International Journal of Engineering Sciences & Emerging Technologies, Dec. 2024
ISSN: 22316604 Volume 11, Issue 4, pp: 907-911 ©IJESET

TRUTHGUARD: TECHNOLOGY FOR RECOGNIZING UNAUTHENTIC TRANSFORMED IMAGES AND VIDEOS

Himanshu Agarwal[1], Tushar Sachdeva[2], Yash Rudra[3],
Vaibhav Malhotra[4], Rishikesh Jha[5]

[1] Associate Professor CS&E Deptt MIT, Moradabad
himanshu.agg2000@gmail.com
[2] B. Tech Research Scholar CS&E Deptt MIT, Moradabad
sachdevatushar025@gmail.com
[3] B. Tech Research Scholar CS&E Deptt MIT, Moradabad
rudrayash007@gmail.com
[4] B. Tech Research Scholar CS&E Deptt MIT, Moradabad
vaibhavm2102@gmail.com
[5] B. Tech Research Scholar CS&E Deptt MIT, Moradabad
jharishikesh007@gmail.com

ABSTRACT

This project aims to tackle the challenges posed by deepfakes by proposing an innovative method for detecting them. It involves using the Xception model, convolutional neural network architecture known for its effectiveness in complex image classification tasks. The choice of Xception Net demonstrates our commitment to leveraging deep learning capabilities to distinguish between videos and those generated by AI. Additionally, we have given importance to the user interface of the project by utilizing the Python QT front end library creating a user-friendly platform. The detection process involves analyzing each video frame, in which Xception excels. By scrutinizing each frame, the model can accurately determine whether the video content is authentic or not. The integration of Python QT enhances the user experience making it easy and intuitive to use our detection application.

In a society where false information and online dangers can lead to real world impacts this project highlights the significance of progress in managing the hazards linked to the use of artificial media. To sum up combining Xception Net and Python QT presents a remedy.

KEYWORDS - *Deep Learning, XceptionNet, Python QT, CNN.*

1. INTRODUCTION

Xception Net, stands for Extreme Inception, represents a sophisticated convolutional neural network (CNN) architecture designed for image classification tasks. Introduced by François Chollet in 2017, Xception Net innovatively replaces the traditional inception modules with depthwise separable convolutions. This modification enhances the network's capacity to capture complex hierarchical features while significantly reducing computational costs. By decoupling the spatial and channel-wise dependencies, Xception Net achieves a balance between expressive power and computational efficiency, making it particularly well-suited for applications with constrained resources [5,6,7].

Python QT, on the other hand, is a powerful framework for developing cross-platform graphical user interfaces (GUIs) in Python. QT stands for "Quality and Technology," and Python QT provides a set of tools and libraries for creating visually appealing and responsive applications. Developed by The Qt Company, Python QT supports the creation of desktop applications with a native look and feel on various operating systems. Its modular architecture facilitates the development of versatile and customizable GUIs, allowing developers to design interfaces for applications ranging from simple tools to complex software solutions.

When integrated into a project, Python QT serves as the front-end library, enabling the creation of a user-friendly interface for the deepfake detection application. Its capabilities extend to designing interactive and visually pleasing graphical elements, ensuring an intuitive experience for users interacting with the deepfake detection system [8]. The choice of Python QT underscores the project's commitment to delivering not just a robust deep learning model with Xception Net but also an accessible and engaging user interface. This combination of a powerful neural network architecture and a versatile front-end library emphasizes the project's dedication to addressing the multifaceted challenges posed by deepfake technology.

2. LITERATURE REVIEW

2.1. Celeb-DF: An Extensive Dataset Designed to challenge Deepfake Forensics

From here we see the emergence of a major problem online – AI generated face swapping videos, also known as Deepfakes. These videos pose a significant threat to the trustworthiness of online information, creating convincing illusions by replacing faces of target individuals with those synthesized by deep neural network (DNN) models. To address the need for robust Deepfake detection methods, here we introduce a new large and challenging Deepfake video dataset called Celeb-DF, with 5,639 high quality Deepfakes of celebrities. This dataset is crucial for developing and evaluating detection algorithms as it addresses the limitations of existing datasets available on the internet that have low visual quality and don't represent the Deepfakes that are out there.

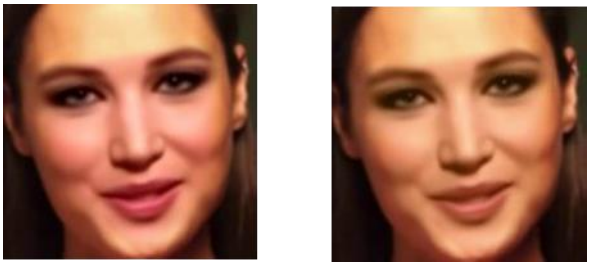

Figure 1: Comparison of Deepfake frames: left without synthesized face, right with color correction.

2.2. Deepfake Detection using Deep Neural Networks

From the above paragraph, we get the concept of deepfakes which are fake images or videos generated through algorithms, image processing and face swap. These computer-generated fakes often merge images to create convincing representations of events, comments, or activities that never took place. The key technology behind deepfakes is the Generative Adversarial Network (GAN) which uses landmark points on faces to map and manipulate facial features.

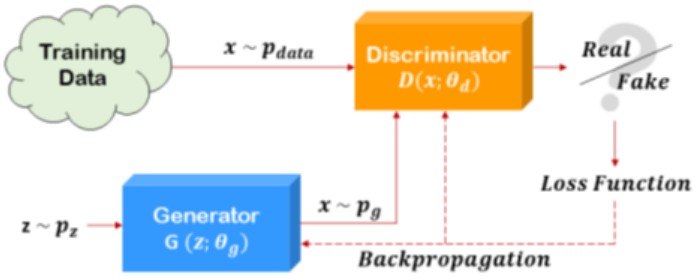

Figure 2: Simplified description of a GAN learning framework.

The goal of this project is to use neural networks to tell the difference between fake and real images, specifically deepfakes. To do this they use the publicly available Flickr Faces High-Quality (FFHQ) dataset. For feature extraction they use pre-trained Convolutional Neural Network (CNN) architectures like EfficientNetB4, InceptionV3 and InceptionResNetV2. Classification tasks are performed using

Long Short-Term Memory (LSTM). The evaluation of the models is done through a Classification Report, including metrics like accuracy and F-1 score.

2.3. Detecting Deepfake Images Using Deep Learning Techniques

From the above paragraph, we learn about the pervasive issue of deepfakes and their disruptive impact on society. Deepfakes involve the use of artificial intelligence (AI), machine learning (ML), and deep learning (DL) to replace a person's likeness in images or videos with that of another. While visual media manipulation is not new, the advent of deepfakes has introduced a more sophisticated and challenging form of fake media. The paragraph emphasizes the societal consequences of such manipulations.

To counter the threat of deepfakes, the study proposes a comprehensive approach for detection using DL methods. It acknowledges the black-box nature of DL systems, which makes them challenging to interpret and trust fully. The solution proposed involves Explainable Artificial Intelligence (XAI), specifically the Local Interpretable Model-Agnostic Explanations (LIME) algorithm, to enhance transparency by interpreting DL predictions.

2.4. Deepfake Detection by Analyzing Convolutional Traces

From the above paragraph, we learn about the significant and growing issue of Deepfake technology, which involves using deep learning algorithms to automatically generate or alter a person's face in images and videos. Deepfakes can produce highly convincing multimedia content that is challenging for the human eye to distinguish as real or fake. The term "Deepfake" encompasses all multimedia content synthetically altered or created by machine learning generative models.

The paragraph highlights various examples of Deepfake instances involving celebrities, such as inserting Nicolas Cage into movies he did not originally act in or creating a video where Jim Carrey plays a role originally portrayed by Jack Nicholson. More concerning examples include a video of ex-US President Barack Obama and another featuring Mark Zuckerberg making false statements. These instances underscore the potential threat that Deepfakes pose to the authenticity of news, politics, companies, and individual privacy.

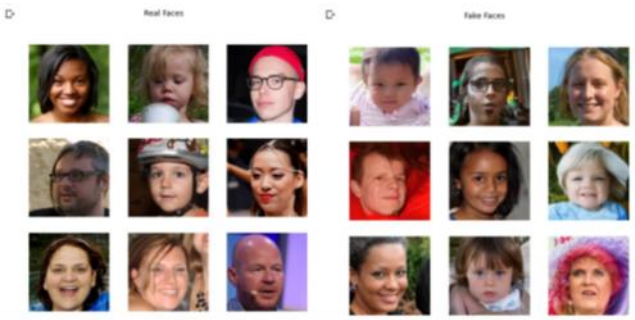

Figure 3: Data Exploration for real and fake Images.

3. PROBLEM STATEMENT

Due to this increasing prevalence of deepfakes, there remains a notable absence of effective and accessible tools for detecting these manipulated media. Current detection methods are often limited in scope, accuracy, and usability, leading to a growing concern about the unchecked spread of deepfakes across digital platforms [9,10].

Therefore, there is a need for the development of a robust and user-friendly deepfake detection system. It should leverage advanced machine learning techniques to

distinguish between authentic and manipulated content, while also providing transparency and interpretability in its decision-making process and this detection system should be easily accessible to

a wide range of users, including journalists, social media platforms, and the public. It should offer seamless integration with existing digital tools and platforms, enabling proactive detection and mitigation of deepfake threats in real-time [11,12,13].

In the upcoming sections, a comprehensive breakdown of each step will be provided for better understanding and clarity:

• Dataset Gathering

In this phase, we have gathered various datasets, both pre-existing and newly collected, to ensure a diverse and comprehensive dataset for accurate detection of various types of videos. To reduce training biases, our dataset consists of an equal distribution of real and fake videos, each accounting for 50% of the dataset [14].

• Preprocessing

During preprocessing, videos undergo refinement to eliminate extraneous elements and noise. Specifically, we focus on detecting and isolating the essential component of the video – the human face. By employing advanced facial detection techniques, we crop and extract the facial regions from the videos, ensuring that only pertinent data is utilized for subsequent analysis.

• Model Training

Our detection model is a Xception CNN model, we extract features at the frame level from the preprocessed videos. These extracted features serve as input to our model, enabling it to classify videos as either deepfake or pristine based on learned patterns and characteristics [15].

• Prediction

In this final module, new videos undergo preprocessing like the training data and are then fed into the trained model for prediction. The model evaluates the input video and provides a prediction regarding its authenticity – whether it is a real or fake video.

Model	Training_loss	Training_Acc
XceptionNet	0.0371	0.9836

4. RESULTS

The output of the model is going to be whether the video is deepfake or a real video along with the confidence of the model.

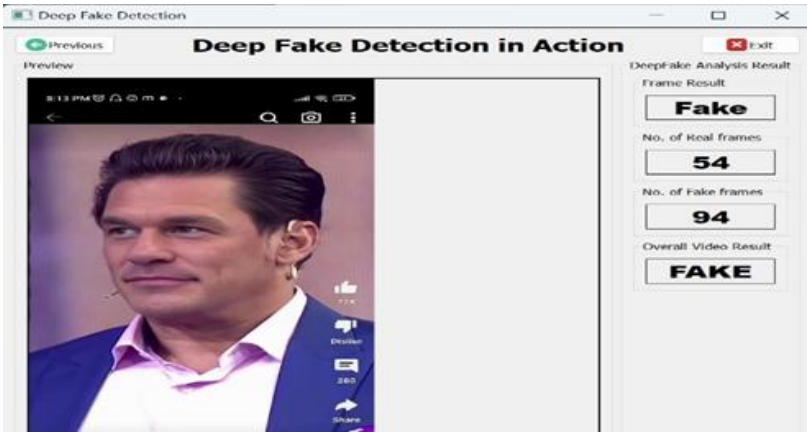

Figure 4: Output Screen.

5. FUTURE SCOPE

The future scope of a deep fake detection project using XceptionNet involves enhancing its accuracy, efficiency, and applicability. This could include refining the model to detect more sophisticated deep fakes, developing real-time detection capabilities, integrating it into social media platforms for automated detection, and adapting it to emerging deep fake techniques. Additionally, there's potential for collaboration with cybersecurity firms, law enforcement agencies, and tech companies to combat the misuse of deep fake technology.

6. CONCLUSION

In summary our continuous effort to improve our detection project, especially by using XceptionNet represents a significant step forward, in enhancing the accuracy and effectiveness of identifying sophisticated manipulated content. By refining our model, implementing state of the art techniques and collaborating with stakeholders we are making progress in strengthening our defense against the misuse of deep fake technology. By combining machine learning algorithms and advanced computer vision methods we aim to enhance the capabilities of our detection system to adapt to evolving techniques. This dedication highlights the importance of research and development in cybersecurity as we strive to combat the spread of content.

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] 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.

[7] David G¨uera and Edward J Delp. Deepfake video detection using recurrent neural networks. In AVSS, 2018.

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In CVPR, 2016.

[9] An Overview of ResNet and its Variants: https://towardsdatascience.com/an-overview-of-resnet-and-its-variants-5281e2f56035

[10] Long Short-Term Memory: From Zero to Hero with Pytorch Available at: https://blog.floydhub.com/long-short-term-memory-from-zero-to-hero-with-pytorch/

[11] Rana, M.S., Nobi, M.N., Murali, B., Sung, A.H. (2022). Deepfake detection: A systematic literature review. IEEE Available at: https://ieeexplore.ieee.org/document/9721302

[12] Mirsky, Y., Lee, W. (2021). The creation and detection of deepfakes. ACM Computing Surveys, 54(1): 7. https://doi.org/10.1145/3425780

[13] Kaggle Deepfake Detection Challenge. Available at: https://www.kaggle.com/c/deepfake-detection-challenge/data

[14] Face Forensics GitHub Repository. Available at: https://github.com/ondyari/FaceForensics

[15] Y. Qian et al. “Recurrent color constancy.” Proceedings of the IEEE International Conference on Computer Vision, pages 5459–5467, Oct. 2017. Venice, Italy.