International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Deepfake Detection Using Deep Learning

Sugavaneshwari. P[1]*, Sreelekha. R*[2]*, Sathya Jothi. M*[3]*, Swetha. S*[4]*, Rudhra. J*[5]

[1-5]Department of Computer Science and Engineering, Government College of Engineering (Anna University), Dharmapuri, Tamil Nadu, India

ABSTRACT:

The proliferation of deep fake technology poses a significant threat to the integrity of digital media, leading to potential information and manipulation. Addressing this challenge requires robust techniques for detecting deep fake videos with high accuracy. This study propose a novel approach for deep fake detection in videos utilizing Convolutional Neural Network (CNN) with the InceptionV3 architecture. This methodology involves preprocessing video frames to extract relevant features and then employing a CNN-based model for classification. This design the CNN architecture to effectively capture discriminative patterns indicative of deep fake manipulation, such as inconsistencies in facial expression, unnatural artifacts, and temporal irregularities. The inceptionV3 architecture is chosen for its effectiveness in image classification tasks and its capability to intricate features at multiple scales

Index terms: Deep Learning, DeepFake videos, InceptionV3,CNN.

INTRODUCTION

Deepfake detection using deep learning involves several key steps. Initially, a sizable dataset comprising both genuine and synthetic media is collected. This dataset undergoes preprocessing to ensure uniformity. Following this, convolutional neural networks (CNNs) or similar architectures are employed to extract pertinent features from the data, enabling the model to discern between authentic and manipulated content. Through extensive training on the extracted features, the deep learning model learns to accurately classify inputs as either real or fake. Evaluation metrics like accuracy, precision, and loss the model's efficency in detecting deepfakes. Once validated, the model can be deployed for real-world application, continuously refined to adapt.

LITERATURE SURVEY

The literature survey offers a comprehensive overview of recent methodologies and advancements in the Deepfake Video Detection**.** Deressa Wodaj(2020) introduced a sophisticated " Deepfake Video Detection Using Convolutional Vision Transformer yielding significant results video detection. Deepfakes open new possibilities in digital media, VR, robotics, education, and many other fields. On another spectrum, these are technologies that can cause havoc and distrust to the general public. In light of this, it has designed and developed a generalized model for Deepfake video detection using CNNs and Transformer, which is named as Convolutional Vison Transformer. CNNs are strong at learning local features, while Transformers can learn from local and global feature maps. This combined capacity enables this model to correlate every pixel of an image and understand the relationship between nonlocal features. It gave equal emphasis on this data preprocessing during training and classification. It used the largest and most diverse dataset for Deepfake detection. The CViT model was trained on a diverse collection of facial images that were extracted from the DFDC dataset. Still, this model has a lot of room for improvement. In the future, It intend to expand on this current work by adding other datasets released for Deepfake research to make it more diverse, accurate, and robust.

CNN

Deepfake detection in videos has become crucial in combating the spread of manipulated content, offering robust accuracy and efficiency in analysing video frames. Deepfake detection algorithms are built to automatically extract intricate features from video sequences through a series of convolutional and recurrent layers. In the context of deepfake detection, these algorithms can discern subtle visual inconsistencies indicative of manipulated content, such as unnatural facial expressions, inconsistent lighting, or mismatched lip movements.

Similar to their application, convolutional neural networks (CNNs) excel in capturing spatial dependencies within video frames, allowing them to identify intricate patterns and artifacts characteristic of deepfake videos. By leveraging convolutional filters, CNNs can effectively extract features across multiple frames, enabling them to detect anomalies and inconsistencies that may signify the presence of deepfake manipulation. Moreover, CNN architectures such as VGG, Inception, and ResNet offer versatility in model design, enabling researchers to tailor network architectures and optimize performance for specific deepfake detection tasks.

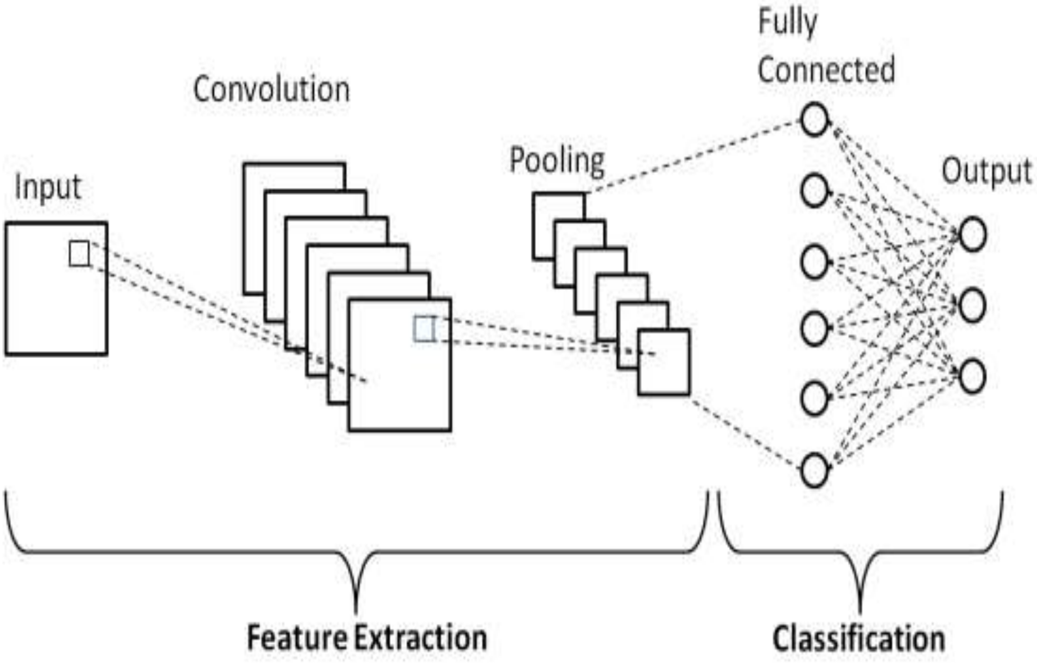

Fig 3.1CNN Architecture

CNNs demonstrate exceptional sensitivity to visual cues indicative of deepfake manipulation, enabling them to identify even subtle alterations in facial features, gestures, or background elements. Additionally, recurrent neural networks (RNNs) can capture temporal dependencies within video sequences, further enhancing the accuracy of deepfake detection by analysing the coherence and consistency of actions and movements over time. With their capacity for automated and accurate deepfake detection, CNNs and RNNs hold immense promise for mitigating the proliferation of manipulated videos across various platforms. By enabling swift identification and flagging of deepfake content, these algorithms contribute to preserving the integrity of visual media and safeguarding against misinformation and manipulation. This capability is crucial for maintaining trust and authenticity in digital content, thereby fostering a safer and more secure online environment.

EXPERIMENTAL RESULTS

The dataset used for experimentation consists of 4600 videos representing 2 categories of real and fake videos. Train the CNN model based on the Inception architecture, on the labeled dataset. During training, the model learns to extract discriminative features that distinguish between real and fake videos. This predicting the probability (real or fake) of the uploaded video based on permanence measure as accuracy, precision, loss and confusion matrix.

Dataset Description

A dataset for deepfake detection in videos would typically include a collection of videos that are classified into two categories: real and deepfake

Real Videos: This subset of the dataset consists of authentic videos captured from various sources such as movies, news broadcasts, social media, or personal recordings. These videos showcase real people and events without any manipulation.

Deepfake Videos: This subset comprises videos that have been artificially generated using deep learning techniques to manipulate the appearance or speech of individuals. Deepfake videos often superimpose one person's likeness onto another's body or alter their facial expressions and lip movements to create a realistic but synthetic representation.

Files

- train_sample_videos.zip - a ZIP file containing a sample set of training videos and a metadata.json with labels. the full set of training videos is available through the links provided above.
- sample_submission.csv - a sample submission file in the correct format.
- test_videos.zip - a zip file containing a small set of videos to be used as a public validation set. To understand the datasets available for this competition, review the Getting Started information.

Metadata Columns

- filename - the filename of the video

- label - whether the video is REAL or FAKE
- original - in the case that a train set video is FAKE, the original video is listed
- split - this is always equal to "train".

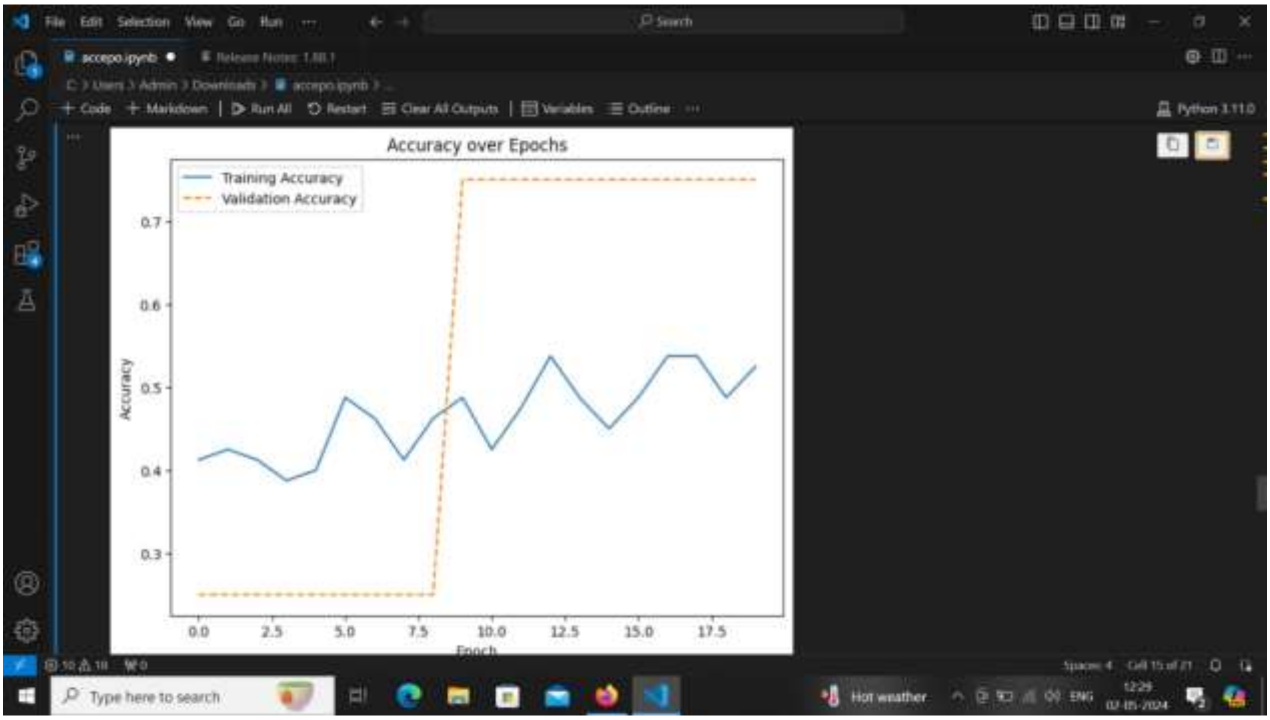

Fig.4.1 Accuracy vs epochs

CONCLUSION

The Custom Convolutional Neural Network that extracts visual artifacts to detect deepfake videos. Instead of using predesigned architecture user can define number of layers, type of layers and their configuration according to the requirement. CNN model learn from local and global image features of a video. This project aims to provide an effective solution for identifying whether the video is real or fake . Those performance measures are accuracy, loss and confusion matrix. This offers a promising approach for accurate and reliable deepfake video detection. The performance of CNN can be analysed by accuracy of the model training and validation set over different epochs. Accuracy increases with number of epochs over videos. This project with an accuracy of 81 percent and minimum loss with a dataset containing 4600 videos

REFERENCES

1. International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 11 Issue V May 2023- Available at www.ijraset.com

2. Deressa Wodajo Jimma University deressa.wodajo@ju.edu.et Solomon Atnafu Addis Ababa University solomon.atnafu@aau.edu.et

3. G.Oberoi. Exploring DeepFakes. Accessed: Jan. 4, 2021.Available: https://goberoi.com/exploring-deepfakes-20c9947c22d9

4. J. Hui. How Deep Learning Fakes Videos (Deepfake) and How to Detect it. Accessed: Jan. 4, 2021. Available: https://medium. com/how-deep-learning-fakes-videos-deepfakes-and-how-to-detect-itc0b50fbf7cb9

5. ''Generative adversarial nets,'' in Proc. 27th I. Goodfellow, J. P. Abadie, M. Mirza, B. Xu, D. W. Farley, S. Ozair, A. Courville, and Y. Bengio, Int. Conf. Neural Inf. Process. Syst. (NIPS), vol. 2. Cambridge, MA, USA: MIT Press, 2014, pp. 2672–2680.

6. 'The state of deepfakes: Reality under attack,'' Deeptrace B.V., Amsterdam, The Netherlands, Annu. Rep. v.2.3., 2018. G. Patrini, F. Cavalli, and H. Ajder, [Online]. Available:https://s3.eu-west2.amazonaws.com/rep2018/2018-the-state-ofdeepfakes.

7. ''Face2Face: Real-time face capture and reenactment of RGB videos,'' J. Thies, M. Zollhofer, M. Stamminger, C. Theobalt, and M. Niessner, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, USA, Jun. 2016, pp. 2387–395, doi: 10.1109/CVPR.2016.262.J

8. Comput. Vis. (ICCV), Y. Zhu, T. Park, P. Isola, and A. A. Efros Venice, Oct. 2017,]-, IEEE Int. Confs pp. 2242–2251, doi: 10.1109/ICCV.2017.244.

9. ‘‘Synthesizing Obama: Learning lip sync from audio,’’ S. Suwajanakorn, M. Seitz, and I. K. Shlizerman, ACM Trans. Graph., vol. 36, no. 4, p. 95, 2017.

10.Artificial Intelligence is Now Fighting Fake Porn. Accessed: Jan. 4, 2021. L.Matsakis.[Online]. Available: https://www.wired.com/story/gfycatartificial-intelligence-deepfakes/

11. A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, ‘‘FaceForensics: A large-scale video dataset for forgery detection in human faces’ L. Matsakis. 2018, arXiv:1803.09179.

12. Deep video portraits, H. Kim, P. Garrido, A. Tewari, W. Xu, J. Thies, M. Niessner, P. Pérez, C. Richardt, M. Zollhöfer, and C. Theobalt, ACM Trans. Graph., vol. 37, no. 4, pp. 1–14, Aug. 2018, doi: 10.1145/3197517.3201283.

13. ‘‘Everybody dance now,’’ C. Chan, S. Ginosar, T. Zhou, and A. A. Efros, 2018, arXiv:1808.07371.

14. A style-based generator architecture for generative adversarial networks,’’ T. Karras, S. Laine, and T. Aila, in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Long Beach, CA, USA, Jun. 2019, pp. 4396–4405, doi: 10.1109/CVPR.2019.00453.

15. ‘‘Performing systematic literature reviews in software engineering,’’ D. Budgen and P. Brereton, in Proc. 28th Int. Conf. Softw. Eng., New York, NY, USA, May 2006, pp. 1051–1052, doi: 10.1145/1134285.1134500.