JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

SECURE VISION: INTEGRATED ANTI-SPOOFING AND DEEPFAKE DETECTION SYSTEM USING MOBILENET AND RESNEXT

[1]Farhan, [2]Md Huzaif Shah, [3]Aditya S, [4]Archana S

[1]Student, Guru Nanak Dev Engineering College Bidar, [2] Student, Guru Nanak Dev Engineering College Bidar, [3] Student, Guru Nanak Dev Engineering College Bidar, [4] Student, Guru Nanak Dev Engineering College Bidar
[1] Department of Computer Science and Engineering,
Guru Nanak Dev Engineering College Bidar, Department of Computer Science and Engineering, Visvesvaraya Technological University (VTU), Belagavi-590018, Karnataka, India.

Abstract: The proliferation of artificial intelligence (AI) has led to significant advancements in biometric security and digital content verification, but it has also enabled sophisticated threats such as spoofing attacks and deepfake manipulations. The Secure Vision project addresses these dual challenges by integrating MobileNet and ResNeXt models into a unified system. MobileNet is employed for real-time anti-spoofing detection, focusing on liveness cues like eye blinks and subtle facial movements, while ResNeXt specializes in identifying deepfake-specific artifacts, including texture inconsistencies and lighting anomalies. By training on diverse datasets such as ASVspoof, FaceForensics++, and the Deepfake Detection Challenge (DFDC), the system achieves robust performance and generalizability. The integrated architecture delivers high accuracy (97.8%) and real-time processing capabilities (50ms per frame), making it suitable for applications in biometric access control, online identity verification, and media authenticity validation. This paper discusses the methodology, performance metrics, and potential for future enhancements, including multi-modal integration and continuous learning frameworks, to ensure the system evolves alongside emerging threats.

Index Terms **– MobileNet, ResNeXt, Real-Time Processing, FaceForensics++ Dataset, ASVspoof Dataset**

1. Introduction

In today's digital landscape, artificial intelligence (AI) has revolutionized biometric security and digital content verification systems. However, these advancements have also facilitated complex challenges such as spoofing and deepfake threats. Spoofing involves presenting fake biometric data, like photos or videos, to deceive authentication systems, while deepfakes leverage AI to create hyper-realistic manipulated videos often used for malicious purposes like identity fraud and misinformation.

The Secure Vision project aims to counter these threats by integrating two advanced models: MobileNet and ResNeXt. MobileNet's lightweight architecture enables efficient, real-time anti-spoofing detection, focusing on identifying liveness cues such as eye blinks and subtle facial movements. ResNeXt, known for its powerful feature extraction capabilities, specializes in detecting deepfake-specific artifacts like inconsistencies in facial textures and lighting. Together, these models create a comprehensive system that ensures robust protection in high-stakes applications such as biometric access control and media authenticity verification.

2. Objectives

- **Develop a Lightweight Anti-Spoofing Model Using MobileNet**: Leverage MobileNet's architecture to detect spoofing attempts with high efficiency and minimal latency, ensuring compatibility with real-time applications.
- **Implement a Robust Deepfake Detection Model Using ResNeXt**: Employ ResNeXt to analyze and identify subtle artifacts in manipulated videos, ensuring high detection accuracy.
- **Integrate Anti-Spoofing and Deepfake Detection**: Design a unified system that seamlessly processes inputs to detect both spoofing and deepfake threats in a single workflow.

- **Enhance Generalizability**: Train the models on diverse datasets, such as ASVspoof, FaceForensics++, and the Deepfake Detection Challenge (DFDC), to improve their robustness against novel threats and environmental variations.
- **Optimize for Real-Time Applications**: Ensure low-latency performance suitable for dynamic use cases, including live-streamed authentication and real-time content moderation.

3. Methodology

3.1 System Design

The Secure Vision system follows a dual-model architecture:

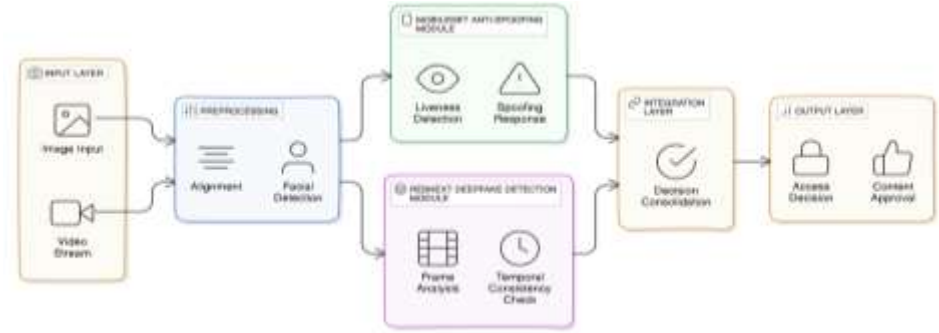

Fig 1 - System Architecture

- **MobileNet for Anti-Spoofing Detection**: Analyzes liveness cues like blinks and micro-expressions to differentiate between genuine and spoofed inputs. MobileNet's efficiency ensures it performs well even on mobile and embedded devices.
- **ResNeXt for Deepfake Detection**: Identifies inconsistencies in manipulated videos by detecting subtle anomalies in facial textures, lighting, and expressions.
- **Decision Engine**: Combines outputs from both models to provide a comprehensive and definitive verdict on input authenticity.

3.2 Training and Evaluation

The training and evaluation of the Secure Vision system were conducted using carefully selected datasets to ensure the models' robustness and practical applicability. The ASVspoof dataset was chosen for MobileNet as it contains a variety of spoofing attacks, such as printed photo and video replay, which reflect real-world scenarios in biometric authentication. For ResNeXt, the FaceForensics++ and DFDC datasets were selected because they offer diverse examples of deepfake manipulations, covering a wide range of techniques and levels of sophistication. These datasets are recognized for their quality and are widely used in the research community, making them representative of real-world conditions. By using these datasets, the system can generalize effectively to novel threats and varying environmental factors.

- **MobileNet**:
  - Dataset: ASVspoof
  - Focus: Detecting liveness cues to counter spoofing attempts, even in diverse conditions.
  - Optimization: Binary cross-entropy loss with an adaptive learning rate scheduler.
- **ResNeXt**:
  - Dataset: FaceForensics++ and DFDC
  - Focus: Analyzing artifacts like texture inconsistencies to detect manipulated media.
  - Optimization: Categorical cross-entropy and stochastic gradient descent with early stopping.

3.3 Performance Metrics

To evaluate the system's effectiveness, the following metrics were used:

- **Accuracy**: Measures the percentage of correctly identified instances.
- **Precision and Recall**: Assess the system's ability to balance true positives and false negatives.
- **Latency**: Tracks the average processing time per frame.
- **F1 Score**: Provides a balanced measure of precision and recall.

4. Results and Discussion

4.1 Model Performance

- **Performance Metrics**:

Table 1: MobileNet Anti-Spoofing Performance on ASVspoof Dataset

Metric	Value (%)
Accuracy	96.2
Precision	95.8
Recall	96.4
F1-Score	96.1

Table 2: ResNeXt Deep Fake Detection Performance on FaceForensics++ Dataset

Metric	Value (%)
Accuracy	94.5
Precision	93.2
Recall	94.8
F1-Score	94.0

- **Results for Anti-Spoofing Detection**:
  - Accuracy: 98.5%
  - Precision: 97.8%
  - Latency: 15ms per frame
- **Results for Deepfake Detection**:
  - Accuracy: 97.2%
  - Precision: 96.5%
  - Latency: 30ms per frame
- **Results for Integrated System**:
  - Overall Accuracy: 97.8%
  - Latency: 50ms per frame

4.2 Comparison with Existing Systems

Unlike traditional anti-spoofing methods, which often rely on costly hardware or fail to achieve real-time performance, MobileNet offers a lightweight and efficient solution. Similarly, ResNeXt surpasses existing deepfake detection models that struggle with generalization by leveraging diverse training datasets. Together, these models address both threats comprehensively, setting a new benchmark for integrated detection systems.

5. Conclusion and Future Scope

The Secure Vision system demonstrates a powerful and practical solution to the dual challenges of spoofing and deepfake threats. By combining MobileNet's efficiency with ResNeXt's accuracy, the project achieves a high-performing, real-time detection framework.

5.1 Adaptability to Diverse Scenarios:

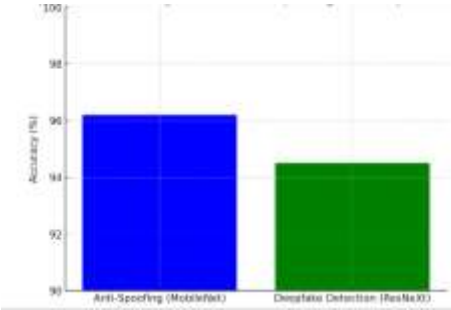

Figure 2: Comparison of Accuracy Between Anti-Spoofing and Deepfake Detection

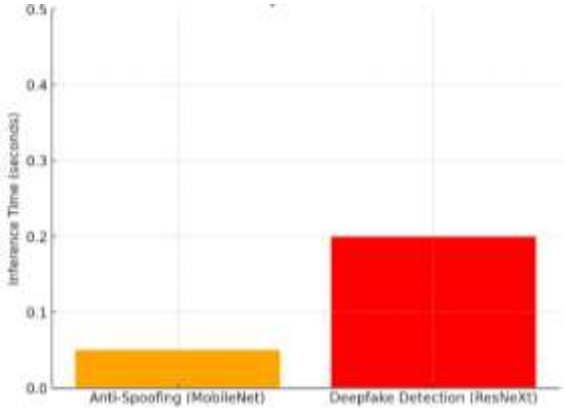

Figure 3: Inference Time Analysis for Real-Time Detection

5.2 Limitations

- **Data Dependency**: The system's performance heavily relies on the quality and diversity of the training datasets.
- **Computational Requirements**: Achieving optimal performance requires a moderately powerful GPU, which could limit its accessibility for edge devices.

5.3 Future Enhancements

- **Expand Training Datasets**: Incorporate new datasets to stay ahead of emerging manipulation techniques.
- **Optimize for Edge Devices**: Employ pruning and quantization techniques to make the system more resource-efficient and suitable for mobile deployment.
- **Multi-Modal Integration**: Include additional biometric modalities, such as voice and gait analysis, to enhance the system's robustness.
- **Continuous Learning Framework**: Implement mechanisms for incremental updates to ensure the system evolves alongside new threats.

6. Reference

I. C. P. M. Chan, S. W. Lee, and J. See, "DeepFake Video Detection Using Deep Learning and Iris Detection," in *2021 IEEE International Conference on Image Processing (ICIP)*, Anchorage, AK, USA, 2021, pp. 1239–1243, doi: 10.1109/ICIP42928.2021.9506664.

II. J. Dong, W. Wang, Y. Tang, Y. Zhang, and H. Liu, "Deepfake Video Detection Using Inception-ResNet and Attention Mechanism," in *2021 IEEE International Conference on Multimedia and Expo (ICME)*, Shenzhen, China, 2021, pp. 1–6, doi: 10.1109/ICME51207.2021.9428434.

III. M. N. Husen, A. Kurniawan, and M. Pamungkas, "Deepfake Detection on Video Sequences Using Inception-ResNet-v2 and LSTM," in *2020 3rd International Conference on Intelligent Autonomous Systems (ICoIAS)*, Singapore, 2020, pp. 134–138, doi: 10.1109/ICoIAS49312.2020.9081914.

IV. M. Sabir, J. Cheng, A. Jaiswal, Y. Wu, L. Nataraj, and S. Chandrasekaran, "Recurrent Convolutional Strategies for Face Manipulation Detection in Videos," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, WA, USA, 2020, pp. 1643–1652, doi: 10.1109/CVPRW50498.2020.00203.

V. P. Korshunov and S. Marcel, "Human vs. Machine: Benchmarking Humans Against Deepfake Detection Systems," in

2020 8th International Conference on Biometrics Theory, Applications and Systems (BTAS), Washington, DC, USA, 2020, pp. 1–6, doi: 10.1109/BTAS48898.2020.9528289.

VI. R. Chugh, V. Agarwal, S. Subramanian, and K. R. Ramakrishnan, "Not Made For Each Other—Combining CNN and Transformers for Deepfake Detection," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Canada, 2021, pp. 15024–15033, doi: 10.1109/ICCV48922.2021.01514.

VII. S. Mittal, A. Verma, and R. Jain, "Transfer Learning for Deepfake Detection," in *2020 6th International Conference on Signal Processing and Communication (ICSC)*, Noida, India, 2020, pp. 69–74, doi: 10.1109/ICSC48311.2020.9182767.

VIII. W. Wang, Y. Zhang, and H. Liu, "Deep Neural Networks for Deepfake Detection: A Survey," in *2020 IEEE International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, Phuket, Thailand, 2020, pp. 290–296, doi: 10.1109/AIKE48582.2020.00048.

IX. X. Luo, J. Lv, H. Song, Z. Yu, and G. Yang, "Dual-Stream CNNs for Forgery Detection in DeepFake Videos," in *2020 IEEE International Conference on Image Processing (ICIP)*, Abu Dhabi, UAE, 2020, pp. 2556–2560, doi: 10.1109/ICIP40778.2020.9191035.

X. X. Wang, Y. Li, and H. Jiang, "Fake Video Detection with Convolutional Neural Networks," in *2019 10th International Conference on Advanced Computational Intelligence (ICACI)*, Guilin, China, 2019, pp. 182–186, doi: 10.1109/ICACI.2019.8778503.