

TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI
VIỆN TOÁN ỨNG DỤNG VÀ TIN HỌC



BÁO CÁO THỰC TẬP KỸ THUẬT

Chuyên ngành: TOÁN TIN
Chuyên sâu: TIN HỌC

Sinh viên thực hiện:	Nguyễn Trọng Tuấn
MSSV:	20176898
Lớp:	CTTN Toán Tin K62

HÀ NỘI-2021

Mục Lục

1 Công ty cổ phần ITSOL	4
1.1 Sơ lược về ITSOL	4
1.2 Tầm nhìn	4
1.3 Sứ mệnh	4
1.4 Giá trị cốt lõi	5
2 Nội dung thực tập	6
2.1 Một số kiến thức mở đầu	6
2.1.1 Khái niệm về học máy và học sâu	6
2.1.2 Phân loại các thuật toán	6
2.1.3 Một số kiến thức cơ bản	7
2.2 Ứng dụng của máy học và bài toán nhận diện mặt người	9
2.2.1 Cấu trúc chương trình	9
2.2.2 Dữ liệu	10
2.2.3 Thuật toán Facenet	10
2.2.4 Lựa chọn triple images input	11
2.2.5 Phân tích lỗi mô hình	11
2.2.6 Kết quả huấn luyện	12

Phần Mở Đầu

Những năm gần đây, AI - Artificial Intelligence (Trí Tuệ Nhân Tạo), và cụ thể hơn là Machine Learning (Học Máy hoặc Máy Học) nổi lên như một bằng chứng của cuộc cách mạng công nghiệp lần thứ tư. Để đón đầu xu thế chung của thế giới, các công ty công nghệ ở Việt Nam cũng đang phát triển, nghiên cứu về lĩnh vực này và ITSOL cũng không phải ngoại lệ. Thật may mắn khi trong thời gian thực tập ở công ty Cổ phần ITSOL, em đã phần nào được tiếp cận và làm quen với Machine Learning và ứng dụng của nó – đó cũng là những nội dung chính của bài báo cáo này. Bài báo cáo ngoài mục lục, lời nói đầu, tài liệu tham khảo thì có các nội dung chính:

- **Phần I:** Đôi nét về công ty cổ phần ITSOL
- **Phần II:** Nội dung thực tập

Để bài báo cáo không quá dài mà vẫn đảm bảo được nội dung, các phương pháp em xin trình bày mới tính chất giới thiệu và tập trung vào ứng dụng đối với bài toán hướng đến mà không đi sâu về cơ sở lý thuyết cơ sở toán của thuật toán. Chi tiết về phần lý thuyết cơ sở toán học có trong phần danh mục tài liệu tham khảo.

Em xin cảm ơn các anh chị trong công ty cổ phần ITSOL đã nhiệt tình giúp đỡ, hướng dẫn để em hoàn thành đợt thực tập và bài báo cáo này. Mặc dù đã cố gắng nhưng chắc chắn bài báo cáo của em vẫn còn nhiều thiếu sót, mong các thầy cô cùng các bạn góp ý để bài báo cáo được hoàn thiện hơn !

1 Công ty cổ phần ITSOL

1.1 Sơ lược về ITSOL

Công ty ITSOL được thành lập từ năm 2009 bởi các sáng lập viên với trên 15 năm kinh nghiệm trong lĩnh vực CNTT, có trụ sở chính tại Hà Nội.

- ITSOL: Công ty cổ phần ITSOL.
- Địa chỉ: Duy Tân, Cầu Giấy, Hà Nội.
- Website: <https://itsol.vn>
- Ngày sinh nhật công ty: 06/06
- Hotline: (84)904.290.583
- Hotmail: Partner@itsol.vn

1.2 Tầm nhìn

Trở thành công ty số một chuyên về phát triển tuyển dụng nguồn CNTT, chuyên gia tư vấn giải pháp công nghệ và gia công phần mềm CNTT.

1.3 Sứ mệnh

Tạo ra giá trị đích thực tối đa cho khách hàng và nhà đầu tư cũng như những tài năng CNTT.

1.4 Giá trị cốt lõi

- ITSOL đối tác tin cậy về Giải Pháp và Nguồn Lực IT.
- Tại ITSOL các cơ hội phát triển nghề nghiệp đảm bảo công bằng minh bạch, luôn giúp CBNV phát huy hết khả năng chuyên môn, sáng tạo và học hỏi.
- Mỗi vị trí sẽ được xác định dựa vào năng lực của CBNV và có những chế độ đãi ngộ tương tự.

2 Nội dung thực tập

2.1 Một số kiến thức mở đầu

2.1.1 Khái niệm về học máy và học sâu

Theo Wikipedia, học máy (machine learning) là một lĩnh vực của trí tuệ nhân tạo liên quan đến việc nghiên cứu và xây dựng các kĩ thuật cho phép các hệ thống ”học tự động từ dữ liệu để giải quyết những vấn đề cụ thể.

Học máy có mối quan hệ mật thiết với AI và Big data. Nếu AI là một mục tiêu của con người thì học máy là phương tiện được kì vọng sẽ giúp con người thực hiện mục tiêu đó. Ở mức độ nhỏ hơn, học máy có những mục tiêu trước mắt như:

- Làm cho máy tính có những khả năng nhận thức cơ bản của con người như nghe, nhìn, hiểu được ngôn ngữ, giải toán, lập trình, ...
- Hỗ trợ con người trong việc xử lý một khối lượng thông tin khổng lồ mà chúng ta phải đối mặt hàng ngày, hay còn gọi là Big Data.

Học sâu (deep learning) là một chi của ngành máy học dựa trên một tập hợp các thuật toán để cố gắng mô hình dữ liệu trừu tượng hóa ở mức cao bằng cách sử dụng nhiều lớp xử lý với cấu trúc phức tạp, hoặc bằng cách khác bao gồm nhiều biến đổi phi tuyến.

2.1.2 Phân loại các thuật toán

Có nhiều cách phân loại các thuật toán các thuật toán machine learning. Trong đó phổ biến nhất là phân loại dựa trên phương thức học (learning style).

- Supervised Learning (Học có giám sát): Classification (Phân loại); Regression (Hồi quy)

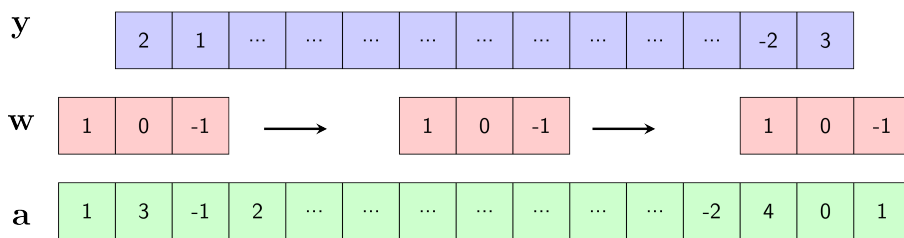
- Unsupervised Learning (Học không giám sát): Clustering (phân nhóm); Association
- Semi-Supervised Learning (Học bán giám sát)
- Reinforcement Learning (Học Củng Cố)

2.1.3 Một số kiến thức cơ bản

Tích chập (Convolution)

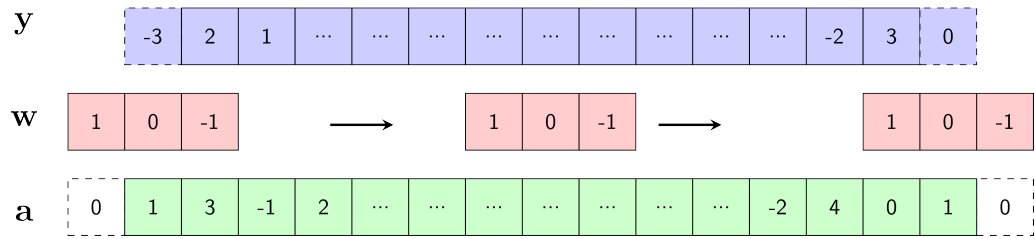
Xét tín hiệu một chiều $a(n)$ và bộ lọc (filter) $w(n)$. Tích chập của tín hiệu và bộ lọc là một tín hiệu một chiều mới $b(n)$ được xác định theo công thức:

$$b(n) = a(n) \times w(n) = \sum_{k=-\infty}^{\infty} a(k)w(n-k)$$



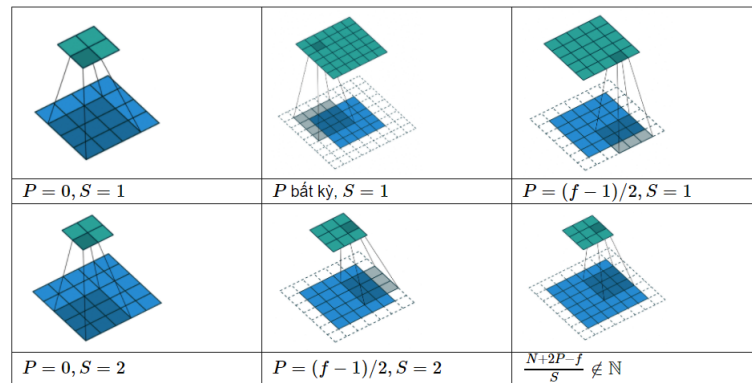
Hình 2.1: Cách tính tích chập một chiều

Nhận thấy rằng kích thước của đầu ra y nhỏ hơn kích thước của đầu vào x vì ta đang giả sử cả bộ lọc phải nằm trọn vẹn trong tín hiệu đầu vào. Trong trường hợp muốn tín hiệu đầu ra có kích thước bằng tín hiệu đầu vào, ta có thể giả sử giả sử tín hiệu đầu vào có thêm các giá trị bằng không ở hai phía.



Hình 2.2: Tích chập với thêm lề

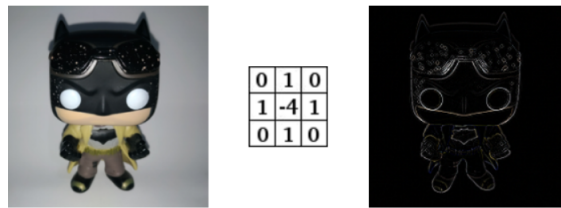
Tích chập được mở rộng ra cho trường hợp dữ liệu nhiều chiều. Dưới đây là hình ảnh minh họa tích chập hai chiều:



Hình 2.3: Tích chập hai chiều

Mạng neurol tích chập (CNN)

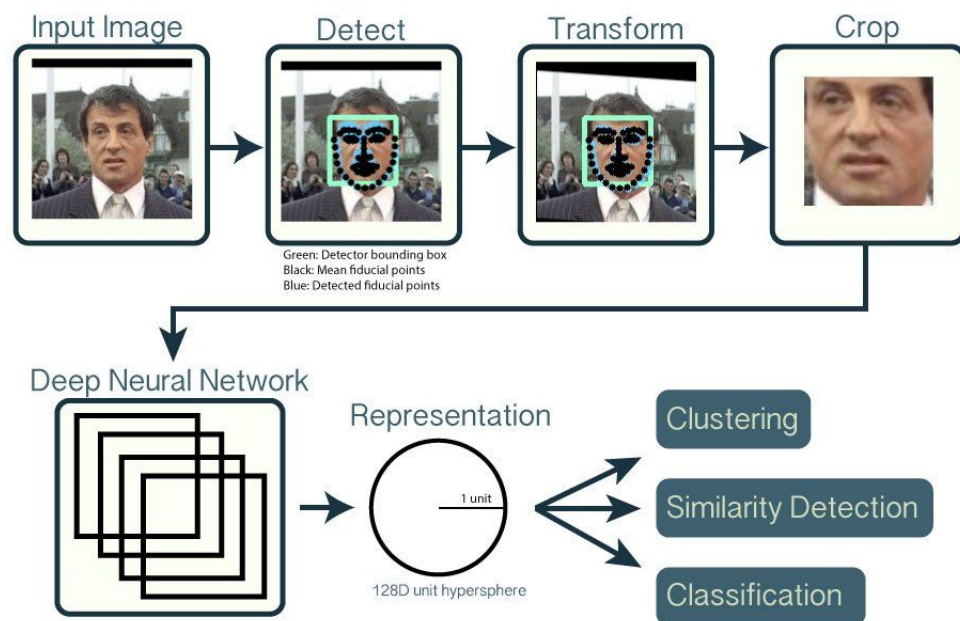
Mạng CNN là một mô hình học sâu sử dụng nhiều trong các bài toán nhận dạng các vật thể trong ảnh bằng cách sử dụng các lớp tích chập (Convolutional) để trích xuất các đặc trưng của vật thể.



Hình 2.4: Ví dụ về mạng CNN

2.2 Ứng dụng của máy học và bài toán nhận diện mặt người

2.2.1 Cấu trúc chương trình



Hình 2.5: Cấu trúc của một chương trình nhận diện mặt người

- Face Detection: Xác định vùng có mặt người trong ảnh
- Face Alignment: Căn chỉnh lại ảnh mặt người để tăng độ hiệu quả
- Face Representation: Trích xuất các đặc trưng của ảnh thành một vecto

2.2.2 Dữ liệu

- CASIA-WebFace: Bộ dữ liệu bao gồm gần 500 nghìn ảnh được thu thập từ khoảng 10 nghìn người.
- VGGFace2: Bộ dữ liệu gồm khoảng 3 triệu ảnh được thu thập từ gần 9 nghìn người.

2.2.3 Thuật toán Facenet

Facenet chính là một dạng siam network có tác dụng biểu diễn các bức ảnh trong một không gian euclidean n chiều (thường là 128) sao cho khoảng cách giữa các véc tơ embedding càng nhỏ, mức độ tương đồng giữa chúng càng lớn.

Hầu hết các thuật toán nhận diện khuôn mặt trước facenet đều tìm cách biểu diễn khuôn mặt bằng một véc tơ embedding thông qua một layer bottle neck có tác dụng giảm chiều dữ liệu.

- Tuy nhiên hạn chế của các thuật toán này đó là số lượng chiều embedding tương đối lớn và ảnh hưởng tới tốc độ của thuật toán. Thường chúng ta phải áp dụng thêm thuật toán PCA để giảm chiều dữ liệu để giảm tốc độ tính toán.
- Hàm loss function chỉ đo lường khoảng cách giữa 2 bức ảnh. Như vậy trong một đầu vào huấn luyện chỉ học được một trong hai khả năng là sự giống nhau nếu chúng cùng 1 class hoặc sự khác nhau nếu chúng khác class mà không học được cùng lúc sự giống nhau và khác nhau trên cùng một lượt huấn luyện.

Facenet đã giải quyết cả 2 vấn đề trên bằng các hiệu chỉnh nhỏ nhưng mang lại hiệu quả lớn:

- Base network áp dụng một mạng convolutional neural network và giảm chiều dữ liệu xuống chỉ còn 128 chiều. Do đó quá trình suy diễn và dự báo nhanh hơn và đồng thời độ chính xác vẫn được đảm bảo.
- Sử dụng loss function là hàm triplet loss có khả năng học được đồng thời sự giống nhau giữa 2 bức ảnh cùng nhóm và phân biệt các bức ảnh không cùng nhóm. Do đó hiệu quả hơn rất nhiều so với các phương pháp trước đây.

2.2.4 Lựa chọn triple images input

Để mô hình khó học hơn và đồng thời cũng giúp mô hình phân biệt chuẩn xác hơn mức độ giống và khác nhau giữa các khuôn mặt, chúng ta cần lựa chọn các input theo bộ 3 khó học (hard triplets).

Ý tưởng là chúng ta cần tìm ra bộ ba (A, N, P) sao cho $d(A, P)$ lớn nhất và $d(A, N)$ nhỏ nhất. Hay nói cách khác với mỗi Anchor A cần xác định:

- Hard Positive: Bức ảnh Positive có khoảng cách xa nhất với Anchor
- Hard Negative: Bức ảnh Negative có khoảng cách gần nhất với Anchor

Việc tính toán các trường hợp Hard Positive và Hard Negative có thể được thực hiện offline và lưu vào checkpoint hoặc có thể tính toán online trên mỗi mini-batch.

2.2.5 Phân tích lỗi mô hình

Ta nhận thấy các bức ảnh dự đoán sai đa phần là rơi vào các trạng thái:

- Nhân vật đang cười.
- Chụp ở một góc nghiêng.
- Đội nón.
- Đeo kính

Wrong predict images



Hình 2.6: Các trường hợp nhận diện sai

Như vậy để tăng cường hiệu quả nhận dạng thì:

- Các bức ảnh được huấn luyện nên đồng nhất về điều kiện chiếu sáng.
- Khuôn mặt của nhân vật cần được giữ ở trạng thái thả lỏng, không nhăn nhó.
- Nên thu thập đa dạng các góc nhìn của khuôn mặt. Không chỉ từ góc thẳng đứng mà còn góc nghiêng và các góc độ khác nhau. Chẳng hạn như ảnh quay tròn khuôn mặt như bên dưới.

2.2.6 Kết quả huấn luyện

Model name	LFW accuracy	Training dataset	Architecture
20180408-102900	0.9905	CASIA-WebFace	Inception ResNet v1
20180402-114759	0.9965	VGGFace2	Inception ResNet v1

Hình 2.7: Kết quả huấn luyện mô hình