Deepfake Detection in Videos

Presented by:

Name: Ali Raafat
ID: 202100348

What is a Deepfake?

- Deepfake technology uses AI, especially **GANs (Generative Adversarial Networks)**, to manipulate videos, making people appear to say or do things they never did.

Problem Statement:

- Fake videos spread misinformation, cause political issues, and create security risks.
- Need **robust detection models** to counter these threats.

Objective:

- Build an AI model that accurately detects **deepfake videos** using machine learning.

Deepfake Generation Techniques

How Are Deepfakes Created?

1. Autoencoders

Uses an encoder-decoder network. Learns to reconstruct faces but replaces the target face with a fake one. Works well but struggles with high-resolution videos.

2. Generative Adversarial Networks (GANs)

Most advanced deepfake technique. Consists of two neural networks:

- Generator (creates fake images).
- Discriminator (tries to detect fakes). The generator improves continuously, making deepfakes harder to detect.

Examples of GANs Used for Deepfakes:

- **DeepFake Autoencoder** – Early face-swapping models.
- **StarGAN** – Can manipulate facial attributes.
- **StyleGAN** – Creates hyper-realistic faces

Research in Deepfake Detection

- **FaceForensics++ (FF++)** – Benchmark dataset for deepfake detection.
- **XceptionNet Model** – CNN-based model for fake detection.
- **Capsule Networks** – Detects inconsistencies in facial movement and lighting.
- **Hybrid Approaches** – Combining CNN with transformers for better accuracy.

1. FaceForensics++ (FF++) – Benchmark Dataset for Deepfake Detection

What is FaceForensics++ (FF++)?
FaceForensics++ is one of the most widely used datasets for deepfake detection. It contains over 1,000 real and fake videos, manipulated using various deepfake techniques. These videos are compressed at different levels to simulate real-world scenarios.

Why is it important?

- **Provides a standardized benchmark for evaluating deepfake detection models.**
- **Includes various deepfake techniques, such as Deepfakes, Face2Face, FaceSwap, and NeuralTextures.**
- **Researchers can test their models on different compression levels (low, medium, high), making their detection algorithms more robust.**

Limitations:

- **Many detection models trained on FF++ struggle to generalize to newer deepfake methods not present in the dataset.**
- **Focuses mainly on face manipulation rather than full-body deepfake detection.**

2. XceptionNet Model – CNN-Based Model for Fake Detection

What is XceptionNet?
XceptionNet is a deep convolutional neural network (CNN) that is widely used for image classification and has been adapted for deepfake detection. It is based on depthwise separable convolutions, making it more efficient than traditional CNNs.

How does it work in deepfake detection?

- **Extracts fine-grained spatial features from frames of videos.**
- **Detects artifacts introduced during deepfake creation (e.g., blending errors, unnatural textures).**
- **Works well on datasets like FaceForensics++.**

Strengths:
Achieves high accuracy (~90%) on FaceForensics++ dataset.
Efficient due to its lightweight architecture.
Detects pixel-level inconsistencies in manipulated images.

Weaknesses:
Struggles with generalization – may not perform well on deepfake videos created with different techniques.
Does not consider temporal inconsistencies (motion artifacts across frames).

3. Capsule Networks – Detecting Inconsistencies in Facial Movement & Lighting

What are Capsule Networks?

Capsule Networks (CapsNets) were introduced as an improvement over traditional CNNs. Unlike CNNs, which focus only on spatial features, Capsule Networks preserve hierarchical relationships between features (e.g., eyes, nose, mouth positioning).

How do they help in deepfake detection?

- **Detect subtle inconsistencies in facial structure and lighting.**
- **Analyze facial expressions and movement patterns to determine if a video is real or fake.**
- **More robust against adversarial attacks compared to traditional CNNs.**

Strengths:

Better at understanding spatial relationships in images.

Can detect distortions in facial expressions and micro-movements.

Weaknesses:

Computationally expensive, requiring high processing power.

Not widely adopted yet due to complex implementation.

4. Hybrid Approaches – Combining CNNs with Transformers for Better Accuracy

Why use a hybrid model?
Traditional CNN-based models, like XceptionNet, focus only on spatial features (single frames). However, deepfake videos contain inconsistencies across frames, which means analyzing sequences (temporal features) is crucial.

Hybrid Approach: CNN + Transformer

- **CNN Component: Extracts spatial artifacts from individual frames.**
- **Transformer Component: Analyzes sequences of frames to capture temporal inconsistencies.**

Example Model:
CNN + Transformer Model

- **The CNN detects pixel-level manipulations in each frame.**
- **The Transformer captures motion irregularities, such as unnatural blinking or head movement inconsistencies.**

Why is this approach better?
Improves detection accuracy across different deepfake datasets.
More robust to new deepfake techniques than CNN-only models.
Captures both spatial and temporal inconsistencies, making it harder for deepfakes to bypass detection.

Challenges:
Requires high computational power.
Needs large datasets for training to avoid overfitting.

Deepfake Evolution

- Older deepfakes were low quality, but modern ones: Have natural facial expressions. Show realistic lighting and shadows. Can synchronize lips with speech.

- **Paper Used**: **DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection**

Key Findings from the Paper:

Deepfake methods rely on **GANs, Autoencoders, and Variational Autoencoders (VAEs)**.

Detection models typically use **Convolutional Neural Networks (CNNs) for spatial analysis**.

Temporal inconsistencies (frame-by-frame differences) are crucial for better accuracy.

Hybrid models (CNN + Transformers) outperform traditional CNN-only models.

We should use a Hybrid CNN + Transformer Model that combines:

CNNs (to detect frame-level artifacts).

Transformers (to capture temporal inconsistencies).

Temporal Analysis (Frame-by-Frame Inconsistencies)

Deepfakes often fail to generate **consistent facial features** across multiple frames.

Blink detection: Many deepfakes fail to replicate natural eye blinks.

Head movement tracking: Checks if head movements match body motion.

Example:

A real video has **natural blinks & smooth head movements**, while a deepfake might show **jerky motion or inconsistent blinking**.

Frequency Analysis (Fourier Transform Methods)

Real images and deepfakes differ in **frequency distribution**. This technique converts images into **frequency space to detect hidden artifacts**.

Best Approach (According to the Paper):

Combining **CNN + Transformers** → Works best for both spatial and temporal deepfake detection.

Why This Model?

CNN extracts fine-grained facial features to detect pixel-level manipulation.

Transformer analyzes sequential frame differences to detect motion inconsistencies.

Best of both worlds – spatial & temporal deepfake detection.

Model Architecture

Preprocessing: Convert videos into frames & normalize data

CNN Feature Extraction: Detects frame-level artifacts.

Transformer Encoder: Captures inconsistencies across multiple frames.

Fully Connected Layers (Classifier): Predicts "Real" or "Fake".

Deepfake Datasets – What Data is Used?

The Paper Reviews Various Deepfake Datasets Used for Training and Testing

Dataset	Size	Type	Best Use Case
FaceForensics++ (FF++)	1,000+ videos	Videos	General deepfake detection
DeepFake Detection Challenge (DFDC)	470GB	Videos	Large-scale training
Celeb-DF (V2)	5,639 videos	Videos	**High-quality deepfake detection**

Best Dataset (According to the Paper): Celeb-DF (V2)

Most realistic deepfake dataset.

Challenging for AI models, making it a strong benchmark.

Conclusion

Deepfakes are **becoming harder to detect**.

Best detection models must analyze **both spatial & temporal features**.

Hybrid CNN + Transformer models perform best for real-world deepfake detection.

The **Celeb-DF (V2) dataset** is one of the most challenging benchmarks.

References:

- **"DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection"**
- **Celeb-DF (V2) Dataset**