Deep Fake Analysis in Compressed Videos

Muhammad Zubair , Saqib Hakak
Faculty of Computer Science , UNB FREDERICTON

Canadian Institute for Cybersecurity

UNB

Introduction

The term "deepfake" comes from the underlying technology "deep learning," which is a form of AI. In terms of digital media, a convincing image or video of someone or something that has been altered to distort or misrepresent someone's actions or words by using deep learning. Following are some face types of deepfake

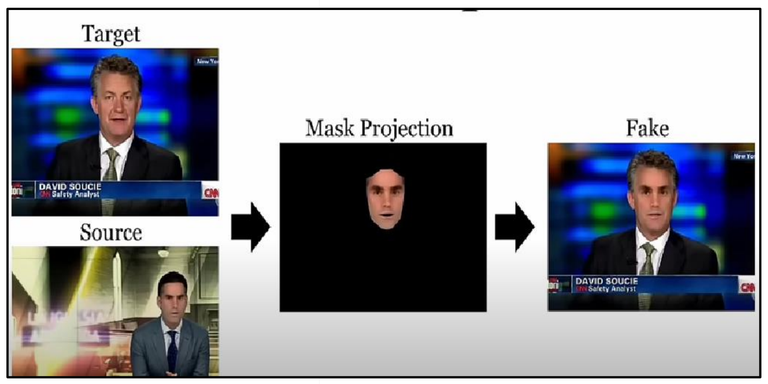

Figure 1. Face Swap

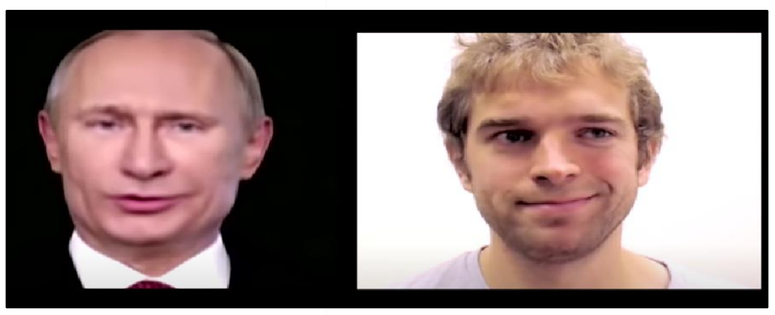

Figure 2. Puppet Master

Application:

- ❖ Positive Impact
  - ❑ Save Cost and Time by using in Entertainment Industry
  - ❑ Most Personalize Content for advertisement
- ❖ Negative Impact
  - ❑ Social Engineering
  - i. Propaganda
  - ii. Adult Industry [3]

Famous Data Sets:

- ❖ DFDC
- ❖ CELEB—DF
- ❖ Face Forensic ++

Problem Statement

Social media is usually used to spread deepfake. In social media platforms videos are compressed before publishing it to network by different compression factors. Current detection algorithms are showing great performance on High Quality Videos(HQ), but they lack good accuracy in detecting deepfake in compressed video.

Figure 3 . Before Compression

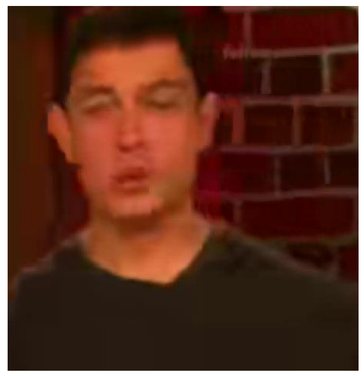

Figure 4 . After Compression

Video Quality Analysis Matrix

There are several matrix available to detect quality of video before and after compression. This experiment is done on different number of video of Celeb-DF V2 data set.[2]

- ❖ PSNR (Peak Signal-to-Noise Ratio) is a way of measuring how much a video or image has been compressed or degraded compared to the original. Higher PSNR indicates better quality.
- ❖ SSIM (Structural Similarity Index) is a measure of how similar two images or videos are in terms of their overall structure, brightness, contrast, and texture. assigning a score between 0 (no similarity) and 1 (perfect similarity).
- ❖ VMAF (Video Multimethod Assessment Fusion) is a method of evaluating the quality of a video by combining multiple objective metrics that aim to replicate how the human eye perceives video quality. VMAF uses machine learning algorithms to predict how a human viewer would rate the video quality based on various factors such as color accuracy, sharpness, and motion smoothness. The resulting score ranges from 0 (worst) to 100 (best), with higher scores indicating better quality.

# of Videos	PSNR	SSIM	VMAF
10	40.58	14.05	66.85
50	40.17	13.87	67.29
100	40.19	13.92	67.19

Table 1. Experiment on CELEB DF V2 Videos

Deep Fake Creation Mechanism

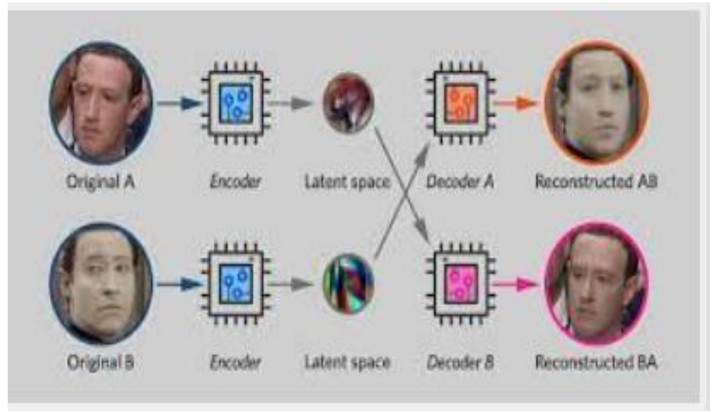

Figure 5. Using Autoencoder and Decoders

Literature Analysis

According to our literature review researchers are adopting different approaches to address this problem.

Some research are using deep learning to solve this problem, in which they are using frame level analysis and temporal level analysis[2] and some researchers are using handcrafted features .[5]

References

1. Xia, Z. *et al.* (2022) "Towards deepfake video forensics based on facial *Sciences*, 607, ptextural disparities in multi-color channels," *Information* p. 654–669. Available at: https://doi.org/10.1016/j.ins.2022.06.003.
2. Hu, J. *et al.* (2022) "Detecting compressed deepfake videos in social networks using frame-temporality two-stream convolutional network," *IEEE Transactions on Circuits and Systems for Video Technology*, 32(3), pp. 1089–1102. Available at: https://doi.org/10.1109/tcsvt.2021.3074259.
3. Kingra, S., Aggarwal, N. and Kaur, N. (2022) "LBPNet: Exploiting texture descriptor for Deepfake detection," *Forensic Science International: Digital Investigation*, 42-43, p. 301452. Available at: https://doi.org/10.1016/j.fsidi.2022.301452.
4. Wang, T. *et al.* (2022) "Deepfake noise investigation and detection," *Forensic Science International: Digital Investigation*, 42, p. 301395. Available at: https://doi.org/10.1016/j.fsidi.2022.301395.
5, Kaddar, B. *et al.* (2022) "On the effectiveness of handcrafted features for Deepfake video detection," *SSRN Electronic Journal* [Preprint]. Available at: https://doi.org/10.2139/ssrn.4264037.