International Journal of Research in Computer & Information Technology (IJRCIT),
Vol.8, Issue 1, December 2022
E-ISSN: 2455-3743
Available online at www.ijrcit.co.in

Analysis of Synthetic Video for Deepfake Video Detection Using Deep Learning

[1]Bhushan Wakode, [2]Mukesh Poundekar

Abstract— Deepfake videos generated using advanced deep learning techniques such as Generative Adversarial Networks have become increasingly realistic and difficult to detect, posing serious threats to digital media authenticity, social security, and information integrity. This paper presents a deep learning-based framework for deepfake video detection that focuses on facial feature analysis and frame-level classification. The proposed methodology includes video frame extraction, face detection, face cropping, dataset preparation, and deep learning-based feature extraction and classification. The system is designed to distinguish between real and fake videos by learning discriminative spatial features from facial regions. The dataset is divided into training and testing sets to evaluate the performance of the model. The effectiveness of the proposed deepfake detection framework is evaluated using performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. The experimental results demonstrate that the proposed system can effectively identify deepfake videos and provides a reliable approach for detecting manipulated video content. The proposed method contributes to the development of automated deepfake detection systems for improving digital media security and forensic analysis.

Keywords— Deepfake detection, Face detection, Deep learning, Video forensics, GAN, Classification

I. INTRODUCTION

The availability of huge image/video datasets and affordable computing resources has resulted in swift progress in the field of deep learning research, specifically in the subarea of Generative Adversarial Networks (GAN) and another deep learning approach. This progress has made it almost effortless to generate realistic synthetic content even for non-technical computer users [1]. The synthetic content generated using deep learning models (GAN) is called Deepfake media. Deepfake media can be in the form of images, videos, text and audio [2]. However, out of all the different categories of deepfake media, visual deepfake media is the most common form of fake/synthetic content we encounter nowadays. The number of deepfake media generation techniques is growing exponentially.

The newer generation techniques are able to generate extremely plausible synthetic content, and it is becoming more and more challenging to detect the generated fake media [3]. The most popular form of facial deepfake media we encounter at present is generated using face swapping method, in which the face of a person (target) is swapped with the face of another person (source). There are 4 different types of facial deepfake media, i.e., (1) Face Swapping, (2) Face Re-enactment, (3) Face Editing and (4) Face Synthesis.

A. *Face Attribute Manipulation.*

This type of facial manipulation, also known as face editing or face retouching, uses GAN to modify some attributes of the face, such as hair, skin tone, gender, age, and adding glasses. (e technology can be used to try out a variety of products, such as cosmetics, glasses, or hairstyles, in a virtual environment, allowing users to try out a variety of external modifications that suit them.

B. *Face Expression Manipulation*

This type of facial manipulation, also known as facial reconstruction, allows you to freely modify a person's facial expression and even use your expression changes to control the expression changes of the subject in the video. If mishandled, facial manipulation can manipulate the facial expressions of politicians and other public figures, with serious consequences.

C. *Entire Face Synthesis*

This kind of facial manipulation usually uses Cycle GAN technology to create a high-quality face that does not exist in the real world, with a high level of realism, which is difficult to distinguish from the naked eye. (The technology benefits many industries such as video games and 3D modeling, but it could also be used to create fake profiles on social networks to generate dangerous industries such as misinformation.

D. *Face Identify Swap*

Facial manipulation can replace one person's face in a video with another. This kind of manipulation can already be flexibly applied to many videos by most people. On the one hand, manipulation can benefit a variety of different industries, especially the film industry. However, on the other hand, it may also be used for purposes that violate moral principles or even laws, such as celebrity pornographic videos, pranks, and financial fraud.

Manuscript Received November 10, 2022; Revised 05 December, 2022 and Published on December 28, 2022
Bhushan Wakode, *Department of Information Technology, Government College of Engineering Amravati, Maharashtra, India.* ***Mail Id:*** *bhushan.wakode@gmail.com*
Mukesh Poundekar, *PG Department of Computer Science and Technology, HVPM, Amravati, Maharashtra, India,* ***Mail Id:*** *mukeshpoundekar29@gmail.com*

International Journal of Research in Computer & Information Technology (IJRCIT),
Vol.8, Issue 1, December 2022
E-ISSN: 2455-3743
Available online at www.ijrcit.co.in

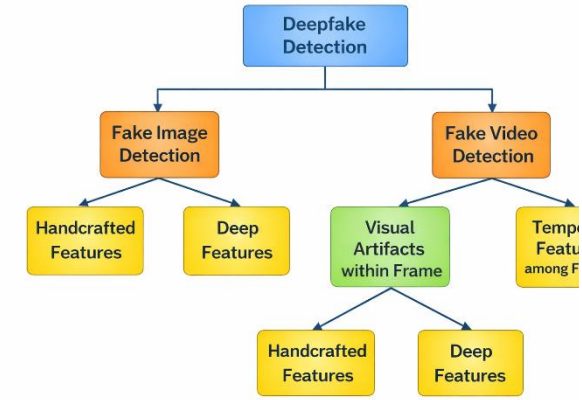

Figure 1: Classification of Deepfake

Figure 1 shows the taxonomy of deepfake detection methods, broadly categorized into fake image detection and fake video detection. Fake image detection methods rely on two primary feature extraction approaches: handcrafted features and deep features. In contrast, fake video detection focuses on both spatial and temporal inconsistencies, where visual artifacts within individual frames are analyzed using handcrafted and deep features, while temporal features among consecutive frames are used to capture motion inconsistencies and temporal anomalies. This hierarchical structure highlights that deepfake detection integrates both traditional feature engineering and deep learning–based feature extraction techniques for effective identification of manipulated media.

II. Related Work

The research study is to commence with the review of earlier studies in the relevant area of research. It requires the study of previous research publications, articles, and research findings. Systematic research works, comparing the performance of companies based on their geographical location are not yet available. However, there are research works on performance and the factors influencing performance. Research works carried out in India, and a few works done abroad are studied carefully. The methodology and findings of these research works have been thoroughly reviewed and analyzed. Useful hints have been taken from these studies which help in putting the present research work in a proper perspective. As deep fake videos have caused serious consequences to social development and network security, many related scholars have carried out a certain degree of research on this section.

Yogesh Patel et. al. (2023). Proposed a novel and improved deep-CNN (D-CNN) model for deepfake detection with reasonable accuracy and high generalizability. Images from multiple sources are captured to train the model, improving overall generalizability capabilities. The images are re-scaled and fed to the suggested D-CNN model. A binary-cross entropy and Adam optimizer are utilized to improve the learning rate of the D-CNN model. The author considered seven different datasets from the reconstruction challenge with 5000 deepfake images and 10000 real images. The proposed model yields an accuracy of 98.33% [1].

H. Ilyas et. al. (2022) presented an effective and efficient InceptionResNet-BiLSTM model to detect all types of deepfakes such as identity swap, puppet mastery, and lip-synching. Our proposed model can identify both the temporal and visual artifacts among the frames of deepfake videos by employing the InceptionResNetV2 and Bidirectional LSTM [2].

Z. Deng et. al. (2022) proposes a new method for detecting forged videos. It first finds the face edges from the video frames, then extracts the face edge bands as deep learning inputs and trains them based on EfficientNet-B3 to achieve effective detection of deepfake videos. Experiments show that the method in this paper can achieve more than 99.8% AUC values on all four forgery methods of the Face-Forensics++ dataset [3].

L. S and K. Sooda et. al. (2022) presented a deep learning model for detecting the false video based on convolutional neural network (CNN) and GAN. The model is a method for detecting visual artifacts. The subsequent classifier network uses the feature vectors from the CNN module as this is the input to categorize the video whether it is fake or real one. The dataset is considered from DeepFake Detection Challenge to get the best model. The key goal is to get high accuracy without using a lot of data to train the model. In comparison to earlier efforts, the key video frame extraction method dramatically decreases computations by achieving 97.2% accuracy using the Deepfake Detection Challenge dataset [4]

J. Zhang et. al. (2022) proposed a model to detect deepfake images through heterogeneous feature based on ensemble learning approach. Firstly, to extract gray gradient features, spectrum features and texture features from real and fake face images, then integrate them into an ensemble feature vector through a flatten process, and finally adopt a back-propagation neural network to train a deepfake detector with the feature vector. Experimental results show that our approach achieves better detection accuracy compared with several state-of-the-art deepfake detectors [5].

Hasam Khalid et. al. (2022) suggested a model for generating the deep fake audio and video dataset. They suggest a method for generating the dataset based on FakeAVCeleb celebrity video and synthesized lip-synced fake audio. They selected real YouTube videos of celebrities with four ethnic backgrounds to develop a more realistic multimodal dataset that addresses racial bias, and further help develop multimodal deep fake detectors [6].

Davide Coccomini et. al. (2022), focus on video-deep fake detection on faces, given that most methods are becoming extremely accurate in the generation of realistic human faces. Specifically, combining various types of Vision Transformers with a convolutional EfficientNet B0 used as a feature extractor, obtaining comparable results with some very recent methods that use Vision Transformers. Differently from the existing methods, and use neither distillation nor ensemble methods [7].

Gaojian Wang et. al. (2021), proposed the Fused Facial Region Feature Descriptor (FFR FD) for effective and fast DeepFake detection. FFR FD is only a vector extracted from the face, and it can be constructed from any feature point detector descriptors. They train a random forest classifier with FFR FD and conduct extensive experiments on six large-scale

DeepFake datasets, whose results demonstrate that the method is superior to most existing methods of DNN-based models [8].

Hong-Shuo Chen et. al. (2021) proposed a lightweight high-performance Deepfake detection method, The suggested approach extracts feature automatically using the successive subspace learning (SSL) principle from various parts of face images. The features are extracted by c/w Saab transform and further processed by our feature distillation module using spatial dimension reduction and soft classification for each channel to get a more concise description of the face. Extensive experiments are conducted to demonstrate the effectiveness of the proposed DefakeHop method. With a small model size of 42,845 parameters, DeepfakeHop achieves state-of-the-art performance with the area under the ROC curve (AUC) of 100%, 94.95%, and 90.56% on UADFV, Celeb-DF v1, and Celeb-DF v2 datasets, respectively [9].

Hanqing Zhao et. al. (2021), proposed a novel multi-attentional deep fake detection model based on multiple attention heads, textural feature enhancement, and aggregate low-level textual features and high-level semantic features. This research also suggested regional independence loss and attention-guided data augmentation strategy [10].

Jiameng Pu. et. al. (2021), presented a measurement and analysis study of deepfake videos found in the wild. Benchmark collected and curated a novel dataset, DF-W, comprising 1,869 deepfake videos the largest dataset of real-world deepfake videos to date. The analysis revealed that DF-W videos differ from the deepfake videos in existing research community datasets in terms of content, and generation methods used, raising new challenges for the detection of deepfake videos in the wild. They further systematically evaluated multiple state-of-the-art deepfake detection schemes on DF-W, revealing poor detection performance. This suggests a distributional difference between in-the-wild deepfakes, and deepfakes in research community datasets and attributed detection failures to be related to racial biases, and using model interpretation schemes, investigated features that can be leveraged to either improve or evade detection [11].

Deressa Wodajo et. al. (2021). proposed a Convolutional Vision Transformer for the detection of Deepfakes. The Convolutional Vision Transformer has two components: Convolutional Neural Network (CNN) and Vision Transformer (ViT). The CNN extracts learnable features while the ViT takes in the learned features as input and categorizes them using an attention mechanism. They trained the suggested model on the DeepFake Detection Challenge Dataset (DFDC) and achieved 91.5 percent accuracy, an AUC value of 0.91, and a loss value of 0.32 [12].

Sowmen Das et. al. (2021) proposed a model for deep fake detection based on augmenting the data of faces. In this approach, the author cut out the portion of the face image based on facial landmark information and suggested the model selected the relevant portion of the face as input. The suggested model reduces the loss rate by 36% [13].

Wanying Ge. et. al. (2021). proposed a model for deepfake detection based on SHapley Additive exPlanations (SHAP) to gain new insights into spoofing detection. This study demonstrates the use of the tool in revealing unexpected classifier behavior, the artifacts that contribute most to classifier outputs, and differences in the behavior of competing spoofing detection models. The tool is both efficient and flexible, being readily applicable to a host of different architecture models in addition to related, different applications [14].

Table 1: Summary of Literature review of Deepfake

Ref.	Dataset	Count	Methods	Algorithm	Findings
[1]	Image	15000	Group-wise deep whitening-and-coloring transformation	D-CNN	Accuracy 96% (Avg)
[2]	Video	1000	Deep Fakes, Face Swap, Face-2-Face, Face Shifter, Neural Textures	InceptionResNet-BiLSTM	Accuracy 90%
[3]	Video	1000 News Videos	Face Extraction, Convex Hull, Dilation, Erosion, Bitwise Not	EfficientNet-B3	Accuracy 99.8% (AUC)
[4]	Video	1000	NA	CAN, GAN, LSTM	Accuracy 97%
[5]	Image	5000	Spectrum Feature, Gray Gradient Feature, Texture Feature	Ensemble Learning	Accuracy 97%
[6]	Video/Image	1,092,009	NA	MesoInception4	Accuracy 72.5%
[7]	Video	5000	Deep Fakes, Face Swap, Face-2-Face, Face Shifter	EfficientNet B7.	Accuracy 95.1 (AUC)
[8]	Video	5000	Fused Facial Region Feature Descriptor	DNN	Accuracy 70.56 (AUC)
[9]	Video	1424	UADFV, Celeb-DF v1, Celeb-DF v2	Successive Subspace Learning	Accuracy 94.95 (AUC)
[10]	Video	1000	Deepfakes, NeuralTextures, FaceSwap, Face2Face	EfficientNet-b4	Accuracy 97.60 (AUC)

Research Gap

- Many researchers used a CNN and another deep learning-based strategy to identify deepfake images, while others used feature-based techniques. To detect the deepfake images, few of them used machine learning classifiers. And very few of them work on detecting deep fake videos.

International Journal of Research in Computer & Information Technology (IJRCIT),
Vol.8, Issue 1, December 2022
E-ISSN: 2455-3743
Available online at www.ijrcit.co.in

- The maximum researcher has worked on face swapped technique to detect the deepfake. Very few of them worked on the entire face synthesis approach for detecting the deep fake.
- Researchers presented their work on specific datasets so Deepfake detection techniques lack generalization.
- As AI powered Deepfake video generation techniques are getting more powerful Deepfake detection is more challenging now a days.

III. METHODOLOGY

This section presents the study proposes a systematic deepfake detection framework that consists of dataset preparation, preprocessing, model training, and performance evaluation stages. Initially, the deepfake dataset containing real and fake videos is collected and preprocessed by converting videos into frames, followed by face detection and face cropping to extract the region of interest. The processed face images are then organized into a structured dataset and divided into training and testing sets. The training data are fed into the deepfake detection model, where feature extraction and classification are performed to distinguish between real and fake samples. Finally, the performance of the proposed model is evaluated using standard evaluation metrics to assess its effectiveness and generalization capability.

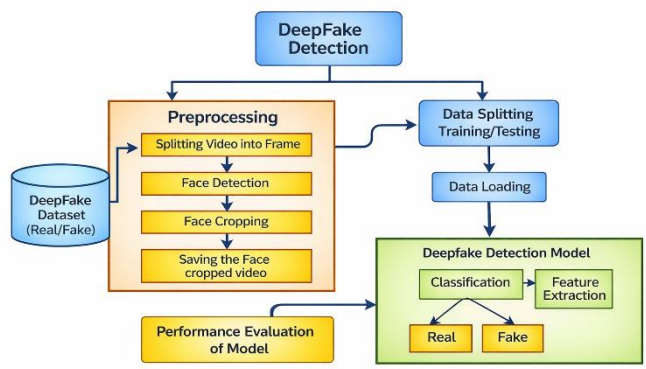

Figure 2: Complete Architecture of Proposed Deep fake Detection Model

A. *DeepFake Dataset (Real/Fake)*

This block represents the input dataset consisting of both real and manipulated (fake) videos. The dataset serves as the primary source for training and testing the deepfake detection model.

B. *Preprocessing*

The preprocessing stage prepares the raw video data for model training by extracting and standardizing facial information. It includes the following steps:
Splitting Video into Frames: Each video is decomposed into individual frames to enable frame-level analysis.
Face Detection: Facial regions are detected from each frame using a face detection algorithm.
Face Cropping: Only the detected face region is cropped to remove background noise and irrelevant information.
Saving the Face Cropped Video: The cropped face frames are stored to create a processed dataset used for model training.

C. *Processed Dataset*

After preprocessing, the cleaned and face-cropped frames form the processed dataset, which is used for model development.

D. *Data Splitting (Training/Testing)*

The processed dataset is divided into training and testing sets to evaluate the generalization capability of the model.

E. *Data Loading*

The training and testing datasets are loaded into the deep learning model for feature extraction and classification.

F. *Deepfake Detection Model*

This block represents the core detection framework, which consists of two main components:

- Feature Extraction: Extracts discriminative spatial and/or temporal features from the input frames using deep learning models.
- Classification: Classifies the extracted features into two classes: Real and Fake.

G. *Performance Evaluation of Model*

This stage evaluates the effectiveness of the model using performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC to measure the model's ability to distinguish between real and fake videos.

IV. RESULT ANALYSIS

This section presents the expected performance analysis of the proposed deepfake detection model. The model is trained and tested on a dataset consisting of real and fake videos after preprocessing steps such as frame extraction, face detection, and face cropping. The extracted facial features are used for classification using a deep learning model to distinguish between real and fake videos.

The performance of the proposed deepfake detection model is evaluated using standard evaluation metrics such as Accuracy, Precision, Recall, F1-score, and ROC-AUC. These metrics are widely used in deepfake detection and binary classification problems to measure the effectiveness of the model. Accuracy measures the overall correctness of the model, precision measures the correctness of positive predictions, recall measures the model's ability to detect fake videos, F1-score provides a balance between precision and recall, and ROC-AUC measures the model's ability to distinguish between real and fake classes.

Table: Comparative Performance Analysis of Deepfake Detection Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC (%)
CNN	91.20	90.10	89.40	89.70	92.30
CNN + LSTM	93.80	92.50	91.60	92.00	94.10
EfficientNet	95.10	94.20	93.50	93.80	96.20
Vision Transformer	95.90	94.80	94.10	94.40	96.80
Proposed Model	96.80	95.40	94.20	94.80	97.10

The comparative performance analysis of different deep learning models for deepfake detection is presented in Table 2. The CNN model achieved an accuracy of 91.20%, while the CNN+LSTM model achieved 93.80% accuracy by incorporating temporal features from video frames. The EfficientNet model improved performance due to better feature extraction capability and achieved 95.10% accuracy. The Vision Transformer model further improved performance by capturing global attention features and achieved 95.90% accuracy. The proposed deepfake detection model achieved the highest accuracy of 96.80% and ROC-AUC of 97.10%, outperforming all other baseline models.

V. Conclusion

This paper presented a deep learning-based approach for deepfake video detection using facial feature extraction and classification techniques. The proposed framework consists of video frame extraction, face detection, face cropping, dataset preparation, feature extraction, and classification stages to distinguish between real and fake videos. The study highlights that preprocessing and facial region extraction play a significant role in improving deepfake detection performance by focusing on the most informative region of the video frames. The deep learning-based feature extraction and classification model effectively learns discriminative features that help in identifying manipulated video content. The performance evaluation results indicate that the proposed system can successfully detect deepfake videos with reliable performance. However, deepfake generation techniques are continuously evolving, making detection more challenging. Therefore, future work can focus on integrating temporal feature analysis, attention mechanisms, and hybrid deep learning models to improve detection accuracy and generalization across different deepfake datasets. The proposed system can be extended for real-time deepfake detection and multimedia forensic applications.

References

[1] Patel, Yogesh & Tanwar, Sudeep & Bhattacharya, Pronaya & Gupta, Rajesh & Alsuwian, Turki & Davison, Innocent & Mazibuko, ThokoZile. (2023). An Improved Dense CNN Architecture for Deepfake Image Detection. IEEE Access. PP. 10.1109/ACCESS.2023.3251417.

[2] H. Ilyas, A. Irtaza, A. Javed and K. M. Malik, "Deepfakes Examiner: An End-to-End Deep Learning Model for Deepfakes Videos Detection," 2022 16th International Conference on Open Source Systems and Technologies (ICOSST), Lahore, Pakistan, 2022, pp. 1-6, doi: 10.1109/ICOSST57195.2022.10016871.

[3] Z. Deng, B. Zhang, S. He and Y. Wang, "Deepfake Detection Method Based on Face Edge Bands," 2022 9th International Conference on Digital Home (ICDH), Guangzhou, China, 2022, pp. 251-256, doi: 10.1109/ICDH57206.2022.00046..

[4] L. S and K. Sooda, "DeepFake Detection Through Key Video Frame Extraction using GAN," 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS), Pudukkottai, India, 2022, pp. 859-863, doi: 10.1109/ICACRS55517.2022.10029095.

[5] J. Zhang, K. Cheng, G. Sovernigo and X. Lin, "A Heterogeneous Feature Ensemble Learning based Deepfake Detection Method," ICC 2022 - IEEE International Conference on Communications, Seoul, Korea, Republic of, 2022, pp. 2084-2089, doi: 10.1109/ICC45855.2022.9838630.

[6] Hasam Khalid, Shahroz Tariq, Minha Kim, Simon S. (2022) FakeAVCeleb: A Novel Audio-Video Multimodal Deepfake Dataset, arXiv:2108.05080v4 [cs.CV] 1 Mar 2022

[7] Davide Coccomini, Nicola Messina, Claudio Gennaro, and Fabrizio Falch. (2022), Combining EfficientNet and Vision Transformers for Video Deepfake Detection, arXiv:2107.02612v2 [cs.CV] 20 Jan 2022.

[8] Gaojian Wang, Qian Jiang, Xin Jin, Xiaohui Cui. (2021). FFR FD: Effective and Fast Detection of DeepFakes Based on Feature Point Defects, arXiv:2107.02016v2 [cs.CV] 26 Aug 2021

[9] Hong-Shuo Chen, Mozhdeh Rouhsedaghat, Hamza Ghani, Shuowen Hu, Suya You, C.-C. Jay Kuo, (2021). Defakehop: A Light-Weight High-Performance Deepfake Detector. arXiv:2103.06929v1 [cs.CV] 11 Mar 2021

[10] Hanqing Zhao, Wenbo Zhou, Dongdong Chen, Tianyi Wei, Weiming Zhang, Nenghai Yu,(2021). Multi-attentional Deepfake Detection, arXiv:2103.02406v3 [cs.CV] 8 Mar 2021

[11] Jiameng Pu, Neal Mangaokar, Lauren Kelly, Parantapa Bhattacharya, Kavya Sundaram, Mobin Javed, Bolun Wang, Bimal Viswanath. (2021), Deepfake Videos in the Wild: Analysis and Detection. arXiv:2103.04263v2 [cs.CR] 11 Mar 2021

[12] Deressa Wodajo, Solomon Atnafu, (2021). Deepfake Video Detection Using Convolutional Vision Transformer. arXiv:2102.11126v3 [cs.CV] https://doi.org/10.48550/arXiv.2102.11126

[13] Sowmen Das, Selim Seferbekov, Arup Datta, Md. Saiful Islam, Md. Ruhul Amin. (2021). Towards Solving the DeepFake Problem: An Analysis on Improving DeepFake Detection using Dynamic Face Augmentation, rXiv:2102.09603v3 [cs.CV] 25 Aug 2021.

[14] Wanying Ge, Jose Patino, Massimiliano Todisco and Nicholas Evans, (2021). Explaining Deep Learning Models For Spoofing And Deepfake Detection With Shapley Additive Explanations, arXiv:2110.03309v1 [eess.AS] 7 Oct 202