REVIEW ON DETECTION OF GAN GENERATED FAKE IMAGES OVER SOCIAL NETWORKS

Vandhana S[1], Vishnupriya J[1], Athira C[1], Saritha C[1], Reshma Mohan A S[2]

[1]UG Scholar, Department of Electronics and Communication Engineering,
Dr. APJ Abdul Kalam Technological University Kerala, India
[2]Head of the Department, Department of Electronics and Communication Engineering,
Dr. APJ Abdul Kalam Technological University Kerala, India

---***---

Abstract - *In the modern world, fake images are easily spreading everywhere, there by the humans are fooled in day to day life. So, here uses a well known technology named GAN. Generative adversarial networks (GANs) can be used to generate a photo-realistic image from a low-dimension random noise. Such types of fake images with unnecessary content can be used on social media networks, it can cause severe and difficult problems. So the aims to successfully detect fake images, an effective and efficient image forgery detector is necessary. Recent advances in Generative Adversarial Networks (GANs) have mainly shown increasing and immediate success in generating photorealistic images. But they can also raise challenges to visual forensics as well as model attribution. Image to image translation based on GAN is one of the dangerous to learn a mapping between images from a source domain and images from a target domain. The enormous application prospect, including image and vision computing, video and language processing, etc. Besides, in this paper also tells that the background of the GAN and its theoretic models and also explains that how to detect GAN generated fake images over social media. It is a most dangerous situation to all.*

Keywords - GAN, Convolutional neural networks, Image to image translation

1. INTRODUCTION

Recently, deep learning-based generative models, such as variational auto encoders and generative adversarial networks (GANs), have been used to synthesize the photo-realistic images partially or whole content of an image as well as video.

For instance, the cycle GAN can be used to synthesize the fake face image in a pornography video [4]. Furthermore, the GANs can also create speech video with the synthesized facial content or expressions of any famous politician, or any others which becomes causing severe problems to the society, political, and other such activities. Therefore, an effective and immediate fake face image detection technique is compulsory. In this paper, our previous study is extended to recognize generated fake images effectively and efficiently.

Since deep neural networks have been widely used in areas such as various recognition tasks. We can also adopt such a deep neural network to detect fake images generated by the GANs. Here, we are studied that the deep learning based approach for fake image detection using supervised learning. Also, fake image detection has been treated as a binary classification problem or model (i.e., fake or real image). For instance, in this case the convolution neural network (CNN) network was used to develop the fake image detector.

Figure1: Generative Adversarial Networks

2. LITERATURE REVIEW

All researchers have an aim to develop which generate and detect GAN generated fake images. Also tried to improve the accuracy of detection of GAN by preprocessing and segmentation feature extraction.

Mr. P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, said that conditional adversarial networks is a basic solution to image to image translation. Many problems in image processing such as graphics, and vision involves translating an input image into aits corresponding output image. So Conditional adversarial networks are a basic general-purpose solution that appears good working well on a wide variety of these situations. This approach is good effective at synthesize photos from labelled maps, to reconstruct objects from edged maps, and colorizing images etc. Here, uses 3 types of GANs mainly: Pixel-GAN, Patch-GAN, Image-GAN. The Pixel-GAN has no effect on spatial sharpness but they increase the colourness of the outputs. Patch-GAN is sufficient to promote sharp and pointed outputs, and also achieves good FCN-scores, but it leads to artifacts. Image-GAN, does not appear to improve the visual quality of the output. Besides, in this paper also suggest that conditional adversarial networks are a promising approach for many image-to-image translation tasks, especially involving highly structural graphical output.

V. Schetinger, M. Oliveira, R. da Silva, and T. Carvalho, proposed that different ways to avoid lucky guess when evaluating users answers. In our world, digital images are spreading everywhere, from social media to news and also in scientific papers. There are 5 possible outcomes, the image can be either T(True) or F(False), the user can either T or F and if the user answers choose F , he can provide either valid or invalid evidence. Here, treat the users answers as a binary classification problem where there are trying to identify F images. Thereby, it shows that the importance of evidence evaluation mask (with each F image from test database we have associated a binary mask).As a result, masks are also created. Take both true and false images and test together. Hence, identify that true images are positive and false images are negative. Masks are created. Also, this paper gives an idea about ability of human to detect forgery in digital images. Its finding suggest that humans can be easily fooled.

N. Rahmouni, V. Nozick, J. Yamagishi&I. Echizeny, This paper is about to check whether an image is a computer graphics or natural photographs. The proposed method is convolutional neutral network. Here distinguishing CG from PG by related to computer graphics performance to generate photo -realistic images. Here done by the process are statistical feature extraction, histogram and filtering. Here using a boosting method used for the estimation of the label. This paper gives an idea about how to distinguish computer graphics from natural images.

M. Villan, A. Kuruvila, K. J. Paul &E. P. Elias ,proposed that how to detect fake images in spreading through the social media. Here implemented a approach that can be determine whether an image is real or fake with the help of machine learning. Here using mainly three steps, first one is metadata analysis for provides information about an image pedigree. Second one is error level analysis and third one machine learning, it proposed using Neuroph library for java. Here using the java programming language for extracting metadata of images. Here take the input and output values of neurons during testing and training. This paper gives brief explanation how to approach to identify the image is real or fake.

K. PMurphy, This paper is about how machine learning helps in datamining method for analysing data increase with increase in amount of data.Machine learning is done by a probabilistic manner. This is implemented in a MATLAB software package called PMTK ie, probabilistic model tool kit. This paper also shows detailed explanation about MATLAB and PMTK.

J. Kos,I. Fischer &D. Song , This paper proposed that the adversarial examples for generative models. Variational auto encoder (VAE)and the VAE GAN are main example. Here some attacks are given to the above deep generative models for the processing. The attacks helps to generate target reconstruction image image from adversarial example, also shows latent representation and explain the differences in source and target latent representation. These attacks are against the networks framed on MNIST, SVHN and teleba. Also detailed about how to build more robust networks and concluded that adversarial examples are general phenomenon for current neural network.

Good fellow, j. Pougetadabie, This paper is completely about GAN. Generative adversarial network has a great role in the field of artificial intelligence. It consist of two parts, which are generator and discriminator for the working the GAN should be trained first. It has so many applications such as image super resolution, image translation, face synthesis etc. This paper also explain about different types of GANs.

A Radford, **L Metz and S Chintala,** This paper is about unsupervised representation learning with deep Convolutional generative adversarial networks has been huge adoption in computer vision applications. Unsuper-vised learning with CNN has received less attention. Training on various image dataset are convincing evidence in DCGANs that hierarchy of representation from object part in generator and descriminator. In this paper detailed representation of image for supervisd learning and generative modeling

O. M Parkhi, A vedaldi and A Zisserman, This paper is about goal of face recognition. This paper propose a two contribution. First large scale dataset can be constructed by semiautomatic condition. Second training procedure that achieve face recognition.This paper about the comparison of dataset. This paper procedure to achieve comparable state of the result on the standard LFW& YTF face benchmark.

Nhu-Tai Do, In-Seop Na, Soo-HyungKim,have been proposed a deep Convolutional neural network to detect forensic face. Here uses 2 type of Gans mainly PG- GAN, DC-GAN. DC- GAN mainly composes of Convolutional layers without max pooling or fully connected layers. PG-GAN models can have the capability of generating photorealistic synthetic images & objects at high resolution. DC-GAN to generate image with size as 64*64and PG-GAN for image size as 256*256 and 1024*1024.Here,this paper only uses 4*4 pixels are used. In this paper proposed good quality fake face images include 20000 image are trained and produce 20000 fake photos. Mission consist of 400 images with 200 fake images & 200 real images.

3. CONCLUSION

In this paper, we have proposed a novel method to detect GAN generated fake images over social networks using deep learning successfully. The proposed technology called GAN helps to detect fake data from real data. And also, the proposed GN network can be used to improve the performance of fake image detection further. With the proposed system, the GAN network should be able to have the ability to detect fake images generated by GAN. The proposed Cycle GAN allows us to train a GAN fake image classifier without need fake images as training data or specific GAN models used for generating fake images.

REFERENCES

[1] V. Schetinger, M. Oliveira, R. da Silva, and T. Carvalho, “Humans are easily fooled by digital images,” Computers & Graphics, vol. 68, pp. 142–151, 2017.

[2] D. Cozzolino, D. Gragnaniello, and L.Verdoliva, “Image forgery detection through residual-based local descriptors and block-matching,” in IEEE Conference on Image Processing (ICIP), October 2014, pp. 5297–5301.

[3] S. Fan, T.-T. Ng, B. Koenig, J. Herberg, M. Jiang, Z. Shen, and Q. Zhao, “Image visual realism: From human perception to machine computation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, in press, 2017.

[4] S. Lyu and H. Farid, “How realistic is photorealistic?” IEEE Transactions on Signal Processing, vol. 53, no. 2, pp. 845 –850, 2005.

[5] R. Wu, X. Li, and B. Yang, “Identifying computer generated graphics via histogram features,” in IEEE ICIP, 2011, pp.1933–1936.

[6] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” in IEEE CVPR, 2017.

[7] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in IEEE ICCV, 2017.

[8] [8] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, andR. Webb, “Learning from simulated and unsupervised imagesthrough adversarial training,” in IEEE CVPR, 2017.

[9] M.-Y. Liu, T. Breuel, and J. Kautz, “Unsupervised image-to-image translation networks,” in NIPS, 2017.

[10] N. Haouchine, F. Roy, H. Courtecuisse, M. Nießner,and S. Cotin, “Calipso: Physics-based image and videoediting through cad model proxies,” arXiv preprintar Xiv:1708.03748, 2017.

[11] M. Zampoglou, S. Papadopoulos, and Y. Kompatsiaris, “Large-scale evaluation of splicing localization algorithms forweb images,” Multimedia Tools and Applications, vol. 76,no. 4, pp. 4801–4834, february 2017.

[12] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative ad-versarial nets,” in Advances in neural information processing systems, 2014,pp.2672–2680.

[13] D. Cozzolino, G. Poggi, and L. Verdoliva, “Recasting residual-based local descriptors as convolutional neural net-works: an application to image forgery detection,” in ACMWorkshop on Information Hiding and Multimedia Security,2017, pp. 1–6.

[14] B. Bayar and M. Stamm, “A deep learning approach touniversal image manipulation detection using a new convolutional layer,” in ACM Workshop on Information Hiding andMultimedia Security, 2016, pp. 5–10.

[15] G. Huang, Z. Liu, L. van der Maaten, and K. Weinberger, “Densely connected convolutional networks,” in IEEE CVPR,2017, pp. 4700–1708.

[16] C. Szegedy, V. Vanhoucke, S.Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision,"in IEEE CVPR, 2016, pp. 2818–2826.

[17] F. Chollet, "Xception: Deep learning with depthwise separable convolutions, "in IEEE CVPR, 2017, pp. 1800–1807.

[18] J. Fridrich and J. Kodovsky, "Rich Models for Steganalysis of ´Digital Images," IEEE Transactions on Information Forensics and Security, vol. 7, no. 3, pp. 868–882, Jun. 2012.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in IEEE CVPR, 2016, pp. 770–778.

[20] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions,"in IEEE CVPR, 2015.