High level Content-based image retrieval

Silvério Pereira, 76505
Pedro Santos, 76532

Introduction

Development of an application to do image retrieval by extracting high level information from the images.

The application is divided in two parts:

- Indexing set of images
- Search by image /test to get the results
- Rank the results

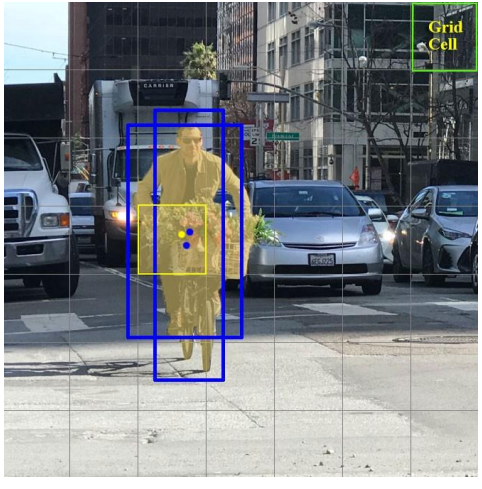
What do we use to score images?

- Object detection
- Visual Saliency
- Face recognition
- Object Character recognition

Yolov3

YOLO (You Only Look Once) is a state of the art object recognition algorithm.

- Divides the image in grid cells and applies a CNN classifier to each, each grid cell has a maximum of objects it can predict.

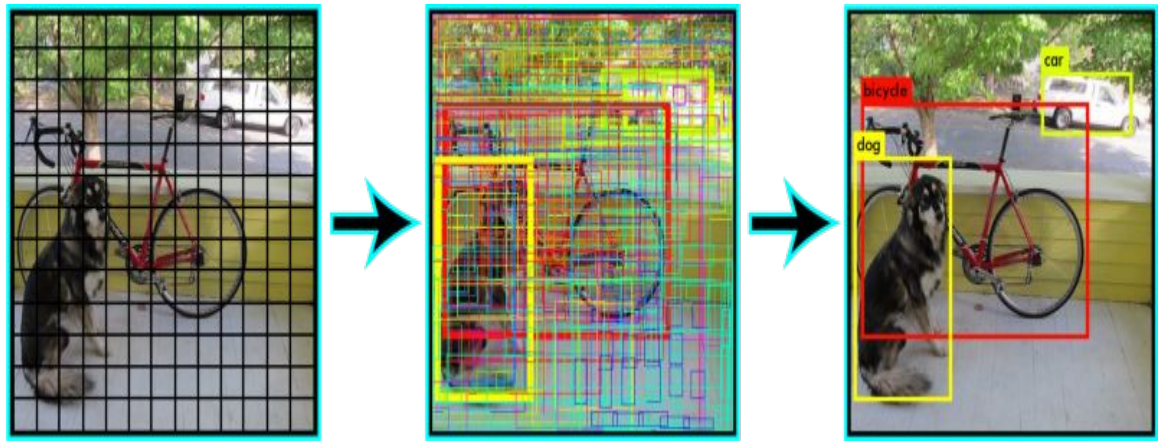


$$\hat{y} = \begin{bmatrix} P_c \\ b_x \\ b_y \\ b_y \\ c_1 \\ c_2 \\ \dots \\ c_n \\ P_c \\ b_x \\ b_y \\ b_y \\ c_1 \\ c_2 \\ \dots \\ c_n \end{bmatrix}$$

1 anchor box

Non-max suppression

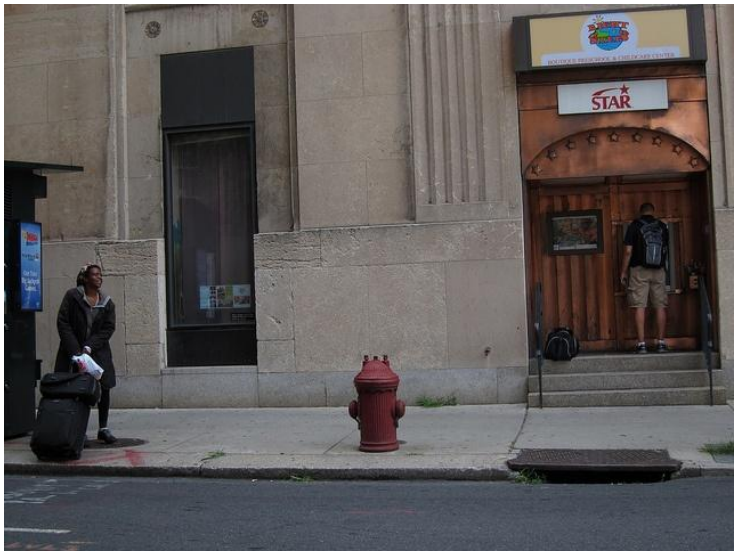
$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

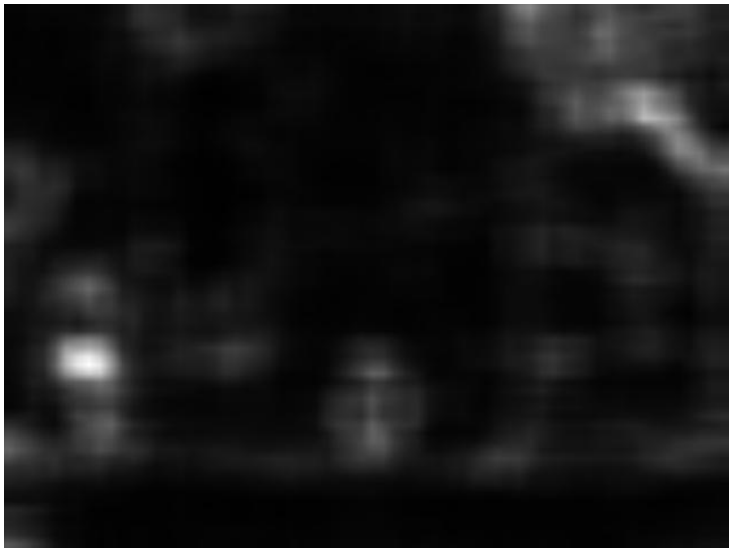


Tunable parameters

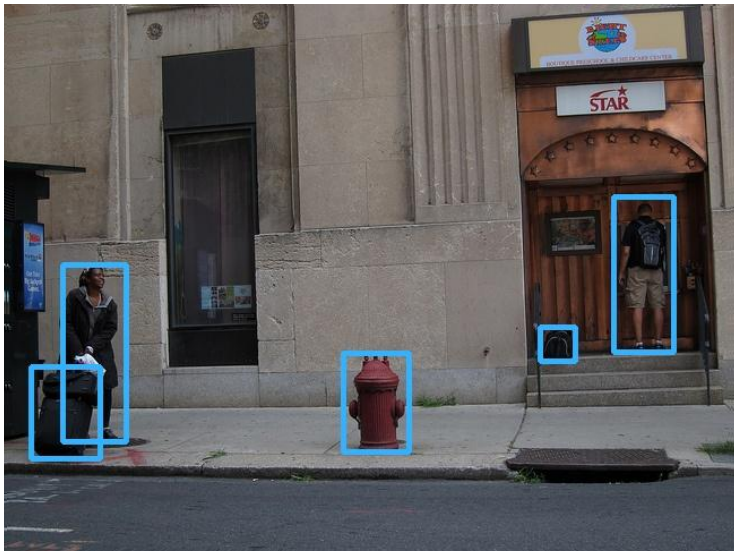
- Confidence threshold
- Non-max suppression threshold (IoU)
- Max number of objects in a cell grid

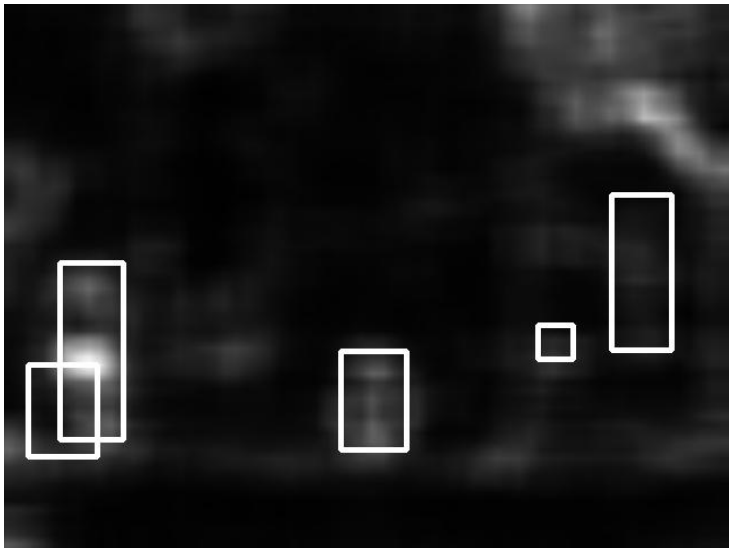
Visual Saliency



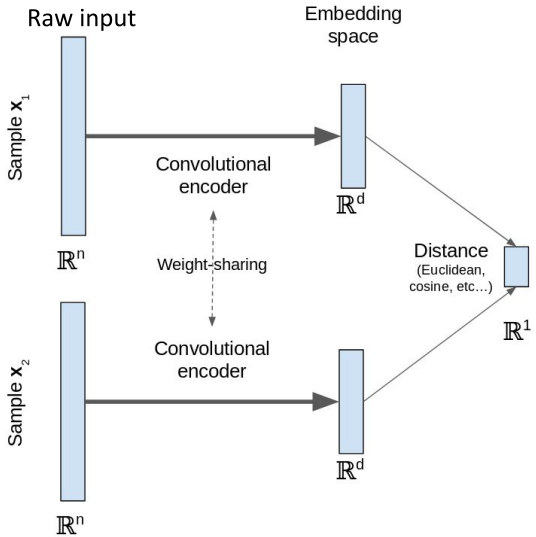


Relevance Calculation



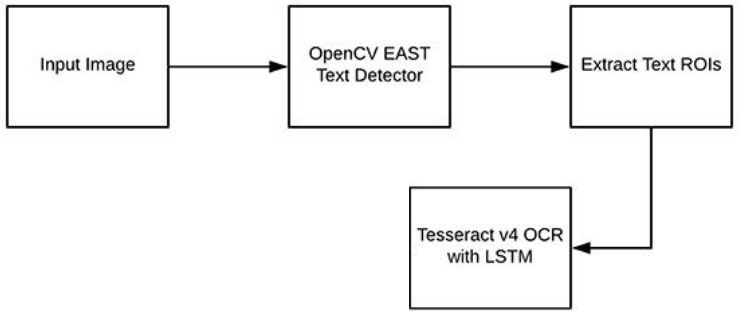


One Shoot Face Recognition

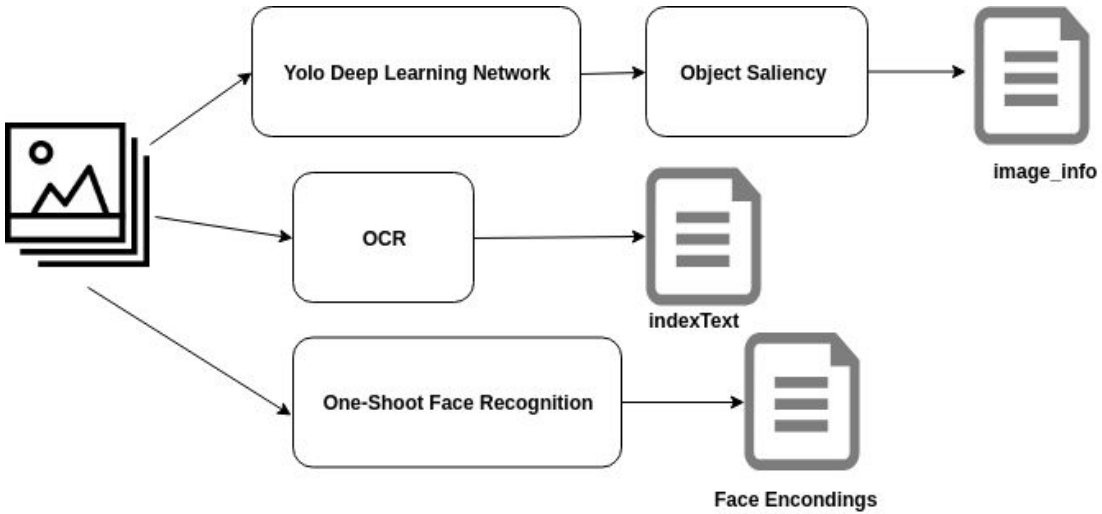


OCR (optical character recognition)

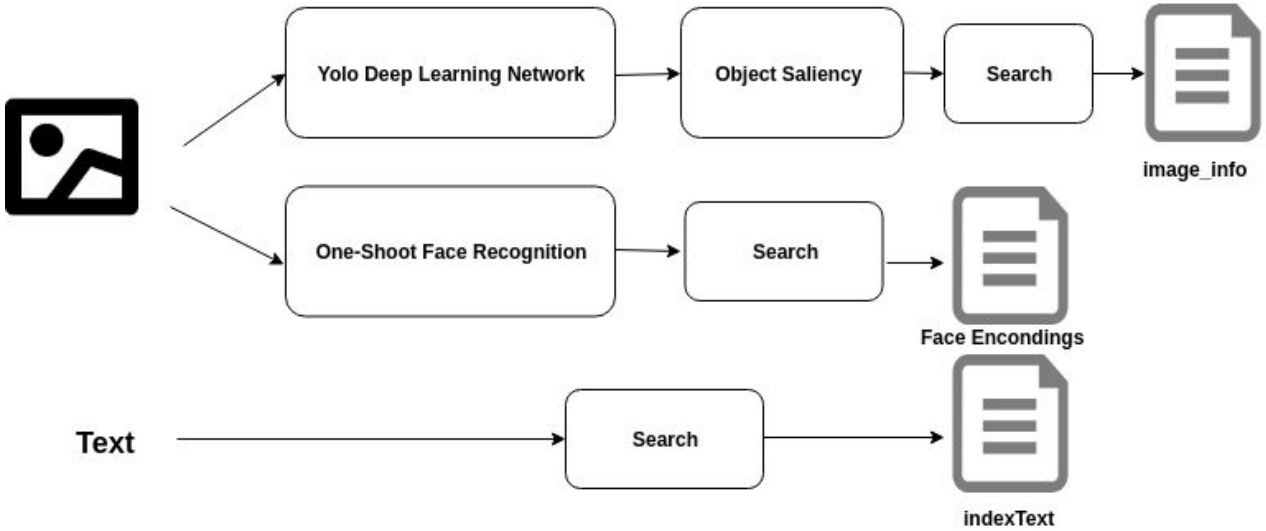
- The function verifyText() is called and it returns a list with the words detected in the given image;
- Verification if it was detected any text in the image;
- Removing Special Characters;
- Performing Stemming;



Indexing Data flow



Search Data flow



Ranking

TABLE I
OBJECT WEIGHT CALCULATION

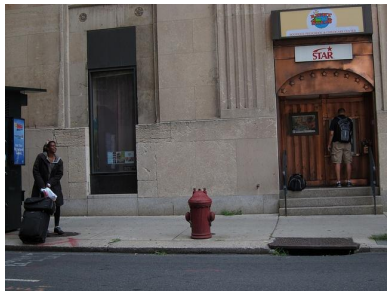
y	$\hat{y}$	cost
0	0	0
0	> 1	$log(1 + \hat{y}) * e^{rel(\hat{y})}$
> 1	0	$1 + log(y) * e^{rel(y)}$
> 1	> 1	$\lvert log(y) - log(\hat{y})\rvert * e^{\lvert rel(y) - rel(\hat{y})\rvert}$

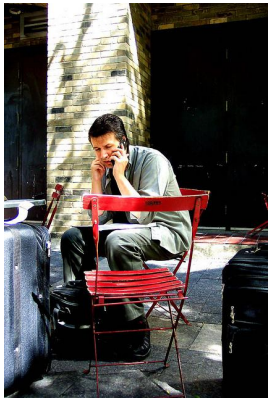
TABLE II
FACE DISCOUNT CALCULATION

$d <= 6$	-10
$0.6 < d < 1$	$min(log((d - 0.6) * 2.5)), -10)$
$d > 1$	0

Results

Query





Results

Query: “nokia”

Query