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

Đặng Thị Huệ - 220104017

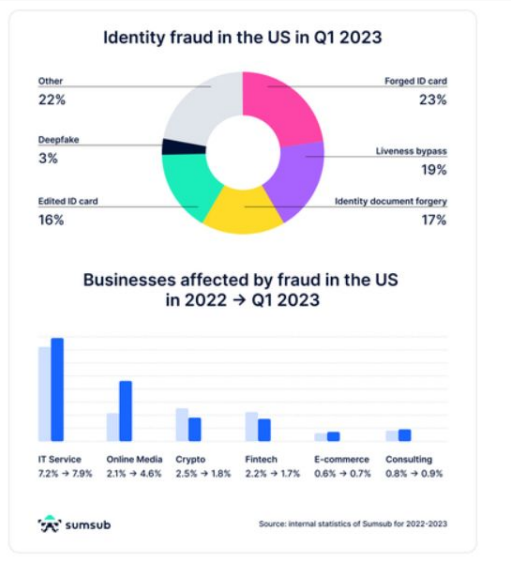
Tóm tắt

- Lớp: CS2205.RM
- Link Github:
- Link YouTube video:

Đặng Thị Huệ

Giới thiệu

- Ở Bắc Mỹ, tỷ lệ deepfake đã tăng gấp đôi từ năm 2022 - quý 1 năm 2023. Tỷ lệ này tăng từ 0,2% lên 2,6% ở Hoa Kỳ và từ 0,1% lên 4,6% ở Canada, tương ứng (theo trang Businesswire).



Giới thiệu

- Kỹ thuật phát hiện giả mạo khuôn mặt mới là Identity Consistency Transformer (ICT) để ngăn chặn việc sử dụng hình ảnh giả mạo.
- Input: Video hoặc hình ảnh chứa khuôn mặt của một người.
- Output:

Khuôn mặt giả mạo đã qua chỉnh sửa hình ảnh

Khuôn mặt thật

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.

Nội dung

- Nghiên cứu sẽ tập trung vào khả năng áp dụng mô hình Transformer 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)
- Đánh giá hiệu quả của mô hình ICT so với các phương pháp khác như: Multi-task, MesoInc4,Capsule.
- Áp dụng mô hình ICT vào thực tế.

Phương pháp

- 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.

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.

Tài liệu tham khảo

[1]. Xiaoyi Dong, Jianmin Bao, Dongdong Chen, Ting Zhang, Weiming Zhang, Nenghai Yu, Dong Chen, Fang Wen, Baining Guo: Protecting Celebrities from DeepFake with Identity Consistency Transformer. CVPR 2022: 9458-9468

[2] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin: Attention Is All You Need. CoRR abs/1706.03762 (2017)

[3] Chen Cao, Yanlin Weng, Shun Zhou, Yiying Tong, and Kun Zhou. Facewarehouse: A 3d facial expression database for visual computing. IEEE Transactions on Visualization and Computer Graphics, 20(3):413–425, 2013

[4] Huy H. Nguyen, Fuming Fang, Junichi Yamagishi, Isao Echizen: Multi-task Learning For Detecting and Segmenting Manipulated Facial Images and Videos. CoRR abs/1906.06876 (2019)

[5] Darius Afchar, Vincent Nozick, Junichi Yamagishi, Isao Echizen: MesoNet: a Compact Facial Video Forgery Detection Network. WIFS 2018: 1-7

[6] Huy H. Nguyen, Junichi Yamagishi, Isao Echizen: Use of a Capsule Network to Detect Fake Images and Videos. CoRR abs/1910.12467 (2019)