High-Capacity Robust Image Steganography via Adversarial Network

Beijing Chen[1,2,3,4]*, Jiaxin Wang[1], Yingyue Chen[5], Zilong Jin[1], Hiuk Jae Shim[1], and Yun-Qing Shi[6]
[1] School of Computer & Software, Nanjing University of Information Science & Technology, Nanjing 210044, China
[2]Engineering Research Center of Digital Forensics, Ministry of Education, Nanjing University of Information Science & Technology, Nanjing 210044, China
[3]Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAEET), Nanjing University of Information Science & Technology, Nanjing 210044, China
[4] Key Laboratory of Computer Network Technology of Jiangsu Province, Southeast University, Nanjing 210096, China
[5]School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China
[6] Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark 07102, USA
*Corresponding author: Beijing Chen [e-mail: nbutimage@126.com.]

Received October 14, 2019; accepted December 18, 2019; published January 31, 2020

Abstract

Steganography has been successfully employed in various applications, e.g., copyright control of materials, smart identity cards, video error correction during transmission, etc. Deep learning-based steganography models can hide information adaptively through network learning, and they draw much more attention. However, the capacity, security, and robustness of the existing deep learning-based steganography models are still not fully satisfactory. In this paper, three models for different cases, i.e., a basic model, a secure model, a secure and robust model, have been proposed for different cases. In the basic model, the functions of high-capacity secret information hiding and extraction have been realized through an encoding network and a decoding network respectively. The high-capacity steganography is implemented by hiding a secret image into a carrier image having the same resolution with the help of concat operations, InceptionBlock and convolutional layers. Moreover, the secret image is hidden into the channel B of carrier image only to resolve the problem of color distortion. In the secure model, to enhance the security of the basic model, a steganalysis network has been added into the basic model to form an adversarial network. In the secure and robust model, an attack network has been inserted into the secure model to improve its robustness further. The experimental results have demonstrated that the proposed secure model and the secure and robust model have an overall better performance than some existing high-capacity deep learning-based steganography models. The secure model performs best in invisibility and security. The secure and robust model is the most robust against some attacks.

Keywords: Steganography, steganalysis, high-capacity, robustness, adversarial network

This work was supported by the NSFC under Grants 61572258, 61602252, and 61772281, the PAPD fund, and the Qing Lan Project of Jiangsu Higher Education Institutions.

http://doi.org/10.3837/tiis.2020.01.020 **ISSN : 1976-7277**

1. Introduction

Steganography involves hiding secret information in an appropriate multimedia carrier, e.g., text, image, audio, and video files, in imperceptible ways such that no one even suspects the existence of the secret information except the sender and receiver [1， 2]. Moreover, the receiver can accurately extract the secret information. Steganography is mainly for covert communication. It is also employed in other applications, e.g., copyright control of materials, smart identity cards where individuals' details are embedded in their photographs, video error correction during transmission [3, 4], privacy protection of authorized people in surveillance system [5], TCP/IP packets (for instance a unique ID can be embedded into an image to analyze the network traffic of particular users) [6]. In particular, the Japanese firm Fujitsu is perfecting steganography to encode data into a picture that is invisible to the human eye but can be decoded by a cell phone camera [7]. This application can be used in doctor's prescriptions, food wrappers, billboards, and business cards. So, steganography plays an important role in people's daily life.

Image steganography is the most popular one in multimedia steganography [8, 9]. Traditional image steganography algorithms can be divided into two categories: spatial-domain models and transform-domain ones. The spatial-domain models mainly hide secret information by changing the brightness value or chrominance value of carrier image. Least significant bit (LSB) model is a simple and representative spatial-domain model. It first converts secret information into binary data, and then replaces the LSBs of some pixels in the carrier image [10, 11]. The transform-domain models need to apply a transform to the carrier image first, and then modify transform-domain coefficients to embed secret information. The commonly used transforms are discrete Fourier transform [12], discrete cosine transform [13], discrete wavelet transform [14], and random transform [15], etc.

However, both spatial-domain and transform-domain models are hand-crafted ones. They consider the positions and strengths of hiding information manually. Moreover, for a given carrier image and secrete information, it is difficult to decide which domain or which transform is the optimal one [16]. Deep learning is an important branch of machine learning. It aims to automatically learn useful and highly abstract features of data by simulating human brain. Therefore, it can learn the basic characteristics of data better than the traditional machine learning methods [17]. So, deep learning has become a common and significant tool in computer vision and other related fields since 2006 [18]. So does the field of steganography. The deep learning-based steganography models [19-28] use an encoding network for steganography and a decoding network for extracting secret information. All the positions and strengths of hiding information as well as hiding domain are automatically achieved by training the networks. However, these works have one or more of the following shortcomings: 1) the colors of the generated steganographic images are distorted [26]; 2) the steganography model does not consider both security against steganalysis and robustness against some attacks during the training phase [26, 27] or only considers security [19-25, 28]. However, security and robustness are also very important in some real applications, where the secret information needs to be correctly extracted from steganographic images subjected to some attacks; 3) hiding capacity is very limited in [19-25]. However, sometimes we need to hide a secret image into a natural carrier image for secret image transmission. Therefore, three models (basic model, secure model, secure and robust model) are proposed in this paper to consider all invisibility, security, and robustness while hiding high-capacity information.

The main contributions of our work are as below: 1) the color distortion problem in [26] is resolved by hiding information in only channel B of color carrier image; 2) all of invisibility,

security, and robustness are considered when hiding high-capacity information in the training phase; 3) the secret image can be still extracted after some attacks. The remaining of the paper is organized as follows. In Section 2, the related works are recalled. In Section 3, three proposed models are presented. The experimental results and analysis are given in Section 4. Finally, the paper is summarized in Section 5.

2. Related Works

In this section, the related works of steganography based on deep learning is first recalled. Since the proposed steganography models in this paper combine the steganalysis based on XuNet, this section also describes the steganalysis based on XuNet.

2.1 Steganography Based on Deep Learning

The steganography models based on deep learning are basically realized by adversarial networks. Volkhonskiy et al. [19] first proposed the steganography model called Steganographic Generative Adversarial Network (SGAN) based on GAN. This model resisted steganalysis and made hidden information secure. Based on SGAN, Shi et al. [20] proposed SSGAN (Secure Steganography Based on GAN) to enhance security against steganalysis. The HayesGAN model proposed by Hayes et al. [21] used adversarial learning to generate steganographic images directly. Then, Hu et al. [22] introduced a steganalysis network into HayesGAN to improve the quality and safety of the generated steganographic images. However, these two works are not guaranteed to extract embedded secret information completely. Zhu et al. [23] proposed another model called Hiding Data With Deep Network (HiDDeN) based on HayesGAN. It can extract embedded information with high accuracy under various attacks, such as Gaussian blur, missing pixels, cropping, and JPEG compression, etc. Tang et al. [24] combined GAN with an adaptive steganography model to find suitable steganographic positions for steganography and proposed Automatic Steganographic Distortion Learning Framework with GAN (ASDL-GAN). Yang et al. [25] modified the ASDL-GAN model by replacing the activation function Ternary Embedding Simulator (TES) with Tanh to improve security.

Although the above-mentioned adversarial network-based steganography models achieve fine performance in steganography and resisting steganalysis, their hiding capacities are very limited. Accordingly, some researchers [26-28] proposed high-capacity steganography models to embed secret images into carrier images with the same resolution. Rehman et al. [26] proposed an end-to-end framework to embed a secret gray image into a color carrier image with the same resolution. Their work realizes high-capacity embedding but distorts the color information of steganographic images. Then, in the loss function, Baluja et al. [27] considered the correlation between a secret image and an error image obtained from a steganographic image and carrier image to improve the invisibility of the steganographic image. Zhang et al. [28] proposed ISGAN (Invisible Steganography via Generative Adversarial Networks) by introducing the steganalysis network proposed by Xu et al. [29] into their basic model to improve its ability to resist steganalysis. In addition, the SteganoGAN model proposed by Zhang et al. [30] used residual structure to improve the quality of steganographic images further. To sum up, these high-capacity works perform well in terms of invisibility and security. However, they are not robust to various attacks. Inspired by works of Rehman [26], Zhang [28] and Zhu [23], this paper tries to consider all invisibility, security, and robustness in the steganography network while preserving high-capacity property.

2.2 Steganalysis Based on XuNet

XuNet is proposed by Xu et al. [29] in 2016. The architecture of XuNet is shown in **Fig. 1**. The network adds a fixed high-pass filter (HPF, using KV kernel) layer at the front end. Since high-frequency noise signal in steganography is symmetric around zero, XuNet uses zero-bias parameter in the first convolutional layer and adds ABS layer to reduce the range of feature map. Batch nomalization (BN) layer is added to improve convergence speed and to avoid falling into local minimum. In addition, XuNet utilizes a hybrid activation function to optimize the model: TanH activation function is used in the network front end to improve the learning ability of features, and ReLU activation function is adopted to reduce the difficulty of back-propagation. Finally, XuNet also uses 1×1 convolution kernels and global average pooling in the later layers of the network to prevent network model overfitting and feature loss. These improvements conducted in XuNet has led to great performance gains.

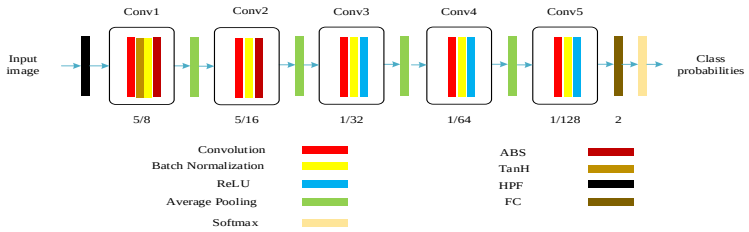


Fig. 1. Architecture of the steganalysis network proposed by *Xu et al*. [29].

3. Proposed Models

This section proposes three models for steganography. The first one is a basic model, which realizes the functions of secret image hiding and secret image extraction. The others are two enhanced models, which improve the security and robustness of the basic model.

3.1 Basic Model

An end-to-end basic model is designed to realize the functions of steganography and secret image extraction. Its main architecture is shown in **Fig. 2**. The basic model consists of two parts: encoding network and decoding network. The encoding network hides a gray secret image into channel B of a color carrier image with the same resolution to obtain a steganographic image. The decoding network is responsible for extracting the secret image from channel B of the steganographic image. Here, the reasons of choosing channel B only are that: 1) considering all channels will distort the colors of the generated steganographic images as shown in [31]; 2) human eyes are less sensitive to the B channel than the R and G channels according to human visual system [32].

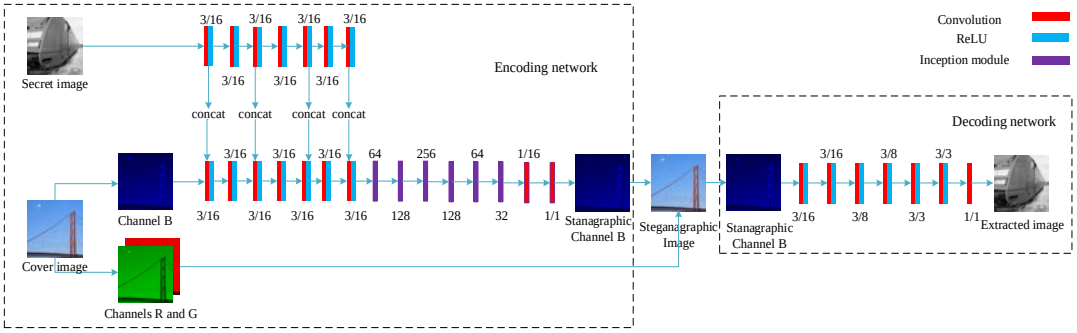


Fig. 2. Architecture of the proposed basic model.

Specifically, as for the encoding network, the gray secret image and the channel B of the color carrier image are respectively used as inputs for two parallel branches. Then, feature extraction is performed on the input image in two parallel branches through a series of convolutional layers and ReLU layers. After each two convolutional layers, the features extracted from the gray secret image branch are superimposed on the features extracted from the color carrier image branch by a total of four concat operations as the work of Rehman et al. [26]. Then, the final superimposed features are processed by six InceptionBlock [33] and two convolutional layers to hide the secret image into the B channel of the cover image completely. Here, different from Ref. [26], InceptionBlock is introduced into the encoding network because InceptionBlock can fuse feature maps with different perceptual field sizes very well by using several convolution kernels with different sizes. Finally, the new channel B is merged with R and G channels to get the final steganographic image. In the decoding network, the channel B of the steganographic image is regarded as an input, and it goes through a series of convolution layers and ReLU layers to obtain an extracted secret image.

For loss function, Rehman et al. [26] considered to minimize the distortion function of both steganographic image and extracted secret image. They used the mean square error (MSE) to evaluate the distortion and achieved a fine performance. So, this paper uses a similar idea. Let θ_E and θ_D denote the parameters of encoding network and decoding network, respectively, $E(\theta_E, c, s)$ be the output of encoding network on the carrier image c and the secret image s, and $D(\theta_D, c')$ be the output of decoding network on the steganographic image c'. Then, the encoding network's loss L_E can be summarized as,

$$\begin{aligned} L_E &= MSE(c, E(\theta_E, c, s)) \\ &= MSE(c, c'). \end{aligned} \tag{1}$$

The decoding network's loss L_D is as follow,

$$\begin{aligned} L_D &= MSE(s, D(\theta_D, c')) \\ &= MSE(s, s'). \end{aligned} \tag{2}$$

where s' represents the corresponding extracted secret image. So, the loss function of the proposed basic model can be given by,

$$L_{Bas} = L_E + L_D. \tag{3}$$

3.2 Two Enhanced Models

In order to improve the security and robustness of the proposed basic model, two enhanced models (secure model, secure and robust model) are proposed. The secure model inserts a steganalysis network into the basic model to enhance security against steganalysis, while the secure and robust model combines the secure model with an attack network to enhance robustness against some attacks further. In these two enhanced models, the encoding network can be regarded as a generator, while the steganalysis network can be seen as a discriminator. Thus, the combination of these two networks can be seen as an adversarial network. The architecture of the secure and robust model is shown in **Fig. 3**, while its attack network is presented in **Fig. 4**. The secure model only needs to remove the attack network.

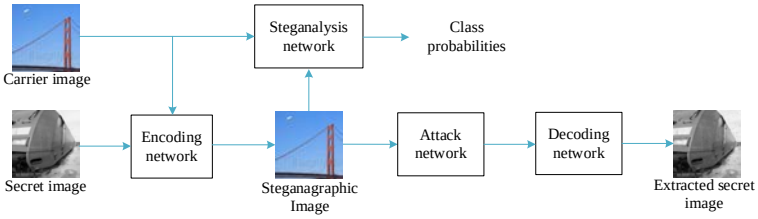


Fig. 3. Architecture of the proposed secure and robust model.

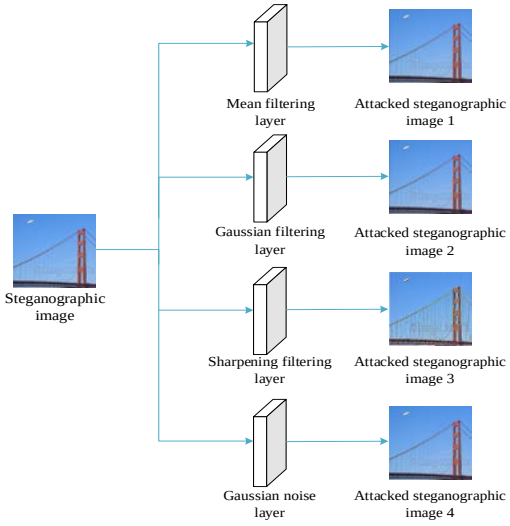


Fig. 4. Architecture of the attack network.

Specifically, as for the encoding network and decoding network, they are the same as those of the basic model given in **Fig. 1**. As for the steganalysis network, since XuNet performs well in steganalysis (please find more detail in Subsection 2.2), it is introduced into the basic model to enhance security against steganalysis. Moreover, inspired by Zhang et al [28], the global average pooling layer in XuNet is replaced by spatial pyramid pool (SPP) module because the SPP module can extract more features from different receptive fields. As for the attack network shown in **Fig. 4**, four types of attacks, i.e., mean filtering, Gaussian filtering, sharpening filtering and Gaussian noise, are simulated through network layers. Moreover,

these four types of attacks are implemented in four parallel branches. More specifically, the first three types of attacks are simulated by three convolutional layers with different convolutional kernels. The Gaussian noise attack is performed by adding a randomly generated Gaussian noise map to a steganographic image. Attack layers are inspired by the work of Zhu et al. [23].

The objectives of both two enhanced models are to optimize three networks: encoding network, decoding network and steganalysis network. So, their loss functions contain three parts corresponding to three networks.

As for encoding network, the loss functions of two models are the same as the loss function of the basic model given in (1).

As for decoding network, the loss function of the secure model is the same as that of the basic model given in (2). However, since the decoding network of the secure and robust model is followed by an attack network, the loss given in (2) should be modified as,

$$\begin{aligned}\tilde{L}_D &= \frac{1}{n}\sum_{i=1}^{n} MSE(s,\ D(\theta_D, c_i')) \\ &= \frac{1}{n}\sum_{i=1}^{n} MSE(s,\ s_i'),\end{aligned} \tag{4}$$

where n represents the number of attacks considered, c_i', $i = 1,2, \ldots, n$, is the steganographic image after attack i, s_i', $i = 1,2, \ldots, n$, is the corresponding extracted secret image.

As for steganalysis network, it is associated with the encoding network, thus the combination of them can be seen as an adversarial network. Let θ_S denote the parameters of steganalysis network and $S(\theta_S, x)$ represent the output of steganalysis network on input image x. As is common with the implementation of discriminator in adversarial network, the steganalysis network's loss L_S is set to be sigmoid cross-entropy loss given by,

$$L_S = -y\log(S(\theta_S, x)) - (1-y)\log(1-S(\theta_S, x)), \tag{5}$$

where $y = 0$ if $x = c'$ and $y = 1$ if $x = c$.

So, the loss function of the proposed secure model is given by,

$$L_{Sec} = L_E + L_D + L_S. \tag{6}$$

The loss function of the proposed secure and robust model is given by,

$$L_{SecRob} = L_E + \tilde{L}_D + L_S. \tag{7}$$

4. Experimental Results and Analysis

In this section, we compare our steganography models with other existing deep learning-based models on ImageNet dataset [34] in terms of invisibility, security, robustness, and capacity. These compared existing deep learning-based models are Rehman's model [25], Zhang's basic model [28] and Zhang's ISGAN [28]. All the deep learning-based models are performed on Python and Tensorflow framework with 11GB GeForce GTX 1080 Ti, 3.20 GHz i7-6900K CPU, and 65GB RAM. Moreover, they are trained by using Adam optimizer with a fixed learning rate of 0.0001, a batch size of 16 and 30000 iterations.

4.1 Evaluation Metric and Experimental Dataset

Peak Signal to Noise Ratio (PSNR) is considered to be the perceptual quality measure of steganographic images and extracted secret images and finally to measure the invisibility and robustness. The higher the PSNR value, the better the invisibility and robustness. Steganalysis

accuracy is used to measure the security against steganalysis. Steganalysis accuracy represents the probability that the steganalyzer can correctly identify whether an image is a steganographic image or not. The lower the accuracy, the better the security against steganalysis. Bits per pixel (bpp), the number of bits hidden in one pixel is used to measure the hiding capacity. The bigger the bpp value, the higher the capacity.

As for the experimental dataset, we randomly select 8000 images from ImageNet dataset. Since the carrier image and secret image need to be the same resolution, the selected 8000 images are adjusted to 300 × 300 pixels. Then, the selected 8000 images are divided into two disjoint sets: a training set contains 3000 image sets and a testing set contains 1000 image sets, where each image set consists of one carrier image and one secret image. Moreover, the secret images are the grayed version of selected images. Some examples of image sets are given in **Fig. 5**.

(a) (b) (c)

Fig. 5. Examples of testing image sets. For each set, the left color image is a carrier image and the right gray one is a secret image.

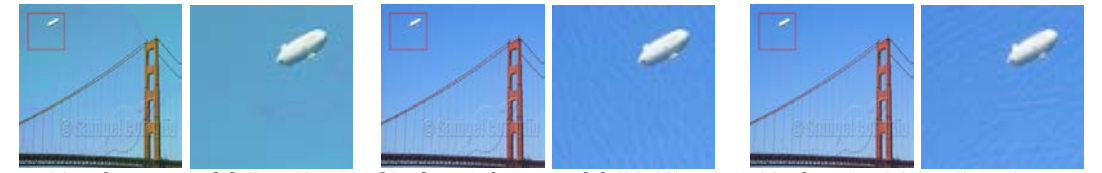
4.2 Experiments and Analysis

Firstly, the invisibility of the proposed models is evaluated. To this end, 1000 testing image sets are input into a well-trained model to generate 1000 steganographic images. The average PSNR values of 1000 steganographic images are provided in **Table 1**. It can be observed from **Table 1** that: (a) our secure model achieves the best performance in invisibility. It is better than our basic model because the adversarial network with steganalysis considered in the secure model enhances the quality of steganographic images to resist steganalysis; (b) the secure and robust model performs worst among our three models because it considers robustness as well as invisibility and security. Hence, the consideration of robustness results in a trade-off between the quality of the steganographic image and the extracted secret image. However, our secure and robust model is still better than Rehman's model [26], Zhang's basic model [28], and Zhang's ISGAN [28].

In order to apprehend the above numerical results better, the visual results (steganographic images) for six models are shown in **Fig. 6**, **Fig. 7**, and **Fig. 8**. These visual results are corresponding to the testing images given in **Fig. 5**. Moreover, in order to show the distortion clearly, these figures also present zoomed regions from the steganographic images. It can be seen from **Fig. 6** to **Fig. 8** that: (a) Rehman's model suffers from color distortion problem the most, while it is not the case for the proposed models mainly because only channel B is used for hiding; (b) In terms of distortion, our secure model shows the least distortion compared to the other five models, especially in the background regions; (c) these visual results are consistent with the numerical results given in **Table 1**

Table 1 PSNR (db) of steganographic images obtained from different models

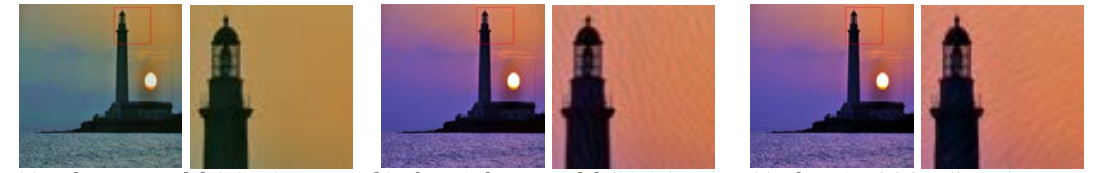
Model	PSNR
Rehman's model [26]	30.47
Zhang's basic model [28]	33.92
Zhang's ISGAN [28]	34.31
Our basic model	36.97
Our secure model	37.35
Our secure and robust model	34.07



(a) Rehman's model (31.53) (b) Zhang's basic model (33.63) (c) Zhang's ISGAN (34.14)

(d) Our basic model (35.45) (e) Our secure model (37.55) (f) Our secure and robust model (34.28)

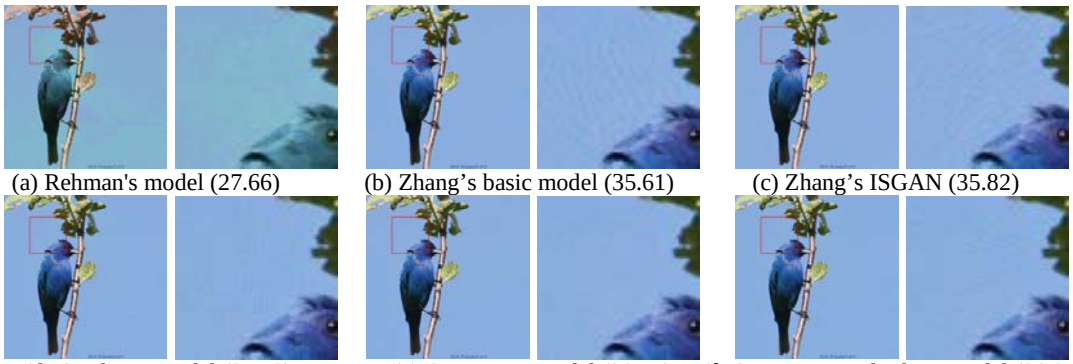
Fig. 6. Steganographic images of **Fig. 5**(a) obtained from different models and their corresponding PSNR values.



(a) Rehman's model (28.42) (b) Zhang's basic model (33.67) (c) Zhang's ISGAN (32.51)

(d) Our basic model (35.87) (e) Our secure model (38.02) (f) Our secure and robust model (33.38)

Fig. 7. Steganographic images of **Fig. 5**(b) obtained from different models and their corresponding PSNR values.



(a) Rehman's model (27.66) (b) Zhang's basic model (35.61) (c) Zhang's ISGAN (35.82)

(d) Our basic model (37.59) (e) Our secure model (39.12) (f) Our secure and robust model (36.82)

Fig. 8. Steganographic images of **Fig. 5**(c) obtained from different models and their corresponding PSNR values.

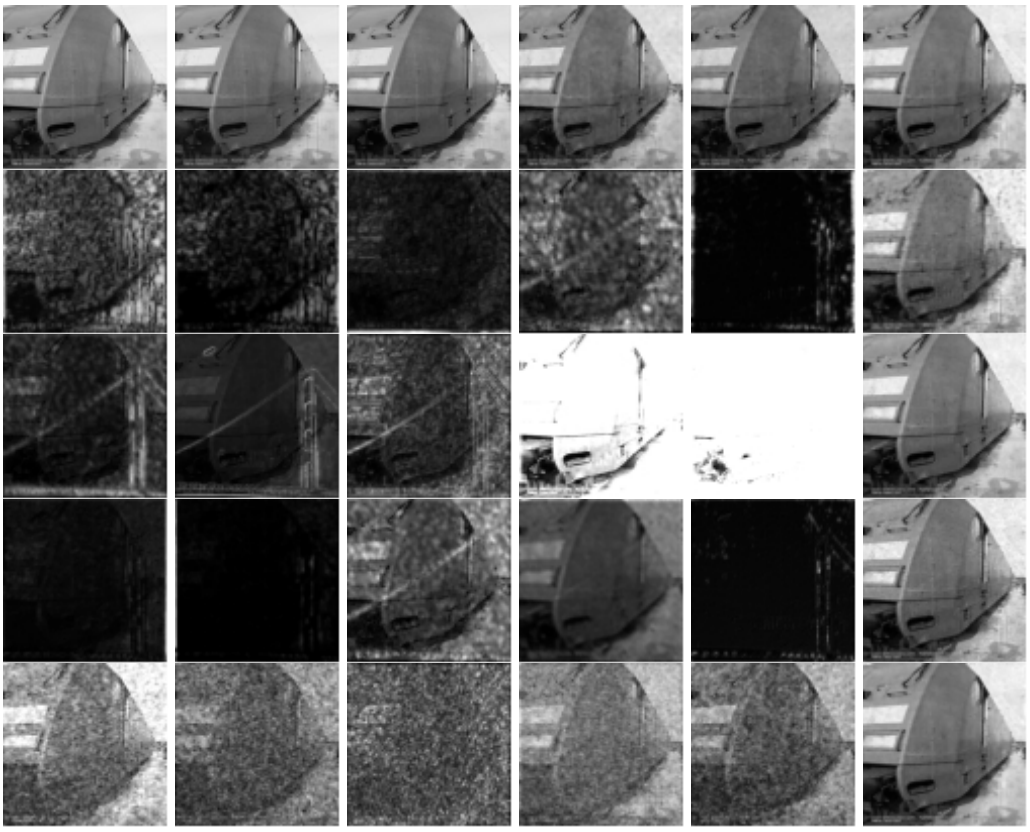
The proposed models are also tested in terms of security against steganalysis. For the purpose, a well-trained basic model is adopted to generate 5000 steganographic images on ImageNet. Then, the 5000 steganographic images and their corresponding carrier images constitute a dataset for retraining an improved XuNet steganalyzer used in [28]. After that, the trained steganalyzer is used to analyze the steganographic images obtained from different models. The average accuracies of the steganalyzer for different models are shown in **Table 2** The lower the accuracy, the better the security against steganalysis. As shown in **Table 2** our secure model performs the best among six models. The main reasons are that: (a) our secure model considers steganalysis during the training phase compared to Rehman's model [26], Zhang's basic model [28] and our basic model; (b) our secure model does not consider robustness during the training phase compared to our secure and robust model.

Table 2. Average accuracy of steganalyzer for steganographic images obtained from different models

Model	Accuracy
Rehman's model [26]	0.7833
Zhang's basic model [28]	0.7702
Zhang's ISGAN [28]	0.7328
Our basic model	0.7682
Our secure model	0.7248
Our secure and robust model	0.7413

Then, the robustness of the proposed models is evaluated. To do so, the steganographic images obtained from different models are attacked by additive Gaussian noise with standard deviation 1.0, sharpening filtering with 3×3 mask, mean filtering with 3×3 mask, and Gaussian filtering with 3×3 mask of standard deviation 1.0, respectively. Then, the secret images are extracted from these attacked steganographic images. Some examples for different models under different types of attacks are shown in **Fig. 9**, **Fig. 10**, and **Fig. 11**, while the average PSNR values of all extracted secret images are presented in **Table 3** It can be observed from **Table 3**and **Figs. 9-11** that: (a) our secure and robust model is significantly more robust to four kinds of attacks than the other five models because it considers these four

kinds of attacks during the training phase; (b) although our secure and robust model is not better than the other five models in the situation without attack, its extracted secret images still maintain high quality.

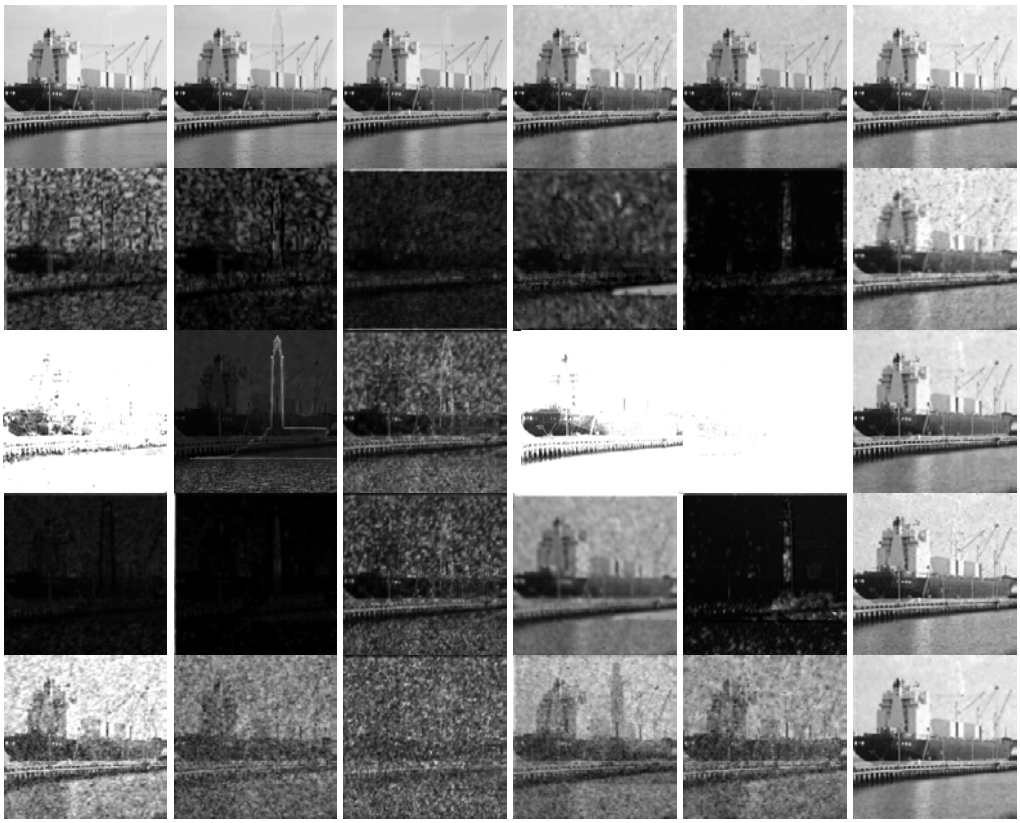


(a) Rehman's model (b) Zhang's basic model (c) Zhang's ISGAN (d) Our basic model (e) Our secure model (f) Our secure and robust model

Fig. 9. Secret images extracted from steganographic images in **Fig. 6** after different attacks. For rows, from top to down, no attack, mean filtering, sharpening filtering, Gaussian filtering, and Gaussian noise

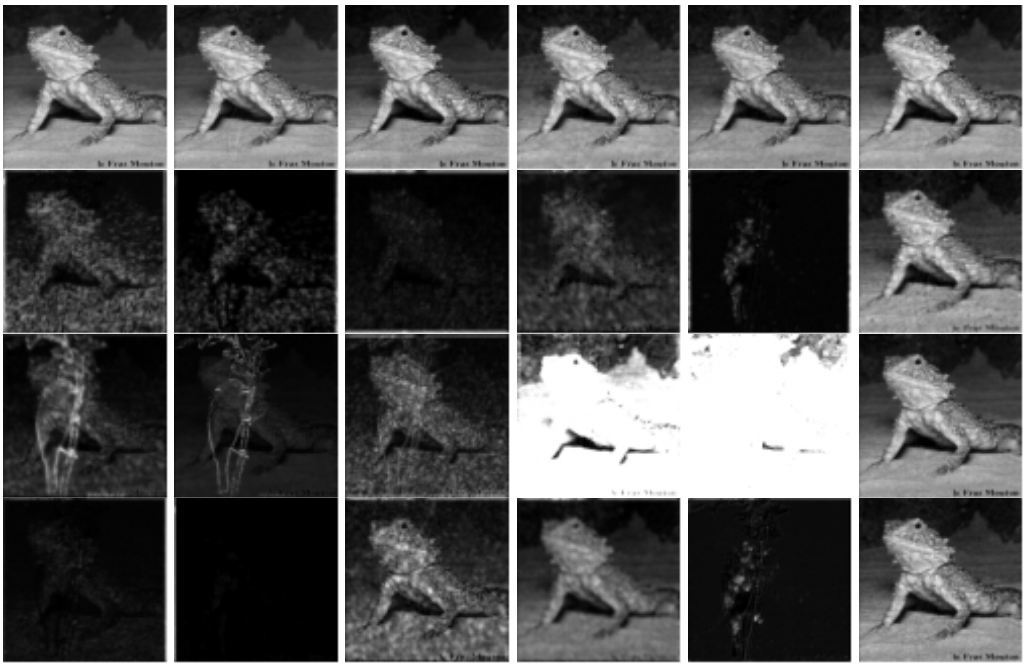
Table 3. PSNR of extracted secret images extracted under different attacks

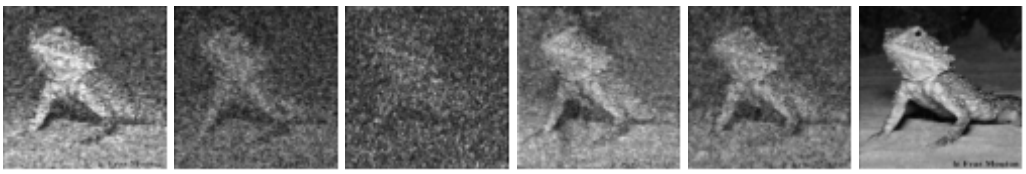
Model	No attack	Mean filtering	Sharpening filtering	Gaussian filtering	Gaussian noise
Rehman's model [26]	36.24	11.15	10.42	7.93	16.76
Zhang's basic model [28]	32.65	8.51	9.27	7.01	14.41
Zhang's ISGAN [28]	32.16	8.07	11.91	15.04	12.53
Our basic model	31.38	10.51	6.14	13.22	15.35
Our secure model	31.08	7.56	3.93	15.01	16.57
Our secure and robust model	29.72	24.65	28.26	26.25	28.49



(a) Rehman's model (b) Zhang's basic model (c) Zhang's ISGAN (d) Our basic model
(e) Our secure model (f) Our secure and robust model

Fig. 10. Secret images extracted from steganographic images in **Fig. 7** after different attacks. For rows, from top to down, no attack, mean filtering, sharpening filtering, Gaussian filtering, and Gaussian noise.





(a) Rehman's model (b) Zhang's basic model (c) Zhang's ISGAN (d) Our basic model (e) Our secure model (f) Our secure and robust model

Fig. 11. Secret images extracted from steganographic images in **Fig. 8** after different attacks. For rows, from top to down, no attack, mean filtering, sharpening filtering, Gaussian filtering, and Gaussian noise

Finally, as for the hiding capacity, our three models achieve 8 bpp by hiding a secret image with 300 × 300 × 8 bits into a carrier image with the resolution 300 × 300. Although the models of Rehman et al. [26] and Zhang et al. [28] also obtain 8 bpp hiding capacity, Rehman's model [26] does not consider security against steganalysis and robustness against attack during the training phase, while Zhang's ISGAN [28] does not consider robustness. So, our secure and robust model performs better than Rehman's model [26] and Zhang's ISGAN [28] in robustness as shown in **Table 3** and **Figs. 9-11**, and is better than Rehman's model [26] in security against steganalysis as shown in **Table 2**. Moreover, Rehman's model [26] suffers from the problem of color distortion as shown in **Figs. 6-8**.

5. Conclusion

Three steganography models, i.e., the basic model, the secure model, and the secure and robust model, have been designed for different cases by using adversarial network. All of them realize the functions of high-capacity secret information hiding and extraction. The secure model can also make steganalysis more difficult, while the secure and robust model not only can make steganalysis more difficult but also is robust against some attacks. The experimental results show that the secure model as well as secure and robust model have an overall better performance than some existing models. The main reasons are that: (a) the proposed three models use the concat operation to combine a carrier image and a secret image with the same resolution, moreover, it hides the secret image into the channel B of carrier image only to resolve the problem of color distortion; (b) the secure model combines the steganalysis network with the basic model to enhance its security against steganalysis; (c) the secure and robust model considers the attack network besides the steganalysis network to enhance its robustness against some attacks further. For future work, more kinds of attacks will be considered to improve the robustness further, such as JPEG compression, geometric attacks and so on.

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, March, 2010. Article (CrossRef Link)

[2] X. Duan, H. Song and C. Qin, "Coverless steganography for digital images based on a generative model," *Computers, Materials & Continua*, vol. 55, no. 3, pp. 483-493, 2018. Article (CrossRefLink)

[3] D. L. Robie and R. M. Mersereau, "Video error correction using steganography," *EURASIP Journal on Advances in Signal Processing*, vol. 1, pp. 164-173, 2002. Article (CrossRef Link)

[4] W. N. Lie, T. I. Lin and C. W. Lin, "Enhancing video error resilience by using data-embedding techniques," *IEEE Transactions on Circuits Systems for Video Technology*, vol.16, no.2, pp.300-308, 2006. Article (CrossRef Link)

[5] W. Zhang, S. Cheung and M. Chen, "Hiding privacy information in video surveillance system," in *Proc. of 12th IEEE International Conference on Image Processing*, pp. 868-871, 2005. Article (CrossRef Link)

[6] N.F. Johnson and S. Jajodia, "Exploring steganography: seeing the unseen," *IEEE Computer*, vol. 31, no.2, pp. 26–34, 1998. Article (CrossRef Link)

[7] BBC News: Hiding messages in plain sight, available from: http://news.bbc.co.uk/go/pr/fr/-/1/hi/technology/6361891.stm

[8] I. J. Kadhim, Pr. Premaratne, P. J. Vial and B. Halloran, "Comprehensive survey of image steganography: techniques, evaluations, and trends in future research," *Neurocomputing*, vol. 335, pp. 299-326, March, 2019. Article (CrossRef Link)

[9] L. Shi, Z. Wang and Z. Qian, "Distortion function for emoji image steganography," *Computers, Materials & Continua*, vol. 59, pp. 943-953, 2019. Article (CrossRefLink)

[10] W. Bender, D. Gruhl, N. Morimoto and A. Lu, "Techniques for data hiding," *IBM Systems Journal*, vol. 35, no. 3.4, pp. 313-336, January, 1996. Article (CrossRef Link)

[11] W. Lie and L. Chang, "Data hiding in images with adaptive numbers of least significant bits based on the human visual system," in *Proc. of IEEE International Conference on Image Processing*, pp. 286-290, February, 1999. Article (CrossRef Link)

[12] M. Ramkumar, A. N. Akansu and A. A. Alatan, "A robust data hiding scheme for images using DFT," in *Proc. of IEEE International Conference on Image Processing*, pp. 211-215, February, 1999. Article (CrossRef Link)

[13] B. Kaur, E. Kaur and J. Singh, "Steganographic approach for hiding image in DCT domain," *International Journal of Advances in Engineering and Technology*, vol.1, no. 3, pp. 72-78, 2011. Article (CrossRef Link)

[14] P. Chen and H. Lin, "A DWT based approach for image steganography," *International Journal of Applied Science and Engineering*, vol. 4, no. 3, pp. 275-290, January, 2006. Article (CrossRef Link)

[15] B. Chen, C. Zhou, B. Jeon, Y. Zheng and J. Wang, "Quaternion discrete fractional random transform for color image adaptive watermarking," *Multimedia Tools and Application*, vol. 77, no. 16, pp. 20809-20837, 2018. Article (CrossRefLink)

[16] Z. Qu, H. Song and C. Qin, "Analysis and improvement of steganography protocol based on bell states in noise environment," *Computers, Materials & Continua*, vol. 59, no. 2, pp. 607-624, 2019. Article (CrossRefLink)

[17] J. Schmidhuber, "Deep learning in neural networks: an overview," *Neural Networks*, vol.61, pp. 85-117, January, 2015. Article (CrossRef Link)

[18] R. Salakhutdinov, A. Mnih and G. Hinton, "Restricted boltzmann machines for collaborative filtering," in *Proc. of 24th ACM International conference on Machine Learning*, pp. 791-798, Junuary, 2007. Article (CrossRef Link)

[19] D. Volkhonskiy, I. Nazarov, B. Borisenko and E. Burnaev, "Steganographic generative adversarial networks," *arXiv:1703.05502 [cs.MM]*, 2019. Article (CrossRef Link)

[20] H. Shi, J. Dong, W. Wang, Y. Qian and X. Zhang, "SSGAN: secure steganography based on generative adversarial networks," *Advances in Multimedia Information Processing*, pp.534-544, 2017. Article (CrossRef Link)

[21] J. Hayes and G. Danezis, "Generating steganographic images via adversarial training," *Advances in Neural Information Processing Systems 2017*, December, 2017. Article (CrossRef Link)

[22] D. Hu, L. Wang, W. Jiang, S. Zheng and B. Li, "A novel image steganography method via deep convolutional generative adversarial networks," *IEEE Access*, vol. 6, pp. 38303-38314, July, 2018. Article (CrossRef Link)

[23] J. Zhu, R. Kaplan, J. Johnson and F. F. Li, "Hidden: hiding data with deep networks," in *Proc. of ECCV 2018*, pp. 657-672, 2018. Article (CrossRef Link)

[24] W. Tang, S. Tan, B. Li and J. Huang, "Automatic steganographic distortion learning using a generative adversarial network," *IEEE Signal Processing Letters*, vol. 24, no. 10, pp. 1547-1551, October, 2017. Article (CrossRef Link)

[25] J. Yang, K. Liu, X. Kang, E. K. Wong and Y. Shi, "Spatial image steganography based on generative adversarial network," *arXiv:1804.07939 [cs.MM]*, 2018. Article (CrossRef Link)

[26] A. Rehman, R. Rahim, M. Nadeem and S. Hussain, "End-to-end trained CNN encode-decoder networks for image steganography," in *Proc. of Computer Vision - ECCV 2018 Workshops*, pp. 723-729, 2018. Article (CrossRef Link)

[27] S. Baluja, "Hiding images in plain sight: deep steganography," in *Proc. of Neural Information Processing Systems*, pp. 2069-2079, December, 2017. Article (CrossRef Link)

[28] R. Zhang, S. Dong and J. Liu, "Invisible steganography via generative adversarial networks." *Multimedia Tools and Applications*, vol. 78, no. 7, pp. 8559–8575, 2019. Article (CrossRef Link)

[29] G. Xu, H. Wu and Y. Shi, "Structural design of convolutional neural networks for steganalysis," *IEEE Signal Processing Letters*, vol. 23, no. 5, pp. 708-712, May, 2016. Article (CrossRef Link)

[30] K. Zhang, A. Cuesta-Infante, L. Xu and K. Veeramachaneni, "SteganoGAN: high capacity image steganography with GANs," *arXiv:1901.03892 [cs.CV]*, 2019. Article (CrossRef Link)

[31] Y. G. Fu and R. M. Shen, "Color image watermarking scheme based on linear discriminant analysis," *Computer Standards and Interfaces*, vol. 30, no. 3, pp. 115-120, 2008. Article (CrossRef Link)

[32] M. Kutter and S. Winkler, "A vision-based masking model for spread-spectrum image watermarking," *IEEE Transactions on Image Processing*, vol. 11, no. 1, pp. 16-25, 2002. Article (CrossRef Link)

[33] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. of CVPR 2016*, pp. 2818-2826, 2016. Article (CrossRef Link)

[34] J. Deng, W. Dong, R. Socher, L. Li and F. Li, "ImageNet: a large-scale hierarchical image database," in *Proc. of CVPR 2009*, pp. 248-255, June, 2009. Article (CrossRef Link)

Beijing Chen received the Ph.D. degree in Computer Science in 2011 from Southeast University, Nanjing, China. Now he is an associate professor in the School of Computer & Software, Nanjing University of Information Science & Technology, China. His research interests include color image processing, image forensics, image watermarking, and pattern recognition.

Jiaxin Wang received the B.S. degree in Computer Science & Technology in 2017 from Binjiang College, Nanjing University of Information Science & Technology. He is currently pursuing the M.S. degree in the School of Computer & Software, Nanjing University of Information Science & Technology, Nanjing, China. His research interest includes information hiding.

Yingyue Chen is currently pursuing a bachelor's degree in School of Internet of Things, Jiangnan University, Wuxi, China. Her research interests include wireless sensor networks, and image processing.

Zilong Jin received the Ph.D. degree in Computer Engineering in 2016 from Kyung Hee University, Korea. He is currently an assistant professor of School of Computer and Software at Nanjing University of Information Science and Technology, China. His research interests include wireless sensor networks, mobile wireless networks, and cognitive radio networks.

Hiuk Jae Shim received the Ph.D. degree in Electronics Electrical Engineering from Sungkyunkwan University, Suwon, Korea, in 2013. He is currently a lecturer in the School of Computer & Software, Nanjing University of Information Science & Technology, China. His main research interests include video compression, and distributed video coding.

Yun-Qing Shi received the Ph.D. degree from the University of Pittsburgh, USA. He has been with the New Jersey Institute of Technology, USA, since 1987. He has published more than 300 papers, and holds 30 U.S. patents. His research interests include data hiding, forensics and information assurance, visual signal processing, and communications. He has served as an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS (II). He serves as an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY.