Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 11, No.1, November : 2024

DEEPFAKE DETECTION IN LOW-RESOLUTION CCTV FOOTAGE

Shweta Tripathi, Research Scholar, Faculty of Computer Engineering, Pacific Academy of Higher Education & Research University, Udaipur
Dr. Mukesh Shrimali, Professor, Pacific Academy of Higher Education & Research University, Udaipur : shwetripathi@gmail.com

Introduction

Deepfakes videos are digitally modified clips that use machine learning algorithm to swap out people or objects for the people in an original picture of video [1]. The use of deepfakes to harm the reputations of public individuals such as politicians, celebrities, and other presents a serious challenge to modern civilization [2]. The widespread distribution of these videos has the risk of causing significant disruption and harm to communities and countries around the globe. Image and video editing becomes accessible to almost anybody with a computer because of the exponential growth of artificial intelligence and deep learning-based accessible modern tools, techniques, and software, such as convolutional neural networks [3], autoencoders, generative adversarial networks [4] etc. Because of the expanding multimedia forgery which is deepfakes it is almost impossible for anyone to state the difference between real and fake content only through human eye inspection. On the other hand, the resources needed to produce deepfakes are publicly accessible on the internet and are open sourced. A well-liked open source-swapping program is FaceApp [6] which lets the user alter features such as age, gender, and hairstyles by applying transformations to face photos. Using the artificial intelligence method, FaceApp allows to construct masks that can change their appearance in videos. Another deep learning-based method that uses generative adversarial networks (GANs) to swap the faces in photos is called FaceSwap [5]. Fig. 1 shows the basic work flow of GANs for deepfakes graphically.

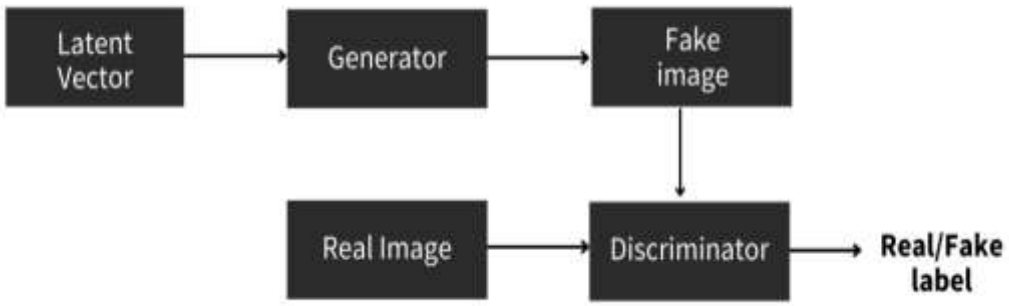

Figure 1. Generative adversarial networks for deepfakes

Majority of the fake detection research has concentrated on images, with videos receiving less attention because the analysis of videos has distinct difficulties such as a huge volume of data and sometimes insufficient data availability. Effective algorithms are also needed for the extremely difficult task of detecting fraudulent faces in low-quality videos.

Related Work

There is still a dearth of study in the rapidly developing subject of deepfake video and image manipulation detection. There is a lot of potential for new discoveries in this developing field. Some of the previous research, for example a model based on recurrent neural networks (RNNs) [8] was proposed in a study in [7] to identify deepfakes. Convolutional LSTM (Conv-LSTM) is a unique technology that is created by combining CNNs with long short-term memory (LSTM) [9] networks. In their method, LSTMs are used for video analysis while CNNs are used to extract information from each frame. To be more precise they used the InceptionV3 [10] model as the CNN's fully linked top layer. The images were cropped to the correct size using an ImageNet [11] model that had already

been trained. The CNN's output was thereafter sent to the LSTM network for additional examination in order to ascertain if the frames were authentic or fraudulent. In this model the dropout rate of 0.5 and configuration of 2048 by 2048 units was used and a 512-unit fully linked layer with a dropout rate of 0.5 percent was then employed.

The researchers of [12] introduced the groundbreaking Deepfake technology, which lets users swap out one person's face in a video for another. The PRNU study found a substantial difference in the average standardized cross-correlation test between authentic videos and Deepfakes for both. The dataset includes real, unbiased footage captured at 1240 x 720 pixels using a Canon Powershot SX210IS. OpenFeedShop is used by the Powershot SX210IS and Deepfeke GUI for Deepfake. With the program from FFPpeig, the videos are produced in a sequence of frames as PNG. The frames are separated into eight groups with a PRNU pattern that is average in size. developed for every group using the "PRNU comper" program and the second-order (FSTV) approach. Next, these eight PRNU patterns are contrasted with each other.

Creation of Deepfakes

The employment of paired encoder-decoder models [13] to create deepfake content, as seen in Fig. 2, it raises serious questions regarding the improper application of a cutting-edge technology. Presently, artificial intelligence generative models have progressed beyond the conventional purview of scholars, posing significant obstacles in the identification of counterfeit facial duplicates. To tackle this problem, the authors provide a novel method for identifying videos that have been synthesized. In order to classify the authenticity of video portraits, this method is based on the analysis of biological signals from faces, which are subsequently processed and retrieved as characteristics. The goal of this strategy is to make a substantial contribution to the field of bogus video detection [13].

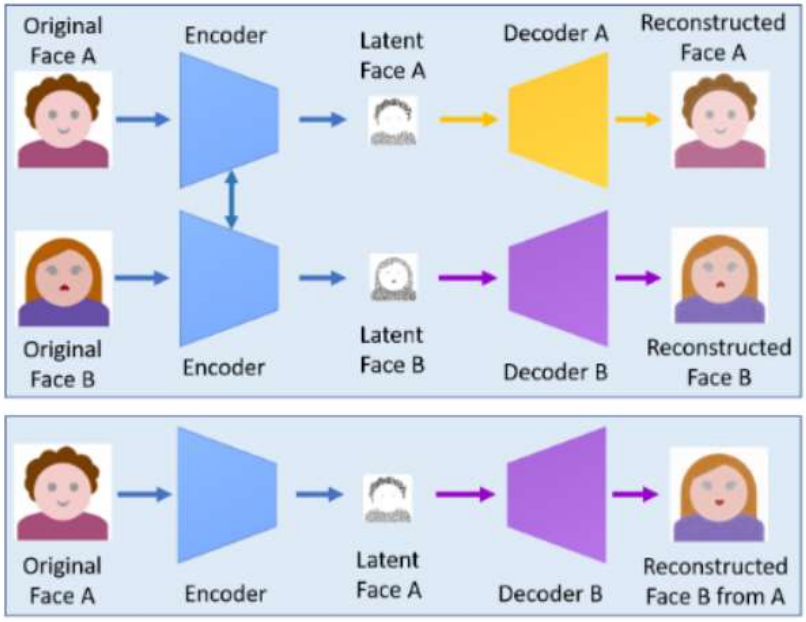

Figure 2. Swapping of face using auto encoders

Since image and video manipulation techniques are developing quickly, it is critical to expose these deepfakes as soon as possible. To identify deepfakes, a variety of techniques are used, including looking for variations in head orientation, position, and facial feature alignment [16]. Segmentation techniques can also be used to identify splicing, copy-move, and removal attacks in order to discover modified images. Semantic segmentation additionally identifies altered areas by returning bounding boxes that emphasize the manipulated areas [16]. An additional method for identifying fake videos is to examine how people blink their eyes; this can reveal information about the legitimacy of the video [15]. Moreover, local noise analysis can reveal areas that have been manipulated in secret, greatly improving the accuracy of fake image detection [14].

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 11, No.1, November : 2024

Detection of Deepfakes

The speed at which fake news spreads across many platforms is making it harder for academics to identify it. While existing deep learning and vision models do a great job at accurately identifying high-quality forged films, they have trouble distinguishing facial landmarks in compressed, low-quality videos. The increasing number of fake videos on the internet makes this problem worse, emphasizing the urgent need for a sophisticated AI model that can identify all kinds of fake videos and stop them from spreading quickly. Facial manipulation techniques include full facial synthesis, identity switching, attribute manipulation, and expression swapping, among others, that can be used to identify fraudulent movies [21]. The accuracy and resilience of fake video detection systems can be enhanced by addressing the particular problems and opportunities presented by these manipulations. Some of these manipulations are shown in figure 3. There are a lot of approaches that are used for manipulations using GAN architectures.

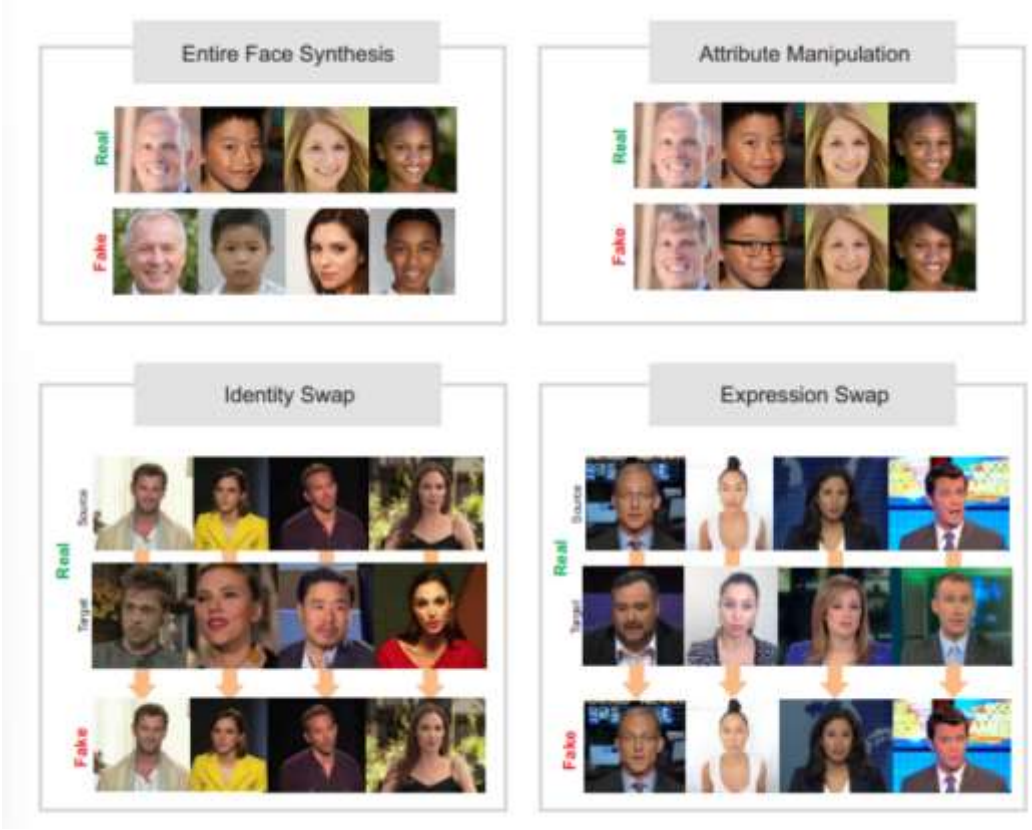

Figure 3. Manipulations of Facial expressions

There are a lot of methods which helps achieving Deepfake detection and the field is always in a developing state. In order to know the difference between artificial and real head motions, these techniques include looking for visual artifacts [21], tracking eye blinking [20], and examining the head pose movements [24]. Model training has been done using a two-stream CNN network [19], which makes use of a classification model that is built on the LeNet architecture. Furthermore, MesoNet [23] is intended to learn discrete features from videos frames because to its shallow architecture that includes an inception module. A deepfake detection method developed by Matern [22] looks for absent reflections and ocular features to identify artificial faces. Since synthesized face regions are frequently spliced, deepfakes that add discrepancies can be identified using 3D head posture estimations. In order to classify facial emotions, Agarwal and Farid [18] suggested an analysis technique that uses a Support Vector Machine (SVM) to measure the head movement distances and the movements of face muscles in the check and nose regions.

One of the most important factors in identifying synthetic recordings is the lack of reflection and illumination details in features such as teeth and eyes. Several studies [25] have shown that deep

structures perform noticeably better in this situation than shallow networks. Different deepfake detection techniques work well on high-quality videos, but they have trouble on the low-quality ones. According to Geura and Delp's theory [24], deepfake videos display temporal and intra-frame anomalies which are harder to spot in the low-quality videos but easier to spot in the high-quality ones. Yisroel stresses that crucial elements for identifying deepfakes include lip movement, gaze direction, posture, facial expression, and body language [25]. Some deepfakes can be easily determined by machine when they cannot be detected by humans directly.

Methodology

In order to detect Deepfakes in an Image or Video, a methodology given in figure 4 could be followed. This methodology can be divided into four major sections that are data pre-processing, spatial feature extraction followed by temporal feature extraction and classification.

(1) Data preprocessing and preparation: We can use a tagged dataset of real and fake videos as the source for deepfake detection method and extract frames at a rate of five frames per second to record temporal changes. We can then recognize and crop the faces using Multi-task Cascaded Convolutional Networks (MTCNN) [26]. To guarantee consistent input for neural networks, the cropped face images could be scaled, normalized and data augmentation can be applied to increase robustness. After dividing the data into training and testing sets in an 80:20 ratio, the pre-trained InceptionV3 model could be used to extract features for additional temporal analysis. This preparation method could improve the performance of a deepfake detection system by ensuring consistent, high-quality data.

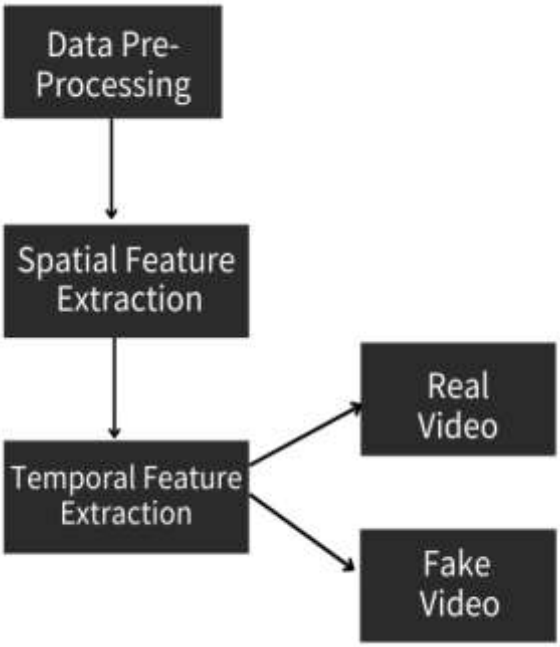

Figure 4. Architecture for detection of Deepfakes

(2) Spatial feature extraction: After preprocessing, spatial feature extraction is essential for deepfake detection in low-resolution CCTV footage after preprocessing. To extract fine-grained spatial features from the facial images, we could employ the InceptionV3 model, which is been pre-trained on ImageNet. The model's specific layers are adjusted to identify features unique to our dataset, such as edges, textures, and face traits. To control complexity, feature maps can be created and dimensionality reduction methods like PCA or t-SNE are used. For consistency, the generated feature vectors are scaled and normalized. To produce a thorough spatial representation, other spatial elements can be combined, such as texture and key point descriptors. By combining these traits with temporal data, analysis is robust and improves our system's detection accuracy.

(3) Temporal feature extraction: In order to detect deepfakes in CCTV footage, temporal feature extraction is crucial in order to capture the changes across video frames. The frame sequences could be examined using Long Short-Term Memory (LSTM) networks in order to find temporal irregularities that might point to deepfake alterations. The LSTM network can identify minute, time-based abnormalities that static analysis could overlook by looking at how features change over time.

(4) Classification: Classification could be performed using the activation function softmax classification that also uses neurons depending on the input.

Testing and Evaluation

In order to evaluate how well a classification model performed on a set of test data we use the following measures, in which TP refers to as True positives, TN as True negatives, False positives as FP and False negatives as FN.

1. Accuracy: It is the measure of how accurate a model is by dividing successfully predicted by the total number of predictions.

$$\text{Accuracy} = \frac{TP + TN}{TP+FN+TN+FP}$$

2. Precision: Precision is the ratio of total Positive samples correctly classified to the total number of samples classified as positive.

$$\text{Precision} = \frac{TP}{TP + FP}$$

3. Recall: It is the total number of Positive samples accurately categorized as Positive divided by the total number of Positive samples is known as recall.

$$\text{Recall} = \frac{TP}{TP + FN}$$

Conclusion and Future Directions

Because false media technologies are getting more advanced and accessible, academics face a big issue when it comes to synthesized media. Using cutting-edge AI methods to identify these frauds is imperative. Our work focuses on efficient facial extraction for the detection of low-resolution deepfake videos. The need to create efficient detection techniques for both low- and high-resolution videos arises from the ongoing improvement in the quality of deepfake videos. Currently, available research often focuses on the shortcomings in deepfake production pipelines, like irregularities in colour, shadows, and facial features.

The Future directions for deepfake detection include creating automated tools that can quickly and reliably identify manipulated media, with an emphasis on deep-learning based strategies that allow for domain generalization. More investigation is required to fill in the current gaps in audio deepfake detection, especially with regard to identifying fakeness in accented speech or ambient noises. The conflict between deepfake generation and detection offers insightful information about the opportunities, trends, and challenges facing the field of deepfake research. Future studies should also concentrate on enhancing the domains of deepfake generation and detection, taking into account the processes involved in creating and detecting deepfakes, as well as their present and potential constraints and directions.

References

1. S. Adee, "What are deepfakes and how are they created?" IEEE Spectrum, Jun 2021. [Online]. Available: https://spectrum.ieee.org/what-is-deepfake
2. L. Nataraj, T. M. Mohammed, B. Manjunath, S. Chandrasekaran, A. Flenner, J. H. Bappy, and A. K. Roy-Chowdhury, "Detecting gan generated fake images using co-occurrence matrices," Electronic Imaging, vol. 2019, no. 5, pp. 532-1, 2019.
3. S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in 2017 International Conference on Engineering and Technology (ICET). Ieee, 2017, pp. 1-6.
4. A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview," IEEE Signal Processing Magazine, vol. 35, no. 1, pp. 53-65, 2018.
5. KORSHUNOVA I., SHI W., DAMBRE J., THEIS L. Fast face-swap using convolutional neural networks. In: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 3677–3685.
6. WIRELESS LAB. FaceApp Open-source software Tool available. 2016. Available from: https://www.faceapp.com.
7. D. Guera and E. J. Delp, "Deepfake video detection using È recurrent neural networks," in 2018 15th IEEE international conference on advanced video and signal based surveillance (AVSS). IEEE, 2018, pp. 1-6.
8. [19] A. Sherstinsky, "Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network," Physica D: Nonlinear Phenomena, vol. 404, p. 132306, 2020.
9. [20] H. Sak, A. W. Senior, and F. Beaufays, "Long short-term memory recurrent neural network architectures for large scale acoustic modeling," 2014
10. 21] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2818-2826.
11. [22] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in 2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009, pp. 248-255.
12. M. Koopman, A. M. Rodriguez, and Z. Geradts, "Detection of deepfake video manipulation," in the 20th Irish machine vision and image processing conference (IMVIP), 2018, pp. 133-136.
13. NGUYEN T.T., NGUYEN C.M., NGUYEN D.T., NGUYEN D.T., NAHAVANDI S. Deep learning for deepfakes creation and detection. arXiv preprint arXiv:1909.11573. 2019, 1.
14. HAN X., MORARIU V., LARRY DAVIS P.I. Two-stream neural networks for tampered face detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 19–27.
15. LI Y., CHANG M.-C., LYU S. In ictu oculi: Exposing ai generated fake face videos by detecting eye blinking. arXiv preprint arXiv:1806.02877. 2018.
16. YANG X., LI Y., LYU S. Exposing deep fakes using inconsistent head poses. In: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019, pp. 8261–8265.
17. AFCHAR D., NOZICK V., YAMAGISHI J., ECHIZEN I. Mesonet: a compact facial video forgery detection network. In: 2018 IEEE International Workshop on Information Forensics and Security (WIFS), 2018, pp. 1–7.
18. AGARWAL S., FARID H., GU Y., HE M., NAGANO K., LI H. Protecting World Leaders Against Deep Fakes. In: CVPR Workshops, 2019, pp. 38–45.
19. HAN X., MORARIU V., LARRY DAVIS P.I. Two-stream neural networks for tampered face detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 19–27.

20. LI Y., CHANG M.-C., LYU S. In ictu oculi: Exposing ai generated fake face videos by detecting eye blinking. arXiv preprint arXiv:1806.02877. 2018.
21. LI Y., LYU S. Exposing deepfake videos by detecting face warping artifacts. arXiv preprint arXiv:1811.00656. 2018.
22. MATERN F., RIESS C., STAMMINGER M. Exploiting visual artifacts to expose deepfakes and face manipulations. In: 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW), 2019, pp. 83–92
23. YANG X., LI Y., LYU S. Exposing deep fakes using inconsistent head poses. In: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019, pp. 8261–8265.
24. GUERA D., DELP E.J. Deepfake Video Detection Using Recurrent Neural Net- ¨works. In: 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2018, pp. 1–6. doi: 10.1109/AVSS.2018.8639163.
25. MIRSKY Y., LEE W. The Creation and Detection of Deepfakes: A Survey. arXiv preprint arXiv:2004.11138. 2020.
26. ZHANG K., ZHANG Z., LI Z., QIAO Y. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. IEEE Signal Processing Letters. 2016, 23(10), pp. 1499–1503, doi: 10.1109/LSP.2016.2603342.