DEEP FAKE IDENTIFICATION

Presented By : Team B

Chetali Bandodkar, Divyanshu Ranjan, Pushkar Aditya, Sunny Kumar

Mentor: Mr. Bappaditya

OUTLINE

ABSTRACT

This presentation covers deepfake detection methods, including CNN-based classification, frequency analysis, and transformer models. It explores key datasets like FaceForensics++ and DFDC, highlighting challenges such as adversarial attacks and real-time detection. Future directions focus on multi-modal integration and robust techniques to improve digital media security.

INTRODUCTION TO DEEPFAKES

- Definition: Synthetic media where a person's likeness is replaced with someone else's using AI
- Created primarily using GANs, Autoencoders, and CNNs
- Growing threat to information integrity, privacy, and security
- Applications: Political manipulation, fraud, misinformation, non-consensual content

RESEARCH WORK

Paper Title	Authors	Publication Year	Research Area	Key Architecture/Method	Reported Accuracy on Key Datasets	Brief Summary of Contribution
MesoNet: a Compact Facial Video Forgery Detection Network	Afchar et al.	2018	Image/Video	Lightweight CNN focused on mesoscopic properties	95.23% on FaceForensics	Demonstrates high accuracy with a shallow network by targeting mid-level frequency features, robust to compression and suitable for mobile devices.
Exposing Deep Fakes Using Inconsistent Head Poses	Yang et al.	2018	Image/Video	CNN head pose estimator + SVM classifier for inconsistencies between face regions	99.1% on UADFV, 97.4% on DeepfakeTIMIT	Detects deepfakes by analyzing discrepancies in 3D head pose estimations from the whole face and the central facial region, does not require temporal data.
ASVspoof 2019: Future Horizons in Spoofed and Fake Audio Detection	Todisco et al.	2019	Audio	Light CNN with max-feature-map activations and squeeze-and-excitation blocks	95.19% on ASVspoof 2019	Focuses on artifacts in the constant-Q transform domain and phase inconsistencies, effective against known and unknown attacks, low computational cost.

IMAGE DEEPFAKE DETECTION

The Problem

Mostly focused on face swapping & face generation

- Uses techniques like:
  - StyleGAN
  - CycleGAN
  - StarGAN
- It Gets harder to detect with the naked eye
- Social media platforms are flooded with fake images

Methodologies

Method I: Frequency Domain Analysis

Method 2: CNN-based Classification

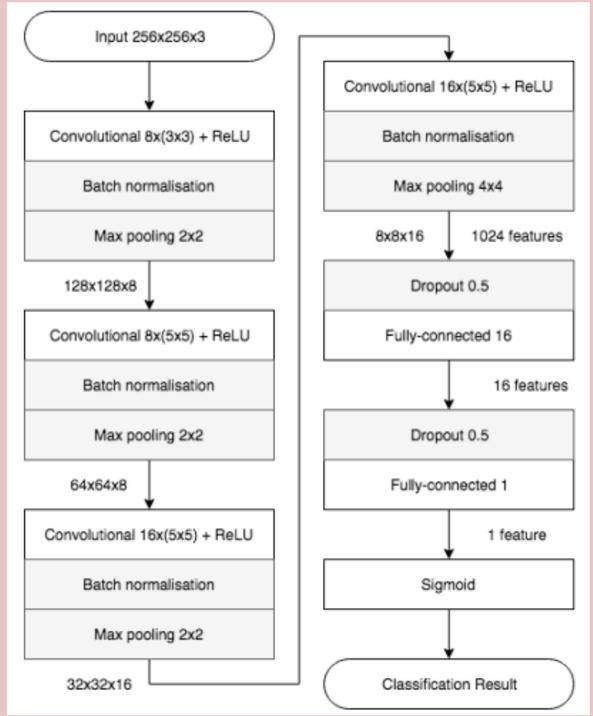
Dataset:

- FaceForensics++: 1000 original videos, manipulated using different methods
- Celeb-DF: 5,639 high-quality deepfake videos of celebrities
- DFFD: Diverse Fake Face Dataset with 100,000+ images

MesoNet: a Compact Facial Video Forgery Detection Network

Architecture:

- Lightweight CNN (4-5 layers)
- Specialized in mesoscopic properties of images



Results:

- 95.23% accuracy on FaceForensics
- Works well on low-resolution images
- Runs on mobile devices

Method:

- Focuses on mesoscopic features
- Lower computational requirements
- Effective against compression artifacts

Challenges

Generalization to new manipulation methods

Robustness against compression

Real-world deployment issues

VIDEO DEEPFAKE DETECTION

Video Deepfake Artifacts

- Unnatural Blinking Patterns
- Lip-Sync Errors
- Facial Boundary Issues
- Lighting & Shadow Inconsistencies
- Head & Eye Movements

Video Deepfake Detection Techniques

- Frame-Level Analysis (Image-Based Detection)
- Temporal Analysis (Motion-Based Detection)
- Audio-Visual Inconsistencies
- GAN Fingerprint Detection
- Transformer-Based Detection

Methods

- Method 1: Frame-by-Frame Analysis
- Method 2: Spatio-Temporal Analysis

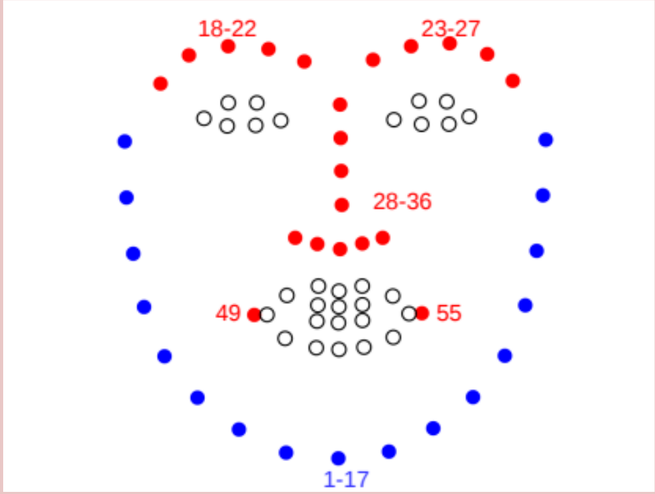
Dataset

- FaceForensics++
- DFDC (DeepFake Detection Challenge)
- Celeb-DF
- DeeperForensics-1.0

Exposing Deep Fakes Using Inconsistent Head Poses

Architecture:

- CNN head pose estimator
- SVM classifier for discrepancy detection



Results:

- 99.1% accuracy on UADFV dataset
- 97.4% accuracy on DeepfakeTIMIT
- Robust against various manipulation methods

Method:

- Analyzes inconsistencies between head pose and facial landmarks
- Detection based on 3D rotation inconsistencies
- Doesn't require temporal data (works on single frames)

Challenges

- High-quality deepfakes are becoming harder to detect as GANs evolve.
- Adversarial attacks can trick deepfake detection models.
- Compressed videos degrade detection accuracy by removing key artifacts.
- Real-time deepfake detection is computationally expensive and requires powerful GPUs.

AUDIO DETECTION

Audio Deepfake Detection Method

- Signal Processing-Based Methods
- Signal Processing-Based Methods
- Deep Learning-Based Methods
- Real-World Testing & Robustness Checks
- Hybrid Approaches

Audio Deepfake Detection Challenges

- Voice cloning & speech synthesis
- Realistic prosody (rhythm, stress, intonation)
- Variable audio quality in real-world settings
- Phone call quality often masks artifacts

Methods

- Method 1: Spectral Analysis
- Method 2: Raw Waveform Analysis

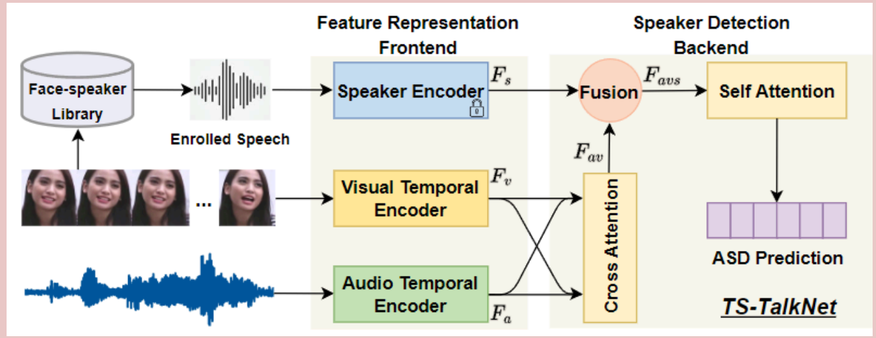
Dataset

- FakeAVCeleb (multimodal dataset)
- DFDC (DeepFake Detection Challenge)
- ASVspoof(2015,2017,2019)

Exposing Deep Fakes Using Inconsistent Head Poses

Architecture:

- Light CNN with max-feature-map activations
- Squeeze-and-excitation blocks for channel attention



Results:

- 95.19% accuracy on ASVspoof 2019 dataset
- Effective against both known and unknown attack types
- Low computational requirements for real-time use

Method:

- Focus on artifacts in the constant-Q transform domain
- Detection of phase inconsistencies

Challenges

- Voice cloning & speech synthesis
- Realistic prosody (rhythm, stress, intonation)
- Variable audio quality in real-world settings
- Phone call quality often masks artifacts
- Easy to distribute (podcasts, phone calls, voice messages)

FUTURE SCOPE

Advancing detection through generalization, multi-modal analysis, and real-world robustness.

Generalization & Novel Attacks:

- Improve detection of unseen forgery/spoofing methods across video & audio.

Multi-Modal Integration:

- Combine audio (lip-sync, voice) & visual (pose, facial features) cues for enhanced accuracy.

Real-World Robustness:

- Address noise, compression, & environmental variations for practical applications.

Real-time & Efficient Models

- Develop lightweight models for mobile and real-time processing.

Enhanced Analysis:

- Refine head pose, 3D facial analysis, and audio forensic tools.
- Improve deepfake voice conversion detection.

FUTURE TECHNOLOGY

State-of-the-Art Detection

Commercial & Research Solutions:

Sensity AI: Multi-modal deepfake detection platform
- Uses transformer-based architecture
- 98.7% accuracy across diverse manipulation types

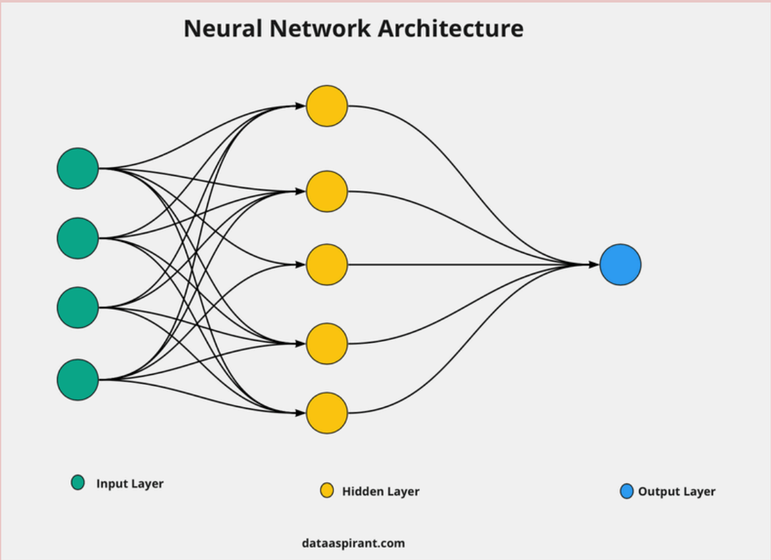
BioID DeepFake Detection: Liveness detection + manipulation analysis
- Focuses on physiological inconsistencies
- Specialized in real-time detection for authentication

Microsoft Video Authenticator: Analyzes facial boundaries and blending
- Provides confidence score for manipulation probability
- Based on Face Forensics research

Key Technologies Used in SOTA Solutions

Neural Network Architectures:

- GANs: Understanding generator patterns
- Autoencoders: For feature extraction and reconstruction errors
- CNNs: Spatial feature analysis
- RNNs/LSTMs: Temporal inconsistency detection
- Transformers: Context-aware feature analysis



Detection Approaches

- Frequency domain analysis
- Biological signal inconsistencies
- Temporal coherence analysis
- Attention-based inconsistency detection

REFERENCES

1. Afchar, D., et al. "MesoNet: a Compact Facial Video Forgery Detection Network"
2. Yang, X., et al. "Exposing Deep Fakes Using Inconsistent Head Poses"
3. Todisco, M., et al. "ASVspoof 2019: Future Horizons in Spoofed and Fake Audio Detection"

THANK YOU

Presented By :

Chetali Bandodkar, Divyanshu Ranjan, Pushkar Aditya, Sunny Kumar

Mentor: Mr. Bappaditya