IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 8, August 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.625

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

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

Deep Fake Face Detection

Mrs. Lipika Rajanandini Sahu[1], Mrs. T Rathidevi[2], Kirana S[3], Manasa J[4], Meghana C V[5], Dr.J.Amutharaj[6]

Department of Information Science and Engineering, Rajarajeswari College of Engineering, Bangalore, Karnataka, India[1,2,3,4,5,6].

ABSTRACT: In light of the escalating threat assailed by the widespread distribution of deepfake content, this paper presents a robust method for detecting such material using a hybrid architecture. This approach combines Residual Networks (ResNet) to extract spatial features and Long Short-Term Memory (LSTM) with Convolutional Neural Networks (CNN) to model temporal dependencies. The ResNet component is adept at identifying complex patterns in facial and contextual information, while the LSTM-CNN module focuses on recognizing dynamic facial expressions and movements across frames. To enhance the model's ability to generalize, transfer learning techniques are employed. This involves pre-training utilizing a huge dataset and then fine-tuning it on deepfake-specific data. The effectiveness of this hybrid architecture is demonstrated through experimental evaluations on various deepfake datasets, showing superior performance regarding precision, and recall. This highlights its capability to address the evolving challenges posed by the increasingly sophisticated techniques used in generating deepfake content**.**

KEYWORDS: Deepfake Detection, Convolutional Neural Network (CNN), Residual Network (ResNet), Long Short-Term Memory (LSTM).

I. INTRODUCTION

The development of hyper-realistic multimedia content that can fool human perception has been made possible by deepfake technology, which has emerged as a result of the rapid advancement of artificial intelligence, especially in the field of deep learning. Deepfakes that are intentionally used to contain different variety of negative effects, from reputational harm and disinformation to possible dangers to national security. It is therefore essential to build efficient deepfake detection techniques as a response to this growing threat. The strengths of Residual Networks (ResNet) and Long Short-Term Memory (LSTM) with Convolutional Neural Networks (CNN) are combined in this paper to produce a hybrid architecture that excels in capturing both spatial andthorough spatial comprehension provided by ResNet in order to overcome the shortcomings of the current deepfake detection methods. Conventional approaches frequently struggle to detect tiny changes in facial characteristics and fail to capture the temporal dynamics inherent in video sequences because to the exponential development in sophistication of deepfake images. As a result, our method not only combines these two potent neural network architectures but also uses transfer learning techniques to improve generalization, allowing the model to adjust to the vast array of dynamic deepfake production techniques. The architecture, technique, and experimental findings are described in detail in the following sections, which also show how effective our suggested solution is at tackling the difficult problem of deepfake detection.

1.1 CREATION:

Convolutional Neural Networks (CNNs) can be trained to produce highly accurate images of human heads. Typically, creating a customized talking head model requires training on a vast quantity of pictures of a single person. However, in many practical situations, these models need to be trained with only one or a few images of the subject. Figure 1.1 demonstrates how few-shot and one-shot learning of neural talking head models for previously unknown individuals is framed as an adversarial training problem, involving high-capacity generators and discriminators. This is accomplished through extensive meta-learning on a large dataset of films. The system's ability to individually initialize the settings for both the discriminator and generator is crucial, as it enables rapid training even when millions of parameters need to be adjusted based on a limited number of images.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

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

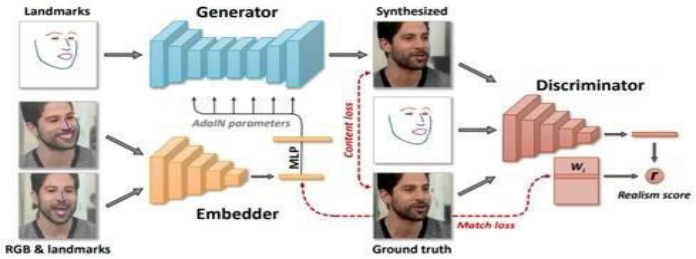


Figure.1.1 Block Diagram For Deepfakes Creation

expressions and identities. While many of these fake faces are created for entertainment, their misuse has caused significant social concern. A hybrid face forensics system, based on a convolutional neural network (CNN), integrates two forensics techniques to improve manipulation detection performance. The proposed model, illustrated in Figure 1.2, is a type of CNN that simultaneously extracts content features and traces attributes from a face image by combining two distinct feature extractor types. Figure 1.2 depicts this model, which improves its ability to detect manipulation by leveraging these dual feature extractors.

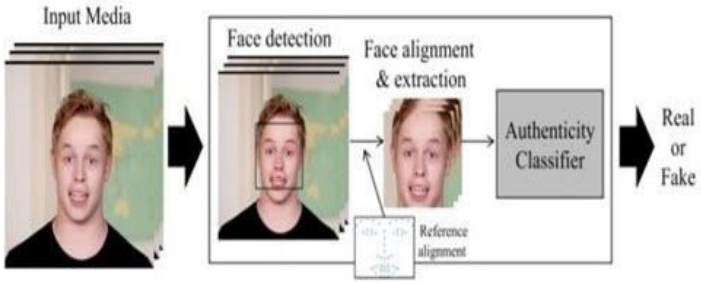


Figure 1.2 Block Diagram For Deepfakes Detection

1.2 DETECTION:
Advances in computer vision and deep learning have led to the rising popularity of AI-generated realistic-looking fake face media, such as Face2Face and Deepfake, which can alter facial

II. RELATED WORK

[1] In this study, Jee-Young Sun et al. surpass conventional techniques in the identification of forgeries by introducing a CNN-based strategy for contrast enhancement (CE) forensics. Their approach demonstrates improved accuracy, especially against counter-forensic attacks, by leveraging gray-Level Co-Occurrence Matrix (GLCM) characteristics. This demonstrates how CNNs work well in CE forensics by providing enhanced forgery detection capabilities.

[2] In their introduction, Andreas Rössler and colleagues present "Face Forensics: A Comprehensive Video Dataset for Human Face Forgery Detection." By utilizing Face2Face and deep learning, they provide an extensive dataset of edited videos that greatly outpaces current compilations. This dataset, which includes Source-to-target and Self-reenactment alterations, is made up of more than 500,000 frames from 1004 films.

[3] Ruixiang Zhang et al.'s "MetaGAN: An Adversarial Approach to Few-Shot Learning" offers a novel paradigm for few-shot learning issues. This approach, called MetaGAN, provides a straightforward yet efficient way to improve few-shot learning models' performance. One particularly adaptable method for handling the difficulties of few-shot learning problems is MetaGAN.

[4] A. Bromme et al.'s research "Fake Face Detection Methods: Can They Be Generalized?" assesses techniques for

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

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

identifying fake faces, including CNN models (e.g., Alex Net, ResNet50) and Local Binary Patterns (LBP). These models outperform other approaches despite not being specifically trained for it, suggesting that CNNs have the ability to recognize phony faces even as technology advances.

PROBLEM STATEMENT

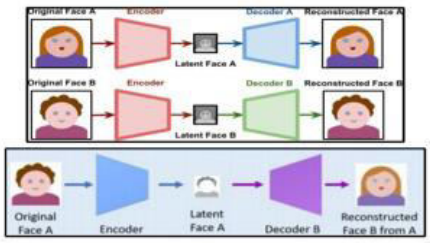
The widespread use of deepfake technology presents a significant risk to the legitimacy of multimedia content, hence requiring reliable methods for identifying altered photos and videos. The increasing complexity of deepfake generation techniques frequently outpaces current deepfake detection tools. The temporal dynamics of video sequences and the spatial complexities of facial features pose tough obstacles that necessitate creative solutions for the timely and accurate detection of deepfake content.

III. METHODOLOGY

The suggested approach combines Long Short-Term Memory (LSTM) with Convolutional Neural Networks (CNN) to represent temporal dependencies and integrates Residual Networks (ResNet) for spatial feature extraction. The goal of this hybrid architecture is to combine the advantages of LSTM-CNN's skill in identifying dynamic temporal changes in video sequences with ResNet's capacity to record complex spatial patterns in a synergistic manner. The model will be pre-trained using a variety of datasets using transfer learning algorithms to provide a solid foundation. It will then be fine-tuned on datasets specifically designed for deepfake production to increase its adaptability to changing deepfake generation techniques. The suggested approach aims to address the shortcomings of current methods by offering a more thorough and flexible response to the complex problems associated with deepfake identification.

Experts use Photoshop and After Effects on a daily basis, but it doesn't mean that installing both programs is all that's needed to create convincing images and videos. Similarly, it is difficult to make reasonable face-traded records. The final product, as with any creative endeavor, is a combination of aptitude, responsibility, and appropriate instruments. The first attempt at creating a deepfake was Fake App, which was made by a Reddit user using the autoencoder-decoder blending structure. According to that method, the autoencoder extracts inactive highlights from facial images, and the decoder is used to recreate the images. Two encoder-decoder sets are required to exchange faces between source and destination images. Each pair is used to prepare an image set, and the encoder's parameters are shared between the two system sets. To put it simply, the encoder network of two pairs is identical. The AI Framework TensorFlow from Google, which was used before the Deep Dream program among other things, is used by the FakeApp software. Unlike the original FakeApp program, there are also open-source alternatives, such as DeepFaceLab, Face Swap (now hosted on GitHub), and FakeApp (currently hosted on Bitbucket). It as three processes involved in creating a deepfake.

- Extraction
- Training
- Creation



www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

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

Extraction:
Huge data sets are necessary for deep learning, which is where the "deep" in "deepfake" originates. To make a deepfake video, thousands of different images are needed. The procedure of removing every frame, locating the face, and aligning them is known as the extraction process. Alignment is a difficult step in the process; all of the faces should have the same size and the neural network is used to switch.

Training:
A specific word from machine learning is "training." In this case, it refers to the process that allows a brain system to switch from one face to another. Even though it requires several hours, the preparatory step should only be carried out once. When it's done, it can switch the face of individual A with that of individual B.

Creation:
When the training is finished, it is at last time to make 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 the first casing.

3.1.1Architecture of Deepfake Creation Model

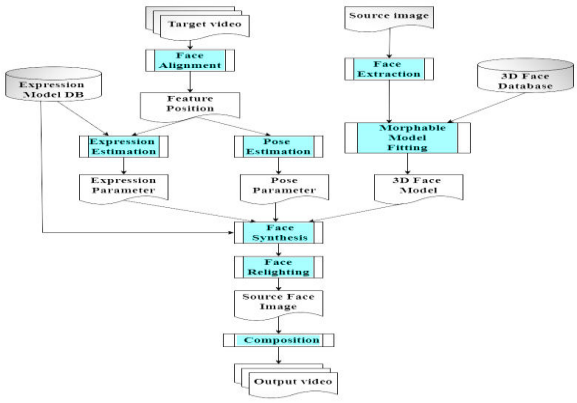


Figure 3.1.1 A proposed creation model

The Generator network, also known as the embedder network, is depicted in Figure 6.1.1. It represents the last step in creating a deepfake after training is finished. At this stage, all external features are removed, and appearances are altered based on a video or photograph. The readymade brain system is then used to change everyone's face. Finally, the face change is consolidated back into the original casing, completing the process.

The fundamental architecture used to create deepfake is an encoder-decoder architecture, in which the encoder obtains the features of the source and target faces, is to assist the decoder in obtaining the encoded feature of the target face in order to produce fake video. High-level processing is used to improve the video's quality and remove leftovers, although certain traces remain that are invisible to the unaided eye. The salient characteristics of our detection model are these remaining traces. InceptionResnetV2 is a component of the suggested model used for feature extraction. A recurrent neural network designed to determine whether or not a video has been altered is trained using these extracted features. Since there is minimal manipulation of the video, the deepfakes occur more quickly. As a result, the video is divided into smaller frames, which are then fed into the detection model.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

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

3.2.1 Architecture of Deepfake Detection Model

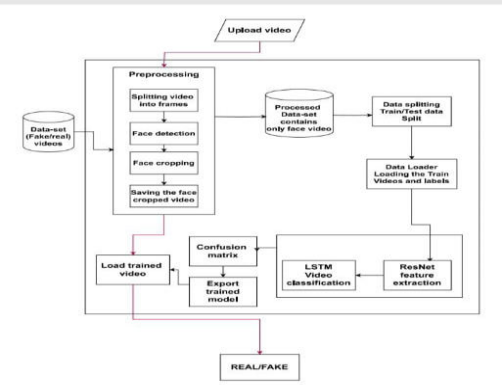


Figure 3.2.1 A proposed detection model

Dataset and Preprocessing:
The dataset was gathered from the Kaggle, Face Forensics, and Celeb-deepfake forensics deepfake detection challenge datasets. There are about 6458 videos in it. These videos also feature actual footage that was altered further by hired actors and turned into a deepfake using various techniques fordeepfake generators. Thirty percent of the dataset was used for system testing, while seventy percent was used for training. Labels for the video files that were supplied to the system during the training phase were also provided into the machine. When a video is transformed into a deepfake, this point is recorded as a frame, which is subsequently examined during the preprocessing stage. Preprocessing of a video extracts an average of 147 frames. We trained the model using a restricted amount of frames because of our low computing capability. Frames are sent on for training and testing in tiny batches following preprocessing.

Modeling Model:
Every frame that is taken from the video is subjected to an image categorization analysis by this system. We combined LSTM with RNN and a pretrained CNN (Convolutional Neural Network) model called InceptinResNetV2. Determining the Loss function, Optimizer, and additional hyper-parameters necessary for the training process are also necessary. To reduce the loss value, the learning rate needs to be modified based on the training model's current state.
The proposed fake face detection framework consists of

(i) Face detection
(ii) Alignment and extraction of the face and
(iii) Authenticity classification.

Face detection: A neural facial landmark detection model automatically identifies the 68 fiducial points around facial components and features, such as the mouth, chin, and eyes, in order to detect the facial region given an input image. Just 51 of those points are utilized, with the exception of 17 points from the chin, as facial manipulation is done inside the inner region of the face.

Alignment and extraction of the face: Since faces in media are rarely frontal or unrotated, the algorithm then aligns the face to meet the reference alignment We apply the affine transformation to the image by finding the one-to-one mapping from the retrieved landmark points to the reference alignment points. The performance of fake face detection can be improved by aligning rotated or profiled faces with respect to the reference alignment using affine transformation. Ultimately, the facial region of the image is cropped by the system and fed into the facial authenticity classifier as shown in the figure below.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

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

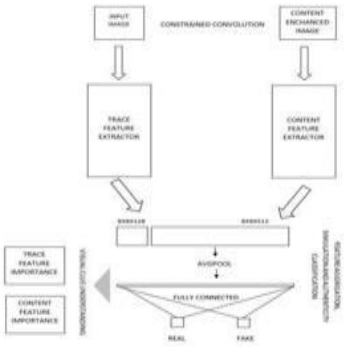


Figure 3.2.2 Authenticity Classifier

Authenticity classification: content feature extractor (CFE) and trace feature extractor (TFE) are combined in the suggested face authenticity classifier. A convolution is represented in the two feature extractors by a square that contains its detail. For instance, the CFE's (content feature extractor) initial convolution produces 64 feature maps and features a 7 × 7 convolutional filter with stride 2.

IV. EXPERIMENTAL RESULTS

The suggested model is an architecture based on Convolutional Neural Networks (CNNs) and intended for the identification of deepfake images. For feature extraction and dimensionality reduction, it uses a sequence of convolutional and pooling layers, respectively. These layers preserve computational efficiency and guard against overfitting while allowing the model to capture complex patterns typical of deepfake images. The flattened layer prepares the feature maps for input into fully connected layers by converting them into a one-dimensional array following feature extraction. In order to guarantee consistent input batch distributions, accelerate convergence, and stabilize the training process, batch Normalization is utilized. Data augmentation techniques include rescaling, rotating, flipping, shearing, and zooming. Strategies that are employed to enhance the model's resilience and generality.

V. CONCLUSION

Since deepfakes are created at a distinct level of abstraction, it has always been difficult to identify them. The issue of categorization has traditionally been viewed as having two possible solutions: actual and deepfake class labels. CNN is a well-known method for identifying deepfake photos. This is what inspired us to propose in this study a CNN-based architecture for the detection of deepfake images. The proposed architecture delivers 97.2% accuracy for deepfake photographs, using images from 5 different data sources and 2 different data sources for real images. Despite the significant disparity in image resolution, the suggested design offers a well-balanced performance across all data sources. The approach can be expanded upon to include the classification of deepfake video footage. Using each extracted video frame, the face is identified, cropped, and supplied to the model to identify deepfake manipulations, this model may be utilized for video deepfake detection. Making a pipeline to process this will make it simple to accomplish this video information. Therefore, with all data augmentation approaches implemented, the suggested CNN-based model performs well and has a fairly balanced performance over the supplied dataset. Moreover, it exhibits strong performance and generalizability over different test sets that have not been encountered.

REFERENCES

[1] Antreas Antoniou, Amos J. Storkey, and Harrison Edwards. Enhancing image classifiifiers with a data augmentation generative adversarial networks. In Artifificial Neural Networks and Machine Learning - ICANN, pages 594–603, 2018.
[2] SercanArik, Jitong Chen, KainanPeng, Wei Ping, and Yanqi Zhou. Neural voice cloning with a few samples. In Proc. NIPS, pages 10040–10050, 2018.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

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

[3] HadarAverbuch-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

[4] Facebook Wants to Stay 'Neutral' on Deepfakes. Congress Might Force it to Act. Accessed: Jun. 14, 2019. [Online]. Available: https://www.vox.com/future-perfect/2019/6/13/18677574/facebook-zuckerbergdeepfakes-congress-house-hearing .

[5] A. K. Jain, A. Ross, and S. Pankanti, ''Biometrics: A tool for information security,'' IEEE Trans. Inf. Forensics Security, vol. 1, no. 2, pp. 125–143, Jun. 2006.

[6] A. K. Jain, A. Ross, and S. Prabhakar, ''An introduction to biometric recognition,'' IEEE Trans. Circuits Syst. Video Technol., vol. 14, no. 1, pp. 4–20, Jan. 2004.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details