Feature Extraction Based Image Steganalysis Using Deep Learning

Shamal Salunkhe
Ramrao Adik Institute of Technology
16 PUBLICATIONS **28** CITATIONS

Surendra Bhosale
Veermata Jijabai Technological Institute
129 PUBLICATIONS **184** CITATIONS

WEENTECH
Proceedings in
Energy
ISSN:2059-2353

Available online at www.weentechpublishers.com

CrossMark

WEENTECH Proceedings in Energy 6 (2) (2019) 1-1

International Conference on Recent Advances in Mechanical Engineering, Department of Mechanical Engineering,
Delhi Technological University, New Delhi, India

Feature Extraction Based Image Steganalysis Using Deep Learning

Shamal Salunkhe [a], Surendra Bhosale [b]

[a] *Research Scholar in Electrical Engineering Department, V. J. T. I., Mumbai, India,*
[a]*Assistant Professor in Instrumentation Engineering Department, R. A. I. T., Navi Mumbai, India,*
[b]*Associate Professor in Electrical Engineering Department, V. J. T. I., Mumbai, India*

[*]*Corresponding author's mail:* sssalunkhe_p18@ee.vjti.ac.in

Abstract

The innovations in advanced information technologies has led to rapid delivery and sharing of multimedia data like images and videos. The digital steganography offers ability to secure communication and imperative for internet. The image steganography is essential to preserve confidential information of security applications. The secret image is embedded within pixels. The embedding of secret message is done by applied with S-UNIWARD and WOW steganography.

Hidden messages are reveled using steganalysis. The exploration of research interests focused on conventional fields and recent technological fields of steganalysis. This paper devises Convolutional neural network models for steganalysis.

Convolutional neural network (CNN) is one of the most frequently experimented deep learning techniques. The Convolutional neural network is used to extract spatio-temporal information or features and classification. We have compared steganalysis outcome with AlexNet and SRNeT with same dataset. The stegnalytic error rates are compared with different payloads.

*Keywords: D*igital Steganography, Deep Learning, Convolutional Neural Network (CNN), Feature Extraction.

https://doi.org/10.32438/WPE.XXXX
Manuscript History
Receipt of completed manuscript: 28 February 2019
Receipt of Revised Manuscript: 11 August 2019
Date of Acceptance: 20 September 2019
Online available from: 22 October 2019

Abbreviations

SRM	Spatial Rich Model
SVM	Support vector machines
FLD	Fisher Linear Discriminant
WOW	Wavelet Obtained Weights
SA	Stegnographic Algorithm
ANN	Artificial Neural Networks
FFT	Fast Fourier Transform
LSB	Least Significant Bit
SGD	Stochastic gradient descent
GAN	Generative Adversarial Networks
BPP	Bit Per Pixel
BOSS	Break Our Stegnographic System

1. Introduction

Steganography is the practice to hide secret messages into multimedia signals such as text, audio, image and video. The technique of revealing the presence of secret messages embedded in the digital media is referred as Steganalysis. Steganalytic techniques are spread over a wide area of applications such as patient's medical imagery records, document authentication, intelligence agencies, military, and organizations for secret data communication, smart identity cards and remote sensing.

There are many different ways to embed secret messages into ordinary data files. The most common embedding technique is the least significant bit (LSB) steganography which is performed in spatial domain. Another embedding technique is transform domain technique. The techniques such as Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT) and Fast Fourier Transform (FFT) are implemented to embed secret data in lower, middle or higher coefficients of the clear images [1]. In the distortion based steganography, distortion threshold is the key parameter which helps to select embedding regions. The concept of spread spectrum is used in embedding technique by which the secret information is hide within a vast frequency bandwidth. In the Statistical technique, various properties of cover image are changed to hide a secrete message. Masking and filtering approach hides the data by marking an image. It is mostly used in watermarking. Image steganography algorithms such as S-UNIWARD, J-UNIWARD, WOW, and HUGO are developed to maintain the security of secret information communication [2].

The presence of the hidden message is detected and disclosed using steganalysis. The correlation of image neighborhoods, distortion at embedding are the factors considered in steganalysis. The Spatial Rich Model (SRM) and Projection based Model (PSRM) are mainly used for the high-order and dimensional features extraction. Selection of right features for stego image classification is studied by many researchers. DCT based steganography can be detected using particular DCTR models. Hash algorithms are also implemented for steganalysis. Researchers focused on deep learning techniques for both steganography and steganalysis. Artificial Neural Network (ANN) is devised for Steganography whereas Convolutional Neural Network (CNN) is popular for steganalysis process [3]. CNN, is a architecture is inspired by human visual cortex. CNN has shown its efficiency in images processing works such as segmentation, classification and detection. The steganalysis algorithm could be two stages or one stage. When feature extraction and classification are performed by two different modules, it is a two stages steganalysis. Within a single module, when feature extraction and classification are performed, it is a one stage steganalysis. CNN is a best option for one stage steganalysis.

.

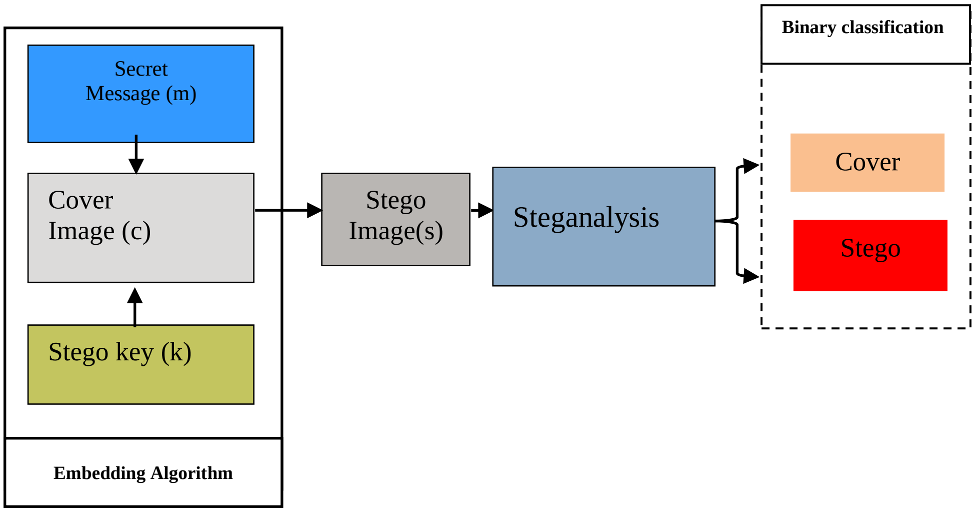


Fig.1 One stage steganalysis

Generalized structure of one stage steganalysis is shown in Fig.1. As shown in diagram, secret message is hiding inside the cover image using embedding algorithm. This stego image is mixed with clear images and transmitted. Steganalysis is aimed to detection stego images. Stego and cover images are classified using steganalysis algorithm.

1.1 Types of Steganalysis:

The steganalysis is divided into two parts, Specific (Targeted) and Generic (Blind). Based on dependency on the Steganographic Algorithm (SA), steganalysis is classified as targeted and blind steganalysis.

Based on detection methods it is further divided on the base of detection tools, deep learning techniques, attacks and embedding methods. The Steganalyst is trying to find out hidden message or suspicious region by attacking. Types of attack based steganalysis is given bellow,

- When the cover image and stego image are both known it represents a Known cover attack.
- In some cases the hidden message may cracked by the attacker, it is a Known message attack.
- Using particular attacking patterns, stego region can be found out. It is difficult sometimes even if message is known.
- When Steganalyst recreates a stego image from some Steganography tool or algorithm from a known data it is known as Chosen message.
- When both the original and stego images are openly available and the steganography algorithm is known, it is called as known stego attack.

1.2 Challenges in Digital Steganalysis:

- Nonlinear residuals generated during embedding are not much utilized in detection.
- Diminishing of features is a one of the limitation in steganalysis.
- Variable embedding rate in JPEG steganalysis can be explored
- Effects of variable quality factors in JPEG images steganalysis could be a research area.
- Because of lossy compression, JPEG steganography/steganalysis is less focused.
- Limited JPEG datasets are available and explored [4].

1.3 Proposed Work:

We have experimented with S-UNIWARD and WOW steganographic techniques on BOSSbase 1.01 dataset and performed steganalysis using CNN (SRNeT and AlexNet). The paper is organized as follows, in the Section 2; related work to CNN (ConvoNet) is discussed. We have described the embedding techniques, and basics of SRNeT and AlexNet, separately, in the Section 3. We have compared the evaluation parameters and analyzed them in the Section 4. The paper is finally end up in the Section 5, with the conclusion.

2 Related Work:

Steganalysis is equivalent to steganography. Targeted steganalysis is effective when embedding algorithm is known. Steganalyst can easily find out traces generated by embedding algorithm. But adaptive steganography algorithms or embedding algorithms are difficult to trace out, so targeted steganalysis become failing. When embedding algorithms are unknown, the blind steganalysis approach is useful though it is less accurate. In the clairvoyant steganalysis, almost all parameters are known to steganalyst so simply by attacking on cover he can easily trace stego image. SRM, MaxSRM, tSRM, σmaxSRM based on statistical parameters and machine learning techniques [5]. Deep learning techniques are implemented in blind steganalysis to classify stego and cover image. CNN structure was first proposed by Fukushima in 1988. Different types of CNN architectures are implemented for multi classification and binary classification. The image steganalysis is a binary classification problem. The deep neural network has different variants other than CNN are Generative Adversarial Networks (GAN) used for images steganography and steganalysis. Other variations of deep learning such as Long Short Term Memory Network (LSTM) and Recurrent Neural Network (RNN) and used when sequential data processing is required [6].

Research based on CNN has catches tremendous rise in recent years, particularly on various image based applications. Various CNN architectures has been developed so far. For image feature detection and classification following CNN architectures are used, AlexNet [7], VggNet, GoogLeNet, ResNet, Ye-Net, Xu-Net, YedroudjNet [8], ZhuNeT, and SRNeT [9] for image classification. In this paper, to study architecture and functioning of CNN models, AlexNet and SRNeT has been chosen.

3 Embedding Techniques and CNN Architectures:

The cover modification approach of image steganography is most popular. Content-adaptive embedding gives more strength to secure steganography. These are base on minimization of embedding distortion function. Following are example of these methods, HUGO , HILL, S/J-UNIWARD, WOW, and

MiPOD. Embedding in spatial domain, directly carried out on pixels. Though it is not secure as compare to frequency domain embedding, has capacity to hide more number of bits.

3.1 S-UNIWARD Embedding Techniques:

Spatial-UNIversal WAvelet Relative Distortion (S-UNIWARD) technique is firstly proposed by Holub et al. in 2014 [10]. The wavelet domain is chose to define distortion function of S-UNIWARD technique.
The Cost map ρ is composed as, $\{\rho_i \in [0,\infty]\}_{i-1}^{n}$. Here ρ_i is given in Eq. (1),

$$\rho_i = \sum_{k=n}^{d} \sum_{u=1}^{n_1} \sum_{v=1}^{n_2} \frac{\left|W_{uv}^{(k)}(x)-W_{uv}^{(k)}(x\sim x_1)\right|}{\sigma+\left|W_{uv}^{(k)}(x)\right|} \qquad (1)$$

3.2 WOW Embedding Techniques:

The WOW (Wavelet Obtained Weights) algorithm is proposed by Holub and Fridrich, in 2012 [11], [13]. WOW technique has resemblance with S-UNIWARD. In WOW technique, three directional wavelet filters are used to compute the embedding costs for each pixel from three directional residuals. It firstly calculates the weighted difference between the residual wavelet coefficients of the clear image, and stego image. Then it aggregates the results and used to construct a cost map. The ρ_i is cost map and ξ_i is embedding capacity are represented by Eq. (2) and Eq. (3),

$$\rho_i = \sum_{k=1}^{d} \frac{1}{\xi_i^{(k)}} \qquad (2)$$

$$\xi_i^{(k)} = \sum_{u=1}^{n_1} \sum_{v=1}^{n_2} \left|W_{uv}^{(k)}(x)\right| * \left|W_{uv}^{(k)}(x) - W_{uv}^{(k)}(x\sim x_i)\right| \qquad (3)$$

Embedding changes are performed by WOW, within in the noisy regions.

3.3 CNN Architecture:

In image steganalysis problem, CNN model used to classify stego and original image. For an input image, i = (i_1,…., i_n), i is either cover image (P_C) or (P_S). Pc is the distribution of the clear images, whereas Ps is distribution of stego images. Three major steps are involved namely, pre-processing, feature extraction and lastly classification. Fig. 2 is showing steganalysis using one stage CNN.
Every layer of CNN carried out feature selection and outcome of every layer of CNN is a feature map. Filtering and pooling are cardinal parts of CNN, which are working on feature maps.

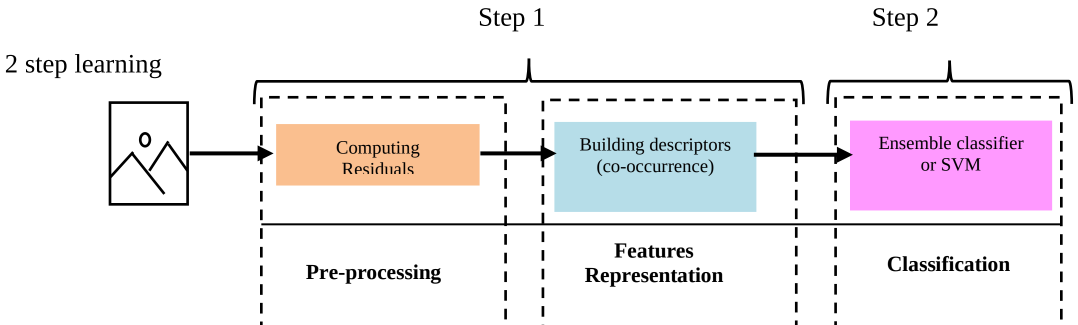


Fig.2 Steganalysis using one stage CNN [8].

- Each input image is passes through a series of convolution stages (pooling, batch normalization) by applying filters (kernels). In filtering operation actually convolutes a filter matrix or weights with the input values. Convolution Layers extract feature information from input images. Convolution operation is followed by pooling operation. Pooling operation, such as max pooling, average pooling used to reduce dimensions of images. The classification is performed with flatten and fully connected layers. Softmax activation function is applied to classify the image with probability between 0 and 1.

3.3.1 AlexNet:

AlexNet is a one of the deep convolutional network. It is proposed by Alex Krizhevsky in 2012. It can handle 60 million parameters. It has total eight layers, five layers of convolutional layers, max-pooling layers, dropout layers, and last three layers are fully-connected. AlexNet just added few more layers onto LeNet-5. They were the first to implement Rectified Linear Units (ReLUs) as activation function. AlexNet is based on the deep neural network such as CNN. It extracts the features from raw image pixels.
AlexNet is divided into two parts; each operation is performed with a separate graphics processing unit to get optimum response. A convolution layer is a first layer with 96 filters. Size of each filter is 11 × 11 × 3. The output of first layer is given to second layer. The second layer is max pooling layer used to reduce the computation complexity. It is followed by convolution layer with 256 filters with size 5 × 5 × 48. Fully connected layer is used in the upcoming three layers but with different size. Sixth layer is a fully connected layer which maps the input matrix to a vector with, 1 × 2048 size. Then, layers seven and eight are used as fully connected layers. A feature vector is the output of the AlexNet, with 1 × 1000 size.

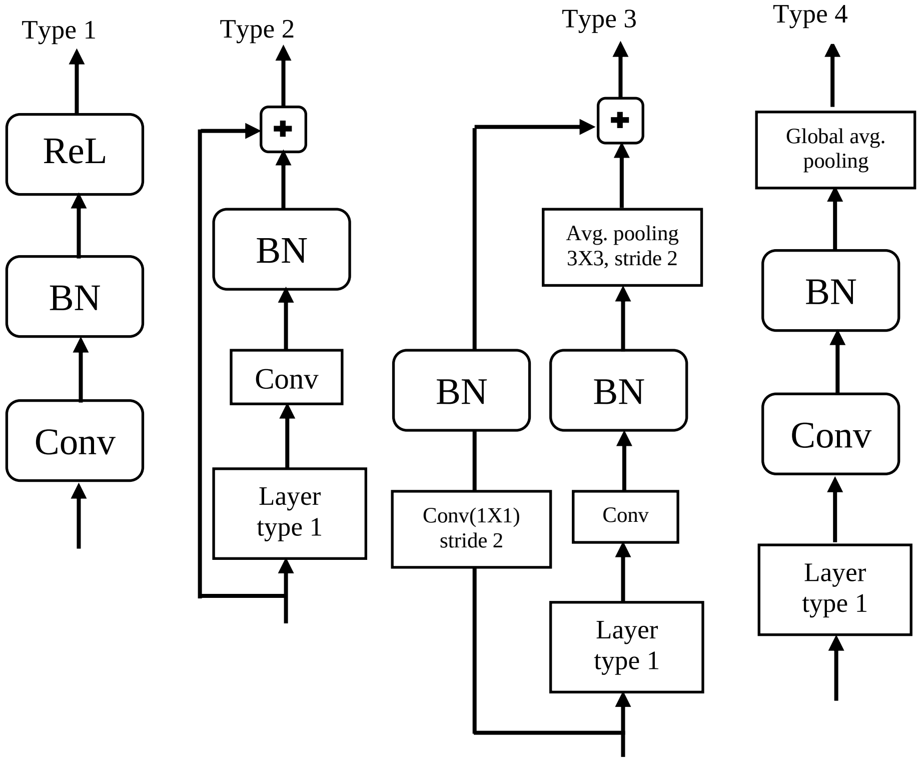


Fig. 3 AlexNet architecture

The AlexNet extract features from high dimensional feature set containing 1000 and 686 features. It may leads to computational cost and can slow down the classifiers performance. AlexNet has better capacity to handle more GPUs as compared to SRMs. Dropout concept is used in AlexNet.

3.3.2 SRNet:

Steganalysis Residual Network (SRNet) is invented by Boroumand et al. in 2019. SRNeT can be utilized for both JPGE and spatial steganalysis. It consists of the convolution part and the classification part. The pre-processing part is not used in SRNet. It can handle only large datasets. In Fig. 4 architecture of SRNeT is shown.

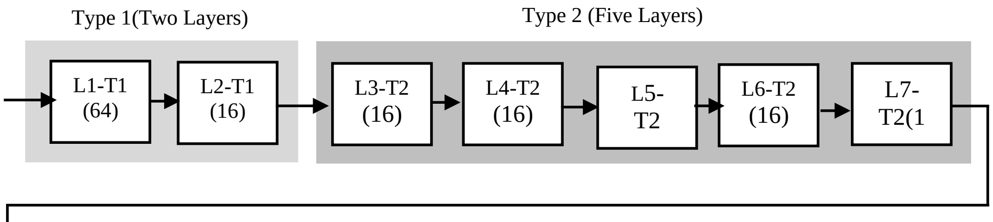


Fig. 4 SRNet architecture

SRNet consists of total twelve Convolutional layers. ReLU is the only one activation function used in all layers. SRNeT model has a classification module with one fully connected layer. A softmax layer is came after fully connected layer. Finally, it is separating two classes. In next version of SRNeT the side channel aware concept is incorporated which performs even better than SRNeT.

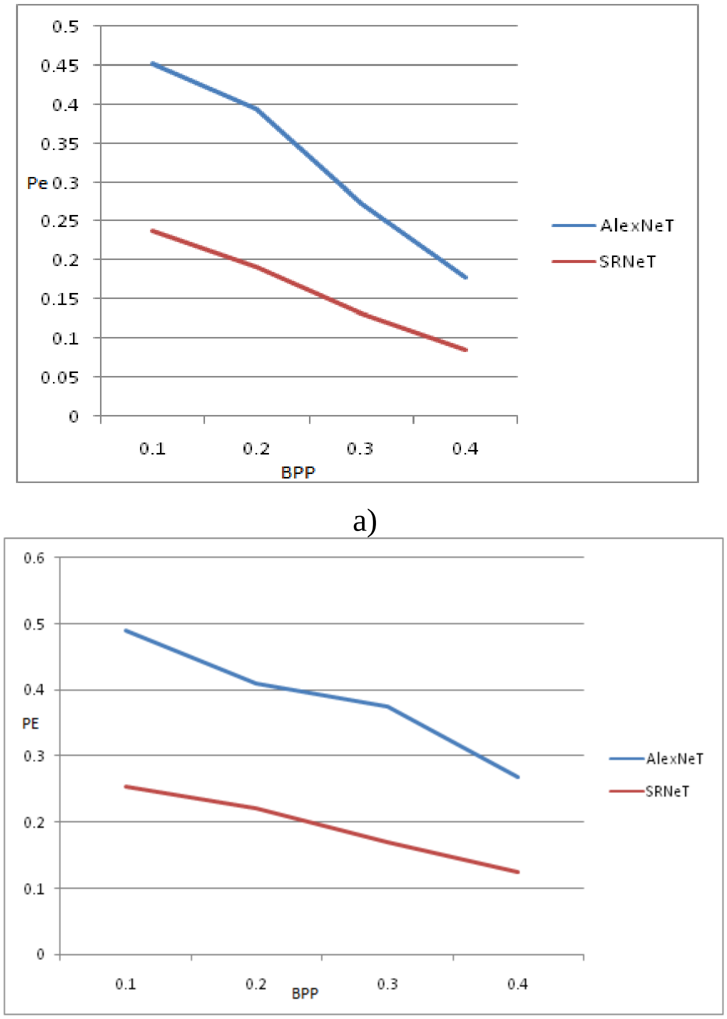
4 Results and discussion:

We have used BOSSBase v1.01 database which is the most widely used dataset in steganography and steganalysis. The dataset is consists of total 10,000 images. These are grey scale images of size 512 × 512. For training 6,000 images are used for training, 3000 used for testing and 1000 used for validation. All images are down sampled as per ConvoNet. There are many optimizers are available such as Adam, SGD, Gradient Descent, Adagrad, Adadelta and Adamax. We have used Adam optimizer. Maximum epochs are fixed to 200. AlexNet and SRNeT are tested on BOSSbase v1.01, with two WOW and S-UNIWARD steganographic algorithms with various embedding rates. The detection error rates are further compared for CNN models using two different steganographic algorithms with different payloads are shown in Table 1.

Table 1. The detection error rates comparison with CNN models using two steganographic methods at different payload.

Embedding Algorithms	Payload (BPP)	AlexNet	SRNeT
WOW	0.1	0.452	0.239
	0.2	0.395	0.192

	0.3	0.272	0.133
	0.4	0.178	0.086
S-UNIWARD	0.1	0.490	0.254
	0.2	0.410	0.221
	0.3	0.375	0.170
	0.4	0.269	0.125



a)

b)

Fig.5 The Detection error rates comparison using CNN models with a) WOW and b) S-UNIWARD steganographic algorithms at different payload.

We have tested steganalysis error with WOW and S-UNIWARD embedding methods at different payload (BPP). Both CNN algorithms are showing superior results with WOW than S-UNIWARD embedding algorithm. Algorithms are working well at payloads. The results are even better when. Finally, SRNet is better performing as compared to AlexNet. It may show even a greater response with additional data set.

5 Conclusions:

We have studied embedding techniques WOW and S-UNIWARD. Overviewed performance of AlexNet and SRNet with different payload, embedding techniques. CNN models could perform better with adaptive optimization methods.

References

[1] Jan Butora and Jessica Fridrich, (2019) Effect of JPEG Quality on Steganographic Security, IHMMSec '19, July 3–5, 2019, TROYES, France Association for Computing Machinery. ACM ISBN 978-1-4503-6821-6/19/06

[2] Ruohan Meng, Qi Cui and Chengsheng Yuan (2018) A Survey of Image Information Hiding AlgorithmsBased on Deep Learning, CMES, vol.117, no.3, pp.425-454.

[3] Dina Bashkirova (2016) Convolutional Neural Networks for Image Steganalysis, BioNanoSci. Springer, DOI 10.1007/s12668-016-0215-z.

[4] M. Chaumont (2020) Deep Learning in steganography and steganalysis from 2015 to 2018. Digital Media Steganography: Principles, Algorithms, Advances, volume abs/1904.01444, page 39. Elsevier.

[5] J. Fridrich and J. Kodovský (2012)Rich Models for Steganalysis of Digital Images. IEEE Transactions on Information Forensics and Security, TIFS, 7(3):868–882.

[6] D. Hu, L. Wang, W. Jiang, S. Zheng, and B. Li. (2018) A novel image steganography method via deep convolutional generative adversarial networks. IEEE Access, 6:38303–38314, 2018.

[7] Alex Krizhevsky, Ilya Sutskever and Geoffrey E. Hinton (2012) ImageNet Classification with Deep Convolutional Neural Networks, NIPS, pp. 1106–1114.

[8] M. Yedroudj, F. Comby, and M. Chaumont (2018) Yedroudj-Net: An Efficient CNN for Spatial Steganalysis. In Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'2018, Calgary, AB, Canada.

[9] M. Boroumand, M. Chen, and J. J. Fridrich (2019) Deep residual network for steganalysis of digital images. IEEE Trans. Information Forensics and Security, 14(5):1181–1193.

[10] V. Holub, J. Fridrich, and T. Denemark (2014) Universal Distortion Function for Steganography in an Arbitrary Domain. EURASIP Journal on Information Security, JIS, 2014(1):1.

[11] V. Holub and J. Fridrich (2012) Designing Steganographic Distortion Using Directional Filters. In Proceedings of the IEEE International Workshop on Information, Forensics and Security, WIFS'2012, pages 234–239, Tenerife, Spain.

[12] V. Holub, J. J. Fridrich, and T. Denemark (2013) Random projections of residuals as an alternative to co-occurrences in steganalysis. In Media Watermarking, Security, and Forensics 2013, Burlingame, CA, USA.

[13] http://dde.binghamton.edu/download/