Deep-fake Detection using rPPG signals

B. Tech. Project Mid Sem Report

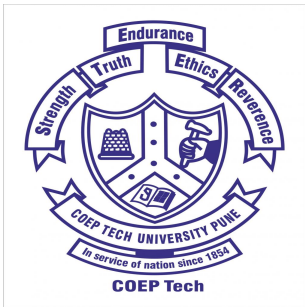
Submitted by

Sahil Deshpande	**112103119**
Rewa Saykar	**112103126**
Sidhesh Lawangare	**112103138**

Under the guidance of

Prof. P. R. Deshmukh

COEP Technological University, Pune



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

COEP TECHNOLOGICAL UNIVERSITY, PUNE-5

March 2024

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING,

COEP TECHNOLOGICAL UNIVERSITY, PUNE-5

CERTIFICATE

Certified that this project titled, "Deep-fake detection using rPPG signals" has been successfully completed by

Sahil Deshpande	**112103119**
Rewa Saykar	**112103126**
Sidhesh Lawangare	**112103138**

and is approved for the partial fulfillment of the requirements for the degree of "B.Tech. Computer Engineering".

SIGNATURE
Prof. Dr. P. R. Deshmukh
Project Guide
Department of CSE
COEP Tech Pune,
Shivajinagar, Pune - 5.

SIGNATURE
Dr. P. K. Deshmukh
Head
Department of CSE
COEP Tech Pune,
Shivajinagar, Pune - 5.

Abstract

Deepfake videos pose significant risks to digital security and misinformation by generating highly realistic synthetic content. Although deep fakes have found applications in entertainment and visual effects, they also pose serious risks, including misinformation, identity theft, and political manipulation. High-profile cases, such as the fake video of Sachin Tendulkar endorsing a political party and deepfake speeches attributed to former U.S. presidents, have demonstrated the growing challenge of distinguishing real content from synthetic fabrications.Traditional detection methods often struggle with advanced deep-fake techniques that convincingly mimic facial expressions and movements.r, we propose a novel deep-fake detection approach utilizing remote photoplethysmography (rPPG) signals to analyze subtle physiological cues such as heart rate variability, which are often missing or inconsistent in deep-fakes. Using a data set of 6,000 real and deep-fake videos, we extract rPPG features and train a deep learning model to distinguish between authentic and manipulated videos. Our approach demonstrates the effectiveness of physiological signal analysis in deep-fake detection.

Contents

Chapter 1

Introduction

Deepfake technology, powered by generative adversarial networks (GANs) and other AI models, has made it increasingly difficult to differentiate real videos from synthetic ones. Although traditional deep-fakefake detection methods focus on inconsistencies in facial expressions, lighting, and artifacts, these approaches can struggle with highly refined deep-fakes.

One promising avenue for detection is remote photoplethysmography (rPPG), a noncontact method that extracts heart rate signals from subtle skin color variations caused by blood flow. Since deep-fake videos lack natural physiological signals or exhibit unrealistic pulse patterns, rPPG-based detection can serve as a robust biometric defense.

In this study, we propose a deepfake detection model using rPPG signals, combined with deep learning techniques, to improve accuracy. We evaluate our method on a dataset of 6,000 videos and compare it with existing approaches. The results highlight the potential of physiological-based deepfake detection for real-world applications in media forensics and security.

Chapter 2

Literature Review

2.1 Deepfake Creation Techniques

Deepfake videos are primarily generated using Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs). These methods synthesize realistic human faces and expressions by training on extensive datasets. Common face manipulation techniques include FaceSwap, Face2Face, DeepFakes, NeuralTextures, and FaceShifter, each employing distinct methodologies to replace or modify facial features.

2.1.1 FaceSwap and Face2Face

Enable real-time facial reenactment, making them popular in entertainment and social media.

2.1.2 NeuralTextures

utilizes learned textures to enhance facial expressions, making detection more difficult.

2.1.3 DeepFakes

rely on autoencoders to swap faces in a convincing manner, often seen in viral videos and manipulated political content.

2.1.4 FaceShifter

Itis an advanced deepfake generation model that improves identity preservation and facial blending by using Adaptive Feature Fusion (AFF) to produce more seamless face-swapped videos, making detection even more challenging.

2.2 Deepfake Detection Approaches

Traditional deepfake detection techniques have relied on pixel-level inconsistencies, unnatural eye blinking patterns, and frame-by-frame inconsistencies. Machine learning-based classifiers, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have been employed to differentiate real and manipulated content. Recent advancements incorporate transformers and attention-based architectures, allowing models to better analyze temporal coherence in deepfake videos. These techniques improve upon basic CNN models by assigning weights to key regions of interest, highlighting areas where deepfake artifacts are likely to occur.

2.3 Biological Signal-Based Detection

Emerging research suggests that physiological signals, such as heartbeat rhythms and facial micro-expressions, provide promising avenues for deepfake detection.

2.3.1 Remote Photoplethysmography

Remote Photoplethysmography (rPPG) has been widely used in biomedical applications for heart rate estimation and is now being explored for deepfake detection.

2.3.2 DeepRhythm

The DeepRhythm approach specifically introduces attention-guided extraction of rPPG maps, making it possible to detect irregular or missing heartbeat patterns in deepfake videos.

2.3.3 Other Biological signals

Other studies have explored eye movement and mouth synchronization inconsistencies as additional biological signals for detection.

2.4 Challenges in Detecting Deepfakes in Compressed Videos

One major challenge in deepfake detection is the impact of video compression algorithms (e.g., H.264, H.265). Compression removes fine details, making many visual detection techniques ineffective. Existing detection models often fail in real-world applications where social media platforms automatically compress videos.

- Many detection methods trained on high-resolution datasets struggle when applied to compressed formats.
- Physiological signals like rPPG remain more resilient to compression, making them an attractive alternative.

- Researchers are now exploring hybrid models that combine spatial-temporal attention with physiological data to improve detection accuracy across different video qualities.

Addressing these challenges, our approach focuses on detecting physiological discrepancies in both high-quality and compressed videos, offering a robust and generalizable solution.

Chapter 3

Research Gaps and Problem Statement

3.1 Research gap

Despite significant advancements in deepfake detection, several challenges remain unresolved:

3.1.1 Impact of Video Compression on Detection Accuracy

Most deepfake detection models are trained on high-quality video datasets, often using c23 compression. However, real-world applications involve higher compression levels (e.g., c40), where crucial visual and physiological features are lost. Current methods struggle to maintain accuracy on highly compressed videos, leading to false negatives. Addressing this gap requires training models on c40-compressed datasets to improve robustness in real-world scenarios.

3.1.2 Enhancing Model Accuracy and Efficiency

While deep learning models such as ResNet-18 and MesoNet have shown promise in detecting deepfakes, there is still room for improvement. Key

areas of enhancement include:

- Hyperparameter tuning to optimize learning rates and batch sizes.
- Architectural modifications such as replacing ResNet-18 with ResNet-101 for improved feature extraction.
- Reducing false positives and false negatives by refining attention-based mechanisms.

3.1.3 A Practical-Knowledge conflict Gap - Deployment of a Web-Based Detection Platform

A major limitation of current research is the lack of user-friendly, accessible solutions for deepfake detection. Developing a web-based deepfake detection platform would allow users to upload videos or live-stream feeds for real-time analysis, making deepfake detection more practical and widely available.

3.1.4 Real-Time Deepfake Detection for Streaming and Live Video Calls

Existing detection methods primarily focus on post-processing analysis, making them ineffective for real-time applications such as video calls, live streams, and social media broadcasts. There is a growing need for real-time deepfake detection systems that can analyze and classify manipulated content during live interactions. Challenges in this domain include:

- Generating and integrating deepfakes in real time for model validation.
- Replacing live webcam feeds with deepfake content using tools like OBS & DeepFaceLive.

- Processing each frame in real-time to identify manipulation before transmission

3.2 Problem Statement

The rapid advancement of deepfake technology poses a significant threat to the authenticity and security of digital media. While substantial progress has been made in detecting manipulated images, deepfake video detection remains a complex challenge, particularly in compressed video formats commonly used on social media platforms. Existing detection methods struggle to maintain accuracy when faced with compression-induced distortions, as they primarily focus on high-quality, uncompressed video data.

Furthermore, many current approaches depend heavily on visual artifacts that become less distinguishable after compression, leading to reduced detection reliability. Additionally, the integration of physiological indicators such as remote Photoplethysmography (rPPG) signals remains underutilized, despite its potential to differentiate real and synthetic videos based on biological inconsistencies.

To address these gaps, this study proposes a robust deepfake detection framework that enhances the effectiveness of rPPG-based detection methods by accounting for compressed video data. The research aims to improve detection accuracy across varying compression levels and explore real-time deepfake detection applications in live streaming and video calls. By leveraging a combination of visual, physiological, and real-time processing techniques, this work seeks to strengthen deepfake detection capabilities and ensure greater reliability in identifying manipulated media.

Chapter 4

Proposed Methodology/ Solution

Our method is designed to detect DeepFake videos by analyzing heartbeat rhythms from facial regions using remote photoplethysmography (rPPG). The methodology follows a systematic flow:

4.1 Dataset Preparation

Datasets available are

- FaceForensics++ (FF++): Contains DeepFake, Face2Face, FaceSwap, NeuralTextures and FaceShifter videos.
- Out of them, we used Real and Deepfake videos for training our model, and evaluated it on Real and Face2Face videos

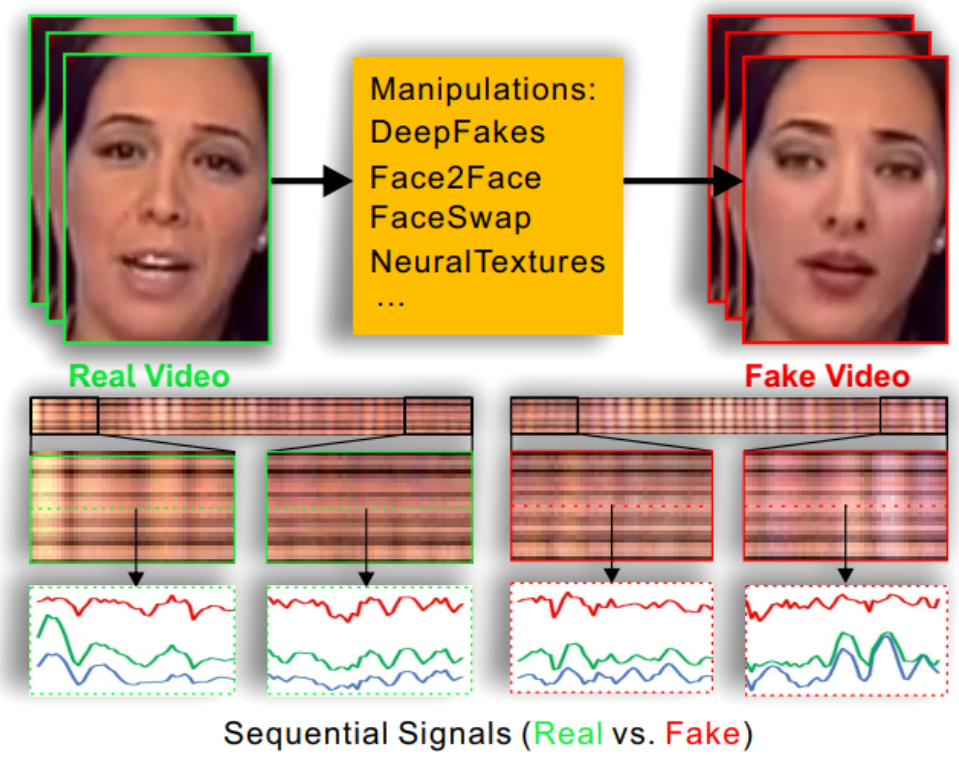


Figure 4.1: An example of a real video and its fake video generated by various manipulations, e. g., Deepfakes, Facce2Face, FaceSwap, etc. It is hard to decide real/fake via the appearance from a single frame. The state-of-the-art Xception fails in this case as in literature review. However, we see that the manipulations easily diminish the sequential signals representing remote heartbeat rhythms.

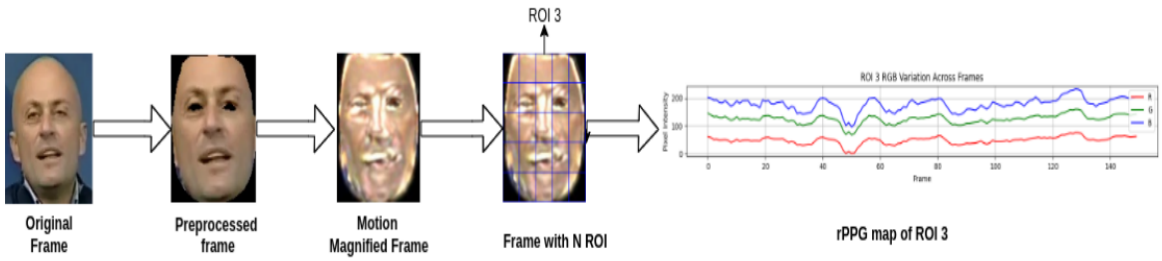


Figure 4.2: Preprocessing flow

4.2 Data Preprocessing and Face Segmentation

4.2.1 Face Detection and Landmark Extraction

- Face Detection: Detect the face in each frame using MTCNN (Multi-task Cascaded Convolutional Networks).
- Facial Landmarks: Identify 81 key landmarks on the face using Dlib.
- Region of Interest (ROI) Selection:
  - Remove eyes and background as they introduce noise.
  - Focus on forehead, cheeks, and under-eye areas, where heartbeat signals are strongest.

4.2.2 Face Tracking and Stabilization

- If multiple faces are detected, retain the one that is closest to the previously detected face.
- Frames without detected faces are discarded (if more than 50 frames are lost, the video is skipped).

4.3 Motion-Magnified Spatial-Temporal Representation (MMSTR)

4.3.1 Motion Magnification for Heartbeat Enhancement

- Why? The heartbeat signal in facial regions is subtle and may not be directly visible.
- How? Apply the Eulerian Video Magnification (EVM) technique to amplify small color changes caused by blood flow.
- Output: A motion-magnified face video where color changes due to the heartbeat are enhanced.

4.3.2 Generating the MMST Map

Divide the face into N non-overlapping blocks (ROIs) (e.g., 5x5 grid = 25 blocks).

Extract the average RGB colour intensity per block over time.

Construct an MMST (Motion-Magnified Spatial-Temporal) Map, where:

- Rows = different facial regions (N blocks)
- Columns = time (frames)
- Values = RGB intensity variations (representing heartbeat patterns)

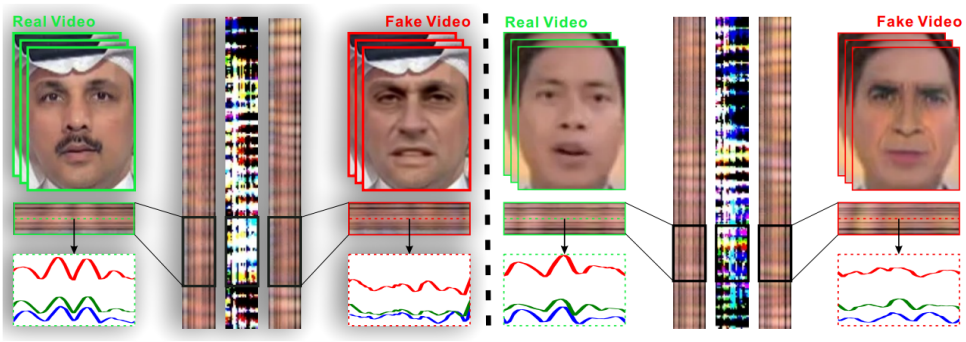


Figure 4.3: Two real-fake video pairs, their MMST maps and the colorful difference maps between real and fake

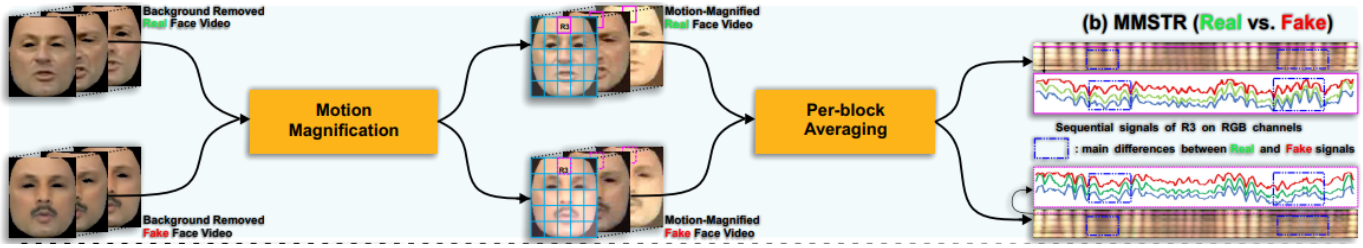


Figure 4.4: (b) MMSTR real vs. fake

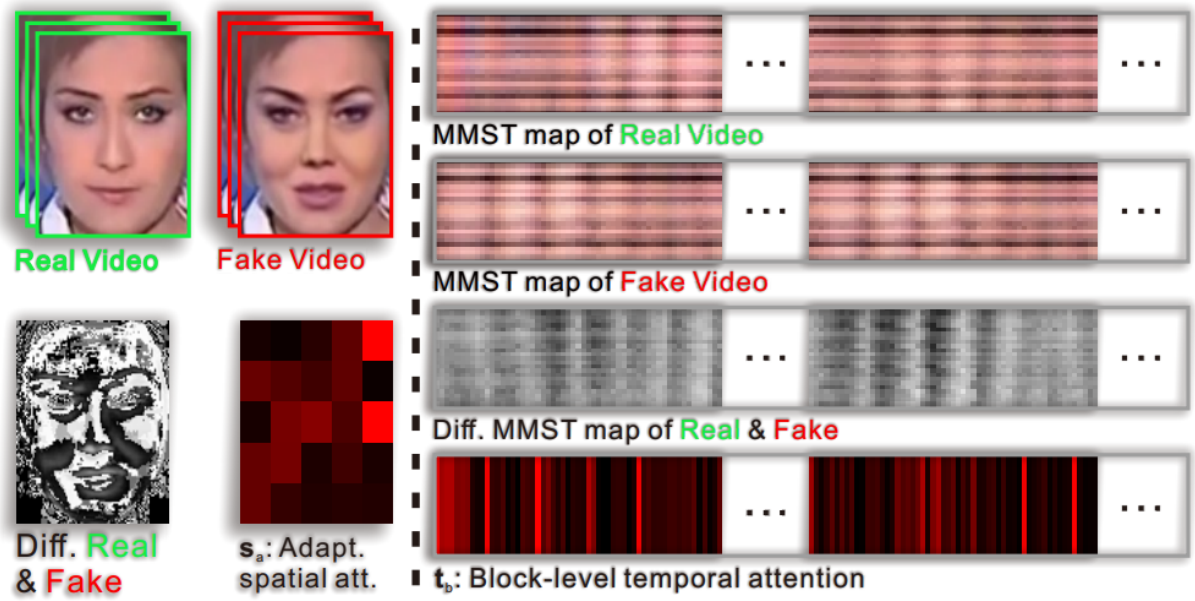


Figure 4.5: : An example of a real video, the corresponding fake video, the difference image between real and fake frames (Diff. Real & Fake), the MMST maps of real and fake videos, the difference map between real and fake MMST maps (Diff. MMST map of Real & Fake), the adaptive spatial attention (sa, i.e., adaptive spatial attention), and the block-level temporal attention (tb).

4.4 Dual-Spatial-Temporal Attentional Network (Dual-ST AttenNet)

DeepRhythm utilizes a dual-attention mechanism to focus on meaningful areas while ignoring noise.

4.4.1 Spatial Attention (Where to Focus?)

Some facial regions provide stronger heartbeat signals than others.

The model applies a Dual-Spatial Attention Mechanism:

- Prior Spatial Attention (Fixed Weights): Focuses on pre-defined robust regions (e.g., under the eyes).
- Adaptive Spatial Attention (Learned Weights): Adjusts dynamically based on video conditions (e.g., lighting changes).

4.4.2 Temporal Attention (When to Focus?)

Some frames contain more distinctive DeepFake artifacts than others.

Two types of temporal attention are applied:

- Block-Level Temporal Attention: Uses LSTM to analyze variations in facial regions over time.
- Frame-Level Temporal Attention: Uses MesoNet (a CNN-based model) to assign importance scores to frames.
- The final weight matrix A = (t * s) X ensures that the model gives higher importance to frames and regions with strong heartbeat signals.

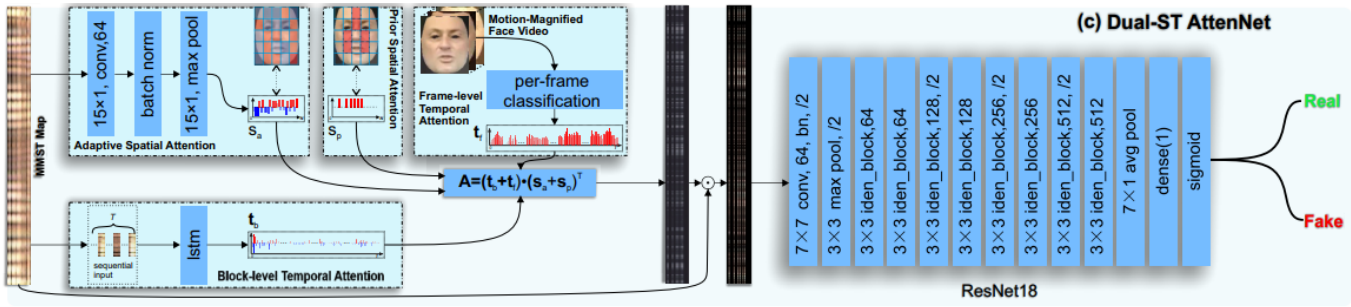


Figure 4.6: (c) Dual-ST AttenNet

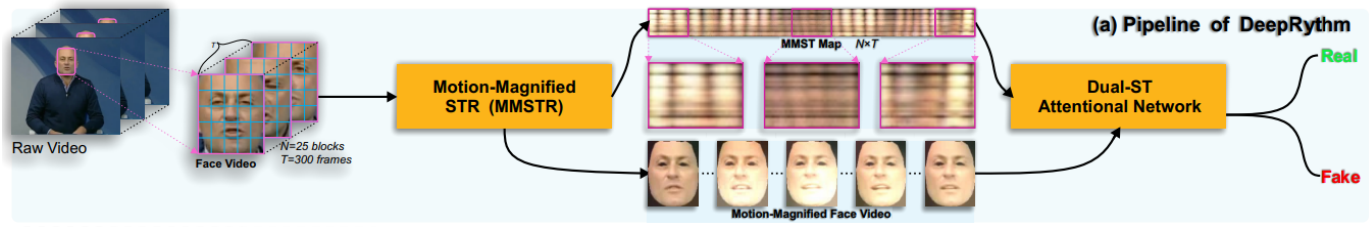


Figure 4.7: (a) Pipeline of DeepRythm

4.5 Deep Neural Network for DeepFake Classification

- The weighted MMST map (with spatial and temporal attention applied) is passed to a deep neural network for classification.
- ResNet18 is used as the final classifier.
- Adam Optimizer is used
- The network is trained to output 1 for fake videos and 0 for real videos.

4.6 Model Training

4.6.1 L2 Regularization:

- Helps in mitigating overfitting by adding a penalty for large weights, preventing the model from relying too heavily on specific features.

- Encourages better generalization by ensuring that the model does not memorize the training data but instead learns meaningful patterns.

4.6.2 Early Stopping:

- Continuously monitors the validation loss during training and stops the process when no further improvement is observed.
- Prevents the model from over-training on the training data, reducing the risk of poor performance on new, unseen data.
- Ensures that the model retains optimal performance without excessive training, leading to better real-world applicability.

4.6.3 Combined Effect:

- Helps in reducing variance by balancing model complexity and performance, preventing overfitting to training data.
- Enhances generalization by ensuring that the learned features remain relevant across different datasets.
- Stabilizes the overall training process, making it more efficient and reducing the risk of unnecessary computations.

4.6.4 Outcome:

- Leads to improved accuracy on the test dataset by preventing the model from becoming too complex or over-specialized.
- Maintains an optimal balance between simplicity and performance, ensuring that the model remains interpretable and effective.

Chapter 5

Results and Discussion

We used the Accuracy metric for the evaluation.

5.1 Accuracy

Accuracy is a measure of the overall correctness of the classification system and is calculated as the ratio of correctly classified instances to the total instances. Accuracy = (TP+TN) / (TP+FN+FP+TN). The model was trained for 500 epochs with a batch size of 32, using the Adam optimizer and binary cross-entropy loss function. To improve generalization and reduce overfitting, L2 regularization was applied to penalize large weight values, and Early Stopping was implemented to halt training when the validation loss stopped improving.

5.1.1 Training and Validation Performance

By Epoch 95, the model achieved:

- Training Accuracy: 65.53%
- Training Loss: 0.9619
- Validation Accuracy: 61.78%

- Validation Loss: 0.9498

The incorporation of L2 regularization helped in reducing model variance by discouraging overly complex weight distributions, while Early Stopping ensured that training was halted before overfitting could occur. The accuracy and loss values fluctuated in the early stages but eventually stabilized. The Accuracy vs. Epochs and Loss vs. Epochs graphs, included in this section, provide a clear visualization of the learning behavior.

5.1.2 Testing Performance on Deepfake

To evaluate real-world generalization, the model was tested on a separate dataset consisting of 20% deepfake data, yielding:

- Test Accuracy: 61.00%
- Test Loss: 0.9789

These results suggest that while the model learned useful patterns from the training data, there is still room for improvement in generalization. The moderate test accuracy could be attributed to class imbalance, feature complexity, and potential overfitting to the training dataset. Further enhancements, such as fine-tuning hyperparameters, adding data augmentation techniques, or exploring alternative architectures, could lead to better performance.

5.1.3 Face2Face (F2F) Testing Results

In addition to deepfake detection, the model was tested on the Face2Face dataset, producing the following results:

- Face2Face Test Accuracy: 61.88%
- Face2Face Test Loss: 0.9943

- DeepRhythm Confidence Score: 58.53%

The performance on Face2Face videos suggests that the model retains its ability to detect manipulated content across different deepfake types, although the accuracy remains moderate.

5.1.4 Graphical Representation

The Accuracy vs. Epochs and Loss vs. Epochs graphs included in this section (Figure 5.1 and 5.2 respectively provide further insight into the model's learning process, showing how accuracy improved and loss decreased over time.

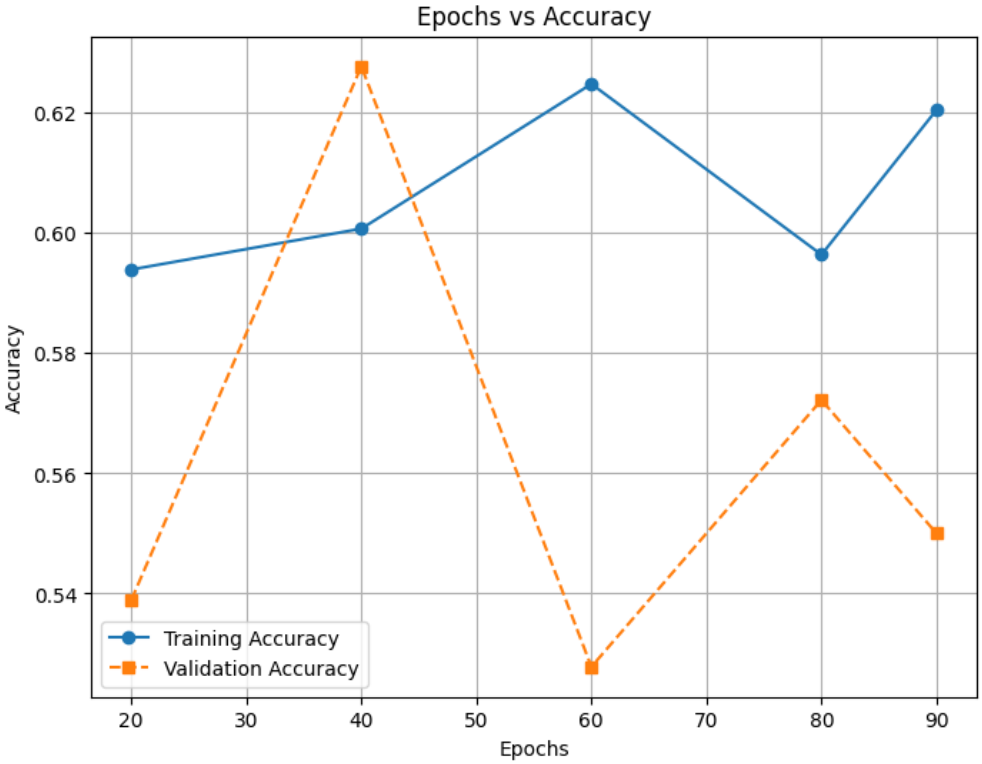


Figure 5.1: Accuracy v/s Epoch

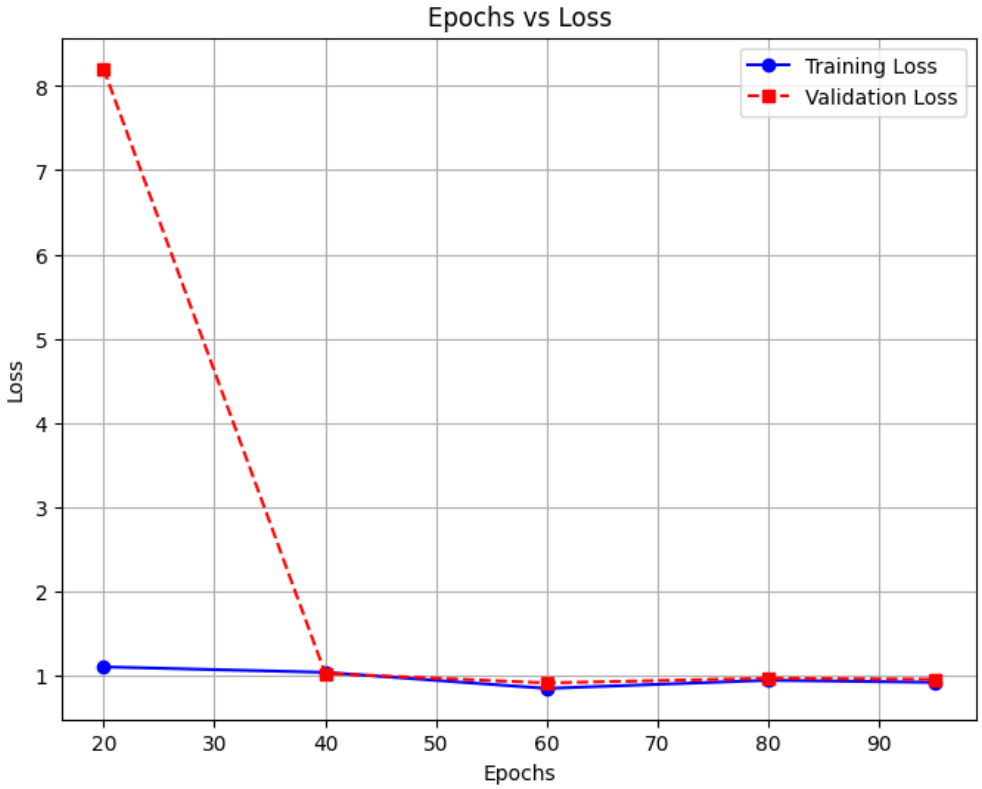


Figure 5.2: Loss v/s Epoch

Chapter 6

Conclusion

In this project, we implemented a deepfake detection model using a ResNet18-based architecture, incorporating L2 regularization and Early Stopping to improve generalization and prevent overfitting. Our approach focused on extracting spatial and temporal features from video frames to classify them as real or fake. The dataset used for training consisted of deepfake and genuine videos, with preprocessing techniques applied to enhance feature extraction.

Achieved the following performance:

- Training Accuracy: 65.53%
- Test Accuracy: 61.00%

Identified challenges related to generalization and robustness, requiring further improvements.

Planned enhancements include:

- Hyperparameter tuning (learning rate, batch size)
- Applying advanced augmentation techniques
- Training on additional datasets like DFDC (Deepfake Detection Challenge)
- Evaluating performance on different video compression levels

Future work will focus on optimizing network architecture, inference speed, and computational efficiency to support real-time deployment. Additionally, we may explore the feasibility of making the model real-time, enhancing its practical applicability for deepfake detection in live scenarios. Optimizing the network architecture, inference speed, and computational efficiency will be key considerations for real-time deployment.

Bibliography

[1] DeepRhythm: Exposing DeepFakes with Attentional Visual Heartbeat Rhythms.

[2] Munawar et al., "Forged Video Detection Using Deep Learning: A Systematic Literature Review," Applied Computational Intelligence and Soft Computing, 2023.

[3] Analysis and Survey on Deepfake Detection and Recognition with Convolutional Neural Networks.

[4] Zhao et al., "Multi-Attentional Deepfake Detection," CVPR 2021.

[5] Maksutov et al., "DeepFake Detection Based on Discrepancies Between Faces and Their Context."

[6] FaceShifter: Towards High Fidelity and Occlusion Aware Face Swapping.

[7] Shruti Agarwal and Hany Farid, "Detecting Deep-Fake Videos from Aural and Oral Dynamics," In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 981–989, 2021.

[8] Giuseppe Boccignone, Sathya Bursic, Vittorio Cuculo, Alessandro D'Amelio, Giuliano Grossi, Raffaella Lanzarotti, and Sabrina Patania, "Deepfakes Have No Heart: A Simple rPPG-Based Method to Reveal Fake Videos," In International Conference on Image Analysis and Processing, pages 186–195. Springer, 2022.

[9] Umur Aybars Ciftci, Ilke Demir, and Lijun Yin, "How Do the Hearts of Deep Fakes Beat? Deep Fake Source Detection via Interpreting Residuals

with Biological Signals," In 2020 IEEE International Joint Conference on Biometrics (IJCB), pages 1–10. IEEE, 2020.

[10] Javier Hernandez-Ortega, Ruben Tolosana, Julian Fierrez, and Aythami Morales, "DeepFakeson-Phys: Deepfakes Detection Based on Heart Rate Estimation," arXiv preprint arXiv:2010.00400, 2020.

[11] Tackhyun Jung, Sangwon Kim, and Keecheon Kim, "DeepVision: Deepfakes Detection Using Human Eye Blinking Pattern," IEEE Access, 8:83144–83154, 2020.

[12] Bahar Uddin Mahmud and Afsana Sharmin, "Deep Insights of Deepfake Technology: A Review," arXiv preprint arXiv:2105.00192, 2021.

[13] Kundan Patil, Shrushti Kale, Jaivanti Dhokey, and Abhishek Gulhane, "Deepfake Detection Using Biological Features: A Survey," arXiv preprint arXiv:2301.05819, 2023.

[14] Jiahui Wu, Yu Zhu, Xiaoben Jiang, Yatong Liu, and Jiajun Lin, "Local Attention and Long-Distance Interaction of rPPG for Deepfake Detection," The Visual Computer, 40(2):1083–1094, 2024.

[15] Yuezheng Xu, Ru Zhang, Cheng Yang, Yana Zhang, Zhen Yang, and Jianyi Liu, "New Advances in Remote Heart Rate Estimation and Its Application to Deepfake Detection," In 2021 International Conference on Culture-Oriented Science & Technology (ICCST), pages 387–392. IEEE, 2021.