Anti-Forensic Against Double JPEG Compression Detection Using Adversarial Generative Network

Kutub Uddin, Yoonmo Yang, Byung Tae Oh
Korea Aerospace University
kutub@kau.kr, yym064@kau.kr, byungoh@kau.ac.kr

이중압축 검출기술에 대한 GAN 기반 안티 포렌식 기술

우딘, 양윤모, 오병태
한국항공대학교

ABSTRACT

Double JPEG compression detection is one of the most important ways of exposing the integrity of the JPEG image in image forensics. Several methods have been proposed for discriminating against the double JPEG image. In this paper, we propose a new method for restoring the JPEG compressed image and making the detector confused by introducing a Generative Adversarial Network (GAN). First, a generator network is designed for restoring the JPEG compressed image and analyzed the quality. Then, the restored image is tested with the double compression detector for evaluating the robustness of the proposed GAN model. The detection accuracy reduces from 98% to 58%.

1. INTRODUCTION

With the revolution of social media like Facebook, Twitter, Instagram, KakaoTalk, etc., people are uploading their daily views as images at an alarming rate for sharing important moments with friends and relatives. At the same time, there are a number of software introduced for manipulating these images for some kinds of malicious purposes. These make the great demand for efficient forensic methods for the authentication of the images. Many researchers have been motivated themselves for the identification of the different kinds of forgery in the digital images for ensuring integrity. Double JPEG compression detection is one of the most important fields of digital image forensics. Double JPEG compression refers to the recompression of the JPEG image once again with the same or different quality factors. There are several methods proposed for detecting the double JPEG compression. Huang et al. [1] proposed a method for double JPEG compression detection with the same quantization matrix using dense CNN feature. With the introduction of the methods for detecting double JPEG compression, some anti-forensic methods against the double JPEG image also proposed. In [2], Luo et al. proposed an anti-forensic method against double compression detection with the same quantization matrix by observing the decrease of the number of DCT coefficients during recompression. A method of detecting double JPEG compression and some related anti-forensic using CNN was motivated by Zhang et al. [3]. Kim et al. [4] tried to reconstruct the median filtered image by removing the traces using an adversarial neural network. In [5], Wang et al. designed a deep duel-domain (D3) model for restoring the JPEG compressed image.

In this work, we introduce a new anti-forensic method for reconstructing the double JPEG compressed image to make the detector fool. First, we design a generator network with an appropriate loss function for restoring the artifact leaving due to the double JPEG compression into the original-like JPEG image and analyzed the quality. Then, the restored image is used for making the detector [3] confused in order to show the robustness of the designed GAN model. The method proposed in [3] mainly based on the detection of the double JPEG image and its related anti-forensic with CNN.

2. PROPOSED METHOD

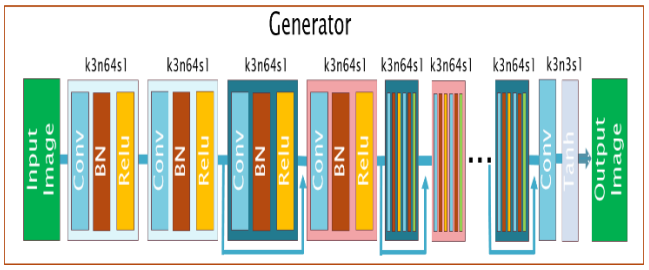
Several methods have been introduced for anti-forensic against double JPEG compression detection mostly based on hand-crafted features. We propose a GAN model for restoring the double JPEG image inspired from [4] consisting of two networks: a generator network (G) and a discriminator network (D). The concept of minimax problem is used for training the generator and the discriminator networks, given as:

$$\min_{\theta_G}\max_{\theta_D} \; \mathbb{E}_{x\sim P_{JPEG(x)}}\left[log D_{\theta_D}(x)\right] + \mathbb{E}_{x'\sim P_{DJPEG(x')}}\left[\log\left(1 - D_{\theta_D}(G_{\theta_G}(x'))\right)\right] \quad (1)$$

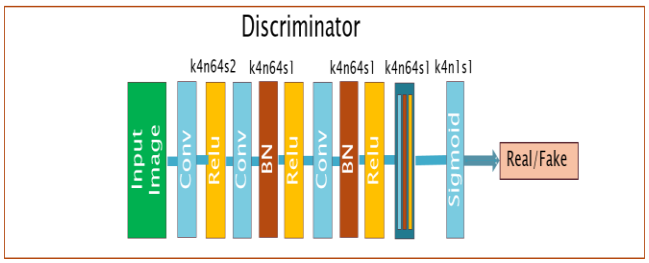
where θ_G and θ_D are the learnable parameters for generator and discriminator respectively. The generator and discriminator networks compete with each other for their jobs best. In this case, the generator network tries to reconstruct the double JPEG image x' like the raw image x and the discriminator network task is to classify between restored and raw images.

A. Generative Adversarial Network

The first part of the GAN model is the generator network. It is composed of the residual block shown in Figure 1(a) including convolution and batch normalization followed by leaky rectified linear unit (LeakyReLU) activation function. There are some skip connections after two blocks for the summation between neighboring blocks. Finally, a tangent activation function is used in the last layer. The generator network is intended to minimize the errors between the JPEG compressed and raw JPEG images in order to generate the reconstructed image similar to the raw JPEG image.



(a)



(b)

Figure 1: Generative Adversarial Network (GAN) (a) Generator and (b) Discriminator

The second part of the GAN model is the discriminator network. It consists of five blocks shown in Figure 1(b) including convolution, batch normalization followed by leaky rectified linear unit (LeakyReLU) activation function. A sigmoid activation function with LeakyReLU is used as the final layer. The main task of the discriminator is to maximize the classification rate between the double JPEG image and the generated image.

There are three different loss functions used in the proposed GAN model shown in Figure 2.

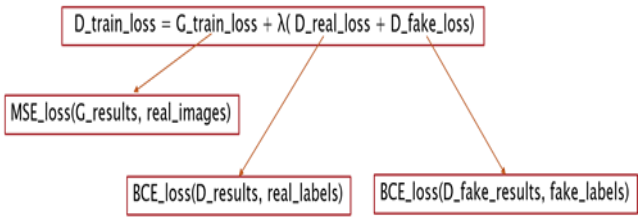


Figure 2: Loss function

First, the Mean Squared Error (MSE) loss function is used for calculating the loss between generated and real images called generator loss. Then, the Binary Cross-Entropy (BCE) loss function is used for calculating discriminator real loss and fake loss. The discriminator real loss is determined between the predicted and real labels. In the same way, the discriminator fake loss is determined between the predicted and fake labels.

B. Double JPEG Image Restoration

Double JPEG compression is the process of recompressing the JPEG image into twice with the same or different quality factors. For the reconstruction of the double JPEG image, the artifacts existing due to the recompression must be eliminated. We use a GAN model for removing the traces leaving due to the double JPEG compression to retrieve the original like JPEG image. Figure 3 shows the overall structure of generating the reconstructed JPEG image by minimizing the generator loss using the raw image. The input to the generator can be either single JPEG compressed, or double JPEG compressed images and it generates raw like image. We train the proposed GAN model with original JPEG and double JPEG compressed images for generating the restored image.

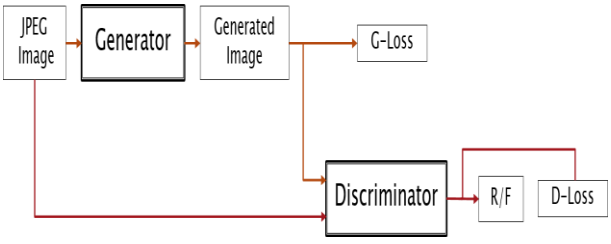


Figure 3: JPEG Image Restoration

3. EXPERIMENTAL REUSLTS

In this part, we present the experimental details of our proposed system. We use the DIV2K dataset [6] and perform resizing of size 512×384×3 on 1600 training and 200 testing images. We conduct two types of experiments. First, we show the qualitative analysis in terms of peak signal-to-noise ratio (PSNR) and structural similarity (SSIM). Then,

the performance is evaluated for showing the robustness of the proposed method.

A. Quality Measurement of Restored JPEG Image

We use five images for illustrating the qualitative judgment. First, we calculate the PSNR and the SSIM between raw JPEG images (Raw) and double JPEG images (DJPEG). Then, the same operations are performed in the case of raw JPEG images and reconstructed JPEG images (RJPEG). The double JPEG compression is done with $QF_1 = 70$ and $QF_2 = 75$. The results are shown in Table 1:

Table 1: Qualitative analysis

Index	Raw vs DJPEG		Raw vs RJPEG	
	PSNR	SSIM	PSNR	SSIM
1	27.78	0.96	29.28	0.91
2	30.03	0.98	30.34	0.92
3	29.24	0.98	29.56	0.92
4	29.72	0.96	29.81	0.94
5	27.26	0.98	27.81	0.93

From Table 1, it is clear that the images generated by the proposed GAN model are more similar to the raw JPEG images than the double JPEG compressed images in terms of PSNRs. But the SSIMs are less in generated images. It is because of the structural properties of the images that are not preserved completely by our GAN model.

B. Performance Evaluation

In order to show the effectiveness of the proposed model, we evaluate the performance using the existing method [3] D^{DD-CNN} and our proposed anti-forensic D^{DD-CNN} ($AF-D^{DD-CNN}$). The detection result is shown in Table 2:

Table 2: Detection accuracies

(QF_1, QF_2)	D^{DD-CNN}[2]	$AF-D^{DD-CNN}$
(70, 75)	98.45	58.00

From Table 2, we can say that the proposed method is much effective to make the discriminator fool as the accuracy is reduced up to 58%.

4. CONCLUSION

In this paper, a new anti-forensic method against double JPEG compression detection is introduced using a GAN model. First, a generator network is designed to reconstruct the artifacts due to the double compression and the quality is analyzed. Then, the reconstructed image is used for testing the existing method [3] for detecting double JPEG compression. The proposed method can make fool the existing method up to 58% times.

ACKNOLEDGEMENT

This research was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of ICT (NRF-2019R1F1A1063229), and by the GRRC program of Gyeonggi Province [2017-B02, Study on image processing and UI platform for mobile media devices].

REFERENCES

[1] X. Huang, S. Wang, and G. Liu, "Detecting double JPEG compression with same quantization matrix based on dense CNN feature." *25th IEEE International Conference on Image Processing (ICIP)*, 3813-3817, 2018.

[2] H. Li, W. Luo, and J. Huang, "Anti-forensics of double JPEG compression with the same quantization matrix." *Multimedia Tools and Applications* 74(17), 6729-6744, 2015.

[3] Li. B. Zhang, H. Luo, H. and Tan, "Detecting double JPEG compression and its related anti-forensic operations with CNN." *Multimedia Tools and Applications*, 78(7), 8577-8601, 2019.

[4] D. Kim, H. Jang, S. M. Mun, S. Choi, and H. K. Lee, "Median filtered image restoration and anti-forensics using adversarial networks." *IEEE Signal Processing Letters* 25(2), 278-282, 2018.

[5] Z. Wang, D. Liu, S. Chang, Q. Ling, Y. Yang, and T. S. Huang, "D3: Deep dual-domain based fast restoration of jpeg-compressed images.", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2764-2772, 2016.

[6] E. Agustsson, and R. Timofte. "NTIRE 2017 challenge on single image super-resolution: Dataset and study." *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 126-135. 2017.