PHÁT HIỆN HÌNH ẢNH KHUÔN MẶT NGƯỜI DEEPFAKE SỬ DỤNG MẠNG NƠ - RON TÍCH CHẬP

Bùi Lê Bảo Trân

Trường ĐH Công Nghệ Thông Tin
- ĐHQG

What ?

Chúng tôi giới thiệu mô hình nhận diện hình ảnh khuôn mặt người deepfake, bao gồm:

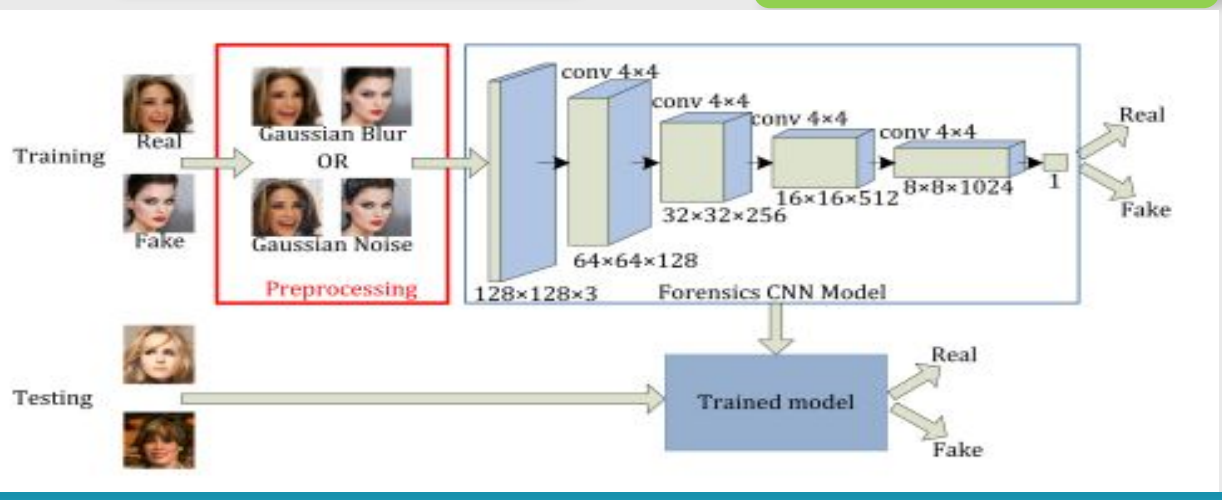
- Xây dựng mô hình Mạng nơ-ron tích chập (CNN) để phân loại hình ảnh "real" hay "deepfake"
- Xây dựng tập dữ liệu khuôn mặt lớn bằng cách thu thập dữ liệu có sẵn và biến đổi chúng để mở rộng.
- Đánh giá, so sánh với phương pháp phát hiện hình ảnh deepfake khác

Why ?

- Hình ảnh deepfake khuôn mặt người là những hình ảnh giả mạo có độ chân thực cao
- Những hình ảnh deepfake này có thể được sử dụng để đưa thông tin sai lệch, giả mạo, vi phạm quyền riêng tư và có khả năng gây hại cho các cá nhân, tổ chức
- Ngăn ngừa các tác động tiêu cực, bảo vệ quyền riêng tư, đảm bảo thông tin đáng tin cậy và tạo môi trường trực tuyến an toàn

Overview

Dữ liệu hình ảnh → Mô hình CNN → Phân loại hình ảnh: "deepfake" hay "real"

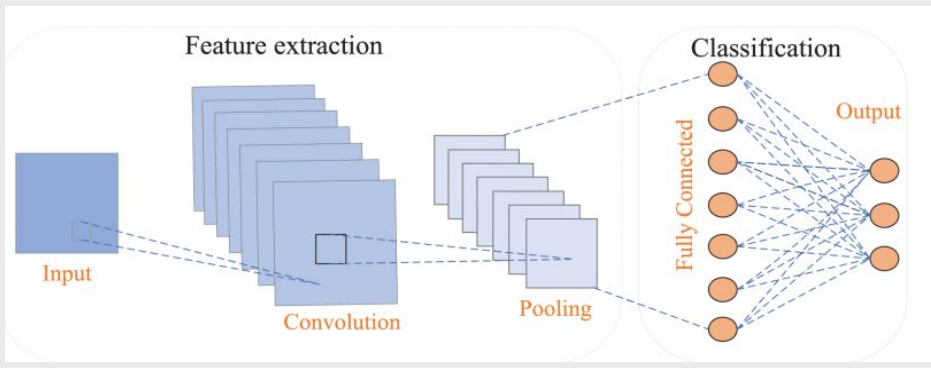


Description

1. Chuẩn bị dữ liệu

- Thu thập dữ liệu từ nhiều nguồn
- Thực hiện các phép biến đổi dữ liệu để mở rộng tập dữ liệu huấn luyện và tránh overfitting
- Chia tập dữ liệu thành tập huấn luyện (80%) và tập kiểm tra (20%). Tập huấn luyện được sử dụng để huấn luyện mô hình, tập kiểm tra để đánh giá hiệu suất

2. Xây dựng mô hình CNN



Kiến trúc cơ bản của CNN.

3. Kết quả mong đợi

2a. Huấn luyện mô hình

- Tiến hành huấn luyện mô hình trên tập dữ liệu huấn luyện.
- Huấn luyện mô hình trong nhiều epoch (vòng lặp) cho đến khi hiệu suất của mô hình không cải thiện thêm trên tập kiểm tra.

2b. Đánh giá mô hình

- Đánh giá mô hình trên tập kiểm tra bằng cách tính toán các độ đo như độ chính xác, độ nhạy và độ đặc hiệu.
- So sánh kết quả với một mô hình đã được công nhận là hiệu quả có thể xác định xem mô hình có đạt được hiệu suất tương đương hay không
- Nếu mô hình chưa đạt hiệu suất mong đợi thì tiến hành tinh chỉnh và cải thiện

NII

Bùi Lê Bảo Trân – Trường Đại học Công nghệ thông tin - Đại học Quốc Gia Thành phố Hồ Chí Minh
TEL : 0981010925 Email : tranblb.17@grad.uit.edu.vn