**Báo cáo lần 1 (30/03/2021)**

*Tên sinh viên* : Nguyễn Hồng Huy

Bùi An Nguyên

Lương Thế Long

Trần Thế Dâng

*Đề tài:* Hệ thống mở cửa tự động dùng nhận diện khuôn mặt.

**Nội dung công việc:**

* Tìm hiểu các keyword “triplet loss”, “facenet”, “one-shot learning”.
* Tìm hướng đi cho việc kết nối giữa output và khoá.
* Chuẩn bị dữ liệu cho việc làm phần face recognition.
* Xác định server sử dụng.

**Tìm hiểu các keyword “triplet loss”, “facenet”, “one-shot learning”:**

* Triplet loss:

Tripless loss hay còn gọi là loss bộ ba A, P, N. Công thức này chúng ta sẽ xét trên độ dài vector của từng cặp A, P (Anchor and Positive) và A, N (Anchor and Negative).

Xét với cặp A, P trước: Đây là hai ảnh Anchor (ảnh gốc) và Positive (ảnh tương đồng), hai ảnh này chúng ta sẽ phân tích thành hai vector Embeeding sau đó xét khoảng cách của hai vector, nếu giá trị khoảng cách của hai vector này càng thấp thì hai ảnh này càng tương đồng.

Xét cặp A, N tiếp theo: Đây là hai ảnh Anchor (ảnh gốc) và Negative (ảnh khác biệt), hai ảnh này cũng phân tích thành hai vector sau đó xét tiếp khoảng cách của hai vector này, nếu như hai vector có khảng cách càng xa nhau thì hai ảnh này càng khác nhau.

Công thức:

Thuật toán Triplet loss này được sử dụng phổ biến vì nó là sự kết hợp giữa phân tích hai hướng cùng lúc đó là xét sự tương đồng nếu hai ảnh cùng class và sự khác nhau nếu không cùng class:

Bức ảnh sẽ là rõ ràng khi mà Triplet loss trả về là bằng 0. Vì vậyf(img) = f(0). Tuy nhiên chúng ta sẽ có trường hợp là: cho ra số âm, và đương nhiên số âm sẽ thường là hơi bị tù túng và khó chịu nên là mình sẽ luôn cho nó là = 0 khi mà kết quả là số âm.

Tuy nhiên chúng ta sẽ có một vài vấn đề nhỏ có thể xảy ra sai số vô cùng bé giữa khoảng cách từ f(A) tới f(P) và khoảng cách từ f(A) tới f(N). Ví dụ và . Sự khác biệt để học hầu như là không có. Vì vậy ta cần một hằng số Margin là để phân biệt được sự tương đồng và khác biệt giữa khoảng cách của hai vector Chung quy lại: Thuật toán Triplet Loss giúp chúng ta nhận diện sự tương đồng và sự khác biệt cùng 1 lúc và hiệu quả hơn.

* Facenet:
  + Khái niệm:

Là thuật toán dùng Siam network để biểu thị 2 vector Embedding đầu ra có thể so sánh được.

Trước khi dùng FaceNet:

* + - Mọi người tìm cách phân tích ảnh sang Vector Embeeding.
    - Nhưng mà để ra được Vector Embedding hoàn chỉnh thì phải thông qua một layer Bottle Neck để giảm chiều dữ liệu so sánh nhưng mà mặt hạn chế của Vector Embedding này là số lượng chiều Vector Embedding này rất là lớn (tầm hơn 1000) sẽ ảnh hưởng rất nhiều tới thời gian thuật toán xử lý, vì vậy họ đã áp dụng thuật toán PCA để giảm chiều Vector Embedding để giảm thời gian xử lý.
    - Tương đương với Loss Function của nó chỉ đo lường được khoảng cách giữa hai bức ảnh vậy đầu vào của học máy chỉ là một trong hai chính là giống nhau nếu cùng class và khác nhau nếu khác class mà không học được giống nhau và khác nhau trên cùng một lượt huấn luyện với lượng lớn dataset.
  + Lợi ích:
    - Facenet dùng chính CNN (Convolutional neutral network) và giảm chiều dữ liệu xuống còn 128 chiều (không gian Eucledean).
    - Do đó quá trình phân tích ảnh và dự báo ảnh sẽ nhanh hơn đồng thời độ chính xác được đảm bảo.
    - Loss Function không còn là phân tích giữa một bức ảnh nữa mà dùng triplet loss có thể đọc được sự giống nhau giữa hai bức ảnh cùng class và phân biệt được sự khác nhau giữa hai bức ảnh không cùng class do đó nó rất là đa dụng so với loss function trước đó.
* One-shot learning:

Là thuật toán mà người dùng chỉ cần một vài hay duy nhất một ảnh để training.

* + Từ đầu vào là một bức ảnh của một người, chúng ta sẽ dùng thuật toán CNN đơn giản để dự báo người đó là ai.
  + Với One-shot learning thì đơn thuần chỉ là dữ liệu rất là ít để có thể nhận dạng và việc nhận dạng này cũng hạn chế vì nếu có một người mới không nằm trong dữ liệu training thì nó sẽ phải cho học lại thuật toán một lần nữa vì khi đó shape của output đã tăng lên 1. Vậy nếu mà trong một môi trường có một lượng người luôn luôn tăng lên như các công ty, xí nghiệp thì điều đó là hơi khó khăn vì khi học lại thuật toán thì lượng thời gian cho học lại hơn 1000 người chắc phải tốn gần một lượng lớn thời gian đấy.

**Hướng đi cho việc kết nối giữa output và khóa:**

* Arduino sẽ đóng vai trò như web client, từ đó nó sẽ dùng những giao thức truyền nhận dữ liệu, cụ thể là HTTP request để gửi và nhận yêu cầu từ web server, sau đó dùng dữ liệu trả về từ web server để điều khiển khóa điện tử, ở đây là loại động cơ khóa điện tử 5V góc quay 90 độ, kết nối với arduino nhằm thực hiện chức năng mong muốn (cụ thể ở đây là đóng, mở cửa). Về việc làm sao arduino có thể đóng vai trò như web client thì nó sẽ phải có thêm một module wifi trợ giúp cho việc kết nối mạng, đó là module ESP 8266 V1.
* Arduino sẽ lập trình cho việc đóng mở khóa. Vấn đề phát sinh là khóa phải được kết nối với arduino để thực hiện được chức năng như mong muốn, dẫn đến việc arduino và khóa đi cùng với nhau -> khá mất thẩm mỹ trong mắt người nhìn. Giải pháp tạm thời đặt ra là kết nối khóa và arduino từ xa (dùng dây dài hơn để kết nối hoặc kết nối bằng module wifi), sau đó giấu (che đi) phần mạch arduino.

**Chuẩn bị dữ liệu cho việc làm face recognition:**

* Hiện tại nhóm đã thực hiện lấy mẫu dữ liệu dựa trên khuôn mặt của 4 thành viên trong nhóm. Trong bộ dữ liệu sẽ bao gồm 1600 ảnh trong đó mỗi người có 200 ảnh lúc không mang kính và 200 ảnh lúc mang kính. Link dataset: <https://drive.google.com/drive/folders/1YImQYJB8YX5YjNjhNNEJfBBVUIwpqtJw?usp=sharing>
* Tuy nhiên, qua tìm hiểu thì nhóm có tìm được một dataset trên mạng phù hợp cho đồ án của nhóm. Link dataset: <https://www.kaggle.com/vasukipatel/face-recognition-dataset>.

Nếu được sự đồng ý từ giáo viên và doanh nghiệp hướng dẫn, nhóm sẽ sử dụng dataset này cho đồ án.

**Xác định server sử dụng:**

Nhóm đã quyết định sẽ chọn Django làm web server, vì qua việc test thử thì nó phù hợp cho việc xử lý task, điều mà nhóm em đang cần vì khối lượng thực hiện công việc của phía AI có thể lâu, nên có thể sẽ bù đắp được một phần về tốc độ. Thêm vào đó, nó có thể mở rộng một cách nhanh chóng và linh hoạt nếu như cơ sở dữ liệu có khả năng lớn lên.

Chạy thử khi kết nối với cơ sở dữ liệu thì cho tốc độ truy xuất khá nhanh:

