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

bộ dữ liệu công khai FER-2013.

- 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 8 tuần, từ tháng 4 đến tháng 5 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

2.1 Tổng hợp các tài liệu, nghiên cứu trước liên quan

2.1.1 Nghiên cứu về nhận diện biểu cảm khuôn mặt (FER)

Nhận diện biểu cảm khuôn mặt (Facial Expression Recognition - FER) là một lĩnh vực trọng điểm trong thị giác máy tính. Từ đầu những năm 2000, các phương pháp truyền thống như LBP, HOG, hoặc SIFT kết hợp với SVM từng chiếm ưu thế. Tuy nhiên, chúng không hiệu quả trong điều kiện ánh sáng thay đổi hoặc góc nhìn khác nhau. Từ năm 2014, học sâu - đặc biệt là CNN - đã nâng cao độ chính xác mô hình FER. Các kiến trúc như VGGNet, ResNet, InceptionNet đạt độ chính xác 70–75% trên FER-2013 nhưng yêu cầu tài nguyên tính toán lớn.

2.1.2 Ảnh hưởng của điều kiện ánh sáng yếu

Zhang et al. (2019), Wang et al. (2020) đã chứng minh rằng ảnh thiếu sáng làm giảm hiệu quả của mô hình FER. GAN-based như EnlightenGAN hoặc RetinexNet giúp cải thiện nhưng đòi hỏi GPU mạnh, không phù hợp với thiết bị thực tế như điện thoại hoặc camera nhúng.

2.1.3 Các kỹ thuật tiền xử lý ảnh tăng cường sáng

Gamma Correction: điều chỉnh độ sáng theo hàm $I_{out} = I_{in}^{\gamma}$. Với $\gamma < 1$, ảnh được làm sáng.

CLAHE: nâng cao độ tương phản cục bộ, phù hợp ảnh có vùng sáng tối không đều. Wang et al. (2021) dùng CLAHE trước FER đạt cải thiện đáng kể.

2.1.4 Mô hình học sâu nhẹ: MobileNetV3

MobileNetV3 (Howard et al., 2019) là CNN nhẹ, tối ưu cho thiết bị di động, gồm kỹ thuật như depthwise separable convolution, squeeze-and-excitation và NAS. MobileNetV3-Small có $\tilde{2}$.5M tham số, cân bằng tốt giữa độ chính xác và hiệu suất, chưa được nghiên cứu sâu trong FER ánh sáng yếu.

2.2 Cơ sở lý thuyết của nghiên cứu

Nghiên cứu kết hợp (1) tiền xử lý ảnh thích ứng theo điều kiện ánh sáng và (2) mô hình CNN nhẹ - MobileNetV3 để tăng hiệu suất FER trong điều kiện ánh sáng yếu.

2.2.1 Tiền xử lý ảnh trong điều kiện ánh sáng yếu

Gamma Correction: hàm phi tuyến giúp làm sáng ảnh thiếu sáng. Ying et al. (2017) chỉ ra rằng γ phù hợp có thể nâng cao chất lượng ảnh mà không gây nhiễu.

CLAHE: phân tích cục bộ từng vùng ảnh, cải thiện chi tiết biểu cảm ở vùng mắt, miệng (Zhu et al., 2018).

Tính thích ứng: thuật toán tự động phân tích histogram và độ sáng trung bình để chọn phương pháp phù hợp (Chen et al., 2021).

2.2.2 Nhận diện biểu cảm bằng mô hình CNN nhẹ – MobileNetV3

MobileNetV3-Small (Howard et al., 2019): $\tilde{2}$.5 triệu tham số, thích hợp cho thiết bị nhúng. Nghiên cứu dùng mô hình này để fine-tune phân loại 7 biểu cảm.

Kỹ thuật chính: Depthwise Separable Convolution (Howard et al., 2017), SE Module (Hu et al., 2018), Hard-Swish Activation.

2.2.3 Pipeline đề xuất trong nghiên cứu

Dựa trên hai thành phần lý thuyết đã trình bày, nghiên cứu đề xuất pipeline xử lý gồm 3 giai đoạn chính như trong Bảng 1:

Bảng 1: Pipeline đề xuất trong nghiên cứu

Giai đoạn	Nội dung
Tiền xử lý ảnh	• Chuyển ảnh sang ảnh grayscale. • Tính độ sáng trung bình μ. • Nếu $\mu < T_1$: áp dụng gamma correction với $\gamma \in [0.4, 0.5]$. • Nếu $T_1 < \mu < T_2$: áp dụng CLAHE. • Nếu $\mu > T_2$: giữ nguyên hoặc áp dụng contrast stretching nhẹ.
Học biểu cảm	• Ảnh sau tiền xử lý được đưa vào mô hình MobileNetV3-Small. • Mô hình được fine-tune để phân loại 7 biểu cảm: vui, buồn, giận, sợ, bất ngờ, ghê tởm, trung tính.
Đánh giá mô hình	• Thực hiện trên tập test có và không áp dụng tăng cường ảnh. • Sử dụng các chỉ số đánh giá: – Accuracy – Precision / Recall – F1-score – Confusion Matrix • So sánh với baseline không áp dụng tăng cường để đánh giá hiệu quả thực sự.

2.3 Phân tích điểm mạnh, điểm yếu của các nghiên cứu trước và hướng kế thừa

2.3.1 Điểm mạnh

- Mô hình học sâu giúp tăng độ chính xác FER (Mollahosseini et al., 2016).
- MobileNetV3 hiệu quả, tiết kiệm tài nguyên (Howard et al., 2019).
- CLAHE giúp tăng sáng hiệu quả, đơn giản (Wang et al., 2020).

2.3.2 Hạn chế

- Chưa chú trọng ánh sáng yếu trong FER (Barsoum et al., 2016).
- Pipeline thiếu bước tăng cường ảnh (Zhou et al., 2021).
- Dùng GAN tăng sáng gây tốn tài nguyên (Chen et al., 2020).

2.3.3 Hướng kế thừa và phát triển

- Chọn MobileNetV3-Small làm backbone (Howard et al., 2019).
- Thiết kế pipeline có bước xử lý ảnh thích ứng đầu vào.
- Mô phỏng tập FER-2013 thiếu sáng để kiểm thử.
- Ưu tiên tăng sáng đơn giản thay vì GAN.

2.4 Cơ sở lý thuyết của thuật toán tăng cường dữ liệu thích ứng

2.4.1 Lý do phát triển thuật toán

Trong bài toán nhận diện biểu cảm khuôn mặt (FER – Facial Expression Recognition) dưới điều kiện ánh sáng yếu, hình ảnh khuôn mặt thường bị suy giảm chất lượng nghiêm trọng do hiện tượng thiếu sáng toàn cục hoặc cục bộ. Điều này dẫn đến hiện tượng mất chi tiết, đặc biệt ở các vùng chứa đặc trưng biểu cảm quan trọng như mắt, miệng, nếp nhăn. Kết quả là mô hình học sâu, vốn phụ thuộc vào độ tương phản và cấu trúc cục bộ, sẽ khó khăn trong việc nhận dạng chính xác.

Các kỹ thuật tăng cường dữ liệu truyền thống như histogram equalization hoặc gamma correction thường được áp dụng đồng loạt cho toàn bộ dữ liệu huấn luyện. Tuy nhiên, cách tiếp cận này bỏ qua tính biến thiên về mức sáng của từng ảnh đầu vào. Cụ thể:

- Với ảnh quá tối, tăng sáng quá mức dễ làm mất chi tiết do bão hòa điểm ảnh.
- Với ảnh sáng vừa đủ, tăng cường không cần thiết có thể làm biến dạng đặc trưng tự nhiên, dẫn đến suy giảm hiệu quả học.

Do đó, nghiên cứu này đề xuất một thuật toán tăng cường dữ liệu thích ứng, có khả năng phân tích đặc trưng ánh sáng riêng của từng ảnh, từ đó lựa chọn kỹ thuật xử lý phù hợp, đơn giản nhưng hiệu quả và phù hợp để huấn luyện với mô hình nhẹ như MobileNetV3-Small.

2.4.2 Các thành phần lý thuyết chính

(a) Phân tích độ sáng của ảnh

Để xác định ảnh đầu vào có cần tăng cường hay không, và nếu cần thì sử dụng phương pháp nào, cần phân tích một số đặc trưng cơ bản về độ sáng:

- Độ sáng trung bình (mean intensity): Được tính trên ảnh chuyển sang thang xám (grayscale) hoặc kênh Y (luminance) trong không gian YUV.

$$\mu = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} I(i,j)$$

- Độ lệch chuẩn (standard deviation): Đánh giá mức độ phân tán sáng tối, cho biết ảnh có sáng đồng đều hay có vùng sáng – vùng tối xen kẽ.
- Histogram phân bố pixel: Dùng để xác định ảnh có độ tương phản thấp (hẹp histogram) hoặc bị lệch về vùng tối.

(b) Các kỹ thuật tăng cường ánh sáng được sử dụng

- Gamma Correction:

$$I_{\text{out}} = I_{\text{in}}^{\gamma}$$

  - $\gamma < 1$: ảnh được làm sáng lên.
  - $\gamma > 1$: ảnh bị làm tối hơn.

  Việc chọn giá trị γ được tính toán dựa trên giá trị độ sáng trung bình μ của ảnh.

- Histogram Equalization (HE): Phân bố lại giá trị pixel để làm tăng độ tương phản tổng thể. Phù hợp khi histogram bị tập trung ở vùng tối (low dynamic range). Tuy nhiên, dễ gây nhiễu ở ảnh có noise.
- Contrast Stretching: Kéo dãn mức độ sáng từ dải cường độ cũ về dải chuẩn 0–255:

$$I_{\text{out}} = \frac{I_{\text{in}} - I_{\text{min}}}{I_{\text{max}} - I_{\text{min}}} \times 255$$

(c) Tính thích ứng của thuật toán

Thuật toán sẽ:

- Tính toán độ sáng trung bình (μ) và độ lệch chuẩn (σ) của từng ảnh đầu vào.
- Dựa vào hai ngưỡng xác định trước T_1 và T_2, phân loại mức độ ánh sáng:
  - $\mu < T_1$ (ảnh rất tối): áp dụng gamma nhỏ (0.3–0.5).
  - $T_1 \leq \mu < T_2$ (tối vừa): áp dụng gamma nhẹ (0.7–0.9) hoặc HE.
  - $\mu \geq T_2$ (sáng đủ): không tăng cường hoặc chỉ contrast stretching nhẹ.

Cách tiếp cận này giúp mỗi ảnh được tăng cường đúng mức, tránh làm hỏng đặc trưng gốc hoặc gây dư sáng.

2.4.3 Nguồn cảm hứng và các nghiên cứu liên quan

Retinex-based methods (Fu et al., 2020) đề xuất kỹ thuật phân tách ảnh thành hai thành phần: phản xạ và ánh sáng chiếu vào, sau đó tái cấu trúc lại ảnh với độ sáng cải thiện. Phương pháp này cho kết quả nâng cao rõ rệt nhưng đòi hỏi thuật toán phức tạp và tài nguyên tính toán lớn, do đó khó triển khai trên các thiết bị nhúng.

GAN-based methods như EnlightenGAN (Jiang et al., 2019) sử dụng mạng sinh ảnh để tạo lại phiên bản ảnh có ánh sáng tốt hơn từ ảnh thiếu sáng ban đầu. Mặc dù đem lại chất lượng thị giác cao, nhưng các mô hình GAN thường yêu cầu GPU mạnh và thời gian xử lý lâu, khiến chúng không phù hợp với các ứng dụng thời gian thực trên thiết bị di động.

Adaptive Augmentation trong học sâu (Zhang et al., 2021) nhấn mạnh tầm quan trọng của việc sử dụng đặc trưng đầu vào để quyết định chiến lược tăng cường dữ liệu phù hợp, thay vì áp dụng cố định một kỹ thuật như truyền thống. Điều này giúp mô hình học sâu đạt hiệu quả tốt hơn trong môi trường đầu vào đa dạng.

Từ các nghiên cứu trên, thuật toán của nhóm đề xuất kế thừa ý tưởng adaptive preprocessing, nhưng được đơn giản hóa để giảm chi phí tính toán và đảm bảo tính linh hoạt, phù hợp với các mô hình nhẹ như MobileNetV3.

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

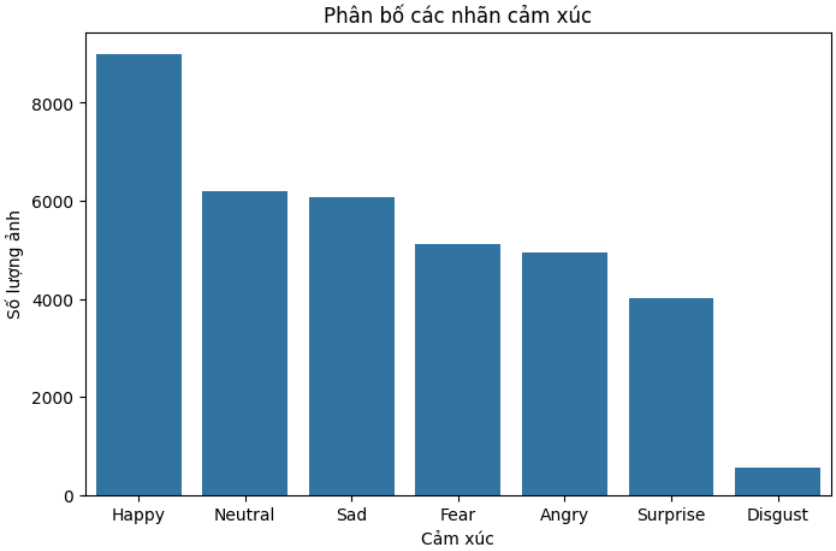
Dữ liệu được thu thập từ tập dữ liệu FER-2013 công khai trên nền tảng Kaggle. Các bước gồm:

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.

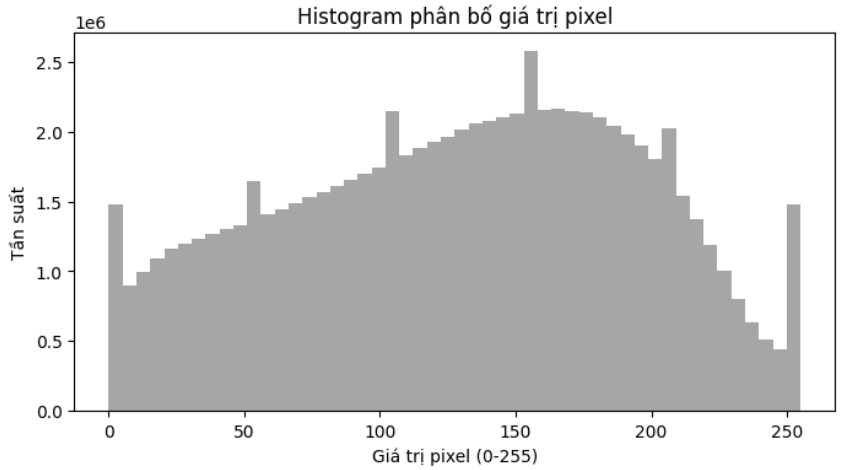


Hình 1: Phân bố nhãn cảm xúc trong tập dữ liệu FER-2013.

Ý nghĩa:

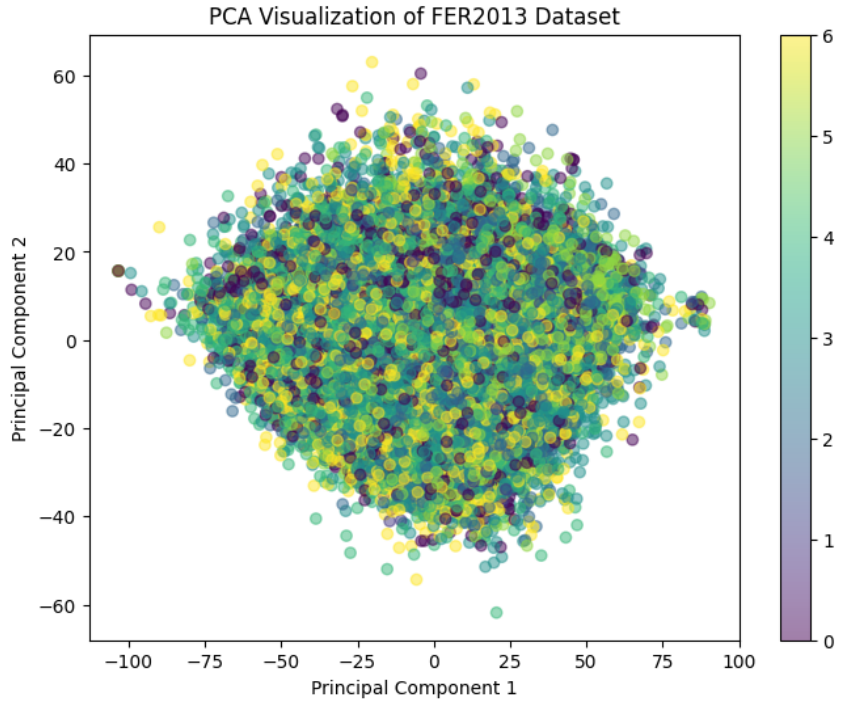
Hình 2: Một số hình ảnh mẫu từ tập dữ liệu FER-2013.

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.