WISERNet: Wider Separate-then-reunion Network for Steganalysis of Color Images

Jishen Zeng, *Student Member, IEEE,* Shunquan Tan*, *Senior Member, IEEE,* Guangqing Liu,
Bin Li, *Senior Member, IEEE,* and Jiwu Huang, *Fellow, IEEE*

Abstract—Until recently, deep steganalyzers in spatial domain have been all designed for gray-scale images. In this paper, we propose WISERNet (the wider separate-then-reunion network) for steganalysis of color images. We provide theoretical rationale to claim that the summation in normal convolution is one sort of "linear collusion attack" which reserves strong correlated patterns while impairs uncorrelated noises. Therefore in the bottom convolutional layer which aims at suppressing correlated image contents, we adopt separate channel-wise convolution without summation instead. Conversely, in the upper convolutional layers we believe that the summation in normal convolution is beneficial. Therefore we adopt united normal convolution in those layers and make them remarkably wider to reinforce the effect of "linear collusion attack". As a result, our proposed wide-and-shallow, separate-then-reunion network structure is specifically suitable for color image steganalysis. We have conducted extensive experiments on color image datasets generated from BOSSBase raw images and another large-scale dataset which contains 100,000 raw images, with different demosaicking algorithms and down-sampling algorithms. The experimental results show that our proposed network outperforms other state-of-the-art color image steganalytic models either hand-crafted or learned using deep networks in the literature by a clear margin. Specifically, it is noted that the detection performance gain is achieved with less than half the complexity compared to the most advanced deep-learning steganalyzer as far as we know, which is scarce in the literature.

Index Terms—steganalysis, steganography, deep learning, convolutional neural network.

I. INTRODUCTION

IN the last decade, the main battleground of spatial image information hiding is at gray-scale cover images. However, the confrontation between color image steganography and its rival, color image steganalysis has drawn ever greater attentions of researchers due to the fact that the vast majority of digital images in real life have colors.

Most of the modern gray-scale steganographic algorithms including famed SUNIWARD [1], HILL [2] and MiPOD [3] adopt the so-called additive embedding distortion minimizing framework [4]. Move a step further, Denemark et al. [5] and Li et al. [6] (named CMD steganography) independently constructed an effective approach to preserve the correlation between neighboring pixels. Generally, gray-scale image steganographic algorithms, such as SUNIWARD and HILL can be directly used for color images, by treating every color band [1] as a gray-scale image and embedding secret bits into the bands independently. There has been no steganographic algorithms aiming at color images until recently. In 2016, inspired by CMD steganography, Tang et al. proposed a so-called CMD-C steganography [7] purposely for color images. CMD-C can preserve not only the correlation within each color band, but also the correlation among three color bands. Therefore it can better resist steganalysis targeted for color images. We denote the CMD-C steganography using SUNIWARD [1] and HILL [2] for initialization as CMD-C-SUNIWARD and CMD-C-HILL, respectively. [2]

The current dominant universal/blind gray-scale steganalytic detectors use *rich models* with tens of thousands of features [8], [9], and an ensemble classifier [10]. Rich models for color image steganalysis had been proposed even before specialized color steganography arose. In 2014, Goljan et al. proposed the first color rich-model steganalytic features set (CRM) [11]. Since most of the digital color images are captured by cameras with one single sensor plus a Color Filter Array (CFA), the CFA demosaicking algorithm used in the generation procedure introduces constraints on the relationship between neighboring color pixels, which can be utilized in steganalysis. In [12], Goljan et al. proposed a CFA-aware rich model for color image steganalysis. In [13], Abdulrahman et al. proposed another CFA-aware rich model. Further on, they fused CRM with features extracted from inter-band geometric transformation measures (GCRM) [14], and features based on steerable Gaussian filters bank (SGRM) [15]. However, the applicability of the features proposed in [12]–[15] remains limited due to the fact that varying post-processing procedures,

[1] In the literature, the red/greeen/blue channels in true-color images are also called "color bands". Throughout the paper, we adopt the term "color bands" to avoid confusion with the term "channels" in deep-learning architectures.

[2] Throughout this paper, the acronyms used for the steganographic and steganalytic algorithms are taken from the original papers. The corresponding full names are omitted for brevity.

e.g. down-sampling and rotating, tend to severely weaken CFA related correlations.

Characteristics of the specific attacking targets [16], [17] and the specific cover sources [18] can be utilized to further improve detection performance of rich models. But, utilizing the knowledge of attacking targets might violate the purpose of universal/blind steganalysis, while utilizing the characteristics of specific cover sources might lead to over-optimized detectors. Therefore in our work we adhere to generic universal steganalysis with neither the knowledge of specific steganographic algorithms (e.g. content-adaptive) nor the knowledge of specific cover sources.

In recent years, deep-learning networks have achieved overwhelming superiority over conventional approaches in many fields [19]. Researchers in image steganalysis have also tried to investigate the potential of deep-learning networks in this field. Started from the pioneer work of Tan and Li [20], researchers have kept trying to promote detection performance of deep-learning steganalyzers [21]–[23]. In 2016, Xu et al. proposed a Convolutional Neural Network (CNN) structure for image steganalysis which outperforms handcrafted steganalytic features [24] (referred as Xu's model #1). In the following years, the literature witnessed deep-learning steganalyzers going deeper and more complicated. Ye et al. proposed a deeper CNN equipped with a new activation function (TLU) which can further boost detection performance [25] (referred as Ye's model). In JPEG domain, Zeng et al. proposed a hybrid CNN steganalyzers equipped with quantization and truncation, which is obviously superior to hand-crafted JPEG steganalytic features [26], [27]. Chen et al. proposed a JPEG-phase-aware deep CNN steganalyzer [28]. Inspired by ResNet [29], Xu proposed a much deeper CNN based JPEG steganalyzer [30] (referred as Xu's model #2). However, all of the above deep-learning steganalyzers are gray-scale image oriented. According to our best knowledge, there is no report to address deep-learning based color image steganalysis.

In this paper, we propose WISERNet, a specific wider separate-then-reunion network for steganalysis of color images. We claim that the summation in normal convolution is one sort of "linear collusion attack" which is the process of forming a linear combination of input bands [31]. It reserves strong correlated patterns while impairs uncorrelated noises. Since the main purpose of the convolutions in the bottom convolutional layer is to suppress correlated image contents, for the bottom convolutional layer we discard summation and introduce channel-wise convolution. On the other hand, in the upper convolutional layers we still adopt normal convolution which retains summation and make them remarkably wider, in order to reinforce the effect of "linear collusion attack". This is because we believe such an upper structure is beneficial to the ability of CFA related "hidden-pattern-aware". We have conducted extensive experiments on color image datasets generated from public true-color raw images as well as a large-scale dataset which contains 100,000 raw images. The results demonstrated the superiority of our proposed network over other state-of-the-art steganalyzers either hand-crafted or learned using deep networks in the literature.

The rest of the paper is organized as follows. In Sect. II, we firstly show the theoretical rationale of our proposed WISERNet, and then describe the detailed structure of WISERNet. Results of experiments conducted on different color image datasets are presented in Sect. III. Finally, we make conclusions in Sect. IV.

II. Our proposed WISERNet

In this section, we firstly introduce convolutional layers, the principal part of CNN as preliminaries. Then we discuss the motivation of our proposed WISERNet in theory. Finally we provide the conceptual architecture of WISERNet, as well as its detailed configuration.

A. Preliminaries

In this paper we only consider RGB true-color model. Given $\mathbf{X}$, a true-color image of size of $M \times N$, it comprises three bands, namely the *red*, the *green*, and the *blue* band. In our research, we do not take the specific characteristic of a band into consideration. Therefore without loss of generality, $\mathbf{X}$ can be represented as $\{\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3\}$, where $\mathbf{X}_i = (x_{i,pq})^{M \times N}$, $x_{i,pq} \in \{0, 1, \cdots, 255\}$, $1 \leq i \leq 3$, $1 \leq p \leq M$ and $1 \leq q \leq N$. All of the state-of-the-art steganographic algorithms which can be applied on color images ([1], [2], [5]–[7]) add zero-mean ± 1 additional stego noise to every target band of a given cover image. Therefore, every band in $\mathbf{X}$ can be further represented as $\mathbf{X}_i = (c_{i,pq} + n_{i,pq})^{M \times N} = \mathbf{C}_i + \mathbf{N}_i$, where $\mathbf{C}_i = (c_{i,pq})^{M \times N}$, $c_{i,pq} \in \{0, 1, \cdots, 255\}$ denotes the corresponding cover band, and $\mathbf{N}_i = (n_{i,pq})^{M \times N}, n_{i,pq} \in \{-1, 0, +1\}$ denotes the additive stego noise matrix added to $\mathbf{X}_i$. For an innocent cover image, $\mathbf{N}_i$, $1 \leq i \leq 3$ are all zero matrices.

All of the state-of-the-art deep-learning steganalyzers [20]–[28], [30] are based on CNN (Convolutional Neural Network). The principal part of CNN is a cascade of alternating convolutional layers, regulation layers (e.g. BN layers [32]) and pooling layers. On top of the principal part, there are optional multiple fully-connected layers. Convolutional layers are the core building blocks of a CNN. For a given convolutional layer L_l, it takes J channels of inputs $\mathbf{Z}_j^{l-1}$, $1 \leq j \leq J$, convolves them with an array of $J \times K$ kernels $\mathbf{W}_{jk}^l$ (usually with learnable weights) and generates K channels of outputs $\mathbf{Z}_k^l$, $1 \leq k \leq K$. In the context of image steganalysis, $\mathbf{Z}_j^{l-1}$, $\mathbf{Z}_k^l$, and $\mathbf{W}_{jk}^l$ are always represented as two-dimensional matrices. The *normal convolution* can be modeled as:

$$\mathbf{Z}_k^l = \sum_{j=1}^J \mathbf{Z}_j^{l-1} * \mathbf{W}_{jk}^l, 1 \leq k \leq K \qquad (1)$$

We utilize *channel-wise convolution*, a variant of normal convolution in our work. In channel-wise convolution, each input channel corresponds to standalone K output channels and is convolved with an array of K kernels. As a result with J input channels we can get $J \times K$ output channels:

$$\begin{aligned} \mathbf{Z}_{k'}^l &= \mathbf{Z}_j^{l-1} * \mathbf{W}_{jk}^l, \\ k' &= (j-1) \times K + k, 1 \leq j \leq J, 1 \leq k \leq K \end{aligned} \qquad (2)$$

TABLE I
MEANS OF THE ABSOLUTE VALUE OF THE CORRELATION BETWEEN THE INTENSITY VALUES OF DIFFERENT COLOR BANDS VERSUS THOSE OF THE CORRESPONDING STEGO NOISES. THE RESULTS ON BOSS-PPG-LAN FOR HILL, CMD-C-HILL, SUNIWARD, AND CMD-C-SUNIWARD WITH 0.4 BPC PAYLOAD ARE REPORTED.

Elements	*red* band vs. *green* band	*red* band vs. *blue* band	*blue* band vs. *green* band
For HILL steganography			
Intensity values	0.9317	0.8297	0.9217
Stego noises	0.0024	0.0023	0.0024
For CMD-C-HILL steganography			
Intensity values	0.9317	0.8297	0.9217
Stego noises	0.2704	0.2619	0.2859
For SUNIWARD steganography			
Intensity values	0.9317	0.8297	0.9217
Stego noises	0.0022	0.0021	0.0022
For CMD-C-SUNIWARD steganography			
Intensity values	0.9317	0.8297	0.9217
Stego noises	0.2980	0.2817	0.2975

The existing deep-learning steganalyzers [20]–[28], [30] incorporate the domain knowledge behind rich models, and initialize the kernels in the bottom convolutional layer as high-pass filters to increase SNR (Signal-to-Noise Ratio) [3]. Fixed weights in the bottom kernels are adopted in most existing deep-learning steganalyzers [20]–[24], [26]–[28], [30], while learnable weights are adopted in Ye's model [25].

B. Rationale of our proposed WISERNet

If we apply normal convolution with K output channels to a true-color image in the bottom convolutional layer, we get:

$$\mathbf{Z}_k^1 = \sum_{j=1}^{3} \mathbf{X}_j * \mathbf{W}_{jk}^1, \ 1 \le k \le K$$

$$= \sum_{j=1}^{3} \mathbf{C}_j * \mathbf{W}_{jk}^1 + \sum_{j=1}^{3} \mathbf{N}_j * \mathbf{W}_{jk}^1, \ 1 \le k \le K \quad (3)$$

Our rationale starts from the following statement about existing steganographic algorithms for color images:

For a true-color stego image, the intensity values in the same location of the three bands exhibit strong correlation. Their means (or expectations) are similar from the perspective of statistics. Conversely, for steganographic algorithms originally oriented to gray-scale images, e.g. [1]–[3], [5], [6], the zero-mean ±1 additional stego noises in the same location of the three bands exhibit no correlation. Even for [7] which is committed to increase the correlation of the stego noises among bands, they still exhibit weak correlation. This is because the well-established rate-distortion bound [33] determines that strong correlation among stego noises and minimal possible distortion are in conflict.

[3]In the literature of steganalysis, image content is "noise" while stego noise is "signal".

To verify the above statement, we analyzed the $10,000$ BOSS-PPG-LAN (see Sect. III-A) stego images generated by HILL, CMD-C-HILL, SUNIWARD, and CMD-C-SUNIWARD, respectively, all with 0.4 bpc (bits per channel/band pixel) embedding rate. In the experiment, means of the absolute value of the correlation between the intensity values of different color bands were calculated. We compared them with those of the corresponding stego noises. From Tab. I, we can see that for all of the four steganographic algorithms, stego noises have no effect on the correlation of the intensity values among bands. They all exhibited notably strong correlation. On the other hand, for HILL and SUNIWARD, the stego noises among bands exhibit nearly zero correlation. Even for CMD-C-HILL and CMD-C-SUNIWARD, they exhibit weak correlation.

As mentioned in Sect. II-A, the main purpose of the convolutions in the bottom convolutional layer is to boost SNR, namely suppress image contents (intensity values) and retain stego noises at the same time. Let $E([\bullet])$ and $Var([\bullet])$ denote the expectation and the variance of the elements in matrix $[\bullet]$; $Corr([\bullet], [\square])$ denotes the correlation of the corresponding elements in matrix $[\bullet]$ and $[\square]$. Following the convention in image processing [34], we define SNR of a given two-dimensional input $[\bullet]$ as:

$$SNR([\bullet]) = \frac{Var([\bullet])}{E^2([\bullet])} \quad (4)$$

According to the linearity of expectation and (3), for a given $\mathbf{Z}_k^1$ we can get:

$$SNR(\mathbf{Z}_k^1) = \frac{Var(\sum_{j=1}^{3} \mathbf{N}_j * \mathbf{W}_{jk}^1)}{E^2(\sum_{j=1}^{3} \mathbf{C}_j * \mathbf{W}_{jk}^1)} = \frac{Var(\sum_{j=1}^{3} \mathbf{N}_j * \mathbf{W}_{jk}^1)}{(\sum_{j=1}^{3} E(\mathbf{C}_j * \mathbf{W}_{jk}^1))^2} \quad (5)$$

Initially, we set $\mathbf{W}_{1k}^1 = \mathbf{W}_{2k}^1 = \mathbf{W}_{3k}^1 = \widetilde{\mathbf{W}}_k$, where $\widetilde{\mathbf{W}}_k$ is a predefined high-pass filter. Since $E(\mathbf{C}_1) \approx E(\mathbf{C}_2) \approx E(\mathbf{C}_3)$, we can get $E(\mathbf{C}_1 * \mathbf{W}_{1k}^1) \approx E(\mathbf{C}_2 * \mathbf{W}_{2k}^1) \approx E(\mathbf{C}_3 * \mathbf{W}_{3k}^1) = \mu$ (as demonstrated later in Tab. XI of Sect. III-D, the assumption still holds along with increasing training iterations of WISERNet). Denote $Var(\mathbf{N}_j * \mathbf{W}_{jk}^1) = \sigma_j^2, \ 1 \le j \le 3$. Existing color image steganography tends to uniformly distribute embedding changes to three color bands, therefore for the sake of simplicity, we further assume that $\sigma_j = \sigma, \ 1 \le j \le 3$. From (5) we can get:

$$SNR(\mathbf{Z}_k^1) = \frac{\sum_{j=1}^{3} Var(\mathbf{N}_j * \mathbf{W}_{jk}^1) + \Delta}{9\mu^2} = \frac{\sum_{j=1}^{3} \sigma_j^2 + \Delta}{9\mu^2} \quad (6)$$

in which:

$$\Delta = 2 \cdot \sum_{1 \le i < j \le 3} Corr(\mathbf{N}_i * \mathbf{W}_{ik}^1, \mathbf{N}_j * \mathbf{W}_{jk}^1) \cdot \sigma_i \sigma_j \quad (7)$$

Please note that a discrete convolution is a linear transform, and a linear transform will never change the correlation between random variables [35]. Therefore:

$$Corr(\mathbf{N}_i * \mathbf{W}_{ik}^1, \mathbf{N}_j * \mathbf{W}_{jk}^1) = Corr(\mathbf{N}_i, \mathbf{N}_j), \ 1 \le i < j \le 3 \quad (8)$$

TABLE II
IMPACT OF GROWING $\rho_{\text{R-G}} = \rho_{\text{R-B}} = \rho$ ON $\frac{\text{MMD}_c}{\text{MMD}_n}$.

$\rho_{\text{r-g}} = \rho_{\text{r-b}} = \rho$										
0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
1.178	1.158	1.124	1.108	1.064	1.038	1.027	1.021	1.013	1.008	1.000

From (7) and (8) we can set $\Delta = 0$ for HILL and SUNI-WARD since the inter-band correlation of the stego noises generated by them is nearly zero. As a result:

$$\text{SNR}(\mathbf{Z}_k^1) = \frac{\sum_{j=1}^3 \sigma_j^2}{9\mu^2} = \frac{1}{3} \cdot \frac{\sigma^2}{\mu^2} \qquad (9)$$

For CMD-C, Δ is unneglectable. However, the inter-band correlation of the stego noises stays weak. Using the statistics reported in Tab. I, we set $\text{Corr}(\mathbf{N}_i, \mathbf{N}_j) \approx 0.3, \ 1 \leq i < j \leq 3$. From (6) we can get:

$$\text{SNR}(\mathbf{Z}_k^1) \approx \frac{3\sigma^2 + 2 \cdot 3 \cdot 0.3 \cdot \sigma^2}{9\mu^2} = \frac{5}{9} \cdot \frac{\sigma^2}{\mu^2} \qquad (10)$$

Please note that:

$$\text{SNR}(\mathbf{X}_j * \mathbf{W}_{jk}^1) = \frac{\text{Var}(\mathbf{N}_j * \mathbf{W}_{jk}^1)}{\text{E}^2(\mathbf{C}_j * \mathbf{W}_{jk}^1)} = \frac{\sigma^2}{\mu^2}, \ 1 \leq j \leq 3 \qquad (11)$$

Compare (9), (10) with (11), we can see for those steganographic algorithms applied on color images, $\text{SNR}(\mathbf{Z}_k^1) < \text{SNR}(\mathbf{X}_j * \mathbf{W}_{jk}^1), \ 1 \leq j \leq 3$. The summation in $\mathbf{Z}_k^1$ actually impairs the SNR which has been boosted by the convolutions in $\mathbf{X}_j * \mathbf{W}_{jk}^1, \ 1 \leq j \leq 3$. In fact, we can regard the summation in normal convolution one sort of "linear collusion attack" [31] which is the process of forming a linear combination of input bands, and as a result can reserves strong correlated patterns, while impairs uncorrelated noises (or weak correlated signals) in input bands. Accordingly, we decide not to apply normal convolution in the bottom convolutional layer.

In practice, there are three solutions to bypass the summation in normal convolution. The first solution is to directly concatenate the three bands of $\mathbf{X}$ to generate a one-band input $\mathbf{X}'$ for the bottom convolutional layer, which is straightforward:

$$\begin{aligned} \mathbf{X}' &= (x'_{pq'})^{M \times 3N}, \\ x'_{pq'} &= x_{i,pq}, \ q' = (i-1) \times N + q, \\ & 1 \leq i \leq 3, \ 1 \leq p \leq M, \ 1 \leq q \leq N \end{aligned} \qquad (12)$$

The second solution is to interleave the intensity values from three bands, e.g. in a fashion that $(x_{1,pq}, \ x_{2,pq}, \ x_{3,pq}, \ x_{1,(p+1)q}, \ x_{2,(p+1)q}, \ x_{3,(p+1)q}, \ \cdots)$. The third solution is channel-wise convolution. We adopt channel-wise convolution in our proposed network since we believe the first two solutions are inferior to the third solution. The argument is as follows.

Given $(x_{1,pq}, x_{2,pq}, x_{3,pq})$, three intensity values in the same location of the three color bands. We have already known that they exhibit strong correlation. If we adopt the first solution, a.k.a. the straightforward solution, then $(x_{1,pq}, x_{2,pq}, x_{3,pq})$

corresponds to $(x'_{pq}, x'_{p(N+q)}, x'_{p(2N+q)})$ in $\mathbf{X}'$. Please note that the column distance of $(x'_{pq}, x'_{p(N+q)}, x'_{p(2N+q)})$ is N, far beyond the usual perception field of the kernels in lower convolutional layers. As a result, the originally strong correlation in $(x_{1,pq}, x_{2,pq}, x_{3,pq})$, after mapped to $(x'_{pq}, x'_{p(N+q)}, x'_{p(2N+q)})$, cannot be catched in lower convolutional layers. Only convolution kernels in the top layers of very deep networks can perceive it. The second solution is also infeasible. You can think of it as a noisy up-sampling procedure of a given band in which the step size is two. Those weak ± 1 stego noises will be by large concealed under the relatively more powerful up-sampling noises.

On the other hand, if we adopt channel-wise convolution, then after convolution, those elements affected by $(x_{1,pq}, x_{2,pq}, x_{3,pq})$ are still allocated at the corresponding locations of the output channels. The strong correlation in them is thus kept, and can be perceived by all of the convolutional layers from bottom to top.

In order to verify that the bottom channel-wise convolutional layer do help to preserve stego noises and consequently boost the SNR, we conducted an evaluation experiment as follows:

Firstly, 40% pixels of every band of every cover image in a dataset used in our experiments (BOSS-PPG-LAN, please refer to Sect. III-A) are stochastically selected to perform random ± 1 modifications. The modifications simulate the effect of naïve LSB matching embedding. Then we apply normal convolution or channel-wise convolution to the cover images and their corresponding pseudo-stego images. Only $\mathbf{K}_5$, one of the 30 spatial-domain rich model kernels [8], is used in the convolution to further reduce the complexity. With $\mathbf{K}_5$, whether normal convolution or channel-wise convolution generates a single output feature map. Next, we extract the 686-dimensional SPAM (Subtractive Pixel Adjacency Matrix) [36] steganalytic feature vector for every output feature map. Finally, we compute the MMD (Maximum Mean Discrepancy) [37] between the cover images and the corresponding pseudo-stego images in the SPAM feature space. The presence of a higher MMD value indicates that it is easier to distinguish stego images from cover images.

Please note that a given stego image can be considered to be a noisy version of the corresponding cover image, and $\mathbf{K}_5$ is a high-pass filter aims at boosting the SNR. Therefore with the same $\mathbf{K}_5$ kernel, the convolution type with relatively higher MMD value helps to preserve stego noises. Let MMD_n and MMD_c denotes the MMD value for the output feature maps generated by normal convolution and channel-wise convolution, respectively. We can get $\text{MMD}_n = 0.01187$ while $\text{MMD}_c = 0.01398$. As a result $\frac{\text{MMD}_c}{\text{MMD}_n} = 1.178$, which implies
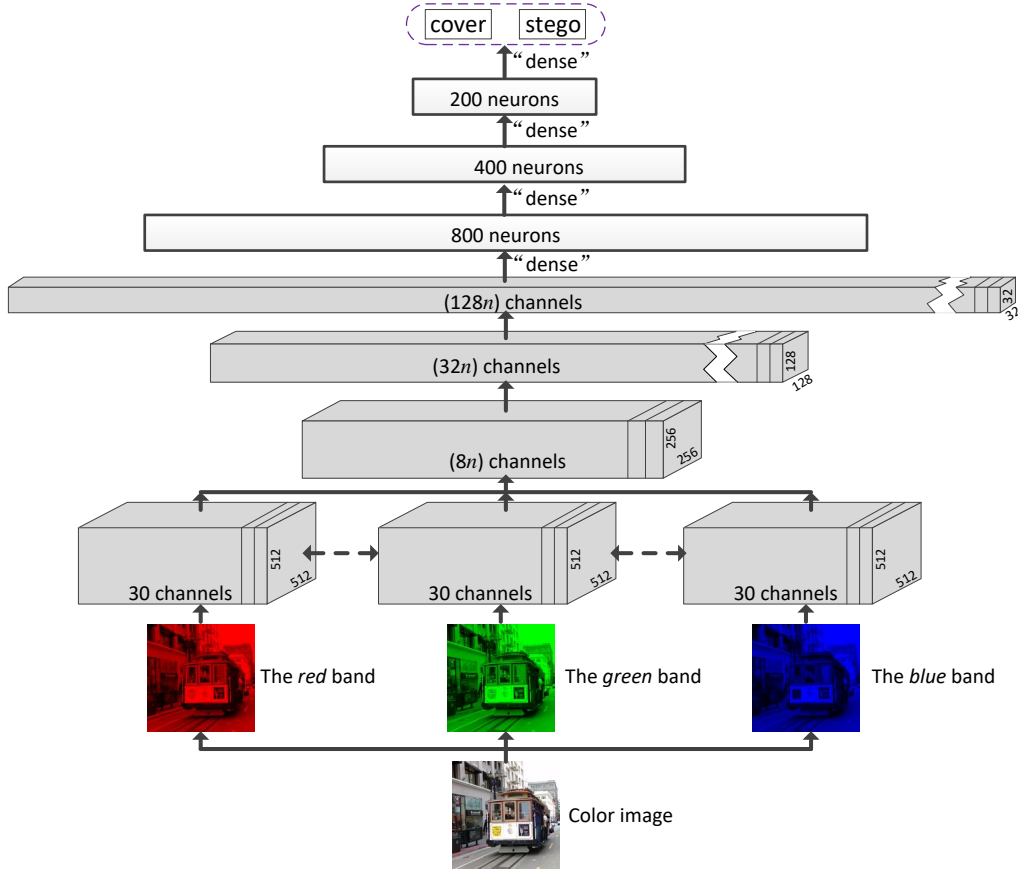
Fig. 1. Conceptual architecture of our proposed WISERNet.

that the bottom channel-wise convolutional layer do help to preserve stego noises and consequently boost SNR when stego noises in separate color bands exhibit no correlation.

Furthermore, we investigate the impact of growing correlation of stego noises in different bands on $\frac{\mathrm{MMD}_c}{\mathrm{MMD}_n}$. For every cover image, we firstly stochastically select 40% pixels of the *red* band to perform random ±1 modifications. Let $\rho_{\text{r-g}}$ denotes the correlation of the ±1 modifications in the *red* band and those in the *green* band. Analogically, The meaning of $\rho_{\text{r-b}}$ and $\rho_{\text{g-b}}$ are self-evident. The ±1 modifications are performed on the *green* band and the *blue* band, to guarantee that $\rho_{\text{r-g}} = \rho_{\text{r-b}} = \rho$, $0 < \rho \leq 1$. It is easy to verify that $2\rho^2 - 1 \leq \rho_{\text{g-b}} \leq 1$ with $\rho_{\text{r-g}} = \rho_{\text{r-b}} = \rho$, and $\rho_{\text{g-b}} \to 1$ along with $\rho \to 1$. Therefore as a compromise, we only investigate the impact of growing $\rho_{\text{r-g}} = \rho_{\text{r-b}} = \rho$ on $\frac{\mathrm{MMD}_c}{\mathrm{MMD}_n}$. As shown in Tab. II, $\frac{\mathrm{MMD}_c}{\mathrm{MMD}_n}$ keeps reducing along with increasing ρ, which implies that the gain introduced by channel-wise convolution gradually decreases with increasing inter-band correlation of stego noises. However, since even for the state-of-the-art steganographic algorithms (e.g. CMD-C) taking into account the correlation of stego noises among bands, the inter-band correlation of stego noises still keeps weak (is approximately equal to 0.3, as shown in Tab. I). Therefore even for those algorithms like CMD-C, the gain introduced by channel-wise convolution still cannot be neglected.

We decide to only adopt channel-wise convolution and withdraw summation in the bottom convolutional layer. In the upper convolutional layers we still adopt normal convolution which retains summation. This is because withdrawal of the summation in convolution is a double-edged sword. Please note that during the generation procedure of true-color images, various kinds of sources, especially the CFA interpolation algorithm, introduces dependencies among pixels and bands. The dependencies can be regarded as hidden patterns stay approximately the same in color bands. Though those hidden patterns are severely weaken in varying post-processing procedures, it is still potential to be catched by deep-learning networks. In a CNN based deep-learning network, since we have already known discrete convolution will never change the correlation between input channels, for a given convolutional layer L_l we can also get:

$$\mathrm{Corr}(\mathbf{Z}_i^{l-1} * \mathbf{W}_{ik}^l, \mathbf{Z}_j^{l-1} * \mathbf{W}_{jk}^l) = \mathrm{Corr}(\mathbf{Z}_i^{l-1}, \mathbf{Z}_j^{l-1}) \qquad (13)$$

Therefore if we model the hidden patterns a certain sort of correlation among bands, it can be passed forward through the convolutional layers from bottom to top, even though disturbed by the nonlinearities in the pipeline. Since every normal convolutional layer can regarded as one sort of "linear collusion attack", the summations in cascaded convolutional layers may enhance the ability of "hidden-pattern-aware" of our proposed network. Therefore only in the bottom convolutional layer, which aims at suppressing contents while retaining stego noises, we withdraw summation and adopt channel-wise convolution.

It has been verified that in "linear collusion attack", the more the objects involved in collusion, the better the strong correlated patterns are reserved, the worse the weak correlated signals are impaired [31]. Since we regard the summation in normal convolution as one sort of "linear collusion attack", it is clear that the more the kernels in a given normal convolutional layer, the more the outputs can be involved in summation, and therefore the better results of the "linear collusion attack" we can expect. Based on the above analysis, our design concept is different from typically making the deep-learning network deeper. We try to promote the performance of our proposed steganalyzer by making the upper convolutional layers wider, namely expanding the number of their convolution kernels.

C. WISERNet: the wider separate-then-reunion network

In light of the rationale presented in Sect. II-B, we propose WISERNet, the WIder SEparate-then-Reunion Network. The conceptual architecture of WISERNet is illustrated in Fig. 1.

WISERNet takes a true-color image as input and applies channel-wise convolution to the red, green, and blue bands of the input image, respectively. Following the recipe of prior works [20]–[28], [30], the weights of the kernels assigned to each channel are initialized with high-pass filters in rich models to increase SNR. Here the thirty filters used in SRM [8] are used. Therefore each band is convolved with thirty 5×5 initialized kernels and thirty corresponding output channels are generated. Please note that in the training procedure, we set the weights in the bottom kernels *learnable*. The bottom channel-wise convolutional layer corresponds to the "separate" stage of our proposed network. The three separate groups of output channels are then concatenated together to form a ninety-channel input of the second convolutional layer.

Started from the second convolutional layer, the upper structure of our proposed network corresponds to the "reunion" stage. It is a united wide and relatively shallow convolutional neural network, which contains convolutional layers with plenty of kernels. We fix the depth, namely the number of cascaded convolutional layers of the upper structure to three, and explicitly increase the capacity of WISERNet by magnifying the number of the kernels in each convolutional layer with a model magnification factor n. With increasing n, WISERNet becomes "wider" and "wider". On top of the convolutional layers there is a four-layer fully-connected ("dense") neural network which makes the final prediction. The successive layers of the fully-connected network contain 800, 400, 200, and 2 neurons, respectively.

The detailed configuration of WISERNet is shown in Tab. III. Assume that the true-color image fed to WISERNet is of size 512×512. From Tab. III we can see the output of the bottom channel-wise convolutional layer, the "separate" stage, is ninety channels of feature maps of size 512×512, which act as the input of the "reunion" stage. In the "reunion" stage, all of the normal convolutional layers are followed by a BN (Batch Normalization) layer, a ReLU (Rectified Linear Unit) layer, and an average pooling layer successively. Specifically, the absolute values of the output of the first normal convolutional layer are fed forward, following the

TABLE III
THE DETAILED CONFIGURATION OF OUR PROPOSED WISERNET.

Type	Kernels size/stride (width×height×depth)/stride	Output size (width×height×channel)
Channel-wise convolution	$(5 \times 5 \times 30)/1$	$512 \times 512 \times 90$
Convolution	$(5 \times 5 \times (8n))/2$	$256 \times 256 \times (8n)$
ABS	/	——
BN	/	——
ReLU	/	——
Average pooling	$(5 \times 5 \times 1)/2$	$128 \times 128 \times (8n)$
Convolution	$(3 \times 3 \times (32n))/1$	$128 \times 128 \times (32n)$
BN	/	——
ReLU	/	——
Average pooling	$(5 \times 5 \times 1)/4$	$32 \times 32 \times (32n)$
Convolution	$(3 \times 3 \times (128n))/1$	$32 \times 32 \times (128n)$
BN	/	——
ReLU	/	——
Average pooling	$(32 \times 32 \times 1)/32$	$1 \times 1 \times (128n)$
Flatten	/	$128n$
Fully connection	/	800
ReLU	/	——
Fully connection	/	400
ReLU	/	——
Fully connection	/	200
ReLU	/	——
Fully connection	/	2
Softmax	/	——

recipe of prior works [25]–[28], [30]. The size of the output feature maps of the normal convolutional layers from bottom to top in this stage is 256×256, 128×128, and 32×32, respectively. In order to roughly preserve the time complexity per layer, the number of the kernels in each convolutional layer is quadrupled accordingly. Consequently, the number of the output feature maps is $8n$, $32n$, and $128n$ with the model magnification factor n, respectively. Ahead of the top-most normal convolutional layer, the output feature maps are pooled with a large stride (step=32) and then flatten to a $128n$-D feature vector, which further acts as the input of the top fully-connected network. In the top fully-connected network, ReLU activation functions are used in all three hidden layers. The final layer contains two neurons which denote "stego" prediction and "cover" prediction. Softmax function is used to output predicted probabilities.

III. EXPERIMENTS

A. Experiment setup

As reported in prior works [11]–[15], different CFA demosaicking algorithms and different down-sampling algorithms greatly affect detection performance of existing color image steganalyzers. Therefore, all experiments in this paper are conducted on different versions of BOSSBase (v1.01) [38]. Starting with the 10,000 full-resolution raw images, firstly we followed the dataset generating process used in [11]. We used ufraw to demosaick the raw images and then used ImageMagick "convert" utility with the default "Lanczos" kernel to down-sample (set the smaller image dimension to 512) and central crop the resulting PPM color images to 512×512. We used two demosaicking algorithms in ufraw, PPG (Patterned Pixel Grouping) and AHD (Adaptive Homogeneity Directed). The corresponding datasets were named BOSS-PPG-LAN and BOSS-AHD-LAN. Following the same process as above, with

TABLE IV
DETECTION PERFORMANCE OF CRM, SGRM, AND GCRM ON FOUR DIFFERENT
DATASETS. IN EACH SUB-TABLE, THE BEST RESULTS FOR 0.2 BPC ARE UNDERLINED,
WHILE THE BEST RESULTS FOR 0.4 BPC ARE IN FRAMED BOXES.

Datasets	Rich models					
	CRM		SGRM		GCRM	
	0.2 bpc	0.4 bpc	0.2 bpc	0.4 bpc	0.2 bpc	0.4 bpc
For HILL steganography						
BOSS-PPG-LAN	0.6826	0.8061	0.68	0.8048	0.6743	0.8005
BOSS-PPG-CRP	0.9632	0.9952	0.9611	0.9947	0.9627	0.9954
BOSS-AHD-LAN	0.6817	0.8075	0.6813	0.8068	0.6775	0.8022
BOSS-AHD-CRP	0.9642	0.9942	0.9606	0.9957	0.9632	0.9951
For CMD-C-HILL steganography						
BOSS-PPG-LAN	0.6325	0.7548	0.6333	0.7538	0.6265	0.748
BOSS-PPG-CRP	0.9337	0.9941	0.9329	0.9931	0.9324	0.9947
BOSS-AHD-LAN	0.6363	0.7558	0.6338	0.7528	0.6314	0.7477
BOSS-AHD-CRP	0.9357	0.9935	0.9317	0.9919	0.9331	0.9942

the down-sampling operation removed, we obtained another two datasets, BOSS-PPG-CRP and BOSS-AHD-CRP. In order to explore the impact of different down-sampling operations on detection performance of color image steganalyzers, we replaced the "convert" utility with correspondent Matlab® script. Pairing PPG, or AHD demosaicking algorithm with "Bicubic", or "Bilinear" kernel in Matlab® function *imresize*, we generated four more datasets, BOSS-PPG-BIC, BOSS-PPG-BIL, BOSS-AHD-BIC and BOSS-AHD-BIL, respectively.

Four color image steganographic algorithms, HILL, SUNIWARD, CMD-C-HILL, and CMD-C-SUNIWARD were our attacking targets in the experiments. [4] Their embedding payloads were set to 0.1, 0.2, 0.3, 0.4, and 0.5 bpc, in which we mainly focused on 0.2 bpc and 0.4 bpc. For HILL and SUNIWARD, the same payload was embedded in every band.

In prior works, color image steganalyzers were evaluated on different scenarios and datasets. In order to select the most challenging scenarios and competitors of our proposed WISERNet, we conducted a preliminary experiment in which the detection performance of three state-of-the-art rich models for color image steganalysis, CRM [11], GCRM [14], and SGRM [15] were evaluated on four different datasets. As shown in Tab. IV, different demosaicking algorithms (PPG or AHD) had little effect on the performance of the three rich models. The down-sampled datasets, e.g. BOSS-PPG-LAN and BOSS-AHD-LAN, were the most challenging scenarios for rich models. When evaluated on down-sampled datasets, the performance of CRM was always the best. The performances of the three rich models on BOSS-PPG-CRP and BOSS-AHD-CRP, the datasets without down-sampling, were similar, and were all approaching to 100%. Therefore for brevity, all the rest experiments reported in this paper were only conducted on down-sampled datasets, and only CRM was selected as the representative rich-model based competitor.

As for deep-learning steganalyzers, we selected Ye's model [25], and Xu's model #2 [30] as the representative competitors in the experiments. For Ye's model, the bottom

convolutional layer was equipped with channel-wise convolution with learnable weights. To be fair, its selection-channel-aware version was not included in the experiments. For Xu's model #2, please note that it is designed for gray-scale JPEG image steganalysis. Therefore in the experiments we adopted an alternative version of Xu's model #2 in which the bottom convolutional layer were with 30 fixed SRM kernels and input channel concatenation. For a detailed discussion regarding to the configurations of the bottom convolutional layer of the competing deep-learning steganalyzers, please refer to Sect. III-C.

Our implementation of WISERNet was based on Caffe toolbox [39]. Unless otherwise specified, the model magnification factor of WISERNet was fixed to $n = 9$. It was trained using mini-batch stochastic gradient descent with "inv" learning rate starting from 0.001 (power: 0.75; gamma: 0.0001; weight_decay: 0.0005) and a momentum fixed to 0.9. The batch size in the training procedure was 16 and the maximum number of iterations was set to 30×10^4. The source codes and auxiliary materials are available for download from GitHub [5].

We used the same batch size and maximum number of iterations in the training procedure of Ye's model and Xu's model #2. The settings of all other hyper-parameters for those two deep-learning based competitors followed what reported in the original papers [25], [30]. In every experiment, 6,000 cover-stego pairs were randomly selected for training. The remaining 4,000 cover-stego pairs were for testing. All experiments were repeated ten times, and the mean of predictive accuracies on testing set over ten repetitions were reported. The experiments involved two types of steganalyzers, the rich-model based and the deep-learning based. For those rich-model based steganalyzers, FLD ensemble classifier with default settings [10] was utilized. For those deep-learning steganalyzers, include our proposed WISERNet, 1,000 cover-stego pairs were further randomly picked out from training set for validation. In each experiment, the model was validated and saved every 1×10^4 iterations. The one with the best validation accuracy was evaluated on the corresponding testing set. [6]

B. Comparison to state of the art

Firstly, we report the results of the experiments conducted on BOSS-PPG-LAN and BOSS-AHD-LAN. From Fig. 2 we can see different demosaicking algorithms (PPG or AHD) had little impact on the performance of color image steganalyzers. All of the three deep-learning based models could obtain significant performance improvement compared with CRM, the rich-model based steganalyzer. As for the three deep-learning steganalyzers themselves, the performance of Xu's model #2 (with 30 fixed SRM kernels) was always better than that of Ye's model, which is reasonable since compared with Ye's model, Xu's model #2 is much deeper and more complicated. However, our proposed WISERNet performed even better than Xu's model #2 although it is a relatively shallow and small model.

---

[4]According to peer feedback, we have fixed a bug in the original implementation of CMD-C [7] and guarantee that different pseudo-random seeds are assigned to each of the simulators corresponding to three color bands.

[5]https://github.com/tansq/WISERNet

[6]All of the deep-learning steganalyzers were trained and tested on a GPU cluster with 80 NVIDIA® Tesla® P100 GPU cards. The rich-model based steganalyzers were trained and tested on a CPU cluster with 200 cores.
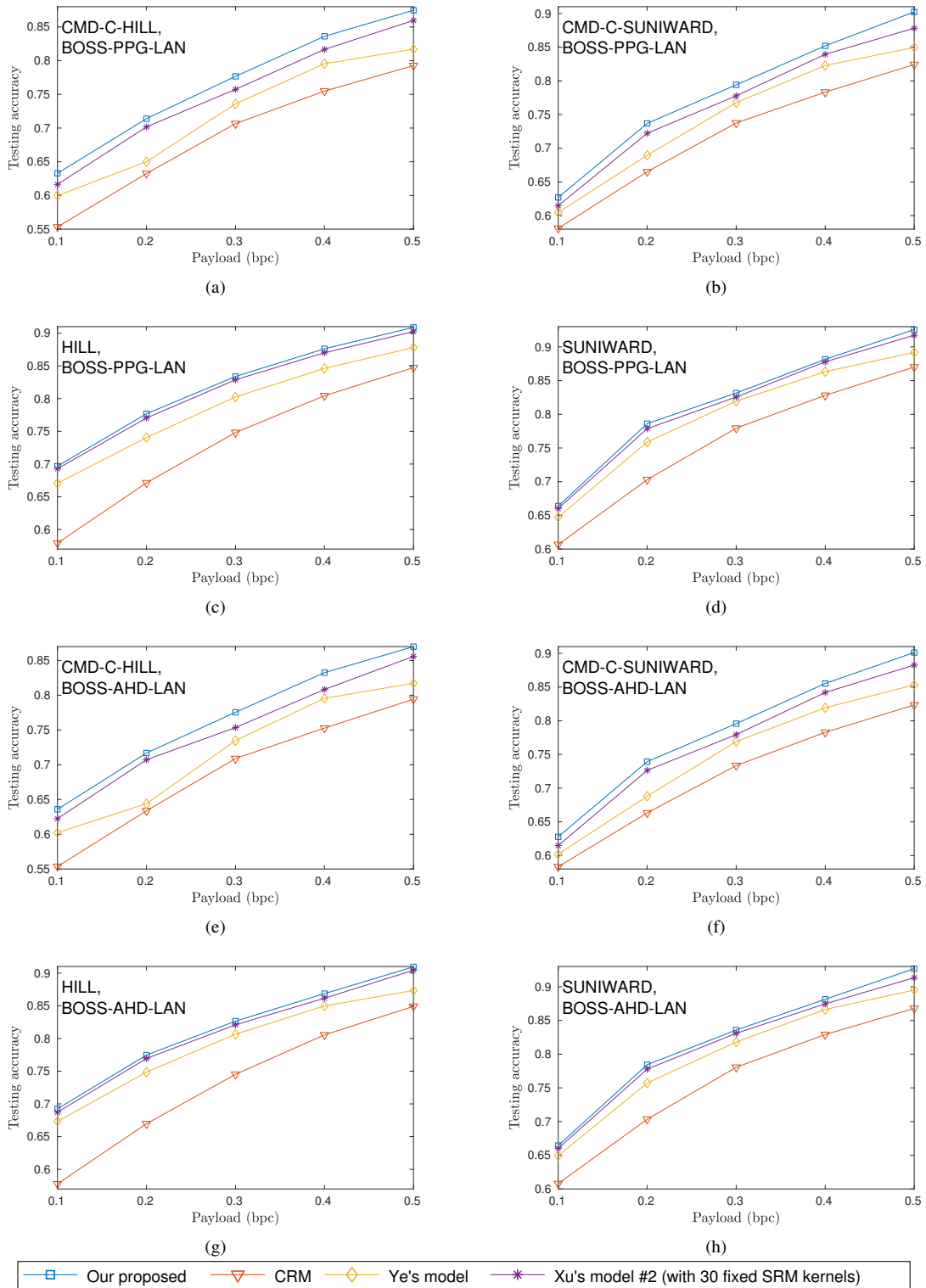
Fig. 2. Comparison of testing accuracy of our proposed WISERNet with state-of-the-art steganalyzers in the literature. (a) and (e) are the results for CMD-C-HILL; (b) and (f) are for CMD-C-SUNIWARD; (c) and (g) are the results for HILL; (d) and (h) are for SUNIWARD. The experiments for (a), (b), (c) and (d) were conducted on BOSS-PPG-LAN, while those for (e), (f), (g) and (h) were conducted on BOSS-AHD-LAN.

From the perspective of color image steganography, it is no doubt that CMD-C better resisted color image steganalysis. However, the superiority of WISERNet was also more obvious when used to attack CMD-C steganography. For CMD-C-

HILL with 0.4 bpc on BOSS-AHD-LAN dataset, WISERNet could further increase detection accuracy by as large as 4% on the basis of Xu's model #2 (with 30 fixed SRM kernels). The more obvious performance improvement for CMD-C implies

TABLE V

Full comparison of testing accuracy of our proposed WISERNet with state-of-the-art steganalyzers in the literature. CMD-C-HILL stego images with 0.2 bpc and 0.4 bpc were included. The results on BOSS-PPG-BIC, BOSS-PPG-BIL, BOSS-AHD-BIC and BOSS-AHD-BIL are given. The results on BOSS-PPG-LAN are also listed here for reference. The best results for 0.2 bpc are underlined, while the best results for 0.4 bpc are in framed boxes.

| Datasets | CRM | | Ye's model | | | | | | Xu's model #2 (with 30 fixed SRM kernels) | | | | | | Our proposed | |
| | | | Channel-wise convolution | | Normal convolution | | Input concatenation | | Channel-wise convolution | | Normal convolution | | Input concatenation | | | |
	0.2 bpc	0.4 bpc	0.2 bpc	0.4 bpc	0.2 bpc	0.4 bpc	0.2 bpc	0.4 bpc	0.2 bpc	0.4 bpc	0.2 bpc	0.4 bpc	0.2 bpc	0.4 bpc	0.2 bpc	0.4 bpc
BOSS-PPG-LAN	0.6325	0.7548	0.6567	0.7965	0.6474	0.7568	0.6187	0.7169	0.6962	0.7941	0.6562	0.7763	0.7083	0.8167	0.7139	0.8361
BOSS-PPG-BIC	0.6551	0.7802	0.6741	0.8069	0.6589	0.7732	0.6332	0.7335	0.7124	0.8068	0.6611	0.7895	0.7294	0.8289	0.7318	0.8435
BOSS-PPG-BIL	0.7556	0.8717	0.7859	0.9019	0.7611	0.8721	0.7445	0.8343	0.7872	0.9045	0.7487	0.863	0.7904	0.9093	0.8033	0.9169
BOSS-AHD-BIC	0.6597	0.7832	0.6711	0.7991	0.6614	0.7728	0.6374	0.7355	0.7105	0.8141	0.6627	0.7922	0.7276	0.8198	0.7369	0.8448
BOSS-AHD-BIL	0.7578	0.8728	0.7804	0.9019	0.7622	0.8738	0.7376	0.837	0.7857	0.9067	0.7647	0.8593	0.7933	0.9061	0.8022	0.9144

TABLE VI

Comparison of testing accuracy of our proposed WISERNet with state-of-the-art steganalyzers in the literature under the scenario of mixed datasets. CMD-C-HILL stego images with 0.2 bpc and 0.4 bpc were included. The best results for 0.2 bpc are underlined, while the best results for 0.4 bpc are in framed boxes. Abbreviations enclosed in braces indicates different involved datasets. For instance, BOSS-PPG-{LAN, BIL, BIC} indicates that the mixed dataset is composed of BOSS-PPG-LAN, BOSS-PPG-BIL, and BOSS-PPG-BIC.

| Mixture of datasets | CRM | | Ye's model (with channel-wise convolution) | | Xu's model #2 (with 30 fixed SRM kernels, and input concatenation) | | Our proposed | |
	0.2 bpc	0.4 bpc	0.2 bpc	0.4 bpc	0.2 bpc	0.4 bpc	0.2 bpc	0.4 bpc
BOSS-PPG-{LAN, BIL, BIC}	0.6951	0.8136	0.7417	0.8379	0.7566	0.8497	0.7658	0.8674
BOSS-AHD-{LAN, BIL, BIC}	0.6973	0.8193	0.7422	0.8402	0.7517	0.8516	0.7649	0.8737
BOSS-{PPG, AHD}-LAN	0.668	0.7878	0.6844	0.8031	0.7055	0.8118	0.725	0.8411
Mixture of all six datasets	0.6926	0.8101	0.7388	0.8401	0.7586	0.8516	0.7633	0.8622

TABLE VII

Comparison of number of parameters and computational complexity for our proposed WISERNet and other state-of-the-art deep-learning based steganalyzers. The computational complexity is measured in terms of FLOPs (floating-point operations).

	Our proposed WISERNet	Xu's model #2 11-layer version (with 30 fixed SRM kernels, and input concatenation)	Ye's model (with channel-wise convolution)	Xu's model #2 (with 30 fixed SRM kernels, and input concatenation)
Parameters	2.12×10^6	2.66×10^6	1.07×10^6	4.87×10^6
FLOPs	4.11×10^9	1.08×10^{10}	5.76×10^9	1.68×10^{10}

that compared with Xu's model #2, the specific shallow-and-wide, separate-then-reunion structure of WISERNet can better utilize the intrinsic statistical characteristics among color bands to attack CMD-C.

For the sake of completeness, for CMD-C-HILL, we give a full comparison of testing accuracy of our proposed WISERNet with CRM, Ye's model and Xu's model #2 in Tab. V. The results for two representative payloads, 0.2 bpc and 0.4 bpc are given. The impacts of another two down-sampling algorithms, "Bicubic", and "Bilinear" are inspected. Therefore the results on BOSS-PPG-BIC, BOSS-PPG-BIL, BOSS-AHD-BIC and BOSS-AHD-BIL are listed. From Tab. V we can see different down-sampling algorithms had huge impact on detection performance of the steganalyzers. The steganalyzers always achieved better performances on the images generated with "Bilinear" down-sampling algorithms. For Ye's model, channel-wise convolution in the bottom convolutional layer was always the best choice. However for Xu's model #2, with input band concatenation was the best choice. Although detection performance of WISERNet was also affected by different

down-sampling algorithms, in all scenarios it performed the best and possessed a clear margin of superiority.

In Tab. VI, we also give a comparison of testing accuracy of the four steganalyzers under the scenario of mixed datasets for CMD-C-HILL. In each experiment, we used the same pseudo-random number series to split every involved dataset into training set, validation set (for deep-learning steganalyzers), and testing set. The corresponding subsets were merged together to form the mixed training set, validation set and testing set, respectively. Therefore we guaranteed that all the cover-stego pairs generated from the same raw image can only be included in one of the mixed subsets, e.g. in the training set. Firstly we fixed the demosaicking algorithm, and mixed the datasets with three different down-sampling options, namely with "Lanczos" kernel of ImageMagick, "Bicubic" and "Bilinear" kernels of Matlab®. Then we fixed the down-sampling option to "Lanczos" and mixed the datasets with PPG and AHD demosaicking algorithms. Finally we mixed the datasets with two demosaick options and three down-sampling options. From Tab. VI we can see the effects of

different demosaicking options and down-sampling options on detection performance of the investigated steganalyzers are similar. But it is amazing that in all of those scenarios our proposed WISERNet always performed the best.

The above experimental results indicates that our proposed WISERNet is with superior performance especially when used to attack CMD-C, the latest steganographic algorithm purposely for color images. The superiority of WISERNet is obvious even compared with the most advanced and complicated deep-learning steganalyzer, namely Xu's model #2. Please note that the detection performance gain is not achieved with deeper, larger or more complicated deep-learning structure. As shown in Tab. VII, WISERNet is only with less than half parameters and about a quarter computational complexity compared with Xu's model #2 (with fixed 30 SRM kernels), the second best performing deep-learning steganalyzer after WISERNet. If we would like to make a fairer comparison, the 11-layer version of Xu's model #2 is a suitable rival, which was also investigated in [30]. The 11-layer version of Xu's model #2 possesses slightly more parameters and more than double computational complexity compared with our proposed WISERNet. But as shown in Tab. VIII, the superiority of WISERNet is obvious. For instance, when the payload of the target CMD-C is 0.4bpc, WISERNet can surpass the 11-layer version of Xu's model #2 by more than four percent.

C. Impact of different components for WISERNet and its deep-learning based competitors

In Tab. IX we further compare the impact of different configurations of the bottom convolutional layer on detection performances of our proposed WISERNet, and another two deep-learning based competitors. From Tab. IX we can see for our proposed WISERNet, channel-wise convolution with learnable weights always achieved the best performance. The results also show that channel-wise convolution with learnable weights was more suitable for Ye's model. Besides, we can clearly observe that with cross-band interleave pre-processing strategy, all deep-learning based steganalyzers suffer severe performance degradation.

Xu's model #2 is the deepest and the most advanced deep-learning steganalyzer among the competitors. However it is designed for gray-scale JPEG image steganalysis, and the 16 4×4 DCT high-pass filters in the bottom convolutional layer adopted in its original version might not suitable for spatial-domain color image steganalysis. Tab. IX clearly shows that the alternative version of Xu's model #2 (in which the 16 DCT filters are replaced with 30 SRM kernels) performed better than the original version. It is noteworthy that with fixed kernel weights, input band concatenation before the bottom convolutional layer is the best option for Xu's model #2. It may be attributed to the very deep structure of Xu's model #2, which makes the convolution kernels in the top layers finally have the opportunity to perceive the cross-band correlation in pixels.

Tab. X shows the impact of different model magnification factor n on our proposed WISERNet. As shown in Tab. X, starting from $n = 1$, detection performance was steadily promoted along with increasing n, and reached its maximum with $n = 9$. Therefore, n was fixed to 9 in our experiments.

D. Imapct of learnable bottom kernels

Firstly, we try to give an explanation why making the weights of the bottom convolution kernels of WISERNet learnable can further improve detection performance.

We agree with the opinion that the main purpose of the convolutions in the bottom convolutional layer is to suppress image contents and retain stego noises at the same time [20], [21], [24], [25]. However, as pointed out in [27], optimization of the bottom convolution kernels in favor of the extraction of stego noises is hard to achieve with gradient-descent based learning. In fact, as mentioned in Sect. II-A, fixed high-pass filters in the bottom convolutional layer of most existing deep-learning steganalyzers also provide evidences to support our argument.

Since the proposal of rich models [8], it is well accepted that model diversity is crucial to the performance of steganalyzers. Therefore we believe the performance improvement of WISERNet can be attributed to the further model diversity brought by continuous learning and optimizing of the kernels in the bottom convolutional layer of WISERNet, although learnable bottom kernels cannot help boost SNR. Refer to Sect. II-B, the bottom convolution kernels of WISERNet can be divided into triples $\{\mathbf{W}_{1k}^1, \mathbf{W}_{2k}^1, \mathbf{W}_{3k}^1\}, 1 \le k \le 30$. Denote the average of pair-wise correlation between weights in the triples as:

$$\overline{C_W} = \frac{\sum_{k=1}^{30} \sum_{1 \le i < j \le 3} \text{Corr}(\mathbf{W}_{ik}^1, \mathbf{W}_{jk}^1)}{90} \quad (14)$$

$\overline{C_W}$ can be used to measure the diversity of the bottom kernels. Initially, we set $\mathbf{W}_{1k}^1 = \mathbf{W}_{2k}^1 = \mathbf{W}_{3k}^1 = \widetilde{\mathbf{W}}_k$, where $\widetilde{\mathbf{W}}_k$ is one of the SRM high-pass filters. Therefore initially $\overline{C_W} = 1$, and it decreases along with increasing diversity of the bottom kernels. In Tab. XI, we give a demonstration. In one training procedure of our proposed WISERNet which aimed at attacking CMD-C-HILL stego images with 0.4 bpc, we could clearly inspect $\overline{C_W}$ steadily decreased with increasing training iterations, which indicates diversity of the bottom kernels increased along with increasing training iterations. $\overline{C_W}$ reached its minimum at around 20×10^4 iterations. It is interesting that WISERNet also achieved its best validation accuracy at 20×10^4 iterations, which implies that the performance of WISERNet was relevant with diversity of the bottom kernels. However, please note that initialized with high-pass filters, the bottom kernels of WISERNet eventually cannot exhibit large diversity even with enormous iterations. In our extensive experiments, we have never observed $\overline{C_W}$ was reduced to below 0.9 even after 100×10^4 iterations. At last, one more remarkable thing is that Xu's model #2, the one with much deeper structure could not gain better performance with learnable bottom kernels. Therefore we believe that the wide and shallow structure of WISERNet might be the determining factor of beneficial learnable bottom convolution kernels.

As mentioned in Sect. II-B, our basic assumption is that $\text{E}(\mathbf{C}_1 * \mathbf{W}_{1k}^1) \approx \text{E}(\mathbf{C}_2 * \mathbf{W}_{2k}^1) \approx \text{E}(\mathbf{C}_3 * \mathbf{W}_{3k}^1)$. One interesting question arises: Is this assumption still holds with more and

TABLE VIII

COMPARISON OF TESTING ACCURACY OF OUR PROPOSED WISERNET WITH THE 11-LAYER VERSION OF XU'S MODEL #2 (WITH 30 FIXED SRM KERNELS). THE EXPERIMENTS WERE
CONDUCTED ON BOSS-PPG-LAN. THE TERMS IN PARENTHESES WITH PRECEDING ↑ DENOTE ACCURACY INCREMENT OF OUR PROPOSED WISERNET COMPARED TO THE 11-LAYER
VERSION OF XU'S MODEL #2.

Steganalyzers	Payload (bpc)				
	0.1	0.2	0.3	0.4	0.5
Our proposed WISERNet	0.6329 (↑**0.0321**)	0.7139 (↑**0.0364**)	0.7767 (↑**0.0376**)	0.8361 (↑**0.0412**)	0.8748 (↑**0.0433**)
11-layer version of Xu's model #2	0.6008	0.6775	0.7391	0.7949	0.8315

TABLE IX

IMPACT OF DIFFERENT CONFIGURATIONS OF THE BOTTOM CONVOLUTIONAL LAYER. THE EXPERIMENTS WERE CONDUCTED ON BOSS-PPG-LAN. ONLY CMD-C-HILL STEGO
IMAGES WITH 0.4 BPC WERE INCLUDED. THE BEST RESULT IN EVERY COLUMN IS UNDERLINED. THE UNDERLINE IN BOLD HIGHLIGHTS THE BEST RESULT AMONG THEM.

How to convolve	Learnable	Our proposed network	Compared deep-learning steganalyzers		
			Ye's model	Xu's model #2 with 16 DCT kernels	Xu's model #2 with 30 SRM kernels
Channel-wise convolution	✓	0.8361	0.7965	0.7244	0.7941
	✗	0.8232	0.7895	0.7269	0.7951
Normal convolution	✓	0.7268	0.7608	0.6916	0.7725
	✗	0.7257	0.7566	0.7085	0.7763
Input concatenation	✓	0.7259	0.7319	0.7672	0.7987
	✗	0.7270	0.7268	0.7916	0.8167
Cross-band interleave	✓	0.7063	0.6935	0.7051	0.7122
	✗	0.6968	0.6848	0.6936	0.7041

TABLE X

IMPACT OF DIFFERENT MODEL MAGNIFICATION FACTOR n. THE EXPERIMENTS WERE CONDUCTED ON BOSS-PPG-LAN. ONLY CMD-C-HILL STEGO IMAGES WITH 0.4 BPC WERE
INCLUDED. THE BEST RESULT IS IN FRAMED BOX.

Model magnification factor n									
1	2	3	4	5	6	7	8	9	10
0.7375	0.8037	0.8208	0.8214	0.8227	0.8262	0.8313	0.8319	0.8361	0.8321

TABLE XI

$\overline{C_W}$, $\overline{|\overset{\circ}{S}|}_C$ AND $\overline{|\overset{\circ}{S}|}_S$ VERSUS INCREASING ITERATIONS OF OUR PROPOSED WISERNET. THE EXPERIMENT WAS CONDUCTED ON BOSS-PPG-LAN. CMD-C-HILL STEGO IMAGES
WITH 0.4 BPC WERE INCLUDED.

Iterations	$\overline{C_W}$	$\overline{|\overset{\circ}{S}|}_C$	$\overline{|\overset{\circ}{S}|}_S$	Iterations	$\overline{C_W}$	$\overline{|\overset{\circ}{S}|}_C$	$\overline{|\overset{\circ}{S}|}_S$
0×10^4	1	0.9888	0.9888	6×10^4	0.9877	0.9886	0.9886
1×10^4	0.9985	0.9888	0.9888	10×10^4	0.9876	0.9884	0.9884
2×10^4	0.9978	0.9886	0.9886	15×10^4	0.9814	0.9881	0.9881
3×10^4	0.9978	0.9886	0.9886	20×10^4	0.9753	0.9889	0.9889
4×10^4	0.9975	0.9890	0.9890	25×10^4	0.9795	0.9886	0.9886
5×10^4	0.9967	0.9891	0.9891	30×10^4	0.9848	0.9887	0.9887

more diverse bottom kernels? For a given image, denote $E(\mathbf{C}_i * \mathbf{W}_{ik}^1) = \mu_{ik}, 1 \leq i \leq 3$. Let $\boldsymbol{\mu}_k = (\mu_{1k}, \mu_{2k}, \mu_{3k})$, $\mathbf{1} = (1, 1, 1)$, and θ_k be the angle between vector $\boldsymbol{\mu}_k$ and vector $\mathbf{1}$. The *cosine similarity* [40] between $\boldsymbol{\mu}_k$ and $\mathbf{1}$ is defined as:

$$S_k = \cos(\theta_k) = \frac{\mu_{1k} + \mu_{2k} + \mu_{3k}}{\sqrt{3} \cdot \sqrt{\mu_{1k}^2 + \mu_{2k}^2 + \mu_{3k}^2}} \quad (15)$$

However, since $\mathbf{W}_{ik}^1, 1 \leq i \leq 3$ are high-pass filters, $\mu_{ik}, 1 \leq i \leq 3$ may be less than zero. As a result, S_k ranges from -1 to 1. For simplicity we use $|S_k|$ to measure the similarity among

$\mu_{ik}, 1 \leq i \leq 3$. The more similar $\mu_{ik}, 1 \leq i \leq 3$ are, the more parallel $\boldsymbol{\mu}_k$ and $\mathbf{1}$ tend to be , the more $|S_k|$ is close to 1.

Fed with one image, let $\overline{|S|}$ denote the average of $|S_k|, 1 \leq k \leq 30$. In Tab. XI, we also analyzed the 1,000 cover-stego pairs in the validation set of BOSS-PPG-LAN. Denote the average of $\overline{|S|}$ of the 1,000 cover images as $\overline{|\overset{\circ}{S}|}_C$, and that of the 1,000 stego images as $\overline{|\overset{\circ}{S}|}_S$. From Tab. XI we can see $\overline{|\overset{\circ}{S}|}_C$ was always equal to $\overline{|\overset{\circ}{S}|}_S$ which means that the impact of stego noises was negligible. Though $\overline{C_W}$ slowly but steadily

decreased with increasing training iterations, both $\overline{|\overset{\circ}{S}|}_C$ and $\overline{|\overset{\circ}{S}|}_S$ kept close to 1. Therefore the experimental evidence indicated that for our proposed WISERNet, the means of the corresponding output channels are still nearly equivalent even with more and more diverse bottom kernels during the training procedure. As a result, our basic assumption in Sect. II-B holds.

E. Impact of the correlations among color bands of the targets

Our major target, CMD-C is a non-additive embedding distortion minimizing framework which can preserve not only the correlation within each color band, but also the correlations among three color bands. It is interesting to observe how the correlations among color bands of CMD-C stego images affect the performance of our proposed WISERNet. We can disable the inter-band correlations in CMD-C stego images via removing elements from other bands in the calculation of the costs [7]. As shown in Tab. XII, when the target is the alternative version of CMD-C in which the correlation among three color bands is disabled, our proposed WISERNet achieves better performance. Furthermore, the gap between Xu's model #2 and WISERNet is wider than when the target is the original CMD–HILL. The wider gap implies that inter-band correlation of stego noises introduced by original CMD-C (even weak) do help it better resist the channel-wise convolution in the bottom "Separate" stage of WISERNet. Our proposed wide-and-shallow, separate-then-reunion network structure shows even greater advantage when used to attack CMD-C without the correlation among three color bands.

F. Performance on large-scale dataset under cover-source mismatching scenarios

We collected another 100,000 diverse raw images and followed the dataset generating process as mentioned in Sect. III-A (with PPG demosaicking algorithm and "Lanczos" down-sampling kernel) to construct a new dataset SZUBASE-PPG-LAN. Fig. 3 is devoted to the comparison of performance of our proposed WISERNet with other state-of-the-art steganalyzers on SZUBASE-PPG-LAN, under cover-source mismatching scenarios. From Fig. 3 we can see that the impact of varing demosaicking algorithms is moderate, while the impact of varing down-sampling kernels is more obvious. However, under such cover-source mismatching scenarios, the impacts on the steganalyzers, either hand-crafted or deep-learning based (including our proposed WISERNet), are similar. Our proposed WISERNet is still the one with the best performance.

IV. Concluding remarks

Along with the arise of steganographic algorithms purposely for color spatial images, the corresponding requirement for powerful color image steganalysis becomes more compelling. In this paper we propose WISERNet, the wider separate-then-reunion network for steganalysis of color images. The major contributions of this work are as follows:

- We have provided theoretical rationale to claim that the summation in normal convolution actually impairs the signal-to-noise ratio, which collides with the main purpose of the bottom convolutional layer.
- We have pointed out that the summation in normal convolution is a "linear collusion attack" which is a double-edged sword for color image steganalysis. Accordingly we have proposed WISERNet, the wider separate-then-reunion network for steganalysis of color images.
- We have conducted extensive experiments on image datasets with different demosaicking and down-sampling opinions. The experimental results demonstrated the superiority of our proposed WISERNet.

Our future work will focus on two aspects: (1) making WISERNet capable of identifying suspicious images under more complex cover source mismatching scenarios; (2) further developing deep-learning architectures suitable for large-scale JPEG color image steganalysis on the basis of WISERNet.

References

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

[2] B. Li, M. Wang, J. Huang, and X. Li, "A new cost function for spatial image steganography," in *Proc. IEEE 2014 International Conference on Image Processing, (ICIP'2014)*, 2014, pp. 4206–4210.

[3] V. Sedighi, R. Cogranne, and J. Fridrich, "Content-adaptive steganography by minimizing statistical detectability," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 2, pp. 221–234, 2016.

[4] J. Fridrich and T. Filler, "Practical methods for minimizing embedding impact in steganography," in *Proc. SPIE, Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents IX*, vol. 6505, 2007, pp. 650 502–1–650 502–15.

[5] T. Denemark and J. Fridrich, "Improving steganographic security by synchronizing the selection channel," in *Proc. 3rd ACM Information Hiding and Multimedia Security Workshop (IH&MMSec' 2015)*, 2015, pp. 5–14.

[6] B. Li, M. Wang, X. Li, S. Tan, and J. Huang, "A strategy of clustering modification directions in spatial image steganography," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 9, pp. 1905–1917, 2015.

[7] W. Tang, B. Li, W. Luo, and J. Huang, "Clustering steganographic modification directions for color components," *IEEE Signal Processing Letters*, vol. 23, no. 2, pp. 197–201, 2016.

[8] 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.

[9] V. Holub and J. Fridrich, "Random projections of residuals for digital image steganalysis," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 12, pp. 1996–2006, 2013.

[10] J. Kodovský and J. Fridrich, "Ensemble classifiers for steganalysis of digital media," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 432–444, 2012.

[11] M. Goljan, J. Fridrich, and R. Cogranne, "Rich model for steganalysis of color images," in *Proc. IEEE 2012 International Workshop on Information Forensic and Security (WIFS'2014)*, 2014, pp. 185–190.

[12] M. Goljan and J. J. Fridrich, "CFA-aware features for steganalysis of color images," in *Proc. IS&T/SPIE Electronic Imaging 2015 (Media Watermarking, Security, and Forensics)*, 2015, pp. 94 090V–1–94 090V–13.

[13] H. Abdulrahman, M. Chaumont, P. Montesinos, and B. Magnier, "Color image stegananalysis using correlations between RGB channels," in *Proc. IEEE 10th International Conference on Availability, Reliability and Security (ARES'2015)*, 2015, pp. 448–454.

[14] ——, "Color images steganalysis using RGB channel geometric transformation measures," *Security and communication networks*, vol. 9, no. 15, pp. 2945–2956, 2016.

[15] ——, "Color image steganalysis based on steerable Gaussian filters bank," in *Proc. 4th ACM Information Hiding and Multimedia Security Workshop (IH&MMSec'2016)*, 2016, pp. 109–114.

TABLE XII
COMPARISON OF TESTING ACCURACY WHEN THE TARGETS ARE CMD-C-HILL AND ITS ALTERNATIVE VERSION IN WHICH THE CORRELATION AMONG THREE COLOR BANDS IS DISABLED. THE EXPERIMENTS WERE CONDUCTED ON BOSS-PPG-LAN. THE TERMS IN PARENTHESES WITH PRECEDING ↑ DENOTE ACCURACY INCREMENT OF OUR PROPOSED WISERNET COMPARED TO XU'S MODEL #2 (WITH 30 FIXED SRM KERNELS).

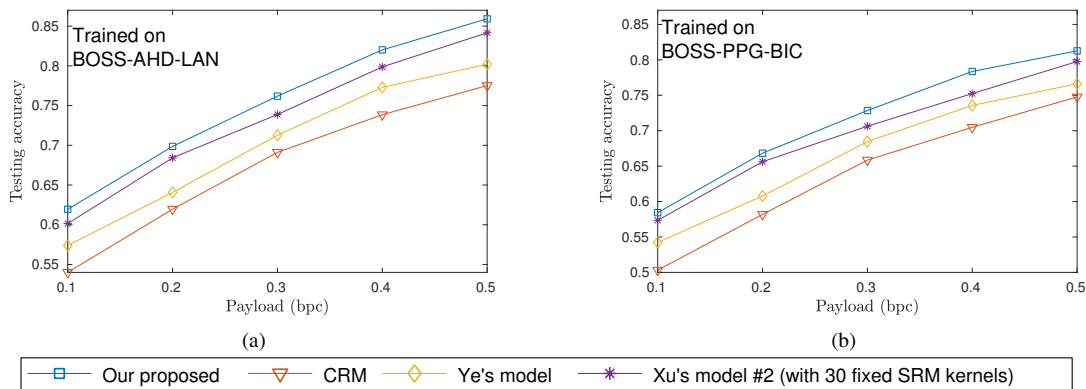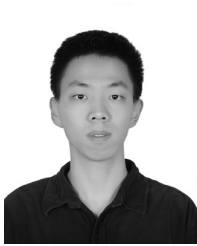| Steganalyzers	Payload (bpc)				
	0.1	0.2	0.3	0.4	0.5
For CMD-C-HILL					
Our proposed WISERNet	0.6329 (↑0.0165)	0.7139 (↑0.0122)	0.7767 (↑0.0193)	0.8361 (↑0.0194)	0.8748 (↑0.0152)
Xu's model #2 (with 30 fixed SRM kernels)	0.6164	0.7017	0.7574	0.8167	0.8596
For the alternative version of CMD-C-HILL (with the correlation among three color bands disabled)					
Our proposed WISERNet	0.6509 (↑0.0172)	0.7584 (↑0.0242)	0.7985 (↑0.0247)	0.8598 (↑0.0266)	0.8934 (↑0.0177)
Xu's model #2 (with 30 fixed SRM kernels)	0.6337	0.7342	0.7738	0.8332	0.8757

Fig. 3. Comparison of performance of our proposed WISERNet with other state-of-the-art steganalyzers under cover-source mismatching scenario. The target is CMD-C-HILL. (a) Trained on BOSS-AHD-LAN, while tested on SZUBASE-PPG-LAN; (b) Trained on BOSS-PPG-BIC, while tested on SZUBASE-PPG-LAN.

[16] T. Denemark, V. Sedighi, V. Holub, R. Cogranne, and J. Fridrich, "Selection-channel-aware rich model for steganalysis of digital images," in *Proc. 6th IEEE International Workshop on Information Forensic and Security (WIFS'2014)*, 2014, pp. 48–53.

[17] W. Tang, H. Li, W. Luo, and J. Huang, "Adaptive steganalysis based on embedding probabilities of pixels," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 4, pp. 734–745, 2016.

[18] M. Boroumand and J. Fridrich, "Applications of explicit non-linear feature maps in steganalysis," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 4, pp. 823–833, 2018.

[19] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.

[20] S. Tan and B. Li, "Stacked convolutional auto-encoders for steganalysis of digital images," in *Proc. Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA'2014)*, 2014, pp. 1–4.

[21] Y. Qian, J. Dong, W. Wang, and T. Tan, "Deep learning for steganalysis via convolutional neural networks," in *Proc. IS&T/SPIE Electronic Imaging 2015 (Media Watermarking, Security, and Forensics)*, 2015, pp. 94 090J–1–94 090J–10.

[22] L. Pibre, P. Jérôme, 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 *Proc. Media Watermarking, Security, and Forensics, Part of IS&T International Symposium on Electronic Imaging (EI'2016)*, 2016, pp. 1–11.

[23] Y. Qian, J. Dong, W. Wang, and T. Tan, "Learning and transferring representations for image steganalysis using convolutional neural network," in *Proc. IEEE 2016 International Conference on Image Processing, (ICIP'2016)*, 2016, pp. 2752–2756.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] M. Chen, V. Sedighi, M. Boroumand, and J. Fridrich, "JPEG-phase-aware 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.

[29] 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.

[30] 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.

[31] K. Su, D. Kundur, and D. Hatzinakos, "Statistical invisibility for collusion-resistant digital video watermarking," *IEEE Transactions on Multimedia*, vol. 7, no. 1, pp. 43–51, 2005.

[32] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. International Conference on Machine Learning (ICML' 2015)*, 2015, pp. 448–456.

[33] T. Filler and J. Fridrich, "Gibbs construction in steganography," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 4, pp. 705–720, 2010.

[34] R. C. Gonzalez and R. E. Woods, *Digital Image Processing (3rd Ed)*. Prentice Hall, New York, 2008, p. 354.

[35] G. Carey, "Linear transformations and linear composites," http://psych.colorado.edu/~carey/Courses/PSYC7291/handouts/transformations.pdf, accessed: 2018-02-06.

[36] T. Pevný, P. Bas, and J. Fridrich, "Steganalysis by subtractive pixel adjacency matrix," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 2, pp. 215–224, 2010.

[37] T. Pevný and J. Fridrich, "Benchmarking for steganography," in *Proc. 10th Information Hiding Workshop (IH'2008)*, 2008, pp. 251–267.

[38] P. Bas, T. Filler, and T. Pevný, "Break our steganographic system—the ins and outs of organizing BOSS," in *Proc. 13th Information Hiding Workshop (IH'2011)*, 2011, pp. 59–70.

[39] Y. Jia *et al.*, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. 22nd ACM International Conference on Multimedia*, 2014, pp. 675–678.

[40] "Cosine similarity — Wikipedia, the free encyclopedia," https://en.wikipedia.org/wiki/Cosine_similarity, accessed: 2018-02-06.

Jiwu Huang (M'98–SM'00–F'16) received the B.S. degree from Xidian University, Xian, China, in 1982, the M.S. degree from Tsinghua University, Beijing, China, in 1987, and the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, in 1998. He is currently a Professor with the College of Information Engineering, Shenzhen University, Shenzhen, China. His current research interests include multimedia forensics and security. He is a member IEEE Signal Processing Society Information Forensics and Security Technical Committee and serves as an Associate Editor for the IEEE Transactions on Information Forensics and Security. He was a General Co-Chair of the IEEE Workshop on Information Forensics and Security in 2013 and a TPC Co-Chair of the IEEE Workshop on Information Forensics and Security in 2018.

Jishen Zeng (S'16) received the B.S degree of electronic information science and technology from Sun Yat-sen University, Guangzhou, China in 2015. He is currently a Ph.D. student in Shenzhen University majoring in information and communication engineering. His current research interests include steganography, steganalysis, multimedia forensics, and deep learning.

Shunquan Tan (M'10–SM'17) received the B.S. degree in computational mathematics and applied software and the Ph.D. degree in computer software and theory from Sun Yat-sen University, Guangzhou, China, in 2002 and 2007, respectively.

He was a Visiting Scholar with New Jersey Institute of Technology, Newark, NJ, USA, from 2005 to 2006. He is currently an Associate Professor with College of Computer Science and Software Engineering, Shenzhen University, China, which he joined in 2007. His current research interests include multimedia security, multimedia forensics, and machine learning.

Guangqing Liu received the B.S degree of computer science and technology from Shenzhen University, Shenzhen, China in 2018. His current research interests include information hiding, multimedia forensics, and deep learning.

Bin Li (S'07–M'09–SM'17) received the B.E. degree in communication engineering and the Ph.D. degree in communication and information system from Sun Yat-sen University, Guangzhou, China, in 2004 and 2009, respectively. He was a Visiting Scholar with New Jersey Institute of Technology, Newark, NJ, USA, from 2007 to 2008. He is currently an Associate Professor with Shenzhen University, Shenzhen, China, where he joined in 2009. He is the director of Shenzhen Key Laboratory of Media Security. He is also a scholar with Peng Cheng Laboratory. He is now a member of IEEE Information Forensic and Security Technical Committee (IFS-TC).

His current research interests include image processing, multimedia forensics, and pattern recognition.