

●
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
khuôn mặt tự động.”**

Sinh viên thực hiện: Nguyễn Ngọc Bảo

Mã sinh viên: B23DCCN077

Lớp: D23CQCN07-B Nhóm: 11

Giảng viên hướng dẫn: TS. Kim Ngọc Bách

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) và Trí tuệ nhân tạo (Artificial Intelligence). Nhận diện cảm xúc giúp máy tính hiểu được trạng thái cảm xúc của con người thông qua biểu cảm khuôn mặt, từ đó mở rộng 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 nâng cao 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 quan trọng trên khuôn mặt như mắt, miệng hoặc lông mày – những vùng đóng vai trò quan trọng trong việc thể hiện cảm xúc.

Trong những năm gần đây, cơ chế Attention được đề xuất nhằm giúp mô hình tập trung vào các đặc trưng quan trọng hơn trong quá trình học. Việc kết hợp CNN với Attention giúp cải thiện khả năng trích xuất đặc trưng và tăng hiệu quả phân loại. Việc lựa chọn đề tài **“Xây dựng và huấn luyện mô hình CNN kết hợp Attention 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 và phân loại hình ảnh.
- Attention giúp mô hình tập trung vào các vùng đặc trưng quan trọng trên khuôn mặt.
- Đề tài giúp hiểu sâu về cấu trúc mạng CNN, cơ chế Attention, hàm mất mát và các chỉ số đánh giá như Accuracy, Precision, Recall, F1-score và Confusion Matrix.
- Việc tự xây dựng và huấn luyện mô hình giúp nâng cao kỹ năng lập trình, xử lý dữ liệu và tối ưu mô hình Deep Learning.

Thông qua đề tài này, người thực hiện không chỉ củng cố kiến thức lý thuyết về Deep Learning mà còn rèn luyện kỹ năng triển khai hệ thống AI thực tế.

1.2. Ý nghĩa và tính ứng dụng

1.2.1. Ý nghĩa khoa học

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

Thứ nhất, đề tài giúp làm rõ nguyên lý hoạt động của mạng CNN trong bài toán phân loại ảnh và cơ chế Attention trong việc tăng cường khả năng tập trung vào đặc trưng quan trọng.

Thứ hai, đề tài giúp nghiên cứu và phân tích các chỉ số đánh giá mô hình phân loại như:

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



Thứ ba, đề tài tạo nền tảng để mở rộng nghiên cứu sang các kiến trúc hiện đại hơn như Vision Transformer hoặc các mô hình lai CNN–Transformer.

1.2.2. Ý nghĩa thực tiễn

Nhận diện cảm xúc có tính ứng dụng cao trong nhiều lĩnh vực:

- Hệ thống giáo dục thông minh (phân tích mức độ tập trung của học sinh)
- Chăm sóc khách hàng (phân tích cảm xúc người dùng) • Hệ thống hỗ trợ tâm lý
- Tương tác người – máy (Human-Computer Interaction)
- Robot và trợ lý ảo

Việc xây dựng thành công mô hình CNN–Attention giúp tạo nền tảng triển khai hệ thống nhận diện cảm xúc thời gian thực.

1.2.3. Dataset sử dụng

Trong đề tài này, sử dụng bộ dữ liệu:

FER2013

FER2013 bao gồm khoảng 35.000 ảnh khuôn mặt kích thước 48×48 pixel, được chia thành 7 lớp cảm xúc:

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

Bộ dữ liệu này được sử dụng rộng rãi trong nghiên cứu học thuật và phù hợp với môi trường đào tạo.

1.3. Giá trị học thuật

Đề tài có giá trị học thuật cao vì nghiên cứu các nội dung sau:

Kiến trúc mạng sâu

- Cấu trúc Convolutional Neural Network
- Batch Normalization
- Dropout
- Residual Learning (nếu mở rộng)
- Attention Mechanism (Channel Attention / Spatial Attention / Multi-Head Attention)

Thuật toán tối ưu

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

Các chỉ số đánh giá

- Confusion Matrix
- Precision – Recall
- F1-score
- ROC Curve (nếu mở rộng)

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

Nhận diện cảm xúc khuôn mặt là bài toán phân loại ảnh, trong đó hệ thống cần:

- Nhận đầu vào là ảnh khuôn mặt

- - Phân loại ảnh vào một trong các lớp cảm xúc

Bài toán này 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, LeakyReLU)
- Pooling layer
- Fully Connected layer

CNN có khả năng:

- Trích xuất đặc trưng không gian
- Nhận diện mẫu
- Học biểu diễn phân cấp từ mức thấp đến mức cao

2.1.3. Cơ chế Attention

Attention giúp mô hình tập trung vào các đặc trưng quan trọng hơn.

Các dạng Attention có thể sử dụng:

- SE Block (Squeeze-and-Excitation)
- CBAM (Convolutional Block Attention Module)
- Multi-Head Self Attention

Attention giúp cải thiện khả năng phân loại bằng cách tăng trọng số cho các vùng biểu cảm quan trọng như mắt và miệng.

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:

- Dynamic computation graph □ GPU acceleration (CUDA)
- Xây dựng mô hình linh hoạt

2.2.3. Thư viện xử lý ảnh

- OpenCV
- Pillow (PIL)
- Matplotlib

3. PHÂN TÍCH YÊU CẦU DỰ ÁN

3.1. Mục tiêu dự án

- Xây dựng mô hình CNN baseline
- Xây dựng mô hình CNN + Attention
- Huấn luyện trên FER2013
- So sánh hiệu năng hai mô hình
- Phân tích kết quả và đề xuất cải tiến

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
- Huấn luyện và lưu checkpoint
- Thực hiện inference
- Vẽ confusion matrix
- Hiển thị biểu đồ loss và accuracy

4. KẾ HOẠCH THỰC HIỆN

Giai đoạn 1 (Tuần 1–2)

- Nghiên cứu CNN và Attention
- Chuẩn bị dataset □ Xây dựng pipeline DataLoader

Giai đoạn 2 (Tuần 3–4)

- Xây dựng CNN baseline
- Huấn luyện và đánh giá

Giai đoạn 3 (Tuần 5–6)

- Thêm Attention module
- Huấn luyện lại mô hình □ So sánh kết quả **Giai đoạn 4 (Tuần 7–8)**

- Vẽ confusion matrix
- Phân tích sai số
- Trực quan hóa Attention (Grad-CAM nếu có) **Giai đoạn 5 (Tuần 9–10)**
- Hoàn thiện báo cáo

- Chuẩn bị demo
- Kiểm tra và tối ưu mô hình

TÀI LIỆU THAM KHẢO

1. A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd ed. Sebastopol, CA, USA: O'Reilly Media, 2019.
2. F. Chollet, *Deep Learning with Python*, 2nd ed. Shelter Island, NY, USA: Manning Publications, 2021.
3. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
4. I. J. Goodfellow et al., "Challenges in representation learning: A report on three machine learning contests," in *Proceedings of the International Conference on Neural Information Processing (ICONIP)*, 2013.
5. A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.