PHÁT HIỆN GIẢ MẠO GƯƠNG MẶT BẰNG CÁCH HỌC MỐI QUAN HỆ GIỮA CÁC ĐƠN VỊ HÀNH ĐỘNG CƠ BẢN CỦA CƠ MẶT

DETECTING FACE FORGERY BY LEARNING THE RELATIONSHIPS BETWEEN BASIC FACIAL ACTION UNITS

GVHD: PGS. TS Lê Đình Duy
Học viên: Đinh Hoàng Thùy Linh - 240101054

Tóm tắt

- Lớp: **CS2205.FEB2025**
- Link Github của nhóm:

https://github.com/thuyLinhUIT/CS2205.FEB2025

- Link YouTube video:

https://www.youtube.com/watch?v=TNkfLoGECo8

- Họ và tên:

Đinh Hoàng Thùy Linh - 240101054

Giới thiệu

- Sự gia tăng cả về số lượng lẫn mức độ tinh vi của các vụ lừa đảo qua video, đặc biệt với sự hỗ trợ của công nghệ sinh ảnh như **GAN**, các kỹ thuật giả mạo khuôn mặt ngày càng trở nên khó phát hiện.

→ Nghiên cứu các phương pháp **phát hiện giả mạo đối kháng**

Vấn đề: Tổng quát hóa đối với các phương pháp giả mạo chưa từng thấy vẫn chưa được đảm bảo. Cụ thể, các nghiên cứu này có thể được chia thành hai hướng chính:

- 1. Biến đổi dữ liệu (data modification)
- 2. Tích hợp nhiệm vụ phụ (auxiliary task integrating)

Các mối quan hệ giữa các đơn vị khuôn mặt (face units) lại ít được khai thác, điều này cản trở việc cải thiện hơn nữa khả năng tổng quát hóa của mô hình.

Mục tiêu

- Tạo được mô-đun Action Units Relation Transformer **ART**: Học quan hệ vùng AU, nâng cao phát hiện giả mạo.
- Tạo được nhiệm vụ phụ Tampered AU Prediction TAP: Tăng khả năng phát hiện giả mạo cục bộ.
- Chứng minh hiệu quả mô hình: Mô hình đạt kết quả tốt trên cả đánh giá nội bộ và chéo tập.

Nội dung và Phương pháp

- **Nội dung**: Khung mô hình được đề xuất có tên là Action Units Relation Learning, bao gồm hai thành phần chính:

1. Action Units Relation Transformer (ART)

Mục tiêu: Học mối quan hệ giữa các vùng liên quan đến các đơn vị hành động (AU) trên khuôn mặt để hỗ trợ phát hiện giả mạo.

2. Tampered AU Prediction (TAP)

Mục tiêu: Tăng khả năng phát hiện giả mạo cục bộ bằng cách tạo ra các vùng bị làm giả và huấn luyện mô hình nhận diện chúng.

Nội dung và Phương pháp

- Phương pháp

Tập dữ liệu huấn luyện:
Sử dụng **FaceForensics++ (FF++)**, gồm 1000 video gốc từ YouTube và các video giả được tạo bằng 4 kỹ thuật: **Deepfakes (DF), Face2Face (F2F), FaceSwap (FS), NeuralTextures (NT)**.

Tiền xử lý:
Phát hiện khuôn mặt bằng **RetinaFace**.
Phát hiện điểm đặc trưng bằng **Dlib**.
Cắt và chuẩn hóa khuôn mặt về kích thước **224×224**.

Huấn luyện mô hình:
Sử dụng **Xception** (đến block 11) làm backbone.
Trọng số khởi tạo từ **ImageNet**.

Đánh giá tổng quát:
Thử nghiệm thêm trên các tập dữ liệu giả mạo khác:
Celeb-DF, DFD, DFDC, DFDCP, Wild-Deepfake, theo cách chia chính thức.

Kết quả dự kiến

Với ảnh đầu vào thật:

- Trả về mặt nạ trắng (không phát hiện chỉnh sửa)
- Tỷ lệ dương tính giả cực thấp (dưới 10% trong thử nghiệm)

Với ảnh giả mạo: Mặt nạ đầu ra 24×24 khớp chính xác với:

- Vị trí vùng bị chỉnh sửa
- Hình dạng và kích thước thao tác

Ưu điểm công nghệ:

Thu thập đủ chi tiết khuôn mặt

Tối ưu tài nguyên tính toán

Mặt nạ 24×24 vẫn đạt độ chính xác cao nhờ:

- Cơ chế tập trung theo vùng (region-aware)
- Mạng tích chập sâu phân giải đa tầng

Khả năng đặc biệt:

Phân biệt rõ thao tác cục bộ (mắt/mũi/miệng) vs toàn bộ khuôn mặt

Nhạy với các thay đổi dưới 5% diện tích khuôn mặt

Thời gian xử lý trung bình 0.2s/ảnh trên GPU thế hệ mới

Tài liệu tham khảo

[*]Weiming Bai, Yufan Liu, Zhipeng Zhang, Bing Li, Weiming Hu. AU-Net: Learning Relations Between Action Units for Face Forgery Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 24709–24719, 2023.

[1] Contributing data to deepfake detection research. https : / / ai . googleblog . com / 2019 / 09 / contributing- data- to- deepfake detection.html. Accessed 2022-11-10.

[2] Deepfakes. https://github.com/deepfakes/ faceswap. Accessed 2022-11-10. 2, 6

[3] Faceswap. https : / / github . com / MarekKowalski/FaceSwap/. Accessed 2022-11 10. 2, 6

[4] Franc ¸ois Chollet. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE con ference on computer vision and pattern recognition, pages 1251–1258, 2017.

[5] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition, pages 248–255. Ieee, 2009.

[6] Jiankang Deng, Jia Guo, Evangelos Ververas, Irene Kot sia, and Stefanos Zafeiriou. Retinaface: Single-shot multi level face localisation in the wild. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 5203–5212, 2020.

[7] Brian Dolhansky, Joanna Bitton, Ben Pflaum, Jikuo Lu, Russ Howes, Menglin Wang, and Cristian Canton Ferrer. The deepfake detection challenge (dfdc) dataset. arXiv preprint arXiv:2006.07397, 2020.

[8] Brian Dolhansky, Russ Howes, Ben Pflaum, Nicole Baram, and Cristian Canton Ferrer. The deepfake de tection challenge (dfdc) preview dataset. arXiv preprint arXiv:1910.08854, 2019.

Tài liệu tham khảo

[9] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Syl vain Gelly, et al. An image is worth 16x16 words: Trans formers for image recognition at scale. arXiv preprint arXiv:2010.11929, 2020.

[10] Paul Ekman and Wallace V Friesen. Facial action coding system. Environmental Psychology & Nonverbal Behavior, 1978.

[11]Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. Commu nications of the ACM, 63(11):139–144, 2020.

[12] Yuezun Li, Xin Yang, Pu Sun, Honggang Qi, and Siwei Lyu. Celeb-df: A large-scale challenging dataset for deep fake forensics. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 3207 3216, 2020.

[13] Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Chris tian Riess, Justus Thies, and Matthias Nießner. Faceforen sics++: Learning to detect manipulated facial images. In Proceedings of the IEEE/CVF international conference on computer vision, pages 1–11, 2019.

[14] Justus Thies, Michael Zollh¨ofer, and Matthias Nießner. De ferred neural rendering: Image synthesis using neural tex tures. ACM Transactions on Graphics (TOG), 38(4):1–12, 2019.

[15] Justus Thies, Michael Zollhofer, Marc Stamminger, Chris tian Theobalt, and Matthias Nießner. Face2face: Real-time face capture and reenactment of rgb videos. In Proceed ings of the IEEE conference on computer vision and pattern recognition, pages 2387–2395, 2016.

[16] Sheng-Yu Wang, Oliver Wang, Richard Zhang, Andrew Owens, and Alexei A Efros. Cnn-generated images are surprisingly easy to spot... for now. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 8695–8704, 2020.

[17] Tianchen Zhao, Xiang Xu, Mingze Xu, Hui Ding, Yuanjun Xiong, and Wei Xia. Learning self-consistency for deepfake detection. In Proceedings of the IEEE/CVF international conference on computer vision, pages 15023–15033, 2021

[