| Bộ Giáo dục và Đào tạo  **Trường Đại Học Sài Gòn** | Khoa Công Nghệ Thông Tin  *Bộ môn PPNCKH trong CNTT* |
| --- | --- |

**ĐỀ CƯƠNG CHI TIẾT**

Ngày thực hiện đề cương:

| **TÊN ĐỀ TÀI** | NGHIÊN CỨU NHẬN DẠNG BIỂU CẢM KHUÔN MẶT BẰNG PHƯƠNG PHÁP HỌC SÂU SỬ DỤNG KIẾN TRÚC RESNET | |
| --- | --- | --- |
| **LĨNH VỰC CHUYÊN NGÀNH** | CÔNG NGHỆ THÔNG TIN | |
| **LOẠI HÌNH NGHIÊN CỨU** | SẢN PHẨM | |
| **NGƯỜI HƯỚNG DẪN** | TS.ĐỖ NHƯ TÀI | |
| ***SINH VIÊN THỰC HIỆN 1*** | Đỗ Đình Thái | 3122410378 |
| ***SINH VIÊN THỰC HIỆN 2*** | Đào Thành Phát | 3122410290 |
| ***SINH VIÊN THỰC HIỆN 3*** | Nguyễn Minh Trí | 3122410424 |
| ***SINH VIÊN THỰC HIỆN 4*** | Nguyễn Trọng Phú | 3122410316 |

**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. |
| Ngày nay, nhận dạng cảm xúc khuôn mặt đang trở thành một lĩnh vực quan trọng trong công nghệ thông tin (CNTT), với tiềm năng ứng dụng rộng rãi trong nhiều ngành nghề như giáo dục, y tế, marketing, và dịch vụ khách hàng. Công nghệ này cho phép máy tính phân tích và nhận diện các biểu cảm trên khuôn mặt con người, từ đó hiểu được trạng thái cảm xúc như vui vẻ, buồn bã, tức giận, hay ngạc nhiên. Điều này không chỉ cải thiện tương tác giữa con người và máy móc mà còn mở ra nhiều cơ hội để cá nhân hóa trải nghiệm người dùng.  Hiện nay, các phương pháp học máy (machine learning) và học sâu (deep learning) đã mang lại những bước tiến vượt bậc trong việc nhận dạng cảm xúc khuôn mặt. Các mô hình mạng nơ-ron tích chập (Convolutional Neural Networks - CNN) như VGG, Inception, hay ResNet được sử dụng rộng rãi để phân tích hình ảnh khuôn mặt và nhận diện cảm xúc với độ chính xác cao. Ví dụ, một số nghiên cứu gần đây đã đạt được độ chính xác trên 70% khi sử dụng các tập dữ liệu lớn như AffectNet hay FER2013. Công nghệ này không chỉ dừng lại ở việc nhận diện cảm xúc mà còn được tích hợp vào các hệ thống thông minh, chẳng hạn như robot giao tiếp, hệ thống giám sát an ninh, hay các ứng dụng hỗ trợ sức khỏe tâm lý.  Tại Việt Nam, nhận dạng cảm xúc khuôn mặt cũng đang được quan tâm và phát triển. Các trường đại học và viện nghiên cứu đã bắt đầu ứng dụng công nghệ này vào các dự án thực tiễn, như xây dựng hệ thống đánh giá cảm xúc của học sinh trong lớp học trực tuyến, hoặc hỗ trợ các doanh nghiệp phân tích phản hồi của khách hàng thông qua biểu cảm khuôn mặt. Tuy nhiên, để áp dụng hiệu quả, cần có thêm các tập dữ liệu đa dạng hơn, phản ánh đặc điểm khuôn mặt và cảm xúc của người Việt, cũng như cải thiện khả năng xử lý trong điều kiện ánh sáng và góc quay phức tạp. | |
|  | |
| 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. |
| Trong những năm gần đây, nhận dạng cảm xúc khuôn mặt đã trở thành một lĩnh vực nghiên cứu nổi bật trong công nghệ thông tin (CNTT), đặc biệt với sự phát triển của học máy (machine learning) và học sâu (deep learning). Trên thế giới, các nghiên cứu về nhận dạng cảm xúc khuôn mặt đã đạt được nhiều thành tựu đáng kể. Các mô hình mạng nơ-ron tích chập (CNN) như ResNet, VGG, và Inception đã được áp dụng rộng rãi, đạt độ chính xác cao trên các tập dữ liệu lớn như FER2013 và AffectNet. Ví dụ, một nghiên cứu tại Mỹ đã sử dụng ResNet50 để nhận diện 7 loại cảm xúc cơ bản (vui, buồn, tức giận, ngạc nhiên, sợ hãi, ghê tởm, trung tính) với độ chính xác lên đến 72% trên tập dữ liệu FER2013. Ngoài ra, các ứng dụng thực tiễn của công nghệ này cũng rất đa dạng, từ phân tích cảm xúc khách hàng trong marketing, hỗ trợ chẩn đoán tâm lý trong y tế, đến cải thiện tương tác người-máy trong robot thông minh.  Tại Việt Nam, nhận dạng cảm xúc khuôn mặt cũng đang nhận được sự quan tâm lớn từ cộng đồng nghiên cứu. Các trường đại học và viện nghiên cứu đã bắt đầu triển khai các dự án ứng dụng công nghệ này vào thực tiễn, như phân tích cảm xúc học sinh trong lớp học trực tuyến hoặc hỗ trợ doanh nghiệp đánh giá mức độ hài lòng của khách hàng. Tuy nhiên, các nghiên cứu tại Việt Nam vẫn còn gặp nhiều thách thức, bao gồm sự thiếu hụt tập dữ liệu lớn phản ánh đặc điểm khuôn mặt và cảm xúc của người Việt, cũng như hạn chế về khả năng xử lý trong điều kiện ánh sáng và góc quay phức tạp. Một số nghiên cứu tại Việt Nam đã thử nghiệm các mô hình học sâu như ResNet18 và đạt độ chính xác khoảng 65% trên tập dữ liệu tự xây dựng, nhưng vẫn cần cải thiện để đáp ứng nhu cầu thực tiễn.  Để thúc đẩy sự phát triển của nhận dạng cảm xúc khuôn mặt tại Việt Nam, cần tập trung vào việc xây dựng các tập dữ liệu đa dạng và chất lượng cao, đồng thời tối ưu hóa các mô hình học sâu để phù hợp với đặc điểm của người Việt. Ngoài ra, việc tích hợp công nghệ này vào các hệ thống thực tiễn như giáo dục trực tuyến (e-learning) cũng là một hướng đi tiềm năng. Ví dụ, một hệ thống e-learning có thể sử dụng nhận dạng cảm xúc để đánh giá mức độ tập trung và hứng thú của học sinh, từ đó điều chỉnh nội dung giảng dạy cho phù hợp. Tuy nhiên, để triển khai hiệu quả, cần giải quyết các vấn đề về quyền riêng tư, bảo mật dữ liệu, và sự chấp nhận của người dùng. | |
|  | |
| 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. |
| Với bài toán nhận diện cảm xúc khuôn mặt, mục tiêu của đề tài là phát triển một hệ thống AI có khả năng phân tích biểu cảm của người dùng thông qua hình ảnh hoặc video theo thời gian thực. Hệ thống này hướng đến các chức năng chính như:  -**Phát hiện và phân loại cảm xúc** (vui, buồn, giận dữ, ngạc nhiên, sợ hãi, chán nản, trung lập) từ khuôn mặt người.  **-Cải thiện độ chính xác của mô hình nhận diện cảm xúc** bằng cách sử dụng thuật toán học sâu (Deep Learning) và tập dữ liệu đa dạng.  **-Tích hợp hệ thống vào các ứng dụng thực tế**, như lớp học trực tuyến, chăm sóc sức khỏe tinh thần, hoặc hỗ trợ giám sát an toàn giao thông.  Về phạm vi nghiên cứu: đề tài khóa luận được nghiên cứu triển khai và thử nghiệm trong các học phần thuộc bộ môn Phương Pháp Nghiên Cứu, Khoa Công nghệ thông tin – Trường Đại học Sài Gòn. | |
|  | |
| GHI CHÚ | |
|  | |

**GIẢ THUYẾT VÀ CÁCH TIẾP CẬN**

| CÂU HỎI GỢI Ý | **Các giả thuyết đặt ra để giải quyết bài toán chính?**  **Các cách tiếp cận để giải quyết bài toán đặt ra?** |
| --- | --- |
| HƯỚNG DẪN | - Đặt ra những giả thuyết, hay vấn đề bài toán con cần phải giải quyết để đạt được mục tiêu nghiên cứu đề tài.  - Các cách tiếp cận (dự kiến) để giải quyết các giả thuyết, bài toán con đã đặt. |
| **1. Giả thuyết về sự đa dạng của biểu cảm khuôn mặt**  * **Bài toán con:**   + Cảm xúc con người có thể biểu hiện khác nhau tùy vào văn hóa, độ tuổi, giới tính, cá nhân.   + Một số biểu cảm có thể bị che khuất (do khẩu trang, kính, tóc, góc nghiêng). * **Cách tiếp cận:**   + Thu thập dữ liệu đa dạng từ nhiều nguồn (FER-2013, AffectNet, CK+).   + Sử dụng các kỹ thuật Data Augmentation (xoay, crop, thay đổi ánh sáng) để tăng tính tổng quát.   + Áp dụng mô hình đa nhiệm (multi-task learning) để nhận diện cảm xúc cùng với các yếu tố khác như tuổi, giới tính.  **2. Giả thuyết về độ phức tạp của các đặc trưng cảm xúc**  * **Bài toán con:**   + Các đặc trưng cảm xúc có thể rất tinh tế (ví dụ: sự khác biệt giữa "hạnh phúc" và "mỉm cười gượng").   + Nhiễu trong ảnh (ánh sáng thay đổi, độ phân giải thấp) làm giảm độ chính xác. * **Cách tiếp cận:**   + Sử dụng các mô hình CNN (ResNet, VGG) hoặc Transformer (ViT) để trích xuất đặc trưng sâu.   + Kết hợp Landmark khuôn mặt (dùng OpenFace, Dlib) để tập trung vào vùng biểu cảm quan trọng (mắt, miệng, lông mày).   + Áp dụng Attention Mechanism để tập trung vào các vùng quan trọng.  **3. Giả thuyết về sự mất cân bằng dữ liệu**  * **Bài toán con:**   + Một số cảm xúc hiếm gặp (ví dụ: "sợ hãi", "ghê tởm") có ít dữ liệu huấn luyện hơn so với "vui", "buồn". * **Cách tiếp cận:**   + Sử dụng kỹ thuật Oversampling (SMOTE) hoặc Undersampling để cân bằng dữ liệu.   + Áp dụng Loss Function điều chỉnh trọng số (Weighted Cross-Entropy, Focal Loss).  **4. Giả thuyết về hiệu suất thời gian thực**  * **Bài toán con:**   + Ứng dụng thời gian thực (real-time) yêu cầu tốc độ xử lý nhanh trên thiết bị edge (điện thoại, camera). * **Cách tiếp cận:**   + Sử dụng các kiến trúc nhẹ (MobileNet, EfficientNet).   + Quantization (chuyển đổi từ FP32 → INT8), Pruning để giảm kích thước mô hình.   + Triển khai trên nền tảng tối ưu (TensorRT, ONNX Runtime).  **5. Giả thuyết về tính khái quát hóa của mô hình**  * **Bài toán con:**   + Mô hình huấn luyện trên lab data có thể hoạt động kém trên dữ liệu thực tế (do khác biệt về môi trường). * **Cách tiếp cận:**   + Sử dụng Transfer Learning từ các mô hình pre-trained (ví dụ: VGG-Face).   + Fine-tuning trên tập dữ liệu gần với ứng dụng thực tế.   + Áp dụng Domain Adaptation (CycleGAN để chuyển đổi phong cách ảnh). | |
|  | |
| 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] Li, S.; Deng, W. (2020). Deep Facial Expression Recognition: A Survey. *IEEE Transactions on Affective Computing*, 11(1), 1-18.  [2] Krizhevsky, A.; Sutskever, I.; Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, 25, 1097-1105.  [3] Mollahosseini, A.; Hasani, B.; Mahoor, M. H. (2019). AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild. *IEEE Transactions on Affective Computing*, 10(1), 18-31.  [4] Goodfellow, I. J.; Erhan, D.; Carrier, P. L.; et al. (2013). Challenges in Representation Learning: A Report on Three Machine Learning Contests. *Neural Networks*, 64, 59-63.  [5] Ekman, P.; Friesen, W. V. (1978). Facial Action Coding System: A Technique for the Measurement of Facial Movement. *Consulting Psychologists Press*. |
| ✻ **TIẾNG ANH**  [1] Goodfellow, I. J.; Erhan, D.; Carrier, P. L.; et al. (2013). Challenges in Representation Learning: A Report on Three Machine Learning Contests. *Neural Networks*, 64, 59-63.  [2] Li, S.; Deng, W. (2020). Deep Facial Expression Recognition: A Survey. *IEEE Transactions on Affective Computing*, 11(1), 1-18.  [3] Mollahosseini, A.; Hasani, B.; Mahoor, M. H. (2019). AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild. *IEEE Transactions on Affective Computing*, 10(1), 18-31.  [4] Ekman, P.; Friesen, W. V. (1978). Facial Action Coding System: A Technique for the Measurement of Facial Movement. *Consulting Psychologists Press*.  [5] Krizhevsky, A.; Sutskever, I.; Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, 25, 1097-1105.  [6] Dosovitskiy et al. (2020) – *"An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale"*  [7] Baltrusaitis et al. (2018) – *"OpenFace 2.0: Facial Behavior Analysis Toolkit"*  [8] Howard et al. (2017) – *"MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications"*  [9] Chawla et al. (2002) – *"SMOTE: Synthetic Minority Over-sampling Technique"*  [10] Lin et al. (2017) – *"Focal Loss for Dense Object Detection"* (xử lý class imbalance)  [11] Han et al. (2015) – *"Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding"*  [12] He, K.; Zhang, X.; Ren, S.; Sun, J. (2016). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770-778.  [13] Simonyan, K.; Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv preprint arXiv:1409.1556*.  [14] Poria, S.; Cambria, E.; Bajpai, R.; Hussain, A. (2017). A Review of Affective Computing: From Unimodal Analysis to Multimodal Fusion. *Information Fusion*, 37, 98-125.  [15] Zhang, Z.; Luo, P.; Loy, C. C.; Tang, X. (2018). From Facial Expression Recognition to Interpersonal Relation Prediction. *International Journal of Computer Vision*, 126(5), 550-569.  [16] Garrison, D. R. (2011). E-Learning in the 21st Century: A Framework for Research and Practice. *Routledge*. | |
|  | |
| GHI CHÚ | |
|  | |

| **NGƯỜI THỰC HIỆN ĐỀ CƯƠNG** | | |
| --- | --- | --- |
| ***SINH VIÊN THỰC HIỆN 1*** | Đỗ Đình Thái | 3122410378 |
| ***SINH VIÊN THỰC HIỆN 2*** | Đào Thành Phát | 3122410290 |
| ***SINH VIÊN THỰC HIỆN 3*** | Nguyễn Minh Trí | 3122410424 |
| ***SINH VIÊN THỰC HIỆN 4*** | Nguyễn Trọng Phú | 3122410316 |

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