iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** III **Month of publication:** March 2026

DOI: https://doi.org/10.22214/ijraset.2026.78356

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

Deepfake Video Detection using Neural Networks

A.S. Nagargoje[1], Disha Thakur[2], Darshana Patil[3], Lenisha Murkar[4], Rutika Madane[5]
[1]*Guide, Department Computer Engineering, AISSMS Polytechnic College, Pune, Maharashtra, India*
[2, 3, 4, 5]*Student, Department of Computer Engineering, AISSMS Polytechnic College, Pune, Maharashtra, India*

Abstract: ***In recent months, advancements in free deep learning-based software tools have made it easier to create convincing face swaps in videos, often leaving minimal traces of manipulation. This phenomenon is commonly known as "DeepFake" (DF) videos. While manipulations of digital videos using visual effects have been demonstrated for decades, recent progress in deep learning has significantly enhanced the realism of fake content and the ease with which it can be generated. These artificially intelligent tools, commonly referred to as AI-synthesized media or DF, have simplified the creation process. However, detecting DeepFake videos poses a significant challenge. Despite the simplicity of generating DF using AI tools, training algorithms to identify them is not straightforward. To address this issue, we have taken a step forward by employing Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) for detection. Our system utilizes a CNN to extract features at the frame level. These features are then employed to train an RNN, which learns to classify whether a video has undergone manipulation. The system is capable of detecting temporal inconsistencies between frames introduced by the tools used in DF creation. To evaluate the effectiveness of our approach, we tested it against a large set of fake videos collected from a standard dataset. The results demonstrate the competitiveness of our system, achieved through a simple architecture.***

Keywords: Deepfake Video Detection , Convolutional Neural Network(CNN), Recurrent Neural Network (RNN)

I. INTRODUCTION

The widespread use of advanced smartphone cameras and the global availability of high-speed internet have significantly expanded the influence of social media and media-sharing platforms. This has made the creation and sharing of digital videos more accessible than ever. The increasing computational power, particularly in deep learning, has surpassed what was once considered impossible just a few years ago. However, as with any transformative technology, this has brought about new challenges. One such challenge is the rise of "DeepFake" (DF) content, produced by sophisticated generative adversarial models capable of manipulating video and audio clips. The dissemination of DeepFake content on social media platforms has become common, leading to issues such as spamming and the spread of misinformation. These instances of DeepFake content can be alarming and pose threats, misleading the general public. To address this situation, the detection of DeepFake content becomes crucial. In response, we introduce a novel deep learning-based method designed to effectively distinguish AI-generated fake videos (DF Videos) from genuine ones. Developing technology capable of identifying and preventing the spread of DeepFake content on the internet is of utmost importance.

Understanding the process by which Generative Adversarial Networks (GAN) create DeepFake content is key to detection. GAN takes a video and an image of a specific individual (the 'target') as input, generating another video with the target's faces replaced by those of a different individual (the 'source'). Deep adversarial neural networks, trained on face images and target videos, play a critical role in automatically mapping the faces and facial expressions from the source to the target. The resulting videos can achieve a high level of realism through post-processing.

Our method for detecting DeepFake content is rooted in the same process used by GAN to create DeepFakes. It leverages specific properties of DeepFake videos, acknowledging that due to limitations in computation resources and production time, the DeepFake algorithm can only synthesize face images of a fixed size. These images undergo affinal warping to match the configuration of the source's face, leaving distinguishable artifacts in the output.

DeepFake video due to resolution inconsistencies between the warped face area and the surrounding context.

Our detection technique compares the generated face areas with their surrounding regions in order to identify

these artefacts. To accomplish this, we divide the video into frames, use a ResNext Convolutional Neural Network (CNN) to extract features, and record temporal inconsistencies generated by GAN during the reconstruction of the DeepFake using a Recurrent Neural Network (RNN) with Long ShortTerm Memory (LSTM). We streamline the procedure by

directly modelling the resolution discrepancy in affine face wrappings in order to train the ResNext CNN model.

©IJRASET: All Rights are Reserved

II. LITERATURE SURVEY

The public's trust, democracy, and justice are seriously threatened by the deepfake video industry's fast expansion and illicit use. As a result, there is a greater need for the analysis, detection, and intervention of fraudulent videos. Some of the related word in deep fake detection are listed below: ExposingDF Videos by Detecting Face Warping Artifacts used an approach to detects artifacts by comparing the generated face areas and their surrounding regions with a dedicated Convolutional Neural Network model. There were two types of face artefacts in this piece. Their approach is based on the observation that the present DF algorithm can only produce images with a certain resolution; these images must then undergo further transformations in order to match the faces that need to be replaced in the original video. A novel technique for exposing false face movies made with deep neural network models is described in Exposing AI Created false movies by Detecting Eye Blinking. The technique relies on the identification of eye blinking in the movies, a physiological signal that is not well displayed in the artificially created videos. When tested against benchmarks of eye-blinking detection datasets, the approach performs admirably when it comes to identifying movies produced by Deep Neural Network-based software, or DF. Their detecting mechanism just looks for the absence of blinking. But in order to identify the deep fake, other factors like facial wrinkles and teeth enchantment must also be taken into account. We offer our strategy taking into account each of these factors. The technique of using a capsule network to identify modified, forged images and videos in various contexts, such as replay attack detection and computer-generated video detection, is known as "capsule network based image and video detection." They have employed random noise in the training phase of their approach, which is not a recommended practice. Even so, the model did well on their dataset; however, noise in the training set could cause it to perform poorly on real-time data. We suggest using real-time and noiseless datasets to train our technique. Biological Signals Based Synthetic Portrait Video Detection: This technique gathers biological signals from the face regions of real and phoney portrait video pairings. Utilise transformations to train a CNN and a probabilistic SVM, compute the temporal consistency and spatial coherence, and represent the signal properties in feature sets and PPG maps. Subsequently, the total authenticity probabilities are used to determine the legitimacy of the video. Fraudulent Catcher accurately identifies fraudulent content regardless of the video's source, content, resolution, or quality. The procedure of developing a differentiable loss function that adheres to the suggested signal processing processes is not simple because the absence of a discriminator results in a loss in their discoveries to preserve biological signals.

III. PROPOSED SYSTEM

Numerous tools exist for creating DeepFake (DF) content, but there is a notable scarcity of tools dedicated to DF detection. In response to this gap, our approach to detecting DeepFakes represents a significant contribution aimed at preventing the proliferation of deceptive content across the World Wide Web. We are developing a user-friendly web-based platform that allows users to upload videos and determine whether they are authentic or fake. This project has the potential for scalability, evolving from a web-based platform to a browser plugin designed for automatic DF detection. Moreover, larger applications such as WhatsApp and Facebook could integrate our project into their platforms, enabling users to identify and prevent the transmission of DeepFake content before sending it to others. An essential objective of our project is to assess its performance and acceptance based on criteria such as security, user-friendliness, accuracy, and reliability. Our method is specifically tailored to detect various types of DeepFake content, including replacement DeepFakes, retrenchment DeepFakes, and interpersonal DeepFakes. The proposed system architecture, as illustrated in Figure 1, outlines a straightforward representation of our system. This architecture encompasses the steps involved in the detection process, emphasizing the simplicity and clarity of our approach. We believe that addressing the challenges posed by DeepFake content detection is not only crucial for maintaining the integrity of digital media but also for fostering a safer online environment for users.

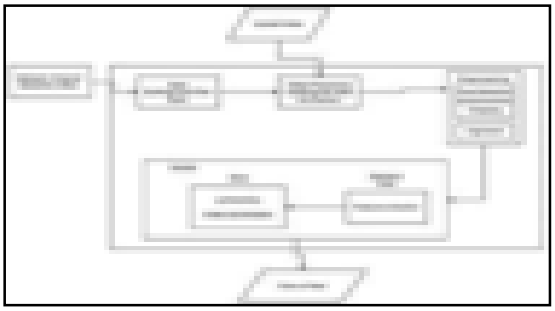


Fig.1:SystemArchitecture

A. Dataset

Our methodology employs a diverse dataset, amalgamating equal proportions of videos from various sources, including YouTube, FaceForensics++, and the Deep Fake Detection Challenge dataset. In addition, we have curated a new dataset comprising an equal distribution of 50% original videos and 50% manipulated deepfake videos. This balanced composition ensures a comprehensive representation of both authentic and manipulated content within our dataset. To facilitate the training and evaluation of our deep fake detection model, we have partitioned the dataset into a 70% training set and a 30% test set. This division allows for a robust training phase on a significant portion of the data, followed by a rigorous evaluation on an independent subset to assess the model's generalization and performance on previously unseen samples. The utilization of diverse datasets and a well-structured split aims to enhance the model's ability to discern between authentic and manipulated content across different sources and scenarios.

B. Preprocessing

The dataset preprocessing stage involves several key steps to enhance the efficiency of subsequent model training. Initially, each video is split into frames, followed by face detection to identify and isolate facial regions within each frame. Subsequently, to maintain consistency in the number of frames across the dataset, the mean frame count is computed. A new processed dataset is then generated, containing frames equal to this computed mean. During this process, frames lacking detected faces are excluded to ensure the dataset's focus on facial features. Recognizing the computational demands associated with processing an entire 10-second video, especially at a frame rate of 30 frames per second (resulting in a total of 300 frames), we propose a pragmatic approach for experimental purposes. Specifically, for training the model, we suggest utilizing only the initial 100 frames from each video. This reduction in frame count aims to alleviate computational requirements during the experimental phase while still providing sufficient data for training and validating the deep fake detection model effectively. This approach allows for a more manageable and resource-efficient exploration of the model's performance within the experimental constraints.

C. Model

The architecture of our model comprises a ResNeXt50_32x4d backbone followed by a single Long Short-Term Memory (LSTM) layer. The Data Loader component is responsible for loading the preprocessed face cropped videos and subsequently dividing them into distinct training and testing sets. In the training phase, frames extracted from the processed videos are fed into the model in mini-batches. This ensures an efficient and iterative approach to training, allowing the model to learn and adapt to patterns within the data. Similarly, during testing, the frames are processed through the trained model to evaluate its performance on the unseen test set. The combination of the ResNeXt50_32x4d backbone and LSTM layer is designed to leverage the capabilities of both convolutional neural networks (CNNs) for spatial feature extraction and LSTMs for capturing temporal dependencies within sequential data. This architecture aims to enhance the model's ability to discern patterns in facial features over time, contributing to its effectiveness in detecting deep fakes. The modular design facilitates a systematic and organized approach to training and evaluating the model's performance on the given preprocessed video datasets.

D. ResNext CNN for Feature Extraction

Rather than implementing a new classifier, we advocate leveraging the ResNext CNN classifier to extract features and effectively identify frame-level characteristics. Our approach involves fine-tuning the network by incorporating additional layers as needed and carefully selecting an appropriate learning rate to ensure the proper convergence of the gradient descent during the training process. The 2048-dimensional feature vectors obtained after the final pooling layers in the ResNext architecture are utilized as input for the sequential Long Short-Term Memory (LSTM) layer. This integration allows the model to capture and analyze temporal dependencies within the sequential data. The choice of a 2048-dimensional feature vector ensures a rich representation of the frame-level features extracted by the ResNext CNN classifier, contributing to the accuracy and effectiveness of the deep fake detection model.

E. LSTM for Sequence Processing

Assume for now that a 2-node neural network receives a sequence of ResNext CNN feature vectors of input frames as input. Determine the likelihood that the sequence is a deepfake video or an intact video. The design of a model to meaningfully process a series recursively is the main issue that needs to be resolved. We are suggesting that a 2048 LSTM unit with a 0.4 dropout probability be used for this challenge in order to accomplish our goal. The sequential processing of the frames using LSTM allows for the comparison of the frame at 't' seconds with the frame at 't-n' seconds, which allows for the temporal analysis of the video.

Where n can be any number of frames before.

F. Predict

The trained model receives a new video to forecast. Additionally, a fresh video is preprocessed to import the format of the learned model. After the video is divided into frames, the faces are cropped, and the cropped frames are sent straight to the trained model for detection rather than being stored locally.



Fig. 2: Training Flow

IV. RESULT

The model's output is designed to provide a determination of whether a given video is a deepfake or an authentic recording. Additionally, the output includes a confidence score, indicating the level of certainty the model has in its classification. To illustrate this concept, an example is presented in Figure 3. In Figure 3, you can observe a representation of the model's output, where it not only categorizes the video as a deepfake or real but also assigns a confidence level to this classification. This visual representation serves as a straightforward way for users to interpret and understand the model's assessment of the authenticity of the video. The higher the confidence score, the more certain the model is in its classification, providing users with valuable insights into the reliability of the model's decision. This approach enhances the transparency and usability of the model's output, ensuring that users can make informed decisions based on both the classification and the associated confidence level.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

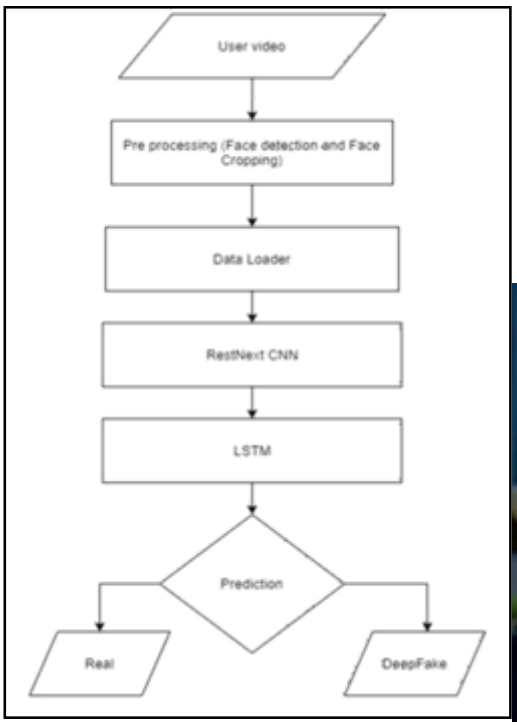


Fig.4:Predict flow





Fig.3:Expected Result

V. CONCLUSION

We introduced an approach that employs a neural network for the classification of videos, determining whether they are deepfakes or genuine recordings. Our method not only categorizes the videos but also provides a confidence score associated with the model's assessment. The inspiration for our method comes from the techniques used in the creation of deepfakes, particularly those generated by Generative Adversarial Networks (GANs) with the assistance of Autoencoders. Our approach focuses on detecting deepfakes at the frame level, utilizing a ResNext Convolutional Neural Network (CNN), and extends to video classification using Recurrent Neural Network (RNN) in conjunction with Long Short-Term Memory (LSTM). By leveraging these techniques, our proposed method demonstrates the capability to identify whether a video is a deepfake or real, based on the parameters outlined in the associated research paper. We are confident that our method will yield a high level of accuracy when applied to real-time data. The combination of frame-level detection and video classification, along with the integration of deep learning components, positions our approach as a robust solution for discerning between authentic and manipulated videos. This has implications for enhancing the accuracy and reliability of real-time video content analysis, contributing to the broader field of video forensics and authentication.

VI. LIMITATIONS

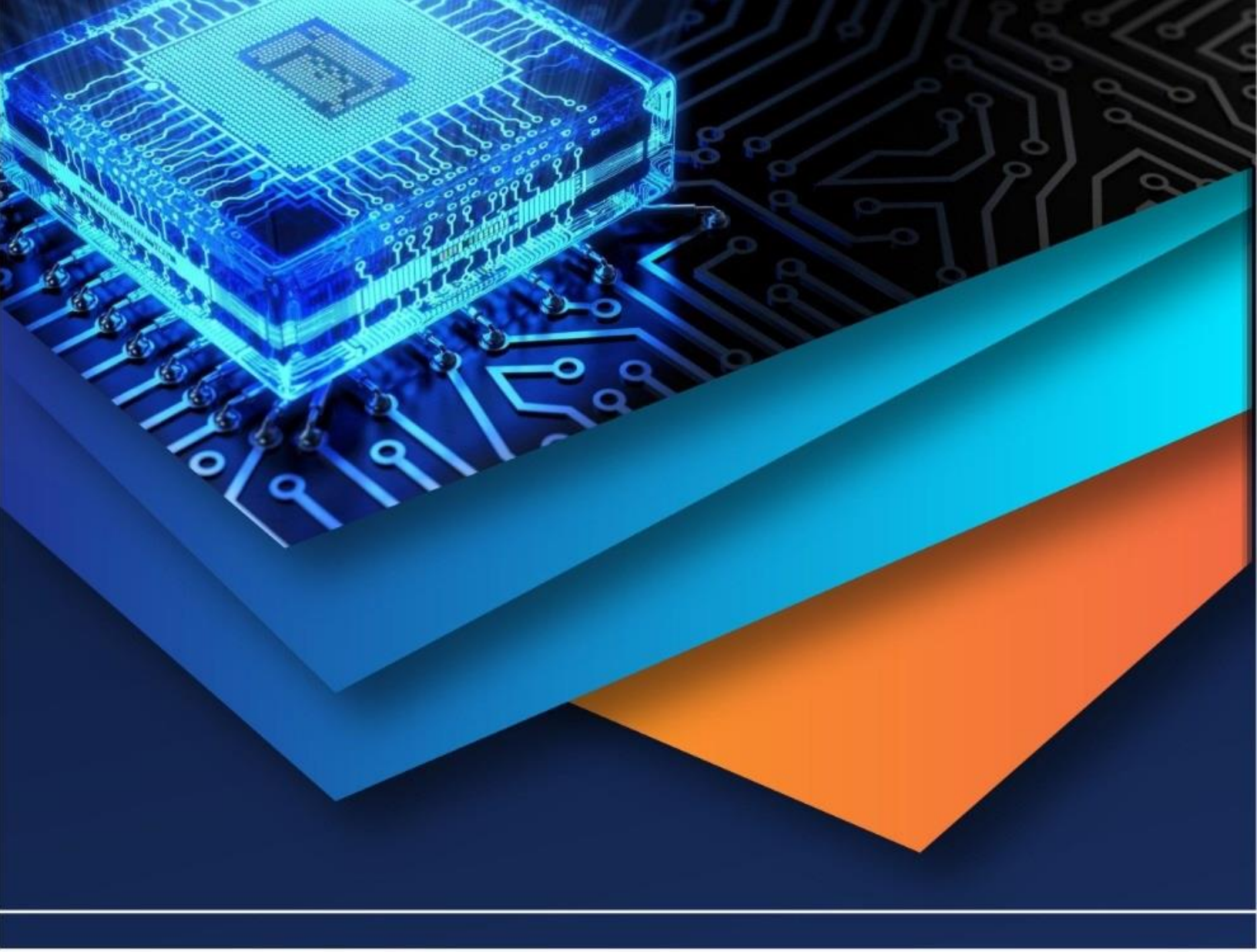
Certainly, it's essential to acknowledge the current limitation in our method, which does not incorporate considerations for audio, thereby making it incapable of detecting audio deep fakes. Recognizing this gap, we are actively working towards enhancing our approach to include audio deep fake detection as a crucial component of our future developments. The significance of addressing audio deep fakes is increasingly evident as technology advances, and malicious actors become more sophisticated. Audio deep fakes pose unique challenges due to their potential to deceive individuals through manipulated voice recordings. Our commitment to advancing detection capabilities extends beyond visual elements to encompass a comprehensive solution that addresses the audio dimension of deep fakes. In our forthcoming work, we plan to explore innovative techniques and integrate state-of-the-art methodologies for audio analysis.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

This proactive approach aligns with our dedication to staying at the forefront of deep fake detection technology. By incorporating audio considerations into our method, we aim to provide a more robust and comprehensive solution that addresses the evolving landscape of synthetic media.As we continue our research and development efforts, we remain committed to refining our method to ensure it remains effective in detecting both visual and audio deep fakes. Through these advancements, we aspire to contribute to the ongoing efforts in safeguarding the integrity of digital content and countering the potential risks associated with deceptive audiovisual manipulations.

REFERENCES

[1] Yuezun Li, Siwei Lyu, “ExposingDF Videos By Detecting Face Warping Artifacts,” in arXiv:1811.00656v3.
[2] Yuezun Li, Ming-Ching Chang and Siwei Lyu “Exposing AI Created Fake Videos by Detecting Eye Blinking” in arxiv.
[3] Huy H. Nguyen, Junichi Yamagishi, and Isao Echizen “Using capsule networks to detect forged images and videos”.
[4] Hyeongwoo Kim, Pablo Garrido, Ayush Tewari and Weipeng Xu “Deep Video Portraits” in arXiv: 1901.02212v2.
[5] Umur Aybars Ciftci, ˙Ilke Demir, Lijun Yin “Detection of Synthetic Portrait Videos using Biological Signals” in arXiv: 1901.02212v2.
[6] Luisa Verdoliva. Media forensics and deepfakes: an overview. arXiv preprint arXiv:2001.06564, 2020.
[7] Martyn Jolly. Fake photographs: making truths in photogra phy. 2003.
[8] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In NIPS, 2014.
[9] David G¨uera and Edward J Delp. Deepfake video detection using recurrent neural networks. In AVSS, 2018.
[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In CVPR, 2016.
[11] An Overview of ResNet and its Variants: https://towardsdatascience.com/an-overview-of-resnet- and-its-variants-5281e2f56035
[12] Long Short-Term Memory: From Zero to Hero with Pytorch: https://blog.floydhub.com/long-short-term- memory-from-zero-to-hero-with-pytorch/
[13] Sequence Models And LSTM [22] F. Song, X. Tan, X. Liu, and S. Chen, “Eyes closeness detection from still images with multi-scale histograms of principal oriented gradients,” Pattern Recognition, vol. 47, no. 9, pp. 2825–2838, 2014. [23] D. E. King, “Dlib-ml: A machine learning toolkit,” JMLR, vol. 10, pp. 1755–1758, 2009. Networks https://pytorch.org/tutorials/beginner/nlp/sequence_mod els_tutorial.html
[14] https://discuss.pytorch.org/t/confused-about-the-image- preprocessing-in-classification/3965
[15] https://www.kaggle.com/c/deepfake-detection- challenge/data
[16] https://github.com/ondyari/FaceForensics
[17] Y. Qian et al. Recurrent color constancy. Proceedings of the IEEE International Conference on Computer Vision, pages 5459–5467, Oct. 2017. Venice, Italy.
[18] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros. Image-to- image translation with conditional adversarial networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 5967–5976, July 2017. Honolulu, HI.
[19] R. Raghavendra, Kiran B. Raja, Sushma Venkatesh, and Christoph Busch, “Transferable deep-CNN features for detecting digital and print-scanned morphed face images,” in CVPRW. IEEE, 2017.
[20] Tiago de Freitas Pereira, Andr´e Anjos, Jos´e Mario De Martino, and S´ebastien Marcel, “Can face anti spoofing countermeasures work in a real world scenario?,”in ICB. IEEE, 2013. [21] Nicolas Rahmouni, Vincent Nozick, Junichi Yamagishi, and Isao Echizen, “Distinguishing computer graphics from natural images using convolution neural networks,” in WIFS. IEEE, 2017.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)