Computers, Materials & Continua
DOI:10.32604/cmc.2022.024862
Article

Tech Science Press

Reinforced CNN Forensic Discriminator to Detect Document Forgery by DCGAN

Seo-young Lim and Jeongho Cho*

Department of Electrical Engineering, Soonchunhyang University, Asan, 31538, Korea
*Corresponding Author: Jeongho Cho. Email: jcho@sch.ac.kr

Received: 02 November 2021; Accepted: 10 December 2021

Abstract: Recently, the technology of digital image forgery based on a generative adversarial network (GAN) has considerably improved to the extent that it is difficult to distinguish it from the original image with the naked eye by compositing and editing a person's face or a specific part with the original image. Thus, much attention has been paid to digital image forgery as a social issue. Further, document forgery through GANs can completely change the meaning and context in a document, and it is difficult to identify whether the document is forged or not, which is dangerous. Nonetheless, few studies have been conducted on document forgery and new forgery-related attacks have emerged daily. Therefore, in this study, we propose a novel convolutional neural network (CNN) forensic discriminator that can detect forged text or numeric images by GANs using CNNs, which have been widely used in image classification for many years. To strengthen the detection performance of the proposed CNN forensic discriminator, CNN was trained after image preprocessing, including salt and pepper as well as Gaussian noises. Moreover, we performed CNN optimization to make existing CNN more suitable for forged text or numeric image detection, which have mainly focused on the discrimination of forged faces to date. The test evaluation results using Hangul texts and numbers showed that the accuracy of forgery discrimination of the proposed method was significantly improved by 20% in Hangul texts and 5% in numbers compared with that of existing state-of-the-art methods, which proved the proposed model performance superiority and verified that it could be a useful tool in reducing crime potential.

Keywords: Digital forensics; CNN; GAN; discriminator; image processing

1 Introduction

Owing to the COVID-19 pandemic, the use of non-face-to-face services through which official documents can be submitted and issued to public institutions has increased recently. According to a press release in 2020 by the Ministry of Security and Public Administration of the Republic of Korea, the number of annual usages of the "Mobile Document 24" application opened by the government has increased from 350,000 to 2,350,000 uses, more than 6.5 times since the first service launch in 2018.

According to the increase in the usage rate of mobile official documents, mobile identifications have also been introduced, which are used to verify adult ages at convenience stores or for vehicle rentals through applications [1,2].

However, the number of forged and falsified official documents continuously increases according to the data from Statistics Korea [2]. Forged academic and grade transcripts of Bachelor's, Master's, and Doctorate degrees are used illegally for young people to get jobs. In illegal entry, residents' registration cards and passports are forged, as are endless crimes of official document forgery and falsification in society, eroding public trust and increasing social unrest [3]. High-resolution digital cameras are now affordable, and software programs and source code are easily accessible, which further lowers the barriers to image manipulation. Thus, many techniques to detect fake images have been proposed to cope with easily generated fake images by adding, removing, or copying people and objects through relatively easy manipulation and editing. However, new forgery attacks are emerging daily [4].

Recently, with the development of machine learning and artificial intelligence, images and videos can be manipulated digitally much easily than before. Moreover, an entirely new image can be created using its original image (Fig. 1). Thus, the limitation of existing image manipulation can now be overcome. In particular, a generative adversarial network (GAN) is an unsupervised learning-based generative model implemented with two networks, which is regarded as one of the most promising technological developments in the image creation and manipulation field [5–8]. Techniques to create data based on a GAN, such as image-to-image conversion [9–11] and Deepfake [12], which changes a person's face into a different face to make a fake video, are frequently used to make fake images or videos. In addition, detailed manipulations can be possible by adjusting data in the image synthesis process, and a style may be applied arbitrarily [13,14].

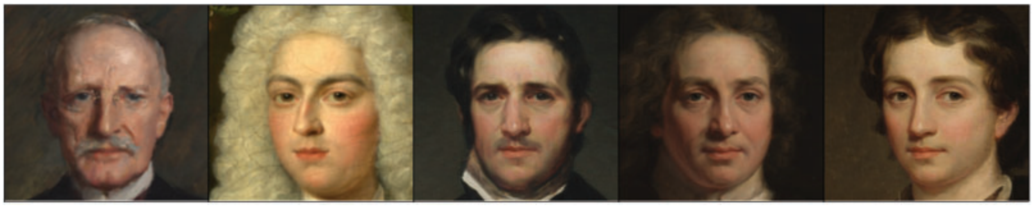
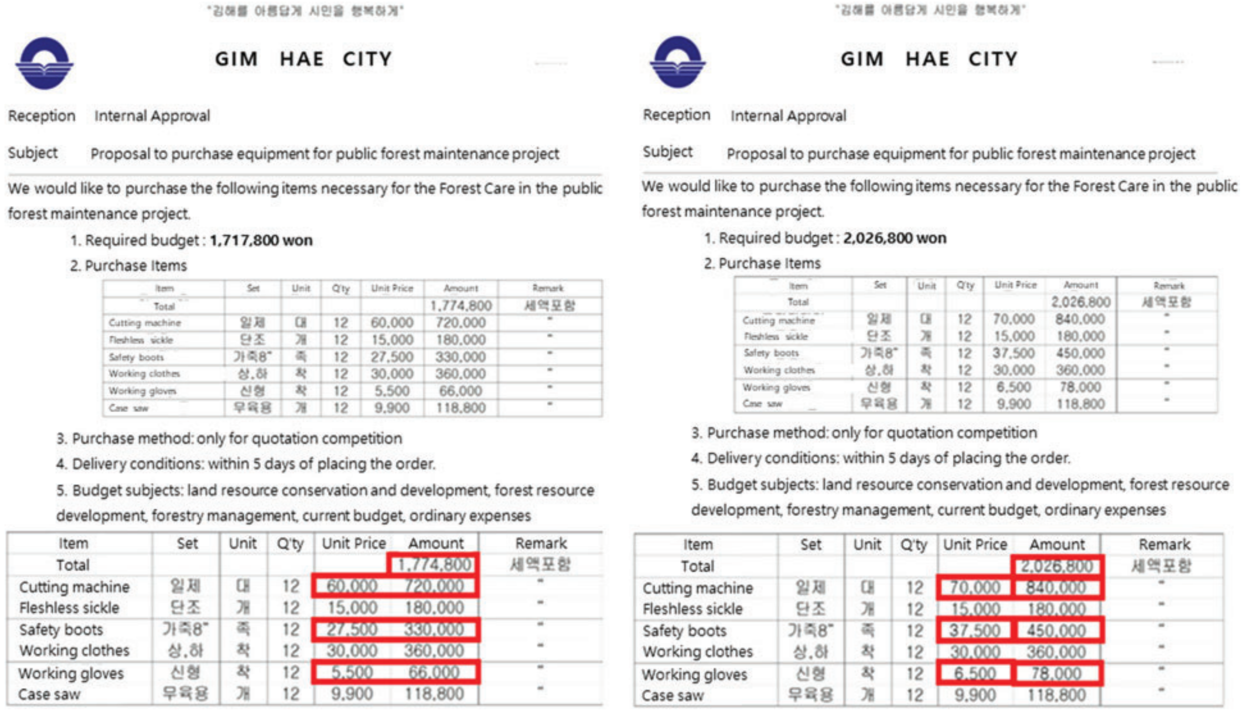


Figure 1: Image generated by stylegan2 ADA: The two images on the left are real images from the training set, and the last three images are generated through stylegan2 ADA

Fig. 2 shows an example of a fake document image for internal approval in the national administration office created using public administration documents of artificial intelligence hub. The left is the original image, and the right is the manipulated image. The total cost was artificially manipulated from KRW 1,774,000 to KRW 2,026,000. In this document, a fake numeric image was created through learning with deep convolutional GANs (DCGANs) using the same style of lettering, which was Gothic. This example demonstrates the possible risk of national budget waste through the forgery of unit prices or amounts in documents of national administration offices. Document forgery can completely change the context and meaning of the document in the real world and is difficult to be detected. Thus, a process to detect them efficiently is required.



"김해를 아름답게 시민을 행복하게"

GIM HAE CITY

Reception Internal Approval

Subject Proposal to purchase equipment for public forest maintenance project

We would like to purchase the following items necessary for the Forest Care in the public forest maintenance project.

1. Required budget : **1,717,800 won**
2. Purchase Items

Item	Set	Unit	Q'ty	Unit Price	Amount	Remark
Total					1,774,800	세액포함
Cutting machine	일제	대	12	60,000	720,000	-
Fleshless sickle	단조	개	12	15,000	180,000	-
Safety boots	가죽8"	족	12	27,500	330,000	-
Working clothes	상.하	착	12	30,000	360,000	-
Working gloves	신형	착	12	5,500	66,000	-
Case saw	무육용	개	12	9,900	118,800	-

3. Purchase method: only for quotation competition
4. Delivery conditions: within 5 days of placing the order.
5. Budget subjects: land resource conservation and development, forest resource development, forestry management, current budget, ordinary expenses

Item	Set	Unit	Q'ty	Unit Price	Amount	Remark
Total					1,774,800	세액포함
Cutting machine	일제	대	12	60,000	720,000	-
Fleshless sickle	단조	개	12	15,000	180,000	-
Safety boots	가죽8"	족	12	27,500	330,000	-
Working clothes	상.하	착	12	30,000	360,000	-
Working gloves	신형	착	12	5,500	66,000	-
Case saw	무육용	개	12	9,900	118,800	-

(a)

"김해를 아름답게 시민을 행복하게"

GIM HAE CITY

Reception Internal Approval

Subject Proposal to purchase equipment for public forest maintenance project

We would like to purchase the following items necessary for the Forest Care in the public forest maintenance project.

1. Required budget : **2,026,800 won**
2. Purchase Items

Item	Set	Unit	Q'ty	Unit Price	Amount	Remark
Total					2,026,800	세액포함
Cutting machine	일제	대	12	70,000	840,000	-
Fleshless sickle	단조	개	12	15,000	180,000	-
Safety boots	가죽8"	족	12	37,500	450,000	-
Working clothes	상.하	착	12	30,000	360,000	-
Working gloves	신형	착	12	6,500	78,000	-
Case saw	무육용	개	12	9,900	118,800	-

3. Purchase method: only for quotation competition
4. Delivery conditions: within 5 days of placing the order.
5. Budget subjects: land resource conservation and development, forest resource development, forestry management, current budget, ordinary expenses

Item	Set	Unit	Q'ty	Unit Price	Amount	Remark
Total					2,026,800	세액포함
Cutting machine	일제	대	12	70,000	840,000	-
Fleshless sickle	단조	개	12	15,000	180,000	-
Safety boots	가죽8"	족	12	37,500	450,000	-
Working clothes	상.하	착	12	30,000	360,000	-
Working gloves	신형	착	12	6,500	78,000	-
Case saw	무육용	개	12	9,900	118,800	-

(b)

Figure 2: Document manipulated with numeric images created through a generative adversarial network (GAN); (a) original image (b) the image manipulated

Recently, deep learning technology has been actively used in object recognition and classification fields, and a convolutional neural network (CNN) is particularly useful in finding a feature pattern for image recognition [15]. As the trace of photo-response nonuniformity pattern remains in a shot image by a camera, various studies have been conducted to detect whether images generated with a GAN are forged using a CNN based on the fact that a specific fingerprint remains in an image created by a GAN [16]. An 80% accuracy was realized in detecting forged images using the VGG16 architecture configured with five-layer blocks, including convolution and max-pooling layers [3], and the detection accuracy was improved to 91% using 13 convolution, 5 pooling, and 3 dense layers by changing the architecture [17]. However, since their proposed networks attempted learning using only forged images through a specific architecture, the detection performance of images forged by other architectures rapidly degraded [16]. [18] employed a high-frequency pass filter (HFPF) to distinguish fake and real faces and determined whether the face is genuine through CNN using a residual. [19] detected fake images through CNN based on high-frequency components obtained through the HFPF after preprocessing via the Gaussian low-frequency pass filter, which suppresses the influence of external noise to identify face images generated by GAN that has undergone postprocessing such as image denoising and sharpening. In addition, the statistical similarity was improved at the pixel level between images through preprocessing using Gaussian blur and Gaussian noise to reinforce the generalization ability of the detection method of fake images created over other architectures, thereby learning more meaningful features inherently by the classifier [20]. Consequently, the accuracy improved by up to 6% compared with that before using data preprocessing. However, the above previous CNN-based forgery discrimination techniques were designed for the digital synthesis image discrimination of human faces and did not verify the effectiveness of the discrimination of forged documents using a GAN, which is the aim of this study. Most recently, based on the fact that GAN-generated fake images have strong spectral correlations, [21] proposed a novel fake image detection strategy based on Discrete Wavelet Transform (DWT); GAN-generated images were transformed into three color components, R, G, and B, and then DWT and standard correlation coefficients were employed. As a result of the test, the detection accuracy was more than 90%, but the generalization performance was not considered as only styleGAN2-synthesized faces were included in the evaluation, and thus it was very limited.

When a text in a forged document using a GAN is compared with that of the original document, it is not easy to distinguish it compared to when it is a face image. Thus, it is necessary to improve existing detection methods. Therefore, in this study, we propose a CNN-based detection strategy of document forgery created through learning with a GAN including preprocessing to improve the generalization ability of forgery discriminator and optimized CNN, by which existing CNN forensic discriminators apply to Hangul and numeric-based document forgery detection. The proposed strategy underwent preprocessing using salt and pepper noise pattern as well as Gaussian noise to improve the detection performance of forged documents. The performance evaluation results verified that the proposed method was effective in the forged text detection by training with a CNN using data containing various features. As a result of comparison with existing CNN-based fake face detection techniques, the proposed method was superior in forgery detection. This result verified the usability in effective discrimination of forged documents generated by a GAN, which are receiving significant attention in the field of digital image forensics recently. Further, the proposed method will significantly contribute to not only document forgery detection but also Deepfake detection research as a basic technology in the future.

Fake news that provides false information through forged images, as well as information such as highly realistic fake profiles in a social network through the forgery and falsification of texts or numbers, can be harmfully used in other applications, which must be solved. Thus, the contributions of this study in this context are as follows:

i. An optimization method of the CNN forensic model is proposed to extend existing discrimination techniques, which were limited to face images generated by a GAN, to the discrimination of Hangul and number forgery detection.
ii. A data preprocessing technique is proposed to improve the diversity of GAN-generated data and the generalization ability of CNN forensic models.

The rest of this paper is organized as follows: In Section 2, the data creation and proposed forgery document detection technique are described; in Section 3, experimental results are presented. Finally, in Section 4, conclusions are presented.

2 Methodology

Owing to the need for developing a detection technology to match with the growth rate of forgery technology, we propose an algorithm to detect texts, particularly Hangul or numeric forged images generated by a GAN. A CNN, which has been used to detect forged images, extracts input image features, and classifies images using the extracted pattern. However, because the generated forged image features may differ depending on the GAN's architecture, sufficient images generated by various GAN models should be included in CNN learning to ensure the generalizability of CNN classifiers, but it is difficult to implement this. To solve this problem, we improved the generalizability of detectors by removing fingerprints of GANs through preprocessing of training images. Further, existing CNN detectors were mainly used as tools for fake face detection. Thus, we optimized CNN classifiers to realize forged texts or numbers using a GAN. The proposed reinforced discrimination strategy of GAN-generated forged document is shown in Fig. 3, which comprises three sections: 1) fake image generator, 2) preprocessor, and 3) CNN forensic discriminator.

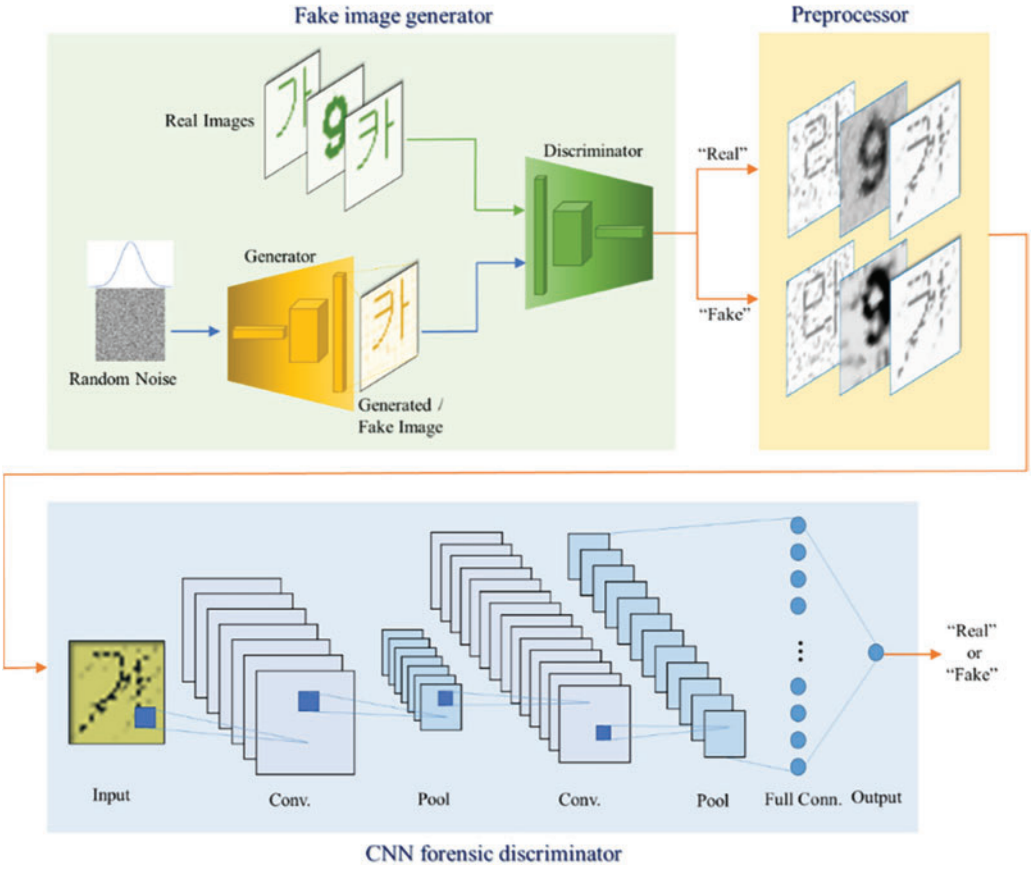


Figure 3: Overall framework of the proposed reinforced CNN forensic discriminator

2.1 Fake Image Generator

Because fake images made by a fake image generator are produced by their original images, original images were first created with a size of 28 × 28 pixels, which were similar fonts supported by the Windows. Here, the numbers were created with a size of 0.7 points and "2" bold using five fonts available in Python: "Italic," "Hershey complex," "Hershey duplex," "Hershey script complex," and "Hershey." For Hangul, five fonts, "Gulim," "Gungseo," "Malgun Gothic," "Batang," and "Headline," were used, and 14 consonant letters, "가," "나," "다," "라," "마," "바," "사," "아," "자," "차," "카," "타," "파," and "하," were created with a size of 20 points. Fig. 4 shows examples of original Hangul and numeric images in various fonts.

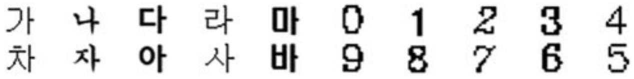


Figure 4: Examples of original Hangul and numeric images in various fonts

The DCGAN was used to create forged images for CNN learning and test evaluations based on the produced original images. The DCGAN is a GAN comprising two networks with a CNN structure. The first network is a generator, which creates fake images with the input of random noise, and the second network is a discriminator, which determines whether the generated image is real or fake. The generated fake image is input to the discriminator along with the original image to determine whether it is genuine or not. Based on this result, the generator is trained adversarially to generate more realistic fake images. Here, the training continues until it can fool the discriminator completely. The loss function is presented in Eq. (1), which aims for complete mapping of latent space Z into data space X.

$$\min_{G}\max_{D} V(D,G) = \mathbb{E}_{x\sim P_{data(x)}}[\log D(x)] + \mathbb{E}_{z\sim P_{z(z)}}[\log(1-D(G(z))], \tag{1}$$

where $G(z)$ is the output of the generator, given a random noise z, $D(x)$ is the estimate from the discriminator, $P_{data(x)}$ is the distribution of the original image, and $P_{z(z)}$ refers to the Gaussian distribution of the latent space. The generator is trained to make the objective function $V(D,G)$ the minimum, i.e., $D(G(z)) = 1$, whereas the discriminator is trained to make the objective function the maximum, i.e., $D(x) = 1,\ D(G(z)) = 0$.

2.2 *Preprocessor*

A fake image generated by the DCGAN provides a realistic image similar to the original image, but it can be easily distinguished from the original image because it has a specific fingerprint, thereby obtaining high accuracy. However, for images created by GAN models other than the DCGAN, which have the same architecture, the performance of discriminating whether a fake image is genuine is significantly degraded. To solve this drawback, salt and pepper as well as Gaussian noises were added to be used in the learning of discriminator to prevent the detection model from being more focused on training artificial fingerprints inserted by the GAN and image background than on a text itself.

Salt and pepper noise randomly alters a certain amount of pixel into two incorrect brightness values, either 0 or 255, which occurs as a form of black or white point in an image. Generally, in the document background, visible watermark, anticounterfeiting pattern, and background noise can be included, which become the background behind a series of texts in a document. Thus, by adding a random salt and pepper noise pattern to the original image, the background is made in the form of noise instead of white, thereby allowing the CNN forensic discriminator to learn by focusing more on the text itself rather than the background. The amounts of salt and pepper were added equally at 50% each, and the added amount was randomly set between 0 and 0.5. Moreover, the statistics of image pixels were modified by adding the Gaussian noise of zero mean unit variance to make it difficult for DCGAN to focus only on detection through artificial fingerprints in training the CNN forensic discriminator. Preprocessed data, which are randomly sampled, are shown in Fig. 5; we determined whether they are genuine as they enter the forensic detection unit.

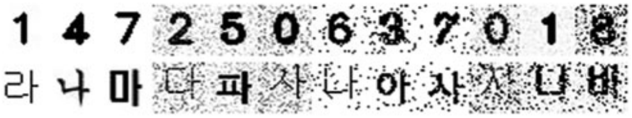


Figure 5: Examples of Hangul and numeric images with noise added through the preprocessor

2.3 CNN Forensic Discriminator of Document Forgery

Deep learning-based technologies have been widely used in object and face recognition as well as image classification recently. Among them, a CNN is particularly outstanding in finding a pattern to recognize images. A filter is applied to the input image, and a feature map, which is the same as that of the input image, is generated through convolution and combination, thereby classifying images using the extracted feature patterns. Owing to the advantage of automatically extracting image features, CNNs have shown the effectiveness to identify forged images [22]. Thus, a CNN was considered the discriminator architecture to determine whether a document was forged, and the CNN detection model for discriminating forged text or number by GANs was optimized with the structure presented in Tab. 1.

Table 1: Proposed CNN architecture for discriminating document forgery

Layer	Layer comment	Output
Input	Input image	28 × 28 × 3
Conv1	3 × 3 × 64	26 × 26 × 64
Pool1	2 × 2 max pooling	13 × 13 × 64
Conv2	3 × 3 × 128	11 × 11 × 128
Pool2	2 × 2 max pooling	5 × 5 × 128
Conv3	3 × 3 × 256	3 × 3 × 256
Flatten	Flatten Conv3 output	2,304
FC1	Full conn. with 1 hidden layer	512
FC2	Full conn. with 2 hidden layer	256
Output	Output node	2

For the network optimization, real and fake images were randomly mixed at a 1:1 ratio and extracted to have 100,000 and 20,000 images for learning and testing, respectively. To shorten the learning time and reduce the possibility of falling into the local minimum, all data were normalized, and the network had three convolution layers. All convolution layers employed 2 × 2 max pooling and "batch normalization." The size of all convolution kernels was 3 × 3; the activation function used in each convolution layer was "ReLU", but the last layer employed "SoftMax." The loss function and optimization algorithm were sparse categorical cross-entropy loss function and adaptive momentum estimation (Adam). The loss function is shown in Eq. (2):

$$L = -\frac{1}{N}\sum_{i=1}^{N} \beta_i \log(\alpha_i) + (1 - \beta_i)\log(1 - \beta_i), \tag{2}$$

where β_i is the actual value, and α_i is the estimated one. Adam was adopted to the optimization because it has few memory requirements as it can perform efficient operations through simple implementation. As the learning progresses, the learning rate was reduced, and the speed was calculated to adjust the learning intensity. In the forensic CNN model's learning process, "Dropout" was used to prevent overfitting of training data.

3 Experimental Results

The forgery performance of the GAN was first verified before the performance evaluation of the proposed GAN-generated forgery document discriminator was performed. The hardware used for learning GAN and others included an Intel Core i7-7700 CPU, NVIDIA GTX 1060 GPU (6 GB), and 16 GB of memory. The software environment comprised Tensorflow 2.2.0, Opencv 4.5.1, CUDA v.10.2, and Cudnn v.7.6.4 on Ubuntu 16.04.5 (4.15.0–38 kernel). The GAN-generated forged images were highly similar to the original images, which could not be clearly distinguishable by naked eyes. Thus, the following Fréchet inception distance (FID), which is a method of measuring the similarity between curves by considering the location and order of points along the curve, was used to calculate a distance of feature vectors between the original and fake images for the quality evaluation of how the forged image was close to the original image.

$$\mathrm{FID} = ||\mu_r - \mu_g||^2 + \mathrm{Tr}\left(\mathrm{Cov}_r + \mathrm{Cov}_g - 2(\mathrm{Cov}_r\mathrm{Cov}_g)^{\frac{1}{2}}\right), \tag{3}$$

where (μ_r, Cov_r) and (μ_g, Cov_g) are the mean and covariance of each original and generated image feature, respectively, and Tr is the trace operation. The lower the FID is, the better the forged image quality is, and the generated image is closer to the original image. The performance of the forged image generated using the DCGAN was compared with that of the Wasserstein GAN (WGAN), which is designed to compensate for the weakness of being difficult to train while maintaining a balance between the discriminator and generator of GAN. WGAN is a structure that uses the critic to deliver the gradient well and trains the critic and generator to the optimal point so that it can play a better role in classification than the discriminator of GAN. The comparative results are summarized in Tab. 2.

Table 2: FID to evaluate GAN-generated fake image performance

	DCGAN	WGAN
Numbers	0.143	0.083
Hangul	0.154	0.106

Both techniques verified that as the FID was closer to zero, they could fake images better, and the FID of WGAN was lower in both number and Hangul than that of DCGAN, indicating that the images created by WGAN were closer to the original images. That is, it can be expected that the image generated by DCGAN will be relatively easier to detect forgery than the image generated by WGAN. Thus, DCGAN-generated images were used for CNN training to have more objective evaluations, whereas, for performance evaluation, WGAN-generated images were used to make closer images to the original images for harder discrimination. A comparison of performance evaluation on forged image detection according to various preprocessing procedures was performed. Tab. 3 summarizes the models used in the comparison evaluation. In this study, based on an existing CNN forensic model optimized to face images, $\mathrm{M}_{\mathrm{face}}$, and a CNN model optimized to Hangul and numeric images, M, a CNN model trained through preprocessing by adding Gaussian noise to model M, a CNN model trained via preprocessing by adding the Gaussian blur noise to model M, and a CNN model trained additionally through preprocessing using the Gaussian noise as well as salt and pepper noise were defined as M_{g}, M_{gb}, and $\mathrm{M}_{\mathrm{g+sp}}$ (proposed model), respectively.

Table 3: Definition of models used in test evaluation

Forensic model	Description
M_{face} [17]	CNN model optimized for face image, no preprocessing
M	CNN model optimized for Hangul and numeric image, no preprocessing
M_{g} [20]	CNN model optimized for Hangul and numeric image, preprocessed with Gaussian noise
M_{gb} [20]	CNN model optimized for Hangul and numeric image, preprocessed with Gaussian blur noise
M_{g+sp} (proposed model)	CNN model optimized for Hangul and numeric image, preprocessed with Gaussian as well as salt and pepper noise

For indices to evaluate the performance of the proposed CNN forensic model, accuracy, which calculates a probability of being classified correctly considering both the case of correctly predicting the real image as "True" and the case of correctly predicting the fake image as "False," true negative rate (TNR), which is a rate of forged images predicted as fakes, and true positive rate (TPR), which is a rate of predicting the original image as the original. They can be calculated as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN}, \tag{4}$$

$$TNR = \frac{TN}{TN + FP}, \tag{5}$$

$$TPR = \frac{TP}{TP + FN}, \tag{6}$$

where TP means true positive, which is correctly determined case of original images, TN means true negative, which is correctly determined case of forged images. FP means false positive, which is a case where original images are wrongly determined as forged images, and FN means false negative, which is a case where fake images are wrongly determined as original images.

Tab. 4 presents the comparison results of detection performance among forensic models with and without preprocessing. Data used in the performance evaluation were numbers and Hangul, and forensic models M and M_{g+sp} were optimized through data comprising fake and original images generated by DCGAN. To verify the generalizability, forged data created using DCGAN, which was the same form used in training the CNN forensic model, and WGAN, which was a new form without participation in training, were employed. The performance results by forensic models with and without preprocessing using test data created by DCGAN showed that M and M_{g+sp} had no significant difference in forgery discrimination at a nearly similar level. However, the accuracy results using test data created by WGAN revealed that for numbers, the accuracies of M and M_{g+sp} were 76.44% and 82.07%, respectively, which showed the proposed model was 5.63% superior, and for Hangul, their respective accuracies were 84.94% and 88.22%, which showed the proposed model was approximately 3% superior. Thus, the above results verified that the proposed forensic model improved the generalizability and detection performance of fake images by incorporating preprocessing, focusing more on characters' feature analysis rather than on fingerprints of GAN or background of Hangul and numeric characters.

Table 4: Performance comparisons of forensic models optimized for Hangul and numbers with and without preprocessing

Data type	Test generative model	Forensic model	Accuracy [%]	TPR [%]	TNR [%]
Number	DCGAN	M	96.25	95.90	93.56
		M_{g+sp}	96.50	97.93	95.07
	WGAN	M	78.43	95.90	57.93
		M_{g+sp}	80.02	97.93	62.10
Hangul	DCGAN	M	94.11	97.09	91.14
		M_{g+sp}	93.55	95.55	91.48
	WGAN	M	84.94	97.09	80.79
		M_{g+sp}	88.22	95.55	80.90

The comparative evaluation was performed to verify how much the proposed forgery discrimination structure was better in character detection than that of other state-of-the-art models. Tab. 5 summarizes the results. M_{face} is an optimized forensic model for face detection, which is widely used in Deepfake detection, and M_g and M_{gb} are preprocessed forgery and face detection models using Gaussian noise and Gaussian blur noise. For numbers, the accuracy of M_{face} using the DCGAN- and WGAN-generated test data was higher than that of M_{gb} and M_g but lower than that of M_{g+sp} by 5.55% and 6.08%, respectively. In terms of execution flows, M_{gb} and M_g were similar to M_{g+sp} in that they performed image preprocessing followed by CNN training, but they were focused on face identification. Thus, their accuracies degraded by 23.46% and 27.99%, compared with that of M_{g+sp} when testing the generalizability usings WGAN-generated images, indicating that they were highly vulnerable to number detection. Moreover, for Hangul, the accuracy of M_{g+sp} in the generalizability test using WGAN-generated images showed 36.59%, 21.8%, and 21.95% better than those of M_{face}, M_g, and M_{gb}, respectively, which verified that the proposed method was much more suitable to Hangul and number detection forged with a GAN, showing robustness to new forms, which were GAN-generated images.

Table 5: Performance comparisons of the proposed discriminator with state-of-the-art CNN forensic models

Data type	Test generative model	Forensic model	Accuracy [%]	TPR [%]	TNR [%]
Number	DCGAN	M_{face}	90.98	90.16	42.71
		M_g	78.01	98.02	58.01
		M_{gb}	86.00	75.97	96.03
		M_{g+sp}	**96.50**	**97.93**	**95.07**

(Continued)

Table 5: Continued

Data type	Test generative model	Forensic model	Accuracy [%]	TPR [%]	TNR [%]
	WGAN	M_{face}	75.99	90.16	42.71
		M_g	54.80	98.02	11.14
		M_{gb}	58.61	75.97	41.26
		M_{g+sp}	**80.02**	**97.93**	**62.10**
Hangul	DCGAN	M_{face}	55.33	80.8	28.87
		M_g	81.49	65.5	97.48
		M_{gb}	78.53	66.03	91.02
		M_{g+sp}	**93.55**	**95.55**	**91.48**
	WGAN	M_{face}	51.63	22.47	80.80
		M_g	66.42	65.50	67.34
		M_{gb}	66.27	66.03	66.51
		M_{g+sp}	**88.22**	**95.55**	**80.90**

Tabs. 6 and 7 present examples of detection results by forensic models for numbers and Hangul, respectively. The first row shows the test data images, and the second row indicates the image is genuine or fake. The original image that is not faked is indicated as T (True) and the forged image as F (False). T or F cannot be distinguished by naked eyes, but the proposed forensic model can discriminate T or F of Hangul and numbers correctly.

Table 6: Examples of detection results by forensic models for numbers

Forensic model										
Label	T	F	F	T	F	F	T	T	F	T
M_{g+sp}	T	F	F	T	F	F	T	T	F	T
M_{face}	T	F	**T**	T	**T**	F	T	T	**T**	T
M_{gb}	T	F	**T**	T	F	F	T	T	**T**	**F**

Table 7: Examples of detection results by forensic models for Hangul

Forensic model										
Label	T	F	F	T	T	F	T	T	F	T
M_{g+sp}	T	F	F	T	T	F	T	T	F	T
M_{face}	T	**T**	**T**	**F**	**F**	F	**F**	T	**T**	**F**
M_g	T	F	F	T	**F**	**T**	T	**F**	**T**	T

4 Conclusions

Recently, considerable attention has been paid to forgery and falsification using a GAN, which is rapidly advancing, and various techniques to detect them have been proposed. However, most studies were focused on face detection only. The forgery of documents using a GAN can completely change the context and meaning of the document and is much easier than that of a fake face, which makes it more difficult to distinguish from the original images. As a result, the crime rate of document forgery is steadily increasing. Thus, a new method to detect document forgery has been constantly demanded. In this context, this study performed an optimization process of the CNN forensic model to detect characters (Hangul and numbers) and increased the reliability of the possibility to detect text forgery. To overcome the shortcoming that the generalizability of models degrades due to specific fingerprints in images when detecting GAN-generated images using a CNN, changes in pixel values were given using the Gaussian Noise Pattern, and an image preprocessing process was proposed by adding a salt and pepper noise pattern, which could play a similar role as visible watermarking or anticounterfeit noise of documents. To verify how much more the proposed forgery discrimination structure was effective in character detection than that of other state-of-the-art models, comparative evaluations were performed on the basis of WGAN-generated numbers and Hangul test data, which have never been used in learning before. The evaluation results showed that the accuracy improved by more than 5% and 20% for numbers and Hangul, respectively, which verified that the proposed forgery detection structure was suitable for detecting Hangul and numbers forged by a GAN, and also demonstrated high robustness against new types of GAN-generated images.

Acknowledgement: The authors thank the editor and anonymous reviewers for their helpful comments and valuable suggestions.

Funding Statement: This research was funded by a National Research Foundation of Korea (NRF) grant funded by the Korean government (MOE) (No. 2021R1I1A3055973) and the Soonchunhyang University Research Fund.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] Ministry of public administration and security of the Republic of Korea. Document 24 the number of uses. 2020. [Online]. Available: https://www.mois.go.kr/frt/bbs/type010/commonSelectBoardArticle.do?bbsId=BBSMSTR_000000000008&nttId=80660.

[2] Statistics Korea. Crime occurrence and arrest status (national). 2021. [Online]. Available: https://kosis.kr/statHtml/statHtml.do?orgId=132&tblId=DT_13204_2011_211&vw_cd=MT_ZTITLE&list_id=132_13204_GKIT659_dike256_eii6&scrId=&seqNo=&lang_mode=ko&obj_var_id=&itm_id=&conn_path=E1.

[3] D. Nhu-Tai, I. S. Na and S. H. Kim, "Forensics face detection from GANs using convolutional neural network," in *Proc. Int. Symp. on Information Technology Convergence (ISITC)*, Jeju-si, Jeju-do, South Korea, 2018.

[4] J. Fridrich, D. Soukal and J. Lukas, "Detection of copy-move forgery in digital image," in *Proc. Digital Forensic Research Workshop (DFRWS)*, Cleveland, Ohio, USA, pp. 55–61, 2003.

[5] I. J. Goodfellow, J. P. A. Badie, M. Mirza, B. Xu, D. W. Farley *et al.,* "Generative adversarial nets," in *Proc. Advances in Neural Information Processing Systems 27: Annual Conf. on Neural Information Processing Systems (NIPS)*, Montreal, Quebec, Canada, pp. 2672–2680, 2014.

[6] A. Radford, L. Metz and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in *Proc. Int. Conf. on Learning Representations (ICLR)*, San Juan, Puerto Rico, 2016.

[7] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin and A. Courville, "Improved training of wasserstein GANs," in *Proc. Neural Information Processing Systems (NIPS)*, Long Beach, CA, USA, pp. 5769–5779, 2017.

[8] T. Karras, S. Laine and T. Aila, "A style-based generator architecture for generative adversarial nerworks," in *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, pp. 4401–4410, 2019.

[9] P. Isola, J. Y. Zhu, T. Zhou and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp. 1125–1134, 2017.

[10] J. Y. Zhu, T. Park, P. Isola and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. on Computer Vision (ICCV)*, Venice, Italy, pp. 2223–2232, 2017.

[11] T. Kim, M. Cha, H. Kim, J. K. Lee and J. Kim, "Learning to discover cross-domain relations with generative adversarial networks," in *Proc. Int. Conf. on Machine Learning*, Sydney, Australia, pp. 1857–1865, 2017.

[12] T. T. Nguyen, Q. V. H. Nguyen, C. M. Nguyen, D. Nguyen, D. T. Nguyen *et al.,* "Deep learning for deepfakes creation and detection: A survey," arXiv:1909.11573v3, 2019.

[13] T. Barras, T. Aila, S. Laine and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," in *Proc. Int. Conf. on Learning Representations (ICLR)*, Vancouver, BC, Canada, pp. 447, 2018.

[14] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen *et al.,* "Training generative adversarial networks with limited data," in *Proc. Conf. on Neural Information Processing Systems (NIPS)*, Vancouver, BC, Canada, 2020.

[15] G. Giacinto and F. Roli, "Design of effective neural network ensembles for image classification purposes," *Image and Vision Computing*, vol. 19, no. 9–10, pp. 699–707, 2001.

[16] F. Marra, D. Gragnaniello, L. Verdoliva and G. Poggi, "Do gans leave artificial fingerprints?," in *Proc. IEEE Conf. on Multimedia Information Processing and Retrieval (MIPR)*, San Jose, CA, USA, pp. 506–511, 2019.

[17] A. Badale, C. Darekar, L. Castelino and J. Gomes, "Deep fake detection using neural networks," *International Journal of Engineering Research & Technology*, vol. 9, no. 3, pp. 349–354, 2021.

[18] H. Mo, B. Chen and W. Luo, "Fake faces identification via convolutional neural network," in *Proc. ACM Workshop on Information Hiding and Multimedia Security*, Innsbruck, Austria, pp. 43–47, 2018.

[19] Y. Fu, T. Sun, X. Jiang, K. Xu and P. He, "Robust GAN-face detection based on dual channel CNN network," in *Proc. Int. Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, Huaqiao, China, pp. 1–5, 2019.

[20] X. Xuan, B. Peng, W. Wang and J. Dong, "On the generalization of GAN image forensics," in *Proc. Chinese Conf. on Biometric Recognition*, Zhuzhou, China, pp. 134–141, 2019.

[21] G. Tang, L. Sun, X. Mao, S. Guo, H. Zhang *et al.,* "Detection of GAN-synthesized image based on discrete wavelet transform," *Security and Communication Networks*, vol. 2021, Article ID 5511435, pp. 1–10, 2021.

[22] S. Albawi, T. Mohammed and S. Alzawi, "Understanding of a convolutional neural network," in *Proc. Int. Conf. on Engineering and Technology (ICET)*, Antalya, Turkey, pp. 1–6, 2017.