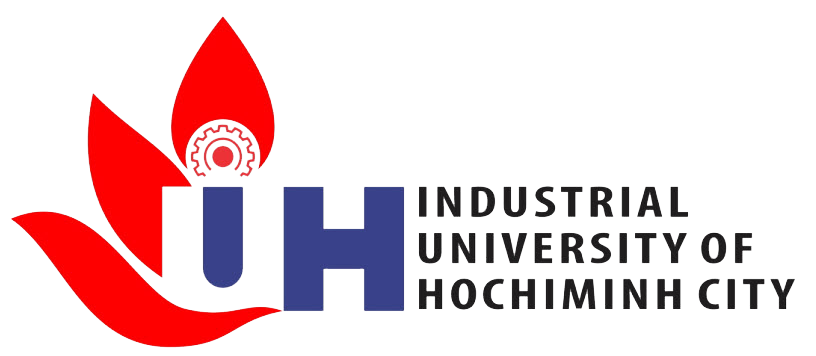
**TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TPHCM**

**KHOA CÔNG NGHỆ THÔNG TIN**



**ĐỒ ÁN MÔN HỌC**

**TÊN ĐỀ TÀI**

**Điểm danh sinh viên bằng mô hình học sâu mạng CNN**

**GIẢNG VIÊN HƯỚNG DẪN: Võ Quang Hoàng Khang**

**SINH VIÊN THỰC HIỆN 1: Nguyễn Hùng Anh**

**MÃ SỐ SINH VIÊN: 21064051**

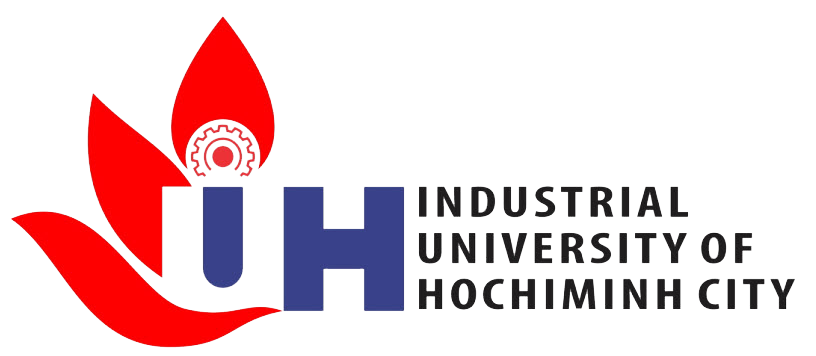
**SINH VIÊN THỰC HIỆN 2: Trương Hải Anh Thắng**

**MÃ SỐ SINH VIÊN: 21062461**

*TP HỒ CHÍ MINH, Tháng 04 năm 2025*

**TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TPHCM**

**KHOA CÔNG NGHỆ THÔNG TIN**



**ĐỒ ÁN MÔN HỌC**

**TÊN ĐỀ TÀI**

**Điểm danh sinh viên bằng mô hình học sâu mạng CNN**

**GIẢNG VIÊN HƯỚNG DẪN: Võ Quang Hoàng Khang**

**SINH VIÊN THỰC HIỆN 1: Nguyễn Hùng Anh**

**MÃ SỐ SINH VIÊN: 21064051**

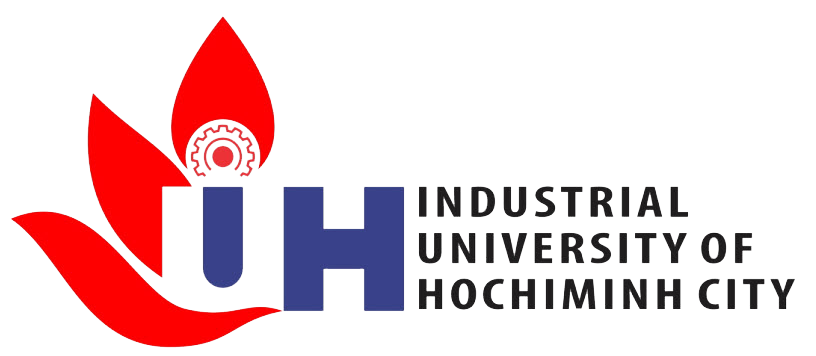
**SINH VIÊN THỰC HIỆN 2: Trương Hải Anh Thắng**

**MÃ SỐ SINH VIÊN: 21062461**

*TP HỒ CHÍ MINH, Tháng 04 năm 2025*

**INDUSTRIAL UNIVERSITY OF HO CHI MINH CITY**

**FACULTY OF INFORMATION TECHNOLOGY**



**GRADUATION THESIS**

**Student attendance with CNN's deep learning model**

**SUPERVISOR: MSc. Vo Quang Hoang Khang**

**STUDENT NAME: Nguyen Hung Anh**

**STUDENT CODE: 21064051**

**STUDENT NAME: Truong Hai Anh Thang**

**STUDENT CODE: 21062461**

*HO CHI MINH CITY, Month 04 year 2025*

**TÓM TẮT**

Hệ thống này nhằm nhận diện khuôn mặt theo thời gian thực để phục vụ cho việc điểm danh sinh viên tự động, sử dụng mô hình học sâu kết hợp với webcam và bộ phát hiện khuôn mặt Haar Cascade. Mô hình được sử dụng trên tập dữ liệu thu thập từ các sinh viên ở trường Đại Học Công Nghiệp, sử dụng các chỉ số như Accuracy, precision, recall, f1-score, support. Kết quả cho thấy mô hình đạt độ chính xác cao, hứa hẹn khả năng ứng dụng trong phục vụ cho việc điểm danh sinh viên tự động tại trường.

**LỜI CẢM ƠN**

Trong suốt quá trình thực hiện đồ án tốt nghiệp, chúng em đã nhận được rất nhiều sự quan tâm, hướng dẫn và hỗ trợ quý báu từ các thầy cô, bạn bè và gia đình. Nhờ đó, chúng em đã có thể hoàn thành khoá luận này một cách trọn vẹn.

Trước hết, chúng em xin bày tỏ lòng biết ơn sâu sắc đến Ban Giám hiệu Trường Đại học Công nghiệp Thành phố Hồ Chí Minh cùng Ban Chủ nhiệm Khoa Công nghệ Thông tin đã tạo điều kiện thuận lợi và môi trường học tập tốt nhất cho chúng em trong suốt thời gian học tại trường.

Đặc biệt, chúng em xin gửi lời cảm ơn chân thành đến Thầy ThS. Võ Quang Hoàng Khang – người đã tận tình hướng dẫn, truyền đạt kiến thức, định hướng và hỗ trợ chúng em trong suốt quá trình thực hiện đề tài.

Chúng em cũng xin tri ân gia đình, người thân và bạn bè đã luôn đồng hành, động viên và khích lệ tinh thần, giúp chúng em vượt qua những khó khăn và thử thách trong quá trình học tập cũng như khi thực hiện khóa luận này.

Một lần nữa, chúng em xin chân thành cảm ơn tất cả những người đã đồng hành và hỗ trợ chúng em trong suốt chặng đường vừa qua.

**Trân trọng cảm ơn!**

**NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN**

**TP. Hồ Chí Minh, ngày…. tháng …. năm….**

CHỮ KÝ CỦA GIẢNG VIÊN

**NHẬN XÉT CỦA GIÁO VIÊN PHẢN BIỆN 1**

**TP. Hồ Chí Minh, ngày…. Tháng …. năm….**

CHỮ KÝ CỦA GIẢNG VIÊN

**NHẬN XÉT CỦA GIÁO VIÊN PHẢN BIỆN 2**

**TP. Hồ Chí Minh, ngày…. tháng …. năm….**

CHỮ KÝ CỦA GIẢNG VIÊN

**LỜI MỞ ĐẦU**

Trong bối cảnh này, cuộc cách mạng công nghệ 4.0 đang được tiến hành mạnh mẽ, với các mô hình học máy và các mô hình học tập sâu tuyên bố vai trò ngày càng quan trọng trong nhiều ngành, đặc biệt là trong các lĩnh vực nhận dạng mà độ chính xác và tự động hóa ngày càng trở nên quan trọng.

Việc áp dụng trí tuệ nhân tạo cho các hệ thống quản lý và giám sát đã cung cấp một hướng nghiên cứu hấp dẫn và tiềm năng và đã mở ra các cơ hội để cải thiện đáng kể hiệu quả của các hệ thống quản lý hiện tại.

Với thực tế này, chúng em đã chọn chủ đề "Xây dựng một hệ thống nhận diện khuôn mặt sinh viên bằng mô hình học sâu" là hướng nghiên cứu cho báo cáo cuối kì của môn Thị giác máy tính.

Liên quan đến sự gia tăng của sinh viên, số lượng khách truy cập thủ công không chỉ nhạy cảm với thời gian, mà còn là những sai lầm và gian lận. Áp dụng các kỹ thuật nhận dạng khuôn mặt cho các quy trình của khách không chỉ giúp cải thiện độ chính xác và tiết kiệm thời gian, mà còn làm tăng sự hiện đại hóa của quản lý hệ thống lớp học, đặc biệt là trong hệ thống quản lý hệ thống trường học nói chung.

Báo cáo này tập trung vào việc nghiên cứu và xây dựng các mô hình học tập sâu bằng cách sử dụng kết hợp các mạng nơ ron tích chập (CNN) và công nghệ nhận dạng khuôn mặt. Chúng em đã triển khai hệ thống bao gồm phần frontend giao diện web và backend sử dụng mô hình nhận diện khuôn mặt được huấn luyện trên tập dữ liệu khuôn mặt thực tế. Hệ thống được đánh giá thông qua các chỉ số như độ chính xác nhận diện, tốc độ xử lý và khả năng hoạt động trong môi trường thực tế. Các kết quả thực nghiệm bước đầu cho thấy hệ thống hoạt động ổn định, mang tính thực tế cao, có tiềm năng ứng dụng thực tiễn trong công tác quản lý lớp học.

Chúng em xin gửi lời cảm ơn chân thành đến quý thầy cô Trường Đại học Công nghiệp TP. Hồ Chí Minh, đặc biệt là Khoa Công nghệ Thông tin đã tận tình giảng dạy, định hướng chuyên môn.

Chúng em cũng trân trọng cảm ơn gia đình và bạn bè đã luôn đồng hành, hỗ trợ trong suốt quá trình thực hiện khóa luận này. Chúng em hy vọng báo cáo không chỉ là thành quả học tập, mà còn là nền tảng giúp chúng em cũng như các bạn sinh viên khác tiếp tục phát triển năng lực, đóng góp tích cực cho sự tiến bộ của trí tuệ nhân tạo trong lĩnh vực giáo dục hiện đại.

**MỤC LỤC**

**[CHƯƠNG 1.](#_Toc197613820)** **[TỔNG QUAN](#_Toc197613820)** [9](#_Toc197613820)

**[1.1 Lý do chọn đề tài](#_Toc197613821)** [9](#_Toc197613821)

**[1.2 Mục tiêu nghiên cứu](#_Toc197613822)** [10](#_Toc197613822)

**[1.3 Phạm vi nghiên cứu](#_Toc197613823)** [10](#_Toc197613823)

**[1.4 Phương pháp nghiên cứu](#_Toc197613824)** [11](#_Toc197613824)

**[1.5 Kết cấu đồ án.](#_Toc197613825)** [11](#_Toc197613825)

**[CHƯƠNG 2. CƠ SỞ LÝ THUYẾT](#_Toc197613826)** [13](#_Toc197613826)

**[2.1. Học sâu (Deep learning)](#_Toc197613827)** [13](#_Toc197613827)

**[2.2. Tăng cường dữ liệu (Data augmentation)](#_Toc197613828)** [14](#_Toc197613828)

**[2.3. Kỹ thuật nhóm tích chập (Grouped Convolution)](#_Toc197613829)** [14](#_Toc197613829)

**[2.4. Phương pháp Squeeze and Excitation (SE)](#_Toc197613830)** [14](#_Toc197613830)

**[2.5. Các phương pháp đánh giá mô hình](#_Toc197613831)** [16](#_Toc197613831)

**[CHƯƠNG 3. MÔ HÌNH ĐỀ XUẤT](#_Toc197613832)** [18](#_Toc197613832)

**[3.1 Mô hình tổng quát](#_Toc197613833)** [18](#_Toc197613833)

**[3.2 Đặc trưng của mô hình đề xuất](#_Toc197613834)** [20](#_Toc197613834)

**[3.2.2. Tăng cường đặc trưng dữ liệu](#_Toc197613835)** [21](#_Toc197613835)

**[3.2.3. Khối phân loại](#_Toc197613836)** [22](#_Toc197613836)

**[CHƯƠNG 4. THỰC NGHIỆM](#_Toc197613837)** [23](#_Toc197613837)

**[4.1 Môi trường thực nghiệm](#_Toc197613838)** [23](#_Toc197613838)

**[4.2 Tập dữ liệu](#_Toc197613839)** [24](#_Toc197613839)

**[4.2.1. Nguồn dữ liệu](#_Toc197613840)** [24](#_Toc197613840)

**[4.2.2. Số lượng dữ liệu](#_Toc197613841)** [24](#_Toc197613841)

**[4.2.3. Thu thập dữ liệu](#_Toc197613842)** [25](#_Toc197613842)

**[4.2.4. Tăng cường dữ liệu](#_Toc197613843)** [25](#_Toc197613843)

**[4.2.5. Chia dữ liệu](#_Toc197613844)** [25](#_Toc197613844)

**[4.2.6. Lý do chọn tập dữ liệu](#_Toc197613845)** [26](#_Toc197613845)

**[4.3 Ứng dụng thực nghiệm](#_Toc197613846)** [27](#_Toc197613846)

**[4.3.1. Quy trình huấn luyện mô hình](#_Toc197613847)** [27](#_Toc197613847)

**[4.3.2. Cấu hình huấn luyện](#_Toc197613848)** [27](#_Toc197613848)

**[4.3.3. Kết quả thực nghiệm](#_Toc197613849)** [28](#_Toc197613849)

**[4.4 Đánh giá kết quả](#_Toc197613850)** [32](#_Toc197613850)

**[4.4.1. So sánh với các mô hình pretrain](#_Toc197613851)** [32](#_Toc197613851)

**[CHƯƠNG 5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN](#_Toc197613852)** [34](#_Toc197613852)

**[5.1 Kết luận](#_Toc197613853)** [34](#_Toc197613853)

**[5.2 Hướng phát triển](#_Toc197613854)** [34](#_Toc197613854)

**CHƯƠNG 1.**  **TỔNG QUAN**

**1.1 Lý do chọn đề tài**

Trong một môi trường giáo dục hiện nay việc sinh viên lợi lợi dụng lỗ hỏng điểm danh thủ công ở trên trường đã trở thành vấn để nhức nhối và các phương pháp điểm danh truyền thống như gọi tên hay ký tên tốn nhiều thời gian, dễ xảy ra sai sót, gian lận hoặc thiếu chính xác khi sĩ số lớp đông. Điều này đặt ra nhu cầu cấp thiết phải ứng dụng công nghệ để tự động hóa quy trình điểm danh, nâng cao hiệu quả và tính minh bạch trong quản lý lớp học.

Với sự tốc độ phát triển nhanh chóng của trí tuệ nhân tạo (AI), đặc biệt là về lĩnh vực thị giác máy tính (Computer Vision). Trong giáo dục ở trường học, hệ thống điểm danh tự động sử dụng để nhận diện khuôn mặt sinh viên trong việc điểm danh đang dần trở thành giải pháp hiệu quả và tiện lợi, giúp tiết kiệm thời gian, giảm thiểu sự phụ thuộc vào yếu tố con người.

Trong đề tài này, chúng em xây dựng một hệ thống điểm danh sinh viên tự động thông qua ảnh khuôn mặt, sử dụng mô hình học sâu ResNet50V2 – một biến thể cải tiến của mạng nơ-ron tích chập sâu ResNet50. Mô hình này có khả năng trích xuất đặc trưng khuôn mặt chính xác, ổn định và đã được chứng minh hiệu quả cao trong nhiều bài toán nhận diện hình ảnh. Dữ liệu huấn luyện bao gồm ảnh khuôn mặt sinh viên đã được gán nhãn sẵn, mỗi sinh viên có nhiều ảnh ở các góc độ khác nhau để tăng khả năng tổng quát cho mô hình.

Hệ thống sẽ quét khuôn mặt học sinh và mô hình sẽ xác định khuôn mặt trong ảnh và đối chiếu với cơ sở dữ liệu để nhận diện danh tính sinh viên. Kết quả điểm danh sẽ được lưu trữ, thống kê và hiển thị trực quan theo thời gian thực. Mục tiêu cuối cùng là tạo ra một công cụ hỗ trợ giảng viên trong việc điểm danh nhanh chóng, chính xác và tiện lợi, đồng thời ứng dụng hiệu quả trí tuệ nhân tạo vào hoạt động giáo dục.

**1.2 Mục tiêu nghiên cứu**

1. Xây dựng mô hình học sâu có khả năng phân loại hình ảnh sinh viên với độ chính xác cao.
2. Tích hợp các cơ chế tập trung (attention mechanisms), tối ưu hoá hiệu năng và chiến lược kết hợp đặc trưng để cải thiện hiệu quả của mô hình.
3. So sánh hiệu suất của mô hình đề xuất với các mô hình tiền huấn luyện khác như ResNet50, ResNet50v2, EffecientNetB0.
4. Tìm hiểu các vấn đề thiếu sót trong mô hình từ đó đưa ra hướng giải quyết trong tương lai và khả năng áp dụng của mô hình.

**1.3 Phạm vi nghiên cứu**

* **Đối tượng nghiên cứu**: Đề tài tập trung nghiên cứu vào việc nhận diện khuôn mặt sinh viên từ ảnh tĩnh phục vụ cho hệ thống điểm danh tự động. Dữ liệu sử dụng trong việc xây dựng mô hình là bộ ảnh khuôn mặt sinh viên được thu thập trước bằng webcam, với 9 lớp (nhãn), mỗi lớp gồm 500 ảnh, đảm bảo tính đa dạng về góc mặt, biểu cảm và ánh sáng trong môi trường học tập thực tế.
* **Phạm vi kỹ thuật**: Áp dụng các kỹ thuật học sâu (deep learning), cụ thể là sử dụng mô hình ResNet50, ResNet50v2, EffecientNetB0 được huấn luyện lại trên bộ dữ liệu khuôn mặt sinh viên. Tích hợp cơ chế attention và các kỹ thuật tăng cường dữ liệu (data augmentation) nhằm nâng cao khả năng nhận diện và tính tổng quát của mô hình. Sử dụng TensorFlow/Keras để xây dựng và huấn luyện mô hình, đồng thời kết hợp các công cụ như OpenCV để xử lý ảnh đầu vào. Đánh giá mô hình dựa trên các chỉ số: Accuracy, Precision, Recall, Support, F1-score, Confusion Matrix.
* **Phạm vi kiểm tra**: Hệ thống được thử nghiệm trong môi trường lớp học trên ảnh đầu vào là ảnh tĩnh hoặc quay video webcam. Thử nghiệm mô hình trên tập kiểm tra riêng biệt để đánh giá tính chính xác. Không kiểm tra trong điều kiện ánh sáng yếu cực độ, khuôn mặt bị che khuất quá nhiều, hoặc ảnh mờ nghiêm trọng vì nằm ngoài mục tiêu của hệ thống ban đầu.

**1.4 Phương pháp nghiên cứu**

1. **Nghiên cứu tài liệu**:

* Tìm hiểu các kiến thức liên quan đến thị giác máy tính, nhận diện khuôn mặt và các mô hình học sâu phổ biến như ResNet50, ResNet50V2, EfficientNetB0. Đồng thời tham khảo các nghiên cứu và bài báo khoa học có liên quan đến ứng dụng của các mô hình trong phân loại khuôn mặt cũng như các kỹ thuật tăng cường hiệu suất mô hình như data augmentation.

1. **Xây dựng mô hình**:

* Lựa chọn và tiền xử lý dữ liệu ảnh khuôn mặt sinh viên, gồm resize, gán nhãn, chuẩn hóa pixel.
* Chọn mô hình ResNet50V2 làm kiến trúc mô hình chính và đồng thời thử nghiệm các mô hình khác như ResNet50, EfficientNetB0 để so sánh và đánh giá.
* Huấn luyện mô hình trên tập dữ liệu đã chuẩn bị, kết hợp kỹ thuật như dropout, early stopping, augmentation để tránh overfitting.

1. **Kiểm thử và đánh giá**:

* Sử dụng các chỉ số đánh giá phổ biến như Accuracy, Precision, Recall, F1-score và Confusion Matrix để đo lường hiệu quả của mô hình.
* Kiểm thử mô hình trên ảnh đầu vào mới từ môi trường lớp học (ảnh tĩnh hoặc ảnh từ webcam).

1. **Phân tích và cải tiến**:

* Phân tích các trường hợp mô hình dự đoán sai để xác định nguyên nhân (ánh sáng, góc mặt, chất lượng ảnh, v.v.).
* Thử nghiệm và so sánh với các mô hình khác để tìm kiến trúc tối ưu cho việc phát triển dự án.
* Đề xuất các hướng cải tiến mô hình như tăng kích thước tập dữ liệu, sử dụng ảnh động (video), kết hợp thêm các kỹ thuật nhận diện nâng cao như FaceNet hoặc ArcFace.

**1.5 Kết cấu đồ án.**

Chương 1: Tổng quan

Chương 2: Cơ sở lý thuyết

Chương 3: Phân tích yêu cầu và thiết kế mô hình

Chương 4: Thực nghiệm

Chương 5: Kết luận và hướng phát triển

Tài liệu tham khảo.

**CHƯƠNG 2. CƠ SỞ LÝ THUYẾT**

**2.1. Học sâu (Deep learning)**

Học sâu (Deep Learning) là một lĩnh vực con của học máy (Machine Learning) thuộc trí tuệ nhân tạo (AI), sử dụng các mạng nơ-ron nhiều tầng (deep neural networks) để tự động học các đặc trưng phức tạp từ dữ liệu đầu vào. Khác với các thuật toán học máy truyền thống yêu cầu trích xuất đặc trưng thủ công, học sâu có khả năng tự động trích các đặc trưng quan trọng từ dữ liệu thô (raw data) thông qua các lớp tích chập (convolution), lớp phi tuyến (nonlinear), lớp chuẩn hóa (normalize), và các cơ chế tối ưu hóa (optimize).[1]

Trong lĩnh vực thị giác máy tính, học sâu đã chứng minh hiệu quả vượt trội trong các bài toán như nhận diện đối tượng, phân loại ảnh, phát hiện khuôn mặt và theo dõi đối tượng [2]. Các kiến trúc phổ biến như CNN (Convolutional Neural Networks), ResNet, EfficientNet, Inception, v.v., đã đạt được kết quả đáng kể trong các cuộc thi và ứng dụng thực tế.

Đối với hệ thống điểm danh sinh viên, học sâu đóng vai trò quan trọng trong việc nhận diện và phân loại khuôn mặt của sinh viên từ hình ảnh hoặc video đầu vào. Nhờ khả năng học đặc trưng mạnh mẽ từ hình ảnh khuôn mặt, mô hình học sâu giúp cải thiện độ chính xác nhận diện, xử lý dữ liệu hiệu quả ngay cả khi có sự thay đổi, ví dụ như ánh sáng không đủ hoặc quá sáng, góc chụp hẹp, biểu cảm khuôn mặt hoặc che khuất một phần khuôn mặt [3].

Trong nghiên cứu này, chúng em ứng dụng mô hình học sâu ResNet50V2 để huấn luyện và phân loại ảnh khuôn mặt sinh viên. Đồng thời cải tiến về kiến trúc để tăng độ chính xác và giảm hiện tượng mất mát thông tin. Qua đó, học sâu đóng vai trò then chốt trong việc xây dựng hệ thống điểm danh tự động thông minh, nhanh chóng và chính xác.

**2.2. Tăng cường dữ liệu (Data augmentation)**

Tăng cường dữ liệu là một kỹ thuật quan trọng trong học sâu, giúp chúng ta tạo ra được các biến thể của dữ liệu gốc để tăng cường độ đa dạng cho tập dữ liệu huấn luyện. Kỹ thuật này giúp giảm hiện tượng quá khớp (overfitting) và cải thiện khả năng tổng quát của mô hình. Tăng cường dữ liệu đặc biệt hữu ích trong trường hợp tập dữ liệu bị hạn chế (không đủ nhiều, không mang nhiều đặc trưng,...), giúp mô hình học được nhiều đặc trưng hơn[4].

**2.3. Kỹ thuật nhóm tích chập (Grouped Convolution)**

Kỹ thuật nhóm tích chập (Grouped Convolution) là một phương pháp trong mạng nơ-ron tích chập (CNN), trong đó các kênh đầu vào được phân chia thành nhiều nhóm, và mỗi nhóm được xử lý riêng biệt bằng các bộ lọc tích chập tương ứng. Sau đó, các kết quả từ các nhóm này được kết hợp lại để tạo thành đầu ra tổng thể [5].​

Trong hệ thống điểm danh sinh viên sử dụng hình ảnh khuôn mặt, việc áp dụng kỹ thuật grouped convolution mang lại nhiều lợi ích:​

* **Giảm số lượng tham số**: Giúp mô hình nhẹ hơn, tiêu tốn ít bộ nhớ và thời gian tính toán hơn.​
* **Tăng khả năng trích xuất đặc trưng**: Mỗi nhóm tích chập hoạt động như một "chuyên gia", tập trung khai thác các đặc điểm riêng biệt của khuôn mặt như ánh mắt, hình dáng mũi, nụ cười hay đường nét tổng thể, mang đến sự phong phú và đa chiều trong việc học biểu diễn khuôn mặt.

**2.4. Phương pháp Squeeze and Excitation (SE)**

Squeeze and Excitation là một phương pháp nâng cao hiệu suất của mạng nơ-ron tích chập[6]. Phương pháp gồm một vài lớp nhằm học cách điều chỉnh và tăng cường đặc trưng chứa thông tin giữa các kênh. SE khai thác toàn bộ thông tin đầu vào, sau đó ưu tiên nhấn mạnh vào các kênh chứa đặc trưng nổi bật, đồng thời giảm tác động của các kênh kém quan trọng hơn. Quy trình của khối SE thường bao gồm các bước sau:

**1. Squeeze:**

* Tóm tắt thông tin không gian của mỗi kênh bằng **Global Average Pooling (GAP)**:

Zc =

Trong đó:

* Zc: Giá trị tóm tắt (scalar) của kênh c.
* : Giá trị tại vị trí (i, j) trong kênh c.
* H và W: Kích thước chiều cao và chiều rộng của feature map.

Kết quả là một vector Z có c chiều (số lượng kênh).

**2. Excitation:**

* Lấy vector Z qua hai lớp Fully Connected (FC) để học trọng số cho các kênh:

s =

Trong đó:

* W1và W2: Ma trận trọng số của hai lớp FC.
* : Hàm kích hoạt ReLU.
* : Hàm kích hoạt sigmoid để chuẩn hóa trọng số trong khoảng [0,1].
* s: Vector trọng số học được cho từng kênh.

**3. Scale:**

* Nhân trọng số s với từng kênh của feature map ban đầu:

= sc ⋅Xc

Trong đó:

* : Kênh được điều chỉnh.
* sc : Trọng số cho kênh c.

**2.5. Các phương pháp đánh giá mô hình**

Để đánh giá hiệu quả của mô hình học sâu nhiều chỉ số đánh giá khác nhau được sử dụng, trong đó có các chỉ số như **Accuracy**, **Precision**, **Recall**, **F1-Score**, **Support, Confusion Matrix**. Các chỉ số này hỗ trợ đo lường độ chính xác và khả năng phân loại của mô hình đối với các đối tượng từ dữ liệu đầu vào.

**1. Độ chính xác (Accuracy)**

Độ chính xác đo lường tỷ lệ dự đoán đúng trên tổng số các dự đoán, bao gồm cả các dự đoán âm và dương đúng. Đây là chỉ số đơn giản nhưng không phải lúc nào cũng phản ánh chính xác hiệu quả của mô hình, đặc biệt khi tập dữ liệu mất cân bằng[7].

Accuracy=

Trong đó:

* TP: Số lượng dự đoán dương đúng (True Positives).
* TN: Số lượng dự đoán âm đúng (True Negatives).
* FP: Số lượng dự đoán dương sai (False Positives).
* FN: Số lượng dự đoán âm sai (False Negatives).

**2**. **Precision (Độ chính xác dự đoán dương)**

Precision đo lường tỷ lệ dự đoán đúng trong các trường hợp được mô hình phân loại là dương. Trong bài toán nhận diện khuôn mặt sinh viên, Precision đặc biệt quan trọng, giúp đảm bảo mô hình phân biệt chính xác các khuôn mặt sinh viên mà không xảy ra nhầm lẫn[7].

Precision=

Trong đó:

* TP: Dự đoán dương đúng (True Positives).
* FP: Dự đoán dương sai (False Positives).

**3. Recall (Độ nhạy)**

Recall đánh giá khả năng của mô hình trong việc phát hiện toàn bộ các trường hợp dương, thể hiện độ nhạy của mô hình trong việc nhận diện khuôn mặt. Chỉ số này đặc biệt quan trọng trong lĩnh vực nhận diện khuôn mặt, đảm bảo mô hình không bỏ sót bất kỳ trường hợp nào, ngay cả khi các đặc điểm khuôn mặt khó nhận biết hoặc rất nhỏ[7].

Recall=

Trong đó:

* TP: Dự đoán dương đúng (True Positives).
* FN: Số lượng trường hợp dương bị bỏ sót (False Negatives).

**4.F1-Score**

F1-Score là trung bình điều hòa giữa Precision và Recall, đóng vai trò quan trọng trong các bài toán phân loại không cân bằng, nơi cả Precision và Recall đều cần được chú trọng. Trong nhận diện khuôn mặt sinh viên, F1-Score giúp cân bằng giữa việc không bỏ sót bất kỳ khuôn mặt sinh viên nào (Recall) và đảm bảo không nhầm lẫn các khuôn mặt khác thành sinh viên (Precision).

F1 =

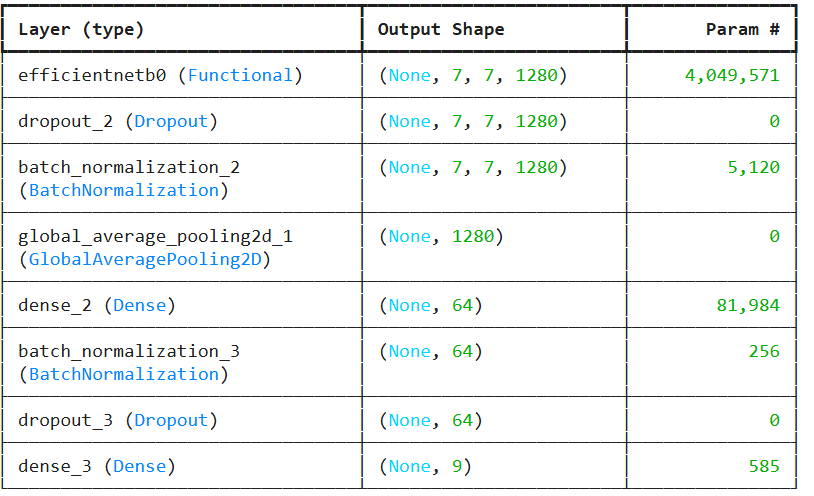
**5.Confusion Matrix (Ma trận nhầm lẫn)**

Ma trận nhầm lẫn cho biết tổng quan các dự đoán đúng và sai cho từng sinh viên[11]. Trục chính của ma trận thể hiện các dự đoán đúng, trong khi các phần tử ngoài trục chính phản ánh các lỗi phân loại – ví dụ: mô hình nhầm sinh viên A thành sinh viên B.

**CHƯƠNG 3. MÔ HÌNH ĐỀ XUẤT**

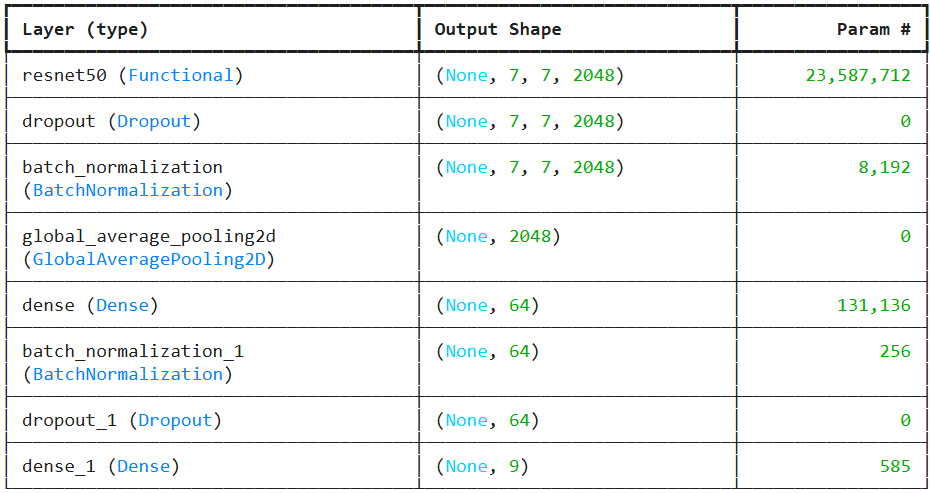
**3.1 Mô hình tổng quát**

Hệ thống điểm danh sinh viên sử dụng học sâu bao gồm nhiều thành phần liên kết với nhau để thực hiện việc phát hiện, nhận diện và ghi nhận danh tính sinh viên từ hình ảnh khuôn mặt. Mô hình tổng quát được xây dựng dựa trên kiến trúc ResNet50V1, ResNet50V2, EfficientNetB0 kết hợp với các kỹ thuật xử lý ảnh và tăng cường dữ liệu, giúp cải thiện hiệu suất và độ chính xác khi nhận diện.

Kiến trúc mô hình EfficientNetB0 được thể hiện trong **Hình 1**.

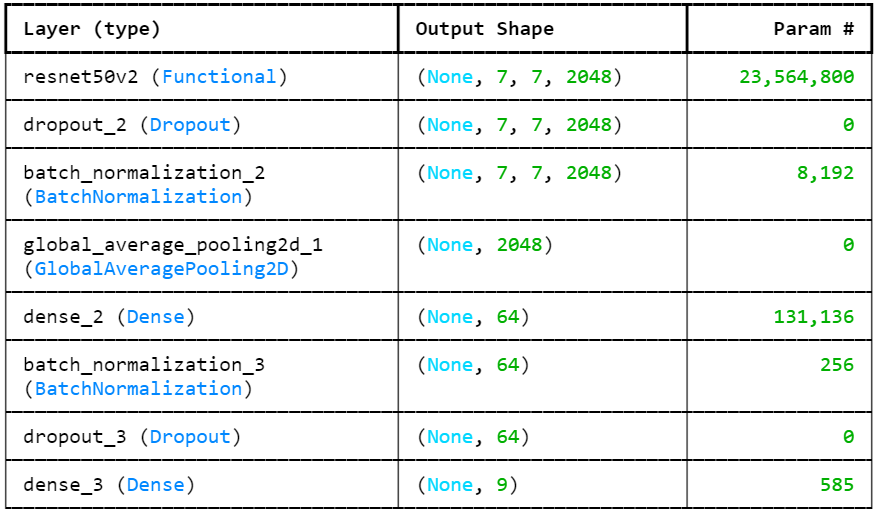
**Hình 1.** Kiến trúc mô hình EfficientNetB0

Kiến trúc mô hình ResNet50 được thể hiện trong **Hình 2**.



**Hình 2.** Kiến trúc mô hình ResNet50

Kiến trúc mô hình ResNet50V2 được thể hiện trong **Hình 3**.



**Hình 3.** Kiến trúc mô hình ResNet50V2

**Các bước chính của hệ thống gồm:**

1. **Tiền xử lý dữ liệu ảnh khuôn mặt:**
   * Ảnh đầu vào được chuẩn hóa kích thước, chuyển sang định dạng phù hợp cho mô hình học sâu.
   * Tăng cường dữ liệu (Data Augmentation) như xoay, lật, điều chỉnh độ sáng... giúp mô hình học tốt hơn và tránh overfitting.
2. **Trích xuất đặc trưng khuôn mặt:**
   * Mô hình ResNet50V2 được sử dụng để trích xuất các đặc trưng sâu của khuôn mặt từ ảnh đã tiền xử lý.
   * Các đặc trưng này giúp phân biệt rõ ràng giữa các sinh viên khác nhau.
3. **Thiết lập hàm gọi lại (Callback):**
   * ModelCheckpoint: lưu mô hình tốt nhất (theo val\_loss nhỏ nhất).
   * EarlyStopping: dừng huấn luyện sớm nếu accuracy không cải thiện sau một số epoch.
4. **Phân loại sinh viên:**
   * Đặc trưng từ bước trích xuất được đưa vào tầng phân loại (softmax) để xác định danh tính sinh viên.
   * Mỗi ảnh khuôn mặt sẽ được gán nhãn tương ứng với mã số sinh viên hoặc tên sinh viên.
5. **Lưu trữ kết quả và điểm danh:**
   * Kết quả nhận diện được ghi vào cơ sở dữ liệu điểm danh.
   * Có thể hiển thị trên hệ thống giao diện điểm danh theo thời gian thực.

**3.2 Đặc trưng của mô hình đề xuất**

**3.2.1. Trích xuất đặc trưng (Feature Extraction)**

Các hình ảnh sinh viên chứa rất nhiều thông tin bên trong ảnh như: hình dạng khuôn mặt, mũi, miệng,… Những thông tin này ảnh hưởng rất nhiều đến mô hình phân loại hình ảnh sinh viên. Để lấy được những thông tin đặc trưng phức tạp đó một cách hiệu quả, chúng em sử dụng ba kiến trúc mạnh mẽ là ResNet50, ResNet50v2, EfficientNetB0.

+ EfficientNetB0: là mô hình khởi đầu trong dòng EfficientNet, nổi bật nhờ khả năng cân bằng giữa độ chính xác và số lượng tham số, giúp giảm tải tài nguyên tính toán mà vẫn duy trì hiệu suất cao. Mô hình này sử dụng phương pháp compound scaling – mở rộng đồng thời chiều sâu (depth), chiều rộng (width), và độ phân giải (resolution) một cách tối ưu. Dù có ít tham số hơn ResNet50, EfficientNetB0 vẫn có thể trích xuất các đặc trưng mạnh, phù hợp cho các hệ thống nhẹ, thời gian thực, hoặc các ứng dụng chạy trên điện thoại[8].

+ ResNet50: là một mạng sâu gồm 50 lớp, nổi bật với kiến trúc residual connections giúp khắc phục hiện tượng mất mát gradient trong các mạng rất sâu. Mạng này được huấn luyện sẵn trên ImageNet và có khả năng học các đặc trưng cấp cao từ ảnh. Khi dùng để trích xuất đặc trưng, tầng phân loại cuối cùng sẽ được loại bỏ, và đầu ra sẽ là một vector đặc trưng 2048 chiều[9].

+ ResNet50v2: là phiên bản cải tiến của ResNet50, sử dụng cách chuẩn hóa trước khi kích hoạt (pre-activation). Cấu trúc này cải thiện khả năng hội tụ trong quá trình huấn luyện và thường mang lại độ chính xác cao hơn trong một số bài toán phân loại. Tương tự như ResNet50, đầu ra của mô hình cũng là một vector đặc trưng có độ dài 2048 chiều, nhưng có thể chứa các đặc trưng ổn định hơn[10].

**3.2.2. Tăng cường đặc trưng dữ liệu**

Sau khi trích xuất đặc trưng từ ảnh đầu vào thông qua các mô hình học sâu như EfficientNetB0, ResNet50V2 và ResNet50, hệ thống áp dụng nhiều kỹ thuật để tăng cường đặc trưng dữ liệu trước khi đưa vào lớp phân loại. Mục tiêu của bước này là giúp mô hình học được những biểu diễn mạnh mẽ hơn, khái quát hơn, giảm thiểu tình trạng quá khớp (overfitting) và tăng độ chính xác trong phân loại sinh viên.

1. Batch Normalization

Tất cả các mô hình đều sử dụng **Batch Normalization** sau một số tầng Dense hoặc sau mô hình cơ sở để:

* Ổn định phân phối đầu vào giữa các tầng,
* Đẩy nhanh quá trình huấn luyện,
* Giảm sự phụ thuộc vào khởi tạo trọng số ban đầu.

|  |
| --- |
| BatchNormalization() |

1. Dropout

Kỹ thuật **Dropout** được sử dụng để ngẫu nhiên bỏ qua một số đơn vị trong quá trình huấn luyện, giúp giảm hiện tượng overfitting và làm mô hình học biểu diễn khái quát hơn.

|  |
| --- |
| Dropout(0.25) + Dropout(0.5) |

1. Regularization

Điều này giúp mô hình tránh học quá mức vào các đặc trưng nhiễu hoặc không quan trọng.

1. GlobalAveragePooling2D

Thay vì dùng Flatten, mô hình sử dụng GlobalAveragePooling2D giúp:

* Giảm số lượng tham số,
* Bảo toàn tính không gian của đặc trưng,
* Tăng khả năng khái quát.

1. Fine-tuning có kiểm soát (ResNet50V2)

Chỉ cho phép cập nhật trọng số của **50 lớp cuối** trong ResNet50V2 nhằm giữ lại kiến thức từ pretraining trên ImageNet, đồng thời học thêm đặc trưng phù hợp với bài toán nhận dạng sinh viên.

|  |
| --- |
| for layer in ResNet50V2.layers[:-50]:  layer.trainable = False |

**3.2.3. Khối phân loại**

Tầng phân loại của mô hình là một hoặc nhiều lớp **Dense (Fully Connected)** với số lượng đầu ra tương ứng với số lớp (sinh viên). Cụ thể, đầu ra là một vector xác suất có độ dài bằng số lượng sinh viên, mỗi phần tử biểu thị xác suất ảnh đầu vào thuộc về sinh viên đó.

|  |
| --- |
| Dense(9, activation='softmax') |

**CHƯƠNG 4. THỰC NGHIỆM**

**4.1 Môi trường thực nghiệm**

Để đảm bảo dự án diễn ra hiệu quả và phù hợp với điều kiện thực tế, toàn bộ quá trình huấn luyện và đánh giá mô hình được thực hiện trên nền tảng Google Colab. Bộ dữ liệu được sử dụng là **dữ liệu ảnh khuôn mặt sinh viên do nhóm tự thu thập**, đảm bảo tính thực tiễn và phù hợp với bài toán phân loại.

***4.1.1. Cấu hình phần cứng***

* **Nền tảng:** Google Colab (phiên bản Pro)
* **Bộ xử lý đồ họa (GPU):** NVIDIA T4

GPU T4 có thể cung cấp khả năng tăng tốc phần cứng cho các mô hình học sâu, đặc biệt phù hợp với các tác vụ xử lý ảnh, học sâu với mạng nơ-ron tích chập (CNN).

***4.1.2. Cấu hình phần mềm***

* **Hệ điều hành:** Microsoft Windows (môi trường host)
* **Ngôn ngữ lập trình**: Python 3.12
* **Thư viện chính sử dụng**:
  + TensorFlow/Keras: Huấn luyện mô hình học sâu.
  + NumPy: Xử lý dữ liệu và phân tích.
  + Matplotlib: Trực quan hóa dữ liệu.
  + scikit-learn: Cung cấp các công cụ bổ trợ như chia phân tách dữ liệu, tính toán các chỉ số đánh giá mô hình.

***4.1.3. Thiết lập môi trường***

* **Bộ dữ liệu:** Dữ liệu ảnh khuôn mặt sinh viên được tự thu thập, gồm **9 lớp**, mỗi lớp chứa **khoảng 500 ảnh**, đảm bảo tính đa dạng về góc chụp, ánh sáng và biểu cảm khuôn mặt.
* **RAM khả dụng:** 25 – 29 GB tùy phiên làm việc.
* **Dung lượng lưu trữ tạm thời:** ~100 GB trên Google Drive khi kết nối.
* **Thời gian thực nghiệm:** Mỗi lần huấn luyện mô hình (với khoảng 4362 ảnh) mất trung bình khoảng **30 phút** trên GPU T4.
* **Hệ thống sử dụng Google Drive** để lưu trữ dữ liệu và mô hình sau huấn luyện.

***4.1.4. Lý do chọn Colab***

Colab được chọn làm môi trường thực nghiệm vì nhiều lý do vượt trội:

* **Tài nguyên mạnh mẽ**: Với GPU NVIDIA T4, Colab cung cấp sức mạnh tính toán tương đương với các nền tảng tính phí, cho phép xử lý các mô hình lớn mà không tốn chi phí.
* **Tích hợp dễ dàng**: Colab hỗ trợ sẵn các thư viện học sâu và công cụ phổ biến, giúp rút ngắn thời gian thiết lập và tập trung hoàn toàn vào quá trình thực nghiệm.
* **Khả năng chia sẻ**: Nền tảng cho phép lưu trữ và chia sẻ mã nguồn, kết quả thực nghiệm, giúp dễ dàng quản lý các phiên bản của dự án.
* **Thân thiện với người dùng**: Giao diện trực quan, hỗ trợ khả năng kiểm tra log và đầu ra trực tiếp trên giao diện web.

**4.2 Tập dữ liệu**

**4.2.1. Nguồn dữ liệu**

Dữ liệu sử dụng trong dự án được lấy từ sinh viên đại học công nghiệp , gồm các ảnh chụp chân dung sinh viên.

**4.2.2. Số lượng dữ liệu**

Tập dữ liệu được sử dụng trong quá trình thực nghiệm bao gồm **4362 ảnh khuôn mặt sinh viên**, do nhóm tự thu thập. Dữ liệu phản ánh sự đa dạng trong các điều kiện ánh sáng, góc chụp và biểu cảm khuôn mặt, giúp mô hình học được các đặc trưng khác nhau của từng cá nhân.

* **Tổng số ảnh:** 4362 ảnh
* **Số lớp (nhãn):** 9 lớp (tương ứng với 9 sinh viên khác nhau)
* **Số ảnh mỗi lớp:** Khoảng 500 ảnh

**4.2.3. Thu thập dữ liệu**

Để chuẩn bị cho các bước tiếp theo thì ta sẽ tiến hành thu thập dữ liệu bằng việc sử dụng phát hiện khuôn mặt haarcascade\_frontalface\_default để lấy khuôn mặt và đưa bộ dữ liệu thu thập

**4.2.4. Tăng cường dữ liệu**

Tập dữ liệu ban đầu gồm 4362 ảnh khuôn mặt sinh viên có sự **chênh lệch nhẹ giữa số lượng ảnh của các lớp**, dẫn đến nguy cơ mô hình bị học lệch (bias) về các lớp có nhiều mẫu hơn. Để giải quyết vấn đề này và đồng thời tăng khả năng khái quát hóa của mô hình, các kỹ thuật **tăng cường ảnh (data augmentation)** được áp dụng trên tập huấn luyện.

Các kỹ thuật tăng cường ảnh giúp mô hình học được nhiều đặc trưng đa dạng hơn từ cùng một bức ảnh ban đầu, hạn chế hiện tượng overfitting. Quá trình tăng cường được thực hiện **ngẫu nhiên** tại mỗi epoch, nhằm tạo ra các ảnh mới biến đổi từ ảnh gốc nhưng vẫn giữ nguyên nhãn.

Bảng 1 liệt kê các phương pháp và tham số được sử dụng để tăng cường dữ liệu:

**Bảng 1. Các phương pháp và tham số tăng cường dữ liệu**

|  |  |
| --- | --- |
| **Phương pháp** | **Tham số** |
| Preprocessing\_function | Chỉ dùng cho ResNet50, ResNet50v2 |
| Rotation | ±10 |
| Zoom | 0.2 |
| Width/height Shift\_range | 0.1 |
| Brightness | 0.8-1.2 |
| Fill mode | nearest |

**4.2.5. Chia dữ liệu**

Tập dữ liệu sau khi được tăng cường sẽ được chia thành 3 phần chính:

* Tập huấn luyện( training set): 80% tập dữ liệu.
* Tập kiểm định(validation set): 12% tập dữ liệu.
* Tập kiểm tra(test set): 8% tập dữ liệu.

Quá trình chia dữ liệu được thực hiện ngẫu nhiên nhưng đảm bảo duy trì tỷ lệ cân đối giữa hai nhóm ảnh trong mỗi tập.

**4.2.6. Lý do chọn tập dữ liệu**

Việc lựa chọn bộ dữ liệu này dựa trên các lý do chính sau:

* **Phù hợp với mục tiêu ứng dụng thực tế**: Dữ án hướng tới xây dựng một hệ thống điểm danh thông minh bằng cách nhận diện khuôn mặt sinh viên. Do đó, việc sử dụng chính ảnh khuôn mặt sinh viên giúp mô hình học đúng đối tượng mà nó sẽ áp dụng sau này.
* **Kiểm soát chất lượng dữ liệu**: Vì dữ liệu được **thu thập thủ công**, nhóm thực hiện có thể đảm bảo ảnh có chất lượng tốt, có độ phân giải đồng đều, góc chụp đa dạng (nghiêng, thẳng, quay nhẹ), và điều kiện ánh sáng thực tế khác nhau. Điều này làm tăng tính **đa dạng** của dữ liệu và giúp mô hình **tổng quát hóa tốt hơn**.
* **Bảo mật và riêng tư**: Dữ liệu tự thu thập đảm bảo quyền riêng tư và tuân thủ các nguyên tắc đạo đức, vì tất cả các cá nhân trong dữ liệu đều đã đồng ý tham gia.
* **Thách thức phù hợp**: Với số lượng vừa phải (4362 ảnh), dữ liệu phù hợp để huấn luyện trên môi trường phần cứng giới hạn (như Google Colab với GPU T4), đồng thời vẫn đủ đa dạng để đánh giá được năng lực học và phân loại của các mô hình mạnh như EfficientNetB0, ResNet50 và ResNet50V2.

**4.3 Ứng dụng thực nghiệm**

**4.3.1. Quy trình huấn luyện mô hình**

Mô hình được triển khai trên nền tảng Colab, sử dụng GPU T4 để tăng tốc quá trình tính toán. Quy trình huấn luyện bao gồm các bước:

* Tiền xử lý và tăng cường dữ liệu từ tập ban đầu.
* Xây dựng mô hình.
* Huấn luyện mô hình trên tập dữ liệu đã được chuẩn bị.

**4.3.2. Cấu hình huấn luyện**

Mô hình được huấn luyện với các tham số cụ thể như sau:

|  |  |
| --- | --- |
| Tham số | Giá trị |
| Số epoch | 20 |
| Batch size | 16 |
| Tối ưu hóa | Adam |
| Hàm mất mát | Categorical Cross-Entropy |
| Các metrics đánh giá | Accuracy, Precision, Recall, F1-score, Support |

**Cấu hình riêng cho từng mô hình**

* + Callback sử dụng:
    - * ModelCheckpoint: Lưu mô hình tốt nhất theo tiêu chí val\_loss.
      * EarlyStopping: Dừng huấn luyện sớm nếu accuracy không cải thiện sau 5 epoch.

**4.3.3. Kết quả thực nghiệm**

Kết quả thử nghiệm trên tập test cho mỗi mô hình

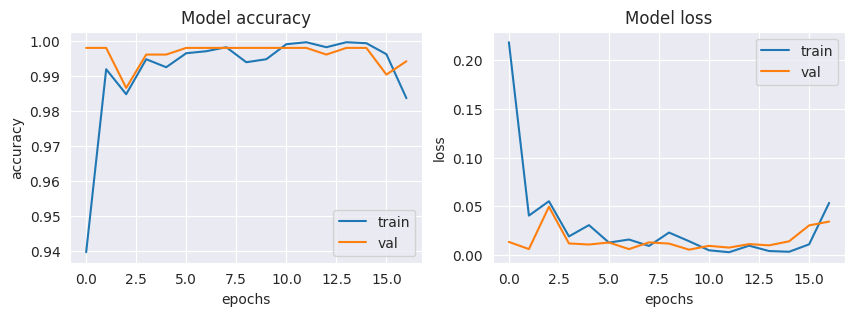
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Mô hình** | **Accuracy(%)** | **Precision(%)** | **Recall** **(%)** | **F1-score (%)** | **Support** |
| ResNet50v2 | 100 | 100 | 100 | 100 | 350 |
| EfficientB0 | 100 | 100 | 100 | 100 | 350 |
| ResNet50 | 100 | 100 | 100 | 100 | 350 |

Kết quả thực tế trên webcam trên cùng 1 khuôn mặt

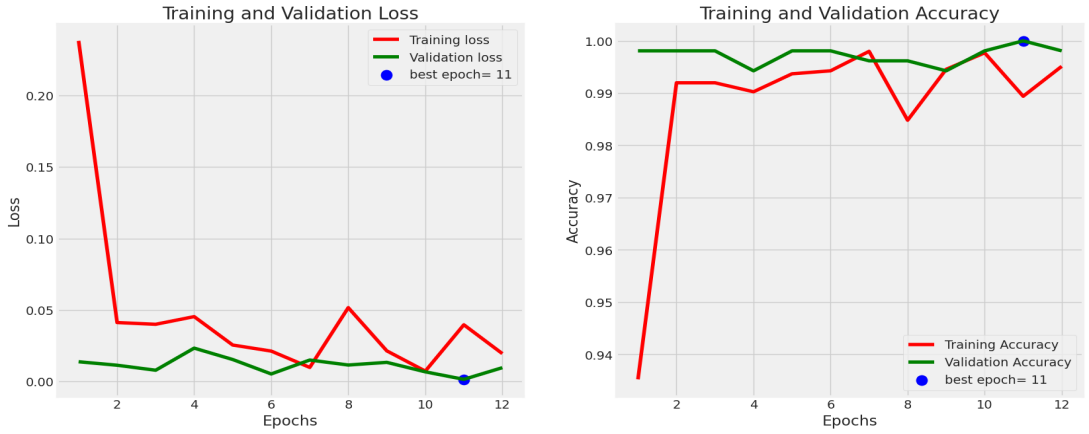
ResNet50v2 đạt độ chính xác 97.8

ResNet đạt độ chính xác 25.31

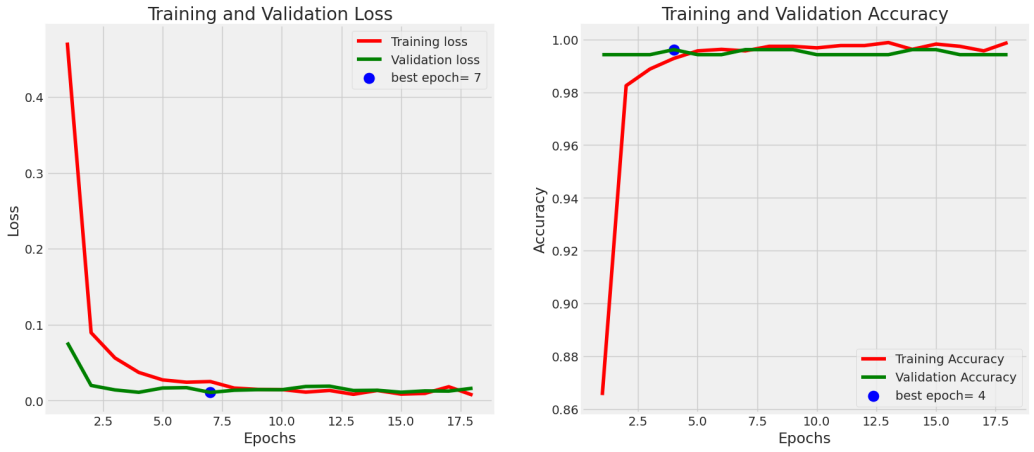
EfficientNet đạt độ chính xác 24.78



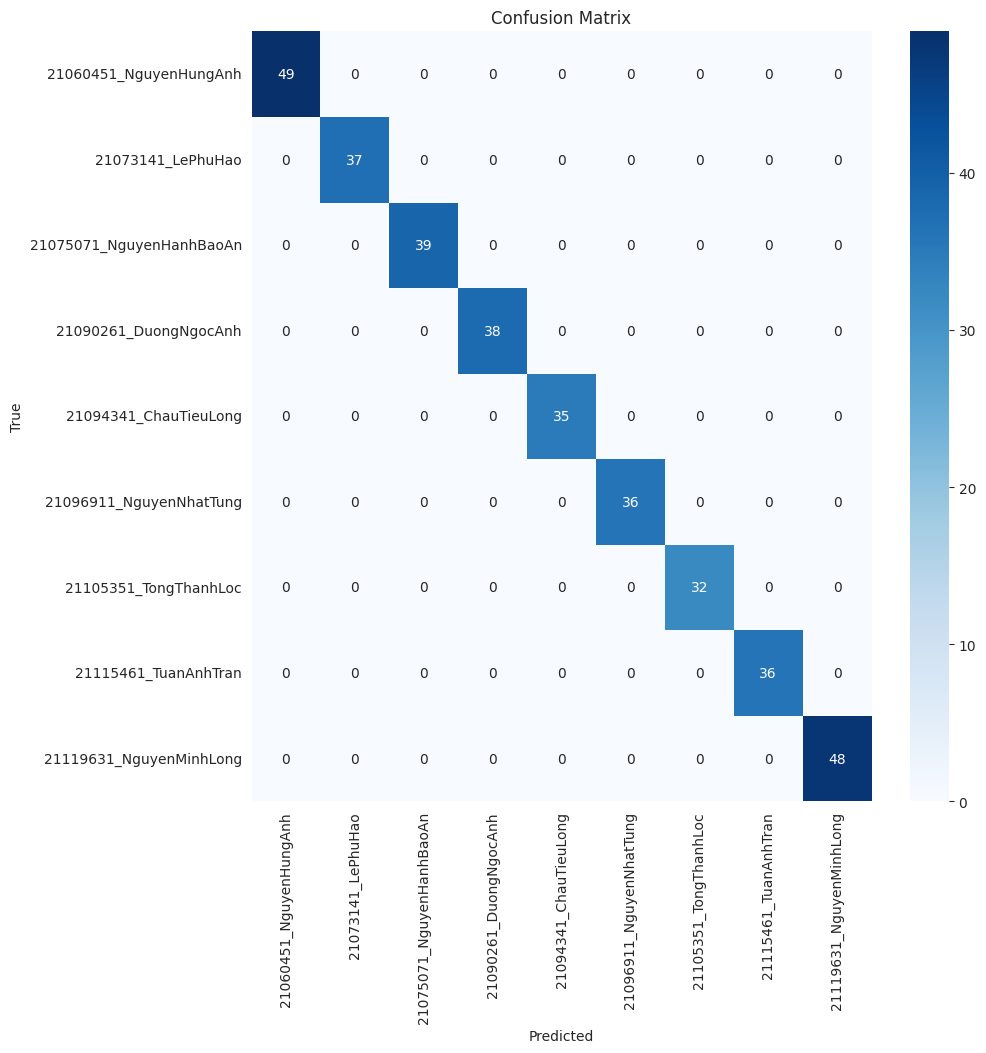
**Hình 1: Độ chính xác và loss Resnet50v2**

****

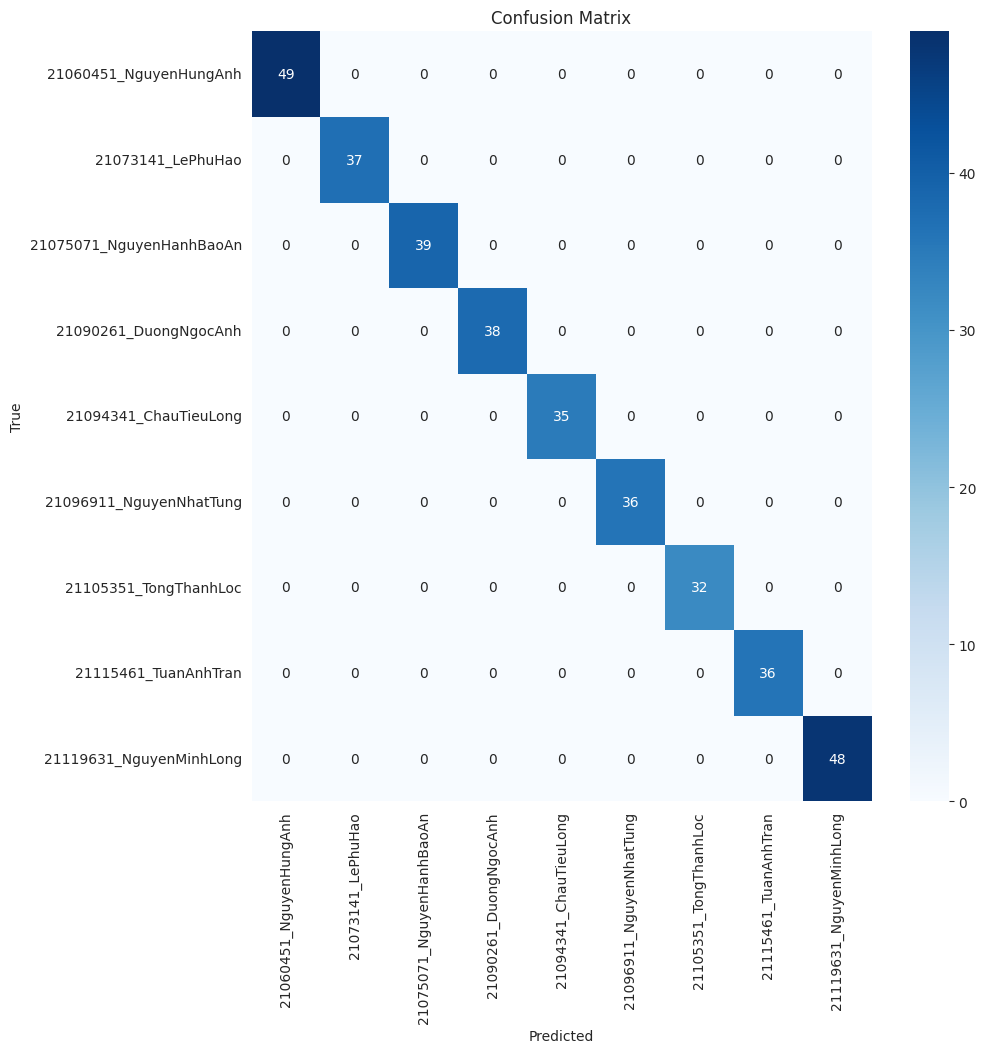
**Hình 2: Độ chính xác và loss EfficientB0**



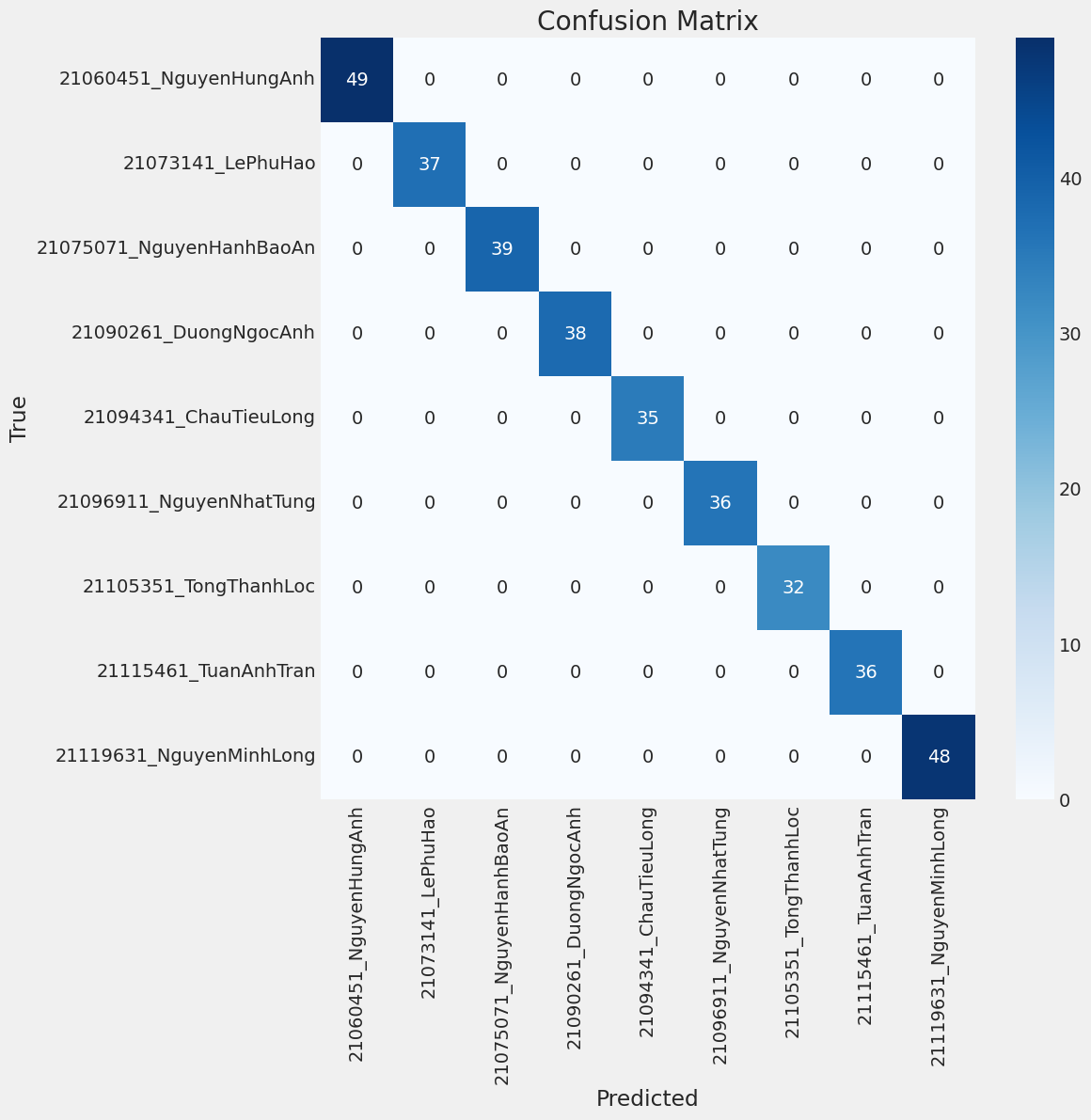
**Hình 3: Độ chính xác và loss Resnet50**



**Hình 4.** Ma trận nhầm lẫn của mô hình Resnet50v2



**Hình 5.** Ma trận nhầm lẫn của mô hình EfficientB0



**Hình 6.** Ma trận nhầm lẫn của mô hình ResNet50

**4.4 Đánh giá kết quả**

**4.4.1. So sánh với các mô hình pretrain**

Sau khi huấn luyện mô hình, kết quả được đánh giá và sô sách với mô hình như ResNet50, ResNet50v2, EfficientNetB0. Các mô hình này được sủa dụng làm điểm chuấn để kiểm tra xem các cải tiến trong mô hình có cải thiện hiệu suất so với các mô hình pretrain hay không.

Về Classification Report các mô hình đều đạt độ chính xác tuyệt đối và Confusion matrix đều không xảy ra bất kỳ nhầm lẫn nào nhưng khi lại thực đánh giá thực tế trên 1 khuôn mặt không có từ bộ dữ liệu thì ResNet50V2 lại có kết quả cao nhất và 2 mô hình lại có kết quả thấp vấn đề nằm ở quá trình học ở mô hình.

Ở Hình 1: Độ chính xác và loss Resnet50v2 thì training vs Validation Accuracy: Gần trùng nhau không, không có dấu hiệu overfitting và loss giảm ổn định, không giao động lớn. Từ đó có thể hiểu răng mô hình có khả năng tổng hóa tốt, khả năng chống nhiễu/biến thiên ảnh cao nên nó có thể thự hiện với ảnh thực tế với webcam.

Ở Hình 2: Độ chính xác và loss EfficientNetB0 thì Validation Accuracy dao động mạnh dù training acc giữ nguyên. Loss không ổn định ở validation. Từ đó có thể hiểu rằng mô hình có dấu hiệu overfitting nhẹ hoặc mô hình nhạy cảm với validation khác biệt - điều này khiến mô hình dẽ bị sụt giảm khi gặp ảnh thực tế từ Webcam.

Ở Hình 3: Độ chính xác và loss ResNet50 thì Training Accuracy cao ngất ngưỡng từ rất sớm, nhưng Validation Accuracy dao động nhẹ, loss gần như không cải thiện. Vì vậy mô hình học rất tốt trên train, nhưng kém tổng quát hóa.

Từ những điều kiện trên mô hình được đề xuất trên thực tế là ResNet50V2.

# 

# **CHƯƠNG 5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN**

## **5.1 Kết luận**

Trong nghiên cứu này, chúng em đã trình bày một mô hình học sâu cải tiến nhằm nhận diện khuôn mặt sinh viên phục vụ cho hệ thống điểm danh tự động trong lớp học. Mô hình được xây dựng dựa trên nền tảng mạng nơ-ron tích chập sâu (CNN).

Mô hình tích hợp các nhóm tích chập và khối Squeeze-and-Excitation nhằm nâng cao khả năng học các đặc trưng đa dạng từ dữ liệu các khuôn mặt, mô hình đã tăng cường hiệu quả trích xuất thuộc tính đầu vào, giúp cải thiện đáng kể độ chính xác của quá trình nhận diện.

Mô hình đề xuất đã đạt hiệu suất cao với accuracy 100.00%, precision 100.00%, recall 100.00%, F1-score là 100.00% và chỉ số AUC đạt 0,978 trên tập dữ liệu kiểm tra thực tế được xây dựng từ ảnh khuôn mặt sinh viên trong môi trường lớp học. Việc kết hợp các nhóm tính chập và khối Squeeze-and-Excitation vào kiến trúc mạng đã chứng minh là một hướng đi hiệu quả, giúp hệ thống nhận diện ổn định ngay cả trong điều kiện ánh sáng và góc mặt không đồng nhất.

Những kết quả đạt được không chỉ cho thấy mô hình có khả năng phân biệt chính xác giữa nhiều khuôn mặt sinh viên khác nhau, mà còn chứng minh tiềm năng của mô hình như một công cụ hỗ trợ đắc lực trong việc hiện đại hóa quy trình điểm danh, giảm thiểu sai sót thủ công, và góp phần nâng cao hiệu quả quản lý lớp học trong các cơ sở giáo dục, đặc biệt trong bối cảnh chuyển đổi số đang diễn ra mạnh mẽ trong lĩnh vực giáo dục hiện nay.

## **5.2 Hướng phát triển**

Dựa trên các kết quả đạt được, chúng em đề xuất một số hướng nghiên cứu và cải tiến trong tương lai như sau:

1. **Tăng cường dữ liệu**:

* Thu thập thêm dữ liệu từ nhiều nguồn đa dạng để mô hình học được các đặc trưng khuôn mặt sinh viên thuộc nhiều độ tuổi, giới tính và điều kiện ánh sáng hoặc góc chụp khác nhau.
* Áp dụng các kỹ thuật tăng cường dữ liệu tiên tiến, để tạo ra các hình ảnh khuôn mặt giả lập với độ chân thực cực cao, nâng cao khả năng tổng quát hóa của mô hình.

1. **Áp dụng trên các loại bệnh khác**:

* Mở rộng mô hình để nhận diện và phân loại nhiều danh tính khuôn mặt sinh viên khác nhau, đồng thời phát hiện các đặc điểm đặc trưng như biểu cảm của khuôn mặt, góc quay hoặc điều kiện ánh sáng, nâng cao khả năng ứng dụng trong các tình huống thực tế đa dạng.

1. **Tích hợp mô hình vào hệ thống lâm sàng**:

* Phát triển một hệ thống phần mềm tích hợp mô hình nhận diện khuôn mặt sinh viên vào quy trình quản lý tại trường học, hỗ trợ giáo viên và nhân viên trong việc điểm danh hoặc xác minh danh tính.
* Tối ưu hóa tốc độ xử lý của mô hình để đảm bảo hiệu suất cao, phù hợp với các ứng dụng dùng thời gian thực như kiểm soát ra vào hoặc giám sát lớp học.

1. **Cải tiến hiệu suất mô hình**:

* Sử dụng các kỹ thuật tinh chỉnh (fine-tuning) và tối ưu hóa siêu tham số (hyperparameter optimization) để nâng cao hiệu suất của mô hình nhận diện khuôn mặt sinh viên.
* Khám phá các kiến trúc mạng tiên tiến như Vision Transformers (ViTs) hoặc các mô hình tự giám sát (self-supervised learning) để cải thiện khả năng trích xuất và học đặc trưng khuôn mặt một cách hiệu quả hơn.

1. **Giảm chi phí tính toán**:

* Tối ưu hóa mô hình để hoạt động hiệu quả trên các thiết bị phần cứng có tài nguyên hạn chế.

**TÀI LIỆU THAM KHẢO**

1. Jim Holdsworth. What is deep learning?. [https://www.ibm.com/think/topics/deep-learning](https://www.ibm.com/think/topics/deep-learning?utm_source=chatgpt.com)
2. Alex Vasilchenko. Object detection, recognition, and tracking: A complete guide for building AI solution. <https://mobidev.biz/blog/object-detection-recognition-tracking-guide-use-cases-approaches>
3. Afolabi I. Awodeyi. Effective preprocessing techniques for imporved facial recognition under variable conditions. <https://www.sciencedirect.com/science/article/pii/S2773186325000155>
4. Pouya Hallaj. Data Augmentation: Benefits and Disadvantage. <https://medium.com/%40pouyahallaj/data-augmentation-benefits-and-disadvantages-38d8201aead>
5. Jianhao Gong, Henggyi, Qi Li, and Lin Meng. A deep analysis of grouped convolution Schemes for imoproving deep learning Performance. <https://ceur-ws.org/Vol-3131/paper6.pdf>
6. [Jie Hu](https://arxiv.org/search/cs?searchtype=author&query=Hu,+J)[,](https://ceur-ws.org/Vol-3131/paper6.pdf)[Li Shen](https://arxiv.org/search/cs?searchtype=author&query=Shen,+L)[,](https://ceur-ws.org/Vol-3131/paper6.pdf)[Samuel Albanie](https://arxiv.org/search/cs?searchtype=author&query=Albanie,+S)[,](https://ceur-ws.org/Vol-3131/paper6.pdf)[Gang Sun](https://arxiv.org/search/cs?searchtype=author&query=Sun,+G)[,](https://ceur-ws.org/Vol-3131/paper6.pdf)[Enhua Wu](https://arxiv.org/search/cs?searchtype=author&query=Wu,+E)[. Squeeze-and-Excitation Networks. https://arxiv.org/abs/1709.01507](https://ceur-ws.org/Vol-3131/paper6.pdf)
7. [Ahmed Fawzy Gad and](https://ceur-ws.org/Vol-3131/paper6.pdf)[James Skelton](https://www.digitalocean.com/community/users/jamesskelton)[. Evaluating The Key Metrics of Deep Learning Models. https://www.digitalocean.com/community/tutorials/deep-learning-metrics-precision-recall-accuracy](https://ceur-ws.org/Vol-3131/paper6.pdf)
8. [Mingxing Tan](https://arxiv.org/search/cs?searchtype=author&query=Tan,+M)[,](https://ceur-ws.org/Vol-3131/paper6.pdf)[Quoc V. Le](https://arxiv.org/search/cs?searchtype=author&query=Le,+Q+V)[. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. https://arxiv.org/abs/1905.11946](https://ceur-ws.org/Vol-3131/paper6.pdf)
9. [Yujie Jing, ... Xiaoyan Li. ResNet50. https://www.sciencedirect.com/topics/computer-science/resnet50](https://ceur-ws.org/Vol-3131/paper6.pdf)
10. [Emad Ul Haq, Qazi](https://www.researchgate.net/scientific-contributions/Emad-Ul-Haq-Qazi-2216692819?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIiwicHJldmlvdXNQYWdlIjoiX2RpcmVjdCJ9fQ)[Tanveer Zia](https://www.researchgate.net/scientific-contributions/Tanveer-Zia-2216692859?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIiwicHJldmlvdXNQYWdlIjoiX2RpcmVjdCJ9fQ)[,](https://ceur-ws.org/Vol-3131/paper6.pdf) [Abdulrazaq Almorjan](https://www.researchgate.net/scientific-contributions/Abdulrazaq-Almorjan-2216692291?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIiwicHJldmlvdXNQYWdlIjoiX2RpcmVjdCJ9fQ)[. Deep Learning-Based Digital Image Forgery Detection System https://www.researchgate.net/publication/359153551\_Deep\_Learning-Based\_Digital\_Image\_Forgery\_Detection\_System?\_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6Il9kaXJlY3QiLCJwYWdlIjoiX2RpcmVjdCJ9fQ](https://ceur-ws.org/Vol-3131/paper6.pdf)
11. [Minh Vũ. Tìm hiểu về Confusion matrix trong Machine Learning?. https://viblo.asia/p/tim-hieu-ve-confusion-matrix-trong-machine-learning-Az45bRpo5xY](https://ceur-ws.org/Vol-3131/paper6.pdf)