**Hãy tìm hiểu cho bên dưới và tóm tắt lạicác nội dung sau:**

**Tên đồ án;** Tác giả Phạm Quí Luận , Nơi công bố: Đại học Bách Khoa HCM, Năm công bố 2019

**Mục tiêu nghiên cứu**: cải tiến độ chính xác trên tập dữ liệu, tác giả thực hiện làm trên 1 bộ dữ liệu về khuôn mặt chứa hình ảnh người Việt Nam.

Cơ sở lý thuyết;

Phương pháp thu thập và xử lý thông tin;

Kết quả đạt được;

Hạn chế của công trình.

Tóm tắt

NHẬN DIỆN CẢM XÚC MẶT NGƯỜI SỬ DỤNG MẠNG HỌC SÂU CÓ CHÚ Ý

Giảng viên hướng dẫn: TS. Trần Tuấn Anh

Giảng viên phản biện: TS. Nguyễn Hồ Mẫn Rạng

Sinh viên thực hiện: Phạm Quí Luận

* Mở đầu bằng lời cam đoan
* Lời cảm ơn
* Tóm tắt luận văn

Luận văn này sẽ sử dụng một mạng học sâu có sử dụng cơ chế chú ý - ResidualMasking Network để phân lớp cảm xúc dựa trên ảnh đầu vào trong môi trường phức tạp và kết hợp với nhiều mô hình hiện đại để tăng cường độ chính xác. Tác giả có xây dụng thêm và phát triện một bộ dữ liệu có mặt người Viêt Nam nhằm góm phần phất triển bài toán này ở nước ta. Kết quả thực nghiệm cho thấy 2 phương pháp tác giả đề xuất đều có kết quả tốt hơn 1 số phương pháp hiện đại trong bài toán nhận diện cảm xúc cho ảnh dầu vào phức tạp và 1 số kết quả được báo cáo trong các nckh. Riêng pp học kết hợp cho ra độ chính xác tốt nhất hiện nay- 76.82% trên tập dữ liệu FER2013

* Mục lục
* Danh sách hình vẽ

Tác giả khát quát danh sách các hình vẽ trong luận văn làm và thiết kế như là một mục lục vậy đi kèm là có các trích dẫn [] để chắc rằng các tài liệu này đã đc chứng minh và tham khảo từ các nguồn có uy tín.

* Danh sách bảng

Giống với danh sách hình vẽ bảng này có mục đích là so sánh các độ chính xác của các mô hình, các kết quả của tập dữ liệu nào đó và Cấu hình chi tiết của kiến trúc Residual Masking Network cho bài toán nhận diện biểu cảm khuôn mặt.

**CHƯƠNG 1 GIỚI THIỆU**

- Nét mặt hay còn gọi là biểu cảm khuôn mặt con người (facial expression) đóng một vai trò rất quan trọng trong giao tiếp xã hội. Một cuộc trò chuyện bình thường bao gồm các yếu tố ngôn ngữ và phi ngôn ngữ. Các yếu tố phi ngôn ngữ bao gồm giao tiếp bằng mắt, cử chỉ, nét mặt, và ngôn ngữ cơ thể, v.v. Một nụ cười cho thấy sự hạnh phúc, một nét mặt buồn bã cho thấy sự mất mát, một nét giận dữ cho thấy sự không vui và một nét mặt bất ngờ cho thấy một điều không đoán trước đã xảy ra. Theo C. Darwin và P. Prodger [7], nét mặt là một trong những tín hiệu phổ quát, tự nhiên và mạnh mẽ của con người để truyền tải ý định và trạng thái cảm xúc của họ. Trong lĩnh vực thị giác máy tính và học máy, rất nhiều nhà khoa học đã tiến hành nghiên cứu về các hệ thống phân tích nét mặt tự động bởi vì những ứng dụng thực tế quan trọng của nó trong các hệ thống tương tác người máy, qua đó, nhiều hệ thống nhận dạng nét mặt đã cố gắng mã hóa biểu cảm từ những thể hiện trên khuôn mặt.

- Từ đầu thế kỷ 20, Ekman and Friesen [11] đã định nghĩa sáu nét mặt cơ bản dựa trên các nghiên cứu của họ, họ khẳng định rằng cảm xúc con người là phổ quát, nghĩa là con người nhận thức các biểu cảm là giống nhau bất kể họ đến từ nền văn hóa nào. Các nét mặt cơ bản là giận dữ, ghê tởm, sợ hãi, hạnh phúc, buồn bã, ngạc nhiên và trạng thái bình thường. Biểu cảm khinh bỉ được thêm vào sau này như một trong các nét mặt cơ bản. Tuy vậy, các nghiên cứu gần đây về khoa học thần kinh và tâm lý học đã lập luận rằng mô hình của sáu nét mặt cơ bản mang tính chất đặc thù văn hóa và không phổ quát[25]. Nhưng vì tính chất lịch sử, các tập dữ liệu đều được xây dựng dựa trên giả thuyết phổ quát của biểu cảm khuôn mặt. Hiện nay có một số nhà nghiên cứu Trung Quốc đã thực hiện thu thập và phát triển những bộ dữ liệu nhận diện biểu cảm cho người Trung Quốc như Ma, Jialin và đồng nghiệp [35], góp phần đẩy mạnh triển khai ứng dụng bài toán này trong công nghiệp của Trung Quốc [24]. Vì đó tôi thực hiện xây dựng một bộ dữ liệu chứa ảnh người Việt Nam cho bài toán này, đồng thời thực nghiệm để so sánh với một số phương pháp hiện đại (xem Chương 6).

- Hiện nay có 2 mô hình hệ thống nhận diện nét mặt con người đầu vào hay là thiết kế của hệ thống ảnh là

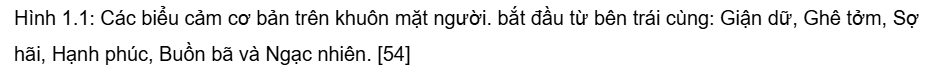
+ Hệ thống dựa vào ảnh tĩnh

+ Hệ thống dựa vào chuỗi ảnh động.

* Hệ thống dựa vào ảnh tĩnh là chỉ phân tích cảm xúc của riêng từng tấm ảnh dựa vào các thông tin sẵn có thông tấm ảnh đó, chủ yếu là trích xuất đặc trưng dựa vào vị trí tương đối của các thành phần có trong tấm ảnh ấy.
* Còn đối với hệ thống dựa vào chuỗi ảnh động, thì hệ thống xem xét cả thông tin về nhịp độ (temporal) của một chuỗi ảnh liên tục, nó không chỉ xem xét sự tương quan về không gian (spatial) trong từng tấm ảnh, và còn xem xét sự tương quan ấy giữa các bức ảnh với nhau.
* Trong luận văn này, tác giả sẽ nghiên cứu phát triển hệ thống dựa vào ảnh tĩnh

1.1 Đặt vấn đề

- Tác giả nói rằng vai trò của nhận diện cảm xúc rất quan trọng trong việc tương tác người máy. Có nhiều phương thức nhận dạng như cảm xúc, giống nói, biểu cảm, cử chỉ … Dưới góc độ khoa học thì việc phân tích cảm xúc còn có thêm yếu tố để phân tích và thấy hiểu con người. Do vậy mà nó có những tác động và ứng dụng rất lớn trong đời sống cũng như trong nghiên cứu của các lĩnh vực khác nhau

* Tác giả đề cập đến việc sử dụng các nhận diện biểu cảm và việc nhận thức chúng như đang nghiên cứu trẻ em bị mắc chứng bệnh rối loạn phổ tự kỷ, nghiên cứu người mắc chứng tâm thần phân liệt, hay một số ứng dụng khác để giám sát cảm xúc của tài xế hoặc lĩnh vực giáo dục, thương mại, quảng cáo, có thể phát triển thêm trong việc quản lý và phân tích đám đông trong 1 số môi trường cụ thể.
* Hiện nay nhận diện cảm xúc khuôn mặt đang được chia làm 2 nhánh dữ liệu khác nhau là:
  + Dữ liệu được điều khiển
  + Dữ liệu thực tế phức tạp
* Tác giả sẽ tập trung thực nghiệm, so sánh và đánh giá trên các tập dữ liệu thức tế phức tạp

1.2 Phạm vi mục tiêu

- Về mặt tiếp cận tác giả tập trung khảo sát, đánh giá, thử nghiện và cải tiến đối với các mô hình học sau và các phương pháp hiện đại

- Về mặt dữ liệu các cơ sở dữ liệu biểu cảm mặt người công khai có sự khác nhau về nhiều mặt bao gồm môi trường thu thập, đầu vào, số lượng, chất lượng và phân phối biểu cảm

- Về mặt đo lường và khảo thí, các nckh khác nhau thường được triển khai rất khác nhau từ giai đoạn tiền xử lý, kiến trúc mạng, quá trình huấn luyện và kiểm thử. Đặc biệt các yếu tố này ảnh hưởng đến hiệu suất

- Về mục tiêu thứ nhất tác giả tập chung cải tiến độ chính xác trên tập dữ liệu, thứ 2 tác giả thực hiện làm trên 1 bộ dữ liệu về khuôn mặt chứa hình ảnh người Việt Nam, phần này tác giả được ảnh hưởng từ 1 vị giáo sư Guoying Zhao đã làm với tập dữ liệu người Trung Quốc.

1.3 Bố cục luận văn

- Gồm 7 chương theo thứ tự và mục đích sau

+ Chương 1: Tác giả trình bày chung về đề tài, mục tiêu và vấn đề giải quyết trong khi làm

+ Chương 2: Các nghiên cứu liên quan đến bài toán được trình bày mang tính tham khảo và dùng để so sánh với cải tiến trong luận văn.

+ Chương 3: Trình bày kiến thức liên quan đến mô hình

+ Chương 4: Trình bày 2 bộ dữ liệu FER2013 và VEMO mà tác giả để tự thu thập và xây dựng

+ Chương 5: Trình bày về mô hình Residual Masking Network là đóng góp chính của tác giả trong luận văn này.

+ Chương 6: Thực nghiệm, gồm phương pháp, quá trình và các kết quả đem đi so sánh  
+ Chương 7: Tóm tắt đóng gop, ưu điểm và nhược điểm trong quá trình thực hiện luận văn. Và nhứng thứ có thể cải tiến trong tương lai

**CHƯƠNG 2 NGHIÊN CỨU LIÊN QUAN**

2.1 Phương pháp sử dụng Học sâu và Máy véc-tơ hỗ trợ tuyến tính

2.1.1 Trình bày sơ lược

- Nghiên cứu của Yichuan Tang tập trung vào việc so sánh các cách khác nhau để xây dựng mô hình học sâu (deep learning), đặc biệt là so sánh hàm Softmax (thường dùng trong phân loại) với Support Vector Machine (SVM) – một phương pháp phân loại khác.

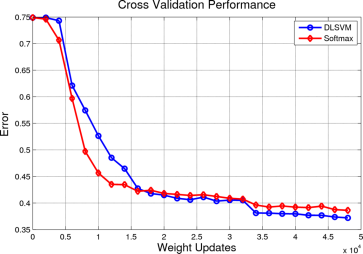
- Kết luận ta thấy được qua tác giả trình bày L2-SVM thường tốt hơn L1-SVM trong các thử nghiệm. và đã dùng L2-SVM vào các thưc nghiệm

- Mô hình của Tang kết hợp SVM với học sâu cho kết quả tốt, đặc biệt trong các bài toán phân loại phức tạp như nhận diện biểu cảm.

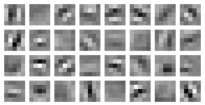
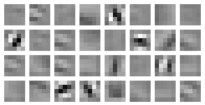
2.1.2 Kết quả

- Chiến thắng Cuộc thi Nhận diện biểu cảm khuôn mặt, năm 2013

- Ytang so sánh khả năng của 2 thuật toán đối với mô hình học sâu



* (Hình 2.1) so sánh giữa việc sử dụng Softmax và L2-SVM như một hàm cập nhật trọng số, điểm lỗi của mô hình DLSVM giảm nhiều hơn ở nửa cuối của quá trình huấn luyện, tuy là không nhiều, nhưng rõ ràng là tốt hơn.
* Ngoài ra, tác giả còn thực hiện thử nghiệm trên 2 tập dữ liệu là MNIST và CIFAR-10, nhưng nó đã vượt ra ngoài đề tài nghiên cứu nên tôi xin phép không nhắc đến ở đây.



2.1.3 Ưu điểm, nhược điểm và khó khăn

- Ưu điểm Y. Tang là một đóng góp nền tảng cung cấp cái nhìn khách quan về việc huấn luyện mô hình học sâu

- Nhược điểm

+ Thiếu đi sự phân tích kỹ lưỡng vào các đặc trưng của biểu cảm khuôn mặt người.

+ Nghiên cứu đa phần tập trung vào việc so sánh giữa việc sử dụng các cách cài đặt mà thiếu đi sự tập trung vào bài toán Nhận diện Biểu cảm

+ Việc trình bày về mô hình được sử dụng trong nghiên cứu vô cùng sơ khai, thiếu đi các chi tiết cụ thể về việc cài đặt các thông số.

2.2 Phương pháp sử dụng chồng chất mạng tích chập đa khu vực

- **MRE-CNN** là một mô hình học sâu dựa trên mạng nơ-ron tích chập (CNN), dùng để nhận diện biểu cảm khuôn mặt (như giận dữ, buồn, vui, ngạc nhiên, v.v.).

- Mô hình này kết hợp nhiều vùng trên khuôn mặt (mắt, mũi, miệng, toàn mặt) để đưa ra dự đoán chính xác hơn.

**- MRE-CNN** cải thiện độ chính xác bằng cách tận dụng thông tin từ nhiều vùng khuôn mặt, thay vì chỉ dùng toàn bộ khuôn mặt. Kiến trúc này lấy cảm hứng từ các mô hình CNN nổi tiếng như AlexNet và VGGNet.

2.3 Phương pháp học kết hợp mạng tích chập đa mức

Skip

**CHƯƠNG 3 CƠ SỞ LÝ THUYẾT**

3.1 Cảm xúc con người thông qua biểu thị nét mặt.

3.1.1 Biểu thị cảm xúc trên khuôn mặt người

**-** 3 yếu tố trong giao tiếp là

+ Từ ngữ dùng trong giao tiếp.

+Cách truyền đạt và giọng điệu khi giao tiếp.

+ Các hành vi phi ngôn ngữ khi giao tiếp.

* Paul Ekman đã nghiên cứu rõ hơn và đưa ra kết luận rằng có một tập hợp các loại cảm xúc luôn được biểu đạt bằng cùng một kiểu nét mặt giống nhau, bất kể giới tính, tuổi tác, nền văn hóa hay lịch sử xã hội [10].

3.1.2 Hệ thống mã hóa cơ mặt

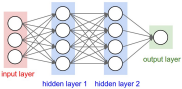
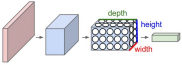
- Hệ thống Mã hóa Hoạt động Cơ mặt (Facial Action Coding System - FACS) được phát triển bởi Ekman và Friesen[14] là hệ thống mô tả đầy đủ những hành vi trên gương mặt con người bằng một tập hợp các đơn vị hoạt động được định nghĩa (Action Unit - AUs).

3.2 Mạng nơ-ron tích chập trong bài báo nhận ảnh

Mạng nơ-ron tích chập (CNNs/ConvNets) rất giống với mạng nơ-ron thông thường, chúng được tạo thành từ các nơ-ron có trọng số và độ chệch mà có thể học được. Mỗi nơ-ron nhận được một số đầu vào, thực hiện một phép tích vô hướng và tùy ý theo sau nó với một lớp phi tuyến tính.

3.2.1 Tổng quan về mạng nowrron tích chập

- Mạng nơ ron tích chập [6] lợi dụng thực tế là đầu vào bao gồm các hình ảnh và những ràng buộc của chúng để thiết kế ra kiến trúc theo một cách hợp lý hơn

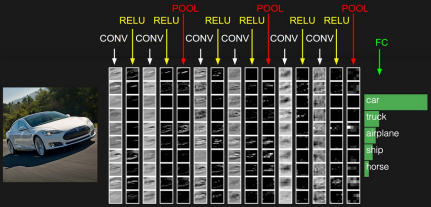
 

So sánh giữa mạng nơ-ron thông thường và mạng nơ-ron tích chập. Một mạng ConvNet được tạo nên từ các lớp. Tất cả các lớp có một API đơn giản: Nó biến đổi một khối đầu vào 3 chiều thành một khối đầu ra 3 chiều bằng một hàm khả vi có thể có hoặc không có tham số.

3.2.2 Các lớp cơ bản trong Mạng Nơ-ron Tích chập.

- Như được đặc tả từ trước, một mạng ConvNets [6] là một chuỗi của các lớp và mỗi lớp của ConvNet biến đổi một khối đặc trưng sang một khối đặc trưng khác thông qua một hàm khả vi. Có ba loại lớp chính để tạo nên một kiến trúc ConvNets: Lớp Tích chập (Convolutional Layer), Lớp Gộp (Pooling Layer) và một Lớp Kết nối Đầy đủ (Fully-Connected Layer). Chồng chất các lớp này chúng ta sẽ nhận được một kiến trúc ConvNet.

- Nói ngắn gọn, một mạng ConvNet à một chồng chất các Lớp nhằm biến đổi một khối ảnh sang thành một khối đầu ra. Mạng này chứa một số lớp cơ bản như CONV/RELU/FC/POOL/... Mỗi lớp này nhận một đầu vào ba chiều và trả về đầu ra ba chiều bằng một hàm khả vi. Mặt khác, mỗi lớp có thể có tham số hoặc không.



3.2.3 Lớp tích chập

- Lớp tích chập rất quan trọng vì đa số các công việc tính toán đều trong kiểu kiến trúc này.