Face Swap Crisis

Gabriella Rodriguez, Jenniffer Rodriguez, Valerie Cua, Maria Gloria Obono, Gokul Venugopal, Mequanint Moges
Engineering Technology Department
University of Houston

Abstract

A few years ago, artificial intelligence networks known as Generative Adversarial Networks (GANs) gave rise to so-called "Deepfakes" (a portmanteau of "deep learning" and "fake"), video imagery designed to either completely impersonate an individual or superimpose their image/voice onto another visage. Deepfake is a technique for human image synthesis based on Artificial Intelligence (AI). Deepfakes have been used to generate celebrity pornography videos, misrepresent well-known politicians, spreading financial frauds and fake news. Due to these capabilities, this technology poses a grave threat to the reputation of individuals and firms. Such abuse of otherwise wonderous technology needs a societal response swiftly before serious problems start emerging.
Our objective is to understand how a Deepfake is generated and to use deep learning algorithms to detect a fake image by comparing it with a pool of images. Deepfakes can be generated from video footage of two or more individuals using the face-swap application. Then fake images and pool of known images are processed and compared to identify if it is a fake or not. Through this project, we hope to improve the chances of fighting Deepfake and reducing the adverse effects that could have caused by Deepfake videos and images.

Introduction

A Deepfake is the product of a deep learning algorithm, known as GAN, trained to recognize patterns in actual audio and/or visual recordings of a person and manipulating it to generate fake audio/visual recording. As stated in the Bloomberg article[1], it is very easy now to make a deep fake and they are acting as one of the biggest security threats. The impact of the deep fake can be huge, and it can affect financially as well as in terms of reputation as mentioned in the Harvard Business Review[2]. As they say, "The criminal is the creative artist; the detective only the critic."-G.K.Chesterton. Deep fakes can be created in many ways. So, in order to know how to detect a deep fake, we needed to know how it is created. Our plan is to generate deep fakes and formulate a solution to fight it. Through this project, we also hope to educate the crowd and raise awareness.

Method

The project is divided into two phases: Creating deep fake and detecting it. To create deep fake, we needed to capture 500-5000 frames/pictures of two or more people. We have used Faceswap, a python-based application used to convert video footage to images and blend pictures using deep learning. Real Videos and Images of ourselves, faculties and famous celebrities from YouTube were used for this purpose. Although this application generated images and stored it separately for each subject, we went through them to delete irrelevant images. This also improved the efficiency of the

deep fake. Using faceswap, we blended one subject's images on to another subject's image to generate deep fakes. The second stage of the project is to detect deep fake. Due to the strict timeline, we decided to compare unknown pictures against the set of images of known personals in the database instead of creating an anti-GAN model as decided earlier. Identification is done by using the face_recognition library built on a deep learning algorithm. The detection process starts training a convolution neural network that would generate 128 points on each face to measure features on the face. The model is trained over 50000 real images several times to reduce the error. Then deep fake images are passed through the model to see which is the closest match. The number of points that matched is converted to a percentage for checking how real the image is. This value is then shown over the boundary on the image.

Figure 1: The picture on the left shows the impersonation of Trump by Alec Baldwin. The picture in the center shows deep fake by infusing Trump on Alec Baldwin's impersonation to make it look real[3]. The picture on the right shows how real are the test images using our algorithm.

Summary and Conclusions

In summary, we have learned and made our own deep fake version and formulated a solution to detect them. While developing this project, we understood that there are numerous ways of generating deep fakes and there is no single solution to tackle all. We also realized that any solution will have difficulty with deep fakes of people whose information is not available on the internet. Despite the success of our solution, we are aware that it is limited to the set of images we have and there is space for improvement. With this project, we hope to make everyone realize that we can't believe everything we see and hear and give everyone that edge against the people who are making the deep fakes.

References

1. Howcroft, E., 2018 – Bloomberg QuickTake - https://www.bloomberg.com/news/articles/2018-09-10/how-faking-videos-became-easy-and-why-that-s-so-scary-quicktake

2. Ovadya, A., Beinstock, H., 2018 – Harvard Business Review - https://hbr.org/2018/11/is-your-company-ready-to-protect-its-reputation-from-deep-fakes
3. IEEE Spectrum, 2019 – Metadeepfakefake - https://www.youtube.com/watch?time_continue=15&v=G3HfbogI92Q&feature=emb_logo

MEQUANINT MOGES
Dr. Moges serves as the Chair of the College of Technology, University of Houston. He is also a professor of the Electrical Power Engineering Technology department. His research topics include design and optimization of wireless sensor networks, job scheduling in parallel and distributed systems and computational grids and Performance Evaluation and Optimization of Computer and Communication Systems.

GOKUL VENUGOPAL
Mr. Venugopal is a student pursuing a master's degree in Computer Engineering at the Cullen College of Engineering, University of Houston.

GABRIELLA RODRIGUEZ
Ms. Rodriguez is a 2019 graduate of the College of Technology, University of Houston with a degree in Computer Engineering Technology.

JENNIFFER RODRIGUEZ
Ms. Rodriguez is a 2019 graduate of the College of Technology, University of Houston with a degree in Computer Engineering Technology.

VALERIE CUA
Ms. Cua is a 2019 graduate of the College of Technology, University of Houston with a degree in Computer Engineering Technology.

MARIA GLORIA OBONO
Ms. Obono is a 2019 graduate of the College of Technology, University of Houston with a degree in Computer Engineering Technology.