Identification of Multiple Image Steganographic Methods Using Hierarchical ResNets

SUMMARY Image deformations caused by different steganographic methods are typically extremely small and highly similar, which makes their detection and identification to be a difficult task. Although recent steganalytic methods using deep learning have achieved high accuracy, they have been made to detect stego images to which specific steganographic methods have been applied. In this letter, a steganalytic method is proposed that uses hierarchical residual neural networks (ResNet), allowing detection (i.e. classification between stego and cover images) and identification of four spatial steganographic methods (i.e. LSB, PVD, WOW and S-UNIWARD). Experimental results show that using hierarchical ResNets achieves a classification rate of 79.71% in quinary classification, which is approximately 23% higher compared to using a plain convolutional neural network (CNN).

1. Introduction

Image steganography is a covert communication technique that conceals secret data by embedding them into digital images; note that the image before data embedding is called a cover image and the image after data embedding is called a stego image [3]. Image steganalysis is a technique for detecting stego images, usually by analysing visual and numerical changes due to the application of image steganography. Current statistical methods, such as RS-Analysis [12], can effectively detect stego images created by early steganographic methods, such as LSB [2] and PVD [14], which involve significant statistical changes. However, to detect stego images created by more sophisticated steganographic methods, such as WOW [6] and S-UNIWARD [7], that minimise statistical changes caused by data embedding, machine learning methods with well-designed filters, such as SPAM [8] and SRM [5], have been proposed. More recently, inspired by the success of AlexNet [10] in image classification, various steganalytic methods using convolutional neural networks (CNN) have been proposed [4], [11], [15]–[17]. CNNs can automatically learn the optimal features for classification through convolution operations, thus achieving much higher accuracy. However, conventional CNN-based steganalytic methods only focus on binary classification between cover and stego images. Thus, these methods cannot be used for multi-class classification, that is, classification of stego images created by different steganographic methods.

In this letter, we propose a steganalytic method that can detect and identify stego images created by multiple steganographic methods (i.e., separating between cover and stego images and identifying the steganographic method by which each stego image has been generated). Unlike conventional CNN-based methods, the proposed method is designed for multi-class classification. A similar method has been proposed for ternary classification of cover, WOW stego and S-UNIWARD stego images using a plain CNN [9]. However, as the number of classes increases, the classification accuracy of this previous method significantly decreases. In this letter, we propose to use hierarchical residual neural networks (ResNet) to classify five classes, i.e. cover, LSB stego, PVD stego, WOW stego and S-UNIWARD stego images, where we split classes into two-layered ResNets to be classified hierarchically. Using two ResNets can reduce the number of classes classified in each ResNet, which reduces the classification complexity. In addition, ResNets are much deeper than CNNs, thus providing higher classification accuracy [13].

2. Identification of Multiple Steganographic Methods Using a Plain CNN or ResNet

The simplest way to detect and classify multiple stego images is to increase the output number of steganalytic CNNs. Thus, we conducted an experiment using a CNN that is a variant of XuNet [15] (refer to Fig. 1 (a)), which has been previously proposed for ternary classification in [9], where the output of the fully-connected layer is extended to 5. As shown in Table 1, the quinary classification accuracy was very low and the PVD stego and S-UNIWARD stego images were not classified (over-fitting occurred and this is why the classification accuracy for WOW stego images is excessively high). This indicates a plain CNN with increased number of output classes cannot identify multiple steganographic methods.

We conducted another experiment using a deep ResNet (refer to Fig. 1 (b)), where the output of the fully-connected layer is extended to 5. The ResNet can prevent gradient loss
in the convolution layer by adding the input to the output of the convolution layer, allowing more convolution layers to be used. Table 1 shows the accuracy of quinary classification using the ResNet. The average classification rate was significantly increased compared to using the CNN; however, the WOW stego and S-UNIWARD stego images were not classified (WOW and S-UNIWARD stego images were misclassified to each other. WOW stego images were also frequently recognized as cover images.). This indicates that increasing the convolution layers is also not an effective solution for identifying multiple steganographic methods.

### 3. Proposed Method

To detect and classify multiple stego images, we can consider an approach to deploy the CNNs or ResNets in Fig. 1 hierarchically (refer to Fig. 2). Hierarchical CNNs or ResNets perform a three-step binary classification that classifies cover and stego images, classifies stego groups and then identifies each stego image; i.e. dividing quinary classification into multiple binary classifications.

However, hierarchical binary classification tends to accumulate classification errors from each layer, thus suffering from decreased accuracy due to a large number of layers (this will be confirmed later).

![Table 1](image)

|          | COVER | LSB  | PVD | WOW  | S-UNIWARD | Average |
|----------|-------|------|-----|------|------------|---------|
| Plain CNN| 88.45 | 64.80| 0.00| 99.87| 30.58      | 56.74   |
| Plain ResNet | 80.22 | 96.12| 99.80| 48.43| 47.76      | 74.47   |

Therefore, we propose to remove the layer that classifies cover and stego images and the layer that classifies the groups of stego images, which reduces the structure to two layers, as shown in Fig. 3. We attempt to classify relatively well-classified LSB stego and PVD stego images together with cover images in the first layer, while WOW stego and S-UNIWARD stego images, which are difficult to classify simultaneously, are classified hierarchically. That is, WOW and S-UNIWARD stego images are classified in the second layer after being recognised in the same class in the first layer.

Note that the CNN should be very deep to classify four classes in the first layer. Therefore, we adopt the ResNet in Fig. 1 (b), where the output number of the fully-connected layer is extended to 4 for quaternary classification. In addition, since it has been shown that ResNets have high accuracy for detecting WOW and S-UNIWARD stego images (95% and 94% for WOW and S-UNIWARD, respectively) [13], we also use the RestNet in Fig. 1 (b) in the second layer, instead of the CNN. Since the WOW stego and S-UNIWARD stego images are very similar, classifying them using the CNN in Fig. 1 (a) is not easy, which will be shown later.

Our ResNet (Fig. 1 (b)) is an extension of the CNN (Fig. 1 (a)) using the residual blocks (Fig. 1 (c), introduced in [13]). Our ResNet has a different configuration from the previous study [13]. In the previous study [13], pairs of cover and stego images were required as input for training and testing the network, which is unusual and impractical. However, our ResNet uses a single input image as in common image steganalytic methods.

### 4. Experimental Results and Discussion

For all the experiments, 10,000 grayscale images of 512 × 512 resolution of BossBase1.01 [1] were used. Each image was partitioned into non-overlapping segments to produce a
total of 40,000 256 × 256 images, and then they were separated into 30,000 for learning and 10,000 for verification. Using the images for learning and verification, we created stego images by applying each steganographic method applied. For LSB and PVD, we used our code written in C++. For WOW and S-UNIWARD, we used the C++ and MATLAB codes provided in [18] and set the payload to 0.4. The Pytorch library [19] was used to implement the neural networks, and each network was trained 200,000 times using a momentum optimiser with a momentum value of 0.9 using cross entropy as the cost function. The mini-batch size was 60 for CNNs and 30 for ResNets.

Table 2 shows the accuracy of binary classification of each CNN in Fig. 2. The classification rates were 82.29%, 98.03%, 99.85% and 68.47% between cover and stego images, between stego groups, between LSB stego and PVD stego images and between WOW stego and S-UNIWARD stego images, respectively. Unlike the others, binary classification between WOW stego and S-UNIWARD stego images had low accuracy due to the high similarity between the two steganographic methods. Table 3 shows the results of quinary classification through hierarchical binary classification. Compared to the results of using the plain CNN in Table 1, the classification accuracy of cover and WOW stego images decreased but the classification accuracy of LSB stego, PVD stego and S-UNIWARD stego images significantly increased. The over-fitting did not occur. As a result, the average classification rate increased by 15.55%. However, the classification rate of WOW stego images was very low due to the poor performance of the CNN in classifying between WOW stego and S-UNIWARD stego images and the accumulation of classification errors caused by the hierarchical classification. Consequently, the average classification rate was lower than that for the plain ResNet in Table 1.

Table 4 shows the results of quinary classification using the proposed two-layered ResNets. By reducing the number of layers, the classification rates of all stego images increased. Moreover, the classification rates of WOW stego and S-UNIWARD stego images further increased by using the ResNet instead of the CNN. From the results in Tables 2 and 4, we found that the ResNet had 8.89% higher classification rates than the CNN in classifying between WOW stego and S-UNIWARD stego images. As a result, the average classification rate increased by 7.42% and 22.97% compared to those of the hierarchical CNNs and the plain CNN, respectively.

On the other hand, the classification rate of cover images resulting from the quaternary classification in Table 4 decreased by 13.04% compared to the binary classification in Table 3. However, since image steganalysis is intended for the detection of hidden secret data, it should be focused more on accurately classifying stego images rather than the cover image.

5. Conclusion

This letter proposed a steganalytic method that can identify multiple steganographic methods by using hierarchical deep ResNets. Four different types of stego images (created by LSB, PVD, WOW and S-UNIWARD) could be detected and were classified with an accuracy of 79.71%. The accuracy was higher by 22.97%, 5.24% and 7.42% compared to those of a plain CNN, a plain ResNet and hierarchical CNNs, respectively.

To make the proposed method more practical, the proposed method will be improved so that the stego images generated by more steganographic methods can be detected and identified.

References

[1] P. Bas, T. Filler, and T. Pevný, “‘Break our steganographic system’: the ins and outs of organizing BOSS,” Lecture Notes in Computer Science, vol.6958, pp.59–70, 2011.
[2] C.-K. Chan and L.M. Cheng, “Hiding data in images by simple LSB substitution,” Pattern Recognition, vol.37, no.3, pp.469–474, 2004.
[3] A. Cheddad, J. Condell, K. Curran, and P.M. Kevitt, “Digital image steganography: survey and analysis of current methods,” Signal Processing, vol.90, no.3, pp.727–752, 2010.
[4] X. Deng, B. Chen, W. Luo, and D. Luo, “Fast and effective global covariance pooling network for image steganalysis,” Proc. ACM Workshop on Information Hiding and Multimedia Security, pp.230–234, 2019.
[5] J. Fridrich and J. Kodovský, “Rich models for steganalysis of digital images,” IEEE Transactions on Information Forensics and Security, vol.7, no.3, pp.868–882, 2012.
[6] V. Holub and J. Fridrich, “Designing steganographic distortion using directional filters,” Proc. IEEE International Workshop on Information Forensics and Security, pp.234–239, 2012.
[7] V. Holub, J. Fridrich, and T. Denemark, “Universal distortion function for steganography in an arbitrary domain,” EURASIP J. Info.
[8] Y. Hou, R. Ni, and Y. Zhao, “Steganalysis to adaptive pixel pair matching using two-group subtraction pixel adjacency model of covers,” Proc. International Conference on Signal Processing, pp.1864–1867, 2014.

[9] S. Kang, H. Park, and J.-I. Park, “CNN-based ternary classification for image steganalysis,” Electronics, vol.8, no.11, 1225, 2019.

[10] A. Krizhevsky, I. Sutskever, and G.E. Hinton, “ImageNet classification with deep convolutional neural networks,” Communications of the ACM, vol.60, no.6, pp.84–90, 2017.

[11] B. Li, W. Wei, A. Ferreira, and S. Tan, “ReST-Net: diverse activation modules and parallel subnets-based CNN for spatial image steganalysis,” IEEE Signal Processing Letters, vol.25, no.5, pp.650–654, 2018.

[12] S. Manoharan, “An empirical analysis of RS steganalysis,” Proc. Third International Conference on Internet Monitoring and Protection, pp.172–177, 2008.

[13] S. Wu, S. Zhong, and Y. Liu, “Deep residual learning for image steganalysis,” Multimedia Tools and Applications, vol.77, no.9, pp.10437–10453, 2018.

[14] D.-C. Wu and W.-H. Tsai, “A steganographic method for images by pixel-value differencing,” Pattern Recognition, vol.24, no.9-10, pp.1613–1626, 2003.

[15] 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, pp.708–712, 2016.

[16] 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.

[17] M. Yedroudj, F. Comby, and M. Chaumont, “Yedroudj-Net: an efficient CNN for spatial steganalysis,” Proc. ICASSP, pp.15–20, 2018.

[18] http://dde.binghamton.edu/download/feature_extractors/

[19] http://pytorch.org