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

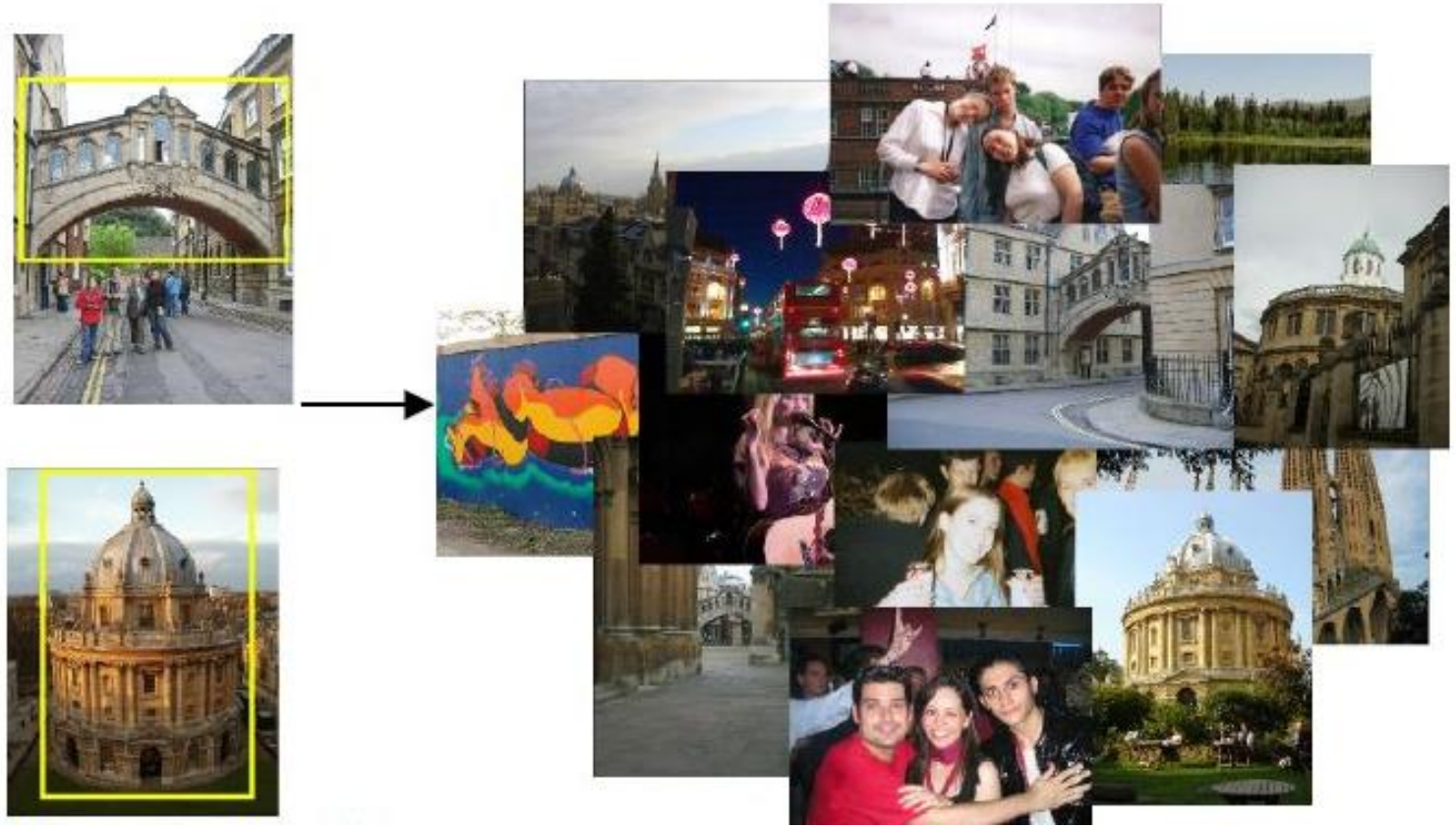
Students:

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

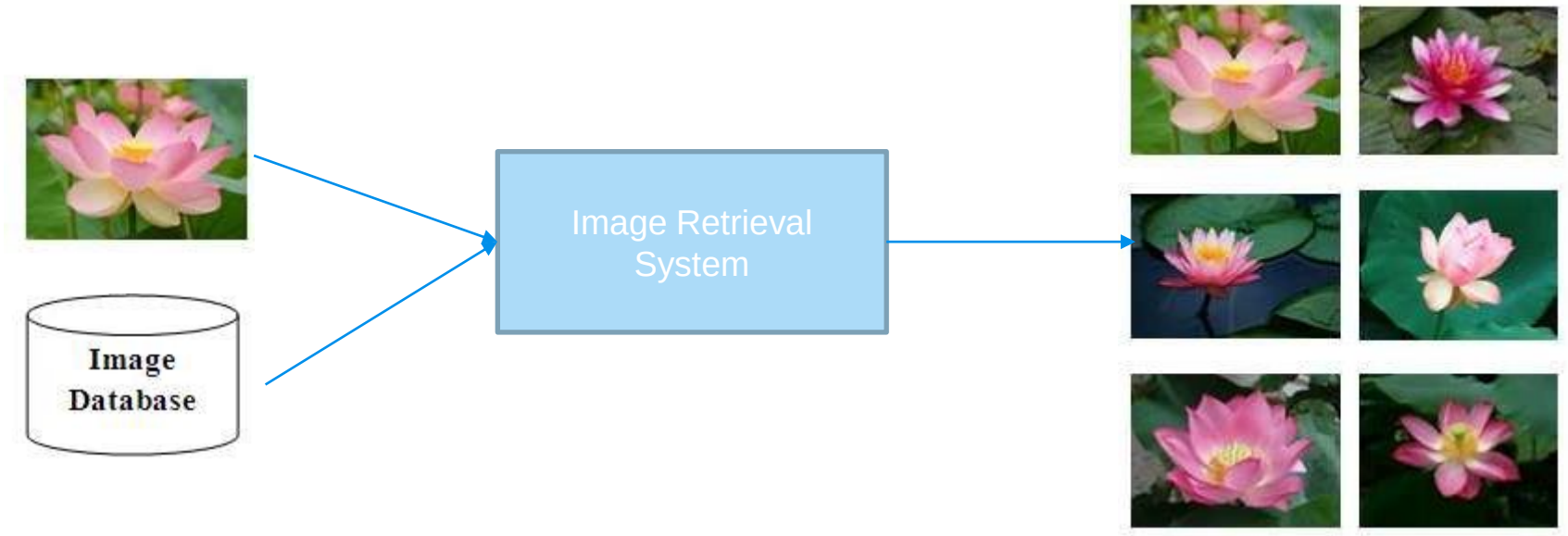
Giới thiệu bài toán



Giới thiệu bài toán

INPUT

OUTPUT



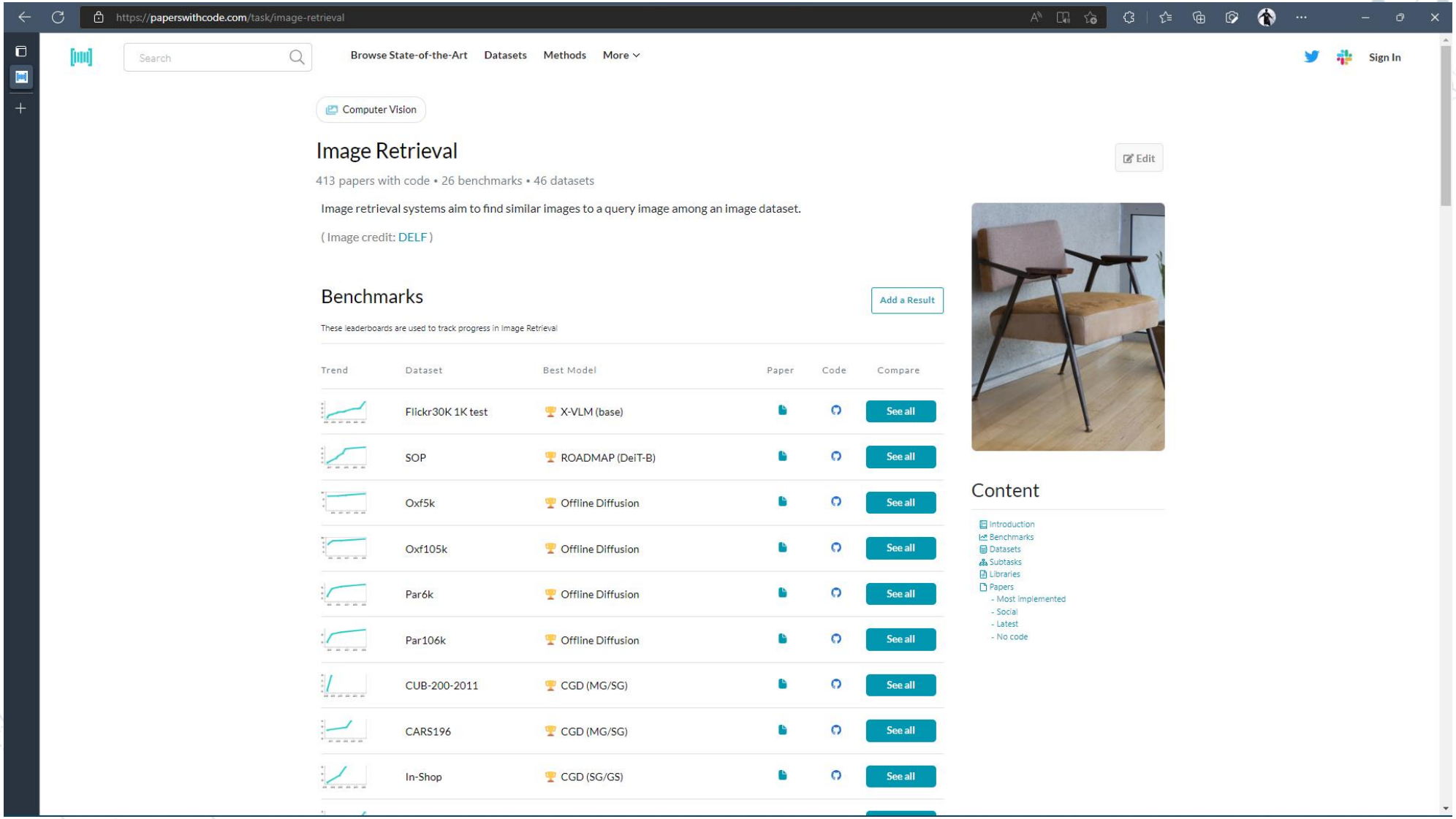
Khảo sát & phân tích

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.

Wei Chen, Yu Liu, Weiping Wang, Erwin M. Bakker, Theodoros Georgiou, Paul Fieguth, Li Liu, and Michael S. Lew, “Deep Learning for Instance Retrieval: A Survey”, 2022

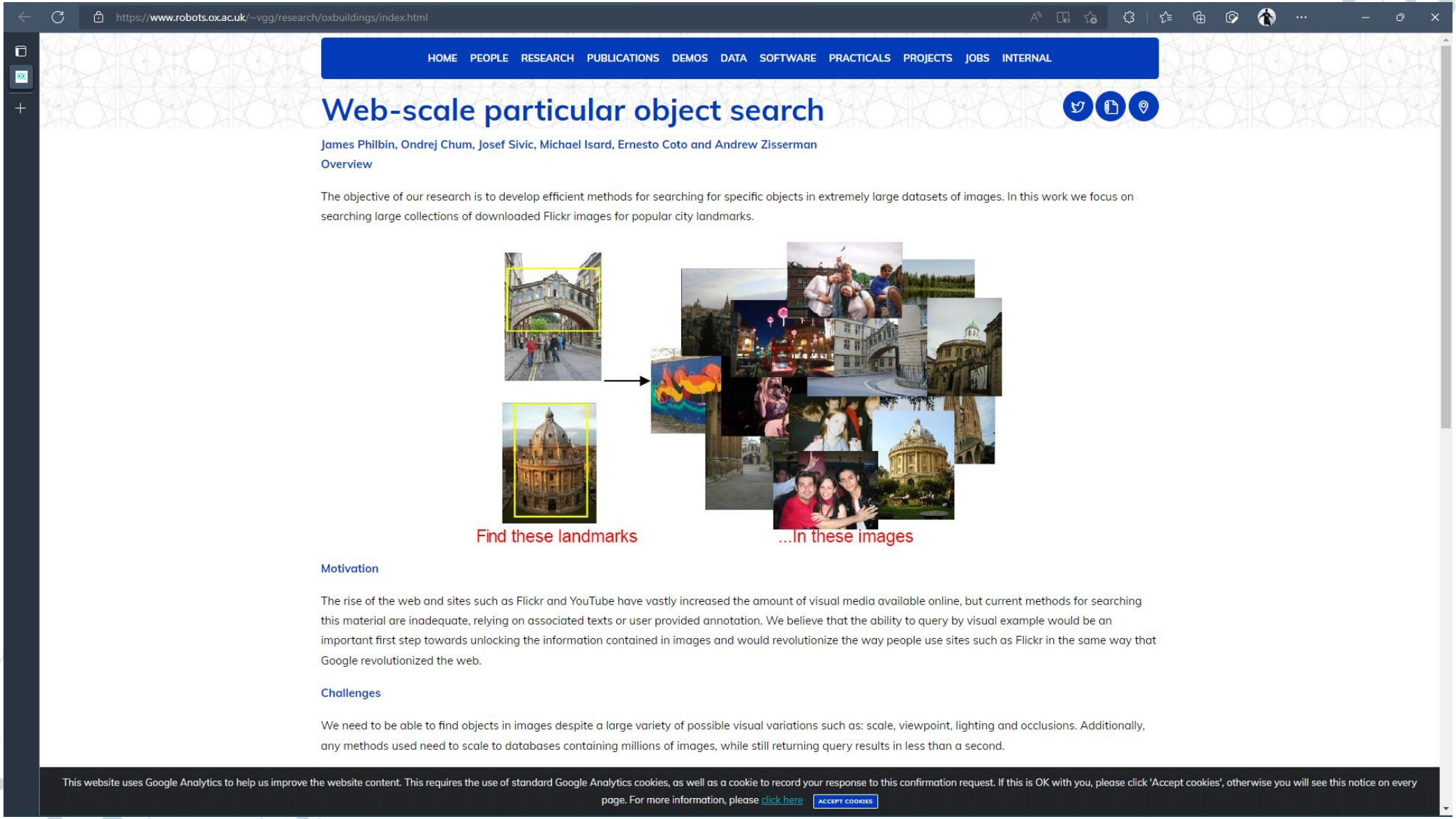
Khảo sát & phân tích

Image Retrieval | Papers With Code



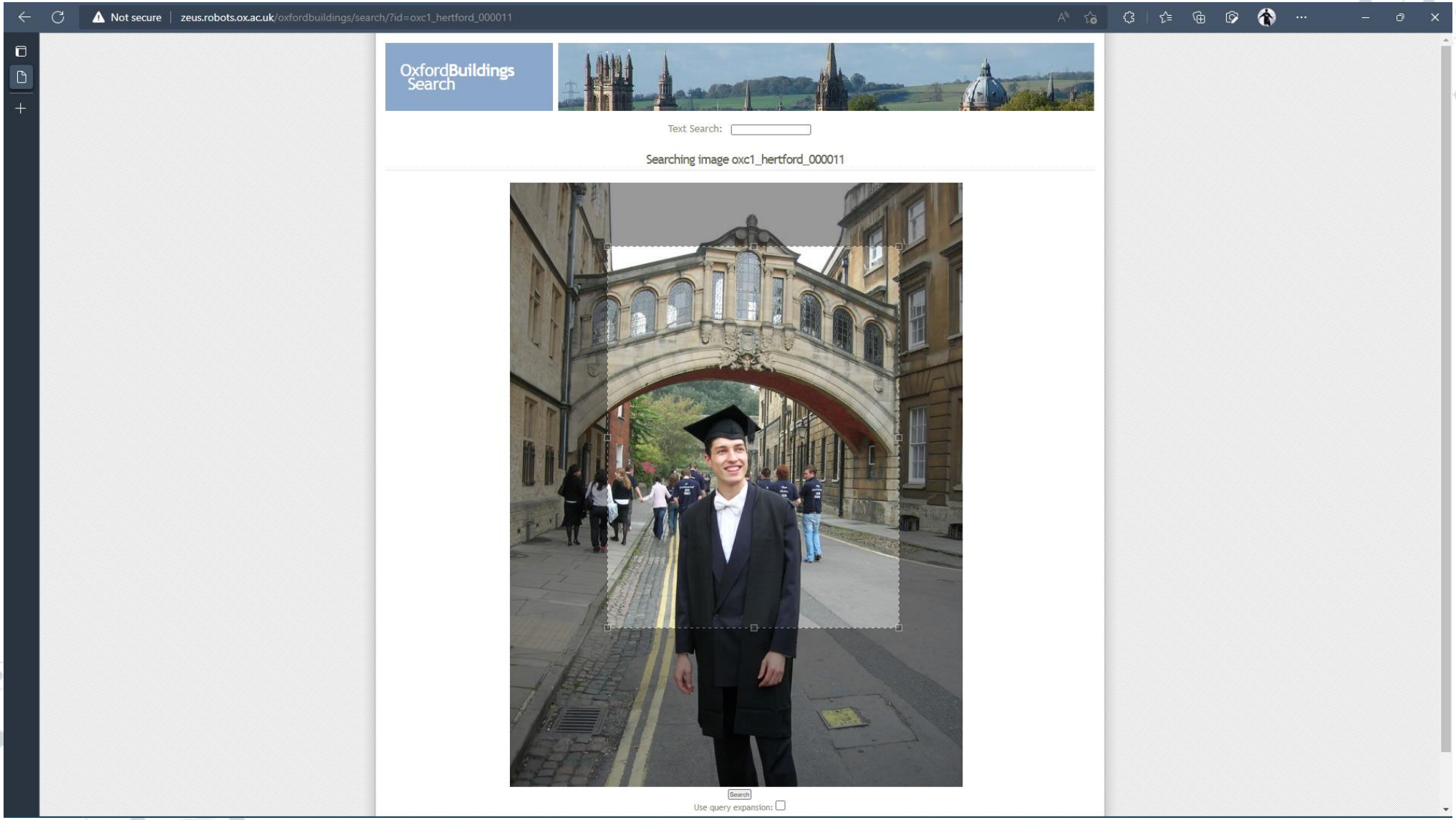
Khảo sát & phân tích

Visual Geometry Group - University of Oxford



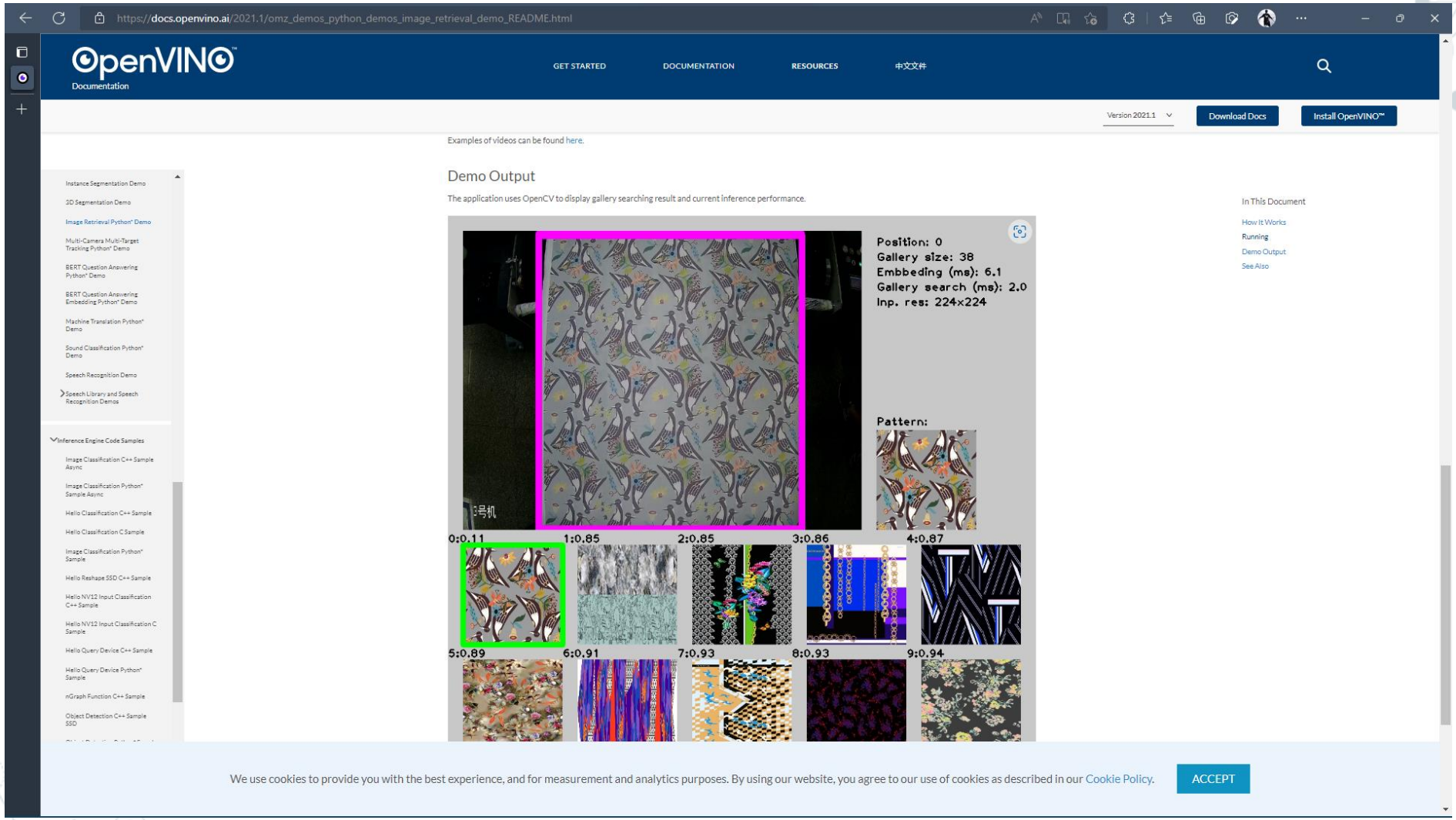
Khảo sát & phân tích

Oxford Buildings Search



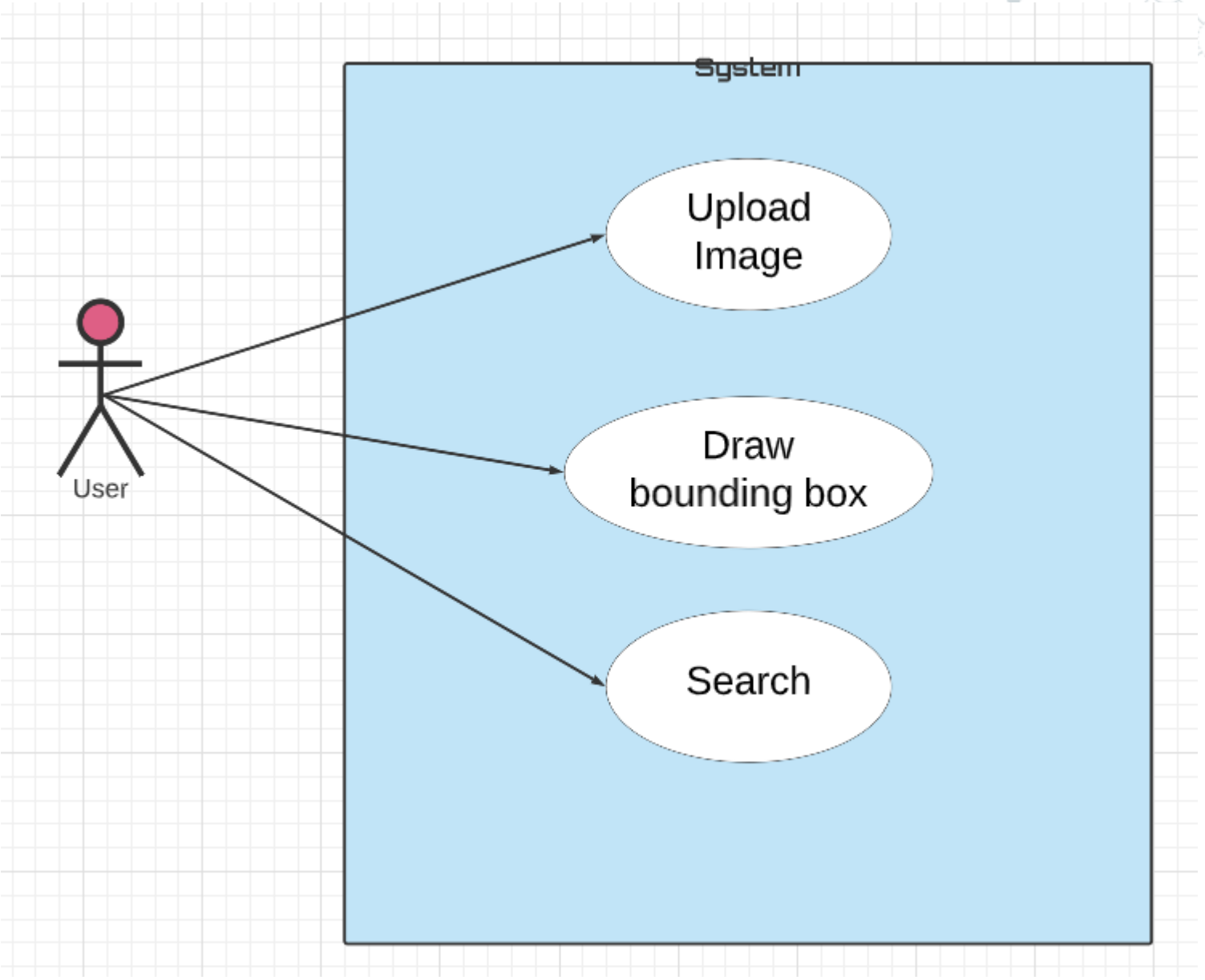
Khảo sát & phân tích

[Image Retrieval Python* Demo - OpenVINO™ Toolkit]

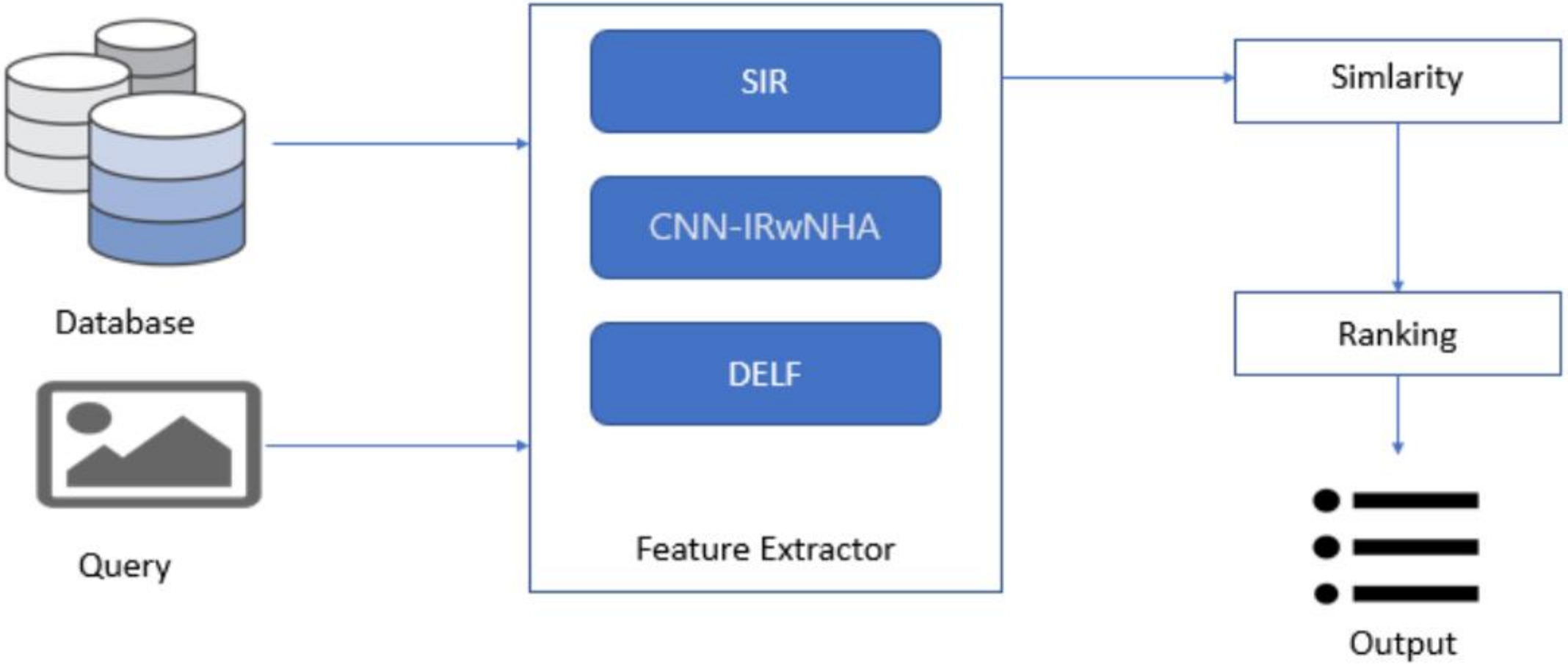


Khảo sát & phân tích

Yêu cầu người dùng

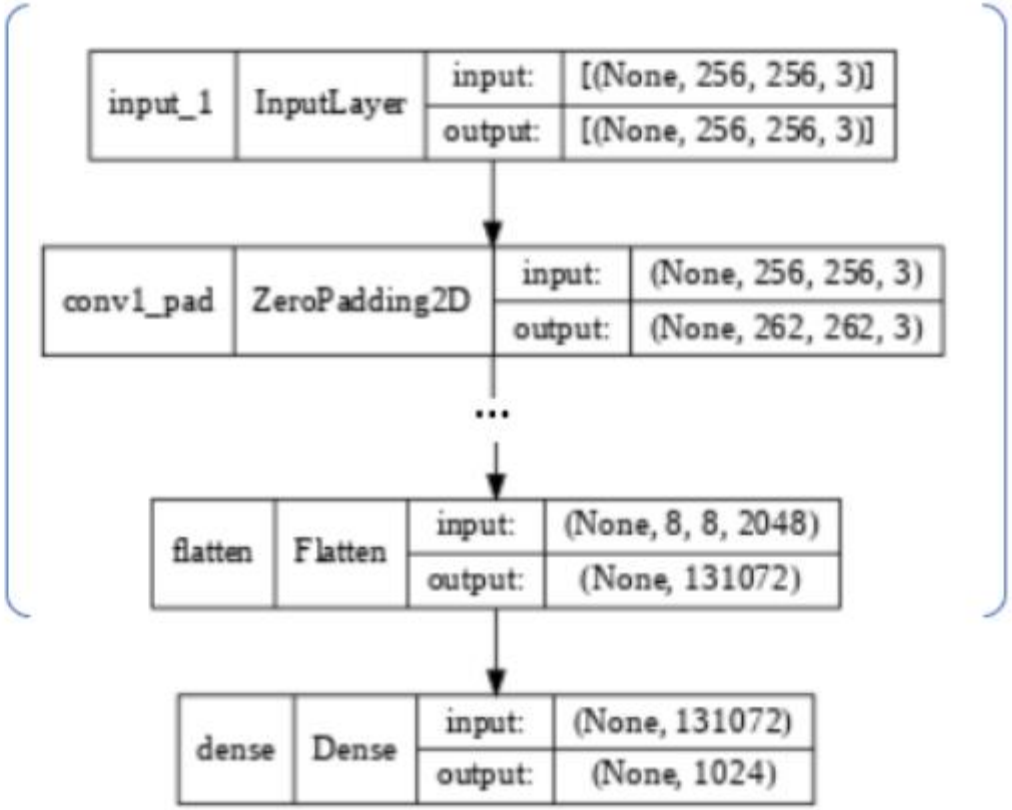


Xây dựng phương pháp: pipeline

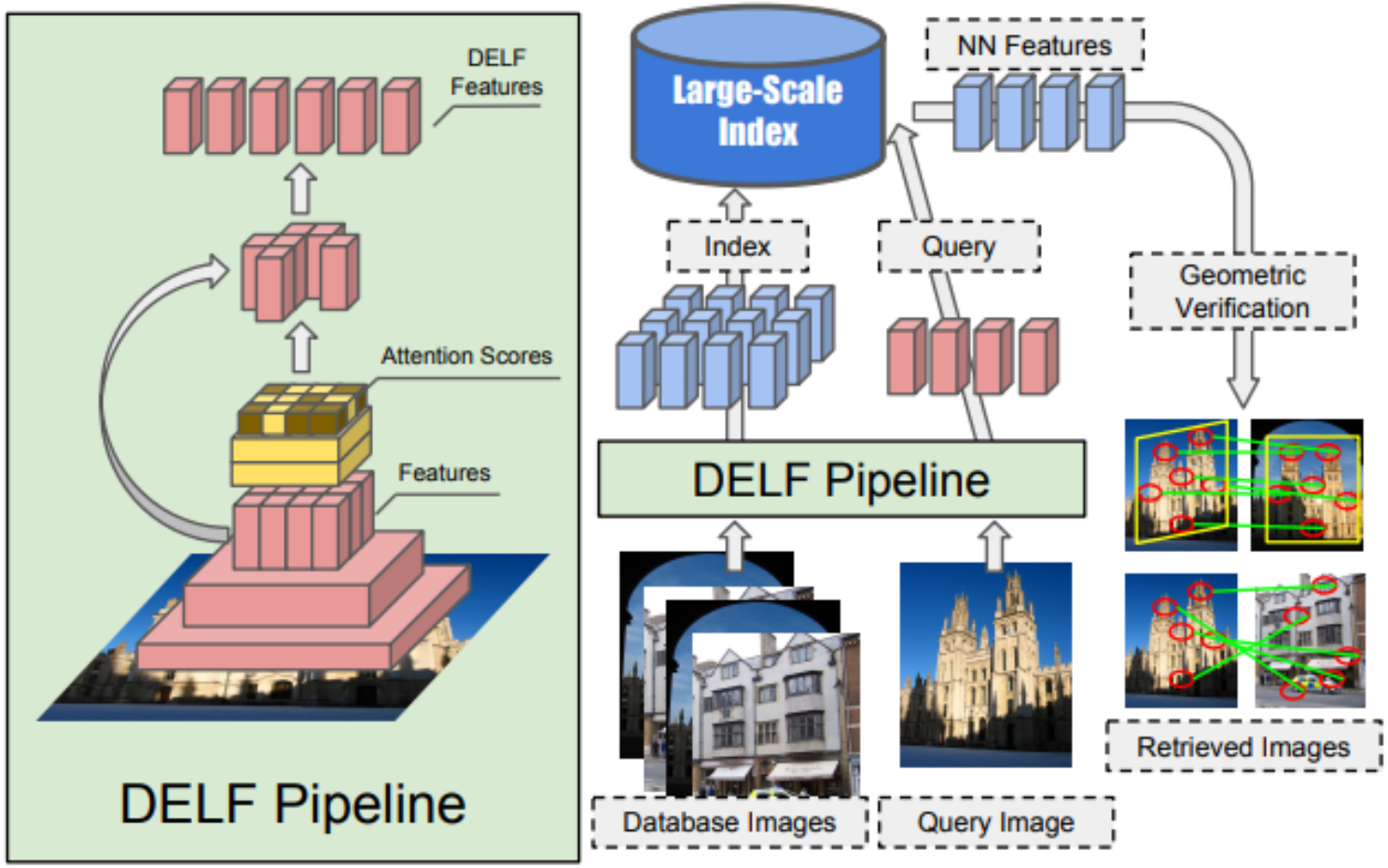


Simple Image Retrieval (SIR)

Resnet152

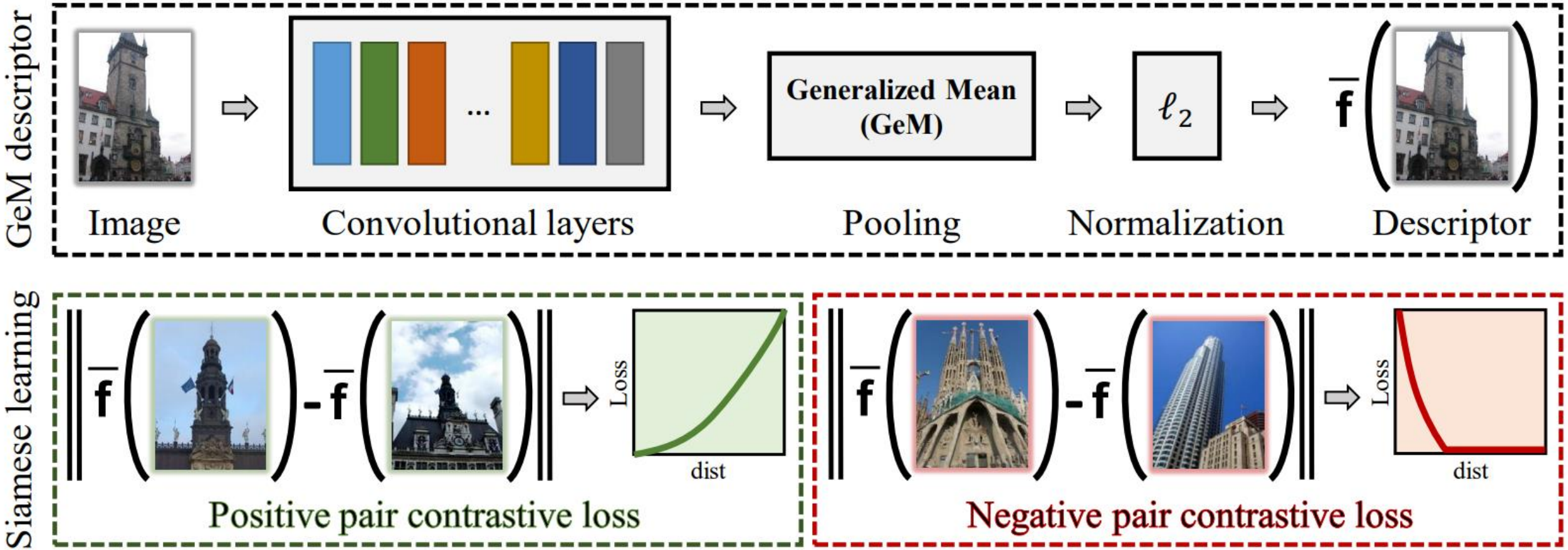


Deep Local Feature (DELF)



[1612.06321] Large-Scale Image Retrieval with Attentive Deep Local Features (arxiv.org)

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



[1711.02512] Fine-tuning CNN Image Retrieval with No Human Annotation (arxiv.org)

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
SIR	14.4	11.85	2.34	32.86	27.71	24.86
DELF	58.56	42.04	16.98	84.29	73.86	66.57
CNN-IRwNHA	85.08	68.65	44.24	92.86	92.29	86.

EVALUATION

Oxford 5k dataset:

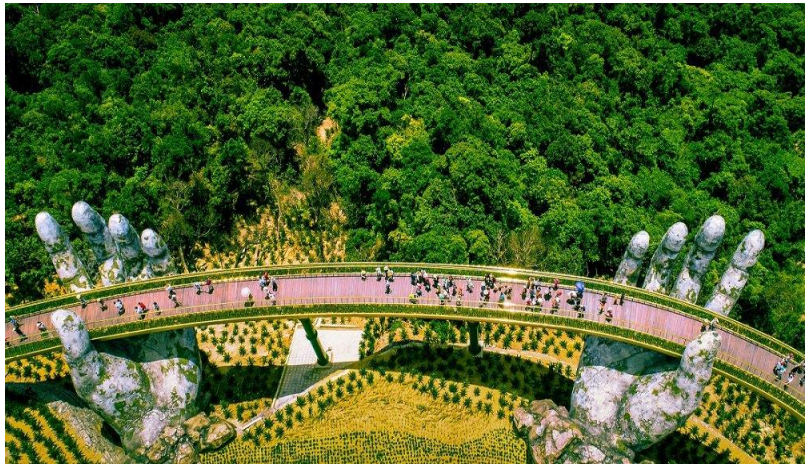
Method	map
SIR	23.34
DELF	66.69
CNN-IRwNHA	82.09

Vietnam Tourism Dataset

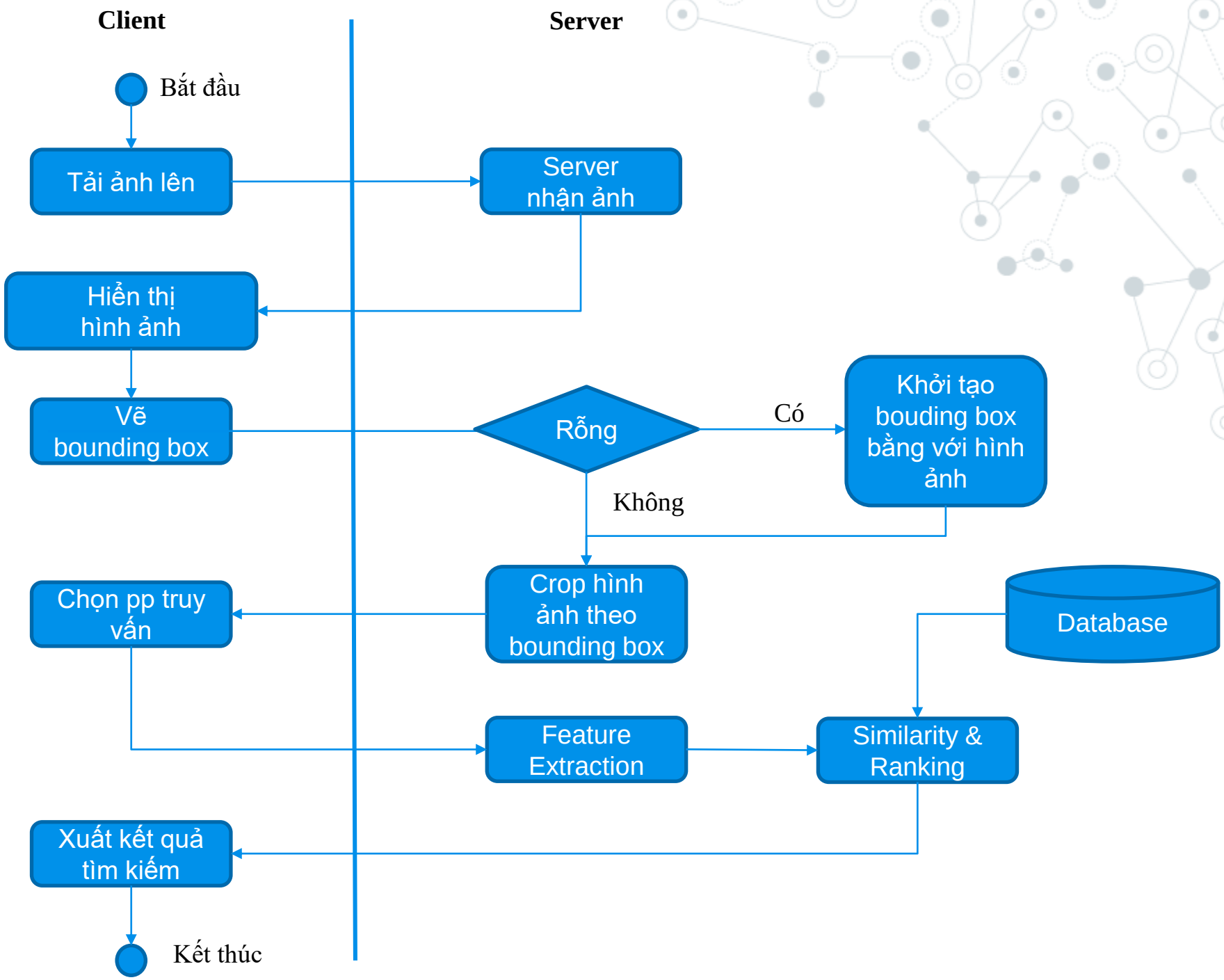






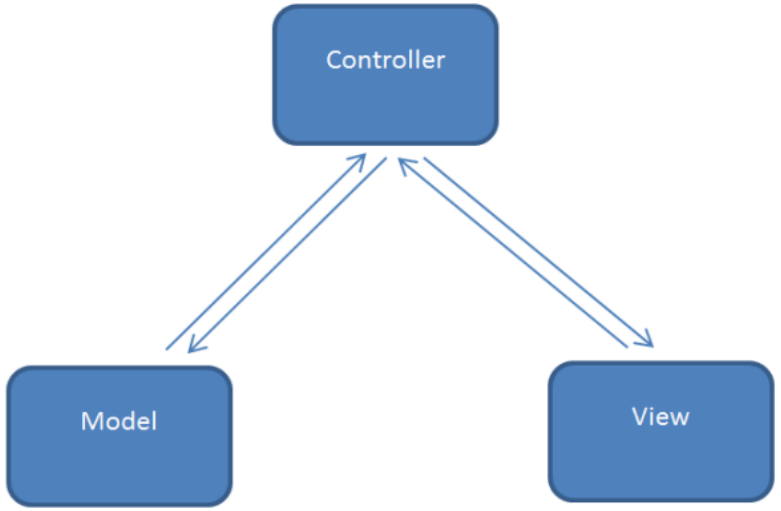



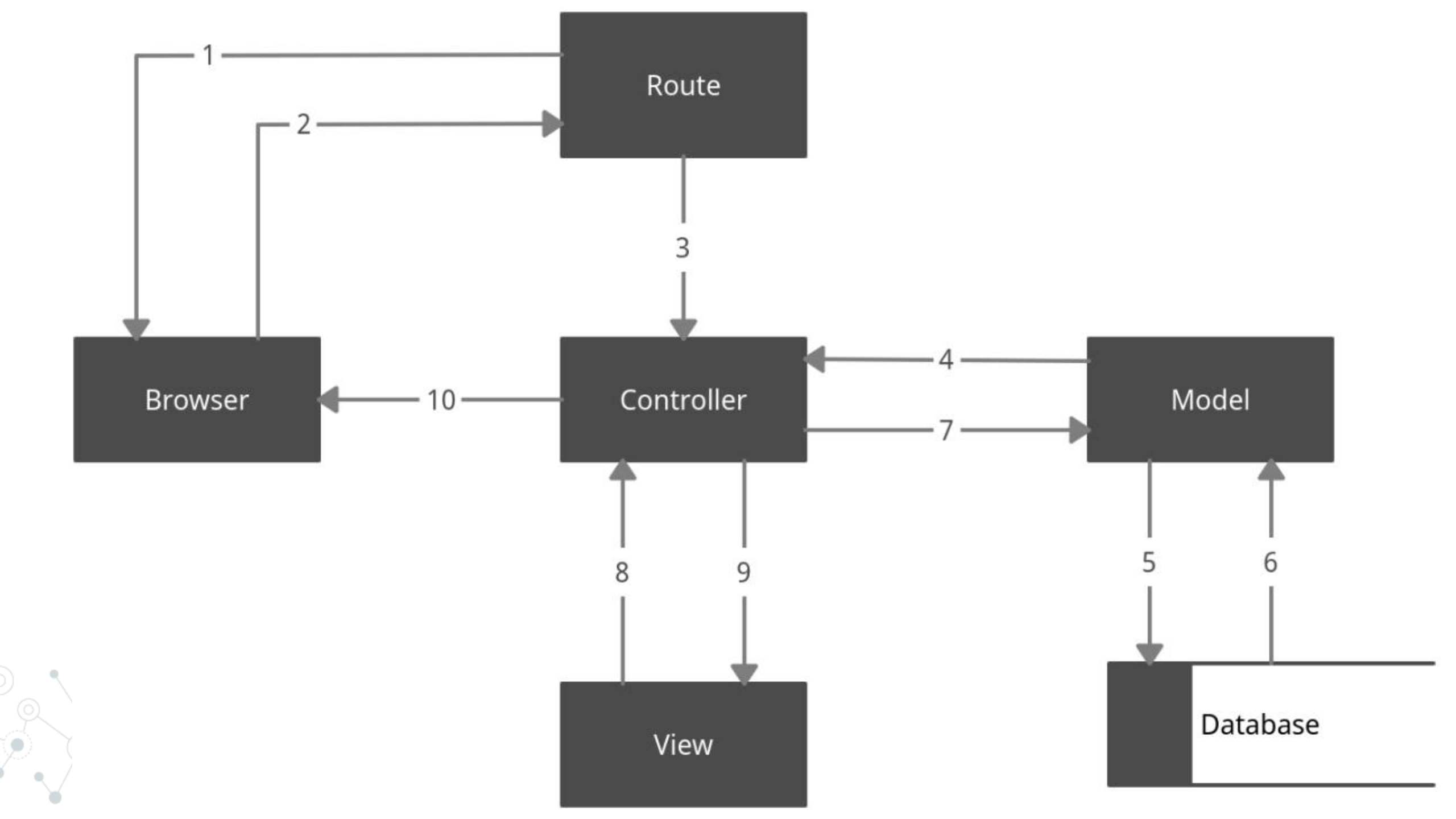

Activity Diagram



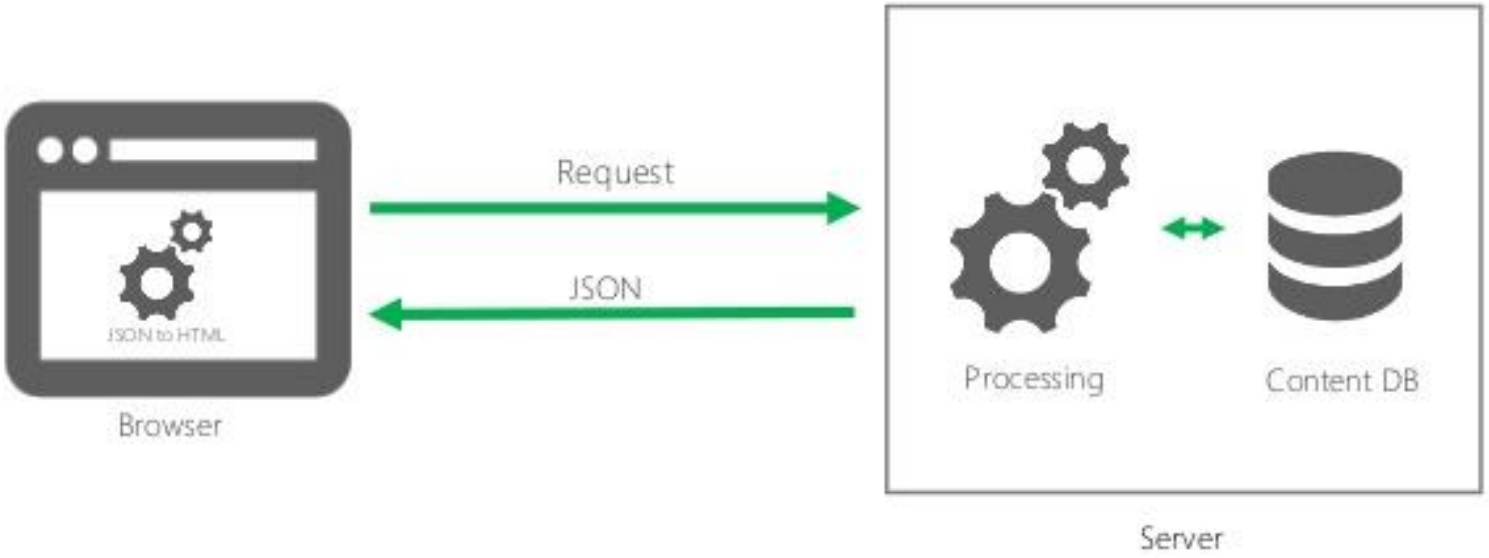
Kiến trúc

- 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.

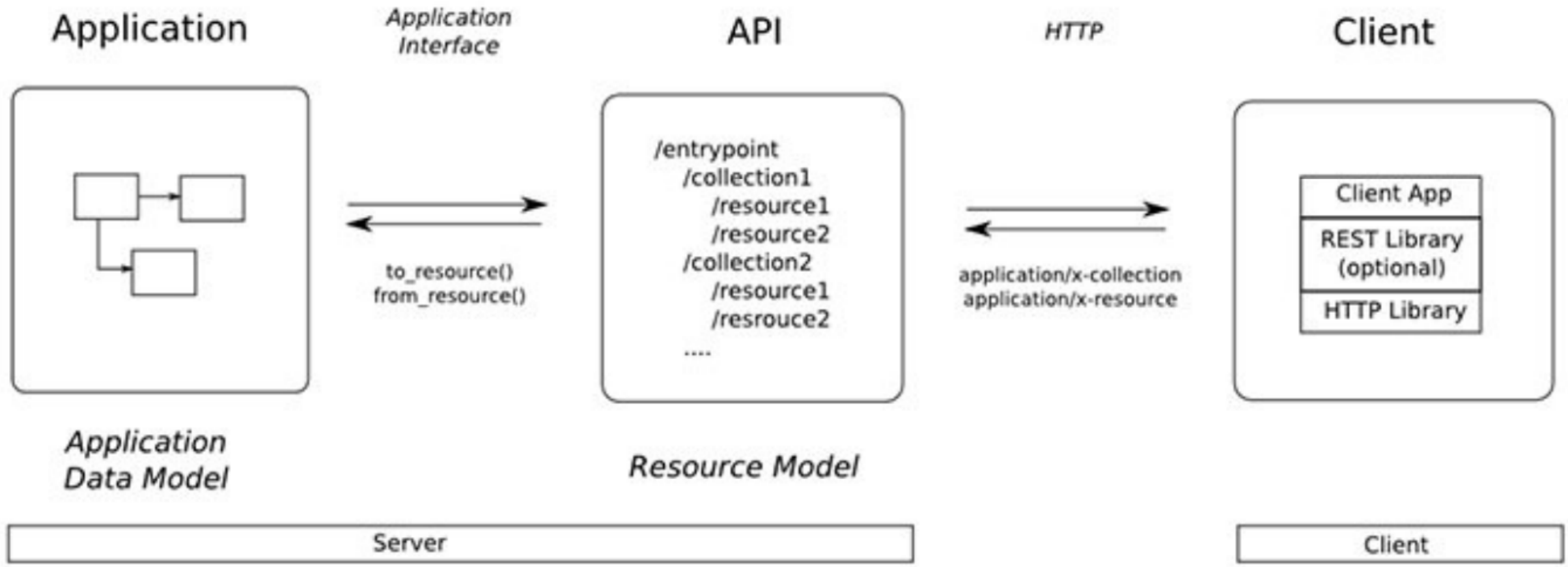



Route
1
2
3
Browser
10
Controller
4
7
Model
8
9
5
6
View
Database

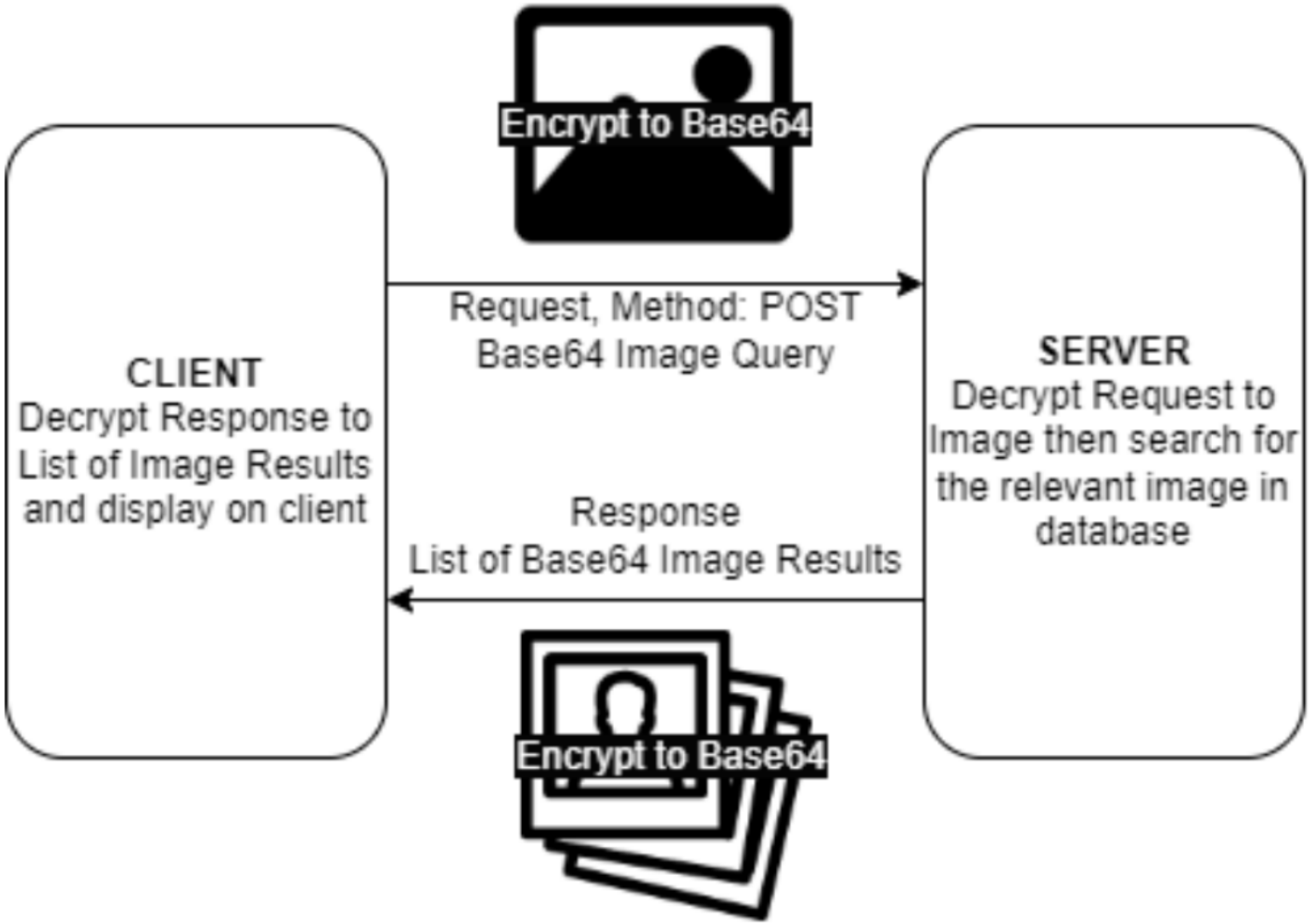
CLIENT-SIDE RENDERING



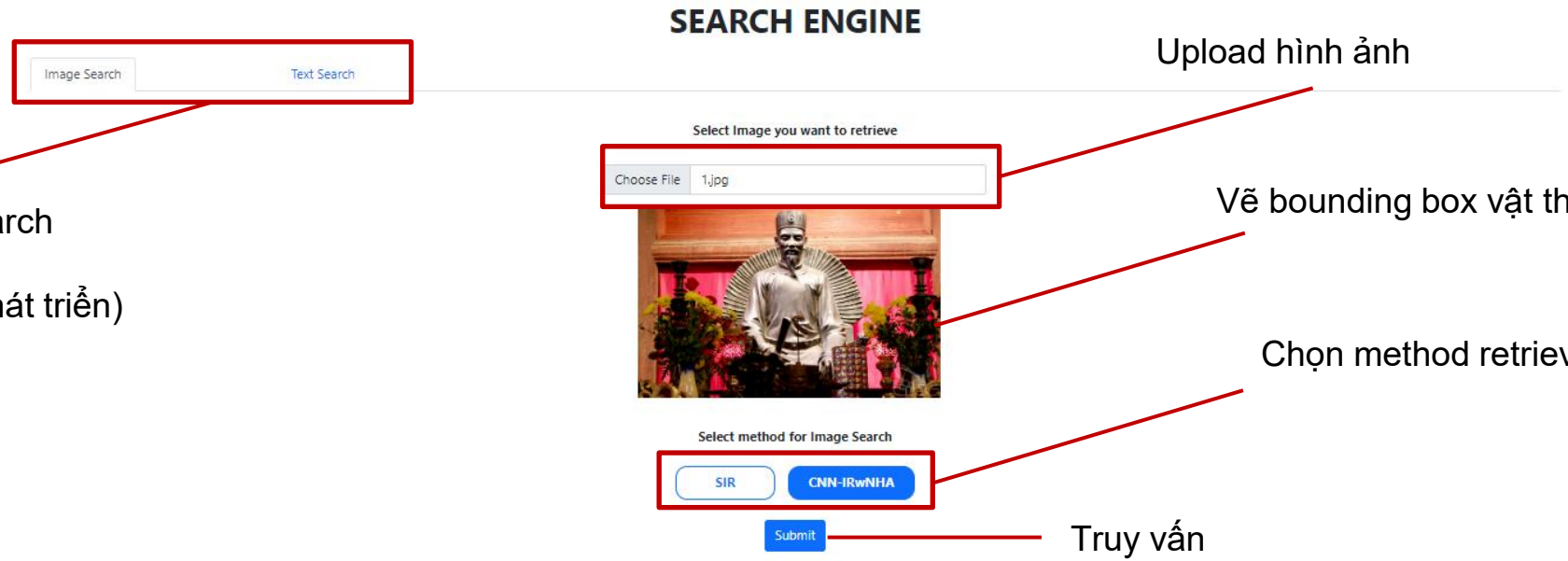
RESTFUL API


Application
Application Interface
API
HTTP
Client
/entrypoint
/collection1
/resource1
/resource2
/collection2
/resource1
/resrouce2
....
to_resource()
from_resource()
application/x-collection
application/x-resource
Client App
REST Library (optional)
HTTP Library
Application Data Model
Resource Model
Server
Client

RESTFUL API


Encrypt to Base64
CLIENT
Decrypt Response to List of Image Results and display on client
Request, Method: POST
Base64 Image Query
Response
List of Base64 Image Results
SERVER
Decrypt Request to Image then search for the relevant image in database
Encrypt to Base64

Thiết kế giao diện



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

Upload hình ảnh

Vẽ bounding box vật thể cần tìm kiếm

Chọn method retrieval

Truy vấn

Xuất kết quả

Conclusion and Demo

Thanks for watching