

ỦY 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 NHE
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

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: 22.968 hình ảnh (64%).
- Tập xác thực (validation): 5.741 hình ảnh (16%).
- Tập kiểm tra: 7.178 hình ảnh (20%).

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).

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.

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:**

- 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.

- **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 tính có CPU Intel Core i5, RAM 16GB. Sử dụng Google Colab để mô phỏng điều kiện tài nguyên thấp. Việc sử dụng CPU thay vì GPU giúp đánh giá thời gian suy luận gần 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.