# TCNN: Two-Way Convolutional Neural Network for Image Steganalysis

Zhili Chen(✉), Baohua Yang, Fuhu Wu, Shuai Ren, and Hong Zhong

School of Computer Science and Technology, Anhui University, Hefei, China
zlchen@ahu.edu.cn, 779541664@qq.com, 12088@ahu.edu.cn, 154209401@qq.com, zhongh@mail.ustc.edu.cn

**Abstract.** Recently, convolutional neural network (CNN) based methods have achieved significantly better performance compared to conventional methods based on hand-crafted features for image steganalysis. However, as far as we know, existing CNN based methods extract features either with constrained (even fixed), or random (i.e., randomly initialized) convolutional kernels, and this leads to limitations as follows. First, it is unlikely to obtain optimal results for exclusive use of constrained kernels due to the constraints. Second, it becomes difficult to get optimal when using merely random kernels because of the large parameter space to learn. In this paper, to overcome these limitations, we propose a two-way convolutional neural network (TCNN) for image steganalysis, by combining both constrained and random convolutional kernels, and designing respective sub-networks. Intuitively, by complementing one another, the combination of these two kinds of kernels can enrich features extracted, ease network convergence, and thus provide better results. Experimental results show that the proposed TCNN steganalyzer is superior to the state-of-the-art CNN-based and hand-crafted features-based methods, at different payloads.

**Keywords:** Steganalysis · Two-way · Convolutional neural network

## 1 Introduction

Steganalysis is a kind of reverse analysis technology against steganography. Its purpose is to judge whether there is hidden information according to the extracted features of the image, and then distinguish the cover and stego.

Recently, deep convolutional neural network (CNN) has been increasingly applied to image steganalysis, and achieved better performance compared to conventional methods based on hand-crafted features. For instance, Xu *et al.* [12] proposed XuNet based on CNN, adding the absolute value (ABS) layer to narrow the range of feature map. In addition, TanH activation [4] was used in the front part of the network to improve the learning ability of features. Ye *et al.* [13] proposed a CNN which marks a significant breakthrough in the field of steganalysis. They initialized the first layer with the high-pass filter set in SRM,


N. Park et al. (Eds.): SecureComm 2020, LNICST 335, pp. 509–514, 2020.
https://doi.org/10.1007/978-3-030-63086-7_29

adopted a new activation function called truncated linear unit (TLU) and introduced the information of selection channel. This scheme greatly improved the detection performance and had obvious advantages over the traditional methods. Boroumand *et al.* [1] proposed a deep residual network called SRNet, which has made the latest achievements in image steganalysis.

As far as we know, existing CNN based steganalyzers extract features through either constrained convolutional kernels [11–13] or random ones [1]. In other words, they use one-way networks of either constrained or random convolutional kernels. The resulted limitations are as follows. First, if constrained convolutional kernels are used, it is unlikely that the optimal result is learnt due to the constraints enforced artificially. Second, if random convolutional kernels are applied, the parameter space to learn will become very large, and it is difficult to learn the optimal result without falling into sub-optimal ones.

In [8], the authors proposed a dual CNN for image steganalysis that consists of two parallel, identical sub-CNNs. Each sub-CNN applies Xu and Wu's design [12]. Two different forms of inputs are fed into the two sub-CNNs, respectively. The authors showed that different forms of inputs would improve the steganalysis performance. However, it is shown that the improvement is quite limited.

Different from the work [8], in this paper, to further improve the performance, we build a two-way network structure, each subnetwork of which is designed differently but complementarily. Specifically, we combine both constrained and random convolutional kernels into a two-way convolutional neural network (TCNN) for image steganalysis, expecting to extract more comprehensive features, and globally optimize the detection in a uniform network. We input images into two sub-networks for feature extraction, respectively. The first sub-network are initialized with all the 30 basic filters (convolutional kernels) used in the computation of residual maps in SRM [3], while the second sub-network are initialized with random filters with the same sizes. The two kinds of initialization are supposed to extract stego noise residual with both empirical convolutional kernels and learnt ones. Both sub-networks are then processed similarly, except with their respective appropriate pooling operations. Finally, the features extracted from both sub-networks are fused together, input to the classifier module, and the whole network is globally optimized. The proposed network combines the very best of constrained and random convolutional kernels, and its depth is shown to be shallow when getting a good result.

The rest of the paper is organized as follows. Section 2 describes the proposed network. Section 3 shows the experimental results and analysis. Finally, the concluding remarks of this paper and future works are given in Sect. 4.

## 2 The Proposed TCNN

### 2.1 Motivation

To illustrate the necessity of the two-way idea, we design 2 two-way networks, and investigate their performances as follows. The first network combines two subnetworks, each of which are initialized by the high-pass filters mentioned in SRM

**Table 1.** Performance comparison among one-way network, HPF+HPF and HPF+RND two-way networks in terms of detection error ($P_E$).

| Model | S-UNIWARD | | WOW | | HILL | |
|---|---|---|---|---|---|---|
| | 0.2 | 0.4 | 0.2 | 0.4 | 0.2 | 0.4 |
| One-way | 0.1670 | 0.0910 | 0.1368 | 0.0764 | 0.1918 | 0.1329 |
| HPF+HPF Two-way | 0.1985 | 0.0983 | 0.1271 | 0.0655 | 0.1805 | 0.1151 |
| HPF+RND Two-way | 0.1583 | 0.0805 | 0.1173 | 0.0585 | 0.1725 | 0.0959 |

[3]. The second one consists a sub-network initialized with the same high-pass filters, and the other sub-network initialized with random filters. Moreover, for each two-way network, the average pooling and maximum pooling algorithms are used in the two sub-networks, respectively. For convenience, we call the first two-way network as HPF+HPF two-way network, and the second one as HPF+RND two-way network.

Table 1 shows the performance comparison among HPF+HPF and HPF+RND two-way networks, and one way network, which is initialized with the SRM high-pass filters and applies average pooling operations, for steganography methods S-UNIWARD, WOW and HILL at payloads 0.2 and 0.4 bpp. From the first two rows, we can see that the performance of HPF+HPF two-way network is slightly better than that of one-way network for WOW and HILL, while slightly worse for S-UNIWARD, especially at low payloads. This indicates that the HPF+HPF two-way network perform comparatively with the one-way network, since many extended features in the HPF+HPF two-way network are repetitive or useless. Table 1 also demonstrates that the performance HPF+RND two-way network is obviously better than both one-way and HPF+HPF two-way networks. This shows that, by combining both constrained and random filters (convolutional kernels), the features extracted become more comprehensive, and the detection results are significantly improved.

### 2.2 TCNN Architecture

From the motivation above, our design takes the HPF+RND two-way structure. As shown in Fig. 1, the proposed TCNN network consists of two sub-networks (A/B), each of which is composed of pre-processing module and feature extraction module. The sub-network A is initialized with constrained convolutional kernels, while the sub-network B is initialized with random ones. Both sub-networks take as images as input, and share the same classification module. The structures of the two sub-networks are similar, but there are some differences detailed as follows.

In sub-network A, the first layer, also known as a pre-processing module, consists of 30 high-pass filters used in SRM [3] and an Absolute Value (ABS) layer. The ReLU function [9] serves as the activation function of the entire sub-network A. The feature extraction module consists of six layers, in which batch

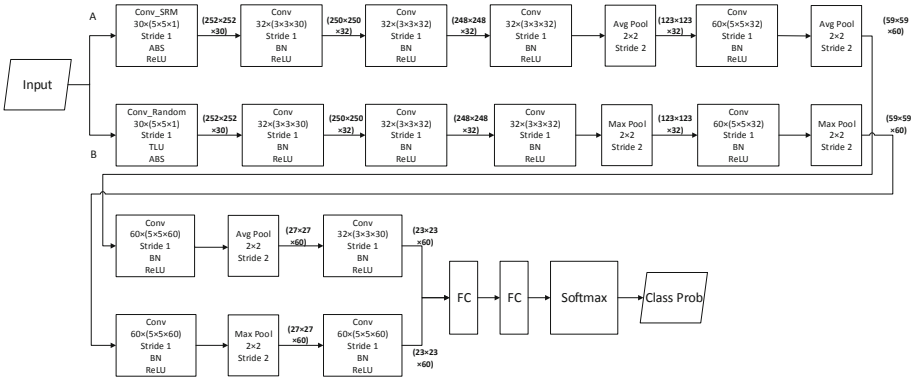


**Fig. 1.** The proposed TCNN architecture.

normalization [7] is added. And the pooling operations from 1st to 3rd layers are suppressed, the average pooling is set from the fourth layer. Unlike sub-network A, the weights of the first layer in sub-network B are initialized at random. Except that the activation function of the first layer is TLU [13], all other layers use the ReLU function. Moreover, we apply the maximum pooling algorithm in sub-network B. In particular, The ABS layer is following the TLU activation function in the pre-processing module. Finally, the outputs of the last layers of the feature extraction module in sub-network A and B are fused, and the fused features are transferred into the classification module. Two fully-connected layers and a softmax function map the vector to classification probability.

## 3 Experiments

### 3.1 Experiment Setup

We perform experiments using our TCNN to detect three spatial domain content-adaptive steganographic algorithms: S-UNIWARD [6], WOW [5] and HILL [10], with embedding rates from 0.1 bpp to 0.5 bpp. The dataset comes from BOSSbase v1.01, which contains 10000 original grayscale images of size $512 \times 512$. Constrained by our available GPU computing platform, all involved images are resized to the ones of size $256 \times 256$. To alleviate overfitting, data augmentation methods including random mirroring and 90 degrees rotation are used. For a given steganography and a payload, we have totally 20, 000 cover-stego image pairs. The training set includes 16,000 image pairs, which are randomly selected. The validation set includes 2000 image pairs. The test set includes the remaining 2000 image pairs.

### 3.2 Comparison with Other State-of-the-Art Methods

In this subsection, we compare the performance of the proposed TCNN model with three state-of-the-art steganalyzers in spatial domain, i.e., maxSRMd2 [2],

**Table 2.** Performance comparison of the steganalyzers in terms of detection error ($P_E$).

| Algorithm | Payload (bpp) | maxSRMd2 | SCA-YeNet | SCA-SRNet | The Proposed TCNN |
|---|---|---|---|---|---|
| S-UNIWARD | 0.1 | 0.3806 | 0.3220 | 0.2969 | **0.2613** |
| | 0.2 | 0.2999 | 0.2224 | 0.1918 | **0.1583** |
| | 0.3 | 0.2542 | 0.1502 | 0.1309 | **0.1098** |
| | 0.4 | 0.2136 | 0.1281 | 0.0935 | **0.0805** |
| | 0.5 | 0.1732 | 0.1000 | 0.0667 | **0.0630** |
| WOW | 0.1 | 0.3163 | 0.2442 | 0.2197 | **0.2139** |
| | 0.2 | 0.2325 | 0.1691 | 0.1401 | **0.1173** |
| | 0.3 | 0.1918 | 0.1229 | 0.0980 | **0.0812** |
| | 0.4 | 0.1536 | 0.0959 | 0.0769 | **0.0585** |
| | 0.5 | 0.1331 | 0.0906 | 0.0578 | **0.0514** |
| HILL | 0.1 | 0.3894 | 0.3380 | 0.3014 | **0.2716** |
| | 0.2 | 0.3226 | 0.2538 | 0.2159 | **0.1725** |
| | 0.3 | 0.2804 | 0.1949 | 0.1664 | **0.1160** |
| | 0.4 | 0.2410 | 0.1708 | 0.1290 | **0.0959** |
| | 0.5 | 0.2115 | 0.1305 | 0.1026 | **0.0645** |

SCA-YeNet [13] and SCA-SRNet [1]. Table 2 shows the performance comparison in terms of detection error ($P_E$) for all the tested schemes. We observe that TCNN model has obvious advantages over other steganalyzers for the involved embedding schemes and tested payloads. And in contrast to SCA-SRNet, the performance gap becomes most pronounced for HILL at 0.3bpp, where the detection error is decreased by 5%. The proposed model also has a better improvement at low payload. For instance, the detection errors of the proposed model for S-UNIWARD and WOW at 0.2 bpp are decreased by 3.3% and 2.3%, respectively. Moreover, the proposed TCNN has higher detection accuracy than other steganalysis methods at high payloads. For example, when the payload is 0.5 bpp, the detection accuracy for WOW has been improved to 94.8%. It is worth noting that the proposed TCNN is the most effective for HILL, in contrast to SCA-SRNet, the detection errors are reduced by more than 3%. The experimental results above demonstrate that the two-way network design combining both constrained and random convolutional kernels together with their respective subnetwork structures contributes to the detection performance improvement over other state-of-the-art steganalysis methods.

## 4 Conclusion

In this paper, we further improve the detection performance of the CNN based steganalysis methods by proposing a two-way convolutional network (TCNN) architecture. The TCNN is the first two-way CNN, which combines both man-made constrained convolutional kernels and freely learnt random ones to extract

stego noise residual signals to learn more complementary, comprehensive features. Furthermore, for different residual signals, different sub-network structures are designed to enhance the performance. Experimental results have shown that the proposed TCNN steganalyzer is superior to the state-of-the-art CNN-based and hand-crafted features-based methods, against steganography algorithms in spatial domain like S-UNIWARD, WOW and HILL. In future, the proposed network may be extended to multi-way networks or combined with selection-channel-aware methods to further improve performance.

**Acknowledge.** This work is supposed by the Special Fund for Key Program of Science and Technology of Anhui Province, China (Grant No. 18030901027).

## References

1. Boroumand, M., Chen, M., Fridrich, J.: Deep residual network for steganalysis of digital images. IEEE Trans. Inf. Forensics Secur. **14**(5), 1181–1193 (2018)
2. Denemark, T., Sedighi, V., Holub, V., Cogranne, R., Fridrich, J.: Selection-channel-aware rich model for steganalysis of digital images. In: 2014 IEEE International Workshop on Information Forensics and Security (WIFS), pp. 48–53. IEEE (2014)
3. Fridrich, J., Kodovsky, J.: Rich models for steganalysis of digital images. IEEE Trans. Inf. Forensics Secur. **7**(3), 868–882 (2012)
4. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, pp. 249–256 (2010)
5. Holub, V., Fridrich, J.: Designing steganographic distortion using directional filters. In: 2012 IEEE International Workshop on Information Forensics and Security (WIFS), pp. 234–239. IEEE (2012)
6. Holub, V., Fridrich, J., Denemark, T.: Universal distortion function for steganography in an arbitrary domain. EURASIP J. Inf. Secur. **2014**(1), 1–13 (2014). https://doi.org/10.1186/1687-417X-2014-1
7. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167 (2015)
8. Kim, J., Kang, S., Park, H., Park, J.I.: Dual convolutional neural network for image steganalysis. In: 2019 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), pp. 1–4. IEEE (2019)
9. Nair, V., Hinton, G.E.: Rectified linear units improve restricted Boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning (ICML 2010), pp. 807–814 (2010)
10. Pevný, T., Filler, T., Bas, P.: Using high-dimensional image models to perform highly undetectable steganography. In: Böhme, R., Fong, P.W.L., Safavi-Naini, R. (eds.) IH 2010. LNCS, vol. 6387, pp. 161–177. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-16435-4_13
11. Qian, Y., Dong, J., Wang, W., Tan, T.: 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)
12. Xu, G., Wu, H.Z., Shi, Y.Q.: Structural design of convolutional neural networks for steganalysis. IEEE Signal Process. Lett. **23**(5), 708–712 (2016)
13. Ye, J., Ni, J., Yi, Y.: Deep learning hierarchical representations for image steganalysis. IEEE Trans. Inf. Forensics Secur. **12**(11), 2545–2557 (2017)