NGHIÊN CỨU VÀ PHÁT TRIỂN HỆ THỐNG PHÁT HIỆN GIẢ MẠO KHUÔN MẶT DỰA TRÊN PHÂN TÍCH ĐƯỜNG BIÊN PHA TRỘN

Nguyễn Hồng Cường

250202005

Tóm tắt

- Lớp: CS2205.SEP2025
- Link Github của nhóm:

  https://github.com/cuongnh20-uit/CS2205.SEP2025-face_x_ray

- Link YouTube video:

  https://www.youtube.com/channel/UCV4wEu25dSKjhfjqMndV3pA

Nguyễn Hồng Cường
250202005

Bối cảnh & Thách thức

Sự bùng nổ: Các kỹ thuật sinh ảnh (GANs, Autoencoders) ngày càng tạo ra khuôn mặt giả tinh vi.

Nguy cơ tiềm ẩn: Lừa đảo xác thực danh tính, tin giả (fake news), tống tiền và bôi nhọ.

Hạn chế hiện tại: Các mô hình thường bị **Overfitting** (quá khớp) vào dữ liệu huấn luyện.

Vấn đề cốt lõi: Hiệu quả giảm sút nghiêm trọng khi đối mặt với các kiểu tấn công mới (Unseen Attacks).

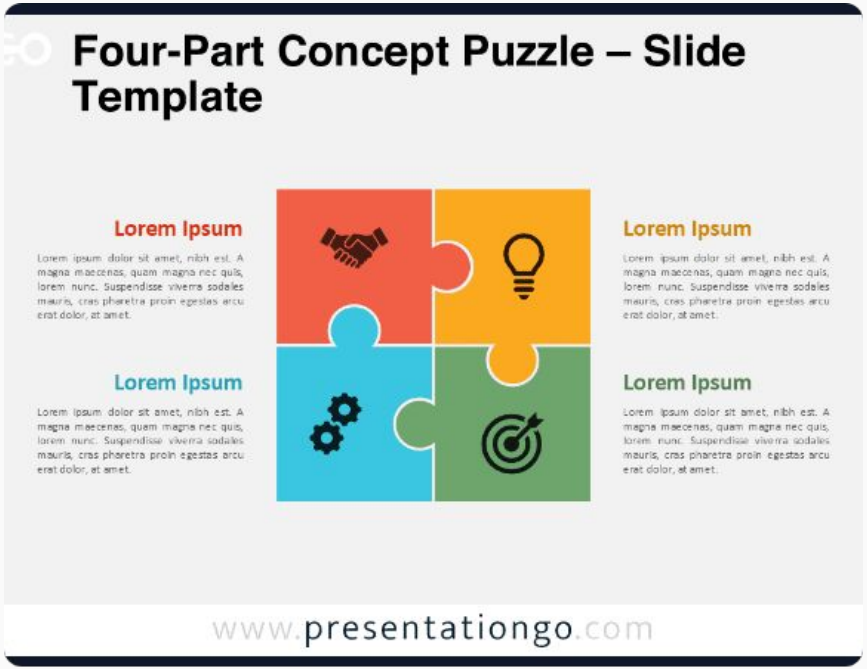
Khoảng trống Nghiên cứu

Câu hỏi nghiên cứu: Làm thế nào để phát hiện Deepfake mà không cần biết thuật toán tạo ra nó?

Quan sát quan trọng: Hầu hết quy trình Deepfake đều phải qua bước **Pha trộn (Blending)** để ghép mặt giả vào ảnh gốc.

Dấu hiệu bất biến: Luôn tồn tại sự khác biệt thống kê tại vùng biên giới pha trộn (Blending Boundary).

Research Gap: Thiếu các phương pháp tập trung vào dấu hiệu biên giới này thay vì các lỗi hình ảnh bề mặt.



Four-Part Concept Puzzle – Slide Template

Lorem Ipsum

Lorem ipsum dolor sit amet, nibh est. A magna maecenas, quam magna nec quis, lorem nunc. Suspendisse viverra sodales mauris, cras pharetra proin egestas arcu erat dolor, at amet.

Lorem Ipsum

Lorem ipsum dolor sit amet, nibh est. A magna maecenas, quam magna nec quis, lorem nunc. Suspendisse viverra sodales mauris, cras pharetra proin egestas arcu erat dolor, at amet.

Lorem Ipsum

Lorem ipsum dolor sit amet, nibh est. A magna maecenas, quam magna nec quis, lorem nunc. Suspendisse viverra sodales mauris, cras pharetra proin egestas arcu erat dolor, at amet.

Lorem Ipsum

Lorem ipsum dolor sit amet, nibh est. A magna maecenas, quam magna nec quis, lorem nunc. Suspendisse viverra sodales mauris, cras pharetra proin egestas arcu erat dolor, at amet.

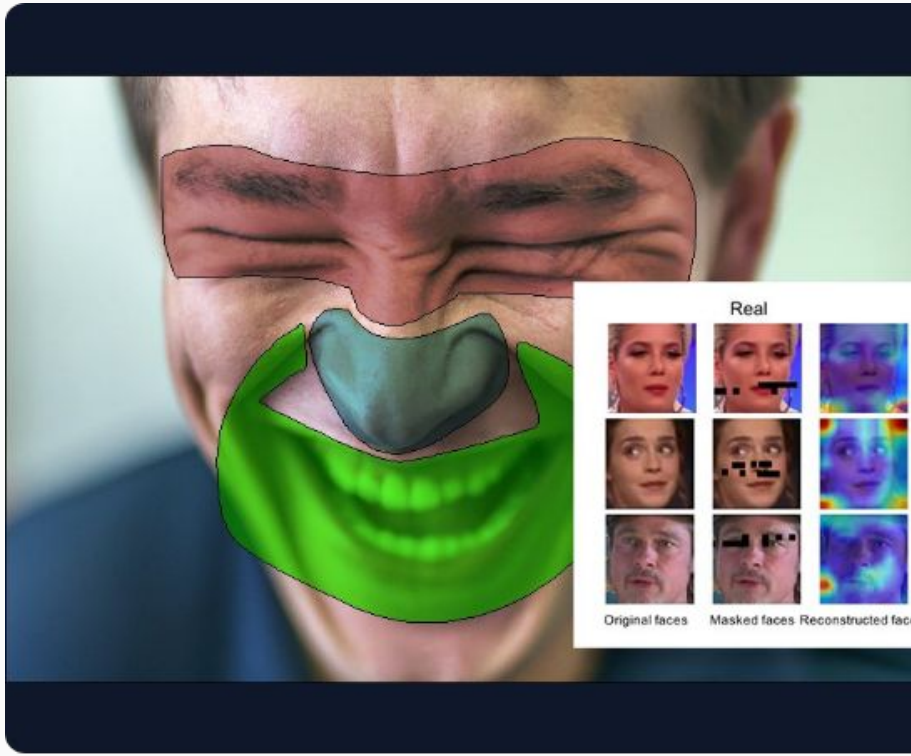
www.presentationgo.com

Giải pháp: Face X-ray

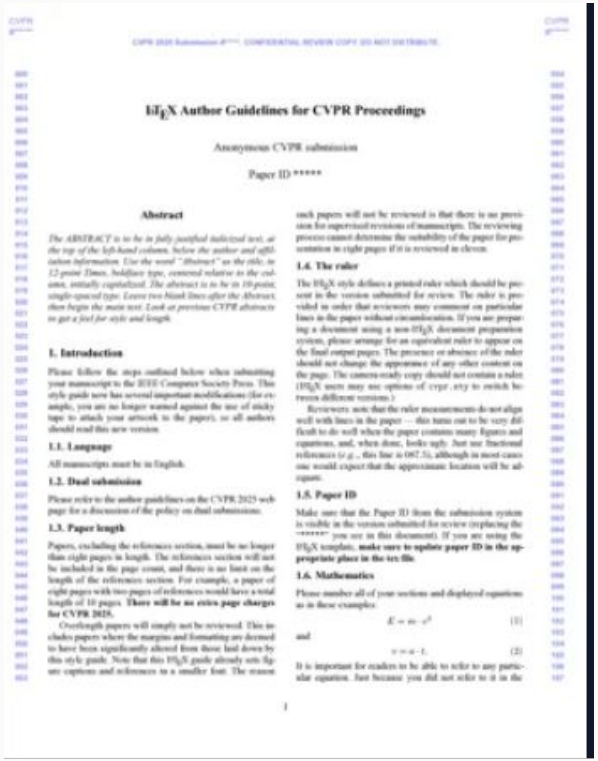
Ý tưởng cốt lõi: Không phân tích nội dung khuôn mặt (mắt, mũi, miệng), chỉ tìm sự không nhất quán tại biên giới pha trộn.

Nguyên lý hoạt động:
Input: Ảnh khuôn mặt đầu vào.
Output: Bản đồ ranh giới (Boundary Map) & Nhãn.

Ưu điểm vượt trội: Khả năng Tổng quát hóa (Generalization) cao, phát hiện được các phương pháp giả mạo chưa biết.



Cơ sở lựa chọn đề tài



LaTeX Author Guidelines for CVPR Proceedings

Anonymous CVPR submission

Paper ID *****

Abstract

1. Introduction

1.1. Language

1.2. Dual submission

1.3. Paper length

1.4. The ruler

1.5. Paper ID

1.6. Mathematics

Bài báo nền tảng: "Face X-ray for More General Face Forgery Detection"

Hội nghị: CVPR 2020 (Top-tier Conference về Computer Vision).

Dataset chuẩn: Sử dụng FaceForensics++ được cộng đồng công nhận rộng rãi.

Tính khả thi:

- Có Source code công khai (Microsoft).
- Có Demo minh họa rõ ràng.
- Kết quả vượt trội so với Baseline (XceptionNet).

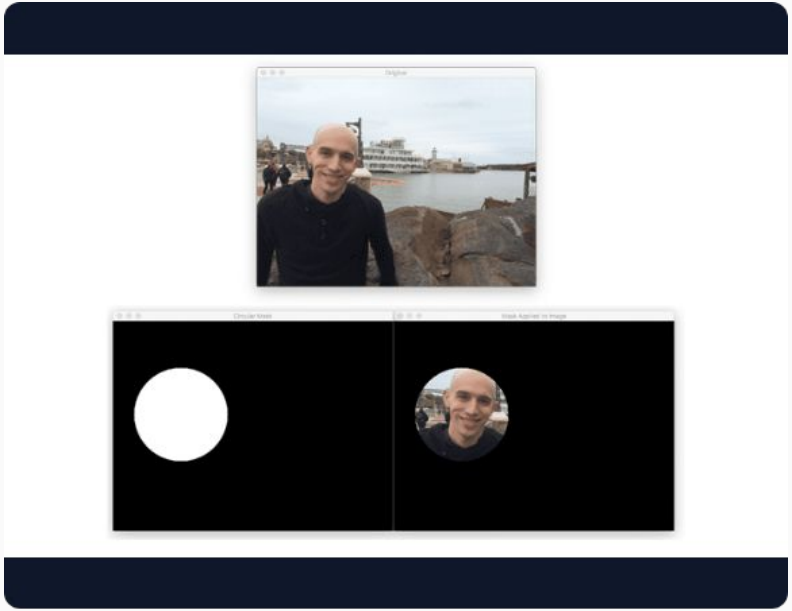
Cơ chế tự sinh dữ liệu (Self-Supervised)

Vấn đề: Rất khó có dữ liệu gán nhãn chính xác cho đường biên pha trộn thực tế.

Giải pháp: Tự cắt ghép và pha trộn ảnh trong quá trình huấn luyện (Dynamic Augmentation).

Kỹ thuật: Ghép ảnh BI vào ảnh I với Mask Gaussian mờ để tạo biên giới nhân tạo.

Mục đích: Ép mô hình học đặc trưng biên giới thay vì học nội dung ảnh.



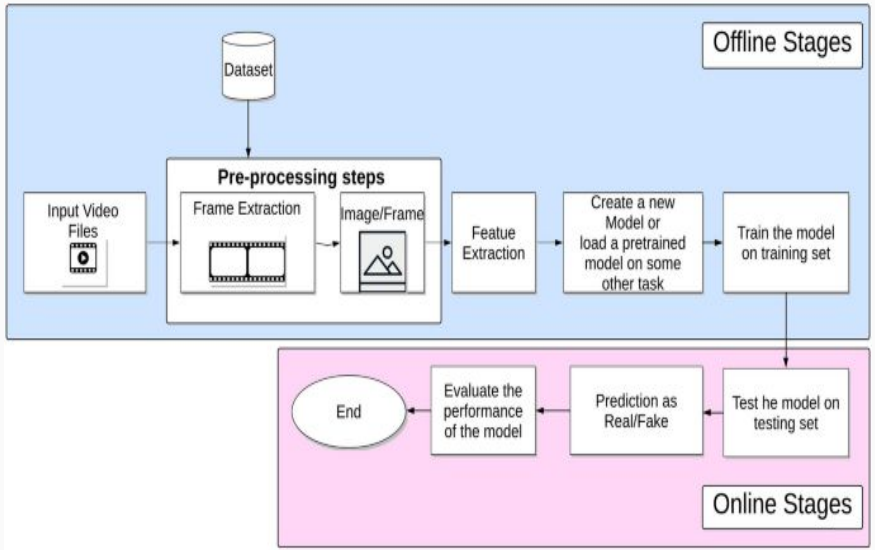
Kiến trúc hệ thống đề xuất

Bước 1: Tiền xử lý - Trích xuất và căn chỉnh khuôn mặt từ video/ảnh đầu vào.

Bước 2: Data Augmentation - Tự tạo mẫu giả mạo (Self-blended images) để huấn luyện.

Bước 3: Feature Extraction - Sử dụng mạng HRNet để trích xuất đặc trưng độ phân giải cao.

Bước 4: Dự đoán - Sinh ra Bản đồ biên (Boundary Map) và Phân loại (Softmax).



Dữ liệu và Kịch bản đánh giá

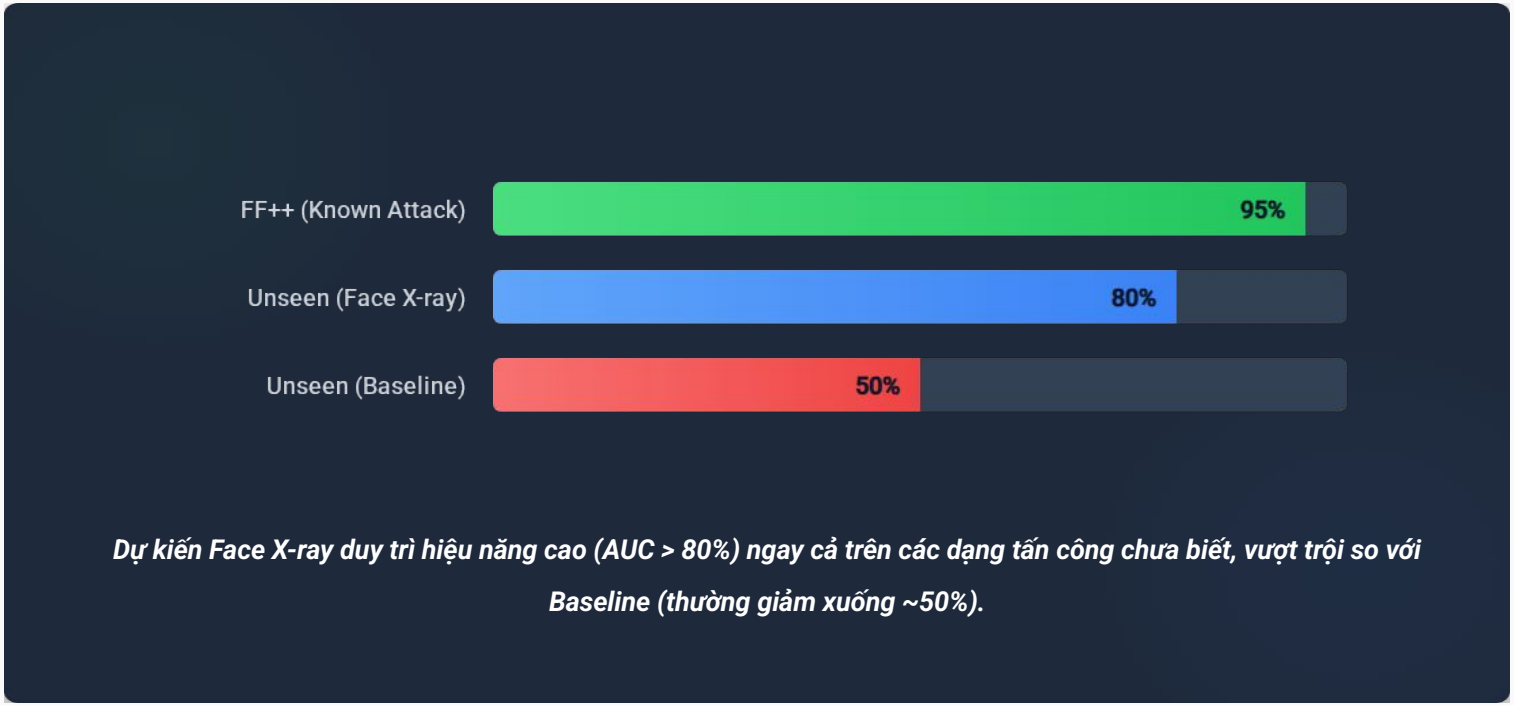
Tập Huấn Luyện (Training)

- **Dataset:** FaceForensics++ (FF++)
- **Quy mô:** 1,000 video gốc và 4,000 video giả mạo.
- **4 Phương pháp giả mạo:**
  - Deepfakes
  - Face2Face
  - FaceSwap
  - NeuralTextures
- **Chất lượng:** Cấu hình nén HQ (High Quality).

Tập Kiểm Thử (Testing)

- **Mục tiêu:** Đánh giá khả năng Tổng quát hóa (Generalization).
- **Kịch bản:** Cross-Dataset Evaluation (Train trên FF++, Test trên dataset khác).
- **Datasets chưa biết (Unseen):**
  - DeepFake Detection Challenge (DFDC)
  - Celeb-DF
- **Chỉ số đo lường:** AUC (Area Under Curve) và Accuracy.

Kết quả Kỳ vọng



Dự kiến Face X-ray duy trì hiệu năng cao (AUC > 80%) ngay cả trên các dạng tấn công chưa biết, vượt trội so với Baseline (thường giảm xuống ~50%).

Tổng kết & Hướng phát triển

Đóng góp

Đề xuất giải pháp phát hiện Deepfake bền vững, tập trung vào bản chất pha trộn hình ảnh thay vì chạy đua theo thuật toán tạo giả.

Ý nghĩa

Nâng cao khả năng an ninh thông tin, giảm sự phụ thuộc vào tri thức về các phương pháp tấn công cụ thể.

Tương lai

Mở rộng nghiên cứu sang tối ưu hóa thời gian thực (Real-time) và phân tích tính nhất quán theo thời gian trong video.

Tài liệu tham khảo

[1] Lingzhi Li, Jianmin Bao, Ting Zhang, Hao Yang, Dong Chen, Fang Wen, and Baining Guo. "Face X-ray for More General Face Forgery Detection." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5001-5010, 2020.

[2] Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Niessner. "FaceForensics++: Learning to Detect Manipulated Facial Images." In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 1-11, 2019.

[3] Yuezun Li and Siwei Lyu. "Exposing DeepFake Videos By Detecting Face Warping Artifacts." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2019.