International Journal of Research Publication and Reviews

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

Deepfake Classification for Image Identification

Atharva Thakkar[1], Aman Vohra[2], Yash Patil[3], Pushkraj Salunke[4], Assistant Professor Shital Bhandare[5]

[1, 2, 3, 4] UG Student, [5]Assistant Professor
[1, 2, 3, 4, 5] Department of Computer Engineering,
K. K. Wagh Institute of Engineering Education and Research,
atharvathakkar@yahoo.com[1], vohraaman2002@gmail.com[2], yashp030303@gmail.com[3], pushkrajsalunke8668@gmail.com[4], ssbhandare@kkwagh.edu.in[5]

ABSTRACT:

In today's digital age, the emergence of deepfake technology has raised concerns about the authenticity of images and videos shared online. To address this challenge, we propose a new method to detect deepfakes using advanced computer algorithms. These algorithms, inspired by how our brains learn, analyze visual content to identify subtle signs of manipulation. Our focus is on understanding how things change over time in videos and images, as this is where deepfakes often reveal themselves. While our approach has not yet been tested, we recommend using the FaceForensics++ dataset for future evaluations. This dataset contains a wide variety of deepfake content, making it an ideal resource for testing our method. By harnessing the power of these intelligent computer models, we hope to make it easier to distinguish between real and fake media online, thus contributing to a more trustworthy digital environment for all users.

Keywords: Deepfakes, Deep learning, Computer algorithms, Image and video manipulation, Detection methods, FaceForensics++ dataset

Introduction:

In today's digital age, the increased use of artificial intelligence has made it easier and faster to generate deepfake content, especially in manipulating images and videos. This deepfake generation relies on advanced concepts of deep learning and machine learning. However, with this ease of creation comes a significant risk of misinformation, fraud, and threats to society. In response, our proposed method focuses on developing a robust approach for identifying deepfake images.

Our goal is to design and propose a deep learning solution that can accurately differentiate between genuine and manipulated images. To achieve this, we suggest utilizing neural network architectures, commonly known as deep-learning models, for both feature extraction and temporal analysis. By harnessing these technologies, we aim to detect the discrepancies present in deepfake images, enabling more effective detection and classification. A crucial aspect of our proposed method is the suggestion of using diverse datasets comprising both real and deepfake images. This approach ensures that any future model developed based on our method would be robust and adaptable by being exposed to a wide range of visual variations and manipulation techniques. Through further research, testing, and experimentation by the wider community, refinements of the model's parameters can optimize performance metrics such as accuracy, precision, and recall.

Additionally, our proposed method not only focuses on addressing technical challenges associated with identifying deepfakes but also aims to raise awareness in society. By combating misinformation and other threats, we aspire to preserve the integrity of digital communication.

While we have not implemented a system yet, our proposed method offers a promising avenue for future research and development in the field of deepfake detection.

LITERATURE SURVEY:

- Yogesh Patel, Sudeep Tanwar, Pronaya Bhattacharya, Rajesh Gupta, Turki Alsuwian, Innocent Ewean Davidson, and Thokozile F. Mazibuko [1] "An Improved Dense CNN Architecture for Deepfake Image Detection" presents a significant advancement in computer vision technology, specifically targeting the detection of deepfake images generated through generative adversarial networks (GANs). The authors address the pressing need for reliable tools to identify synthetic media that can manipulate public opinion and spread misinformation. Their proposed deep convolutional neural network (D-CNN) architecture stands out for its ability to capture inter-frame dissimilarities in media streams by training on images from multiple sources, thereby enhancing its generalizability. Leveraging techniques like binary-cross entropy and Adam optimizer, the model achieves impressive accuracy rates ranging from 94.67% to 99.33% across various deepfake datasets. By meticulously analyzing existing methodologies and introducing their innovative D-CNN approach, Patel et al. significantly contributes to the advancement of deepfake detection, offering a solution that addresses key challenges such as robustness, interpretability, and cross-domain applicability.

- Van-Nhan Tran, Bo-Sung Kim, Suk-Hwan Lee, Ki-Ryong Kwon, Hoanh-Su Le Explored Generalization Deepfake Detection for Face Forgery.[2] Tran et al.'s "Generalization Deepfake Detector" addresses the urgent need to discern manipulated facial media amidst the proliferation of deepfake technology. With the rise of techniques like Generative Adversarial Networks (GANs), producing convincing fake images and videos has become alarmingly easy, undermining trust in digital content. Conventional deepfake detection models, predominantly based on convolutional neural networks (CNNs), excel within specific datasets but struggle when confronted with unseen domains, where their performance plummets. In response, the authors propose GDD, a model designed to generalize across diverse domains without necessitating updates. They introduce novel loss functions and leverage meta-learning to enhance model adaptability. By harnessing various datasets and employing domain separation strategies, GDD exhibits promising efficacy in detecting manipulated content across varying contexts. This research contributes to the burgeoning efforts aimed at mitigating the societal risks associated with deepfake technology by advancing the capabilities of face forgery detection models.
- Alben Richards MJ, Prakash P, Kaaviya Varshini E, Kasthuri P, Diviya N Explored Deep Learning Techniques for Deepfake Face Detection [6] In their research on "Deep Fake Face Detection using Convolutional Neural Networks," Richards et al. tackle the growing concern of deepfake technology, which produces highly realistic but fabricated media, posing significant threats like identity theft and misinformation. They develop a Convolutional Neural Network (CNN)-based model to discern manipulated facial images. Leveraging the power of CNNs, the model learns to differentiate between real and fake faces by extracting features like texture, shape, and color. Through extensive training on datasets comprising both authentic and manipulated images, the model achieves promising results, demonstrating high accuracy and effectiveness in identifying deepfake content. By employing deep learning techniques and comprehensive datasets, the study contributes to the ongoing efforts to combat the adverse effects of deepfake technology, highlighting the importance of robust detection methods in safeguarding against digital manipulation and preserving trust in media content.
- "Jiaxin Ai, Zhongyuan Wang, Baojin Huang, Zhen Han, Qin Zou's paper [8] 'Deepfake Face Provenance for Proactive Forensics' tackles the rising concern of deepfake technology by proposing a proactive approach to forensic analysis. Unlike traditional methods that focus solely on detecting the authenticity of deepfake images, the authors delve into the innovative realm of deepfake inversion. By leveraging a disentangling reversing network, they aim to trace the original faces behind manipulated images, thereby enhancing the interpretability and traceability of evidence. This pioneering work opens up new avenues for addressing the challenges posed by deepfake technology, emphasizing the importance of proactive measures in forensic analysis and highlighting the need for further research in this domain."

PROPOSED SYSTEM:

ALGORITHM:

1. ***Start: Begin the process.***
2. ***Collection of Datasets:*** *Gather a diverse collection of datasets containing both genuine and deepfake videos. For example, utilize datasets like FaceForensics++ which contains 6000 videos.*
3. ***Preprocessing of Videos:***
    - *Detect faces within the videos using face detection algorithms.*
    - *Split each video into frames.*
    - *Crop the faces from each frame to focus on facial features.*
    - *Reconstruct the videos by merging the cropped faces, ensuring each frame contains a face.*
    - *Select a desired number of frames from each video (e.g., 150 frames) to maintain consistency and reduce computational complexity.*
4. ***Development of Model Using Deep Learning Models:***
    - *Design a deep learning model utilizing convolutional neural networks (CNN) for feature extraction and deep learning models for temporal analysis.*
5. ***Training the Model:***
    - *Train the deep learning model using the preprocessed dataset collected in step 3.*
6. ***Testing the Model:***
    - *Evaluate the performance of the trained model using a separate testing dataset.*
7. ***Perform Prediction on Specific Video:***
    - *Provide an unseen video to the trained model for prediction.*
8. ***Calculation of Accuracy:***
    - *Calculate the accuracy of the model by comparing the predicted labels with the ground truth labels for the testing dataset.*
9. ***Display Results:***

- *Display the results of the classification of the video (i.e., whether it is identified as genuine or deepfake) along with the accuracy achieved by the model.*

10. Stop: *End the process.*

This algorithm outlines the step-by-step procedure for preprocessing videos, training a deep learning model using CNNs and models for temporal analysis, testing the model's performance, making predictions on unseen videos, calculating accuracy, and displaying the classification results.

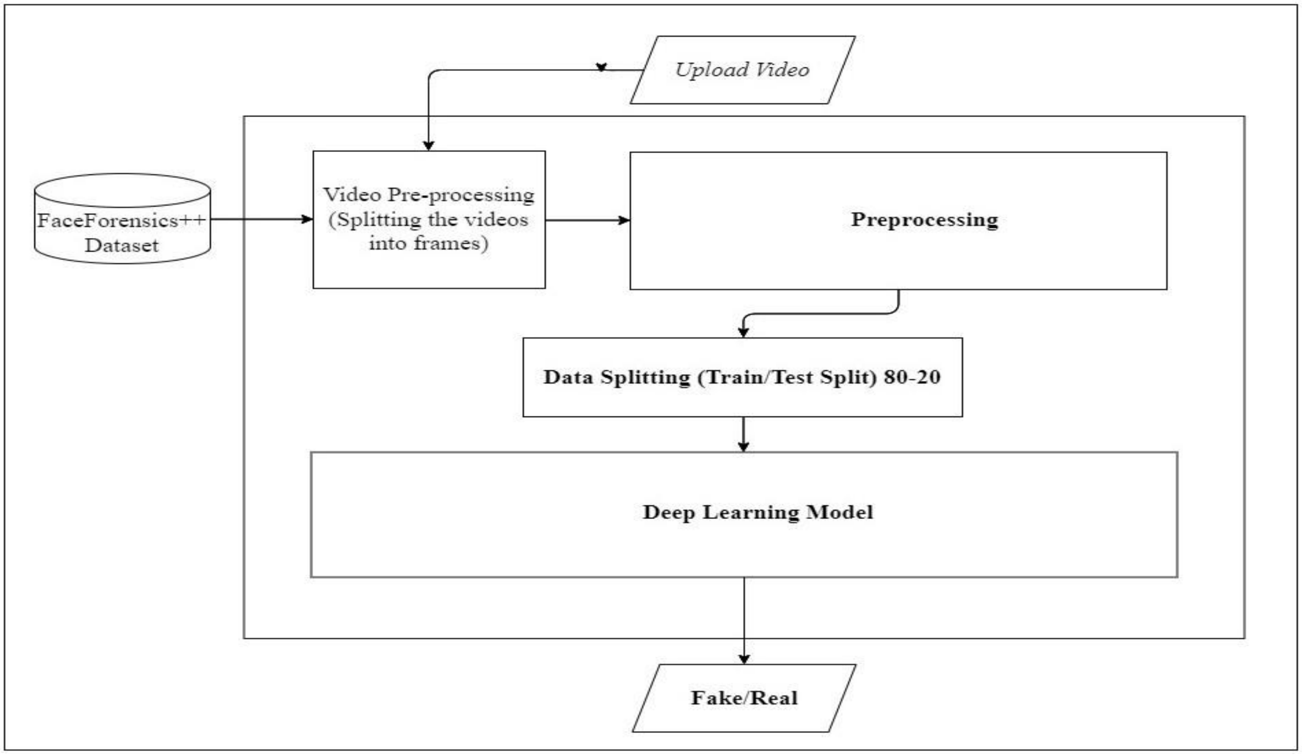
Upload Video
FaceForensics++ Dataset
Video Pre-processing (Splitting the videos into frames)
Preprocessing
Data Splitting (Train/Test Split) 80-20
Deep Learning Model
Fake/Real

Figure 1. System architecture diagram

Architecture is designed to effectively tackle the challenges posed by deepfake detection, employing a comprehensive approach that encompasses data preprocessing, model training, and prediction.

A. Dataset Used:

The system will utilize the FaceForensics++ dataset, ensuring a diverse representation of both original and manipulated deepfake videos.

B. Video Processing (Splitting the Videos into Frames):

Videos will be split into frames to facilitate frame-level analysis, an essential step in preprocessing.

C. Preprocessing of Video Frames:

Face detection algorithms will be applied to identify and crop facial regions within each frame, ensuring consistency. Frames lacking detected faces will be discarded, with computational constraints limiting training to the first 100 frames of each video.

D. Data Splitting for Testing and Training:

The processed frames will be aggregated back into videos and split into an 80% training set and a 20% test set for future experimentation.

E. Deep Learning Model Building:

The model architecture will comprise a deep learning model backbone followed by a layer for temporal analysis. Frames from the processed videos will be fed into the model in mini-batches for both training and testing.

F. Prediction:

During the prediction phase, new videos will undergo preprocessing to align with the trained model's input format. Frames will be extracted, faces will be cropped, and the processed frames will be inputted into the model for efficient and accurate detection of deepfake content.

Dataset Used:

Table1. FaceForensics++ [Ref: Rossler et al. 2019. "FaceForensics++: Learning to Detect Manipulated Facial Images."] https://arxiv.org/abs/1901.08971

Classes	**Total No. of Videos**	**Train (80%)**	**Test (20%)**	**Total Frames**	**Image Size**	**Video Quality**
DeepFakes	1000	800	200	148	112	C24
FaceShifter	1000	800	200			
FaceSwap	1000	800	200			
NeuralTextures	1000	800	200			
Face2Face	1000	800	200			
Original	1000	800	200			
Total	***6000***	***4800***	***1200***			

Figure 2. Attributes of FaceForensics++ Dataset

Conclusion

In conclusion, our proposed system outlines a comprehensive approach to address the challenge of detecting deepfake videos. By leveraging preprocessing techniques and deep learning models, we aim to preprocess videos, extract facial features, and analyze temporal patterns. Although the system has not been implemented yet, the proposed methodology holds promise in distinguishing between genuine and manipulated videos. Through future experimentation and refinement, we anticipate that our proposed approach will contribute significantly to the ongoing efforts to combat misinformation and maintain the integrity of digital media.

REFERENCES:

Research Papers:

1. Patel, Y., Tanwar, S., Bhattacharya, P., Gupta, R., Alsuwian, T., Davidson, I. E., & Mazibuko, T. F. (2023). An Improved Dense CNN Architecture for Deepfake Image Detection. IEEE Access, 10.1109/ACCESS.2023.3251417
2. Tran, V. N., Kim, B. S., Lee, S. H., Kwon, K. R., & Le, H. S. (2023). Learning Face Forgery Detection in Unseen Domain with Generalization Deepfake Detector. In 2023 IEEE International Conference on Consumer Electronics (ICCE) (pp. 1-6). IEEE. DOI: 10.1109/ICCE56470.2023.10043436
3. Theerthagiri, P., & Nagaladinne, G. B. (2023). Deepfake Face Detection Using Deep InceptionNet Learning Algorithm. In 2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS) (pp. 1-6). IEEE. DOI: 10.1109/SCEECS57921.2023.10063128
4. Vajpayee, H., Jhingran, S., Yadav, N., & Raj, A. (2023). Detecting Deepfake Human Face Images Using Transfer Learning: A Comparative Study. In 2023 IEEE International Conference on Contemporary Computing and Communications (InC4) (pp. 1-6). IEEE. DOI: 10.1109/InC457730.2023.1026321
5. Khan, H. A., Alwakid, G. N., Tehsin, S., & Humayun, M. (2023). Detection of Facial Forgery in Digital Images. In 2023 International Conference on Business Analytics for Technology and Security (ICBATS) (pp. 1-6). IEEE. DOI: 10.1109/ICBATS57792.2023.10111318
6. Richards, A. M. J., Prakash, P., Varshini, K., Kasthuri, P., Diviya, N., & Sasithradevi, A. (2023). Deep Fake Face Detection using Convolutional Neural Networks. In 2023 12th International Conference on Advanced Computing (ICoAC) (pp. 1-6). IEEE. DOI: 10.1109/ICoAC59537.2023.10250107
7. Guarnera, L., Giudice, O., & Battiato, S. (2020). Fighting Deepfake by Exposing the Convolutional Traces on Images. IEEE Access, 10.1109/ACCESS.2020.3023037
8. Ai, J., Wang, Z., Huang, B., Han, Z., & Zou, Q. (2023). Deepfake Face Provenance for Proactive Forensics. In 2023 IEEE International Conference on Image Processing (ICIP) (pp. 1-6). IEEE. DOI: 10.1109/ICIP49359.2023.10222669