PhD thesis subject

# FACESWAPPING VIDEO DETECTION

Keywords : Machine learning, forged videos, forensics

## Context

The GREYC laboratory (UMR CNRS 6072) is located in Normandy, and has over 230 members, including around 100 permanent researchers and lecturers, in 6 research teams covering many areas of computer science, automation and electronics.

GREYC's SAFE cybersecurity research team, conducts research activities in the field of computer security around 3 themes: 1) Biometrics, 2) Architecture and security model and 3) Forensics.

The team is offering a PhD internship on deepfake video detection.

## Problem statement

The rapid advancement of deep learning has facilitated the creation of highly realistic face-swapping techniques, which are increasingly used in malicious contexts, such as disinformation campaigns, identity theft, and other forms of fraud. These manipulated images and videos, often referred to as deepfakes, present significant challenges for existing detection systems, especially with the proliferation of easy-to-use generative AI tools.

Current state-of-the-art face-swapping methods leverage architectures such as StyleGAN and neural rendering techniques, achieving high visual fidelity while maintaining temporal and spatial consistency. However, detection systems often struggle with manipulated content due to compression artifacts, low-resolution inputs, and the lack of generalizability across diverse datasets. Recent research highlights the importance of incorporating both spatial and frequency-domain features to detect manipulation effectively under such challenging conditions [1,2]. Spatial features capture texture and pixel inconsistencies, while frequency-domain analysis identifies subtle spectral artifacts, such as those introduced by compression or neural network-based manipulations like StyleGAN. Residual signal analysis further refines detection by exposing differences between manipulated and original content, effectively highlighting hidden artifacts [3]. State-of-the-art models employ hybrid approaches, such as attention-based U-Nets and spectral feature learning, to generalize across manipulation types and maintain efficiency for real-world applications [4,5]. These advancements demonstrate that combining spatial and frequency insights is pivotal for overcoming challenges like compression and low-resolution input.

The computational demands of many existing deepfake detection models also pose a barrier to their deployment in real-world applications where lightweight, real-time solutions are essential. There is an urgent need to design detection architectures that balance robustness and efficiency while capitalizing on residual signal analysis to capture subtle manipulation artifacts.

This thesis aims to address these challenges by developing a lightweight deep learning framework for face-swapping detection. The approach will integrate signal residual analysis and spectral feature extraction to enhance the detection of manipulation artifacts. By prioritizing efficient architectures, the research will contribute to accessible and scalable solutions for deepfake detection.

## Objectives

1. Literature Review : Overview of face-swapping technologies and associated challenges, current deepfake detection methods and lightweight architectures and residual signal analysis technique

2. Develop a neural network architecture optimized for size and complexity, leveraging techniques such as MobileNet, EfficientNet, or quantized variants of convolutional networks.

3. Investigate residual signals (differences between the original image and a filtered version) to detect face-swapping artifacts using image processing approaches such as:
   - Frequency domain analysis (wavelets, Fourier transform).
   - Detection of gradients and local inconsistencies.
   - Residual artifacts from compression.

4. Integrate these residual signals with deep learning architectures to enhance detection accuracy and robustness.

5. Validate the approach using standardized datasets (e.g., FaceForensics++, DeepFakeDetection) and novel manipulations.

Particular attention will be paid to the scalability of the methods developed, *i.e.* the ability to detect modifications independently of the length of the video, and their positioning in the video.

## LOCATION

Université Caen Normandie
GREYC lab., SAFE Team
Bat F, ENSICAEN
6, bd Maréchal Juin, 14 032 – Caen

## SKILLS REQUIRED

- A solid background in machine learning
- Solid knowledge and experience in image/video processing, deep learning and programming (Python, TensorFlow/PyTorch)
- Interest in ethical and regulatory considerations surrounding video manipulation.

## HOW TO CANDIDATE?

To apply, send an email to the supervisors with a dossier including your CV, cover letter, academic transcripts for the last two years of your studies, and any additional documents that could strengthen your application (e.g., recommendation letters).

## CONTACT

Christophe Charrier (christophe.charrier@unicaen.fr)
Emmanuel Giguet (emmanuel.giguet@unicaen.fr)

## REFERENCES

[1] Kaur, A., Noori Hoshyar, A., Saikrishna, V. *et al.* Deepfake video detection: challenges and opportunities. *Artif Intell Rev* **57**, 159 (2024). https://doi.org/10.1007/s10462-024-10810-6

[2] Lee, H., Lee, C., Farhat, K., Qiu, L., Geluso, S., Kim, A. and Etzioni, O., 2024. The Tug-of-War Between Deepfake Generation and Detection. *arXiv preprint arXiv:2407.06174*.

[3] Nguyen, T., Chen, H., & Lee, D., 2023. Hybrid spatial-spectral approaches for robust deepfake detection. *Springer Open Journal*.

[4] Zhao, X., Wang, Y., and Li, J. (2023). *Multi-Modal Deepfake Detection with Spatial-Frequency Fusion*. IEEE Transactions on Information Forensics and Security.

[5] Liu, Z., Wang, F., & Zhang, X., 2023. Multi-stream learning for deepfake detection: Integrating spatial and frequency features. *IEEE Transactions on Information Forensics and Security*.