TRƯỜNG ĐẠI HỌC ĐẠI NAM
KHOA CÔNG NGHỆ THÔNG TIN

Hệ Thống Nhận Diện Biểu Cảm Khuôn Mặt Kết Hợp Phản Hồi Âm Thanh

Vũ Đức Anh, Nguyễn Quang Hiệp, Nguyễn Xuân Thuận, Lê Đức Mạnh
Giảng viên hướng dẫn: Lê Trung Hiếu, Nguyễn Văn Nhân
Dainam University, Hanoi, Vietnam

https://github.com/Snookari/facial-expression-recognition

Giới thiệu

Mục tiêu:
- Nhận diện biểu cảm khuôn mặt trong thời gian thực.
- Phát phản hồi âm thanh thích ứng dựa trên cảm xúc.

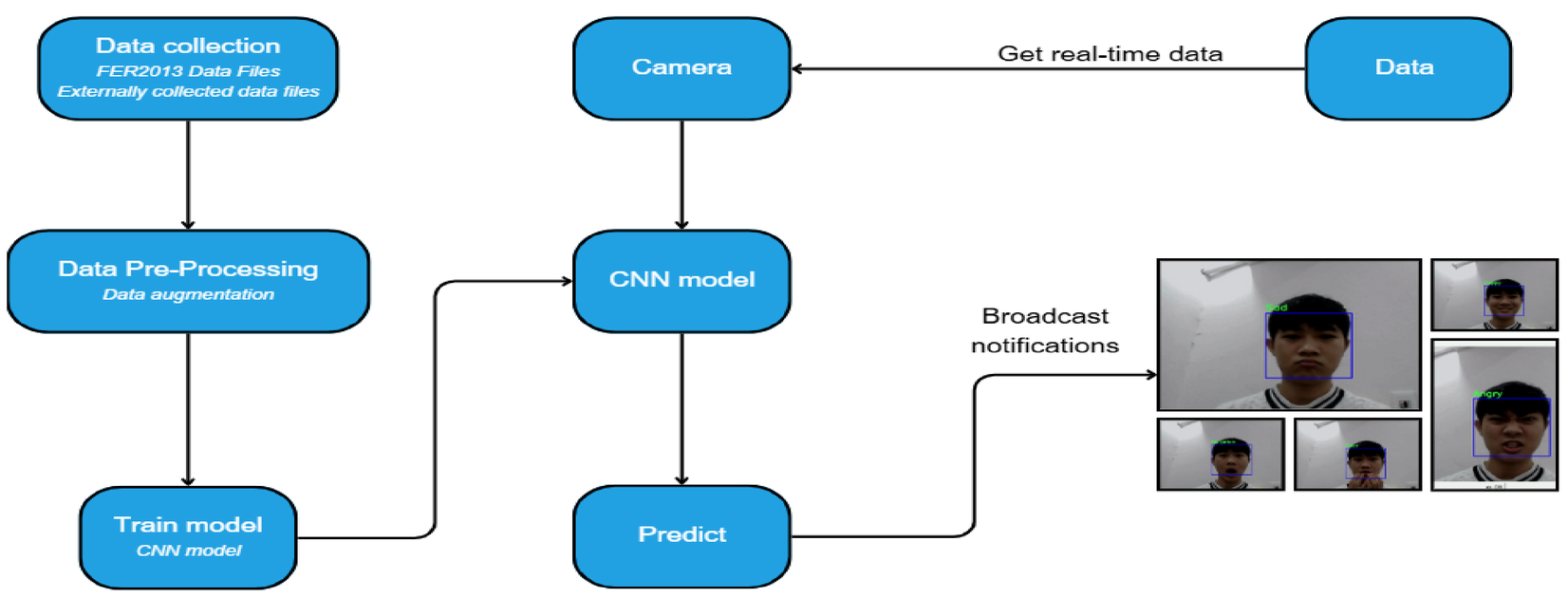
Thành phần hệ thống:
- **Dữ liệu:** Sử dụng **FER2013** và dữ liệu thu thập từ bên ngoài.
- **Mô hình:** Áp dụng **CNN** để nhận diện cảm xúc.
- **Phản hồi âm thanh:** Chuyển đổi tin nhắn văn bản thành giọng nói bằng **Text-to-Speech (TTS)**.

Quy trình hoạt động:
- **Nhận diện khuôn mặt:** Trích xuất khuôn mặt từ luồng video hoặc ảnh.
- **Dự đoán cảm xúc:** Phân loại cảm xúc (Vui, Buồn, Giận dữ, Ngạc nhiên, Trung tính,...) bằng CNN.
- **Phát phản hồi âm thanh:** Tạo giọng nói thích ứng với cảm xúc nhận diện được.

Ứng dụng:
- **Trợ lý ảo thông minh** có thể phản hồi theo cảm xúc người dùng.
- **Theo dõi sức khỏe tâm lý** và hỗ trợ giao tiếp cho người gặp vấn đề về tâm lý.
- **Hệ thống giao tiếp thông minh** trong các thiết bị và nền tảng số.

Lợi ích:
- Tăng cường tương tác người-máy.
- Cải thiện khả năng tiếp cận và giao tiếp tự nhiên.
- Thiết kế linh hoạt, có thể mở rộng và triển khai thực tế.

Sơ đồ hệ thống



Phương pháp sử dụng

1. Thành phần hệ thống

•Thu thập và tiền xử lý dữ liệu:
- Nguồn dữ liệu: **FER2013** và tập dữ liệu thu thập từ bên ngoài.
- Tiền xử lý ảnh: chuẩn hóa, cân bằng dữ liệu, tăng cường dữ liệu (xoay, lật, thay đổi độ sáng).

•Huấn luyện mô hình nhận diện cảm xúc:
- Sử dụng **CNN** để trích xuất đặc trưng và phân loại cảm xúc.
- Tối ưu hóa mô hình cho thời gian thực.
- Đánh giá mô hình bằng độ chính xác (**accuracy**), tốc độ xử lý (**FPS**) và thời gian suy luận (**inference time**).

•Phát thông báo âm thanh thích ứng:
- Xác định phản hồi dựa trên cảm xúc nhận diện được.
- Sử dụng công nghệ **Text-to-Speech (TTS)** để chuyển văn bản thành giọng nói.
- Tùy chỉnh giọng nói và nội dung phản hồi theo ngữ cảnh.

2. Quy trình hoạt động của hệ thống

1.Thu thập dữ liệu đầu vào:
1. Nhận diện khuôn mặt từ luồng video hoặc ảnh.
2. Tiền xử lý ảnh để chuẩn bị cho mô hình CNN.

2.Nhận diện cảm xúc:
1. Dự đoán cảm xúc từ ảnh khuôn mặt.
2. Phân loại cảm xúc (Vui, Buồn, Giận dữ, Ngạc nhiên, Trung tính,...)

3.Phát phản hồi âm thanh:
1. Chọn nội dung phản hồi dựa trên cảm xúc nhận diện được.
2. Sử dụng **TTS** để tạo phản hồi giọng nói phù hợp.

3. Ứng dụng tiềm năng

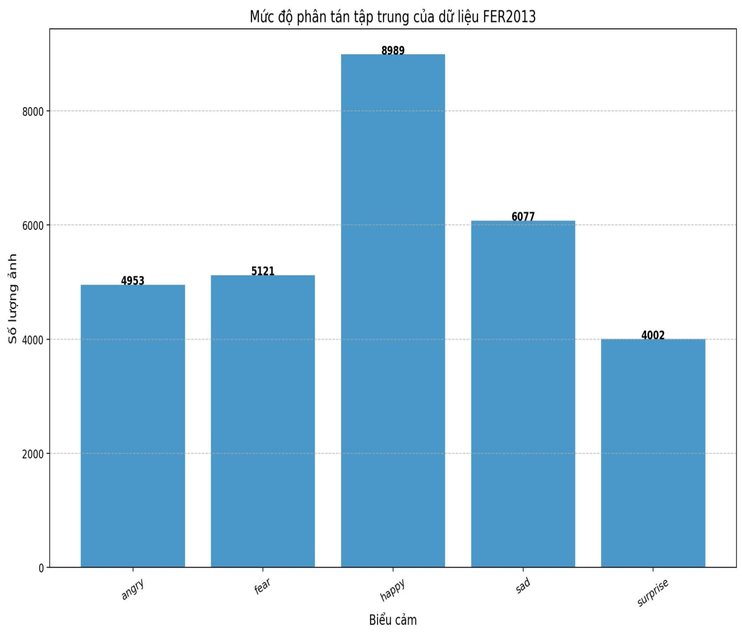
•Trợ lý ảo thông minh có khả năng phản hồi theo cảm xúc người dùng.

•Hỗ trợ tâm lý cho người gặp vấn đề về sức khỏe tinh thần.

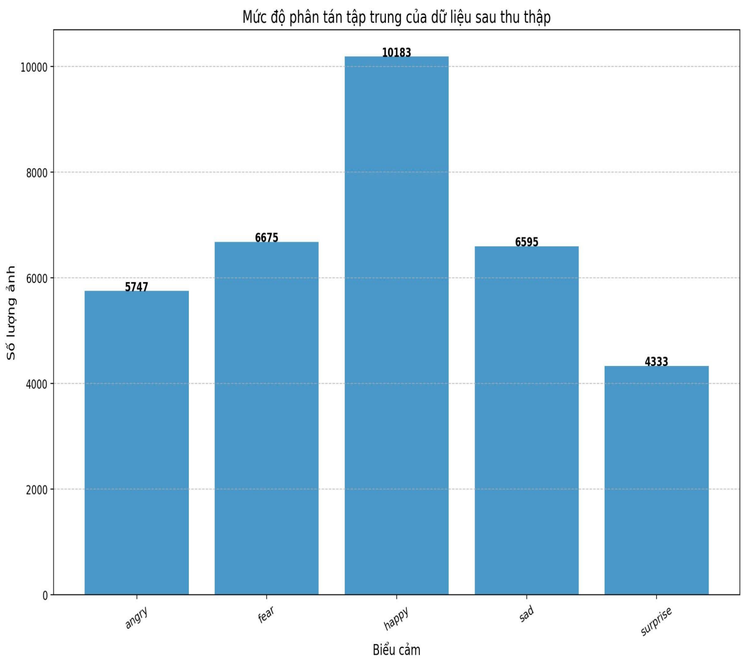
•Tích hợp vào hệ thống giao tiếp thông minh để cải thiện trải nghiệm người dùng.

Dataset

Dữ liệu FER2013



Dữ liệu thu thập

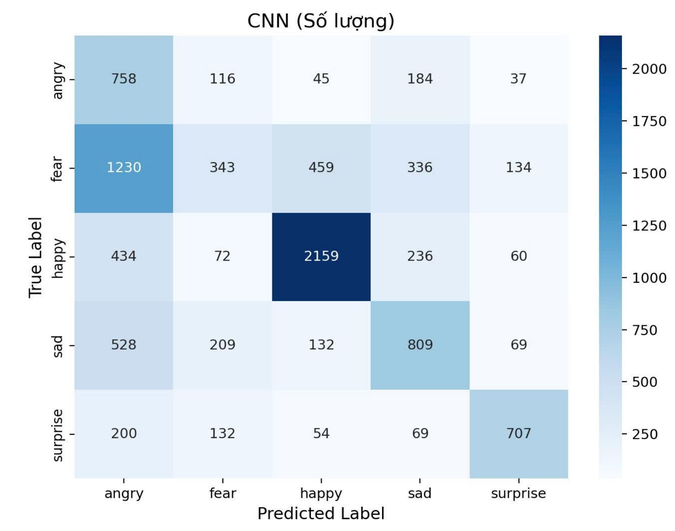


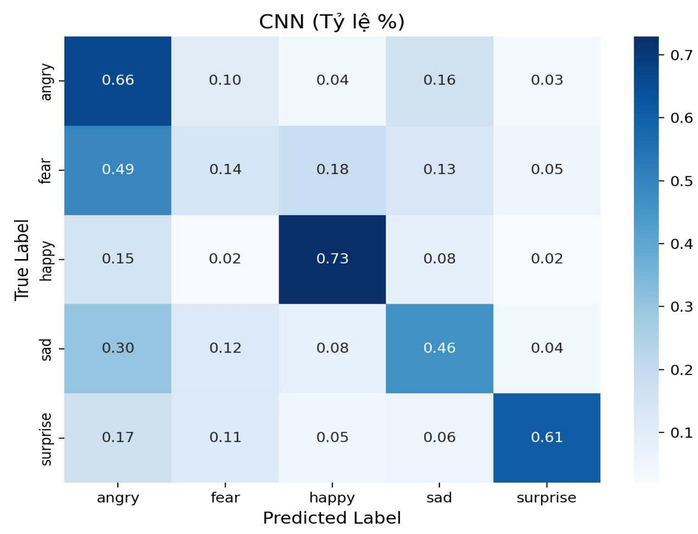
Kết quả

Đánh giá mô hình

Accuracy	0.50
Precision	0.54
Recall	0.50
F1-score	0.49

Ma trận nhầm lẫn





Kết luận và phương hướng tương lai

Kết luận:

Phương pháp đề xuất được kiểm chứng trên hai bộ dữ liệu công khai, cho thấy hiệu suất cải thiện và mô hình gọn nhẹ hơn so với các phương pháp khác.

Hướng phát triển:

Triển khai vào ứng dụng thực tế để nhận diện cảm xúc liên tục trong tương tác người - máy.