Bộ giáo dục và đào tạo Khoa Công Nghệ Thông Tin

**Trường Đại Học Sài Gòn** *Môn: PP luận nghiên cứu khoa học*

**ĐỀ CƯƠNG CHI TIẾT**

Ngày thực hiện đề cương : 25/03/2025

| **TÊN ĐỀ TÀI** | Face Recognition Using Convolutional Neural Networks | |
| --- | --- | --- |
| **LĨNH VỰC CHUYÊN NGÀNH** | Công nghệ phần mềm | |
| **LOẠI HÌNH NGHIÊN CỨU** | Sản phẩm | |
| **NGƯỜI HƯỚNG DẪN** | TS. Đỗ Như Tài | |
| ***SINH VIÊN THỰC HIỆN 1*** | Lê Quang Vinh | 3120411176 |
| ***SINH VIÊN THỰC HIỆN 2*** | Vũ Thị Thanh Vân | 3120411173 |

**GIỚI THIỆU**

| **CÂU HỎI GỢI Ý** | **Lĩnh vực và nghiên cứu liên quan đã và đang phát triển như thế nào?**  **Các vấn đề, và bài toán đặt ra cần giải quyết là gì?** |
| --- | --- |
| **HƯỚNG DẪN** | Giới thiệu tổng quan về đề tài – những vấn đề và lĩnh vực liên quan đến đề tài. |
| Convolutional Neural Networks (CNN) là một trong những mô hình mạng Học sâu phổ biến nhất hiện nay [1-3], có khả năng nhận dạng và phân loại hình ảnh với độ chính xác rất cao, thậm chí còn tốt hơn con người trong nhiều trường hợp. Mô hình này đã và đang được phát triển, ứng dụng vào các hệ thống xử lý ảnh lớn của Facebook, Google hay Amazon… cho các mục đích khác nhau, như các thuật toán gắn thẻ tự động, tìm kiếm ảnh hoặc gợi ý sản phẩm cho người tiêu dùng.  Sự ra đời của mạng CNN là dựa trên ý tưởng cải tiến cách thức các mạng nơ ron nhân tạo truyền thống [4] học thông tin trong ảnh. Do sử dụng các liên kết đầy đủ giữa các điểm ảnh vào node, các mạng nơ ron nhân tạo truyền thẳng (Feedforward Neural Network) [5-7] bị hạn chế rất nhiều bởi kích thước của ảnh, ảnh càng lớn thì số lượng liên kết càng tăng nhanh, kéo theo sự bùng nổ khối lượng tính toán. Ngoài ra, sự liên kết đầy đủ này cũng là sự dư thừa với mỗi bức ảnh, các thông tin chủ yếu thể hiện qua sự phụ thuộc giữa các điểm ảnh với những điểm xung quanh nó mà không quan tâm nhiều đến các điểm ảnh ở cách xa nhau. Mạng CNN với kiến trúc thay đổi, có khả năng xây dựng liên kết chỉ sử dụng một phần cục bộ trong ảnh kết nối đến node trong lớp tiếp theo thay vì toàn bộ ảnh như trong mạng nơ ron truyền thẳng.  Một hệ thống nhận diện mặt người thông thường bao gồm bốn bước xử lý sau:  1. Phát hiện khuôn mặt (Face Detection).  2. Phân đoạn khuôn mặt (Face Alignment hay Segmentation).  3. Trích chọn đặc trưng (Feature Extraction).  4. Nhận diện (Recognition) hay Phân lớp khuôn mặt (Face Classification).  Bên cạnh những bước chính nêu trên, còn có thể áp dụng thêm một số bước khác như tiền xử lý, hậu xử lý nhằm làm tăng độ chính xác cho hệ thống. Sau bước phát hiện khuôn mặt, ta có thể thực hiện bước tiền xử lý (Preprocessing) bao gồm các bước căn chỉnh ảnh (face image alignment) và chuẩn hóa ánh sáng (illumination normalization). Do một vài thông số như: tư thế khuôn mặt, độ sáng, điều kiện ánh sáng,… phát hiện khuôn mặt được đánh giá là bước khó và quan trọng nhất so với các bước còn lại của hệ thống. | |
| **GHI CHÚ** | |

**TÌNH HÌNH NGHIÊN CỨU**

| **CÂU HỎI GỢI Ý** | **Lĩnh vực và nghiên cứu liên quan đã và đang phát triển như thế nào?**  **Các vấn đề, và bài toán đặt ra cần giải quyết là gi?̀** |
| --- | --- |
| **HƯỚNG DẪN** | Tìm hiểu các nghiên cứu đã công bố gần nhất (5 năm trở lại) về lĩnh vực liên quan.  Phân tích các kết quả đóng góp, nhận xét các hạn chế còn tồn tại. |
| 1. **Các kết quả đóng góp**  * *Hiệu quả của mô hình Lightweight CNN* * Nghiên cứu năm 2024 đã phát triển một mô hình CNN nhẹ với số lớp nhỏ, đạt độ chính xác kiểm thử trung bình 94,19% trên dữ liệu chưa thấy trước đó. * Kết quả này cho thấy rằng các mô hình CNN nhẹ có thể đạt hiệu suất cao mà không cần cấu trúc phức tạp, phù hợp cho các ứng dụng trên thiết bị có tài nguyên hạn chế.​ * *Hệ thống nhận diện khuôn mặt thời gian thực* * Một nghiên cứu đã thiết kế và đánh giá hệ thống nhận diện khuôn mặt thời gian thực sử dụng CNN, đạt độ chính xác 98,75% trên tập dữ liệu chuẩn và 98,00% trong điều kiện thời gian thực. * Điều này chứng minh khả năng ứng dụng của CNN trong các hệ thống thực tế. * *Tối ưu hóa hiệu suất và độ chính xác* * Nghiên cứu giới thiệu RobFaceNet, một mô hình CNN được thiết kế để tối ưu hóa độ chính xác trong khi vẫn duy trì hiệu suất tính toán, bằng cách kết hợp các cơ chế chú ý và trích xuất đặc trưng đa cấp. * Mô hình này đạt được sự cân bằng giữa hiệu suất và độ chính xác, phù hợp cho các ứng dụng yêu cầu thời gian thực.  1. **Các hạn chế còn tồn tại**  * *Thiên lệch chủng tộc và giới tính* * Nghiên cứu của NIST năm 2019 phát hiện rằng nhiều thuật toán nhận diện khuôn mặt có sai số cao hơn đối với người da đen, châu Á và phụ nữ, so với nam giới da trắng. * Điều này cho thấy sự cần thiết phải cải thiện tính công bằng trong các hệ thống nhận diện khuôn mặt. * *Chi phí tính toán cao* * Một phân tích chỉ ra rằng mặc dù CNN đã cải thiện độ chính xác trong nhận diện khuôn mặt, nhưng việc triển khai thời gian thực vẫn là một thách thức do chi phí tính toán cao. * Điều này hạn chế khả năng triển khai trên các thiết bị có tài nguyên hạn chế. * *Thiếu tính minh bạch và khả năng giải thích* * Nghiên cứu về hệ thống nhận diện khuôn mặt dựa trên AI có thể giải thích cho thấy rằng người dùng thường không biết mô hình dựa vào đặc trưng nào để nhận diện khuôn mặt. * Việc thiếu khả năng giải thích này gây khó khăn trong việc kiểm định và giải trình trong các ứng dụng nhạy cảm. | |
| **GHI CHÚ** | |

**MỤC TIÊU VÀ PHẠM VI NGHIÊN CỨU**

| **CÂU HỎI GỢI Ý** | **Mục tiêu nghiên cứu chính của đề tài là gì?**  **Phạm vi nghiên cứu là gì** |
| --- | --- |
| **HƯỚNG DẪN** | Đặt bài toán giải quyết và trình bày mục tiêu nghiên cứu chính của đề tài.  Nêu phạm vi nghiên cứu của đề tài, bao gồm việc giới hạn phạm vi nghiên cứu và triển khai, các giả định ban đầu đối với nghiên cứu. |
| 1. **Phạm vi nghiên cứu của đề tài :**  * **Nhiệm vụ chính :** Huấn luyện lại mô hình Inception-v3 của Google cho một nhiệm vụ mới là nhận diện khuôn mặt (facial recognition), thay vì phân loại các đối tượng chung trong ImageNet. * **Phạm vi triển khai :** * Không huấn luyện toàn bộ mạng nơ-ron (vì quá tốn tài nguyên), chỉ huấn luyện lại lớp cuối cùng (last layer). * Sử dụng mô hình đã được huấn luyện trước (pre-trained) để tận dụng khả năng nhận diện các đặc trưng cơ bản như hình dạng, màu sắc, kết cấu… * Dataset sử dụng: "Labeled Faces in the Wild" (LFW), giới hạn lại chỉ 10 người có nhiều hình ảnh nhất (mỗi người ít nhất 50 ảnh). * Sử dụng phiên bản hình ảnh đã được căn chỉnh bằng kỹ thuật deep funneling nhằm giảm sự khác biệt nội bộ trong cùng một lớp*(intra-class variability)***.** * **Môi trường thực hiện :** * Thực thi trên laptop cá nhân. * Gợi ý sử dụng Google Cloud GPU để tăng tốc độ nếu có điều kiện.  1. **Các giả định ban đầu đối với nghiên cứu:**  * Các lớp dưới của Inception-v3 đã đủ tốt để nhận diện đặc trưng thấp cấp (low-level features) nên chỉ cần thay đổi lớp trên cùng cho nhiệm vụ mới. * Chỉ cần một lượng dữ liệu huấn luyện nhỏ (10 lớp, mỗi lớp 50 ảnh) để mô hình học được nhiệm vụ phân loại khuôn mặt. * Hình ảnh trong cùng một lớp (cùng một người) nên được căn chỉnh (aligned) để mô hình dễ dàng học các đặc trưng thực sự của khuôn mặt thay vì bị nhiễu bởi góc chụp, ánh sáng, hay nền ảnh. * Mặc dù mạng Inception-v3 ban đầu không được huấn luyện với ảnh người, nhưng ta giả định rằng những đặc trưng học được từ ImageNet vẫn có thể thích ứng hiệu quả cho bài toán nhận diện người. | |
| **GHI CHÚ** | |

**TÀI LIỆU THAM KHẢO**

| **CÂU HỎI GỢI Ý** | **Các tài liệu tham khảo thuộc lĩnh vực nghiên cứu liên quan là gì?̀** |
| --- | --- |
| **HƯỚNG DẪN** | [1] A. Canziani, A. Paszke and E. Culurciello (2016), “An analysis of deep neural network models for practical applications”, arXiv preprint arXiv:1605.07678.ss  [2] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell Caffe (2014), “Convolutional Architecture for Fast Feature”,  Embedding arXiv:1408.5093.  [3] Y. Sun, X. Wang and X. Tang (2014), “Deep learning face representation by joint identification-verification”, CoRR, abs/1406.4773.  [4] Đoàn Hồng Quang, Lê Hồng Minh, Chu Anh Tuấn (2015), “Nhận dạng bàn tay bằng mạng nơ ron nhân tạo”, Tuyển tập báo cáo Diễn đàn “Đổi mới - Chìa khóa cho sự phát triển bền vững”, Viện Ứng dụng Công nghệ, Bộ Khoa học và Công nghệ.  [5] Đoàn Hồng Quang, Lê Hồng Minh (2014), “Dùng RFNN kết hợp khử mùa và khử xu hướng để dự báo chỉ số giá vàng trên thị trường”, Tuyển tập báo cáo Diễn đàn “Đổi mới - Chìa khóa cho sự phát triển bền vững”, Viện Ứng dụng Công nghệ, Bộ Khoa học và Công nghệ.  [6] Nguyễn Quang Hoan, Đoàn Hồng Quang (2014), “Dự báo chỉ số giá chứng khoán bằng RFNN”, Tạp chí Khoa học và Công nghệ, Trường Đại học Sư phạm Kỹ thuật Hưng Yên, 1, tr.52-56.  [7] Nguyễn Quang Hoan, Dương Thu Trang, Đoàn Hồng Quang (2018), “Dự báo số học sinh nhập trường bằng mạng nơ ron nhân tạo”, Tạp chí Khoa học và Công nghệ, Trường Đại học Sư phạm Kỹ thuật Hưng Yên, 18, tr.1-8.  Link tài liệu đã kham thảo :   1. <https://www.researchgate.net/publication/358161310_Facial_Recognition_Using_a_Lightweight_Deep_Neural_Networks> 2. <https://www.nist.gov/news-events/news/2019/12/nist-study-evaluates-effects-race-age-sex-face-recognition-software?utm_source=chatgpt.com> 3. <https://link.springer.com/article/10.1007/s11042-024-19521-0?utm_source=chatgpt.com> 4. <https://williamkoehrsen.medium.com/facial-recognition-using-googles-convolutional-neural-network-5aa752b4240e> |
|  | |
| **GHI CHÚ** | |

| **NGƯỜI THỰC HIỆN ĐỀ CƯƠNG** | | |
| --- | --- | --- |
| **SINH VIÊN THỰC HIỆN 1** | Lê Quang Vinh | 3120411176 |
| **SINH VIÊN THỰC HIỆN 2** | Vũ Thị Thanh Vân | 3120411173 |

| **XÁC NHẬN CỦA GIẢNG VIÊN HƯỚNG DẪN** | | |
| --- | --- | --- |
| **NGƯỜI HƯỚNG DẪN** | TS. Đỗ Như Tài | |
| **Ý KIẾN** |  | |
|  | ◻ đồng ý hướng dẫn | ◻ không đồng ý hướng dẫn |