e-ISSN:2582-7219

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 8, August 2024

INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.521

© 2024 IJMRSET | Volume 7, Issue 8, August 2024| DOI: 10.15680/IJMRSET.2024.0708037

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Deep Fake Image and Video Detection using Deep Learning

Rashmi B R, Ruchitha S, Ruchitha S, Vidhya S

Assistant Professor, Department of Computer Science Engineering, Rajarajeswari College of Engineering, Bangalore, Karnataka, India

U.G. Student, Department of Computer Science Engineering, Rajarajeswari College of Engineering, Bangalore, Karnataka, India

ABSTRACT: Strong detection systems are needed since deepfake content is becoming more and more common, which poses a serious threat to multimedia integrity. Here, we present a comprehensive method that tackles deepfake technology's generation and detection. Personalized and extremely realistic talking head models are produced from a small number of photos by our system by utilizing the latest developments in few-shot learning techniques. With the assistance of deep Convolutional Neural Networks (ConvNets) trained on an extensive video dataset, our system can generate convincing video sequences of face mimics and vocal expressions from a single photo. The system initializes the parameters of both the generator and discriminator in a person-specific manner through vast meta-learning and adversarial training, allowing for quick adaptability and training despite the complexity of the task. Based on this basis, we provide a unique deepfake detection framework that combines convolutional neural networks (CNNs) for modeling temporal dependencies and residual networks (Resnets) for extracting spatial features and long short-term memory (LSM). This hybrid architecture skillfully blends lstm-cnn's capacity to recognize dynamic facial expressions and movements across successive frames with resnet's strengths in capturing complex facial patterns and contextual information. Additionally , transfer learning techniques are used to improve model generalization, including pre-training on a dataset and fine-tuning on data unique to deepfake.

KEYWORDS: Residual Networks (ResNet), Convolutional Neural Networks (ConvNets), and Deepfake Detection.

I. INTRODUCTION

Despite its revolutionary potential deep learning has also facilitated controversial applications like deepfake technology deep learning a branch of machine learning has garnered a lot of attention for its ability to use artificial neural networks to derive valuable insights from complex data these networks characterized by their layered structures enable learning from intricate and unstructured datasets leading to significant advancements in various fields the term deepfake refers to the automated creation of synthetic media that frequently involves the replacement or alteration of elements in video content this technology has gained notoriety due to its potential for misuse including the spread of false or misleading information and cyberbullying to address these issues this project suggests an integrated system at the the foundation of our strategy is a face forensics model that combines create novel detection methods to address the social problems caused by deepfake manipulation our goal is to use our combined expertise in deep learning and picture forensics to support continuing efforts to mitigate the harmful effects of synthetic media conventional image forensic procedures with cutting-edge approaches designed to identify phony facial photos by utilizing knowledge from both fields the objective is to improve deepfake detections accuracy and dependability furthermore we present a convolutional method to identify indications of media manipulation or manipulation of convolutional neural networks by fusing these components into a coherent framework cnns who are renowned for their skill in image analysis are well-suited to detect anomalies or inconsistencies that point to deepfake interference we provide a strong solution for locating and preventing the spread of deepfake content through thorough assessment and validation we also show the efficacy and dependability of our approach in addressing the risks related to deepfake technology.

1.1CREATION OF DEEPFAKE

Traditionally large datasets of photographs of a single person are used to train convolutional neural networks CNN to produce very realistic human head images nonetheless our suggested method seeks to get over this restriction by utilizing few-shot learning strategies our system can learn from little data inputs and swiftly adjust to new people by doing meta-learning on a vast dataset of videos this makes it possible to create customized talking head models using a single image or a small number of image views our method initializes the generator and discriminator parameters in a person- specific way by including adversarial training procedures even with tens of millions of parameters to tweak this customized initialization makes training with sparse image data more efficient as a result our method facilitates quick convergence and adaptability guaranteeing the creation of incredibly lifelike talking head models for novel people the outcomes of our experiments show how well our strategy works to create customized talking head models that perform on par with more conventional techniques furthermore our method has a lot going for it in terms of efficiency and scalability which makes it a viable option for a lot of different applications like entertainment virtual assistants and human-computer interaction.

1.2DETECTION OF DEEPFAKE

Deepfake and face2face are two examples of the ai algorithms that have produced increasingly realistic fake face material thanks to the advancements in computer vision and deep learning these technologies alter face expressions or identities producing media that is nearly identical to authentic content even though these synthetic faces were first a de for amusement their improper use has given rise to serious issues and contributed to social discontent we suggest a hybrid face forensics framework to solve the issues brought up by the widespread use of altered media this framework detects and identifies altered information by utilizing the power of convolutional neural networks CNN a kind of deep learning model our method improves manipulation detection overall by integrating two different forensic techniques into a single CNN architecture through utilizing with insights from we offer a potent instrument to counter the propagation of falsified media and reduce the societal hazards connected with it all inside a unified CNN ramework cutting-edge approaches specifically designed to detect fakeface photos as well as traditional image forensic techniques our hybrid framework provides a comprehensive solution for manipulation detection our method can detect small signs and anomalies that point to manipulation even in very realistic fake face material since it integrates various forensic approaches our frameworks CNN based architecture makes manipulation detection effective and scalable which makes it appropriate for real-world applications where prompt and precise detection are critical we show the efficacy and resilience of our method in identifying altered material in a range of situations and contexts by means of thorough assessment and testing to sum up our hybrid face forensics framework is a major development in the manipulation detection space by merging the advantages of various forensic methodologies.

II. LITERATURE SURVEY

[1]The study by Jee-Young Sun et al. introduces a CNN-based approach for contrast enhancement (CE) forensics, outperforming traditional methods in forgery detection. By utilizing gray-Level Co- Occurrence Matrix (GLCM) features, their method shows enhanced accuracy, particularly against counter-forensic attacks. This highlights the effectiveness of CNNs in CE forensics, offering improved forgery detection capabilities.

[2]Andreas Rössler et al. introduce "FaceForensics: A Large-scale Video Dataset for Forgery Detection in Human Faces." Leveraging deep learning and Face2Face technology, they present a vast dataset of manipulated videos, surpassing existing collections significantly. This dataset comprises over 500,000 frames from 1004 videos, covering Source-to-target and Self-reenactment manipulations.

[3]The study by Jee-Young Sun et al. introduces a CNN- based approach for contrast enhancement (CE) forensics, outperforming traditional methods in forgery detection. By utilizing gray-Level Co- Occurrence Matrix (GLCM) features, their method shows enhanced accuracy, particularly against counter-forensic attacks. This highlights the effectiveness of CNNs in CE forensics, offering improved forgery detection capabilities.

[4]Andreas Rössler et al. introduce "FaceForensics:A Large- scale Video Dataset for Forgery Detection inHuman Faces." Leveraging deep learning and Face2Face technology, they present a vast dataset of altered videos, surpassing

existing collections significantly. This dataset comprises over 500,000 frames from 1004 videos, covering Source-to-target andSelf-reenactment manipulations.

[5]"MetaGAN: An Adversarial Approach to Few- Shot Learning" by Ruixiang Zhang et al. presents a novel framework for few-shot learning problems. Thisframework, MetaGAN, offers a simple yet effective method to boost the performance of few-shot learning models. MetaGAN stands out as a versatile and flexible solution for addressing the challenges of few-shot learning tasks.

[6]The study "Fake Face Detection Methods: Can TheyBe Generalized?" by A. Bromme et al. evaluates methods like Local Binary Patterns (LBP) and CNN models (e.g., AlexNet, ResNet50) for detecting fake faces. Despite not being trained for it, these models outperform other methods, indicating CNNs' potential for identifying fake faces even with evolving technology.

III. PROBLEM STATEMENT

Deepfake technologies quick development has made it a ubiquitous hazard threatening the veracity of multimedia material and giving rise to worries about disinformation and fraud existing detection systems are unable to keep up with the alarming rate at which deepfake techniques are evolving leaving digital media open to manipulation and abuse the task is made more difficult by the complex temporal dynamics present in video sequences and the minute spatial details of facial characteristics as a result new and reliable methods are needed for the timely and accurate detection of deepfake content it is vital to create practical solutions to stop the spread of deepfakes since failing to do so could jeopardize the integrity of digital media and erode public confidence in online information sources.

IV. EXISTING SYSTEM

One approach to detecting deepfakes leverages visible discrepancies in physical or physiological features within images or videos. Researchers examine aspects such as improper shadows, irregular geometry, or inconsistencies in facial details like teeth, ears, and eye colors to determine if content is synthetic orauthentic. Li et al. used the blink patterns of eyes in videos to detect abnormalities that might indicate deepfake content. Similarly, other methods focus on inconsistencies in head movement relative to body movement. These approaches aim to identify patterns that are difficult for deepfake generation tools to accurately mimic, providing a basis for distinguishingreal from synthetic content.

V. PROPOSED SYSTEM

Our suggested approach involves modeling temporal relationships using convolutional neural networks CNN and long short-term memory LSTM in conjunction with residual networks ResNet for spatial feature extraction this hybrid architecture seeks to combine the strengths of LSTM-CNN which can identify dynamic temporal changes in video sequences with ResNet skill in capturing complex spatial patterns we intend to use techniques for transfer learning pre-training the model on a variety of datasets at first to provide a strong basis is followed by fine-tuning on datasets unique to deepfake production to improve adaptability to changing deepfake generation methods our suggested approach seeks to overcome the shortcomings of current methods by providing a thorough and flexible resolution to the complex problems presented by deepfake identification.

VI. IMPLEMENTATION

6.1METHODOLOGY FOR DEEPFAKE CREATION

Photoshop and After Effects are utilized each day by experts, yet that doesn't imply that simply introducing both of them is everything necessary to make photorealistic pictures and recordings. In like manner, making reasonable face-traded recordings is hard. The primary endeavor of FakeApp pioneered deepfake created by a Reddit client utilizing autoencoder-decoder blending structure. To trade faces between source pictures what's more, target pictures, there is a need of two encoder-decoder sets where each pair is utilized to prepare on a picture set, and the encoder's parameters are shared between two system sets. In other words, two sets share the identical encoder network. The FakeApp software uses the AI Framework TensorFlow of Google, which in addition to other things was at that point utilized for the program DeepDream. There are additionally open-source options in contrast to the firstFakeApp program, as DeepFaceLab, FaceSwap and FakeApp. Regardless of whichever application we use to generate a deepfake the process involves mainly three steps.

•Extraction
•Training
•Creation

Extraction:
Deep learning which as we all know involves massive data sets is where the deep- in deepfake originates a deepfake video requires thousands of different images to be produced the procedure of removing all frames locating the face and lining them up is called the extraction process a crucial step in the process is alignment all of the faces should be the same size as the neural network is swapped out.

Training:
Training is a specialized term acquired from Machine Learning. For this situation, it alludes to the procedure which permits a neural system to change over a face into another. In spite of the fact that it takes a few hours, the preparation stage ought to be carried out just a single time. When finished, it can change over a face from individual A to individual

Creation:
When the training is finished, it is at last time to create a deepfake. Beginning from a video or an image, all casings are removed and all appearances are adjusted. At that point, everyone is changed over-utilizing the prepared neural system. The last advance is to consolidate the change over the face once again into thefirst casing. While this seems like a simple errand, it is really where most face-trade applications turn out badly. As already been told that autoencoders are employed to create a deepfake.

ARCHITECTURE OF DEEPFAKECREATION MODEL

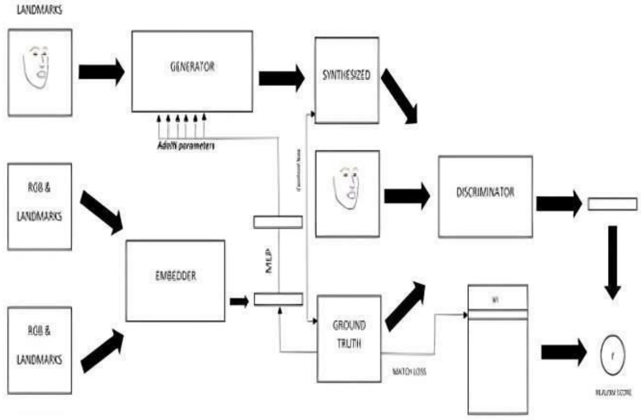

fig 6.1.1 A proposed creation model

When the training is finished, it is at last time to makea deepfake. Beginning from a video or an image, all casings are removed and all appearances are adjusted. At that point, everyone is changed over-utilizing the prepared neural system.

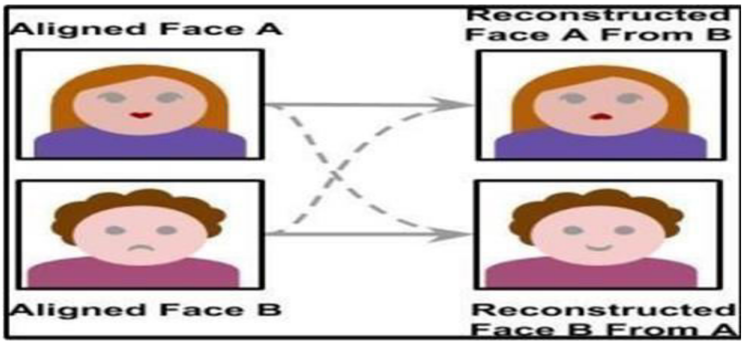

Fig 6.1.2 Training of an image from face A Face B

6.2 METHODOLOGY FOR DEEPFAKE DETECTION

Deepfakes are a kind of artificial media where a person's face is replaced in pre-existing images or videos with the likeness of another person. The underlying architecture frequently uses the encoder-decoder model, wherein encoder collects data from both the target desired face and the source original face photos. Given that deepfakes are often shorter in length, we split the video into two parts, and the decoder unit uses these features to create a fake video reconstruction of the desired face. High-level processing enhances the caliber of the video and removes visible artifacts, but subtle traces remain invisible to the unaided eye. Interestingly, these subtle traces are important characteristics for our detection model, which we suggest using the recursive neural network inceptionresnetv2 for feature extraction. Deepfakes use state-of-the-art face- manipulation techniques like generative adversarial networks, gans, and autoencoders, while our detection model focuses on identifying minute traces left behind during the process. This is an interesting field with both creative potential and moral dilemmas. into smaller frames and feed them as input to the detection model.

ARCHITECTURE FOR DEEPFAKE DETECTION MODEL

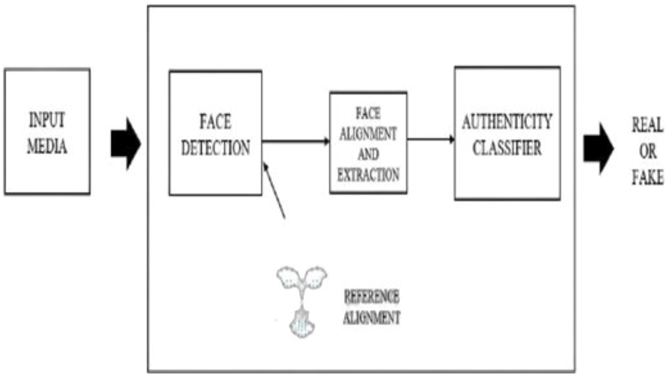

Fig 6.2.1 A proposed detection model

Dataset and Preprocessing:

The dataset was meticulously curated from various sources including the dataset for deepfake detection challenge on kaggle face forensics and celeb-deepfake forensics in total it comprises 401 videos interestingly these videos encompass both genuine footage and manipulated content the manipulation process involved paid actors altering real videos which were then transformed into deepfakes using various generator methods for our model we split the dataset 70 for training and 30 for testing during the training phase we provided the machine with labels corresponding to the video files crucially we pinpointed the exact frame where the original video transitioned into a deepfake this frame was then subjected to thorough analysis during preprocessing generally due to computational limitations we were only able to extract 147 frames from each video during preprocessing these frames were additionally subdivided into small batches for training and testing our goal to create an efficient deepfake detection system capable of identifying subtle manipulations within videos.

Modelling Model

For the system conducts an image categorizationanalysis on each frame extracted from the video. We Used a pertained CNN model named Inceptin ResNetV2 [12] and RNN along with LSTM. We also need to define Loss function, Optimizer and other Hyper-parameters required for the training procedure.Depending on the state of training model, the learningrate should be adjusted to minimize the loss value.

i)Face detection

Given an input image, the facial zone is detected by a neural facial landmark detection model that automatically localizes the 68 fiducial facial landmark points around facial components and facial contours such as eyes, mouth, and chin. Among those points, only 51 points are used excluding 17 points from chin because facial manipulation is performed inside the inner facial area.

ii)Face alignment and extraction

Then, the system aligns the face to fit the reference alignment because faces appearing in media are rarely frontal or unrotated. We apply the affine transformation on the image by finding the one-to-onemapping from the extracted landmark points to the reference alignment points. Through affine transformation, rotated or profile faces can be alignedaccording to the reference alignment, which helps to enhance the fake face detection performance. Finally, the system crops the facial region from the image and feed it to the facial authenticity classifier.

iii)Authenticity classification

The proposed face authenticity classifier combines content feature extractor (CFE) and trace feature extractor (TFE). A convolution is depicted by a square containing its detail in the two feature extractors. For example, the first convolution in the CFE has 7 × 7 convolutional filter with stride 2 and outputs 64 feature maps.

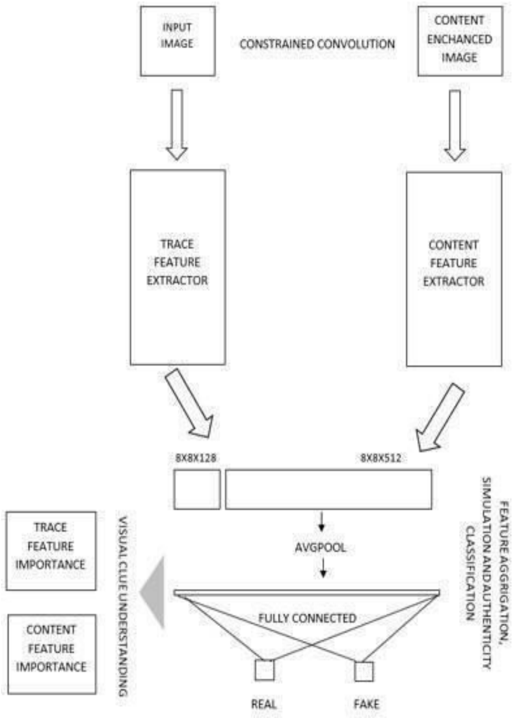

Fig 6.2.2 Authenticity Classifier

VII. CONCLUSION

Advances in artificial intelligence, particularly in deep learning, have given rise to deepfake technology, which allows highly realistic multimedia material to be generated with the intention of misleading viewers. There are a numerous risks associated with this, such as false information, damage to one's reputation, and even threats to national security. The importance of creating efficient detection techniques increases with the threat of deepfake. The new deepfake detection method presented in this paper integrates Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Residual Networks (ResNet). By capturing both geographical and temporal data, this hybrid design overcomes the drawbacks of previous methods. In order to enhance generalization and make the model more flexible in response to the quickly evolving deepfake environment, transfer learning is utilized. Deepfakes have made it harder to distinguish between authentic and altered content, which has damaged public confidence in the media. They could provide energy. The hazards posed by political instability, hate speech, and misinformation are exacerbated by social media's quick spread. Deepfakes can cause havoc by focusing on particular demographics rather than requiring a huge audience. Consequently, media literacy and the capacity to recognize altered materials become extremely important, particularly in court situations when digital content is presented as proof. Using a convolutional neural network, the suggested hybrid forensic framework combines face-specific analysis with general-purpose image forensics offering a strong defense against the deepfake danger. The approach's validity is confirmed by the experimental results, which emphasize the necessity of continuous research in deepfake detection to safeguard people's privacy and maintain global, political, and societal security.

Future research on the suggested model attempts to decrease computational expenses and increasing velocity. This could entail processing high-resolution photos without downscaling them using neural networks equipped with global pooling. Furthermore, investigating various image upscaling techniques and assessing their effects on performance may provide insightful information for improving accuracy while balancing the trade-off with time latency.

REFERENCES

[1]Oleg Alexander, Mike Rogers, William Lambeth, Jen- Yuan Chiang, Wan-Chun Ma, Chuan-Chang Wang, and Paul Debevec. The Digital Emily project: Achieving a photorealistic digital actor.
[2]Antreas Antoniou, Amos J. Storkey, and Harrison Edwards. Augmenting image classifiers using data augmentation generative adversarial networks. In Artificial Neural Networks and Machine Learning -
ICANN, pages 594–603, 2018. 2
[3]Sercan Arik, Jitong Chen, Kainan Peng, Wei Ping, and Yanqi Zhou. Neural voice cloning with a few samples. In Proc. NIPS, pages 10040–10050, 2018. 2
[4]Hadar Averbuch-Elor, Daniel Cohen-Or, Johannes Kopf, and Michael F Cohen. Bringing portraits to life. ACM Transactions on Graphics (TOG), 36(6):196, 2017. 1, 14
[5]Facebook Wants to Stay 'Neutral' on Deepfakes. Congress Might Force it to Act. Accessed: Jun. 14, 2019. [Online]. deepfakes-congress-house-hearing
[6]A. K. Jain, A. Ross, and S. Pankanti, ''Biometrics: A tool for information security,''
[7]A. K. Jain, A. Ross, and S. Prabhakar, ''An introduction to biometric recognition,''

INTERNATIONAL JOURNAL OF
MULTIDISCIPLINARY RESEARCH
IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com