IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

Received August 7, 2020, accepted August 28, 2020, date of publication September 2, 2020, date of current version September 24, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3021103

A Novel Grayscale Image Steganography Scheme Based on Chaos Encryption and Generative Adversarial Networks

QI LI[1], (Member, IEEE), XINGYUAN WANG[1], XIAOYU WANG[1], (Member, IEEE), BIN MA[2], (Member, IEEE), CHUNPENG WANG[2], (Member, IEEE), YONGJIN XIAN[1], AND YUNQING SHI[3], (Life Fellow, IEEE)

[1]School of Information Science and Technology, Dalian Maritime University, Dalian 116026, China
[2]School of Cyber Security, Qilu University of Technology, Jinan 250353, China
[3]Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ 07102, USA

Corresponding authors: Xingyuan Wang (xywang@dlmu.edu.cn), Xiaoyu Wang (qluwxy@163.com), and Bin Ma (sddxmb@126.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61872203, Grant 61802212, and Grant 61672124; in part by the Password Theory Project of the 13th Five-Year Plan National Cryptography Development Fund under Grant MMJJ20170203; in part by the Liaoning Province Science and Technology Innovation Leading Talents Program Project under Grant XLYC1802013; and in part by the Key Research and Development Projects of Liaoning Province under Grant 2019JH2/10300057.

ABSTRACT This article proposes a novel grayscale image steganography scheme that is capable of hiding an encrypted secret image into a cover image of the same size. The secret image is encrypted using the chaos encryption technology before being hidden. The encrypted image and cover image are then transformed into the steg image by a convolutional neural network (CNN). Also, a generative adversarial network (GAN) is adopted to produce a more realistic steg image, whose appearance is difficult to distinguish from the cover image. The steganography scheme is composed of three CNNs which are regarded as hiding network, discriminative network, and extracting network, respectively. Additionally, a new weight allocation mechanism is introduced to ensure the balanced training procedure of hiding-extracting networks. The obtained experimental results show that the proposed scheme not only solved the problems of secret information leakage and color distortion of the current mainstream methods but also maintain high embedding capacity. Compared with other methods, the average values of the PSNR and SSIM of our steg images can reach 0.987 and 42.3, respectively. And the quality of the reconstructed secret images is also advantageous.

INDEX TERMS Grayscale image steganography, chaos encryption Technology, convolutional neural network (CNN), generative adversarial network (GAN).

I. INTRODUCTION

In today's information era, cloud computing provides enough cyberspace for both individuals and organizations to store their data, such as documents, videos, and images, simultaneously providing a convenient way for people to access and share the data through the network [1]–[3]. Since the stored data can be personal, commercial, or even highly-confidential, it is crucial to prevent data leakage, which is commonly done by two following methods: information hiding and encryption technology. Information hiding is to embed secret data into a cover medium wherein the secret data are undetectable. For information hiding, steganography is the main component and can be further classified into two categories: spatial domain and transform domain [4], [5].

Steganography conceals data into a cover medium in a way that the embedded data are undetectable. It focuses on the imperceptivity of both the hidden data and the act of data embedding, i.e., the embedded data are not only imperceptible to human eyes but also to potential analyzers [6]. The steganography has two fundamental components, secret information to be hidden, and cover image. The secret information can be audio [7], text [8], image [9], or even video data. In the traditional steganography, images denote the first choice of a cover image due to their rich texture and edge information, which make it difficult to be detected after embedding the secret data. Currently, most of the steganography methods embed only a small amount of secret information [10].

The associate editor coordinating the review of this manuscript and approving it for publication was Alessandra De Benedictis.

In order to improve the embedding capacity of the steganography algorithm, Lin *et al.* [11] proposed a new image hiding technology, which can be applied to embed a gray-scale image within a color image, but the distortion of steg images embedded with secret images are too serious to be applied in practice. Li *et al.* [12] proposed a novel image-hiding scheme, which can improve the quality of steg image by exploring the block correlation between the cover image and secret image. Hu [13] presented a novel grayscale image hiding scheme. In this scheme, the modulus function and the image property are employed to improve the quality of steg image. Zhang *et al.* [14] proposed a novel technique for hiding arbitrary binary data in images using generative adversarial networks. Although the quality of steg images is very satisfactory, the embedding capacity can only reach 0.4 bpp. Yang *et al.* [15] proposed a secure steganography algorithm by using adversarial training. In this scheme, the cover image can be translated into the embedding change probability, so the embedding capacity has been improved. Zhang *et al.* [16] proposed a generative reversible data hiding (GRDH) scheme, which can implement the task of secret information deliver by using image-to-image translation. Islam *et al.* [17] proposed a novel image steganography technique based on most significant bits (MSB) of image pixels. And this scheme is not only secure, but computationally efficient as well. Rehman *et al.* [18] proposed an efficient steganography method, which represents the cover image using the Fibonacci sequence. This scheme has achieved very good robustness.

Although the traditional image steganography method has made a lot of progress, the embedding capacity has never been able to make a breakthrough [19]–[21]. With the rapid development of convolutional neural networks, many researchers have attempted to apply CNN to steganography and obtained considerable achievements [22]–[24]. Among them, Baluja [25] proposed a steganography scheme to place a full-size color image within another image of the same size. This scheme consists of three components: preparation network, hiding network, and reveal network. The bit rates of this scheme are 10× −40× higher than those of the traditional methods. Rehman *et al.* [26] proposed a CNN-based encoder-decoder architecture for embedding the images as payload. In this approach, a cover image is a color image, and a secret image is a grayscale image. Although a deep neural network can be used to hide images into images of the same size without interfering with the appearance of the cover image significantly, there are still many problems to be addressed. In the steganography method proposed in [25], assuming that an attacker can obtain the original cover image, the linear operation on the original cover image and steg image can still expose the information of the transmitted secret image, indicating that this method has the additive property of the traditional steganography. Since the cover images are in color and a secret image is hidden into all the pixel bits of the cover image, there is also the problem of color distortion in the generated image. The steganography method proposed in [26] can reduce the embedding capacity by nearly a half, but the problem of color distortion cannot be solved because the cover image is in color. The problems of the existing steganography methods are illustrated in Fig. 1.

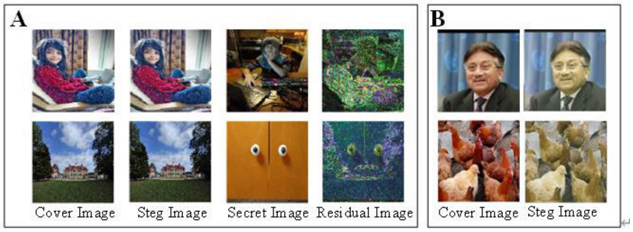


FIGURE 1. Illustration of the problems of the existing steganography methods. (A): Secret information leakage. (B): Color distortion in the cover image.

Different from the deep steganography scheme [25], [26], in this work, CNN is adopted to hide an encrypted secret image into a cover image of the same size. To the best of authors' knowledge, this is the first time that such a method for grayscale-in-grayscale image hiding based on a generative adversarial network (GAN) is reported.

The major contributions of our work can be summarized as follows:

(1) In order to solve the problem of secret information leakage, the secret image to be hidden is firstly encrypted by the chaos encryption technology, and then the cover image and encrypted secret image are integrated to a steg image. Since the semantic content of secret images is scrambled before they are embedded into cover images, the confidentiality of secret information is well protected.

(2) The cover images used by the existing steganography schemes are all color images, so their luminance channels are damaged after embedding a large amount of secret information. Therefore, the current steganography schemes suffer from low secret information security. In this article, the steganography scheme uses a grayscale image as a cover image, and it is needed to consider only the integrity of semantic content without considering the color distortion.

(3) Different from the existing deep steganography schemes, in this article, the steganography scheme adds a discriminative network except the hiding and extracting networks. The discriminative network consists of a CNN-based steganalysis model, which is developed based on the XuNet [27].

(4) The hiding network adopts the skip connections to concatenate low-level features with high-level features, which contributes to hiding the details of the secret image. In addition, a new weight allocation mechanism is introduced to ensure the balanced training procedure of hiding-extracting networks.

The remainder of this article is organized as follows. Section II introduces the related work on deep learning and steganography. Section III describes the proposed method. Section IV presents the experimental results and analysis in detail. Section V concludes the paper.

II. STEGANOGRAPHY FOR HIDING IMAGE WITHIN IMAGE

A. DEEP STEGANOGRAPHY

Baluja (2017) proposed a deep steganography method, which can embed a full-size color image within another image of the same size. The DNNs are simultaneously trained to create hiding and revealing processes and are designed to specifically work as a pair. The steganography procedure is shown in Fig. 2. The steganographic scheme consists of three components: preparation network, hiding network, and reveal network. The preparation network is adopted to perform the conversion on an image to be hidden, so as to use the texture and edge of the cover image more reasonably in encoding the secret image. The hiding network is composed of deep convolutional neural networks, whose inputs represent the combination of the cover image and the preparation network output. Different from the traditional steganography scheme, CNN is adopted to encode the secret information into all available bits of the cover image. There is no obvious difference between the cover image and the generated container image in deep steganography method. In fact, the hiding network is equivalent to compressing the cover image and secret image and generating a three-channel container image. The last step of the steganography process is to extract the secret image, which is conducted by a receiver to recover the secret image, thus accomplishing the transmission of secret information.

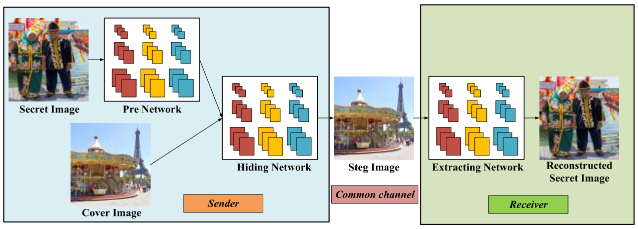


FIGURE 2. The architecture of deep steganography.

The deep steganography relies largely to the idea of auto-encoder, but, in contrast to the auto-encoder, it encodes the cover and secret images simultaneously in order to make the intermediate image look as similar as possible to the cover image. Simultaneously, a loss function is proposed to reconstruct the secret image. The formula is shown below.

$$L(c, c', s, s') = ||c - c'|| + \alpha ||s - s'|| \quad (1)$$

where c denotes the cover image, s denotes the secret image, and α represents the reconstruction loss weight of the secret image. The whole steganography scheme aims at minimizing the loss function, making it as small as possible. The loss function indicates the difference between the steg image and the cover image, as well as between the original secret image and the reconstructed secret image. The maximum amount of information that can be hidden by the traditional steganography algorithms, such as HUGO, S-UNIWARD [28], and MiPOD [5], is up to 0.4 bpp. The deep steganography uses deep neural networks to place a full-size color image within another image of the same size, and compared to the traditional steganography algorithms, the embedded capacity is increased by 10–40 times. Due to the increase in embedded capacity, it is impossible to avoid the detection of the existing steganalysis algorithms completely. In the proposed method, the neural networks are used for the first time to place a full-size color image within another image, thus providing a new idea to the image steganography.

B. ENCODER-DECODER NETWORKS FOR IMAGE STEGANOGRAPHY

Rehman *et al*. proposed a CNN based encoder-decoder architecture for embedding the images as a payload. The idea of auto-encoder is adopted completely. This method utilizes characteristics of encoder to learn from low-level features of a cover image to high-level features to effectively hide a secret gray image, while decoder is responsible for recovering the secret image to the greatest possible extend. The architecture of the steganography procedure is shown in Fig. 3.

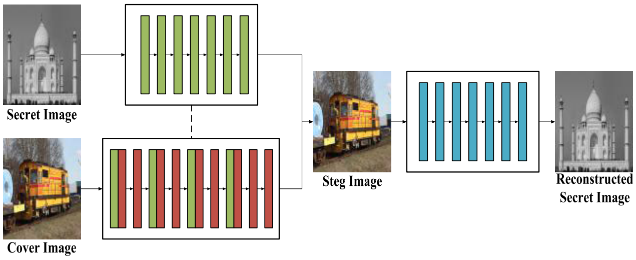


FIGURE 3. The architecture of the encoder-decoder networks for image steganography.

As presented in Fig. 3, the steganography scheme consists of an encoder and a decoder. The encoder consists of two parallel networks. The upper layer consists of seven-layer CNN, including convolution and ReLU layers, and it is responsible for decomposing the secret image into a series of low-level image features and high-level semantic features. The lower layer consists of nine-layer CNN, including convolution and ReLU layers, and it is responsible for decomposing the cover image into a series of low-level image features and high-level semantic features. The merging operation is utilized to fuse the cover and secret images. In order to hide the secret image the best possible and make the steg image indistinguishable from the cover image, a new loss function is proposed for the encoder and decoder during training, and it is given by:

$$Loss(c, c', s, s') = \alpha ||c - c'||^2 + \beta ||s - s'||^2 + \lambda(||w_1||^2 + ||w_2||^2) \quad (2)$$

where c and c' represent the cover image and the steg image, respectively; s and s' represent the secret image and the reconstructed secret image, respectively; α and β represent the control parameters of encoder and decoder, respectively; and lastly, w_1 and w_2 represent the parameter values learned from encoder and decoder, respectively. The training goal of the steganography scheme is to minimize the loss function value. This method shows excellent experimental results on a series of datasets, such as ImageNet, CIFAR10, and MNIST.

III. PROPOSED SCHEME

In order to solve the problems of the existing steganography methods [25], [26], this article proposes a novel steganography method, wherein a grayscale image is adopted as a cover image to eliminate the problem of color distortion in the cover image. The chaotic encryption technique is used before the secret, and cover images are merged. Namely, the content of a secret image is scrambled by the chaotic encryption algorithm, and then the encrypted secret image and the cover image are merged, thereby solving the problem of secret image information leakage during transmission. Thus, the generated steg image is double-encrypted. Besides, in order to enhance the generated image quality further, the idea of generative adversarial networks is adopted to train discriminative network using distributions of generated images and real images. When a receiver receives the steg image, first, the encrypted secret image is extracted by the extracting network, and then the encrypted image is recovered to get the secret image. The security of the steganography scheme is improved by the method of double encryption and decryption. The architecture of the proposed steganography scheme is shown in Fig. 4.

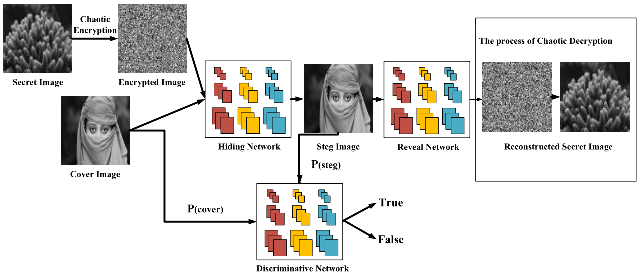


FIGURE 4. **The architecture of the proposed image steganography scheme.**

A. CHAOTIC ENCRYPTION ALGORITHM

Compared to the traditional image encryption methods, the chaotic image encryption denotes an encryption technique based on a chaotic sequence generated by the chaotic system [29]. Because the proposed steganography scheme both encrypts a secret image and completely decrypt and encrypt the secret image after the extraction process. The robustness of encryption algorithms is crucial. Since an encryption method based on the chaotic encryption technique has excellent initial sensitivity and can effectively resist various kinds of attacks, it is reasonable to use the chaotic encryption technique in our steganography scheme.

In the chaotic state, the Chen system can iterate out three chaotic sequences, which are further applied to different steps of the encryption algorithm in this article. The Chen system is defined by:

$$\begin{cases} \dot{x} = 35(y-x) \\ \dot{y} = -7x - xz + 28y \\ \dot{z} = xy - 3z \end{cases} \tag{3}$$

In this work, the size of a secret image is supposed to be 512 × 512 pixels.

The proposed encryption algorithm mainly consists of four steps, which are as follows:

Step 1. Denote the image matrix as $P = [p]_{512\times512}$.Select the initial values $K' = [k'_0, k'_1, k'_2, k'_3]$, where $k'_0 \leq 512\times512$; the number of iterations of the scrambling algorithm is set to 50; a triplet $[k'_1, k'_2, k'_3]$ denotes the initial values of the chaotic Chen system, and it is set to [0.0663598, 0.45679, 0.9256] randomly. The attack resistance ability of the encryption algorithm is enhanced by this random operation.

Step 2. Obtain the chaotic sequences C_1, C_2, C_3 by the Chen system with the key group. The three sequences consist of bits 1001 to 1000 $+M$ $\times N$ to discard for chaos. ($M \times N$ denote the size of secret image)

Step 3. Let A be the helical scan sort matrix; A' is calculated based on the number of scrambling iterations as follows:

$$A'_i = \begin{cases} \text{rotation}(A, 90), & 0 < C_1(i) < 0.25 \\ \text{rotation}(A, 180), & 0.25 \leq C_1(i) < 0.5 \\ \text{rotation}(A, 270), & 0.5 \leq C_1(i) < 0.75 \\ \text{rotation}(A, 270), & 70.75 \leq C_1(i) < 1 \end{cases} \tag{4}$$

Let $P = [p]_{512\times512} = [p_0]_{512\times512}$, and get the index S of the chaotic sequence C_2 in order. Then, $P'_i = [p'_i]_{512\times512}$ is obtained, and it holds that:

$$p'_i(A'_i(S(j))) = p_{i-1}(j), \quad j = 1, 2, \ldots, 512 \times 512 \tag{5}$$

Setting based on the Equation 5, $i = 1, 2, \ldots, k'_0$, and the scrambling process is repeated k'_0 times. Finally, $P' = P'_{k'_0}$.

Step 4. Let matrix $P'' = [p'']_{512\times512}$ be diffused by $P' = [p']_{512\times512}$ with C_3 using the chaotic diffusion as follows:

$$p''(j) = \begin{cases} p'(j) \oplus (\text{mod}(\lfloor C_3(j) \times 1000 \rfloor, 256)), j = 1 \\ p'(j) \oplus p'(j-1) \oplus (\text{mod}(\lfloor C_3(j) \times 1000 \rfloor, 256)), \\ \qquad j = 2, 3, \ldots, M \times N \end{cases} \tag{6}$$

By these four steps, a cipher secret image P'' is obtain, while a key sequence K', which is used to encrypt the original image and decrypt the encrypted image. The decryption algorithm is the opposite of the encryption algorithm, and will not be repeated due to the limitation on the paper length.

B. HIDING NETWORK

In the schemes presented in [25], [26], the original and cover images are first fed to the input of the hiding network, and then the steg image is generated. However, in this article, the hiding network input denotes the combination of encrypted secret image and cover image in order to prevent the secret information leakage. The architecture of the hiding network relies largely on the U-Net architecture, and it mainly consists of two parts: the compression stage and the expansion stage. In the compression stage, seven convolutional layers with a convolution kernel size of 3 × 3 are used. Each convolution layer is followed by a ReLU layer and an average pooling layer. The ReLU layer is adopted because it can enhance the nonlinear fitting ability of a model, so as to integrate the details of the encrypted secret image and cover image better. On the other hand, the average pooling layer can preserve the background of a cover image to the greatest extend and avoid the distortion of the generated image content. The training speed of the steganography model is further accelerated by using the average pooling operation that reduces a large number of training hyper-parameters. Finally, the input is transformed into 1024-feature maps with a size of 4 × 4 in the compression stage. In the expansion stage, the de-convolution operation is adopted to transform the 1024-feature maps into a single-channel image. In this stage, seven convolutional layers with a convolution kernel size of 4×4 are utilized to de-convolve the feature maps. Each de-convolution layer is followed by a ReLU layer, and the last convolutional layer is followed by a Tanh layer. In order to prevent the information loss, instead of using the average pooling layer, each layer in the compression stage is copied to the expansion stage while keeping its size unchanged, which ensures that the secret information is perfectly encoded into steg image. In the last layer, 32 feature maps with a size of 512 × 512 are mapped to the steg image with a size of 512 × 512, which has no difference in appearance with the cover image. The architecture of the hiding network is shown in Fig. 5.

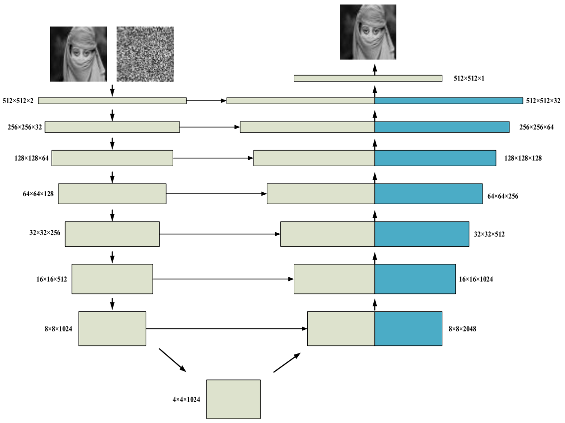


FIGURE 5. The architecture of the hiding network.

C. EXTRACTING NETWORK

The main function of the extracting network is to extract the encrypted secret image from the steg image. The quality of the extracted secret image has a great impact on the decryption effect of a secret image. The extracting network used in this article is a CNN with ten convolutional layers, and this CNN architecture was chosen based on the experimental results as the optimal one. In the CNN, the main purpose of drop out and pooling operations is to eliminate redundant training parameters, thereby reducing the model complexity and possibility of overfitting, while simultaneously enhancing the nonlinear fitting ability of the model. In order to minimize the loss of hidden information, in the extracting network, each convolution layer is followed by a ReLU layer and a batch normalization (BN) layer, but there is no pooling layer. In this way, not only the nonlinear fitting ability of the model is guaranteed, but also the hidden information is recovered to the greatest extent. The kernel size of each convolutional layer is set to 3 × 3, and the last convolutional layer is a Tanh layer, which maps 32 feature maps with a size of 512 × 512 into the reconstructed encrypted secret image. The architecture of the extracting network is shown in Fig. 6.

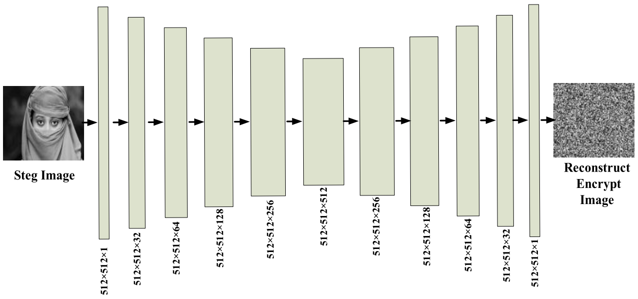


FIGURE 6. The architecture of the extracting network.

D. DISCRIMINATIVE NETWORK

In the proposed steganography scheme, a discriminative network module is introduced, and the adversarial idea of GAN is used to improve the generated image quality [30]. The goal of a GAN is to estimate the potential distribution of real data and accordingly generate new data samples with the same distribution. The GAN model consists of a generator G and a discriminator D. By establishing a minimax confrontation process between generator G and discriminator D, the ability of generator G to generate samples is optimized. Theoretically, the distribution of samples generated by generator G trained by the GAN model can gradually approach to the distribution of real data samples. The fundamental goal of the GAN is to construct a generator G using real data X. Finally, the sample distribution P_g generated by G and the sample distribution of the real data P_d are indistinguishable. The input of generator G is random noise z, and the generated sample distribution is considered as $P_g = G(z)$. Then, a discriminator D is used to distinguish the generated samples from the real data continuously, thus gradually improving the performance of generator G. The objective function of the model is given by:

$$\underset{G}{min}\,\underset{D}{max} = E_{x\sim P_d}[(log(D(x)] + E_{z\sim P_z}[(1 - D(G(z)))] \quad (7)$$

The steganography and steganalysis correspond to each other, and the relationship between them is antagonistic. Xu *et al.* [27] proposed a steganalysis architecture based on a CNN, which takes the absolute values of elements in the feature maps generated from the first convolutional layer to facilitate and improve the statistical modeling in the subsequent layers. The discriminator proposed in designed based on its CNN architecture, but it does not include the high-pass filtering (HPF) and the absolute activation layer (ABS). The architecture of the discriminating network is shown in Fig. 7. By adding a discriminator to the steganography scheme, the distribution and appearance of a steg image become more similar to those of the cover image. The loss function is given by:

$$\mathcal{L}_{disc} = \mathbb{E}_{c\sim P_c}[logD(c)] + \mathbb{E}_{c\sim P_c, s\sim P_s}[\log(1 - D(H(c, s))] \tag{8}$$

where P_c and P_s represent the distributions of the cover image and secret image, respectively; H (c, s) represents the steg image generated by the generator. As shown in Fig. 7, the discriminator network consists of six convolutional layers, and each of them is followed by the average pooling layer and the BN layer except the last convolutional layer. Finally, 128 feature maps are mapped into the softmax function by fully-convolutional layers, and then the probabilities of true and false samples are obtained.

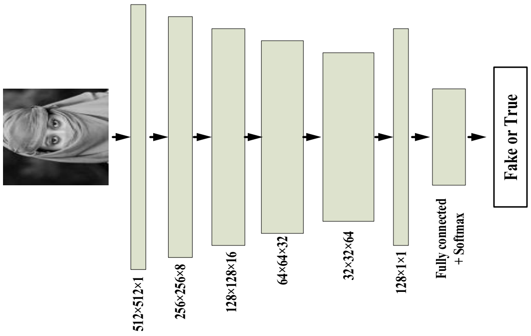


FIGURE 7. The architecture of the discriminative network.

E. LOSS FUNCTION

In order to minimize the loss of the steg image c' and original cover image c, as well as the extracted secret image s' and original secret image s, the hiding network and extrating network are optimized to minimize the reconstruction errors in the training procedure. The mean square error (MSE) is usually utilized as a cost function, but it only punishes the pixel error between the two images, and ignores the basic properties of the image (such as texture, structure, etc.). To this end, we put forward a new cost function, which is minimized to optimize the model:

$$\underset{H_\theta}{argmin}\,\frac{1}{n}\sum_{i=1}^{n}(1 - SSIM(c_i, H_\theta(c_i, s_i))) \tag{9}$$

where s represents the steg image or the reconstructed secret image, c represents the cover image or the original secret image, and n represents the number of training samples. The loss function proposed in [25] considers the ratio of the loss term of the cover image and steg image to the loss term of the secret image and reconstructed secret image is 1:1. In [26], it was considered that only the loss term of the secret image passed through the entire steganography scheme in the back-propagation process of model training, so the ratio of loss function was set to 4:3. In this article, the adversarial network assists in the image generation stage, so the weight of loss term of the secret image needs to be balanced. The loss function proposed in this article is given by:

$$\mathcal{L}_{steg} = \lambda_1||c - c'|| + \lambda_2||s - s'|| + \lambda_3\mathcal{L}_{disc} \tag{10}$$

where $||c - c'||$ represents the loss term between the cover image and the steg image, $||s - s'||$ represents the loss term between the original secret image and the reconstructed secret image and each λ is utilized to balance the weight of individual objective. The extensive experiments show that the best results can be achieved when the value of α and β are 0.65 and 0.85.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this work, 80000 training images and 10000 test images were gathered from the ImageNet dataset, which were utilized for training the network models. In order to optimize the model parameters, the Adam optimization method was used to adjust the learning rate in the training process automatically. The initial learning rate was set to 0.0001, and the hyper-parameters α and β were set to 0.65 and 0.85, respectively. The number of images per batch was set to 64, and the maximal number of training iterations was set to 250. In the experiments, a PC with GPU NVIDIA GeForce Tesla V100 32G was used, and the experimental environment Pytorch 1.1 and Python 3.7 were adopted. The experimental results obtained from randomly selected images from the ImageNet dataset by the trained model are presented in Fig. 8.

The first column in Fig. 8 represents the cover images, which were the grayscale images with a size of 512 × 512; the second column represents the secret images to be hidden, which were also grayscale images with a size of 512×512; the third column shows the results of the encrypted secret images, and the fourth column represents the steg images, which denoted the composition of the cover image and encrypted secret image, and which were obtained after the merging process. In terms of experimental visual effects, after the encrypted secret image was embedded into the cover image, the appearance of the generated image was almost as same as that of the cover image. The fifth column in Fig. 8 represents the reconstructed encrypted secret images that were obtained after the extraction process. The quality of the reconstructed encrypted secret images directly affected the restoration results of the secret image. The sixth column in Fig. 8 represents the decrypted secret images. In terms of experimental visual effects, the reconstructed secret images could recover almost the entire semantic content of the original secret image. However, the decrypted secret image could be affected by noise due to the encryption algorithm.

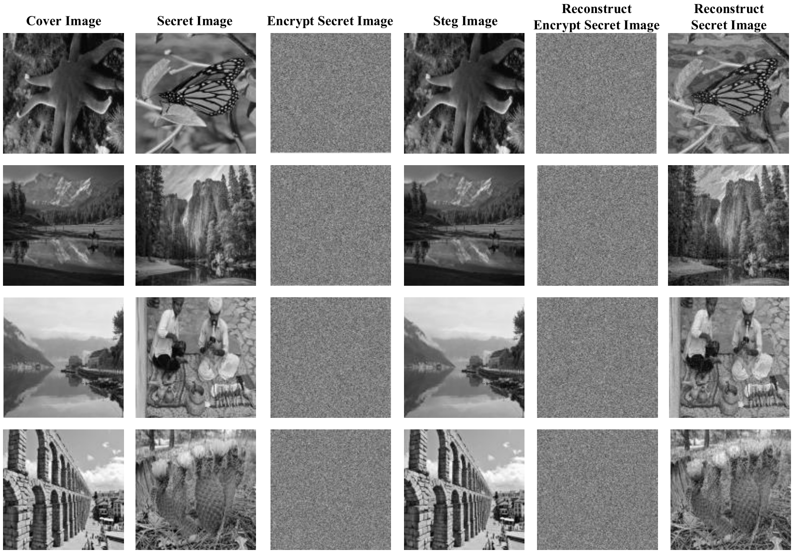


FIGURE 8. The experimental results obtained from randomly selected images from the ImageNet dataset by the proposed steganography scheme.

In order to compare the proposed scheme with the scheme proposed in [25], the same images used in [25] were utilized to test the proposed steganography scheme. The comparison results are shown in Fig. 9. The first two rows in Fig. 9 show the drawbacks of the method proposed in [25]. The first column in Fig. 9 represents the original cover images, the second column represents the steg images, the third column represents the secret images to be hidden, and the fourth column represents the residual image obtained by the linear processing of the original cover images and steg images, which had been magnified by 20 times. As presented in Fig. 9, the method proposed in [25] had a great disadvantage regarding the secret information security — much semantic information of the secret image existed in the residual image. The last two rows in Fig. 9 show the processing results of the same image by the proposed steganography scheme. As shown in Fig. 9, the information on the secret image did not exist in the residual image. Thus, at the same embedding rate, the proposed steganography method greatly improved security compared to the method proposed in [25].

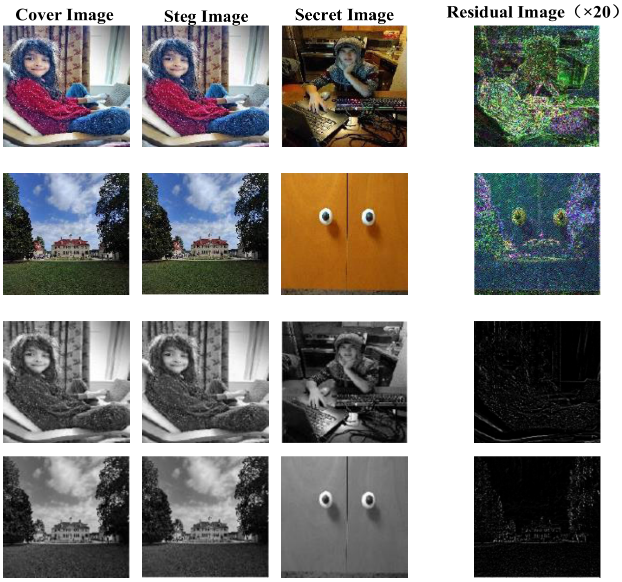


FIGURE 9. The comparison results of the proposed scheme and the method proposed in [25].

The proposed method was also compared with the method proposed in [26]. The same images used in [26] were utilized to test the proposed steganography scheme. The comparison results are shown in Fig. 10. The first two rows in Fig. 10 show the comparison results on the LFW dataset. When the embedding rate of the proposed method was three times larger than that of the method proposed in [26], the color distortion problem that existed in [26] could be avoided, and the difference between the cover image and the steg image was almost indistinguishable by the naked eye. The middle two rows in Fig. 10 show the experimental results on the PASCAL-VOC12 dataset. The synthetic steg image still performed well, but the content distortion appeared in the reconstructed secret image (refer to the reconstructed secret image 02 in Fig. 10). The last two rows in Fig. 10 show the experimental results on the ImageNet dataset. Compared to the method proposed in [26], under the premise that the embedded information had tripled, the quality of the generated image obtained by the proposed method was still better. However, the results show that the proposed scheme sometimes caused content distortion in the secret image reconstruction stage. The content distortion can be seen in the last reconstructed image in the fourth row in Fig. 10, and in the third reconstructed image in the last row in the same figure. This was caused by the low robustness of the decryption algorithm.

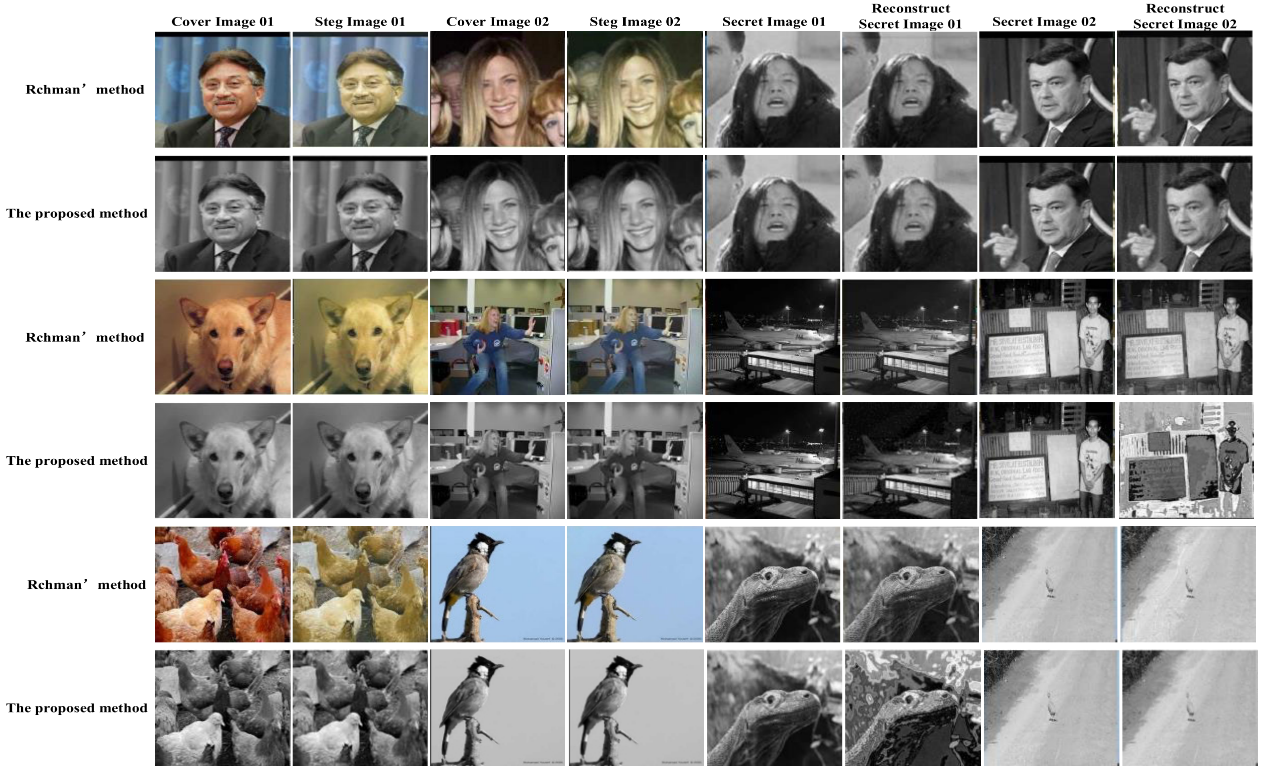


FIGURE 10. The comparison results of the proposed scheme and method proposed in [26]. The odd-numbered rows represent the experimental results of [26], including two examples. The even-numbered rows represent the experimental results of our scheme. The contrast content of the experiment is composed of secret images and reconstructed secret images.

As an objective measure of image quality, the *PSNR* evaluates image quality by calculating the error between the corresponding pixels. The larger the *PSNR* is, the smaller the image distortion. The *PSNR* is calculated by:

$$PSNR = 10\log_{10}\left(\frac{2^n-1}{MSE}\right)^2 \tag{11}$$

where *MSE* denotes the mean square error between the original image and the evaluated image, $(2^n - 1)^2$ is the square of the maximum signal value, and n is the number of bits of each sample value.

In order to evaluate the performance of the proposed method further, the *SSIM* (structural similarity index measure) was employed to measure the quality of the processed images, so as to obtain the optimal parameters. The *SSIM* evaluates the processed image quality by comparing the brightness, contrast, and structural similarity with the original image. The simplified form of *SSIM* is as follows:

$$SSIM(x,y) = \frac{(2u_xu_y + C_1)(2\sigma_{xy} + C_2)}{(u_x^2 + u_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{12}$$

where x represents the image blocks in the original cover image or secret image, y represents the image blocks in the steg image or reconstructed secret image, which take the same position and correspond to each other. The luminance similarity of images was measured by the average gray values u_x and u_y; the contrast similarity of images was estimated by the deviation information σ_x and σ_y; the structural similarity σ_{xy} was determined by the correlation between the image blocks x and y; C_1 and C_2 denoted the constants sets, and they were calculated as $C_1 = (K_1L)^2$ and $C_2 = (K_2L)^2$, respectively.

In order to reflect the difference between the proposed steganography scheme and other steganography schemes numerically, a comparison from three aspects: payload, *PSNR,* and *SSIM*, was conducted. The comparison results are given in Table 1. In terms of payload, the embedding rate of the proposed scheme was the same as that of the scheme proposed in [25], the difference was that the cover image and secret image in the proposed paper were grayscale images. the *SSIM* and *PSNR* of the generated steg image obtained The color images are utilized in [25]. The embedding rate in this article is three times that of the proposed method in [26]. Based on the comparison results, the proposed method is the highest. Although our embedding capacity has increased, the *SSIM* and *PSNR* of the secret image and the reconstructed secret image still have advantages compared with the method in [26]. Although the embedded information was small, the *SSIM* and *PSNR* of the generated steg image were the lowest, which was closely related to the network architecture and loss function. Therefore, in [26], both the embedding capacity and the color distortion needs to improvement for steganography. The proposed scheme performed better than the other two methods in image generation, which was closely related to the utilization of the generative adversarial nets. The secret image was encrypted before the embedment,

TABLE 1. Payload, SSIM and PSNR comparison for differenct steganography schemes.

Steganography scheme	Payload	Steg image	Reconstruct Secret	Steg image	Reconstruct Secret
		(SSIM)	(SSIM)	(PSNR)	(PSNR)
Ref [25]	100%	0.98	0.97	41.2	37.6
Ref [26]	33.3%	0.937	0.93	32.5	34.76
Ref [31]	100%	0.985	0.981	40.45	37.32
Ref [32]	100%	0.97	0.984	40.47	36.92
Ours	100%	**0.987**	0.953	**42.3**	**38.45**

TABLE 2. The comparison of SSIM and PSNR of differenct de-nosing methods.

	PSNR				SSIM			
	Original	Mean	Median	Gaussian	Original	Mean	Median	Gaussian
01	28.1376	31.1293	30.7698	33.1923	0.8877	0.9267	0.9176	**0.9771**
02	34.2862	35.7626	33.9022	36.2985	0.9544	0.9607	0.8927	0.9658
03	27.3930	28.5123	30.2345	30.9453	0.7547	0.8526	0.8147	0.9003
04	35.5256	35.0449	36.5364	37.1780	0.9591	0.9571	0.9733	**0.9834**
05	35.6794	36.5727	37.4394	37.6169	0.9510	0.9535	0.9745	0.9754
06	34.5574	35.5254	35.8898	37.1554	0.9279	0.9678	0.9773	**0.9804**

therefore, the encrypted secret image recovered by the extracting network required further decryption to extract secret image, which also led to the conclusion that *SSIM* and *PSNR* of the proposed scheme were lower than those of the method proposed in [25] in the reconstruction of the secret image, but compared to the method proposed in [26], there was a great improvement in terms of extraction of secret image. At the same time, we also compared with the newly proposed methods [31], [32]. For the methods proposed by [31], [32], the cover image and the secret image are all color image and have same size. The experimental results of the comparison are also shown in Table 1, it can be seen from the experimental results that our method surpasses the other four steganography methods in other aspects except that it is not dominant on the SSIM of the secret image. The noise points existed in the extracted secret image due to the decrypted operation, and different image de-noise methods were used to handle the decrypted secret image. The results are shown in Fig. 11.

As shown in Fig. 11, six groups of experiments were randomly selected to verify whether the de-noise process can improve the quality of the decrypted secret image. The mean filter, median filter, and Gaussian filter were adopted to handle the decrypted secret image. In Fig. 11, from the left to the right, the original secret image, decrypted secret image, secret image after mean filtering, secret image after median filtering, and secret image after Gaussian filtering are presented. From the perspective of visual effects, the mean and median filters could effectively restrain the noise of the decrypted secret image, but the semantic information of the secret image was blurred. On the other hand, the Gaussian filter could protect the details of the secret image while removing the noise.

In order to reflect the improvement effect brought by the de-noise operation more intuitively, a comparison from two aspects, *PSNR* and *SSIM*, was conducted. The comparison results are shown in Table 2, where it can be seen that after de-noising the decrypted secret image, *PSNR* and *SSIM* were improved. Also, *PSNR* and *SSIM* of the most secret images after Gaussian filtering obtained by the proposed scheme exceeded those of the method proposed in [25], which indicated that the de-noise operation could compensate for the shortcoming of the proposed steganography scheme.

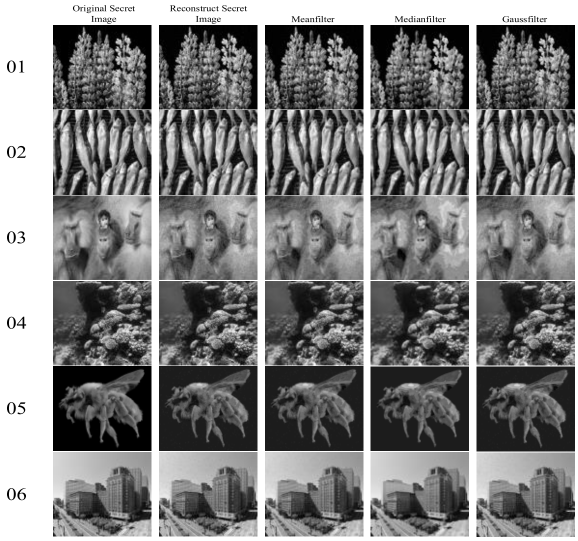


FIGURE 11. The experimental results after mean, median, and Gaussian filtering from left to right.

V. CONCLUSION

In this article, an encrypted steganography system based on chaos encryption and generative adversarial nets is proposed. The specific contributions are as follows: (1) The steganography system uses grayscale image as the cover image and only needs to consider the integrity of semantic content without considering color distortion. The problem of color distortion in the current method is perfectly solved. (2) To solve the problem of secret information leakage, the secret image to be hidden is firstly encrypted by chaos encryption technology, then the cover image and encrypted secret image are integrated to steg image. (3) The steganography system adds a discriminative network except hiding network and extracting network. (4) The hiding network adopts skip connections to concatenate the low-level features with the high-level features, which contributes to hiding the details of secret image preferably. In addition, a new weight allocation mechanism is introduced to ensure training procedure of hiding-extracting networks. The experimental results indicate that the steganography system proposed in this article can reasonably utilize all bits of the cover image to embed the secret image and there is no obvious appearance difference between cover image and steg image. The steganography system proposed in this article performs well on ImageNet, PASCAL-VOC12 and LFW image dataset. The experimental results also prove that steganography of double encryption can ensure the information of secret image will not be leaked during the process of transmission, and it can reconstruct the secret image perfectly.

REFERENCES

[1] C.-W. Lee, "Multipurpose protection for numeric data with capabilities of self-authentication and ownership declaration," *IEEE Access*, vol. 6, pp. 71152–71167, 2018.

[2] C. Qin, W. Zhang, F. Cao, X. Zhang, and C.-C. Chang, "Separable reversible data hiding in encrypted images via adaptive embedding strategy with block selection," *Signal Process.*, vol. 153, pp. 109–122, Dec. 2018.

[3] G. Gando, T. Yamada, H. Sato, S. Oyama, and M. Kurihara, "Fine-tuning deep convolutional neural networks for distinguishing illustrations from photographs," *Expert Syst. Appl.*, vol. 66, pp. 295–301, Dec. 2016.

[4] T. Pevný, T. Filler, and P. Bas, "Using high-dimensional image models to perform highly undetectable steganography," in *Proc. Int. Workshop Inf. Hiding*, in Lecture Notes in Computer Science, vol. 6387. Berlin, Germany, 2010, pp. 161–177.

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

[6] C. Mo, V. Sedighi, M. Boroumand, and J. Fridrich, "JPEG-phase-aware convolutional neural network for steganalysis of JPEG images," in *Proc. 5th ACM Workshop Inf. Hiding Multimedia Secur. (IH&MMSec)*, 2017, pp. 75–84.

[7] K. Bhowal, D. Bhattacharyya, A. J. Pal, and T.-H. Kim, "A GA based audio steganography with enhanced security," *Telecommun. Syst.*, vol. 52, no. 4, pp. 2197–2204, Apr. 2013.

[8] E. Satir and H. Isik, "A Huffman compression based text steganography method," *Multimedia Tools Appl.*, vol. 85, no. 10, pp. 2385–2394, 2014.

[9] W. Tang, S. Tan, B. Li, and J. Huang, "Automatic steganographic distortion learning using a generative adversarial network," *IEEE Signal Process. Lett.*, vol. 24, no. 10, pp. 1547–1551, Oct. 2017.

[10] X. Li, W. Zhang, X. Gui, and B. Yang, "A novel reversible data hiding scheme based on two-dimensional difference-histogram modification," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 7, pp. 1091–1100, Jul. 2013.

[11] M.-H. Lin, Y.-C. Hu, and C.-C. Chang, "Both color and gray scale secret images hiding in a color image," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 16, no. 6, pp. 697–713, Sep. 2002.

[12] S.-L. Li, K.-C. Leung, L. M. Cheng, and C.-K. Chan, "A novel image-hiding scheme based on block difference," *Pattern Recognit.*, vol. 39, no. 6, pp. 1168–1176, Jun. 2006.

[13] Y.-C. Hu, "High-capacity image hiding scheme based on vector quantization," *Pattern Recognit.*, vol. 39, no. 9, pp. 1715–1724, Sep. 2006.

[14] K. A. Zhang, A. Cuesta-Infante, L. Xu, and K. Veeramachaneni, "SteganoGAN: High capacity image steganography with GANs," 2019, *arXiv:1901.03892*. [Online]. Available: https://arxiv.org/abs/1901.03892

[15] J. Yang, K. Liu, X. Kang, E. K. Wong, and Y. Q. Shi, "Spatial image steganography based on generative adversarial network," 2018, *arXiv:1804.07939*. [Online]. Available: https://arxiv.org/abs/1804.07939

[16] Z. Zhang, G. Fu, F. Di, C. Li, and J. Liu, "Generative reversible data hiding by image to image translation via GANs," *Secur. Commun. Netw.*, vol. 2019, 2019, doi: 10.1155/2019/4932782.

[17] A. U. Islam, F. Khalid, M. Shah, Z. Khan, T. Mahmood, A. Khan, U. Ali, and M. Naeem, "An improved image steganography technique based on MSB using bit differencing," in *Proc. 6th Int. Conf. Innov. Comput. Technol.*, Aug. 2016, pp. 265–269.

[18] A. Rehman, T. Saba, T. Mahmood, Z. Mehmood, M. Shah, and A. Anjum, "Data hiding technique in steganography for information security using number theory," *J. Inf. Sci.*, vol. 45, no. 6, pp. 767–778, Dec. 2019.

[19] K. Chen, H. Zhou, W. Zhou, W. Zhang, and N. Yu, "Defining cost functions for adaptive JPEG steganography at the microscale," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 4, pp. 1052–1066, Apr. 2019.

[20] D.-C. Wu, C.-Y. Hsiang, and M.-Y. Chen, "Steganography via MIDI files by adjusting velocities of musical note sequences with monotonically non-increasing or non-decreasing pitches," *IEEE Access*, vol. 7, pp. 154056–154075, 2019.

[21] J. Wu, B. Chen, W. Luo, and Y. Fang, "Audio steganography based on iterative adversarial attacks against convolutional neural networks," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 2282–2294, 2020.

[22] J. Butora, Y. Yousfi, and J. Fridrich, "Turning cost-based steganography into model-based," in *Proc. ACM Workshop Inf. Hiding Multimedia Secur. (IH&MMSec)*, 2020, pp. 151–159.

[23] C. Li, Y. Jiang, and M. Cheslyar, "Embedding image through generated intermediate medium using deep convolutional generative adversarial network," *Comput. Mater. Continua*, vol. 56, no. 2, pp. 313–324, 2018.

[24] J. Wen, X. Zhou, M. Li, P. Zhong, and Y. Xue, "A novel natural language steganographic framework based on image description neural network," *J. Vis. Commun. Image Represent.*, vol. 61, pp. 157–169, May 2019.

[25] S. Baluja, "Hiding images in plain sight: Deep steganography," in *Proc. Neural Inf. Process. Syst.*, 2017, pp. 2066–2076.

[26] A. U. Rehman, R. Rahim, M. S. Nadeem, and S. U. Hussain, "End-to-end trained CNN encoder-decoder networks for image steganography," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 723–729.

[27] G. Xu, H.-Z. Wu, and Y.-Q. Shi, "Structural design of convolutional neural networks for steganalysis," *IEEE Signal Process. Lett.*, vol. 23, no. 5, pp. 708–712, May 2016.

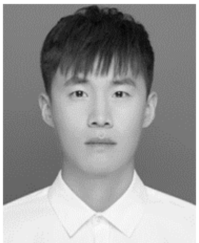
[28] V. Holub, J. Fridrich, and T. Denemark, "Universal distortion function for steganography in an arbitrary domain," *EURASIP J. Inf. Secur.*, vol. 2014, no. 1, pp. 1–13, Dec. 2014.

[29] X.-B. Kang, G.-F. Lin, Y.-J. Chen, F. Zhao, E.-H. Zhang, and C.-N. Jing, "Robust and secure zero-watermarking algorithm for color images based on majority voting pattern and hyper-chaotic encryption," *Multimedia Tools Appl.*, vol. 79, nos. 1–2, pp. 1169–1202, Jan. 2020.

[30] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

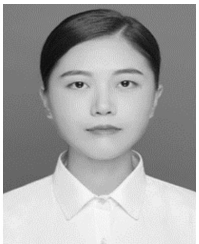
[31] X. Duan, K. Jia, B. Li, D. Guo, E. Zhang, and C. Qin, "Reversible image steganography scheme based on a U-Net structure," *IEEE Access*, vol. 7, pp. 9314–9323, 2019.

[32] D. Xintao and L. Nao, "Hide the image in FC-DenseNets to another image," 2019, *arXiv:1910.08341*. [Online]. Available: http://arxiv.org/abs/1910.08341



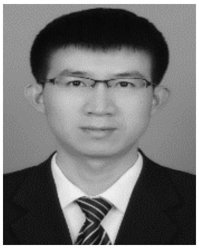
QI LI (Member, IEEE) was born in Jinan, China, in 1993. He received the B.S. degree in computer science and technology and the M.S. degree in computer application from the Qilu University of Technology (Shandong Academy of Sciences), Jinan, in 2016 and 2019, respectively. He is currently pursuing the Ph.D. degree in computer science and technology with Dalian Maritime University, Dalian, China. His main research interests include computer vision, machine vision, and information hiding.

XINGYUAN WANG received the Ph.D. degree in computer software and theory from Northeast University, China, in 1999. From 1999 to 2001, he was a Postdoctoral Researcher with Northeast University. He is currently a Professor of information science and technology with Dalian Maritime University, China. He has published three books and more than 400 scientific articles in refereed journals and proceedings. His research interests include nonlinear dynamics and control, image processing, chaos cryptography, systems biology, and complex networks.



XIAOYU WANG (Member, IEEE) was born in Heze, China, in 1994. She received the B.S. degree in computer science and technology and the M.S. degree in computer application from the Qilu University of Technology (Shandong Academy of Sciences), Jinan, China, in 2016 and 2019, respectively. She is currently pursuing the Ph.D. degree in computer science and technology with Dalian Maritime University, Dalian, China. Her main research interests include reversible data hiding, machine vision, and image processing.

BIN MA (Member, IEEE) received the M.S. and Ph.D. degrees from Shandong University, Jinan, China, in 2005 and 2008, respectively. From 2008 to 2013, he was an Associate Professor with the School of Information Science, Shandong University of Political Science and Law, Jinan. He visited the New Jersey Institute of Technology, Newark, NJ, USA, as a Visiting Scholar, from 2013 to 2015. He is currently an Associate Professor with the School of Information Science, Qilu University of Technology, Shandong, China. His research interests include reversible data hiding, multimedia security, and image processing.



CHUNPENG WANG (Member, IEEE) was born in Heze, China, in 1989. He received the B.E. degree in computer science and technology from Shandong Jiaotong University, China, in 2010, the M.S. degree from the School of Computer and Information Technology, Liaoning Normal University, China, in 2013, and the Ph.D. degree from the School of Computer Science and Technology, Dalian University of Technology, China, in 2017. He is currently a Teacher with the School of Cyber Security, Qilu University of Technology (Shandong Academy of Sciences), Jinan. His main research interests include image watermarking and signal processing.

YONGJIN XIAN was born in Dalian, China, in 1989. He received the B.E. degree in information and computing science from Nanjing Technology University, China, in 2012, and the M.S. degree in mathematics from Dalian Jiaotong University, China, in 2018. He is currently pursuing the Ph.D. degree in computer application technology with Dalian Maritime University, China. His main research interests include information security, information encryption, and information hiding.

YUNQING SHI (Life Fellow, IEEE) received the M.S. degree from Shanghai Jiao Tong University, China, and 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 authored/coauthored more than 300 articles, one book, and five book chapters. He was an Editor of ten books, three special issues, and 13 proceedings. He holds 30 U.S. patents. His research interests include data hiding, forensics and information assurance, visual signal processing, and communications. He is a member of a few IEEE technical committees. He was the Technical Program Chair of IEEE ICME 2007 and IEEE MMSP 2005, the Co-General Chair of IEEE MMSP02, and a Distinguished Lecturer of IEEE CASS. He has been the Co-Technical Chair of IWDW since 2006. He has served as an Associate Editor for the IEEE Transactions on Signal Processing and the IEEE Transactions on Circuits and Systems—II: Express Briefs. He serves as an Associate Editor for the IEEE Transactions on Information Forensics and Security, and an editorial board member for a few journals.