Hindawi
Security and Communication Networks
Volume 2021, Article ID 6663486, 8 pages
https://doi.org/10.1155/2021/6663486

Research Article

Anti-Forensics of Image Contrast Enhancement Based on Generative Adversarial Network

Hao Zou,[1,2] **Pengpeng Yang**,[1,2] **Rongrong Ni**,[1,2] **and Yao Zhao**[1,2]

[1]*Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China*
[2]*Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing 100044, China*

Correspondence should be addressed to Rongrong Ni; rrni@bjtu.edu.cn and Yao Zhao; yzhao@bjtu.edu.cn

Received 31 December 2020; Revised 5 February 2021; Accepted 16 February 2021; Published 25 March 2021

Academic Editor: Nanrun Zhou

In the multimedia forensics community, anti-forensics of contrast enhancement (CE) in digital images is an important topic to understand the vulnerability of the corresponding CE forensic method. Some traditional CE anti-forensic methods have demonstrated their effective forging ability to erase forensic fingerprints of the contrast-enhanced image in histogram and even gray level cooccurrence matrix (GLCM), while they ignore the problem that their ways of pixel value changes can expose them in the pixel domain. In this paper, we focus on the study of CE anti-forensics based on Generative Adversarial Network (GAN) to handle the problem mentioned above. Firstly, we exploit GAN to process the contrast-enhanced image and make it indistinguishable from the unaltered one in the pixel domain. Secondly, we introduce a specially designed histogram-based loss to enhance the attack effectiveness in the histogram domain and the GLCM domain. Thirdly, we use a pixel-wise loss to keep the visual enhancement effect of the processed image. The experimental results show that our method achieves high anti-forensic attack performance against CE detectors in the pixel domain, the histogram domain, and the GLCM domain, respectively, and maintains the highest image quality compared with traditional CE anti-forensic methods.

1. Introduction

With the development of computer techniques, digital images are widely used in our world. Accordingly, potential security problems in digital images have been emerging in recent years. Many manipulated digital images are threats to our forensic systems. To deal with this issue, researchers studied a large number of image forensic methods. However, numerous forensic methods have their limitations in robustness. To understand their limitations and weakness, anti-forensic studies in digital images were developed as well.

In recent years, many anti-forensic studies were proposed [1–14]. The common practice of digital image anti-forensics [1–3, 5, 7–11, 14] is to introduce a minimum distortion in the digital image to eliminate or change the corresponding fingerprints that forensic methods rely on, which can lead to the successful anti-forensic attack against forensic methods. In this case, the anti-forensic image is visually close to the attacked image. Recently, a different kind of anti-forensic practice in operation image anti-forensics is to model the image's anti-forensic problem as an image translation or restoration problem based on GAN, such as median filtering anti-forensics [4], JPEG compression anti-forensics [6], and multi-operation image anti-forensics [12, 13]. This kind of practice translates the operated image to its unaltered one, which can result in the disappearance of the operation fingerprint. Besides, Chen et al. [14] proposed a GAN-based camera model anti-forensic study [14] to deceive camera model detectors as well as preserving the visual effect of the attacked image. Up to now, it is still developable to study operation image anti-forensics based on GAN under the condition of preserving the visual effect of the operation. We decided to focus on the study of a single operation anti-forensic task, such as CE anti-forensics, for the first attempt at this issue.

CE anti-forensics, as one of the tasks of anti-forensics, was developed to counter CE forensic methods in recent years. Early CE anti-forensic strategies [8–10] were studied against histogram-based CE forensic methods. Cao et al. [8]

proposed the method of local random dithering (LRD), which aims at removing the peak-gap artifacts that appeared in the gray level histogram of the contrast-enhanced image. Barni et al. [10] proposed a universal anti-forensic method against histogram-based forensic detectors. After that, to further remove the artifacts in both histogram and gray level cooccurrence (GLCM), Ravi et al. [11] proposed an effective anti-forensic CE technique by solving an optimization problem.

Although these methods can deceive histogram-based detectors and even GLCM-based CE detectors, they ignore the fact that their ways of pixel value changes would expose them in the pixel domain. So far, it is still a challenging CE anti-forensic task that counters CE forensic detectors in the pixel domain, the histogram domain, and the GLCM domain simultaneously. To both solve this problem and dig the potential capability of GAN, we propose a novel GAN-based CE anti-forensic method in this paper. We exploit GAN to process the contrast-enhanced image and make the processed image indistinguishable from the unaltered one in the pixel domain. Meanwhile, a specially designed histogram-based loss is introduced to enhance the attack effectiveness in the histogram domain and the GLCM domain. Besides, we use a pixel-wise loss to keep the visual enhancement effect of the processed image. We follow the mean-shifted Gaussian-functions-based method in [15] to calculate the differentiable histogram suitable for deep learning training procedure. The experimental results show that our method successfully deceives three deep-learning-based CE forensic detectors [6, 16] in the pixel domain, the histogram domain, and the GLCM domain, respectively, and keeps the highest image quality compared to traditional CE anti-forensic methods.

Our contributions are summarized as follows:

(1) We exploit GAN to accomplish CE anti-forensics in the condition of preserving the visual effect of CE to a large extent. To the best of our knowledge, this is the first attempt to use GAN for CE anti-forensics in the condition of preserving the visual effect of CE.

(2) We introduce a specially designed histogram-based loss to enhance the attack effectiveness in the histogram domain and the GLCM domain.

(3) Our method shows high anti-forensic attack performance in terms of the pixel domain, the histogram domain, and the GLCM domain.

(4) Compared with traditional CE anti-forensic methods, our method achieves the highest image quality.

The rest of the paper is organized as follows. In Section 2, we describe the background. In Section 3, we describe the proposed method. In Section 4, we present the details of our experiment. Finally, we summarize the work and look into the future development in Section 5.

2. Background

2.1. Generative Adversarial Network and Anti-Forensics. GAN is a deep learning framework proposed by Goodfellow et al. [17] to generate visually realistic images. Classical GAN includes two networks, a generator G and a discriminator D. G tries to generate an image x' and make its distribution $p_g(x')$ close to the distribution $p_r(x)$ of the real image x as much as possible. D tries to distinguish the real image x from the generated image x'. The two networks are trained alternately in a competitive way by optimizing the following minimax problem:

$$\min_G \max_D \mathbb{E}_{x\sim p_r(x)}[\log(D(x))] + \mathbb{E}_{x'\sim p_g(x')}[\log(1-D(x'))]. \tag{1}$$

The two networks in GAN are opposite, which is similar to the relationship between the attacker and the forensic investigator. Thus, it is a proper way to study the anti-forensic method based on GAN [18].

2.2. CE Artifacts and CE Detection in Histogram Domain. In the early age of CE forensic studies, Stamm and Liu [19] studied a blind CE forensic method in digital images based on the fact that the histogram of the unaltered image is smooth, while the corresponding histogram of contrast-enhanced one is with artifacts of peak-gap, as is illustrated in Figure 1.

Specifically, CE operation used to be a nonlinear pixel mapping, for example, Gamma Correction, which can be separated into locally contractive mapping and locally expansive mapping. The rounding after locally contractive mapping and locally expansive mapping results in the appearance of peaks and gaps, respectively.

The peak-gap artifacts in the histogram are a kind of high-frequency signal, which can lead to the addition of the high-frequency component in the Fourier domain, as is shown in Figure 1. Meanwhile, there does exist a similar phenomenon in high-end and low-end saturated unaltered images. In the corresponding histograms, there exists the impulsive peak at the pixel value of 255 and 0. The Fourier transform of an impulse is a constant function, which also results in the addition of the high-frequency component. To avoid this effect, Stamm and Liu [19] proposed a pinch-off function to process the histogram as follows:

$$y(x) = p(x)h(x), \tag{2}$$

where x represents the image, $p(x)$ is the pinch-off function, and $h(x)$ represents the histogram of x.

Then, the high-frequency measure F is calculated by the following formula:

$$F = \frac{1}{N}\sum_{\omega} |\beta(\omega)Y(\omega)|, \tag{3}$$

where $Y(\omega)$ is the Fourier transform of $y(x)$ and $\beta(\omega)$ is a weighting function that takes values between 0 and 1 to deemphasize low-frequency regions of $Y(\omega)$. $\beta(\omega)$ is formulated as

$$\beta(\omega) = \begin{cases} 1, & \omega \geq c \\ 0, & \omega < c \end{cases} \tag{4}$$

where c is a cutoff frequency.

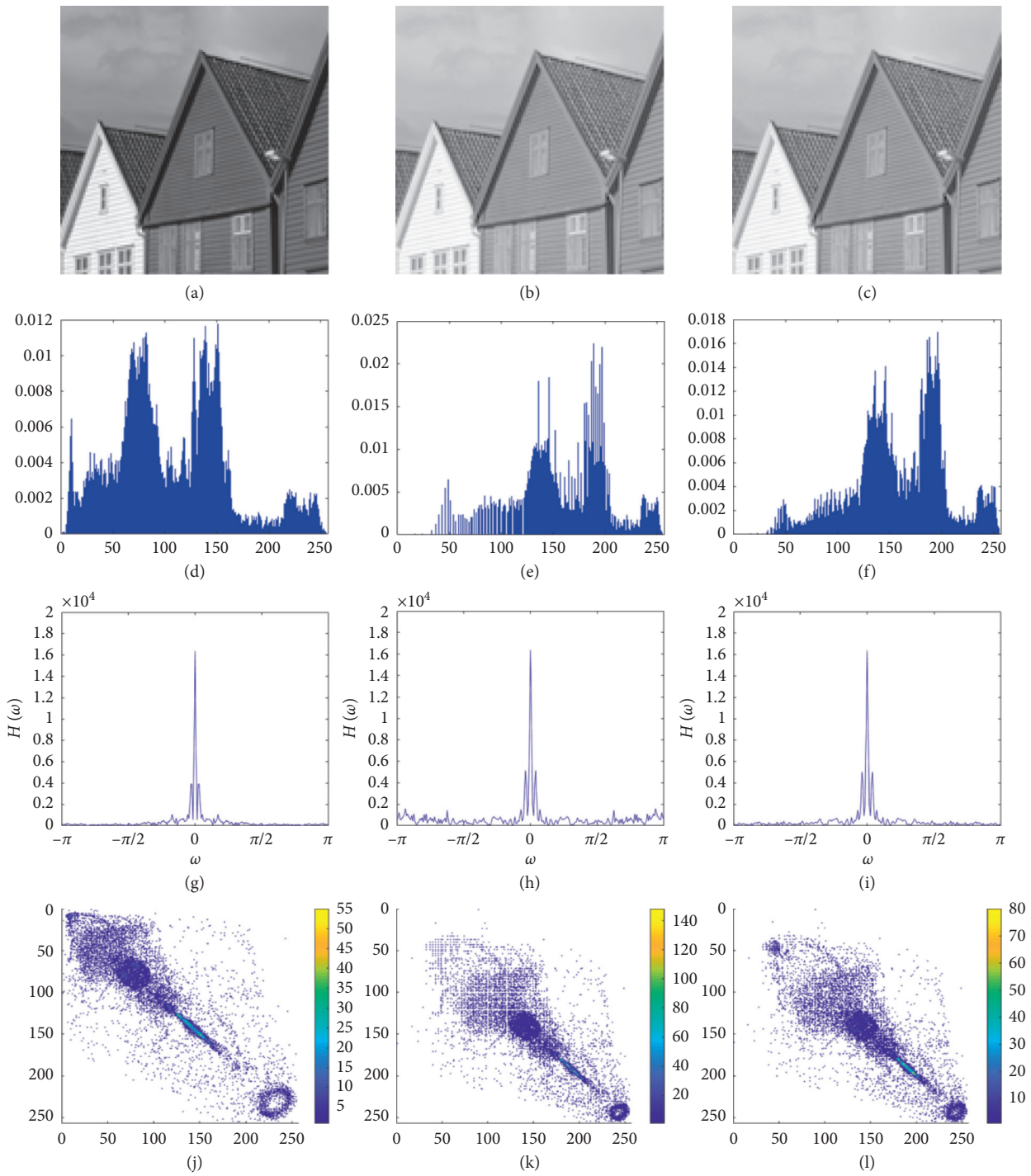

FIGURE 1: An example of visualization. The first column shows an unaltered image (a), its histogram (d), the magnitude (g) of the Fourier transform of the histogram, and its GLCM (j) from top to bottom. The second column shows a contrast-enhanced image (b) using gamma correction with $\gamma = 0.5$, its histogram (e), the magnitude (h) of the Fourier transform of the histogram, and its GLCM (k) from top to bottom. The third column shows an anti-forensic image (c) of our method, its histogram (f), the magnitude (i) of the Fourier transform of the histogram, and its GLCM (l) from top to bottom.

Finally, Stamm and Liu [19] performed a threshold test to identify CE operation. From the point of anti-forensics, lowering the high-frequency component of the histogram of the contrast-enhanced image is a possible solution.

2.3. CE Artifacts in GLCM Domain. GLCM is used to describe texture features from the pixel correlation of gray. De Rosa et al. [20] firstly discovered empty rows and columns appearing in GLCM of the contrast-enhanced image, while

there does not exist this kind of artifacts in the GLCM of the unaltered one, as is illustrated in Figure 1. The empty rows and columns in the GLCM correspond to the gaps in the histogram because of the absence of the corresponding pixel values. From the point of anti-forensics, empty rows and columns can be removed under the circumstance that the artifacts of peak-gap in the histogram are eliminated.

2.4. Histogram Calculation for Convolutional Neural Network. Shifted step functions centered on the corresponding histogram bins can be used to calculate the histogram without any information loss, while they are useless for CNN to learn the histogram feature due to the fact that their derivative is zero everywhere except for the edges. Towards this issue, Sedighi and Fridrich [15] proposed a method to approximate the histogram with mean-shifted Gaussian functions, as is illustrated in Figure 2. With this method, the histogram bins can be calculated by the following formula:

$$H(I,k) = \sum_{i=1}^{W}\sum_{j=1}^{H} e^{-\left(\left(I_{ij}-k\right)^2/\sigma^2\right)}, \tag{5}$$

where I_{ij} represents the pixel value of the image I in the location i, j, W, and H represent the corresponding width and height of I, k denotes the mean value, and σ denotes the standard deviation.

Mean-shifted Gaussian functions are continuously differentiable and their derivatives are not always 0. With this property, it can obtain a valid back-flow of gradients for the update of CNN parameters.

3. Proposed Method

3.1. GAN-Based CE Anti-Forensic Framework. In this section, we propose a GAN-based framework for CE anti-forensics in the condition of preserving the visual effect of CE. Given a contrast-enhanced image x, our goal is to reconstruct it with the capability of attacking CE forensic methods as well as maintaining the visual effect of x.

Figure 3 shows the overall architecture of our framework. Our framework is composed of three portions. In the blue portion, to enhance the attack effectiveness in the pixel domain, the generator G is used to transform the contrast-enhanced image to the generated one capable of falsifying the discriminator D by optimizing the adversarial loss $\mathscr{L}_G^{\text{adv}}$. In the green portion, we approximate the histogram of the generated image and then process the histogram using Fourier transform (FT) and finally calculate the corresponding high-frequency measure F as the histogram-based loss $\mathscr{L}_G^F$. Considering the fact that the high-frequency measure of the contrast-enhanced image histogram is higher than that of the unaltered one and that the CE artifacts in the histogram and the GLCM are interrelated, as is mentioned in Section 2, we can enhance the attack effectiveness of our method in the histogram domain and the GLCM domain by minimizing $\mathscr{L}_G^F$. In the orange portion, we use a pixel-wise loss $\mathscr{L}_G^{\text{pixel}}$ to lower the visual difference between the contrast-enhanced image and the generated image.

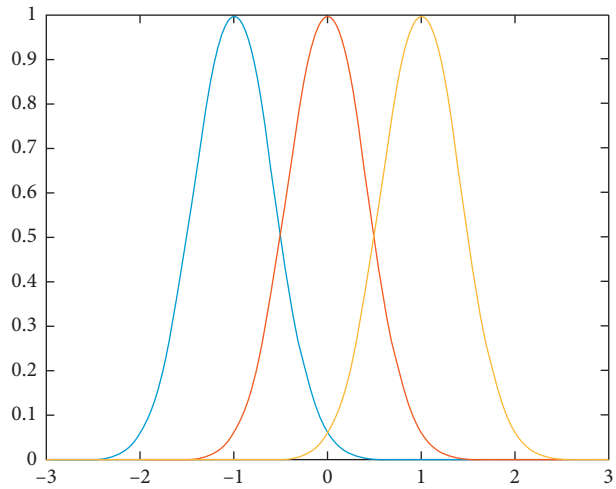

Figure 2: An example of mean-shifted Gaussian functions.

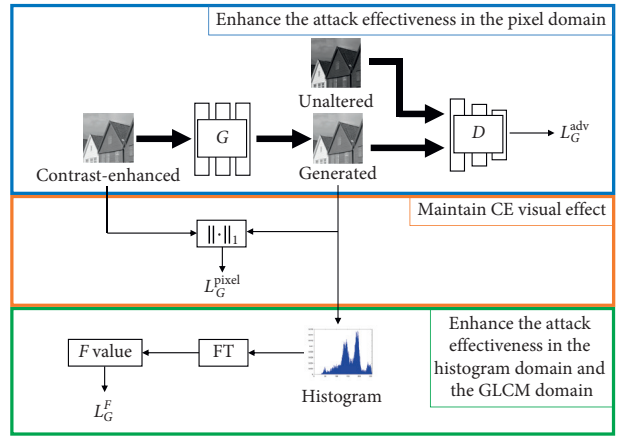

Figure 3: The overall framework of our method.

The details of our network and loss function are described in the next two subsections.

3.2. CE Anti-Forensic Network. We design our generator for the anti-forensic processing of the contrast-enhanced image able to falsify the discriminator. Our generator network is illustrated in Figure 4. We introduce a skip connection from the input position to the position before the last clamp layer, which is a type of residual learning strategy to accelerate the network training [21]. The last clamp layer is used to restrict the maximum and minimum pixel values for keeping the consistency of the pixel range between the generated image and the contrast-enhanced image.

The backbone network is composed of several groups. The first group includes a 3×3 convolution layer with output of 16 feature maps, a batch normalization (BN) layer, and a leaky rectified linear unit (LeakyReLU). Then, three same residual blocks (ResBlocks) are connected. Each ResBlock has two repetitive parts, which include a 3×3 convolution layer with output of 16 feature maps, a BN layer, and a LeakyReLU layer. Besides, there exists a skip connection between the input of Resblock and the position

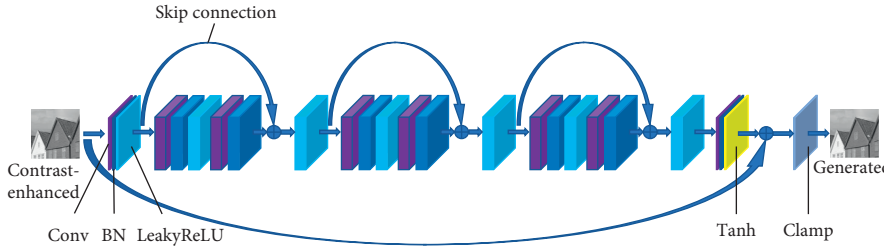

Figure 4: The generator network.

before the second LeakyReLU of ResBlock. The last group includes a 3×3 convolution layer with a single channel output, a BN layer, and an activation layer of the hyperbolic tangent (Tanh).

To ensure that the discriminator is enough capable of detecting CE in the generated image, the discriminator directly adopts the structure of P-CNN in [16], which was proposed for CE forensics.

3.3. Loss Function. Our ultimate goal is to generate anti-forensic images that can deceive CE forensic methods and are visually close to the corresponding contrast-enhanced images. To achieve this goal, we set the loss functions for the generator and the discriminator, respectively, and optimize their parameters for both of them by minimizing the loss functions during the training procedure. The details of our loss functions are as follows.

3.3.1. Generator Loss. The generator loss function is

$$\mathscr{L}_G = \lambda_1 \mathscr{L}_G^{\text{adv}} + \lambda_2 \mathscr{L}_G^F + \lambda_3 \mathscr{L}_G^{\text{pixel}}, \tag{6}$$

where $\mathscr{L}_G^{\text{adv}}$ represents the adversarial loss for fooling the discriminator, $\mathscr{L}_G^F$ represents high-frequency measure F loss based on the image histogram in the Fourier domain, and $\mathscr{L}_G^{\text{pixel}}$ represents pixel-wise image quality loss of the generated image compared with the contrast-enhanced image. The coefficients λ_1, λ_2, and λ_3 represent the corresponding weights of each loss term.

$\mathscr{L}_G^{\text{adv}}$ is to ensure that our generator G can falsify the discriminator D. We calculate $\mathscr{L}_G^{\text{adv}}$ by the following formula:

$$\mathscr{L}_G^{\text{adv}} = \log(1 - D(G(x))), \tag{7}$$

where x represents the contrast-enhanced image, $G(x)$ represents the generated image, and $D(\cdot)$ denotes the output of the discriminator.

$\mathscr{L}_G^F$ is to lower the high-frequency component in the histogram for the better attack effectiveness in the histogram domain and the GLCM domain. Before calculating this loss, we need to calculate the histogram of the generated image. To ensure that the back-flow of gradients is not blocked due to the calculation of the image histogram, we follow the method of using mean-shifted Gaussian functions to approximate it [15]. Considering that the pixel values of the generated image are not integers and they cannot be rounded for keeping the back-flow of gradients in the training procedure, we introduce a bias term b to the mean value of mean-shifted Gaussian functions to center on fractional pixels. The different values of b correspond to different fractional pixels. The corresponding histogram bin of the generated image $G(x)$ is calculated by the following formulation:

$$H(G(x), k, b) = \sum_{i=1}^{W} \sum_{j=1}^{H} e^{-\left(\left(G(x)_{ij} - (k+b) \right)^2 / \sigma^2 \right)}, \tag{8}$$

where $k = 0, 1, \ldots, 255$, $-1 < b < 1$, $k + b$ represents the center of the histogram bin, and $\sigma = 0.3$. We can get the final histogram $V(G(x), b)$ by concatenating the 256 histogram bins. After that, we calculate $\mathscr{L}_G^F$ by the following formula:

$$\mathscr{L}_G^F = \sum_{i=1}^{N} M(V(G(x), b_i)), \tag{9}$$

where N denotes the number of bias terms, b_i represents the i-th bias term, $V(G(x), b_i)$ represents the calculated histogram, and $M(V(G(x), b_i))$ represents the high-frequency measure of the histogram $V(G(x), b_i)$ in the Fourier domain. We follow the method [19] mentioned in Section 2 to calculate the high-frequency measure.

$\mathscr{L}_G^{\text{pixel}}$ is to ensure that the generated image is visually close to the contrast-enhanced image. We calculate the absolute mean difference between the generated image $G(x)$ and the contrast-enhanced image x. The formula of $\mathscr{L}_G^{\text{pixel}}$ is as follows:

$$\mathscr{L}_G^{\text{pixel}} = \frac{1}{WH} \sum_{i=1}^{W} \sum_{j=1}^{H} \left| x_{ij} - G(x)_{ij} \right|, \tag{10}$$

where i and j denote the pixel indexes and W and H denote the width and height of the image, respectively.

3.3.2. Discriminator Loss. The discriminator is trained to identify the generated image from the unaltered one by optimizing the traditional discriminator loss function [17], which is as follows:

$$\mathscr{L}_D = -\log(D(y)) - \log(1 - D(G(x))), \tag{11}$$

where y represents the unaltered image and $G(x)$ represents the corresponding contrast-enhanced image.

4. Experiment

4.1. Experiment Setup. In our experiment, we chose the public BOSSbase dataset [22] as the original dataset, which contains 10,000 grayscale images of size 512×512 in png format. Considering the limited hardware configuration, we decided to launch our experiment with images of size 128×128. Each image in the original dataset was cropped with no overlapping to get eight 128×128 patches. In this way, we obtained the unaltered dataset containing 80,000 images. Accordingly, we created 80,000 contrast-enhanced images using gamma correction.

We chose four γ values of 0.5, 0.8, 1.2, and 1.5, while the number of images for each γ value is 20,000. Therefore, we got 80,000 pairs of unaltered and contrast-enhanced images. We divided these image pairs into the training set and the testing set at a ratio of 4 : 1 for training and testing, respectively. The proposed network was implemented by PyTorch framework [23] and trained on one GPU, NVIDIA RTX 2080 Ti.

During each iteration for training our network, the generator was trained with 40 contrast-enhanced images and the discriminator was trained with 40 pairs of images, including contrast-enhanced images and the corresponding unaltered ones. The generator and the discriminator were alternately trained in iterations. Our training procedure was divided into two parts. Firstly, we trained our network for 35 epochs. The learning rates for the generator and the discriminator were fixed to 5×10^{-5} and 1×10^{-4}, respectively. We set the coefficients of $\lambda_1 = 1$, $\lambda_2 = 0$, and $\lambda_3 = 100$ in generator loss. Then, we continued to train our network for 5 epochs. The learning rates for the generator and the discriminator were both fixed to 1×10^{-6}. We set the coefficients of $\lambda_1 = 1$, $\lambda_2 = 0.35$, and $\lambda_3 = 100$ in generator loss. The cut-off frequency was set to 0.875π. Besides, we set four bias terms of $b_1 = -0.25$, $b_2 = 0$, $b_3 = 0.25$, and $b_4 = 0.5$ in $\mathscr{L}_G^F$.

We used Adam as the optimizer with $\beta_1 = 0.5$, $\beta_2 = 0.999$, and $\varepsilon = 0.5$ for the generator and used SGD as the optimizer with momentum = 0.9 and weight_decay = 5×10^{-4} for the discriminator. After the training procedure, we input 16,000 contrast-enhanced images in the testing set into the well-trained generator model to obtain 16,000 anti-forensic images.

4.2. Evaluation. Before evaluating CE anti-forensic algorithms, we trained three deep-learning-based CE forensic detectors proposed by [16, 24]. Two detectors of P-CNN and H-CNN, the input data of which are in the form of images and histograms, respectively, were proposed in [16]. Another detector was proposed in [24]. For convenience, we refer to it as GLCM-CNN, as it classifies contrast-enhanced images from unaltered images by analyzing the GLCM of images. The performance of the three detectors under the testing set is shown in Table 1.

We evaluated CE anti-forensic methods in two aspects, attack effectiveness and image quality. Firstly, we carried out anti-forensic attacks using four types of anti-forensic images, which were obtained by our method and three other traditional methods [8, 10, 11], against three trained CE forensic detectors. The detection accuracies of each detector for these four types of anti-forensic images are shown in Table 2. The lower detection accuracy indicates the better attack effectiveness of the corresponding anti-forensic methods. The average detection accuracy of P-CNN for our anti-forensic images is 0.1304, which is the lowest compared with the other three anti-forensic methods. This is because our method considers the anti-forensic attack in the pixel domain, while other methods do not take it into account. The detection accuracies of H-CNN and GLCM-CNN to our method are still at low levels because we consider enhancing the anti-forensic attack performance in terms of the histogram domain and the GLCM domain by introducing a histogram-based loss $\mathscr{L}_G^F$. Even if they are not the lowest, these results indicate that our method is still effective enough to deceive H-CNN and GLCM-CNN. In general, our method successfully deceives P-CNN [16], H-CNN [16], and GLCM-CNN [24].

Secondly, to verify the image quality of these four CE anti-forensic methods in the condition of keeping the contrast-enhanced visual effect, we calculated PSNR and SSIM between anti-forensic images and the corresponding contrast-enhanced ones. The higher values of PSNR and SSIM mean the better image quality. The average PSNR and average SSIM of these four anti-forensic images are shown in Table 3. Our method achieves the highest image quality, 49.0258 dB of PSNR, and 0.9926 of SSIM. To summarize, our method can still keep good anti-forensic attack effectiveness with the highest image quality.

For visualization, we present an example that contains an unaltered image, a contrast-enhanced image, an anti-forensic image of our method, and the corresponding histograms and GLCM, shown in Figure 1. We can hardly find the visible distortion in the anti-forensic image compared to the contrast-enhanced one. The artifacts of peak-gap and empty rows and columns of the contrast-enhanced image in the histogram and the GLCM, respectively, are successfully erased. Besides, the high-frequency component in the Fourier transform of our anti-forensic image histogram is at a low level, which is close to the unaltered one.

Finally, we evaluated the impact of the histogram-based loss $\mathscr{L}_G^F$. In Figure 5, we can find that the loss term of $\mathscr{L}_G^F$ is beneficial to enhance the attack ability against P-CNN, H-CNN, and GLCM-CNN. In particular, the enhancement of the attack effectiveness against H-CNN and GLCM-CNN is obvious, which is in accord with our idea of enhancing the attack effectiveness in the histogram domain and the GLCM domain by using $\mathscr{L}_G^F$.

5. Conclusions

In this paper, we propose a novel CE anti-forensic method based on GAN. Our method shows the high anti-forensic attack performance against deep-learning-based CE detection techniques in terms of the pixel domain, the histogram domain, and the GLCM domain. The image quality of our anti-forensic images is also superior to other

Table 1: The average accuracies of three CE forensic detectors under the testing set.

Detector	$\gamma = 0.5$	$\gamma = 0.8$	$\gamma = 1.2$	$\gamma = 1.5$	Average
P-CNN [16]	0.9006	0.8211	0.7750	0.8463	0.8357
H-CNN [16]	0.9938	0.9914	0.9771	0.9876	0.9875
GLCM-CNN [24]	0.9942	0.9928	0.9822	0.9882	0.9893

Table 2: The average detection accuracies of three CE forensic detectors to four types of anti-forensic images. The lower accuracy indicates the better attack effectiveness of the anti-forensic method. The best attack results are marked in bold.

Attack	Detector	$\gamma = 0.5$	$\gamma = 0.8$	$\gamma = 1.2$	$\gamma = 1.5$	Average
Cao et al. [8]	P-CNN [16]	0.8055	0.6953	0.6120	0.6793	0.6980
	H-CNN [16]	**0.0000**	**0.0000**	**0.0165**	**0.0358**	**0.0131**
	GLCM-CNN [24]	**0.0013**	**0.0018**	**0.0208**	**0.0410**	**0.0162**
Barni et al. [10]	P-CNN [16]	0.9942	0.9667	0.9725	0.9887	0.9805
	H-CNN [16]	0.5353	0.2025	0.1690	0.2853	0.2980
	GLCM-CNN [24]	0.7620	0.5920	0.5097	0.6657	0.6324
Ravi et al. [11]	P-CNN [16]	0.6378	0.5498	0.6845	0.7075	0.6449
	H-CNN [16]	0.1350	0.0513	0.0658	0.1133	0.0914
	GLCM-CNN [24]	0.0598	0.0335	0.0543	0.0753	0.0557
Our method	P-CNN [16]	**0.1704**	**0.1159**	**0.0900**	**0.1452**	**0.1304**
	H-CNN [16]	0.1722	0.0108	0.0262	0.0709	0.0700
	GLCM-CNN [24]	0.0338	0.0179	0.0255	0.0436	0.0302

Table 3: The average PSNR and average SSIM of four types of anti-forensic images. The higher values of PSNR and SSIM mean the better image quality. The best results are marked in bold.

Image quality	Cao et al. [8]	Barni et al. [10]	Ravi et al. [11]	Our method
PSNR (dB)	42.0392	45.1696	37.3918	**49.0258**
SSIM	0.9629	0.9896	0.9718	**0.9926**

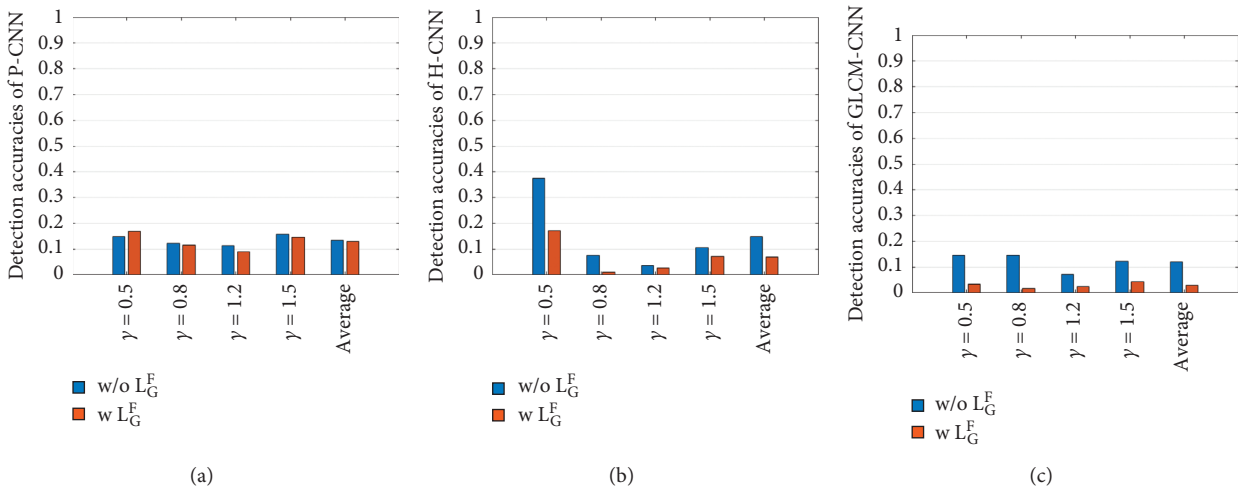

Figure 5: The detection accuracies of (a) P-CNN [16], (b) H-CNN [16], and (c) GLCM-CNN [24] to our anti-forensic method under the circumstances of using $\mathscr{L}_G^F$ and not using $\mathscr{L}_G^F$. The lower accuracy indicates the better attack effectiveness of the anti-forensic method.

traditional methods. In the future, we attempt to study a general visual effect preserved operation image anti-forensic method based on GAN for more tasks of operation image anti-forensic.

Data Availability

All data included in this study are available upon request to the corresponding author.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported in part by the National Key Research and Development of China (2018YFC0807306), the National Natural Science Foundation of China (U1936212 and 61672090), and Beijing Fund-Municipal Education Commission Joint Project (KZ202010015023).

References

[1] M. Kirchner and R. Bohme, "Hiding traces of resampling in digital images," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 4, pp. 582–592, 2008.

[2] M. C. Stamm, S. K. Tjoa, and W. S. Lin, "Anti-forensics of JPEG compression," in *Proceedings of the 2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1694–1697, IEEE, Dallas, TX, USA, March 2010.

[3] M. C. Stamm and K. J. R. Liu, "Anti-forensics of digital image compression," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 1050–1065, 2011.

[4] D. Kim, H.-U. Jang, S.-M. Mun, S. Choi, and H.-K. Lee, "Median filtered image restoration and anti-forensics using adversarial networks," *IEEE Signal Processing Letters*, vol. 25, no. 2, pp. 278–282, 2018.

[5] W. Fan, K. Wang, and F. Cayre, "Median filtered image quality enhancement and anti-forensics via variational deconvolution," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 5, pp. 1076–1091, 2015.

[6] Y. Luo, H. Zi, and Q. Zhang, "Anti-forensics of JPEG compression using generative adversarial networks," in *Proceedings of the 2018 26th European Signal Processing Conference (EUSIPCO)*, pp. 952–956, Rome, Italy, September 2018.

[7] L. Lu, G. Yang, and M. Xia, "Anti-forensics for unsharp masking sharpening in digital images," *International Journal of Digital Crime & Forensics*, vol. 5, no. 3, pp. 53–65, 2013.

[8] G. Cao, Y. Zhao, and R. Ni, "Anti-forensics of contrast enhancement in digital images," in *Proceedings Of the 12th ACM Workshop on Multimedia And Security (MM&Sec 10)*, pp. 25–34, New York, NY, USA, February 2010.

[9] C. W. Kwok, O. C. Au, and S. H. Chui, "Alternative anti-forensics method for contrast enhancement," in *Proceedings of the 10th International Conference on Digital-Forensics And Watermarking (IWDW'11)*, pp. 398–410, Springer-Verlag, Berlin, Germany, June 2011.

[10] M. Barni, M. Fontani, and B. Tondi, "A universal technique to hide traces of histogram-based image manipulations," in *Proceedings of the on Multimedia And Security (MM&Sec 12)*, pp. 97–104, New York, NY, USA, February 2012.

[11] H. Ravi, A. V. Subramanyam, and S. Emmanuel, "ACE-an effective anti-forensic contrast enhancement technique," *IEEE Signal Processing Letters*, vol. 23, no. 2, pp. 212–216, 2016.

[12] J. Wu, Z. Wang, and H. Zeng, "Multiple-operation image anti-forensics with WGAN-GP framework," in *Proceedings of the 2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pp. 1303–1307, Lanzhou, China, November 2019.

[13] J. Wu and W. Sun, "Towards multi-operation image anti-forensics with generative adversarial networks," *Computers & Security*, vol. 100, Article ID 102083, 2021.

[14] C. Chen, X. Zhao, and M. C. Stamm, "MISLGAN: an anti-forensic camera model falsification framework using a generative adversarial network," in *Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 535–539, Athens, Greece, October 2018.

[15] V. Sedighi and J. Fridrich, *"Histogram Layer, Moving Convolutional Neural Networks towards Feature-Based Steganalysis,"* Electronic Imaging, Media Watermarking, Security, and Forensics, Binghamton, NY, USA, 2017.

[16] P. Yang, R. Ni, and Y. Zhao, "Robust contrast enhancement forensics using pixel and histogram domain CNNs," 2019, https://arxiv.org/abs/1803.04749.

[17] I. Goodfellow, J. Pouget-Abadie, and M. Mirza, "Generative adversarial nets," in *Proceedings of the 27th International Conference on Neural Information Processing Systems-(NIPS'14)*, pp. 2672–2680, MIT Press, Cambridge, MA, USA, December 2014.

[18] M. Barni, M. C. Stamm, and B. Tondi, "Adversarial multimedia forensics: overview and challenges ahead," in *Proceedings of the 2018 26th European Signal Processing Conference (EUSIPCO)*, pp. 962–966, Rome, Italy, September 2018.

[19] M. Stamm and K. J. R. Liu, "Blind forensics of contrast enhancement in digital images," in *Proceedings of the 2008 15th IEEE International Conference on Image Processing*, pp. 3112–3115, San Diego, CA, USA, October 2008.

[20] A. De Rosa, M. Fontani, M. Massai, A. Piva, and M. Barni, "Second-order statistics analysis to cope with contrast enhancement counter-forensics," *IEEE Signal Processing Letters*, vol. 22, no. 8, pp. 1132–1136, 2015.

[21] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[22] P. Bas, T. Filler, and T. Pevný, "Break our steganographic system: the ins and outs of organizing BOSS," in *Proceedings of the 13th International Conference on Information Hiding (IH'11)*, pp. 59–70, Springer-Verlag, Berlin, Germany, May 2011.

[23] A. Paszke, S. Gross, and F. Massa, "Pytorch: an imperative style, high-performance deep learning library," *Advances In Neural Information Processing Systems*, pp. 8024–8035, 2019.

[24] J.-Y. Sun, S.-W. Kim, S.-W. Lee, and S.-J. Ko, "A novel contrast enhancement forensics based on convolutional neural networks," *Signal Processing: Image Communication*, vol. 63, pp. 149–160, 2018.