BMP Color Images Steganographer Detection Based on Deep Learning

Shuaipeng Yang[1], Yang Yu[1(✉)], Xiaoming Liu[2], Hong Zhang[3], and Haoyu Wang[1]

[1] Beijing University of Posts and Telecommunications, Beijing, China
{2018140769,yangyu}@bupt.edu.cn
[2] CNCERT/CC Chaoyang District, Beijing, China
liuxm@cert.org.cn
[3] Tianjin Branch of CNCERT/CC Nankai District, Tianjin, China
zhangh@cert.org.cn

Abstract. A user who achieves covert communication by embedding secret information in the original image is called steganographer. Steganographer detection determines which user sent a secured image with a secret message. Existing steganographer detection algorithms take gray images as the main research content. To better adapt to the reality, we propose a WiserNet-based steganograph detection algorithm for the characteristics of BMP color images, and the process is divided into the following three steps: feature extraction through each channel convolution structure, prevent the conventional convolution structure destroy the correlation between the color image channel operation, reduce the number of the extraction of feature dimension. The use of a per-channel convolution structure makes it easier to extract color image features, and the low-dimensional feature vector reduces the time required for subsequent clustering algorithms, which improves the efficiency of steganographer detection. Simulation experiments are conducted for the classification of feature extractors, detection of different steganographic rates, and detection of different image scales. First, the steganalysis binary classification results of this algorithm are compared with similar algorithms, and the classification accuracy is 84.90% when the steganalysis rate is 0.4 BPC, which is 1.11% higher than Ye-Net and 0.83% higher than Xu-ResNet. Since there is very little published research on steganography detection of color images, four feature extractors, Ye-Net, Xu-ResNet, SRNet, and WiserNet, will be used in this experiment to replace the WiserNet-100 feature extractor in the steganography detection algorithm. The results show that the detection accuracy of the algorithm proposed in this paper reaches 93% when the embedding rate is 0.2 BPC, and the detection accuracy reaches 100% when the embedding rate is greater than 0.2 BPC. The steganographic detection accuracy reaches 84% when the graph scale is 60% and the steganographic rate is 0.2 BPC. In terms of detection time, the WNCISD-100 is 7.79 s, which is 50% less time-consuming compared to SRSD.

Keyword: WiserNet · BMP color image · Steganographer detection · DBSCAN

This work is supported by the National Key R&D Program of China under Grant No. 2016YFB0801004.

J. Xiong et al. (Eds.): MobiMedia 2021, LNICST 394, pp. 602–612, 2021.
https://doi.org/10.1007/978-3-030-89814-4_44

1 Introduction

Multimedia, such as texts, images, audio, and video, is considered to be one of the most effective mediums of communication and information sharing [1]. Image steganography ensures secret communication by embedding secret information in ordinary images [2, 3]. The most commonly used steganographic medium is the digital image [4]. As opposed to image steganography, image steganalysis is the art of revealing the secret information that is hidden by the steganographer in the images.

Currently, most image steganalysis techniques follow the pattern of separating suspicious images into overlay images or steganographic images. This problem is known as the steganographer detection problem. With the large amount of image data generated by social media and the development of steganographic algorithms, some users attempt to deliver secret information by image steganography through innocent users of social networks [5–7]. The detection performance of traditional image steganalysis can be significantly degraded. The effect of information theft forensics is not evident. Therefore, an increasing number of scholars have shifted the focus of image security detection from image steganalysis to locating the user who first sent the data carrying the hidden image, i.e., the steganographer, or criminals [8–10]. How to locate those users who send digital steganographic images carrying the hidden message and identify the steganographer from many common users is a major challenge for steganographer detection. Steganographer detection can effectively detect illegal users and avoid information being stolen,which we believe will play a key role in many important multimedia security applications in the future.

Steganographer detection aims to locate criminals among a large number of innocent users who may be carrying secret information using the steganography technique. The difficulty of this task is in the collection of useful evidence, that is, to detect secret messages generated by an unknown steganography method and the payloads embedded in suspicious images and to identify criminals based on the image features. Existing steganalysis methods are algorithms that rely on the binary classification of known data sets and payloads. And the detection performance decreases significantly in the case of unknown payloads. In the inference phase, the learned model is used to extract discriminative features, thus capturing the differences between illegal and innocent users. A series of experimental results show that the method performs well in both the spatial and the frequency domains even with low embedded payloads. The method has good robustness and offers the possibility to solve the payload mismatch problem.

In this paper, deep learning-based steganographer detection is studied for color images, and the WiserNet(Wider Separate-Then-Reunion Network) steganographer detection method for color images is proposed. The main findings of the paper are as follow:

This is the first study on BMP color image steganographer detection. The experimental results show that the model can identify the steganographer accurately in terms of various steganography rates.

In this paper, we introduce DBSCAN (Density-Based Spatial Clustering of Applications with Noise) for steganographer detection after feature extraction, which can effectively identify users with different steganography image ratios. The simulation and experiment show that this method has fair good robustness.

The proposed method is validated by simulation and experiments taking on a standard dataset, showing that our method achieves a low detection error rate on the spatial domain.

The rest of the paper is organized as follows: In Sect. 2, we give a detailed overview of the current framework of grayscale graph steganographer detection methods. In Sect. 3, we describe the proposed WiserNet-based color images steganographer detection framework in detail. In Sect. 4, we perform a series of comprehensive experiments to verify the performance of our proposed method. In Sect. 5, we summarize our proposed work and outline future work.

2 Related Work

This section will give a brief overview of the latest results on grayscale graph steganographer detection. So far, steganographer detection on color images has not been developed. Therefore, we introduce the steganographer approach based on BMP color images.

2.1 Color Images Steganalysis

The main battlefield for information hiding in spatial domain images is on grayscale images. However, the confrontation between color image steganography algorithms and color image steganalysis algorithms has also received increasing attention from researchers since most images in real life come with color. The mainstream grayscale image steganography algorithms, including the well-known S-UNIWARD [11], HILL [12], and MiPOD [13], employ the so-called minimizing additive embedding distortion framework. Later, Denemark and Fridrich [14] and Li [15] went a step further from additive distortion algorithms and constructed effective non-additive distortion algorithms to embed images using the correlation between adjacent pixels. Among them, Li [16] proposed the CMD(Clustering Modification Directions) steganography algorithm to achieve excellent steganography performance.

To illustrate, grayscale image steganography algorithms (such as S-UNIWARD and HILL) can be directly applied to color images. This is usually done by treating each color channel as separate grayscale images and embedding secret information into each color channel separately. The general practice is to embed bits of secret information independently into each color channel by treating each color channel as a separate grayscale image.

Subsequently, inspired by the CMD steganography algorithm, Tang [17] proposed a non-additive steganography algorithm for color images, CMD-C. The CMD-C color images steganography algorithm preserves not only the correlation of pixels within each color channel, but also the correlation between the three color channels, and uses these correlations for embedding. Therefore, the performance of resisting the steganalysis algorithms for color images is even better and obtains good color image steganography performance.

The currently dominant steganalysis algorithms are steganalysis detectors built using rich models with multi-dimensional features [18] and integrated classifiers [19]. A separation, followed by aggregation network, was proposed in the paper about WiserNet.

The authors considered the weighted summation operation in the conventional convolutional structure, i.e., the process of forming a linear combination of the input color channels, as a "linear complicity attack" [20]. It retained the strongly correlated content while weakening the irrelevant noise in the input, which was more favorable for the determination of the steganalysis results.

2.2 Steganographer Detection

Currently, far too little attention has been paid to the task of steganographer detection, which can be divided into two main categories based on the way of detecting the steganographer: steganographer detection based on clustering and steganographer detection based on anomaly detection.

In 2011, Ker et al. [21] first transformed steganographer detection into a clustering problem study. They first extract 274-dimensional PEV features for each image of the user, which was composed of 193-dimensional DCT coefficient features and 81-dimensional calibrated Markov features.

Subsequently, based on the extracted PEV features, Ker et al. used the MMD(maximum mean discrepancy) to calculate the distance between feature sets for each pair of users as a similarity metric between users. Finally, an aggregated hierarchical classification algorithm based on the similarity metric was used to distinguish steganographers from the many non-steganographer.

In 2012 and 2014, Ker et al. [22, 23] further improved their work by defining a steganographer as an anomaly among the communicating users and proposed to identify the steganographer using anomaly detection methods. Unlike the previous work, they used the method of local anomaly factor [24] to calculate the anomaly degree value of each user and rank them, and the user with the highest anomaly value would be identified as the steganographer.

In 2018 and 2019, after calculating the MMD distance of every two users, Zheng used the conjoint hierarchical clustering algorithm to detect the steganographer, and combined objects by establishing a hierarchical tree. Finally, all non-steganographer were grouped into one class, while the steganographers were separately classified into one class [20]. In 2019, Zheng enhanced feature extraction by multi-scale embedding and then selected the steganographer by Gaussian voting [27]. In 2020, Zheng detects steganographers based on LOF (local outlier factor) and selective strategy [28].

3 Method

The proposed framework for steganographer detection based on BMP color images is shown in Fig. 1. The framework mainly consists of three parts. In the first step, each color images of each user are extracted with a special cross using the trained WiserNet; second, based on the extracted feature vectors, the steganographic images and non-steganographic images are classified using the DBSCAN clustering method; in the third step, the steganographer is detected using the ranking determination method. Unlike the existing grayscale graph-based steganographic detection framework, the steganographer detection algorithm proposed here is further optimized based on WiserNet in the feature

extraction step to cut the number of feature vectors to 100, which reduces the time consumption of clustering computation. The proposed method will be introduced in detail in the subsequent sections.

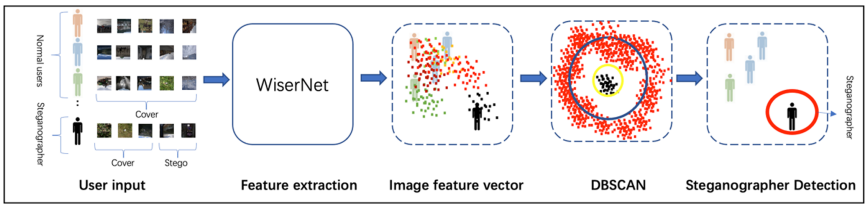

Fig. 1. Color image steganographer detection framework based on WiserNet.

3.1 Feature Extraction

The main object of this paper is true-color images, and we consider only the RGB true-color model, given a color image X of size M × N. It contains three color channels, namely, the red R channel, the green G channel, and the blue B channel. In this section, we do not consider the specific features of each color.

Thus, X can be expressed without loss of generality as $\{X_1, X_2, X_3\}$, where $X_i = (x_{i,pq})M \times N$, $x_{i,pq} \in \{0, 1, \cdots, 255\}$, $1 \leq i \leq 3$, $1 \leq p \leq M$, $1 \leq q \leq N$. In this paper, the feature extraction of color images is based on WiserNet. The first part of this network takes depthwise convolution, in which each input channel is convolved with a matrix of 30 convolution kernels to obtain the corresponding 30 independent output channels with 3 input channels and 30 convolution kernels, and 90 output feature maps will be obtained after passing the depthwise convolution structure. The second part performs the regular convolutional layer operation, which contains the BN (batch normalization) layer, ReLU (rectified linear unit) layer, and average pooling layer. The input feature image size of the regular convolution layer is reduced to 256 * 256, 128 * 128, and 32 * 32 in that order. The third part performs the pooling layer for processing, and the feature vector dimension is reduced by four times of fully connected network, that is, in descending order to 800, 400, 200, and 100. The final output is a 100-dimensional feature vector (Fig. 2).

3.2 DBSCAN Clustering Sorting to Detect Steganographer

After feature extraction to obtain a 100-dimensional feature vector, we take the DBSCAN algorithm based on density clustering to detect steganographer. Unlike the KMeans algorithm, This method does not require determining the number of clusters, but rather inferring the number of clusters based on the data, and can generate clusters for arbitrary shapes. DBSCAN is one of the classes that achieve the final clustering by the set of samples connected by the maximum density derived from the density reachability relation. The parameter (ϵ, MinPts) is used to describe the closeness of the neighborhood sample distribution. Where ϵ describes the neighborhood distance threshold of a sample

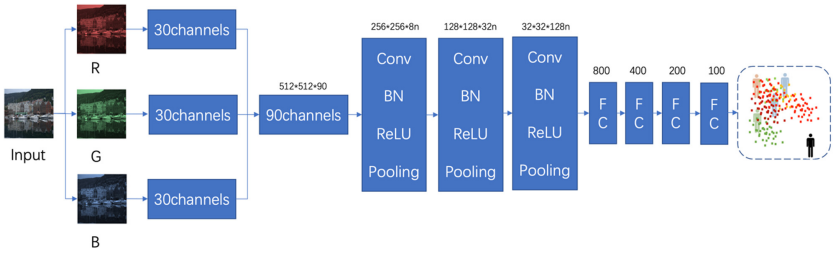

Fig. 2. Feature extraction framework based on WiserNet.

and MinPts describes the threshold of the number of samples in the neighborhood of a sample with distance. There can be one or more core objects inside the clusters of DBSCAN. If there is only one core object, all other non-core object samples in the cluster are in the ϵ-neighborhood of this core object; if there are multiple core objects, there must be one other core object in the ϵ-neighborhood of any one core object in the cluster, otherwise, these two core objects cannot be density reachable. The set of all samples in the ϵ-neighborhood of these core objects forms a DBSCAN clustering cluster.

4 Experiments

4.1 Data Set and Experimental Setup

All experiments in this paper were conducted on the BossBase (v1.01) dataset. Following the data set generation process used in the article [24], we obtained the BossBMP data set from 10,000 full-resolution original images of BossBase through subsampling operation.

BossBase dataset was also used in all experiments in this paper. In this chapter, the dataset generation process in the article [29] is used as a reference, and the original DNR format image is decontaminated using UFRaw software. The BMP format image is obtained by subsampling using ImageMagick, and then it is centered and cropped to 512 $\times$ 512 to obtain the BossBMP dataset. In the following illustration, the advanced CMD-C-Hill color image steganography algorithm is selected as the detection object, and the load is set to 0.1, 0.2, 0.3, 0.4, and 0.5 BPC (bits per color channel) for steganography operation, so as to obtain the carryover image corresponding to the carrier image.

4.2 Evaluation of the Effectiveness of Feature Extraction Methods

The proposed color image steganographer detection in this paper consists of two main parts, of which feature extraction based on WiserNet is the most critical technique. Therefore the effectiveness of the proposed feature extraction method for the steganalysis task is first tested before the performance evaluation of the steganographer detection task. Since the compared methods are based on the binary classification training models of steganograms and non-steganograms, in hopes of a fair comparison, the experiment processes the 100-dimensional features for binary classification results and compares them with the classical Ye's model, Xu's model, and SRNet networks (Tables 1 and 2).

Table 1. Performance evaluation parameters of feature extraction network.

Accuracy	Precision	Recall	F1
84.90%	82.04%	84.65%	82.80%

Table 2. Comparison of the performance and feature latitude of different color images steganography analysis methods.

Method	Classification precision (%)	Dimensional feature
Ye-Net [30]	83.79	144
Xu-Resnet [26]	84.97	128
SRNet [31]	87.06	512
WiserNet [34]	86.42	200
WiserNet-100	84.90	100

What stands out in the above tables is the marked improvement in the precision of our WiserNet-100 feature extraction network based on WiserNet. There is a 1.11% performance improvement compared with Ye-Net and equal to Xu-Resnet. It is also worth mentioning that there is a slight decrease in precision compared with the SRNet network, but SRNet outputs 512 feature latitude and WiserNet-100 has only 100 feature latitude. In specific scenarios, WiserNet-100 may reduce the time consumption of the whole experiment owing to the reduced latitude.

4.3 Effectiveness of Steganographer Detection

In practice, the steganographer may use different embedding rates to embed different hidden information in different images. Therefore, in our experiments, Ye-Net, Xu-Resnet, and SRNet are used as control groups for comparison regarding embedding rate.

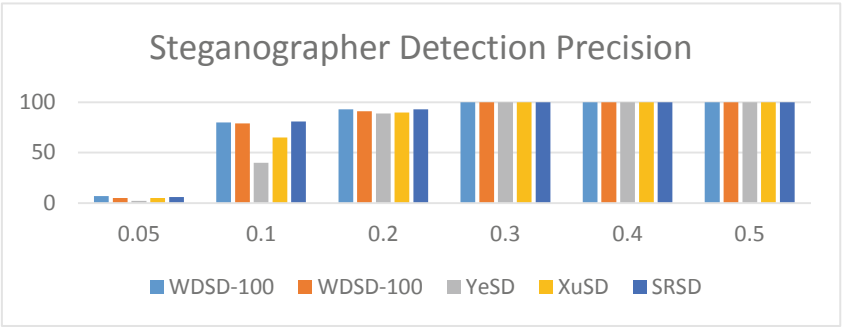

Fig. 3. Performance comparison of different steganographer detection methods with different embedding rates in the same steganography

The tested embedding probabilities contain 0.05, 0.1, 0.2, 0.3, 0.4, and 0.5. The above steganography algorithms use CMD-C-HILL to control the variables uniformly.

As can be seen from Fig. 3, the improved WiserNet-100 in this chapter has similar accuracy to the original WiserNet-200 in terms of Steganography analysis and detection. At 0.1 BPC, the detection accuracy of the improved WiserNet-100 is the same as that of WiserNet-200, and slightly higher than that of SRSD. At 0.2 BPC, the proposed detection algorithm maintains high accuracy, which is higher than traditional YESD and XUSD. After 0.3BPC, the detection rate of all the five detection algorithms reaches 100%.

Table 3. Comparison of the time consumption of different color image steganographer analysis methods

Method	Dimensional feature	Time(S)
Ye-Net	144	9.62
Xu-Resnet	128	9.10
SRNet	512	16.41
WiserNet	200	10.74
WiserNet-100	100	7.79

As can be seen from Table 3, when the performance of WiserNet-100 is the same as that of WiserNet-200, WiserNet-100 has a 2.95 s improvement in time consumption. The maximum detection time for SRSD with a higher feature dimension is 16.41Ss, which decreases as the feature dimension decreases. However, the detection time of WNCISD-100 is only 7.79S. In comparison, the time consumption is reduced to 50%. The reason for the similar accuracy of steganographer detection despite the difference in classification accuracy is that the experiments assume that there is only one steganographer, and the final identification of steganographers is done by tag sorting. Therefore, even though there is some difference in classification accuracy, a relatively high detection accuracy can still be achieved. Overall, the steganographer detection algorithm proposed in this chapter has an ideal university in specific scenarios.

4.4 Validity Assessment with Different User Graph Percentages

So far, all the above experiments are based on 100 users with 10 graphs each, of which 99 users with all cover images and the images of the steganographer are all steganographs, However, in practical life, it is likely that there is a user who sends a hidden message in the form of multiple messages and only a few steganographs at a time. Therefore, this experiment focuses on the percentage of steganographer who do not use steganography. The following results are obtained by repeating the experiment 100 times.

As can be seen from Fig. 4, the detection accuracy reaches 100% when the BPC is 0.5 and the figure share is greater than or equal to 40%; the detection accuracy reaches 89% when the BPC is 0.4 and the share is greater than or equal to 40% and reaches 100% detection when the figure share is 80%; the detection accuracy increases with the

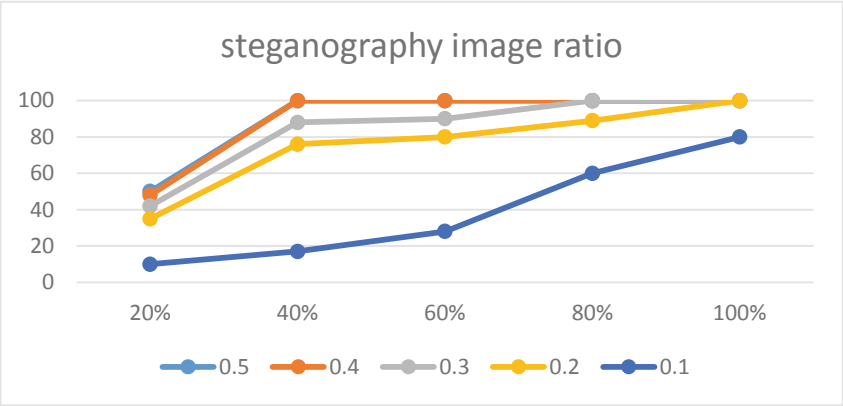

Fig. 4. WDSD performance comparison at different graph occupancy ratios and steganography rates.

increasing figure share and BPC. The performance is average when the BPC is 0.1, which is related to the low steganography rate resulting in the inconspicuous feature extraction. When the graph occupancy ratio is greater than or equal to 40%. The framework performs well in terms of detection performance.

5 Conclusion

This paper proposes a novel BMP color image steganographer detection method based on WiserNet. The first step is to extract the feature vectors of images from the WiserNet deep learning model through all three channels of BMP color images. The second step is to perform DBSCAN density clustering based on the extracted image feature vectors. The third step is to sort the clustering results and identify the steganographer.

Under the condition of the same detection accuracy, at present, relevant literature and case studies on color images are scarce. We use SRNET network with better feature extraction effect to carry out steganographer detection, which consumes 16.41 s. The algorithm proposed in this paper consumes 7.79 s, which reduces the time by half. In addition. This paper also verifies the performance of this method in complex realistic environments such as different embedding rates and different proportions of graphs, and the experiments show that this method achieves impressive performance in all the above cases.

In the future, the proposed color image steganographer detection will be applied to other image types, such as JPEG, so as to expand its application range.

References

1. Yang, J., Li, S.: Steganalysis of joint codeword quantization index modulation steganography based on codeword bayesian network. Neurocomputing **313**, 316–323 (2018)
2. Anderson, R.J., Petitcolas, F.A.P.: On the limits of steganography. IEEE J. Sel. Areas Commun. **16**(4), 474–481 (1998)

3. Islam, S., Gupta, P.: Effectof morphing on embedding capacity and embedding effiffifficiency. Neurocomputing **137**, 136–141 (2014)
4. Guo, SL., Yi, T.C., Wen, N.L.: A framework of enhancing images steganography with picture quality optimization and anti-steganalysis based on simulated annealing algorithm. IEEE Trans. Multimedia **12**(5), 345–357 (2010)
5. Zhang, X., Peng, F., Long, M.: Robust coverless images steganography based on DCT and LDA topic classification. IEEE Trans. Multimedia **20**(12), 3223–3238 (2018)
6. Zhou, H., Chen, K., Zhang, W., et al.: Comments on "steganography using reversible texture synthesis." IEEE Trans. Images Process. **26**(4), 1623 (2017)
7. Zheng, M., Zhong, S.-h., Wu, S., et al.: Steganographer detection via deep residual network. In: Proceedings of the IEEE International Conference on Multimedia and Expo, pp. 235–240 (2017).
8. Ker, A.D.: Batch steganography and the threshold game. In: Security, Steganography, and Watermarking of Multimedia Contents IX. SPIE, pp. 401–413 (2007)
9. Li, F., Kui, W., Lei, J., et al.: Steganalysis over large-scale social networks with high-order joint features and clustering ensembles. IEEE Trans. Inf. Forens. Secur. **11**(2), 344–357 (2017)
10. Kodovsky, J., Fridrich, J.: Ensemble classifiers for steganalysis of digital media. IEEE Trans. Inf. Forensics Secur. **7**(2), 432–444 (2012)
11. Goljan, M., Fridrich, J., Cogranne, R.: Rich model for steganalysis of color images. In: Proceedings IEEE International Workshop Information Forensics and Security (WIFS), pp. 185–190, December 2014
12. Goljan, M., Fridrich, J.: CFA-aware features for steganalysis of color images, Proc. SPIE **9409**, 94X090V-1–94090V-13 (2015)
13. Abdulrahman, H., Chaumont, M., Montesinos, P., et al.: Color images steganalysis using correlations between RGB channels. In: Proceedings IEEE 10th International Conference Availability, Reliability and Security (ARES), pp. 448–454, August 2015
14. Abdulrahman, H., Chaumont, M., Montesinos, P., et al.: Color images steganalysis using RGB channel geometric transformation mea- sures. Secur. Commun. Netw. **9**(15), 2945–2956 (2016)
15. Abdulrahman, H., Chaumont, M., Montesinos, P., et al.: Color images steganalysis based on steerable Gaussian filters bank. In: Proceedings 4th ACM Information Hiding Multimedia Security Workshop (IH&MMSec), pp. 109–114 (2016)
16. Denemark, T., Sedighi, V., Holub, V., et al.: Selection-channel-aware rich model for steganalysis of digital images. In: Proceedings 6th IEEE International Workshop Information Forensic Security (WIFS), pp. 48–53, December 2014
17. Tang, W., Li, H., Luo, W., et al.: Adaptive steganalysis based on embedding probabilities of pixels. IEEE Trans. Inf. Forensics Secur. **11**(4), 734–745 (2016)
18. Boroumand, M., Fridrich, J.: Applications of explicit non-linear feature maps in steganalysis. IEEE Trans. Inf. Forensics Secur. **13**(4), 823–833 (2018)
19. Schmidhuber, J.: Deep learning in neural networks: an overview. Neural Netw. **61**, 85–117 (2015)
20. Tan, S., Li, B.: Stacked convolutional auto-encoders for steganalysis of digital images. In: 2014 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), pp. 1–4 (2014)
21. Qian, Y., Dong, J., Wang, W., et al.: Deep learning for steganalysis via convolutional neural networks. Proc. SPIE **9409** (2015). Art. no. 94090J
22. Pibre, L., Pasquet, J., Ienco, D., et al.: Deep learning is a good steganalysis tool when embedding key is reused for different images, even if there is a cover sourcemismatch. In: Proceedings Electron (2016)

23. Qian, Y., Dong, J., Wang, W., et al.: Learning and transferring representations for images steganalysis using convolutional neural network. In: Proceedings IEEE International Conference Images Processing (ICIP), pp. 2752–2756, September 2016
24. Zheng, M., Zhong, S.h.-h., Wu, S., et al.: Steganographer detection based on multiclass dilated residual networks. In: ACM International Conference on Multimedia Retrieval (ICMR) (2018)
25. Holub, V.: Content adaptive steganography-design and detection. PhD thesis. Department of Electricaland Computer Engineering, Binghamton University (2014)
26. Song, X., Liu, F., Yang, C., et al.: Steganalysis of adaptive JPEG steganography using 2D gabor filters. In: Proceedings of the 3rd ACM Workshop on Information Hiding and Multimedia Security -IH&MMSec 20‘15. ACM Press (2015)
27. Xu, G., Wu, H.Z., Shi, Y.Q.: Structural design of convolutional neural networks for steganalysis. IEEE Signal Process. Lett. **23**(5), 708–712 (2016)
28. Ye, J., Ni, J., Yi, Y.: Deep learning hierarchical representations for image steganalysis. IEEE Trans. Inf. Forensics Secur. **12**(11), 2545–2557 (2017)
29. Suykens, J.A., Vandewalle, J.: Least squares support vector machine classifiers. Neural Process. Lett. **9**(3), 293–300 (1999)
30. Boroumand, M., Chen, M., Fridrich, J.: Deep residual network for steganalysis of digital images. IEEE Trans. Inf. Forensics Secur. **14**(5), 1181–1193 (2019)
31. Tang, W., Li, B., Luo, W., et al.: Clustering steganographic modification directions for color components. IEEE Sig. Process. Lett. **23** (2), 197–201 (2016)