TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI
Viện Công nghệ thông tin & Truyền thông
----- oOo -----

BÁO CÁO

THỊ GIÁC MÁY TÍNH

Giáo viên hướng dẫn: TS. Nguyễn Thị Oanh

Họ và tên	MSSV	Lớp
Lê Khánh Duy	20160760	CNTT1.01-K61
Nguyễn Tiến Dũng	20160686	CNTT1.02-K61
Nguyễn Trọng Dương	20160851	CNTT1.01-K61
Vũ Minh Hào	20161272	CNTT1.02-K61
Nguyễn Bá Quân	20173319	KTMT.06-K62

Hà Nội, ngày 24 tháng 06 năm 2021.

TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI
Viện Công nghệ thông tin & Truyền thông
----- oOo -----

BÁO CÁO

THỊ GIÁC MÁY TÍNH

Giáo viên hướng dẫn: TS. Nguyễn Thị Oanh

Họ và tên	MSSV	Lớp
Lê Khánh Duy	20160760	CNTT1.01-K61
Nguyễn Tiến Dũng	20160686	CNTT1.02-K61
Nguyễn Trọng Dương	20160851	CNTT1.01-K61
Vũ Minh Hào	20161272	CNTT1.02-K61
Nguyễn Bá Quân	20173319	KTMT.06-K62

Hà Nội, ngày 24 tháng 06 năm 2021

Mục lục

Danh mục hình ảnh

Chương I: Truy xuất hình ảnh bằng đặc trưng toàn cục

Là việc truy xuất các hình ảnh tương tự từ tập dữ liệu bằng cách cung cấp hình ảnh dưới dạng truy vấn.

1. Bộ dữ liệu và thuật toán

1.1. Bộ dữ liệu

Coil-100

Bộ dữ liệu chứa 7200 ảnh màu của 100 đối tượng (72 ảnh cho mỗi đối tượng). Các đối tượng có nhiều đặc điểm hình học và phản xạ phức tạp. Các đối tượng được đặt trên một bàn xoay cơ giới trên nền đen. Bàn xoay được xoay 360 độ để thay đổi tư thế đối tượng đối với máy ảnh màu fxed. Hình ảnh của các đối tượng được chụp ở các khoảng cách tư thế là 5 độ, tương ứng với 72 tư thế cho mỗi đối tượng.

Hình 1: Một số đối tượng trong bộ dữ liệu coil-100

1.2. Thuật toán trích xuất tính năng

Một số kỹ thuật thị giác máy tính được sử dụng để trích xuất năng toàn cục như: Hist, hog, … Thuật toán học sâu được sử dụng như convAE.

Histogram

Dùng để lấy ra các đặc trưng về không gian màu toàn cục. Cho phép tìm kiếm các hình ảnh có màu sắc tương tự.

Hu Moment

Được sử dụng để trích xuất hình dạng của một đối tượng quan trọng từ hình ảnh.

Hog

Là một kỹ thuật phát hiện cạnh hoạt động trên nguyên tắc biểu đồ. Nó xác định sự thay đổi rõ nét về cường độ của các pixel và trả về hướng chuyển màu.

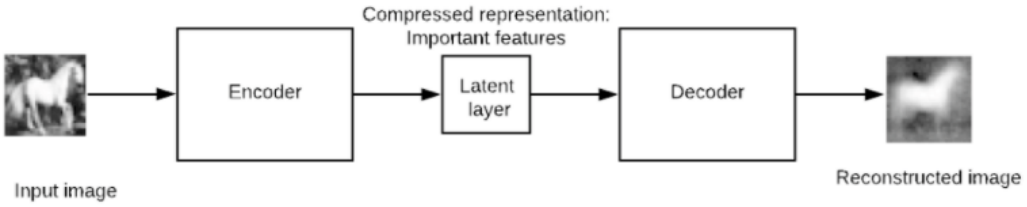
Haralick

Được sử dụng để định lượng một hình ảnh dựa trên kết cấu.

Convolutional Autoencoder (convAE)

Đây là mô hình học sâu không giám sát vì đầu vào của mạng chỉ cần ảnh chứ không cần nhãn. Giống với mạng autoencoder, mạng gồm một khối mã hóa và một khối giải mã nhưng bên nhánh mã hóa sử dụng các lớp tích chập để trích xuất tính năng và bên nhánh giải mã sử dụng các lớp giải tích chập để tái tạo lại hình ảnh.

Cơ chế hoạt động của convAE là buộc mạng phải mã hóa các đặc trưng quan trọng vào không gian latent có kích thước nhỏ.



Hình 2: Kiến trúc autoencoder

1.3. Thuật toán truy vấn

Vì không gian tìm kiếm cực kỳ lớn và truy vấn sẽ mất nhiều thời gian. Trong tìm kiếm, thời gian cũng là một hạn chế vì người dùng sẽ không muốn chờ đợi quá lâu, thà nhận được một số kết quả sai mà nhanh còn hơn là nhận được kết quả chính xác mà lâu. Để giảm thời gian tìm kiếm chúng ta cần giảm không gian tìm kiếm. Phân cụm là một cách tốt hơn để giảm không gian tìm kiếm.

Đầu tiên ta sẽ so sánh ảnh truy vấn với các cụm trung tâm, sau đó tìm kiếm các hình ảnh trong cụm tương tự nhất và các cụm còn lại được loại bỏ.

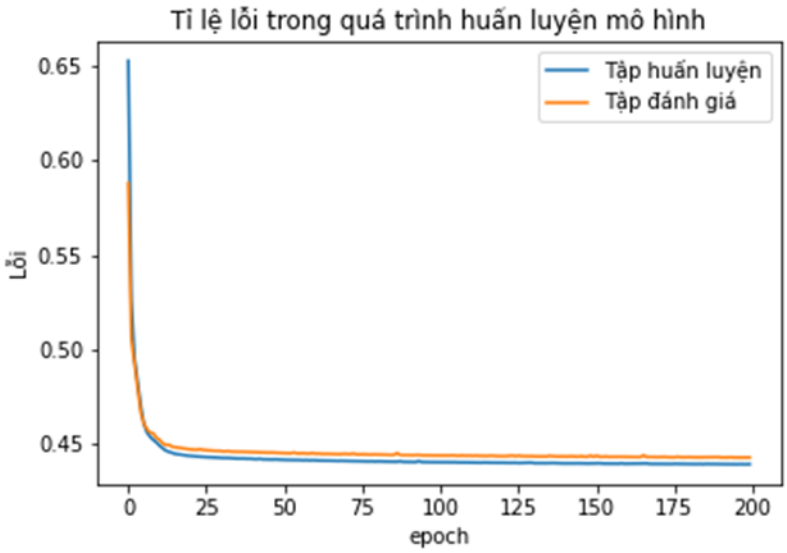
1.4. So sánh tính năng

Tính năng được trả về như là một véc tơ đặc trưng, một số độ đo có thể được sử dụng để so sánh độ tương đồng giữa 2 véc tơ như: Độ tương tự cosine, khoảng các Euclid

2. Đào tạo mô hình

Ảnh đầu vào có kích thước cố định, ta chỉ cần chuẩn hóa hình ảnh về khoảng (0, 1). Số mẫu đào tạo và số mẫu đánh giá được chia theo tỉ lệ 9:1.

Mạng được đào tạo với ảnh màu 3 kênh RGB. Thuật toán tối ưu adam với hàm lỗi được sử dụng là binary_crossentropy, số epoch là 200.

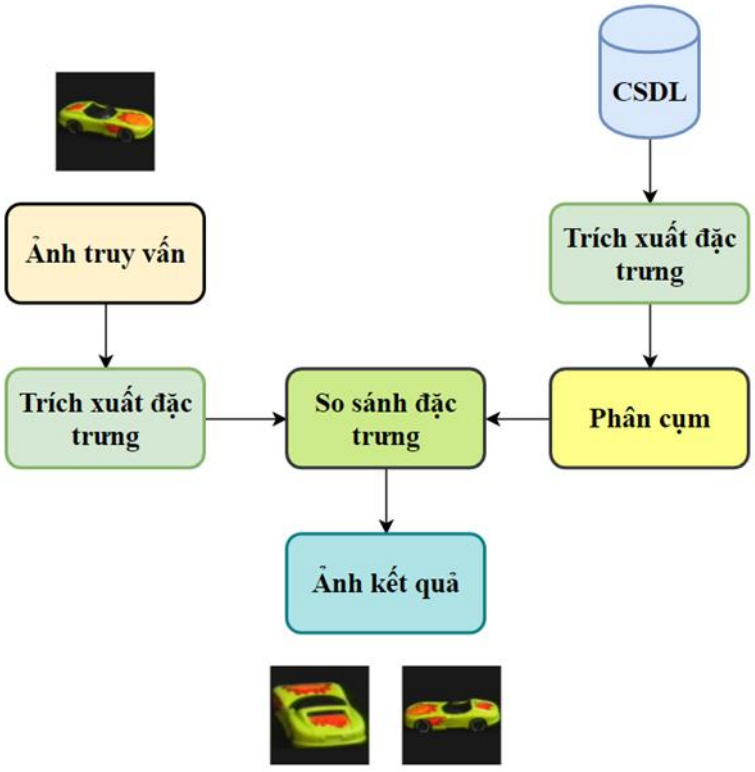


Hình 3: Tỉ lệ lỗi trong quá trình huấn luyện mô hình

Sau khi đào tạo xong mô hình, trọng số mô hình được lưu lại để dùng trong lúc truy vấn.

3. Truy vấn và kết quả

3.1. Kiến trúc tổng quan



Hình 4: Kiến trúc tổng quan

3.2. Đánh giá kết quả

Kết quả được đánh giá trên hai tiêu chí, đó là độ chính xác của kết quả trả về và thời gian truy vấn. Với độ chính xác của kết quả trả về ta đánh giá như sau, nếu kết quả trả về có nhãn giống với nhãn của hình ảnh truy vấn thì ta coi là kết quả đúng, nếu khác với nhãn của hình truy vấn thì là kết quả sai. Độ chính xác được tính là tổng số trường hợp đúng trên tổng số trường hợp.

3.3. Kết quả

Bảng 1 Kết quả thử nghiệm với các đặc trưng khác nhau

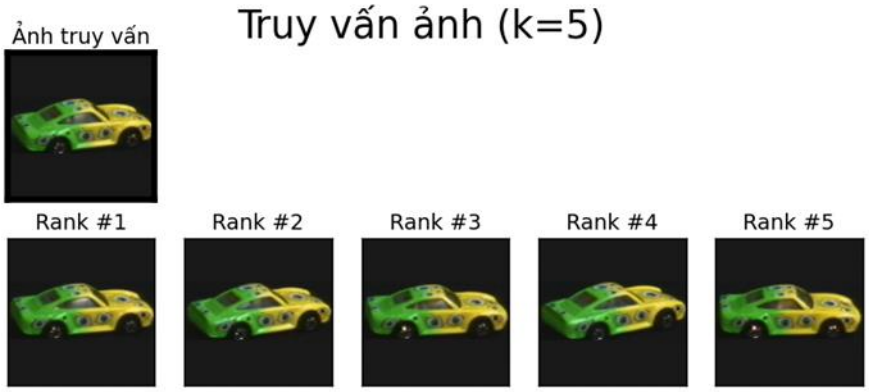
Tính năng / Số lượng truy vấn	HIST		Hu moment		HOG		convAE	
	Độ chính xác (%)	Thời gian truy vấn (s)	Độ chính xác (%)	Thời gian truy vấn (s)	Độ chính xác (%)	Thời gian truy vấn (s)	Độ chính xác (%)	Thời gian truy vấn (s)
1	98.0	3.27	52.0	0.17	98.0	135.61	**100.0**	11.62
5	95.0	3.77	42.6	0.17	89.2	134.98	**97.0**	9.07
10	89.8	3.80	38.5	0.15	77.8	129.91	**92.1**	9.30
20	86.0	3.79	34.7	0.15	69.8	132.95	**88.5**	8.68

3.4. Nhận xét, đánh giá

Với bộ dữ coil-100, ta thấy đặc trưng khi sử dụng mô hình convAE cho kết quả chính xác cao nhất, tiếp đến là đặc trưng hist, hog, hu moment. Ta thấy đặc trưng hist cho kết quả tương đối cao cho thấy các đối tượng trong bộ dữ liệu có sự khác biệt khá rõ rệt về màu sắc.

Thời gian truy vấn phụ thuộc vào độ dài của véc tơ đặc trưng, véc tơ càng dài, kích thước càng lớn thì thời gian tính toán càng lâu.

3.5. Kết quả một số truy vấn



Hình 5: Ví dụ truy vấn 1

Hình 6: Ví dụ truy vấn 2

4. Mã nguồn

Mã nguồn chương trình tại: https://github.com/dung98pt/image_retrieval đã bao gồm cách thức sử dụng và cài đặt.

Chương II: Phát hiện đối tượng chuyển động, theo vết đối tượng

1. Theo vết đối tượng

Object Tracking là bài toán theo dõi một hoặc nhiều đối tượng chuyển động theo thời gian trong một video. Đối tượng được xử lí không đơn giản là một hình ảnh mà là một chuỗi các hình ảnh: video. Theo vết đối tượng đảm bảo một số yếu tố:

- ID của các đối tượng luôn đảm bảo không đổi qua các frame.
- Khi đối tượng bị che khuất hoặc bị biến mất sau vài frame, hệ thống cần đảm bảo nhận diện lại được đúng ID khi đối tượng xuất hiện.
- Các vấn đề liên quan đến tốc độ xử lí để đảm bảo realtime và tính ứng dụng cao

2. Phân loại

Phân chia theo số lượng đối tượng cần theo dõi:

- **Single object tracking:** theo dõi một đối tượng duy nhất trong toàn bộ video. Và tất nhiên, để biết được cần theo dõi đối tượng nào, việc cung cấp 1 bouding box từ ban đầu là việc buộc có.
- **Multiple object tracking:** Cố gắng phát hiện đòng thời và theo dõi tất cả các đối tượng có trong tầm nhìn.

Phân chia theo phương pháp sử dụng số lượng frame (thời gian):

- **Online tracking:** chỉ sử dụng frame hiện tại và frame ngay trước để tracking.
- **Offline tracking:** thường sử dụng toàn bộ frame của video.

Phân chia theo phương pháp phát hiện đối tượng:

- **Detection based tracking:** tập trung và mối liên hệ chặt chẽ giữa object detection và object tracking, từ đó dựa vào kết quả của detect để theo dõi đối tượng qua các frame.
- **Detection free tracking:** coi video như 1 dạng dữ liệu chuỗi, từ đó áp dụng phương pháp xử lý dành cho chuỗi như RNN, LSTM, …

3. Metric đánh giá

Về metric đánh giá, chúng ta cần quan tâm các metric sau:

- **FP** (False Positive): tổng số lần xuất hiện một đối tượng được phát hiện mặc dù không có đối tượng nào tồn tại
- **FN** (False Negative): tổng số lần mà đối tượng hiện có không được phát hiện.
- **ID Switches**: tổng số lần 1 đối tượng bị gán cho 1 ID mới trong suốt quá trình tracking video.
- **MOTA**: Mutiple Object Tracking Accuracy:

$$MOTA = 1 - \frac{\sum_t FN_t + FP_t + id\_sw_t}{\sum_t GT_t}$$

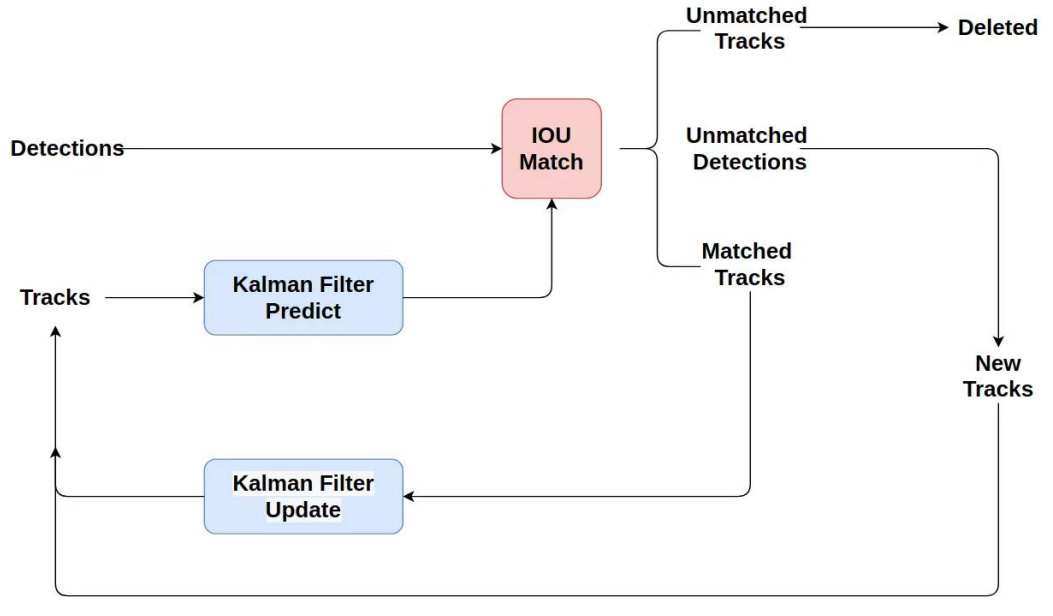
- **MOTP**: Mutiple Object Tracking Precision: là sự khác biệt trung bình giữa dự đoán đúng và mục tiêu đúng.

$$MOTP = \frac{\sum_{i,t} d_t^i}{\sum_t c_t}$$

- **MT** (Most Tracked Target): tính trong ít nhất 80% video.
- **ML** (Most Lost Target): tính trong 20% video
- **Hz (FPS)**: Tốc độ tracking

4. Xử lý bài toán

Nhóm thực hiện giải quyết bài toán trên phương pháp "detection based tracking" có thể phát hiện nhiều đối tượng cùng lúc, có thể chạy gần như realtime (phụ thuộc vào tốc độ xử lý của máy tính).



Hình 7: Luồng thực thi

Cụ thể khối detections sử dụng thuật toán YOLO – You Only Look Once, với đầu ra gồm các mảng chứa vị trí, nhãn và độ chắc chắn của đối tượng. Sau đó, khối tracking sử dụng thuật toán SORT (Simple Online Realtime Object Tracking) nhận đầu vào gồm vị trí và độ chắc chắn của đối tượng đã được dự đoán từ khối detections, thực hiện dự đoán vị trị mới của đối tượng dựa vào các frame trước, liên kết các vị trí detected với vị trí dự đoán để gán ID tương ứng và cho đầu ra vẫn gồm vị trí nhưng có thêm ID để định danh (theo vết) đối tượng.

5. Đánh giá khả năng theo vết

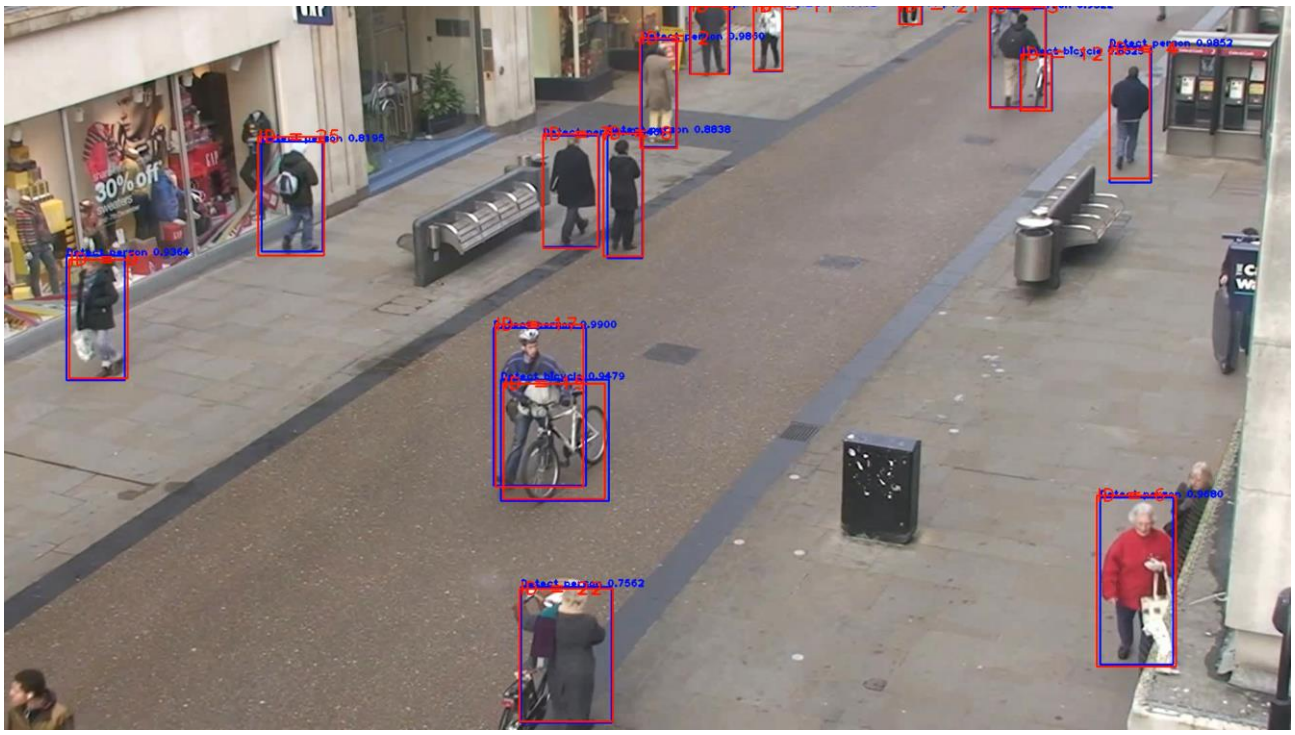
Tổng quan:

- Tốt: đối tượng chuyển động đều và luôn xuất hiện trong khung hình thì thuật toán luôn theo vết đúng đối tượng với mã định danh ID cố định.
- Tệ: đối tượng ra khỏi khung hình, bị che khuất, độ sáng thấp, … thì mặc dù vẫn có khả năng phát hiện đối tượng, nhưng khả năng theo vết đối tượng giảm đáng kể thể hiện ở mã định danh ID bị thay đổi.

Đối với SORT:

- Giả định tuyến tính: SORT đang sử dụng Linear Kalman Filter trong thuật toán cốt lõi, điều này trong thực tế là chưa phù hợp. Để cải thiện vấn đề này, chúng ta cần quan tâm đến các Kalman Filter phức tạp hơn, như Extended Kalman filter, Unscented Kalman filter, ...
- ID Switches: Đây là vấn đề lớn nhất của SORT hiện tại. Do việc liên kết giữa detection và track trong SORT chỉ đơn giản dựa trên độ đo IOU (tức SORT chỉ quan tâm đến hình dạng của đối tượng), điều này gây ra hiện tượng số lượng ID Switches của 1 đối tượng là vô cùng lớn khi đối tượng bị che khuất, khi quỹ đạo trùng lặp, ...

6. Kết quả



Hình 8: Kết quả theo về đối tượng (Khung xanh thể hiện kết quả phát hiện vật thể - gồm tên đối tượng và độ chắc chắn; khung đỏ thể hiện kết quả theo vết – ID đối tượng)

7. Mã nguồn

Mã nguồn chương trình tại: https://github.com/quannar178/python_yolov3_sort đã bao gồm cách thức sử dụng và cài đặt.

Tài liệu tham khảo

[1] Simple Online Realtime Object Tracking - Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, Ben Upcroft - 2016

[2] SORT - Deep SORT: Một góc nhìn về Object Tracking (https://viblo.asia/p/sort-deep-sort-mot-goc-nhin-ve-object-tracking-phan-1-Az45bPooZxY)

[3] YOLOv3: An Incremental Improvement - Joseph Redmon, Ali Farhadi - 2018