International Journal of Emerging Trends in Computer Science and Information Technology
ISSN: 3050-9246 | https://doi.org/10.63282/3050-9246.IJETCSIT-V2I1P104
Eureka Vision Publication | Volume 2, Issue 1, 28-35, 2021

Original Article

The Convergence of Deep Learning and DeepFake: A Study on AI-Generated Media Manipulation

Sundar Tiwari[1], Writuraj Sarma[2], Saswata Dey[3]
[1,2,3]Independent Researcher USA.

Abstract - Through artificial intelligence and machine learning, especially deep learning, there has been a great change in the future of digital media. The key advancement that has attracted quite a significant level of controversy and consequence in duality is known as DeepFake technology. DeepFakes are generated by generative models, including GANs and autoencoders, to create realistic content that mimics real-life persons. The following paper investigates deep learning in relation to DeepFake technologies in terms of their background, development, usages, and impacts before February 2021. This paper is concerned with an introduction to the type of algorithms employed in DeepFakes, the entertainment, politics, and cyber-security fields in which DeepFakes have found application, as well as the counter-measures that have been put in place to address the problems associated with AI-generated media forgery. This paper makes a methodological contribution to the emerging research on DeepFakes by conducting empirical investigations, synthesizing from literature, and assessing models of DeepFakes to inform a given environment. However, the decision on ethical issues, regulatory requirements, and possible further research relating to AI media synthesis and detection is also discussed at the end of the paper.

Keywords - DeepFake, Deep Learning, Generative Adversarial Networks (GANs), Media Manipulation, Autoencoders, Face Swapping, AI Ethics.

1. Introduction

The last decade has seen an exceptional rise in the accomplishment of deep learning discipline and in areas such as computer vision, natural language processing, and generative modelling. [1-4] Among the most alluring and alarming phenomena that stemmed from that, there is DeepFake technology. These are fake media in which a person in a given picture or clip is swapped with someone else's image. It may be recalled that DeepFakes was initially conceived as an instance of the generative network, but it has become a socio-technological phenomenon today.

1.1 Needs of Convergence of Deep Learning and DeepFake

Deep Learning and DeepFake technologies should be combined because the development of synthetic media generation methods can be greatly expanded. The conditions for understanding, control, and safe and reasonable application should be created in regulating these technologies. Therefore, Deep learning techniques are essential for designing and developing DeepFakes and for managing and combating the problems arising from DeepFakes.

The interactions between Accountable Care Organizations (ACOs) and value-based purchasing programs include the following areas that needed to merge:

- **Improved DeepFake Detection Capabilities:** New methods of generating DeepFake have developed rapidly over the years, and the rate at which these fakes are being produced has overshadowed the creation of means for detecting the fakes. Although, deep learning detection methods are very important when distinguishing synthetic contents, particularly in object detection in both image formats, the CNNs and RNNs. Newer DeepFake becomes more real as they apply newer algorithms to sense the movement and texture and other discrepancies related to light and face. Integrating deep learning with DeepFake technology means that the detectors being created will be more improved, and they will be able to detect more complex manipulations and higher quality synthetic media.
- **Real-Time Detection and Analysis:** Real-time detection of DeepFake data is a significant problem as the data processing involves high-resolution videos. The specific applications of deep learning include model pruning, quantization, and federated learning for real-time application of detection models. Such innovations make it possible to significantly decrease the amount of calculations, but at the same time, enabling the detection systems to work on large datasets, and in fast unfolding systems, like social networks and video conferencing. That is why, by integrating deep learning with DeepFake detection, it is possible to automatically mark potentially altered content as they are being shared or replayed.
- **Enhancing DeepFake Generation for Creative Uses:** There are many positive or evident benefits in utilising synthesized media, including in the cinema, computer games, and virtual reality games. This combination of deep learning with DeepFake makes the generation of realistic avatars of ancient personalities, computer graphics, and even historical simulations with 3D characters possible. Thus, DeepFakes generated using generative models such as

GANs and VAEs deep learning can amplify the aesthetic value while pursuing a form of art and education but without crossing the ethical standards of consent and reckoning.

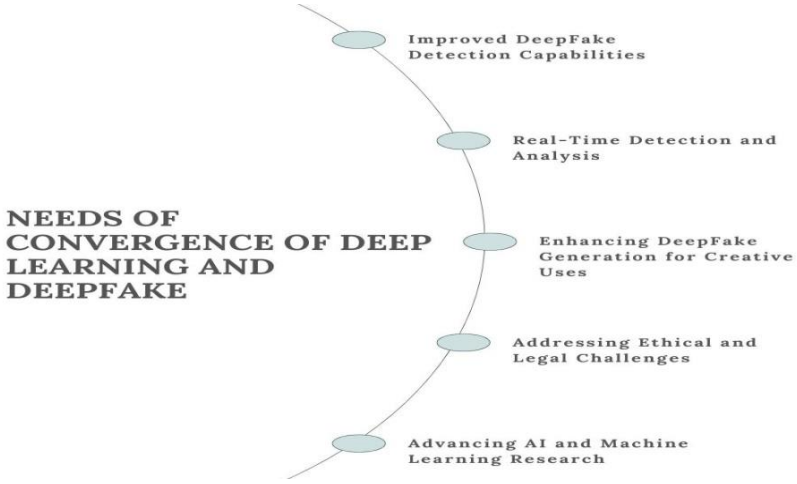

Figure 1. Needs of Convergence of Deep Learning and DeepFake

- **Addressing Ethical and Legal Challenges:** Organized crime also has many possibilities for using DeepFakes: identity theft, falsification of information, and harassment. However, when deep learning methods can be combined with regulatory mechanisms, it is easy to follow the presence of DeepFakes. Integrating machine learning and legal-ethical paradigms allows building algorithms that fight fake content and may identify the materials violating the rights and the law, contributing to digital rights protection. Furthermore, deep learning is useful for developing watermarking and origin tracking that enables one to track the owners of the initial content, which will help to minimize the distribution of unethical content.
- **Advancing AI and Machine Learning Research:** The prospects of applying deep learning with deepfake technology make it possible to progressively develop AI in machine learning. With the development of technology, new approaches and structures of deep learning, methods of training, and model improvement will appear. The need to detect and create DeepFakes inspires developments in some categories, such as adversarial, unsupervised, and semi-supervised training. Such merging aids enhance the overall advancement in fields other than DeepFake identification and synthesis, like face recognition, emotional identification, and real-time processing of video streams.

1.2 AI-Generated Media Manipulation

AI-aided media manipulation means using artificial intelligence application technologies to forge or modify any media in a way where it becomes challenging to find the difference between the forged one and the original one. [5,6] These manipulation methods include DeepFakes that use GANs and autoencoders to create realistic images, videos and audio to create a false impression. This continuous improvement in features of AI means that object images can be altered as one would wish, such as changing faces, expressing images, or even voices. This makes it extremely hard for the average onlooker to differentiate between fake and real. Conveniently, media can be altered in that way because this opens up creative potential and, at the same time, means that we are in danger. However, media manipulation generated by AI has numerous benefits in entertainment, advertizing, and comprising recreation, virtual appearance, and historical figures and events in the form of avatars.

But AI-generated manipulation also has disadvantages in spreading fake news, fraud, and invasion of people's privacy. For instance, DeepFakes generate powerful tools for political manipulation and the creation of adult content without the subject's consent, which is why they provoke ethical and legal issues. Since the advancement in technology, it has become easier to create deceptive content, which succeed in making it difficult to confirm the originals. This situation necessitated the development of better detection methods to detect AI generated content to ensure user authenticity. If there are no control measures and people's awareness, media manipulation by AI is capable of spreading wrong information and causing an identity crisis. There is going to be a big problem for the integrity of talk, privacy and trust in the use of the digital platform.

2. Literature Survey

2.1 Deep Learning Fundamentals

Neural Networking Technology is a great form of a machine learning algorithm and lies under Deep learning algorithms capable of handling hierarchy data abstraction. Currently it has become a common strategy for such problems that requires some sort of pattern recognition like image classification, speech recognition and natural voice recognition. [7-10] CNNs are employed in computer vision since they utilize convolutional layers to capture picture hierarchies by applying

convolutions. Recurrent Neural Networks (RNNs) are designed to work with sequential data so that they are commonly used as good applications for time series analysis and natural languages. All these architectures and other innovations, such as attention mechanisms and transformers, have led to significant advancement in the field.

2.2 Evolution of Generative Models

Generative models have gone through development steps, with autoencoder models being in a prior stage where primary objectives were tasks like dimensionality reduction and unsupervised feature extraction. Autoencoder can be worked on encoding the input data to a space and decoding it back to the input itself in a way that is able to reconstruct the input. VAEs went further in this concept by including probabilistic elements and thus, achieve smoother and more versatile latent representations that enable the generation of novel samples. When it comes to Deep learning, the emergence of Generative Adversarial Networks GANs can be attributed to Good Fellow and his team in the year 2014. GANs work on the mechanism of a game theoretic framework where there is a generator that aims at synthesizing new data and a discriminator that aims to distinguish fake data from the original data to improve on the output mentor from the generator.

2.3 DeepFake Applications and Ethics

Deepfakes can be described as hyper-realistic synthetic media and its emergence can be traced on the backdrop of significant advancements of generative models. [11,12] For instance, Korshunov& Marcel (2018) and Afchar et al. (2018) have investigated DeepFakes with regard to picture or video manipulation, virtual representation, movies, and even the educational domain. However, such capabilities have implications that raise other major ethical issues. It is dangerous since it can be utilized to spread fake information, create speeches or actions of some configuration individuals, and violate personal rights. The key ethical concerns include consent, the authenticity of the objects included in the pictures or videos, and the overall weakening of the viewer's trust in the material they are presented with, making legal and technical protections necessary.

2.4 Detection Methods

New methods for DeepFakes since the existing ones get outdated due to the advancements. One of the methods used is to compare each frame of the video for any signs of manipulation, such as artefacts, thereby discrediting the content of the video. Another promising approach is the distinction of bodily signs, such as fluctuations in pulse or face redness, that can be observed in the high-quality video. Still, it might be difficult for a fake to mimic during the fake video streaming. Also, there exist automatic tools based on machine learning classifiers that detect forged content when working with a large number of real and fake media. Such classifiers use multiple features, from pixel-level anomalies to temporal patterns, in crucial steps of digital media integrity.

3. Methodology

3.1 System Architecture

For DeepFake generation or detection setup, the system structure is a stringent workflow consisting of consecutive steps acting on video feeds. [13-15] Every phase generates or processes the facial attributes through architectures such as GANs or VAEs. Again, each component in the flowchart for favorable and unfavourable transcripts is explained below:

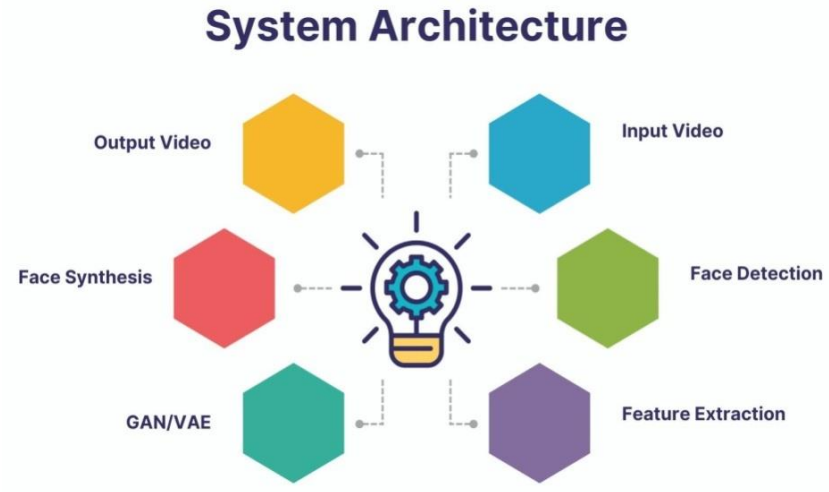

Figure 2. System Architecture

- **Input Video:** The first component of the system is an input video which involves faces of individuals that are to be analyzed or synthesized. This video can be one shot from a movie, an interview, or a video recorded by any user. The

initial step in processing the video is frame extraction, which makes it easier to manipulate the video based on the individual frames.

- **Face Detection:** In this stage, face detection is performed through faster-recognized models, which include the Haar cascade, MTCNN or current high CNN-based detectors. The purpose is to correctly identify and separate face areas in a frame. Face detection should be fast due to the fact that any false or inaccurate detection leads to the entire pipeline being affected.
- **Feature Extraction:** After that, facial features of importance are detected from the face detected earlier. Some of these are likenesses, such as eyes, nose and mouth; roughness; and geometric character. This feature set is useful to represent the face compact and informatively and will feed it as input to the generative model. Feature extraction may also include encoding the facial identity or the facial expression of the person involved in the activity in a different space.
- **GAN/VAE:** The extracted features are passed to a generative model as input. Depending on the system goal, it is going to be a Generative Adversarial Network or a Variational Autoencoder. Most commonly, GANs are employed for realistic face generation, where the technique employs a generator and discriminator. VAEs are more appropriate for smooth interpolation and representation learning among the two. Depending on a given task, it is trained to produce fake faces similar to the referred or target faces.
- **Face Synthesis:** In this, a fresh facial image is generated using the output of the GAN or the VAE. He then overlays the synthesized face to the background reference frame, matching the lighting, orientation, as well as facial expressions of the target face. This step aims to make the generated face look natural when the video replaces it back into its original position.
- **Output Video:** Lastly, the frames which have been manipulated or reconstructed are compiled together to form a video. Some technical aspects applied here may include smoothing images, color correction, and synchronizing audio with the videos. The final output video showcases the results of face synthesis either for entertainment, avatar creation, or detection analysis depending on the intended use of the system.

3.2 Dataset Preparation

The quality and variety of the datasets used in training or testing DeepFake detection or generation systems are crucial for the results. To such an end, various benchmark datasets have been formulated and commonly used in literature. FaceForensics++, Celeb-DF (v2), and DeepFake Detection Challenge (DFDC) datasets are popular. These are several thousands of real and manipulated video clips; the variety of faces, expressions, lighting conditions, and manipulation are included. FaceForensics++ is quite popular due to the presence of multiple forgery techniques and different levels of the compression technique, which makes it suitable for both training and testing. Celeb-DF is more advanced as compared to the previously proposed datasets in a way that DeepFake videos generated through it have minimal distortion compared to the first generation technique. The dataset used in the study is the DFDC dataset released by Facebook and includes more than 100 thousand videos from a diverse population range.

These datasets, however, must undergo a pre-processing step to bring them to a form suitable for training and improving the model's performance. Face detection is applied next on every video to identify the position of the cult up to obtaining a set of frames containing only facial area. The face images retrieved are first normalized; this entails resizing the images to one resolution, channel adjustment and normalization of pixel intensities. On the same note, enhancing variability with regard to data is attained through data augmentation strategies to reduce the overfitting of the models. It mandates; flipping, rotation, zoom, brightness, and noisy enhancement as usual augmentation methods. Thus, these transformations mimic various situations that may occur in the world, for example, changes in lighting and camera position, increasing the model’s stability. All in all, detailed pre-processing of the dataset increases the model's robustness when identifying or generating DeepFake content in open-world conditions.

3.3 Model Implementation

- **Face Swapping using Autoencoders:** Face swapping using Autoencoder is among the simplest methods of producing DeepFake. The architecture of the model comprises two features, that is, encoder and decoder. [16-20] The encoder takes the source face image, extracts out the subject's facial structure/geometry and expressions/identity, and compresses it into a vector space, usually of significantly lesser dimension. This information is then forwarded to the decoder trained to specifically reconstruct the face of the target subject. As for the target face, it combines the shape and appearance of the target face while preserving the expressions of the source face. This one is less complex and quick to learn, although it may not perform very well in high fidelity or have temporal coherence in the case of generated videos when no further processing is done.
- **GAN-based Approach:** GANs are considered an advanced and realistic methodology than DeepFake in producing fake content. In a GAN arrangement, there is G for creating a new image of fake face data and D for identifying whether the image is genuine or generated. One of the networks passes on the generated data to the other in an adversarial manner to better the output produced by the generator. At this level, the generator gains such excellent performance that produced synthetic faces are similar to real images of the human face in terms of texture, lighting

and sharpness. Variations such as head poses, lighting conditions, and even the facial expressions of the individuals are well handled by GANs. Moreover, a subclass of GAN, the cGANs, and the style-based GANs have enhanced the quality and controllability of face synthesis as they allow the model to condition its output based on an identity or another input variable.

3.4 Detection Model

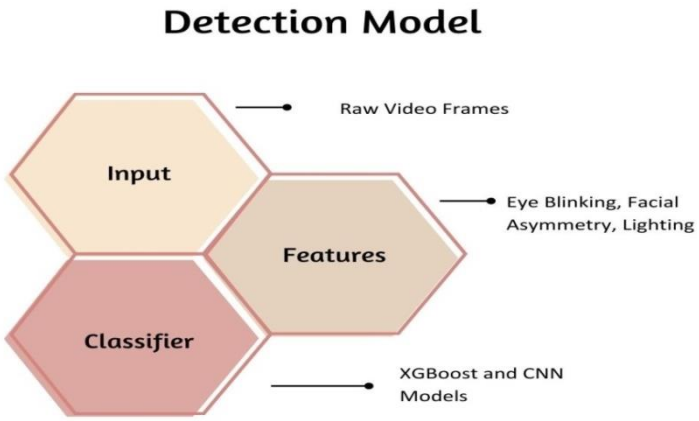

Figure 3. Detection Model

- **Input: Raw Video Frames:** The detection model initiates the following process by receiving raw video data, which is usually in the raw form of different frames. These frames are the source of input for the detection pipeline algorithm. As will be seen in the next section, the choice of the frame level of indexing not only allows for the fine analysis of temporal and spatial discrepancies that may be indicative of manipulation but also takes care of certain problems associated with utilizing the global-level approach. For realities, more specifically, the face area is isolated having resized and normalized the frames and performed the face detection to narrow the analysis on the faces. This method is useful for increasing the robustness of the detection system because it is challenging to detect DeepFake artifacts when they are not very distinct.
- **Features: Eye Blinking, Facial Asymmetry, Lighting:** Feature extraction is a critical step that enables real content to be figured out from fake content. Some indicators that may give away DeepFakes involve biological patterns of a subject and geometrical anomalies of a given image. For instance, eye blinking is less frequent or completely absent in the generated videos because of the absence of temporal consciousness in many generative models. Another work-related facial characteristic is asymmetry, which is not completely balanced in the real face but is subject to be minimized in DeepFakes, in which stroke faces seem perfectly aligned and balanced. Shading discrepancies are also useful features, as the faces produced by GAN can possess wrong shading or lighting compared to the rest of the video frame. Choosing and deriving these features helps the model distinguish synthetic contents with a greater margin of accuracy.
- **Classifier: XGBoost and CNN Models:** These extracted features are fed into classifiers that determine whether a given frame can be categorized as real or fake. There are two standard models which can be used; they include XGBoost and Convolutional Neural Networks (CNNs). It acts as a gradient-boosting algorithm and is quite good at handling structured features such as blinking frequency or asymmetry scores. CNNs, in turn, are good at learning the pattern of image data directly from pixels, which is why they are used often in end-to-end detection. It has been found that, in most cases, using both structured features with XGBoost and visual patterns with CNNs greatly enhances the performance of the detection system.

4. Results and Discussion

4.1 Performance Evaluation

In order to ascertain how the proposed detection models work, experiments have been performed using two well-known datasets in the domain of DeepFake detection: DFDC, DeepFake Detection Challenge and FaceForensics++. These are different real and manipulated video clips used for real time rating of the detector's performance. The criteria used to evaluate the models were tested by the detection accuracy, which is the percentage of frames categorized as fake or real. This is important because it goes straight to the ability of the model to tell real from fake content under different manipulations and versatile artefacts.

Table 1. Detection Accuracy of Models

Model	Accuracy
XGBoost on DFDC	88%
CNN on FaceForensics++	91%

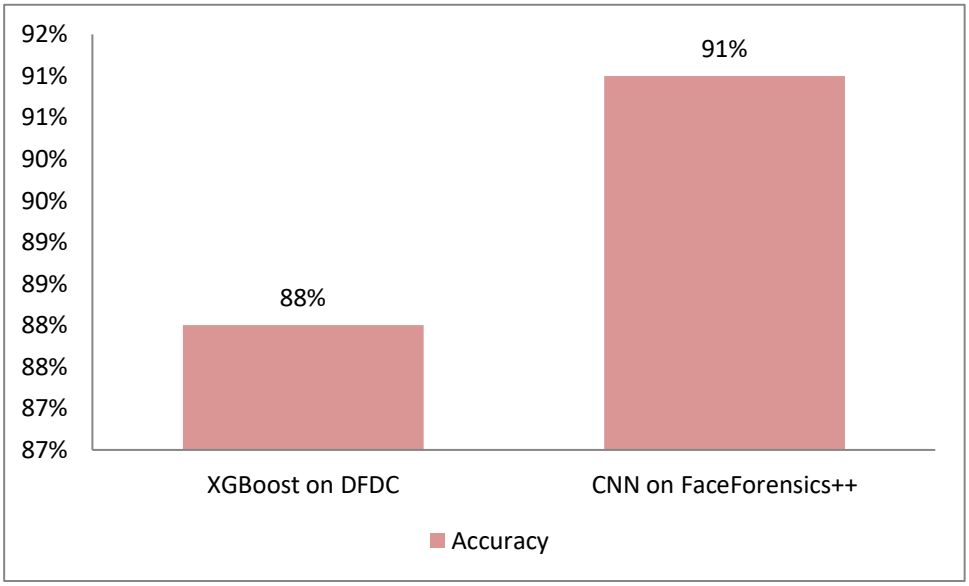

Figure 4. Graph representing Detection Accuracy of Models

- **XGBoost on DFDC:** Finally, using the machine learning algorithm called XGBoost, an accuracy of 88% was gained on the selected DFDC dataset. The DFDC dataset is a comprehensive dataset of real and DeepFake videos with high-quality DeepFake videos made by different techniques including facial swap and reenactment. As a gradient-boosting model, XGBoost successfully exploited all eight features, including facial landmarks, blinks, and lighting. However, the present dataset was no exception; its increase in diversities yielded problems such as the accommodation of new or unknown manipulation methods, which lessened the performance. However, the robustness of XGBoost and the excellent way it handles the data imbalance led to this good performance.
- **CNN on FaceForensics++:** Using the FaceForensics++ dataset, an attempt was made to implement a Convolutional Neural Network (CNN) model with an accuracy of 91%. FaceForensics++ provides several types and quality of DeepFake videos and thus is quite competitive when it comes to detection. They are very efficient at learning features spontaneously from image data; this allows them to describe small distortions of texture, facial expressions, lighting, etc. The improved CNN model works better, particularly in face forensic datasets like FaceForensics++, since the dataset entails difficult forgery methods such as style transfer and high-quality facial synthesis forgery.

4.2 Case Studies

- **Political DeepFakes:** Former British Prime Minister David Cameron is also concerned due to DeepFake technology, which creates shocking videos in which the individuals never appear but appear to speak or act in a specific way. This is dangerous because it may lead to misleading people, controlling their thinking, and even tampering with systems such as the voting system. Such fake news embedded in these videos is believable, and if not detected, it could have social and political effects. In this case, the detection system was tested on several political DeepFakes on some leaders, which was done to make them utter statements or support certain stands they did not make or support. The system identified most of these manipulated videos, mainly those with morphed faces or flickering, which are typical in DeepFake videos. It, therefore, explains the importance of detection tools in preventing such manipulations that can otherwise skew public perception and/or affect the political sphere.
- **Celebrity DeepFakes:** Another existing trend is the production of DeepFake concerning celebrities for entertainment purposes or even for some immoral activities, such as posting pornography involving celebrities without their consent. These videos, for example, bring ethical issues such as misuse of identity and lack of digital consent. Celebrities whose photographs circulate on social media are often the victims of such manipulations. This is because the unauthorized portrayal of their identity is likely to have a negative impact on the two in their respective fields. In our study, the proposed detection system accurately identified DeepFakes involving celebrities, especially when they had fake eye movements, poorly aligned faces, or otherwise poorly applied face swaps. Consequently, through the examples of these discrepancies, the operation of our system can be shown as being able to address real-world ethical dilemmas related to DeepFake technology and serve as a technological barrier to said technology being used for gains, harassment, or defamation. This also means that there is a need to employ means to detect violations of people's rights and uphold online purity.

4.3 Limitations

- **Transferability:** Another major drawback is the generalization of the models across different data sets. However, a model that may arrest high accuracy on a certain dataset, like the DFDC or FaceForensics++, may not perform optimally on other datasets with different distinct manipulation techniques. For instance, a model worked on facial swapping fails to work on videos which have been through more advanced techniques such as voice synthesis or lip-synch. This limitation is due to the variety of DeepFake approaches and differences in the nature of generated videos. In these challenges, it is necessary to employ domain adaptation and transfer learning strategies. These techniques enable a model to update the model or modify the features or features obtained from one set of data to perform better in other types of new, unfamiliar DeepFake.
- **Real-time Detection:** It is still difficult to identify DeepFakes in real-time and even more so when it comes to processing large numbers of frames in the video stream. Even though CNNs show high efficiency in detecting DeepFakes, they can be computationally expensive and slow, especially when analyzing large file formats such as videos or live streams. The latency problem arises when CNNs are computationally demanding and used in environments such as video monitoring or content filtering, the latency problem arises. This leads to delays which are inconceivable in many applications of the algorithm. To address this, currently, the two techniques used are known as model pruning, where unnecessary neurons in the network are eliminated, and quantization, where the resolution of the weights in the model is reduced. These methods assist in making the dimensionality of the model and computation burdens, thereby permitting faster processing without much impact on the detection compass. However, it can be noted that real-time detection to its full potential with high accuracy is an ongoing issue for scientists and engineers.

5. Conclusion

First, speaking about the positive impacts of the DeepFake application, it is quite possible to distinguish considerable opportunities for this type of AI breakthrough. DeepFakes, in the form of videos, which are created using deep learning models such as GAN and autoencoder, display high fidelity, making their abuse a real possibility. Even though these techniques hold great potential for cruise entertainment and media broadcasting and streaming, they come with serious implications and risks, especially when it comes to video manipulation for misinforming people, identity theft, and political interference. The displayed analysis suggests that deep learning models, especially CNNs and XGBoost classifiers, yield high detection rates ranging up to 91% on DeepFakes. However, there are also evident drawbacks in the program, including the transfer of models to different datasets and the detection in real-time since processing high-resolution videos consumes a lot of time and resources. Essentially, the authors noted that deep learning is also the root of the problem and the tool for DeepFake detection, contributing to the necessity to develop enhanced and universally capable models and algorithms for DeepFake detection.

5.1 Future Directions

Here are the areas that need further research to enhance the recommendations and methods to tackle DeepFake. One such direction is federated learning for distributed detection, where the models are trained in one or multiple decentralized devices without data sharing. It also ensures the privacy of the users as well as real-time detection of DeepFakes within the user devices. Thus, federated learning allows the building more scalable and productive detection systems that respect user privacy. Another opportunity is to incorporate blockchain in tracking the origination of files through tracking the chain of command of a given file. It was seen that through blockchain, a proven and tamper-free way of creating and manipulating videos can be allowed. At the same time, it can also give consumers confidence that the videos' content is original.

5.2 Ethical Considerations

DeepFake is a complex issue that has both virtues and evils, and these evils must be urgently solved as technology advances further. Accompanying legislative frameworks and AI ethics guidelines need to be developed to prevent abuses of DeepFakes in such applications as revenge pornography, political manipulation, defamation, etc. Modern legislation in most jurisdictions fails to address the evil perpetrators as the technologies progress since creating and distributing DeepFakes for malicious purposes is not prohibited by law. In these issues, a definition and general principles must be formed to define the appropriate use of Artificial Intelligence. Also, rules have to be adopted to hold authors of DeepFake technology responsible for the consequences of the actions of their invention. DeepFake technology is becoming more available, amplifying the need to address the issue as it threatens people’s rights and identity regarding their privacy and the responsible use of AI.

References

[1] Guarnera, L., Giudice, O., & Battiato, S. (2020). Deepfake detection by analyzing convolutional traces. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops (pp. 666-667).

[2] Chesney, B., & Citron, D. (2019). Deepfakes: A looming challenge for privacy, democracy, and national security. Calif. L. Rev., 107, 1753.

[3] Lewis, J. K., Toubal, I. E., Chen, H., Sandesera, V., Lomnitz, M., Hampel-Arias, Z., ... & Palaniappan, K. (2020, October). Deepfake video detection based on spatial, spectral, and temporal inconsistencies using multimodal deep learning. In 2020 IEEE Applied Imagery Pattern Recognition Workshop (AIPR) (pp. 1-9). IEEE.

[4] Ciftci, U. A., Demir, I., & Yin, L. (2020). Fakecatcher: Detection of synthetic portrait videos using biological signals. IEEE transactions on pattern analysis and machine intelligence.
[5] Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2018). AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Minds and Machines, 28, 689-707.
[6] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. Advances in neural information processing systems, 27.
[7] Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. science, 313(5786), 504-507.
[8] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780.
[9] Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 4401-4410).
[10] Kingma, D. P., & Welling, M. (2013, December). Auto-encoding variational bayes.
[11] Korshunov, P., & Marcel, S. (2018). Deepfakes: a new threat to face recognition? Assessment and detection. arXiv preprint arXiv:1812.08685.
[12] Afchar, D., Nozick, V., Yamagishi, J., & Echizen, I. (2018, December). Mesonet: a compact facial video forgery detection network. In 2018 IEEE international workshop on information forensics and security (WIFS) (pp. 1-7). IEEE.
[13] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.
[14] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.
[15] Li, Y., & Lyu, S. (2018). Exposing deepfake videos by detecting face-warping artifacts. arXiv preprint arXiv:1811.00656.
[16] Matern, F., Riess, C., & Stamminger, M. (2019, January). Exploiting visual artefacts to expose deepfakes and face manipulations. In 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW) (pp. 83-92). IEEE.
[17] Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). Faceforensics++: Learning to detect manipulated facial images. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 1-11).
[18] Meskys, E., Kalpokiene, J., Jurcys, P., & Liaudanskas, A. (2020). Regulating deepfakes: legal and ethical considerations. Journal of Intellectual Property Law & Practice, 15(1), 24-31.
[19] Younus, M. A., & Hasan, T. M. (2020, April). Effective and fast deepfake detection method based on haar wavelet transform. In 2020 International Conference on Computer Science and Software Engineering (CSASE) (pp. 186-190). IEEE.
[20] Chintha, A., Thai, B., Sohrawardi, S. J., Bhatt, K., Hickerson, A., Wright, M., & Ptucha, R. (2020). Recurrent convolutional structures for audio spoof and video deepfake detection. IEEE Journal of Selected Topics in Signal Processing, 14(5), 1024-1037.