International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Advancements in Video Deepfake Detection: Integration of ResNet50, EfficientNetB7, and Efficient NetAutoAtt B4 Models

Shwetambari Borade[1], Nilakshi Jain[2], Bhavesh Patel[3], Vineet Kumar[4], Mustansir Godhrawala[5], Shubham Kolaskar[6], Yash Nagare[7], Pratham Shah[8], Jayan Shah[9]

Submitted: 29/01/2024 **Revised:** 08/03/2024 **Accepted:** 15/03/2024

Abstract*:* The study aims to foster responsible advances in facial manipulation techniques by developing reliable detection methods for deepfakes. The proposed model, based on the average of three separate frameworks, was trained using various datasets and all false positives and negatives. It outperforms other deepfake detection techniques in detecting both dynamic and static deepfakes. The model's architecture is robust, requires minimal computing power, and is adaptable to unique features of deepfake detection, making it suitable for real-world deployment. The model also aids in further research to optimize ensemble models and evaluate advanced training methods like knowledge distillation. Its efficacy, efficiency, and scalability provide a feasible approach in the fight against deepfakes and serve as a foundation for future research and development in deepfake detection technologies. The novelty lies in the model's potential for real-life implementation and its proven effectiveness in mitigating challenges related to deepfake video detection.

Keywords: *Convolutional neural networks, Deepfakes, Deep Learning, Efficient Net, GAN*

1. Introduction

In the rapidly evolving digital world, the advent of deepfake technology has brought forth both awe-inspiring possibilities and alarming challenges. Deepfakes, sophisticated artificial intelligence-based techniques used to manipulate or fabricate visual and audio content with a high potential for deception, have raised significant concerns about the integrity and authenticity of digital media [1]. This underscores the urgent need for robust and reliable detection methods to counteract these manipulations.

This research paper introduces a novel approach to deepfake detection, utilizing the combined strengths of three distinct, state-of-the-art models: ResNet50, EfficientNetB7, and EfficientNetAutoAttB4. Each of these models has been chosen for their unique capabilities and contributions to the overall detection system. ResNet50, renowned for its deep residual networks, excels at learning from large-scale datasets. Its ability to effectively handle complex patterns makes it a powerful tool for identifying subtle manipulations in images, a crucial aspect in deepfake detection. On the other hand, EfficientNetB7 brings to the table a new paradigm in scaling up neural networks. By balancing the network depth, width, and resolution, it achieves superior performance, making it an invaluable asset in our detection toolkit. Lastly, EfficientNetAutoAttB4, with its automated search for the best neural architecture, ensures optimal performance by dynamically adapting to the task at hand. This adaptability is key in the ever-evolving landscape of deepfakes, where new techniques and variations continually emerge.

By integrating these models into a cohesive system, this research aims to develop a deepfake detection method that is not only highly accurate but also efficient and adaptable to various forms of deepfakes. The system's design also takes into consideration real-world deployment, ensuring that it can be effectively implemented even in devices with limited computing power.

The subsequent sections of this paper will delve deeper into the specifics of each model, the training process, and the performance results. This comprehensive overview aims to shed light on this cutting-edge approach to deepfake detection, providing valuable insights and paving the way for future advancements in this critical field. The novelty of this research lies not just in the integration of these three models, but also in its practical application. The model's potential for real-life implementation and its demonstrated effectiveness in mitigating the challenges related to deepfake video detection underscore the promising role of this technology in the ongoing fight against deepfakes. This research, therefore, serves as a significant contribution to the body of knowledge in deepfake detection and a beacon guiding future research in this area.

2. Literature review

This research [2] proposes a hybrid strategy combining ResNet-50 (Residual Network-50) and Long Short-Term Memory (LSTM) algorithms for detecting anomaly video activity. The technique achieved 96.48% accuracy using the UCSD Ped 1 dataset.

This study [3] systematically studies model scaling and identifies that carefully balancing network depth, width, and resolution can lead to better performance.

This research [4] proposes a novel framework to detect fake videos through the utilization of transfer learning in autoencoders and a hybrid model of convolutional neural networks (CNN) and Recurrent neural networks (RNN).

This research [5] introduces the DFDC dataset, the largest currently and publicly available face swap video dataset, with over 100,000 total clips sourced from 3,426 paid actors.

This work [6] presents Celeb-DF, a new large-scale challenging

DeepFake video dataset, which contains 5,639 high-quality DeepFake videos of celebrities generated using an improved synthesis process.

This research [7] examines the realism of state-of-the-art image manipulations, and how difficult it is to detect them, either automatically or by humans.

This study [8] introduces EfficientNetV2, a new family of convolutional networks that have faster training speed and better parameter efficiency than previous models.

This research [9] proposes a novel strategy to remove GAN “fingerprints” from synthetic fake images based on autoencoders in order to spoof facial manipulation detection systems while keeping the visual quality of the resulting images.

[10] proposes a framework called digital forensics capability analyzer to help organizations assess their readiness to handle digital forensics.

[11] tackles the problem of face manipulation detection in video sequences targeting modern facial manipulation techniques. It studies the ensembling of different trained Convolutional Neural Network (CNN) models.

3. Methodology

3.1. Resnet50

The proposed architecture is based on ResNet50, which is a 50-layer deep neural network recognized for its ability to identify facial expressions with high accuracy. It uses 32x4 dimensions and is designed to recognize various biases and characteristics, particularly in the context of deepfake detection. The model takes video frames as input and undergoes preprocessing steps such as normalization and data augmentation to enhance its deepfake detection capabilities. Using its 50 layers, including convolutional, pooling, and fully connected layers, the model extracts hierarchical features from the preprocessed frames, utilizing residual blocks to learn and extract complex visual patterns and features [12]**.**

During training, the model is trained on a large dataset containing labeled video frames of both real and deepfake content, enabling it to distinguish between the two. Inference involves calculating the probabilities of each frame being part of a deepfake video, then using post-processing techniques such as temporal analysis and consensus decision-making to determine the overall likelihood of the video being a deepfake. The video frames are resized to 256×256 pixels, and textural features are extended to optical flow fields for analysis.

To ensure the necessary input dimensions for feature extraction, preprocessing involves removing each frame from the source video, resizing them, and utilizing a rescaling technique. The ResNet-50 model is utilized for feature extraction, and a 1538 vector of features is used as input for anomaly detection classification.

Figure 1 shows the performance metrics for the proposed model of ResNet50.The system combines the use of ResNet-50 and LSTM techniques for video abnormality detection. The LSTM helps the network learn difficult features, and the input & forget gate network within the LSTM regulates the internal memory cells, addressing the problem of disappearing or exploding gradients in RNN. The system uses various performance metrics such as F1-score, AUC, Accuracy, Precision, Specificity, and Recall to evaluate its efficacy in detecting anomalies and reducing false alarms.

The ResNet50 model undergoes training using a large dataset of labeled images or video frames, encompassing both real and deepfake content. The training process involves presenting the training data to the network in batches, generating predictions, computing discrepancies between predicted outputs and actual labels using a loss function, and employing backpropagation to update the network’s weights in a direction that minimizes the loss. [13]. This process is repeated over multiple epochs to enhance the model’s capability to differentiate between real and deepfake content.

The unique architecture of ResNet50 enables the model to effectively capture and process intricate features during training, contributing to its ability to make precise predictions during inference.

Fig. 4. Real image from DFDC Dataset

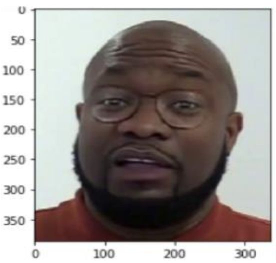


Fig. 5. Fake image from DFDC Dataset

The model was tested and trained using the Celeb-DF and DFDC datasets, resulting in an efficient model training process with a high accuracy percentage of 97%. The Celeb-DF dataset consists of 795 synthesized videos, 408 real videos that were used to create deepfakes, and 250 real videos from YouTube, and it played a crucial role in training the model to yield optimal results. The usage of these realistic datasets not only facilitated the training process but also enhanced training accuracy and speed, leading to a more efficient model training process with fewer epochs required. As we can see in figure 2 and figure 3 the deepfakes of CelebDf dataset are phenomenal and not easy to identify. This is used to train our model to differentiate between the deepfakes and real images. Similarly figures 4 and 5 show the images classified from the DFDC dataset as deepfake image and real image.

3.2. EfficientNetAutoAttB4

The proposed strategy for enhancing the accuracy of predictions involves ensembling several CNN-based classifiers to gather distinct kinds of high-level semantic data that improve each other and enhance the ensemble as a whole. To achieve this, two improvements are suggested for an EfficientNet architecture. The first enhancement involves the incorporation of an attention mechanism for identifying the most informative portion of the video for classification, while the second improvement explores the use of Siamese training methodologies to extract additional data. The attention mechanism is implemented on the EfficientNetB4 model, utilizing the ImageNet dataset. By focusing on facial data instead of using the entire frame as input, the model generates a feature vector of 1792 elements, enhancing the accuracy of classification. The attention mechanism selects feature

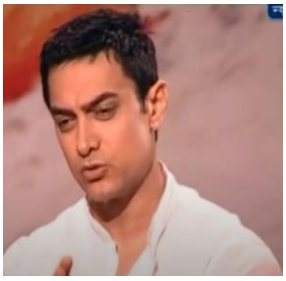


Fig. 2. Real Image classified by ResNet50 model

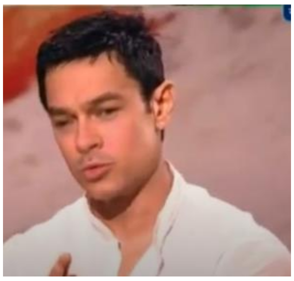


Fig. 3. Fake Image classified by ResNet50 model

maps from the EfficientNetB4 up to a specific layer and processes them using a single convolutional layer with a kernel size of 1, followed by a Sigmoid activation function to obtain a single attention map. This attention map is then multiplied with each ofthe feature maps, enabling the network to focus on the most relevant parts of the input. This mechanism not only improves the network's focus on the most relevant portions of the feature maps but also provides deeper insights into the network's informative assumptions about the input. [14] The deepfake detection model, based on the EfficientNetAutoAttB4 architecture, underwent comprehensive evaluation and achieved commendable results across various metrics.

Confusion Matrix:

- True Positives (Correctly classified real videos): 1798
- True Negatives (Correctly classified fake videos): 1837
- False Positives (Wrongly classified real videos as fake): 455

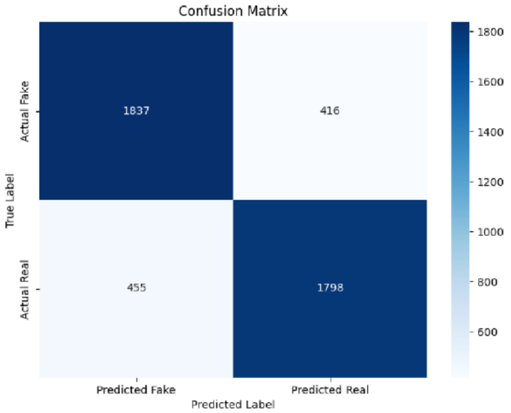


Fig. 6. Confusion matrix for EfficientNetAutoAttB4

- False Negatives (Wrongly classified fake videos as real): 416

The confusion matrix provides a detailed breakdown of the model's performance, showcasing a high number of correct classifications for both real and fake videos. The limited occurrences of false positives and false negatives further underline the model's reliability.

The confusion matrix is depicted in figure 6 above.

In parallel to the traditional training strategy, the Siamese training leverages the network's ability to generalize to create a feature descriptor emphasizing the similarity between samples from the same class. The goal is to learn a representation in the network's encoding space that effectively separates samples into real and fake classes. Both end-to-end and Siamese training methods are employed on different datasets using specific data split strategies for each dataset. Data augmentation techniques, including downscaling, horizontal flipping, random brightness contrast, hue saturation, noise addition, and JPEG compression, are used on the input faces during training and validation to increase robustness. [15] The training process is designed to prevent overfitting and maximize efficiency. The models are trained using the Adam optimizer with specific hyperparameters, and data processing is reduced by focusing on the region where the subject's face is located. The BlazeFace extractor is utilized to extract faces from each frame, resulting in the input for the networks being a squared color image of size 224 × 224 pixels.The feature extractor is trained using the same number of iterations, validation routine, and learning rate scheduling as the end-to-end training, with the difference lying in the loss function used and the composition of the batch. For the feature extractor, the batch is composed of 12 triplets of samples selected across all videos of the set considered. The training duration is determined by a maximum number of iterations or until the validation loss plateaus. Validation is performed at regular intervals, and if the validation loss doesn't decrease, the initial learning rate is reduced. The entire methodology is illustrated in a flowchart to visually represent the process. In summary, the proposed methodology focuses on enhancing the EfficientNet architecture with attention mechanisms and Siamese training methodologies, employing specific training and validation strategies, data augmentation techniques, and efficient data processing to improve accuracy and efficiency in deepfake detection.

3.3. EfficientNetB7

The proposed NS-Efficient B7 model presents a sophisticated method for deepfake video verification, combining the capabilities of EfficientNet B7 and multistage facial analysis to meticulously examine every video frame for face characteristics and temporal anomalies. In stage 1, the preprocessing and face detection involve accepting raw video frames, optional resampling for efficient processing, and utilizing powerful algorithms such as Multi-task Cascaded Convolutional Networks (MTCNNs) to accurately recognize faces and build bounding boxes around them. During stage 2, facial feature extraction and tamper detection occur through landmark generation and tamper detection modules. The EfficientNet B7 network systematically processes the cropped images, comparing local image patches within the crop to generate multiple Structural Similarity Index Measure (SSIM) masks. Attention mechanisms within the EfficientNet B7 network are employed to emphasize the features of the face most amenable to manipulation, while temporal analysis, utilizing recurrent neural networks (RNNs) or other temporal modeling techniques, is used to capture temporal anomalies across frames. In stage 3, fusion and aggregation gather features from SSIM masks, EfficientNet B7 features, and temporal analysis results, which are then utilized by an advanced binary classification module to determine the probability of each frame being a deepfake. This stage also involves generating confidence scores for each frame's deepfake

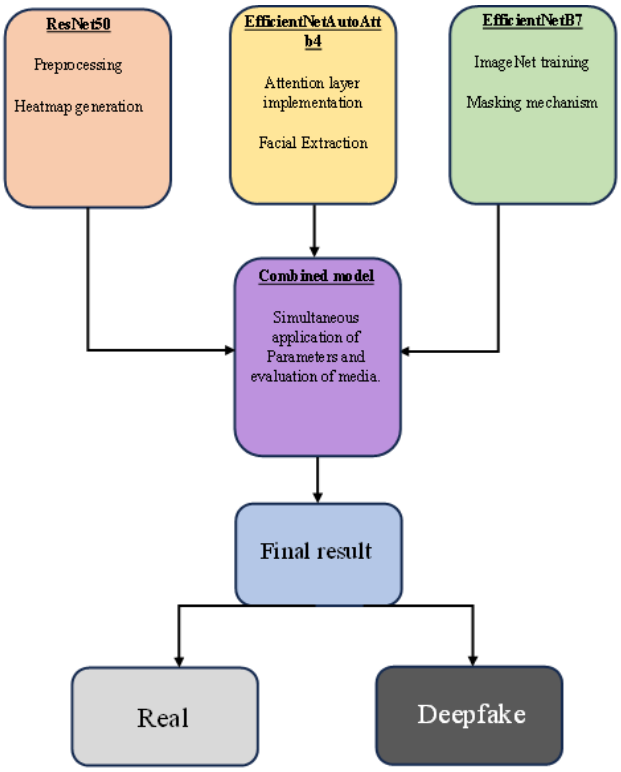


Fig. 9. Concurrent working of the 3 proposed models

prediction. Additional considerations include training with noisy students, which leverages both labelled and unlabelled video data, and ensemble learning to further improve the precision and dependability of the system, making it stronger against different deepfake methods.

The attention mechanism, as seen in figure 8, explains the neural network's ability to monitor the eye movements of the chopped faces of the individuals of interest while searching for anomalies in that particular region. Such a procedure adds an additional layer of detection which improves the model's overall reliability and accuracy. NS-Efficient B7 combines EfficientNet B7's feature extraction with targeted facial analysis and temporal modeling, showcasing great promise for safeguarding against digital manipulation. The model applies compound scaling, reversed residual blocks, and squeeze-and-excitation (SE) blocks to enhance its feature extraction capabilities, and operates alongside a multi-stage facial analysis pipeline. The noisy student training technique further improves the model's identifying abilities by utilizing both labelled and unlabelled data [16].

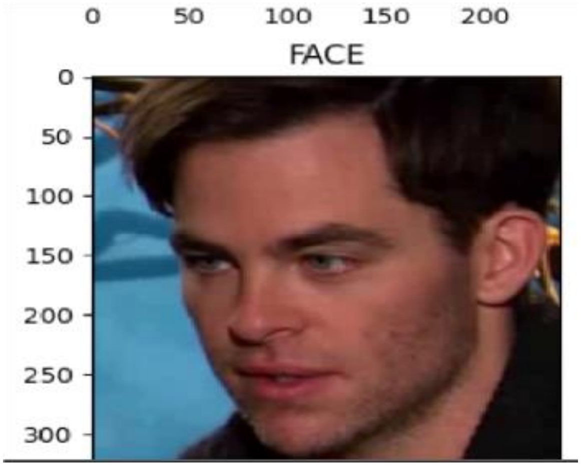


Fig. 7. Real Face from DFDC dataset

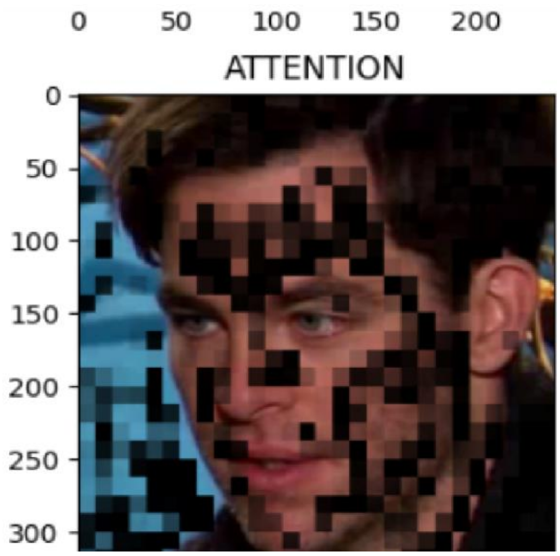


Fig. 8. Attention Layer on the face for deepfake detection by EfficientnetB7 model

The three models individually analyze the uploaded video and extract the frames, create a rgb heatmap around the suspicious areas and overlay another attention layer on the video. Initially Three different scores are assigned to the media, then a threshold value is considered and then the media is evaluated as a deepfake or a real video. The process is showed in the above figure 9.

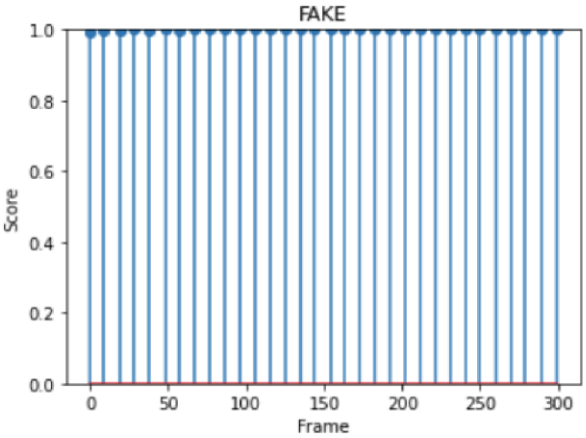


Fig. 14. Graph of the result for a fake Image by the model EfficientNetAutoAttB4

4. Results

4.1. ResNet50

The ResNet50 model uses rgb heatmap to distinguish between the real and fake images and classifies the areas of deepfake images. The utilization of heatmaps in identifying prominent areas on the face shown in media to determine whether it is a deepfake makes it more accessible for non-technical individuals to identify the manipulated details. This method, adopted from a referenced
pa hen
pr n as
th ore
in can
vi gher
in gest
m ent-
w ient
of ese
gı ow
th can
be and

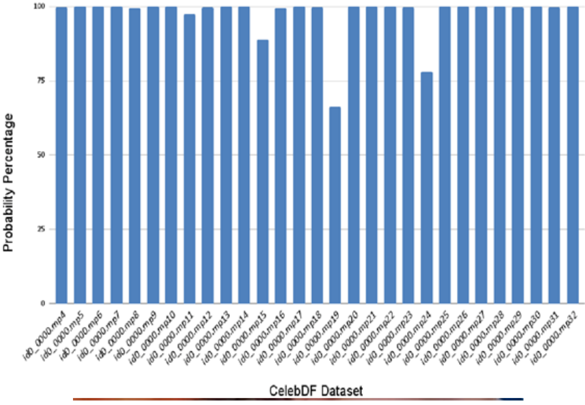


Fig. 12. Prediction results from ResNet50 model

Fig. 10. Original photo from CelebDF dataset

trustworthiness, particularly when training more complex models like Convolutional Neural Networks (CNNs). Heatmaps enable the detection of regions where the model indicates an uneven surface of the face and other characteristics by which it classifies the content as a deepfake.

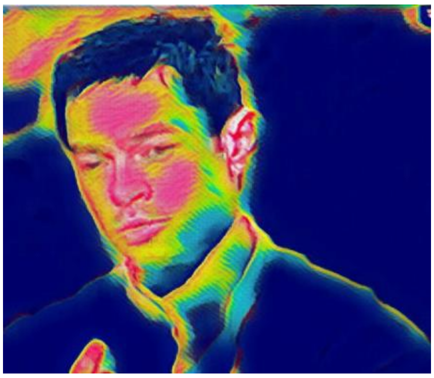


Fig. 11. Heatmap identification on image from CelebDF dataset

In figure 12 the X Axis contains the content of the Celebdf Dataset While the Y Axis shows Probability of The Video To Be a Deepfake. If the given probability is near to 1 which is near to 100% then the video is said to be deepfake.

4.2. EfficientNetAutoAttB4

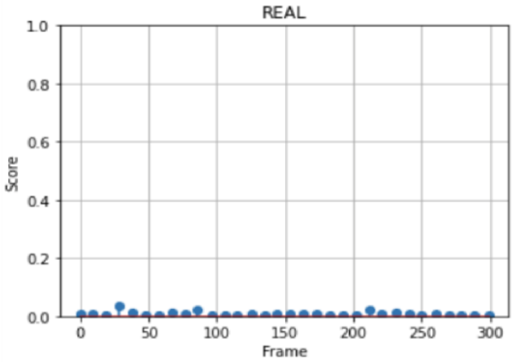


Fig. 13. Graph of the result for a Real Image by the model EfficientNetAutoAttB4

Figure 13 is the graph of the real video and 14 is of the fake video respectively. Real video means lines will be closer to the x axis. Every line is depicting a different feature of the video frame. And if the line is closer to 1 meaning the video is fake.

As shown in figure 15, if the particular video has a score near to ‘0’ the video will be real. Score near to ‘1’ will show that the video is fake. The score is determined by taking into account the combination of EfficientNet-B4 architecture and the attention layer.

4.3. Efficientnet B7

The structural similarity index measure (SSIM) is a method utilized to assess the perceived quality of digital images and videos, as well as to determine the resemblance between two images. Unlike other metrics such as mean squared error (MSE) or peak signal-to-noise ratio (PSNR) that estimate absolute errors, SSIM focuses on evaluating image degradation by considering perceived changes in structural information and integrating key perceptual phenomena such as brightness and contrast masking elements. This approach prioritizes the importance of an unaltered, distortion-free image as the baseline for quality evaluation. Structural information in images, which signifies the interrelatedness of pixels, especially those in close proximity, plays a critical role in conveying vital details about the visual scene's arrangement.

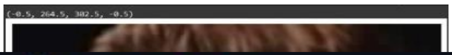
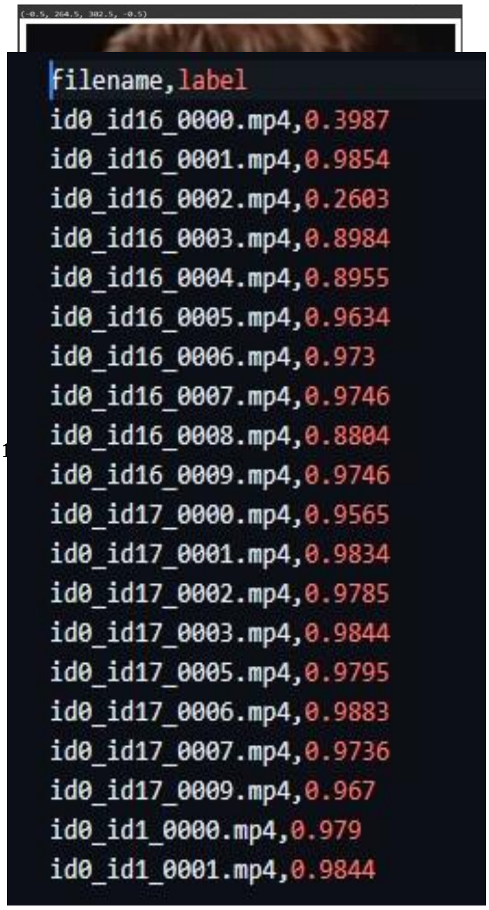


Fig. 1 model



```
filename,label
id0_id16_0000.mp4,0.3987
id0_id16_0001.mp4,0.9854
id0_id16_0002.mp4,0.2603
id0_id16_0003.mp4,0.8984
id0_id16_0004.mp4,0.8955
id0_id16_0005.mp4,0.9634
id0_id16_0006.mp4,0.973
id0_id16_0007.mp4,0.9746
id0_id16_0008.mp4,0.8804
id0_id16_0009.mp4,0.9746
id0_id17_0000.mp4,0.9565
id0_id17_0001.mp4,0.9834
id0_id17_0002.mp4,0.9785
id0_id17_0003.mp4,0.9844
id0_id17_0005.mp4,0.9795
id0_id17_0006.mp4,0.9883
id0_id17_0007.mp4,0.9736
id0_id17_0009.mp4,0.967
id0_id1_0000.mp4,0.979
id0_id1_0001.mp4,0.9844
```

Fig. 17. Prediction Output for CelebDF dataset in ResNet50 model

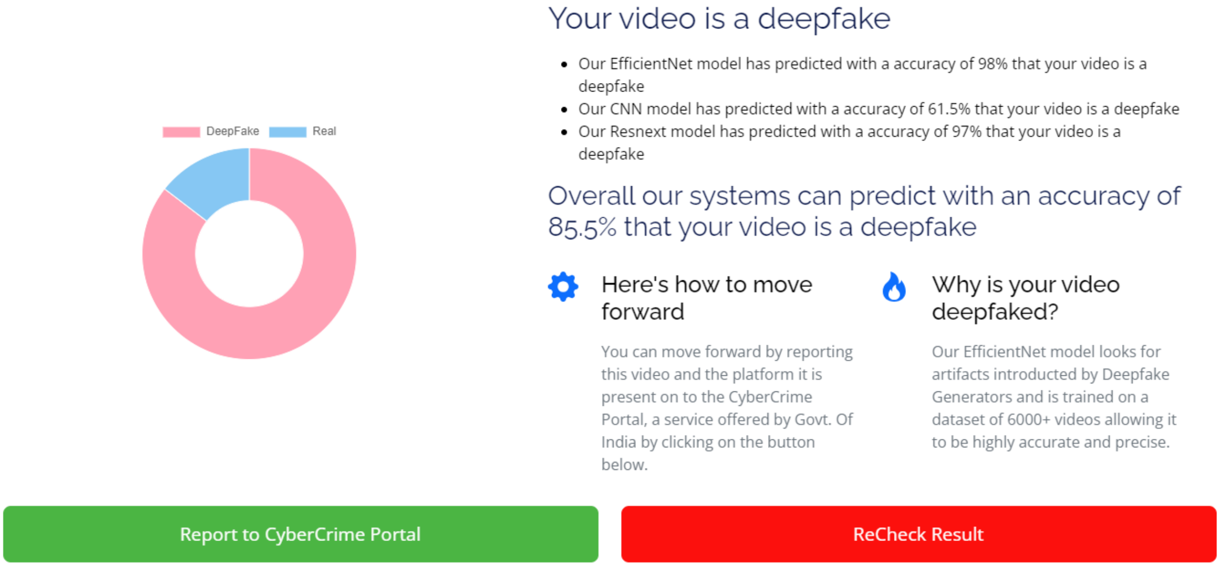
Moreover, SSIM accounts for contrast masking, where distortions become less noticeable in areas with high activity or texture, and luminance masking, where distortions are less visible in bright components of the image.

Figure 17 illustrates how we provided the models an array of test videos to process in the last stage of the detection phase with one another. The final results were saved into a Comma Separated Value (CSV) file, which appears below. Whenever the label's trustworthiness for the test's content reaches 1, it indicates a deep fake. It's a real video if the score is closer to 0.

The three proposed models work together for video deepfake detection and are currently deployed on Trusttrace.net.in Here the results are evaluated and presented as shown in the below figure 18. The above image contains the result for the uploaded video media. Figure 3 is a still from the video uploaded on the website for eligibility evaluation. As mentioned earlier the video is a deepfake and the results from the above image also conclude the same. The deepfake video models of EfficientNetB7 and EfficientNetAutoAttB4 have contributed in identification of improper facial topology and the resnet50 model accurately detects the facial movements of the video. Thus the 3 models effectively evaluate the media to be considered as a deepfake video.

5. Conclusion

In conclusion, the deployment model for the detection of deepfake videos is portrayed as a significant progression in the realms of computer vision and deep learning. The model, with its robust architecture comprising 50 layers and residual blocks, has showcased remarkable abilities in identifying intricate visual patterns and features that differentiate real content from deepfake content within videos. By capitalizing on a diverse and well-annotated dataset, the model can be effectively trained to accurately pinpoint and flag deepfake content, thereby contributing to the ongoing efforts to curb the proliferation of misleading and potentially harmful manipulated videos.
The use of deep learning frameworks such as TensorFlow, Keras, or PyT has provided a sturdy and flexible environment for the implementation of the model, facilitating efficient training, validation, and deployment processes. As deepfake techniques continue to evolve, the model serves as a foundation for enhancing deepfake detection capabilities, including the incorporation of ongoing advancements in data preprocessing, model optimization, and continuous learning from new and emerging deepfake content. Moreover, the application of the model in a real-world context has the potential to bolster the trust and reliability of video content across various domains, including media, journalism, and law enforcement. Furthermore, it's crucial to recognize the ethical considerations associated with the use of deepfake detection technologies, ensuring responsible deployment and safeguarding individual privacy and rights. Sustained research and collaboration in this area are vital to further refine and advance the capabilities of deepfake detection models, ultimately contributing to a more trustworthy and secure digital landscape.



Your video is a deepfake

- Our EfficientNet model has predicted with a accuracy of 98% that your video is a deepfake
- Our CNN model has predicted with a accuracy of 61.5% that your video is a deepfake
- Our Resnext model has predicted with a accuracy of 97% that your video is a deepfake

Overall our systems can predict with an accuracy of 85.5% that your video is a deepfake

Here's how to move forward

You can move forward by reporting this video and the platform it is present on to the CyberCrime Portal, a service offered by Govt. Of India by clicking on the button below.

Why is your video deepfaked?

Our EfficientNet model looks for artifacts introducted by Deepfake Generators and is trained on a dataset of 6000+ videos allowing it to be highly accurate and precise.

Figure 18. Results generated model on Trusttrace platform

Acknowledgements

We would like to express our sincere gratitude to the CyberPeace Foundation for their invaluable assistance and support in our research project on deepfake detection. Their generous allowance facilitated the execution of this project, enabling us to delve into the intricate complexities of combating digital misinformation. The CyberPeace Foundation's commitment to promoting cybersecurity and fostering awareness of emerging threats has been instrumental in shaping our research endeavors. We are truly thankful for their partnership and dedication to advancing the cause of CyberPeace.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] N. Jain *et al.*, "Deepfake Technology and Image Forensics: Advancements, Challenges, and Ethical Implications in Synthetic Media Detection," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 16s, pp. 49–58, Feb. 2024, [Online]. Available: https://www.ijisae.org/index.php/IJISAE/article/view/4782

[2] S. R. Krishnan and P. Amudha, "International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING Hybrid ResNet-50 and LSTM Approach for Effective Video Anomaly Detection in Intelligent Surveillance Systems." [Online]. Available: www.ijisae.org

[3] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," May 2019, doi: https://doi.org/10.48550/arXiv.1905.11946.

[4] A. Seth and A. K. Gogineni, “Detection of Deep-fakes in Videos using CNN and Transformers”, doi: 10.13140/RG.2.2.23238.60480.

[5] B. Dolhansky *et al.*, “The DeepFake Detection Challenge (DFDC) Dataset,” Jun. 2020, doi: https://doi.org/10.48550/arXiv.2006.07397.

[6] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, “Celeb-DF: A Large-scale Challenging Dataset for DeepFake Forensics,” Sep. 2019, doi: https://doi.org/10.48550/arXiv.1909.12962.

[7] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Niessner, “FaceForensics++: Learning to detect manipulated facial images,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2019. doi: 10.1109/ICCV.2019.00009.

[8] M. Tan and Q. V. Le, “EfficientNetV2: Smaller Models and Faster Training,” Apr. 2021, [Online]. Available: http://arxiv.org/abs/2104.00298

[9] J. C. Neves, R. Tolosana, R. Vera-Rodriguez, V. Lopes, H. Proença, and J. Fierrez, “GANprintR: Improved Fakes and Evaluation of the State of the Art in Face Manipulation Detection,” *IEEE Journal on Selected Topics in Signal Processing*, vol. 14, no. 5, 2020, doi: 10.1109/JSTSP.2020.3007250.

[10] P. Reeva, D. Siddhesh, G. Preet, S. Pratik, and N. Jain, “Digital Forensics Capability Analyzer: A tool to check forensic capability,” in *2019 International Conference on Nascent Technologies in Engineering, ICNTE 2019 - Proceedings*, 2019. doi: 10.1109/ICNTE44896.2019.8945960.

[11] N. Bonettini, L. Bondi, E. D. Cannas, P. Bestagini, S. Mandelli, and S. Tubaro, “Video face manipulation detection through ensemble of CNNs,” in *Proceedings - International Conference on Pattern Recognition*, Institute of Electrical and Electronics Engineers Inc., 2020, pp. 5012–5019. doi: 10.1109/ICPR48806.2021.9412711.

[12] A. Berroukham, K. Housni, M. Lahraichi, and I. Boulfrifi, “Deep learning-based methods for anomaly detection in video surveillance: a review,” *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 1, 2023, doi: 10.11591/eei.v12i1.3944.

[13] H. T. Duong, V. T. Le, and V. T. Hoang, “Deep Learning-Based Anomaly Detection in Video Surveillance: A Survey,” *Sensors*, vol. 23, no. 11. 2023. doi: 10.3390/s23115024.

[14] S. Atas, I. Ilhan, and M. Karakse, “An Efficient Deepfake Video Detection Approach with Combination of EfficientNet and Xception Models Using Deep Learning,” in *2022 26th International Conference on Information Technology, IT 2022*, 2022. doi: 10.1109/IT54280.2022.9743542.

[15] S. Ganguly, A. Ganguly, S. Mohiuddin, S. Malakar, and R. Sarkar, “ViXNet: Vision Transformer with Xception Network for deepfakes based video and image forgery detection,” *Expert Syst Appl*, vol. 210, Dec. 2022, doi: 10.1016/j.eswa.2022.118423.

[16] L. Bondi, E. Daniele Cannas, P. Bestagini, and S. Tubaro, “Training Strategies and Data Augmentations in CNN-based DeepFake Video Detection,” in *2020 IEEE International Workshop on Information Forensics and Security, WIFS 2020*, Institute of Electrical and Electronics Engineers Inc., Dec. 2020. doi: 10.1109/WIFS49906.2020.9360901.