Numerical Analysis and Optimization Application Trend of Deepfake Video Inspection Technology

Jiyu Xia and Man Hua*
School of Computer Science, Civil Aviation Flight University of China, Guanghan, Sichuan, China
*Corresponding author email: hua.man@163.com

Abstract. Science and technology can be a double-edged sword. Novel digital technologies not only allow people to enjoy more digital achievements, but also make it increasingly difficult to distinguish the authenticity of digital information. This review begins with a brief introduction of the deepfake technology, followed by an explanation of the possible hazards of deepfake, and finally, an overview of the detection technology and research status of deepfake.

Keywords: Deepfake; deep learning, generative adversarial networks, detection technology.

1. Introduction
The rapid development of the Internet has brought convenience to people's lives and at the same time and the frequency that people use the Internet are increasing day by day, our relevant data are becoming less private but easier for others to collect, and we have access to deepfake datasets through different ways. If these data are used by someone with an ulterior motive, deepfakes can and will be used for dark purposes.

Table 1 lists the common deepfake datasets.

| Datasets   | Link                                               |
|------------|----------------------------------------------------|
| UADFV      | https://github.com/danmohaha/WIFS2018_In_Ictu_Oculi |
| Celeb-DF   | https://github.com/danmohaha/celeb-deepfakeforensics |
| FF++       | https://github.com/ondyari/FaceForensics           |
| DF-TIMIT   | https://www.idiap.ch/dataset/deepfaketimit          |
| DFDC       | https://www.deepfakedetectionchallenge.ai/          |

In the recent years, the issue of fake news has gradually threatened the right of public discourse and even human society.[1] In terms of the individual, the misuse of deepfakes may impinge on citizen's reputation and privacy. For society, deepfakes may have an impact on social stability by creating fake news. As for the national level, the deepfake technology offers certain possibilities for bad guys, making it possible to use it in bad ways. For example, to tell lies, to incite violence, to discredit leaders and institutions, and even to subvert elections.[3] Therefore, to propose technologies that can effectively detect these fake videos is indispensable.

The structure of the article is as follows. After the introduction, in part II will explain what are deepfakes, because only we know deepfakes better that we are more capable to deal with problems it
brings. The potential threats of deepfakes will be discussed in part III. And part IV is mainly about Developments in deepfakes video detection technology. Finally, summarize the full text and look forward to the current trends in deepfakes video detection technology.

2. What Are Deepfakes?
Deepfake is a new word that combines the words “deep learning” and “forgery”, and this new technology, which was appeared with the free access to public databases and the development of deep learning techniques, uses a deep learning models called Generative Adversarial Networks (GAN) to create vivid fake images or videos. In late 2017, a Reddit user named “deepfakes” claimed online that he had developed a machine learning algorithm that could transpose celebrity faces into pornographic videos and this is how the word “deepfakes” originated[2]. The Reddit user has not only posted a video of the face swap, but has also posted the original source code of the algorithm on the Internet. And widespread social concern and re-tweets were drawn to the event.

![Figure 1. The structure of GAN. GAN is made up of a generator and a discriminator.](image)

The generator adds noise to the original image, while the discriminator is trained continuously to try to discriminate between the real and fake images in the post-processing training set.

Deepfakes uses a neural network to learn and imitate a person's facial expressions, behavior, and voice after analyzing a massive sample set of data. In other words, deepfakes uses the facial mapping technology and artificial intelligence to replace one person's face with another person's face in the video.

As for the technology deepfakes uses, GAN (Generative Adversarial Networks), is able to review up to a couple of thousand photos of a person and then generate a fresh look which is very similar to the previous one, rather than an exact copy of any one photo. Researchers suggest that as deep learning continues to evolve, it will be harder to detect deeply faked videos.

3. What the Threats of Deepfakes Technology Are

3.1. Security Concerns
On the one hand, deepfakes technology threatens personal privacy and security. With the generation of this technology, it is possible for everyone to wake up and find their face spreading across the Internet in some deepfakes video, not just their reputation will be severely damaged, but also their privacy can be violated at any time, which may ruin a person's work, life, and even their life.

On the other hand, deepfakes technology can also have serious implications on both national security and public safety. In 2018, 20 people were violently killed in India as a result of online rumours of child abduction or involvement in other crimes. Recently, researcher Watts pointed out that deepfakes technology makes it possible for false information to be presented to the public in a highly credible manner, thereby manipulating people’s emotions and arousing widespread distrust in society.

3.2. Trust Concerns
Nowadays, people are getting more and more influenced by spam generated by artificial intelligence, as well as fake news based on fake videos and numerous conspiracy theories. A phenomenon known as
"reality coldness" results from the frequent contact with misinformation that makes people feel that lots of the information they are exposed to is not credible, which makes it no longer easy for people to believe everything they see and hear. This also coincides with the "trust-decline effect" that Citron, a law professor at the University of Maryland, suggested at a hearing on deep falsification of artificial intelligence.[4] Government credibility is the foundation of social trust. Nonetheless, the most damaging aspect of deepfakes technology is the increased reduction in credibility. What’s more, the existence of the technology would adversely affect its probative value in court.[5] If the evidence given is no longer credible, can the criminals still be brought to justice?

4. Discussion on Ways to Detect Deepfake Videos
In response to the range of potential problems posed by the development of deepfake technology, to develop an effective method to detect deepfake videos becomes extremely important. In 2018, the U.S. Defense Advanced Research Projects Agency launched a media forensics research program to develop tools that can detect deepfake videos.[6] In this project, Professor Siwei Lu's team has successfully developed a very simple identification technique, i.e. face generated by the deepfake technique rarely blinks, but when it does, the eye movements are unnatural. Similar techniques are being captured by other teams in the project to identify deepfake videos.[7]

But the truth is that there are far more available research resources and people who are working on developing techniques to produce deepfake videos than there are people who are detecting these fake videos. In addition, deepfake developers often use the published results of deepfake research to improve their techniques and bypass new detection systems. For instance, these deepfake developers have already fixed this unnatural eye blinking.

It is no coincidence that researchers at the University of Southern California and the University of California, Berkeley, have collaborated to develop new AI tools that also identify deepfake videos by the specific details of the movements of each facial organ when a person speaks or makes an expression.[8] But the body has more variability and is more complex to deal with, and it is a quite difficult thing for the ordinary person to make the subtle differences in body parts. There are many more detection techniques like this. Sabir[9] et al. has proposed a recurrent convolutional model (RCM) to detect deepfakes videos using the time lag between frames.

![Figure 2](image-url)

**Figure 2.** The overall approach to manipulation detection can be divided into two steps: cropping and aligning the face from the video frame, and then performing manipulation detection on the pre-processed face area.

Güera[10] et al. use convolutional neural networks (CNN) to extract frame-level features which are then distributed to long-short-term memory (LSTM) networks to construct temporal sequence descriptors that are useful for classification.
Li\[11\] et al. used CNN to discover the artifact based on the property that the resolution of the facial region in the deepfakes video is not consistent with the surrounding environment.

Nguyen\[12\] et al. use a capsule network based on dynamic routing algorithms to classify fake and authentic videos.

Yang\[13\] et al. extracted facial marker location anomalies and inputted svm to detect deepfake video.

Matern\[14\] et al. have proposed a method for detecting visual features based on eyes, teeth, and facial features.

Koopman\[15\] et al. used local optical responses uneven pattern differences between real and fake videos to detect deepfake videos and achieved a better result.

But there is no single technological solution that can completely solve the problem. Detection of the authenticity of a video is not a complete way to identify the video, we can also go from the source of the problem to the originality of the video.

For example, Create digital watermarks or digital signatures systems throughout the entire network. In
this method, the basic idea is to integrate a watermark or digital signature into the audio/video capture device, so that even counterfeiting cannot alter the original video. But this is a partial idealistic approach. To create such a system across the network, two preconditions would first have to be met. First, the device has the ability to create watermarks or signatures. Second, the digital content posted to internet social platforms must be digitally watermarked or signed. Due to these two prerequisites, this approach is still in its infancy, as it is too limited in scope to be used.

And for the first condition, blockchain technology can be used. The blockchain is a distributed database system in which nodes participate. It is immutable and non-forgable, and can be understood as a ledger system. The blockchain can create a unique and unchangeable blockchain of metadata, as well as a decentralized way of allocating transactions between accounts. In other words, Blockchain technology can help in verifying the origins and distribution of videos by creating and storing digital signatures in a ledger that is almost impossible to manipulate. As for the second requirement, this is not just a technical issue, but also needs to be supported by relevant legal regulations, as the Internet is not a place outside the law.

5. Conclusion

Science and technology is a double-edged sword. In order to deal with a series of problems caused by deepfake false information, this study discusses the current research on the detection technology of deep false information, and discusses the possible harm caused by deep false information in terms of security issues and trust issues respectively, and discusses the corresponding solutions based on the problems in these two aspects. They are digital watermark and blockchain respectively. However, the detection of deep false information is not a simple problem, and no single detection technology can completely solve the problems caused by deep false information. Therefore, we should combine multiple methods to try to solve the problems caused by deep false videos, such as combining theoretical methods and regulations to manage the problem from theoretical and technical perspectives, or using the integration of multiple technologies to manage the problem from different perspectives.

Acknowledgments

I would like to express my sincere gratitude to my supervisor, Professor Hua, for his constant encouragement and guidance, without which this paper would not be possible.

References

[1] Qayyum, A., Qadir, J., Janjua, M. U. & Sher, F. 2019. Using Blockchain to Rein in the New Post-Truth World and Check the Spread of Fake News. IT Professional, 21(4):16–24.
[2] BBC Bitesize, “Deepfakes: What Are They and Why Would I Make One?” 2019. [Online].
[3] CHESNEY R, CITRON D. Deepfakes and the New Disinforma - tion War: The Coming Age of Post - truth Geopolitics [J]. Forei gn Aff., 2019, 98:147.
[4] TIAN Shaobo, TIAN Qian. Deepfake Technology: Ethical Challenges and Reflections[J]. Journal of Foshan Institute of Science and Technology (Social Science Edition), 2020, 38(04):15-21.
[5] MARAS Marie-Helen, ALEXANDROU Alex. Determining Authenticity of Video Evidence in the Age of Artificial Intelligence and in the Wake of Deepfake Videos [J]. The International Journal of Evidence & Proof; 2019, 23(3) : 255-262.
[6] Matt T. MediaForensics (MediFor) [EB/OL]. (2020-02-26) [2020-02-26].
[7] Feng, Weidong. The U.S. has developed a "deepfake" video forensics tool [N]. Science and Technology Daily, 2018-08-09.
[8] Li Pengfei. U.S. research institute develops AI tool to identify fake DeepFakes videos [J]. Technology China, 2019(7): 103-106.
[9] Sabir E, Cheng J, Jaiswal A, et al. Recurrent convolutional strategies for face manipulation detection in videos [J]. Interfaces (GUI), 2019, 3(1):80-87
[10] Güera D, Delp E J. Deepfake video detection using recurrent neural networks [C] //Proc of the 15th IEEE Int Conf on Advanced Video and Signal Based Surveillance(AVSS). Piscataway, NJ: IEEE, 2018:1-6
[11] Li Y, Lyu S. Exposing deepfake videos by detecting face warping artifacts [J]. arXiv preprint, arXiv: 1811.00656, 2018

[12] Nguyen H H, Yamagishi J, Echizen I. Capsule-forensics: Using capsule networks to detect forged images and videos [C] // Proc of 2019 IEEE Int Conf on Acoustics, Speech and Signal Processing (ICASSP). Piscataway, NJ: IEEE, 2019:23072311

[13] Yang X, Li Y, Lyu S. Exposing deep fakes using inconsistent head poses [C] //Proc of 2019 IEEE Int Conf. on Acoustics, Speech and Signal Processing (ICASSP).Piscataway, NJ: IEEE, 2019:8261-8265

[14] Matern F, Riess C, Stamminger M. Exploiting visual artifacts to expose deepfakes and face manipulations [C] //Proc of 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW). Piscataway, NJ: IEEE, 2019:83-92

[15] Koopman M, Rodriguez A M, Geradts Z. Detection of deepfake video manipulation [C] //Proc of the 20th Irish Machine Vision and Image Processing Conf (IMVIP). Belfast, Northern Ireland: IMVIP2018, 2018:133-136

[16] Westerlund, M. 2019. The Emergence of Deepfake Technology: A Review. Technology Innovation Management Review, 9(11): 40-53.