Deep Learning based Effective Steganalysis

John Babu G, Sridevi Rangu

Abstract: ***There is an evident paradigm shift in steganalysis techniques with discovery of deep learning networks. As steganalysis is a classification task, it is done by machine learning classifiers and ensembles of them. But with the proliferation of deep learning and Convolutional Neural Networks in many areas, the performance of steganalysis techniques have jumped up to a another high, because of the application of Convolutional Neural Networks. The traditional steganalysis techniques consists two important steps, i.e., feature extraction and classification; where as deep learning networks learn the features automatically, eliminating the need of extraction of handcrafted features. Because of this feature CNNs were highly successful in image recognition and image classification techniques. In addition to that, feature extraction and classification are combined together in deep learning hence classification would be more effective because of the learning of the features which are really important for classification. But in Steganalysis the task is to detect very subtle and weak noise created by the hidden data with steganography techniques. We have designed a deep CNN architecture customized for steganalysis task based on existing residual neural networks frame. We have introduced a descriptor to capture the inter pixel dependencies and which acts as an indicator for weightage of a particular feature maps. Thus the classifier can give more weightage to effective feature maps instead of treating all the feature maps equally. We have also used a gating mechanism by using sigmoid function after nonlinear activation function sandwiched between two fully connected layers. This enhancement to the existing deep residual neural networks has given better results in terms of error detection rate compared to the other deep learning based steganalysis techniques.***

Keywords**:** ***Classification, Convolutional Neural Networks, Deep Learning, Steganalysis.***

I. INTRODUCTION

The word Steganography is derived from Greek, which means masked writing. The secret data to be communicated is hidden in a carrier such as text, image, audio or video. The objective of steganalysis is to conceal the very existence of presence of the secret data in the carrier. Whereas in Cryptography the objective is to decrypt the encrypted data. Steganography is an one time method as if the opponent knows the method of hiding the data, that method is useless and cannot be used any more at all.

On the other hand Steganalysis is opposite of Steganography, i.e., discovering the hidden data in a given carrier, even if the steganalyzer may not be able to extract the hidden information or decode the data, detection of the very presence of the secret data in a carrier can be termed as successful steganalysis technique. Though many carriers are in use for the steganographic techniques images have become very suitable choice for the majority of steganographers, because of the availability of digital images over Internet and the redundancy in the image data. In our research we focused on Image steganalysis.

Steganalysis can be described as a binary classification problem. Given an image the steganalysis technique should be effective enough to label the image either as an innocuous image which is called Cover image or image containing the secret data called Stego image.

II. TRADITIONAL STEGANALYSIS

As Steganalysis is a binary classification problem it was done using machine learning techniques until recently. At first the properties of images which will be altered by the hiding of the data, which are called features are to be discovered. This step is called feature extraction. Then to reduce the dimensionality of the problem a few features are selected, from the extracted features, it is called feature selection step. Then a classifier is trained to identify the stego images and cover images.

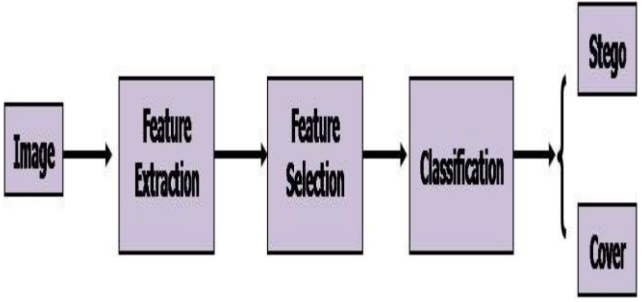

Fig. 1.Traditional Steganalysis

These Steganalysis techniques can be classified into two categories 1. Targeted Steganalysis techniques 2. Universal or blind Steganalysis techniques. Targeted steganalysis techniques are designed to detect any particular steganographic technique where as Universal steganalysis techniques are designed to detect the stego images irrespective of the steganographic technique. The major step in these steganalysis techniques is the extraction of features, which requires a great deal of expertise and knowledge of statistical properties of image which will be affected by hiding a secret data.

The features that are used in various steganalysis techniques are image quality metrics[4], wavelet statistical moments statistics [5] [6] [7] [8] [9] [10] [11][12][13] [14], cooccurence matrices[15] [16] [17], Binary similarity measures[18], Histogram based features[20], DCT features & Markov features[24] rich models, calibrated features and merged features[26].

Primary disadvantage with traditional steganalysis techniques is that features are to be handpicked, which is a very difficult task.

Revised Manuscript Received on March 02, 2020.

Mr. John Babu, Ph.D, Department of Computer Science & Engineering, JNTUH College of Engineering, Hyderabad, India.

Dr. Sridevi Rangu, Professor and Head, Computer Science & Engineering, JNTUH College of Engineering, Hyderbad, India.

Secondly the feature extraction step and classification steps are separate, as a result the insights from classification step cannot be used in the feature extraction step. Because of these reasons, the performance measured in terms of detection accuracy is low in traditional steganalysis. Till now performance of traditional methods is maximum in ensemble classifier with rich model features.

III. CNN BASED STEGANALYSIS

Convolutional Neural Networks are a category of deep learning networks which are a successful breakthrough in image classification. In Convolutional Neural Networks the feature extraction and classification are combined in a single step. i.e. features are not handpicked by the steganalyzer but features are extracted and selected automatically. Any feature missed in the feature extraction can be obtained in classification step as both are combined.

A CNN has four parts 1. Convolutional Layers 2. Pooling 3. Activation function 4. Fully connected layer

CNN automatically extracts features from images eliminating the step of handpicking of features. CNN learns the features where as conventional classifier is trained with features. Hence the CNN outperforms the of machine learning based classifiers and even ensemble classifiers.

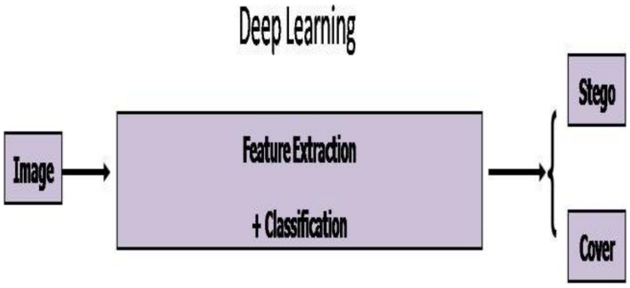

Fig. 2.CNN based Steganalysis

IV. RELATED WORKS

Tan and Li [1] used convolutional auto-encoders for steganalysis. This CNN consisted nine layers and three blocks used Stacked Convolutional Auto Encoders (SCAE). The performance of this CNN was far below the performance of SRM steganalysis.[2]. Qian et al.[3] designed a CNN with Gaussian function as the activation function. They have used a preprocessing filter of size 5x5 before sending the image to the convolutions layer. This CNN can be considered as a base for many other CNNs for steganalysis designed later. The preprocessing filter used by Qian has become a standard preprocessing filter for all the successive CNN based steganalysis techniques. Xu[4] incorporated domain knowledge of features in the training of CNN. This CNN architecture consisted five blocks of convolutions, average pooling and non-linear activation functions TanH and ReLU are used. Batch Normalization is used after each convolution. The fully connected layer to which 128-D features are passed produced the output label. The performance of this CNN surpassed that of SRM steganalysis technique. Qian and Dong[5] used the learning with huge payload images in Steganography and transferred this learning for steganalysis with low payload images. This methodology of this learning is termed transfer learning. Couchot et al. [6] designed a CNN for a case of steganalysis where the same key is used in steganography. Qian[7] have tried to enhance the noise created by steganogrpahy by passing the input image through a high pass filter and included domain knowledge the CNN architecture. Ye et al.[8] designed a CNN with huge number of layers and used a new activation function Truncated Linear Unit which has proven to be a better performed. Zeng et al[9],[10] have constructed a hybrid CNN with truncation and quantization which has outperformed the conventional steganalysis methods. Chen et al. have built a JPEG-phase –aware deep CNN for steganalysis. In JPEG domain ,they proposed a hybrid CNN steganalyzers equipped with quantization and truncation, which is obviously superior to hand-crafted JPEG steganalytic features. Chen et al. proposed a JPEG-phase-aware deep CNN steganalyzer [11]. Xu [12] proposed a CNN based on the successful image classification CNN architecture –ResNet[13]. Many other deep learning based CNNs have been devised by tweaking the hyper parameters of CNN[[14][15][16][17][18].

V. PROPOSED METHOD

Different variants of CNN architectures for image classification have been proposed in the recent past such as LeNet-5[19], AlexNet [20], ZFNet[21], GoogleNet[22], VGGNet[23], ResNet[24], ResNext[25]. We have chosen the residual network architecture for our research work for the following reasons. Firstly, The basis of residual network is very similar to that of steganalysis. As shown in the figure F(x) can be likened to the pure image and x can be likened to the secret data. F(x)+x represents the Stego image and F(x) represents the cover image. Secondly the residual networks support very deep architecture with less computational complexity.

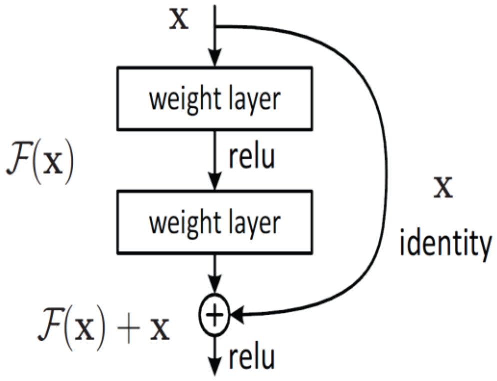

Fig. 3.Residual Network block

ResNet uses shortcut connections in the architecture thus avoiding the downgrading of performance at higher depths also. In general stacking up more and more layers would result in vanishing gradient problem. Back propagation is done through identity function. And the identity matrix forwards the input data without loss of information.

Generally in convolution feature maps obtained represent the local

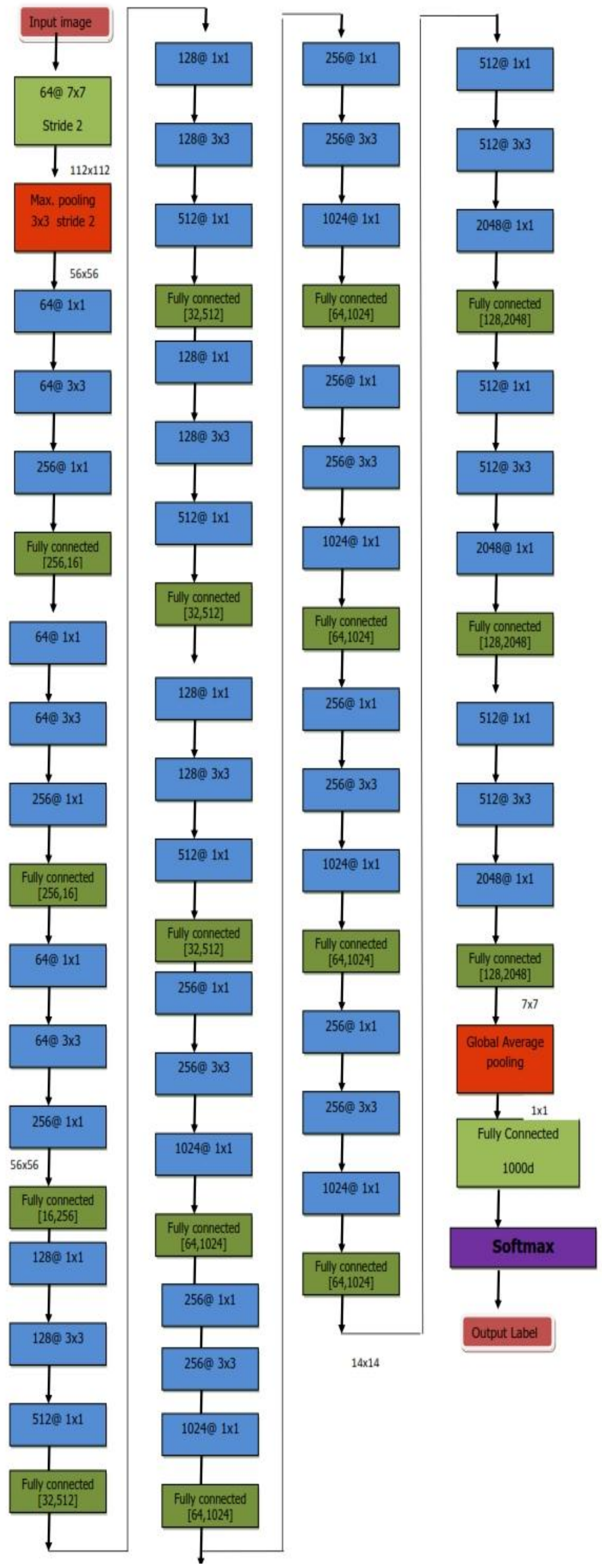

Fig. 4.Proposed Method

descriptor which will be index of local features. In order to capture the inter pixel dependencies more effectively we have included a global descriptor at the end of each convolution block. This global descriptor is obtained by global average pooling of the feature map. An effective steganalysis method have to characterize the inter pixel dependencies; the global descriptor does that work perfectly. The global descriptor indicates the weightage of that particular feature map. For non-lineaity we used ReLU activation function sandwiched between two fully connected layers. Then the smooth gating can be obtained by sigmoid function. This setup of global descriptor, fully connected layer, ReLU, fully connected layer followed by sigmoid acts as an booster to effectively capture the features that are sensitive to the hiding of data by steganography. We have included Batch Normalization and used dropout also for effectiveness of CNN.

VI. EXPERIMENTS & RESULTS

Tensorflow with Keras has been used on NVIDIA GPU machine. Learning rate of 0.001, momentum of 0.1, and weight decay of 0.0001 are used. SGD with mini batch size of 10 is considered. We have used the BOSSbase1.01version which consists of 10000 images. We have used crop function to generate 40000 cover images and 40000 stego images, each of size 256x256. For generating stego images we have used WOW, S-UNIWARD, HILL techniques. The detection error rate at 0.4bpp payload has been tabulated below in comparison with state of art SRM model and other deep learning based methods.

Table- I: Error in Detection rate of Proposed method

Method of Steganography	WOW	S-UNIWARD	HILL
Qians model	29.30%	30.90%	-
Xu's model	-	19.70%	20.70%
DRN model	4.30%	6.30%	10.40%
Proposed Method	2.80%	4.90%	7.50%

The results show that the proposed method has shown better performance than other contemporary techniques, due to the modifications made such as global descriptors and gating function.

REFERENCES

1. S. Tan and B. Li, “Stacked convolutional auto-encoders for steganalysis of digital images,” in Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific, pp. 1–4, IEEE, 2014.
2. J. Kodovsky`, J. J. Fridrich, and V. Holub, “Ensemble classifiers for steganalysis of digital media.,” IEEE Trans. Information Forensics and Security, vol. 7, no. 2, pp. 432–444, 2012.
3. Y. Qian, J. Dong, W. Wang, and T. Tan, “Deep learning for steganalysis via convolutional neural networks,” in Media Watermarking, Security, and Forensics 2015, vol. 9409, p. 94090J, International Society for Optics and Photonics, 2015.
4. G. Xu, H.-Z. Wu, and Y.-Q. Shi, “Structural design of convolutional neu- ral networks for steganalysis,” IEEE Signal Processing Letters, vol. 23, no. 5, pp. 708–712, 2016.
5. Y. Qian, J. Dong, W. Wang, and T. Tan, “Learning and transferring representations for image steganalysis using convolutional neural network,” in Image Processing (ICIP), 2016 IEEE International Conference on, pp. 2752–2756, IEEE, 2016.
6. J.-F. Couchot, R. Couturier, C. Guyeux, and M. Salomon, “Steganalysis via a convolutional neural network using large convolution filters for em- bedding process with same stego key,” arXiv preprint arXiv:1605.07946
7. Y. Qian, J. Dong, W. Wang, and T. Tan, “Feature learning for steganalysis using convolutional neural networks,” Multimedia Tools and Applications, pp. 1–25, 2017.
8. J. Ye, J. Ni, and Y. Yi, “Deep learning hierarchical representations for image steganalysis,” IEEE Transactions on Information Forensics and Security, vol. 12, no. 11, pp. 2545–2557, 2017.

9. J. Zeng, S. Tan, and B. Li, "Pre-training via fitting deep neural network to rich-model features extraction procedure and its effect on deep learning for steganalysis," in Proc. Media Watermarking, Security, and Forensics, Part of IS&T International Symposium on Electronic Imaging (EI'2017), 2017, pp. 44–49.
10. J. Zeng, S. Tan, B. Li, and J. Huang, "Large-scale JPEG steganalysis using hybrid deep-learning framework," IEEE Transactions on Information Forensics and Security, vol. 13, no. 5, pp. 1242–1257, 2018.
11. M. Chen, V. Sedighi, M. Boroumand, and J. Fridrich, "JPEG-phaseaware convolutional neural network for steganalysis of JPEG images," in Proc. 5th ACM Information Hiding and Multimedia Security Workshop (IH&MMSec'2017), 2017, pp. 75–84
12. G. Xu, "Deep convolutional neural network to detect J-UNIWARD," in Proc. 5th ACM Information Hiding and Multimedia Security Workshop (IH&MMSec'2017), 2017, pp. 67–73.
13. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR' 2016), 2016, pp. 770–778.
14. Y. Ma, X. Luo, X. Li, Z. Bao, Y. Zhang, Selection of Rich Model Steganalysis Features Based on Decision Rough Set a-Positive Region Reduction, IEEE Transactions on Circuits and Systems for Video Technology 29 (2) (2019) 336–350.doi:10.1109/TCSVT.2018.2799243.
15. M. Boroumand, M. Chen, J. Fridrich, Deep Residual Network for Steganalysis of Digital Images, IEEE Transactions on Information Forensics and Security 14 (5) (2019) 1181–1193. doi:10.1109/TIFS.2018.2871749.
16. M. Chen, M. Boroumand, J. Fridrich, Reference Channels for Steganalysis of Images with Convolutional Neural Networks, in: ACM Information Hiding and Multimedia Security Workshop, IH&MMSec '19, 2019, pp. 188–197. doi:10.1145/3335203.3335733.
17. J. Zeng, S. Tan, G. Liu, B. Li, J. Huang, WISERNet: Wider SeparateThen-Reunion Network for Steganalysis of Color Images, IEEE Transactions on Information Forensics and Security 14 (10) (2019) 2735–2748. doi:10.1109/TIFS.2019.2904413.
18. R. Zhang, F. Zhu, J. Liu, G. Liu, Depth-wise separable convolutions and multi-level pooling for an efficient spatial CNN-based steganalysis, IEEE Transactions on Information Forensics and Security (2019). doi:10.1109/TIFS.2019.2936913.
19. Y. LeCun, L. Bottou, Y. Bengio and P. Haffner: Gradient-Based Learning Applied to Document Recognition, Proceedings of the IEEE, 86(11):2278-2324, November 1998
20. Krizhevsky, A., Sutskever, I., and Hinton, G. E. ImageNet classification with deep convolutional neural networks. In NIPS, pp. 1106–1114, 2012.
21. Zeiler, M. D. and Fergus, R. Visualizing and understanding convolutional networks. CoRR, abs/1311.2901, 2013. Published in Proc. ECCV, 2014.
22. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1–9,2015.
23. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. CoRR. abs/1409.1556, (2014).
24. K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778, 2016.
25. S. Xie, R. Girshick, P. Dollar, Z. Tu and K. He. Aggregated Residual Transformations for Deep Neural Networks. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017

AUTHORS PROFILE

Mr. John Babu, is research scholar pursuing is Ph.D at Department of Computer Science & Engineering at JNTUH College of Engineering, Hyderabad, India. He is also working as an Assistant Professor at Sreekavitha Engineering College, Karepalli, Khammam, Telangana, India. His research interests include Steganography, Steganalysis & Deep Learning.

Dr. Sridevi Rangu is a Professor and Head of department of Computer Science & Engineering at JNTUH College of Engineering, Hyderbad, India. She has completed her Bachelors degree in CSE from Madras University, Masters from Andhra University and Ph.D from JNTUH. She is guiding many Postgraduate students and research scholars in the areas Network Security & Cryptography, Computer Networks,Data Structures