Deep Fake Image/Video detection Using Deep Learning

Kollati Kasi
PG scholar, Department of MCA, CDNR collage, Bhimavaram, Andhra Pradesh.
A.Durga Devi
(Assistant Professor), Master of Computer Applications, DNR collage, Bhimavaram, Andhra Pradesh.

ABSTRACT

The proliferation of deepfake technology, which uses artificial intelligence to create highly realistic synthetic videos and images, poses significant threats to privacy, security, and trust in digital media. Traditional methods for detecting these manipulations often fall short due to the sophisticated nature of deepfake algorithms. This paper proposes a novel approach for deepfake face detection using Deep Learning (DL) well-suited for sequential data analysis. Our method leverages the temporal dependencies and patterns inherent in video sequences to identify subtle inconsistencies and artifacts introduced by deepfake generation processes. By analyzing frames in a sequence rather than in isolation, the DL can capture dynamic facial features and movements that are difficult to replicate accurately in deepfakes. The proposed model is trained on a comprehensive dataset of real and deepfake videos, incorporating various scenarios and levels of manipulation. Experimental results demonstrate that our DL-based approach achieves superior accuracy and robustness compared to state-of-the-art deepfake detection techniques, particularly in challenging cases with high-quality deepfakes. Furthermore, the model exhibits strong generalization capabilities across different datasets and deepfake generation methods. This research highlights the potential of DL for enhancing the detection of deepfake content, contributing to the development of more secure and trustworthy digital media platforms.

INTRODUCTION

Temporal Analysis of Video Sequences: To utilize DL for analyzing the sequential frames in videos, capturing dynamic facial features and movements that are challenging to replicate accurately in deepfakes. Enhancement of Detection Accuracy: To improve the accuracy and robustness of deepfake detection methods by leveraging the capabilities of DL in identifying subtle inconsistencies and artifacts introduced by deepfake generation processes.

Comprehensive Dataset Utilization: To train and validate the LSTM-based model on a diverse and extensive dataset of real and deepfake videos, ensuring the model's effectiveness across various scenarios and manipulation levels. Generalization Across Datasets: To ensure the model's strong generalization capabilities by testing its performance on different datasets and deepfake generation methods, demonstrating its applicability in real-world scenarios.

Contributing to Digital Media Security: To enhance the security and trustworthiness of digital media platforms by providing a reliable tool for detecting and mitigating the spread of deepfake content.

Data Collection and Preprocessing: Compilation of a comprehensive dataset consisting of real and deepfake videos from various sources. Preprocessing of video frames to ensure uniformity in size, format, and quality for effective training and evaluation of the LSTM model. Model Development: Design and implementation of an LSTM network architecture tailored for temporal analysis of video sequences. Integration of additional neural network layers (e.g., convolutional layers) if necessary to enhance feature extraction and improve detection performance.

LITERATURE SURVEY

[1] Nguyen, T.T., Nguyen, Q.V.H., Nguyen, D.T., Nguyen, D.T., Huynh-The, T., Nahavandi, S., Nguyen, T.T., Pham, Q.V. and Nguyen, C.M., 2022. Deep learning for deepfakes creation and detection: A survey. Computer Vision and Image Understanding, 223, p.103525.

Deep learning has been successfully applied to solve various complex problems ranging from big data analytics to computer vision and human-level control. Deep learning advances however have also been employed to create software that can cause threats to privacy, democracy and national security. One of those deep learning-powered applications recently emerged is deepfake. Deepfake algorithms can create

fake images and videos that humans cannot distinguish them from authentic ones.

[2] Westerlund, M., 2019. The emergence of deepfake technology: A review. Technology innovation management review, 9(11).

Novel digital technologies make it increasingly difficult to distinguish between real and fake media. One of the most recent developments contributing to the problem is the emergence of deepfakes which are hyper-realistic videos that apply artificial intelligence (AI) to depict someone say and do things that never happened. Coupled with the reach and speed of social media, convincing deepfakes can quickly reach millions of people and have negative impacts on our society.

While scholarly research on the topic is sparse, this study analyzes 84 publicly available online news articles to examine what deepfakes are and who produces them, what the benefits and threats of deepfake technology are, what examples of deepfakes there are, and how to combat deepfakes. The results suggest that while deepfakes are a significant threat to our society, political system and business, they can be combatted via legislation and regulation, corporate policies and voluntary action, education and training, as well as the development of technology for deepfake detection, content authentication, and deepfake prevention.

[4] Indolia, S., Goswami, A.K., Mishra, S.P. and Asopa, P., 2018. Conceptual understanding of convolutional neural network-a deep learning approach. Procedia computer science, 132, pp.679-688.

Deep learning has become an area of interest to the researchers in the past few years. Convolutional Neural Network (CNN) is a deep learning approach that is widely used for solving complex problems. It overcomes the limitations of traditional machine learning approaches. The motivation of this study is to provide the knowledge and understanding about various aspects of CNN. This study provides the conceptual understanding of CNN along with its three most common architectures, and learning algorithms. This study will help researchers to have a broad comprehension of CNN and motivate them to venture in this field. This study will be a resource and quick reference for those who are interested in this field.

With the rapidly growing demand for learnable machines for solving many complex problems, deep learning has evolved itself as an area of interest to the researchers in the past few years. As researchers tend to mimic human behavior, a major question arises that how do the humans acquire knowledge? The answer to this question is an essential ability of humans i.e. learning, which needs to be incorporated in machines, hence the term machine learning was coined.

PROPOSED METHOD

The requirements gathering process takes as its input the goals identified in the high-level requirements section of the project plan. Each goal will be refined into a set of one or more requirements. These requirements define the major functions of the intended application, define operational data areas and reference data areas, and define the initial data entities. Major functions include critical processes to be managed, as well as mission critical inputs, outputs and reports. A user class hierarchy is developed and associated with these major functions, data areas, and data entities. Each of these definitions is termed a Requirement. Requirements are identified by unique requirement identifiers and, at minimum, contain a requirement title and textual description.

These requirements are fully described in the primary deliverables for this stage: the Requirements Document and the Requirements Traceability Matrix (RTM). The requirements document contains complete descriptions of each requirement, including diagrams and references to external documents as necessary. Note that detailed listings of database tables and fields are *not* included in the requirements document.

The title of each requirement is also placed into the first version of the RTM, along with the title of each goal from the project plan. The purpose of the RTM is to show that the product components developed during each stage of the software development lifecycle are formally connected to the components developed in prior stages.

In the requirements stage, the RTM consists of a list of high-level requirements, or goals, by title, with a listing of associated requirements for each goal, listed by requirement title. In this hierarchical listing, the RTM shows that each requirement developed during this stage is formally linked to a specific product goal. In this format, each requirement can be traced to a specific product goal, hence the term requirements traceability.

The outputs of the requirements definition stage include the requirements document, the RTM, and an updated project plan.

The most critical section of the project plan is a listing of high-level product requirements, also referred to as

goals. All of the software product requirements to be developed during the requirements definition stage flow from one or more of these goals. The minimum information for each goal consists of a title and textual description, although additional information and references to external documents may be included.

RESULT

In this project DL algorithm is used to detect Deepfake faces detection from video input.

To train above model we have used Deepfake faces dataset from KAGGLE repository which contains more than 95000 images and this dataset can be downloaded from below URL

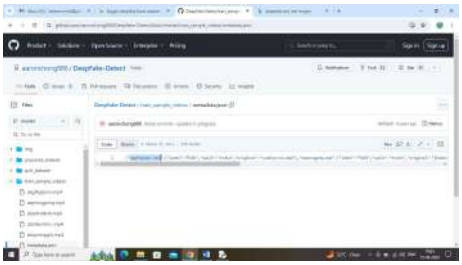
https://www.kaggle.com/datasets/dagnelies/deepfake-faces

Above dataset contains two different class labels such as Fake and Real

To detect Deepfake faces we have downloaded videos from below URL

https://github.com/aaronchong888/DeepFake-Detect/blob/master/train_sample_videos/metadata.json

From above URL file we can see fake and real videos which we downloaded and tested with our model



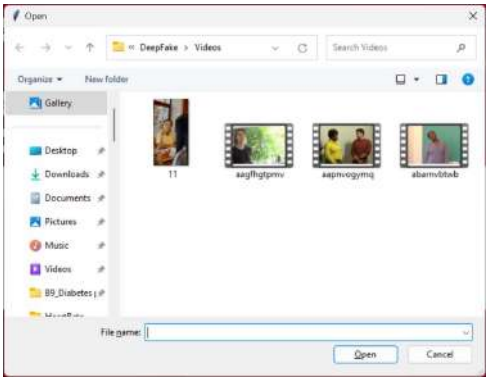
In above URL screen you can see 'aagfhgtpmv.mp4' video is fake and 'abarnvbtwb.mp4' is the real video and this model is successfully predicting this videos.

To implement this project we have designed following modules

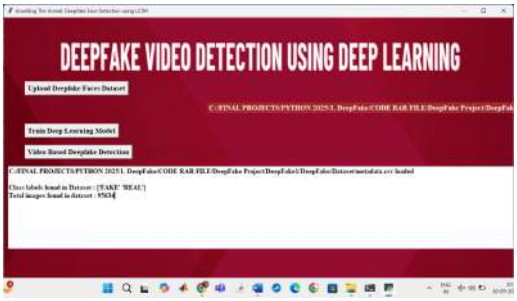
1) Upload Deepfake Faces Dataset: using this module will upload dataset images to application and then application will read all images and then resize to equal sizes and then creating X and Y training array
2) Train DL Model: this module will shuffle, normalize and then split all images into 80:20 percent train and test ratio. 80% images will be input to DL algorithm to train a model and this model will be applied on 20% test data to calculate prediction accuracy
3) Video Based Deepfake Detection: using this module will upload test video and then DL model will analyse faces from each frame slowly and then predict video as Real or Deepfake. Once after prediction will get video playing output with result as fake or real.

To run project double click on run.bat file to get below screen

In above screen click on 'Upload Deepfake Faces Dataset' button to load dataset and get below page

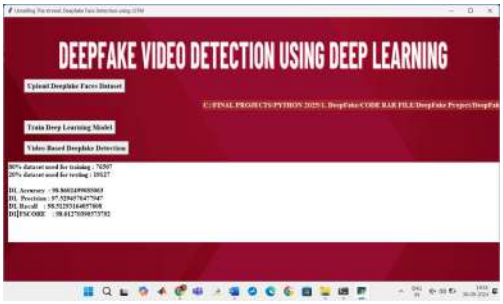


In above screen selecting and uploading dataset annotation file and then click on 'Open' button to get below output

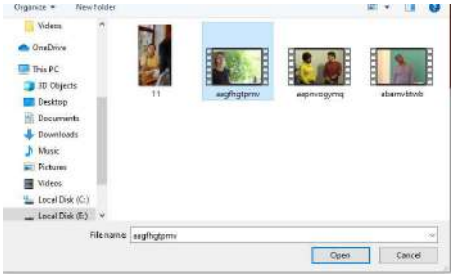


In above screen can see dataset loaded and can see different class labels found in dataset and then can see number of images

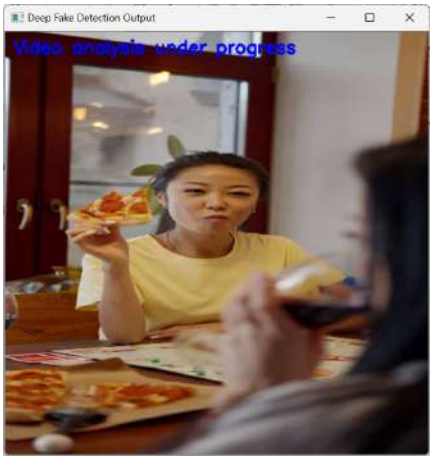
found in dataset and now click on 'Train DL Model' button to train algorithm and get below page



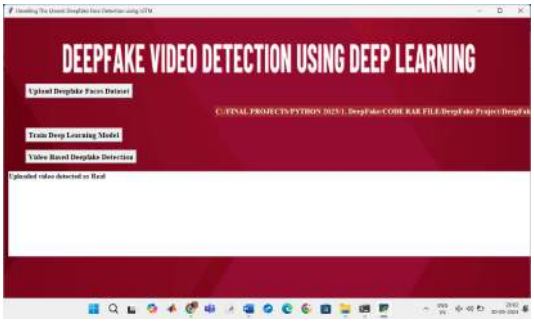
In above screen can see train and test dataset size and then can see DL got 99% accuracy and can see other metrics like precision, recall and FSCORE. Now click on 'Video Based Deepfake Detection' button to upload test video and get below page

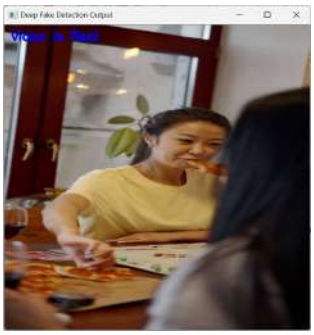


In above screen selecting and uploading 11.mp4 video and then click on 'Open' button to start analysing video and get below page

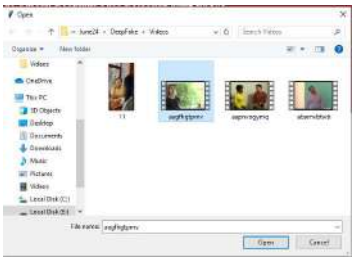


In above screen in blue colour text can see video analysis started and after thorough analysis by DL will get below output

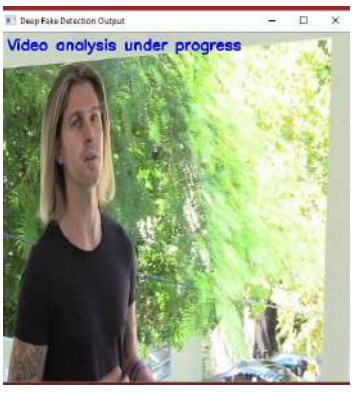




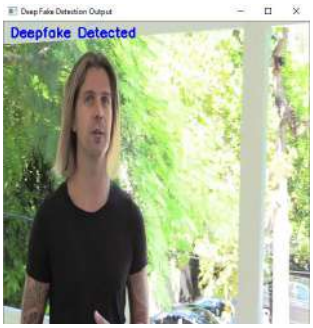
In above screen uploaded video predicted as 'Real' and now test with other fake video



In above screen uploading 'aagfhgtpmv.mp4' video and then click on 'Open' button to load video and get below output



In above screen video analysis under progress and once after analysis will get below output



In above screen uploaded video is detected as Deepfake and similarly you can upload and test other video

CONCLUSION

The conclusion of deep fake face detection from videos using deep learning (DL) emphasizes the importance of leveraging advanced neural networks to detect subtle facial manipulations. Deep learning models, such as CNNs and RNNs, have shown promising results in accurately identifying fake faces by analyzing temporal inconsistencies, texture anomalies, and facial landmarks across frames. These approaches help mitigate the spread of misinformation, improve security in digital platforms, and enable automated real-time detection, proving essential in combating the evolving threats posed by deep fakes.

REFERENCE

[1] Nguyen, T.T., Nguyen, Q.V.H., Nguyen, D.T., Nguyen, D.T., Huynh-The, T., Nahavandi, S., Nguyen, T.T., Pham, Q.V. and Nguyen, C.M., 2022. Deep learning for deepfakes creation and detection: A survey. Computer Vision and Image Understanding, 223, p.103525.

[2] Westerlund, M., 2019. The emergence of deepfake technology: A review. Technology innovation management review, 9(11).

[3] Thippanna, G., Priya, M.D. and Srinivas, T.A.S., An Effective Analysis of Image Processing with Deep Learning Algorithms. International Journal of Computer Applications, 975, p.8887.

[4] Indolia, S., Goswami, A.K., Mishra, S.P. and Asopa, P., 2018. Conceptual understanding of convolutional neural network-a deep learning approach. Procedia computer science, 132, pp.679-688.

[5] Ralf C. Staudemeyer, "Understanding LSTM – a tutorial into Long Short-Term Memory Recurrent Neural Networks", arXiv:1909.09586v1 [cs.NE] 12 Sep 2019

[6] Güera, D. and Delp, E.J., 2018, November. Deepfake video detection using recurrent neural networks. In 2018 15th IEEE international conference on advanced video and signal based surveillance (AVSS) (pp. 1-6). IEEE.

[7] Mallet, J., Dave, R., Seliya, N. and Vanamala, M., 2022, November. Using deep learning to detecting deepfakes. In 2022 9th International Conference on Soft Computing & Machine Intelligence (ISCMI) (pp. 1-5). IEEE.

[8] Abir, W.H., Khanam, F.R., Alam, K.N., Hadjouni, M., Elmannai, H., Bourouis, S., Dey, R. and Khan, M.M., 2023. Detecting Deepfake Images Using Deep Learning Techniques and Explainable AI Methods. Intelligent Automation & Soft Computing., pp.2151-2169.

[9] Gong, D., Kumar, Y.J., Goh, O.S., Ye, Z. and Chi, W., 2021. DeepfakeNet, an efficient deepfake detection method. International Journal of Advanced Computer Science and Applications, 12(6).

[10] Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J. and Nießner, M., 2019. Faceforensics++: Learning to detect manipulated facial images. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 1-11).

[11] DFDC data from Kaggle:- https://www.kaggle.com/competitions/deepfake-detection-challenge (Accessed on 13/09/2023)

[12] Li, Y., Yang, X., Sun, P., Qi, H. and Lyu, S., 2020. Celeb-df: A large-scale challenging dataset for deepfake forensics. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 3207-3216).