

# Face Processing in Video

Tác giả thứ nhất<sup>1,3</sup>

Tác giả thứ 2<sup>2</sup>

Tác giả thứ 3<sup>3,1</sup> and Tác giả thứ 4<sup>3,1</sup>

<sup>1</sup> Trường ĐH.....

<sup>2</sup> University of Science  
HCMC, Vietnam

<sup>3</sup> National Institute of Informatics

## What ?

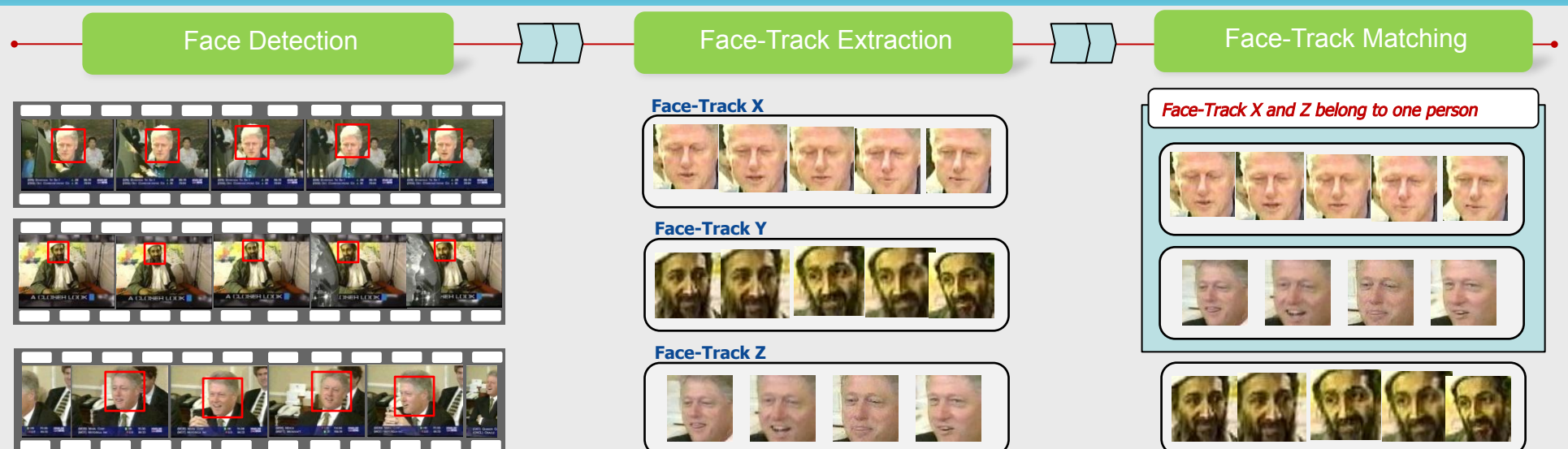
We introduce a framework to process and retrieve faces in video, in which we have:

- Proposed a robust method to extract face-tracks in news video.
- Built the largest face database compare to current popular worldwide face databases.
- Evaluated several face-track matching methods.

## Why ?

- The human face is one of the most important objects in video since it provides rich information for spotting people of interest and is the basis for interpreting facts. Therefore, detecting and recognizing faces appearing in video are essential tasks of video indexing and retrieval applications.
- Most studies have focused on static images rather than **large-scale** and **real video dataset**.

## Overview



## Description

### 1. Face Detection

- The face detector implemented in OpenCV based on Viola method was used for detecting frontal views of faces in every frame of our video sequences.
- A high threshold was used to reduce the number of false positives.

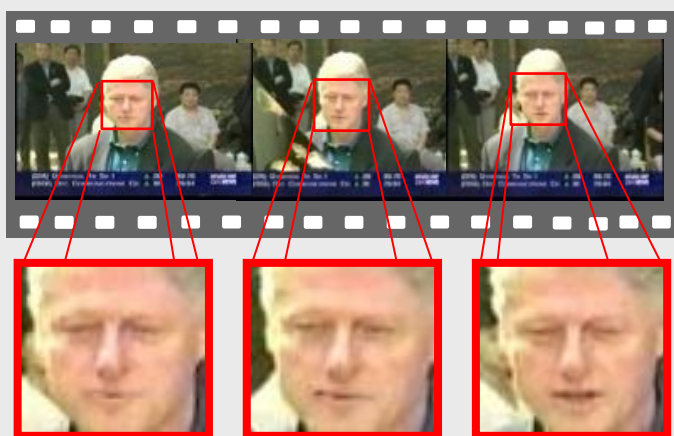


Figure 1. Detection results with a high threshold.

### 3. Face-Track Matching

- Face-track matching is done by applying similarity estimation methods (e.g. Min-Min distance) on the feature space.

### 2. Face-Track Extraction

- We used Kanade–Lucas–Tomasi (KLT) method to create and track key/interest points between frames.
- The number of key points that pass through pairs of faces in consecutive frames was computed to make decision on grouping faces into face-tracks.
- Several treatments are proposed to handle tracking traps in news video:
  - Flash-frame detector.
  - Adaptively generating key points.

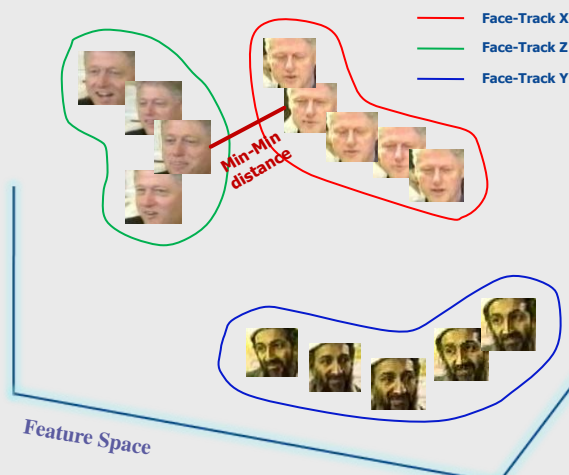


Figure 4. Apply Min-Min method for face-track matching.

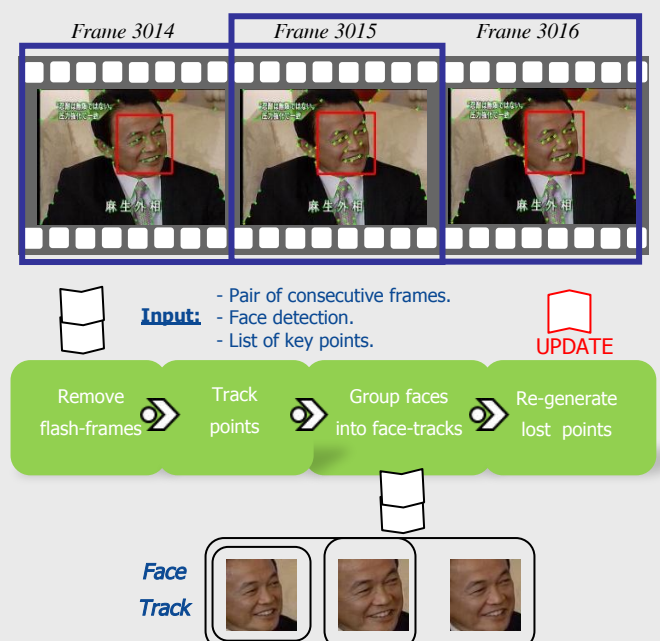


Figure 2. Process-flow of face tracker.

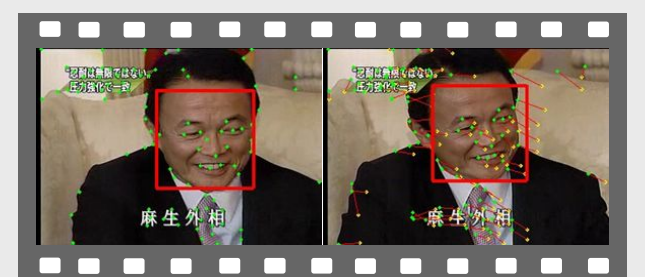


Figure 3. Key/Interest points (plotted as green dots) in the left frame are tracked in the right frame. Small lines from green dots are motion of these points. Two faces in these frames share 22/23 points.

# MỘT KHUÔN KHỔ HỖ TRỢ LỰA CHỌN DỊCH VỤ CỦA CÁC NHÀ CUNG CẤP ĐIỆN TOÁN Đám Mây DỰA TRÊN ĐIỂM CHUẨN CẤU HÌNH MÁY ẢO

Nguyễn Hồng Sơn<sup>1,2</sup>

<sup>1</sup> Trường ĐH Công nghệ Thông tin

<sup>2</sup> Đại học Quốc gia Thành phố Hồ Chí Minh

## Mục tiêu ?

Giới thiệu khuôn khổ điểm chuẩn và trực quan hóa thông số máy ảo của các nhà cung cấp điện toán đám mây :

- Khả năng tạo số lượng lớn máy ảo và chạy điểm chuẩn bằng các phần mềm sẵn có.
- Tạo bộ dữ liệu điểm chuẩn theo thời gian các loại máy ảo của nhiều nhà cung cấp như GCP và AWS
- Thực hiện ứng dụng trực quan hóa trên website.

## Why ?

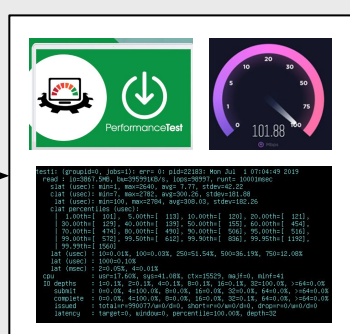
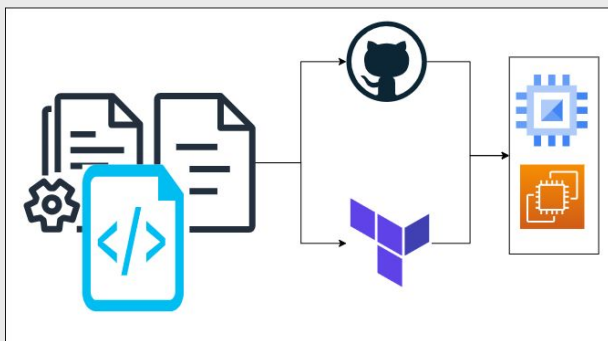
- Sự phát triển nhanh chóng của các công ty cung cấp điện toán đám mây tạo ra rất nhiều dịch vụ. Việc hiểu hết và đánh giá đúng chất lượng, giá cả là một thách thức.
- Các công ty chỉ giới thiệu sản phẩm của mình mà chưa có các tổng hợp so sánh với đối thủ cạnh tranh.
- Cung cấp đánh giá **khách quan, dễ hiểu** sẽ giúp người dùng dễ lựa chọn và tiết kiệm chi phí hơn.

## Overview

Tạo tài nguyên

Thực hiện điểm chuẩn

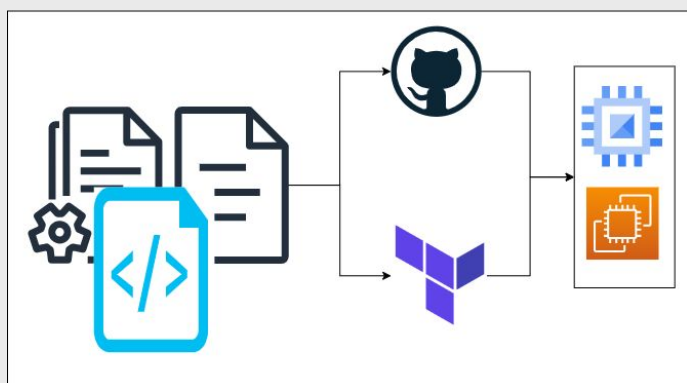
Biểu đồ trực quan dữ liệu



## Description

### 1. Tạo tài nguyên

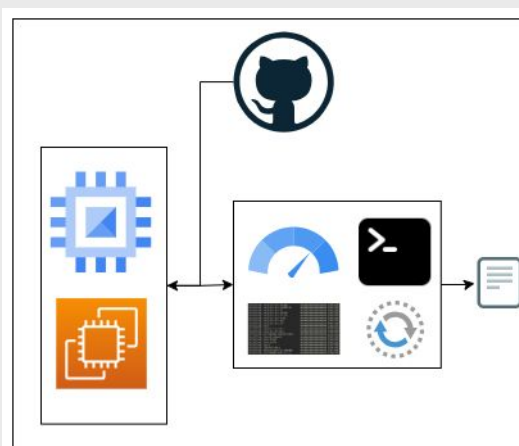
- Sử dụng thư viện được cung cấp bởi Terraform có tích hợp đồng thời cả AWS và GCP, là nền tảng IaC có thể dễ dàng mở rộng và quản lý linh hoạt.
- Các tài nguyên được tạo nhanh chóng, tự động thực thi phần mềm điểm chuẩn đã thiết kế từ trước.



### 2. Điểm chuẩn máy ảo

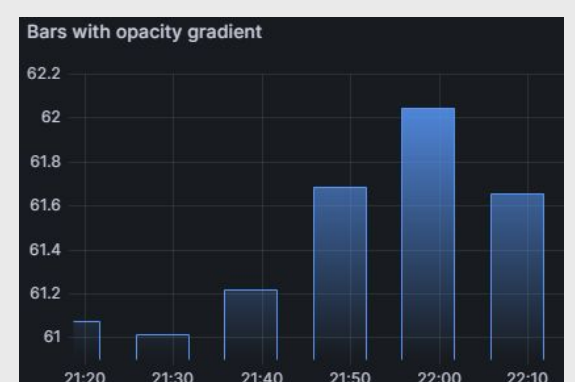
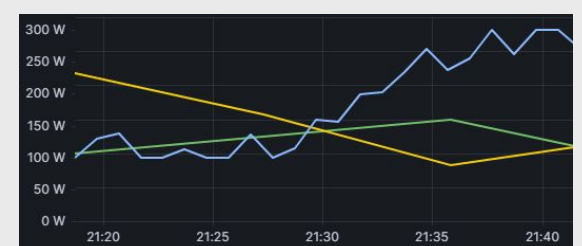
- Passmark PerformanceTest, FIO, IOzone, Speedtest là những công cụ điểm chuẩn miễn phí, có nhiều người sử dụng, kết quả có tham chiếu so sánh.

- Các tham số truyền vào phù hợp với mục đích là kiểm tra hiệu năng tài nguyên máy ảo cung cấp
- Cài đặt các thư viện, phần mềm bằng script. Kết quả trả về cụ thể và đúng như mong muốn
- Sau khi có được dữ liệu thô, xử lý để có dữ liệu sạch và gửi về máy chủ trung tâm để lưu trữ.
- Dự án sử dụng MySQL làm cơ sở dữ liệu, để sử dụng và tăng tính mở rộng, đồng thời thương tích với website.



### 3. Kết quả thực hiện

- Ứng dụng website hiển thị các thông số theo thời gian
- Người dùng có thể tương tác với các biểu đồ để xem chi tiết thông số hoặc khoảng thời gian truy vấn.





# ĐÁNH GIÁ TÁC ĐỘNG CỦA ĐIỀU KIỆN THỜI TIẾT ĐẾN VIỆC TRÌ HOÃN CÁC CHUYẾN BAY TRÊN THẾ GIỚI

Nguyễn Cường Phát<sup>1</sup>

<sup>1</sup> Trường Đại học Công nghệ Thông tin  
ĐHQG TP HCM

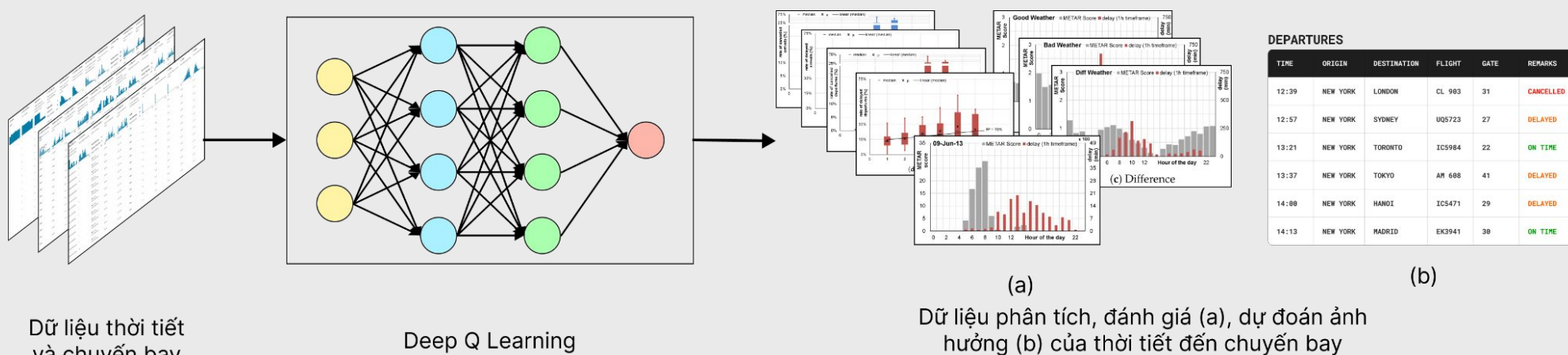
## Mục tiêu

- Đánh giá tác động của điều kiện thời tiết đến việc trì hoãn các chuyến bay trên toàn cầu và xây dựng một mô hình DQN để ước lượng mức độ tác động.
- So sánh kết quả của mô hình DQN với các phương pháp khác như hồi quy tuyến tính và mô hình Random Forest
- Cung cấp cái nhìn sâu sắc về tác động của điều kiện thời tiết đến hoạt động hàng không và đóng góp vào việc tối ưu hóa quản lý và vận hành chuyến bay trên toàn cầu.

## Lý do chọn đề tài

- Điều kiện thời tiết có thể ảnh hưởng đến an toàn của các chuyến bay. Việc đánh giá tác động của điều kiện thời tiết đến trì hoãn chuyến bay sẽ giúp nhận biết các yếu tố nguy hiểm và đưa ra các biện pháp để tăng cường an toàn trong các điều kiện thời tiết khắc nghiệt.
- Các nghiên cứu trước đây đã tìm hiểu về tác động của điều kiện thời tiết đến việc trì hoãn chuyến bay cho một số khu vực, tuy nhiên, còn hạn chế về việc đánh giá chi tiết cho các chuyến bay trên toàn thế giới.

## Overview



## Description

### 1. Nội dung

- Khảo sát và nghiên cứu bộ dữ liệu "2019 Airline Delays w/Weather and Airport Detail"
- Xây dựng mô hình Deep Q-Network (DQN) ước lượng tác động của điều kiện thời tiết đến việc trì hoãn chuyến bay.
- Đánh giá mức độ tương quan giữa điều kiện thời tiết và trì hoãn chuyến bay dựa trên dữ liệu thực tế
- Đề xuất khuyến nghị và thông tin hữu ích nhằm giảm thiểu trì hoãn chuyến bay dựa trên thông tin về điều kiện thời tiết.

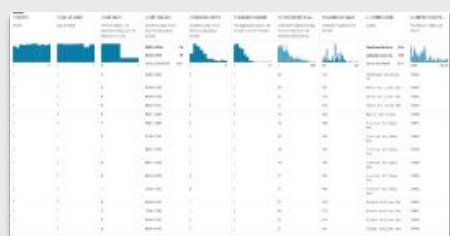


Figure 1. 2019 Airline Delays w/Weather and Airport Detail dataset

### 2. Phương pháp nghiên cứu

- Thu thập và xử lý bộ dữ liệu "2019 Airline Delays w/Weather and Airport Detail".

- Phân tích và khảo sát bộ dữ liệu để hiểu về cấu trúc và thông tin liên quan đến chuyến bay, điều kiện thời tiết, và sân bay.
- Xây dựng mô hình DQN, huấn luyện mô hình trên bộ dữ liệu đã chuẩn bị, sử dụng các thuật toán tối ưu hóa như thuật toán Q-learning để tối đa hóa hiệu suất dự đoán.
- Phân tích mức độ tương quan giữa các yếu tố thời tiết và thời gian trì hoãn chuyến bay bằng các phương pháp thống kê và mô hình hóa.
- So sánh với kết quả dự đoán của mô hình hồi quy tuyến tính, mô hình Random Forest.
- Dựa trên kết quả phân tích, đề xuất các khuyến nghị và thông tin hữu ích cho các quyết định vận hành và quản lý hàng không, nhằm giảm thiểu tác động của điều kiện thời tiết đến trì hoãn chuyến bay.

### 3. Kết quả dự kiến

- Kết quả đánh giá chi tiết tác động của điều kiện thời tiết đến việc trì hoãn chuyến bay trên toàn cầu.
- Mô hình DQN được xây dựng để ước lượng mức độ tác động một cách chính xác và hiệu quả.

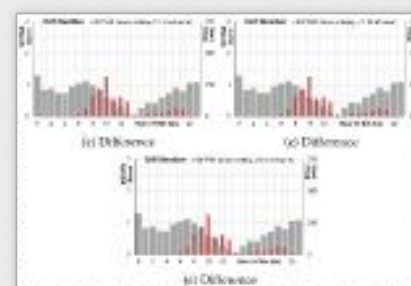


Figure 2. Đánh giá chi tiết tác động của thời tiết đến việc trì hoãn chuyến bay

- Cung cấp thông tin quan trọng về mối quan hệ giữa điều kiện thời tiết và trì hoãn chuyến bay, đóng góp vào việc hiểu rõ hơn về ảnh hưởng của yếu tố thời tiết đối với hoạt động hàng không trên toàn cầu.
- Hỗ trợ các nhà quản lý hàng không, các chuyên gia dự báo thời tiết và các bên liên quan trong việc ra quyết định và lập kế hoạch vận hành, từ đó giảm thiểu trì hoãn chuyến bay và cải thiện trải nghiệm khách hàng.
- Kết quả so sánh hiệu quả của mô hình DQN với các phương pháp khác như hồi quy tuyến tính và mô hình Random Forest trong việc dự đoán và ước lượng tác động của điều kiện thời tiết.

# DATA MANAGEMENT TRADITIONAL AND BLOCKCHAIN DATABASES

Tác giả: Trần Giang Nam

Trường ĐH Công nghệ thông tin TP.HCM

## What ?

So sánh, đánh giá ưu điểm và khuyết điểm Hệ quản trị cơ sở dữ liệu quan hệ MySQL và Cơ sở dữ liệu với Công nghệ Blockchain.

Cung cấp thông tin giúp nhà phát triển phần mềm ra quyết định chọn loại cơ sở nào để triển khai hệ thống phần mềm của họ

## Why ?

Sự phát triển đa dạng về công nghệ lưu trữ dẫn đến một số khó khăn trong việc lựa chọn loại cơ sở dữ liệu để lưu trữ dữ liệu tối ưu. Nghiên cứu với mục tiêu cung cấp thêm thông tin giúp Nhà phát triển phần mềm tương lai có góc nhìn tốt hơn khi ra quyết định chọn loại Cơ sở dữ liệu nào để triển khai cho các hệ thống phần mềm của mình.

## Overview

### MySQL

Nghiên cứu Cơ sở dữ liệu với MySQL: định nghĩa cấu trúc dữ liệu, cài đặt, triển khai, truy vấn, thao tác, dữ liệu.

### Blockchains

Nghiên cứu Cơ sở dữ liệu với Blockchain: định nghĩa cấu trúc dữ liệu, cài đặt, triển khai, truy vấn, thao tác, dữ liệu.

### Experimental

Triển khai thực nghiệm với bài toán giả định "Energy Trading Platform" (ETP).

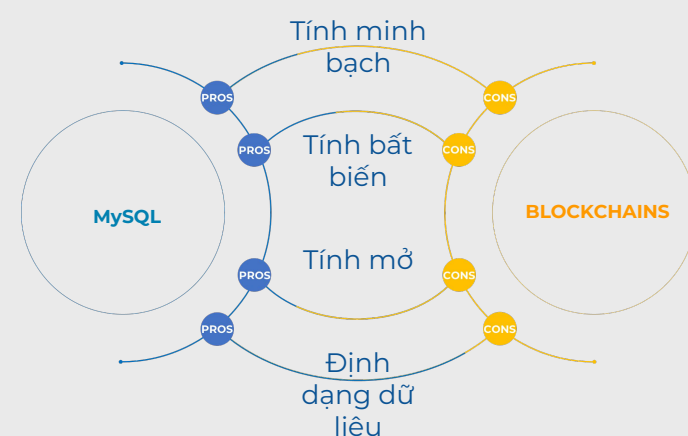
### Comparison

Tổng hợp, so sánh, thông tin ưu/nhược điểm của 2 loại cơ sở dữ liệu, từ đó đưa ra đánh giá nhận định có giá trị

## Description

**MySQL:** Thực nghiệm với cơ sở dữ liệu quan hệ MySQL: Định nghĩa cấu trúc dữ liệu lưu trữ, triển khai và cài đặt dữ liệu, truy vấn thông tin và thao tác với dữ liệu.

**Kết quả mong đợi:** Thông tin so sánh, đánh giá về các đặc điểm, ưu điểm-nhược điểm của hai loại Cơ sở dữ liệu quan hệ MySQL (truyền thống) và Cơ sở dữ liệu với công nghệ Blockchain trên các khía cạnh bên dưới.



1

DEFINE RELATIONS

Cấu trúc dữ liệu lưu trữ.  
Chuẩn hóa dữ liệu  
Thiết lập quan hệ  
(Công cụ: DBMS)

2

DEPLOY RELATIONS

Thực hiện các Query khởi tạo các Relations: Users, Energy Type, Trade Account, Buyer, Sellers,...  
(Công cụ: DBMS)

3

STORE DATA VIA RELATIONS

Thực hiện lưu trữ dữ liệu vào các Relations: Users, Energy Type, Trade Account, Buyer, Sellers,...  
(Công cụ: DBMS/Webpage)

4

RETRIEVE DATA TO FORM A TRANSITION

Thực hiện truy vấn dữ liệu: Users, Energy Type, Trade Account, Buyer, Sellers,...  
(Công cụ: DBMS/Webpage)

**Blockchains:** Thực nghiệm với cơ sở dữ liệu quan hệ MySQL: Định nghĩa cấu trúc dữ liệu lưu trữ, cách triển khai và cài đặt dữ liệu, cách truy vấn thông tin và thao tác với dữ liệu.

1

WRITE SMART CONTRACT

Tạo Smart Contract để lưu trữ dữ liệu.  
(Công cụ: Remix)

2

DEPLOY SMART CONTRACT

Triển khai Smart Contract trên Ethereum Test Network (Ropsen)  
(Công cụ: Ropsen, Metamask, Ether toping up sites)

3

STORE DATA VIA CONTRACT ADDRESS

Thực hiện lưu trữ dữ liệu các Transactions vào Smart Contract  
(Công cụ: Web3, ReactJS, Manually confirm payment of GAS)

4

RETRIEVE DATA TO FORM A TRANSITION

Thực hiện truy vấn dữ liệu blockchain thông qua Smart Contract Address  
(Công cụ: Web3.Provider, Filter on Timestamp for each period)



# GÁN NHÃN CHO HÌNH ẢNH

Trương Thanh Luân

Trường ĐH Công Nghệ Thông Tin

## What ?

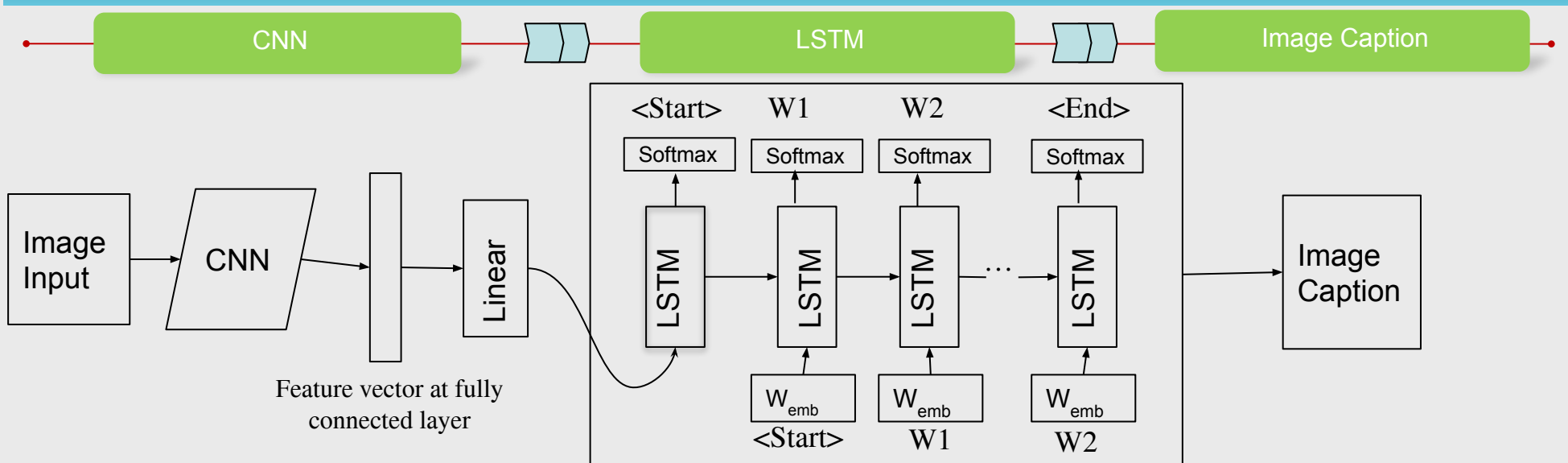
Kết hợp mô hình CNN và LSTM để gán nhãn cho hình ảnh:

- CNN được sử dụng để trích xuất các đặc trưng hình ảnh.
- LSTM sử dụng thông tin từ CNN để tạo mô tả cho hình ảnh.
- Gán nhãn cho hình ảnh.

## Why ?

- Giúp những người khiếm thính hoặc các em nhỏ có thể biết được cảnh vật xung quanh. Ví dụ: Image -> text -> voice.
- Quản lý và tìm kiếm hình ảnh theo nội dung hình ảnh.

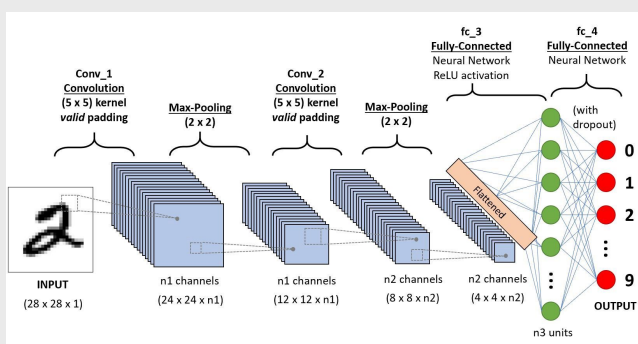
## Overview



## Description

### 1. Mô hình CNN

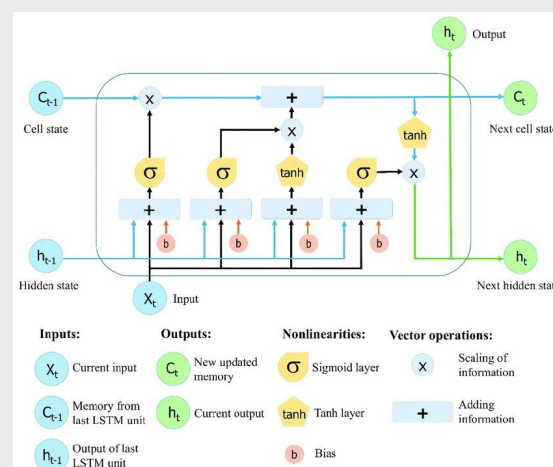
- Các lớp tích chập của CNN giúp tìm ra các đặc trưng từ hình ảnh. Ví dụ: cấu trúc như cạnh, góc, hoặc hình dạng của các đối tượng trong hình ảnh.
- Các đặc trưng hình ảnh sẽ được trích xuất từ InceptionV3, mô hình CNN được đào tạo trên bộ dữ liệu Imagenet. Sử dụng các lớp tích chập và lớp pooling, CNN có khả năng tự động học và trích xuất các đặc trưng quan trọng từ hình ảnh.



Hình 1. Kiến trúc mô hình CNN

### 2. Mô hình LSTM

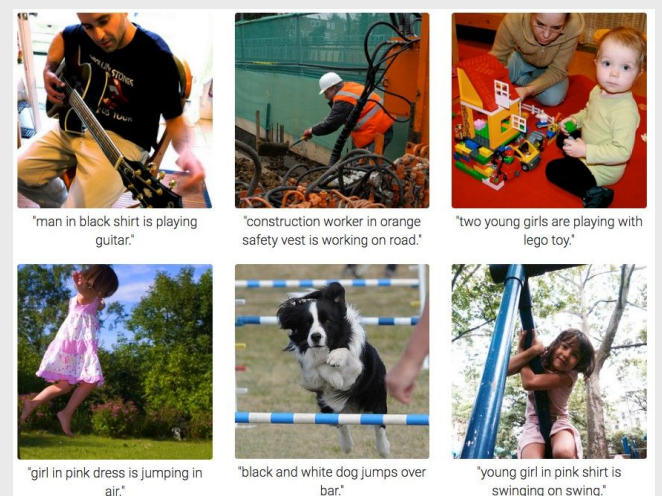
- Khả năng ghi nhớ và xử lý thông tin từ quá khứ.
- Phân tích chuỗi hình ảnh liên tiếp để dự đoán kết quả trong tương lai.
- Chống hiện tượng "Vanishing Gradient".



Hình 2. Kiến trúc mô hình LSTM

### 3. Kết quả mong đợi

- Độ chính xác cao.
- Tự động hóa quá trình gán nhãn.
- Xử lý hình ảnh phức tạp và chuỗi hình ảnh.
- Tự động hóa quá trình gán nhãn.



Hình 3. Ảnh được gán nhãn từ mô hình CNN và LSTM

# Improving Spam Detection Performance With BERT

Nguyễn Duy Phương

Trường Đại Học Công Nghệ Thông Tin - DHQG TP.HCM

## What ?

Nghiên cứu này nhằm tận dụng sức mạnh của BERT để:

- Tận dụng sức mạnh của BERT để xây dựng một hệ thống phát hiện thư rác hiệu quả và chính xác.
- Nâng cao hiệu suất phân loại giữa thư rác và email hợp lệ bằng cách sử dụng BERT và tinh chỉnh mô hình.
- Cải thiện quy trình mã hóa và đại diện đầu vào cho BERT.

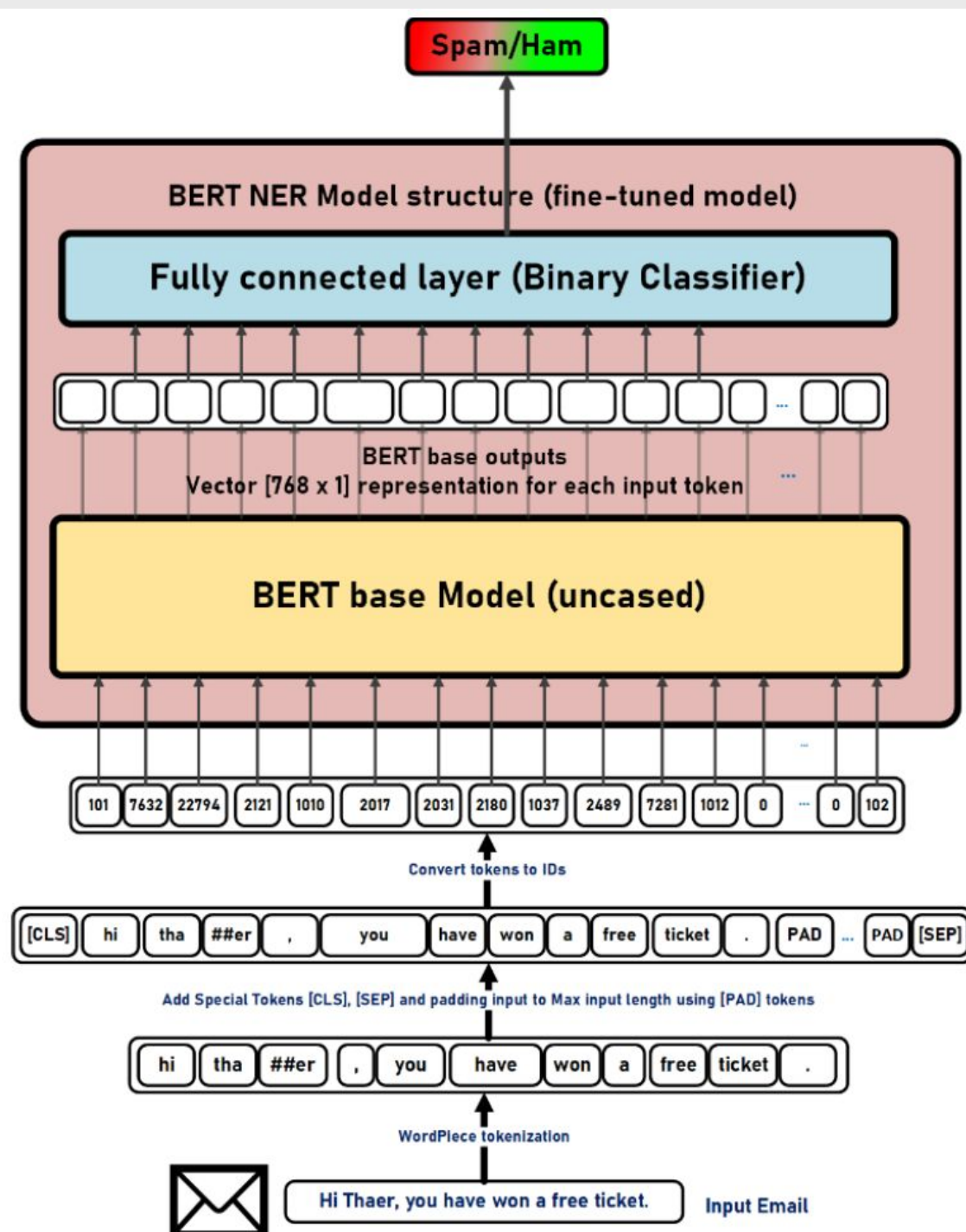
## Why ?

- Phát hiện thư rác là một nhiệm vụ quan trọng vì nó bảo vệ người dùng khỏi nội dung độc hại, tiết kiệm tài nguyên mạng và chống lại tội phạm mạng. Nó ngăn chặn việc tiết lộ thông tin cá nhân và tài sản, giảm lãng phí tài nguyên bằng thông và dung lượng lưu trữ không cần thiết, và ngăn chặn các hình thức tấn công và lừa đảo trực tuyến, bảo vệ người dùng và tổ chức.

## Overview



## Description



## 4. Đánh giá

- Bộ dữ liệu đánh giá được sử dụng để đánh giá hiệu suất của mô hình được đào tạo.

## 3. Đào tạo mô hình

- Mô hình cơ sở BERT được đào tạo bằng PyTorch. Email và nhãn tương ứng được cung cấp cho mô hình theo lô 16.

## 2. Tạo tập dữ liệu đào tạo

- Tập hợp dữ liệu được xử lý trước được chia thành tập dữ liệu đào tạo và đánh giá. Tập dữ liệu đào tạo được sử dụng để đào tạo mô hình phát hiện thư rác.

## 1. Tiền xử lý dữ liệu

- Dữ liệu email từ SpamAssassin, SMS Spam Collection v.1, Enron corpus và Ling-Spam corpus được làm sạch bằng cách xóa URL, ký tự không mong muốn và thẻ HTML.