Fake/Doctored image identification using Deep Learning

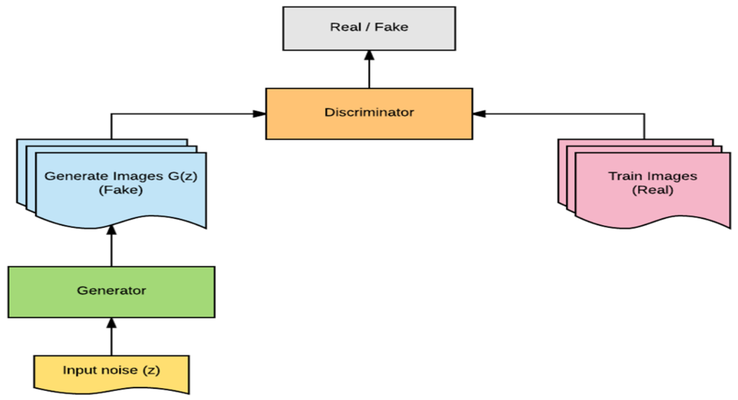
UNIVERSITY of STIRLING

Fidelis Henry Kamunde
MSc in Big Data

Introduction

Over many years, Images and videos has been considered as a proof of existence of something or a memory of actions that has taken place. But after the emergence of technologies that can manipulate digital media, should we still believe what we see? Technologies like Adobe Photoshop can be used to create fake images or videos. In recent years, AI has learned to create synthetic images/videos called deepfake using a deep learning technique. These doctored images/videos are the main sources of fake news and are also used in a malicious way to cause trouble in a society

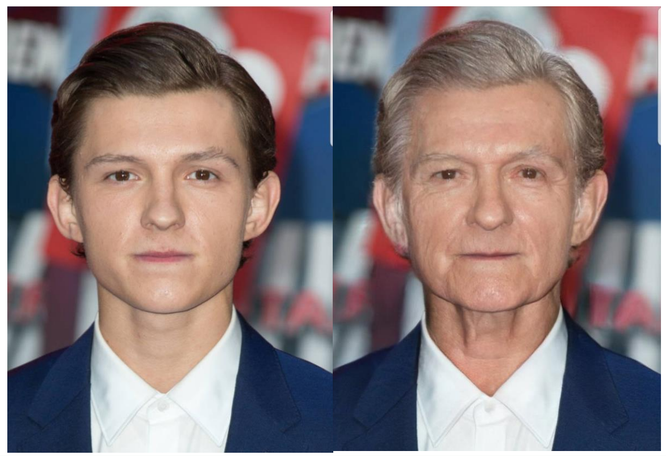
My project will focus on deepfake detection

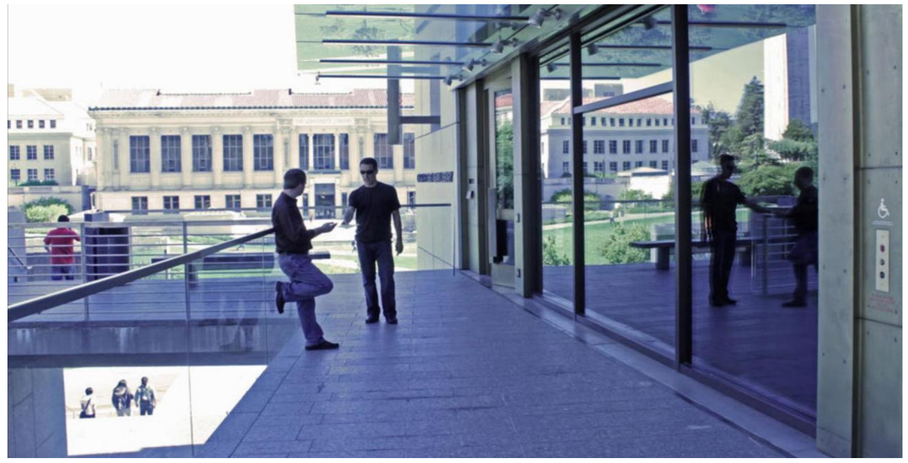


How fake images are generated

Classical methods for creating fake images involves copy/pasting and slicing.

Deepfake is generated by using a deep learning technique known as generative adversarial network (GAN). GAN is an algorithm consisting of two neural networks that acts as two players competing against each other, these networks are trained in adversarial manner.





Current methods to detect fake images

Feature-based — where there is a kind of artifact in a certain (or multiple) types of forgeries — mostly applicable for classic methods.

Supervised learning — using deep learning classifiers (mostly CNN) to learn certain types of fake images applicable for classical methods as well as generative models, e.g GANs.

Unsupervised learning — an attempt to capture some essence of a genuine image, to detect new kinds of forgeries (that the model hasn't seen before). It can be seen as a kind of anomaly detection.

For further information contact: *Fidelis Henry Kamunde* *fhk00001@students.stir.ac.uk*