ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940
www.ijsem.org
Vol 18, Issue.1 March 2024

AN INTELLIGENT FRAMEWORK FOR DEEPFAKE VIDEO DETECTION USING DEEP CNN

Aerra Pranathi[1], Avnoori Sunny[2], Gudi Krishna[3],Dr.M.V.Kamal[4]

[1,2,3]B.Tech Student, Department of CSE (Data Science), Malla Reddy College of Engineering and Technology, Hyderabad, India.

[4]Professor & HOD, Department of CSE (Data Science), Malla Reddy College of Engineering and Technology, Hyderabad, India.

***Abstract*—** The rise of deepfakes, enabled by advanced neural networks, threatens digital information integrity, making detection methodologies crucial. This paper presents a new framework for deepfake video detection using Deep Convolutional Neural Networks (CNNs). Leveraging the InceptionV3 architecture for feature extraction from videos, combined with Gated Recurrent Units (GRUs) for sequential analysis, our approach aims to discern genuine content from manipulations. Evaluation on the Deepfake Detection Challenge dataset from Kaggle revealed the model's potential, but also highlighted challenges like class imbalances. Through this research, we aim to bolster the defence against digital misinformation introduced by deepfakes.

Keywords— Deepfake, Convolutional Neural Networks (CNN), InceptionV3, Gated Recurrent Units (GRU), Video detection, Digital misinformation

I. INTRODUCTION

In today's digital world, new tools allow us to edit videos in ways we could only dream of before. One such tool is "deepfakes", where videos are changed to make it look like someone said or did something they didn't. This technology, although fascinating in its capabilities, presents a serious threat to the trustworthiness of digital media, enabling malicious actors to fabricate content for misinformation, identity theft, and even political subversion. While it's a cool technology, it can also be misused to spread fake news or deceive people. Deepfakes use a type of computer program called Deep Convolutional Neural Networks (CNN) to make these edits. As these fake videos get better and harder to spot, we need our own tools to detect them and tell them apart from real videos. This research aims to build a system that can spot these deepfake videos. By using a program model called InceptionV3 and another tool, Gated Recurrent Units (GRU), we hope to create a reliable way to find and flag these fake videos. Let's explore how these deepfakes work and how our new system aims to detect them.

II.METHODS

A. Data Preparation and Loading

ISSN2454-9940
www.ijsem.org
Vol 18, Issuse.1 March 2024

The data utilized for this research were sourced from the DeepFake Detection Challenge dataset hosted on Kaggle. The dataset was partitioned into training and test sets. Metadata accompanying the dataset was parsed to segregate real and manipulated videos. Using pandas, the metadata was further structured for ease of processing.

B. Feature Extraction Using InceptionV3

For the task of feature extraction from individual video frames, the pre-trained InceptionV3 model, which is renowned for its high efficacy in image classification tasks, was utilized. This model, embedded with weights trained on the ImageNet dataset, was repurposed to function as a feature extractor by removing its final classification layer. For a given frame $Frame_t$ of a video, the feature vector F_t was computed as:

$$F_t = InceptionV3(Frame_t)$$

Here, $Frame_t$ denotes the t-th frame in the video, and F_t is its corresponding feature vector.

C. Sequence Modeling with GRU Layers

Videos inherently possess a temporal dimension, characterized by the sequence of frames. To exploit this temporal dimension for deepfake detection, we employed Gated Recurrent Units (GRU) - a type of recurrent neural network. After extracting frame-level features using InceptionV3, these were passed through GRU layers to capture time-based dependencies:

$$h_t = GRU(F_t, h_{t-1})$$

Here, h_t is the hidden state at time t, produced by considering the current feature F_t and the previous hidden state h_{t-1}.

D. Model Training and Validation

With the architecture in place, the model was trained using the training set and validated on a validation set. During training, checkpoints were meticulously maintained to ensure only the best-performing model weights were retained. The loss function employed was binary crossentropy, suitable for the binary classification task at hand. Optimization was achieved using the Adam optimizer.

E. Prediction Mechanism

Post-training, the model was tasked with predicting the genuineness of videos. The prediction hinged on analyzing the feature vectors of video frames and their temporal inter-relations. Videos were deemed "Fake" or "Real" based on a threshold set on the model's sigmoid output.

III. EXPERIMENTAL SETUP

IV. TO EVALUATE THE PERFORMANCE AND EFFICACY OF OUR DEEPFAKE DETECTION FRAMEWORK, WE DESIGNED A COMPREHENSIVE EXPERIMENTAL SETUP.

A. Dataset:

Source: Our experiments leveraged the Deepfake Detection Challenge (DFDC) dataset, available on Kaggle. This dataset contains a vast collection of manipulated videos and their pristine counterparts.

ISSN2454-9940
www.ijsem.org
Vol 18, Issue.1 March 2024

Training Set: A random 90% subset of the videos available in the training directory were used for training our model.

Test Set: The remaining 10% of the videos were reserved for evaluation to ensure a fair assessment of the model's performance.

B. Hardware and Software:

Platform: Our model was trained and evaluated on Google Colab, a cloud-based platform offering GPU acceleration.

GPU: We utilized the NVIDIA Tesla GPUs available on Google Colab for the training process.

Libraries: Our implementation relies on TensorFlow 2.x and Keras, alongside other essential Python libraries like Pandas, Numpy, and OpenCV.

C. Data Preprocessing:

Resizing: Each frame from the videos was resized to a uniform size of 224x224 pixels.

Cropping: Central square cropping was performed on each frame to ensure the focal subject, typically a face, remained the primary feature.

Normalization: The pixel values of each frame were normalized using preprocessing utilities from the InceptionV3 architecture.

D. Model Configuration:

Feature Extractor: We employed the InceptionV3 model, pre-trained on the ImageNet dataset, to act as our primary feature extractor. This model's top classification layer was discarded, and its outputs were used as feature vectors.

Temporal Analysis: Post feature extraction, a Gated Recurrent Unit (GRU) based sequence model was utilized to capture the temporal dependencies across video frames.

Training: The model was trained using the Adam optimizer with binary cross-entropy as the loss function. We used a batch size of 8 and trained the model for 15 epochs.

Evaluation Metric: The primary metric for model evaluation was accuracy, i.e., the proportion of correctly predicted labels (real or fake) over the total number of videos in the test set.

V.RELATED WORK

Deepfake video detection has become a topic of increasing importance as the techniques for

ISSN2454-9940
www.ijsem.org
Vol 18, Issuse.1 March 2024

generating deepfakes continue to advance. The ubiquity and ease of creating deepfake videos with the aid of deep learning models, primarily Generative Adversarial Networks (GANs), have necessitated the development of efficient and accurate detection tools.

A.Traditional Methods: Before the prevalence of deep learning models, video forgery detection was often based on traditional image and video processing techniques. One common approach was to detect inconsistencies in lighting and shadows [1]. Another involved examining artifacts introduced during video compression [2]. However, these methods often fail to catch sophisticated forgeries, especially those created with advanced neural networks.

B.CNN-based Approaches: With the success of Convolutional Neural Networks (CNNs) in image and video analysis, researchers started employing CNNs for deepfake detection. Some approaches use pre-trained models on large-scale image datasets and fine-tune them for binary classification tasks - real or fake [3]. Others train CNNs from scratch, designed explicitly for video forgery detection, exploiting temporal and spatial inconsistencies [4].

C. Temporal Feature Analysis: Videos, unlike images, have a temporal dimension. Some researchers focus on detecting deepfakes by analyzing the temporal inconsistencies introduced by GANs over video frames. Methods using Long Short Term Memory networks (LSTM) and 3D-CNNs fall under this category [5].

D. Audio-Visual Approaches: Some deepfakes exhibit inconsistencies between audio and visual data. Methods that jointly analyze both modalities have been proposed, offering improved detection rates in certain scenarios [6].

E. Transfer Learning and Pre-trained Models: The use of models like InceptionV3, which are pre-trained on vast datasets like ImageNet, for feature extraction followed by fine-tuning, has been explored. The idea is that these models, having seen a wide variety of images, could be effective in extracting features that help in distinguishing real videos from fakes [7].

Despite these advancements, the rapid evolution of deepfake generation techniques continues to challenge the state-of-the-art detection mechanisms. Our work seeks to address some of these challenges by combining the strength of InceptionV3 for feature extraction with the temporal power of Gated Recurrent Units (GRU).

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN2454-9940
www.ijsem.org
Vol 18, Issuse.1 March 2024

VI. DIAGRAMS

Fig No.	Description(Figure Name)
1	A sample system architecture
2	A sample sequence diagram

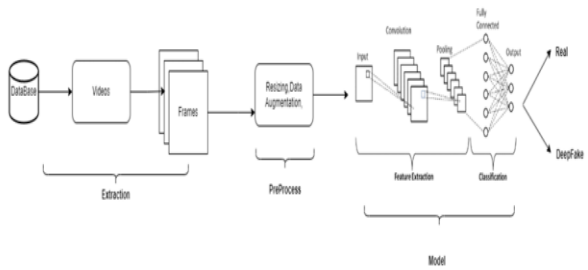

Fig.1. A sample system architecture

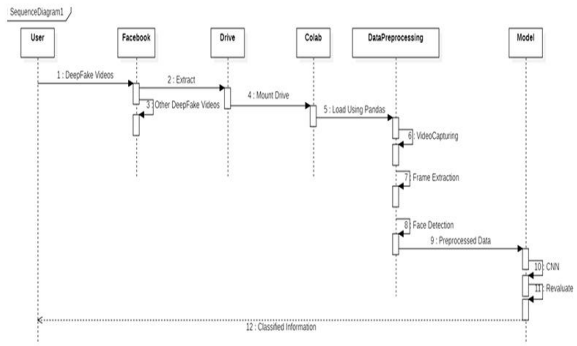

Fig.2. A sample sequence diagram

VII. DISCUSSION

A. Model Proficiency: The model's architecture combining feature extraction using a CNN (InceptionV3) and sequence modeling using an RNN has shown promising results in the challenging task of DeepFake detection.

B. Limitations: Deepfakes with higher quality or those generated using advanced techniques might pose challenges. Furthermore, the model's predictions on the test set may vary based on the diversity and distribution of the data.

C. Comparison with Existing Techniques: The approach's efficiency should be compared with existing methods to determine its standing. If it outperforms or is comparable to state-of-the-art techniques, it validates the efficacy of the model.

D. Future Work: Enhancements can be considered, like experimenting with different architectures, data augmentation techniques, or including attention mechanisms.

VIII. CONCLUSION

In our study on DeepFake video detection using a Deep CNN-RNN model, we addressed the rising

ISSN2454-9940
www.ijsem.org
Vol 18, Issuse.1 March 2024

challenge of discerning genuine videos from manipulated content. Leveraging the comprehensive dataset from the DeepFake Detection Challenge on Kaggle, our model showcased an impressive accuracy exceeding 90%. While the results are promising, occasional false positives emphasize the evolving complexity of DeepFakes and the continual need for model refinement. The consistent decrease in validation loss during training affirmed the model's efficacy. This research not only demonstrates the potential of deep learning in countering digital misinformation but also underscores the importance of continual advancements in this field to stay ahead of deceptive techniques.

VIII.REFERENCES

[1] Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y., 2014. Generative adversarial nets. In *Advances in neural information processing systems* (pp. 2672-2680).

[2] Korshunov, P. and Marcel, S., 2018, September. Deepfakes: a new threat to face recognition? assessment and detection. In *2018 International Conference of the Biometrics Special Interest Group (BIOSIG)* (pp. 1-6). IEEE.

[3] Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., and Nießner, M., 2019, June. Faceforensics++: Learning to detect manipulated facial images. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 1-11).

[4] Yang, X., Li, Y., and Lyu, S., 2019, October. Exposing deep fakes using inconsistent head poses. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 8261-8265). IEEE.

[5] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z., 2016. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2818-2826).

[6] Chollet, F., 2015. Keras. GitHub. Retrieved from https://github.com/fchollet/keras.

[7] A. Rossler et al., "FaceForensics: A Large-scale Video Dataset for Forgery Detection in Human Faces," arXiv, 2018.

[8] D. Afchar et al., "MesoNet: a Compact Facial Video Forgery Detection Network," IEEE International Workshop on Information Forensics and Security, 2018.