2018 18th IEEE International Conference on Communication Technology

Spatial Image Steganalysis Based on ResNeXt

Akash Sharma
Birla Institute of Technology & Science, Pilani
Delhi, India
e-mail: f2012771@pilani.bits-pilani.ac.in

Sunil Kumar Muttoo
Department of Computer Science
University of Delhi
Delhi, India
e-mail: skmuttoo@cs.du.ac.in

***Abstract*—With the success of Convolutional Neural Networks (CNN) in computer vision tasks, Steganalysis, the technique of detecting hidden secret messages within images, is moving away from Feature Engineering to Network Engineering. Deep neural networks are being proposed to model and capture the weak embedded signals, in such a low Signal-to-Noise (SNR) scenario. In this paper, we propose a novel Convolutional Neural Network based on aggregated residual transformations, which generate stronger image representations helpful for steganalysis. The architecture has very few hyper-parameters to set and focus on increasing the classification accuracy while keeping the depth and number of parameters fixed. The residual skip connections further help preserve the weak embedded signals and improve the gradient flow. We evaluated our proposed CNN on BOSSbase against S-UNIWARD and HILL steganographic algorithms with different payloads. Comparing with the state-of-the-art Deep Residual Learning (DRN) based on Residual Learning and the SRM plus Ensemble, our proposed CNN gives a better classification Accuracy.**

Keywords-steganalysis; Convolutional Neural Networks (CNN); deep learning

I. INTRODUCTION

Steganography is the technique of hiding data within digital covers such that no one understands its existence except the sender and the receiver. Its counter process, image steganalysis, tries to look for such hidden embedding in images, compressed or uncompressed and detects its presence. Usually formulated as a binary classification problem, steganalysis is a two-step process, that is, feature extraction and cover or stego classification. This feature extraction forms an integral part of image steganalysis and if hand-crafted manually, requires an in-depth knowledge of steganography algorithms.

The stego images have a low Signal-to-Noise-Ratio (SNR) making it difficult to detect such a weak signal. Also, in the spatial domain, some of the steganographic algorithms, such as S-UNIWARD (spatial uniward) [1], HILL [2] and Mipod [3] are adaptive to the image content and hide these low signals in complex textures. So, steganalysis methods have to be sophisticated enough to capture such issues. Current steganalysis techniques use high pass filters for preprocessing images, quantization and truncation followed by an Ensemble classifier [4-6].

With the success of deep learning algorithms in computer vision problems, different deep steganalysis methods are being proposed [7, 8-13]. Convolutional Neural network (CNN) automatically extracts the image features and provides a better alternative to handcrafting complex features. With this advent, steganalysis started transforming from Feature engineering to Network Engineering. A spatial Convolutional Neural Network was proposed by Qian [8] and Pibre [9] in 2015. In 2016, the classification accuracy was known to reach the state of the art by using an ensemble of CNN [10]. Later models started incorporating the high pass filtering from rich models, in CNN architectures [11]. Steganalysis models have also been proposed for JPEG domain. To incorporate JPEG-phase influence Chen et al [12] proposed architecture with a phase-split module. In [13] Zeng.et.al. proposed a hybrid architecture based on DCT filters.

Inspired from Residual Learning, Wu proposed a deep steganalysis residual learning based Network (DRN). The architecture is deep, and captures the statistical properties of the weak stego signal. Also, it makes use of shortcut connections connecting the input and output making it preserve the embedded signal. DRN achieved better classification accuracy as compared to the other proposed CNN models and the rich models.

Recently, Xie. Et al. [14] proposed a deep neural network, called the ResNeXts based on aggregated residual transformations for computer vision. ResNeXts are a clever combination of repeating block strategy of Resnets [15] and the split-transform-merge strategy of Inception [16,17,18]. The network takes care of the vanishing gradient descent problem in deep networks, using residual connections and has very few parameters to train. It also proposed a new hyper-parameter called Cardinality. Increasing Cardinality, has been more fruitful in increasing accuracy, than making the network deeper or wider.

In this work, we propose a novel CNN architecture inspired from the ResNeXts. Our architecture uses aggregated residual transformations for spatial image steganalysis. The skip connections enable the network to preserve weak stego signals, improving classification accuracy. Accuracy can be effectively improved by increasing the cardinality rather than making the network deeper or wider, hence keeping the number of parameters fixed.

978-1-5386-7635-6/18/$31.00 ©2018 IEEE

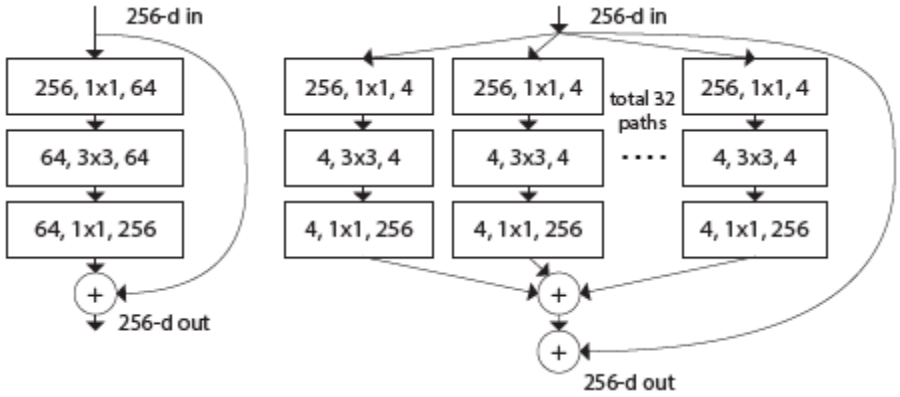

Figure 1: **Left:** Example building block of DRN **right**: Example Building block of proposed architecture with Aggregated Residual transformations, Cardinality =32 and skip connections between input and output. Each layer (#input channels, filter size, #output channels).

TABLE I. THE PROPOSED CNN ARCHITECTURE. BRACKETS REPRESENT THE SHAPE OF A RESIDUAL BLOCK (EQUIVALENT TO GROUP CONVOLUTIONAL FORM OF FIG. 1 (RIGHT)). A LAYER IS SHOWN AS (FILTER SIZE, #OUTPUT CHANNELS).

Group	Output size	Process	Times
Group 1	256 X 256	High PassFiltering	X 1
Group 2	128 X 128	7×7, 64, stride 2	X 1
Group 3	64 x 64	3×3, Max Pool stride 2	X 1
		(1X1, 128 3X3, 128,C=32 1X1, 256)	X 3
Group 4	32 x 32	(1X1, 256 3X3, 256,C=32 1X1, 512)	X 4
Group 5	16 x 16	(1X1, 512 3X3, 512,C=32 1X1, 1024)	X 6
Group 6	8 x 8	(1X1, 1024 3X3, 1024,C=32 1X1, 2048)	X 3
Group 7	1 x 1	Global Average Pool 1000-d Fully connected, Softmax	X 1

The rest of this paper is organized as follows. Details of the proposed CNN architecture are described in Section 2. Experiments and analysis are given in Section 3. The paper is finally concluded in Section 4.

II. PROPOSED NETWORK ARCHITECTURE

Table I describes the proposed architecture. It contains a pre-processing layer having a high pass filter, the residual learning layers focused on aggregated transformations and classification layers containing a fully connected layer followed by a Softmax, to get the stego or cover probability.

A. High Pass Filtering

In steganography, the embedded signal or noise is very weak compared to the image content, that is, the signal-to-noise-ratio is very low. Therefore, rather than feeding the original image, we extract the noise components of the image using a High Pass Filter (HPF), and then input it in the network. Previous studies [8-9] indicate that this pre-processing largely suppresses the image content, narrows the dynamic range, and thus increases the signal-to-noise ratio between the weak stego signal (if present) and the image signal. We follow the previous works, and use the KV kernel (Fig. 2) for pre-processing (Group 1).

B. Residual Learning Module

Our proposed architecture is inspired from the success of ResNeXt in computer vision domain. The basic building block of our architecture is the residual bottleneck block (Fig 1 right), which uses the split-transform-merge-strategy of Inception models. Each block splits the incoming vector into low dimensional embeddings, performs a set of transformations on each and finally aggregates the outputs by summation.

After preprocessing, the network (Group 2) generates many Activation maps using convolutional filters (64 filters of size 7 X 7). Group 3-6 corresponds to the residual module, where each group is a stack of the above described residual Block (Fig. 1 right). Each residual block has three convolution layers with two filter sizes 1X1 and 3X3. Because of the GPU memory, each convolution layer in the residual block is followed by batch normalization (BN) [19] and a RELU activation layer. All the transformations in our building block are of the same topology (Fig. 1) and the size of such transformations (called Cardinality) is 32. The bottlenet width of each transformation is 4d, making our architecture follow a 32X4d template (ResNeXt). The features extracted from the residual module feed the fully connected layer (Group 7) with 1000 neurons, which is further followed by a softmax activation function. The activation function gives probability distribution over the two class labels.

$$KV = \frac{1}{12}\begin{pmatrix} -1 & 2 & -2 & 2 & -1 \\ 2 & -6 & 8 & -6 & 2 \\ -2 & 8 & -12 & 8 & -2 \\ 2 & -6 & 8 & -6 & 2 \\ -1 & 2 & -2 & 2 & -1 \end{pmatrix}$$

Figure 2. KV Kernel

C. *Proposed CNN and DRN*

This section focuses on the similarities and differences between the proposed CNN and the DRN [7] architectures. Our proposed CNN outperforms the current best Residual Learning based steganalysis method DRN, with same number of parameters. Fig. 1 compares the building blocks of both the architectures. Ours is a multi-branch architecture incorporating the new concept of Cardinality, as an effective hyper-parameter. Following the results of ResNeXt architecture, the aggregated residual transformations used, prove to generate stronger representations as compared to the Residual learning based DRN.

Both the architectures use skip connections for preserving the weak stego signals, helpful for steganalysis classification. Enhancement of depth or width to improve the classification accuracy of a neural network, at times, reduces the model performance due to problems like vanishing gradient descent. Increasing the Cardinality, that is the number of transformations in a building block, while keeping the networks depth and width fixed, causes a surge in our networks classification accuracy. Only a few hyper-parameters need to be set, making the extension of this architecture fairly simple.

III. Experiments and Analysis

A. *Dataset and Software Platform*

We used the standard BOSSbase dataset 1.01 version [20], which is commonly used for steganography and steganalysis experiments. It contains 10,000 cover images (grayscale) of size 512 X 512. All the images are cropped to 256 X 256 using the Matlab imresize() function. In order to increase the training dataset size, data augmentation was applied by horizontally rotating the images by 90 degrees and mirroring them. Out of the 40,000 images obtained, 10000 are used for validation while 30000 images are used for training the model.

Due to computational constraints, the networks effectiveness was evaluated only on S-UNIWARD, a well known steganography method in the spatial domain. It uses distortion functions to hide the message. The online Matlab implementation with a random key for each embedding, rather than a fixed key, have been used to avoid use of any wrong code.

B. *Hyper-Parameters*

Mini batch Stochastic Gradient descent (SGD) was used to train our network. The size of mini batch was set to 10 while the momentum and weight decay was set to 0.9 and 0.001 respectively. Zero mean Gaussian distribution was used to initialize the convolutional kernels, with the standard deviation fixed at 0.01. Xavier initialization [21], that is, the weights follow a Gaussian distribution and are chosen so that the variance for both input and output among each layer remains the same, was used to initialize the fully connected layers. The bias was disabled.

C. *Classification Results*

We have compared our CNN with DRN (ResNet based steganalyser) and the conventional SRM plus Ensemble model [22]. Only S-UNIWARD adaptive steganographic algorithm was used with payloads 0.2 and 0.4. Both the proposed architecture and the Residual Network based steganalyser has almost equal number of parameters. To get the accurate results, all the steganalysis methods are tested on the same set of images. The error detection rate has been improved while keeping the number of parameters almost the same.

TABLE II. Steganalysis Error Probability Comparison for S-UNIWARD on BOSSBASE

Payload	Proposed	DRN	SRM + EC
0.4bpp	**5.98%**	6.71%	19.37%
0.2 bpp	**11.79%**	13.29%	27.81%

In Table II above, we report the error probability for S-UNIWARD (Spatial-UNIWARD) on our proposed CNN, the DRN and the feature engineering based SRM with Ensemble classifier. In Table III, we report the error probability for HILL on similar steganalysis models. Fig. 3 displays the error detection rates of our proposed CNN and DRN on BOSSBASE for S-UNIWARD at 0.4 bpp. For S-UNIWARD ,our proposed CNN has an error probability 13% lower than SRM + EC at 0.4 bpp, while it is around 16% at 0.2 bpp. For HILL algorithm, our proposed CNN has an error probability 12% lower than SRM + EC at 0.4 bpp, while it is around 15% at 0.2 bpp.

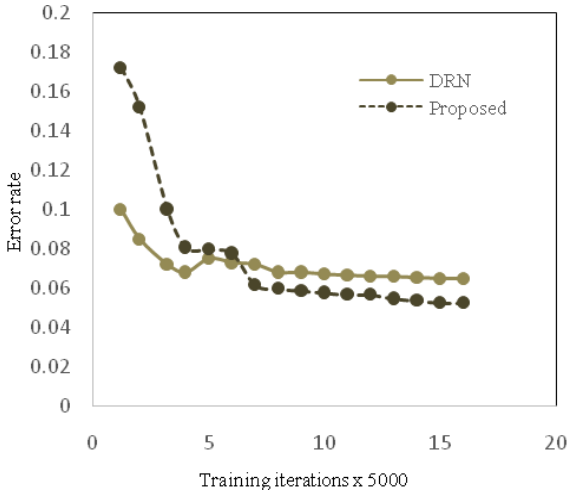

Figure 3. Detection error rates of proposed CNN and DRN on BOSSBASE for S-UNIWARD at 0.4 bit per pixel (bpp)

TABLE III. STEGANALYSIS ERROR PROBABILITY COMPARISON FOR HILL ON BOSSBASE

Payload	Proposed	DRN	SRM + EC
0.4bpp	**8.93%**	10.68%	22.01%
0.2 bpp	**16.24%**	18.60%	31.29%

IV. CONCLUSION

This paper proposes a new Convolutional Neural Network (CNN) for spatial image steganalysis. It is based on aggregated residual transformations which generate stronger image representations. The CNN focuses on increasing the size of transformations (Cardinality) to increase the classification accuracy rather than increasing the networks depth and width. There are very few hyper-parameters due to the same topology of all the transformations. Also, the shortcut connections inside a residual block, enables the preservation of weak embedded signals and improves the gradient flow. The proposed architecture is compared with the state of the art, resNet based DRN and SRM + Ensemble on BOSSbase dataset. With the same number of parameters, our proposed network produces better classification accuracy. Future work can be extending the above methodology in the JPEG domain.

REFERENCES

[1] Vojtech Holub, Jessica Fridrich and Tomas Denemark. Universal distortion function for steganography in an arbitrary domain. EURASIP Journal on Information Security, 2014(1):1,2014.

[2] Bin Li, Ming Wang, Jiwu Huang, and Xiaolong Li. A new cost function for spatial image steganography. In Image Processing (ICIP), 2014 IEEE International Conference on, pages 42064210. IEEE, 2014.

[3] Vahid Sedighi, Remi Cogranne, and Jessica Fridrich. Content-adaptive steganography by minimizing statistical detectability.IEEE Transactions on Information Forensics and Security, 11(2):221 234, 2016.

[4] Jessica Fridrich and Jan Kodovsky. Rich models for steganalysis of digital images. IEEE Transactions on Information Forensics and Security, 7(3):868882, 2012.

[5] Tomas Denemark, Vahid Sedighi, Vojtech Holub, Remi Cogranne, and Jessica Fridrich. Selection-channel-aware rich model for steganalysis of digital images. In In- formation Forensics and Security (WIFS), 2014 IEEE International Workshop on, pages 48 53.IEEE, 2014.

[6] Weixuan Tang, Haodong Li, Weiqi Luo, and Jiwu Huang. Adaptive steganalysis against wow embedding algorithm. In Proceedings of the 2nd ACM workshop on Information hiding and multimedia security, pages 9196. ACM, 2014.

[7] Songtao Wu, Shenghua Zhong, and Yan Liu. Deep residual learning for image steganalysis. Multimedia Tools and Applications, pages 117, dec 2017.

[8] Yinlong Qian, Jing Dong, Wei Wang, and Tieniu Tan, Deep Learning for Steganalysis via Convolutional Neural Networks, in Proceedings of Media Watermarking, Security, and Forensics 2015, MWSF2015, Part of IST/S IE Annual Symposium on Electronic Imaging, SPIE2015, San Francisco, California, USA, Feb. 2015,vol. 9409, pp. 94090J94090J10.

[9] L. Pibre, J. Pasquet, 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 Proceedings of Media Watermarking, Security, and Forensics, MWSF 2016, Part of IST International Symposium on Electronic Imaging, EI2016, San Francisco, California, USA, Feb. 2016, pp. 1-11.

[10] Guanshuo Xu,Han-Zhou Wu,and Yun Q. Shi, Ensemble of CNNs for Steganalysis: An Empirical Study, in Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security,Vigo,Galicia, Spain, June 2016, IHMMSec16, pp. 103107.

[11] Jishen Zeng, Shunquan Tan, Bin Li, and Jiwu Huang, Pre- training via fitting deep neural network to rich-model features extraction procedure and its effect on deep learning for steganalysis, in Proceedings of Media Watermarking, Security, and Forensics 2017, MWSF2017, Part of IST Symposium on Electronic Imaging, EI2017, Burlingame, California, USA, Jan. 2017, p. 6.

[12] Mo Chen, Vahid Sedighi, Mehdi Boroumand, and Jessica Fridrich. Jpeg-phase-aware convolutional neural network for steganalysis of jpeg images. In 5th ACM Workshop Inf. Hiding Multimedia Security. (IH- MMSec), 2017.

[13] Jishen Zeng, Shunquan Tan, Bin Li, and Jiwu Huang. Large-scale jpeg steganalysis using hybrid deep-learning framework. arXiv preprint arXiv:1611.03233, 2016.

[14] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He. Aggregated residual transformations for deep neural networks. In CVPR, 2017.

[15] He K, Zhang X, Ren S, Sun J (2015) Deep residual learning for image recognition. arXiv:1512.03385v1

[16] C. Szegedy, S. Ioffe, and V. Vanhoucke. Inception-v4, inception-resnet and the impact of residual connections on learning. In ICLR Workshop, 2016.

[17] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In CVPR, 2015.

[18] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In CVPR, 2016.

[19] S.Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In ICML,2015.

[20] Patrick Bas, Tomas Filler, and Tomas Pevn ‘y. break our steganographic system: The ins and outs of organizing boss. In Information Hiding, pages 5970. Springer, 2011.

[21] Xavier Glorot and Yoshua Bengio, Understanding the difficulty of training deep feedforward neural networks, in Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, AISTATS2010, Chia Laguna Resort,Sardinia, Italy, May 2010, vol. 9 of Proceedings of Machine Learning-Research,pp.249256.

[22] J. Fridrich and J. Kodovsky, Rich Models for Steganalysis of Digital Images, IEEE Transactions on Information Forensics and Security, TIFS, vol. 7, no. 3, pp. 868882, June 2012.