Identifying fake image using Machine learning

Zubair Baqai

DIAG, Universita di Roma “La Sapienza”, Italy

ABSTRACT

In this digital era, where people use Social media on daily bases to share their photos, or view the Images/Videos of their friends and family. Facebook and Instagram are the most widely used Social networks , according to stats, daily 350 million images are uploaded on Facebook alone , Among which could be personal photos of individual / family / friends / political parties campaign . An easy access to such photos can becomes easier for someone to fabricate with state of art techniques such as DeepFakes .

I. INTRODUCTION

The Emergence of Social Media , has changed the way of interaction of people and their daily routines . Facebook has over 2.6 billion monthly active users , based on the report of 2019 , Facebook removed 2.2 Billion fake accounts . Many of these fake accounts had motives to effect the lives of other individual by sharing falsified information . And based on another stat Facebook average took 4.8 days to detect and delete each account . This implies that there are 119 million fake accounts at any instance. This delay in removal or detection of an account is enough for one to cause damage to individual / group or whole community with fake Content .Apart from social media aspect, there are many other domains where Fake images causes huge negative impacts ,e.g Politics campaign , Court of law etc . To tackle such problems many different techniques have been proposed but in this research we will discuss about the state of art neural based Approaches

Figure 1 - Error level Analysis of a fake Image

II. NATIVE TECHNIQUES FOR GENERATION OF FAKE IMAGES

Criminals have been coming with new techniques and algorithms to forge the images . Few of the techniques to forge Images / Documents are following , i)Content-Aware fill and Patch Matching ii) Content-Aware Healing iii)Clone-Stamping iv)seam carving v)Image painting vi)Alpha Matting . These have been classic techniques by which criminals have been successful making fake images , but they can be classified with Real / fake as they leave traceable marks on image which can be identified by tools such as ‘FotoForensics’ which underneath uses advances algorithms like Error level Analysis (ELA) which aims to find patches that have different error levels than other patches , as JPEG images are supposed to have roughly the same

error level across whole image as demonstrated on Figure 1

III. Neural Based Techniques for Generation of Fake Contents

Development in generating Fake images specially human face has been moving at a startling pace , the first Neural based technique to generate faces was published in 2014 [**1**] by Ian J.Goodfellow . Which introduced the AI tool today know as Generative Adversarial Network (GAN) . The images that were generated were satisfactory at that time , yet they looked not as perfect and could be easily distinguishable as the faces looked very generic similar to taking an average of multiple images (Dataset) and the generation were in Greyscale images .Since then a lot research had been done , and after 6 years of improvements on Base Model of GAN , it can be seen from Figure 2 , that AI has been drastically improved and its very hard to distinguish between real and fake images generated by GAN . The images generated by recent development in GAN generates images not only in full – color but they cover generation of faces of most of the ethnicity and Age groups . Figure 3 shows few of the results from StyleGAN 2 [2], which is one of state of art algorithm for Image generation. And as we may observe, it would be hard to identify if any of these images are generated from DeepFakes . Rise of such advanced technology brings many unethical misuses , For example Generating Videos against Politicians on which they can be seen giving statements which they normally wouldn't , The following video was generated in which Obama can be seen talking on critical topic [3] . Generation of Such content doesn't require much of expertise as Most of such algorithms have been deployed as standalone application as web

Figure 2 – Evolution of Deep Fake Images

apps , or a simple GUI based programs . For example

i)thispersondoesnotexist - Which is Face Generation using Style-GAN 2

ii) FaceSwap – Which is an openSource Application that can be used to Swap the Faces and generate Videos.

iii)TalkToTransformer – A web app that utilizes the Transformer , which GPT-2 leverages

iv)Voice Cloning – A GUI python Application, that can Clone a voice of your choice at real time .

With that being said , it is necessary to discuss on how to identify such Fake contents be it Image, Video or Fake News [4 , 5], As it is evident how easily fake content can be generated , there is a need to discuss how we can counter these Falsified contents created with negative intention . In this Research we will talk briefly how we can use neural network based Techniques to identify Fake images generated by Deep fakes, So that while utilizing these phenomenal advancements on positive aspects , we can stop its misuses .

IV. *Identifying Deep Fake Images*

As the Technology for generation of Deep fake images progress , the identification of Deep fake images are getting complicated as the detail in images are Improving and hard for human to identify any flaws leave alone the native fake image detection algorithms . In 2018 Researchers discovered that human face images / Videos generated by Deep fake don't show human blinking normally, this is due to that fact that human blink randomly in time, and there is no specific pattern, furthermore the frames retrieved from videos or images have high count of human eyes being opened rather than blinking. Hence algorithm didn't learn to blink. But within few months new research had solved this issue as well , and deep fake Frames were found to be blinking normal enough to not being identified as fake . Similarly many different flaws have been found with each version of Generators , and Discriminators were able to identify them . Yet soon new research were able to improve them and discriminators were fooled .
One deep fake detection algorithm which was unveiled in early May 2019 by Ekraam Sabir [6], boasted 97 percent accuracy by using **Recurrent Convolution Strategies** . In Figure 4 , we can observe their results they have achieved on images from DeepFake , Face2Face and Face2 Swap .

Another digital forensics technique promises to protect president Trump & other world leaders and celebrities against deepfakes . The research is being funded by Google and DARPA to improve and maintain enhancements in identifying deepfakes .[7]

Spotting the anomalies in Deep Fake images cannot a Simple task for human eye , however observing following details may help .

Figure 3 – Results From Style-GAN 2

Manipulation	Frames	FF++ [34]	ResNet50	DenseNet	ResNet50 + Alignment	DenseNet + Alignment	ResNet50 + Alignment + BiDir	DenseNet + Alignment + BiDir
Deepfake	1	93.46	94.8	94.5	96.1	96.4	-	-
	5	-	94.6	94.7	96.0	96.7	94.9	**96.9**
Face2Face	1	89.8	90.25	90.65	89.31	87.18	-	-
	5	-	90.25	89.8	92.4	93.21	93.05	**94.35**
FaceSwap	1	92.72	91.34	91.04	93.85	96.1	-	-
	5	-	90.95	93.11	95.07	95.8	95.4	**96.3**

Figure 4 : Recurrent Convolution Results

i. Strange Blinkings – So far even the best Deep fake algorithms haven't mastered the motion of Human eye blinking . Focusing strongly on Blinking patterns may signal Fakeness of a video / frame
ii. Facial movement and masculature may appear jerky for instance movement of jaw and head may not be moving normally
iii. Sudden change in Skin tone and environment lightening, especially when there are rapid optical flow motions (Hand Gesture , Turning of face , etc)
iv. Symmetry of Face (Jaw Dropping or Eyes not appearing symmetrical)
v. Looking at the direction of Nose respective to the Face as Demonstrated in Fig 5 below

Figure 5 : Direction of nose W.R.T Face

V. *CONCLUSIONS*

Deep Fake technology is developing with a virus vs. antivirus dynamics, as time progress new viruses are being developed and antivirus identify and prevent them, but for each new virus there have to be newer version of antivirus to counter the virus. Hence it can never be assured with certainty that our community will be safe from all future enhancements in DeepFakes .However with the use of Advance machine learning algorithms as discussed in our research could help us fight the deep fakes by identifying them as early as possible and breaking the chain of fake scandals and information that could harm someone.

VI. *REFERENCES*

1. Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2* (*NIPS'14*). MIT Press, Cambridge, MA, USA, 2672–2680.
2. *Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2019). Analyzing and Improving the Image Quality of StyleGAN. ArXiv, abs/1912.04958.*
3. https://www.youtube.com/watch?v=bE1KWpoX9Hk
4. *Haidar, M.A., & Rezagholizadeh, M. (2019). TextKD-GAN: Text Generation Using Knowledge Distillation and Generative Adversarial Networks. Canadian Conference on AI.*
5. https://openai.com/blog/openai-api/
6. Sabir, Ekraam & Cheng, Jiaxin & Jaiswal, Ayush & AbdAlmageed, Wael & Masi, Iacopo & Natarajan, Prem. (2019). Recurrent Convolutional Strategies for Face Manipulation Detection in Videos.
7. http://www.hao-li.com/publications/papers/cvpr2019workshopsPWLADF.pdf