

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

KHOA CÔNG NGHỆ THÔNG TIN



Môn: THỰC TẬP CƠ SỞ

BÁO CÁO DỰ ÁN

Xây dựng một hệ thống nhận diện cảm xúc con người.

Giảng viên hướng dẫn: TS. Kim Ngọc Bách

Sinh viên thực hiện: Nguyễn Ngọc Bảo

Mã sinh viên: B23DCCN077

Lớp: D23CQCN07-B Nhóm: 11

Hà Nội 2026

LỜI CẢM ƠN

"Lời đầu tiên, nhóm chúng em xin gửi lời cảm ơn chân thành đến Học viện Công nghệ Bưu chính Viễn thông đã tạo điều kiện để chúng em tiếp cận môn học Thực tập cơ sở — một học phần quan trọng giúp kết nối lý thuyết và thực hành, môn học là hành trang quan trọng để chúng em hiểu rõ hơn về ngành học cũng như những công việc sẽ phải làm sau này.

Chúng em đặc biệt bày tỏ lòng biết ơn tới **thầy Kim Ngọc Bách**. Dù mới ở giai đoạn định hướng đề tài, thầy đã tận tình dẫn dắt, giúp chúng em định hình ý tưởng và phương pháp tiếp cận đúng đắn. Những kinh nghiệm quý báu của thầy sẽ là cơ sở để chúng em xây dựng đề tài này một cách nghiêm túc và khoa học.

Do bước đầu làm quen với nghiên cứu và thực tiễn, cũng như lần đầu thử sức ở lĩnh vực liên quan đến AI, đề tài chắc chắn không tránh khỏi những thiếu sót. Chúng em rất mong nhận được sự góp ý của thầy để em có thể hoàn thiện kế hoạch và triển khai giai đoạn thực tập tiếp theo một cách hiệu quả nhất."

1. GIỚI THIỆU DỰ ÁN

1.1. Lý do chọn đề tài

Trong những năm gần đây, bài toán nhận diện cảm xúc khuôn mặt (Facial Expression Recognition – FER) đã trở thành một hướng nghiên cứu quan trọng trong lĩnh vực Thị giác máy tính (Computer Vision). Nhận diện cảm xúc giúp máy tính hiểu được trạng thái tâm lý của con người thông qua biểu cảm khuôn mặt, từ đó nâng cao khả năng tương tác người – máy.

Sự phát triển mạnh mẽ của Deep Learning, đặc biệt là các mạng Convolutional Neural Network (CNN), đã giúp cải thiện đáng kể độ chính xác trong các bài toán phân loại ảnh. Tuy nhiên, CNN truyền thống vẫn còn hạn chế trong việc tập trung vào các vùng đặc trưng quan trọng trên khuôn mặt như mắt, miệng hoặc lông mày.

Gần đây, cơ chế Attention được đề xuất nhằm giúp mô hình học tập trung hơn vào các đặc trưng quan trọng. Một trong những phương pháp Attention hiệu quả là CBAM (Convolutional Block Attention Module), kết hợp Channel Attention và Spatial Attention.

Việc lựa chọn đề tài “Xây dựng và huấn luyện mô hình CNN kết hợp CBAM cho bài toán nhận diện cảm xúc khuôn mặt” xuất phát từ các lý do sau:

- CNN là kiến trúc nền tảng trong xử lý ảnh
- CBAM giúp cải thiện khả năng tập trung vào vùng biểu cảm quan trọng
- Đề tài giúp hiểu sâu về Attention Mechanism
- Rèn luyện kỹ năng xây dựng và tối ưu mô hình Deep Learning

1.2. Ý nghĩa và tính ứng dụng

1.2.1. Ý nghĩa khoa học

Đề tài có ý nghĩa khoa học quan trọng trong lĩnh vực Thị giác máy tính và Học sâu.

Thứ nhất, đề tài giúp làm rõ cơ chế hoạt động của CNN trong bài toán phân loại ảnh khuôn mặt.

Thứ hai, nghiên cứu cơ chế Channel Attention và Spatial Attention trong CBAM giúp hiểu sâu hơn về cách mô hình học trọng số đặc trưng.

Thứ ba, đề tài góp phần phân tích các chỉ số đánh giá mô hình phân loại như:

- Accuracy
- Precision
- Recall
- F1-score
- Confusion Matrix

1.2.2. Ý nghĩa thực tiễn

Nhận diện cảm xúc có nhiều ứng dụng thực tế:

- Giáo dục thông minh (phân tích mức độ tập trung)
- Hệ thống chăm sóc khách hàng
- Robot và trợ lý ảo
- Phân tích hành vi người dùng
- Hệ thống tương tác người – máy

Việc triển khai mô hình CNN + CBAM giúp tạo nền tảng xây dựng hệ thống nhận diện cảm xúc thời gian thực.

● 1.3. Giá trị học thuật

Đề tài giúp nghiên cứu các nội dung:

➤ Kiến trúc mạng sâu

- Convolutional Neural Network
- Batch Normalization
- Dropout
- Attention Mechanism (CBAM)
- Hoặc có thể :Residual Learning

➤ Thuật toán tối ưu

- Cross-Entropy Loss
- Backpropagation
- Adam Optimizer
- Learning Rate Scheduler

➤ Các chỉ số đánh giá

- Accuracy
- Precision – Recall
- F1-score
- Confusion Matrix

2. CƠ SỞ LÝ THUYẾT VÀ CÔNG NGHỆ SỬ DỤNG

2.1. Cơ sở lý thuyết

2.1.1. Tổng quan về Facial Expression Recognition

FER là bài toán phân loại ảnh, trong đó:

- Đầu vào: ảnh khuôn mặt
- Đầu ra: nhãn cảm xúc

Bài toán thuộc nhóm Image Classification trong Computer Vision.

2.1.2. Mạng nơ-ron tích chập (CNN)

CNN gồm các thành phần chính:

- Convolution layer
- Activation function (ReLU)
- Pooling layer
- Fully Connected layer

CNN có khả năng trích xuất đặc trưng phân cấp từ thấp đến cao.

2.1.3. Cơ chế Attention và CBAM

CBAM (Convolutional Block Attention Module) gồm hai thành phần:

- a) Channel Attention: Giúp mô hình xác định kênh đặc trưng quan trọng.
- b) Spatial Attention: Giúp mô hình tập trung vào vị trí quan trọng trong ảnh.

CBAM giúp tăng cường khả năng biểu diễn đặc trưng mà không làm tăng quá nhiều số tham số.

2.2. Công nghệ sử dụng

2.2.1. Ngôn ngữ lập trình

- Python

2.2.2. Thư viện Deep Learning

- PyTorch

PyTorch hỗ trợ xây dựng mô hình linh hoạt, dynamic computation graph và GPU acceleration (CUDA).

2.2.3. Thư viện xử lý ảnh

- OpenCV
- Pillow (PIL)
- Matplotlib

2.2.4. Dataset sử dụng

FER2013 gồm khoảng 35.000 ảnh 48×48 grayscale, chia thành 7 lớp:

- Angry
- Disgust
- Fear
- Happy
- Sad
- Surprise
- Neutral

Bộ dữ liệu này phù hợp cho môi trường học thuật và đánh giá mô hình phân loại cảm xúc.

3. PHÂN TÍCH YÊU CẦU DỰ ÁN

3.1. Mục tiêu dự án

- Xây dựng mô hình CNN baseline
- Tích hợp CBAM vào CNN
- Huấn luyện trên FER2013
- So sánh hiệu năng CNN và CNN+CBAM
- Phân tích Confusion Matrix
- Trực quan hóa Attention (Grad-CAM nếu có)

3.2. Yêu cầu chức năng

Hệ thống cần:

- Đọc và xử lý dữ liệu ảnh
- Chia tập train/validation/test
- Xây dựng mô hình CNN và CNN+CBAM
- Huấn luyện mô hình
- Lưu và tải checkpoint
- Thực hiện inference
- Hiển thị nhãn dự đoán
- Vẽ confusion matrix
- Vẽ biểu đồ loss và accuracy

3.3. Yêu cầu về dữ liệu

Dataset cần:

- Ảnh khuôn mặt
- Nhãn cảm xúc

Dữ liệu được chia thành:

- Training set
- Validation set
- Test set

4. KẾ HOẠCH THỰC HIỆN ĐỀ TÀI

Giai đoạn 1 (Tuần 1–2): Nghiên cứu lý thuyết

- Tìm hiểu CNN
- Tìm hiểu CBAM
- Chuẩn bị dataset
- Xây dựng DataLoader

Giai đoạn 2 (Tuần 3–4): Xây dựng CNN Baseline

- Thiết kế kiến trúc CNN
- Huấn luyện mô hình
- Đánh giá accuracy

Giai đoạn 3 (Tuần 5–6): Tích hợp CBAM

- Cài đặt Channel Attention
- Cài đặt Spatial Attention
- Tích hợp vào CNN
- Huấn luyện lại mô hình

Giai đoạn 4 (Tuần 7–8): Đánh giá và phân tích

- Tính Precision, Recall, F1-score
- Vẽ Confusion Matrix
- So sánh CNN và CNN+CBAM
- Trực quan hóa Attention

● Giai đoạn 5 (Tuần 9–10): Hoàn thiện báo cáo

- Tổng hợp nội dung
- Bổ sung bảng biểu và hình ảnh
- Viết kết luận
- Kiểm tra và tối ưu mô hình

TÀI LIỆU THAM KHẢO

[1] Ian Goodfellow, Yoshua Bengio, Aaron Courville (2016). *Deep Learning*. MIT Press.

[2] Alex Krizhevsky, Ilya Sutskever, Geoffrey Hinton (2012). ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems.

[3] Sanghyun Woo, Jongchan Park, Joon-Young Lee, In So Kweon (2018). CBAM: Convolutional Block Attention Module. Proceedings of the European Conference on Computer Vision.

[4] FER2013 Facial Expression Recognition Dataset.

[5] PyTorch Documentation.

[6] OpenCV Documentation.