Deep Fake Creation by Deep Learning

Harshal Vyas

Abstract: Machine Learning techniques are escalating technology's sophistication. Deep Learning is also known as Deep Structured Learning which is a part of the broader family of Machine Learning. Deep Learning architectures such as Deep Neural Networks, Deep Belief Networks, Convolutional Neural Networks(CNN) have been used for computer visions and for language processing. One of those deep learning-powered applications recently emerged is "deep fake".

Deep fake is a technique or technology where one can create fake images and videos which are difficult for humans to detect. This paper deals with how deep fakes are created and what kind of algorithms are used in it. This will help people to get to know about the deep fakes that are being generated on a daily basis, a way for them to know what is real and what isn't.

1. Introduction:

Deep fake (a portmanteau of "deep learning" and "fake") is a technique for human image synthesis based on artificial intelligence. Deep learning models such as Autoencoders and Generative Adversarial Networks have been applied widely in the computer vision domain to solve various problems. All these models have been used by deepfake algorithms to examine facial expressions and moments of an individual person and later synthesize facial images of another person making comparable expressions and moments. Generally, deepfake algorithms require a large amount of image and video data set to train models. Deepfakes were used to swap faces of Celebrities or Politicians with porn.

The first Deepfake video emerged in 2017. It caused popularity when one Reddit user named "Deep Fake" shared a deep fake created by him. An example of a deep fake is where people witnessed Barack Obama talking trash about Donald Trump.

Deepfakes are also used in a positive way to generate the voice of people who lost their own voice. It is also a threat to people in many ways. A release of the software called **DeepNudes** which can transform a person into non-consensual porn. Likewise, an app called Snapchat also started a Face swap feature where two different persons can swap their faces. To overcome the face swap technique the United States Defense Advanced Research Project Agency(DARPA) initiated a research scheme in media forensics to automatically detect deepfake. The only way to solve problems created by Deepfake is to use technology to detect fake videos or improve media literacy. The technical solution is to try and detect deepfake by using the same kind of AI that is used to create them.

2. Creation of Deepfake:

Photoshop and After Effects are utilized each day by experts, yet that doesn't imply that simply introducing both of them is everything necessary to make photorealistic pictures and recordings. In like manner, making reasonable face-traded recordings is hard. Like any imaginative undertaking, the conclusive outcome is a blend of ability, duty, and right tools. The primary endeavor of deepfake creation was FakeApp, created by a Reddit client utilizing autoencoder-decoder blending structure. In that strategy, the autoencoder extricates inert highlights of face pictures and the decoder is utilized to remake the face pictures. To trade faces between source pictures what's more, target pictures, there is a need of two encoder-decoder sets where each pair is utilized to prepare on a picture set, and the encoder's parameters are shared between two system sets. In simple words, two pairs have the same encoder network. The FakeApp software uses the AI Framework TensorFlow of Google, which in addition to other things was at that point utilized for the program DeepDream. There are additionally open-source options in contrast to the first FakeApp program, as DeepFaceLab, FaceSwap (right now facilitated on GitHub), and FakeApp (as of now facilitated on Bitbucket). Regardless of whichever application we use to create a deepfake the process involves mainly three steps.

- Extraction
- Training
- Creation

2.1 Extraction:

The deep- in deepfake comes from deep learning and as we know that deep learning requires large data sets. Thousands of different pictures are required to create a deepfake video. The extraction process refers to the process of extracting all frames, identifying the face, and aligning them. The alignment is a critical process, the neural network is performed to swap and all the face should have the same size.

2.2 Training:

Training is a specialized term acquired from Machine Learning. For this situation, it alludes to the procedure which permits a neural system to change over a face into another. In spite of the fact that it takes a few hours, the preparation stage should be done just a single time. When finished, it can change over a face from individual A to individual B.

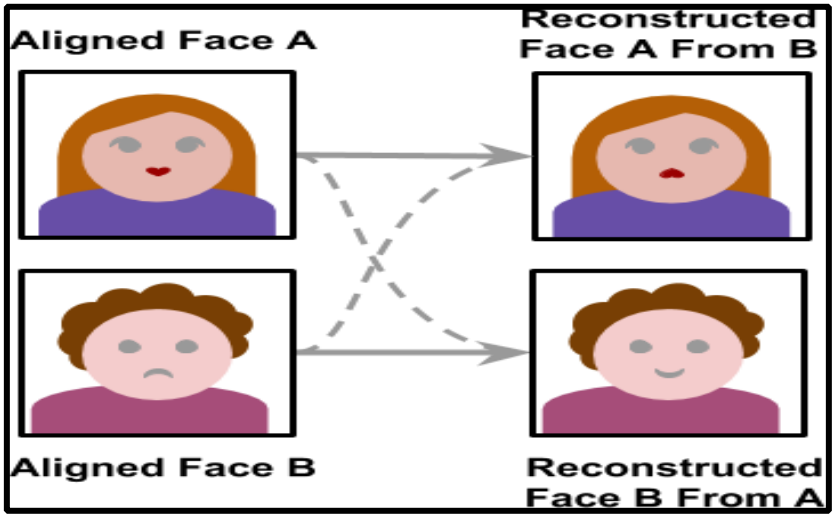

Fig.1: Shows the training of an image from Face A to Face B

2.3 Creation:

When the training is finished, it is at last time to make a deepfake. Beginning from a video or an image, all casings are removed and all appearances are adjusted. At that point, everyone is changed over-utilizing the prepared neural system. The last advance is to consolidate the change over the face once again into the first casing. While this seems like a simple errand, it is really where most face-trade applications turn out badly. As already been told that autoencoders are used to create a deepfake.

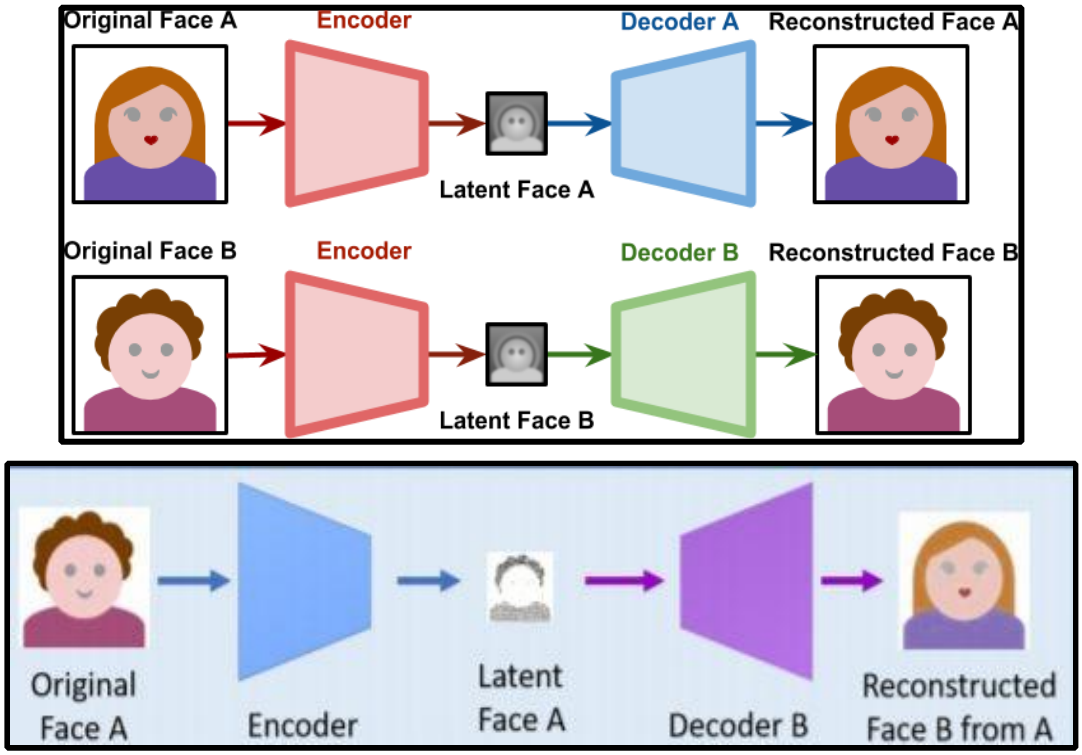

Fig.2: A deep fake creation model using two encoder-decoder pairs. Two networks use the same encoder but different decoder for the training process(top). An image of face A is encoded and decoded by decoder B to create a deep fake (bottom).

In the same way, the cross-section of the face can be done by passing the latent face A to the Decoder B and latent face B to Decoder A.

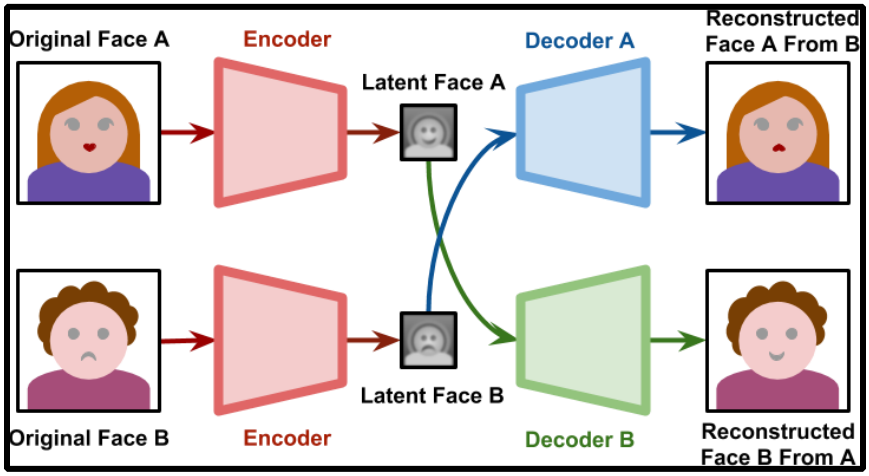

Fig.3: Cross Section of a Face where Latent Face A is passed to Decoder B and Latent Face B to Decoder A

If a network has generalized that is well enough to make a face, the latent space will represent facial expressions and orientations. This means that generating a face for Subject B by using the same expressions and orientations of subject A. What is significant is that the two subjects utilized in the preparation share whatever number likenesses as could reasonably be expected. This is to guarantee that the mutual encoder can sum up to significant highlights that are anything but difficult to move.

3. Deep fake detection Algorithm:

Some researchers have demonstrated a new algorithm for detecting deepfake images, those modified vaguely by AI frameworks, possibly for repulsive purposes. Starting trial of the calculation selected fake from undoctored pictures down to the individual pixel level with somewhere in the range of 71 and 95 percent precision, contingent upon the example informational collection utilized.

The calculation has not yet been extended to incorporate the discovery of deepfake recordings. Neural networks have been trained by showing them both deepfake and genuine images, to learn pixels present in the images. Later when we show a deep fake image the algorithm compares it pixel by pixel and with higher-level encoding analysis. At the point when these equal investigations trigger warnings over a similar area of a picture, it is then labeled as a potential deepfake.

A Deepfake Detection challenge(DFDC) was started by top leading companies like Facebook, Microsoft, Amazon, and Partnership of AI. It invites people around the world to build their innovative new technologies that can help detect deepfakes. All the large data sets are available on Kaggle and can be used to create deep face detection algorithms.

Concerns, Responses, and Conclusion:

Deepfakes have started to dissolve the trust of individuals in media substance as observing them is not, at this point proportionate with putting stock in them. They could make the pain and negative impacts those focused on, increase disinformation and abhor discourse, and even could animate political strain, excite general society, savagery, or war. This is particularly basic these days as the advancements for making deepfakes are progressively agreeable furthermore, online networking stages can spread those phony substances rapidly. Sometimes deepfakes do not need to be spread to the massive audiences to cause detrimental effects. People who create deepfakes with a malicious purpose only need to deliver them to target audiences as part of their sabotage strategy without using social media. People need to be more careful about what they see and they should have the ability to understand whether the content is fake or real.

Recordings and photographs have been broadly utilized as confirmations in police examination and equity cases. They might be presented as confirmations in an official courtroom by computerized media criminology specialists who have a foundation in PC or law implementation and involvement with gathering, looking at, and breaking down advanced data. The improvement of ML and AI advances may have been utilized to adjust this computerized substance and in this manner, the specialists' suppositions may not be sufficient to confirm these confirmations in light of the fact that even specialists can't perceive controlled substances.

References:

- https://en.wikipedia.org/wiki/Deepfak e
- Schwartz, Oscar (12 November 2018). "You thought fake news was bad? Deep fakes are where the truth goes to die". The Guardian. Retrieved 14 November 2018.
- "Experts fear face-swapping tech could start an international showdown". The Outline. Retrieved 28 February 2018.
- Roose, Kevin (4 March 2018). "Here Come the Fake Videos, Too". The New York Times. ISSN 0362-4331. Retrieved 24 March 2018.
- Bregler, Christoph; Covell, Michele; Slaney, Malcolm (1997). "Video Rewrite: Driving Visual Speech with Audio". Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques. 24: 353–360 – via ACM Digital Library
- "It took us less than 30 seconds to find banned 'deep fake' AI smut on the internet". Retrieved 20 February 2018.