

BÁO CÁO THỰC HIỆN ĐỀ TÀI ĐỊNH VỊ VÙNG TRỌNG TÂM CỦA ẢNH

Nhóm: Nguyễn Tuấn Anh, Phạm Hoàng Anh, Đinh Trường An

Ngày 7 tháng 3 năm 2025

1 Các vấn đề

1.1 Phát biểu bài toán

Bài toán thực hiện phát hiện cảm xúc từ hình ảnh khuôn mặt bằng cách sử dụng mô hình học sâu và nhận diện biểu báo giao thông. Dữ liệu đầu vào là các hình ảnh khuôn mặt hoặc biển báo, đầu ra là nhãn cảm xúc hoặc loại biển báo tương ứng.

1.2 Dataset sử dụng

Để thực hiện nghiên cứu, chúng tôi sử dụng hai bộ dữ liệu:

1.2.1 Bộ dữ liệu CK+

Bộ dữ liệu **CK+** để xây dựng mô hình nhận diện cảm xúc.

- Bao gồm 981 hình ảnh với 7 nhãn cảm xúc: Anger, Contempt, Disgust, Fear, Happiness, Sadness, và Surprise.
- Hình ảnh được lưu trong các thư mục tương ứng với nhãn.
- Dữ liệu có thể tải về từ: <https://www.kaggle.com/davilsena/ckdataset>



Hình 1: Một số hình ảnh mẫu từ bộ dữ liệu CK+

1.2.2 Bộ dữ liệu GTSRB

Bộ dữ liệu **GTSRB** - **German Traffic Sign Recognition Benchmark** được sử dụng để nhận diện biển báo giao thông.

- Bài toán phân loại ảnh đa lớp với hơn 40 loại biển báo khác nhau.
- Gồm hơn 50.000 hình ảnh biển báo thực tế.
- Được sử dụng trong cuộc thi tại hội nghị *International Joint Conference on Neural Networks* (IJCNN) năm 2011.
- Dữ liệu có thể tải về từ: https://benchmark.ini.rub.de/gtsrb_news.html



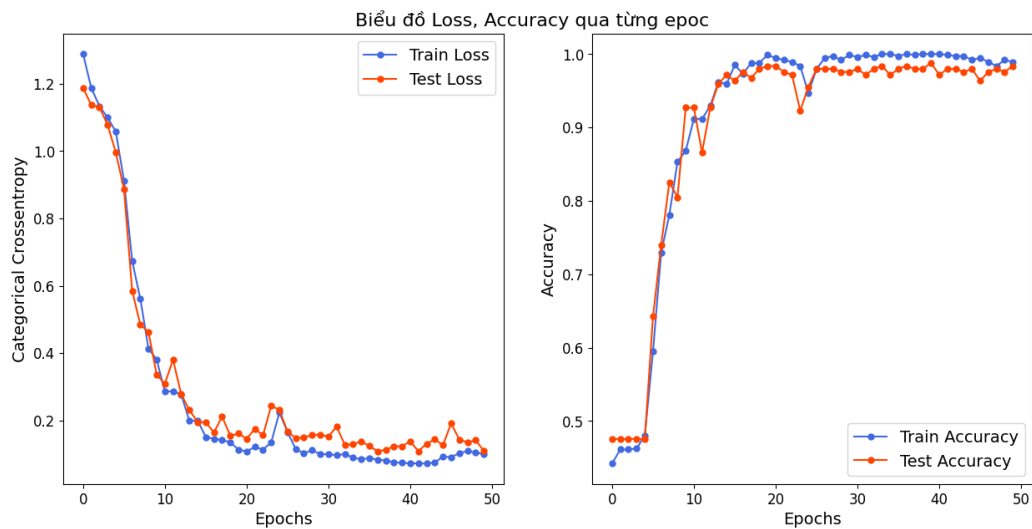
Hình 2: Một số hình ảnh mẫu từ bộ dữ liệu GTSRB

1.3 Xây dựng chương trình nhận diện

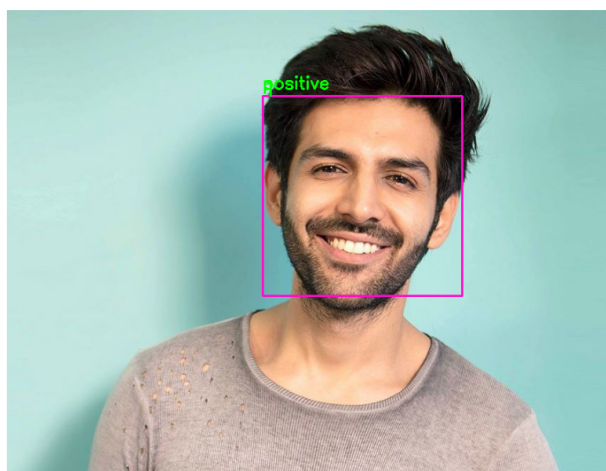
1.3.1 Nhận diện cảm xúc

Cấu trúc mô hình CNN sử dụng:

- Bốn lớp tích chập (Conv2D) với số filters tăng dần: 32, 64, 128, 256.
- Sử dụng ReLU activation, Dropout 0.1 sau mỗi lớp.
- Lớp Flatten để chuyển đổi dữ liệu trước khi đưa vào Dense layers.
- Một lớp ẩn với 128 neurons, Dropout 0.2, đầu ra 3 lớp với hàm kích hoạt sigmoid.



Hình 3: Biểu đồ Accuracy và Loss cho mô hình nhận diện cảm xúc



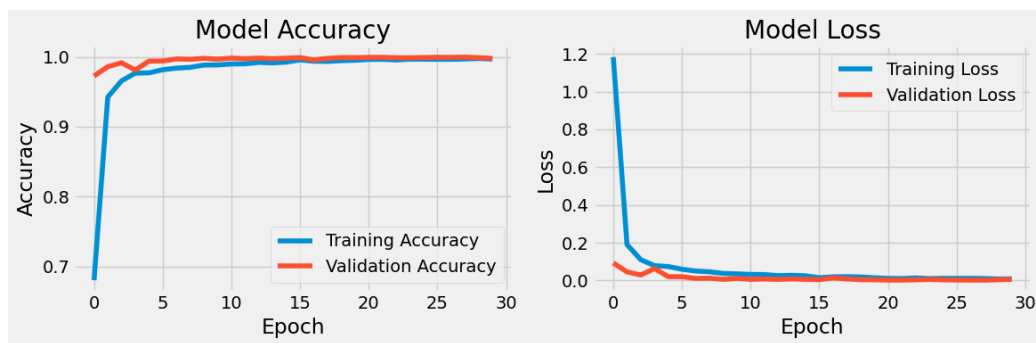
Hình 4: Dự đoán thử trên hình ảnh mẫu

Đánh giá: Mô hình đạt độ chính xác tốt trên tập kiểm tra, nhưng có thể cải thiện bằng cách tăng cường dữ liệu.

1.3.2 Nhận diện biển báo giao thông

Cấu trúc mô hình CNN sử dụng:

- Hai khối Conv2D với số filters tăng dần: 16, 32, 64, 128.
- MaxPooling2D và BatchNormalization sau mỗi khối.
- Lớp Fully Connected với 512 neurons, Dropout 0.5.
- Đầu ra có 43 lớp với Softmax activation.



Hình 5: Biểu đồ Accuracy và Loss cho mô hình nhận diện biển báo



Hình 6: Dự đoán thử trên hình ảnh mẫu

Đánh giá: Từ biểu đồ, có thể kết luận rằng mô hình CNN được huấn luyện đã hội tụ ổn định, không xảy ra overfitting nghiêm trọng, và cho kết quả phân loại cảm xúc khả quan trên bộ dữ liệu thử nghiệm.

1.4 Phương pháp FovEx: Định vị vùng trọng tâm

FovEx (**Foveated Explanation**) là một phương pháp tối ưu hóa nhằm xác định các vùng quan trọng trong hình ảnh, giúp mô hình học sâu tập trung vào các đặc điểm có ý nghĩa nhất.

1.4.1 Nguyên lý hoạt động

Quá trình thực hiện bao gồm:

- **Tạo không gian foveation:** Sử dụng mặt nạ Gaussian để mô phỏng mức độ tập trung của mô hình vào các vùng quan trọng.
- **Tính toán độ quan trọng:** Dựa trên gradient mất mát (*loss gradient*) để xác định vị trí tối ưu cần quan sát.
- **Tối ưu hóa scanpath:** Cập nhật vị trí nhìn dựa trên thuật toán gradient descent nhằm tối ưu hóa khả năng phân loại.
- **Sinh bản đồ nhiệt (heatmap):** Tổng hợp các vị trí nhìn để tạo ra bản đồ thể hiện mức độ quan trọng của từng vùng ảnh.

1.4.2 Ứng dụng trong nhận diện cảm xúc

Trong bài toán nhận diện cảm xúc từ khuôn mặt, sử dụng FovEx để định vị vùng trọng tâm của ảnh trong mô hình. Các bước cụ thể bao gồm:

- Tải và chuyển trọng số của mô hình đã huấn luyện từ Keras sang PyTorch (Vì Fovex không hoạt động với mô hình Keras/TensorFlow).
- Áp dụng mô hình FovEx trên hình ảnh khuôn mặt để xác định các vùng quan trọng nhất.
- Sử dụng bản đồ nhiệt để trực quan hóa các khu vực ảnh hưởng lớn đến quyết định của mô hình.



Hình 7: Bản đồ nhiệt FovEx cho mô hình nhận diện cảm xúc

1.4.3 Ứng dụng trong nhận diện biển báo giao thông

FovEx cũng có thể áp dụng để định vị vùng trong tâm của ảnh biển báo giao thông, tập trung vào các đặc điểm quan trọng như hình dạng và biểu tượng của biển báo. Quá trình này cũng tương tự như việc thực hiện với nhận diện cảm xúc.



Hình 8: Bản đồ nhiệt FovEx cho mô hình nhận diện biển báo giao thông

2 Tình hình thực hiện

2.1 Tiến độ thực hiện

Dưới đây là tiến độ thực hiện từng bước của dự án:

- Tìm hiểu về bộ dữ liệu CK+ và GTSRB (Hoàn thành)
- Xây dựng mô hình nhận diện cảm xúc và biển báo (Hoàn thành)
- Đánh giá và tinh chỉnh mô hình (Hoàn thành)
- Sử dụng phương pháp FovEx trên mô hình nhận diện cảm xúc (Hoàn thành)
- Sử dụng phương pháp FovEx trên mô hình nhận diện biển báo (Đang tối ưu)

2.2 Nhận xét và đánh giá

- Hiện tại báo cáo đã thu thập được 2 dataset và xử lý, đưa chúng vào mô hình. Việc thực hiện xây dựng mô hình CNN trên cả 2 dataset đều cho ra kết quả nhận diện, phân loại tốt.
- Việc áp dụng phương pháp FovEx lên mô hình nhận diện cảm xúc cũng đạt được kết quả khá tốt khi thử với nhiều ảnh khác nhau.
- Tuy nhiên với mô hình nhận diện biển báo, khi thực hiện áp dụng FovEx thì sẽ xảy ra vấn đề độ chính xác của mô hình giảm 1 cách đáng kể, mặc dù vẫn định vị được vùng trọng tâm trong ảnh
- Lý do của việc này có thể là do cách triển khai, code lỗi.
- Vẫn chưa nắm vững được hoàn toàn cách hoạt động của FovEx.

3 Dự kiến

Trong thời gian tới, báo cáo sẽ thực hiện các bước sau:

- Tìm hiểu sâu hơn cách FovEx hoạt động
- Hoàn thiện chương trình FovEx trên mô hình nhận diện biển báo
- Tiến hành thử nghiệm với 1 mô hình có hiệu suất thấp, áp dụng FovEx để cải thiện hiệu suất của mô hình.
- Tìm hiểu và so sánh kết quả với các phương pháp khác (Gradcam).

4 Tài liệu tham khảo

- FovEx: Human-inspired Explanations for Vision Transformers and Convolutional Neural Networks
- Traffic Sign Detection Using Convolutional Neural Networks