Forensics of Content Adaptive Steganography Techniques

SAURABH AGARWAL
Amity School of Engineering & Technology, Amity University Uttar Pradesh, Noida, India
Department of Cyber Security, Kyungil University, 50 Gamasil-gil, Hayang-eup, Gyeongsan-si, Gyeongbuk 38428, Republic of Korea

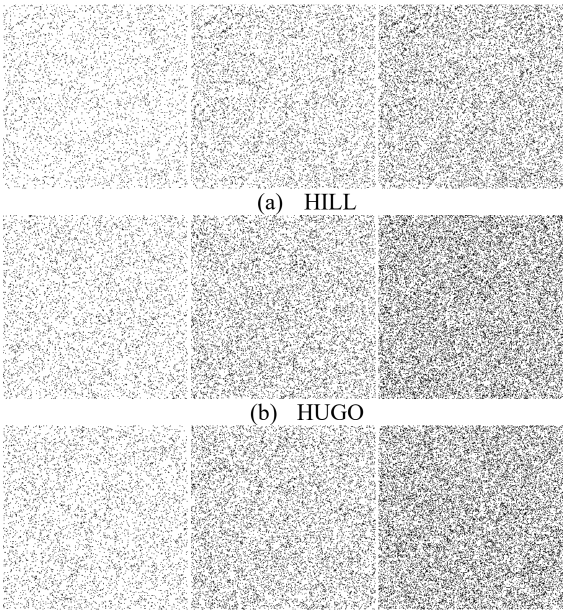
KI-HYUN JUNG
Department of Cyber Security, Kyungil University, 50 Gamasil-gil, Hayang-eup, Gyeongsan-si, Gyeongbuk 38428, Republic of Korea

Digital images are versatile due to their universal acceptability. Sometimes image are used to send secret messages. Several steganographic techniques exist to hide the messages in unnoticeable way. Out of several categories content adaptive steganographic techniques are quite popular. Content adaptive techniques hide the message in edge and texture regions so that it cannot be traced easily. However, steganalysis is performed to verify the robustness of steganographic techniques. In some case steganalysis is also applied to find the secret message that is sent for wrong purpose. Traditional feature extraction and deep learning based techniques are available. In feature extraction, co-occurrence matrix based steganalysis techniques are robust against content adaptive steganographic techniques. Some techniques also exploited the texture operators with co-occurrence for steganalysis. Most of the existing deep learning model requires high end computing resources due to large number of layers in network and preprocessing requirements. Therefore, there is need of better feature extraction techniques and low computational deep network.

Additional Keywords and Phrases: Steganography, Steganalysis, Image forensics

1 INTRODUCTION

Steganography techniques are applied to hide the information in the image. Steganalysis is performed to decide whether image is stego or cover. In many cases, blind steganalysis is performed as the steganography technique is unknown. Blind or universal image steganalysis can be performed on any type of image steganography technique. Universal image steganalysis can also be used to assess the robustness of steganography techniques. Numerous type of steganography techniques exist based on different approaches. In Least Significant Bits (LSB) based methods, message is hide in LSB's. Presently, content adaptive techniques are more popular. In these techniques, message is concealed in texture, edge or dense regions where detection of message is challenging. Some popular content adaptive steganography techniques are Highly Undetectable steGO (HUGO) [1], Wavelet Obtained Weights (WOW) [2], Spatial-UNIversal WAvelet Relative Distortion (S-UNIWARD) [3], (High-pass, Low-pass, and Low-pass (HILL) [4], and Minimizing the Power of Optimal Detector (MiPOD) [5]. Techniques [1-4] follow syndrome trellis code (STC) for information hiding. STC decreases the embedding distortion. HUGO [1] has large message embedding capacity around seven times than LSB matching methods. HUGO technique is derived from co-occurrence of pixels as considered in SPAM. WOW [2] technique can embed two times more information than HUGO. Multiple directional filters provide the information of edges and texture regions. The embedding is performed on these detected regions. UNIWARD [3] is suggested for spatial (S-UNIWARD) and JPEG (J-UNIWARD) domain. UNIWARD technique is based on residual of wavelet filters. HILL [4] high-pass filter recognizes the unusual predictable regions in the image. Low cost embedding is executed in textural areas by applying two low pass filters. HILL is faster than HUGO, UNIWARD and WOW. MiPOD [5] tries to capture non-stationary properties of image to minimize the effect of message embedding. Gaussian distribution is considered to reduce the difference between cover and stego image. A blank image (white) of size 256x256 pixels is considered to see the effect of HILL, HUGO and S-UNIWARD with different payloads. In first row of Fig. 1 HILL steganography effect with payloads 0.4, 0.6, and 0.8 is shown (left to right). Similarly for HUGO and S-UNIWARD, stego images are displayed in second and third rows, respectively. The stego grayscale image is converted into binary image for better visibility. The changes in pixels are performed uniformly if the image region is smooth.



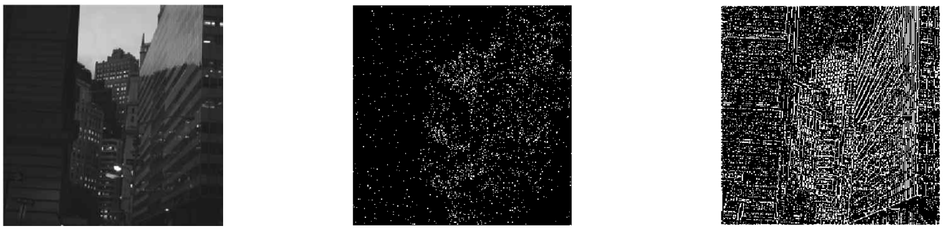
(a) HILL

(b) HUGO

(c) S-UNIWARD

Figure 1: HILL, HUGO and S-UNIWARD images with payload 0.4, 0.6, 0.8

The main attribute of steganography techniques is to hide information in noticeable way and in random order by changing pixel value by +1 or -1. However, in other image operations like median filtering the structure of the image can be visualized in median residual array. In Figure 2 (a) an image is taken from BOWS2 dataset. The residual of S-UNIWARD (payload=0.4) image with Figure 2(a) is shown in Figure 2(b).The residual of median filtered image (3x3) with Figure 2(a) is displayed in Figure 2(c).



(a) An image (b) Residual of S-UNIWARD (c) Residual of median filtering

Figure 2: Image and its residuals

2 STEGANOGRAPHY FORENSICS

Steganalysis techniques can be broadly divided into classical feature extraction based and deep learning based techniques. Co-occurrence matrix based [6-9], co-occurrence matrix and texture operator hybrid [10-12] are popular as feature extraction techniques. Convolutional neural network based steganalysis techniques [13-23] demand high computational resources. The realization of results in CNN is difficult as multiple parameters need to be tuned. Since the dimension of feature vector is very large in feature extraction based techniques [6-12], Some classifiers [24, 25] are suggested to handle large size feature dimension especially for steganalysis. Kodovský and Fridrich [24] proposed the ensemble classifier to reduce the training complexity. Multiple based learners work independently to classify. The base learner randomly selects a subspace for the feature space. The final results are calculated by aggregating results of each subspace. Cogranne et al. [25] suggested that Fisher Linear Discriminant can also provide good results with low computation using optimization techniques and ridge regression regularization. In Penvy et al. [6]'s method, SPAM steganalysis technique considered the difference array in multiple directions. The co-occurrence between difference array elements is calculated. SPAM is also found suitable in other classification tasks. SRM [7] and its variants [8, 9] are the backbone of steganalysis. Most of the existing feature extraction based techniques and deep learning based techniques directly or indirectly utilizes the SRM. Multiple linear and non-linear high-pass filters are applied for better statistical information. 106 sub-models are considered for extracting information in total. Feature extraction based hybrid technique [11, 12] suggested the threshold local binary pattern (TLBP) and combined TLBP features with SRM features for better results. TLBP considers the some sub-models of SRM while considering neighbors. Bernoulli distribution is taken into consideration while deciding the threshold value. Rotation invariant uniform LBP is considered for concise and better feature set. Generally, operators are applied on the whole image uniformly. However, the steganography techniques hide the message irregularly in the image. In the first time, CNN model i.e., Gaussian-Neuron CNN [13, 14] was proposed for steganalysis by Qian et al. The effectiveness of the proposed model was verified for HUGO, S-UNIWARD and WOW techniques. The images are preprocessed with one high pass kernel of size 5x5 before inserting to the CNN. However, the performance of the CNN method is inferior to SRM. Multiple types of layers like as convolutional, pooling, and fully connected are utilized. In similar type of work [15], large dataset is considered for experimental analysis. Method claims improvement in performance by using the one embedding key in network training. Xu et al. [16] applied the 5x5 kernel [13, 14] for preprocessing. One new layer i.e., absolute activation (ABS) layer is introduced to consider the symmetry of residuals. Despite conventional ReLU layer, tanh layer is utilized. Detailed experimental analysis is given for S-UNIWARD and HILL [4]. Wu et al. [17, 18] proposed the deep network with convolutional, max pooling, batch normalization, ReLU and fully connected layers. The images are preprocessed by same filter as in [13, 14] and further residual arrays are considered as input to the network. The authors claim better detection capability than SRM and previous CNN techniques. Yuan et al. [19] applied three high pass filters to improve the detection accuracy. Ye et al. [20] technique proved the effectiveness on WOW, S-UNIWARD, and HILL techniques. Image preprocessing is performed using SRM filters and their residuals are considered in CNN. In spite of the conventional ReLU, the truncated linear unit is utilized in CNN. The accuracy is improved by measuring the probability of each pixel embedding i.e., selection channel. A popular network Steganalysis Residual Network (SRNet) [21] is proposed by Boroumand et al. for S-UNIWARD, WOW, HILL, J-UNIWARD, and UED-JC [26] steganography algorithm. The network contains broadly three types of layers. First type (layer 1 & 2) of layers doesn't have any pooling and residual shortcut. Second type of layers (layers 3–7) has residual shortcuts with no pooling. In third type of layers (layers 8–11), pooling and residual shortcuts both are considered. In last, statistical moments are calculated and passed to the classifier. The performance is verified on both uncompressed and compressed images. Yedroudj et al. [22] suggested a network similar to techniques [16, 20] for S-UNIWARD, and WOW. However, thirty SRM filters are used for preprocessing to improve the performance significantly. Wu et al. [27] introduced the shared batch normalization layer to extract better statistical information. Images are preprocessed using twenty SRM kernels. Kim et al. [23] increased the stego weak signal strength by additional embedding. Images are preprocessed using Qian's kernel and followed by dual CNN. In deep networks, the requirement of preprocessing using SRM filters is compulsory in most of the methods. Also multiple layers and residual connections increased the computational requirements exponentially. In the next section, SPAM and SRM feature extraction techniques are applied for analyzing the content adaptive steganographic techniques.

3 EXPERIMENTAL RESULTS

In this section, some experiments are shown for steganalysis using SPAM and SRM. SPAM and SRM generate the feature vectors of size 686 and 34,671, respectively. Experiments are performed on BOWS2 database [28]. BOWS2 is mostly used in steganalysis and contains 10,000 images of size 512x512 pixels. In experiments, 5,000 images are taken form BOWS2 database randomly. Block of size 256x256 pixels are cropped from original images in random order. Therefore, 5,000 cover images and 5,000 stego images are considered in each set. Four content adaptive algorithms i.e., HUGO, WOW, S-UNIWARD and HILL are analyzed with payload 0.4, 0.6 and 0.8 bit per pixel (bpp). In training set, 3,500 cover and 3,500 stego images are taken. Testing set contain remaining 1,500 cover images and 1,500 stego images. Two classifiers abbreviated as ENSE [24] and LCL [25] are used for classification. Results are shown in terms of percentage detection accuracy. In Figure 3, results are shown for SPAM steganalysis technique. As the payload decreases, the detection becomes more challenging. On HILL steganography with 0.6 bpp payload, the detection accuracies are 66.40, and 66.75 with ENSE and LCL classifiers respectively. The detection of S-UNIWARD steganography is less challenging when compared wtih other techniques. The detection of HILL steganography is comparatively more difficult.

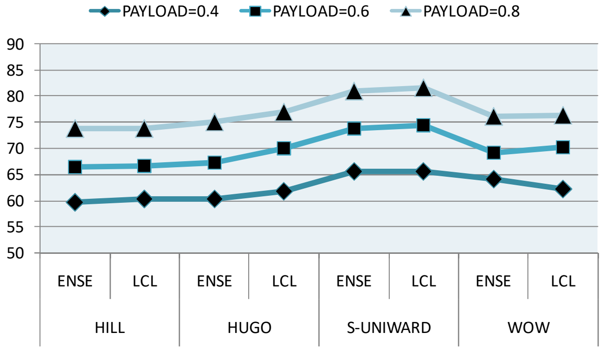


Figure 3: SPAM steganalysis

As evident from Figure 4, SRM gives the better performance than SPAM. However, detection performance depends upon the payload. Still there is significant improvement. For payload 0.4 bpp, the detection of HILL steganography is most challenging than other steganography techniques. Detection performance SRM on HUGO, S-UNIWARD and WOW steganography techniques is similar.

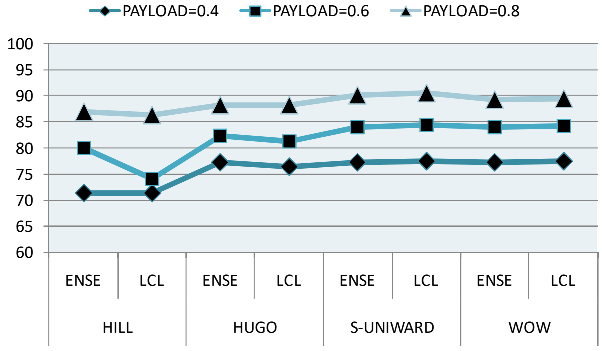


Figure 4: SRM steganalysis

In Figure 3 and 4, it can be seen that there is significant difference between SPAM and SRM detection accuracies for payload 0.4 bpp. More than 11% difference is occurred in the detection accuracies of both the techniques. For payload 0.6 bpp, more than 10% difference is given in the detection accuracies of both techniques except in one case. SRM gives 84.55% detection accuracy for S-UNIWARD using LCL classifier. For payload 0.8 bpp, there is 9% difference in the detection accuracies of SPAM and SRM. SRM gives 86.90% detection accuracy for HILL using ENSE classifier.

4 CONCLUSION

In this paper, some popular content adaptive steganography techniques have been analyzed. In comparison with other steganography techniques, the content adaptive techniques were more robust. High dimensional feature vector or very deep convolutional neural network was required to classify cover and content adaptive stego images. SRM steganalysis technique gave satisfactory results. The size of feature vector of SRM was very large like as 34,671. In the most of the deep learning methods, preprocessing requirement with SRM high pass filters was necessary. Some experiments have been performed on four content adaptive steganography techniques. HILL steganography technique has been found most challenging for forensics. It can be concluded that there is scope of developing more robust texture operators based techniques for steganalysis and explainable deep learning need to be incorporated to understand complex deep learning networks.

ACKNOWLEDGMENTS

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2018R1D1A1A09081842) and Korea Research Fellowship Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (2019H1D3A1A01101687).

REFERENCES

[1] T. Pevný, T. Filler, and P. Bas, “Using High-Dimensional Image Models to Perform Highly Undetectable Steganography,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6387 LNCS, 2010, pp. 161–177.

[2] V. Holub and J. Fridrich, “Designing steganographic distortion using directional filters,” in *WIFS 2012 - Proceedings of the 2012 IEEE International Workshop on Information Forensics and Security*, 2012, pp. 234–239, doi: 10.1109/WIFS.2012.6412655.

[3] V. Holub, J. Fridrich, and T. Denemark, “Universal distortion function for steganography in an arbitrary domain,” *EURASIP J. Inf. Secur.*, vol. 2014, no. 1, p. 1, Dec. 2014, doi: 10.1186/1687-417X-2014-1.

[4] B. Li, M. Wang, J. Huang, and X. Li, “A new cost function for spatial image steganography,” in *2014 IEEE International Conference on Image Processing, ICIP 2014*, 2014, pp. 4206–4210, doi: 10.1109/ICIP.2014.7025854.

[5] V. Sedighi, R. Cogranne, and J. Fridrich, “Content-Adaptive Steganography by Minimizing Statistical Detectability,” *IEEE Trans. Inf. Forensics Secur.*, vol. 11, no. 2, pp. 221–234, Feb. 2016, doi: 10.1109/TIFS.2015.2486744.

[6] T. Pevny, P. Bas, and J. Fridrich, “Steganalysis by Subtractive Pixel Adjacency Matrix,” *IEEE Trans. Inf. Forensics Secur.*, vol. 5, no. 2, pp. 215–224, Jun. 2010, doi: 10.1109/TIFS.2010.2045842.

[7] J. Fridrich and J. Kodovsky, “Rich Models for Steganalysis of Digital Images,” *IEEE Trans. Inf. Forensics Secur.*, vol. 7, no. 3, pp. 868–882, Jun. 2012, doi: 10.1109/TIFS.2012.2190402.

[8] T. Denemark, V. Sedighi, V. Holub, R. Cogranne, and J. Fridrich, “Selection-channel-aware rich model for Steganalysis of digital images,” in *2014 IEEE International Workshop on Information Forensics and Security, WIFS 2014*, 2015, pp. 48–53, doi: 10.1109/WIFS.2014.7084302.

[9] W. Tang, H. Li, W. Luo, and J. Huang, “Adaptive steganalysis against WOW embedding algorithm,” in *Proceedings of the 2nd ACM workshop on Information hiding and multimedia security - IH&MMSec ’14*, 2014, pp. 91–96, doi: 10.1145/2600918.2600935.

[10] Y. Q. Shi, P. Sutthiwan, and L. Chen, “Textural features for steganalysis,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7692 LNCS, 2013, pp. 63–77.

[11] B. Li, Z. Li, S. Zhou, S. Tan, and X. Zhang, “New Steganalytic Features for Spatial Image Steganography Based on Derivative Filters and Threshold LBP Operator,” *IEEE Trans. Inf. Forensics Secur.*, vol. 13, no. 5, pp. 1242–1257, May 2018, doi: 10.1109/TIFS.2017.2780805.

[12] P. Wang, F. Liu, and C. Yang, “Towards feature representation for steganalysis of spatial steganography,” *Signal Processing*, vol. 169, p. 107422, Apr. 2020, doi: 10.1016/j.sigpro.2019.107422.

[13] Y. Qian, J. Dong, W. Wang, and T. Tan, “Deep learning for steganalysis via convolutional neural networks,” in *Media Watermarking, Security, and Forensics 2015*, 2015, p. 94090J, doi: 10.1117/12.2083479.

[14] Y. Qian, J. Dong, W. Wang, and T. Tan, “Learning and transferring representations for image steganalysis using convolutional neural network,” in *Proceedings - International Conference on Image Processing, ICIP*, 2016, doi: 10.1109/ICIP.2016.7532860.

[15] L. Pibre, J. Pasquet, D. Ienco, and M. Chaumont, “Deep learning is a good steganalysis tool when embedding key is reused for different images, even if there is a cover sourcemismatch,” *IS T Int. Symp. Electron. Imaging Sci. Technol.*, vol. 2016, no. 8, pp. 1–11, Feb. 2016, doi: 10.2352/ISSN.2470-1173.2016.8.MWSF-078.

[16] G. Xu, H.-Z. Wu, and Y.-Q. Shi, “Structural Design of Convolutional Neural Networks for Steganalysis,” *IEEE Signal Process. Lett.*, vol. 23, no. 5, pp. 708–712, May 2016, doi: 10.1109/LSP.2016.2548421.

[17] S. Wu, S. H. Zhong, and Y. Liu, “Steganalysis via deep residual network,” in *Proceedings of the International Conference on Parallel and Distributed Systems - ICPADS*, 2016, vol. 0, pp. 1233–1236, doi: 10.1109/ICPADS.2016.0167.

[18] S. Wu, S. Zhong, and Y. Liu, “Deep residual learning for image steganalysis,” *Multimed. Tools Appl.*, vol. 77, no. 9, pp. 1–17, May 2017, doi: 10.1007/s11042-017-4440-4.

[19] Y. Yuan, W. Lu, B. Feng, and J. Weng, “Steganalysis with CNN Using Multi-channels Filtered Residuals,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017, pp. 110–120.

[20] J. Ye, J. Ni, and Y. Yi, “Deep Learning Hierarchical Representations for Image Steganalysis,” *IEEE Trans. Inf. Forensics Secur.*, vol. 12, no. 11, pp. 2545–2557, Nov. 2017, doi: 10.1109/TIFS.2017.2710946.

[21] M. Boroumand, M. Chen, and J. Fridrich, “Deep Residual Network for Steganalysis of Digital Images,” *IEEE Trans. Inf. Forensics Secur.*, vol. 14, no. 5, pp. 1181–1193, May 2019, doi: 10.1109/TIFS.2018.2871749.

[22] M. Yedroudj, F. Comby, and M. Chaumont, “Yedroudj-Net: An Efficient CNN for Spatial Steganalysis,” in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, vol. 2018-April, pp. 2092–2096, doi: 10.1109/ICASSP.2018.8461438.

[23] J. Kim, H. Park, and J. Il Park, “CNN-based image steganalysis using additional data embedding,” *Multimed. Tools Appl.*, vol. 79, no. 1–2, pp. 1355–1372, Jan. 2020, doi: 10.1007/s11042-019-08251-3.

[24] J. Kodovský and J. Fridrich, “Steganalysis in high dimensions: fusing classifiers built on random subspaces,” in *Media Watermarking, Security, and Forensics III*, 2011, p. 78800L, doi: 10.1117/12.872279.

[25] R. Cogranne, V. Sedighi, J. Fridrich, and T. Pevny, “Is ensemble classifier needed for steganalysis in high-dimensional feature spaces?,” in *2015 IEEE International Workshop on Information Forensics and Security (WIFS)*, 2015, pp. 1–6, doi: 10.1109/WIFS.2015.7368597.

[26] Linjie Guo, Jiangqun Ni, and Yun Qing Shi, “Uniform Embedding for Efficient JPEG Steganography,” *IEEE Trans. Inf. Forensics Secur.*, vol. 9, no. 5, pp. 814–825, May 2014, doi: 10.1109/TIFS.2014.2312817.

[27] S. Wu, S. Zhong, and Y. Liu, “A Novel Convolutional Neural Network for Image Steganalysis With Shared Normalization,” *IEEE Trans. Multimed.*, vol. 22, no. 1, pp. 256–270, Jan. 2020, doi: 10.1109/TMM.2019.2920605.

[28] P. Bas and T. Furon, “Break Our Watermarking System,” *Available http//bows2.ec-lille.fr/ 2nd*, 2008.