

**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**  
**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**



**ĐỒ ÁN MÔN HỌC**  
**NHẬN DIỆN CẢM XÚC GƯƠNG MẶT**  
**CS111.K21.KHTN**

**Thành viên:**

- 1. Nguyễn Anh Khoa – 18520923**
- 2. Trần Thị Phương Thảo – 18521422**
- 3. Võ Quốc An – 18520440**

**TP. Hồ Chí Minh, 2020**

# Mục lục

I.	Giới thiệu đề tài: .....	3
II.	Nội dung đề tài:.....	3
III.	Hướng tiếp cận:.....	5
IV.	Kết quả đạt được:.....	9
V.	Nhận xét: .....	11
VI.	Xây dựng hệ thống demo nhận diện cảm xúc:.....	11
VII.	Tài liệu tham khảo: .....	11

## **I. Giới thiệu đề tài:**

- Bạn mong đợi nhân viên có trí tuệ cảm xúc cao khi tương tác với khách hàng. Giờ đây, nhờ những tiến bộ trong Học sâu, bạn sẽ sớm mong đợi phần mềm của mình cũng làm được như vậy.
- Nghiên cứu đã chỉ ra rằng hơn 90% giao tiếp của chúng ta có thể là phi ngôn ngữ, nhưng công nghệ đã phải vật lộn để theo kịp và mã truyền thông nói chung là kém hiểu được ngữ điệu và ý định của chúng ta. Nhưng nhận dạng cảm xúc - còn được gọi là Tính toán theo cảm xúc - đang trở nên dễ tiếp cận hơn với nhiều loại nhà phát triển hơn.
- Bạn là một nhà doanh nghiệp. Làm thế nào để đánh giá độ hài lòng của khách hàng với dịch vụ của bạn mà không cần thông qua khảo sát khách hàng. Đề tài của chúng tôi hướng đến là giải quyết vấn đề trên bằng cách xây dựng ứng dụng nhận biết cảm xúc của khách hàng thông qua hình ảnh lấy được từ camera. Ngoài ra đề tài của chúng tôi còn hướng đến việc ứng dụng cho robot tiếp khách có khả năng nhận biết cảm xúc của người đối diện để lựa chọn ngôn ngữ giao tiếp một cách phù hợp.

## **II. Nội dung đề tài:**

### **1. Nhận dạng cảm xúc gương mặt là gì ?**

- Nhận dạng cảm xúc trên khuôn mặt là quá trình phát hiện cảm xúc của con người từ các biểu hiện trên khuôn mặt. Bộ não của con người nhận biết cảm xúc một cách tự động, và phần mềm hiện đã được phát triển để có thể nhận ra cảm xúc. Công nghệ này ngày càng chính xác hơn mọi lúc, và cuối cùng sẽ có thể đọc được cảm xúc giống như bộ não của chúng ta.
- AI có thể phát hiện cảm xúc bằng cách tìm hiểu ý nghĩa của từng biểu hiện trên khuôn mặt và áp dụng kiến thức đó vào thông tin mới được đưa ra. Trí tuệ nhân tạo cảm xúc, hay AI cảm xúc, là một công nghệ có khả năng đọc, bắt chước, diễn giải và phản ứng với các biểu hiện và cảm xúc trên khuôn mặt của con người.

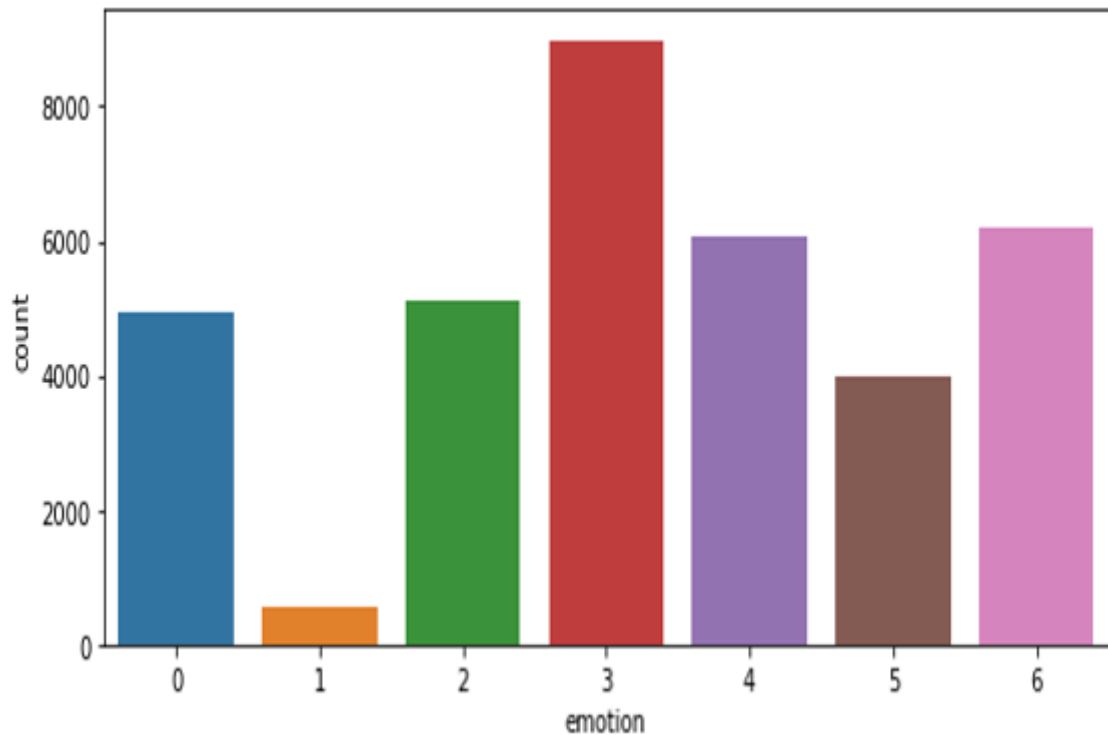
### **1. Bộ dữ liệu**

#### **1.1 Tổng quan về bộ dữ liệu**

- Fer2013 bộ dữ liệu được lấy từ Kaggle. Fer2013 chứa hơn 30.000 hình ảnh biểu cảm được nhóm thành 7 danh mục:

- Giận dữ 0
- Ghê tởm 1
- Sợ hãi 2
- Hạnh phúc 3
- Bình thường 4
- Buồn 5
- Ngạc nhiên 6

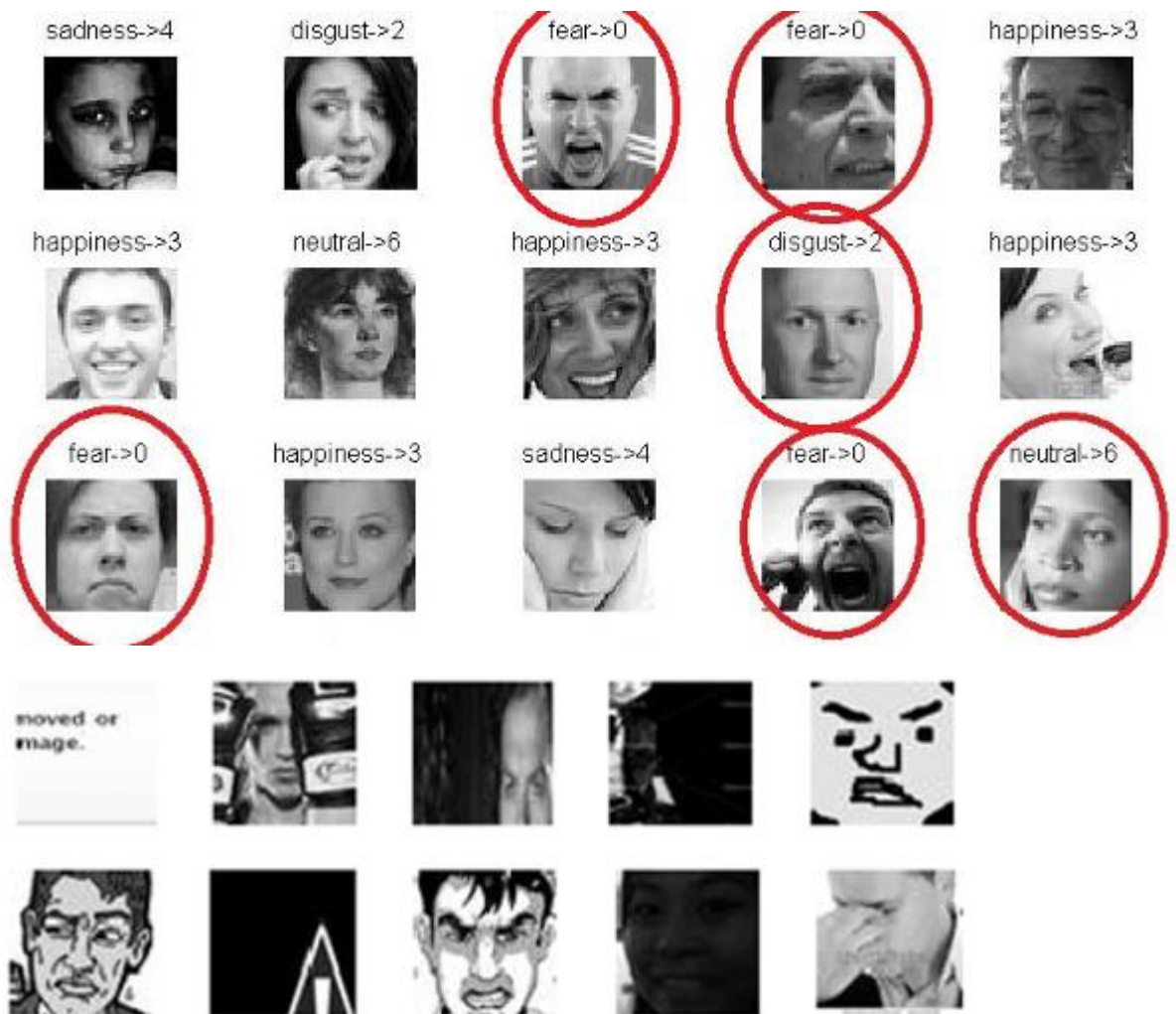
Dữ liệu bao gồm hình ảnh thang độ xám 48 \* 48 pixel của khuôn mặt.



Hình ảnh thống kê bộ dữ liệu

## 1.2 Thách thức của bộ dữ liệu:

- Fer2013 là một tập dữ liệu đầy thách thức. Các hình ảnh không được căn chỉnh và một số trong số chúng được dán nhãn không chính xác như chúng ta có thể thấy từ các hình ảnh sau đây. Hơn nữa, một số mẫu không chứa khuôn mặt.



- Điều này làm cho việc phân loại khó hơn vì mô hình phải tổng quát hóa tốt và mạnh mẽ đối với dữ liệu không chính xác.

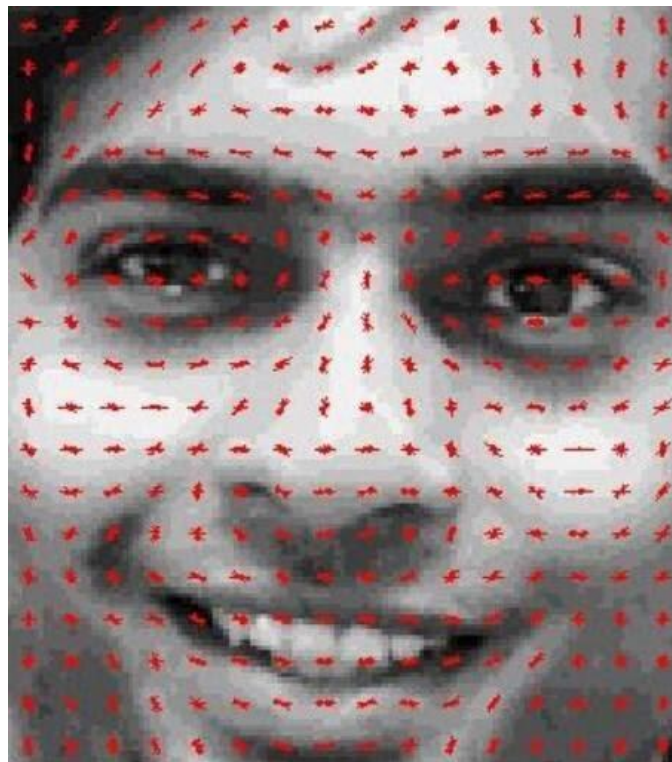
### III. Hướng tiếp cận:

- Sử dụng phương pháp rút trích đặc trưng Hog, và các điểm mốc trên gương mặt để đào tạo model SVM nhận diện cảm xúc qua hình ảnh
- Sử dụng CNN
- Sử dụng CNN + Hog + facial landmarks để train model

#### 1. Đặc trưng HOG

- HOG là viết tắt của Histogram of Oriented Gradient - một loại “feature descriptor”. Mục đích của “feature descriptor” là trừu tượng hóa đối tượng bằng cách trích xuất ra những đặc trưng của đối tượng đó và bỏ đi những thông tin không hữu ích. Vì vậy, HOG được sử dụng chủ yếu để mô tả hình dạng và sự xuất hiện của một đối tượng trong ảnh.

- Bản chất của phương pháp **HOG** là sử dụng thông tin về **sự phân bố** của các **cường độ gradient** (intensity gradient) hoặc của **hướng biên** (edge directions) để mô tả các đối tượng cục bộ trong ảnh.
- Các toán tử **HOG** được cài đặt bằng cách chia nhỏ một bức ảnh thành các vùng nhỏ, được gọi là “**tế bào**” (cells) và với mỗi cell, ta sẽ tính toán một histogram về các hướng của gradients cho các điểm nằm trong cell. Ghép các histogram lại với nhau ta sẽ có một biểu diễn cho bức ảnh ban đầu.
- Để tăng cường hiệu năng nhận dạng, các histogram cục bộ có thể được chuẩn hóa về độ tương phản bằng cách tính một ngưỡng cường độ trong một vùng lớn hơn cell, gọi là các khối (blocks) và sử dụng giá trị ngưỡng đó để chuẩn hóa tất cả các cell trong khối. Kết quả sau bước chuẩn hóa sẽ là một vector đặc trưng có tính bất biến cao hơn đối với các thay đổi về điều kiện ánh sáng.



HOG có khuôn mặt mẫu (Nguồn: eInfochips)

## 2. Đặc trưng landmark

### 2.1 Các mốc trên khuôn mặt (facial landmarks) là gì?

- Phát hiện các điểm mốc trên khuôn mặt là một tập con của bài toán dự đoán hình dạng. Đưa ra một hình ảnh đầu vào (và thường là ROI chỉ định đối tượng quan tâm), công cụ dự đoán hình dạng cố gắng xác định vị trí các điểm quan tâm chính dọc theo hình dạng.
- Trong bối cảnh các điểm mốc trên khuôn mặt, mục tiêu của chúng tôi là phát hiện các cấu trúc quan trọng trên khuôn mặt bằng phương pháp dự đoán hình dạng.
- Do đó, phát hiện các điểm mốc trên khuôn mặt là một quá trình gồm hai bước:
  - **Bước 1:** Nhận diện khuôn mặt
    - Điều này có thể được thực hiện bằng cách sử dụng một số kỹ thuật khác nhau, nhưng thường liên quan đến các tầng Haar hoặc các bộ dò SVM tuyến tính HOG + (nhưng bất kỳ cách tiếp cận nào tạo ra một hộp giới hạn xung quanh mặt đều được).
  - **Bước 2:** Phát hiện các cấu trúc chính trên khuôn mặt
    - Có nhiều loại máy dò mốc khuôn mặt, nhưng tất cả các phương pháp về cơ bản đều cố gắng xác định vị trí và gắn nhãn các vùng trên khuôn mặt sau:
      - Miệng
      - Lông mày phải
      - Lông mày trái
      - Mắt phải
      - Mắt trái
      - Mũi
      - Hàm
    - Máy dò mốc khuôn mặt có trong thư viện dlib là sự triển khai của Căn chỉnh khuôn mặt một phần nghìn giây với một bài báo về Cây hồi quy của Kazemi và Sullivan (2014).
    - Phương pháp này bắt đầu bằng cách sử dụng:
      1. Tập huấn luyện các điểm mốc trên khuôn mặt được gắn nhãn trên hình ảnh. Những hình ảnh này được tự dán

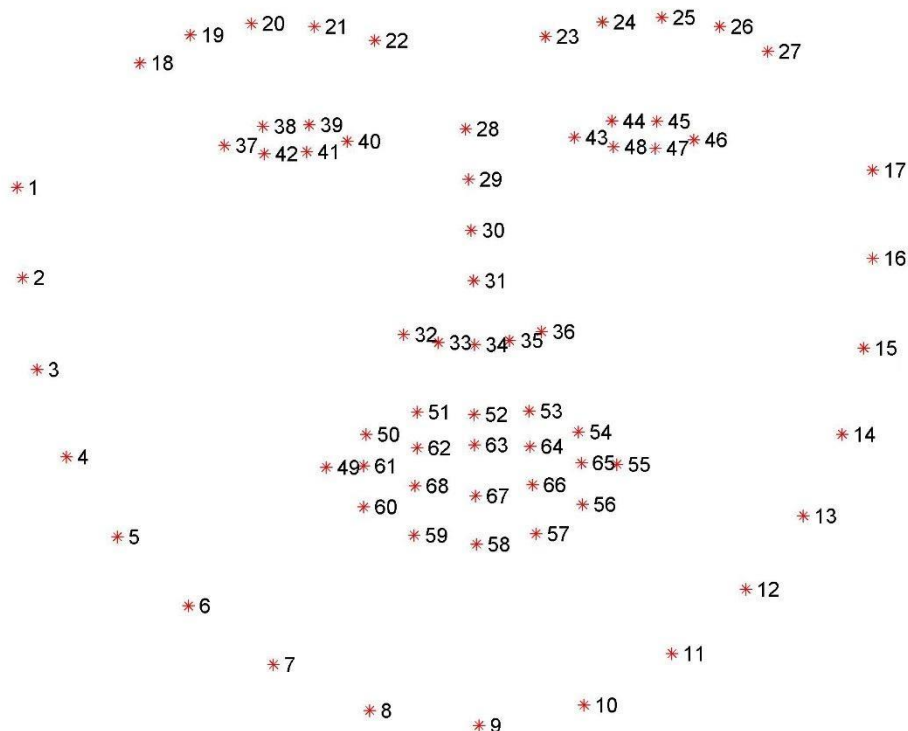
nhãn , xác định cụ thể (x, y) -coordinates của khu vực xung quanh mỗi cấu trúc khuôn mặt.

2. Trước đó , cụ thể hơn là xác suất về khoảng cách giữa các cặp pixel đầu vào.

- Với dữ liệu đào tạo này, một tập hợp các cây hồi quy được đào tạo để ước tính các vị trí mốc trên khuôn mặt trực tiếp từ chính cường độ pixel (nghĩa là không diễn ra "trích xuất đối tượng").
- Kết quả cuối cùng là máy dò mốc trên khuôn mặt có thể được sử dụng để phát hiện các mốc trên khuôn mặt trong thời gian thực với các dự đoán chất lượng cao.

## 2.2. Tìm hiểu về máy dò mốc khuôn mặt của dlib

- Máy dò mốc trên khuôn mặt đã được đào tạo trước bên trong thư viện dlib được sử dụng để ước tính vị trí của tọa độ 68 (x, y) ánh xạ đến các cấu trúc trên khuôn mặt.
- Các chỉ số của 68 tọa độ có thể được hình dung trên hình ảnh dưới đây:



Hình dung 68 tọa độ mốc khuôn mặt từ tập dữ liệu iBUG 300-W

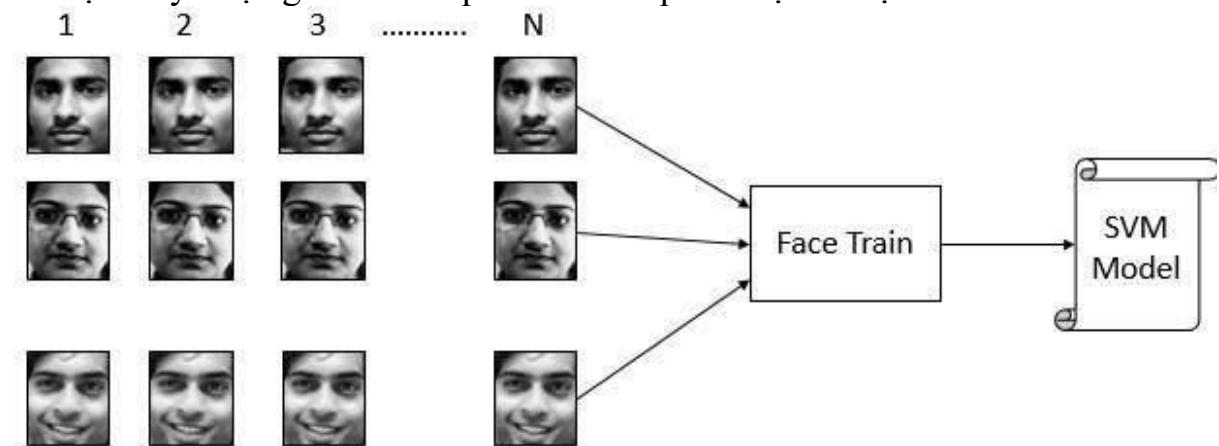


- Các chú thích này là một phần của tập dữ liệu 68 điểm iBUG 300-W mà công cụ dự báo mốc khuôn mặt dlib đã được đào tạo.
- Điều quan trọng cần lưu ý là tồn tại các hướng vị khác của máy dò mốc khuôn mặt, bao gồm cả mô hình 194 điểm có thể được đào tạo trên tập dữ liệu HELEN .

## IV. Kết quả đạt được:

### 1. Bộ phân loại SVM

- Support Vector Machine (SVM) là gì? Máy vector hỗ trợ (SVM) là mô hình học máy được giám sát để phân chia và phân loại dữ liệu.



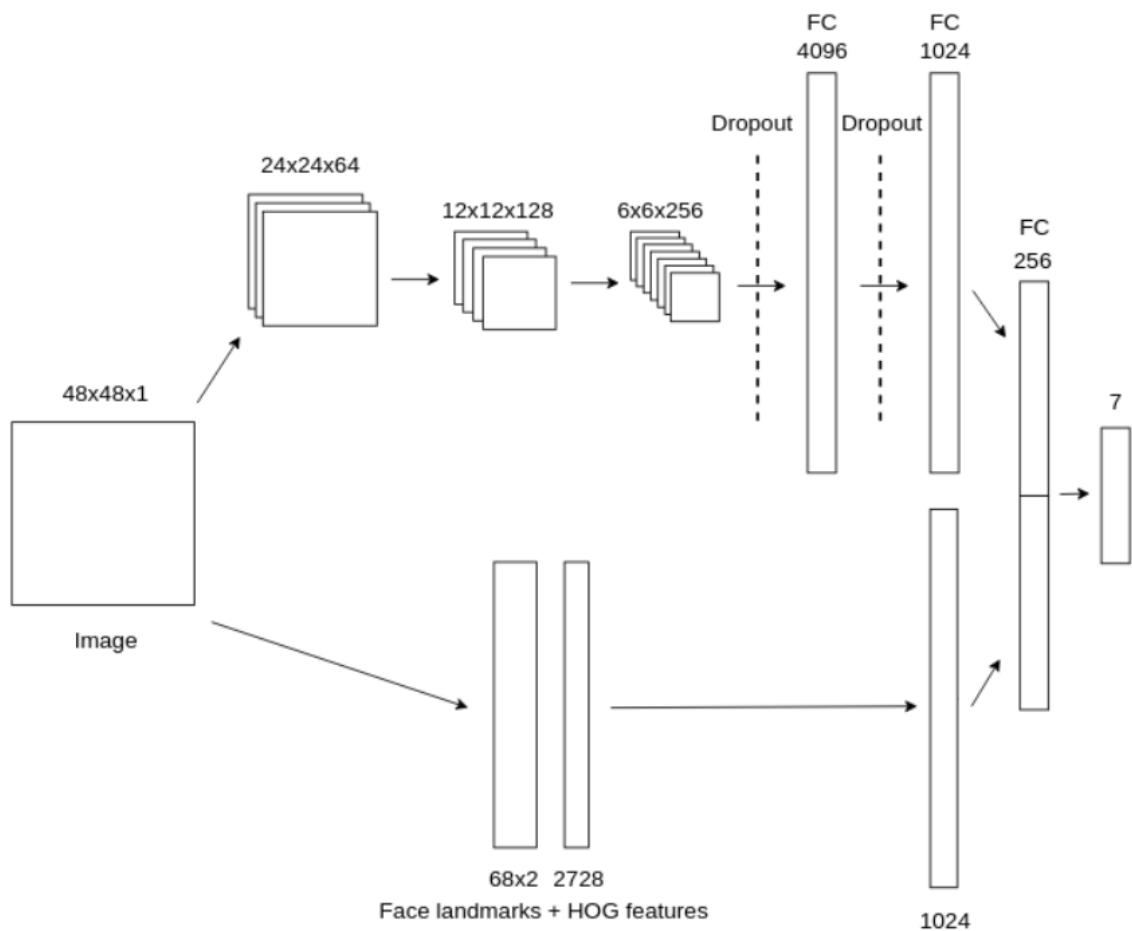
Sơ đồ khối của đào tạo khuôn mặt (Nguồn: eInfochips)

- SVM được sử dụng rộng rãi cho các ứng dụng như nhận diện khuôn mặt, phân loại hình ảnh, nhận dạng chữ viết tay, v.v. Một mô hình SVM có thể được coi là một không gian điểm trong đó nhiều lớp được tách biệt bằng cách sử dụng siêu máy bay.
- Accuracy model SVM

Model	Accuracy
Hog	50.02%
Landmarks	52.86%
Hog + Landmarks	57.19%

## 2. Model CNN

- Mục đích là để có được một đường cơ sở nhanh chóng để so sánh xem kiến trúc CNN hoạt động tốt hơn khi nó chỉ sử dụng các pixel thô của hình ảnh để đào tạo hoặc nếu tốt hơn là cung cấp một số thông tin bổ sung cho CNN (chẳng hạn như các mốc khuôn mặt hoặc các tính năng HOG). Kết quả cho thấy rằng thông tin bổ sung giúp CNN hoạt động tốt hơn.
- Các khuôn mặt được phát hiện đầu tiên bằng opencv, sau đó chúng tôi trích xuất các mốc khuôn mặt bằng cách sử dụng dlib. Chúng tôi cũng trích xuất các tính năng HOG và nhập dữ liệu hình ảnh thô với các mốc khuôn mặt + con heo vào một mạng nơ-ron phức hợp.



Model CNN được sử dụng (\*)

(\*): kiến trúc mạng CNN tham khảo tại paper <https://arxiv.org/abs/1612.02903>

- Accuracy CNN

Model	Accuracy
CNN	57.98%
CNN + Landmarks	61.5%

## V. Nhận xét:

- Mô hình bị thiên về lớp có nhiều dữ liệu.
- Mô hình CNN cho kết quả tốt hơn SVM
- Kết hợp nhiều tính năng hơn như Face Landmarks và HOG, cải thiện thêm độ chính xác của model.

## VI. Xây dựng hệ thống demo nhận diện cảm xúc:

- Sau quá trình training ta được 1 model nhận diện khuôn mặt
- Để demo kết quả với input là ảnh hoặc video real time, trước tiên cần xác định được gương mặt, trong bài này, nhóm mình sử dụng pretrained model **Haar Cascade** để lấy ra được ảnh gương mặt
- Từ ảnh gương mặt, tiến hành nhận diện với model đã train

### Mã nguồn:

<https://colab.research.google.com/drive/1Hmf91jM2Ks6bHn1CDA6k3w619QxmbEdp>

- Hướng dẫn chạy:
  - o Mount drive
  - o Run all cell
  - o Chọn ảnh và up load để tiến hành demo

## VII. Tài liệu tham khảo:

- [1] <https://arxiv.org/abs/1612.02903>
- [2] <https://github.com/amineHorseman/facial-expression-recognition-using-cnn>
- [3] <https://github.com/amineHorseman/facial-expression-recognition-svm>
- [4] <https://viblo.asia/p/nhan-dien-cam-xuc-khuon-mat-don-gian-voi-keras-V3m5WvRwlO7>
- [5] <https://www.miai.vn/2019/09/11/face-recog-2-0-nhan-dien-khuon-mat-trong-video-bang-mtcnn-va-facenet/>