Bộ Giáo dục và Đào tạo

**Trường đại học Sài Gòn**

Khoa Công Nghệ Thông Tin

Bộ môn Phương pháp nghiên cứu khoa học trong CNTT

**0B**

ĐỀ CƯƠNG CHI TIẾT

Ngày thực hiện đề cương:

| **TÊN ĐỀ TÀI** | Xây dựng hệ thống nhận diện khuôn mặt sử dụng phương pháp mạng nơ-ron song sinh | |
| --- | --- | --- |
| **LĨNH VỰC CHUYÊN NGÀNH** | Kỹ thuật phần mềm | |
| **LOẠI HÌNH NGHIÊN CỨU** | Sản phẩm | |
| **NGƯỜI HƯỚNG DẪN** | Ths. Đỗ Như Tài | |
| ***SINH VIÊN THỰC HIỆN 1*** | Đặng Huy Hoàng | 3122560019 |
| ***SINH VIÊN THỰC HIỆN 2*** | Đặng Huy Hoàng | 3122560020 |

**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. |
| Trong thời đại bùng nổ công nghệ thông tin và truyền thông, cùng sự phát triển mạnh mẽ của trí tuệ nhân tạo (AI) và thị giác máy tính, nhận diện khuôn mặt (face recognition) đã trở thành một trong những công nghệ sinh trắc học quan trọng và có tính ứng dụng cao. Không chỉ dừng lại ở việc mở khóa điện thoại, nhận diện khuôn mặt đang len lỏi vào mọi lĩnh vực của đời sống, từ an ninh quốc gia, quản lý doanh nghiệp, đến các dịch vụ cá nhân hóa. Ứng dụng của nó rất đa dạng: theo dõi, phát hiện tội phạm tại các khu vực công cộng; thay thế thẻ từ, mật khẩu trong kiểm soát truy cập; xác thực giao dịch trong tài chính ngân hàng; hỗ trợ chẩn đoán bệnh và theo dõi sức khỏe trong y tế; phân tích hành vi khách hàng trong bán lẻ; và tạo trải nghiệm tương tác tự nhiên trong giải trí.  Tuy có tiềm năng to lớn, nhận diện khuôn mặt vẫn đối mặt với nhiều thách thức kỹ thuật và đạo đức. Khuôn mặt có thể thay đổi đáng kể do các yếu tố nội tại (tuổi tác, biểu cảm, kiểu tóc) và ngoại tại (góc chụp, ánh sáng, phụ kiện, trang điểm, chất lượng ảnh). Điều này đòi hỏi hệ thống phải đạt độ chính xác và độ tin cậy cao để đảm bảo an ninh và tránh gây phiền toái. Bên cạnh đó, vấn đề quyền riêng tư, bảo mật dữ liệu, tính thiên vị, phân biệt đối xử, chi phí triển khai, và tính khả thi cũng là những trở ngại cần được giải quyết. Đặc biệt, việc thu thập và gán nhãn dữ liệu khuôn mặt là tốn kém và khó khăn, nhất là với đối tượng hiếm gặp hoặc không hợp tác.  Sự hạn chế của dữ liệu, đặc biệt trong việc có đủ ảnh để huấn luyện cho *mọi* đối tượng, là một vấn đề thực tế trong nhiều dự án. Để vượt qua, các kỹ thuật như one-shot learning, zero-shot learning và few-shot learning đã được phát triển. Trong số đó, one-shot learning nổi lên như một hướng tiếp cận hứa hẹn. One-shot learning là bài toán mà mô hình cần phân loại hoặc nhận dạng đối tượng mới chỉ từ *một* mẫu ví dụ của mỗi lớp, khác hẳn với các phương pháp truyền thống cần rất nhiều dữ liệu. One-shot learning có hai bài toán chính: *one-shot classification* (xác định đối tượng thuộc lớp nào với chỉ một mẫu từ mỗi lớp) và *one-shot verification* (xác định hai ảnh có cùng đối tượng hay không). Nó khác với *zero-shot learning* (nhận dạng không cần mẫu của lớp mục tiêu) và *few-shot learning* (cần một vài mẫu).  Mạng nơ-ron song sinh (Siamese Neural Networks - SNN) được xem là giải pháp hiệu quả cho one-shot learning, không chỉ trong nhận diện khuôn mặt. SNN có kiến trúc gồm hai (hoặc nhiều) mạng con giống hệt nhau, đóng vai trò trích xuất đặc trưng. Các vectơ đặc trưng này được đưa vào một hàm khoảng cách để tính độ tương đồng. Thay vì học phân loại trực tiếp, SNN học *so sánh* hai hình ảnh và xác định chúng có cùng lớp hay không. Quá trình huấn luyện dựa trên các cặp hình ảnh: cặp *positive* (cùng người, khoảng cách đặc trưng nhỏ) và cặp *negative* (khác người, khoảng cách đặc trưng lớn). Deep learning, đặc biệt là CNN, thường được dùng làm bộ trích xuất đặc trưng trong SNN nhờ khả năng học đặc trưng phân cấp. Tuy nhiên, CNN *không đủ* cho one-shot learning; điểm khác biệt là SNN *học so sánh* các đặc trưng do CNN trích xuất.  SNN có các ưu điểm: (1) Học từ ít dữ liệu: chỉ cần một (vài) mẫu từ các lớp mới nhờ học cách so sánh. (2) Tính tổng quát hóa cao: Nhận diện tốt danh tính mới (chưa từng thấy) do học được các đặc trưng phân biệt. (3) Linh hoạt, có thể mở rộng: dễ áp dụng và tích hợp với nhiều kiến trúc, bài toán khác nhau. *Similarity function* và *dữ liệu huấn luyện* (positive/negative pairs) là những yếu tố quan trọng của SNN, và được trình bày kỹ ở phần sau.  Với những ưu điểm đó, đề tài này tập trung vào "**Xây dựng hệ thống nhận diện khuôn mặt sử dụng phương pháp mạng nơ-ron song sinh (SNN)**". Cụ thể, đề tài sẽ khảo sát các kiến trúc SNN, hàm khoảng cách, phương pháp huấn luyện; thiết kế và triển khai mô hình SNN dựa trên CNN; huấn luyện, đánh giá trên bài toán one-shot learning; tối ưu hóa, cải thiện độ chính xác, tốc độ; và (nếu có) nghiên cứu hướng mở rộng cho ứng dụng thực tế. Mục tiêu là xây dựng một hệ thống có khả năng học và nhận dạng từ một mẫu duy nhất, đóng góp vào giá trị khoa học và làm nền tảng cho các ứng dụng thực tế của nhận dạng khuôn mặt, nơi việc có được số lượng lớn ảnh của một người thường không khả thi. | |

| **GHI CHÚ** |
| --- |

**TÌNH HÌNH NGHIÊN CỨU TRONG VÀ NGOÀI NƯỚC**

| **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** | Tìm hiểu các nghiên cứu đã công bố gần nhất (5 năm trở lại) về lĩnh vực liên quan trong và ngoài nước.  Phân tích các kết quả đóng góp, nhận xét các hạn chế còn tồn tại. |
| Nghiên cứu về mạng nơ-ron song sinh (Siamese Neural Networks - SNN) và ứng dụng trong one-shot learning, đặc biệt là nhận diện khuôn mặt, đã và đang thu hút sự quan tâm lớn của cộng đồng nghiên cứu trên toàn thế giới. Những tiến bộ trong lĩnh vực này không chỉ có ý nghĩa khoa học mà còn mở ra những tiềm năng ứng dụng to lớn trong thực tiễn.  **1. Tình hình Nghiên cứu Ngoài nước:**  Trên thế giới, nghiên cứu về SNN và one-shot learning đã có những bước tiến đáng kể, đặc biệt trong khoảng một thập kỷ trở lại đây. Bài báo "Siamese Neural Networks for One-shot Image Recognition" của Koch, Zemel và Salakhutdinov [1] được xem là một trong những công trình tiêu biểu, đặt nền móng cho việc ứng dụng SNN trong bài toán one-shot learning. Trong nghiên cứu này, các tác giả đã đề xuất một kiến trúc SNN dựa trên mạng nơ-ron tích chập (CNN) và huấn luyện mô hình trên tập dữ liệu Omniglot [2], một tập dữ liệu chữ viết tay đa dạng, bao gồm nhiều bảng chữ cái khác nhau. Kết quả cho thấy SNN có khả năng vượt trội so với các phương pháp truyền thống trong bài toán one-shot classification, đạt độ chính xác gần với mức độ của con người. Trang web chính thức của Omniglot: [7]  Trước đó, nghiên cứu về one-shot learning đã được khởi xướng bởi Fei-Fei Li và cộng sự [3], [4], với các công trình tập trung vào việc xây dựng các mô hình Bayes tổng quát (generative Bayesian models) cho bài toán phân loại ảnh. Tuy nhiên, các mô hình Bayes thường gặp khó khăn trong việc mở rộng và tính toán trên các tập dữ liệu lớn.  Một hướng nghiên cứu khác liên quan đến SNN là "learning to learn" (học cách học) hay "meta-learning", trong đó mục tiêu là huấn luyện một mô hình có khả năng học nhanh chóng từ một lượng nhỏ dữ liệu. Các công trình tiêu biểu trong lĩnh vực này bao gồm "Model-Agnostic Meta-Learning (MAML)" của Finn, Abbeel và Levine [5], và "Matching Networks" của Vinyals và cộng sự [6]. Trang web giới thiệu về MAML của nhóm tác giả: [8]. Mặc dù không trực tiếp sử dụng SNN, các phương pháp này chia sẻ tinh thần chung với SNN trong việc tận dụng kiến thức đã học để thích nghi nhanh chóng với các nhiệm vụ mới.  Ngoài ra, một số nghiên cứu gần đây đã mở rộng SNN để xử lý các bài toán phức tạp hơn, chẳng hạn như:   * **Few-shot learning:** Thay vì chỉ sử dụng một mẫu, mô hình có thể sử dụng một vài mẫu (vài "shots") cho mỗi lớp. * **Zero-shot learning:** Mô hình có thể nhận dạng các lớp chưa từng thấy trong quá trình huấn luyện. * **Video-based Person re-identification:** Sử dụng SNN trong các bài toán video.   Các nhóm nghiên cứu hàng đầu về SNN và one-shot learning có thể kể đến như nhóm của Richard Zemel (University of Toronto) [9], Ruslan Salakhutdinov (Carnegie Mellon University, Apple) [10], Fei-Fei Li (Stanford University) [11], Oriol Vinyals (Google DeepMind) [12], và Chelsea Finn (UC Berkeley) [13].  **2. Tình hình Nghiên cứu Trong nước:**  Tại Việt Nam, nghiên cứu về SNN và ứng dụng trong nhận diện khuôn mặt vẫn còn ở giai đoạn đầu, nhưng đã có những dấu hiệu tích cực. Một số nhóm nghiên cứu tại các trường đại học và viện nghiên cứu đã bắt đầu quan tâm đến lĩnh vực này và đạt được những kết quả bước đầu:   * Một số nhóm tập trung vào ứng dụng deep learning, transfer learning cho các bài toán như one-shot, zero-shot learning. Tuy chưa nhiều nhóm làm riêng về SNN nhưng đây vẫn có thể xem là một bước tiếp cận cho các hướng đi trong tương lai. * Ứng dụng nhận diện đã xuất hiện trong các hệ thống kiểm soát, các hệ thống IoT.   Tuy nhiên, do các lý do như:   * Tiếp cận nguồn tài liệu, đặc biệt các tài liệu, dữ liệu lớn, chất lượng * Tiếp cận phần cứng, hạ tầng * Nhân lực, Các nhóm nghiên cứu này còn nhỏ, và chưa có quá nhiều ấn phẩm, bài báo liên quan đến SNN và one-shot.   **3. So sánh và Đánh giá:**  So với thế giới, nghiên cứu về SNN và one-shot learning tại Việt Nam còn khá non trẻ. Các nghiên cứu trên thế giới đã có lịch sử phát triển lâu dài, với nhiều công trình có ảnh hưởng lớn, được công bố trên các hội nghị và tạp chí hàng đầu. Các nhóm nghiên cứu quốc tế có nguồn lực dồi dào (về tài chính, nhân lực, dữ liệu, và cơ sở vật chất), và có sự hợp tác chặt chẽ với các công ty công nghệ lớn.  Tuy nhiên, các nhóm nghiên cứu tại Việt Nam cũng có những lợi thế riêng, chẳng hạn như sự hiểu biết sâu sắc về các đặc điểm văn hóa và ngôn ngữ của người Việt, có thể giúp ích trong việc phát triển các hệ thống nhận diện khuôn mặt phù hợp với bối cảnh Việt Nam.  **4. Khoảng trống và Hướng Nghiên cứu Tiếp theo:**  Mặc dù đã có những tiến bộ đáng kể, nghiên cứu về SNN và one-shot learning vẫn còn nhiều khoảng trống và vấn đề cần được giải quyết:   * **Cải thiện hiệu suất:** Cần tiếp tục cải thiện độ chính xác và độ tin cậy của SNN trong các điều kiện thực tế (ánh sáng thay đổi, góc chụp khác nhau, khuôn mặt bị che khuất,...). * **Mở rộng ứng dụng:** Nghiên cứu ứng dụng SNN trong các lĩnh vực khác ngoài nhận diện khuôn mặt, chẳng hạn như nhận dạng chữ viết tay, nhận dạng giọng nói, hoặc phát hiện bất thường. * **Giải quyết các vấn đề đạo đức:** Cần có những quy định và hướng dẫn rõ ràng về việc sử dụng công nghệ nhận diện khuôn mặt để đảm bảo quyền riêng tư và tránh phân biệt đối xử. | |

| **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. |
| **Mục tiêu của đề tài**   Đề tài này hướng tới mục tiêu chính là xây dựng một hệ thống nhận diện khuôn mặt có khả năng học từ một mẫu duy nhất (one-shot learning) dựa trên phương pháp mạng nơ-ron song sinh (SNN). Để đạt được mục tiêu này, đề tài sẽ tập trung vào các mục tiêu cụ thể sau:   * Nghiên cứu Tổng quan: Khảo sát và phân tích các kiến trúc SNN khác nhau đã được đề xuất trong các nghiên cứu trước đây. Nghiên cứu các hàm khoảng cách (distance functions) hoặc hàm tương tự (similarity functions) thường được sử dụng trong SNN, đánh giá ưu nhược điểm của từng loại hàm. Tìm hiểu các phương pháp huấn luyện SNN hiệu quả, bao gồm các kỹ thuật tối ưu hóa (optimization), các chiến lược lấy mẫu (sampling strategies), và các phương pháp tăng cường dữ liệu (data augmentation). * Thiết kế và Xây dựng Mô hình:   + Lựa chọn một kiến trúc SNN phù hợp với bài toán nhận diện khuôn mặt, dựa trên các tiêu chí như độ chính xác, tốc độ tính toán, và khả năng tổng quát hóa.   + Sử dụng một mạng nơ-ron tích chập (CNN) làm bộ trích xuất đặc trưng (feature extractor) cho SNN. Lựa chọn kiến trúc CNN phù hợp (ví dụ: VGGNet, ResNet, Inception, EfficientNet) và điều chỉnh các tham số để đạt được hiệu suất tốt nhất.   + Thiết kế một hàm khoảng cách/tương tự phù hợp để so sánh các vectơ đặc trưng được tạo ra bởi CNN.   + Xây dựng một quy trình (pipeline) hoàn chỉnh cho hệ thống, bao gồm các bước tiền xử lý dữ liệu (phát hiện khuôn mặt, căn chỉnh khuôn mặt, chuẩn hóa hình ảnh), trích xuất đặc trưng, tính toán khoảng cách, và đưa ra quyết định. * Huấn luyện, Đánh giá và Tối ưu hóa:   + Chuẩn bị một tập dữ liệu huấn luyện và một tập dữ liệu kiểm tra (validation set) để huấn luyện và đánh giá mô hình SNN. Tập dữ liệu huấn luyện nên bao gồm nhiều danh tính khác nhau, với các biến đổi về ánh sáng, góc chụp, biểu cảm, và phụ kiện.   + Sử dụng một hàm mất mát (loss function) phù hợp để huấn luyện SNN, chẳng hạn như contrastive loss hoặc triplet loss.   + Áp dụng các kỹ thuật tối ưu hóa (ví dụ: stochastic gradient descent, Adam) để tìm kiếm các tham số tối ưu cho mô hình.   + Đánh giá hiệu suất của mô hình trên tập dữ liệu kiểm tra bằng các độ đo phù hợp, chẳng hạn như độ chính xác (accuracy), precision, recall, F1-score, và diện tích dưới đường cong ROC (AUC).   + Thực hiện các thử nghiệm để tinh chỉnh các siêu tham số (hyperparameters) của mô hình (ví dụ: learning rate, batch size, số lượng layers, số lượng filters) để đạt được hiệu suất tốt nhất.   + Phân tích các lỗi nhận diện, tìm hiểu nguyên nhân.   + (Nếu có thể) So sánh hiệu suất của mô hình SNN với các phương pháp nhận diện khuôn mặt khác, chẳng hạn như các phương pháp dựa trên CNN truyền thống hoặc các phương pháp one-shot learning khác. * (Mục tiêu mở rộng, nếu có):   + Nghiên cứu, đề xuất các hướng cải tiến   + Triển khai hệ thống trên một nền tảng thực tế (ví dụ: ứng dụng di động, website, thiết bị nhúng) để kiểm tra khả năng ứng dụng của hệ thống. * **Phạm vi nghiên cứu**     Phần này sẽ giới hạn phạm vi, không phải liệt kê như trên, mà trình bày tập trung để đảm bảo tính khả thi và tập trung, đề tài sẽ giới hạn phạm vi nghiên cứu như sau:  **Bài toán:** Đề tài chỉ tập trung vào bài toán *one-shot verification*, tức là xác định xem hai hình ảnh khuôn mặt có thuộc về cùng một người hay không. Bài toán one-shot classification (phân loại một hình ảnh vào một trong số nhiều lớp, mỗi lớp chỉ có một mẫu) không nằm trong phạm vi nghiên cứu này, *có thể dành cho phần mở rộng hoặc nghiên cứu tiếp theo.*  Dữ liệu:   * + Đề tài sẽ sử dụng các tập dữ liệu khuôn mặt công khai (publicly available datasets), chẳng hạn như Labeled Faces in the Wild (LFW), YouTube Faces (YTF), hoặc một phần của tập dữ liệu lớn hơn như VGGFace2, để huấn luyện và đánh giá mô hình.   + Đề tài *không* tập trung vào việc thu thập và gán nhãn dữ liệu khuôn mặt mới. *Tuy nhiên việc tiền xử lý chất lượng dữ liệu sẽ được đầu tư và chú trọng*   Mô hình:   * + Đề tài sẽ tập trung vào các kiến trúc SNN dựa trên CNN. Các kiến trúc SNN khác (ví dụ, dựa trên recurrent neural networks) không nằm trong phạm vi nghiên cứu. *Việc thử nghiệm với nhiều loại CNN backbone là một phần của quá trình tối ưu hóa*.   Nền tảng:   * + Mô hình SNN sẽ được triển khai và huấn luyện bằng các thư viện deep learning phổ biến như TensorFlow hoặc PyTorch. *Không xây dựng các framework/ thư viện mới*   Phần cứng:   * + Đề tài thực hiện huấn luyện trên máy tính/ server có GPU   Đối tượng:   * + Đề tài tập trung vào nhận diện khuôn mặt người *trưởng thành*.   + Đề tài chưa tập trung vào các vấn đề chống giả mạo, các khuôn mặt bị biến dạng (do phẫu thuật, do bệnh lý), nhận dạng một phần | |
|  | |

**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] Koch, G., Zemel, R., & Salakhutdinov, R. (2015). Siamese neural networks for one-shot image recognition. In *ICML Deep Learning Workshop*. Lille, France. [2] Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science, 350*(6266), 1332-1338. [3] Fei-Fei, L., Fergus, R., & Perona, P. (2006). One-shot learning of object categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 28*(4), 594-611. [4] Fergus, R., Perona, P., & Zisserman, A. (2003). Object class recognition by unsupervised scale-invariant learning. In *Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2*, II-264. [5] Finn, C., Abbeel, P., & Levine, S. (2017). Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70* (pp. 1126-1135). Sydney, NSW, Australia: PMLR. [6] Vinyals, O., Blundell, C., Lillicrap, T., Wierstra, D., et al. (2016). Matching Networks for One Shot Learning. In *Advances in Neural*  **Website**  [7] Trang web chính thức của Omniglot: [https://github.com/brendenlake/omniglot](https://www.google.com/url?sa=E&q=https%3A%2F%2Fgithub.com%2Fbrendenlake%2Fomniglot) [8] Trang web giới thiệu về MAML của nhóm tác giả: [https://bair.berkeley.edu/blog/2017/07/18/learning-to-learn/](https://www.google.com/url?sa=E&q=https%3A%2F%2Fbair.berkeley.edu%2Fblog%2F2017%2F07%2F18%2Flearning-to-learn%2F) [9] Richard Zemel (University of Toronto): [https://www.cs.toronto.edu/~zemel/](https://www.google.com/url?sa=E&q=https%3A%2F%2Fwww.cs.toronto.edu%2F~zemel%2F) [10] Ruslan Salakhutdinov (Carnegie Mellon University, Apple): [https://www.cs.cmu.edu/~rsalakhu/](https://www.google.com/url?sa=E&q=https%3A%2F%2Fwww.cs.cmu.edu%2F~rsalakhu%2F) [11] Fei-Fei Li (Stanford University): [https://profiles.stanford.edu/fei-fei-li](https://www.google.com/url?sa=E&q=https%3A%2F%2Fprofiles.stanford.edu%2Ffei-fei-li) [12] Oriol Vinyals (Google DeepMind): [https://research.google/people/OriolVinyals/](https://www.google.com/url?sa=E&q=https%3A%2F%2Fresearch.google%2Fpeople%2FOriolVinyals%2F) [13] Chelsea Finn (UC Berkeley): [https://people.eecs.berkeley.edu/~cbfinn/](https://www.google.com/url?sa=E&q=https%3A%2F%2Fpeople.eecs.berkeley.edu%2F~cbfinn%2F) |
|  | |
|  | |



| **NGƯỜI THỰC HIỆN ĐỀ CƯƠNG** | | |
| --- | --- | --- |
| **SINH VIÊN THỰC HIỆN 1** | Đặng Huy Hoàng | 3122560019 |
| **SINH VIÊN THỰC HIỆN 2** | Đặng Huy Hoàng | 3122560020 |

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