

VIET NAM NATIONAL UNIVERSITY HO CHI MINH
UNIVERSITY OF INFORMATION TECHNOLOGY



IMAGE RETRIEVAL

Subject: Computer Vision in Human-Computer Interaction
Class: CS532.M21.KHCL
Lecturer: Do Van Tien

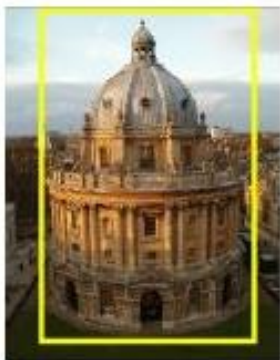
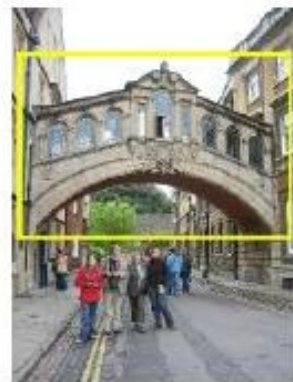
Students:

Thai Tran Khanh Nguyen	19520188
Doan Nguyen Nhat Quang	19520235
Nguyen Pham Vinh Nguyen	19520186
Nguyen Khanh Nhu	19520209

Nội dung

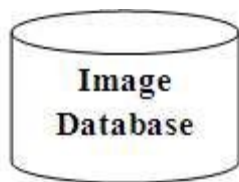
1. Giới thiệu bài toán
2. Khảo sát và phân tích
3. Xây dựng phương pháp
 - Simple Image Retrieval (SIR) – (our)
 - CNN Image Retrieval with No Human Annotation (CNN-IRwNHA)
 - Deep Local Feature (DELF)
4. Đánh giá kết quả
5. Xây dựng demo
 - Thiết kế kiến trúc
 - Thiết kế giao diện
 - Vietnam Tourism Dataset
 - Restful API
 - Kiểm thử
6. Kết luận và hướng phát triển

Giới thiệu bài toán



Giới thiệu bài toán

INPUT



OUTPUT



Khảo sát & phân tích

Type	Method	Backbone DCNN	Output Layer	Embedding Aggregation	Feature Dimension	Loss Function	Holidays	UKB	Oxford5k (+100k)	Paris6k (+100k)	Brief Conclusions and Highlights
Supervised Fine-tuning	DELF [5]	ResNet-101	Conv4 Block	Attention + PCA _w	CE Loss	2048	—	—	83.8 (82.6)	85.0 (81.7)	Exploring the FCN to extract region-level features and construct feature pyramids of different sizes.
	Neural codes [40]	AlexNet	FC6	PCA	CE Loss	128	78.9	3.29 (N-S)	55.7 (52.3)	—	The first work which fine-tunes deep networks for image retrieval. Compressed neural codes and different layers are explored.
	Nonmetric [41]	VGG16	Conv5	PCA _w	Regression Loss	512	—	—	88.2 (82.1)	88.2 (82.9)	Visual similarity learning of similar and dissimilar pairs is performed by a neural network, optimized using regression loss.
	Faster R-CNN [96]	VGG16	Conv5	MP / SP	Regression Loss	512	—	—	75.1 (—)	80.7 (—)	RPN is fine-tuned, based on bounding box coordinates and class scores for specific region query which is region-targeted.
	SIAM-FV [42]	VGG16	Conv5	FV + PCA _w	Siamese Loss	512	—	—	81.5 (76.6)	82.4 (—)	Fisher Vector is integrated on top of VGG and is trained with VGG simultaneously.
	SIFT-CNN [127]	VGG16	Conv5	SP	Siamese Loss	512	88.4	3.91 (N-S)	—	—	SIFT features are used as supervisory information for mining positive and negative samples.
	Quartet-Net [129]	VGG16	FC6	PCA	Siamese Loss	128	71.2	87.5 (mAP)	48.5 (—)	48.8 (—)	Quartet-net learning is explored to improve feature discrimination where double-margin contrastive loss is used.
	NetVLAD [44]	VGG16	VLAD Layer	PCA _w	Triplet Loss	256	79.9	—	62.5 (—)	72.0 (—)	VLAD is integrated at the last convolutional layer of VGG16 network as a plugged layer.
	Deep Retrieval [87]	ResNet-101	Conv5 Block	MP + PCA _w	Triplet Loss	2048	90.3	—	86.1 (82.8)	94.5 (90.6)	Dataset is cleaned automatically. Features are encoded by R-MAC. RPN is used to extract the most relevant regions.
Unsupervised Fine-tuning	MoM [133]	VGG16	Conv5	MP + PCA _w	Siamese Loss	64	87.5	—	78.2 (72.6)	85.1 (78.0)	Exploring manifold learning for mining dis/similar samples. Features are tested globally and regionally.
	GeM [47]	VGG16	Conv5	GeM Pooling	Siamese Loss	512	83.1	—	82.0 (76.9)	79.7 (72.6)	Fine-tuning CNNs on an unordered dataset. Samples are selected from an automated 3D reconstruction system.
	SfM-CNN [45]	VGG16	Conv5	PCA _w	Siamese Loss	512	82.5	—	77.0 (69.2)	83.8 (76.4)	Employing Structure-from-Motion to select positive and negative samples from unordered images.
	IME-CNN [46]	ResNet-101	IME Layer	MP	Regression Loss	2048	—	—	92.0 (87.2)	96.6 (93.3)	Graph-based manifold learning is explored within an IME layer to mine the matching and non-matching pairs in unordered datasets.
	MDP-CNN [137]	ResNet-101	Conv5 Block	SP	Triplet Loss	2048	—	—	85.4 (85.1)	96.3 (94.7)	Exploring global feature structure by modeling the manifold learning to select positive and negative pairs.

Khảo sát & phân tích

Image Retrieval | Papers With Code

←

↺

https://paperswithcode.com/task/image-retrieval

🔍

🏠

📁

🔖

🌐

🔗

👤

...

—

🖨

✖

📊

🔍 Search

🏠

📁

+

🔗

Browse State-of-the-Art

Datasets

Methods

More ▾

👤

Sign In

🖨

Edit

📊

Computer Vision

Image Retrieval

413 papers with code • 26 benchmarks • 46 datasets

Image retrieval systems aim to find similar images to a query image among an image dataset.

(Image credit: DELF)

Benchmarks

These leaderboards are used to track progress in Image Retrieval

Trend	Dataset	Best Model	Paper	Code	Compare
	Flickr30K 1K test	🏆 X-VLM (base)			See all
	SOP	🏆 ROADMAP (DeiT-B)			See all
	Oxf5k	🏆 Offline Diffusion			See all
	Oxf105k	🏆 Offline Diffusion			See all
	Par6k	🏆 Offline Diffusion			See all
	Par106k	🏆 Offline Diffusion			See all
	CUB-200-2011	🏆 CGD (MG/SG)			See all
	CARS196	🏆 CGD (MG/SG)			See all
	In-Shop	🏆 CGD (SG/GS)			See all

Content

- 📖 Introduction
- 📊 Benchmarks
- 📁 Datasets
- 👤 Subtasks
- 📁 Libraries
- 📖 Papers
 - Most implemented
 - Social
 - Latest
 - No code

Khảo sát & phân tích

[Visual Geometry Group - University of Oxford](https://www.robots.ox.ac.uk/~vgg/research/oxbuildings/index.html)

←

↻

🔒

https://www.robots.ox.ac.uk/~vgg/research/oxbuildings/index.html

⌕

🔧

🌟

📁

🖼️

👤

⋮

—

📄

✕

🏠

📄

+

HOME

PEOPLE

RESEARCH

PUBLICATIONS

DEMOS

DATA

SOFTWARE

PRACTICALS

PROJECTS

JOB

INTERNAL

🐦

📄

📍

Web-scale particular object search

James Philbin, Ondrej Chum, Josef Sivic, Michael Isard, Ernesto Coto and Andrew Zisserman

Overview

The objective of our research is to develop efficient methods for searching for specific objects in extremely large datasets of images. In this work we focus on searching large collections of downloaded Flickr images for popular city landmarks.



Find these landmarks



...In these images

Motivation

The rise of the web and sites such as Flickr and YouTube have vastly increased the amount of visual media available online, but current methods for searching this material are inadequate, relying on associated texts or user provided annotation. We believe that the ability to query by visual example would be an important first step towards unlocking the information contained in images and would revolutionize the way people use sites such as Flickr in the same way that Google revolutionized the web.

Challenges

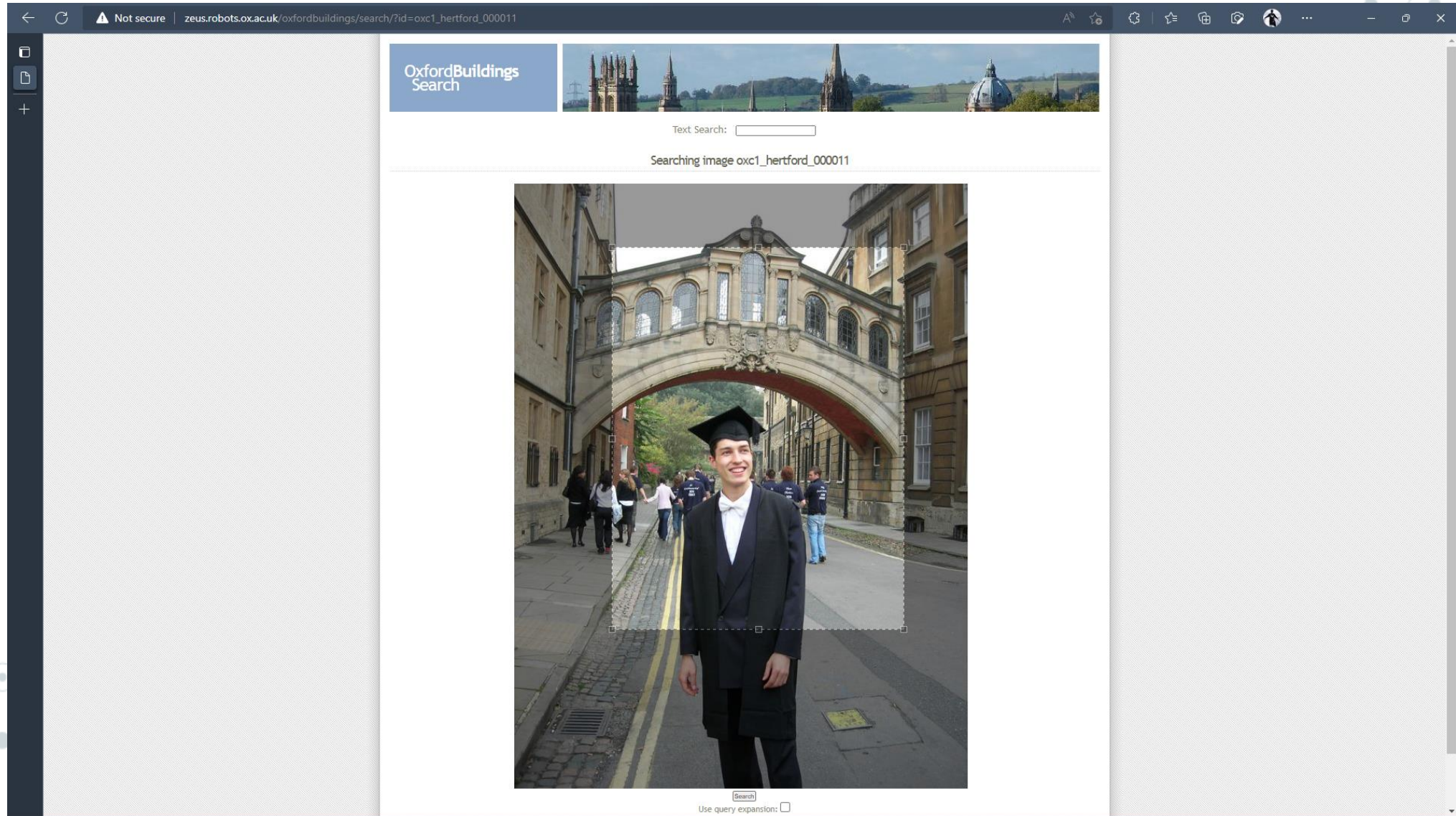
We need to be able to find objects in images despite a large variety of possible visual variations such as: scale, viewpoint, lighting and occlusions. Additionally, any methods used need to scale to databases containing millions of images, while still returning query results in less than a second.

This website uses Google Analytics to help us improve the website content. This requires the use of standard Google Analytics cookies, as well as a cookie to record your response to this confirmation request. If this is OK with you, please click 'Accept cookies', otherwise you will see this notice on every page. For more information, please [click here](#)

ACCEPT COOKIES

Khảo sát & phân tích

Oxford Buildings Search



Khảo sát & phân tích

Image Retrieval Python* Demo - OpenVINO™ Toolkit

The screenshot shows the OpenVINO documentation website. The browser address bar displays the URL: https://docs.openvino.ai/2021.1/omz_demos_python_demos_image_retrieval_demo_README.html. The page header includes the OpenVINO logo, navigation links (GET STARTED, DOCUMENTATION, RESOURCES, 中文文件), and a search icon. A sidebar on the left lists various demos, with 'Image Retrieval Python* Demo' selected. The main content area is titled 'Demo Output' and includes a link to find videos. It features a large image of a patterned fabric with a bounding box, a 'Pattern:' section showing a smaller version of the pattern, and a grid of 10 retrieved images with their respective inference times. A footer at the bottom contains a cookie consent message and an 'ACCEPT' button.

OpenVINO™
Documentation

GET STARTED DOCUMENTATION RESOURCES 中文文件

Version 2021.1 Download Docs Install OpenVINO™

Examples of videos can be found [here](#).

Demo Output

The application uses OpenCV to display gallery searching result and current inference performance.

Position: 0
Gallery size: 38
Embedding (ms): 6.1
Gallery search (ms): 2.0
Inp. res: 224x224

Pattern:

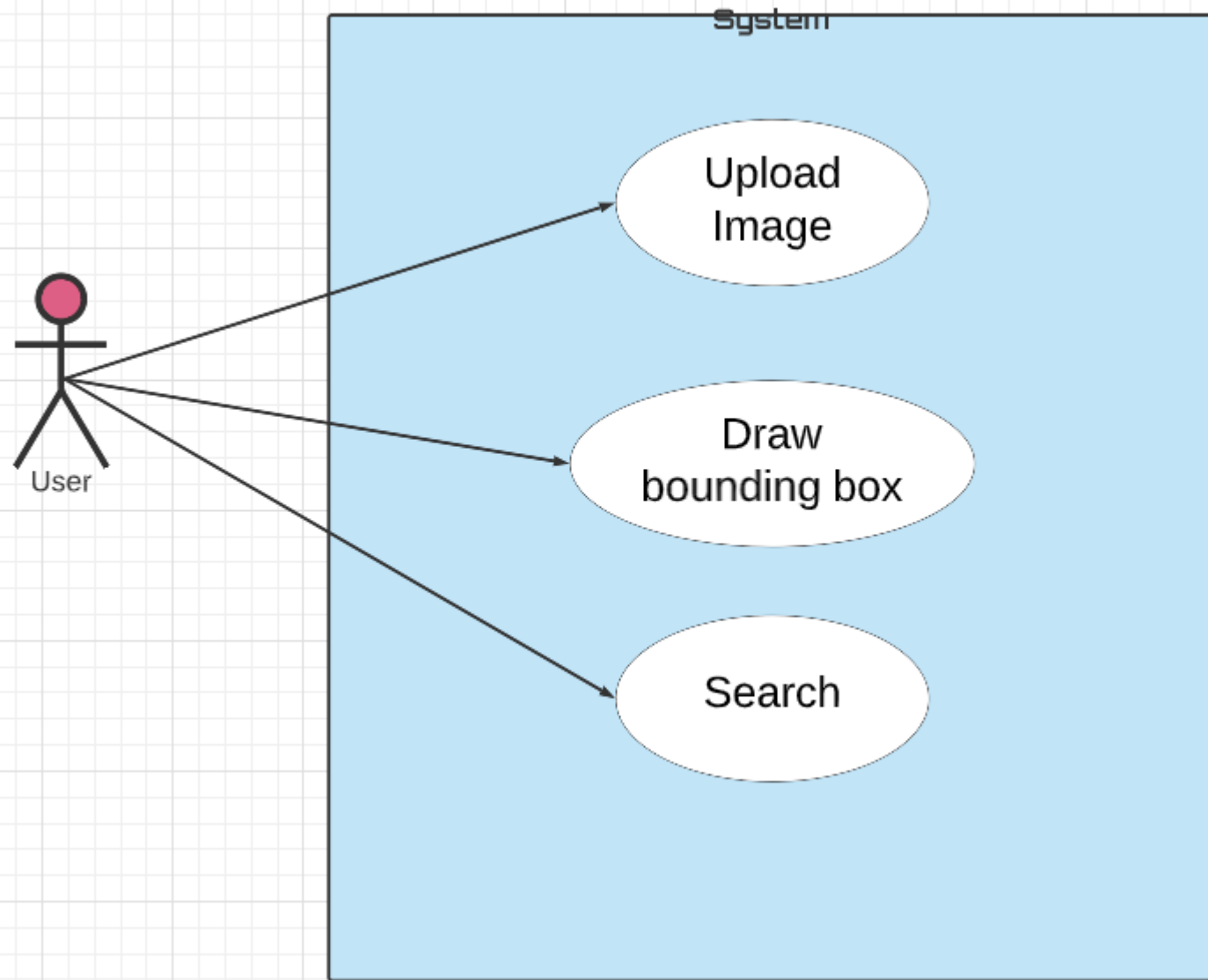
0:0.11 1:0.85 2:0.85 3:0.86 4:0.87
5:0.89 6:0.91 7:0.93 8:0.93 9:0.94

In This Document
[How It Works](#)
[Running](#)
[Demo Output](#)
[See Also](#)

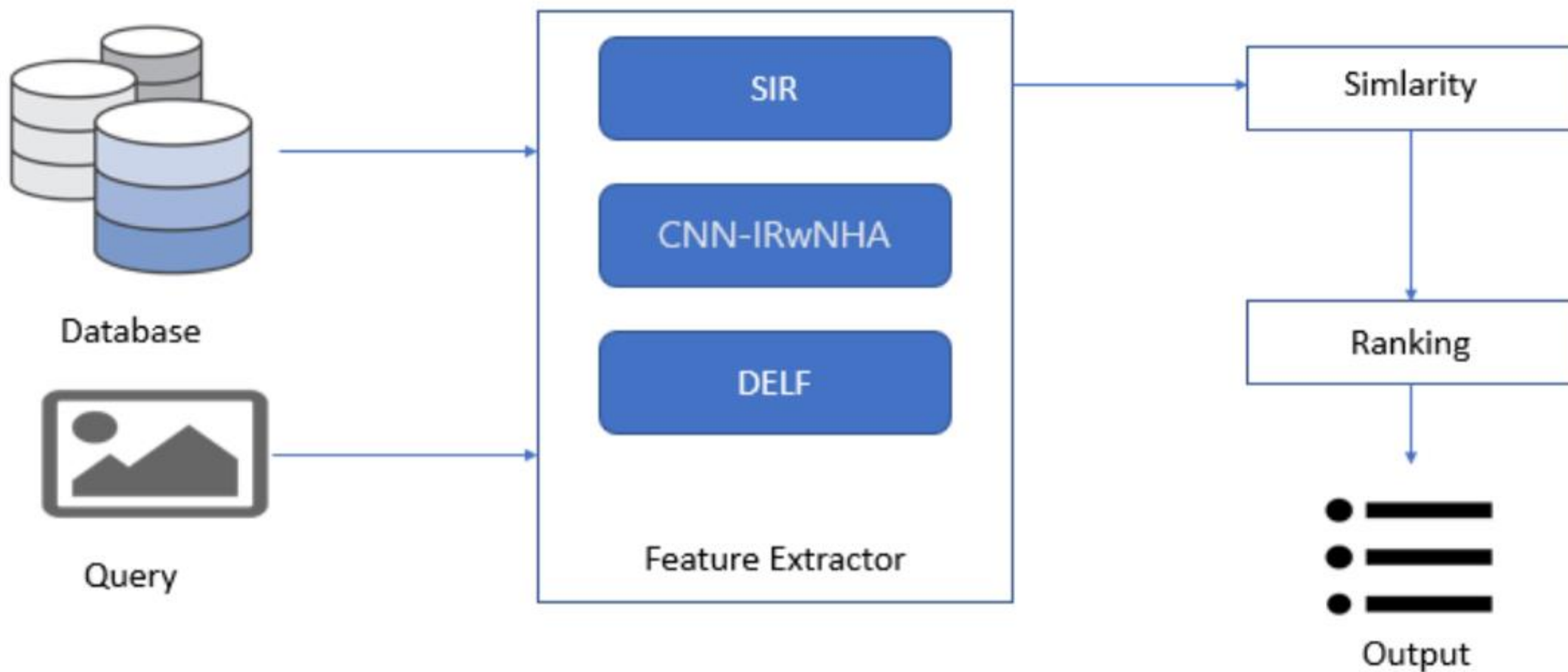
We use cookies to provide you with the best experience, and for measurement and analytics purposes. By using our website, you agree to our use of cookies as described in our [Cookie Policy](#). [ACCEPT](#)

Khảo sát & phân tích

Yêu cầu người dùng

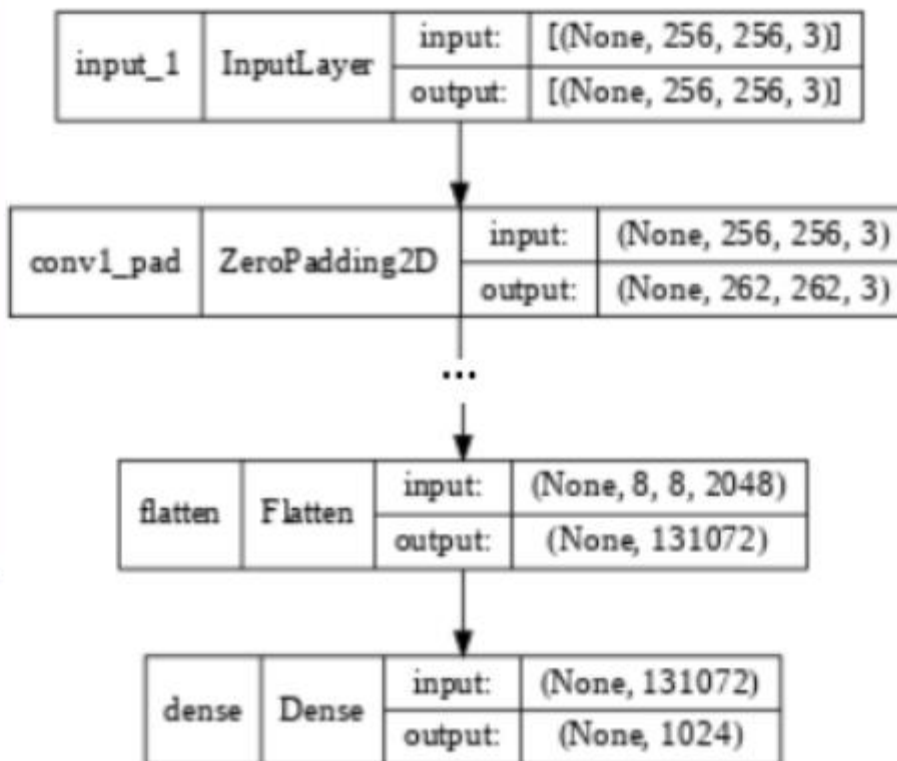


Xây dựng phương pháp: pipeline

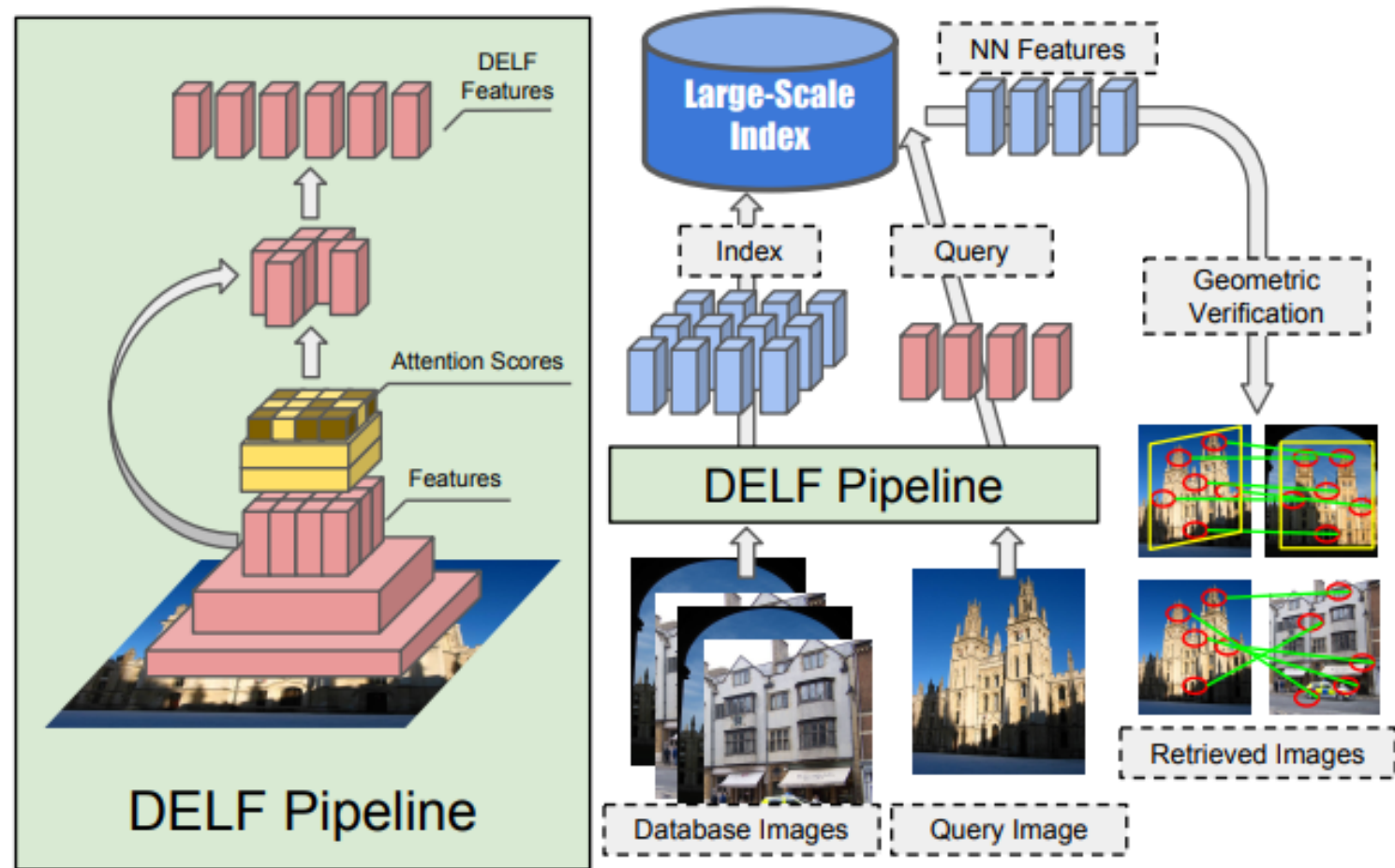


Simple Image Retrieval (SIR)

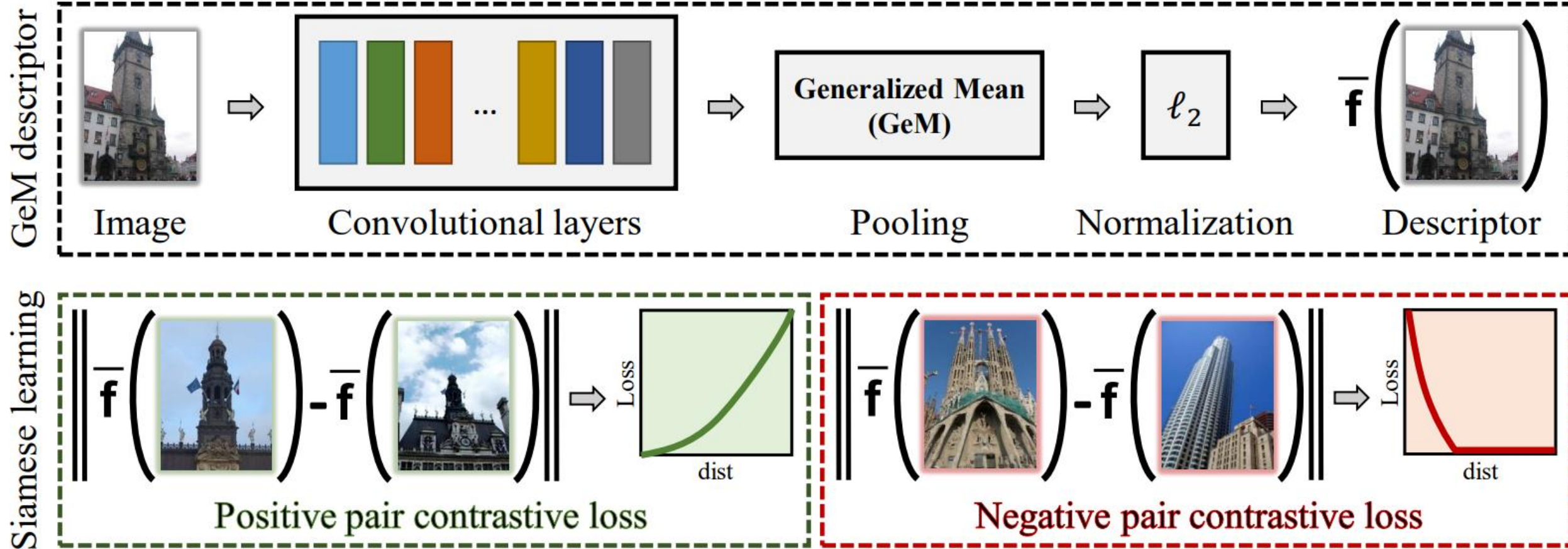
Resnet152



Deep Local Feature (DELF)



CNN Image Retrieval with No Human Annotation (CNN-IR_wNHA)



EVALUATION

Roxford 5k dataset ([Revisiting Oxford](#)) [4]:

Author revisits and address issues with Oxford 5k and Paris 6k image retrieval benchmarks. New annotation for both datasets is created with an extra attention to the reliability of the ground truth and three new protocols of varying difficulty are introduced. Author additionally introduces 15 new challenging queries per dataset and a new set of 1M hard distractors.


Method	map			map@5		
	map E	map M	map H	map E	map M	map H
<i>SIR</i>	14.4	11.85	2.34	32.86	27.71	24.86
<i>DELF</i>	58.56	42.04	16.98	84.29	73.86	66.57
<i>CNN-IR_wNHA</i>	85.08	68.65	44.24	92.86	92.29	86.

EVALUATION



Oxford 5k dataset:

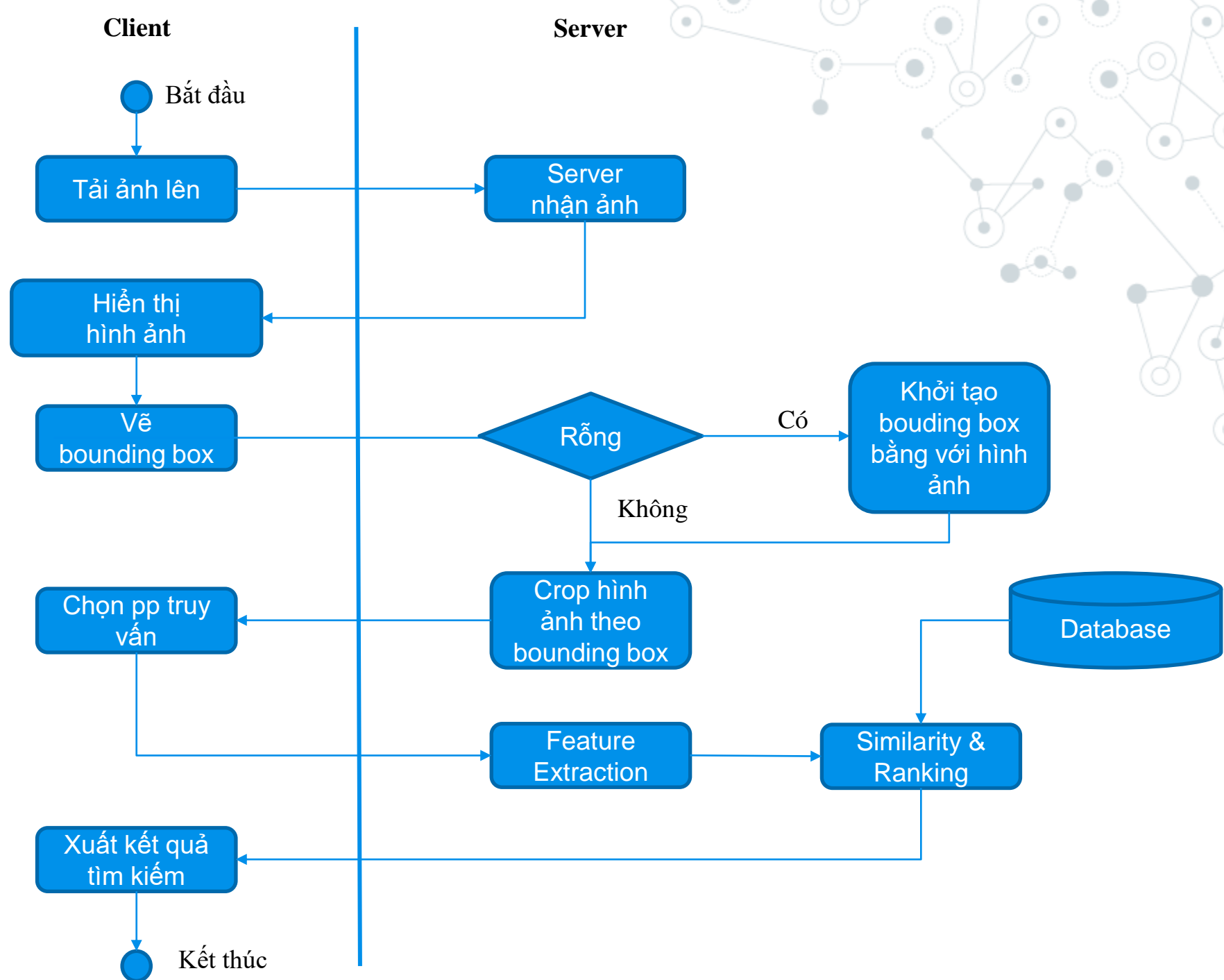
Method	map
SIR	23.34
DELF	66.69
CNN-IRwNHA	82.09



Vietnam Tourism Dataset

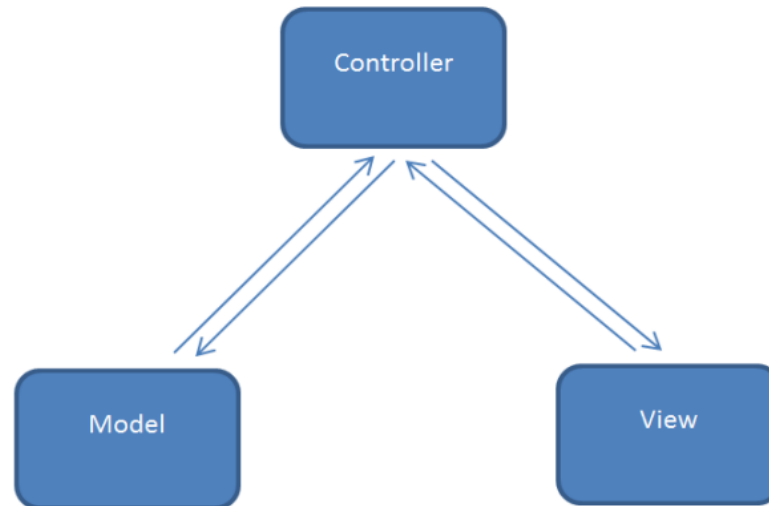


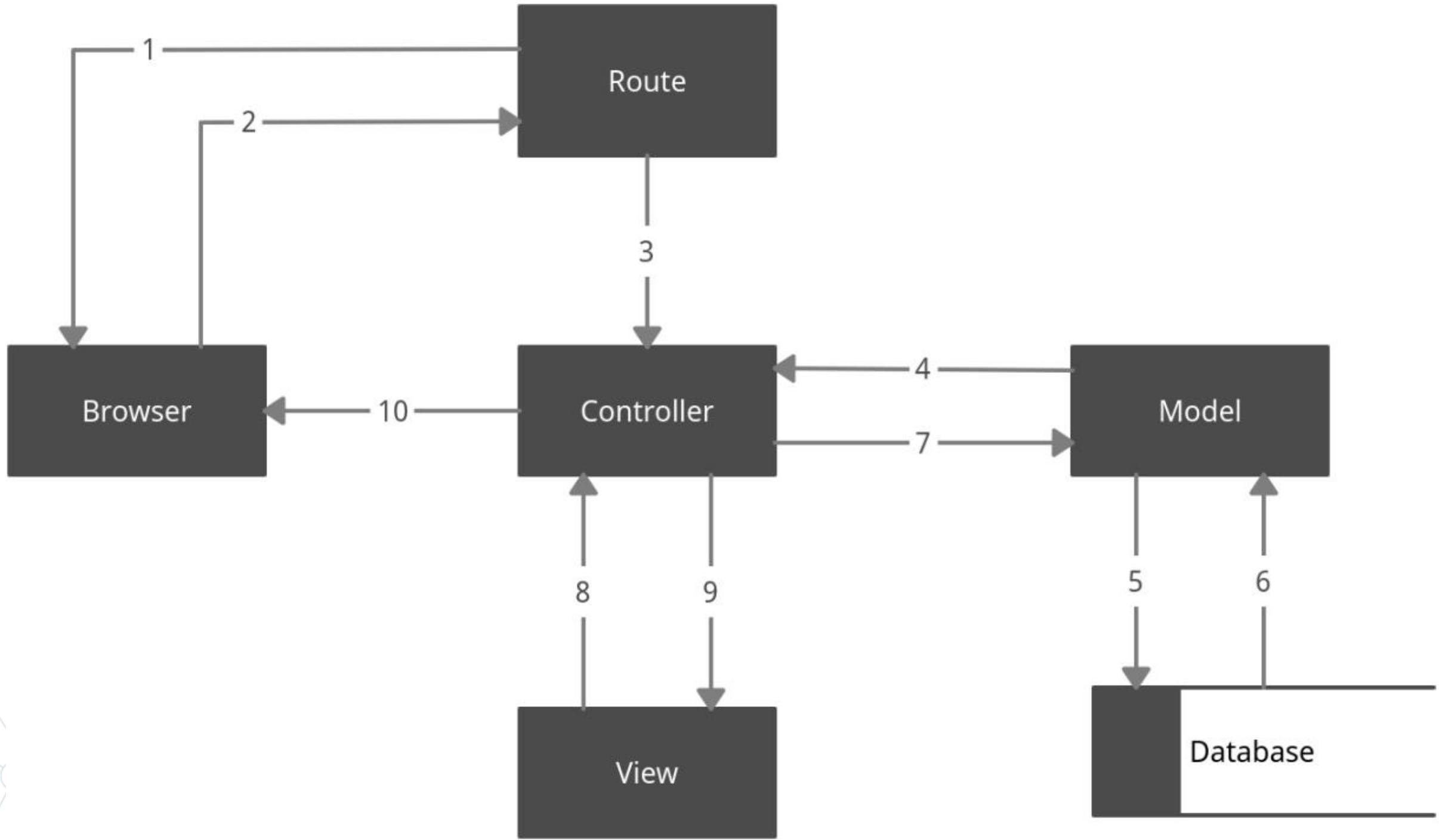
Activity Diagram



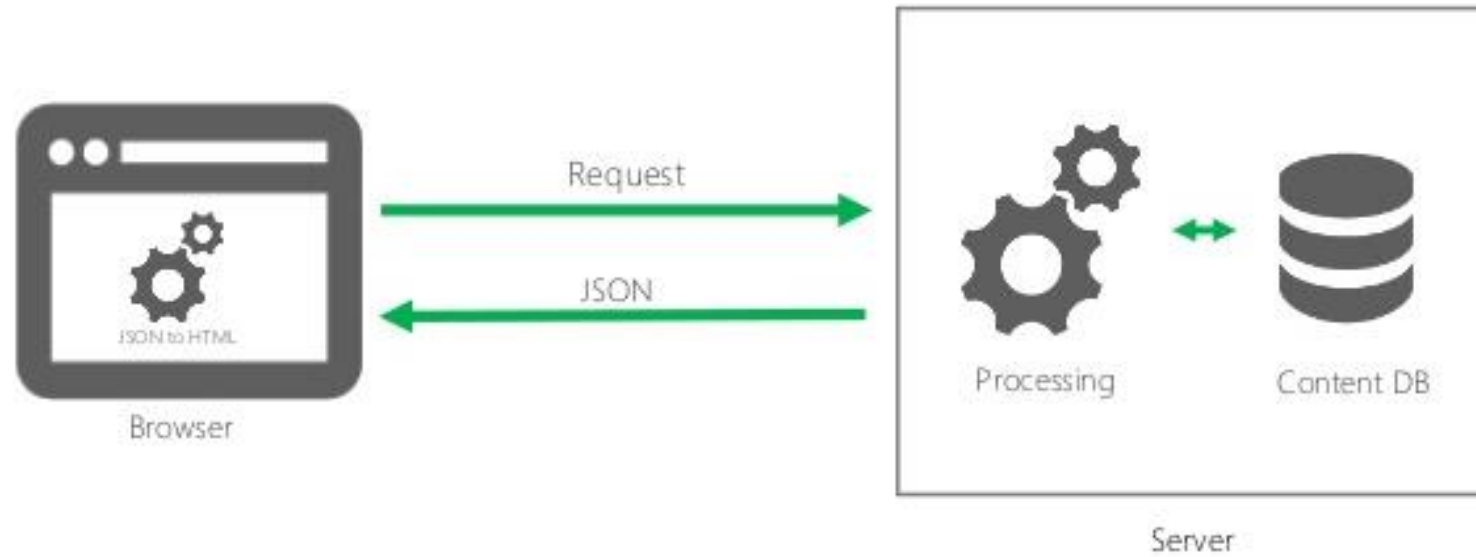
Kiến trúc

- Dễ dàng tiếp cận.
- Trình tự xử lý rõ ràng.
- Không phụ thuộc môi trường, nền tảng xây dựng hay ngôn ngữ lập trình phát triển.
- Mô hình đơn giản, dễ hiểu và dễ dàng triển khai với các dự án vừa hoặc nhỏ.
- Thuận lợi trong việc phát triển, quản lý, vận hành, bảo trì.
- Tạo được các chức năng chuyên biệt hóa đồng thời kiểm soát được luồng xử lý.

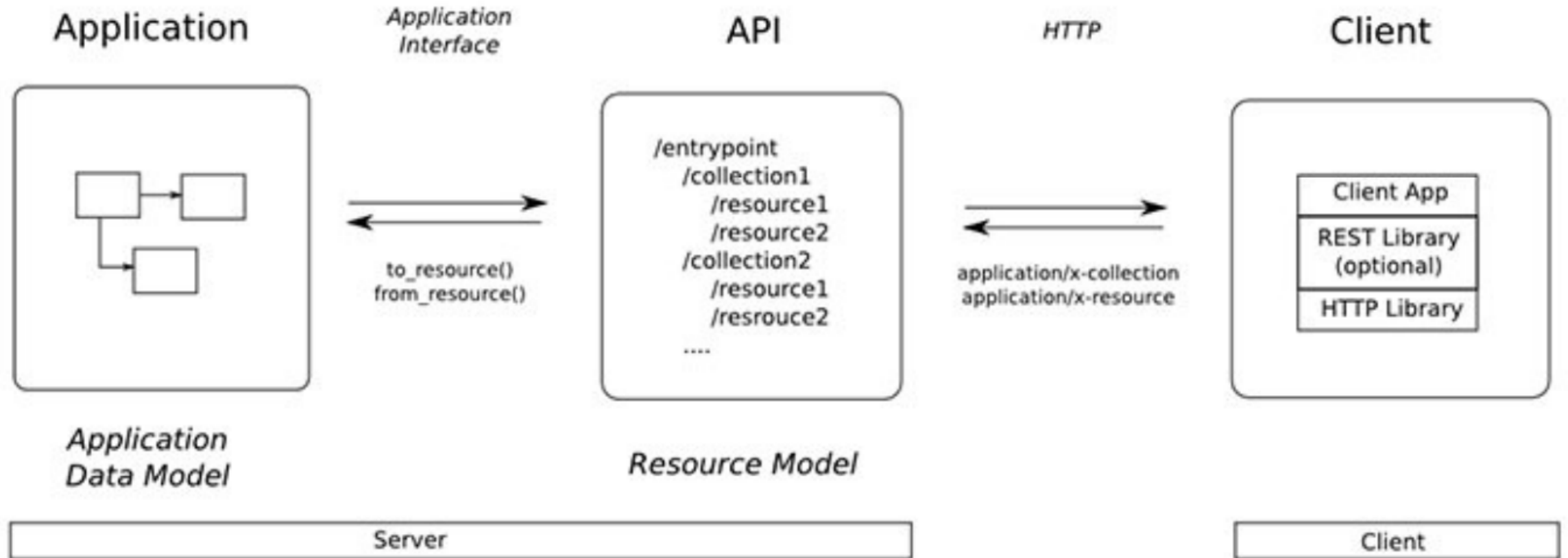




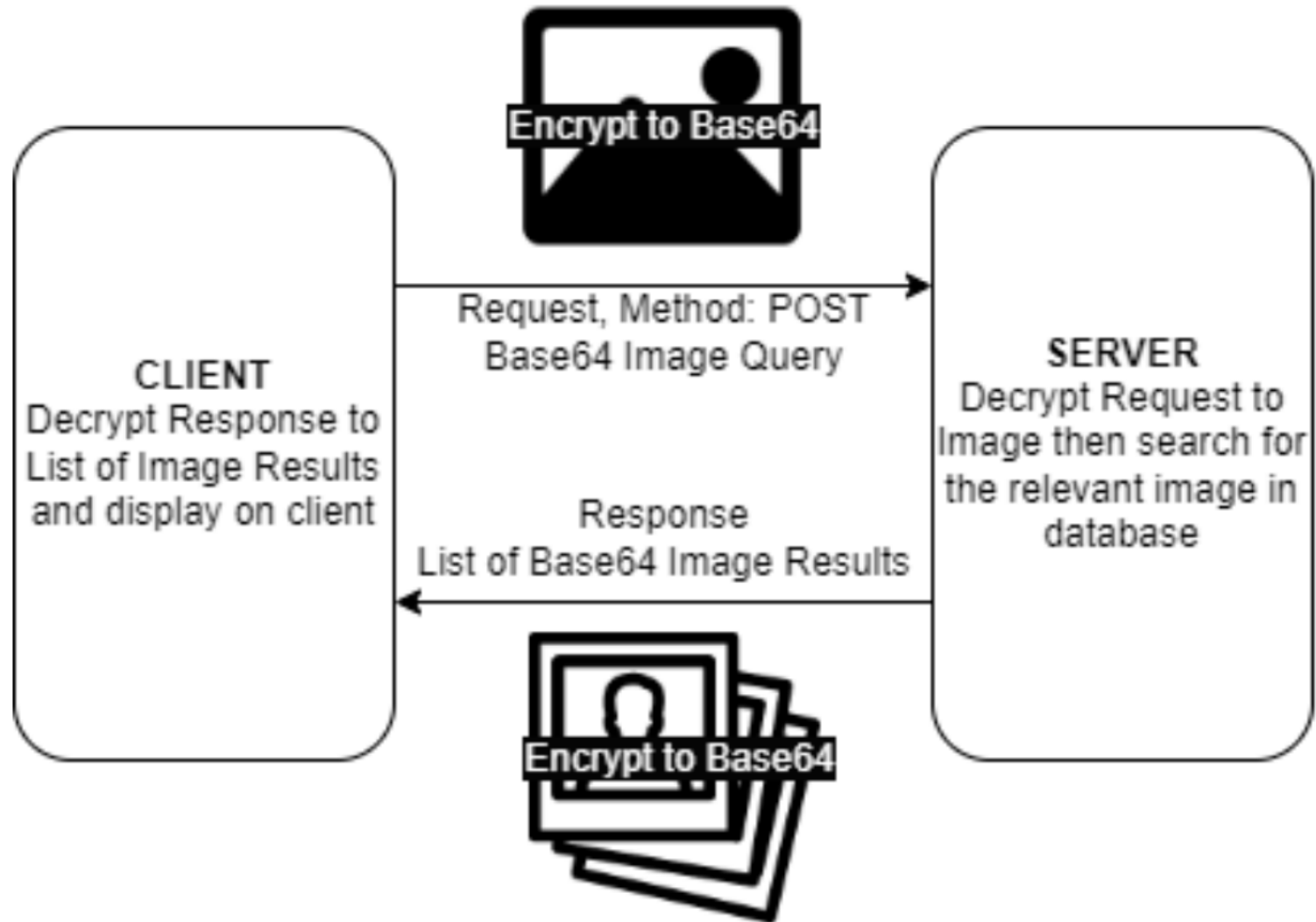
CLIENT-SIDE RENDERING



RESTFUL API



RESTFUL API



Thiết kế giao diện

SEARCH ENGINE

Image Search ☐ Text Search ☐

Chọn chế độ search
+ Image
+ Text (đang phát triển)

Upload hình ảnh

Select Image you want to retrieve

Choose File

Vẽ bounding box vật thể cần tìm kiếm



Chọn method retrieval

Select method for Image Search

☐ SIR ☒ CNN-IRwNHA

Submit

Truy vấn



Xuất kết quả



Conclusion and Demo

A decorative background featuring a network diagram. It consists of numerous nodes, represented by small circles, some of which are solid blue, some are grey with a blue outline, and others are plain grey. These nodes are interconnected by thin, light grey lines, forming a complex web-like structure that is more dense on the left and right sides of the slide.

Thanks for watching