BẢO VỆ NGƯỜI NỔI TIẾNG KHỎI VẤN ĐỀ DEEPFAKE VỚI MÔ HÌNH IDENTITY CONSISTENCY TRANSFORMER

Đặng Thị Huệ[1]

[1] Trường Đại học Công nghệ Thông tin
ĐHQG TPHCM

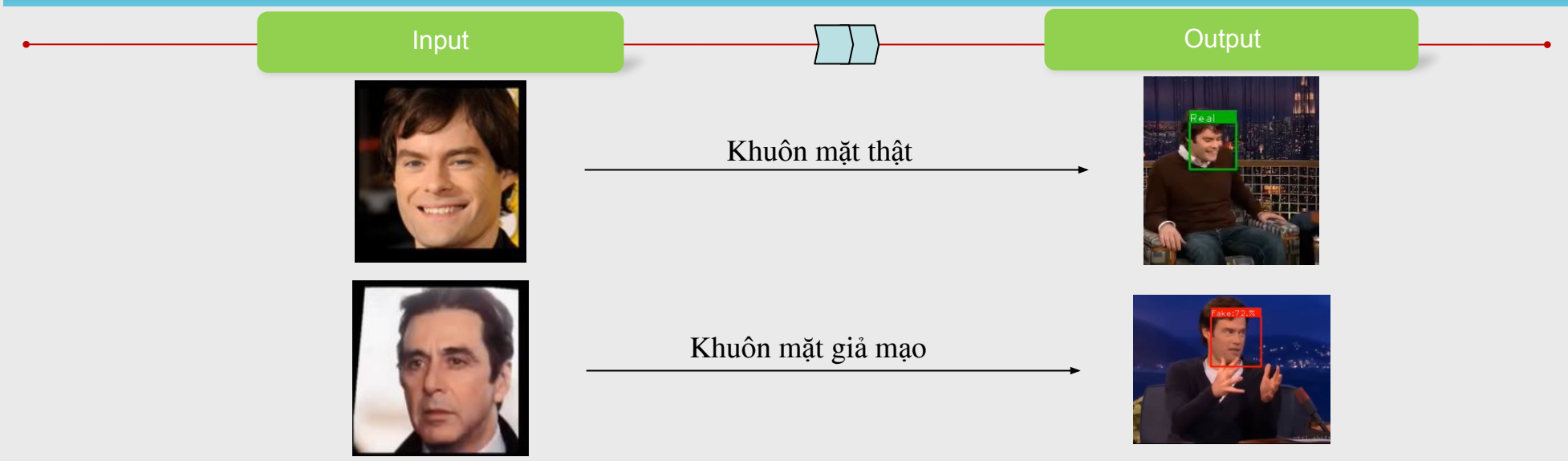
Mục tiêu

- Áp dụng mô hình Transformer để xử lý dữ liệu hình ảnh.
- Xây dựng mô hình Identity Consistency Transformer để phát hiện giả mạo khuôn mặt trên video độ phân giải thấp.
- Áp dụng mô hình Identity Consistency Transformer để cảnh báo về giả mạo khuôn mặt người nổi tiếng trên các nền tảng mạng xã hội.

Lý do chọn đề tài

- Tỷ lệ deepfake đã tăng gấp đôi từ năm 2022 - quý 1 năm 2023 ở Bắc Mỹ, tăng từ 0,2% lên 2,6% ở Hoa Kỳ.
- Các phương pháp phát hiện deepfake hiện tại chỉ hiệu quả đối với các video bị nghi ngờ tính chính xác và các video bị giảm chất lượng hình ảnh. Tuy nhiên, với các video được chỉnh sửa tinh vi, ghép nối chuyên nghiệp và có chất lượng cao, việc phát hiện giả mạo trở nên khó khăn.

Overview



Description

1. Nội dung

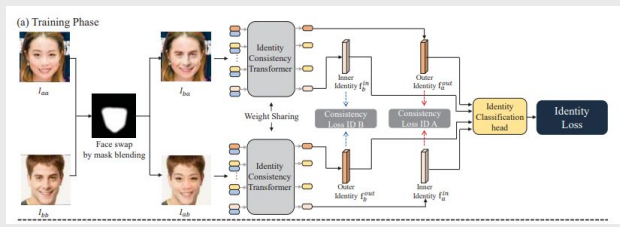
- Tập trung vào khả năng áp dụng mô hình Transformer, một mô hình học sâu dựa trên cơ chế chú ý (attention mechanism), để trích xuất thông tin danh tính đồng nhất trên dữ liệu dạng hình ảnh
- Xây dựng một mô hình phát hiện giả mạo khuôn mặt, gọi là Identity Consistency Transformer (ICT), dựa trên kiến trúc và cơ chế chú ý của mô hình Transformer.
- Đánh giá hiệu quả của mô hình ICT trong việc phát hiện deepfake đồng thời so sánh với các phương pháp phát hiện deepfake khác như: Multi-task, MesoInc4,Capsule.
- Áp dụng mô hình ICT vào thực tế.

2. Phương pháp nghiên cứu

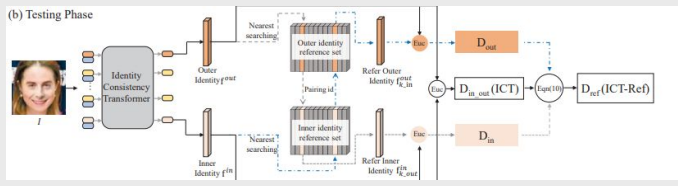
- Sử dụng dữ liệu hình ảnh khuôn mặt từ MS-Celeb và mô hình Transformer để trích xuất thông tin danh tính.
- Xây dựng mô hình Identity Consistency Transformer (ICT) để phát hiện giả mạo khuôn mặt.
- Đánh giá hiệu suất của mô hình ICT trên bộ dữ liệu FaceForensics+ và so sánh với các phương pháp phát hiện deepfake khác bằng chỉ số AUC (%).
- Phát triển tiện ích mở rộng trên Google Chrome để cảnh báo người dùng về nội dung có giả mạo khuôn mặt trên internet.

3. Kết quả dự kiến

- Mô hình ICT phát hiện và phân biệt được được các khuôn mặt thật hay giả mạo với chất lượng video đầu vào có độ phân giải thấp (480p, 360p).
- Mô hình ICT mang lại hiệu suất và độ chính xác cao hơn so với các phương pháp phát hiện deepfake đang có hiện nay.
- Tiện ích mở rộng trên trình duyệt Chrome nhận diện được giả mạo khuôn mặt trên mạng xã hội Facebook và Youtube.



Mô phỏng giai đoạn training mô hình ICT



Mô phỏng giai đoạn testing mô hình ICT

Đặng Thị Huệ – Trường Đại học Công nghệ Thông tin - ĐHQG TPHCM
TEL : 0328289909 Email : huedt.17@grad.uit.edu.vn