze the statistical relationship between the gradients and embedding costs, and try other effective rules for selecting embedding units to be modified. Besides image residual distance, other rules for stego selection would be considered, such as using the steganalytic feature distance in a high dimension space and the predicted probability with some pre-trained steganalytic classifers.

5. REFERENCES

[1] Vojtech Holub and Jessica Fridrich, "Designing steganographic distortion using directional filters," *IEEE Workshop on Information Forensic and Security*, pp. 234–239, 2012.

[2] Vojtěch Holub, Jessica Fridrich, and Tomáš Denemark, "Universal distortion function for steganography in an arbitrary domain," *EURASIP Journal on Information Security*, vol. 2014, no. 1, pp. 1–13, 2014.

[3] Bin Li, Ming Wang, Jiwu Huang, and Xiaolong Li, "A new cost function for spatial image steganography," in *IEEE International Conference on Image Processing*, 2014, pp. 4206–4210.

[4] Vahid Sedighi, Remi Cogranne, and Jessica Fridrich, "Content-adaptive steganography by minimizing statistical detectability," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 2, pp. 221–234, 2016.

[5] Linjie Guo, Jiangqun Ni, Wenkang Su, Chengpei Tang, and Yunqing Shi, "Using statistical image model for jpeg steganography: Uniform embedding revisited," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 12, pp. 2669–2680, 2015.

[6] Bin Li, Ming Wang, Xiaolong Li, Shunquan Tan, and Jiwu Huang, "A strategy of clustering modification directions in spatial image steganography," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 9, pp. 1905–1917, 2015.

[7] Zichi Wang, Zhenxing Qian, Xinpeng Zhang, Min Yang, and Dengpan Ye, "On improving distortion functions for jpeg steganography," *IEEE Access*, vol. 6, pp. 74917–74930, 2018.

[8] Zichi Wang, Zhaoxia Yin, and Xinpeng Zhang, "Asymmetric distortion function for jpeg steganography using block artifact compensation," *International Journal of Digital Crime and Forensics*, vol. 11, no. 1, pp. 90–99, 2019.

[9] Xinzhi Yu, Kejiang Chen, Yaofei Wang, Weixiang Li, Weiming Zhang, and Nenghai Yu, "Robust adaptive steganography based on generalized dither modulation and expanded embedding domain," *Signal Processing*, vol. 168, 2020.

[10] Weixuan Tang, Bin Li, Shunquan Tan, Mauro Barni, and Jiwu Huang, "Cnn-based adversarial embedding for image steganography," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 8, pp. 2074–2087, 2019.

[11] Solène Bernard, Tomás Pevnỳ, Patrick Bas, and John Klein, "Exploiting adversarial embeddings for better steganography," *ACM Workshop on Information Hiding and Multimedia Security*, pp. 216–221, 2019.

[12] Solène Bernard, Patrick Bas, John Klein, and Tomás Pevnỳ, "Explicit optimization of min max steganographic game," *IEEE Transactions on Information Forensics and Security*, vol. 16, no. 2, pp. 812–823, 2021.

[13] Huaxiao Mo, Tingting Song, Bolin Chen, Weiqi Luo, and Jiwu Huang, "Enhancing jpeg steganography using iterative adversarial examples," *IEEE International Workshop on Information Forensics and Security*, pp. 1–6, 2019.

[14] Yiwei Zhang, Weiming Zhang, Kejiang Chen, Jiayang Liu, Yujia Liu, and Nenghai Yu, "Adversarial examples against deep neural network based steganalysis," *ACM Workshop on Information Hiding and Multimedia Security*, pp. 67–72, 2018.

[15] Bolin Chen, Weiqi Luo, and Jiwu Huang, "Universal stego post-processing for enhancing image steganography," *arXiv: Multimedia,* `http://arxiv.org/abs/1912.03878`, 2019.

[16] Patrick Bas, Tomáš Filler, and Tomás Pevnỳ, "Break our steganographic system: the ins and outs of organizing boss," *International Workshop on Information Hiding*, pp. 59–70, 2011.

[17] Patrick Bas and Teddy Furon, "Bows-2," `http://bows2.ec-lille.fr//`, 2007.

[18] Xiaoqing Deng, Bolin Chen, Weiqi Luo, and Da Luo, "Fast and effective global covariance pooling network for image steganalysis," *ACM Workshop on Information Hiding and Multimedia Security*, pp. 230–234, 2019.

[19] Jessica Fridrich and Jan Kodovsky, "Rich models for steganalysis of digital images," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 3, pp. 868–882, 2012.

[20] Tomas Denemark, Vahid Sedighi, Vojtech Holub, Remi Cogranne, and Jessica Fridrich, "Selection-channel-aware rich model for steganalysis of digital images," *IEEE International Workshop on Information Forensics and Security*, pp. 48–53, 2014.

[21] Mehdi Boroumand, Mo Chen, and Jessica Fridrich, "Deep residual network for steganalysis of digital images," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 5, pp. 1181–1193, 2019.