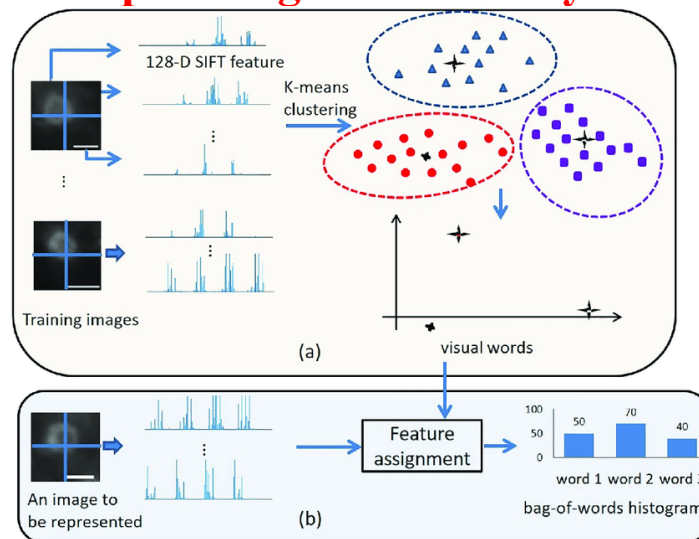


ĐẠI HỌC QUỐC GIA TP.HCM
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN



BÀI BÁO CÁO
Build Simple Image Retrieval System Part2



Sinh viên thực hiện:
Trịnh Đức Thiên - 20127330

2022-2023

Mục lục

I.	MÔ HÌNH VISUAL BAG OF WORD	3
II.	CÁCH TÍNH PRECISION, RECALL, F1 SCORE.....	4
1.	BỘ DỮ LIỆU.....	4
2.	PRECISION.....	5
3.	RECALL	5
4.	TƯƠNG QUAN GIỮA PRECISION VÀ RECALL	5
5.	F1 SCORE	5
III.	HƯỚNG DẪN CHẠY CODE.....	6
IV.	TÀI LIỆU THAM KHẢO	7

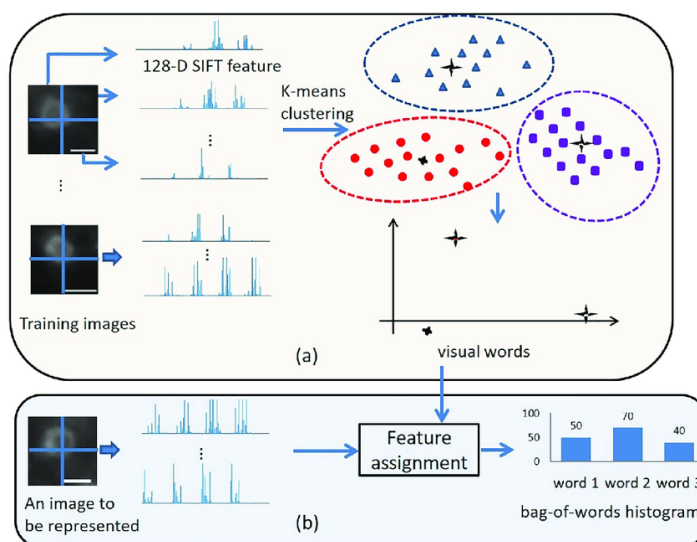
I. Mô hình Visual Bag of Word

Visual Bag of Words (BoW) là một phương pháp được sử dụng trong xử lý ảnh và phân loại hình ảnh dựa trên tần suất xuất hiện của các đặc trưng hình ảnh trong tập dữ liệu.

Quá trình xây dựng Visual BoW bao gồm các bước sau:

1. Trích xuất đặc trưng: Sử dụng các phương pháp như SIFT hoặc SURF, các đặc trưng hình ảnh được trích xuất từ tập dữ liệu. Điều này có thể làm bằng cách tìm các điểm đặc trưng quan trọng trong ảnh và mô tả chúng thành các vector đặc trưng.
2. Xác định từ vựng: Từ các đặc trưng được trích xuất, tập từ vựng được xây dựng bằng cách áp dụng các thuật toán như K-means clustering để phân cụm các đặc trưng thành các nhóm. Mỗi cụm sẽ đại diện cho một từ trong từ vựng.
3. Biểu diễn hình ảnh: Mỗi ảnh được biểu diễn dưới dạng vector BoW. Quá trình này bắt đầu bằng việc xác định các đặc trưng trong ảnh và gán chúng vào từng cụm trong từ vựng. Sau đó, vector BoW được tạo ra bằng cách đếm số lượng đặc trưng thuộc từng cụm.
4. Phân loại hình ảnh: Vector BoW của mỗi ảnh được sử dụng như đặc trưng đầu vào cho các mô hình học máy, như Support Vector Machines (SVM) hoặc Random Forest, để phân loại hình ảnh.

Visual BoW đã được áp dụng thành công trong nhiều ứng dụng như nhận dạng đối tượng, phân loại hình ảnh và gợi ý hình ảnh. Tuy nhiên, nó cũng có một số hạn chế, bao gồm mất mát thông tin không gian và thứ tự trong hình ảnh ban đầu. Các phương pháp nâng cao như Spatial Pyramid Matching (SPM) có thể được sử dụng để cải thiện hiệu suất của Visual BoW bằng cách xem xét cấu trúc không gian của hình ảnh.



II. Cách tính Precision, Recall, F1 Score

1. Bộ dữ liệu

Giả định rằng chúng ta đang xây dựng một mô hình phân loại nợ xấu. Nhãn của các quan sát sẽ bao gồm GOOD (thông thường) và BAD (nợ xấu). Kích thước của các tập dữ liệu như sau:

- Tập train: 1000 hồ sơ bao gồm 900 hồ sơ GOOD và 100 hồ sơ BAD.
- tập test: 100 hồ sơ bao gồm 85 hồ sơ GOOD và 15 hồ sơ BAD.

Predict / Actual		Actual	
		Positive	Negative
Predict	Positive	55 TP – True Positive	50 FP – False Positive
	Negative	45 FN – False Negative	850 TN – True Negative
Total		100	900

Các chỉ số TP, FP, TN, FN lần lượt có ý nghĩa là :

- TP (True Positive): Tổng số trường hợp dự báo khớp Positive.
- TN (True Negative): Tổng số trường hợp dự báo khớp Negative.
- FP (False Positive): Tổng số trường hợp dự báo các quan sát thuộc nhãn Negative thành Positive.
- FN (False Negative): Tổng số trường hợp dự báo các quan sát thuộc nhãn Positive thành Negative.

2. Precision

$$Precision = \frac{TP}{Total\ predicted\ positive} = \frac{TP}{TP + FP} = \frac{55}{55 + 50} = 52.4\%$$

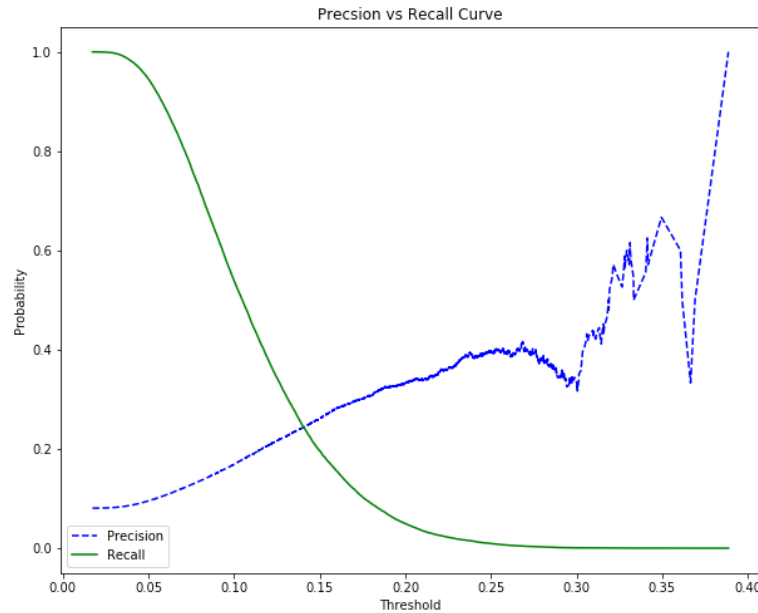
3. Recall

$$Recall = \frac{TP}{Total\ actual\ positive} = \frac{TP}{TP + FN} = \frac{55}{55 + 45} = 55\%$$

4. Tương quan giữa Precision và Recall

Thông thường các model sẽ lựa chọn một ngưỡng mặc định là 0.5 để quyết định nhãn. Nếu ta có một hàm f_0 thì nhãn dự báo dựa trên độ lớn xác suất:

$$\begin{cases} f_0(x) \geq 0.5, label = 1 \\ f_0(x) < 0.5, label = 0 \end{cases}$$



5. F1 Score

$$F_1 = \frac{2}{precision^{-1} + recall^{-1}} = \frac{2}{0.524^{-1} + 0.55^{-1}} = 53.7\%$$

Trong trường hợp $Precision = 0$ hoặc $Recall = 0$ ta qui ước $F_1 = 0$

Ta chứng minh được giá trị của F_1 score luôn nằm trong khoảng của precision và recall

$$F_1 = \frac{(2precision \times recall)}{(precision + recall)} \leq \frac{2 precision \times recall}{(2 \min(precision, recall))} = \max(precision, recall)$$

Tương tự:

$$F_1 = \frac{(2 \text{ precision} \times \text{recall})}{(\text{precision} + \text{recall})} \geq \frac{2 \text{ precision} \times \text{recall}}{(2 \max(\text{precision}, \text{recall}))} = \max(\text{precision}, \text{recall})$$

Ta có kết quả:

Predict / Actual		Actual		
		Positive	Negative	
Predict	Positive	55 TP – True Positive	50 FP – False Positive	Precision = 55/(55+5) = 91,6%
	Negative	45 FN – False Negative	850 TN – True Negative	
Total		100	900	
		recall = 55/(55+45)=55%		

III. Hướng dẫn chạy code

Đầu tiên, ta sử dụng lệnh:

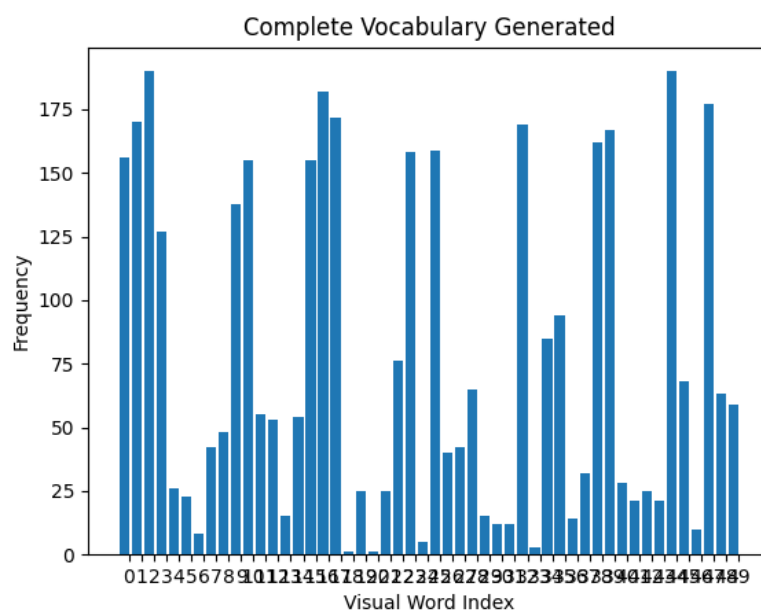
cd Code Files

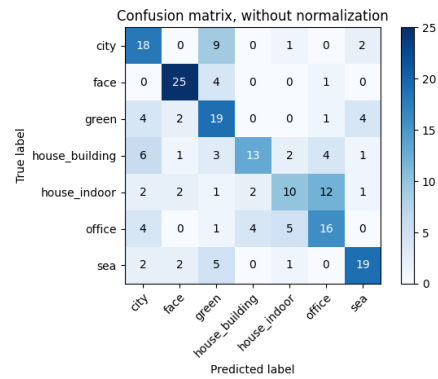
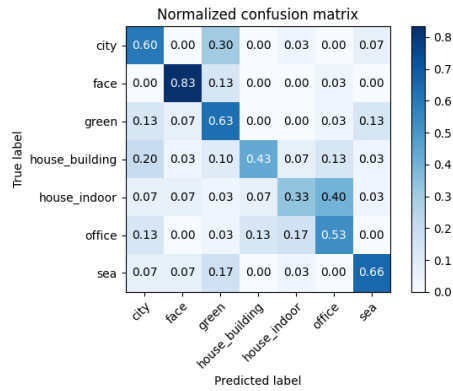
Sau khi đã truy cập vào files ta tiếp tục câu lệnh

python3 BoW.py

Để chạy chương trình

Chương trình sẽ đọc dữ liệu từ tập train và cho ra kết quả như hình:





IV. Tài liệu tham khảo

[1] [Bài 46 - Đánh giá mô hình phân loại trong ML](#)