Forensics Face Detection From GANs Using Convolutional Neural Network

Nhu-Tai Do[1], In-Seop Na[2], Soo-Hyung Kim[1]

[1]School of Electronics and Computer Engineering,
Chonnam National University
77 Yongbong-ro, Buk-gu, Gwangju 500 – 757, Korea
donhutai@gmail.com, shkim@chonnam.ac.kr

[2]Software Convergence Education Institute
Chosun University
375 Seosuk-dong, Dong-gu, Gwangju, Korea
ypencil@hanmail.net

***Abstract*—The rapid development of Generative Adversarial Networks (GANs) brings the new challenge in anti-forensics face techniques. Many applications use GANs to create fake images/videos leading identity theft and privacy breaches. In this paper, we proposed a deep convolutional neural network to detect forensics face. We use GANs to create fake faces with multiple resolutions and sizes to help data augments. Moreover, we apply a deep face recognition system to transfer weight to our system for robust face feature extraction. In additional, the network is fined tuning suitable for real/fake image classification. We experimented on the validation data from AI Challenge and achieved good results.**

Keywords—GANs; fake face detection; forensics image; Deep Convolution Neural Netwrok

I. INTRODUCTION

In the rapid development of social networking as well as the popularity of digital cameras through mobile phones, anti-forensics techniques are one of the key challenges to identify the truthfulness of digital publications on the social in front of the powerful development of image/video editing software, especially artificial intelligence techniques [1].

Previous anti-forensic techniques often focused on the analysis of specific correlated cues or patterns at the stages of the digital image creation/manipulation process such as image acquisition, storage and editing. In the image acquisition step, the features will be observed at signal levels such as lens aberrations [2], color filter arrays (CFA) artifacts [3], etc. At the image storage, the features focused on the property of image coding, particularly the lossy data compression methods such as JPEG with jpeg ghosts, or artificial blocks [4]. At the editing step, the physical level view will focus on the properties of light conditions, shadows and light reflections [5], as well as local filters such as median filter, un-sharp masking [6], etc. Besides, the semantic level view will find the abnormalities of similarity and consistency among the image patches.

Among various types of fake image detection methods, machine-based techniques play an important role. These techniques are modeled as binary classification problems. It receives hand-crafted features, explores hidden knowledge, and distinguishes fake images from editing operation such as enhancement (histogram equalization, color change, etc.), geometry changes (rotation, cropping, shearing), and content changes (copy-move, cut-paste, etc.).

Fig.1. Forensics in AI Challenge Contentest [11] with image size 1024x1024, 256x256, 64x64

Following the success of deep convolution neural networks (DNNs) in image classification, object detection, etc. end-to-end learning solutions based on DNNs are designed to take advantage of automated learning and features extracting. Ouyang et al. [7] proposed a method based on DNNs to resolve copy-move forgery detection. Kim et al. [8] proposed median-filter forensic method based on DNN. Bayar et. al. [9] built network architecture to detect image editing at multiple times.

Deep learning, however, also promotes forensics face artificial intelligence techniques from the advancement of new generations of GANs. These models have been researched to generate large-scale and diversity dataset helping for training DNNs as well as features extraction. However, it is also a powerful tool to generate fake data, especially face to digital contents to spread with bad content.

Based on GANs, many fake face software such as DeepFake, FakeApp [10] uses them to create fake face videos replacing original faces into the specific faces. Using a huge database up to hundred thousand of face images, these software easily generate new fake faces in the images and videos resulting in identity theft and privacy breaches, which causes serious legal consequences.

In the paper, our purpose is to propose a deep convolutional neural network for detecting real image or fake image from GANs. The results of proposed method are based on evaluation from the AI Challenge contest [11] shown in Fig.1. The challenges in the contest come from data issues. It has very fewer samples for validation task. In additionally, the data is

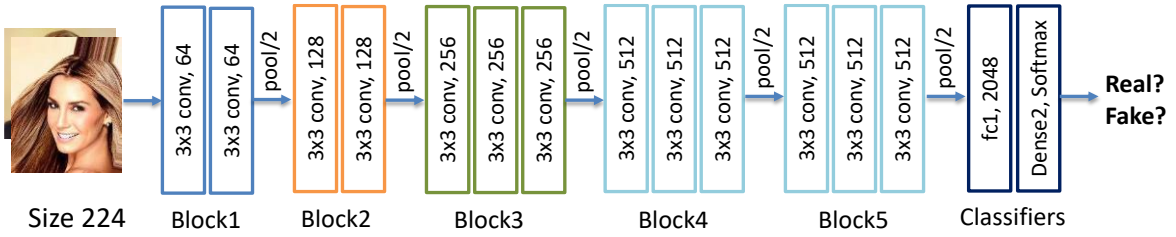

Fig.2. Forensics Face Detection Architecture

created from many GANs with many image sizes and resolutions.

We summarize the three main contributions of the paper. Firstly, we build training data sets that can be adapted to the test data set of the AI Challenge contest. Secondly, we build a deep learning network based on face recognition networks to extract face features. Thereby, we use fine-tuning to make face features suitable for real/fake face classification. Finally, we obtained good results from the contest validation data.

The rest paper consists of three parts. Firstly, the second part will describe the propose method for forensics face detection. Next, the third part will focus on experiments and results. The final part is conclusion of paper.

II. Proposed Method

A. *Forensics Face Data Generation from GANs*

The original idea of GANs came from Goodfellow et al. [12] for proposing adversarial nets, which contain a pair of models. The generator learns to generate a fake pattern against distinguishing from the discriminator through the training pattern. This model has successfully generated similar patterns in the MNIST dataset.

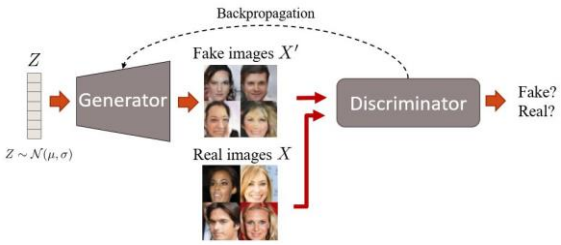

Fig.3. GAN Architecture Overview

Radford et al. [13] proposed deep convolutional generative adversarial network (DC-GAN), which contains the constrains based on convolutional layer leading to more stable training. It also sets the specific properties for generation representation as arithmetic operation, smooth transitions. They are also applied successfully in the face dataset dbpedia.

Karras et al. [14] proposed progressive training (PG-GANs) through starting with low resolution, adding new layers increasingly in the training process to increase detail. Since then, the authors have successfully reproduced forensics faces with high resolution like real image from the face dataset CelebA.

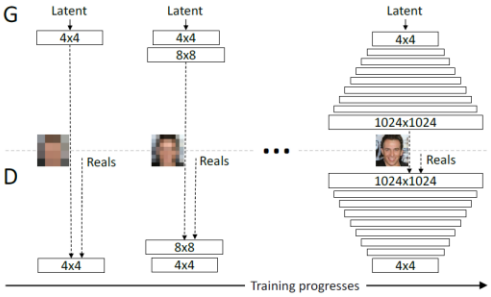

Fig.4. Training Processes in PG-GAN [14]

In this paper, we choose the DC-GANs to generate images with size 64x64, and PG-GANs for image size 256x256, 1024x1024. The purpose helps training data to fit with various image size and quality.

B. *Deep Face Representation*

Deep face recognition systems usually contain three main modules such as face processing, deep feature extraction and face matching. Firstly, face processing is responsible for face normalization to frontal view to increase accuracy. Next, deep feature extraction is a deep learning network architecture that is trained on large-scale face data sets to identity the specific people with appropriate loss functions. After that, these networks usually remove the fully connected layer in order to derive feature vectors that represent faces as well as normalize these vectors. Finally, the face matching module will take these face feature vectors and through deep metric learning to create a suitable model to calculate the distance between the face feature vectors for face identification and verification. [15]

Face recognition systems often use AlexNet, VGG-Net, ResNet, SENet, etc. for backbone network architectures to extract face representation. In this paper, we choose Deep Feature Extraction in VGGFace [16] for deep face extraction through the VGG-Net described in Fig. 2.

The VGG-Net consists of five-layer blocks including convolutional and max-pooling layers in each block for feature extraction task. Finally, the fully connected layer connects to the K-way softmax (where K is the number of classes) to output the correct probability of the corresponding face identities. We use

pre-train weights from VGG-Face without the fully connected layer. The length of face feature vector is 512.

C. Fine-Tuning for Fake Face Classification

After extracting the face feature, we performed fine-tuning by adding a new fully connected layer after feature representation blocks, which is connected to a 2-way softmax for real/fake binary classification as Fig.2.

When fine-tuning the deep neural network, we need to adjust the weight of the classifier layer or top presentation blocks next to the classifier. Smaller learning rates are used to adjust gradually to the appropriate weights for classification on fake/real images.

Moreover, to balance with amount of image between training (over two hundred thousand of images) and validation (only two hundred images) dataset, we apply data augment techniques as randomly flip, rotation, etc. Besides, we choose the number of training samples in an epoch with 80/20 ratio to number of validating samples. It helps the distribution domain of training data suitable for validation data. Additionally, it is not also reduce generalization of training data on small validation data.

III. Experiments and Results

A. Environments

We built the system on a Windows environment with Python 3.5. In the system, we use the Keras library based on Tensorflow for developing Deep Learning models. The training machine has the core i5-3470 CPU with a GTX 1080 graphics card 16 GB of RAM.

B. Forensics Face Data Generation for training

We use the CelebA face dataset with 292,599 images. For forensics fake, we download a set of GAN face images from PG-GAN with a high resolution of 1024x1024 called Celeb-A HQ [17]. A high resolution and good quality fake face images include 200,000 images as shown in Fig.5.

Fig.5. Forensics Faces PG-GAN with high quality and image size 1024x1024

In addition, the paper applies DC-GAN to train and produce 200,000 fake photos from the Celeb-A set with each 64x64 image as shown in Fig.6.

Fig.6. Forensics Faces DC-GAN with image size 64x64

We also use PG-GAN to produce 256x256 images with 200,000 images as shown in Fig.7.

Fig.7. Forensics Faces PG-GAN with image size 256x256

C. Evaluation Dataset

We use the evaluation data from the first mission of the AI Challenge contest [11] to test the performance of fake image classifiers from GANs. The first mission consists of 400 images with 200 fake images and 200 real images as Fig.1. The images have multiple sizes such as 64x64, 256x256, and 1024x1024. Fake images quality is different. It is difficult for fake images to distinguish. Otherwise, some fake images have noise, so it is easy to recognize.

TABLE I. Evaluation Data from AI Challenge Contest

	Real Images	**Fake Images**
Evaluation Data	200	200

D. Assesment Method

The performance of a solution determined by Area under the ROC Curve (AUROC). ROC curve is a graphical plot to show the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings.

E. Results

Our proposed method uses VGG16 architecture as Fig.2 with pre-train weights from VGG-Face. Besides, we also do some experiments with ResNet architecture and pre-train weights from ImageNet.

The accuracy of the methods in the paper shows in Table 2:

TABLE II. The Performance of Methods

Method Name	**Accuracy**	**AUROC**
VGG-Face VGG16	**80%**	**0.807**
ImageNet VGG16	76%	0.765
VGG-Face ResNet50	73%	0.766

The RPC Curve shows the performance of evaluation methods as Fig. 8:

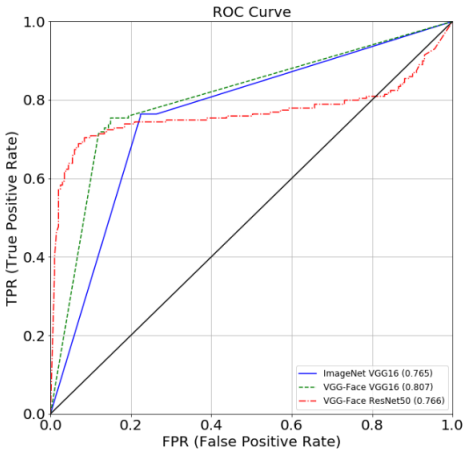

Fig. 8. The ROC Curve of evaluation methods.

The system has the accuracy 80% and AUROC 0.807. This result shows the method has good performance in forensics detection.

IV. Conclusion

In summary, we present a GAN forensics face detection. Forensics face generation with multiple sizes and resolutions uses GANs such as PG-GAN, DC-GAN. The network architecture applies convolution neural network in deep feature extraction and binary classification. Besides, we suggest fine-tuning method to train with validation data in AI Challenge contest with the good performance.

Acknowledgment

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(NRF-2017R1A4A1015559), and Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIT) (No.2017-0-00383,Smart Meeting: Development of Intelligent Meeting Solution based on Big Screen Device).

References

[1] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, “FaceForensics: A Large-scale Video Dataset for Forgery Detection in Human Faces,” arXiv:1803.09179, 2018.

[2] K. S. Choi, “Source camera identification using footprints from lens aberration,” Int. Soc. Opt. Photonics, 2006.

[3] P. Ferrara, T. Bianchi, A. De Rosa, and A. Piva, “Image forgery localization via fine-grained analysis of CFA artifacts,” IEEE Trans. Inf. Forensics Secur., 2012.

[4] Y. L. Chen and C. T. Hsu, “Detecting recompression of JPEG images via periodicity analysis of compression artifacts for tampering detection,” IEEE Trans. Inf. Forensics Secur., 2011.

[5] M. K. Johnson and H. Farid, “Exposing digital forgeries by detecting inconsistencies in lighting,” in Proceedings of the 7th workshop on Multimedia and security - MM&Sec ’05, 2005.

[6] F. Ding, G. Zhu, J. Yang, J. Xie, and Y. Q. Shi, “Edge perpendicular binary coding for USM sharpening detection,” IEEE Signal Process. Lett., 2015.

[7] J. Ouyang, Y. Liu, and M. Liao, “Copy-move forgery detection based on deep learning,” Proc. - 2017 10th Int. Congr. Image Signal Process. Biomed. Eng. Informatics, CISP-BMEI 2017, 2018.

[8] 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 Process. Lett., 2018.

[9] B. Bayar and M. C. Stamm, “A Deep Learning Approach to Universal Image Manipulation Detection Using a New Convolutional Layer,” in Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security - IH&MMSec ’16, 2016.

[10] Wikipedia, “Deepfake,” https://en.wikipedia.org/wiki/Deepfake, 2018. .

[11] ″AI Challenge,” http://airndchallenge.com/g5/, 2018. .

[12] I. Goodfellow et al., “Generative Adversarial Nets,” Adv. Neural Inf. Process. Syst. 27, 2014.

[13] A. Radford, L. Metz, and S. Chintala, “Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks,” in ICLR, 2016.

[14] T. Karras, T. Aila, S. Laine, and J. Lehtinen, “Progressive Growing of GANs for Improved Quality, Stability, and Variation,” arXiv: 1710.10196, 2017.

[15] M. Wang and W. Deng, “Deep Face Recognition: A Survey,” arXiv:1804.06655v3, 2018.

[16] O. M. Parkhi, A. Vedaldi, and A. Zisserman, “Deep Face Recognition,” in Procedings of the British Machine Vision Conference 2015, 2015.

[17] GitHub, “Celeb-A HQ,” https://github.com/tkarras/progressive_growing_of_gans, 2018.

NOTICE to authors

On behalf of the ISITC'2018 Organizing Committee, we would like to kindly ask authors to let us know authors' preferences in addition to paper submission. We deeply appreciate authors' cooperation for better preparation of ISITC'2018.

Specify **priorities** in the brackets for **two preferred tracks** to be assigned in the conference program in the case of acceptance.

[x] Signal and Image Processing for IT Convergence
[ ] Web and Database Technology for IT Convergence
[x] IT Convergence in Bio-inspired Intelligence
[ ] IT Convergence in Health Care
[ ] IT Convergence in Robotics
[ ] IT Convergence in Transportation System
[ ] Internet of Things
[ ] Human-Computer Interaction
[ ] Virtual Reality
[ ] Embedded Systems
[ ] Wireless Technology
[ ] IT and Cultural Innovation
[x] Convergence Applications