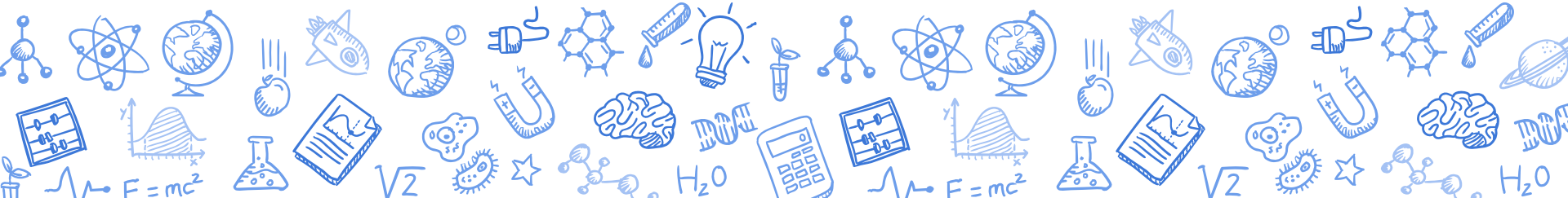
BÁO CÁO ĐỒ ÁN CUỐI KỲ

Môn học: CS519 - PHƯƠNG PHÁP LUẬN NCKH
Lớp: CS519.N11
GV: PGS.TS. Lê Đình Duy
Nhóm: AKH

Trường ĐH Công Nghệ Thông Tin, ĐHQG-HCM



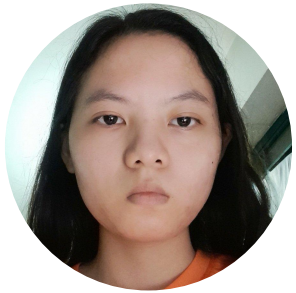
BẢO VỆ NGƯỜI NỔI TIẾNG VỚI IDENTITY CONSISTENCY TRANSFORMER

Nguyễn Quốc Khánh 20521452
Nguyễn Trần Anh Minh 20520394
Lê Nguyễn Bảo Hân 20520174

Tóm tắt

- Link Github của nhóm:
  https://github.com/nqkhanh2002/CS519.N11
- Link YouTube video:
  https://www.youtube.com/watch?v=LxhwHdXHndw

Nguyễn Quốc Khánh



Nguyễn Trần Minh Anh

Lê Nguyễn Bảo Hân

Giới thiệu

Deepfake là công nghệ sử dụng trí tuệ nhân tạo (AI) để lấy hình ảnh, giọng nói của một người ghép vào video của người khác.

Mục tiêu

1. Nghiên cứu, khảo sát các hướng tiếp cận hiện có cho bài toán nhận dạng Deepfake

2. Xây dựng mô hình phát hiện giả mạo khuôn mặt có tên là Identity Consistency Transformer (ICT) dựa trên thông tin ngữ nghĩa cấp cao

3. Đánh giá và cải tiến mô hình ICT và xây dựng chương trình demo trực quan hóa nghiên cứu

Nội dung và Phương pháp

Nội dung 1: Nghiên cứu, khảo sát các hướng tiếp cận hiện có cho bài toán nhận dạng Deepfake

- Khảo sát các kết quả nghiên cứu đã có về chủ đề Deepfake, phát hiện Deepfake, chống lại phát hiện Deepfake.
- Kỹ thuật hoán đổi danh tính (identity swap) xếp hạng cao nhất về mức độ phổ biến và nguy hiểm trong 4 loại Deepfake.
- Hầu hết các phương pháp hiện có nhằm mục đích phân biệt hình ảnh giả bằng cách khai thác kết cấu cấp thấp và tìm kiếm các tạo phẩm tạo tác bên dưới.

Nội dung và Phương pháp

Nội dung 2: Xây dựng mô hình phát hiện giả mạo khuôn mặt có tên là Identity Consistency Transformer (ICT) dựa trên thông tin ngữ nghĩa cấp cao

- Sử dụng kiến trúc Transformer để xây dựng mô hình ICT - đây là một loại mạng thần kinh thường được sử dụng trong các tác vụ xử lý ngôn ngữ tự nhiên.
- Mô hình ICT được thiết kế để tập trung vào việc duy trì các đặc điểm nhận dạng của người nổi tiếng, thay vì chỉ phát hiện các điểm bất thường về hình ảnh trong video.
- Mô hình ICT sẽ từ đó tự sinh ra không gian đặc trưng trích xuất được của người nổi tiếng, từ đó đối chiếu sự khác biệt và đưa ra nhận định.

Nội dung và Phương pháp

Nội dung 3: Đề xuất cải tiến thuật toán Transformer để kiểm chứng hiệu quả với module bài toán phân loại khuôn mặt

- Thử nghiệm các thước đo tính nhất quán nhận dạng khác nhau
- Nghiên cứu thuật toán Transformer truyền thống và huấn luyện mô hình ICT trên tập dữ liệu MS-Celeb-1M
- Đánh giá và so sánh hiệu quả của mô hình ICT
- Nghiên cứu các kỹ thuật tăng cường dữ liệu
- Xây dựng chương trình ứng dụng minh họa với Python

Kết quả dự kiến

- Tạo ra được mô hình mới tập trung vào ngữ nghĩa cấp cao, tận dụng thông tin danh tính bổ sung
- Báo cáo phương pháp và kỹ thuật của mô hình đã phát triển, kết quả thực nghiệm, đánh giá
- Cải thiện, khuyến khích nhiều kết quả nghiên cứu hơn
- Một chương trình demo để trực quan hóa nghiên cứu

Tài liệu tham khảo

[1] Ali Raza, Kashif Munir and Mubarak Almutairi. “A Novel Deep Learning Approach for Deepfake Image Detection”.

[2]. Maryam Taeb and Hongmei Chi**.** “Comparison of Deepfake Detection Techniques through Deep Learning”.

[3]. “Countering Malicious DeepFakes: Survey, Battleground, and Horizon”.

[4] Jianmin Bao, Dong Chen, Fang Wen, Houqiang Li, and Gang Hua. Towards open-set identity preserving face synthesis, 2018. 1, 2.

[5]. Darius Afchar, Vincent Nozick, Junichi Yamagishi, and Isao Echizen. Mesonet: a compact facial video forgery detection network, 2018. 1, 2, 6.