

ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN



ĐỒ ÁN MÔN HỌC
CS114 – Machine Learning

ĐỀ TÀI
NHẬN DẠNG CẢM XÚC TRONG ẢNH
(EMOTION DETECTION)

Giảng viên hướng dẫn : **LÊ ĐÌNH DUY, PHẠM NGUYỄN TRƯỜNG AN**
Sinh viên thực hiện: **TRƯỜNG LÊ VĨNH PHÚC**

LÊ THÀNH TIẾN
CS114.O11

Lớp:

TP. Hồ Chí Minh, tháng 1 năm 2024

TÓM TẮT

Link Repository: <https://github.com/sloweyyy/CS114.O11-FinalProject>

1. Mô tả bài toán

1.1. Ngữ cảnh ứng dụng

Trong thời đại số hóa, ảnh chân dung ngày càng trở thành một phương tiện quan trọng trong việc giao tiếp, thu thập dữ liệu, và được sử dụng trong nhiều lĩnh vực như bán hàng online và chăm sóc sức khỏe tâm thần. Đồ án tập trung vào việc phân loại cảm xúc từ ảnh chân dung để ứng dụng trong các cơ sở bán hàng và bệnh viện tâm thần.

1.2. Input & Output:

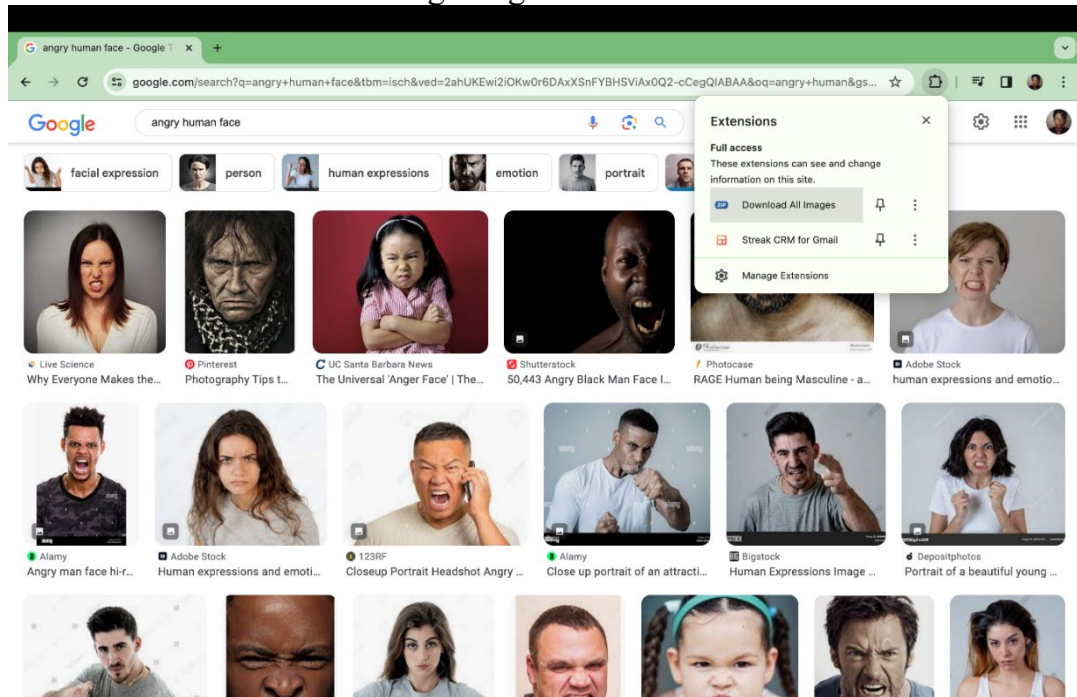
- Input: Ảnh chứa mặt của 01 người
- Output: Ảnh cắt mặt và mô tả cảm xúc

2. Mô tả bộ dữ liệu:

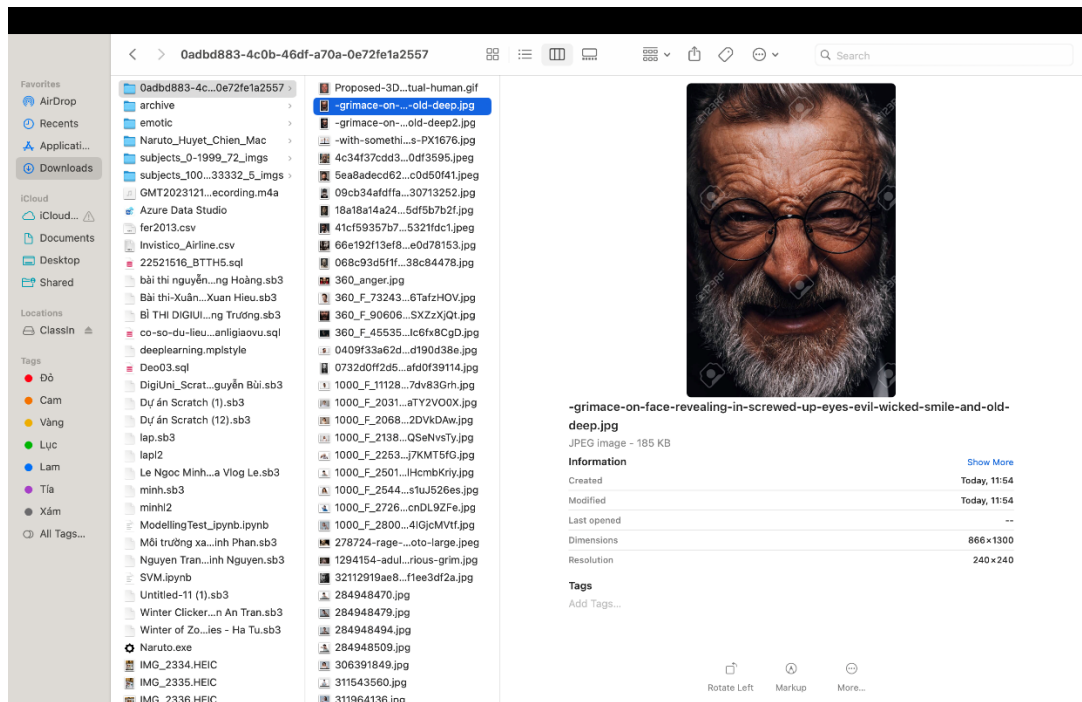
2.1. Thu thập dữ liệu

2.1.1. Cách thức thu thập dữ liệu

- Tìm kiếm các keywords của cảm xúc liên quan (angry face human, happy face,...) trên các trang tìm kiếm hình ảnh của google, bing, unsplash.com, freepik.com,...
- Sử dụng extension của trình duyệt Google Chrome (Download All Images) để tải về tất cả các ảnh có trong trang.








- Giải nén folder vừa tải về và xóa bớt các bức ảnh sai chủ đề.



2.1.2. Khó khăn trong việc thu thập dữ liệu

- Khi sử dụng nhiều trang để tìm kiếm cùng một loại cảm xúc, nhiều trang có các tấm ảnh giống nhau.
- Một tấm ảnh có thể tìm thấy ở 2 cảm xúc khác nhau (VD: fear và neutral).
- Nhiều tấm ảnh không liên quan xuất hiện chung trang, gây mất thời gian phân loại.

2.2. Số lượng, độ đa dạng

ID	Tên cảm xúc	Số lượng	Hình ảnh
0	Angry	1000	
1	Fear	1000	
2	Happy	1000	
3	Neutral	1000	
4	Sad	1000	

2.3. Các thao tác tiền xử lý dữ liệu

- Lọc ảnh trùng
- Xóa các ảnh không phải ảnh người thật
- Xóa các ảnh khó phân biệt với cảm xúc khác

2.4. Phân chia dữ liệu

- Sau khi label và lọc ảnh, còn lại 5000 ảnh. Tiến hành chia train/val với tỉ lệ 8/2:
 - o Train: 4000 ảnh.
 - o Validate: 1000 ảnh.

3. Mô tả về đặc trưng

3.1. Feature Engineering

- Sử dụng Dlib để phát hiện khuôn mặt
- Cắt ảnh chỉ giữ lại vùng xung quanh khuôn mặt
- Resize ảnh về 48x48 pixel.

3.2. Data Processing Pipeline

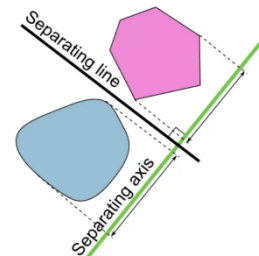
- Chuyển ảnh thành ma trận pixel kích thước 48x48 pixel
- Ghi ma trận vào file csv cùng với ID của label

4. Mô tả thuật toán máy học

4.1. Tổng quan thuật toán SVM (Support vector machine)

- Support Vector Machines (có tài liệu dịch là Máy véctor hỗ trợ) là một trong số những thuật toán phổ biến và được sử dụng nhiều nhất trong học máy trước khi mạng nơ ron nhân tạo trở lại với các mô hình deep learning. Nó được biết đến rộng rãi ngay từ khi mới được phát triển vào những năm 1990.

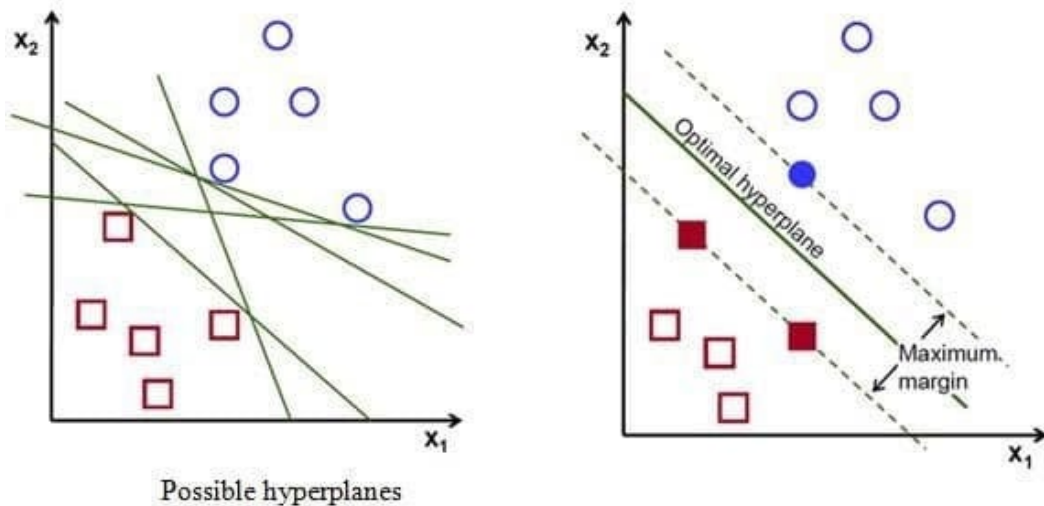
- Mục tiêu của SVM là tìm ra một siêu phẳng trong không gian N chiều (ứng với N đặc trưng) chia dữ liệu thành hai phần tương ứng với lớp của chúng. Nói theo ngôn ngữ của đại số tuyến tính, siêu phẳng này phải có lề cực đại và phân chia hai bao lồi và cách đều chúng.



4.2. Nguyên lý hoạt động

- Trong không gian N chiều, một siêu phẳng là một không gian con có kích thước N-1 chiều. Một cách trực quan, trong một mặt phẳng (2 chiều) thì siêu phẳng là một đường thẳng, trong một không gian 3 chiều thì siêu phẳng là một mặt phẳng. Để phân chia hai lớp dữ liệu, rõ ràng là có rất nhiều siêu phẳng có thể làm được điều này. Mặc dù vậy, mục tiêu của chúng ta là tìm ra siêu phẳng có lề rộng nhất tức là có khoảng cách tới các điểm của hai lớp là lớn nhất. Hình dưới đây là một ví dụ trực quan về điều đó.

- Để phân chia hai lớp dữ liệu, rõ ràng là có rất nhiều siêu phẳng có thể làm được điều này. Mặc dù vậy, mục tiêu của chúng ta là tìm ra siêu phẳng có lề rộng nhất tức là có khoảng cách tới các điểm của hai lớp là lớn nhất. Hình dưới đây là một ví dụ trực quan về điều đó.



5. Cài đặt, tinh chỉnh thông số

5.1. Tham số C

- C là tham số đối với biên mềm trong mô hình SVM. Nó kiểm soát sự đồng nhất của biên quyết định.
- Giá trị C lớn sẽ tạo ra một biên quyết định chặt chẽ, có thể dẫn đến việc mô hình không tổng quát hoá tốt (overfitting).
- Giá trị C nhỏ hơn sẽ tạo ra một biên mềm, cho phép mô hình tổng quát hoá tốt hơn nhưng có thể không xác định các điểm nhiều.

5.2. Tham số Gamma

- Gamma đặc trưng mức độ ảnh hưởng của một điểm dữ liệu đến việc tạo ra biên quyết định. Giá trị gamma càng cao, mức độ ảnh hưởng càng giảm.
- Đối với kernel “rbf”, “poly”, và “sigmoid”, gamma quyết định hình dạng của siêu phẳng quyết định.

5.3. Tham số Kernel

- Kernel quyết định loại hàm nhân được sử dụng trong SVM để biến đổi không gian dữ liệu. Các giá trị phổ biến cho kernel là “linear”, “rbf” (Radial Basis Function), “poly” (Polynomial), và “sigmoid”.

6. Đánh giá kết quả, kết luận

6.1. Đánh giá kết quả

Sau khi thực hiện train model, để xác định model của chúng ta có đủ tốt hay chưa cũng như đảm bảo khả năng nhận diện trong tương lai ta cần có một phương pháp đánh giá với tiêu chí cụ thể. Đối với bài toán Classification, model thường được đánh giá dựa trên Precision, Recall, F1-score.

Class	Tên cảm xúc	Precision	Recall	F1-Score
0	Angry	0.5561	0.5278	0.5416

1	Fear	0.5251	0.4947	0.5095
2	Happy	0.7767	0.8333	0.8040
3	Neutral	0.8191	0.7762	0.7971
4	Sad	0.4692	0.5156	0.4913

6.2. Kết luận

- Mô hình có hiệu suất khá tốt trong việc nhận diện cảm xúc “Happy” và “Neutral”, với các độ đo hiệu suất đều ở mức cao.
- Cần cải thiện thêm hiệu suất cho các cảm xúc “Angry”, “Sad” và “Fear” để đảm bảo hiệu suất giữa các lớp.

7. Hướng phát triển trong tương lai

- Hệ thống sẽ trả về thông tin phân loại trực tiếp trên video, không chỉ ảnh.
- Việc phân loại cần yếu tố thời gian, do đó yêu cầu thiết bị có cấu hình mạnh. Nhưng đa phần các loại điện thoại di động bây giờ không được thiết kế để thực hiện tác vụ này. Chúng em sẽ triển khai model lên web, người dùng sẽ giao tiếp với hệ thống thông qua web. Khi đó ta chỉ cần quan tâm tới tốc độ mạng (Vấn đề về mạng thì dễ giải quyết hơn).
- Hướng phát triển tiếp theo là kết hợp model này với Zoom, Google Meet, vì được sử dụng rộng rãi hiện nay ở Việt Nam cho nhiều mục đích.

CHƯƠNG 0: UPDATE SAU KHI VẤN ĐÁP

1. So sánh với các thuật toán khác

1.1. VGG16:

- VGG16 là một mô hình Convolutional Neural Network (CNN) có kiến trúc sâu với 16 lớp trọng số.
- Mỗi lớp convolution được thiết kế với các filter nhỏ (3x3) và sử dụng hàm kích hoạt ReLU.
- Sử dụng các lớp max-pooling để giảm kích thước của feature map.

Đặc điểm nổi bật:

- Kiến trúc đơn giản và dễ hiểu.
- Sử dụng các filter nhỏ giúp học được các đặc trưng chi tiết.

Ưu điểm và nhược điểm:

- *Ưu điểm:*
 - Hiệu suất tốt trên các tập dữ liệu ảnh cỡ lớn.
 - Dễ triển khai và sử dụng.
- *Nhược điểm:*
 - Số lượng tham số lớn, tốn kém về tài nguyên.
 - Cần nhiều dữ liệu để đạt được kết quả tốt.

Ứng dụng trong phân loại ảnh:

- Phù hợp cho các bài toán phân loại ảnh tổng quát.
- Sử dụng rộng rãi trong các cuộc thi và ứng dụng thực tế.

1.2. ResNet

- ResNet sử dụng khái niệm "residual learning" để xử lý vấn đề của việc mô hình trở nên khó huấn luyện khi càng sâu.
- Các "residual blocks" giữ lại thông tin ban đầu và thêm vào đó thông tin học được trong quá trình lan truyền tiến.

Keypoints về hiệu suất:

- Khả năng huấn luyện mô hình sâu hơn mà không gặp vấn đề "vanishing gradient".
- Hiệu suất tốt trên các tập dữ liệu lớn và đặc trưng phức tạp.

Ưu điểm và nhược điểm:

- *Ưu điểm:*
 - Xử lý được các mô hình sâu mà không gặp vấn đề độ suy giảm gradient.
 - Hiệu suất cao trên nhiều nhiệm vụ thị giác máy tính.
- *Nhược điểm:*
 - Đòi hỏi nhiều tài nguyên, cần phải có kích thước dữ liệu lớn.

Ứng dụng trong phân loại ảnh:

- Sử dụng rộng rãi trong các ứng dụng nhận diện và phân loại ảnh.
- Đặc biệt hiệu quả trên các nhiệm vụ yêu cầu mô hình sâu.

1.3. So sánh kết quả

model	accuracy	precision	recall
svm	0.54	0.62	0.56
VGG16	0.19	0.19	0.19
ResNet50	0.20	0.20	0.20

Có một số nguyên nhân có thể giải thích tại sao SVM có thể có số điểm cao hơn so với VGG16 và ResNet50 trên một tập dữ liệu chân dung có 5 loại cảm xúc:

- **Kích thước Dữ liệu:**
 - SVM thường có hiệu suất tốt khi kích thước dữ liệu nhỏ. Do bộ data chỉ có 5000 ảnh chân dung, đó là một tập dữ liệu khá nhỏ, và SVM có thể hoạt động tốt trên các tập dữ liệu nhỏ hơn.
- **Feature Engineering:**
 - SVM yêu cầu tính đặc trưng (feature engineering) chủ động từ dữ liệu đầu vào. Nếu thiết kế các đặc trưng tốt cho bài toán cụ thể, SVM có thể tận dụng được thông tin đó.
- **Hyperparameter Tuning:**
 - SVM có một số siêu tham số (hyperparameters) cần được điều chỉnh, như kernel, C, gamma, v.v. Nếu đã tinh chỉnh siêu tham số cẩn thận, SVM có thể thích ứng tốt với dữ liệu.
- **Khả năng Tính toán:**
 - Trong một số trường hợp, đặc biệt là với dữ liệu nhỏ, các mô hình đơn giản như SVM có thể được đào tạo nhanh chóng và hiệu quả hơn so với các mô hình phức tạp như VGG16 và ResNet50.
- **Overfitting:**
 - Có thể mô hình VGG16 và ResNet50 có kích thước lớn và chứa nhiều tham số. Trong trường hợp dữ liệu nhỏ, có khả năng mô hình sẽ bị overfitting, đặc biệt nếu không có các biện pháp chống overfitting được áp dụng.
- **Phức Tạp Của Mô Hình:**
 - VGG16 và ResNet50 là các mô hình phức tạp với nhiều lớp và tham số. Trong một số trường hợp, sự phức tạp này có thể làm tăng khả năng overfitting, đặc biệt khi dữ liệu là nhỏ.