International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

# DETECTION OF DEEPFAKE VIDEOS USING LONG DISTANCE ATTENTION

**Putta Srivani[1],E. Sai sreeja[2], P.Harshitha[3], S. Meghana [4]**

[1] Associate Professor, Department of IT ,Malla Reddy Engineering College For Women(Autonomous Institution), Maisammaguda, Dhulapally,Secunderabad,Telangana-500100

[234]UG Scholar, Department of CS,Malla Reddy Engineering College for Women, (Autonomous Institution), Maisammaguda,Dhulapally,Secunderabad,Telangana-500100

Email : Pulla.srivani@gmail.com

## ABSTRACT

Deepfake techniques have reached an incredibly fast pace lately, making highly deceptive videos. Detecting such forging videos is urgently needed; it is also a rather challenging task. Unlike in the traditional approach, when this problem was treated more like a binary classification type, the approach here considered this problem as fine-grained classification due to the minute differences present between the real and the fake faces.Observations depict that most face forgery methods leave artifacts in spatial and temporal domains, that is, generative defects within individual frames and interframe inconsistencies. To catch these artifacts, a new spatial-temporal model, with two components designed through a long-distance attention mechanism, is proposed. The former component focuses on the detection of the artifacts in single frames and the latter one on capturing interframe inconsistencies. The attention mechanism produces patch-based attention maps that lead the network to pay attention to critical facial regions, enhancing the integration of global context and local statistical information.Experimental results on a variety of public datasets demonstrate that the proposed method achieves state-of-the-art performance. The long-distance attention mechanism effectively identifies pivotal forgery traces, providing a robust solution to the challenges of deepfake detection.

**Keywords**-Deepfake Detection;Video Forgery;Spatial-Temporal Model;Long-Distance Attention Mechanism;Binary Classification;Fine-Grained Classification.

## I. INTRODUCTION

Deepfake technology has advanced rapidly in recent years, enabling the creation of highly realistic videos by replacing the face of one person with another. Powered by sophisticated generative models, deepfake videos have become increasingly difficult to distinguish from authentic content, raising concerns over their potential misuse. These videos can be easily created and shared, which opens the door for spreading misinformation, rumors, and hate, causing significant harm to individuals and society. In the era of the internet, the ubiquity of such technology presents a serious threat to public trust and safety.The process of creating a deepfake video typically involves extracting the face from a video, converting it into a target face using generative models, and then reassembling the video by inserting the manipulated face back into the corresponding frames. During this process, two types of defects are inevitably introduced: **spatial defects (artifacts within individual frames) and temporal inconsistencies (mismatches across consecutive frames). These artifacts are a result of imperfections in the generative models and the lack of global constraints when stitching frames together to create the final video.While many deepfake detection methods have focused on identifying spatial defects—such as abnormal facial features or mismatched details—these methods are often fragile and may fail when deepfake videos do not exhibit the typical semantic abnormalities. The reliance on semantic cues alone limits the effectiveness of

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

detection methods, especially as deepfake technology improves and produces videos with fewer visible artifacts.

Thus, there is an urgent need for more robust detection techniques that can reliably identify deepfake videos even when traditional semantic-based methods are insufficient. The goal is to develop detection systems that address both the spatial and temporal aspects of deepfakes, offering a more comprehensive solution to this growing problem.

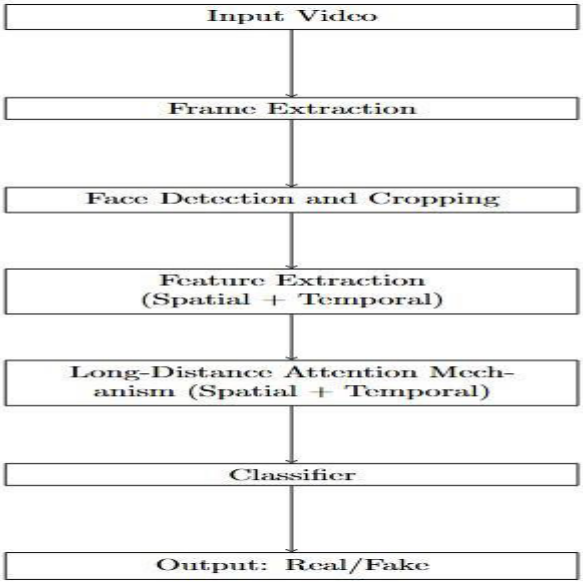


**Fig 1: System Architecture**

## II. RELATED WORK

### 1.Generative Adversarial Networks (GANs)

**Goodfellow et al. in 2014.** Two neural networks comprising this framework are the generator and discriminator, which create artificial data, and the process tries to distinguish between what is real and what is fake. This adversarial process brings about the increase of realness in produced data. In the context of deepfakes, GANs are used to create highly realistic videos by swapping faces, thus making it a powerful tool for generating deepfakes. The generator's ability to create near-identical faces is a major challenge for detection, as it becomes harder to distinguish between real and fake faces in deepfake videos.

### 2.Variational Autoencoders (VAEs)

**Kingma and Welling in 2013**, describe a method for learning compressed representations of data that later can be used to generate new samples. VAEs are not originally designed for deepfakes but have been modified to generate faces and modify images in ways that match deepfake techniques. The fact that VAEs could be used to generate very convincing face images has led to the proliferation of deepfakes. These models can easily manipulate faces, which is at the core of deepfake videos. However, when the generated faces are highly realistic, detection becomes challenging, which is why advanced techniques such as attention mechanisms are needed to spot subtle flaws in deepfakes.

### 3. Face Frontalization and Recognition with GANs

**Duan and Zhang (2021)** introduced a technique that can handle face recognition in difficult conditions such as occlusions or misalignment using GAN-based models. The work is known as BoostGAN, focusing on improving the frontalization of faces, making it easier for the recognition systems to process faces which may be rotated or partially occluded. This method has implications for deepfake detection because it allows for the reconstruction of realistic frontal faces that can be analyzed more effectively for manipulation. The key idea is to recognize faces more accurately, even when they contain distortions or mismatches, as is typical in deepfake videos. This enhances the ability of detection systems to identify manipulated content by focusing on face alignment and recognition.

### 4.Generative Adversarial Networks GANs

**Goodfellow et al. in 2014** is composed of a generator and a discriminator. GANs are the most famous type of deepfake model currently and has been used for training purposes that generates

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

images mimicking real data while the discriminator learns the difference between real and fake. Due to rapid improvements in GANs, deepfakes are getting more challenging to identify since they produce faces and videos with very high realism. Increasing realism of deepfakes underscores the necessity of further enhanced methods for detecting these videos rather than the mere visual inspection .

### 5.Variational Autoencoders (VAEs)

**Kingma and Welling (2013)** proposed Variational Autoencoders, which are another variant of generative models that have been used to generate faces and images in a realistic fashion. VAEs use the encoder-decoder architecture to produce high-quality images that resemble actual photographs. These models, although not developed for deepfakes, are part of a larger landscape of generative models that are contributing to the proliferation of fake media. Deepfakes are detected by a discrepancy between the real content and the generated content, where the statistical difference between generated facial features and real facial features can be captured by a deep learning model .

## III. IMPLEMENTATION

Several are the key components of deepfake detection implementation through a spatial-temporal model. First, face detection is done by OpenCV or MTCNN to extract and crop each frame of the video. This will help the system concentrate on the manipulated face. To extract the features, a pre-trained CNN such as VGG16 captures spatial features, thereby capturing fine details within the face, and 3D-CNNs or Recurrent Neural Networks (RNNs) are used to analyze the temporal features, where inconsistencies are detected between consecutive frames. A crucial aspect of the model is the long-distance attention mechanism. The spatial attention layer focuses on the important regions of the face, that are eyes and mouth in this case, to detect artefacts, while the temporal attention layer focuses on artefacts between frames to ensure the consistency of the whole. Finally, the model outputs a real or fake video by making use of both spatial and temporal features and passing them to a classifier. The dense layer is used for the final classification. The performance of the model is evaluated by comparing the predictions with the ground truth. This approach effectively exploits the spatial and temporal aspects of deepfakes, which makes it a robust solution for even the most sophisticated manipulations.

## IV. ALGORITHM

### Step 1: Input Video Preprocessing

**Extract frames from the video**: Divide the input video into individual frames for easier analysis.

### Step 2: Face Detection

**Detect faces in each frame:** Use a face detection algorithm (e.g., MTCNN, OpenCV) to identify the locations of faces in the frames.

### Step 3: Face Cropping

**Crop the face region:** For each detected face, crop the region containing the face from the frame. This will isolate the manipulated face for further analysis.

### Step 4: Feature Extraction

**Spatial Feature Extraction:**

- Utilize a pre-trained Convolutional Neural Network (CNN) like VGG16 to extract spatial features from the cropped faces.
- These features capture local facial details such as texture, shape, and imperfections.

**Temporal Feature Extraction:**

- Utilize a 3D-CNN or RNN to identify temporal connections between consecutive frames.
- This would facilitate the identification of inconsistencies or unnatural changes in frames that might signify tampering.

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

**Step 5: Attention Mechanism**

**Spatial Attention:**

- Implement spatial attention on important facial areas such as eyes, mouth, and so on.
- This step detects localized artifacts or minor anomalies in these areas.

**Temporal Attention:**

- Use a temporal attention mechanism to identify inconsistencies or mismatches between consecutive frames.
- This step detects problems such as jittering or unnatural facial movements that arise from frame-to-frame inconsistencies.

**Step 6: Feature Combination**

- Combine Spatial, as well as Temporal Features Together: Combine the spatial features obtained from facial analysis together with temporal features obtained from interframe analysis. This will create a fully robust feature set for classification purpose

**Step 7: Classification**

**Classify video:**

- feed the combined features to classification layers-a completely connected neural network-to tell whether the video is genuine or fake.
- Use a sigmoid activation function to output a binary result: real (0) or fake (1).

**Step 8: Output**

Output whether the video is real or fake.

- Optionally, give a confidence score along with the result to indicate the model's certainty.
- Optionally, return the facial regions or frame inconsistencies that led to the detection decision.

This step-by-step process ensures that the deepfake detection system captures the spatial artifacts and temporal inconsistencies inside a video quite efficiently, thus improving the detection accuracy.

## RESULTS


**Fig 1: User Login**


**Fig 2:Tested Results**

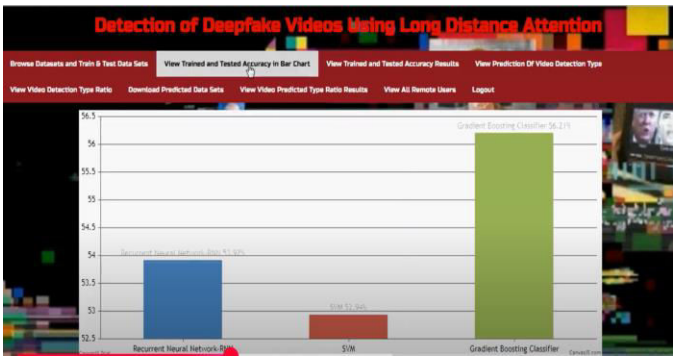


**Fig 3:Accuracy Bar Graph**

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

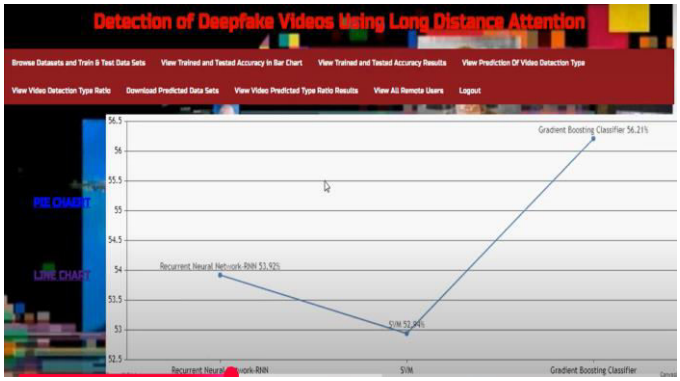


**Fig 4:Line Graph**

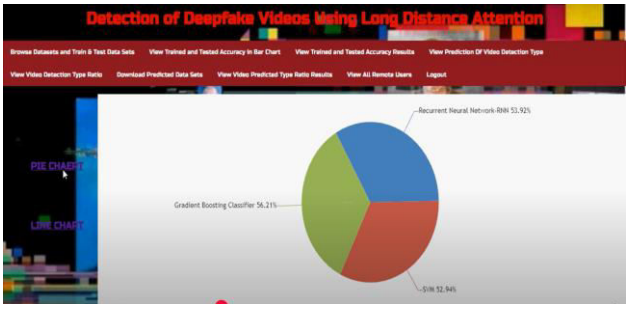


**Fig 5: Pie Chart**

## CONCLUSION

In summary, deepfake video detection is one of the key challenges of advanced generative models in the age of very realistic forgeries. By employing a spatial-temporal model with attention mechanisms, this approach is effective in addressing both spatial artifacts (defects in individual frames) and temporal inconsistencies (discrepancies between frames). Both CNNs and RNNs or 3D-CNNs are used for extracting key features from facial regions of each frame and the relationships between frames. Spatial and temporal attention mechanisms further enhance the capability to focus on critical facial areas and to detect subtle inconsistencies. A combination of these features, along with a classification model applied to the system, can accurately determine whether a video is real or fake. This approach provides a strong solution to the emerging threat of deepfakes to counter the misuse of this technology in spreading false information or leading to harm..

## REFERENCES

I.Goodfellow et al., "Generative adversarial nets", Proc. Adv. Neural Inf. Process. Syst. (NIPS), pp. 2672-2680, 2014.

2.D. P. Kingma and M. Welling, "Auto-encoding variational Bayes", arXiv:1312.6114, 2013.

3.Q. Duan and L. Zhang, "Look more into occlusion: Realistic face frontalization and recognition with BoostGAN", IEEE Trans. Netw. Learn. Syst., vol. 32, no. 1, pp. 214-228, Jan. 2021.

4.Deepfake, Sep. 2019, [online] Available: https://www.github.com/deepfakes/.

5.Faceswap, Sep. 2019, [online] Available: https://www.github.com/MarekKowalski/.

6.F. Matern, C. Riess and M. Stamminger, "Exploiting visual artifacts to expose deepfakes and face manipulations", Proc. IEEE Winter Appl. Comput. Vis. Workshops (WACVW), pp. 83-92, Jan. 2019.

7.D. Afchar, V. Nozick, J. Yamagishi and I. Echizen, "MesoNet: A compact facial video forgery detection network", Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS), pp. 1-7, Dec. 2018.

8.X. Yang, Y. Li, H. Qi and S. Lyu, "Exposing GAN-synthesized faces using landmark locations", Proc. ACM Workshop Inf. Hiding Multimedia Secur., pp. 113-118, Jul. 2019.

9.D.-T. Dang-Nguyen, G. Boato and F. G. De Natale, "Discrimination between computer generated and natural human faces based on asymmetry information", Proc. 20th Eur. Signal Process. Conf., pp. 1234-1238, Aug. 2012.

10.P. Zhou, X. Han, V. I. Morariu and L. S. Davis, "Two-stream neural networks for tampered face detection", Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), pp. 1831-1839, Jul. 2017.

11.B. Bayar and M. C. Stamm, "A deep learning approach to universal image manipulation detection using a new convolutional layer", Proc. 4th ACM Workshop Inf. Hiding Multimedia Secur., pp. 5-10, Jun. 2016.

12.U. A. Ciftci, I. Demir and L. Yin, "FakeCatcher: Detection of synthetic portrait videos using biological signals", IEEE Trans. Pattern Anal. Mach. Intell., Jul. 2020.

13.M. Li, B. Liu, Y. Hu and Y. Wang, "Exposing deepfake videos by tracking eye movements", Proc. 25th Int. Conf. Pattern Recognit. (ICPR), pp. 5184-5189, Jan. 2021.

14.Y. Li, M.-C. Chang and S. Lyu, "In ictu oculi: Exposing AI created fake videos by detecting eye blinking", Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS), pp. 1-7, Dec. 2018.

15.C.-Z. Yang, J. Ma, S. Wang and A. W.-C. Liew, "Preventing deepfake attacks on speaker authentication by dynamic lip movement analysis", IEEE Trans. Inf. Forensics Security, vol. 16, pp. 1841-1854, 2021.

16.S. Fernandes et al., "Predicting heart rate variations of deepfake videos using neural ODE", Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop, pp. 1721-1729, Oct. 2019.

17.I. Amerini, L. Galteri, R. Caldelli and A. Del Bimbo, "Deepfake video detection through optical flow based CNN", Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW), pp. 1205-1207, Oct. 2019.

18.G. Wang, J. Zhou and Y. Wu, "Exposing deep-faked videos by anomalous co-motion pattern detection", arXiv:2008.04848, 2020.

19.H. Zhao, T. Wei, W. Zhou, W. Zhang, D. Chen and N. Yu, "Multi-attentional deepfake detection", Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 2185-2194, Jun. 2021.

20.T. Hu, H. Qi, Q. Huang and Y. Lu, "See better before looking closer: Weakly supervised data augmentation network for fine-grained visual classification", arXiv:1901.09891, 2019.

21.X. Yang, Y. Li and S. Lyu, "Exposing deep fakes using inconsistent head poses", Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP), pp. 8261-8265, May 2019.

22.S. Agarwal, H. Farid, Y. Gu, M. He, K. Nagano and H. Li, "Protecting world leaders against deep fakes", Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops, pp. 1-8, Jun. 2019.