IJMLNCE JOURNAL

International Journal of Machine Learning and Networked Collaborative Engineering

Journal Homepage: http://www.mlnce.net/home/index.html

DOI : https://doi.org/10.30991/IJMLNCE.2018v02i02.001

Steganalysis for Reversible Data Hiding Based on Neural Networks and Convolutional Neural Networks

[a]Ho Thi Huong Thom[*], [b]Nguyen Kim Anh, [c]Bui Dinh Vu

[a,b,c]Faculty of Information Technology, Vietnam Maritime University, Vietnam

[a]thomhth@vimaru.edu.vn, https://orcid.org/0000-0002-1825-9577,

[b]Kimanh@vimaru.edu.vn, https://orcid.org/0000-0001-7725-7945

[c]buidv@vimaru.edu.vn,https://orcid.org/0000-0002-2914-9613

Abstract

Lossless data hiding techniques is a technique that is very interesting. In which there is a large amount of reversible information hidden technologies. This technique makes it possible to restore the original image after extracting the information from the stego image. The stego image (hidden image with secret data) is hardly detected by any variable. There are many studies for this field are published. Secret information is hidden on the pixel space, frequency (cosine, wavelet) coefficient space or difference image coefficient space. However, by analyzing meticulously between the cover image and the stego image on these space, one can detect abnormal signs. In a previous work, a steganalytic techniques produced that was based on analysis of the transform coefficient histogram with the correct detection ratio between 88% and 92%. In this article, proposing another method to improve the detection ratio of that steganalysis based on Neural Networks (NNs) and Convolutional Neural Networks (CNNs). The test results show 96% correct detection rates for NNs and 94% for CNNs, this is a better result than our previous method. This proposed approach can be applied to detect stego images on spatial and other frequency domain.

Keywords

Steganography,

Steganalysis,

Cover Image,

Stego Image,

Histogram Shifting,

Lossless Data Hiding,

Neural Networks and

Convolutional Neural Networks

1 Introduction

Steganography is the science of hiding the existence of information. Its purpose is to convey the message secretly so except the sender and receiver, no one knows about the existence of the message. Steganography actually is a form of security by concealment. digital image, video, sound or any other files can be used as "cover" to carry secret messages, after that it is called stego.

[*] Corresponding Author

Ho Thi Huong Thom

Email: thomhth@vimaru.edu.vn

Steganalysis is opposite of Steganography. It is the science of detecting the message hidden using steganography. The main purpose task of it is to distinguish between cover image and stego image.

In recent years, many lossless data hiding techniques have been proposed for stego images. Lossless data can be embedded in the spatial domain [1, 2, 3] or in the transform domain [4, 5]. Xuan et al [5] proposed a method by a histogram shifting in integer wavelet transform domain (IWH method). This method hides messages into high frequent sub-bands of integer wavelet coefficients.

In [7], we offered a new steganalytic method based on integer wavelet transform that can detect stego images using Xuan's method. Besides capability of detecting the hidden image, the algorithm can estimate the length of embedded data reliably. To increase detection ratio, we also research another method based on Neural Networks (NNs) and Convolutional Neural Networks (CNNs). Recent CNN works focused on problems about computer vision, such as identification of 3D objects, natural images, and traffic signs [8, 9, 10] image denoising [11] and image segmentation [12]. Convolutional architectures also seem to benefit of unsupervised learning algorithms used for analyzing image data [13, 14, 15].

In the next section, we describe again the Xuan's steganography method. In section 3, 4 we review again our proposed steganalytic methods and introduce new method on NNs and CNN's. Our experimental results are displayed in section 5. Finally, the conclusion is given in section 6.

2 Lossless Data Hiding based on Integer Wavelet Histogram Shifting

In this section, we describe Xuan's IWH algorithm introduced in [5]. This algorithm does not cause distortion by hiding information on the integer wavelet domain. The image space domain after being transformed to integer wavelet domain will be divided into four sub-bands. Xuan et al hid secret information on three high-frequency domains. The details of the IWH algorithm are summarized as follows:

Suppose there are M bits of secret information into a high frequency band. The IWH algorithm performs the following steps:

Step 1: Choose the threshold T (T> 0), so that number of the coefficient in the range [-T, T] is greater than M. Set Peak = T.

Step 2: In the histogram of the wavelet coefficient, shift the histogram column (the value of the histogram is greater than the Peak) to the right by a unit to create a Zero column at the Peak +1 position. Secret information is hidden in this location. Scanning all the coefficients of the high frequency band, if the coefficient is equal to Peak, the secret bit is 1, add 1 to the Peak coefficient to become Peak + 1, the secret bit is 0, the value of coefficient doesn't change.

Step 3: Keeping the secret information, change Peak = - Peak, shift the histogram column (the value of histogram is smaller than Peak) to the left by one unit to create the Zero column at (-Peak – 1). Information is hidden at this point.

Step 4: If all M bits have been hidden, the algorithm stops here and records the stop position S = Peak. In contrast, set Peak = - Peak -1, go back to step 2 to continue to hide secret bits remains.

3 PreviousSteganalytic Methods

The IWH algorithm is a Lossless Data Hidingalgorithm, but if you compare the integer wavelet coefficients, the histogram of the original image and the IWH stego image, you will see this difference. This is the key to estimate the information hidden in the stego image.

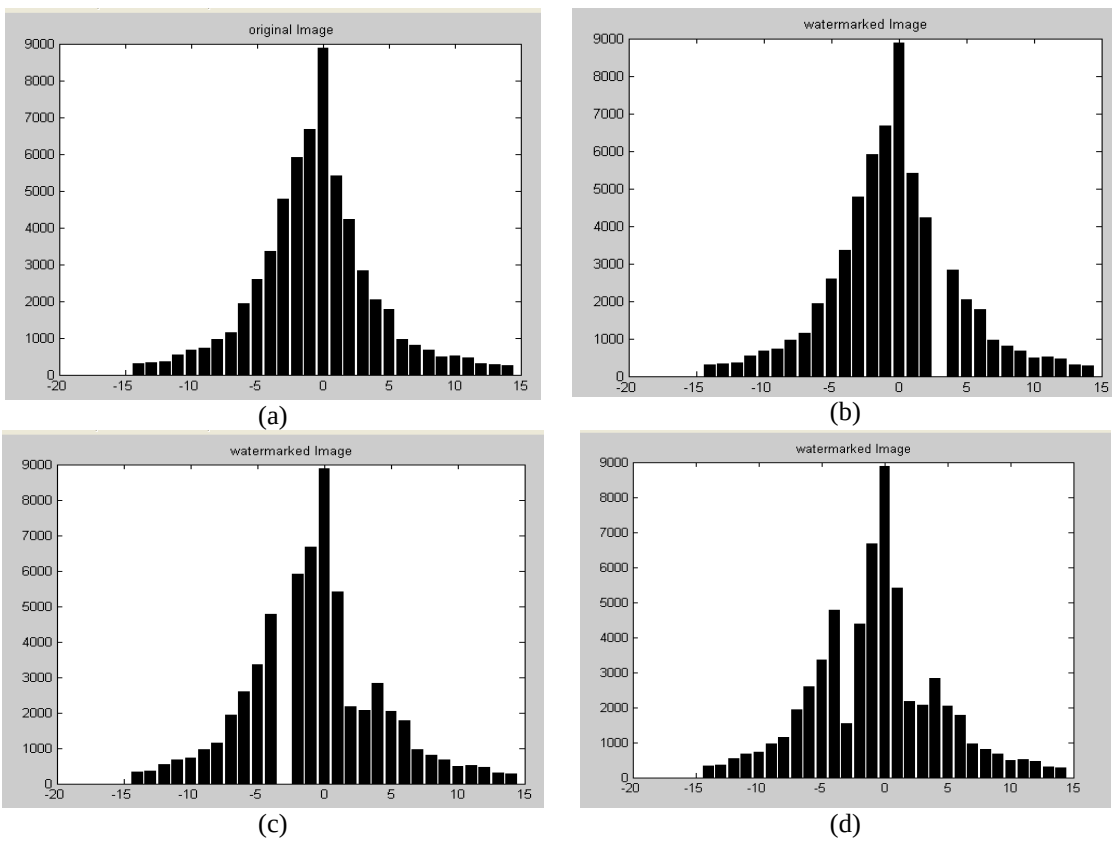
We first give analysis of occurrences in watermarking process as the three following experiments:

In the first experiment, we use Lena image of size 512 x 512 pixels (see Fig. 1. (b)) and Logo image of size 128 x 56 pixels to test (see Fig. 1. (a)). After integer wavelet transform, we calculate the histograms of high frequency sub bands (see Fig. 2. (a)). We next embed payload data (that is the binary sequence from Logo image) into the high frequency sub bands with T=2 using IWH method. We get S= -2 and calculate again histogram of the high frequency sub bands that is shown in Fig. 2. (d). The data embedding process performs via some steps: the first and second step embeds data in the point 2 and -2 (see Fig.2. (b), (c)) but

there are to been bedded data remaining, the process performs the third and fourth step with T= -2 to embed data (see Fig.2. (c), (d)).

a) b)

Fig.1: Test images: a) Lena original image, b) Binary logo image.



(a) (b)

(c) (d)

Fig.2: An example showing how a zero point is generated and payload data embedding process: (a) original histogram, (b) histogram after a zero point is created, (c) histogram after data embedding at Peak =2 and then a new zero point is created at new next Peak, (d) histogram after data remaining embedding with new Peak.

In the second experiment, we use also the Lena original image and Logo watermark with T=4, we then get S=3. In this case, the histogram is changed much that is shown in Fig.3. (a).

In the third experiment, we use the same input with T=6, we then get S=-5. In this case, the histogram is changed clearly that is shown in Fig.3. (b).

From the above three experiments we find that the wavelet coefficient histogram is slightly variable. In the natural image, the value column of the pair of coefficients (h_i, h_{-i}) is approximately equal and symmetric across the value h_0 column, the h_i value is usually greater than h_{i-1} (Fig. 2. (a)). However, after hiding it broke the nature (Fig. 2. (d), Fig. 3. (a), (b)).

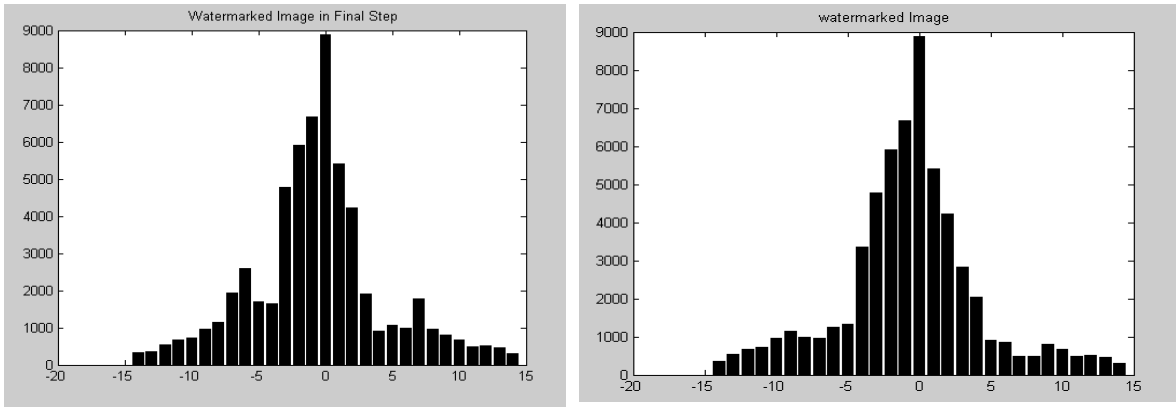


Fig.3: Another example showing how payload data embedding process: (a) the histogram after data embedding with chosen T=4, (b) histogram after data embedding with chosen T=6.

From the analysis we can estimate length of the information be hidden in the stego images. Details of the algorithm show the following as:

Step 1: Set L=0 (length of message), scan all column h_i (i >0 and i <=max (all integer wavelet coefficient of high sub bands)), if the first $\frac{(h_i+h_{i+1})}{2} < h_{i+2}$ is met, stop scanning, set Peak = i be first location to estimate data length.

Step 2: if $h_{Peak} \approx h_{Peak+1}$, $L=L+h_{Peak}+h_{peak+1}$; set Peak = -Peak and perform next step 3. Contrariwise, perform step 4.

Step 3: if $h_{Peak} \approx h_{Peak+1}$, $L=L+h_{Peak}+h_{peak+1}$; set Peak = -Peak - 1 and return step 2. Otherwise, perform step 4.

Step 4: if $h_{Peak+1}<h_{Peak+2}$ and $h_{Peak+1}<h_{Peak}$ then $L= L+2*h_{Peak+1}$. The process stops here.

Applying the proposed algorithm for the three experiments we can estimate the length of information hidden in the image according to Table 1.

Table 1: Estimating the length of information hidden on Lena image

Embedded data length	Chosen threshold T	Gotten Stop value S	Estimated data length
7168	2	-2	7231
7168	4	3	6998
7168	6	-5	7177

4 Steganlysis based on Neural Networks

Base on the analysis in section 2, 3, we found out classify images by Artificial Neural Networks – NNs and Convolutional Neural Networks - CNNs.

On NNs, the same architecture is used for experiments on 1000 cover images which were downloaded from [16, 17] and 500 stego images from 1000 cover images. We transform all pixel of cover image and stego image to integer wavelet frequency domain, then calculate histogram of the wavelet coefficients h=[h_{-128}, h_{-127}, …, 0,…, h_{126}, h_{127}], each h_i divided by max(h) to decrease h_i in value range [0,1] to increase the accuracy of the network training process. We consider the vector h is the characteristic vector as the input neuron (each h_i being an input neuron) for training the network. Using a hidden layer with 20 neurons, the neurons of the output layer are two neurons, using the sigmoid function to summary the output values of

each class, the weights of the neurons in each class are initialized accordingly so that the smallest possible output error (Fig. 4).

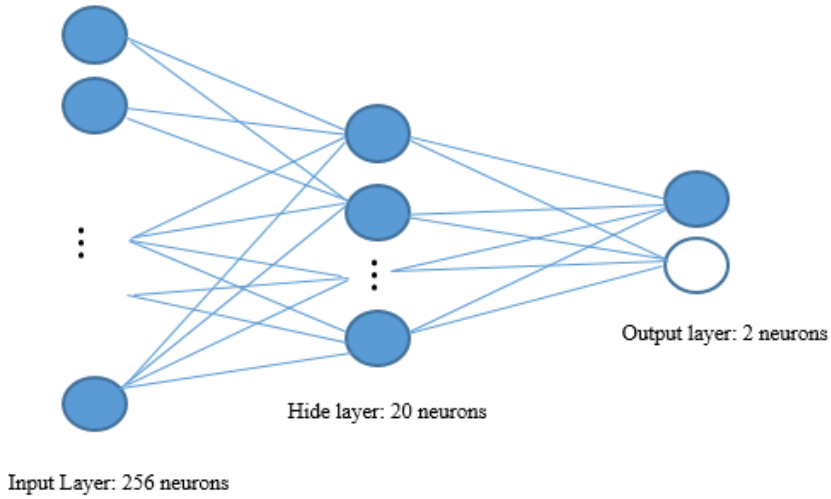


Fig.4: ANN architecture for the proposed method

On CNNs, use CNNs as base classifiers [11]. This network consists of 6 layers, 1 input layer, 2 convolution layers, 2 max – pooling layers and one layer of output. The input layer consists of 16 x 16 neurons (corresponding to 16 x 16 = 256 value of the wavelet coefficient histogram). The first hiding layer (the first convolution layer) consists of 6 maps of 12 x 12 neurons with 5 x 5 filter windows. The next hidden layer (max-pooling) is the output of the first convolution layer. Six mappings of 6 x 6 neurons use a 2 x 2 filter core.

The next convolution consists of 12 mappings of 6 x 6 neurons with 5 x 5 kernel. The second max - pooling layer consists of 12 mappings of 2 x 2 neurons using the 2 x 2 filter kernel. The output layer has a neuron per layer (corresponding to the original image layer and the stego image) (Fig. 5).We pick the trained CNN with the lowest validation error, and evaluate it on the corresponding test set.

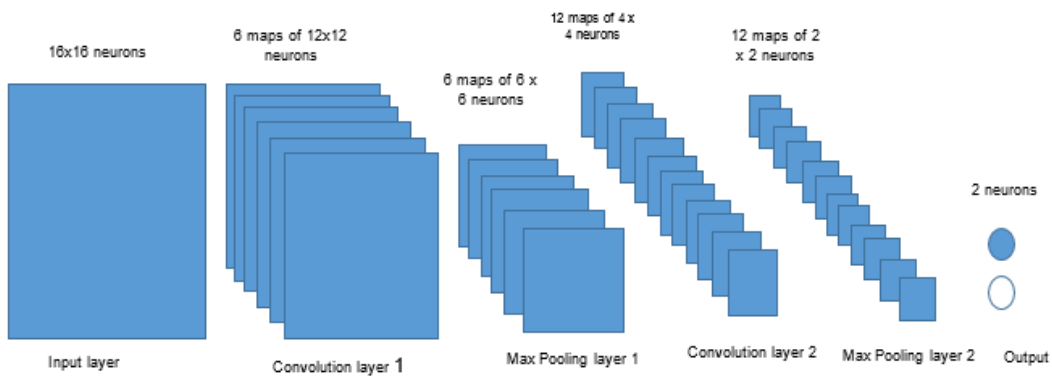


Fig. 5: CNN Structure for the proposed method

Prior to training, we give the detail of choosing input neurons of CNNs for the above steganographies: In a cover image or stego image in train dataset transforms integer wavelet transform, calculate the histograms h_w of high frequency sub bandswith values in set {-128… 127} (including 256 values of the wavelet coefficient histogram), divide h_D by max value of them to reduce the value 0 – 1, then reshape h into a vector of 16 x 16 elements. We use the vector as 16x16 input neurons of CNNs.

5 Experimental Results

Having a set of images, it includes 2088 images.They were downloaded from [15], [16] and they were created from my digital camera, all images are then converted to grayscale images by Photoshop CS2 software. This set of photos is created into two subsets they will be used to test the proposed detection or classification methods in above session as follows:

The first data set is used to detect by the estimation method includes:

+ 2088 cover images

+ 2088 stego images with hidden information is a 2000 bits binary string generated randomly.

The second data set is used for two neural networks includes:

+ The training set: 1500 images including 500 cover images and 1000 stego images which are embedded the randomly secret binary sequence of 2000 bits and 6000 bits into corresponding 500 original images by IWH method.

+ The testing set: 3676 cover images and stego images which are embedded the randomly secret binary sequence of 2000 bits or 6000 bits.

Proceeing to test three scenarios for two proposed approaches to detect hidden images using IWH hiding techniques:

+ Case 1: using estimating the information hidden in the wavelet coefficient domain of the image for the first data set, the images are hidden with the amount of information is 2000 bits, we obtain the estimation results as shown in Fig. 5. There the horizontal axis represents image number # and the vertical axis represents the embedded data length corresponding image number #.

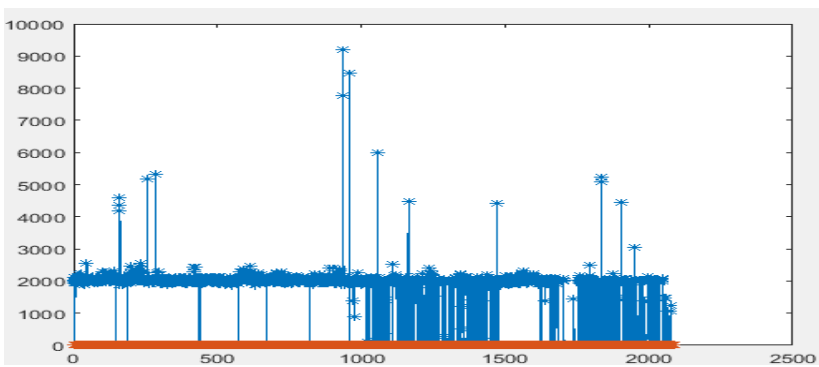


Fig.6: Estimate the hidden information for 2088 images with the hidden message of 2000 bits

Using estimating the information hidden in the wavelet coefficient domain of the image, the images are hidden with the amount of information is 6000 bits, we obtain the estimation results as shown in Fig. 6.

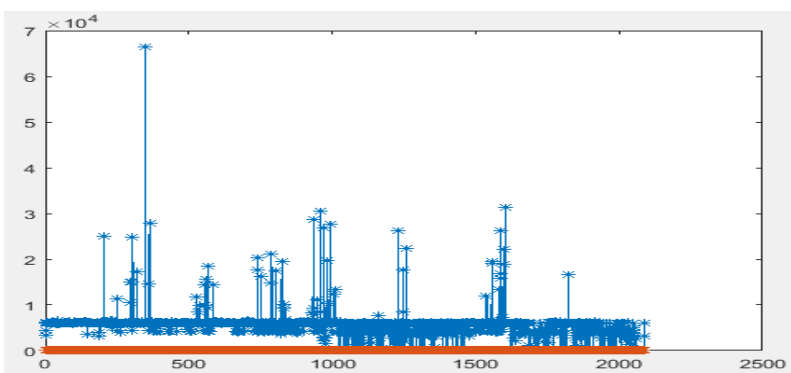


Fig. 7: Estimate the hidden information for 2088 images with the hidden message of 6000 bits

According to the first method the results are as follows:
- Original image (estimated information hidden by 0) - false detection of 440 image sat 21.07%
- Stego images with embedded 2000 bits (correct detection) rate reached 88. 93%, the average estimate is 1581.42 bits with a deviation of 933.39.The execution time is 147.43 seconds.

- Stego images with embedded 6000 bits (correct detection) rate reached 91.23%, the average estimate is 5301 bits with a deviation of 3165.The execution time is 156.65 seconds.

+ Case 2: using NNs to classify images for the second data set with epoch size options into training with 100 neurons (Number of batch), the results are as shown in the following table 2.

Table 2: The image classification resulting using NNs

Epoch	Training time (second)	classifying time (second)	True rate
10	0.27	0.19	94.31
50	1.332	0.22	96.32
100	2.81	0.18	96.21
200	5.37	0.2	95.72
300	8.12	0.2	95.51
500	14.14	0.2	95.23
Average		0.2	95.55

According to Table 2, the average detection rate for the data set is 95.55%, the classifying time is 0.2 seconds.

+ Case 3: using CNNs to classify images with epoch size options into training with 100 neurons (Number of batch), the results are as shown in the following table 3.

According to Table 3, the average detection rate for the data set is 93.05%, the classifying time is 0.38 seconds. However, the training time of CNNs is higher than the training time of NNs. From the three test cases, the results obtained by the method of message estimation are lower than those found in the Neuron network. In fact, CNNs are rated better than NNs, but in this case, NNs are better at detecting CNNs. This may be because the number of entry layers of CNNs in this case is much lower than those used in [11].In addition, the training time of CNNs is many times higher than that of NNs.

Table 3: The image classification resulting using CNNs

Epoch	Training time (second)	classifying time (second)	True rate
10	7.53	0.38	91.31
50	35.54	0.35	94.32
100	78.73	0.40	94.21
200	142.72	0.36	93.72
300	189.31	0.39	92.51
500	329.69	0.37	92.23
Average		0.38	93.05

6 Conclusions

In this paper, introducing two methods by using NNs and CNNs for better results. The two results show 96% correct detection rates for NNs and 94% for CNNs that indicates the reliability of the methods, this is a better result than old method [7]. Combining the old method [7] and new method to classify cover and stego images using IWH method. The first, using the new method (using on NNs or CNNs) to detect, then using the old method to estimate the hidden information.

However, it is hard to detect stego image with two factors which are shown in section 5. Noting that, there are many elements in this algorithms that can be changed or replaced with other elements. This research can be used to detect hidden images on spatial, frequency, or other domain.

References

[1]. O. Egger, M. Kunt, Embedded zerotree based lossless image coding, Proceedings., International Conferenceon Image Processing, Washington,DC,USA,USA,2015, https://doi.org/10.1109/icip.1995.537710

[2]. Ni, Z., Shi, Y., Ansari, N., Su, W.: Reversible data hiding. Proc. ISCAS (2003) 912–915. https://doi.org/10.1109/iscas.2003.1206123

[3]. Sang-Kwang Lee, Young-Ho Suh, and Yo-Sung Ho, Lossless Data Hiding Based on Histogram Modification of Difference Images, PCM 2004, LNCS 3333 ,2004, 340–347. https://doi.org/10.1007/978-3-540-30543-9_43

[4]. Xuan, G., Zhu, J., Chen, J., Shi, Y., Ni, Z., Su, W.: Distortionless data hiding basedon integer wavelet transform. IEEEElectrionics Letters ,2002, 1646–1648.https://doi.org/10.1049/el:20021131

[5]. GuorongXuan, Qiuming Yao, ChengyunYang, JianjiongGao, Peiqi Chai, Yun Q. Shi, Zhicheng Ni, Lossless Data Hidding Using Histogram Shifting Method Based on Integer Wavelets,Proc. 5th Digital watermarking workshop, IWDW 2006, Korea, vol. 4283, pp. 323-332.https://doi.org/10.1007/11922841_26

[6]. Ho Thi Huong Thom, Ho Van Canh, Trinh Nhat Tien, Steganalysis to Reversible Data Hiding, Proceedings of FGIT 2009 (the Future Generation Information Technology Conference) on Database Theory and Application, Springer-Verlag, Jeju Island, Korea (2009), pp. 1- 6. https://doi.org/10.1007/978-3-642-10583-8_1

[7]. F.-J. Huang and Y. LeCun, Large-scale learning with svm and convolutional nets for generic object categorization, in Proc. Computer Vision and Pattern Recognition Conference (CVPR'06). IEEE Press, 2006. https://doi.org/10.1109/cvpr.2006.164

[8]. D. C. Ciresan, U. Meier, J. Masci, and J. Schmidhuber, A committee of neural networks for traffic sign classification," in International Joint Conference on Neural Networks, to appear, 2011. https://doi.org/10.1109/ijcnn.2011.6033458

[9]. Stamatios Lefkimmiatis, Non-local Color Image Denoising with Convolutional Neural Networks, in IEEE Conference on Computer Vision and Pattern Recognition 2017 (CVPR 2017). https://doi.org/10.1109/cvpr.2017.623

[10]. Y.A. Alotaibi, High performance Arabic digits recognizer using neural networks, in Proceedings of the International Joint Conference on Neural Networks, 2003. https://doi.org/10.1109/ijcnn.2003.1223444

[11]. H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations, in Proceedings of the 26th International Conference on Machine Learning, 2009, pp.609–616. https://doi.org/10.1145/1553374.1553453

[12]. M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, Deconvolutional Networks, in Proc. Computer Vision and Pattern Recognition Conference (CVPR 2010), 2010. https://doi.org/10.1109/cvpr.2010.5539957

[13]. Hyeonseob Nam, Bohyung Han, Learning Multi-domain Convolutional Neural Networks for Visual Tracking, in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR2016), https://doi.org/10.1109/cvpr.2016.465

[14]. D. C. Ciresan, U. Meier, L. M. Gambardella, and J. Schmidhuber. Convolutional neural network committees for handwritten character classification. In International Conference on Document Analysis and Recognition (2011), pages 1250–1254. https://doi.org/10.1109/icdar.2011.229

[15]. CBIR image database, University of Washington, available at: http://www.cs.washington.edu/research/imagedatabase/groundtruth/

[16]. USC-SIPI Image Database, http://sipi.usc.edu/services/database/Database.html.

Author'sBiography

Ho Thi Huong Thom: Ho Thi Huong Thom received the B.S. degree of Information Technology department from Haiphong Private University, the M.S. degree and PhD degree in Information Systems from College of Technology, Vietnam National University in Vietnam, in 2001, 2005 and 2012, respectively. She started her career as Lecturer in Department of Information Technology in Haiphong Private University, Vietnam and served 12 years. From 2014, She has been teaching at Vietnam maritime university (VMU).

Nguyen Kim Anh: Nguyen Kim Anh obtained her master's degree in 2009 in Computer Science and Engineering from Centre for Development of Advanced Computing, Noida, India and her bachelor's degree in 2005 in Information Technology form Vietnam Maritime University. She is currently working as a lecture in Information Technology Department in Vietnam Maritime University.

Bui Dinh Vu: Bui Dinh Vu obtained a bachelor's degree in Information Technology from Hanoi National University in 1998. He obtained his master's degree in Computer Science from Hanoi Military Technical Academy in 2006. He currently working at the Faculty of Information Technology, Vietnam Maritime University.

How to Cite

Thom, Ho Thi Huong, Anh, Nguyen Kim and Vu,Bui Dinh "Steganalysis for Reversible Data Hiding Based on Neural Networks and Convolutional Neural Networks*",International Journal of Machine Learning and Networked Collaborative Engineering*, Vol. 02 No. 02, 2018,pp.40-48. doi: https://doi.org/10.30991/IJMLNCE.2018v02i02.001.