Article

An Image Style Transfer Network Using Multilevel Noise Encoding and Its Application in Coverless Steganography

Shanqing Zhang [1], Shengqi Su [1], Li Li [1,*], Qili Zhou [1], Jianfeng Lu [1] and Chin-Chen Chang [2]

[1] School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China
[2] Department of Information Engineering and Computer Science, Feng Chia University,
    Taichung 40724, Taiwan
*   Correspondence: lili2008@hdu.edu.cn

check for
updates

Abstract: Most of the existing image steganographic approaches embed the secret information imperceptibly into a cover image by slightly modifying its content. However, the modification traces will cause some distortion in the stego-image, especially when embedding color image data that usually contain thousands of bits, which makes successful steganalysis possible. A coverless steganographic approach without any modification for transmitting secret color image is proposed. We propose a diversity image style transfer network using multilevel noise encoding. The network consists of a generator and a loss network. A multilevel noise to encode matching the subsequent convolutional neural network scale is used in the generator. The diversity loss is increased in the loss network so that the network can generate diverse image style transfer results. Residual learning is introduced so that the training speed of network is significantly improved. Experiments show that the network can generate stable results with uniform texture distribution in a short period of time. These image style transfer results can be integrated into our coverless steganography scheme. The performance of our steganography scheme is good in steganographic capacity, anti-steganalysis, security, and robustness.

Keywords: convolutional neural networks; coverless steganography; image style transfer; multilevel noise encoding

1. Introduction

Information hiding technology hides secret information in the host signals in an invisible way and extracts it when needed [1]. Information hiding realizes the concealment of communication and copyright protection and has received increasing attention. Digital images are often used as a cover for information hiding. The traditional image information hiding is divided into spatial-based methods and transform domain methods according to different methods. Spatial-based information hiding mainly modifies the pixel data of an image, for example, replacing the least significant bit (LSB) of an image [2,3] or pixel-value difference (PVD) [4], modifying the statistical characteristics of the host image information [5]. Transform domain information hiding converts image into a corresponding transform domain, performing information hiding such as QT hiding method [6], discrete fourier transform (DFT) domain information hiding method [7], and discrete wavelet transform (DWT) domain information hiding method [8]. What the traditional information hiding methods do is hide the secret information by modifying the cover.

The traditional information hiding method inevitably leaves modification traces on the cover, causing statistical abnormality of the cover image, so that it cannot resist the steganalysis tools.

Coverless steganography has been proposed to resist steganalysis. Coverless does not mean that no cover is required, instead, the generation of secret-embedded information is directly driven by the secret information. The transmission of the secret information without modifying the cover becomes the basis of the coverless steganography. Zhang et al. [9] divided coverless steganography into semi-constructed coverless steganography and fully constructed coverless steganography. The semi-constructive steganography specifies a preset condition of cover construction and generates a dense cover according to the secret information while following certain construction rules. Different from the semi-constructive steganography, fully constructed steganography directly uses multiple objects in different normal images. It is driven by secret data, the objects are selected to construct the secret-embedded cover with reasonable content and reasonable statistical features directly.

A digital image contains not only the information of pixels, but also many characteristics such as brightness, color, texture, edge, contour, and high-level semantics. It is possible to organize the seemingly unrelated image information in a reasonable way, by hiding the secret information through some mapping relationship, if these characteristics are described in certain way. If the corresponding original image can be found through certain mapping relationship, the secret information transmission is achieved through the original image, so that the coverless steganography without modifying the cover is realized. Therefore, the key to coverless steganography is the mapping of image information to a codebook. The construction of coverless steganography codebook is shown in Figure 1. Different secret-embedded covers are generated by a certain mapping relationship from an image, one set (one or more secret-embedded covers) of secret-embedded covers is indexed to a codeword, and all covers construct the codewords of coverless steganography. The secret information transmission is realized by transmitting a secret-embedded cover corresponding to codewords.
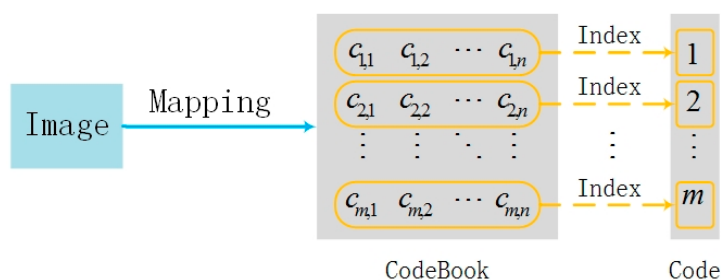

Figure 1. Construction of coverless steganography codewords.

The existing image coverless steganography methods use different mapping methods to construct the codewords corresponding to secret information. A codeword of coverless steganography is constructed using the mapping of the visual words (VW) of the encoding block of bag-of-words (BOW) model [10]. Otori et al. [11] mapped image texture to codewords using LBP encoding, Wu et al. [12] constructed codewords using block sorting to synthesize texture. Liu et al. [13] constructed the codewords of secret-embedded cover using generative adversarial networks (GAN).

Although the above new coverless steganography methods have different mapping methods to construct codewords, they have three characteristics in common: (1) Diversity: In order to avoid suspicion from network analysts, the secret-embedded covers are rich in forms according to the real scene; (2) Difference: There is difference between the secret-embedded covers corresponding to different codewords to ensure identifiability; (3) Completeness: There is at least one secret-embedded cover in each group in the codewords to ensure one-to-one correspondence with codeword index.

Deep learning brings new insight into image synthesis. Gatys et al. [14] proposed the image style transfer based on convolutional neural network (CNN) and found that CNN can be used to separate the content characteristics and the style characteristics of an image. By independently processing these high-level abstracted characteristics to achieve the image style transfer effectively, a rich artistic result is obtained. However, when the input is the same, the difference between the

multiple generated results is too small. This image style transfer is not suitable for constructing codewords for coverless steganography.

In order to realize the diversity and difference of the image style transfer results, Li et al. [15] proposed a simple method which used whitening and coloring transforms (WCTs) for universal style transfer, which enjoys the style-agnostic generalization ability with marginally compromised visual quality and execution efficiency. To increase diversity and enhance visual effects, Li et al. [16] proposed an image style transfer algorithm with added random noise based on CNN. Only one-dimensional random noise is used as the network input, the noise difference at different levels are not considered. Based on their work, we propose a diversity image style transfer network using of multilevel noise encoding and integrate it into a coverless steganography scheme. The network generates not only artistic effect image style transfer results, but also texture synthesis results with the expected difference and diversity, without changing the network structure. The network is used to construct coverless steganography codewords with various texture characteristics. The coverless steganography for secret information transmission is realized.

The main contributions of this paper include:

A generator structure of multilevel noise encoding that matches the subsequent visual geometry group-19 (VGG-19) network scale, the ability to synthesize multiple textures is enhanced.

- Diversity loss is used to prevent the network from falling into local optimization and allows the network to generate diversity image style transfer results.
- Residual learning is introduced to improve the network training speed significantly.
- Coverless steganography and image style transfer are combined, a coverless steganography scheme is presented. The performance of our coverless steganography scheme is good in steganographic capacity, anti-steganalysis, security, and robustness.

2. MNE-Style: Image Style Transfer Network Using Multilevel Noise Encoding

The structure of the diversity image style transfer network using multilevel noise encoding (MNE-Style) is shown in Figure 2. A generator and a loss network are connected in series. The generator uses a multilevel noise encoding that matches the scale of the loss network. The random details are generated by directly introducing the same scale explicit noise of the loss network in each layer to the generator. The loss network has two inputs, one for the output of the generator, the other is the content image. Different from the loss functions of other image style transfer networks, our loss function not only includes the style loss and content loss, but also increases the diversity loss. It improves the synthesis of diversity image style transfer results. The VGG-19 training network is used.
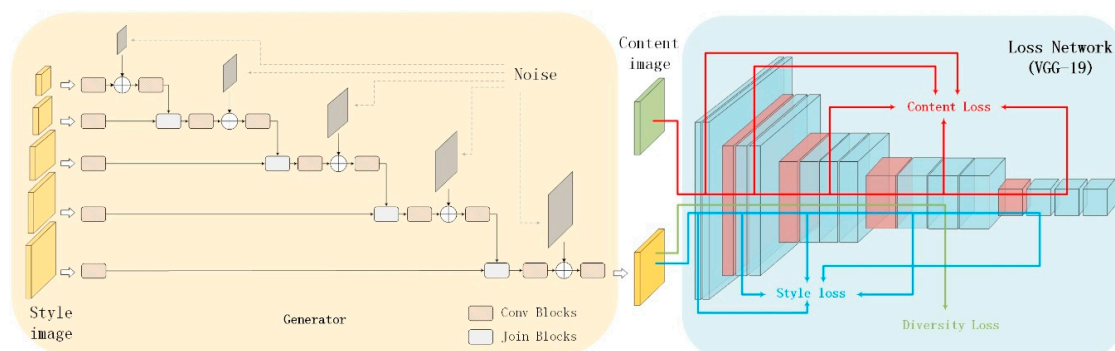

Figure 2. Structure of multilevel noise encoding(MNE)-Style.

2.1. Generator using Multilevel Noise Encoding

Inspired by Li et al. [16] and GAN network [17], we achieve the diversity of image style transfer results by adding multilevel noise to the generator. There are two problems in encoding noise in image

style transfer network: (1) In the network, different styles statistical characteristics represent different characteristic descriptions on different characteristic layers. It is difficult to add noise uniformly due to big differences among them. (2) The noise vector in the network is over-fitting due to easy marginalization, so that the network output is uniform.

The generated texture synthesis networks usually have repeated structure (as shown in Figure 3) by early texture synthesis networks [18,19] although two different scales of noises are introduced to add randomness to the texture synthesis results. It is difficult to generate pseudo-random numbers in the network that do not have significant periodic repetition but still maintain spatial variations. We solved this problem by adding pixel-by-pixel noise to the characteristic images of different scales in the network.
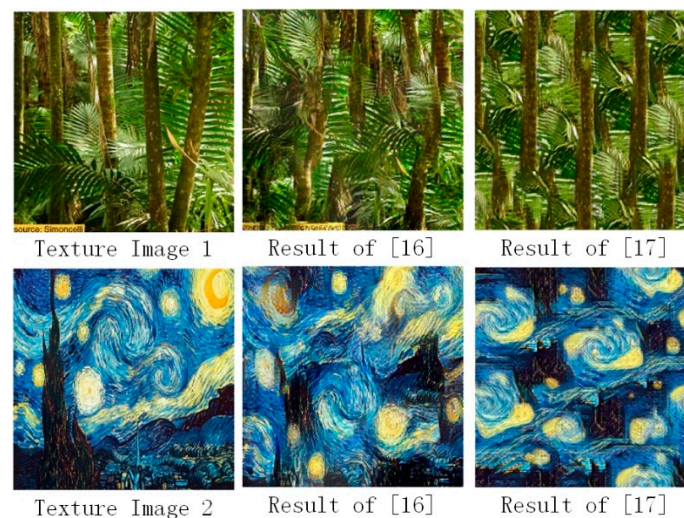
Figure 3. Periodic repeated texture from early texture synthesis networks.

StyleGAN [20] is a GAN network for synthesizing high quality facial images. In order to generate diversity facial images, the synthetic network uses the adaptive instance normalization (AdaIN) to add a scaled display noise to the synthetic results of different scales. Inspired by styleGAN, we propose a new generator using multilevel noise encoding, as shown in Figure 2. The structure first down-samples the style image to obtain a set of down-sampled results of the same size as the feature map of loss network. Then, the explicit noise inputs of the same scale are added to create random variations. Next, the results of the different scales and convolution operations are used to obtain the noise-coded style image. In the generator, the join blocks contain three parts: up-sample, batch norm, and concat. In the convolution blocks, one round has four part: padding, convolution, regularization and rectified linear unit (ReLU). The convolution blocks contain three rounds with kernel sizes of 3×3, 3×3 and 1×1, respectively. The generator uses multilevel noise encoding to solve the problem of local effects on the synthesis result, while introducing diversity into the style image.

2.2. Loss Function

Inspired by Li et al. [16], the loss function in this study consists of content loss, style loss, and diversity loss.

2.2.1. Content Loss

Like the definition of content loss by Gatys et al. [14], the advanced characteristic representation calculated by the VGG network extracts the content characteristics of the image. This content characteristic representation does not rely on pixel-by-pixel contrast but can preserve the content and structure of the image in high-level characteristics, but not the color texture and precise shape. The

content loss is used to ensure the image style transfer results close to the content image C. The content loss function for each layer is shown in (1):

$$E_{content}^l = \frac{1}{2} \sum \left(F_{ij}^l - C_{ij}^l\right)^2. \tag{1}$$

where F_{ij}^l is the activation of the *i*-th filter at *j*-th position in the *l*-th layer of the loss network of the image style transfer result, C_{ij}^l is the activation of the *i*-th filter at *j*-th position in the *l*-th layer of the loss network of the image style transfer result.

The total content loss is shown in (2):

$$L_{content} = \sum_{l=0}^{L} \varphi_l E_{content}^l. \tag{2}$$

where φ_l is the weight coefficient of different layers contributing to the total loss, L is the number of model layers used for content loss. In the experiment, the content loss takes the activation of the Conv1_2, Conv2_2, Conv3_2, Conv4_2, and Conv5_2 layers of the VGG-19 model, $L = 5$.

2.2.2. Style Loss

In the traditional image style transfer network, the style loss based on the difference of the texture statistical characteristics (Gram matrix). Since the scales of Gram matrices of different textures are different, we use the improved gram matrix in Li et al. [16] to calculate (3):

$$\overline{G}_{ij}^l = \sum_k \left(F_{ik}^l - \overline{F}^l\right)\left(F_{jk}^l - \overline{F}^l\right). \tag{3}$$

where F_{ik}^l is to the activation of the *i*-th filter at the *k*-th position in the current *l*-th layer of the loss network, and $\overline{F}^l$ is the average value of all the activations in the current *l*-th layer of the lost network.

The style loss for each layer is shown in (4):

$$E_{style}^l = \frac{1}{4M_l^2 N_l^2} \sum_{ij} (\overline{G}_{ij}^l - \overline{S}_{ij}^l). \tag{4}$$

where $\overline{G}_{ij}^l$ is the improved gram matrix of the generated style transfer image in the *l*-th layer, $\overline{S}_{ij}^l$ is the improved gram matrix of the style image in the *l*-th layer, M_l and N_l is the scale size of the feature map for the current *l*-th layer.

The total style loss is shown in (5):

$$L_{style} = \sum_{l=0}^{L} \omega_l E_{style}^l. \tag{5}$$

where ω_l is the weight coefficient contributing to the total loss from different layers, L is the number of used model layers. In experiment, style loss is the activation of layer Conv1_1, Conv2_1, Conv3_1, Conv4_1, and Conv5_1 of the VGG model, $L = 5$

2.2.3. Diversity Loss

Image style transfer networks that use only content loss and style loss usually produce image style transfer results with similar visual effects, because content loss and style loss are only for the synthesis results with highly similar contents and styles, but do not guarantee the diversity of output. The preservation of content and style characteristics by loss function surpasses the effect from the

noise. It does not help to obtain the synthesis of diverse image style transfer. Li et al. [16] proposed a diversity loss that aims to synthesize image style transfer results related to the input noise.

Diversity loss produces diversity output results by measuring differences among high-level characteristics of the same texture under different noises. There are N input sample in a layer of the network, and N output result is generated after the layer is processed, are $\{P_1, P_2, \ldots, P_N\}$. The diversity loss calculates the difference between any two outputs in high-level characteristics. $\{Q_1, Q_2, \ldots, Q_N\}$ described as a rearrangement sequence of $\{P_1, P_2, \ldots, P_N\}$, and satisfied $P_i \neq Q_i$. The diversity loss is shown in (6):

$$L_{style} = \sum_{l=0}^{L} \omega_l E_{style}^l.$$ (6)

where λ is the step size of diversity loss. The smaller the step size is, the greater is the difference in the diversity of the synthetic results. In order to make the generated results have a high level of diversity, the feature map of Conv4_2 is selected to calculate the diversity loss.

The image style transfer network of the multilevel noise encoding can realize diversity texture synthesis without adjusting the network structure (refer to Section 2.4). The effect of the diversity loss function on the diversity texture synthesis results is shown in Figure 4. The input texture image is in the left side, the texture synthesis result with diversity loss function is in the upper-right corner, the texture synthesis result without diversity loss function is in the bottom-right corner. The difference between adjacent results with diversity loss function is obvious, and the texture synthesis results are diverse.
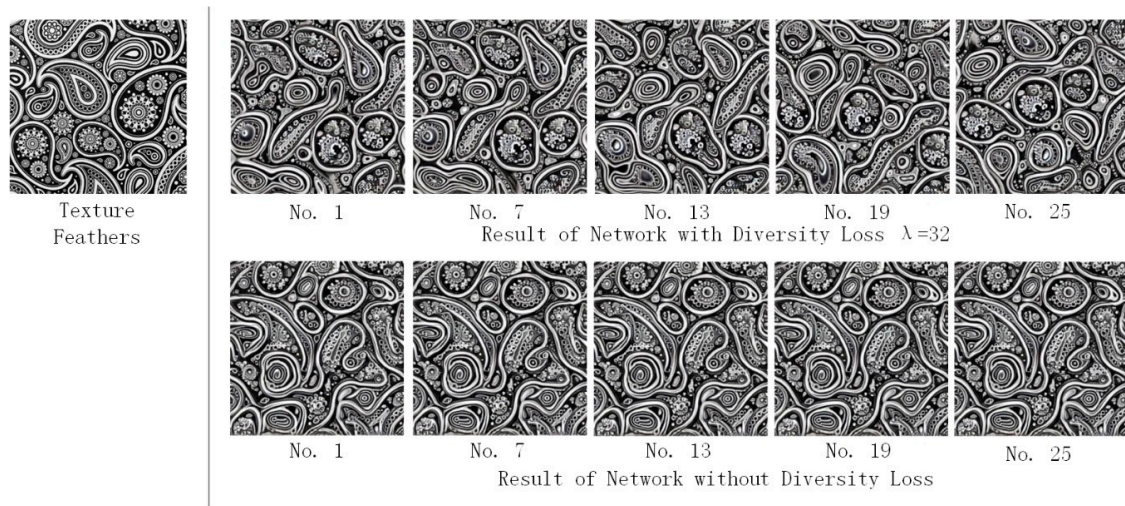

Figure 4. The effect of diversity loss on style transfer.

The step size of diversity loss λ also has an impact on the diversity of the synthesis results. The structural similarity index (SSIM) between adjacent synthesis results is shown in Figure 5. The SSIM is shown in (7).

$$SSIM(x, y) = \frac{\left(2\mu_x\mu_y + c_1\right)\left(2\sigma_{xy} + c_2\right)}{\left(\mu_x^2 + \mu_y^2 + c_1\right)\left(\sigma_x^2 + \sigma_y^2 + c_2\right)}.$$ (7)

where μ_x is the average of x, μ_y is the average of y, σ_x^2 is the variance of x, σ_y^2 is the variance of y, σ_{xy} is the covariance of x and y, $c_1 = (k_1 L)^2$ and $c_2 = (k_2 L)^2$ are constants used to maintain stability. L is the dynamic range of pixel values. Here $k_1 = 0.01$ and $k_2 = 0.03$.
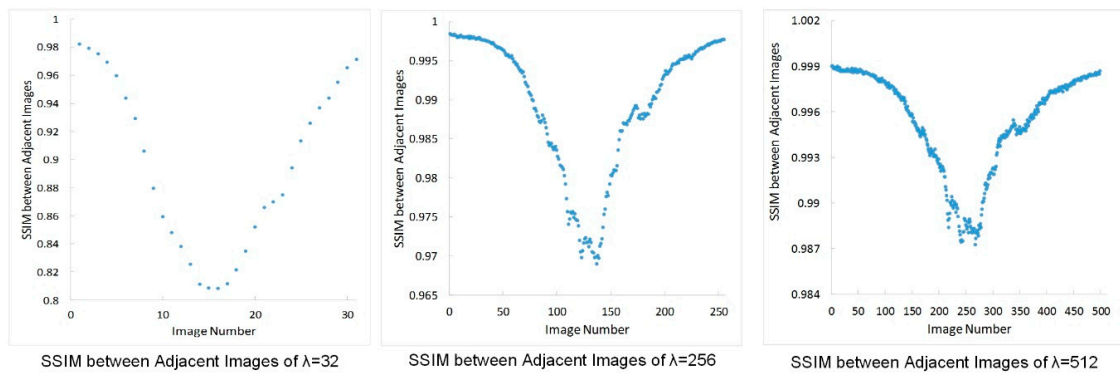
Figure 5. The relationship between structural similarity index (SSIM) and step size of diversity loss between adjacent synthesis results.

As shown in Figure 5, the smaller the step size λ is, the greater the difference between adjacent texture synthesis results will be.

2.2.4. Total Loss

The sum of the above-mentioned content loss, style loss, and diversity loss form a total loss function. In order to make the generated image style transfer result in line with human visual habits, a total variation regularization term [21–23] is added to the total loss function, as shown in (8):

$$L_{TV} = \sum_{i=1}^{h-1} \sum_{j=1}^{w} \sum_{k=1}^{d} \left(I_{i+1,j,k} - I_{i,j,k}\right)^2 + \sum_{i=1}^{h} \sum_{j=1}^{w-1} \sum_{k=1}^{d} \left(I_{i,j+1,k} - I_{i,j,k}\right)^2. \tag{8}$$

where L is the image style transfer result, its size is $h \times w \times d$.

The total loss function of MNE-Style network is:

$$L_{total} = \alpha L_{content} + \beta L_{style} + \gamma L_{diversity} + \tau L_{TV}. \tag{9}$$

where α, β, γ, and τ are undetermined constants. Used in this experiment, $\alpha = 3.0 \times 10^{-3}$, $\beta = 1.0$, $\gamma = -0.6$, and $\tau = 1.0 \times 10^{-4}$.

2.3. Residual Learning

He et al. [24] proposed image classification using residual learning to train deep learning networks, which proves that residual learning can guide the network to learn specific tasks easily. Johnson et al. [20] introduced residual learning into the image style transfer network. They presented a residual block (as shown in Figure 6a) to replace the basic convolutional block in VGG-16 (as shown in Figure 6b). Deep learning network used in this study is VGG-19 which is more suitable than VGG-16 for training with residual learning. The residual block of Johnson et al. [22] is used for training MNE-Style.
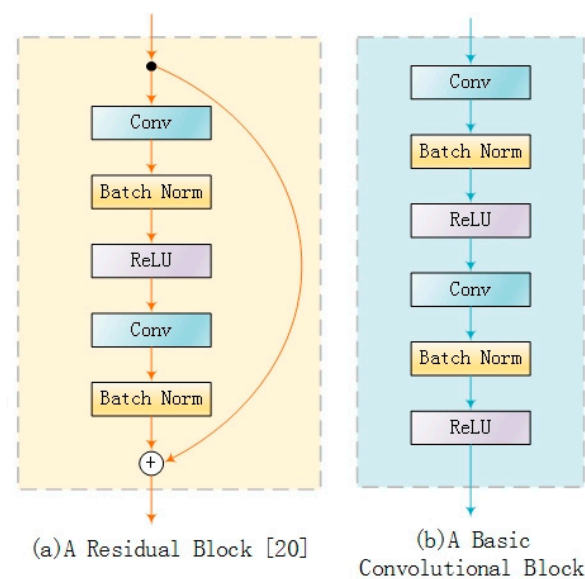
Figure 6. A residual block and an equivalent convolution block.

Our experiment showed improved convergence speed of the image style transfer network with residual learning. The results are shown in Table 1.

Table 1. Time of the convergence of network training.

Method	Time (min)
MNE-Style	24
MNE-Style without residual learning	53

2.4. Diversified Texture Synthesis

The diversity texture synthesis network and the image style transfer network need to adjust the structure of the generator network [16]. No operation is performed on the content image, only the style image and noise are re-encoded in the generator in this study. It is not necessary to reconstruct the generator when obtaining various texture synthesis results. The style image serves as the content image for the network and the content loss weight is set $\alpha = 0$ when the loss network produces a variety of texture synthesis. In short, the network easily separates the diversity texture synthesis and image style transfer operations without reconstructing the network structure. It provides a convenient condition for constructing the information hiding codewords.

3. The Integration Scheme of Coverless Steganography

As a branch of coverless steganography, image coverless steganography needs to consider security and anti-steganalysis. The key to the coverless steganography is the mapping between the image information and the codewords corresponding to the secret information. Codewords generated by different mapping methods have different characteristics such as diversity, difference, and completeness. The MNE-Style proposed in Section 2 can synthesize rich texture results and image style transfer results in a relatively short time. These results can be re-generated stably through the network with a few specific parameters. These rich and varied image style transfer results are suitable for codewords with coverless steganography. We propose a new coverless steganography method based on MNE-Style (as shown in Figure 7).
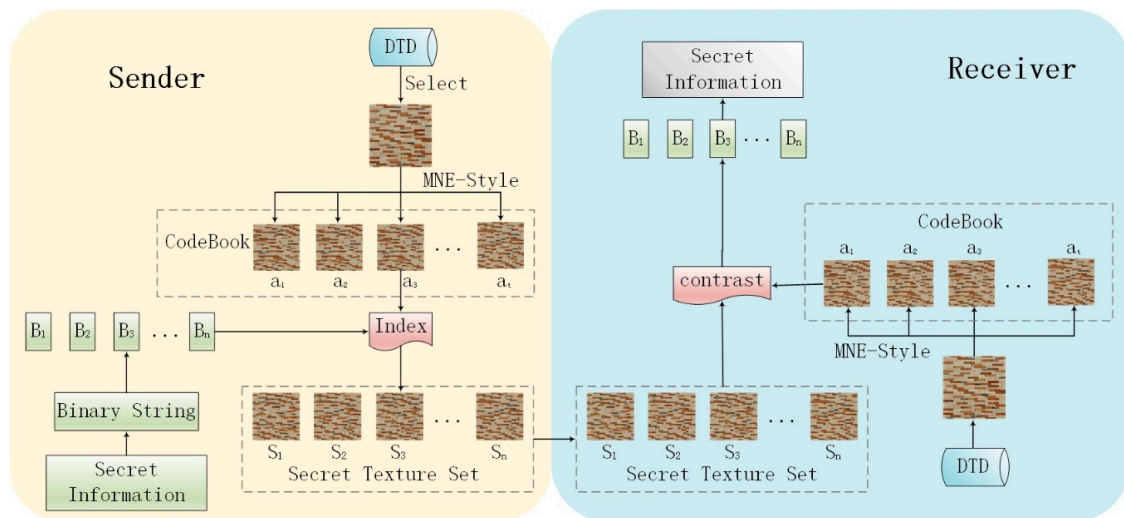
Figure 7. Coverless steganography method based on MNE-Style.

3.1. Hiding Method

The hidden method and extraction method of secret information is the focus of steganography method. When hiding is performed, we should consider how to encode the secret-embedded cover generated by MNE-Style, and how to convert the secret texts into image. The image style transfer result generated by the image style transfer network contains not only style information but also content information. In order to reduce the amount of data that needs to be transmitted, the describable textures dataset (DTD) image dataset is selected in this study. In the image style transfer network, only the texture synthesis result is used for the coverless information hiding. As shown in the sender side of Figure 7, the specific steps of the hiding are as follows:

Step 1. One texture image from the public DTD dataset is selected and input into the MNE-Style network. The image step size of diversity loss λ, diversity loss weight γ, and style loss weight β are specified. The diversify texture synthesis is performed. The collection of these diverse texture synthesis results is used as a codebook.

Step 2. According to the step size of diversity loss, the various texture synthesis results generated by the MNE-Style are numbered in order. For example, $\lambda = 256$, the diversity texture synthesis result is sequentially marked as 0000 0000, 0000 0001, 0000 0010, 0000 0011, ..., 1111 1111 according to the output numbers 1, 2, 3, 4, ..., 256. These 8-bit code streams are mapped to codes corresponding to the codebook.

Step 3. The secret information is segmented according to the codeword number in codeword book. The segmentation results are B_1, B_2, ..., B_n, the corresponding image set is found in the codeword book according to the results of B_1, B_2, ..., B_n. In the order of B_1 to B_n, the to-be-transmitted information hiding codeword set constitutes.

3.2. Extraction Method

The extraction process is as shown in the receiver of Figure 7. The extraction steps are as follows:

Step 1. The closest original texture image is selected in the DTD according to the received image information of the secret texture image set.

Step 2. The selected DTD texture image is used as an input of the MNE-Style network, and the coverless steganography codeword book is synthesized via network according to received network parameters. Network parameters include the image step size of diversity loss λ, diversity loss weight γ, and style loss weight β.

Step 3. The various texture synthesis results generated by the MNE-Style network are numbered in order according to the image diversity generation step size.

Step 4. The peak signal to noise ratio (PSNR) of each image in the secret texture image set and each image in the codeword book are calculated. The image is considered to correspond to the secret information represented by the current image of the codeword book when the value is the largest (or infinite). All secret information segments are connected in order to obtain the hidden texts.

4. Experimental Results

Two image transformation tasks, namely image style transfer and diversity texture synthesis, are completed (the details are in Sections 4.1 and 4.2). In addition, the stability, capacity, anti-steganalysis, and security of the coverless steganography based on the MNE-Style are analyzed (the details are in Section 4.3, Section 4.4and Section 4.5).

4.1. Diversity of Image Style Transfer Results

The goal of image style transfer is to synthesize an image that has the same contents of the original content image and the same style of the original style image. The MNE-Style is trained by Microsoft COCO (MSCOCO) dataset [25]. 10,000 images are randomly selected from the verification set of MSCOCO2014 as the training set. There are two cycles in the training, each cycle has 1000 iterations (refer to Figure 8, 99.91% of the style image iterations converge about 950 times). The initial learning rate is 0.001. The activation of Conv1_1, Conv2_1, Conv3_1, Conv4_1, and Conv5_1 of the VGG-19 are selected as the style. The activation is selected as the content characteristic of the Conv1_2, Conv2_2, Conv3_2, Conv4_2, and Conv5_2 layers of the VGG-19. The activation of the Conv4_2 is used to calculate the diverse texture loss. Torch7 and cuDNN9.1 are used in our experiment. It took about 24 minutes of training on the NVIDIA GeFoce GTX 1060 6GB.
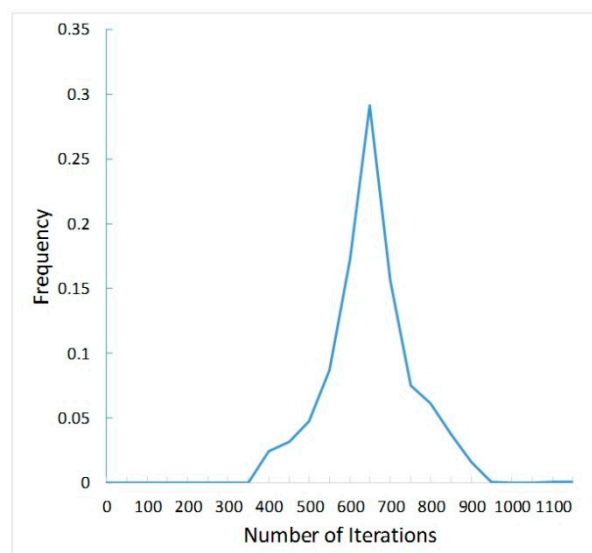
Figure 8. Statistics of the number of iterations required for network training to convergence.

Several different style images are selected for training, the results are compared to a representative parametric image style transfer method and a nonparametric image style transfer method. As shown in Figure 9, our results are not only richer than the image style transfer network results represented by the parameter method [14], but they also have better visual effect than those by Chen et al. [26]. The non-parametric image style transfer network results represent a more artistic style, and the content and style of an image is well balanced.
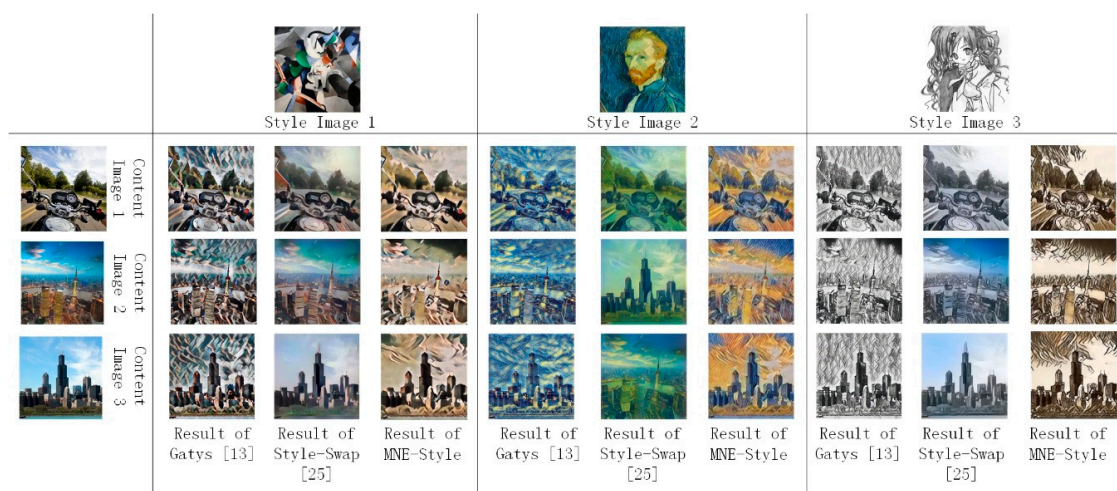
Figure 9. Comparison of the visual effects of the MNE-Style and the existing image style transfer network.

4.2. Network Speed

The MNE-Style is used to construct a codeword book with coverless steganography, so network training speed and image style transfer speed should be considered.

As for network training speed, Gatys et al. [14] did not use any network optimization, so the time of calculation is long. Johnson et al. [22] introduced network optimization based on Gatys et al. [14]. Although the time consumption is decreased to some extent, certain repetitions occurred, which led to image quality being decreased. Li et al. [16] decreased the data size by optimizing the style representation parameters and decreased the time consumption while achieving better visual effects. MNE-Style use the residual learning based on Li et al. [14] and further decrease the network training time. This provides a time guarantee for constructing steganography codewords. The comparison of results is shown in Table 2.

Table 2. Average training time for a style transfer network.

| Method	N.Iters	Time/Iter. (min)	Time (min)
Gatys et al. [14]	10,000	0.099	990
Johnson et al. [22]	2000	0.126	252
Li et al. [16]	1000	0.078	78
MNE-Style	1000	0.024	24

The computational time required for image style transfer with Gates et al. [14], Johnson et al. [22], and Li et al. [16] and MNE-Style are shown in Table 3. These four methods are all parametric image style transfer methods. Ours are faster than all the others.

Table 3. Average computational time for a style transfer network.

| Method	N.Iters	Time/Iter. (s)	Time (s)
Gatys et al. [14]	800	0.045	26
Johnson et al. [22]	300	0.050	15
Li et al. [16]	200	0.023	4.6
MNE-Style	200	0.007	1.4

4.3. Diversity Texture Synthesis Results and Robustness of Steganography

In order to make sure that a set of diversity texture images generated by the MNE-Style meet the requirements of the unsupported steganography codewords, there should be differences among different secret-embedded covers. In order to balance the texture diversity synthesis results and the robustness of steganography, the step size and robustness are balanced through experiment. Five hundred texture images are randomly selected from the DTD dataset, synthesized into the diversity texture images as codewords of coverless steganography by the MNE-Style. In order to encode the binary secret information with different lengths, the number of diversity texture synthesis is 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, respectively. The obtained diversity texture results are attacked by salt and pepper noise, brightness adjustment, and grayscale conversion with different parameters. The effects of one synthesis texture result after attack are shown in Figure 10. The relationship between the correct extraction rate and step size of diversity under attacks with different parameters is shown in Figure 11. Our method is robust to grayscale conversion, and it can maintain good robustness and encode more secret information in the cases of salt and pepper noise and brightness adjustment when step size of diversity loss $\lambda = 256$ is generated. The step size of diversity loss is 256. This setting remains in the experiment in the next sections.
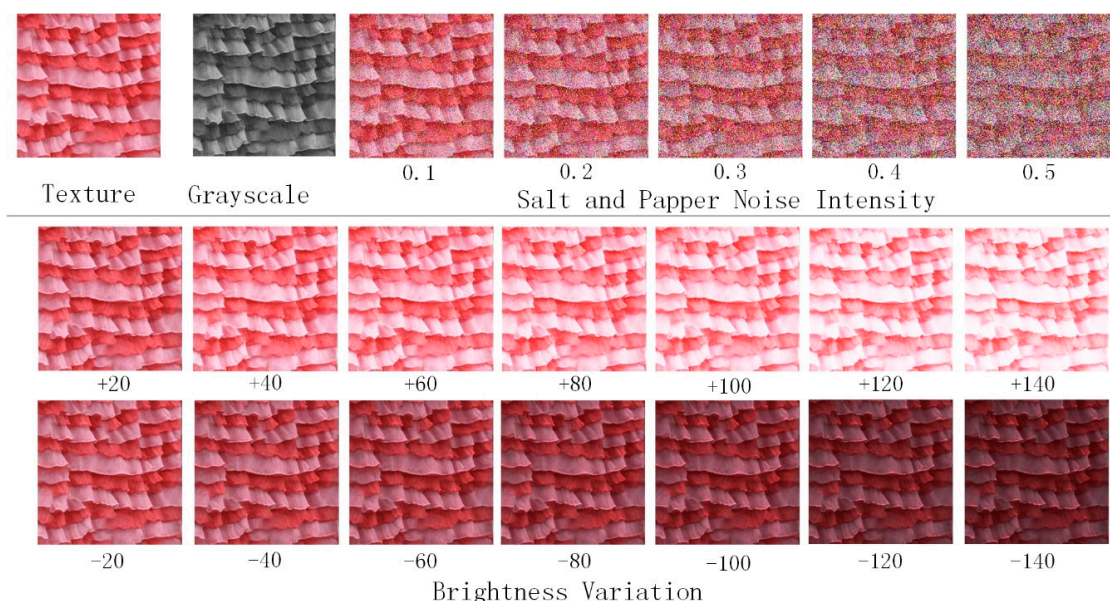

Figure 10. The effect of different attack.

The comparison of SSIM of the adjacent composite images between our MNE-Style network and the network in Li et al. [16] is shown in Figure 12. There are differences among the images generated by our network, the SSIM of the adjacent images is approximately gaussian distributed. The textures generated with different standard texture images by our network are shown in Figure 13. MNE-Style not only stably maintains the texture characteristics of color, shape, and structure, but also generates rich texture images.
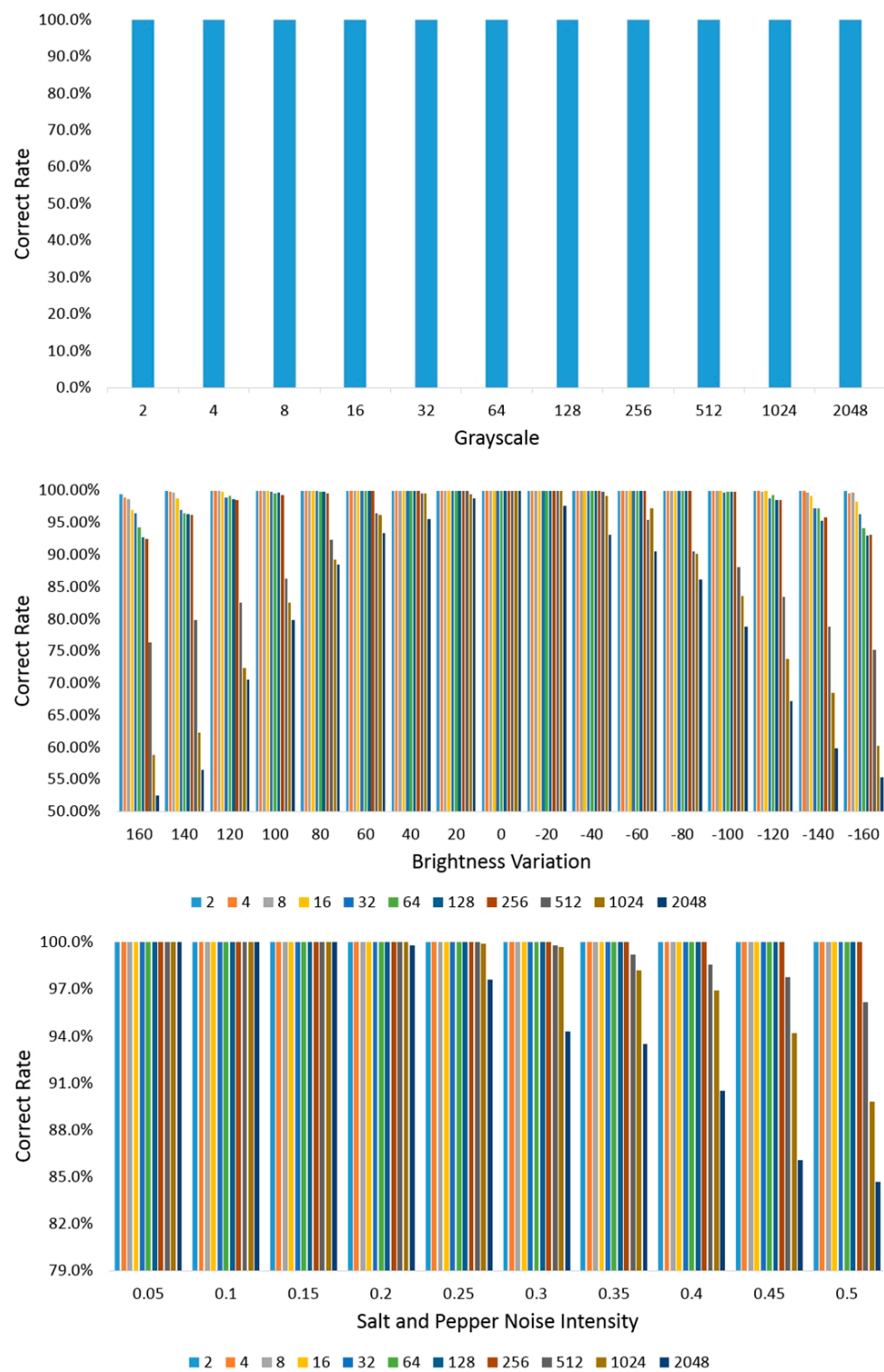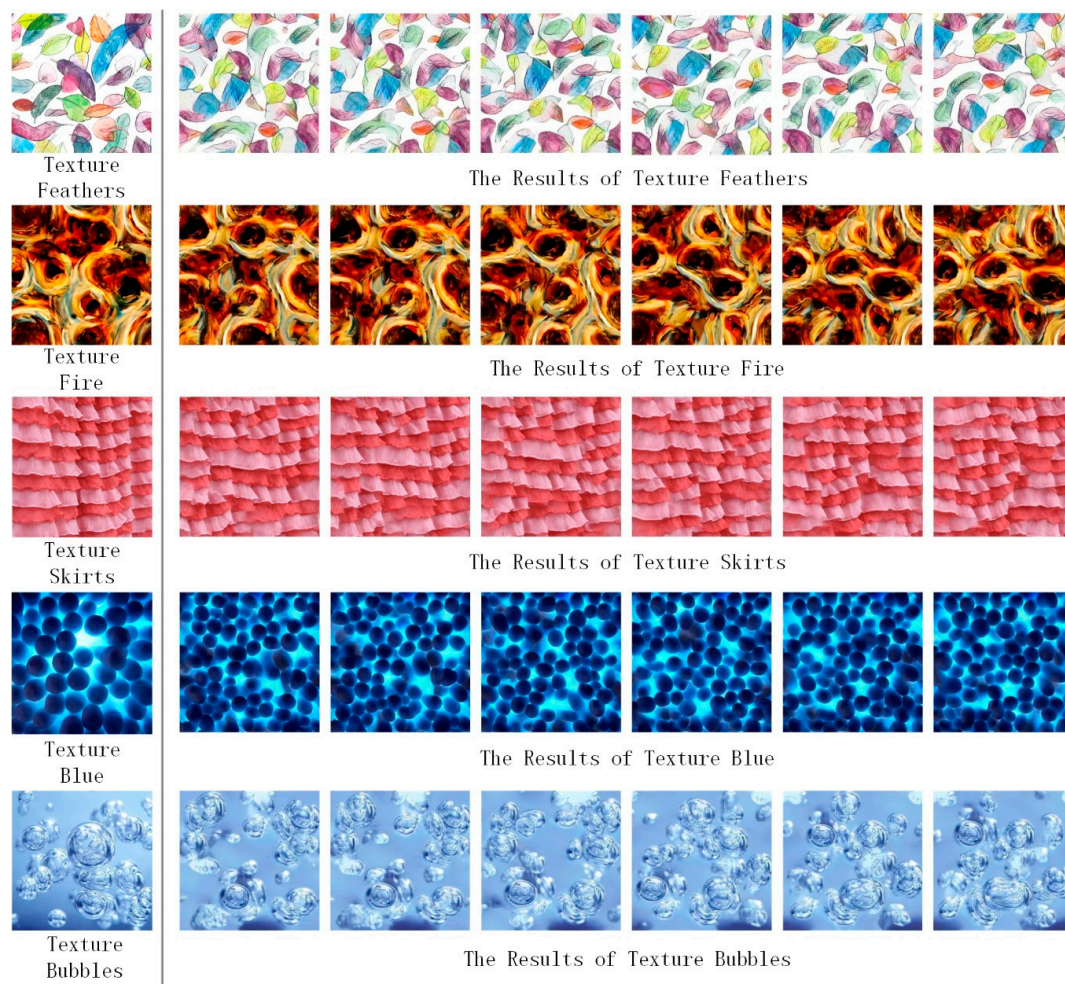
Figure 11. The relationship between the correct extraction rate and step size of diversity loss under attacks with different parameters.
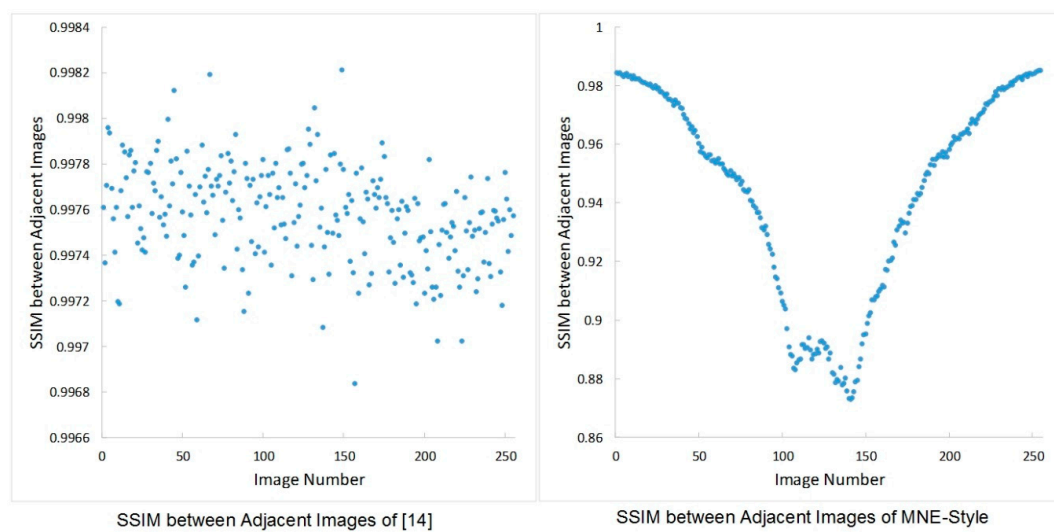
Figure 12. SSIM between adjacent images of different networks.

Figure 13. Diversity texture synthesis results of different texture images.
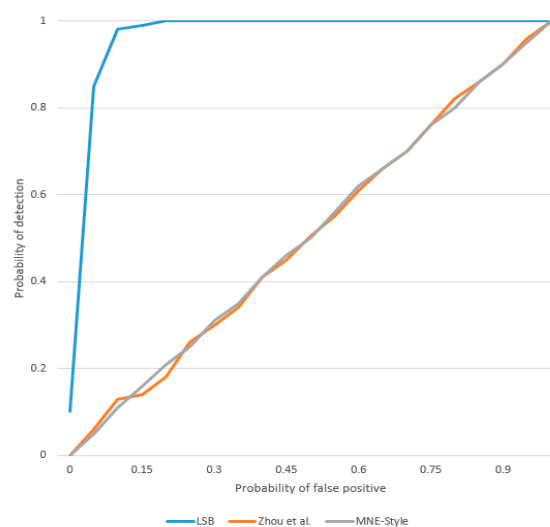
4.4. The Capacity of Coverless Steganography

Since the current image coverless steganography is not mature and is not comparable to traditional steganography in capacity, the comparison is performed only among coverless steganography themselves. Cao et al. [27] relied on the number of masks of Mask R-CNN, did not consider codewords, and used the most accurate mask as the secret information. They can hide 5-bit secret information. Zhou et al. [28] pointed out that an image can hide 8-bit secret information. Zhou et al. [10] used the BOW model, each image can hide 16–32 bits secret information without considering the codewords. The number of secret information bits tightly depends on codewords information [13,29], so we do not estimate their case nor compare. The capacity in our method is 8 bits when $\lambda = 256$. If the step size of diversity loss λ increases, the number of the encoded digits increases but the difference between adjacent images decreases accordingly. Select $\lambda = 256$ according to the experiment in Section 4.3. The specific comparison results are shown in Table 4.

Table 4. Comparison of capacity of coverless steganography.

Method	Capacity(bits)
Cao et al. [27]	5
Zhou et al. [28]	8
MNE-Style	8
Zhou et al. [10]	16–32

4.5. Anti-Steganalysis and Security

The mapping between the secret information and the coverless steganography codewords is generated by the MNE-Style in this study. The secret information hiding is transformed into the search for the diversity textures that meet the same condition. No modification of the secret-embedded image is performed in the process of the secret information transmission. The receiver operating characteristic (ROC) curve is used to test the resistance of different methods to steganalysis algorithms. We apply typical steganalysis algorithms which are improved by the standard Pairs method [30] to detect LSB replacement, Zhou et al. [10] and MNE-Style. In the experiment, we used 10% LSB steganography, and the step size of MNE-Style is 256. As shown in Figure 14, MNE-Style and Zhou et al.'s method [10] resist the detection from statistics-based steganalysis algorithms.

Figure 14. The receiver operating characteristic (ROC) curves of different methods.

This paper studies the coverless steganography, and the diversified texture image which is mapped to the secret information is the texture image directly generated by the MNE-Style. The texture image generated in this study maintains good texture characteristics of the original image (refer to Section 4.3) and can be disguised as an ordinary image for transmission. Comparing with the traditional encryption and steganography methods, our method is difficult to be attacked because our images are less suspicious to attackers.

Even if an attacker suspects the transmitted image contains secret information, he/she does not have the same diversity image style transfer network model nor parameters as the authentic communication partners have, he/she cannot generate the coverless steganography, thus cannot get the secret information that is the unique mapping of the codewords. Therefore, security is ensured.

5. Conclusions

A new steganography method with coverless steganography and image style transfer method based on combined CNN is proposed. A diversity image style transfer network using multilevel noise encoding mechanism (MNE-Style) is designed. The network consists of a generator and a loss network. A multilevel noise encoding matching the subsequent VGG-19 network scale is proposed in the generator to improve the sensitivity to noise. In the loss network, diversity image style transfer results are generated using diversity loss. The residual learning is introduced to improve the network training speed significantly. Our network can realize diversity texture synthesis and image style transfer without modifying the structure operating. Experiments show that our network can quickly generate stable diversity results with guaranteed visual quality. These results can be used as coverless steganography codewords and meet the requirements of security and anti-steganalysis. However, we found that in the experiment, the images with particularly dense texture are not suitable for dataset of steganography. How to filter texture images that meet the requirements will be in our future study.

References

1. Shen, C.; Zhang, H.; Feng, D.; Cao, Z.; Huang, J. Survey of information security. *Sci. China Ser. F Inf. Sci.* **2007**, *50*, 273–298. [CrossRef]
2. Wu, H.C.; Wu, N.I.; Tsai, C.S.; Hwang, M.S. Image steganographic scheme based on pixel-value differencing and LSB replacement methods. *IEE Proc. Vis. Image Signal Process.* **2005**, *152*, 611–615. [CrossRef]
3. Zakaria, A.; Hussain, M.; Wahab, A.; Idris, M.; Abdullah, N.; Jung, K.H. High-Capacity Image Steganography with Minimum Modified Bits Based on Data Mapping and LSB Substitution. *Appl. Sci.* **2018**, *8*, 2199. [CrossRef]
4. Bender, W.; Gruhl, D.; Morimoto, N.; Lu, A. Techniques for data hiding. *IBM Syst. J.* **1996**, *35*, 313–336. [CrossRef]
5. Ni, Z.; Shi, Y.Q.; Ansari, N.; Su, W. Reversible data hiding. *IEEE Trans. Circuits Syst. Video Technol.* **2006**, *16*, 354–362.
6. Li, X.; Wang, J. A steganographic method based upon JPEG and particle swarm optimization algorithm. *Inf. Sci.* **2007**, *177*, 3099–3109. [CrossRef]
7. Mckeon, R.T. Strange fourier steganography in movies. In Proceedings of the IEEE International Conference on Electro/information Technology, Chicago, IL, USA, 17–20 May 2007.
8. Chen, W.Y. Color image steganography scheme using set partitioning in hierarchical trees coding, digital Fourier transform and adaptive phase modulation. *Appl. Math. Comput.* **2007**, *185*, 432–448. [CrossRef]

9.   Zhang, X.; Qian, Z.; Li, S. Prospect of Digital Steganography Research. *J. Appl. Sci.* **2016**, *34*, 475–489.

10. Zhou, Z.; Cao, Y.; Sun, X. Coverless Information Hiding Based on Bag-of-WordsModel of Image. *J. Appl. Sci.* **2016**, *34*, 527–536.

11. Otori, H.; Kuriyama, S. Texture Synthesis for Mobile Data Communications. *IEEE Comput. Graph. Appl.* **2009**, *29*, 74–81. [CrossRef] [PubMed]

12. Wu, K.C.; Wang, C.M. Steganography using reversible texture synthesis. *IEEE Trans. Image Process.* **2014**, *24*, 130–139. [PubMed]

13. Liu, M.M.; Zhang, M.Q.; Liu, J.; Zhang, Y.N.; Ke, Y. Coverless Information Hiding Based on Generative Adversarial Networks. *arXiv* **2017**, arXiv:1712.06951.

14. Gatys, L.A.; Ecker, A.S.; Bethge, M. A Neural Algorithm of Artistic Style. *arXiv* **2015**, arXiv:1508.06576. [CrossRef]

15. Li, Y.; Fang, C.; Yang, J.; Wang, Z.; Lu, X.; Yang, M.H. Universal Style Transfer via Feature Transforms. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2017.

16. Li, Y.; Fang, C.; Yang, J.; Wang, Z.; Lu, X.; Yang, M.H. Diversified Texture Synthesis with Feed-Forward Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.

17. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2014; pp. 2672–2680.

18. Gatys, L.A.; Ecker, A.S.; Bethge, M. Texture Synthesis Using Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2015. [CrossRef]

19. Ulyanov, D.; Lebedev, V.; Vedaldi, A.; Lempitsky, V.S. Texture Networks: Feed-forward Synthesis of Textures and Stylized Images. In Proceedings of the 33rd International Conference on Machine Learning, New York, NY, USA, 19–24 June 2016.

20. Karras, T.; Laine, S.; Aila, T. A Style-Based Generator Architecture for Generative Adversarial Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–21 June 2019.

21. Aly, H.A.; Dubois, E. Image up-sampling using total-variation regularization with a new observation model. *IEEE Trans. Image Process.* **2005**, *14*, 1647–1659. [CrossRef] [PubMed]

22. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. In Proceedings of the European conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016. [CrossRef]

23. Mahendran, A.; Vedaldi, A. Understanding Deep Image Representations by Inverting Them. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015. [CrossRef]

24. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016. [CrossRef]

25. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common Objects in Context. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014. [CrossRef]

26. Chen, T.Q.; Schmidt, M. Fast Patch-based Style Transfer of Arbitrary Style. *arXiv* **2016**, arXiv:1612.04337.

27. Yi, C.; Zhou, Z.; Yang, C.N.; Leqi, J.; Chengsheng, Y.; Xingming, S. Coverless information hiding based on Faster R-CNN. In Proceedings of the International Conference on Security with Intelligent Computing and Big-Data Services, Guilin, China, 14–16 December 2018.

28. Zhou, Z.; Sun, H.; Harit, R.; Chen, X.; Sun, X. Coverless Image Steganography Without Embedding. In Proceedings of the International Conference on Cloud Computing and Security, Nanjing, China, 13–15 August 2015. [CrossRef]

29. Wang, Y.; Yang, X.; Liu, J. Cross-domain image steganography based on GANs. In Proceedings of the International Conference on Security with Intelligent Computing and Big-Data Services, Guilin, China, 14–16 December 2018.

30. Ker, A.D. Improved Detection of LSB Steganography in Grayscale Images. In Proceedings of the 6th International Conference on Information Hiding, Toronto, ON, Canada, 23–25 May 2004.