

**BÁO CÁO KẾT QUẢ XÂY DỰNG CHƯƠNG TRÌNH THỰC NGHIỆM  
ĐỀ TÀI NHẬN DIỆN CẢM XÚC KHUÔN MẶT TRÊN BỘ DỮ LIỆU FER2013**

**\* Em xem ảnh nào bị sai, xử lý theo 2 hướng thử xem:**

**1. Lấy đem cho vào tập train**

**2. Bỏ thử ra khỏi dữ liệu**

**=> Em đã thử theo hướng thứ 1, cụ thể:**

Đối với bộ dữ liệu bộ dữ liệu PublicTest mô hình nhận nhầm **937** ảnh; PrivateTest mô hình nhận nhầm **898** ảnh.

-> Em đã đưa **937 + 898** ảnh trên **bổ sung** vào tập dữ liệu huấn luyện; đồng thời, đưa **937 + 898** ảnh đầu tiên của bộ dữ liệu huấn luyện ra 02 bộ dữ liệu test tương ứng để đảm bảo số lượng ảnh huấn luyện và ảnh kiểm tra vẫn được giữ nguyên với bộ dữ liệu gốc:

- **27.765** ảnh thuộc tập training
- **3.457** ảnh thuộc tập PublicTest
- **3.472** ảnh thuộc tập PrivateTest

Kết quả sau khi huấn luyện và áp dụng phương pháp tính trung bình có trọng số thu được độ chính xác cao nhất đạt **93.434%**, cụ thể:

- Bảng trọng số:

| Bộ dữ liệu                                  | Mắt trái | Mắt phải | Má trái | Má phải | Miệng  | Mũi    | Trán   | Cả mặt        |
|---|----------|----------|---------|---------|--------|--------|--------|---------------|
| Public (test-<br>bộ test cho<br>huấn luyện) | 65.722   | 66.387   | 73.706  | 75.991  | 73.966 | 67.949 | 57.044 | <b>84.437</b> |
| Private<br>(validation -<br>bộ kiểm thử)    | 67.079   | 66.561   | 74.856  | 75.403  | 73.041 | 68.404 | 58.237 | <b>85.369</b> |

- Độ chính xác khi kết hợp các phần với mô hình VggNet sử dụng trọng số:

| Bộ dữ liệu                              | Public (test- bộ test cho huấn luyện) | Private (validation - bộ kiểm thử) |
|---|---------------------------------------|------------------------------------|
| Độ chính xác trên 7 phần + cả khuôn mặt | <b>93.434</b>                         | <b>93.433</b>                      |

## KẾT QUẢ TRƯỚC ĐÓ

### 1. Tách các phần trên khuôn mặt

Để phát hiện và tách khuôn mặt thành 7 phần (mắt trái, mắt phải, mũi, miệng, má trái, má phải, trán), học viên đã sử dụng thư viện *face\_alignment* để phát hiện **68 key points** từ đó phân tách thành các vùng bộ phận trên khuôn mặt:

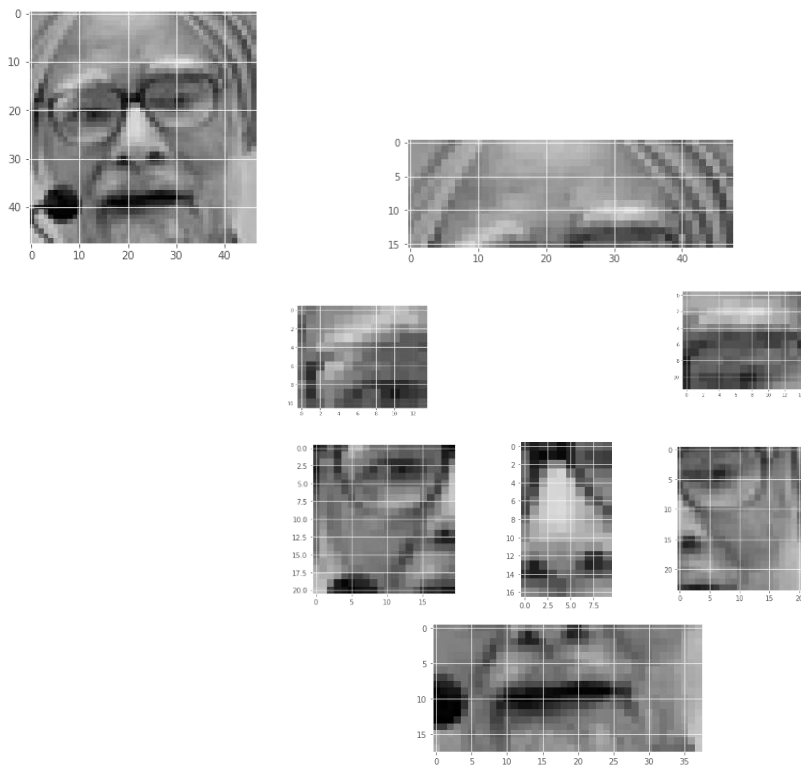
—> mức độ phát hiện và phân chia các vùng có chính xác 100% ko?

- Hệ thống không phát hiện được 1.193 trên 32.298 ảnh mặt (chiếm 3,6%) do khuôn mặt bị khuất hoặc chỉ chứa 1 phần khuôn mặt. Phương pháp này hiệu quả hơn sử dụng thư viện *mtcnn* (không phát hiện được hơn 22.000 ảnh) chỉ phát hiện được 5 key points (2 mắt, đỉnh mũi, 2 khóe miệng).

→ khuôn mặt nào ko phát hiện được vì lý do đã nêu, loại bỏ khỏi dữ liệu luôn để hiệu suất phát hiện khuôn mặt lên 100%, những khuôn mặt bị thiếu ko đủ để nhận dạng cảm xúc cũng bỏ đi luôn, chỉ giữ lại khuôn mặt đủ thành phần, sau đó đánh giá độ chính xác lại

-> Sau khi kiểm tra lại và làm sạch dữ liệu và kiểm thử trên một số tỉ lệ trọng số các phần khác nhau, độ chính xác cao nhất của hệ thống đã tăng thành **74.136%** (kết quả trước đó đạt 73.4% thu được là được tính toán trên bộ dữ liệu em đã làm sạch ạ)

- Phân chia các vùng chính xác 100%, một số mẫu cụ thể như sau:



## 2. Kết quả thực nghiệm

### 2.1. Em có thử điều chỉnh các trọng số chưa?

\* Trước khi điều chỉnh các trọng số, độ chính xác tốt nhất đạt **73.675%** trên bộ PrivateTest và **72.057%** trên bộ PublicTest.

=> Sau khi điều chỉnh trọng số, độ chính xác tốt nhất đạt **74.136%** trên bộ PrivateTest và **73.212%** trên bộ PublicTest.

Cụ thể:

- Độ chính xác trên các phần và cả khuôn mặt (bảng trọng số):

| Bộ dữ liệu                            | Mắt trái | Mắt phải | Má trái | Má phải | Miệng  | Mũi    | Trán   | Cả mặt |
|---------------------------------------|----------|----------|---------|---------|--------|--------|--------|--------|
| Public (test- bộ test cho huấn luyện) | 58.345   | 57.883   | 65.259  | 66.156  | 64.304 | 60.775 | 51.085 | 71.71  |
| Private (validation - bộ kiểm thử)    | 58.497   | 59.044   | 67.166  | 65.985  | 66.187 | 60.225 | 52.65  | 73.416 |

- Độ chính xác khi kết hợp các phần:

| Bộ dữ liệu                            | Kết quả trên 7 phần | Kết quả trên 7 phần + cả khuôn mặt |
|---------------------------------------|---------------------|------------------------------------|
| <b>Mô hình VggNet</b>                 |                     |                                    |
| <b>Không sử dụng trọng số</b>         |                     |                                    |
| Public (test- bộ test cho huấn luyện) | 69.540              | 71.507                             |
| Private (validation - bộ kiểm thử)    | 70.622              | 73.041                             |
| <b>Sử dụng trọng số</b>               |                     |                                    |
| Public (test- bộ test cho huấn luyện) | 69.800              | 72.346                             |

|  |        |               |
|--|--------|---------------|
| Private<br>(validation -<br>bộ kiểm thử) | 70.795 | <b>74.136</b> |
|--|--------|---------------|

| <b>Mô hình VGG19</b>                        |        |        |
|---|--------|--------|
| <b>Không sử dụng trọng số</b>               |        |        |
| Public (test-<br>bộ test cho<br>huấn luyện) | 70.061 | 71.478 |
| Private<br>(validation -<br>bộ kiểm thử)    | 70.334 | 72.437 |
| <b>Sử dụng trọng số</b>                     |        |        |
| Public (test-<br>bộ test cho<br>huấn luyện) | 70.466 | 71.796 |
| Private<br>(validation -<br>bộ kiểm thử)    | 70.622 | 72.408 |

***\* Sau khi điều chỉnh trọng số, độ chính xác tối đa đạt: 74.136%***

Cụ thể: Nhận thấy phần: miệng và toàn bộ khuôn mặt ảnh hưởng lớn nhất tới kết quả nhận diện; trong đó, miệng là thành phần thể hiện cảm xúc rõ nhất, do đó, học viên đã tăng trọng số:

- Miệng tăng **5** lần
- Toàn bộ khuôn mặt tăng **10** lần

Qua đó đã thu được kết quả độ chính xác trung bình tăng lên **74.136%**

=> Việc sử dụng phương pháp trung bình có trọng số thu được hiệu quả tốt, tăng độ chính xác của hệ thống.

## 2.2. Em thử nghiệm trên bộ dữ liệu khác:

**- Chọn bộ khó và bộ dễ (ko huấn luyện, mà dùng kết quả đã huấn luyện để test)**

Học viên đã kiểm tra trên bộ dữ liệu CK+48 gồm 1.181 ảnh của 07 loại cảm xúc. Sau khi tách 07 phần của mắt, mũi, miệng... học viên đã kiểm tra độ chính xác trên 02 phương pháp:

**\* Kiểm tra việc phân lớp các ảnh của bộ CK+48 dựa trên bộ trọng số thu được từ việc huấn luyện trên bộ FER2013:**

Học viên sử dụng bộ trọng số từ bộ dữ liệu Public data của FER2013:

| Bộ dữ liệu                            | Mắt trái | Mắt phải | Má trái | Má phải | Miệng  | Mũi    | Trán   | Cả mặt |
|---------------------------------------|----------|----------|---------|---------|--------|--------|--------|--------|
| Public (test- bộ test cho huấn luyện) | 58.345   | 57.883   | 65.259  | 66.156  | 64.304 | 60.775 | 51.085 | 71.71  |

Kết quả kiểm tra trên bộ CK+48 đạt độ chính xác là **63.985%**.

**\* Huấn luyện lại từ đầu nhưng cùng kiến trúc mô hình:**

Chia bộ dữ liệu CK+48 thành 02 phần, gồm 80% ảnh mỗi lớp cảm xúc cho huấn luyện và 20% ảnh mỗi lớp cảm xúc cho kiểm thử.

Bộ trọng số CK+48 thu được khi huấn luyện trên từng phần (250 epochs, patient=40):

| Bộ dữ liệu                            | Mắt trái | Mắt phải | Má trái | Má phải | Miệng | Mũi   | Trán  | Cả mặt |
|---------------------------------------|----------|----------|---------|---------|-------|-------|-------|--------|
| Public (test- bộ test cho huấn luyện) | 44.44    | 48.48    | 63.63   | 68.18   | 75.76 | 56.57 | 45.96 | 85.35  |

Kết quả: Qua thực nghiệm thay đổi các tỉ lệ tăng trọng số khác nhau, học viên thu được độ chính xác tối đa (khi áp dụng 7 phần + cả mặt) đạt **86.364%**, trong đó, trọng số theo tỉ lệ: má trái = má phải = **1.5** lần; miệng = **4** lần; cả mặt = **6** lần, các bộ phận còn lại = 1 lần.

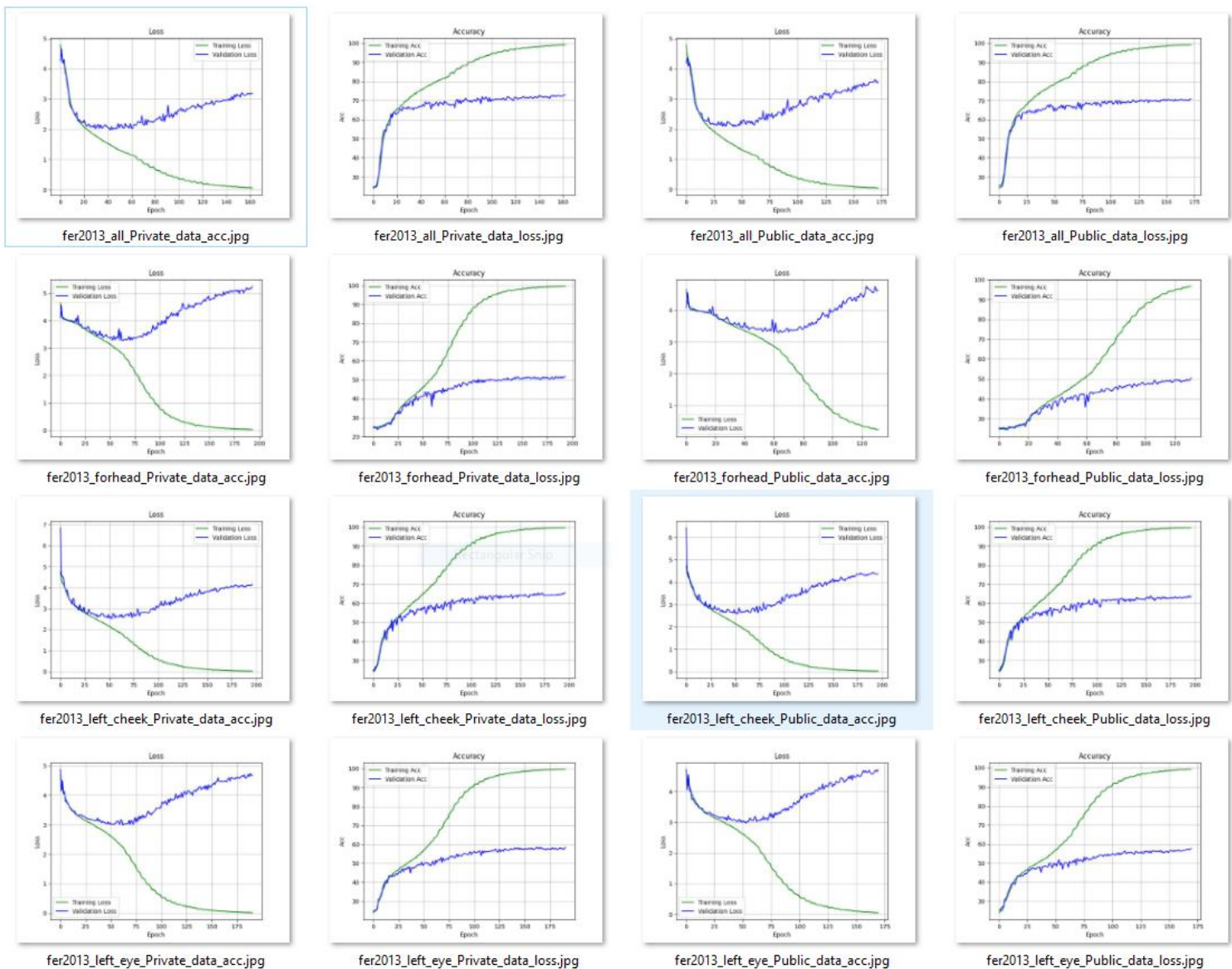
Nhận xét:

- Kết quả thu được là **xếp thứ 4** về độ chính xác trong nhận diện cảm xúc khuôn mặt trên bộ dữ liệu Fer2013 (theo *paperwithcode*) (SATO).
- Bộ dữ liệu Fer2013 bộc lộ hạn chế do **độ phân giải** không cao dẫn đến các đặc trưng thu được khi chia khuôn mặt thành các phần (mắt, mũi, trán...) hạn chế, chưa có hiệu quả cao trong phân lớp cảm xúc tổng thể.
- Phương pháp sử dụng trọng số đã làm tăng độ chính xác, là đóng góp lớn của đề tài.

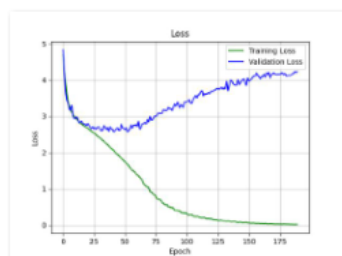
### 3. Huấn luyện trên bộ dữ liệu của các phần và toàn thể khuôn mặt

Học viên tiến hành huấn luyện trên bộ dữ liệu 7 vùng + 1 khuôn mặt với 02 kiến trúc mạng học sâu là: VGG19 và VggNet

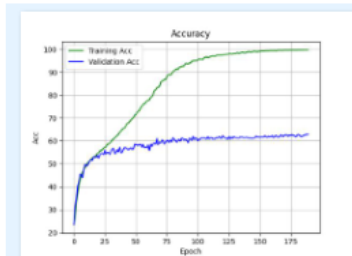
1.1. Trên kiến trúc VGG19 (Hình biểu đồ Loss và Accuracy của các phần khi huấn



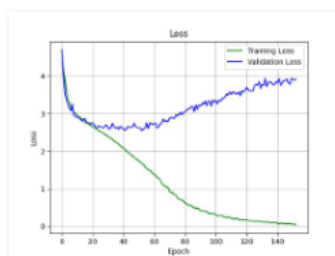
luyện)



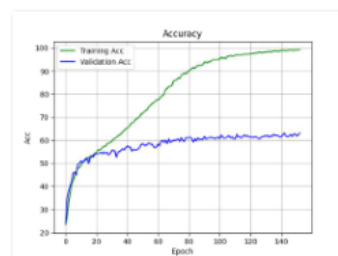
fer2013\_mouth\_Private\_data\_acc.jpg



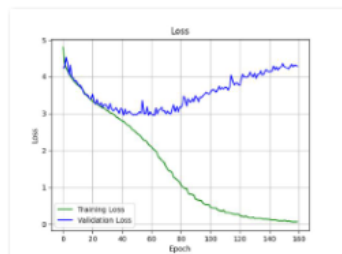
fer2013\_mouth\_Private\_data\_loss.jpg



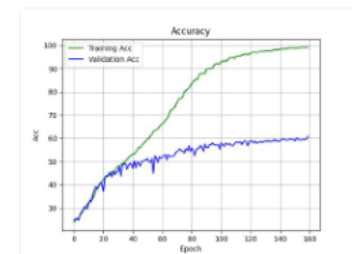
fer2013\_mouth\_Public\_data\_acc.jpg



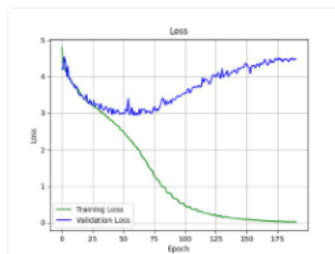
fer2013\_mouth\_Public\_data\_loss.jpg



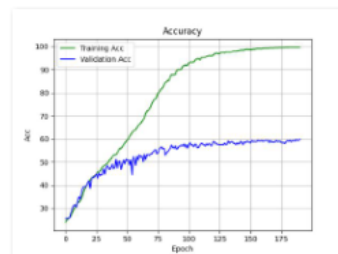
fer2013\_nose\_Private\_data\_acc.jpg



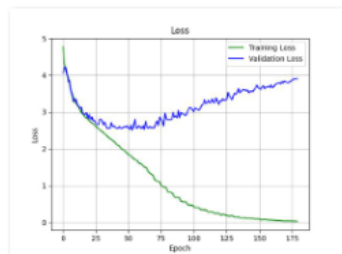
fer2013\_nose\_Private\_data\_loss.jpg



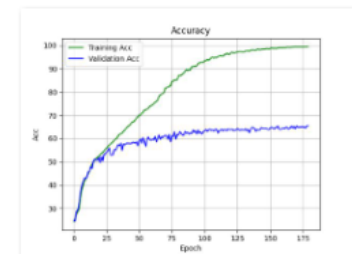
fer2013\_nose\_Public\_data\_acc.jpg



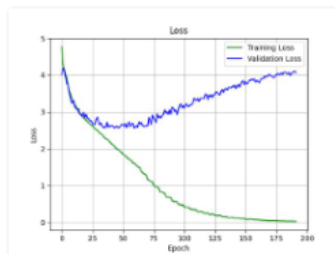
fer2013\_nose\_Public\_data\_loss.jpg



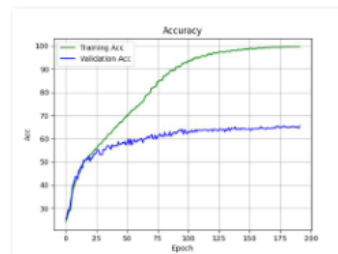
fer2013\_right\_cheek\_Private\_data\_acc.jpg



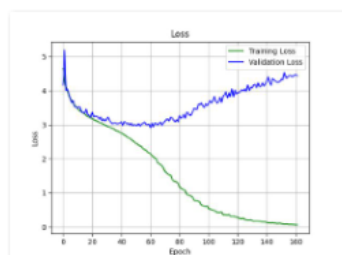
fer2013\_right\_cheek\_Private\_data\_loss.jpg



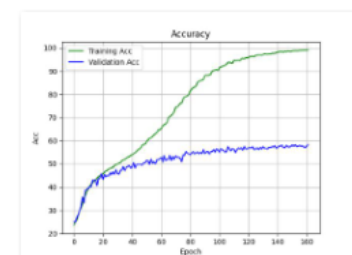
fer2013\_right\_cheek\_Public\_data\_acc.jpg



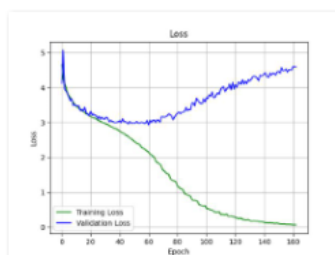
fer2013\_right\_cheek\_Public\_data\_loss.jpg



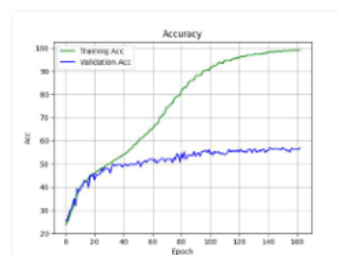
fer2013\_right\_eye\_Private\_data\_acc.jpg



fer2013\_right\_eye\_Private\_data\_loss.jpg



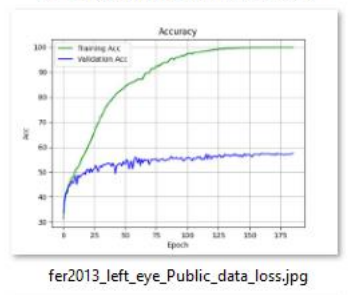
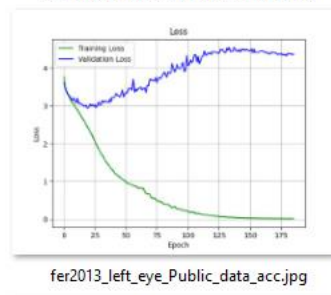
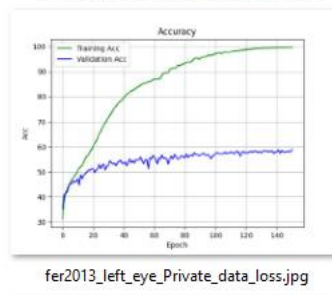
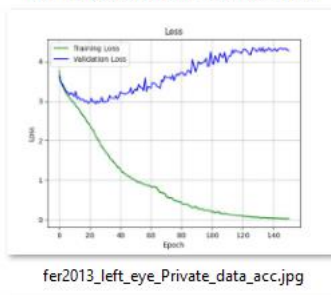
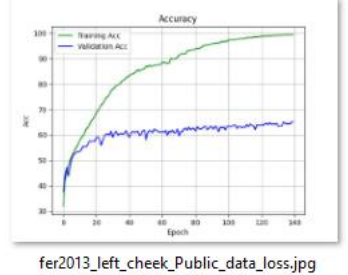
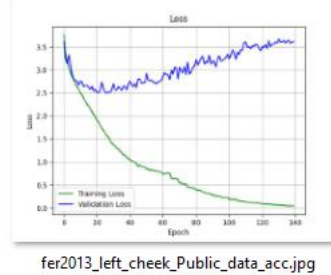
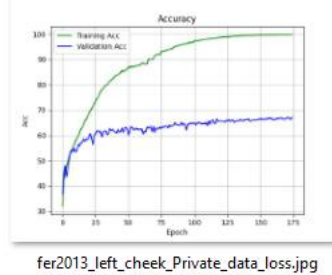
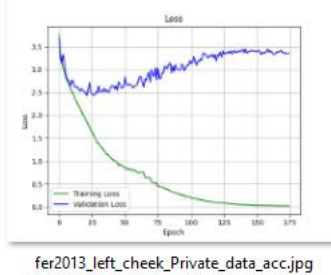
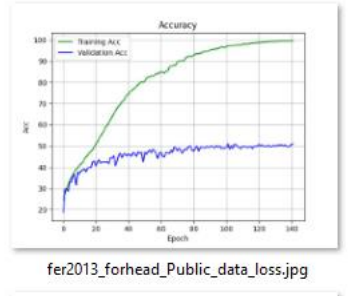
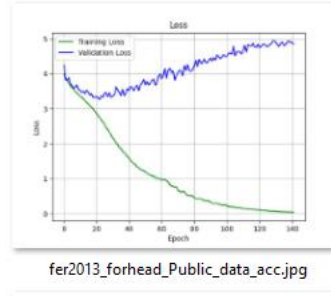
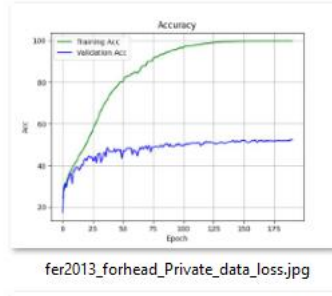
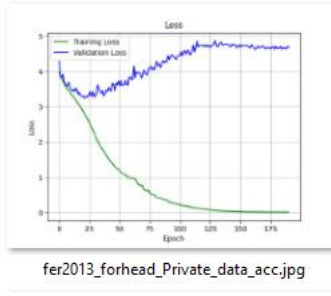
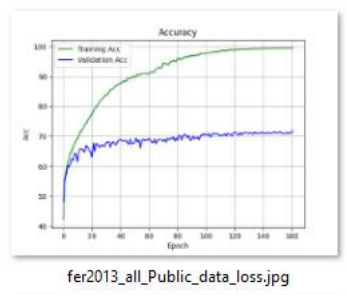
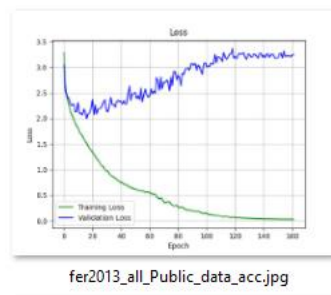
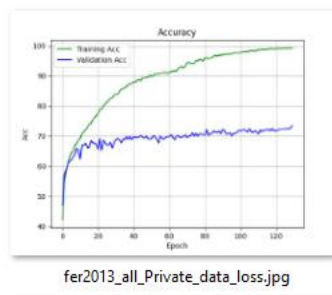
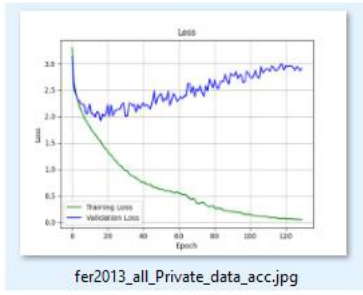
fer2013\_right\_eye\_Public\_data\_acc.jpg



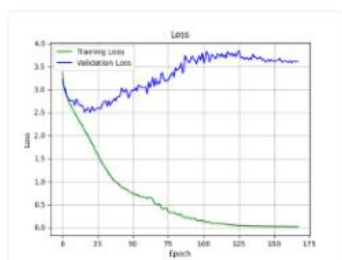
fer2013\_right\_eye\_Public\_data\_loss.jpg



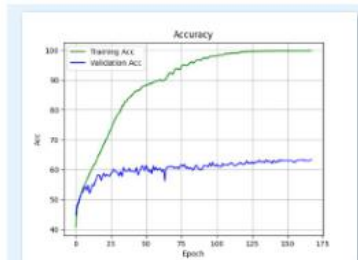
## 1.2. Trên kiến trúc VggNet



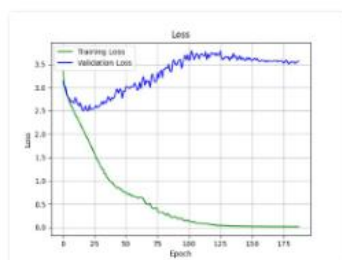




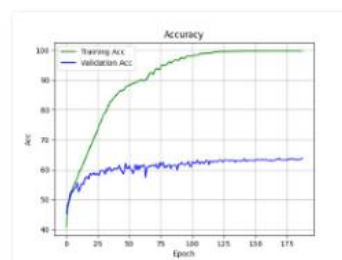
fer2013\_mouth\_Private\_data\_acc.jpg



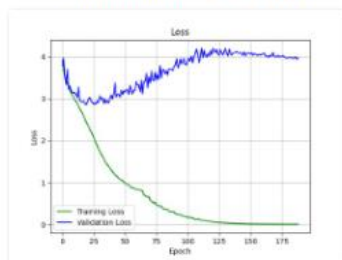
fer2013\_mouth\_Private\_data\_loss.jpg



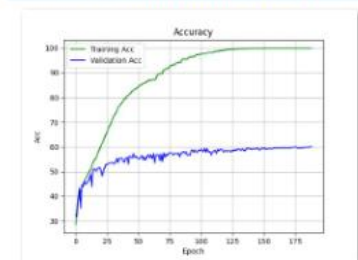
fer2013\_mouth\_Public\_data\_acc.jpg



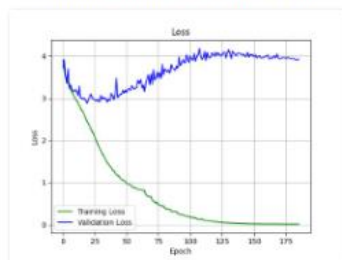
fer2013\_mouth\_Public\_data\_loss.jpg



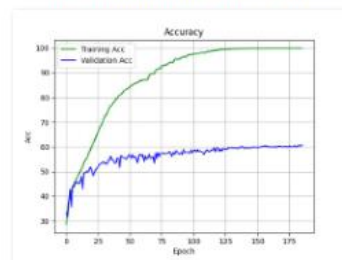
fer2013\_nose\_Private\_data\_acc.jpg



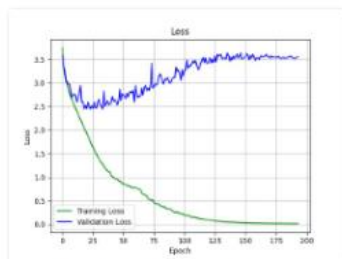
fer2013\_nose\_Private\_data\_loss.jpg



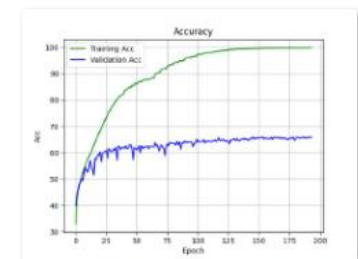
fer2013\_nose\_Public\_data\_acc.jpg



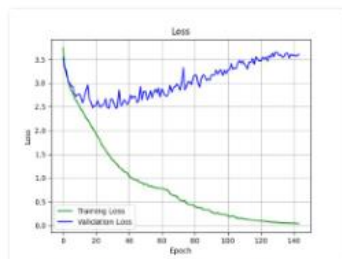
fer2013\_nose\_Public\_data\_loss.jpg



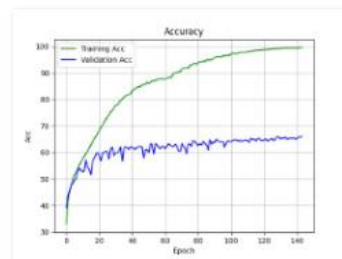
fer2013\_right\_cheek\_Private\_data\_acc.jpg



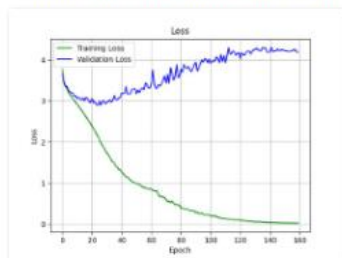
fer2013\_right\_cheek\_Private\_data\_loss.jpg



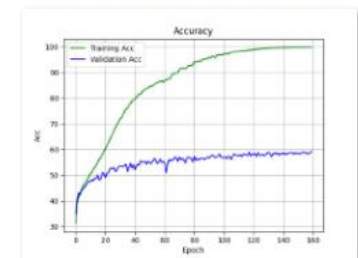
fer2013\_right\_cheek\_Public\_data\_acc.jpg



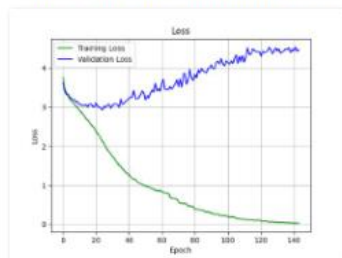
fer2013\_right\_cheek\_Public\_data\_loss.jpg



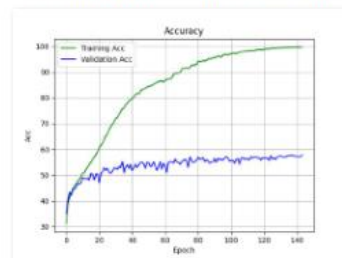
fer2013\_right\_eye\_Private\_data\_acc.jpg



fer2013\_right\_eye\_Private\_data\_loss.jpg



fer2013\_right\_eye\_Public\_data\_acc.jpg



fer2013\_right\_eye\_Public\_data\_loss.jpg