UỶ BAN NHÂN DÂN THÀNH PHỐ HỒ CHÍ MINH

TRƯỜNG ĐẠI HỌC SÀI GÒN

NHẬN DIỆN BIỂU CẢM KHUÔN MẶT TRONG ĐIỀU KIỆN ÁNH SÁNG YẾU SỬ DỤNG CNN NHẸ KẾT HỢP KỸ THUẬT TĂNG CƯỜNG DỮ LIỆU THÍCH ỨNG

LUẬN VĂN MÔN HỌC NCKH TRONG CNTT

NGÀNH: CÔNG NGHỆ THÔNG TIN

Nhóm sinh viên thực hiện:

Họ và tên	MSSV
Văn Tuấn Kiệt	3122410202
Mai Phúc Lâm	3122410207
Nguyễn Đức Duy Lâm	3122410208
Nguyễn Hữu Lộc	3122410213

Giáo viên hướng dẫn: Đỗ Như Tài

TP.HCM, 2025

BÁO CÁO LUẬN VĂN

1 Tổng quan vấn đề

1.1 Lý do chọn đề tài

Nhận diện biểu cảm khuôn mặt (Facial Expression Recognition - FER) đóng vai trò quan trọng trong các ứng dụng thực tiễn như giao tiếp người-máy, giám sát an ninh, và phân tích hành vi. Tuy nhiên, trong các điều kiện ánh sáng yếu, chẳng hạn như môi trường ban đêm hoặc khu vực thiếu sáng, hiệu quả của các hệ thống FER giảm đáng kể do chất lượng hình ảnh thấp. Các nghiên cứu gần đây (2020–2025) chủ yếu tập trung vào điều kiện ánh sáng lý tưởng, trong khi các giải pháp cho ánh sáng yếu thường phức tạp, đòi hỏi tài nguyên tính toán lớn hoặc không tối ưu cho các thiết bị nhúng.

Việc phát triển một phương pháp nhận diện biểu cảm hiệu quả trong điều kiện ánh sáng yếu, sử dụng mô hình CNN nhẹ (như MobileNetV3) và kỹ thuật tăng cường dữ liệu thích ứng, không chỉ đáp ứng nhu cầu thực tiễn mà còn mang lại giá trị khoa học thông qua việc cải tiến các kỹ thuật hiện có. Đề tài này được chọn vì tính khả thi trong thời gian nghiên cứu (6 tuần), tính mới trong việc kết hợp các phương pháp đơn giản nhưng hiệu quả, và tiềm năng ứng dụng trong các hệ thống thực tế như camera giám sát hoặc thiết bị IoT.

1.2 Vấn đề nghiên cứu

Trong điều kiện ánh sáng yếu, các mô hình nhận diện biểu cảm khuôn mặt truyền thống thường gặp khó khăn do độ tương phản thấp, nhiễu ảnh, và mất chi tiết khuôn mặt. Các phương pháp hiện tại như sử dụng GAN (Generative Adversarial Networks) hoặc Retinex-based preprocessing tuy hiệu quả nhưng phức tạp, yêu cầu thời gian huấn luyện lâu và tài nguyên tính toán lớn, không phù hợp với các ứng dụng thời gian thực hoặc thiết bị có tài nguyên hạn chế. Ngoài ra, các kỹ thuật tăng cường dữ liệu cố định (fixed augmentation) không tối ưu vì không thích nghi với mức độ ánh sáng yếu khác nhau của từng ảnh.

Vấn đề nghiên cứu được đặt ra là: Làm thế nào để phát triển một hệ thống

nhận diện biểu cảm khuôn mặt trong điều kiện ánh sáng yếu, sử dụng mô hình CNN nhẹ và kỹ thuật tăng cường dữ liệu thích ứng, nhằm đạt được độ chính xác cao, tốc độ xử lý nhanh, và khả năng triển khai trên các thiết bị nhúng?

1.3 Mục tiêu nghiên cứu

Mục tiêu tổng quát của nghiên cứu là xây dựng một hệ thống nhận diện biểu cảm khuôn mặt hiệu quả trong điều kiện ánh sáng yếu, sử dụng mạng nơ-ron tích chập nhẹ (MobileNetV3) kết hợp với kỹ thuật tăng cường dữ liệu thích ứng. Các mục tiêu cụ thể bao gồm:

1. Phát triển một pipeline tăng cường dữ liệu thích ứng, tự động điều chỉnh các kỹ thuật tăng cường dựa trên mức độ ánh sáng yếu của từng ảnh.
2. Huấn luyện và tinh chỉnh mô hình MobileNetV3 để nhận diện biểu cảm khuôn mặt trong điều kiện ánh sáng yếu với độ chính xác cao.
3. Đánh giá và so sánh hiệu quả của phương pháp đề xuất với các kỹ thuật tăng cường dữ liệu cố định và các mô hình CNN khác (nếu khả thi).

1.4 Câu hỏi nghiên cứu

Nghiên cứu tập trung trả lời các câu hỏi sau:

1. Làm thế nào để thiết kế một pipeline tăng cường dữ liệu thích ứng, hiệu quả trong việc cải thiện chất lượng ảnh ánh sáng yếu cho nhận diện biểu cảm khuôn mặt?
2. Mô hình MobileNetV3 có thể đạt được độ chính xác tương đương hoặc vượt trội so với các kỹ thuật tăng cường dữ liệu cố định trong điều kiện ánh sáng yếu không?
3. Các kỹ thuật tăng cường dữ liệu thích ứng ảnh hưởng như thế nào đến hiệu suất của mô hình CNN nhẹ trong nhận diện biểu cảm khuôn mặt?

1.5 Phạm vi nghiên cứu

- Đối tượng nghiên cứu: Các biểu cảm khuôn mặt (ví dụ: vui, buồn, tức giận, ngạc nhiên) trong điều kiện ánh sáng yếu, được mô phỏng hoặc thu thập từ

các bộ dữ liệu công khai như FER-2013 hoặc RAF-DB.

- Phạm vi không gian: Nghiên cứu tập trung vào xử lý hình ảnh tĩnh (static images), không bao gồm dữ liệu video hoặc dữ liệu đa phổ.
- Phạm vi thời gian: Nghiên cứu được thực hiện trong 6 tuần, từ tháng 5 đến tháng 6 năm 2025, với các thí nghiệm dựa trên dữ liệu công khai và mô hình pre-trained.
- Phạm vi kỹ thuật: Sử dụng mô hình CNN nhẹ (MobileNetV3) và các kỹ thuật tăng cường dữ liệu như gamma correction, histogram equalization, được triển khai bằng Python với các thư viện TensorFlow/Keras và OpenCV.

2 Lược khảo tài liệu

3 Phương pháp nghiên cứu

3.1 Thiết kế nghiên cứu

Nghiên cứu được thiết kế theo phương pháp định lượng, tập trung vào việc xây dựng và đánh giá hiệu suất của các mô hình học sâu trong bài toán nhận diện biểu cảm khuôn mặt (Facial Expression Recognition - FER) trong điều kiện ánh sáng yếu. Phương pháp định lượng được chọn vì mục tiêu nghiên cứu là đo lường các chỉ số hiệu suất cụ thể (Accuracy, Precision, Recall, F1-score và thời gian suy luận) của hai mô hình CNN: MobileNetV3 (mô hình nhẹ) và ResNet18 (mô hình sâu hơn), khi áp dụng kỹ thuật tăng cường dữ liệu thích ứng.

Quá trình nghiên cứu bao gồm ba giai đoạn chính:

- Tiền xử lý dữ liệu: Sử dụng tập dữ liệu FER-2013, áp dụng các kỹ thuật tăng cường dữ liệu thích ứng để mô phỏng điều kiện ánh sáng yếu.
- Huấn luyện và tối ưu mô hình: Triển khai MobileNetV3 và ResNet18, tinh chỉnh các tham số để phù hợp với bài toán FER.
- Đánh giá và so sánh: So sánh hiệu suất và thời gian suy luận của các mô hình khi có và không áp dụng kỹ thuật tăng cường dữ liệu thích ứng.

3.2 Đối tượng và mẫu nghiên cứu

3.2.1 Đối tượng nghiên cứu

Đối tượng nghiên cứu là các kỹ thuật nhận diện biểu cảm khuôn mặt trong điều kiện ánh sáng yếu, với trọng tâm vào:

- Mô hình học sâu: MobileNetV3 và ResNet18 dùng để phân loại 7 biểu cảm khuôn mặt (vui, buồn, tức giận, sợ hãi, ngạc nhiên, ghê tởm, trung lập).
- Kỹ thuật tăng cường dữ liệu thích ứng: Các phương pháp như gamma correction, contrast stretching và histogram equalization, được điều chỉnh dựa trên đặc trưng ánh sáng của hình ảnh.

3.2.2 Mẫu nghiên cứu

Mẫu nghiên cứu là tập dữ liệu FER-2013, chứa 35.887 hình ảnh khuôn mặt (48x48 pixel, ảnh xám) được phân loại thành 7 biểu cảm. Tập dữ liệu được chia như sau:

- Tập huấn luyện: 28.709 hình ảnh (80%).
- Tập xác thực (validation): 3.589 hình ảnh (10.00%).
- Tập kiểm tra: 3.589 hình ảnh (10.00%).

Nhằm mô phỏng điều kiện ánh sáng yếu, một tập dữ liệu phụ được tạo ra bằng cách giảm độ sáng của ảnh gốc. Quá trình này thực hiện bằng cách chuyển ảnh sang không gian màu HSV, giảm kênh độ sáng (V) theo một hệ số cố định, sau đó chuyển lại về không gian RGB. Cụ thể, độ sáng được giảm xuống 10% so với ảnh ban đầu.

3.3 Cách thu thập dữ liệu

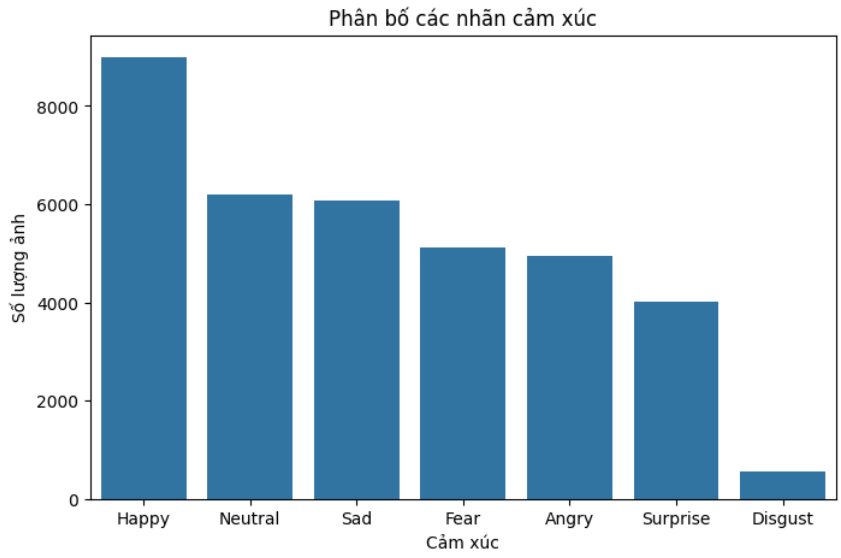
Dữ liệu được thu thập từ tập dữ liệu FER-2013 công khai trên nền tảng Kaggle. Các bước gồm:

Thu thập dữ liệu

- Tải tập dữ liệu FER-2013 từ Kaggle.
- Kiểm tra tính toàn vẹn (số lượng ảnh, định dạng, chất lượng).

3.3.1 EDA dữ liệu

Phân tích dữ liệu khám phá (EDA) được thực hiện trên tập dữ liệu FER-2013 nhằm hiểu rõ cấu trúc, phân phối và đặc trưng của dữ liệu trước khi áp dụng các mô hình học sâu. Dữ liệu được lưu trữ dưới dạng tệp CSV với ba cột chính: cột emotion (nhãn cảm xúc, giá trị từ 0 đến 6), cột pixels (tập hợp các giá trị pixel của ảnh dưới dạng chuỗi số), và cột Usage (chỉ định tập huấn luyện, xác thực hoặc kiểm tra). Kích thước tổng cộng của tập dữ liệu là 35.887 mẫu, trong đó mỗi hình ảnh có độ phân giải 48x48 pixel.

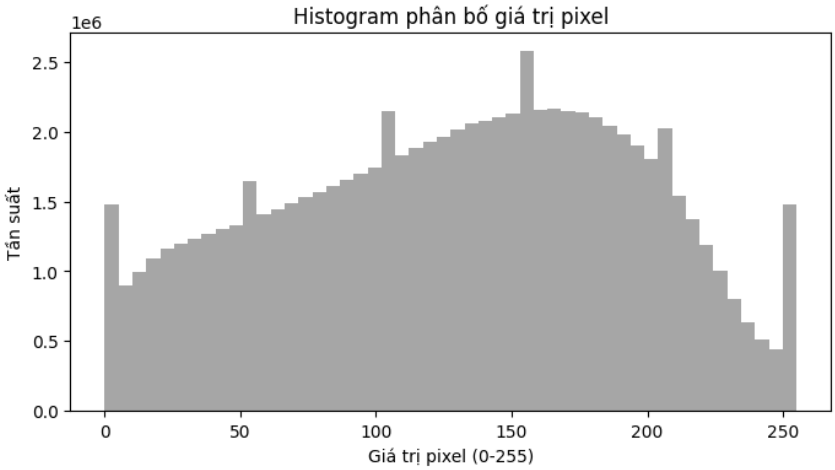


Hình 1: Phân bố nhãn cảm xúc trong tập dữ liệu FER-2013.

Ý nghĩa:

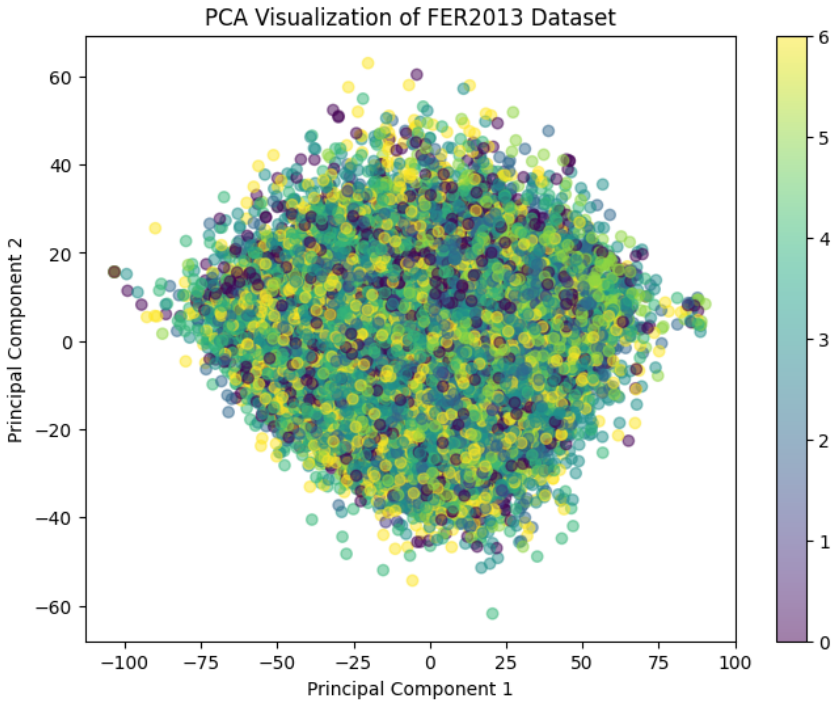
Hình 2: Một số hình ảnh mẫu từ tập dữ liệu FER-2013.

Ý nghĩa:



Hình 3: Phân bố pixel trong tập dữ liệu FER-2013.

Ý nghĩa:



Hình 4: Trực quan hóa dữ liệu FER-2013 bằng PCA.

Ý nghĩa:

3.3.2 Tiền xử lý dữ liệu

Các bước tiền xử lý được thực hiện nhằm cải thiện chất lượng ảnh đầu vào và mô phỏng các điều kiện môi trường khác nhau, cụ thể như sau:

- Chuẩn hóa hình ảnh: Loại bỏ nhiễu và đảm bảo định dạng đồng nhất (kích thước ảnh, không gian màu), giúp mô hình huấn luyện ổn định hơn.
- Mô phỏng điều kiện ánh sáng yếu: Để mô phỏng môi trường có ánh sáng yếu, hình ảnh được chuyển sang không gian màu HSV và kênh độ sáng (V) được giảm xuống còn 10% so với ảnh gốc. Sau đó, ảnh được chuyển lại về không gian RGB để sử dụng trong huấn luyện.

3.3.3 Tăng cường dữ liệu thích ứng

Áp dụng các phép biến đổi linh hoạt dựa trên đặc trưng ánh sáng của từng ảnh. Việc tăng cường được thực hiện bằng Python với OpenCV và NumPy.

3.4 Phân tích dữ liệu

3.4.1 Công cụ và phần mềm

- Python: xử lý dữ liệu và huấn luyện mô hình.
- TensorFlow/Keras: xây dựng và đánh giá mô hình.
- OpenCV: tiền xử lý ảnh.
- NumPy, Pandas: quản lý dữ liệu.
- Matplotlib, Seaborn: trực quan hóa kết quả.

3.4.2 Quy trình phân tích

- Huấn luyện mô hình MobileNetV3Small:
  - Sử dụng mô hình MobileNetV3Small với trọng số ImageNet, loại bỏ phần fully-connected gốc (include_top=False).
  - Chỉ tinh chỉnh 30 lớp cuối cùng trong mạng, các lớp còn lại được đóng băng để giữ lại đặc trưng đã học.
  - Kiến trúc phần đầu ra gồm: Global Average Pooling, hai lớp Dense (128 và 64 nodes, activation ReLU), kèm Dropout 0.3, kết thúc bằng lớp Softmax với 7 nhãn đầu ra.
  - Hàm mất mát: Categorical Crossentropy.
  - Tối ưu hóa bằng Adam (learning rate mặc định).
  - Số epoch: 10, sử dụng Early Stopping với patience = 3 để tránh overfitting.
- Huấn luyện mô hình với ResNet18:
  - Sử dụng mô hình ResNet18 với trọng số đã được huấn luyện sẵn trên tập ImageNet (ResNet18_Weights.IMAGENET1K_V1).
  - Điều chỉnh lại lớp Fully Connected cuối cùng thành nn.Linear(..., 7) để phù hợp với bài toán phân loại 7 cảm xúc trên tập dữ liệu FER2013.
  - Hàm mất mát sử dụng là CrossEntropyLoss, phù hợp với phân loại đa lớp.
  - Trình tối ưu hóa: Adam với learning rate 0.001.

  - Mô hình được huấn luyện trong 20 epoch.
  - Trong quá trình huấn luyện, độ chính xác và mất mát (loss) trên tập huấn luyện và tập xác thực được theo dõi để đánh giá hiệu quả mô hình. Mô hình tốt nhất được lưu lại sau mỗi epoch nếu có cải thiện.
- Đánh giá mô hình:
  - Các chỉ số đánh giá: Accuracy, Precision, Recall, F1-score.
  - Đo thời gian suy luận trung bình trên CPU (per image).
  - Kích cỡ mô hình sau huấn luyện.
- So sánh mô hình:
  - MobileNetV3 (cơ bản vs. tăng cường).
  - ResNet18 (cơ bản vs. tăng cường).
  - So sánh giữa MobileNetV3 và ResNet18.
- Phân tích kết quả:
  - Ma trận nhầm lẫn, biểu đồ Accuracy theo epoch.
  - Quan sát các trường hợp dự đoán sai.

3.4.3 Thiết bị triển khai

Thực nghiệm được thực hiện trên máy MacBook Air M1, được trang bị chip Apple M1 và RAM 8GB. Ngoài ra, Google Colab cũng được sử dụng để mô phỏng điều kiện tài nguyên thấp, với việc chỉ sử dụng CPU thay vì GPU nhằm đánh giá thời gian suy luận, phù hợp với môi trường nhúng.

3.5 Phương pháp so sánh

Nghiên cứu tiến hành so sánh định lượng qua các chỉ số hiệu suất (Accuracy, Precision, Recall, F1-score) và thời gian suy luận giữa:

- MobileNetV3 cơ bản vs. tăng cường.
- ResNet18 cơ bản vs. tăng cường.
- So sánh giữa MobileNetV3 và ResNet18.

Kết quả được trình bày dưới dạng bảng và biểu đồ để làm rõ hiệu quả của các kỹ thuật và sự phù hợp của mô hình trong ứng dụng thực tế.

4 Thực nghiệm và thảo luận

5 Kết luận và hướng phát triển

6 Danh mục tài liệu tham khảo

[1] S. Kusal et al., "A review on text-based emotion detection—techniques, applications, datasets, and future directions," arXiv preprint, arXiv:2205.03235, 2022.

[2] W. Wu, J. Weng, P. Zhang, X. Wang, W. Yang, and J. Jiang, "URetinex-Net: Retinex-based deep unfolding network for low-light image enhancement," in Proc. IEEE CVPR, 2022, pp. 5901–5910.

[3] M. Bie et al., "DA-FER: Domain adaptive facial expression recognition," Appl. Sci., vol. 13, no. 10, p. 6314, 2023, doi: 10.3390/app13106314.

[4] L. A. Al Hak, W. A. Ali, and S. J. Saba, "Facial expression recognition using data augmentation and transfer learning," Ingénierie des Systèmes d'Information, vol. 29, no. 3, pp. 1219–1225, 2024, doi: 10.18280/isi.290338.

[5] A. G. Howard et al., "Searching for MobileNetV3," in Proc. IEEE ICCV, 2019, pp. 1314–1324, doi: 10.1109/ICCV.2019.00140.

[6] X. Liang, J. Liang, T. Yin, and X. Tang, "A lightweight method for face expression recognition based on improved MobileNetV3," IET Image Process., vol. 17, no. 8, pp. 2375–2384, 2023, doi: 10.1049/ipe2.12798.

[7] S. B. R. Prasad and B. S. Chandana, "MobileNetV3: A deep learning technique for human face expressions identification," Int. J. Inf. Technol., 2023, doi: 10.1007/s41870-023-01380-x.