

The background of the entire image is a dark blue field filled with a pattern of red dots. These dots are arranged in a way that they form a large, faint, stylized 'H' shape that frames the central text. The dots are of varying sizes and are more densely packed in some areas, creating a sense of depth and movement.

# HUST

**ĐẠI HỌC BÁCH KHOA HÀ NỘI**  
HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

ONE LOVE. ONE FUTURE.



**ĐẠI HỌC  
BÁCH KHOA HÀ NỘI**  
HANOI UNIVERSITY  
OF SCIENCE AND TECHNOLOGY

# HỌC SÂU VÀ ỨNG DỤNG

*Bài toán gán ảnh đa nhãn: Hanoi Tourist*

**GVHD: TS. Trịnh Anh Phúc**

**Nhóm sinh viên: Nhóm 2**

ONE LOVE. ONE FUTURE.

1. Giới thiệu bài toán
2. Phân công công việc
3. Thu thập bộ dữ liệu huấn luyện
4. Tiền xử lý dữ liệu
5. Huấn luyện mô hình

# 1. Giới thiệu bài toán

## • Bài toán:

Phân biệt và gán đa nhãn cho ảnh 2 địa điểm du lịch tại Hà Nội: **Hồ Gươm** (HG) và **Hồ Tây** (HT).

- Các nhãn (đối tượng) cần phát hiện trong ảnh:
  - Tháp Rùa (TR),
  - Cầu Thê Húc (CTH),
  - Bưu điện (BD),
  - Vườn hoa (VH),
  - Chùa Trấn Quốc (CTQ),
  - Đền Quán Thánh (DQT),
  - Khách sạn (KS),
  - Vườn hoa (VH).
  - Công viên nước (CNV)

# 1. Giới thiệu bài toán

- Ví dụ:



- Hồ Gươm
- Cầu Thê Húc



- Hồ Tây
- Chùa Trấn Quốc

## 2. Phân công công việc

- **Các công việc đặt ra:**

- a) Thu thập bộ dữ liệu huấn luyện
- b) Tiền xử lý dữ liệu
- c) Gán nhãn cho bộ dữ liệu
- d) Huấn luyện, kiểm thử mô hình phân loại

## 2. Phân công công việc

### a) Thu thập bộ dữ liệu huấn luyện

- Thu thập ảnh hồ Tây:
  - Phạm Duy Tùng
  - Trần Hữu Hiếu
  - Hồ Viết Đức Lương
- Thu thập ảnh hồ Gươm:
  - Phạm Thành Dũng
  - Trương Đăng Biên

## 2. Phân công công việc

### a) Thu thập bộ dữ liệu huấn luyện

- **Yêu cầu chung:**

- Chứa các thông tin đủ để phân biệt ảnh tại hai vị trí phân biệt
- Có đối tượng được chụp nằm giữa khung hình
- Có khung cảnh – background
- Chụp tự nhiên
- Người xem có thể gán nhãn và phân biệt các đối tượng dựa theo thông tin có sẵn trong ảnh.
- **KHÔNG** thu thập các ảnh/dữ liệu mang tính **đánh đổ** người gán nhãn hay mô hình phân loại



## 2. Phân công công việc

### b) Tiền xử lý dữ liệu

- Đặng Tiến Đạt: Lọc, cắt ảnh hồ Gươm.
- Trần Anh Quốc: Lọc, cắt ảnh hồ Tây.
- Nguyễn Minh Hiếu: Làm giàu dữ liệu, resize ảnh, quản lí sheet gán nhãn.

### c) Gán nhãn

- Đảm bảo tính khách quan: mô phỏng lại hình thức bồi thẩm đoàn: toàn bộ 12 người độc lập đánh giá xem một ảnh trong bộ dữ liệu mang nhãn gì và lựa chọn kết quả theo số đông.

## 2. Phân công công việc

### d) Huấn luyện mô hình phân loại

- Nguyễn Duy Long
- Lê Đức Minh
- Vũ Anh Đức
  
- Nhóm trưởng: Nguyễn Trí Lâm

# 3. Thu thập bộ dữ liệu huấn luyện

## a) Thu thập ảnh Hồ Tây: 3 buổi

- Buổi 1:
  - Thời gian: buổi chiều, 19/10/2023
  - Thực hiện: Lương + Tùng
  - Nội dung:
    - Xung quanh hồ Tây
    - Đền Quán Thánh
    - Chùa Trấn Quốc
  - Kết quả: ~150 bức ảnh (thô)

# 3. Thu thập bộ dữ liệu huấn luyện

## a) Thu thập ảnh Hồ Tây:

- Buổi 1:



# 3. Thu thập bộ dữ liệu huấn luyện

## a) Thu thập ảnh Hồ Tây: 3 buổi

- Buổi 2:
  - Thời gian: buổi sáng, 20/10/2023
  - Thực hiện: Lương + Tùng
  - Nội dung:
    - Xung quanh hồ Tây
    - Đền Quán Thánh
    - Công viên nước hồ Tây
  - Kết quả: ~150 bức ảnh (thô)



# 3. Thu thập bộ dữ liệu huấn luyện

## a) Thu thập ảnh Hồ Tây: 3 buổi

- Buổi 2:



# 3. Thu thập bộ dữ liệu huấn luyện

## a) Thu thập ảnh Hồ Tây: 3 buổi

- Buổi 3:
  - Thời gian: buổi sáng, 21/10/2023
  - Thực hiện: Hữu Hiếu
  - Nội dung:
    - Xung quanh hồ Tây
    - Đền Quán Thánh
    - Vườn hoa
    - Khách sạn và bưu điện
  - Kết quả: ~127 bức ảnh (thô)

# 3. Thu thập bộ dữ liệu huấn luyện

## a) Thu thập ảnh Hồ Tây: 3 buổi

- Buổi 3:





# 3. Thu thập bộ dữ liệu huấn luyện

## a) Thu thập ảnh Hồ Tây: 3 buổi

- Buổi 3:



# 3. Thu thập bộ dữ liệu huấn luyện

## a) Thu thập ảnh Hồ Gươm: 2 buổi

- Buổi 1:

- Thời gian: buổi chiều, 20/10/2023
- Thực hiện: Dũng + Biễn
- Nội dung:
  - Xung quanh hồ Gươm
  - Tháp Rùa, Cầu Thê Húc
- Kết quả: ~210 bức ảnh (thô)

# 3. Thu thập bộ dữ liệu huấn luyện

## a) Thu thập ảnh Hồ Gươm: 2 buổi

- Buổi 1:



# 3. Thu thập bộ dữ liệu huấn luyện

## a) Thu thập ảnh Hồ Gươm: 2 buổi

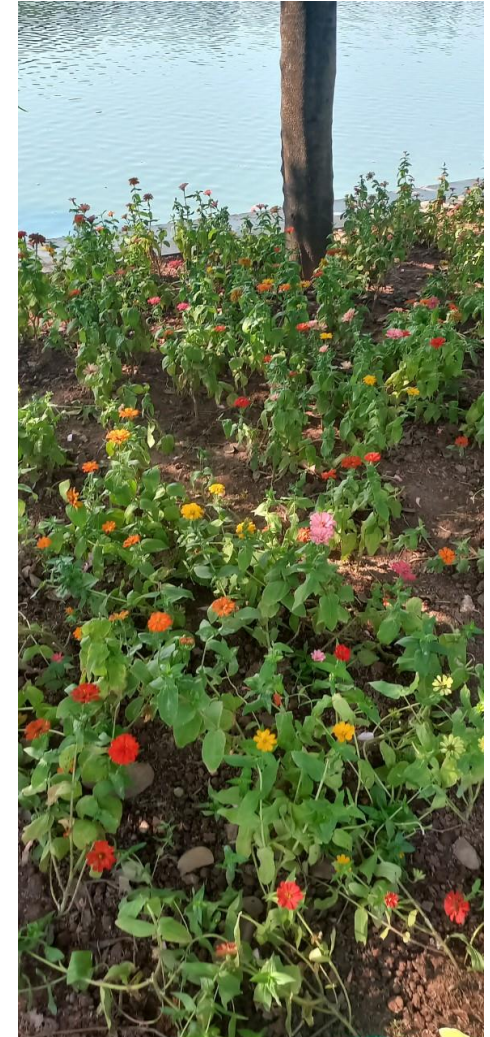
- Buổi 2:
  - Thời gian: buổi sáng, 27/10/2023
  - Thực hiện: Dũng + Biên
  - Nội dung:
    - Xung quanh hồ Gươm, bồn hoa
    - Bưu điện Hà Nội
    - Vườn hoa
  - Kết quả: ~120 bức ảnh (thô)



# 3. Thu thập bộ dữ liệu huấn luyện

## a) Thu thập ảnh Hồ Gươm: 2 buổi

- Buổi 2:



# 4. Tiền xử lý dữ liệu

## a) Lọc ảnh

- Loại bỏ các ảnh:
  - Không thể gán nhãn. (Vd: ảnh chụp thùng rác)
  - Trùng lặp hoặc được chụp liên tiếp.
  - Chủ thể bị chắn. (Vd: chụp vườn hoa nhưng có người đứng chắn)
  - Nhiều quá lớn. (Vd: chụp mặt hồ nhưng quá nhiều người lọt vào khung hình)
  - Không phù hợp với yêu cầu. (Vd: Đề bài yêu cầu chụp bưu điện Hà Nội nhưng ảnh là Viettel post).

## b) Cắt ảnh

- Cắt bỏ phần rìa các ảnh nếu chứa nhiều, vd: người, xe cộ,...

## 4. Tiền xử lý dữ liệu

### c) **Đảm bảo phân phối nhãn:**

- Yêu cầu nhóm chụp ảnh chụp thêm ảnh những nhãn có số lượng thấp (bưu điện, vườn hoa, khách sạn), xử lý thủ công phần ảnh bổ sung.
- Cân bằng tỉ lệ phân phối giữa các nhãn trong tập dữ liệu.

### d) **Chọn lọc và resize:**

- Chọn đúng 200 ảnh và nén ảnh về kích thước như yêu cầu của project (file 1-200.jpg, <20kB)

## 4. Tiền xử lý dữ liệu

### e) Gán nhãn:

- Tạo sheet gán nhãn từng ảnh cho 12 thành viên.
- Tạo sheet tổng hợp nhãn tự động từ 12 sheet thành viên (gán nhãn = 1 nếu ít nhất 6 thành viên gán nhãn đó, ngược lại = 0).



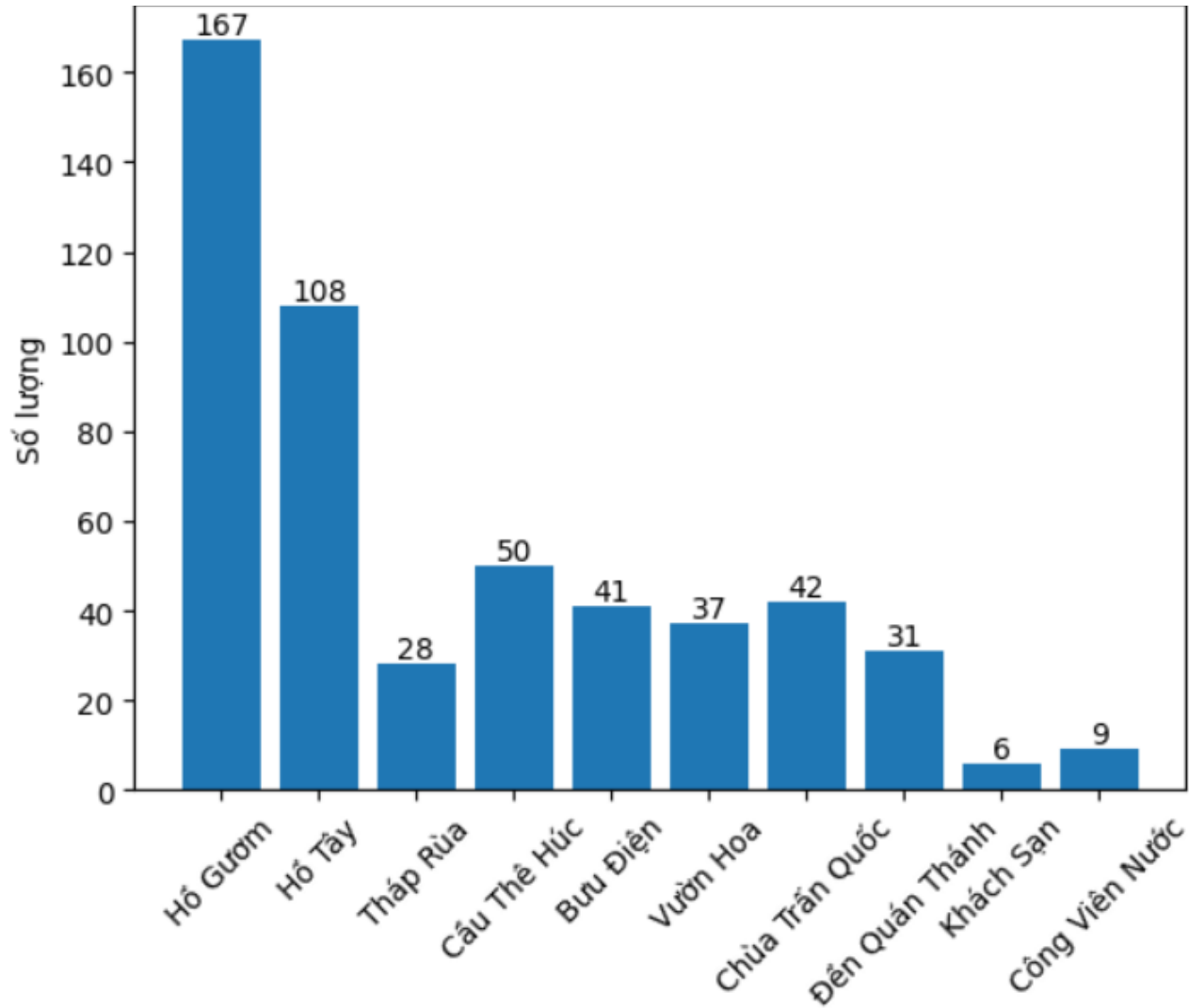
## 4. Tiền xử lý dữ liệu

### Kết quả:

Số lượng ảnh	Hồ Gươm	Hồ Tây
Thô	~350 ảnh	~450 ảnh
Đã xử lý	~170 ảnh	~180 ảnh
Kết quả nộp GV	~100 ảnh	~100 ảnh

# 4. Tiền xử lý dữ liệu

Phân phối nhãn của dataset



## 4. Tiền xử lý dữ liệu

Tập dữ liệu sử dụng:

- Số lượng ảnh train: 1088
- Số lượng ảnh validation: 115
- Số lượng ảnh test: 200

# 5. Huấn luyện mô hình

Thử nghiệm transfer-learning > 30 kiến trúc mô hình học sâu có kết quả tốt trên ImageNet.

-> **Kết quả:** Chọn mô hình **NFNet** (Google, 2021)

- 54M tham số
- Thành công trong nhiều nhiệm vụ/tập dữ liệu các bài toán xử lý ảnh
- Xử lý được nhiều tác vụ trong xử lý ảnh

# 5. Huấn luyện mô hình

Training dataset



NFNet  
(Pretrained in  
ImageNet)



Evaluation  
&  
Save model

Testing dataset



NFNet  
(Best F1)



Inference  
Result

# 5. Huấn luyện mô hình

Mô hình **NFNet** có các đặc trưng:

- Không sử dụng batch-norm, giúp đơn giản model và giảm bộ nhớ.
- Tốc độ đào tạo nhanh ~8 lần so với các model cùng tham số
- Sử dụng Adaptive Gradient Clipping
- Đạt kết quả tốt khi sử dụng transfer-learning

Các đặc trưng trên của mô hình **NFNet** phù hợp với bài toán này vì:

- Kết quả tốt khi sử dụng transfer-learning của model
- Bộ dữ liệu nhỏ (batch-size ảnh hưởng lớn)
- Adaptive Gradient Clipping giúp mô hình hội tụ dễ dàng
- Tốc độ đào tạo và kiểm thử nhanh

# 5. Huấn luyện mô hình

Batch-norm:

- Thường sử dụng trong các mô hình học sâu, chuẩn hóa dữ liệu input của các lớp, giúp mô hình dễ dàng hội tụ, tránh mất mát hay bùng nổ gradient.
- Tuy nhiên, nó có nhược điểm khi chiếm nhiều tài nguyên tính toán, gây ra những khác biệt khi đào tạo và kiểm thử, cũng như đánh mất sự độc lập giữa các mẫu.
- Với bộ dữ liệu nhỏ, BN thường ít hiệu quả, và dễ dẫn tới overfit.

Adaptive Gradient Clipping:

- Một kỹ thuật sử dụng để scale gradient nhằm tránh bùng nổ dữ liệu trong quá trình huấn luyện.

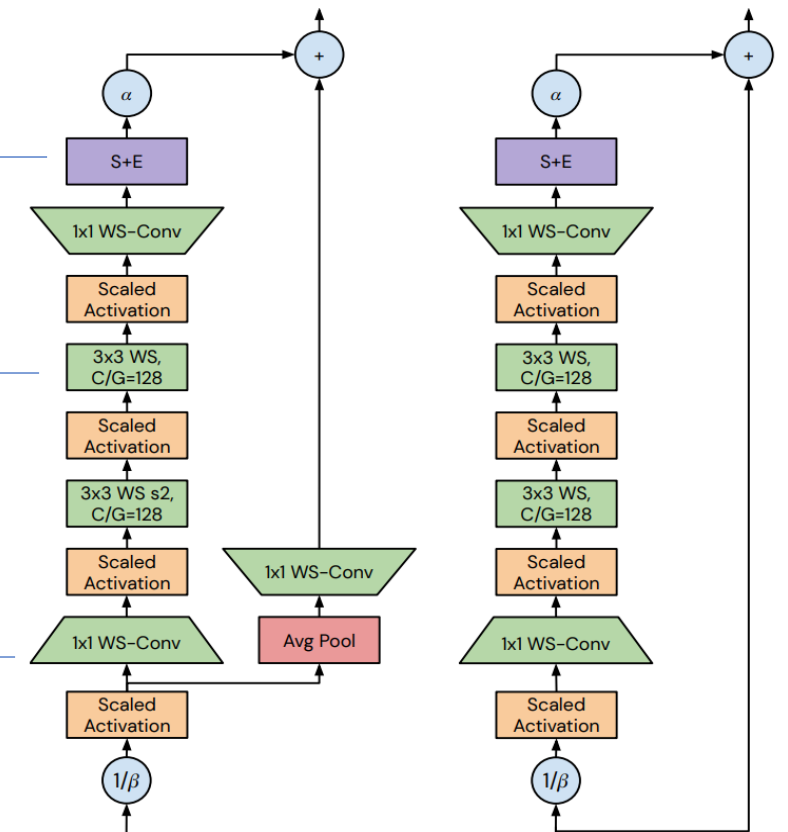
# 5. Huấn luyện mô hình

Mô hình **NFNet** gồm 4 stage, số lượng block mỗi stage là 1, 2, 6, 3.  
Mỗi stage bắt đầu bởi 1 transition block, theo sau là các residual block.

Squeeze and Excite Layer gán trọng số quan trọng cho từng kênh

Các lớp DepthWise Separable CNN được cài đặt với Scaled Weight Standardization chuẩn hóa trọng số theo chiều kênh.

Transition Block sử dụng để giảm kích thước feature map; các residual block sử dụng để trích xuất feature



a) Transition block

b) Residual Block



# 5. Huấn luyện mô hình

## Quá trình huấn luyện

- Loss: Binary Cross Entropy Loss

$$-\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^L [y_{ij} \cdot \log(\hat{y}_{ij}) + (1 - y_{ij}) \cdot \log(1 - \hat{y}_{ij})]$$

In this formula:

- $N$  is the number of instances.
- $L$  is the number of labels.
- $y_{ij}$  is the true label for instance  $i$  and label  $j$ .
- $\hat{y}_{ij}$  is the predicted probability for instance  $i$  and label  $j$ .

# 5. Huấn luyện mô hình

## Quá trình huấn luyện

- Tỷ lệ train/valid: 0.9 – 0.1
- Data Augmentation: RandomResizedCrop, RandomAugment...
- Optimizer: AdamW
- Pretrain Data: ImageNet
- Hyperparameters:
  - Learning rate:  $3e-4$
  - Scheduler: Cosine Annealing
  - Weight decay:  $1e-4$
  - Batch size: 16
  - Epochs: 200



## 5. Huấn luyện mô hình

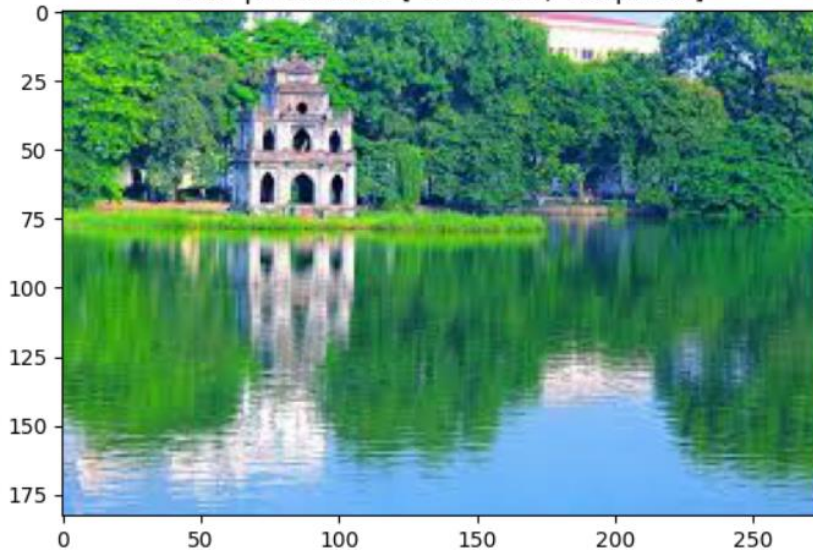
**Kết quả trên tập validation:**

Metrics	Value
Accuracy	0.9875
Precision	0.9749
Recall	0.9593
F1-Score	0.9670

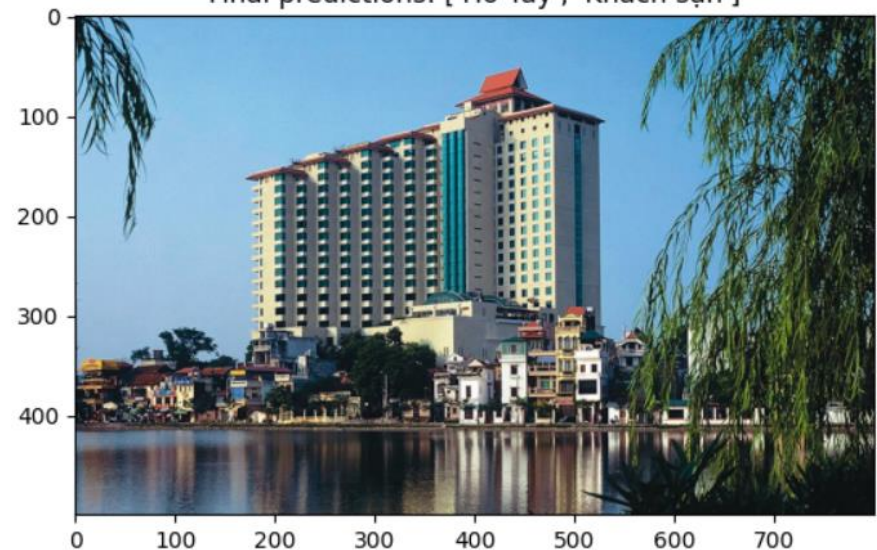
# 5. Huấn luyện mô hình

## Một số dự đoán trên tập test GV cung cấp:

NFNet's predictions: ['Hồ Gươm', 'Tháp Rùa']  
Final predictions: ['Hồ Gươm', 'Tháp Rùa']



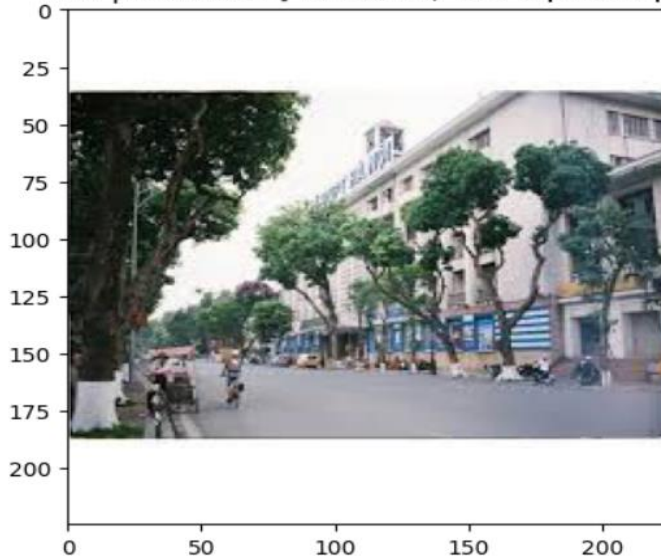
NFNet's predictions: ['Hồ Tây', 'Khách sạn']  
Final predictions: ['Hồ Tây', 'Khách sạn']



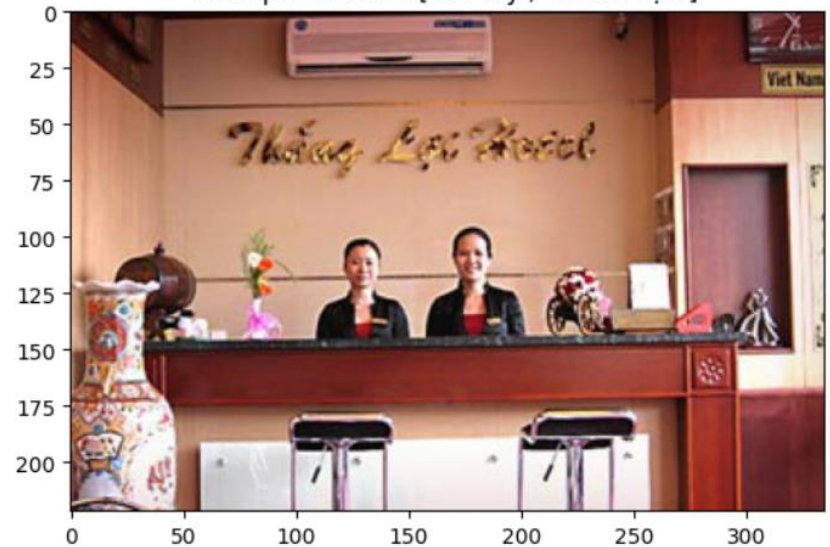
# 5. Huấn luyện mô hình

## Một số dự đoán trên tập test GV cung cấp:

NFNet's predictions: ['Hồ Gươm', 'Hồ Tây', 'Bưu điện Hà Nội']  
Final predictions: ['Hồ Gươm', 'Bưu điện Hà Nội']



NFNet's predictions: ['Hồ Gươm', 'Khách sạn']  
Final predictions: ['Hồ Tây', 'Khách sạn']





# 5. Huấn luyện mô hình

## Một số dự đoán trên tập test GV cung cấp:

