PHÁT HIỆN VÀ PHÂN LOẠI NỘI DUNG GIẢ MẠO KHUÔN MẶT SỬ DỤNG MÔ HÌNH MẠNG NEURAL ĐỒ THỊ (GRAPH NEURAL NETWORK)

Học viên: Phạm Huy Hùng - 240104034
GVHD: PGS. TS Lê Đình Duy

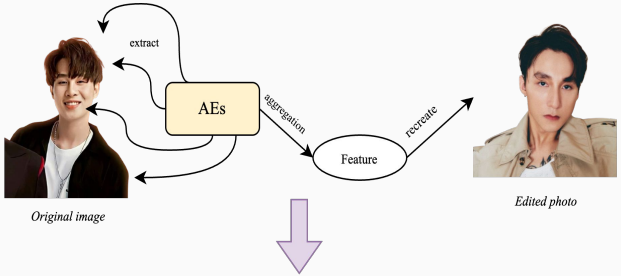
UIT-HCM

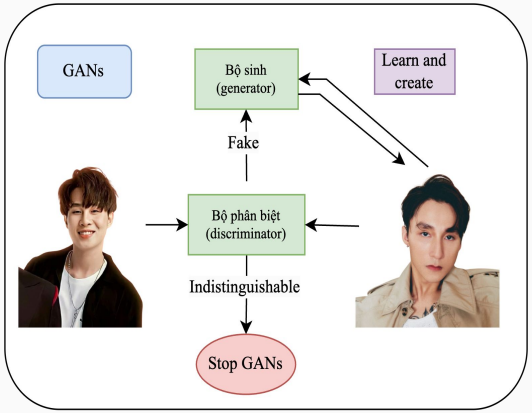
Tóm tắt

- Họ và tên: **Phạm Huy Hùng**
- Lớp: **CS2205.NOV2024**
- Link Github của nhóm:
  https://github.com/hpham16/CS2205.CH2023-01-NOV2024
- Link YouTube video:
  https://www.youtube.com/watch?v=TdouTgy86lM

Giới thiệu

- Ngày nay, **Deepfake** [1] đã được coi là **một mối đe dọa toàn cầu** đối với an ninh thông tin.
- Các hình ảnh, video giả mạo được tạo ra bởi các mô hình học sâu như **GANs và autoencoders** có độ chân thực cao khó phân biệt [2, 3].
- **CNN truyền thống** gặp hạn chế trong việc **tổng quát hoá**, không giải thích được về **đặc trưng** của từng nội dung, **khó kiểm soát & cải tiến** trong thực tiễn.[4]
- Ứng dụng lý thuyết vào việc nhận diện deepfake.
- Sử dụng **mô hình GNN**. [5]

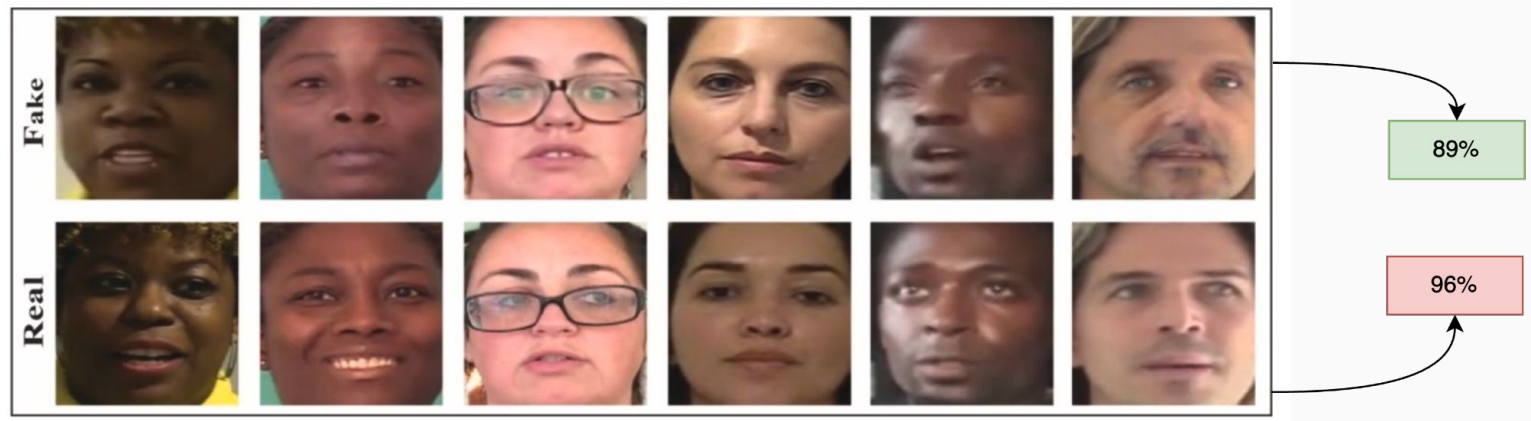




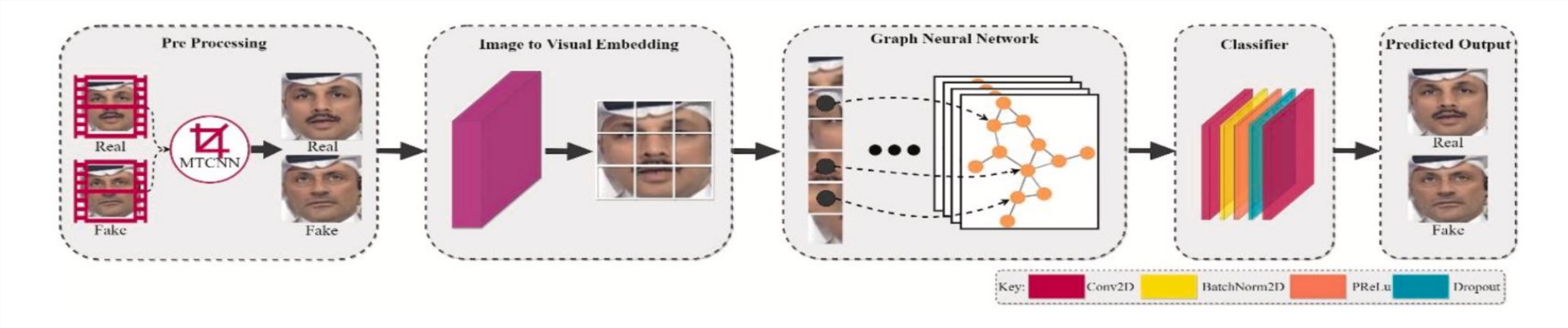
Giới thiệu

Input: Dữ liệu Video hoặc hình ảnh chứa khuôn mặt cần nhận diện.

Output: Nhãn phân loại, dự đoán xem khuôn mặt thuộc video/hình ảnh là thật hay là deepfake, kèm theo xác suất dự đoán. Như hình minh hoạ.



Giới thiệu



*(a) Hình ảnh được trích xuất từ công trình của tác giả **Khalid & Fatimal** trong nghiên cứu về GNN [5].*

Mục tiêu

- Phát triển và hiệu chỉnh mô hình GNN nhằm phát hiện và phân loại nội dung deepfake khuôn mặt.
- Thực nghiệm đánh giá hiệu năng của GNN, qua việc so sánh với các phương pháp phát hiện deepfake truyền thống dựa trên CNN với nhiều tập dữ liệu khác nhau.
- Xây dựng hệ thống prototype thực nghiệm minh hoạ.

Nội dung và phương pháp

Về Nội Dung

- Nghiên cứu tổng quan về **deepfake**, các tác động tiêu cực.
- Phân tích các những hạn chế của các **mô hình CNN.**
- Nghiên cứu về **Graph Neural Network (GNN).**
- Nghiên cứu phương pháp sử dụng **mô hình GNN** nhằm phát hiện nội dung deepfake.
- Đọc hiểu và xử lý các bộ dữ liệu deepfake từ các nguồn công khai như **FaceForensics++, DFDC, Celeb-DF và World Leaders Dataset (WLRD)**.
- Tiến hành huấn luyện **mô hình GNN** và **mô hình tiền nhiệm là CNN.**
- **So sánh đánh giá** các phương pháp đã sử dụng.
- Xây dựng ứng dụng minh hoạ.

Nội dung và phương pháp

Về Phương Pháp

- Tìm hiểu lý thuyết về **deepfake**, các tác động tiêu cực, phân tích các hạn chế của **mô hình CNN** trong việc nhận diện nội dung **deepfake** trên các diễn đàn công nghệ.
- Nghiên cứu về **Graph Neural Network (GNN)** và cách triển khai và tối ưu trên các hội nghị, tạp chí đã công bố [5, 6, 7].
- **Tiền xử lý dữ liệu** deepfake từ các nguồn công khai (FaceForensics++, DFDC, Celeb-DF, WLRD) bằng cách sử dụng **MTCNN.**
- **Huấn luyện** mô hình **GNN** và **CNN** sau đó **so sánh đánh giá** nhiều lần qua các chỉ số như Accuracy, AUC, F1-score.
- Phát triển ứng dụng prototype trên nền tảng mạng xã hội, để **kiểm chứng.**

Kết quả dự kiến

- Báo cáo chi tiết về hiệu năng của mô hình **GNN và CNN.**
- Đưa ra kết quả thực nghiệm chạy trên các bộ dữ liệu cùng với các chỉ số đánh giá.
- Xây dựng ứng dụng prototype trên nền tảng mạng xã hội cho phép người dùng đăng tải hình ảnh và tham gia các cuộc gọi video.
- Chạy thử nghiệm trên các nền tảng cần độ chính xác cao như **ngân hàng** trong việc **nhận diện sinh trắc học.**

Tài liệu tham khảo

[1]. Chadha, Anupama, et al. “Deepfake: an Overview.” In Proceedings of Second International Conference on Computing, Communications, and Cyber-Security: IC4S 2020, Springer Singapore, 2021.

[2]. Tov, Omer, et al. “Designing an Encoder for StyleGAN Image Manipulation.” ACM Transactions on Graphics (TOG), vol. 40, no. 4, 2021, pp. 1–14.

[3]. Bank, Dor, Noam Koenigstein, and Raja Giryes. “Autoencoders.” In Machine Learning for Data Science Handbook: Data Mining and Knowledge Discovery Handbook, 353–374, 2023.

[4]. Amerini, Irene, et al. “Deepfake Video Detection Through Optical Flow Based CNN.” In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019.

[5] Khalid, Fatima, et al. “DFGNN: An Interpretable and Generalized Graph Neural Network for Deepfakes Detection.” Expert Systems with Applications, vol. 222, 2023, Art. no. 119843.

[6] Xiang, J., and G. Zhu. “Joint Face Detection and Facial Expression Recognition with MTCNN.” In 2017 4th International Conference on Information Science and Control Engineering (ICISCE), 424–427. IEEE, 2017.

[7] Rafique, Rimsha, et al. “Deepfake Detection Using Error Level Analysis and Deep Learning.” In 2021 4th International Conference on Computing & Information Sciences (ICCIS), IEEE, 2021.