Journal of Applied Computer Science & Artificial Intelligence

Research Paper Vol.-2, Issue- 1, June 2021 ISSN: 2582-8673

Spatial domain steganalysis using convolutional neural network features

S.T.Veena[1*], S.Arivazhagan[2]

[1] Department of CSE, Mepco Schlenk Engineering College, Sivakasi, Tamilnadu, India, Email ID:veena_st@mepcooeng.ac.in
2 Department of ECE, Mepco Schlenk Engineering College, Sivakasi, Tamilnadu, India

Available online at: http://jacsai.org/

***Abstract*—** Recent studies have indicated the use of deep convolutional neural networks (CNNs) as a successful framework for steganalysis. However the cost of deep CNN is high due to its GPU and RAM and it is not possible in a lower and system. Therefore In this paper, a shallow CNN architecture for rich discriminant feature extraction is proposed. The CNN of a single layers depth is proposed with preprocessing, convolution and pooling layers. The features are extracted from this CNN and are then used for classification purpose. The features contain macroscopical information hidden in the image and are at large useful for steganalysis. Since steganalysis process requires content-less information and is provided by CNN. The extracted features are then classified using an ensemble fisher linear discriminant (FLD) classifier. The experimentation is performed for three content adaptive spatial domain algorithms and the results are much better than the state-of-the-art steganalysers.

***Keywords*— convolutional neural network , spatial domain steganalysis , ensemble classifier, shallow CNN, deep features**

I. INTRODUCTION

Steganalysis is an art of detecting the presence of malicious information in an innocuous cover media. Digital images provide large area and easy access to data hiding as the cover medium. Spatial domain steganographic approaches challenge the steganalysts with their relatively lower embedding change rate than its JPEG counterpart [1]. The state-of-the-art spatial domain steganographic algorithms are content adaptive in nature. These robust content adaptive schemes embed data adaptively into the most crucial data (textural region) of the cover medium using a minimum distortion strategy. Algorithms like Highly Undetectable steGanOgraphy (HUGO) [2], Wavelet Obtained Weights (WOW) [3], Spatial UNIversal WAvelet Relative Distortion (S-Uniward or SW) [4] fall under this discipline. The current state-of-the-art steganalysis is a two phase process. The first phase deals with building feature models from various residuals. Spatial Rich Model (SRM) [5], Projected SRM (PSRM) [6] are a few among them and the success of steganalysis lies on these hand picked features. The second phase deals with classifying the extracted features using a classifier. The classifier may be a support vector machine (SVM), an artificial neural network (ANN) or an ensemble classifier. The ensemble classifier with reduced complexity than SVM or ANN has the advantage to handle huge feature dimensionality. Recently developing or steganalysis of spatial ing.

*Corresponding Author

Deep learning models are a class of hierarchical machine learning models that learn features automatically under the supervision of the classifier. Deep learning has been exploited in various artificial tasks like identification and classification of objects, processing of natural languages, etc. [7, 8, 9]. However, their application in steganalysis is very limited because of the following reasons. 1) The stego signal is a weak noise which is usually ignored by the deep models. 2) Training needs to be on large scale images say (256 _ 256) rather than on small sub-sampled ones which erase stego noise. Due to these limitations, steganalysis has been applied mainly to a ”clairvoyant scenario” where the steganalyst has a good distribution of the cover images used by the steganographer along with the knowledge of the algorithm being embedded and the size of the payload used. This lab scenario matches Kerchoff’s law. In addition, a hypothesis that the same key is used to create the stego images (i.e) the simulator uses the same stego key is assumed.

In this paper, a shallow CNN architecture is proposed for a clairvoyant scenario steganalysis where the features trained from a single convolution layer output are extracted and classified by an ensemble classifier. This proposed model does not require any high end system or processor as other CNNs and achieves better accuracy of more than 99% accuracy in lesser time. The organisation of the paper is as follows: Section 2 introduces the various current works related to the discussion. Section 3 proposes a CNN model for spatial

© 2021, JACSAI All Rights Reserved

steganalysis. The ensemble classifier is briefly explained in Section 4. The experiments conducted are detailed in Section 5. Section 6 concludes with suggestions for future scope.

II. RELATED WORK

The initial work in this domain by Qian et al.[10] was a Gaussian neural CNN (GNCNN) architecture. The image processing layer of the GNCNN was a preprocessing layer where a high pass filtering with a kernel of size 5×5 was done to strengthen the weak stego signal and reduce the cover image content. In each of the five convolution layers, convolution filtering was followed by non-linearity activation function and pooling. The authors used 16 filters in each layer followed by Gaussian non-linearity activation function and overlapped average pooling with window 3×3 and step size 2. The first and the fifth convolution layer filters were of size 5×5 while the second to fourth layer filters were of 3×3 sizes. And the classification layer consisted of three fully connected layers. The first two fully connected layers output from 128 neurons were activated by Rectified Linear Units (ReLUs). And the last layer was connected to softmax function for binary classification. The GNCNN was applied to resized Bossbase v1.01 and ImageNet databases trained on a PC with Intel Xeon E5-2650 2.0GHz CPU and Tesla K40 12G GPU. The average training time on Bossbase database is about 2 hours and on ImageNet database is about 52 hrs. They tested the CNN for Bossbase images embedded with 0.3, 0.4 and 0.5 bpp payloads using HUGO, WOW and SW algorithms and for ImageNet images with only 0.4 bpp payload. The results were compared against SRM and SPAM features of size 34,671 and 686 respectively. The authors were able to surpass SPAM but obtained a detection error 2% to 5% higher than SRM features and the feature vector dimension did not compare favourably with respect to deep learning.

Pibre et al.[11] tested a larger number of CNNs for clairvoyant and cover-source mismatch scenarios and found that the most efficient network was a shallow CNN. It consisted of only two convolution layers, followed by three fully connected layers. They experimented with both cropped Bossbase and the LIRMMBase databases with images of size 256×256. The preprocessing layer was same as Qian et al.[10] followed by two layers of convolution. Each convolution with no sub-sampling was followed by ReLU activation function and normalisation across maps. The first convolution layer consisted of 64 filters of size 7×7 with stride 2 and the second layer consisted of 16 filters of size 5×5. The classification layer consisted of three fully connected layers with 1000 neurons and the last layer of them had only 2 neurons with softmax activation. In comparison to the Qian et al. network, the number of filters in the convolution layer were increased while the number of CNN layers were reduced. The increase in height provided a reduction of more than 17% in the classification error for SW algorithm 0.4 bpp payload in Bossbase database. They reasoned that the increase in layer number leads to a loss of information, probably due to the negative impact of the pooling step (sub-sampling).

Couchot et al.[12] proposed a CNN where the preprocessing was replaced by a convolution layer with a single filter of size 3×3, followed by another convolution layer of 64 filters with zero padding and stride equal to 1. The pooling operation was dropped out and the hyperbolic tangent activation function was used. The classification layer consisted of a single output layer of two softmax neurons. The authors conducted experiments on Bossbase and Raise databases with 512×512 size images embedded with 0.4 and 0.1 bpp payloads using HUGO and WOW algorithms and obtained an accuracy of 97%. The recent work by Xu et al, [13] proposed a deep CNN with a random filter for preprocessing followed by 5 convolution layers and linear classification module. The convolution layer was customised with absolute values of feature maps to improve statistical modelling and hyperbolic tangent activation was used to prevent over fitting. Though this model works on a single steganographic algorithm, the performance was good enough against SRM features.

III. PROPOSED METHOD

A. CNN Model

The Convolutional Neural Network (CNN) is one of the most notable deep learning approaches where multiple layers are trained in a robust manner. Generally, a CNN consists of three main neural layers, which are convolution layers, pooling layers, and fully connected layers. Different kinds of layers play different roles and in steganalysis image preprocessing layer is used additionally to enhance the stego signal embedded inside the stego image. Also, the stego noise added to the cover is a kind of very weak signal and is greatly impacted by the cover image content. A high pass filtering operation as in Eqn 1, helps to strengthen the weak stego signal and reduce the impact of the cover.

$$PI = I \times K \tag{1}$$

where I is the image, K is the high pass filter kernel used and PI is the preprocessed image. This can drive the network to good initialization and helps to achieve good performance as compared to random initialization. In the preprocessing layer, as suggested by Qian et al.[10] a high pass filtering with a 5×5 kernel K with no padding is done.

$$K = \frac{1}{12}\begin{bmatrix} -1 & 2 & -2 & 2 & -1 \\ 2 & -6 & 8 & -6 & 2 \\ -2 & 8 & -12 & 8 & -2 \\ 2 & -6 & 8 & -6 & 2 \\ -1 & 2 & -2 & 2 & -1 \end{bmatrix} \tag{2}$$

This is followed by the convolution layer. Generally, a convolution layer is made of three steps i.e. convolution, activation, and pooling. These three consecutive steps provide

© 2021, JACSAI All Rights Reserved

a link between the feature maps from a layer to the previous one .

$$CI = pool(f(b + \sum_{i=1}^{N} F_i)) \quad (3)$$

where CI is the convolved image, Fk denotes the kth filter, N is the number of filters in the layer (N is also the number of feature map outputs by the layer), b is the bias to the convolution, f() is the activation function applied to the filtered image, and pool() denotes the pooling operation. In the proposed architecture, a single layer of convolution is employed. The convolution consists of 16 kernels of size 5×5 with the stride equal to 2.

According to Pibre et al.[11], activation is done to break the linearity property resulting from the linear filtering and activation by ReLU (f (x) = max(0,x)) is best for steganalysis. Similarly, the authors suggest pooling operation of computing average on local neighbourhood helps to reduce the variance obtained during convolution and is better than Max pooling in steganalysis since the stego signal is weak. Thus, in the proposed architecture, ReLU and average pooling are considered. The pooling is accompanied by sub sampling with the stride equal to 2 in order to reduce the size of the obtained feature map and can be visualised as down-sampling accompanied with low pass filtering. This is useful to reduce the memory occupation in the GPU (Graphical Processing Unit) where a high end GPU is not available. Further, experimentation with two different pool sizes of 2 and 5 are done. The last layer is a single fully connected layer for binary classification followed by a softmax layer. Thus, a very shallow CNN is proposed. Once the CNN is designed and trained, the feature maps from the output of the convolution layer are extracted and fed to the ensemble classifier. This is because the exploitation of deep features is sometimes much better than full trained CNN system. Also, this method reduces the classification time [14, 15]. The pipeline of the proposed CNN architecture is shown in Fig. 1.

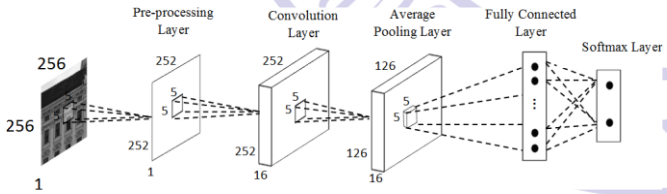


Figure 1. Example of a figure caption.

B. ENSEMBLE FLD CLASSIFIER

Ensemble classifiers are in general more favourable than the SVMs and ANNs becauseof their less complexity and better performance quality. Here in the paper, an ensemble classifier model proposed by Kodovsky et al.[16] is used. The basic learners are the Fischer Linear Discriminants (FLDs), which are of low complexity. The ensemble random forest of basically weak learners is combined together to form a strong classifier. This ensemble classifier is an efficient classifier for binary classification. Also, the classifier works on a random subspace of features which helps it to handle even very large dimensional feature models. Thus, in this ensemble FLD, the random set of base learners are trained on the bootstrap samples of the training set and each base learner is a decision tree whose split is done on a random small subset of the original features. The final prediction is formed by the majority vote of base learners. The schematic diagram of the ensemble classifier is given in Fig.2

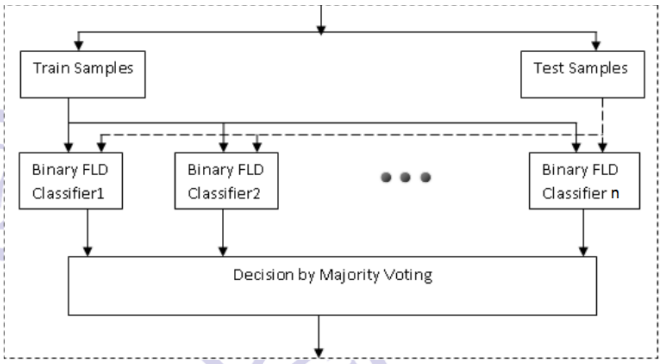
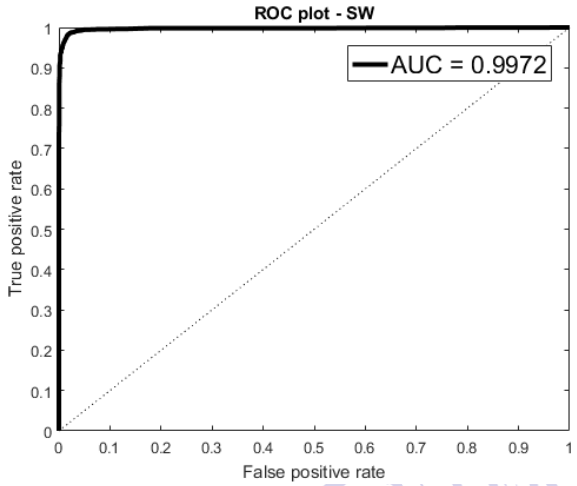


Figure 2. Schematic diagram of ensemble FLD used

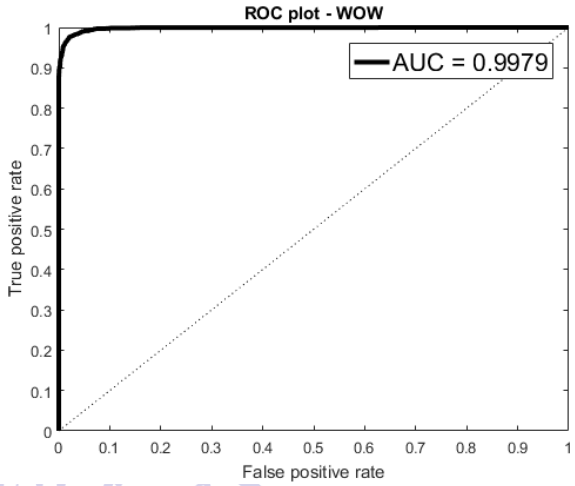
IV. RESULTS AND DISCUSSION

This section analyses the experiments conducted using the proposed CNN and ensemble classifier on a cropped Bossbase v1.01 database [17]. This database contains 10,000 images acquired by seven digital cameras in raw format of size 512×512 pixels. In our experiments, the images are cropped to the size of 256×256 pixels. This processing of cropping is only due to the limitation of the computational capabilities. The cropping was done by randomly choosing the region of the source. To evaluate the effectiveness of the developed model for steganalysis, experiments were conducted on the images embedded with a random fixed payload of 0.4 bpp using the three state-of-the-art spatial domain steganographic algorithms HUGO, WOW and S-UNIWARD (SW). The proposed CNN network was trained on a PC with Intel Xeon E5-1607 3.0GHz CPU and Quadro K2000 2GB GPU. This is a comparatively a very low end GPU than the GPUs used in other related works. The preprocessing layer produces an output image of size 252×252. Sixteen convolutions applied to the input image, generates 16 feature maps, each of size 126 126 leading to feature of dimension 61,504 and 57600 with pool size 2 and 5 respectively. The train test ratio is set at 80:20. The main parameters are given for the reproducibility of experiments as follows. 'WeightLearnRateFactor' and 'BiasLearn- RateFactor' of the neural network are kept at 0.001 and 0.002 respectively. The images are grayscale images and therefore the number of channels is one. The other parameters like 'MiniBatchsize', 'Moment' and 'maximum epochs' are set as 128, 0.9 and 30 respectively. The extracted features from the convolution layer of the trained CNN are now fed to the ensemble classifier. The output of the results is shown in Fig 3.

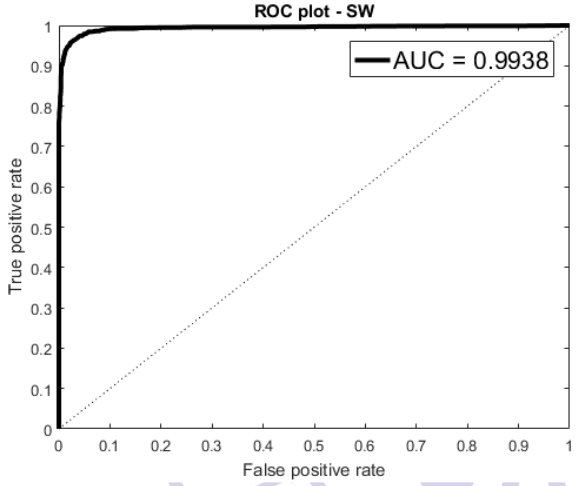
© 2021, JACSAI All Rights Reserved



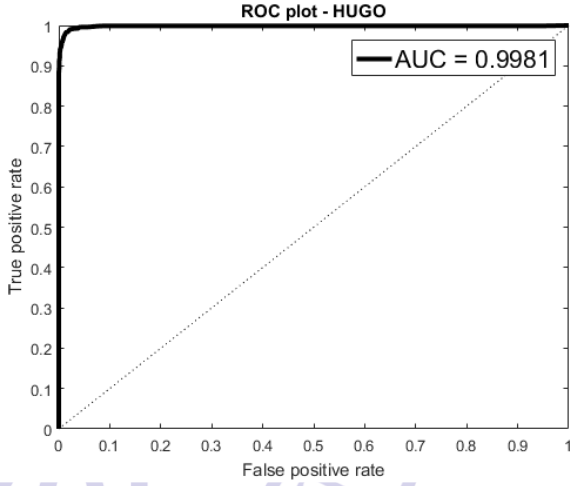
(a) Pool size=2



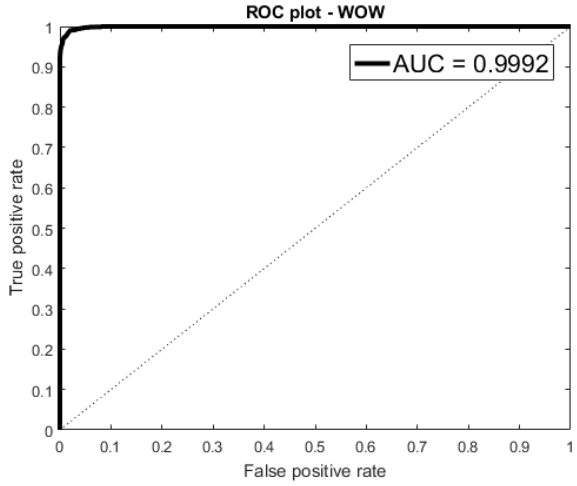
(d)Pool size=5



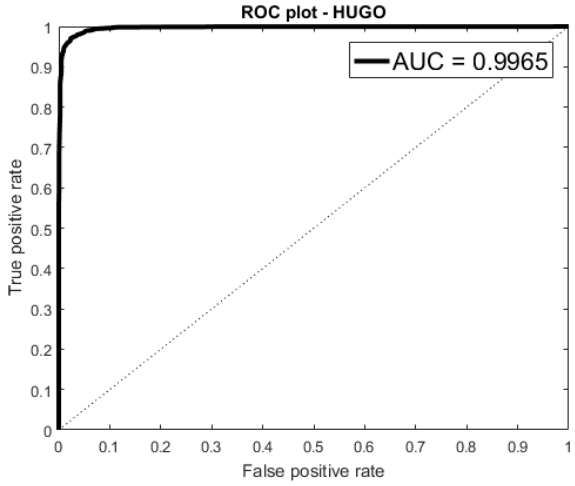
(b) Pool size =5



(e) Pool size=2



(c) Pool size=2 (f)

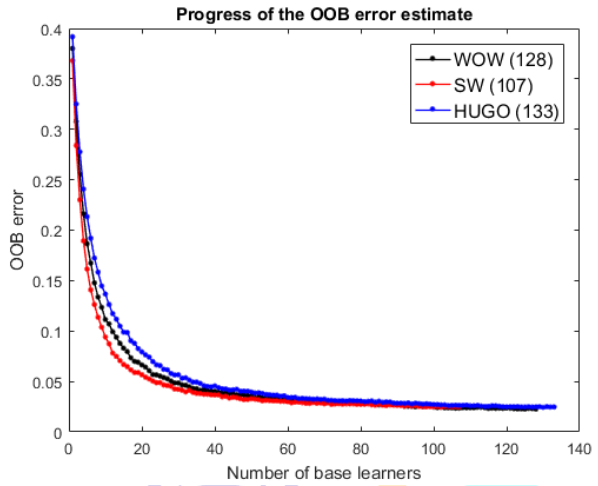


(f) Pool size=5

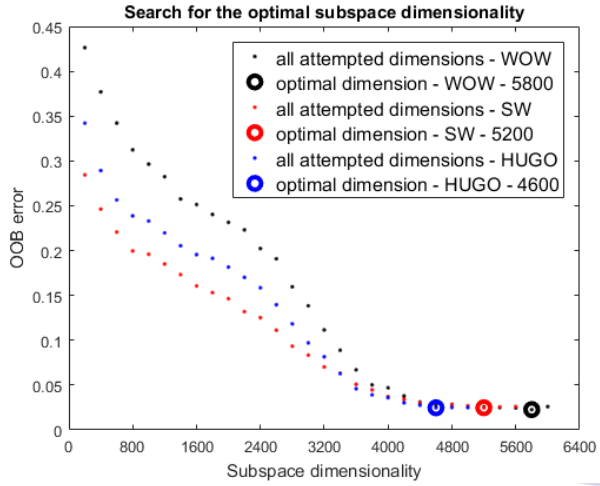
© 2021, JACSAI All Rights Reserved

Figure 3: Steganalysis of Proposed CNN on various algorithms for a payload of 0.4 bpp of pool size 2 and 5

It can be noted that CNN with pool size equal to 2 is better in steganalysis than its counterpart. This is because strong dependencies exist between the immediate neighbours and are captured better by pool size 2. The convergence of subspaces and base learners of the ensemble classifier with respect to the out of bag (OOB) error is given by Fig. 4.



(a) Base learners



(b) Dimensionality

Figure 4: Variation of base learners and dimension using ensemble FLD

Thus, in a random feature space of 4600 to 5800 (original feature dimension is 61504) an accuracy of 99% has been achieved. Also, the number of base learners required is as low as 133. The maximum time taken for classification is 4782.79 seconds with a minimum of 30 iterations for CNN. The comparison of results with other works is given in Table.1.

Table 1: Comparison of the percentage of detection error obtained for a payload of 0.4 bpp and various algorithms with existing works

	HUGO	WOW	SW
Proposed Method	1.58	1.47	1.82
Qian et al.[10]	25.7	29.3	30.9
Pibre et al.[11]	-NA-	-NA-	7.4
Couchot et al.[12]	2.91	4.56	-NA-
Xu et al.[13]	-NA-	-NA-	19.76

V. Conclusion and Future Scope

In this paper, use of shallow CNN architecture has been studied for clairvoyant scenario steganalysis and the extracted rich features from the proposed CNN are classified with an ensemble classifier (with FLD as its base) which produces better results than the state-of-the-art CNN for steganalysis. This proposed CNN can be deployed in low end GPU or CPU driven systems also with a reduced runtime. It has been tested on three state-of-the-art spatial steganography for a payload of 0.4 bpp and an accuracy of 99% is obtained in a time period (both training and testing) of maximum 1.39 hrs (approximately 4,000 seconds for CNN+1,000 seconds for ensemble classification). The future scope will concentrate on improving some network parameters, and on gaining insight into the network behaviour. Moreover, further experiments have to be done with different payload sizes, and different databases.

Acknowledgment

The authors would like to thank Defence Research and Development Organisation for providing the Quadro K2000 GPU. They would also like to express their gratitude to the anonymous editors and reviewers for their helpful suggestions and constructive comments. Also, the authors would like to express their sincere thanks to the Management and Principal of MSEC for providing the necessary facilities and support to carry out this research work.

References

[1] J. Fridrich, M. Goljan, and R. Du, “Detecting LSB steganography in color, and grayscale images,” IEEE MultiMedia, Vol. 8, no. 4, Oct. 2001, pp. 22–28.

[2] T. Pevn´y, T. Filler, and P. Bas, “Using high-dimensional image models to perform highly undetectable steganography,” in International Workshop on Information Hiding. Springer, 2010, pp. 161–177.

[3] V. Holub and J. Fridrich, “Designing steganographic distortion using directional filters,” in IEEE InternationalWorkshop on Information Forensics and Security. IEEE, 2012, pp. 234–239.

[4] V. Holub and J. Fridrich, “Digital image steganography using universal distortion,” in Proceedings of the first ACM workshop on Information hiding and multimedia security. ACM, 2013, pp. 59–68.

[5] J. Fridrich and J. Kodovsky, “Rich Models for Steganalysis of Digital Images,” IEEE Transactions on Information Forensics and Security, Vol. 7, no. 3, Jun 2012, pp. 868–882.

[6] V. Holub, J. Fridrich, and T. Denemark, “Random projections of residuals as an alternative to co-occurrences in steganalysis,” in Media Watermarking, Security, and Forensics, Vol. 8665, 2013, pp. 86 650L–86 650L–11.

© 2021, JACSAI All Rights Reserved

[7] C. Gao, P. Li, Y. Zhang, J. Liu, and L. Wang, "People counting based on head detection combining Adaboost and CNN in crowded surveillance environment," Neurocomputing, Vol. 208, Oct. 2016, pp. 108–116.

[8] Y. Dong, Y. Liu, and S. Lian, "Automatic age estimation based on deep learning algorithm," Neurocomputing, Vol. 187, Apr. 2016, pp. 4–10.

[9] A. Soleimani, B. N. Araabi, and K. Fouladi, "Deep Multitask Metric Learning for Offline Signature Verification," Pattern Recognition Letters, Vol. 80, Sep. 2016, pp. 84–90.

[10] Y. Qian, J. Dong, W. Wang, and T. Tan, "Deep learning for steganalysis via convolutional neural networks," A. M. Alattar, N. D. Memon, and C. D. Heitzenrater, (eds.), Mar. 2015, pp. 94 090J–1–10.

[11] L. Pibre, P. J´er˘ome, 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 source-mismatch," in Electronic Imaging, 2016.

[12] J.-F. Couchot, R. Couturier, C. Guyeux, and M. Salomon, "Steganalysis via a Convolutional Neural Network using Large Convolution Filters for Embedding Process with Same Stego Key," arXiv preprint arXiv:1605.07946, 2016.

[13] G. Xu, H.-Z. Wu, and Y.-Q. Shi, "Structural Design of Convolutional Neural Networks for Steganalysis," IEEE Signal Processing Letters, Vol. 23, no. 5, May 2016, pp. 708–712.

[14] L. Hertel, E. Barth, T. Kster, and T. Martinetz, "Deep convolutional neural networks as generic feature extractors," in International Joint Conference on Neural Networks. IEEE, 2015, pp. 1–4.

[15] G. Farias, S. Dormido-Canto, J. Vega, G. Ratt´a, H. Vargas, G. Hermosilla, L. Alfaro, and A. Valencia, "Automatic feature extraction in large fusion databases by using deep learning approach," Fusion Engineering and Design, Vol. 112, Nov. 2016, pp. 979–983.

[16] J. Kodovsky, J. Fridrich, and V. Holub, "Ensemble Classifiers for Steganalysis of Digital Media," IEEE Transactions on Information Forensics and Security, Vol. 7, no. 2, Apr. 2012, pp. 432–444.

[17] P. Bas, T. Filler, and T. Pevn´y, "Break Our Steganographic System: The Ins and Outs of Organizing BOSS," in Proceedings of the 13th International Conference on Information Hiding, ser. IH'11. Berlin, Heidelberg: Springer-Verlag, May 2011, pp. 59–70.

Authors Profile

Dr. S.T.Veena received her B.Sc degree in Computer Science from Madurai Kamaraj University in 1995, her M.C.A degree in Computer Applications from Alagappa University in 2001, her M.E degree in Computer Science and Engineering from Anna University in 2010 and completed her Ph.D in Information and Communication Engineering. Her interests include Steganalysis, Digital Image Processing, Steganography, Visual Cryptography, Pattern Recognition.

*Dr. S. Arivazh*agan born in Sivakasi obtained his B.E degree in Electronics and Communication Engineering, M.E degree in Applied Electronics and Ph.D in Image Processing. He is currently working as a Professor of Electronics and Communication Engineering at MSEC. Image Processing and Computer Communications are his major areas of interest in which he has published 56 International Journals. He is also a reviewer of many Journals. He is actively involved in research projects funded by different organisations (DRDO, DST, ISRO). He is a governing council member in IETE as well as member of ISTE and CSI. His interests include Digital Image Processing, Networking.

© 2021, JACSAI All Rights Reserved