FACE FORGERY DETECTION USING BLENDING BOUNDARY ANALYSIS (FACE X-RAY)

Nguyen Hong Cuong [1,2]

[1] University of Information Technology, Ho Chi Minh City, Vietnam
[2] Vietnam National University, Ho Chi Minh City, Vietnam

What ?

We introduce **Face X-ray**, a novel framework for detecting forged faces (Deepfakes) in videos. Unlike traditional methods that try to learn specific artifacts of a generation model (which leads to overfitting), our approach focuses on the **blending step**.

- We assume that most face forgeries involve merging a fake face into a background image.
- Face X-ray detects the **blending boundary** caused by the discrepancies in noise statistics or resolution between the two sources.
- This allows the system to detect **unknown/unseen attacks** effectively.

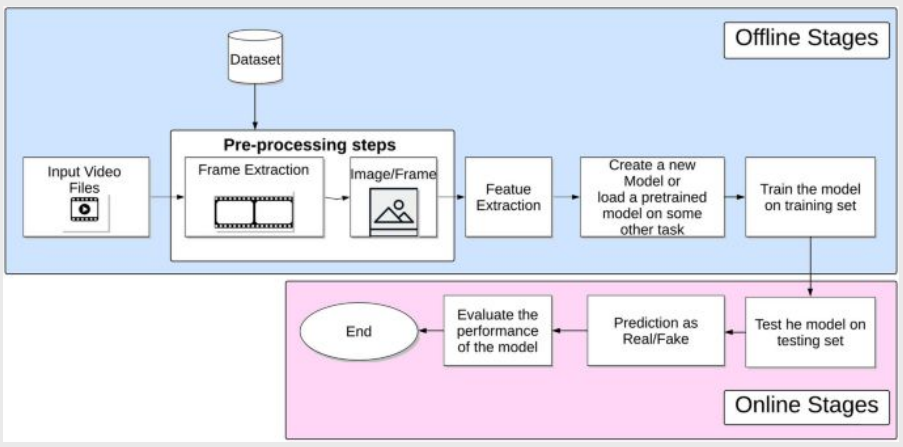
Why ?

- **The Threat:** Deepfakes created by GANs and Autoencoders are becoming increasingly realistic, posing severe threats to identity verification (eKYC) and information integrity (fake news).
- **The Problem:** Current detectors (like XceptionNet) perform well on known attacks but fail significantly (accuracy drops to ~50%) on **unseen attacks** (new generation methods).
- **Our Solution:** We need a generalizable approach. By focusing on the fundamental "blending" evidence rather than specific visual artifacts, our method achieves high generalization capability.

Overview

The proposed system takes a video frame as input and decomposes it into a grayscale signal called "Face X-ray". This signal reveals the boundary where a fake face has been spliced into the background.

Input Image -> Face X-ray (Grayscale) -> Classifier -> Real/Fake.


Offline Stages

Dataset

Pre-processing steps

Input Video Files

Frame Extraction

Image/Frame

Featue Extraction

Create a new Model or load a pretrained model on some other task

Train the model on training set

Online Stages

Test he model on testing set

Prediction as Real/Fake

Evaluate the performance of the model

End

Description

1. Face Detection & Preprocessing

The system extracts frames from the input video. We use a standard face detector to locate the Region of Interest (ROI) and align the face.

- *Key action:* Identifying the facial area to ignore background noise.

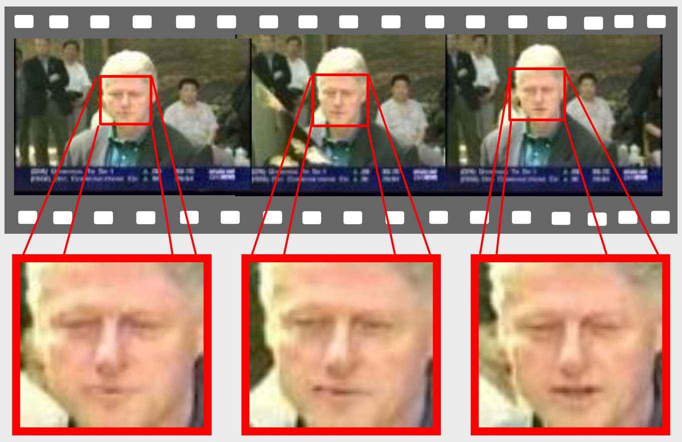


Figure 1. Detection results with a high threshold.

3. Classification

- The generated Face X-ray signal is fed into a lightweight classifier (e.g., HRNet or ResNet).

2. Blending Boundary Prediction (The Core)

This is the most critical step. Since a fake face is a composition of two different images (source face and target background), they have different noise signatures/error levels.

- **Face X-ray Generation:** The model predicts a grayscale mask where white pixels represent the **blending boundary**.
- **Real Face:** The image is consistent, so the Face X-ray is fully black.
- **Fake Face:** A distinct white outline appears around the face or manipulated region.

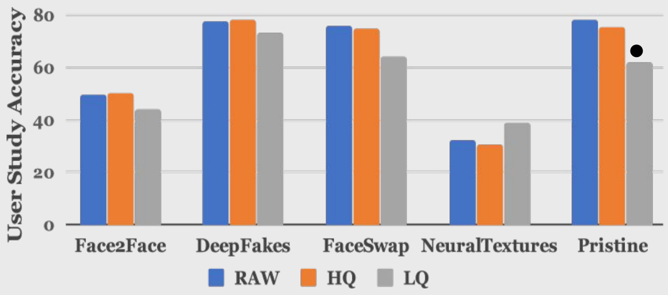


Figure 3. FaceForensics++ Quoted from ar5iv.labs.arxiv.org/html/1901.08971

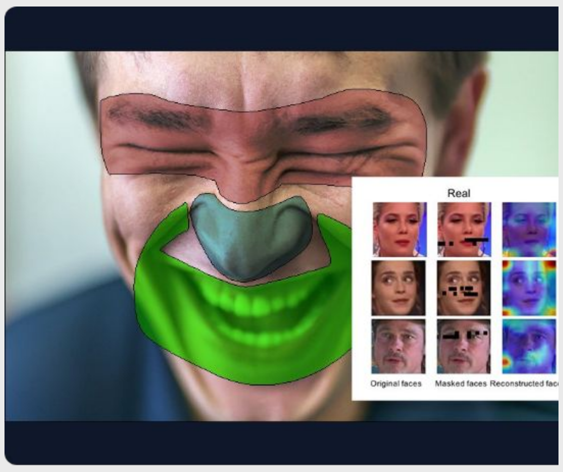


Figure 4. Face X-ray blending boundary examples.

- Based on the existence and shape of the boundary, the system classifies the video as **Real** or **Fake**.
Result: Achieved **>90% Accuracy** on FaceForensics++ and maintained high AUC (>70%) on unseen datasets like DFDC.

NII

Nguyễn Hồng Cường – University of Information Technology, Ho Chi Minh City, Vietnam
Email : cuongnh.20@grad.uit.edu.vn