

**TRƯỜNG ĐẠI HỌC SÀI GÒN**  
**KHOA CÔNG NGHỆ THÔNG TIN**



**HỌC PHẦN: NCKH TRONG**  
**CNTT**

**PHÂN TÍCH LUẬN VĂN TỐT NGHIỆP**

**Nhóm sinh viên thực hiện:**

| <b>Họ và tên</b>   | <b>MSSV</b> |
|--------------------|-------------|
| 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

## Phân công công việc

| STT | MSSV       | Họ và Tên          | Phân Công                                   | Thái Độ     |
|-----|------------|--------------------|---|-------------|
| 1   | 3122410202 | Văn Tuấn Kiệt      | Các bài toán liên quan                      | Rất tốt     |
| 2   | 3122410207 | Mai Phúc Lâm       | Nguồn gốc và vị trí của tập dữ liệu         | Tích cực    |
| 3   | 3122410208 | Nguyễn Đức Duy Lâm | Kết quả đạt được, độ đo, khảo sát, kết luận | Nhiệt tình  |
| 4   | 3122410213 | Nguyễn Hữu Lộc     | Tóm tắt, giới thiệu, nguồn gốc dữ liệu      | Trách nhiệm |

Bảng 1: Phân công công việc nhóm

## Mục lục

|          |  |          |
|----------|--|----------|
| <b>1</b> | <b>Giới thiệu</b>                              | <b>3</b> |
| <b>2</b> | <b>Mục tiêu nghiên cứu</b>                     | <b>3</b> |
| <b>3</b> | <b>Cơ sở lý thuyết</b>                         | <b>3</b> |
| <b>4</b> | <b>Phương pháp thu thập và xử lý thông tin</b> | <b>4</b> |
| <b>5</b> | <b>Kết quả đạt được</b>                        | <b>4</b> |
| 5.1      | Trên tập FER2013 . . . . .                     | 4        |
| 5.2      | Trên tập VEMO . . . . .                        | 4        |
| 5.3      | Trực quan hoá GradCAM . . . . .                | 4        |
| <b>6</b> | <b>Hạn chế công trình</b>                      | <b>5</b> |

# Phân tích luận văn tốt nghiệp

Tác giả: Văn Tuấn Kiệt<sup>1</sup>, Mai Phúc Lâm<sup>1</sup>, Nguyễn Đức Duy Lâm<sup>1</sup>, Nguyễn Hữu Lộc<sup>1</sup>

<sup>1</sup>Trường Đại học Sài Gòn, Ngày: 17 tháng 4 năm 2025

## 1 Giới thiệu

- **Tên đề án:** Nhận diện cảm xúc mặt người sử dụng mạng học sâu có chú ý
- **Tác giả:** Phạm Quý Luân
- **Nơi công bố:** Trường Đại học Bách Khoa, Đại học Quốc gia Thành phố Hồ Chí Minh, Khoa Khoa học - Kỹ thuật Máy tính
- **Năm công bố:** THÁNG 12/2019

## 2 Mục tiêu nghiên cứu

Mục tiêu của luận văn là phát triển và đánh giá một phương pháp nhận diện cảm xúc trên khuôn mặt người sử dụng mạng học sâu tích hợp cơ chế chú ý (Residual Masking Network) để phân loại cảm xúc từ hình ảnh đầu vào trong môi trường phức tạp, đồng thời sử dụng phương pháp học kết hợp nhiều mô hình hiện đại nhằm nâng cao độ chính xác.

## 3 Cơ sở lý thuyết

- Luận văn dựa trên các nghiên cứu về nhận diện cảm xúc khuôn mặt, một lĩnh vực đã được phát triển nhiều năm với các lợi ích trong nhiều ứng dụng thực tiễn.
- Sử dụng mạng nơ-ron tích chập (CNN) và cơ chế chú ý để tập trung vào các đặc trưng quan trọng trên khuôn mặt liên quan đến cảm xúc (theo Hệ thống mã hóa hành động khuôn mặt - FACS).
- Các mô hình hiện đại như VGG19, ResNet, DenseNet, GoogLeNet, Inception v3,
- Dữ liệu được lấy từ các tập dữ liệu chuẩn như FER2013 và VEMO, chứa các hình ảnh khuôn mặt với nhãn cảm xúc cơ bản (giận dữ, ghê tởm, sợ hãi, hạnh phúc, buồn bã, ngạc nhiên, trung lập).

## 4 Phương pháp thu thập và xử lý thông tin

- Sử dụng tập dữ liệu FER2013 và VEMO. Các hình ảnh được xử lý để phát hiện khuôn mặt bằng phương pháp của OpenCV 3.4 nếu chưa được xử lý trước.
- Áp dụng hai phép biến đổi chính: lật ảnh theo trục dọc (Flip Left Right) và xoay ảnh từ -30 đến 30 độ, sử dụng thư viện `imgaug`.
- Điều chỉnh kích thước ảnh về 224x224 pixel, chuyển đổi sang tensor, và nhân bản thành ba kênh màu
- Sử dụng lớp Dataset trong PyTorch để tải và quản lý dữ liệu, với hàm `getitem` để truy xuất ảnh và nhãn.
- Sử dụng mạng Residual Masking Network (ResMaskingNet) với cơ chế chú ý để phân loại cảm xúc.
- Sử dụng GradCAM và gộp theo chiều kênh (Average Pooling) để trực quan hóa các vùng chú ý của mô hình trên khuôn mặt.

## 5 Kết quả đạt được

### 5.1 Trên tập FER2013

- Mạng ResMaskingNet đạt độ chính xác 74.14%, vượt qua nhiều mô hình hiện đại như VGG19 (70.8%), ResNet18 (72.9%), DenseNet121 (73.16%), và CBAM\_ResNet50 (73.39%).
- Khi kết hợp ResMaskingNet với 6 mô hình CNN khác, độ chính xác đạt 76.82%, là kết quả tốt nhất so với các phương pháp được báo cáo khoa học (ví dụ: Ensemble 8 CNNs đạt 75.2%).

### 5.2 Trên tập VEMO

- ResMaskingNet đạt độ chính xác 65.94%, vượt qua ResNet18 (63.94%), ResNet34 (64.84%), và ResAttNet56 (60.82%).

### 5.3 Trực quan hoá GradCAM

- Chúng đã tập trung vào vùng mặt người, có chú ý vào các bộ phận có ảnh hưởng trực tiếp đến cảm xúc như được miêu tả trong FACS
- Mô hình được thử nghiệm trên các hình ảnh thực tế (ví dụ: ảnh từ phim "Mất Biếc") và cho kết quả dự đoán chính xác cảm xúc như hạnh phúc.

## 6 Hạn chế công trình

- Chỉ sử dụng 2 phương pháp tiền xử lí dữ liệu
- Khi trực quan hoá bằng gộp kênh thì chúng không thể hiện rõ ràng sự khác nhau giữa các đặc trưng cho ta thấy
- Qua ma trận nhầm lẫn , việc dự đoán cảm xúc trên các tập dữ liệu dễ bị nhầm lẫn fear , disgust