Available online @ https://jjem.jnnce.ac.in
https: www.doi.org/10.37314/JJEM.SP0267
Indexed in International Scientific Indiexing (ISI)
Impact factor: 1.395 for 2021-22
Published on: 08 December 2024

Unmasking DeepFake

Sumanth M V[1*], Arunkumar K L[2]

[1] Student, Department of Computer Application, JNNCE, Shimoga
[2] Assistant Professor, epartment of Computer Application, JNNCE, Shimoga
Sumanthmv411@gmail.com, arunkumarkl@jnnce.ac.in

Abstract

This paper presents a method to automatically and efficiently detect face tampering in videos, deepfake has emerged as a significant challenge in the digital age. The system utilizes the convolutional neural network (CNN) to extract frame-level features and detect the fake videos that are created. While deepfakes hold potential for beneficial applications in entertainment, education, and creative arts, their misuse poses serious threats like spread misinformation in the society. These threats encompass the circulate of misinformation, defamation, erosion of trust in digital media, political manipulation, and the potential for unprecedented levels of cybercrimes, Various techniques are employed to distinguish real from fake media, such as analyzing face swapping indicators, detecting behavioral anomalies, and identifying inconsistencies in a person's face expression, this deepfake detector can be adopted by police to find if the video is legit or fake.

Keywords: *Deepfake, Synthetic, convolution neural network (CNN),Long Short-Term Memory (LSTM).*

1. Introduction

The improving resolution of the smart phone cameras and the widespread availability of high-speed internet have greatly expanded the use of social networking and media-sharing platforms, making it easier to create and stream digital videos. At the same time, advances in computing power have significantly enhanced deep learning capabilities, enabling technologies that seemed impossible just a few years ago. However, these sophisticated technologies come with new challenges. One such challenge is the emergence of "DeepFakes," which are created using advanced GAM that can modify video and audio. The spread of DeepFakes on social media platforms has become a significant issue, contributing to the increase of wrong information and deceptive content. These malicious DeepFakes pose serious threats by misleading and potentially harming ordinary people in society. In this generation deepfake has been used to defame a person identity by using the targeted persons face and swap to a different persons face and make them mentally weak and spoil they reputation, deepfake is also used in criminal cases to swap identity and make the innocent person suffer mentally due to this deepfake technology

Figure. 1: Person with face swapped.

Deepfake technology, which includes various face modification techniques, uses advanced tools like computer vision and advanced machine learning. Face manipulation is categorized into four types: expression swap, full-face synthesis, attribute manipulation, and identity swap. Identity swap, commonly known as face swap, is a prevalent form of

deepfake video where the faces of original individuals are replaced with the faces of target individuals. Deepfake news can also combine forged videos and photos, making it hard for users to tell if they are real or fake. This type of deepfake can easily spread on social media and negatively impact people's lives. Although some deepfakes can be generated using conventional visual effects or computer graphics, the sophistication of current technology makes them more convincing and harder to detect. The first-generation Deepfake videos were easily identified by the naked eye, but the third-generation Deepfake videos are difficult to identify, which may cause problems in society. The deepfake not only have disadvantage but also use full like in the entertainment industry, deepfake technology offers innovative possibilities. Filmmakers and video game developers can use it to create lifelike characters and seamless special effects, reducing production costs and time. Actors can be digitally de-aged, and deceased performers can bring back to life on screen, providing new creative avenues for storytelling. Additionally, deepfakes can be utilized for dubbing foreign films, matching the lip movements of actors to translated dialogue, thereby enhancing the viewing experience.

2. Literature Survey

In this part, we will discuss the various research works in the domain of Deepfake creation and detection. Deressa at al.(2021) [18]“A Large-scale Challenging Dataset for DeepFake Forensics”. Deepfakes technology is a serious threat when used for harmful purposes such as phishing, scams, and identity theft, as they reduce the trustworthiness of public data. To detect deepfakes, this project uses a convolutional vision transformer (CViT). The project integrates a CNN module into the ViT architecture because CNNs are effective at extracting features, like facial features in images, which the ViT then uses to classify the images. The datasets used include Face Forensics Face swap, Face Forensics Deepfake detection, Darius Afchar at al.(2018) [2]" MesoNet:a Compact Facial Video Forgery Detection Network", Agreed that the huge use of images that are edited by the software digital to change image contents, using editing software like Photoshop. The digital image forensics research field is dedicated to detecting fake images to regulate the circulation of such fake content. Providing two possible network architectures (Meso 4 and Meso Inception 4) to detect such forgeries efficiently with a low computational cost. The dataset used is the Deepfake dataset, the Face2Face dataset. Both networks have reached close scores, around 90 %, considering each frame independently. A higher score is not expected as some images have facial extractions with a very low resolution. It is observed a decline of scores at the strong video compression level. The image aggregation significantly enhanced both detection rates. It even rose greater. Xin Yang at al. (2018) [19]“Exposing Deep Fakes Using Inconsistent Head Poses", Deepfakes have a major impact on our environment today, as they are created by inserting faces into original images or videos utilizing deep neural networks. Along with other forms of misinformation spread through digital social networks, digital impersonations created by deepfakes have become a serious problem with negative social consequences. Therefore, valid methods for detecting deepfakes utilized for this study are UADFV and a subset of the DARPA dataset. The findings indicate that SVM classifier achieves an AUROC of 0.89, indicating that the difference between head poses evaluated from the central region and the whole face is a useful feature for identifying deepfake images. For the DARPA GAN Challenge dataset, the SVM classifier achieves an AUROC of 0.843. This lower score is because of the synthesized faces in the DARPA GAN challenges often being blurry, making it difficult to accurately predict facial landmark positions and evaluate head poses. Performance was also evaluated by analyzing individual videos within the UADFV dataset. dataset by averaging the classification predictions on frames across individual videos. Yuezun Li at al.(2020) [10]“Celeb-DF: A Large-scale Challenging Dataset for DeepFake Forensics" Deepfakes, a blend of 'deep learning' and

'fake,' refer to synthetic media where a person's likeness in an image or video is replaced with another's. The rise of deepfakes has sparked significant concerns due to their potential misuse in spreading misleading or harmful content. As a result, the field of digital image forensics has advanced to develop effective methods for detecting these forgeries. This literature review explores different deepfake detection techniques, with a focus on two network architectures Meso4 and MesoInception4 and their application to widely-used datasets like Deepfake and Face2Face datasets. Zeina Ayman at al.(2023) [1]“DeepFake DG: A Deep Learning Approach for Deep Fake Detection and Generation”, The advent of deep learning technologies has significantly impacted various fields, including digital media. One of the most notable and controversial applications is the creation of deepfakes Another significant approach involves using Support Vector Machine (SVM) classifier to identify deepfakes by evaluating head poses. Research utilizing the UADFV and DARPA datasets demonstrated that the SVM classifier could achieve an Area Under the Receiver Operating Characteristic (AUROC) of 0.89 for the UADFV dataset and 0.843 for the DARPAGAN Challenge dataset. This method leverages the inconsistencies between the head pose calculated from the central facial region and the entire face as a key feature for detection. Kimaya Kulkarni at al. (2022), [9]“DeepFake Detection: A survey of countering malicious Deep-Fakes” The ResNext50 model is utilized to extract features. and precisely find them at the level of each frame. The CNN will be fine-tuned by adopting extra layers and choosing an appropriate training rate to ensure it converges with the gradient. After the final pooling layers, there are embeddings for vectors of 2048 dimensions that will be utilized as input for the sequential LSTM. The model comprises of ResNext50 32x4d and an LSTM layer. The Data Loader splits preprocessed, face-cropped videos into learning and testing sets. Frames that are extracted from these videos are then fed into the model in mini-batches for exercise and testing. Yuezun Li at al. (2018), [11]“In Ictu Oculi: Exposing AI Generated Fake Face Videos by Detecting Eye Blinking”. Exposing AI Created Fake Videos by Detecting Eye Blinking" describes a new method to find fake face videos generated with deep neural networks. This method focuses on detecting eye blinking, a physiological signal often missing in synthetic videos. The technique was tested on eye-blinking detection datasets and showed promising results in identifying videos created with deepfake software.The LRCN model is developed using image datasets of eye open states. The algorithm is then tested for detecting eye blinking in authentic and fake videos generated with the DeepFake algorithm. Huy H. Nguyen at al. (2019) [14]“Using capsule networks to detect forged images and videos”. Describes a method that utilizes a capsule network to detect forged and altered images and videos in various scenarios, such as replay attack detection and computer-generated video detection. However, they used random noise during the training phase, which is not ideal. While the model performed well on their dataset, it might fail on real-time data because of the noise in training. In contrast, this approach is intended to be trained on data without any noise. and real-time datasets. Capsule networks use fewer parameters than traditional convolutional neural networks (CNNs) while maintaining similar performance. Their application to forensics is demonstrated through detailed analysis and visualization. Sharves at al. (2024) [17]“A Multimodal Approach Harnessing AI to Expose Digital Deceptions”. The mechanisms behind deep fake generation, particularly the utilization of Generative Adversarial Networks (GANs) coupled with auto encoders, involve a multi-stage process. Initially, frame-level detection is conducted using a ResNet Convolutional Neural Network (CNN), analyzing individual frames of the video to identify potential manipulations or anomalies characteristic of deep fakes. Subsequently, the video undergoes classification using RNN, specifically LSTM networks. Raja al. (2017) [16]“Transferable deep-CNN features for detecting digital and print-scanned morphed face images”. In this study, a novel approach and framework were

introduced for detecting morphed face images. The method leverages transferable features derived from pre-trained D-CNN to identify both digitally manipulated and print-scanned altered facial images. Comprehensive testing was conducted. on a newly constructed database containing 352 legit unedited and 431 morphed face images, created from 104 unique subjects. Arya Shah at al. (2024) [17]“Review Paper on Deepfake Video Detection using Neural Network“. approach focuses on detecting deepfakes at the frame-by-frame level, utilizing a ResNext CNN. It extends to video categorization using RNN in conjunction with LSTM algorithm which identify weather the video is legit or fake.

3. Methodology

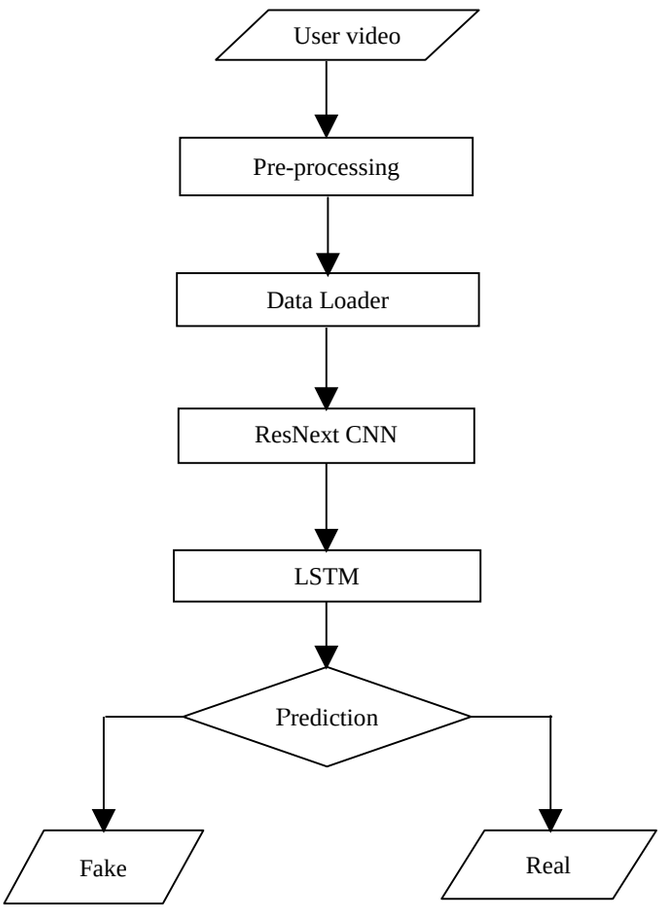

Figure. 2: Testing Work Flow

3.1 Input

As shown in the above work flow the user gives the input i.e deepfake and original video file which are collected from the Kaggle. The videos are fed to the pre-processing in the zip format. and in the next step the video zip file is extracted, in this dataset both real and deepfake videos are preset.

3.2 Pre-processing

In this step, videos undergo preprocessing to remove unnecessary elements and noise. The focus is on detecting and cropping faces from the video. In the first step, the video is partitioned into individual frames. For each frame, the face is recognized and the frame is cropped to include only the face. These trimmed frames are then recombined to form a new video. This procedure is added to all videos, giving a processed dataset containing only face-centric videos. Frames without detectable faces are discarded during initial processing. In the preprocessing “face-recognition” module is applied to identify face in the uploaded video and crop only the face of the person.

3.3 Data Loader

The data loader plays a very major role in the deepfake identification pipeline by managing the initial processing and feeding of video data to the model for learning and evaluation. It is used to load the videos and their labels with a group size of 4 videos at a time and send them to the model by doing. This improves the robustness of the model and the speed of processing the data, the data loader may perform data augmentation. This involves applying various transformations to the frames, like rotations, flips, and color adjustments. Augmentation helps in increasing the diversity of the training data which result in detecting the deepfake video and precision of the model is increased.

Cv2 is Used for video capturing and processing. The cv2.VideoCapture function is used to open videofiles, and to count the number of frames in the“cv2.CAP_PROP_FRAME_COUNT” property is used to collect the count of frames

in each video.

3.4 RestNext CNN

The model combines a CNN and an RNN. We use a pre-trained ResNext CNN model to extract features from each video frame. These appearances are then used by an LSTM network to classify the video as either deepfake or real. The Data Loader helps load and fit the video labels into the model for training. The pre-trained ResNext model is used for feature extraction. ResNext is a type of Residual CNN optimized for deep neural networks. The specific model used is resnext50_32x4d, which has 32x4 dimension setup. For the experiment, the network is finely tuned by adding extra layers and selecting an appropriate learning rate to help the model learn correctly. The 2048-dimensional feature vectors derived from the final pooling layer of ResNext were utilized as input for the LSTM network.

3.4.1 Feature Extraction using CNN

The ResNext50 model is utilized to extract features and precisely detect them in the frame level. The CNN will be fine-tuned by including extra layers and choosing an appropriate training rate to ensure it converges with the gradient. After the final pooling layers, there are 2048-dimensional feature vectors that will be used as input for the sequential LSTM.

3.5 LSTM for sequential data processing

Imagine a 2-layer neural network designed to process a sequence of ResNext CNN feature vectors. from input frames and predicts if the video is a deepfake or unaltered. The main challenge is designing a model that can process this series in an effective way. To address this, a 2048-unit LSTM with a 40% dropout rate is suggested. The LSTM processes the frames in sequence, allowing for temporal analysis by validating the frame at time t-n seconds where n is any frame number preceding t.

3.6 Prediction

The trained model receives a new video for prediction. the system breaks down a new video into individual images, extracts face in every image, and feeds them directly to the model for analysis without saving them first. The trained neural network performs the prediction and returns if the video is a real or fake along with the confidence of the prediction. The prediction of the video depends on the type of the model we have, more the sequence more the accuracy

4. Result

The model's output will indicate whether the video is a deepfake or a genuine video, along with the associated confidence score. This confidence score reflects the model's certainty in its classification decision.

Figure. 3: Model predicting video as fake with the confidence rate.

In the above figure 3 the video is split into frames and identify the face in the video, it has predicted the video is fake with the confidence rate

Figure. 4: Model predicting video as real with the confidence rate.

Here the above figure 4 video has been split into frames and identify the face in the video

using the face recognition model and detects it as fake, the video file should of size 200 M B or less than 200 M B

5. Proposed Method

Model Name	Dataset	Accuracy
model_90_acc_20_frames_FF_data	FaceForensic++	90.9%
Support Vector Machines (SVMs)	NIST MFC2018	70.0%
Support Vector Machines (SVMs)	InterFaceGAN (Generative Adversarial Network),StyleGAN	84.7%

Table. 1: Models accuracy comparison.

Here the model_90_acc_20_frames_FF_data is used which is trained after the initial processing of the datasets, this model utilizes CNN for the feature extraction and LSTM for the processing. The model benifits the 20 frames series length in the video and predicts the output with the precision of 90.9% which is good compare to the SVM model with the precision rate of 70%,the accuracy of the output depends of all the attributes of the model like dataset and the attribute of the datasets.

The GAN (Generative Adversarial Network) which architecture they have used the InterFaceGAN and Style GAN datasets to train the model, the accuracy rate of this architecture is 80%.Considering the different architecture and models to find the deepfake video, the model here that is used to fine the deepfake video is model_90_acc_20_frames_FF_data is excellent at detecting the deepfake with the accuracy rate of 90.9%, here model is not a pre-trained model, which takes lot to time and resource to train this model.

6. Conclusion

Our approach leverages neural networks to classify videos as either deepfakes or genuine, providing associated confidence scores. Inspired by the method deepfakes are generated using GANs and auto encoders, our method performs frame-level detection using a ResNext CNN. Additionally, video classification is achieved using RNN in conjunction with LSTM. We anticipate high accuracy when applied to real-time data.

References

1. AbdElminaam, Diaa S., et al. "DeepFakeDG: A Deep Learning Approach for Deep Fake Detection and Generation." Journal of Computing and Communication 2.2 (2023): 31-37.

2. Afchar, Darius, et al. "Mesonet: a compact facial video forgery detection network." 2018 IEEE international workshop on information forensics and security (WIFS). IEEE, 2018.

3. Arunkumar K L, Ajit Danti, "Recognition of Vehicle using geometrical features of a tail light in the night vision-National Conference on Computation Science and Soft Computing (NCCSSC-2018).

4. Arunkumar K L, Ajit Danti. "A novel approach for vehicle recognition based on the tail lights geometrical features in the night vision", International Journal of Computer Engineering and Applications, Volume XII, Issue I, Jan. 18, www.ijcea.com ISSN 2321-3469.

5. Arunkumar, K. L., Ajit Danti, and H. T. Manjunatha. "Estimation of vehicle distance based on feature points using monocular vision." 2019 International Conference on Data Science and Communication (IconDSC). IEEE, 2019.

6. Arunkumar, K. L., et al. "Classifica-

tion of Vehicle Type on Indian Road Scene Based on Deep Learning." Recent Trends in Image Processing and Pattern Recognition: Third International Conference, RTIP2R 2020, Aurangabad, India, January 3–4, 2020, Revised Selected Papers, Part I 3. Springer Singapore, 2021.

7. Arya Shah, Ashwin Thakur, Atharva Kale, Harsh Bothara, Prof. D. C. Pardeshi. “Review Paper on Deepfake Video Detection using Neural Networks”, ijsrdv, (2024).

8. Aunkumar K L, Ajit Danti, Manjunatha H T. "Classification of Vehicle Make Based on Geometric Features and Appearance-Based Attributes Under Complex Background", Springer 1035 (CCIS), pp 41-48.

9. Kimaya Kulkarni, Sahil Khanolkar, Yash Walke, Rahul Sonkamble. “DeepFake Detection: A survey of countering malicious DeepFakes” IJRASET, (2022).

10. Li, Yuezun, et al. "Celeb-df: A large-scale challenging dataset for deepfake forensics." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.

11. Li, Yuezun, Ming-Ching Chang, and Siwei Lyu. "In ictu oculi: Exposing ai created fake videos by detecting eye blinking." 2018 IEEE International workshop on information forensics and security (WIFS). Ieee, 2018.

12. Manjunatha, H. T., Ajit Danti, and K. L. ArunKumar. "A novel approach for detection and recognition of traffic signs for automatic driver assistance system under cluttered background." Recent Trends in Image Processing and Pattern Recognition: Second International Conference, RTIP2R 2018, Solapur, India, December 21–22, 2018, Revised Selected Papers, Part I 2. Springer Singapore, 2019.

13. Manjunatha, H. T., et al. "Indian Road Lanes Detection Based on Regression and clustering using Video Processing Techniques." Recent Trends in Image Processing and Pattern Recognition: Third International Conference, RTIP2R 2020, Aurangabad, India, January 3–4, 2020, Revised Selected Papers, Part I 3. Springer Singapore, 2021.

14. Nguyen, Huy H., Junichi Yamagishi, and Isao Echizen. "Capsule-forensics: Using capsule networks to detect forged images and videos." ICASSP 2019-2019 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, 2019.

15. Nrupatunga, C. M., and K. L. Arunkumar. "Peruse and Recognition of Old Kannada Stone Inscription Characters." Recent Trends in Image Processing and Pattern Recognition: Third International Conference, RTIP2R 2020, Aurangabad, India, January 3–4, 2020, Revised Selected Papers, Part I 3. Springer Singapore, 2021.

16. Raja, Kiran, Sushma Venkatesh, and R. B. Christoph Busch. "Transferable deep-cnn features for detecting digital and print-scanned morphed face images." Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2017.

17. Sharves, Shobana, Sri Dharrshan, Senthil Kumar. “A Multimodal Approach Harnessing AI to Expose Digital Deceptions” IJNIET, (2024).

18. Wodajo, Deressa, and Solomon Atnafu. "Deepfake video detection us-

ing convolutional vision transformer." arXivpreprint arXiv:2102.11126 (2021).

19. Yang, Xin, Yuezun Li, and Siwei Lyu. "Exposing deep fakes using inconsistent head poses." ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019.