Jozac Academic Voice (JAV)
https://journals.jozacpublishers.com/jav
Vol 4, No 1, 2024

OPEN ACCESS

Deepfake digital face manipulation: A rapid literature review

Genesis Gregorious Genelza[1*]

[1]*College of Teacher Education & Junior High School Department/ University of Mindanao Tagum College, Philippines, genesis.genelza@umindanao.edu.ph*

*Corresponding author: genesis.genelza@umindanao.edu.ph

Abstract: "Deepfake" refers to the complex processing of audiovisual content using generative adversarial networks (GANs) and other deep learning methods. It enables the smooth superimposition of one person's face or voice over another, producing remarkably lifelike fake films or images. However, the online distribution of the modified images could result in major moral, ethical, and legal issues if misused. This rapid literature review aimed to establish a clear perspective regarding deepfake digital face manipulation. The paper emphasized the potential risks of this technology while also discussing the benefits of looking into deepfakes. The study contributes to the field's understanding. It addresses the abuse of deepfake technology by offering insights into the most current advancements in identifying and creating manipulated facial images. Strict ethical standards and rules regulating its application must be developed to maximize its potential. To maximize the benefits and minimize the risks of this technology, a multifaceted strategy that considers ethical concerns, legal frameworks, technological innovation, and public awareness is necessary. With the development of AI-driven detection systems, the adverse effects of modified content on both people and society can be addressed.

Keywords: Artificial Intelligence, Deepfake, Digital Face Manipulation, Face Cloning, Generative Adversarial Networks

1. Introduction

Due to developments in deep learning methods and the accessibility of ample, accessible databases, even people without technical backgrounds can now edit or create realistic facial samples for benign and malevolent reasons. Face audiovisual information that has been artificially generated or digitally modified using deep neural networks is referred to as "DeepFakes." Even with significant advancements in artificial intelligence, physics, and traditional and sophisticated computer vision, there is still a massive arms race between attackers, offenders, and adversaries (for example, ways for creating DeepFakes) and defenders (for example, methods for detecting DeepFakes). DeepFakes, or artificial intelligence-generated or digitally altered face samples, threaten the accuracy of face recognition software and online data security. Though there has been some progress, many problems still need to be fixed before highly efficient and universal generation and defensive procedures can be achieved. Reliable deepfake and face manipulation detection frameworks still need to be revised in deepfakes, necessitating multidisciplinary research efforts across various fields, including machine learning, computer vision, human vision, and psychophysiology (Akhtar, 2023).

Furthermore, the creation of hyper-realistic facial photographs, which are hard to identify and may quickly reach millions of people, has been transformed by the release of large-scale datasets that are publicly available and developments in generative adversarial networks (GANs). These developments have adverse effects on society. The altered face image recognition and production field is still in its infancy and research phases. Multimedia material in cyberspace has increased significantly due to the prevalence of affordable and sophisticated mobile devices like digital cameras, cell phones, and portable PCs. These multimedia data come in various formats, such as audio, video, and image. This trend is driven by social media's dynamic and ever-changing landscape, making it the perfect platform for people to swiftly and easily share their multimedia material with the world and contribute to the exponential expansion of such content. With this,

such innovative methods are necessary to maintain the security and reliability of digital media in the future and to remain ahead of the rapidly changing field of face alteration techniques (Dang & Nguyen, 2023).

Following the introduction of social networking services (SNSs), there has been a noticeable rise in the need to manipulate multimedia content, like images on Instagram or videos on TikTok, to draw in more viewers. Before now, ordinary users found multimedia data manipulation to be extremely difficult. This was primarily because of the obstacles professional graphics editor programs such as Adobe and the GNU Image Manipulation Program (GIMP) created, in addition to the lengthy editing process itself. However, technological developments have simplified multimedia data manipulation and produced more realistic results. Deep learning (DL) technology has advanced significantly, bringing with it complex designs like generative adversarial networks (GANs) (Radford et al., 2015). Hence, the research field has made significant efforts to introduce novel algorithms that may reliably and rapidly detect signals of manipulated multimedia data in response to the growing threat of increasingly realistic and advanced edited facial photographs. Hence, these days, everyone agrees on the importance of education as a tool or act of social innovation and progress. In addition to learning new information and abilities, we must be trained to be mature citizens and responsible individuals. It takes a big commitment for people to make every individual aware of the everyday situation in society (Genelza, 2022).

With this, this article aimed to establish a clear perspective regarding deepfake digital face manipulation following a rapid literature review as methodology. A different approach to a systematic literature review (SLR) that can expedite the examination of recently released data is a rapid literature review (RLR) (Chukwuere, 2023). The project aimed to find and compile information about various definitions of RLR and the methods used to carry out these assessments (Smela et al., 2023). Rapid reviews are a type of knowledge synthesis where information is produced quickly by simplifying or removing steps from the systematic review process (Tricco et al., 2015). The study also discussed the advantages of examining deepfakes while underlining the possible risks associated with this technology. Several research directions are suggested for future studies to successfully address the issues currently facing the discipline. The study advances knowledge and fights the misuse of deepfake technology by providing insights into the state of the art for manipulated facial picture identification and production.

2. Background:

What is Deepfake Digital Face Manipulation?

The term "deepfake" describes the intricate manipulation of audiovisual material through generative adversarial networks (GANs) and other deep learning techniques. It allows the seamless superimposition of one person's face or voice onto another, creating phony movies or photographs that are incredibly realistic. The possibility of misleading viewers, invasions of privacy, and false information are all serious issues brought up by these altered media. Deepfakes have spurred debates concerning the moral ramifications of manipulating digital media and the difficulties in identifying fake information in the digital era. Creating sophisticated detection techniques and increasing public awareness of the prevalence of manipulated media are two steps taken in the fight against deepfake technology (Dang & Nguyen, 2023).

After the debut of the avatarify program, a model of facial cloning prompted the creation of deep fake digital face animation, a virtual animation creative that uses the First Order Motion Method and styleGAN to toonify the photos and forgive the fake facial movements. Here, the design was made to motivate people and draw more attention. Here, we will initially insert an image serving as our source file. The cartoon dataset we have gathered will match all facial vital points and choose the image that best fits the detected vital point. These two will then be combined, and the model will cut the image to resemble the cartoon character before producing the cartoon image. Following execution, the cartooned image will function as the source file, and we must insert a driver file—a video file—with the image. Once the image and the video have been detected, the source and driver files are inserted, and an essential point detection process is carried out. The intended result is for the source file to function as a driver file, allowing the deep animator to create a cartoonized face animation clone and manipulation (Nithin & Bargavi, 2021).

One helpful method that can be used to modify photographs is image tampering. Three standard methods of manipulating images are copy-move, image splicing, and image retouching. While the image splicing approach joins two or more images to form a composite image, the copy-move method involves copying specific portions of a source image and inserting them into a target image. However, the picture

retouching method uses several computer vision (CV) technologies to enhance some aspects of the original image and produce a new one. Following the application of these techniques, improvements are made to the image's boundaries, form, scaling, and illumination to reduce faults and increase the difficulty of identifying the tampered portions. Generative adversarial networks (GANs), a new area of artificial intelligence (AI) that focuses on unsupervised learning, have recently been popular, adding to the task's difficulty. It can produce images that have realistic features and appear to human viewers to be somewhat authentic (Carvalho et al., 2015; Antipov et al., 2017; Karras et al., 2017).

Image forging, which involves retouching, merging two photos into one, or shifting a portion of one image into another, is a popular topic in digital image manipulation. Furthermore, it is now more difficult for humans to identify the manipulated one due to recent advancements in generative adversarial networks (GANs), which are used to create human face images. If the altered photos are misused, the online dissemination of those pictures may give rise to serious moral, ethical, and legal problems. Because of this, a great deal of research has been done in the last few years to identify facial image manipulation using machine learning algorithms applied to manipulated face datasets (Dang et al., 2020).

3. Discussions

In the present, according to Dang and Nguyen (2023), an upsurge of face manipulation apps, such as FaceApp and FaceSwap, have appeared on the scene. To exacerbate the situation, the release in June 2019 of the intelligent undressing app Deepnude sent shockwaves through the global community. Regular users now find it more challenging to weed out altered content because multimedia content may spread like wildfire online, with dire implications such as defamation, election manipulation, and scenarios that incite conflict. Furthermore, things have gotten worse due to the recent emergence of strong, sophisticated, and easy-to-use mobile manipulation programs like FaceApp, Snapchat, and FaceSwap, which make it much harder to authenticate and confirm the accuracy of photos and videos.

In addition, digital forensics, scientific research, health, media, and many more professions depend heavily on digital images. Digital image usage and social media sharing are commonplace these days. Digital photos are regarded as one of the primary information sources. Given the widespread usage of picture sharing on social media sites like Reddit, WhatsApp, Instagram, and Telegram, it can be difficult to distinguish between authentic and fake photographs. The proliferation of image manipulation tools is making it harder and harder to tell an authentic image from a fake one. The modern era's technological breakthroughs in all fields contribute to data misuse. Therefore, finding these modified data types and differentiating the actual data from the manipulated ones presents a complex problem for researchers. One of the most popular methods for manipulating digital images is splicing, which involves pasting a chosen portion copied from the same or a different image into another image. Digital image authenticity can be reliably confirmed by image forgery detection (Qazi et al., 2022).

Yerushalmy et al. (2011) proposed an innovative method for identifying image forgeries. This method does not compare the photographs for training and testing or add digital watermarking to the images. The authors suggested that the image's characteristics taken out at the acquisition stage serve as independent evidence of the image's legitimacy. The unaided eye may frequently see these characteristics. To be more precise, it employs irregularity-induced artifacts as identifiers to assess the veracity of an image. A color filter array method for identifying picture manipulation was presented by Dirik and Memon (2009). It computes a threshold-based classifier and a single feature. The authors tested their methodology using natural, manipulated, and computer-generated photographs. The examination of the experiment revealed minimal error rates.

With all that said and done, it has also made exemplary contributions to humanity. For example, using artificial intelligence (AI) to produce lifelike fake films or pictures, or "deepfake" digital face modification, has various benefits. Its potential for the entertainment sector is one of its main advantages; it can let filmmakers recreate the lives of deceased actors or create flawless special effects without the need for costly CGI or makeup, although this benefit may also lead to severe issues and concerns. Furthermore, deepfake technology can be used in research and development to create lifelike training simulations, including medical simulations or disaster response drills, which can be helpful to humanity these days. It can also help enhance AI by offering a problematic standard for identifying corrupted information, pushing the frontiers of AI research.

In contrast, there are also severe disadvantages to deepfake face manipulation. Misuse of this technology can create political propaganda or fake revenge pornographic content, which can harm people's reputations and propagate false information. This poses a severe threat to privacy and security. Deepfake technology is getting more sophisticated, making it harder to distinguish modified content from actual content, adding to worries about identity theft and other fraudulent activities. When someone creates a phony film or image without getting consent, ethical questions come up because there could be social and legal fallout. From an educational perspective, Genelza (2022) mentioned that teachers and students take on specific identities and roles that help them grasp what makes up the learning process and the subject matter that needs to be learned, especially in distinguishing what is real and not in integrating technological advancement in the classroom.

While deepfake digital face manipulation offers significant benefits for study in education and entertainment, its misuse raises ethical and security problems that require strict restrictions and excellent education about its hazards and potential. Utilizing innovation while ensuring responsible use and protecting against possible danger is essential to maximizing benefits and minimizing drawbacks from this rapidly developing technology.

4. Recommendation

Deepfake digital face manipulation is a potent tool that needs to be used with caution and morality. It is imperative to develop stringent ethical norms and regulations governing its application to realize its potential fully. Both governments and organizations should work together to create laws that specify permitted applications to prevent deepfake technology from being used for nefarious activities such as fraud, deception, or defamation. Furthermore, research and development on deepfake detection techniques is essential to counteract the detrimental effects of its exploitation. The damaging impact of altered content on people and society can be minimized with advances in AI-driven detection systems.

Managing deepfake technology is greatly aided by education and awareness campaigns. Public awareness campaigns regarding the existence and dangers of deepfakes can assist people in developing greater media literacy. Resources for teaching critical thinking can be found in schools, media literacy initiatives, and internet platforms. These tools help people see possible manipulations and take action to confirm the legitimacy of content. It is essential to stress that modern online learning environments are the only tools to help us, not a substitute for the face-to-face or virtual teaching methods we employ in the classroom (Genelza, 2023). Additionally, deepfake misuse can be lessened, and people can be empowered to traverse the digital landscape safely by encouraging digital literacy among users and emphasizing responsible online conduct. Technology companies, researchers, legislators, and advocacy groups must work together to address the complex issues raised by deepfake digital face manipulation. Developing an interdisciplinary strategy that considers ethical issues, regulatory frameworks, technological innovation, and public awareness is essential to maximizing the advantages of this technology while reducing its risks. Unless we work together, we cannot guarantee that deepfake manipulation is applied morally and sensibly in our increasingly digital society.

ORCID

Genesis Gregorious Genelza https://orcid.org/0000-0001-5577-7480

References

1. Akhtar, Z. (2023). Deepfakes Generation and Detection: A Short Survey. *Journal of Imaging*, 9(1), 18.
2. Antipov, G., Baccouche, M., & Dugelay, J. L. (2017). Face aging with conditional generative adversarial networks. *In 2017 IEEE international conference on image processing (ICIP)* (pp. 2089-2093). IEEE.
3. Carvalho, T., Faria, F. A., Pedrini, H., Torres, R. D. S., & Rocha, A. (2015). Illuminant-based transformed spaces for image forensics. *IEEE transactions on information forensics and security*, 11(4), 720-733.
4. Chukwuere, J. E. (2023). Exploring literature review methodologies in information systems research: a comparative study.
5. Dang, M., & Nguyen, T. N. (2023). Digital Face Manipulation Creation and Detection: A Systematic Review. *Electronics*, 12(16), 3407.

6. Dang, L. M., Min, K., Lee, S., Han, D., & Moon, H. (2020). Tampered and computer-generated face images identification based on deep learning. *Applied Sciences*, 10(2), 505.
7. Dirik, A. E., & Memon, N. (2009). Image tamper detection based on demosaicing artifacts. *In 2009 16th IEEE International Conference on Image Processing (ICIP)* (pp. 1497-1500). IEEE.
8. Genelza, G. G. (2022). Higher education's outcomes-based education: Bane or boon?. *West African Journal of Educational Sciences and Practice*, 1(1), 34-41.
9. Genelza, G. G. (2022). Internship program and skills development of fourth year bachelor of secondary education major in English. Galaxy International Interdisciplinary Research Journal, 10(2), 496-507.
10. Genelza, G. G. (2023). Quipper utilization and its effectiveness as a learning management system and academic performance among BSED English students in the new normal. *Journal of Emerging Technologies,* 3(2), 75-82.
11. Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2017). Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv*:1710.10196.
12. Nithin Y. & Bargavi, M. (2021). Deep Fake Face Animation Cloning. *International Journal for Scientific Research & Development*, 9(3), 277-280.
13. Qazi, E. U. H., Zia, T., & Almorjan, A. (2022). Deep learning-based digital image forgery detection system. *Applied Sciences,* 12(6), 2851.
14. Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*.
15. Smela, B., Toumi, M., Świerk, K., Francois, C., Biernikiewicz, M., Clay, E., & Boyer, L. (2023). Rapid literature review: definition and methodology. *Journal of Market Access & Health Policy,* 11(1), 2241234.
16. Tricco, A. C., Antony, J., Zarin, W., Strifler, L., Ghassemi, M., Ivory, J., ... & Straus, S. E. (2015). A scoping review of rapid review methods. *BMC Medicine*, 13(1), 1-15.
17. Yerushalmy, I., & Hel-Or, H. (2011). Digital image forgery detection based on lens and sensor aberration. *International journal of computer vision*, 92, 71-91.