RESIDUAL CONVOLUTION NETWORK BASED STEGANALYSIS WITH ADAPTIVE CONTENT SUPPRESSION

Songtao Wu[1,2] *Sheng-hua Zhong*[1,*] *Yan Liu*[2]

[1] College of Computer Science and Software Engineering, Shenzhen University
[2] Department of Computing, The Hong Kong Polytechnic University
Email: csstwu@szu.edu.cn, csshzhong@szu.edu.cn, csyliu@comp.polyu.edu.hk

ABSTRACT

Image steganalysis is to discriminate innocent images and those suspected images with hidden messages. In this paper, we propose a unified Convolutional Neural Network (CNN) model for this task. In order to reliably detect modern steganographic algorithms, we design the proposed model from two aspects. For the first, different from existing CNN based steganalytic algorithms that use a predefined high-pass kernel to suppress image content, we integrate the high-pass filtering operation into the proposed network by building a content suppression subnetwork. For the second, we propose a novel sub-network to actively preserve the weak stego signal generated by secret messages based on residual learning, making the successive network capture the difference between cover images and stego images. Extensive experiments demonstrate that the proposed model can detect states-of-the-art steganography with much lower detection error rates than previous methods.

Index Terms— Image steganalysis, convolutional neural network, residual learning, adaptive content suppression

1. INTRODUCTION

Steganography is the technique to hide secret messages into multimedia signals such as audio, image or video, etc [1]. Steganalysis, from an opponent's perspective, is the art of revealing the presence of secret messages embedded in digital media [2]. Among all steganalytic techniques, image steganalysis plays an important role in many security systems and attracts increasing interests in recent years [3].

Designing effective features that are sensitive to message embedding is key to image steganalysis. Traditional methods use handcrafted features to detect steganography [4-6]. However, the feature design is a difficult task which needs strong domain knowledge about steganography and steganalysis. Recently, deep neural networks have attracted increasing interests in many image related tasks [7-8, 15]. Based on deep CNN models, several interesting works have been proposed to detect steganography. Compared with traditional methods that extract handcrafted features, CNN based steganalysis directly learns effective features using various network architectures for discriminating cover images and stego images. Tan and Li [9] first proposed to detect the presence of secret messages based on a deep stacked convolutional auto-encoder network. In [10], Qian *et al.* proposed a model for steganalysis using the standard CNN architecture with Gaussian activation function. Xu *et al.* in [11] designed a new CNN structure with *tanh* activation function and absolute operation after the first convolutional layer. Pibre *et al.* in [12] proposed a novel CNN model for image steganalysis, which obtains a low detection error in the scenario that the message is hidden with a same embedding key. Wu *et al.* in [13-14] proposed a novel CNN model for image steganalysis, achieving much better performances than previous methods.

However, two problems remain unsolved for the existing CNN based steganalytic methods. Firstly, they only use a fixed highpass kernel to preprocess the input image for content suppression, which may constrain useful information that contains the difference between cover images and stego images to come into the following network. Secondly, there are no effective learning methods to preserve the weak stego signal generated by secret messages, making the network hard to learn useful features to discriminate covers and stegos.

To address these difficulties, this paper proposes a unified CNN model for steganalysis. On one hand, unlike previous methods that separately preprocess the input image and extract features for classification, we integrate the highpass filtering operation into the proposed network by building a content suppression subnetwork. The highpass kernels in the sub-network is adaptively updated in the network training, allowing more powerful discriminative features come into the subsequent network than that of CNN models with a prede-

This work was supported by the National Natural Science Foundation of China (No. 61502311, No. 61373122), the Natural Science Foundation of Guangdong Province (No. 2016A030310053), the Science and Technology Innovation Commission of Shenzhen under Grant (No. JCYJ20150324141711640), the Special Program for Applied Research on Super Computation of the NSFC-Guangdong Joint Fund (the second phase), the Shenzhen high-level overseas talents program, the Shenzhen University research funding (201535), and the Tencent Rhinoceros Birds - Scientific Research Foundation for Young Teachers of Shenzhen University.

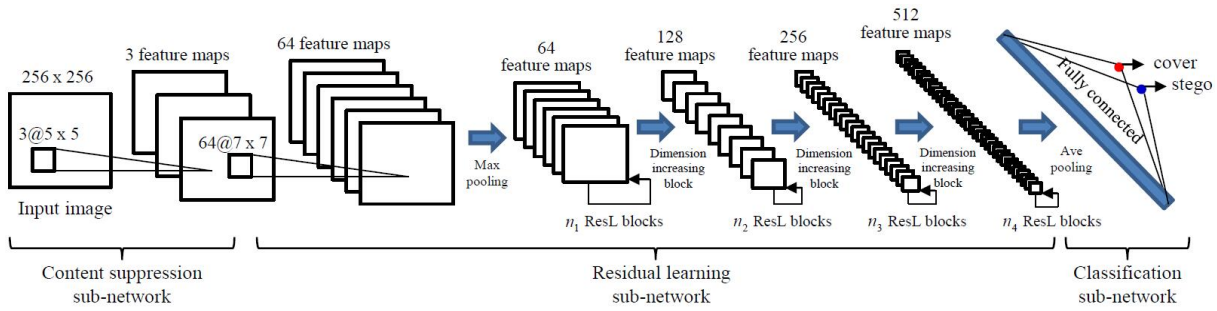

Fig. 1. The proposed network for image steganalysis. The network contains the content suppression sub-network, the residual learning sub-network and the classification sub-network, which are used for the noise component extraction, feature learning and label mapping respectively. In this figure, $p@q \times q$ denotes p convolutional kernels with the size of $q \times q$.

fined kernel. To the best of our knowledge, this is the first CNN model that unifies image preprocessing and feature learning in a whole network for steganalysis. On the other hand, we propose a novel learning scheme to actively preserve the weak stego signal generated by secret message by incorporating residual learning [15] in our network. This learning scheme has demonstrated superior performance than previous CNN based steganalytic methods. In theory, we have proved that shortcut connections in residual learning can effectively preserve the weak stego signal for the network.

In the rest of this paper, we first introduce the proposed network model in section 2. In section 3, we provide a theoretical analysis to explain the rationality of residual learning for image steganalysis. In section 4, we validate the effectiveness of the proposed model on several states of the art steganographic algorithms. The paper is closed with the conclusion in section 5.

2. PROPOSED NETWORK MODEL

2.1. Network Architecture

Fig.1 shows the overall architecture of the proposed network model. The network contains three sub-networks, which are introduced in the following parts.

The content suppression sub-network is to extract the noise component [1] of input cover/stego images. Three 5×5 kernels are used to convolve the input image. To pledge that the sub-network is to extract the noise components, each of three kernels is initialized by a highpass kernel, i.e. the KV kernel [10]. These kernels are updated in model training.

The residual learning sub-network is to extract effective features for steganalysis. The concept of residual learning was originally proposed by He *et al.* in [15]. We follow the idea of residual learning to design our network. In the residual learning sub-network, 64 filters with the size of 7×7 are used to convolve the noise components generated by the content

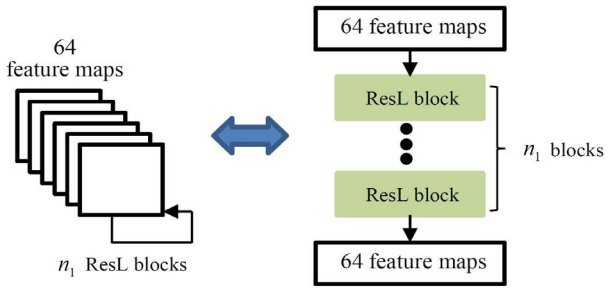

Fig. 2. Feature maps followed by several ResL blocks.

suppression sub-network. Following the convolutional layer is a batch normalization layer, a ReLU activation layer and a max pooling layer[2]. Then, the network uses two kinds of blocks to process the data: the residual learning (ResL) block and the dimension increasing block. A ResL block consists of two convolutional layers, each of which is also followed by a batch normalization layer and a ReLU layer. The size of convolutional kernels in the block is 3×3 and the number of kernels is equal to the number of input feature map (details about the block are described in [15]). A shortcut path connects the input and the output of the block, acting as the identical mapping. For a dimension increasing block, the only difference to a ResL block is that the number of feature maps is doubled and each feature map is down-sampled for the output. In general, there are several ResL blocks before a dimensional increasing block in the residual learning sub-network. We use Fig.2 to represent feature maps followed by several ResL blocks to make the figure of overall network compact. For economical considerations, a bottleneck version for residual learning and dimension increasing is developed for very deep networks. Different from the non-bottleneck version with two convolutional layers, a bottleneck version has three convolutional layers [15].

The classification sub-network maps extracted features into binary labels. Two output nodes are fully connected to fea-

[1]The noise component of an image denotes the image filtered by a high-pass filter. This is a general preprocessing for image steganalysis.

[2]The batch normalization and ReLU are not shown in the figure.

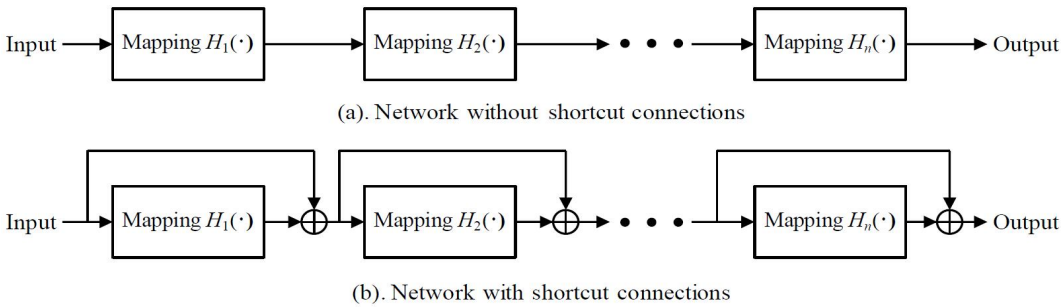

Fig. 3. (a). A typical CNN model can be abstracted as a cascaded network. (b). A residual learning network can be abstracted as a cascaded network, where each building block has a shortcut connecting its input and output.

ture maps that are pooled in residual learning sub-network.

2.2. Network Training

Parameters of the proposed network are learned by minimizing the softmax loss function:

$$L(\mathbf{x}_i, \theta) = -\sum_{k=1}^{K} 1\{y_i = k\} \cdot log\left(\frac{e^{o_{i,k}(\mathbf{x}_i,\theta)}}{\sum_{k=1}^{K} e^{o_{i,k}(\mathbf{x}_i,\theta)}}\right) \quad (1)$$

where θ denotes the parameters of the network, including weight matrices $\mathbf{W}$ and the bias vectors $\mathbf{b}$. K is the number of labels, where $K = 2$ in our model. y_i is the label of $\mathbf{x}_i$, $1\{\cdot\}$ is the indicator function. $o_{i,k}(\mathbf{x}_i, \theta)$ represents the output of the network for the sample $\mathbf{x}_i$. $\mathbf{W}$ and $\mathbf{b}$ of the network parameter θ are updated by the mini-batch stochastic gradient descending (SGD):

$$\mathbf{W}(t+1) = \mathbf{W}(t) - \alpha\frac{1}{N}\sum_{i\in B}\frac{\partial L(\mathbf{x}_i,\theta)}{\partial \mathbf{W}} \quad (2)$$

$$\mathbf{b}(t+1) = \mathbf{b}(t) - \alpha\frac{1}{N}\sum_{i\in B}\frac{\partial L(\mathbf{x}_i,\theta)}{\partial \mathbf{b}} \quad (3)$$

where N is the size of a mini-batch B, α is the learning rate.

3. RATIONALITY OF RESIDUAL LEARNING FOR STEGANALYSIS

He *et al.* in [15] presented a deep residual network for large scale image classification. In this work, a new learning method called the residual learning has been proposed. To approximate an underlying function $H(\mathbf{x})$, residual learning does not fit it directly but turns to fit its residual mapping: $F(\mathbf{x}) := H(\mathbf{x}) - \mathbf{x}$. As indicated in his paper, it is easier to optimize the residual mapping $F(\mathbf{x})$ than the unreferenced mapping $H(\mathbf{x})$, especially when $H(\mathbf{x})$ is an identical mapping or a near identical mapping.

For steganalysis, we believe that residual learning can preserve the weak stego signal. Actually, the task of steganalysis is to classify an input image as a cover or a stego:

$$\mathbf{y} = \begin{cases} \mathbf{x} + \mathbf{0}, \text{ cover} \\ \mathbf{x} + \mathbf{s}, \text{ stego} \end{cases} \quad (4)$$

where $\mathbf{x}$ represents the innocent cover image, $\mathbf{0}$ is zero vector and $\mathbf{s}$ denotes the weak stego signal. By feeding $\mathbf{y}$ into a residual learning network, the identity mapping of the network puts forward $\mathbf{x}$ to the output of the block, while the residual mapping $F(\cdot)$ fits $\mathbf{0}$ or $\mathbf{s}$. Since both $\mathbf{0}$ and $\mathbf{s}$ are small signals, they can be effectively modeled by the residual mapping $F(\cdot)$. Consequently, $\mathbf{s}$ can be effectively captured by the residual network. Therefore, the stego signal is expected to be preserved.

In the following part, we provide a mathematical analysis to prove that weak stego signal can be effectively preserved in a CNN model with residual learning.

3.1. Feature Diminishing in a Typical CNN Model

For steganalysis, the *feature diminishing* becomes a major problem. This indicates that the weak stego signal added to a cover image would be attenuated as it travels the whole CNN model, making the later network hardly capture effective features to discriminate cover images and stego images. To illustrate this phenomenon, we perform a mathematical derivation as follows. Assume we have a cover image $\mathbf{x}$ and its stego version $\mathbf{y}$, where the stego image $\mathbf{y}$ can be represented as:

$$\mathbf{y} = \mathbf{x} + \mathbf{s} \quad (5)$$

Generally, a CNN can be abstracted as a typical model of Fig.3(a). In this abstracted form, the mapping $H_i(\mathbf{x})$ ($i = 1, 2, \cdots, n$) could be a convolutional layer, a nonlinear activation layer, a pooling layer or their combinations. By feeding $\mathbf{x}$ and $\mathbf{y}$ into a typical CNN model, we obtain their outputs:

$$Z_t^n(\mathbf{x}) = H_n(\cdots H_2(H_1(\mathbf{x}))) \quad (6)$$

$$Z_t^n(\mathbf{y}) = H_n(\cdots H_2(H_1(\mathbf{y}))) \quad (7)$$

Compared with the cover image $\mathbf{x}$, the stego signal $\mathbf{s}$ is often very weak. Therefore, we iteratively use Taylor expansion for $Z_t^n(\mathbf{y})$ and get the following result:

$$Z_t^n(\mathbf{y}) = Z_t^n(\mathbf{x}) + \left(\prod_{i=1}^{n} F^{(i)}(\mathbf{x})\right)\mathbf{s} + O(||\mathbf{s}||^2) \quad (8)$$

where $O(||\mathbf{s}||^2)$ is the expansion remainder, $F^{(i)}(\mathbf{x})$ is:

$$F^{(i)}(\mathbf{x}) = \begin{cases} H_i'(\mathbf{x}), & i = 1 \\ H_i'(\cdots H_1(\mathbf{x})), & i < 1 \leq n \end{cases} \quad (9)$$

In above equation, $H_i'(\mathbf{x})$ denotes the derivative of the mapping $H_i(\mathbf{x})$:

$$H_i'(\mathbf{x}) = \frac{\partial H_i(\mathbf{x})}{\partial \mathbf{x}} \quad (10)$$

In a CNN model, each element f_i in the derivative matrix $F^{(i)}(\mathbf{x})$ satisfies the following inequality:

$$|f_i| \leq 1 \quad (11)$$

We explain this result for each basic operation in a typical CNN model:

- For a convolutional layer, f_i actually relates to the sum of image pixels multiplied by weights in a convolutional kernel. To ensure the stability of learning CNN models, existing methods initialize convolutional kernels with small weights (e.g. Gaussian random values with zero mean and 0.01 standard derivation) and set learning rate for parameter updating to a small value. These settings ensure that elements in convolutional layers are small during the learning phase. In addition, the size of a convolutional kernel in CNN models is often small. Consequently, f_i of a convolutional layer is small and Eq.(11) can be satisfied in most cases;

- For a nonlinear activation layer, f_i in $F^{(i)}(\mathbf{x})$ is ensured to be smaller than 1. This is because for existing activation functions, e.g. the sigmoid, tanh or ReLU, their slopes are smaller than or equal to 1 anywhere.

- For a pooling layer, either the average pooling or the maximum pooling does not increase the absolute value of each element in a feature map $H_i(\mathbf{x})$, thus Eq.(11) is satisfied.

With the property as the Eq.(11), $\prod_{i=1}^{n} F^{(i)}(\mathbf{x})$ decays exponentially as n increases. This will make the difference between $\mathbf{x}$ and $\mathbf{y}$ very small for large n. Under this case, the later layers in a CNN model can hardly discriminate cover images and stego images.

3.2. Preserve Weak Stego Signal with Residual Learning

For image steganalysis, a network with shortcut connections (the model as Fig.3(b) shows) can effectively overcome the feature diminishing phenomenon. Same to the analysis as Fig.3(a), we feed the cover image $\mathbf{x}$ and its stego image $\mathbf{y}$ into the network as Fig.3(b) and obtain their outputs:

$$Z_s^n(\mathbf{x}) = R_n(\cdots R_2(R_1(\mathbf{x}))) \quad (12)$$

$$Z_s^n(\mathbf{y}) = R_n(\cdots R_2(R_1(\mathbf{y}))) \quad (13)$$

where $R_i(\mathbf{x})$ denotes:

$$R_i(\mathbf{x}) = H_i(\mathbf{x}) + \mathbf{x},\ 1 \leq i \leq n \quad (14)$$

Similarly, we perform the Taylor expansion for Eq.(13):

$$Z_s^n(\mathbf{y}) = Z_s^n(\mathbf{x}) + \left[\prod_{i=1}^{n}\left[1 + H_i'(F_R^{(i)}(\mathbf{x}))\right]\right]\mathbf{s} + O(||\mathbf{s}||^2) \quad (15)$$

where $F_R^{(i)}(\mathbf{x})$ is:

$$F_R^{(i)}(\mathbf{x}) = \begin{cases} \mathbf{x}, & i = 1 \\ R_{i-1}(\mathbf{x}) & i = 2 \\ R_{i-1}(\cdots R_1(\mathbf{x})), & i > 2 \end{cases} \quad (16)$$

Unlike the case as Fig.3(a), the coefficient matrix of the stego signal $\mathbf{s}$, $\prod_{i=1}^{n}\left[1 + H_i'(F_R^{(i)}(\mathbf{x}))\right]$, does not exponentially decay as the depth increases. To better understand the advantage of the network with shortcut connections, we factorize $Z_s^n(\mathbf{x})$ and $Z_s^n(\mathbf{y})$ when n is 2. For $Z_s^n(\mathbf{x})$, we have:

$$Z_s^2(\mathbf{x}) = R_2(R_1(\mathbf{x})) = \mathbf{x} + H_1(\mathbf{x}) + H_2(H_1(\mathbf{x}) + \mathbf{x}) \quad (17)$$

For $Z_s^n(\mathbf{y})$, we iteratively use Taylor expansion and obtain:

$$\begin{aligned} Z_s^2(\mathbf{y}) &= R_2(R_1(\mathbf{y})) \\ &= R_2(R_1(\mathbf{x})) + [R_2'(R_1(\mathbf{x})) \cdot R_1'(\mathbf{x})]\,\mathbf{s} + O(||\mathbf{s}||^2) \\ &= \mathbf{x} + H_1(\mathbf{x}) + H_2(H_1(\mathbf{x}) + \mathbf{x}) + \mathbf{s} + [H_1'(\mathbf{x})]\,\mathbf{s} \\ &\quad + [H_2'(H_1(\mathbf{x}) + \mathbf{x})]\,\mathbf{s} + [H_2'(H_1(\mathbf{x}) + \mathbf{x}) \cdot H_1'(\mathbf{x})]\,\mathbf{s} \\ &\quad + O(||\mathbf{s}||^2) \end{aligned} \quad (18)$$

We compared the factorization Eq.(18) with Eq.(8), and find two advantages of the network with shortcut connections:

- The coefficient matrix for the stego signal $\mathbf{s}$ does not decay as the network's depth increases. Unlike the Eq.(8), the coefficient matrices for $\mathbf{s}$ in Eq.(18), i.e. the coefficient matrices for $\mathbf{s}$, $[H_1'(\mathbf{x})]\,\mathbf{s}$, $[H_2'(H_1(\mathbf{x}) + \mathbf{x})]\,\mathbf{s}$, are always kept to be no larger than order one, independent of network's depth. This property ensures that the stego signal $\mathbf{s}$ does not decays as the network's depth increases;

- The Signal-to-Noise Ratio (SNR) between the stego signal and the image content does not decay as the network's depth increases. By checking Eq.(18), we find that each term of $\mathbf{x}$ is always accompanied with a corresponding term of $\mathbf{s}$, i.e. $\mathbf{s}$ for $\mathbf{x}$, $[H_1'(\mathbf{x})]\,\mathbf{s}$ for $H_1(\mathbf{x})$, and $[H_2'(H_1(\mathbf{x})+\mathbf{x})]\,\mathbf{s}$ for $H_2(H_1(\mathbf{x})+\mathbf{x})$. For a well learned network, this property ensures a non-decaying SNR between the stego signal $\mathbf{s}$ and the image content $\mathbf{x}$ even the depth n is large.

4. EXPERIMENTS

This section is to validate the effectiveness of the proposed model. The dataset used for validation is the BOSSbase 1.01 [16], which is a standard database for evaluating steganography and steganalysis. The original BOSSbase contain 10,000 natural images with the size of 512×512. In our experiments, each image in the dataset is cropped into 4 non-overlapping 256×256 cropped images. Therefore, we have a cropped BOSSbase with 40,000 images.

For the proposed model, in the residual learning sub-network and the classification sub-network, weight matrices $\mathbf{W}$ are initialized by a zero mean Gaussian distribution with the standard derivation 0.01 and biases vectors $\mathbf{b}$ are initialized to zeros. The momentum and the weight decay in two sub-networks are set to 0.9 and 0.0001 respectively. The learning rate α for $\mathbf{W}$ and $\mathbf{b}$ starts from 0.0005 and is divided 10 every 50 training epoches. For the content suppression sub-network, three kernels are initialized to the KV kernel and no bias vectors are incorporated in this sub-network. The learning rate for the content suppression sub-network is set to 0.0001. The size of mini-batch SGD, N, is set to 20 and the number of training epoch is set to 200.

4.1. Demonstration of Adaptive Content Suppression

In this experiment, we investigate the influence of adaptive content suppression for image steganalysis. The proposed model is compared with a baseline network that has fixed highpass kernels in the content suppression sub-network. To make the result comparable, three highpass kernels in the baseline network are set to the KV kernel. For the architecture, we set $[n_1, n_2, n_3, n_4]$ (the number of ResL blocks illustrated in Fig.1) of both networks to $[2, 2, 1, 1]$. Thus, the number of total convolutional layers in two networks is 20. We use the MATLAB version Spatial domain UNIversal Wavelet Relative Distortion (S-UNIWARD) stegnography [17], which can be downloaded from Fridrich's homepage [3]. The payload is set to 0.4 bit-per-pixel (bpp). 30,000 randomly selected cover images from the cropped BOSSbase and their stegos are used as the training set. The rest images and their stegos are used for testing.

[3] http://dde.binghamton.edu/download/stego_algorithms/

Table 1 shows the training error and testing error for the proposed network and the baseline network. The result demonstrates that an adaptively learned content suppression sub-network can improve the performance obviously.

Table 1. Detection errors for the network with and without adaptive content suppression on SUNIWARD at 0.4 bpp.

Configuration	Adaptive kernel	Fixed kernel
Detection error	5.97%	8.21%

4.2. Performance Comparisons with Prior Arts

We conduct a comprehensive experiment to demonstrate the effectiveness of the proposed network. We compare the proposed network with the classical Spatial Rich Model based steganalysis (SRM) [5] and its select-channel-aware version, the maxSRMd2 steganalysis [6]. Four states of the art steganographic algorithms, including the Wavelet Obtained Weights steganography (WOW) [18], S-UNIWARD [17], the HIgh-pass Low-pass Low-pass steganography (HILL) [19] and the Minimizing the Power of Optimal Detector steganography (MiPOD) [20], are used for validation. Unlike Pibre in [12] used a same embedding key for all images, all steganographic algorithms hide secret messages for each cover image with different embedding keys. Same to the setting in the previous experiment, $[n_1, n_2, n_3, n_4]$ of the model is set to $[2, 2, 1, 1]$. 30,000 randomly selected images and their stegos are for training the model, the rest 10,000 and their stegos are for testing.

Table 2 gives the detection error rates of the proposed network. The results demonstrate that our network achieves much lower detection error rates than the rich model based steganalysis over all settings. In addition, we compare the proposed network with four states of the art CNN based steganalytic algorithms, including Qian's network [10] and Xu's network [11]. Since Alexnet [21] is a representative CNN model for image classification, we train an Alexnet model for image steganalysis and report its performnace here. **Table 3** gives the detection error rates of these CNN models. We find that the proposed model achieves lowest detection errors.

5. CONCLUSIONS

This paper introduced a unified convolutional neural network for image steganalysis. The proposed network has two improvements over previous CNN based steganalytic methods. On one hand, our network unifies image preprocessing and feature learning in a whole model. On the other hand, we proposed a novel subnetwork to actively preserve the weak stego signal based on residual learning. The experiment demonstrates that adaptive content suppression can improve the detection accuracy obviously. In addition, theoretical analysis

Table 2. Detection error rates for SRM, maxSRMd2 and the proposed network on four steganographic algorithms.

Steganography	WOW		S-UNIWARD		HILL		MiPOD	
	0.3 bpp	0.4 bpp	0.3 bpp	0.4 bpp	0.3 bpp	0.4 bpp	0.3 bpp	0.4 bpp
SRM	24.76%	20.08%	25.51%	20.70%	29.39%	23.57%	26.12%	22.26%
maxSRMd2	17.98%	15.20%	22.35%	18.84%	25.71%	21.63%	24.19%	20.38%
Proposed network	**7.72%**	**3.46%**	**9.92%**	**5.97%**	**9.17%**	**6.23%**	**8.73%**	**3.95%**

and experimental results show that residual learning can effectively overcome the feature diminishing phenomenon in steganalysis, thus enables the proposed model to detect modern steganographic algorithms at high accuracies.

Current network shows promising performances on detecting spatial domain steganography. In future works, we will extend this network to detect steganography in the compressed domain.

Table 3. Detection error rates for CNN models on three steganographic algorithms at payload 0.4 bpp. '\' denotes that the result is not reported in the paper.

Steganography	WOW	S-UNIWARD	HILL
Qian's network [10]	29.30%	30.90%	\
Xu's network [11]	\	19.76%	20.76%
Alexnet network [21]	29.03%	28.22%	30.88%
Proposed network	**3.46%**	**5.97%**	**6.23%**

6. REFERENCES

[1] 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.

[2] H. Wang and S. Wang, "Cyber warfare: steganography vs. steganalysis," *Communications of the ACM*, vol. 47, no. 10, pp.76-82, 2004.

[3] N. Provos and P. Honeyman, "Detecting steganographic content on the internet," *NDSS*, 2002.

[4] G. S. Lin, C. H. Yeh, C. J. Kuo, "Data hiding domain classification for blind image steganalysis," *ICME*, 2004.

[5] J. Fridrich and J. Kodovsky, "Rich models for steganalysis of digital images," *IEEE Tansactions on Information Forensics and Security*, vol. 7, no. 3, pp.868-882, 2012.

[6] T. Denemark, V. Sedighi, V. Holub, R. Cogranne, and J. Fridrich, "Selection-channel-aware rich model for steganalysis of digital images," *WIFS*, 2014.

[7] S. Zhong, Y. Liu, B. Li, and J. Long, "Query-oriented unsupervised multi-document summarization via deep learning model," *Expert Systems with Applications*, 42(21): 8146-8155, 2015.

[8] S. Zhong, Y. Liu, K. A. Hua, "Field effect deep networks for image recognition with incomplete data," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 12, 2016.

[9] S. Tan and B. Li, "Stacked convolutional auto-encoders for steganalysis of digital images," *APSIPA*, 2014, pp. 1-4.

[10] Y. Qian, J. Dong, W. Wang, and Tieniu Tan, "Deep learning for steganalysis via convolutional neural networks," *SPIE Media Watermarking, Security, and Forensics*, vol. 9409, 2015.

[11] G. Xu, H. 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.

[12] 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 sourcemismatch," *SPIE Media Watermarking, Security, and Forensics*, 2016.

[13] S. Wu, S. Zhong, and Y. Liu, "Steganalysis via deep residual network," *ICPADS*, 2016.

[14] S. Wu, S. Zhong, and Y. Liu, "Deep residual learning for image steganalysis," *Multimedia Tools and Applications*, 2017.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CVPR*, 2016.

[16] P. Bas, T. Filler, and T. Pevny, "Break our steganographic system: the ins and outs of organizing BOSS," *Information Hiding Conference*, 2011.

[17] V. Holub, J. Fridrich, and T. Denemark, "Universal distortion function for steganography in an arbitrary domain," *EURASIP Journal on Information Security*, vol. 1, no. 1, pp.1-13, 2014.

[18] V. Holub and J. Fridrich, "Designing Steganographic Distortion Using Directional Filters," *WIFS*, 2012.

[19] B. Li, M. Wang, J. Huang, and X. Li, "A new cost function for spatial image steganography," *ICIP*, 2014, pp.4206-4210.

[20] V. Sedighi, R. Cogranne, and J. Fridrich, "Content-Adaptive steganography by minimizing statistical detectability," *IEEE Tansactions on Information Forensics and Security*, vol. 1, no. 2, pp.221-234, 2016.

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *NIPS*, 2012.